2014 IEEE International Conference on Robotics & Automation (ICRA)
Hong Kong Convention and Exhibition Center
May 31 - June 7, 2014. Hong Kong, China

Robust Calibration of an Ultralow-Cost Inertial Measurement Unit and a Camera: Handling of Severe System Uncertainty

Chang-Ryeol Lee, Ju Hong Yoon, and Kuk-Jin Yoon

***Abstract*— Recently, mobile devices such as smart phones and quad-copters are being equipped with inertial measurement units (IMUs) because of advances in micro-electro-mechanical systems technology. This has increased the importance of IMU–camera fusion for vision-based applications. However, ultralow-cost IMUs take much less accurate measurements than low-cost and high-cost IMUs. This uncertainty degrades the accuracy and reliability of IMU–camera calibration, which is the most important step for IMU–camera fusion technology. In this paper, we propose three effective algorithms for robust IMU–camera calibration with uncertain measurements: boundary constraint, adaptive prediction, and angular velocity constraint. These algorithms incorporate a Bayesian filtering framework to estimate calibration parameters more efficiently. The experimental results on both simulation and real data demonstrated the superiority of the proposed algorithms.**

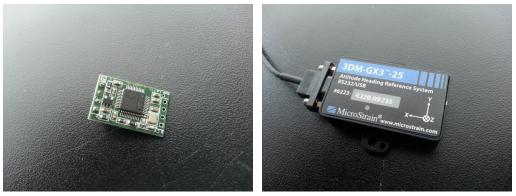
I. INTRODUCTION

Inertial measurement unit (IMU)–camera fusion has great advantages for various vision-based applications such as mobile robot localization because the two sensors can complement each other. For example, an IMU can be used to estimate the motion of a device even when the motion is fast. However, motion estimation eventually drifts after long sequences because the estimation errors from the IMU measurements accumulate. With a camera, motions can be estimated more robustly even for long sequences. However, when the device moves fast or the camera captures only a textureless scene, the motion estimation becomes unreliable. Because of the complementary properties of the IMU and camera, many researchers have tried fusing them to achieve better motion estimation.

Recently, mobile devices such as smart phones and quad-copters are being equipped with ultralow-cost IMUs and cameras, which has made IMU–camera fusion more important and common than ever. The most important step to fuse two heterogeneous sensors is calibration, where a six degrees-of-freedom (DOF) transformation (3-DOF rotation and 3-DOF translation) is estimated to represent the relation between the IMU and camera frames. Without the calibration, the sensors cannot be used together for motion estimation.

However, measurements from ultralow-cost IMUs typically contain more severe noise than those from low-cost and high-cost IMUs. Fig. 1 presents ultralow-cost and low-cost IMUs. Although the former is about four times smaller than the latter and much cheaper, it generates 10 times more noise, as described in Table I. Since the IMU measurement noise critically affects the IMU–camera calibration, the noise needs to be alleviated or avoided. We considered four types of noises, which can be categorized as internal or external noise.

C. Lee, J. Yoon, and K. Yoon are with the School of Information and Communications at the Gwang-ju Institute of Science and Technology, Gwang-ju, Korea {crlee, jhyoon, kjyoon}@gist.ac.kr



(a) Ultralow-cost IMU: price is $50. (b) Low-cost IMU: price is about $1000.

Fig. 1. Ultralow-cost IMU vs. low-cost IMU.

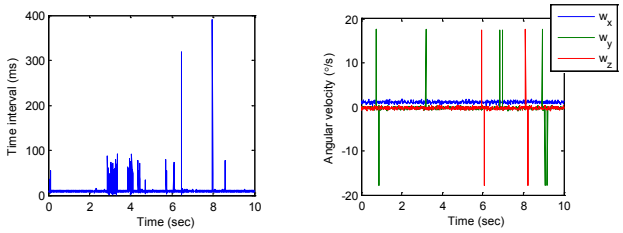
TABLE I
COMPARISON OF GAUSSIAN NOISE OF ULTRALOW-COST AND LOW-COST IMUS.

		Ultralow-cost IMU	Low-cost IMU
Accelerometer (m/s^2)			
Std. dev.	$\sigma_{\mathbf{a}_x}$	0.2550	0.0076
	$\sigma_{\mathbf{a}_y}$	0.2450	0.0080
	$\sigma_{\mathbf{a}_z}$	0.2550	0.0092
Gyroscope ($^\circ/s$)			
Std. dev.	$\sigma_{\mathbf{g}_x}$	0.1780	0.0031
	$\sigma_{\mathbf{g}_y}$	0.2590	0.0029
	$\sigma_{\mathbf{g}_z}$	0.2360	0.0032

Internal noise: There are two kinds of internal noise: natural sensor noise, which is typically assumed to be additive Gaussian noise and slowly time varying bias [1]; and errors of scale, bias, and misalignment parameters that are caused by deficient intrinsic calibration of an IMU [2].

External noise: General data acquisition problems include missing data and time-delayed measurements. With time-delayed IMU measurements, it is not possible to estimate the motion of a device accurately, and missing data because of unknown sensor failure act like outliers [3]. Both kinds of noise can be considered to be non-Gaussian, as shown in Fig. 2.

To achieve robust IMU–camera calibration that considers various types of IMU noise, we propose the following three effective algorithms. First, we use the boundary constraint (BC) to prevent the estimated states from drifting because of severe IMU noise. Second, we use adaptive prediction (AP)



(a) Abrupt delay of sampling period.

(b) Outlier in angular velocity measurement.

Fig. 2. Non-Gaussian noise in IMU measurements.

to avoid unreliable prediction because of external noise. AP exploits two motion models and selects one by considering the error magnitude of IMU measurements. Third, we impose an angular velocity constraint (AVC) on the IMU's orientation to lessen the accumulation error. These three sub-algorithms are incorporated into a Bayesian filtering framework to estimate the calibration parameters efficiently. In this study, we adopted a well-known Bayesian filter: the unscented Kalman filter (UKF) [4].

