First International Workshop on Augmented Reality, Nov. 1, 1998, San Francisco.

Fusion of Data from Head-Mounted and Fixed Sensors[1]

William A. Hoff
Engineering Division, Colorado School of Mines
Golden, Colorado

Abstract

A methodology is developed to explicitly fuse sensor data from a combination of fixed and head-mounted sensors, in order to improve the registration of objects in an augmented reality system. The methodology was applied to the analysis of an actual experimental augmented reality system, incorporating an optical see-through head-mounted display, a head-mounted CCD camera, and a fixed optical tracking sensor. The purpose of the sensing system was to determine the position and orientation (pose) of a movable object with respect to the head-mounted display. A typical configuration was analyzed and it was shown that the hybrid system produces a pose estimate that is significantly more accurate than that produced by either sensor acting alone. Using only the fixed sensor, the maximum translational error in the location of an object with respect to the head-mounted display in any direction was 8.23 mm (corresponding to a 97% 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. In order to fuse the pose estimates, the uncertainties are explicitly calculated, in the form of covariance matrices. A capability was also developed to visualize the uncertainties as 3D ellipsoids.

1 Introduction

Where is an object of interest with respect to the user's head? In augmented reality systems that use head-mounted displays (HMD's), knowing the relative position and orientation (pose) between object and head is crucial in order to display a virtual object that is aligned with the real object. If the estimated pose of the object is inaccurate, the real and virtual objects may not be registered correctly. Registration inaccuracy is one of the most important problems limiting augmented reality applications today [1].

---

[1] This work was supported by a grant from Johnson & Johnson Professional, Inc.

To determine the pose of an object with respect to the user's head, tracking sensors are necessary. Optical sensors use cameras or photo-effect sensors to track optical targets, such as light emitting diodes (LED's) or passive fiducial markings [2] [3] [4]. Using two or more sensors (stereo vision), the three-dimensional (3D) position of a target point can be determined directly via triangulation. The accuracy of locating the point is improved by increasing the separation (baseline) between the sensors. The full six degree-of-freedom (DOF) pose of a rigid body can be determined by measuring three or more target points on the body, assuming the geometry of the points on the body is known. This procedure is known as the "absolute orientation" problem in the photogrammetry literature. Alternatively, a single sensor can be used to measure the 2D (image) locations of three or more target points on a rigid body. If the geometry of the points is known, the full 6 DOF pose of the rigid body can be estimated, by a procedure is known as "exterior orientation" [5].

One 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" [3]). 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 3-D positions of LED markers to an accuracy of less than 0.25 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 one it is impossible to put multiple cameras with a large baseline separation on the head. Either a small baseline separation must be used, or alternatively a single camera can be used with the exterior orientation algorithm. Either method gives rise to large translation errors along the line of sight of the cameras.

A question arises – is it possible to fuse the data from a head-mounted sensor and a fixed sensor, and derive a more accurate estimate of object-to-HMD pose? If the data from these two types of sensors are complementary, then the resulting pose can be much more accurate than that from each sensor used alone. Effectively, we can create a hybrid system that combines the "inside-out" and "outside-in" approaches.

This paper describes a methodology 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 methodology to the registration problem in augmented reality. It is shown that a hybrid sensing system, combining both head-mounted and fixed sensors can improve registration accuracy.

2 Background on Pose Estimation

2.1 Representation of Pose

The notation in this section follows that of Craig [6]. The pose of a rigid body {A} with respect to another coordinate system {B} can be represented by a six element vector ${}^{B}_{A}X = \left({}^{B}x_{Aorg}, {}^{B}y_{Aorg}, {}^{B}z_{Aorg}, \alpha, \beta, \gamma\right)^{T}$, where ${}^{B}P_{Aorg} = \left({}^{B}x_{Aorg}, {}^{B}y_{Aorg}, {}^{B}z_{Aorg}\right)^{T}$ 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 xyz angles and quaternions is straightforward. Equivalently, pose can be represented by a 4x4 homogeneous transformation matrix:

$$ {}^{B}_{A}H = \begin{pmatrix} {}^{B}_{A}R & {}^{B}P_{Aorg} \\ 0 & 1 \end{pmatrix} \qquad (1) $$

where ${}^{B}_{A}R$ is a 3x3 rotation matrix. In this paper, we shall use the letter X to designate a six-element pose vector and the letter H to designate the equivalent 4x4 homogeneous transformation matrix.

