Object Motion Segmentation: Advantages from Event Data

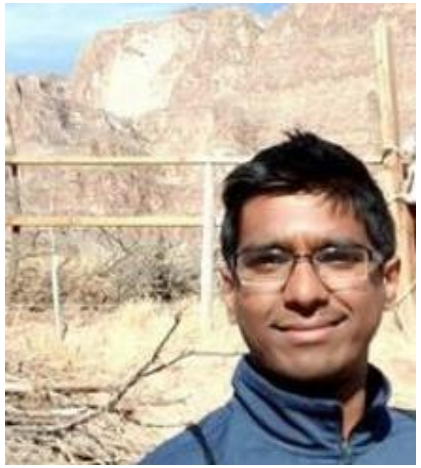
Cornelia Fermüller

Computer Vision Laboratory, UMIACS

University of Maryland

Collaborators

Anton Mitrokhin

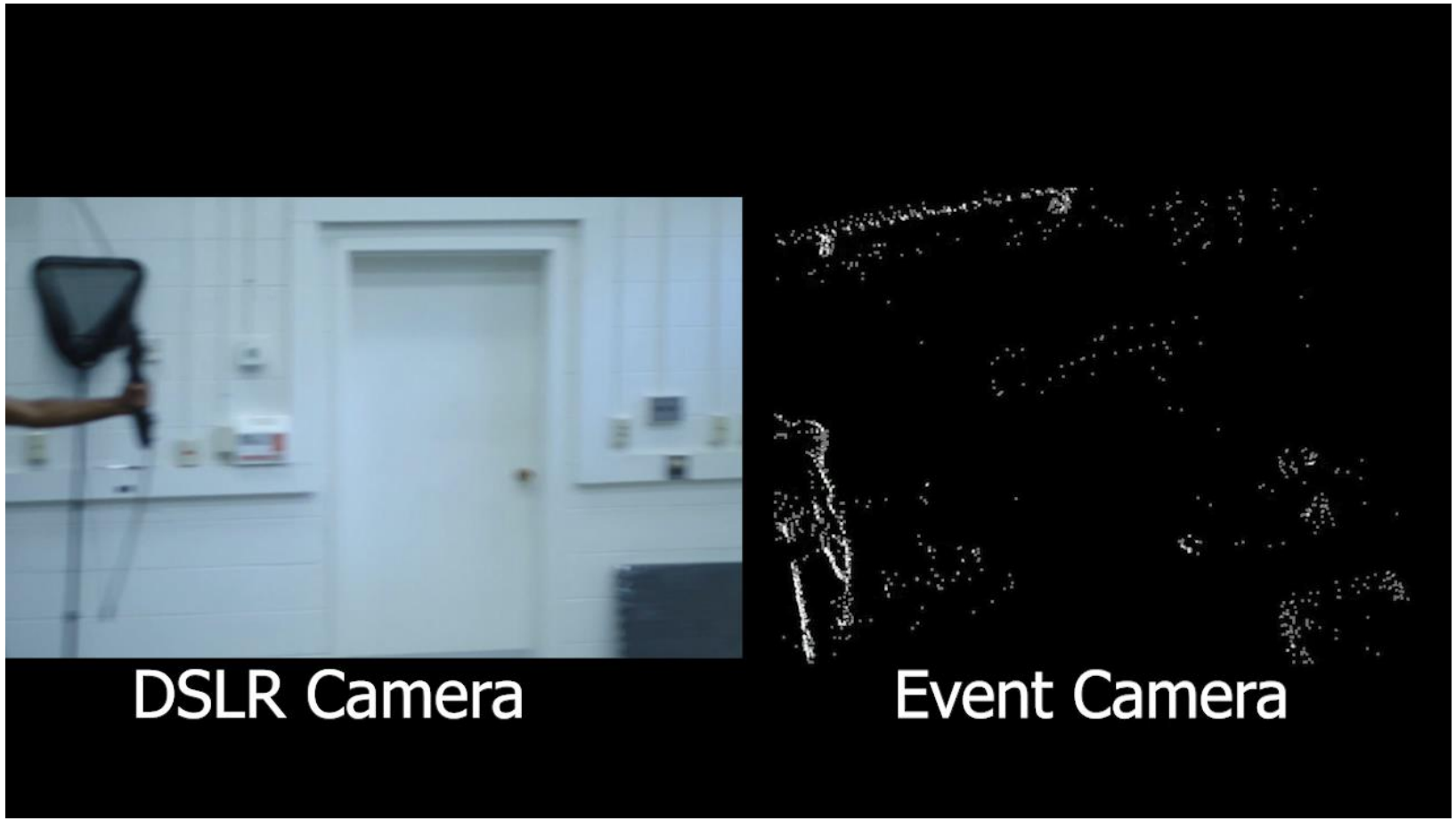


Chethan Parameshwara

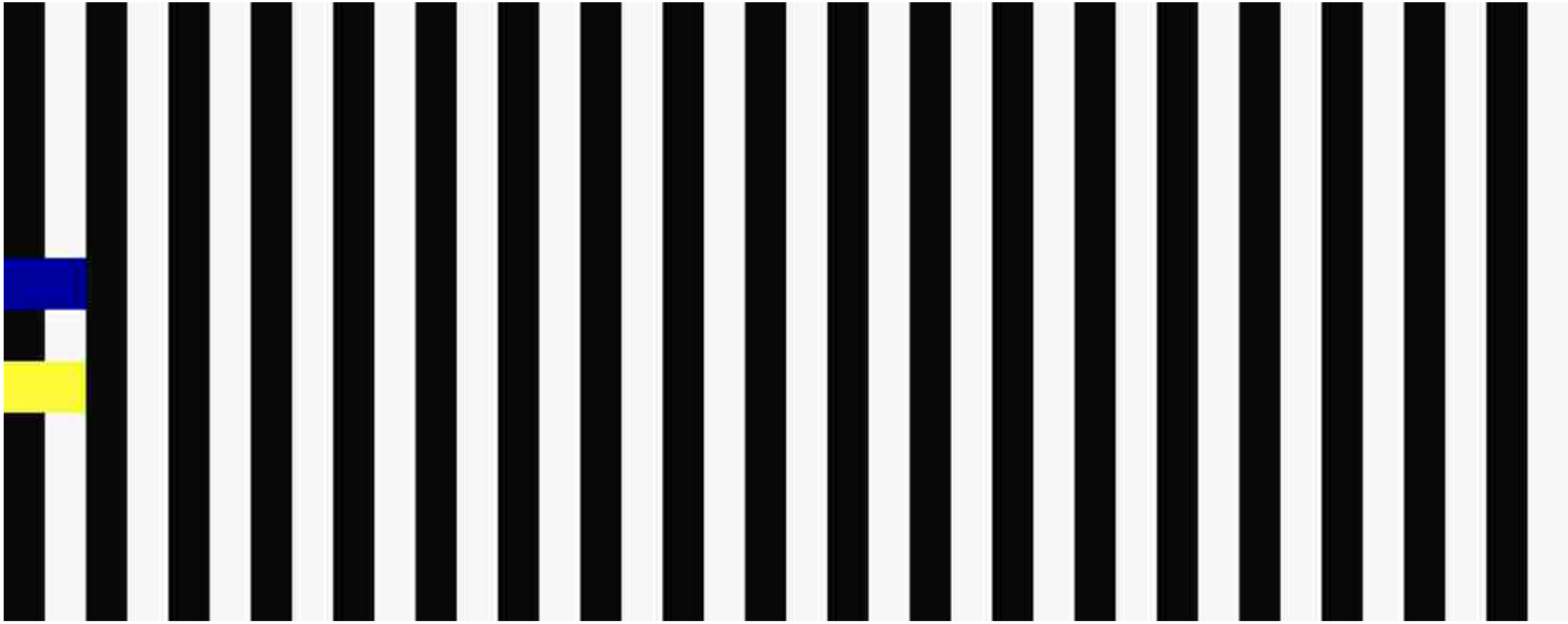
ChengXi Ye

Yiannis Aloimonos

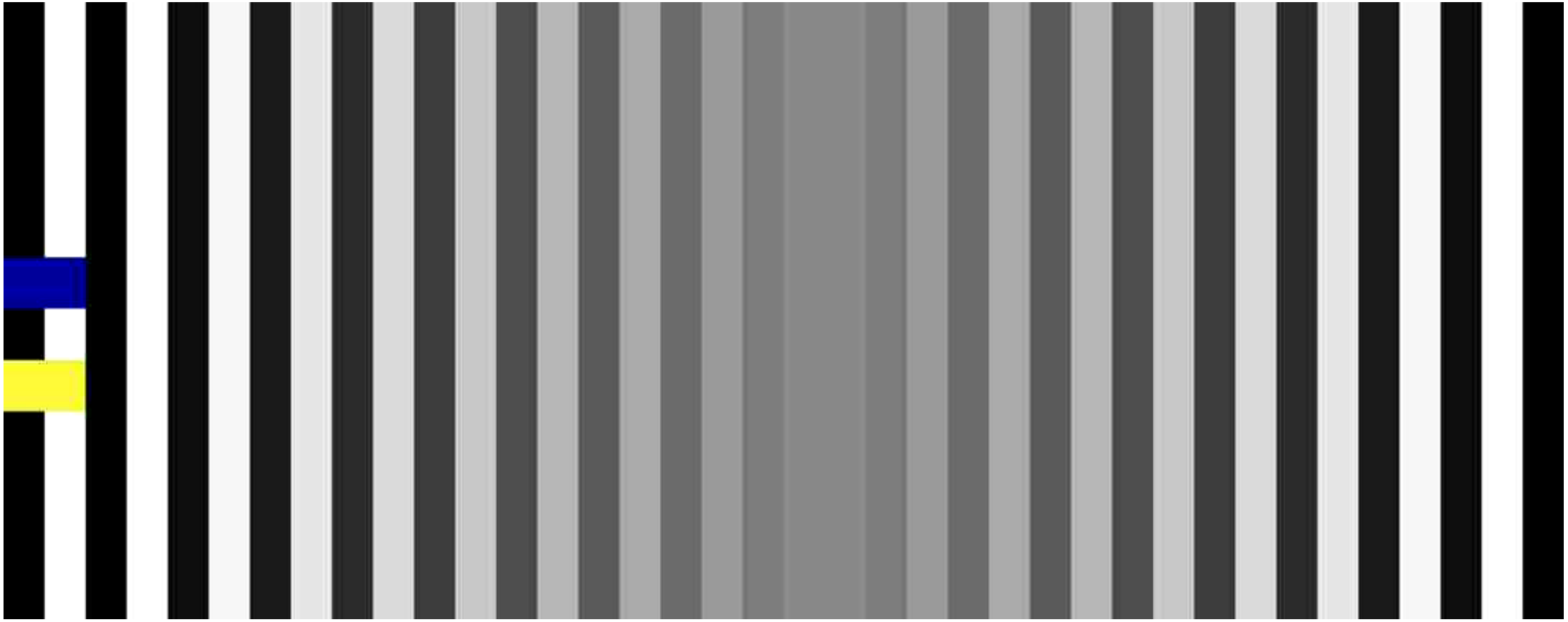
Fast events aid in segmentation



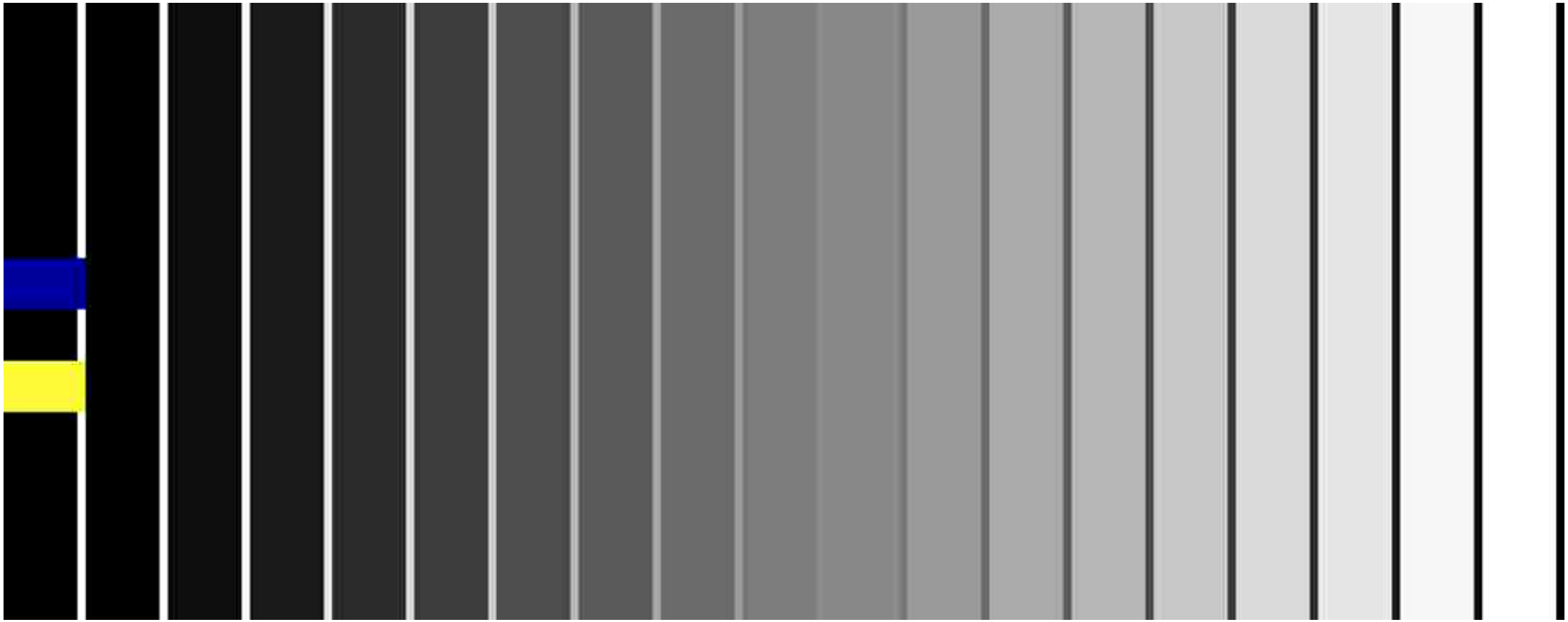
Stepping Feet Illusion



Variation of Stepping Feet Illusion



Variation of Stepping Feet Illusion

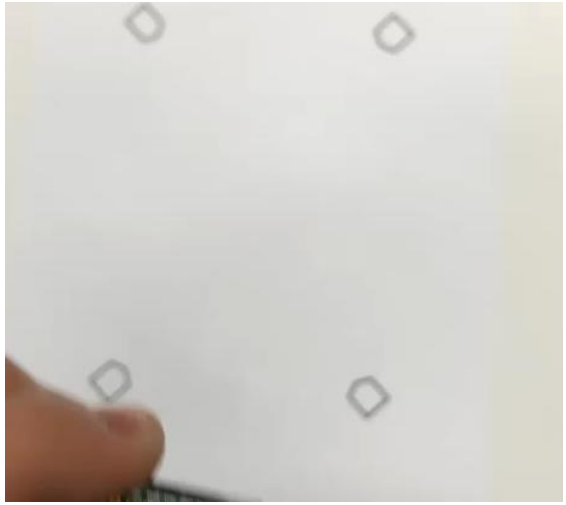


Overview

