Analysis of Head Pose Accuracy in Augmented Reality

William Hoff, *Member*, *IEEE*, and Tyrone Vincent, *Member*, *IEEE*

Abstract—A method is developed to analyze the accuracy of the relative head-to-object position and orientation (pose) in augmented reality systems with head-mounted displays. From probabilistic estimates of the errors in optical tracking sensors, the uncertainty in head-to-object pose can be computed in the form of a covariance matrix. The positional uncertainty can be visualized as a 3D ellipsoid. One useful benefit of having an explicit representation of uncertainty is that we can fuse sensor data from a combination of fixed and head-mounted sensors in order to improve the overall registration accuracy. The method was applied to the analysis of an experimental augmented reality system, incorporating an optical see-through head-mounted display, a head-mounted CCD camera, and a fixed optical tracking sensor. The uncertainty of the pose of a movable object with respect to the head-mounted display was analyzed. By using both fixed and head mounted sensors, we produced a pose estimate that is significantly more accurate than that produced by either sensor acting alone.

Index Terms—Augmented reality, pose estimation, registration, uncertainty analysis, error propagation, calibration.

1 Introduction

Augmented reality is a term used to describe systems in which computer-generated information is superimposed on top of the real world [1]. One form of enhancement is to use computer-generated graphics to add virtual objects (such as labels or wire-frame models) to the existing real world scene. Typically, the user views the graphics with a head-mounted display (HMD), although some systems have been developed that use a fixed monitor (e.g., [2], [3], [4], [5]). The combining of computer-generated graphics with real-world images may be accomplished with either optical [6], [7], [8] or video technologies [9], [10].

A basic requirement for an AR system is to accurately align virtual and real-world objects so that they appear to coexist in the same space and merge together seamlessly. This requires that the system accurately sense the position and orientation (pose) of the real world object with respect to the user's head. If the estimated pose of the object is inaccurate, the real and virtual objects may not be registered correctly. For example, a virtual wire-frame model could appear to float some distance away from the real object. This is clearly unacceptable in applications where the user is trying to understand the relationship between real and virtual objects. Registration inaccuracy is one of the most important problems limiting augmented reality applications today [11].

This paper shows how one can estimate the registration accuracy in an augmented reality system, based on the characteristics of the sensors used in the system. Only quasi-static registration is considered in this paper; that is, objects are stationary when viewed, but can freely be moved. We develop an analytical model and show how the model can be used to properly combine data from multiple sensors to improve registration accuracy and gain insight into the effects of object and sensor geometry and configuration. A preliminary version of this paper was presented at the First International Workshop on Augmented Reality [12].

1.1 Registration Techniques in Augmented Reality

To determine the pose of an object with respect to the user's head, tracking sensors are necessary. Sensor technologies that have been used in the past include mechanical, magnetic, acoustic, and optical [13]. We concentrate on optical sensors (such as cameras and photo-effect sensors) since they have the best overall combination of speed, accuracy, and range [7], [14], [15].

There has been much work in the past in the photogrammetry and computer vision fields on methods for object recognition and pose estimation from images. Some difficult problems (which are not addressed here) include how to extract features from the images and determine the correspondence between extracted image features and features on the object. In many practical applications, these problems can be alleviated by preplacing distinctive optical targets, such as light emitting diodes (LEDs) or passive fiducial markings, in known positions on the object. The 3D locations of the target points on the object must be carefully measured, in some coordinate frame attached to the object. In this paper, we will assume that point features have been extracted and the correspondences known so that the only remaining problem is to determine the pose of the object with respect to the HMD.

One issue is whether the measured points are two-dimensional (2D) or three-dimensional (3D). Simple passive optical sensors, such as video cameras and photo-effect sensors, can only sense the direction to a target point and not its range. The measured data points are 2D, i.e., they

- *The authors are with the Engineering Division, Colorado School of Mines, 1500 Illinois St., Golden, CO 80401.*
  E-mail: {whoff, tvincent}@mines.edu.

Manuscript received 1 Feb. 1999; revised 6 July 2000; accepted 10 July 2000.
For information on obtaining reprints of this article, please send e-mail to: tvcg@computer.org, and reference IEEECS Log Number 109094.

represent the locations of the target points projected onto the image plane. On the other hand, active sensors, such as laser range finders, can directly measure direction and range, yielding fully 3D target points. Another way to obtain 3D data is to use triangulation; for example, by using two or more passive sensors (stereo vision). The accuracy of locating the point is improved by increasing the separation (baseline) between the sensors.

Once the locations of the target points have been determined (either 2D or 3D), the next step is to determine the full six degree-of-freedom (DOF) pose of the object with respect to the sensor. Again, we assume that we know the correspondence of the measured points to the known 3D points on the object model. If one has 3D point data, this procedure is known as the "absolute orientation" problem in the photogrammetry literature. If one has 2D target points, this procedure is known as the "exterior orientation" problem [16].

Another issue is where to locate the sensor and target. One possibility is to mount the sensor at a fixed known location in the environment and put targets on both the HMD and on the object of interest (a configuration called "outside-in" [14]). We measure the pose of the HMD with respect to the sensor, and the pose of the object with respect to the sensor, and derive the relative pose of the object with respect to the HMD. Another possibility is to mount the sensor on the HMD and the target on the object of interest (a configuration called "inside-out"). We measure the pose of the object with respect to the sensor and use the known sensor-to-HMD pose to derive the relative pose of the object with respect to the HMD. Both approaches have been tried in the past and each has advantages and disadvantages.

With a fixed sensor (outside-in approach), there is no limitation on size and weight of the sensor. Multiple cameras can be used, with a large baseline, to achieve highly accurate 3D measurements via triangulation. For example, commercial optical measurement systems, such as Northern Digital's Optotrak, have baselines of approximately 1 meter and are able to measure the 3D positions of LED markers to an accuracy of approximately 0.15 mm. The orientation and position of a target pattern is then derived from the individual point positions. A disadvantage with this approach is that head orientation must be inferred indirectly from the point positions.

The inside-out approach has good registration accuracy because a slight rotation of a head-mounted camera causes a large shift of a fixed target in the image. However, a disadvantage of this approach is that large translation errors occur along the line of sight of the camera. To avoid this, additional cameras could be added with lines of sight orthogonal to each other.

1.2 Need for Accuracy Analysis and Fusion

In order to design an augmented reality system that meets the registration requirements for a given application, we would like to be able to estimate the registration accuracy for a given sensor configuration. For example, we would like to estimate the probability distribution of the 3D error distance between a generated virtual point and a corresponding real object point. Another measure of interest is the overlay error; that is, the 2D distance between the projected virtual point and the projected real point on the HMD image plane, which is similar to the image alignment error metrics that appear in other work [7], [9], [17].

Another reason to have an analytical representation of uncertainty is for fusing data from multiple sensors. For example, data from head-mounted and fixed sensors might be combined to derive a more accurate estimate of object-to-HMD pose. The uncertainties of these two sensors might be complementary so that, by combining them, we can derive a pose that is much more accurate than that from each sensor used alone. In order to do this, a mathematical analysis is required of uncertainties associated with the measurements and derived poses. Effectively, we can create a hybrid system that combines the "inside-out" and "outside-in" approaches.

1.3 Relationship to Past Work and Specific Contributions

Augmented reality is a relatively new field, but the problem of registration has received ample attention, with a number of authors taking an optical approach. Some researchers have used photocells or photo-effect sensors which track light-emitting diodes (LEDs) placed on the head, object of interest, or both [7], [14], [15]. Other researchers have used cameras and computer vision techniques to detect LEDs or passive fiducial markings [5], [8], [18], [19], [20], [21]. The resulting detected features, however they are obtained, are used to determine the relative pose of the object to the HMD. A number of researchers have evaluated their registration accuracy experimentally [17], [7], with Monte-Carlo simulations [19], or both [18]. However, no one has studied the effect of sensor-to-target configuration on registration accuracy. In this paper, we develop an analytical model to show how sensor errors propagate through to registration errors, given a statistical distribution of the sensor errors and the sensor-to-target configuration.

Some researchers avoid the problem of determining pose altogether and instead concentrate on aligning the 2D image points using affine projections [22], [23]. Although this approach works well for video-based augmented reality systems, in optical see-through HMD systems, it would not work as well because the image as seen by the head-mounted camera may be different than the image seen by the user directly through the optical combiner.

A number of researchers have developed error models for HMD-based augmented reality systems. Some researchers have looked at the optical characteristics of HMDs in order to calculate viewing transformations and calibration techniques [24], [25]. Holloway [17] analyzed the causes of registration error in a see-through HMD system, due to the effects of misalignment, delay, and tracker error. However, he did not analyze the causes of tracker error, merely its effect on the overall registration accuracy. This work, on the other hand, focuses specifically on the tracker error and does not look at the errors in other parts of the system, or attempt to derive an overall end-to-end error model.

In the computer vision field, the problem of determining the position and orientation from a set of given point or line correspondences has been well-studied. Some researchers have developed analytical expressions for the uncertainty of a 3D feature position as derived from image data [26]. Other researchers have evaluated the accuracy of pose estimation algorithms using Monte Carlo simulations [27], [28], [29], [30]. Few researchers have addressed the issue of error propagation in pose estimation. We follow the method suggested by Haralick and Shapiro [16], who outline how to derive the uncertainty of an estimated quantity (such as a pose) from the given uncertainties in the measured data.

Kalman filtering [31] is a standard technique for optimal estimation. It has been used to estimate head pose in augmented and virtual reality applications [7], [32], [33]. From a sequence of sensor measurements, these techniques also estimate the uncertainty of the head pose. This is similar to the work described in this paper in the sense that a Kalman filter can be interpreted as a method for obtaining a maximum likelihood estimate of the state in a dynamic system, given input-output data [34]. Our system is static and, so, we do not have a model of the state dynamics. We fuse data from two measurements, rather than data from a measurement and a prediction from past data.

In this work, a method is developed to explicitly compute uncertainties of pose estimates, propagate these uncertainties from one coordinate system to another, and fuse pose estimates from multiple sensors. The contribution of this work is the application of this method to the registration problem in augmented reality. Specifically:

- The method shows how to estimate the uncertainty of object-to-HMD pose from the geometric configuration of the optical sensors and the pose estimation algorithms used. To help illustrate the method, we describe its application to a specific augmented reality system.
- We show how data from multiple different sensors can be fused, taking into account the uncertainties associated with each, to yield an improved object-to-HMD pose. In particular, it is shown that a hybrid sensing system combining both head-mounted and fixed sensors can improve registration accuracy over that from either sensor used alone.
- We demonstrate mathematically some insights regarding the characteristics of registration sensors. In particular, we show that the directions of greatest uncertainty for a head-mounted and fixed sensor are nearly orthogonal and that these can be fused in a simple way to improve the overall accuracy.

The remainder of this paper is organized as follows: Section 2 provides a background on pose estimation, with a description of the terminology used in the paper. Section 3 develops the method for estimating the uncertainty of a pose, transforming it from one coordinate frame to another, and fusing two pose estimates. Section 4 describes the particular experimental augmented reality system that was used to test the registration method—that of a surgical aid. Section 5 illustrates the application of the method to the surgical aid system. A typical configuration is analyzed and the predicted accuracy of the combined (hybrid) pose estimate is found to be much improved over that obtained by either sensor alone. Finally, Section 6 provides a discussion.

