Technische Universität Berlin

Some examples of Visual SLAM

Guillermo Gallego

17.12.2019

Latest survey paper: IEEE Trans. Robotics, 2016

Past, Present, and Future of Simultaneous Localization And Mapping: Towards the Robust-Perception Age

Cesar Cadena, Luca Carlone, Henry Carrillo, Yasir Latif,
Davide Scaramuzza, José Neira, Ian Reid, John J. Leonard

Abstract*—Simultaneous Localization And Mapping (SLAM) consists in the concurrent construction of a model of the environment (the *map*), and the estimation of the state of the robot moving within it. The SLAM community has made astonishing progress over the last 30 years, enabling large-scale real-world applications, and witnessing a steady transition of this technology to industry. We survey the current state of SLAM and consider future directions. We start by presenting what is now the *de-facto* standard formulation for SLAM. We then review related work, covering a broad set of topics including robustness and scalability in long-term mapping, metric and semantic representations for mapping, theoretical performance guarantees, active SLAM and exploration, and other new frontiers. This paper simultaneously serves as a position paper and tutorial to those who are users of SLAM. By looking at the published research with a critical eye, we delineate open challenges and new research issues, that still deserve careful scientific investigation. The paper also contains the authors' take on two questions that often animate discussions during robotics conferences: *Do robots need SLAM?* and *Is SLAM solved?

***Index Terms*—Robots, SLAM, Localization, Mapping, Factor graphs, Maximum a posteriori estimation, sensing, perception.**

I. Introduction

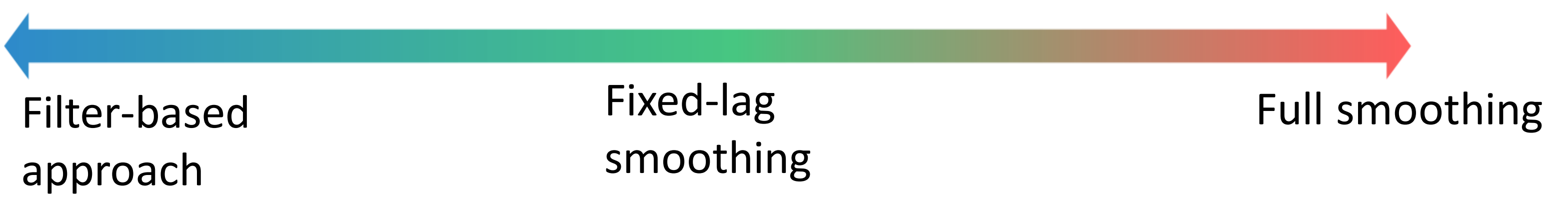
SLAM comprises the simultaneous estimation of the state of a robot equipped with on-board sensors, and the construction of a model (the *map*) of the environment that the sensors are perceiving. In simple instances, the robot state is described by its pose (position and orientation), although other quantities may be included in the state, such as robot velocity, sensor biases, and calibration parameters. The map, on the other hand, is a representation of aspects of interest (e.g., position of landmarks, obstacles) describing the environment in which the robot operates.

The need to use a map of the environment is twofold. First, the map is often required to support other tasks; for instance, a map can inform path planning or provide an intuitive visualization for a human operator. Second, the map allows limiting the error committed in estimating the state of the robot. In the absence of a map, dead-reckoning would quickly drift over time; on the other hand, using a map, e.g.,

What’s different in vSLAM?

- vSLAM, structure-wise, is the same as LiDAR SLAM
- Sensor model (cameras, IMUs): features vs. direct
- Motion model (6-DOF: rotation + translation)

