Vision Algorithms for Mobile Robotics

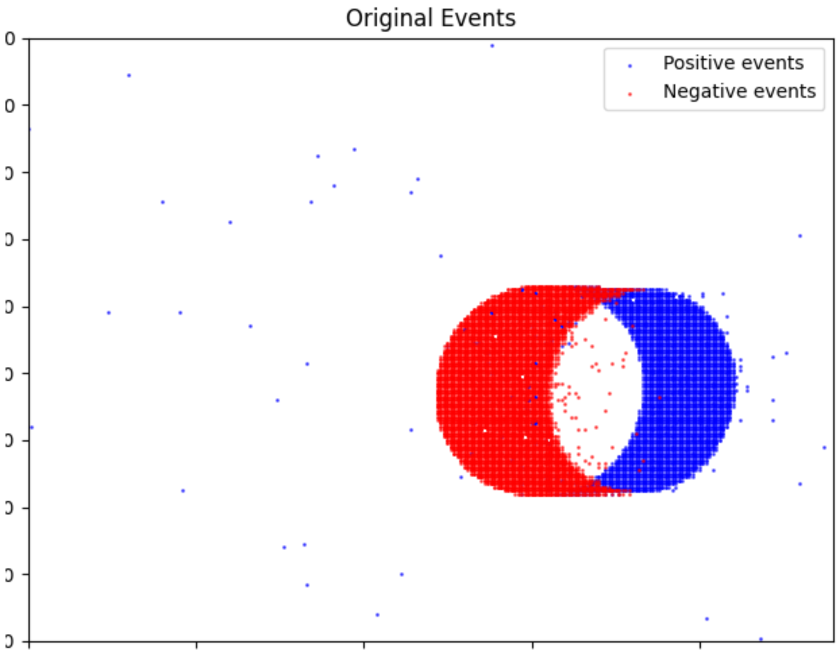
Lecture 14
Event-based Vision

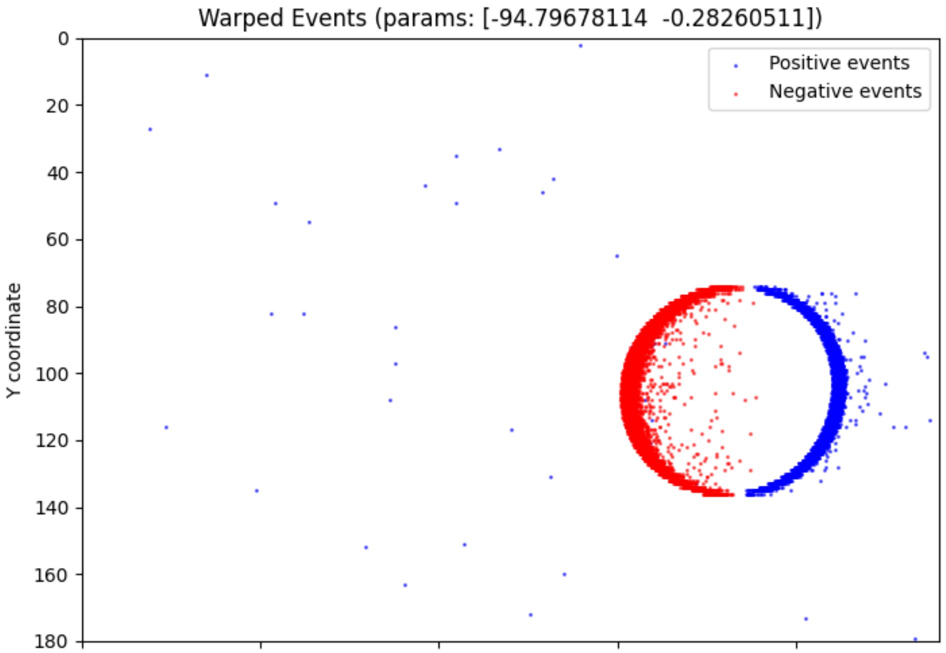
Davide Scaramuzza

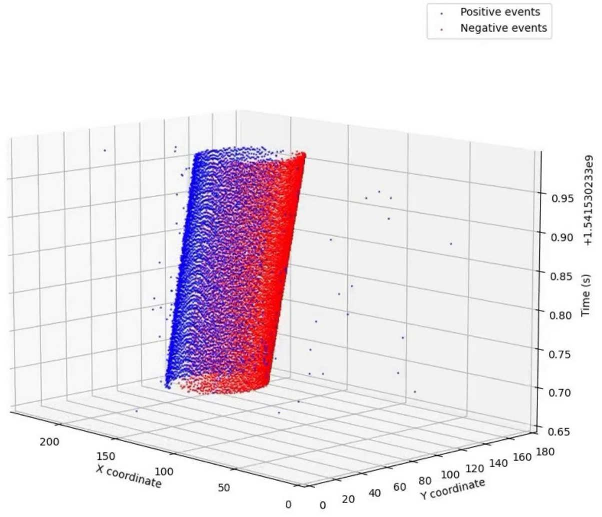
https://rpg.ifi.uzh.ch

Lab Exercise 11 – Event-based Vision

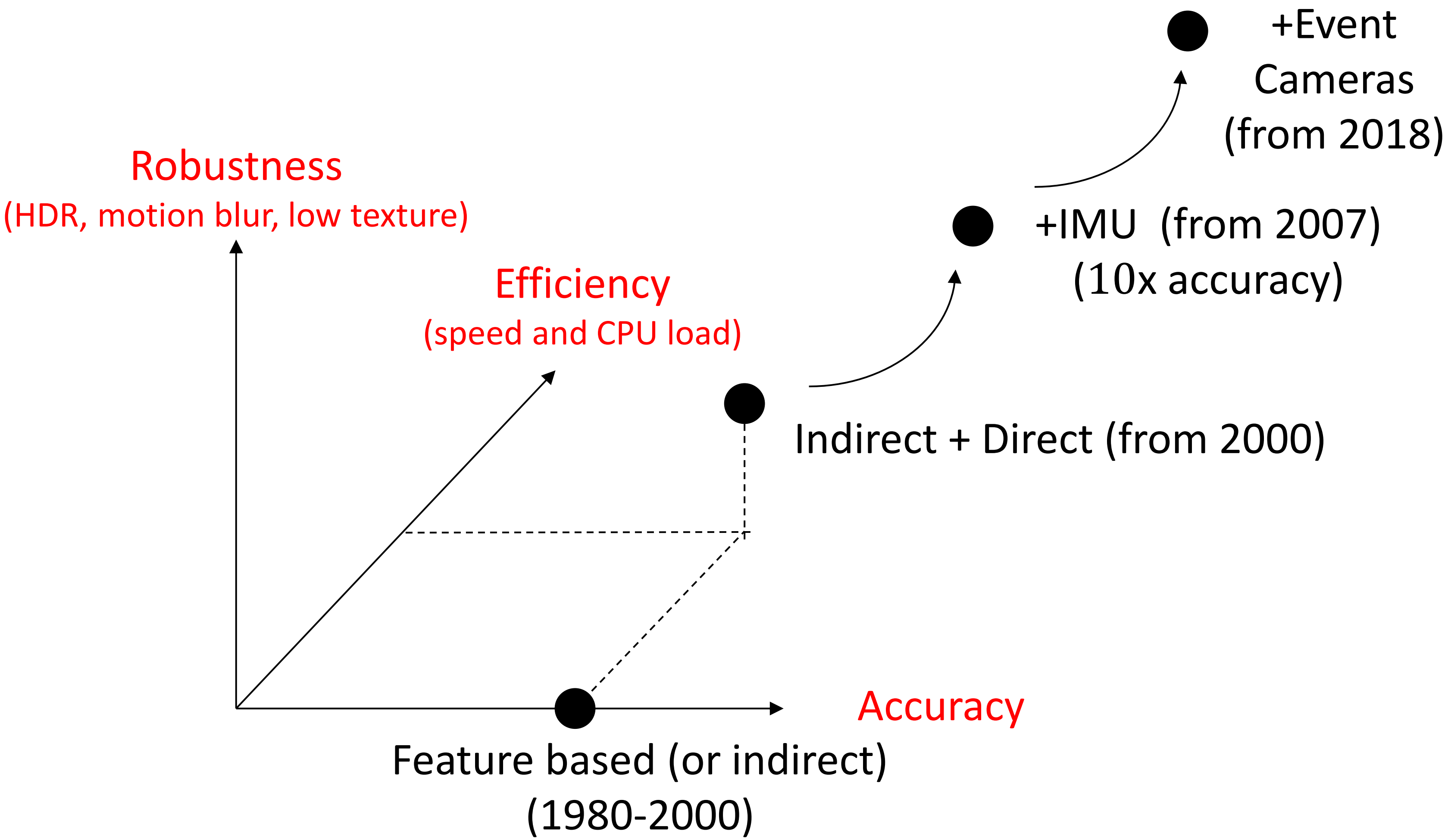
Followed by departure to visit our lab







A Taxonomy of the Last 44 Years of VIO



Open Challenges in Computer Vision

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

Motion blur

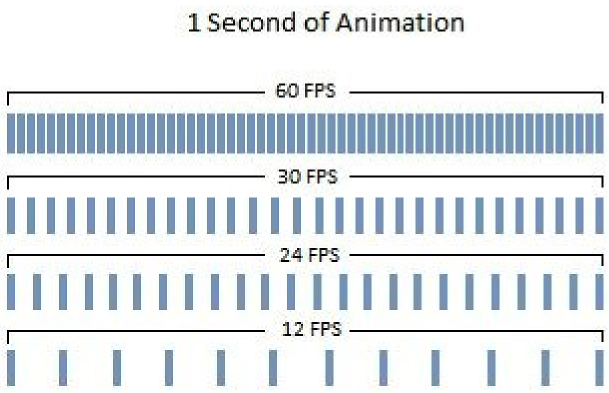
Dynamic Range

Bandwidth-Latency tradeoff

Open Challenges in Computer Vision

Standard cameras suffer from the **bandwidth-latency tradeoff**:

- A **high framerate reduces** perceptual **latency** but introduces significant **bandwidth overhead** for downstream tasks
- A **low framerate reduces the bandwidth** but at the cost of increasing the latency, thus missing important scene dynamics for safety-critical tasks.



Example grayscale VGA camera:

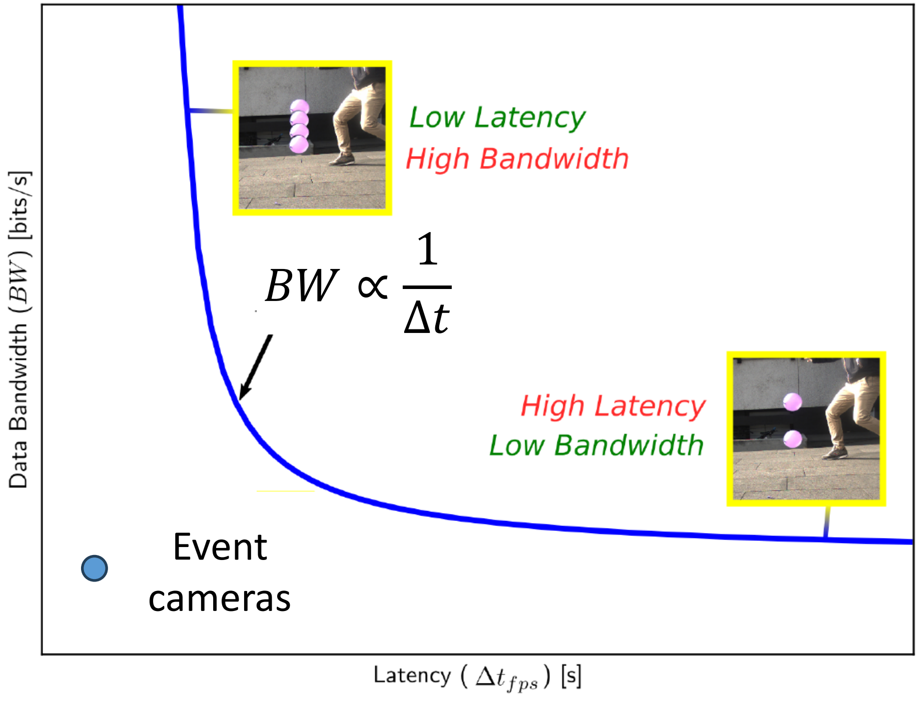
- **30 fps**:
  - Latency: **33 ms**
  - Bandwidth: **70 Megabits/s**
- **1,000 fps** :
  - Latency: **1 ms**
  - Bandwidth: **3,000 Megabits/s**
- **VGA event camera**:
  - Latency: **0.2 ms**
  - Bandwidth: **<10 Megabits/s**

Bandwidth-Latency tradeoff

Open Challenges in Computer Vision

Standard cameras suffer from the **bandwidth-latency tradeoff**:

- A **high framerate reduces** perceptual **latency** but introduces significant **bandwidth overhead** for downstream tasks
- A **low framerate reduces the bandwidth** but at the cost of increasing the latency, thus missing important scene dynamics for safety-critical tasks.



Bandwidth-Latency tradeoff

What is an Event Camera

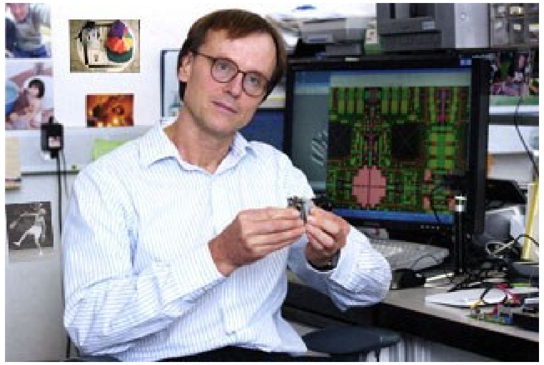
First commercialized by Prof. T. Delbruck in 2008 at the Institute of Neuroinformatics of UZH & ETH under the name of Dynamic Vision Sensor (DVS)

Advantages

- **Sub millisecond latency with micro-second resolution**
- **High updated rate** (1 MHz)
- **Negligible motion blur**
- **High-dynamic range (HDR)** (140 dB instead 60 dB)
- **Low power** (1mW instead 1W)

Challenges

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



Prof. [Tobi Delbruck](), UZH & ETH Zurich

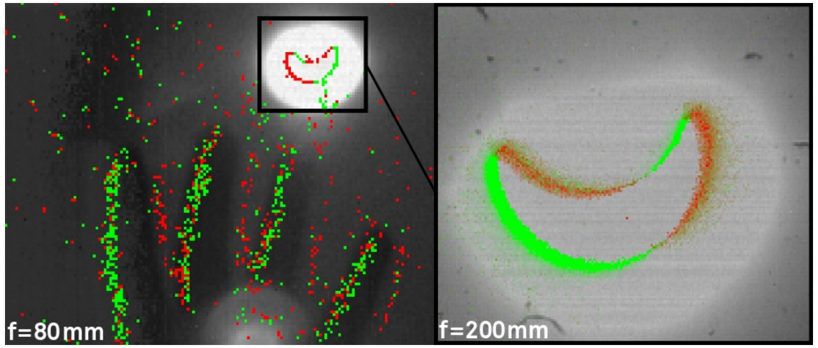
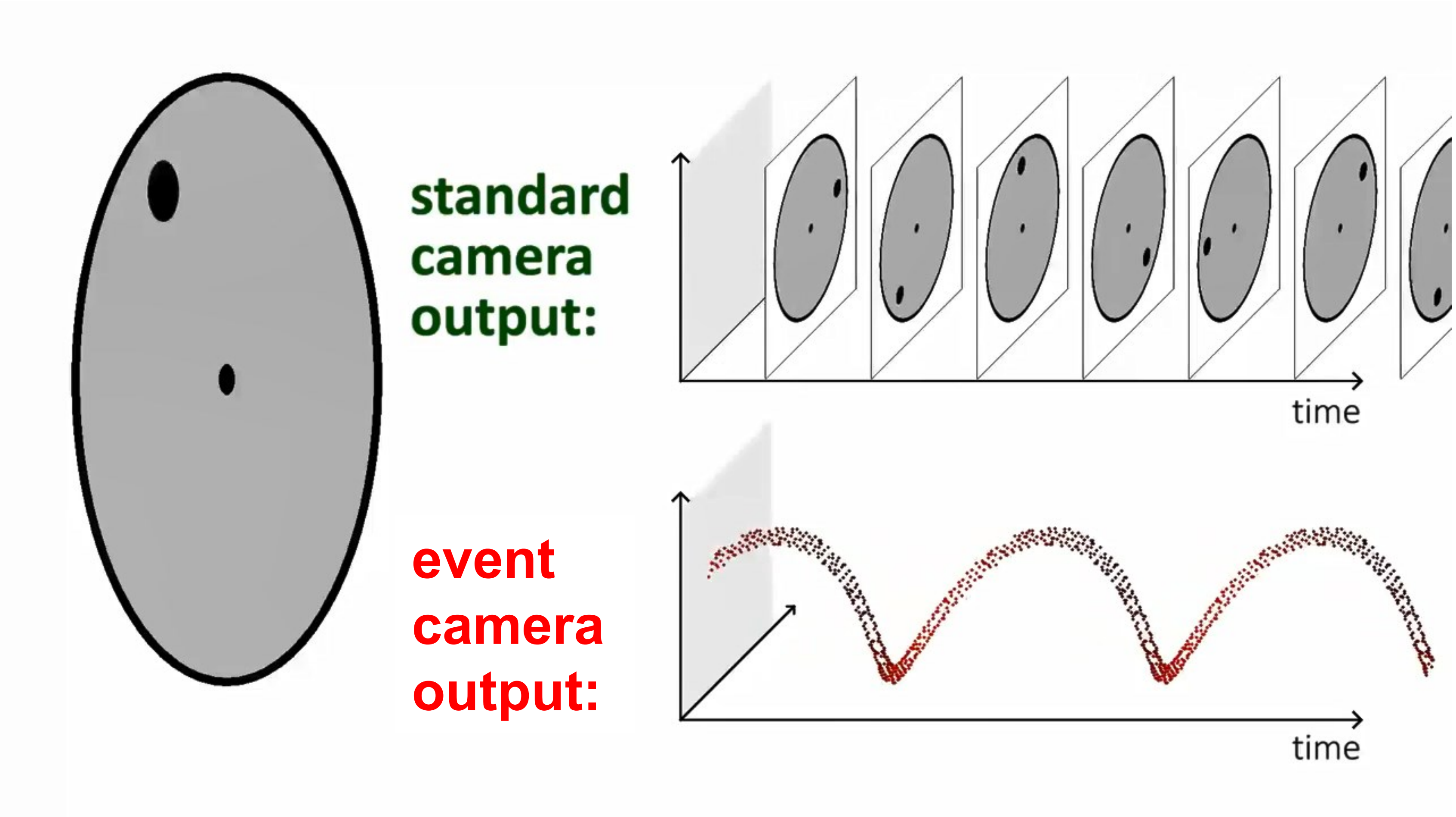


Image of solar eclipse captured by an event camera without black filter

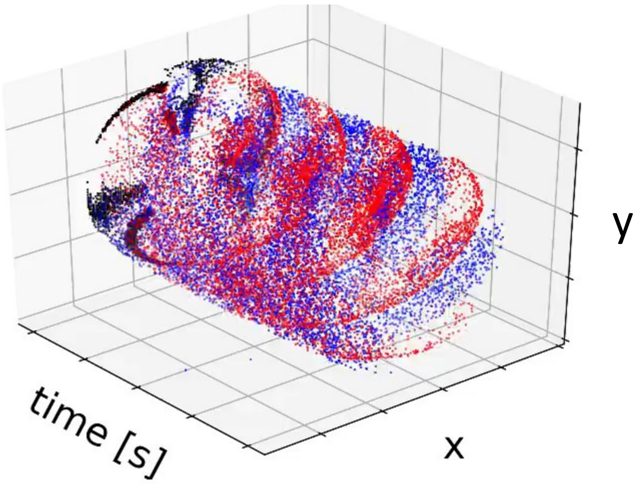
Lichtsteiner, Posch, Delbruck, A 128x128 120 dB 15µs Latency Asynchronous Temporal Contrast Vision Sensor, IEEE Journal of Solid-State Circuits, 2008. [PDF]()

Animation of an Event Camera Output



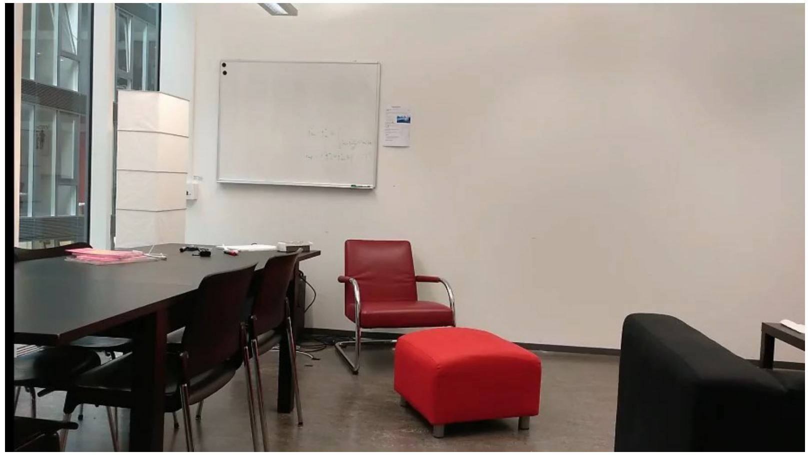
Video from here: https://youtu.be/LauQ6LWTkxM?t=30

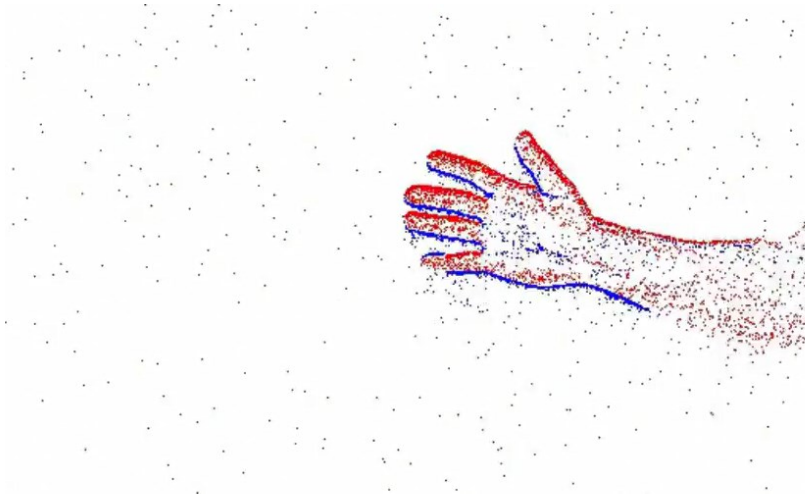
Conventional frames



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

Conventional frames





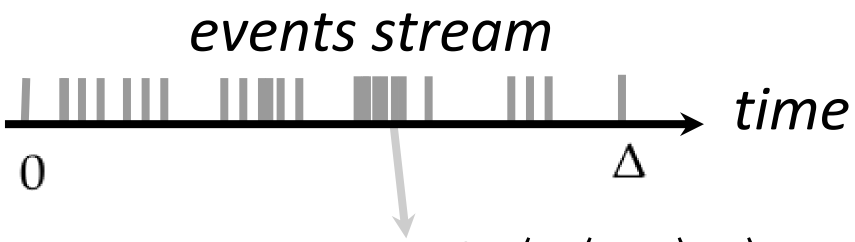
Events in the **image domain** (x, y)