2 Background on Pose Estimation

2.1 Representation of Pose

The pose of a rigid body {A} with respect to another coordinate system {B} can be represented by a six element vector ${}^{B}_{A}\mathbf{x} = ({}^{B}x_{Aorg}, {}^{B}y_{Aorg}, {}^{B}z_{Aorg}, \alpha, \beta, \gamma)^T$, where ${}^{B}\mathbf{p}_{Aorg} = ({}^{B}x_{Aorg}, {}^{B}y_{Aorg}, {}^{B}z_{Aorg})^T$ is the origin of frame {A} in frame {B}, and (α, β, γ) are the angles of rotation of {A} about the (z, y, x) axes of {B}. An alternative representation of orientation is to use three elements of a quaternion; the conversion between Euler angles and quaternions is straightforward [35].

Equivalently, pose can be represented by a 4 × 4 homogeneous transformation matrix [35]:

$$ {}^{B}_{A}\mathbf{H} = \begin{pmatrix} {}^{B}_{A}\mathbf{R} & {}^{B}\mathbf{p}_{Aorg} \\ 0 & 1 \end{pmatrix}, \tag{1} $$

where ${}^{B}_{A}\mathbf{R}$ is the 3 × 3 rotation matrix corresponding to the angles (α, β, γ). In this paper, we shall use the letter $\mathbf{x}$ to designate a six-element pose vector and the letter $\mathbf{H}$ to designate the equivalent 4 × 4 homogeneous transformation matrix.

Homogeneous transformations are a convenient and elegant representation. Given a homogeneous point ${}^{A}\mathbf{p} = ({}^{A}x_P, {}^{A}y_P, {}^{A}z_P, 1)^T$, represented in coordinate system {A}, it may be transformed to coordinate system {B} with a simple matrix multiplication ${}^{B}\mathbf{p} = {}^{B}_{A}\mathbf{H}\,{}^{A}\mathbf{p}$. The homogeneous matrix representing the pose of frame {B} with respect to frame {A} is just the inverse of the pose of {A} with respect to {B}, i.e., ${}^{A}_{B}\mathbf{H} = {}^{B}_{A}\mathbf{H}^{-1}$. Finally, if we know the pose of {A} with respect to {B} and the pose of {B} with respect to {C}, then the pose of {A} with respect to {C} is easily given by the matrix multiplication ${}^{C}_{A}\mathbf{H} = {}^{C}_{B}\mathbf{H}\,{}^{B}_{A}\mathbf{H}$.

2.2 Pose Estimation Algorithms

The 2D-to-3D pose estimation problem is to determine the pose of a rigid body, given an image from a single camera (this is also called the "exterior orientation" problem in photogrammetry). Specifically, we are given a set of 3D known points on the object (in the coordinate frame of the object) and the corresponding set of 2D measured image points from the camera, which are the perspective projections of the 3D points. The internal parameters of the camera (focal length, principal point, etc.) are known. The goal is to find the pose of the object with respect to the camera, ${}^{cam}_{obj}\mathbf{x}$. There are many solutions to the problem; in this work, we used the algorithm described by Haralick and Shapiro [16], which uses an iterative nonlinear least squares method. The algorithm effectively minimizes the squared error between the measured 2D point locations and the predicted 2D point locations.

The 3D-to-3D pose estimation problem is to determine the pose of a rigid body, given a set of 3D point measurements[1] (this is also called the "absolute orientation" problem in photogrammetry). Specifically, we are given a set of 3D known points on the object $\{{}^{obj}\mathbf{p}_i\}$ and the

[1] These 3D point measurements may have been obtained from a previous triangulation process using a stereo vision sensor.

corresponding set of 3D measured points from the sensor $\{{}^{sen}\mathbf{p}_i\}$. The goal is to find the pose of the object with respect to the sensor, ${}^{sen}_{obj}\mathbf{x}$. There are many solutions to the problem; in this work we used the solution by Horn [36], which uses a quaternion-based method.[2] The algorithm effectively minimizes the squared error between the measured 3D point locations and the predicted 3D point locations.

3 Determination and Manipulation of Pose Uncertainty

Given that we have estimated the pose of an object using one of the methods above, what is the uncertainty of the pose estimate? We can represent the uncertainty of a six-element pose vector $\mathbf{x}$, by a 6 × 6 covariance matrix $\mathbf{C_x} = E(\Delta\mathbf{x}\Delta\mathbf{x}^T)$, which is the expectation of the square of the difference between the estimate and the true vector.

This section describes methods to estimate the covariance matrix of a pose, given the estimated uncertainties in the measurements, transform the covariance matrix from one coordinate frame to another, and combine two pose estimates.

3.1 Computation of Covariance

Assume that we have n measured data points from the sensor $\{\mathbf{p}_i\}$ and the corresponding points on the object $\{\mathbf{q}_i\}$. The object points $\mathbf{q}_i$ are 3D; the data points $\mathbf{p}_i$ are either 3D (in the case of 3D-to-3D pose estimation) or 2D (in the case of 2D-to-3D pose estimation). We assume that the noise in each measured data point is independent and that the noise distribution of each point is given by a covariance matrix $\mathbf{C_p}$.

Let $\mathbf{p}_i = \mathbf{g}(\mathbf{q}_i, \mathbf{x})$ be the function which transforms object points into measured data points for a hypothesized pose $\mathbf{x}$. In the case of 3D-to-3D pose estimation, this is just a multiplication of $\mathbf{q}_i$ by the corresponding homogeneous transformation matrix. In the case of 2D-to-3D pose estimation, the function is composed of a transformation followed by a perspective projection. The pose estimation algorithms described above solve for $\mathbf{x}_{est}$ by minimizing the sum of the squared errors. Assume that have we solved for $\mathbf{x}_{est}$ using the appropriate algorithm (i.e., 2D-to-3D or 3D-to-3D). We then linearize the equation about the estimated solution $\mathbf{x}_{est}$:

$$\mathbf{p}_i + \Delta\mathbf{p}_i = \mathbf{g}(\mathbf{q}_i, \mathbf{x}_{est} + \Delta\mathbf{x}) \approx \mathbf{g}(\mathbf{q}_i, \mathbf{x}_{est}) + \left[\frac{\partial\mathbf{g}}{\partial\mathbf{x}}\right]^T_{\mathbf{q}_i, \mathbf{x}_{est}} \Delta\mathbf{x}. \tag{2}$$

Since $\mathbf{p}_i \approx \mathbf{g}(\mathbf{q}_i, \mathbf{x}_{est})$, the equation reduces to

$$\Delta\mathbf{p}_i = \left[\frac{\partial\mathbf{g}}{\partial\mathbf{x}}\right]^T_{\mathbf{q}_i, \mathbf{x}_{est}} \Delta\mathbf{x} = \mathbf{M}_i\Delta\mathbf{x}, \tag{3}$$

where $\mathbf{M}_i$ is the Jacobian of $\mathbf{g}$, evaluated at $(\mathbf{q}_i, \mathbf{x}_{est})$. Combining all the measurement equations:

$$\begin{pmatrix} \Delta\mathbf{p}_1 \\ \vdots \\ \Delta\mathbf{p}_n \end{pmatrix} = \begin{pmatrix} \mathbf{M}_1 \\ \vdots \\ \mathbf{M}_n \end{pmatrix} \Delta\mathbf{x} \Rightarrow \Delta\mathbf{P} = \mathbf{M}\Delta\mathbf{x}. \tag{4}$$

Solving for $\Delta\mathbf{x}$ in a least squares sense, we get $\Delta\mathbf{x} = \left(\mathbf{M}^T\mathbf{M}\right)^{-1}\mathbf{M}^T\Delta\mathbf{P}$. The covariance matrix of $\mathbf{x}$ is given by the expectation of the outer product:

$$\begin{aligned}
\mathbf{C_x} &= E\left(\Delta\mathbf{x}\,\Delta\mathbf{x}^T\right) \\
&= E\left[\left(\mathbf{M}^T\mathbf{M}\right)^{-1}\mathbf{M}^T\Delta\mathbf{P}\Delta\mathbf{P}^T\left(\left(\mathbf{M}^T\mathbf{M}\right)^{-1}\mathbf{M}^T\right)^T\right] \\
&= \left(\mathbf{M}^T\mathbf{M}\right)^{-1}\mathbf{M}^T E\left(\Delta\mathbf{P}\Delta\mathbf{P}^T\right)\left(\left(\mathbf{M}^T\mathbf{M}\right)^{-1}\mathbf{M}^T\right)^T \\
&= \left(\mathbf{M}^T\mathbf{M}\right)^{-1}\mathbf{M}^T \begin{pmatrix} \mathbf{C_p} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \mathbf{C_p} \end{pmatrix} \left(\left(\mathbf{M}^T\mathbf{M}\right)^{-1}\mathbf{M}^T\right)^T.
\end{aligned} \tag{5}$$

Note that we have assumed that the errors in the data points are independent, i.e., $E(\Delta\mathbf{p}_i\Delta\mathbf{p}_j^T) = 0$, for $i \neq j$. If the errors in different data points are actually correlated, our simplified assumption could result in an underestimate of the actual covariance matrix. Also, the above analysis was derived assuming that the noise is small. However, we computed the covariance matrices for the configuration described in Section 4, using both (5) and using a Monte Carlo simulation, and found (5) is fairly accurate even for noise levels much larger than in our application. For example, using input noise with variance 225 mm^2 (compared to the actual 0.0225 mm^2 in our application) the largest deviation between the variances of the translational dimensions was 5.5 mm^2 (out of 83 mm^2).

3.2 Transformation of Covariance

We can transform a covariance matrix from one coordinate frame to another. Assume that we have a six-element pose vector $\mathbf{x}$ and its associated covariance matrix $\mathbf{C_x}$. Assume that we apply a transformation, represented by a six-element vector $\mathbf{w}$, to $\mathbf{x}$ to create a new pose $\mathbf{y}$. Denote $\mathbf{y} = \mathbf{g}(\mathbf{x}, \mathbf{w})$. A Taylor series expansion yields $\Delta\mathbf{y} = \mathbf{J}\Delta\mathbf{x}$, where $\mathbf{J} = (\partial\mathbf{g}/\partial\mathbf{x})$. The covariance matrix $\mathbf{C_y}$ is found by:

$$\begin{aligned}
\mathbf{C_y} &= E\left(\Delta\mathbf{y}\Delta\mathbf{y}^T\right) = E\left[(\mathbf{J}\Delta\mathbf{x})(\mathbf{J}\Delta\mathbf{x})^T\right] \\
&= \mathbf{J}E\left(\Delta\mathbf{x}\Delta\mathbf{x}^T\right)\mathbf{J}^T = \mathbf{J}\mathbf{C_x}\mathbf{J}^T.
\end{aligned} \tag{6}$$

A variation on this method is to assume that the transformation $\mathbf{w}$ also has an associated covariance matrix $\mathbf{C_w}$. In this case, the covariance matrix $\mathbf{C_y}$ is:

$$\mathbf{C_y} = \mathbf{J_x}\mathbf{C_x}\mathbf{J_x}^T + \mathbf{J_w}\mathbf{C_w}\mathbf{J_w}^T, \tag{7}$$

where $\mathbf{J_x} = (\partial\mathbf{g}/\partial\mathbf{x})$ and $\mathbf{J_w} = (\partial\mathbf{g}/\partial\mathbf{w})$. The above analysis was verified with Monte Carlo simulations, using both the 3D-to-3D algorithm and the 2D-to-3D algorithm.

