Fast Relative Pose Calibration for Visual and Inertial Sensors

Jonathan Kelly and Gaurav S. Sukhatme

Abstract Accurate vision-aided inertial navigation depends on proper calibration of the relative pose of the camera and the inertial measurement unit (IMU). Calibration errors introduce bias in the overall motion estimate, degrading navigation performance - sometimes dramatically. However, existing camera-IMU calibration techniques are difficult, time-consuming and often require additional complex apparatus. In this paper, we formulate the camera-IMU relative pose calibration problem in a filtering framework, and propose a calibration algorithm which requires only a planar camera calibration target. The algorithm uses an unscented Kalman filter to estimate the pose of the IMU in a global reference frame and the 6-DoF transform between the camera and the IMU. Results from simulations and experiments with a low-cost solid-state IMU demonstrate the accuracy of the approach.

1 Introduction

Many mobile robotics tasks require a robot to accurately estimate its pose, or position and orientation in a reference frame, over time. Recent work has shown that visual and inertial sensors, in combination, can provide very accurate pose estimates [1, 2]. This makes them ideal for use in environments where global pose information, from e.g. GPS, is either unavailable or unreliable.

In particular, progress in the fabrication of micro-electrical mechanical systems has led to the development of reliable, low-cost Inertial Measurement Units (IMUs). An IMU typically consists of three orthogonal, single-axis accelerometers and three orthogonal angular rate gyroscopes. Solid-state IMUs

Department of Computer Science
University of Southern California
{jonathsk,gaurav}@usc.edu
http://robotics.usc.edu/resl/

Revised Sept. 21, 2010

are designed for use in a 'strapdown' configuration, with the accelerometers and gyroscopes fixed to a common chassis and not actively gimbaled to maintain a fixed orientation [3]. The change in pose of a strapdown IMU can, in theory, be determined by double-integrating the accelerometer outputs over time, using rate information from the gyroscopes to determine orientation. In reality, inertial sensors are subject to low-frequency drift, and therefore some form of aiding is normally required to maintain the integrity of the pose information.

When a camera is used for aiding (in *vision-aided inertial navigation*, or V-INS), it is important that the relative pose of the sensors be accurately known so that measurements can be properly fused. Errors in the estimated rotation or translation between the camera and the IMU will introduce unmodeled biases in the pose estimate – over time, these biases can cause the overall pose error to grow without bound. For applications in which accuracy is important, calibration is therefore a necessity. However, existing calibration techniques are time-consuming, and require the use of additional apparatus [4].

In this paper, we formulate the camera-IMU relative pose calibration problem in a filtering framework, and use the constraints imposed by rigid body dynamics to solve for the transform between the sensor reference frames. If the sensors are rigidly attached to the same platform, then any change in the orientation of one sensor is accompanied by an equivalent change in the orientation of the other sensor, measured in the appropriate sensor-specific frame. Likewise, the length of the moment arm between the sensors determines how the two may move in relation to each other.

To solve for the six degrees-of-freedom (6-DoF) relative pose, we record measurements from the IMU and images from the camera while the camera views a planar calibration target. We track known target points, and use this information in an unscented Kalman filter (UKF) to estimate the transform between the sensors. Our choice of the UKF is motivated by its superior performance compared to the extended Kalman filter (EKF) for many nonlinear problems. The calibration algorithm also provides a measure of the uncertainty associated with the relative pose estimate (i.e. a covariance matrix), and can therefore be easily integrated with other estimators.

2 Related Work

Vision-aided inertial navigation is an active research area in robotics [5]. The complimentary frequency response and noise characteristics of cameras and IMUs make the sensors suitable for use in combination to accurately estimate the ego-motion of a robot over time [6]. Advances in computing hardware have recently enabled the development of practical, online V-INS systems [2, 7].

For vehicle applications, an initial alignment procedure is usually carried out to determine the orientation of the IMU in the vehicle navigation frame. The absolute positioning information required for alignment can be obtained from e.g. a GPS receiver [8]. Similarly, for V-INS, it is important to know the relative alignment between the camera and the IMU, and the relative translation between the sensors, so that measurements from both devices can be correctly fused in the navigation frame.

