Lab Visit and Exercise - Today

- **Lab visit with live demos** (@Robotics and Perception Group):
  - We will take Tram 10 to Bahnhof Oerlikon Ost
  - Lab address: Andreasstrasse 15, 2nd floor, 2.11
  - Visit starts at 12:30hrs
  - Duration of the visit: 1.5-2 hours (feel free to leave at any time)
  - Afterwards, chocolates and drinks in the lab lounge

- **Lunch**: Sandwiches will be served. You can eat them during the visit

- **Exercise** Session: **Q&A on final VO integration**
  - Room **UZH BIN 0.B.06** from **14:30 to 17:00 hrs**
    Address: Binzmuehlestrasse 14, 8050 Zurich

Exams Questions

- The oral exam will last 30 minutes
- It will consist of one application question followed by two theoretical questions
- This document contains a "**non exhaustive**" list of possible application questions and an "**exhaustive**" list of all the topics that you should learn about the course, which will be subject of discussion in the theoretical part: http://rpg.ifi.uzh.ch/docs/teaching/2019/Exam_Questions.pdf

University of Zurich[UZH] | ETH zürich

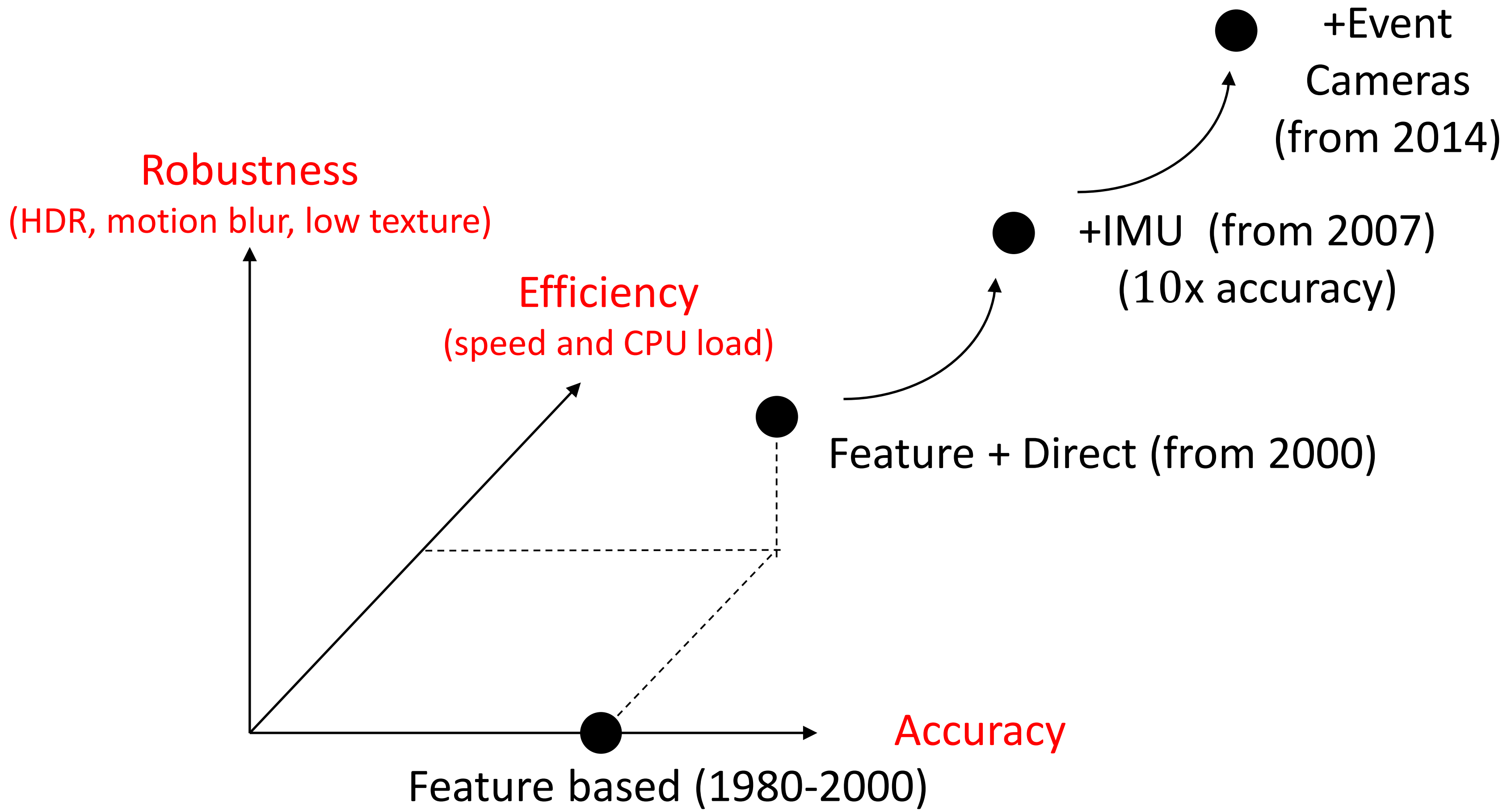
ROBOTICS & PERCEPTION GROUP

Institute of Informatics – Institute of Neuroinformatics

Lecture 14
Event based vision

Davide Scaramuzza
http://rpg.ifi.uzh.ch

A Short Recap of the last 30 years of VIO



Robustness: Challenges of Vision for SLAM

- IMU alone only helpful for short motions; **drifts very quickly** without visual constraint
- Biggest challenges for vision today is robustness to:
  - **High Dynamic Range (HDR)**
    - Can be handled with Active Exposure Control or Event cameras
  - High-speed motion (i.e., **motion blur**)
    - Can be handled with event cameras
  - **Low-texture** scenes
    - Can be handled with Dense Methods, or with Depth cameras (laser projector) or by getting closer to the scene, or by using context (e.g., machine learning)
  - **Dynamic environments**
    - Can be handled with an IMU, using context (e.g., machine learning)
- Current VO algorithms and sensors have **large latencies** (50-200 ms)
  - Can we reduce this to much below a 1ms?
    - Can be handled with event cameras

Event-based Cameras

References

- **Tutorial** paper:
  G. Gallego, T. Delbruck, G. Orchard, C. Bartolozzi, B. Taba, A. Censi, S. Leutenegger, A. Davison, J. Conradt, K. Daniilidis, D. Scaramuzza,
  Event-based Vision: A Survey, arXiv, 2019. PDF

- List of event camera papers, codes, datasets, companies: https://github.com/uzh-rpg/event-based_vision_resources

- Event-camera simulator: http://rpg.ifi.uzh.ch/esim.html

- More on our research: http://rpg.ifi.uzh.ch/research_dvs.html

Open Challenges in Computer Vision

The past 60 years of research have been devoted to frame-based cameras ...but they are not good enough!

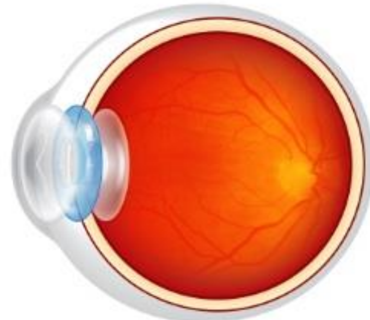
Latency & Motion blur

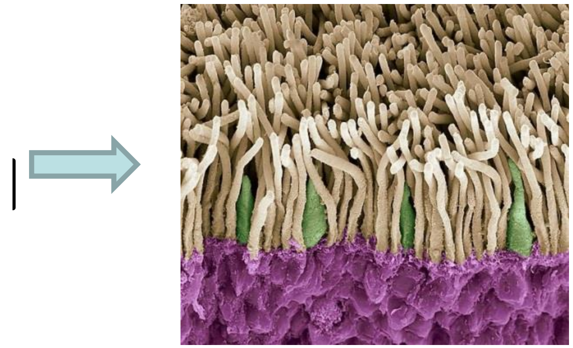
Dynamic Range

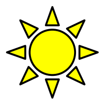
Event cameras do not suffer from these problems!

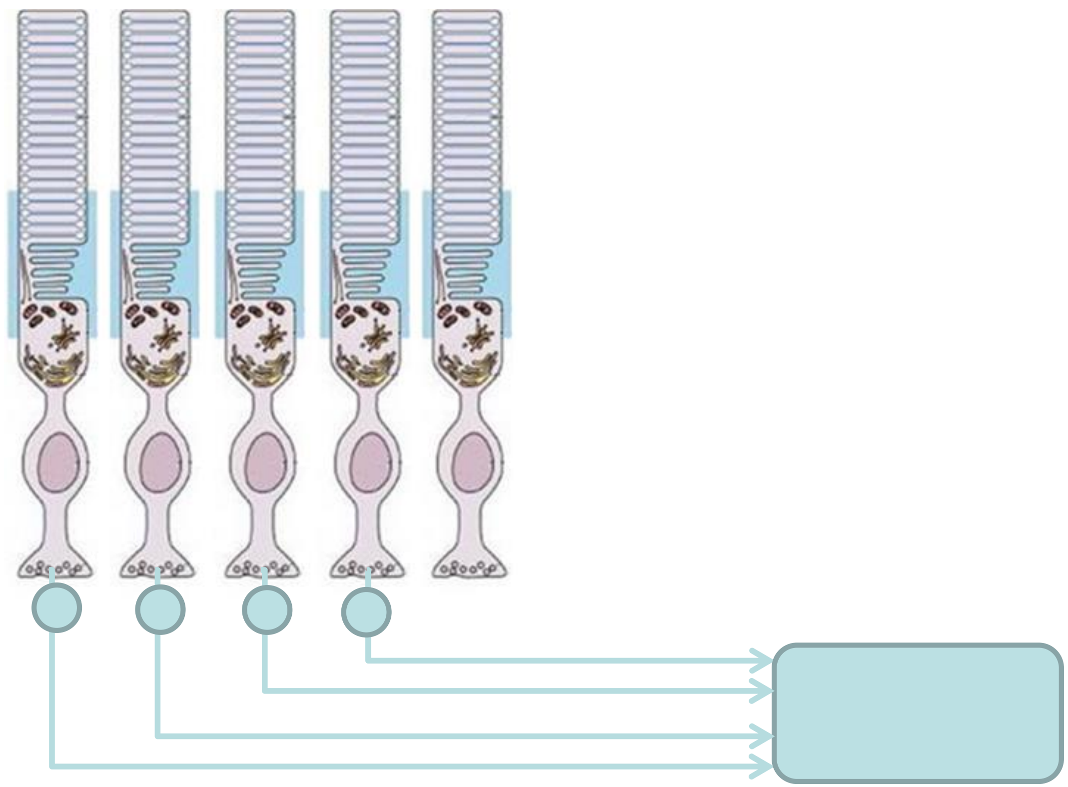
Human Vision System

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

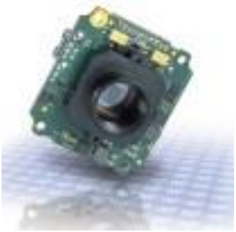








Dynamic Vision Sensor (DVS)



DVS from inilabs.com

First commercialized by Prof. T. Delbruck in 2008 at the Institute of Neuroinformatics of UZH & ETH

Advantages

- **Low-latency** (~1 micro-seconds)
- **High-dynamic range (HDR)** (140 dB instead 60 dB)
- **High updated rate** (1 MHz)
- **Low power** (10mW instead 1W)

Challenges

- **Paradigm shift**: Requires totally **new vision algorithms** because:
  - **Asynchronous** pixels
  - **No intensity information** (only binary intensity changes)

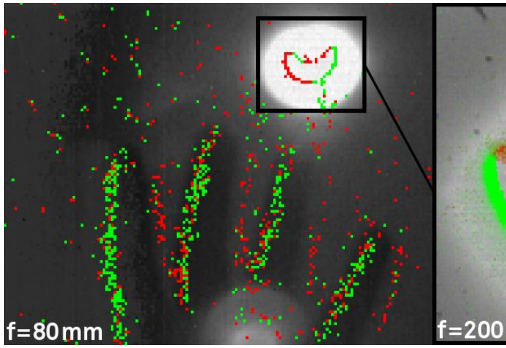
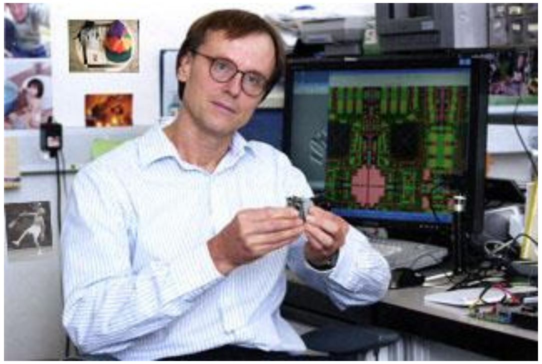


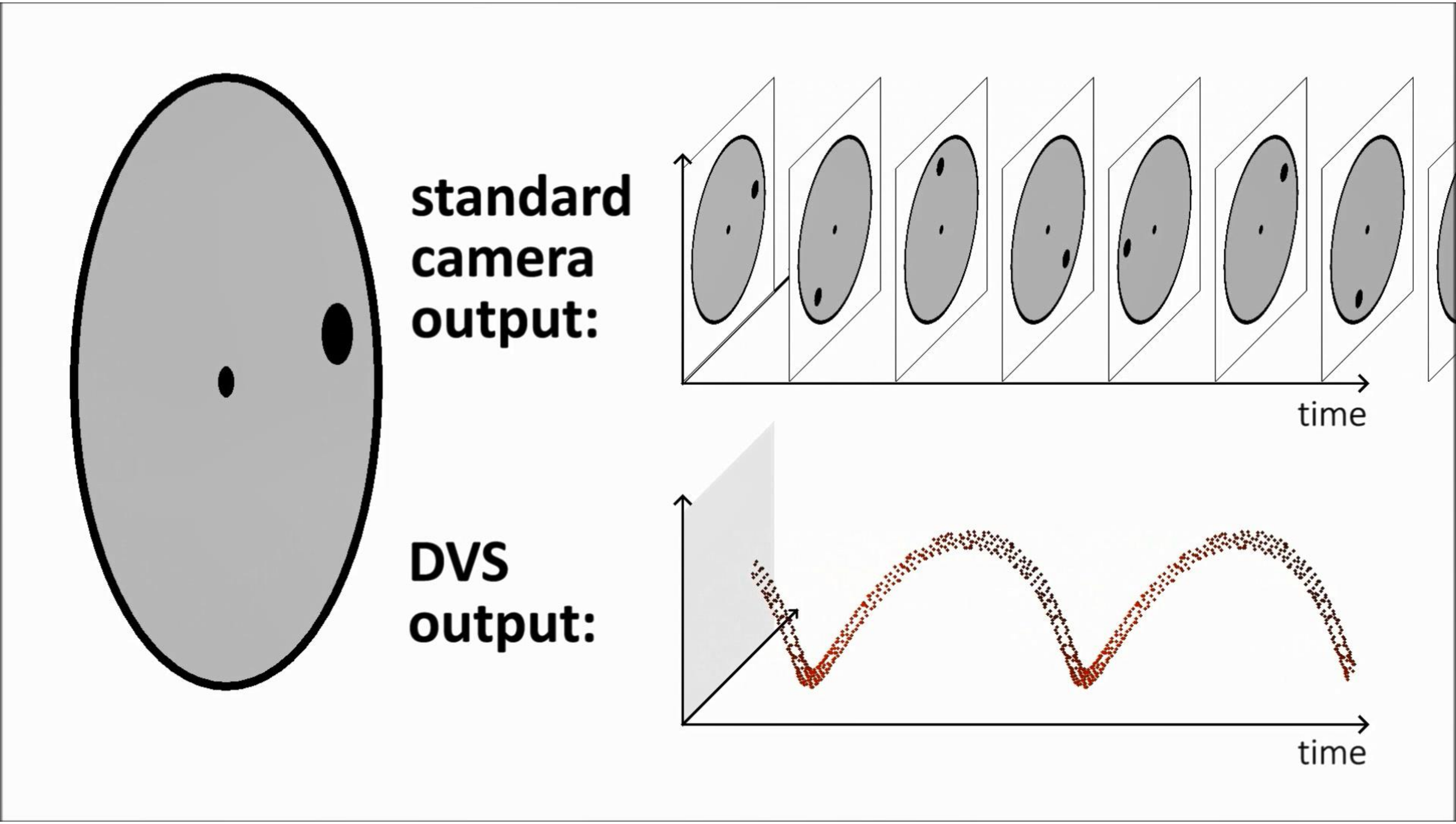
Image of solar eclipse captured by a DVS, without black filter!



Prof. Tobi Delbruck, UZH & ETH Zurich

Lichtsteiner et al., A 128x128 120 dB 15µs Latency Asynchronous Temporal Contrast Vision Sensor, 2008

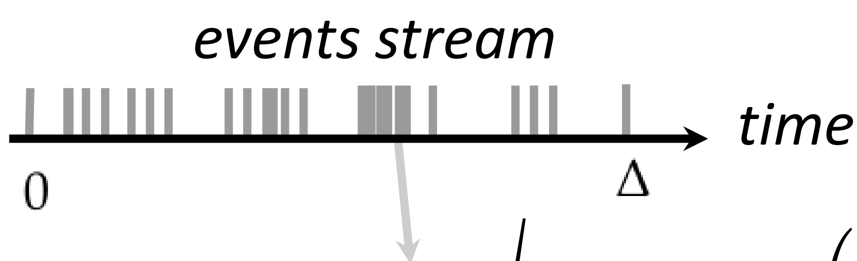
Camera vs Dynamic Vision Sensor



Dynamic Vision Sensor (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 detects a change of intensity



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

Timestamp (μs)

Pixel coordinates

Event polarity (or sign) (*+1 or -1*): *positive or negative change*

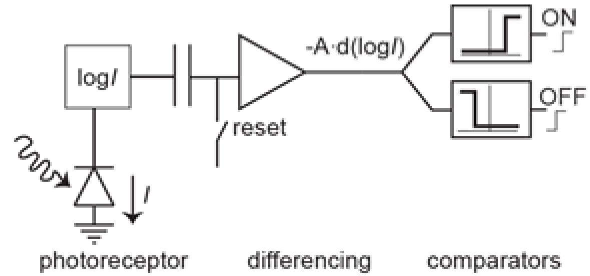
Lichtsteiner, Posch, Delbruck. *A 128x128 120 dB 15μs Latency Asynchronous Temporal Contrast Vision Sensor.* 2008

What is an event camera, **precisely**?

