The EuRoC micro aerial vehicle datasets

Article *in* The International Journal of Robotics Research · January 2016

DOI: 10.1177/0278364915620033

8 authors, including:

Michael Burri
ETH Zurich
30 PUBLICATIONS **1,976** CITATIONS
SEE PROFILE

Pascal Gohl
Hexagon Geosystems
9 PUBLICATIONS **822** CITATIONS
SEE PROFILE

Thomas Schneider
ETH Zurich
20 PUBLICATIONS **800** CITATIONS
SEE PROFILE

Markus Wilhelm Achtelik
ETH Zurich
34 PUBLICATIONS **2,930** CITATIONS
SEE PROFILE

The EuRoC MAV Datasets

Michael Burri, Janosch Nikolic, Pascal Gohl, Thomas Schneider,
Joern Rehder, Sammy Omari, Markus W. Achtelik and Roland Siegwart*

January 29, 2016

Abstract

This paper presents visual-inertial datasets collected on-board a Micro Aerial Vehicle (MAV). The datasets contain synchronized stereo images, IMU measurements, and accurate ground truth. The first batch of datasets facilitates the design and evaluation of visual-inertial localization algorithms on real flight data. It was collected in an industrial environment and contains millimeter accurate position ground truth from a laser tracking system. The second batch of datasets is aimed at precise 3 D environment reconstruction and was recorded in a room equipped with a motion capture system. The datasets contain 6 D pose ground truth and a detailed 3 D scan of the environment. Eleven datasets are provided in total, ranging from slow flights under good visual conditions to dynamic flights with motion blur and poor illumination, enabling researchers to thoroughly test and evaluate their algorithms. All datasets contain raw sensor measurements, spatio-temporally aligned sensor data and ground truth, extrinsic and intrinsic calibrations, and datasets for custom calibrations.

*The research leading to these results has received funding from the European Community's Seventh Framework Programme (FP7) under grant-agreement n.608849 (EuRoC). Michael Burri, Janoch Nikolic and Helen Oleynikova are PhD students, Markus Achtelik is a senior researcher at the Autonomous Systems Lab (ASL), and Roland Siegwart is a full professor at the ETH Zurich and head of the ASL. (email:{michael.burri, janosch.nikolic, helen.oleynikova, markus.achtelik}@mavt.ethz.ch, r.siegwart@ieee.org).

1 Introduction

The datasets presented in this paper were recorded in the context of the European Robotics Challenge (EuRoC)[1], to assess the contestant's visual-inertial SLAM and 3 D reconstruction capabilities on MAVs. It has been tested by more than 20 teams and was used for evaluation in several publications (Marzat et al., 2015; Ait-Jellal and Zell, 2015; Fu et al., 2015; Oleynikova et al., 2015).

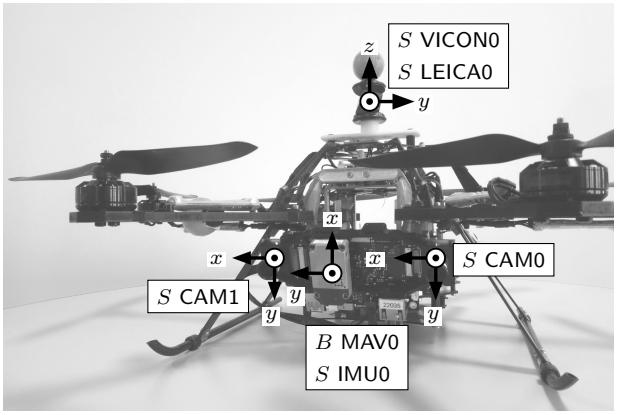


Figure 1: The AscTec "Firefly" hex-rotor helicopter used for dataset collection. A visual-inertial sensor unit provided stereo images and data from an inertial measurement unit. Position ground truth was recorded with a Leica total station, and 6 D pose ground truth with a Vicon system. The body frame is defined at the IMU sensor frame.

The datasets were recorded on-board an As-

[1] http://www.euroc-project.eu/

cTec Firefly hex-rotor helicopter, using a front-down looking stereo camera, hardware-synchronized with an IMU (see Figure 1). The content of the datasets ranges from slow flights in a small, cluttered workspace to dynamic flights in a large industrial machine hall. This allows for testing and tuning under various conditions.

