Estimating Motion Parameters of Head by Using Hybrid Extended Kalman Filter

Sejong Heo, Oksik Shin and Chan Gook Park
Seoul National University, South Korea

BIOGRAPHY
Sejong Heo is a M.S. student in the School of Mechanical and Aerospace Engineering of Seoul National University, South Korea. He received his B.S. degrees in Mechanical and Aerospace Engineering of Seoul National University in 2008. His current research concerns the hybrid head/eye tracking system development and the computer vision.

Oksik Shin is a M.S. student in the School of Mechanical and Aerospace Engineering of Seoul National University, South Korea. He received his B.S. degrees in Mechanical and Aerospace Engineering of Sejong University in 2009. His current research concerns the head tracking system development and the fault tolerant system.

Prof. Chan Gook Park is a professor of School of Mechanical and Aerospace Engineering of Seoul National University. He received his B.S., M.S., and Ph.D. degrees in Control and Instrumentation Engineering from Seoul National University in 1985, 1987, and 1993, respectively. He worked as a postdoctoral fellow at the Engineering Research Center for Advanced Control and Instrumentation, Seoul National University in 1993. From 1994 to 2003, he was with Kwangwoon University, Seoul, Korea, as an associate professor. He is currently in charge of IEEE AES Korean Society. His current research topics include the development of Inertial Navigation Systems, GPS/INS integration, low-cost IMU application, positioning for Ubiquitous Sensor Network and advanced filtering techniques.

ABSTRACT
In this paper, we introduce an optical and inertial helmet tracker system that is required to estimate the motion parameters of the helmet. In this case, the motion parameters consist of the position, velocity, acceleration, attitude and angular velocity. This helmet tracker system consists of two infrared CCD sensors, several infrared LEDs on the helmet, an inertial measurement unit (IMU) and a computer for tracking.

Our Helmet tracking system has a framework with a two-channel motion filter structure. The two channels, one for the optical measurement from the image sequence of CCD sensors and the other for the inertial measurement from the IMU, process independently with different sampling rates. With this hybrid system, we can overcome the failure in tracking a rapid motion and the accumulated error from the inertial sensors. Because of the nonlinearity in the state model, we implemented the system with the Extended Kalman Filter (EKF). The EKF has two channels for measurement that share a common prediction module. We implemented the real system and conducted the simulation with the real sensor data.

INTRODUCTION
Due to the rapid development of the avionic system with modern high techniques, aircraft pilots face a problem in that the complexity in using the system reduces the pilots’ ability in concentrating on attacking targets in battles or controlling aircraft. Many technologies have been developed to address this problem. If a pilot can control the targeting system or the avionic system without seeing the control panel, the pilot can greatly improve his capability to focus on the problem at hand. In this case, the key solution of this problem is to estimate the motion of the pilot, especially the motion of the pilot’s head. The helmet tracker system has been developed to estimate the motions of a head in a military fields. The helmet tracker system is very useful not only in military fields, such as a weapon cueing system, but also in the fields of augmented reality such as head-mounted display.

We can describe the helmet tracker with sensors mounted in the system. The magnetic helmet tracker has the advantages of small head-mounted sensors and high accuracy; however it is sensitive to the distortion by metallic objects. The optical helmet tracker has the advantage of low cost; however it is hard to implement because of the high computational expense. With the advancements in computing technology and current digital imaging equipment, researchers have actively studied the optical helmet tracker system. However this system still has problems in robustness that cause some serious failures and in high computational expenses. Especially in the case of tracking a rapid motion, the low
sample frequency (lower than 30 Hz) can cause failures in tracking.

To overcome this rapid motion problem, we have extended our previous work of the optical helmet tracker system to the hybrid tracer system with two measurement channels of optical and inertial sensors. The inertial sensors are widely used for motion tracking. These sensors are self-contained and can be sampled at high rates (less than 1 kHz). This second feature of the inertial sensors makes them suitable for sensing rapid motion. However, inertial sensors measure angular rates and accelerations, so the data from the inertial sensors should be integrated to produce the position or the orientation of the device. The noise or bias in the sensor signal integration produces a drift in the attitude computation that accumulates with elapsed time. To correct this accumulated drift, periodic measurements from other sensors need to provide the absolute pose data. In this paper, some image processing algorithms are used to produce this absolute pose data.

