ROBOTICS & PERCEPTION GROUP

University of Zurich[UZH]

Vision-controlled Flying Robots
From Frame-based to Event-based Vision

Davide Scaramuzza

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

Research Background

[ICVS'06, IROS'06, PAMI'13]

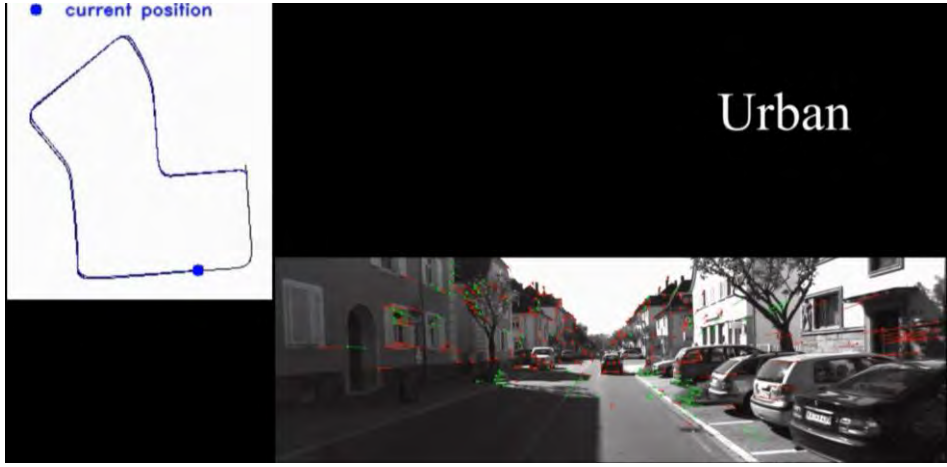
Computer Vision

- Visual Odometry and SLAM
- Sensor fusion
- Camera calibration

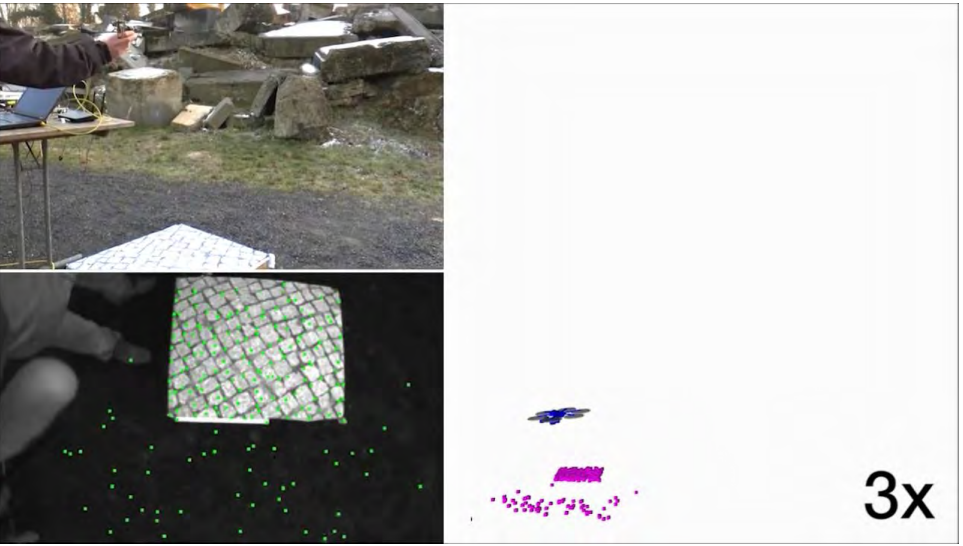
Autonomous Robot Navigation

- Self driving cars
- Micro Flying Robots

[ICCV'09, CVPR'10, JFR'11, IJCV'11]



[JFR'10, AURO'11, RAM'14, JFR'15]



My Research Group

ROBOTICS &
PERCEPTION
GROUP

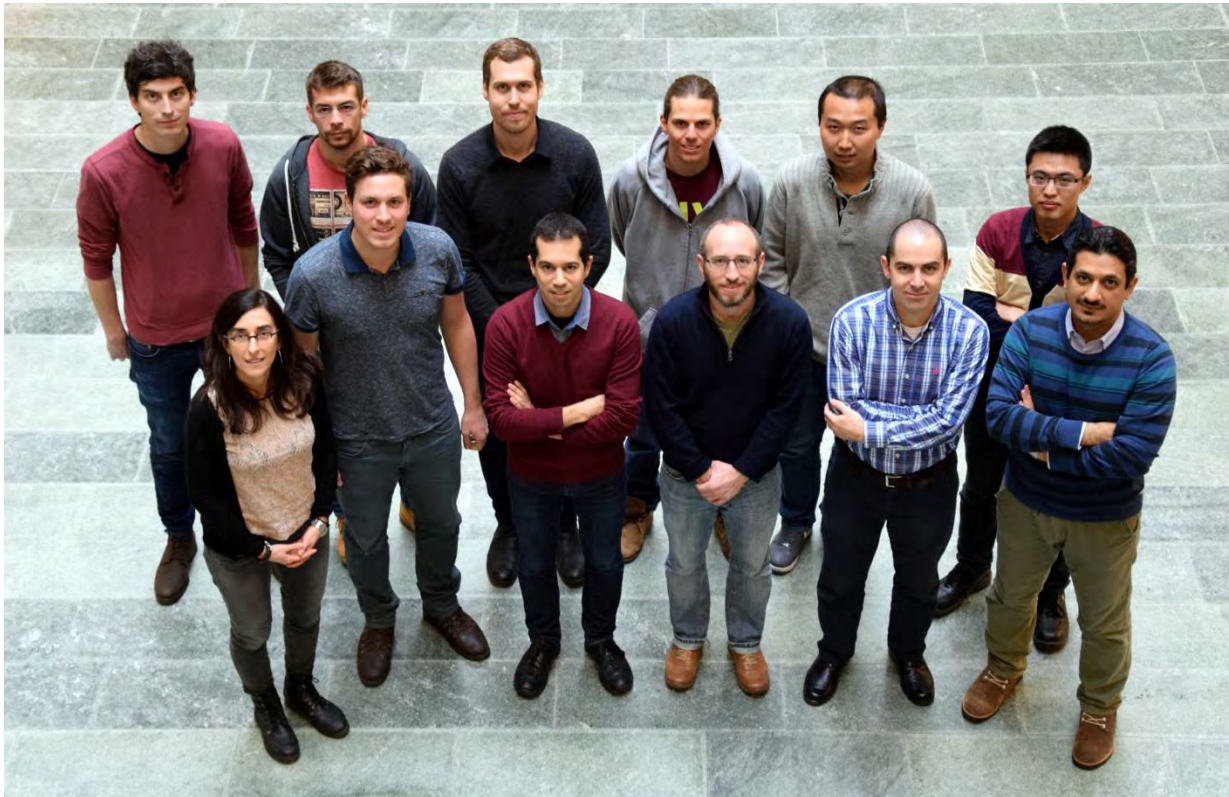
http://rpg.ifi.uzh.ch

erc

Google
Google
Faculty Research Awards

FNSNF
SWISS NATIONAL SCIENCE FOUNDATION

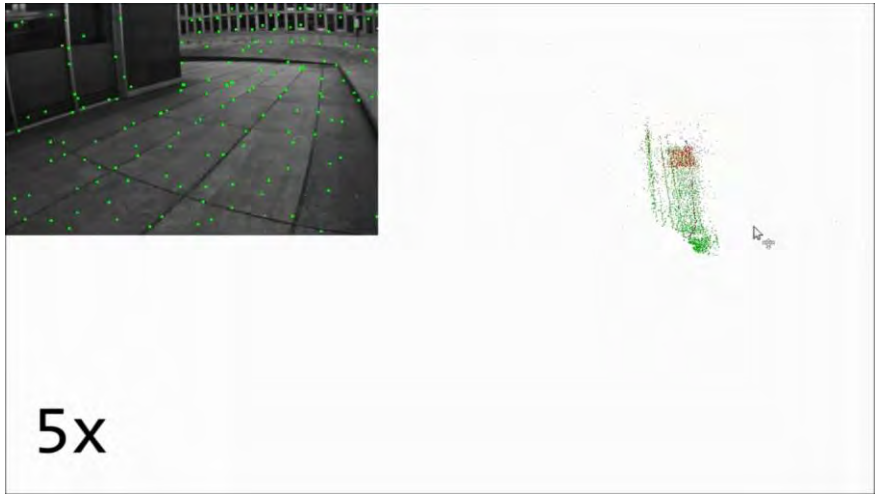
robotics
Swiss National
Centre of Competence
in Research



Current Research

Visual-Inertial State Estimation

[T-RO'08, IJCV'11, PAMI'13, RSS'15]

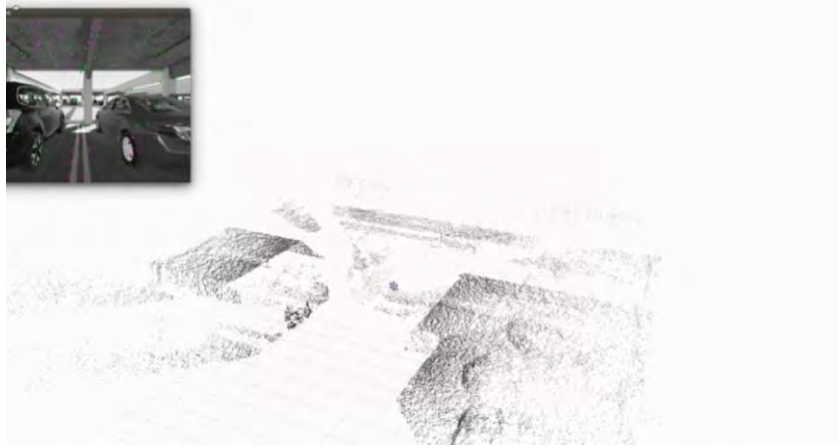


Autonomous Navigation of Flying Robots

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

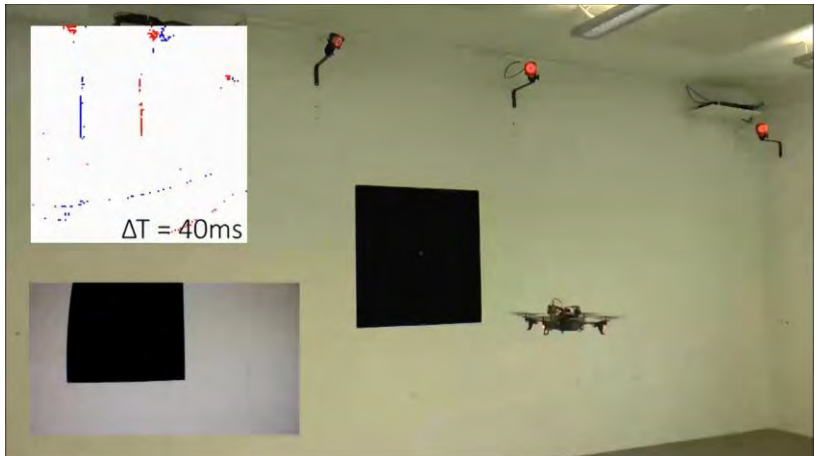
Probabilistic, Dense Reconstruction

[ICRA'14, JFR'15]



Event-based Vision for Agile Flight

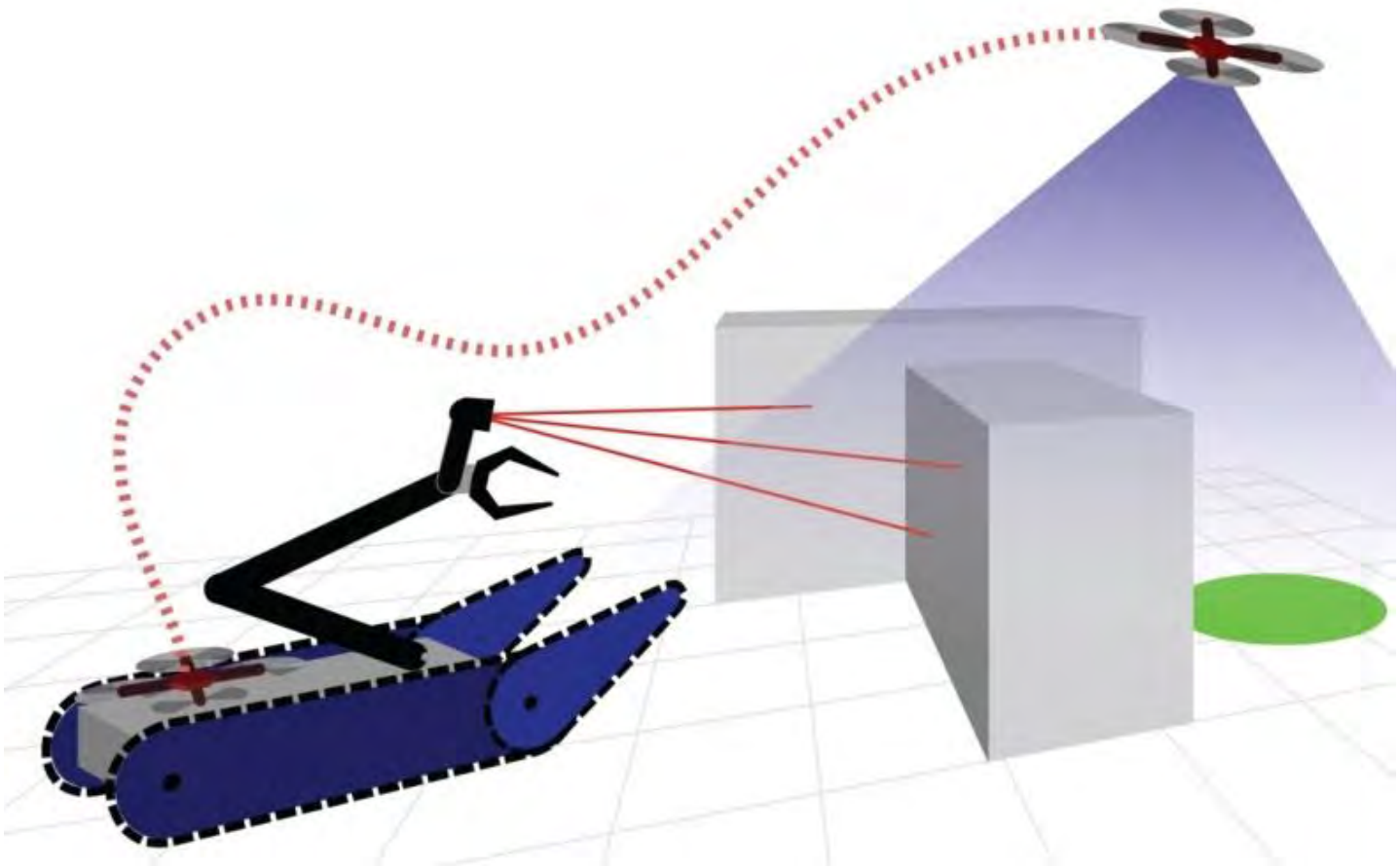
[IROS'3, ICRA'14, RSS'15]



Other research topics (not shown in this presentation)

Aerial-guided navigation of a Ground Robot among Movable Obstacles

[IROS'13, SSRR'14, JFR'15]



Other research topics (not shown in this presentation)

Aerial-guided navigation of a Ground Robot among Movable Obstacles

[IROS'13, SSRR'14, JFR'15]

Other research topics (not shown in this presentation)

Autonomous trail following in the forests using Deep Learning

[Submitted to IEEE RA-L]

Today’s Applications

Today's Applications of MAVs

Transportation

Search and rescue

Aerial photography

Inspection

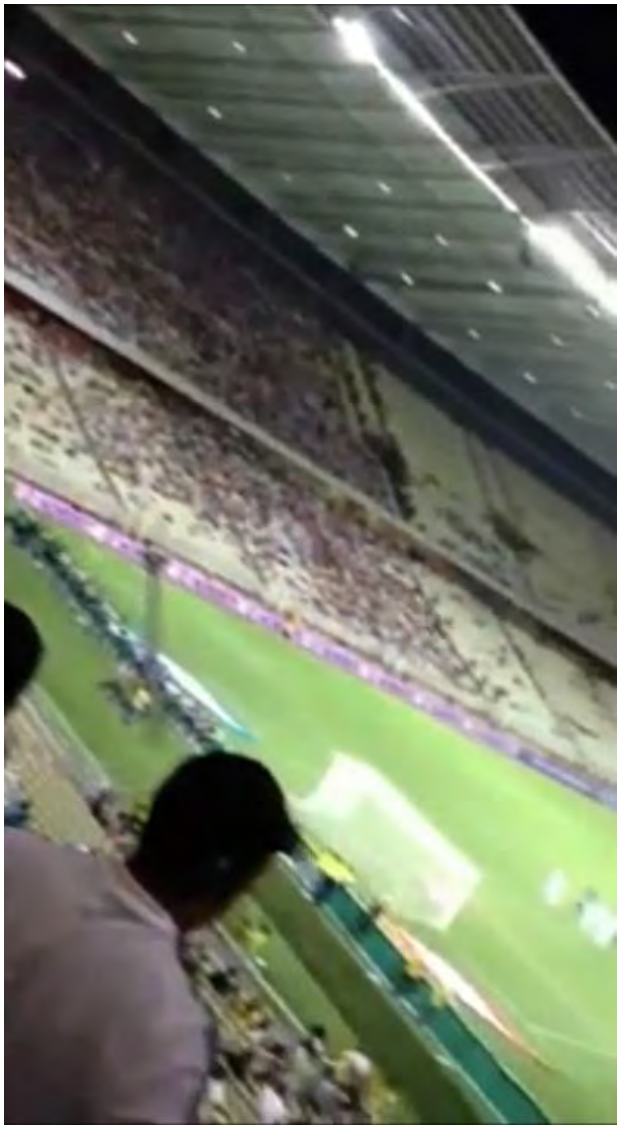
Law enforcement

Agriculture

How to fly a drone

- **Remote control**
  - Requires line of sight or communication link
  - Requires skilled pilots
- **GPS-based navigation**
  - Doesn’t work indoors
  - Can be unreliable outdoors

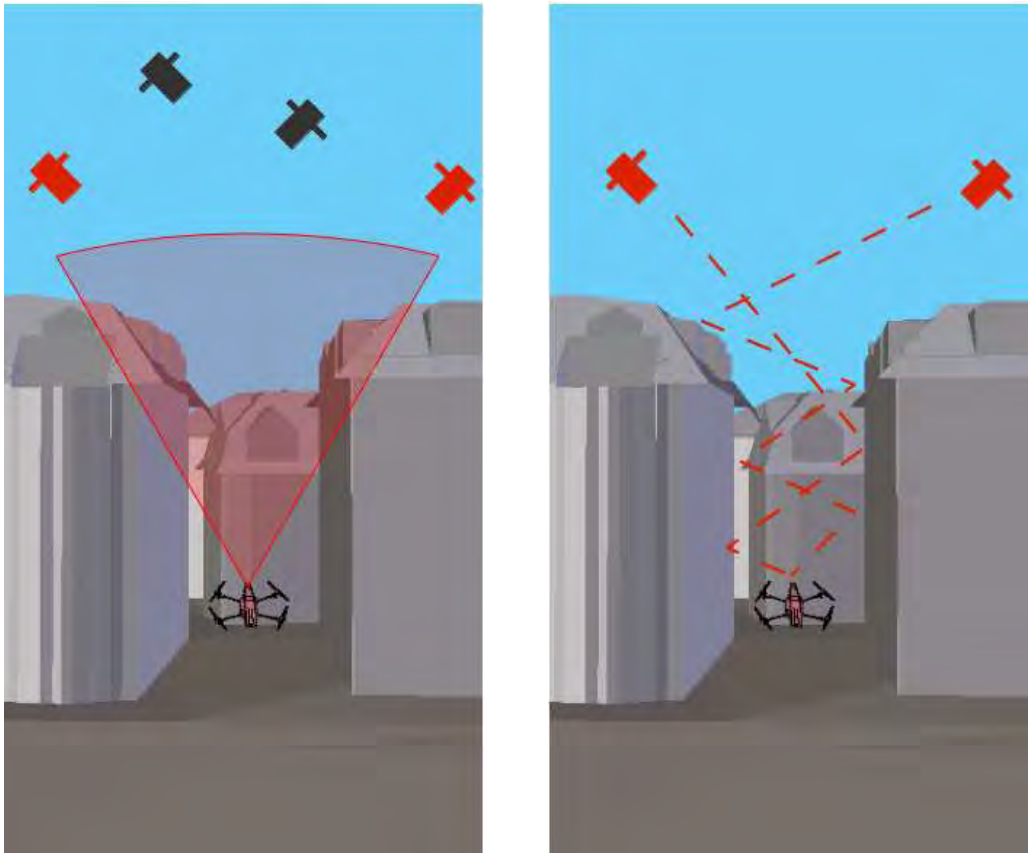
Interior of an earthquake-damaged building in Japan



Drone crash during soccer match, Brasilia, 2013

Problems of GPS

- Does not work indoors
- Even outdoors it is not a reliable service
  - Satellite coverage
  - Multipath problem



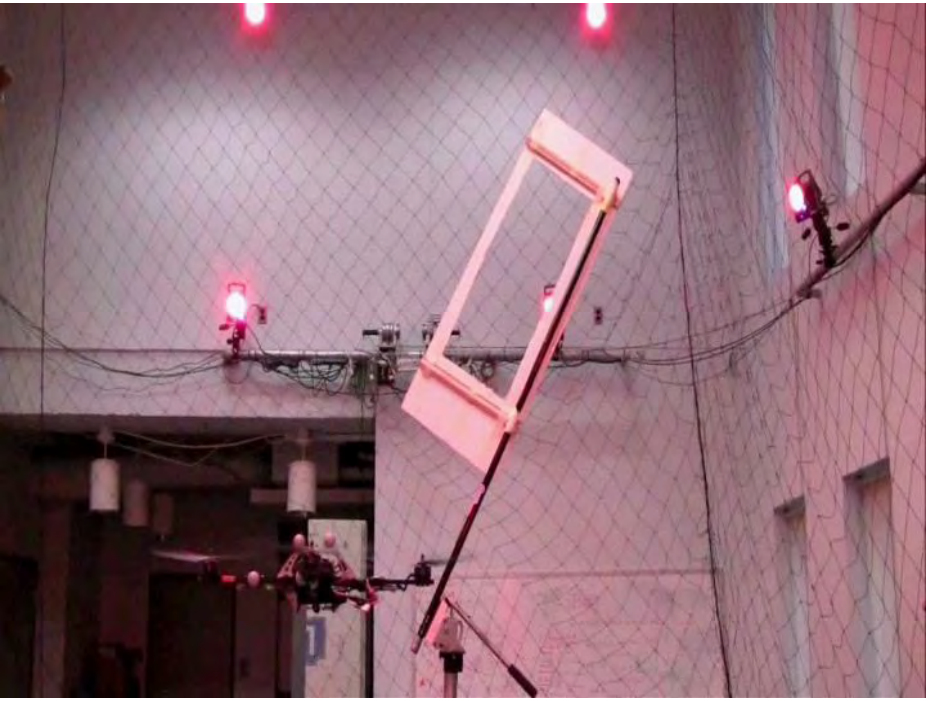
Why do we need autonomy?

Autonomous Navigation is crucial for:

Search and Rescue

Remote Inspection

How do we Localize without GPS ?



