University of Zurich[UZH] | ETH zürich

Institute of Informatics – Institute of Neuroinformatics

ROBOTICS & PERCEPTION GROUP

Autonomous, Agile, Vision-controlled Drones:

From Frame to Event Vision

Davide Scaramuzza

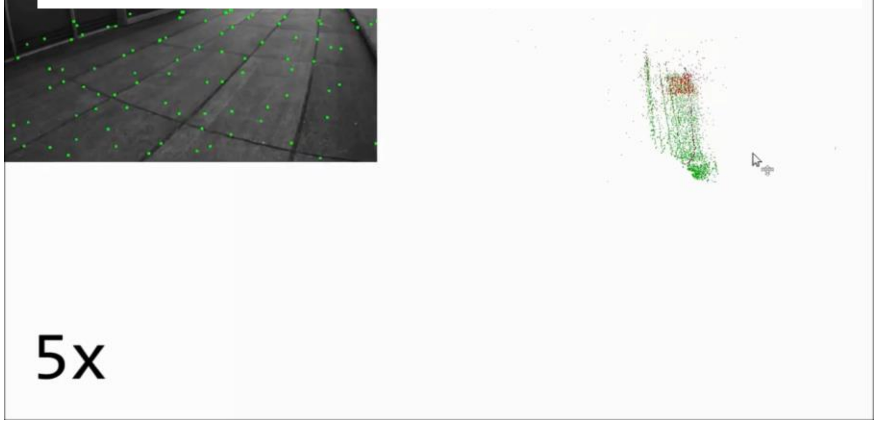
- My lab homepage: http://rpg.ifi.uzh.ch/
- Publications: http://rpg.ifi.uzh.ch/publications.html
- Software & Datasets: http://rpg.ifi.uzh.ch/software_datasets.html
- YouTube: https://www.youtube.com/user/ailabRPG/videos

Research Overview

Real-time, Onboard Computer Vision and Control for **Autonomous, Agile** Drone **Flight**

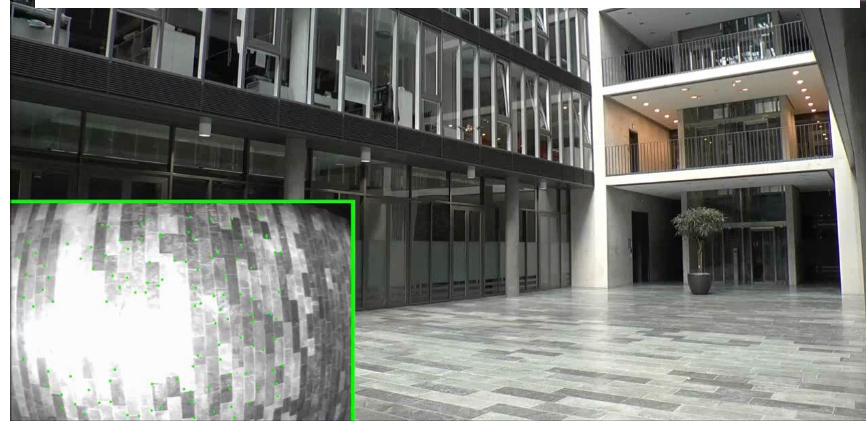
Visual-Inertial State Estimation (~SLAM)

http://rpg.ifi.uzh.ch/research_vo.html



Autonomous Flight

http://rpg.ifi.uzh.ch/research_mav.html

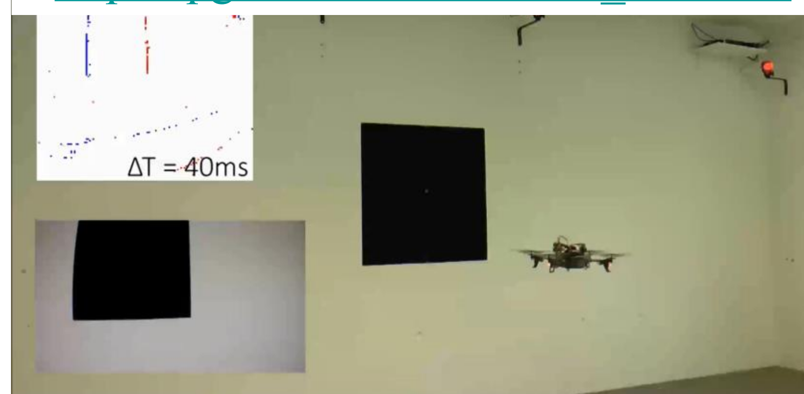


Learning-aided Autonomous Navigation

http://rpg.ifi.uzh.ch/research_learning.html

Event-based Vision for Low-latency Control

http://rpg.ifi.uzh.ch/research_dvs.html



Today's Main Applications of Drones (PWC report 2017)

Infrastructure Inspection

Agriculture

Aerial photography

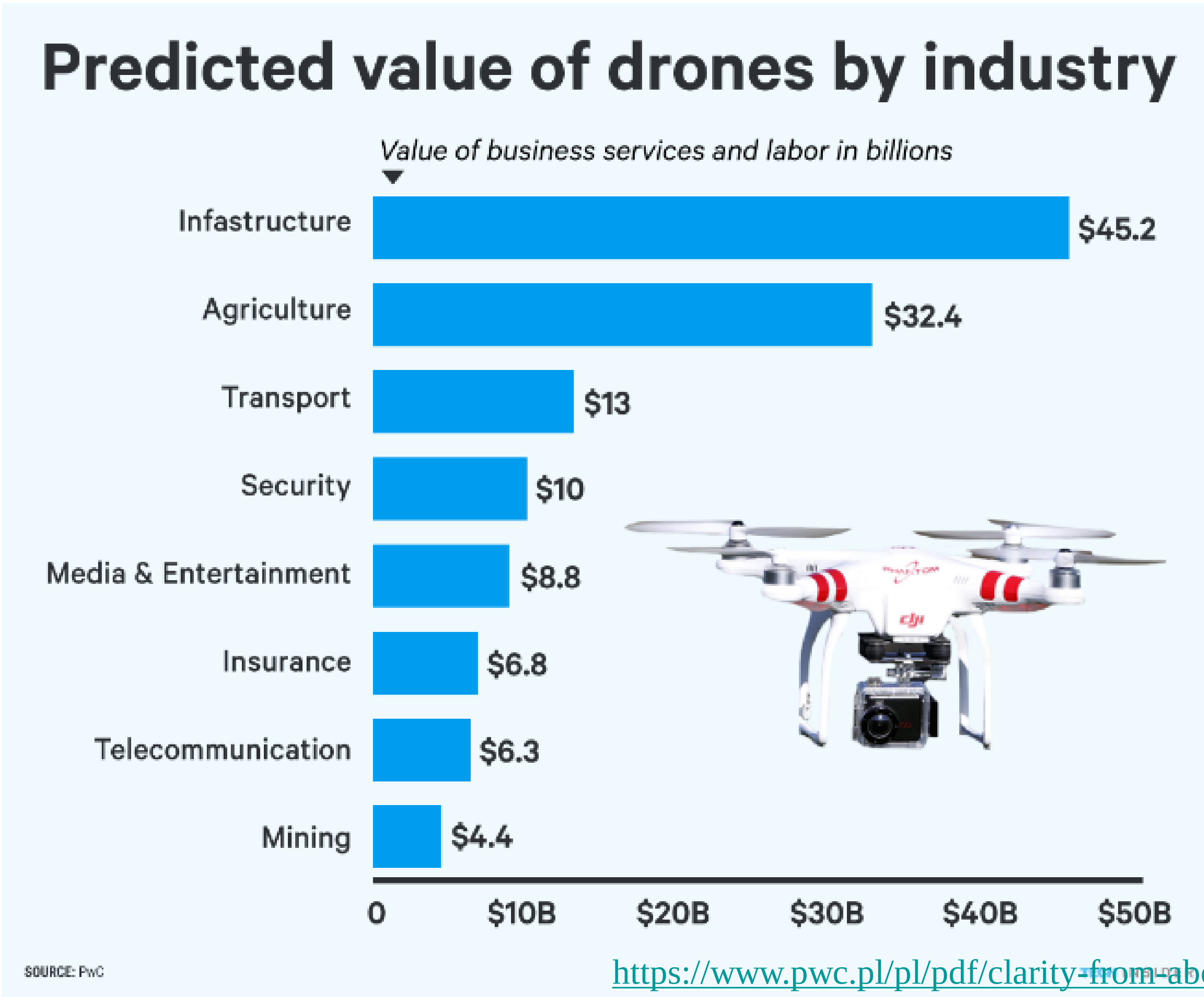
Transportation

Security & Search and rescue

Law enforcement

https://www.pwc.pl/pl/pdf/clarity-from-above-pwc.pdf

The Drone Market (current value: 130B$) [PWC report 2017]

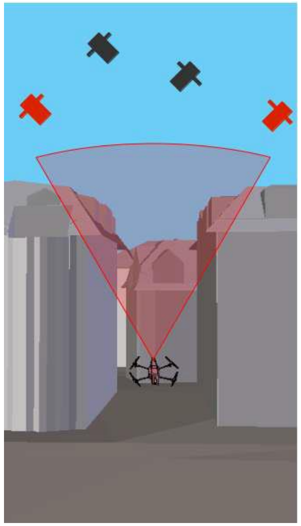


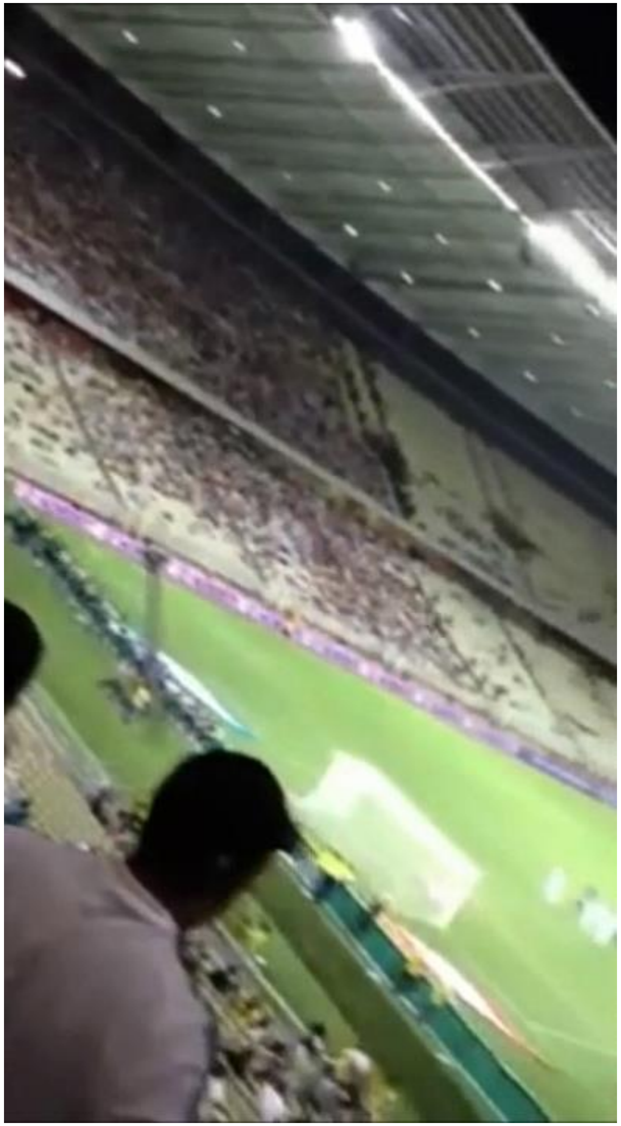
https://www.pwc.pl/pl/pdf/clarity-from-above-pwc.pdf

How to fly a drone

- **Remote control**
  - Requires **line of sight** or **communication link**
  - Requires **skilled pilots**

- **GPS-based navigation**
  - Does not work indoors!
  - Can be unreliable outdoors (e.g., low altitude in urban canyons, under bridges, etc.)

