University of Zurich[UZH] | ETH zürich
Institute of Informatics – Institute of Neuroinformatics
ROBOTICS & PERCEPTION GROUP

Tutorial on Event-based Cameras:

Davide Scaramuzza

Code, datasets, simulators, papers, and videos:

http://rpg.ifi.uzh.ch/research_dvs.html

Research updates:

@davsca1

@davidescaramuzza

Reference

Event-based Vision: A Survey

Guillermo Gallego, Tobi Delbrück, Garrick Orchard, Chiara Bartolozzi, Brian Taba, Andr
Stefan Leutenegger, Andrew Davison, Jörg Conradt, Kostas Daniilidis, Davide Scaramuzza

Abstract— Event cameras are bio-inspired sensors that work radically different from traditional cameras. Instead of capturing images at a fixed rate, they measure per-pixel brightness changes asynchronously. This results in a stream of events, which encode the time, location and sign of the brightness changes. Event cameras posses outstanding properties compared to traditional cameras: very high dynamic range ($140\,\mathrm{dB}$ vs. $60\,\mathrm{dB}$), high temporal resolution (in the order of $\mu\mathrm{s}$), low power consumption, and do not suffer from motion blur. Hence, event cameras have a large potential for robotics and computer vision in challenging scenarios for traditional cameras, such as high speed and high dynamic range. However, novel methods are required to process the unconventional output of these sensors in order to unlock their potential. This paper provides a comprehensive overview of the emerging field of event-based vision, with a focus on the applications and the algorithms developed to unlock the outstanding properties of event cameras. We present event cameras from their working principle, the actual sensors that are available and the tasks that they have been used for, from low-level vision (feature detection and tracking, optic flow, etc.) to high-level vision (reconstruction, segmentation, recognition). We also discuss the techniques developed to process events, including learning-based techniques, as well as specialized processors for these novel sensors, such as spiking neural networks. Additionally, we highlight the challenges that remain to be tackled and the opportunities that lie ahead in the search for a more efficient, bio-inspired way for machines to perceive and interact with the world.

Index Terms—Event Cameras, Bio-Inspired Vision, Asynchronous Sensor, Low Latency, High Dynamic Range, Low Power.

1 Introduction and Applications

"THE *brain is imagination, and that was exciting to me; I wanted to build a chip that could imagine something*[1]." that is how Misha Mahowald, a graduate student at Caltech in 1986, started to work with Prof. Carver Mead on the stereo problem from a joint biological and engineering perspective. A couple of years later, in 1991, the image of a cat in

as well as new computer vision and robotic tasks. Sight is, by far, the dominant sense in humans to perceive the world, and, together with the brain, learn new things. In recent years, this technology has attracted a lot of attention from both academia and industry. This is due to the availability of prototype event cameras and the advantages that these devices offer to tackle problems that are currently unfeasible

http://rpg.ifi.uzh.ch/docs/EventVisionSurvey.pdf

Open Challenges in Computer Vision

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

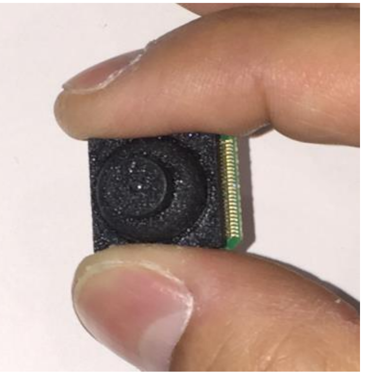
Latency

Motion blur

Dynamic Range

Event cameras do not suffer from these problems!

What is an event camera?

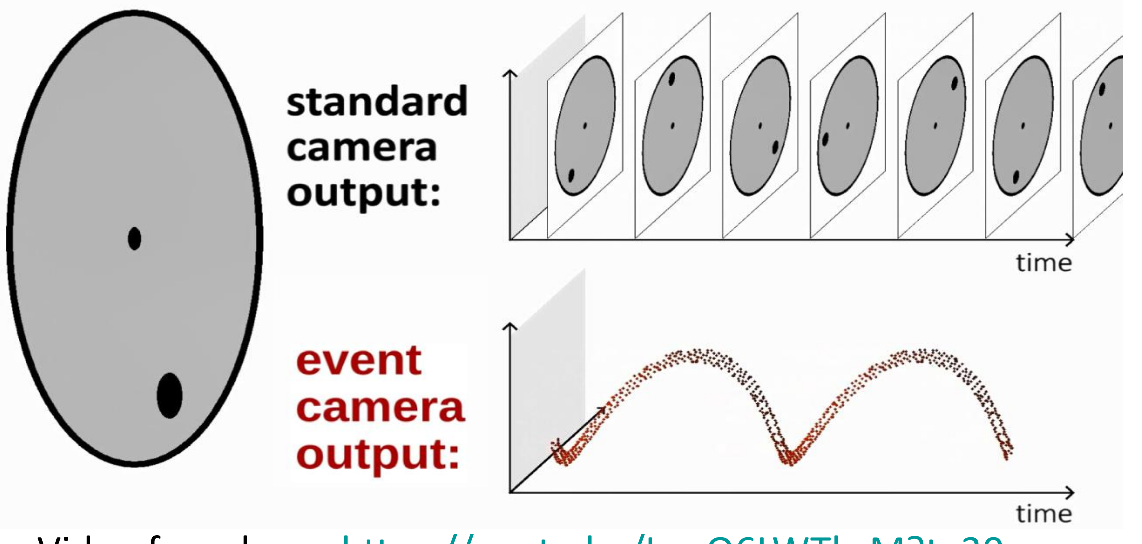


Mini DVS sensor from IniVation.com

- Novel sensor that measures only **motion in the scene**
- **First commercialized in 2008** by T. Delbruck (UZH&ETH) under the name of Dynamic Vision Sensor (DVS)
- **Low-latency** (~ 1 μs)
- **No motion blur**
- **High dynamic range** (140 dB instead of 60 dB)
- **Ultra-low power** (mean: 1mW vs 1W)

Traditional vision algorithms cannot be used because:

- **Asynchronous** pixels
- **No intensity information** (only binary intensity changes)



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

What is an event camera?

- Novel sensor that measures only **motion in the scene**
- **First commercialized in 2008** by T. Delbruck (UZH&ET
  under the name of Dynamic Vision Sensor (DVS)
- **Low-latency** (~ 1 μs)
- **No motion blur**
- **High dynamic range** (140 dB instead of 60 dB)
- **Ultra-low power** (mean: 1mW vs 1W)

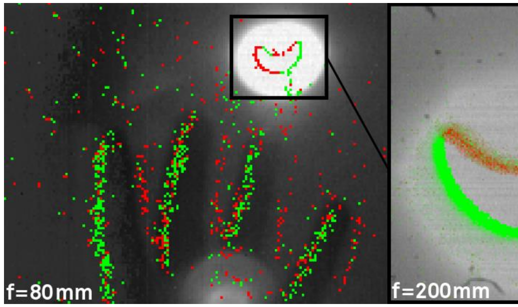
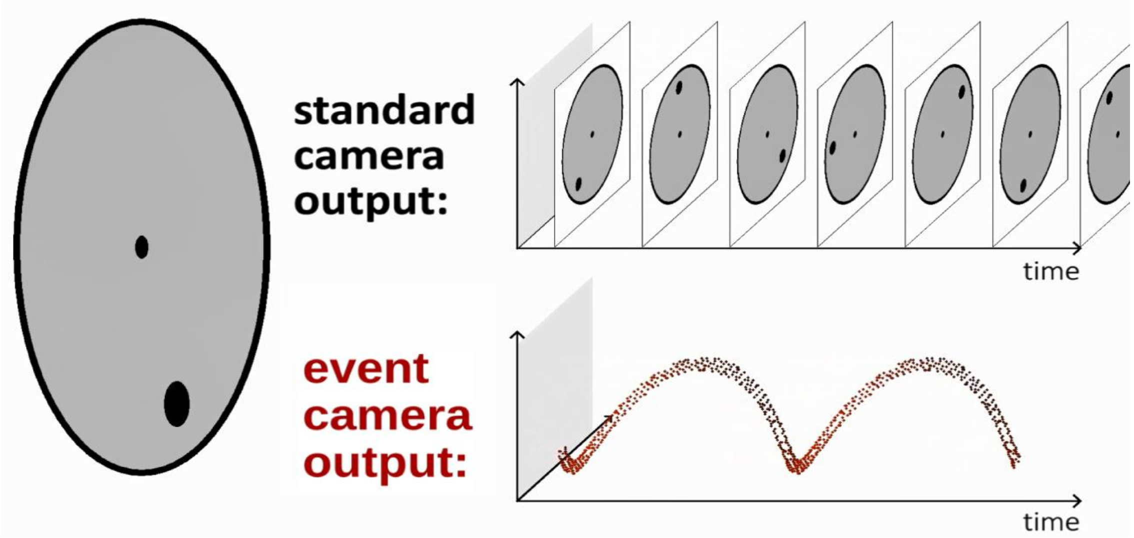


Image of the solar eclipse captured by a DVS

Traditional vision algorithms cannot be used because:

- **Asynchronous** pixels
- **No intensity information** (only binary intensity changes)

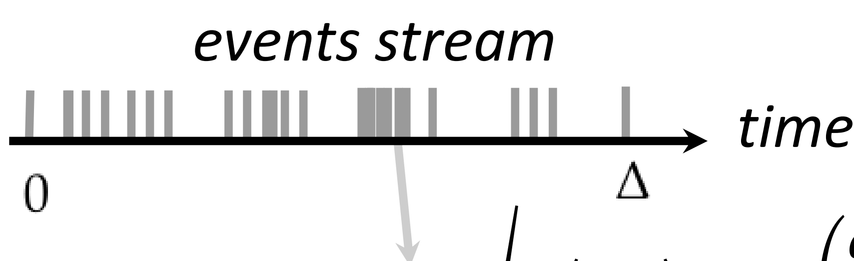


Video from here: https://youtu.be/LauQ6LWTkxM?t=30

Camera vs Event Camera

- 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 an intensity changes value



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

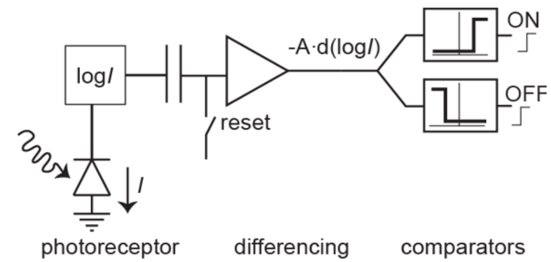
Timestamp (μs)

Pixel coordinates

Event polarity (or sign) (-1 *or* 1): *increase or decrease of brightness*

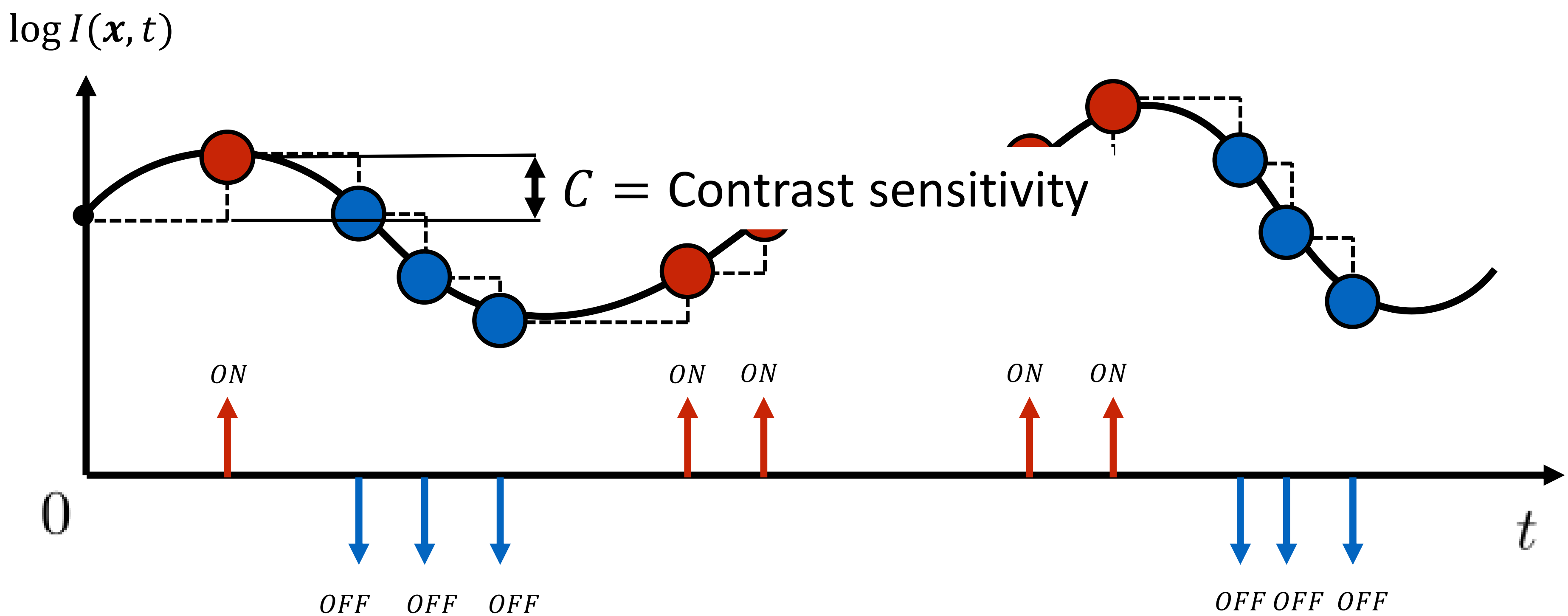
[Lichtsteiner, Posch, Delbruck, A 128x128 120 dB 15µs Latency Asynchronous Temporal Contrast Vision Sensor, 2008

Generative Event Model



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

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



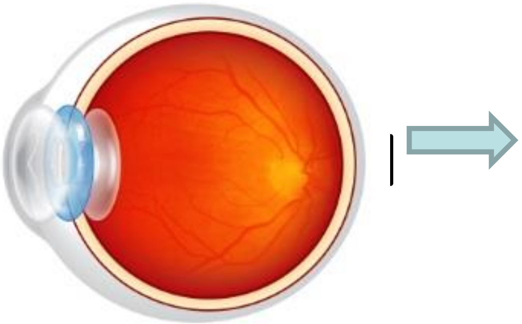
Events are triggered **asynchronously**

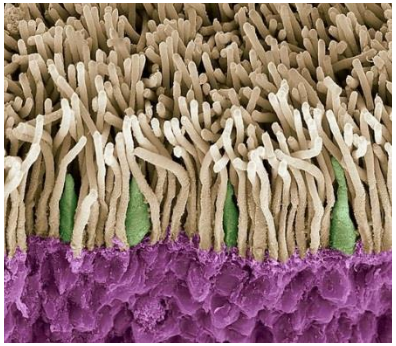
Lichtsteiner et al., A 128x128 120 dB 15μs Latency Asynchronous Temporal Contrast Vision Sensor, IEEE Journal of Solid-State Circuits, 2008

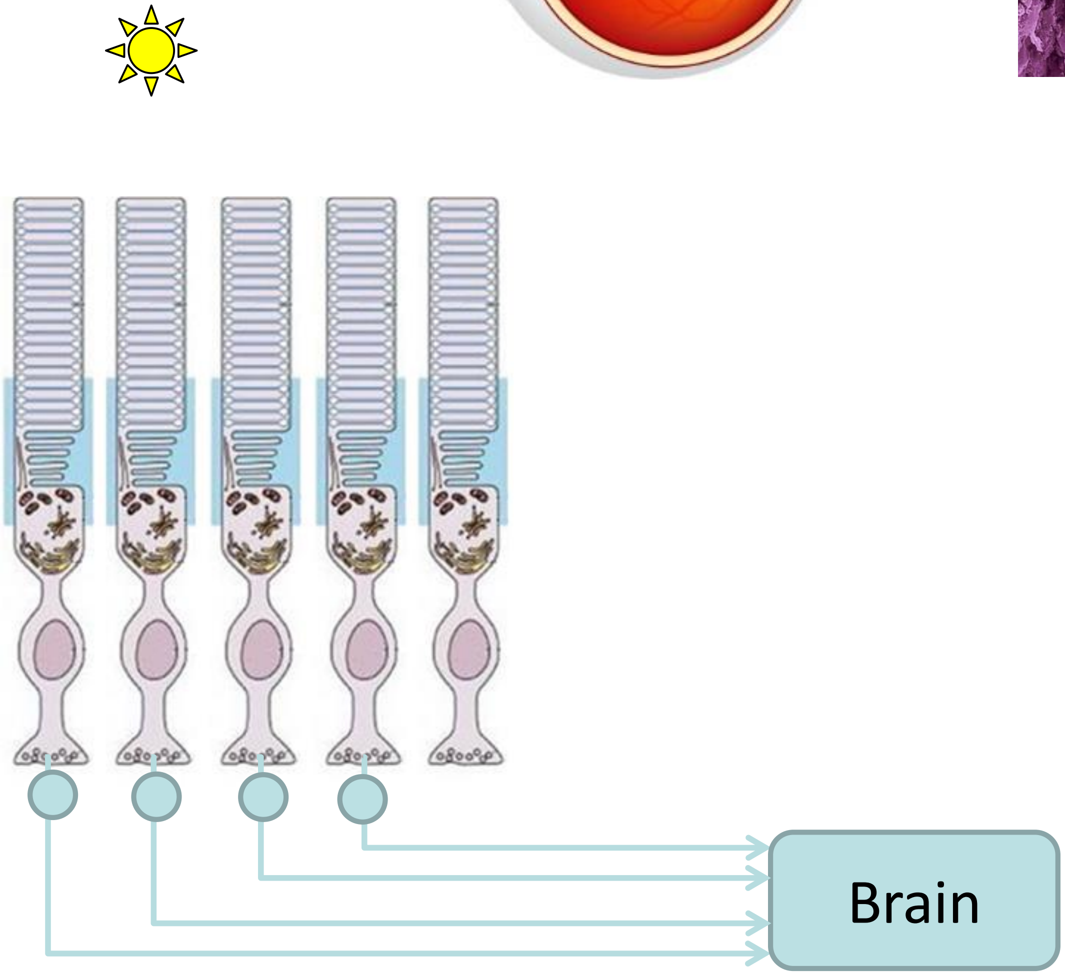
Event cameras are inspired by the Human Eye

Human retina:

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







Event Camera with Static Motion

Without motion, only background noise is output

Standard Camera

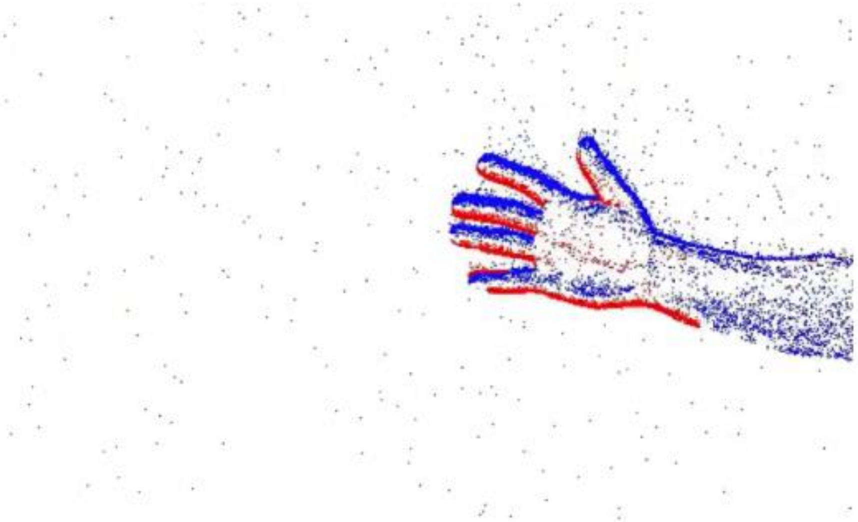
Event Camera (ON, OFF events)

ΔT = 40 ms

Event Camera output with Motion

Standard Camera

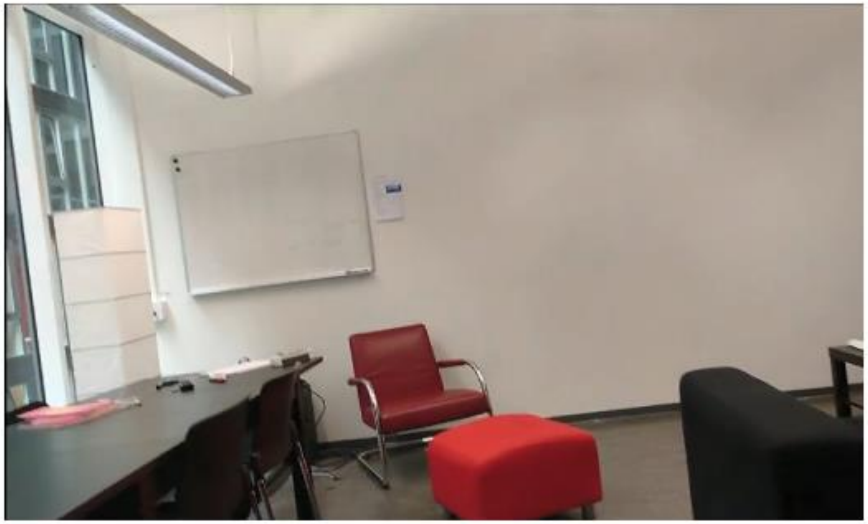
Event Camera (ON, OFF events)



