Joint calibration of an inertial measurement unit and coordinate transformation parameters using a monocular camera

Dave Zachariah and Magnus Jansson
ACCESS Linneaus Centre, KTH Royal Institute of Technology
Stockholm, Sweden. Email: {dave.zachariah, magnus.jansson}@ee.kth.se

Abstract—An estimation procedure for calibration of a low-cost inertial measurement unit (IMU), using a rigidly mounted monocular camera, is presented. The parameters of a sensor model that captures misalignments, scale and offset errors are estimated jointly with the IMU-camera coordinate transformation parameters using a recursive Sigma-Point Kalman Filter. The method requires only a simple visual calibration pattern. A simulation study indicates the filter's ability to reach subcentimeter and subdegree accuracy.

I. Introduction

Progressive developments in micro-electro-mechanical system (MEMS) technology have lowered the cost of accelerometers and gyroscopes and enabled the use of inertial information in various applications from navigation to augmented reality. The performance of such systems are dependent on the accuracy of the calibration of the accelerometers and gyroscopes in the IMU. For low-cost applications, it is of interest to find simple, repeatable calibration procedures that do not require precisely controlled mechanical platforms that excite the IMU outputs.

In [1], a triad of accelerometers are calibrated by observing the magnitude of the output in static positions and then numerically solving a nonlinear Least-Squares problem. The gyroscope calibration parameters, however, are not observable. A different approach is presented in [2] based on optically tracking LEDs attached to the IMU and allows for estimating its position and orientation. The parameters are solved using a Gauss-Newton search.

IMUs are used in conjunction with cameras in vision-aided inertial navigation, robotics and augmented reality, cf. [3]–[7]. In this context it is also important to estimate the coordinate transformation parameters between the inertial and camera frames. A procedure using turntables was given in [8] but [9], [10] use a considerably simpler setup with a monocular camera, which is an inexpensive and passive sensor, along with a visual pattern typically used to calibrate the internal camera parameters in many computer vision applications [11]. An example is shown in Figure 1.

Here we use this information for calibrating the IMU itself, along with the coordinate transformation parameters, exploiting the fact that the orientation of the planar calibration pattern



Fig. 1. Calibration pattern, 28×28 cm.

relative to the gravitational field is known to a high degree of accuracy, e.g. by placing it on a horizontal plane perpendicular to the gravity vector, using a digital spirit level. We present a sequential linear filter, based on the Sigma-Point approach, that jointly estimates the IMU sensor model parameters and refines initial estimates of the coordinate transformation parameters between the inertial and camera frames.

II. SENSOR MODEL

Sensors that measure inertial information contain errors arising from nonlinearities, such as saturation, manufacturing imperfections, and other unmodeled random noise. Figure 2 illustrates a model of a sensor that measures the sought physical quantity \mathbf{u}^b , such as specific force $\mathbf{f}^b \in \mathbb{R}^3$ or angular velocity $\boldsymbol{\omega}^b \in \mathbb{R}^3$, in an ideal inertial frame $\{b\}$. For navigation in three-dimensional Euclidean space using Newtonian mechanics, the sensitivity axes of the sensors should ideally be orthogonal and with equally scaled outputs but manufacturing imperfections cause them to be non-orthogonal with scale errors, modeled by the matrices \mathbf{T} and \mathbf{K} , respectively. Offset error or bias $\delta \mathbf{u}^b$ is modeled as a slow-varying term and remaining errors, such as quantization noise, as an additive random white noise process \mathbf{n} with mutually uncorrelated components.

Using the sensor model above, the accelerometer output is written as

$$\tilde{\mathbf{f}}^b = \mathbf{A}\mathbf{f}^b + \delta\mathbf{f}^b + \mathbf{n_f},\tag{1}$$

978-1-4244-5864-6/10\$26.00 © IEEE

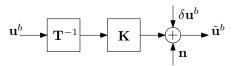


Fig. 2. Model of sensor measuring physical quantity \mathbf{u} in inertial frame $\{b\}$.

where $\mathbf{A} \triangleq \mathbf{K}_a (\mathbf{T}_a^b)^{-1} \in \mathbb{R}^{3\times 3}$. By defining the ideal inertial coordinate frame $\{b\}$ to be aligned to the sensor's x-axis and the plane spanned by the xy-axes, using a small angle assumption of the misalignment, we can write

$$\mathbf{A}^{-1} = \mathbf{T}_{a}^{b} \mathbf{K}_{a}^{-1}$$

$$= \begin{pmatrix} 1 & -\alpha_{yz} & \alpha_{zy} \\ 0 & 1 & -\alpha_{zx} \\ 0 & 0 & 1 \end{pmatrix} \operatorname{diag}\left(\frac{1}{k_{x}^{a}}, \frac{1}{k_{y}^{a}}, \frac{1}{k_{z}^{a}}\right) \quad (2)$$