- **Asynchronous**: all pixels are *independent* from one another
- Implements ***level-crossing*** sampling rather than uniform time sampling
- Reacts to ***logarithmic*** brightness changes

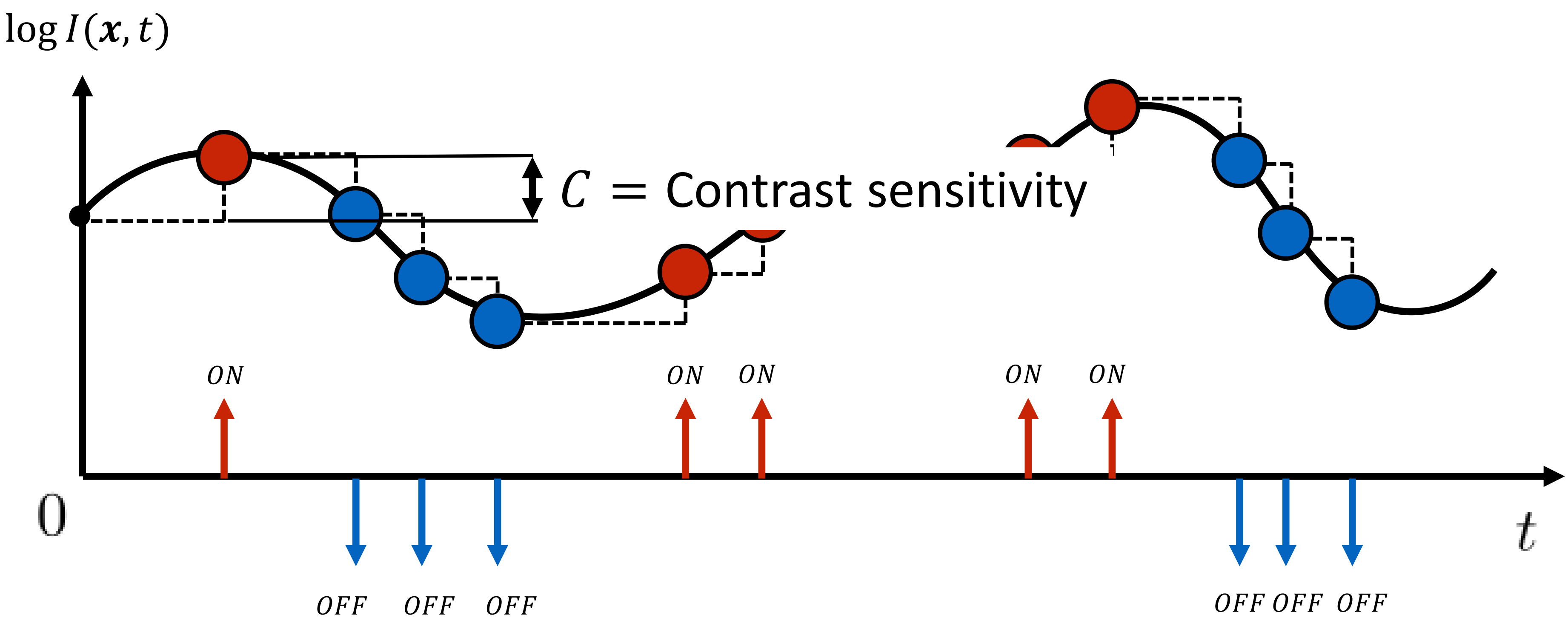
Let´s look at how this works for one pixel in detail

Generative Event Model



Consider the intensity at a **single pixel...**

$$\pm C = \log I(\boldsymbol{x}, t) - \log I(\boldsymbol{x}, t - \Delta t)$$

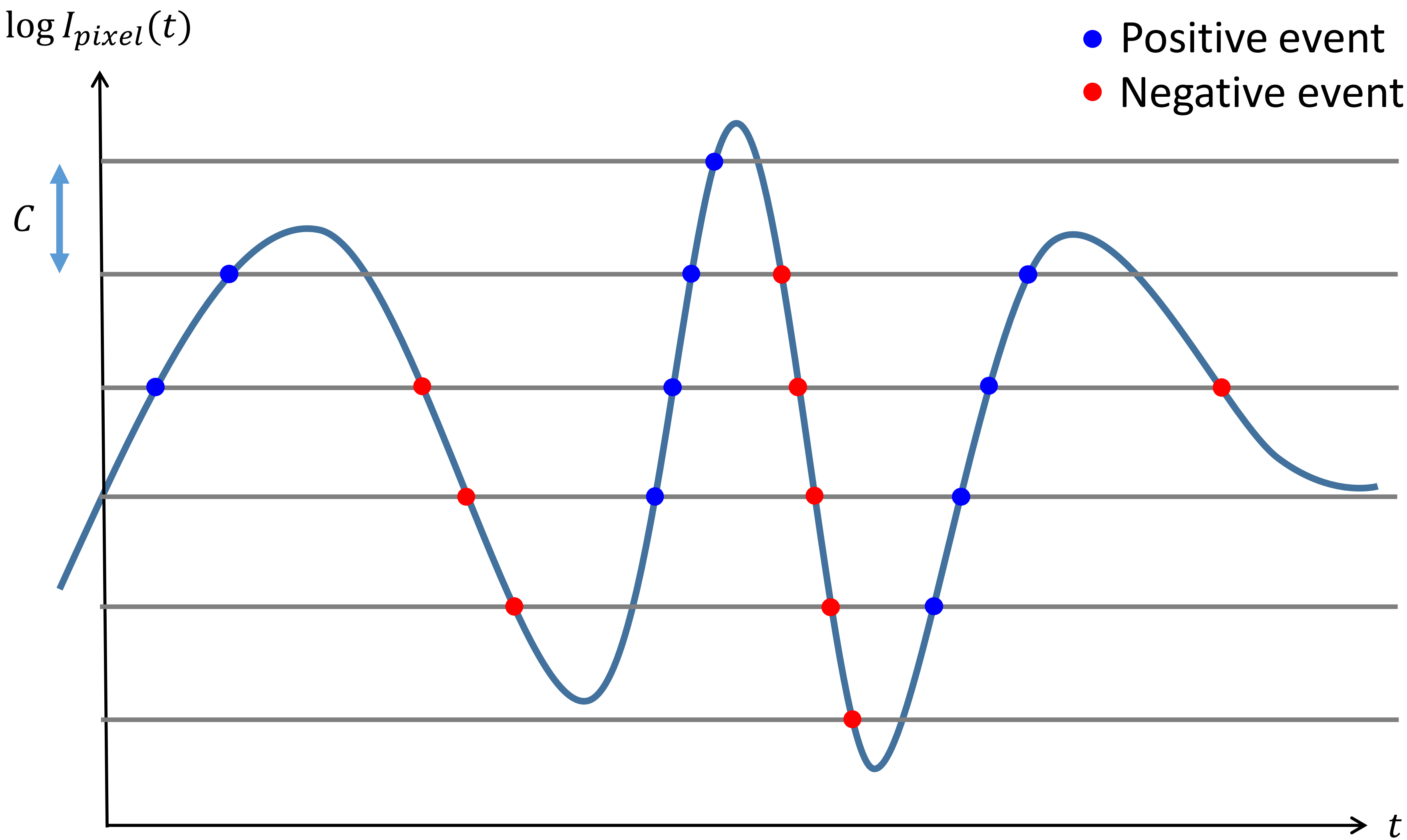


Events are triggered **asynchronously**

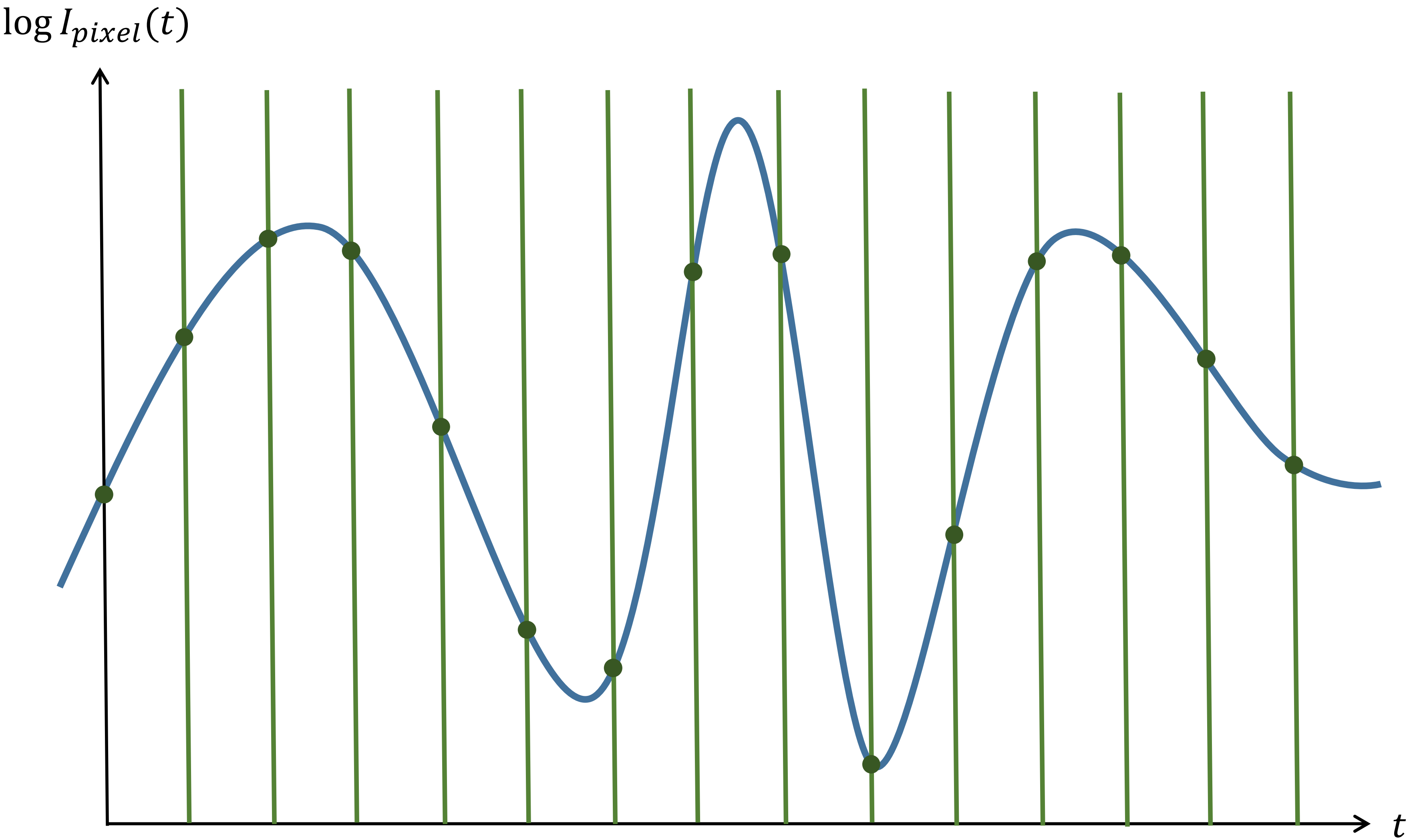
Lichtsteiner, Posch, Delbruck, A 128x128 120 dB 15μs Latency Asynchronous Temporal Contrast Vision Sensor, IEEE Journal of Solid-State Circuits, 2008. [PDF](PDF).

Event cameras sample intensity when this crosses a threshold (**Level-crossing sampling**)

- An **event** is generated when the signal *change* equals C



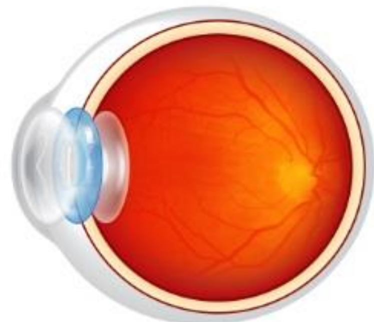
Standard cameras sample intensity at uniform time intervals (**uniform time sampling**)

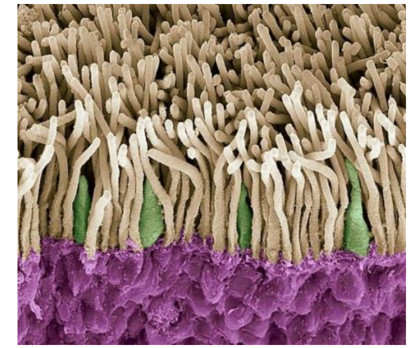


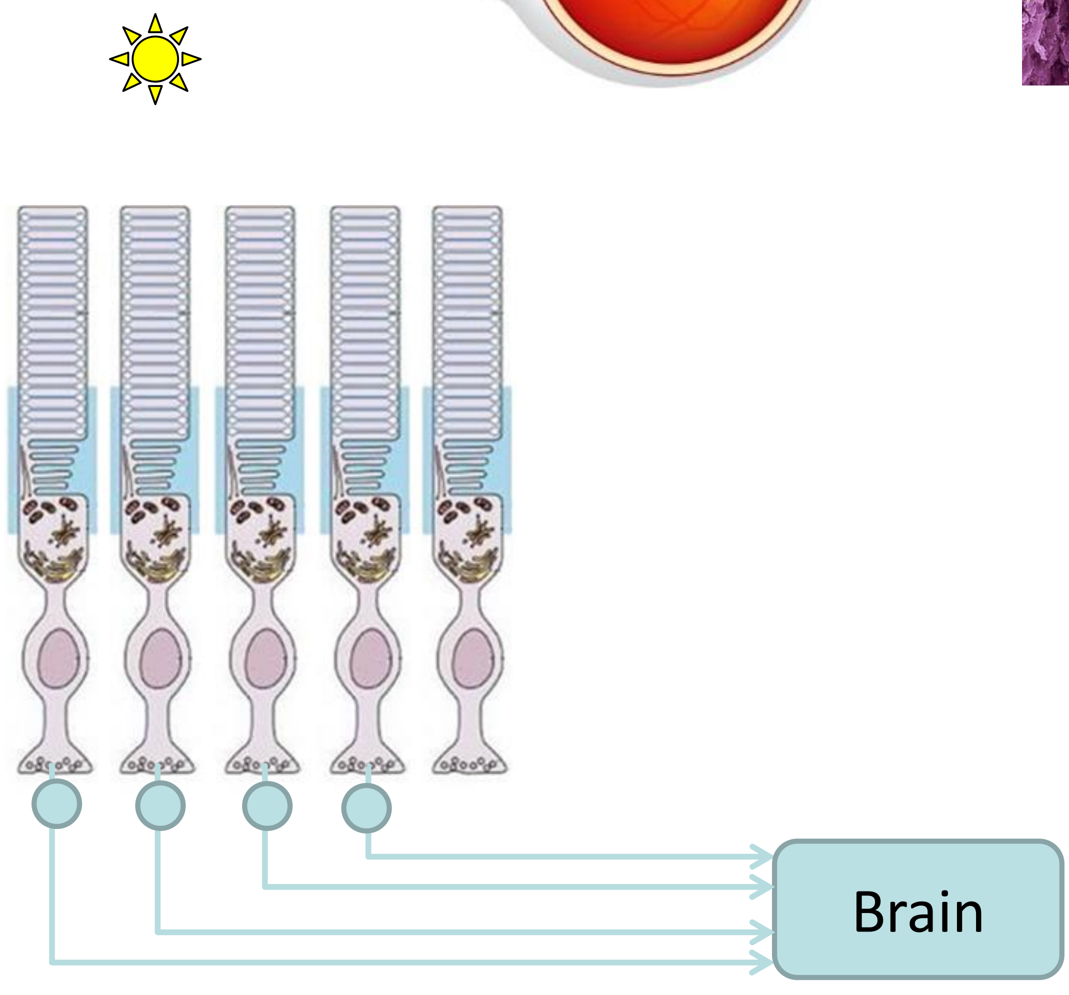
Event cameras are inspired by the Human Eye

Human retina:

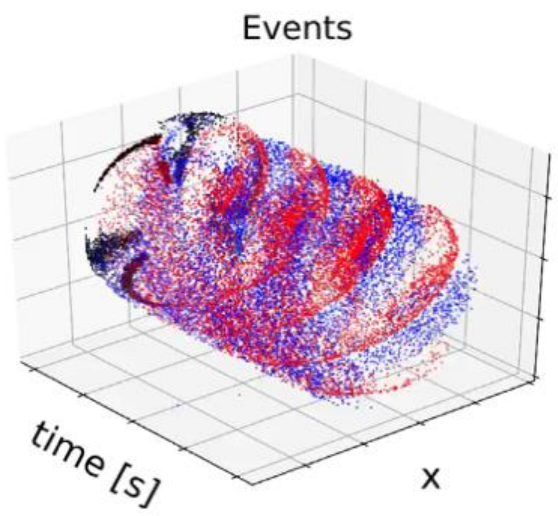
- 130 million **photoreceptors**
- But only 2 million **axons**!







