Self-Calibration of Inertial and Omnidirectional Visual Sensors for Navigation and Mapping

Jonathan Kelly and Gaurav S. Sukhatme

Abstract— **Omnidirectional cameras are versatile sensors that are able to provide a full 360-degree view of the environment. When combined with inertial sensing, omnidirectional vision offers a potentially robust navigation solution. However, to correctly fuse the data from an omnidirectional camera and an inertial measurement unit (IMU) into a single navigation frame, the 6-DOF transform between the sensors must be accurately known. In this paper we describe an algorithm, based on the unscented Kalman filter, for self-calibration of the transform between an omnidirectional camera and an IMU. We show that the IMU biases, the local gravity vector, and the metric scene structure can also be recovered from camera and IMU measurements. Further, our approach does not require any additional hardware or prior knowledge about the environment in which a robot is operating. We present results from calibration experiments with an omnidirectional camera and a low-cost IMU, which demonstrate accurate self-calibration of the 6-DOF sensor-to-sensor transform.**

I. Introduction

Vision is a rich sensing modality that is useful for a wide variety of robot navigation tasks. In systems which employ standard refractive camera lenses, however, the camera's limited field of view can be a weakness – there may be a complete loss of tracked features or objects when the camera rotates or translates quickly. One solution is to use an omnidirectional lens, effectively trading resolution for a 360-degree field of view. In order to image such a large azimuth range, omnidirectional lenses are often *catadioptric*, i.e., they incorporate both reflective and refractive elements [1], [2].

Although omnidirectional cameras have the advantage of a wide field of view, they can be more prone to problems caused by motion blur and resolution limitations. Additionally, absolute depth and scale information cannot be recovered using measurements from a single camera alone. Obtaining this depth information, which is required for metric mapping and in some cases for path planning, requires another sensor, such as an inertial measurement unit (IMU). Recent work has shown that visual and inertial sensors can be used effectively in combination to estimate egomotion [3], [4]. However, the six degrees-of-freedom (6-DOF) transform between the camera and the IMU must be accurately known for measurements to be properly fused in the navigation frame. Calibration of this transform is usually a difficult process, which requires additional equipment and which must be repeated whenever the sensors are repositioned.

Instead of manual calibration, a more efficient and convenient approach is to *self-calibrate* the sensor-to-sensor transform. Self-calibration refers to the process of using (noisy) measurements from the sensor themselves to improve estimates of the related system parameters. In this paper, we extend our previous work on camera-IMU calibration [5] to omnidirectional sensors and demonstrate *true* self-calibration, i.e., calibration without the use of a known calibration object. We formulate calibration as a filtering problem, and show that is it possible to estimate the 6-DOF camera-IMU transform, the time-varying IMU biases, the local gravity vector and the scene structure *at the same time*. That is, we simultaneously calibrate, localize the camera-IMU platform and build a map of the environment (i.e., perform SLAM). We emphasize that, unlike in monocular camera-only systems, we are able to recover the scene scale.

The remainder of the paper is organized as follows. We review prior work in Section II. In Sections III and IV, we describe our system model and our filtering algorithm, respectively. We then give an overview of our calibration experiments in Section V, and present results from those experiments in Section VI. Finally, we offer some conclusions and directions for future work in Section VII.

II. Related Work

There is a very large body of literature on vision for robotics applications, which includes the use of omnidirectional sensors. We focus specifically on visual-inertial sensor calibration in this section. For example, Lobo and Dias [6] describes a camera-IMU calibration procedure in which the relative orientation and relative translation of the sensors are determined independently. The procedure requires a pendulum unit and a turntable, making it impractical for larger robot platforms.

More closely related to our work is the approach developed by Hol et al. [7] for calibrating the relative transform between a spherical camera and an IMU. A similar algorithm was proposed by Mirzaei and Roumeliotis [8] for standard perspective camera systems. These algorithms operate by tracking corner points on a planar calibration target and fusing the image measurements with IMU data.

Jones, Vedaldi and Soatto present an observability analysis in [9] which shows that the camera-IMU relative pose, gravity vector and scene structure can be recovered from camera and IMU measurements only. However, their work assumes that the IMU biases are static over the calibration

This work was funded in part by the National Science Foundation under grants CNS-0325875, CNS-0540420, CNS-0520305 and CCF-0120778 and by a gift from the Okawa Foundation. Jonathan Kelly was supported by an Annenberg Fellowship from the University of Southern California and by an NSERC postgraduate scholarship from the Government of Canada.