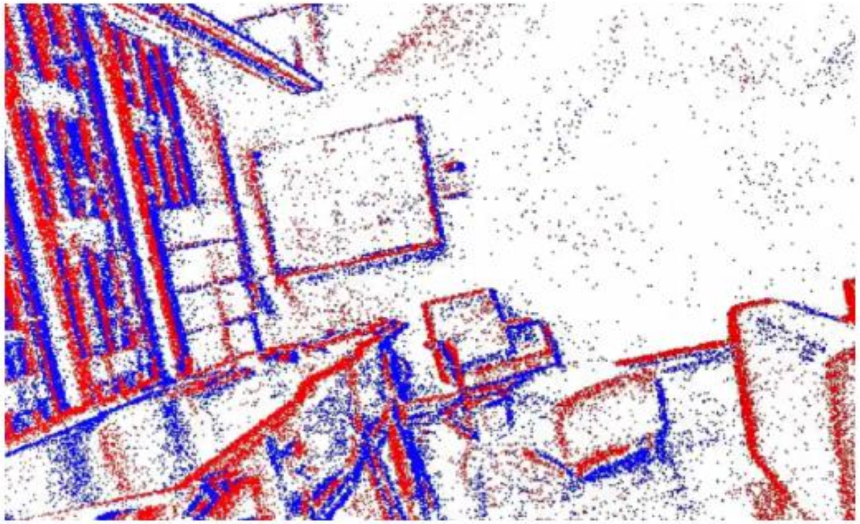
ΔT = 10 ms

Event Camera output with Motion

Standard Camera

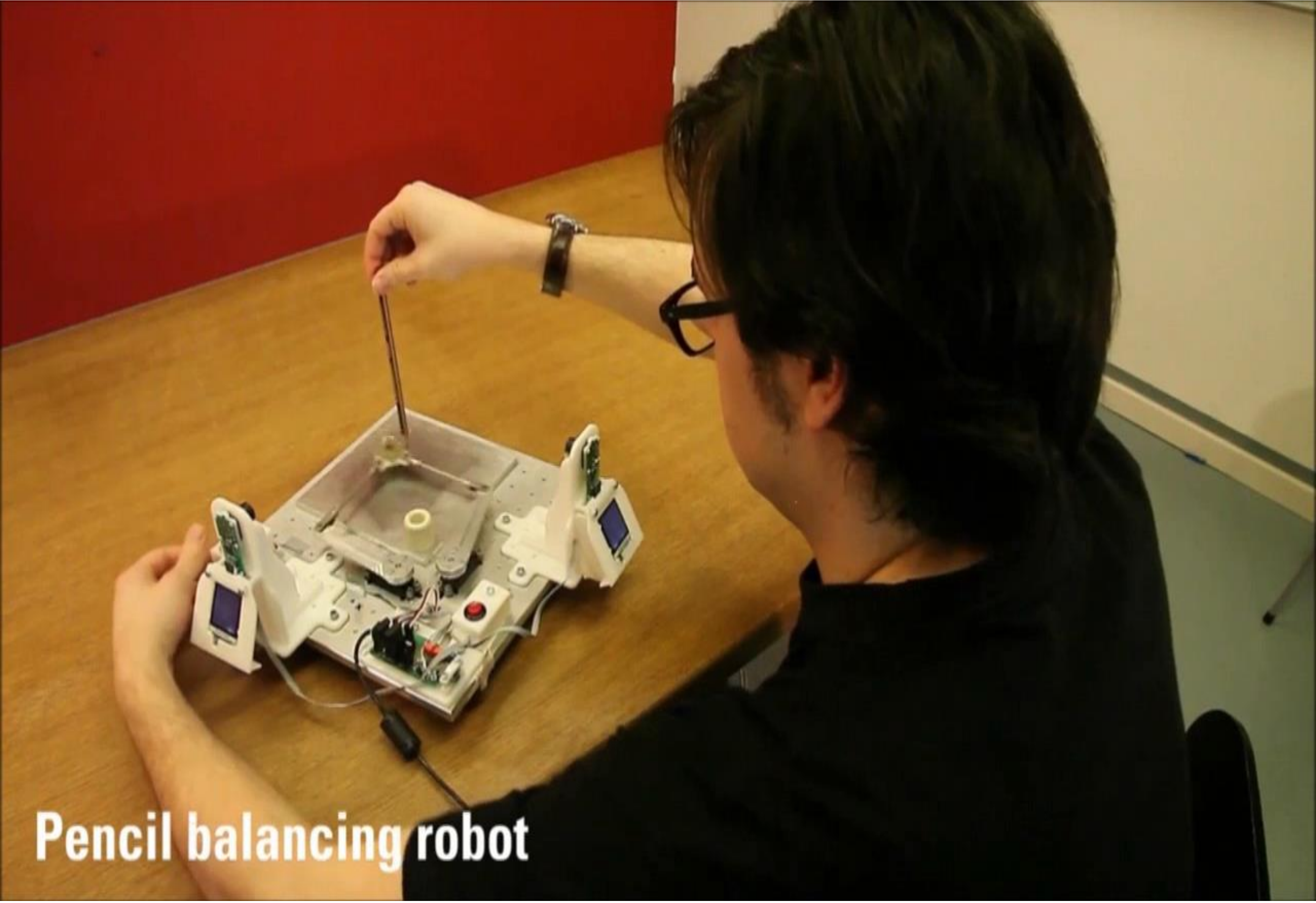


Event Camera (ON, OFF events)



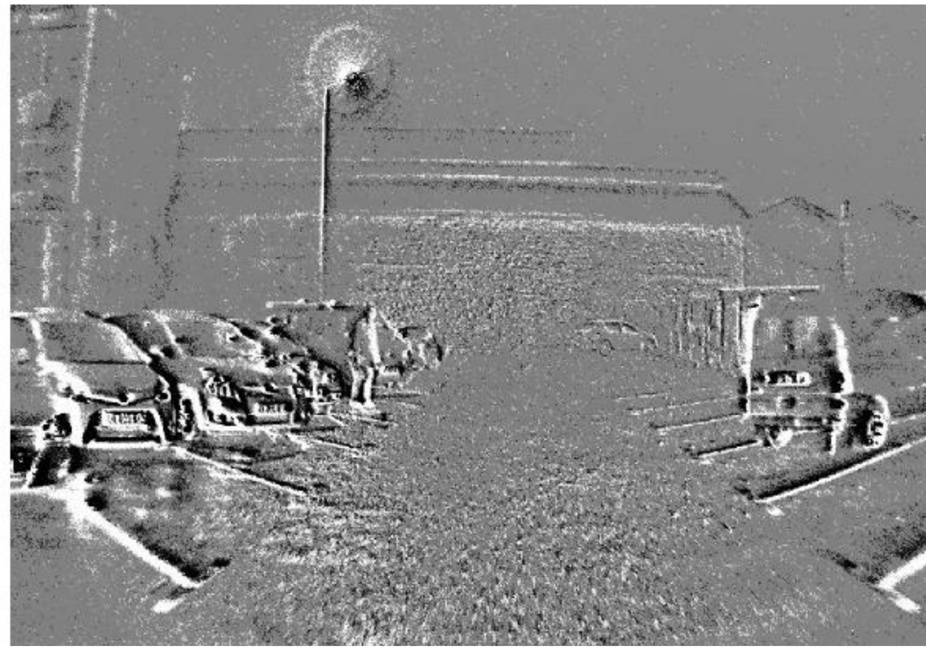
ΔT = 40 ms

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)

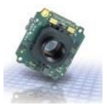
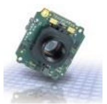


GoPro Hero 6

Event Camera by Prophesee
White = Positive events
Black = Negative events

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**

Comparison of current event cameras

	DVS128 [95]	ATIS [135]	DAVIS240 [16]	DAVIS346 [31]	Gen3 CD [138]	DVS-Gen3 [166]	CeleX-V [23]	DVS-Gen4 [166]
Supplier	iniVation	Prophesee	iniVation	iniVation	Prophesee	Samsung	CelePixel	Samsung
Year	2008	2011	2014	2017	2017	2018	2018	2019
Resolution (pixels)	128×128	304×240	240×180	346×260	640×480	640×480	1280×800	1280×960
Dynamic range (dB)	120	143	120	120	> 120	> 90	> 120	> 90
Die power consumption (mW)	23	50 - 175	5 - 14	10 - 170	36 - 95	65	-	140
Camera Max. Bandwidth (Meps)	1	-	12	12	66	-	100	-
Pixel size (μm^2)	40×40	30×30	18.5×18.5	18.5×18.5	15×15	9×9	9.8×9.8	4.95×4.95
Grayscale output	no	yes	yes	yes	no	no	yes	no
IMU output	no	no	1 kHz	1 kHz	1 kHz	1 kHz	no	no

[95] P. Lichtsteiner, C. Posch, and T. Delbruck. "A 128×128 120 dB 15 μs latency asynchronous temporal contrast vision sensor". In: *IEEE J. Solid-State Circuits,* 2008, http://dx.doi.org/10.1109/JSSC.2007.914337

[135] C. Posch, D. Matolin, and R. Wohlgenannt. "A QVGA 143 dB Dynamic Range Frame-Free PWM Image Sensor With Lossless Pixel-Level Video Compression and Time-Domain CDS". In: *IEEE J. Solid-State Circuits* 46.1, 2011, http://dx.doi.org/10.1109/JSSC.2010.2085952

[16] C. Brandli, R. Berner, M. Yang, S.-C. Liu, and T. Delbruck. "A 240x180 130dB 3us Latency Global Shutter Spatiotemporal Vision Sensor". In: *IEEE J. Solid-State Circuits* 49.10 (2014), pp. 2333–2341. http://dx.doi.org/10.1109/JSSC.2014.2342715

[31] https://inivation.com/wp-content/uploads/2019/07/2019-07-09-DVS-Specifications.pdf

[138] https://www.prophesee.ai/event-based-evk/

[166] http://rpg.ifi.uzh.ch/docs/CVPR19workshop/CVPRW19_Eric_Ryu_Samsung.pdf

[23] http://rpg.ifi.uzh.ch/docs/CVPR19workshop/CVPRW19_CelePixel.pdf

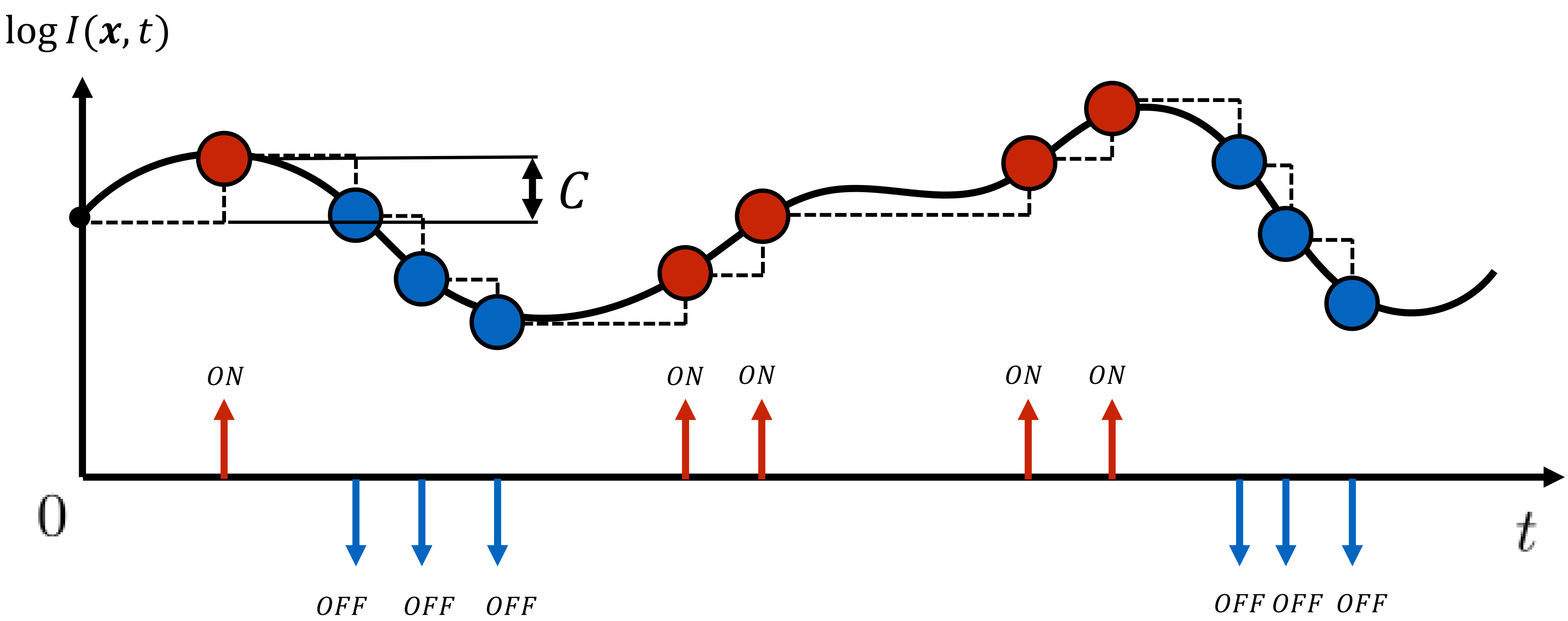
How do we unlock the outstanding potential of event cameras:

- Low latency
- High dynamic range
- No motion blur

Recall the Generative Event Model

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

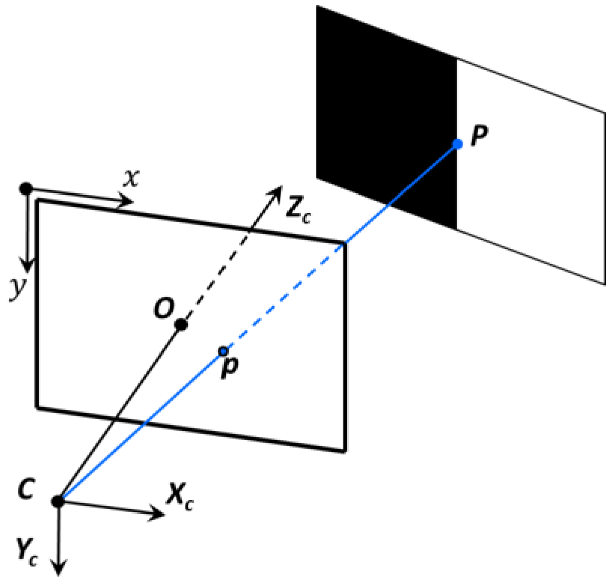
$$\log I(\boldsymbol{x}, t) - \log I(\boldsymbol{x}, t - \Delta t) = \pm C$$

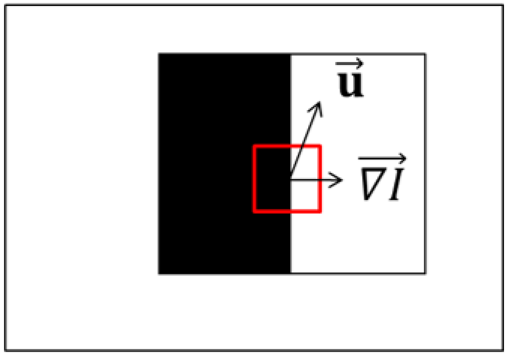


1st Order Approximation

- 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:

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

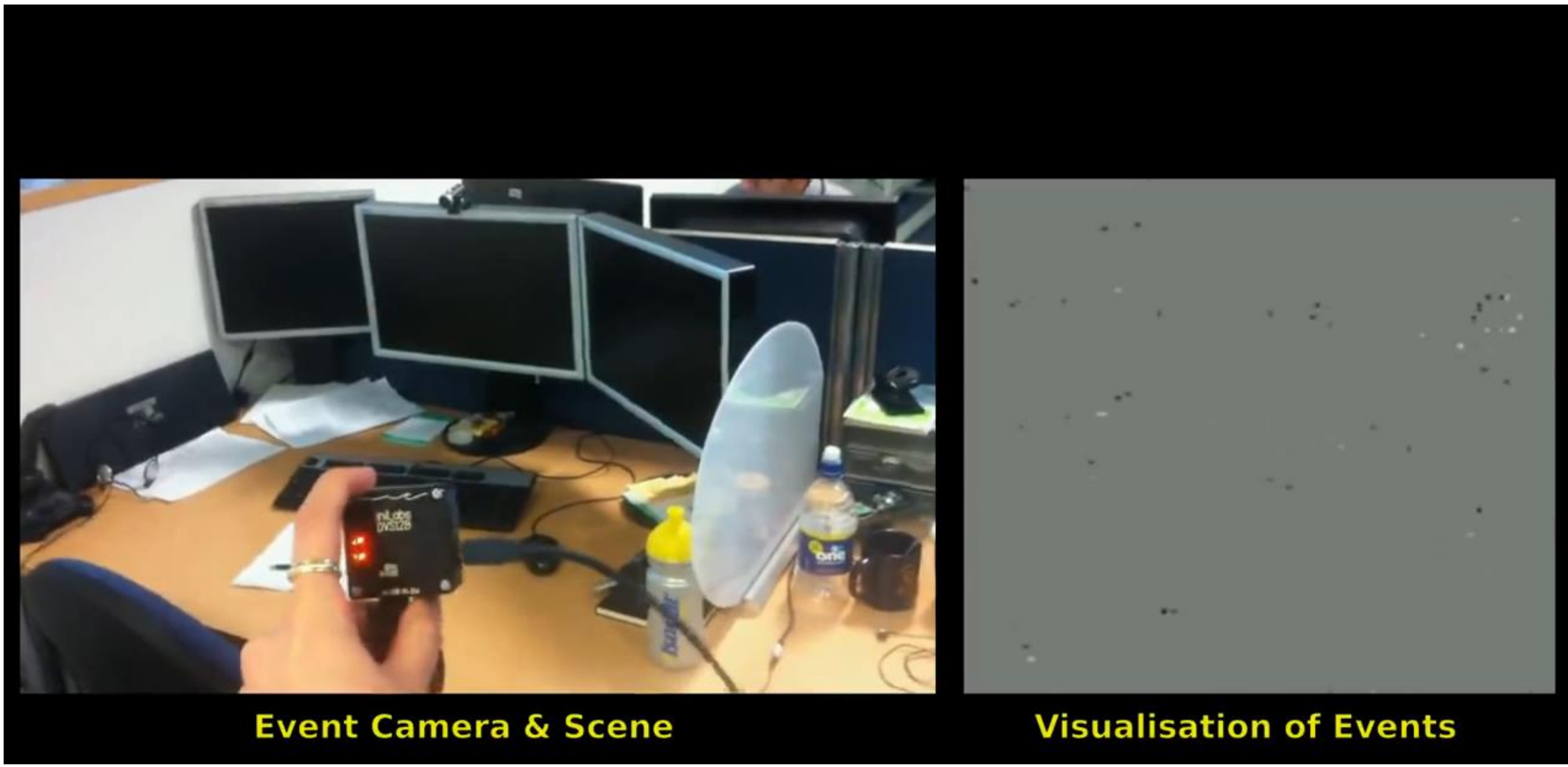
$$\Rightarrow \quad L(x, y, t + \Delta t) - L(x, y, t) = -\frac{\partial L}{\partial x} u - \frac{\partial L}{\partial y} v$$

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

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.

