pISSN:2093-274x eISSN:2093-2480
DOI:10.5139/IJASS.2010.11.1.031

Motion and Structure Estimation Using Fusion of Inertial and Vision Data for Helmet Tracker

Sejong Heo*, Ok shik Shin* and Chan Gook Park**
School of Mechanical and Aerospace Engineering Institute of Advanced Aerospace Technology,
Seoul National University, Seoul, 151-744, Korea

Abstract

For weapon cueing and Head-Mounted Display (HMD), it is essential to continuously estimate the motion of the helmet. The problem of estimating and predicting the position and orientation of the helmet is approached by fusing measurements from inertial sensors and stereo vision system. The sensor fusion approach in this paper is based on nonlinear filtering, especially expended Kalman filter(EKF). To reduce the computation time and improve the performance in vision processing, we separate the structure estimation and motion estimation. The structure estimation tracks the features which are the part of helmet model structure in the scene and the motion estimation filter estimates the position and orientation of the helmet. This algorithm is tested with using synthetic and real data. And the results show that the result of sensor fusion is successful.

Key Words : Sensor Fusion, Motion and Structure Estimation, Helmet Tracker

Introduction

Due to the complexity of the avionic system with modern high techniques, aircraft pilots face a problem in that the pilots don't concentrate on attacking targets in battles or control the aircraft. Many technologies have been developed to solve 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 estimates the motions of a head in military of pilot's head, we can estimate the line of sight more exactly. As a result, the helmet tracker 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[1].

The vision sensors are used successfully for this kind of tracking problems. But they are restricted to a small acceleration and rotation. On the other hand, inertial sensors are widely used for tracking rapid motion. But they suffer from the accumulated error with time, because of the double integration performed to compute the position. The fusion of inertial and vision sensors enables to make the helmet tracker system which is able to estimate the position and orientation of the helmet with long term stability, due to the periodic correction from the vision system and to track the rapid motion successfully, due to the properties of inertial sensors.

The combination of vision and inertial sensors has been used previously in mobile robotics, computer vision and augmented reality .The most popular solution is the EKF. The disadvantage of the EKF is that the computational load is very heavy, if the nonlinearity of the system is severe. Because of the high nonlinearity of the vision system, our previous helmet tracker

* Graduate Student
** Professor
E-mail : chanpark@snu.ac.kr
Tel : +82-2-880-1675 Fax : +82-2-880-7958

system suffers from high computational expenses[2]. To reduce the computation and improve the performance of the vision processing, the motion and structure estimation is used. We compute both structure and motion of the helmet simultaneously similar to Chai[3] which separates the motion estimation and structure estimation into two different EKFs. But rather than combining the positions of the features into a single large state vector, we separate them each other. As a result, a bank of EKF is used for structure estimation and each EKF uses the position of the feature as a state vector.

The term “motion” refers the position, velocity, acceleration, angular velocity and orientation of the helmet in the earth frame. We will use the term “pose” which refers the position and orientation of the helmet. The term “structure” refers the 3D positions of the features which are the part of the helmet model structure. The helmet model structure consists of the 3D location of LED features on the helmet. We identify the features in the image with comparison of this model structure.

In this paper, we will present a concept and evaluation of how to combine the structure estimation and motion estimation with two measurement channels. This is done by use of one EKF for the motion estimation which fuses the vision and inertial data and a filter bank for structure estimation. Each sensor system has a different measurement equation and the fusion of the two measurements is accomplished with common dynamic system model. We briefly introduce the hybrid helmet tracker system and the sensor fusion algorithm with structure and motion estimation. And results with synthetic and real sensor data will be introduced.

Hybrid Helmet Tracker System

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. Fig 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. Inertial data sampling rate is 100Hz and vision processing rate is 15Hz.

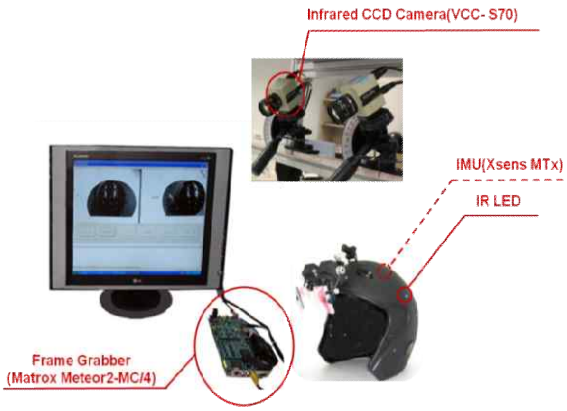

Fig. 1. Composition of Hybrid Helmet Tracker

Inertial sensors

The MTx IMU contains 5g MEMS accelerometers and 1200 degree/s MEMS rate gyroscopes. MEMS sensors are chosen because of their dramatically reduced size and low cost as compared to alternatives. The signals from the inertial sensors are synchronously measured at 100 Hz using a 16 bit A/D converter. A temperature sensor is added to compensate for the temperature dependency.