Jonathan Kelly and Gaurav S. Sukhatme are with the Robotic Embedded Systems Laboratory, University of Southern California, Los Angeles, California, USA 90089-0781 Email: {jonathsk, gaurav}@usc.edu

Proceedings of the 2nd. Workshop on Omnidirectional Robot Vision
A workshop of the 2010 IEEE International Conference on Robotics and Automation (ICRA2010)
Anchorage, Alaska, USA, May 7, 2010, ISBN 978-88-95872-02-5
pp. 1-6

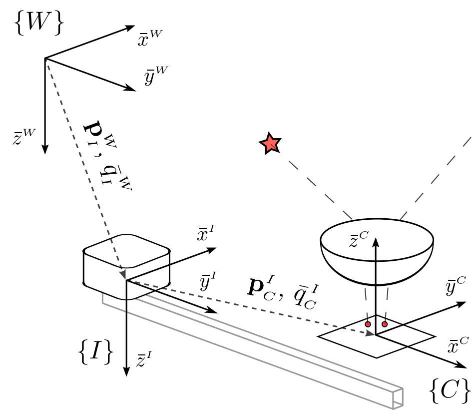


Fig. 1. Relationship between the world $\{W\}$, IMU $\{I\}$, and omnidirectional camera $\{C\}$ reference frames. The goal of the calibration procedure is to determine the transform $(\mathbf{p}_C^I, \bar{q}_C^I)$ between the camera and IMU, using accelerometer and gyroscope data and camera images of static point landmarks (red stars) in the immediate environment.

interval – for the low-cost inertial sensors that we consider in this paper, drift in the biases can be significant even over short periods of time.

Importantly, our technique does not require *any* additional apparatus in the general case (although a planar calibration target can be used, if one is available), and explicitly models uncertainty in the gravity vector and in the gyroscope and accelerometer biases.

III. Calibration Approach

We begin by describing our system model below, and then briefly discuss its observability properties in Section III-B. The calibration problem involves three separate reference frames:

1) the *camera frame* $\{C\}$, with its origin at the effective optical center of the omnidirectional 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 resolved, and
3) the *world frame* $\{W\}$, which serves as an absolute reference for both the camera and the IMU.

The relationship between these frames is illustrated in Figure 1. We treat the world frame as an inertial frame, and choose the origin of this frame to coincide with the initial camera position.

A. System Model

We use an unscented Kalman filter (UKF) to simultaneously estimate the pose of the IMU in the world frame, the IMU biases, the gravity vector, and the position and orientation of the omnidirectional camera with respect to the IMU. The 26×1 *sensor* portion of the state vector is

$$\mathbf{x}_s(t) = \big[(\mathbf{p}_I^W(t))^T \quad (\bar{q}_I^W(t))^T \quad (\mathbf{v}^W(t))^T \quad \ldots \\ (\mathbf{b}_g(t))^T \quad (\mathbf{b}_a(t))^T \quad (\mathbf{g}^W)^T \quad (\mathbf{p}_C^I)^T \quad (\bar{q}_C^I)^T\big]^T \tag{1}$$

where $\mathbf{p}_I^W$ is the position of the IMU in the world frame, $\bar{q}_I^W$ is the (unit quaternion) orientation of the IMU frame relative to the world frame, $\mathbf{v}^W$ is the linear velocity of the IMU in the world frame, $\mathbf{b}_g$ and $\mathbf{b}_a$ are the gyroscope and accelerometer biases, respectively, and $\mathbf{g}^W$ is the gravity vector in the world frame. The remaining entries, $\mathbf{p}_C^I$ and $\bar{q}_C^I$, are static parameters which define the position and the (unit quaternion) orientation of the camera frame relative to the IMU frame.

During self-calibration, we also estimate the positions of a series of point landmarks in the environment. The complete state vector is

$$\mathbf{x}(t) = \big[\mathbf{x}_s^T(t) \quad \mathbf{x}_m^T\big]^T, \quad \mathbf{x}_m = \big[(\mathbf{p}_{l_1}^W)^T \quad \ldots \quad (\mathbf{p}_{l_n}^W)^T\big]^T \tag{2}$$

where $\mathbf{x}_m$ is the *map* portion of the state vector. Each entry in $\mathbf{x}_m$ is a 3×1 vector, $\mathbf{p}_{l_i}^W$, that defines the position of landmark i in the world frame, $i = 1, \ldots, n$. The complete state vector has size $(26 + 3n) \times 1$.