Mellinger, Michael, **Kumar**

Penn University of Pennsylvania GRASP Laboratory

Fontana, Faessler, **Scaramuzza**

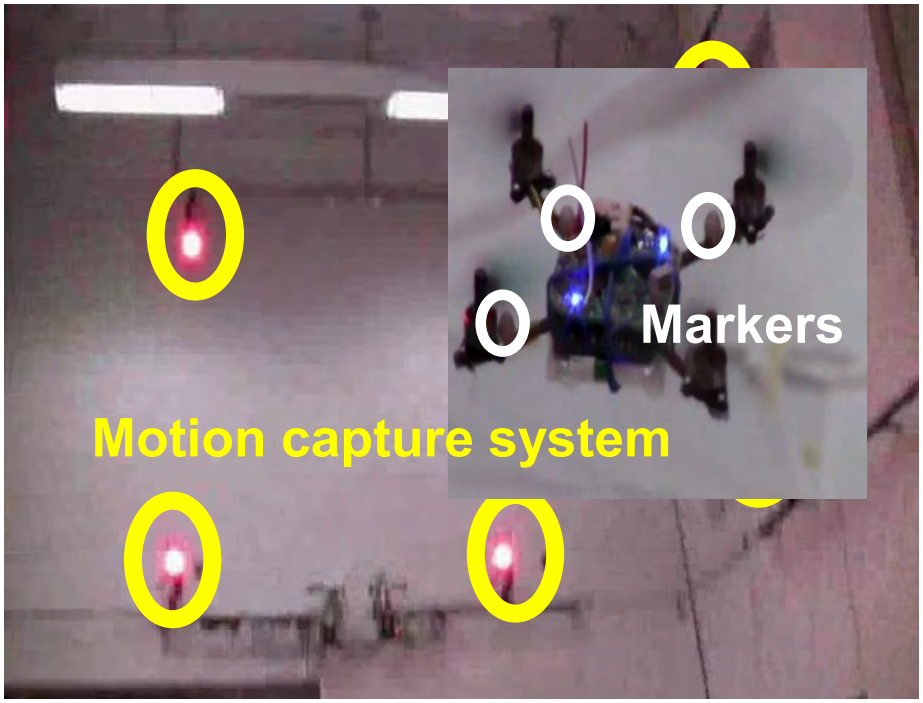
Universität Zürich UZH

How do we Localize without GPS ?

This robot is «***blind***»

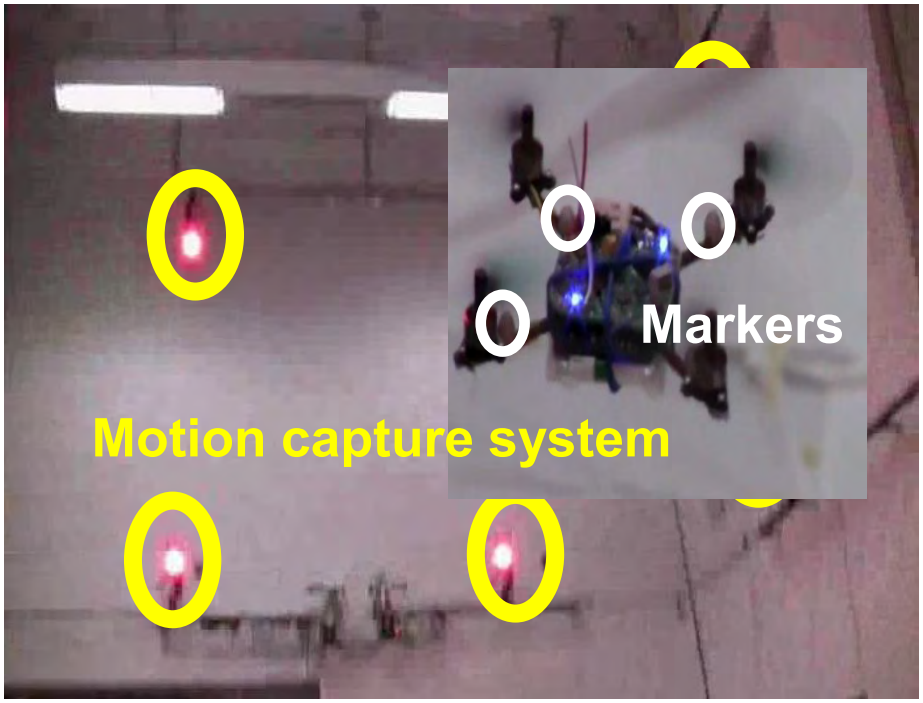
How do we Localize without GPS ?

This robot is «***blind***»



How do we Localize without GPS ?

This robot is «***blind***»



This robot can «***see***»

Autonomous Vision-based Navigation in GPS-denied Environments

[Scaramuzza, Achtelik, Weiss, Fraundorfer, et al., Vision-Controlled Micro Flying Robots: from System Design to Autonomous Navigation and Mapping in GPS-denied Environments, **IEEE RAM, 2014**]

Problems with Vision-controlled MAVs

Quadrotors have the **potential** to navigate quickly but…

- Autonomous operation is currently **restricted to controlled environments**
- Vision-based maneuvers still **slow** and **inaccurate** compared to VICON

Why?

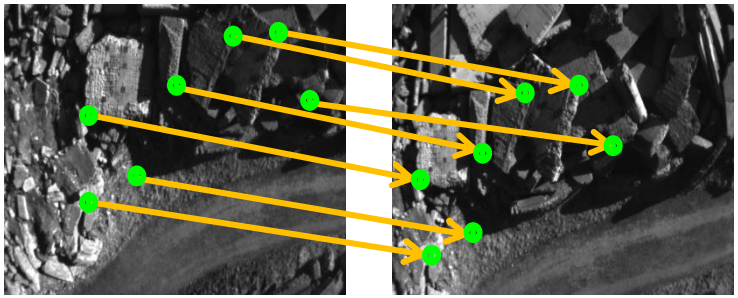
- Perception algorithms are mature but **not robust**
  - Unlike lasers and Vicon, **localization accuracy** depends on **depth & texture**!
  - Algorithms and sensors have **big latencies** (50-200 ms)
  - Sparse models instead of **dense environment models**
  - Control & perception have been mostly **considered separately**

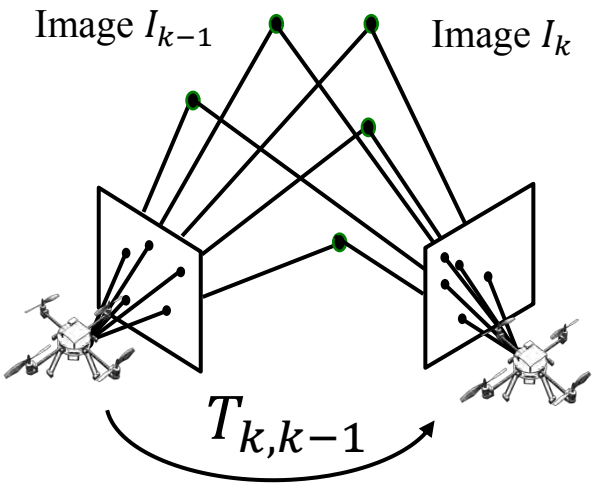
Outline

- Visual-inertial state estimation
- From sparse to dense models
- Active vision and control
- Event-based Vision for agile flight

Vision-based, GPS-denied Navigation

Visual Odometry

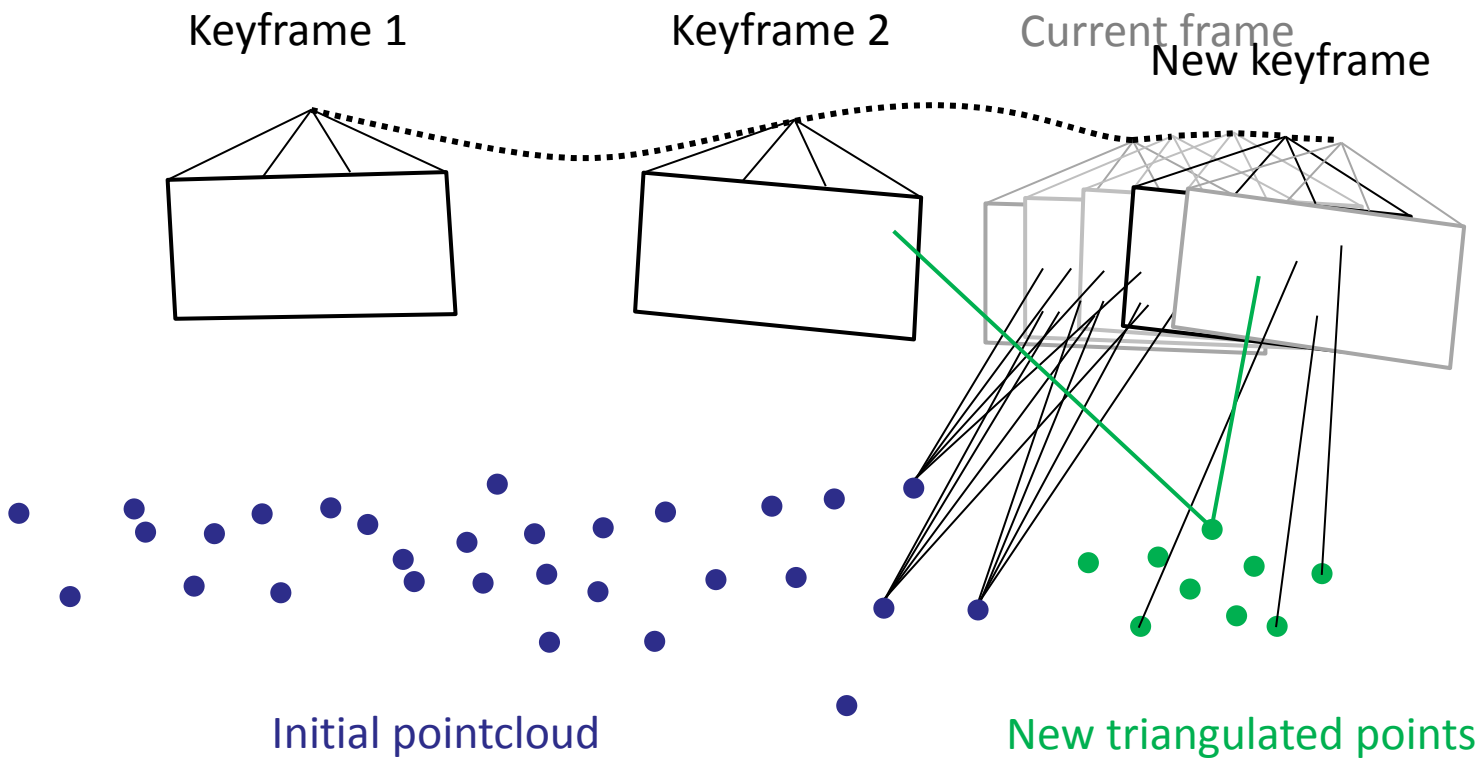




$$\mathbf{T}_{k,k-1} = \arg\min_{\mathbf{T}} \iint_{\bar{\mathcal{R}}} \rho\, I_k\left(\pi\left(\mathbf{T}\cdot\pi^{-1}(\mathbf{u}, d_{\mathbf{u}})\right)\right) - I_{k-1}(\mathbf{u})\, d\mathbf{u}$$

1. Scaramuzza, Fraundorfer. Visual Odometry, *IEEE Robotics and Automation Magazine*, 2011
2. D. Scaramuzza. 1-Point-RANSAC Visual Odometry, International Journal of Computer Vision, 2011

Keyframe-based Visual Odometry



PTAM (Parallel Tracking & Mapping) [Klein, ISMAR'08]

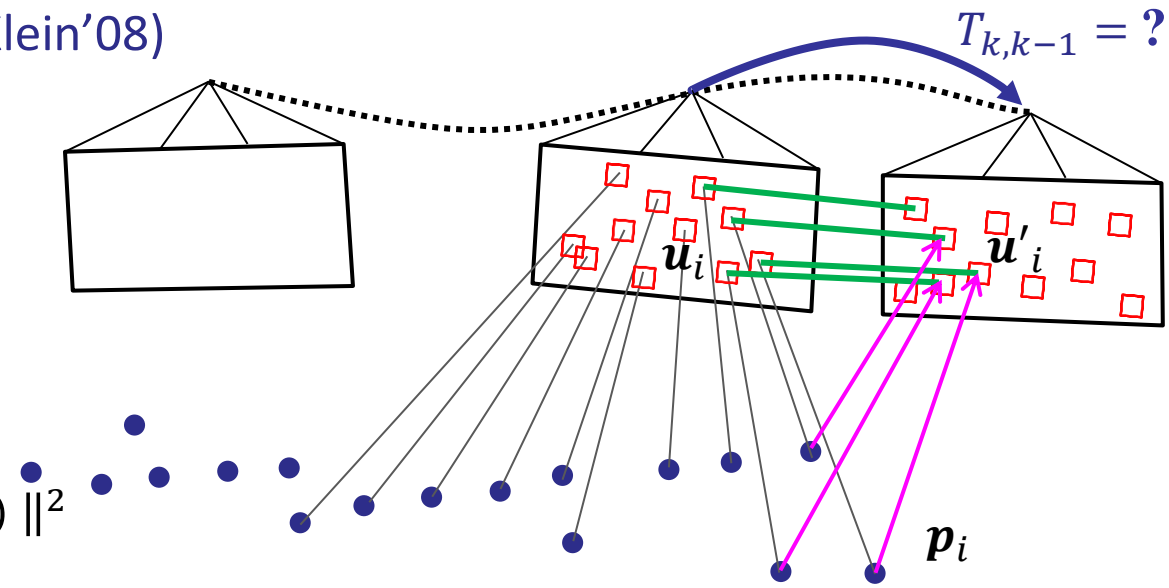
Scaramuzza, Fraundorfer. Visual Odometry, *IEEE Robotics and Automation Magazine*, 2011

Feature-based vs. Direct Methods

Feature-based (e.g., PTAM, Klein'08)

1. Feature extraction
2. Feature matching
3. RANSAC + P3P
4. **Reprojection error** minimization

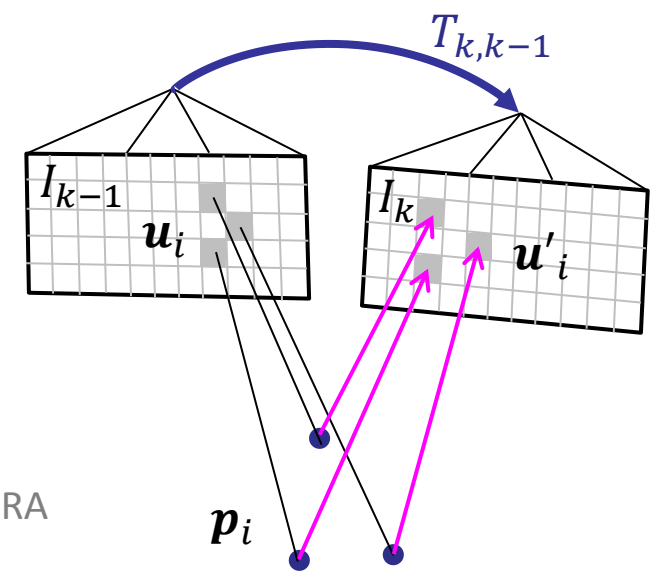
$$T_{k,k-1} = \arg\min_T \sum_i \|\boldsymbol{u}'_i - \pi(\boldsymbol{p}_i)\|^2$$



Direct approaches (e.g., Meilland'13)

1. **Minimize photometric error**

$$T_{k,k-1} = \arg\min_T \sum_i \|I_k(\boldsymbol{u}'_i) - I_{k-1}(\boldsymbol{u}_i)\|^2$$



[Soatto'95, Meilland and Comport, IROS 2013], DVO [Kerl et al., ICRA 2013], DTAM [Newcombe et al., ICCV '11], ...

Feature-based vs. Direct Methods

Feature-based (e.g., PTAM, Klein'08)

1. Feature extraction
2. Feature matching
3. RANSAC + P3P
4. **Reprojection error** minimization

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

- ✓ **Large frame-to-frame motions**
- × **Slow (20-30 Hz) due to costly feature extraction and matching**
- × **Not robust to high-frequency and repetive texture**

Direct approaches

1. **Minimize photometric error**

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

- ✓ **Every pixel in the image can be exploited (precision, robustness)**
- ✓ **Increasing camera frame-rate reduces computational cost per frame**
- × **Limited to small frame-to-frame motion**

Feature-based vs. Direct Methods

Feature-based (e.g., PTAM, Klein'08)

1. Feature extraction
2. Feature matching
3. RANS

✓ **Large frame-to-frame motions**

× **Slow (20-30 Hz) due to costly feature extraction and matching**

Our solution:

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

Combines feature-based and direct methods

1. M

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

✓ **Increasing camera frame-rate reduces computational cost per frame**

× **Limited to small frame-to-frame motion**

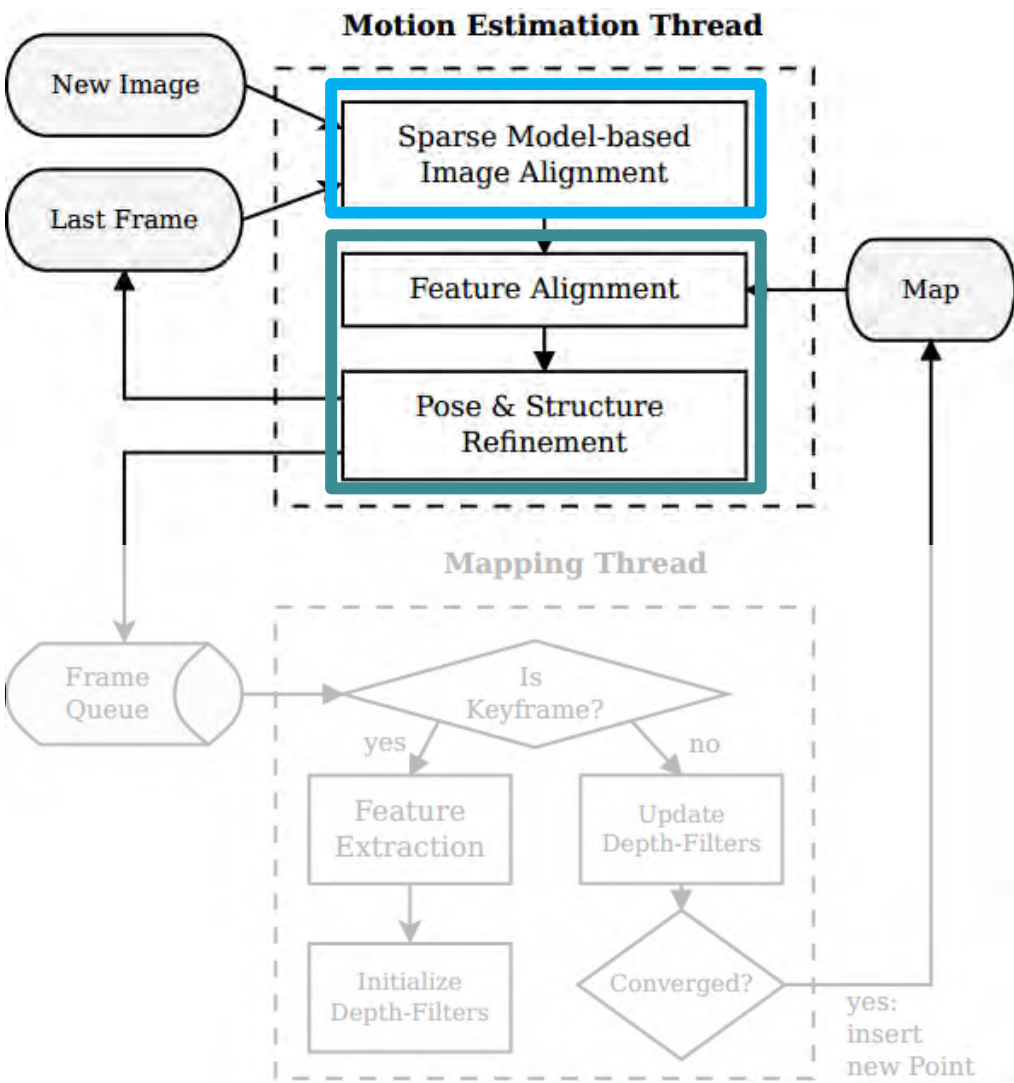
SVO: Semi-Direct Visual Odometry [ICRA'14]

Direct

- Frame-to-frame motion estimation

Feature-based

- Frame-to-Keyframe pose refinement



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

SVO: Semi-Direct Visual Odometry [ICRA'14]

Direct

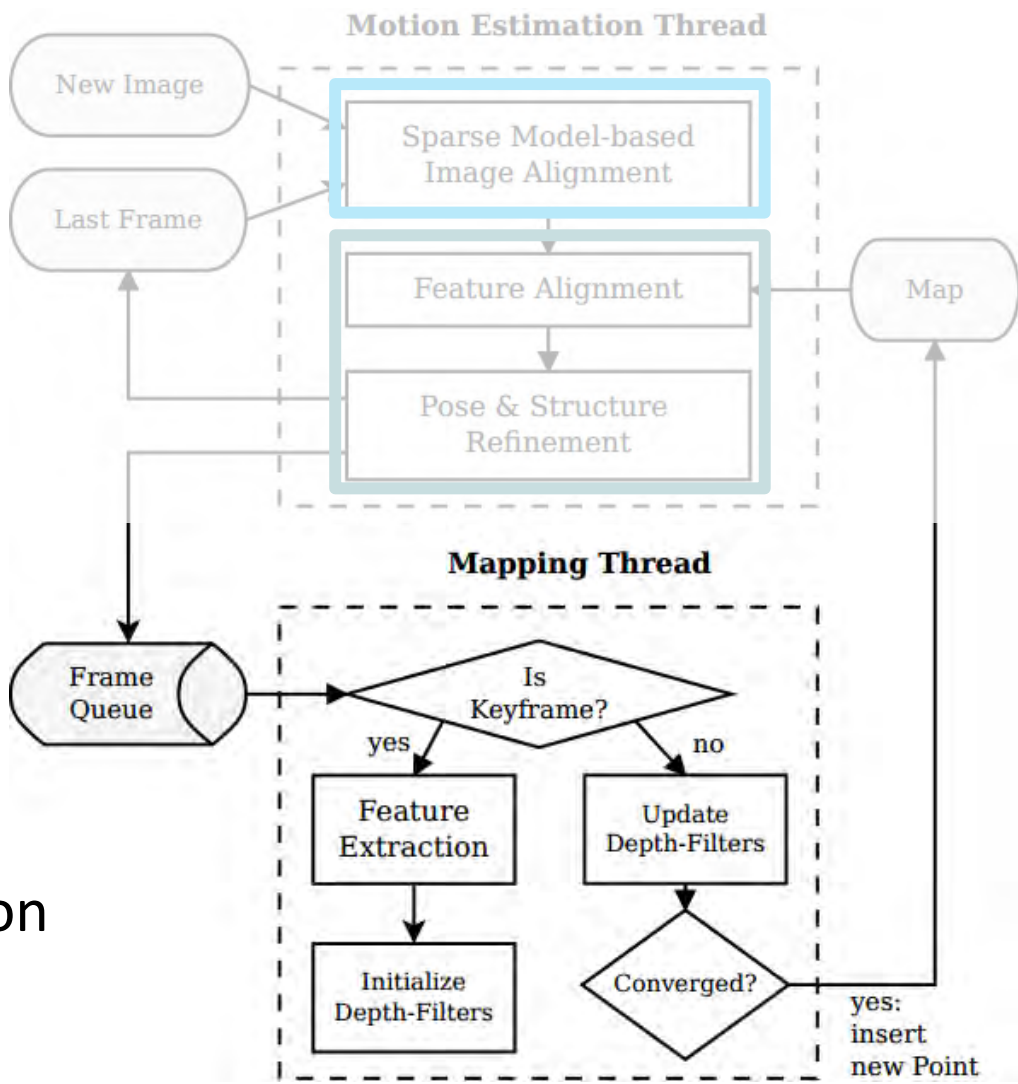
- Frame-to-frame motion estimation

Feature-based

- Frame-to-Kreyframe pose refinement

Mapping

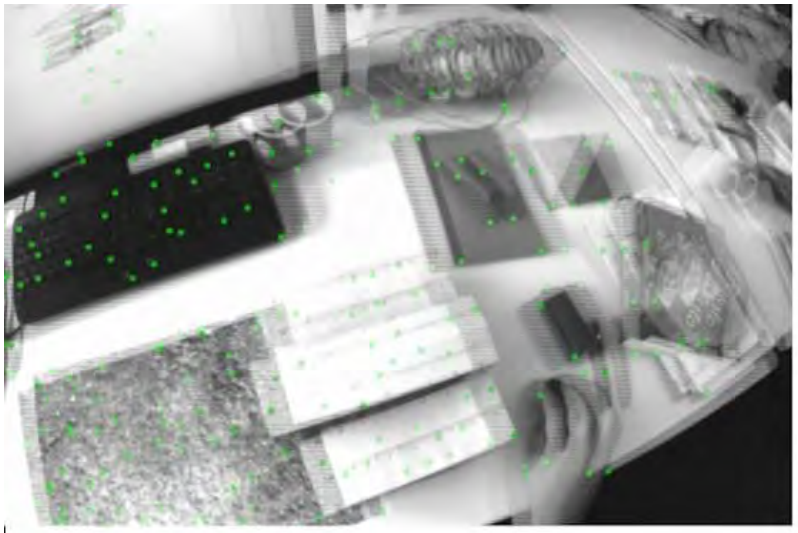
- ➢ Feature extraction only for every keyframe
- ➢ **Probabilistic depth** estimation of 3D points