Integration time can be arbitrary: from 1 microsecond to infinity

Standard Camera vs. Event Camera

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

- By contrast, an event camera outputs **asynchronous events** at **microsecond resolution**. An event is generated each time a single pixel detects a change of intensity



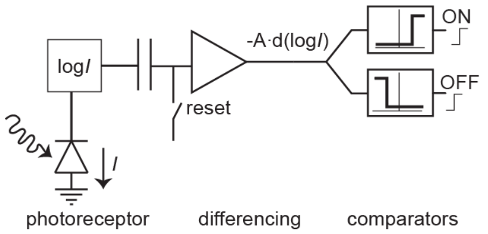
event: $\langle t, \langle x, y \rangle, p \rangle$

Timestamp (μs)

Pixel coordinates

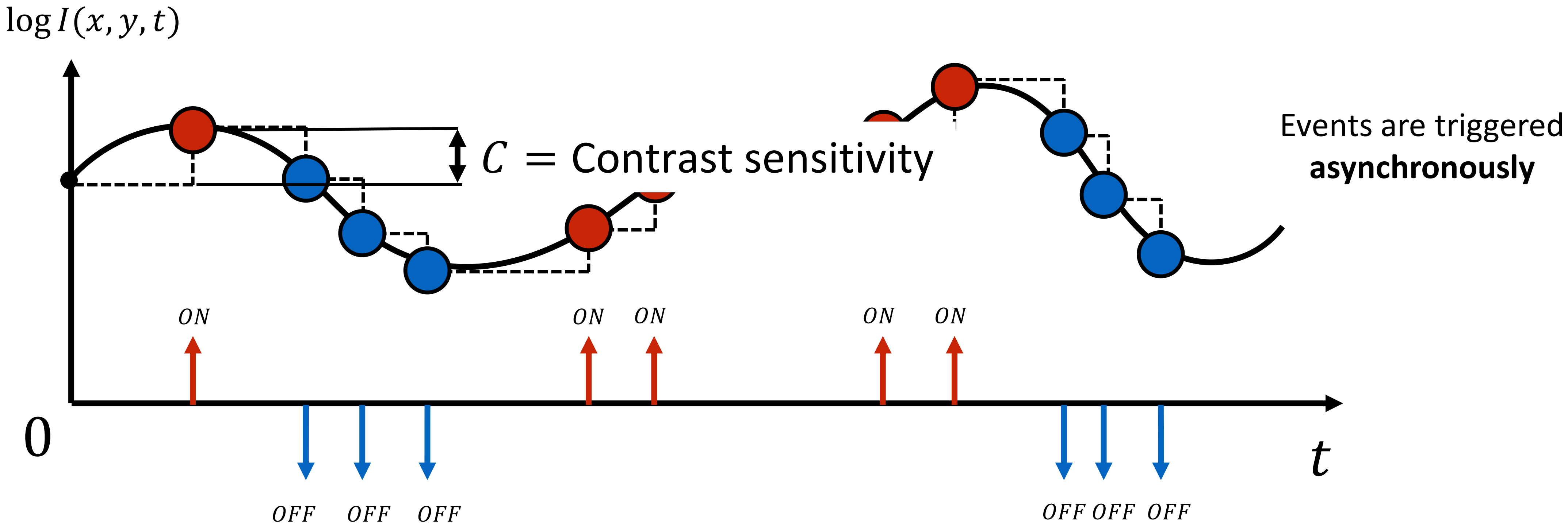
Event polarity (or sign) (+1 or -1): *positive or negative change*

Generative Event Model



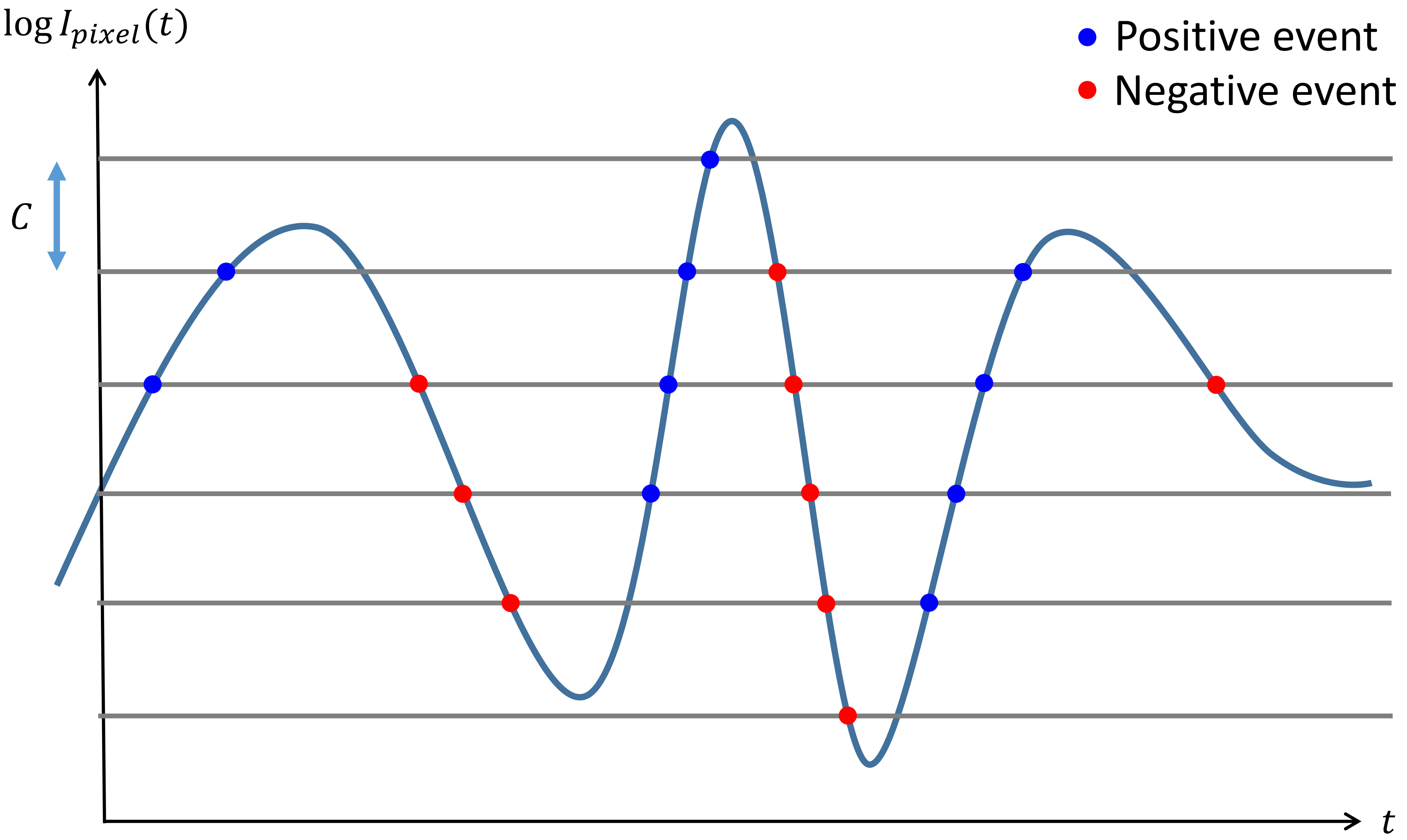
- Consider the intensity at a **single pixel** (x, y). An event is generated when the following condition is satisfied:

$$\log I(x, y, t + \Delta t) - \log I(x, y, t) = \pm C$$

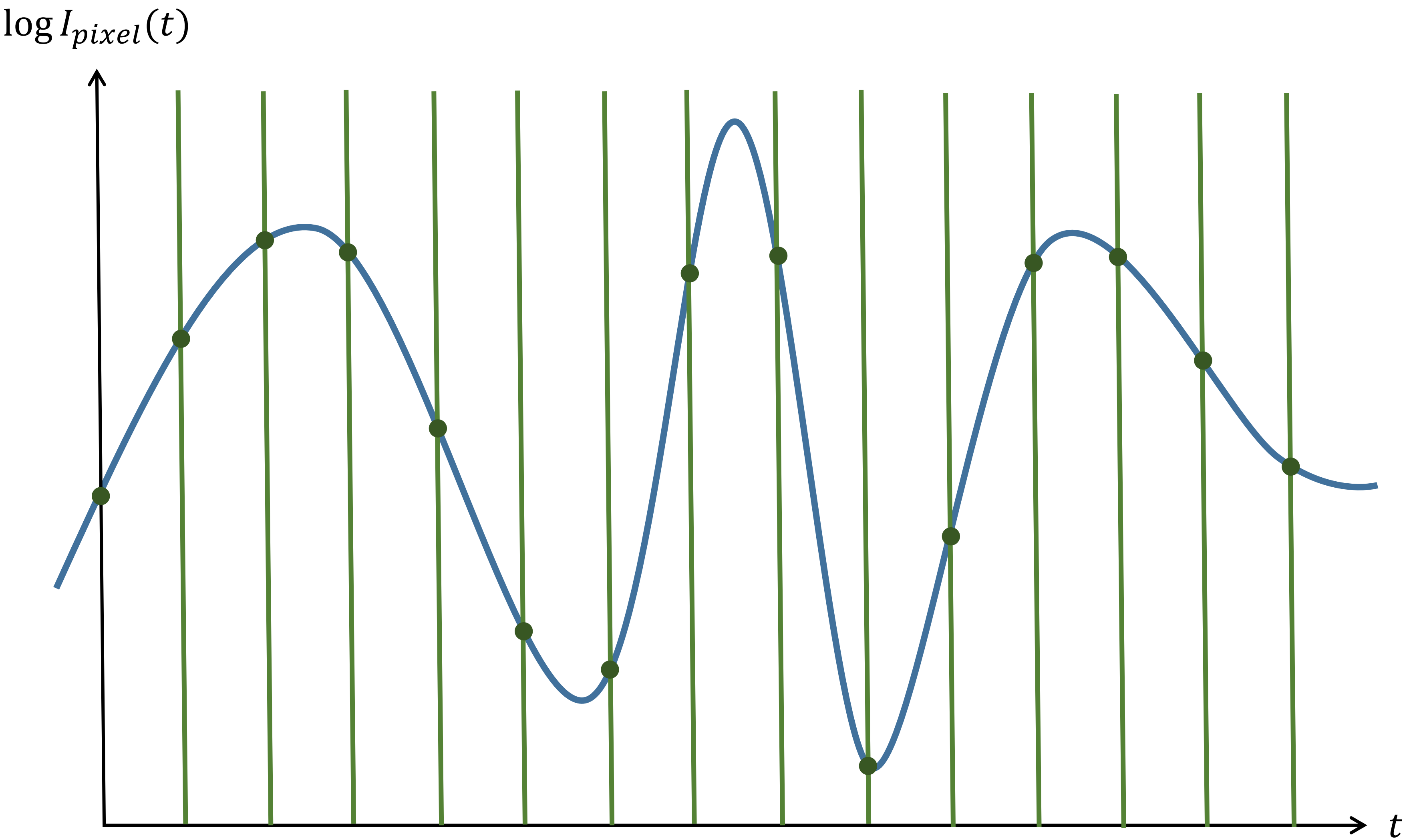


Can we reconstruct the pixel intensity? $\log(I(x, y, t)) = \log(x, y, 0) + \Sigma_{k=1}^{N_t} p_k C$

Event cameras sample the signal when the signal deviates from the last sampled value by a threshold **(level-crossing sampling)**



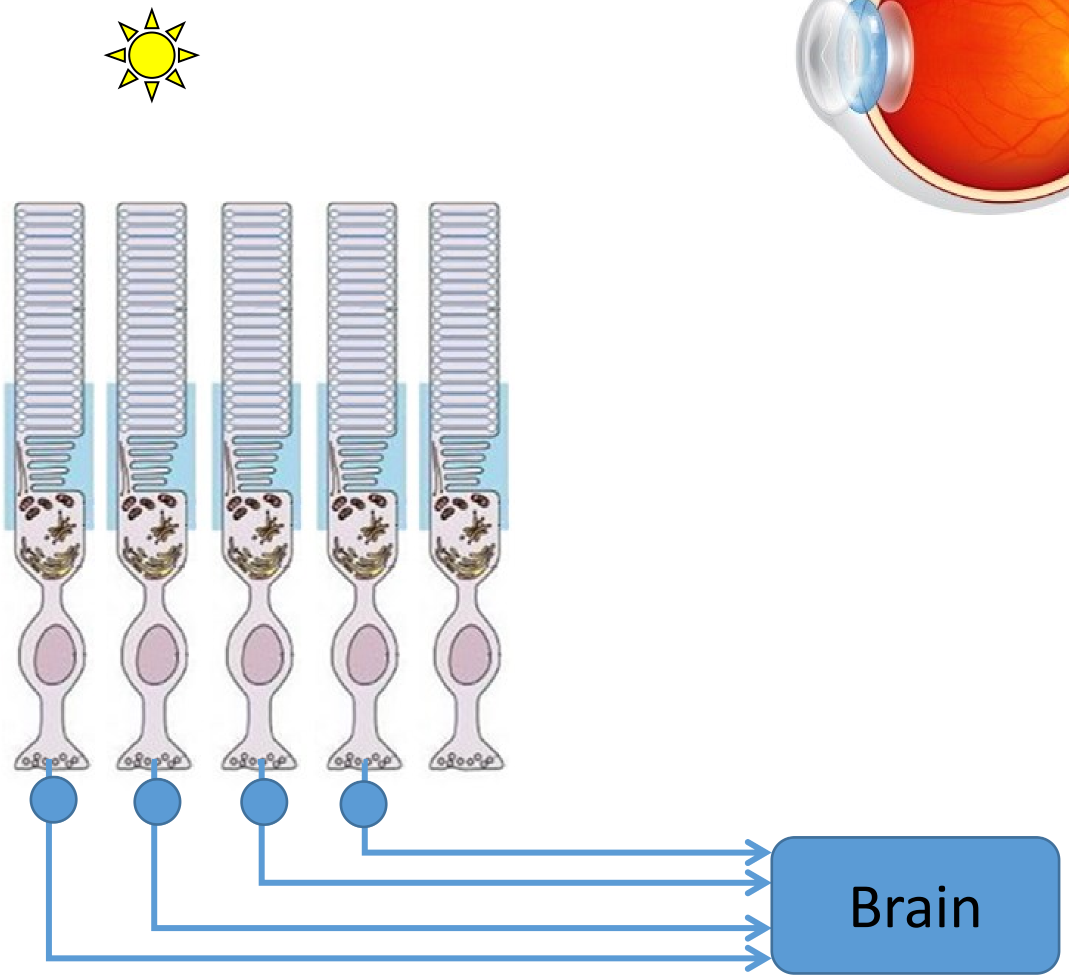
By contrast, standard cameras sample the signal at uniform time intervals **(uniform time sampling)**

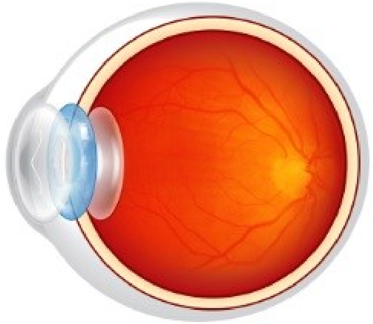


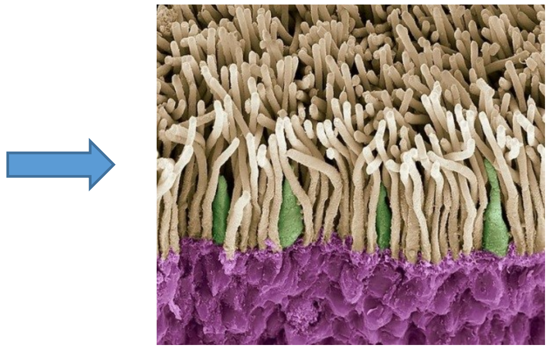
Event cameras are inspired by the Human Eye

Human retina:

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

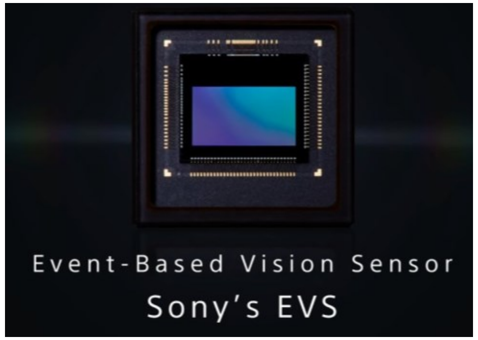


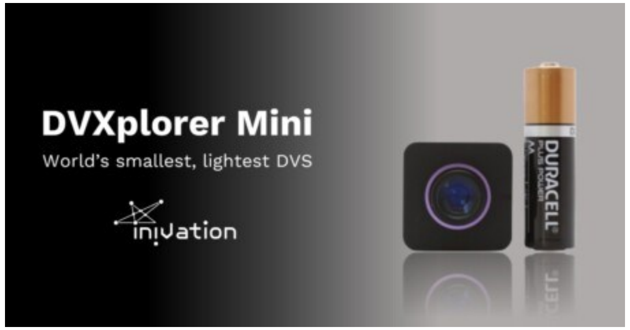




Who sells event cameras and how much are they?

- Prophesee & SONY:
  - Resolution: **1M pixels**
- Inivation & Samsung
  - Resolution**: VGA** (640x480 pixels)
- CelePixel Technology & Omnivision:
  - Resolution: **1M pixels**

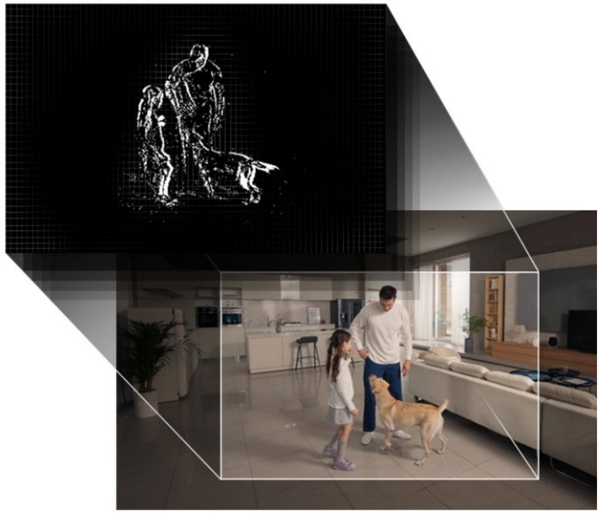




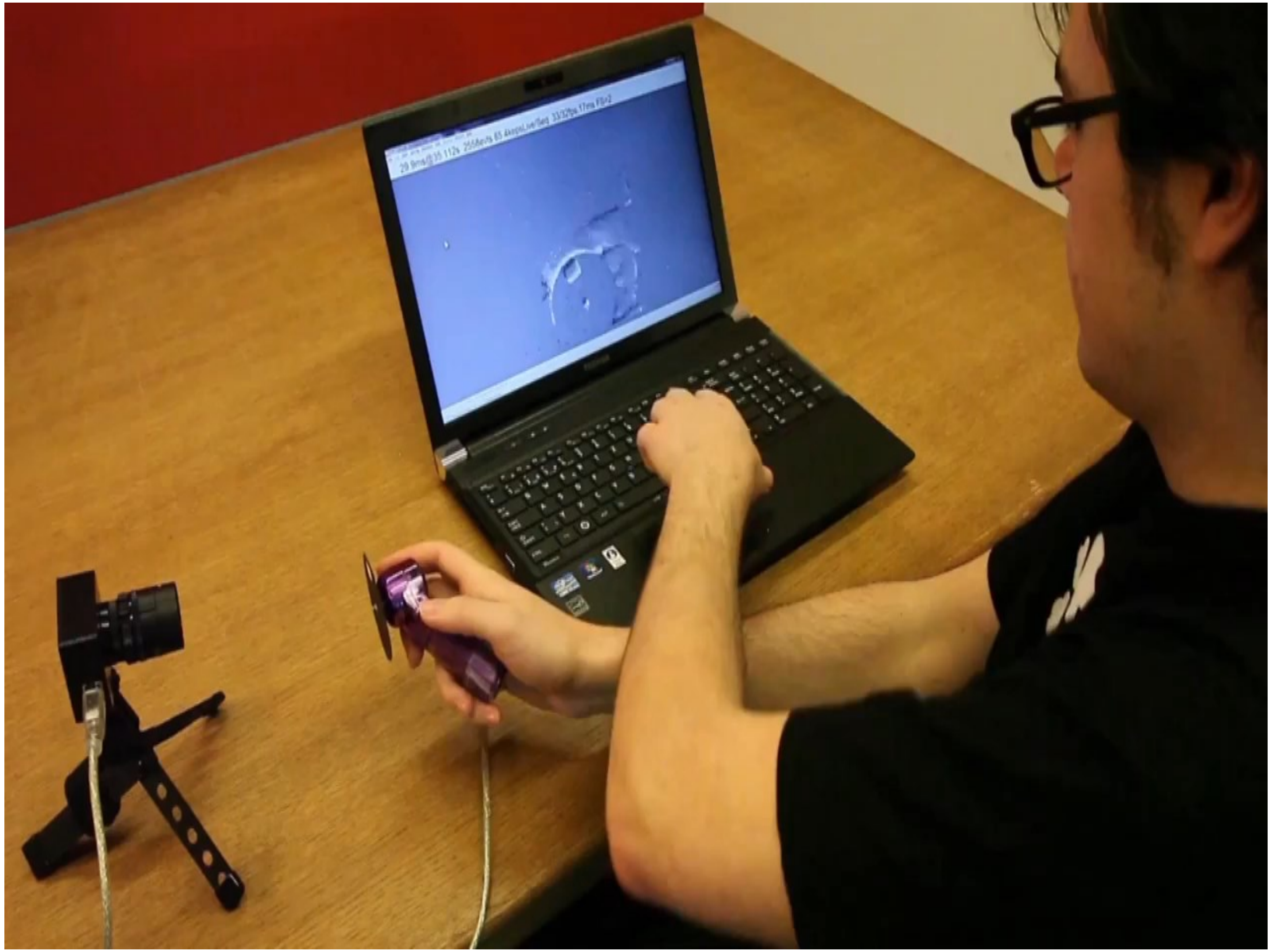
Who sells event cameras and how much are they?