The calibrated gyroscope signal $y_{\omega,t}$ contains measurements of the angular velocity $\omega_{eb,t}^{b}$ from the body to earth expressed in body coordinate system.

$$y_{\omega,t} = \omega_{eb,t}^{b} + e_{\omega,t}^{b} \tag{1}$$

The measurements of the MEMS rate gyroscope are not accurate enough to pick up the rotation of the earth. This means that the earth coordinate system can be considered as an inertial frame.

The calibrated gyroscope signal $y_{\omega,t}$ contains measurements of the angular velocity $\omega_{eb,t}^{b}$ from the body to earth expressed in body coordinate system.

$$y_{a,t} = \ddot{b}_t^{b} - g^{b} + e_{a,t}^{b} \tag{2}$$

Gravity vector is constant in the earth coordinate system. But gravity vector in the body coordinate system depends on the orientation of the sensor unit. This means that we have to know the orientation of the sensor unit to estimate the acceleration.

Stereo Vision System

From Stereo CCD cameras, gray images with a resolution of 640x480 pixels at a frame rate with 15 Hz are received to a PC using frame grabber. Infrared LEDs on the helmet can be seen as a white bulb on dark background, because of the use of infrared cameras. Extracting the features and computing the 3D positions of the features is a known and well studied in computer vision. The key ingredient is to find the correspondence, relations between the features found in the two images which come from stereo cameras. Using two cameras side by side, stereo vision produces instantaneous estimation of the 3D position of features in the scene. As a result, stereo vision is highly effective for segmenting the objects or features from a background. The epipolar constraint is applied to find the relation between the features found in two images and reject outlier matches. Fig. 2 describes the epipolar geometry of the stereo vision system. The epiploar constraint states that if a point p in the image of the left camera and a point q in the image of the right camera correspond to the same 3D point in the camera coordinate system. They satisfy the following equation.

$$\mathbf{q}^{\mathrm{T}}\mathbf{F}\mathbf{p} = 0 \tag{3}$$

where $\mathbf{F}$ is the *fundamental matrix* that is the algebraic representation of epipolar geometry between the left and right images, and $\mathbf{p}$, $\mathbf{q}$ is the homogeneous representation of the position of the feature in the image plane. This equation means that the point q should have pass through the epipolar line defined as $\mathbf{Fp}$ in the right image plane. By applying this constraint to the stereo image pair, outliers are easily rejected and the 3D positions of the features on the helmet are estimated easily. Fundamental matrix $\mathbf{F}$ can be obtained from the camera calibration. Bouguet's camera calibration toolbox is used to obtain the internal parameters of each camera and the extrinsic relation between the two cameras[4].

Coordinate Systems

When working with a helmet containing inertial sensors and LED features and two cameras, several coordinate systems have to be considered.

- Earth coordinate system($\mathbf{e}$):

The pose of the helmet is estimated with respect to earth coordinate system. It is fixed to earth and the features of the scene are modeled in this coordinate system.

- Camera coordinate system($\mathbf{C}$):

This coordinate system attached to a fixed camera. Its origin is located in the optical center of the left camera with $\mathbf{z}$ axis aligned along the optical axis. The camera needs image plane ($\mathbf{i}$) to represent a projective image. This plane is perpendicular to the optical axis and is located at an offset from the optical center of the camera. This offset is called as focal length.

- Helmet coordinate system($\mathbf{H}$):

This coordinate system attached to a moving helmet. The 3D positions of the LED features represented in the helmet coordinate system are obtained after 3D scanning. These 3D positions organize the model structure of the helmet. Body coordinate system ($\mathbf{B}$) which is the coordinate system of the IMU is fixed to helmet coordinate system. They are separated by a constant translation and rotation.

