Fusion of Vision and Gyro Tracking for Robust Augmented Reality Registration

Suya You and Ulrich Neumann

Integrated Media Systems Center
University of Southern California
Los Angeles, CA 90089-0781
{suyay|uneumann}@imsc.usc.edu

Abstract

A novel framework enables accurate AR registration with integrated inertial gyroscope and vision tracking technologies. The framework includes a two-channel complementary motion filter that combines the low-frequency stability of vision sensors with the high-frequency tracking of gyroscope sensors, hence, achieving stable static and dynamic six-degree-of-freedom pose tracking. Our implementation uses an Extended Kalman filter (EKF). Quantitative analysis and experimental results show that the fusion method achieves dramatic improvements in tracking stability and robustness over either sensor alone. We also demonstrate a new fiducial design and detection system in our example AR annotation systems that illustrate the behavior and benefits of the new tracking method.

1. Introduction

An augmented reality (AR) contains a mixture of real and synthetic scene elements. In contrast, virtual reality immerses users in a completely computer-generated world. AR systems enhance a user's perception of the real world with information that is not naturally part of the scene, thereby offering an intuitive and natural means for helping people to navigate and work effectively in the real world [3].

A key challenge for creating an augmented reality is to maintain accurate registration between real and computer-generated objects. As users move their viewpoints, the graphic virtual elements must retain their alignment with the observed 3D positions and orientations of real objects. This alignment depends completely on accurate tracking of the viewing pose (position and orientation), relative to either the environment or the annotated objects. The tracked viewing pose defines the virtual camera used to project 3D graphics onto the real world image, so pose tracking accuracy directly determines the visually-perceived accuracy of AR alignment and registration [5].

Vision-based tracking is commonly employed for AR. Unlike other active and passive sensing technologies, vision methods can estimate camera pose directly from the same imagery observed by the user. The pose obtained is therefore relative to the viewed objects of interest, not a separate sensor or emitter attached to the environment. This has several advantages: a) tracking may occur relative to moving objects; b) tracking measurements made from the viewing position often minimize the visual alignment error; and c) tracking accuracy varies in proportion to the visual size (or range) of the objects in the image [15]. However, vision also suffers from a notorious lack of robustness, end-to-end system delay, and high computational expense. Since vision sensors (cameras) nominally sample at video rates (30Hz), they are most appropriate for measuring *low-frequency* pose variations. Rapid or abrupt camera rotations or motions can cause vision tracking failures or instabilities.

Inertial sensors are widely used for motion tracking [4, 8, 10, 12, 18]. These sensors are self-contained, sourceless, and can be sampled at high rates (~1KHz). The latter characteristic makes them suitable for sensing the rapid motions that create *high-frequency* pose variations. However, inertial sensors only measure motion rates or accelerations; their signals must be integrated to produce position or orientation. Noise or bias in the sensor signal integration produces a drift in the attitude computation that accumulates with elapsed time. To correct the accumulated drift, periodic measurements from other sensors must provide absolute pose data.

In this paper, we extend our previous work of pure vision based tracking [16] to hybrid 6DOF pose tracking. We present a new framework for data fusion that combines the low-frequency stability of vision sensors with the high-frequency motion tracking of inertial rate-gyroscope sensors. The resulting system produces stable and accurate static and dynamic 6DOF pose tracking.

It may appear natural to simply apply a low-pass filter to the vision system and a high-pass filter to the integrals of the gyro sensors and combine the results, but the problems we confront are the selection of cutoff frequency and attenuation parameters for the filters and the method of fusing the two resulting data streams.

A number of theoretical and experimental studies lead us to a flexible framework with a two-channel motion-filter structure. The two processing channels, one for the low-rate vision measure and another for the high-rate inertial gyro, are complementary in that each compensates for the weaknesses in the other. The two channels process data independently, allowing for different sample rates of the sensor systems and reducing the end-to-end system delay.

We implement of framework with an Extended Kalman Filter (EKF). Novel aspects of the implementation allow for varied gyro data formats (angular-rate or relative-angles) as measurement input, and variations of the filter structure to adapt to application timing requirements. We describe these properties in later sections of the paper.

2. Related research

A wealth of research, employing a variety of sensing technologies, focus on motion tracking and registration for augmented reality applications. Recent hybrid tracking systems attempt to overcome the limits of any single technology [18].

State et al. [17] produced a hybrid system with fiducial-based vision and a magnetic tracker. Auer et al. [1] also proposed a similar vision-magnetic hybrid system using corners as visual features. The data from the magnetic tracker is used to predict feature positions and bound the result from the vision system.