Several visual-inertial calibration techniques have been proposed in the literature. Lang and Pinz [9] uses a constrained nonlinear optimization algorithm to solve for the rotation between a camera and an IMU. The camera is mounted on a tripod and manually rotated while gyroscope data is recorded. By measuring the relative angle to several external markers viewed in the camera image, and comparing these values with the integrated gyro outputs, the algorithm determines the rotation which best aligns the sensor frames. The approach ignores the relative translation between the sensors, however.

Lobo and Dias [4] describes a camera-IMU calibration procedure in which the relative orientation and relative translation between the sensors are determined independently. First, the rotational offset of the camera frame relative to the IMU frame is found by rotating the sensors using a pendulum unit while the camera views a planar calibration target. The relative translation is then determined by spinning the camera and the IMU on a turntable, while positioning the IMU such that its measured horizontal acceleration is zero. A drawback of the technique is that separately calibrating the relative rotation and translation decouples the estimates and therefore ignores any correlations that exist between the parameters.

Our technique, in contrast, does not require additional apparatus beyond a camera calibration target, provides a measure of the uncertainty associated with the transform, and can be used *online* to quickly re-calibrate the relative pose between the sensors if either the camera or the IMU is repositioned.

3 Calibration Algorithm

The goal of the calibration procedure is to accurately determine the 6-DoF rigid body transform between the camera optical center and the translation and rotation center of the IMU. We describe our system model below, and then briefly discuss our implementation of the unscented Kalman filter. Three separate reference frames are considered:

1. the *camera frame* $\{C\}$, with its origin at the optical center of the camera and with the z-axis aligned with the optical axis of the lens,
2. the *IMU frame* $\{I\}$, with its origin at the center of the IMU body, in which linear accelerations and angular rates are measured, and
3. the *global frame* $\{G\}$, with its origin at the upper left-hand corner of the calibration target.

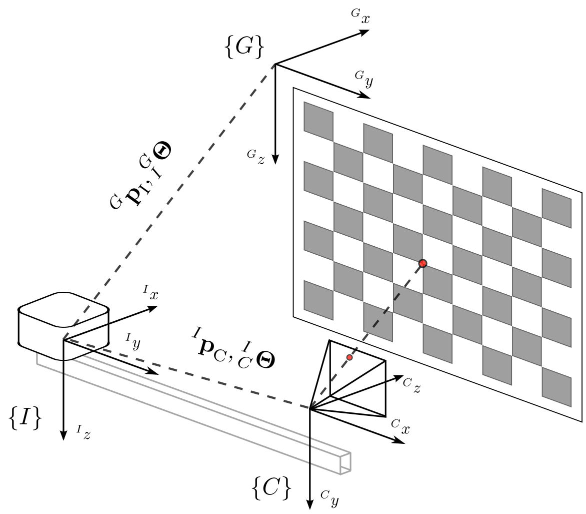


Fig. 1: Global, $\{G\}$, IMU, $\{I\}$, and camera, $\{C\}$, reference frames. The transform from the IMU frame to the global frame is defined by the $({}^{G}\mathbf{p}_{\mathrm{I}}, {}^{G}_{I}\boldsymbol{\Theta})$ translation and rotation pair. The (unknown) transform from the camera frame to the IMU frame is defined by the $({}^{I}\mathbf{p}_{\mathrm{C}}, {}^{I}_{C}\boldsymbol{\Theta})$ translation and rotation pair.

The global frame is an inertial frame and serves as an absolute reference for both the camera and the IMU.[1] Figure 1 illustrates the relationship between the frames.

