Event-based Algorithms for Robust and High-speed Robotics

Davide Scaramuzza

All my research on event-based vision is summarized on this page:
http://rpg.ifi.uzh.ch/research_dvs.html

My Dream Robot

Agile, lightweight drones rapidly navigating to accomplish a given task

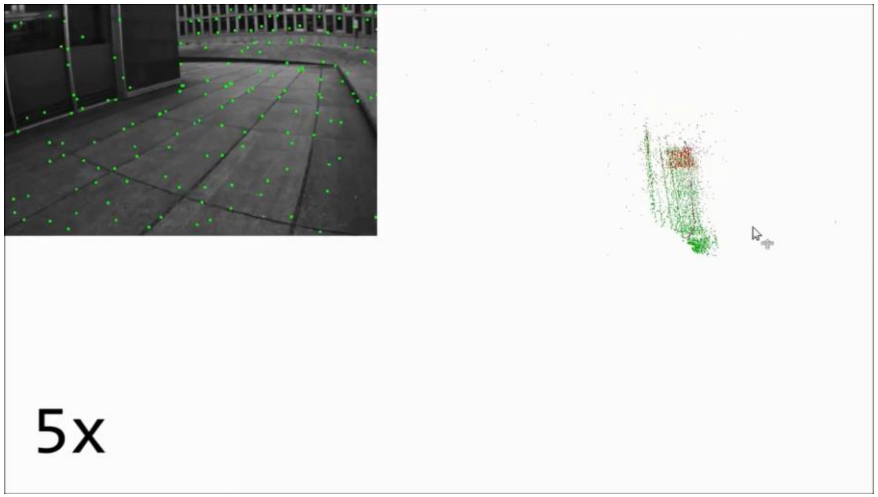
- **Challenges**: fast, lightweight, agile (*low-latency perception & control*)

Video credit: LEXUS commercial, 2013

Our Research Areas

Visual-Inertial State Estimation (SVO)

[IJCV'11, PAMI'13, RSS'15, TRO'16-17]



Vision-based Navigation of Flying Robots

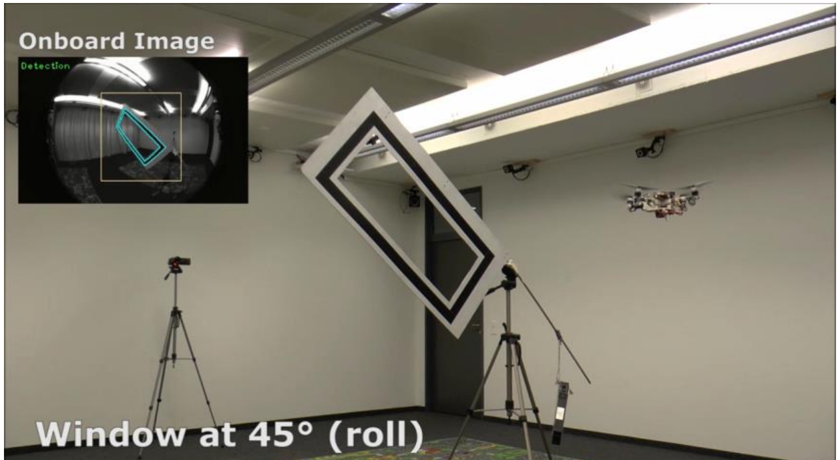
[ICRA'10, AURO'12, RAM'14, JFR'15]

Deep Learning for End-to-End Navigation

[RAL'16-17]

Low-latency vision for Aggressive Flight

[IROS'3, ICRA'14, RSS'15, BMVC'16, RAL'17]



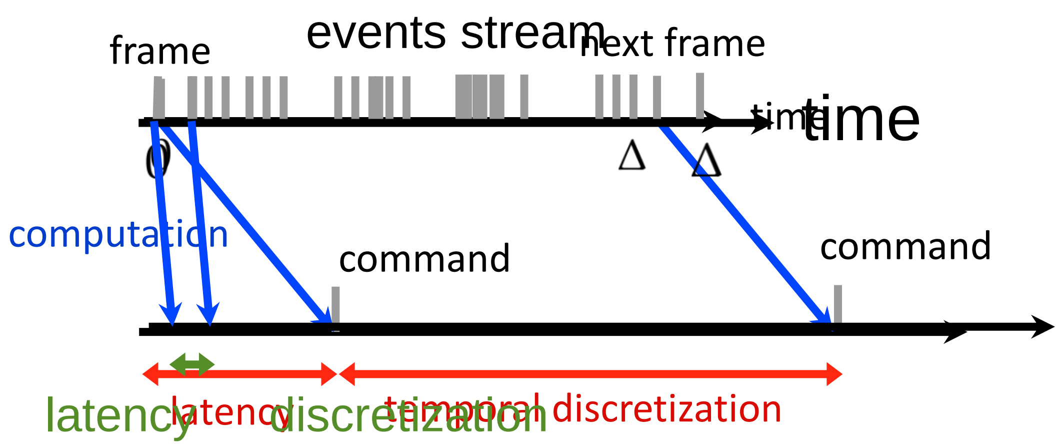
The Challenge of Vision Controlled Drones

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

A sparrowhawk catching a garden bird (National Geographic)

To go faster, we need faster sensors! [IROS'13]

- The agility of a robot is limited by the latency and temporal discretization of its sensing pipeline.



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

- Event cameras enable **low-latency sensory motor control (<< 1ms)**

A. Censi, J. Strubel, C. Brandli, T. Delbruck, D. Scaramuzza,
Low-latency localization by Active LED Markers tracking using a Dynamic Vision Sensor, IROS'13

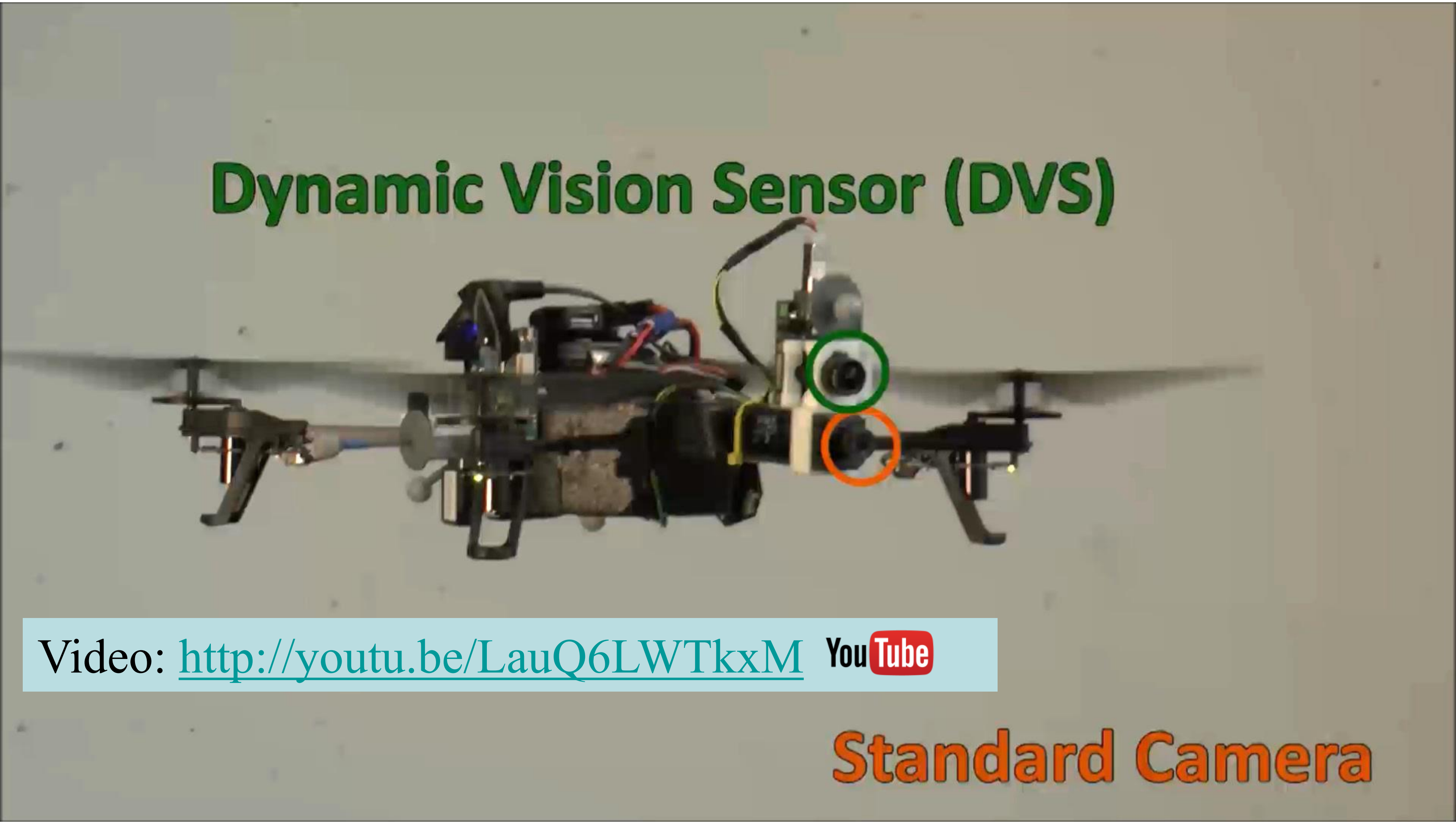
Event-based, 6-DOF Pose Tracking from Line-based Maps