Our method estimates the rotation and translation simultaneously by using a Bayesian filter. Therefore, it can be applied to any calibration algorithm using a nonlinear state estimation filter such as the EKF, UKF, or gray-box system identification. To the best of our knowledge, this is the first work on handling the severe uncertainty of an ultralow-cost IMU for IMU–camera calibration.

II. PREVIOUS WORKS

In the preliminary stages of this study, we estimated the 6-DOF transformation between an IMU and a camera using a CAD plot or range sensor. Johnson *et al*. used a three-dimensional (3D) laser scanner for calibration purposes [5][6]. To avoid the use of high-cost equipment, Lobo and Dias proposed a two-step algorithm [7]. Their inputs were several still images and the corresponding IMU measurements. This algorithm is easy and intuitive, but the setup is tricky because the IMU needs to be at the exact center of the turntable. Moreover, errors in the rotation estimates are propagated to the translation estimates.

To overcome these drawbacks, several algorithms have been proposed that estimate rotation and translation. These take a video-based strategy and only need a checkerboard or scene with many textures. Popular methods are to exploit nonlinear state estimation techniques such as the extended Kalman filter (EKF) and UKF [8]. These approaches are a classic way to deal with system identification issues [9]. The calibration problem can be transformed to state estimation by augmenting parameters into the state.

Mirzaei and Roumeliotis were the first to present filter-based IMU–camera calibration [6]. They exploited an error-state (indirect) Kalman filter and analyzed the observability of the algorithm using Lie derivatives [10]. One drawback is information loss caused by first-order approximation of the EKF. Hol and Gustafsson adopted gray-box system identification where they combined a filter-based framework and optimization technique [11]. They minimized the innovation (prediction error) of the filter-based method. Kelly and Sukhatme proposed a general calibration algorithm that does not require additional equipment [12]. They employed the UKF to handle the nonlinearity of the model well. Furthermore, they considered gravity as a state to be estimated because it depends on geo-location. Panahandeh *et al*. exploited a mirror instead of a checkerboard to acquire measurements from a camera [13]. They also estimated the intrinsic parameters of a camera for more accurate calibration. Similarly, Zachariah and Jansson estimated the intrinsic parameters of an IMU and the transformation between an IMU and a camera simultaneously [14]. Brink and Soloviev extended this framework to an IMU and multi-camera system for an autonomous vehicle [15].

Several approaches based on optimization techniques have been used to initialize the estimator. Dong-Si and Mourikis formulated rotation calibration as a convex problem [16]. Fleps *et al*. exploited the alignment between each trajectory of an IMU and a camera as a cost function [17]. They were the first to compare various calibration algorithms using real datasets.

III. NOTATION

Before presenting the problem formulation and proposed algorithms, we briefly present some notations to make our formulation more clear. $\{I\}$ denotes the IMU frame, and $\{W\}$ represents the world frame. Special subscripts (i.e., ${}^{W}_{I}$ and ${}^{I}_{C}$) explain the parameter frame explicitly. For example, when $\mathbf{p}$ denotes the 3D position, ${}^{W}_{I}\mathbf{p}$ represents the IMU 3D position with respect to the world frame $\{W\}$; here, the subscript I represents an IMU. When $\mathbf{t}$ denotes 3D translation, ${}^{I}_{C}\mathbf{t}$ represents camera translation with respect to the IMU frame $\{I\}$; here, the subscript C represents a camera.

IV. SYSTEM DESCRIPTION

We estimate the state $\mathbf{x}$ in (1) by using the measurement $\mathbf{z}$ in (5) with the UKF because both the state transition and state measurement models are nonlinear functions.

A. State Vector

We divide a state vector $\mathbf{x}$ into three vectors:

$$\mathbf{x} = \begin{bmatrix} \mathbf{x}_{calib}^{\top} & \mathbf{x}_{imu}^{\top} & \mathbf{x}_{con}^{\top} \end{bmatrix}^{\top}, \tag{1}$$

where $\mathbf{x}_{calib}$ is a calibration state, $\mathbf{x}_{imu}$ is an IMU state, and $\mathbf{x}_{con}$ is a control input state. The calibration state comprises fundamental states for IMU–camera calibration. The IMU state consists of the IMU's intrinsic parameters. The control input state is composed of the inertial state.

The calibration state vector represents 3D poses of an IMU, the transformation between the IMU and camera, and gravity:

$$\mathbf{x}_{calib} = \begin{bmatrix} {}^{W}_{I}\mathbf{p}^{\top} & {}^{W}_{I}\mathbf{v}^{\top} & {}^{W}_{I}\mathbf{q}^{\top} & {}^{W}\mathbf{g}^{\top} & {}^{I}_{C}\mathbf{t}^{\top} & {}^{I}_{C}\mathbf{q}^{\top} \end{bmatrix}^{\top},$$

where ${}^{W}_{I}\mathbf{q} \in \Re^4$ is a unit quaternion representing the orientation, ${}^{W}_{I}\mathbf{p} \in \Re^3$ is the position, and ${}^{W}_{I}\mathbf{v} \in \Re^3$ is the velocity. ${}^{I}_{C}\mathbf{t} \in \Re^3$ denotes the translation between an IMU and a camera, ${}^{I}_{C}\mathbf{q} \in \Re^4$ is a unit quaternion representing the rotation between an IMU and a camera, and ${}^{W}\mathbf{g} \in \Re^3$ is the gravity for the world frame.