- Prophesee & SONY:
  - Resolution: **1M pixels**
- Inivation & Samsung
  - Resolution: **VGA** (640x480 pixels)
- CelePixel Technology & Omnivision:
  - Resolution: **1M pixels**

$180

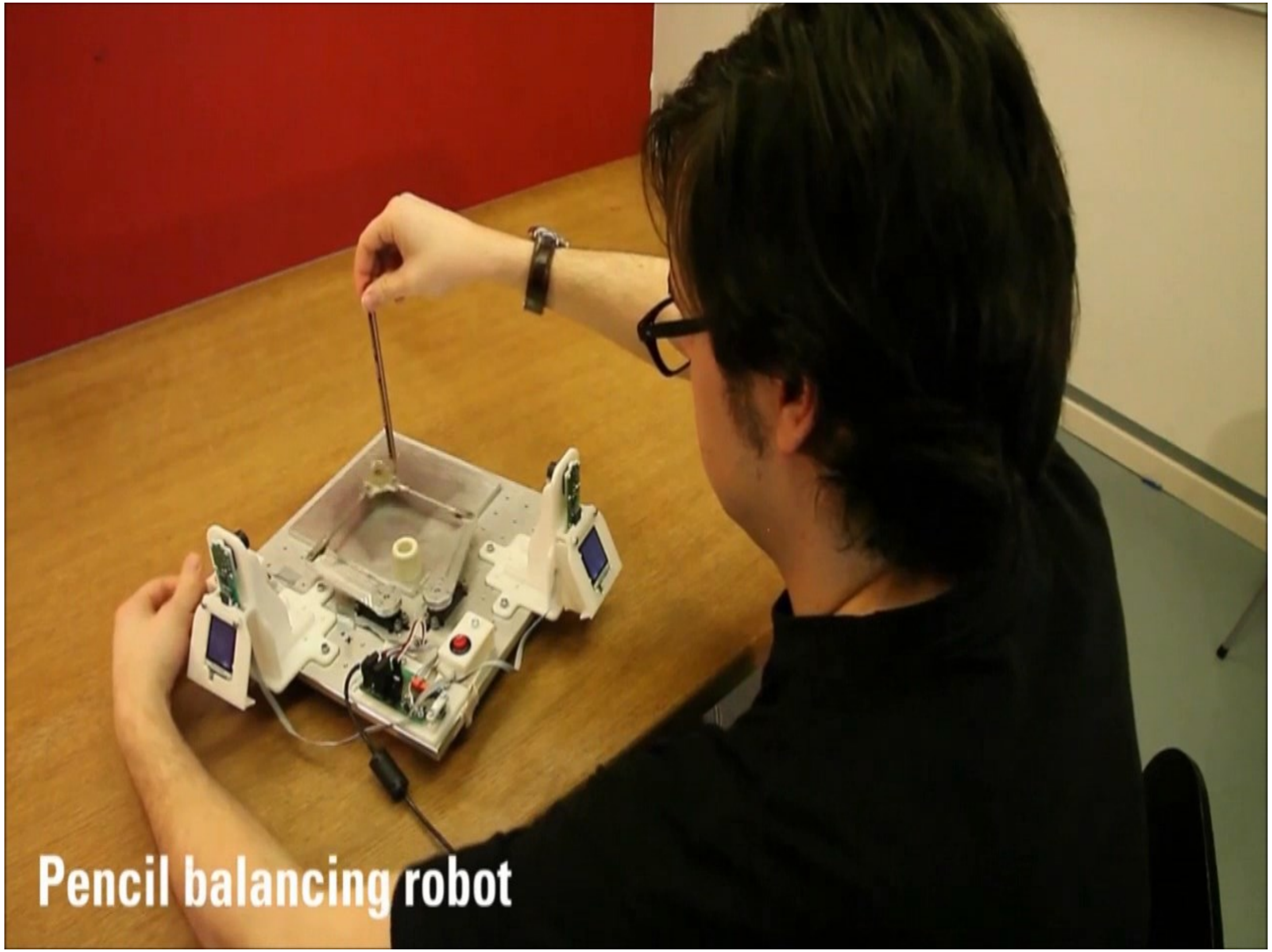


Event Camera Demo



https://youtu.be/QxJ-RTbpNXw

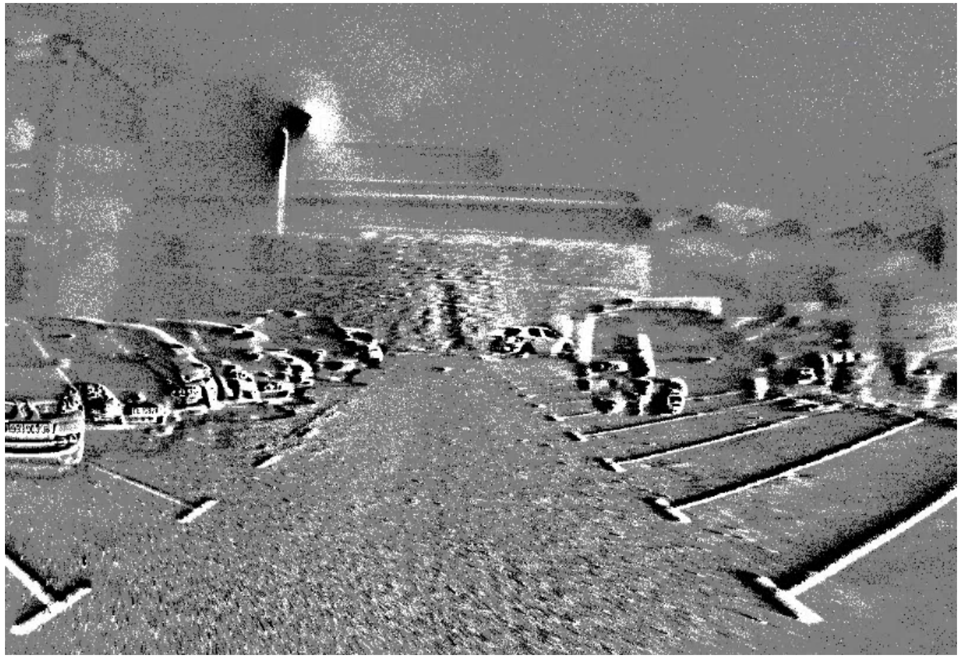
Event Camera Demo



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. [PDF](PDF).

Low-light Sensitivity (night drive)

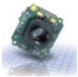
GoPro Hero 6

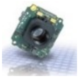


Aggregated event image
(pixel intensity equal to the sum of positive (+1) and negative (-1) events in a given time interval)

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

High-speed Camera vs. Event Camera





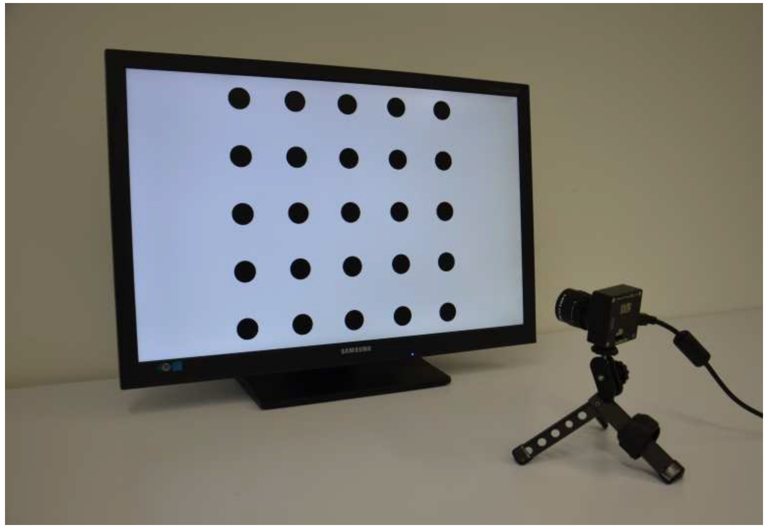
	High speed camera	Standard camera	Event Camera
Max fps or measurement rate	**Up to 1MHz** (watch the Slow Mo Guys on YouTube)	100-1,000 fps	**1MHz**
Resolution at max fps	640x64 pixels	**>1Mpxl**	**>1Mpxl**
Bits per pixels (event)	12 bits	8-10 per pixel	**~40 bits/event** {t,(x,y),p)}
Weight	6.2 Kg	**30 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 & contrast threshold)**
Mean power consumption	150 W + external light	1 W	**1 mW**
Dynamic range	not specified	60-140 dB depending on the quality	**140 dB**

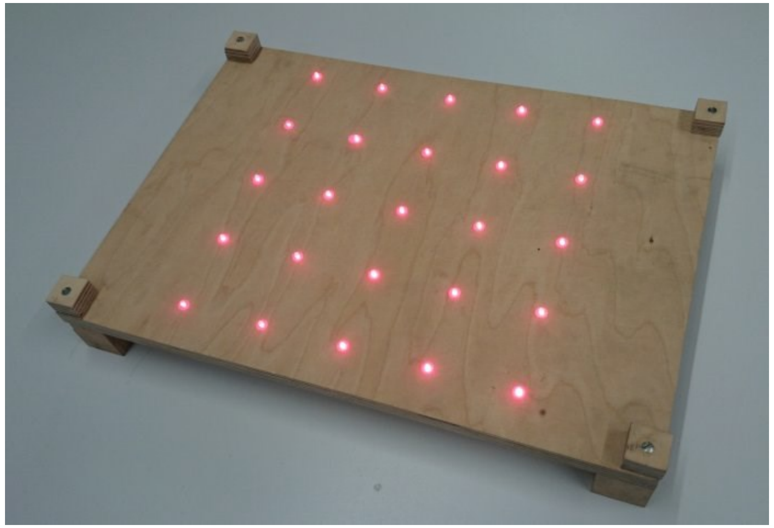
Current commercial applications

- **Monitoring and surveillance**
  - Action and gesture recognition in HDR scenes
- **Industrial automation**
  - Fast object counting
- **Computational photography**
  - Deblurring, super resolution, HDR, slow-motion video
- **High-speed robotics and Automotive**:
  - low-latency detection, object classification, low-power and low-memory storage

Calibration of an Event Camera

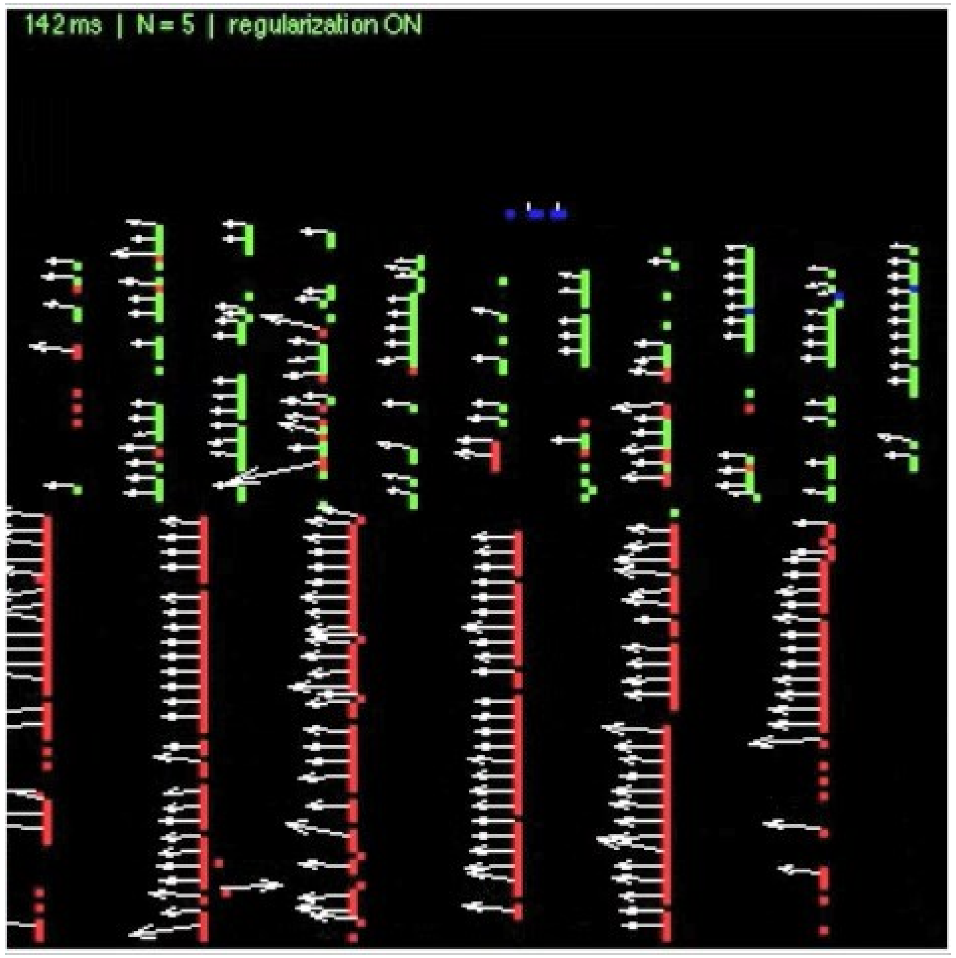
- Standard **pinhole camera model** still valid (same optics)
- Standard passive calibration patterns **cannot be used**
  - need to move the camera → inaccurate corner detection
- **Blinking patterns** (computer screen, LEDs)
- ROS DVS driver + intrinsic and extrinsic mono & stereo calibration: https://github.com/uzh-rpg/rpg_dvs_ros





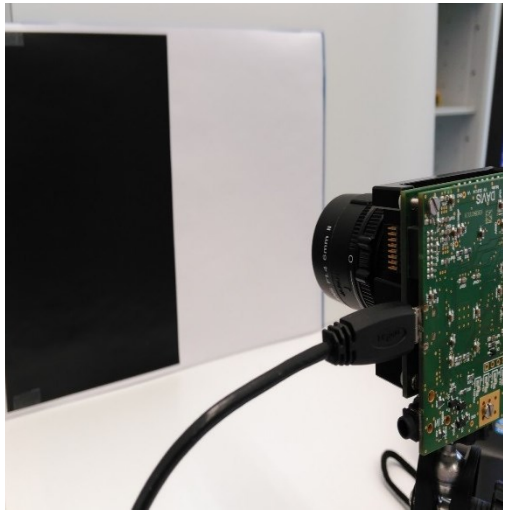
Mueggler, Huber, Scaramuzza, *Event-based 6-DOF Pose Tracking for High-Speed Maneuvers*, IEEE/RSJ International Conference on Robotics and Intelligent Systems (IROS), 2014. PDF.

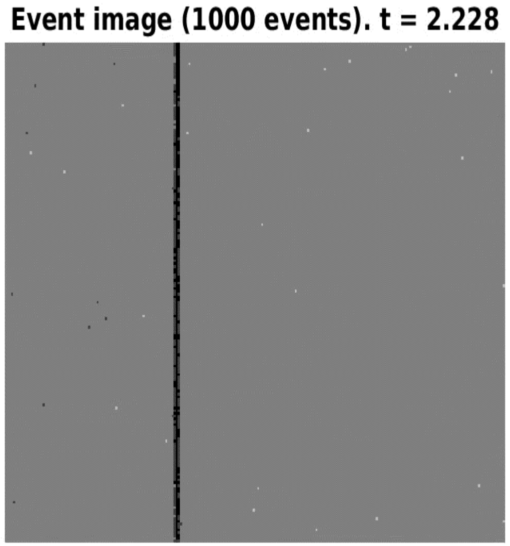
A Simple Optical Flow Algorithm

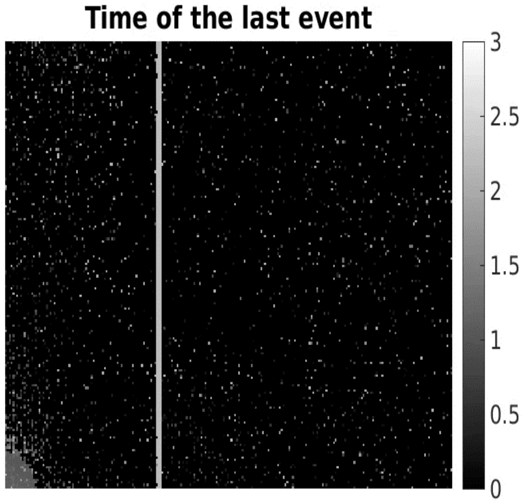


A Simple Optical Flow Algorithm

- Let’s assume pure horizontal left-to-right motion of binary pattern in front of the camera
- White pixels become black → brightness decrease → negative events (-1, i.e., in black color)



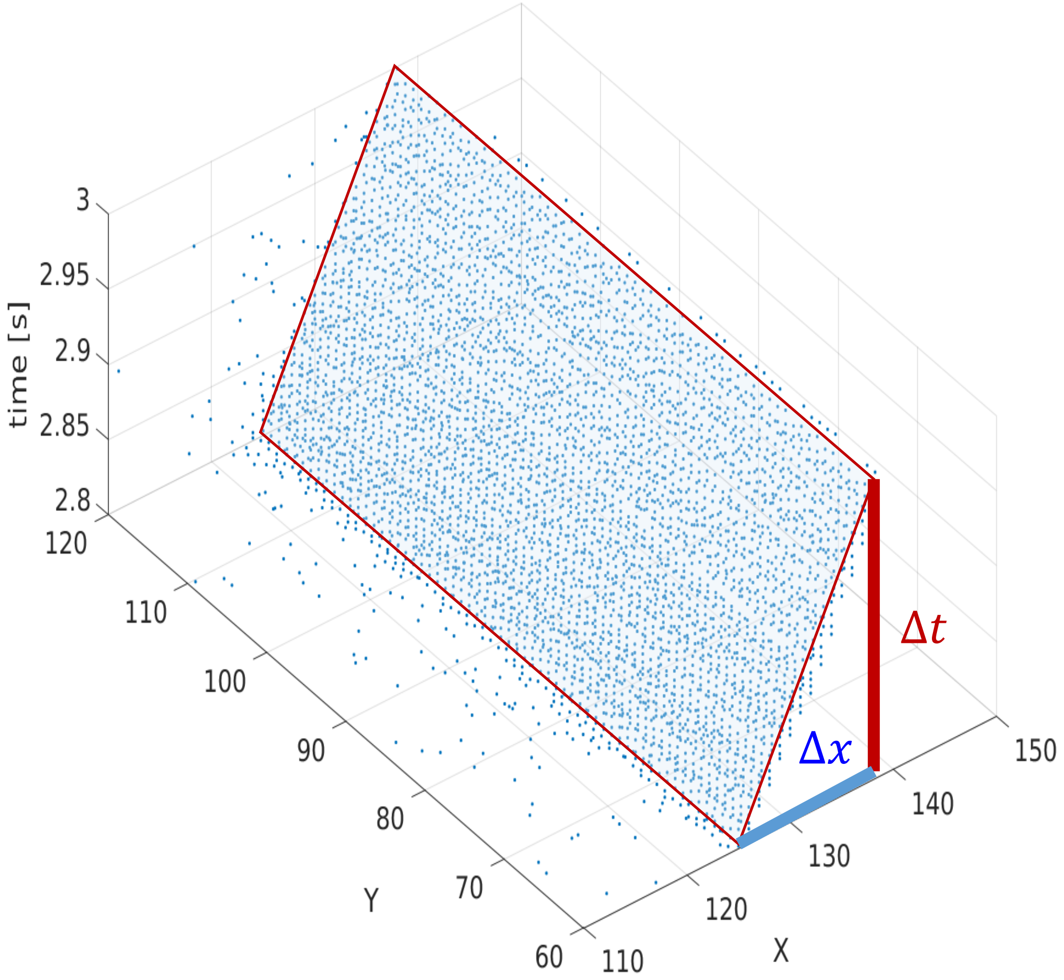




Negative events: -1 (black)
No events: 0 (gray)
Positive events: +1 (white)

A Simple Optical Flow Algorithm

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



The edge is moving at a speed of:

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

How do we unlock the outstanding potential of event cameras?

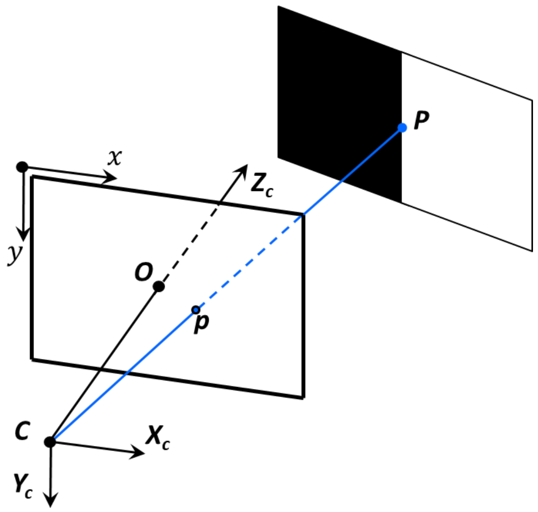
- Low latency
- High dynamic range
- No motion blur

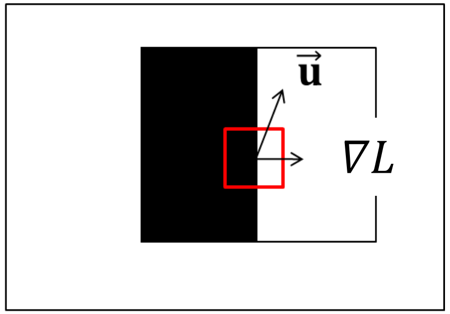
1st order approximation of the Generative Event Model

- An event is generated when the following condition is satisfied:

$$\log I(x, y, t + \Delta t) - \log I(x, y, t) = \pm C$$

- For many applications, it is convenient to derive a 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 $\boldsymbol{P}$



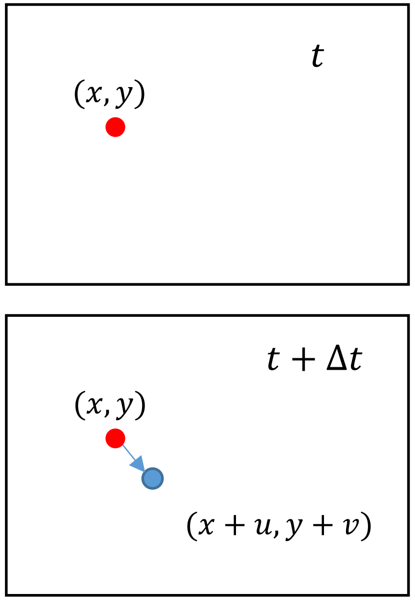


1st order approximation of the Generative Event Model

- Let's apply 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 with its 1st order approximation at $t + \Delta t$, we get:



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

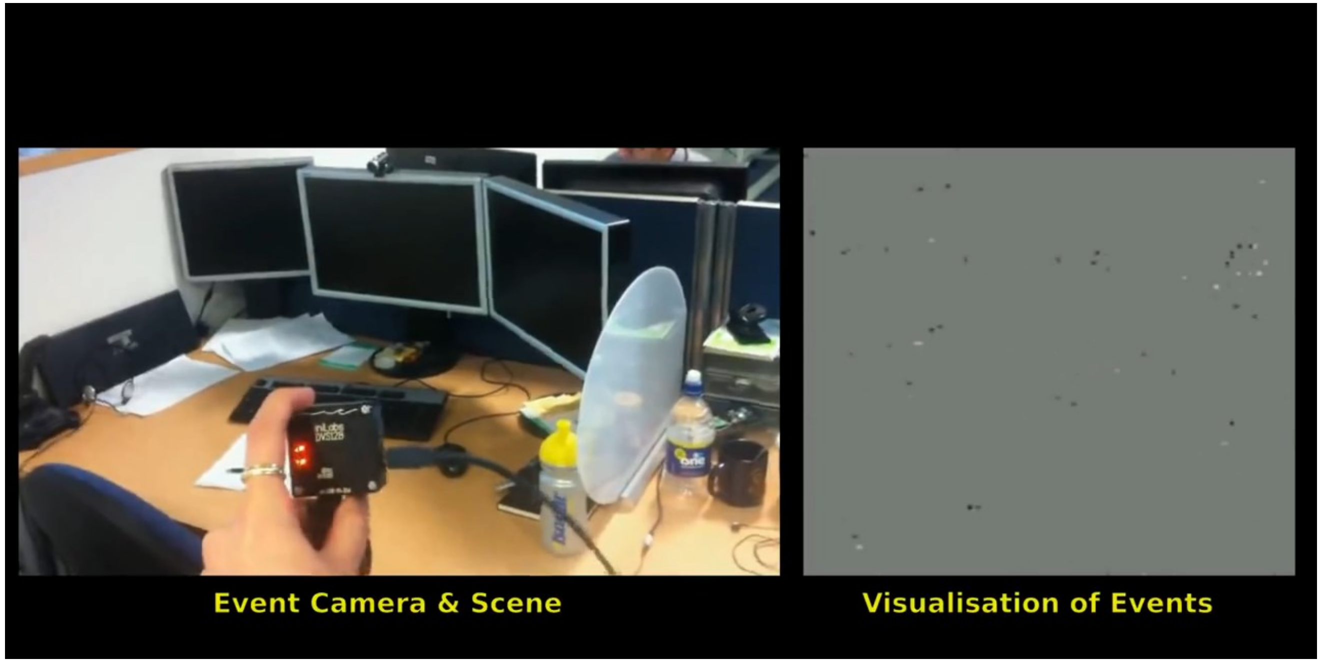
$$\Rightarrow \quad L(x, y, t+\Delta t) - L(x, y, t) = -\frac{\partial L}{\partial x} u - \frac{\partial L}{\partial y} v$$

$$\Rightarrow \boxed{\pm C = -\nabla L \cdot \boldsymbol{u}}$$

- This formula shows that **maximum generation of events** (i.e., higher event rate) occurs when the **relative motion of the camera is perpendicular to the edge** and is **minimum when parallel** to the edge.

