Tracking in Unprepared Environments for Augmented Reality Systems

Ronald Azuma
HRL Laboratories
3011 Malibu Canyon Road, MS RL96
Malibu, CA 90265-4799, USA
azuma@HRL.com

Jong Weon Lee, Bolan Jiang, Jun Park,
Suya You, and Ulrich Neumann
Integrated Media Systems Center
University of Southern California
Los Angeles, CA 90089-0781, USA
suyay@graphics.usc.edu

ABSTRACT

Many Augmented Reality applications require accurate tracking. Existing tracking techniques require *prepared* environments to ensure accurate results. This paper motivates the need to pursue Augmented Reality tracking techniques that work in *unprepared* environments, where users are not allowed to modify the real environment, such as in outdoor applications. Accurate tracking in such situations is difficult, requiring hybrid approaches. This paper summarizes two 3DOF results: a real-time system with a compass – inertial hybrid, and a non-real-time system fusing optical and inertial inputs. We then describe the preliminary results of 5- and 6-DOF tracking methods run in simulation. Future work and limitations are described.

MOTIVATION

An Augmented Reality (AR) display system superimposes or composites virtual 3-D objects upon the user's view of the real world, in real time. Ideally, it appears to the user as if the virtual 3-D objects actually exist in the real environment [1]. One of the key requirements for accomplishing this illusion is a tracking system that accurately measures the position and the orientation of the observer's location in space. Without accurate tracking, the virtual objects will not be drawn in the correct location and the correct time, ruining the illusion that they coexist with the real objects. The problem of accurately aligning real and virtual objects is called the *registration* problem.

The best existing Augmented Reality systems are able to achieve pixel-accurate registration, in real time; see [15] for one example. However, such systems only work indoors, in *prepared* environments. These are environments where the system designer has complete control over what exists in the environment and can modify it as needed. For traditional AR applications, such as medical visualization and display manufacturing instructions, this assumption is reasonable. But many other potential AR applications would become feasible if accurate tracking was possible in *unprepared* environments. Potential users include hikers navigating in the woods, soldiers out in the field, and drivers operating vehicles. Users operating outdoors, away from carefully prepared rooms, could use AR displays for improved situational awareness, navigation, targeting, and information selection and retrieval. AR interfaces might be a more natural means of controlling and interacting with wearable computers than the current WIMP standard.

Beside opening new application areas, tracking in unprepared environments is an important research direction because it will reduce the need to prepare environments, making AR systems easier to set up and operate. Compared to Virtual Environment systems, AR systems are rarely found outside research laboratories. Preparing an environment for an AR system is hard work, requiring a significant amount of measurement and calibration. If AR systems are to become more commonplace, they must become easier for end users to set up and operate. Today's systems require expert users to set up and calibrate the environment and the system. If accurate tracking could be achieved without the need to carefully prepare the environment in advance, that would be a major step in reducing the difficulty of operating an AR system. The ultimate goal is for the AR system to support accurate tracking in arbitrary environments and conditions: indoors, outdoors, anywhere the user wants to go. We are far from that goal, but by moving AR systems into unprepared, outdoor environments, we take an important step in that direction.

Tracking in unprepared environments is difficult for three reasons. First, if the user operates outdoors and traverses long distances, the resources available to the system may be limited due to mobility constraints. In a prepared environment where the user stays within one room, bandwidth and CPU constraints are limited more by the budget than any physical factors. But if the user moves around outdoors, especially if he wears all the equipment himself, then size, weight, and power constraints all become concerns. Second, the range of operating conditions is greater than in prepared environments. Lighting conditions, weather, and temperature are all factors to consider in unprepared environments. For example, the display may not be bright enough to see on a sunny day. Visual landmarks that a video tracking system relies upon may vary in appearance under different lighting conditions or may not be visible at all at night. Third and most importantly, the system designer cannot control the environment. It may not be possible to modify the environment. For example, many AR tracking systems rely upon placing special fiducial markers at known locations in the environment; an example is [11]. However, this approach is not practical in most outdoor applications. We cannot assume we can cover the landscape with billboard-sized colored markers. Also, we may not be able to accurately measure all objects of interest in the

environment beforehand. The inability to control the environment also restricts the choice of tracking technologies. Many trackers require placing active emitters in the environment. These three differences between prepared and unprepared environments illustrate the challenge of accurate tracking in unprepared environments.