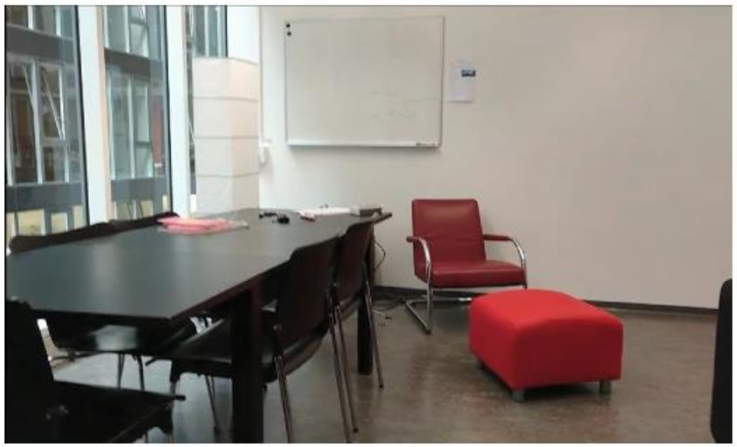
Event Camera output with Motion: Space-time domain



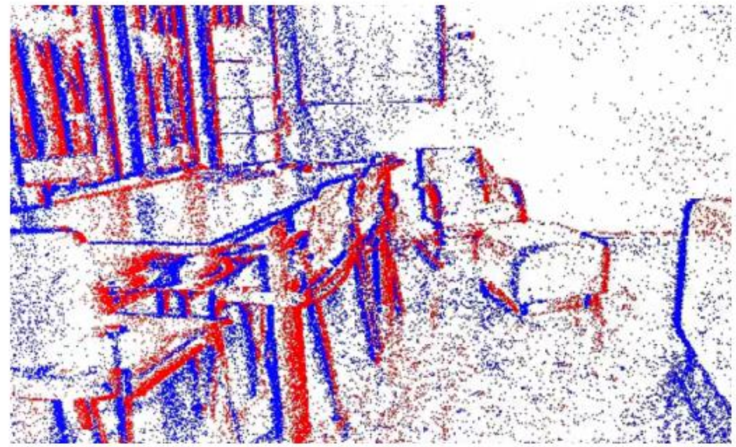
Events in the **space-time** domain (x, y, t)

Event Camera output with Motion: image domain

Standard Camera



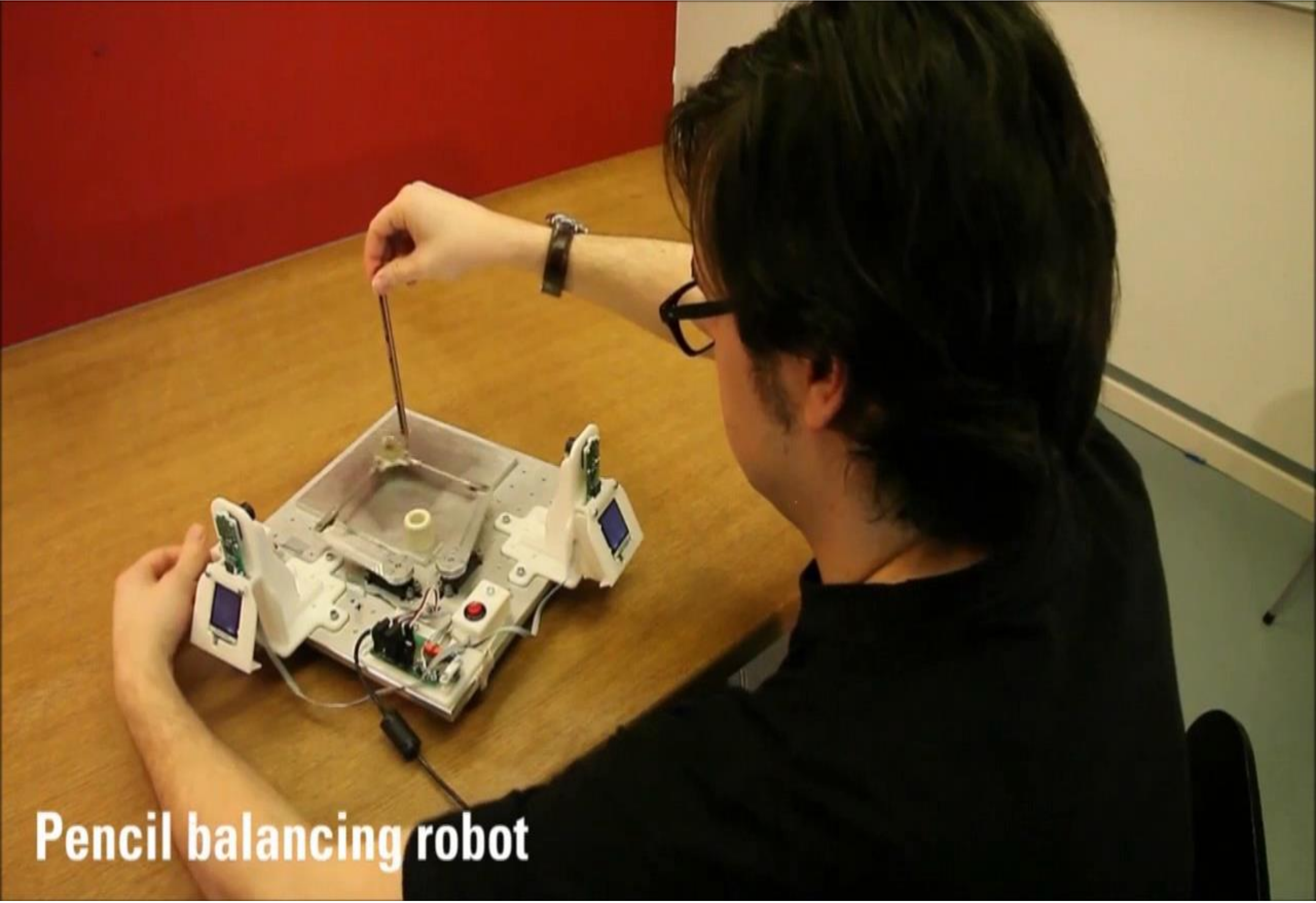
Event Camera (ON, OFF events)



ΔT = 40 ms

Events in the **image domain** (x, y)
Integration time can be arbitrary: from 1 microsecond to infinity)

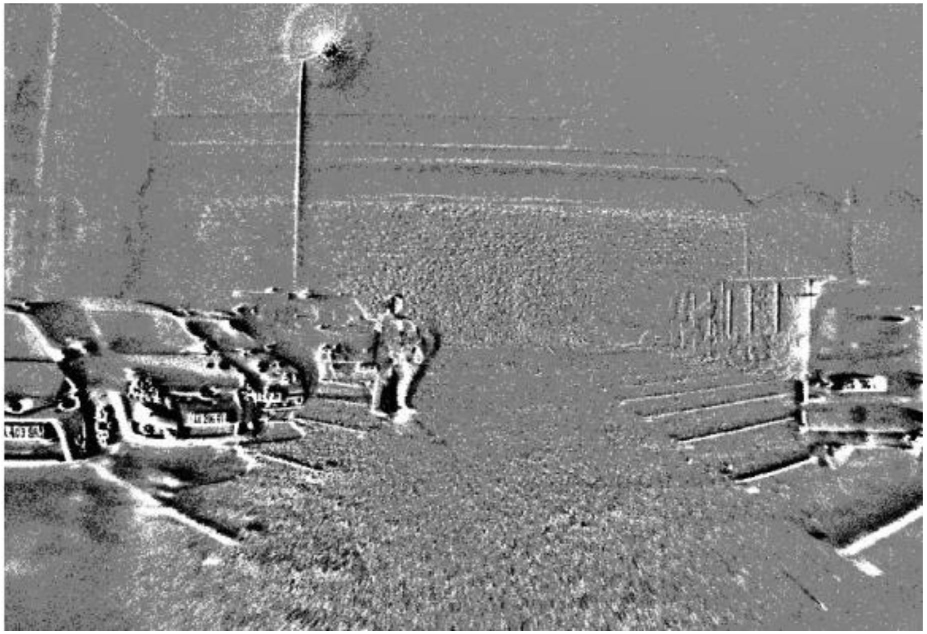
Examples



Conradt, Cook, Berner, Lichtsteiner, Douglas, Delbruck, **A pencil balancing robot using a pair of AER dynamic vision sensors**, IEEE International Symposium on Circuits and Systems. 2009

Low-light Sensitivity (night drive)

GoPro Hero 6



Aggregated event image
(pixel intensity equal to the sum of positive (+1) and negative (-1) events in a given time interval)

Video courtesy of Prophesee: https://www.prophesee.ai

High-speed vs Event Cameras

	High speed camera	Standard camera	Event Camera
Max fps or measurement rate	**Up to 1MHz**	100-1,000 fps	**1MHz**
Resolution at max fps	64x16 pixels	**>1Mpxl**	**>1Mpxl**
Bits per pixels (event)	12 bits	8-10 per pixel	~**40 bits/event** {t,(x,y),p)}
Weight	6.2 Kg	**30 g**	**30 g**
Active cooling	yes	**No cooling**	**No cooling**
Data rate	1.5 GB/s	32MB/s	**~1MB/s on average (depends on dynamics)**
Mean power consumption	150 W + external light	1 W	**1 mW**
Dynamic range	n.a.	60 dB	**140 dB**

Current commercial applications

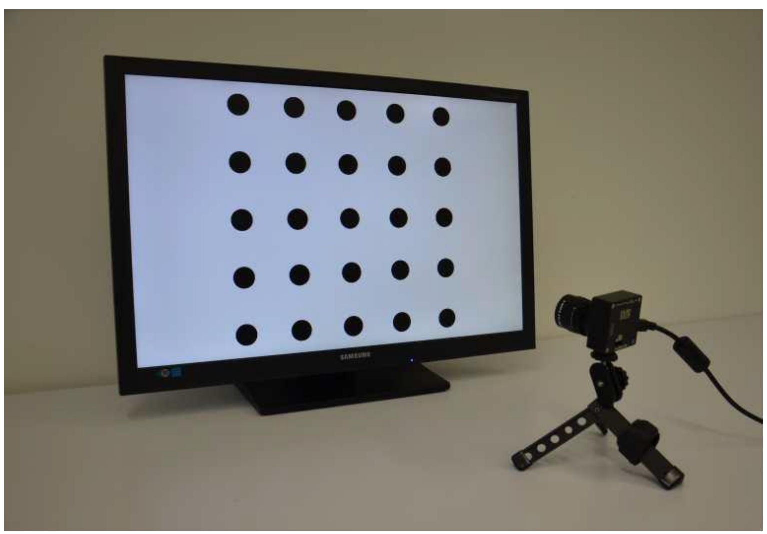
- **Internet of Things (IoT)**
  - Low-power, always-on devices for monitoring and surveillance
- **Automotive**:
  - low-latency, high dynamic range (HDR) object detection
  - low-power training & inference
  - low-memory storage
- **AR/VR**
  - low-latency, low-power tracking
- **Industrial automation**
  - Fast pick and place

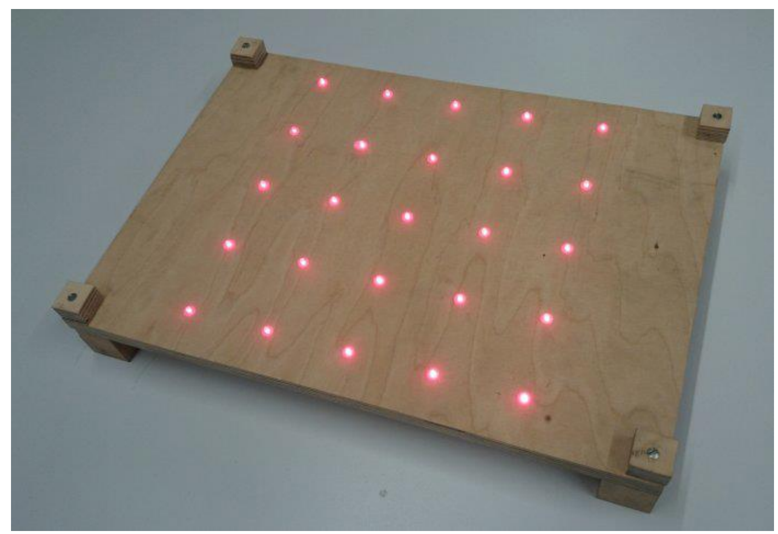
Who sells event cameras and how much are they?

- Inivation:
  - **DAVIS sensor**: **frames, events, IMU**.
  - Resolution**:** ~QVGA (346x260 pixels)
  - **Cost: 6,000 USD**
- Insightness:
  - **RINO sensor**: frames, events, IMU.
  - **Resolution:** ~QVGA (320x262 pixels)
  - **Cost: 6,000 USD**
- Prophesee:
  - **ATIS sensor:** events, IMU, absolute intensity at the event pixel
  - Resolution: 1M pixels
  - **Cost: 4,000 USD.**
- CelexPixel Technology:
  - **Celex One:** events, IMU, absolute intensity at the event pixel
  - Resolution: 1M pixels
  - **Cost: 1,000 USD.**
- **Samsung Electronic**s
  - Samsung DVS: events, IMU
  - Resolution: up to 1Mpxl
  - **Cost: not listed**

Calibration of a DVS [IROS’14]

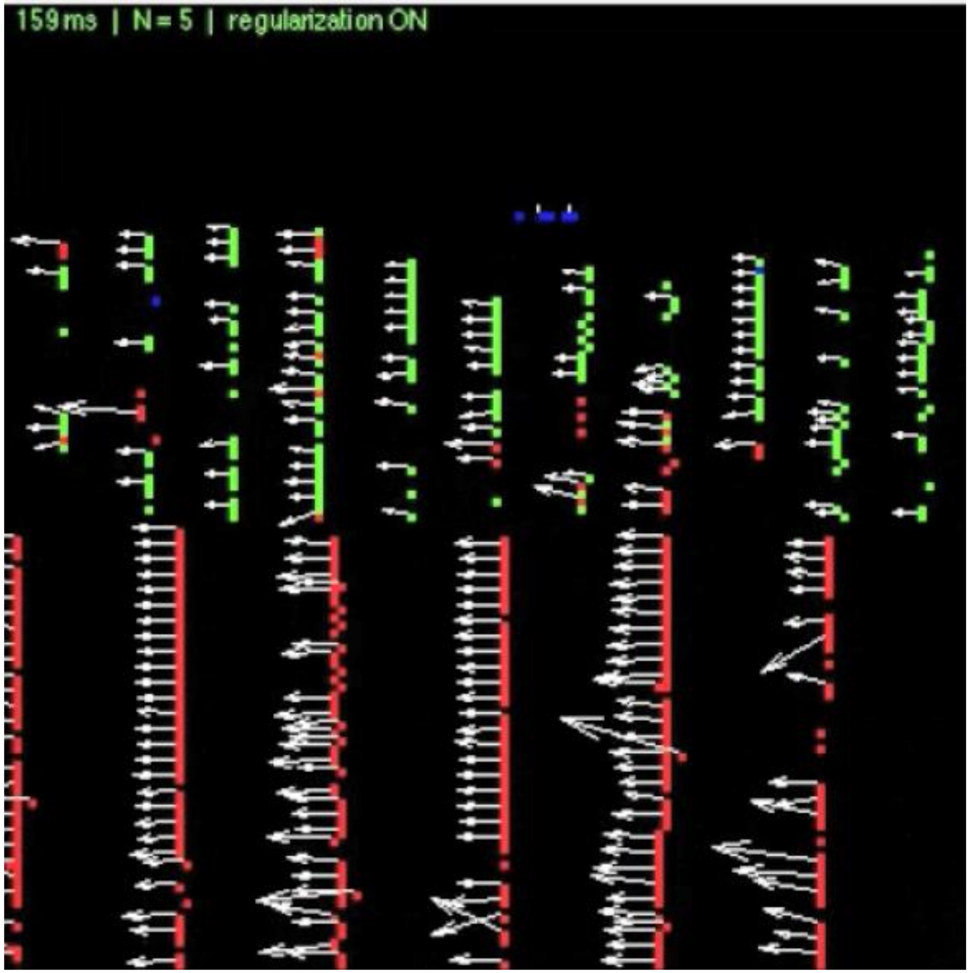
- Standard **pinhole camera model** still valid (same optics)
- Standard passive calibration patterns **cannot be used**
  - need to move the camera → inaccurate corner detection
- **Blinking patterns** (computer screen, LEDs)
- ROS DVS driver + intrinsic and extrinsic stereo calibration **open source**: https://github.com/uzh-rpg/rpg_dvs_ros





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

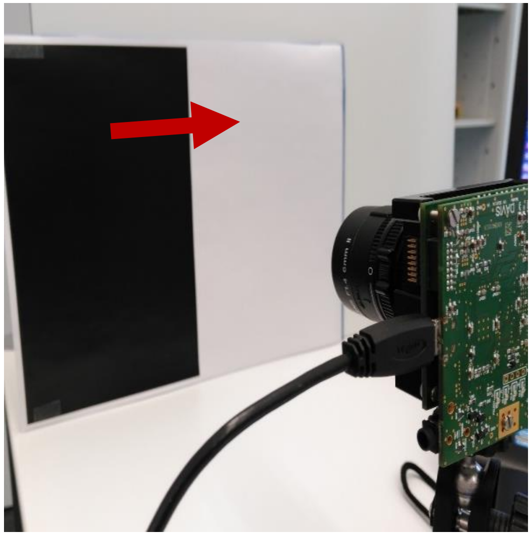
A Simple Optical Flow Algorithm

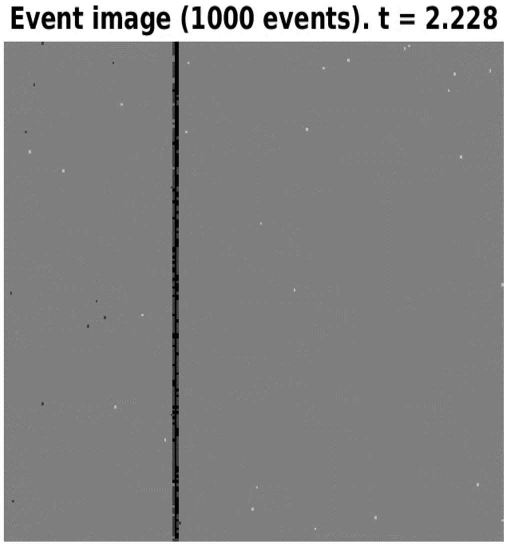


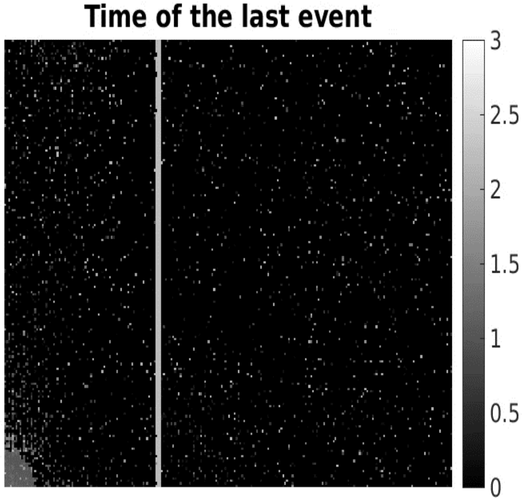
A moving edge

Horizontal motion

White pixels become black → brightness decrease → negative events (in black color)