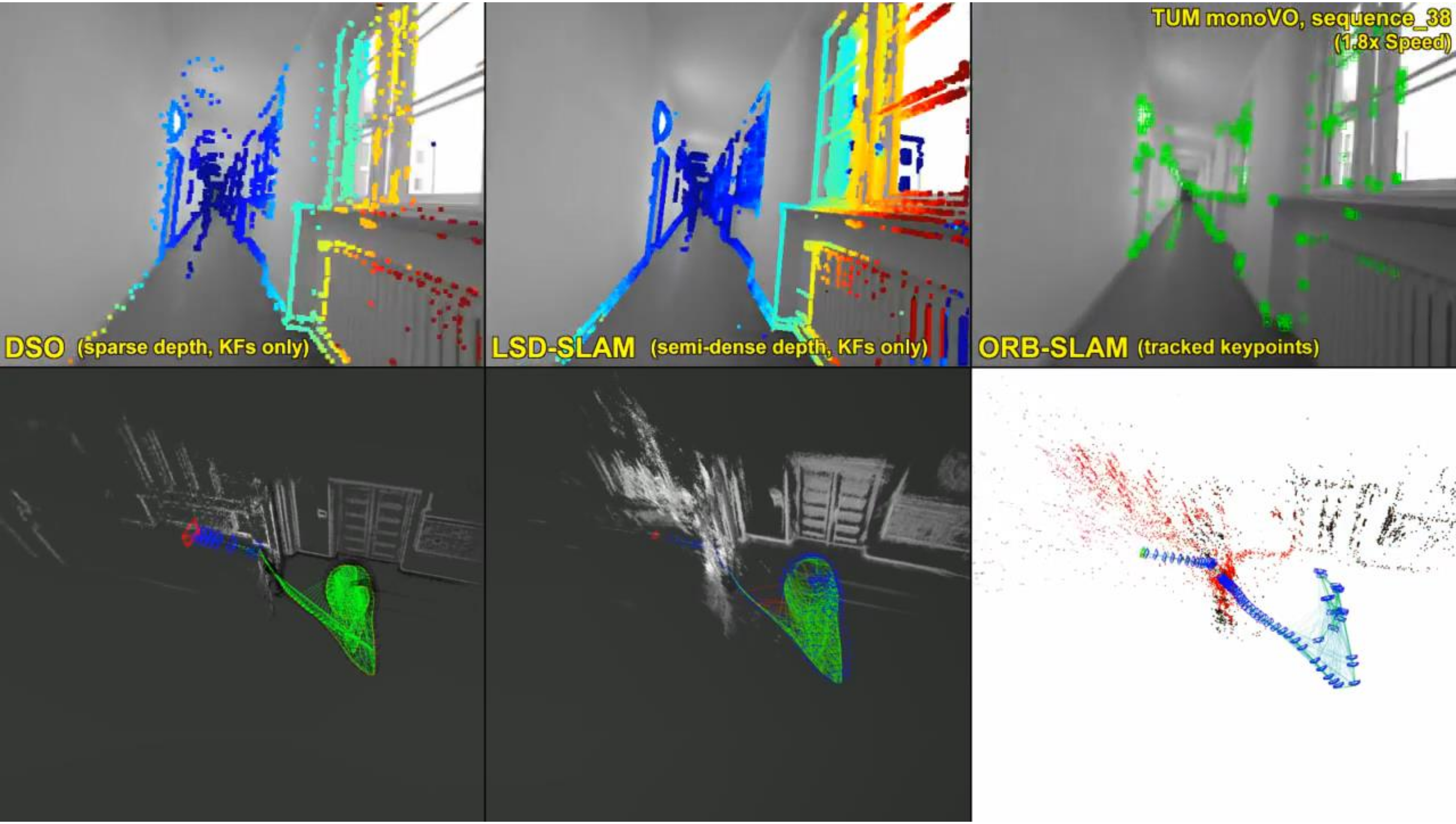
- Three main problem formulations:



  - Filter-based (estimate most recent pose and local map)
  - Fixed-lag smoothing (i.e., sliding window)
  - Full smoothing (estimate full trajectory and map)

DSO (state of the art VO)

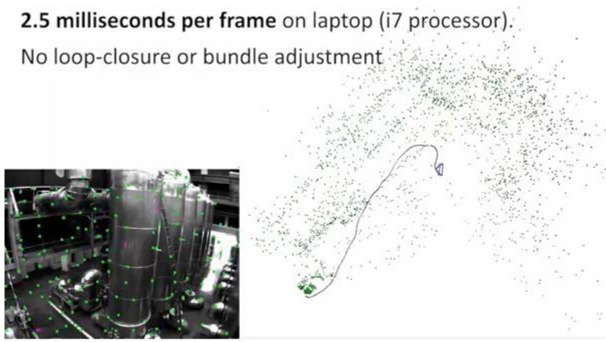
https://youtu.be/C6-xwSOOdqQ



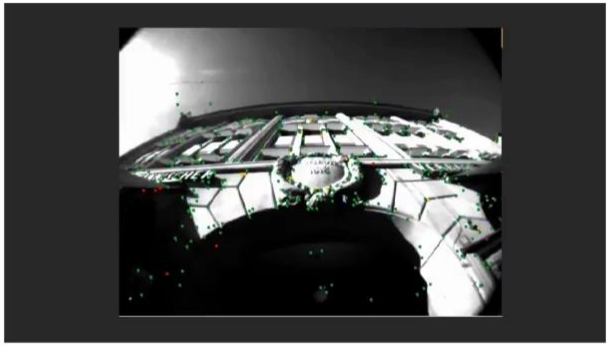
Direct Sparse Odometry, Trans. PAMI 2018

Visual-Inertial SLAM (Camera + IMU)