3.3 Interpretation of Covariance

A useful interpretation of the covariance matrix is obtained by assuming that the errors are jointly Gaussian. The joint probability density for n-dimensional error vector $\Delta\mathbf{x}$ is [37]:

[2] This is the algorithm used in the Northern Digital Optotack sensor, described in Section 4.

$$p(\Delta\mathbf{x}) = \left(|2\pi|^{N/2}|\mathbf{C_x}|^{1/2}\right)^{-1}\exp\left(-\tfrac{1}{2}\Delta\mathbf{x}^T\mathbf{C_x^{-1}}\Delta\mathbf{x}\right). \quad (8)$$

If we look at surfaces of constant probability, the argument of the exponent is a constant, given by the relation $\Delta\mathbf{x}^T\mathbf{C_x^{-1}}\Delta\mathbf{x} = z^2$. This is the equation of an ellipsoid in n dimensions. For a given value of z, the cumulative probability of an error vector being inside the ellipsoid is P. For n = 3 dimensions, the ellipsoid defined by z = 3 corresponds to a cumulative probability P of approximately 97 percent.[3]

For a six-dimensional pose **x**, the covariance matrix $\mathbf{C_x}$ is 6 × 6 and the corresponding ellipsoid is six-dimensional (which is difficult to visualize). However, we can select only the 3D translational component of the pose and look at the covariance matrix corresponding to it. Specifically, let **z** = $(x, y, z)^T$ be the translational portion of the pose vector **x** = $(x, y, z, \alpha, \beta, \gamma)^T$. We obtain **z** from **x** using the equation **z = M x**, where **M** is the matrix

$$\mathbf{M} = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \end{pmatrix} \quad (9)$$

The covariance matrix for **z** is given by $\mathbf{C_z} = \mathbf{M}\ \mathbf{C_x}\ \mathbf{M}^T$ (which is just the upper left 3 × 3 submatrix of $\mathbf{C_x}$). We can then visualize the uncertainty in position using the three-dimensional ellipsoid corresponding to the set $\{\mathbf{z}|(\mathbf{z}-\bar{\mathbf{z}})^{\mathbf{T}}\mathbf{C_z^{-1}}(\mathbf{z}-\bar{\mathbf{z}})9\}$.

We can visualize the uncertainty in the rotational component of the pose by finding the uncertainties in the directions of the x, y, z axes of the coordinate frame relative to the world frame. The orientation of a particular axis **a** of the coordinate frame is found using $\mathbf{a} = \mathbf{R}(\alpha,\beta,\gamma)\mathbf{e}$, where $\mathbf{R}(\alpha,\beta,\gamma)$ is the rotation matrix of the coordinate frame in the world and **e** is the relevant unit vector in the world frame. Using the results of the previous section, the covariance of **a** is given by $\mathbf{C_a} = \frac{\partial}{\partial\mathbf{e}}[\mathbf{R}(\alpha,\beta,\gamma)\mathbf{e}]\mathbf{C_e}\frac{\partial}{\partial\mathbf{e}}[\mathbf{R}(\alpha,\beta,\gamma)\mathbf{e}]^T$, where $\mathbf{C_e}$ is the 3 × 3 lower right submatrix of $\mathbf{C_x}$ corresponding to the angular uncertainty and $\frac{\partial}{\partial\mathbf{e}}[\mathbf{R}(\alpha,\beta,\gamma)\mathbf{e}]$ is the Jacobian of $\mathbf{R}(\alpha,\beta,\gamma)\mathbf{e}$ with respect to α,β,γ. $\mathbf{C_a}$ is of rank 2 and the ellipsoid associated with it will be "flat" in the direction perpendicular to **a**. For visualization, these ellipsoids define the bases of cones drawn about each axis and show how the ends of the axis would move given the variation in the Euler angles.

To illustrate these concepts, a simulation of a pose estimation process was performed. A simulated target pattern was created, attached to a coordinate frame A. The pose of coordinate frame A with respect to a sensor S, ${}^S_A\mathbf{H}$, was estimated using a 3D-to-3D algorithm. The covariance matrix of the resulting pose, $\mathbf{C_A}$, was computed using (5). Fig. 1 shows a rendering of the ellipsoid corresponding to the uncertainty of the translational component of the pose. The ellipsoid is shown centered at the origin of frame A. The rotational uncertainty is depicted as elongated cones about each axis. Note that, although the ellipsoid (representing the translational uncertainty) is almost spherical, the cones (representing the rotational uncertainty) are asymmetrical. The uncertainty is greatest for rotations about the long axis of the target pattern and, so, the cones perpendicular to that axis are elongated. This is because the shorter dimension of the target pattern provides less orientation constraint than the longer dimension.

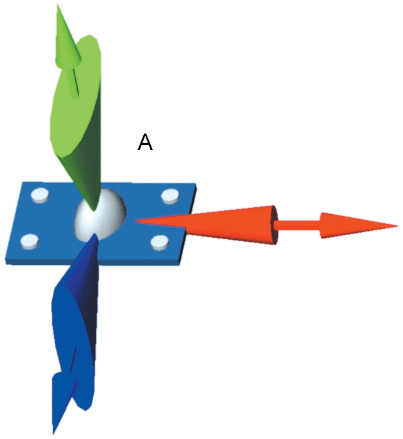


Fig. 1. A visualization of the uncertainty of the pose of a coordinate frame. The ellipsoid, shown centered at the origin of the coordinate frame, represents the uncertainty in the translational component of the pose. The rotational uncertainty is depicted as elongated cones about each axis.

To illustrate the effect of transformations on covariance matrices, another simulation of a pose estimation process was performed. The target pattern used in Fig. 1 was attached to coordinate frame A and the uncertainty of the pose of A with respect to sensor S was computed. As shown in Fig. 2, the translational component of the uncertainty is represented by a small ellipsoid centered at A and the rotational component of the uncertainty is represented by elongated cones about each axis of A. Next, two other objects with coordinate frames B and C were rigidly attached to A, at known poses with respect to A. The poses of B and C with respect to S were derived via ${}^S_B\mathbf{H} = {}^S_A\mathbf{H}{}^A_B\mathbf{H}$ and ${}^S_C\mathbf{H} = {}^S_A\mathbf{H}{}^A_C\mathbf{H}$, respectively. The covariance matrices of these poses, $\mathbf{C_B}$ and $\mathbf{C_C}$, were then estimated using (6). The uncertainties of the translational components of $\mathbf{C_B}$ and $\mathbf{C_C}$ are shown by the ellipsoids centered at B and C, respectively.

Note that the ellipsoids for $\mathbf{C_B}$ and $\mathbf{C_C}$ are much larger than the ellipsoid for $\mathbf{C_A}$, even though the relative poses of B and C with respect to A are known exactly. This is due to the orientation uncertainty in the pose of A with respect to S, which gives rise to an uncertainty in the location of B and C. The uncertainty is greatest in the plane perpendicular to the line to object A—hence, the flattened shapes of the ellipsoids associated with $\mathbf{C_B}$ and $\mathbf{C_C}$. Note that the shape of the flattened ellipsoids corresponds to the shape of the cones about the axes perpendicular to the flattened parts.

In general, the component of translational uncertainty in a frame B that is caused by the orientation error in A can be estimated by $\Delta P = d\ \Delta\theta$, where $\Delta\theta$ is the orientation error and d is the distance between A and B. Thus, the uncertainty in the derived location of B grows with the orientation uncertainty in A and also with the distance between A and B. If one needs to track an object using a

[3] The exact formula for the cumulative probability in N dimensions is $1 - P = \frac{N}{2^{N/2}\Gamma(N/2+1)}\int_z^\infty X^{N-1}e^{-X^2/2}dX$ [37].

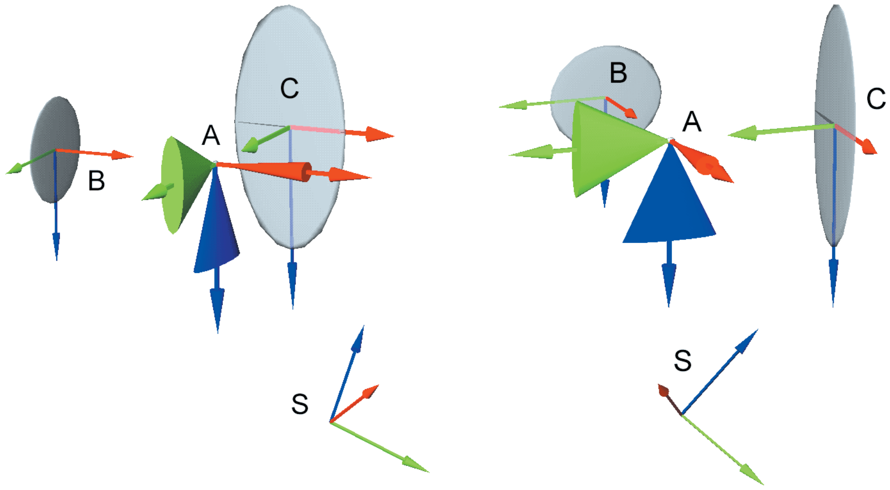


Fig. 2. The orientation uncertainty in the pose of a corrdinate frame A (with respect to sensor S) gives rise to translational uncertainties in the poses of coordinate frames B and C. Two views are shown, taken from slightly different viewpoints. Note the highly flattened shapes of the ellipsoids for frames B and C.

sensor and attaches a target pattern to the object in order to do this, then the target pattern should be placed as close as possible to the object to minimize the derived pose error.

Holloway [17] also noted a similar effect in his analysis of registration errors in augmented reality systems. He found that small errors in measuring the orientation of a (fixed) tracking sensor in a world coordinate system could lead to large errors in the derived position of a point in the scene, due to the large "moment arm" of the tracker-to-point distance.

3.4 Combining Pose Estimates

In this section, we develop a formula to combine two pose estimates, weighted by their covariance matrices, following the approach outlined by Bevington [38]. Let $\mathbf{x_1}$, $\mathbf{x_2}$ be two n-dimensional vectors, representing the measured values of a quantity (e.g., a pose). Let $\mathbf{C_1}$, $\mathbf{C_2}$ be their estimated $n \times n$ covariances. We wish to find the most probable estimate of the mean $\mathbf{x}$.

According to (8), the probability densities of the deviations of $\mathbf{x_1}$ and $\mathbf{x_2}$ from the unknown mean $\mathbf{x}$ are given by (assuming Gaussian distribution of errors):

$$p(\Delta\mathbf{x}_1) = \left(|2\pi|^{N/2}|\mathbf{C}_1|^{1/2}\right)^{-1}\exp\left(-\frac{1}{2}(\mathbf{x_1}-\mathbf{x})^T\mathbf{C}_1^{-1}(\mathbf{x_1}-\mathbf{x})\right)$$
$$p(\Delta\mathbf{x}_2) = \left(|2\pi|^{N/2}|\mathbf{C}_2|^{1/2}\right)^{-1}\exp\left(-\frac{1}{2}(\mathbf{X_2}-\mathbf{x})^T\mathbf{C}_2^{-1}(\mathbf{x_2}-\mathbf{x})\right) \tag{10}$$

Assuming that $\mathbf{x_1}$, $\mathbf{x_2}$ are uncorrelated, the joint probability is just the product of the above two expressions. The maximum likelihood estimate of the parameter $\mathbf{x}$ is given by maximizing the probability density function with respect to $\mathbf{x}$, which is equivalent to minimizing the argument of the exponential:

$$\frac{d}{d\mathbf{x}}\left((\mathbf{x_1}-\mathbf{x})^T\mathbf{C}_1^{-1}(\mathbf{x_1}-\mathbf{x}) + (\mathbf{x_2}-\mathbf{x})^T\mathbf{C}_2^{-1}(\mathbf{x_2}-\mathbf{x})\right) = 0. \tag{11}$$

Taking the derivative and using the fact that the covariance matrices are symmetric:

$$\mathbf{C_1^{-1}}(\mathbf{x_1}-\mathbf{x}) + \mathbf{C_2^{-1}}(\mathbf{x_2}-\mathbf{x}) = 0. \tag{12}$$

The most probable value of $\mathbf{x}$ is therefore given by:

$$\mathbf{x} = \mathbf{C_2}(\mathbf{C_1}+\mathbf{C_2})^{\mathbf{-1}}\mathbf{x_1} + \mathbf{C_1}(\mathbf{C_1}+\mathbf{C_2})^{\mathbf{-1}}\mathbf{x_2}. \tag{13}$$