Mueggler, Huber, Scaramuzza,

IROS'14

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

Quadrotor Flip (1,200 deg/s) [IROS'14, RSS'15]



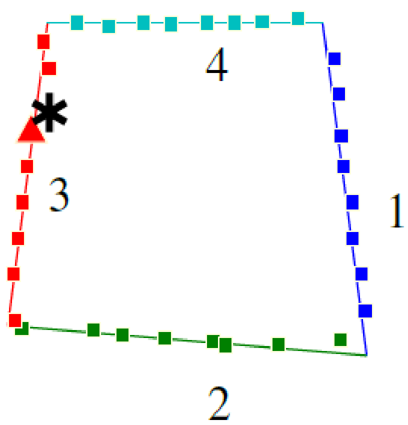
Video: http://youtu.be/LauQ6LWTkxM YouTube

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

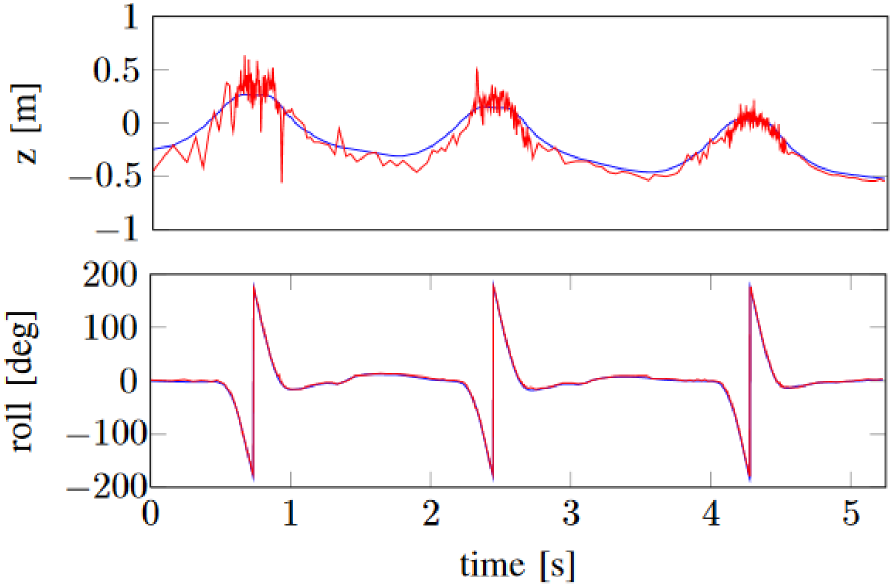
Event-based 6-DOF pose Tracking

- Optimization-based (minimizes reprojection error)
- Assumption: line-based maps

$$P^* = \arg\min_P \sum_{l=1}^{4} \sum_{i=1}^{N} \| d\left(\pi(L_l, P), e_{l,i}\right) \|^2$$



Estimated motion



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

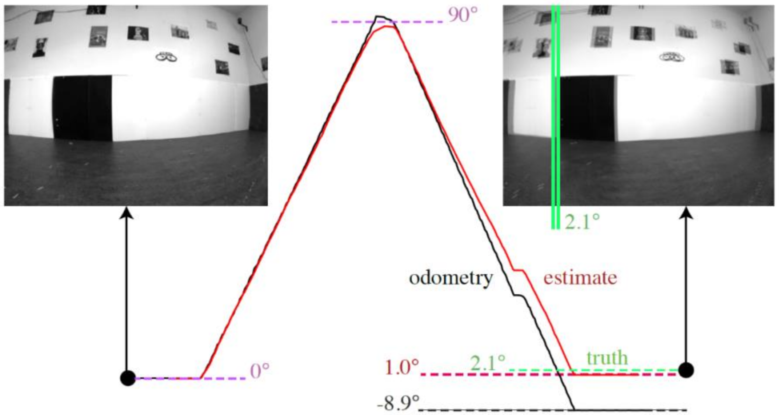
Event-based, Pose Tracking from High-Contrast Scenes

Censi, Scaramuzza

ICRA'14

Censi, Scaramuzza,
Low-Latency Event-Based Visual Odometry, ICRA'14

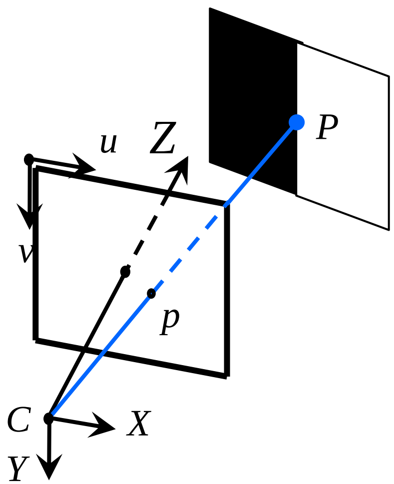
Pose Estimation from from High Contrast Scenes [ICRA'14]

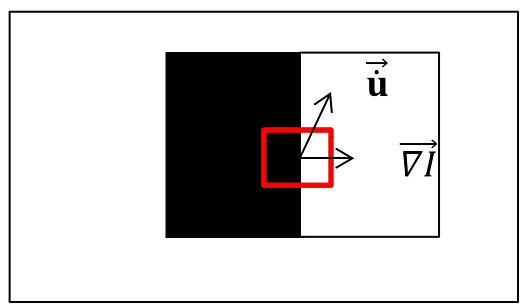


3 DOF Tracking (planar motion)

- Planar motion & known map
- Recursive estimation
- Measurement model:

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





Censi, Scaramuzza,
Low-Latency Event-Based Visual Odometry, ICRA'14

Event-based, 6-DOF Pose Tracking from Photometric Depth Maps

Gallego, Lund, Mueggler, Rebecq,
Delbruck, Scaramuzza

ArXiv, 2016
submitted to PAMI

[Gallego et al., Event-based, 6-DOF Camera Tracking for High-Speed Applications, Arxiv'16]

Problem statement

- Given a **photometric depth map**, track the 6-DOF **pose** of the DVS **event by event**

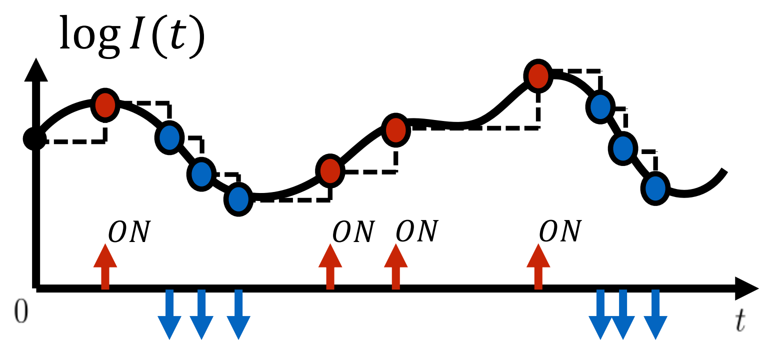
- How to get a photometric depth map?
  - Dense reconstruction based on standard cameras (DTAM, REMODE)
  - RGB-D cameras
  - Reference frames plus depth