In our experiments, we have found it sufficient to use Cartesian coordinates to specify landmark positions. We initialize each landmark at a nominal depth and with a large variance along the camera ray axis, at the camera position where the landmark is first observed. If the true landmark depths vary significantly, it may be more appropriate to use an inverse-depth parameterization [10].

Note also that, because the world frame is defined with respect to the initial camera pose, the relationship between the frame and the local gravity vector can be arbitrary, i.e., it will depend entirely on how the camera is initially oriented. This is one reason why we estimate the gravity vector in the world frame during calibration.

1) Process Model: Our filter process model uses the IMU linear acceleration and angular velocity measurements as control inputs [11]. We model the IMU gyroscope and accelerometer biases as Gaussian random walk processes driven by the white, zero-mean noise vectors $\mathbf{n}_{gw}$ and $\mathbf{n}_{aw}$. The gyroscope and accelerometer measurements are corrupted by zero-mean, white Gaussian noise, defined by the vectors $\mathbf{n}_g$ and $\mathbf{n}_a$, respectively. The evolution of the system state in continuous time is described by

$$\dot{\mathbf{p}}_I^W = \mathbf{v}^W \qquad \dot{\mathbf{v}}^W = \mathbf{a}^W \qquad \dot{\bar{q}}_I^W = \frac{1}{2}\boldsymbol{\Omega}(\boldsymbol{\omega}^I)\,\bar{q}_I^W \tag{3}$$

$$\dot{\mathbf{b}}_g = \mathbf{n}_{gw} \qquad \dot{\mathbf{b}}_a = \mathbf{n}_{aw} \qquad \dot{\mathbf{g}}^W = \mathbf{0}_{3\times1} \tag{4}$$

$$\dot{\mathbf{p}}_C^I = \mathbf{0}_{3\times1} \qquad \dot{\bar{q}}_C^I = \mathbf{0}_{4\times1} \tag{5}$$

The term $\boldsymbol{\Omega}(\boldsymbol{\omega}^I)$ above is the 4×4 quaternion kinematical matrix, which relates the time rate of change of the orientation quaternion to the IMU angular velocity. Vectors $\mathbf{a}^W$ and $\boldsymbol{\omega}^I$ are the linear acceleration of the IMU in the world frame and the angular velocity of the IMU in the IMU frame, respectively. These quantities are related to the *measured* IMU angular velocity, ω_m, and linear acceleration, $\mathbf{a}_m$, by

$$\boldsymbol{\omega}_m = \boldsymbol{\omega}^I + \mathbf{b}_g + \mathbf{n}_g \tag{6}$$

$$\mathbf{a}_m = \mathbf{C}^T(\bar{q}_I^W)\,(\mathbf{a}^W - \mathbf{g}^W) + \mathbf{b}_a + \mathbf{n}_a \tag{7}$$

where $\mathbf{C}(\bar{q}_I^W)$ is the direction cosine (rotation) matrix that describes the orientation of the IMU frame with respect to the world frame. We propagate the state forward in time using fourth-order Runge-Kutta integration of (3) to (5) above.

2) Measurement Model: As the sensor platform moves, the camera captures images of tracked landmarks in the environment, over a full 360 degrees of azimuth. We use the omnidirectional camera model described in [12], and calibrate the camera intrinsic parameters prior to our mapping or navigation experiments. Measurement $\mathbf{z}_i$ is the projection of landmark l_i, at position $\mathbf{p}_{l_i}^C = \begin{bmatrix} x_i & y_i & z_i \end{bmatrix}^T$ in the camera frame, onto the image plane

$$\mathbf{p}_{l_i}^C = \begin{bmatrix} x_i \\ y_i \\ z_i \end{bmatrix} = \mathbf{C}^T(\bar{q}_C^I)\,\mathbf{C}^T(\bar{q}_I^W)\left(\mathbf{p}_{l_i}^W - \mathbf{p}_I^W\right) - \mathbf{C}^T(\bar{q}_C^I)\,\mathbf{p}_C^I$$

$$\mathbf{z}_i = \begin{bmatrix} u_i \\ v_i \end{bmatrix} = \begin{bmatrix} x_i' \\ y_i' \end{bmatrix} + \boldsymbol{\eta}_i, \qquad \begin{bmatrix} x_i' \\ y_i' \\ 1 \end{bmatrix} = \boldsymbol{\mathcal{P}} \begin{bmatrix} x_i/z_i \\ y_i/z_i \\ 1 \end{bmatrix} \tag{8}$$

where $\begin{bmatrix} u_i & v_i \end{bmatrix}^T$ is the vector of observed image coordinates, $\boldsymbol{\mathcal{P}}$ is the 3×3 camera intrinsic matrix, and $\boldsymbol{\eta}_i$ is a Gaussian measurement noise vector with covariance matrix $\mathbf{R}_i = \sigma_i^2\,\mathbf{I}_{2\times 2}$. The projection function is modeled as a fourth-order polynomial, plus an affine transform that maps between the camera CCD and the (virtual) image plane.