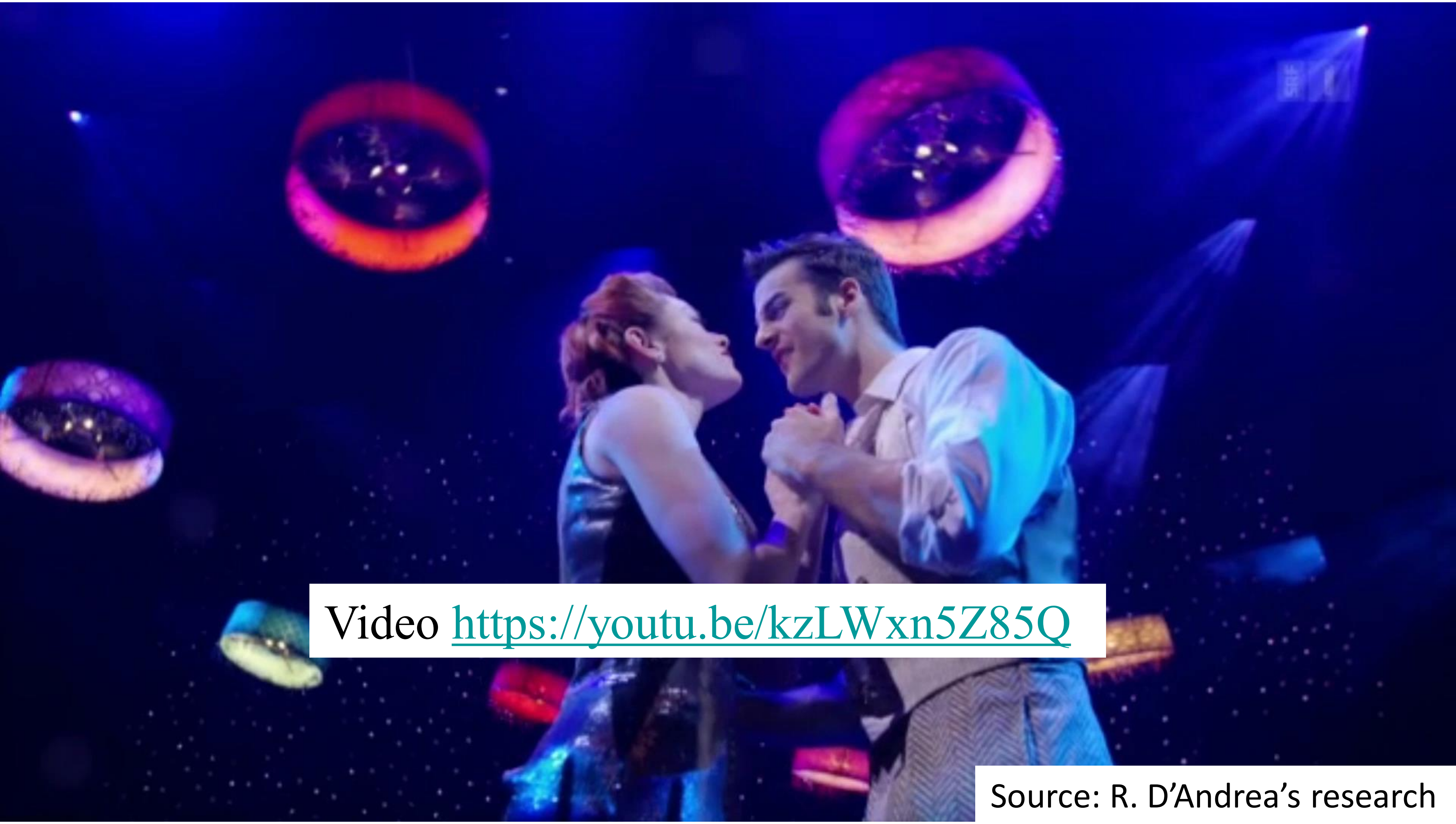




Drone crash during soccer match, Brasilia, 2013

My Goal: Flying Robots to the Rescue!

State of the Art on Autonomous Drone Navigation

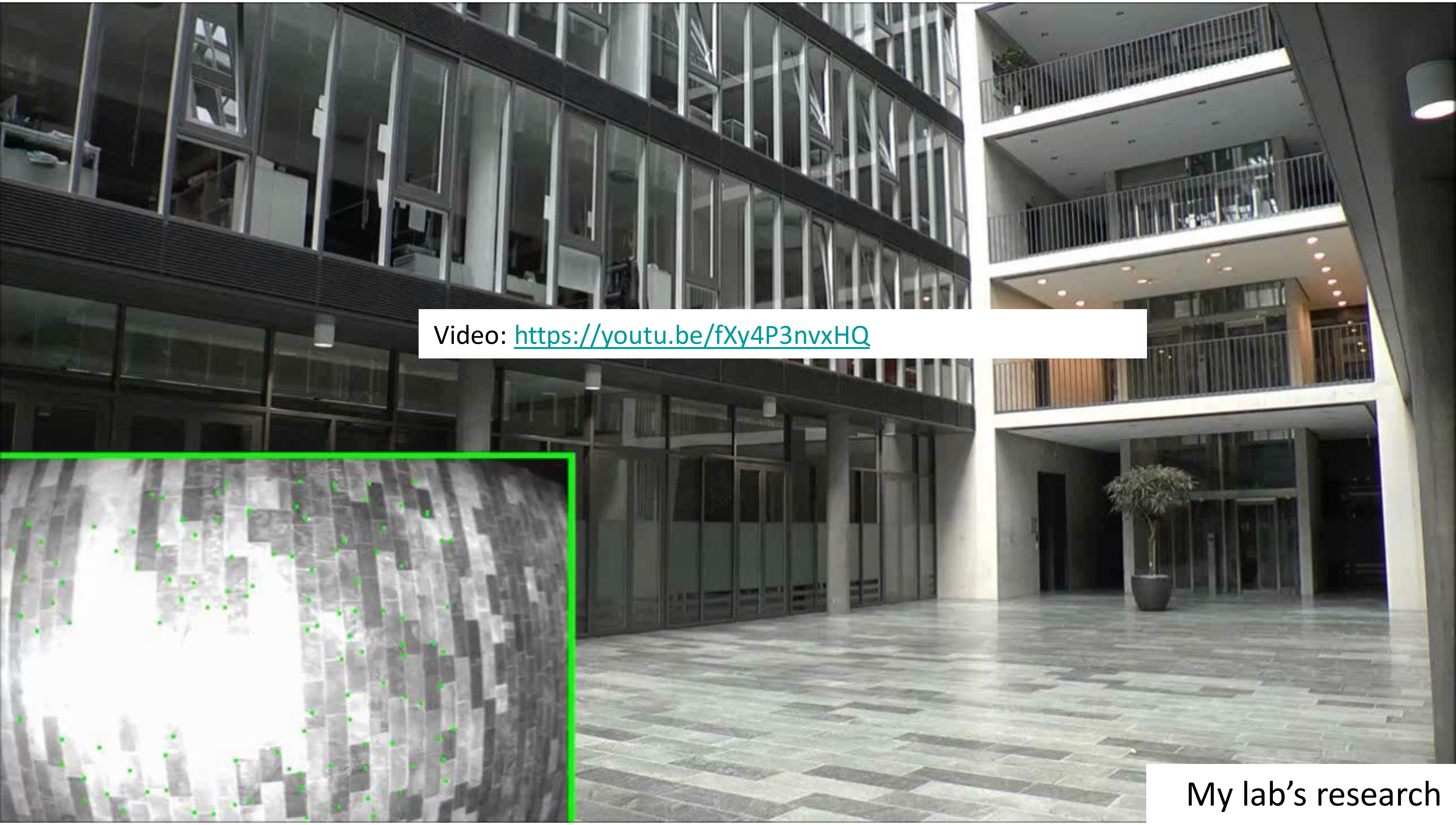


Video https://youtu.be/kzLWxn5Z85Q

Source: R. D’Andrea’s research

… but these robots are completely “**blind**”