We take the total derivative of each side:

$$\Delta\mathbf{x} = \mathbf{C_2}(\mathbf{C_1}+\mathbf{C_2})^{\mathbf{-1}}\Delta\mathbf{x_1} + \mathbf{C_1}(\mathbf{C_1}+\mathbf{C_2})^{\mathbf{-1}}\Delta\mathbf{x_2}. \tag{14}$$

To find the covariance of $\Delta\mathbf{x}$, we take the expectation of $\Delta\mathbf{x}\Delta\mathbf{x}^{\mathrm{T}}$:

$$\begin{aligned} &E\left(\Delta\mathbf{x}\Delta\mathbf{x}^{\mathbf{T}}\right) = \\ &= \mathbf{C_2}(\mathbf{C_1}+\mathbf{C_2})^{\mathbf{-1}}E\left(\Delta\mathbf{x}_1\Delta\mathbf{x}_1^T\right)(\mathbf{C_1}+\mathbf{C_2})^{\mathbf{-1}}\mathbf{C_2} \\ &\qquad + \mathbf{C_1}(\mathbf{C_1}+\mathbf{C_2})^{-1}E\left(\Delta\mathbf{x}_2\Delta\mathbf{x}_2^T\right)(\mathbf{C_1}+\mathbf{C_2})^{\mathbf{-1}}\mathbf{C}_1 \\ &= \mathbf{C_2}(\mathbf{C_1}+\mathbf{C_2})^{\mathbf{-1}}\mathbf{C_1}(\mathbf{C_1}+\mathbf{C_2})^{\mathbf{-1}}\mathbf{C_2} \\ &\qquad + \mathbf{C_1}(\mathbf{C_1}+\mathbf{C_2})^{\mathbf{-1}}\mathbf{C_2}(\mathbf{C_1}+\mathbf{C_2})^{\mathbf{-1}}\mathbf{C}_1. \end{aligned} \tag{15}$$

Since all the matrices are symmetric, we can rearrange and simplify to get:

$$\mathbf{C} = \mathbf{C_2}(\mathbf{C_1}+\mathbf{C_2})^{\mathbf{-1}}\mathbf{C}_1. \tag{16}$$

Therefore, this is a method of sensor fusion in an augmented reality system. If the pose of the object with respect to the HMD can be estimated using data from one sensor (e.g., head-mounted) and the same pose can be estimated from another sensor (e.g., fixed), then a combined estimate can be produced using (13) and (16).

When combining pose estimates, we use a quaternion-based representation of orientation, rather than xyz angles or Euler angles. The reason is that xyz angles have a problem for orientations where one angle is close to 180°. In this case, one of the pose vectors may have a value for the angle close to +180° and the other vector may have a value close to −180°. Even though the two vectors represent very similar orientations, the combined vector would represent a

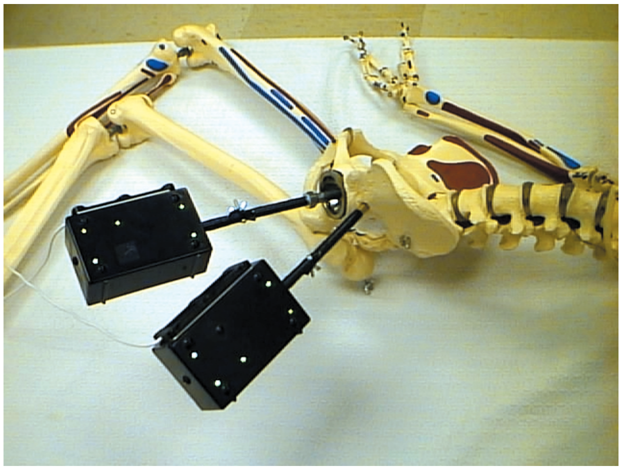


Fig. 3. An artificial hip joint includes a hemispherical metal cup that is implanted into the patient's pelvis. To facilitate tracking, we attached optical targets to the cup (upper black box) and to a peg inserted into the pelvis (lower black box). Not the different geometric patterns of the two 5-LED targets.

wildly different orientation. Quaternions do not have this problem.

4 Description of Experimental AR System

The method described above was applied to an experimental augmented reality system, developed for a surgical aid application, specifically, total hip joint replacement. The purpose of the augmented reality system is to assist the surgeon in implanting the acetabular component (a hemispherical metal cup) into the patient's pelvis, in a prescribed orientation. Placing the acetabular component accurately is important because an incorrect orientation can lead to impingement, accelerated wear, and dislocation. We concentrated on the part of the system that tracks the implant and displays a graphical overlay on the surgeon's HMD that is registered with the implant. To enable tracking of the implant, an optical target was attached to the acetabular implant, as shown in Fig. 3. To enable tracking of the patient, another target was attached to a peg that was rigidly attached to the pelvis (hammered into the bone). Such pegs are normally used in hip surgeries anyway to aid in leg length measurements, so this does not impose an additional burden.

The prototype augmented reality system incorporates an optical see-through HMD (Virtual i-o i-glasses) mounted on a helmet (Fig. 4). A video signal (VGA format) is generated by a desktop PC and transmitted to the HMD through a cable tether. The field-of-view of the HMD is 30 degrees in each eye. In Fig. 4, the helmet is shown mounted on a bust. The bust incorporates a video camera at the location of the user's left eye so that we can record the actual view (with overlays) that would be seen by a user (Fig. 5).

The hybrid sensor system incorporates two sensors: a fixed sensor (Optotrak) and a head-mounted sensor (video camera), described below.

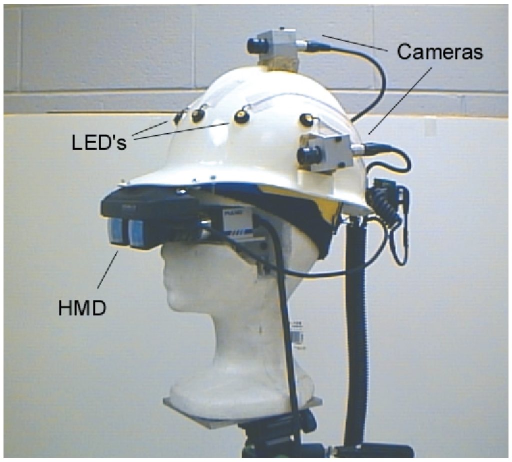


Fig. 4. The augmented reality system incorporates an HMD, cameras, and LED targets, all mounted on a helmet.

4.1 Fixed Sensor

The Optotrak 3020 system (Northern Digital Inc., Waterloo, Canada) is a position tracking device which tracks infrared LEDs by three fixed linear array CCD cameras. The sensor was fastened to one wall of the laboratory. The distance between the outermost two cameras is approximately 90 cm.

At a range of 2.25 m, the stated RMS[4] accuracy of locating a single LED is 0.15 mm in the direction along the (Z) axis of the sensor and 0.1 mm in the directions perpendicular to the axis. At 4 m, the accuracy is 0.45 mm in Z and 0.3 mm in X, Y. The accuracy reported by the manufacturer has been confirmed by Rohling et al. [39]. In this project, the range between the sensor and the objects to be tracked varied between 2.5 and 3 m. To simplify the analysis, we assumed that the errors in 3D point measurements were independent, normally distributed, and had a constant standard deviation of 0.15 mm in all three directions (X, Y, Z).

The LEDs are illuminated sequentially under the control of the Optotrak control unit so that there is no ambiguity about which LED is being observed at any instant. For each LED target that is illuminated, the controller acquires the data from the cameras and calculates (via triangulation) the 3D location of the target with respect to the sensor. From the set of 3D point positions on a target body, the controller also calculates the pose of the body with respect to the sensor, using the quaternion-based algorithm described in Section 2. The update rate of the system is dependent on the number of target points being tracked. For 18 target points, we measured an update rate of approximately 4 Hz.

An optical target, consisting of a set of six infrared LEDs, is fastened to each object of interest, such as the acetabular cup. Fig. 6 shows the six-point rectangular pattern, attached by a white wire, mounted on the front side of the plastic box attached to the cup. Infrared LEDs were also placed on the helmet to form an optical target. Six LEDs were mounted in a semicircular ring around the front half of the helmet, as shown in Fig. 4. In most typical head poses, only four LEDs were visible at one time to the sensor.

[4] RMS = " root mean square."

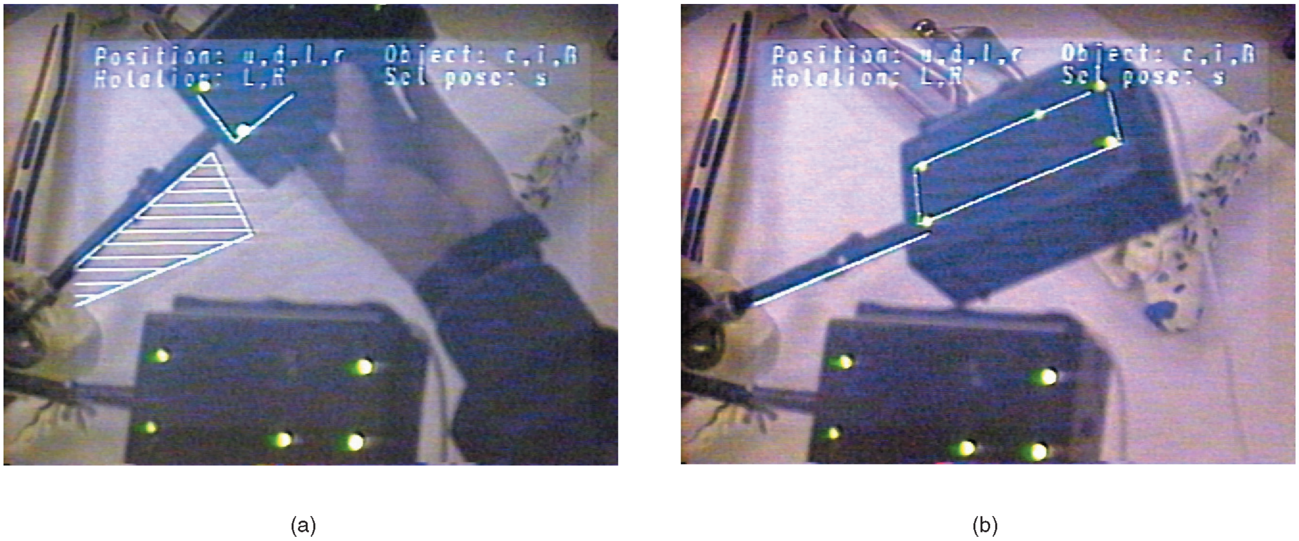


Fig. 5. (a) A view through the HMD (recorded from the "eye-piece" camera), showing a graphical overlay aligned with the optical target attached to the acetabular implant component. The user adjusts the orientation of the implant to zero out the error between the current alignment and the desired alignment (displayed as a crosshatched region). (b) The implant in the correct orientation.

4.2 Head-Mounted Sensor

Three micro-head CCD TV color cameras (Panasonic GP-KS162) were mounted on the helmet. The camera lenses have a nominal field of view of 44 degrees. For the work described in this paper, we only used one camera (the left camera). To simplify the analyses described below, we assumed a square field of view.[5] The NTSC-format video signal from the camera is transmitted to the PC through a coaxial cable tether.

An optical target is affixed to each object of interest. For this work, we used a pattern of five passively illuminated green LEDs in a rectangular planar configuration (see top side of box in Fig. 6). The distinctive geometric pattern of the LEDs enables the correspondence to be easily determined [8].