Foxlin [12] developed an inertial orientation tracker with three orthogonal rate-gyro sensors. To correct for the drift caused by the gyro data integration, he employed inclinometers and a compass. A complementary Kalman Filter performs data fusion and error compensation. This system only tracks 3DOF orientations. More recently, Foxlin et al. [13] developed a 6DOF tracking system with a similar orientation tracking system aided by ultrasonic time-of-flight range measurements to a constellation of wireless transponder beacons.

Azuma et al. [2] developed a 6DOF tracking system using linear accelerometers and rate-gyroscopes to improve the dynamic registration of an optical-beacon ceiling tracker. Recently, Yokokohji et al. [19] demonstrated a hybrid system combining six linear accelerometers and a vision tracker. The accelerometers predict head motion to compensate for system delay and an EKF is used for data fusion.

Chai et al. [8] uses an adaptive pose estimator with vision and inertial sensors for head tracking. The EKF estimator used multiple models of the expected user head motion to allow for variations in the range of expected motions. Emura et al. [10] suggests a hybrid tracking system consisting of a magnetic tracker and gyro sensor that compensates for the latency in the magnetic tracker.

AR tracking in unconstrained environments, especially outdoors, is a challenge problem. In [11], Columbia's Touring Machine tracks outdoors by combining a differential GPS with a compass and tilt sensor. Azuma et al. [4] demonstrates a hybrid orientation tracker that stabilizes an outdoor AR display by combining gyro sensors with a compass and tilt sensor.

Our previous work [18] combines a natural-feature vision system with three gyro sensors to provide accurate 3DOF orientation tracking in outdoor environments. The fusion approach is based on the SFM (structure from motion) algorithm, in which approximate feature motion is derived from the inertial data, and vision feature tracking corrects and refines these estimates in the image domain. Furthermore, the inertial data also serves as an aid to the vision tracking by reducing the search space and providing tolerance to interruptions.

Behringer [6] also suggests an outdoor system that combines vision with GPS. In his vision algorithm, terrain silhouettes are visual features matched to DEM (Digital Elevation Model) maps for orientation registration in a well-structured terrain. A GPS is used for position estimation.

3. Motion model and system dynamics

The motion model of our tracking system is illustrated in Figure 1, where three 3D coordinate frames are defined. The first frame is a world coordinate system **W** fixed on the reference ground. Second is a camera coordinate system **C**, which uses the camera's focal center as its

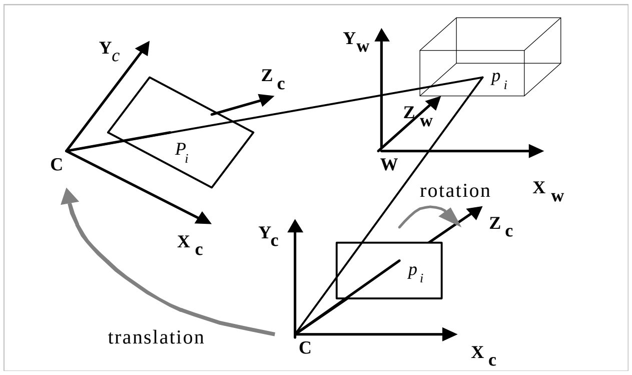

Figure 1: Motion model and the related coordinate

origin, and the third one is the inertial-body fixed coordinate frame **I**. Since the inertial sensor is rigidly mounted to the camera, the transformation between frames **C** and **I** is pre-calibrated [18] and constant.

At time t_0, coordinate frames **W** and **C** are considered coincident. As the camera and its frame **C** move, the frame **W** remains fixed. As shown in Figure 1, the camera's motion between two image frames can be uniquely decomposed into a translation component and a rotation component about the focal center of camera.

Camera motion at time t_i can be represented by

$$M_i = \left(r_i, v_i, y_i, w_i\right) \tag{1}$$

where the position r_i and the translation velocity v_i represent the camera position parameters, and y_i, the related rotation angle, and w_i, the instantaneous rotation velocity, represent camera rotation parameters. As the sampling period $\Delta T = t_{i+1} - t_i$ is small, we can use a constant velocity model for the system dynamics

$$M_{i+1} = \begin{bmatrix} r_{i+1} \\ v_{i+1} \\ y_{i+1} \\ w_{i+1} \end{bmatrix} = \begin{bmatrix} r_i + v_i \Delta T \\ v_i + n_v \\ y_i + w_i \Delta T \\ w_i + n_w \end{bmatrix} \tag{2}$$

where n_v and n_w are the random distribution noise components of translation velocity and instantaneous rotation velocity, respectively.