HYBRID HELMET TRACKING SYSTEM

Hardware Design

The hybrid helmet tracker system consists of two infrared CCD cameras (VCC-S70) and infrared LEDs attached on the helmet, XSens MTx inertial measurement unit(IMU), Matrox Meteor2-MC/4 frame grabber and a computer for the tracking algorithm. Figure 1 illustrates the hybrid helmet tracker system. The data from the IMU and the video images from the cameras are transmitted to the desktop computer through the cables. The video images from the two cameras are digitized to a resolution of 640x480 pixels.

Measurements

Our hybrid head tracker system has two measurements, one from the image sequence and the other from the IMU. The measurements from the image sequence are the attitude and the position of the helmet, not the position of features in the image plane. The infrared stereo cameras detect the infrared LED features on the helmet, and by using the infrared LEDs, the lighting condition and the viewing direction can be controlled independently. The features are separated from the background by the threshold in gray image space. Due to the high contrast of the features, we can simply and robustly extract the features from the images. After the processing with these extracted features, we can obtain the 3-D position data of the point sets indexed with the model that we obtained by 3-D scanning of the helmet. We can then estimate the attitude and the position from the comparison with the 3-D position data of the point sets of the model. We use these attitudes and positions as image measurements.

The inertial sensors deliver linear acceleration and angular velocity in three orthogonal coordinate directions aligned with the IMU. We directly use these angular velocity and acceleration as inertial measurements. To estimate the velocity and the position, integration over time has to be performed. When we use the inertial sensors, we need to know the scale factor and the bias. Then we can obtain these parameters with calibration. In this study, however, instead of calibrating IMU by ourselves, we use the calibrated data offered by MTx.

These two measurements process independently with different sample rates. The sample rate of the image processing is 30 Hz, and the sample rate of the IMU is 100 Hz in our experimental setup.

IMAGE PROCESSING ALGORITHM

Figure 2 shows the flow chart of our image processing algorithm. In preprocessing, we have to calibrate the camera. We extract the feature with the threshold and with the masking technique of the image plane. Then we can obtain the 2-D point sets in the image plane. By using the epipolar line, the Hausdorff distance and the camera calibration information, we can transform the 2-D point sets of two images to 3-D point sets. After the model indexing, we can obtain ‘N’ reliable 3-D point sets indexed with the model. We can then calculate the attitude and the position of the helmet frame. A more detailed image processing technique is introduced later in this section.

Figure 1: Composition of the hybrid helmet tracker

Figure 2: Flow chart of the image processing algorithm
Camera Calibration

Camera calibration has been studied extensively, and the standard technique has been established. In this study, we use the stereo camera calibration with Bouguet’s camera calibration toolbox (Bouguet, 2006). The calibration uses images of a checkered target in several positions and recovers the cameras’ intrinsic parameters, as well as the extrinsic parameters between the two cameras.

Feature Segmentation

Figure 3 shows the image from the camera and the LED-marked feature at an enlarged pixel resolution. The LED-marked features are easily recognizable and are separated from the background without the need of defined lighting conditions. A further advantage is that the rotationally symmetric LEDs do not change their appearances much in the observations from different directions. Because of the high contrast, features are separated from the background by applying the threshold in gray image space. After the dilation process for robustness, Gaussian mask is used to determine the center of the segmented features. The equations for this are as follows:

\[
g(u,v) = \sum_{s=-m}^{m} \sum_{t=-n}^{n} G(m,n)f(u+s,v+t)
\]

\[
g(C_u,C_v) = \arg \max_{u,v} (g(u,v))
\]

where \((m,n)\) is the size of the mask, \(G()\) is the Gaussian mask and \((u,v)\) is the 2-D coordinate of the image.

Projective Reconstruction

In order to obtain the 3-D information, we should know the point correspondence between the images from two stereo cameras. Because we do not have the point correspondence, we have to find the stereo correspondence between the two point sets. We use epipolar correspondence to achieve the candidate points. Epipolar correspondence describes the relationship between the pair of images. We can obtain the epipolar line of points in one image plane from the fundamental matrix, which we can obtain from the camera calibration. However the noises in the image and the errors in the cameras’ calibration process result in the corresponding epipolar line not passing through the matching point.