$$= f_{acc}(\boldsymbol{\alpha}, \mathbf{k}^{a}).$$

The gyroscope output is modeled in a similar fashion,

$$\tilde{\boldsymbol{\omega}}^b = \mathbf{G}\boldsymbol{\omega}^b + \delta\boldsymbol{\omega}^b + \mathbf{n}_{\boldsymbol{\omega}},\tag{3}$$

where $\mathbf{G} \triangleq \mathbf{K}_g (\mathbf{T}_g^b)^{-1} \in \mathbb{R}^{3\times 3}$. Its three sensitivity axes are related by six small angle rotations to the ideal inertial frame $\{b\}$, which is determined by the accelerometer cluster's xy-axes as described above, and allows us to write,

$$\begin{aligned} \mathbf{G}^{-1} &= \mathbf{T}_{g}^{b} \mathbf{K}_{g}^{-1} \\ &= \begin{pmatrix} 1 & -\gamma_{yz} & \gamma_{zy} \\ \gamma_{xz} & 1 & -\gamma_{zx} \\ -\gamma_{xy} & \gamma_{yx} & 1 \end{pmatrix} \operatorname{diag} \left(\frac{1}{k_{\phi}^{g}}, \frac{1}{k_{\theta}^{g}}, \frac{1}{k_{\psi}^{g}} \right) \quad (4) \\ &= f_{\text{gyro}}(\boldsymbol{\gamma}, \mathbf{k}^{g}). \end{aligned}$$

Ideally, all scale factors in $(\mathbf{k}^a, \mathbf{k}^g)$ equal 1 and misalignments in (α, γ) equal 0.

III. PROCESS MODEL

The unknown sensor parameters cause the position, velocity and attitude computed using the inertial navigation system (INS) equations to deviate from their ideal values. The deviations are estimated using a mapping between a visual calibration pattern and the image plane of the rigidly mounted camera and are correlated with the sensor parameters. The framework used here is based on a feedback approach that increases the tractability of the estimation problem by simplifying the process model.

A. Navigation error states

The INS equations provide estimates in a navigation frame $\{n\}$, centered on the calibration pattern. The deviations or error states $(\delta \mathbf{p}^n, \delta \mathbf{v}^n, \delta \boldsymbol{\theta})$ are caused by the deterministic and stochastic sensor errors, modeled in the previous section, and evolve according to

$$\delta \dot{\mathbf{p}}^n = \delta \mathbf{v}^n \tag{5}$$

and

$$\delta \dot{\mathbf{v}}^{n} = \hat{\mathbf{R}}_{b}^{n} (\mathbf{I}_{3} - \mathbf{A}^{-1}) \tilde{\mathbf{f}}^{b} + [\hat{\mathbf{R}}_{b}^{n} \mathbf{A}^{-1} \tilde{\mathbf{f}}^{b}]_{\times} \delta \boldsymbol{\theta} + \hat{\mathbf{R}}_{b}^{n} \mathbf{A}^{-1} (\delta \mathbf{f}^{b} + \mathbf{n}_{\mathbf{f}}),$$
(6)

when neglecting higher-order terms. $[\cdot]_{\times}:\mathbb{R}^3\to\mathbb{R}^{3\times3}$ denotes the skew-symmetric matrix representation of the cross-product operation. Similarly the attitude error dynamics are described by

$$\delta \dot{\boldsymbol{\theta}} = \hat{\mathbf{R}}_b^n (\mathbf{G}^{-1} - \mathbf{I}_3) \tilde{\boldsymbol{\omega}}^b - \hat{\mathbf{R}}_b^n \mathbf{G}^{-1} (\delta \boldsymbol{\omega}^b + \mathbf{n}_{\boldsymbol{\omega}}). \tag{7}$$

B. IMU-to-camera transformation parameters

The affine transformation between IMU and camera coordinate frames, denoted $\{b\}$ and $\{c\}$, is parameterized by an offset vector $\mathbf{p}_c^b \in \mathbb{R}^3$ and rotation matrix $\mathbf{R}_b^c \in \mathbb{R}^{3 \times 3}$. These are however taken as estimates since the attachment of the camera to the IMU may not occur under precisely controlled circumstances. Moreover, in certain scenarios the camera will be removed after calibration of the IMU and hence the method must be repeatable without having to assume highly accurate estimates.

While an initial offset estimate $\hat{\mathbf{p}}_c^b$ can be given relatively easily using rudimentary measurement equipment, providing an initial rotation estimate $\hat{\mathbf{R}}_b^c$ can be somewhat harder. However, one can exploit static measurements of the gravitational field which is known in $\{n\}$. The images of the planar calibration pattern provide estimates of \mathbf{R}_n^c [11] which gives an estimate of the gravity vector in $\{c\}$, $\hat{\mathbf{g}}^c = \hat{\mathbf{R}}_n^c \mathbf{g}^n$. Since static accelerometers measure \mathbf{g}^b , we have $\hat{\mathbf{g}}^c \approx \mathbf{R}_b^c \hat{\mathbf{g}}^b$ which lends the estimation of \mathbf{R}_b^c to a Least Squares-problem, by minimizing