State of the Art on Autonomous Drone Navigation

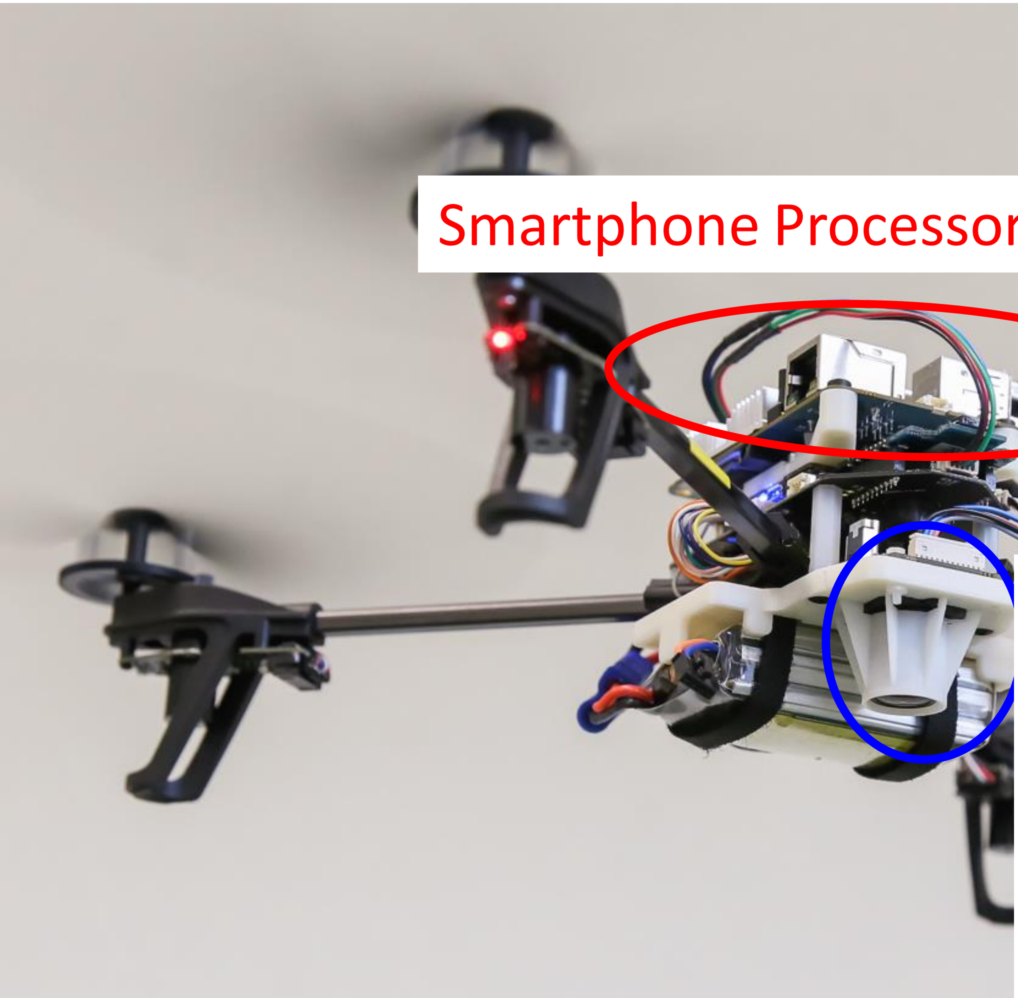


Video: https://youtu.be/fXy4P3nvxHQ

My lab's research

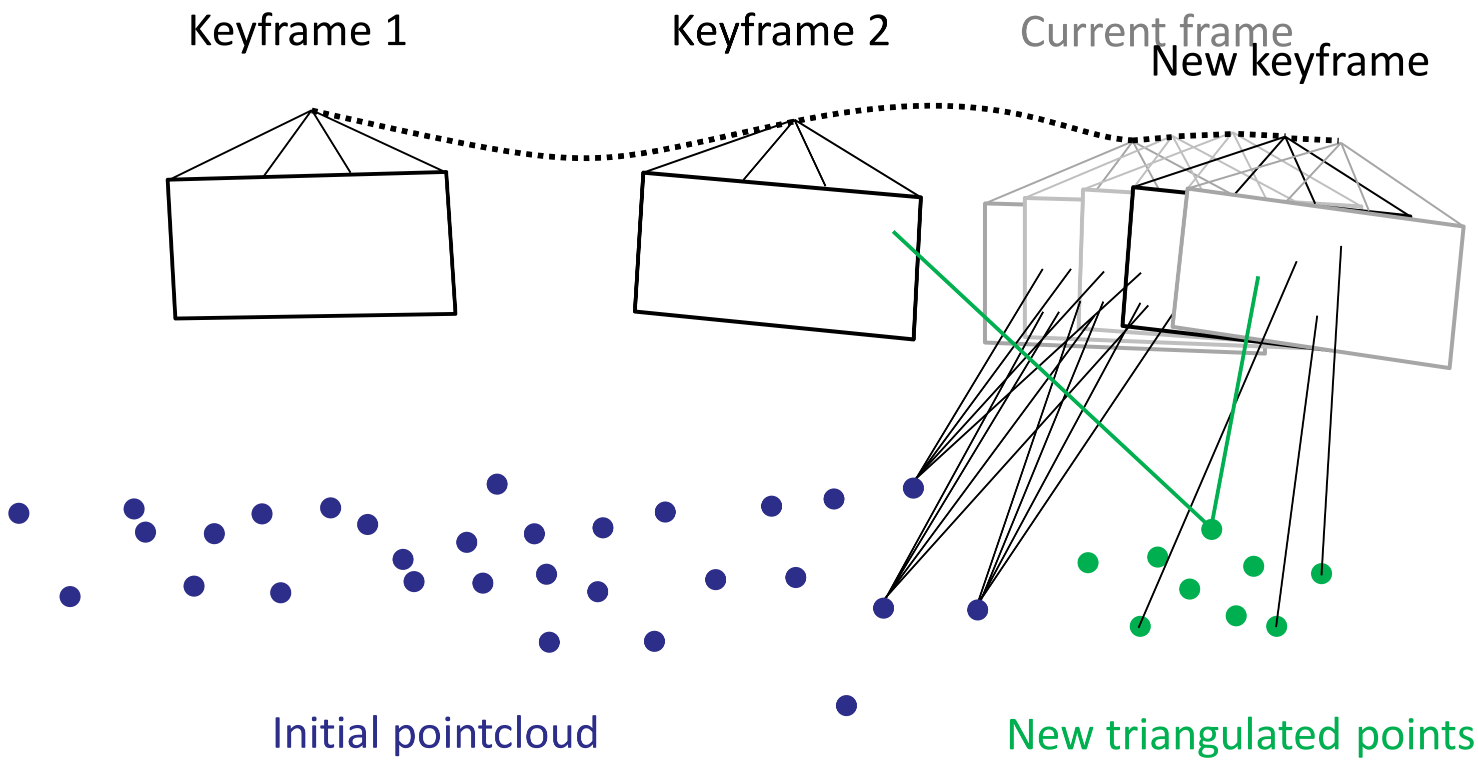
…while this robot can "**see**"

System Overview



450 grams

Simultaneous Localization And Mapping (SLAM)



Also used in several open-source monocular systems:
PTAM, LSD-SLAM, ORBSLAM, OKVIS, DSO, SVO

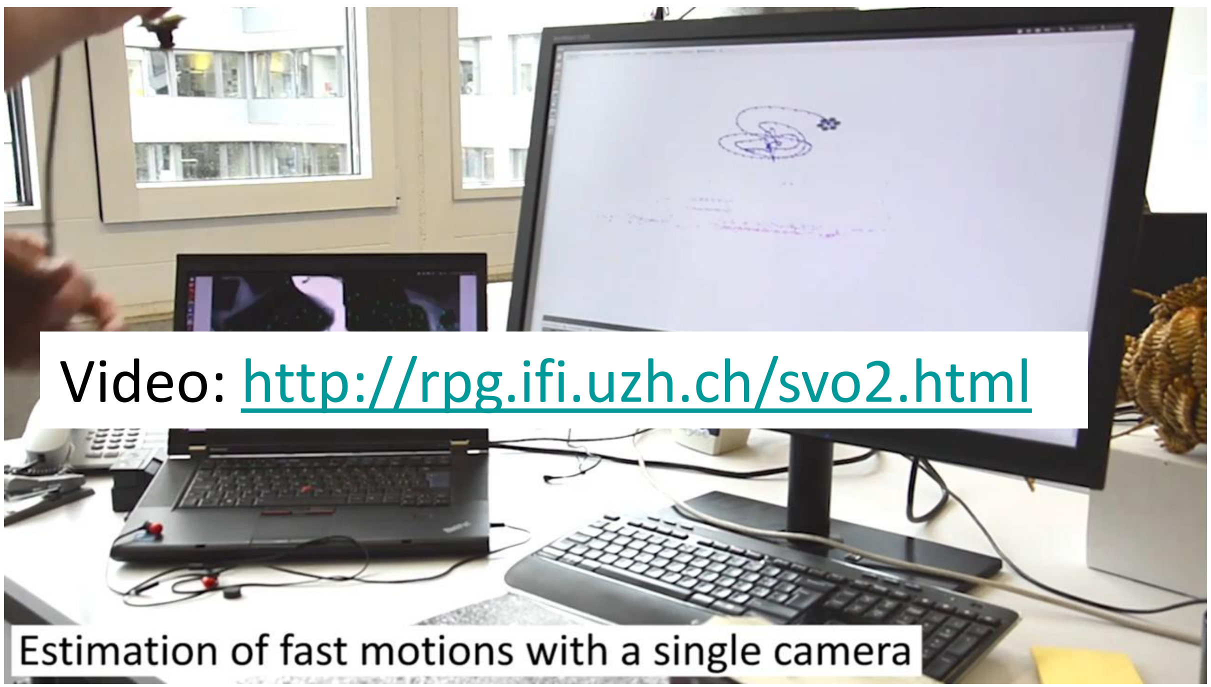
SVO: Semi-direct Visual Odometry [ICRA'14, TRO'17]

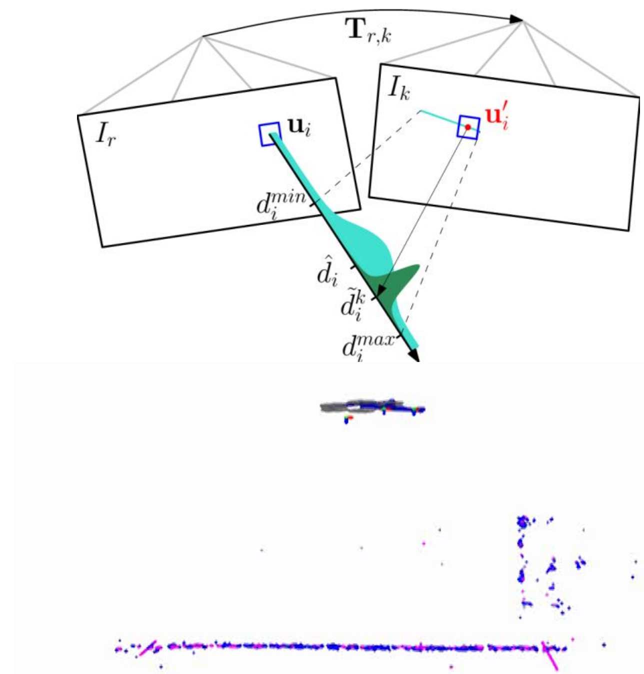
- **Jointly tracks features and 6DoF motion** under mostly-rigid world assumption
- Minimizes both **photometric and geometric** error

