Event-based Robot Vision

Prof. Dr. Guillermo Gallego
Chair: Robotic Interactive Perception

guillermo.gallego@tu-berlin.de

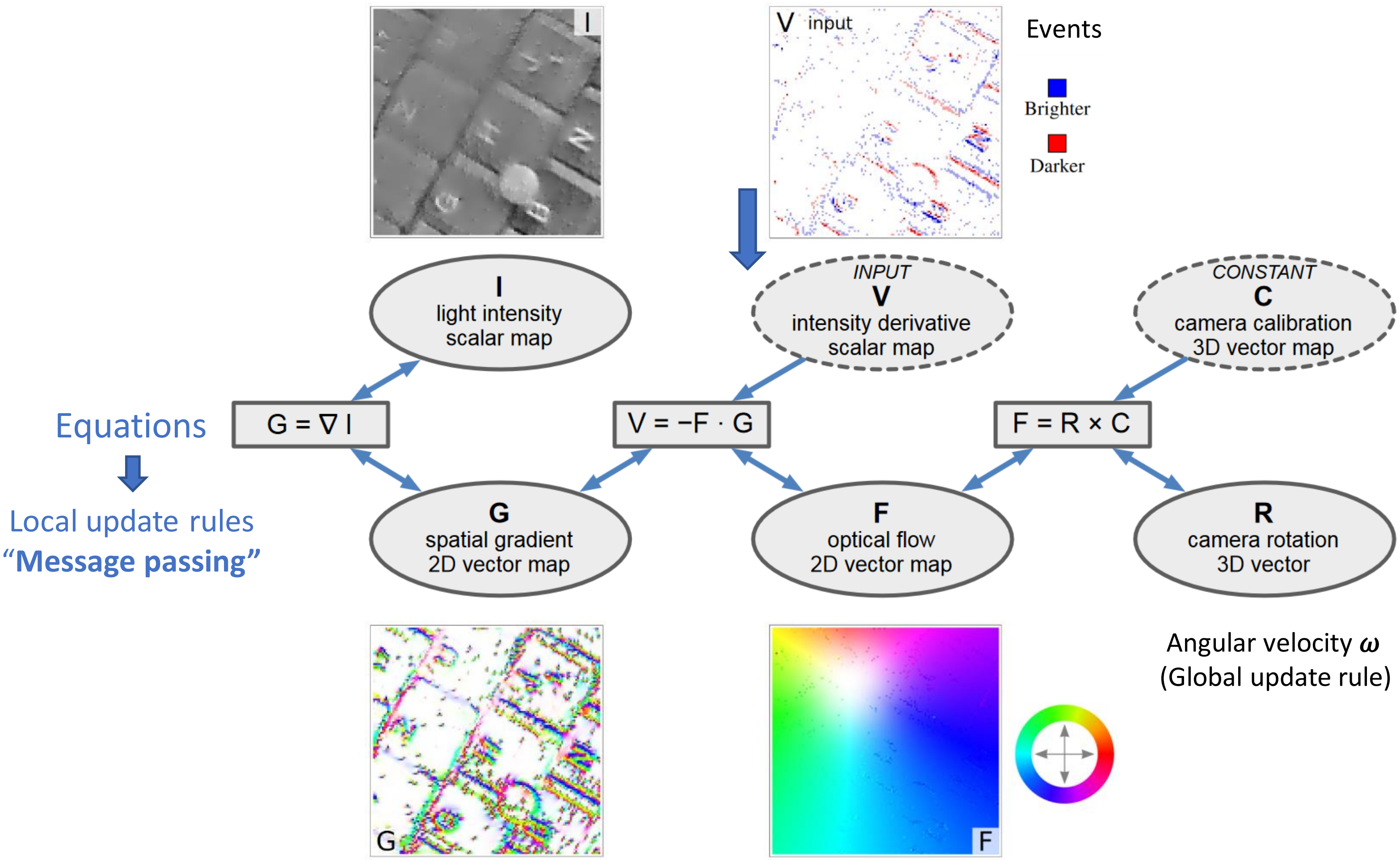
http://www.guillermogallego.es

Image (intensity) Reconstruction

Literature Review

“Interacting Maps”

- Simultaneous estimation of multiple visual quantities with a *rotating* event camera



Cook et al., *[Interacting maps for fast visual interpretation](#)*, IJCNN 2011. https://youtu.be/irX3Nd5U0hY

"Interacting Maps"

- Simultaneous estimation of multiple visual quantities with a *rotating* event camera

Image intensity Image gradient magnitude

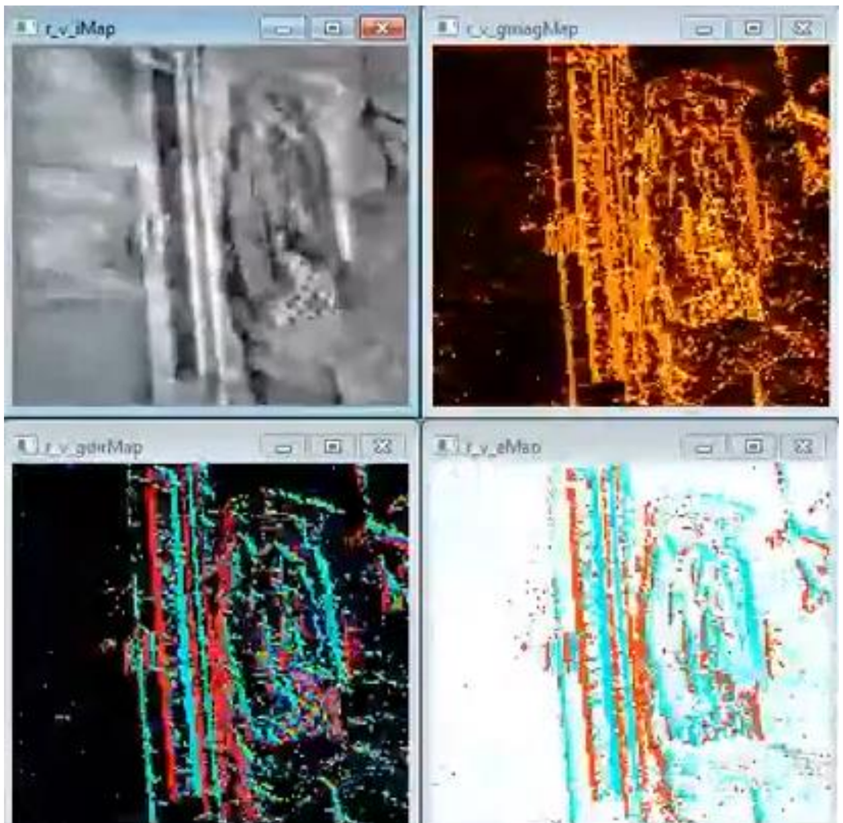


Image gradient direction Events (input)

Inspired by the primary visual cortex.

From the events, jointly estimate:

- Rotation (3 DOF ego-motion)
- Optical flow
- Image gradient
- Intensity reconstruction ←

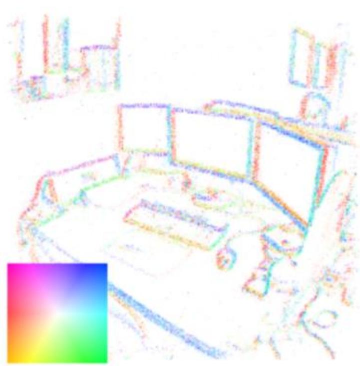
Pure rotation: no translation or depth

Cook et al., *Interacting maps for fast visual interpretation*, IJCNN 2011. https://youtu.be/irX3Nd5U0hY

Simultaneous mosaicing and tracking

- Parallel tracking and mapping
- *Rotating* event camera (no translation or depth)

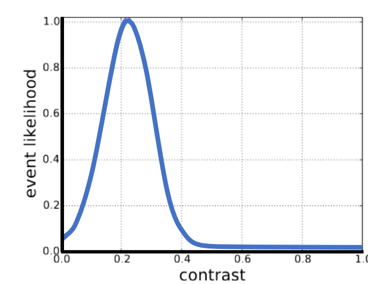
- **Mapping** = mosaicing (panoramic imaging)
  - Get intensity gradient g using pixel-wise EKF: $h(e|R) = \frac{g \cdot v}{C}$ should equal the pixel event rate $\frac{1}{\Delta\tau}$ (uses linearized event generation model)
  - Poisson reconstruction to obtain intensity



