

An Optical-Inertial Tracking System for Fully-enclosed VR Displays

A. Hogue, M. R. Jenkin, R. S. Allison
{hogue,jenkin,allison}@cs.yorku.ca

Centre for Vision Research and Department of Computer Science
York University, Toronto, Ontario, Canada.

Abstract

This paper describes a hybrid optical-inertial tracking technology for fully-immersive projective displays. In order to track the operator, the operator wears a 3DOF commercial inertial tracking system coupled with a set of laser diodes arranged in a known configuration. The projection of this laser constellation on the display walls are tracked visually to compute the 6DOF absolute head pose of the user. The absolute pose is combined with the inertial tracker data using an extended Kalman filter to maintain a robust estimate of position and orientation. This paper describes the basic tracking system including the hardware and software infrastructure.

1. Introduction

For a virtual reality system to provide a realistic visual display to the user, it is often necessary to know the location and orientation of the user’s head in order to project the correct images on all sides of the display. If this is not done correctly, the user is more likely to experience discomfort (headaches, nausea, disorientation; symptoms collectively known as cybersickness[13]). In non-fully-enclosed displays, it is possible to use commercial head tracking systems since the tracking equipment can be positioned in such a way that it does not interfere with the user’s view of the scene (i.e. behind the user). Such an approach is not possible in a fully enclosed environment. Tracking a user within a fully-enclosed projection-based display such as COSMOS[16, 3], HyPi-6[6, 10], PDC VR-CUBE[2], C6[7], ALICE[12], and IVY[11] is a more complex task. The user is fully-enclosed in the display volume, and there is no reasonable place for visible tracking equipment as it interferes with the display’s immersive effect.

Given this constraint, the tracking technology of choice for current fully-enclosed displays is electro-

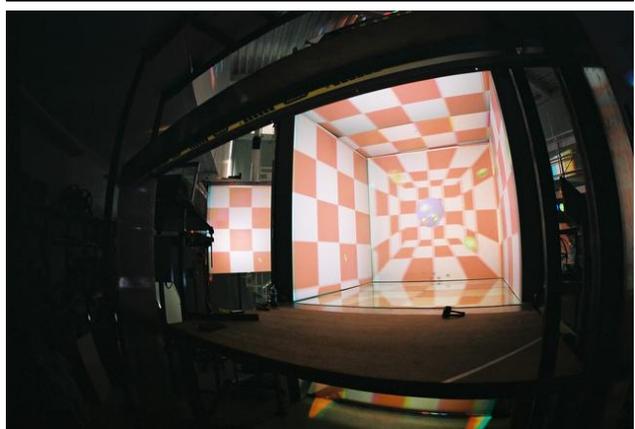


Figure 1. IVY: The Immersive Visual environment at York. IVY is shown here with the rear (entry) wall removed in order to show the structure of the device more clearly.

magnetic technology. However, electromagnetic tracking systems behave poorly in the presence of metallic objects[8, 9], accuracy and signal strength degrades with distance from the base emitter and the user is typically tethered to the equipment. Although various tracking technologies have been tried in fully immersive environments, no completely effective tracking technology currently exists for fully-enclosed immersive environments. Although the nature of the fully-enclosed environment limits the applicability of existing tracking technologies, the fully-enclosed nature of the hardware can also be exploited in order to track the user. Specifically, the surface that surrounds the user can be used as a screen upon which an optical tracking system can be deployed.

2. Hybrid Tracking Approach

In order to overcome the limitations of existing magnetic trackers, we have developed a novel “outside in”

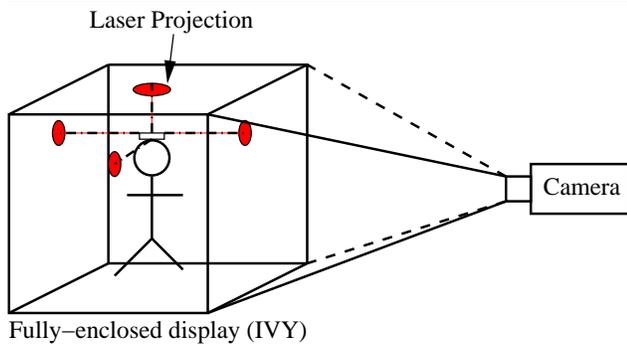
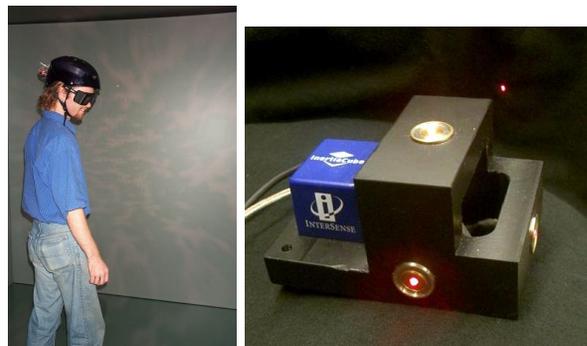