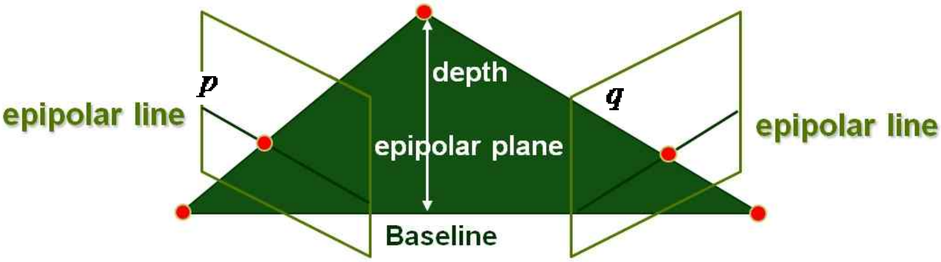

Fig. 2. Epipolar Geometry of Stereo Vision System

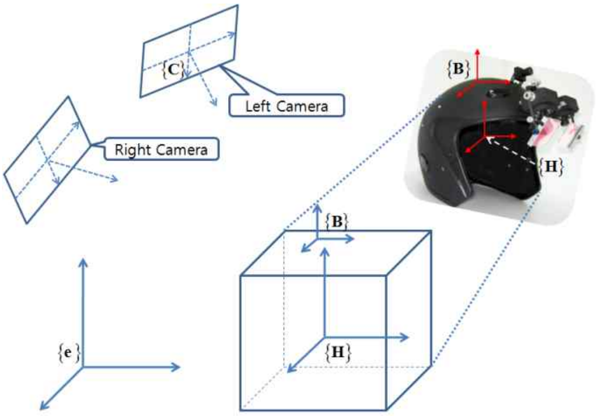

Fig. 3. Geometry of Coordinate Systems

These coordinate systems are used to denote geometric quantities in various coordinate systems and the relation between the coordinate systems. Acceleration and angular velocity vector received from the IMU are represented in the body coordinate system $\{\mathbf{B}\}$. And the 3D positions of the LED features from the vision system are represented in the camera coordinate system $\{\mathbf{C}\}$. Fig 3 describes these coordinate systems and their relationship.

Motion & Structure Estimation for Sensor Fusion

Algorithm Overview

The inertial and vision sensor have complementary properties. Vision processing by using structure estimation gives accurate absolute pose information at a slow motion, but has problems during rapid motion. The IMU provides high rate relative pose information regardless of the speed of motion, but suffers from a rapid accumulated error with elapsed time. By fusing the information from the two sensor systems, it is possible to obtain robust tracking result.

Fusing the inertial and vision data is possible in several ways. In this paper, the indirect method to combine the two types of sensors is proposed with the parallel use of structure estimation and motion estimation. The vision processing using structure estimation can be used for error correction of an inertial navigation system (INS). This is similar to how global positioning system (GPS) are used to correct the accumulated error in an INS.

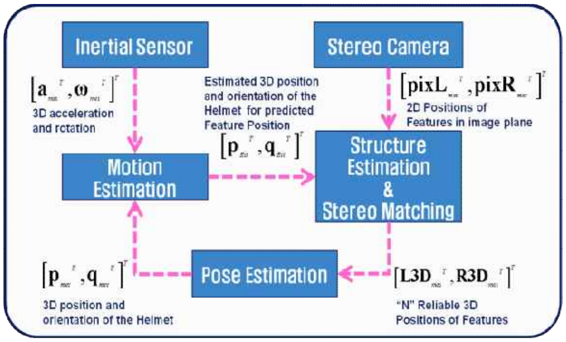
Inertial Sensor
Stereo Camera
$[\mathbf{a}^T, \boldsymbol{\omega}^T]^T$
3D acceleration and rctation
Estimated 3D position and orientation of the Helmet for predicted Feature Position
$[\mathbf{p}^T, \mathbf{q}^T]^T$
$[\mathbf{pixL}^T, \mathbf{pixR}^T]^T$
2D Positions of Features in image plane
Motion Estimation
Structure Estimation & Stereo Matching
$[\mathbf{p}^T, \mathbf{q}^T]^T$
3D position and orientation of the Helmet
Pose Estimation
$[\mathbf{L3D}^T, \mathbf{R3D}^T]^T$
"N" Reliable 3D Positions of Features

Fig. 4. Flow Chart of the Sensor Fusion Algorithm

The result of the structure estimation is the reliable 3D positions of the LED features on the helmet. And then the pose of the helmet is able to be obtained from the algorithm which involves with the eigen system of the matrix related to unit quaternion. This pose information is transmitted to the motion estimation filter and is fused with the inertial sensor data. Thus, the input from the vision system is a directly pose information of the helmet, not image points. This helps to simplify the filter equation of the motion estimation [5] [6].

The result of the motion estimation is used to predict the 2D positions of the feature in the left and right image plane. These predictions can be used to determine where in the image the LED features indexed with helmet model structure are to be expected. The term “indexed” means that we know what point in the model structure is matched with the LED feature in the image. As a result, the stereo matching and the tracking process of each LED feature are performed easily. Fig. 4 is the flow chart of the sensor fusion algorithm. In our application, there is no “main” sensor or “aiding” sensor. Both vision processing and inertial sensors are equivalent in the sense that they provide the information about the pose of the helmet.