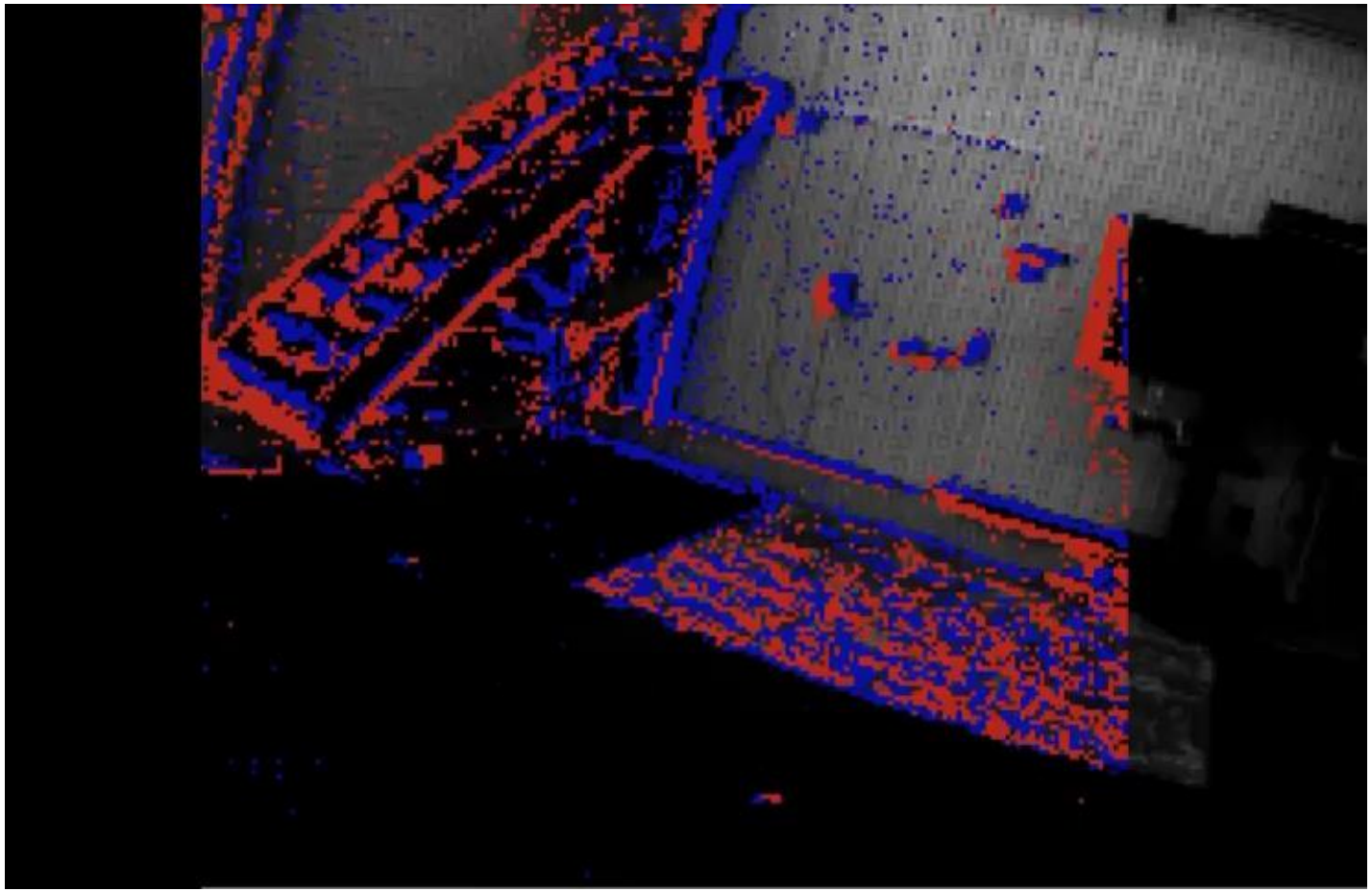
I. Optimization approach for event alignment
II. Self-supervised deep learning for SLAM
III. Supervised/Unsupervised deep learning for motion segmentation and a new dataset
IV. EVDodge: Motion detection as input to control dodging

Properties of this sensor

+ High temporal resolution
+ High dynamic range
+ Low Bandwidth signal
+ Low latency
- High noise

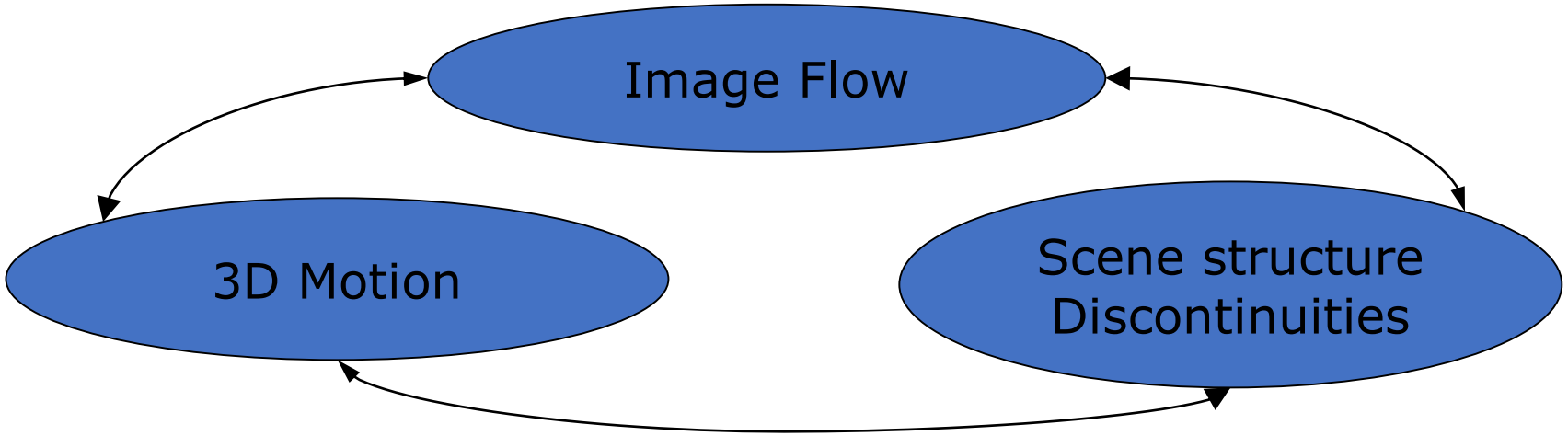


I. Egomotion+ Independent Motion

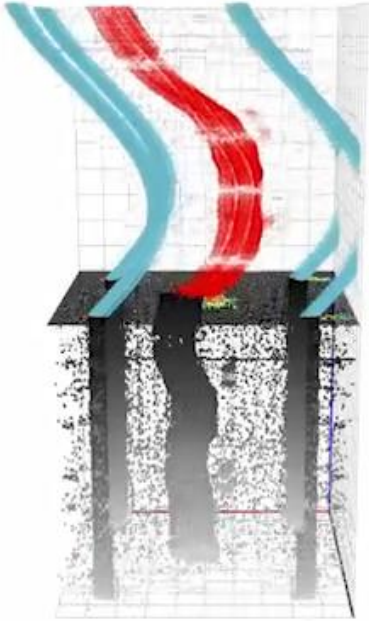


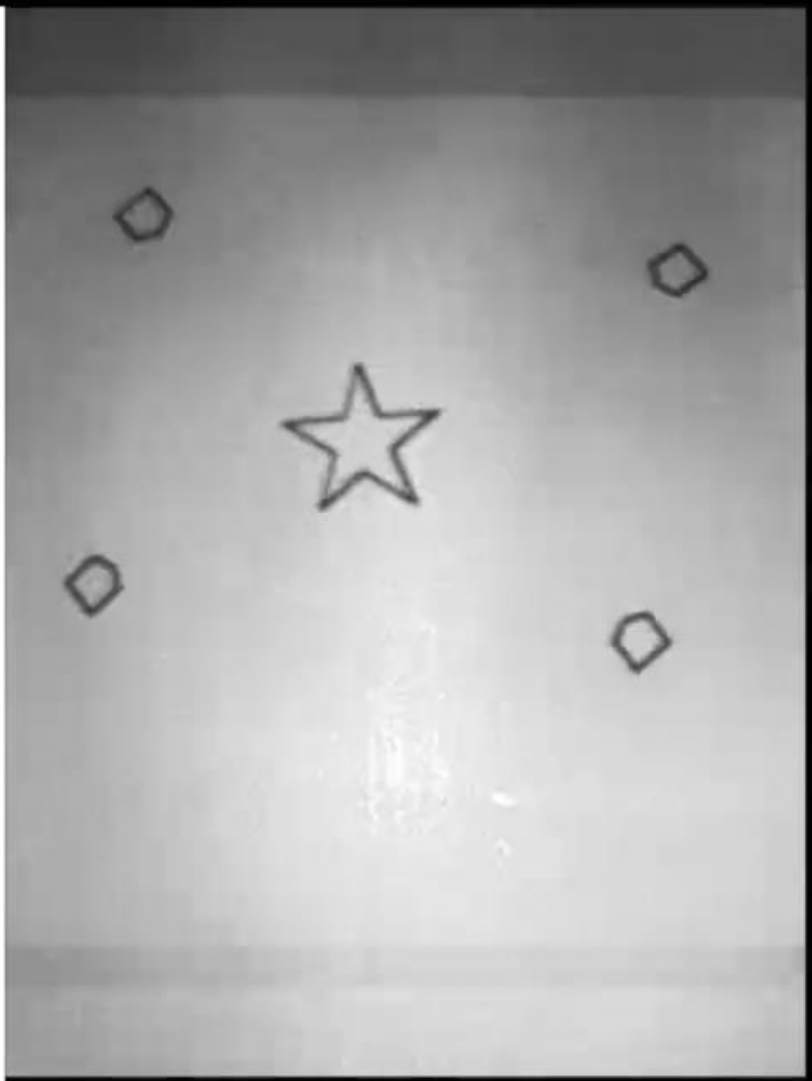
What is the problem?

- All the components are related.