$$C(\mathbf{R}) \triangleq \sum_{i=1}^{N} ||\hat{\mathbf{g}}_i^c - \mathbf{R}\hat{\mathbf{g}}_i^b||_2^2 \in \mathbb{R}^+$$
 (8)

subject to $\mathbf{R}^*\mathbf{R} = \mathbf{I}_3$. The solution to such a problem was given in [12]. Let

$$\mathbf{H} \triangleq \sum_{i=1}^{N} \hat{\mathbf{g}}_{i}^{b} (\hat{\mathbf{g}}_{i}^{c})^{*} \in \mathbb{R}^{3 \times 3},$$

compute the SVD, $\mathbf{H} = \mathbf{U}\Lambda\mathbf{V}^*$, and form $\mathbf{X} = \mathbf{V}\mathbf{U}^*$. If $\det(\mathbf{X}) = +1$ then $\hat{\mathbf{R}}_b^c = \mathbf{X}$ is the solution to the Least Squares-problem.

The initial estimates of the transformation parameters are subsequently refined by estimating the offset error $\delta \mathbf{p}_c^b \in \mathbb{R}^3$ and orientation error $\delta \boldsymbol{\varphi} \in \mathbb{R}^3$.

C. IMU calibration parameters

In accordance with the sensor model in (1), the misalignments and scale factors of the accelerometers are denoted as $\boldsymbol{\alpha} = [\alpha_{yz}, \alpha_{zy}, \alpha_{zx}]^{\top}$ and $\mathbf{k}^a = [k_x^a, k_y^a, k_z^a]^{\top}$, respectively. Similarly for the gyroscopes (3), $\boldsymbol{\gamma} = [\gamma_{yz}, \gamma_{zy}, \gamma_{xz}, \gamma_{zx}, \gamma_{xy}, \gamma_{yx}]^{\top}$ denotes misalignments and $\mathbf{k}^g = [k_{\phi}^g, k_{\theta}^g, k_{\psi}^g]^{\top}$ scale factors. The parameters are assumed to be constant during the calibration procedure.

D. Discrete-time process model

Writing the error state space model in discrete-time form yields

$$\begin{split} \delta\mathbf{p}_{k+1}^{n} &= \delta\mathbf{p}_{k}^{n} + dt\delta\mathbf{v}_{k}^{n} \\ \delta\mathbf{v}_{k+1}^{n} &= \delta\mathbf{v}^{n} + dt\hat{\mathbf{R}}_{b,k}^{n}(\mathbf{I}_{3} - \mathbf{A}_{k}^{-1})\tilde{\mathbf{f}}_{k}^{b} \\ &+ dt[\hat{\mathbf{R}}_{b,k}^{n}\mathbf{A}_{k}^{-1}\tilde{\mathbf{f}}_{k}^{b}]_{\times}\delta\boldsymbol{\theta}_{k} + dt\hat{\mathbf{R}}_{b,k}^{n}\mathbf{A}_{k}^{-1}(\delta\mathbf{f}_{k}^{b} + \mathbf{n}_{\mathbf{f},k}) \\ \delta\boldsymbol{\theta}_{k+1} &= \delta\boldsymbol{\theta}_{k} + dt\hat{\mathbf{R}}_{b,k}^{n}(\mathbf{G}_{k}^{-1} - \mathbf{I}_{3})\tilde{\boldsymbol{\omega}}_{k}^{b} \\ &- dt\hat{\mathbf{R}}_{b,k}^{n}\mathbf{G}_{k}^{-1}(\delta\boldsymbol{\omega}_{k}^{b} + \mathbf{n}_{\boldsymbol{\omega},k}). \end{split} \tag{9}$$

The slow-varying biases are modeled as Brownian motion,

$$\delta \mathbf{f}_{k+1}^b = \delta \mathbf{f}_k^b + dt \mathbf{n}_{\delta \mathbf{f},k}$$

$$\delta \boldsymbol{\omega}_{k+1}^b = \delta \boldsymbol{\omega}_k^b + dt \mathbf{n}_{\delta \boldsymbol{\omega},k}.$$
(10)

Let the total state vector be denoted

$$\mathbf{x} = [(\delta \mathbf{x}^{\text{ins}})^{\top} \quad (\delta \mathbf{x}^{\text{cam}})^{\top} \quad (\mathbf{x}^{\text{imu}})^{\top}]^{\top} \in \mathbb{R}^{36}$$

where

$$\begin{split} \delta \mathbf{x}^{\text{ins}} &\triangleq \text{vec}([\delta \mathbf{p}^n \quad \delta \mathbf{v}^n \quad \delta \boldsymbol{\theta} \quad \delta \mathbf{f}^b \quad \delta \boldsymbol{\omega}^b]) \\ \delta \mathbf{x}^{\text{cam}} &\triangleq \text{vec}([\delta \mathbf{p}^b_c \quad \delta \boldsymbol{\varphi}]) \\ \mathbf{x}^{\text{imu}} &\triangleq [\boldsymbol{\alpha}^\top \quad (\mathbf{k}^a)^\top \quad \boldsymbol{\gamma}^\top \quad (\mathbf{k}^g)^\top]^\top, \end{split}$$