Motion Estimation

The proposed head motion model assumes constant angular velocity and translational acceleration, implying that acceleration and angular velocity are included in the state vector. And position and orientation of the helmet are included in the state vector. Quaternion is used to describe the coordinate transformation between the two coordinate systems. We represent the head motion with the 16x1 state vector, $\mathbf{X}_{head} = \left[\mathbf{p}^T, \mathbf{v}^T, \mathbf{a}^T, \mathbf{q}^T, \boldsymbol{\omega}^T\right]^T$ where $\mathbf{p}$ is the position vector, $\mathbf{v}$ is the velocity vector, $\mathbf{a}$ is the acceleration vector of the helmet represented in $\{\mathbf{e}\}$. $\mathbf{q}$ is the quaternion representing the orientation of $\{\mathbf{H}\}$ with respect to $\{\mathbf{e}\}$ (this can avoid the jump from 2π to 0 by using Euler angles) and $\boldsymbol{\omega}$ is the angular velocity around the body coordinate system axis. The discretized dynamic system model is represented as follows.

$$\mathbf{p}_{k+1}^{e} = \mathbf{p}_{k}^{e} + \Delta t \mathbf{v}_{k}^{e} + \frac{\Delta t^{2}}{2}\mathbf{a}_{k}^{e} + \frac{\Delta t^{3}}{6}\mathbf{j}_{k}^{e} \tag{4.a}$$

$$\mathbf{v}_{k+1}^{e} = \mathbf{v}_{k}^{e} + \Delta t \mathbf{a}_{k}^{e} + \frac{\Delta t^{2}}{2}\mathbf{j}_{k}^{e} \tag{4.b}$$

$$\mathbf{a}_{k+1}^{e} = \mathbf{a}_{k}^{e} + \Delta t \mathbf{j}_{k}^{e} \tag{4.c}$$

$$\mathbf{q}_{k+1}^{Be} = \mathbf{q}(k+1\,|\,k) \otimes \mathbf{q}_{k}^{Be} = \exp\left(\frac{\Delta\theta}{2}\right) \otimes \mathbf{q}_{k}^{Be} \tag{4.d}$$

$$\boldsymbol{\omega}_{k+1}^{B} = \boldsymbol{\omega}_{k}^{B} + \Delta t \boldsymbol{\alpha}_{k}^{B} \tag{4.e}$$

where Δt is the time interval between the measurements, $\Delta\boldsymbol{\theta}_k = \boldsymbol{\omega}_k^B \Delta t + \frac{1}{2}\boldsymbol{\alpha}_k^B \Delta t^2$ and $\mathbf{exp}$ denote the the quaternition exponential map defined by Ude (1999) as follows[7].

$$\exp(\mathbf{r}) = \begin{cases} \begin{bmatrix} \cos(|\mathbf{r}|) \\ \sin(|\mathbf{r}|)\dfrac{\mathbf{r}}{|\mathbf{r}|} \end{bmatrix}, & |\mathbf{r}| \neq 0 \\ \begin{bmatrix} 1 \\ \mathbf{0}_{3\times1} \end{bmatrix}, & |\mathbf{r}| = 0 \end{cases} \tag{5}$$

The system noise is composed of jerks $\mathbf{j}_k^e$ and angular acceleration $\boldsymbol{\alpha}_k^B$. The advantage of this formulation is that all entries of system noise are dependent on the time interval Δt and the infulence of noise will change with the time between the two measurements. If the entries of the process noise is independent on the time, the system noise will not change and the noise will be too low or too high at the worst case.

We have the two types of measurement , one is from the inertial system and the other is from the vision system. Each type of sensor system has associated with the each measurement models. The relationship between the state and the measurement of the inertial system is as follows.

$$\mathbf{y}_{a,k} = \mathbf{q}_k^{Be}\left(\mathbf{a}_k^e - g^e\right)\mathbf{q}_k^{Be*} + \mathbf{e}_{a,k}^B = \mathbf{R}_k^{Be}\left(\mathbf{a}_k^e - g^e\right) + \mathbf{e}_{a,k}^B \tag{6}$$

$$\mathbf{y}_{\omega,k} = \omega_{eb,k}^{b} + \mathbf{e}_{\omega,k}^{b} \tag{7}$$

Note that the rotation matrix $\mathbf{R}_k^{be}$ is constructes from $\mathbf{q}_k^{be}$. The relationship between the state and the measurement from the vision processing is as follows.

$$\mathbf{y}_{p,k} = P_k^e + \mathbf{e}_{p,k}^e \tag{7}$$

$$\mathbf{y}_{q,k} = \mathbf{q}_k^{Be} + \mathbf{e}_{q,k} \tag{8}$$

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

$$\mathbf{y}(k) = \mathbf{H} \cdot \mathbf{x}(k) + \mathbf{n}(k)$$

The matrix $\mathbf{H}$ describes the dependency between the measurement and the states. Two measurement matrices, $\mathbf{H}_I$ for the IMU and $\mathbf{H}_V$ for the vision system are defined as follows:

$$\mathbf{H}_I = \begin{bmatrix} \mathbf{0}_{3\times3} & \mathbf{0}_{3\times3} & \mathbf{R}_{3\times3}^{Be} & \mathbf{0}_{3\times4} & \mathbf{0}_{3\times3} \\ \mathbf{0}_{3\times3} & \mathbf{0}_{3\times3} & \mathbf{0}_{3\times3} & \mathbf{0}_{3\times4} & \mathbf{I}_{3\times3} \end{bmatrix} \tag{9}$$

$$\mathbf{H}_V = \begin{bmatrix} \mathbf{0}_{3\times3} & \mathbf{0}_{3\times3} & \mathbf{I}_{3\times3} & \mathbf{0}_{3\times4} & \mathbf{0}_{3\times3} \\ \mathbf{0}_{4\times3} & \mathbf{0}_{4\times3} & \mathbf{0}_{4\times3} & \mathbf{I}_{4\times4} & \mathbf{0}_{4\times3} \end{bmatrix} \tag{10}$$

