Camera-IMU Calibration Using a Tilted Calibration Board

MARKUS KLEINERT[1], UWE STILLA[2]

Zusammenfassung: Für mobile Einsatzkräfte ist die Kenntnis der eigenen Position eine große Hilfe, wenn sie an unbekannten Orten zeitkritische oder gefährliche Aufgaben erledigen sollen. Eine automatische Bestimmung der eigenen Position ist jedoch schwierig, wenn Funksignale abgeschattet sind und gängige Verfahren der Satellitenortung somit nicht verwendet werden können. Um hier Abhilfe zu schaffen, wollen wir eine Kombination aus Inertialmesseinheit (IMU) und Kamera verwenden, um fortwährend eine Koppelnavigationslösung zu berechnen. Voraussetzung für diese Art der Sensordatenfusion ist die Kenntnis der relativen Lage (Translation und Rotation) der Sensorkoordinatensysteme. Diese kann beispielsweise durch Beobachtung der Punkte eines Kalibriermusters bestimmt werden. Dabei wird oft davon ausgegangen, dass das verwendete Kalibriermuster an der Richtung der Schwerkraft ausgerichtet ist, damit die notwendige Kompensation der Gravitation in den Beschleunigungsmessungen erfolgen kann. In unserer Arbeit wird ein solches Kalibrierverfahren vorgestellt, wobei im Unterschied zu gängigen Verfahren aus der Literatur nicht davon ausgegangen wird, dass das Kalibriermuster an der Gravitationsrichtung ausgerichtet ist. Stattdessen wird die Richtung der Gravitation bezüglich des Kalibriermusters in den Ausgleichungsprozess einbezogen. Weiterhin wird die Kalibrierung auch für ein erweitertes IMU-Fehlermodell durchgeführt.

1 Introduction

1.1 Envisaged application

Being able to orient themselves in unknown environments is crucial for first responders in order to successfully accomplish their task. Failing here has potentially severe consequences, e.g. firefighters could become trapped in closed parts of a building or rescuers do not arrive at the site of an accident in time in extended buildings. For these tasks it may already be helpful to see one's position and heading direction laid over an aerial image of the surrounding area because this enables the user to roughly infer his or her position within a building. If the location of the target, e.g. the site of an accident, is also marked on the map or aerial image this should noticeably facilitate the navigation task.

In open environments the determination of position and heading is usually easily achieved using a satellite navigation system, such as GPS, and possibly a compass. However, due to occlusions and multipath effects, satellite signals are usually not available or severely distorted in built environments. Likewise, soft iron effects disturb the heading estimates obtained with a compass near reinforced concrete, especially in buildings.

One possibility to overcome these problems is to build up radio beacons whose position can be determined accurately, as suggested in (MCCROSKEY *et al.* 2007). Using such beacons, the position of a pedestrian can be determined by at least four range measurements.

1) Markus Kleinert, Fraunhofer IOSB, Gutleuthausstr. 1, 76275 Ettlingen
2) Uwe Stilla, Technische Universität München, Arcisstr. 21, 80333 München

1.2 Camera-IMU sensor data fusion and calibration

The main drawback of the aforementioned method is that it requires a significant effort to build up the external infrastructure needed for positioning. Additionally, reinforced concrete may hamper the distribution of radio signals inside the building even if the beacons are located in windows or door openings. Therefore, it is of interest to develop alternative positioning techniques which do not rely on external infrastructure at all. To this end, we aim at exploiting the complementary sensor measurements of an inertial measurement unit (IMU) and a camera. While the camera as a bearings sensor provides a way to accurately estimate relative motion up to scale, the IMU measures acceleration and angular velocity which yield a short-time accurate estimate of displacement and scale when integrated over time.

In order to perform sensor data fusion of inertial and visual measurements, the relative 6-DOF transformation between the two sensor frames needs to be known. This problem is typically addressed by moving the sensor system in front of a planar calibration pattern. Based on the known locations of the points observed in the images, the sensor-to-sensor transformation as well as the pose of the sensor system can be estimated similar to what is done when calibrating a camera. For this purpose, inertial measurements are integrated to obtain estimates of relative motion between subsequent image exposures. To perform this integration, the direction of the gravity vector relative to the sensor frame needs to be known in order to compensate the effect of gravity on acceleration measurements. Therefore, most calibration methods rely on a checkerboard which is aligned with gravity. In this work we want to address this problem by augmenting the estimation problem such that the direction of the gravity vector in the checkerboard's frame is also estimated during calibration. This approach only requires an initial estimate of the direction of gravity and thereby facilitates lab calibrations. The relatively low-cost IMUs we deal with are usually subject to measurement bias and axis misalignment errors. Therefore, we also examine if the calibration can be improved by estimating these offsets too.