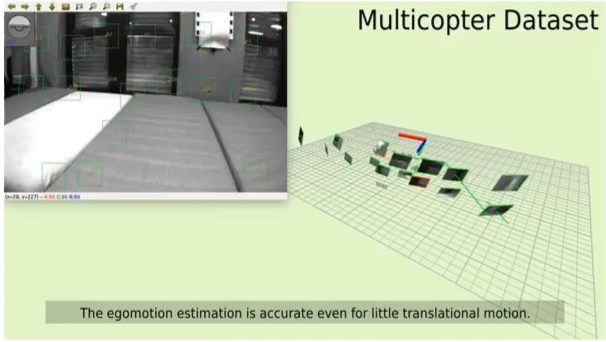
https://youtu.be/ymI3FmwU9AY



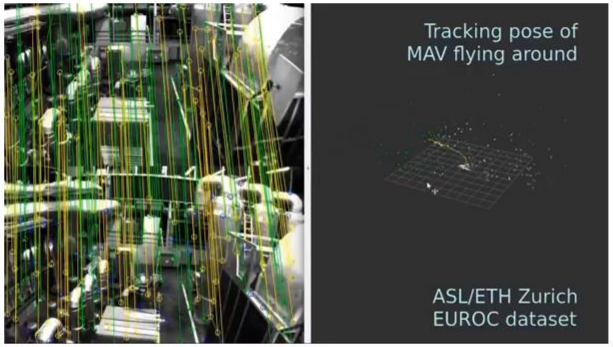
SVO+MSF



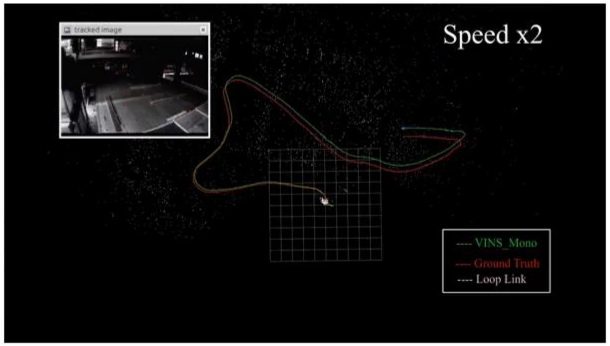
MSCKF



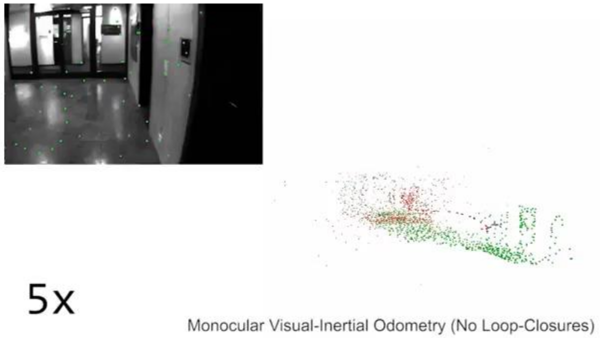
ROVIO



OKVIS



VINS-Mono

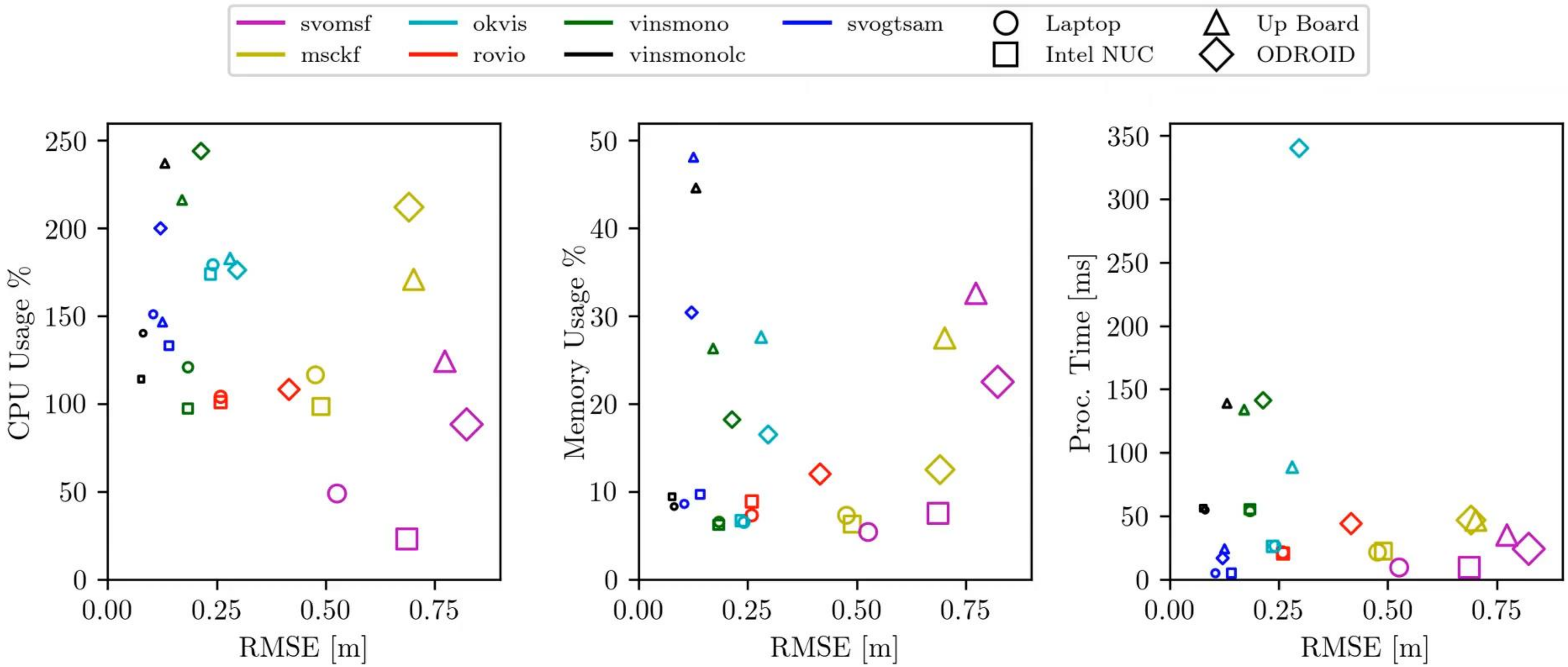


SVO+GTSAM

A Benchmark Comparison of Monocular Visual-Inertial Odometry Algorithms for Flying Robots, ICRA 2018