We use the modified Hausdorff distance (MHD) for robust and correct stereo matching. The HD measures how far the two subsets are from each other in a metric space. In the computer vision, the HD can be used to find a given template in an arbitrary target image. The candidate points with the minimal HD can be considered the best candidate for stereo matching. However HD is very sensitive to noise and overlap. To overcome this sensitivity, we use the MHD.

\[
H(X,Y) = \max \{h_{\text{MHD}}(X,Y), h_{\text{MHD}}(X,Y)\}
\]

\[
h_{\text{MHD}}(X,Y) = 1/N \sum_{x \in X \cap Y} \inf_d d(x,y)
\]

The MHD allows us obtain the robustness and a stable stereo matching result between the features of the left and the right camera. After projective reconstruction using the stereo matching correspondence, we can obtain 3-D position.

Model Indexing

After recovering the 3-D points from the 2-D correspondence of stereo camera images, the resulting 3-D point set is examined to find sub-structures that are similar to pre-calibrated model structures. If we can obtain the 3-D point set indexed with respect to the model structure, we can compute the position and the orientation of the helmet in a camera frame. The basic idea behind our model-indexing algorithm is to formulate the indexing problems in a way that allows for the application of structure alignment. We use the Geometric Hashing to find the correspondence between the 3-D points and the pre-calibrated model structure. The Geometric Hashing was originally developed in the computer vision area to match geometric features against a database of model features. Geometric Hashing consists of two steps, preprocessing and recognition. In the preprocessing step, the objects are encoded by each triplet of points as a geometric basis. The remaining points can be represented in invariant coordinates with respect to this basis. All such invariant coordinates of object points are stored in a 3-D table. In the recognition step, randomly selected triplets of 3-D points from image processing are considered as candidate bases. For each candidate basis, the remaining data points are encoded according to the basis and
possible correspondence from the object is found in the table, which is constructed in the preprocessing step. The candidate basis is determined if a sufficiently large number of the data points index an object basis. Figure 5 shows the indexing result of the Geometric Hashing.

**Pose Computation**

Determining the relationship between the two coordinate systems, by using the sets of corresponding feature measurements, is known as the absolute orientation problem. It has numerous applications in the fields of photogrammetry and robotics. In this paper, we use the well-known algorithm, developed by Horn (1987), which involves computing the eigensystem of a matrix related to representing the rotational component as a unit quaternion. With this algorithm, we can obtain the 3-D rigid body transformation between the helmet frame and the camera frame.

**MOTION PARAMETR ESTIMATION**

To satisfy accurate and robust helmet tracking, we chose to develop a hybrid system that is composed of a computer vision system with inertial sensors. The inputs of this hybrid system for motion parameter estimation are the measurements from the IMU and the pose computation results from the computer vision system. The measurements of the IMU are not (and will hardly be in most cases) synchronized with the measurements of the vision system. We assume that both sensors have a fixed sampling rate. Whenever there is a new measurement from either the inertial or the visual sensors, it updates the state and is fused in the motion estimation process. The prediction from one time step to the next is performed with the same equation independent of the type of measurement, but the dependencies between system states and measurements are totally different.

**Reference Frames**

When combining the two types of sensors, the frames in which measurements are made need to be taken into account. Figure 6 shows the several frames defined to be considered.

---

**Figure 5: Indexing result of the geometric hashing**

**Figure 6: Related reference frames of the hybrid system**

We define the camera frame \( \{C\} \) for the camera, IMU frame \( \{I\} \) for the IMU, helmet frame \( \{H\} \) for the helmet, world frame \( \{W\} \). The camera frame and world frame are fixed frames, and the IMU frame and helmet frame are moving frames with respect to the camera and the world frame, but they have a fixed transformation relationship with each other.

Since the inertial measurements performed by inertial sensors are given in the IMU frame \( \{I\} \), not in the camera frame \( \{C\} \), the rigid body transformation between the two has to be taken into account. This transformation can be expressed by the unit quaternion \( \mathbf{q} \) that rotates inertial measurements in the IMU frame \( \{I\} \) to the camera frame \( \{C\} \) and the translation vector \( \mathbf{r} \) from the IMU frame \( \{I\} \) to the camera frame \( \{C\} \).