Intensity gradient

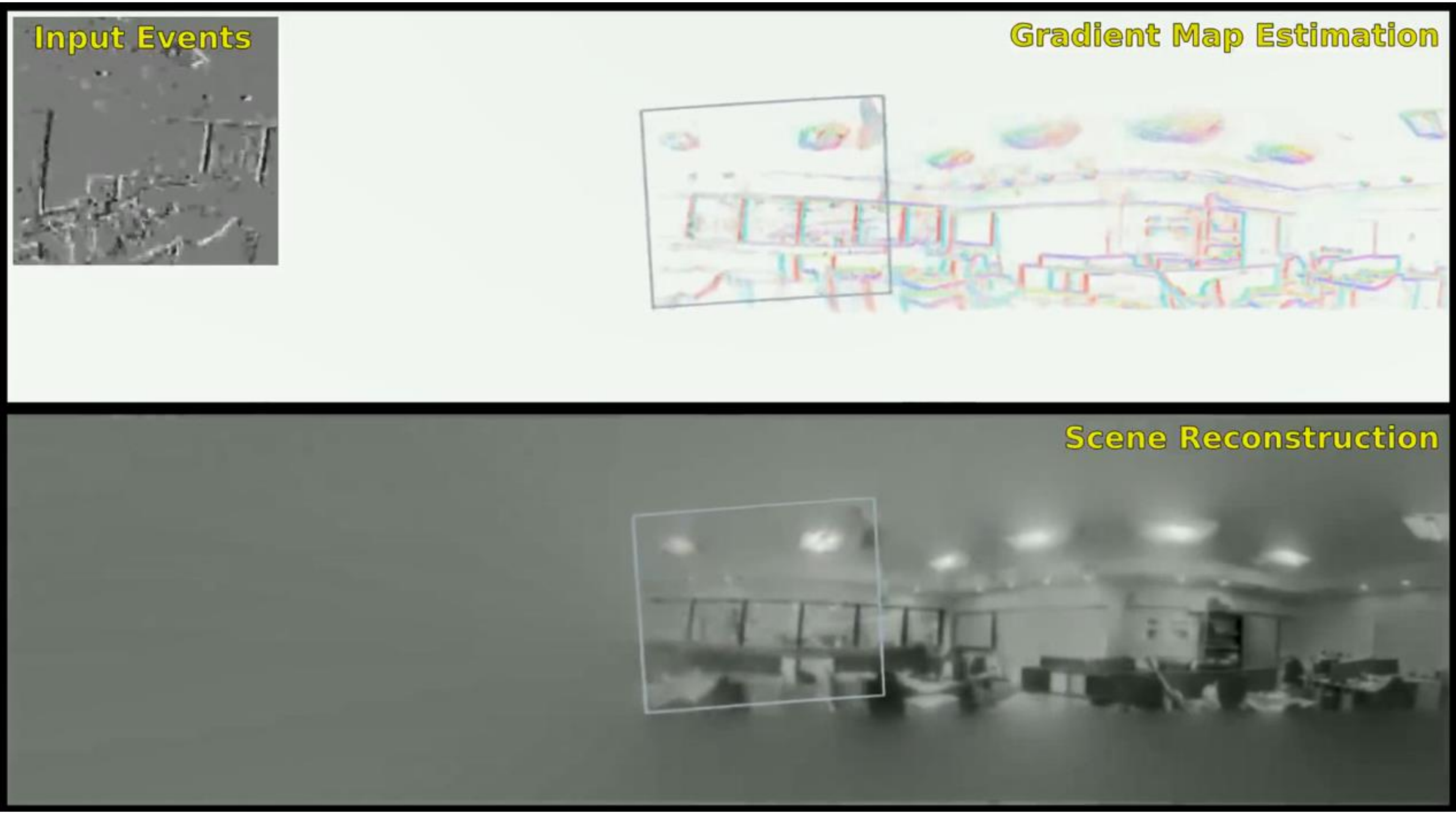
Image intensity

- **Tracking** (ego-motion estimation)
  - Random diffusion in motion space
  - Particle filter: particle weights updated using the map: $\text{contrast} = |\log I(t) - \log I(t - \Delta t)|$



Kim et al., *Simultaneous Mosaicing and Tracking with an Event Camera*, BMVC 2014

Simultaneous mosaicing and tracking

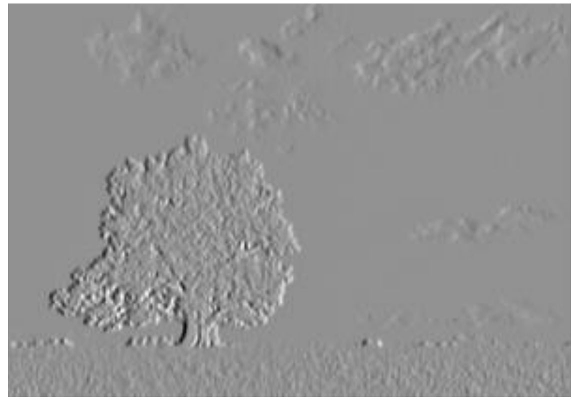


Kim et al., [*Simultaneous Mosaicing and Tracking with an Event Camera*](), BMVC 2014. https://youtu.be/l6qxeM1DbXU

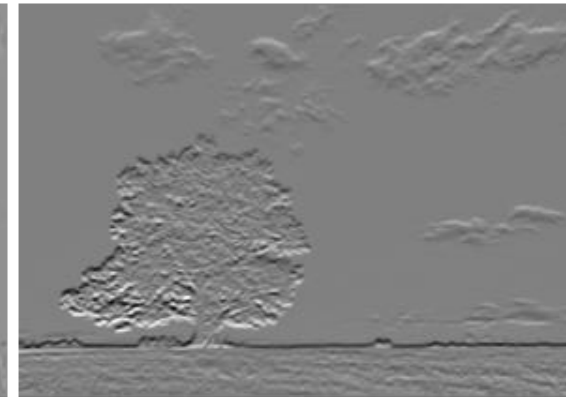
Image reconstruction by Poisson integration

Integrate gradient map $\boldsymbol{g}$ to get absolute brightness M

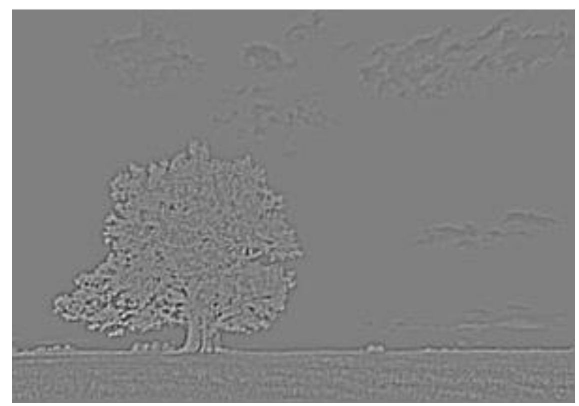
Original Image



Gradient in *x* direction
$(g_\mathrm{x} = \partial_\mathrm{x} I)$



Gradient in *y* direction
$(g_\mathrm{y} = \partial_\mathrm{y} I)$



Divergence
$(\operatorname{div} \boldsymbol{g} = \partial_\mathrm{x} g_\mathrm{x} + \partial_\mathrm{y} g_\mathrm{y})$

2D integration

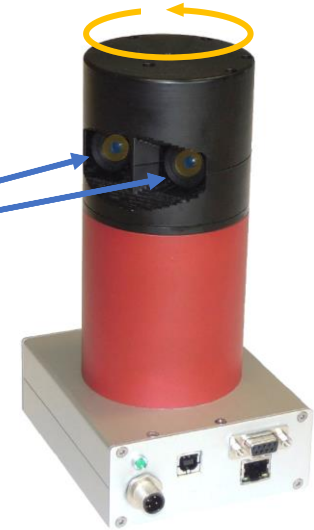
Solve Poisson eq:
$(\Delta \tilde{I} = \operatorname{div} \boldsymbol{g})$
fast using the FFT