Example 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



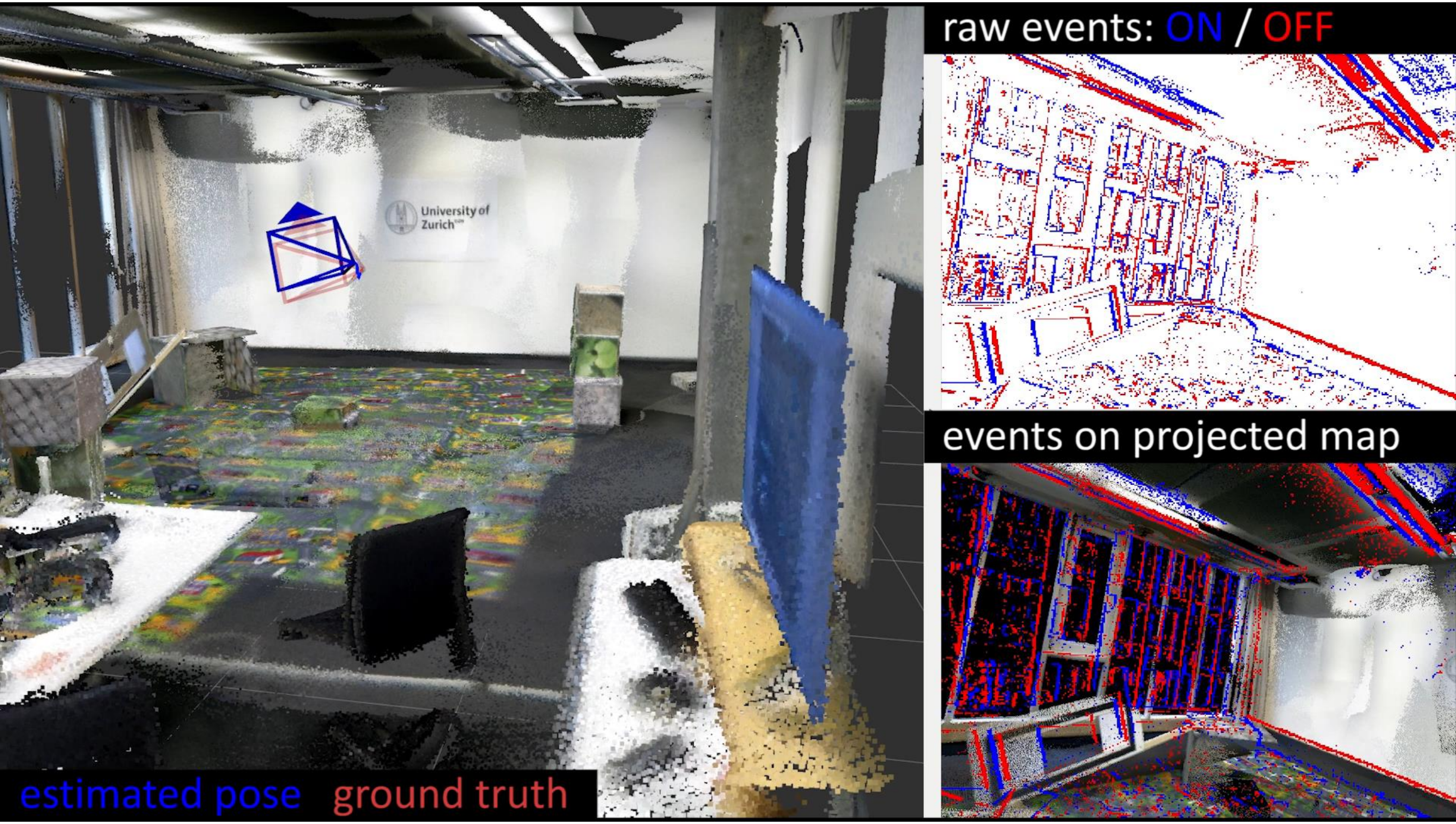
Kim et al., Simultaneous Mosaicing and Tracking with an Event Camera, BMVC'14

Example 2: 6DoF Tracking from Photometric Map

- Probabilistic, simultaneous motion & contrast 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

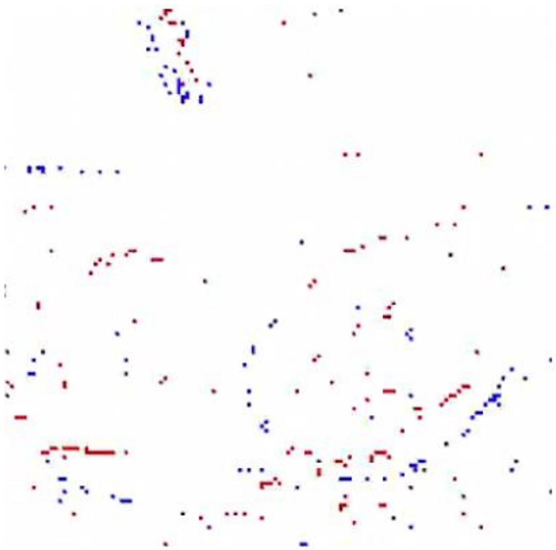
Gallego et al., Event-based 6-DOF Camera Tracking from Photometric Depth Maps, **T-PAMI**, 2018. [PDF](PDF) Video

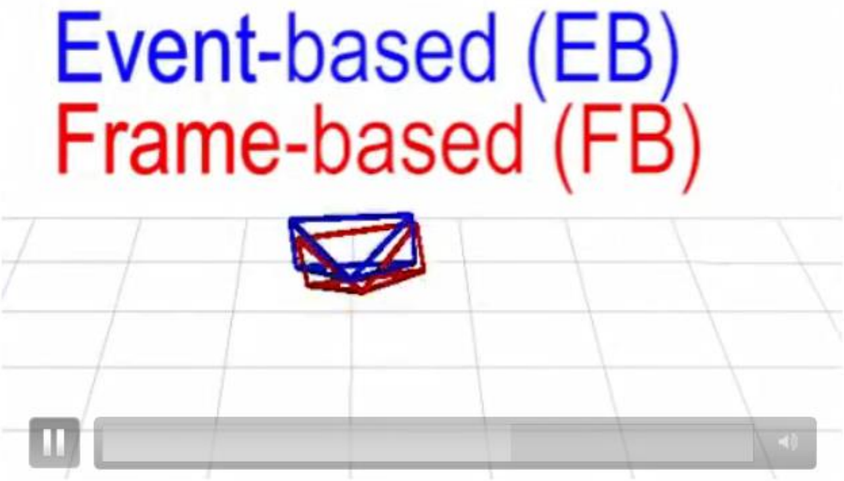
Example 2: 6DoF Tracking from Photometric Map



Bryner et al., Event-based, Direct Camera Tracking from a Photometric Depth Map, ICRA'19. [PDF](PDF) Video

Event camera



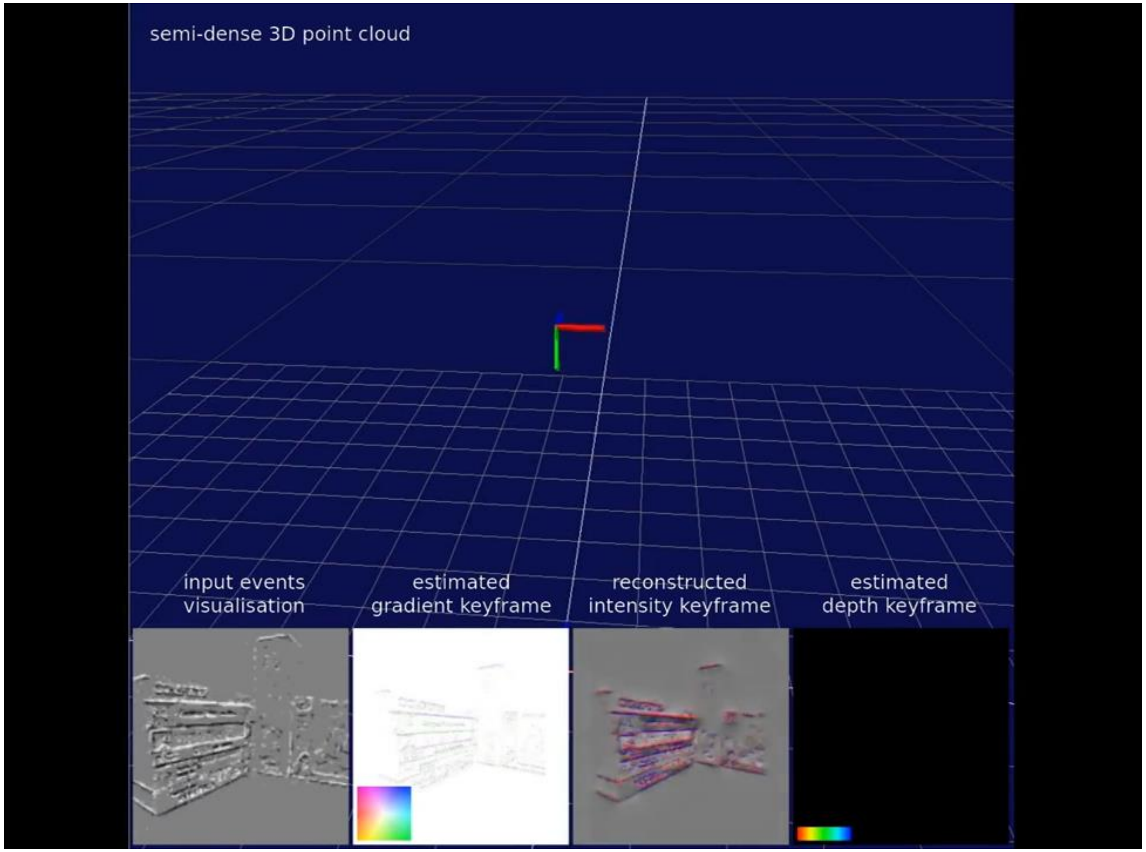


Standard camera

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

Example 3: Parallel Tracking & Mapping (SLAM)

- Tracking: EKF in 6 DOF pose
  - Uses random walk model & inverse depth
  - Use 1st order approximation of generative event model to update pose
- Runs in real time on a GPU

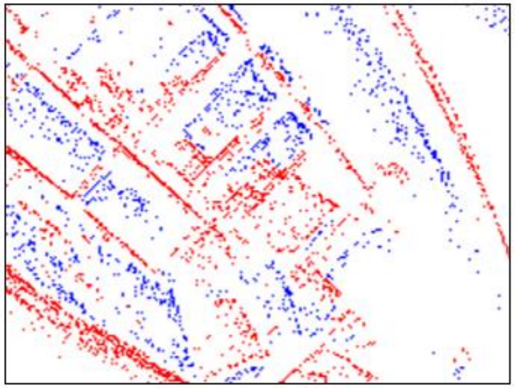


Kim et al., Real-Time 3D Reconstruction and 6-DoF Tracking with an Event Camera, ECCV'16

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

Why combining them?

< 10 years research

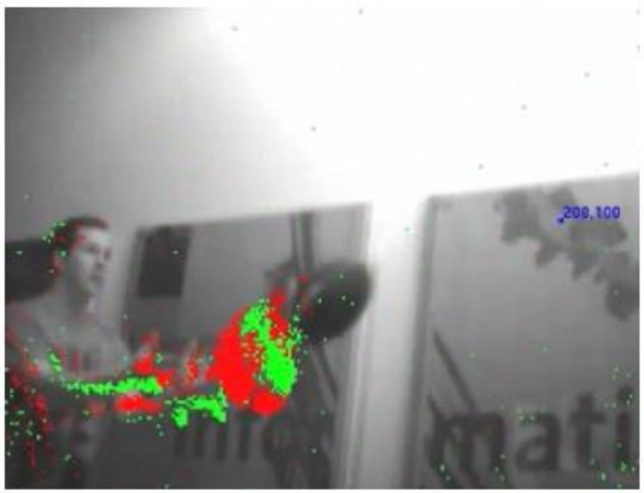


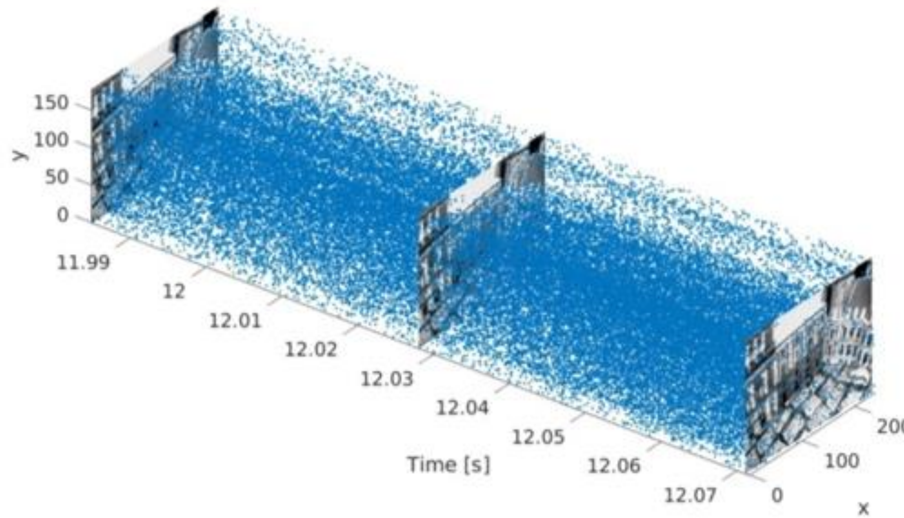
> 60 years of research!

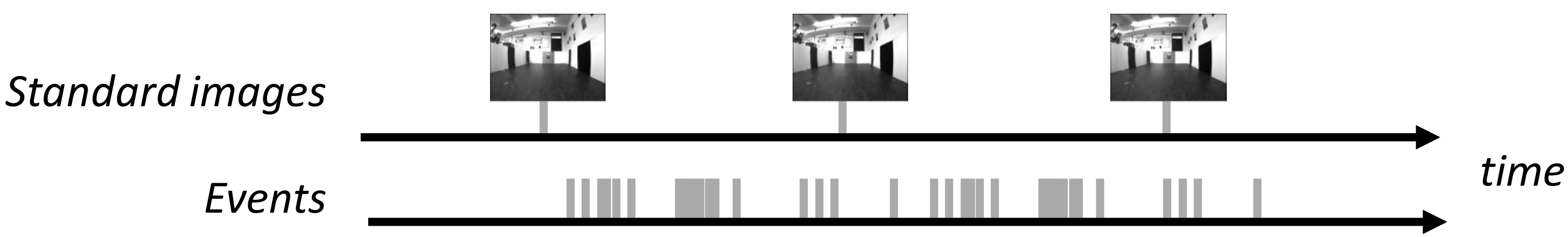
	Event Camera	**Standard Camera**
Update rate	High (asynchronous): 1 MHz	Low (synchronous)
Dynamic Range	High (140 dB)	Low (60 dB)
Motion Blur	No	Yes
Static motion	No (event camera is a high pass filter)	Yes
Absolute intensity	No (reconstructable up to a constant)	Yes
Maturity	< 10 years of research	>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



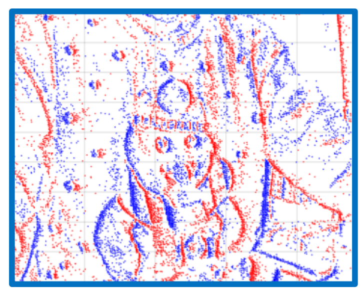




Brandli et al. *A 240x180 130dB 3us latency global shutter spatiotemporal vision sensor*. IEEE JSSC, 2014

Example 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

Example 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

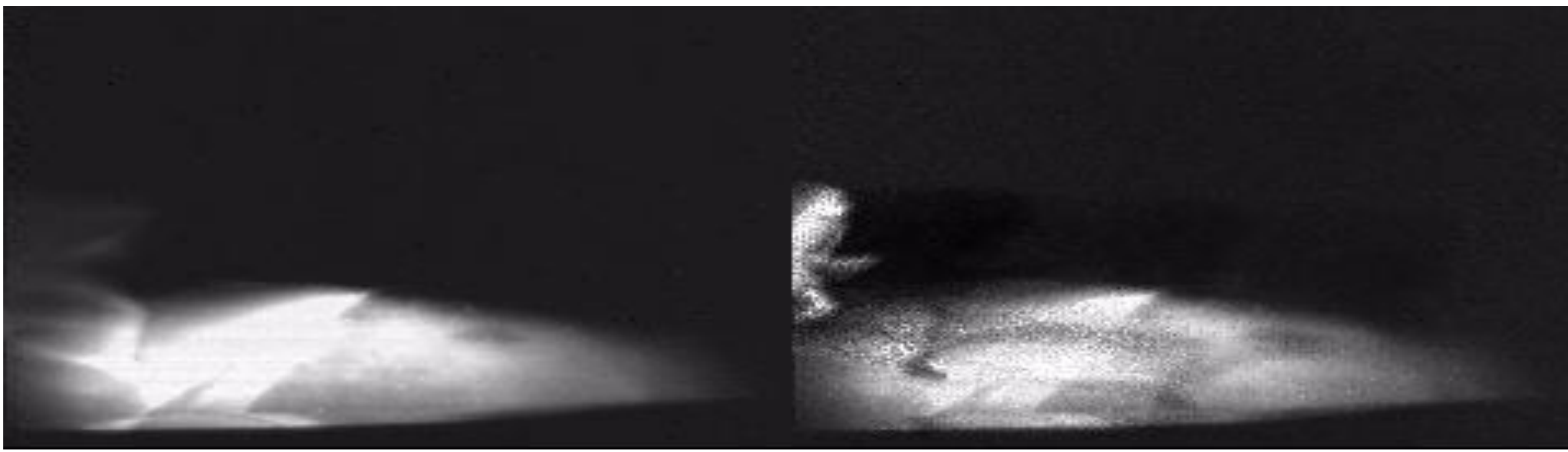
Input blur image

Output sharp video

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

Example 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

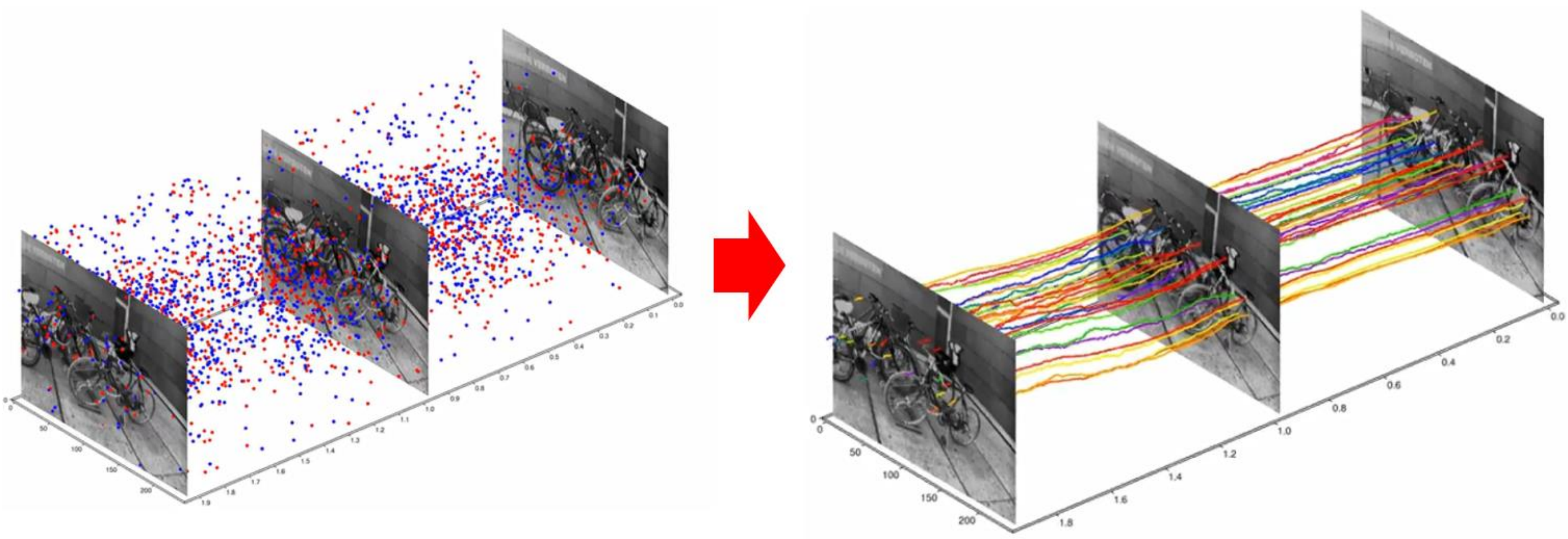
What about an asynchronous Luca-Kanade-Tomasi (KLT) Tracker for Event Cameras?

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

Asynchronous, Photometric Feature Tracking using Events and Frames

- **Goal**:Extract features on **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 far enable **asynchronous**, **low-latency** (~10µs) algorithmic update on an **event-by-event fashion**
- However:
  - Event-by-event update requires **GPU for real-time processing**
  - Additionally, they make use of the **generative event model**

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

  or its 1st order approximation

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

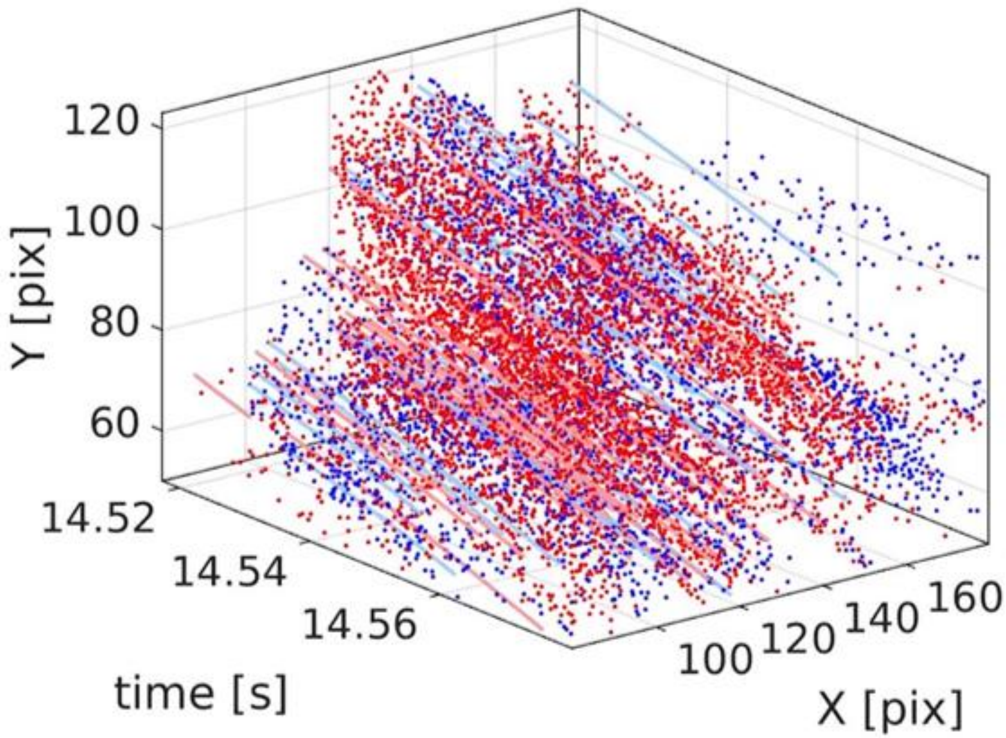
  which **requires knowledge of the contrast sensitivity** C (which is **scene dependent** and might **differ from pixel to pixel**)

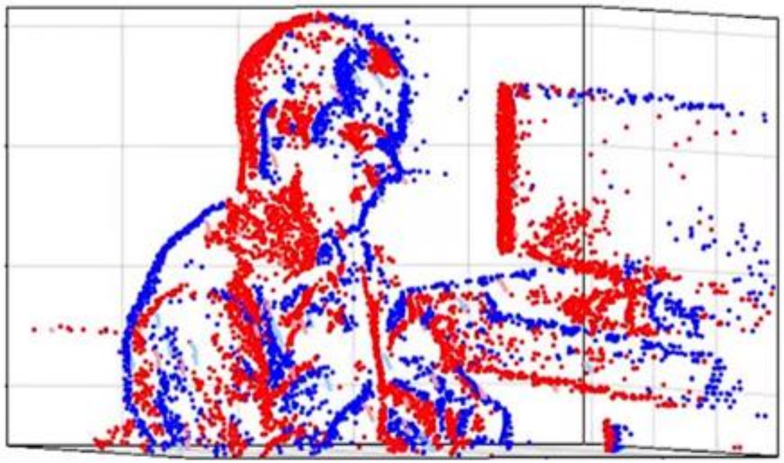
Focus Maximization for:

- 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.