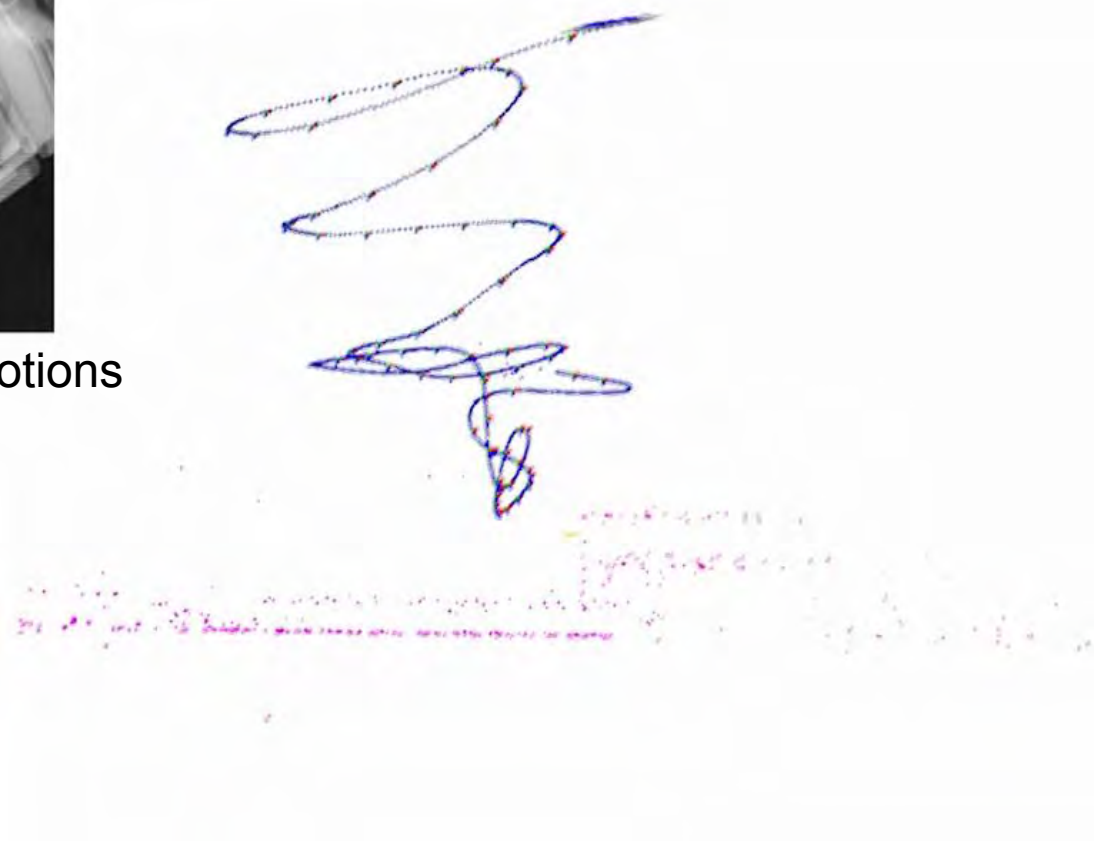
[Forster, Pizzoli, Scaramuzza, «SVO: Semi Direct Visual Odometry», ICRA'14]

SVO: Experiments in real-world environments

Video: YouTube
https://www.youtube.com/watch?v=2YnIMfw6bJY



Robust to fast and abrupt motions



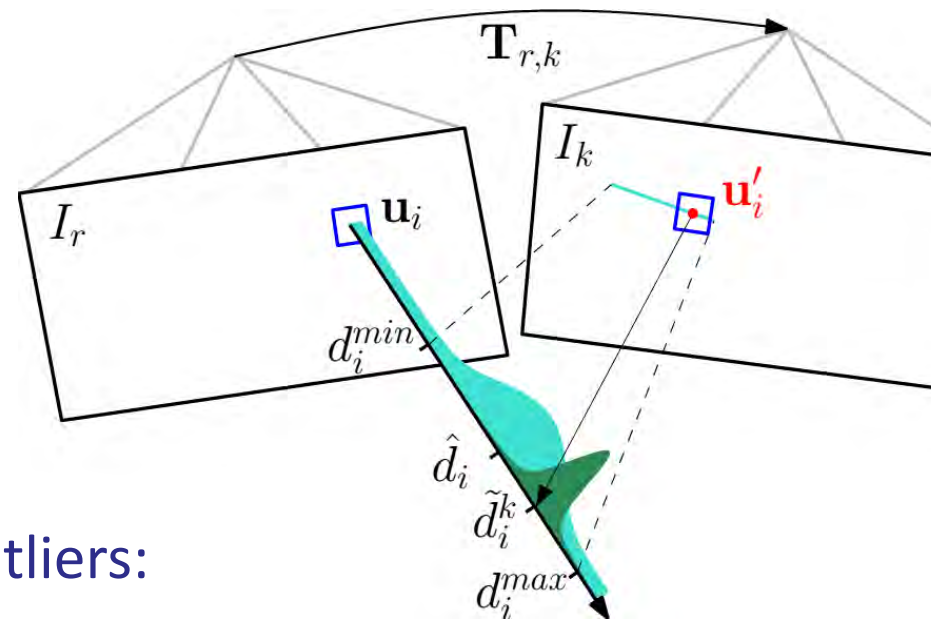
Camera at 70fps

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

Probabilistic Depth Estimation in SVO

Depth-Filter:

- Depth-filter for every new feature
- Recursive Bayesian depth estimation
- Epipolar search using ZMSSD



Measurement Likelihood models outliers:

$$p(\tilde{d}_i^k | d_i, \rho_i) = \rho_i \mathcal{N}(\tilde{d}_i^k | d_i, \tau_i^2) + (1 - \rho_i)\mathcal{U}(\tilde{d}_i^k | d_i^{\min}, d_i^{\max})$$

- 2-Dimensional distribution: Depth $\boldsymbol{d}$ and inliner ratio $\boldsymbol{\rho}$
- Mixture of Gaussian + Uniform
- Inverse depth

[Vogiatzis and Hernández, *«Video-based, Real-Time Multi View Stereo, Image and Vision Computing»*, vol. 29, no. 7, 2011]

Probabilistic Depth Estimation in SVO

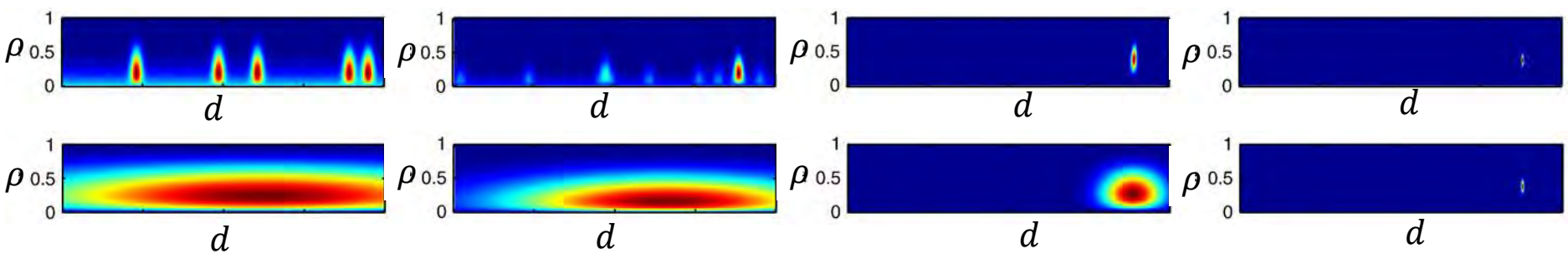
- Based on the model by [Vogiatzis &Hernandez, 2011] but with inverse depth

$$p(\hat{d}, \rho | d_{r+1}, \ldots, d_k) \propto p(\hat{d}, \rho) \prod_k p(d_k | \hat{d}, \rho) \qquad (1)$$

$$p(d_k | \hat{d}, \rho) = \rho \mathcal{N}(d_k | \hat{d}, \tau_k^2) + (1 - \rho) \mathcal{U}(d_k | d_{min}, d_{max}) \qquad (2)$$

- The posterior in (1) can be approximated by

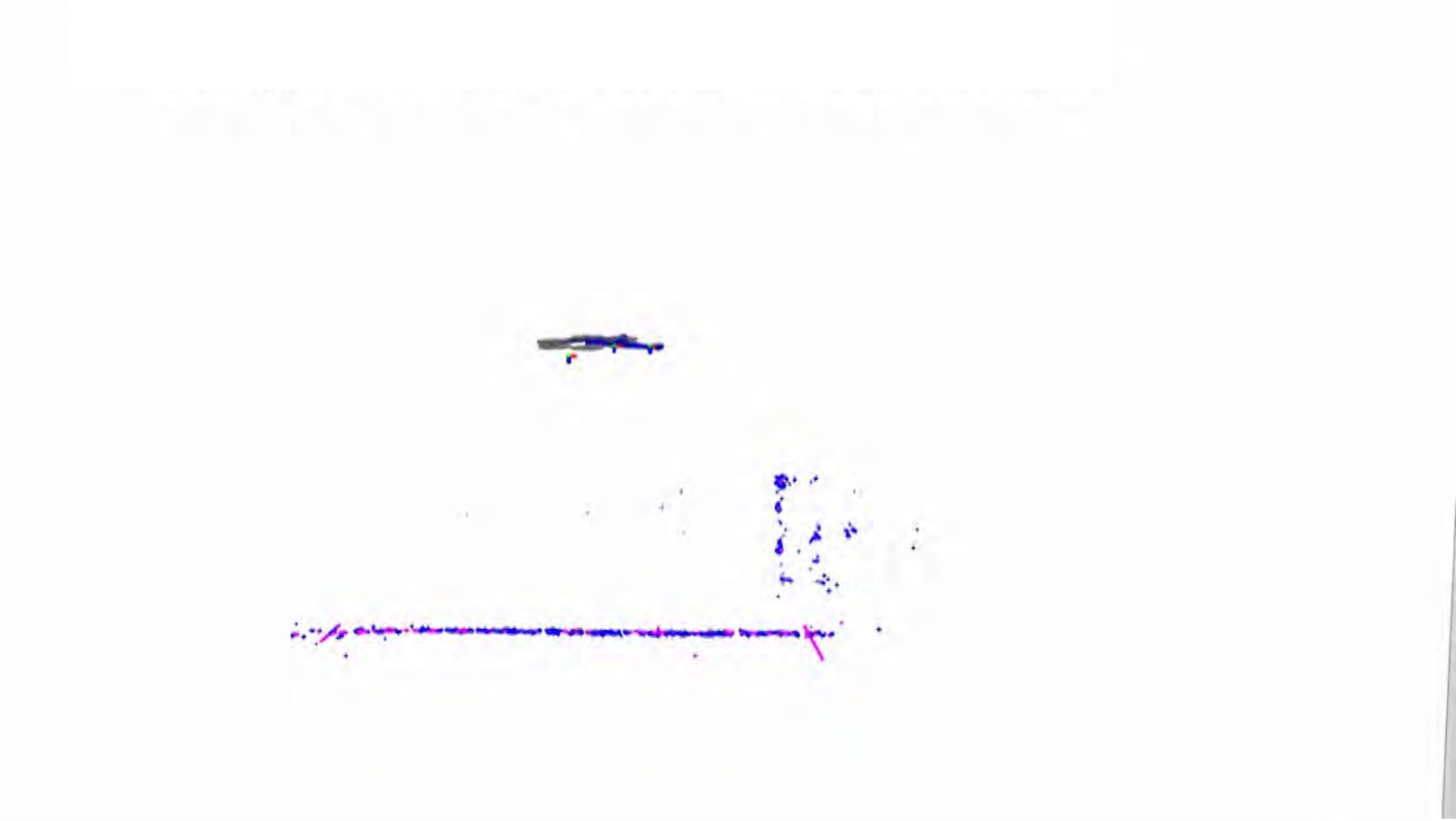
$$q(\hat{d}, \rho | a_k, b_k, \mu_k, \sigma_k^2) = Beta(\rho | a_k, b_k) \mathcal{N}(\hat{d} | \mu_k, \sigma_k^2) \qquad (3)$$



[Vogiatzis and Hernández, "Video Mapping: Real-Time Multi View Stereo, Image and Vision Computing», Vol. 29, no. 7

The parametric model $\{a_k, b_k, \mu_k, \sigma_k^2\}$ describes the pixel depth at time *k*.

Probabilistic Depth Estimation in SVO



[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

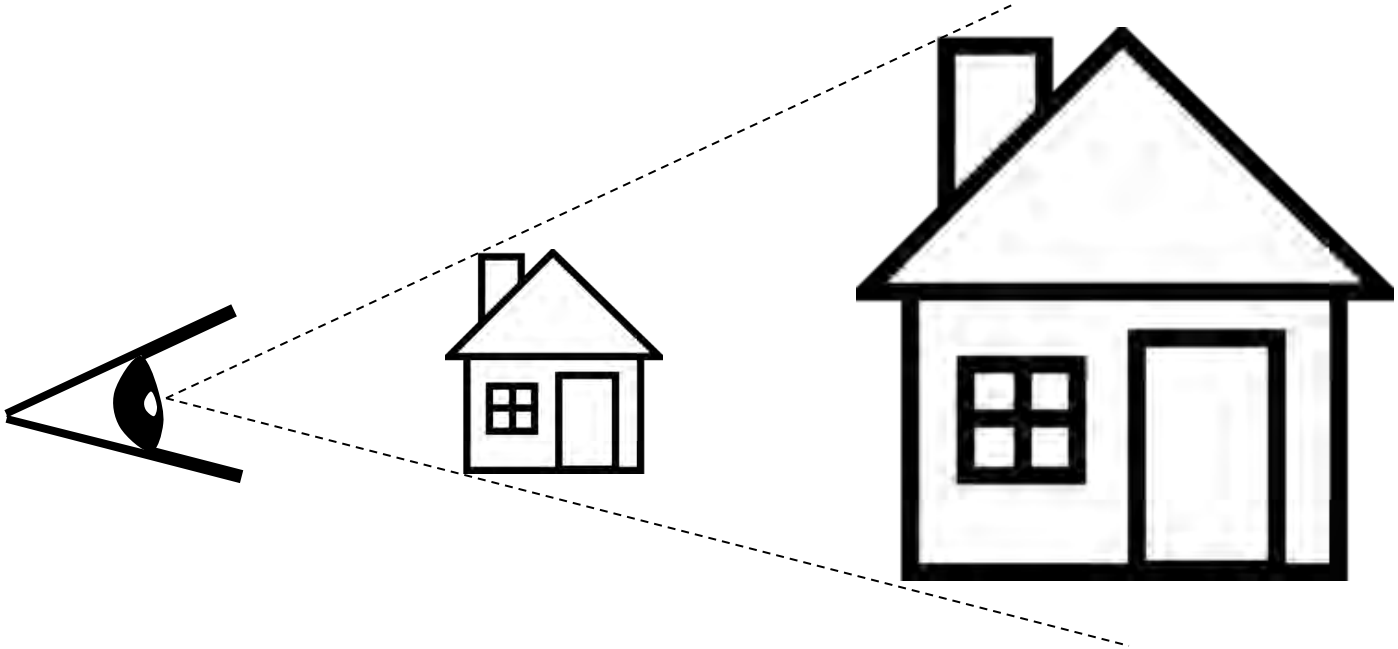
Source Code

- Open Source available at: **github.com/uzh-rpg/rpg_svo**
- Works **with and without ROS**
- **Closed-Source professional edition** available for companies

Absolute Scale Estimation

Scale Ambiguity

- With a single camera, we only know the relative scale
- No information about the *metric scale*



Absolute Scale Estimation

- The absolute pose x is known up to a scale s, thus

$$x = s\tilde{x}$$

- IMU provides accelerations, thus

$$v = v_0 + \int a(t)dt$$

- By derivating the first one and equating them

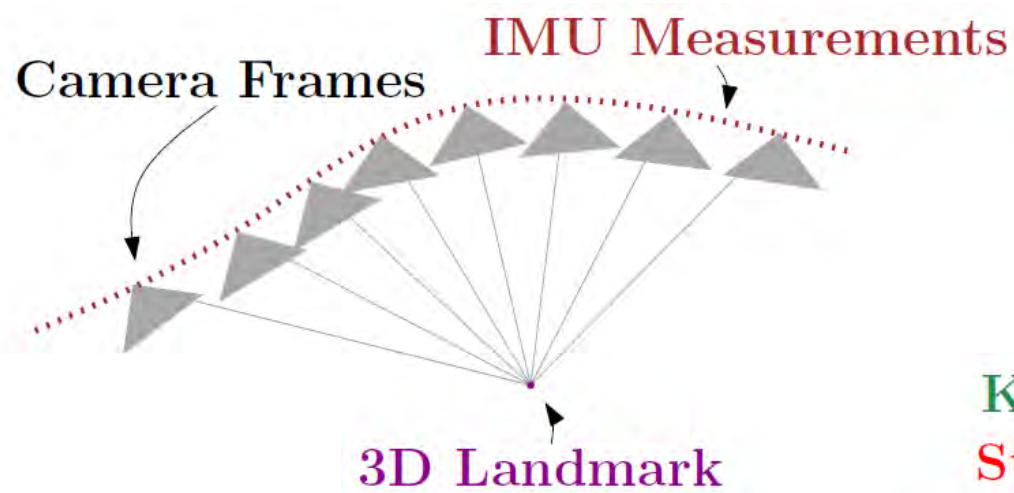
$$s\dot{\tilde{x}} = v_0 + \int a(t)dt$$

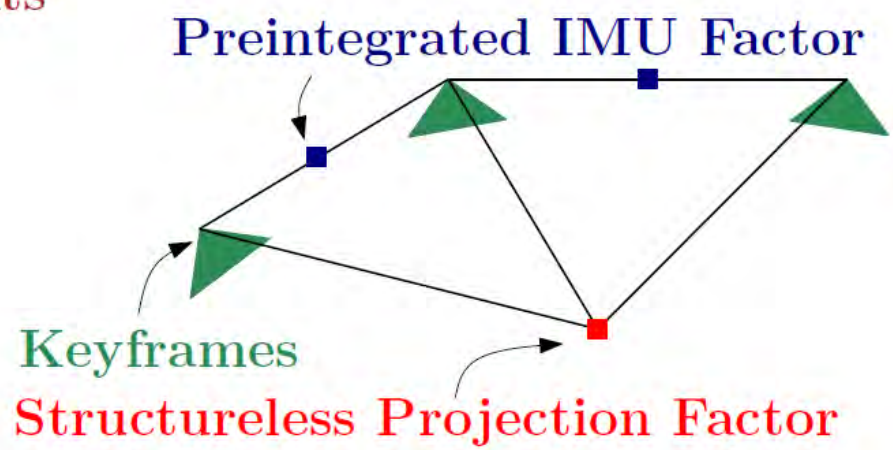
- As shown in [Martinelli, TRO'12], for 6DOF, both s and v_0 can be determined in closed form from a **single feature observation and 3 views**

The scale and velocity can then be tracked using

- **Filter-based approaches**
  - **Losely-coupled** approaches [Lynen et al., IROS'13]
  - **Tightly-coupled** approached (e.g., Google TANGO) [Mourikis & Roumeliotis, TRO'12]
- **Optimization-based approaches** [Leutenegger, RSS'13], [Forster, Scaramuzza, RSS'14]

Visual-Inertial Fusion [RSS'15]





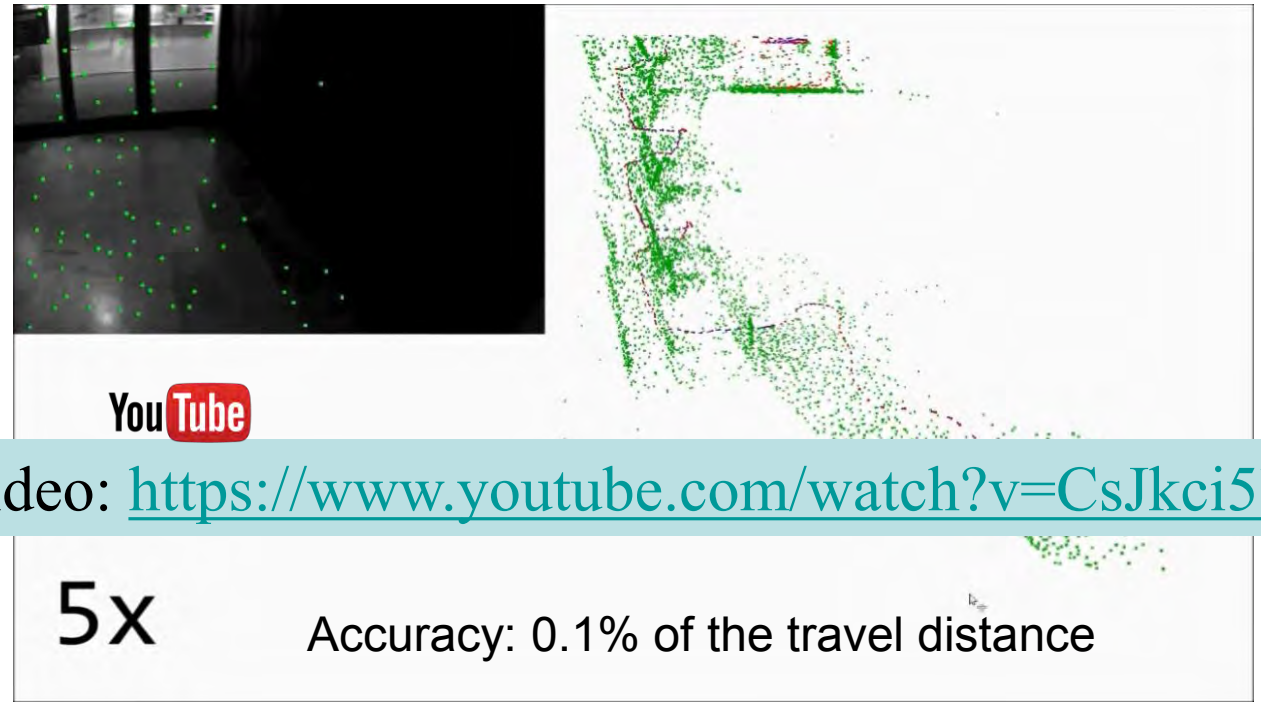
- Fusion is solved as a non-linear optimization problem (no Kalman filter):
- Increased accuracy over filtering methods