Reconstructed Image

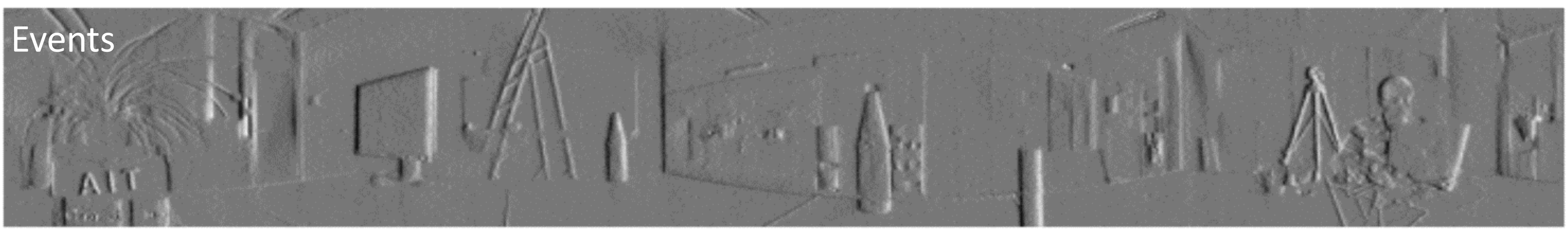
Slide by T. Horstschäfer

TUCO-3D Panoramic Imaging

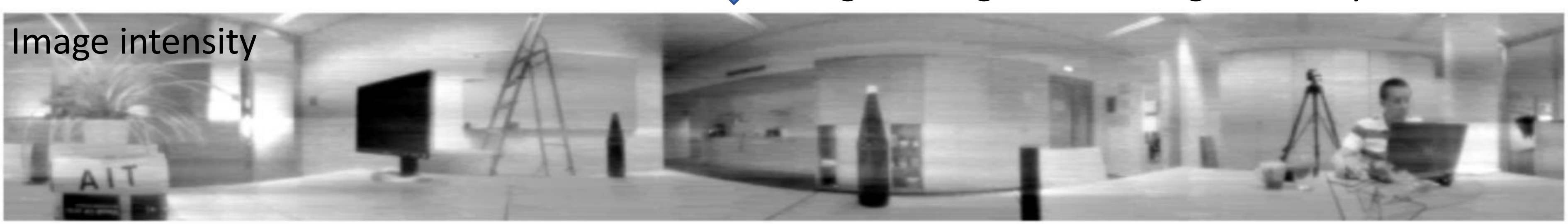
- 360 deg Panoramic Imaging
- Exploit constrained motion
  - Two rotating 1D event cameras (stereo)
  - Event integration must satisfy periodic boundary conditions
- It also provides depth (3D)



TUCO-3D by AIT



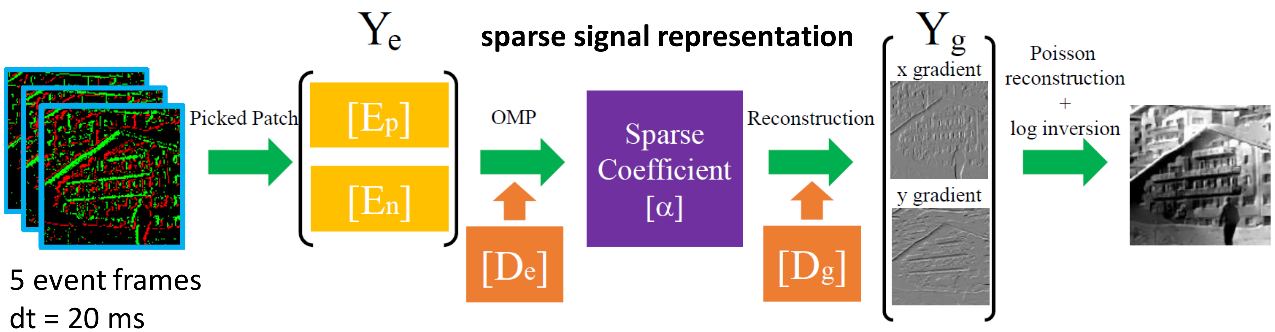
Integrate brightness changes, line by line

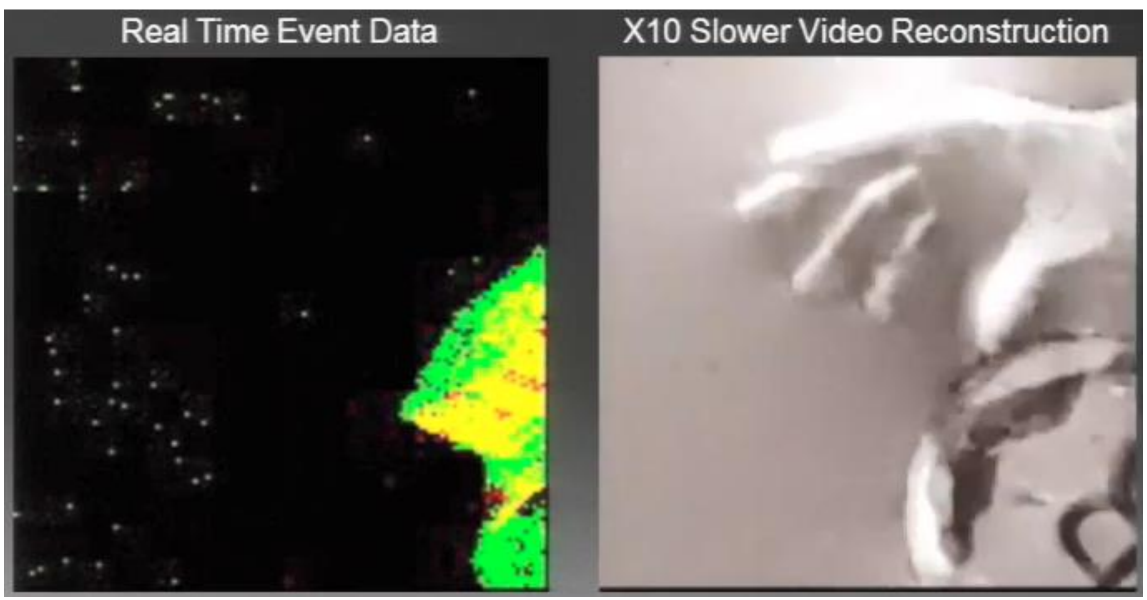


Belbachir et al., *A Novel HDR Depth Camera for Real-time 3D 360-degree Panoramic Vision*, CVPR Workshops, 2014

Image Reconstruction using a Sparse Dictionary

- Shows that there is **no need to estimate motion** for reconstruction
- **Patch-based dictionary of events** learnt from simulated data





Green: positive
Red: negative
Yellow: both events

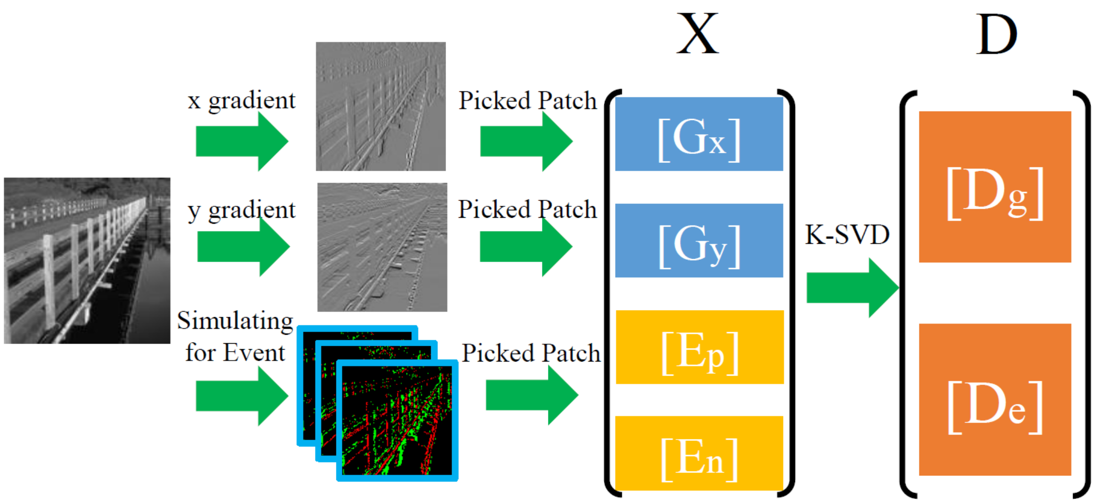
Video at **2 kHz**

Patches of 9 x 9 pixels

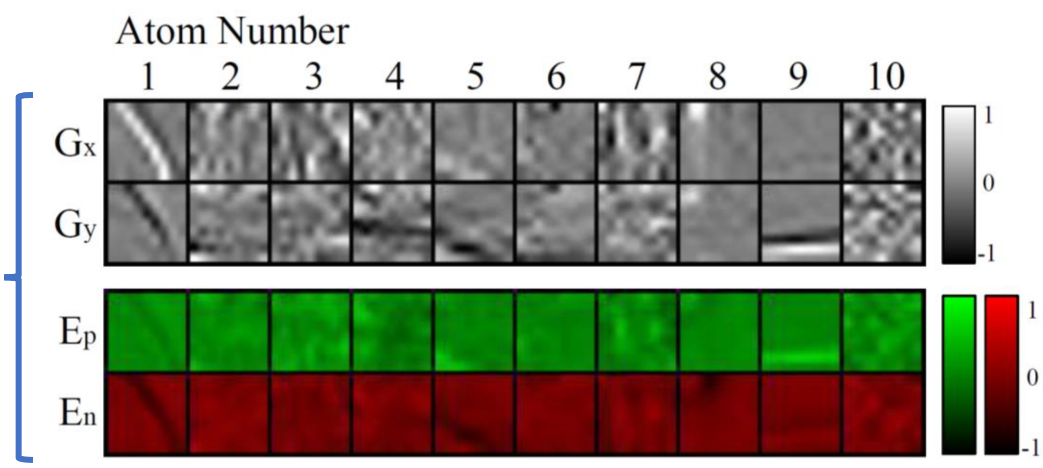
Barua et al., *[Direct face detection and video reconstruction from event cameras](#)*, WACV 2016. [YouTube](#)

Image Reconstruction using a Sparse Dictionary

- How is the dictionary computed (i.e., learned)?