Achieves lowest latency & CPU load

- 2.5ms (400 fps) on i7 laptops
- 10ms (100 fps) on smartphones

	Mean	St.D.	CPU@20 fps
SVO Mono	2.53	0.42	55 ±10%
ORB Mono SLAM (No loop closure)	29.81	5.67	187 ±32%
LSD Mono SLAM (No loop closure)	23.23	5.87	236 ±37%



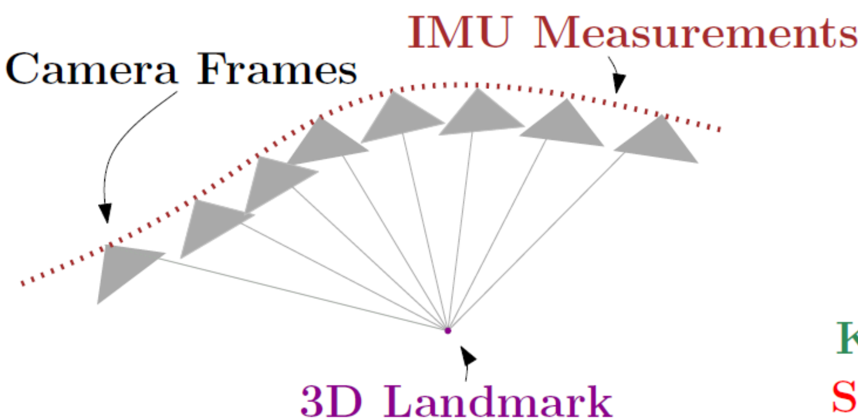


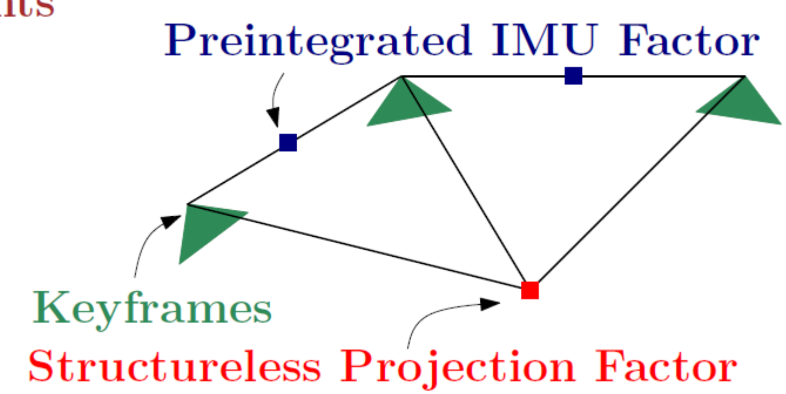
Probabilistic Depth Estimation

Download from : http://rpg.ifi.uzh.ch/svo2.html

Visual-Inertial Odometry via Full-Smoothing

- Full smoothing methods **estimate the entire history of the states** (camera trajectory and 3D landmarks), by solving a large nonlinear optimization problem
- **Superior accuracy over filtering** methods, which only update the last state
- Solved using **phactor graphs** (iSAM): only update variables affected by a new measurement





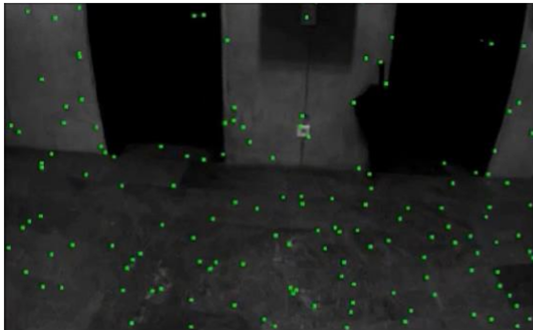
$$\sum_{(i,j)\in\mathcal{K}_k} \|\mathbf{r}_{\mathcal{I}_{ij}}\|^2_{\Sigma_{ij}} + \sum_{i\in\mathcal{K}_k}\sum_{l\in\mathcal{C}_i} \|\mathbf{r}_{\mathcal{C}_{il}}\|^2_{\Sigma_{\mathcal{C}}}$$

IMU residuals *Reprojection residuals*

Open Source
https://bitbucket.org/gtborg/gtsam

1. Forster, Carlone, Dellaert, Scaramuzza, *On-Manifold Preintegration for Real-Time Visual-Inertial Odometry,* IEEE Transactions on Robotics 2017**, TRO'17 Best Paper Award**. PDF, Video
2. Delmerico, Scaramuzza**,** A Benchmark Comparison of Monocular Visual-Inertial Odometry Algorithms, ICRA'18, PDF, Video

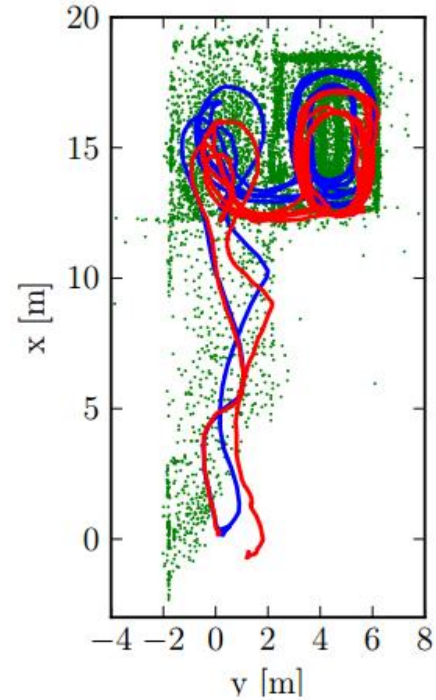
Comparison to Google Tango and OKVIS

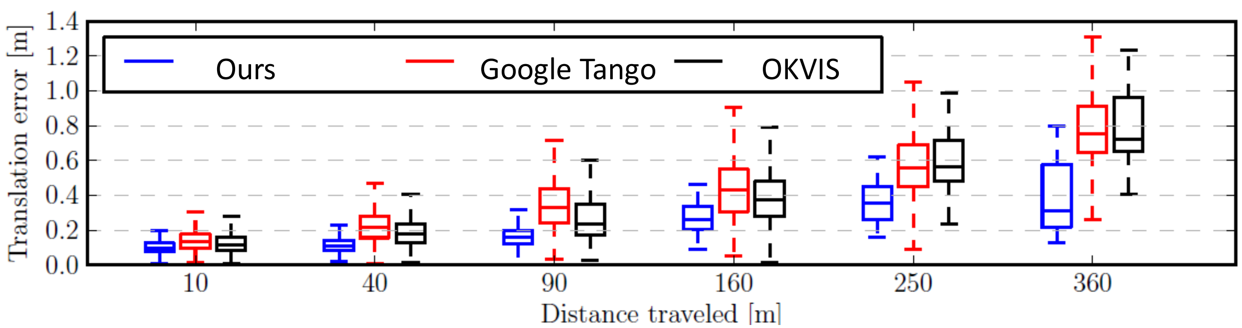


Video: https://youtu.be/CsJkci5lfco

5x

Accuracy: 0.1% of the travel distance





1. Forster, Carlone, Dellaert, Scaramuzza, *On-Manifold Preintegration for Real-Time Visual-Inertial Odometry,* IEEE Transactions on Robotics 2017**, TRO'17 Best Paper Award**. PDF, Video
2. Delmerico, Scaramuzza**,** A Benchmark Comparison of Monocular Visual-Inertial Odometry Algorithms, ICRA'18, PDF, Video

Autonomous Visual Navigation

Video: https://youtu.be/_-p08o_oTO4

Scaramuzza, Fraundorfer, Pollefeys, Siegwart, Achtelick, Weiss, et al., *Vision-Controlled Micro Flying Robots: from System Design to Autonomous Navigation and Mapping in GPS-denied Environments*, RAM'14, PDF

DARPA FLA Program (June 2015 – June 2018)

Mohta, Loianno, Scaramuzza, Daniilidis, Taylor, Kumar, Fast, Autonomous Flight in GPS-denied and Cluttered Environments, *Journal of Field Robotics*, 35 (1), 2018, PDF, Video

Robustness to “Strong Disturbances”!

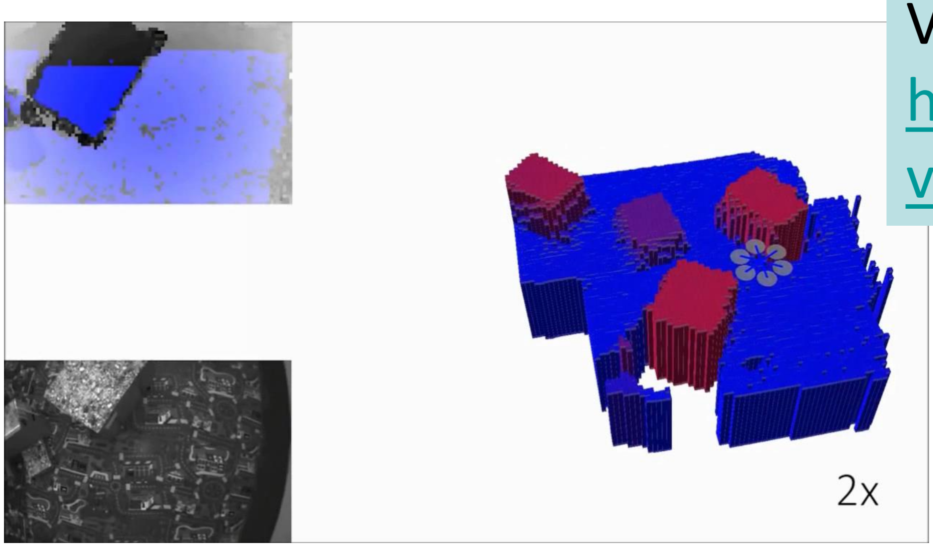
Video: https://youtu.be/pGU1s6Y55JI

Faessler, Fontana, Forster, Scaramuzza, Automatic Re-Initialization and Failure Recovery for Aggressive Flight with a Monocular Vision-Based Quadrotor, ICRA’15. **Featured in IEEE Spectrum.**

Autonomus, Live, Dense Reconstruction

REMODE: probabilistic, REgularized, MOnocular DEnse reconstruction in real time [ICRA'14]
State estimation with SVO 2.0

Video: YouTube
https://www.youtube.com/watch?v=7-kPiWaFYAc



Running at 25Hz onboard (Odroid U3) - Low res.

Running live at 50Hz on laptop GPU – HD res.

