2019 International Conference on Robotics and Automation (ICRA)
Palais des congres de Montreal, Montreal, Canada, May 20-24, 2019

Are We Ready for Autonomous Drone Racing? The UZH-FPV Drone Racing Dataset

Jeffrey Delmerico, Titus Cieslewski, Henri Rebecq, Matthias Faessler, and Davide Scaramuzza.

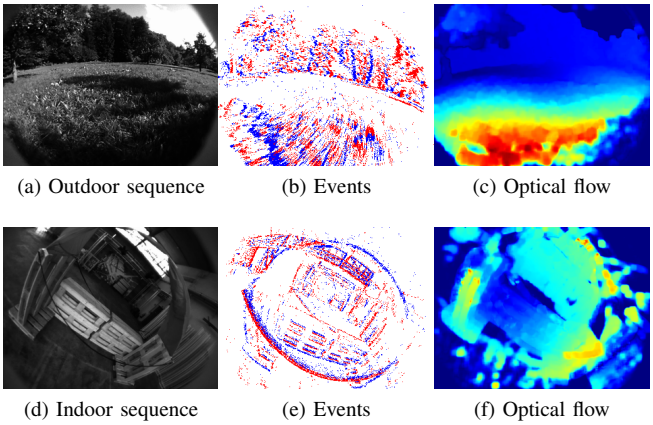
***Abstract*— Despite impressive results in visual-inertial state estimation in recent years, high speed trajectories with six degree of freedom motion remain challenging for existing estimation algorithms. Aggressive trajectories feature large accelerations and rapid rotational motions, and when they pass close to objects in the environment, this induces large apparent motions in the vision sensors, all of which increase the difficulty in estimation. Existing benchmark datasets do not address these types of trajectories, instead focusing on slow speed or constrained trajectories, targeting other tasks such as inspection or driving. We introduce the UZH-FPV Drone Racing dataset, consisting of over 27 sequences, with more than 10 km of flight distance, captured on a first-person-view (FPV) racing quadrotor flown by an expert pilot. The dataset features camera images, inertial measurements, event-camera data, and precise ground truth poses. These sequences are faster and more challenging, in terms of apparent scene motion, than any existing dataset. Our goal is to enable advancement of the state of the art in aggressive motion estimation by providing a dataset that is beyond the capabilities of existing state estimation algorithms.**

Supplementary Material

The dataset is available at `http://rpg.ifi.uzh.ch/uzh-fpv`.

I. Introduction

High-quality, large-scale, and task-driven benchmarks are key to pushing the research community forward. A well-known and compelling example can be found in autonomous driving, where the introduction of multiple datasets and benchmarks ([1], [2], [3], [4]) triggered drastic improvements of various low-level algorithms (visual odometry, stereo, optical flow), leading to impressive results on these benchmarks. These improvements ended up finding applications not only in autonomous driving, but also in many other tasks, benefiting the vision community as a whole. Yet, can we conclude that low-level vision is solved? Our opinion is that the constraints of autonomous driving—which have driven the design of the current benchmarks—do not set the bar high enough anymore: cars exhibit mostly planar motion with limited accelerations, and can afford a high payload and compute. So, what is the next challenging problem? We posit that drone racing represents a scenario in which low level vision is not yet solved. The fast, six-degree-of-freedom trajectories that occur in drone races, with high accelerations and rapid rotations, are beyond the capabilities of the current state of the art in state estimation. The purpose of this dataset is to spur innovation toward solutions to state estimation under these challenging conditions.



(a) Outdoor sequence (b) Events (c) Optical flow

(d) Indoor sequence (e) Events (f) Optical flow

Fig. 1: We present a drone racing dataset containing synchronized IMU, camera and event camera data recorded in indoor and outdoor environments. The dataset exhibits the largest optical flow magnitudes (pixel displacement per second) among all visual-inertial datasets to date. Figs. 1a,1d: preview images from the high-quality onboard fisheye camera. Figs. 1b,1e: visualization of the asynchronous event stream for in the outdoor (resp. indoor) sequence, obtained by integrating events over a temporal window of $\Delta_t = 30\ ms$ (resp. $\Delta_t = 10\ ms$) (blue: positive events, red: negative events). Figs. 1c,1f: color-coded magnitude of the optical flow (blue is small, red is large).

Very recently, there has been tremendous enthusiasm for autonomous drone racing in the research community due to the research challenges posed by agile flight, including the now annual IROS Autonomous Drone Race [5], [6]. More generally, high-speed robot navigation in cluttered, unknown environments is currently a very active research area [7], [8], [9], [10], [11], [12], [13] and funding of over 50 million US dollars has been made available through the DARPA Fast Lightweight Autonomy Program (2015-2018) and the DARPA Subterranean Challenge (2018-2021).

However, drone racing did not begin in research or industrial labs, but rather grew from a community of passionate amateurs. These First Person View (FPV) hobbyists devel-

This work was supported by the National Centre of Competence in Research Robotics (NCCR) through the Swiss National Science Foundation, the SNSF-ERC Starting Grant, and the DARPA Fast Lightweight Autonomy Program.