Two types of datasets are provided: the first contains 3 D position ground truth from a Leica Multistation and focuses on evaluation of visual-inertial SLAM algorithms in a realistic industrial scenario. It was recorded in a machine hall at ETH Zürich. The environment was unstructured and cluttered, which renders the dataset challenging to process.

The second type of datasets contains 6 D pose ground truth from a Vicon motion capture system and an accurate 3 D point cloud of the environment (captured with a Leica 3 D laser scanner, see Figure 5). Obstacles were placed in the room to render the reconstruction more challenging and to provide more visual texture.

Airborne inspection of industrial facilities marks a promising field for applications of MAVs, and this dataset is geared towards assessing two key capabilities enabling the deployment of MAVs in this use-case: accurate motion estimation and 3 D reconstruction. Rigid mechanical design and exposure-compensated synchronization ensure optimal sensor data. At the same time, accurate spatio-temporal alignment of the sensor data to ground truth allows for rigorous benchmarking. It thus enables users to incorporate quantitative assessments into the development of novel approaches to visual-inertial SLAM and 3 D reconstruction pipelines. To our knowledge, this work is unique in its scope and the only publicly available dataset that provides both motion and structure ground truth recorded for an airborne platform.

The paper is organized as follows: first, the sensor setup, coordinate frame conventions, and the notation are introduced. Then, a detailed description of the two different types of datasets is given, followed by the documentation of the format in which the data is stored and how to access it. Finally, we describe the methods used to align sensor data and ground truth measurements spatially and temporally, how the sensors were calibrated and some known issues of the datasets.

2 Sensor Setup

An AscTec Firefly MAV[2], equipped with a visual-inertial sensor unit (Nikolic et al., 2014), was used for data collection. The VI sensor unit was mounted in a front-down looking position, which provided an unobstructed stereo view on the environment. The images of the two global-shutter, monochrome cameras were logged at a rate of 20 Hz. Angular rates and specific force measurements from the IMU were logged with 200 Hz. Note that this is the IMU of the visual-inertial sensor unit and not the IMU of the MAV autopilot. IMU and cameras were hardware time-synchronized such that the middle of the exposure aligned with the IMU measurements.

For ground truth, two different systems were used. A Leica Nova MS50[3] laser tracker measured the position of a prism, which was mounted on top of the MAV. The measurements were millimeter accurate, and available at a rate of approximately 20 Hz. For some of the datasets, a Vicon motion capture system[4] provided 6 D pose measurements of a coordinate frame, which was defined by a set of reflective markers mounted on the MAV. The pose measurements were recorded at a rate of 100 Hz.

Visual and inertial data was logged and timestamped on-board the MAV, while ground truth was logged on the base station. A maximum likelihood estimator was used to align the data temporally, and to calibrate the position of the ground truth body coordinate frames with respect to the sensor unit. Both raw and spatio-temporally aligned data is provided with the datasets.

Figure 2 depicts the sensors and the transformations that link them. Table 1 provides an overview of the sensors and reference systems for which measurements are present in the datasets.

[2] http://www.asctec.de/en/uav-uas-drone-products/asctec-firefly

[3] http://www.leica-geosystems.com/en/Leica-Nova-MS50_103592.htm

[4] http://www.vicon.com/products/camera-systems/

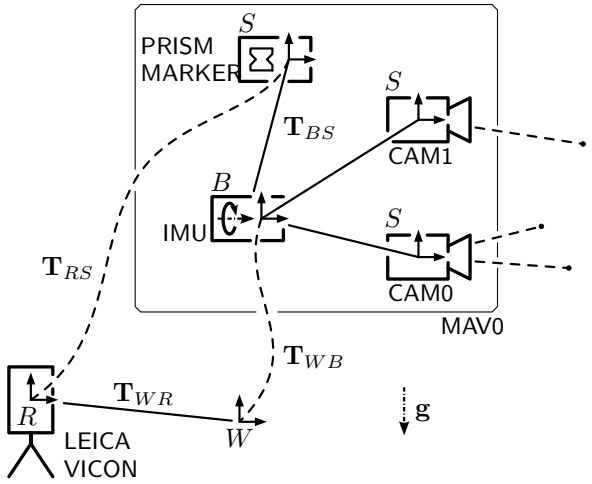


Figure 2: The sensor system that was used to capture the datasets. Each sensor reports the measurements in its own reference frame S. Raw data from ground truth instruments is reported with respect to the respective prism or marker coordinate frames. A calibration for all extrinsic parameters linking the sensors to the body frame B and intrinsic parameters are provided with the datasets. The body frame is defined to be aligned with the IMU sensor frame.