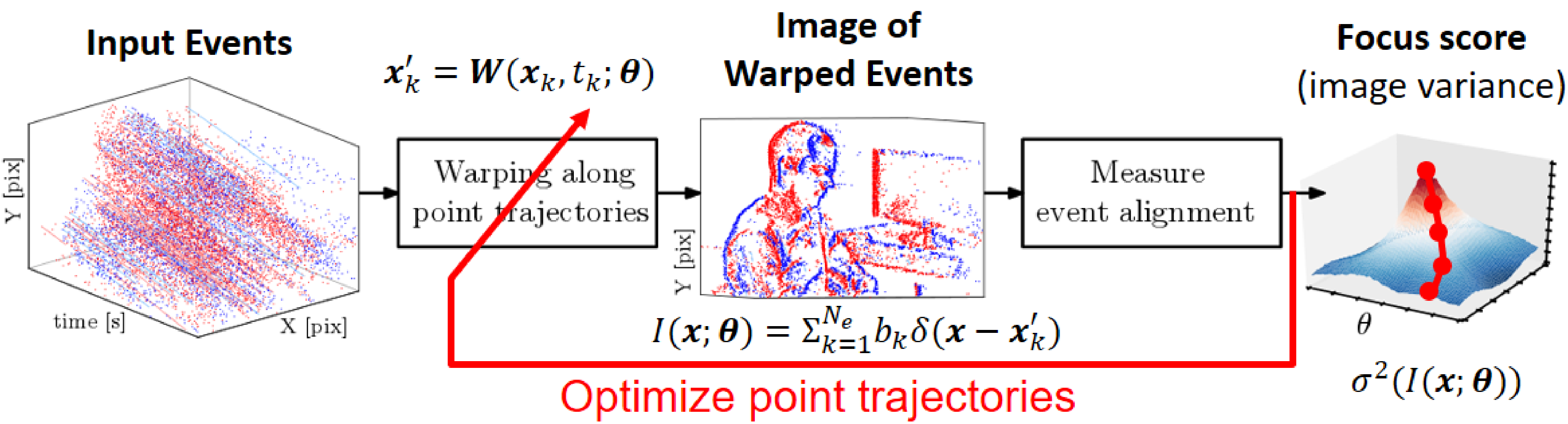
Idea: Warp spatio-temporal volume of events to **maximize focus** (e.g., sharpness) of the resulting image





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

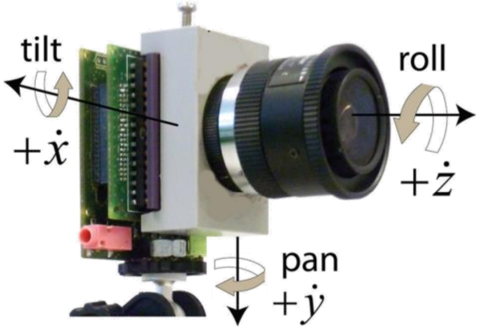


Can be implemented in a sliding-window fashion to enable per low-latency, per-event update rate

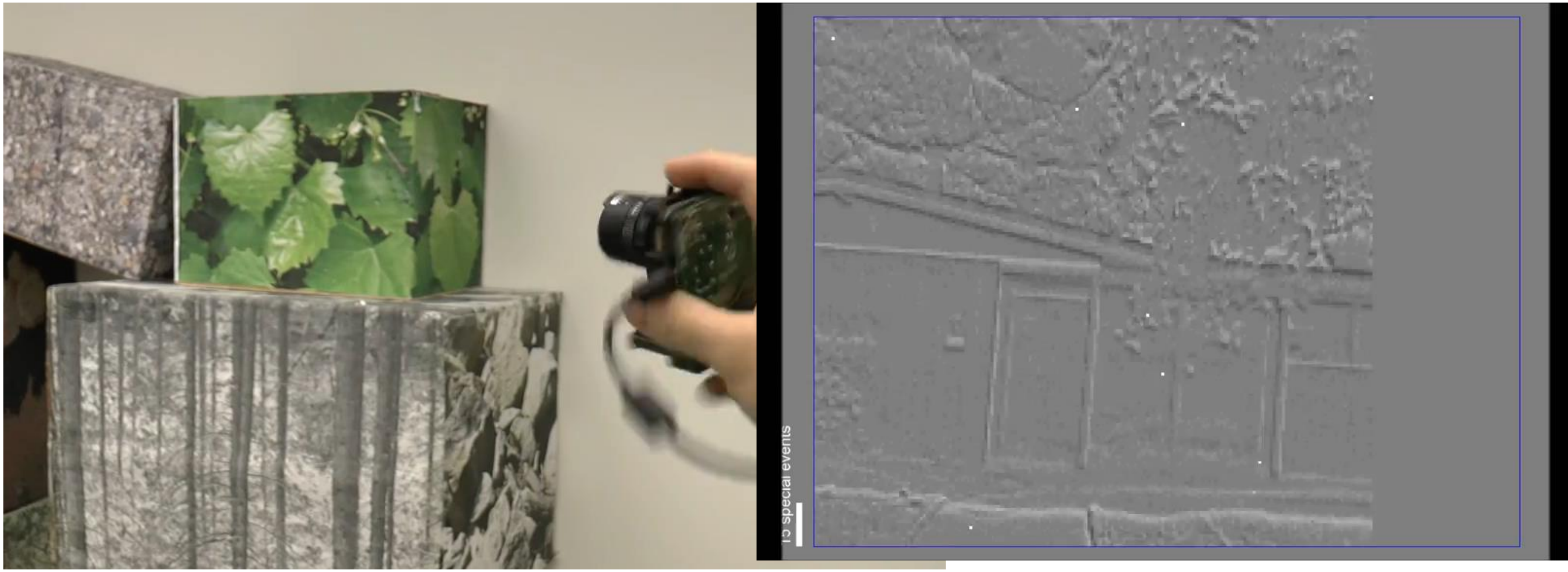
Runs in real time on a CPU

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.

Application 1: Image Stabilization

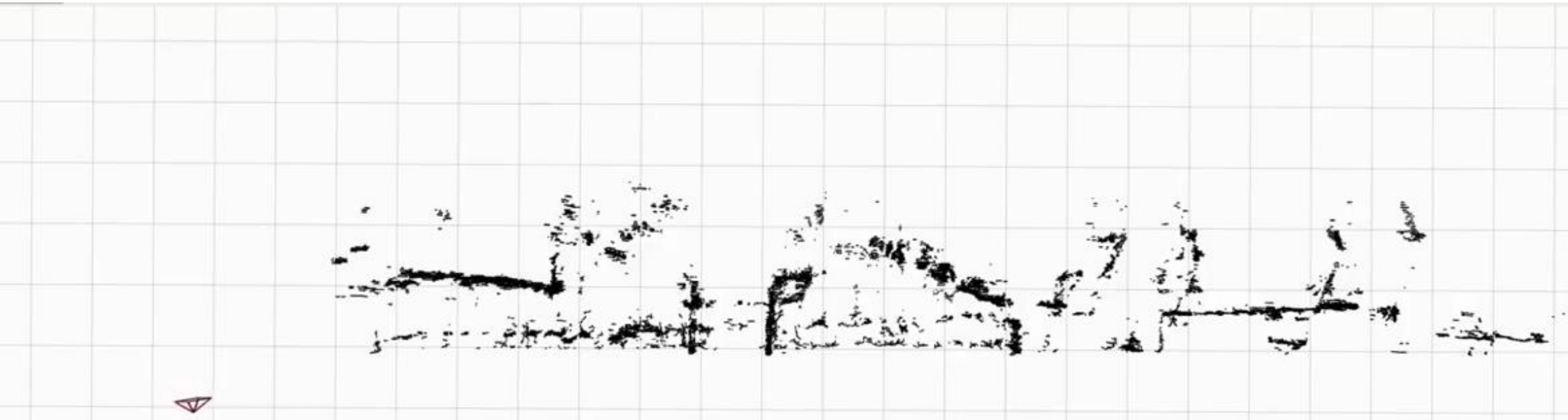


- Problem: Estimate rotational motion (3DoF) of an event camera
- Can process millions of events per second in real time on a smartphone CPU (OdroidXU4)
- Works up to over ~1,000 deg/s



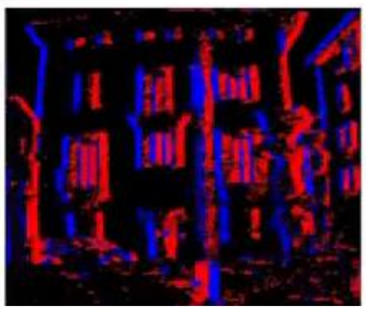
Gallego et al., Accurate Angular Velocity Estimation with an Event Camera, IEEE RAL'16. PDF. Video.

Application 3: 3D Reconstruction from a Train at 200km/h

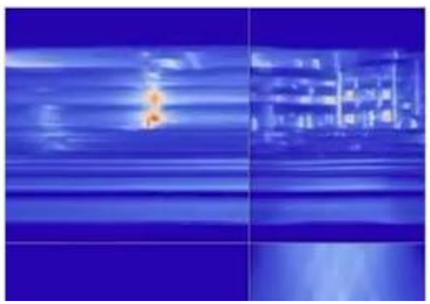


Video: https://www.youtube.com/watch?v=fA4MiSzYHWA

DVS Events



DSI



External camera

In collaboration with **SIEMENS** and the Swiss Railway company, SBB

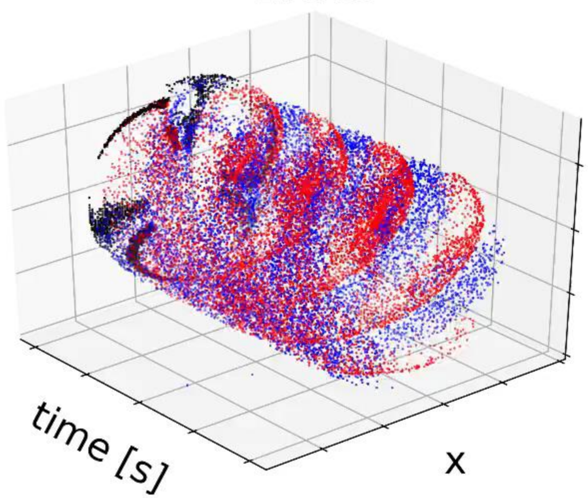
SBB CFF FFS

Rebecq et al., EVO: A Geometric Approach to Event-based 6-DOF Parallel Tracking and Mapping, RAL'17. PDF. Video
Rebecq et al., EMVS: Event-based Multi-View Stereo, IJCV, 2018. PDF Video Source Code

Application 4: Motion Segmentation

Conventional Frames

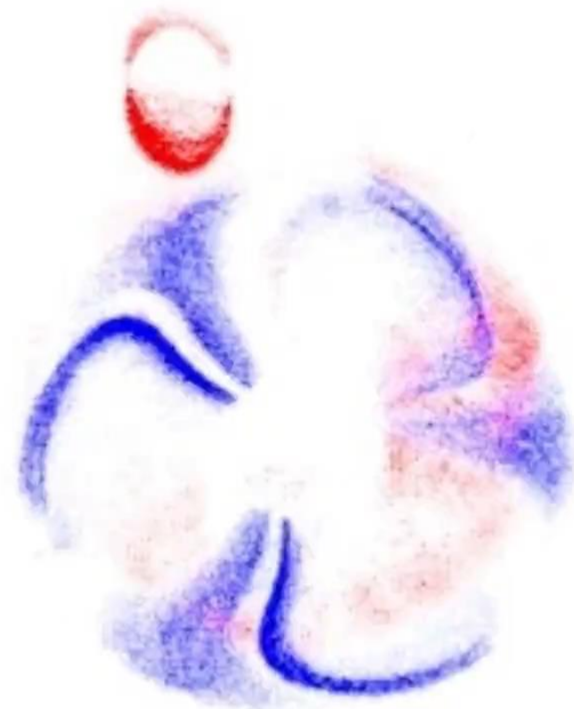
Events



Sequence: Fan and Coin

One motion model is used per cluster; one for the fan, modelling rotation, one for the coin, modelling optic flow

Motion-Compensated Segmented Events

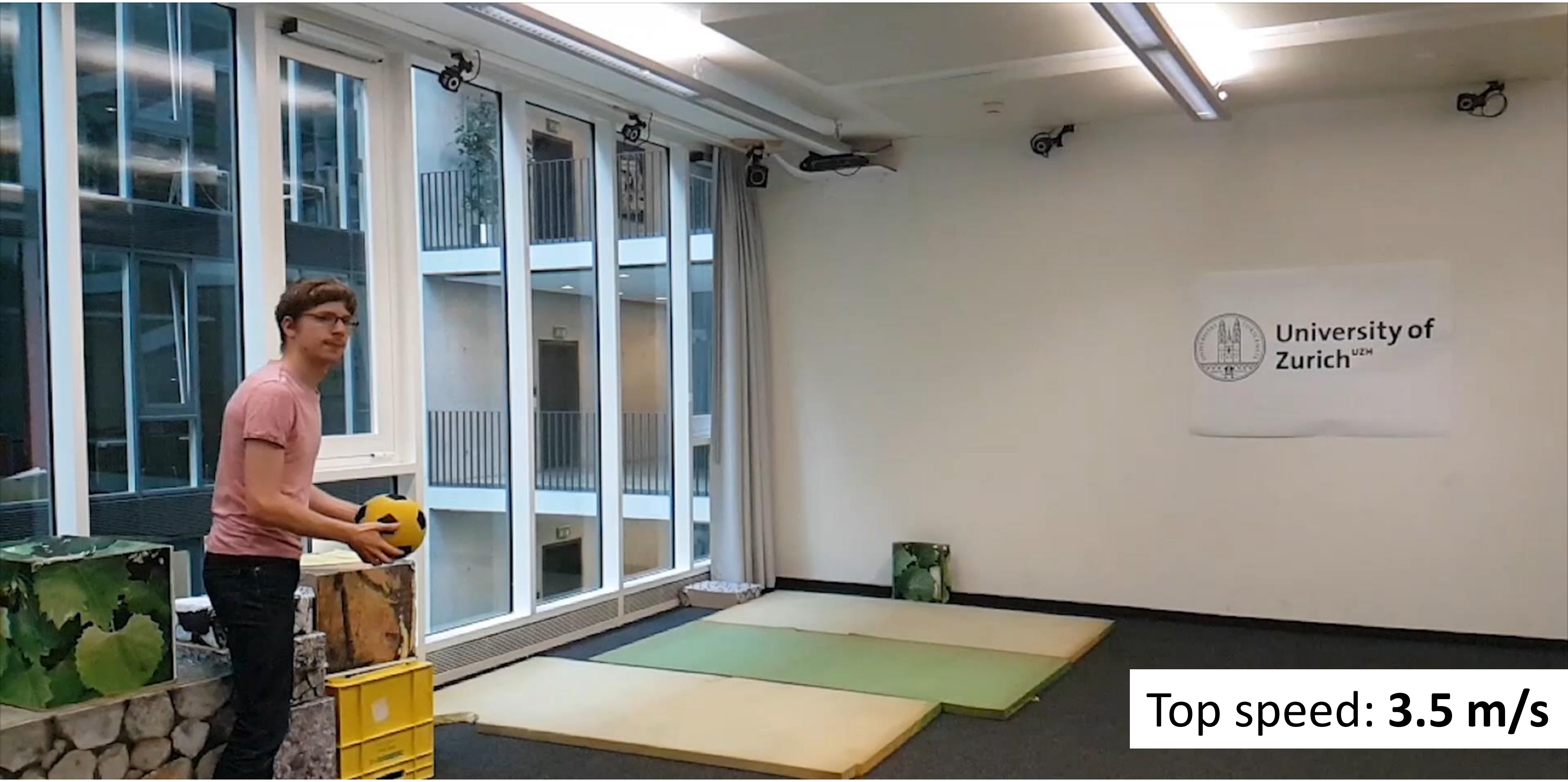


Stoffregen et al., Motion Segmentation by Motion Compensation, arXiv 2019. [PDF](). [Video]().

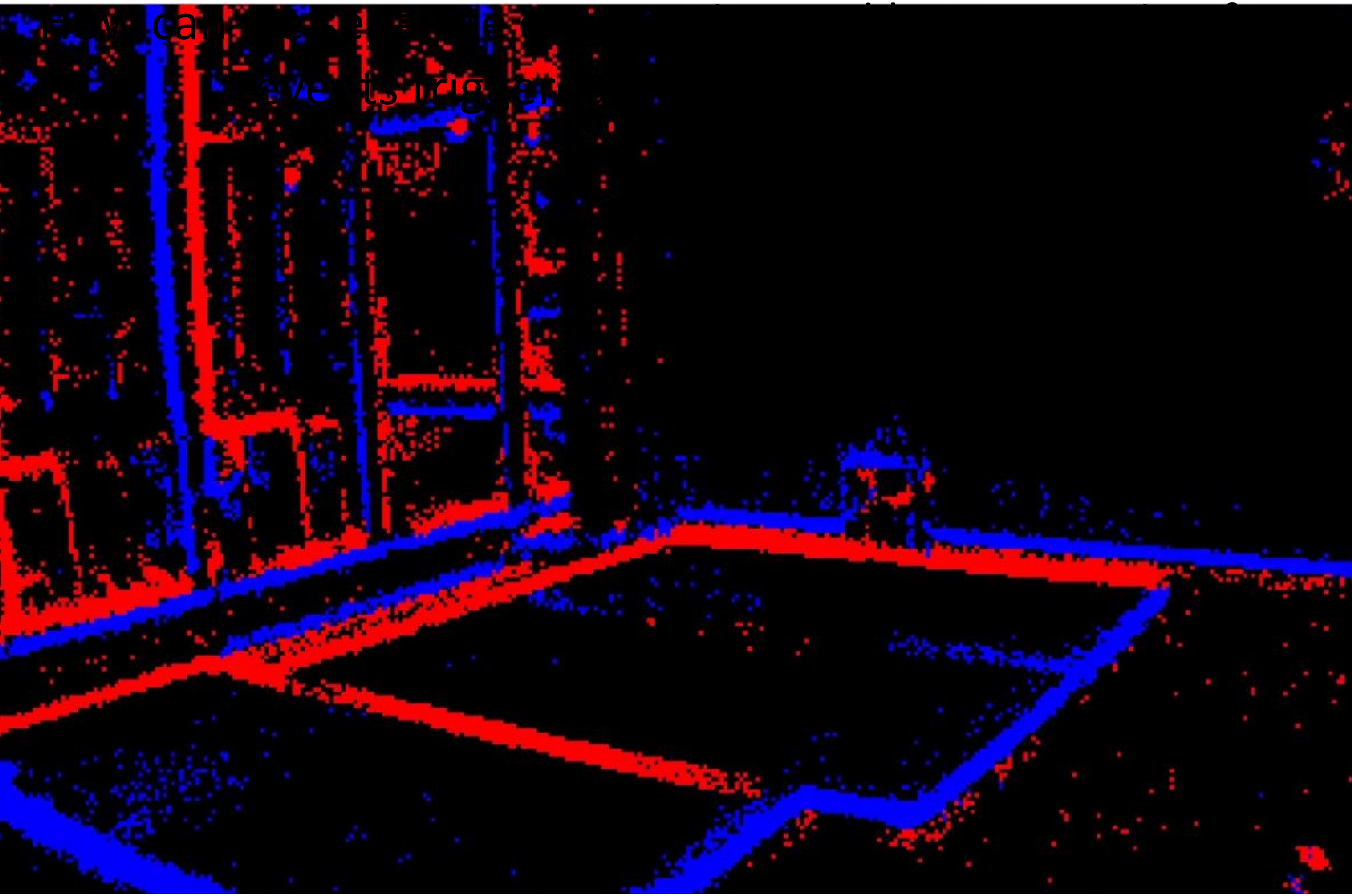
Application 5:
Drone Dodging Dynamic Obstacles