The authors are with the Robotics and Perception Group, Dept. of Informatics, University of Zurich, and Dept. of Neuroinformatics, University of Zurich and ETH Zurich, Switzerland—`http://rpg.ifi.uzh.ch`. Jeffrey Delmerico is now with Microsoft Mixed Reality and AI Lab, Zurich, Switzerland. Matthias Faessler is now with Verity Studios, Zurich, Switzerland.

oped an expertise in the design of agile quadrotor platforms flown by trained human pilots, beginning a new sport now known as drone racing. The growth of the sport has been rapid, with a corresponding boom in the business of racing as well [14] , including the introduction of professional leagues and an international competition with a million dollar prize bounty for human pilots [15]. This expansion of the sport is beginning to benefit the research community, with crossover projects such as the recently announced AlphaPilot Innovation Challenge, a competition sponsored jointly by Lockheed Martin and the Drone Racing League where autonomous drones must navigate race courses, featuring over 2 million US dollars in cash prizes [16].

Expert human FPV pilots are proof that autonomous drone racing should be possible with minimal sensing. However, we are still far from achieving human-level performance with autonomous drones. The constraints are truly demanding: drone racing requires platforms with a limited payload for sensors and limited computational power to track aggressive and agile 6-DOF maneuvers with low latency, while controlling for complex aerodynamics. For these reasons, we believe that pushing the limits of drone racing is a compelling research problem to move the research community forward. The algorithms needed to achieve good visual tracking in the challenging drone racing scenario will highly benefit robotics in general, and will contribute to multiple areas such as autonomous transportation, disaster relief, or even space exploration. However, there exist no drone racing datasets with ground truth trajectories available to the community at the current time. In fact, the only existing drone flight datasets (EuRoC [17], MVSEC [18], Zurich Urban MAV Dataset [19]) were not designed for aggressive flight, mostly because the mobile recording platforms for these datasets mount numerous bulky sensors (LIDAR, stereo cameras, etc.), which did not allow them to perform the aggressive and agile maneuvers that are required for drone racing.

The recently released Blackbird Dataset [20] does target aggressive flight, using a motion capture system for closed-loop control of fast UAV trajectories in an indoor flying room, while rendering photorealistic images of synthetic scenes to pair with the IMU measurements onboard the UAV. While Blackbird improves on the previous datasets in terms of speed and aggressiveness, and moves the state of the art in the direction of autonomous drone racing, its trajectories are constrained by the dimensions of the motion capture room, limiting their maximum speed. The Blackbird dataset offers some appealing properties, including a large number of sequences, a variety of simulated environments, and high-rate motion capture ground truth, but the UZH-FPV Drone Racing dataset significantly increases the challenge and realism for drone racing, with trajectories that are faster and longer (see Sec. II), and more importantly, from real-world scenarios (i.e., not synthetic ones).

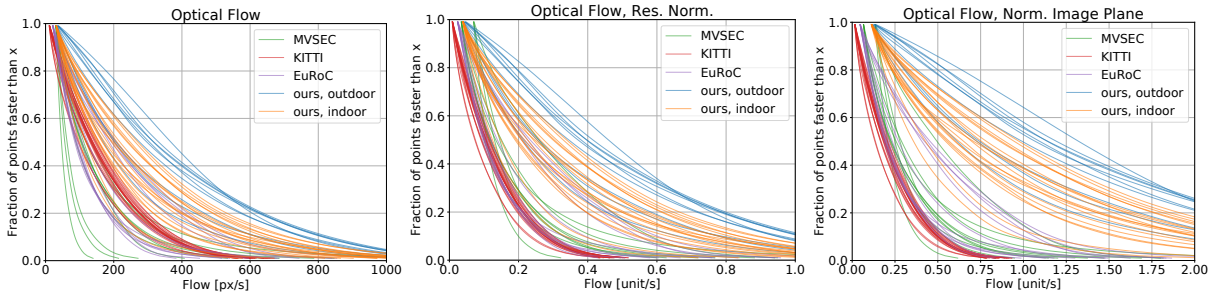
A. Contributions

In this work, we propose to bridge this gap by introducing the first "research-ready" dataset specifically targeted at fast, aggressive, and agile quadrotor maneuvers such as those required for drone racing. To that end, we combine recent developments in hardware and sensors, and build on top of the wisdom the hobbyist FPV community has developed, in order to bring high speed drone racing into the research world. Specifically, our dataset contains sequences recorded by sensors onboard a drone racing quadrotor flown aggressively by an expert human pilot in various indoor and outdoor race courses with the sequences split between forward and 45° down camera orientations. For each sequence, we provide the ground truth 6-DOF trajectory flown, together with onboard images from a high-quality fisheye camera, inertial measurements, and *events* from an event camera. Event cameras are novel, bio-inspired sensors which measure *changes* of luminance asynchronously, in the form of *events* encoding the sign and location of the brightness change on the image plane (more details can be found in Section III-B.0.a). Their outstanding properties (high dynamic range, robustness to motion blur, low latency) make them perfectly suited for estimating high speed camera motions, but research into state estimation algorithms with event cameras is not yet well developed. With this dataset, we also intend to foster further research in this exciting area.

