


# The Zurich Urban Micro Aerial Vehicle Dataset

Journal Title  
XX(X):1-5  
©The Author(s) 2016  
Reprints and permission:  
sagepub.co.uk/journalsPermissions.nav  
DOI: 10.1177/ToBeAssigned  
www.sagepub.com/  


András L. Majdik<sup>1</sup> Charles Till<sup>2</sup> and Davide Scaramuzza<sup>2</sup>

## Abstract

This paper presents a dataset recorded on-board a camera-equipped Micro Aerial Vehicle (MAV) flying within the urban streets of Zurich, Switzerland, at low altitudes (i.e., 5-15 meters above the ground). The 2 km dataset consists of time synchronized aerial high-resolution images, GPS and IMU sensor data, ground-level street view images, and ground truth data. The dataset is ideal to evaluate and benchmark appearance-based localization, monocular visual odometry, simultaneous localization and mapping (SLAM), and online 3D reconstruction algorithms for MAVs in urban environments.

## Keywords

visual localization, air-ground matching, aerial robotics

## Supplementary material

The dataset is available at:

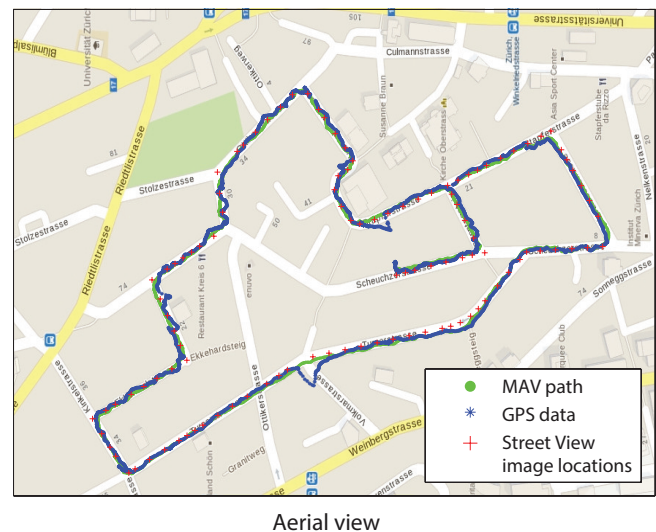
<http://rpg.ifi.uzh.ch/zurichmavdataset.html>

## Introduction

New applications of Micro Aerial Vehicles (MAVs) are envisioned by several companies, e.g., good delivery (e.g., Amazon Prime Air, DHL, Alibaba, Matternet, Swiss Post), inspection and monitoring (e.g., SenseFly, Skycatch), medications and blood samples transportation (e.g., Matternet, Flirtey, Wingtra, RedLine), first-response and telepresence in case of accidents (e.g., Drone Adventures, Microdrones).

Accurate localization is indispensable and is a prerequisite for the successful completion of these tasks in a real-life environment. For above-rooftop flight, even consumer-grade standard and differential GPS receivers provide sufficiently accurate localization (less than 1 meter). However the accuracy and reliability of GPS sensing fundamentally depends on the number of visible satellites which are in the line of sight of the receiver. In urban areas, the availability of GPS signals is often reduced if compared to unobstructed terrain, or even completely unavailable in case of restricted sky view. So-called urban canyons tend to shadow the GPS signals, and building facades reflect the signals violating the underlying triangulation assumption that signals travel along a direct line of sight between the satellite and the receiver (i.e., multipath). Thus, in urban streets vision-based localization and position estimation algorithms are needed.

Recently, the TorontoCity dataset was proposed in Wang et al. (2016) that consists of data captured from: (i) overhead perspective (images and airborne LIDAR captured by airplanes and drones); (ii) ground perspective (street view panoramas, stereo images, Velodyne LIDAR, and GoPro captured by cars); (iii) high-precision maps (buildings and roads, 3D buildings, property meta-data). However, the



**Figure 1.** Bird-eye view of the urban test area. The red plus signs mark the locations of the ground Google Street View images. The blue asterisks mark the GPS labels of the aerial MAV images measured on-board. The green dots represent the ground truth path of the MAV.