When several landmarks are visible in one image, we stack the individual measurements to form a single measurement vector $\mathbf{z} = \begin{bmatrix} \mathbf{z}_1^T & \dots & \mathbf{z}_n^T \end{bmatrix}^T$ and the associated block-diagonal covariance matrix $\mathbf{R} = \mathtt{diag}\begin{pmatrix} \mathbf{R}_1 & \dots & \mathbf{R}_n \end{pmatrix}$. This vector is processed by the filtering algorithm in a single step.

B. Observability of Self-Calibration

In order to calibrate the sensor-to-sensor transform between the camera and the IMU, the relevant system states must be *observable*. That is, we must be able recover the state values from the measured system outputs, the control inputs, and a finite number of their time derivatives [13]. Prior work on the observability of camera-IMU relative pose calibration includes [8] and [9].

The general problem of estimating both camera motion and scene structure has been studied extensively in robotics and in computer vision. Chiuso et al. [14] shows that monocular structure-from-motion is observable up to an unknown similarity transform from image measurements alone. If we choose the initial camera position as the origin of our world frame, and fix[1] the initial camera orientation (relative to three or more noncollinear points on the image plane), then following [9], [14], scene structure is observable up to an unknown scale.

We prove as our main result in [15] that, if we are willing to lock down the initial camera orientation, it is possible to observe the relative pose of the camera and the IMU, the gyroscope and accelerometer biases, the gravity vector *and* the local scene structure. Our analysis is based on a differential geometric characterization of observability and relies a matrix rank test originally introduced by Hermann and Krener [16]. The result holds as long as the IMU measures two nonzero angular rotation rates and two nonzero linear accelerations (i.e., along at least two axes).

IV. Unscented Filtering

Measurements from the camera and the IMU are fused in an unscented Kalman filter to estimate the system state, including the calibration parameters. The UKF is a Bayesian filtering algorithm which propagates and updates the state vector using a set of deterministically-selected sample points called *sigma points* [17]. These points, which lie on the covariance contours in state space, capture the mean and covariance of the state distribution. The filter applies the *unscented transform* to the sigma points, propagating each point through the (nonlinear) process and measurement models, and then computes the weighted averages of the transformed points to determine the posterior state mean and state covariance. This is a form of statistical local linearization, which produces more accurate estimates than the analytic local linearization employed by the extended Kalman filter (EKF).

We use the continuous-discrete formulation of the UKF, where the sigma points are propagated forward by integration, while measurement updates occur at discrete time steps. Our filter employs the scaled form of the unscented transform [18], which uses a scaling term to control the spread of the sigma points around the mean. Process noise is incorporated by augmenting the state vector and state covariance matrix with a noise component [17].

The standard UKF algorithm computes the predicted state vector as the weighted barycenteric mean of the sigma points. For unit quaternions, however, the barycenter of the transformed sigma points will often *not* represent the correct mean. In particular, the weighted average of several unit

Fig. 2. Pioneer 2-AT mobile robot, equipped for calibration experiments. The NetVision 360 omnidirectional camera lens is visible in the upper right.

[1] That is, we treat the initial camera orientation as fixed relative to the directions to three or more world points. These directions are considered as absolute and are not estimated in our filter.

quaternions may not be a unit quaternion. There are several ways to enforce the quaternion unit norm constraint within the UKF [19]. We follow the method described in [20] and reparameterize the state vector to incorporate a multiplicative, three-parameter orientation *error state* vector, separate from the unit quaternions $\bar{q}_I^W$ and $\bar{q}_C^I$. This approach, called the USQUE (UnScented QUaternion Estimator) filter in [20], uses a multiplicative local error quaternion and a three-component vector of modified Rodrigues parameters (MRPs), derived from the error quaternion. The MRP vector is an unconstrained rotation representation, which is singular at 2π. Throughout the calibration procedure, the filter maintains an estimate of the full 26×1 sensor state vector and a 24×24 error state covariance matrix. For the orientation quaternions $\bar{q}_I^W$ and $\bar{q}_C^I$, we store the covariance matrices for the MRP error state representations. We can then recover the appropriate quaternion for each sigma point from the mean quaternion and the MRP error covariance matrix.