Homogeneous transformations are a convenient and elegant representation. Given a homogeneous point ${}^{A}P = \left({}^{A}x_{P}, {}^{A}y_{P}, {}^{A}z_{P}, 1\right)^{T}$, represented in coordinate system {A}, it may be transformed to coordinate system {B} with a simple matrix multiplication ${}^{B}P = {}^{B}_{A}H \ {}^{A}P$. 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.*, ${}^{A}_{B}H = {}^{B}_{A}H^{-1}$. 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 ${}^{C}_{A}H = {}^{C}_{B}H \; {}^{B}_{A}H$.

2.2 Pose Estimation Algorithms

The problem of determining the pose of a rigid body, given an image from a single camera, is 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, ${}^{cam}_{obj}X$. There are many solutions to the problem; in this work we used the algorithm described by Haralick [5], which uses an iterative non-linear least squares method. The algorithm effectively minimizes the squared error between the measured 2D point locations and the predicted 2D point locations.

The problem of determining the pose of a rigid body, given a set of 3D point measurements, is called the "absolute orientation" problem in photogrammetry. These 3D point measurements may have been obtained from a previous triangulation process, using a sensor consisting of multiple cameras. Specifically, we are given a set of 3D known points on the object $\{{}^{obj}P_i\}$, and the corresponding set of 3D measured points from the sensor $\{{}^{sen}P_i\}$. The goal is to find the pose of the object with respect to the sensor, ${}^{sen}_{obj}X$. There are many solutions to the problem; in this work we used the algorithm described by Horn [7] which uses a quaternion-based method.

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? Knowing the uncertainty is critical to fusing measurements from multiple sensors. We can represent the uncertainty of a six-element pose vector X, by a 6x6 covariance matrix $C_X = E\left(\Delta X \; \Delta X^T\right)$, 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, 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 $\{P_1, P_2, ..., P_N\}$, and the corresponding points on the object $\{Q_1, Q_2, ..., Q_N\}$. The object points Q_i are 3D; the data points 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 the noise distribution of each point is given by a covariance matrix C_P.

Let $P_i = H(Q_i, X)$ be the function which transforms object points into data points. In the case of 3D-to-3D pose estimation, this is just a multiplication of 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. An algorithm that solves for X_{est} minimizes the sum of the squared errors. Assume that have we solved for X_{est} using the appropriate algorithm (*i.e.*, 2D-to-3D or 3D-to-3D). We then linearize the equation about the estimated solution X_{est}:

$$P_i + \Delta P_i = H(Q_i, X_{est} + \Delta X) \approx H(Q_i, X_{est}) + \left[\frac{\partial H}{\partial X}\right]^T_{Q_i, X_{est}} \Delta X \qquad (2)$$

Since $P_i = H(Q_i, X_{est})$, the equation reduces to

$$\Delta P_i = \left[\frac{\partial H}{\partial X}\right]^T_{Q_i, X_{est}} \Delta X = M_i \Delta X \qquad (3)$$

where M_i is the Jacobian of H, evaluated at (Q_i, X_{est}). Combining all the measurement equations, we get the matrix equation:

$$-\begin{pmatrix} \Delta P_1 \\ \vdots \\ \Delta P_N \end{pmatrix} = \begin{pmatrix} M_1 \\ \vdots \\ M_N \end{pmatrix} \Delta X \quad \Rightarrow \quad -\Delta P = M \Delta X \qquad (4)$$