<sup>1</sup>MTA SZTAKI, Institute for Computer Science and Control, Hungarian Academy of Sciences, Hungary

<sup>2</sup>Department of Informatics, University of Zurich, Switzerland

## Corresponding author:

András L. Majdik, Machine Perception Research Laboratory, MTA SZTAKI, Kende u. 13-17, 1111 Budapest, Hungary—<http://www.sztaki.hu>

Charles Till and Davide Scaramuzza, Robotics and Perception Group, University of Zurich, Andreasstrasse 15, 8050 Zurich, Switzerland—<http://rpg.ifi.uzh.ch>

Email: majdik@sztaki.hu; charles.till@gmail.com; davide.scaramuzza@ieee.org

**Table 1.** Dataset structure.

Folder/File name	Description
./Log Files/	Folder containing on-board log files and ground truth data.
--BarometricPressure.csv	Log data of the on-board barometric pressure sensor.
--OnboardGPS.csv	Log data of the on-board GPS receiver.
--OnboardPose.csv	Log of the on-board Pixhawk PX4 autopilot pose estimation.
--RawAccel.csv	Log data of the on-board accelerometer.
--RawGyro.csv	Log data of the on-board gyroscope.
--GroundTruthAGL.csv	Ground truth MAV camera positions.
--GroundTruthAGM.csv	Ground truth matches of image IDs between the aerial and ground level images.
--StreetViewGPS.csv	GPS tags of the ground level Street View images.
./MAV Images/	Folder with 81'169 images recorded by the MAV in the city of Zurich, Switzerland.
./MAV Images Calib/	30 images with calibration pattern to compute the intrinsic MAV camera parameters.
./Street View Img/	Folder with 113 Google Street View images covering the area of the data collection.
./calibration_data.npz	Internal camera parameters computed using the images from ./MAV Images Calib/
./loadGroundTruthAGL.m	This script is used by <i>plotPath.m</i> to load the data into Matlab.
./plotPath.m	Script to visualize the GPS and ground truth path in Matlab, similarly to Figure 3.
./write_ros_bag.py	Script to write the data into a ROS— <a href="http://ros.org">http://ros.org</a> bag file.
./readme.txt	More detailed descriptions about the files listed above.

TorontoCity benchmark contains only downwards facing images captured at high altitudes and lacks aerial footage captured by MAVs flying at low altitudes within urban streets. A visual-inertial dataset was proposed in [Burri et al. \(2016\)](#) for autonomous navigation of MAVs in indoor industrial environments (i.e., large machine hall and Vicon room). Conversely, the proposed dataset is meant to benchmark algorithms in outdoor environments and include aerial footage captured by a MAV flying at low altitudes within urban streets.

The dataset detailed in this paper is ideal to evaluate view-point invariant, image-based localization algorithms for GPS-denied MAVs, as we recently proposed in [Majdik et al. \(2015\)](#). In that paper, we described how to match low-altitude airborne images against Google Street View images to localize a quadrotor within urban streets. Furthermore, the dataset is a challenging benchmark to test visual odometry, SLAM, and online 3D reconstruction algorithms for MAV navigation in urban environments.

The 2 km dataset was recorded in January 2015 with a Fotokite\* quadrotor equipped with a GoPro Hero 4 camera, flying in a downtown area of Zurich at low altitudes (i.e., 5-15 meters above the ground). The Fotokite is a tethered MAV that enables aerial filming in confined environments, such as cluttered city streets with GPS signals that suffer from inaccuracy. The tether is connected to the user, who controls the 3D position of the drone either by maneuvering the tether or through a smartphone interface. The smartphone interface also allows the user to further change the yaw angle of the drone and the pitch angle of the camera. Battery power is provided through the tether for up to 45 minutes of non-stop flight. The tether also renders the Fotokite drone safe and legal for data collection in urban streets. All these advantages make the Fotokite drone the ideal platform to record our dataset.