3.1 System Description

Our approach uses an UKF to simultaneously estimate the pose of the IMU in the global frame and position and orientation of the camera relative to the IMU. The 21×1 system state vector is:

$$\mathbf{x} = \begin{bmatrix} {}^{G}\mathbf{p}_{\mathrm{I}}^{T} & {}^{G}_{I}\boldsymbol{\Theta}^{T} & {}^{G}\mathbf{v}_{\mathrm{I}}^{T} & \mathbf{b}_{a}^{T} & \mathbf{b}_{g}^{T} & {}^{I}\mathbf{r}_{\mathrm{C}}^{T} & {}^{I}_{C}\boldsymbol{\Theta}^{T} \end{bmatrix}^{T} \tag{1}$$

where ${}^{G}\mathbf{p}_{\mathrm{I}}$ is the position of the IMU in the global frame, ${}^{G}_{I}\boldsymbol{\Theta} = \begin{bmatrix} \alpha \; \beta \; \gamma \end{bmatrix}^{T}$ is the vector of roll, pitch and yaw Euler angles which define the orientation of the IMU frame with respect to the global frame, ${}^{G}\mathbf{v}_{\mathrm{I}}$ is the linear velocity

[1] In fact, the global frame is not strictly an inertial frame, since it is attached to the surface of the rotating Earth. However, the effects of the Earth's rotation are very small over the calibration time interval, and we therefore ignore them.

of the IMU in the global frame, and $\mathbf{b}_a$ and $\mathbf{b}_g$ are the accelerometer and gyroscope biases, respectively. The remaining entries, ${}^I\mathbf{r}_\mathrm{C}$ and ${}^I_C\mathbf{\Theta}$, define the position and orientation of the camera frame relative to the IMU frame. These values are *parameters*, or static quantities, which we seek to estimate as part of the calibration process. Recent work by Mirzaei and Roumeliotis [10] has shown that the system is fully observable given camera and IMU measurements alone.

3.1.1 Process Model

The filter process model is driven by the IMU linear acceleration and angular velocity measurements. IMU accelerometer and gyroscope biases are modeled as Gaussian random walk processes driven by the white noise vectors $\mathbf{n}_{aw}$ and $\mathbf{n}_{gw}$ [11]. The accelerometer and gyroscope measurements are assumed to be corrupted with zero-mean Gaussian noise vectors $\mathbf{n}_a$ and $\mathbf{n}_g$, respectively. The time evolution of the system state is described by:

$$
{}^G_I\dot{\mathbf{\Theta}} = \mathbf{\Gamma}({}^G_I\mathbf{\Theta})\,{}^I\omega \tag{2}
$$

$$
{}^G\dot{\mathbf{p}}_\mathrm{I} = {}^G\mathbf{v}_\mathrm{I} \qquad {}^G\dot{\mathbf{v}}_\mathrm{I} = \mathbf{C}({}^G_I\mathbf{\Theta})\,{}^I\mathbf{a}_\mathrm{I} - {}^G\mathbf{g} \tag{3}
$$

$$
\dot{\mathbf{b}}_a = \mathbf{n}_{aw}, \qquad \dot{\mathbf{b}}_g = \mathbf{n}_{gw} \tag{4}
$$

$$
{}^I\dot{\mathbf{r}}_\mathrm{C} = \mathbf{0}_{3\times1} \qquad {}^I_C\dot{\mathbf{\Theta}} = \mathbf{0}_{3\times1} \tag{5}
$$

where $\mathbf{\Gamma}$ is the Euler kinematical matrix which relates the rate of change of the Euler angles to the IMU angular velocity, $\mathbf{C}$ is a direction cosine matrix and ${}^G\mathbf{g}$ is the gravity vector in the global frame. The vectors ${}^I\mathbf{a}_\mathrm{I}$ and ${}^I\omega$ are the linear acceleration and angular velocity, respectively, of the IMU in the IMU frame. These values are related to the *measured* IMU linear acceleration, $\mathbf{a}_m$, and angular velocity, ω_m, by:

$$
\mathbf{a}_m = {}^I\mathbf{a}_\mathrm{I} + \mathbf{b}_a + \mathbf{n}_a = \mathbf{C}^T({}^G_I\mathbf{\Theta})({}^G\mathbf{a}_\mathrm{I} + {}^G\mathbf{g}) + \mathbf{b}_a + \mathbf{n}_a \tag{6}
$$

$$
\omega_m = {}^I\omega + \mathbf{b}_\omega + \mathbf{n}_\omega \tag{7}
$$

After each IMU measurement, the state is propagated forward in time until the next camera or IMU update using fourth-order Runge-Kutta integration of Equations 2 through 5 above.