It is worth noting that, in (1), the camera position parameters (position r_i and translation velocity v_i) are defined in the world coordinate frame, but the rotation parameters (related rotation angle y_i, and instantaneous rotation velocity w_i) are defined in the inertial body coordinate frame. The benefits of choosing the inertial body frame for defining these rotation parameters are: a) avoiding complex nonlinear and matrix inversion computations (see Equations (7) and (8)); and b) convenience in interfacing with rate gyro data, which is defined in the inertial body coordinate frame.

Let $R(y_i)$ be the rotation matrix defined in world coordinate, and let $M(y_i)$ be the related rotation matrix in the inertial body frame, then

$$R(y_{i+1}) = R(y_i) \cdot M(y_i) \tag{3}$$

This equation defines the transformation to the absolute orientation attitude.

4. Fusion of vision and gyro measurements

4.1 Image measurement

The vision system measures the image positions of targets known in world coordinates. Assume the camera observes a static scene[*]. Suppose n features are detected and tracked in the scene. $p_k = \left(p_x^k, p_y^k, p_z^k\right)$ is the k^{th} feature point in the world coordinate frame. Then, under perspective projection, at time t_i, its projection $P_{k,i} = \left(P_{x,i}^k, P_{y,i}^k\right)$ on the image plane is

$$\begin{bmatrix} P_{x,i}^k \\ P_{y,i}^k \end{bmatrix} = \begin{bmatrix} f\,\dfrac{R_1(y_i)\cdot[P_k - r_i]}{R_3(y_i)\cdot[P_k - r_i]} + n_x \\ f\,\dfrac{R_2(y_i)\cdot[P_k - r_i]}{R_3(y_i)\cdot[P_k - r_i]} + n_y \end{bmatrix} \tag{4}$$

where f is the focal length of the camera, and R_i is the i^{th} row vector of rotation matrix $R(y)$. n_x and n_y model the measurement noise of feature detection.

Equation (4) describes the relationship between the motion model parameters and the image measurement. Given a number of measurement pairs (≥ 3), the camera pose can be calculated. Since we want to track camera pose in each video frame, fast feature tracking and correspondence between the 2D features and 3D world points are key components of a real-time AR system. Our robust landmark detection and recognition algorithm achieves 28-frames/second on modest PC hardware. This system is detailed in section 5.

4.2 Inertial gyro measurement

Our inertial sensor consists of three orthogonal rate gyroscopes to sense angular rates of rotation along three perpendicular axes. The gyroscopes are analog devices, so a 16-bit A/D card is used for sampling and digital conversion. Since rate gyroscopes only measure the angular rate of rotation, we implement a low-level A/D driver library with time-integration and calibration algorithms to achieve a 1KHz-sampling rate.

Let $w_i = \left(w_{x,i}, w_{y,i}, w_{z,i}\right)$ represent the angular rates measured from the gyros with random-distribution noise n_{w_i}

$$w_i = \overline{w}_i + n_{w_i} \tag{5}$$

where $\overline{w}_i$ is the true noiseless angular rate. In the integration update interval ΔT, the related rotation angle $y_i = \left(y_{x,i}, y_{y,i}, y_{z,i}\right)$ in the inertial body coordinates can be calculated as

[*] This assumption can be removed by the addition of a moving object detection algorithm. We demonstrate this in section 6.3.

$$y_i = \sum_{k=0}^{\Delta T-1} w_{i,k} = \sum_{k=0}^{\Delta T-1} \overline{w}_{i,k} + n_{y_i} \tag{6}$$

where n_{y_i} is random-distribution integration noise. Note that both equation (5) and (6) are measured in the inertial body coordinate frame. For the absolute rotation angle (Euler angle) $q_i = \left(q_{x,i}, q_{y,i}, q_{z,i}\right)$ in the world coordinate, the relationship to the angular rate is

$$w_i = W^{-1}(q_i)\dot{q}_i \tag{7}$$

where $W(q_i)$ is the Jacobian matrix that relates the absolute rotation angle to the angular rate, defined by

$$W(q_i) = \begin{bmatrix} 1 & \sin q_{z,i} \tan q_{y,i} & \cos q_{z,i} \tan q_{y,i} \\ 0 & \cos q_{z,i} & -\sin q_{z,i} \\ 0 & \sin q_{z,i} / \cos q_{y,i} & \cos q_{z,i} / \cos q_{y,i} \end{bmatrix} \tag{8}$$

Equations (7) and (8) are used for calculating the attitude of sensor body [7]. However, since the gyro data is noisy, directly solving (7) will produce drift in the attitude computation that accumulates with elapsed time. In our experiments, we added a third-order term approximating the drift measured by integrating at the 1kHz sampling rate. The measured drift rate is about 0.7degree/min. To correct for accumulated drift, it is necessary to periodically reset the integrator output with a measurement from other sensors that provide correct reference information. (See our previous work in [18] for an analysis of the error sensitivity of an inertial gyroscope within a video AR system.)