Figure 2. Optical Tracking Approach. The user wears low power laser diodes whose projections on the screen surfaces are tracked via cameras outside of the display. The head pose is determined from these projections alone.

optical tracking system (briefly described in [11]) for tracking users within a fully enclosed projective immersive environment and have coupled it with a commercial 3DOF inertial head tracker to maintain tracking in the presence of fast head motions. The optical tracker utilizes commercial cameras and computers, is capable of obtaining 6 DOF pose estimates of the user within the environment at 15-20Hz in its current implementation, and can be used as either a standalone tracking system or as part of a hybrid optical-inertial tracking system[5].

The basic idea (see Figure 2 for an illustration) is to use the projective surfaces outside of the view of the user to track the user within the environment. A fixed arrangement of low power laser diodes is attached to a helmet worn by the user (see Figure 3). Cameras are positioned behind the screens such that they can view the entire projection surface. By tracking the projections of the laser beams as they strike the projective surfaces, we are able to exploit the known geometry of the lasers, apply constraints to the pose of the device, and are thus able to compute and track the user’s correct head pose. The absolute pose information obtained by the vision system is combined with relative inertial data via the inertial sensor. An extended Kalman filter[15] is employed to predict, gate, smooth, and integrate the inertial and optical pose estimates to obtain the final pose of the user.

Various configurations of laser diodes could be used to localize the user. Our implementation uses a simple arrangement of four laser diodes in the geometric configuration shown in Figure 4. Two of the laser diodes are arranged to project in opposite directions along a



(a) (b)

Figure 3. User with helmet (a) and hybrid inertial-optical tracking device (b). Note that the lasers are mounted behind the user so that the laser beams strike walls outside of the user’s view and cannot be seen.

single line, and the other two diodes are arranged so that they project orthogonal to each other and orthogonal to this line. The projection directions of all four laser diodes intersect at a single point, P_0 . Given the projections of the four laser diodes on the exterior walls of the environment it is possible to obtain strong constraints on P_0 and to define a unique 3D coordinate frame centered at this point.

To demonstrate this, we break the problem down into two parts. The first is to determine P_0 and the coordinate system aligned with P_0 given that one can identify the three-dimensional position at which the beam from *specific* diodes strike the various walls, and the second is to determine which laser spot on a wall corresponds to which laser emitter.

For the remainder of this discussion, $P_1...P_4$ are the 3D positions at which the laser beams from the respective laser diodes strike the walls of the environment. P_0 lies at the intersection of P_1P_2 with a perpendicular line that passes through point P_3 . The point P_0 can be found quite easily by noting that P_0 lies along the line defined by $P_1 + \lambda(P_2 - P_1)$ and $P_1P_2 \cdot P_0P_3 = 0$. Solving these equations for P_0 yields

$$P_0 = P_1 + \frac{(P_3 - P_1) \cdot (P_2 - P_1)}{\|P_2 - P_1\|^2} (P_2 - P_1)$$

This defines the origin of the frame, P_0P_3 defines the forward direction vector for the frame, and the normal of the plane is defined by points P_1, P_2, P_3 ; $\vec{n} = P_0P_1 \times P_0P_3$, which determines the direction of the up

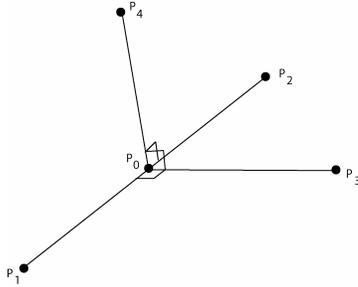


Figure 4. Basic laser geometry. The four lasers are established so that lines drawn through their beams would intersect at a common point P_0 , and $P_3P_0 \cdot P_1P_2 = P_4P_0 \cdot P_1P_2 = P_3P_0 \cdot P_4P_0 = 0$.