3.1.2 Measurement Model

As the sensor beam moves through space, the camera captures images of known points on the calibration target. The points are tracked using the KLT feature tracker [12], after each image has been rectified to remove lens distortions. Projections of the target points in the camera images can be used to determine the absolute position and orientation of the camera in the global frame. In our case, the measurement residuals consist of the difference between the observed positions of the target points and their predicted positions.

We use an ideal projective (pinhole) camera model, and assume that the intrinsic calibration parameters of the camera and lens are known. Measurement $\mathbf{z}_i$ is the projection of target point p_i, at position ${}^{C}\mathbf{p}_{p_i}$ in the camera frame, onto the image plane:

$$ {}^{C}\mathbf{p}_{p_i} = \begin{bmatrix} x_i \\ y_i \\ z_i \end{bmatrix} = \mathbf{C}^T({}^{I}_{C}\boldsymbol{\Theta})\,\mathbf{C}^T({}^{G}_{I}\boldsymbol{\Theta})\left({}^{G}\mathbf{p}_{p_i} - {}^{G}\mathbf{p}_{\mathrm{I}}\right) - \mathbf{C}^T({}^{I}_{C}\boldsymbol{\Theta})\,\mathbf{r}_{\mathrm{C}} \tag{8} $$

$$ \mathbf{z}_i = \begin{bmatrix} u_i \\ v_i \end{bmatrix} = \mathbf{h}(\mathbf{x}, {}^{G}\mathbf{p}_{p_i}) + \boldsymbol{\eta}_i = \begin{bmatrix} x'_i \\ y'_i \end{bmatrix} + \boldsymbol{\eta}_i, \quad \begin{bmatrix} x'_i \\ y'_i \\ 1 \end{bmatrix} = \mathbf{K} \begin{bmatrix} x_i/z_i \\ y_i/z_i \\ 1 \end{bmatrix} \tag{9} $$

where $\begin{bmatrix} u_i, v_i \end{bmatrix}^T$ is the vector of observed image coordinates, $\mathbf{K}$ is the 3×3 camera intrinsic calibration matrix, and η_i is a Gaussian measurement noise vector. The initial position of the camera with respect to the target is determined using an iterative least squares computation. We use a homography-based check and a RANSAC procedure to remove outliers due to tracking errors.

3.2 Unscented Filtering

The UKF [13] captures the mean and covariance of a probability distribution with a set of deterministically-selected sample points called *sigma points*, which lie on the covariance contours of the N-dimensional state space. To update the system state, the UKF generates a set of $2N + 1$ sigma points – each point is propagated through the (nonlinear) system process and measurement models to compute the posterior state mean and covariance. This approach is attractive because it avoids many of the problems that can result from linearization (in, e.g., the EKF), and has third-order accuracy for Gaussian error distributions. An additional benefit of the UKF is that it does not require the computation of Jacobian matrices.

The most straightforward implementation of the UKF augments the state vector and state covariance matrix with process and measurement noise components. In our case, we stack the IMU bias and noise vectors to form a 12×1 process noise vector, and add a 2×1 camera measurement noise vector, producing an augmented 35×1 system state vector. Computing the sigma points requires a matrix square root, which is found by Cholesky decomposition. The individual sigma points are then propagated through the process model and combined in a weighted average to generate an updated state mean and state covariance matrix.

When a measurement arrives (i.e. in our case, an observation of one of the corners of the calibration target projected into the camera image plane), we determine the predicted measurement value by propagating each sigma point through the (nonlinear) measurement function. We then perform a state update by computing the Kalman gain matrix and the *a posteriori* state vector and state covariance matrix.

4 Simulation Studies

In order to examine the performance of the calibration algorithm, we initially performed a series simulation studies with ground truth available. We modeled a sensor beam, 50 cm in length, moving through space according to accurate rigid body kinematics. Simulated IMU updates occurred at a rate of 60 Hz, while camera images arrived at 30 Hz; these update rates match those of the actual hardware used for the experiments described in Section 5. The process noise for the simulated IMU was the same as the estimated noise for the IMU available in our laboratory.