Application 1: Image Reconstruction from events

- Probabilistic **simultaneous gradient reconstruction and rotation estimation** from $\pm C = -\nabla L \cdot \boldsymbol{u}$
- Obtain **image intensity from gradient** via Poisson reconstruction
- The reconstructed image has **super-resolution and High Dynamic Range** (HDR)
- Can run in **real time on a GPU**



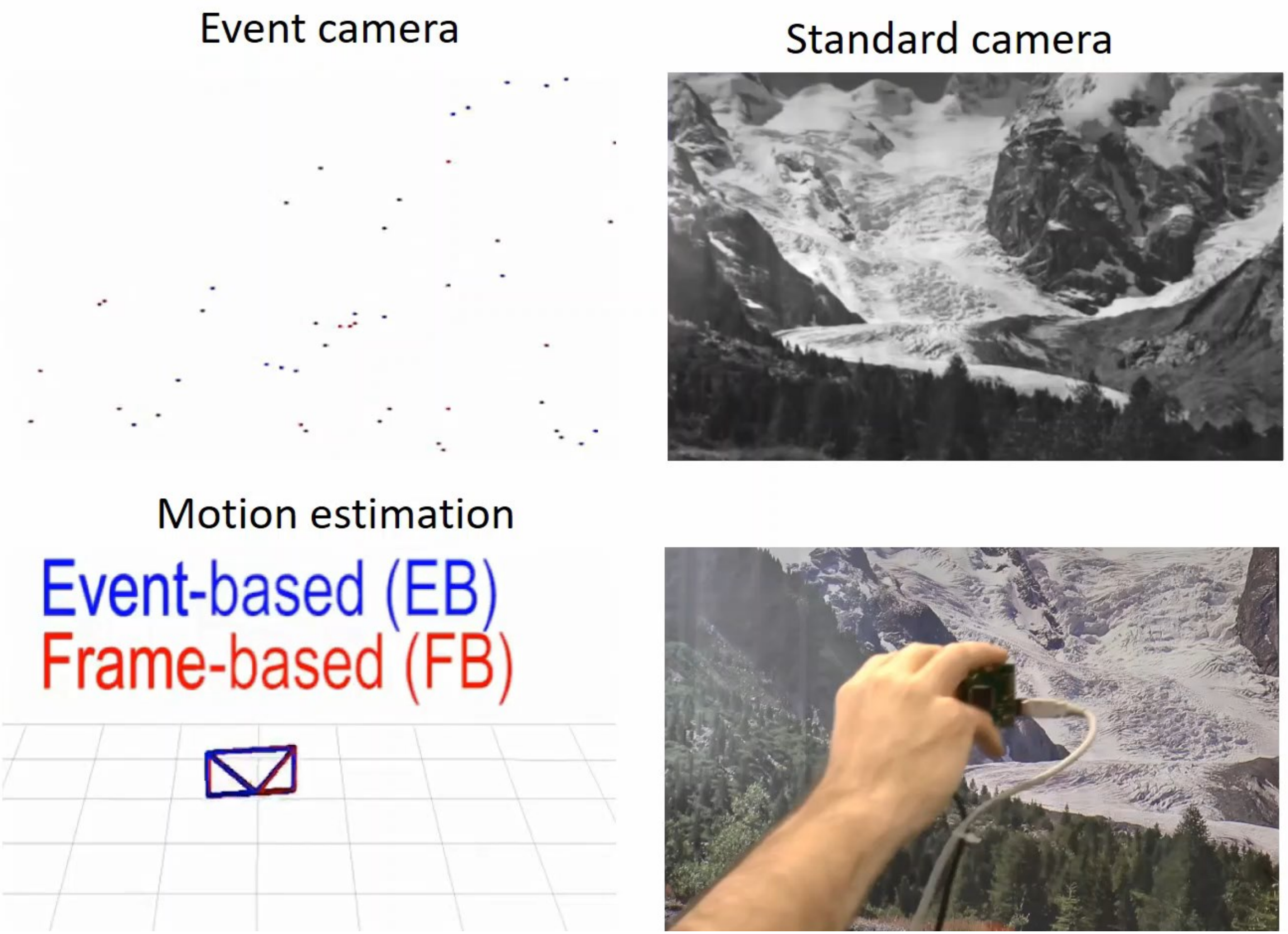
Kim, Handa, Benosman, Ieng, Davison, *Simultaneous Mosaicing and Tracking with an Event Camera*, British Machine Vision Conference (BMVC), 2014. PDF.

Application 2: 6DoF Tracking from Photometric Map

- Probabilistic **6DoF motion estimation** from $\pm C = -\nabla L \cdot \boldsymbol{u}$
- Assumes **photometric map** (x, y, z, grayscale Intensity) is **given**
- Useful for **VR/AR applications** (low-latency, HDR, no motion blur)
- Can run in **real time on a GPU**

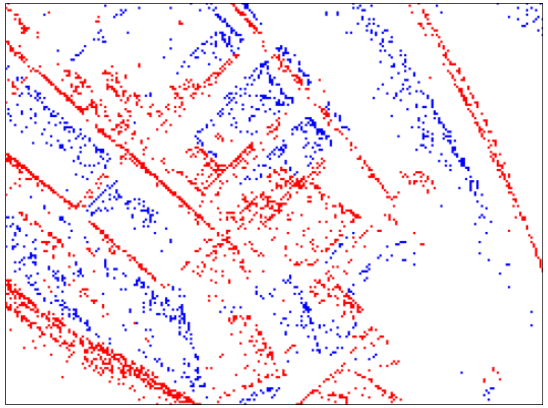
Gallego, Lund, Mueggler, Rebecq, Delbruck, Scaramuzza, *Event-based 6-DOF Camera Tracking from Photometric Depth Maps*, IEEE Transactions on Pattern Analysis and Machine Intelligence (T-PAMI), 2018. PDF. Video.

Application 2: 6DoF Tracking from Photometric Map

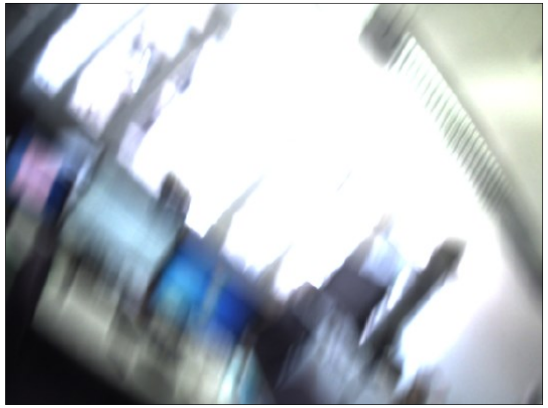


Gallego, Lund, Mueggler, Rebecq, Delbruck, Scaramuzza, *Event-based 6-DOF Camera Tracking from Photometric Depth Maps*, IEEE Transactions on Pattern Analysis and Machine Intelligence (T-PAMI), 2018. PDF. Video.

Combining Standard Cameras with Event Cameras



Event Camera

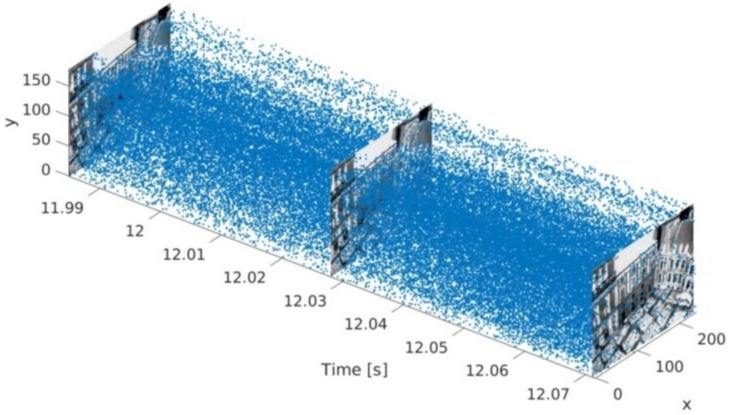


Standard Camera

	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!

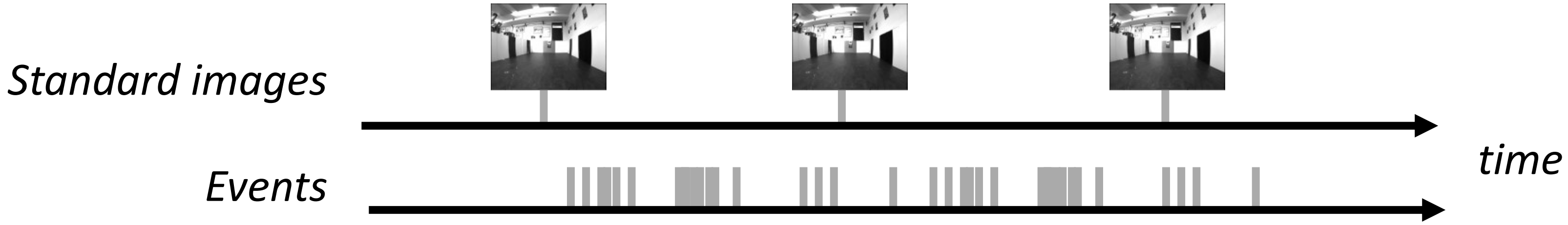
DAVIS sensor: Events + Images + IMU

- 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

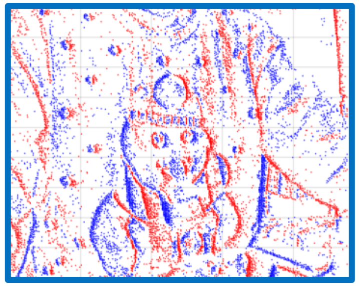
Temporal aggregation of events overlaid on a DAVIS frame



Brandli, Berner, Yang, Liu, Delbruck, *A 240x180 130dB 3us latency global shutter spatiotemporal vision sensor*. IEEE Journal on Solid State Circuits, 2014. PDF.

Application 1: Deblurring a blurry video

- **Idea**: 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**
- **Solution**: sharp image obtained by subtracting the double integral of event from input image

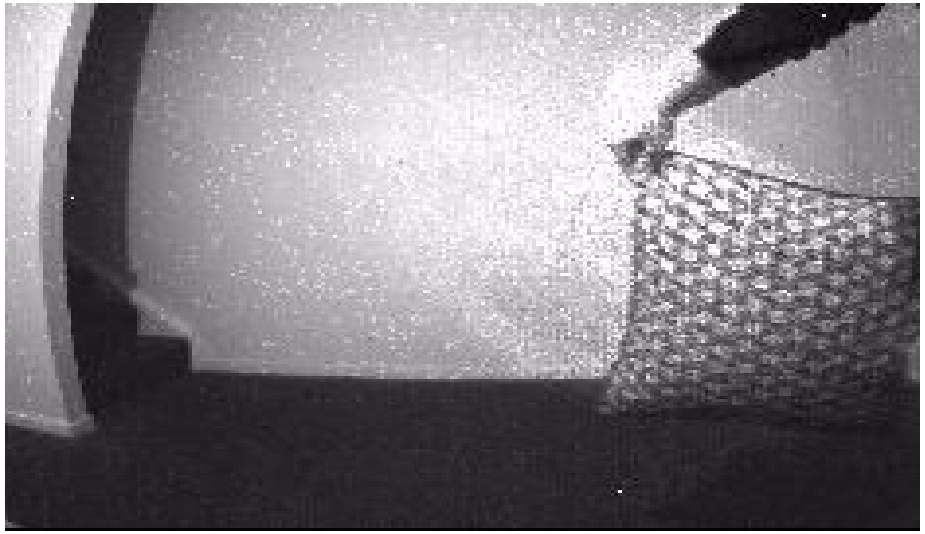
$\log$  $- \iint$  $= \log$ 

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

Pan, Scheerlinck, Hartley, Liu, Dai, Bringing a Blurry Frame Alive at High Frame-Rate with an Event Camera, International Conference on Computer Vision and Pattern Recognition, (CVPR), 2019. PDF.

Application 1: Deblurring a blurry video

- **Idea**: 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**
- **Solution**: sharp image obtained by subtracting the double integral of event from input image

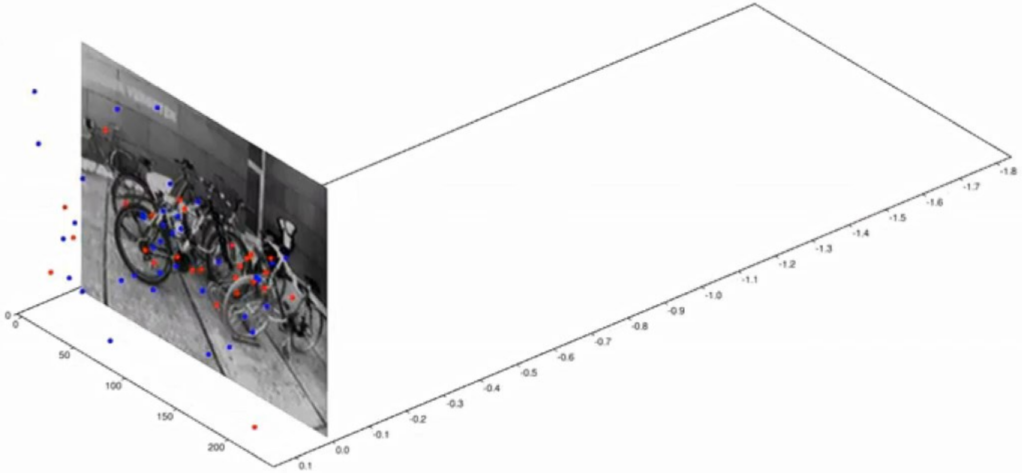


Input blur image **Output sharp video**

Pan et al., Bringing a Blurry Frame Alive at High Frame-Rate with an Event Camera, International Conference on Computer Vision and Pattern Recognition, (CVPR), 2019. [PDF](PDF).

Application 3: Event-based KLT Tracking

- **Goal**: Extract **features from standard frames** and track them using only **events** in the **blind time between two frames**
- Uses the **1st order approximation of event generation model** via joint estimation of patch warping and optic flow



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

Gehrig, Rebecq, Gallego, Scaramuzza, *EKLT: Asynchronous, Photometric Feature Tracking using Events and Frames*, International Journal of Computer Vision (IJCV), 2019. PDF. Video. Code

Recap

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

$$\log I(x, y, t + \Delta t) - \log I(x, y, t) = \pm C$$

- or its 1st order approximation

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

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

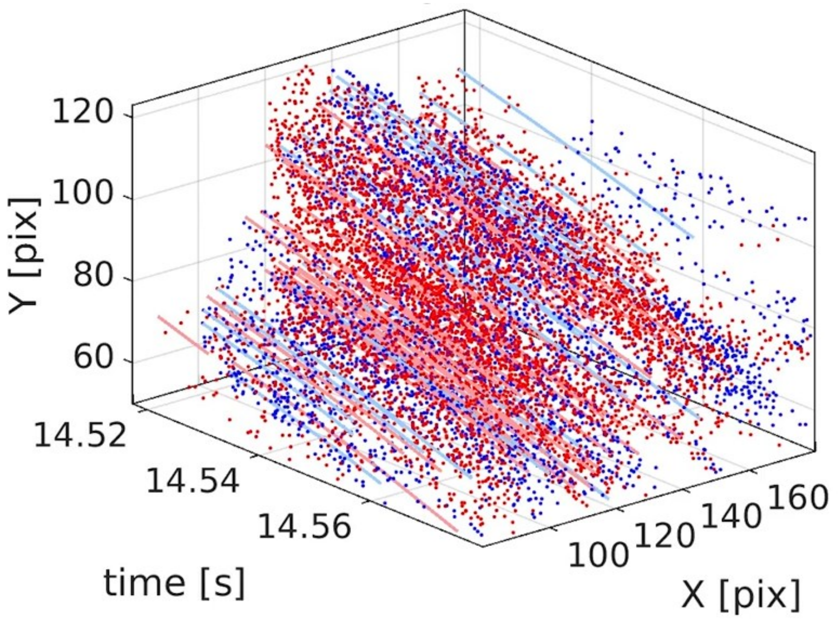
- Unfortunately, $\boldsymbol{C}$ **is scene dependent** and might **differ from pixel to pixel**
- **Alternative approach: Contrast maximization framework**

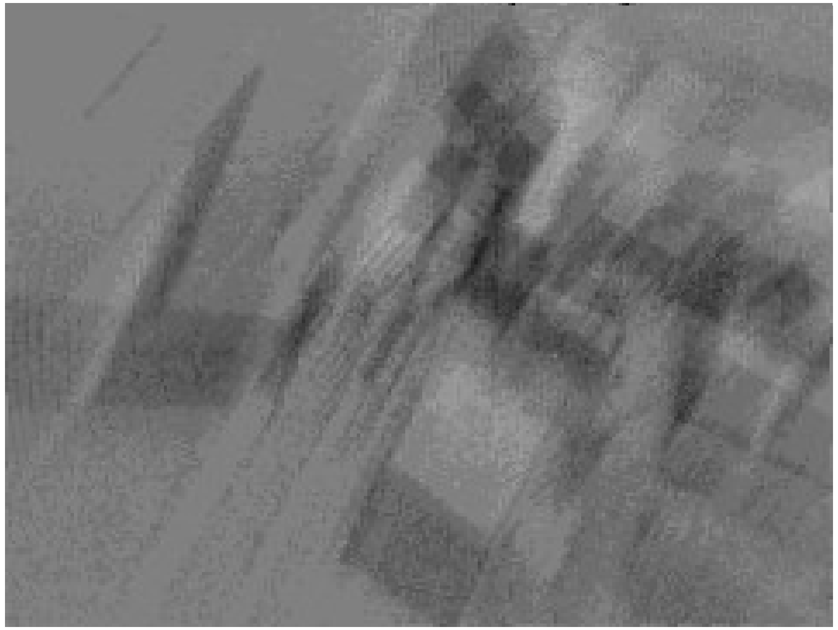
Contrast Maximization Framework

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

Contrast Maximization Framework

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

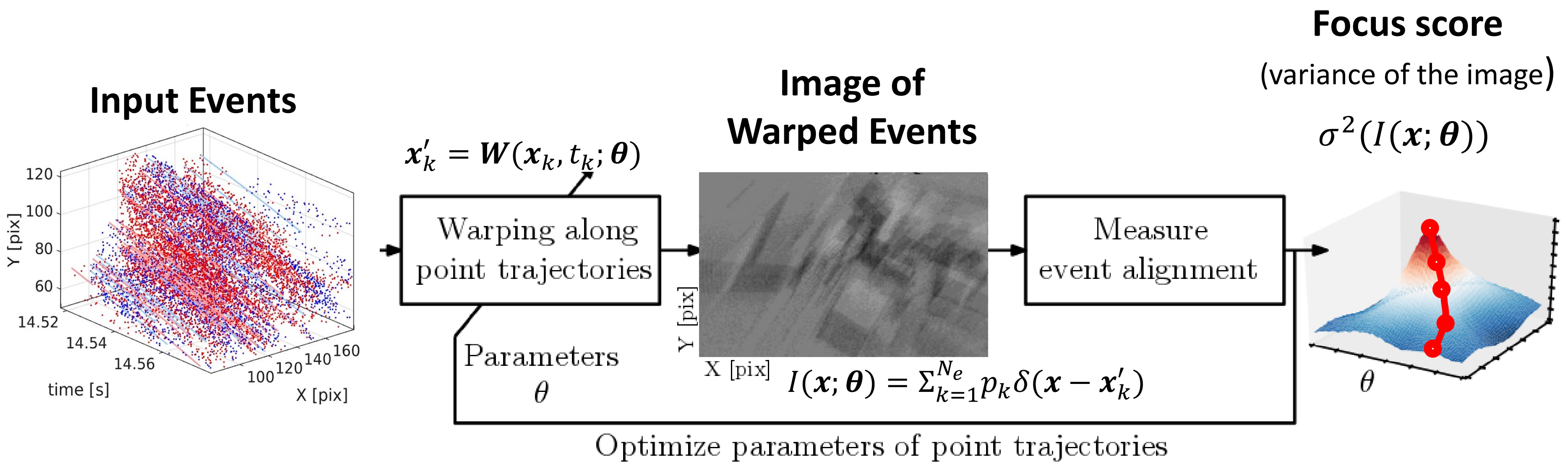




Aggregated image
without motion correction

Gallego, Rebecq, Scaramuzza, *A Unifying Contrast Maximization Framework for Event Cameras*, CVPR18, [PDF](), [Video]()
Gallego, Gehrig, Scaramuzza, *Focus Is All You Need: Loss Functions for Event-based Vision*, CVPR19, [PDF]().