A. Filter Initialization

At the start of calibration, we compute an initial estimate of the sensor state vector. We use the initial camera position as the origin of the world frame, and set the initial camera orientation arbitrarily.[2] The initial camera pose has zero uncertainty in this case, since it corresponds to the base reference frame.

We also require an initial estimate of the pose of the camera relative to the IMU. For the work described here, we use hand measurements of the relative pose – however this information may in many cases be available from CAD drawings or other sources. Using the estimate of the relative pose of the camera with respect to the IMU, we can then compute an initial estimate of the IMU pose in the world frame

$$\hat{\mathbf{p}}_I^W = -\mathbf{C}(\hat{\bar{q}}_C^W)\,\mathbf{C}^T(\hat{\bar{q}}_C^I)\,\hat{\mathbf{p}}_C^I \tag{9}$$

$$\hat{\bar{q}}_I^W = \hat{\bar{q}}_C^W \otimes \left(\hat{\bar{q}}_C^I\right)^{-1} \tag{10}$$

We compute the covariance matrix for the IMU pose using the Jacobians of (9) and (10), after converting to the MRP orientation representation.

We also estimate the initial map state (landmark positions in the world frame). Typically, approximately 40 to 50 landmarks are selected as static references for calibration. At present, we do not add new landmarks when the camera moves (since we expect to perform calibration in a limited area to obtain faster convergence). As noted in Section III-A, we parameterize the landmark positions in the state vector using Cartesian coordinates, and initialize the positions using a nominal depth value, along the associated camera ray axes. The covariance matrix for each landmark is computed by propagating the image plane uncertainty (1 to 2 pixels) and a large initial variance along the camera ray into 3-D, yielding a covariance ellipsoid in the world frame.

[2] We usually choose an initial orientation that results in the z axis of the IMU pointing downwards, approximately aligned with the local gravity vector.

As a last step, we select three or four highly salient and widely disbursed landmarks (90 degrees to 120 degrees apart, if possible) as anchors, to fix the orientation of the world frame. The covariance ellipsoids for these points are initialized using very small image plane uncertainties. This effectively locks down the initial camera pose and ensures that the full system state is observable.

B. Feature Detection and Matching

We attempt to select a set of well-localized point features as landmarks, using a feature selection algorithm such as SIFT [21]. Feature matching between camera frames is performed by comparing the descriptors for all points that lie within a bounded image region. The size of the search region is determined from the integrated IMU measurements during the interval between camera image updates. For the experiments presented here, we used an empirically-determined threshold on the Euclidean distance between SIFT feature descriptors as a validation gate for matching.

V. Experiments

We performed a series of experiments to quantify the accuracy and performance of calibration algorithm. Although we have tested self-calibration in a variety of unstructured environments, in this paper we restrict ourselves to describing an experimental trial in our laboratory, where landmarks were positioned between one and three meters away from the camera. Further details on the sensor platform and our experimental procedure are given below.

A. Hardware Platform

We use a black and white Flea FireWire camera from Point Grey Research (640×480 pixel resolution), coupled to a NetVision 360 omnidirectional lens manufactured by Remote Reality (shown in Figure 2). Images are captured at a rate of 15 Hz. Our IMU is a Microstrain 3DM-GX3 unit, which provides three-axis angular rate and linear acceleration measurements at 100 Hz. Axis scale and non-orthogonality effects are compensated for internally by the IMU. Both sensors are attached to a rigid 40 cm-long aluminum beam.

For our omnidirectional calibration experiments, the sensor beam was attached to a Directed Perception PTU-D46 pan-tilt unit that was mounted on a Pioneer 2-AT mobile robot. The use of a pan-tilt unit made it possible to excite two degrees of rotational and two degrees of translational freedom while the robot moved on a planar surface (i.e., a flat floor).

B. Experimental Procedure

At the start of each calibration experiment, we initialized the IMU biases by keeping the sensor beam stationary for approximately 10 seconds. After this settling time, we enabled the pan-tilt unit, which followed an oscillating roll trajectory (at constant angular velocity), between the angles of 15°and -15°. We then commanded the Pioneer follow a circular path with a radius of approximately 1 m. Image processing and filtering were performed offline.