Figure 3: Representative image of the machine hall, where the first batch of datasets was collected.

Table 1: Sensors and ground truth instruments.

Sensor	**Type**	**Rate**	**Characteristics**
Cameras	MT9V034	2×20 Hz	WVGA, global shutter
IMU	ADIS16448	200 Hz	MEMS, intr. calibrated
Position	Leica MS50	20 Hz	Accuracy: ≈ 1 mm
Pose	Vicon	100 Hz	6 D
Structure	Leica MS50	–	Accuracy: ≈ 1 mm

3 Datasets

Two types of datasets are provided: The first batch of datasets was recorded in a large machine hall shown in Figure 3 and aims at testing visual-inertial motion estimation algorithms or SLAM frameworks. For ground truth, a 3 D position was provided by a laser tracker. A second batch of datasets was recorded in our vicon room equipped with a motion capture system, whith an approximate size of $8\,\text{m} \times 8.4\,\text{m} \times 4\,\text{m}$. For these datasets, additional reference point clouds as shown in Figure 5 are available that enable the assessment of multi-view reconstruction approaches. All datasets were recorded using an AscTec Firefly MAV, equipped with a VI-Sensor as shown in Figure 1 and a short summary of the trajectories is given in Table 2.

3.1 Industrial Machine Hall

The machine hall at ETH, shown in Figure 3, represents a challenging industrial environment for SLAM. Five different datasets were recorded. The datasets are increasingly difficult to process in terms of flight dynamics and lighting conditions. To get an impression of the trajectories, flight paths of each difficulty level are shown in Figure 4. Ground truth position measurements were provided by a Leica Nova MS50 laser tracker and recorded on the base station. For facilitated use of the datasets, we provide an estimate

`bonita`

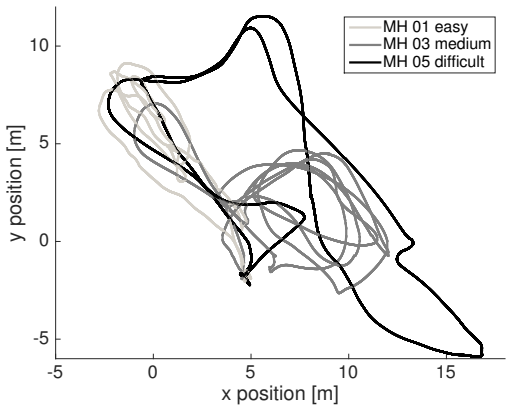


Figure 4: Top down view of three flight trajectories, recorded in the machine hall. The complexity of the machine hall datasets varies in terms of trajectory length, flight dynamics, and illumination conditions.

Table 2: Dataset characteristics

Name	**Length / Duration**	**Avg. Vel. / Angular Vel.**	**Note**
MH_01_easy	80.6 m 182 s	0.44 m s^{-1} 0.22 rad s^{-1}	good texture, bright scene
MH_02_easy	73.5 m 150 s	0.49 m s^{-1} 0.21 rad s^{-1}	good texture, bright scene
MH_03_medium	130.9 m 132 s	0.99 m s^{-1} 0.29 rad s^{-1}	fast motion, bright scene
MH_04_difficult	91.7 m 99 s	0.93 m s^{-1} 0.24 rad s^{-1}	fast motion, dark scene
MH_05_difficult	97.6 m 111 s	0.88 m s^{-1} 0.21 rad s^{-1}	fast motion, dark scene
V1_01_easy	58.6 m 144 s	0.41 m s^{-1} 0.28 rad s^{-1}	slow motion, bright scene
V1_02_medium	75.9 m 83.5 s	0.91 m s^{-1} 0.56 rad s^{-1}	fast motion, bright scene
V1_03_difficult	79.0 m 105 s	0.75 m s^{-1} 0.62 rad s^{-1}	fast motion, motion blur
V2_01_easy	36.5 m 112 s	0.33 m s^{-1} 0.28 rad s^{-1}	slow motion, bright scene
V2_02_medium	83.2 m 115 s	0.72 m s^{-1} 0.59 rad s^{-1}	fast motion, bright scene
V2_03_difficult	86.1 m 115 s	0.75 m s^{-1} 0.66 rad s^{-1}	fast motion, motion blur

of the full pose (i.e. including attitude) at the IMU sampling rate, as explained in the calibration section. To aid the initialization of monocular SLAM frameworks, sufficient motion was present at the beginning of each dataset.