and $\operatorname{vec}(\cdot): \mathbb{R}^{n\times m} \to \mathbb{R}^{nm\times 1}$ is the vectorization operation, then the discrete-time process can be written as a state-space model with a nonlinear and linear part,

$$\delta \mathbf{x}_{k+1}^{\text{ins}} = \mathbf{f}_{k}(\delta \mathbf{x}_{k}^{\text{ins}}, \mathbf{x}_{k}^{\text{imu}}, \mathbf{n}_{k})$$

$$\delta \mathbf{x}_{k+1}^{\text{cam}} = \delta \mathbf{x}_{k}^{\text{cam}}$$

$$\mathbf{x}_{k+1}^{\text{imu}} = \mathbf{x}_{k}^{\text{imu}}.$$
(11)

The covariance matrix $\mathbf{Q} \in \mathbb{R}^{12 \times 12}$ of the augmented noise vector $\mathbf{n}_k \triangleq \text{vec}([\mathbf{n}_{\mathbf{f},k} \quad \mathbf{n}_{\boldsymbol{\omega},k} \quad \mathbf{n}_{\delta\mathbf{f},k} \quad \mathbf{n}_{\delta\boldsymbol{\omega},k}])$ is assumed to be time-invariant.

IV. MEASUREMENT MODEL

A feature point i on the calibration pattern, with a known coordinate $\mathbf{d}_i^n \in \mathbb{R}^3$ in $\{n\}$, i.e. relative to the gravitation field, is projected on the image plane according to a pinhole camera model,

$$\mathbf{z}^{(i)} = \bar{\mathbf{K}}(\begin{bmatrix} 0 & 0 & 1 \end{bmatrix} \mathbf{d}_i^c)^{-1} \mathbf{d}_i^c + \mathbf{v}^{(i)} \in \mathbb{R}^2, \tag{12}$$

where $\mathbf{v}^{(i)}$ is measurement noise that arises from the image caption process, $\bar{\mathbf{K}} \in \mathbb{R}^{2 \times 3}$ is a truncated camera matrix containing its intrinsic parameters, and

$$\begin{split} \mathbf{d}_{i}^{c} &= \mathbf{R}_{b}^{c} \big[\mathbf{R}_{n}^{b} (\mathbf{d}_{i}^{n} - \mathbf{p}^{n}) - \mathbf{p}_{c}^{b} \big] \\ &= \mathbf{C} (\delta \boldsymbol{\varphi})^{\top} \hat{\mathbf{R}}_{b}^{c} \big[\hat{\mathbf{R}}_{n}^{b} \mathbf{C} (\delta \boldsymbol{\theta}) (\mathbf{d}_{i}^{n} - \hat{\mathbf{p}}^{n} + \delta \mathbf{p}^{n}) - \hat{\mathbf{p}}_{c}^{b} + \delta \mathbf{p}_{c}^{b} \big]. \end{split}$$

Concatenating M observed feature points into one vector yields the measurement function:

$$\bar{\mathbf{z}}_k = \mathbf{h}_k(\delta \mathbf{x}_k^{\text{ins}}, \delta \mathbf{x}_k^{\text{cam}}) + \bar{\mathbf{v}}_k \in \mathbb{R}^{2M}. \tag{13}$$

If the measurement noise from each projected feature is assumed to be mutually uncorrelated, the covariance matrix $\mathbf{R}_k \in \mathbb{R}^{2M \times 2M}$ of $\bar{\mathbf{v}}_k$ has a block-diagonal structure.

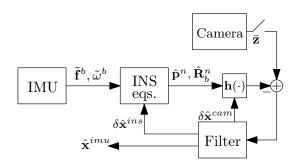


Fig. 3. Feedback formulation.

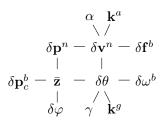


Fig. 4. Correlation diagram of states, parameters and observations.

V. ESTIMATION FRAMEWORK

The process model (11) and measurement model (13) form a state-space system,

$$\mathbf{x}_{k+1} = \bar{\mathbf{f}}_k(\mathbf{x}_k, \mathbf{n}_k)$$
$$\bar{\mathbf{z}}_k = \mathbf{h}_k(\mathbf{x}_k) + \bar{\mathbf{v}}_k,$$
 (14)

that is used in an estimation approach shown in Figure 3. The high-rate IMU propagates the INS state estimates that deviate from their ideal trajectories when the system is in motion. These deviations $\delta \mathbf{x}^{\text{ins}}$, caused by the sensor imperfections, are periodically estimated and corrected by exploiting the known visual calibration pattern. Thus the observations $\bar{\mathbf{z}}$ provide information about the sensor model parameters \mathbf{x}^{imu} that give rise to the deviations. In addition the errors of the coordinate transformation parameters $\delta \mathbf{x}^{\text{cam}}$ are estimated and used for correction. The relationship is illustrated in Figure 4.