The EKF is used for motion estimation. A block diagram of the motion estimation filter is shown in Fig.5. The inputs to this filter are the measurement of the inertial system and the pose estimation result from the vision processing. There is no need to synchronize the two measurements. The prediction from one measurement to the next measurement is performed with the same equation independent with the type of measurement. Whenever the new measurement is received, it updates thestate regardless of whether the type of measurement is inertial or vision, and is fused in motion estimation filter[8].

Structure Estimation

The helmet model structure of the helmet consists of ‘L’ 3D points which represent the LED features on the helmet, and the structure from the image is represented by ‘N’ reliable 3D points. N is the number of LED features tracked in the image sequence. In this paper, the helmet structure is estimated by a bank of simple EKF. Each filter has a 3x1 state vector, $\mathbf{X}_{i,structure} = \left[x_i^C, y_i^C, z_i^C\right]^T$, which is the 3D position of the LED feature represented in the camera coordinate frame. Even though new features appear or some features disappear in the scene, proposed structure estimation algorithm can track the features robustly. As you can see in Algorithm 1, we start with the computation of the predicted 2D positions of the helmet structure in the image based on the result of the motion estimation filter. Second, we find the predicted point which is nearby the feature point segmented from input image under certain threshold and match the feature point with the points in the model structure. Line 2 provides what points of the model structure are in the left and right image. As a result, we can match the points in the two image plane if they are related with the same structure point. The 3D positions of the features indexed with the model structure are obtained by structure estimation. Line 4 is the pose estimation process by using the algorithm which involves with the eigen system of the matrix related to unit quaternion[9]. And then the pose estimation results are transmitted to the motion estimation filter.

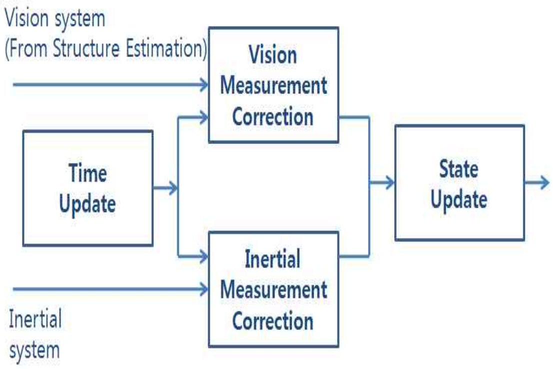
Vision system
(From Structure Estimation)
Vision Measurement Correction
Time Update
State Update
Inertial Measurement Correction
Inertial system

Fig. 5. Block Diagram of the Motion Estimation Filter

Algorithm 1. Vision Processing Algorithm with Structure Estimation

1. Compute $\mathbf{X}_{L,Backprojeted} = \left[\mathbf{x}^{i}_{1,L}, \mathbf{x}^{i}_{2,L}, \cdots, \mathbf{x}^{i}_{n,L} \right]$ $\mathbf{X}_{R,Backprojeted} = \left[\mathbf{x}^{i}_{1,R}, \mathbf{x}^{i}_{2,R}, \cdots, \mathbf{x}^{i}_{n,R} \right]$

   which are the predicted 2D points of the helmet structure in the left and right image from the motion estimation filter. n is the number of the structure points.

2. Find the correspondence between the points of the LED features $\left(\mathbf{p}_{L,input} = \left[\mathbf{p}^{i}_{1,L}, \mathbf{p}^{i}_{2,L}, \cdots, \mathbf{p}^{i}_{m_1,L} \right], \mathbf{p}_{R,input} = \left[\mathbf{p}^{i}_{1,R}, \mathbf{p}^{i}_{2,R}, \cdots, \mathbf{p}^{i}_{m_2,R} \right] \right)$ in the input image

   and the predicted points from 1 $\left(\mathbf{X}_{L,Backprojeted}, \mathbf{X}_{R,Backprojeted} \right)$. See Fig.6

   (a) Find the predicted point which has the minimum distance from the feature point. Check the threshold and the minimum distance.

   (b) Find the correspondence between $\mathbf{p}^{\mathbf{i}}_{j,L}$ and $\mathbf{p}^{\mathbf{i}}_{k,R}$ which are matched to the same structure point.

3. Estimate the structure of the helmet.

   (a) Compare the previous structure and the current structure.