3.2 Vicon Room

The second batch of datasets aims at evaluating multi-view reconstruction performance. Two different scenarios were prepared in the Vicon room, with different obstacle configurations. As an additional challenge there are some moving curtains visible in some parts of the dataset. Each scenario consists of three datasets with increasing complexity.

The 3 D point clouds were recorded using the Leica Nova MS50 laser scanner functionality from multiple locations, providing ground truth of the structure with millimeter accuracy. Figure 5 shows a visualization of the point cloud. Similar to the machine hall datasets, an estimate of the full pose of the MAV is provided at the IMU sampling rate as explained in the calibration section.

4 Dataset Format

This section specifies the format and conventions in which sensor data, ground truth, and calibration parameters are reported.

4.1 Conventions and Notation

All sensors are assumed to be rigidly attached to a "sensor system", which we denote as the "body frame" B, see Figure 2. B is moving with respect to a quasi inertial "world" frame W. All sensor systems share the same world frame, i.e. it is unique per dataset.

Raw sensor data is expressed in each sensor's reference frame S. The transformation between S and B (the "extrinsics") is reported in the sensor's yaml file. If a sensor measures with respect to an external reference frame R, for example a Vicon tracking system, its raw data is reported with respect to this frame.

Figure 5: Ground truth point cloud of the Vicon room in configuration 1 (top) and 2 (bottom). The size of the room is 8 m × 8.4 m × 4 m. Point cloud data was recorded with a Leica Multistation, fusing full 3 D scans from seven different positions.

The reference frame is introduced to support multiple ground truth systems, and the transformation between R and W is provided in the sensor's yaml file. In many cases, R and W coincide. For improved usability of the datasets, post-processed ground truth data is provided that refers directly to the state of B with respect to W.

Units and Timestamps All measurements are reported in SI units. The only exception are timestamps, which are in integer nanoseconds POSIX (i.e. nanoseconds since 1. January 1970, 00:00 UTC).

4.1.1 Rotations and Transformations

Rotations are represented as Hamiltonian unit quaternions $\mathbf{q}$. The quaternions are written as:

$$\mathbf{q} = \begin{bmatrix} q_w & q_x & q_y & q_z \end{bmatrix}^T = \begin{bmatrix} q_w \\ \overline{\mathbf{q}} \end{bmatrix}, \tag{1}$$

where q_w denotes the real and $\overline{\mathbf{q}}$ the imaginary part of $\mathbf{q}$.

The direction cosine matrix $\mathbf{C}_{WB}(\mathbf{q}_{WB})$, which transforms vectors from frame B to frame W, can be computed from $\mathbf{q}$ as

$$\mathbf{C}(\mathbf{q}) = q_w^2 \mathbf{I}_{3\times 3} + 2q_w \lfloor \overline{\mathbf{q}} \times \rfloor + \lfloor \overline{\mathbf{q}} \times \rfloor^2 + \mathbf{q}\mathbf{q}^T, \tag{2}$$

where $\lfloor \cdot \times \rfloor$ denotes the skew-symmetric operator

$$\lfloor \mathbf{r} \times \rfloor = \begin{bmatrix} 0 & -r_2 & r_1 \\ r_2 & 0 & -r_0 \\ -r_1 & r_0 & 0 \end{bmatrix}. \tag{3}$$

A vector $\overrightarrow{SL}$ expressed in B, ${}_B\mathbf{p}_{SL}$, is expressed in W as follows:

$${}_W\mathbf{p}_{SL} = \mathbf{C}_{WB}(\mathbf{q}_{WB}) {}_B\mathbf{p}_{SL}. \tag{4}$$

The 4×4 homogeneous transformation matrix $\mathbf{T}_{WB}$ is defined as

$$\mathbf{T}_{WB} = \begin{bmatrix} \mathbf{C}_{WB} & {}_W\mathbf{p}_{WB} \\ \mathbf{0} & 1 \end{bmatrix}. \tag{5}$$