The video signal is digitized by an image processing board (Sharp GPB-1) in a 486-based PC to a resolution of 512 × 480 pixels, with 8 bits of gray scale intensity per pixel. This board also performs low-level image processing (thresholding and connected-component labeling) to extract the image locations (centroids) of the LED targets. The accuracy of measuring the 2D image location of a point in an image has been well studied by many researchers, including the author [40]. The error in the centroid of a circular image feature due to quantization in the image plane is approximately 0.1 to 0.3 pixels [41]. In addition, there may be an additional 0.1 pixels or so of random horizontal jitter introduced during the video digitization process [42]. Therefore, we assumed a standard deviation of 0.5 pixels for the measured 2D image point locations. To simplify the analysis, we also assumed that the errors were uncorrelated and isotropic.

Pose estimation is done by the PC, using a 2D-to-3D algorithm. The throughput currently achieved with the system described above is approximately 120 ms per update iteration, or 8.3 Hz.

4.3 Coordinate Frames

To analyze the accuracy of our system, we focused on the restricted problem of determining the pose of the acetabular cup implant with respect to the head-mounted display. Note that, in actual operation, the system also determines the pose of the pelvis with respect to the HMD, in order to compute the desired pose of the implant in the pelvis.

The principal coordinate frames used in the system are listed and described in Table 1 and depicted schematically in Fig. 7. Although this figure shows all frames as coplanar, the transformations between frames are actually fully six-dimensional (i.e., three translational and three rotational components).

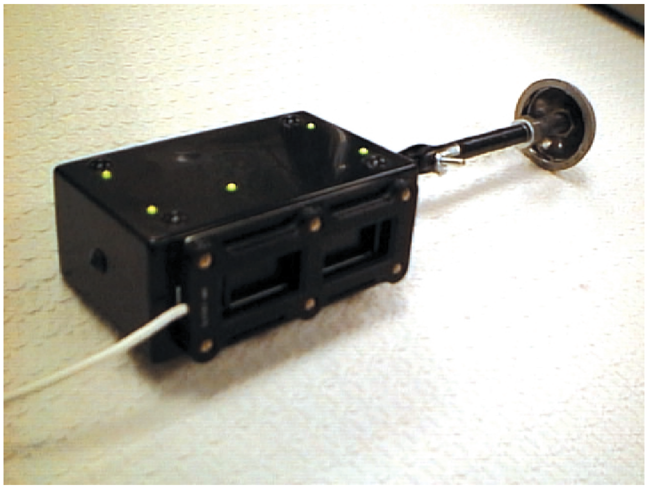


Fig. 6. A rectangular pattern of six infrared LEDs (front site, attached by cable) forms an optical target for the Optotrak sensor. A pattern of five passive LEDs (top side) forms the optical target for the head mounted camera.

[5] Assuming a square field of view artificially restricts the usable image area, but does not affect the accuracy results.

TABLE 1
Principal Coordinate Frames in the System

Frame	Description
HMD	Centered at left eyepiece of display
Implant	Centered on implant component
HMD target	Optical target mounted on helmet, tracked by fixed sensor
Camera	Camera mounted on helmet
Implant target	Optical target attached to implant, tracked by fixed sensor
Camera target	Optical target attached to implant, tracked by head-mounted camera

To aid in visualizing these coordinate frames, a 3D graphical display system was developed using a Silicon Graphics computer and the "Open Inventor" graphics software package. Fig. 8a shows a simplified representation of the coordinate frames on the head: the HMD, the HMD target, and the head-mounted camera. These coordinate frames are rigidly mounted with respect to each other on the helmet. The real helmet assembly is shown in Fig. 4. Fig. 8b shows a simplified representation of the coordinate frames attached to the implant: the implant, the implant target, and the camera target. These coordinate frames are also rigidly mounted with respect to each other. The real implant assembly is shown in Fig. 6. The coordinate axes of all frames are also shown.

Fig. 9a shows the entire room scene, consisting of the fixed sensor on the back wall, the observer with the HMD, and the patient on the table with the hip implant. Fig. 9b shows a 3D visualization of the same scene.

5 Accuracy Analysis

The analysis method described earlier was applied to the experimental augmented reality system to estimate the accuracy of the derived implant-to-HMD pose. The analytical model was implemented using the software application *Mathematica*. The calculations consist of three main steps. First, an estimate of implant-to-HMD pose is derived using data obtained from the Optotrak (fixed) sensor alone. Second, an estimate of implant-to-HMD pose is derived using data obtained from the head-mounted camera alone. Finally, the two estimates are fused to produce a single, more accurate estimate. These steps are described in detail below.

As an example, we determined numeric values for the registration accuracy for a "typical" configuration of the patient, surgeon, and sensors. The configuration shown in Fig. 9 was analyzed and key numerical data describing the configuration are given in Table 2. This is a "typical" configuration in the sense that the distance from the implant to the surgeon's HMD is nominally arm's length (~70 cm). Also, the Optotrak sensor is placed reasonably close to the patient (2.5 m) without interfering with the surgery.

5.1 Analysis of Accuracy from Fixed Sensor

In this section, we analyze the accuracy of pose and overlays using data from the fixed sensor alone. Using data from the fixed sensor (Optotrak), we estimated the pose of the HMD target (${}^{Optotrak}_{HmdTarg}\mathbf{H}$) with respect to the sensor, using the 3D-to-3D algorithm described earlier. From the estimated error in each 3D-point measurement

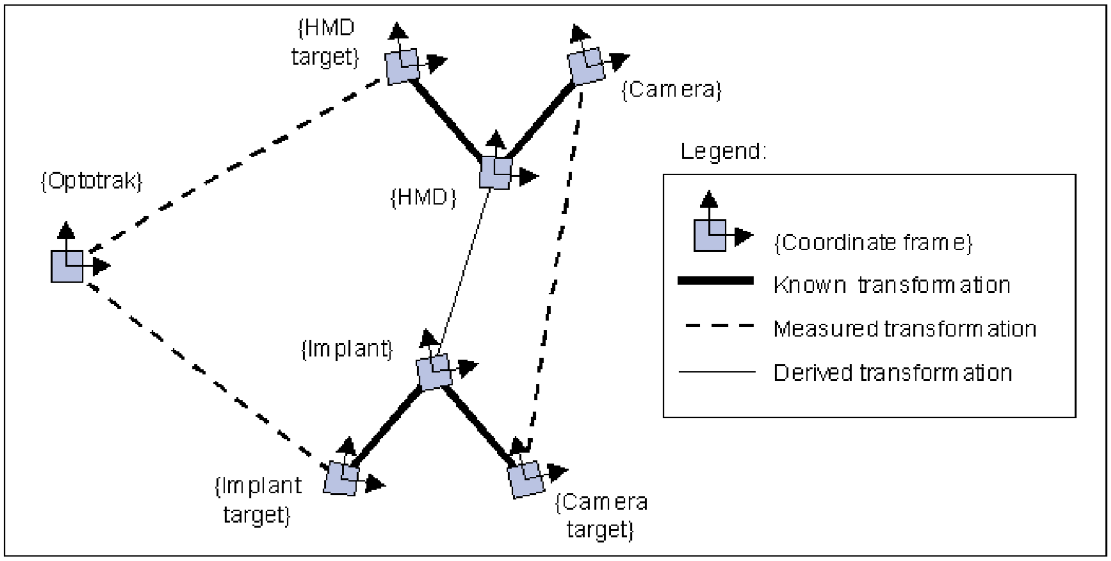
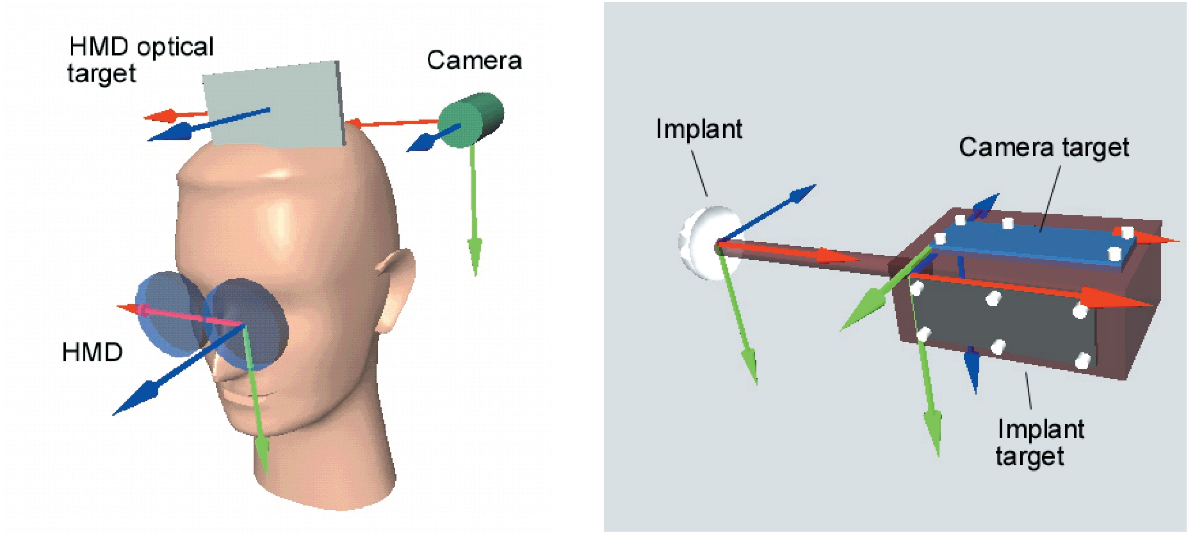
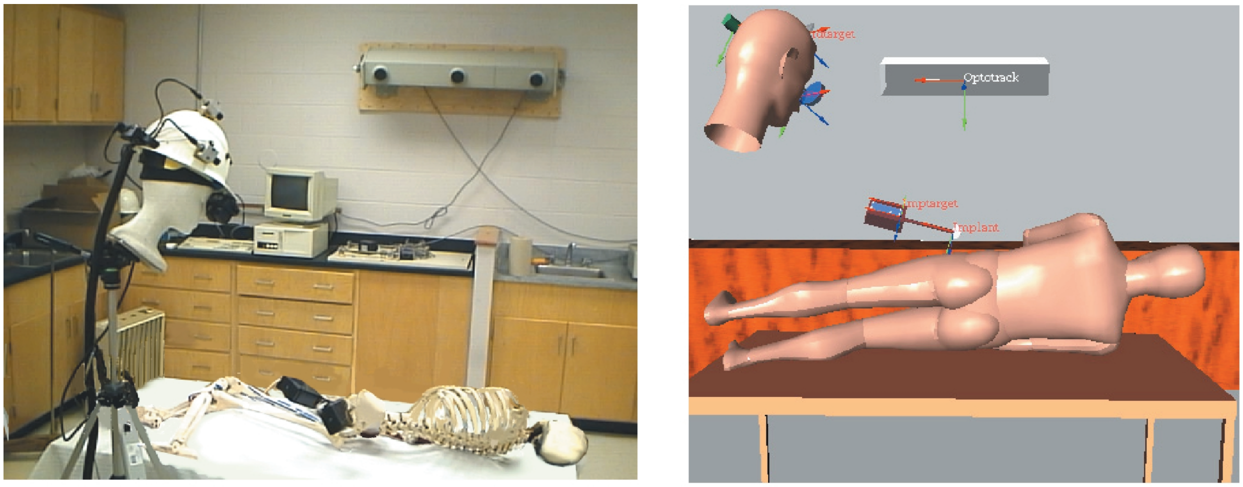


Fig. 7. The principal coordinate frames are shown along with the transformations between them.



(a) (b)

Fig. 8. A 3D visualization of coordinate frames was developed. (a) The coordinate frames on the head are shown: the HMD (two circular disks), the HMD target (flat plate), and the head-mounted camera (cylinder). (b) The coordinate frames on the implant: the acetabular implant (cup), the implant target (six dot pattern), and the camera target (five dot pattern).