4.3 Fusion filter

The goal of the fusion filtering is to estimate the camera pose parameters of (1) from the measurements of the vision and inertial gyro sensors. Since the vision and gyro sensors have different sample rates, we implement a complementary motion estimate filter as shown in Figure 2. The filter has a predictor-corrector structure. There are two independent correction channels sharing a common prediction module: one is for vision measurements; and another one is for gyro measurements. At each new frame time, we first predict the filter state based on the prior state and the system dynamic model (2), and then correct the prediction using the observed vision and gyro measurements given by equation (4) and (5) (or (6)) in each channel. Finally, the state is updated and output by converting it to the required pose format. The advantages of the complementary filtering structure are:

(a) Combining the low-frequency characteristics of vision tracking with the high-frequency characteristics of inertial gyro sensors to compensate for the weaknesses in each component.

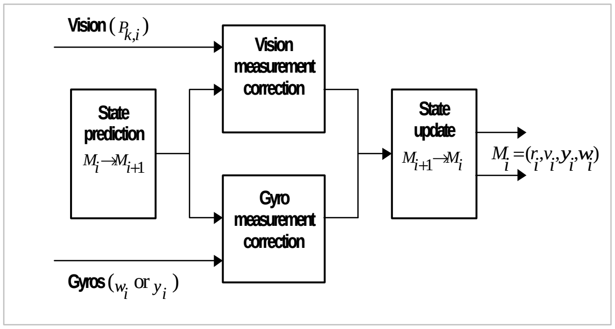

Figure 2: Fusion filter framework

(b) Splitting the correction system into two separate processing channel to accommodate the different sample rates of vision and gyro sensors and reduce the end-to-end system latency.

(c) Independent channel processing handles incomplete information measurements. For example, when no vision measurement is available (due to occlusions, for example), the overall system maintains tracking by only using the gyro corrected channel (and *vice versa)*.

Currently, an EKF is used for the filter implementation. The complementary filtering structure is a variation of two parallel EKF banks sharing one common state prediction module, i.e.

State prediction (common):

$$\begin{aligned} M_{i+1}^- &= A_i M_i \\ P_{i+1}^- &= A_i P_i A_i^T + Q_i \end{aligned} \tag{9}$$

Measurement correction (vision or gyro):

$$\begin{aligned} K_i &= P_i^- H_i^T \left(H_i P_i^- H_i^T + R_i\right) \\ P_{i+1} &= (I - K_i H_i) P_i^- \\ M_{i+1} &= M_{i+1}^- + K_i \left(m_i - \hat{m}_i\right) \end{aligned} \tag{10}$$

where

A_i : the state transition matrix from equation (2)

P_i : the state covariance matrix

Q_i : the process noise covariance matrix

R_i : the measurement covariance matrix

H_i : the Jacobian matrix of measurement

K_i : the Kalman gain

m_i : observed measurement from vision or gyro tracking

$\hat{m}_i$: predicted measurement given current state estimate

For vision measurement, we use the 2D image coordinates of calibrated landmarks given by equation (4). These landmarks are tracked and recognized automatically by a

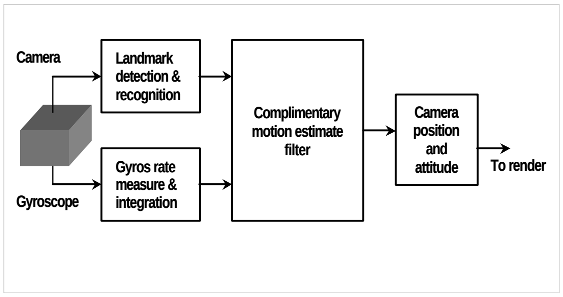
Camera
Gyroscope
Landmark detection & recognition
Gyros rate measure & integration
Complimentary motion estimate filter
Camera position and attitude
To render

Figure 3: System dataflow

robust vision tracker in real-time. For the inertial gyro measurement, as mentioned in Section 3, we directly use the gyro rate given by equation (5), or the relative rotation angle given in equation (6), where both are defined in the inertial body coordinate. We implemented both variations and found their performance almost identical, but the relative rotation angle method is much faster, since the numerical integration of equation (6) is done by the low-level gyro processing library, rather then in the EKF.

5. System

Our AR tracking system consists of a robust vision landmark tracker, inertial gyro sensors, and the complementary fusion filter described above. Figure 3 illustrates the system dataflow and Figure 4 shows the hardware sensor configuration.