Solving for ΔX, we get $\Delta X = (M^T M)^{-1} M^T \Delta P$. The covariance matrix of X is given by the expectation of the outer product:

$$\begin{aligned} C_X &= E(\Delta X \, \Delta X^T) = \\ &= E\left[(M^T M)^{-1} M^T \Delta P \Delta P^T \left((M^T M)^{-1} M^T \right)^T \right] \\ &= (M^T M)^{-1} M^T \; E(\Delta P \Delta P^T) \left((M^T M)^{-1} M^T \right)^T \\ &= (M^T M)^{-1} M^T \begin{pmatrix} C_P & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & C_P \end{pmatrix} \left((M^T M)^{-1} M^T \right)^T \end{aligned} \qquad (5)$$

Note that we have assumed that the errors in the data points are independent; *i.e.*, $E(\Delta P_i \Delta P_j^T)=0$, 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.) The above analysis was verified with Monte Carlo simulations, using both the 3D-to-3D algorithm and the 2D-to-3D algorithm.

3.2 Interpretation of Covariance

A useful interpretation of the covariance matrix can be obtained by assuming that the errors are jointly Gaussian. The joint probability density for N-dimensional error vector ΔX is [8]:

$$p(\Delta X) = \left(\left|2\pi\right|^{N/2} \left|C_X\right|^{1/2}\right)^{-1} \exp\left(-\tfrac{1}{2}\Delta X^T C_X^{-1} \Delta X\right) \quad (6)$$

If we look at surfaces of constant probability, the argument of the exponent is a constant, given by the relation $\Delta X^T C_X^{-1} \Delta X = z^2$. 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 of approximately 97% [2].

For a six-dimensional pose X, the covariance matrix C_X is 6x6, 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

$$M = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \end{pmatrix} \quad (7)$$

The covariance matrix for Z is given by $C_Z = M\, C_X\, M^T$ (which is just the upper left 3x3 submatrix of C_X). We can then visualize the three dimensional ellipsoid corresponding to C_Z.

3.3 Transformation of Covariance

We can transform a covariance matrix from one coordinate frame to another. Assume that we have a six-element pose vector X and its associated covariance matrix C_X. Assume that we apply a transformation, represented by a six-element vector W, to X to create a new pose Y. Denote Y = g(X, W). A Taylor

---

[2] The exact formula for the cumulative probability in N dimensions is $1-P = \frac{N}{2^{N/2}\Gamma(N/2+1)} \int_z^\infty X^{N-1} e^{-X^2/2} dX$ [8].

series expansion yields $\Delta Y = J\,\Delta X$, where J = (∂g/∂X). The covariance matrix C_Y is found by:

$$C_Y = E\left(\Delta Y \Delta Y^T\right) = E\left[\left(J\,\Delta X\right)\left(J\,\Delta X\right)^T\right] = J E\left(\Delta X\,\Delta X^T\right) J^T = J C_X J^T \quad (8)$$

A variation on this method is to assume that the transformation W also has an associated covariance matrix C_W. In this case, the covariance matrix C_Y is:

$$C_Y = J_X C_X J_X^T + J_W C_W J_W^T \quad (9)$$

where J_X = (∂g/∂X) and J_W = (∂g/∂W). The above analysis was verified with Monte Carlo simulations, using both the 3D-to-3D algorithm and the 2D-to-3D algorithm.

3.4 Combining Pose Estimates

Two vector quantities may be fused by averaging them, weighted by their covariance matrices. Let X_1, X_2 be two N-dimensional vectors, and C_1, C_2 be their NxN covariance matrices. Assuming X_1 and X_2 are uncorrelated, then the combined estimate X and the combined covariance matrix C may be found by the following equations[3]:

$$\begin{aligned} X &= C_2\left(C_1 + C_2\right)^{-1} X_1 + C_1\left(C_1 + C_2\right)^{-1} X_2 \\ C &= C_2\left(C_1 + C_2\right)^{-1} C_1 \end{aligned} \quad (10)$$

Therefore, this is a method of sensor fusion in the hybrid augmented reality system. If the pose of an object with respect to the HMD can be estimated using data from the head-mounted sensor, and the same pose can be estimated using data from the fixed sensor, then a combined estimate can be produced using Equation 10.