   (b) If the structure point is in the both structure, estimate the 3D position of the structure point by using the information of the previous estimation. $(\mathbf{P}_{k-1})$

   (c) If new structure point is added in the current structure, add a new filter and estimate the 3D position with the initial error covariance $(\mathbf{P}_{ini})$.

   (d) If the structure point is discarded, discard the filter which is related with the structure point.

4. Estimate the pose information from the algorithm using the unit quaternion. Transmit the information to the motion estimation filter.

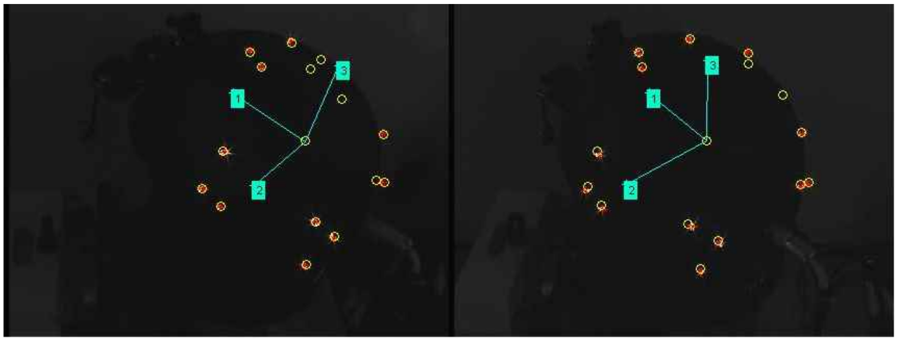

Fig. 6. Projected model structure (yellow circle)and the segmented features(red *)

Experimental Results

Synthetic Data

We examined the performance of the algorithm on purely synthetic data. Three sample motion of the helmet were described and the synthetic sensor outputs were generated. The input images are synthesized by calculating the back-projected 2D image point of the helmet model structure. And white Gaussian noise is added to this data. The inertial data are also generated, which has zero mean white Gaussian noise. The standard deviation for gyro data is $0.01 rad/\sec$ and the standard deviation for accelerometer is $0.02\ m/\sec^2$.

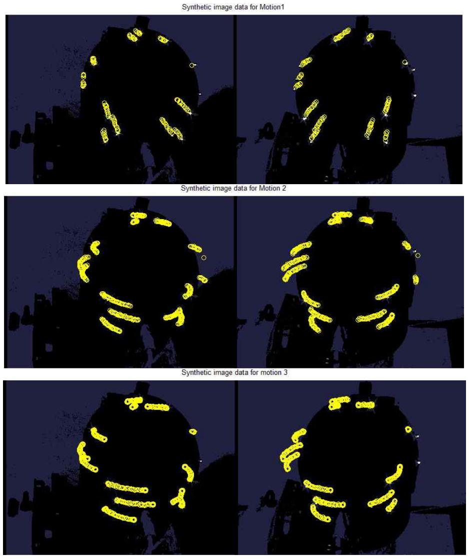
Synthetic image data for Motion1
Synthetic image data for Motion 2
Synthetic image data for motion 3

Fig. 7. Synthetic image data: (a) motion1 (b) motion2 (c) motion3

The first motion is pure translation of $75mm$ along the $\mathbf{x}$ axis with constant velocity $75mm/\sec$ and the second motion is pure rotation of 45 degree about $\mathbf{z}$ axis of the earth coordinate system. The third motion is complex rotation motion with the 3 degree change in roll, 6 degree change in pitch and 45 degree change in yaw. Fig.6 shows the synthetic image data for three motions. The estimation results are as follows.

Fig. 7, 8, 9 shows the estimation results of each motion sequentially. In the each figure, the solid line is the estimation result of sensor fusion algorithm, the dash-dot line is the