K-SVD is a generalization of the **k-means clustering** method (i.e., **unsupervised**)

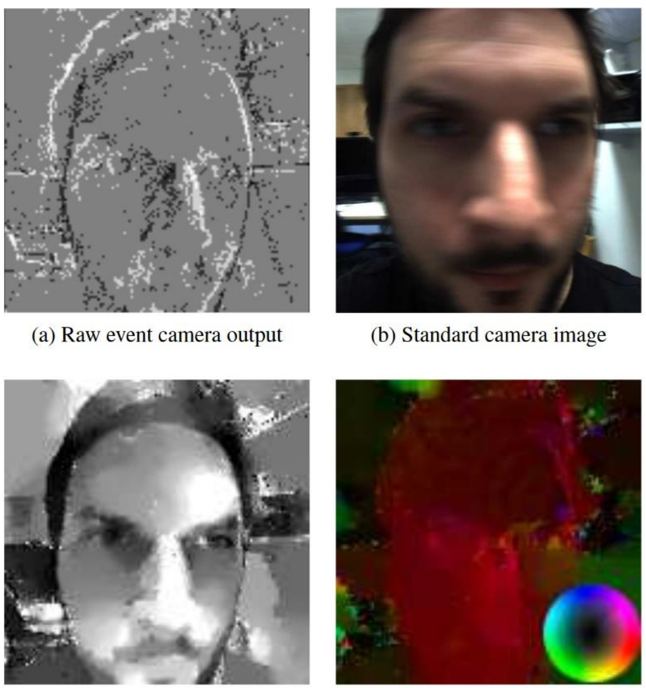


These are the 4 components in the dictionary.

Given positive (Ep) and negative (En) events, we have corresponding Gx, Gy

SOFIE: Simultaneous Optical Flow & IE

- **Joint optimization** over image reconstruction and **optical flow** to explain a volume of events (voxel grid)
- More relaxed than SLAM methods; just need optical flow, i.e., works on **dynamic scenes** with **little assumptions about motion**.
- Solve a variational optimization problem:



(a) Raw event camera output (b) Standard camera image

(c) Intensity estimate from events (d) Optical flow from events

$$\min_{\mathbf{u}, L} \int_\Omega \int_T \Big(\lambda_1 \|\mathbf{u}_\mathbf{x}\|_1 + \lambda_2 \|\mathbf{u}_t\|_1 + \lambda_3 \|L_\mathbf{x}\|_1 + \lambda_4 \|\langle L_\mathbf{x}, \delta_t \mathbf{u} \rangle + L_t\|_1 + \lambda_5\, h_\theta(L - L(t_p)) \Big) \mathrm{d}t\, \mathrm{d}\mathbf{x} + \int_\Omega \sum_{i=2}^{|P(\mathbf{x})|} \|L(t_i) - L(t_{i-1}) - \theta \rho_i\|_1 \mathrm{d}\mathbf{x},$$

Smoothness terms

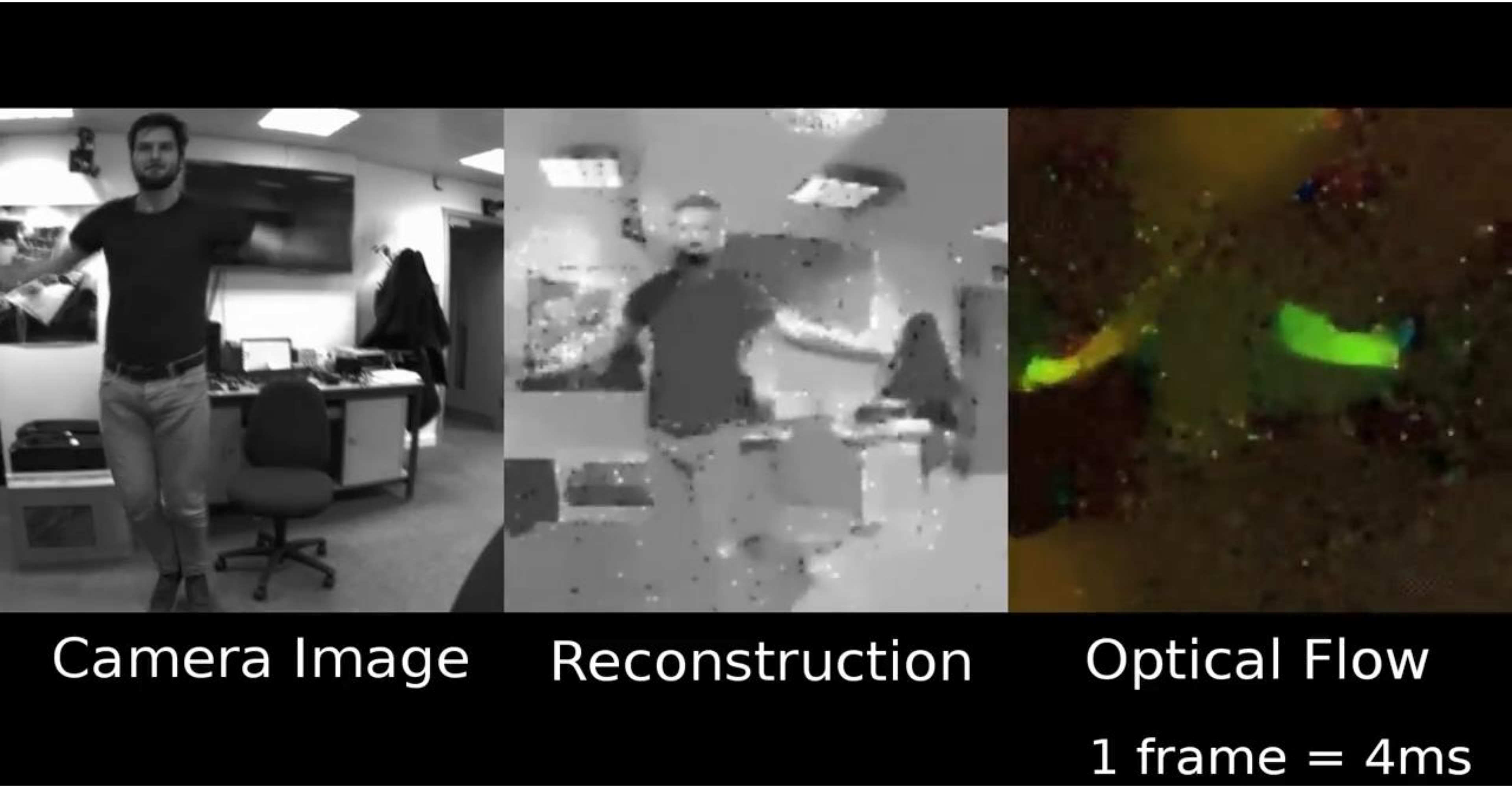
Optical flow term (brightness constancy)

No-event term

Event term

Bardow et al., *Simultaneous Optical Flow and Intensity Estimation from an Event Camera*, CVPR 2016. YouTube

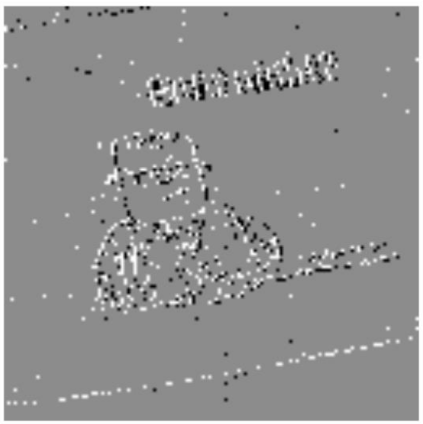
SOFIE: Simultaneous Optical Flow & IE



Bardow et al., *Simultaneous Optical Flow and Intensity Estimation from an Event Camera*, CVPR 2016. YouTube