2 Related work

In one of the first articles on this topic, (LOBO & DIAS 2007) use a two-step approach to obtain the parameters describing the relative motion between a camera and an IMU. In order to obtain the leverarm, i.e. the translation vector, they rotate the system around perpendicular axes passing through the center of the IMU's coordinate frame while inferring the camera's motion from checkerboard observations. This enables the estimation of the component of the translation vector lying in the plane perpendicular to the rotation axis. Thus, rotations about two perpendicular directions suffice in principle to estimate the leverarm. For an independent estimate of the rotation between the two sensors the direction of acceleration is observed during standstill in the IMU's frame while the nadir direction is estimated in the camera's frame by observing the vanishing points of a checkerboard pattern which is lying on a ground plane whose normal direction is parallel to the direction of gravity. Repeating this process several times yields a set of corresponding direction vectors which are subsequently used to estimate the rotation. The main drawback of this method is that it seems to require a significant amount of user interaction during calibration. However it provides a way to compute initial values of relative

rotation and the leverarm for subsequent refinement if little is known about the sensor system, e.g., when the mechanical specification is not available.

A state-of-the-art Kalman-filter based approach to camera-IMU calibration is given in (MIRZAEI & ROUMELIOTIS 2008). To perform the calibration, the sensor system only has to be moved in front of a planar calibration grid which is aligned with the nadir direction. The authors present a non-linear observability (determinability) analysis to show that the state variables related to the transformation between the two sensors are completely observable if the sensor system is rotating about at least two perpendicular axes. In addition, it is shown that the uncertainty of estimated parameters decreases monotonically and that the observability does not depend on linear accelerations, i.e., it is sufficient to rotate the sensor system during calibration. The authors compare the results obtained with their extended Kalman filter (EKF) with the batch algorithm described in (MIRZAEI & ROUMELIOTIS 2007). The batch solution performs a Gauss-Newton optimization to simultaneously minimize reprojection error pertaining to the observations of checkerboard corners and pose-pose constraints created from inertial measurements. They report similar performance for both algorithms, although the batch solution seems to be slightly more accurate according to the presented results. This may be expected because during the batch optimization all relevant Jacobians are updated when state estimates improve and a new iteration of the estimation algorithm takes advantage of the improved Jacobians.

(HOL *et al.* 2010) present a hybrid estimation scheme where an EKF is used to estimate the system's motion given fixed values for the parameters of the calibration and additional nuisance parameters. In a second step these additional parameters are estimated using an iterative estimation procedure such as Gauss-Newton. Similar to the approach taken in this work, they estimate the direction of gravity as well as the IMU's biases as additional nuisance parameters.

A number of authors estimate the camera-IMU transformation parameters during operation of the system, i.e., along with the sensor pose required for navigation. This usually also requires simultaneous estimation of the location of observed features. Therefore this task is more challenging than the lab calibration. For instance, (JONES & SOATTO 2011) employ an EKF to simultaneously estimate the pose of the sensor system, the position of observed features, the direction of gravity w.r.t. the first camera frame, and the camera-IMU calibration parameters. They also discuss the observability for different kinds of trajectories by examining which configurations are not discernible based on sensor readings. In this way they arrive at the conclusion that all parameters are observable if the acceleration and the direction of the rotation axis are not constant. To alleviate this, they suggest switching between an autocalibration mode when the motion allows to estimate all parameters and a normal mode otherwise.

Similarly, (LUPTON & SUKKARIEH 2012) estimate the direction of gravity w.r.t. the first sensor frame, the camera-IMU calibration parameters and the position of tracked point features also in the first sensor frame using a sliding-window batch optimization (bundle adjustment) incorporating inertial and visual measurements. It is argued, that defining the datum by fixing the pose of the first camera and estimating the direction of gravity in that frame improves the linearization because the Jacobians are calculated with values close to the true trajectory in this case.