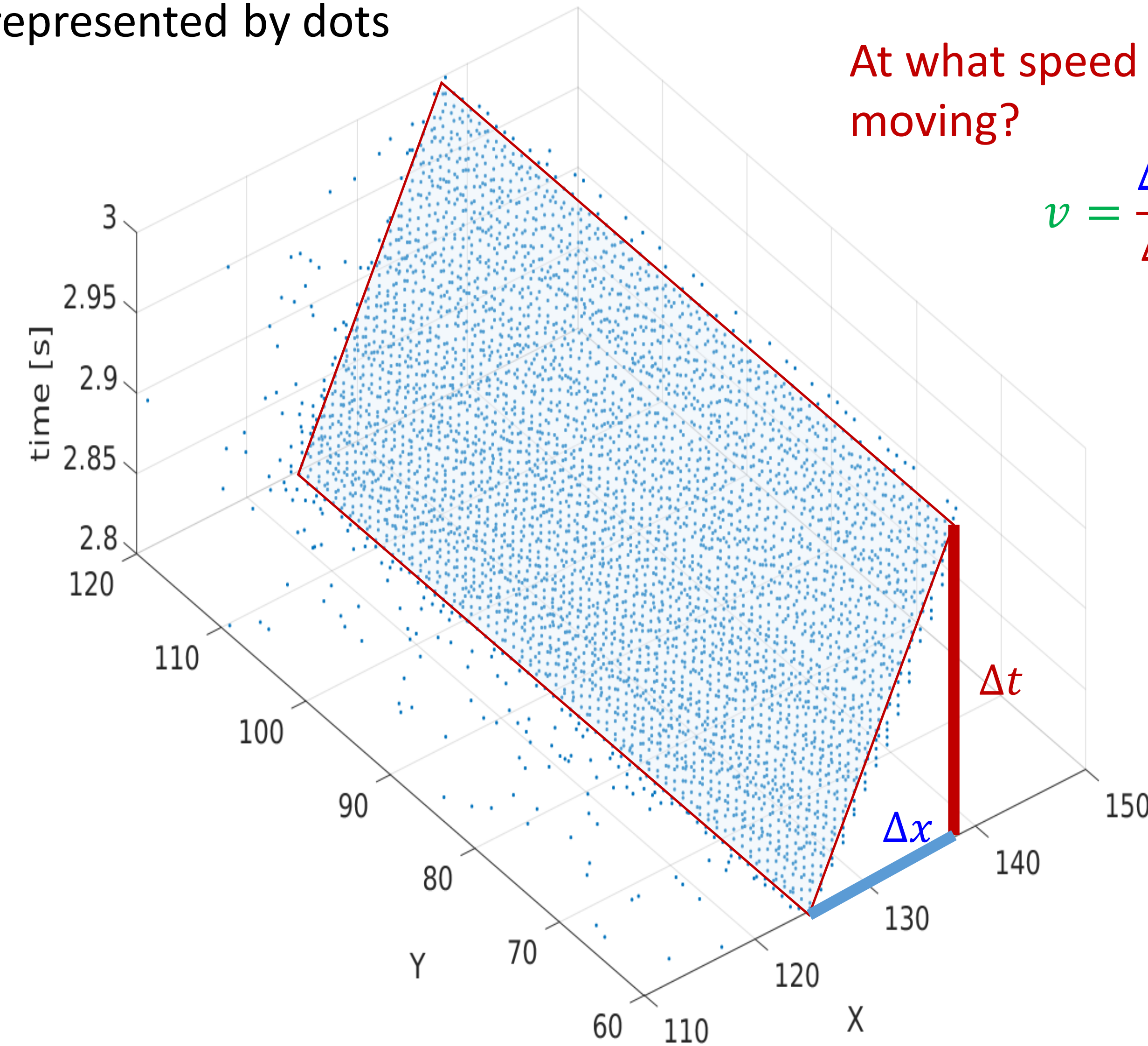


A moving edge

The same edge, visualized in space-time.
Events are represented by dots

At what speed is the edge moving?

$$v = \frac{\Delta x}{\Delta t}$$



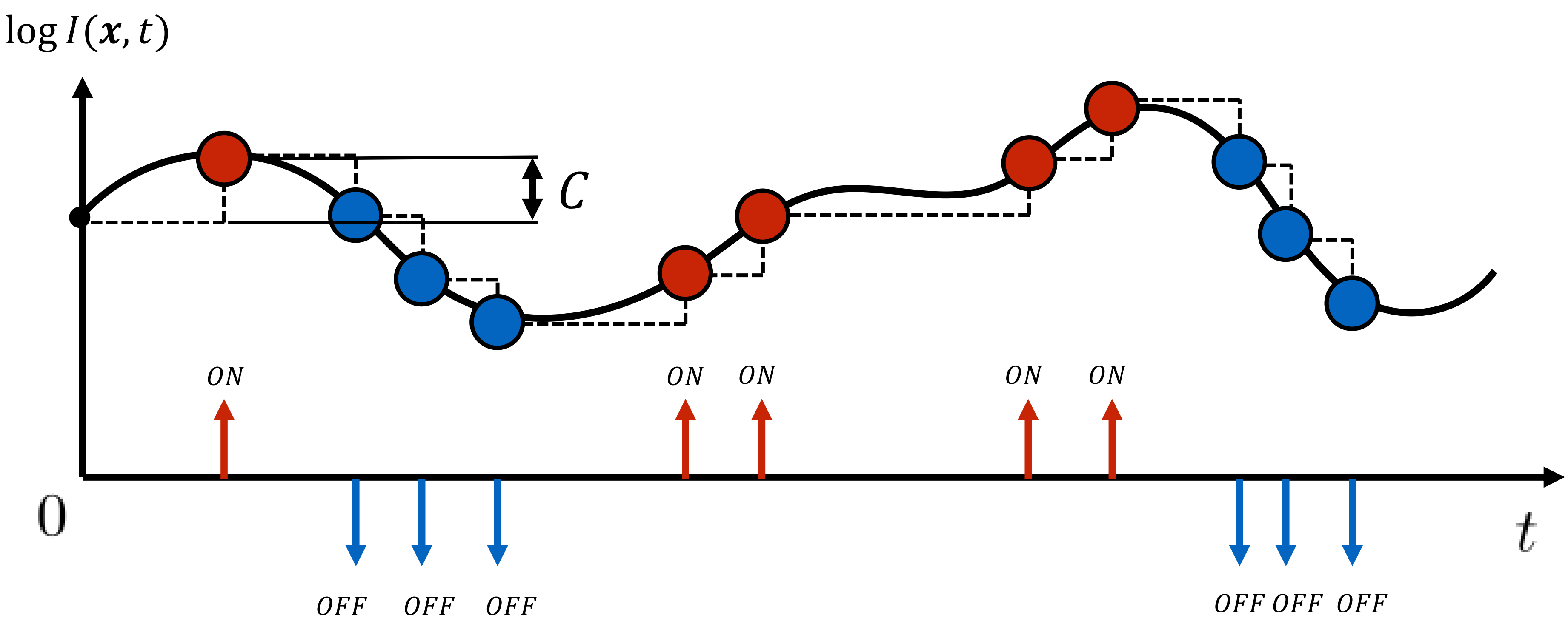
How do we unlock the outstanding potential of event cameras:

- Low latency
- High dynamic range
- No motion blur

Recall the Generative Event Model

An event is triggered at a **single pixel** if

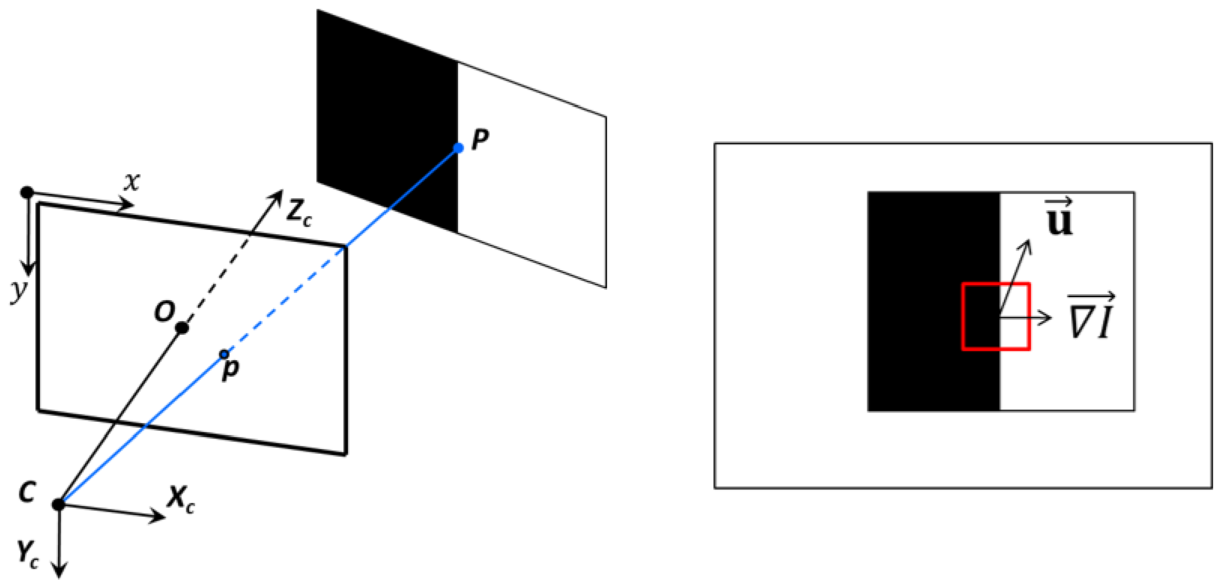
$$\log I(\boldsymbol{x}, t) - \log I(\boldsymbol{x}, t - \Delta t) = \pm C$$



1st Order Approximation

- Let us define $L(x, y, t) = Log(I(x, y, t))$
- Consider a given pixel $p(x, y)$ with gradient $\nabla L(x, y)$ undergoing the motion $\boldsymbol{u} = (u, v)$ in pixels, induced by a moving 3D point $\mathbf{P}$.
- Then, it can be shown that:

$$-\nabla L \cdot \mathbf{u} = C$$



Gallego et al., Event-based Vision: A Survey, arXiv, 2019. PDF

Proof

The proof comes from the ***brightness constancy assumption***, which says that the intensity value of p, before and after the motion, must remain unchanged:

$$L(x, y, t) = L(x + u, y + v, t + \Delta t)$$

By replacing the right-hand term by its 1st order approximation at $t + \Delta t$, we get:

$$L(x, y, t) = L(x, y, t + \Delta t) + \frac{\partial L}{\partial x} u + \frac{\partial L}{\partial y} v$$

$$\Rightarrow \; L(x, y, t + \Delta t) - L(x, y, t) = -\frac{\partial L}{\partial x} u - \frac{\partial L}{\partial y} v$$

$$\Rightarrow \; \Delta \mathrm{L} = \boxed{C = -\nabla L \cdot \mathbf{u}}$$

This equation describes the **linearized** event generation equation for an event generated by a gradient ∇L that moved by a motion vector $\mathbf{u}$ (optical flow) during a time interval Δt.

Application 1: Image Reconstruction from events

- Probabilistic simultaneous, gradient & rotation estimation from $C = -\nabla L \cdot \mathbf{u}$
- Obtain intensity from gradients via Poisson reconstruction
- The reconstructed image has super-resolution and high dynamic range (HDR)
- In real time on a GPU

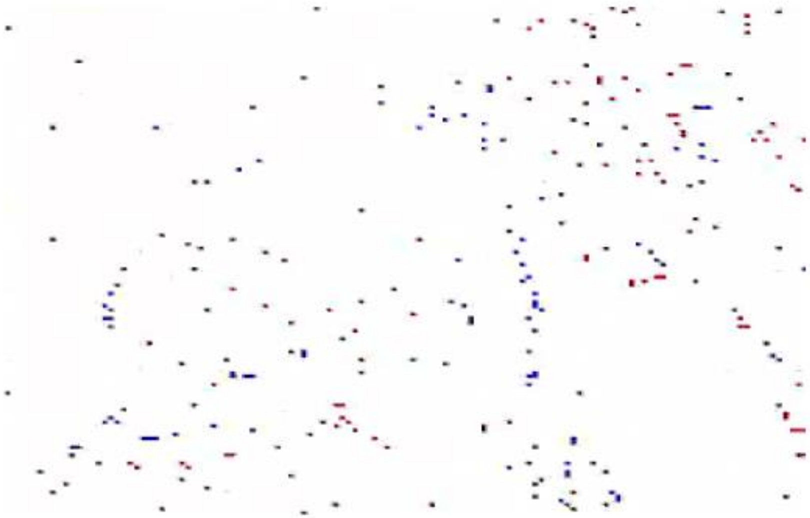
Kim et al., Simultaneous Mosaicing and Tracking with an Event Camera, BMVC’14

Application 2: 6DoF Tracking from Photometric Ma

- Probabilistic, motion estimation from $C = -\nabla L \cdot \mathbf{u}$
- Assumes photometric map (x,y,z, grayscale Intensity) is given
- Useful for VR/AR applications (low-latency, HDR, no motion blur)
- Requires GPU to run in real time

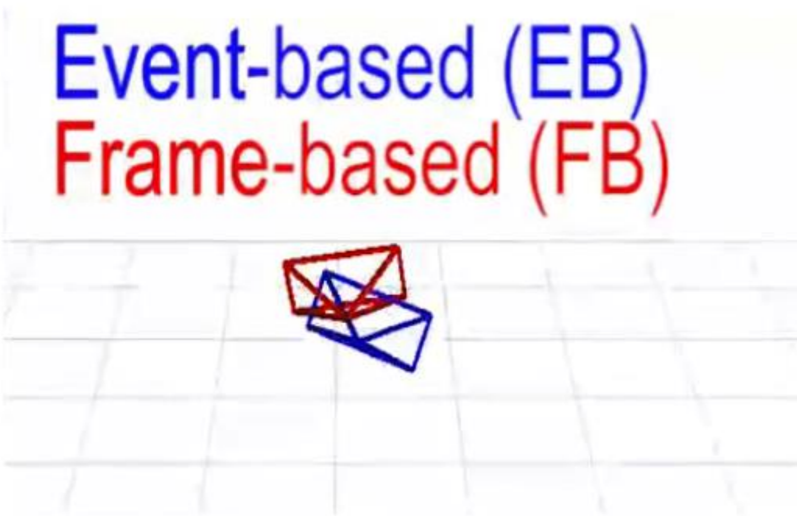
Gallego et al., Event-based 6-DOF Camera Tracking from Photometric Depth Maps, **T-PAMI**'18. [PDF]() [Video]()

Event camera



Standard camera

Motion estimation

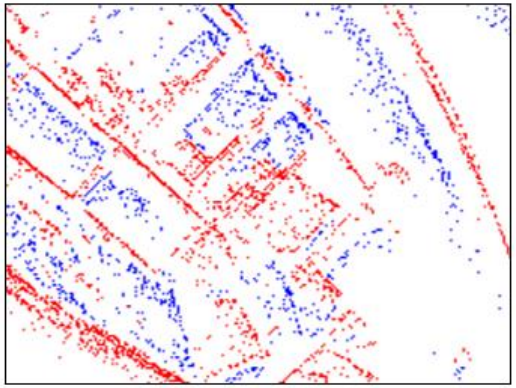


Gallego et al., Event-based 6-DOF Camera Tracking from Photometric Depth Maps, **T-PAMI**'18. [PDF](#) [Video](#)

What if we combined the complementary advantages of event and standard cameras?

Why combining them?

< 10 years research

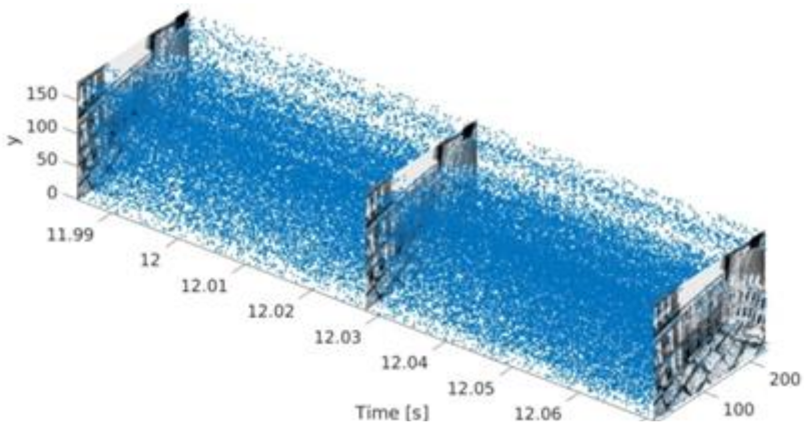


> 60 years of research!

	Event Camera	**Standard Camera**
Update rate	High (asynchronous): 1 MHz	Low (synchronous)
Dynamic Range	High (140 dB)	Low (60 dB)
Motion Blur	No	Yes
Static motion	No (event camera is a high pass filter)	Yes
Absolute intensity	No (but reconstructable up to a constant)	Yes
Maturity	< 10 years of research	> 60 years of research!