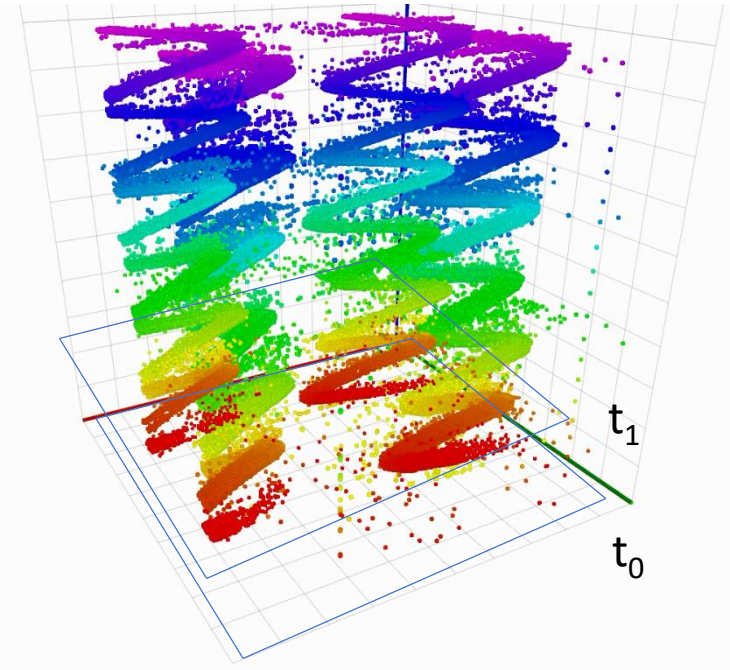


*





Treat events as point clouds



Warp field $\Phi(d_x, d_y, d_r, d_\phi) : (x, y, t) \rightarrow (x + u\Delta t, y + v\Delta t, t)$

d_x, d_y Shift in x and y

d_r, d_θ Radial expansion, and rotation around z-axis
Derived from divergence and curl

Approximation of 3D Motion Estimation

$$\boldsymbol{u}(\boldsymbol{x},\boldsymbol{y})\Delta t = \begin{pmatrix} u_0 \\ v_0 \end{pmatrix} + \left\{ \frac{1}{2} curl_g \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} + \frac{1}{2} div_g \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \right\} \begin{pmatrix} x \\ y \end{pmatrix} \Delta t$$

d_x, d_y $\quad$ d_θ $\quad$ d_z

Approximates rigid movement of fronto-parallel plane

How to compute it?

- **Density** (from Event Count image)

$$\xi_{ij} = \{\{x', y', t\} : \{x', y', 0\} \in C',\ i = x',\ j = y'\}$$

$$D = \frac{\sum i,j|\xi_{i,j}|}{\#I}$$

Sum of events over all pixels / Number of occupied pixels

- **Average time (f**rom Time Stamp image)

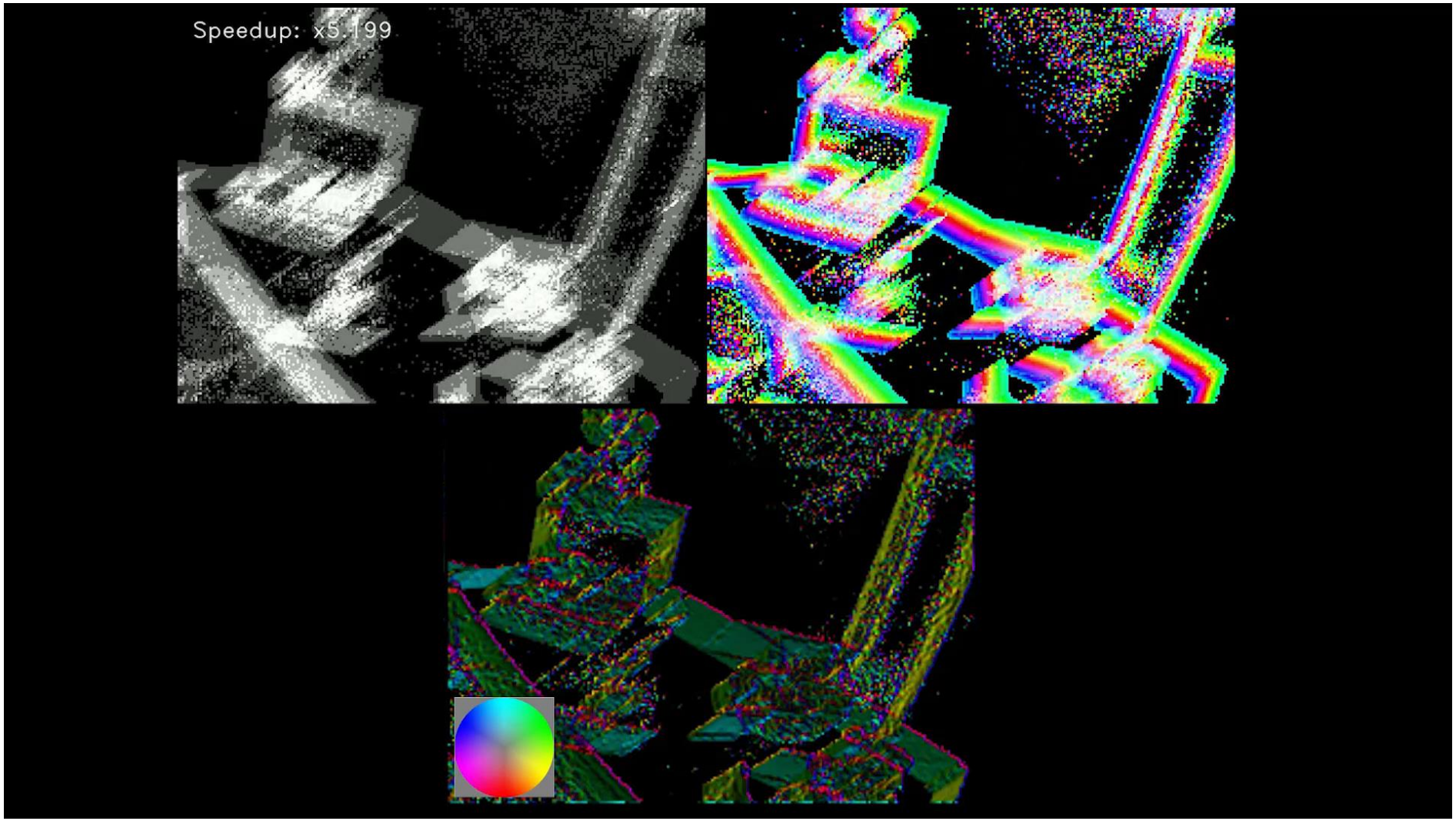
$$\mathcal{T}_{ij} = \frac{1}{\mathcal{I}_{ij}} \sum t : t \in \xi_{ij}$$

Solve minimization efficiently via gradients on time stamp images

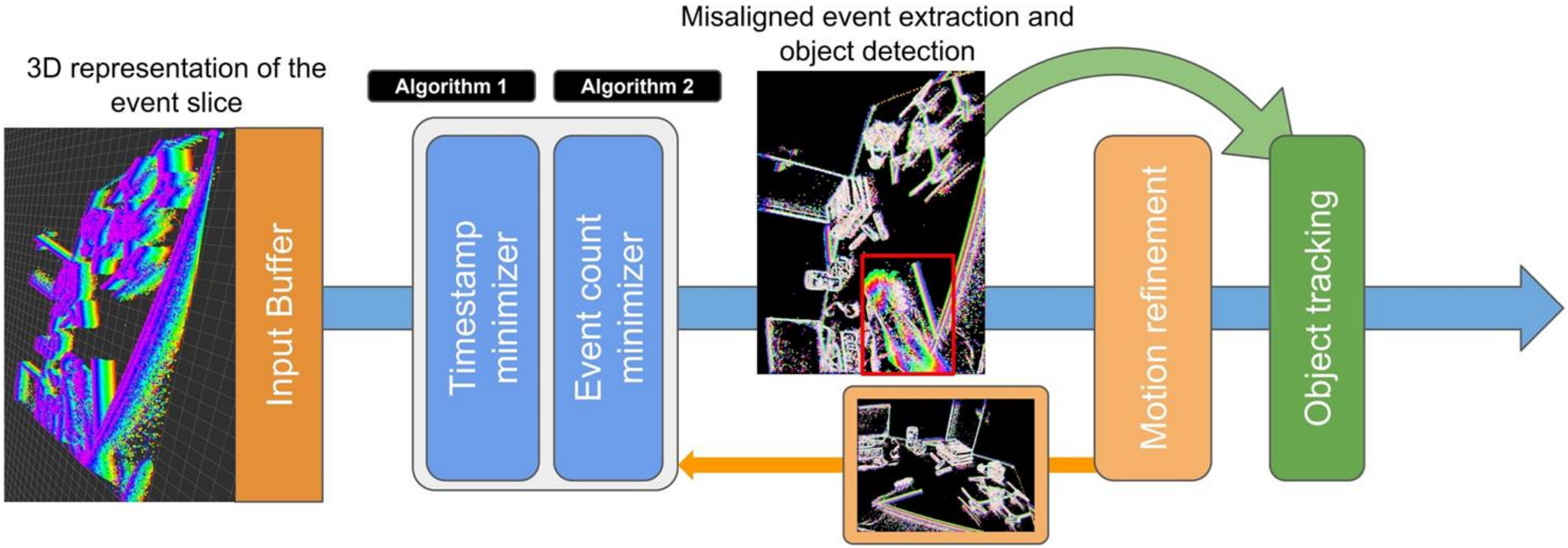
$$d_x = \frac{\sum G_x[i,j]}{\#\mathcal{I}}, \qquad d_y = \frac{\sum G_y[i,j]}{\#\mathcal{I}}$$

$$d_z = \frac{\sum (G_x[i,j], G_y[i,j]) \cdot (i,j)}{\#\mathcal{I}}$$

$$d_\theta = \frac{\sum (G_x[i,j], G_y[i,j]) \times (i,j)}{\#\mathcal{I}}$$

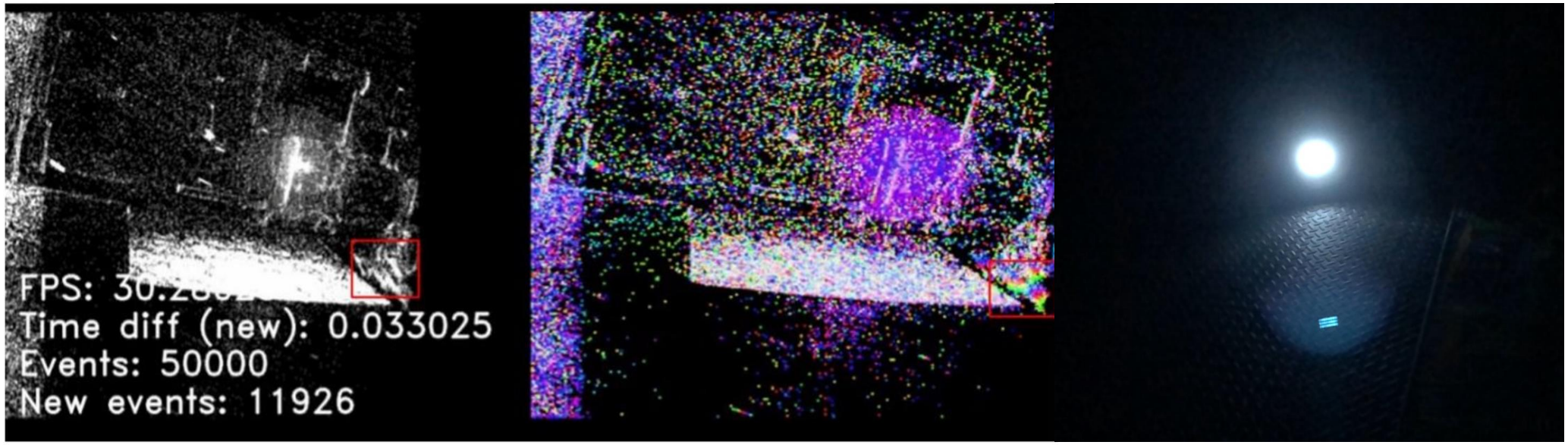

Speedup: x5.199

Algorithm



A Mitrokhin, C Fermüller, C Parameshwara, Y Aloimonos. Event-based moving object detection and tracking, IROS 2018.