The IMU state vector includes misalignment, scale, and bias of the IMU measurements, as described in [14]. Because of the incomplete manufacturing process, the acceleration and gyroscope coordinates of the ultralow-cost IMUs are different. Hence, to estimate the difference, we include ${}^{G}_{A}\mathbf{q}^{\mathrm{T}}$ in the IMU state vector, which is a unit quaternion representing the rotation between the acceleration frame $\{A\}$ and gyroscope frame $\{G\}$. Here, $\{G\}$ is equal to the IMU frame $\{I\}$.

$$\mathbf{x}_{imu} = \begin{bmatrix} {}^{G}_{A}\mathbf{q}^{\top} & \mathbf{m}_a^{\top} & \mathbf{m}_g^{\top} & \mathbf{s}_a^{\top} & \mathbf{s}_g^{\top} & \mathbf{b}_a^{\top} & \mathbf{b}_g^{\top} \end{bmatrix}^{\top},$$

where $\mathbf{m}_a \in \Re^3$ and $\mathbf{m}_g \in \Re^6$ comprise misalignment matrices, $\mathbf{s}_a \in \Re^3$ and $\mathbf{s}_g \in \Re^3$ comprise scale matrices, and $\mathbf{b}_a \in \Re^3$ and $\mathbf{b}_g \in \Re^3$ are bias vectors.

The control input state vector comprises the acceleration and angular velocity of the IMU.

$$\mathbf{x}_{con} = \begin{bmatrix} {}^{W}\mathbf{a}^{\top} & {}^{I}\mathbf{w}^{\top} \end{bmatrix}^{\top},$$

where ${}^{W}\mathbf{a} \in \Re^3$ is the acceleration in the world frame and ${}^{I}\mathbf{w} \in \Re^3$ is the angular velocity in the IMU frame.

B. State Transition Model

The state transition model is formulated as

$$\mathbf{x}_{k+1} = \begin{bmatrix} \mathbf{x}_{calib_{k+1}} \\ \mathbf{x}_{imu_{k+1}} \\ \mathbf{x}_{con_{k+1}} \end{bmatrix} = f(\mathbf{x}_k) + \mathbf{n}_k = \begin{bmatrix} f_{calib}(\mathbf{x}_k) \\ f_{imu}(\mathbf{x}_k) \\ f_{con}(\mathbf{x}_k, \mathbf{u}_k) \end{bmatrix} + \mathbf{n}_k, \quad (2)$$

where the modeling noise $\mathbf{n}_k \in \Re^{51}$ is assumed to be the white Gaussian noise $\mathbf{n}_k \sim \mathcal{N}(0, \mathbf{Q})$ and $\mathbf{Q} \in \Re^{51\times 51}$, and $\mathbf{u}_k \in \Re^6$ is a control input obtained from the IMU.

The calibration state transition model from (2) is formulated with $\mathbf{x}_{calib}$ and $\mathbf{x}_{con}$ based on the basic law of uniformly accelerated motion [11], and ${}^{W}\mathbf{g}$, ${}^{I}_{C}\mathbf{t}$, and ${}^{I}_{C}\mathbf{q}$ are modeled to be constant.

$$\begin{aligned} \mathbf{x}_{calib_{k+1}} &= f_{calib}(\mathbf{x}_k) \\ &= \begin{bmatrix} {}^{W}_{I}\mathbf{p}_{k+1} \\ {}^{W}_{I}\mathbf{v}_{k+1} \\ {}^{W}_{I}\mathbf{q}_{k+1} \\ {}^{W}\mathbf{g}_{k+1} \\ {}^{I}_{C}\mathbf{t}_{k+1} \\ {}^{I}_{C}\mathbf{q}_{k+1} \end{bmatrix} = \begin{bmatrix} {}^{W}_{I}\mathbf{p}_k + {}^{W}_{I}\mathbf{v}_k\Delta T + {}^{W}_{I}\mathbf{a}_k\frac{\Delta T^2}{2} \\ {}^{W}_{I}\mathbf{v}_k + {}^{W}_{I}\mathbf{a}_k\Delta T \\ e^{-{}^{W}_{I}\mathbf{w}_k\frac{\Delta T}{2}} \odot {}^{W}_{I}\mathbf{q}_k \\ {}^{W}\mathbf{g}_k \\ {}^{I}_{C}\mathbf{t}_k \\ {}^{I}_{C}\mathbf{q}_k \end{bmatrix}, \end{aligned} \quad (3)$$

where the operator $\odot$ denotes the quaternion product.

In (2), the IMU state was modeled as Brownian motion $\mathbf{x}_{imu_{k+1}} = f_{imu}(\mathbf{x}_k) = \mathbf{x}_{imu_k}$ because the misalignment and scale terms do not have a dynamic motion model, and the bias term of the IMU measurements varies slowly with time.

The control input state transition model is formulated with $\mathbf{x}_{calib}$, $\mathbf{x}_{imu}$, and $\mathbf{u}_k$. We utilized two transition models: an IMU-based motion model $f_{con,1}(\mathbf{x}_k, \mathbf{u}_k)$ and constant motion model $f_{con,2}(\mathbf{x}_k)$.

<u>*IMU-based motion model*</u>: The control input state vector is propagated with measurements from the IMU through the inverse of the process [18]:

$$\begin{aligned} \mathbf{x}_{con_{k+1}} &= f_{con,1}(\mathbf{x}_k, \mathbf{u}_k) = \begin{bmatrix} {}^{W}\mathbf{a}_{k+1} \\ {}^{I}\mathbf{w}_{k+1} \end{bmatrix} \\ &= \begin{bmatrix} \mathbb{R}({}^{W}_{I}\mathbf{q}_k)\left[\mathbf{S}_a^{-1}\mathbf{M}_a^{-1}[\mathbf{a}_{m,k} - \mathbf{b}_{a,k} - \mathbf{n}_a]\right] + {}^{W}\mathbf{g}_k \\ \mathbf{S}_w^{-1}\mathbf{M}_w^{-1}[\mathbf{w}_{m,k} - \mathbf{b}_{w,k} - \mathbf{n}_w] \end{bmatrix}, \end{aligned} \quad (4)$$

where $\mathbf{n}_w$ and $\mathbf{n}_a$ are the angular and acceleration noise, respectively, and assumed to be white Gaussian noise, $\{\mathbf{a}_{m,k}, \mathbf{w}_{m,k}\} \subset \mathbf{u}_k$ are the measurements from the IMU, $\mathbf{M}_a$ and $\mathbf{M}_w$ are misalignment matrices composed of states $\mathbf{m}_a$ and $\mathbf{m}_w$, and $\mathbf{S}_a$ and $\mathbf{S}_w$ are scale matrices from states $\mathbf{s}_a$ and $\mathbf{s}_w$ [14], the matrix $\mathbb{R}(\cdot)$ denotes a direct cosine matrix converted from the unit-quaternion $\mathbf{q}$.

