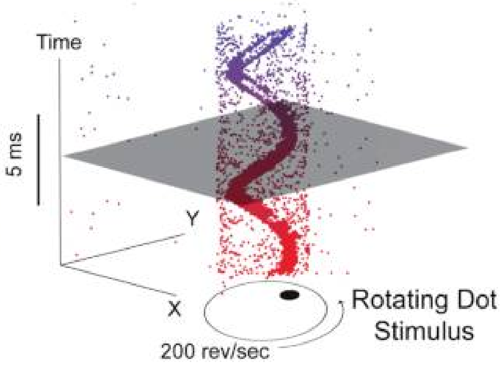
Event-based Vision 2025

CVPR Nashville JUNE 11-15, 2025



Event-based Non-rigid 3D Reconstruction and Novel-View Synthesis

12.06.2025

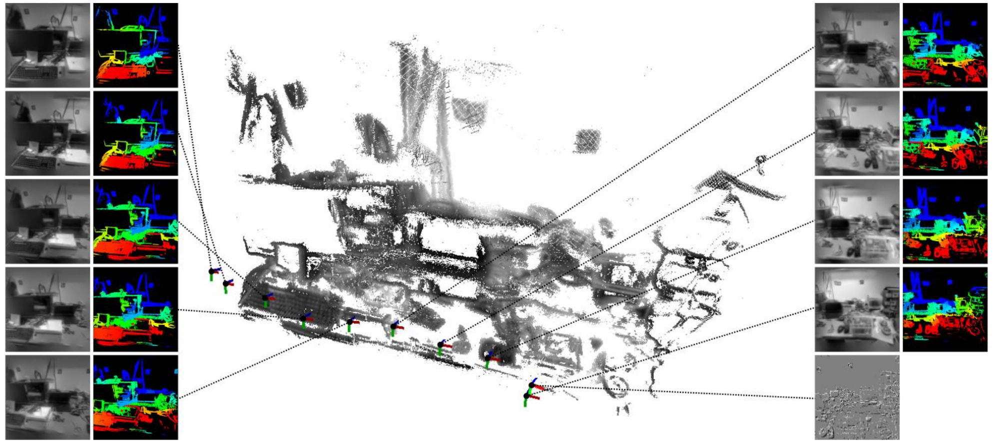
Vladislav Golyanik

golyanik@mpi-inf.mpg.de

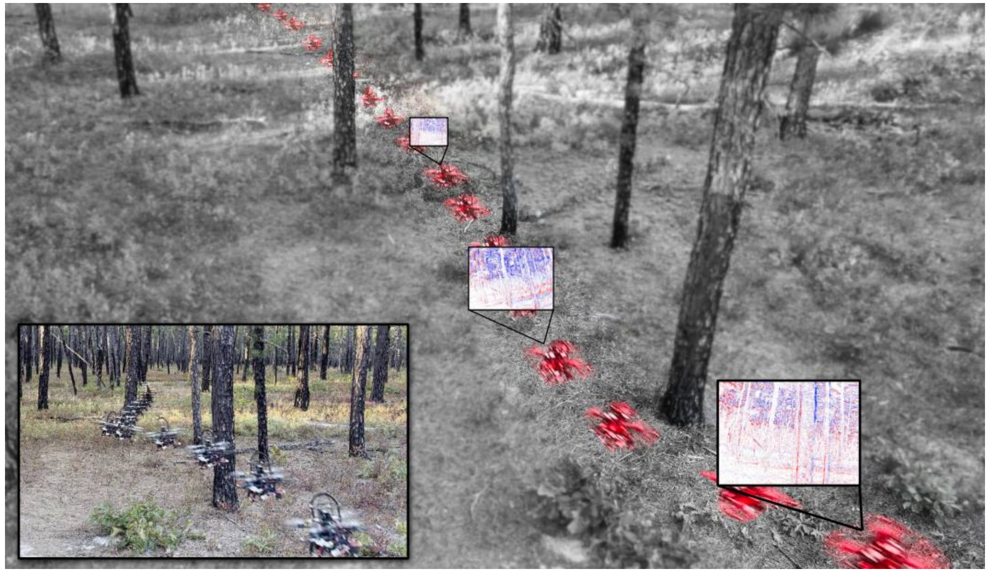
4D and Quantum Vision Group

