Vol agile avec des micro-robots volants contrôlés par vision

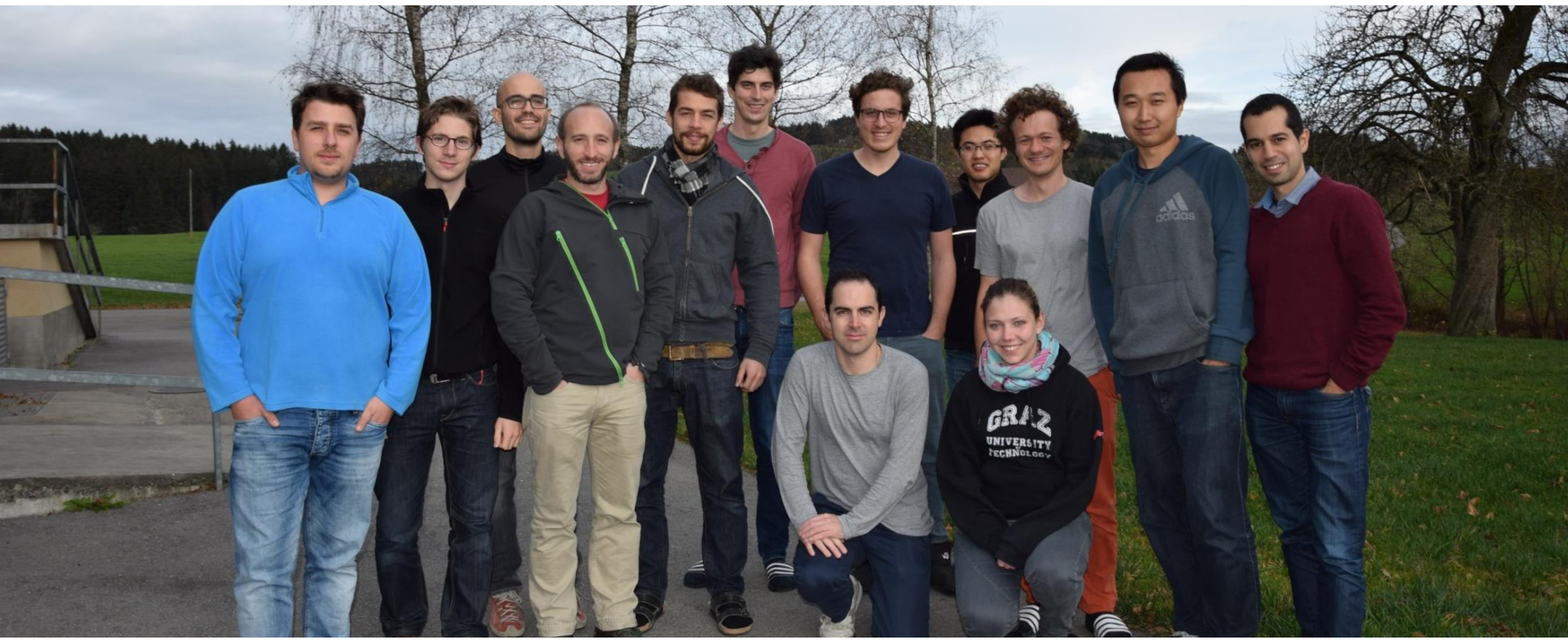
From Active Perception to Event-based Vision

Henri Rebecq
from Prof. Davide Scaramuzza's lab

GT UAV – 17 Novembre 2016, Paris

My Research Group

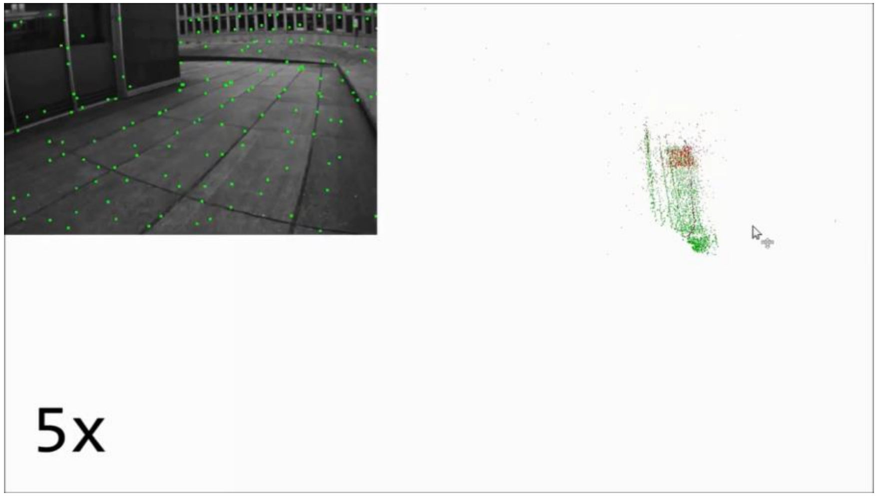
ROBOTICS &
PERCEPTION
GROUP



Our Research Areas

Visual-Inertial State Estimation

[IJCV'11, PAMI'13, RSS'15]



Vision-based Navigation of Flying Robots

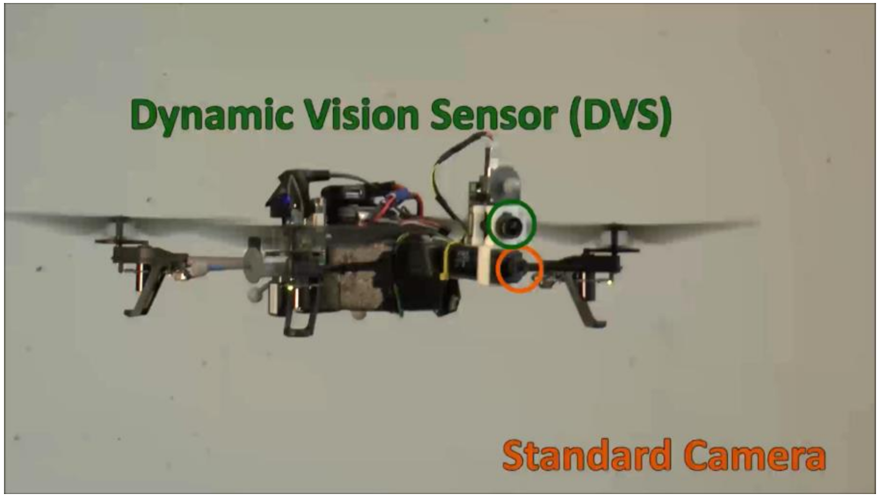
[AURO'12, RAM'14, JFR'15]

Dense Reconstruction

[ICRA'14, JFR'15]

Event-based Vision

[IROS'14, RSS'15]



Our Vision: Flying Robots to the Rescue!

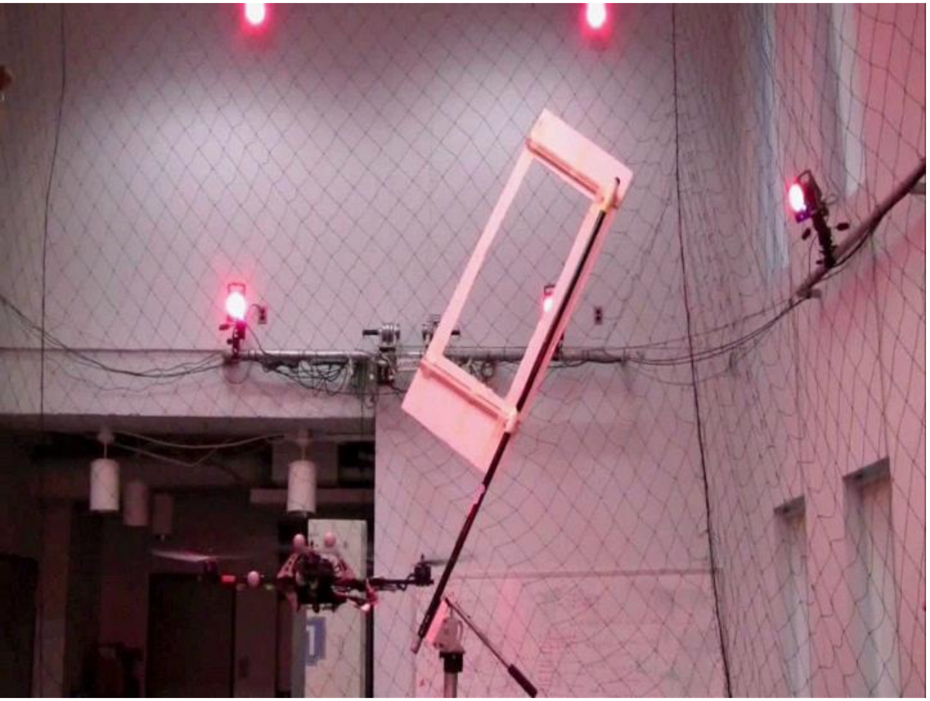
Our Dream Robot

Agile, lightweight drones rapidly navigating to accomplish a given task

- **Challenges**: fast, lightweight, agile (low-latency perception & control)

Video credit: LEXUS commercial, 2013

State of the Art on GPS-denied Navigation

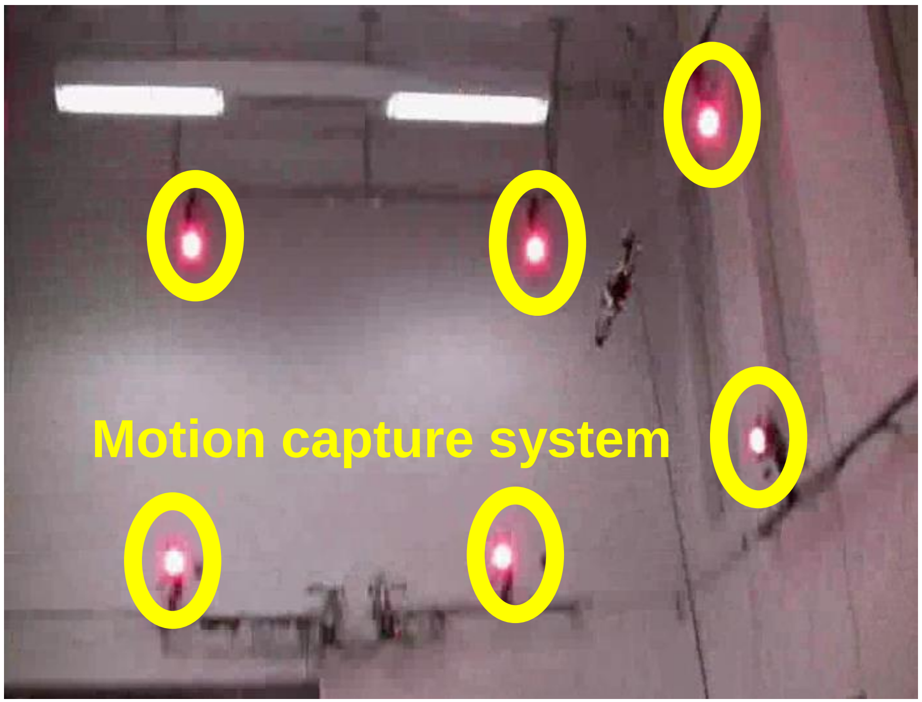


Mellinger, Michael, **Kumar**

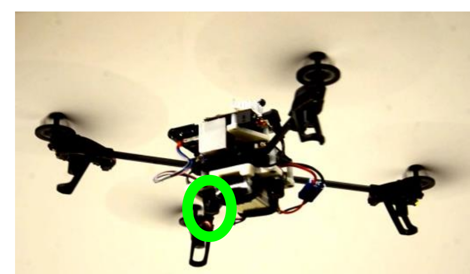
Fontana, Faessler, **Scaramuzza**

How do we Localize without GPS ?

This robot is «***blind***»



This robot can «***see***»



Problems with Vision-controlled Drones

Drones have the **potential to navigate quickly** through unstructured environments but

- Autonomous operation is currently **restricted to controlled environments**
- **Vision-based** maneuvers still **slow** and **inaccurate** wrt motion-capture systems

Why?