If we survey tracking technologies for how well they operate in unprepared environments, we find that no single technology will offer the required performance in the near future [2]. The Global Positioning System (GPS) can measure the position of any point on the Earth from which enough satellites can be seen. Ordinary GPS measurements have typical errors around 30 meters; differential GPS systems can reduce the typical error to around 3 meters. Carrier phase systems can achieve errors measured in the centimeters under certain conditions. However, GPS does not directly measure orientation and does not work when the user can't see enough of the sky (indoors, near buildings, in canyons, etc.) In military circumstances, GPS is relatively easy to jam. Inertial and dead reckoning sensors are self-contained, sourceless technologies. Their main problem is drift. Cost and size restrictions also limit the performance of units suitable for man-portable applications, although MEMS technologies may change that in the future. Because recovering position requires doubly integrating acceleration, getting accurate position estimates from accelerometers is extremely difficult (due to the confounding effects of gravity). Many commonly used trackers (optical, magnetic, and ultrasonic) rely on active sources, which may not be appropriate for unprepared environments. Passive optical (video tracking) is an applicable approach and can generate a full 6-D solution. It needs a clear line-of-sight. Computer vision algorithms can be brittle and computationally intensive. Electromagnetic compasses combined with tilt sensors are trackers commonly used in inexpensive Head-Mounted Displays (HMDs). They do not measure position and are vulnerable to distortions in the Earth's magnetic field, requiring extensive calibration efforts. Even in magnetically clean environments, we have measured 2-4 degree peak-to-peak distortions in a high quality electronic compass. From this analysis, we see that no single tracking technology by itself appears to offer a complete solution.

Therefore, our approach has been to develop hybrid-tracking technologies that combine multiple sensors in ways that compensate for the weaknesses of each individual component. In particular, we have been pursuing hybrids that combine optical and inertial sensors. The complementary nature of these two sensors makes them good candidates for research [16]. Our colleagues Gary Bishop, Leandra Vicci and Greg Welch at the University of North Carolina at Chapel Hill are also developing hybrid sensors of this type for this problem area.

PREVIOUS WORK

Virtually all existing Augmented Reality systems work indoors, in prepared environments. Few AR systems operate in unprepared environments where the user cannot modify or control the real world. The first known system of this type is Columbia's Touring Machine [7]. This uses commercially available sourceless orientation sensors, typically a compass and a tilt sensor, combined with a differential GPS. The most recent version uses an orientation sensor from InterSense. Developing accurate trackers has not been the focus of the Touring Machine project, and the system can exhibit large registration errors when the user moves quickly. Concurrent with our research, a group at the Rockwell Science Center has been developing a method for registration in outdoor environments that is based on detecting the silhouette of the horizon line [5]. By comparing the silhouette against a model of the local geography, the system can determine the user's current location and orientation.

Our research in developing new trackers for unprepared environments is directed by the philosophy that hybrid approaches are the only ones that offer a reasonable chance of success. Some previous AR tracking systems have used hybrid systems. [3] added rate gyroscopes to an optical tracker to aid motion prediction. [8] uses a set of sensors (rate gyroscopes and a compass and tilt sensor) that is similar to our initial base system. The differences in our results are the different mathematics to combine the sensor inputs, our distortion compensation methods, and our actual demonstration of accurate tracking in an AR system.

CONTRIBUTION

The rest of this paper summarizes results from two systems that we have built for tracking in unprepared environments and describes recent new results that tackle the position problem. The first two systems focus on the orientation component of tracking, so we call them 3 degree of freedom (3DOF) results. The initial *base system* combines a compass and tilt sensor with three rate gyroscopes to stabilize the apparent motion of the virtual objects. This system runs in real time and has been demonstrated in the field. While the registration is not perfect, typical errors are reduced significantly from using the compass by itself. Next, the *inertial-optical hybrid* adds input from a video tracker that detects natural 2-D features in the video sequence to reduce errors to a few pixels. We tested this tracker on real data, but because of computational requirements it does not yet run in real time. Finally, the 5- and 6-DOF simulations are our first steps toward actively tracking position as well as orientation (beyond solely relying upon GPS). The results demonstrate video-tracking algorithms that detect features at initially unknown locations and incorporate them into the position estimate as the user moves around a wide area in an unprepared environment.