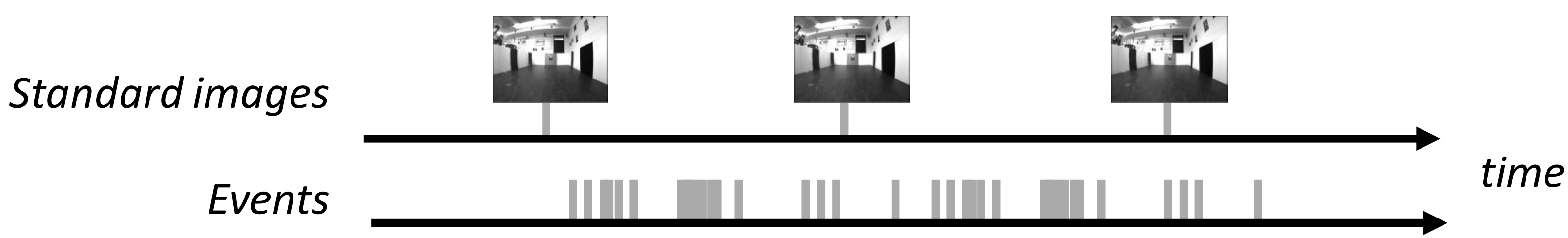
DAVIS sensor: Events + Images + IMU

- Combines an **event and a standard** camera in **the same pixel array** (→ the same pixel can both trigger events and integrate light intensity).
- **It also has an IMU**



Spatio-temporal visualization
of the output of a DAVIS sensor

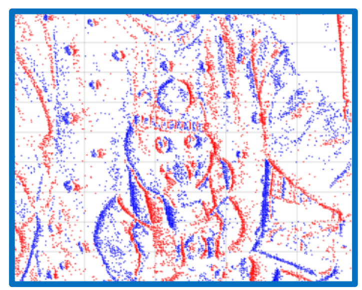
Temporal aggregation of events
overlaid on a DAVIS frame



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

Application 1: Deblurring a blurry video

- A **blurry image** can be regarded as the **integral of a sequence of *latent images*** during the exposure time, while the **events** indicate the **changes between the latent images**.
- **Finding**: sharp image obtained by subtracting the double integral of event from input image

$\log$  $- \iint$  $= \log$ 

Input blur image **Input events** **Output sharp image**

Pan et al., Bringing a Blurry Frame Alive at High Frame-Rate with an Event Camera, CVPR'19

Application 1: Deblurring a blurry video

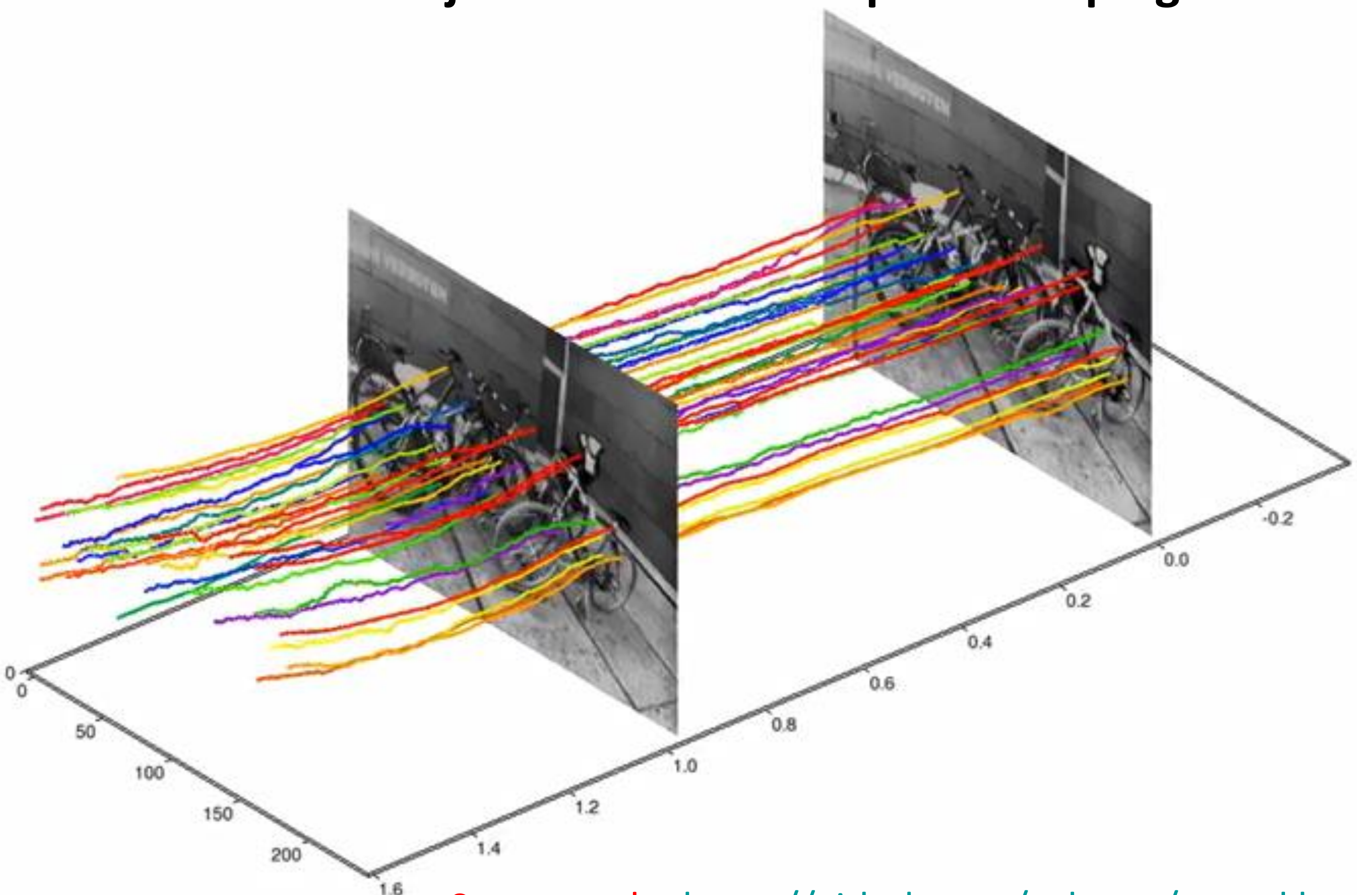
- A **blurry image** can be regarded as the **integral of a sequence of *latent images*** during the exposure time, while the **events** indicate the **changes between the latent images**.
- **Finding**: sharp image obtained by subtracting the double integral of event from input image

Input blur image **Output sharp video**

Pan et al., Bringing a Blurry Frame Alive at High Frame-Rate with an Event Camera, CVPR’19

Application 3: Lucas-Kanade Tracking using Events and Frames

- **Goal**: Extract features from **standard frames** and track them using only **events** in the **blind time** between two **frames**
- Uses the event generation model via **joint estimation of patch warping and optic flow**



Source code: https://github.com/uzh-rpg/rpg_eklt

Gehrig et al., EKLT: Asynchronous, Photometric Feature Tracking using Events and Frames, IJCV 2019.
PDF, YouTube, Evaluation Code, Tracking Code

Recap

- All the approaches seen so use the **generative event model**

$$\pm C = \log I(\boldsymbol{x}, t) - \log I(\boldsymbol{x}, t - \Delta t)$$

  or its 1st order approximation

$$\pm C = -\nabla L \cdot \mathbf{u} \quad ,$$

  which **requires knowledge of the contrast sensitivity C.**

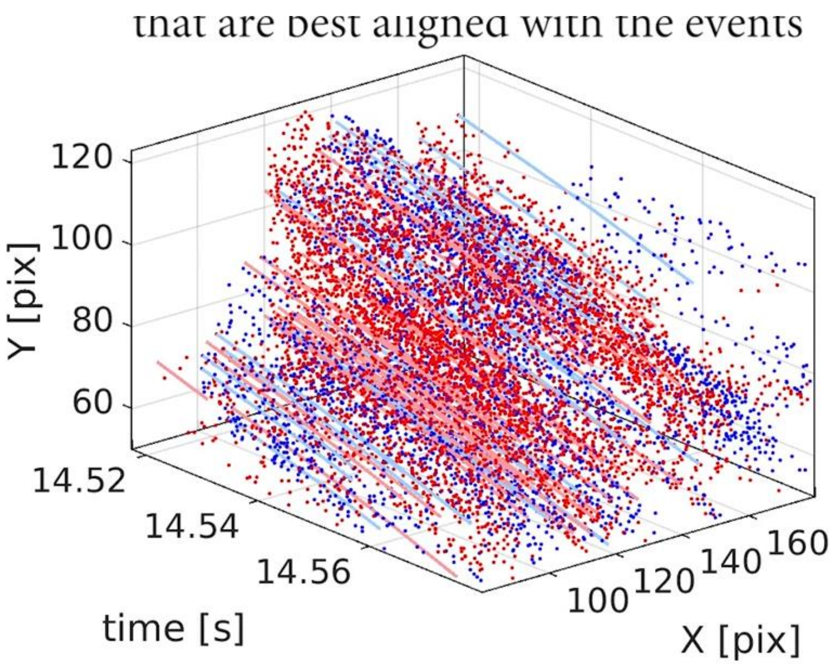
- Unfortunately, **C is scene dependent** and might **differ from pixel to pixel**.
- **Alternative approach: Focus maximization framework**

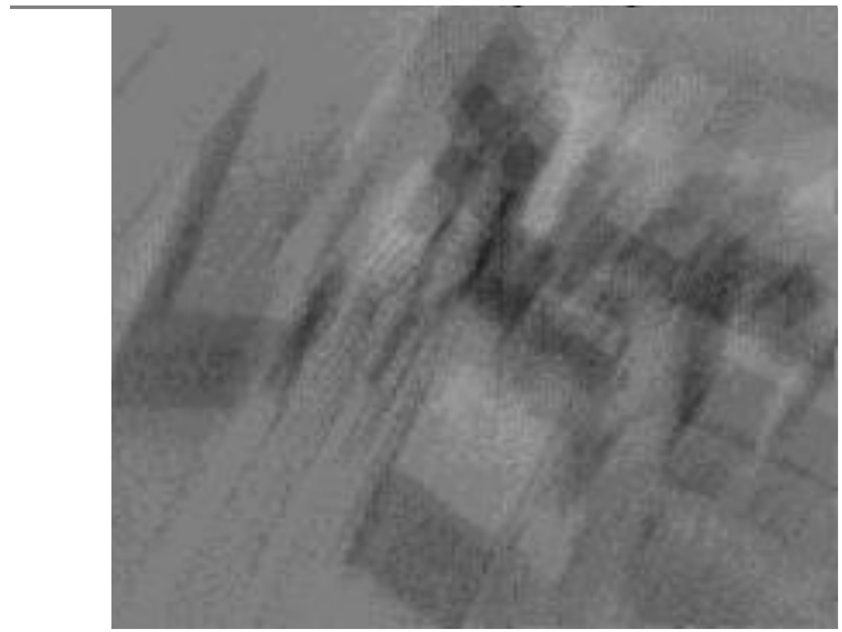
Focus Maximization for:

- Motion estimation
- 3D reconstruction
- SLAM
- Optical flow estimation
- Feature tracking
- Motion segmentation
- Unsupervised learning

Gallego et al., A Unifying Contrast Maximization Framework for Event Cameras, CVPR18, PDF, YouTube
Gallego et al., Focus Is All You Need: Loss Functions for Event-based Vision, CVPR19, PDF.

Idea: Warp spatio-temporal volume of events to **maximize focus** (e.g., sharpness) of the resulting image

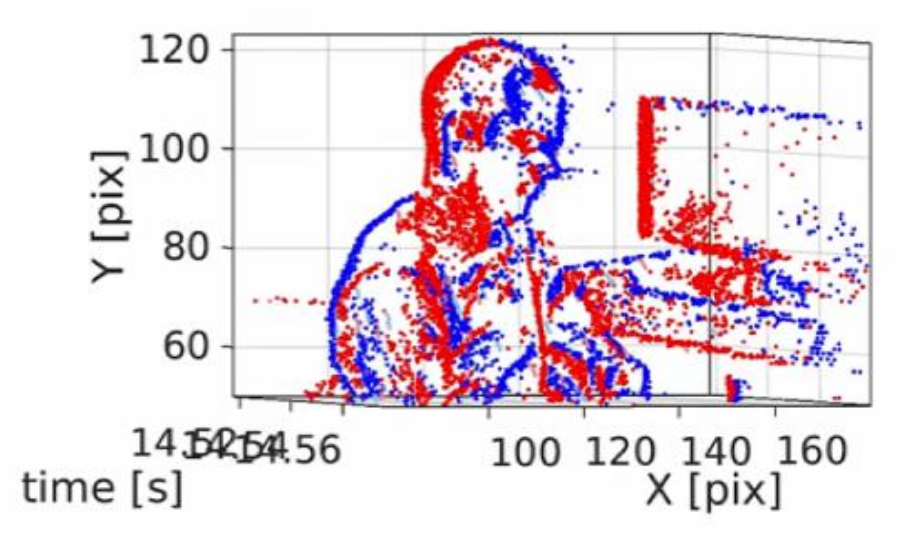


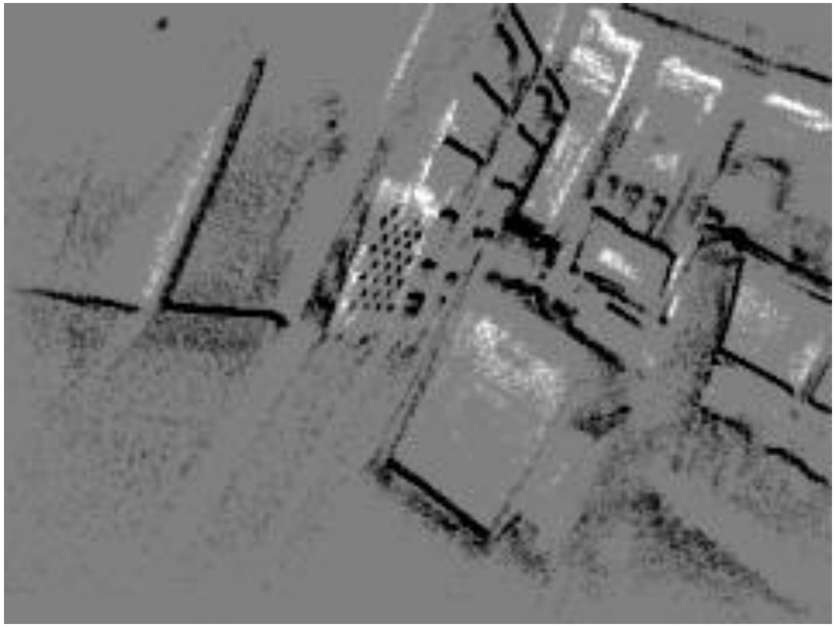


Aggregated image **without** motion correction

Gallego et al., A Unifying Contrast Maximization Framework for Event Cameras, CVPR18, PDF, YouTube
Gallego et al., Focus Is All You Need: Loss Functions for Event-based Vision, CVPR19, PDF.

Idea: Warp spatio-temporal volume of events to **maximize focus** (e.g., sharpness) of the resulting image

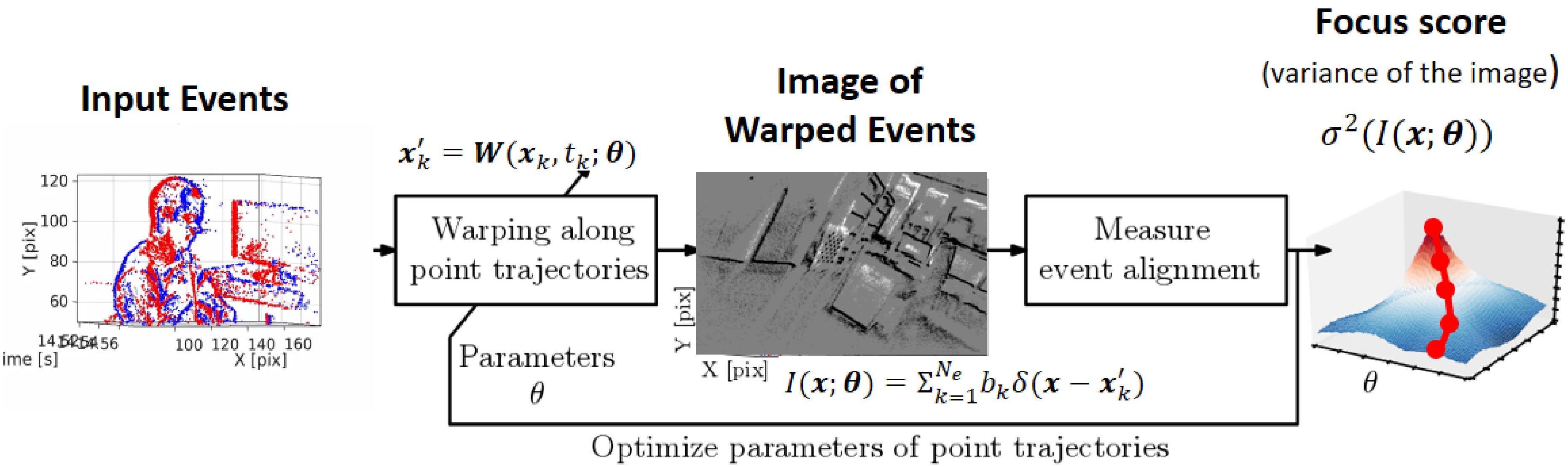




Aggregated image
with motion correction

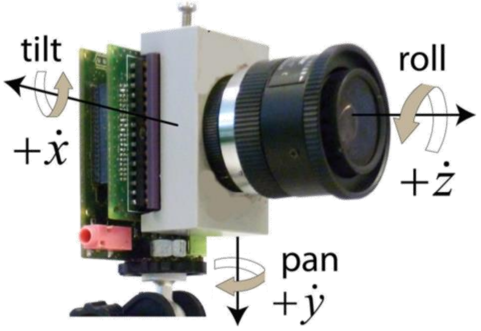
Gallego et al., A Unifying Contrast Maximization Framework for Event Cameras, CVPR18, PDF, YouTube
Gallego et al., Focus Is All You Need: Loss Functions for Event-based Vision, CVPR19, PDF.