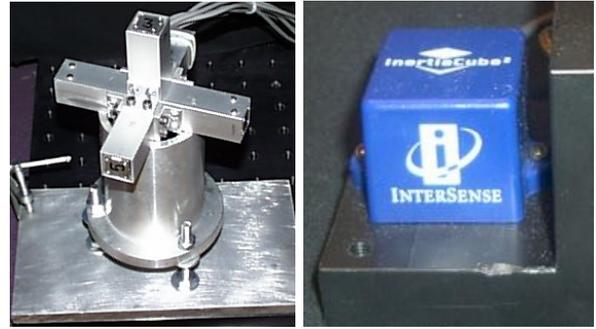
vector. Although P_4 is not required in order to compute this frame (provided that the assignment of laser spots to diodes is known), P_4 will prove useful by providing an additional constraint in terms of determining the appropriate mapping from laser spots to emitters. In terms of the geometry it is important to note that P_0P_4 is perpendicular to the plane defined by points $P_1P_2P_3$.

These calculations assume that we know the correspondence between each laser diode and each laser projection. In practice this may be accomplished using different wavelengths, or by pulsing the lasers at known frequencies. In our current implementation we take a more algorithmic approach and use the geometry to place constraints on the finite number of possible emitter-laser spot labellings.

We must determine the appropriate labelings of the tracked laser projections P_i, P_j, P_k , and P_l with the actual laser points P_1, P_2, P_3 , and P_4 . There are 24 possible assignments of the laser points to the emitters. Of all 24 possible assignments, only four are consistent with the geometry of the emitters[5]. Figure 6 shows examples of the possible labelings and the impact this has on the pose computation.

Although there are four configurations that are consistent with the geometry of the laser diodes, the three incorrect assignments are sufficiently distant from the correct pose to be easily disambiguated using temporal coherence. If the correct assignment is $(P_i, P_j, P_k, P_l) \rightarrow (P_1, P_2, P_3, P_4)$, then the three incorrect assignments are

1. $(P_i, P_j, P_k, P_l) \rightarrow (P_2, P_1, P_4, P_3)$. This configuration has the same P_0 as the correct configuration, but is rotated by 180 degrees. With a 15Hz sampling rate, the user would have to rotate at roughly 1350 deg/sec before this configuration can be confused with the correct one.



(a)

(b)

Figure 5. (a) The original inertial device composed of six accelerometers (b) The Inertiacube² from Intersense®

2. $(P_i, P_j, P_k, P_l) \rightarrow (P_3, P_4, P_2, P_1)$. This incorrect assignment and the final remaining assignment have a different P_0 , and an orientation change of at least 90 degrees. This configuration, like the following configuration, is extremely unstable and can only occur under extremely unusual conditions[5]. With a 15Hz sampling rate, the user would have to rotate at roughly 675 deg/sec before this configuration can be confused with the correct one.
3. $(P_i, P_j, P_k, P_l) \rightarrow (P_4, P_3, P_1, P_2)$. This incorrect assignment is similar to the one above. It has a different P_0 as well as at least a 90 degree orientation change.

A simple temporal tracking system coupled with gating is used to discard these incorrect assignments. Although these constraints allow us to keep a consistent pose, there is still an issue of estimating the initial pose. In our current implementation, the initial correct assignment is chosen manually. Limiting the tracked rotation to less than 500 deg/sec eliminates the three incorrect assignments. Inertial data can also be used to aid in disambiguation.

3. Implementation Details

We previously experimented with a 6DOF inertial system[5] (See Figure 5) comprised of six linear accelerometers and used differential measurements from the accelerometers to estimate the angular accelerations. Unfortunately due to calibration issues the pose estimates were valid only for a fraction of a second

before large errors were accumulated. We then opted to purchase a commercially available 3DOF orientation tracker from Intersense®. The InertiaCube² has a raw data update rate of 180Hz, 0.01° angular resolution and 1° RMS angular accuracy. This has resulted in a considerable improvement in terms of performance of the inertial system. The optical system is comprised of eight digital FirewireTM cameras (capturing 640x480 resolution grayscale images) situated outside the immersive display aimed at each of the rear-projection screens which allow us to track the multiple laser projections. Each camera is equipped with an optical wavelength bandpass filter with a peak response at 650nm (the laser diode wavelength). This allows us to simplify the image processing routine speeding up the response of the tracking system as a whole. To find the centroid of the laser dot in each image, we employ a sub-pixel peak detector[1].