The signal processor applied to solve the joint state and model parameter estimation problem is a recursive Sigma-Point Kalman filter which uses a linear combination of the observations to provide an estimate based on the Mean Square Error criterion, by means of statistical linearization [13]. The algorithm is given below but implemented in square-root form for numerical stability [14], [15]. Since the process noise **n** is not additive, one needs to augment the Sigma-Point state vector with noise states as described in Algorithm 1.

VI. SIMULATIONS

A. Sensor signals and parameters

For validation purposes the proposed Sigma-Point Kalman filter is tested with ground truth by means of simulations. The movement of the IMU is generated as a user-defined trajectory $\{\mathbf{p}_k^n\}_{k=0}^T$ using cubic splines to ensure continuous

Algorithm 1 Estimation using Sigma-Point Kalman Filter, with weights $\{w^c\}$, $\{w^m\}$ and η . Augmented sigma points $\bar{\mathcal{X}}_l = [(\mathcal{X}_l)^\top \ (\mathcal{X}_l^n)^\top]^\top$ with N = 36 + 12 states.

```
1: Initialize \hat{\mathbf{x}}_0^-, \mathbf{P}_0^-
  2: for k = 0, ... do
                    Update INS equations
  3:
                    if feature points \{\mathbf{z}^{(i)}\} observed then
  4:
                              Form \bar{\mathbf{z}}_k and \mathbf{R}_k
   5:
                               %Generate sigma points and prediction:
   6:
   7:
                               \boldsymbol{\mathcal{Z}}_{l,k}^- = \mathbf{h}_k(\boldsymbol{\mathcal{X}}_{l,k}^-)
                              \hat{\bar{\mathbf{z}}}_{k}^{-} = \sum_{l=0}^{2N} w_{l}^{m} \mathcal{Z}_{l,k}^{-}
  8:
                               %Measurement update:
   9:
                              \mathbf{R}_{\mathbf{e},k} = \sum_{l=0}^{2N} w_l^c \left( \mathbf{\mathcal{Z}}_{l,k}^- - \hat{\bar{\mathbf{z}}}_k^- \right) \left( \mathbf{\mathcal{Z}}_{l,k}^- - \hat{\bar{\mathbf{z}}}_k^- \right)^* + \mathbf{R}_k
10:
                              \mathbf{R}_{\mathbf{x},\mathbf{e},k} = \sum_{l=0}^{2N} w_l^c (\boldsymbol{\mathcal{X}}_{l,k}^- - \hat{\mathbf{x}}_k^-) (\boldsymbol{\mathcal{Z}}_{l,k}^- - \hat{\mathbf{z}}_k^-)^*
11:
                              \mathbf{K}_k = \mathbf{R}_{\mathbf{x},\mathbf{e},k} \mathbf{R}_{\mathbf{e},k}^{-1}
12:
                              \hat{\mathbf{x}}_k = \hat{\mathbf{x}}_k^- + \mathbf{K}_k (\bar{\mathbf{z}}_k - \hat{\mathbf{z}}_k^-) 
 \mathbf{P}_k = \mathbf{P}_k^- - \mathbf{K}_k \mathbf{R}_k \mathbf{K}_k^* 
Use \delta \hat{\mathbf{x}}_k^{\text{ins}} and \delta \hat{\mathbf{x}}_k^{\text{cam}} to correct state estimates
13:
14:
15:
                               \mathbf{s}_k := \hat{\mathbf{x}}_k
16:
                              S_k := P_k
17:
                     else
18:
                    egin{aligned} \mathbf{s}_k &:= \hat{\mathbf{x}}_k^- \ \mathbf{S}_k &:= \mathbf{P}_k^- \ \mathbf{end} \ \mathbf{if} \end{aligned}
19:
20:
21:
                     %Generate sigma points and prediction
22:
                     Form \bar{\mathbf{s}}_k = [\mathbf{s}_k^{\mathsf{T}} \quad \mathbf{0}_{1\times 12}^{\mathsf{T}}]
23:
                    Form augmented covariance matrix \bar{\mathbf{S}}_k = \mathbf{S}_k \oplus \mathbf{Q}
\bar{\mathbf{X}}_{l,k} = \bar{\mathbf{s}}_k \pm \eta \cdot \operatorname{col}[\bar{\mathbf{S}}_k^{1/2}]_l
24:
25:
                     oldsymbol{\mathcal{X}}_{l,k+1}^- = ar{\mathbf{f}}_k(oldsymbol{\mathcal{X}}_{l,k},oldsymbol{\mathcal{X}}_{l,k}^n)
26:
                    27:
28:
29:
```

TABLE I IMU sensor model parameters.

Misalignment [°]						
α	1	-1	1	_	_	_
γ	-1	1	0	1	1	-1
Scale factor [-]						
\mathbf{k}^a	1.01	0.95	1.04			
\mathbf{k}^{g}	1.02	0.97	0.98			

acceleration. The attitude and hence the angular rate is determined by directing the camera towards the calibration pattern. The outputs are ideal accelerometer and gyroscope signals $\{\mathbf{f}_k^b, \boldsymbol{\omega}_k^b\}_{k=0}^T$ sampled at 100 [Hz].