Reconstruction using "Manifold Regularisation"

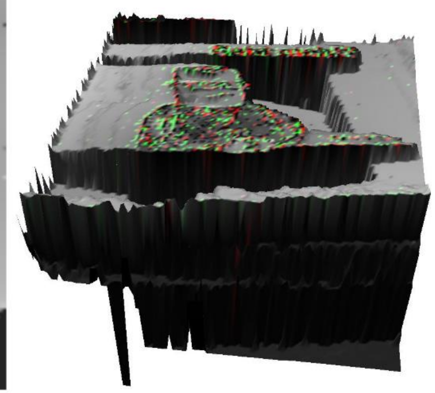
- Does not need to estimate motion
- Reconstruction is posed as **variational nonlinear image denoising**, using the time surface (event timestamps) to guide the denoising



Events

(*Data fidelity*)

Reconstructed image



Event manifold

(*Smoothness*)

$$\hat{I} = \arg\min_{f}(D(f, I_{di}) + R(f))$$

Image from direct integration of events

Reinbacher et al., *Real-Time Intensity-Image Reconstruction for Event Cameras Using Manifold Regularisation*, BMVC 2016 & IJCV 2018. YouTube

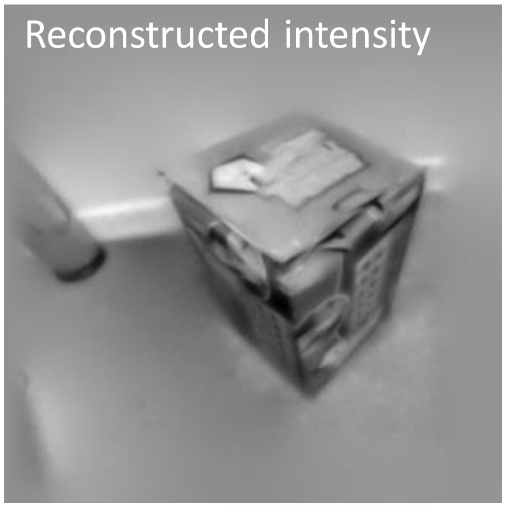
Reconstruction using “Manifold Regularisation”

- Does not need to estimate motion
- Reconstruction is posed as **variational nonlinear image denoising**, using the time surface (event timestamps) to guide the denoising

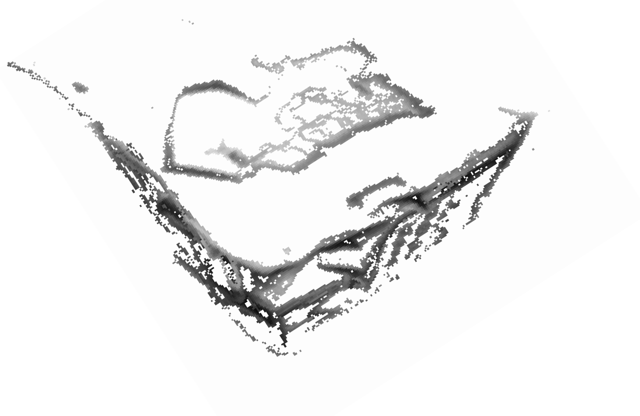
Reinbacher et al., *Real-Time Intensity-Image Reconstruction for Event Cameras Using Manifold Regularisation*, BMVC 2016 & IJCV 2018. YouTube

Event-based 6-DOF SLAM with 3 parallel filters

Parallel 6 DOF Tracking & Mapping in real time on a GPU



3D map of the scene



- **3 Kalman Filters** running in parallel.
  Each filter needs the output from the others.

- **Tracking**: EKF in 6 DOF pose
  - Motion model: constant position
  - Use contrast $h_{\mathbf{x}}(\mathbf{x}^{(t|t-\tau)}) = \mathrm{I}_l\left(\mathbf{p}_w^{(t)}\right) - \mathrm{I}_l\left(\mathbf{p}_w^{(t-\tau_c)}\right)$ to update pose (needs depth and intensity)

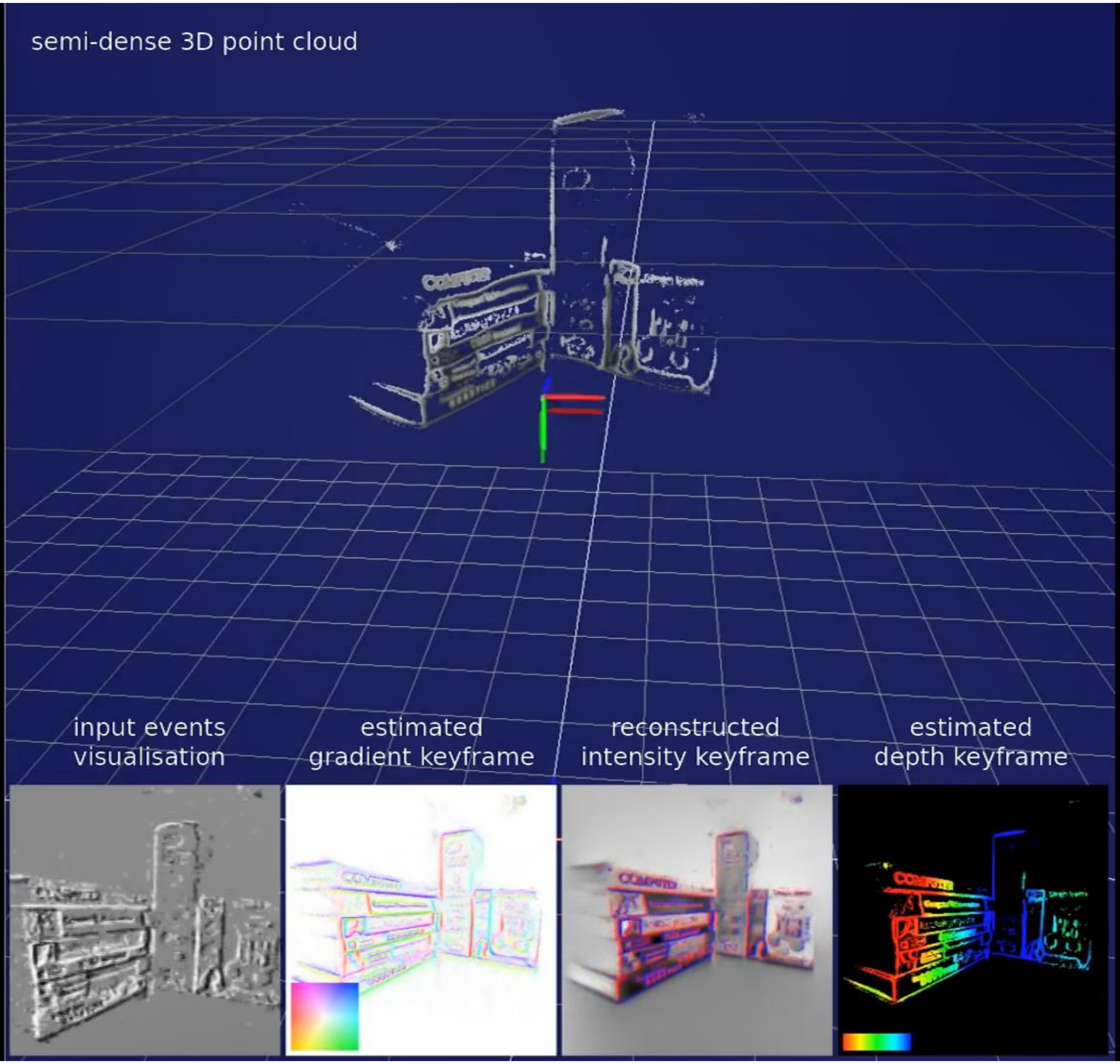
- **Intensity reconstruction**: pixel-wise EKF like Kim'14 & robust Poisson (Huber norm)

- **Mapping**: pixel-wise EKF on inverse depth, using contrast $h_{\boldsymbol{\rho}} = \mathrm{I}_l\left(\mathbf{p}_w^{(t)}\right) - \mathrm{I}_l\left(\mathbf{p}_w^{(t-\tau_c)}\right)$

Kim et al. *Real-Time 3D Reconstruction and 6-DoF Tracking with an Event Camera*, ECCV 2016