Methodology

- **Probabilistic** approach (Bayesian filter): $p(s|e) = p(e|s)p(s)$

  Posterior Likelihood Prior

- **State vector**: $s = (R, T, C, \sigma_C, \rho)$
  - pose (R,T),
  - contrast mean value C
  - uncertainty σ_C,
  - inlier ratio ρ



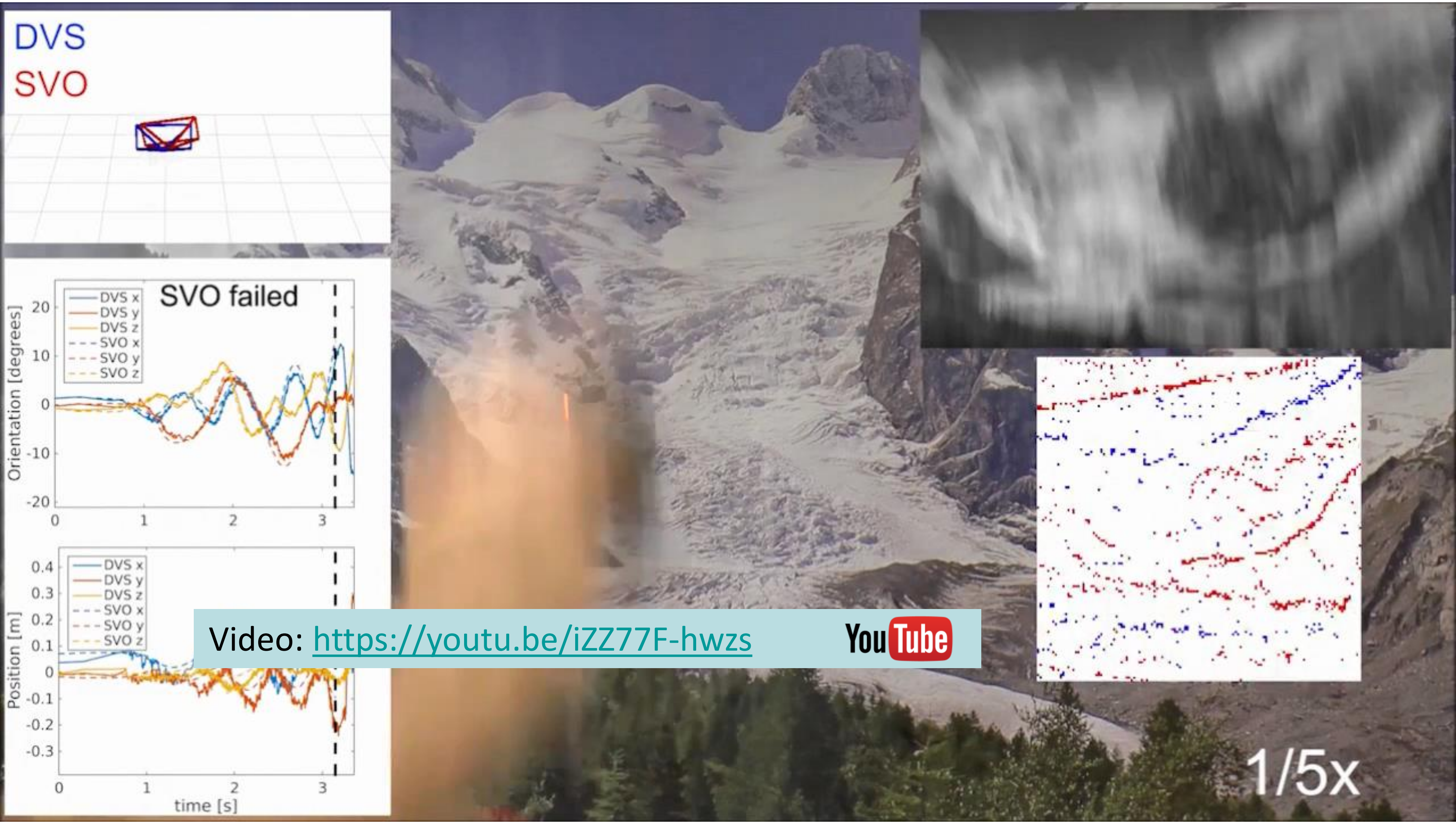
- Motion model: **random walk**
- **Robust sensor model** (likelihood)
  - **Measurement function** derived from generative event model:

$$\log I(t) - \log I(t - \Delta t) = C \quad \Rightarrow \quad M(e|s) = \frac{\Delta\, log I}{C} - 1$$

  - **Mixture model**: heavy-tail Gaussian distribution (i.e., Gaussian + Uniform):

$$p(e|s) = \rho N(M(e|s), 0, \sigma) + (1 - \rho) U(M_{min}, M_{max})$$

Results: high-speed motion



Video: https://youtu.be/iZZ77F-hwzs

[Gallego et al., Event-based, 6-DOF Camera Tracking for High-Speed Applications, Arxiv'16]

EVO:
A Geometric Approach to Event-based 6-DOF Parallel Tracking and Mapping in Real-time

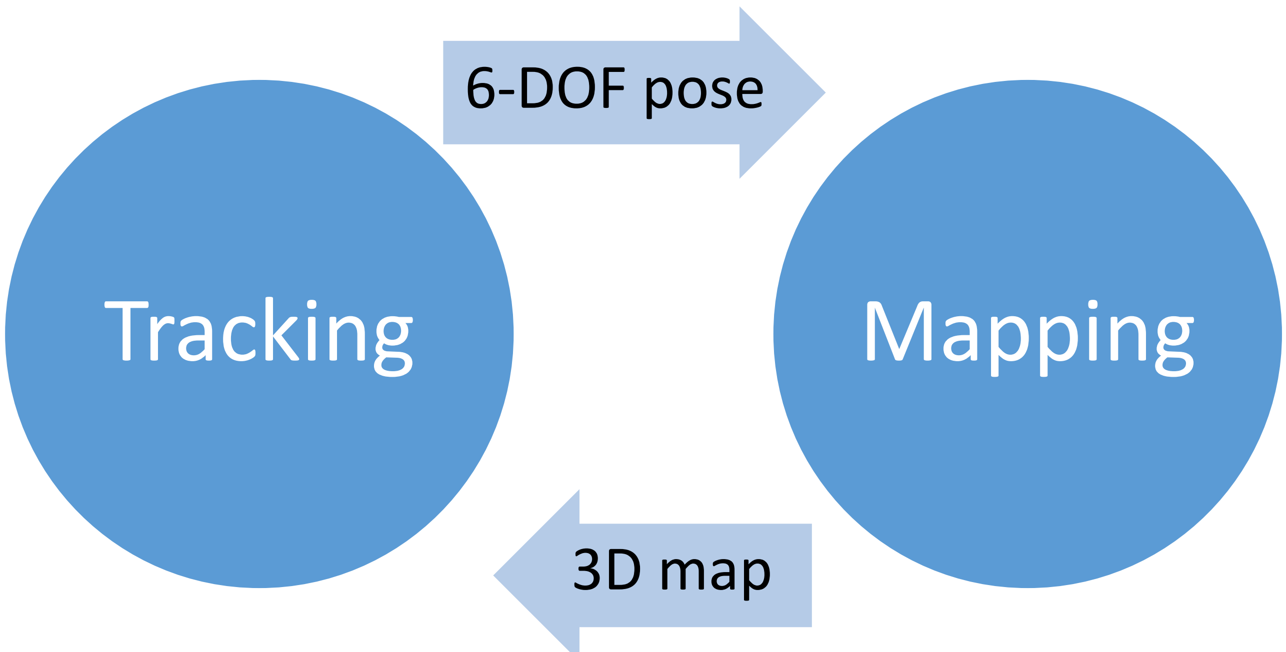
Rebecq, Horstschäfer, Gallego, Scaramuzza

IEEE Robotics & Automation Letters, 01/2017 (presented at ICRA'17)