MAX-PLANCK-INSTITUT FÜR INFORMATIK

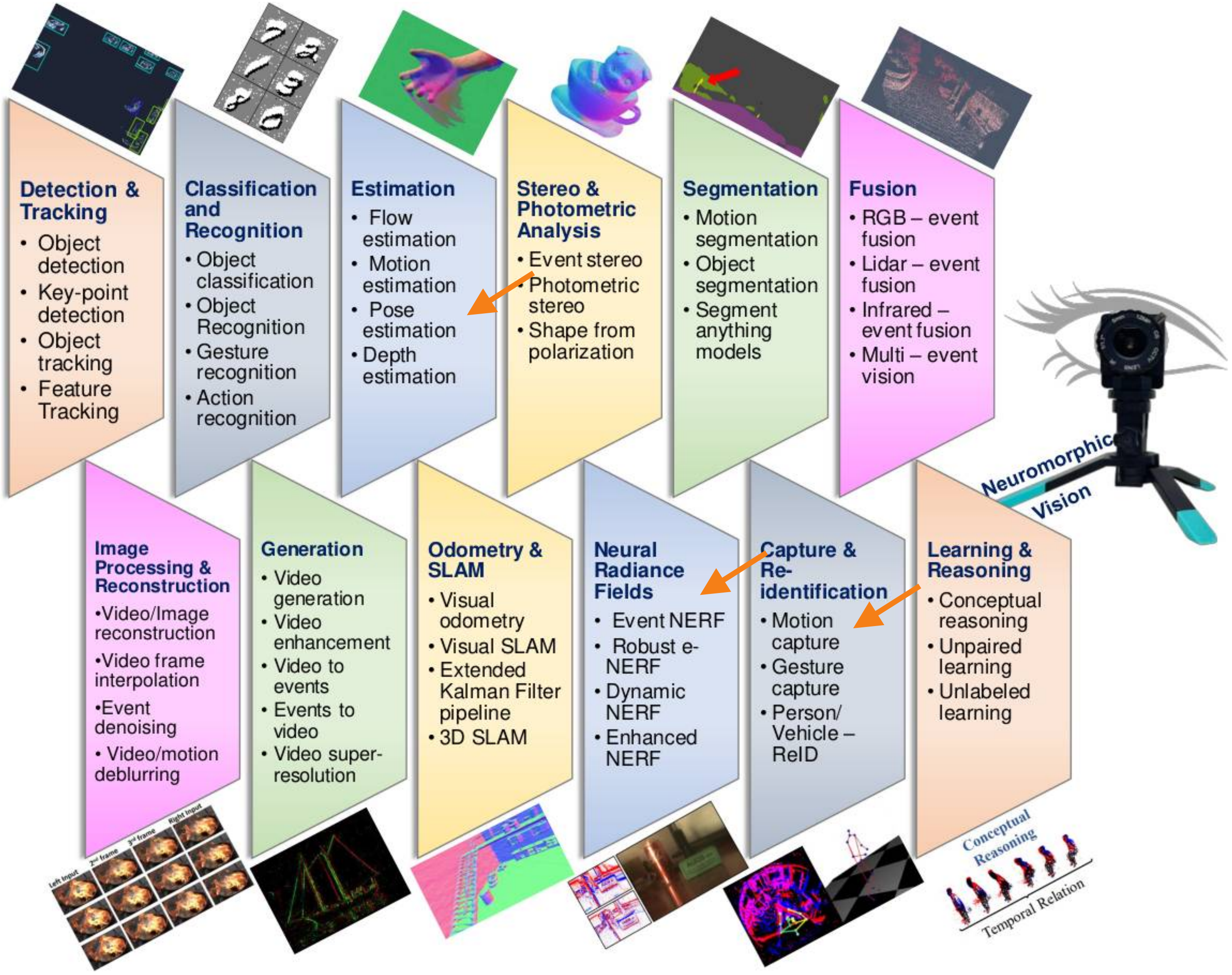
Application Areas of Event Cameras



Event-based SLAM (Kim et al., ECCV 2016)



Bhattacharya et al., CoRL, 2024.



Chakravarthi et al., ECCVW, 2024.

Why is 4D Reconstruction Useful?

Akada et al., ECCV 2022.

AR/VR/Gaming

Mendiratta et al., ToG 2023.