(a) (b)

Fig. 9. A visualization of the entire scene, showing the fixed sensor on the wall, the HMD, and the object of interest, which is the hip implant. (a) The real scene. (b) A 3D visualization.

(0.15 mm), the covariance matrix of the resulting pose was determined. Using the known pose of the HMD with respect to the HMD target (${}^{HmdTarg}_{Hmd}\mathbf{H}$), the pose of the HMD with respect to the sensor was estimated, using the equation ${}^{Optotrak}_{Hmd}\mathbf{H} = {}^{Optotrak}_{HmdTarg}\mathbf{H}\,{}^{HmdTarg}_{Hmd}\mathbf{H}$. The covariance matrix of the resulting pose was also estimated. The ellipsoids corresponding to the uncertainties in the translational components of the poses are shown in Fig. 10. In all figures in this paper, the ellipsoids are drawn corresponding to a cumulative probability of 97 percent. However, during rendering, the ellipsoids are scaled up by a factor of 15 in order to make them easily visible. The major axis of the (unscaled) small ellipsoid in Fig. 10 is actually 0.32 mm; that of the larger ellipsoid is 1.84 mm.

Note that this is the estimated error due only to the random noise in the sensor data. This analysis does not take into account systematic errors due to calibration, misalignment of sensors, or other sources. Therefore, the error in an actual system could be higher.

Next, the fixed sensor estimated the pose of the implant target (${}^{Optotrak}_{ImpTarg}\mathbf{H}$) with respect to the sensor, along with the corresponding covariance matrix. Using the known pose of the implant with respect to the implant target (${}^{ImpTarg}_{Implant}\mathbf{H}$), the pose of the implant with respect to the sensor was estimated, using ${}^{Optotrak}_{Implant}\mathbf{H} = {}^{Optotrak}_{ImpTarg}\mathbf{H}\,{}^{ImpTarg}_{Implant}\mathbf{H}$. The covariance matrix of the resulting pose was also estimated. The ellipsoids corresponding to the uncertainties in the translational components of the poses are shown in Fig. 11. The

TABLE 2
Parameters for the Typical Configuration

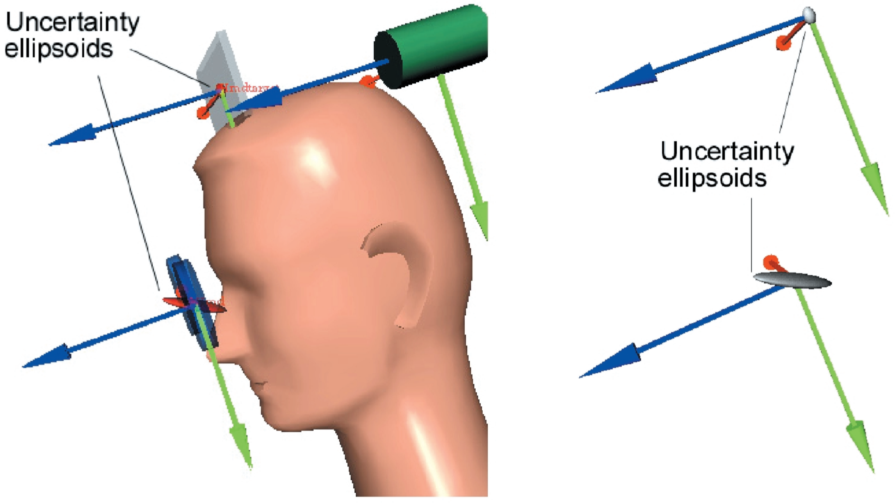
Parameter	Value
Pose of implant with respect to Optotrak	-20.0°, 0.0°, -15.0°, 0 mm, 25 mm, 2500 mm
Pose of implant target with respect to implant	0.0°, 0.0°, 0.0°, 170 mm, -15 mm, -37 mm
Pose of the camera target with respect to implant	-90.0°, 0.0°, 0.0°, 170 mm, -20 mm, 16 mm
Pose of the camera with respect to the HMD	0.0°, 0.0°, 0.0°, -100 mm, -100 mm, -150 mm
Pose of the implant with respect to the HMD	-67.0°, 137.5°, -5.0°, 0 mm, 0 mm, 689 mm
Pose of the HMD target with respect to HMD	0.0°, 0.0°, 0.0°, 0 mm, -125 mm, -60 mm
HMD field of view (assume square image)	30°
HMD resolution (one side)	300 pixels
Head-mounted camera field of view (assume square image)	40°
Head-mounted camera resolution (one side)	512 pixels

Angles represent the rotation about the XYZ axes, respectively. Translations are along the XYZ axes.

major axis of the small ellipsoid in Fig. 11 is 0.34 mm; that of the larger ellipsoid is 1.12 mm.

Finally, the pose of the implant with respect to the HMD was estimated via ${}^{Hmd}_{Implant}\mathbf{H}^{(opto)} = {}^{Hmd}_{Optotrak}\mathbf{H}\,{}^{Optotrak}_{Implant}\mathbf{H}$. The covariance matrix of this pose was estimated and the corresponding ellipsoid is shown in Fig. 12a. The major axis of this ellipsoid is 8.23 mm. Note that the direction of greatest uncertainty is nearly perpendicular to the line of sight from the HMD to the implant, due to the orientation uncertainty in the HMD. In fact, if the HMD target and the HMD coordinate frames were co-located, the direction of greatest uncertainty would be exactly perpendicular to the HMD line of sight.

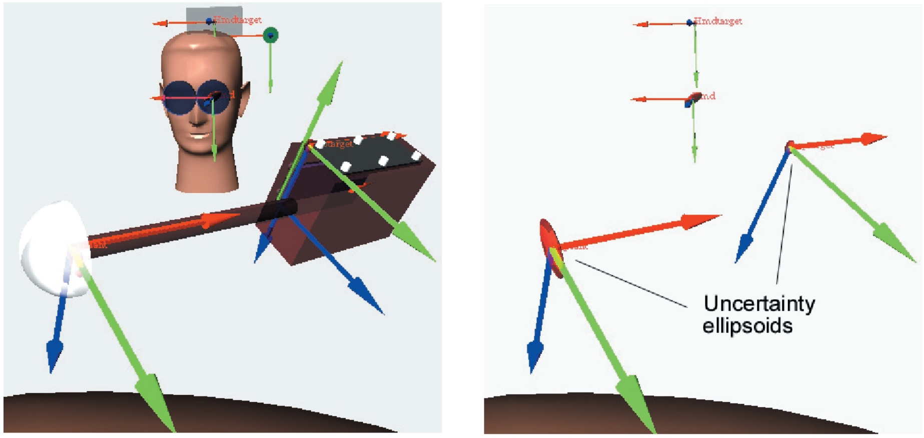
Finally, we performed an analysis of the accuracy of the 2D-image overlay in the HMD, using a Monte Carlo simulation. The pose of the implant with respect to the HMD was calculated from noisy sensor data for 500 trials. In each trial, we added random Gaussian-distributed noise to the measurements from the Optotrak (using a 0.15 mm standard deviation). The derived image point location of the implant origin was then recorded. A cumulative plot of the overlay points is shown in Fig. 12b. The uncertainty ellipsoid corresponding to ${}^{Hmd}_{Implant}\mathbf{H}^{(opto)}$ was also projected onto the image plane (using the 97 percent cumulative probability surface, but drawn with scale factor = 1 rather than 15). The distribution of the overlay points matches the predicted distribution closely. The standard deviation of the overlay points in the vertical direction is 2.19 pixels. This is not a large error, but one that would easily be visible in a high resolution HMD.



(a)

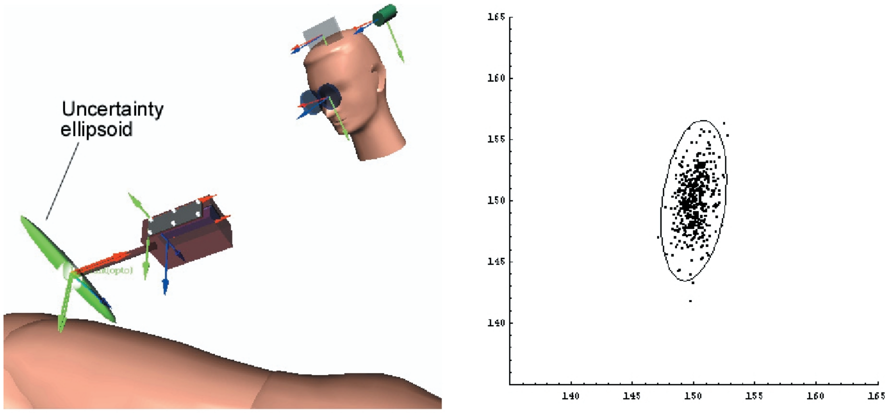
(b)

Fig. 10. The fixed sensor estimates the pose of the HMD target and its covariance matrix (small ellipsoid, barely visible). The image in (a) is redrawn in (b) without the HMD and head models so that the ellipsoids are more easily visible. Using the known pose of the HMD with respect to the HMD target, the pose of the HMD with respect to the sensor is then estimated, along with its covariance matrix (larger ellipsoid).



(a) (b)

Fig. 11. The fixed sensor estimates the pose of the implant target and its covariance matrix (small ellipsoid, barely visible). The image in (a) is redrawn in (b) without the implant and target models so that the ellipsoids are more easily visible. Using the known pose of the implant with respect to the implant target, the pose of the implant with respect to the sensors is then estimated, along with its covariance matrix (larger flattened ellipsoid).



(a) (b)

Fig. 12. (a) Using data from the fixed sensor alone, the pose of the implant with respect to the HMD is derived, along with its covariance matrix (large ellipsoid). (b) The center of the 2D image from the HMD, showing the projection of the predicted uncertainty ellipsoid and a cumulative plot of graphical overlay points from 500 random trials. The standard deviation of the overlay points in the vertical direction is 2.19 pixels.

5.2 Analysis of Accuracy from Head-Mounted Sensor

In this section, we analyze the accuracy of pose and overlays using data from the head-mounted sensor alone for the same typical configuration of patient, surgeon, and sensors. Using data from the head-mounted camera, we estimated the pose of the camera target (${}^{Camera}_{CamTarg}\mathbf{H}$) with respect to the camera, using the 2D-to-3D algorithm described earlier. From the estimated error in each 2D-point measurement (0.5 pixel), the covariance matrix of the resulting pose was determined. Then, using the known pose of the implant with respect to the camera target (${}^{CamTarg}_{Implant}\mathbf{H}$), the pose of the implant with respect to the camera was estimated via ${}^{Camera}_{Implant}\mathbf{H} = {}^{Camera}_{CamTarg}\mathbf{H}{}^{CamTarg}_{Implant}\mathbf{H}$. The covariance matrix of the resulting pose was also estimated. The ellipsoids corresponding to the uncertainties in the translational components of the poses are shown in Fig. 13.