The bird-eye view of the urban test area is shown in Figure 1. To record the dataset, the flying vehicle was piloted close to the center of the streets and the MAV camera was always kept facing the buildings. In Figure 1 the trajectory estimated by the on-board GPS is marked in blue. The red plus signs mark the locations of the ground

Google Street View images. In order to estimate the actual trajectory of the MAV (marked in green) we performed an accurate photogrammetric 3D reconstruction using the Pix4D<sup>†</sup> software, c.f. Figure 5. Note that the GPS signal was shadowed by the surrounding buildings, therefore a root-mean-square geo-location error of 2.22 meters in X, 3.76 meters in Y, and 5.46 meters in Z, exists relative to the actual path of the MAV.

## Dataset format

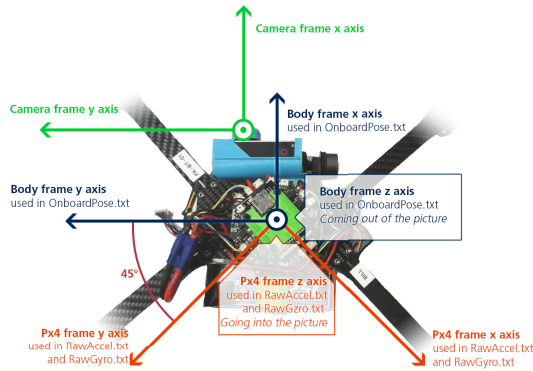
The dataset contains time-synchronized high-resolution images (1920 x 1080 x 24 bits), GPS, IMU, and ground-level Google-Street-View images. The high-resolution aerial images were captured with a rolling shutter GoPro Hero 4 camera that records each image frame line by line, from top to bottom with a readout time of 30 millisecond. A summary of the enclosed files is given in Table 1.

The data from the on-board barometric pressure sensor *BarometricPressure.csv*, accelerometer *RawAccel.csv*, gyroscope *RawGyro.csv*, GPS receiver *OnboardGPS.csv*, and pose estimation *OnboardPose.csv* is logged and time-synchronized using the clock of the PX4 autopilot board. The on-board sensor data was spatially and temporally aligned with the aerial images. The delta time period was set only once at the beginning of the recording and was not changed for every individual image. Its spatio-temporal accuracy was checked by examining fixed and well-identifiable points along the Fotokites path. This was done by comparing the motion from the images with respect to the GPS and IMU measurements. The frame rate of 30Hz of the images combined with the slow and stable speed, at which the tethered Fotokite moves, made it possible to achieve a high spatio-temporal accuracy in the alignment of the images and the GPS/IMU measurements.

The first column of every file contains the timestamp when the data was recorded expressed in microseconds. In the next columns the sensor readings are stored. The second column

\*Fotokite MAV: <http://fotokite.com>

<sup>†</sup>Pix4D image processing software: <https://pix4d.com/>



**Figure 2.** Different sensor coordinate systems used to record the data.

in *OnbordGPS.csv* encodes the identification number (ID) of every aerial image stored in the */MAV Images/* folder. The first column in *GroundTruthAGL.csv* is the ID of the aerial image, followed by the ground truth camera position of the MAV and the raw GPS data. The second column in *GroundTruthAGM.csv* is the ID of the aerial image, followed by the ID of the first, second and third best match ground-level street view image in the */Street View Img/* folder.