Contrast Maximization Framework

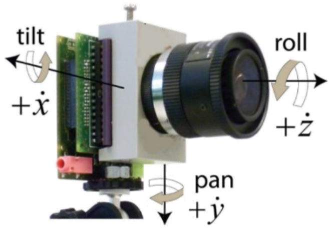


- $\boldsymbol{x}'_k = \boldsymbol{W}(\boldsymbol{x}_k, t_k; \boldsymbol{\theta})$: This warps the (x, y) pixels coordinates of each event, not their time. Possible warps: roto-translation, affine, homography.

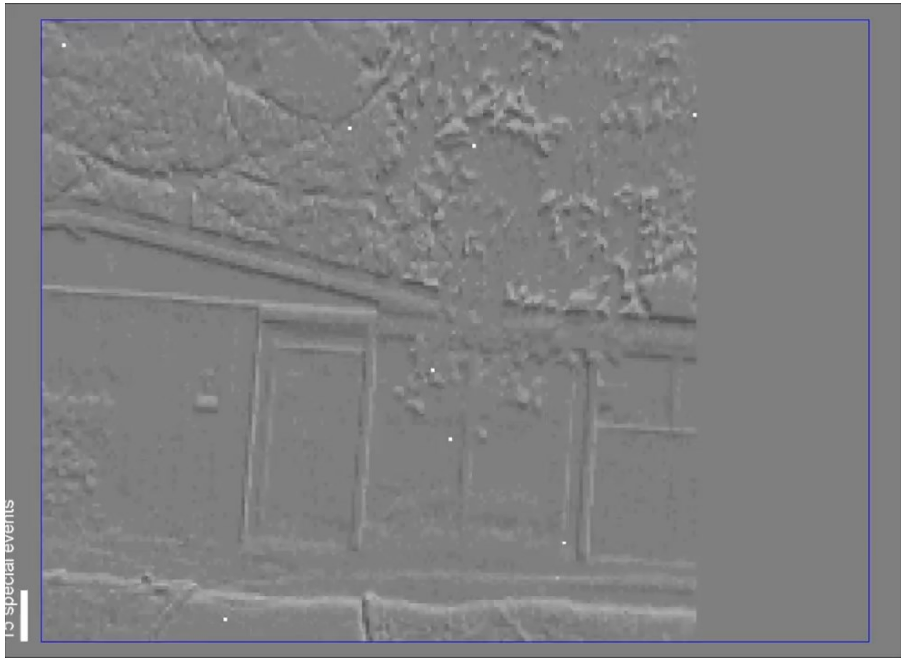
- $I(\boldsymbol{x}; \boldsymbol{\theta}) = \Sigma_{k=1}^{N_e} p_k \delta(\boldsymbol{x} - \boldsymbol{x}'_k)$: This builds a grayscale image, where the intensity of each pixel at the warped location (x', y') is equal to the summation of the polarity p (i.e., positive and negative events $(+1, -1)$)

- $\sigma^2(I(\boldsymbol{x}; \boldsymbol{\theta}))$: 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

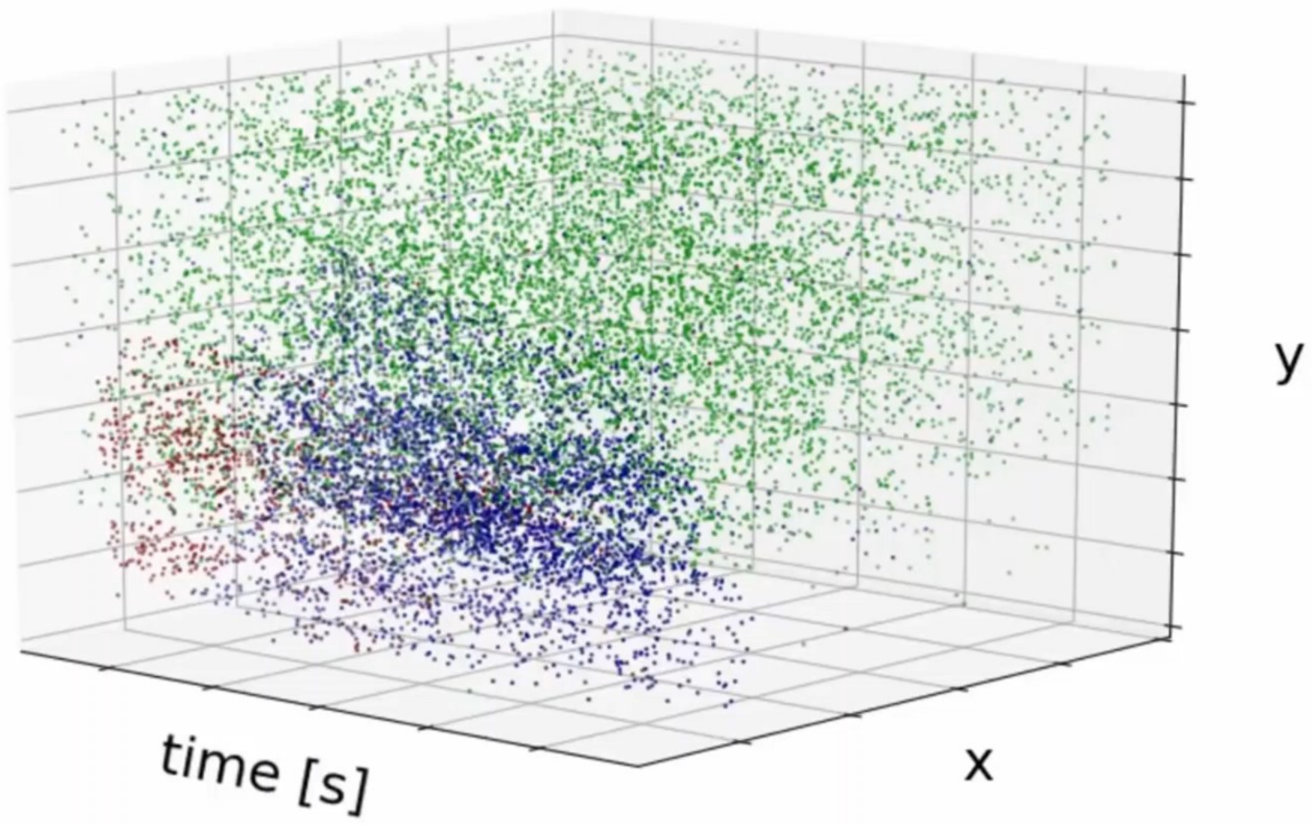


- Goal: **Estimate rotational motion (3DoF)** of an event camera
- Can process millions of events per second in real time on a smartphone PC (e.g., OdroidXU4)
- Works up to over ~1,000 deg/s



Gallego, Scaramuzza, Accurate Angular Velocity Estimation with an Event Camera, IEEE Robotics and Automation Letters (RA-L), 2016. [PDF](). [Video]().

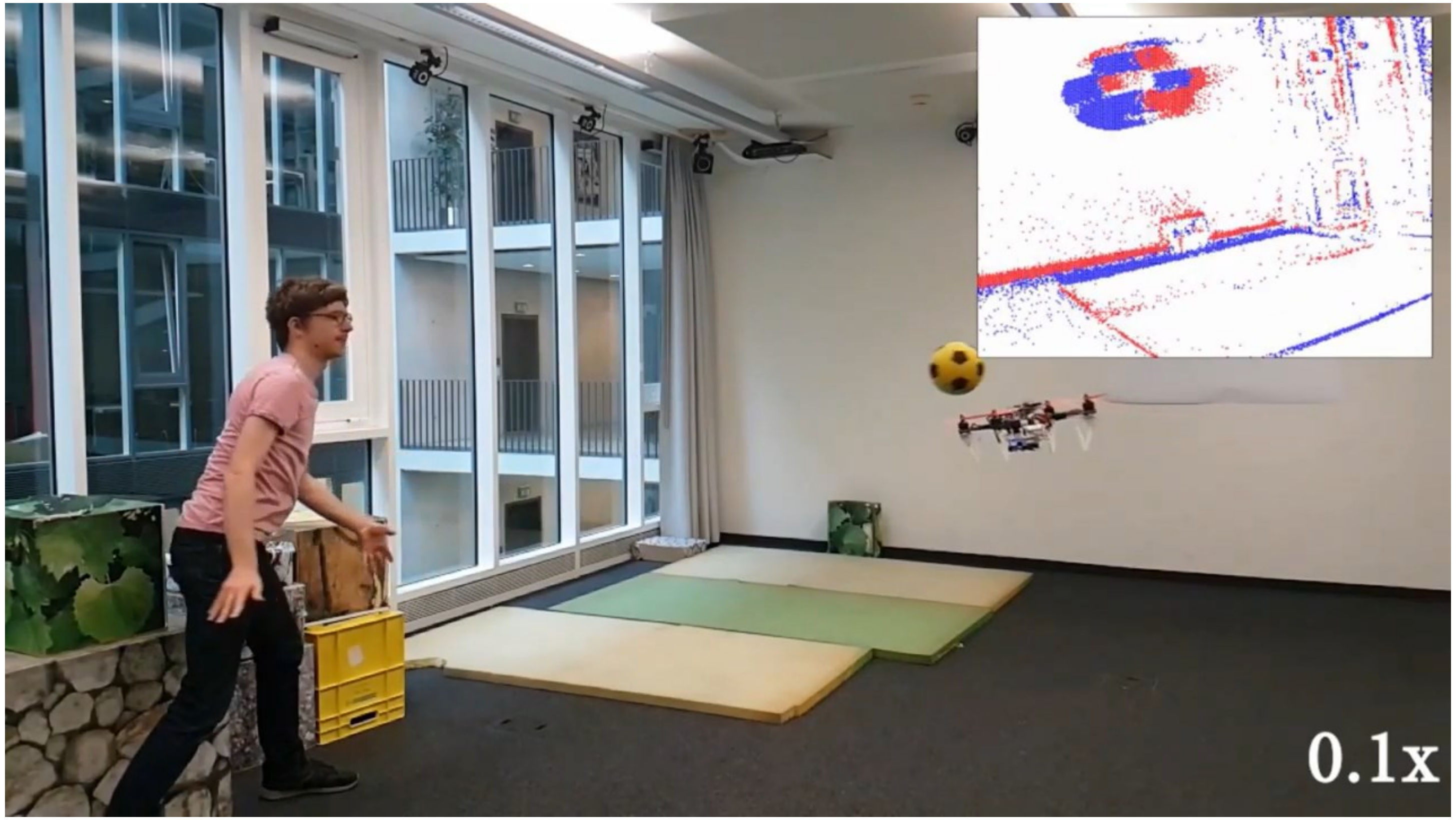
Application 2: Motion Segmentation



Stoffregen, Gallego, Drummond, Kleeman, Scaramuzza, *Motion Segmentation by Motion Compensation,* International Conference on Computer Vision (ICCV), 2019. [PDF](). [Video]().

Application 3: Dynamic Obstacle Avoidance

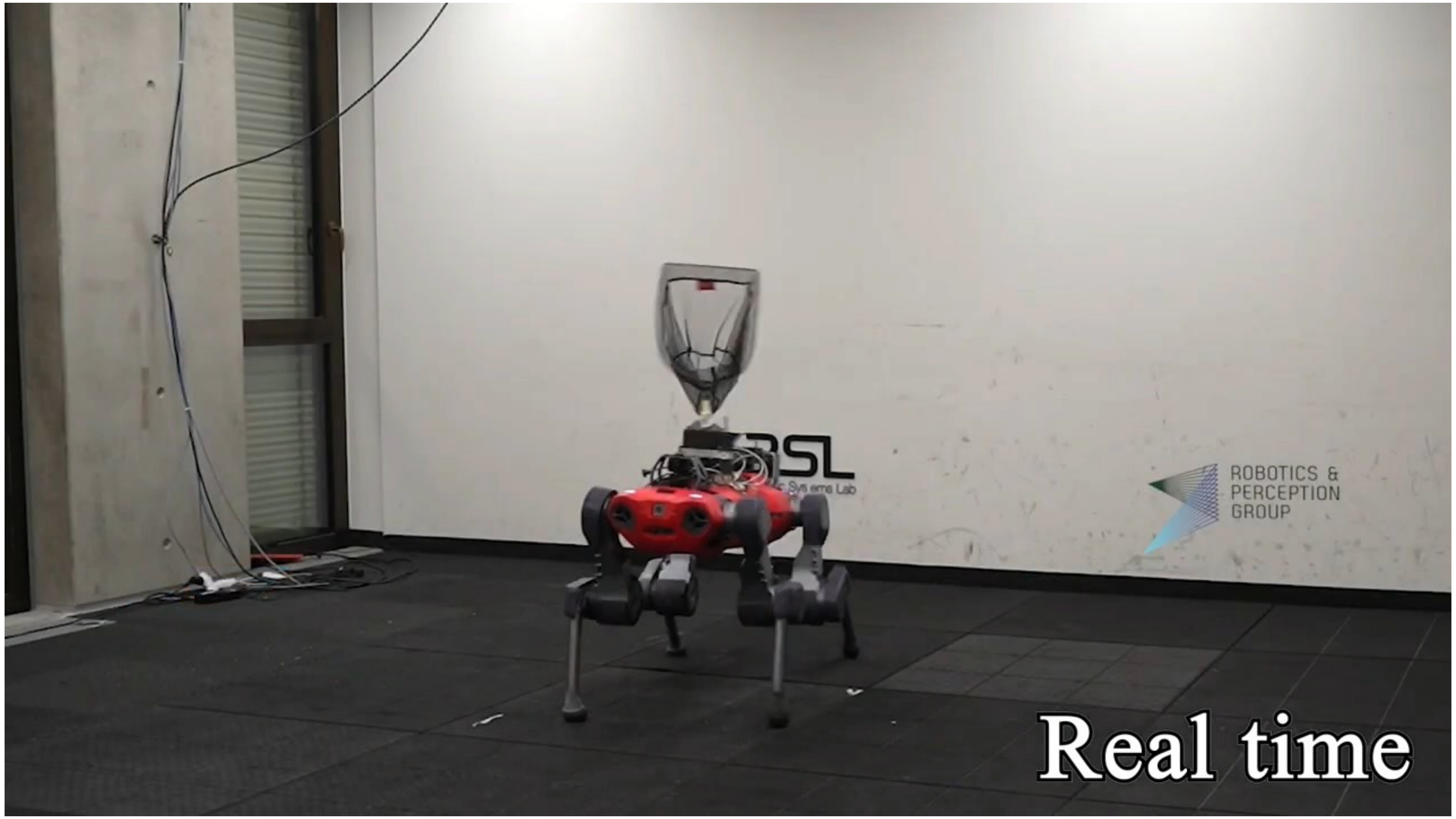
- Works with relative speeds of up to **10 m/s**
- Perception **latency: 3.5 ms**



Falanga, Kleber, Scaramuzza, *Dynamic Obstacle Avoidance for Quadrotors with Event Cameras*, Science Robotics**,** 2020. [PDF](). [Video]()

Catching Dynamic Objects

- Perception latency: **3.5 ms**
- Works with relative speeds of up to **15 m/s**



Forrai, Miki, Gehrig, Hutter, Scaramuzza, *Event-based Agile Object Catching with a Quadrupedal Robot*, **ICRA'23.** PDF. Video

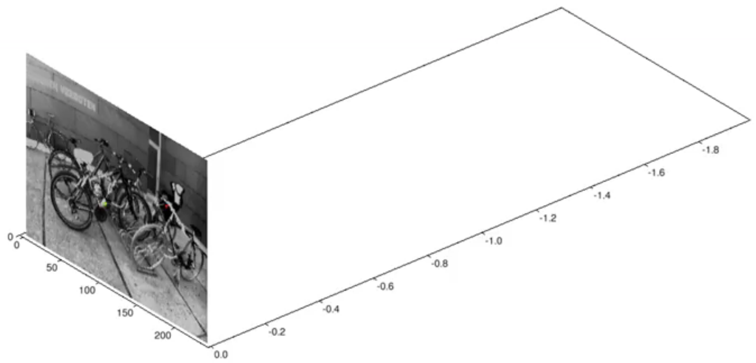
Application 4: "Ultimate SLAM"

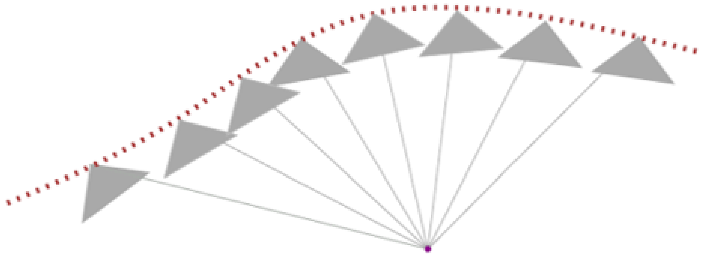
Goal: combining **events**, **images**, and **IMU** for robust visual SLAM in HDR and high speed scenarios

Front End:
Feature tracking from Events and Frames

→

Back-End
State-of-the-art
Non-linear-optimization-based VIO



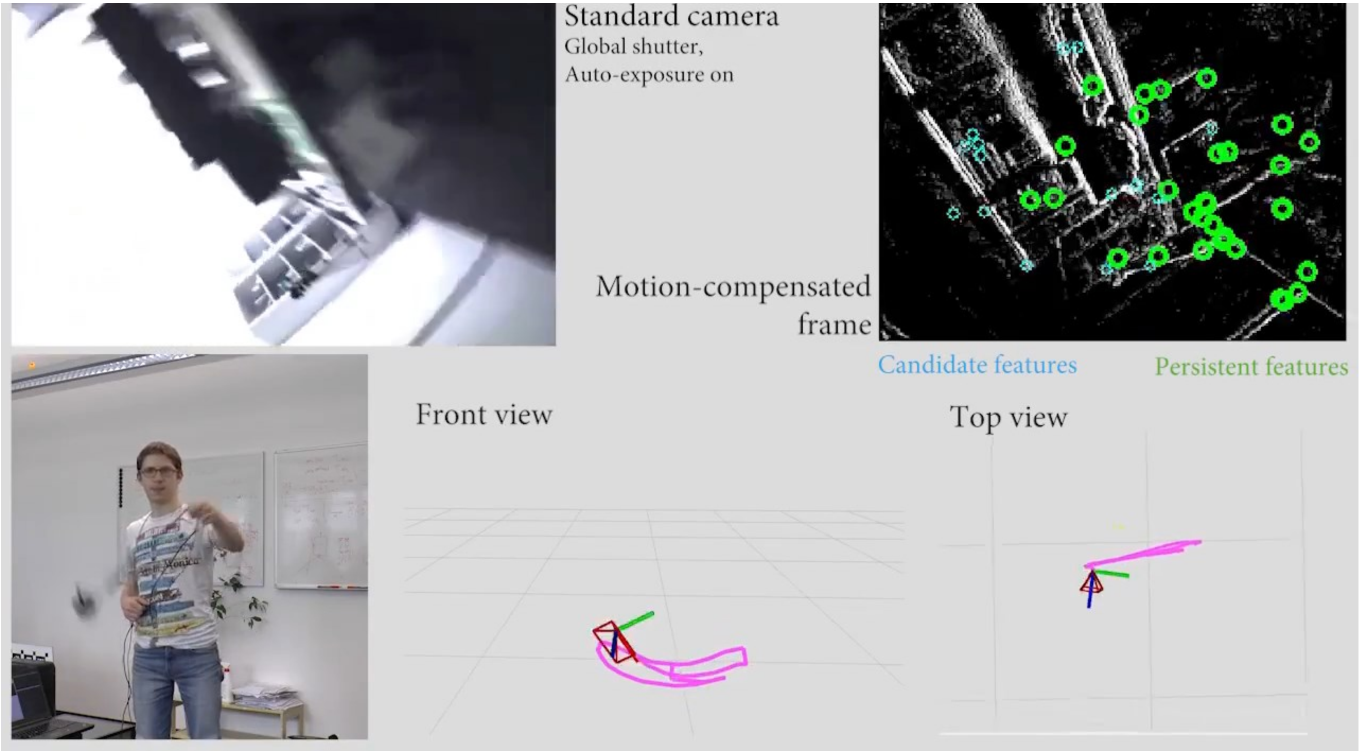


Rosinol-Vidal, Rebecq, Horstschaefer, Scaramuzza, *Ultimate SLAM? Combining Events, Images, and IMU for Robust Visual SLAM in HDR and High Speed Scenarios*, IEEE Robotics and Automation Letters (RAL), 2018 – PDF. Video. Best Paper Award Honorable Mention

Application 4: "Ultimate SLAM"

- **85%** accuracy gain over standard VIO in **HDR and high speed scenarios**

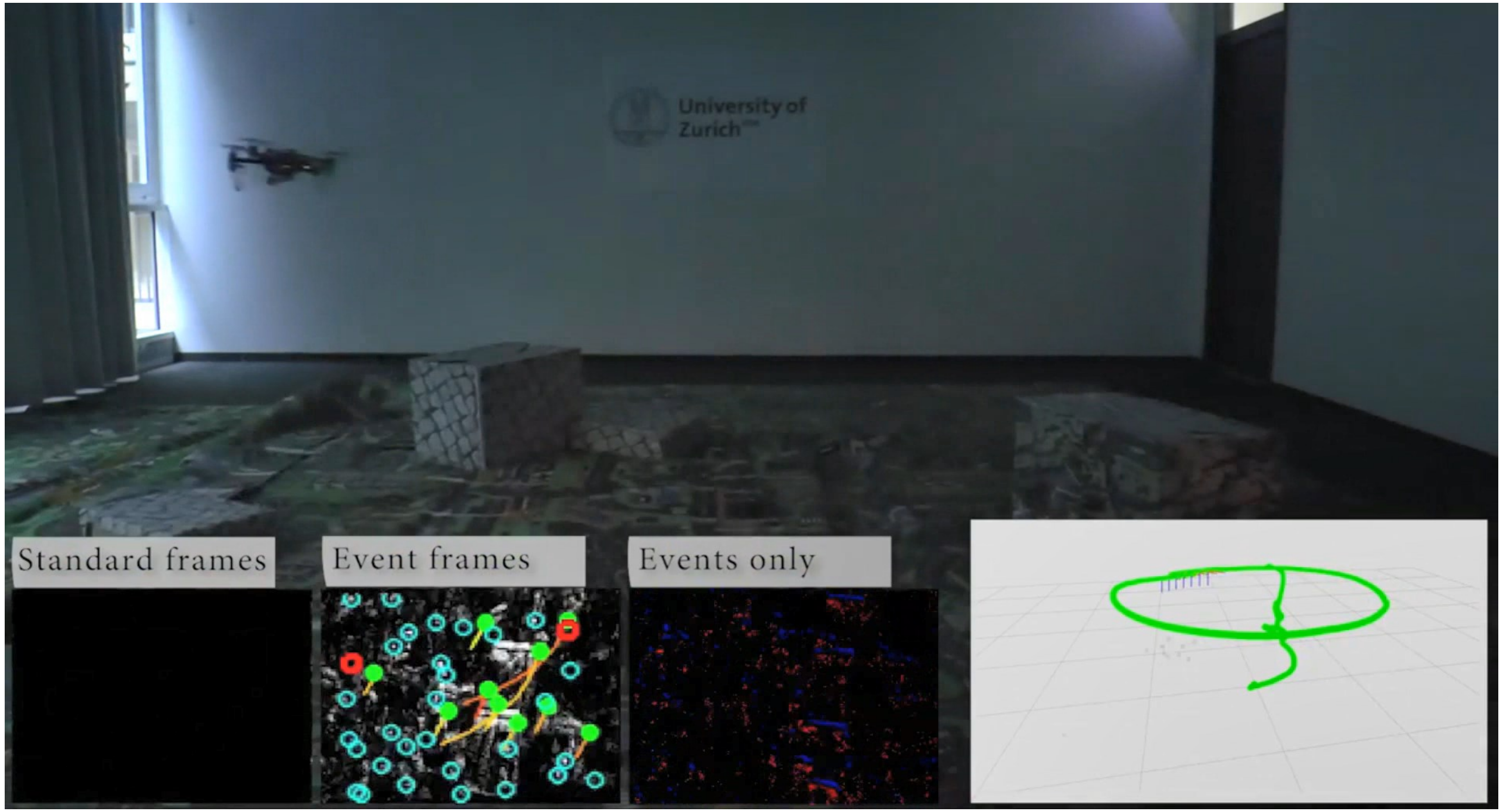
Standard camera | Event camera



Rosinol-Vidal, Rebecq, Horstschaefer, Scaramuzza, *Ultimate SLAM? Combining Events, Images, and IMU for Robust Visual SLAM in HDR and High Speed Scenarios*, IEEE Robotics and Automation Letters (RAL), 2018 – [PDF](). [Video](). Best Paper Award Honorable Mention