Focus Maximization Framework

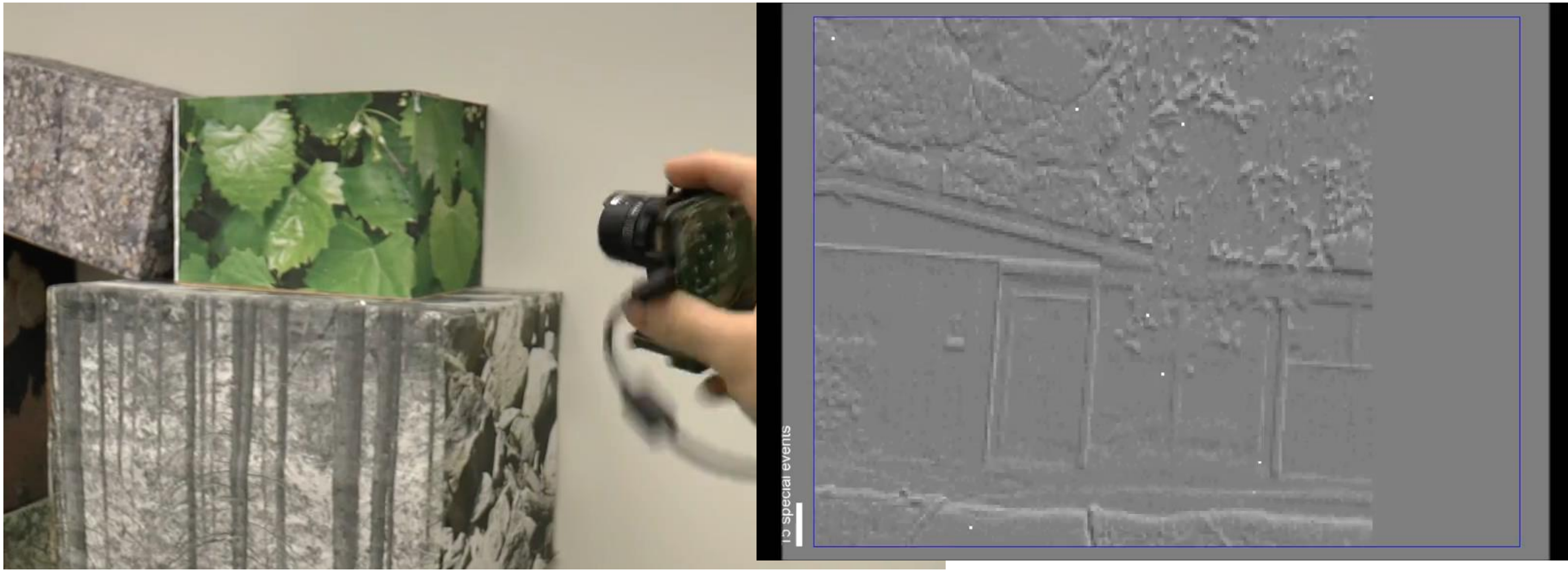


- $\boldsymbol{x}'_k = \boldsymbol{W}(\boldsymbol{x}_k, t_k; \boldsymbol{\theta})$: This warps the (x, y) pixels coordinates of each event, not their time. Possible warps: roto-translation, affine, homography.
- $I(\boldsymbol{x}; \boldsymbol{\theta}) = \Sigma_{k=1}^{N_e} b_k \delta(\boldsymbol{x} - \boldsymbol{x}'_k)$: This builds a grayscale image, where the intensity of each pixel at the warped location (x', y') is equal to the summation of the positive and negative events (+1, -1)
- $\sigma^2(I(\boldsymbol{x}; \boldsymbol{\theta}))$: The assumption here is that if an image contains ***high variance*** then there is a wide **spread of responses, both edge-like and non-edge like**, representative of a normal, in-focus image. But if there is ***very low variance***, then there is a tiny spread of responses, indicating there are very little edges in the image. As we know, the more an image is blurred, *the less edges there are*.

Application 1: Image Stabilization

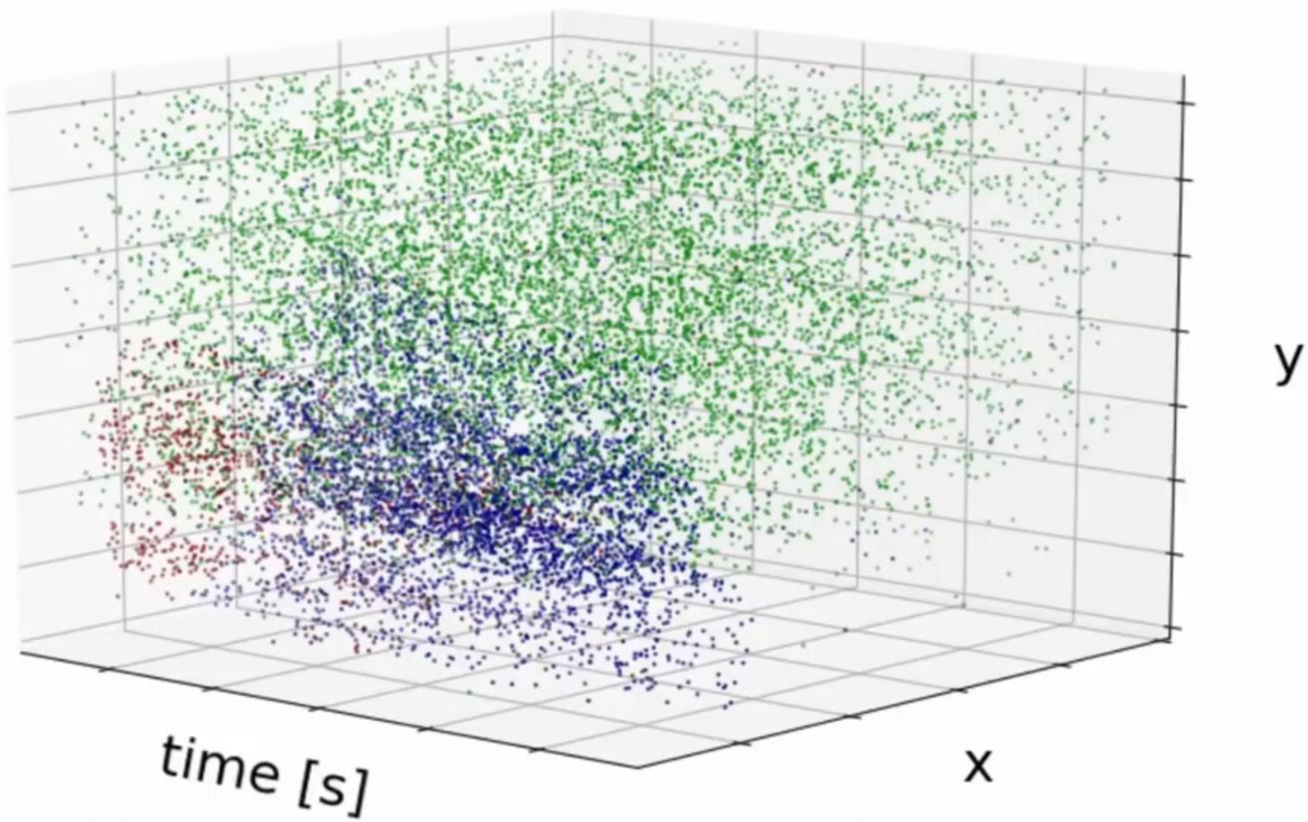


- Problem: Estimate rotational motion (3DoF) of an event camera
- Can process millions of events per second in real time on a smartphone CPU (OdroidXU4)
- Works up to over ~1,000 deg/s



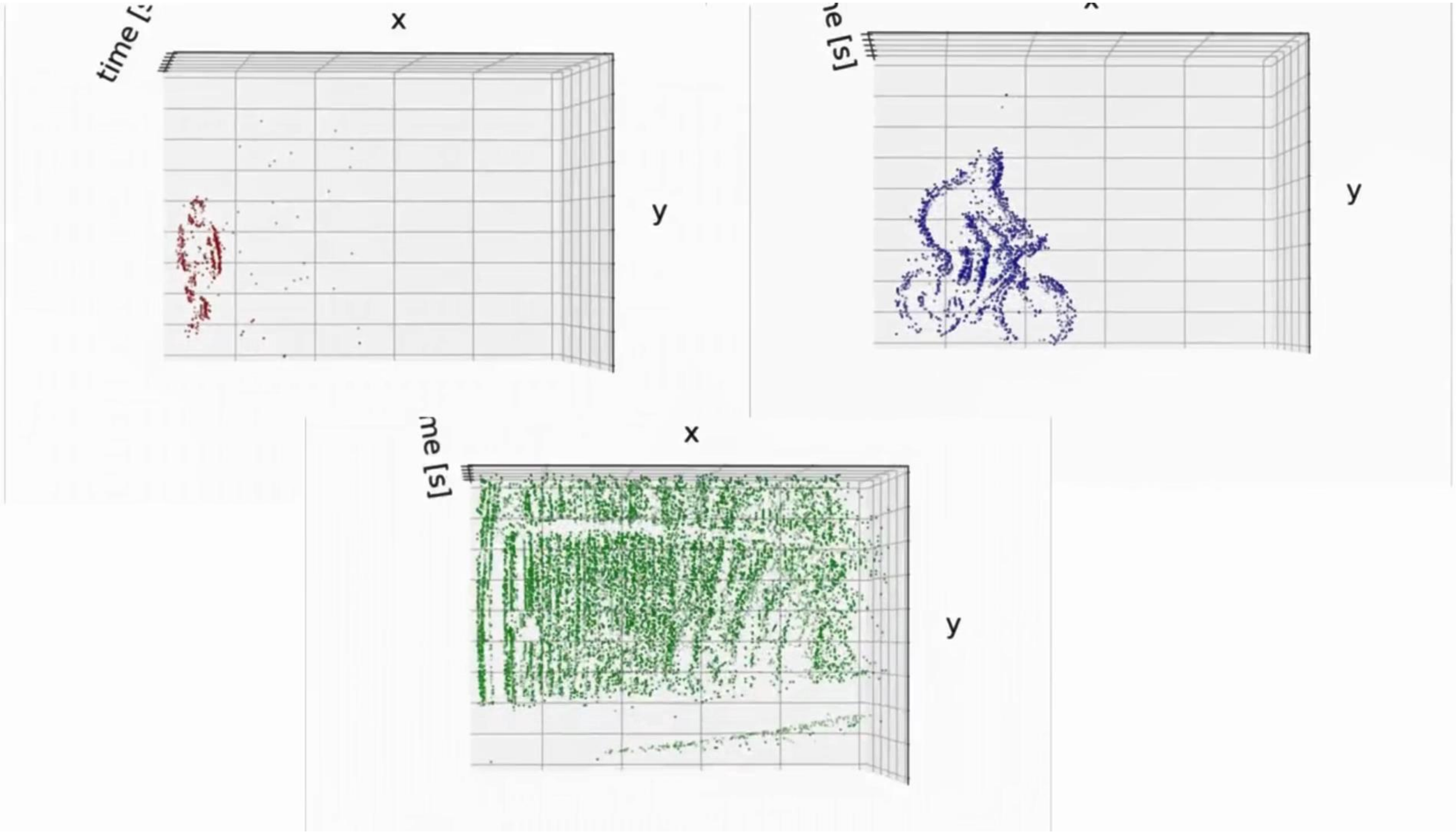
Gallego et al., Accurate Angular Velocity Estimation with an Event Camera, IEEE RAL'16. [PDF](). [Video]().

Application 2: Motion Segmentation



Stoffregen et al., Motion Segmentation by Motion Compensation, ICCV’19. [PDF](). [Video]().

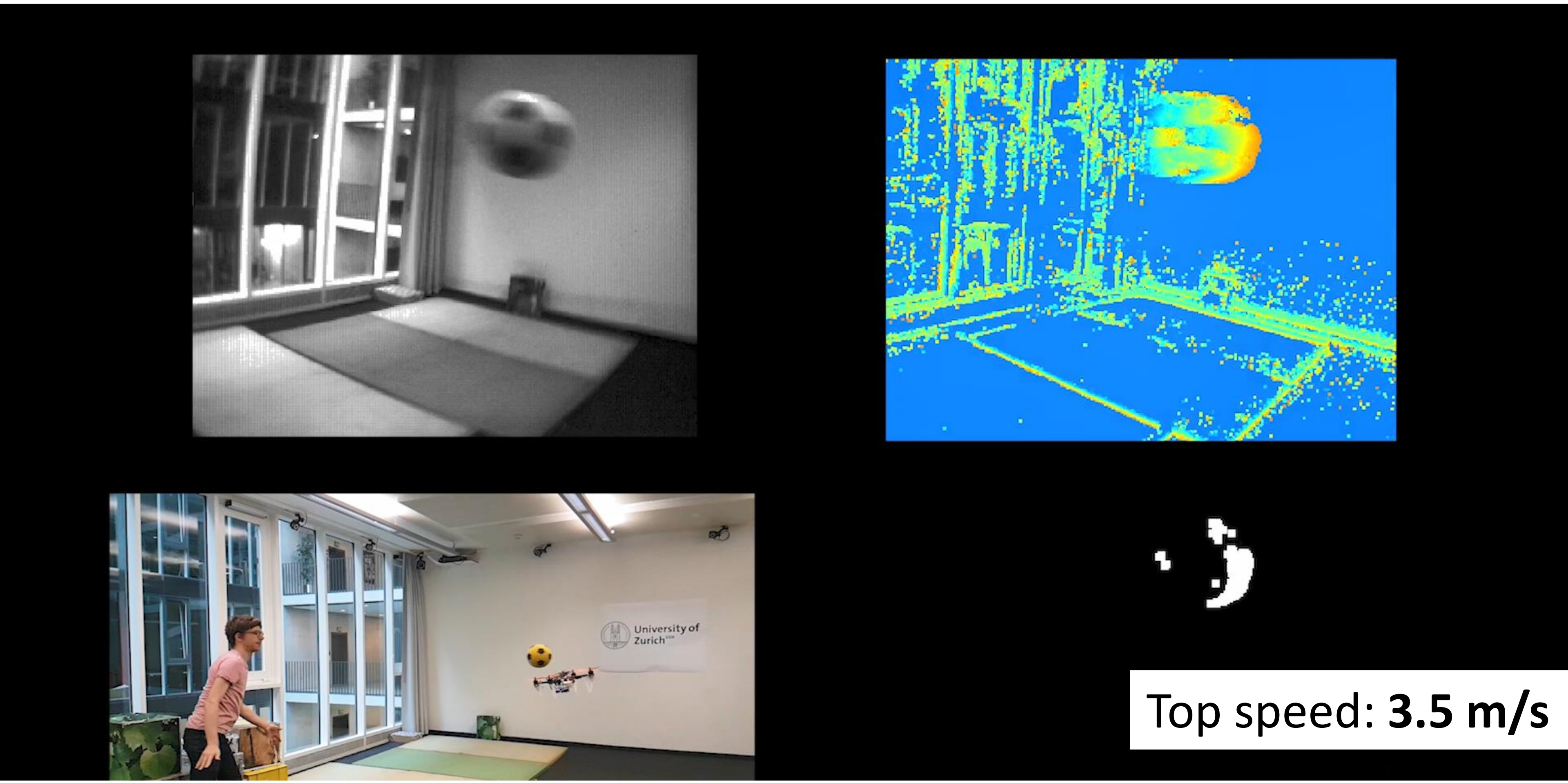
Application 2: Motion Segmentation



Stoffregen et al., Motion Segmentation by Motion Compensation, ICCV'19. PDF. Video.

Application 3: Dynamic Obstacle Detection & Avoidance

- Top speed: **3.5 m/s**
- Object detection runs at 100Hz onboard



Falanga et al. *How Fast is too fast? The role of perception latency in high speed sense and avoid*, RAL'19.
PDF. Video. Featured in IEEE Spectrum.

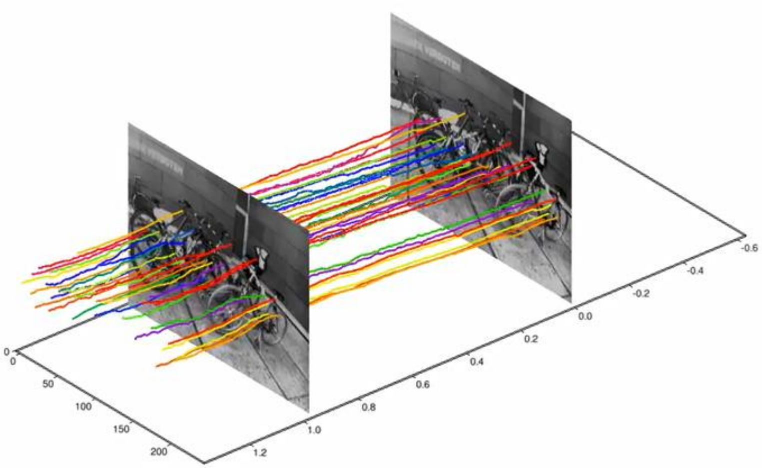
Application 4:
UltimateSLAM:

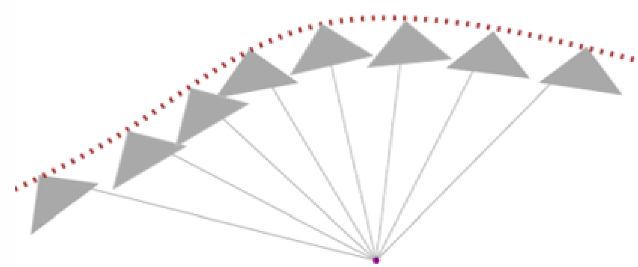
combining **events**, **images**, and **IMU** for robust visual SLAM in HDR and High Speed Scenarios

Application 4: UltimateSLAM: combining Events + Frames + IMU

Front End:
Feature tracking from
Events and Frames

Back-End
State-of-the-art
Sliding-Window
Visual-inertial Fusion

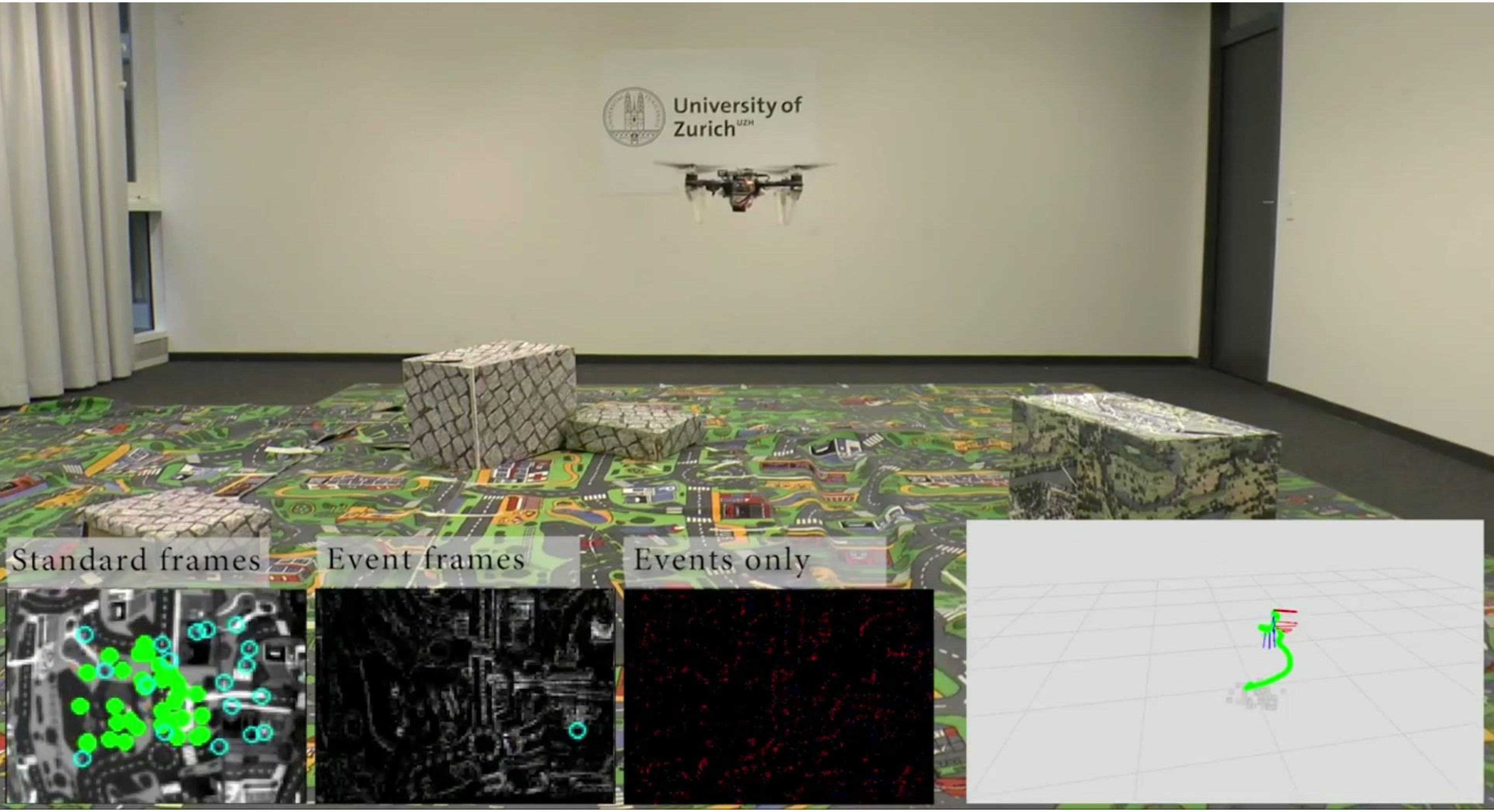




Rosinol et al., Ultimate SLAM? **RAL'18 – Best RAL'18 Paper Award Honorable Mention** PDF. Video. IEEE Spectrum.