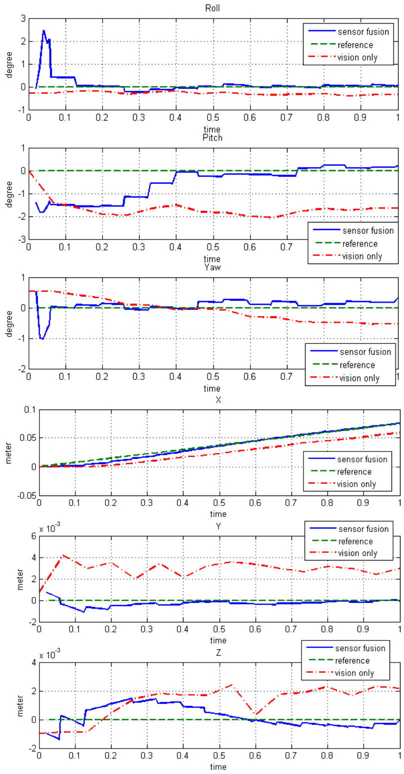

Fig. 8. Motion1 Estimation Result: (a) Orientation (b) Position

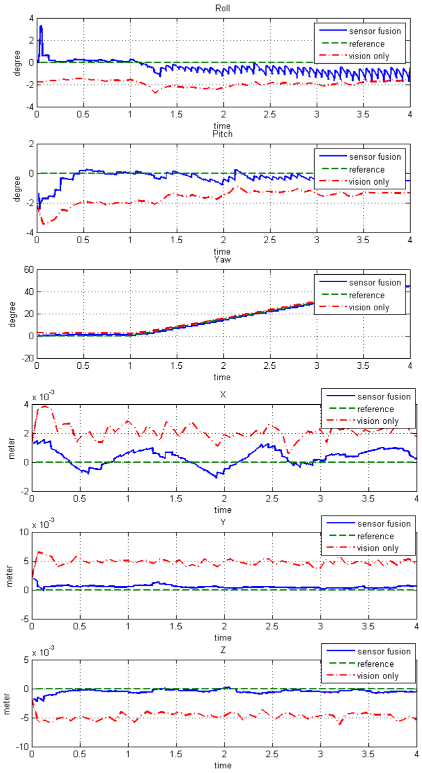

Fig. 9. Motion2 Estimation Result: (a) Orientation (b) Position

estimation result when the vision data is only used and the dash line is the reference. We can see the estimated result is very close to the reference using the sensor fusion algorithm and is better than the estimation results when the vision data is only used.

Real Data

Next we test the algorithm with real image data and the real inertial sensor data. With single axis rate table, we test the rotation motion with real data. The experimental set-up is shown in Fig.10(a). The camera calibration is preformed using Bouguet’s camera calibration toolbox (Bouguet 2006). The toolbox uses a pinhole camera model with nonlinear radial and tangential distortion compensation. The calibration uses images of a checkered target in several positions and recovers the camera’s intrinsic parameters, as well as the extrinsic parameters between two cameras as show in Fig. 10(b). The alignment between the camera frame and the earth frame is performed with the image of checker board similar to the camera calibration.

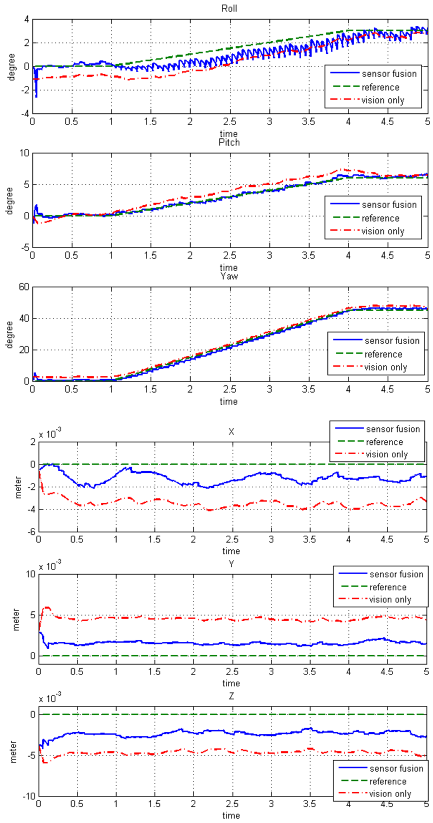

Fig. 10. Motion3 Estimation Result: (a) Orientation (b) Position

In the case of real data, the alignment error between the camera coordinate systems and earth coordinate systems affects the result of the estimation as a bias. And the alignment error between the body coordinate system and the helmet coordinate system affects the inertial measurement data. And also, our experimental system is not accurate, because we cannot align the helmet exactly along the coordinate system which is located at the center of rotating plate of the rate table. As a result, the reference is not accurate as we think.

The estimation result shows on the Fig.11. The solid line is the estimated result and the dash line is the reference. The helmet is rotated along the **z** axis with the high angular velocity of 75deg/sec . Because we assume the constant angular velocity model, the error increases when the angular velocity is changed suddenly, especially at roll and pitch estimation result. But the result of yaw estimation result is accurate. The position estimation result shows the origin of the helmet moves on the circle, because the alignment error of the helmet origin and the center of rotation. But the result of **z** axis is accurate.