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 wildly different orientation. Quaternions do not have this problem.

4 Experiments

The methodology described in the previous sections was applied to an actual experimental augmented reality system developed in our lab. The purpose of

3 These equations can be derived from the discrete Kalman filter update equations, using X_1 as the *a priori* estimate, X_2 as the measurement, and X as the *a posteriori* estimate.

the system is to display a graphical overlay on an HMD, such that the overlay is registered to a movable object in the scene. Only quasi-static registration is considered in this paper; that is, objects are stationary when viewed, but can freely be moved. The system incorporates both head-mounted and fixed sensors. The hybrid system was developed for a surgical aid application, but its capabilities are such that it could be used in many other applications.

The first sub-section below describes the experimental setup, including the sensor and display characteristics. Additional details of the system are described in [9]. In the second sub-section, the task itself is described. Finally, an analysis of registration accuracy is performed.

4.1 Description of Experimental AR System

The prototype augmented reality system incorporates a see-through HMD (Virtual I-O i-glasses ^{TM}) mounted on a helmet (Figure 1, left). A CCD camera with a field of view of 44 degrees is also mounted on the helmet. The NTSC-format video signal from the camera is transmitted to a PC through a cable tether, which digitizes and processes the image. An optical target is affixed to the object of interest (Figure 1, right). For this work, we used a pattern of 5 green LED's, in a rectangular planar configuration. The distinctive geometric pattern of the LED's enables the correspondence to be easily determined [9].

The PC performs low-level image processing to extract the image locations of the LED targets. The noise in the 2D measured image point locations was assumed to be isotropic, with an estimated standard deviation of 0.5 pixels. Pose estimation is done using a 2D-to-3D algorithm. The throughput currently achieved with the system is approximately 8.3 Hz.

Our fixed sensor was an optical measurement system (Northern Digital Optotrak 3020) fastened to one wall of the laboratory. The sensor consists of three linear array CCD cameras. An optical target, consisting of a set of six infrared LED's, is fastened to each object of interest. The cameras detect each LED and calculate (via triangulation) its 3D location with respect to the sensor. From the resulting set of 3D point positions on a target body, the controller also calculates the pose of the body with respect to the sensor. For 18 target points, we measured an update rate of approximately 4 Hz.

Infrared LED's were also placed on the helmet, to form an optical target. A set of 6 LED's were mounted in a semi-circular ring around the front half of the helmet (Figure 1, left). Typically, only 4 LED's were visible at any one time. The measurement noise was assumed to be isotropic, with $\sigma = 0.15$ mm.

4.2 Description of Task

The hybrid augmented reality system was developed for a surgical aid application; specifically, total hip joint replacement. The purpose of the augmented reality system is to track a hip implant and display a graphical overlay on the HMD, that is registered to the implant. Optical targets were attached to the implant to enable sensor tracking, as shown in Figure 1 (right). Separate LED targets were used for the head-mounted and fixed sensors.

The principal coordinate frames used in the system are listed and described in Table 1, and depicted schematically in Figure 2. Even though this figure shows all frames as co-planar, the transformations between frames are actually fully six-dimensional *(i.e.*, three translational and three rotational components).

To aid in visualizing these coordinate frames, a 3D graphical display system was developed using a Silicon Graphics workstation and the "Open Inventor" graphics package. Figure 3 (left) 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. Figure 3 (right) 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 helmet and implant assemblies were shown in Figure 1.) The coordinate axes of all frames are also shown.

Figure 4 (left) 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. Figure 4 (right) shows a 3D visualization of the same scene.

4.3 Analysis of Registration Accuracy

A simulation was implemented, using the software application *Mathematica*, to estimate the accuracy of the derived implant-to-HMD pose. The processing consists 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.