Open Source
https://github.com/uzh-rpg/rpg_open_remode

1. Pizzoli et al., *REMODE: Probabilistic, Monocular Dense Reconstruction in Real Time*, ICRA'14]
2. Forster et al., *Appearance-based Active, Monocular, Dense Reconstruction for Micro Aerial Vehicles*, RSS' 14
3. Forster et al., *Continuous On-Board Monocular-Vision-based Elevation Mapping Applied ...*, ICRA'15.
4. Faessler et al., *Autonomous, Vision-based Flight and Live Dense 3D Mapping ...*, JFR'16

Industrial Applications of my Research

Parrot Albris (2016)

Parrot senseFly

- 5 wide angle cameras (one is pan-tilt) + one thermal camera
- “Tap and Fly” Tablet Interface

Video: https://youtu.be/mYKrR8pihAQ

YouTube

5 vision sensors

Powered by SVO 2. (Download: http://rpg.ifi.uzh.ch/svo2.html)

Dacuda 3D (now Magic Leap Zurich)

- Fully immersive VR (running on iPhone)

Dacuda's 3D divison ➔ magic leap

Video: https://youtu.be/mYKrR8pihAQ

YouTube

Zurich-Eye (now Oculus Zurich)

Vision-based Localization and Mapping Solutions for Mobile Robots

Created in Sep. 2015, **became Facebook-Oculus Zurich in Sep. 2016**

Video: https://youtu.be/TSw75JCtjSE

What’s next?

My Dream Robot: Fast, Lightweight, Autonomous!

WARNING!

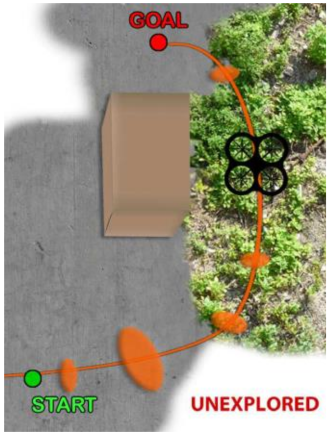
There are 50 drones in this video but 40 are CGI and 10 are controlled via a Motion Capture System ☺

But this is just a vision!
How to get there?

Open Challenges

Perception algorithms are **mature but not robust**

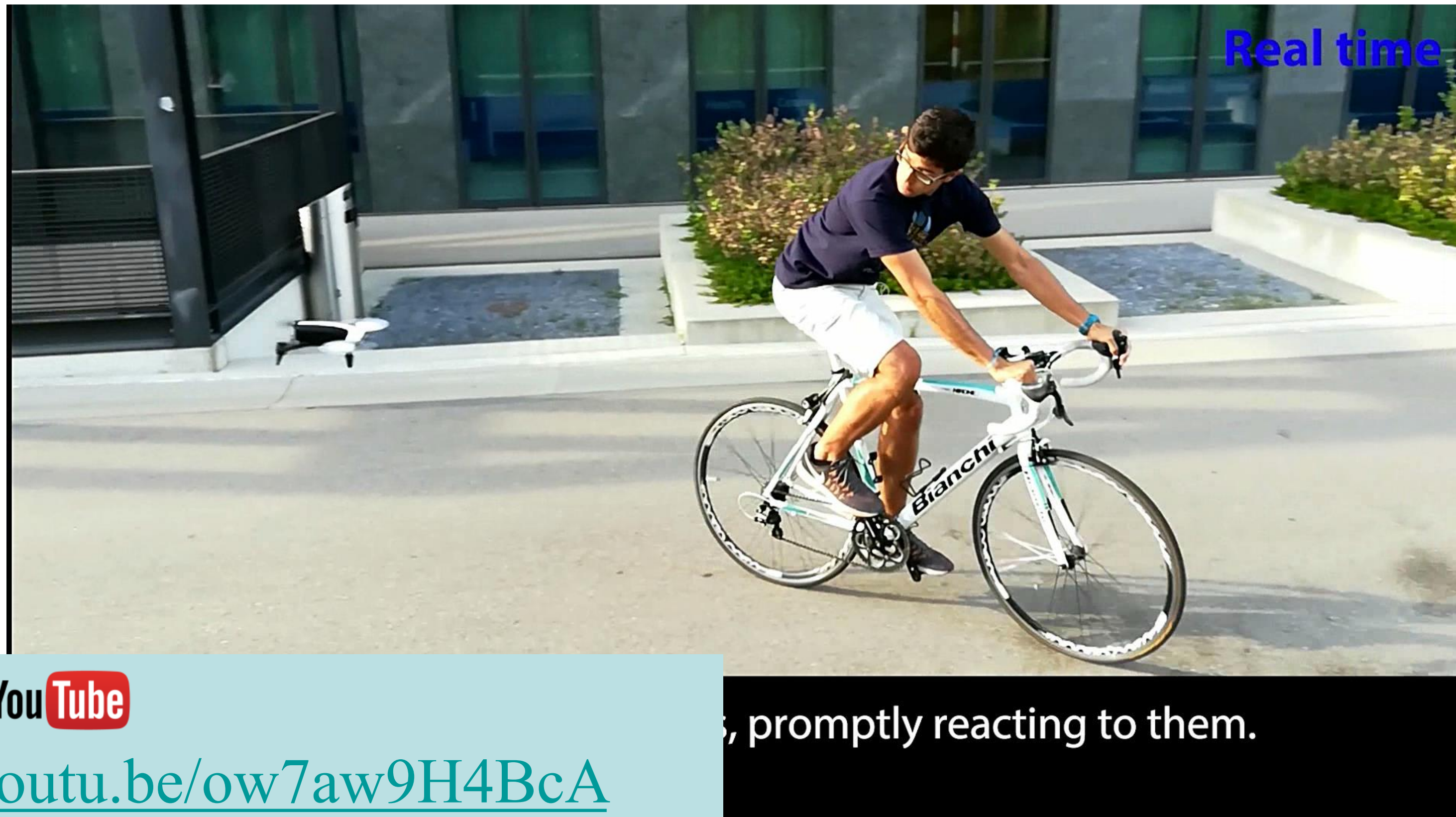
- Problems with **low texture, HDR scenes, motion blur**
- Algorithms and sensors have **big latencies** (50-200 ms) → need faster sensors
- Need **accurate models** of the **sensors** and the **environment**
- **Control & Perception** are often **considered separately** (e.g., perception, state estimation, and planning are treated as separate blocks)



Deep-Learning based Navigation

DroNet: Learning to Fly by Driving

- DroNet learns to follow streets autonomously, without interaction
- The hardest problem in Machine Learning is **data collection**
- **Our idea: learn to fly autonomously by mimicking cars and bicycles!**



Video: YouTube
https://youtu.be/ow7aw9H4BcA

[Loquercio, DroNet: Learning to Fly by Driving, IEEE RAL'18
PDF. **Featured on IEEE Spectrum, MIT Technology Review,** and **Discovery Channel Global**

Code & Datasets:
http://rpg.ifi.uzh.ch/dronet.html

Low-latency, Event-based Vision

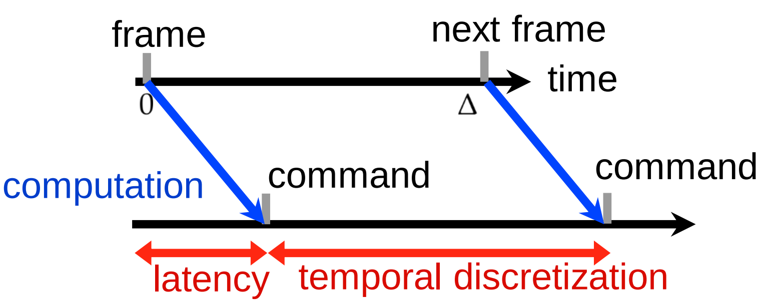
Latency and Agility are tightly coupled!

Current flight maneuvers achieved with onboard cameras are still to slow compared with those attainable by **birds**. We need **faster sensors and algorithms**!

A sparrowhawk catching a garden bird (National Geographic)

What does it take to fly like an eagle?

- Standard cameras are not good enough!



Latency

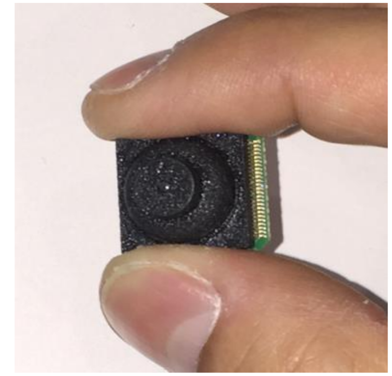
Motion Blur

Low Dynamic Range

- Tasks that need to be done *reliably*, and with *low latency:*
  - Visual odometry (for control)
  - Obstacle detection
  - Recognition

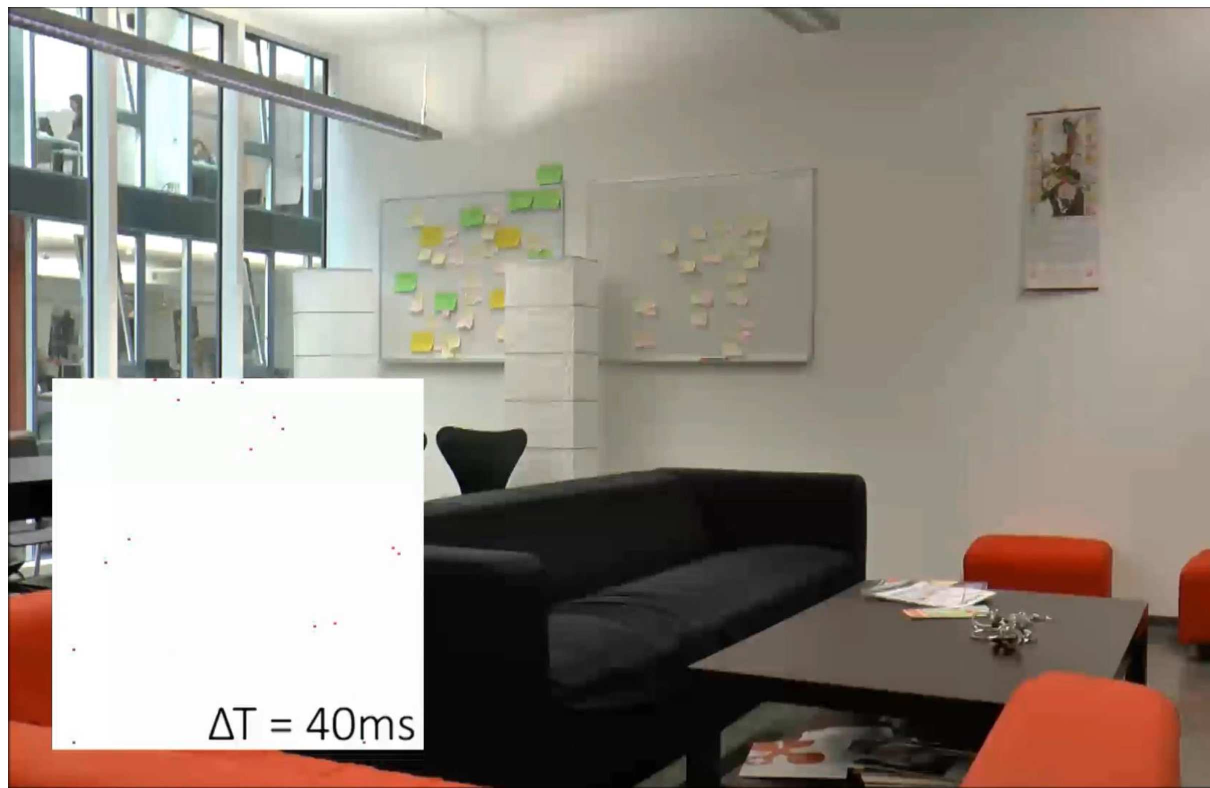
- **Event cameras** promise to solve these three problems!