Calibration of the optical system is performed offline. The transformation between the screen surface and each camera is modeled as a 2D planar homography and is computed using the Discrete Linear Transform algorithm described by Hartley and Zisserman[4]. The final calibration step is to determine the relationship between each of the screen surfaces to the world coordinate system. This is accomplished by defining the world coordinate system as the center of the display and physically measuring the rigid-body transformation separately for each screen.

3.1. Discarding Invalid Configurations

Once the tracked 2D laser points are available, it is possible to determine the pose of the object being tracked. This is done using the above computation for each possible labeling of the four laser dots (24 possibilities). In order to determine which is the correct labeling, we impose geometric constraints on the solution and compute an error function, $\epsilon(i)$, that allows us to determine which labeling is the correct one. Using three constraints, we are able to determine the correct solution (up to a reflection, see Figure 6(b)) that corresponds properly to the pose of the device. Define

$$\epsilon(i) = \alpha\epsilon_{\perp}(i) + \beta D(i) + \gamma\mathcal{F}(i) \quad (1)$$

where i is the current permutation of laser points, α , β , and γ are weightings denoting the importance of each error function (in the results reported here, for simplicity $\alpha = \beta = \gamma = 1$), $\epsilon_{\perp}(i)$ (the perpendicular error) is the sum of the dot products of vectors that should be perpendicular in this configuration (we take the absolute value of each dot product to ensure an increasing function), and $D(i)$ is the distance between the com-

puted position P_0^i and $P_0^{i'}$ where $P_0^{i'}$ is computed using $P_4^i P_2^i P_1^i$ rather than $P_3^i P_2^i P_1^i$. This eliminates many of the incorrect labelings since P_0^i and $P_0^{i'}$ will not coincide if the plane normal is computed with the incorrect points. $\mathcal{F}(i)$ is a binary function that is 1 only when our computed plane normal is in the same direction as P_4^i . The perpendicular error, $\epsilon_{\perp}(i)$ is defined as

$$\epsilon_{\perp}(i) = |(P_0^i P_4^i \cdot P_0^i P_1^i)| + |(P_0^i P_4^i \cdot P_0^i P_2^i)| + |(P_0^i P_4^i \cdot P_0^i P_3^i)| + |(P_0^i P_1^i \cdot P_0^i P_3^i)| + |(P_0^i P_2^i \cdot P_0^i P_3^i)| \quad (2)$$

$D(i)$ is defined as

$$D(i) = \|P_0^i - P_0^{i'}\|^2 \quad (3)$$

and

$$\mathcal{F}(i) = \begin{bmatrix} 1 & \vec{n} \cdot (P_4^i - P_0^i) > 0 \\ 0 & \text{otherwise} \end{bmatrix} \quad (4)$$

After evaluating $\epsilon(i)$ for each possible labeling, the results are sorted in ascending order according to this error function and the first 2 solutions are taken as the correct pose and reflection. Although there exist four possible solutions, two of these do not occur in practice as they correspond to extreme head positions/orientations within the environment, and are extremely unstable. It is still necessary to distinguish between the final two solutions.

This is accomplished by applying a temporal coherence property on possible body rotational velocity. Given the correct quaternion at the previous time step, \hat{q}_{t-1} , and the two possible orientations \hat{q}_1, \hat{q}_2 , the error quaternions can be computed by

$$\hat{q}_{\epsilon 1} = \hat{q}_{t-1} \hat{q}_1^{-1} \quad (5)$$

$$\hat{q}_{\epsilon 2} = \hat{q}_{t-1} \hat{q}_2^{-1} \quad (6)$$

and the correct orientation is determined as the error quaternion with the smallest associated rotation angle, i.e. take the orientation with the smallest $\cos^{-1}(\hat{q}_{\epsilon i}[0])$.

4. Results

Figure 7 shows basic tracking within IVY. Here, the tracker was moved by hand in the horizontal plane along various simple trajectories. Although these results demonstrate that the system can track a moving target within IVY, they do not speak to the accuracy of the tracking system. Tracking accuracy for orientation and position is addressed in the following sections.