II. Aggressive flight

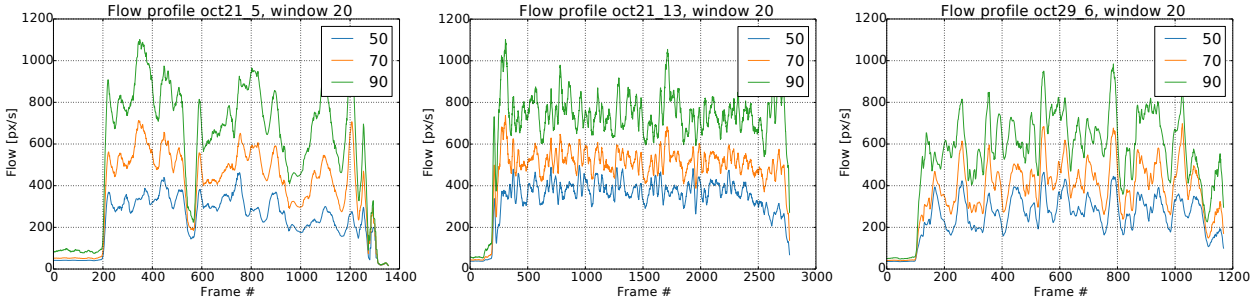
In previous datasets for vision-based state estimation, the metric to quantify the difficulty for this task has been linear and rotational velocity [22], [18]. However, the challenge in visual navigation is not so much the absolute velocity of the camera, but rather the apparent motion of the scene in the image plane. For example, a drone equipped with a downward-looking VGA sensor (640x480 pixels) with a 90 deg FOV lens, flying at $10\frac{m}{s}$ will be able to estimate its pose much easier at an altitude of $10m$ (resulting optical flow magnitude: $\simeq 320\ px/s$) than at an altitude of $10cm$ ($\simeq 32000\ px/s$!). Hence, we propose to use the magnitude of the optical flow as a metric to quantify the level of difficulty of different datasets. We hope that we can motivate the community to adopt this metric, which in our view is a more objective metric for expressing how "fast" a visual dataset is. Aside from more directly expressing the difficulty of visual motion estimation, it also allows us to combine translational and rotational motion in a single quantity.

However, optical flow varies across the image, especially when the camera is oriented in the direction of motion. In that case, even the fastest (forward) motion would have close to 0 optical flow in parts of the image (specifically, in the region close to the focus of expansion). Therefore, we report the distribution of the optical flow in: a) pixels per second, b) pixels normalized by the image diagonal per second, and c) pixels normalized by the camera focal length per second. Pixels per second is the most direct way of measuring the optical flow, but depends on the image resolution of the sensor. To be more objective for different sensor configurations, we normalize over the image resolution. We use the image diagonal to account for the fact that different datasets use cameras with different aspect



(a) Optical Flow (b) Resolution-normalized Flow (c) Focal-length-normalized Flow

Fig. 2: Optical flow comparison to [17], [18], [21] using the Snapdragon camera from our dataset. We report three metrics: (a) standard flow in pixels per second, (b) resolution-normalized flow (flow in pixels per second divided by the image diagonal in pixels) and (c) focal-length-normalized flow (flow divided by the focal length of the camera in pixels). Resolution-normalized flow depends on the field of view and is related to how long features stay in the image. Focal-length-normalized flow is independent of the resolution and field of view. For all three metrics, a higher value means more optical flow.



(a) Outdoor 5 (b) Outdoor 13 (c) Indoor 6

Fig. 3: Evolution of the optical flow distribution over time. We report the $(50, 70, 90)$ percentiles per image, averaged over a window of 20 subsequent images to make the plots easier to read.

ratios. This metric for optical flow is loosely coupled to how long features that are tracked remain in the image. However, resolution-normalized optical flow still depends on the field of view of the camera. To obtain a metric that is completely independent of the camera, we also consider the optical flow normalized by the focal length of the camera, which expresses the optical flow in the normalized image plane.

When comparing datasets, we combine the distribution of optical flow over the images and the distribution over time into a single, large distribution (see Fig. 2). For this, we plot what fraction of pixels moves faster than x, given a flow speed x. The higher the resulting curve, the larger the overall optical flow. To summarize the distribution, we use the area under this curve (AUC). For the analysis of single sequences, we also report how the distribution of optical flow over the image evolves with time (see Fig. 3). This representation may help users of this dataset to correlate algorithm failures with the optical flow at the time of failure. To compute these metrics, the dense optical flow is measured using the OpenCV implementation of the Farnebäck method [23]. For this analysis, we exclude all points with an optical flow below one pixel per frame, as such points would mostly not contribute to state estimation, as they are generally without texture. As we can see in Fig. 2, our sequences exhibit the largest optical flow among similar datasets to date, especially the outdoor sequences, independently of the normalization metric used. A few of the more challenging EuRoC datasets exhibit *unnormalized* optical flow that is comparable to our indoor sequences (see Fig. 2a). However, the Snapdragon camera that we use has a larger field of view at a slightly lower resolution than the VISensor that was used in EuRoC, so we can see that our sequences exhibit higher *normalized* optical flows (see Fig. 2b and 2c). We do not consider the sequences from the Blackbird dataset [20] in Fig. 2 since the trajectories can be made to have arbitrary optical flow by changing the synthetic scenes that are rendered.