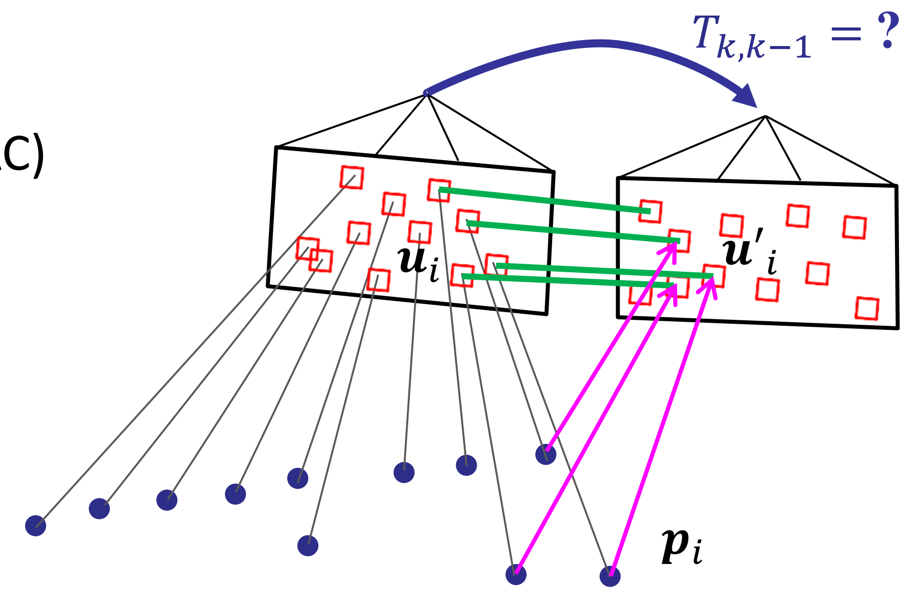
- Perception algorithms are **mature but not robust**
  - Unlike mocap systems, **localization accuracy** depends on **distance & texture**!
  - **Control & Perception** have been mostly **considered separately!**
  - Algorithms and sensors have **big latencies** (50-200 ms) → need faster sensors!

Outline

- Visual-Inertial State Estimation
- Active Vision
- Low-latency, Event-based Vision

Visual-Inertial State Estimation

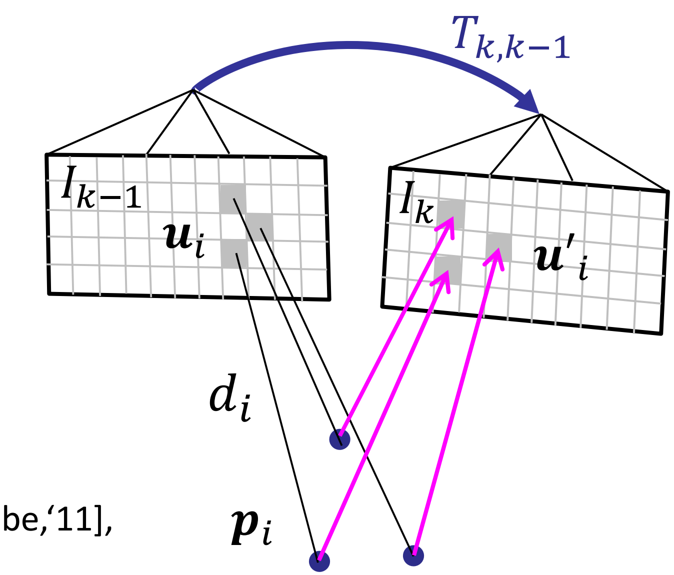
Feature-based methods



1. Extract & match features (+RANSAC)
2. Minimize **Reprojection error** minimization

$$T_{k,k-1} = \arg\min_{T} \sum_{i} \| \boldsymbol{u}'_i - \pi(\boldsymbol{p}_i) \|^2_{\Sigma}$$

Direct methods



1. Minimize **photometric error**

$$T_{k,k-1} = \arg\min_{T} \sum_{i} \| I_k(\boldsymbol{u}'_i) - I_{k-1}(\boldsymbol{u}_i) \|^2_{\sigma}$$

where $\boldsymbol{u}'_i = \pi\left(T \cdot (\pi^{-1}(\boldsymbol{u}_i) \cdot d)\right)$

[Jin,Favaro,Soatto'03] [Silveira, Malis, Rives, TRO'08], [DTAM, Newcombe,'11], [LSD-SLAM, Engel'14], [SVO, Forster'14]

Feature-based methods

1. Extract & match features (+RANSAC)
2. Minimize **Reprojection error** minimization

✓ Large frame-to-frame motions

× Slow due to costly feature extraction and matching

Our solution:

SVO: *Semi-direct* Visual Odometry [ICRA'14]

Combines feature-based and direct methods

$$T_{k,k-1} = \arg\min_{T} \sum_{i} \| I_k(\boldsymbol{u}'_i) - I_{k-1}(\boldsymbol{u}_i) \|_{\sigma}^{2}$$

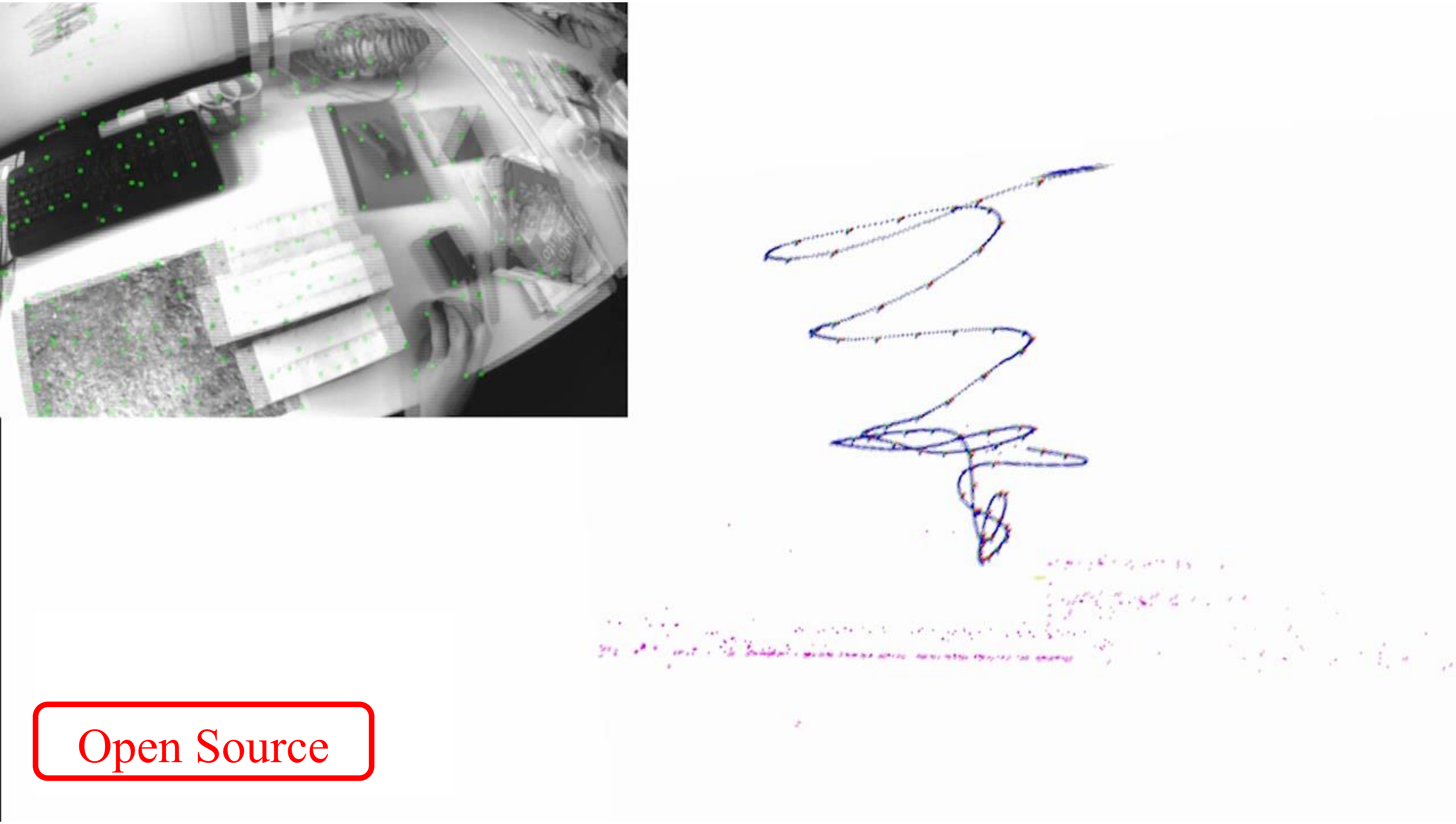
where $\boldsymbol{u}'_i = \pi\big(T \cdot (\pi^{-1}(\boldsymbol{u}_i) \cdot d)\big)$

per frame

× Limited frame-to-frame motion

[Jin,Favaro,Soatto'03] [Silveira, Malis, Rives, TRO'08], [DTAM, Newcombe'11, [LSD-SLAM, Engel'14], [SVO, Forster'14]

SVO Results: Fast and Abrupt Motions



Open Source

[Forster, Pizzoli, Scaramuzza, «SVO: Semi Direct Visual Odometry», ICRA'14]

Processing Times of SVO

Laptop (Intel i7, 2.8 GHz)

400 frames per second

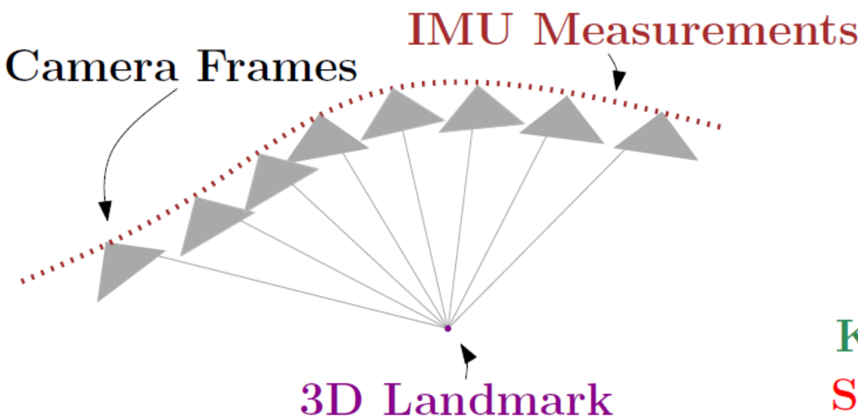
Embedded ARM Cortex-A9, 1.7 GHz

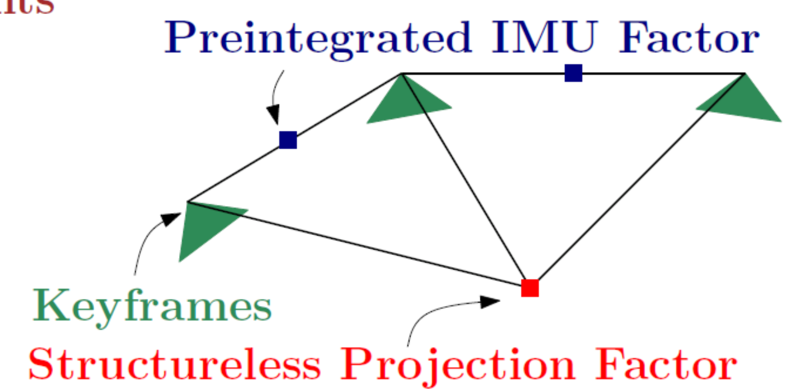
Up to 70 frames per second

[Forster et al., «SVO: Semi Direct Visual Odometry for Monocular and Multicamera Systems», TRO’16]

Visual-Inertial Fusion [RSS'15]

- Fusion solved as a *non-linear optimization problem* (no filters)
- Increased accuracy over filtering methods