What is an event camera?

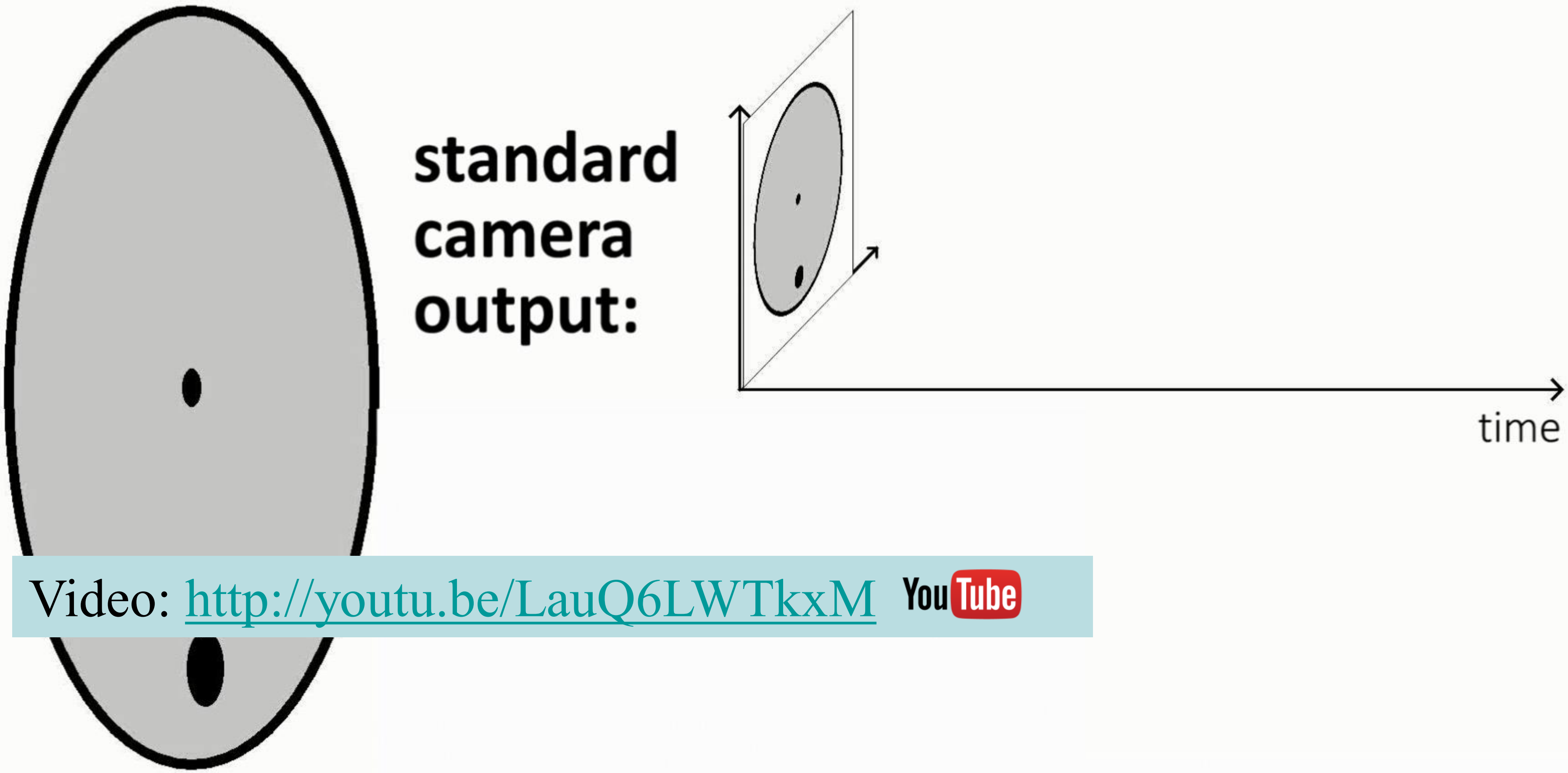


Mini DVS sensor from IniVation.com
Check out their booth in the exhibition hall

- Novel sensor that measures only motion in the scene
- **Low-latency (~ 1 μs)**
- **No motion blur**
- **High dynamic range (140 dB instead of 60 dB)**
- Well-suited for visual odometry
- **But traditional vision algorithms for standard cameras cannot be used!**



Camera vs Event Camera



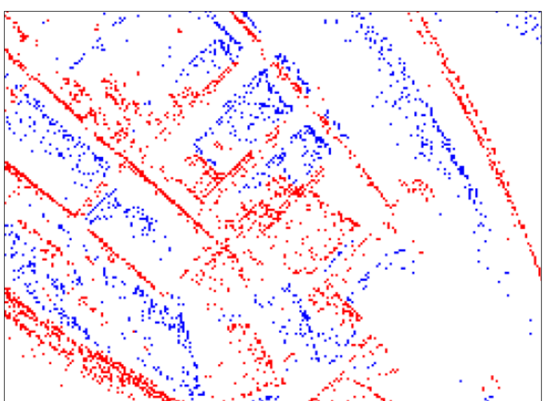
Video: http://youtu.be/LauQ6LWTkxM

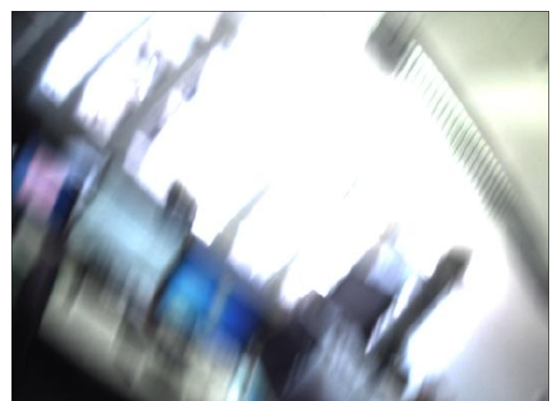
	Event Camera	Standard Camera
Update rate	1MHz and asynchronous	100Hz (synchronous)
Dynamic Range	High (140 dB)	Low (60 dB)
Motion Blur	No	Yes
Absolute intensity	No	Yes
Contrast sensitivity	Low	High

Our idea: combine them!

< 10 years research

> 60 years of research!