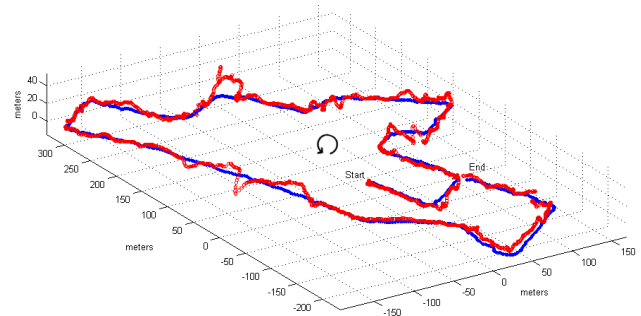
The coordinate frame conventions used to record the data are shown in Figure 2. The translation between the body frame coordinate system (blue on Figure 2) to the camera frame coordinate system (green on Figure 2) expressed in millimeters is: 75.66 mm x-axis (forward); 29.68 mm y-axis (left of center); -32.27 mm z-axis (below the top face of the PX4 PCB). The on-board GPS data *OnbordGPS.csv* and the GPS tags of the Street View images *StreetViewGPS.csv* use the international WGS 84 (GPS) coordinate system. The ground truth MAV camera positions *GroundTruthAGL.csv* are in the International WGS 84 / UTM zone 32N coordinate system. Next, we present tools that can be used by the reader to index the data programmatically.

## Parsing and indexing

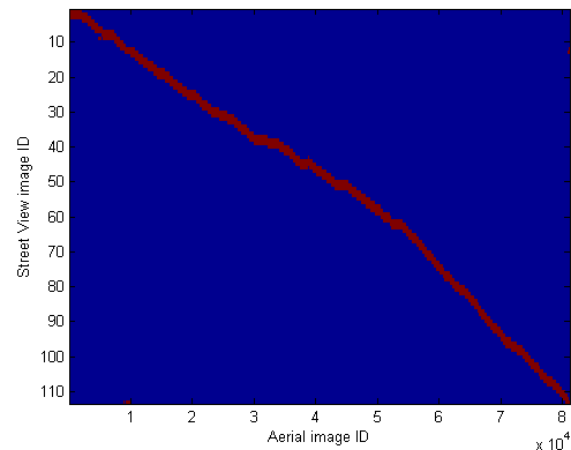
Beside the raw data measured during the flight, we provide a script to visualize the ground truth path of the Fotokite in comparison with the recorded GPS location data in MATLAB, c.f. Figure 3. Also, we provide a script to parse and write all the data into a rosbag file in order to be easily viewed and replayed using the ROS ecosystem. The logged numerical data is saved in human readable tables, the high quality images have jpeg format.

## Calibration

In order to compute the intrinsic parameters of the on-board camera, we used a calibration checkerboard with known dimensions. The size of the calibration checkerboard is nine squares wide and seven high, whereas the length of a single side of one square is 2,45 cm. To compute the enclosed camera parameters we used the pinhole camera model and the calibration tools from the OpenCV library<sup>‡</sup>.



**Figure 3.** Comparison of the estimated GPS (red) and ground truth (blue) trajectories.



**Figure 4.** Ground truth confusion matrix. This plot shows the correct matches (red) between the aerial MAV images and the ground Street View images. Note that for every aerial image three ground-level images are accepted as correct matches, based on the closest geometric distances computed from the GPS tags.

The calibration parameters and data are also included in the dataset.