Because cameras are fixed with respect to world, the inertial sensed acceleration of the helmet can be represented as follows:

\[
\mathbf{a}^C = \mathbf{q} \mathbf{a}^I \mathbf{q}^* \quad (3)
\]

In equation (3), \( \mathbf{a}^I \) is a sensed acceleration in the IMU frame.
Motion Model and System Equation

Our helmet motion model assumes constant angular velocity and constant translational acceleration. An Extended Kalman Filter (EKF) is used to integrate measurements from the IMU and vision system. To continuously estimate the motion parameters in 3-D space, the following discrete time and non linear model is used:

\[ x(k+1) = f[x(k), w(k)] \]

\[ y(k) = h(x(k)) + n(k) \]

where \( x \) is the state vector, \( f \) is the system dynamics, \( w \) describes system noise vector, \( y \) is the output vector, \( h \) is the mapping function from the state to the output and \( n \) is the measurement noise vector.

The state vector \( x \), the output vector \( y \), the system noise vector \( w \), and the measurement noise vector \( n \) are defined as follows:

\[ x(k) = \begin{bmatrix} p^T, v^T, a^T, q^T, \omega^T \end{bmatrix}^T \]

\[ y(k) = \begin{bmatrix} p_{mes}^T, v_{mes}^T, a_{mes}^T, \omega_{mes}^T \end{bmatrix}^T \]

\[ w(k) = \begin{bmatrix} j^T, a^T \end{bmatrix}^T \]

\[ n(k) = \begin{bmatrix} n_p^T, n_q^T, n_a^T, n_\omega^T \end{bmatrix}^T \]

where \( p \) is the position vector, \( v \) is the velocity vector, \( a \) is the acceleration vector of the helmet with respect to the world frame \{W\}, \( q \) is the quaternion representing the orientation of the helmet frame \{H\} with respect to the world frame \{W\} (this can avoid the jump from \( 2\pi \) to 0 by using Euler angles) and \( \omega \) is the angular velocity around the IMU frame axis. The output vector consists of position \( p_{mes} \) and quaternion \( \omega_{mes} \) from the vision system and acceleration \( a_{mes} \) and angular velocity \( \omega_{mes} \) from the IMU. The system noise is composed of jerks \( j \) and angular acceleration \( a \).

The acceleration of a system that describes a rotation and translation is \( a(k) = a_t(k) + \omega(k) \times v(k) \), where \( a_t(k) \) means the tangential acceleration and \( \omega(k) \times v(k) \) means the centripetal acceleration. The derivative of the acceleration is as follows:

\[ \frac{da(k)}{dt} = \frac{da_t(k)}{dt} + \frac{d\omega(k)}{dt} \times v(k) + \omega(k) \times \frac{dv(k)}{dt} \]

The general dynamic equations have the form as follows:

\[ p(k+1) = p(k) + \Delta v(k) + \frac{\Delta^2 v}{2} a(k) + \frac{\Delta^3 v}{6} j(k) \]

\[ v(k+1) = v(k) + \Delta t a(k) + \frac{\Delta^2 a}{2} j(k) \]

\[ a(k+1) = a(k) + \Delta t (j(k) + a_t(k) + \omega(k) \times v(k) + \omega(k) \times a(k)) \]

\[ q(k+1) = q(k | k+1) \otimes q(k) = \exp\left(\frac{\Delta \theta}{2}\right) \otimes q(k) \]

\[ \omega(k+1) = \omega(k) + \Delta t a(k) \]

where \( \Delta \theta(k) = \omega(k) \Delta t + \frac{1}{2} a(k) \Delta t^2 \). The term \( \otimes \) in equation (11) is the quaternion multiplication defined by Chou(1991) and the \( \exp(r) \) defined as the exponential map defined by Ude (1999) as follows:

\[ \exp(r) = \begin{bmatrix} \cos(|r|) & -|r| \sin(|r|) & 0 & 0 \\ 0 & \cos(|r|) & 0 & 0 \\ 0 & 0 & \cos(|r|) & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \]

The translational part of the state vector where \( x_t(k) = [p^T, v^T, a^T]^T \) can be computed as follows:

\[ x_t(k+1) = \begin{bmatrix} I_{3 \times 3} & \Delta t \cdot I_{3 \times 3} & \Delta t^2 \cdot I_{3 \times 3} \\ 0_{3 \times 3} & I_{3 \times 3} & \Delta t \cdot I_{3 \times 3} \\ 0_{3 \times 3} & 0_{3 \times 3} & I_{3 \times 3} \end{bmatrix} \begin{bmatrix} p(k) \\ v(k) \\ a(k) \end{bmatrix} + \begin{bmatrix} 0_{3 \times 3} \\ 0_{3 \times 3} \\ \frac{\Delta t^3}{6} \end{bmatrix} \]