Event-based Dynamic Obstacle Detection & Avoidance

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

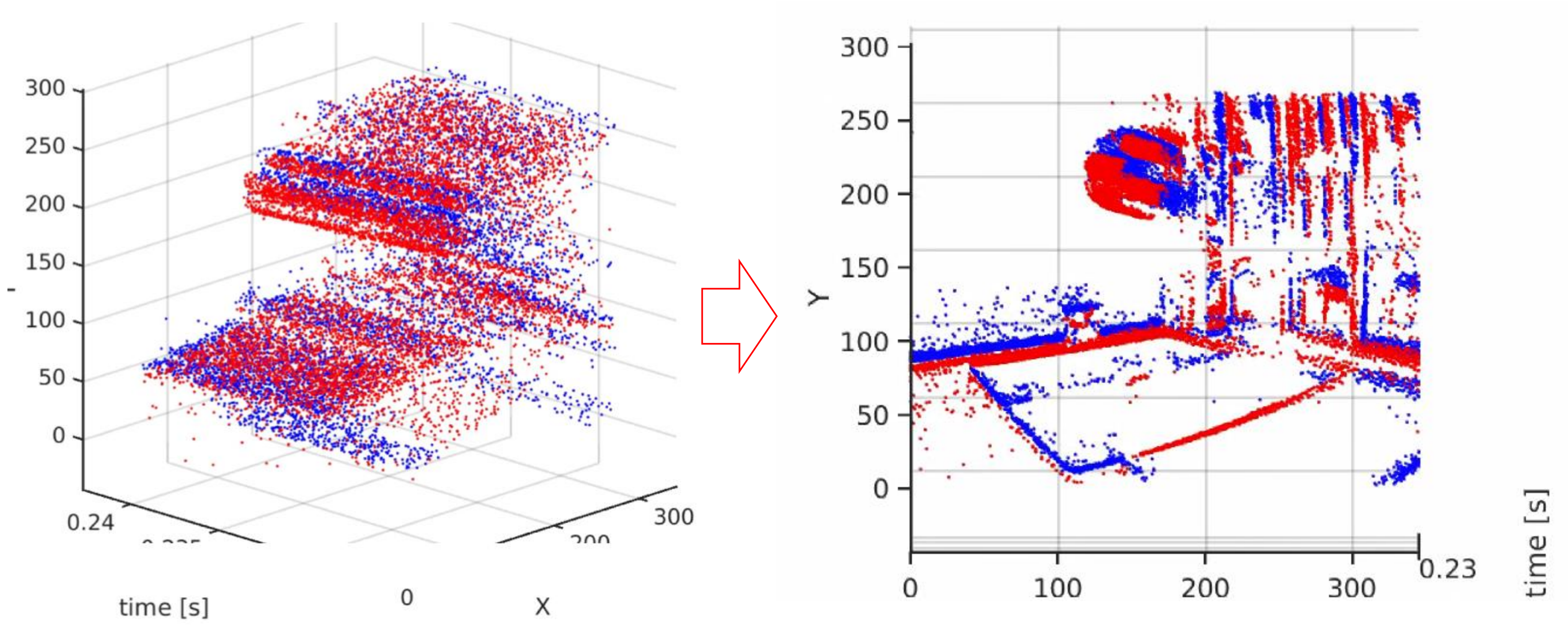


Falanga et al. *How Fast is too fast? The role of perception latency in high speed sense and avoid*, RAL'19.
PDF. Video. Featured in IEEE Spectrum.



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]().

Idea: Warp spatio-temporal volume of events to maximize contrast of the resulting image: Static objects will appear sharp, while moving ones will appeared blurred.



Falanga et al. How Fast is too fast?, IEEE RAL'19. [PDF](). [Video](). Featured in [IEEE Spectrum]().
Gallego et al., A Unifying Contrast Maximization Framework for Event Cameras, CVPR'18, [PDF](), [Video]()

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.

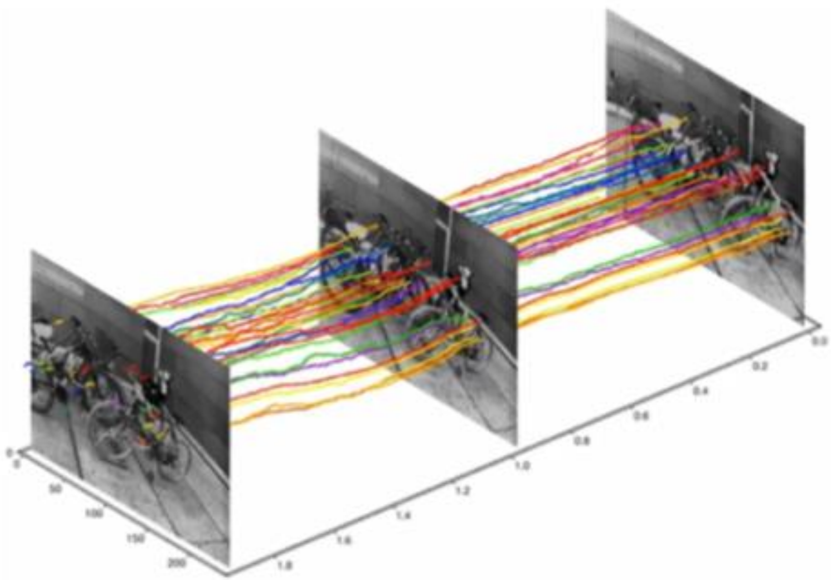
UltimateSLAM:

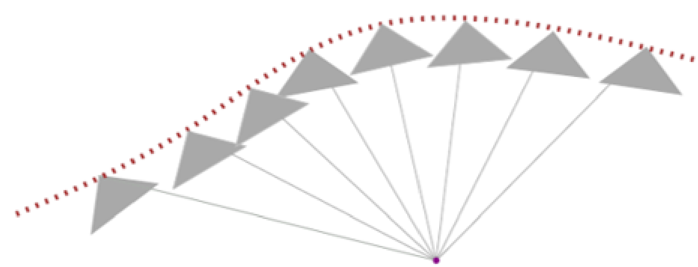
combining **events**, **images**, and **IMU** for robust visual SLAM in HDR and High Speed Scenarios

UltimateSLAM: combining Events + Frames + IMU

Front End:
Feature tracking from
Events and Frames

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

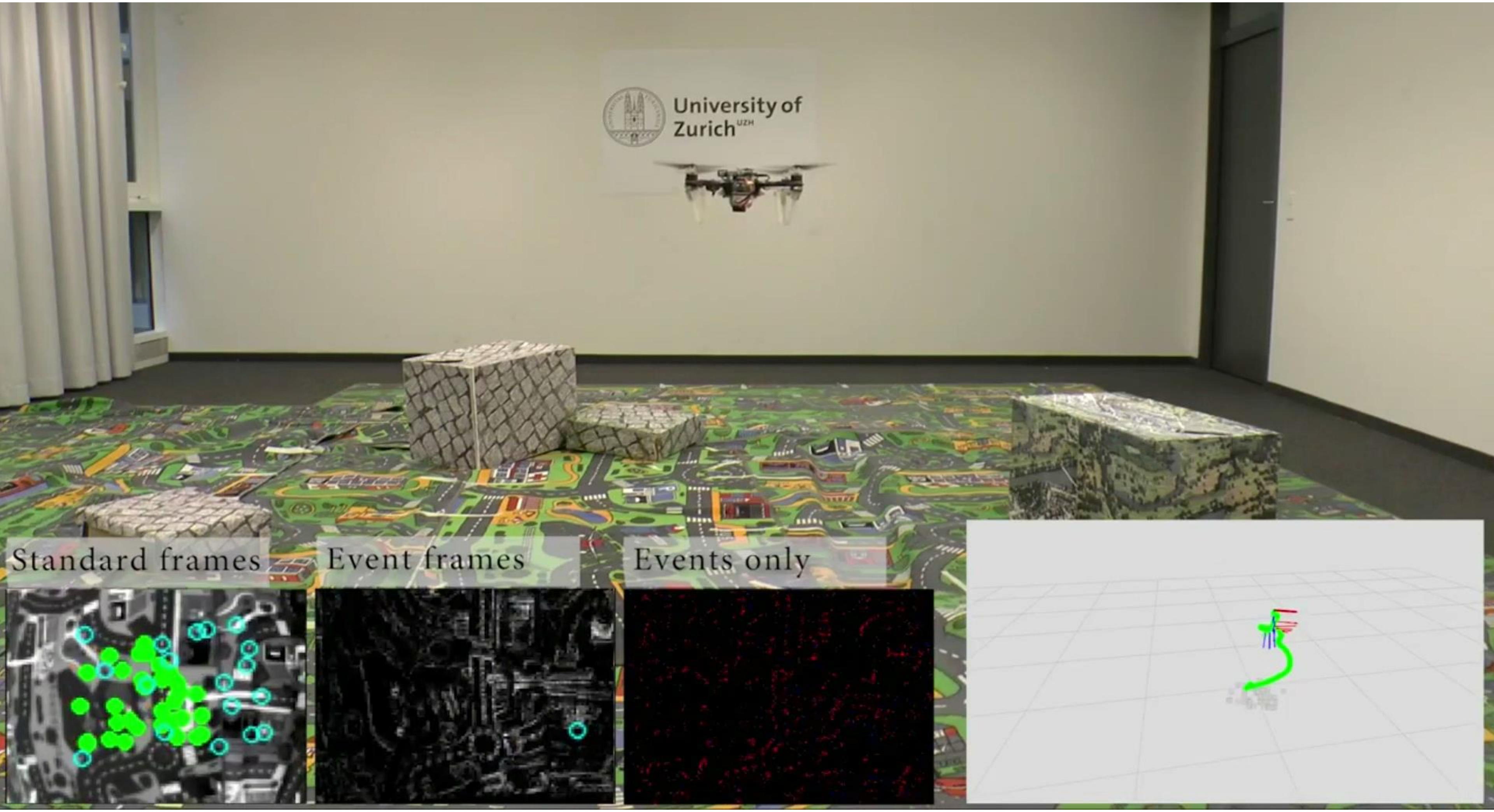




Rosinol et al., Ultimate SLAM? **RAL'18 – Best RAL'18 Paper Award Honorable Mention** [PDF]. [Video]. [IEEE Spectrum].
Mueggler et al., Continuous-Time Visual-Inertial Odometry for Event Cameras, **TRO'18**. [PDF]

Application: Autonomous Drone Navigation in Low Light

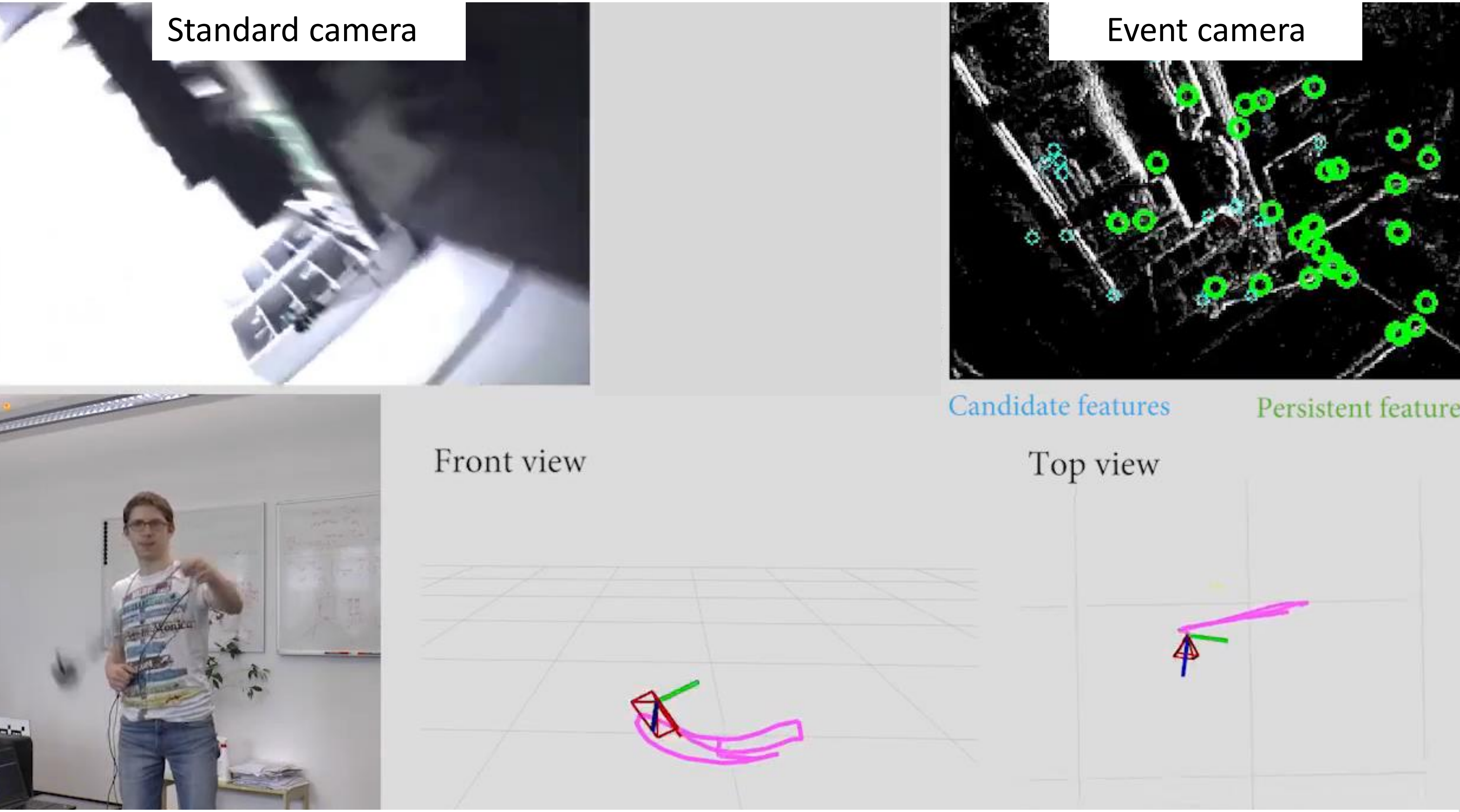
UltimateSLAM running on board (CPU: Odroid XU4)



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



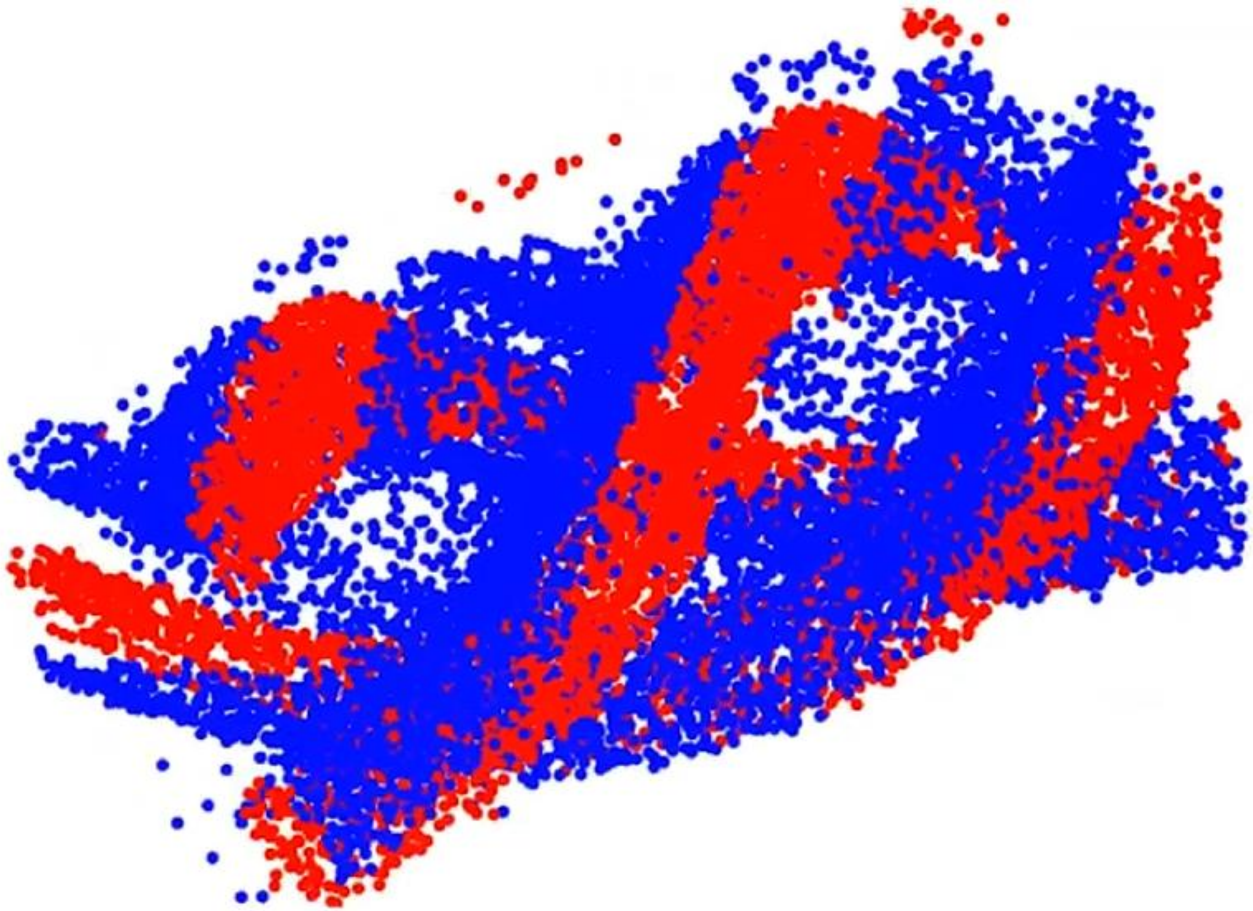
Rosinol et al., Ultimate SLAM? **RAL'18 – Best RAL'18 Paper Award Honorable Mention** PDF. Video. IEEE Spectrum.
Mueggler et al., Continuous-Time Visual-Inertial Odometry for Event Cameras, **TRO'18**. PDF

Learning with Event Cameras

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

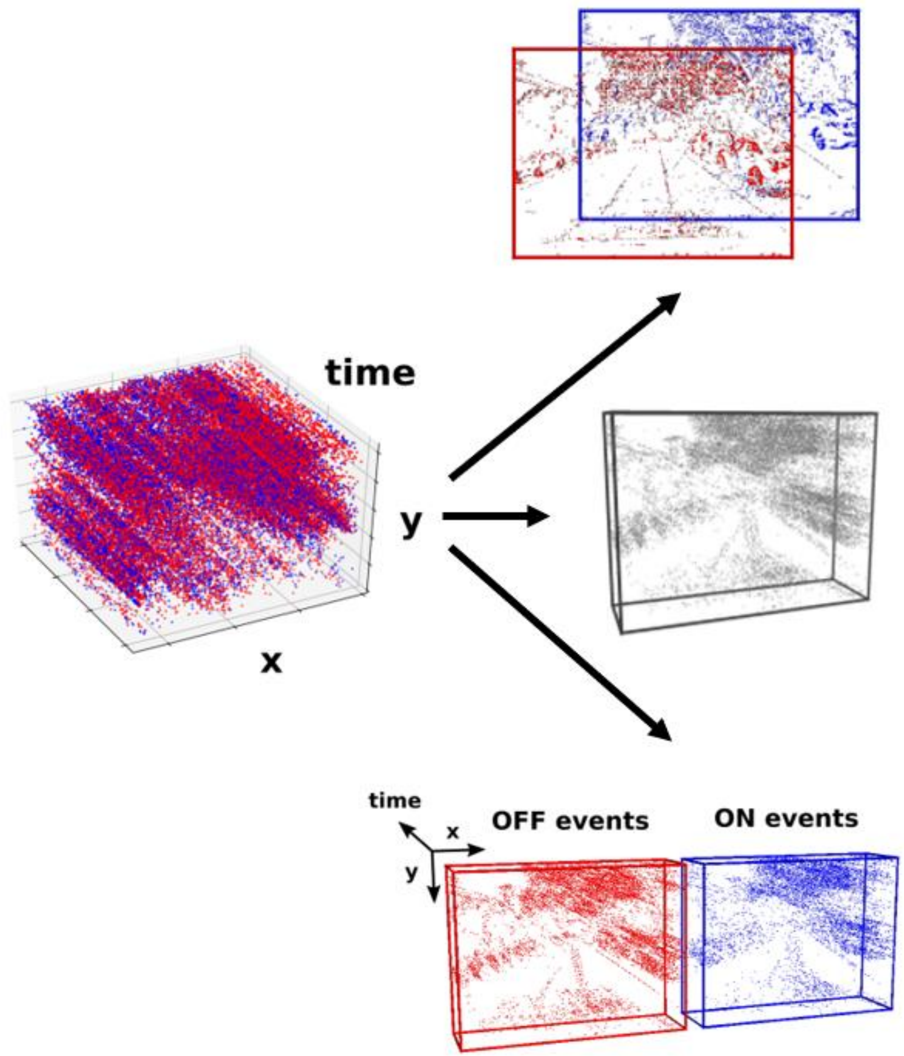
Input representation

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



Video from Zhu et al. (link)