During each simulation, the sensor beam moved along a rotating, spiral trajectory to excite the full six degrees of rotational and translational freedom. We constrained the beam trajectory to ensure that all points on the calibration target remained visible throughout the entire simulation run.

For each simulated camera image, we projected the known target points into the image plane and added Gaussian noise with a standard deviation of 1.0 pixels in the u and v directions to the image coordinates. The simulated camera had a horizontal field of view of $45°$ and a resolution of 640×480 pixels.

Results from one simulation trial are shown in Figure 2 and Table 1; although the data is for a single trial only, these results are typical of the majority of our simulations. For the trial indicated, the initial translation error was five centimeters along each axis and the initial rotation error was eight degrees in all three axes (roll, pitch and yaw).

The plots in Figure 2 show rapid convergence of the orientation estimates, although there is a small residual error in the pitch value. For the translation, the Y and Z axes estimates also converge, although more slowly, to the approximate true values, and the error remains within the 3σ bounds for the entire simulation. The X axis error, in contrast, converges very slowly, and has a residual bias of approximately 1.27 cm, or 25% of the initial error. This suggests that, for the given beam trajectory, the X offset of the camera is only weakly observable. We are currently investigating this issue.

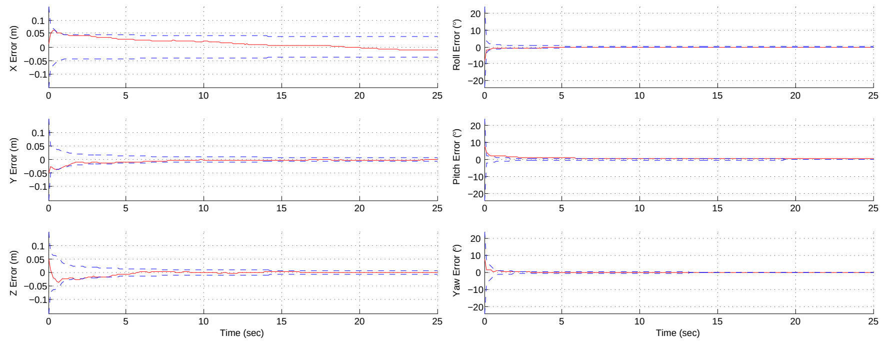


Fig. 2: State error and 3σ bounds for the camera-IMU translation (left) and rotation (right) estimates over the simulation time interval of 25 seconds. The initial position error for the camera relative to the IMU was [-5.0, -5.0, 5.0] cm in the X, Y and Z directions, and the initial orientation error was [-8.0, 8.0, 8.0] degrees in roll, pitch and yaw.

Table 1: Results from one simulation trial. The initial absolute error for the estimate of the camera-IMU transform is shown on the first line of the table. The error in the estimate after the simulation has run for 25 seconds is shown on the second line.

Error	X (cm)	Y (cm)	Z (cm)	Roll (°)	Pitch (°)	Yaw (°)
Initial	5.00	5.00	5.00	8.00	8.00	8.00
Final	1.27	0.19	0.05	0.04	0.56	0.01

5 Experiments

To verify the accuracy of our calibration algorithm with real hardware, we performed several experiments in our laboratory using a calibration test rig. The rig consists of a 60 cm long sensor beam, with an IMU mounted near one end and a camera near the other. For the experiments, we placed the beam approximately two meters in front of a wall-mounted planar camera calibration target, such that the entire target was visible in the camera image plane.

At the start of each experimental trial, we initialized the filter by holding the beam stationary while recording measurements from the IMU and images from the camera. We then manually rotated and translated the beam, while ensuring that the calibration target remained within the camera's field of view. The camera-IMU transform parameters were initialized using hand measurements of the relative position and orientation of the sensors. A subset of 100 images acquired during the visual-inertial procedure were used to calibrate the camera intrinsic parameters.