<u>*Constant motion model*</u>: In visual SLAM, constant velocity and constant angular velocity models that assume that the velocity and angular velocity have Gaussian profiles are commonly used because of the smooth motion [19]. We used the constant acceleration and constant angular velocity model $\mathbf{x}_{con_{k+1}} = f_{con,2}(\mathbf{x}_k) = \mathbf{x}_{con_k}$ because the IMU–camera setup also has to move smoothly because of the constraint of the camera's frame rate. The constant acceleration model is more sophisticated because it is a second-order approximation.

C. State Measurement Model

We used two-dimensional (2D) image feature points as measurements. When we utilize M 2D feature points, the measurement consists of a stacked vector of M 2D features:

$$\mathbf{z} = \begin{bmatrix} \mathbf{z}_1{}^{\top} \ldots \mathbf{z}_M{}^{\top} \end{bmatrix}^{\top}, \quad (5)$$

where each 2D feature point is extracted from an image [6]. Therefore, the measurement model describes the transformation of 3D feature points (in the world frame) to 2D feature points. In the model, 3D feature points are first transformed to the IMU frame and then transformed to the camera frame. Therefore, the measurement is formulated as a nonlinear function.

$$\begin{aligned} \mathbf{z}_i = \begin{bmatrix} u_i \\ v_i \end{bmatrix} &= h(\mathbf{x}, \mathbf{f}_i^W) + \mathbf{m}_i = \begin{bmatrix} x_i/z_i \\ y_i/z_i \end{bmatrix} + \mathbf{m}_i, \\ \begin{bmatrix} x_i \\ y_i \\ z_i \end{bmatrix} &= \mathbf{K}\mathbb{R}({}^{I}_{C}\mathbf{q})\left[\ \mathbb{R}({}^{W}_{I}\mathbf{q})\left[\mathbf{f}_i^W - {}^{W}_{I}\mathbf{p}\right] - {}^{I}_{C}\mathbf{p}\ \right], \end{aligned} \quad (6)$$

where $\mathbf{f}_i^W$ is the i-th 3D feature point in the world frame, $\mathbf{m}_i$ is white Gaussian noise with the noise covariance matrix $\mathbf{R}_i = (\sigma_{R_i})^2\mathbf{I}_2$, and $\mathbf{K}$ is a camera-intrinsic matrix. Since the measurement is a stacked vector of M 2D feature points, its noise covariance is expressed as $\mathrm{diag}(\mathbf{R}_1, \ldots, \mathbf{R}_M)$.

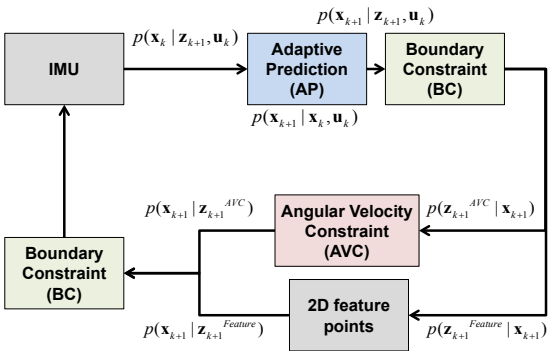
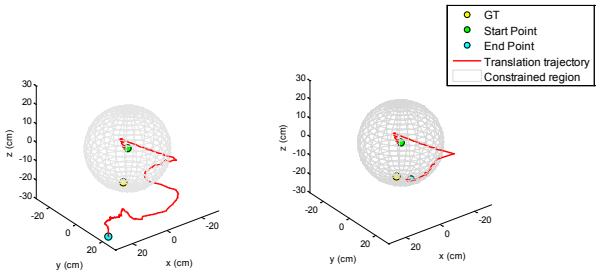


Fig. 3. Block diagram of framework.



(a) Standard calibration (b) BC calibration

Fig. 4. Effect of boundary constraints on calibration under severe uncertainty of IMU.

V. PROPOSED ALGORITHMS

To achieve robust calibration, we incorporate three sub-algorithms into the Bayesian filtering framework as described in Fig. 3: boundary constraint (BC), adaptive prediction (AP), and angular velocity constraint (AVC).

A. Boundary Constraint (BC)

To prevent states from drifting, we impose an inequality constraint on states. The prior knowledge on the transformation between the IMU and camera is expressed as an inequality constraint.

The translation ${}^I_C\mathbf{t}$ cannot exceed a certain boundary since the IMU and camera are rigidly fastened to each other. This inequality is expressed as

$$\left\| {}^I_C\mathbf{t} - {}^I_C\mathbf{t}_0 \right\|_2 \leq \tau_1. \tag{7}$$

Equation (7) expresses a constraint region defined as a sphere from the initial translation ${}^I_C\mathbf{t}_0$. The rotation ${}^I_C\mathbf{q}$ has a true value near the initial rotation $\mathbf{q}_0$ because we use the simple Lobo's algorithm [7], which estimates the rotation between an IMU and a camera. The inequality is given as

$$\mathrm{dist}({}^I_C\mathbf{q}, \mathbf{q}_0) \leq \tau_2. \tag{8}$$

The distance between two quaternions is defined by the geodesic distance in rotation space [20].