TABLE I

SELF-CALIBRATION RESULTS. THE INITIAL HAND-MEASURED (HM) ESTIMATE OF THE CAMERA-IMU TRANSFORM (x, y, z TRANSLATION AND ROLL, PITCH, YAW ORIENTATION) AND THE FINAL SELF-CALIBRATED ESTIMATE ARE LISTED, ALONG WITH THEIR RESPECTIVE 3σ ERROR BOUNDS.

	$p_x \pm 3\sigma$ (cm)	$p_y \pm 3\sigma$ (cm)	$p_z \pm 3\sigma$ (cm)	Roll $\pm 3\sigma$ (°)	Pitch $\pm 3\sigma$ (°)	Yaw $\pm 3\sigma$ (°)
HM	0.00 ± 18.00	-15.00 ± 30.00	0.00 ± 18.00	0.00 ± 15.00	0.00 ± 15.00	0.00 ± 15.00
SC	5.34 ± 0.74	-14.13 ± 0.40	3.15 ± 0.61	-11.79 ± 0.14	-2.39 ± 0.10	-0.38 ± 0.16

The camera-IMU transform parameters were initialized using hand measurements of the relative position and orientation of the sensors. We assumed that each image measurement was corrupted by independent, white Gaussian noise with a standard deviation of 2.0 pixels along both image axes. The intrinsic parameters of the camera were calibrated using the Omnidirectional Camera Calibration Toolbox for MATLAB [12].

Self-calibration requires us to first anchor the orientation of the world reference frame by fixing the directions to three or more points on the image plane. We chose to fix the directions to four widely distributed SIFT features in the first camera image. The initial covariance matrices for these landmarks were computed using very small image plane uncertainties, after averaging the image coordinates over 450 frames (30 seconds) to reduce noise. This averaging was performed while the sensor beam was stationary, before the start of an experimental trial. We selected an initial depth of 2.0 m for all of the landmarks, along the corresponding camera ray, with a standard deviation of 0.6 m.

VI. RESULTS AND DISCUSSION

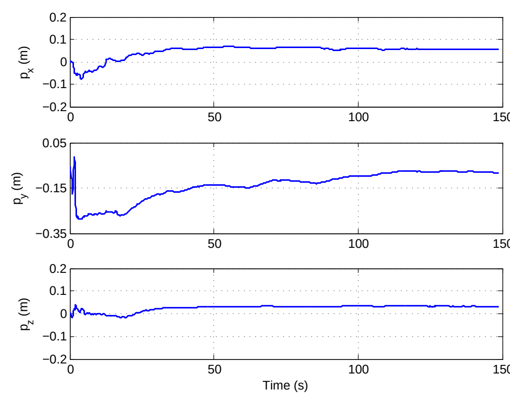
We analyzed the performance of the self-calibration algorithm using a dataset consisting of 14,905 IMU measurements (i.e., angular rates and linear accelerations) and 2,232 camera images, acquired over 150 seconds. A total of 44 SIFT features were tracked, with an average of 13.6 features visible per frame. The maximum rotation rate of the IMU was 75.3°/s, and the maximum linear acceleration (after accounting for gravity) was 11.6 m/s^2. Table I lists the initial hand-measured (HM) camera-IMU relative pose estimate and the final self-calibrated (SC) relative pose estimate. Plots for the time-evolution of the corresponding system states are shown in Figure 4.

Since ground truth measurements of the relative pose parameters were difficult to obtain, we instead evaluated the pixel reprojection errors for the hand-measured and self-calibrated relative pose estimates; the results are shown in Figure 5. We computed these residuals by running the UKF using the respective parameters listed in Table I, without estimating the parameters in the filter. To further emphasize the improvement produced by calibration, we also reduced the camera frame rate to 1.875 Hz (i.e., we processed every eighth image). For our hand-measured estimate, the RMS residual error was 12.29 pixels over all of the images; for the self-calibrated estimate, the RMS residual was 3.89 pixels. Note that the overall RMS residual for self-calibration is significantly lower (by more than a factor of three) than for the hand-measured value, indicating that the calibration is accurate. These results agree with earlier simulation studies.