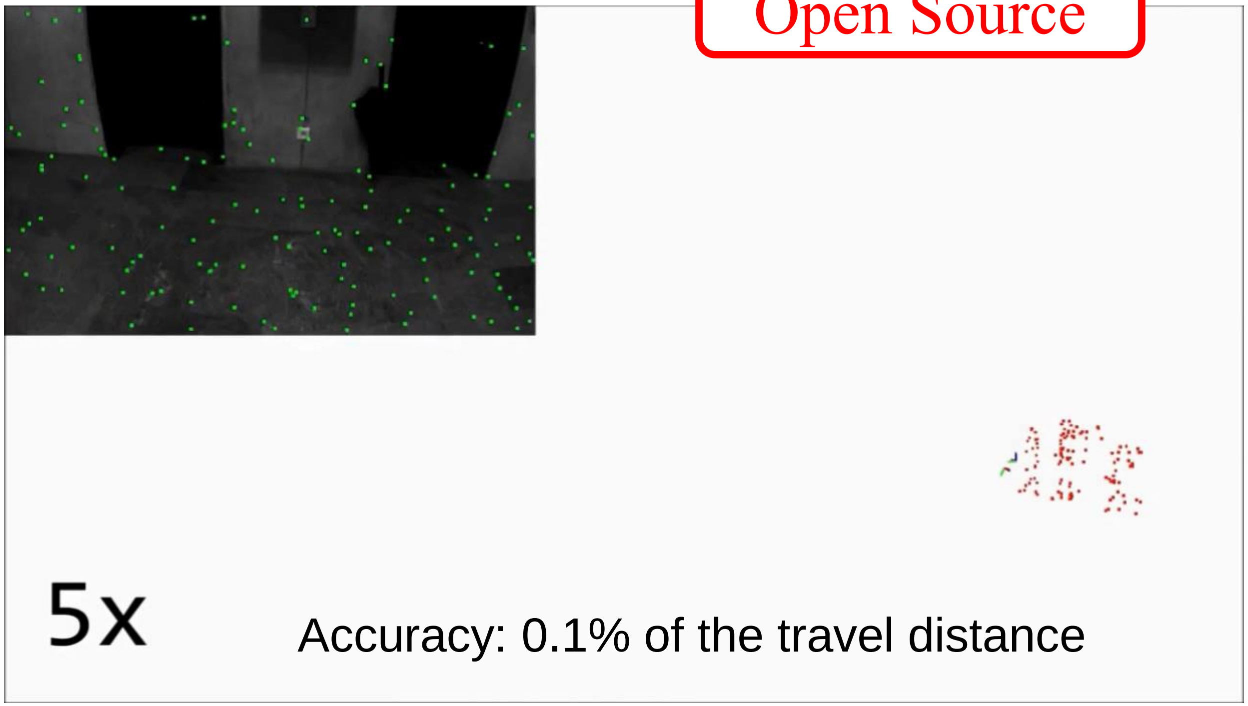
$$\sum_{(i,j)\in\mathcal{K}_k} \|\mathbf{r}_{\mathcal{I}_{ij}}\|^2_{\Sigma_{ij}} + \sum_{i\in\mathcal{K}_k}\sum_{l\in\mathcal{C}_i} \|\mathbf{r}_{\mathcal{C}_{il}}\|^2_{\Sigma_{\mathcal{C}}}$$

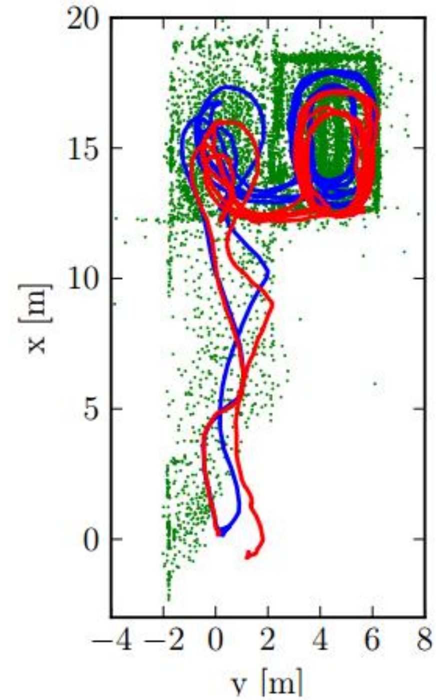
IMU residuals *Reprojection residuals*

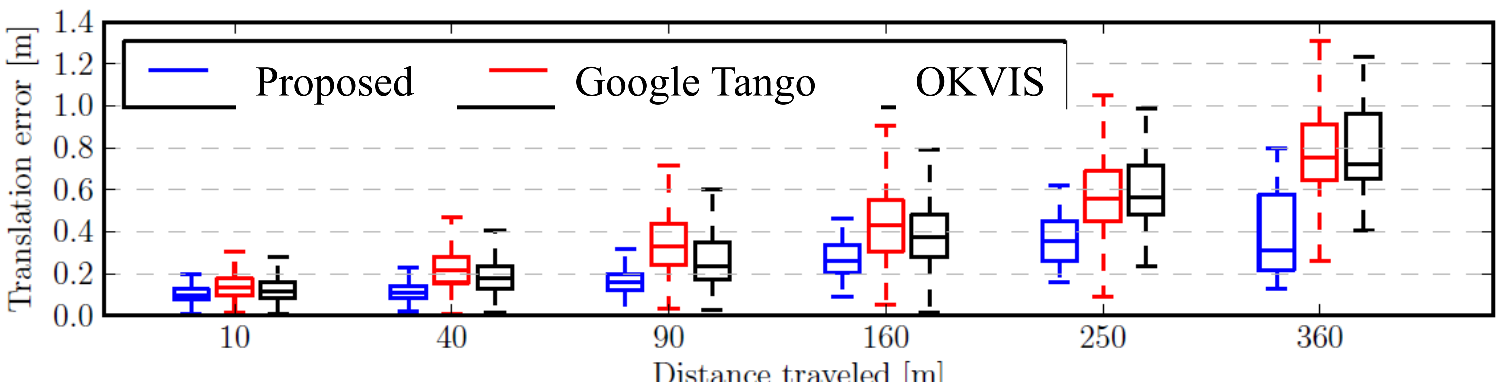
Forster, Carlone, Dellaert, Scaramuzza, IMU Preintegration on Manifold for efficient Visual-Inertial Maximum-a-Posteriori Estimation, *Robotics Science and Systens*'15, **Best Paper Award Finalist**

Comparison with Previous Works

Open Source



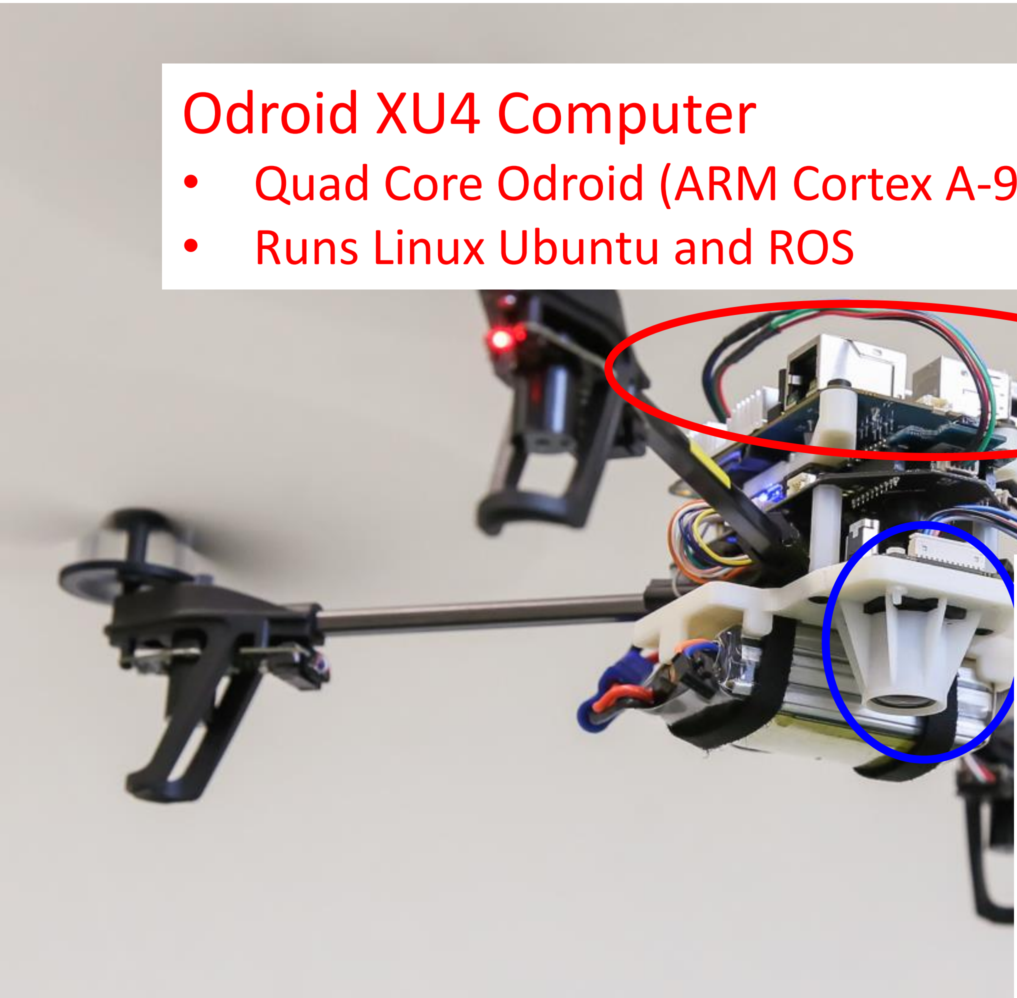




Forster, Carlone, Dellaert, Scaramuzza, IMU Preintegration on Manifold for efficient Visual-Inertial Maximum-a-Posteriori Estimation, *Robotics Science and Systens*'15, **Best Paper Award Finalist**

Integration on a Quadrotor Platform

Quadrotor System



Odroid XU4 Computer

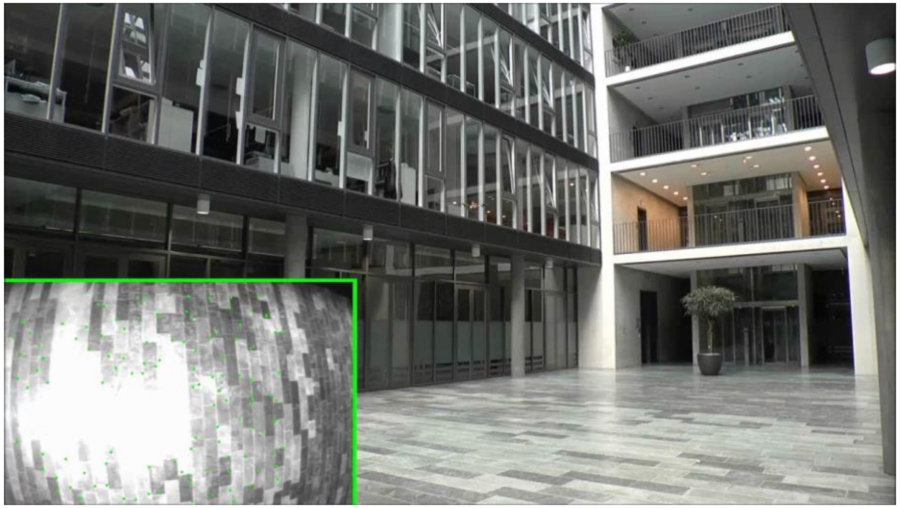
- Quad Core Odroid (ARM Cortex A-9) used in Samsung Galaxy S4 phones
- Runs Linux Ubuntu and ROS

450 grams

Indoors and outdoors experiments

Position error: 5 mm, height: 1.5 m – Down-looking camera

Speed: 4 m/s, height: 1.5 m – Down-looking camera



Faessler, Fontana, Forster, Mueggler, Pizzoli, Scaramuzza, Autonomous, Vision-based Flight and Live Dense 3D Mapping with a Quadrotor Micro Aerial Vehicle, **Journal of Field Robotics, 2015**.