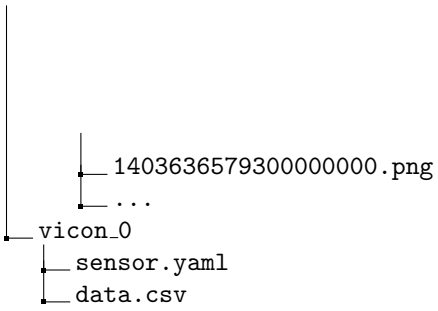
4.2 Data Format

Each dataset contains one or more "sensor systems" (one MAV in our case), and each sensor system can contain several "sensors". Every sensor comes with a `sensor.yaml` file that specifies its calibration parameters and a `data.csv` file that either contains the sensor measurements or points to data files in the optional `data` folder:

```
<sensor_system_id>
└─ <sensor_id>
   ├─ sensor.yaml
   ├─ data.csv
   └─ data
```

A ground truth source, raw or post-processed, is treated like any other sensor. For example:

```
mav_0
├─ imu_0
│  ├─ sensor.yaml
│  └─ data.csv
└─ cam_0
   ├─ sensor.yaml
   ├─ data.csv
   └─ data
      └─ 1403636579250000000.png
```



4.2.1 Yaml Files

Every `sensor.yaml` file contains a `sensor_type` field, where `sensor_type` is a unique identifier {`imu|camera|position|pose|pointcloud|visual-inertial`}. A 4×4 homogeneous transformation matrix `T_BS` specifies the pose of the sensor with respect to the sensor system, the "extrinsics". Further type-specific entries in the yaml files are listed in the description of the individual sensors.

4.2.2 CSV Data Files

The first row of the Comma Separated Values (CSV) data files contains the field names followed by the corresponding SI units in squared brackets. The subsequent rows contain the sensor data. For example:

```
#timestamp [ns], p_RS_R_x [m], ...   CR LF
1403636579250000000, -1.786008, ... CR LF
1403636579260000000, -1.785634, ... CR LF
```

4.3 Sensors

This section specifies the contents of the yaml and CSV data files for each sensor type.

4.3.1 Camera

The `data.csv` file of a camera lists timestamps, followed by the file name of the corresponding camera image. The images are stored in the `data` folder in a lossless compression format (PNG).

As with any other sensor, a camera's yaml file contains the transform between the sensor and the sensor system, commonly denoted as the extrinsics. Additionally, camera yaml files contain the following fields related to the camera intrinsic calibration parameters: `camera_model`, `intrinsic_coefficients`, `distortion_model`, `distortion_coefficients`, `resolution`. The reported parameters correspond to coefficients K_1, K_2, P_2, P_1 and K_3 in the distortion model proposed by Brown (1971).

4.3.2 IMU

The `data.csv` file of a type `imu` sensor contains the following fields (in this order): timestamp, $\boldsymbol{\omega}_S$ [rad s^{-1}], $\mathbf{a}_S$ [m s^{-2}]. $\boldsymbol{\omega}_S$ denotes the angular rate measurements (i.e. the gyroscope's x, y, and z readings, in this order), and $\mathbf{a}_S$ denotes the specific force measurements (i.e. the accelerometer readings), both expressed in the sensor's frame of reference.

The IMU's yaml file contains the gyroscope and accelerometer noise model parameters. The parameters relate to a standard inertial sensor noise model as used for example in (Leutenegger et al., 2015). The noise processes that corrupt the rate and acceleration measurements can be written as follows:

$$\mathbf{n}_g(t) = \mathbf{b}_g(t) + \mathbf{w}_g(t) \tag{6a}$$
$$\mathbf{n}_a(t) = \mathbf{b}_a(t) + \mathbf{w}_a(t), \tag{6b}$$

where $\mathbf{n}_g$ and $\mathbf{n}_a \in \mathbb{R}^3$ denote the gyroscope and accelerometer noise processes. $\mathbf{w}_g$ and $\mathbf{w}_a$ denote continuous–time, white Gaussian noise processes of strength σ_g and σ_a,

$$\mathrm{E}[\mathbf{w}_g(t_1)\mathbf{w}_g(t_2)] = \sigma_g^2 \mathbf{I}\delta(t_1 - t_2) \tag{7a}$$
$$\mathrm{E}[\mathbf{w}_a(t_1)\mathbf{w}_a(t_2)] = \sigma_a^2 \mathbf{I}\delta(t_1 - t_2), \tag{7b}$$

where $\delta(\cdot)$ denotes the Dirac delta function. $\mathbf{b}_g(t)$ and $\mathbf{b}_a(t)$ denote the slowly varying bias processes, with

$$\dot{\mathbf{b}}_g = \mathbf{w}_{bg} \tag{8a}$$
$$\dot{\mathbf{b}}_a = \mathbf{w}_{ba}, \tag{8b}$$

where $\mathbf{w}_{bg}$ and $\mathbf{w}_{ba}$ are white noise processes of strength σ_{bg} and σ_{ba}, the bias "diffusions". In other words, gyroscope and accelerometer noise is modelled as a combination of white noise and a standard random walk.