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

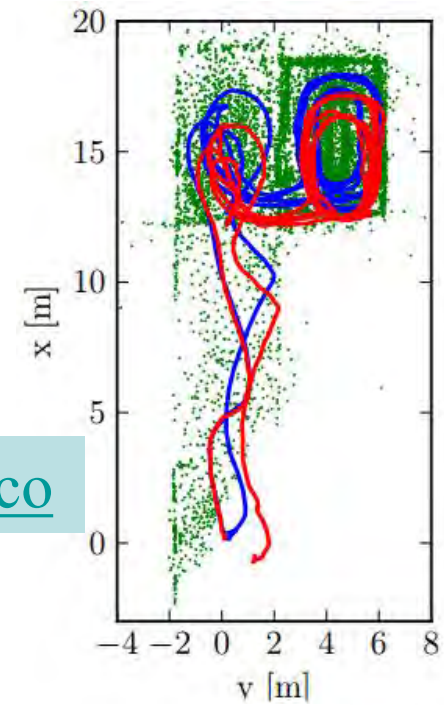
IMU residuals *Reprojection residuals*

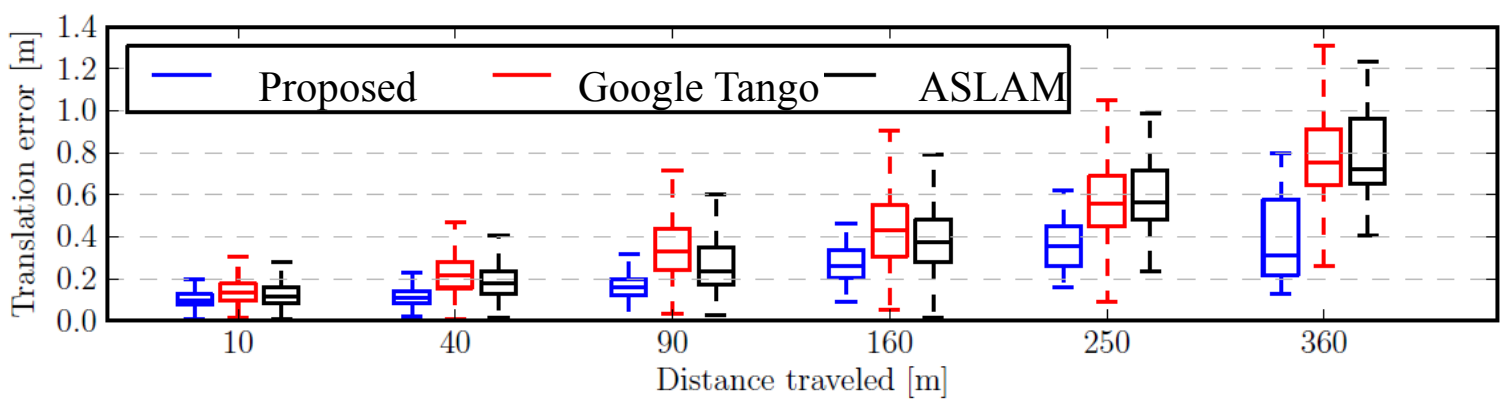
[Forster, Carlone, Dellaert, Scaramuzza, IMU Preintegration on Manifold for efficient Visual-Inertial Maximum-a-Posteriori Estimation, RSS'15, **Best Paper Award Finalist**]

Comparison with Previous Works



Video: https://www.youtube.com/watch?v=CsJkci5lfco





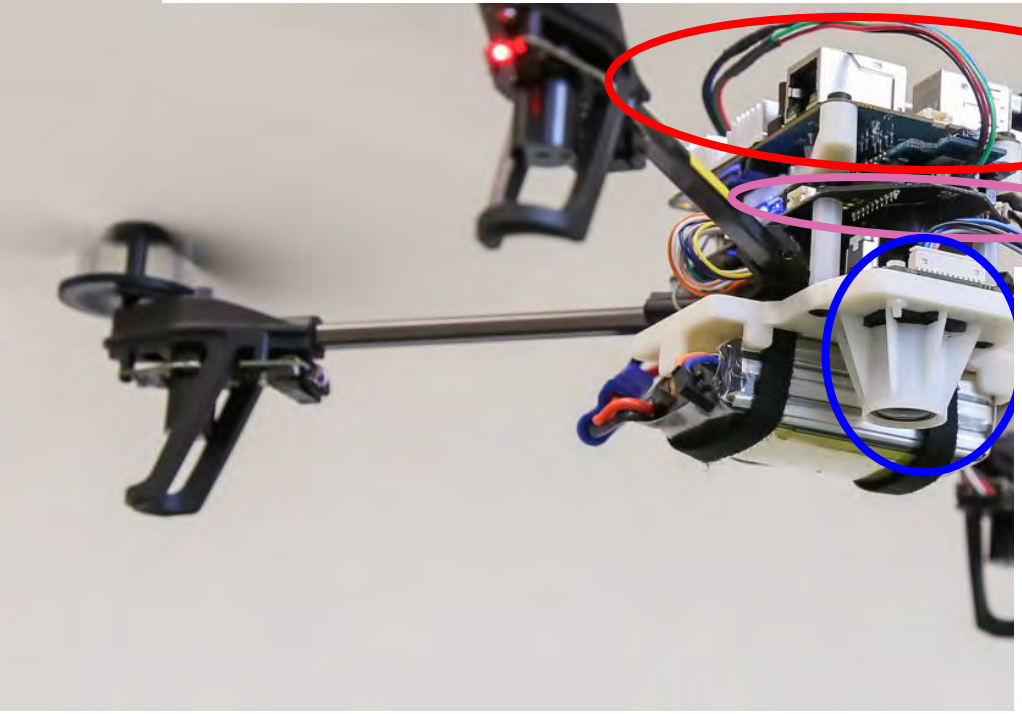
[Forster, Carlone, Dellaert, Scaramuzza, IMU Preintegration on Manifold for efficient Visual-Inertial Maximum-a-Posteriori Estimation, RSS'15, **Best Paper Award Finalist**]

Integration on a Quadrotor Platform

Quadrotor System

Odroid U3 Computer

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



450 grams

Flight Results: Hovering

RMS error: 5 mm, 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**.

Flight Results: Indoor, aggressive flight

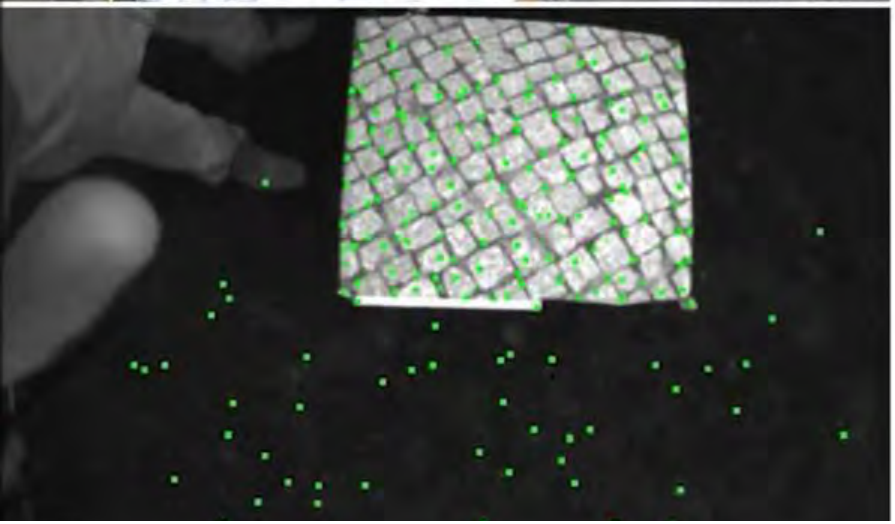
Speed: 4 m/s, height: 1.5 m – Down-looking camera

Video: https://www.youtube.com/watch?v=l3TCiCe_T3g

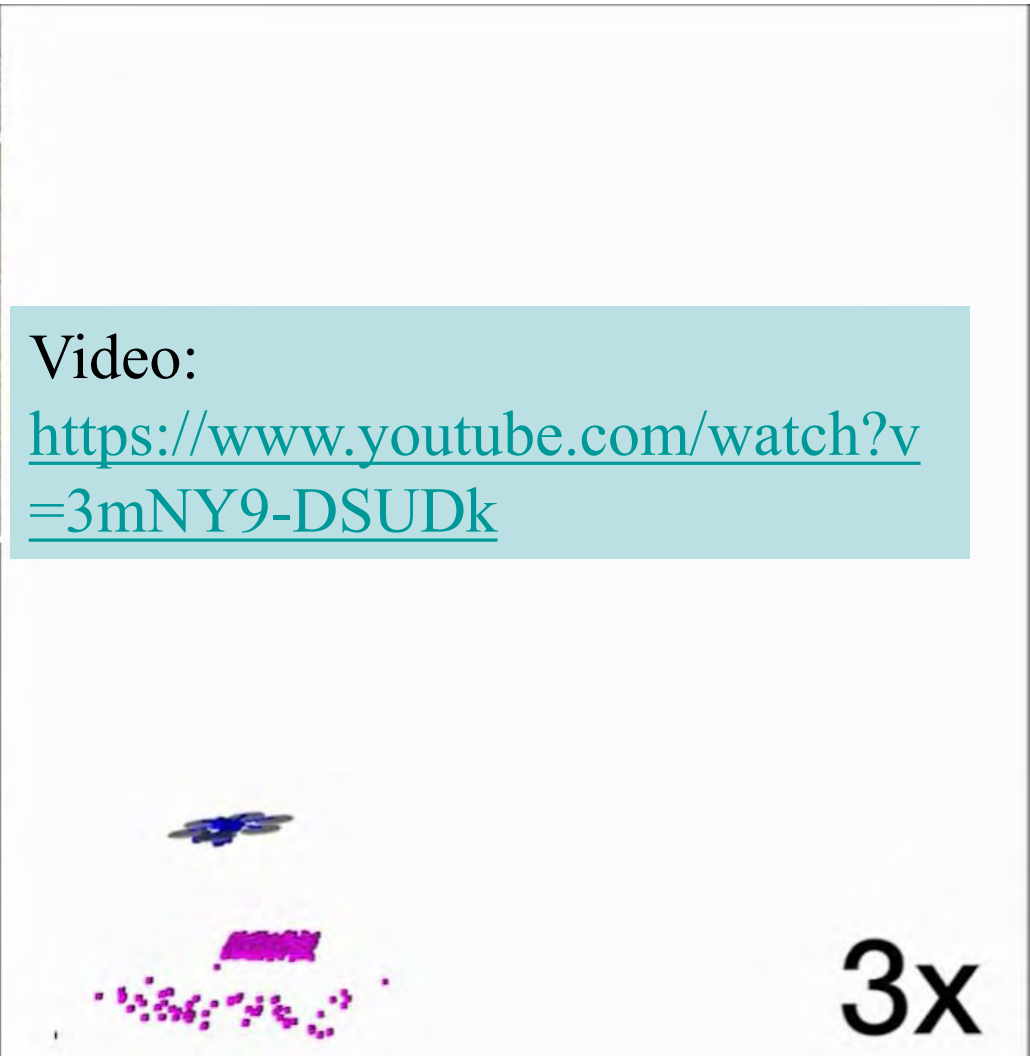
Faessler, Fontana, Forster, Mueggler, Pizzoli, Scaramuzza, Autonomous, Vision-based Flight and Live Dense 3D Mapping with a Quadrotor Micro Aerial Vehicle, **Journal of Field Robotics, 2015**.

Autonomous Vision-based Flight over Mockup Disaster Zone

Firefighters' training area, Zurich



Video:
https://www.youtube.com/watch?v=3mNY9-DSUDk

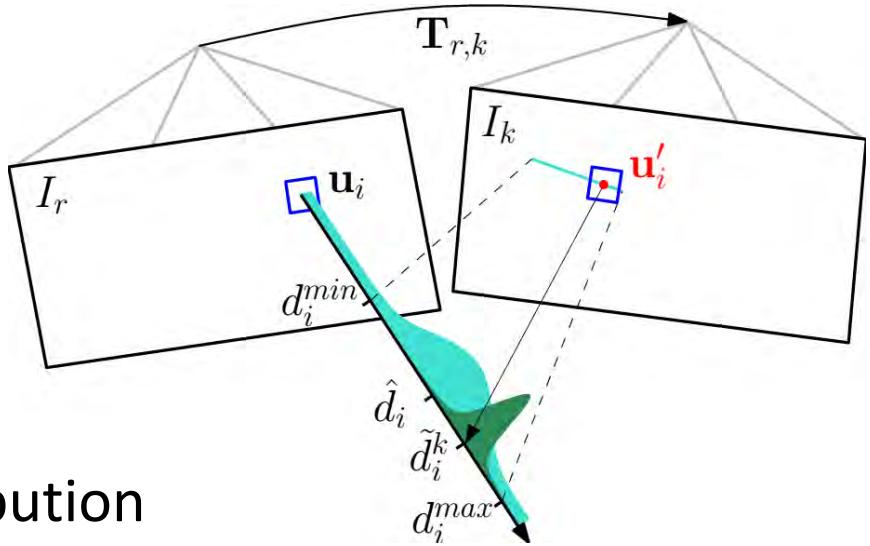


Faessler, Fontana, Forster, Mueggler, Pizzoli, Scaramuzza, Autonomous, Vision-based Flight and Live Dense 3D Mapping with a Quadrotor Micro Aerial Vehicle, **Journal of Field Robotics, 2015**.

Probabilistic Depth Estimation

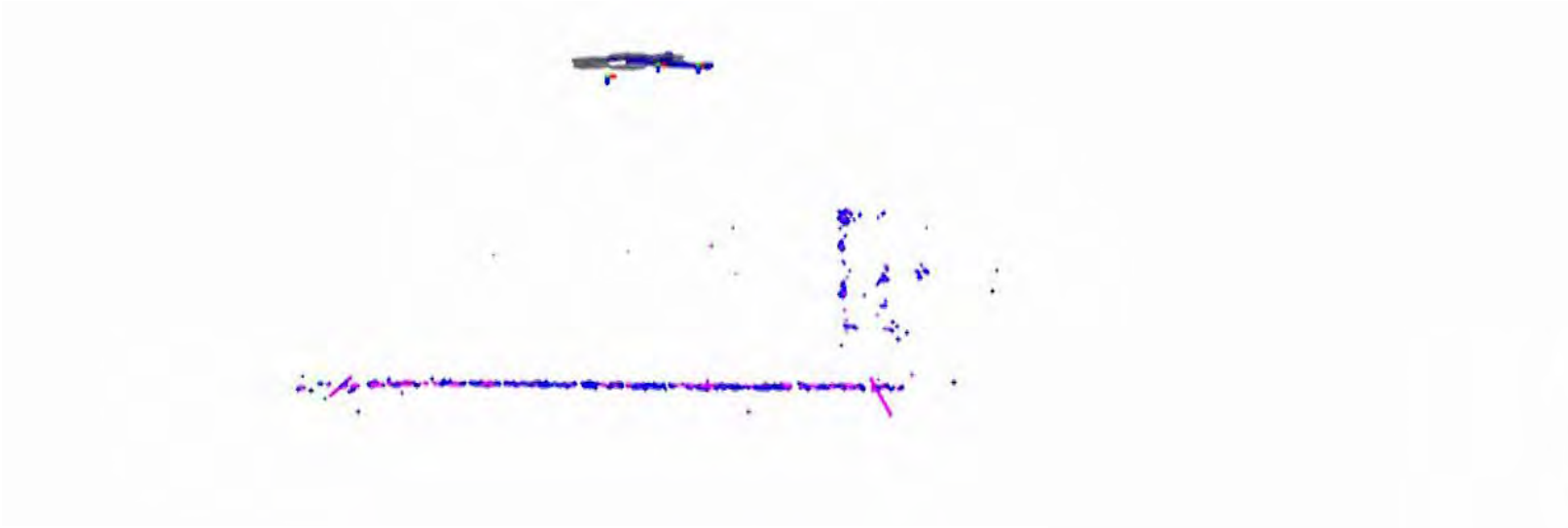
Depth-Filter:

- **Depth Filter** for every feature
- **Recursive Bayesian** depth estimation



Mixture of Gaussian + Uniform distribution

$$p(\tilde{d}_i^k | d_i, \rho_i) = \rho_i \mathcal{N}(\tilde{d}_i^k | d_i, \tau_i^2) + (1 - \rho_i)\mathcal{U}(\tilde{d}_i^k | d_i^{\min}, d_i^{\max})$$



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

Robustness to Dynamic Objects and Occlusions

- Depth uncertainty is crucial for safety and robustness
- Outliers are caused by wrong data association (e.g., moving objects, distortions)
- Probabilistic depth estimation models outliers

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**.

Failure Recovery [ICRA'15]

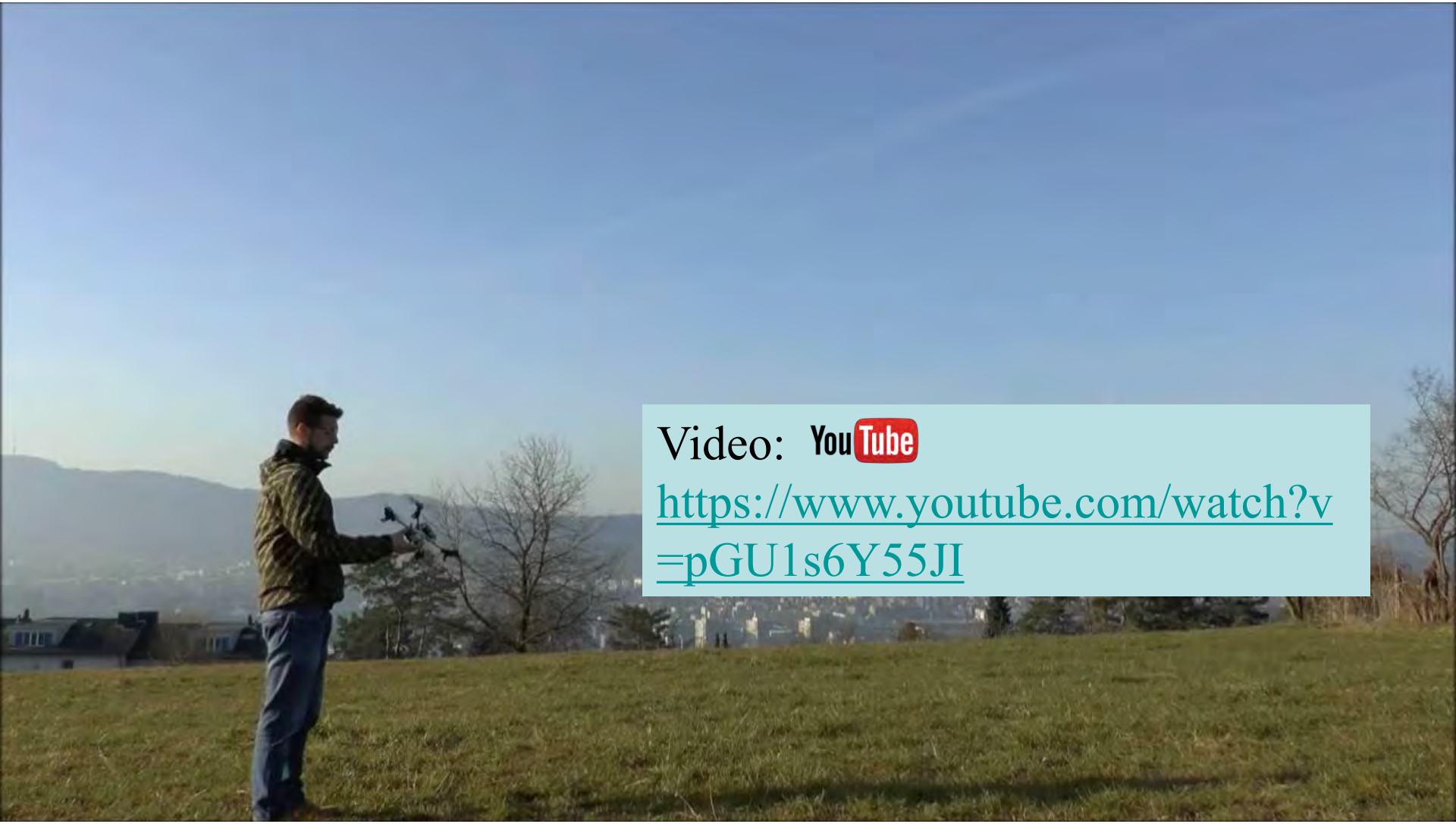
- Loss of GPS
- From aggressive flight
- Visual tracking

IEEE SPECTRUM

Article: http://spectrum.ieee.org/automaton/robotics/aerial-robots/aggressive-flight-quadrotor-recovery

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**.

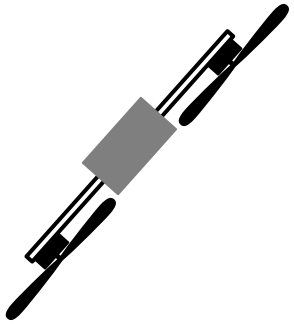
Automatic Failure Recovery from Aggressive Flight [ICRA'15]



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**.