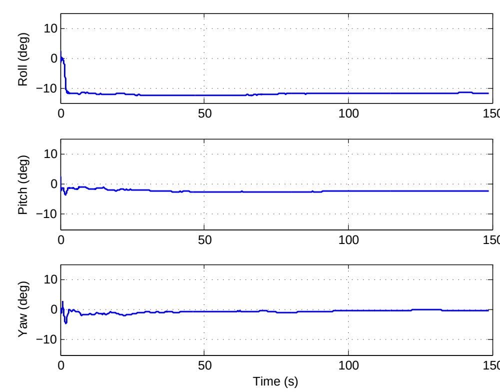
VII. CONCLUSIONS AND FUTURE WORK

This paper presented a relative pose self-calibration algorithm for omnidirectional visual and inertial sensors. As part of the calibration process we also localize the camera-IMU platform and build a sparse metric map of the environment. Our results show that it is possible to accurately calibrate the 6-DOF transform between the sensors *without* the need for a known calibration object. This is a step towards building power-on-and-go robotic systems that are able to self-calibrate during normal operation.

As part of our future work, we are exploring the possibility of self-calibrating the camera intrinsic parameters along with

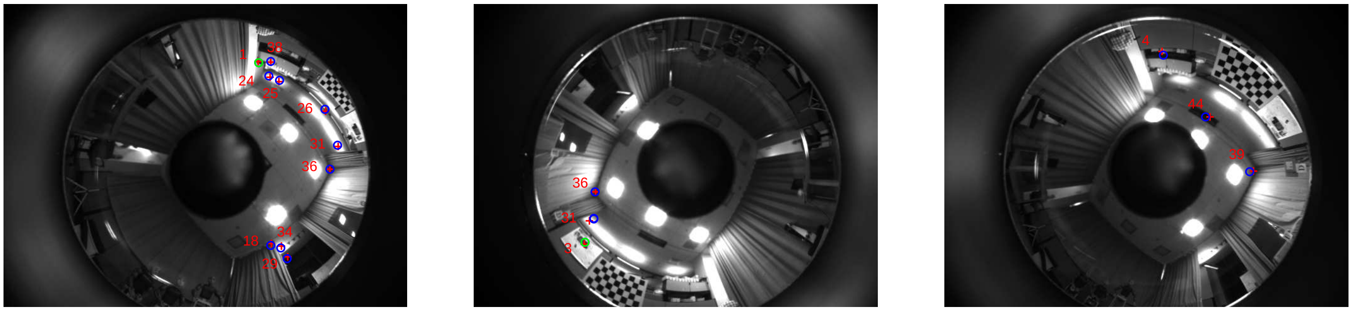


(a)



(b)

Fig. 4. (a) Evolution of the IMU-to-camera translation estimate over the calibration time interval (along the x, y and z axes of the IMU frame, from top to bottom) for the self-calibration procedure. (b) Evolution of the IMU-to-camera orientation estimate (for the roll, pitch and yaw angles that define the orientation of the camera frame relative to the IMU frame, from top to bottom) for the self-calibration procedure.



(a) t = 4.8 s (b) t = 69.6 s (c) t = 123.7 s

Fig. 3. Images from the self-calibration data set, at times t = 4.8 s, t = 69.6 s and t = 123.7 s from left to right, respectively. The measured image locations of the identified landmarks are shown as red crosses, and the predicted locations are shown as blue circles. Green circles identify anchor landmarks, which we used to fix the orientation of the world frame. The planar calibration target was used only to determine the initial and final positions of the camera.

the relative sensor pose. This would enable full camera-IMU self-calibration, and allow the sensors to operate together for significant durations without any manual recalibration.

REFERENCES

[1] R. Benosman and S. B. Kang, Eds., *Panoramic Vision*, 1st ed. Springer, May 2001.

[2] P. I. Corke, D. Strelow, and S. Singh, "Omnidirectional Visual Odometry for a Planetary Rover," in *Proc. IEEE/RSJ Int'l Conf. Intelligent Robots and Systems (IROS'04)*, vol. 4, Sendai, Japan, Sept. 2004, pp. 4007–4012.

[3] D. Strelow, "Motion Estimation from Image and Inertial Measurements," Ph.D. dissertation, Carnegie Mellon University, Pittsburgh, USA, Nov. 2004.

[4] 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.

[5] J. Kelly and G. S. Sukhatme, "Visual-Inertial Simultaneous Localization, Mapping and Sensor-to-Sensor Self-Calibration," in *Proc. IEEE Int'l Symp. Computational Intelligence in Robotics and Automation (CIRA'09)*, Daejeon, Korea, Dec. 2009, pp. 360–368.