Event-based 6-DOF SLAM with 3 parallel filters

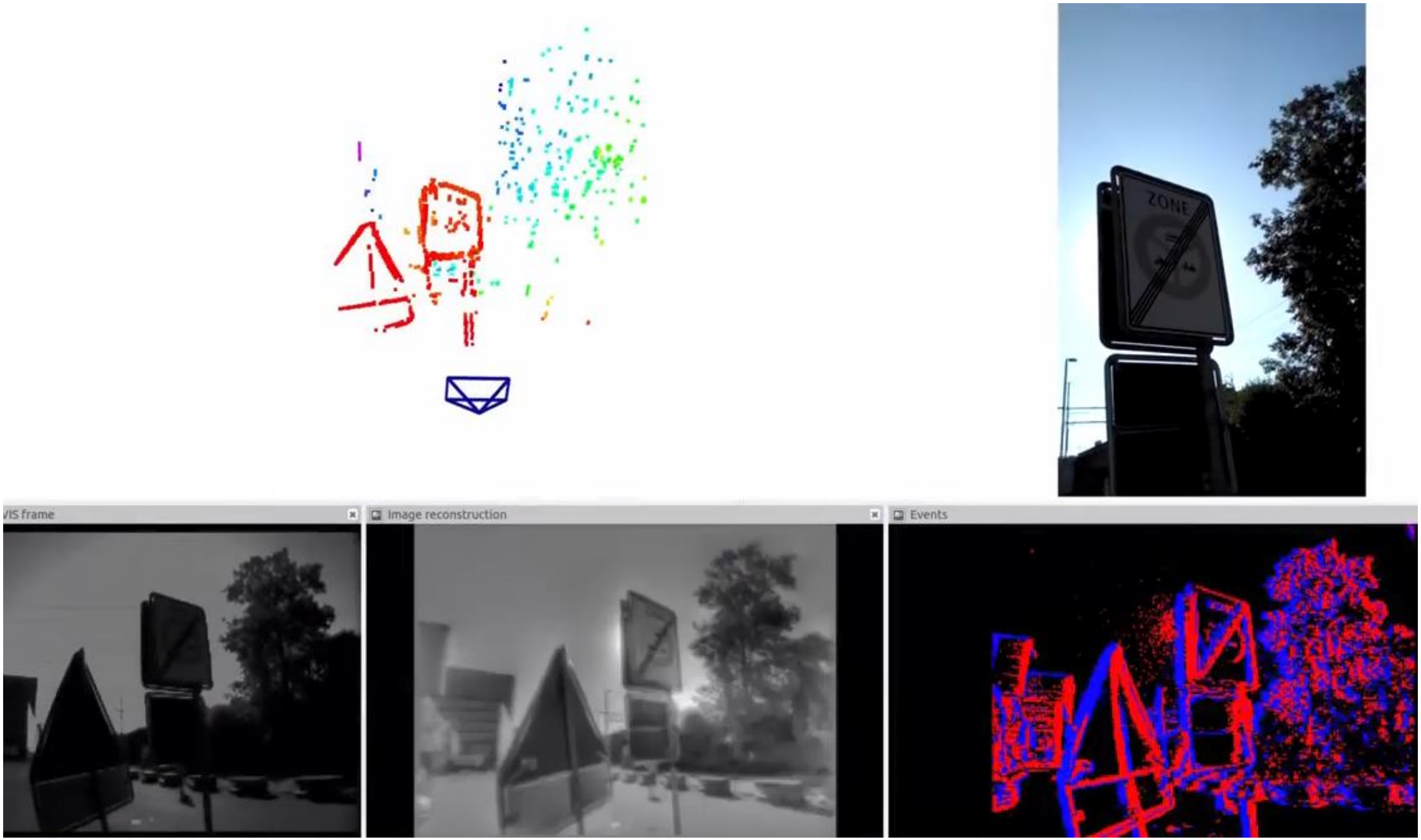
Parallel 6 DOF Tracking & Mapping in real time on a GPU



Kim et al. *[Real-Time 3D Reconstruction and 6-DoF Tracking with an Event Camera](#)*, ECCV 2016. https://youtu.be/yHLyhdMSw7w

Event-based 6-DOF SLAM on a CPU

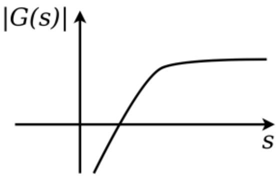
HDR image reconstruction from the output of SLAM



Rebecq et al., *EVO: A Geometric Approach to Event-based 6-DOF Parallel Tracking and Mapping in Real-time*, RAL'17. https://youtu.be/bYqD2qZJlxE

Reconstruction by Temporal Filtering

- Replace **pixel-wise** direct integration with a **high-pass temporal filter** to remove accumulated event noise
- No spatial filtering needed



Direct integration of events

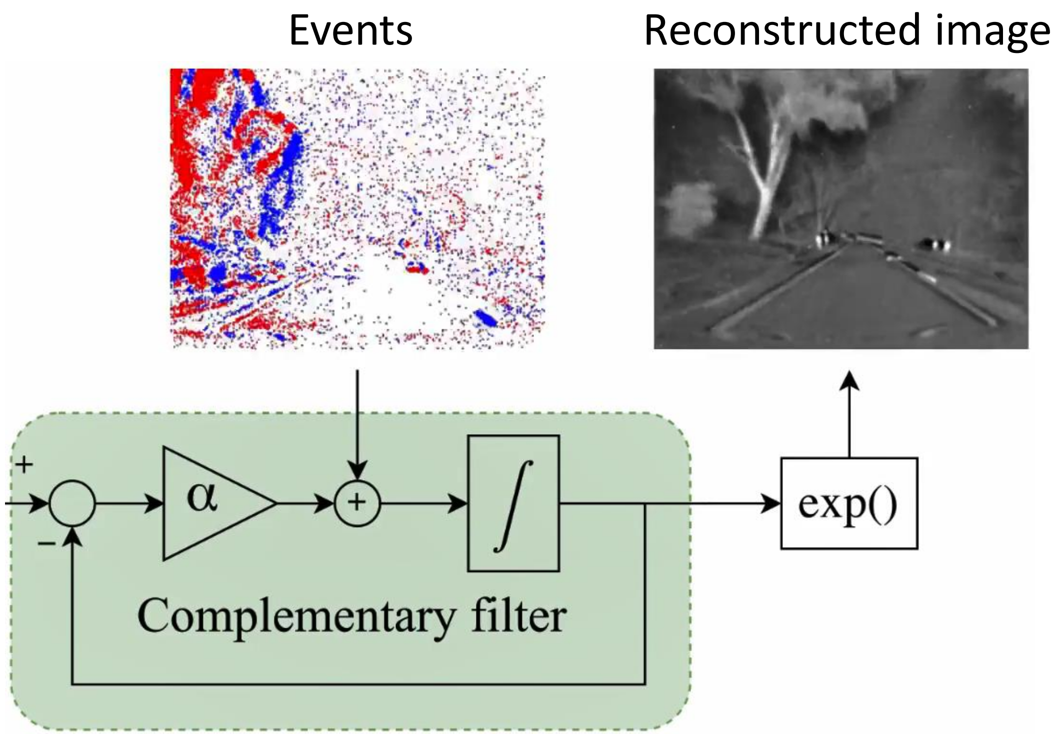
High-pass filtered (in time)

But slow-time information (low freq.) (static background) is lost.
⇒ fuse with grayscale frames from DAVIS (complementary filter)

Scheerlinck et al., *Continuous-time Intensity Estimation Using Event Cameras*, ACCV 2018. https://youtu.be/bZ0ZKido0Ag