Application 5: Autonomous Navigation in Low Light

- UltimateSLAM running on board (CPU: Odroid XU4)



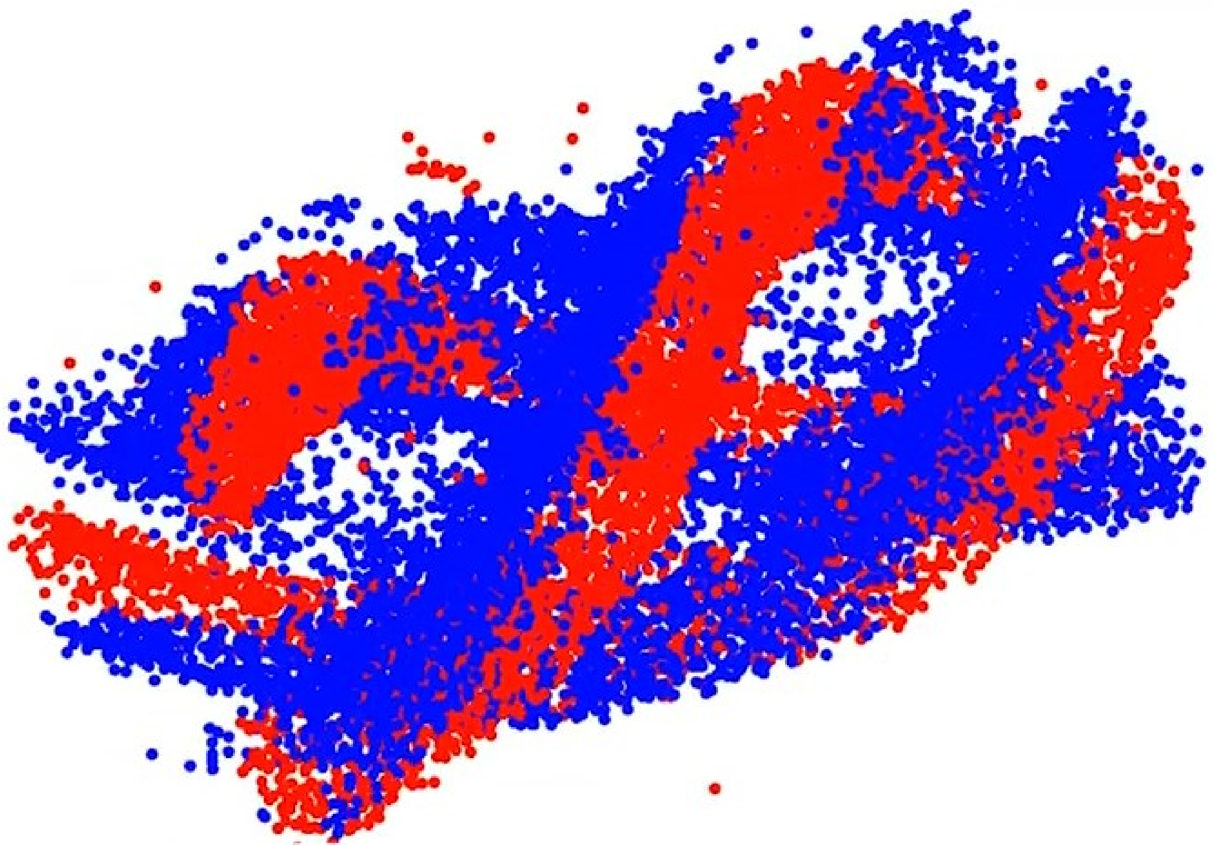
Rosinol-Vidal, Rebecq, Horstschaefer, Scaramuzza, *Ultimate SLAM? Combining Events, Images, and IMU for Robust Visual SLAM in HDR and High Speed Scenarios*, IEEE Robotics and Automation Letters (RAL), 2018 – PDF. Video. Best Paper Award Honorable Mention

Learning with Event Cameras

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

Input representation

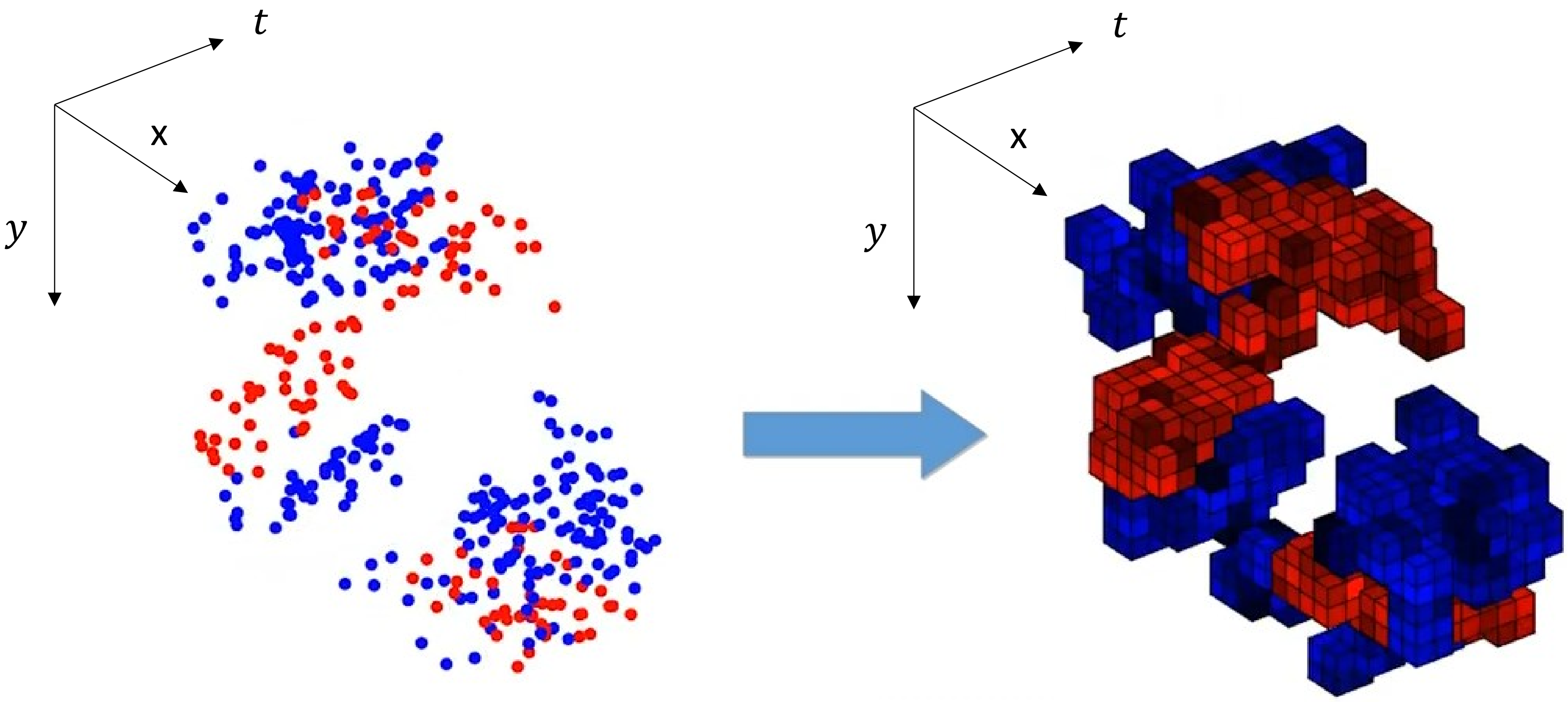
How do we pass sparse events into a convolutional neural network designed for standard images?



Video from here

Input representation

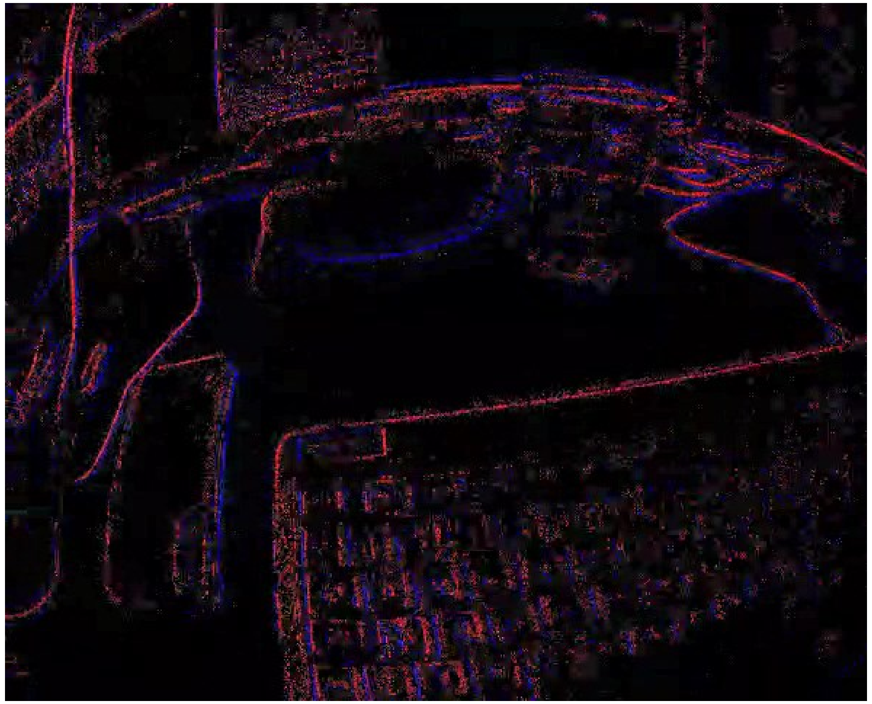
Represent events in space-time into a 3D voxel grid (x, y, t): each voxel contains sum of positive and negative events falling within the voxel



Video from [here]

Application 1: Image Reconstruction from Events

Events



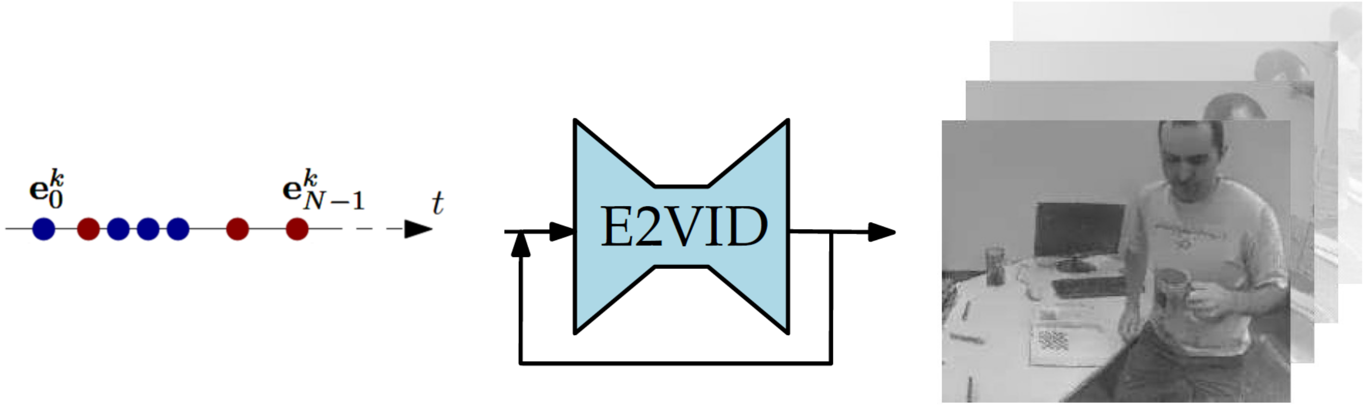
Reconstructed image from events

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

Rebecq, Ranftl, Koltun, Scaramuzza, *High Speed and High Dynamic Range Video with an Event Camera*, T-PAMI, 2019. PDF Video Code

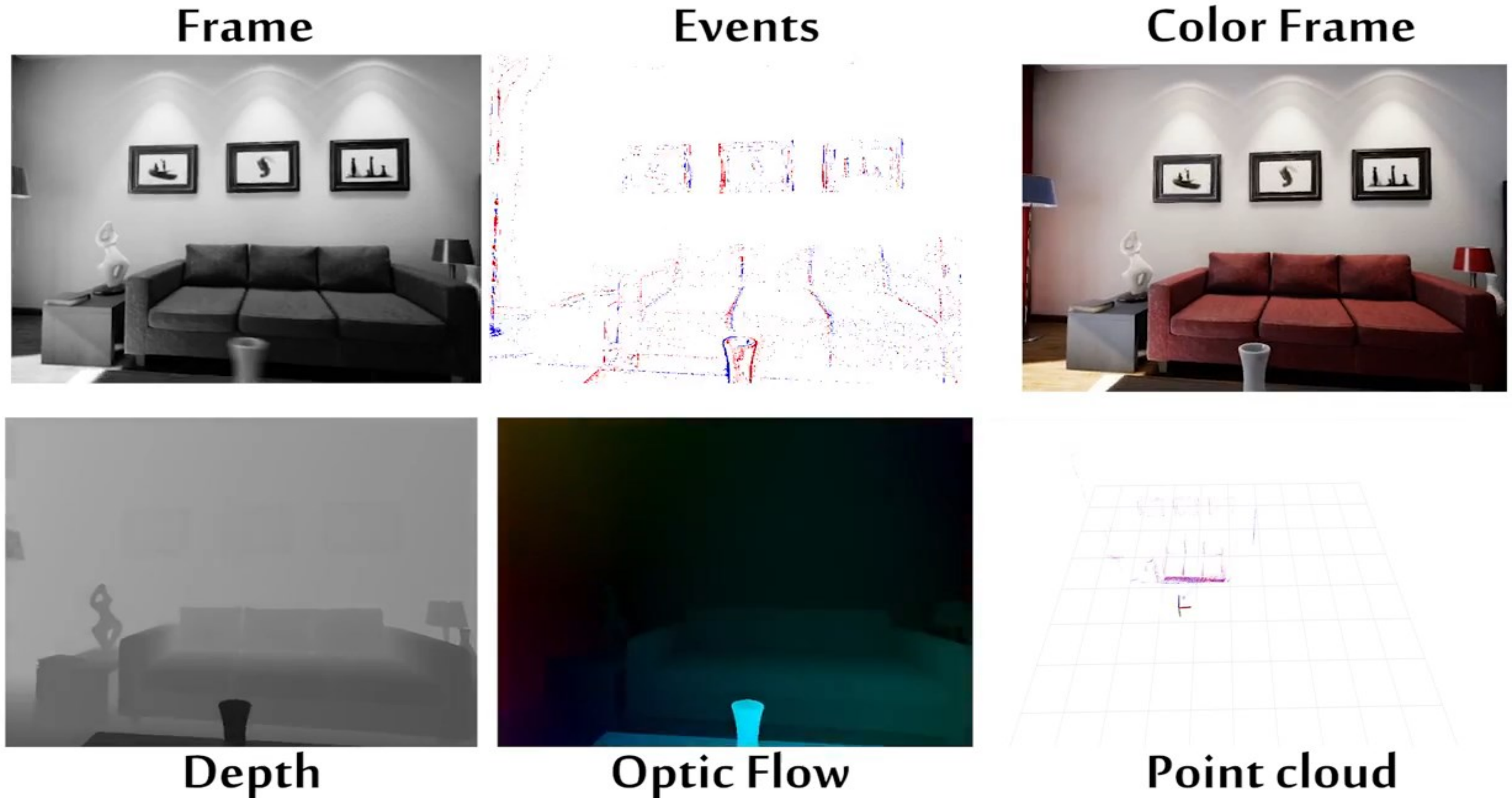
Overview

- **Recurrent neural network** (main module: Unet)
- Input: sequences of *event tensors* (3D spatio-temporal volumes of events[3])
- **Trained in simulation only**, without seeing a single real image
- To improve robustness **we randomize the contrast sensitivity** during simulation.
- Event camera simulator (ESIM): http://rpg.ifi.uzh.ch/esim.html



Rebecq, Ranftl, Koltun, Scaramuzza, *High Speed and High Dynamic Range Video with an Event Camera*, T-PAMI, 2019. PDF Video Code

ESIM: Event Camera Simulator



Open Source: http://rpg.ifi.uzh.ch/esim.html

Rebecq, Gehrig, Scaramuzza, *ESIM: an Open Event Camera Simulator*, Conference on Robot Learning (CORL), 2018. 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)

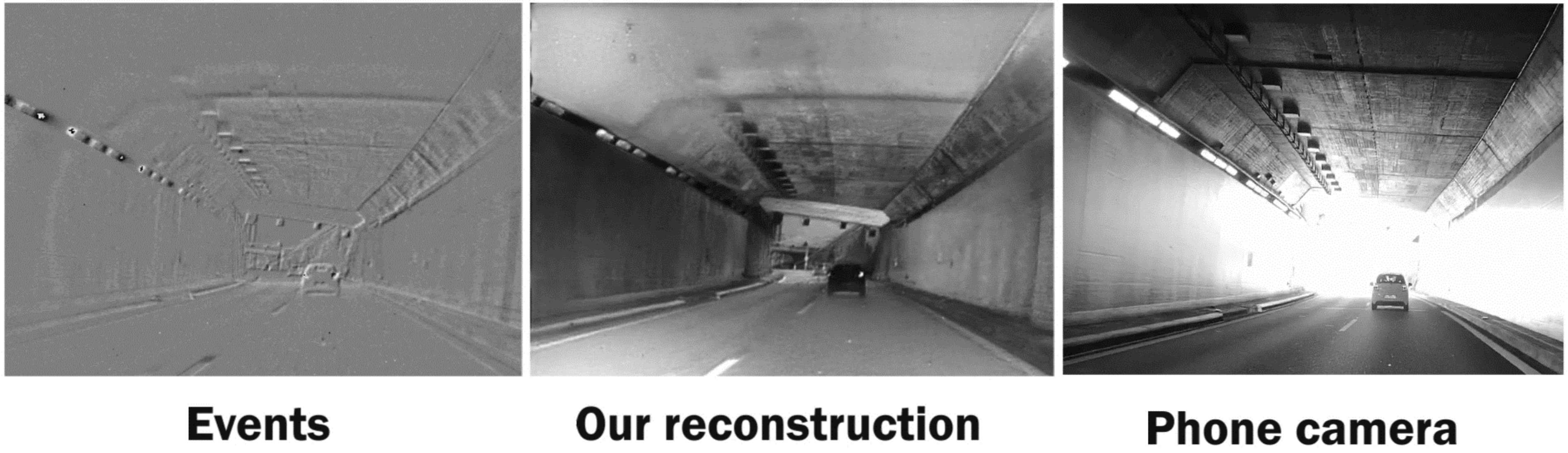
100 x slow motion

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

Rebecq, Ranftl, Koltun, Scaramuzza, *High Speed and High Dynamic Range Video with an Event Camera*, T-PAMI, 2019. PDF Video Code

HDR Video: Driving out of a tunnel

Recall: trained in simulation only!

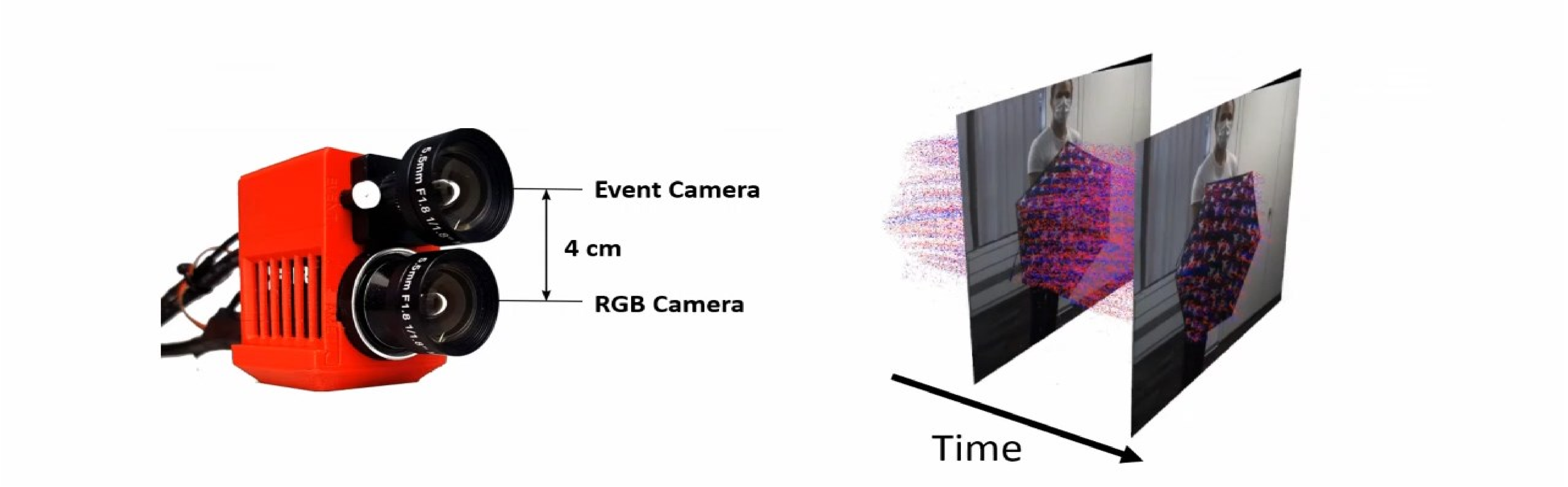


Code & datasets: https://github.com/uzh-rpg/rpg_e2vid

Rebecq, Ranftl, Koltun, Scaramuzza, *High Speed and High Dynamic Range Video with an Event Camera*, T-PAMI, 2019. PDF Video Code

Application 2: Slow Motion Video

- We can combine an event camera with an HD RGB camera
- We use events to **upsample low-framerate video** by over **50 times** with **only 1/40th of the memory** footprint!