4.1. Orientation

The laser diode housing was placed on a rotational stage in the center of IVY roughly 4' above the floor

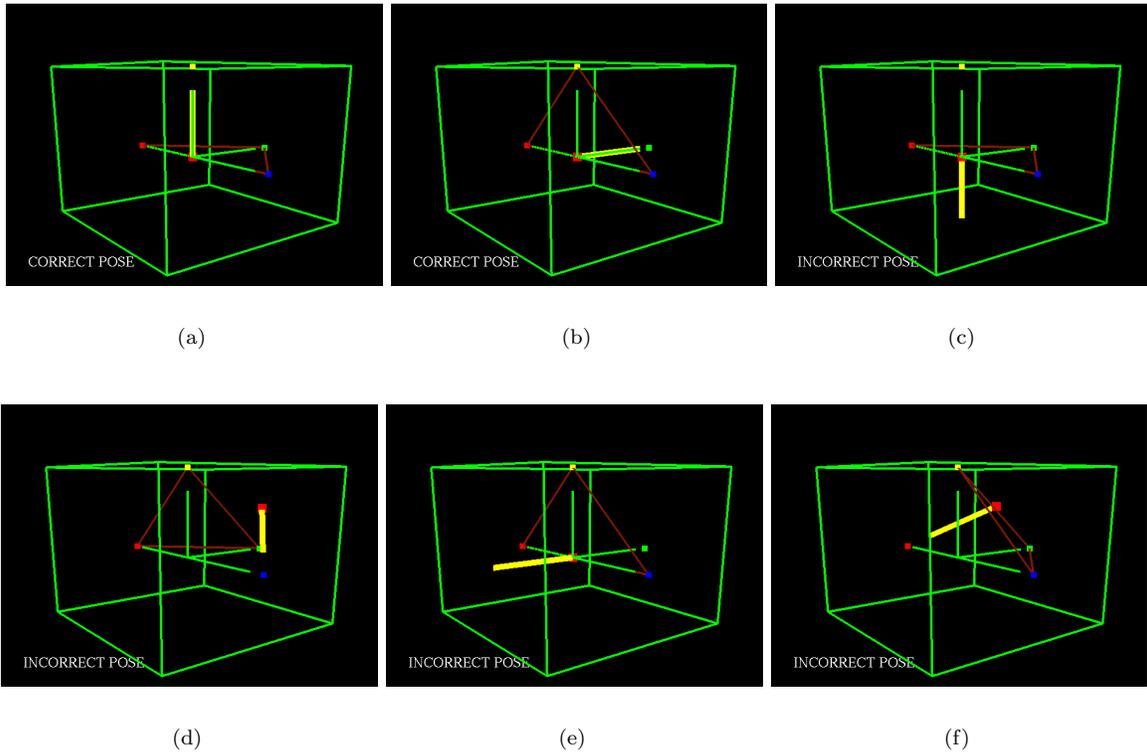


Figure 6. Six examples of the 24 possible labelings and their associated computed pose. Shown here are screenshots from a simulator designed to test the configuration constraints on the laser geometry. The surrounding cube is an analogue of IVY while the smaller dots on the sides of the cube are the associated laser projections. The thick line (shown in yellow) is the computed Up vector, and the computed position is the large dot (shown in red). The connecting lines between laser points indicate which lasers were used to compute the plane for orientation. Each image is labeled with the text “CORRECT POSE” or “INCORRECT POSE” which is automatically computed using only static constraints. (a) is the correctly computed pose while (b) is incorrect but cannot be distinguished using only static constraints. Note that (b) is actually the correct pose rotated by 180 degrees around an oblique axis. A simple temporal mechanism is used to distinguish between these two solutions.

that allowed us to rotate the device precisely at 1° intervals. Figure 8(a) shows the raw data points for a full 360° rotation on the azimuth at 5° intervals. For each direction vector, points on the unit circle are drawn at both the measured and correct orientation in the same colour (note that due to the accuracy of the measurement these points appear almost coincident). A fundamental limitation of this tracking system within a cube-shaped immersive display can be seen when the lasers shine into the corners. At this time no data can be collected since the laser projections cannot be seen, and tracking is lost until the lasers shine onto the screen. Note that these gaps can be filled in by using the secondary inertial system. In a second orientation experiment, rotational data were collected over a 10 degree range at 1 degree intervals on the azimuth. The rela-

tive angle, shown in Table 1, was computed between the direction vectors X-Z components and the first reported direction vector. The mean error of this exercise was approximately 0.1° while the max error was approximately 0.3° .