Camera + IMU (VIO) is a mature field

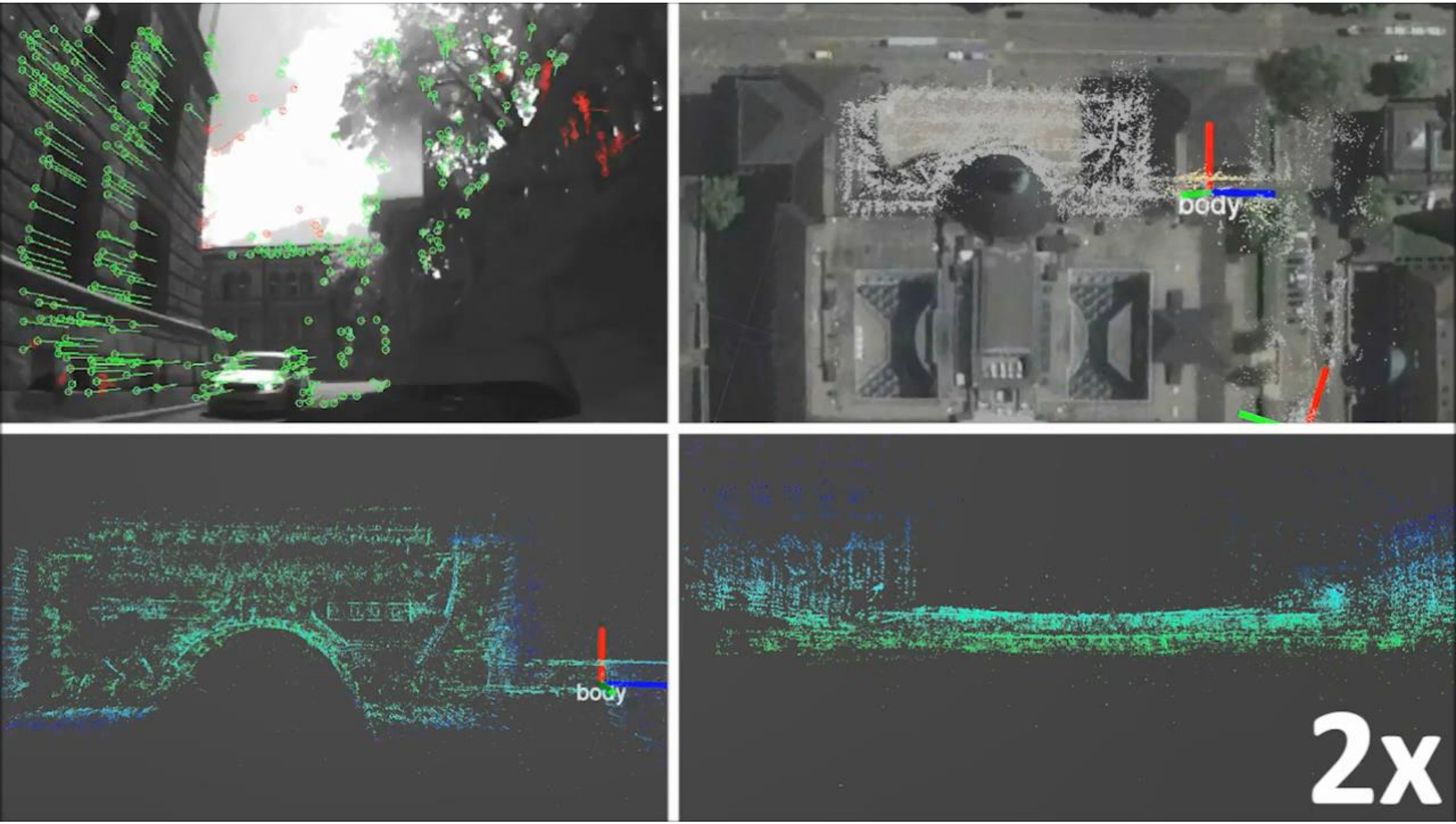
https://youtu.be/ymI3FmwU9AY



RMSE and resource usage give insight into the tradeoffs between accuracy and efficiency.