The sensor module contains a CCD video camera (Sony XC-999 with 6mm lens), and three orthogonal rate gyroscopes (GyroChip II QRS14-500-103, from Systron Donner), which are tightly covered by a foam block to provide shock protection and a stable temperature environment from the sensors. The video camera provides a 30 Hz video stream, while the three gyroscopes are sampled at 1kHz via a 16-bit A/D converter (National Instruments DAQPCI-AI-16XE-20).

5.1 Fiducial system

The fiducial system consists of calibrated landmarks, which is crucial for vision tracking. Different type of landmarks have been suggested in several systems, including corner features, square shape markers, circular markers, and multi-ring color markers [9][15][17]. In our work, we designed landmarks whose appearance simplifies robust detection and recognition. After evaluating many options, we selected the square shape marker similarly to [9] with unique patterns inside (Figure 5). This design is easily downloaded and printed on any BW printer. Its simple geometry lends itself to mathematical perspective projection compensation, leading to robust landmark detection and recognition.

5.2 Landmark detection and recognition

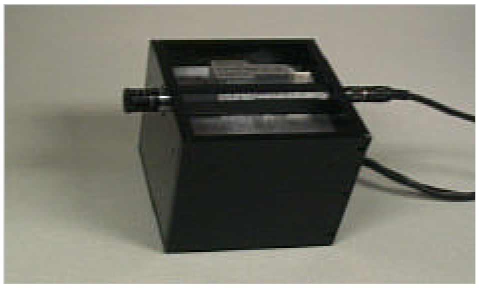

Figure 4: Hybrid vision and gyroscope sensor hardware

Since we want camera pose for each video frame, fast and efficient landmark detection and recognition are crucial. We developed a Principal Component Analysis (PCA) method that robustly detects and recognizes the square landmarks in real-time.

Principal Component Analysis is a statistical signal analysis technique that represents signals efficiently as a weighted linear combination of object measurements (called eigenvectors), which express the greatest statistical variance over all of the data sets. It has also been shown that identifiable signals can be reconstructed by using only a subset of the eigenvectors, i.e., those with the largest eigenvalues. This low-dimensional representation is optimal in that it minimizes the squared error between the representation of the signal and the original signal [14]. These abilities of data compaction and feature extraction (the most significant components) lead to efficient landmark representation and matching.

Landmark training (offline)

We first train the recognizer with all possible markings to extract the most significant components as base vectors that span the landmark feature space. Then the landmarks are projected to the feature space, and their projection coefficients are saved in the database as matching templates.

Landmark detection

The detection procedure is performed in three steps:

Coarse detection: using a predicted position (from the previous frame) we find the regions of the image where the gradients are high over an expected-size image area.

Projection compensation: use a perspective-imaging model to compensate for geometric deformations of the extracted regions. To compensate for lighting variations we use the gradient for intensity normalization.

Fine detection: fit the extracted candidate landmarks to the defined models. The best fit determines the detected landmark.

Landmark recognition

For each detected landmark, we first project it into the PCA feature space, and its projection coefficients are extracted. Then the extracted coefficients are compared with the stored training set saved in the database by using template matching. The best match determines the recognized marking.

This detection and recognition approach is robust and fast. It achieves 28 frames/sec on a 450 MHz PC. The system detects and discriminates between dozens of uniquely marked landmarks.

5.3 Gyroscope calibration

Although the manufacturer provides the bias value for each gyro, directly using the specified value produces significant drift. Drift also arises from the zero-voltage offsets of the A/D converter. We calibrate the combined bias and offset for each gyroscope and A/D channel at startup by averaging several seconds of output while the sensor is still. For scale factors, we use the specified values provided by the manufacturer's test sheets.

6. Experiments

6.1 Real-time landmark detection and recognition

This experiment evaluates the accuracy and robustness of the landmark detection and recognition method. Due to the PCA feature extraction abilities, markings can be any symbols, from simple ASCII characters to complex gray image. Our landmarks are made with a simple image editor, and printed on white paper. The size of the squares varies depending on the lens focal length and the intended viewing range.

We trained about 50 patterns. To handle variations of viewpoint, for each pattern, we capture the frontal view and two views about 45-degrees off-axis in the horizontal and vertical directions. The result is a total of five images for each pattern. These images are used for PCA training, and their projection coefficients are saved as the database for matching templates.