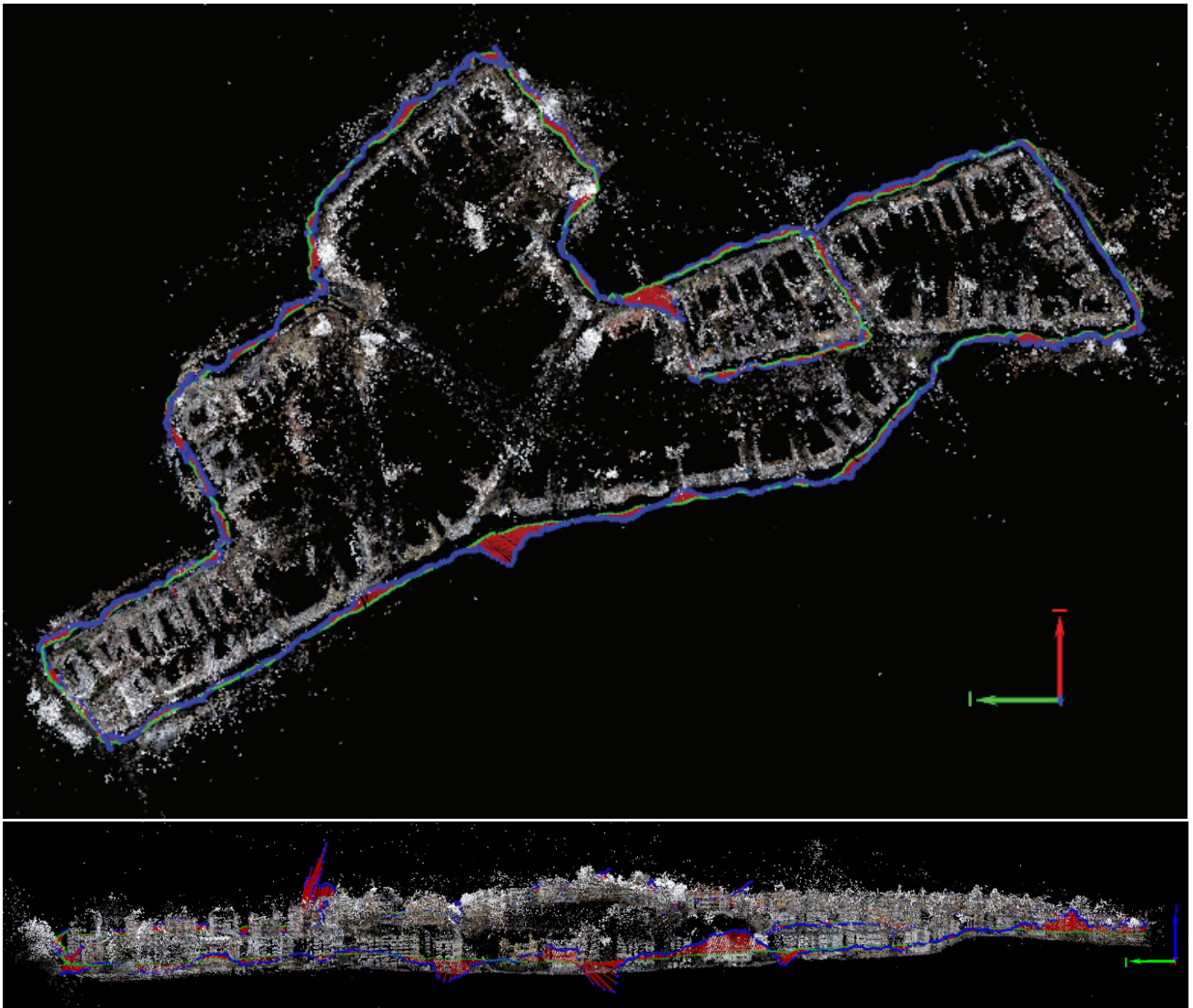
## Ground truth

Two types of ground truth data are provided in order to evaluate and benchmark different vision-based localization algorithms. Firstly, appearance-based topological localization algorithms, that match aerial images to street level ones, can be evaluated in terms of *precision rate* and *recall rate*. Secondly, metric localization algorithms, that computed the ego-motion of the MAV using monocular visual SLAM tools, can be evaluated in terms of *standard deviations* from the ground truth path of the vehicle.