EU Patent 2017

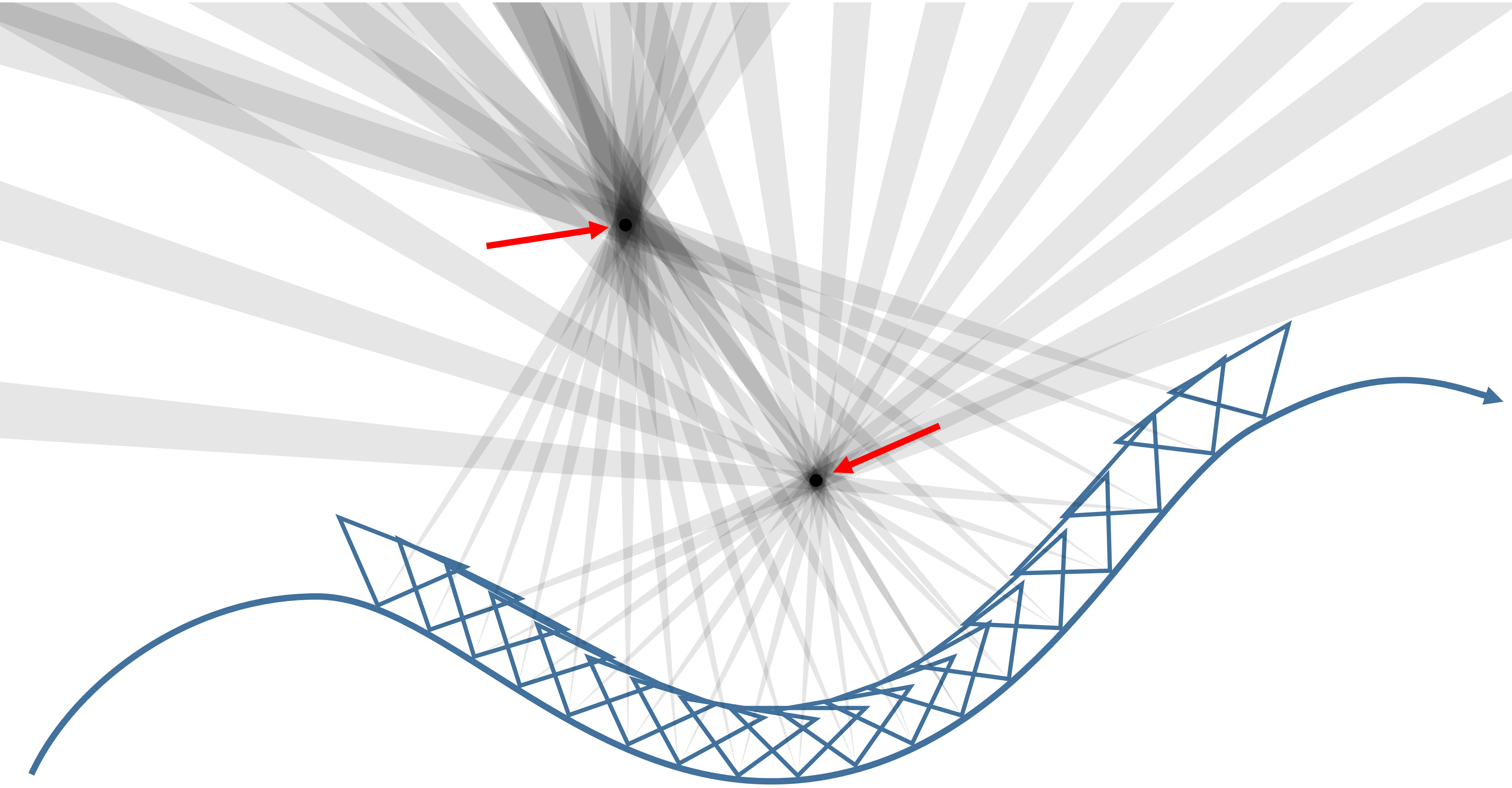
Rebecq, Horstschaefer, Gallego, Scaramuzza, "EVO: Geometric Approach to 6-DoF Event-based Tracking and Mapping in Real Time", IEE Robotics and Automation Letters, 2017. **Also EU Patent**

Parallel Tracking and Mapping



Rebecq, Horstschaefer, Gallego, Scaramuzza, “EVO: Geometric Approach to 6-DoF Event-based Tracking and Mapping in Real Time”, IEE Robotics and Automation Letters, 2017. **Also EU Patent**

How the 3D mapping works

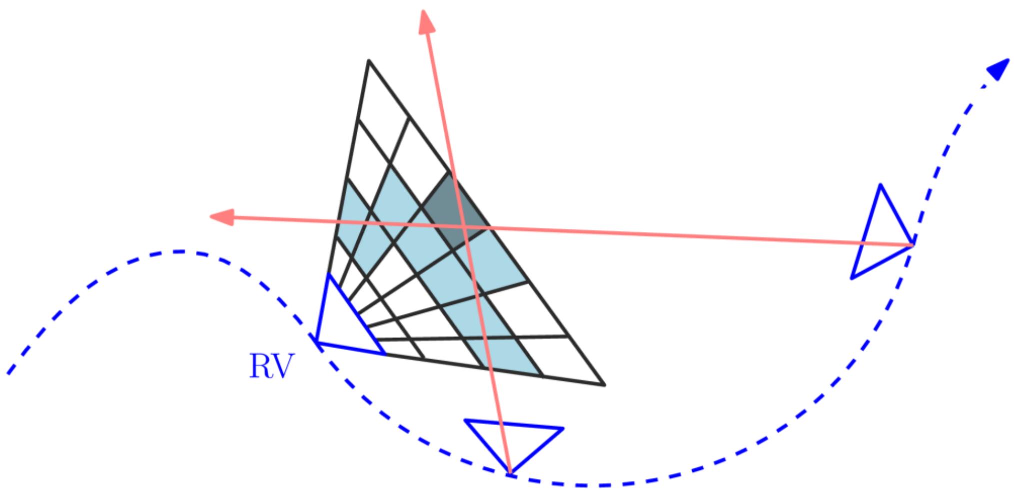


An event camera reacts to strong gradients in the scene

Areas of high ray-density likely indicate the presence of 3D structures

How the 3D mapping works

- Ray-density: Disparity Space Image (DSI)
- Projective sampling grid (DSI) + adaptive thresholding

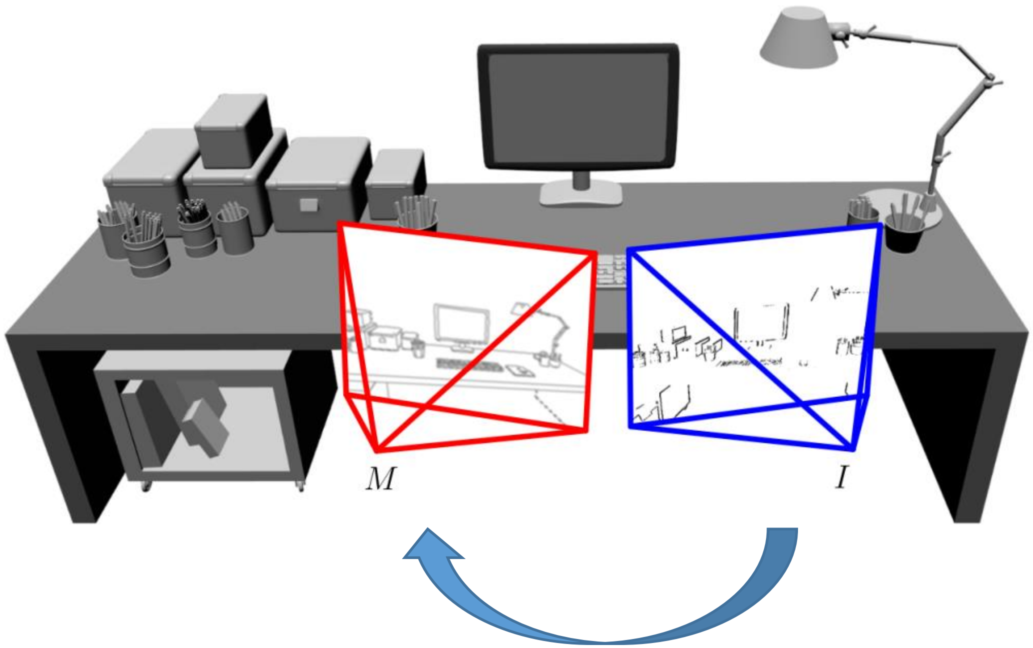


Non-uniform, projective grid, centered on a reference viewpoint

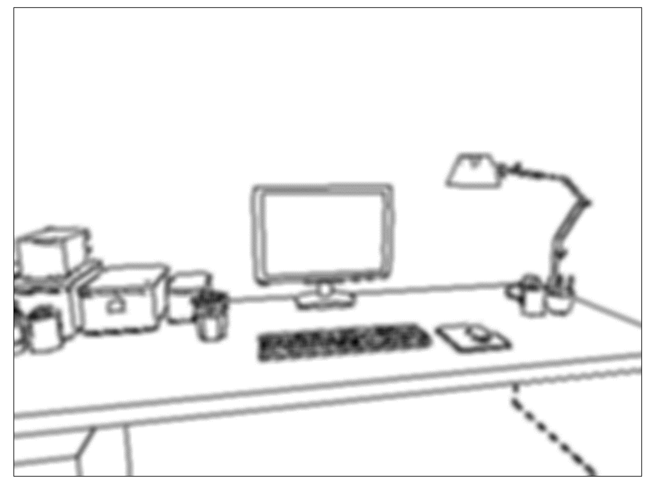
240 x 180 x 100 voxels

“EMVS: Event-based Multi-View Stereo”, Rebecq, Gallego, Scaramuzza, **BMVC’16, Best Industry Paper Award**