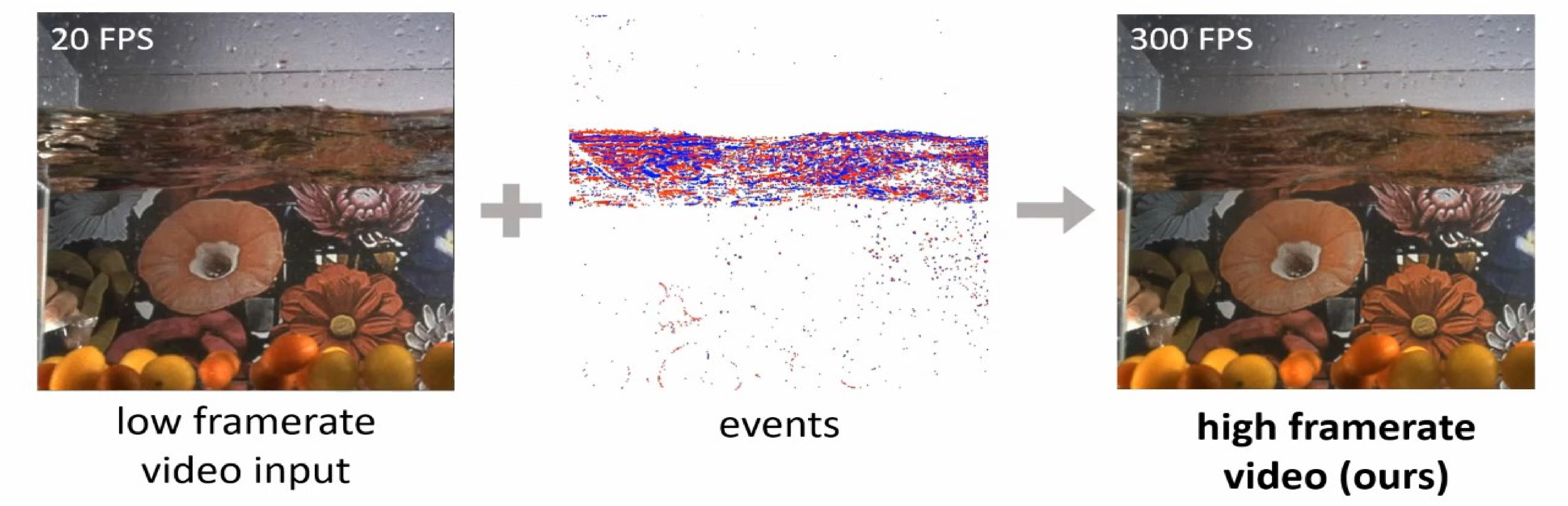


Code & Datasets: http://rpg.ifi.uzh.ch/timelens

Tulyakov et al., TimeLens: *Event-based Video Frame Interpolation*, CVPR'21. PDF. Video. Code.

Application 2: Slow Motion Video

- We can combine an event camera with an HD RG camera
- We use events to **upsample low-framerate video** by over **50 times** with **only 1/40th of the memory** footprint!

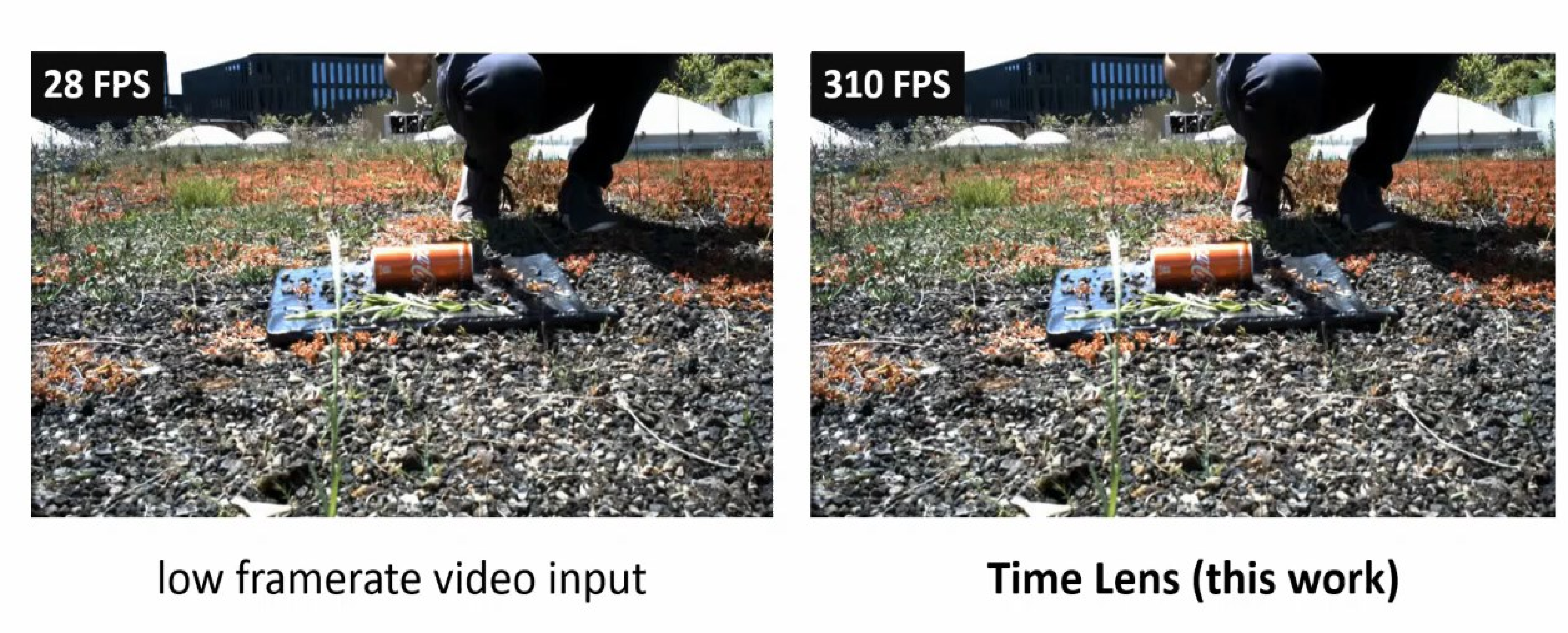


Code & Datasets: http://rpg.ifi.uzh.ch/timelens

Tulyakov et al., TimeLens: *Event-based Video Frame Interpolation*, CVPR'21. PDF. Video. Code.

Application 2: Slow Motion Video

- We can combine an event camera with an HD RG camera
- We use events to **upsample low-framerate video** by over **50 times** with **only 1/40th of the memory** footprint!



Code & Datasets: http://rpg.ifi.uzh.ch/timelens

Tulyakov et al., TimeLens: *Event-based Video Frame Interpolation*, CVPR'21. PDF. Video. Code.

The Evolution of Event Cameras

2008 — First event camera by University of Zurich
Resolution: **128×128 pxl**
Pixel size: **40 microns**

2014 — First event camera commercialized by IniVation
Resolution: **640×480 pxl**
Pixel size: **15 microns**

2019

United States Patent

EVENT CAMERA FOR GENERATION OF EVENT-BASED IMAGES
Applicant: Facebook Technologies, LLC, Menlo Park, CA (US)

METHOD AND DEVICE FOR EYE TRACKING USING EVENT CAMERA DATA
Applicant: Apple Inc., Cupertino, CA (US)

2021 — SONY, SAMSUNG, OmniVision

First Full-HD event sensors:

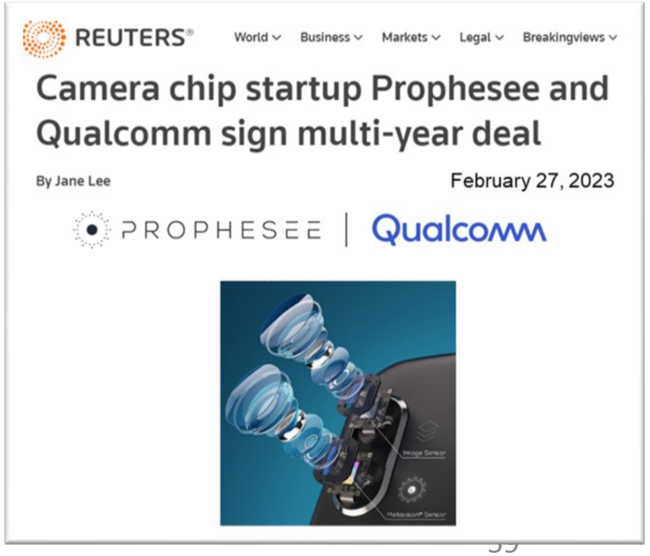
Resolution: **1280×720 pxl**
Pixel size: **5 microns**

2022 — Meta

Meta opens Event-based Sensing Lab

First event cameras in space

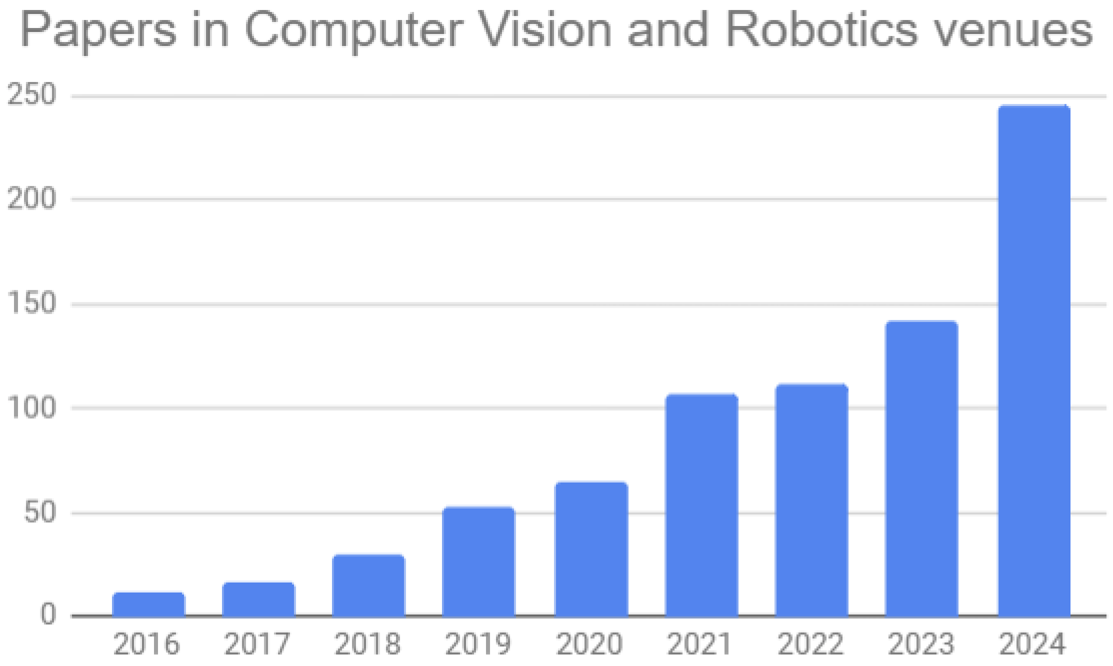
2023


REUTERS® World Business Markets Legal Breakingviews

Camera chip startup Prophesee and Qualcomm sign multi-year deal

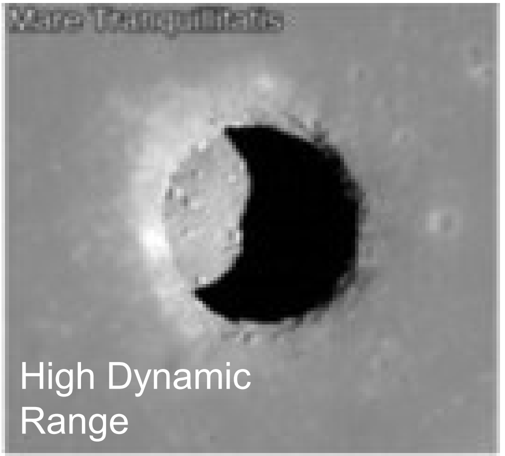
By Jane Lee February 27, 2023

PROPHESEE | Qualcomm



Collaboration with NASA for future space missions

- Future planetary astrobiology missions aim at using drones for the exploration of lava tunnels as a priority objective for investigations
- Lava tunnels host ice, which potentially hosts life
- Lava tunnels can be used as shelters for future Mars missions
- More info here

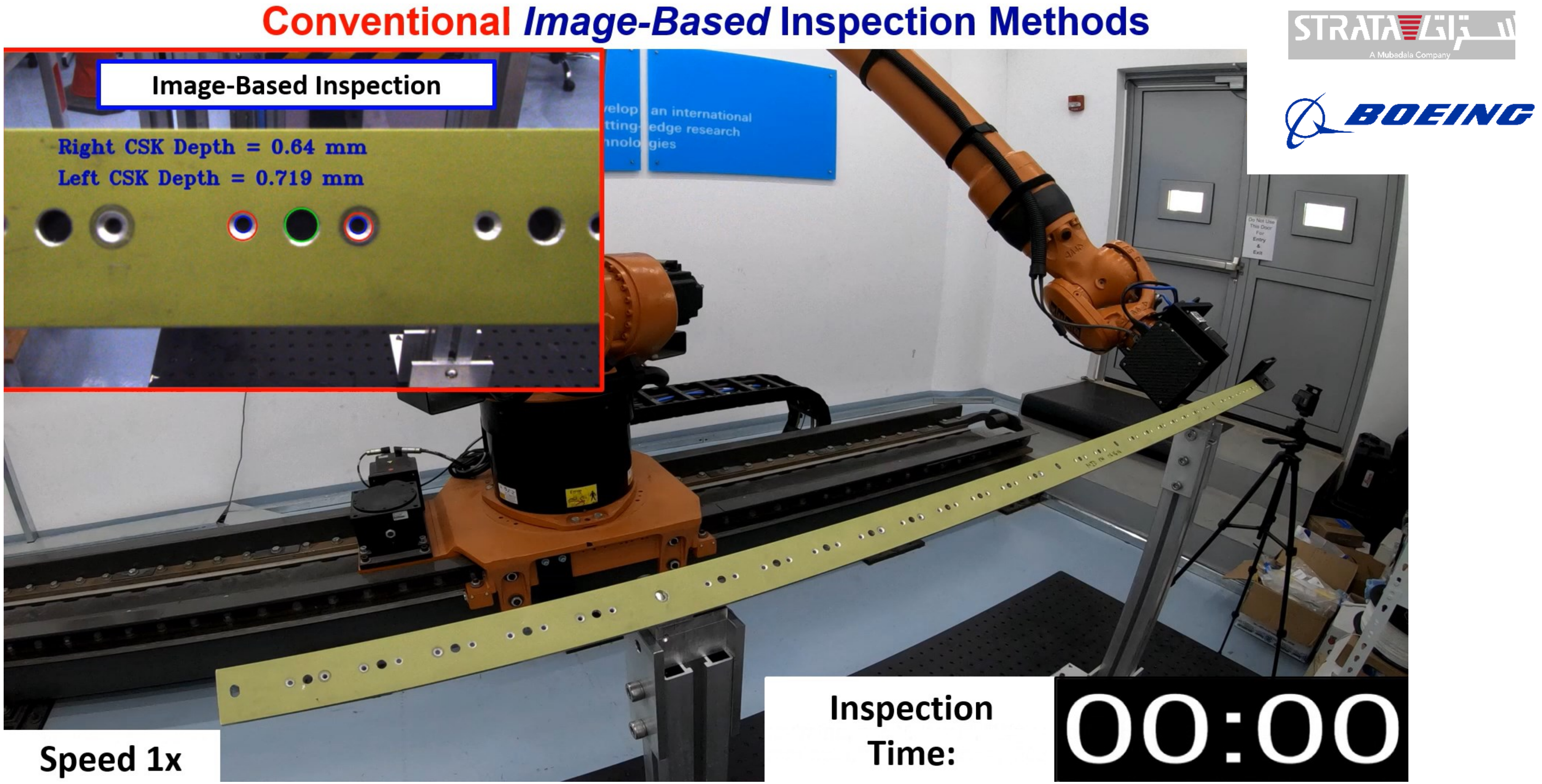


Joint paper with **NASA JPL**:

Mahlknecht, Gehrig, Nash, Rockenbauer, Morrell, Delaune, Scaramuzza

Exploring Event Camera-based Odometry for Planetary Robots, **RAL'22**. PDF. Data & Code

Application 5: High-Speed Inspection of Countersinks



Salah et al, Zweiri, *High speed neuromorphic vision-based inspection of countersinks in automated manufacturing processes*, **Journal of Intelligent Manufacturing**

Other Applications

5,000 fps

Tulyakov, Gehrig, et al., TimeLens: *Event-based Video Frame Interpolation*, CVPR’21

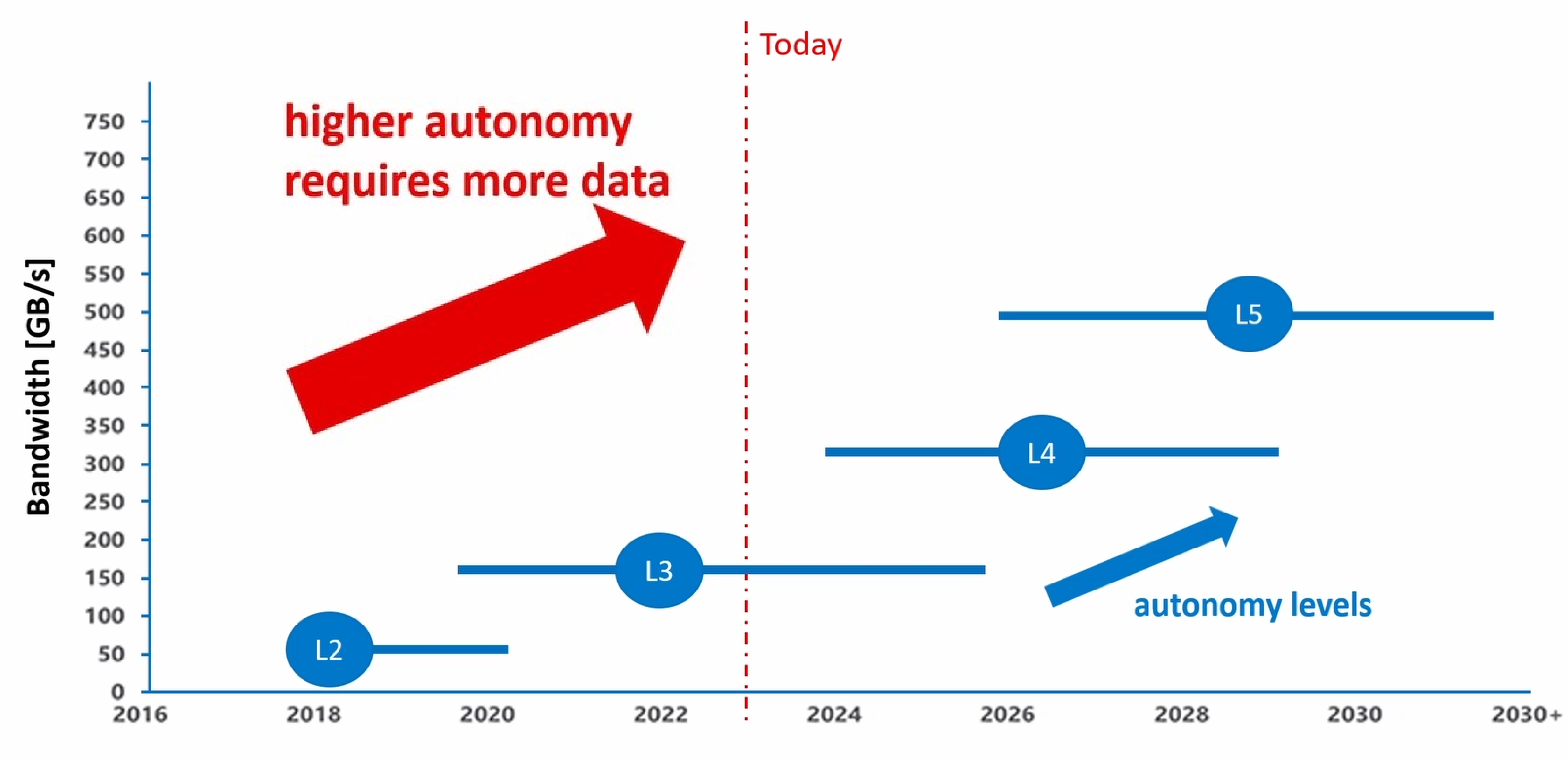
Application 2: Deblurring a Blurry Video

Credit: Prophesee

Advanced Driver Assistance Systems (ADAS)

Tesla Vision System

Memory Bandwidth Requirements by ADAS level

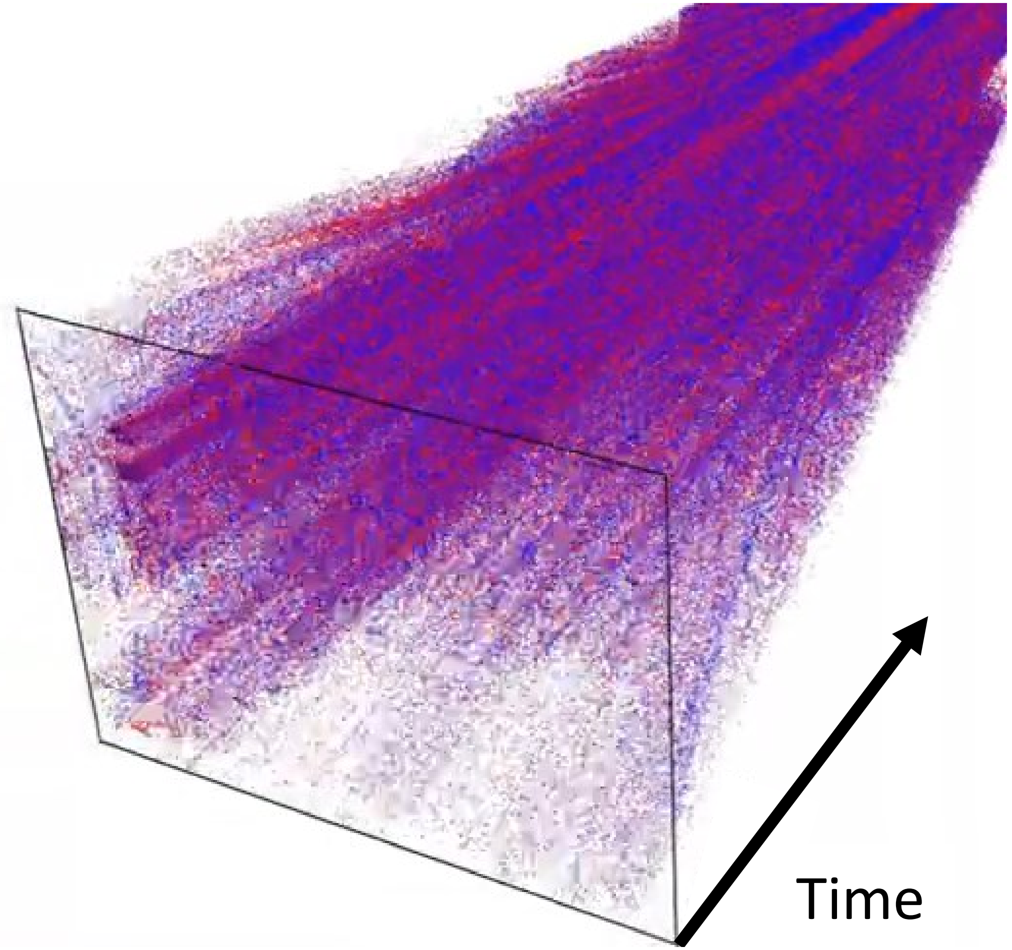


https://www.electronicspecifier.com/industries/automotive/pushing-the-envelope-for-adas-with-advanced-memory-technologies

Can we transfer this to Automotive?