Recovery Stages

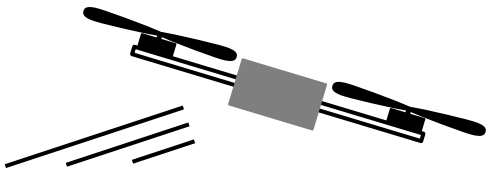
Throw



IMU

Recovery Stages

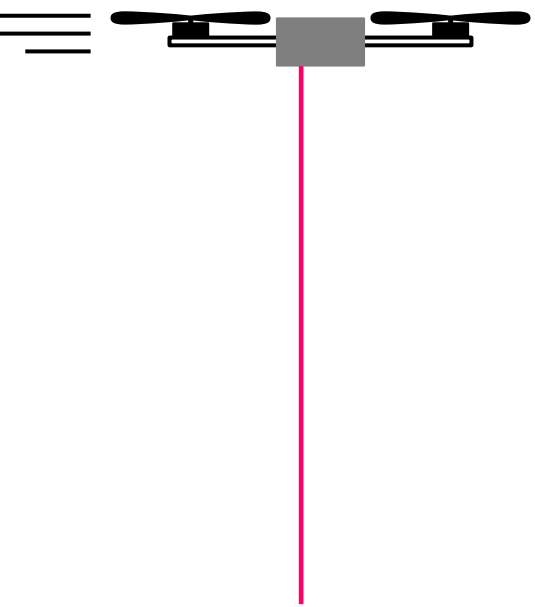
Attitude Control



IMU

Recovery Stages

Attitude + Height Control

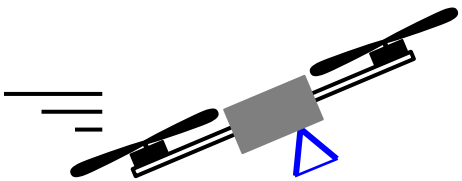


IMU

Distance Sensor

Recovery Stages

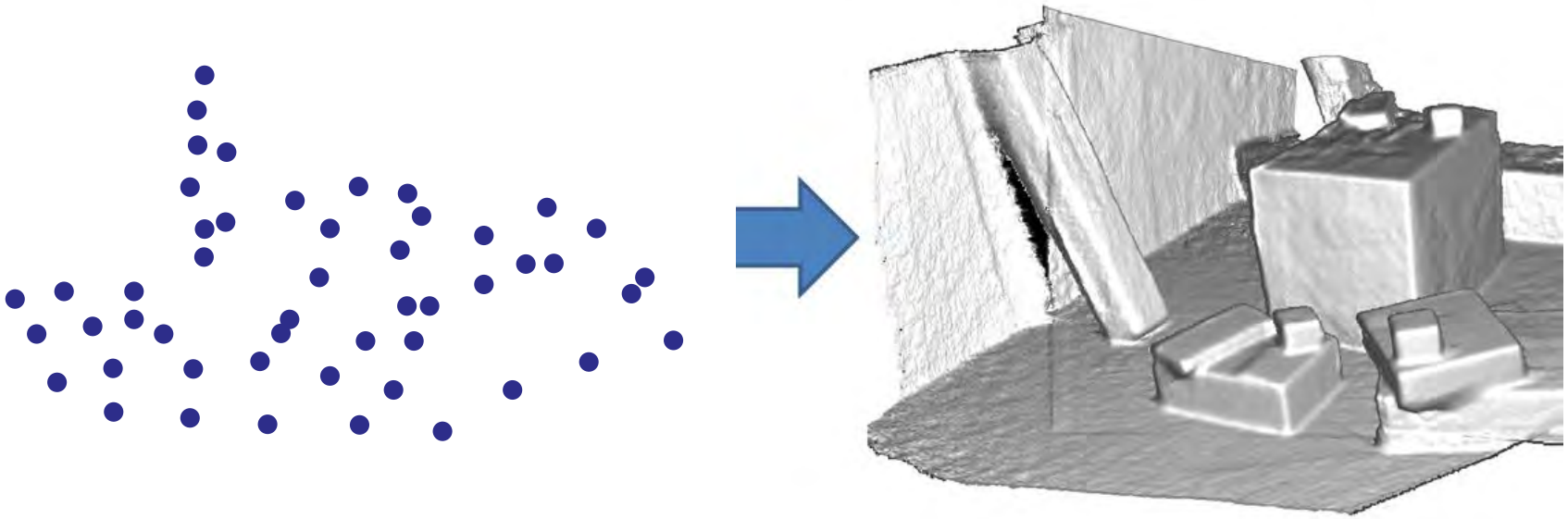
Break Velocity



IMU

Camera

From Sparse to Dense 3D Models

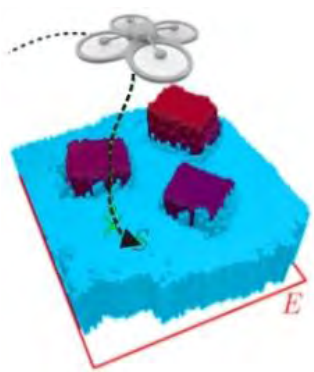


[M. Pizzoli, C. Forster, D. Scaramuzza, REMODE: Probabilistic, Monocular Dense Reconstruction in Real Time, ICRA'14]

Dense Reconstruction in Real-Time

Goal: estimate depth of every pixel in real time

- **Pros**:
  - Advantageous for **environment interaction** (e.g., *collision avoidance, landing, grasping, industrial inspection*, etc)
  - Higher position accuracy
- **Cons**: computationally expensive (requires GPU)



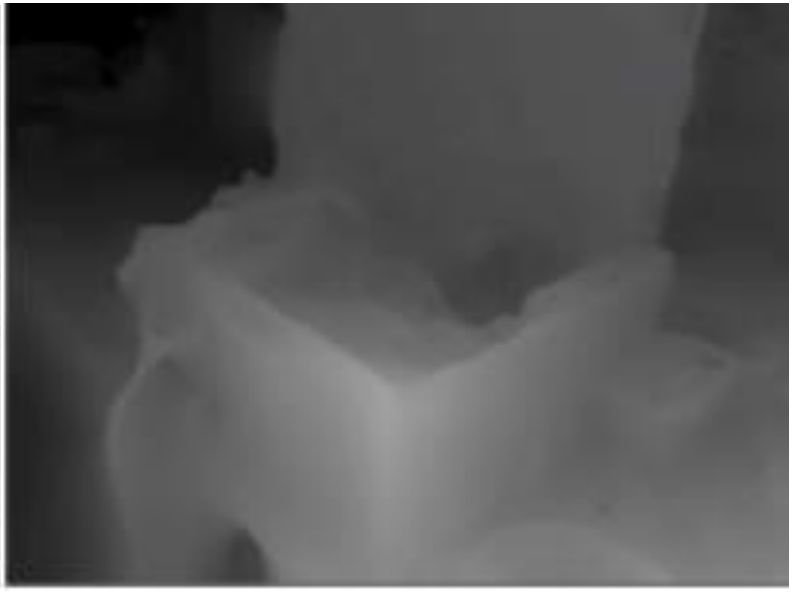
[ICRA'15]

[IROS'13, SSRR'14]

Dense Reconstruction Pipeline

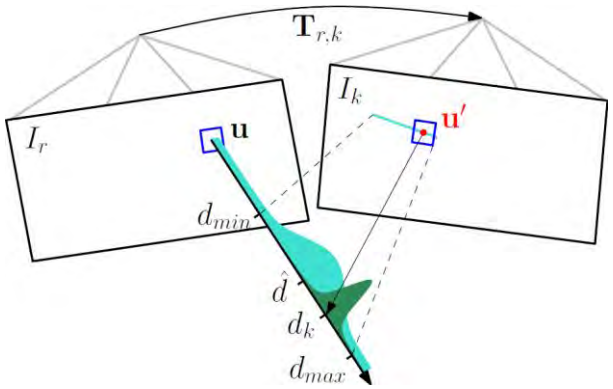
- **Local methods**
  - Estimate depth for every pixel independently using **photometric cost aggregation**
- **Global methods**
  - Refine the depth surface as a whole by enforcing smoothness constraint (“*Regularization*”)

$$E(d) = \underbrace{E_d(d)}_{\text{Data term}} + \underbrace{\lambda E_s(d)}_{\text{Regularization term: penalizes } \textit{non-smooth} \text{ surfaces}}$$



[Newcombe et al. 2011]

REMODE: Probabilistic Monocular Dense Reconstruction [ICRA'14]



- **Pose estimation done by SVO**
- **Track independently** every pixel using the same recursive **Bayesian depth estimation of SVO**

$$p(\tilde{d}_i^k | d_i, \rho_i) = \rho_i \mathcal{N}(\tilde{d}_i^k | d_i, \tau_i^2) + (1 - \rho_i)\mathcal{U}(\tilde{d}_i^k | d_i^{\min}, d_i^{\max})$$

- A ***regularized depth map*** *F*(u) is computed from the noisy depth map *D*(u) as

$$\min_F \int_\Omega \{G(\mathbf{u}) \|\nabla F(\mathbf{u})\|_\epsilon + \lambda \|F(\mathbf{u}) - D(\mathbf{u})\|_1\} \, d\mathbf{u}$$

where

$$G(\mathbf{u}) = \mathbb{E}_\rho[q](\mathbf{u}) \frac{\sigma^2(\mathbf{u})}{\sigma_{max}^2} + \{1 - \mathbb{E}_\rho[q](\mathbf{u})\}$$

Minimization is done using [Chambolle & Pock, 2011]

[M. Pizzoli, C. Forster, D. Scaramuzza, REMODE: Probabilistic, Monocular Dense Reconstruction in Real Time, ICRA'14]

REMODE: Probabilistic Monocular Dense Reconstruction [ICRA'14]

Running at 50 Hz on GPU on a Lenovo W530, i7

Video: YouTube
https://www.youtube.com/watch?v=QTKd5UWCG0Q

Monocular dense reconstruction
in real-time from a hand-held camera

Open source

[M. Pizzoli, C. Forster, D. Scaramuzza, REMODE: Probabilistic, Monocular Dense Reconstruction in Real Time, ICRA'14]

Autonomus, Flying 3D Scanning [JFR'15]

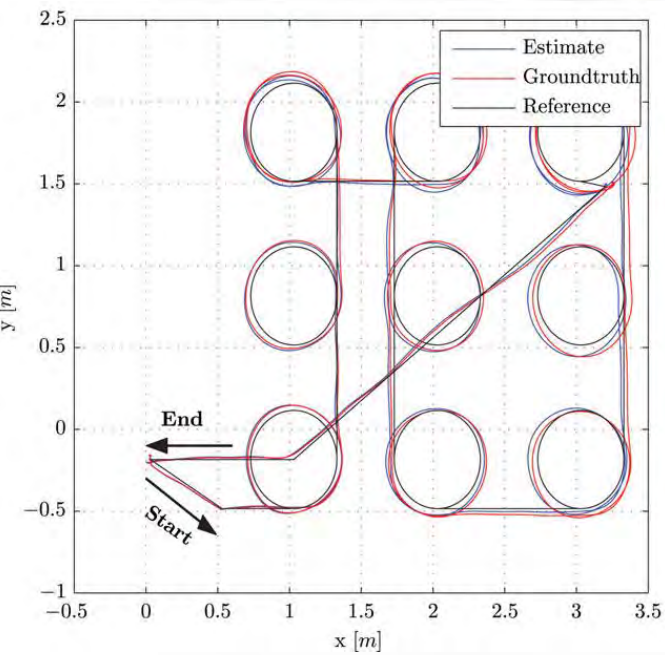
- Sensing, control, state estimation run onboard at 50 Hz (Odroid U3, ARM Cortex A9)
- Dense reconstruction runs live on video streamed to laptop (Lenovo W530, i7)

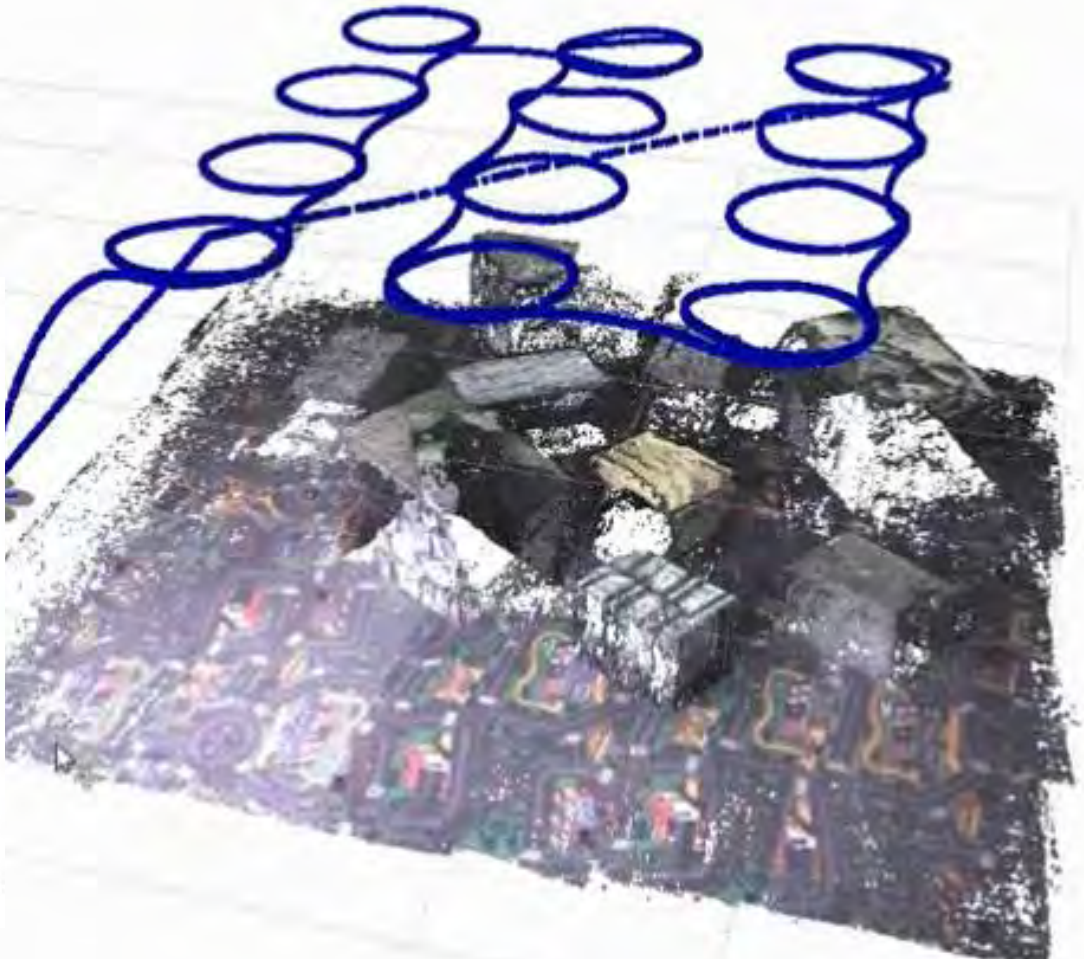
Video: YouTube
https://www.youtube.com/watch?v=7-kPiWaFYAc

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**.

Autonomus, Flying 3D Scanning [JFR'15]

- Sensing, control, state estimation run onboard at 50 Hz (Odroid U3, ARM Cortex A9)
- Dense reconstruction runs live on video streamed to laptop (Lenovo W530, i7)





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**.

Applications: Industrial Inspection

Industrial collaboration with Parrot-SenseFly targets:

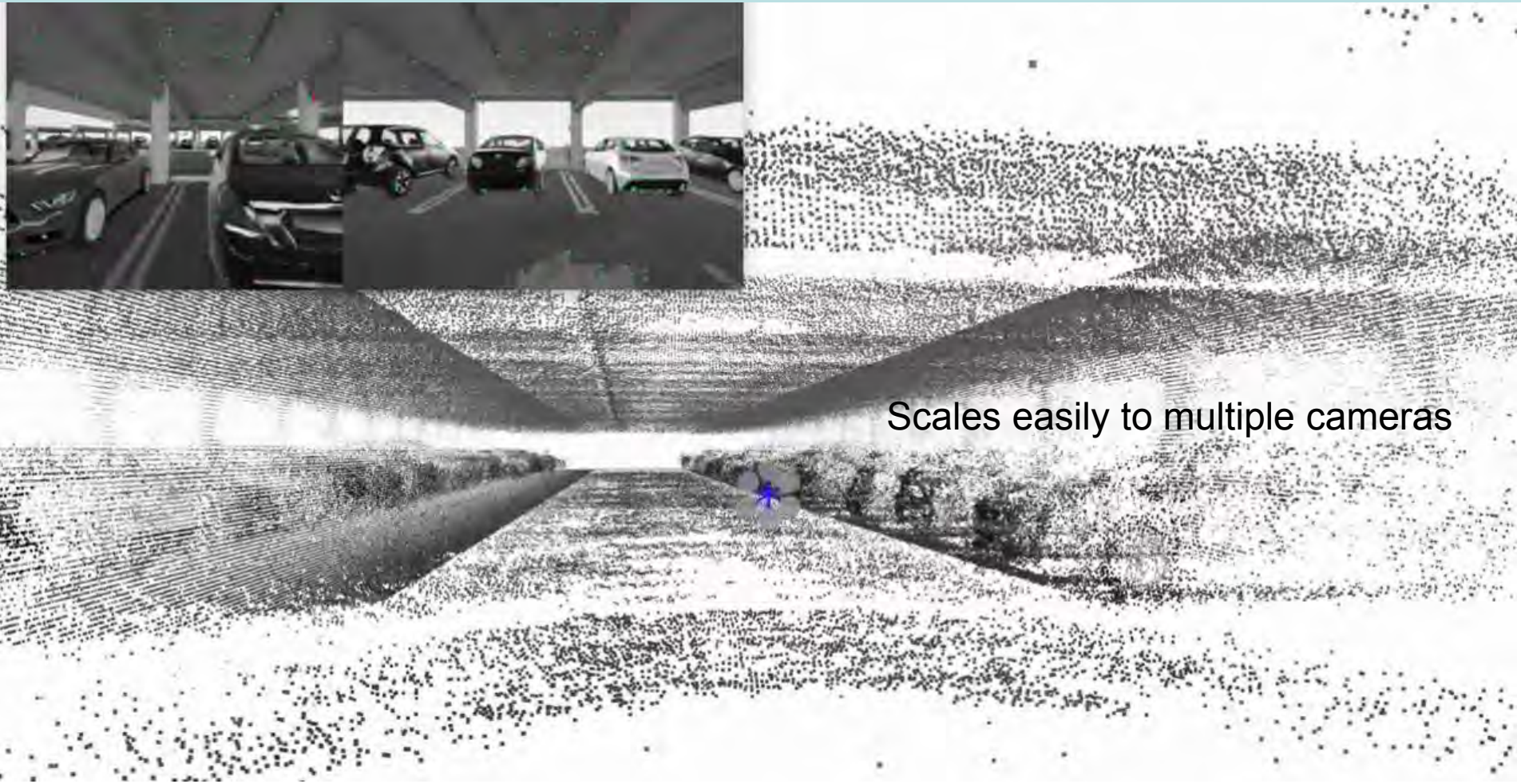
- Real-time dense reconstruction with 5 cameras
- Vision-based navigation
- Dense 3D mapping in real time

Video: https://www.youtube.com/watch?v=gr00Bf0AP1k

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**.

Automotive: 4 fisheye Cameras

Video: https://www.youtube.com/watch?v=gr00Bf0AP1k

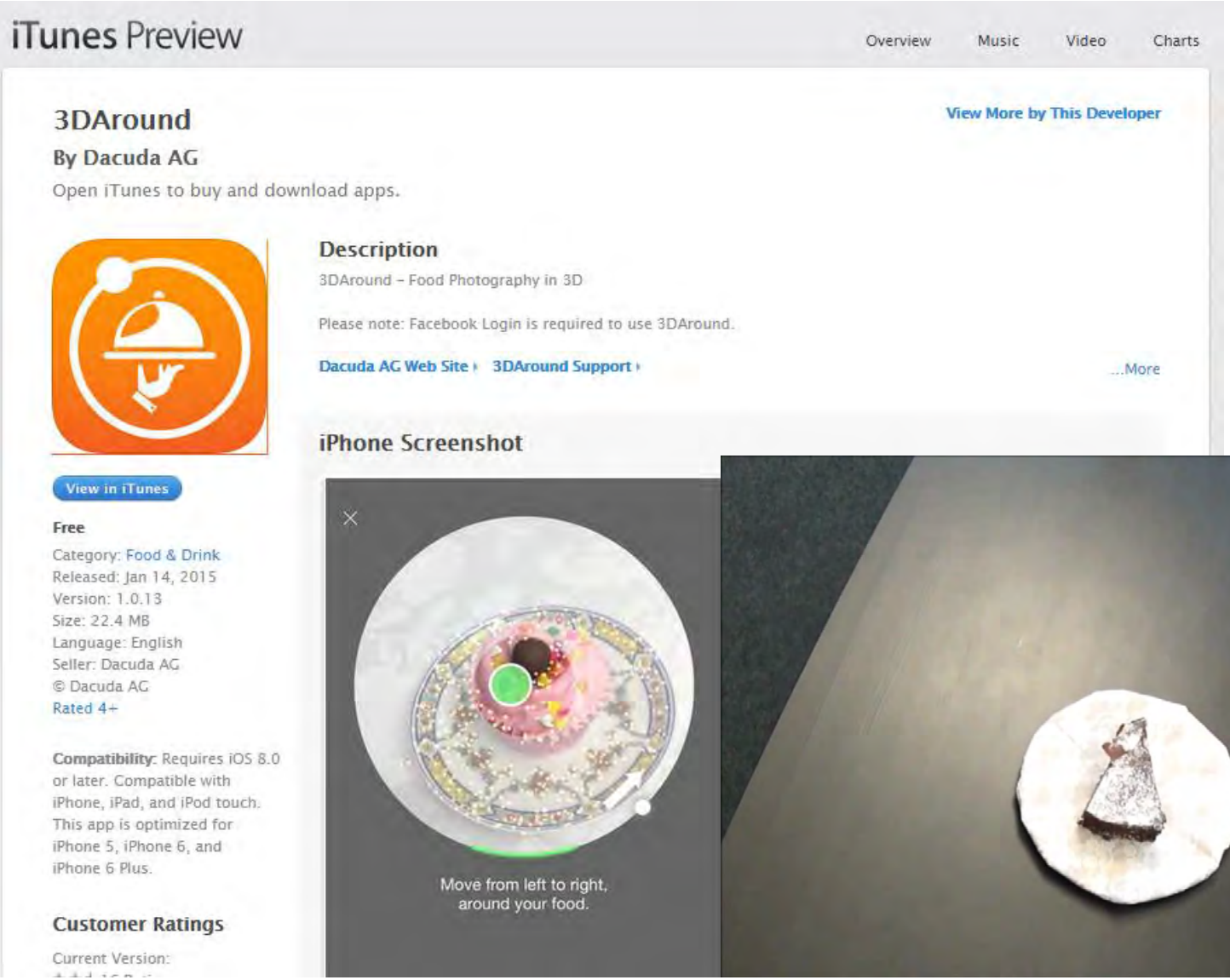


[Forster, Pizzoli, Scaramuzza, «SVO: Semi Direct Visual Odometry», ICRA'14]

Inspection of CERN tunnels

- **Problem**: inspection of CERN tunnels currently done by technicians, who expend much of their annual quota of safe radiation dose
- **Goal**: inspection of LHC tunnel with autonomous drone
- **Challenge**: low illumination, cluttered environment

Try our iPhone App: *3DAround*

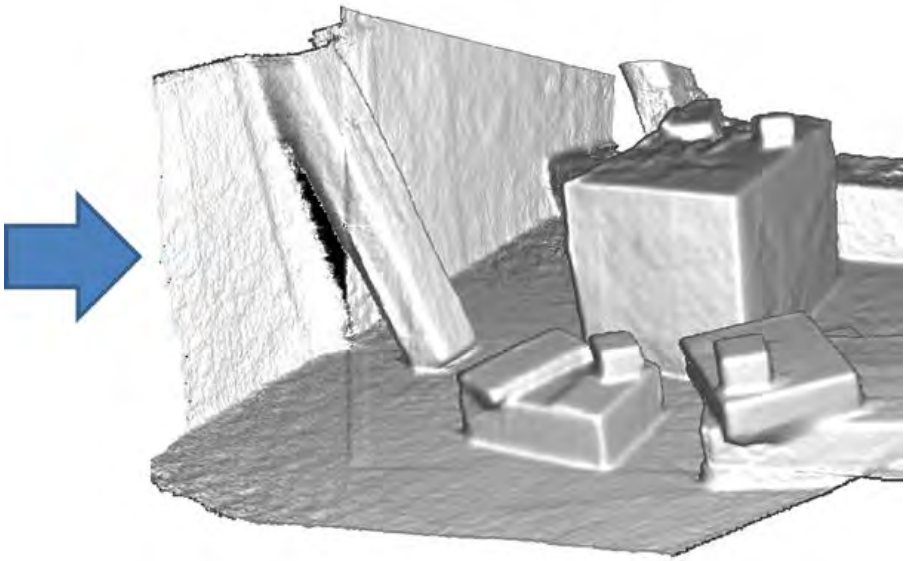


Dacuda

Appearance-based Active Vision

Active Dense Reconstruction [RSS'14]

What's the **optimal motion** to **reconstruct a scene** from a **monocular camera** attached to a **flying robot**?



Forster, Pizzoli, Scaramuzza, Appearance-based Active, Monocular, Dense Reconstruction for Micro Aerial Vehicles, RSS 2014.