In addition to the optical flow metrics that we propose, we also collect more traditional properties of available drone-based visual inertial datasets in Table I, which show that the UZH-FPV Drone Racing dataset offers longer and faster trajectories than other datasets aimed at aggressive flight.

III. Datasets

A. Flying Platform

We now describe the components of the platform that we used for recording aggressive flying data (see Fig. 4). Drawing on the experience of several FPV racing enthusiasts, we designed a quadrotor platform based on a popular FPV racing frame (the Lumenier QAV-R carbon fiber frame) with custom-designed, 3D-printed parts for mounting computers

	EuRoC MAV [17]	UPenn Fast Flight [24]	Zurich Urban MAV [19]	Blackbird [20]	**UZH-FPV Drone Racing**
Environments	2	1	3	**5**^{a}	2
Sequences	11	4	1	**186**	27
Camera (Hz)	20	40	20	**120**	30/50^d + events
IMU (Hz)	200	200	10	100	**500/1000**^{e}
Motor Encoders (Hz)	n/a	n/a	n/a	**~190**	n/a
Max. Distance (m)	130.9	700	**2000**	860.8	340.1/923.5^f
Top Speed (m/s)	2.3	17.5	3.9^b	7.0	**12.8/23.4**^{g}
mm Ground Truth (Hz)	20/100^c	n/a	n/a	**360**	20

TABLE I: Comparison of visual-inertial datasets. Notes: a - Visual environments are rendered in photorealistic simulation. b - Velocity from GPS. c - Rate of precise ground truth from Leica in Machine Hall sequences/Vicon Room sequences. d - Frame rate of Snapdragon Flight/mDAVIS. e - IMU rate of Snapdragon Flight/mDAVIS. f - Distance indoor/outdoor sequences. g - Speed for indoor/outdoor sequences, both are larger than existing indoor or outdoor sequences, respectively.

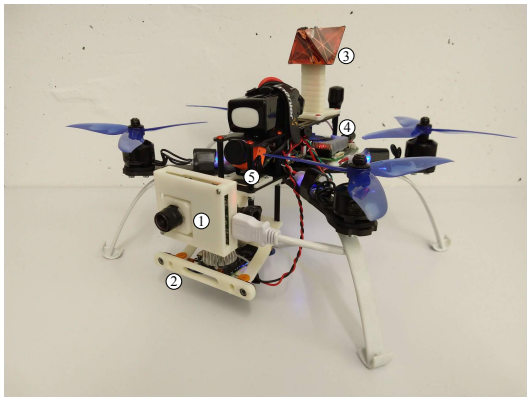


Fig. 4: The recording platform: ① mDAVIS ② Snapdragon Flight with fisheye stereo cameras ③ Omnidirectional reflector that provides the ground truth position. ④ Up Board used for recording the DAVIS data. ⑤ FPV camera used by the human pilot.

and sensors. Thrust is provided by Cobra CM-2208/20 2000KV motors with a 4S battery, and Dalprop T6040C 6" 3-blade propellers. The overall mass of the platform is 1110 g, and based on manufacturer specifications for this motor and propeller combination, the thrust-to-weight-ratio is approximately 3.5.

In order to measure ground truth position for trajectories in large indoor spaces (without a Vicon system) or outdoors, we utilize a laser tracking system in which a ground-based device continuously measures the 3D position of a reflective prism mounted to the platform as it moves. The tracking prism was mounted as far to the rear of the platform as possible, and stand off from the body of the frame, in order to minimize self-occlusions when the platform undergoes large pitch or roll motions. More information about the tracking system can be found in Sec. III-C. The quadrotor platform was flown by an expert FPV pilot.

B. Onboard Data

To keep the platform as lightweight as possible to achieve aggressive flight, we chose to use only lightweight visual-inertial sensors on our platform. Specifically, we mounted a miniDAVIS346 (mDAVIS), which provides events, frames, and IMU, and a Qualcomm Flight board, producing high-quality camera frames and IMU data.

a) Event Camera: The miniDAVIS346 (mDAVIS) is a very recent sensor developed by iniVation [25], with a spatial resolution of 346×260 pixels (this is the largest resolution commercially available on the event-camera market). It provides events with microsecond temporal resolution and can also record standard grayscale frames at 50 Hz. The events and the frames from the mDAVIS are spatially aligned and time-synchronized with the IMU directly on the hardware. We equipped the mDAVIS with a wide-angle lens (FOV: 120°), and mounted it either forward-facing on the quadrotor (see Fig. 4) or angled at 45° down. The sensor was isolated from vibrations with rubber dampers and was connected over USB to a companion computer (an Up Board [26], which receives and logs the sensor readings in ROS bag files [27]. This arrangement, with separate computers for the event-based sensor and standard camera, was done in order to avoid USB bus saturation and enable real-time logging.