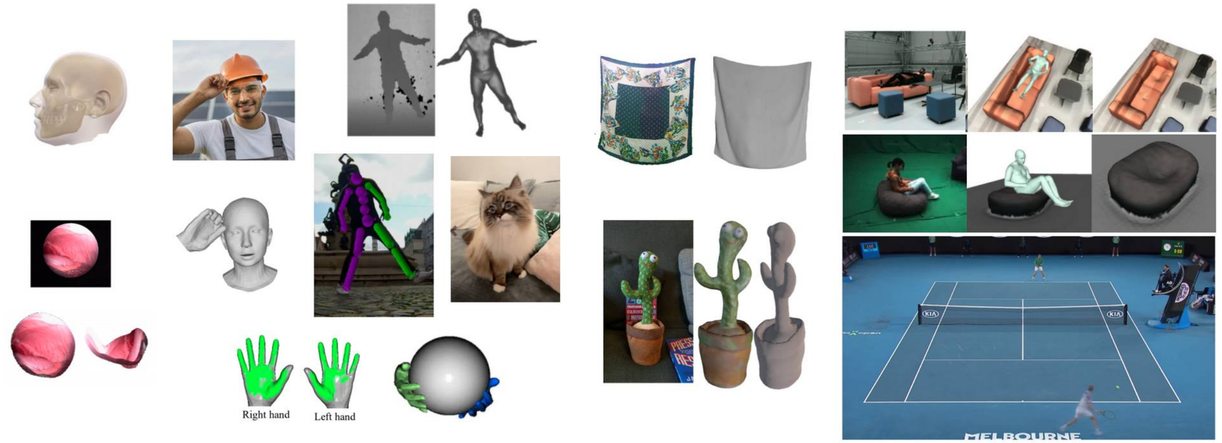
Shiratori et al., FG 2004.

Monocular Video **4D Reconstruction**

Kappel et al., EG 2025.

AR/VR/Gaming image: https://thegamemarketer.com/insight-posts/into-the-metaverse-how-ar-and-vr-are-revolutionising-gaming

Why is 4D Reconstruction Challenging?



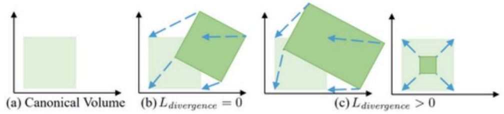
inside living organisms

animals, humans and human body parts

everyday/general objects

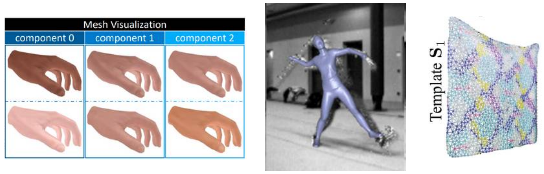
humans in environments

$$L_{divergence} = \frac{1}{|C|} \sum_{j \in C} w'_j \cdot |\mathrm{div}(\mathbf{b}(\bar{\mathbf{r}}(j), \mathbf{l}))|^2$$