4.2. Position

Several experiments were performed to evaluate the position reported from the tracking system. To estimate the accuracy of the position estimates, we placed the device at 20 known locations within IVY and recorded the tracker output. The raw data in this test is illustrated in Figure 8(b). The mean absolute position error was modest at 1.13cm but there were several cases where the error was nearly 5cm. We believe

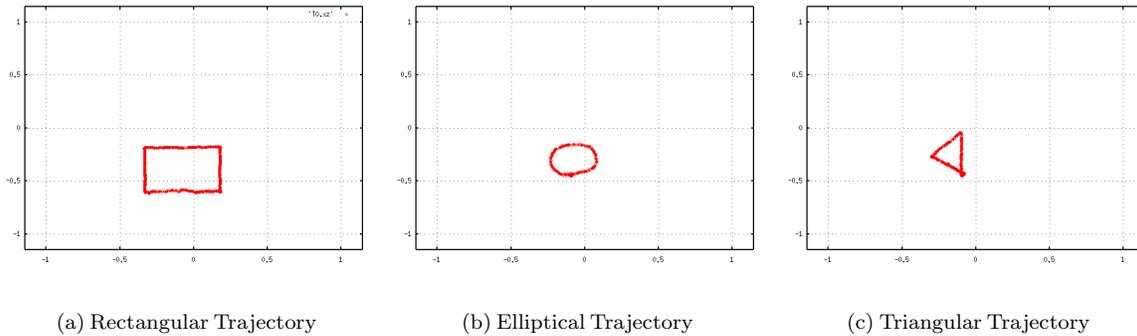


Figure 7. Basic Tracking Results. The above graphs were created by moving the tracker by hand with different trajectories on the XZ plane.

Rotational Stage	Computed Angle
0°	0.0000°
1°	0.9229°
2°	1.9101°
3°	3.2703°
4°	4.1654°
5°	5.0992°
6°	6.2851°
7°	7.0167°
8°	8.3210°
9°	9.1814°
10°	9.8664°

Table 1. Computed angles between the reported direction vectors at 1° increments.

the maximum error in the system is due to the off-axis placement of one of the ceiling cameras which as a result has a large perspective distortion and much lower screen resolution along one side of the immersive display.

The noise covariance of each position estimate was also computed using a Linear Kalman filter with variance of 1cm^2 on position (A typical example can be seen in Figure 9(a)). The small covariance (approximately 0.5cm) in the position is attributable to the noise in each laser position estimate due to the limited resolution of the cameras. Since we are acquiring 640x480 resolution images from the cameras, the 2.29m screen is imaged at approximately 500 pixels, making 1 camera pixel correspond to approximately 0.5cm on the screen surface. Using higher resolution images would increase the precision of the tracking system since it would allow us to make better estimates of the laser po-

sitions. Since the walls of the display are fabric walls, the screens vibrate and move slightly when in the presence of large motion. Concerned as to how this would affect the position estimate, we placed the device in a stationary position and recorded data while violently moving the screen fabric on all walls. The covariance of the estimate can be seen in Figure 9(b). The system reacts well with a spread of approximately 1.5cm even in the presence of large motion of the screen surfaces.

5. Summary and future work

The optical tracking approach for fully immersive environments presented here has many advantages over existing approaches. The accuracy achieved is not a function of the distance of the user from a base station. The system performance is not degraded by metallic objects or other interference. The user is untethered, which is inherent to the laser system since it is fully self contained and the inertial system is untethered from a base computer due to the use of a hand-held PDA and a wireless link to obtain the inertial estimates. Also, the user is not required to wear a large encumbering device which could compromise their immersive experience. Using off-the-shelf FireWire® digital video cameras allows the tracking system to evolve with the commercial market making it possible to increase the resolution and framerate as new camera technology becomes available. Our current implementation of the optical system is limited to approximately 15Hz due to computing power and camera limitations while work continues on increasing this to 30Hz. Increasing the framerate would have a positive impact on the system performance. It would be easier to disambiguate the invalid pose estimates since it would limit