3DOF BASE SYSTEM

For our first attempt at a hybrid tracker in an unprepared environment, we focused on a subset of the problem. We wanted a system that addresses problems within that subset and provides a base to build upon. First, we assume that all objects are distant (several hundred meters away). This allows us to rely solely on differential GPS for position tracking and focus our research efforts on the orientation problem. The largest errors come from distortions in the orientation sensor and dynamic errors caused by system latency. Therefore, the main contributions of the base system are in *calibrating* the electronic compass and other sensors and *stabilizing* the displayed output with respect to user motion. To do this, we built a hybrid tracker that combines a compass and tilt sensor with three rate gyroscopes (Figure 1).

Effectively fusing the compass and gyroscope inputs required careful calibration and development of sensor fusion algorithms. We measured significant distortions in the electronic compass, using a custom built non-magnetic turntable. Furthermore, there is a 90 millisecond delay between the measurements in the two sensors, due to inherent sensor properties and communication latencies. The compass is read at 16 Hz while the gyroscopes are sampled at 1 kHz. The filter fuses the two at the gyroscope update rate. It is not a true Kalman filter, to make the tuning easier and reduce the computational load. However, it provides the desired properties as seen in Figure 2. First, the filter output leads the raw compass measurement, showing that the gyros compensate for the slow update rate and long latency in the compass. Second, the filter output is much smoother than the compass, which is another benefit of the gyroscope inputs. Third, the filter output settles to the compass when the motion stops. Since the gyros accumulate drift, the filter uses the absolute heading provided by the compass to compensate for the inertial drift. A simple motion predictor compensates for delays in the rendering and display subsystems.

The base system operates in real time, with a 60 Hz update rate. It runs on a PC under Windows NT4, using a combination of OpenGL and DirectDraw 3 for rendering. We have run the system in four different geographical locations. Figure 3 shows a sample image of some virtual labels identifying landmarks at Pepperdine University, as seen from HRL Laboratories.

The base system is the first motion-stabilized outdoor AR system, providing the smallest registration errors of current outdoor real-time systems. Compared to using the compass by itself, the base system is an enormous improvement both in registration accuracy and smoothness. Without the benefits provided by the hybrid tracker, the display is virtually unreadable when the user moves around. However, the registration is far from perfect. Peak errors are typically around 2 degrees, with average errors under 1 degree. The compass distortion can change with time, requiring system recalibration. For more details about this system, please read [4].

3DOF INERTIAL-OPTICAL HYBRID

The next system improves the registration even further by adding video tracking, forming our first inertial-optical hybrid. Fusing these two types of sensors offers significant benefits. Integrating the gyroscopes yields a reasonable estimate of the orientation, providing a good initial guess to reduce the search space in the vision processing algorithms. Furthermore, during fast motions the visual tracking may fail due to blur and large changes in the images, so the system relies upon the gyroscopes then. However, when the user stops moving, the video tracking locks on to recognizable features and corrects for accumulated drift in the inertial tracker.

The inertial-optical hybrid performs some calibration to map the relative orientations between the compass-inertial tracker and the video camera's coordinate system. 2-D differences between adjacent images are mapped into orientation differences and used to provide corrections. The 2-D vision tracking does not rely on fiducials at known locations. Instead, it searches the scene for a set of 10-40 features that it can robustly track. These features may be points or 2-D features (Figure 4). The selection is automatic. Thus, the 2-D vision tracking is suitable for use in unprepared environments.

There are two methods for fusing the inputs from the vision and inertial trackers. The first method is to use the integrated gyro orientation as the vision estimate. This means that the inertial estimate will drift with time but it has the advantage that any errors in the vision tracking will not propagate to corrupt the inertial estimate, so this method may be more robust. The second method uses the incremental gyroscope motion becomes the visual estimate. This corrects the gyroscope drift at every video frame, but now visual errors can affect the orientation estimate, so the visual feature tracking *must* be robust. In practice, the second method yields much smaller errors than the first, so that is what we use.

Overall, the inertial-optical hybrid greatly reduces the errors seen in the base system (Figure 5). During fast motions, the errors in the base system can become significant. Figure 6 shows an example of the correction provided by incorporating the video input. The inertial-video output stays within a few pixels of the true location. In actual operation the labels appear to stick almost perfectly.

Because the computer vision processing is computationally intensive, it does not run in real time. However, we emphasize that our result is *not* a simulation. We recorded several motion sequences of real video, compass, and inertial data. The method then ran, offline, on this real unprocessed data to achieve the results. We are investigating ways of incorporating this video-based

correction into the real-time system. For more details about this system, see [18].