b) Qualcomm Flight Board: The Qualcomm Flight is a highly integrated single board computer designed for drone and aerial robotic applications [28]. It provides high-quality, 640×480 grayscale frames from wide-angle cameras, as well as hardware-synchronized IMU measurements. Normally, the Qualcomm Flight uses a downward-looking wide-angle VGA camera for visual state estimation, but we adapted the camera mountings to orient the same camera modules (Sunny MD-102-A 186°FOV VGA camera units using an Omnivision OV7251 global shutter sensor) so that they were forward-looking or angled at 45° down in a stereo configuration. The board was mounted below the platform and was isolated from vibrations in the frame with rubber dampers. The camera images and IMU measurements were recorded onboard to ROS bag files.

C. External Data

Ground truth position measurements were provided by a Leica Nova MS60 TotalStation [29] laser tracker. This device is designed for very accurate and precise range and bearing measurements in near-static conditions, such as building construction and surveying. However, for ease-of-use, it also features the ability to lock to a reflective target and provide continuous measurements. When paired with a MRP122 360° prism, it provides position measurements with ± 1 mm accuracy in a gravity-aligned world reference frame. The MS60 was mounted to a tripod and leveled,

Fig. 5: Indoor and outdoor environments used for data collection. The indoor space is a former airplane hangar, with approximately 30m $\times$ 15m $\times$ 6m of open space. The outdoor environment was a large field sloping gently toward some woods, with several isolated trees.

providing a consistent reference point for measurement throughout each trajectory. Once locked to the prism, the MS60 tracker can provide position measurements at up to 20 Hz, and is capable of rotating up to 180° per second while tracking the prism. However, occlusions, aggressive motions, or high speeds can cause it to lose tracking or fail to obtain some samples. In practice, missing measurements occurred frequently, but typically for only a few samples at a time. However, in a quest to record the fastest and most aggressive trajectories possible, we often pushed the envelope of the MS60's tracking capabilities, so we typically flew each trajectory until an unrecoverable loss of tracking occurred during aggressive maneuvers, although a few of the sequences conclude with a landing back near the start of the trajectory. Considering that the MS60 is not designed for such high-speed applications, it nonetheless provided high-accuracy position measurements for the quadrotor platform during many aggressive maneuvers.

D. Data Collection

For each trajectory, the quadrotor platform was initialized at rest on the ground, and we acquired a tracking lock on the prism with the MS60. Once we began recording the ground truth position of the prism, we began logging the sensor data, and the FPV pilot then took off and flew a predetermined trajectory. The sequence ended when the tracker lost its lock on the prism or the quadrotor completed its trajectory and landed again on the ground.

In addition to the trajectories, calibration sequences were recorded for both the mDAVIS and Qualcomm Flight, for both intrinsic and camera-imu extrinsic calibration. High-resolution RGB video data from the FPV camera used for piloting was also logged, but this data is uncalibrated and unsynchronized with the inertial sensors or ground truth.

We captured data in two environments: one indoor, in an airplane hangar, and one outdoor, in a large field (see Fig. 5).

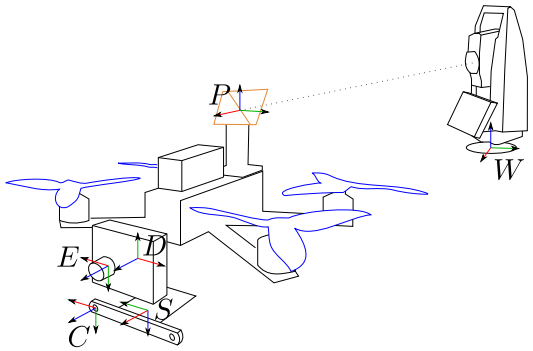


Fig. 6: Coordinate frame definition. W: World frame, P: Center of Leica prism, the MS60 tracks the position of P relative to W. E: event camera and D: IMU on the mDAVIS sensor. C: standard camera and S: IMU on the Snapdragon Flight. For each frame, red, green and blue arrows represent the unit vectors in x, y and z, respectively.

The indoor hall included an approximately 30m $\times$ 15m $\times$ 6m open space, that is part of a larger hangar building. In this space, different race courses were set up, including some with FPV racing gates, obstacles, or just open space to allow free-form trajectories. The outdoor environment was a large open field (approximately 80 m $\times$ 50 m) edged by forest, with a slight slope. Several isolated trees stood in the field and provided obstacles for trajectories that included circles, figure eights, slaloms between the trees, and long, straight, high-speed runs. Details of the individual sequences that comprise the dataset can be found in the supplementary material at `http://rpg.ifi.uzh.ch/uzh-fpv`.