Misalignments (α, γ) and scale factors $(\mathbf{k}^a, \mathbf{k}^g)$ are added according to the sensor model, with values set in Table I. The biases $(\delta \mathbf{f}^b, \delta \boldsymbol{\omega}^b)$ are set as constants and the additive white noise terms $\{\mathbf{n}_{\mathbf{f},k}, \mathbf{n}_{\boldsymbol{\omega},k}\}_{k=0}^T$ are generated by Gaussian distributions with standard deviations $(\sigma_f, \sigma_{\boldsymbol{\omega}})$. The magnitudes, applied here equally to each axis, were set using data from a low-cost IMU and given in Table II.

The planar visual calibration pattern consists of a grid of 5×5 feature points $\{\mathbf{d}_i^n\}_{i=1}^{25}$ spaced 7×10^{-2} [m] apart and placed perpendicular to the gravitation field. The camera has a focal length $f=5\times 10^{-3}$ [m], pixel width $\Delta_x=\Delta_y=6\times 10^{-6}$ [m] and a sample rate of 10 [Hz]. Whilst feature

TABLE II IMU sensor noise parameters.

	Bias	Std.
Acc. [m/s ²]	2×10^{-2}	6×10^{-3}
Gyro. [rad./s]	5×10^{-3}	3×10^{-3}

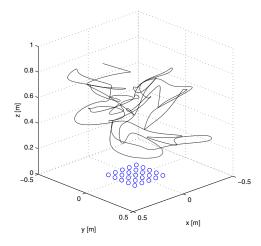


Fig. 5. Trajectory spanning 120 seconds over a 28×28 cm grid of feature points.

point detectors are capable of interpolating image coordinates to subpixel level we set the pixel noise standard deviation σ_v to a conservative value of 2 pixels that takes into account possible motion blur. The camera offset is $\mathbf{p}_c^b = [1, -5, 10]^{\mathsf{T}} \cdot 10^{-2}$ [m] and rotation parameterized in Euler angles, $\boldsymbol{\varphi} = [-90^\circ, 0^\circ, -90^\circ]^{\mathsf{T}}$.

Figure 5 illustrates the motion of the system during two minutes. The camera is directed down towards the pattern during the entire calibration procedure.

B. Estimation setup

The weights in the Sigma-Point filter are set as $w_l^c=w_l^m=\frac{1}{2(N+\lambda)}$ for $l=1,\ldots,2N.$ For l=0, $w_0^c=\frac{\lambda}{N+\lambda}+(1-\alpha^2+\beta)$ and $w_0^m=\frac{\lambda}{N+\lambda}.$ Here $\lambda=\alpha^2(N+\kappa)-N,$ with parameters $\alpha=0.1,$ $\beta=2$ and $\kappa=0$ that also determine the spread of the sigma points through the weight $\eta=\sqrt{N+\lambda}.$

The initial error state estimates $\delta \hat{\mathbf{x}}_0^{\text{ins},-}$ are zero, with standard deviations of initial estimation error of position, $2 \cdot 10^{-2}$ [m], velocity, $1 \cdot 10^{-3}$ [m/s], and attitude errors, 2 [°], respectively. The standard deviations for the bias error estimates were set to $2 \cdot 10^{-2}$ [m/s²] and $4 \cdot 10^{-3}$ [rad./s], respectively. The *actual* initial errors $(\delta \mathbf{p}_0^n, \delta \mathbf{v}_0^n, \delta \boldsymbol{\theta}_0)$ were generated by a zero-mean Gaussian distribution with corresponding covariance matrix. The initial coordinate transformation error estimates $\delta \hat{\mathbf{x}}_0^{\text{cam},-}$ were zero, with standard deviations $1 \cdot 10^{-2}$ [m] and 2 [°]. Similarly, the *actual* errors were generated by a Gaussian distribution. Finally, the initial misalignment estimates are all zero and scale factor estimates set to one, with standard deviations of 1 [°] and $1 \cdot 10^{-2}$ [-], respectively, and form $\hat{\mathbf{x}}_0^{\text{inu},-}$.

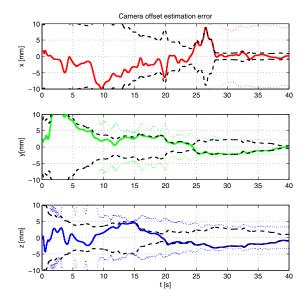


Fig. 6. Estimation errors of $\mathbf{p}_c^b = [p_{c,x}^b, p_{c,y}^b, p_{c,z}^b]^{\top}$.

The standard deviations form together a diagonal error covariance matrix \mathbf{P}_0^- .