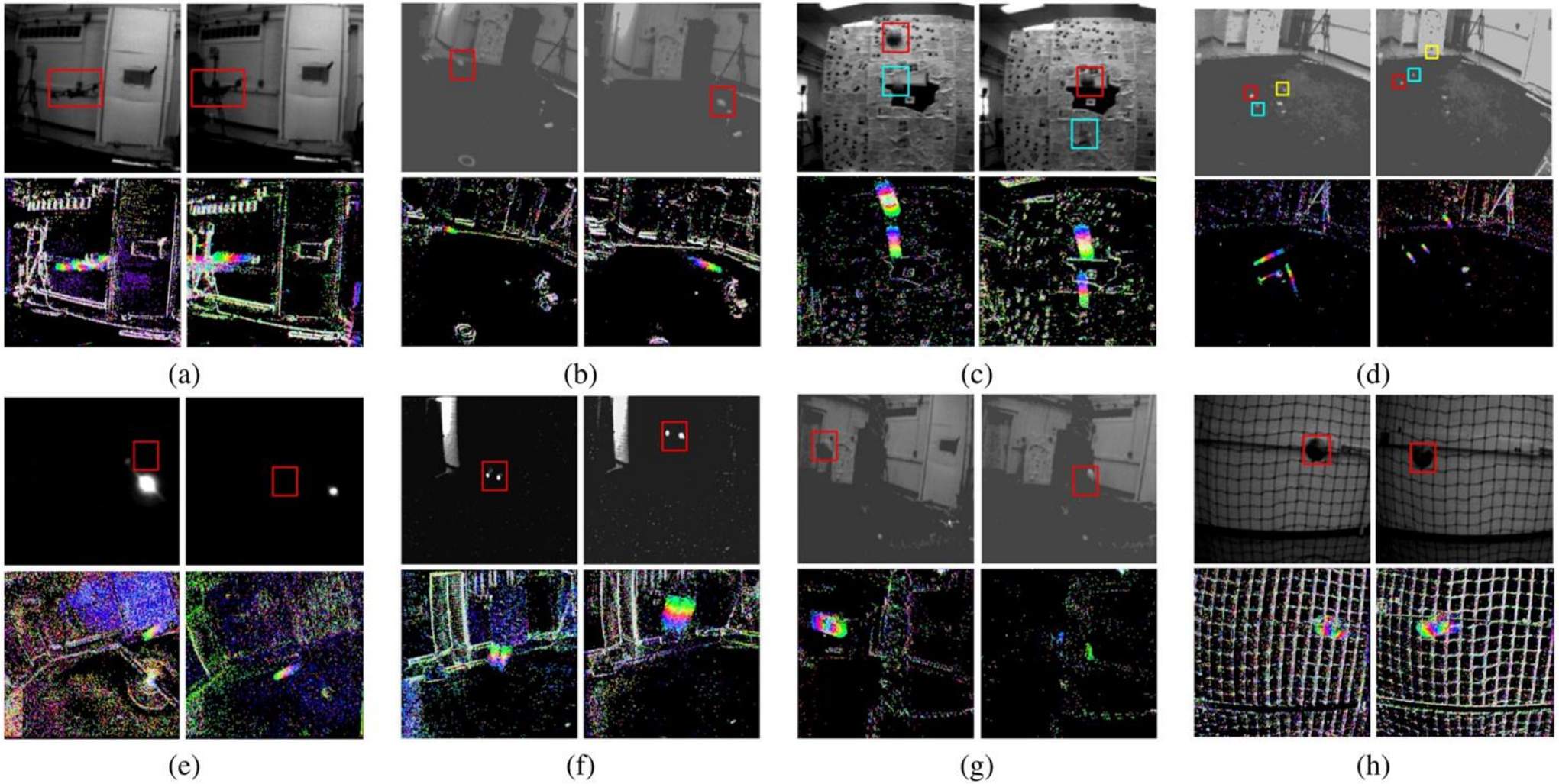
Dataset



Drone used in the dataset collection. *1 - mounted DAVIS240B camera, 2 - Customized Qualcomm Flight platform.*

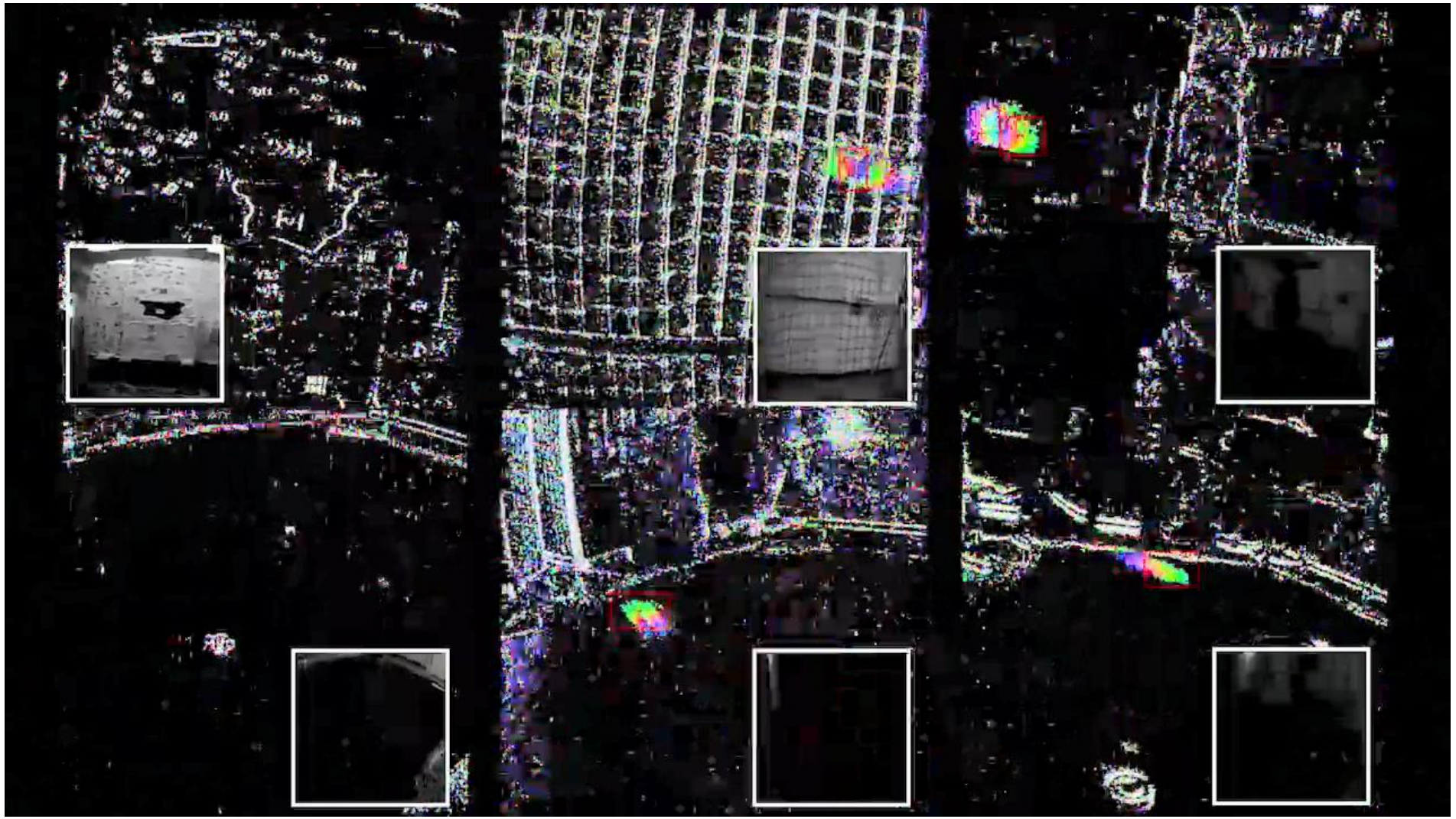
The overall weight of the fully loaded platform is ≈ 500g. and it is equipped with the Snapdragon APQ8074 ARM CPU, featuring 4 cores with up to 2.3GHz frequencies.

Dataset



Fast motion, Multiple Objects, Lighting Variations, Occlusion

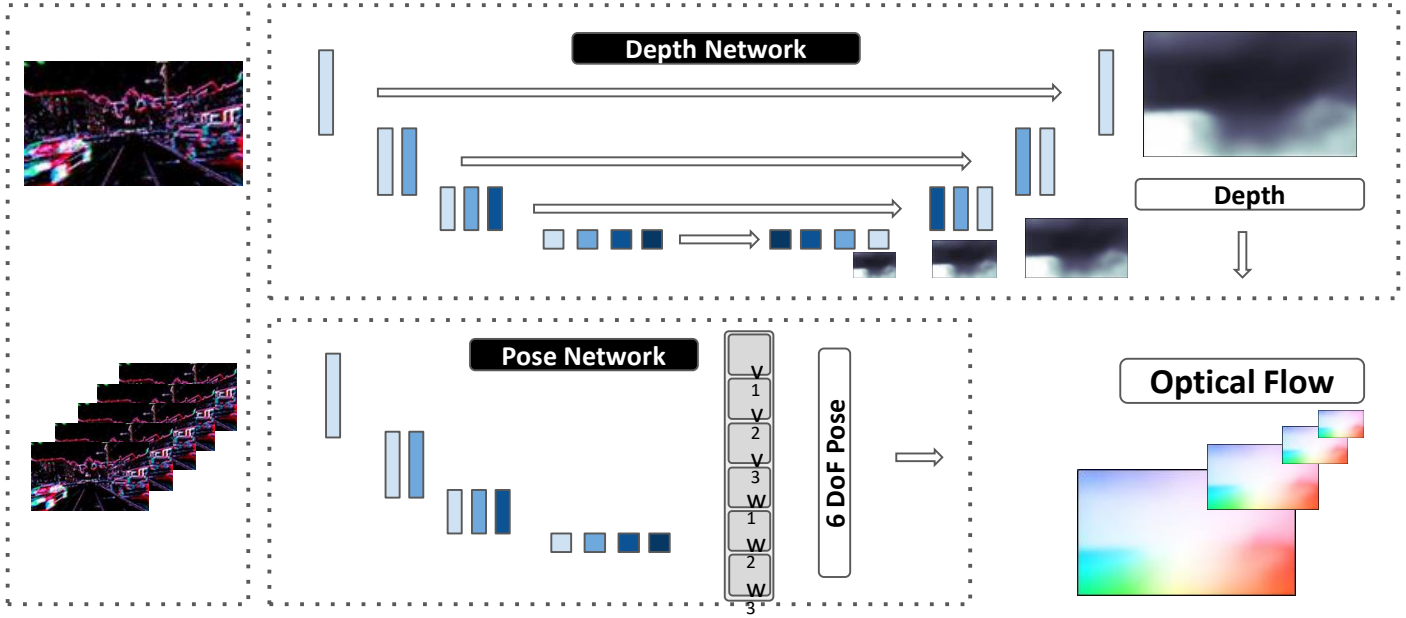
Results



II. Flow Depth and 3D Motion Estimation

$$\begin{pmatrix} u \\ v \end{pmatrix} = \frac{1}{Z} \begin{pmatrix} -1 & 0 & x \\ 0 & -1 & y \end{pmatrix} \begin{pmatrix} v_x \\ v_y \\ v_z \end{pmatrix} + \begin{pmatrix} xy & -1-x^2 & y \\ 1+y^2 & -xy & -x \end{pmatrix} \begin{pmatrix} \omega_x \\ \omega_y \\ \omega_z \end{pmatrix}$$

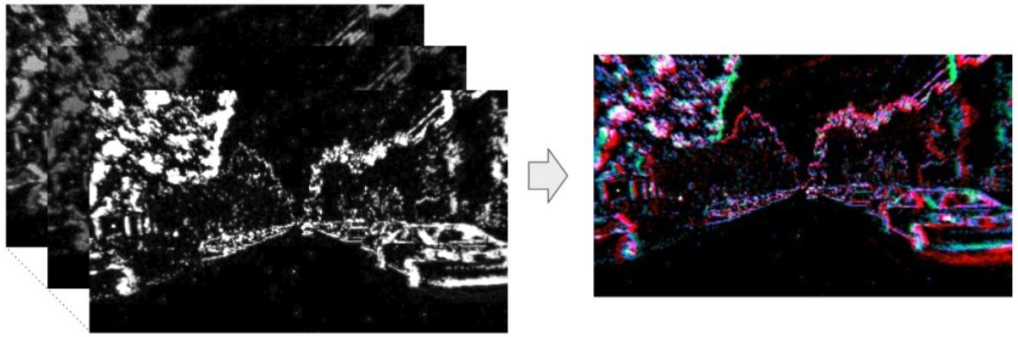
flow depth translation 3D motion rotation

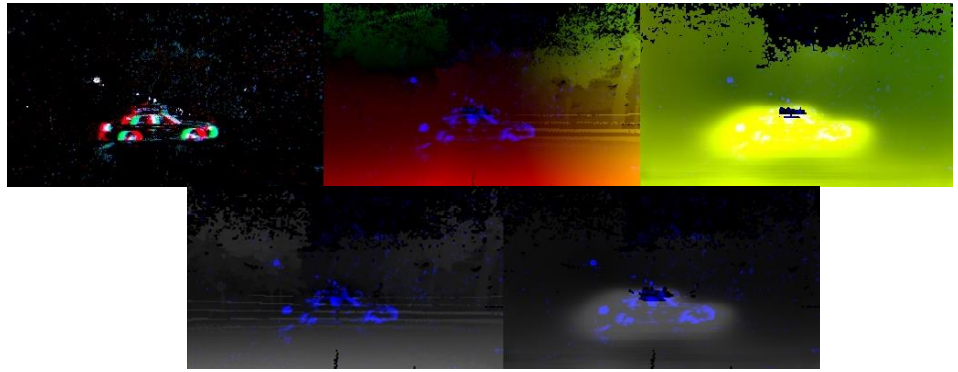


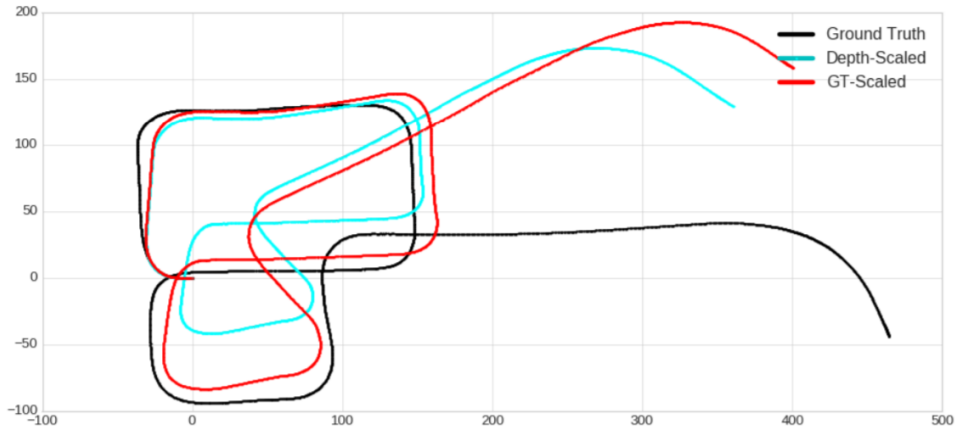
Ye, C., Mitrokhin, A., Fermüller, C., Yorke, J. A., and Aloimonos Y. "Unsupervised Learning of Dense Optical Flow and Depth from Sparse Event Data." arXiv preprint arXiv:1809.08625 (2018).