How the tracking works



I: Event image
(accumulating events)



M: = Projected map

Global image alignment through 6 DOF warp $\boldsymbol{W}$

$$\boldsymbol{W}(\boldsymbol{u}; T) \coloneqq \pi(T \cdot \pi^{-1}(\boldsymbol{u}, d_{\boldsymbol{u}}))$$

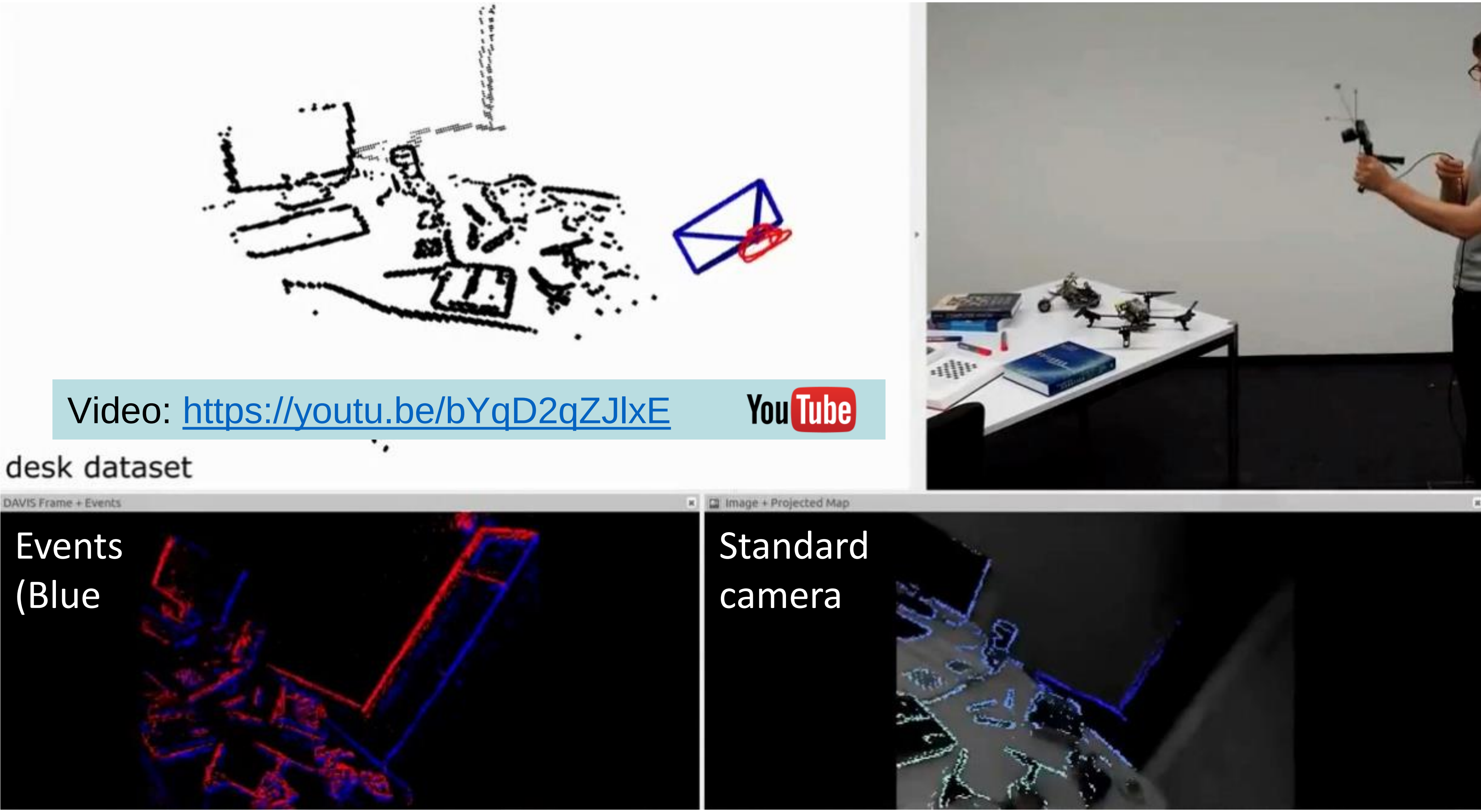
Rigid-body transformation T minimizes alignment error:

$$T = argmin \sum_{\boldsymbol{u}} \Big(M(\boldsymbol{u}) - I\big(\boldsymbol{W}(\boldsymbol{u}; T)\big)\Big)^2$$

Rebecq, Horstschaefer, Gallego, Scaramuzza, “EVO: Geometric Approach to 6-DoF Event-based Tracking and Mapping in Real Time”, IEE Robotics and Automation Letters, 2017. **Also EU Patent**

Results

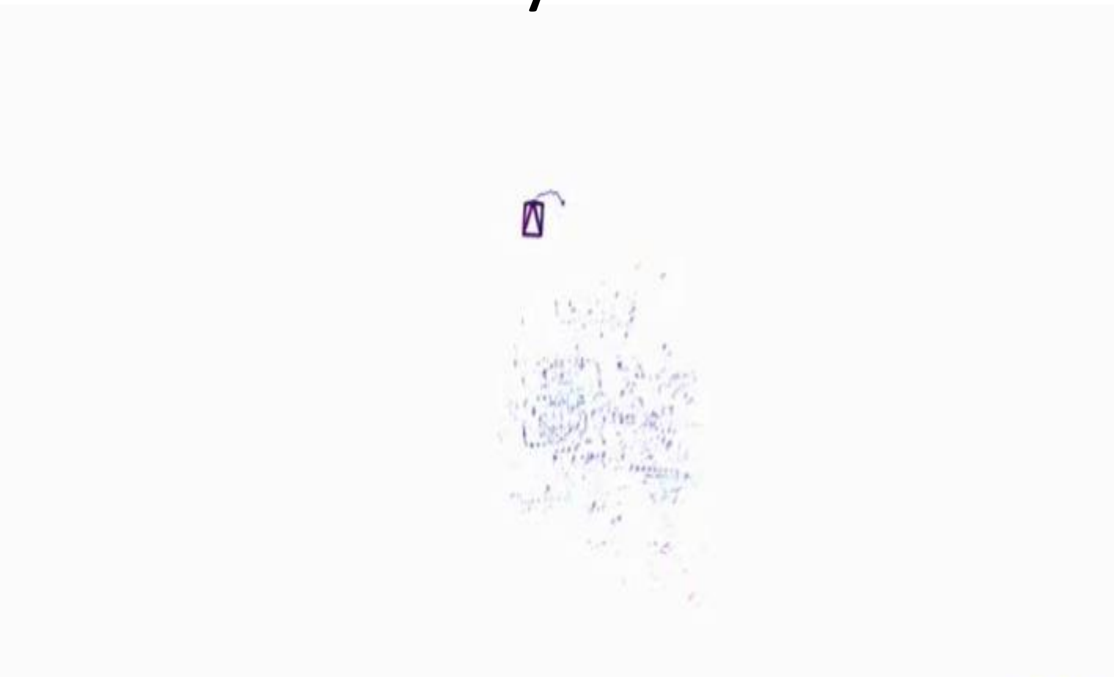
High-Speed tracking



Video: https://youtu.be/bYqD2qZJlxE YouTube

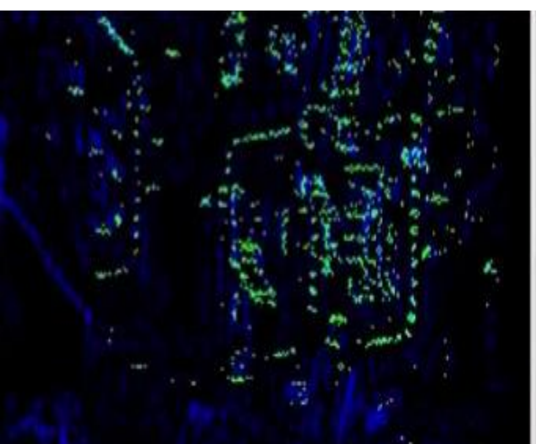
Rebecq, Horstschaefer, Gallego, Scaramuzza, "EVO: Geometric Approach to 6-DoF Event-based Tracking and Mapping in Real Time", IEE Robotics and Automation Letters, 2017. **Also EU Patent**