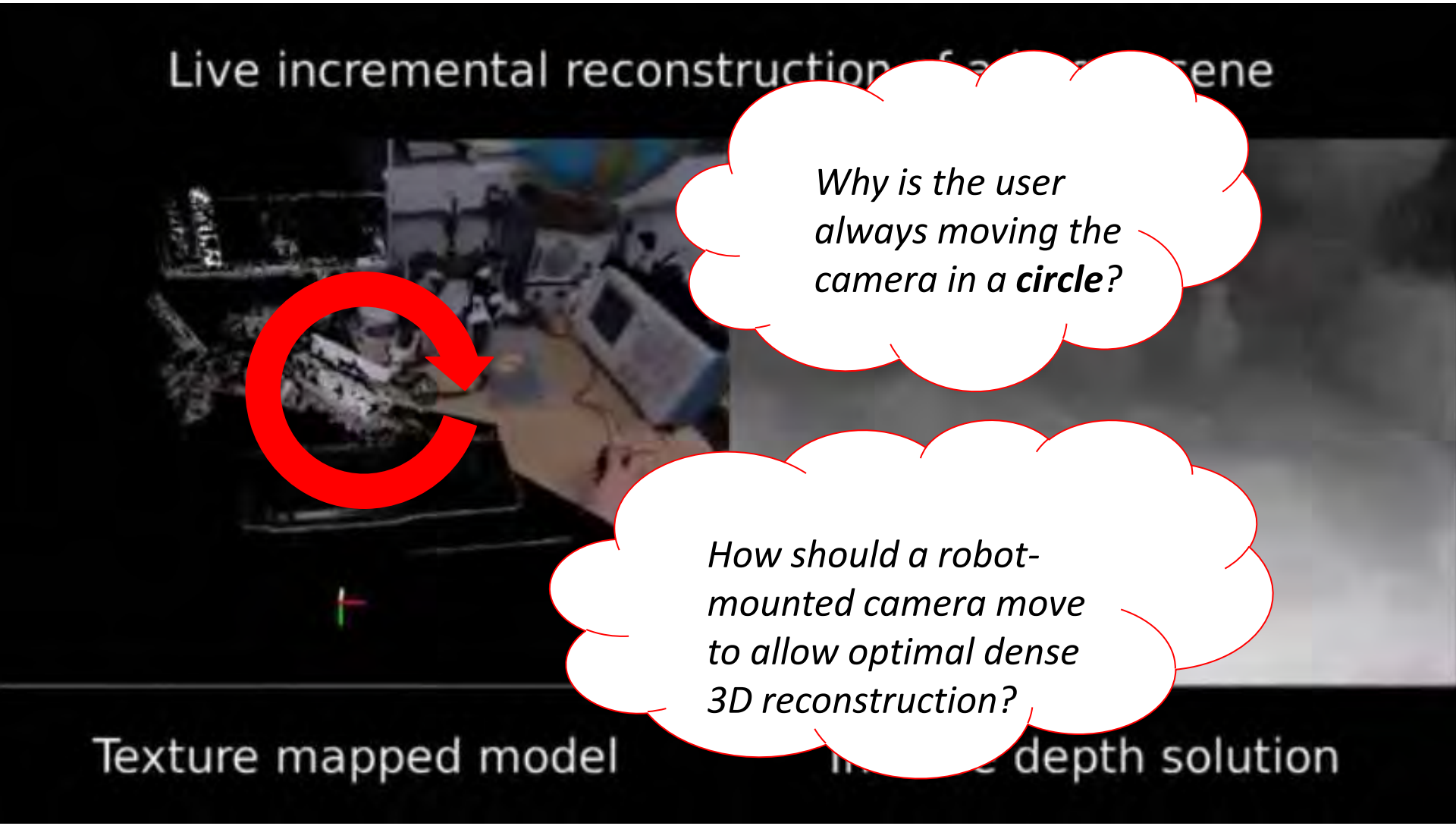
Let’s have a look at **passive** dense reconstruction with **hand-held cameras**

REMODE [Pizzoli, Forster, Scaramuzza 2014]

Monocular dense reconstruction
in real-time from a hand-held camera

Let's have a look at **passive** dense reconstruction with **hand-held cameras**

DTAM [Newcombe et al., ICCV 2011]



*Why is the user always moving the camera in a **circle**?*

How should a robot-mounted camera move to allow optimal dense 3D reconstruction?

Related Work on Active Perception

- View Path Planning, Next-Best-View *[Bajcsy'88, Blake'88]*
- Active SLAM and Exploration *[Davison & Murray'02, Stachniss'05, Vidal-Calleja'10, Dissanayake'12]*

Limitation:

State-of-the-art approaches retain **only geometric information** while discarding the **photometric information** (i.e., texture)

Our solution:

Maximize the expected information gain (i.e., map accuracy), on the basis of scene structure and **photometric information** (i.e., texture).

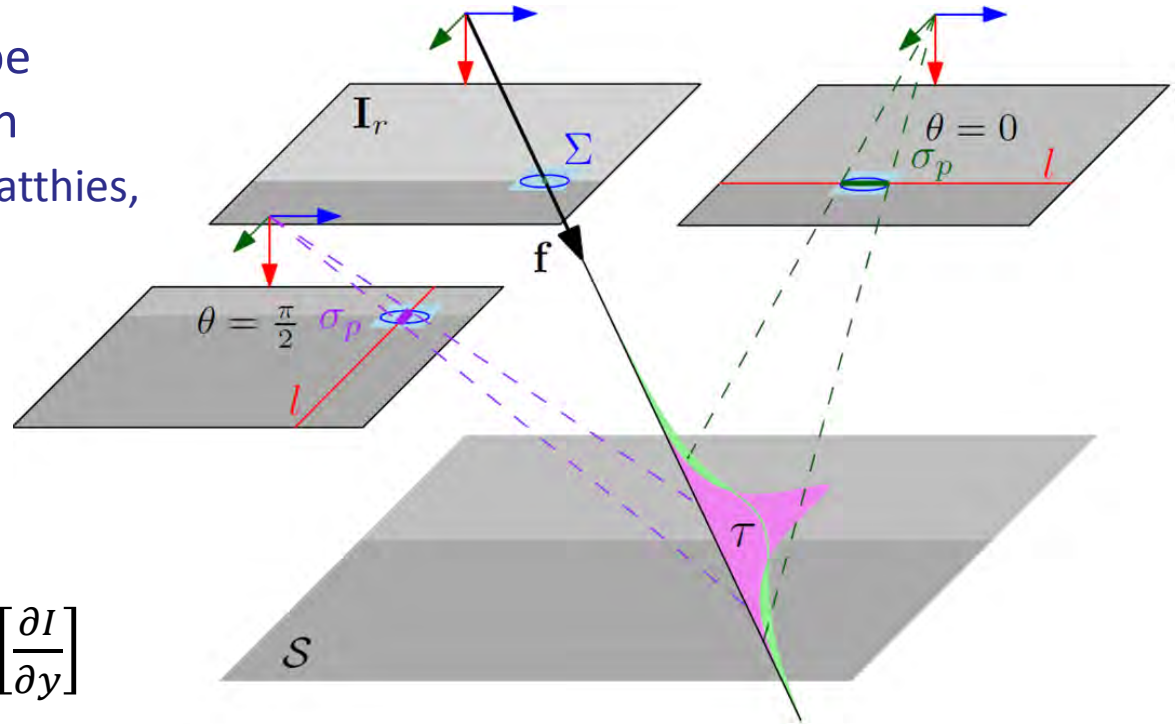
Photometric Disparity Uncertainty

The matching uncertainty can be modeled as a bivariate Gaussian distribution with covariance [Matthies, CVPR'88]

$$\Sigma = 2\sigma_i^2 \begin{bmatrix} {I_x}^2 & I_x I_y \\ I_x I_y & {I_y}^2 \end{bmatrix}^{-1}$$

where σ_i^2 = Image noise,

$$I_x = \sum_P \left[\frac{\partial I}{\partial x}\right], I_x = \sum_P \left[\frac{\partial I}{\partial y}\right]$$



Disparity uncertainty along epipolar line l

$$\sigma_p = f(\boldsymbol{T}_{r,k}, \Sigma)$$

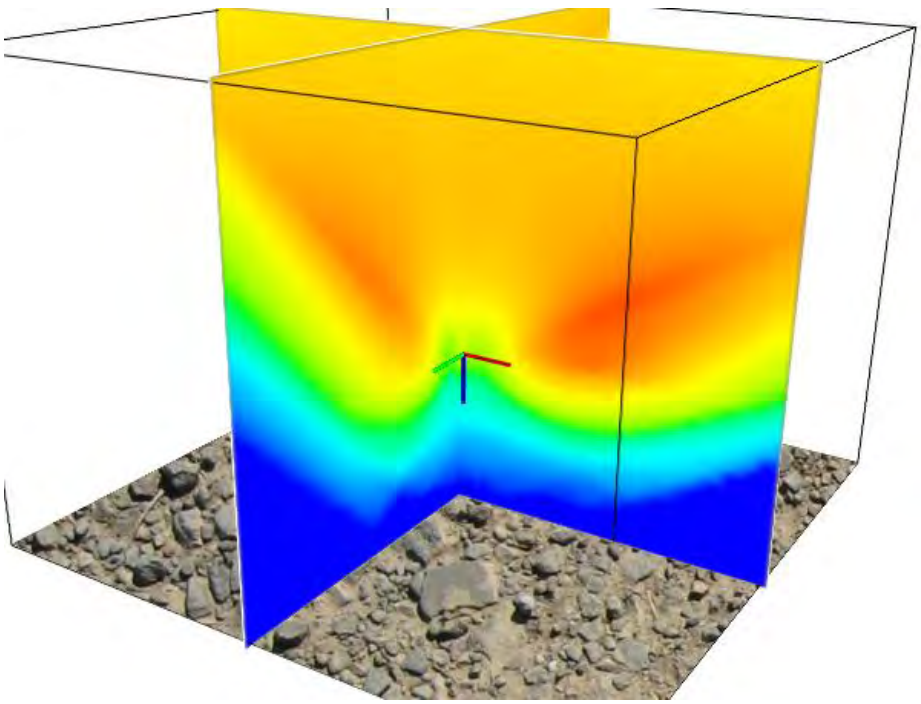
Patch appearance is predicted using reference patch

Forster, Pizzoli, Scaramuzza, Appearance-based Active, Monocular, Dense Reconstruction for Micro Aerial Vehicles, RSS 2014.

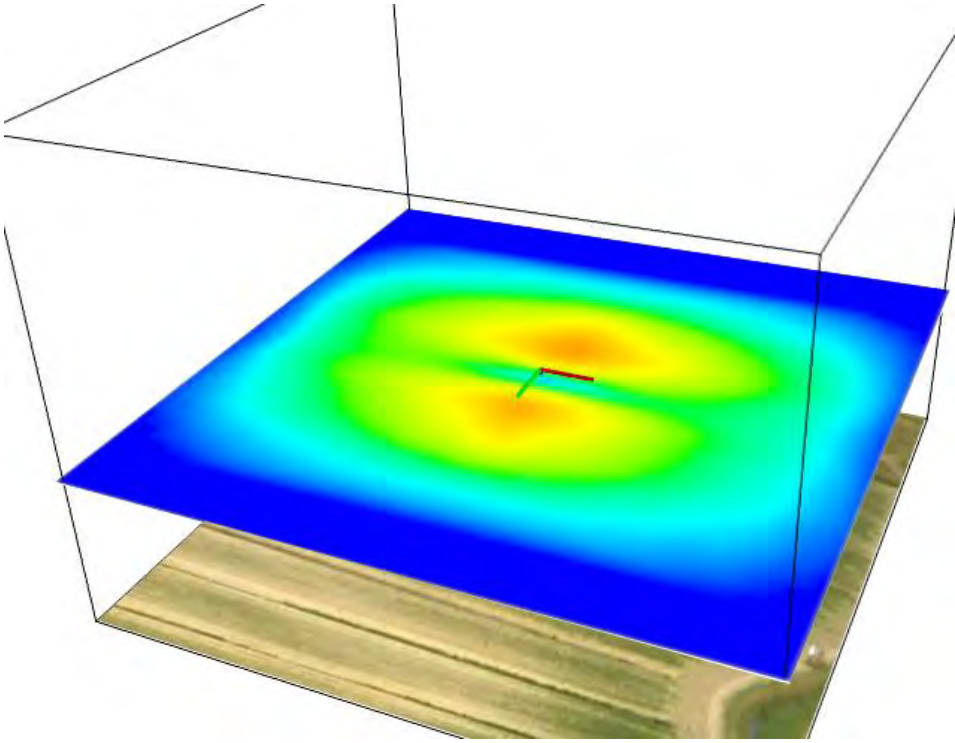
Information gain as a function of the Texture

Information Gain $\mathcal{I}_{k,k+1} = \mathcal{H}_k - \mathcal{H}_{k+1}$ where $\mathcal{H}_{k+1} = \frac{1}{2}\sum \ln(2\pi e\, \sigma^2)$

Isotropic Texture



Dominant Gradient Direction



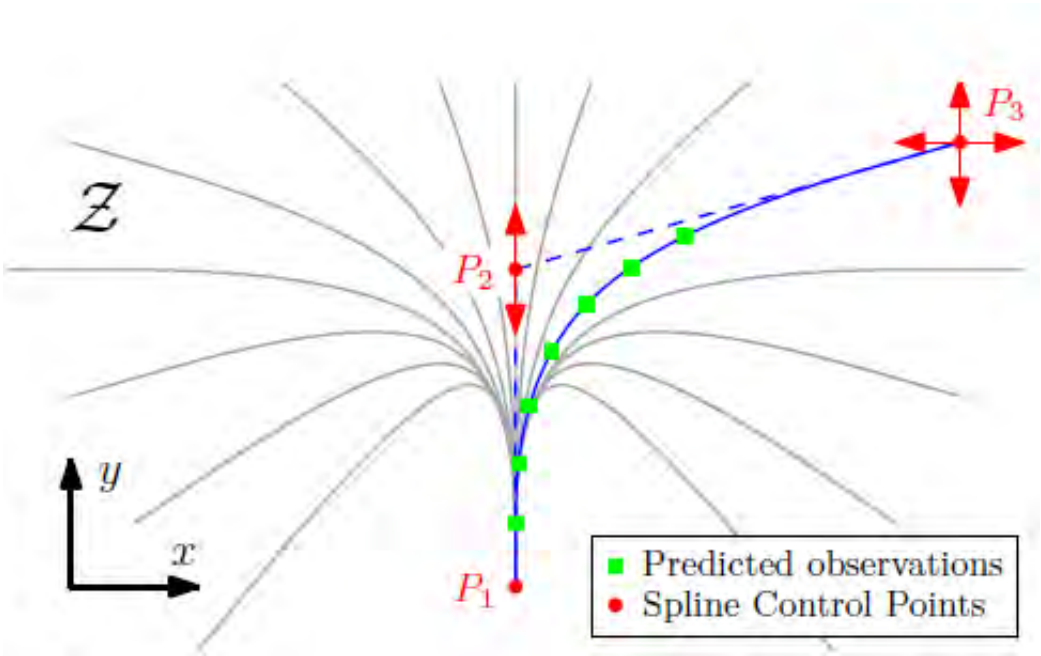
Forster, Pizzoli, Scaramuzza, Appearance-based Active, Monocular, Dense Reconstruction for Micro Aerial Vehicles, RSS 2014.

Receding Horizon Control – Next Best N Views

- The "best" trajectory is selected as the one that maximizes the gain in information (i.e., the map accuracy) over the next robot poses:

$$\boldsymbol{\phi}_k = \arg\max_{\boldsymbol{\phi}} \sum_{i=k}^{k+N} \mathcal{I}_{i,i+1}(\boldsymbol{T})$$

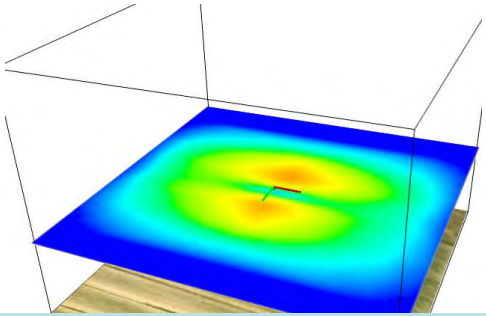
- The "next pose" heavily depends on the texture of the environment



Forster, Pizzoli, Scaramuzza, Appearance-based Active, Monocular, Dense Reconstruction for Micro Aerial Vehicles, RSS 2014.

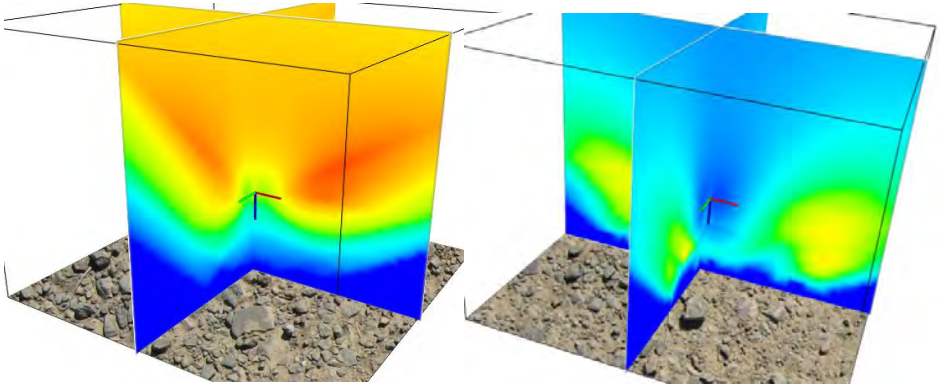
Active Mon. Dense Reconstruction in Real-time [RSS'15]

Information gain for **striped texture**

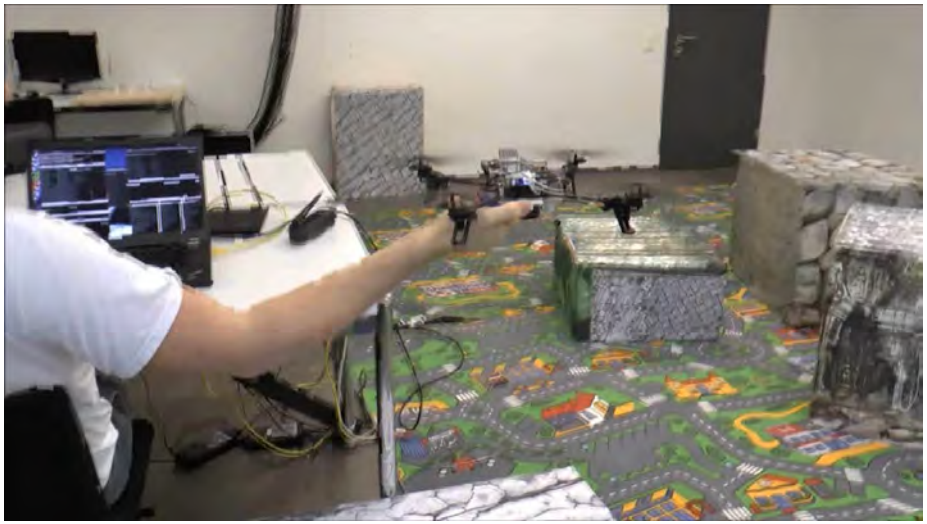


Video: https://www.youtube.com/watch?v=uAc1pL_c-zY

Information gain for **isotropic texture**



After 1 iteration After 10 iterations



Forster, Pizzoli, Scaramuzza, Appearance-based Active, Monocular, Dense Reconstruction for Micro Aerial Vehicles, RSS 2014.

Autonomous Landing-Spot Detection and Landing [ICRA'15]

Video: https://www.youtube.com/watch?v=phaBKFwfcJ4

Forster, Faessler, Fontana, Werlberger, Scaramuzza, Continuous *On-Board Monocular-Vision-based Elevation Mapping Applied to Autonomous Landing of Micro Aerial Vehicles*, ICRA'15.

Having an **autonomous** landing-spot detection can really help!

The Philae lander while approaching the comet on November 12, 2014

Event-based Vision

for High-Speed Robotics

[IROS'13, ICRA'14, RSS'15]

FNSNF
Swiss National Science Foundation
erc
Google
Google
Faculty Research Awards

Open Problems and Challenges with Micro Helicopters

Current flight maneuvers achieved with onboard cameras are **still slow** compared with those attainable with Motion Capture Systems

Penn Mellinger, Kumar

ETH Mueller, D'Andrea

How fast can we go with an onboard camera?

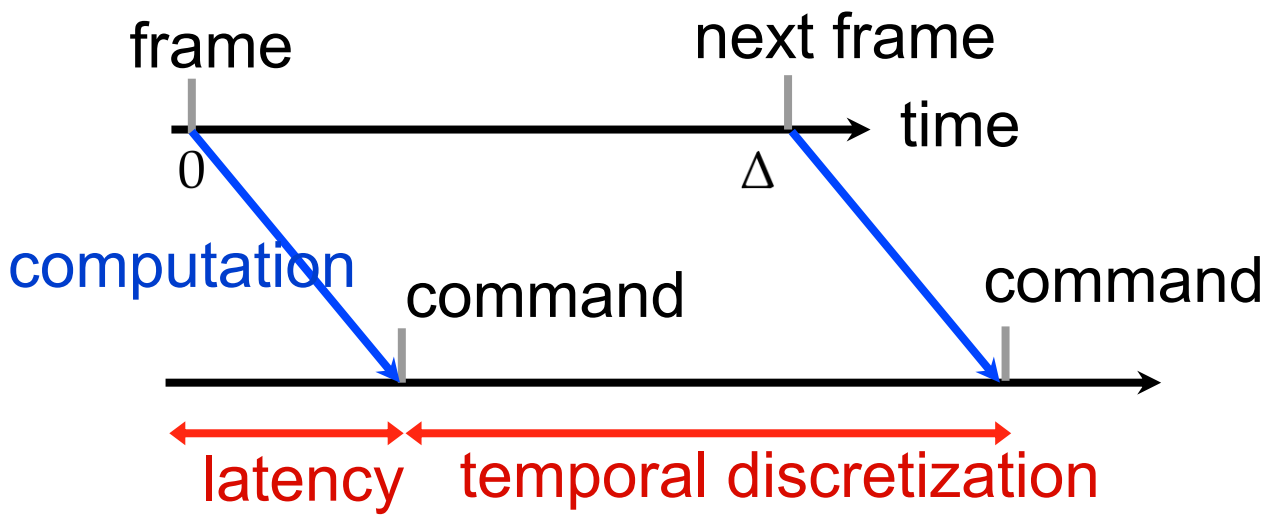
Let's assume that we have perfect perception

Can we achieve the same flight performances
atteinable with motion capture systems or go even faster?

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.



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.

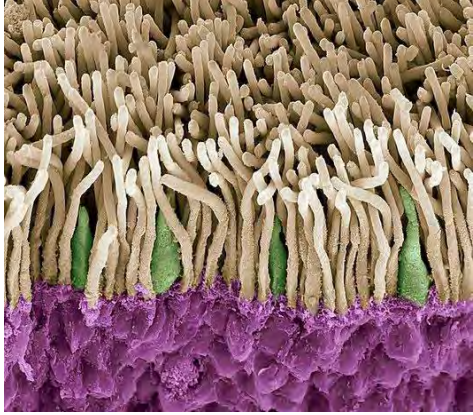
- **Can we create low-latency, low-discretization perception architectures?**

Yes...

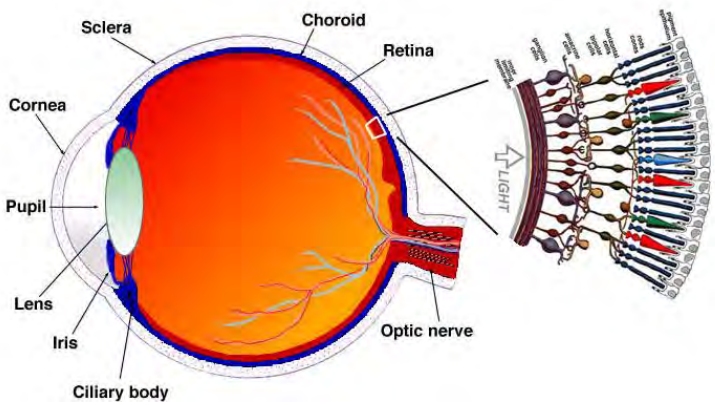
...if we use a camera where pixels do not spike all at the same time

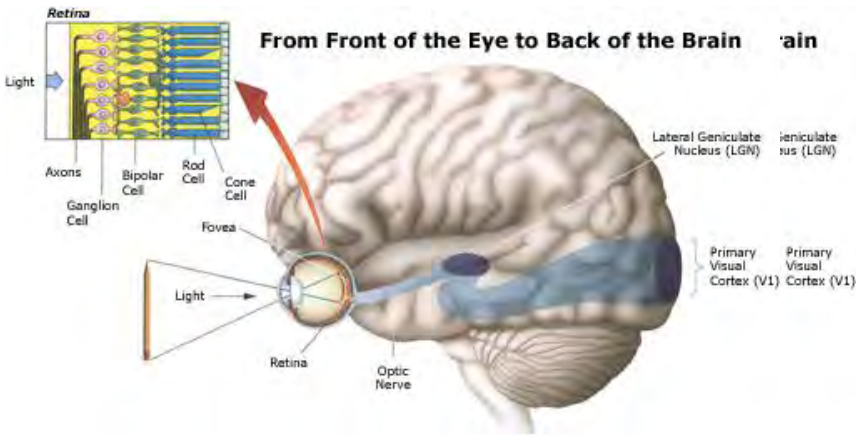
...in a way as we humans do..