Application: Autonomous Drone Navigation in Low Light

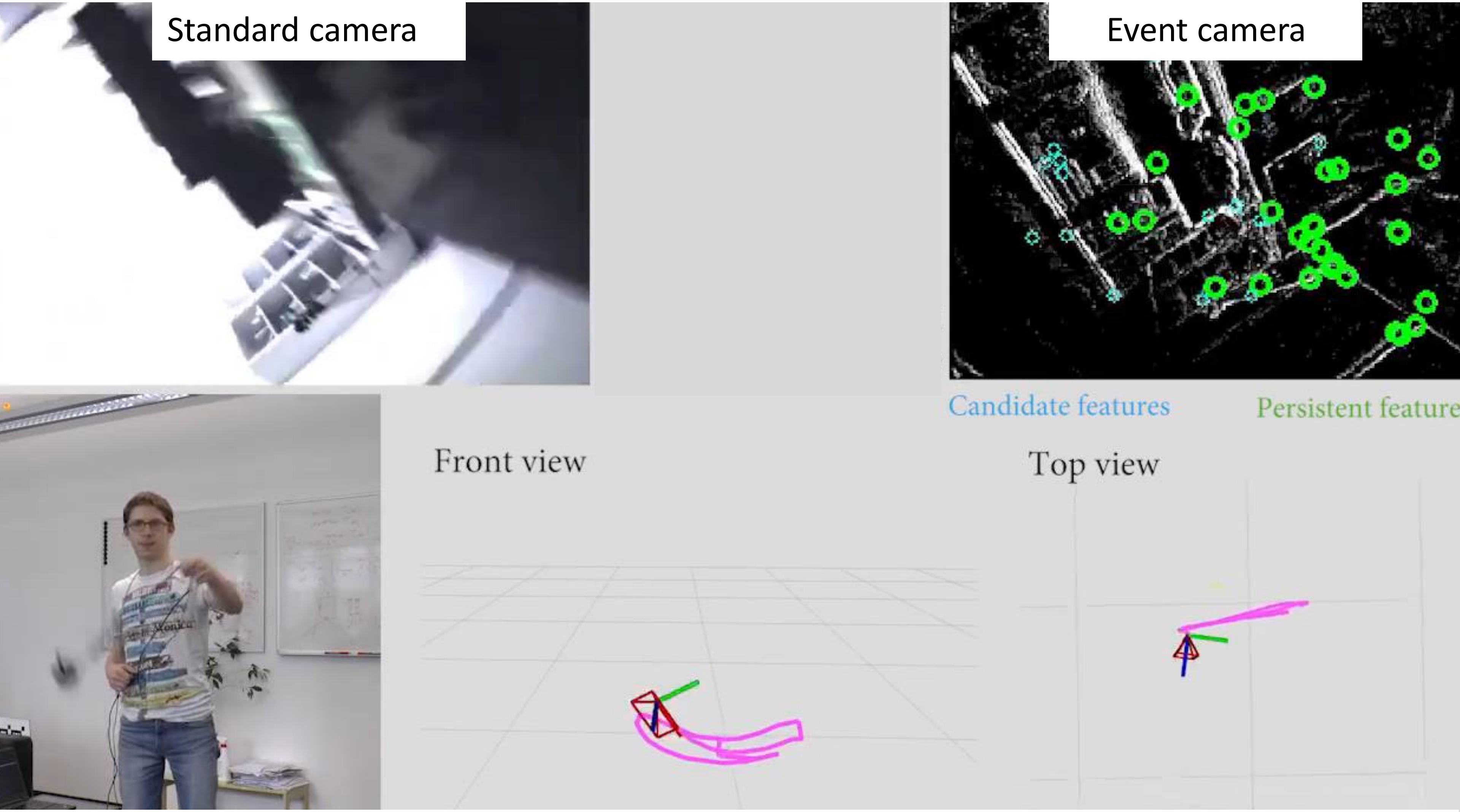
UltimateSLAM running on board (CPU: Odroid XU4)



Rosinol et al., Ultimate SLAM? **RAL'18 – Best RAL'18 Paper Award Honorable Mention** PDF. Video. IEEE Spectrum.
Mueggler et al., Continuous-Time Visual-Inertial Odometry for Event Cameras, **TRO'18**. PDF

UltimateSLAM: Frames + Events + IMU

85% accuracy gain over standard Visual-Inertial SLAM in HDR and high speed scenes



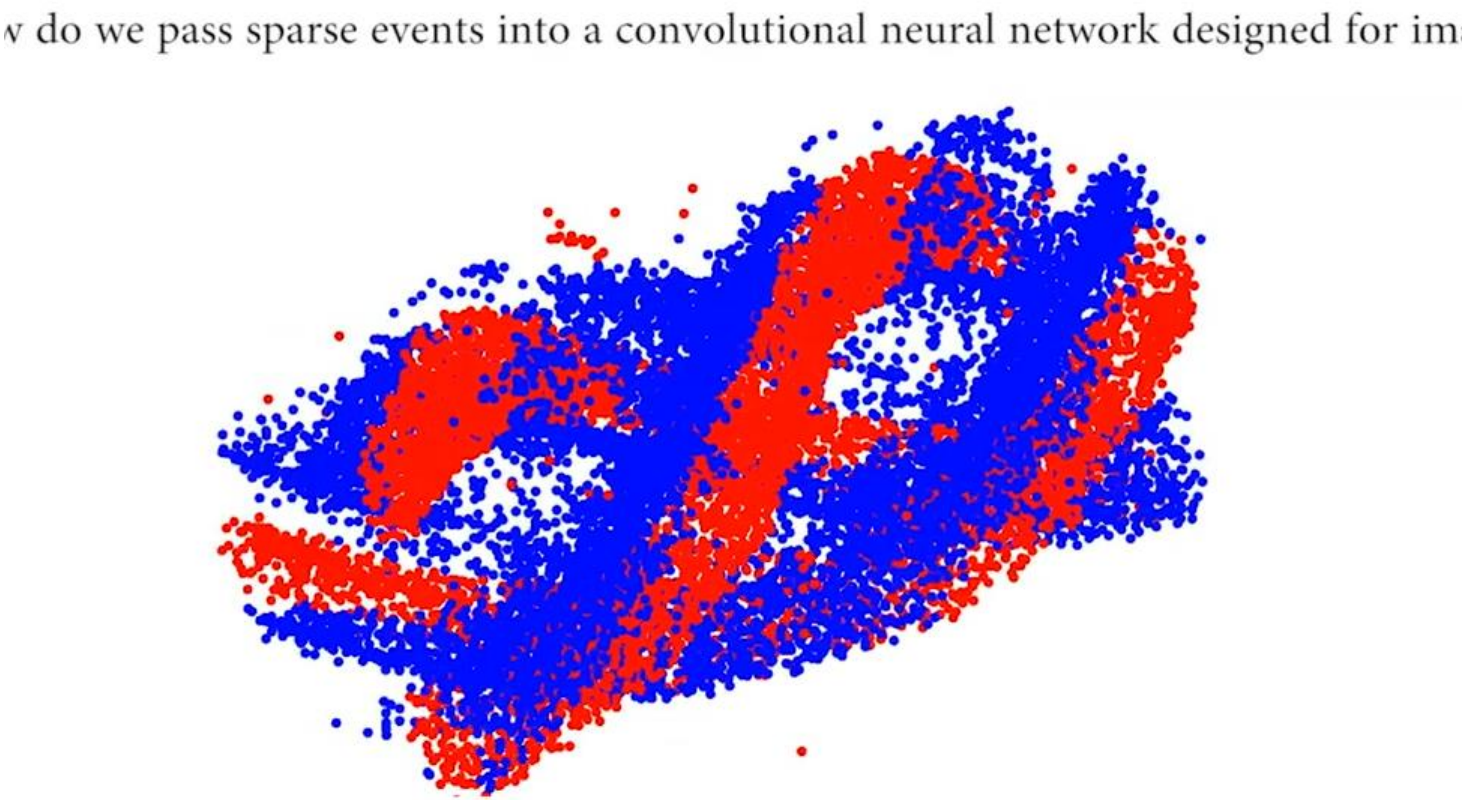
Rosinol et al., Ultimate SLAM? **RAL'18 – Best RAL'18 Paper Award Honorable Mention** PDF. Video. IEEE Spectrum.
Mueggler et al., Continuous-Time Visual-Inertial Odometry for Event Cameras, **TRO'18**. PDF

Learning with Event Cameras

- Approaches using synchronous, Artificial Neural Networks (ANNs) designed for standard images
- Approaches using asynchronous, Spiking neural networks (SNNs)

Input representation

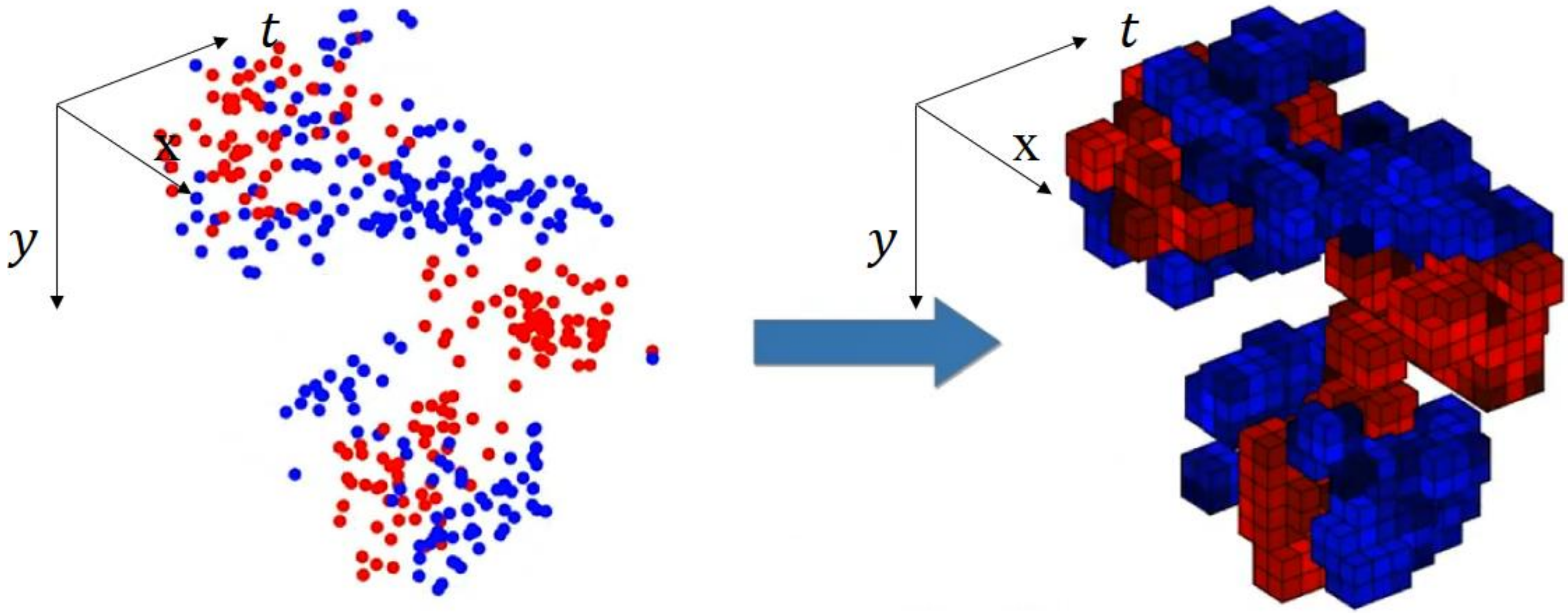
How do we pass sparse events into a convolutional neural network designed for standard images?



Video from Zhu et al. (link)

Input representation

- Represent events in space-time into a 3D voxel grid (x, y, t): each voxel contains sum of positive and negative events falling within the voxel (events are inserted into the volume with trilinear interpolation, resulting in minimal loss in resolution

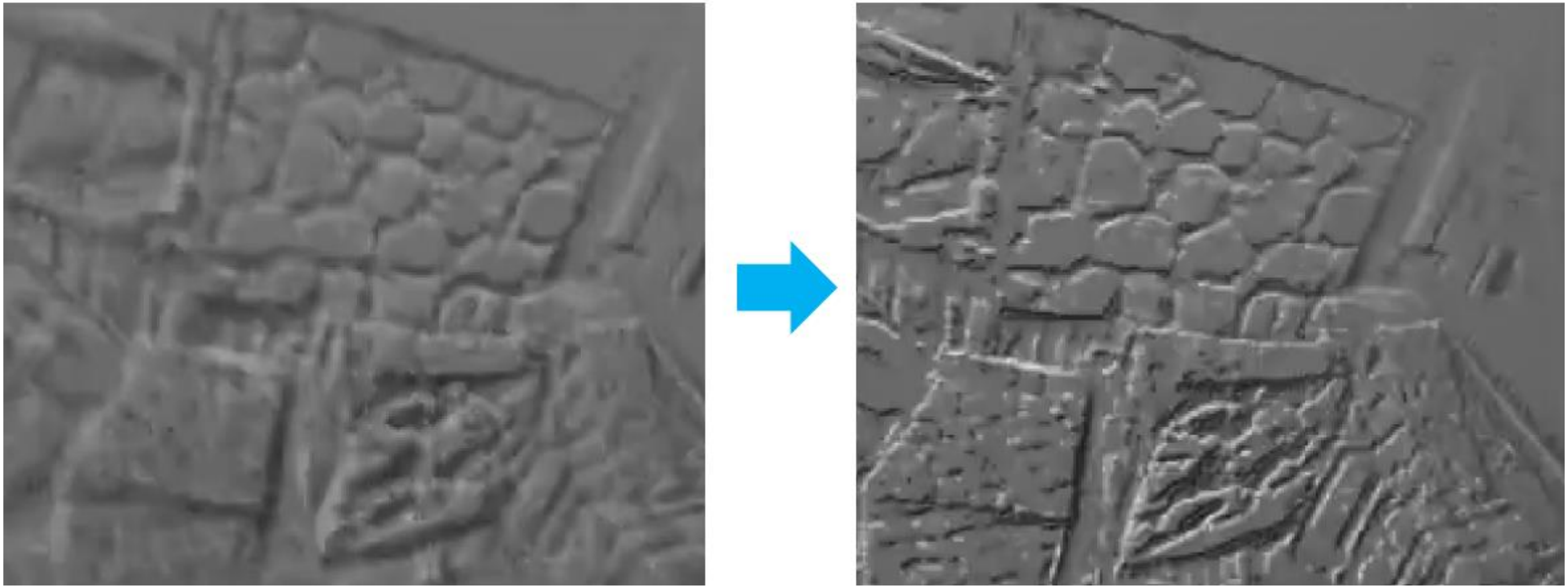


Video from [Zhu et all, CVPR'19]

[Zhu, ECCVW'18], [Zhu, CVPR'19], [Gehrig, ICCV'19], [Rebecq, CVPR'19]

Focus as Loss Function for Unsupervised Learning

Focus used as loss: maximize sharpness of the aggregated event image.