5.1 Hardware

Our camera is a black and white Flea model from Point Grey Research, with a resolution of 640×480 pixels, mated to a 4 mm Navitar lens (58° horizontal field of view). We capture images from the camera at 30 Hz. The IMU is a 3DM-G model manufactured by Microstrain, which provides angular rate and linear acceleration updates at approximately 60 Hz. The planar camera calibration target is 1.0 m $\times$ 0.8 m in size; each white or black square is 104 mm on a side. For the experiments described here, the target was mounted on a wall and the vertical edges of the calibration squares were aligned with the local gravity vector.

5.2 Results

In this section, we present results, shown in Figure 3 and Table 2, from a single experiment; we selected results which are typical of those obtained across multiple trials. Although we measured the relative pose of the sensors by hand, and therefore did not have high-quality ground truth available, we observed a decrease in the residuals for the pixel reprojection error over time, indicating that the calibration was accurate.

The duration of the experiment was approximately 90 seconds, excluding 60 seconds of setup time to initialize the IMU accelerometer and gyroscope biases in the filter. Of the 2706 images captured, the entire calibration target was visible in all except 22 frames – we simply discarded these frames.

The results from the experiment are generally in good agreement with those from the simulation. We observed that a larger residual uncertainty remains for the camera offset along the X axis relative to the IMU. This again is possibly due to the trajectory followed by the beam during the experiment – we varied the beam's position in the X direction less than in other directions. Also, solid-state MEMS-based IMUs such as the 3DM-G exhibit inherently high drift rates, which may also contribute to the poorer outcome for the X offset.

6 Conclusions and Future Work

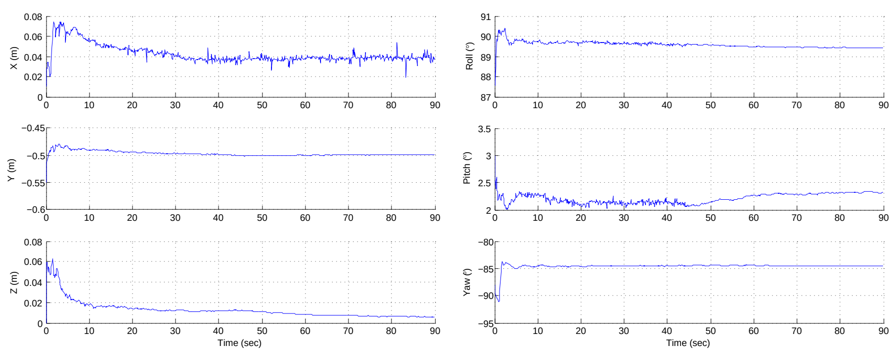
We presented an online camera-IMU relative pose calibration algorithm, which can be used to accurately determine the 6-DoF transform between the sensors. The algorithm employs an unscented Kalman filter to estimate the relative pose of the sensors and the motion of the IMU over time. This online

Table 2: Results for one experimental trial.

	$X \pm 3\sigma$ (cm)	$Y \pm 3\sigma$ (cm)	$Z \pm 3\sigma$ (cm)
Initial	2.00 ± 6.00	55.00 ± 9.00	5.00 ± 6.00
Final	4.08 ± 0.54	49.8 ± 0.30	0.59 ± 0.32
	Roll $\pm 3\sigma$ (°)	Pitch $\pm 3\sigma$ (°)	Yaw $\pm 3\sigma$ (°)
Initial	90.00 ± 6.00	3.00 ± 6.00	-88.00 ± 6.00
Final	89.42 ± 0.10	2.32 ± 0.16	-84.54 ± 0.08

approach is considerably easier than other published calibration methods, and does not require turntables or other complex apparatus. Additionally, our technique enables rapid re-calibration when the positions of the sensors must be changed (e.g. for operation in different environments).