3 Problem formulation

After a review of relevant coordinate systems, we describe sensor measurement models relating the measurements to the system's motion and the parameters to be estimated. Based on these models, error terms are derived, which can be minimized using Gauss-Newton optimization. The relationship between the different parameters and the measurements are compactly expressed as a graphical model.

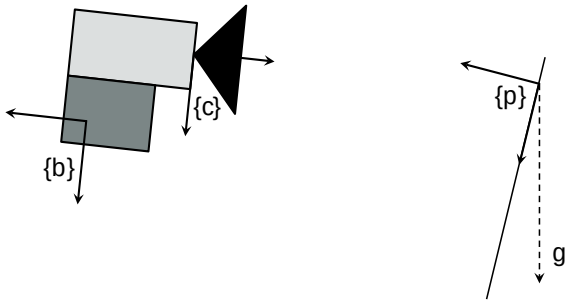

Fig. 1: Coordinate frames relevant for this work.

3.1 Coordinate frames

The relevant coordinate systems used in this work are shown in Fig. 1. As a bearings sensor, a camera measures directions relative to its coordinate frame {c}. Inertial measurements are obtained by the IMU in the body frame {b}. However, since IMUs usually consist of an assembly of many sensors, it cannot necessarily be expected that the measurement axes are orthogonal and equally scaled, as discussed in the following section. Finally, a coordinate system {p} is associated with the checkerboard pattern such that its axes are aligned with the pattern's rows and columns and its origin is coincident with the upper right corner. For the calibration task the six parameters describing the pose of the checkerboard pattern are fixed to define the datum, and the system's motion as well as the direction of gravity are estimated relative to this coordinate system. The main purpose of system calibration is to find the parameters of the rigid body transformation from frame {c} to frame {b} or vice versa.

3.2 IMU measurement model

An IMU measures acceleration and angular velocity in its own coordinate frame. These measurements are in general corrupted by systematic errors arising from sensor axis misalignment, additive bias, scale factor errors, and scale factor non-linearity, cf. (Farrell & Barth 1999). We do not attempt to estimate the 18 non-linearity coefficients for either sensor. From the datasheets provided by the manufacturer of one of the inertial sensors used in our work (Analog Devices ADIS16405), we conclude that axis misalignment and scale factor errors are negligible for the gyroscopes given the level of accuracy we expect to need. For the accelerometers however, the axis misalignment error in the order of 0.5° seems to be too severe to ignore. Therefore, two accelerometer error models were implemented: First, a simple model only taking white noise and bias into account. Second, a more elaborate model that also considers scale factor and misalignment errors. Thus, the accelerometer and angular rate measurements are related to their true values as follows:

$$\begin{aligned} {}^{m}a &= M^{-1}a + b_a + n_a \\ {}^{m}\omega &= \omega + b_\omega + n_\omega \end{aligned} \quad (1)$$

Here, the leading superscript m indicates measured values. The vector of true accelerations is denoted by a, while the angular rate vector is denoted by ω. The vectors b_a, b_ω contain the biases, which change in time according to a random walk process. n_a, n_ω are zero-mean, white Gaussian measurement noise terms.

The matrix M compensates scale factor and misalignment errors of the accelerometer sensor triad. For the simple acceleration measurement model M is a unit matrix of dimension three. Otherwise it is given by

$$M^{-1} = \begin{pmatrix} s_x & -m_{xz} & m_{xy} \\ m_{yz} & s_y & -m_{yx} \\ -m_{zy} & m_{zx} & s_z \end{pmatrix}, \quad (2)$$

where diagonal elements are scale factors and the remaining entries are misalignment parameters. Note that the misalignment parameters also take into account that the axes may not be orthogonal. Therefore, it is assumed that the six off-diagonal elements are independent, and M contains nine independent entries.