Reconstruction by Temporal Filtering

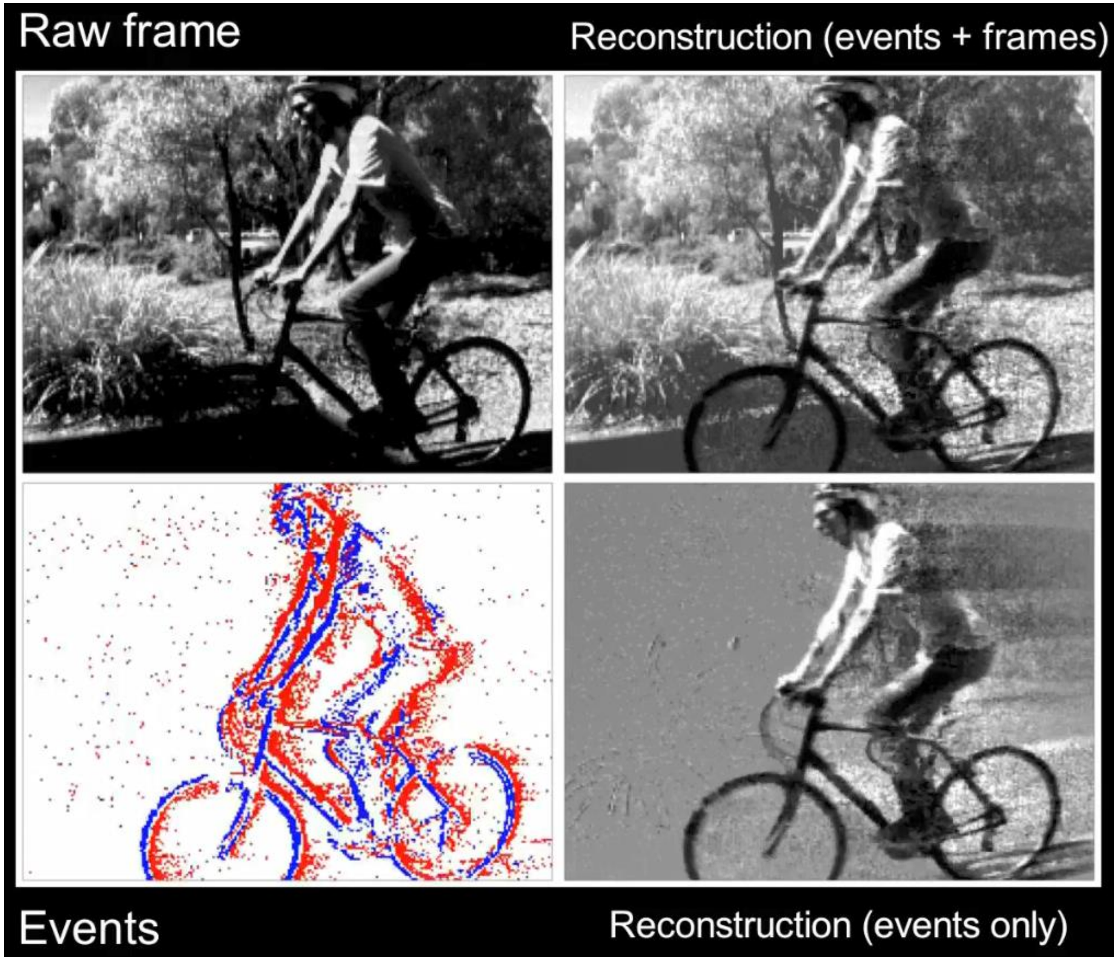
- Replace **pixel-wise** direct integration with a **high-pass filter** to remove accumulated event noise
- The internal **state of the filter** is an **image** that is updated **asynchronously, per-pixel** with each incoming event



Scheerlinck et al., *Continuous-time Intensity Estimation Using Event Cameras*, ACCV 2018. https://youtu.be/bZ0ZKido0Ag

Reconstruction by Temporal Filtering

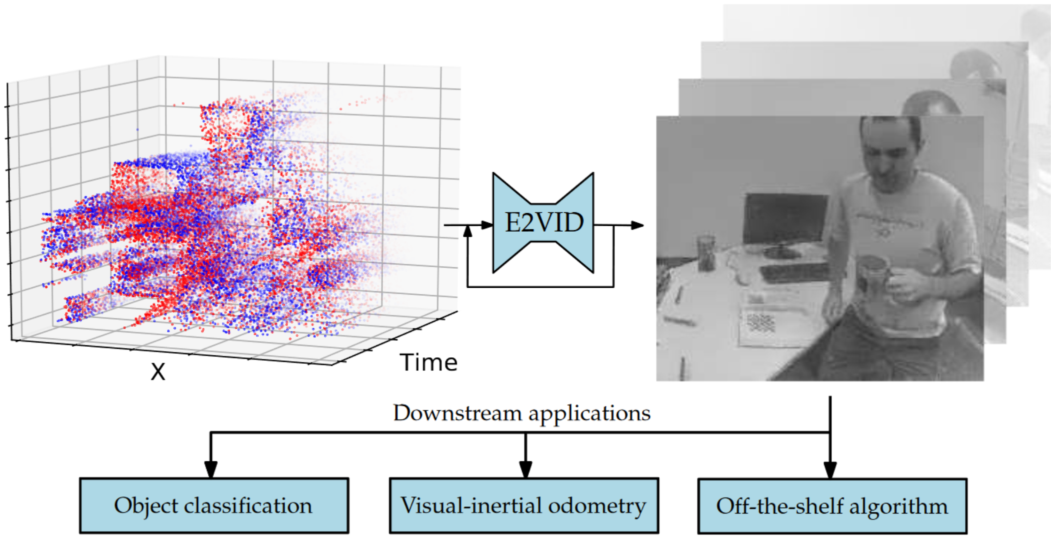
- Replace **pixel-wise** direct integration with a **high-pass filter** to remove accumulated event noise



Scheerlinck et al., *Continuous-time Intensity Estimation Using Event Cameras*, ACCV 2018. https://youtu.be/bZ0ZKido0Ag

Reconstruction using Deep-Learning

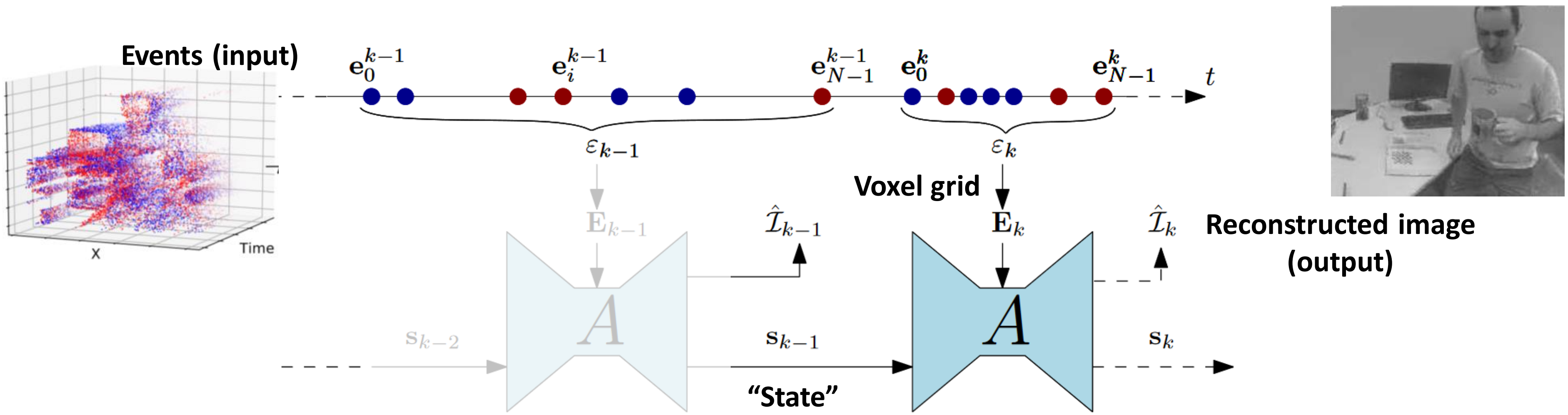
- **Events-to-Video (E2VID)**
  - Deep Learning method: **recurrent network** (with a U-Net)
  - **Loss function: perceptual** (LPIPS) + temporal consistency
  - Trained on simulation, transfers well to real-world data
  - Shows a **big improvement** with respect to previous methods
  - Shows reconstructed images can be used on **off-the-shelf computer vision methods** designed for image data



Rebecq et al., *High Speed and High Dynamic Range Video with an Event Camera*, PAMI 2020. YouTube