[6] 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.

[7] J. D. Hol, T. B. Schon, and F. Gustafsson, "Relative Pose Calibration of a Spherical Camera and an IMU," in *Proc. 7th IEEE/ACM Int'l Symp. Mixed and Augmented Reality (ISMAR'08)*, Cambridge, United Kingdom, Sept. 2008, pp. 21–24.

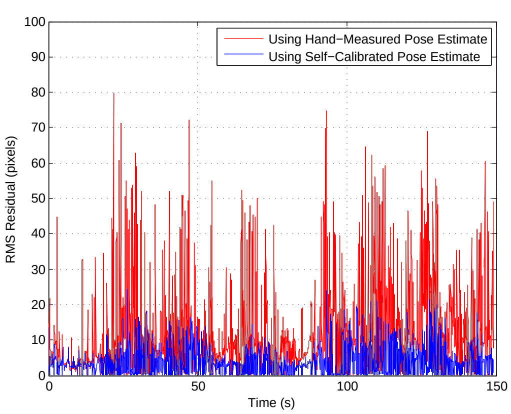


Fig. 5. Averaged RMS pixel reprojection errors (residuals) for all image frames, over 150 seconds. The residuals for the hand-measured relative pose estimate are shown in red, and the residuals for the self-calibrated estimate are shown in blue.

[8] F. M. Mirzaei and S. I. Roumeliotis, "A Kalman Filter-Based Algorithm for IMU-Camera Calibration: Observability Analysis and Performance Evaluation," *IEEE Trans. Robotics*, vol. 24, no. 5, pp. 1143–1156, Oct. 2008.

[9] E. Jones, A. Vedaldi, and S. Soatto, "Inertial Structure From Motion with Autocalibration," in *Proc. IEEE Int'l Conf. Computer Vision Workshop on Dynamical Vision*, Rio de Janeiro, Brazil, Oct. 2007.

[10] J. M. M. Montiel, J. Civera, and A. J. Davison, "Unified Inverse Depth Parametrization for Monocular SLAM," in *Proc. Robotics: Science and Systems (RSS'06)*, Philadelphia, USA, Aug. 2006.

[11] A. B. Chatfield, *Fundamentals of High Accuracy Inertial Navigation*, ser. Progress in Astronautics and Aeronautics, P. Zarchan, Ed. American Institute of Aeronautics and Astronautics, Sept. 1997, vol. 174.

[12] D. Scaramuzza, A. Martinelli, and R. Siegwart, "A Toolbox for Easily Calibrating Omnidirectional Cameras," in *Proc. IEEE/RSJ Int'l Conf. Intelligent Robots and Systems (IROS'06)*, Beijing, China, Oct. 2006, pp. 5695–5701.

[13] G. Conte, C. H. Moog, and A. M. Perdon, *Algebraic Methods for Nonlinear Control Systems*, 2nd ed. Springer, Dec. 2006.

[14] A. Chiuso, P. Favaro, H. Jin, and S. Soatto, "Structure from Motion Causally Integrated Over Time," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 24, no. 4, pp. 523–535, Apr. 2002.

[15] J. Kelly, "On the Observability and Self-Calibration of Visual-Inertial Navigation Systems," University of Southern California, Los Angeles, USA, Tech. Rep. CRES-08-005, Nov. 2008.

[16] R. Hermann and A. J. Krener, "Nonlinear Controllability and Observability," *IEEE Trans. Automatic Control*, vol. AC-22, no. 5, pp. 728–740, Oct. 1977.

[17] 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.

[18] S. J. Julier, "The Scaled Unscented Transform," in *Proc. IEEE American Control Conf. (ACC'02)*, vol. 6, Anchorage, USA, May 2002, pp. 4555–4559.

[19] S. J. Julier and J. J. L. Jr., "On Kalman Filtering With Nonlinear Equality Constraints," *IEEE Trans. Signal Processing*, vol. 55, no. 6, pp. 2774–2784, June 2007.

[20] J. L. Crassidis and F. L. Markely, "Unscented Filtering for Spacecraft Attitude Estimation," in *Proc. AIAA Guidance, Navigation and Control Conf. (GN&C'03)*, no. AIAA-2003-5484, Austin, USA, Aug. 2003.

[21] D. G. Lowe, "Distinctive Image Features from Scale-Invariant Keypoints," *Int'l J. Computer Vision*, vol. 2, no. 60, pp. 91–110, Nov. 2004.