Application: Autonomous Inspection of Bridges and Power Masts

Project with Parrot: Autonomous vision-based navigation

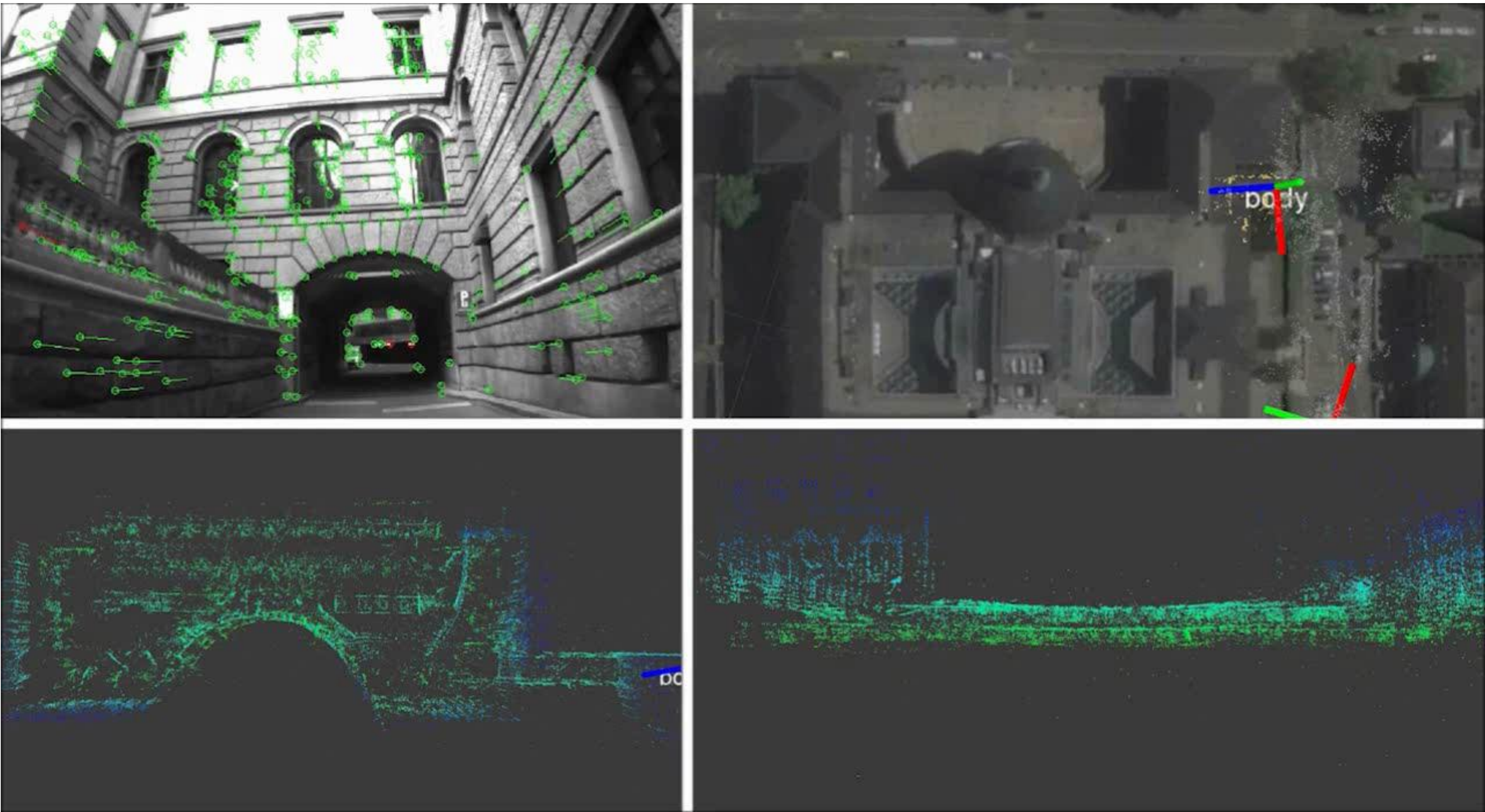
Parrot senseFly

Albris drone

Startup: Zurich-Eye – *www.zurich-eye.com*

Vision-based Localization and Mapping Solutions for Mobile Robots

Founded in Sep. 2015, **became part of Facebook-Oculus in Sep. 2016**

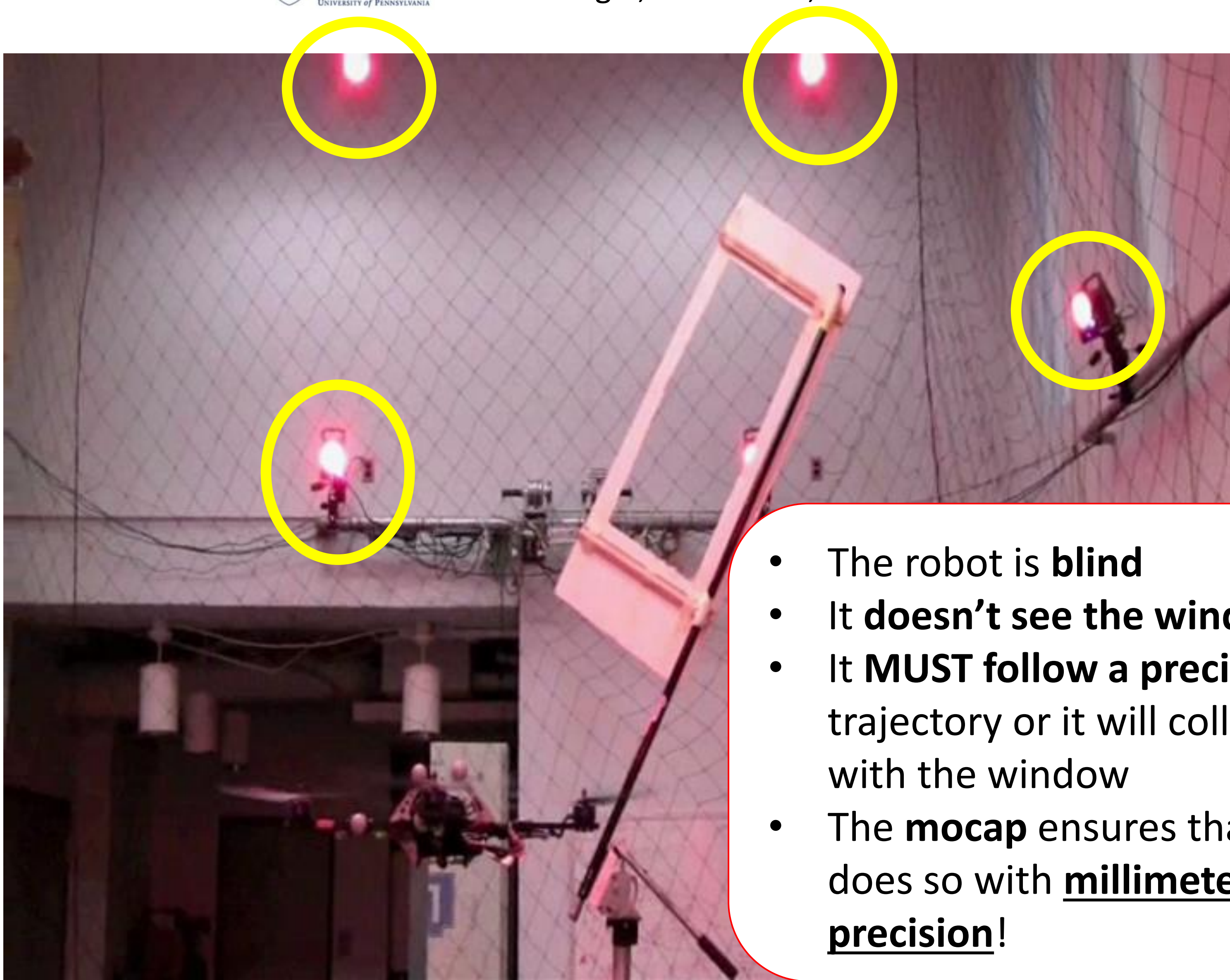


Active Perception

Flight through Narrow Windows

Autonomous Flight through Narrow Windows

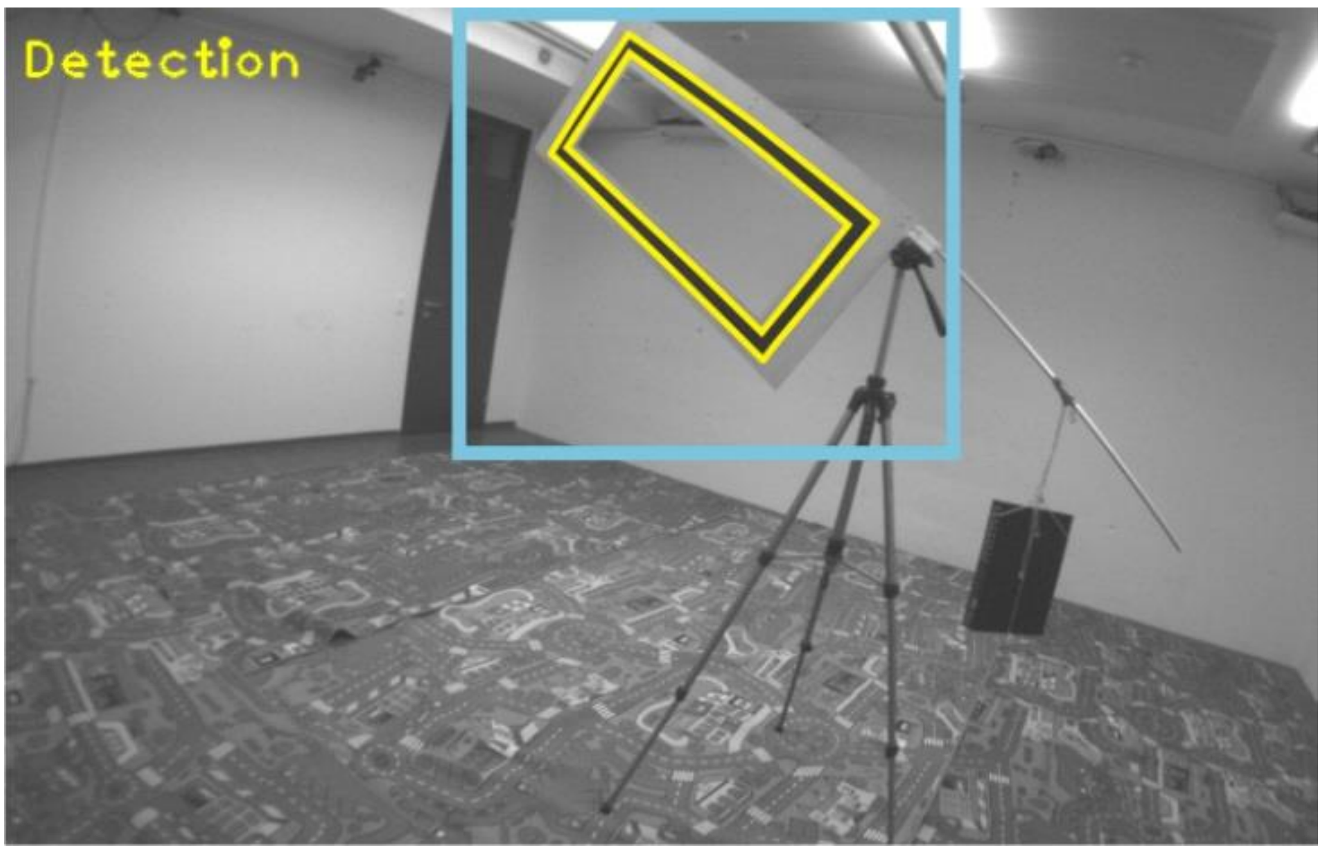
Penn University of Pennsylvania D. Mellinger, N. Michael, **V. Kmar**