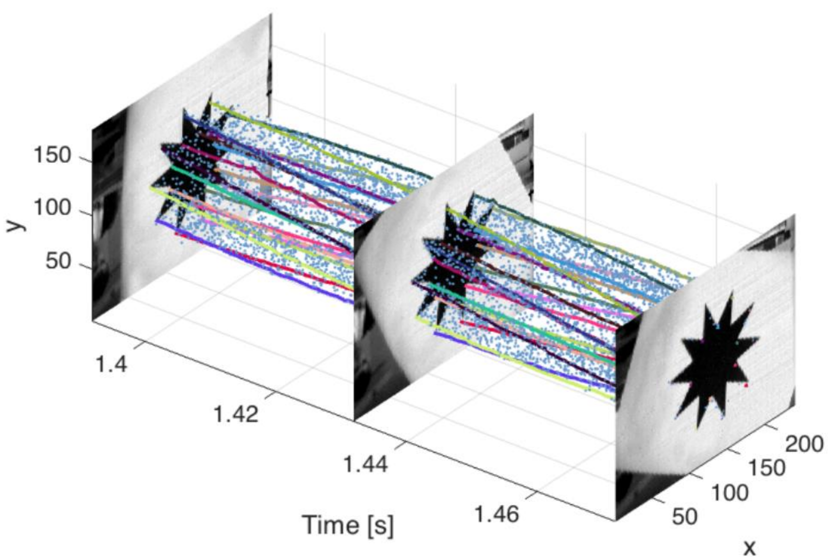


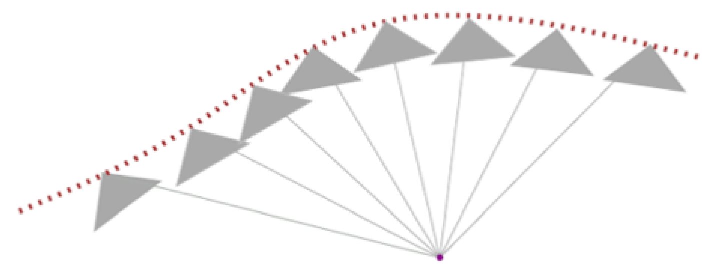


UltimateSLAM: Visual-inertial SLAM with Events + Frames + IMU

Feature tracking from Events and Frames

Visual-inertial Fusion

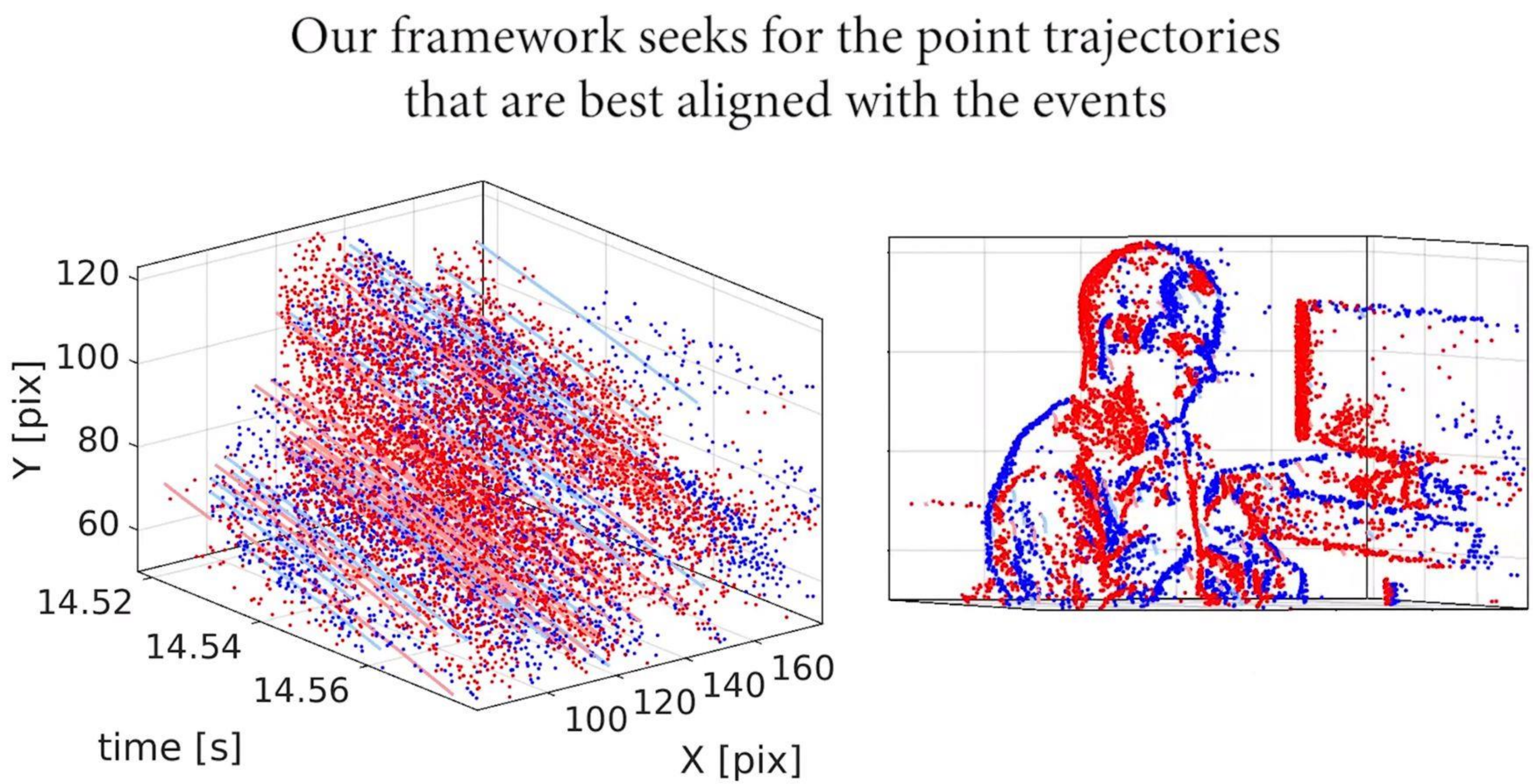




Rosinol, Ultimate SLAM? Combining Events, Images, and IMU for Robust Visual SLAM in HDR and High Speed Scenarios, IEEE RAL'18

Tracking by Contrast Maximization [CVPR’18]

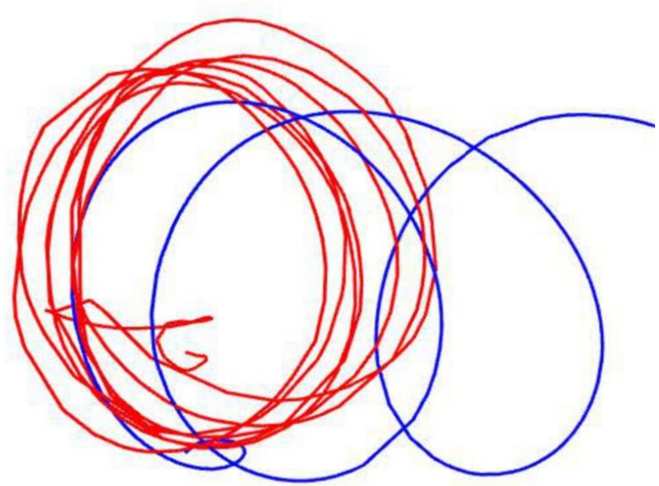
- Directly estimate the motion curves that align the events

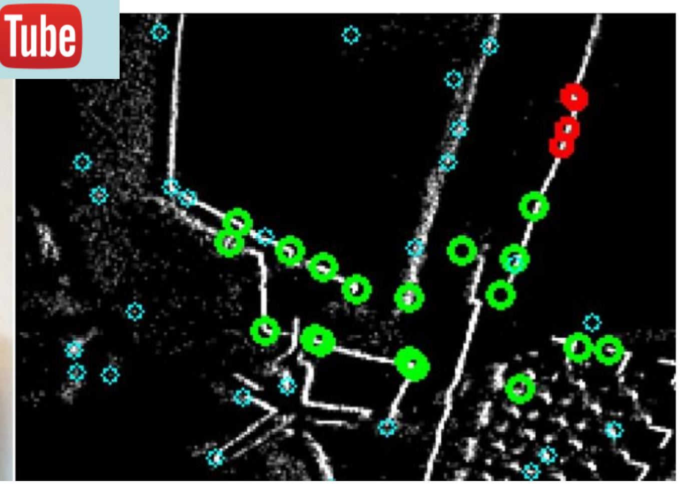


Gallego, Rebecq, Scaramuzza, A Unifying Contrast Maximization Framework for Event Cameras, with Applications to Motion, Depth, and Optical Flow Estimation, CVPR’18, Spotlight talk, PDF, YouTube

UltimateSLAM: Events + Frames + IMU

Tightly coupled fusion. Runs in real time on a smartphone processor.



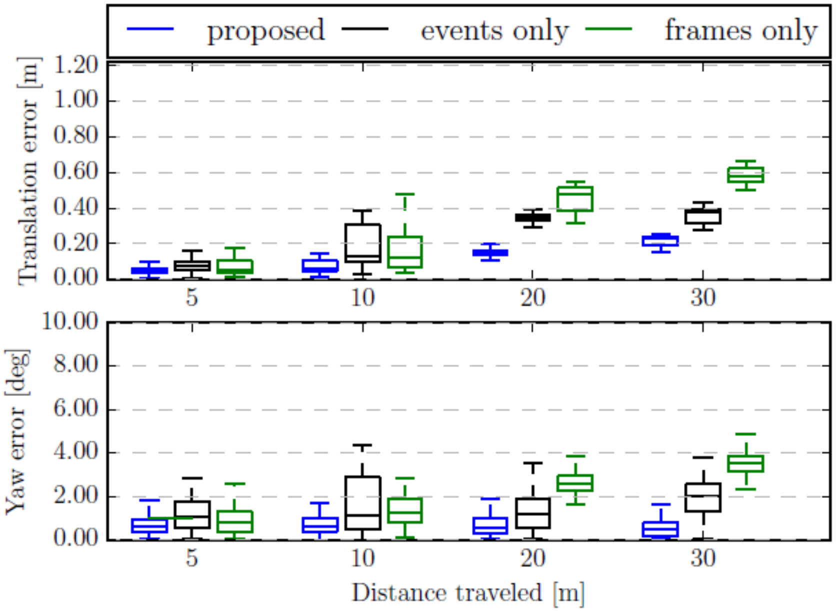


Video: https://youtu.be/DyJd3a01Zlw YouTube

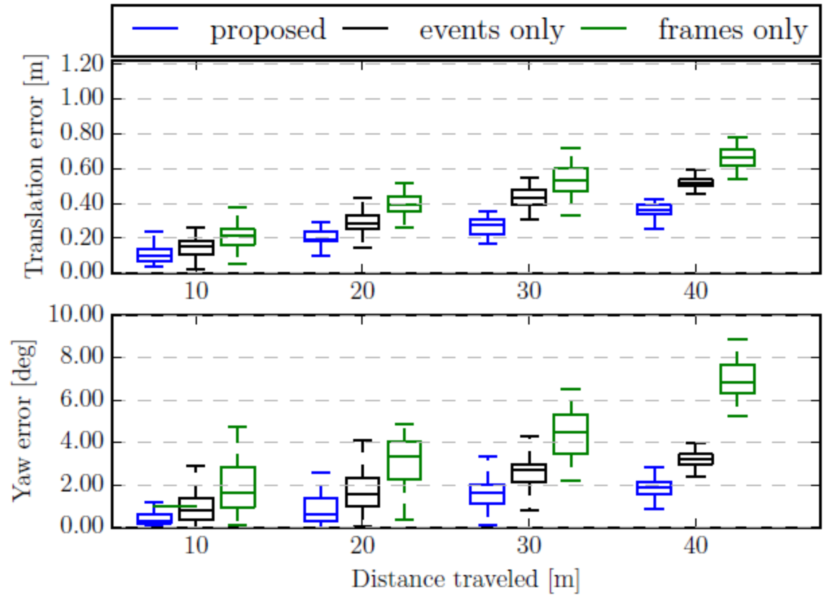
Rosinol et al., Ultimate SLAM? Combining Events, Images, and IMU for Robust Visual SLAM in HDR and High Speed Scenarios, IEEE RAL'18, PDF

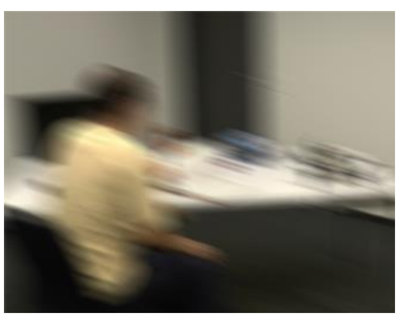
85% accuracy gain over frame-based visual-inertial odometry