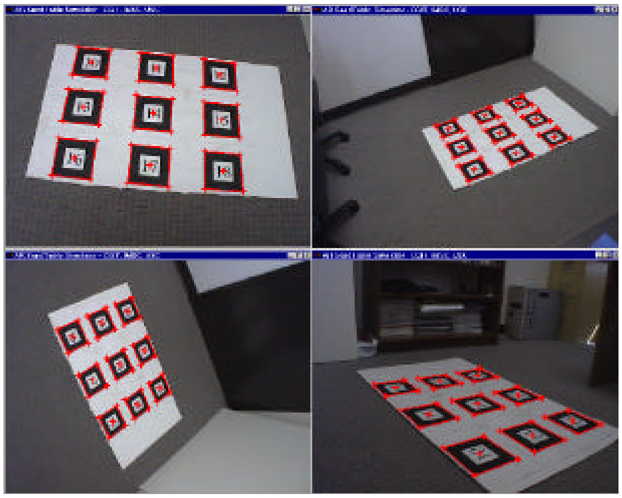

Figure 5: Landmark detection and recognition results

Figure 5 illustrates four cases of landmark detection and recognition from different view distances and orientations. The tracked landmarks are denoted with red boxes and cross hairs at their corners and centers.

To evaluate the detection accuracy, we manually select the corners and centers of the landmarks in each frame, using the distances between the manually selected positions and their detected positions as an accuracy metric. We repeat this measurement several times, and the resulting average RMS accuracy converges to 0.87-pixels. We also measured the recognition rate under various viewpoint and lighting conditions. For all 50 trained patterns, the recognition rate is above 96%, allowing the viewpoint to varying up to 70-degrees off-axis. Recognition ambiguity often occurs for easily confused character pairs, such as 0/8, and 6/9. After several experiments, we simply removed these pairs from the database.

6.2 Pose tracking

This test evaluates the performance of the complementary fusion filter. Figure 6 shows snapshots of tracking in a real scene. In these images, the red boxes and cross hairs identify the detected and recognized landmarks. The green cross hairs denote the projections of the known 3D landmark positions under the estimated camera pose.

During this experiment, the camera undergoes arbitrary 6DOF motions, viewing the calibrated landmark-board from different positions and orientations. Camera pose is continually computed by the fusion filter in real time.

To evaluate the tracking accuracy dynamically, we again use a projected distance metric. By projecting the 3D positions (corners and center) of detected landmarks on to the image plane, we can compare the average differences between the projected image positions and the tracked landmark positions. Since the image-space distances are related to pose tracking errors, this method dynamically measures the accuracy of tracking system. This projected distance measure is appropriate for an AR system since the final measures of an augmented reality are the perceived image. Figure 7 shows the dynamic tracking accuracy for the scene shown in Figure 6. The average error is 2.18 pixels, and the maximum error is 9.93 pixels in 640x480 images.

To evaluate the static registration, we tested the system by keeping the tracker running up to 10 hours in a static state. It exhibits no perceptible drift or jitter, and the resulting registration accuracy remains within 1.3 pixels. Without

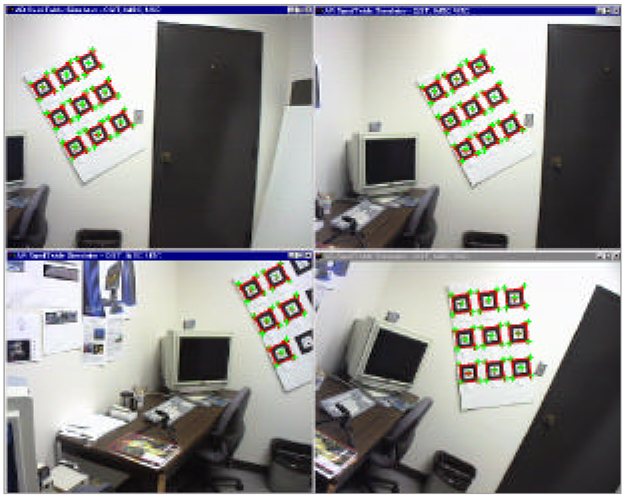

Figure 6: Snapshots of tracking sequence showing the hybrid tracking results

vision compensation, the gyroscope drifts at about 0.7 degree/min.

As described in Section 4, the gyro measurement can be either the angular rate or relative rotation angle. We evaluated the difference between using the two measurements and found that their tracking accuracy were almost identical, but using the relative rotation angle is much more efficient, achieving a frame rate at 23 frames/second on a 450 MHz Pentium III. Figure 7 shows the results for both cases.