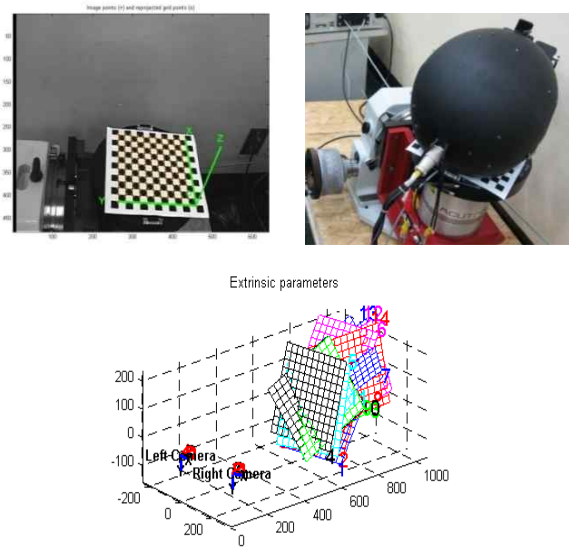

Fig. 11. (a) Experimental set−up (b) Alignment (c) Camera Calibration

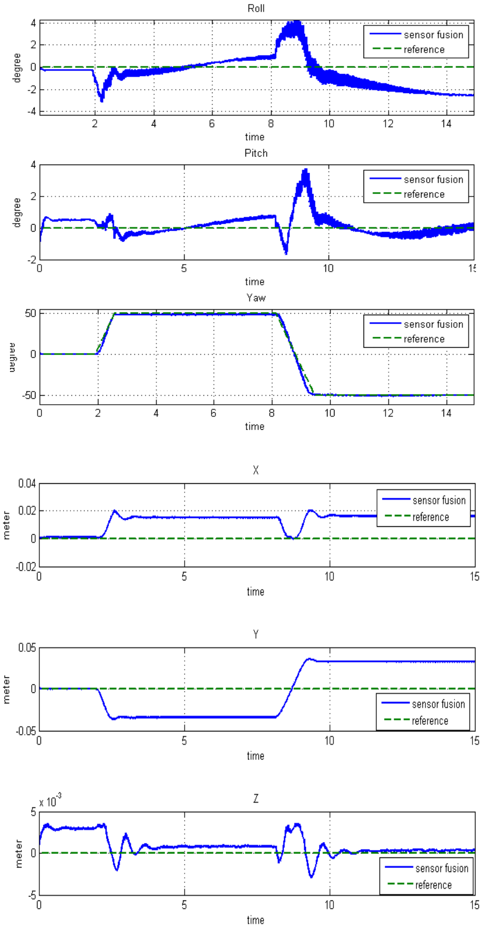

Fig. 12. Real Data Estimation Result: (a) Orientation (b) Position

Although the filter converges for real data, the errors are larger than for synthetic image data. Another reason for larger error especially in the roll and pitch is probably due to increase of errors in vision processing. When we rotate the head along the **z** axis, some points disappear and appear suddenly. This increases the error of the structure estimation. And the error of the depth on the projective reconstruction affects roll and pitch more than yaw.

Conclusion

We proposed a sensor fusion algorithm that includes motion estimation of the pose of user’s head and structure estimation of the 3D positions of helmet structure. The structure estimation reduces the computation time and improve the performance in vision processing. . We do not address the many other problems in detail, such as computing the initial pose for motion estimation and calibration. We experimentally found the good estimation result with synthetic data and the real data. But the sensor fusion algorithm can be more advanced by the adaptive technique for the type of motion, because optimal filter parameters for slow and fast motion are little different. And another shape of helmet structure will be tested to reduce the errors in the roll and pitch.

Acknowledgement

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

1. E. Foxlin, Y. Altshuler, L. Naimark and M. Harington, “FlightTracker: A Novel Optical/ Inertial Tracker for Cockpit Enhance Vision”, IEEE/ACM International Sysmposium on Mixed and Augmented Reality November2−5, 2004, Washington.D.C.

2. Y.I. Kim, Y.J. Lee and C.G. Park, “Hybrid Head Tracker System to Compose Optical /Inertial

Head Tracker System", ICSIIT 2007,Bali, Indonesia, July 26 ~27, 2007.

3. Lin Chai, William A. Hoff, Tyrone Vincent, "Three−dimensional motion and structure estimation using inertial sensors and computer vision for augmented reality", Presence: Teleoperators and Virtual Environments, v.11 n.5, p. 474−492, October 2002.

4. Richard Hartley, Andrew Zisserman, "Multiple View Geometry in Computer Vision" (2nd edition), Cambridge University Press, 2004

5. Peter Gemeiner, Peter Einramhof and Markus Vincze, Zienkiewicz, "Simultaneous Motion and Structure Estimation by Fusion of Inertial and Vision Data", The international Journal of Robotics Research, Vol. 26, No. 6, June 2007, pp. 591−605.

6. S.G. Chroust and M. Vincze, "Fusion of Vision and Inertial Data for Motion and Structure Estimation" , The Journal of Robotic Systems, Vol. 21, No. 2, 2004, pp. 73−83.

7. A. Ude, "Filtering in a Unit Quaternion Space for Model−Based Object Tracking", The Journal of Robotics and Autonomous Systems, Vol. 28, No. 2, August, 1999, pp. 163−170.

8. Greg Welch and Gary Bishop, "An Introduction to the Kalman Filter", SIGGRAPH 2001, Course 8.

9. Berthold K.P. Horn, "Closed form solution of absolute orientation using unit quaternion", Journal of Optical Society of America, Vol. 4, April, 1987, pp. 629−642.