Human Vision System

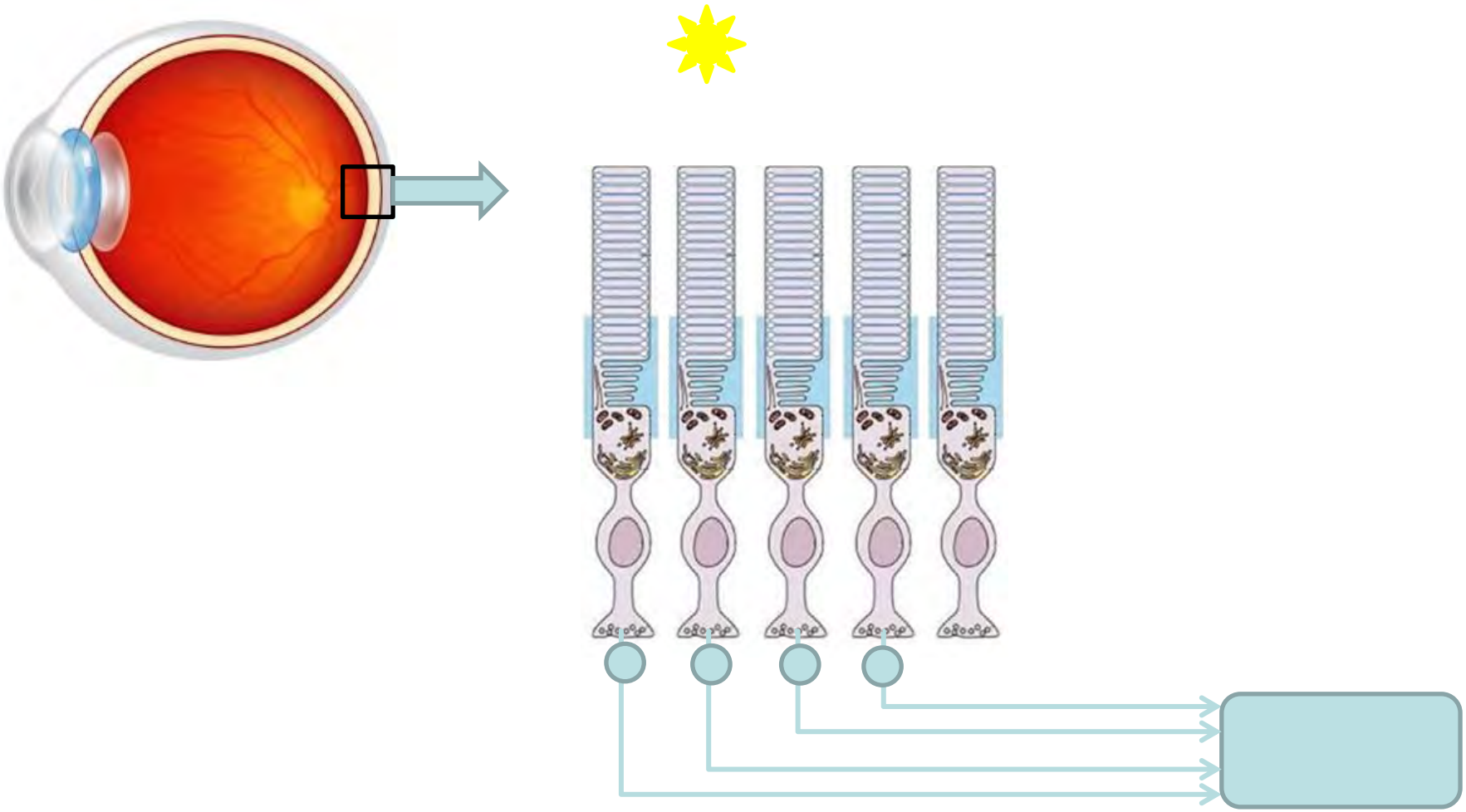


- Retina is ~1000mm^2
- 130 million **photoreceptors**
  - 120 mil. rods and 10 mil. cones for color sampling
  - 1.7 million axons

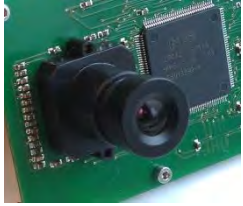




Human Vision System



Dynamic Vision Sensor (DVS)



- **Event-based camera** developed by Tobi Delbruck's group (ETH & UZH).
- Temporal resolution: **1 μs**
- High dynamic range: **120 dB**
- Low power: **20 mW**
- Cost: 2,500 EUR

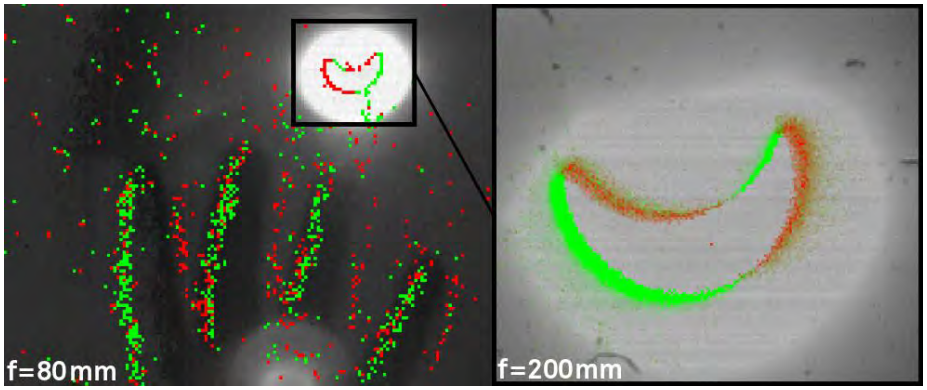
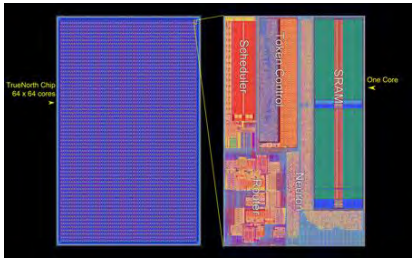


Image of the solar eclipse (March'15) captured by a DVS (courtesy of IniLabs)

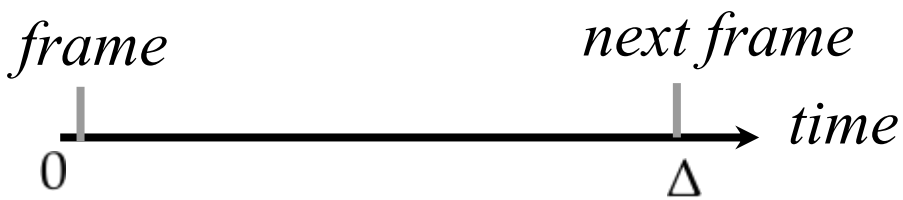


DARPA project Synapse: 1M neuron, brain-inspired processor: IBM TrueNorth

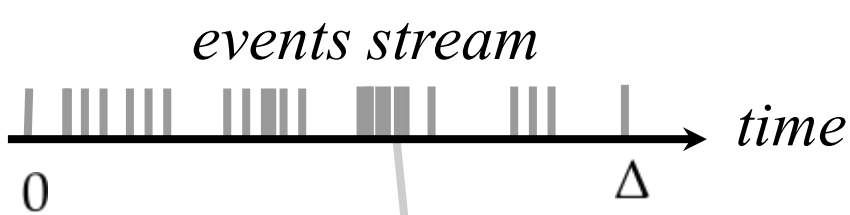
[Lichtsteiner, Posch, Delbruck. A 128x128 120 dB 15μs Latency Asynchronous Temporal Contrast Vision Sensor. 2008]

Camera vs DVS

- A **traditional camera** outputs frames at **fixed time intervals**:



- By contrast, a **DVS** outputs **asynchronous events** at ***microsecond* resolution**. An event is generated each time a single pixel changes value

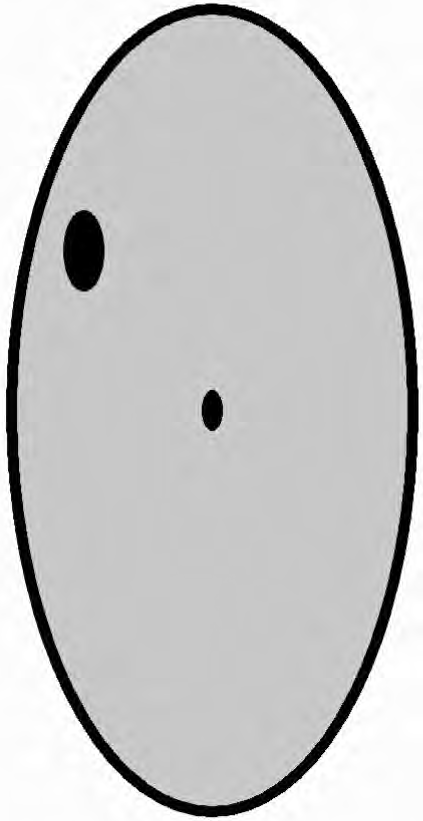


$$event: \left\langle t, \langle x, y \rangle, sign\left(\frac{d}{dt}\log(I_t(x,y))\right)\right\rangle$$

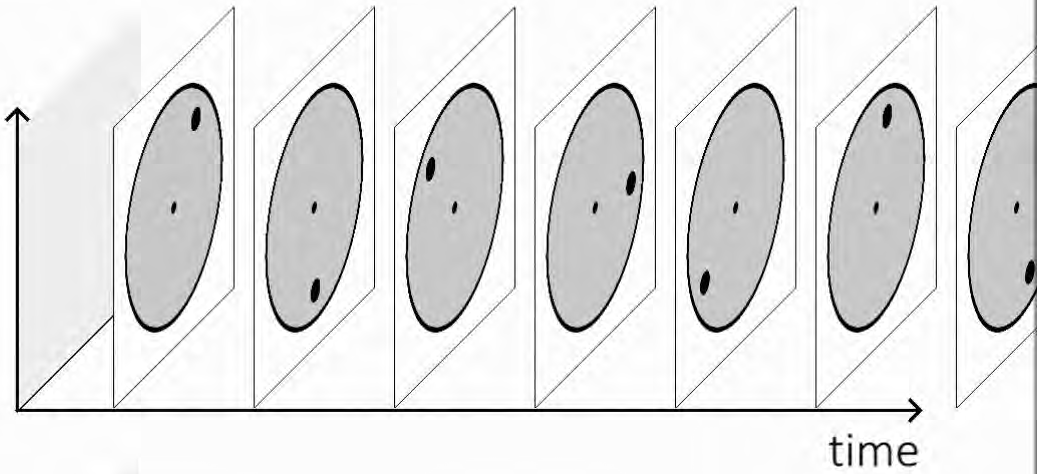
sign (+1 or -1)

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

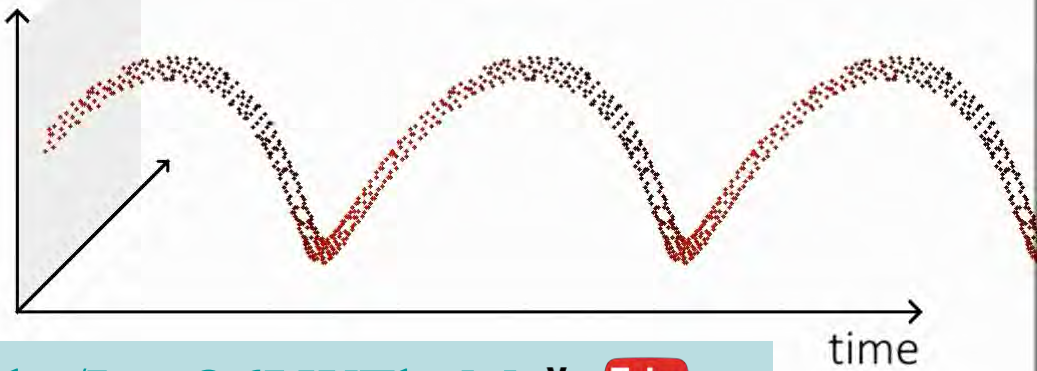
Camera vs Dynamic Vision Sensor



standard camera output:

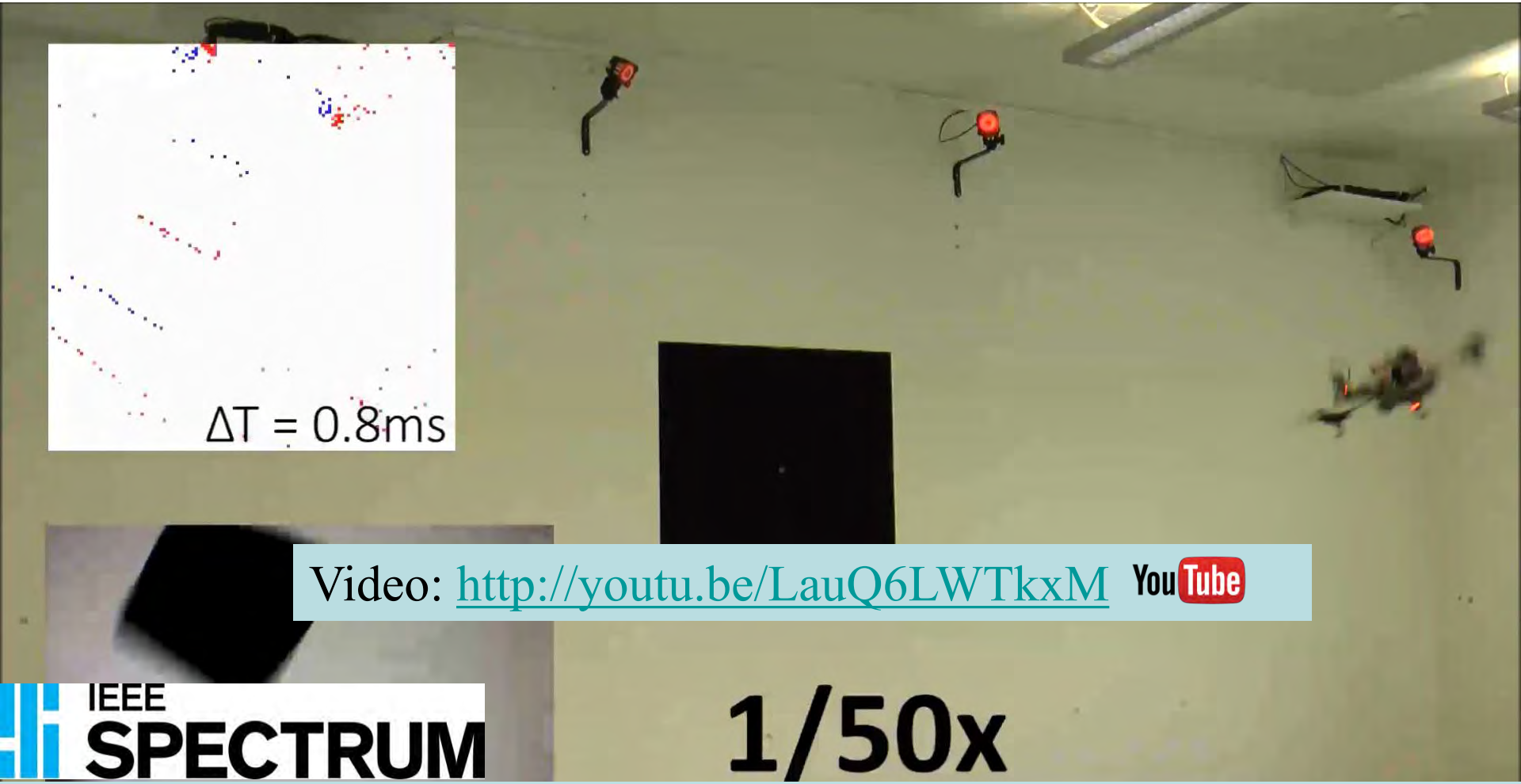


DVS output:



Video: http://youtu.be/LauQ6LWTkxM

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

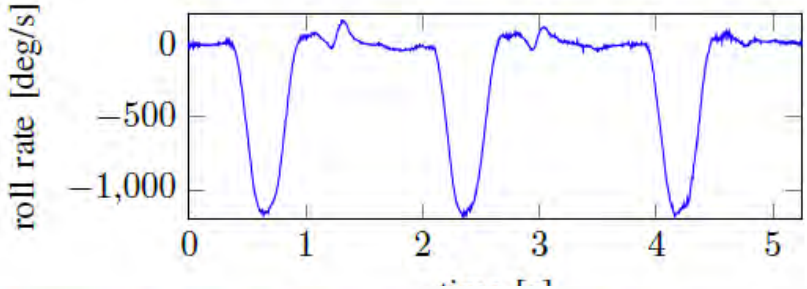


Video: http://youtu.be/LauQ6LWTkxM

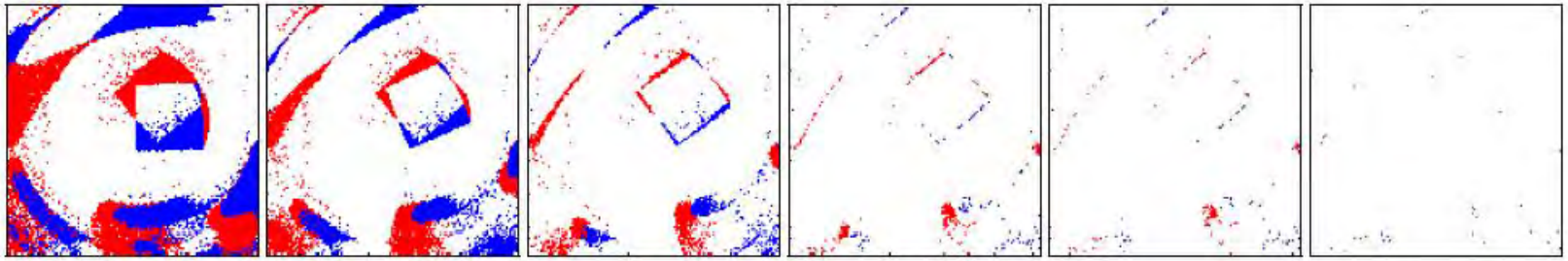
Article: http://spectrum.ieee.org/automaton/robotics/robotics-hardware/dynamic-vision-sensors-enable-high-speed-maneuvers-with-robots

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

Camera and DVS renderings



Peak Angular Speed:
1,200 deg/s



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

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

Frame-based vs Event-based Vision

- Naive solution: **accumulate events** occurred over a certaint time interval and adapt *standard* vision algorithms.
  - Drawback: it **increases latency**

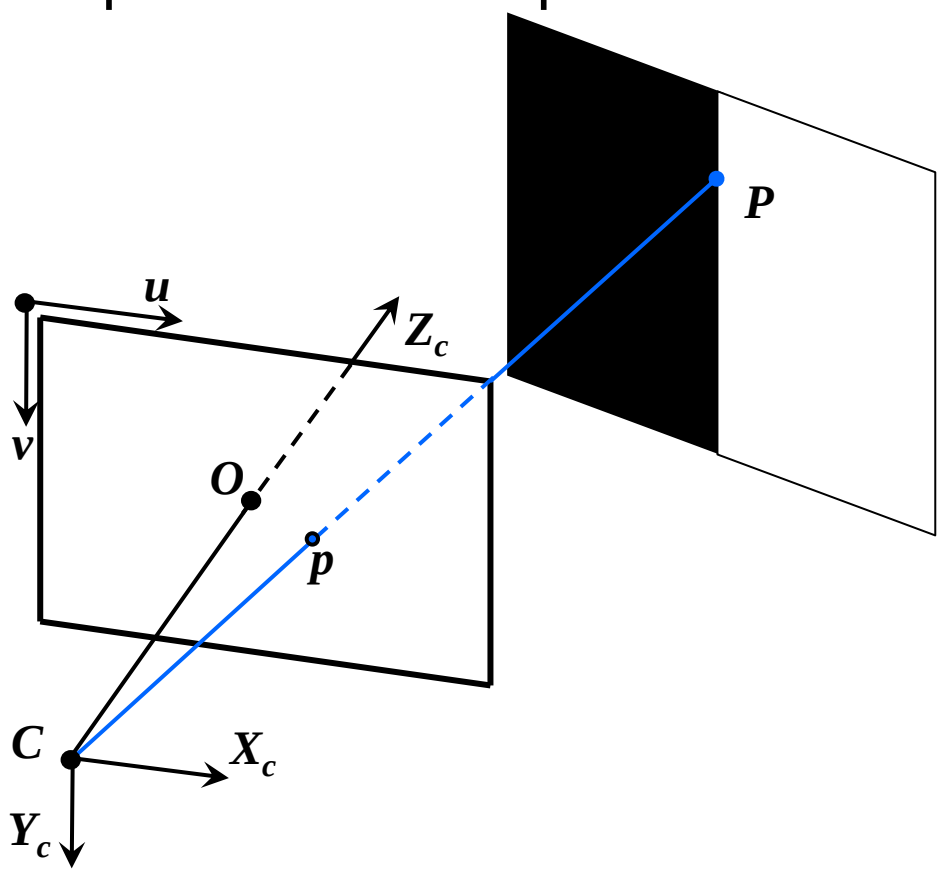
- Instead, we want **each single event** to be used **as it comes!**

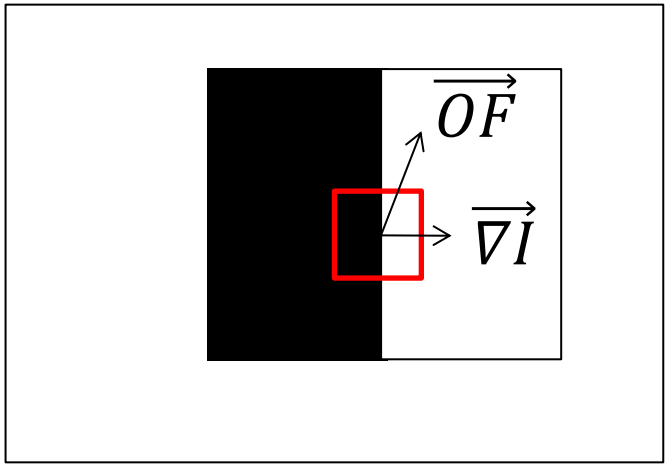
- Problems
  - DVS output is a sequence of **asynchrnous events** rather than a standard image
  - Thus, a **new paradigm shift** is needed to deal with its data

Probabilistic Measurement Model of a DVS

$$p(e_{t,u,v}) \propto \left|\frac{dI(u,v)}{dt}\right| = \left|\overrightarrow{\nabla I} \cdot \overrightarrow{OF}\right|$$