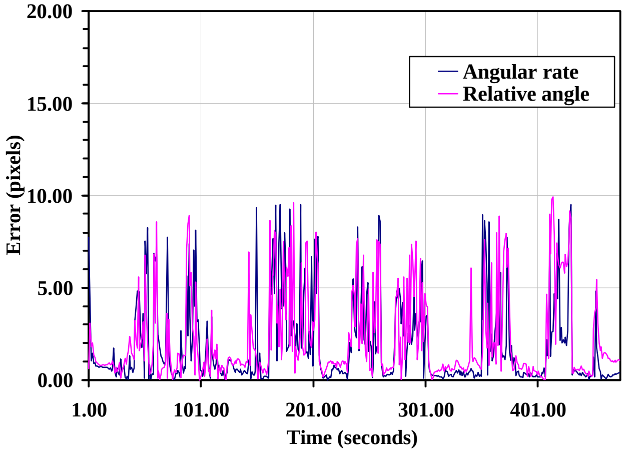

Figure 7: Tracking errors shown in the image plane: blue line denotes the tracking errors when directly using Gyro angular rate. The resulting max tracking error: 9.5 pixels; average error: 1.84 pixels; and error covariance: 5.32 pixels. The red line denotes the errors when using the integrated relative angle. In this case, the max tracking error: 9.93 pixels; average error: 2,18 pixels; and error covariance: 5.62 pixels.

6.3 A virtual sand table

We implemented an AR demonstration system using the hybrid tracker to illustrate its utility in AR or VR application. The system creates a "Virtual Sand Table", simulating the behavior of rear-projection Sand Table displays used in architecture and military applications.

We created a landmark board to serve as the projection screen. The board is a white poster board with several calibrated landmarks on it. In one case, we printed 9 unique numbered landmarks – one per 8.5x11 sheet of plain paper arranged in a 3x3 grid. The choice of nine landmarks is arbitrary. Fewer landmarks give useable results, even one gives useful pose information. When the board is viewed, the camera pose is computed and a virtual terrain model (obtained from the National Geography Association) with annotation is displayed on the board. As the camera moves while viewing the board the virtual model is displayed as if it were attached to the board (Figure 8). The system also allows partial occlusion or large camera rotations that put the board out of view. In the latter case, there are no vision measurements available temporarily, but the gyro data allows the system to

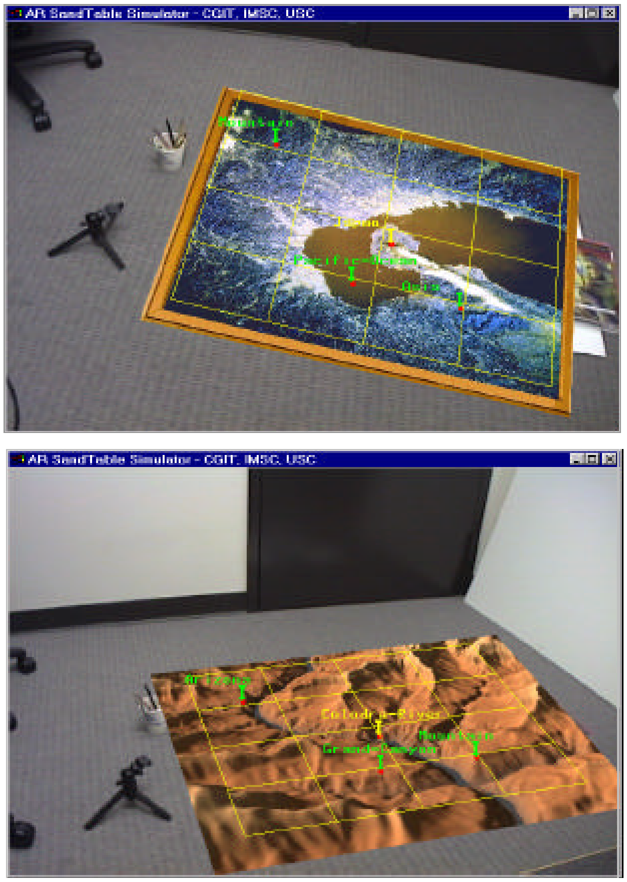
AR SandTable Simulator - CGIT, IMSC, USC
Mountain
Japan
Pacific-Ocean
Asia
AR SandTable Simulator - CGIT, IMSC, USC
Arizona
Colorado-River
Grand-Canyon
Mountain

Figure 8: Images of "Virtual Sand Tables"

continue estimating pose.

The user can interact with the Sand Table scene by using single landmarks attached to a short rod or wand. We associate graphic labels or cursors to each of the movable landmarks. As users move the landmarks over the board, the corresponding graphic is displayed over the scene. In this case, we need to remove the motions of the moving landmarks by subtracting them from the camera motion. The poses of the moving landmarks are calculated relative to the camera and used for rendering the graphic labels.

We also use a laser pointer as an interaction tool. We track the red laser dot on the landmark board and calculate its image position. This tracked 2D image coordinate is then projected into the board coordinates based on the camera pose. When the user points to a "hot" position on the terrain model, for example, related information is annotated for the selected position. The method for tracking the red laser dot is based on a color ring detection approach described in [15]. Video clips of the above three tests can be found at [20].