E. Ground Truth Alignment

Our collected data features sensor measurements recorded on two different onboard computers, as well as external position measurements from the Leica tracker, recorded to a third computer. In order to produce useful ground truth for the trajectories, we aligned the sensor data with our external measurements using a full maximum-likelihood batch optimization, similar to the procedure performed for the EuRoC MAV dataset [17], which featured a similar tracking setup.

The batch optimization procedure utilized all position measurements from the tracker and all IMU data from the onboard sensors. We estimated the following states:

$$\mathbf{x} = [\,\mathbf{q} \quad \mathbf{p} \quad \mathbf{v} \quad \mathbf{b}_g \quad \mathbf{b}_a\,]^\top$$

throughout the trajectory for both the Qualcomm Snapdragon Flight board (sensor frame $\mathbf{S}$) and mDAVIS (sensor frame $\mathbf{D}$). Here, $\mathbf{q}$ is the attitude, $\mathbf{p}$ is the position, and $\mathbf{v}$ is the velocity of the IMU frame, for the sensor frame in question. The gyroscope and accelerometer biases of the frame's IMU are represented by $\mathbf{b}_g$ and $\mathbf{b}_a$, respectively, and are modeled as Wiener processes. We additionally estimated the unknown translation $\mathbf{p}$ from the tracking prism (frame $\mathbf{P}$) to each IMU frame ($\mathbf{S}$ or $\mathbf{D}$) measured in the respective IMU frame, as well as the temporal offset (Δt) between each sensor system's measurements and the Leica tracking data:

$$\theta = [\,\mathbf{p}_{SP} \quad \mathbf{p}_{DP} \quad \Delta t_S \quad \Delta t_D\,]$$

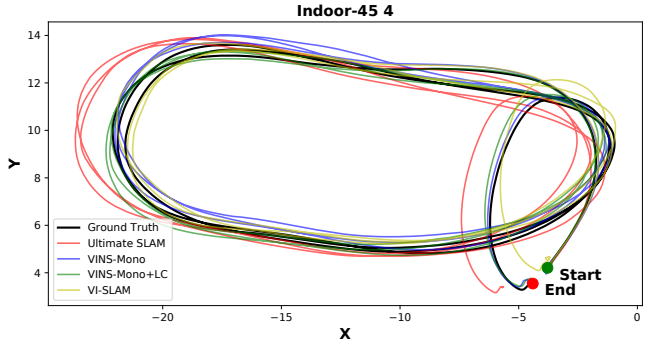


Fig. 7: Visual-inertial odometry results on one of the slower indoor sequences, performing several laps around a race track with gates. Methods shown include Ultimate SLAM [30] (an event-based VIO), VINS-Mono with and without loop closing [31], and Qualcomm's proprietary VI-SLAM algorithm.

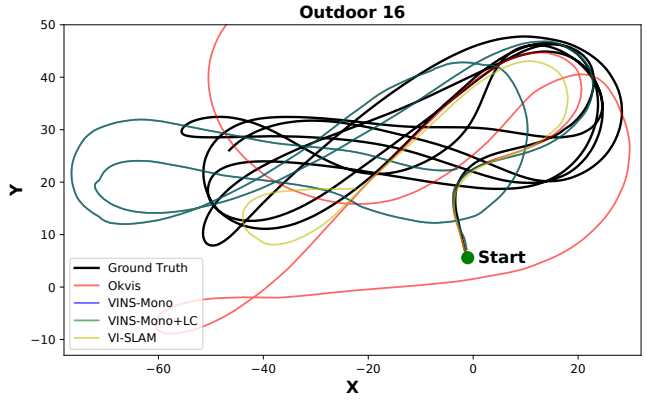


Fig. 8: Visual-inertial odometry results on the Outdoor 16 sequence. This is one of the faster outdoor sequences, flying several large figure eight laps around the open field. In this case, all of the algorithms (including Okvis [32] and VINS-Mono [31]) begin to drift on the second turn, and rapidly drift away from the ground truth trajectory. Qualcomm's VI-SLAM is able to track closer to the true trajectory for a bit longer, but fails before completing a full lap.

where the time offsets are taken to be constant over our short trajectories.

IV. Experiments

We performed a few qualitative experiments to evaluate the level of difficulty that our sequences pose for current state estimation algorithms. We selected one easier trajectory (Indoor45 4) and one more challenging sequence (Outdoor 16), and ran several visual-inertial odometry algorithms on the Snapdragon Flight and mDAVIS data (camera and IMU). Indoor45 4 has a low maximum velocity ($|\vec{v}|_{\max}$) and low area under the optical flow curve (AUC) relative to the other sequences in the dataset, while Outdoor 16 has a high maximum velocity and high AUC. For each sequence, we aligned the beginning of each algorithm's estimated trajectory and plotted the X-Y positions against the ground truth in Figures 7 and 8. Our goal in these experiments was not to perform a comprehensive benchmark evaluation of the state of the art, but rather to support our claim that the UZH-FPV Drone Racing dataset is more challenging for state estimation than existing datasets.