5DOF SIMULATION

The two previous results focused on orientation tracking, relying upon GPS for acquiring position. However, there are many situations where GPS will not be available, and when objects of interest get close to the user, errors in GPS may appear as significant registration errors. Therefore, we need to explore methods of doing position tracking in unprepared environments. This section describes an approach to tracking relative motion direction in addition to rotation, based on observed 2-D motions.

The inertial-optical hybrid AR system described above uses a traditional planar-projection perspective camera. This optical system has several weaknesses when used to measure linear camera motion (translation). First, there is a well known ambiguity in discriminating between the image motions caused by small pure translations and small pure rotations [6]. Second, a planar projection camera is very sensitive to noise when the direction of translation lies outside of the field of view. Panoramic or panospheric projections reduce or eliminate these problems. Their large field of view makes the uncertainty of motion estimation relatively independent of the direction of motion. We compared planar and panoramic projections using the 8-point algorithm that uses essential matrix and co-planarity conditions among image points and observer's position [9, 10]. Figure 7 shows a result from our simulations. We plot the camera rotation and translation direction errors against pixel noise to show that for moderate tracking noise (<0.4 pixel), the panoramic projection gives more accurate results. We show one of several general motion paths tested. In this case, yaw = 8 degrees, pitch = 6 degrees, and translation (up) for 2 cm. Below 0.4 pixel noise levels, the panoramic projection shows superior accuracy (over a planar perspective projection) in terms of lower error and standard deviation of the error.

6DOF SIMULATION

The range of most vision-based tracking systems is limited to areas where a minimum number of calibrated features (landmarks or fiducials) are in view. Even partial occlusion of these features can cause failure or errors in tracking. More robust and dynamically extendible tracking can be achieved by dynamically calibrating the 3-D positions of uncalibrated fiducials or natural features [12, 13], however the effectiveness of this approach depends on the behavior of the pose calculations.

Experiments show that tracking errors propagate rapidly for extendible tracking when the pose calculation is sensitive to noise or otherwise unstable. Figure 8 shows how errors in camera position increase as dynamic pose and calibration errors propagate to new scene features. In this simulated experiment, the system starts tracking with 6 calibrated features. The camera is then panned and rotated while the system estimates the positions of 94 initially uncalibrated features placed in a 100"x30"x20" volume.

The green line in Figure 8 shows the errors in the camera positions computed from the estimated features, using the 3-point pose estimation method described in [11]. After about 500 frames (~16 seconds) the five-inch accumulated error exceeds 5% of the largest operating volume dimension. This performance may be adequate to compensate for several frames of fiducial occlusion, but it does not allow significant camera motion or tracking area extension. More accurate pose estimates are needed to reduce the error growth rate.

To address the pose problem we developed two new pose computation methods that significantly improve the performance of dynamic calibration and therefore increase the possibility of achieving 6DOF tracking in unprepared environments. One method is based on robust averages of 3-point solutions. The other is based on an iterative Extended Kalman Filter (iEKF) and SCAAT (Single Constraint At A Time) filter [17]. Both methods are designed specifically for low frame rates and over-constrained measurements per frame that characterize video vision systems. The pink and blue lines in Figure 8 show the results obtained by these two methods. These initial tests show significant improvements that lead us to believe that autocalibration methods will be an important approach to tracking in unprepared environments. For more details on the two pose estimation methods see [14].

FUTURE WORK

Much remains to be done to continue developing trackers that work accurately in arbitrary, unprepared environments. The results we described in this paper are a first step but have significant limitations. For example, the visual tracking algorithms assume a static scene, and we must add compass calibration routines that compensate for the changing magnetic field distortion as the user walks around. We currently assume that viewed objects are distant to minimize the effect of position errors; as we progress down the 6DOF route we will need to include real objects at a variety of ranges.

Future AR systems that work in unprepared environments must also address the size, weight, power, and other issues that are particular concerns for systems that operate outdoors.

ACKNOWLEDGMENTS

Most of this paper is based on an invited presentation given by Ron Azuma at the 5th Eurographics Workshop on Virtual Environments (with a special focus on Augmented Reality) in June 1999. This work was mostly funded by DARPA ETO Warfighter Visualization, contract N00019-97-C-2013. We thank Axel Hildebrand for his invitation to submit this paper.

REFERENCES

1. R. Azuma, A Survey of Augmented Reality. *Presence: Teleoperators and Virtual Environments* **6** (4), 355-385 (August 1997).

2. R. Azuma, The Challenge of Making Augmented Reality Work Outdoors. In *Mixed Reality: Merging Real and Virtual Worlds*, Y. Ohta and J. Tamura (Eds.), Springer-Verlag, 379-390 (1999).