7. Conclusion

We presented a hybrid filter framework for accurate AR pose tracking with inertial gyroscope and vision tracking technologies. The framework is a two-channel complementary motion filter. The two channels operate independently to allow for different sensor sample rates and reduced system latency.

We implemented the framework using an Extended Kalman Filter (EKF). In this case, the complementary filtering structure is treated as two parallel EKF banks sharing one common state prediction module. We quantitatively evaluated the system under dynamic conditions, and the experimental results showed that the fusion method achieves high tracking stability and robustness, and image projection accuracy is under 10-pixels over a range of test conditions.

For vision tracking, accurate and fast landmark tracking and identification is critical. We presented a PCA method that automatically detects and recognizes black-white landmarks in real time.

We implemented AR annotation systems to illustrate the operation of the hybrid tracking system

References

[1] T. Auer and A Pinz. "Building a Hybrid Tracking System: Integration of Optical and Magnetic Tracking", *Proc.* of *IEEE International Workshop on Augmented Reality*, pp. 13-22, 1999.

[2] R. Azuma and G. Bishop, "Improving Static and Dynamic Registration in an Optical See-through HMD", *Proc. of SIGGRAPH 95*, 1995.

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

[4] R. Azuma, B. Hoff, H. Neely III and R. Sarfaty. "A Motion-stabilized Outdoor Augmented Reality System", *Proc. of IEEE VR'99*, pp. 252-259, 1999.

[5] M. Bajura and U. Neumann. "Dynamic Registration Correction in Augmented Reality Systems", *Proc. of IEEE Virtual Reality Annual International Symposium*, pp. 189-196, 1995.

[6] R. Behringer, "Registration for Outdoor Augmented Reality Applications Using Computer Vision Techniques and Hybrid Sensors", *Proc. of IEEE VR'99*, pp. 244-251, 1999.

[7] K. Britting, "Inertial Navigation System Analysis", Wiley Interscience, New York, 1971.

[8] L. Chai, K. Nguyen, B. Hoff, and T. Vincent. "An Adaptive Estimator for Registration in Augmented Reality", *Proc.* of *IEEE International Workshop on Augmented Reality*, pp. 23-32, 1999.

[9] M. Billinghurst and H. Kato, "Collaborative Mixed Reality", Prof. of the first international sysposium on Mixed Reality, pp. 261 – 284, 1999.

[10] S. Emura and S. Tachi, "Multisensor Integrated Prediction for Virtual Reality", *Presence,* Vol 7. No 4, pp. 410-422, 1998.

[11] S. Feiner, B. MacIntyre, and T. Hollerer, "A Touring Machine: Prototyping 3D Mobile Augmented Reality Systems for Exporing the Urban Environment", *Proc. of First International Symposium on Wearable Computers*, pp. 74-81, July 1997.

[12] E. Foxlin. "Inertial Head-Tracker Sensor Fusion by a Complementary Separate-Bias Kalman Filter", *Proc. of IEEE Virtual Reality Annual International Symposium*, pp. 184-194, 1996.

[13] E. Foxlin, M. Harrington, and G. Pfeifer. "Constellation: A Wide-Range Wireless Motion-Tracking System for Augmented Reality and Virtual Set Applications", *Prof. of GRAPHICS 98*, 1998.

[14] B. Kumar, D. Casasent, and H. Murakami, "Principal Component Imagary for Statistical Pattern Recognition Correlators", Optical Engineering, Vol. 21, No. 1, Jan/Feb 1982.

[15] U. Neumann and Y. Cho, "A Self-Tracking Augmented Reality System", *Proc. of ACM Virtual Reality Software and Technology*, pp. 109-115, 1996.

[16] J. Park, B. Jiang, and U. Neumann. "Vision-based Pose Computation: Robust and Accurate Augmented Reality Tracking", *Proc of IEEE International Workshop on Augmented Reality*, pp. 3-12, 1999.

[17] A. State, G. Hirota, D. T. Chen, B. Garrett, M. Livingston. "Superior Augmented Reality Registration by Integrating Landmark Tracking and Magnetic Tracking", *Proc. of SIGGRAPH'96*, pp. 429-438, 1996.

[18] S. You, U. Neumann, and R. Azuma, "Hybrid Inertial and Vision Tracking for Augmented Reality Registration", *Proc. of IEEE VR'99*, pp. 260-267, 1999.

[19] Y. Yokakohji, et. al., "Accurate Image Overlay on Video See-Through HMDs Using Vision and Accelerometers", *Proc. of IEEE VR 2000*, pp. 247-254, 2000.

[20] http://deimos.usc.edu/~suyay/demo.html