II. Highlights

Unsupervised Learning of Dense Optical Flow, Depth and Egomotion from Sparse Event Data





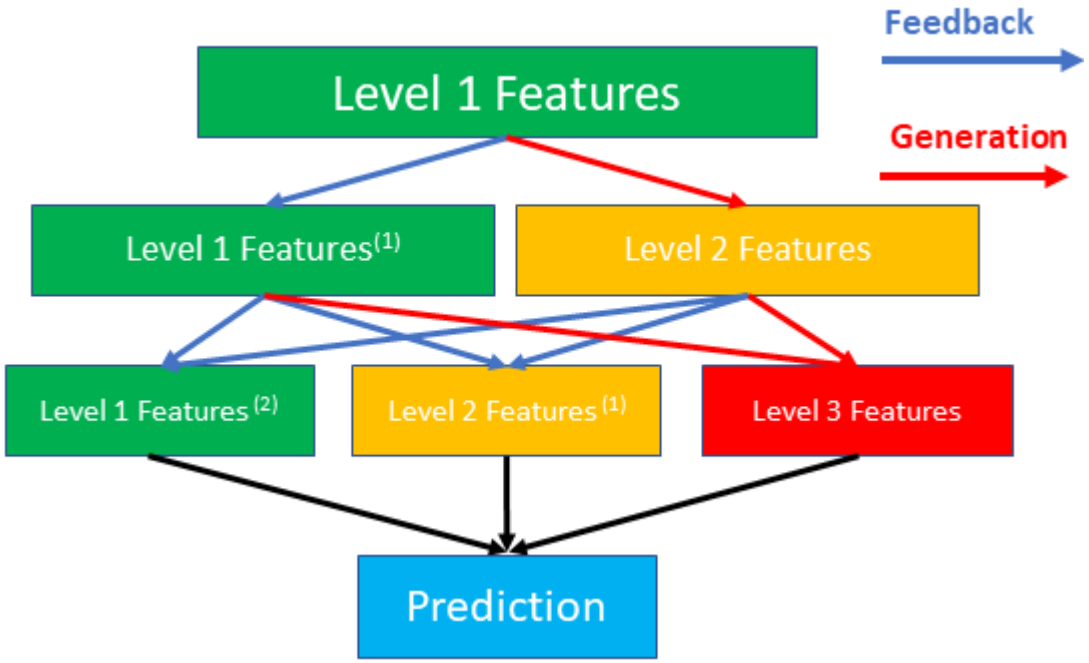


- Transfers from day to night!
- Fixes data sparsity
- Good results

http://prg.cs.umd.edu/ECN.html

II. Highlights

- A new light-weight architecture ECN



II. Highlights

- Sparsity constraint that promotes non-local information propagation

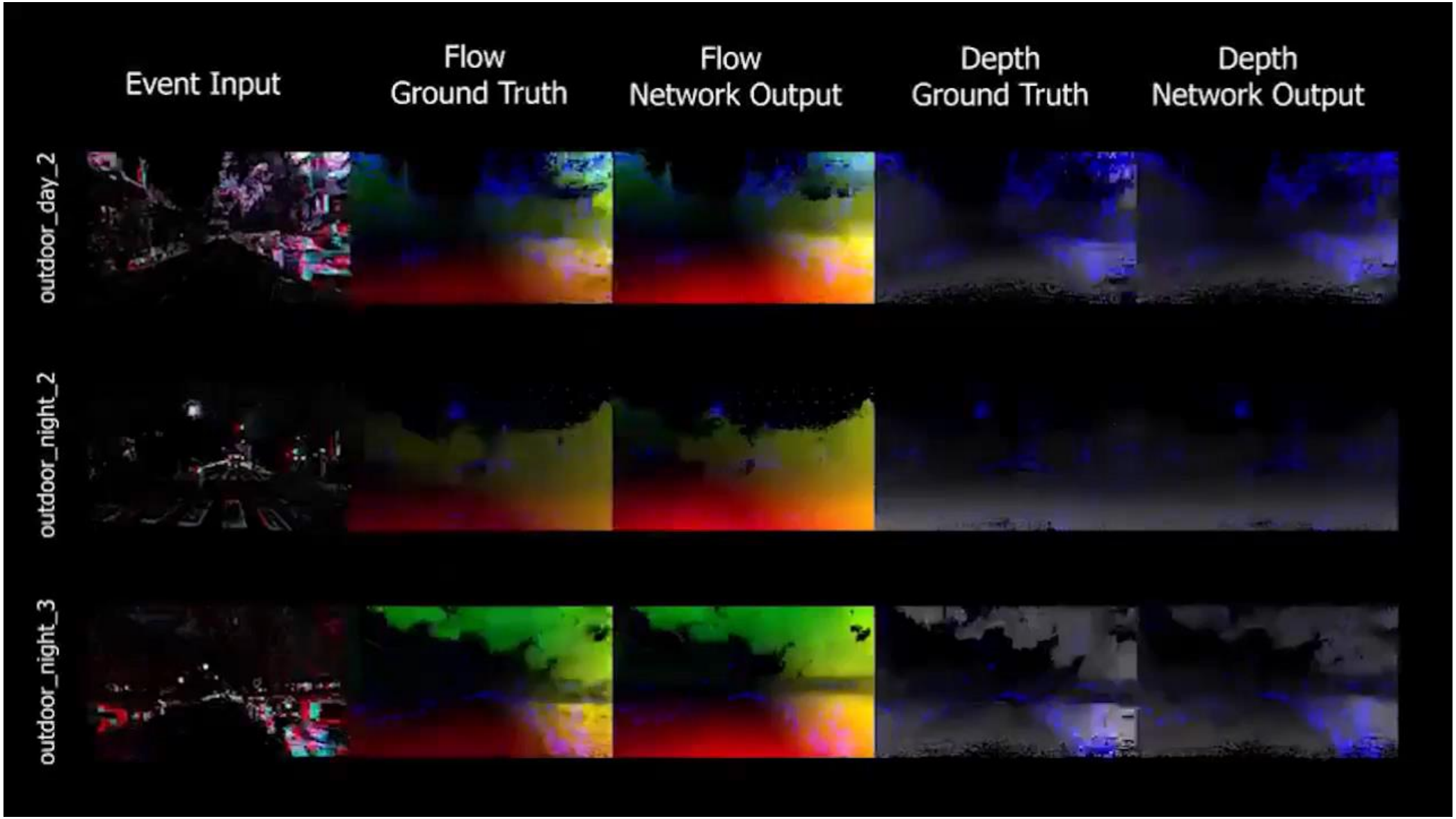
$$Loss_{smooth}(I) = \sum_i \sum_{j \in N(i)} |I_j - I_i|^p$$

With $0 < p < 1$

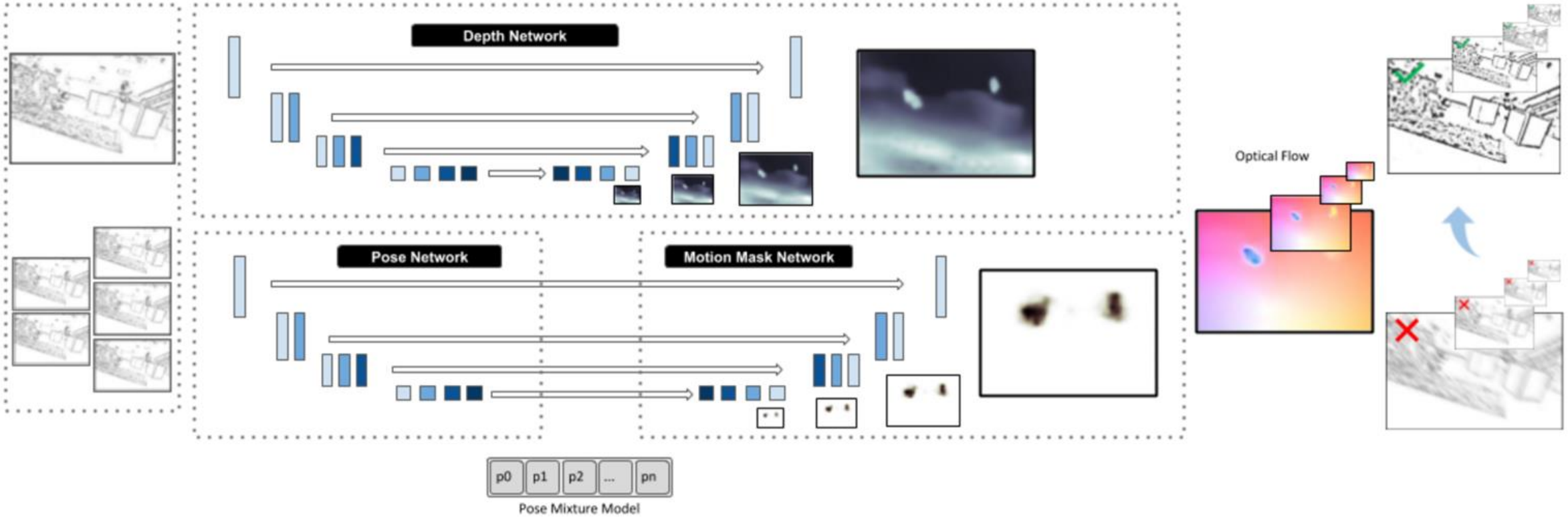
$$= \sum_i \sum_{j \in N(i)} |I_j - I_i|^{p-2} |I_j - I_i|^2 = \sum_i \sum_{j \in N(i)} w_{ij} |I_j - I_i|^2$$

$$Loss = Loss_{warp} + \lambda Loss_{smooth}$$

Loss applied to first order derivative of depth estimate

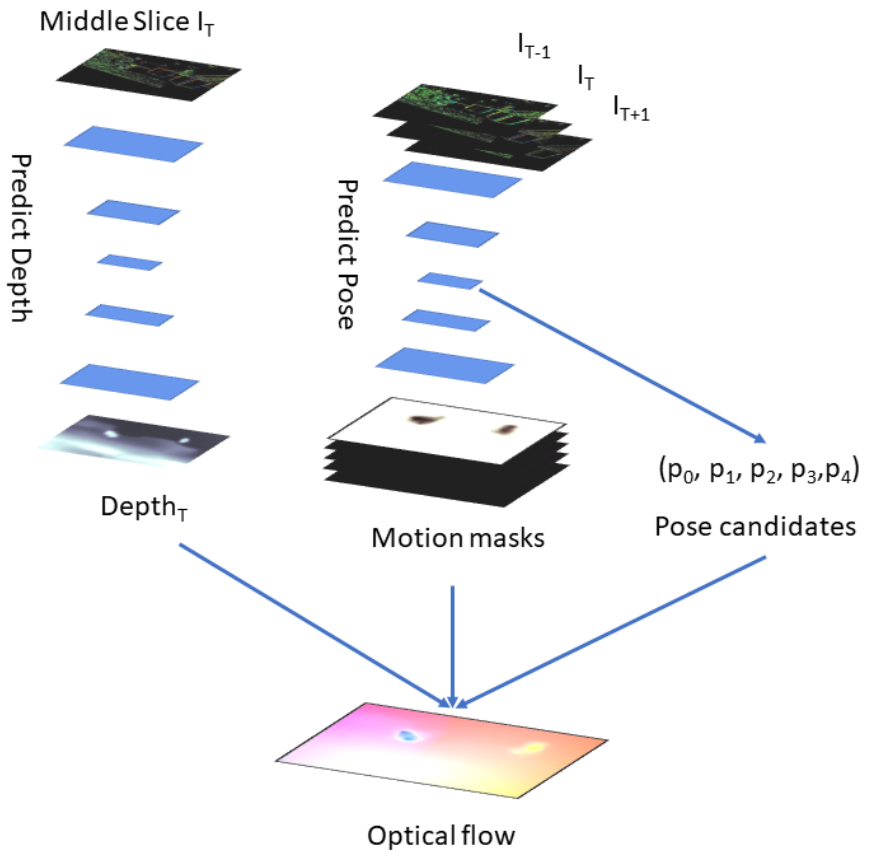

Event Input
Flow Ground Truth
Flow Network Output
Depth Ground Truth
Depth Network Output
outdoor_day_2
outdoor_night_2
outdoor_night_3