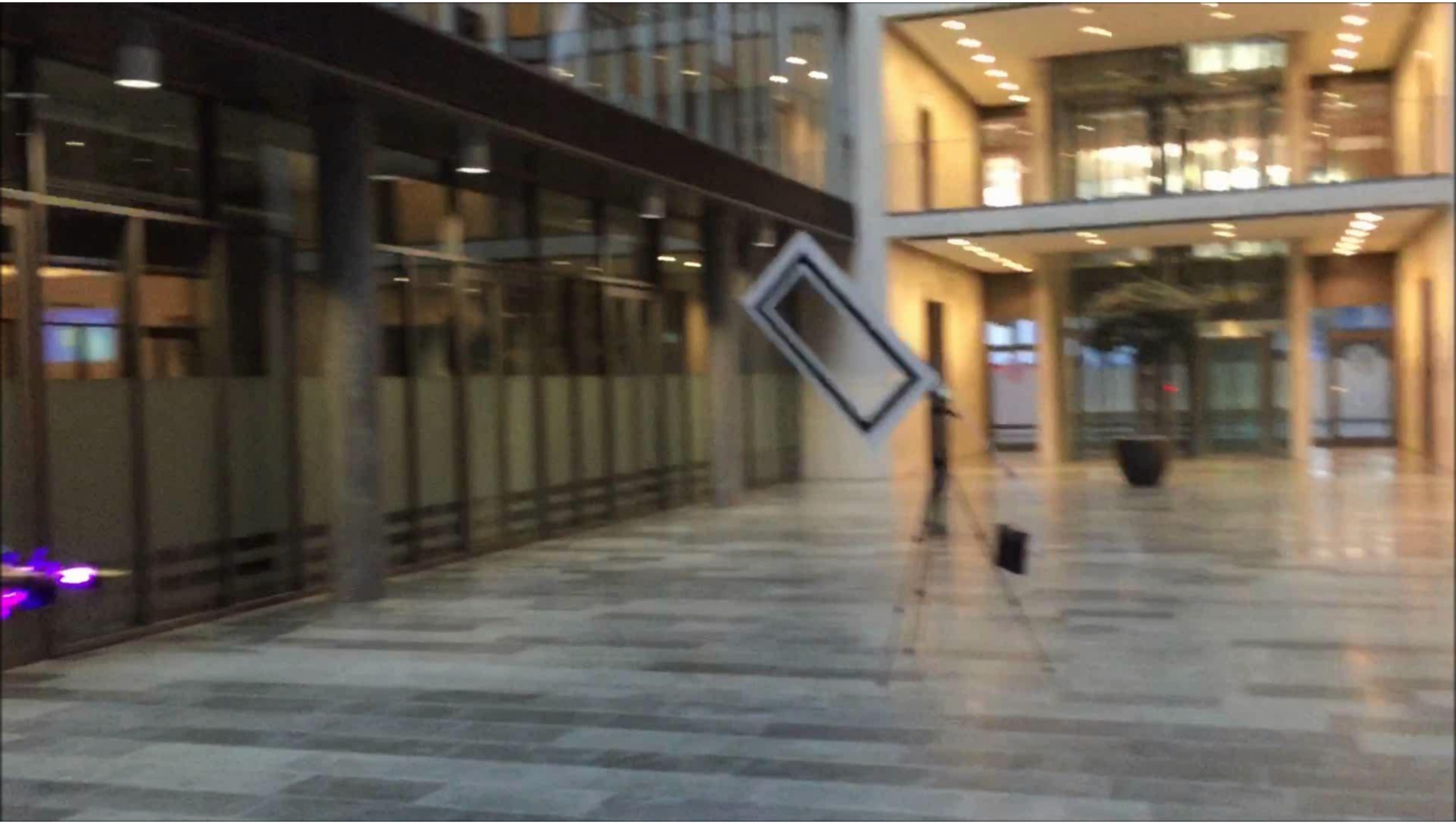
- The robot is **blind**
- It **doesn't see the window**
- It **MUST follow a precise** trajectory or it will collide with the window
- The **mocap** ensures that it does so with **millimeter precision**!

Vision-based Flight through Narrow Windows

Can we pass through narrow windows using only a single onboard camera?

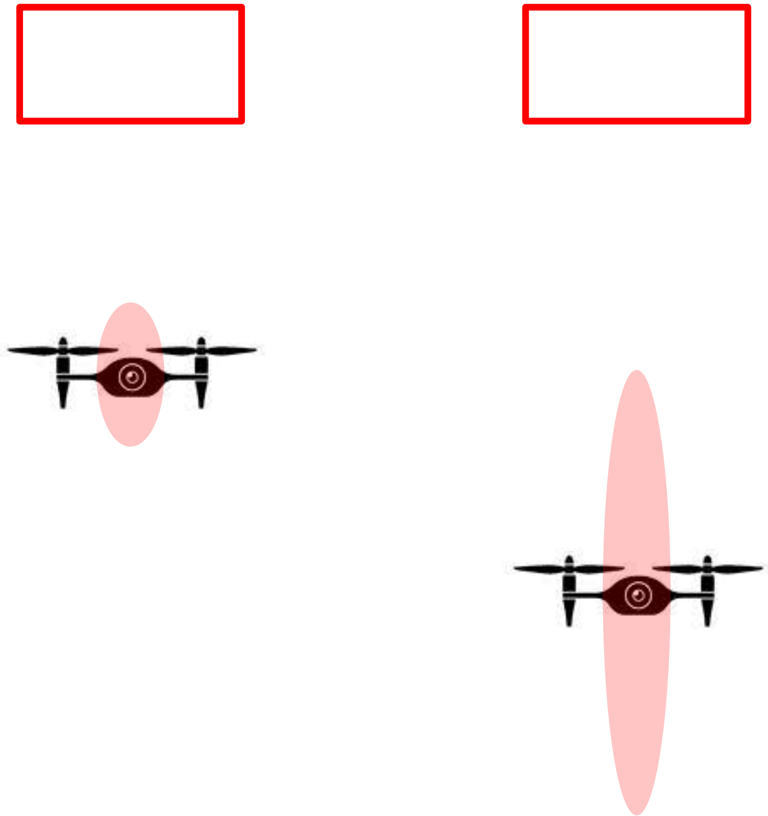


How would a professional drone pilot perform?

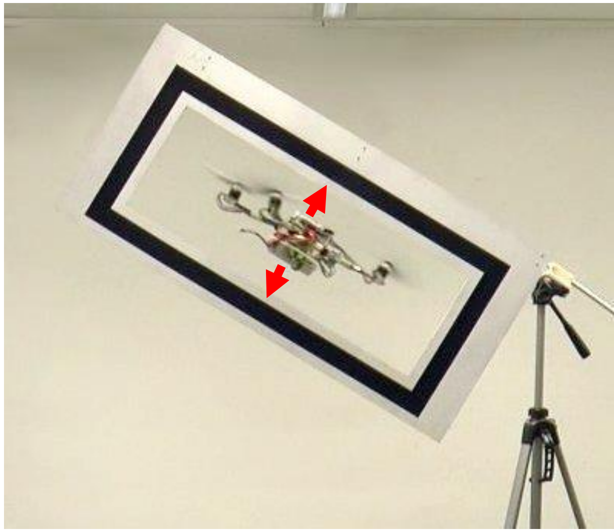


Challenge 1

- How to localize?
  - Use a **window detector** plus **IMU**
  - Pose uncertainty increases **quadratically with the distance**: $\boldsymbol{\sigma} \propto \boldsymbol{d}^2$



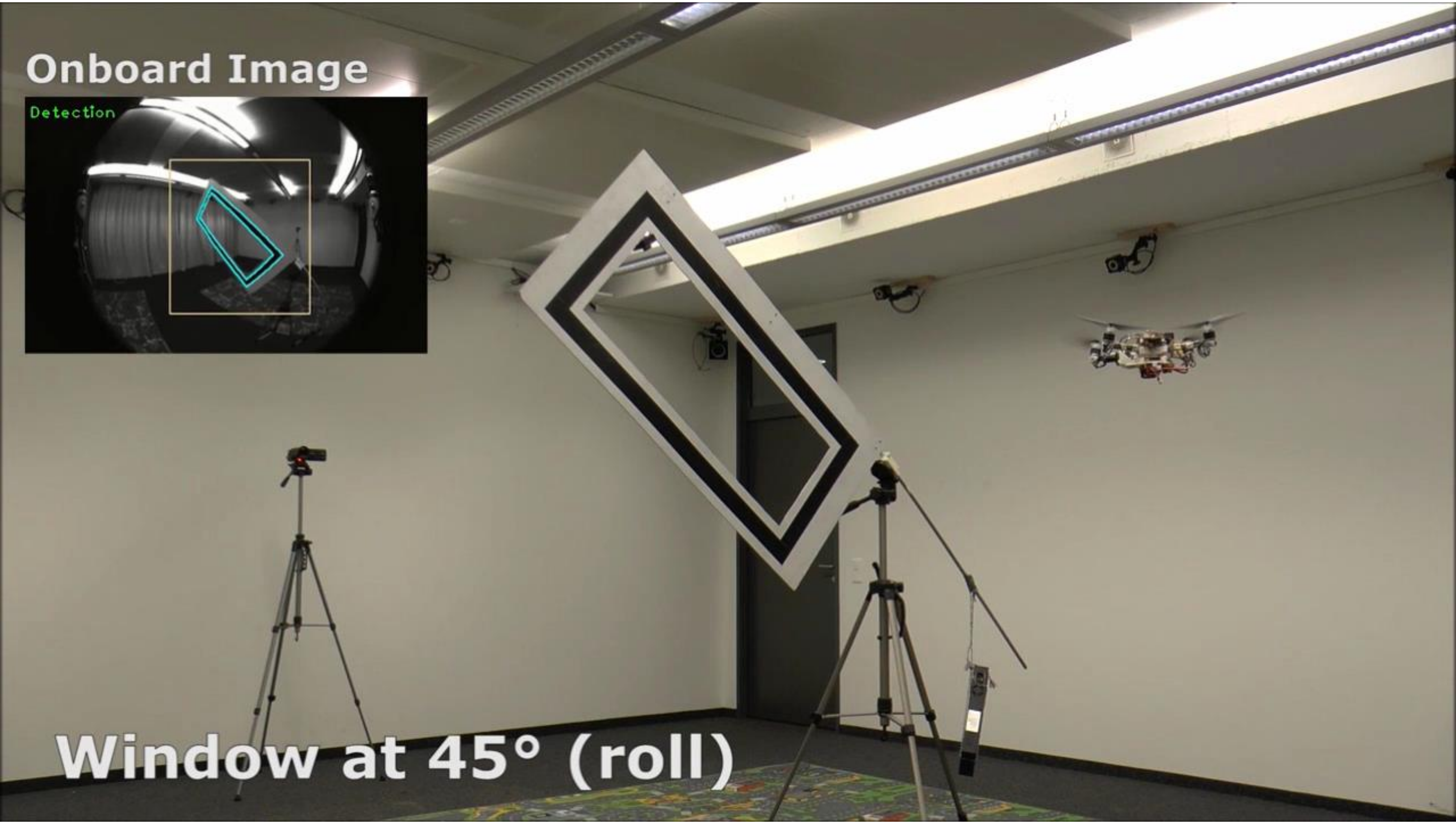
Challenge 2



➢ What is the optimal trajectory in (x, v, a)?

- **Dynamic constraints**
  - Continuous up to the Jerk
  - Minimize the snap (energy)
- **Geometric constraints**
  - Maximize the distance from the edges of the window
- **Perception constraints**
  - Always face the window to enable localization (active vision!)
  - Challenge: how can we **plan in the position and angle domains**?

Autonomous Flight through Narrow Gaps



[Falanga, Mueggler, Faessler, Scaramuzza, «Aggressive Quadrotor Flight through Narrow Gaps with onboard Sensing and Computing», Submitted, ICRA'16] **Featured on IEEE Spectrum**

Low-latency, Event-based Vision

Open Problems and Challenges with Micro Helicopters

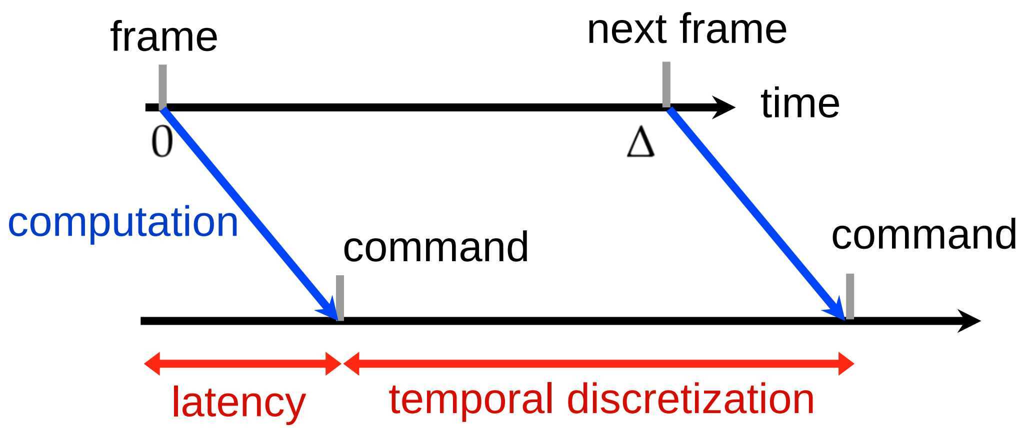
Current flight maneuvers achieved with onboard cameras are still to slow compared with those attainable by birds or FPV pilots

FPV-Drone race

To go faster, we need faster sensors!