4.3.1 Pose Estimation from Fixed Sensor

Using data from the fixed sensor (Optotrak), we estimated the pose of the HMD target (${}^{Optotrak}_{HmdTarg}H$) with respect to the sensor, using the 3D-to-3D algorithm described earlier. From the estimated error in each 3D point measurement (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 (${}^{HmdTarg}_{Hmd}H$), the pose of the HMD with respect to the sensor was estimated, using the equation ${}^{Optotrak}_{Hmd}H = {}^{Optotrak}_{HmdTarg}H \; {}^{HmdTarg}_{Hmd}H$. 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 Figure 5 (left). In all figures in this paper, the ellipsoids are drawn corresponding to a normalized distance of z=3; *i.e.*, corresponding to a cumulative probability of 97%. However, during rendering the ellipsoids are scaled up by a factor of 10 in order to make them more easily visible. The major axis of the small ellipsoid in Figure 5 (left) is actually 0.32 mm; that of the larger ellipsoid is 1.84 mm.

Next, the fixed sensor estimated the pose of the implant target (${}^{Optotrak}_{ImpTarg}H$) with respect to the sensor, along with the corresponding covariance matrix. Using the known pose of the implant with respect to the implant target (${}^{ImpTarg}_{Implant}H$), the pose of the implant with respect to the sensor was estimated, using ${}^{Optotrak}_{Implant}H = {}^{Optotrak}_{ImpTarg}H \; {}^{ImpTarg}_{Implant}H$, along with its covariance matrix.

Finally, the pose of the implant with respect to the HMD was estimated via. ${}^{Hmd}_{Implant}H^{(opto)} = {}^{Hmd}_{Optotrak}H \; {}^{Optotrak}_{Implant}H$. The covariance matrix of this pose was estimated using Equation 9. The corresponding ellipsoid is shown in Figure 5 (right). The major axis of this ellipsoid is 8.23 mm. Note that the shape of this ellipsoid is elongated in the plane perpendicular to the line of sight, due to the orientation uncertainty in the HMD.

4.3.2 Pose Estimation Using Head-Mounted Sensor

Using data from the head-mounted camera, we estimated the pose of the camera target (${}^{Camera}_{CamTarg}H$) 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 (${}^{CamTarg}_{Implant}H$), the pose of the implant with respect to the camera was estimated, via ${}^{Camera}_{Implant}H = {}^{Camera}_{CamTarg}H \; {}^{CamTarg}_{Implant}H$. The covariance matrix of the resulting pose was also estimated. The ellipsoids corresponding to the translational uncertainties are shown in Figure 6 (left). The major axis of the ellipsoid corresponding to ${}^{Camera}_{CamTarg}H$ is 24.6 mm. The major axis of the ellipsoid corresponding to the derived pose, ${}^{Camera}_{Implant}H$, is 19.9 mm.

Note the large uncertainty of ${}^{Camera}_{CamTarg}H$ 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. 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.

Next, the pose of the implant with respect to the HMD is estimated, via. ${}^{Hmd}_{Implant}H^{(cam)} = {}^{Hmd}_{Camera}H\ {}^{Camera}_{Implant}H$. The covariance matrix of this pose was estimated using Equation 9. The corresponding ellipsoid is shown in Figure 6 (right). The major axis of this ellipsoid is 19.9 mm.

4.3.3 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. Using Equation 10, we produce a combined estimate of the implant-to-HMD pose, along with the covariance matrix. The ellipsoids corresponding to the three poses, ${}^{Hmd}_{Implant}H^{(opto)}$, ${}^{Hmd}_{Implant}H^{(cam)}$, and ${}^{Hmd}_{Implant}H^{(hybrid)}$ are shown in Figure 7. Note that the large ellipsoids, corresponding to ${}^{Hmd}_{Implant}H^{(opto)}$ and ${}^{Hmd}_{Implant}H^{(cam)}$, are nearly orthogonal. The ellipsoid corresponding to the combined pose, ${}^{Hmd}_{Implant}H^{(hybrid)}$, is much smaller and is contained within the intersection volume of the larger ellipsoids. The right image of Figure 7 is a wire-frame rendering of the ellipsoids, which allows the smaller interior ellipsoid to be seen. The major axis corresponding to the uncertainty of the combined pose is only 1.47 mm.