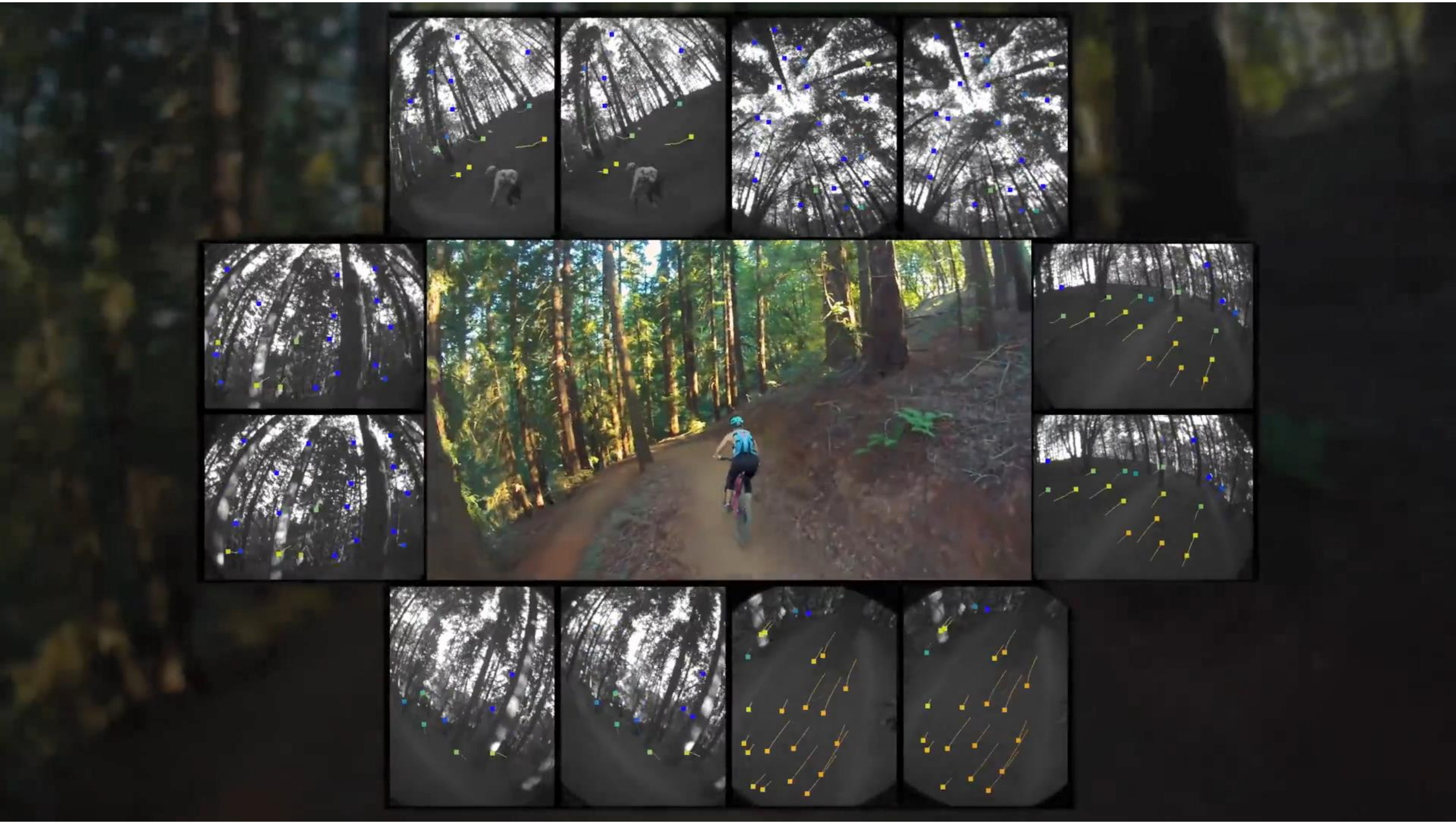
A Benchmark Comparison of Monocular Visual-Inertial Odometry Algorithms for Flying Robots, ICRA 2018