- At the current state, the agility of a robot is limited by the latency and temporal discretization of its sensing pipeline.

- Currently, the average robot-vision algorithms have latencies of 50-200 ms. This puts a hard bound on the agility of the platform.

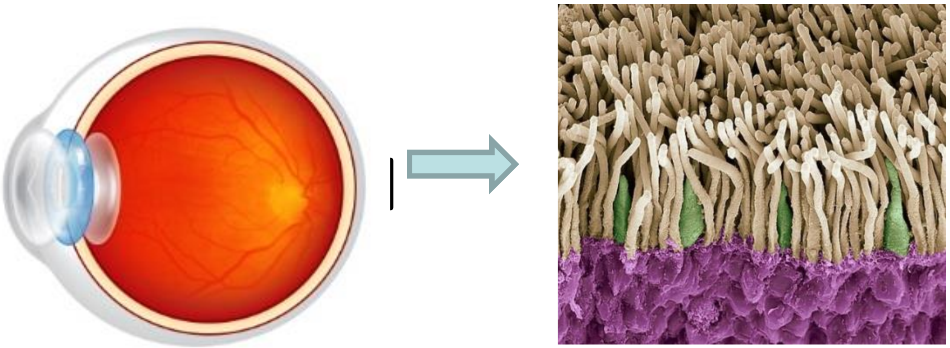


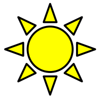
- **Can we create a low-latency and low-discretization perception pipeline?**
  - Yes, if we use **event-based cameras**

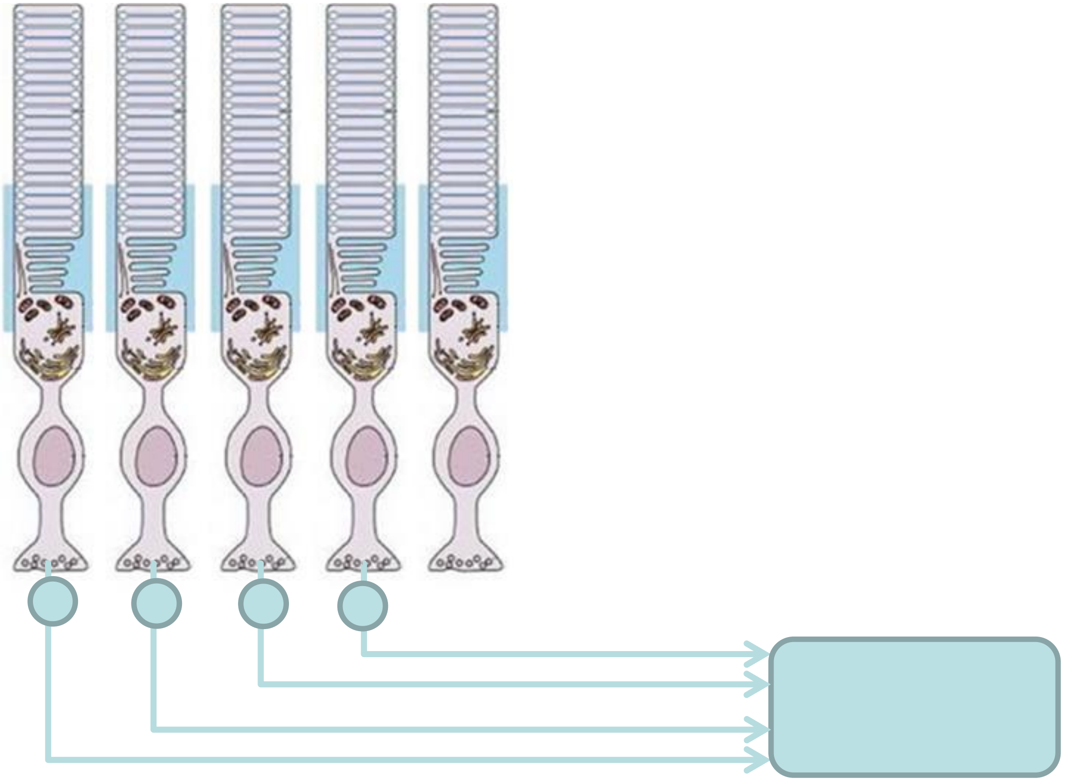
[Censi & Scaramuzza, «Low Latency, Event-based Visual Odometry», ICRA'14]

Human Vision System

- 130 million **photoreceptors**
- But only 2 million **axons**!

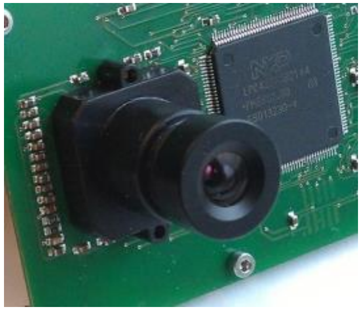






Dynamic Vision Sensor (DVS)

- **Event-based camera** developed by Tobi Delbruck's group (ETH & UZH).
- Temporal resolution: **1 μs**
- High dynamic range: **140 dB**
- Low power: **20 mW**
- Cost: 2,500 EUR



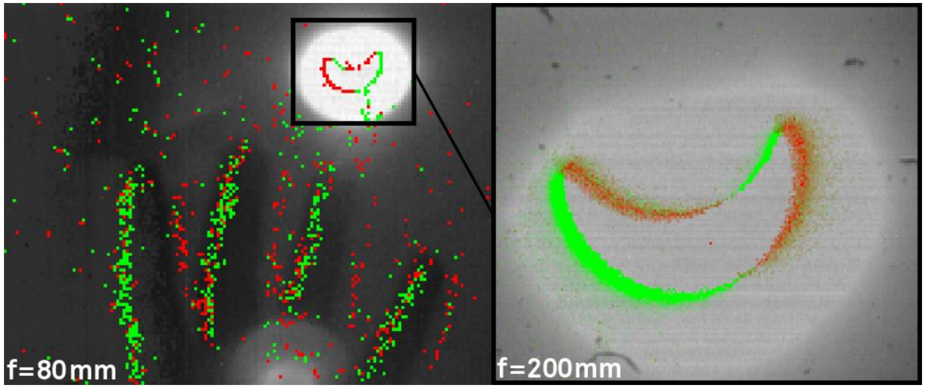
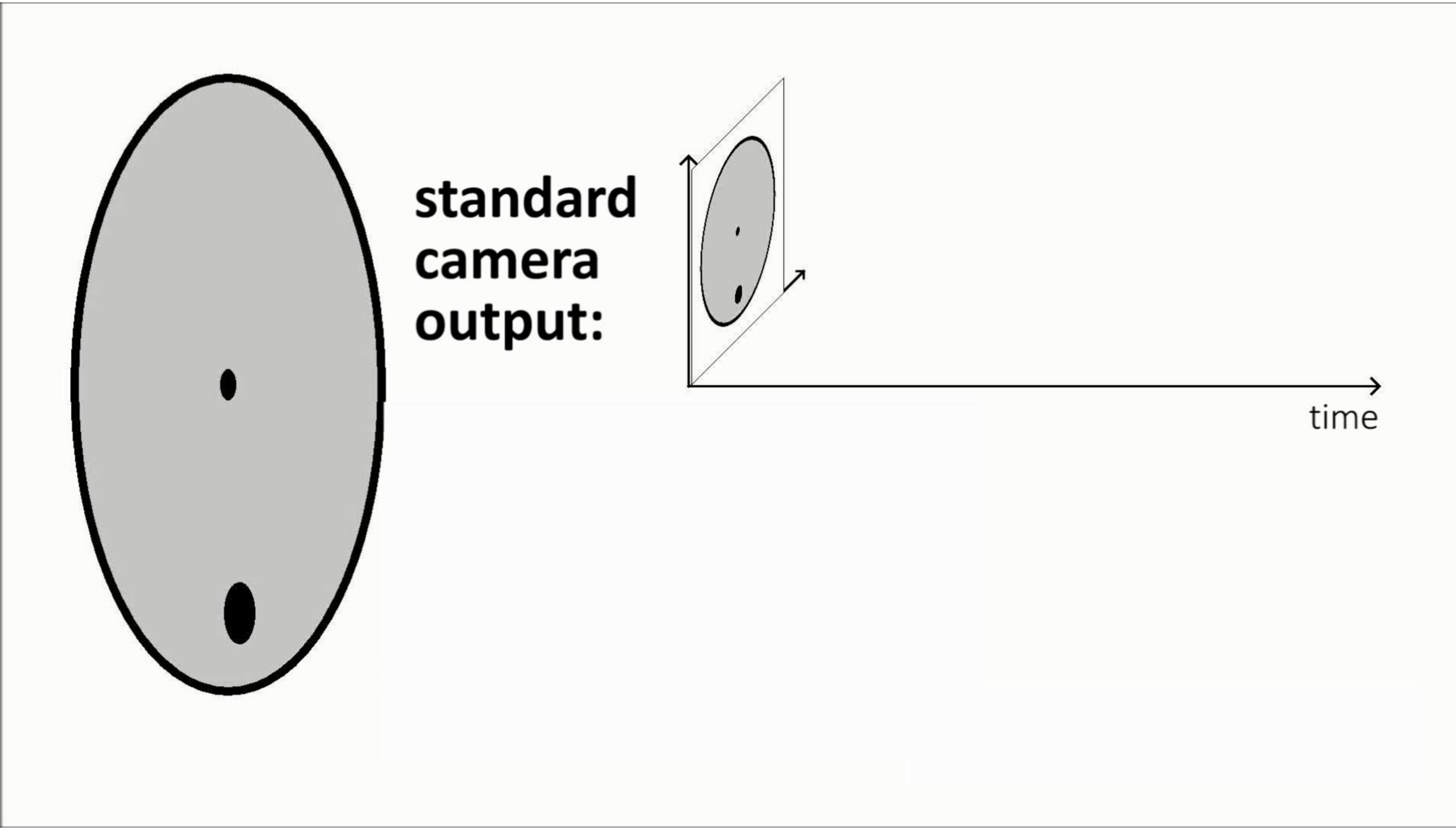


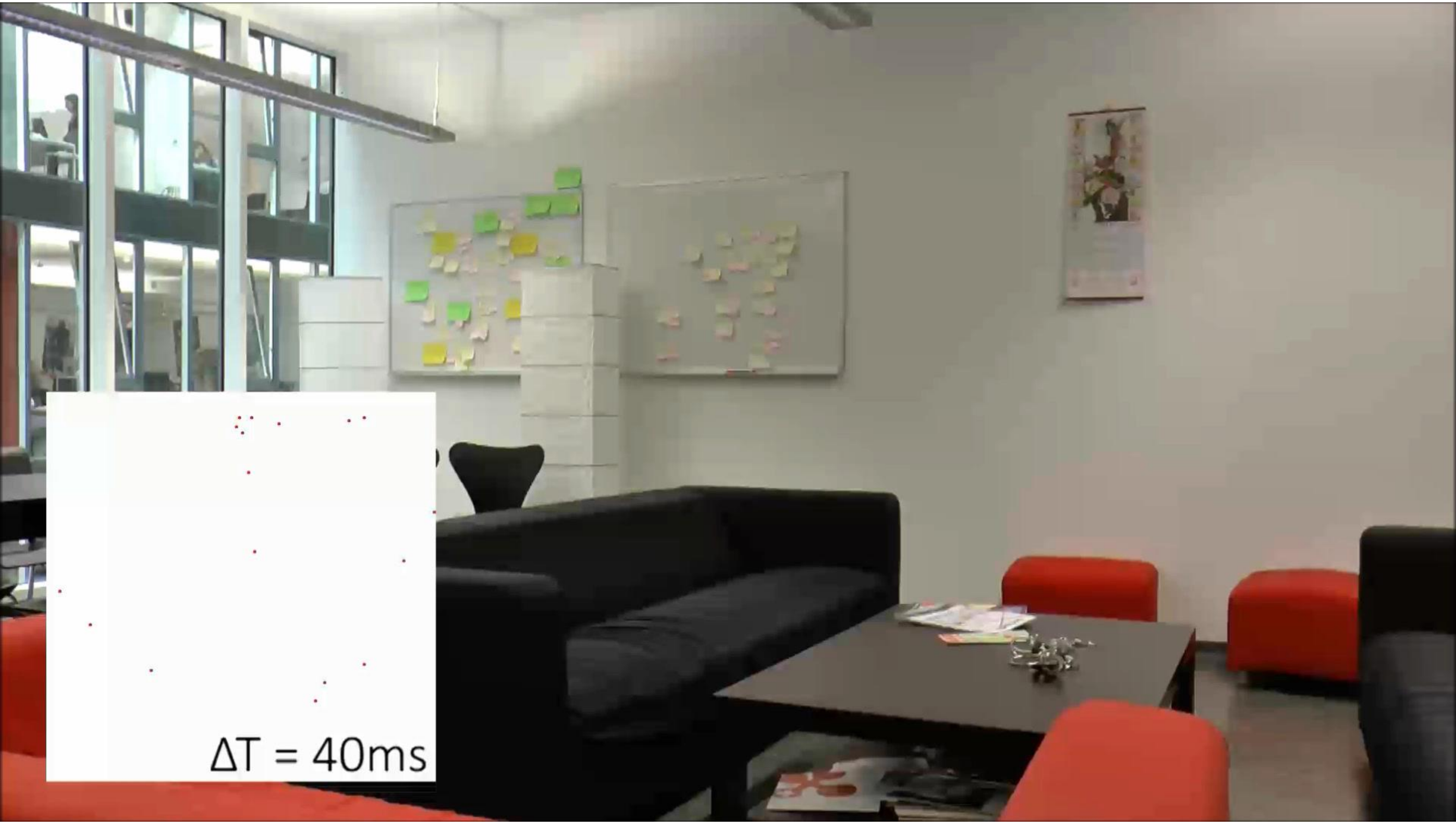
Image of the solar eclipse (March'15) captured by a DVS (courtesy of IniLabs)