These inequality constraints are placed on the constrained region by the projection scheme [21]. Technically, the states outside the constrained region are moved to the boundary of the constrained region. We carry out the projection by interpolating the initial states and estimates. The translation and rotation between the IMU and camera are defined in Cartesian coordinate and four-dimensional rotation space, respectively. Therefore, linear and spherical interpolations are required. The projection function $proj({}^I_C\mathbf{t}, {}^I_C\mathbf{q})$ is defined as follows with (9)-(10).

$${}^I_C\tilde{\mathbf{t}} \triangleq \begin{cases} {}^I_C\mathbf{t} & \text{if } \left\| {}^I_C\mathbf{t} - {}^I_C\mathbf{t}_0 \right\|_2 \leq \tau_1 \\ (1-\tau_1){}^I_C\mathbf{t}_0 + \tau_1 {}^I_C\mathbf{t} & \text{otherwise} \end{cases}, \tag{9}$$

$${}^I_C\tilde{\mathbf{q}} \triangleq \begin{cases} {}^I_C\mathbf{q} & \text{if } \mathrm{dist}({}^I_C\mathbf{q}, {}^I_C\mathbf{q}_0) \leq \tau_2 \\ \left(\frac{\sin(1-\tau_2)\alpha}{\sin\alpha}\right){}^I_C\mathbf{q}_0 + \left(\frac{\sin(\tau_2)\alpha}{\sin\alpha}\right){}^I_C\mathbf{q} & \text{otherwise} \end{cases}, \tag{10}$$

where $\alpha = \cos^{-1}({}^I_C\mathbf{q}_0 \cdot {}^I_C\mathbf{q}_k)$. The inequality constraint is applied to states after sigma point generation of the prediction and update [21]. As a consequence, since we use the UKF, the state covariance is determined within the filtering process. In particular, this constraint is more important during prediction because the IMU uncertainty has a large effect on the system model.

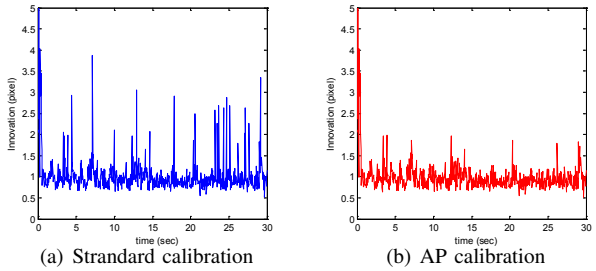
The BC restricts the solution space of the estimates. This effect can be visually displayed because of its geometric meaning. The translation ${}^I_C\mathbf{t}$ can be represented as a point in 3D space. Then, its BC can be expressed as a sphere. Fig. 4 shows how the BC works for translation estimation. The estimated translation is indicated as a trajectory in the translation space. Unfortunately, we cannot visualize the effect on rotation estimation because the rotation ${}^I_C\mathbf{q}$ is defined as a quaternion. However, more accurate calibration results would indicate a similar effect on the rotation estimation.

B. Adaptive Prediction (AP)

As noted in Section I, the IMU measurement sometimes contains severe external noise. To avoid wrong predictions because of outliers, we select one of two transition models using IMU-based motion or constant motion. The criterion for selection is the innovation ε, which the system uncertainty largely affects [22]. Therefore, it can be used to judge whether IMU measurements are reliable or not.

$$\varepsilon = \mathbf{z}_{k+1} - \hat{\mathbf{z}}_{k+1}, \tag{11}$$

$$\hat{\mathbf{z}}_{k+1} = h\begin{pmatrix} f_{calib}(\mathbf{x}_k) \\ f_{imu}(\mathbf{x}_k) \\ f_{con,1}(\mathbf{x}_k, \mathbf{u}_k) \end{pmatrix}, \mathbf{f}^W \Bigg). \tag{12}$$



(a) Standard calibration (b) AP calibration

Fig. 5. Comparison of innovations for same dataset with and without AP.

The threshold for selection was determined to be three times the measurement noise because innovation over the threshold means that the IMU measurements have extremely severe noise.

$$f_{con}(\mathbf{x},\mathbf{u}) \triangleq \begin{cases} f_{con,1}(\mathbf{x},\mathbf{u}) & \text{if } \ \varepsilon \leq 3\sigma_R \\ f_{con,2}(\mathbf{x}) & \text{otherwise.} \end{cases} \tag{13}$$

Fig. 5 shows that the adaptive prediction provides relatively consistent innovation. In contrast, standard calibration has a number of abrupt changes in innovation.

C. Angular Velocity Constraints (AVC)

Although the AP and BC handle uncertainty well based on the geometric distance relation, they are limited in handling accumulation of error by severe noise. To moderate the error, we add a constraint on the IMU's orientation. As pointed out in [1], the orientation estimates of the IMU have a critical impact on trajectory estimation of the IMU. For this reason, the noise of the angular velocity has to be handled carefully. The angular velocity from the camera has much less noise; this attenuates the noise of the IMU measurements.