In order to utilize inertial measurements for displacement estimation, the effects of misalignment, scale factors, and biases have to be compensated to obtain an estimate of the true values whose integration yields an estimate of the sensor's displacement. This is achieved by the following equations:

$$\begin{aligned} \hat{a} &= \hat{M}({}^{m}a - \hat{b}_a) \\ \hat{\omega} &= {}^{m}\omega - \hat{b}_\omega \end{aligned} \quad (3)$$

Where the hat denotes estimated values. The integration is performed by the subsequent equations:

$$\begin{aligned} \hat{C}^{p}_{b,t+\tau} &= \hat{C}^{p}_{b,t} C(\hat{\omega}\tau) \\ {}^{p}\hat{a} &= \hat{C}^{p}_{b,t}\hat{a} + {}^{p}\hat{g} \\ \hat{p}_{t+\tau} &= \hat{p}_t + {}^{p}\hat{v}_t\tau + \frac{1}{2}\,{}^{p}\hat{a}\tau^2 \end{aligned} \quad (4)$$

In the above equations $C(\varphi)$ is the rotation matrix belonging to the Rodrigues vector φ, $C^{p}_{b,t}$ describes the rotation between the IMU and the pattern, ${}^{p}v$ is the IMU's velocity relative to the pattern, and τ is the time duration between samples. Eqs. (3)-(4) provide a simple first-order Euler integration for inertial measurements that can be used to calculate the sensor's motion between successive images. Note that the integration in Eq. (4) depends on the gravity estimate ${}^{p}\hat{g}$.

3.3 Image measurement model

The image measurement model describes how the known locations of checkerboard corners are related to the coordinates of their projections in the image plane. It consists of two parts: First, the projection, distortion, and sensor models, which are mainly determined by the kind of lens and the properties of the CCD-chip inside the camera (Hartley & Zisserman 2000). Second, the coordinate transformation from the calibration pattern to the camera's frame. Here it is assumed that the parameters of the projection model are known from a camera calibration procedure done

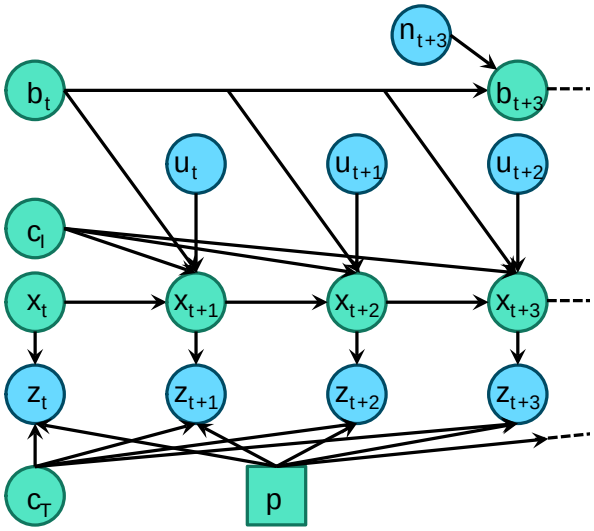

Fig. 2: Bayes network graph for the calibration problem.

beforehand. Since the specific properties of the projection model are not of importance for the calibration task at hand, the projection is just denoted by a function π and not considered further.
In this work the pose of the whole sensor system is identified with the IMU's pose. Therefore, the coordinate transformation between the pattern and the camera is a composition of the transformations from the pattern to the body (T_p^b) and from body to camera (T_b^c). Thus, the image measurement model directly depends on the sought camera-IMU calibration parameters. It is given by:

$$ {}^{m}z \quad = \quad \pi(T_b^c T_p^b \, {}^{p}X) + v \qquad (5) $$

Here, v is a vector with white, Gaussian noise terms and ${}^{p}X$ is the known position of a corner on the calibration pattern.

3.4 Graphical model formulation

The dynamic Bayesian network shown in Fig. 2 gives an overview of the relationship between the parameters in the model equations (green nodes) and the measurements (blue nodes). The rectangular node (p) contains the corner positions on the calibration pattern, which are fixed due to datum definition. They are connected to poses and the camera-IMU calibration (c_T) via measurements (z_t) according to Eqn. (5). Inertial measurements (u_t), bias estimates (b_t), and IMU calibration parameters (c_I) restrict the motion between successive frames according to Eqn. (4). The direction of gravity relative to the calibration pattern is also part of node (c_I). The bias changes according to a random walk process whose noise is given by (n_t). Note that the bias is assumed to be almost constant for a number of timesteps. Thus, new bias nodes are only introduced after time intervals whose duration depends on bias instability, a parameter obtained, e.g., from the manufacturer's datasheet.
Summing up all of the constraints contained in the above model gives rise to an error term, which depends on all the parameters to be estimated. This error term can be minimized iteratively using Gauss-Newton optimization.