EVO: Multi-Keyframe Scene



Video: https://youtu.be/bYqD2qZJlxE YouTube

Observed and reprojected events



Intensity reconstruction from events

Frame of a standard camera plus events

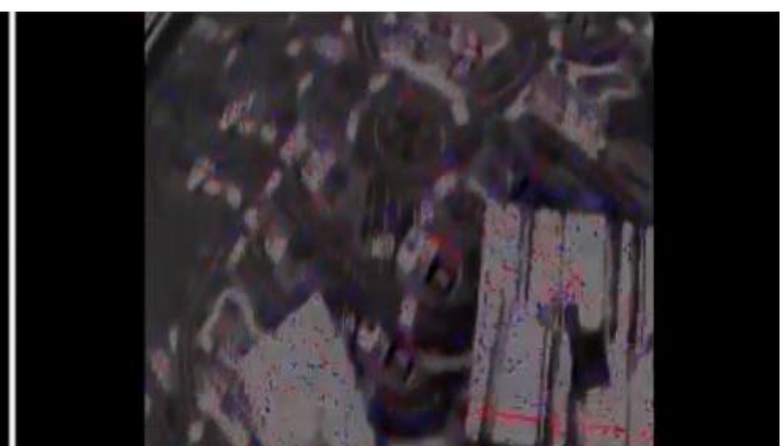
Light On and OFF Experiment

Video: https://youtu.be/bYqD2qZJlxE YouTube

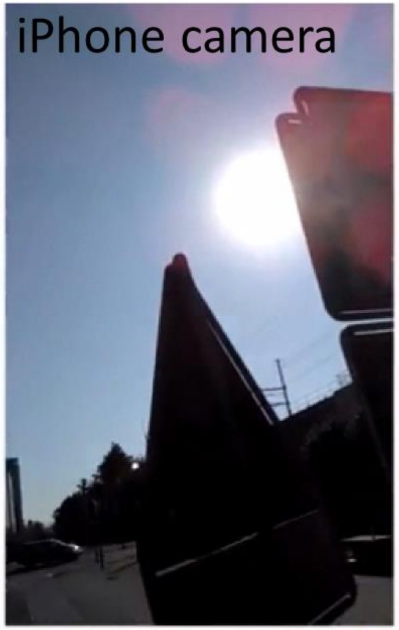
Observed and reprojected events

Intensity reconstruction from events

Frame of a standard camera plus events



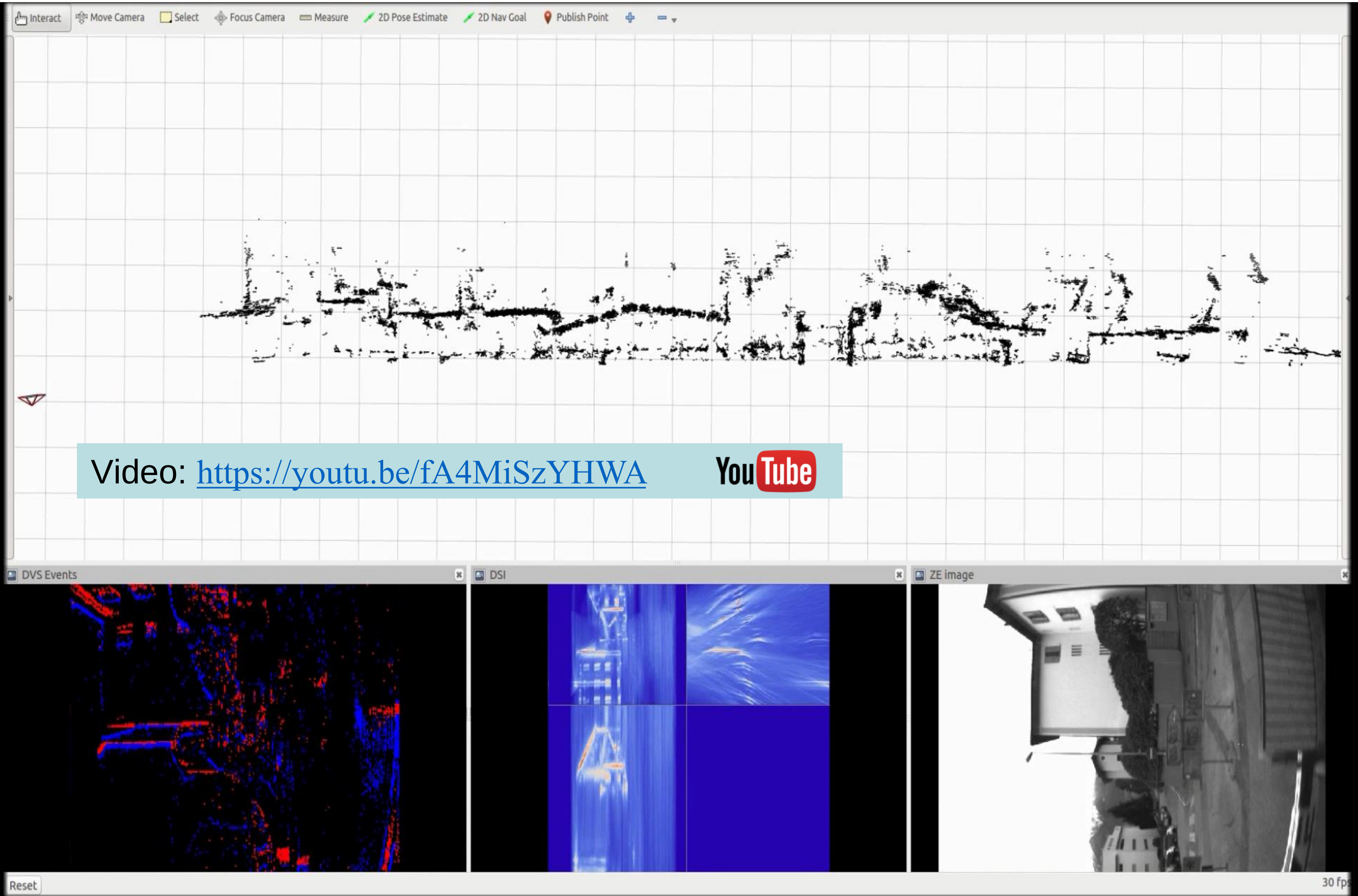
EVO Robustness to High-Dynamic Range Scenes



Video: https://youtu.be/bYqD2qZJlxE

Rebecq, Horstschaefer, Gallego, Scaramuzza, "EVO: Geometric Approach to 6-DoF Event-based Tracking and Mapping in Real Time", IEEE Robotics and Automation Letters, 2017. **Also EU Patent**

3D Mapping from a Train



Video: https://youtu.be/fA4MiSzYHWA

Summary of EVO

- Very **simple** to implement
- Works even in **high-speed** and **HDR** scenes, where standard cameras fail
- **Real-time** even on a **smartphone CPUs (Odroid XU4)**!
- **Intensity reconstruction** not needed but available
- **Come and see our live demo!**