Autonomous driving



Drones. Ex: Skydio

https://www.skydio.com

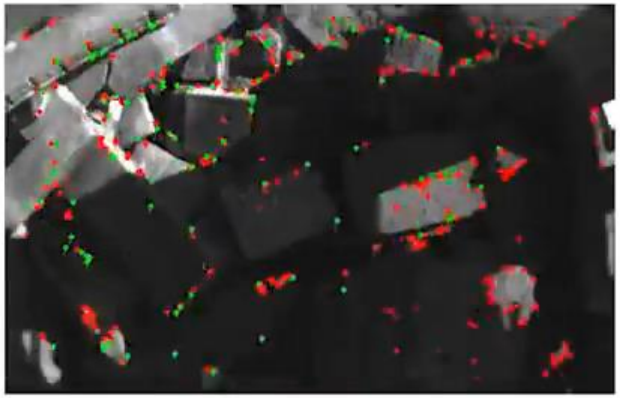


https://tech.fb.com/the-story-behind-oculus-insight-technology

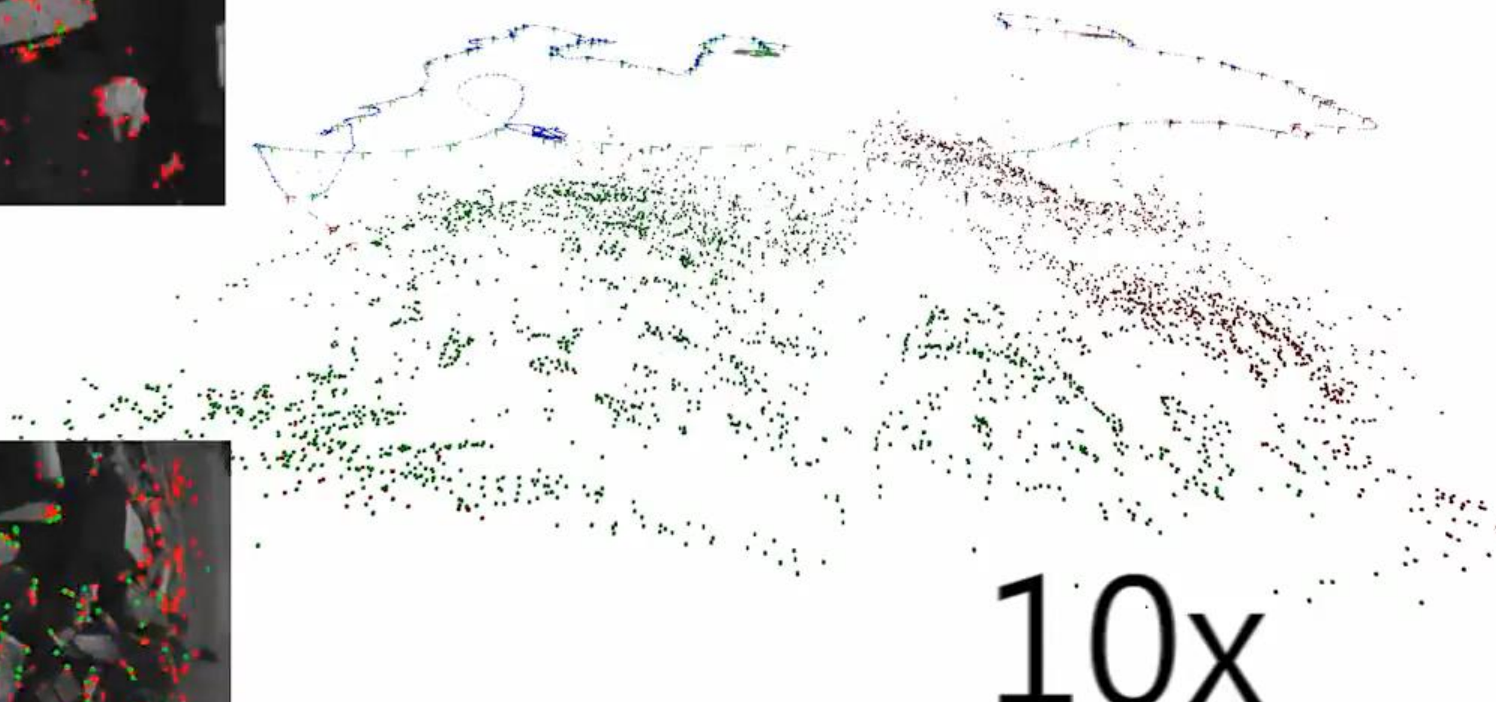
SLAM for AR/VR and Gaming

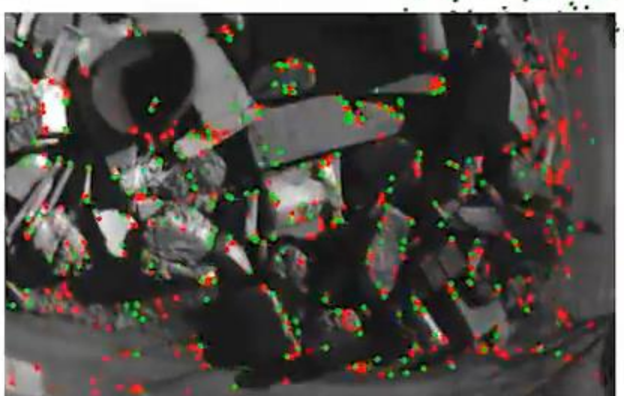
Examples: Microsoft Hololens, Oculus Rift / Quest, etc.