III:EV-IMO: Motion Segmentation Dataset and Learning Pipeline for Event Cameras

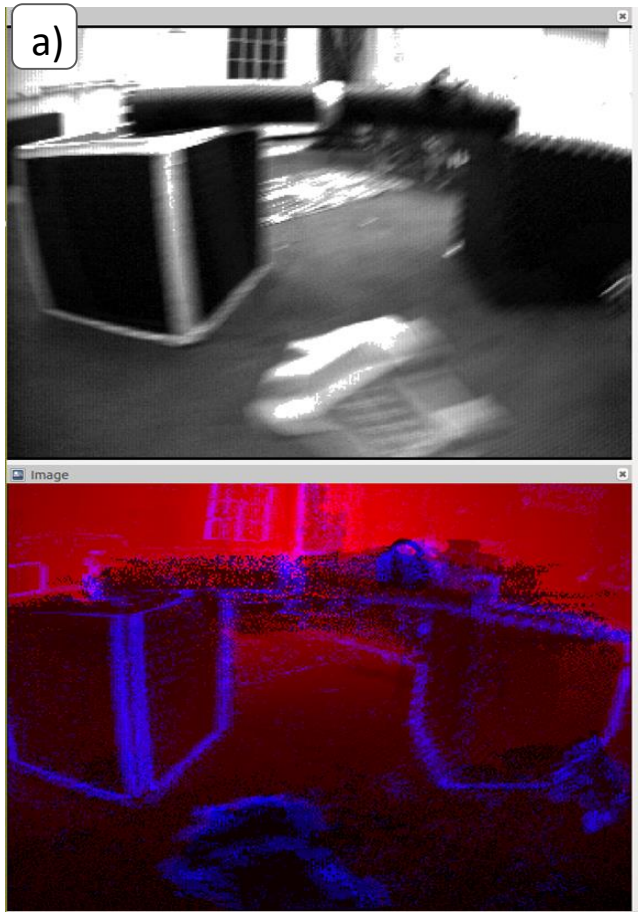


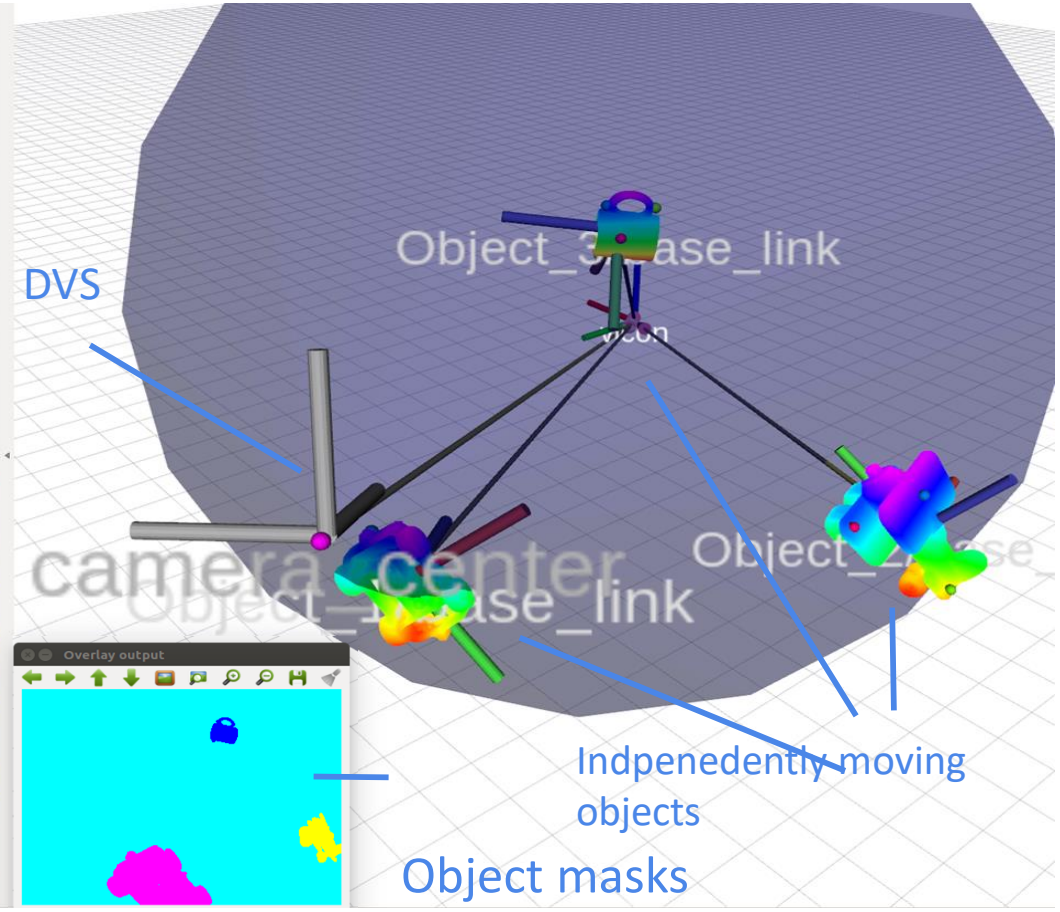
Ye, C., Mitrokhin, A., Fermüller, C., Aloimonos Y and Delbruck, T. "EV-IMO: Motion Segmentation Dataset and Learning Pipeline for Event Cameras." arXiv preprint arXiv:1903.07520 (2019).

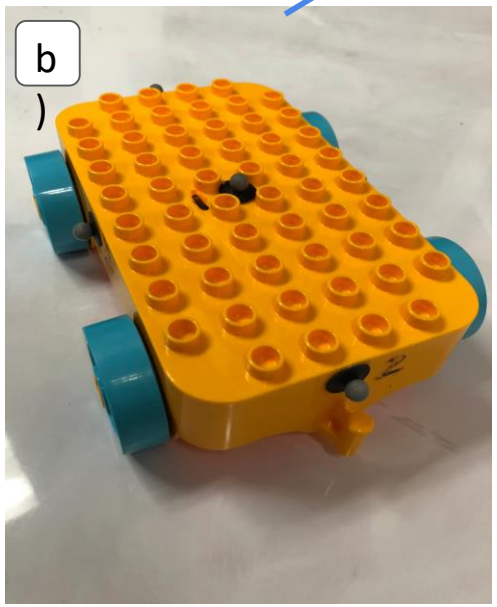
Using motion masks to learn a pose mixture model