Reconstruction using Deep-Learning

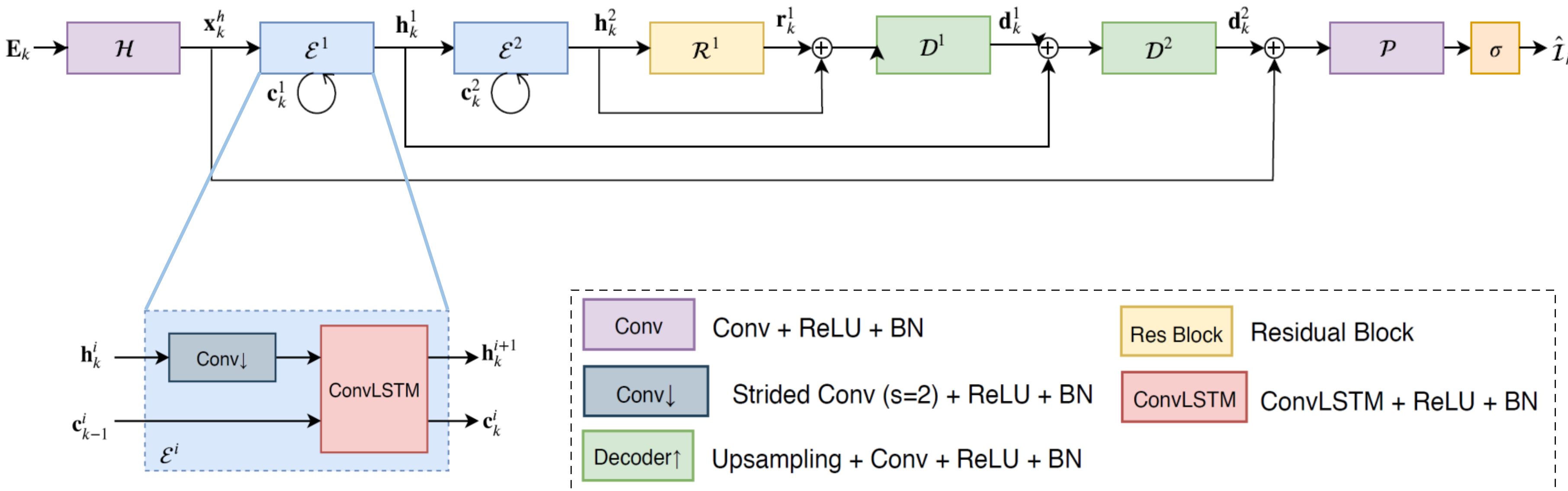
- **Events-to-Video (E2VID)**
  - Deep Learning method: **recurrent network** (with a U-Net)
  - **Loss function: perceptual** (LPIPS) + temporal consistency
  - Trained on simulation, transfers well to real-world data
  - Shows a **big improvement** with respect to previous methods
  - Shows reconstructed images can be used on **off-the-shelf computer vision methods** designed for image data



Rebecq et al., *High Speed and High Dynamic Range Video with an Event Camera*, PAMI 2020. YouTube

Network architecture

- Recurrent U-Net architecture



Rebecq et al., *High Speed and High Dynamic Range Video with an Event Camera*, PAMI 2020. YouTube
Ronneberger et al., *U-Net: Convolutional Networks for Biomedical Image Segmentation*, MICCAI'15

E2VID - Results

Huawei P20 Pro (240 FPS)

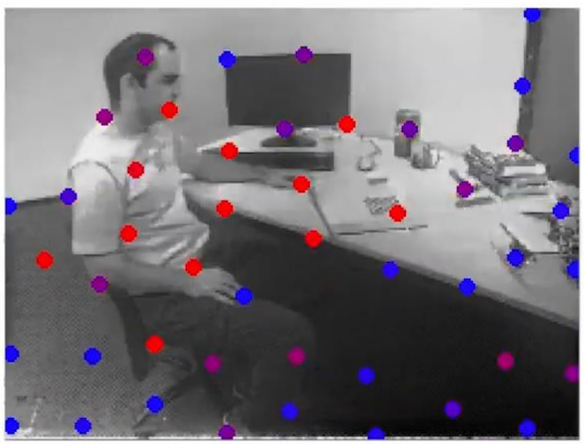
Our reconstruction (5400 FPS)

100 x slow motion

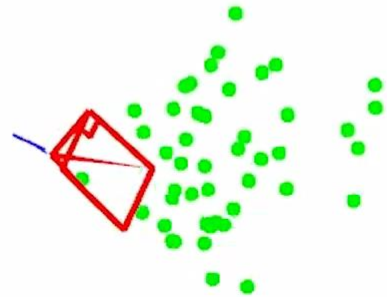
Rebecq et al., *[High Speed and High Dynamic Range Video with an Event Camera]()*, PAMI 2020. [YouTube]()

E2VID - Applications of Reconstructed images

Events



Our reconstruction
+ tracked features

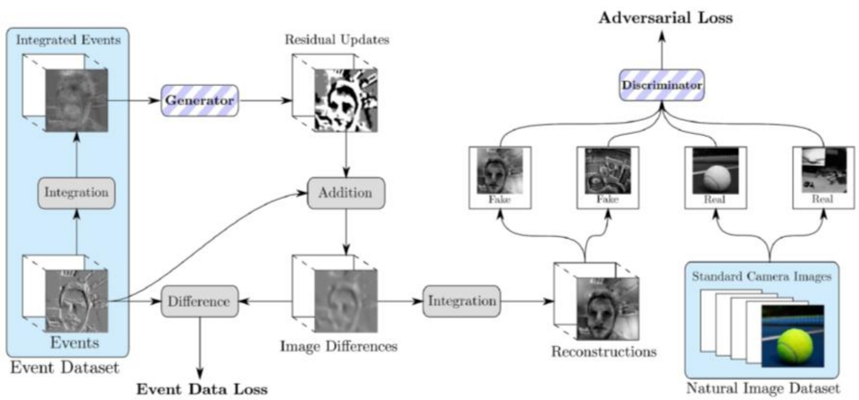


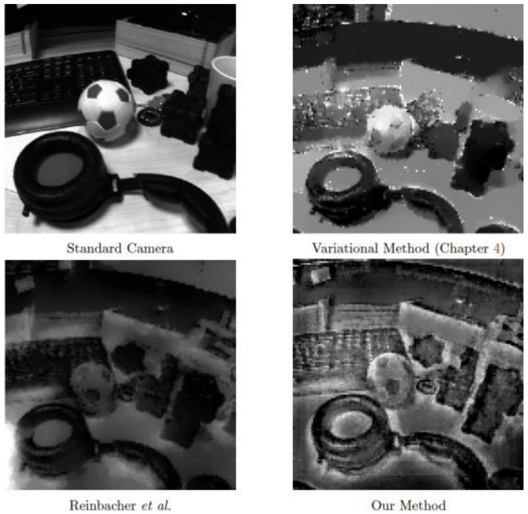
VINS-Mono running on our reconstruction
from events

Rebecq et al., *High Speed and High Dynamic Range Video with an Event Camera*, PAMI 2020. YouTube

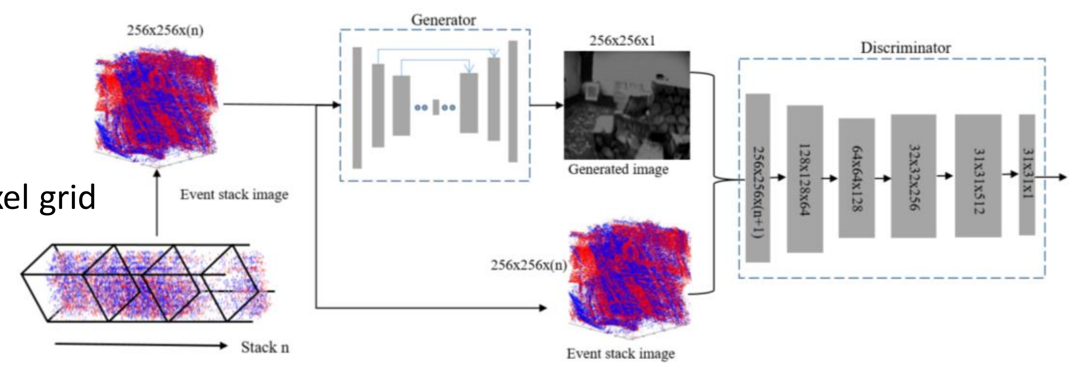
Unsupervised Deep Learning, using GANs

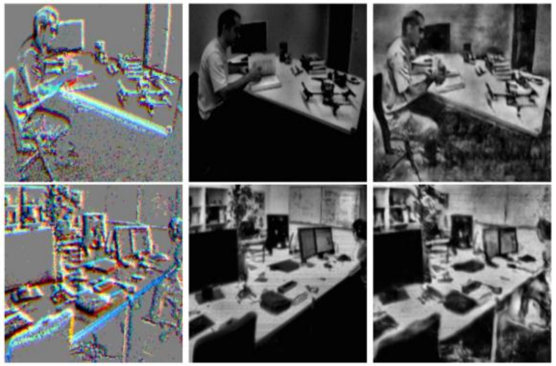
- Bardow (Ch. 6): Reconstruction using only Natural Image Priors





- Mostafavi CVPR 2019





Mostafavi et al., *Event-based High Dynamic Range Image and Very High Frame Rate Video Generation using Conditional GANs*, CVPR 2019
Bardow, P., *Estimating General Motion and Intensity from Event Cameras*, Ph.D. Thesis 2019