As a constraint, we exploit the fact that the angular velocities of the camera and IMU have identical magnitudes. This is suitable for the IMU–camera calibration problem because it does not require transformation between the two units. The angular velocity of the camera is computed by the homography-based pose estimation scheme [23].

$$\mathbf{z}^{AVC} = \left\| {}^{C}\mathbf{w} \right\|_2 . \tag{14}$$

The predicted angular velocity $\hat{\mathbf{z}}^{AVC}$ of the IMU is determined by the L-2 norm of the angular velocity state ${}^{I}\mathbf{w} \in \mathbf{x}_{con}$:

$$\hat{\mathbf{z}}^{AVC} = h^{AVC}({}^{I}\mathbf{w}) + \boldsymbol{\varepsilon}^{\mathbf{AVC}} = \left\| {}^{I}\mathbf{w} \right\|_2 + \boldsymbol{\varepsilon}^{\mathbf{AVC}}, \tag{15}$$

$$\boldsymbol{\varepsilon}^{\mathbf{AVC}} = \left[\ \mathbf{z}^{AVC} - h^{AVC}(\mathbf{x}_{con}) \ \right]. \tag{16}$$

The innovation vector $\boldsymbol{\varepsilon}^{\mathbf{AVC}}$ describes the condition that must be maintained between the visual and inertial observations. Consequently, IMU–camera calibration can avoid divergence because of noisy angular velocities. The AVC is incorporated into the measurement update to augment the measurement model [24].

Fig. 6 shows that the angular velocity constraint increases the accuracy of the estimated IMU pose. This is a representative indicator that the AVC helps IMU–camera calibration.

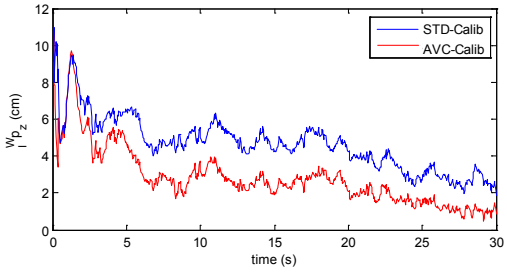


Fig. 6. Comparison of error in position (z-axis) estimates for the same dataset with and without AVC.

VI. EXPERIMENTS

We evaluated the performance of the proposed algorithms using both simulated and real data. For the calibration problem, the convergence of states is not a good evaluation tool. Therefore, the results of the simulated data, which used ground truth, were more valuable than those of the real data for analysis of the advantages and disadvantages of the proposed algorithms.

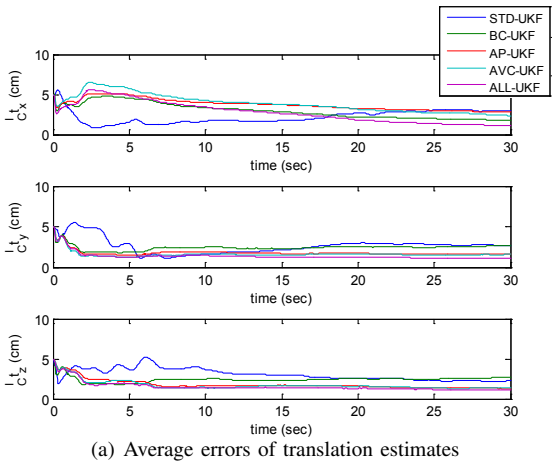
A. Simulation

For the experiments, we generated a smooth IMU–camera motion by combining sinusoidal functions. The three-axis acceleration and angular velocity of the IMU were computed by differentiating the IMU trajectory. The 2D feature points of an image sequence were computed by projecting 3D feature points in the frame $\{W\}$ with the given intrinsic parameters. The camera followed the pinhole model with 320 $\times$ 240 pixel image resolution and a field of view of about 50°. The IMU and camera measurements were received at 100 and 25 Hz with synchronization. The rotation and translation between the IMU and camera were set to [50 -50 50]° and [5 5 -5]$^{\mathrm{T}}cm$, respectively. For estimation, the initial translation and rotation was set to [0 0 0]$^{\mathrm{T}}cm$ and to be rotated from the ground truth up to [-5 5 5]°$^{\mathrm{T}}$, respectively. Zero-initialization was reasonable because guessing the initial IMU–camera translation is difficult.

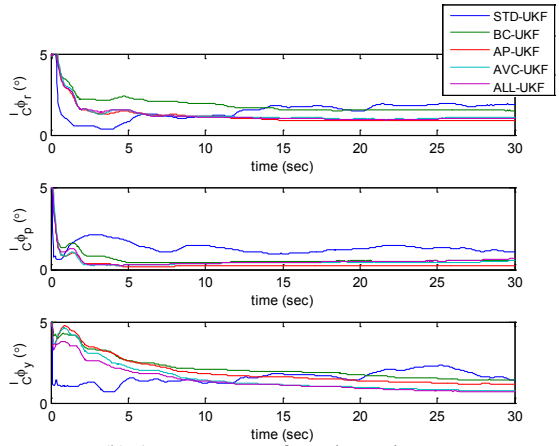
Intrinsic parameters such as the scale, misalignment, and bias were uniformly determined within limit ranges ($\mathbf{s}_{a,w}$ = [1.05 - 1.1], $\mathbf{m}_{a,g}$ = [0.05 - 0.1], $\mathbf{b}_a$ = [0.1 - 0.3] m/s^2, $\mathbf{b}_g$ = [0.3 - 0.5] $^\circ/s$). Furthermore, we added the Gaussian noise (σ_{n_a} = 0.25m/s^2, σ_{n_g} = 0.26°$/s$) and non-Gaussian noise (missing data at 1%) into the IMU measurement. The noise strength was determined from real data by the algorithm described in [1]. The 2D feature points had a measurement error, which was white Gaussian noise with a low standard deviation of 1 pixel, because they were extracted from a checkerboard.

We intensively drew a comparison between the three proposed sub-algorithms and standard UKF-based calibration [12] because the UKF is a popular and good estimator for model nonlinearity. Fig. 7 plots the average error of the estimated rotation and translation between an IMU and a camera over 100 Monte Carlo simulations. The quaternion ${}^{I}_{C}\mathbf{q}$, which expresses the rotation between an IMU and a camera, was converted to the corresponding Euler angle ${}^{I}_{C}\Phi$ for more intuitive understanding. The standard calibration produced an unstable estimation, whereas the proposed algorithms produced a stable estimation. In particular, the average errors of the estimates were lowest when all of the sub-algorithms were combined together. This means that each sub-algorithm affected individual factors for the estimation of the IMU uncertainty with no redundancy.