Multi-robot SLAM



https://youtu.be/taD3XF2w7A0

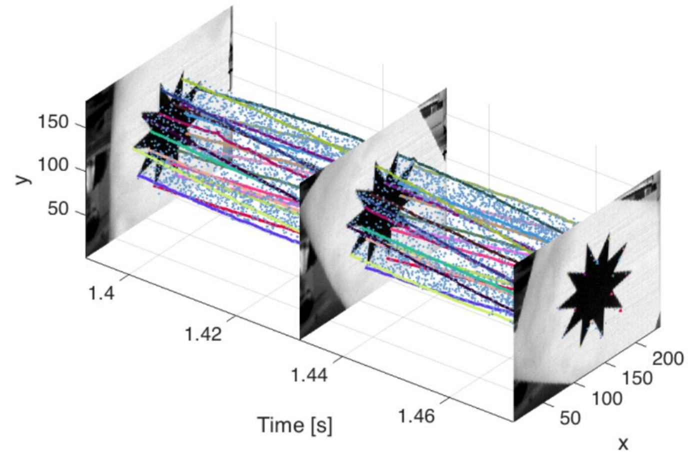


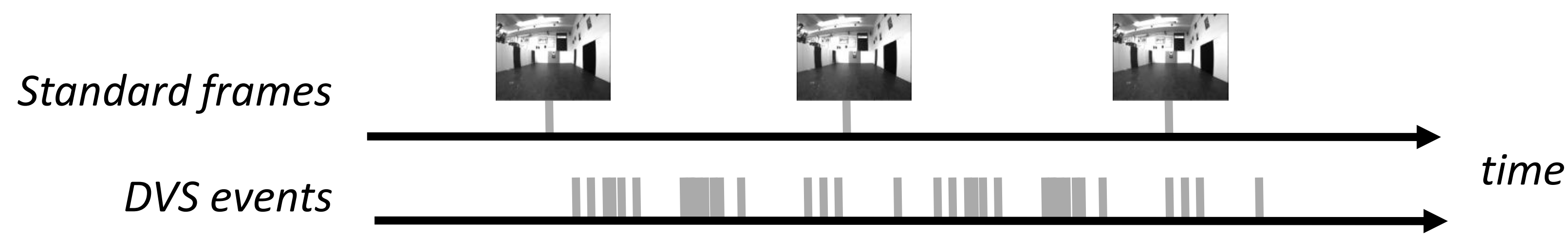


DAVIS sensor: DVS+ Standard Camera

- ➢ Combines DVS and a **standard** camera in the same pixel chip, plus an IMU
- ➢ **Output**: Frames + Events + IMU

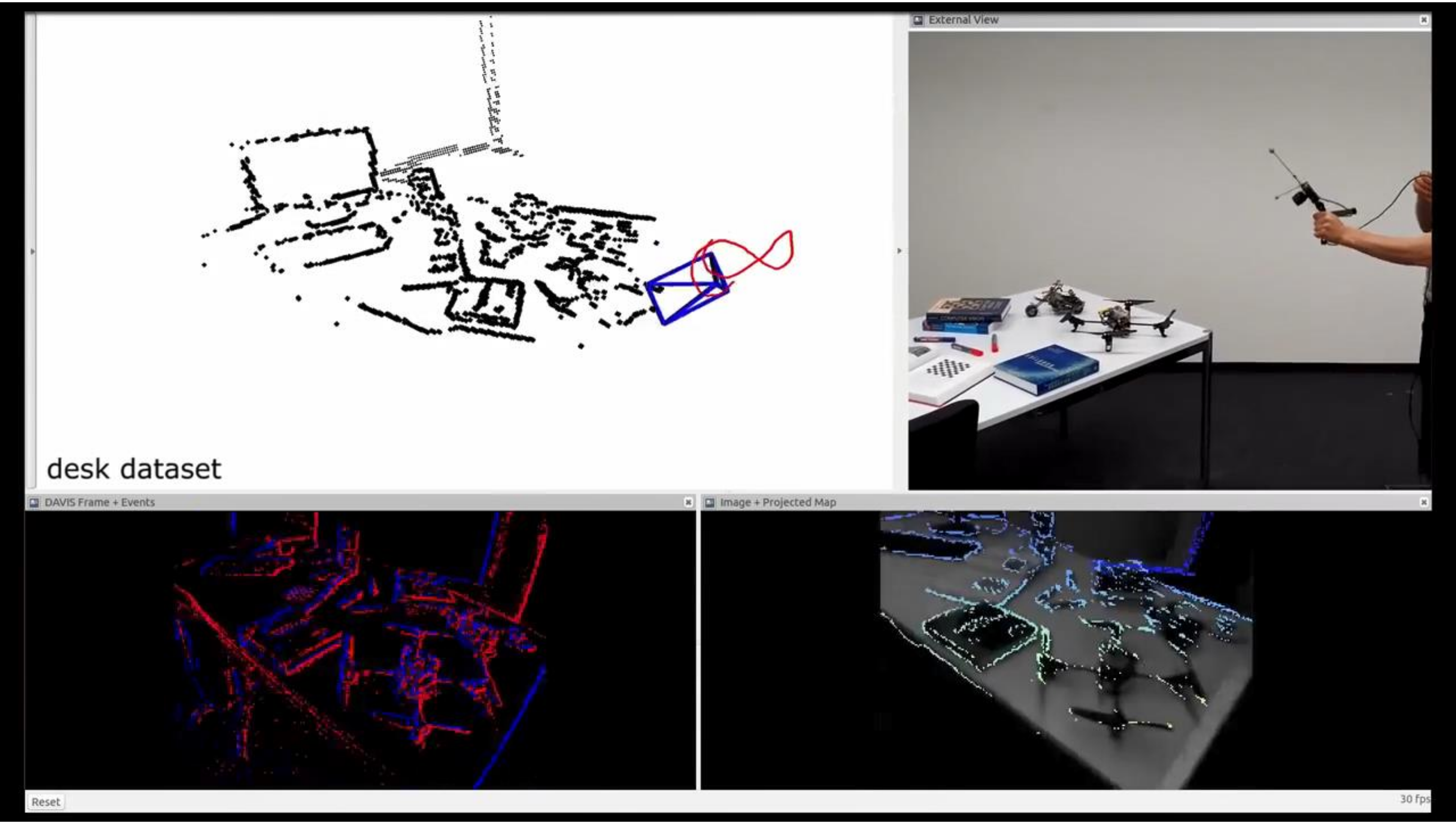
https://inivation.com/solutions/videos





Brandli et al. *A 240x180 130dB 3us latency global shutter spatiotemporal vision sensor*. IEEE JSSC, 2014