5 Summary

This paper has developed a methodology to explicitly 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 methodology was applied to an actual experimental augmented reality system. A typical configuration was analyzed and it was shown that the hybrid system produces a pose estimate that is significantly more accurate than that produced by either sensor acting alone.

6 Acknowledgments

The author would like to thank Dr. Tyrone Vincent for many helpful discussions, Khoi Nguyen for implementing many of the components of the

experimental augmented reality system, and the anonymous reviewers for their helpful comments.

7 References

[1] R. T. Azuma, "A Survey of Augmented Reality," *Presence*, Vol. 6, No. 4, pp. 355-385, 1997.

[2] R. Azuma and G. Bishop, "Improving static and dynamic registration in an optical see- through HMD," *Proc. of 21st International SIGGRAPH Conference*, ACM, pp. 197-204, 1994.

[3] J.-F. Wang, et al, "Tracking a Head-Mounted Display in a Room-Sized Environment with Head-Mounted Cameras," *Proc. of Helmet-Mounted Displays II*, Vol. 1290, SPIE, pp. 47-57, 1990.

[4] D. Kim, S. W. Richards, and T. P. Caudell, "An optical tracker for augmented reality and wearable computers," *Proc. of IEEE 1997 Annual International Symposium on Virtual Reality*, pp. 146-50, 1997.

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

[6] J. Craig, *Introduction to Robotics, Mechanics, and Control*, 2nd ed., Addison Wesley, 1990.

[7] B. K. P. Horn, "Closed-form solution of absolute orientation using unit quaternions," *J. Optical Soc. of America*, Vol. 4, No. 4, pp. 629-642, 1987.

[8] H. L. V. Trees, *Detection, Estimation, and Modulation Theory*, New York, Wiley, 1968.

[9] W. A. Hoff, T. Lyon, and K. Nguyen, "Computer Vision-Based Registration Techniques for Augmented Reality," *Proc. of Intelligent Robots and Computer Vision XV*, Vol. 2904, SPIE, pp. 538-548, 1996.

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

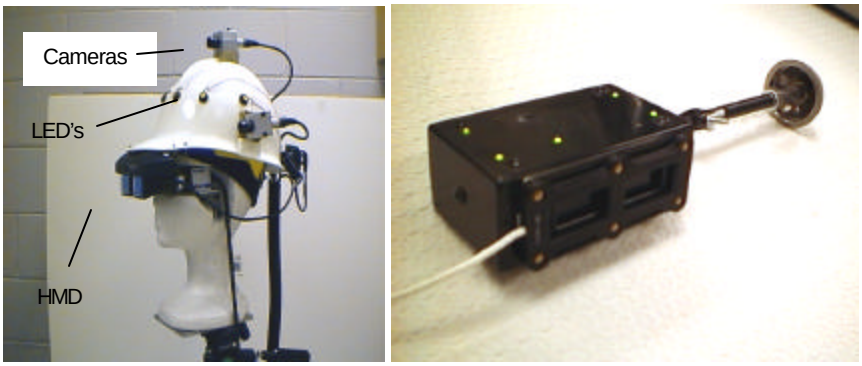

Figure 1 (Left) Prototype augmented reality system. Only one of the cameras was used. (Right) Five green LED's (top surface) form the optical target for the head mounted camera. Six infrared LED's (front surface) form an optical target for the Optotrak sensor. Both targets are mounted on a box, which is attached to a hip implant component.