High-speed sequence



HDR sequence

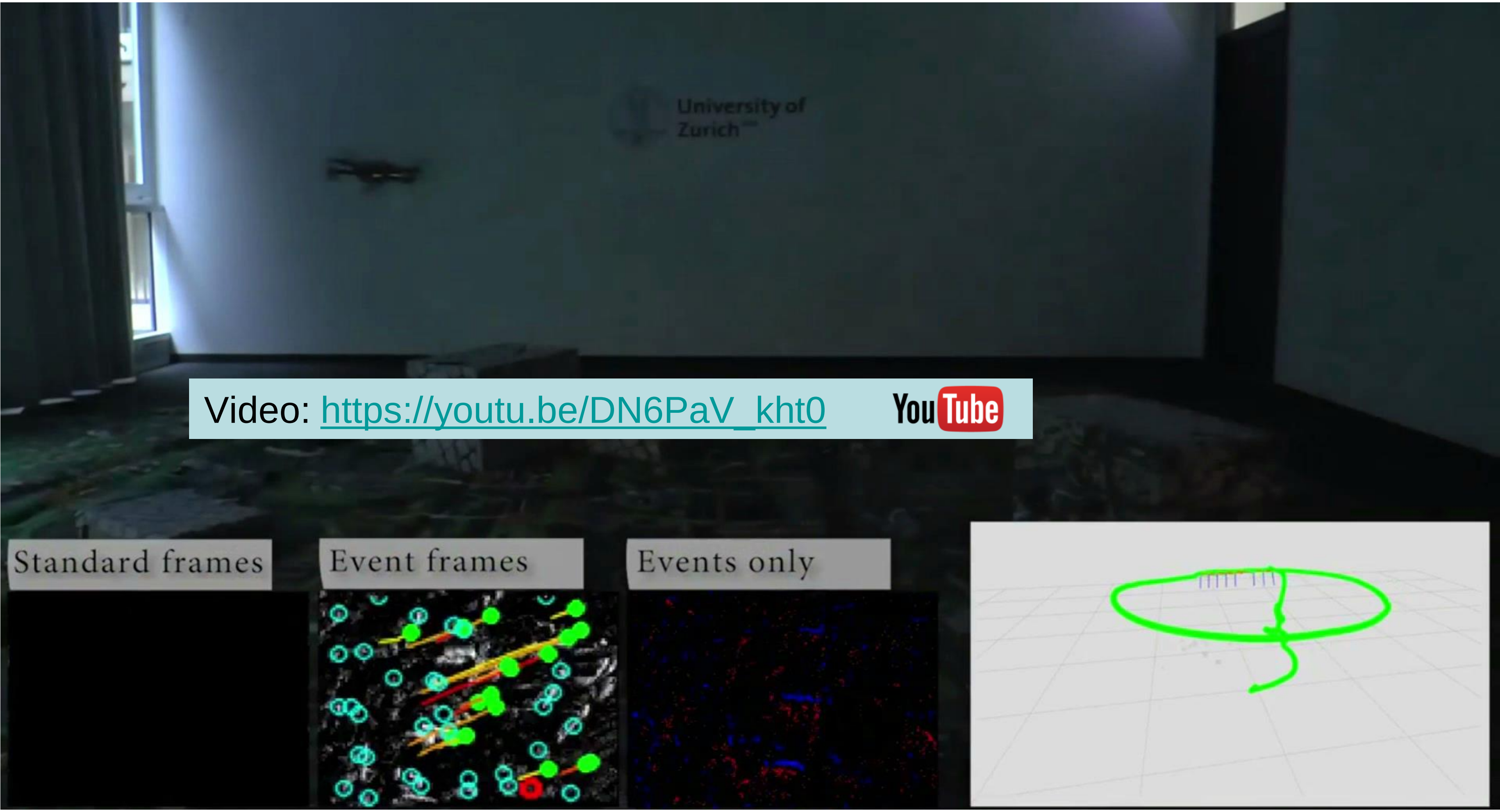




Rosinol et al., Ultimate SLAM? Combining Events, Images, and IMU for Robust Visual SLAM in HDR and High Speed Scenarios, IEEE RAL'18, [PDF]

UltimateSLAM: Autonomous Navigation in Low Light

Fully onboard (Odroid), event camera + IMU, tightly coupled



Video: https://youtu.be/DN6PaV_kht0 YouTube

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

Low-latency Obstacle Avoidance (ongoing work)

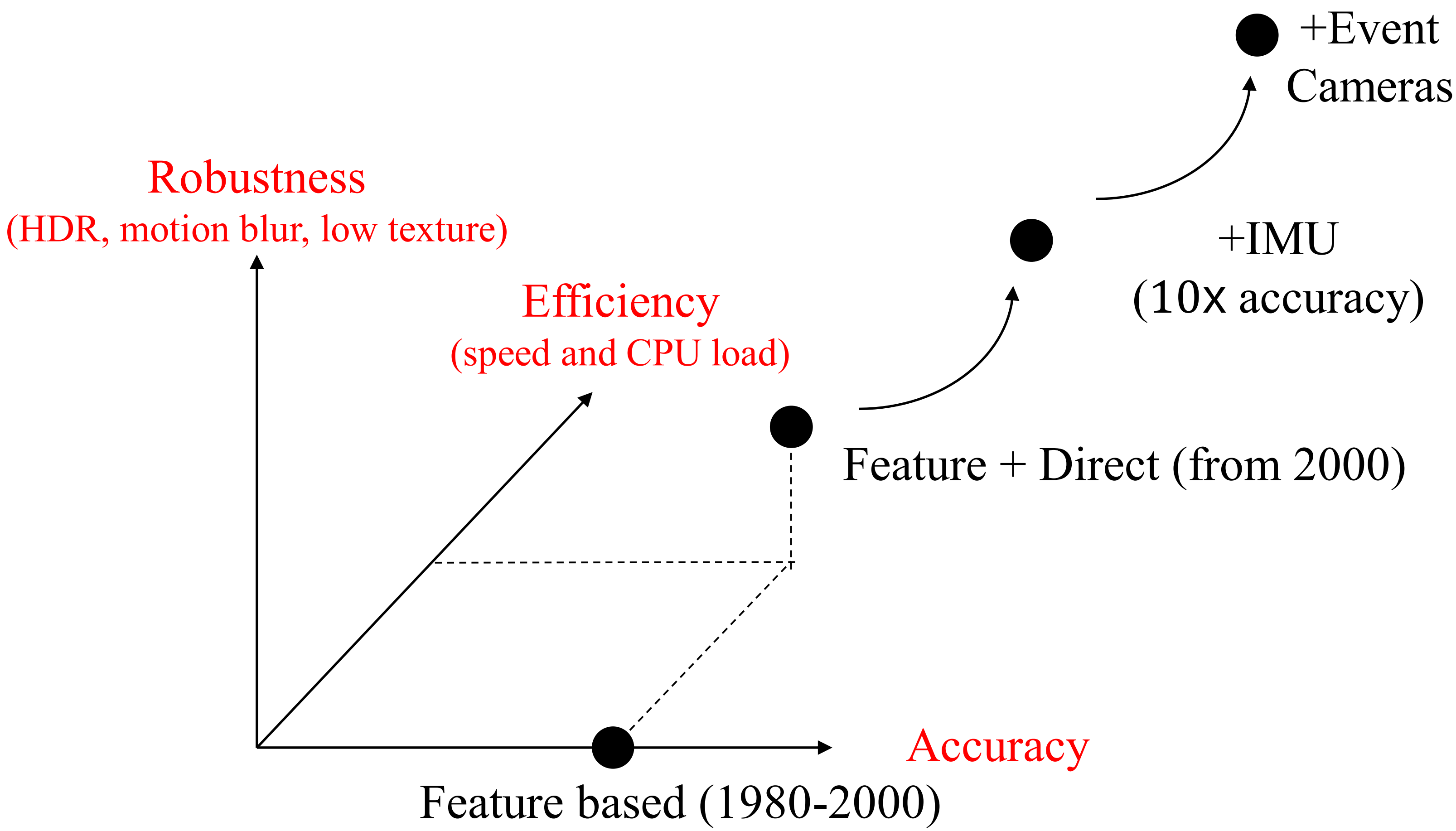
In collaboration with Insightness company (makes event cameras and collision avoidance systems for drones)

Video: https://youtu.be/6aGx-zBSzRA

Conclusions

- Agile flight (**like birds**) is still far (10 years?)
- **Perception and control** need to be considered **jointly!**
- SLAM theory is **well established**
  - Biggest challenges today are **reliability and robustness** to**:**
    - High-dynamic-range scenes
    - High-speed motion
    - Low-texture scenes
    - Dynamic environments
- **Machine Learning can exploit context & provide robustness and invariance to nuisances**
- **Event cameras** are revolutionary and provide:
  - Robustness to **high speed motion** and **high-dynamic-range scenes**
  - Allow **low-latency** control (ongoing work)
  - **Intellectually challenging**: standard cameras have been studied for 50 years! → need of a change!

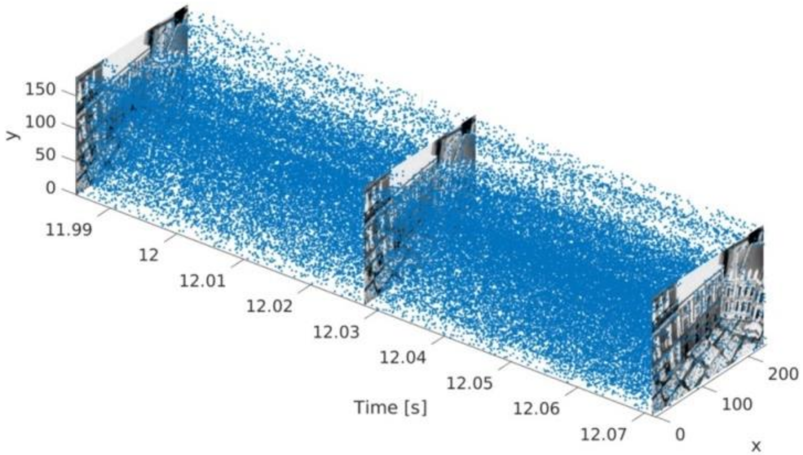
A Short Recap of the last 30 years of Visual Inertial SLAM



C. Cadena, L. Carlone, H. Carrillo, Y. Latif, D. Scaramuzza, J. Neira, I.D. Reid, J.J. Leonard
Past, Present, and Future of Simultaneous Localization and Mapping: Toward the Robust-Perception Age
IEEE Transactions on Robotics, 2016.

Event Camera Dataset and Simulator [IJRR'17]

- **Publicly available**: http://rpg.ifi.uzh.ch/davis_data.html
- **First event camera dataset** specifically made **for VO and SLAM**
- **Many diverse scenes:** HDR, Indoors, Outdoors, High-speed
- **Blender simulator of event cameras**
- Includes
    - **IMU**
    - **Frames**
    - **Events**
    - **Ground truth** from a motion capture system



➢ Complete of code, papers, videos, companies:

- https://github.com/uzh-rpg/event-based_vision_resources

Mueggler, Rebecq, Gallego, Delbruck, Scaramuzza,
The Event Camera Dataset and Simulator: Event-based Data for Pose Estimation, Visual Odometry, and SLAM, International Journal of Robotics Research, IJRR, 2017.

Thanks!

Code, datasets, publications, videos: http://rpg.ifi.uzh.ch

- My lab homepage: http://rpg.ifi.uzh.ch/
- Publications: http://rpg.ifi.uzh.ch/publications.html
- Software & Datasets: http://rpg.ifi.uzh.ch/software_datasets.html
- YouTube: https://www.youtube.com/user/ailabRPG/videos