Standard camera

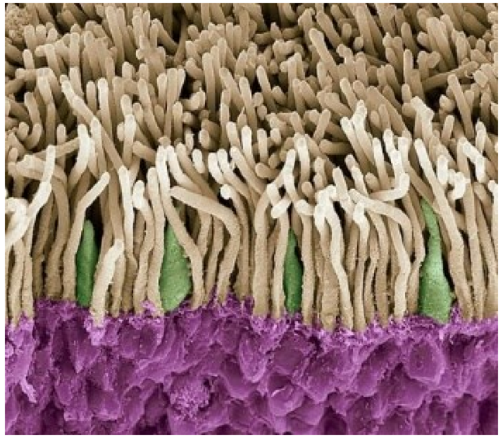
Event camera

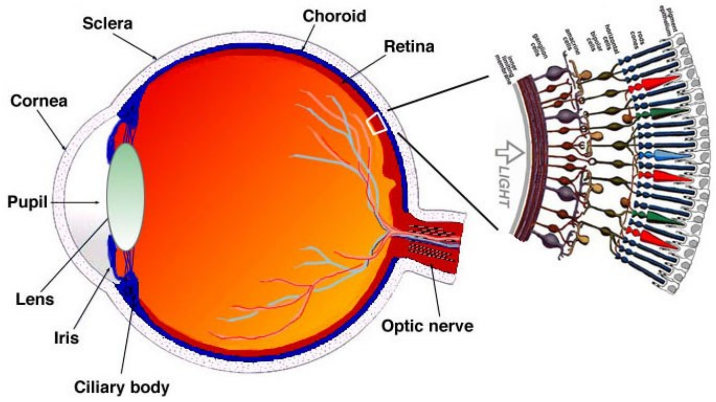


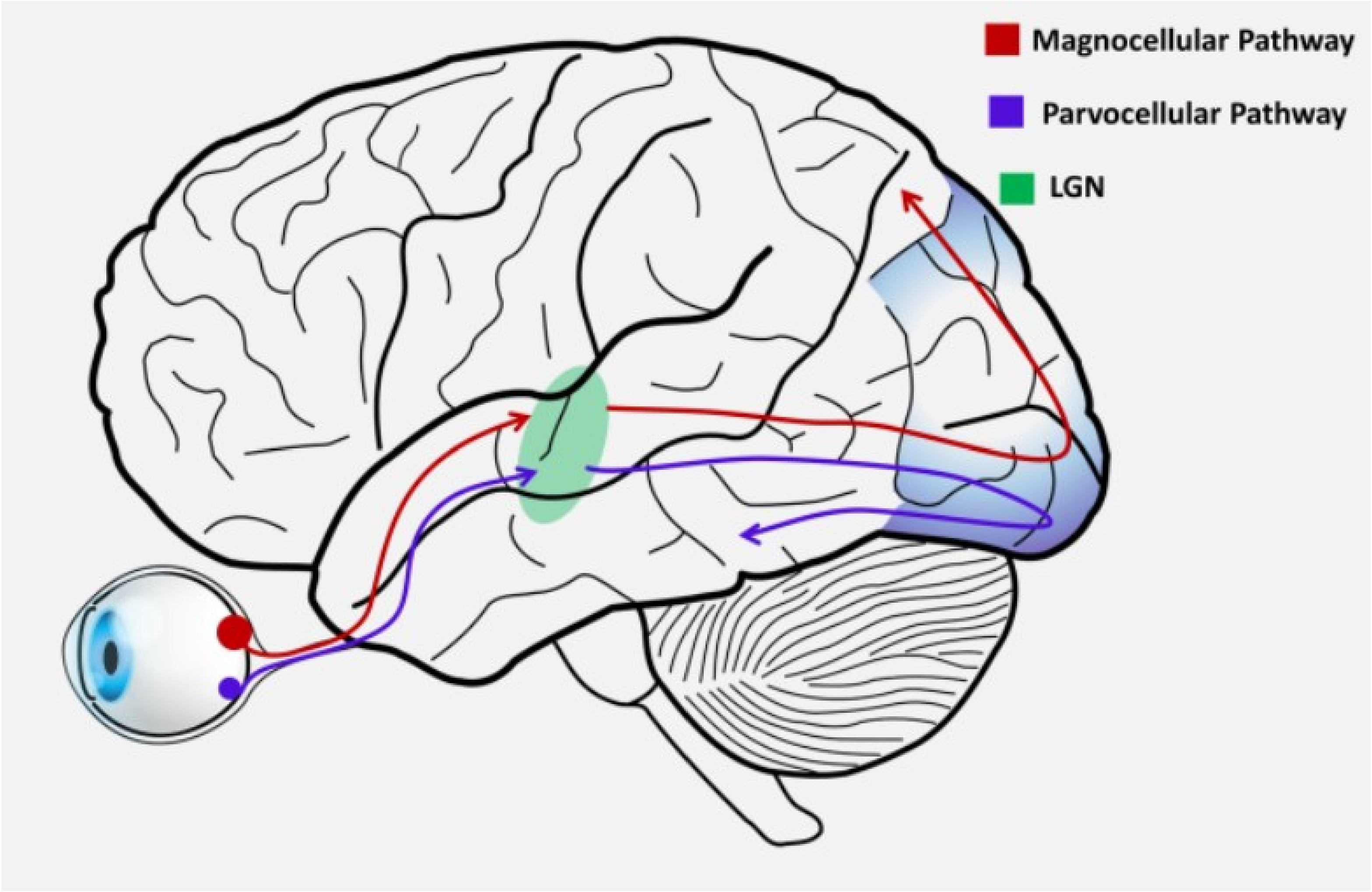
Low Latency Automotive Vision

Gehrig, Scaramuzza, *Low Latency Automotive Vision with Event Cameras*, **Nature**, 2024

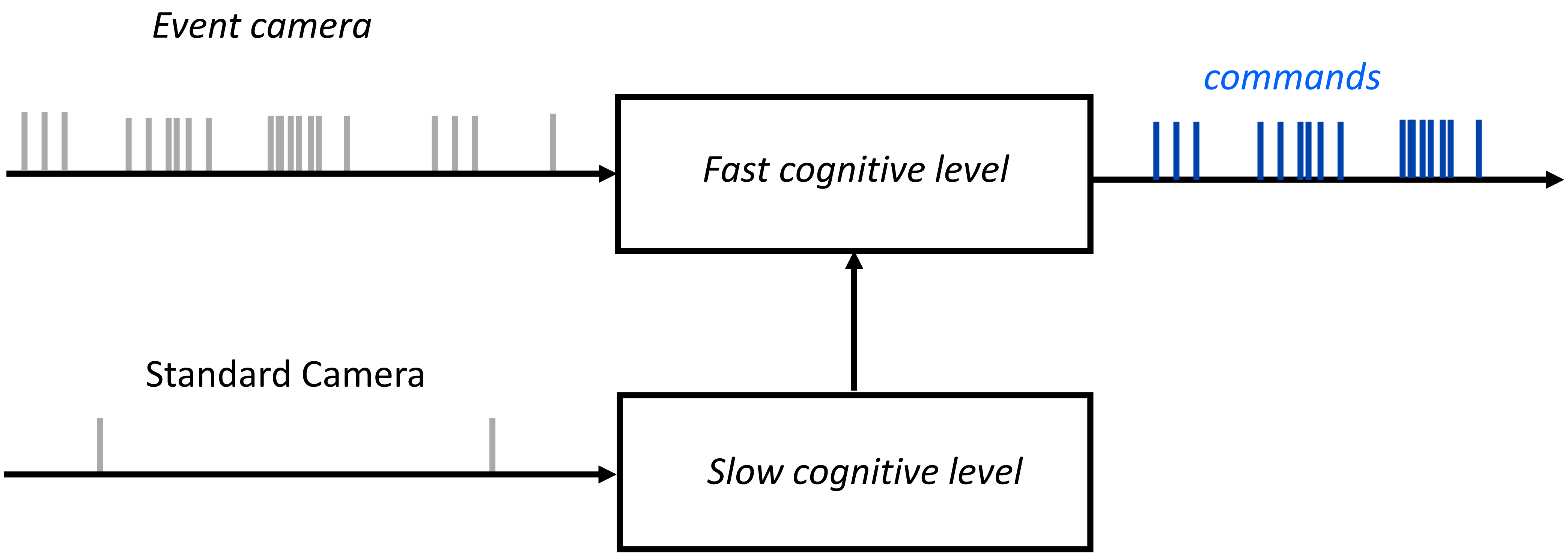
Magno and Parvo Pathways of the Primate Visual System







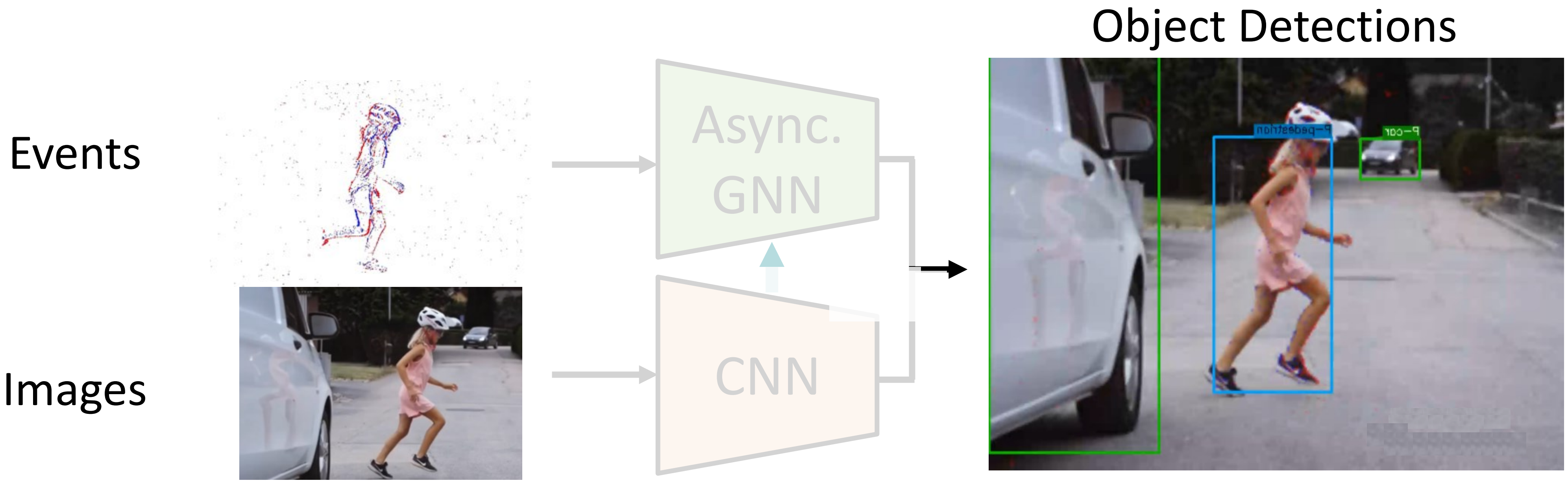
Hybrid Asynchronous Object Detection



Gehrig, Scaramuzza, *Low Latency Automotive Vision with Event Cameras*, **Nature**, 2024

Hybrid Asynchronous Object Detection

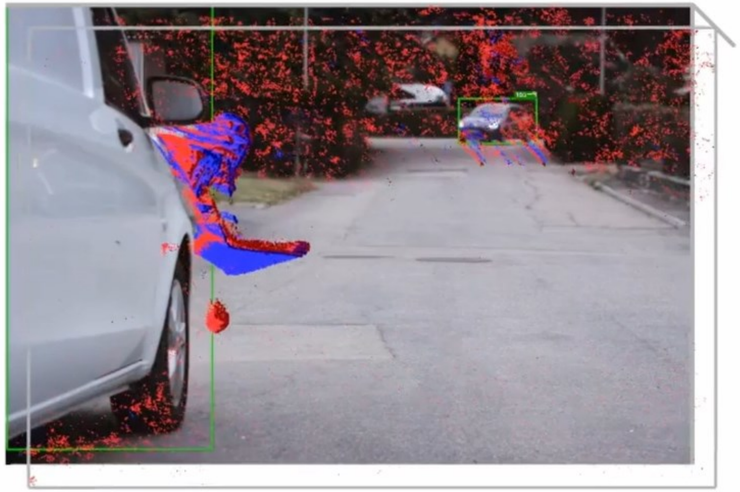
Use a CNN to provide image features to an **asynchronous object detection network.** These features are reused **asynchronously, and thus enable object detection in the blind-time between frames**



This enables early object detection, which cuts down perceptual latency!

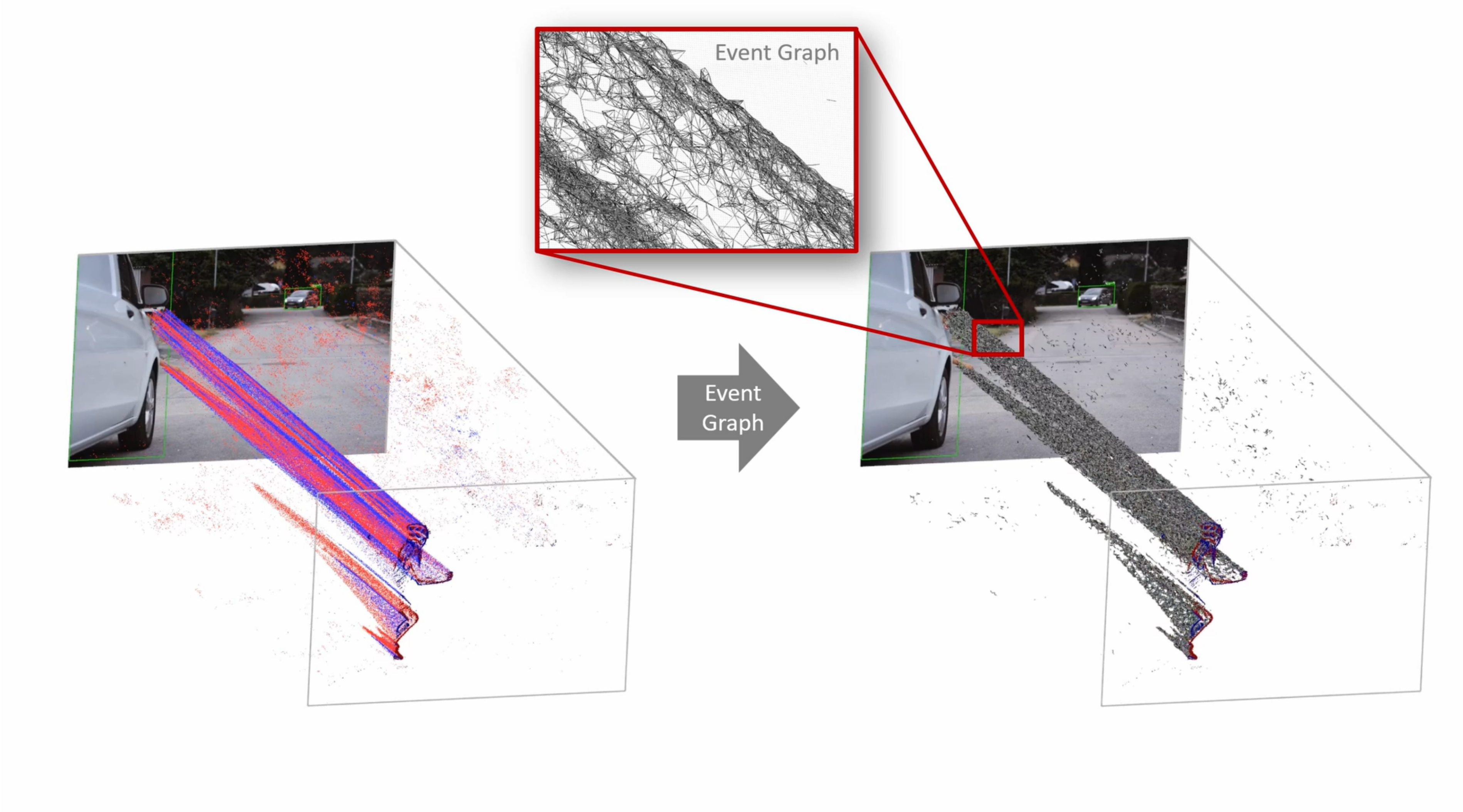
Gehrig, Scaramuzza, *Low Latency Automotive Vision with Event Cameras*, **Nature**, 2024

Low Latency Automotive Vision



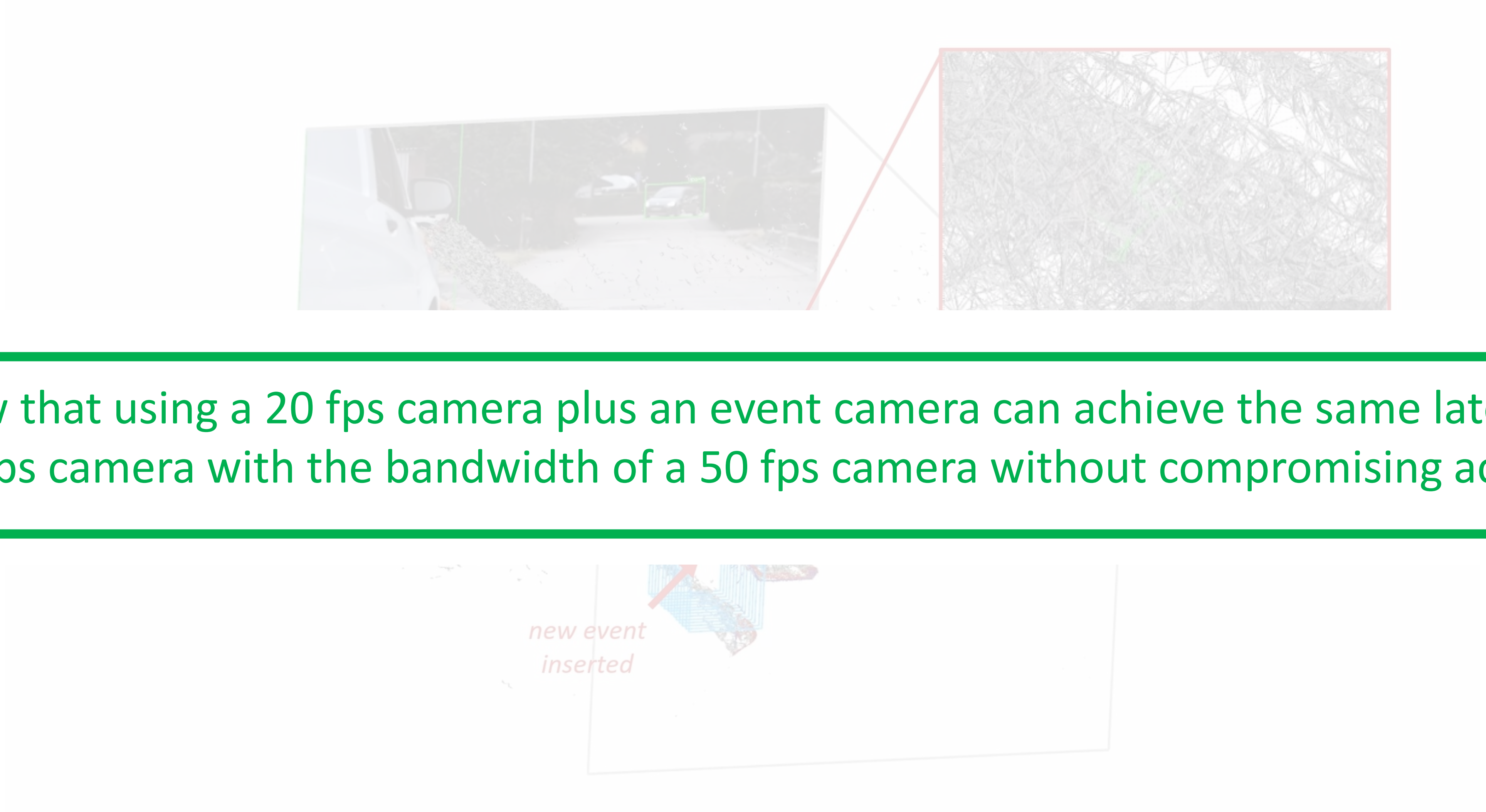
Gehrig, Scaramuzza, *Low Latency Automotive Vision with Event Cameras*, **Nature**, 2024

Low Latency Automotive Vision



Gehrig, Scaramuzza, *Low Latency Automotive Vision with Event Cameras*, **Nature**, 2024

Low Latency Automotive Vision



We show that using a 20 fps camera plus an event camera can achieve the same latency as a 5,000 fps camera with the bandwidth of a 50 fps camera without compromising accuracy.

Gehrig, Scaramuzza, *Low Latency Automotive Vision with Event Cameras*, **Nature**, 2024

Readings

- **Tutorial** paper:
  Gallego, Delbruck, Orchard, Bartolozzi, Taba, Censi, Leutenegger, Davison, Conradt, Daniilidis, Scaramuzza, **Event-based Vision: A Survey**, IEEE Transactions of Pattern Analysis and Machine Intelligence, 2020. 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 event camera research: http://rpg.ifi.uzh.ch/research_dvs.html

Understanding Check

Are you able to answer the following questions?

- What is an event camera 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?
- What is the generative model of an event camera (formula). Can you derive its 1^{st} order approximation?
- Could you intuitively explain why we can reconstruct the intensity from a grayscale frame plus events and from events alone? What are the assumption? What are the failure modes?
- 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?