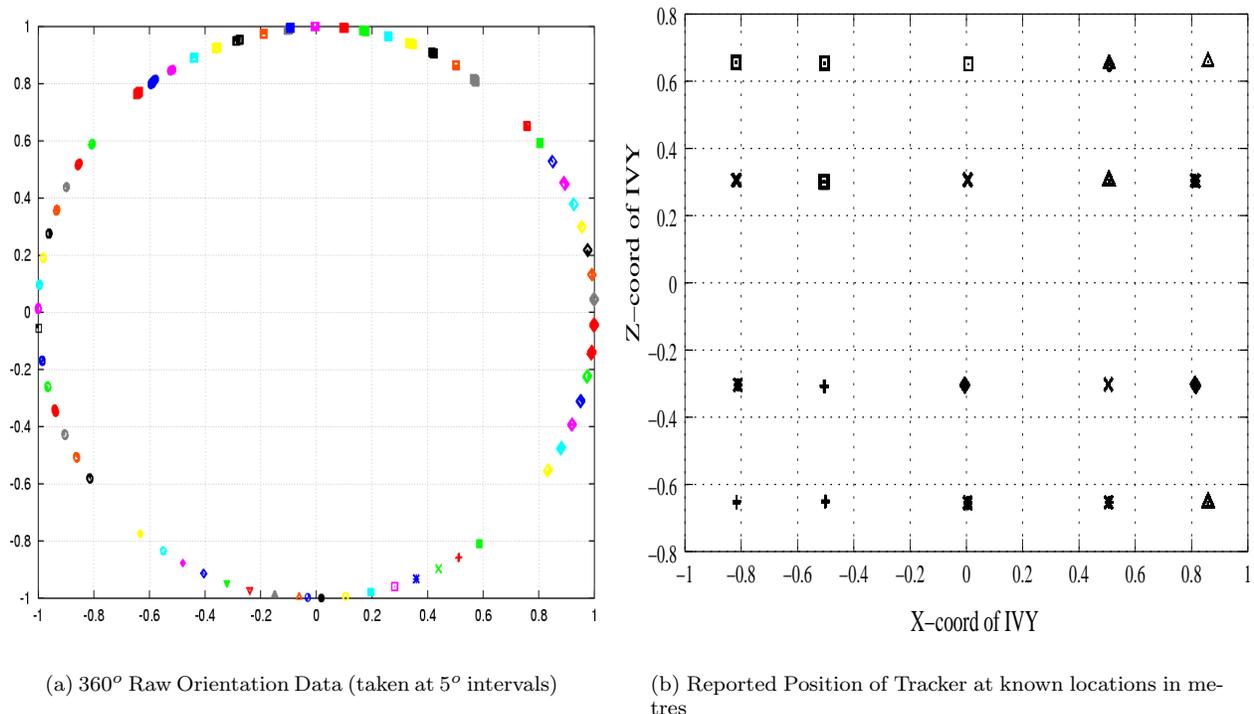


Figure 8. (a) Optical Tracker Orientation Experiment. Unit vectors are plotted, with the same symbol, in the recovered and measured directions. The plotted X-axis is the X-coordinate of the unit vector and the plotted Y-axis is the Z-coordinate of the unit vector. Note: the four large holes indicate positions where the lasers were shining into the corners of IVY and thus could not be tracked. (b) Optical Tracker Position Experiment. Different symbols are used only to distinguish different measurements.

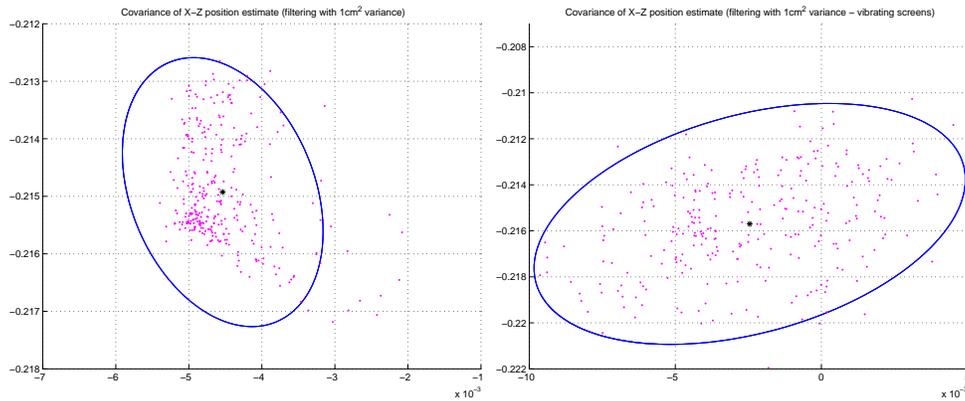
the amount of motion that could occur between updates even more. Work progresses on making the system more robust when the measurements are unreliable or unavailable, e.g. occluded laser dots and lasers projecting into corners make one or more laser projections unavailable and thus increase the delay between updates in this configuration (and increase the chance of choosing an incorrect pose). We are currently developing a SCAAT[14] tracking filter which will allow the user to be tracked consistently even if a laser dot measurement is unavailable for a given frame. Using this algorithm should also decrease the total latency of the system since we will not need to wait until four laser measurements are available to estimate the pose. Although a complete end-to-end latency analysis has not yet been performed, minimum system latency – due to camera capture, initial data processing etc. – is approximately 0.035s.