Input representation



[Maqueda CVPR'18], [Zhu'RSS'18]

- **Aggregate positive and negative events into separate channels**
- Discards temporal information

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

- Represent events in space-time into a 3D voxel grid (x,y,t)
- Each voxel contains sum of ON and OFF events falling within the voxel
- Preserves temporal information but discards polarity information

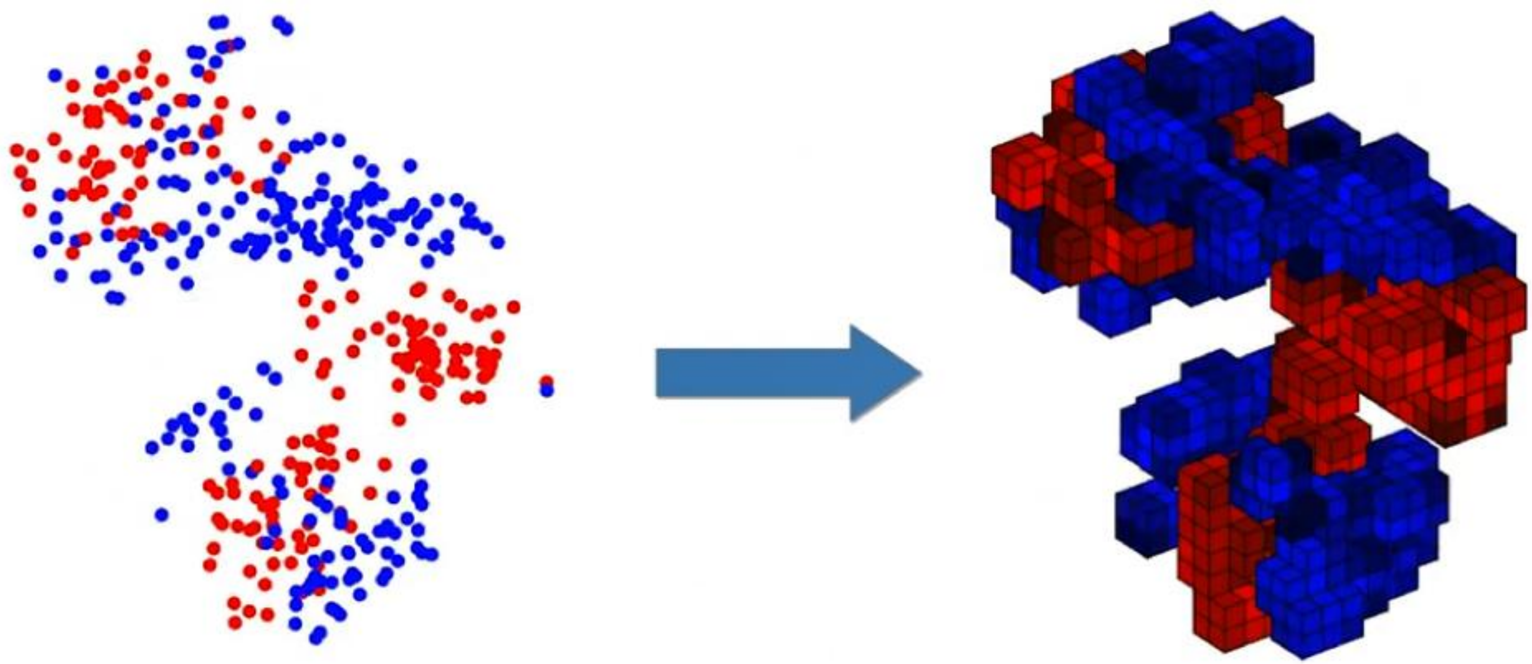
[Gehrig Arxiv'19]

- Represent events in space-time as a 4D Event Spike Tensor (x,y,t,p)
- Polarity information is preserved

Gehrig et al., End-to-End Learning of Representations for Asynchronous Event-Based Data arXiv, 2019. PDF YouTube Project Page

Input representation

Discretized 3D volume (x,y,t): events are inserted into the volume with trilinear interpolation, resulting in minimal loss in resolution

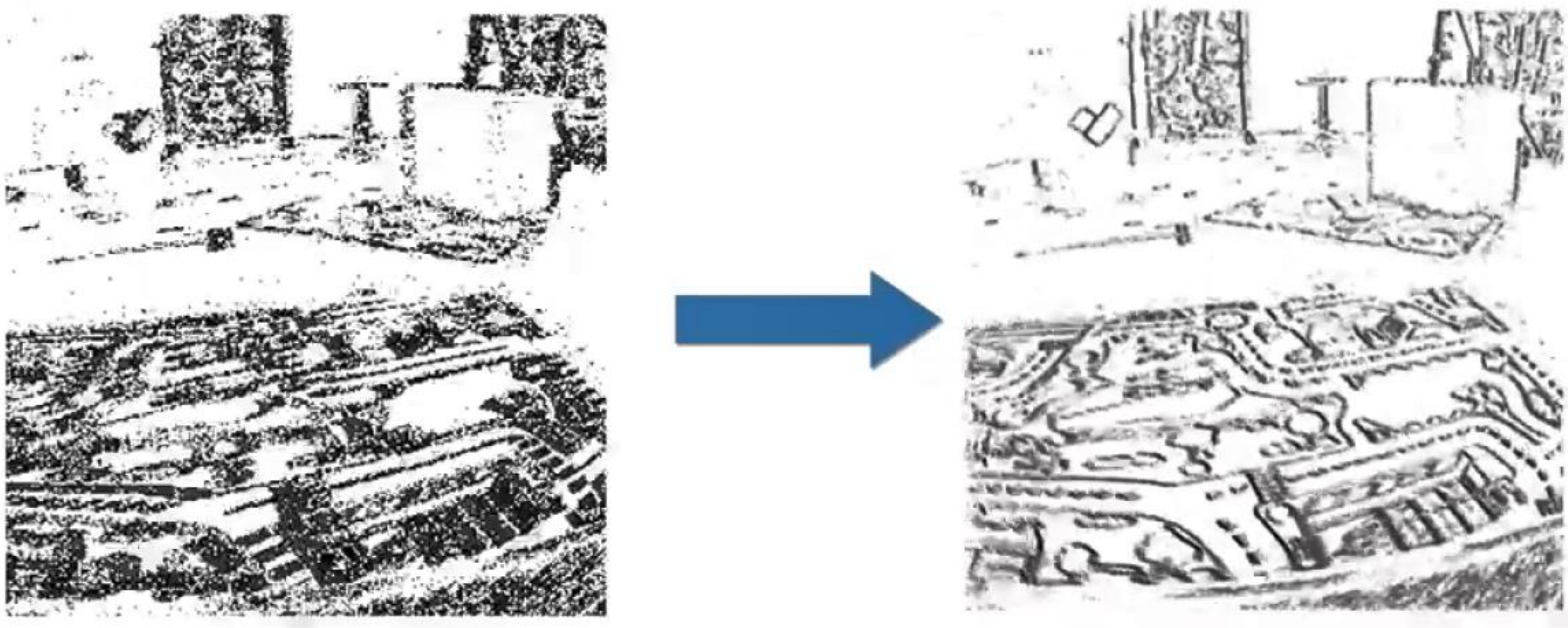


[Video from Zhu et al. (link)]

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

Focus as Loss Function for Unsupervised Learning

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



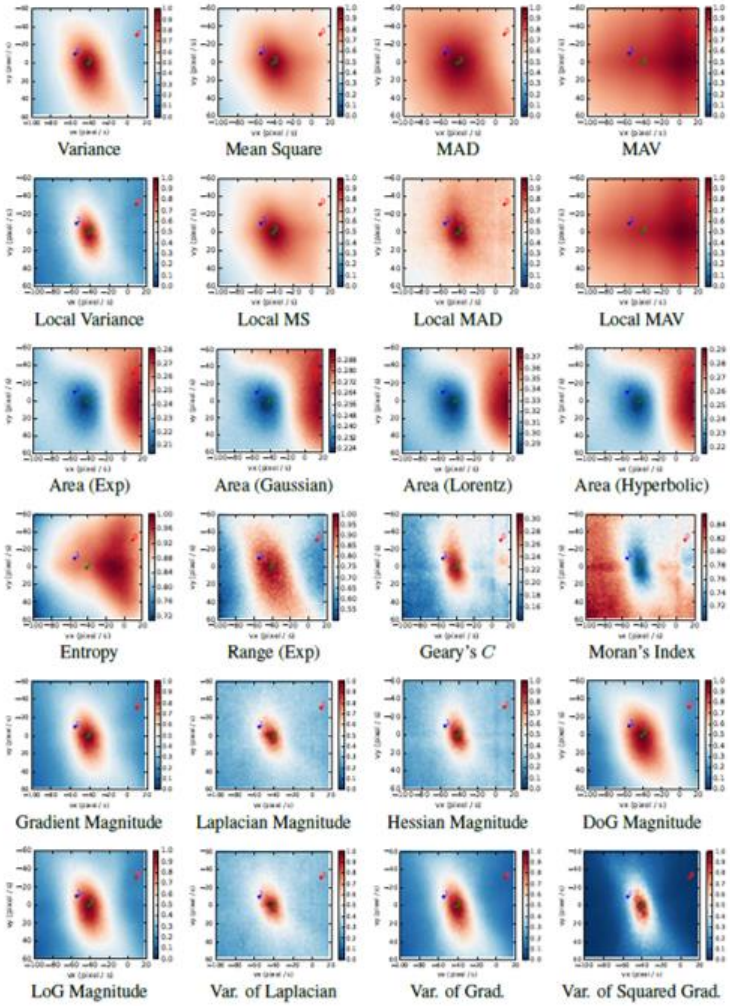
[Video from here]

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].

Focus as Loss Function for Unsupervised Learning

- We proposed and benchmarked **22 focus loss functions**

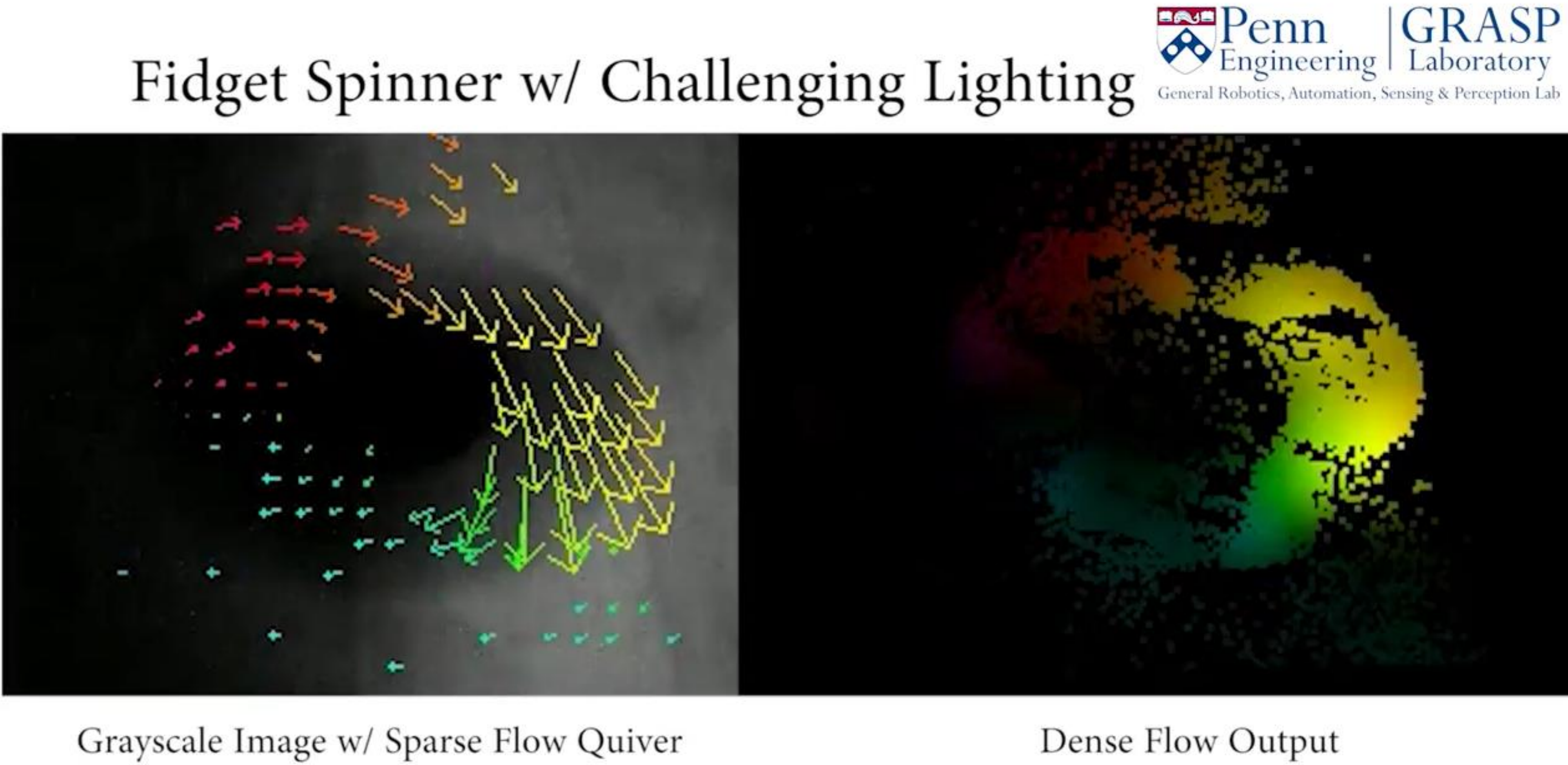
- **Focus is the "data fidelity" term**



Gallego et al., Focus Is All You Need: Loss Functions for Event-based Vision, CVPR19, PDF.

Application: Unsupervised Learning of Optical Flow, Depth and Ego Motion

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

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

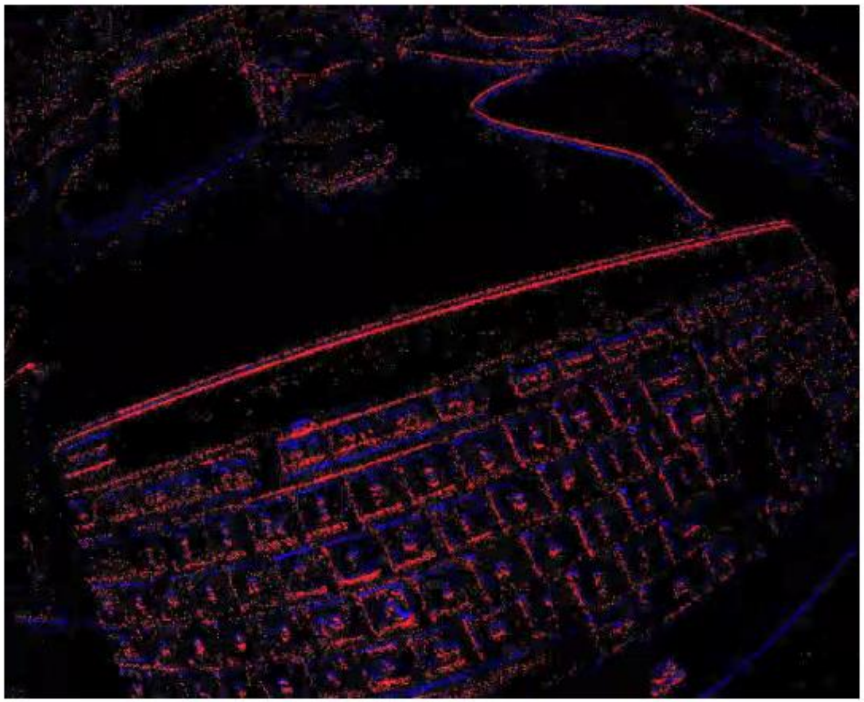
Code & datasets: https://github.com/uzh-rpg/rpg_e2vid

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

Image Reconstruction from Events

Events

Reconstructed image from events (Samsung DVS)

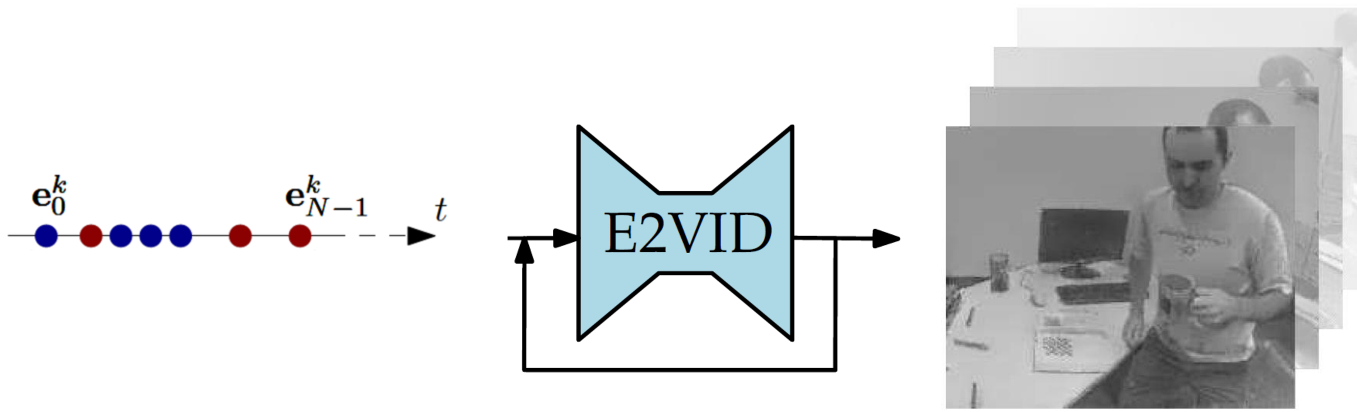


Code & datasets: https://github.com/uzh-rpg/rpg_e2vid

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

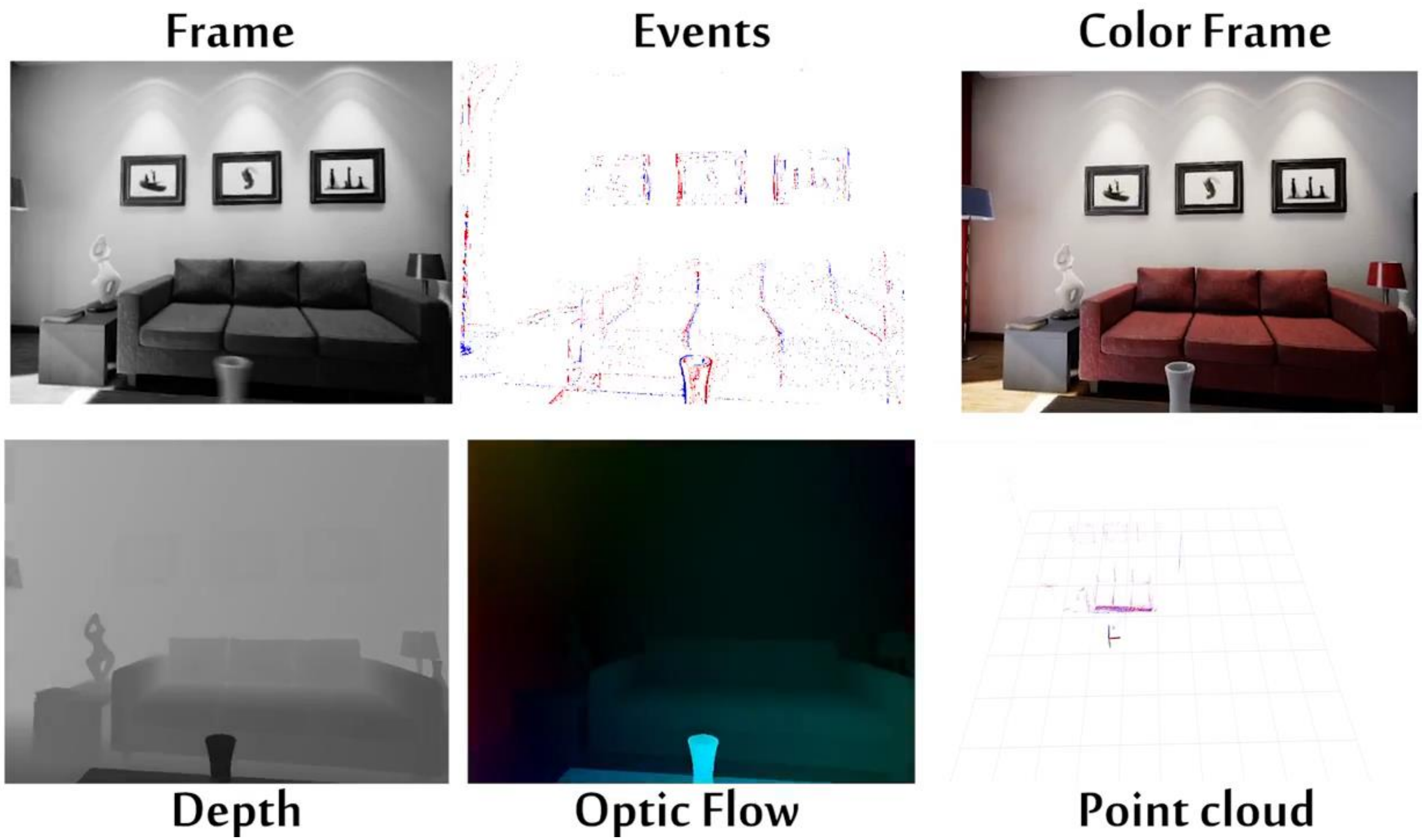
- **Recurrent neural network** (main module: Unet)
- Input: last reconstructed frame + **sequences of *event tensors*** (spatio-temporal 3D voxels grid: each voxel contains sum of ON and OFF events falling within the voxel)
- Network processes **last N events** (10,000)
- **Trained in simulation only**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