### *Evaluation of topological localization algorithms*

In order to establish the ground truth confusion map, c.f., Figure 4, that shows the correct matches between the MAV images and the Street View images, we computed the three

<sup>‡</sup>OpenCV open source computer vision library: <http://opencv.org/>



**Figure 5.** Dense 3D point cloud obtained through Pix4D software using the airborne images of the dataset. The GPS is shown in blue, the camera position with green dots, while the red line that connects the dots shows the error (for further detail, check out the accompanying video on the dataset webpage).

closest geometric distances between the aerial and ground level images using the enclosed GPS tags. The ground truth matches were then verified visually. Thus, the performance of different air-ground image-matching based localization algorithms can be evaluated and compared in terms of precision and recall.

### *Evaluation of metric localization algorithms*

For the dataset proposed in this paper, it is not feasible to track the position of the MAV with a fully independent external reference systems (such as a VICON motion capture systems) to establish the ground truth. However, in order to compute the actual metric path of the MAV we performed an accurate photogrammetric 3D reconstruction using the Pix4D software. To obtain the best result and to reduce the accumulated error in consecutive measurements we recorded the data in such a manner to include loop-closure situations after long flight paths. To perform the reconstruction we sub-sampled the data at 1 fps. The GPS position was used as initial position of the images. Next, a total of

5'237'298 2D keypoint observations and 1'382'274 3D points were used for the block bundle adjustment in order to iteratively refine the camera positions. The mean reprojection error in pixels is 0.216531. In order to obtain a fully consistent 3D reconstruction additional 2D control points were marked manually on the images. For an overview of the reconstructed streets, check out the accompanying video on the dataset webpage.

In Figure 5 we show the top and the side view of the reconstruction. The accurate camera positions are marked with green dots, the measured GPS is shown in blue, the red line that connects the dots shows the error. In Figure 5 the metric camera positions, the reconstructed 3D model, and the GPS locations are all in the same frame of reference, the International WGS 84 / UTM zone 32N coordinate system. The actual trajectory of the MAV and the GPS trajectory is also compared in Figure 3. Note that in the figure the first image location is chosen as origin of the world coordinate system. Hence, different visual odometry, monocular SLAM,

and online 3D reconstruction algorithms can be evaluated in terms of standard deviations using the proposed dataset.

<sup>§</sup>DETEC Ordinance on Special Category Aircraft, 24th November 1994, Sec. 7, and Regulation of the operation of model aircraft on public ground by the Stadtpolizei of Zurich, 8th July 1983, Nr. 651/82, Sec. 1

## Acknowledgements

This dataset was recorded with the help of Karl Schwabe, Mathieu Noiroot-Cosson, Yves Albers-Schoenberg and the Zurich police. To record the dataset we used a Fotokite MAV offered to our disposal by Perspective Robotics AG—<http://fotokite.com>.

The data recording was done in compliance with the Zurich cantonal law<sup>§</sup>, upon approval of the Zurich police.

## Funding

This work was supported by the National Centre of Competence in Research Robotics (NCCR) through the Swiss National Science Foundation and by the Hungarian Scientific Research Fund (No. OTKA/NKFIH 120499).

## References

- Wang S, Bai M, Mattyus G, Chu H, Luo W, Yang B, Liang J, Cheverie J, Fidler S, Urtasun R (2016) *TorontoCity: Seeing the World with a Million Eyes*, December 2016. arXiv:1612.00423 [cs.CV].
- Burri M, Nikolic J, Gohl P, Schneider T, Rehder J, Omari S, Achtelik M, Siegwart R (2016) *The EuRoC micro aerial vehicle datasets*, The International Journal of Robotics Research, Volume 35, Issue 10, pages 11571163, January 2016. DOI: 10.1177/0278364915620033 SAGE Publications.
- Majdik A, Verda D, Albers-Schoenberg Y and Scaramuzza D (2015) *Air-ground Matching: Appearance-based GPS-denied Urban Localization of Micro Aerial Vehicles*, Journal of Field Robotics, Volume 32, Issue 7, pages 10151039, October 2015. DOI: 10.1002/rob.21585 Wiley Periodicals, Inc.