Acknowledgements: We would like to acknowledge Matt Robinson for helping to create the software infras-

tructure, Jeff Laurence and James Zacher and Andrew German for helping us build the hardware, and Urszula Jasiobedzka for her help and support. The financial support of NSERC Canada, the Canadian Foundation for Innovation, SGI, Precarn, IRIS, and the Centre for Research in Earth and Space Technology (CRESTech) is gratefully acknowledged.

References

- [1] R. B. Fisher and D. K. Naidu. *A Comparison of Algorithms for Subpixel Peak Detection*. Springer-Verlag, Heidelberg, 1996. <http://citeseer.nj.nec.com/482699.html>.
- [2] Center for Parallel Computing. Primeur: Advancing European Technology Frontiers, World's first fully immersive VR-CUBE installed at PDC in Sweden, 1998.
- [3] K. Fujii, Y. Asano, N. Kubota, and H. Tanahashi. User interface device for the immersive 6-screens display "COSMOS". In *Proceedings 6th international conference on virtual systems and multimedia (VSMM'00)*, Gifu, Japan, Oct. 4-6 2000.



(a) Typical Noise Covariance of Stationary X-Z position in metres (filtering with 1cm^2 variance). This shows a spread of 6mm on the Z-axis and 3mm on the X-axis.

(b) Typical Noise Covariance of Stationary X-Z position in metres (filtering with 1cm^2 variance). This shows that even in the presence of large motion of the screen surfaces (screen movement due to fast motion within IVY), the spread of the measurement is 1.5cm on the X-Axis and 1cm on the Z-axis.

Figure 9. Results from collected data. (a) shows the typical noise covariance of a stationary position, and (b) shows how the noise increases when the fabric screens vibrate due to fast motion.

- [4] R. Hartley and A. Zisserman. *Multiple View Geometry*. Cambridge University Press; ISBN:0521623049, 2000.
- [5] A. Hogue. MARVIN: a Mobile Automatic Realtime Visual and INertial tracking system. Master's thesis, York University, March 2003.
- [6] Fraunhofer Institute IAO. <http://vr.iao.fhg.de/6-Side-Cave/index.en.php>.
- [7] Virtual Reality Applications Center Iowa State University. <http://www.vrac.iastate.edu/about/labs/c6>.
- [8] V. Kindratenko. A survey of electromagnetic position tracker calibration techniques. In *Virtual Reality: Research, Development, and Applications*, 2000. vol.5, no.3, pp. 169-182.
- [9] V. Kindratenko. A comparison of the accuracy of an electromagnetic and hybrid ultrasound-inertia position tracking system. *Presence: Teleoperators and Virtual Environments*, 10(6):657-663, 2001.
- [10] I. Rötzer. Fraunhofer Magazine, Synthetic worlds within six walls 2:2001.
- [11] M. Robinson, J. Laurence, J. Zacher, A. Hogue, R. Allison, L. R. Harris, M. Jenkin, and W. Stuerzlinger. IVY: The Immersive Visual environment at York. In *6th International Immersive Projection Technology Symposium, March 24-25, 2002, Orlando, Fl.*, 2002.
- [12] Beckman Institute University of Illinois, Integrated Systems Laboratory. A laboratory for immersive cognitive experiments. http://www.isl.uiuc.edu/Virtual%20Tour/Tour-Pages/meet_alice.htm.
- [13] Joseph J. Jr. La Viola. A discussion of cybersickness in virtual environments. In *SIGCHI Bulletin*, January 2000. vol.32 no.1, pp.47-56.
- [14] G. Welch. *SCAAT: Incremental Tracking with Incomplete Information*. PhD thesis, Chapel Hill, NC, 1996.
- [15] G. Welch and G. Bishop. An introduction to the kalman filter. Technical Report TR95-041, University of North Carolina at Chapel Hill, 1995.
- [16] T. Yamada, M. Hirose, and Y. Isda. Development of a complete immersive display: COSMOS. In *Proc. VSMM'98*, pages 522-527, 1998.