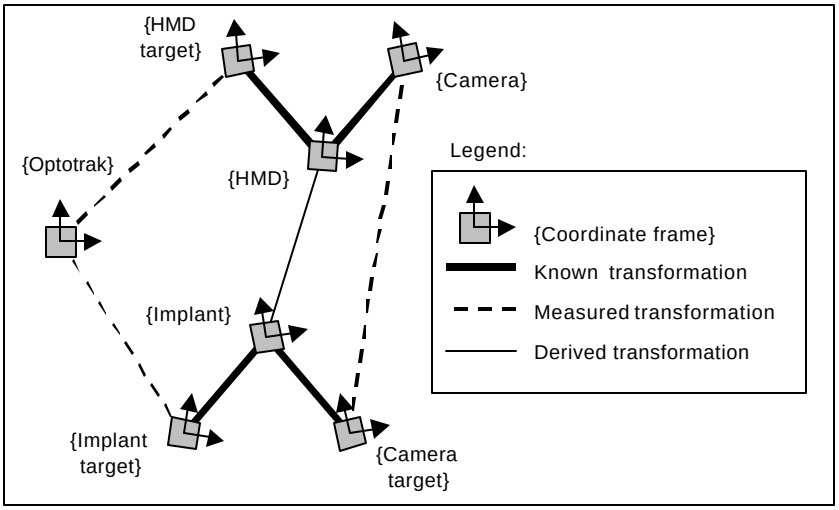

Figure 2 The principal coordinate frames in the system are shown, along with the transformations between them.

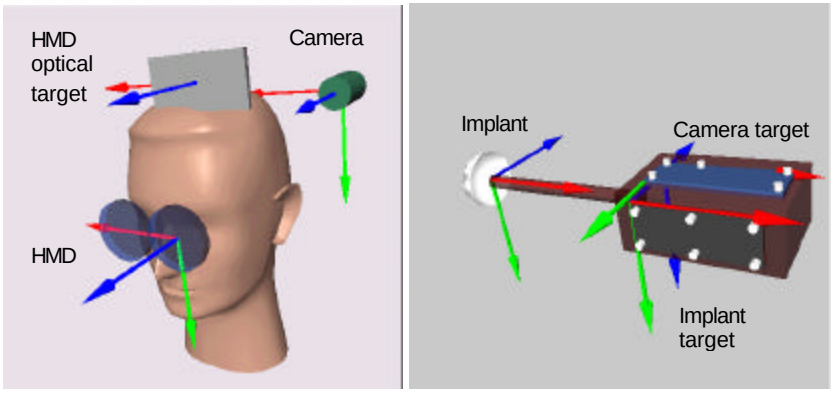

Figure 3 The coordinate frames on the head (left) and on the implant (right).

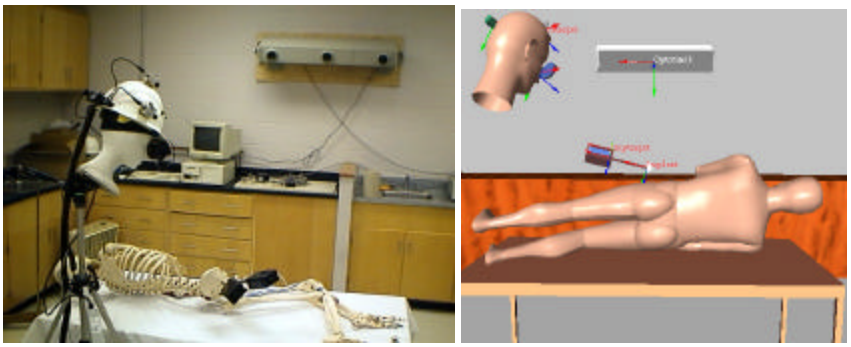

Figure 4 A visualization of the entire scene, showing the fixed sensor on the wall, the HMD, and the hip implant. (Left) The real scene. (Right) A 3D visualization.