[Lichtsteiner, Posch, Delbruck. A 128x128 120 dB 15μs Latency Asynchronous Temporal Contrast Vision Sensor. 2008]

Camera vs Dynamic Vision Sensor



Camera vs Dynamic Vision Sensor



Dynamic Vision Sensor (DVS)

Advantages

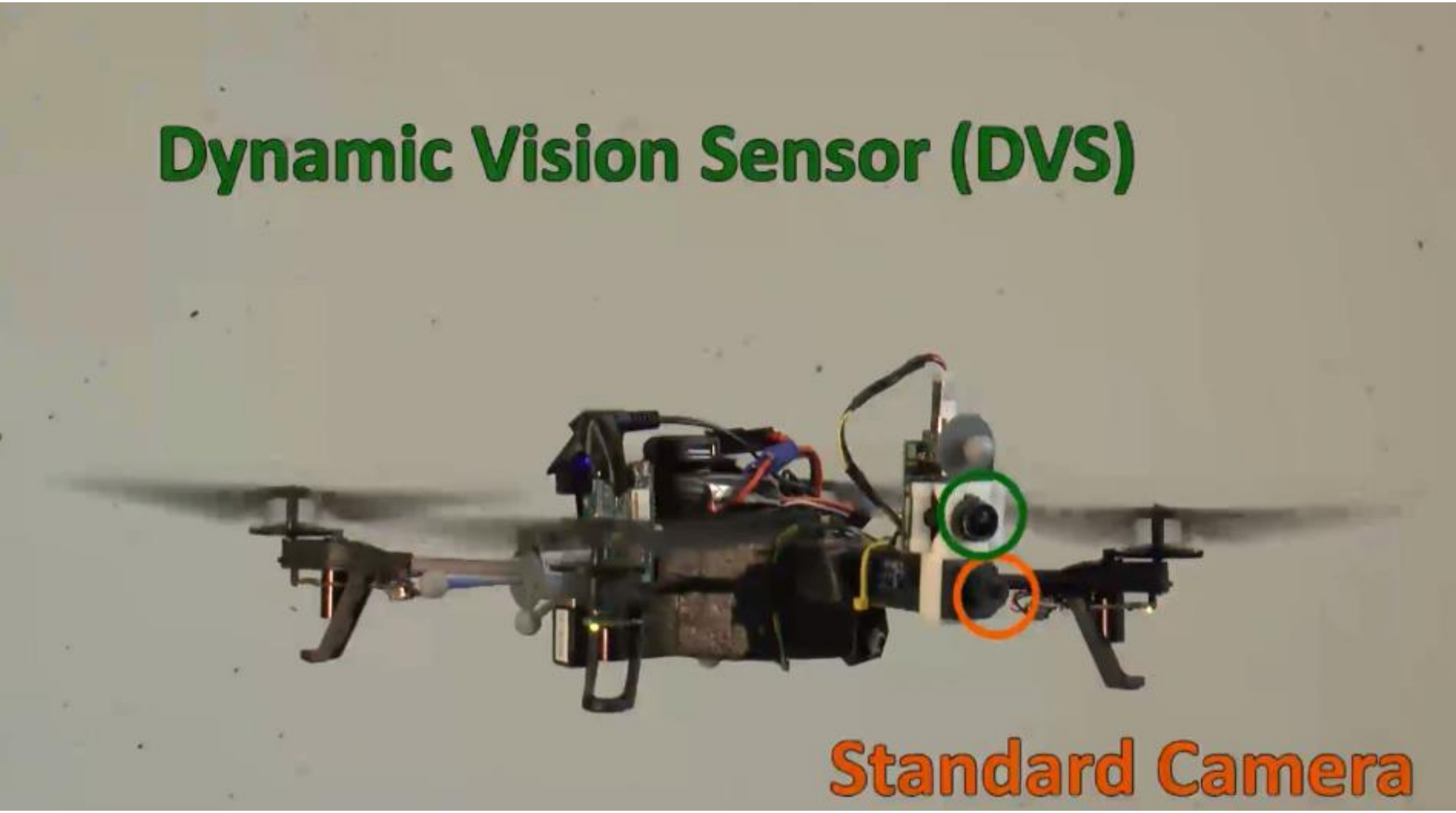
- **low-latency** (~1 micro-second)
- **high-dynamic range** (140 dB instead 60 dB)
- **Low storage capacity, processing time, and power**

Disadvantages

- Require totally **new vision algorithms**
- **No intensity information** (only binary intensity changes)

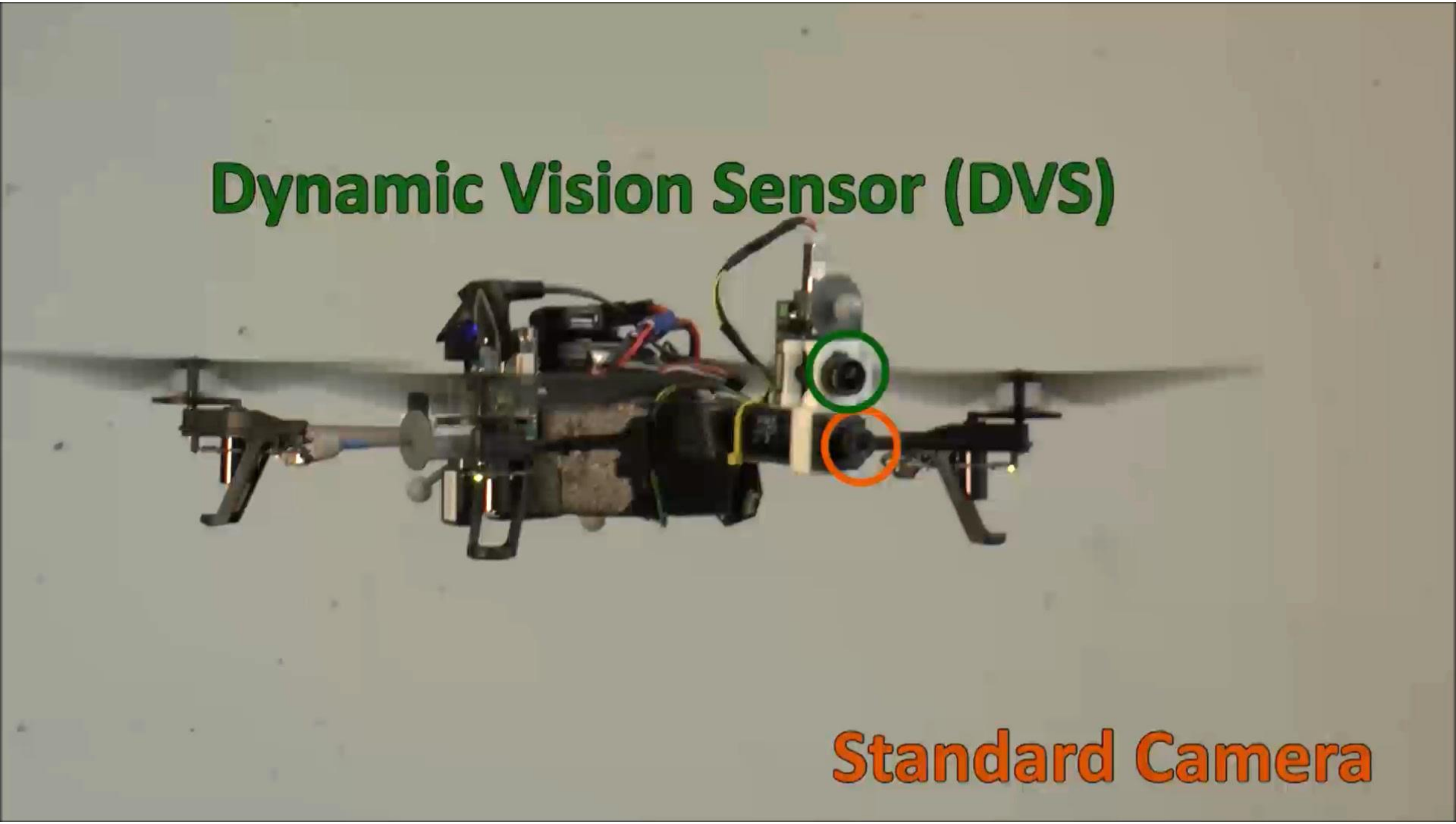
Lichtsteiner, Posch, Delbruck. *A 128x128 120 dB 15μs Latency Asynchronous Temporal Contrast Vision Sensor.* 2008

DVS mounted on a quadrotor AR Drone



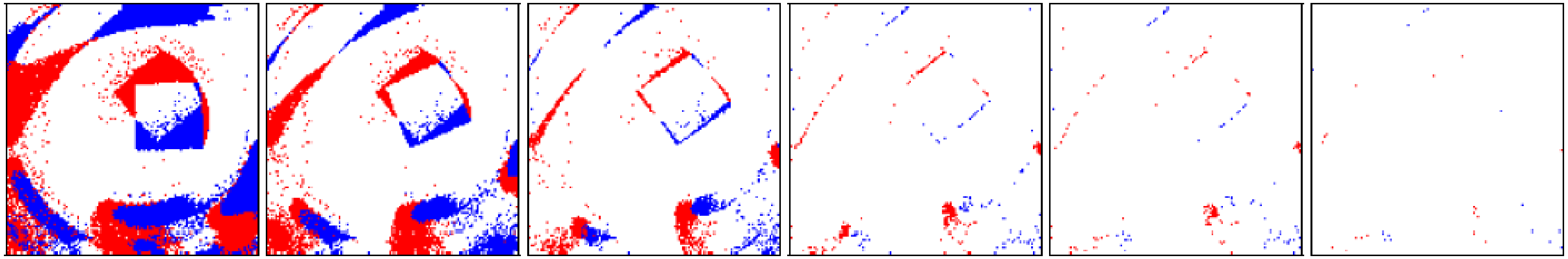
[Mueggler, Huber, Scaramuzza, Event-based, 6-DOF Pose Tracking for High-Speed Maneuvers, IROS’14]

Application Experiment: Quadrotor Flip (1,200 deg/s)



[Mueggler, Huber, Scaramuzza, Event-based, 6-DOF Pose Tracking for High-Speed Maneuvers, IROS'14]