Fig. 3: The camera-IMU system used in this work.

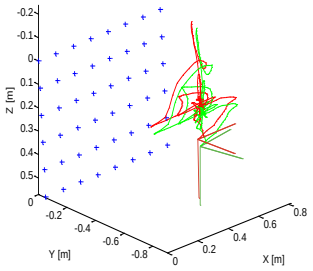

Fig. 4: Trajectory calculated from checkerboard observations (red) and smoothed trajectory (green).

4 Results

4.1 Experimental setup

This section presents preliminary results obtained with the calibration procedure described in the previous section. The system used for data acquisition is shown in Fig. 3. It comprises an ADIS16405 IMU and a BASLER SCOUT SCA-1400 industrial camera equipped with a 2.7 mm Fisheye lens. To obtain a synchronized data stream, the camera was triggered by the IMU (hardware trigger) with approximately 5 Hz.

During calibration, the camera was moved in front of a checkerboard calibration pattern, mainly performing rotations about at least two axes. The corners of the checkerboard were extracted and refined using the checkerboard detection and subpixel refinement algorithms provided by the publicly available OpenCV library. Zero-mean Gaussian noise of 2.0 pixels was assumed for the checkerboard corner locations extracted this way. To obtain noise parameter values for the IMU, the corresponding values from the IMU's datasheet were increased by approximately 25% to take account of modelling errors. Initial values to build up the model shown in Fig. 2 were obtained by first resectioning the exterior camera orientation using the observed checkerboard corners and their known position in space. In a second step these first estimates for exterior camera orientation and the initial estimates for the camera-IMU transformation were combined to obtain initial values for the IMU's pose and velocity. The initial estimates for the camera-IMU transformation were taken from the mechanical specifications. Gaussian noise with a standard deviation of 0.1 m and 15° for the relative translation and rotation of the two sensors was added to these values in order to evaluate how the solution depends on the parameters used for initialization.

Fig. 4 shows the IMU's trajectory before (red), and after (green) applying Gauss-Newton minimization for one dataset. The IMU's final coordinate frame is shown for both trajectories in a slightly darker color. Small blue crosses mark the location of checkerboard corners. Since the z-axis points in the estimated direction of gravity, the angle between the plane containing the corners and the z-axis corresponds to the estimated angle between the checkerboard and the direction of gravity around the long side of the checkerboard.

Tab. 1 presents the results obtained from 11 calibration runs using the extended error model, which includes accelerometer misalignment parameters. The rows contain the mean of all estimates ($\bar{v}$), the mean of the standard deviations estimated for the parameters ($\bar{\sigma}$), the empirical standard deviation (σ_e), and the mean deviation from the mechanical setup ($\overline{\Delta v}$) for the leverarm and the relative rotation error both given in the camera's frame. Additionally the mean estimated scale factors and mean misalignment parameters are shown with the corresponding uncertainty estimates (empirical standard deviation) in Tab. 2 as quantities without units.
Results for the simple accelerometer measurement model are shown in Tab. 3. In this case the misalignment matrix is simply the identity and thus not restated here.

Tab. 1: Calibration results (averaged) for the camera-IMU transformation parameters (extended IMU error model).

	Axis	$\bar{v}$	$\bar{\sigma}$	σ_e	$\overline{\Delta v}$
${}^{c}p_b$ [cm]	X	-0.64	0.18	0.29	-0.64
	Y	4.09	0.19	0.26	-0.91
	Z	-14,23	0.23	0.21	-0.67
${}^{c}\Psi_b$ [°]	X	1.25	0.033	0,08	1.25
	Y	-0.49	0.036	0,062	-0.49
	Z	0.05	0.066	0,093	0.05

Tab. 2: Calibration results (averaged) for accelerometer scale and misalignment parameters.

$$\widehat{M}^{-1} = \begin{bmatrix} 1.001 & -0.0004 & -0.0026 \\ 0.0037 & 1.003 & 0.0027 \\ -0.0003 & -0.0003 & 0.999 \end{bmatrix} \qquad \sigma_e(\widehat{M}^{-1}) = \begin{bmatrix} 0.0022 & 0.0016 & 0.0018 \\ 0.0021 & 0.0012 & 0.0027 \\ 0.0035 & 0.0035 & 0.0017 \end{bmatrix}$$

Tab. 3: Calibration results (averaged) for the camera-IMU transformation parameters (simple IMU error model).