Table II lists the average RMSE of the final estimates for calibration. The calibrations with each sub-algorithm (BC, AP, AVC) provided more accurate estimates for both translation and rotation than standard calibration. In particular, the accuracy of the translation estimates showed



(a) Average errors of translation estimates



(b) Average errors of rotation estimates

Fig. 7. Time history of calibration estimates with each sub-algorithm. 1) STD-UKF: standard UKF-based calibration, 2) BC-UKF: calibration with boundary constraint, 3) AP-UKF: calibration with adaptive prediction, 4) AVC-UKF: calibration with angular velocity constraint, 5) ALL-UKF: calibration with all sub-algorithms.

Fig. 8. IMU–camera experimental setup (PointGrey Grasshoper camera with Navitar wide angle lens and E2box IMU).

marked improvement. Interestingly, when all sub-algorithms were combined, the estimator usually delivered an estimate near the most accurate calibration of the individual sub-algorithms.

B. Real Data

The experimental setup consisted of a tightly connected FireWire camera and ultralow-cost IMU (Fig. 8). We used a camera with a resolution of 320 × 240 pixels with a frame rate of 30 Hz. The camera (PointGrey Grasshopper) was mated to a 3.5 mm Navitar lens (94°×73° field of view). The mounted ultralow-cost IMU provided three-axis acceleration and angular velocity measurements at a rate of 100 Hz. The IMU and camera were synchronized by a B-spline interpolation scheme for the measurements as the time stamp [17]. The IMU–camera moved smoothly around 1 m over the calibration pattern while recording images of the pattern over 40 s. The checkerboard pattern was drawn with a 5 × 6 pattern, and the grid size was 3 cm. It was placed horizontally to the ground. The intrinsic parameters of the camera and radial distortion of the lens were computed using Bouguet's camera calibration toolbox [25]. The angular velocities for the AVC were calculated by differentiating the rotational trajectory of the input sequence obtained from this toolbox. The initial rotation between the IMU and camera was determined by a modified version of [7]; the initial translation was set to $[\ 0\ 0\ 0\ ]^{\mathrm{T}}$.

TABLE II

ERROR STATISTICS OF FINAL ESTIMATES FOR TRANSLATION AND ROTATION BETWEEN THE IMU AND CAMERA.

Average RMSE	Translation (cm)			
	${}^{I}_{C}\mathbf{t}_x$	${}^{I}_{C}\mathbf{t}_y$	${}^{I}_{C}\mathbf{t}_z$	${}^{I}_{C}\mathbf{t}_{avg.}$
STD-UKF	3.29	2.31	3.58	3.04
BC-UKF	1.87	2.96	1.51	2.11
AP-UKF	2.90	1.77	1.52	2.06
AVC-UKF	2.65	1.90	1.68	2.07
ALL-UKF	1.37	1.38	1.44	1.39
Initial error	5	5	5	5

Average RMSE	Rotation (°)			
	${}^{I}_{C}\boldsymbol{\phi}_r$	${}^{I}_{C}\boldsymbol{\phi}_p$	${}^{I}_{C}\boldsymbol{\phi}_y$	${}^{I}_{C}\boldsymbol{\phi}_{avg.}$
STD-UKF	2.10	1.21	1.55	1.62
BC-UKF	2.02	0.73	1.71	1.48
AP-UKF	0.94	0.29	1.22	0.81
AVC-UKF	1.18	0.64	0.90	0.90
ALL-UKF	1.12	0.76	0.86	0.91
Initial error	5	5	5	5

In the real data experiments, the measurement performance of the IMU–camera calibration algorithms was unclear because there was no ground truth. Instead, we used these experiments to judge the reliability and consistency, which were expressed as the variance of the estimated transformation of the IMU–camera.

We tested the proposed algorithms against the standard calibration based on the UKF [12]. Fig. 9 illustrates the calibration results with each sub-algorithm using box and whisker diagrams. Similar to the simulation results, calibration with the proposed algorithms outperformed standard calibration. When all sub-algorithms were combined, the estimator produced the most consistent results. The standard calibration produced good rotation estimates but poor translation estimates. With our framework, the accuracy of the translation estimates was notably improved.

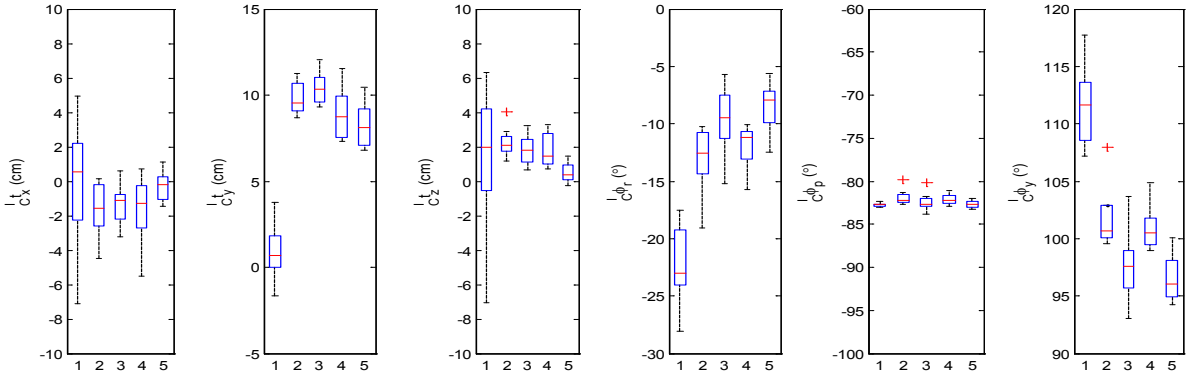


Fig. 9. Comparison of calibration results with each sub-algorithm (1: STD-UKF, 2: BC-UKF, 3: AP-UKF, 4: AVC-UKF, 5: ALL-UKF).