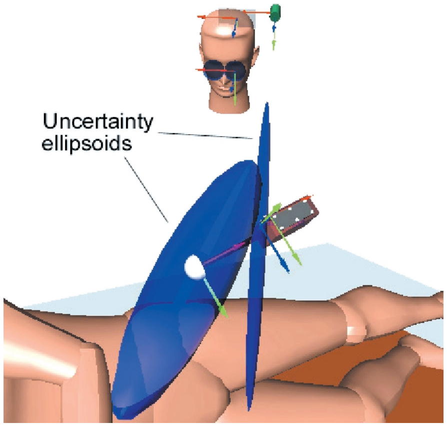
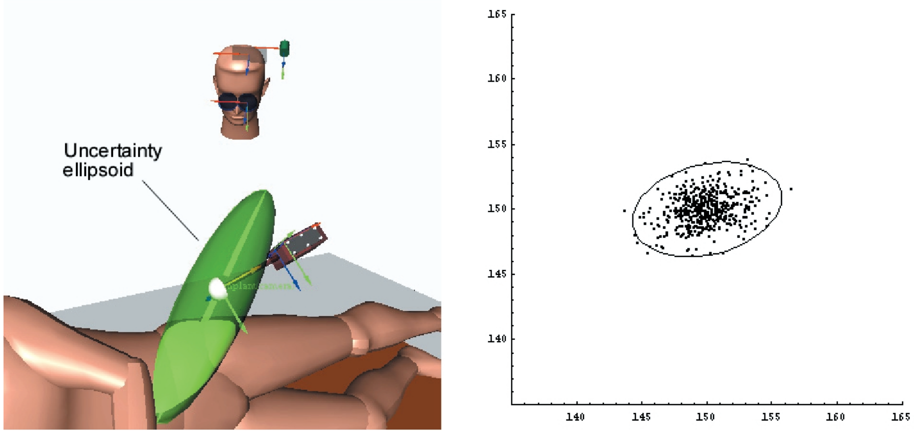


Fig. 13. Using the data from the head-mounted camera, the pose of the camera target with respect to the camera is estimated, along with its covariance matrix (long narrow ellipsoid). Using the known pose of the implant with respect to the camera target, the pose of the implant with respect to the camera is then estimated, along with its covariance matrix (wide ellipsoid to the left).



(a)

(b)

Fig. 14. (a) Using data from the head-mounted camera alone, the pose of the implant with respect to the HMD is computed, along with its covariance matrix (large ellipsoid). (b) The center of the 2D image from the HDM, showing the projection of the predicted uncertainty ellipsoid and a cumulative plot of graphical overlay points from 500 random trials. The standard deviation of the overlay points in the horizontal directions is 1.92 pixels.

Note the large uncertainty of the ellipsoid representing ${}^{Camera}_{CamTarg}\mathbf{H}$ along the line of sight to the camera and very small uncertainty perpendicular to the line of sight. This is typical of poses that are estimated using the 2D-to-3D method. The direction of greatest uncertainty of the camera target is exactly aligned with the direction to the camera. Intuitively, this may be explained as follows: A small translation of the object parallel to the image plane results in an easily measurable change in the image, meaning that the uncertainty of translation is small in this plane. However, a small translation of the object perpendicular to the image plane generates only a very small image displacement, meaning that the uncertainty of translation is large in this direction. The major axis of the ellipsoid corresponding to ${}^{Camera}_{CamTarg}\mathbf{H}$ is 24.6 mm. The major axis of the ellipsoid corresponding to the derived pose, ${}^{Camera}_{Implant}\mathbf{H}$, is 19.9 mm.

Next, the pose of the implant with respect to the HMD was estimated, via ${}^{Hmd}_{Implant}\mathbf{H}^{(cam)} = {}^{Hmd}_{Camera}\mathbf{H}{}^{Camera}_{Implant}\mathbf{H}$. The covariance matrix of this pose was estimated, and the corresponding ellipsoid is shown in Fig. 14a. The major axis of this ellipsoid is 19.9 mm. This ellipsoid was projected onto the 2D image of the HMD, as shown in Fig. 14b.

Also shown in Fig. 14b are the results of a Monte Carlo simulation of the 2D image overlay points. As in the previous section, the pose of the implant with respect to the HMD was calculated from noisy sensor data for 500 trials. In each trial, we added random Gaussian-distributed noise to the measurements from the head-mounted camera (using a 0.5 pixel standard deviation). The derived image point locations were then recorded. The standard deviation of the overlay points in the horizontal direction is 1.92 pixels.

It is interesting to note that the overlay accuracy using data from the head-mounted camera is comparable (actually better) to the overlay accuracy using data from the Optotrak, even though the 6 DOF pose is much less accurate. The reason is that the uncertainty ellipsoid is oriented primarily along the line of sight from the HMD so that the projected uncertainty in the image plane is quite small.

5.3 Fusion of Data from Fixed and Head-Mounted Sensors

The two pose estimates which were derived from the fixed and head-mounted sensors can now be fused. We produced a combined estimate of the implant-to-HMD pose, along with its covariance matrix. The ellipsoids corresponding to the three poses, ${}^{Hmd}_{Implant}\mathbf{H}^{(opto)}$, ${}^{Hmd}_{Implant}\mathbf{H}^{(cam)}$, and ${}^{Hmd}_{Implant}\mathbf{H}^{(hybrid)}$ are shown in Fig. 15a. Note that the large ellipsoids, corresponding to ${}^{Hmd}_{Implant}\mathbf{H}^{(opto)}$ and ${}^{Hmd}_{Implant}\mathbf{H}^{(cam)}$, are nearly orthogonal. The ellipsoid corresponding to the combined pose, ${}^{Hmd}_{Implant}\mathbf{H}^{(hybrid)}$, is much smaller and is contained within the intersection volume of the larger ellipsoids. Fig. 15b is a wire-frame rendering of the ellipsoids, which allows the smaller interior ellipsoid to be seen more easily. The major axis corresponding to the uncertainty of the combined pose is only 1.47 mm.

Finally, we performed an analysis of the accuracy of the 2D image overlay in the HMD, using a Monte Carlo simulation. As in the previous sections, we added random Gaussian-distributed noise to the measurements from the head-mounted camera and the Optotrak sensor, using the same distributions as before. The derived image point location of the resulting overlay was recorded for 500 runs. A cumulative plot of the overlay points is shown in Fig. 15c, along with the projection of the predicted uncertainty ellipsoid corresponding to ${}^{Hmd}_{Implant}\mathbf{H}^{(hybrid)}$. The standard deviation of the overlay points in the direction of maximum error is 0.37 pixels.

5.4 Comparison to Measured Overlay Accuracy

We performed a partial experimental validation of the accuracy analysis. We digitized actual images through the left eyepiece of the head-mounted display while the HMD was stationary and measured the repeatability of the implant overlay position. We did not measure the absolute accuracy of the overlay since this is dependent on many other factors, including camera calibration, HMD calibration, model dimension errors, etc. We recorded the 2D position of a point on the implant overlay for 124 consecutive images similar to Fig. 5. The standard deviations in

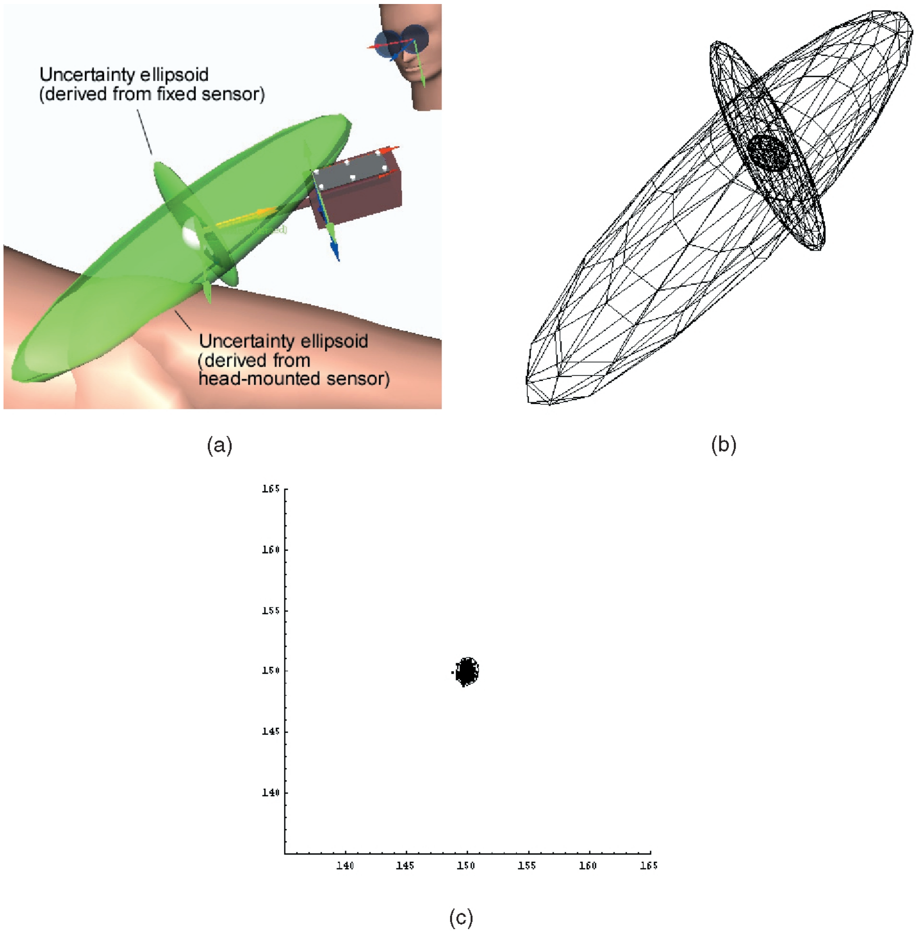


Fig. 15. (a) The ellipsoids from the fixed sensor and the head-mounted sensor are nearly orthogonal. The ellipsoid corresponding to the combined estimate is much smaller and is contained in the volume of intersection. (b) The wire-frame rendering of the uncertainty ellipsoids allows the smaller (combined estimate) ellipsoid to be seen. (c) The 2D image from the HMD, showing the projection of the predicted uncertainty ellipsoid and a cumulative plot of graphical overlay points from 500 random trials. The standard deviation of the overlay points in the direction of maximum error is 0.37 pixels.

the location of the overlay point were 0.29 pixels and 0.48 pixels for the X and Y positions, respectively. This closely matches the predicted maximum standard deviation of 0.37 pixels, as computed in the previous section.

6 Discussion

This paper has developed a method to analyze the head pose accuracy in augmented reality systems. This can be used to fuse sensor data from a combination of fixed and head-mounted sensors in order to improve the registration of objects with respect to a HMD. The method was applied to an actual experimental augmented reality system for the particular application of an orthopedic (hip) surgical aid. A typical configuration was analyzed and it was shown that the hybrid system produces a pose estimate (for the implant with respect to the HMD) that is significantly more accurate than that produced by either sensor acting alone. Using only the fixed sensor, the maximum translational error in any direction was 8.23 mm (corresponding to a 97 percent confidence interval). Using only the head-mounted sensor, the maximum translational error in any direction was 19.9 mm. By combining data from the two sensors, the maximum translational error was reduced to 1.47 mm.

The errors in the 2D image graphical overlays are also significantly reduced in the hybrid system. For the fixed sensor alone, the standard deviation of the predicted image overlay error was 2.19 pixels. For the head-mounted sensor alone, the standard deviation was 1.92 pixels. For the combined (hybrid) system, the standard deviation of the overlay error was 0.37 pixels.

In order to fuse the pose estimates, the uncertainties are explicitly calculated, in the form of covariance matrices. By visualizing the uncertainties as 3D ellipsoids, new insights can be gained. For example, given the measured pose of an object (such as an optical target), in many cases it is necessary to compute the pose of a second object attached to the first object. From the calculated ellipsoids, it is easy to see how the uncertainty in the position of the second object can grow because of the orientation uncertainty in the measured pose of the first object. This uncertainty grows with the distance between the first and second objects. Therefore, it is important to mount optical targets as close as possible to the object of interest.

Pose estimates produced from either sensor acting alone have uncertainties that are not isotropic. The uncertainty of the pose derived from the fixed sensor has its largest component perpendicular to the line of sight from the

HMD. The uncertainty of the pose derived from the head-mounted sensor has its largest component along the line of sight from the HMD. This orthogonality greatly reduces the uncertainty of the fused pose estimate.