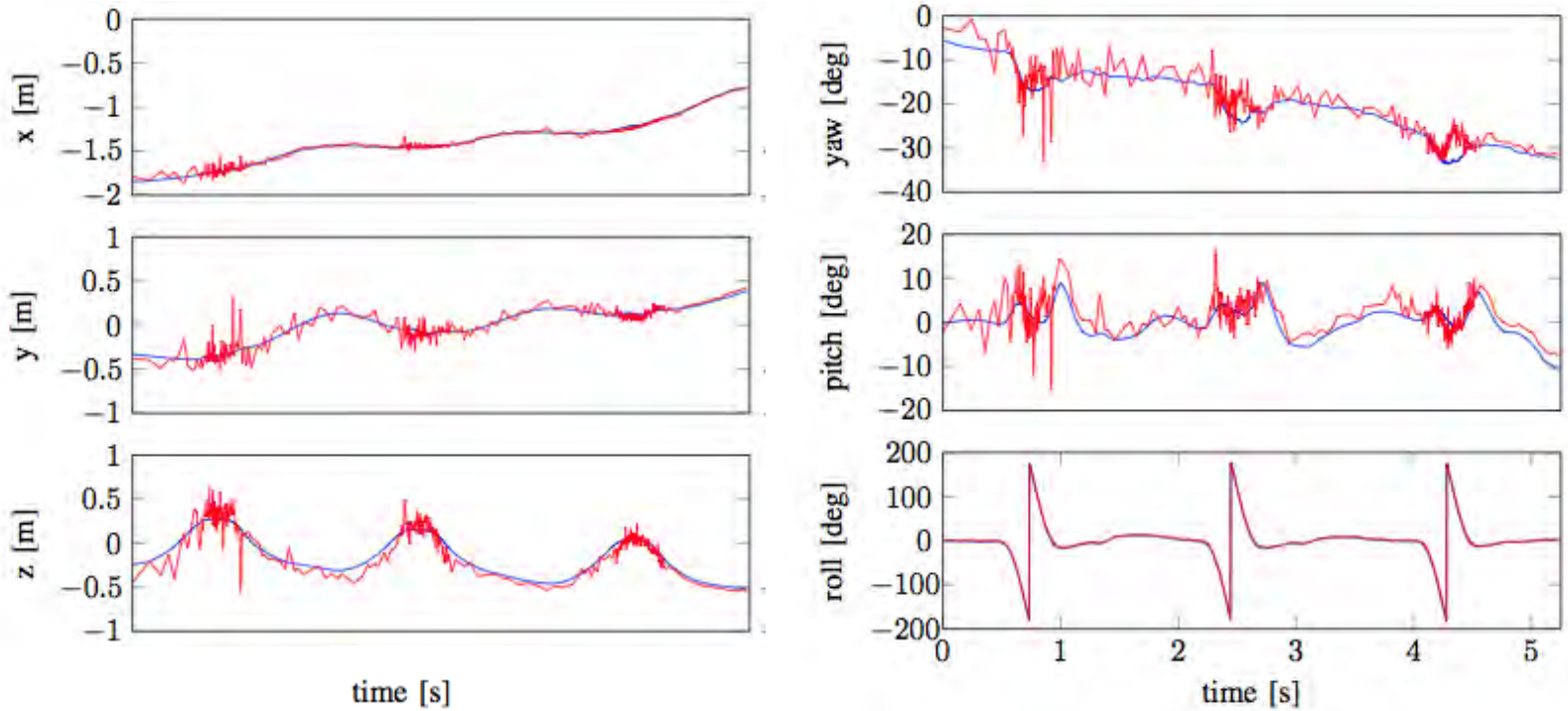
Example: consider a planar scene with a black to white transition





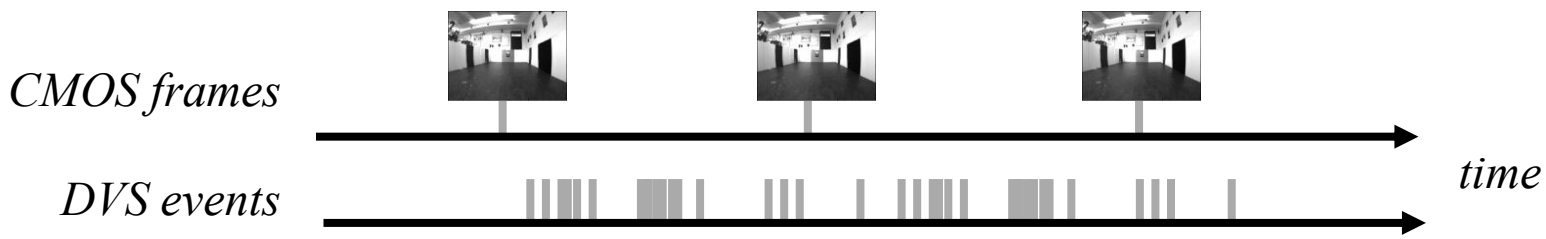
[Censi & Scaramuzza, «Low Latency, Event-based Visual Odometry», ICRA'14]]

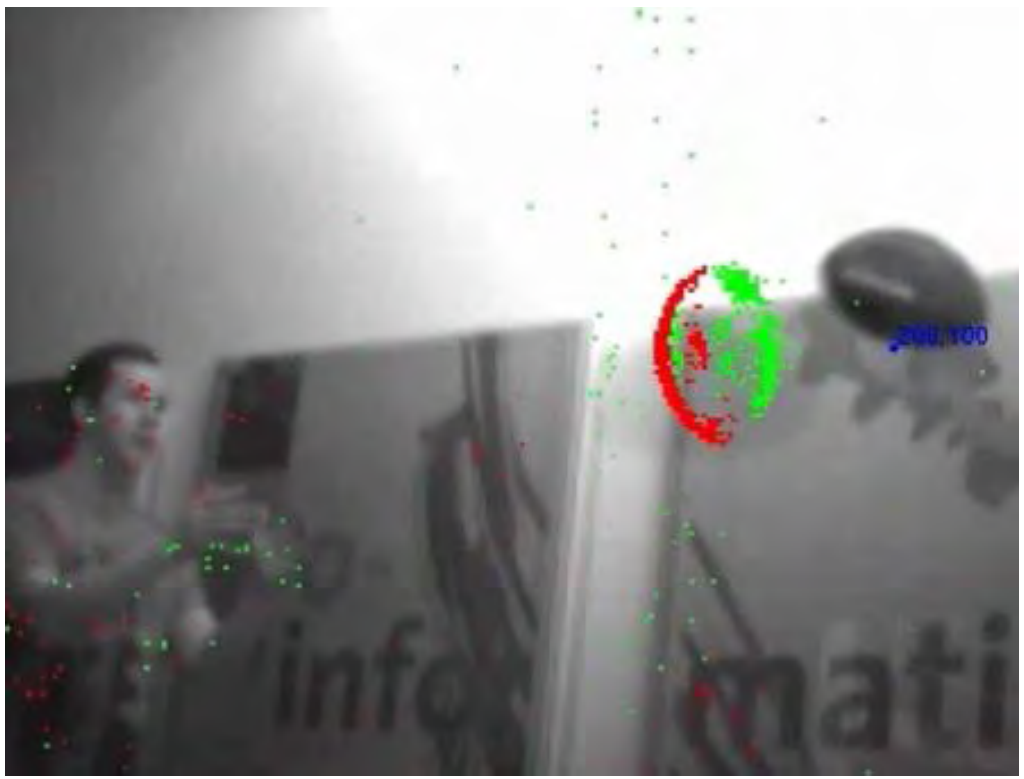
6DoF Pose-Estimation Results at **1MHz** [IROS'14, RSS'15]



[Mueggler, Gallego, Scaramuzza, Continuous-Time Trajectory Estimation for Event-based Vision Sensors, RSS'15]
[Mueggler, Huber, Scaramuzza, Event-based, 6-DOF Pose Tracking for High-Speed Maneuvers, IROS'14]

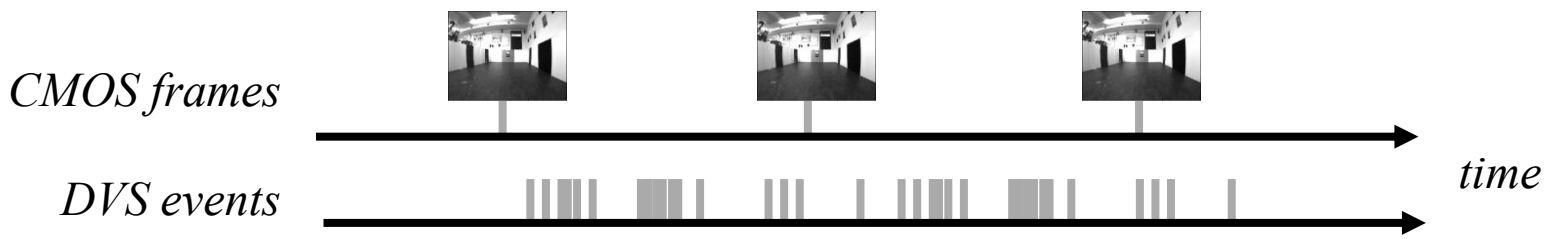
DAVIS: Dynamic and Active-pixel Vision Sensor



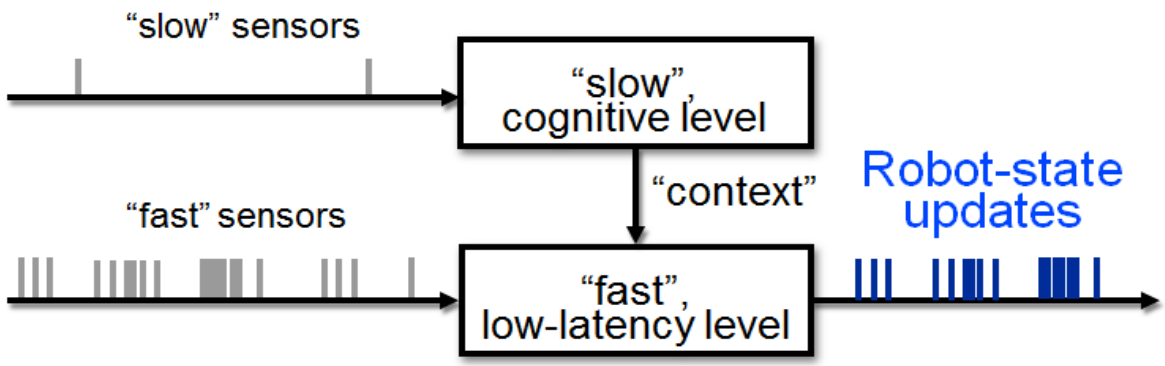


Brandli, Berner, Yang, Liu, Delbruck, "*A 240× 180 130 dB 3 μs Latency Global Shutter Spatiotemporal Vision Sensor.*" IEEE Journal of Solid-State Circuits, 2014.

DAVIS: Dynamic and Active-pixel Vision Sensor

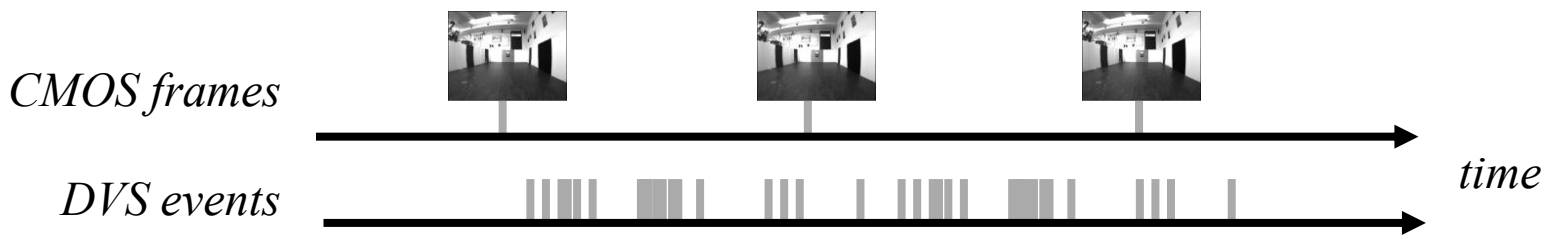


Possible future computer-vision architectures

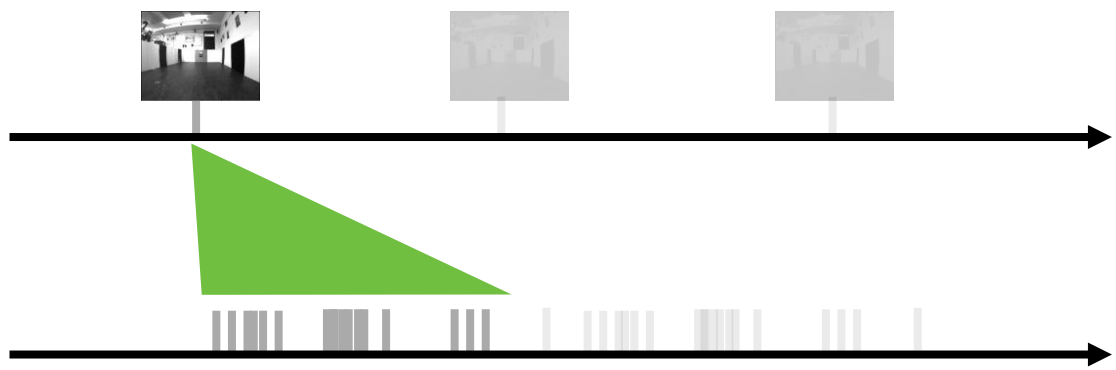


Brandli, Berner, Yang, Liu, Delbruck, "*A 240× 180 130 dB 3 μs Latency Global Shutter Spatiotemporal Vision Sensor.*" IEEE Journal of Solid-State Circuits, 2014.

Inter-frame, Event-based Pose Estimation [ICRA'14]



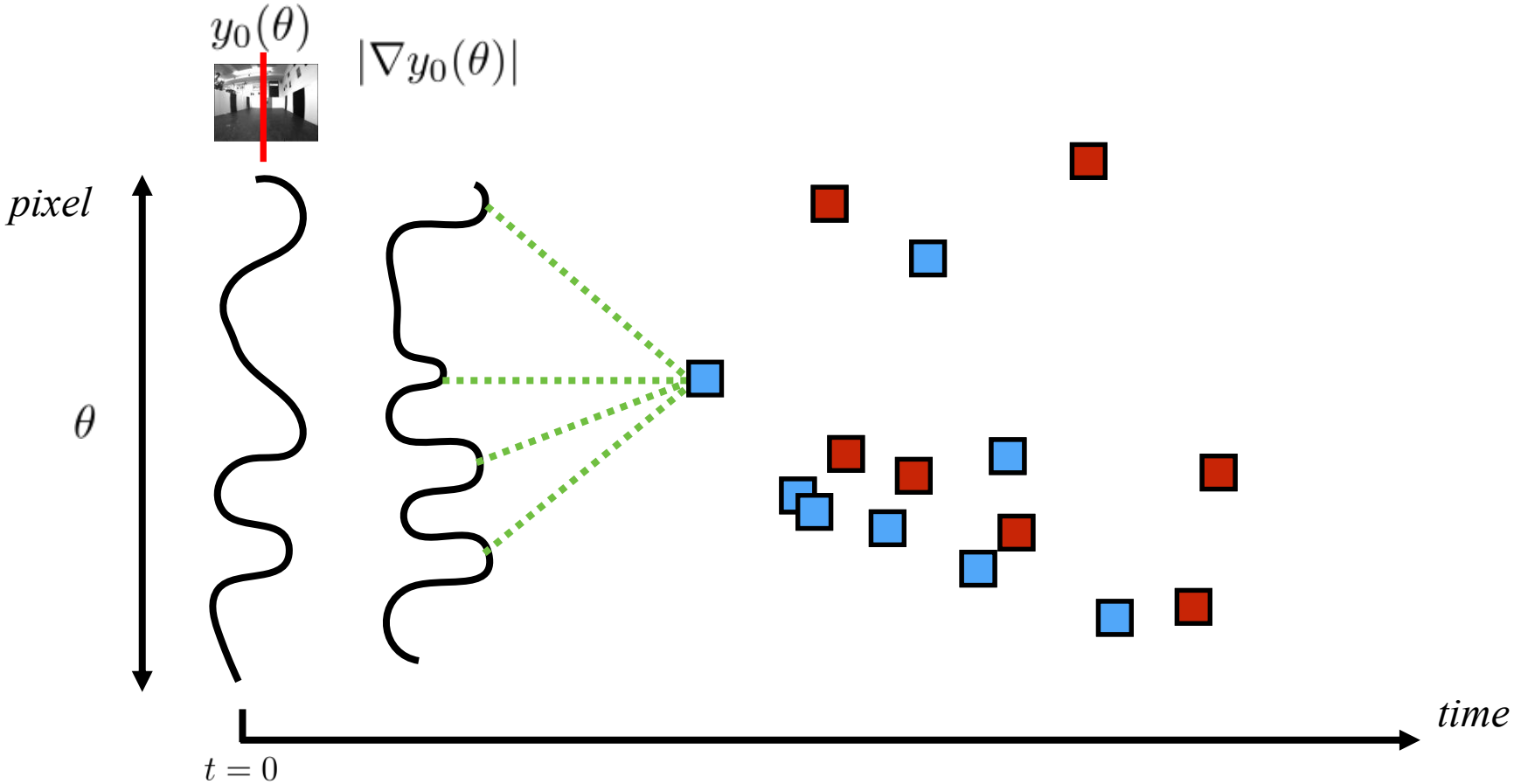
- **Idea**: reduce the problem to "localization" wrt the previous CMOS frame.

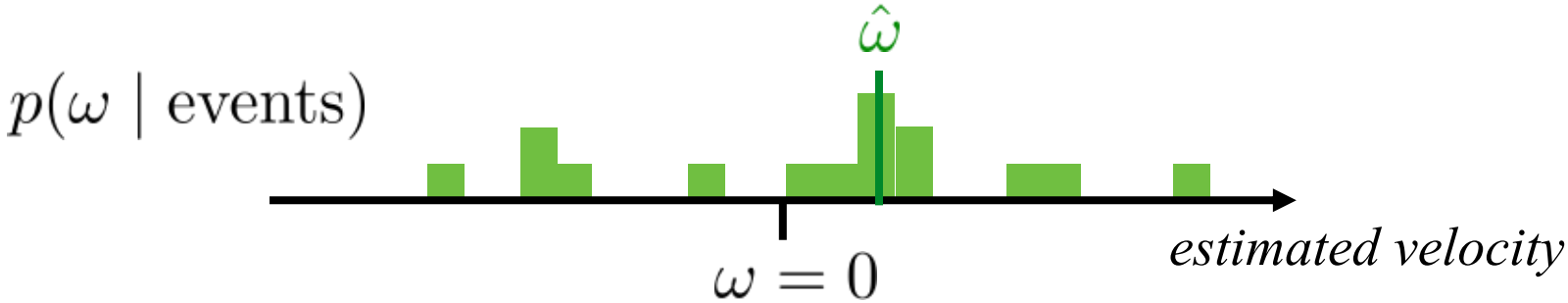


[Gallego, Forster, Mueggler, Scaramuzza, Event-based Camera Pose Tracking using a Generative Event Model, 2015, ArxiV preprint]

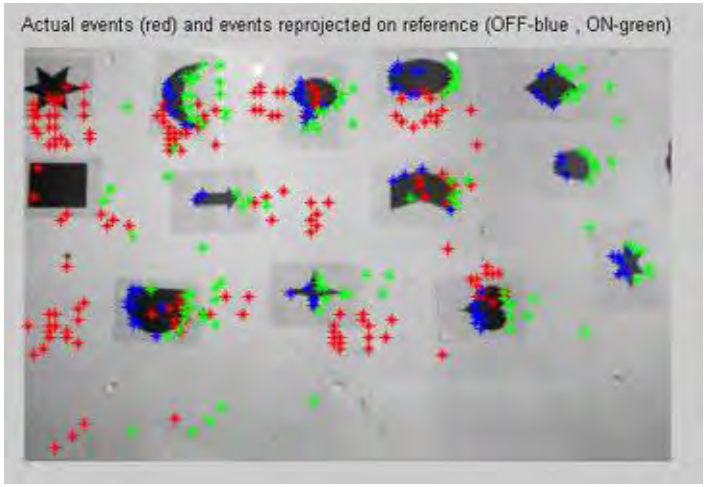
[Censi & Scaramuzza, *Low Latency, Event-based Visual Odometry*, ICRA'14]

Event-based Pose Estimation, 1D Example (pure rotation)

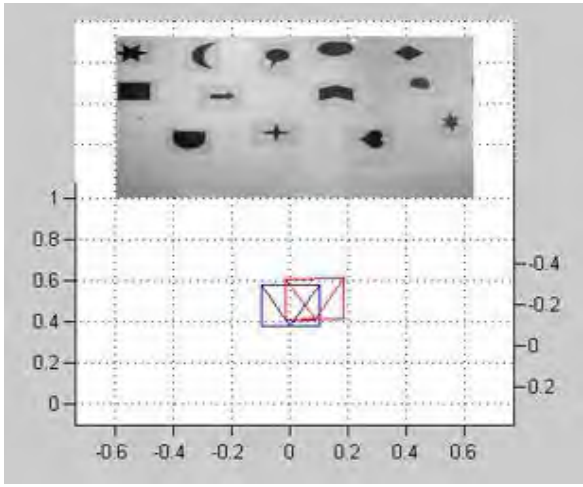




Event-based 6DoF Pose Estimation Results



RED: observed events;
GREEN, BLUE: reprojected events (ON, OFF)



Estimated 6DoF pose

[Gallego, Forster, Mueggler, Scaramuzza, Event-based Camera Pose Tracking using a Generative Event Model, 2015, ArxiV preprint]

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

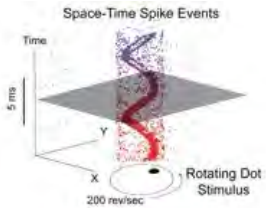
Conclusions and home messages

- Combination of **feature-based and direct methods** yield high frame rate and robustness
- Recursive **Bayesian depth estimation** allows adaptively choosing the baseline of monocular systems
- **Pre-integrated IMU factors** allows high-speed visual inertial fusion
- Modeling depth uncertainty in both sparse and dense methods as a **mixture of uniform and Gaussian** distribution yield
  - Almost no outliers
  - Robustness to dynamic objects
- DVS & DAVIS: **revolutionary sensors** for vision and robotics:
  1. **low-latency** (~1 micro-second)
  2. **high-dynamic range** (120 dB instead 60 dB)
  3. Very **low bandwidth** (only intensity changes are transmitted)

Open Source Software

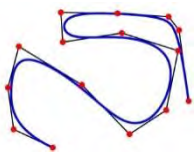
- My lab GitHub repository: **github.com/uzh-rpg**
  - SVO: Semi-direct Visual Odometry
  - REMODE: Regularized, Probabilistic, Monocular Dense Reconstruction
  - DVS: ROS driver and calibration tools for single and stereo event cameras
- BORG lab repository
  - GTSAM (iSAM) with pre-integrated IMU factors

References on Event Vision

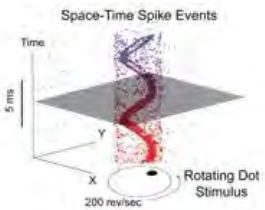


A. Censi, D. Scaramuzza,
Low-Latency Event-Based Visual Odometry
IEEE International Conference on Robotics and Automation (ICRA), Hong Kong, 2014.
PDF

E. Mueggler, B. Huber, D. Scaramuzza
Event-based, 6-DOF Pose Tracking for High-Speed Maneuvers
IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), Chicago, 2014.
PDFYouTube



E. Mueggler, G. Gallego, D. Scaramuzza
Continuous-Time Trajectory Estimation for Event-based Vision Sensors
Robotics: Science and Systems (RSS), Rome, 2015.
PDF



Mueggler, C. Forster, N. Baumli, G. Gallego, D. Scaramuzza
Lifetime Estimation of Events from Dynamic Vision Sensors
IEEE International Conference on Robotics and Automation (ICRA), Seattle, 2015.
PDF

IEEE SPECTRUM

Article: http://spectrum.ieee.org/automaton/robotics/robotics-hardware/dynamic-vision-sensors-enable-high-speed-maneuvers-with-robots

Massachusetts Institute of Technology

Article: http://newsoffice.mit.edu/2014/think-fast-robot-0530

Resources

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

Thanks! Questions?

Funding

FNSNF
Swiss National Science Foundation

erc

robotics
Swiss National Centre of Competence in Research

Google
Google Faculty Research Awards

Schweizerische Eidgenossenschaft
Confédération suisse
Confederazione Svizzera
Confederaziun svizra

Kommission für Technologie und Innovation KTI

ROBOTICS & PERCEPTION GROUP

rpg.ifi.uzh.ch