Camera and DVS renderings



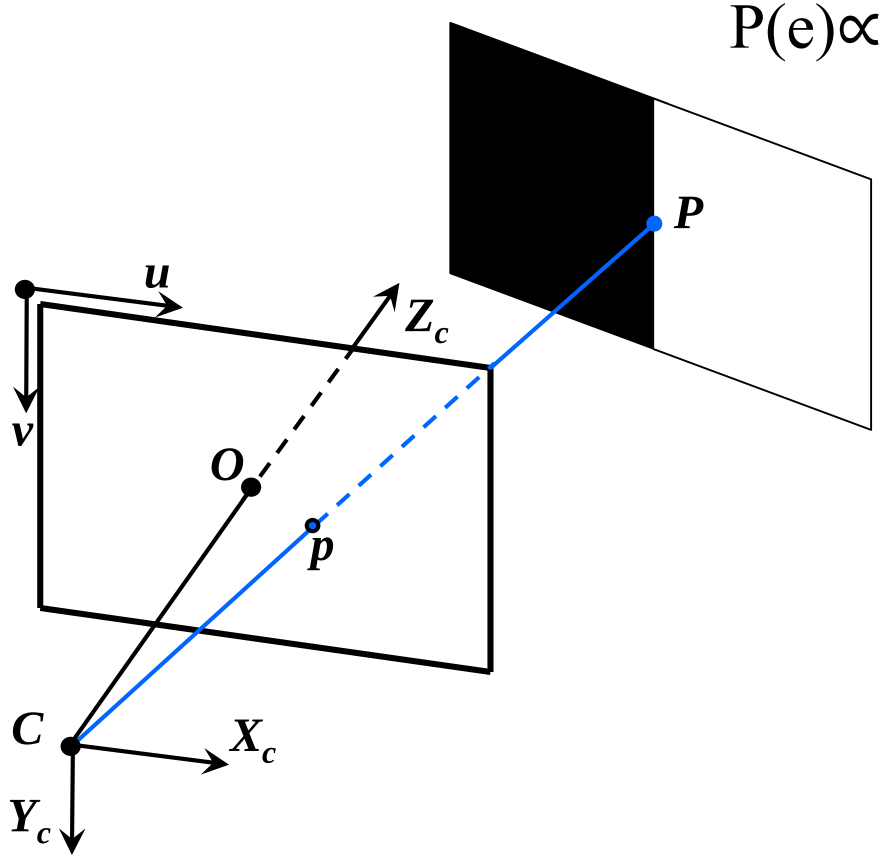
(a) $\Delta T = 33\,\mathrm{ms}$ (b) $\Delta T = 15\,\mathrm{ms}$ (c) $\Delta T = 5\,\mathrm{ms}$ (d) $\Delta T = 1\,\mathrm{ms}$ (e) $\Delta T = 0.5\,\mathrm{ms}$ (f) $\Delta T = 0.1\,\mathrm{ms}$

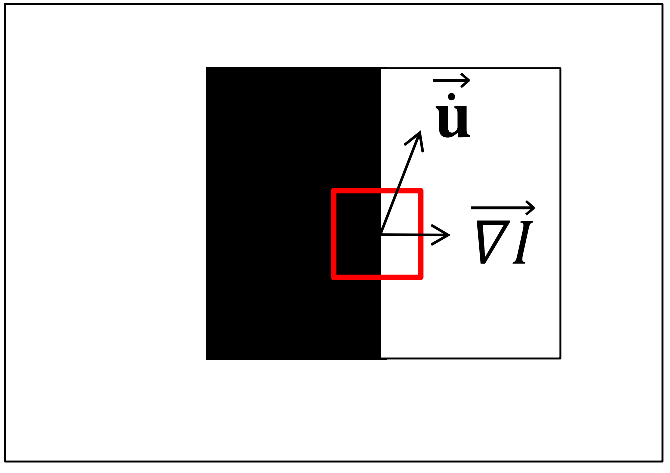
[Mueggler, Huber, Scaramuzza, Event-based, 6-DOF Pose Tracking for High-Speed Maneuvers, IROS'14]

Generative Model [Censi & Scaramuzza, ICRA'14]

The generative model tells us that the **probability** that an event is generated depends on the **scalar product** between the gradient ∇I and the apparent motion $\dot{\mathbf{u}}\Delta t$

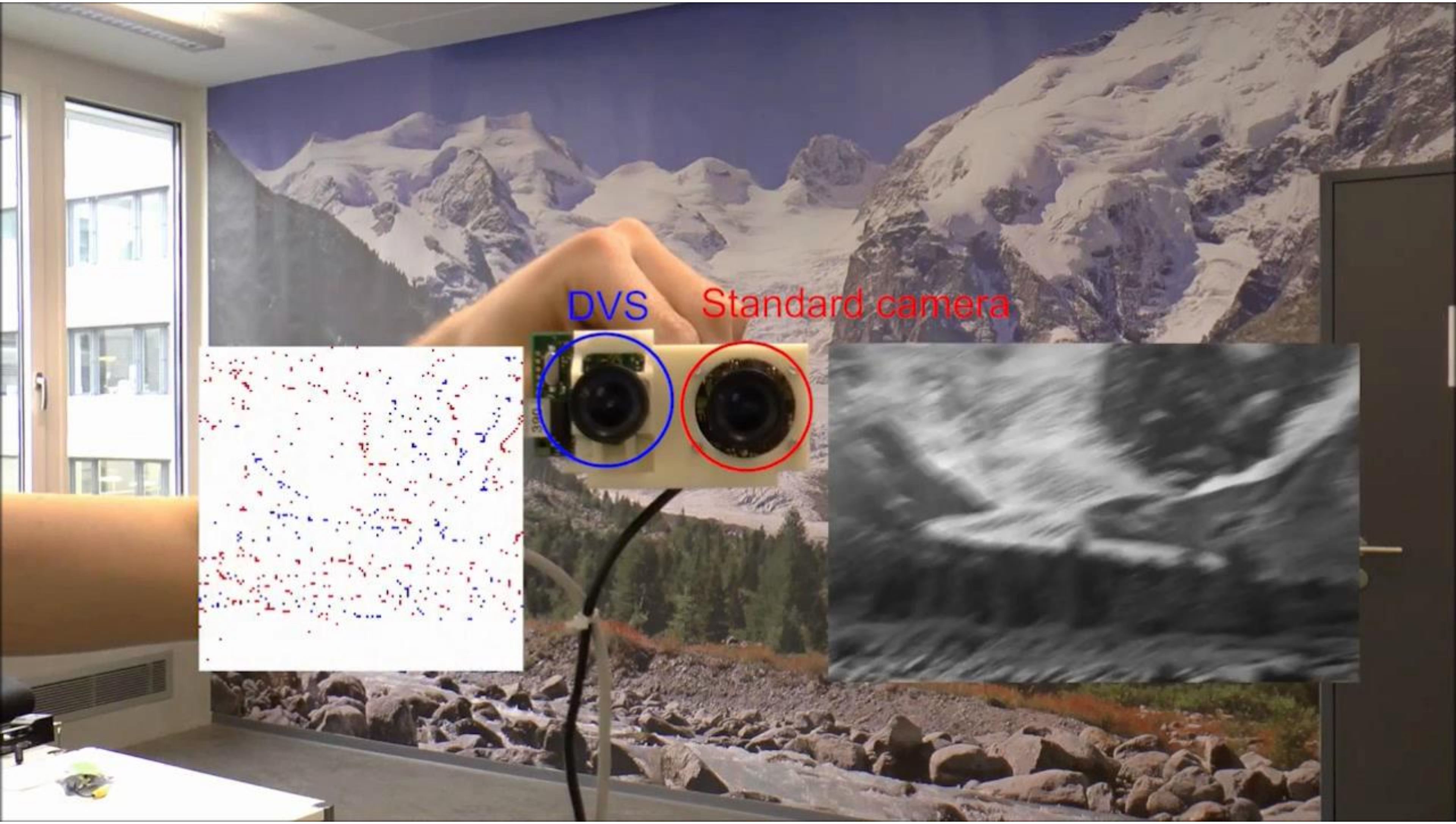
$$P(e) \propto |\langle \nabla I, \dot{\mathbf{u}}\Delta t \rangle|$$





[Censi & Scaramuzza, *Low Latency, Event-based Visual Odometry*, ICRA'14]

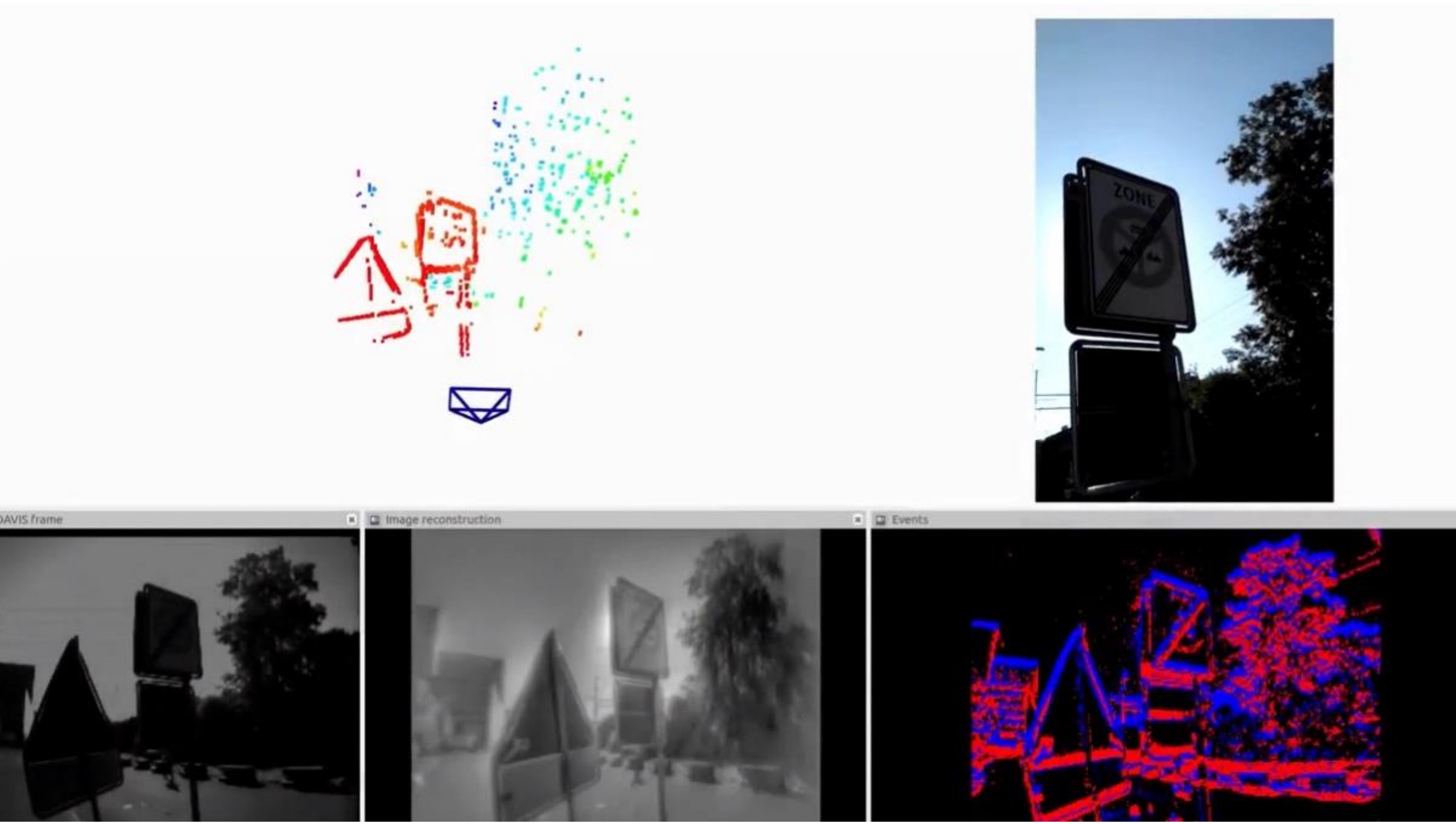
Event-based 6DoF Pose Estimation Results



[Event-based Camera Pose Tracking using a Generative Event Model, Arxiv]

[Censi & Scaramuzza, Low Latency, Event-based Visual Odometry, ICRA'14]

Robustness to Illumination Changes and High-speed Motion



BMVC’16 : EMVS: Event-based Multi-View Stereo, **Best Industry Paper Award**

Conclusions

- Agile flight (**like birds**) is still far (10 years?)
- Agile flight requires success at different levels
  - **perception, planning, and control**
- **Perception and control** need to be considered **jointly!**
- **Event cameras** open enormous possibilities!
  - Ideal for high speed motion estimation and robustness to illumination changes