The noise model parameters are specified in the corresponding yaml file as follows:

σ_g	`gyroscope_noise_density`
σ_{bg}	`gyroscope_random_walk`
σ_a	`accelerometer_noise_density`
σ_{ba}	`accelerometer_random_walk`.

This model reflects only the stochastic errors in the inertial data. Its parameters were obtained from the IMU at rest. It fails to capture deterministic errors which appear in flight due to residual scale factor errors and axes misalignment. Even though the IMU is intrinsically calibrated, a small increase in these parameters may therefore be beneficial for post–processing the data.

4.3.3 Position, Pose

The `data.csv` file of a type `pose` sensor contains the following fields (in this order): timestamp, ${}_R\mathbf{p}_{RS}$, $\mathbf{q}_{RS}$. ${}_R\mathbf{p}_{RS}$ denotes the 3 D position estimate of the sensor with respect to the reference frame R, expressed in R. $\mathbf{q}_{RS}$ lists the four elements of the active unit quaternion corresponding to the orientation estimate of the sensor.

The corresponding `sensor.yaml` file contains the additional field `T_WR`, a 4×4 homogeneous transformation matrix that describes the pose of the sensor's external reference frame R with respect to the "world" frame W.

The same format is used for a type `position` sensor, but without orientation measurements. The orientation extrinsics are set to identity.

4.3.4 Pointcloud

3 D point clouds are provided in the Polygon File Format (PLY, ASCII), rather than in a custom file format, to facilitate efficient processing. Hence for point clouds, `data.csv` is replaced with a `data.ply` file. The points are expressed in the reference frame R, and `T_WR` is reported in the yaml file.

5 Calibration and Synchronization

The datasets contain raw data, extrinsic and intrinsic calibration parameters, and spatio-temporally aligned ground truth. This section briefly outlines how the systems were calibrated and synchronized.

5.1 Visual-Inertial Sensor Unit

The visual-inertial sensor unit was calibrated with *Kalibr*[5] (Furgale et al., 2013) prior to dataset collection. The calibration included the intrinsics of the cameras and the camera-IMU extrinsics. The corresponding calibration datasets (containing checkerboard and AprilTag (Olson, 2011) grid sequences) are provided with the datasets. This data enables proprietary calibration approaches to be used with our datasets, given that these employ standard visual targets.

5.2 Ground-Truth Alignment

To provide a useful ground truth, the measurements from the tracking system (Leica laser tracker and Vicon motion capture system) were spatially and temporally aligned with the sensor system (i.e. the MAV's body or IMU frame B). This required:

- Calibration of the tracking system's reference frame S (i.e. the position of the Leica prism or the Vicon marker frame) with respect to B.
- Time synchronization between ground truth and the sensor system.

To determine these parameters, we used a classical maximum likelihood (ML) state estimator (Maybeck, 1979). Our formulation is similar to (Li and Mourikis, 2014), but we employed a batch estimator in an offline procedure to obtain the full ML solution. The smoother incorporated all ground truth and IMU measurements to estimate the following states:

$$\mathbf{x} = \begin{bmatrix} \mathbf{q}_B & \mathbf{p}_B & \mathbf{v}_B & \mathbf{b}_g & \mathbf{b}_a & \boldsymbol{\theta} \end{bmatrix}^T, \tag{9}$$

with

$$\boldsymbol{\theta} = \begin{bmatrix} \mathbf{q}_S & \mathbf{p}_S & \Delta t_S \end{bmatrix}. \tag{10}$$

$\mathbf{q}_B$ denotes the attitude, $\mathbf{p}_B$ the position, and $\mathbf{v}_B$ the velocity of the body B with respect to the earth-fixed reference frame W. $\mathbf{b}_g$ and $\mathbf{b}_a$ denote the gyroscope and accelerometer biases, modelled as Wiener processes as it is standard in the literature. $\mathbf{q}_S$ and $\mathbf{p}_S$ denote the unknown transform between the ground

[5] `github.com/ethz-asl/kalibr`.

truth reference frame S and B. For the Leica laser tracker, only the translation $\mathbf{p}_S$ was estimated. Δt_S denotes the time-varying temporal offset between the sensor and the ground truth systems.