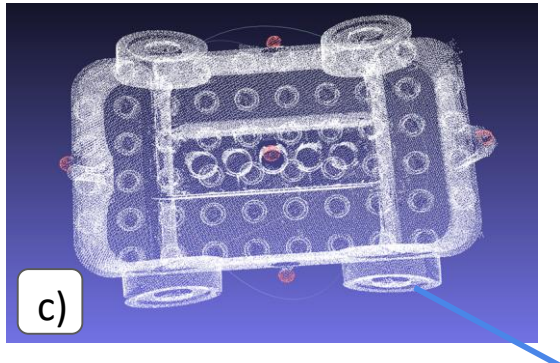


Our Dataset: EV-IMO







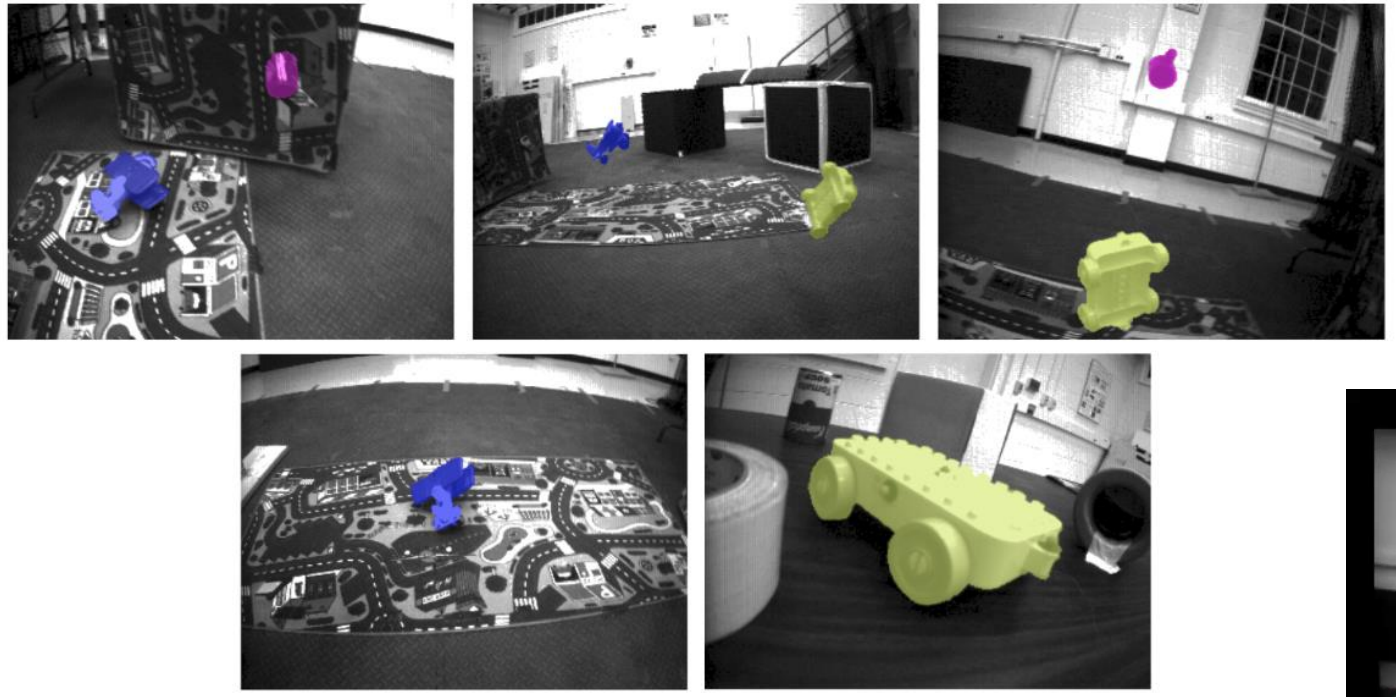


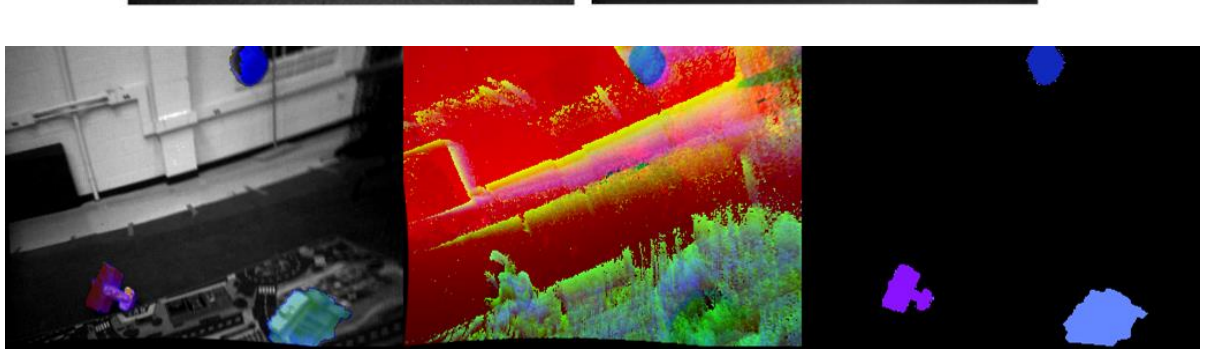
Example object

Depth from static room scan

Scan of object

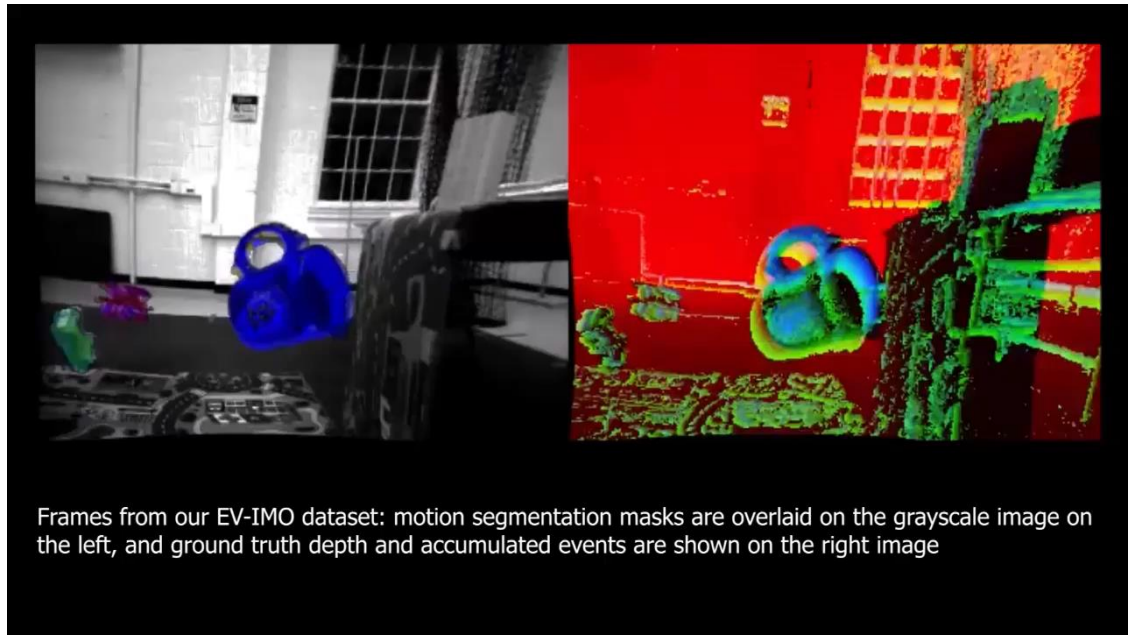
Our Dataset: EV-IMO





First dataset featuring

- Pixelwise object masks
- Depth ground truth
- Object and Camera trajectories



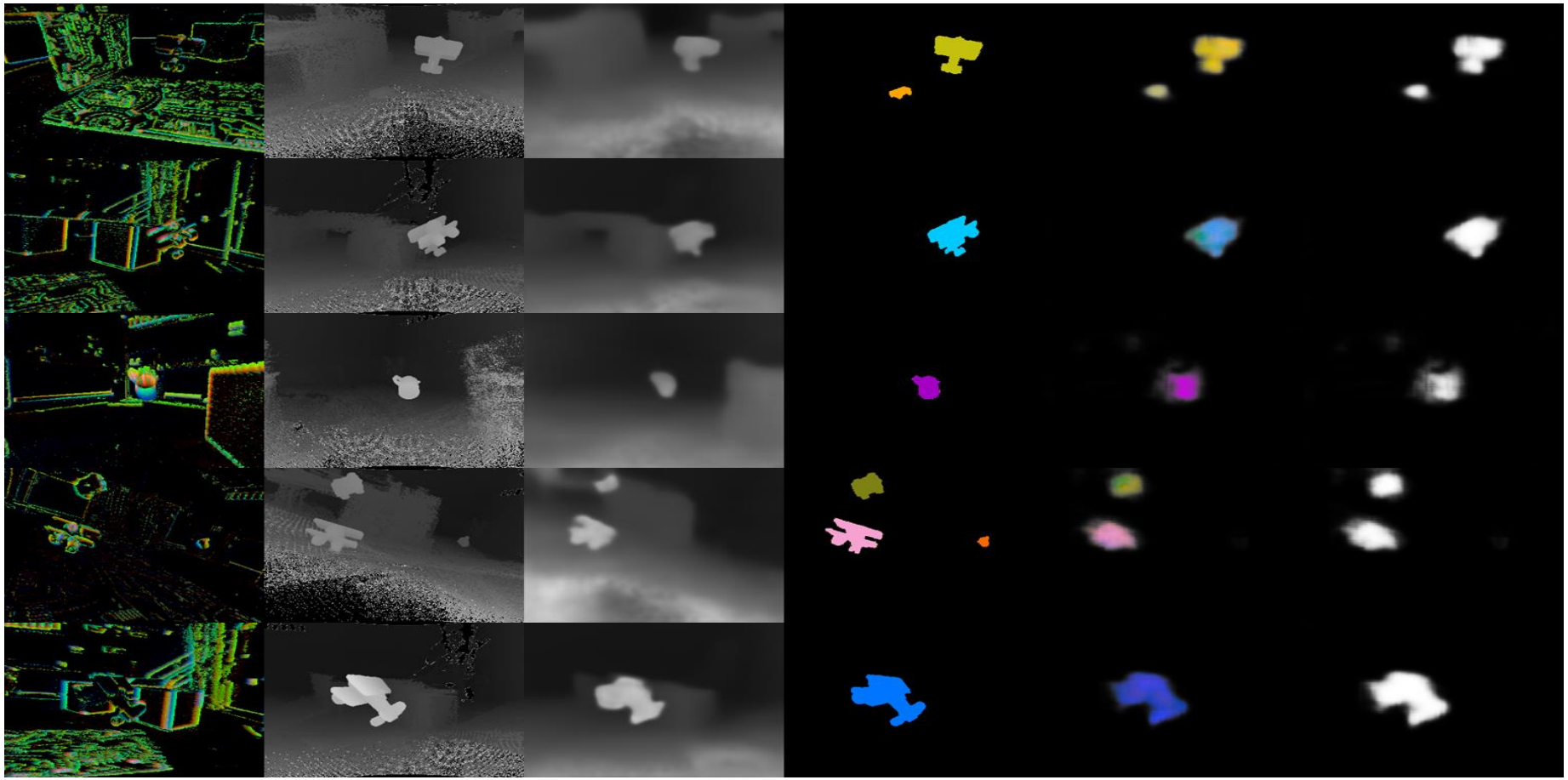
Frames from our EV-IMO dataset: motion segmentation masks are overlaid on the grayscale image on the left, and ground truth depth and accumulated events are shown on the right image

Scene Motion With Event-Based Vision: Learning (II)

- First Work ever to estimate 3D Object Motion and Evaluate it.

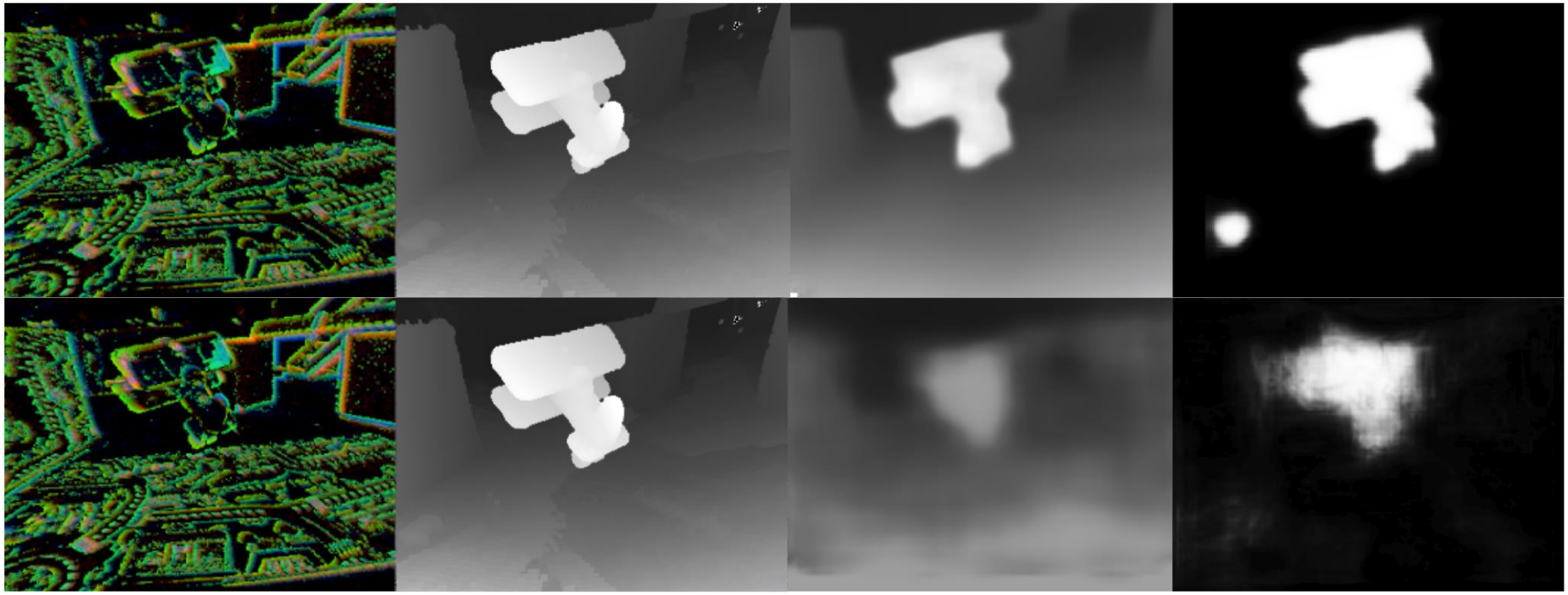
- Supervised (mask and depth)

- Warping done on tiny subslices (closer to 3D)



http://prg.cs.umd.edu/EV-IMO.html

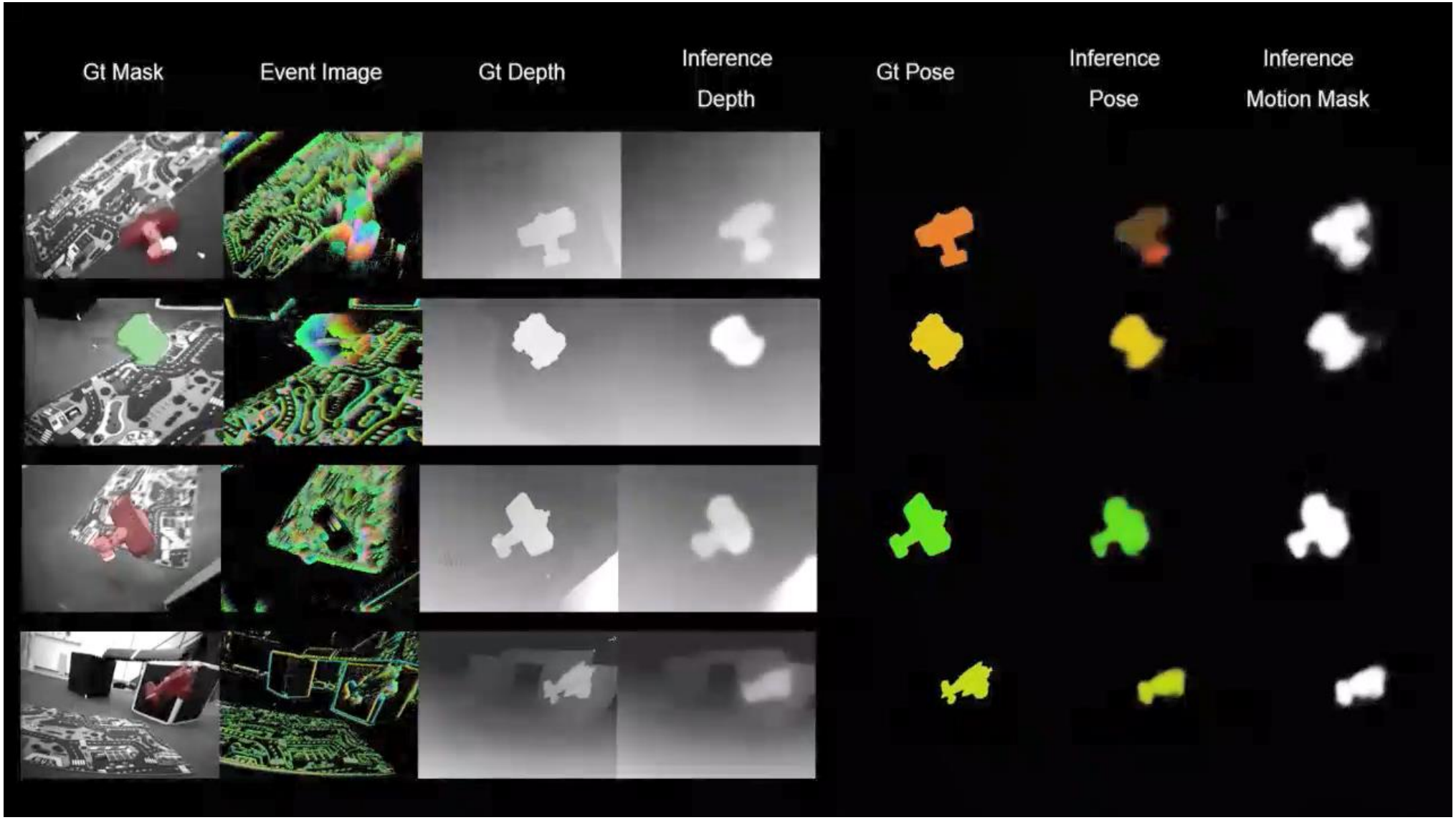
Comparison of full and small network
(2000K Vs 40K parameters)



full

small

Event image | Ground Truth Depth | Estimated Depth | Estimated mask


Gt Mask
Event Image
Gt Depth
Inference Depth
Gt Pose
Inference Pose
Inference Motion Mask