On the easier sequence, Indoor45 4 (see Fig. 7), all of the methods, including Qualcomm's proprietary VI-SLAM, as well as an event-based VIO method [30], exhibit some slow drift but track the complete trajectory successfully. The method based on event data [30] does not outperform the other methods. While we included this baseline for completeness, we point out that a direct comparison between [30] (which operates on data from the mDAVIS sensor [33]), and the other methods such as VI-SLAM is not fair. Indeed, these methods operate on much better hardware: cameras with a higher resolution (640×480 pixels versus 346×260 pixels), higher contrast sensitivity, better signal-to-noise ratio, and more accurate camera-to-IMU synchronization. Despite this, [30] can track the entire trajectory with limited drift. We view these results as promising, and hope to spur further research in this direction with our dataset. None of the algorithms performed well on the more challenging sequence, Outdoor 16 (see Fig. 8), all beginning to drift after a few turns at high speed, and accumulating significant drift and scale errors as the sequence progressed.

These results indicate that while the current state of the art in visual-inertial odometry algorithms can accurately track the full trajectory of slower datasets like EuRoC (see benchmark comparison in [34]), the UZH-FPV Drone Racing dataset is more challenging for these algorithms due to the larger apparent motions in the image stream. Our easier sequences are within reach of existing algorithms, but still non-trivial, while the challenging sequences will require advancement of the state of the art to successfully track.

V. Conclusion

We present a novel dataset for high speed, 6DoF state estimation, in which we leveraged the expertise of the FPV drone racing community to develop a sensor platform capable of the fastest and most agile trajectories yet released for the purpose of benchmarking. We additionally propose new metrics for assessing the level of difficulty of VIO datasets using optical flow statistics. Based on the growing interest in drone racing, our intention is to continue expanding this dataset with more environments, sequences, and sensor modalities. We hope that this data will spur the advancement of the state of the art in high speed state estimation.

Acknowledgments

The authors would like to thank Stefan Gächter, Zoltan Török, and Thomas Mörwald of Leica Geosystems for their support in gathering our data, and Innovation Park Zürich and the Fässler family for providing experimental space. Additional thanks go to iniVation AG and Prof. Tobi Delbruck for their support and guidance with the mDAVIS sensors, and to Francisco Javier Pérez Grau for his help in the evaluation.

REFERENCES

[1] A. Geiger, P. Lenz, and R. Urtasun, "Are we ready for autonomous driving? the KITTI vision benchmark suite," in *CVPR*, 2012.

[2] W. Maddern, G. Pascoe, C. Linegar, and P. Newman, "1 year, 1000 km: The Oxford RobotCar dataset." *Int. J. Robot. Research*, vol. 36, no. 1, pp. 3–15, 2017.

[3] F. Yu, W. Xian, Y. Chen, F. Liu, M. Liao, V. Madhavan, and T. Darrell, "BDD100K: A diverse driving video database with scalable annotation tooling," *CoRR*, vol. abs/1805.04687, 2018. [Online]. Available: http://arxiv.org/abs/1805.04687

[4] A. Gaidon, Q. Wang, Y. Cabon, and E. Vig, "Virtual worlds as proxy for multi-object tracking analysis," in *IEEE Int. Conf. Comput. Vis. Pattern Recog. (CVPR)*, Jun. 2016, pp. 4340–4349.

[5] "Iros 2018 autonomous drone racing competition," 2018. [Online]. Available: https://www.iros2018.org/competitions

[6] H. Moon, Y. Sun, J. Baltes, and S. J. Kim, "The IROS 2016 competitions," *IEEE Robot. Autom. Mag.*, vol. 24, no. 1, pp. 20–29, 2016.

[7] S. Karaman and E. Frazzoli, "High-speed flight in an ergodic forest," in *IEEE International Conference on Robotics and Automation*, 2012, pp. 2899–2906.

[8] A. Censi and D. Scaramuzza, "Low-latency event-based visual odometry," in *IEEE Int. Conf. Robot. Autom. (ICRA)*, 2014, pp. 703–710.

[9] C. Richter, W. Vega-Brown, and N. Roy, "Bayesian learning for safe high-speed navigation in unknown environments," in *Proc. Int. Symp. Robot. Research (ISRR)*, A. Bicchi and W. Burgard, Eds., 2018, pp. 325–341.

[10] K. Mohta, M. Watterson, Y. Mulgaonkar, S. Liu, C. Qu, A. Makineni, K. Saulnier, K. Sun, A. Zhu, J. Delmerico *et al.*, "Fast, autonomous flight in gps-denied and cluttered environments," *Journal of Field Robotics*, vol. 35, no. 1, pp. 101–120, 2018.

[11] C. Richter and N. Roy, "Safe visual navigation via deep learning and novelty detection," 2017.

[12] A. J. Barry, P. R. Florence, and R. Tedrake, "High-speed autonomous obstacle avoidance with pushbroom stereo," *Journal of Field Robotics*, vol. 35, no. 1, pp. 52–68, 2018.