3. R. Azuma and G. Bishop, Improving Static and Dynamic Registration in an Optical See-Through HMD. *Proceedings of SIGGRAPH '94*, 197-204 (July 1994).

4. R.Azuma, B. Hoff, H. Neely III, and R. Sarfaty, A Motion-Stabilized Outdoor Augmented Reality System. *Proceedings of IEEE Virtual Reality '99*, 252-259, (March 1999).

5. R. Behringer, Registration for Outdoor Augmented Reality Applications Using Computer Vision Techniques and Hybrid Sensors. *Proceedings of IEEE Virtual Reality '99*, 244-251, (March 1999).

6. K. Daniilidis and H.-H. Nagel, The Coupling of Rotation and Translation in Motion Estimation of Planar Surfaces. *IEEE Conference on Computer Vision and Pattern Recognition*, 188-193 (June 1993).

7. S. Feiner, B. MacIntyre, and T. Höllerer. A Touring Machine: Prototyping 3D Mobile Augmented Reality Systems for Exploring the Urban Environment. *Proceedings of First International Symposium on Wearable Computers*, 74-81 (October 1997).

8. E. Foxlin, M. Harrington, and G. Pfeiffer, Constellation: A Wide-Range Wireless Motion-Tracking System for Augmented Reality and Virtual Set Applications. *Proceedings of SIGGRAPH '98*, 371-378 (July 1998).

9. R. I. Hartley. In Defense of the 8-Point Algorithm. *5th International Conference on Computer Vision*, 1064-1070 (1995).

10. T. S. Huang and A. N. Netravali. Motion and Structure from Feature Correspondences: A Review. *Proceedings of the IEEE* **82** (2), 251-268 (February 1994).

11. U. Neumann and Y. Cho, A Self-Tracking Augmented Reality System. *Proceedings of ACM Virtual Reality Software and Technology*, 109-115 (July 1996).

12. U. Neumann and J. Park, Extendible Object-Centric Tracking for Augmented Reality. *Proceedings of IEEE Virtual Reality Annual International Symposium1998*, 148-155 (March 1998).

13. J. Park and U. Neumann, Natural Feature Tracking for Extendible Robust Augmented Realities. *International Workshop on Augmented Reality (IWAR)'98* (November 1998).

14. J. Park, B. Jiang, and U. Neumann, Vision-based Pose Computation: Robust and Accurate Augmented Reality Tracking. To appear in proceedings of *International Workshop on Augmented Reality (IWAR)'99* (October 1999).

15. A. State, G. Hirota, D. Chen, B. Garrett, and M. Livingston, Superior Augmented Reality Registration by Integrating Landmark Tracking and Magnetic Tracking. *Proceedings of SIGGRAPH '96*, 429-438 (August 1996).

16. G. Welch, Hybrid Self-Tracker: An Inertial/Optical Hybrid Three-Dimensional Tracking System. *UNC Chapel Hill Dept. of Computer Science Technical Report TR95-048* (1995).

17. G. Welch and G. Bishop, SCAAT: Incremental Tracking with Incomplete Information. *Proceedings of Siggraph97*, 333-344 (August 1997).

18. S. You, U. Neumann, and R.Azuma, Hybrid Inertial and Vision Tracking for Augmented Reality Registration. *Proceedings of IEEE Virtual Reality '99*, 260-267, (March 1999).

FIGURES

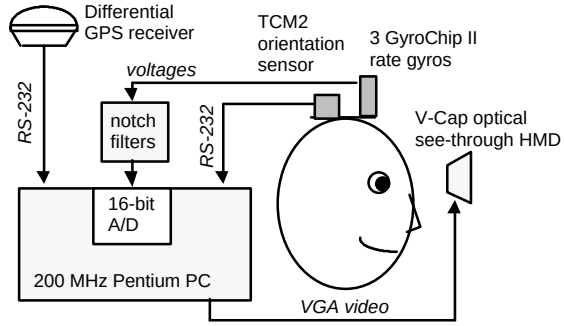

Figure 1: Base System dataflow diagram

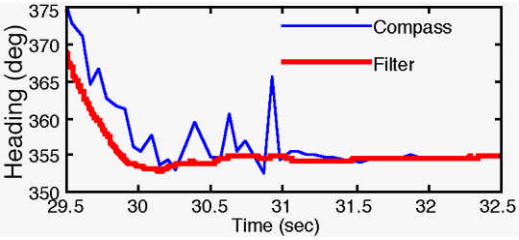

Figure 2: Sequence of heading data comparing the compass input vs. the filter output

Figures 3 and 4: (Left) Virtual labels over outdoor landmarks at Pepperdine University, as seen from HRL Laboratories. (Right) Example of 2-D video features automatically selected and tracked.