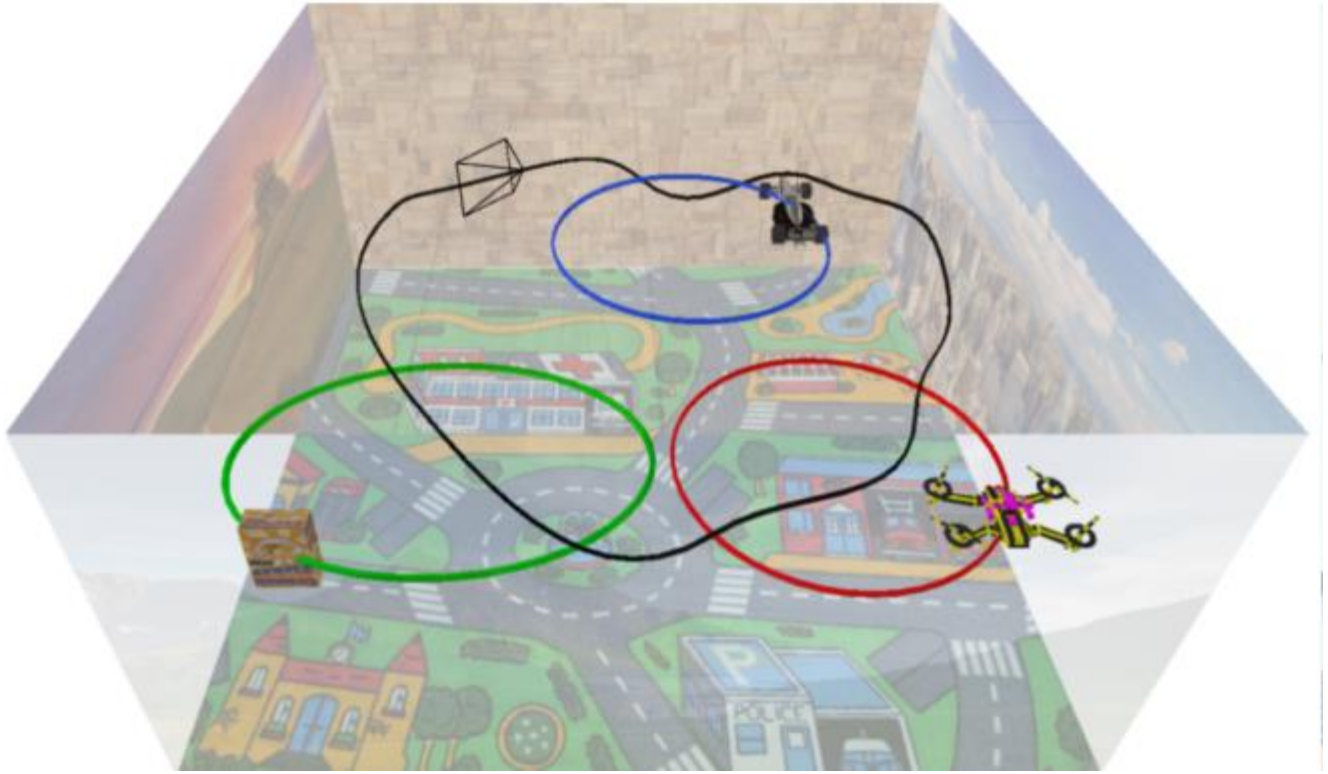
EVDodge

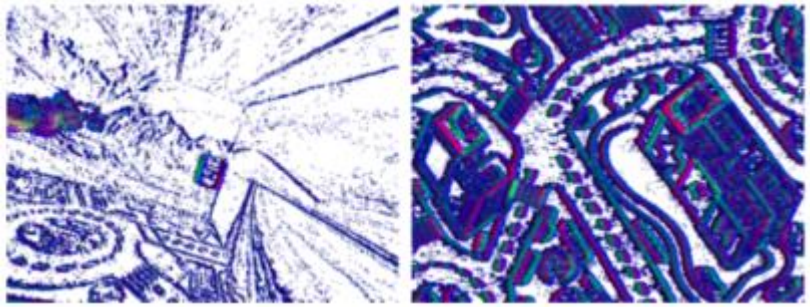
Camera equipped with down- and front-facing DVS, down facing sonar and IMU

All computations done online on a NVIDIA TX2 CPU+GPU

EVDodge: Embodied AI For High-Speed Dodging On A Quadrotor Using Event Cameras. ArXiv
N. Sanket1 , C.. Parameshwara , C.D.Singh, A.. Kuruttukulam1 , C., Fermüller, D. Scaramuzza , Y. Aloimonos

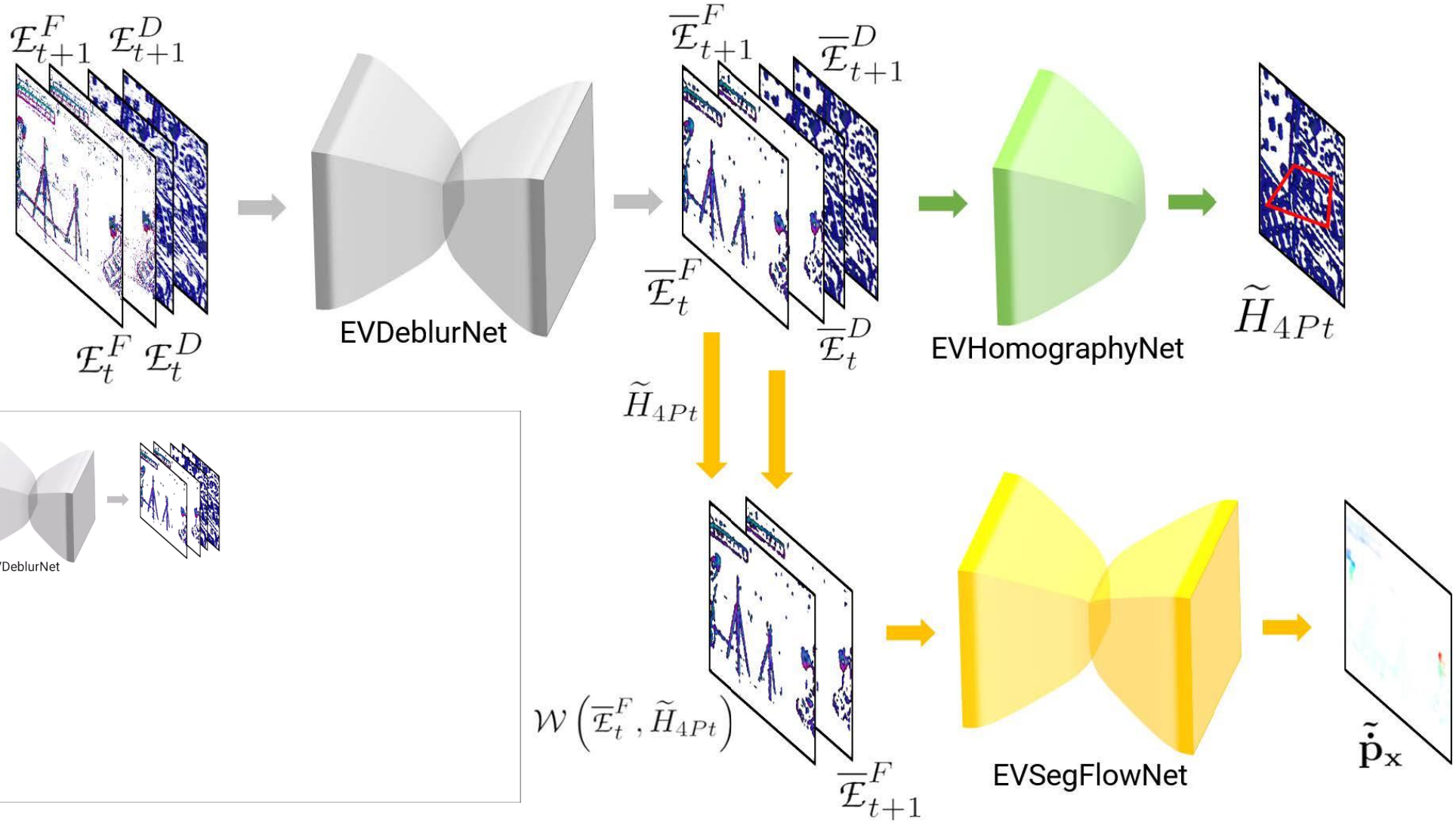
Training in Simulation Environment

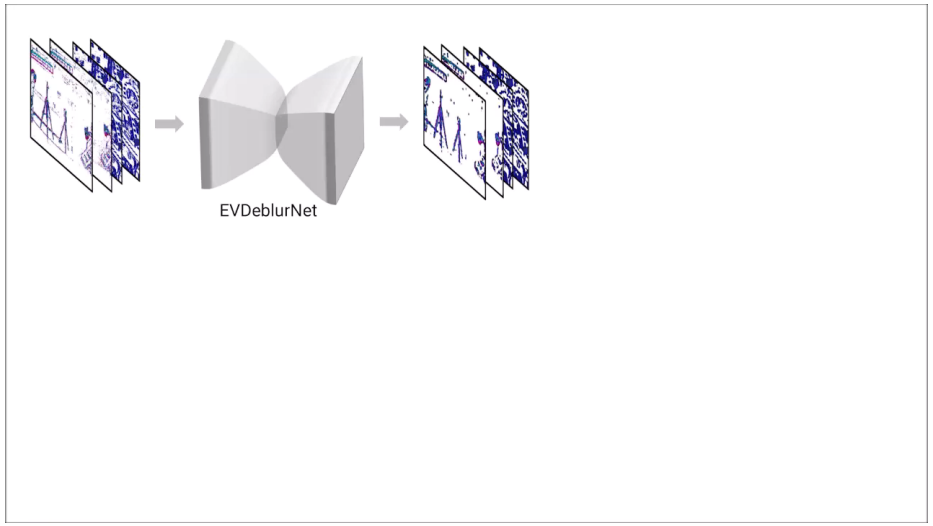


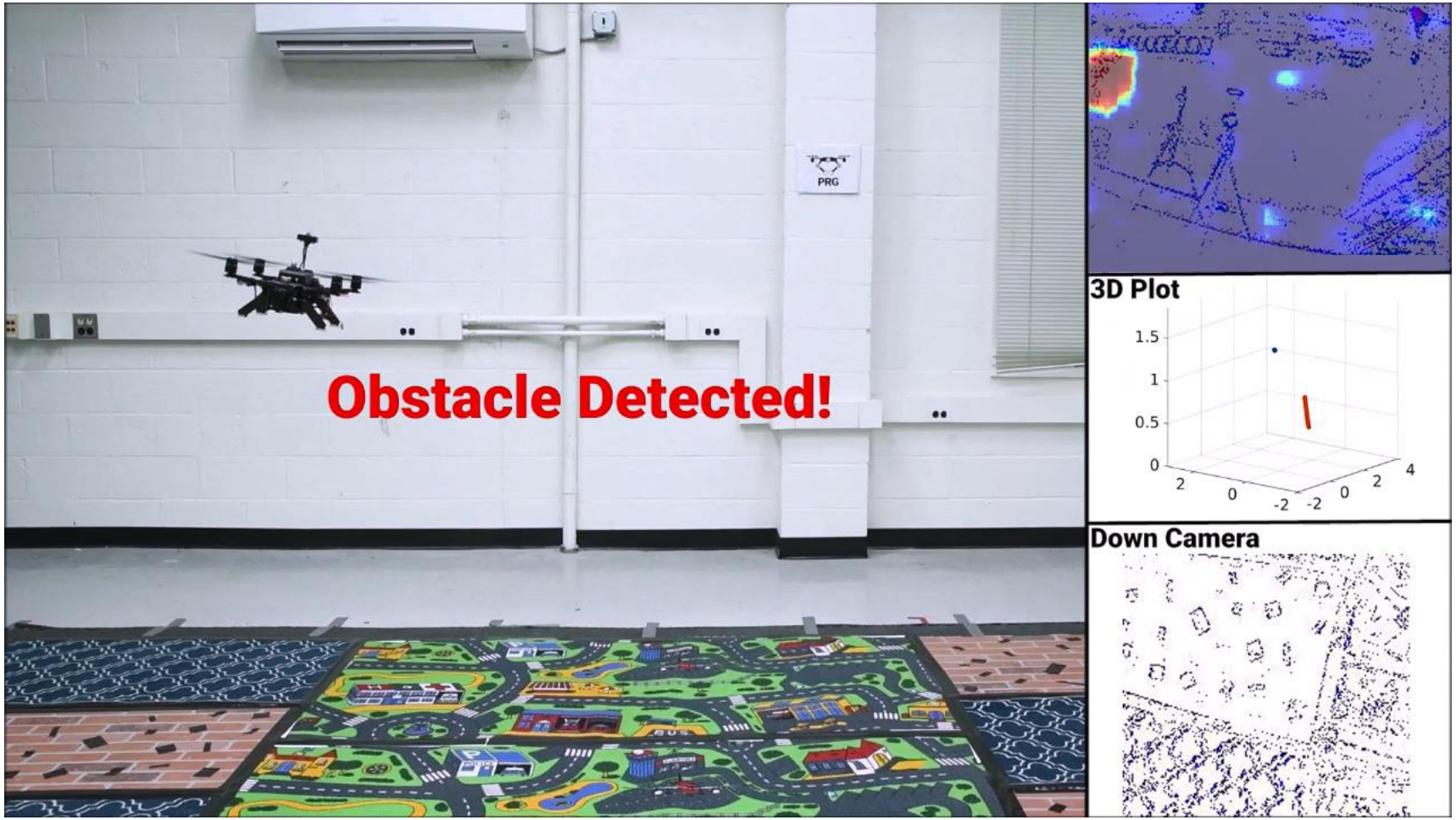


Front and Down-facing simulated events

AI Navigation Stack for Dodging Objects






Obstacle Detected!
PRG
3D Plot
1.5
1
0.5
0
2
0
-2
-2
0
2
4
Down Camera

Summary:

- Events have an important role for robust motion segmentation
- Classic and Neural Network Approaches using event cloud alignment
- New Dataset for Object Motion Segmentation/Estimation
- Approaches for Standard Neural Network Learning with Events
  - Light-weight architecture
  - Cost-functions adapting to Sparseness
  - Matching at multiple time-scales
  - Learning with Simulated data, Deblurring Real Data
- Demonstrated Real Time Actions on Drones