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

Event Camera Simulator

- Event Camera Simulator (ESIM): http://rpg.ifi.uzh.ch/esim.html



Rebecq, ESIM: an Open Event Camera Simulator, CORL'18. PDF, YouTube, Project page

High Speed Video Reconstruction Results

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

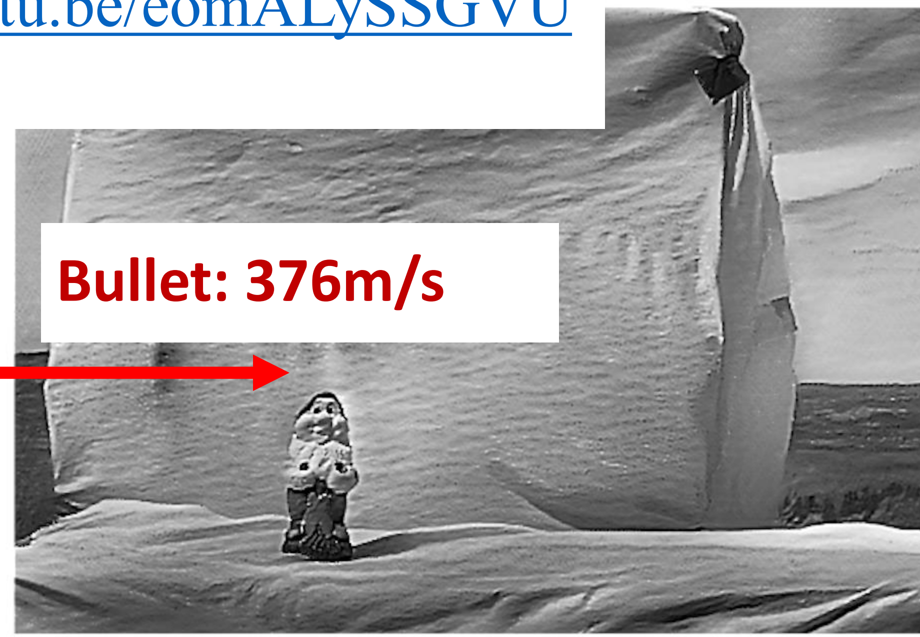
Rebecq et al., "Events-to-Video: Bringing Modern Computer Vision to Event Cameras", CVPR19. PDF Video.
Rebecq et al., "High Speed and High Dynamic Range Video with an Event Camera", PAMI, 2019. PDF Video Code

Popping a water balloon

Recall: trained in simulation only!

YouTube https://youtu.be/eomALySSGVU

Huawei P20 Pro (240 FPS)



Our reconstruction (5400 FPS)

We used Samsung DVS

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

Real time

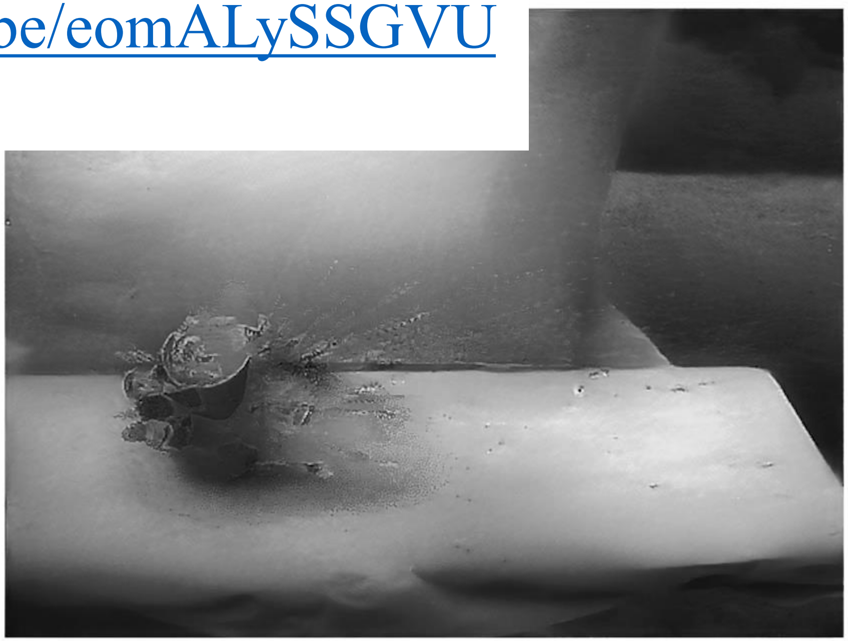
Rebecq et al., "Events-to-Video: Bringing Modern Computer Vision to Event Cameras", CVPR19. PDF Video.
Rebecq et al., "High Speed and High Dynamic Range Video with an Event Camera", PAMI, 2019. PDF Video Code

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

Recall: trained in simulation only!

YouTube https://youtu.be/eomALySSGVU

Huawei P20 Pro (240 FPS)



Our reconstruction (4800 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

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

Recall: trained in simulation only!

https://youtu.be/eomALySSGVU

Huawei P20 Pro (240 FPS)

Our reconstruction (4800 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

Popping a water balloon

Recall: trained in simulation only! Never saw water in simulation

YouTube https://youtu.be/eomALySSGVU

Apple iPad (120 FPS) Our reconstruction (4800 FPS)

* different sequences, recorded in identical conditions 100 x slow motion

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 Reconstruction Results

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’19. PDF Video Code

HDR Video: Driving out of a tunnel

Recall: trained in simulation only!

https://youtu.be/eomALySSGVU

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'19. 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

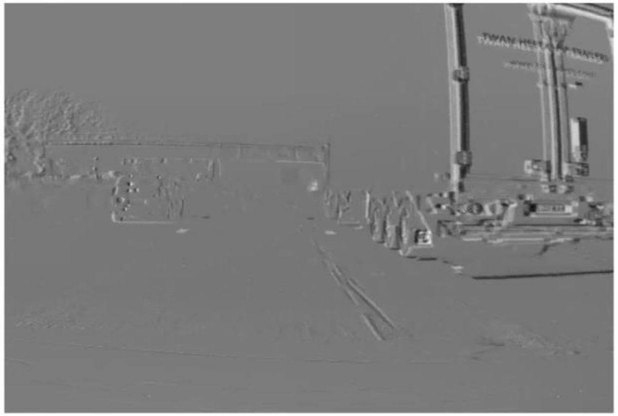
Downstream Applications: What if we input the reconstructed frames to state of the art ML algorithms?

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

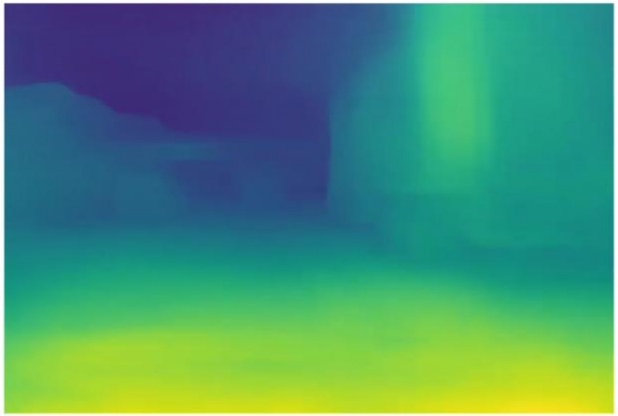
Monocular Depth Estimation

https://youtu.be/eomALySSGVU



Events

Our reconstruction



Monocular depth

Monocular depth estimation (Megadepth) applied on the reconstructed frames

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

Object detection

https://youtu.be/eomALySSGVU

Events

Our reconstruction + object detections (YOLOv3)

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

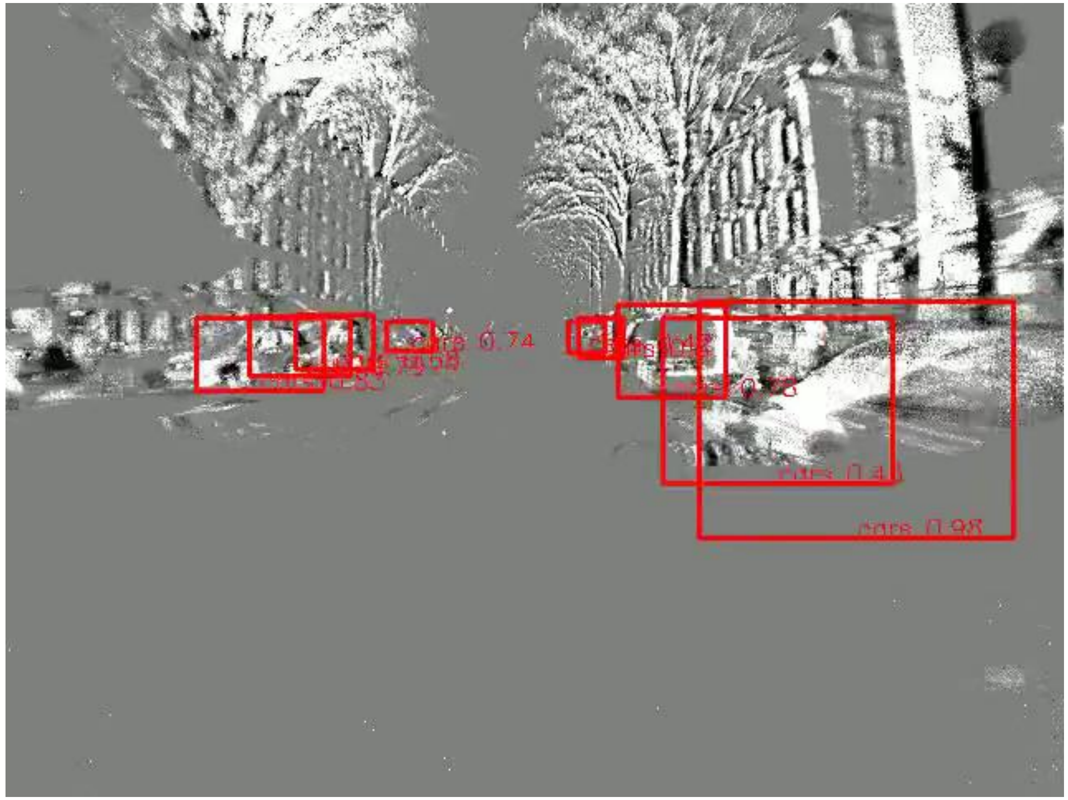
Rebecq et al., "Events-to-Video: Bringing Modern Computer Vision to Event Cameras", CVPR19. PDF Video.
Rebecq et al., "High Speed and High Dynamic Range Video with an Event Camera", PAMI, 2019. PDF Video Code

Does it mean that in order to use event cameras we must first reconstruct an image?

NO!

These results are only meant to show that it should be possible to design algorithms that process events **end-to-end without passing through image reconstruction**!

Example: End-to-End Object Classification



- Dataset from here: https://www.prophesee.ai/dataset-n-cars/
- Collected by PROPHESEE (largest event-camera company)
- Contains: Event, Images, car and pedestrian annotations

Sironi et al., “HATS: Histograms of Averaged Time Surfaces for Robust Event-based Object Classification”. CVPR’18

Spiking Neural Networks (SNN)

- Common processing units based on **Von Neumann architectures** (**CPU and GPU)** are **inefficient & very power consuming for event-by-event processing** [1]
- There exists very efficient, **specialized hardware for event-by-event inference**: **IBM TrueNorth** [1], **Intel Loihi** [2], DynapSE & Speck (AiCTX) [3]
- Promising for Robotics, IoT, VR/AR/MR
  - Low power
  - Low latency
  - Leverage event-based sensing
- **Promise ultra-low carbon footprint!**

[1]: Merolla et. al. A million spiking-neuron integrated circuit with a scalable communication network and interface. *Science*. 2014
[2]: Davies M. et. al. Loihi: A neuromorphic manycore processor with on-chip learning. IEEE Micro. 2018
[3]: Moradi S. et. al.. A scalable multicore architecture with heterogeneous memory structures for dynamic neuromorphic asynchronous processors. IEEE transactions on biomedical circuits and systems. 2017. https://aictx.ai/

The Cost of Current Computer Technologies is Not Sustainable

- **In 2017**, **> 10 zettabytes** of data were produced.
- **IT infrastructures and consumer electronics** absorbed **> 10% of the global electricity supply**.
- By **2025**, over **50 billion of Internet-of-Things** (IoT) devices will be interconnected.
- **Over 180 zettabytes of data will be generated annually**, potentially leading to a consumption of one-fifth of global electricity (source, Nature, Feb., 2018)

nature

EDITORIAL · 06 FEBRUARY 2018

Big data needs a hardware revolution

Artificial intelligence is driving the next wave of innovations in the semiconductor industry.

"Software companies make headlines but research on computer could bring bigger rewards. "

MIT Technology Review

Training a single AI model can emit as much carbon as five cars in their lifetimes

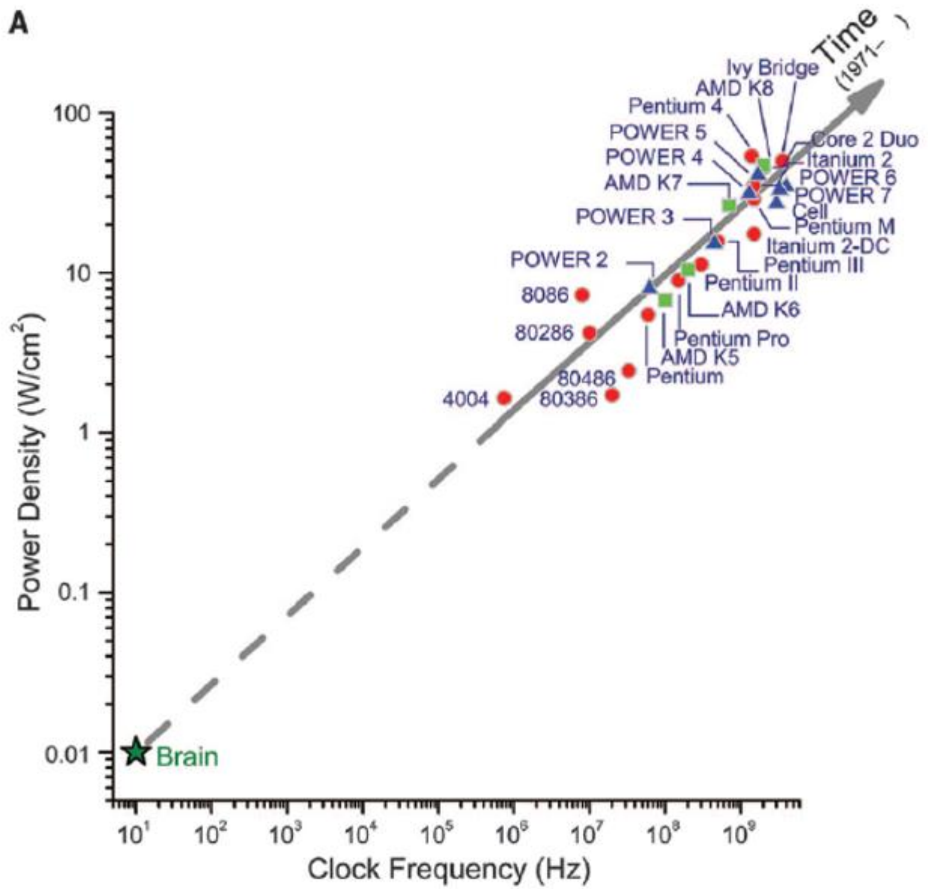
Deep learning has a terrible carbon footprint.

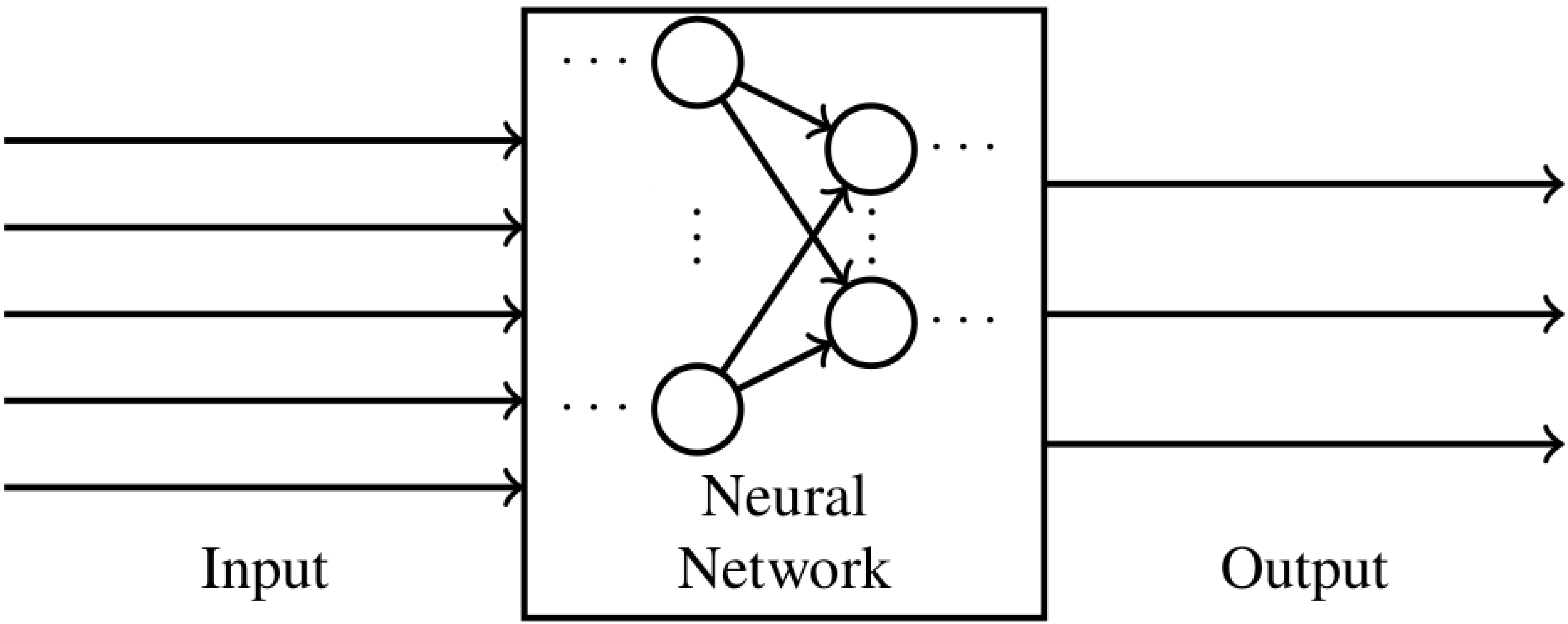
by **Karen Hao** Jun 6, 2019

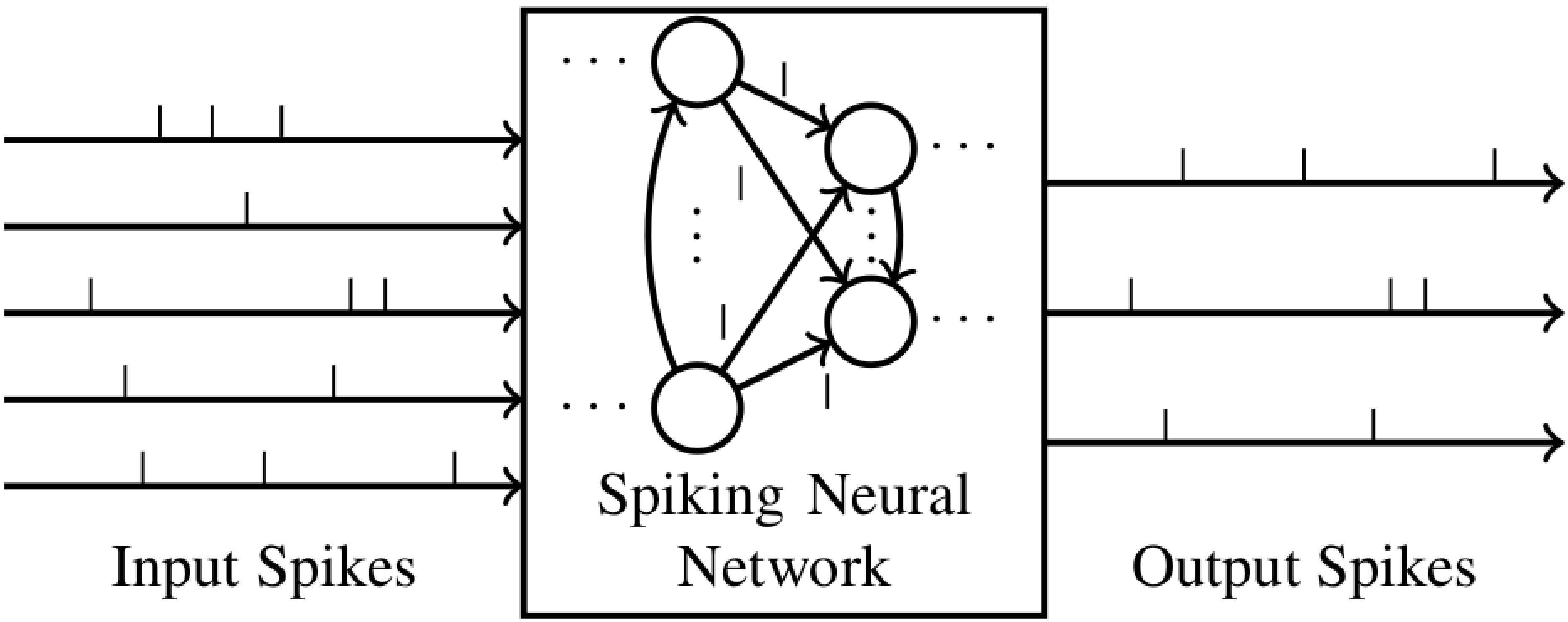
The artificial-intelligence industry is often compared to the oil industry: once mined and refined, data, like oil, can be a highly lucrative commodity. Now it seems the metaphor may extend even further. Like its fossil-fuel counterpart, the process of deep learning has an outsize environmental impact.