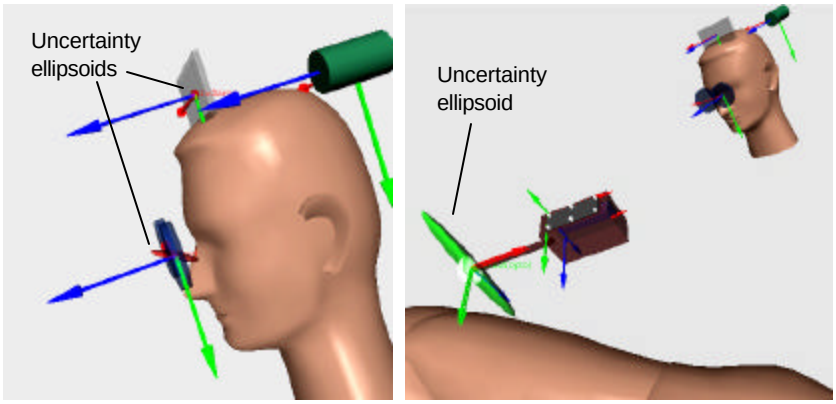

Figure 5 Uncertainties of poses derived from fixed sensor: (Left) HMD target (small ellipsoid) and HMD (large ellipsoid). (Right) Implant with repsect to HMD.

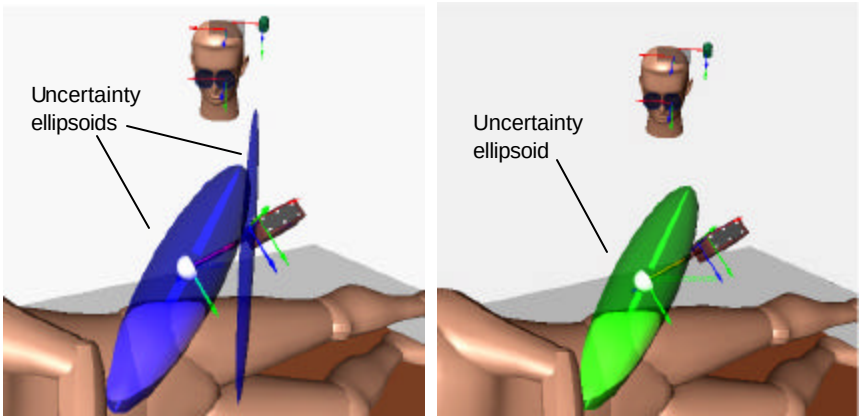

Figure 6 Uncertainties of poses derived from head camera: (Left) Camera target (long narrow ellipsoid) and implant with respect to camera (wide ellipsoid). (Right) Implant with respect to HMD.

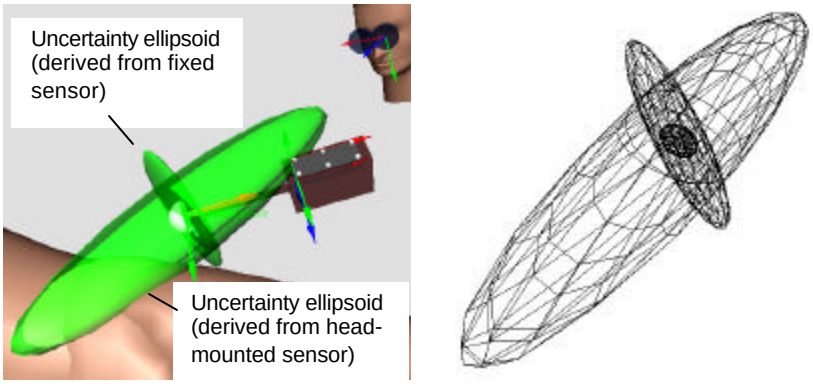

Figure 7 (Left) This figure depicts the fusion of the data. Note that the ellipsoids from the fixed sensor and the head-mounted sensor are nearly orthogonal. The ellipsoid corresponding to the resulting pose estimate is much smaller and is contained in the volume of intersection. (Right) This wire-frame rendering of the uncertainty ellipsoids allows the smaller (combined estimate) ellipsoid to be seen, which is contained in the intersection of the two larger ellipsoids.