C. Results

The performance of the estimator is dependent on the dynamics of the system, i.e. on the sensor signals that are excited. The example trajectory used here is shown in Figure 5. It exhibits roll $\phi \in [-90^{\circ}, 90^{\circ}]$ with a period of 8 seconds and has a peak speed of about 0.40 [m/s] and angular rate of 2.9 [rad./s].

The time-varying estimation error statistics during the initial 40 seconds are shown in figures 6 to 12, using 100 Monte Carlo simulations. The estimated mean errors and Root Mean Square Error (RMSE) bounds are plotted in colored lines and dashed black lines respectively. The figures also show the 3σ -bounds in dotted lines, obtained from the diagonal elements of the filter's error covariance matrix, which provide a representation of its uncertainty of the estimates. We see that while the approximation of the Sigma-Point method cannot fully capture the statistics of the highly non-linear system, the signal processor is capable of providing estimates that quickly reach subcentimeter and subdegree accuracy, as well as scale factor deviations by less that 1 percentage point.

Results for the entire 120 second trajectory from 100 Monte Carlo simulations are shown in Table III. These include the final RMSE of the estimates as well as the means and standard deviations of the errors for the coordinate transformation parameters, sensor biases, misalignments and scale factors, respectively. As can be seen the estimate biases tend to be of the same order of magnitude as the standard deviation of errors.

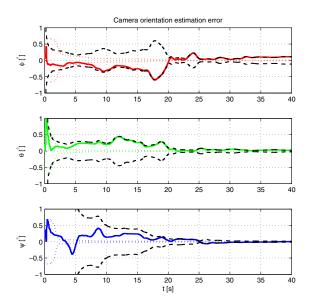


Fig. 7. Estimation errors of $\varphi = [\phi, \theta, \psi]^{\top}$.

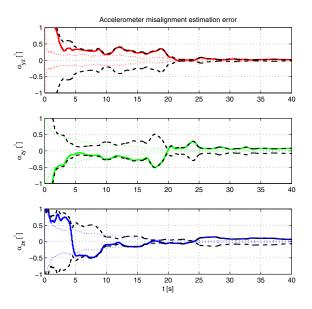


Fig. 8. Estimation errors of $\alpha = [\alpha_{yz}, \alpha_{zy}, \alpha_{zx}]^{\top}$.

VII. CONCLUSIONS

A simple and repeatable calibration procedure for low-cost inertial measurement units has been proposed, using a monocular camera with a visual calibration pattern. The sensor models capture misalignments, scale and offset errors as well as IMU-camera orientation and offset errors which are estimated by a recursive Sigma-Point Kalman filter. A simulation study indicates that it is capable of reaching subcentimeter and subdegree accuracy and scale factor deviations below 0.1 percentage points. Future work for validation would have use

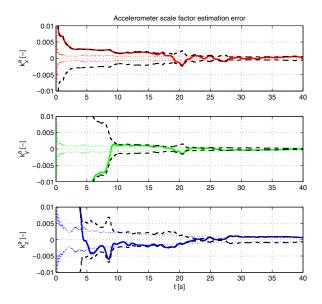


Fig. 9. Estimation errors of $\mathbf{k}^a = [k_x^a, k_y^a, k_z^a]^{\top}$.

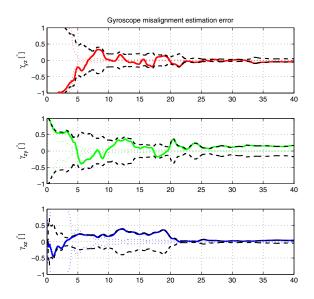


Fig. 10. Estimation errors of $[\gamma_{yz}, \gamma_{zy}, \gamma_{xz}]$.

real inertial and visual sensor data, implement feature point extraction, construct tests for whiteness of the innovations and use detailed information about the sensor properties from an IMU manufacturer.

REFERENCES

- [1] I. Skog and P. Händel, "Calibration of a mems inertial measurement unit," in *Proc. IMEKO XVII*, Sept. 2006.
- [2] A. Kim and M. Golnaraghi, "Initial calibration of an inertial measurement unit using an optical position tracking system," in *Proc. Position Location and Navigation Symposium (PLANS)*, pp. 96–101, 2004.

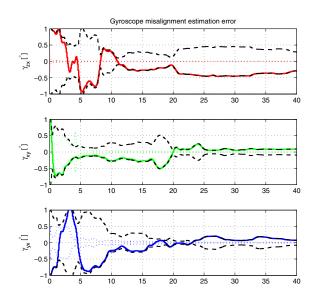


Fig. 11. Estimation errors of $[\gamma_{zx}, \gamma_{xy}, \gamma_{yx}]$.

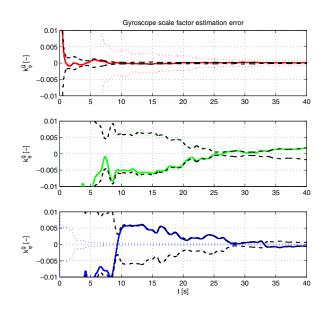


Fig. 12. Estimation errors of $\mathbf{k}^g = [k_\phi^g, k_\theta^g, k_\psi^g]^\top$.