Because the output equation is linear with the state, it can be written as follows:

\[ y(k) = H \cdot x(k) + n(k) \]

The matrix \( H \) describes the dependency between the measurement and the states. Two measurement matrices, \( H_I \) for the IMU and \( H_V \) for the vision system are defined as follows:

\[ H_I = \begin{bmatrix} 0_{3 \times 3} & 0_{3 \times 3} & I_{3 \times 3} & 0_{3 \times 4} & 0_{3 \times 3} \\ 0_{3 \times 3} & 0_{3 \times 3} & 0_{3 \times 4} & I_{3 \times 3} \end{bmatrix} \]

\[ H_V = \begin{bmatrix} 0_{4 \times 3} & 0_{4 \times 3} & I_{4 \times 3} & 0_{4 \times 4} & 0_{4 \times 3} \end{bmatrix} \]

The estimation of the actual motion parameters and its covariances is performed with EKF. The EKF equation is as follows:

\[ x(k+1) = f[x(k-1), 0] \]
\[ P'(k) = \nabla f_{k-1} P(k-1) \nabla f_{k-1}^T + \nabla w_{k-1} Q(k-1) \nabla w_{k-1}^T \]  
(18) 
\[ K(k) = P'(k) \nabla h_k \left[ \nabla h_k P'(k) \nabla h_k^T + R(k) \right]^{-1} \]  
(19) 
\[ \hat{x}(k) = \hat{x}(k) + K(k) \left[ y(k) - h(\hat{x}(k)) \right] \]  
(20) 
\[ P(k) = (I - K(k) \nabla h_k) P'(k) \]  
(21)

where \( P'(k) \) is the covariance of the predicted state \( \hat{x}'(k) \) and \( P(k) \) is the covariance of the estimated state \( \hat{x}(k) \) corrected with the measurement at time \( k \). \( \nabla f_{k-1} \) is the Jacobian matrix of partial derivatives of \( f \) with respect to \( x \), that is \( \nabla f_{k-1} = \left[ \frac{\partial f}{\partial x}(k) \right] \). \( \nabla w_{k-1} \) is the Jacobian matrix of partial derivatives of \( f \) with respect to \( w \), that is \( \nabla w_{k-1} = \left[ \frac{\partial f}{\partial w}(k) \right] \). \( \nabla h_k \) is the Jacobian matrix of partial derivatives of \( h \), that is \( \nabla h_k = \left[ \frac{\partial h}{\partial x}(k) \right] \).

EXPERIMENTAL RESULTS

This section describes the simulation results and the off-line test with the real data. The experimental setup is shown in Figure 8. The helmet is on the rate table and the stereo camera system is placed beside the rate table. We use a single-axis rate table (Acutronic AC1120S). The rate table rotates 100 degrees with the angular velocity 10deg/s counterclockwise around the \( z \) axis of the world frame. Figure 9 show the orientation estimation result. The orientation is represented by the quaternion. The red line is the reference. The blue line is the result of the proposed EKF, and the black line is the result when the only vision measurement is used. As shown in the enlarged plot in Figure 9, the proposed EKF shows the more accurate result than the case when the vision measurement is used.

CONCLUSION

This research demonstrates the hybrid helmet tracker system with the EKF, which fuses the vision and inertial information. If we use 2-D image points as the measurement of the vision system, the output equation becomes nonlinear and the filter equation becomes complicated. We use the position and orientation of the helmet as the measurement of the vision system to make the filter equation simple. There is also no need to synchronize the two measurements. Whenever there is a new measurement from either the IMU or the vision system, it updates the state by the same propagate equation. Our experimental results show that the proposed EKF can improve the accuracy of the tracking.

ACKNOWLEDGMENTS

This work was supported by Defense Acquisition Program Administration and Agency for Defense Development under the contract UD070041AD and the Ministry of Education, Science and Technology of Republic of Korea under NSL (National Space Lab, S10801000163-08A0100-16310).
REFERENCES