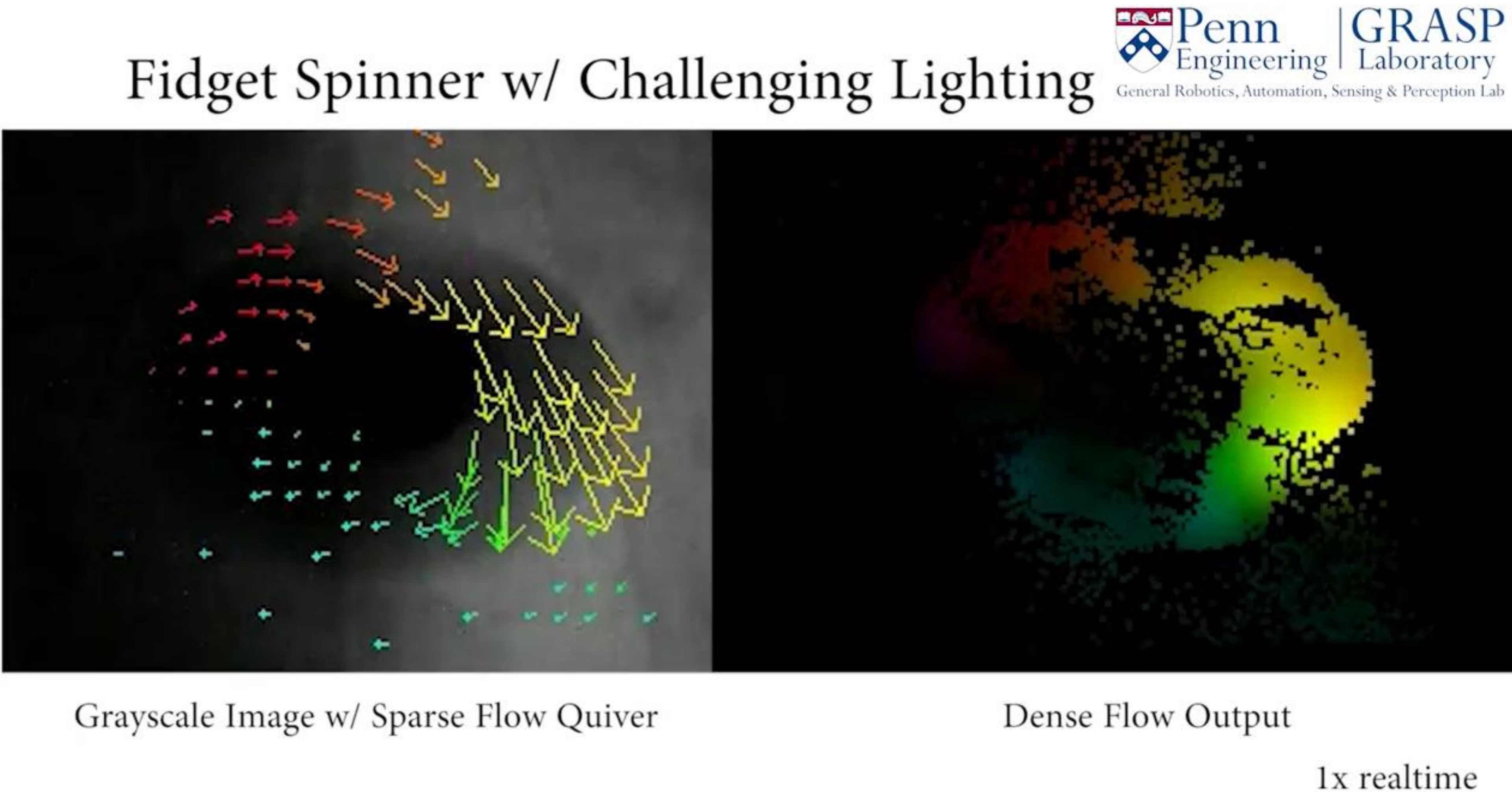


[Video from here]

Zhu, Unsupervised Event-based Learning of Optical Flow, Depth and Egomotion, CVPR 19
Gallego et al., Focus Is All You Need: Loss Functions for Event-based Vision, CVPR19, [PDF].

Application1: Unsupervised Learning of Optical Flow

Focus used as loss: maximize sharpness of the aggregated event image.



Zhu et al., Unsupervised Learning of Optical Flow, Depth and Ego Motion, CVPR'19

Application2: Learning High-speed and HDR Video Rendering from an Event Camera

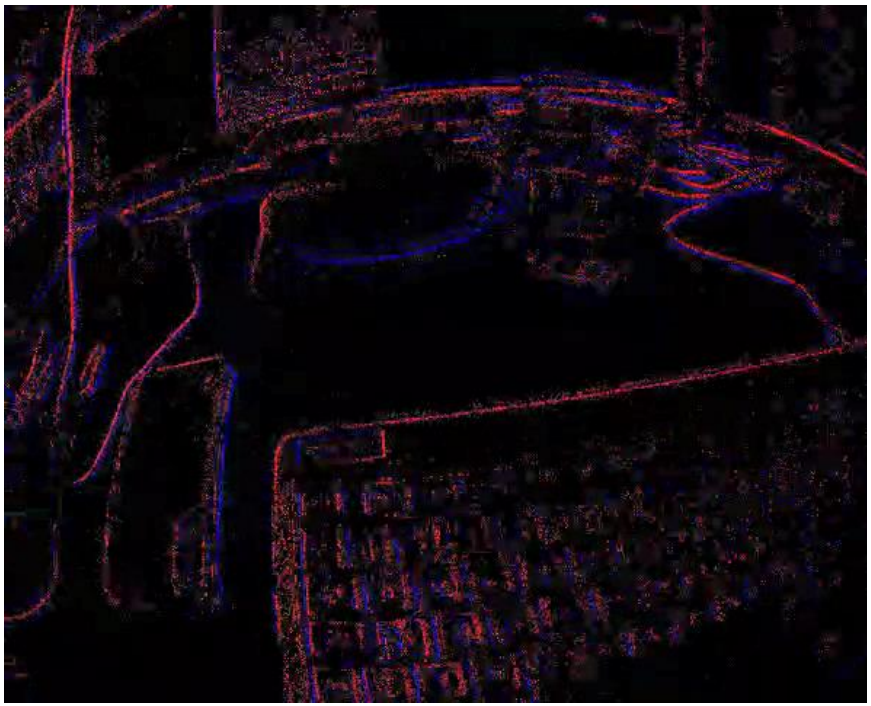
Code & datasets: https://github.com/uzh-rpg/rpg_e2vid

Code & datasets: https://github.com/uzh-rpg/rpg_e2vid

Rebecq et al., "Events-to-Video: Bringing Modern Computer Vision to Event Cameras", CVPR19. PDF Video.
Rebecq et al., "High Speed and High Dynamic Range Video with an Event Camera", PAMI, 2019. PDF Video Code

Image Reconstruction from Events

Events



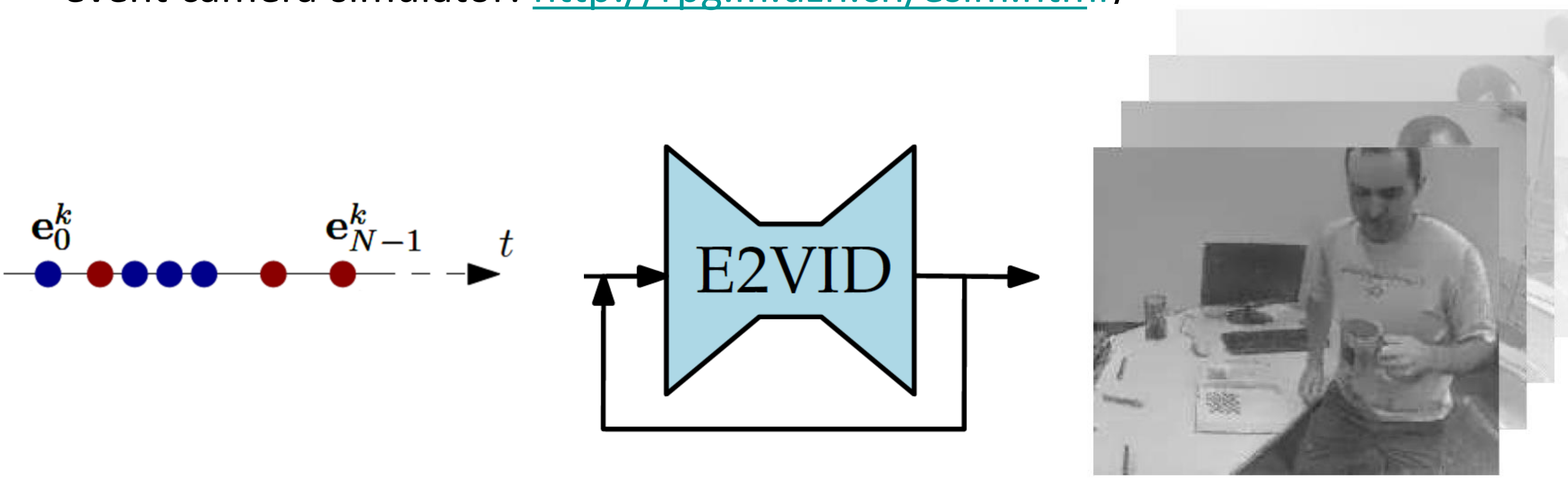
Reconstructed image from events (Samsung DVS)

Code & datasets: https://github.com/uzh-rpg/rpg_e2vid

Rebecq et al., “Events-to-Video: Bringing Modern Computer Vision to Event Cameras”, CVPR19. PDF Video.
Rebecq et al., “High Speed and High Dynamic Range Video with an Event Camera”, PAMI, 2019. PDF Video Code

Overview

- **Recurrent neural network** (main module: Unet)
- Input: last reconstructed frame + **sequences of *event tensors*** (spatio-temporal 3D voxels grid: each voxel contains sum of ON and OFF events falling within the voxel)
- Network processes **last N events** (10,000)
- **Trained in simulation only** (without seeing a single real image) (we used our event camera simulator: http://rpg.ifi.uzh.ch/esim.html)

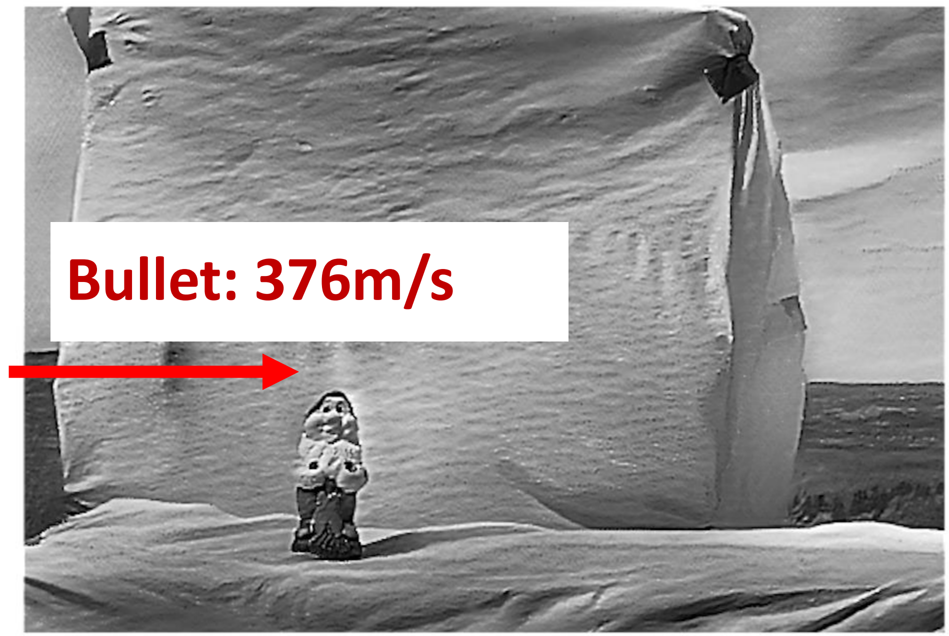


Rebecq et al., "Events-to-Video: Bringing Modern Computer Vision to Event Cameras", CVPR19. PDF Video.
Rebecq et al., "High Speed and High Dynamic Range Video with an Event Camera", PAMI, 2019. PDF Video Code

Bullet shot by a gun (376m/s (=1,354km/h)

Recall: trained in simulation only!

Huawei P20 Pro (240 FPS)



Our reconstruction (5400 FPS)

We used Samsung DVS

Code & datasets: https://github.com/uzh-rpg/rpg_e2vid

Real time

Rebecq et al., "Events-to-Video: Bringing Modern Computer Vision to Event Cameras", CVPR19. PDF Video.
Rebecq et al., "High Speed and High Dynamic Range Video with an Event Camera", PAMI, 2019. PDF Video Code

Bullet shot by a gun (1,300 km/h)

Recall: trained in simulation only!

Huawei P20 Pro (240 FPS)

Our reconstruction (5400 FPS)
We used Samsung DVS

Code & datasets: https://github.com/uzh-rpg/rpg_e2vid

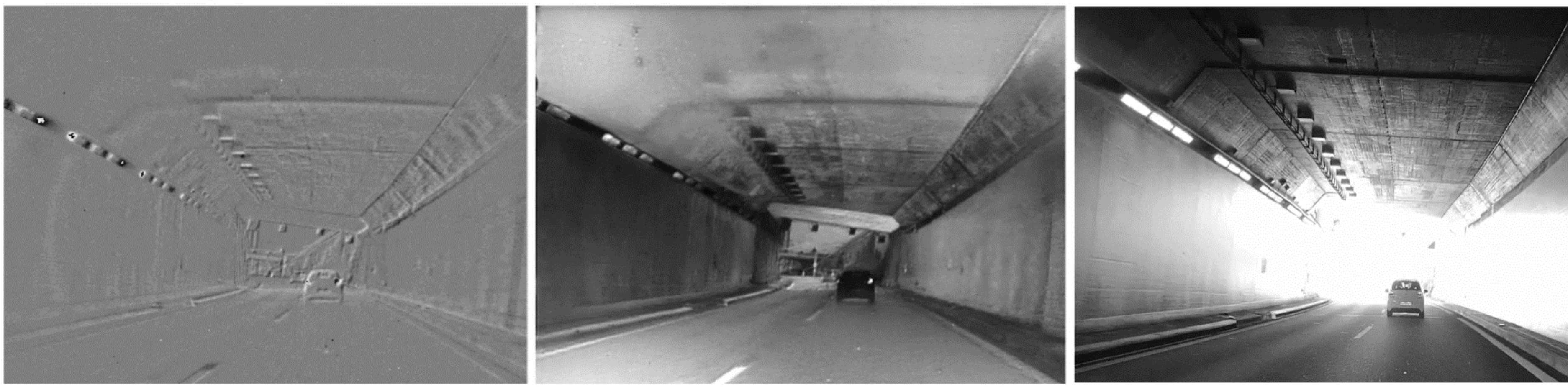
100 x slow motion

Rebecq et al., "Events-to-Video: Bringing Modern Computer Vision to Event Cameras", CVPR19. PDF Video.
Rebecq et al., "High Speed and High Dynamic Range Video with an Event Camera", PAMI, 2019. PDF Video Code

HDR Video: Driving out of a tunnel

Recall: trained in simulation only!

Driving out of a tunnel



Events | **Our reconstruction** | **Phone camera**

Code & datasets: https://github.com/uzh-rpg/rpg_e2vid

Rebecq et al., “Events-to-Video: Bringing Modern Computer Vision to Event Cameras”, CVPR19. PDF Video.

Rebecq et al., “High Speed and High Dynamic Range Video with an Event Camera”, PAMI, 2019. PDF Video Code

HDR Video: Night Drive

Recall: trained in simulation only!

Our reconstruction from events

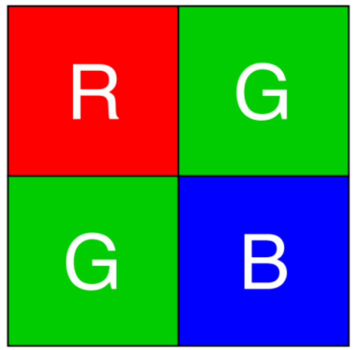
GoPro Hero 6

Code & datasets: https://github.com/uzh-rpg/rpg_e2vid

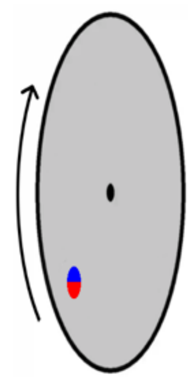
Rebecq et al., "Events-to-Video: Bringing Modern Computer Vision to Event Cameras", CVPR19. PDF Video.
Rebecq et al., "High Speed and High Dynamic Range Video with an Event Camera", PAMI, 2019. PDF Video Code

Color Event Camera

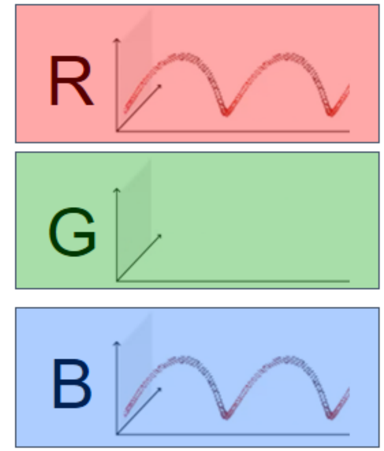
DAVIS346 Red Color



Bayer pattern



Input

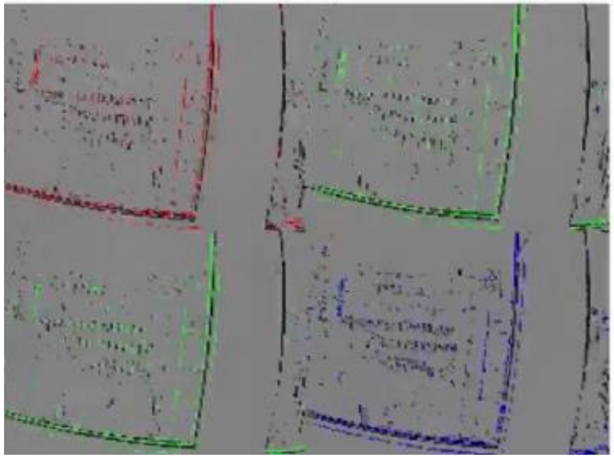


Output

- Each pixel is sensitive to either **red, green or blue** light.
- Transmits **brightness changes** in each color channel

Taverni et al., Front and back illuminated Dynamic and Active Pixel Vision Sensors comparison, TCS'18

Color Event Camera Reconstruction (HDR)



Color events

Our reconstruction

Color frame

Color Event Camera Datasets: http://rpg.ifi.uzh.ch/CED.html

Scheerlinck, Rebecq, Stoffregen, Barnes, Mahony, Scaramuzza
CED: Color Event Camera Dataset. CVPRW, 2019. PDF YouTube Dataset

Conclusions

- ➢ Visual Inertial 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
  - Active sensor parameter control (on-the-fly tuning)
- ➢ **Event cameras** are revolutionary and provide:
  - Very **low latency** (1 μs) and **robustness** to **high speed motion** and **high-dynamic-range scenes**
  - Standard cameras studied for 50 years
    - event cameras offer have plenty of room for research
  - **Open problems on event cameras**: noise modeling, asynchronous feature and object detection and tracking, sensor fusion, asynchronous learning & recognition, low latency estimation and control, low power computation

Understanding Check

Are you able to answer the following questions?

- What is a DVS and how does it work?
- What are its pros and cons vs. standard cameras?
- Can we apply standard camera calibration techniques?
- How can we compute optical flow with a DVS?
- Could you intuitively explain why we can reconstruct the intensity?
- What is the generative model of a DVS and how to derive it?
- What is a DAVIS sensor?
- What is the focus maximization framework and how does it work? What is its advantage compared with the generative model?
- How can we get color events?