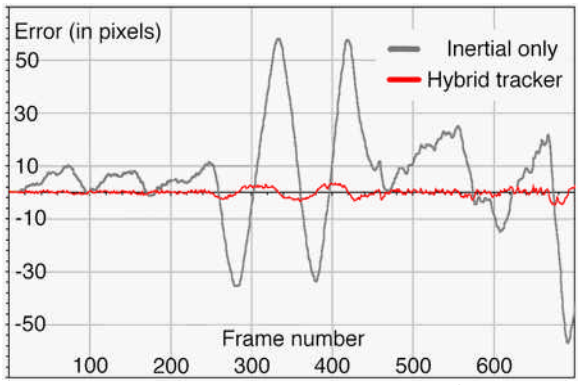

Figure 5: Graph comparing registration errors from base system (gray) vs. the inertial-optical hybrid running the second method (red)

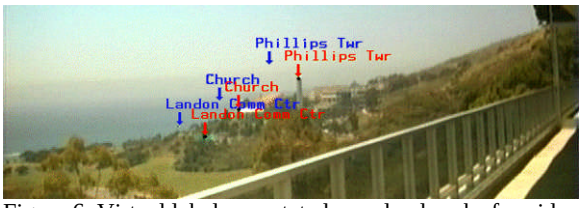

Figure 6: Virtual labels annotated over landmarks for video sequences showing vision-corrected (red labels), and inertial only (blue labels) tracking results.

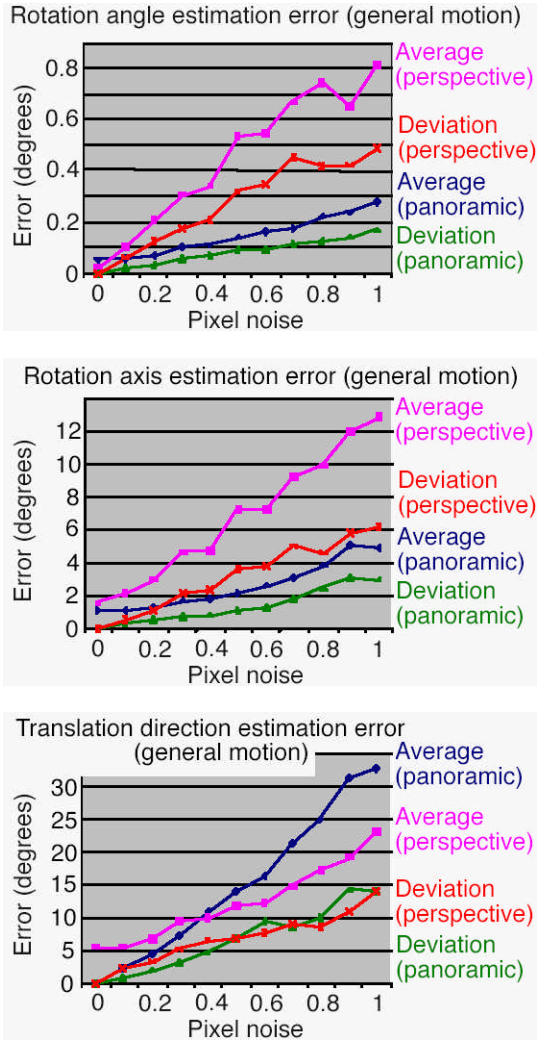

Figure 7: A comparison of 5DOF errors as a function of image-feature tracking pixel noise.

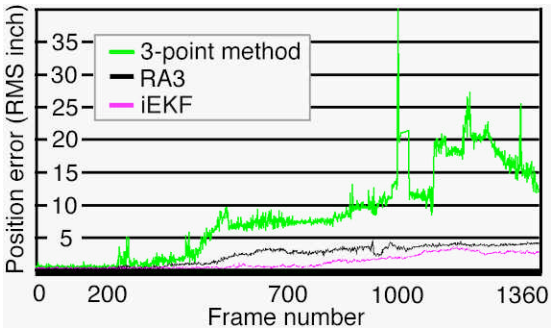

Figure 8: Propagated camera pose errors in 6DOF autocalibration experiment