- [3] S. Ebcin and M. Veth, "Tightly-coupled image-aided inertial navigation using the unscented kalman filter," in *Proc. 20th Int. Technical Meeting* of the Satellite Division of the Institute of Navigation ION GNSS, pp. 1851–1860, Sept. 2007.
- [4] M. George and S. Sukkarieh, "Inertial navigation aided by monocular camera observations of unknown features," in *Proc. IEEE Int. Conf. Robotics and Automation*, pp. 3558–3564, Apr. 2007.
- [5] D. Zachariah and M. Jansson, "Camera-aided inertial navigation using epipolar points," in *Proc. IEEE/ION Position Location and Navigation* Symposium (PLANS), May 2010.
- [6] P. Pinies, T. Lupton, S. Sukkarieh, and J. Tardos, "Inertial aiding of inverse depth slam using a monocular camera," in *Proc. IEEE Int. Conf. Robotics and Automation*, pp. 2797–2802, Apr. 2007.
- [7] J. Hol, T. Schon, F. Gustafsson, and P. Slycke, "Sensor fusion for

 $\begin{tabular}{ll} TABLE~III\\ Error~statistics, 100~Monte~Carlo~simulations. \end{tabular}$

		DIAGE		0.1
		RMSE	Mean	Std.
$p_{c,x}^b$		0.747	0.466	0.585
$p_{c,y}^b$	$\times 10^{-3}$ [m]	1.737	-1.727	0.189
$p_{c,z}^b$		1.430	1.415	0.206
φ_{ϕ}		6.756	6.628	1.308
φ_{θ}	$\times 10^{-2} [^{\circ}]$	1.960	-1.752	0.880
φ_{ψ}		1.727	0.999	1.410
δf_x		6.447	5.994	2.374
δf_y	$\times 10^{-3} [\text{m/s}^2]$	1.064	-0.801	0.700
δf_z		5.436	-5.257	1.384
$\delta\omega_{\phi}$		0.104	-0.100	0.028
$\delta\omega_{ heta}$	$\times 10^{-3}$ [rad./s]	0.846	0.846	0.033
$\delta\omega_{\psi}$		0.157	-0.154	0.033
α_{yz}		0.528	0.057	0.524
α_{zy}	$\times 10^{-2} [^{\circ}]$	5.754	5.662	1.029
α_{zx}		0.574	0.023	0.574
γ_{yz}		2.975	-2.720	1.205
γ_{zy}		9.020	8.865	1.666
γ_{xz}	$\times 10^{-2} [^{\circ}]$	2.133	2.050	0.587
γ_{zx}		11.837	-11.758	1.371
γ_{xy}		6.030	5.959	0.920
γ_{yx}		3.192	-3.028	1.010
k_x^a		5.801	5.236	2.497
k_u^a	$\times 10^{-4}[-]$	1.018	-0.352	0.955
$\frac{k_z^g}{k_z^g}$		8.372	8.268	1.318
O O		0.469	0.256	0.393
$k_{ heta}^{ar{g}}$	$\times 10^{-4}[-]$	9.765	9.536	2.103
$k_{ heta}^{ar{g}} \ k_{\psi}^{g}$		3.848	-2.753	2.688

augmented reality," in *Proc. 9th Int. Conf. Information Fusion*, pp. 1 –6, 2006.

- [8] J. Lobo and J. Dias, "Relative pose calibration between visual and inertial sensors," *Int. J. Robotics Research*, vol. 26, June 2007.
- [9] J. Hol, T. Schon, and F. Gustafsson, "A new algorithm for calibrating a combined camera and imu sensor unit," in *Proc. 10th Int. Conf. Control, Automation, Robotics and Vision (ICARCV)*, pp. 1857–1862, 2008.
- [10] F. Mirzaei and S. Roumeliotis, "A kalman filter-based algorithm for imucamera calibration: Observability analysis and performance evaluation," *IEEE Trans. Robotics*, vol. 24, pp. 1143–1156, Oct. 2008.
- [11] J.-Y. Bouguet, "Camera calibration toolbox." Computational Vision at Caltech, www.vision.caltech.edu/bouguetj/, 2008. Last accessed on June 23 2010
- [12] K. S. Arun, T. S. Huang, and S. D. Blostein, "Least-squares fitting of two 3-d point sets," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. PAMI-9, pp. 698–700, Sept. 1987.
- [13] S. Julier and J. Uhlmann, "A new extension of the kalman filter to nonlinear systems," in *Proc. Signal Processing, Sensor Fusion, and Target Recognition VI*, pp. 182–193, Apr. 1997.
- [14] T. Kailath, A. Sayed, and B. Hassibi, *Linear Estimation*. Prentice Hall, 2000.
- [15] R. Van der Merwe and E. Wan, "The square-root unscented kalman filter for state and parameter-estimation," in *Proc. IEEE Int. Conf. Acoustics*, *Speech, and Signal Processing (ICASSP)*, pp. 3461–3464, May 2001.