VII. CONCLUSIONS

In this paper, we presented new algorithms to calibrate an IMU and camera to handle severe system uncertainty from the IMU. This represents one of the first attempts to accurately calibrate an ultralow-cost IMU and a camera. Each sub-algorithm is designed to handle internal and external factors that cause the noise for an IMU. The proposed calibration algorithms provided more reliable performance for the fusion of ultralow-cost inertial and visual sensors. Furthermore, they can be used to deal with system uncertainty in inertial-visual navigation systems. The simulation and experimental results demonstrated the accuracy and potential industrial value of the proposed algorithms.

ACKNOWLEDGMENT

This work was supported by the Global Frontier R&D Program on <Human-centered Interaction for Coexistence> funded by the National Research Foundation of Korea grant funded by the Korean Government(MSIP) (2012M3A6A3055690).

This research was supported by Basic Science Research Program through the National Research Foundation of Korea(NRF) funded by the Ministry of Science, ICT & Future Planning (2012R1A1A1010871).

REFERENCES

[1] O. J. Woodman, "An introduction to inertial navigation," University of Cambridge, Computer Laboratory, Tech. Rep., 2007.

[2] G. Panahandeh, I. Skog, and M. Jansson, "Calibration of the accelerometer triad of an inertial measurement unit, maximum likelihood estimation and cramer-rao bound," in *International Conference on Indoor Positioning and Indoor Navigation*, 2010.

[3] J. Mattingely and S. Boyd, "Real-time convex optimization in signal processing," *IEEE Signal Processing Magazine*, vol. 27, no. 3, pp. 50–61, 2010.

[4] R. van der Merwe, "Sigma-point kalman filters for probabilistic inference in dynamic state-space models," Ph.D. dissertation, Oregon Health & Science University, 2004.

[5] A. Johnson, R. Willson, J. Goguen, J. Alex, and D. Meller, "Field testing of the mars exploration rovers descent image motion estimation system," 2005.

[6] F. M. Mirzaei and S. I. Roumeliotis, "A kalman filter-based algorithm for imu-camera calibration: Observability analysis and performance evaluation," *IEEE Transactions on Robotics and Automation*, vol. 24, no. 5, pp. 1143–1156, 2008.

[7] J. Lobo and J. Dias, "Relative pose calibration between visual and inertial sensors." *International Journal of Robotics Research*, vol. 26, no. 6, pp. 561–575, 2007.

[8] D. Simon, *Optimal State Estimation: Kalman, H Infinity, and Nonlinear Approaches*. Wiley-Interscience, 2006.

[9] L. Ljung and T. Soderstrom, *Theory and Practice of Recursive Identification*. MIT Press, 1987.

[10] R. Hermann and A. J. Krener, "Nonlinear controllability and observability," *IEEE Transactions On Automatic Control*, vol. 22, no. 5, pp. 728–740, 1977.

[11] J. D. Hol, T. B. Schön, and F. Gustafsson, "Modeling and calibration of inertial and vision sensors," *International Journal of Robotics Research*, vol. 29, no. 2, pp. 231–244, 2010.

[12] J. Kelly and G. S. Sukhatme, "Visual-inertial sensor fusion: Localization, mapping and sensor-to-sensor self-calibration," *International Journal of Robotics Research*, vol. 30, no. 1, pp. 56–79, 2011.

[13] G. Panahandeh, D. Zachariah, and M. Jansson, "Mirror based imu-camera and internal camera calibration," in *International Conference on Robot,Vision and Signal Processing*, 2011.

[14] D. Zachariah and M. Jansson, "Joint calibration of an inertial measurement unit and coordinate transformation parameters using a monocular camera," in *International Conference on Indoor Positioning and Indoor Navigation*, 2010.

[15] K. B. A. Soloviev, "Filter-based calibration for an imu and multi-camera system," in *IEEE/ION Position, Location and Navigation Symposium*, 2012.

[16] T. Dong-Si and A. I. Mourikis, "Initialization in vision-aided inertial navigation with unknown camera-imu calibration," in *IEEE/RSJ International Conference on Intelligent Robots and Systems*, 2012.

[17] M. Fleps, E. Mair, O. Ruepp, M. Suppa, and D. Burschka, "Optimization based imu camera calibration." in *IEEE/RSJ International Conference on Intelligent Robots and Systems*, 2011.

[18] A. B. Chatfield, *Fundamentals of high accuracy inertial navigation*. Reston, VA. American Institute of Aeronautics and Astronautics, Inc., 1997.

[19] A. J. Davison, "Real-time simultaneous localisation and mapping with a single camera," in *IEEE International Conference on Computer Vision*, 2003.

[20] R. I. Hartley and F. Kahl, "Global optimization through rotation space search," *International Journal of Computer Vision*, vol. 82, no. 1, pp. 64–79, 2009.

[21] R. Kandepu, L. Imsland, and B. A. Foss, "Constrained state estimation using the unscented kalman filter," in *Mediterranean Conference on Control and Automation*, 2008.

[22] P. Ruckdeschel, "Optimally robust kalman filtering," Fraunhofer ITWM, Tech. Rep., 2010.

[23] Z. Zhang, "Flexible camera calibration by viewing a plane from unknown orientations," in *IEEE International Conference on Computer Vision*, 1999.

[24] D. Simon, "Kalman filtering with state constraints: a survey of linear and nonlinear algorithms," *IET Control Theory & Applications*, vol. 4, no. 8, pp. 1303–1318, 2010.

[25] J. Y. Bouguet. (2008) Camera calibration toolbox for matlab. [Online]. Available: http://www.vision.caltech.edu/bouguetj/calib_doc/