An interesting result from this work is that the performance of the head-mounted sensor alone is comparable (in terms of overlay error) to the performance of the fixed sensor alone. This is surprising because pose estimation accuracy of the head-mounted sensor system is not nearly as good. However, the uncertainty of pose estimates from the head-mounted sensor is primarily *along the line of sight* from the HMD. When projected onto the 2D image of the HMD, the apparent overlay errors are relatively small. This is opposed to the errors from the fixed sensor which are primarily perpendicular to the line of sight.

In our analytical model, we assume independent, normally distributed errors in measured point locations. This does not take into account systematic or correlated errors, perhaps due to calibration errors. An alternative analysis which does not depend on Gaussian or independence properties of the noise would be to consider unknown but bounded uncertainty, such as the theory developed in [43]. This would provide a worst case set description of the parameters given a priori bounds on the magnitude of the uncertainty, but without requiring a distribution function. The worst case analysis for augmented reality is the subject of current research.

Another direction for future research is dynamic registration. This paper has only considered quasi-static registration accuracy; that is, where objects are stationary when viewed, but can freely be moved. It would be useful to extend this work to enable accurate registration while objects are moving. Inertial sensors (accelerometers and gyroscopes) could be used to improve dynamic registration.

ACKNOWLEDGMENTS

We are grateful to Johnson & Johnson, especially Dr. Jorge Ochoa, for their support of this work. Tory Lyon and Khoi Nguyen implemented many of the components of the experimental augmented reality system. Rex Rideout of the CSM Engineering Division built the optical target apparatus and mounting jigs. Also, Dr. Doug Dennis (MD) of the Rose Musculoskeletal Research Lab provided useful insight into the hip surgery application and allowed the author to view an actual hip surgery. This work was supported by a grant from Johnson & Johnson Professional, Inc.

REFERENCES

[1] W. Robinett, “Synthetic Experience: A Proposed Taxonomy,” *Presence: Teleoperators and Virtual Environments,* vol. 1, no. 2, pp. 229-247, 1992.

[2] W.E.L. Grimson, T. Lozano-Perez, S.J. White, W.M. Wells III, R. Kikinis, and G.J. Ettinger, “An Automatic Registration Method for Frameless Stereotaxy, Image Guided Surgery, and Enhanced Reality Visualization,” *Proc. IEEE Conf. Computer Vision and Pattern Recognition,* pp. 430-436, June 1994.

[3] P. Milgram, A. Rastogi, and J.J. Grodski, “Telerobotic Control Using Augmented Reality,” *Proc. Fourth IEEE Int'l Workshop Robot and Human Communication (RO-MAN '95),* pp. 21-29, July 1995.

[4] M. Tuceryan, D.S. Greer, R.T. Whitaker, D.E. Breen, C. Crampton, E. Rose, and K.H. Ahlers, “Calibration Requirements and Procedures for a Monitor-Based Augmented Reality System,” *IEEE Trans. Visualization and Computer Graphics,* vol. 1, no. 3, pp. 255-273, 1995.

[5] R. Sharma and J. Molineros, “Computer Vision-Based Augmented Reality for Guiding Manual Assembly,” *Presence,* vol. 6, no. 3, pp. 292-317, 1997.

[6] S. Feiner, B. MacIntyre, and D. Seligmann, “Knowledge-Based Augmented Reality,” *Comm. ACM,* vol. 36, no. 7, pp. 52-62, 1993.

[7] R. Azuma and G. Bishop, “Improving Static and Dynamic Registration in an Optical See-Through HMD,” *Proc. 21st Int'l SIGGRAPH Conf.,* pp. 197-204, July 1994.

[8] W.A. Hoff, T. Lyon, and K. Nguyen, “Computer Vision-Based Registration Techniques for Augmented Reality,” *Proc. Intelligent Robots and Computer Vision XV,* pp. 538-548, Nov. 1996.

[9] M. Bajura and U. Neumann, “Dynamic Registration Correction in Video-Based Augmented Reality Systems,” *IEEE Computer Graphics and Applications,* vol. 15, no. 5, pp. 52-60, 1995.

[10] A. State, M.A. Livingston, W.F. Garrett, G. Hirota, M.C. Whitton, E.D. Pisano, and H. Fuchs, “Technologies for Augmented Reality Systems: Realizing Ultrasound-Guided Needle Biopsies,” *Proc. 23rd Int'l Conf. Computer Graphics and Interactive Techniques (SIGGRAPH '96),* pp. 439-446, Aug. 1996.

[11] R.T. Azuma, “A Survey of Augmented Reality,” *Presence,* vol. 6, no. 4, pp. 355-385, 1997.

[12] W.A. Hoff, “Fusion of Data from Head-Mounted and Fixed Sensors,” *Proc. First Int'l Workshop Augmented Reality,* Nov. 1998.

[13] K. Meyer et al., “A Survey of Position Trackers,” *Presence,* vol. 1, no. 2, pp. 173-200, 1992.

[14] J.-F. Wang, R. Azuma, G. Bishop, V. Chi, J. Eyles, and H. Fuchs, “Tracking a Head-Mounted Display in a Room-Sized Environment with Head-Mounted Cameras,” *Proc. Helmet-Mounted Displays II,* pp. 47-57, Apr. 1990.

[15] D. Kim, S.W. Richards, and T.P. Caudell, “An Optical Tracker for Augmented Reality and Wearable Computers,” *Proc. IEEE 1997 Ann. Int'l Symp. Virtual Reality,* pp. 146-150, Mar. 1997.

[16] R. Haralick and L. Shapiro, *Computer and Robot Vision.* Addison-Wesley, 1993.

[17] R.L. Holloway, “Registration Error Analysis for Augmented Reality,” *Presence,* vol. 6, no. 4, pp. 413-432, 1997.

[18] A. State, G. Hirota, D.T. Chen, W.F. Garrett, and M.A. Livingston, “Superior Augmented Reality Registration by Integrating Landmark Tracking and Magnetic Tracking,” *Proc. 23rd Int'l Conf. Computer Graphics and Interactive Techniques (SIGGRAPH '96),* pp. 429-438, Aug. 1996.

[19] A. Janin, K. Zikan, D. Mizell, M. Banner, and H. Sowizral, “A Videometric Head Tracker for Augmented Reality Applications,” *Proc. Telemanipulator and Telepresence Technologies,* pp. 308-315, 1994.

[20] J.P. Mellor, “Enhanced Reality Visualization in a Surgical Environment,” A.I. Technical Report 1544, Massachusetts Inst. of Technology, Cambridge, Mass., Jan. 1995.

[21] T. Starner, S. Mann, B. Rhodes, J. Levine, J. Healey, D. Kirsch, R. Picard, and A. Pentland, “Augmented Reality through Wearable Computing,” *Presence,* vol. 6, no. 4, pp. 386-398, 1997.

[22] M. Uenohara and T. Kanade, “Vision-Based Object Registration for Real-Time Image Overlay,” *Computers in Biology and Medicine,* vol. 25, no. 2, pp. 249-260, 1995.

[23] K.N. Kutulakos and J. Vallino, “Calibration-Free Augmented Reality,” *IEEE Trans. Visualization and Computer Graphics,* vol. 4, no. 1, pp. 1-20, Jan.-Mar. 1998.

[24] W. Robinett and J. Rolland, “A Computational Model for the Stereoscopic Optics of a Head-Mounted Display,” *Presence: Teleoperators and Virtual Environments,* vol. 1, no. 1, pp. 45-62, 1991.

[25] A.L. Janin, D.W. Mizell, and T.P. Caudell, “Calibration of Head-Mounted Displays for Augmented Reality Applications,” *Proc. IEEE Virtual Reality Ann. Int'l Symp.,* pp. 246-255, Sept. 1993.

[26] O.D. Faugeras, *Three-Dimensional Computer Vision—A Geometric Viewpoint.* Cambridge, Mass.: MIT Press, 1993.

[27] 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. Systems, Man, and Cybernetics,* vol. 19, no. 6, pp. 1,426-1,445, 1989.

[28] J. Weng, P. Cohen, and M. Herniou, “Camera Calibration with Distortion Models and Accuracy Evaluation,” *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 14, no. 10, pp. 965-980, Oct. 1992.

[29] X. Zhuang and Y. Huang, “Robust 3-D-3-D Pose Estimation,” *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 16, no. 8, pp. 818-824, Aug. 1994.

[30] R. Kumar and A. Hanson, “Robust Methods for Estimating Pose and a Sensitivity Analysis,” *CVGIP: Image Understanding,* vol. 60, pp. 313-342, 1994.

[31] A. Gelb, *Applied Optimal Estimation.* Cambridge, Mass.: MIT Press, 1974.

[32] E. Foxlin, “Inertial Head-Tracker Sensor Fusion by a Complementary Separate-Bias Kalman Filter,” *Proc. Virtual Reality Ann. Int’l Symp.,* pp. 185-194, 1996.

[33] K. Nguyen, “Inertial Data Fusion Using Kalman Filter Methods for Augmented Reality,” MS thesis, Eng. Division, Colorado School of Mines, Golden, Colo., 1998.

[34] R.G. Brown and P.Y.C. Hwang, *Introduction to Random Signals and Applied Kalman Filtering,* second ed. New York: John Wiley & Sons, 1992.

[35] J. Craig, *Introduction to Robotics: Mechanics and Control,* second ed. Reading, Mass.: Addison-Wesley, 1990.

[36] B.K.P. Horn, “Closed-Form Solution of Absolute Orientation Using Unit Quaternions,” *J. Optical Soc. Am.,* vol. 4, no. 4, pp. 629-642, 1987.

[37] H. L. Van Trees, *Detection, Estimation, and Modulation Theory.* New York: Wiley, 1968.

[38] P. Bevington, *Data Reduction and Analysis for the Physical Sciences.* New York: McGraw-Hill, 1969.

[39] R. Rohling, P. Munger, J.M. Hollerbach, and T. Peters, “Comparison of Relative Accuracy between a Mechanical and an Optical Position Tracker for Image-Guided Neurosurgery,” *J. Image Guided Surgery,* vol. 1, pp. 30-34, 1995.

[40] C. Sklair, W. Hoff, and L. Gatrell, “Accuracy of Locating Circular Features Using Machine Vision,” *Proc. Cooperative Intelligent Robotics in Space,* W. Stoney, ed., Nov. 1991.

[41] C. Bose and I. Amir, “Design of Fiducials for Accurate Registration Using Machine Vision,” *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 12, no. 12, pp. 1,196-1,200, Dec. 1990.

[42] R.K. Lenz and R.Y. Tsai, “Techniques for Calibration of the Scale Factor and the Image Center for High Accuracy 3D Machine Vision Metrology,” *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 10, no. 5, pp. 713-720, 1988.

[43] F.C. Schweppe, *Uncertain Dynamic Systems.* Englewood Cliffs, N.J.: Prentice Hall, 1973.

William A. Hoff received a BS degree in physics from the Illinois Institute of Technology in 1978, an MS degree in physics, and a PhD degree in computer science from the University of Illinois-Urbana in 1981 and 1987, respectively. After employment at Lockheed-Martin Corp., he joined the faculty of the Colorado School of Mines in 1994. His research interests include augmented reality, computer vision, robotics, and interactive systems. He is a member of the IEEE.

Tyrone Vincent received the BS degree in electrical engineering from the University of Arizona, Tucson, in 1992, and the MS and PhD degrees in electrical engineering from the University of Michigan, Ann Arbor, in 1994 and 1997, respectively. He is currently an assistant professor at the Colorado School of Mines. His research interests include nonlinear estimation, system identification, and fault detection with applications in materials processing, robotics, and power systems. He is a member of the IEEE.