Angular rate and acceleration measurement errors were modelled as the sum of the bias processes and white noise. Errors in the Leica and Vicon measurements were modelled as white.

6 Known Issues

Despite careful design and execution of the data collection experiments, we are aware of different issues which pose additional challenges for processing and limit achievable accuracy when comparing to ground-truth.

These are

- The visual-inertial sensor employs an automatic exposure control that is independent for both cameras. This resulted in different shutter times and in turn in different image brightnesses, rendering stereo matching and feature tracking more challenging. Since the mid-exposure times of both cameras were temporally aligned, synchronization was not affected by different shutter times.

- Some of the datasets exhibit very dynamic motions, which are known to deteriorate the measurement accuracy of the laser tracking device. The numbers reported by the manufacturer may be overly optimistic for these events, which complicates the interpretation of ground truth comparisons for highly accurate visual odometry approaches. The effect on sections with less dynamic motion—particularly the start and end of each dataset—is assumed to be negligible.

- The accuracy of the synchronization between sensor data and motion ground truth is limited by the fact that both sources were recorded on different systems and that device timestamps were unavailable for the Vicon system. This issue is mitigated by estimating the temporal offset Δt_S as a state, addressing both a fixed temporal offset and clock drift. Where available, device timestamps were employed to avoid jitter. Raw data is available for assessing alternative synchronization schemes.

7 Data Access Methods

The datasets can be downloaded from http://projects.asl.ethz.ch/datasets/doku.php?id=kmavvisualinertialdatasets. Furthermore, some example tools to load and plot the data in MATLAB are provided and can be used as a starting point for implementations in other languages.

References

Radouane Ait-Jellal and Andreas Zell. A fast dense stereo matching algorithm with an application to 3d occupancy mapping using quadrocopters. In *Advanced Robotics (ICAR), 2015 International Conference on*, pages 587–592. IEEE, 2015.

Duane C. Brown. Close-range camera calibration. *Photogrammetric Engineering & Remote Sensing*, 37:855–866, 1971.

Changhong Fu, Adrian Carrio, and Pascual Campoy. Efficient visual odometry and mapping for unmanned aerial vehicle using arm-based stereo vision pre-processing system. In *Unmanned Aircraft Systems (ICUAS), 2015 International Conference on*, pages 957–962. IEEE, 2015.

P. Furgale, J. Rehder, and R. Siegwart. Unified temporal and spatial calibration for multi-sensor systems. In *Intelligent Robots and Systems (IROS), 2013 IEEE/RSJ International Conference on*, pages 1280–1286, Nov 2013. doi: 10.1109/IROS.2013.6696514.

Stefan Leutenegger, Simon Lynen, Michael Bosse, Roland Siegwart, and Paul Furgale. Keyframe-based visualinertial odometry using nonlinear optimization. *The International Journal of Robotics*

Research, 34(3):314–334, 2015. doi: 10.1177/0278364914554813.

Mingyang Li and Anastasios I Mourikis. Online temporal calibration for camera–imu systems: Theory and algorithms. *The International Journal of Robotics Research*, 33(7):947–964, 2014.

Julien Marzat, Julien Moras, Aurélien Plyer, Alexandre Eudes, and Pascal Morin. Vision-based localization, mapping and control for autonomous mav: Euroc challenge results. In *15th ONERA-DLR Aerospace Symposium (ODAS 2015)*, 2015.

Peter S. Maybeck. *Stochastic models, estimation, and control*. Academic Press, New York, 1979.

Janosch Nikolic, Joern Rehder, Michael Burri, Pascal Gohl, Stefan Leutenegger, Paul T Furgale, and Roland Siegwart. A synchronized visual-inertial sensor system with fpga pre-processing for accurate real-time slam. In *Robotics and Automation (ICRA), 2014 IEEE International Conference on*, pages 431–437. IEEE, 2014.

Helen Oleynikova, Michael Burri, Simon Lynen, and Roland Siegwart. Real-time visual-inertial localization for aerial and ground robots. In *Intelligent Robots and Systems (IROS), 2015 IEEE/RSJ International Conference on*, pages 3079–3085. IEEE, 2015.

Edwin Olson. AprilTag: A robust and flexible visual fiducial system. In *Proceedings of the IEEE International Conference on Robotics and Automation (ICRA)*, pages 3400–3407. IEEE, May 2011.