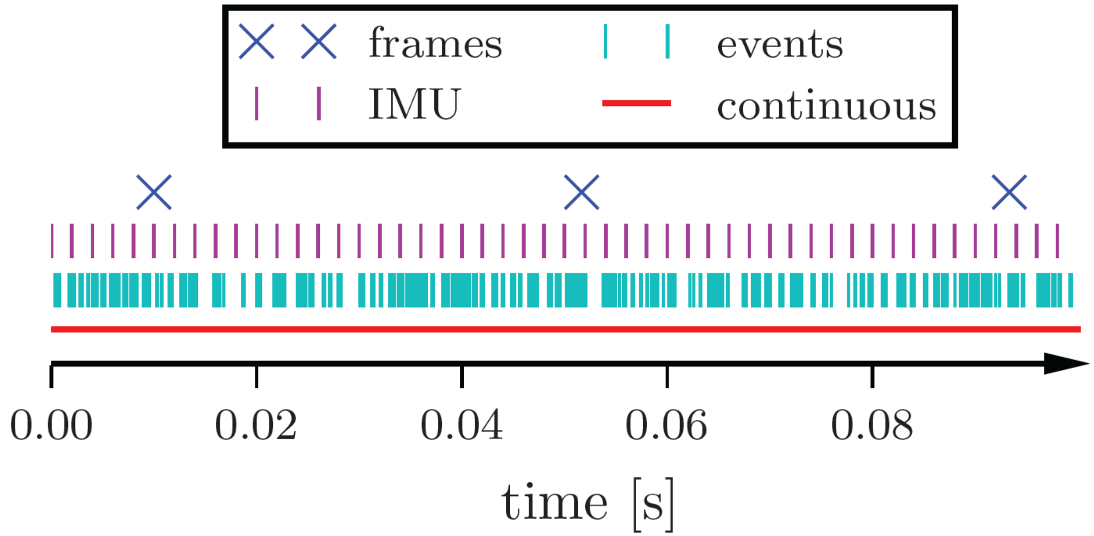
Continuous-Time Visual-Inertial Trajectory Estimation with Event Cameras

Mueggler, Gallego, Rebecq, Scaramuzza

ArXiv, 2017
submitted to TRO

[Mueggler, Gallego, Scaramuzza: *Continuous-Time Trajectory Estimation for Event-based Vision Sensors*, RSS'15]
[Mueggler, Gallego, Rebecq, Scaramuzza: *Continuous-Time Visual-Inertial Trajectory Estimation with Event Cameras*, under review, on arXiv, submitted to TRO'17]

Continuous-Time Trajectory Estimation [RSS'15, Arxiv'17]



- Event stream is asynchronous and high-frequency (almost continuous)
- A single event is ambiguous and does not constrain a pose

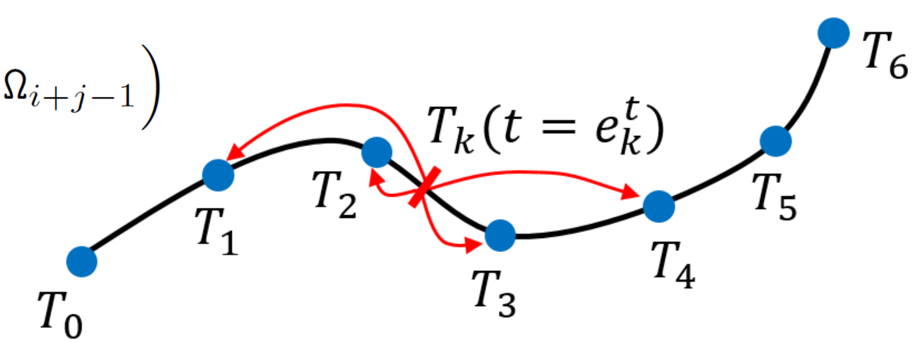
[Mueggler, Gallego, Scaramuzza: *Continuous-Time Trajectory Estimation for Event-based Vision Sensors*, RSS'15]
[Mueggler, Gallego, Rebecq, Scaramuzza: *Continuous-Time Visual-Inertial Trajectory Estimation with Event Cameras*, under review, on arXiv, submitted to TRO'17]

Continuous-Time Trajectory Estimation [RSS’15, Arxiv’17]

- ➢Estimate trajectory instead of discrete poses: $T_1, T_2, T_3, \ldots \rightarrow T(t)$
- ➢Advantages
  - Pose (and its derivatives) is well-defined at *any* time
  - Can handle asynchronous, high-frequency data naturally
- ➢**Spline Fusion** [Lovegrove, IJCV’15]
  - Trajectory is represented with **B-splines**
  - **Cumulative basis functions** on SE(3), free from singularities:

$$\mathrm{T}_{w,s}(u(t)) \doteq \mathrm{T}_{w,i-1} \prod_{j=1}^{3} \exp\left(\tilde{\mathbf{B}}_j(u(t))\, \Omega_{i+j-1}\right)$$

$$\Omega_q = \log(\mathrm{T}_{w,q-1}^{-1} \mathrm{T}_{w,q})$$



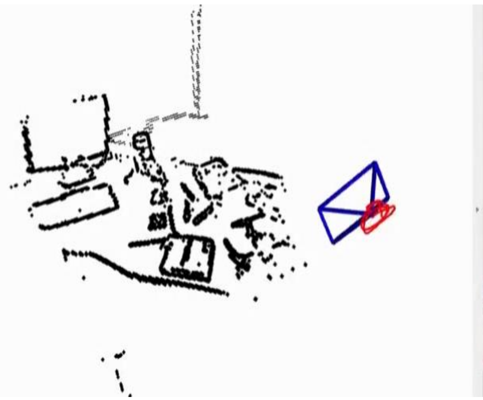
Optimization

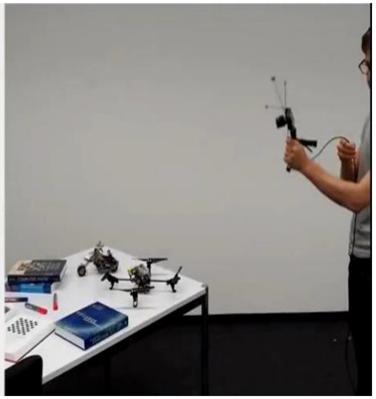
- Find control poses $T^*_{w,i}$ and parameters θ^* (**map scale, gravity alignment, and IMU biases**) *such that* reprojection error of all events and the inertial residuals is minimized:

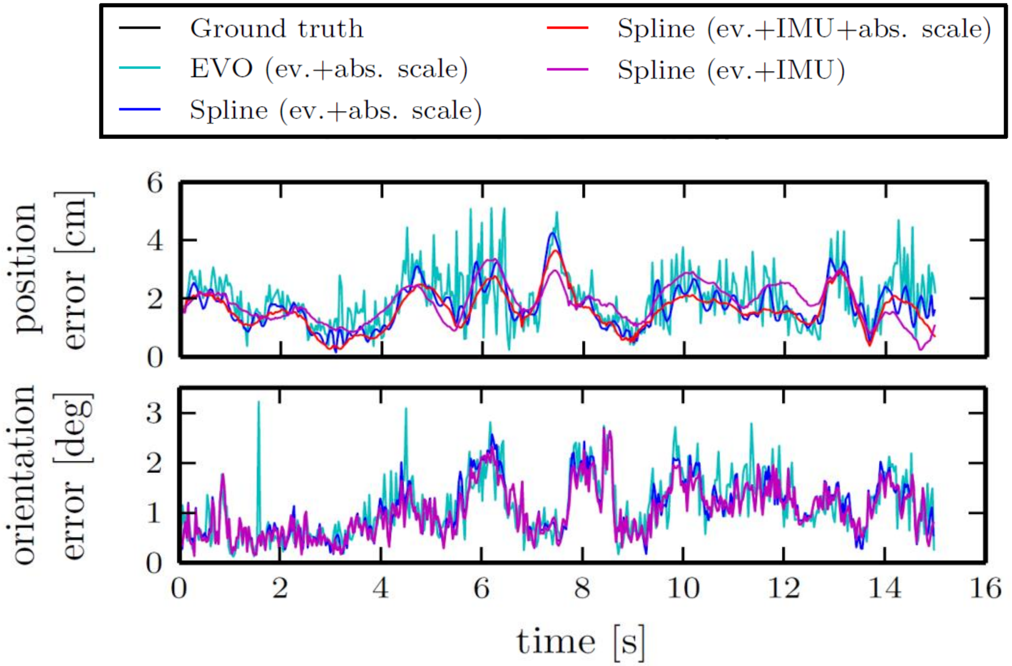
$$\{\mathrm{T}^*_{w,i}, \boldsymbol{\theta}^*\} = \arg\min_{\mathrm{T},\boldsymbol{\theta}} \overbrace{\sum_k \frac{1}{\sigma_k^2} \|\mathbf{e}_k - \hat{\mathbf{e}}_k(\mathbf{x}(t_k), \mathcal{M})\|^2}^{\text{Events}} \underbrace{+ \sum_i \frac{1}{\sigma_i^2} \|\boldsymbol{\omega}_i - \hat{\boldsymbol{\omega}}_i(\mathbf{x}(t_i))\|^2 + \sum_j \frac{1}{\sigma_j^2} \|\mathbf{a}_j - \hat{\mathbf{a}}_j(\mathbf{x}(t_j))\|^2}_{\text{Inertial Measurements}}$$