Radical paradigm shift in computer hardware technologies

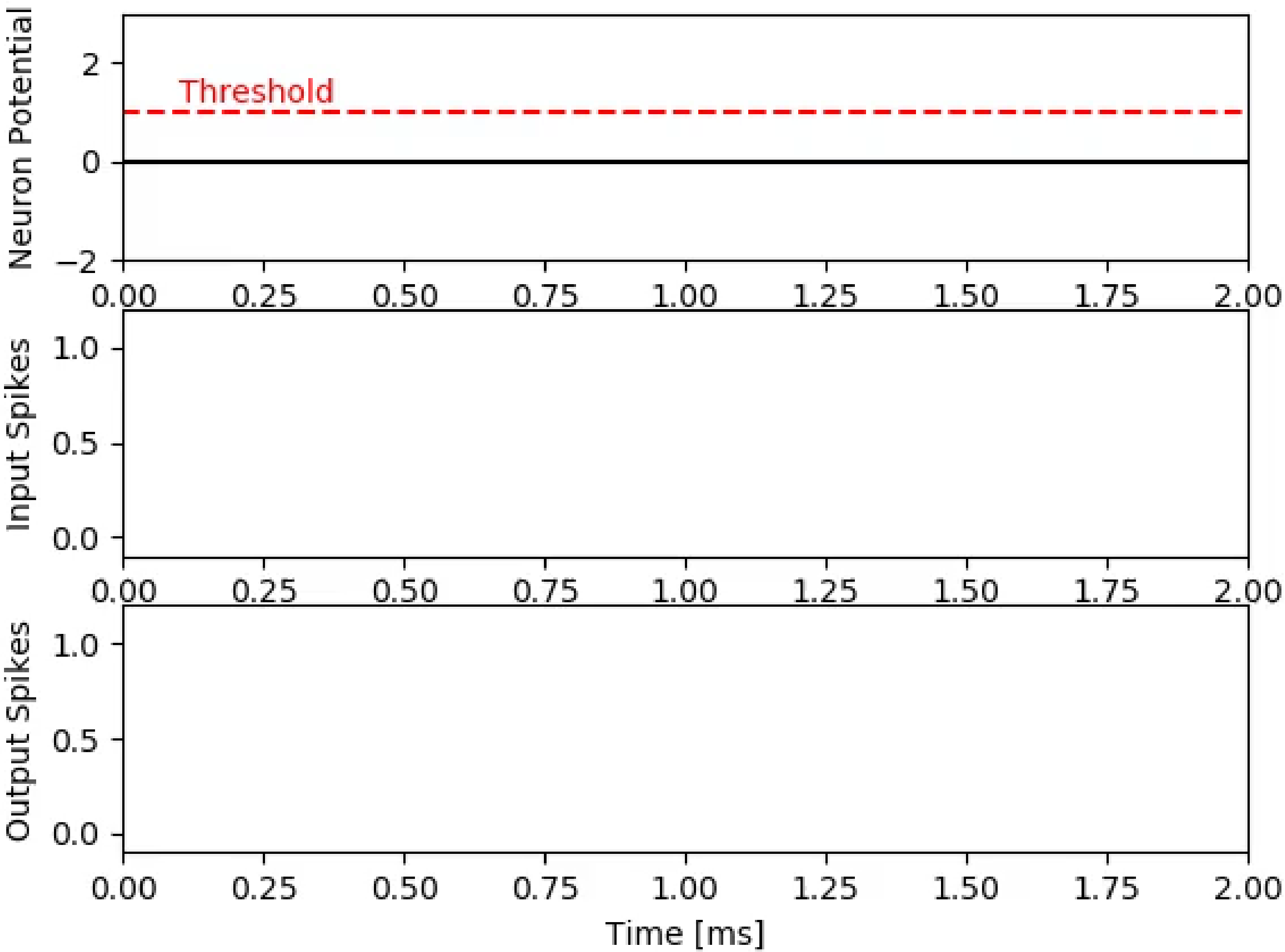
- Our brain is slow, noisy (**“speed” is not a requirement**)
- Massively parallel distributed computation, local connectivity(minimize wiring)
- Real-time interaction with the environment
- **Complex spatio-temporal pattern recognition**

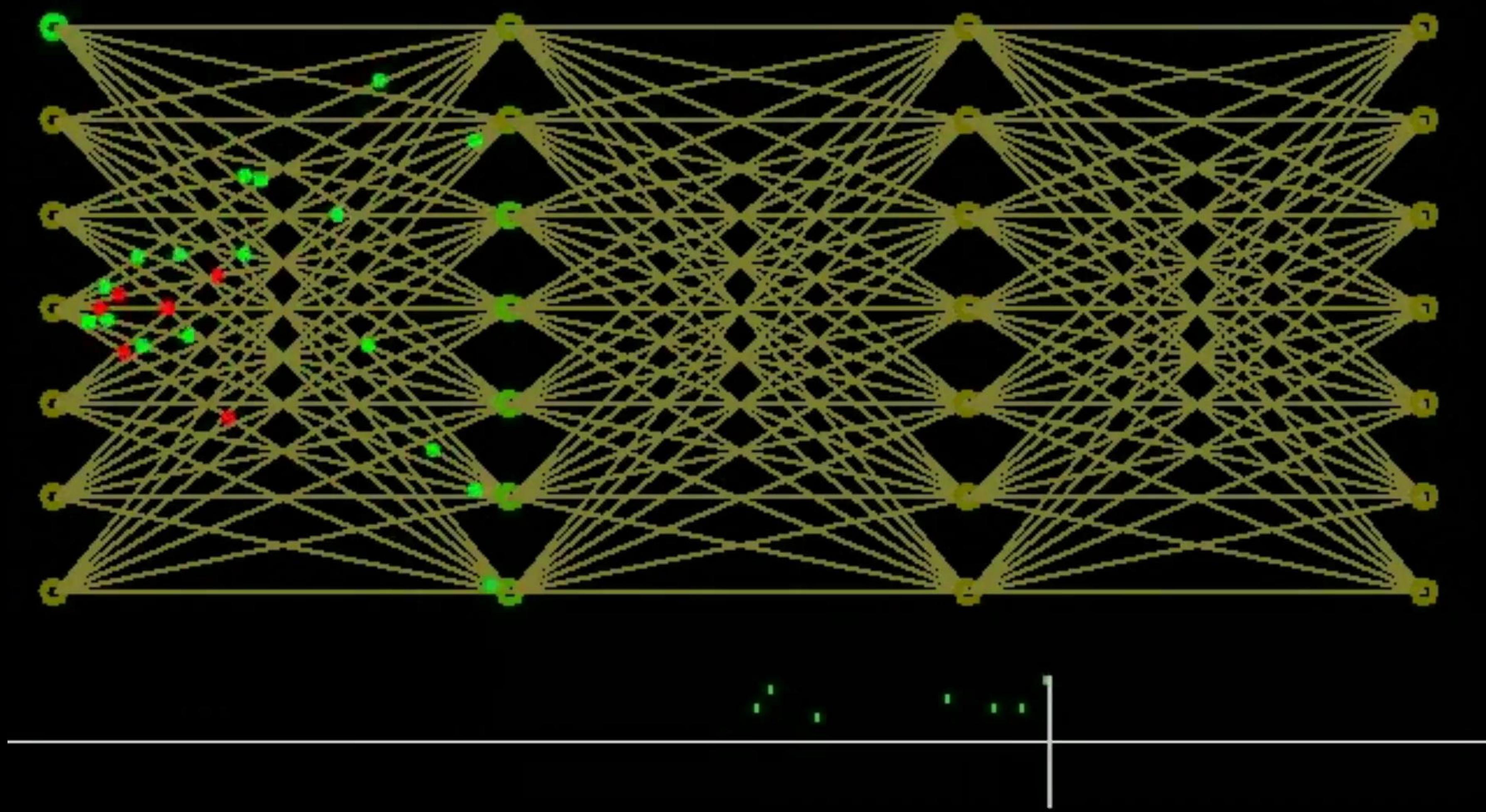



Neural
Network
Input
Output


Spiking Neural
Network
Input Spikes
Output Spikes

Model of a Spiking Neuron





Mior Roberto (2010). Spiking Neural Network. https://youtu.be/lhIdisK7akI

Origins of Spiking Neural Networks

- **First model** (integrate-and-fire) of **a spiking neuron in 1907** by Louis Lapicque [1]
- **First computational model for neural networks in 1943** [2]: Neural network research split into biological processes in the brain and the application for artificial intelligence
- First scientific model of biological spike propagation by Hodgkin and Huxley in **1951** [3] (**Nobel Prize in Physiology**)
- A range of more general spiking neuron models are available nowadays [4]

[1]: Lapicque L. Recherches quantitatives sur l'excitation electrique des nerfs traitee comme une polarization. *Journal de Physiologie et de Pathologie Generale*. 1907
[2]: McCulloch WS, Pitts W. A logical calculus of the ideas immanent in nervous activity. *The bulletin of mathematical biophysics*. 1943
[3]: Hodgkin AL, Huxley AF. A quantitative description of membrane current and its application to conduction and excitation in nerve. *The Journal of physiology*. 1952
[4]: Gerstner W. Time structure of the activity in neural network models. *Physical review*. 1995

SNNs: Current Applications and Demos

- IBM TrueNorth
  - Targeting ultra low-power gesture control, video-surveillance and IoT with SNN on **digital processor**
  - At CVPR 2017 gesture recognition demo (10 gestures)
  - 96.5 % recognition accuracy
  - **200 mW power consumption** (event-camera + processing)
- Intel Loihi:
  - Targeting ultra low-power surveillance and IoT with **SNN** on **analog processor**
- aiCTX (Zurich-based startup):
  - Targeting ultra low-power surveillance and IoT with **SNN** on **analog processor**
  - At CES'19 and CVPR'19, they demonstrated face recognition from event data on an SNN processor; total power consumption: **1mW**

[1]: https://aictx.ai/

[2]: Merolla et. al. A million spiking-neuron integrated circuit with a scalable communication network and interface. *Science*. 2014

IBM-TrueNorth + Samsung DVS, live gesture recognition demo at CVPR 2017 (10 gestures)

Recap

- **Event cameras** have many **advantages**:
  - high dynamic range (HDR)
  - high speed
  - low latency
  - low power

- **Current commercial applications**
  - IoT
    - monitoring and surveillance
  - Automotive:
    - low-latency detection, object classification, low-power and low-memory storage
  - AR/VR
    - low-latency, inter-frame pose estimation, low-power
  - Industrial automation
    - Fast pick and place

Research Challenges with Event Cameras

- Quantify the **trade-offs**:
  - **Latency vs. power consumption** and **accuracy**
  - **Sensitivity vs. bandwidth** and **processing** capacity
- **Active parameter adaptation**
- **Hardware:**
  - pairing event cameras with dedicated hardware (SNN hardware, e.g., Intel Loihi, aiCTX Speck)
  - How do we make sparse convolution in space and time efficient?
- **Learning** with event cameras:
  - How do we **exploit knowledge from image-based learning to event cameras**?
  - **Asynchronous** inference
  - **Where do we find learning data**? Event data is much more rare than frames. Potential solutions: unsupervised Learning, learning in simulation, transfer learning from frames to events

Event-based Vision: A Survey

Guillermo Gallego, Tobi Delbrück, Garrick Orchard, Chiara Bartolozzi, Brian Taba, Andrea Censi, Stefan Leutenegger, Andrew Davison, Jörg Conradt, Kostas Daniilidis, Davide Scaramuzza

Abstract— Event cameras are bio-inspired sensors that work radically different from traditional cameras. Instead of capturing images at a fixed rate, they measure per-pixel brightness changes asynchronously. This results in a stream of events, which encode the time, location and sign of the brightness changes. Event cameras posses outstanding properties compared to traditional cameras: very high dynamic range ($140\,\mathrm{dB}$ vs. $60\,\mathrm{dB}$), high temporal resolution (in the order of μs), low power consumption, and do not suffer from motion blur. Hence, event cameras have a large potential for robotics and computer vision in challenging scenarios for traditional cameras, such as high speed and high dynamic range. However, novel methods are required to process the unconventional output of these sensors in order to unlock their potential. This paper provides a comprehensive overview of the emerging field of event-based vision, with a focus on the applications and the algorithms developed to unlock the outstanding properties of event cameras. We present event cameras from their working principle, the actual sensors that are available and the tasks that they have been used for, from low-level vision (feature detection and tracking, optic flow, etc.) to high-level vision (reconstruction, segmentation, recognition). We also discuss the techniques developed to process events, including learning-based techniques, as well as specialized processors for these novel sensors, such as spiking neural networks. Additionally, we highlight the challenges that remain to be tackled and the opportunities that lie ahead in the search for a more efficient, bio-inspired way for machines to perceive and interact with the world.

Index Terms—Event Cameras, Bio-Inspired Vision, Asynchronous Sensor, Low Latency, High Dynamic Range, Low Power.

1 Introduction and Applications

"THE *brain is imagination, and that was exciting to me; I wanted to build a chip that could imagine something*[1]." that is how Misha Mahowald, a graduate student at Caltech in 1986, started to work with Prof. Carver Mead on the stereo problem from a joint biological and engineering perspective. A couple of years later, in 1991, the image of a cat in

as well as new computer vision and robotic tasks. Sight is, by far, the dominant sense in humans to perceive the world, and, together with the brain, learn new things. In recent years, this technology has attracted a lot of attention from both academia and industry. This is due to the availability of prototype event cameras and the advantages that these devices offer to tackle problems that are currently unfeasible

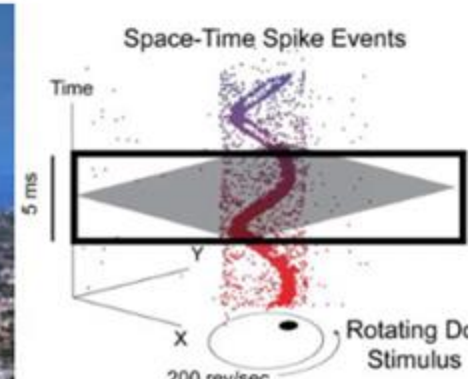
http://rpg.ifi.uzh.ch/docs/EventVisionSurvey.pdf

CVPR19 Workshop on Event-based Vision

- Full-day workshop with talks by 23 researchers on event-based cameras, including Samsung, Intel, and event-camera companies
- Slides and video recordings: http://rpg.ifi.uzh.ch/CVPR19_event_vision_workshop.html

Second International Workshop on Event-based Vision and Smart Cameras
June 17, Long Beach

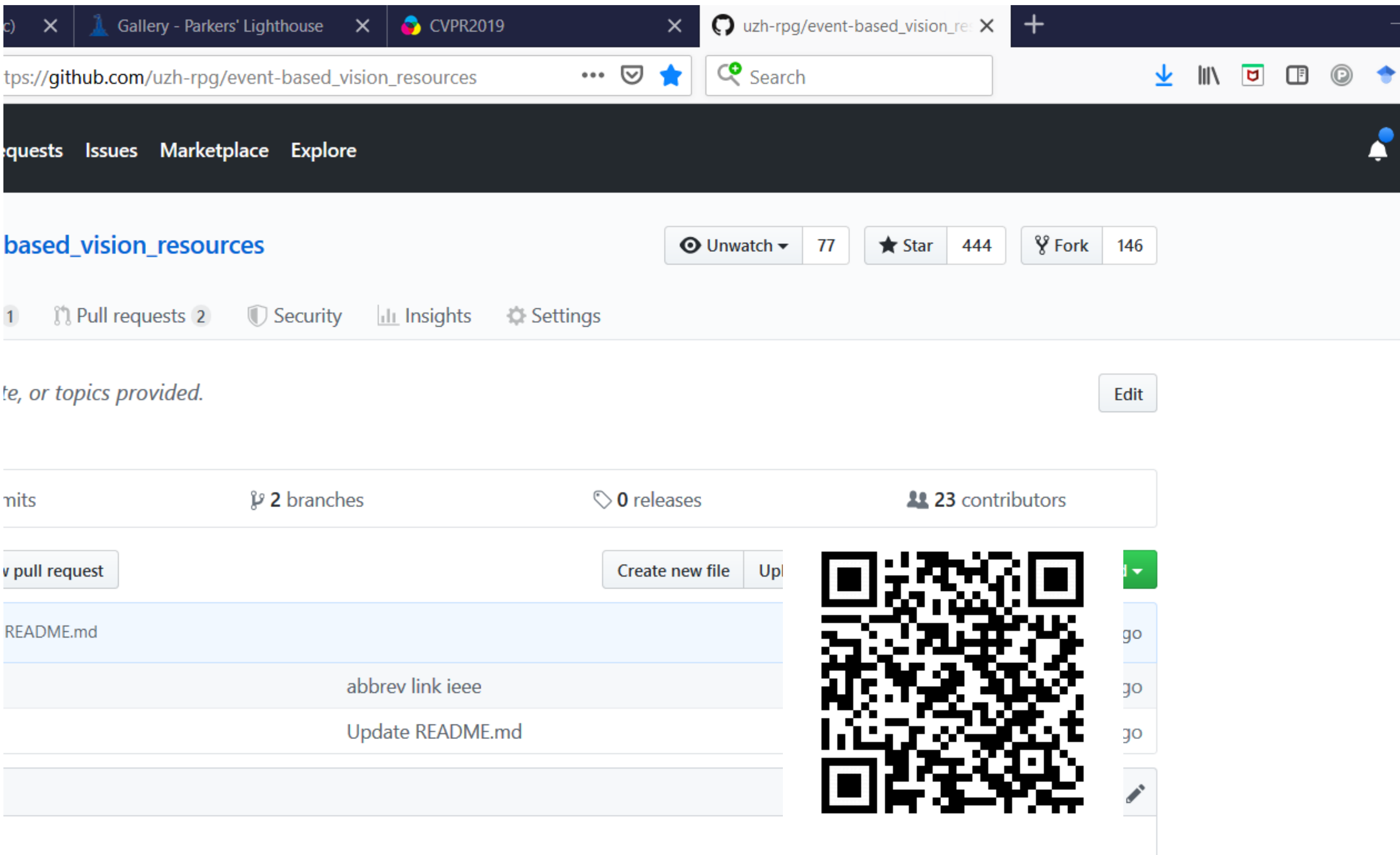
Held in conjuction with the IEEE Conference on Computer Vision and Pattern Recognition, Long Beach.



List of Event-based Vision resources

Code, datasets, papers, videos, companies on event cameras

https://github.com/uzh-rpg/event-based_vision_resources



UZH-FPV Drone Racing Dataset & Competition

- Recorded with a drone flown by a **professional pilot up to over 20m/s**
- Contains over 30 sequences with **images, events, IMU**, and **ground truth from a robotic total station**: http://rpg.ifi.uzh.ch/uzh-fpv.html
- **IROS 2019 Drone Racing VIO competition**: https://github.com/uzh-rpg/IROS2019-FPV-VIO-Competition **Win $1,000 plus invited talk** . **Submission deadline: Sep. 1, 2019**

Delmerico et al. "Are We Ready for Autonomous Drone Racing? The UZH-FPV Drone Racing Dataset" ICRA'19
PDF. Video. Datasets.

Thanks!

Code, datasets, simulators, papers, and videos:

http://rpg.ifi.uzh.ch/research_dvs.html

Research updates:

@davsca1

@davidescaramuzza