	Axis	$\bar{v}$	$\bar{\sigma}$	σ_e	$\overline{\Delta v}$
${}^{c}p_b$ [cm]	X	-0.79	0.17	0.26	-0.79
	Y	3.86	0.18	0.25	-1.14
	Z	-14,3	0.23	0.2	-0.74
${}^{c}\Psi_b$ [°]	X	1.28	0.028	0.077	1.28
	Y	-0.43	0.027	0.076	-0.43
	Z	0.13	0.062	0.11	0.11

4.2 Discussion

For the extended as well as for the simple accelerometer error model, the empirical standard deviation calculated from estimated calibration parameters is considerably lower than the standard deviation of the noise that was added to the initial values at the start of each calibration run. This indicates that coarse initial estimates of the parameters can be improved by the calibration. The rotations about the X- and Y-axes of the camera's coordinate system and the translation in Y-direction differ from the values taken from the mechanical specifications by more than three standard deviations, when considering the empirical standard deviation.

Therefore, it can be expected to obtain better results by using the values obtained by the calibration procedure. This is not the case for the remaining parameters, whose estimates seem to be compatible with the values from the datasheets. Moreover, the estimation procedure is probably overconfident as can be seen by comparing the empirical standard deviations with the estimated ones (The estimated standard deviations displayed in Tab. 1 and Tab. 3 were not scaled by the mean of the estimated variance factor, which is 0.34 for the extended IMU error model and 0.35 for the simple error model). This suggests, that the estimation process is still biased, possibly due to unmodeled errors, e.g., arising from device synchronization problems.
When comparing the estimates obtained with the extended error model in Tab. 1 with the estimates for the simple model in Tab. 3, we note that they lie within one standard deviation of each other (again using the empirical standard deviation). Thus, we conclude that it is not possible to improve the calibration using the extended accelerometer error model with the procedure presented in this paper and the particular sensor system that was used in this work.

5 References

MCCROSKEY, R. ;SAMANANT, P.;HAWKINSON,W; HUSETH,S.&HARTMAN,R., 2010: GLANSER – An Emergency Responder Locator System for Indoor and GPS-Denied Applications. 23rd International Technical Meeting of the Satellite Division of the ION.

LOBO, J.&DIAS, J., 2007: Relative Pose Calibration Between Visual and Inertial Sensors. The International Journal of Robotics Research, Vol. 26, No. 6, pp. 561-575.

MIRZAEI, F.&ROUMELIOTIS, S., 2008:A Kalman Filter-Based Algorithm for IMU-Camera Calibration: Observability Analysis and Performance Evaluation. IEEE Transactions on robotics, Vol. 24, No. 5, pp. 1143-1156.

MIRZAEI, F.&ROUMELIOTIS, S., 2007: IMU-camera calibration: Bundle adjustment. Dept. Comput. Sci. & Eng., Univ. Minnesota, Tech. Rep.

HOL, J.;SCHÖN,T.&GUSTAFSSON,F., 2010: Modeling and Calibration of Inertial and Vision Sensors. International Journal of Robotics Research, Vol. 29, No. 2-3, pp. 231-244.

JONES, E.&SOATTO,S., 2011:Visual-inertial navigation, mapping and localization: A scalable real-time causal approach. International Journal of Robotics Research, Vol. 30, No. 4, pp. 407-430.

LUPTON, T.&SUKKARIEH,S., 2012:Visual-Inertial-Aided Navigation For High-Dynamic Motion in Built Environments Without Initial Conditions. IEEE Transactions on robotics, Vol. 28, No. 1, pp. 61-76.

FARRELL, J.&BARTH,M., 1999:The Global Positioning System & Inertial Navigation. McGraw-Hill

HARTLEY, R.&ZISSERMAN,A., 2000:Multiple View Geometry in Computer Vision. Cambridge University Press