- Few control poses are needed: **1 control pose per 10,000 events**

Results







	Position error (abs. [cm] and rel. [%])						Orientation error [°]		
	μ	%	σ	%	max	%	μ	σ	max
EVO (ev.+abs. scale)	1.08	0.50	0.53	0.25	4.64	2.17	1.31	0.68	3.55
Spline (ev.+abs. scale)	0.78	0.36	0.40	0.19	2.30	1.08	0.98	0.58	3.56
Spline (ev.+IMU+abs. scale)	0.69	0.32	0.36	0.17	1.67	0.78	0.95	0.57	3.49
Spline (ev.+IMU)	0.78	0.36	0.47	0.22	2.07	0.97	0.95	0.57	3.49

Relative errors are given with respect to the mean scene depth.

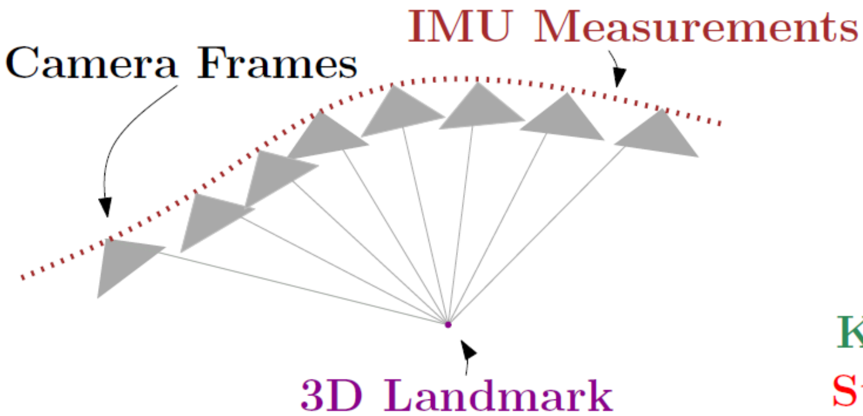
Real-time Visual-Inertial Odometry for Event Cameras using Keyframe-based Nonlinear Optimization

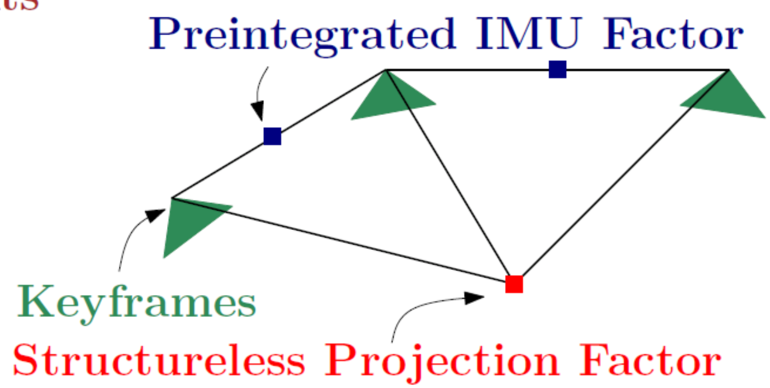
Rebecq, Horstschäfer, Scaramuzza

submitted to BMVC'17

Visual-Inertial Fusion via Non-linear Optimization

- Fusion solved as a *non-linear optimization problem*
- 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, On-Manifold Preintegration for Real-Time Visual-Inertial Odometry, RSS'15, TRO 17

Leutenegger, Lynen, Bosse, Siegwart, Furgale, Keyframe-based visual–inertial odometry using nonlinear optimization, RSS'13, IJRR'15

In the following sequences, we show the trajectory estimated by our pipeline, using the events and IMU only.

The standard camera images are **not** used, and shown for illustration only.

Computation time per *event frame:*
8ms on an i7 Lenovo quadcore **CPU** and
~**30ms** on smartophone CPU (Odroid XU4).

Come and see our live demo!

	Time (ms)
synthesize event frame	4.23
feature detection	0.69
feature tracking	0.90
two-point RANSAC	0.08
add frame to back-end	1.47
wait for back-end	1.29
total time	**8.23**

Conclusions

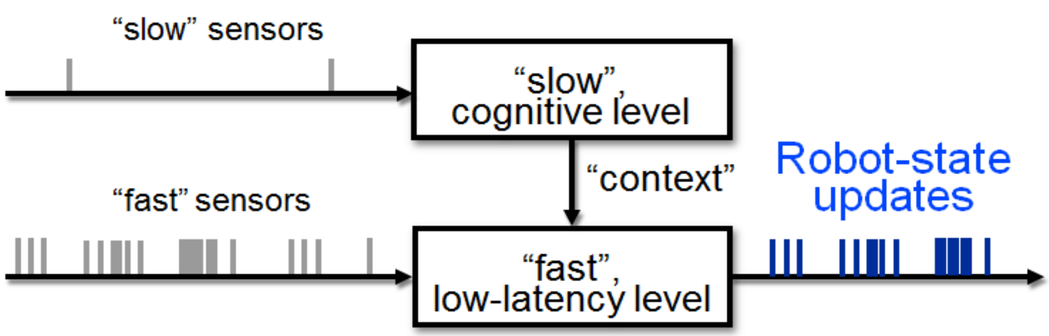
Event cameras are revolutionary and open enormous possibilities!

- Robustness to **high speed motion** and **high-dynamic-range scenes**
- Standard cameras have been studied for 50 years! → need of a change!
- Challenges: asynchronous & binary output, **complete noise & model characterization still missing** (e.g., memory effects and other non idealities)

Future: low-latency perception and control via **a two-level sensing architecture**

➢ **Fast low-latency level**: where *agile behavior* is obtained by a low-latency control action that uses data from fast sensors (e.g., DVS, IMU)

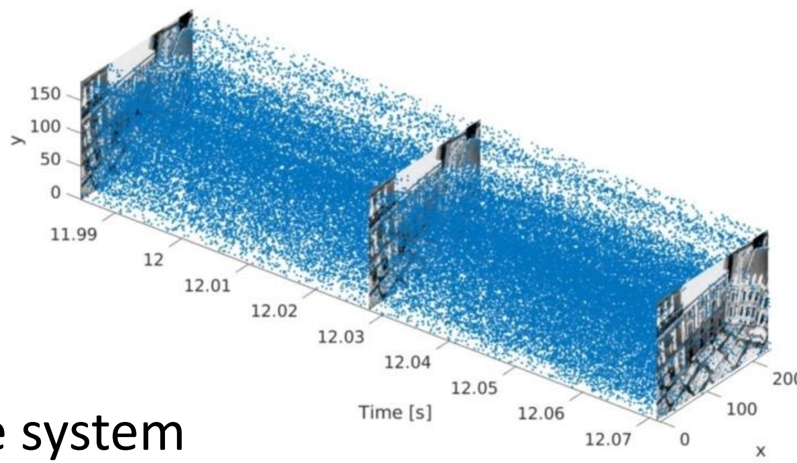
➢ **Slow cognitive level**: tasks, such as recognition, mapping, & loop closing, are done based on slower traditional sensors (cameras, lidars)



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

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



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!

DARPA

Google
Faculty Research Awards

erc

Schweizerische Eidgenossenschaft
Confédération suisse
Confederazione Svizzera
Confederaziun svizra

Kommission für Technologie und Innovation KTI

FNSNF
Swiss National Science Foundation

robotics
Swiss National
Centre of Competence
in Research

Dr. Guillermo Gallego

Elias Mueggler

Henri Rebecq

Timo Horstschäfer

Resources

- **My research on event-based vision:**
  http://rpg.ifi.uzh.ch/research_dvs.html
- **Event camera dataset and simulator:**
  http://rpg.ifi.uzh.ch/davis_data.html
- Lab homepage: http://rpg.ifi.uzh.ch/
- Other 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