There are several directions for future work. We are currently investigating the effect of the beam's trajectory on calibration accuracy and convergence time, in order to define trajectories which rapidly produce accurate calibration results. We are also working with stereo cameras and applying structure-from-motion algorithms to estimate the locations of landmark points as part of the calibration procedure. This will allow us to fully calibrate the relative pose of the stereo cameras and the IMU without the need for the calibration target. This work is part of an ongoing project to develop automated calibration methods for our aerial, aquatic and humanoid robot platforms.



(a) (b)

Fig. 3: (a) Estimated relative camera translation during experimental trial. (b) Estimated relative camera rotation during experimental trial.

Acknowledgments

This work was funded in part by the National Science Foundation under grants CNS-0619937, CNS-0325875, CNS-0520305 and CCF-0120778 and gifts from the Okawa Foundation and Futtock Foundation. Jonathan Kelly was supported by an Annenberg Fellowship from the University of Southern California.

References

1. D. Strelow and S. Singh, "Optimal Motion Estimation from Visual and Inertial Data," in *Proc. 6th IEEE Workshop on the Applications of Computer Vision (WACV'02)*, Orlando, USA, Dec. 2002, pp. 314–319.
2. A. I. Mourikis and S. I. Roumeliotis, "A Multi-State Constraint Kalman Filter for Vision-aided Inertial Navigation," in *Proc. IEEE Int'l Conf. Robotics and Automation (ICRA'07)*, Rome, Italy, Apr. 2007, pp. 3565–3572.
3. D. Titterton and J. Weston, *Strapdown Inertial Navigation Technology*, 2nd ed., ser. IEE Radar, Sonar and Navigation Series 17. London, United Kingdom: The Institution of Electrical Engineers, Feb. 2004, vol. 5.
4. J. Lobo and J. Dias, "Relative Pose Calibration Between Visual and Inertial Sensors," *Int'l J. Robotics Research*, vol. 26, no. 6, pp. 561–575, June 2007.
5. P. I. Corke, J. Lobo, and J. Dias, "An Introduction to Inertial and Visual Sensing," *Int'l J. Robotics Research*, vol. 26, no. 6, pp. 519–535, June 2007.
6. S. I. Roumeliotis, A. E. Johnson, and J. F. Montgomery, "Augmenting Inertial Navigation with Image-Based Motion Estimation," in *Proc. IEEE Int'l Conf. Robotics and Automation (ICRA'02)*, vol. 4, Washington D.C., USA, May 2002, pp. 4326–4333.
7. D. Strelow and S. Singh, "Online Motion Estimation from Visual and Inertial Measurements," in *Proc. 1st Workshop on Integration of Vision and Inertial Sensors (INERVIS'03)*, Coimbra, Portugal, June 2003.
8. E. M. Nebot and H. F. Durrant-Whyte, "Initial Calibration and Alignment of Low-Cost Inertial Navigation Units for Land Vehicle Applications," *J. Robotic Systems*, vol. 16, no. 2, pp. 81–92, Feb. 1999.
9. P. Lang and A. Pinz, "Calibration of Hybrid Vision / Inertial Tracking Systems," in *Proc. 2nd Workshop on Integration of Vision and Inertial Sensors (INERVIS'05)*, Barcelona, Spain, Apr. 2005.
10. F. M. Mirzaei and S. I. Roumeliotis, "A Kalman Filter-based Algorithm for IMU-Camera Calibration," in *Proc. IEEE/RSJ Int'l Conf. Intelligent Robots and Systems (IROS'07)*, San Diego, USA, Oct./Nov. 2007, pp. 2427–2434.
11. A. B. Chatfield, *Fundamentals of High Accuracy Inertial Navigation*, ser. Progress in Astronautics and Aeronautics, P. Zarchan, Ed. Cambridge, USA: American Institute of Aeronautics and Astronautics, Sept. 1997, vol. 174.
12. J. Shi and C. Tomasi, "Good Features to Track," in *Proc. IEEE Conf. Computer Vision and Pattern Recognition (CVPR'94)*, Seattle, USA, June 1994, pp. 593–600.
13. S. J. Julier and J. K. Uhlmann, "Unscented Filtering and Nonlinear Estimation," *Proceedings of the IEEE*, vol. 92, no. 3, pp. 401–422, Mar. 2004.