Event-based Visual Odometry

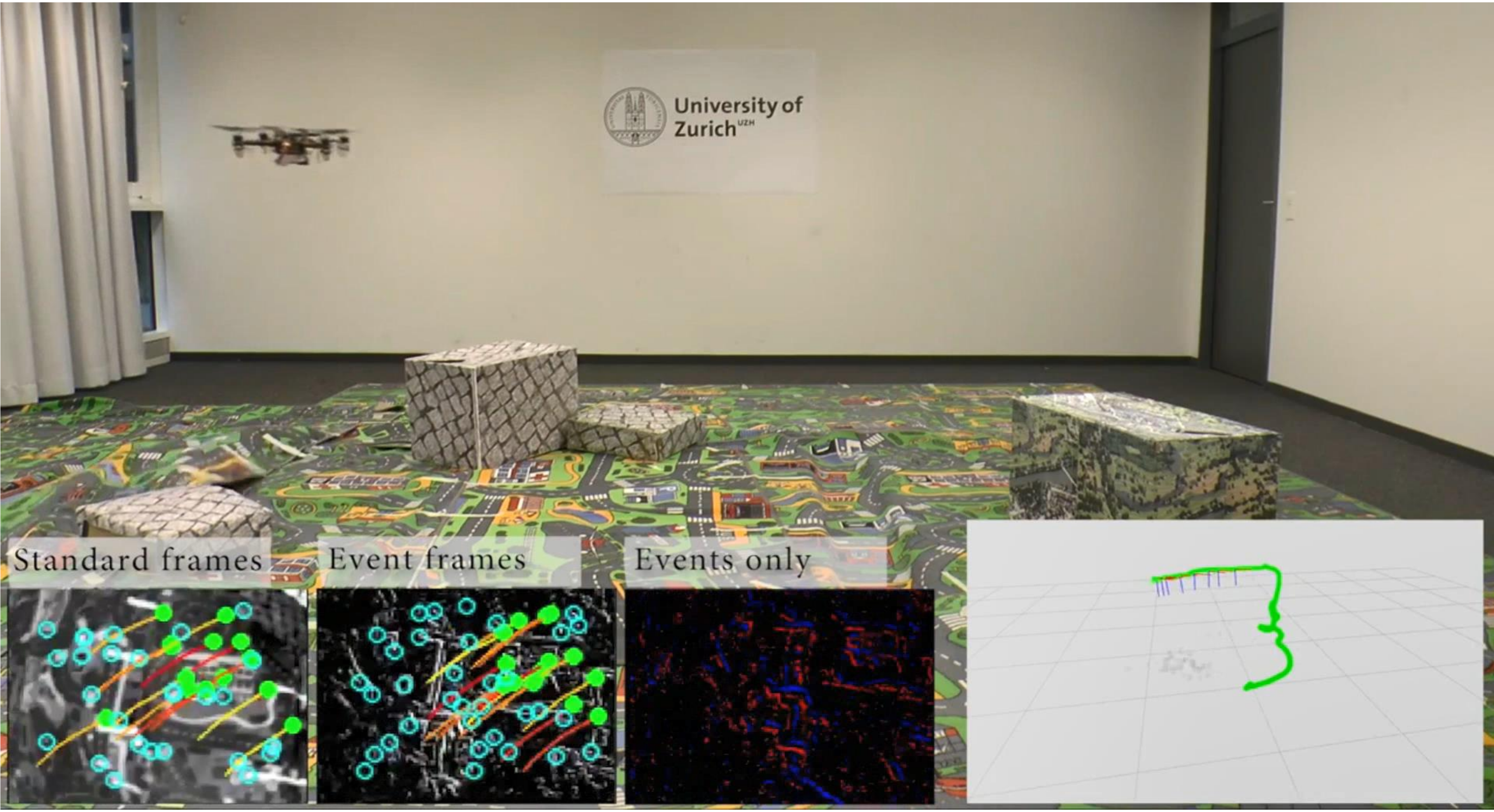


Rebecq et al., EVO: A Geometric Approach to Event-based 6-DOF Parallel Tracking and Mapping in Real-time," IEEE RAL'17. https://youtu.be/bYqD2qZJlxE

UltimateSLAM: Events + frames + IMU

Tightly coupled fusion. Fully onboard.

https://youtu.be/jIvJuWdmemE



Rosinol, Ultimate SLAM? Combining Events, Images, and IMU for Robust Visual SLAM in HDR and High Speed Scenarios, IEEE RAL'18