[13] S. Jung, S. Hwang, H. Shin, and D. H. Shim, "Perception, guidance, and navigation for indoor autonomous drone racing using deep learning," *IEEE Robot. Autom. Lett.*, vol. 3, no. 3, pp. 2539–2544, Jul. 2018.

[14] "Business is booming for the leaders in drone racing," 2018. [Online]. Available: https://www.forbes.com/sites/darrenheitner/2017/08/08/business-is-booming-for-the-leaders-in-drone-racing/#3543f4a85a37

[15] "The super bowl of drone racing will offer $1 million in prize money," 2018. [Online]. Available: http://fortune.com/2016/03/03/world-drone-prix-offers-1-million-prize-money/

[16] "Lockheed martin and drone racing league launch groundbreaking ai innovation challenge," 2018. [Online]. Available: https://news.lockheedmartin.com/2018-09-05-Lockheed-Martin-and-Drone-Racing-League-Launch-Groundbreaking-AI-Innovation-Challenge

[17] M. Burri, J. Nikolic, P. Gohl, T. Schneider, J. Rehder, S. Omari, M. W. Achtelik, and R. Siegwart, "The EuRoC micro aerial vehicle datasets," *Int. J. Robot. Research*, vol. 35, pp. 1157–1163, 2015.

[18] A. Z. Zhu, D. Thakur, T. Ozaslan, B. Pfrommer, V. Kumar, and K. Daniilidis, "The multivehicle stereo event camera dataset: An event camera dataset for 3D perception," *IEEE Robot. Autom. Lett.*, vol. 3, no. 3, pp. 2032–2039, Jul. 2018.

[19] A. L. Majdik, C. Till, and D. Scaramuzza, "The Zurich urban micro aerial vehicle dataset," *Int. J. Robot. Research*, 2017.

[20] A. Antonini, W. Guerra, V. Murali, T. Sayre-McCord, and S. Karaman, "The blackbird dataset: A large-scale dataset for uav perception in aggressive flight," in *2018 International Symposium on Experimental Robotics (ISER)*, 2018.

[21] A. Geiger, P. Lenz, C. Stiller, and R. Urtasun, "Vision meets robotics: The KITTI dataset," *Int. J. Robot. Research*, vol. 32, no. 11, pp. 1231–1237, 2013.

[22] E. Mueggler, H. Rebecq, G. Gallego, T. Delbruck, and D. Scaramuzza, "The event-camera dataset and simulator: Event-based data for pose estimation, visual odometry, and SLAM," *Int. J. Robot. Research*, vol. 36, pp. 142–149, 2017.

[23] G. Farnebäck, "Two-frame motion estimation based on polynomial expansion," in *Image Analysis*. Springer Berlin Heidelberg, 2003, pp. 363–370.

[24] K. Sun, K. Mohta, B. Pfrommer, M. Watterson, S. Liu, Y. Mulgaonkar, C. J. Taylor, and V. Kumar, "Robust stereo visual inertial odometry for fast autonomous flight," *IEEE Robotics and Automation Letters*, vol. 3, no. 2, pp. 965–972, 2018.

[25] "inivation," 2018. [Online]. Available: https://inivation.com/buy/

[26] "Up board," 2018. [Online]. Available: http://www.up-board.org/

[27] M. Quigley, K. Conley, B. Gerkey, J. Faust, T. Foote, J. Leibs, R. Wheeler, and A. Y. Ng, "ROS: an open-source Robot Operating System," in *ICRA Workshop Open Source Softw.*, vol. 3, no. 2, 2009, p. 5.

[28] "Qualcomm flight," 2018. [Online]. Available: https://developer.qualcomm.com/hardware/qualcomm-flight

[29] "Leica nova ms60," 2018. [Online]. Available: https://leica-geosystems.com/products/total-stations/multistation/leica-nova-ms60

[30] A. Rosinol Vidal, H. Rebecq, T. Horstschaefer, and D. Scaramuzza, "Ultimate SLAM? combining events, images, and IMU for robust visual SLAM in HDR and high speed scenarios," *IEEE Robot. Autom. Lett.*, vol. 3, no. 2, pp. 994–1001, Apr. 2018.

[31] T. Qin, P. Li, and S. Shen, "Vins-mono: A robust and versatile monocular visual-inertial state estimator," *IEEE Transactions on Robotics*, vol. 34, no. 4, pp. 1004–1020, 2018.

[32] S. Leutenegger, S. Lynen, M. Bosse, R. Siegwart, and P. Furgale, "Keyframe-based visual-inertial SLAM using nonlinear optimization," *Int. J. Robot. Research*, vol. 34, no. 3, pp. 314–334, 2015.

[33] C. Brandli, L. Muller, and T. Delbruck, "Real-time, high-speed video decompression using a frame- and event-based DAVIS sensor," in *IEEE Int. Symp. Circuits Syst. (ISCAS)*, Jun. 2014, pp. 686–689.

[34] J. Delmerico and D. Scaramuzza, "A benchmark comparison of monocular visual-inertial odometry algorithms for flying robots," in *IEEE International Conference on Robotics and Automation (ICRA)*, 2018.