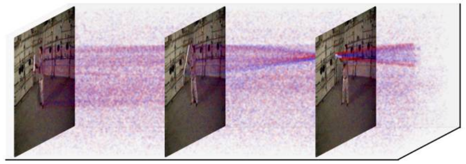


Isometry/Volume Preservation

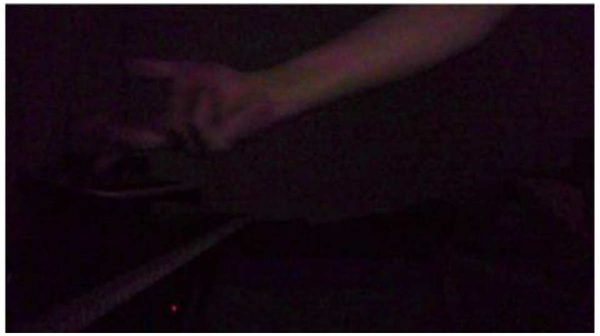
3D Scene Representations



Geometric Priors

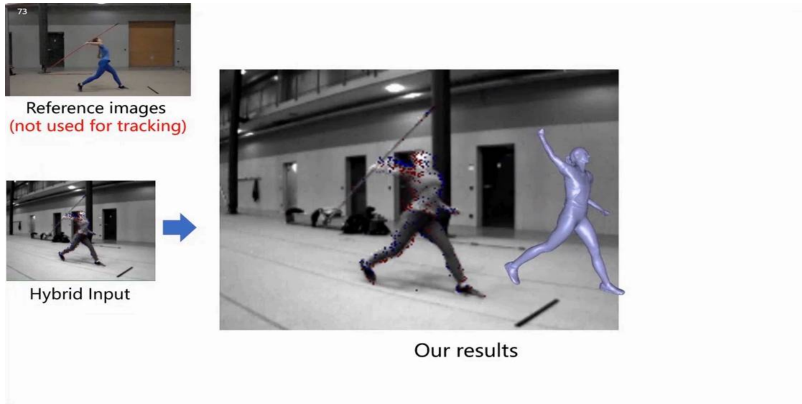


Event Cameras



low lighting

motion blur/high speed

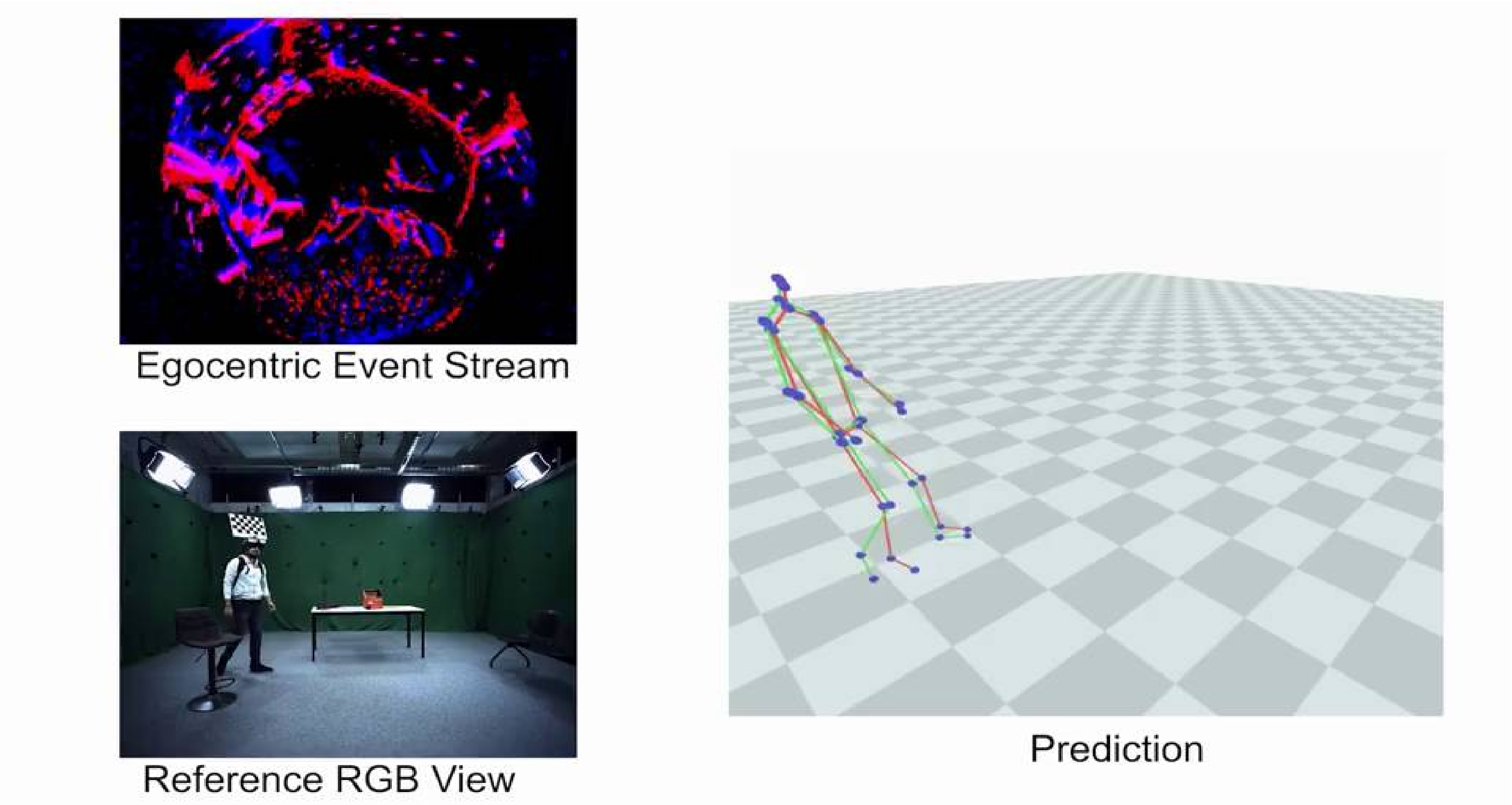


Xu et al. CVPR, 2020.

3D Human Pose Estimation

3D Pose Estimation from an HMD

Head-Mounted Device

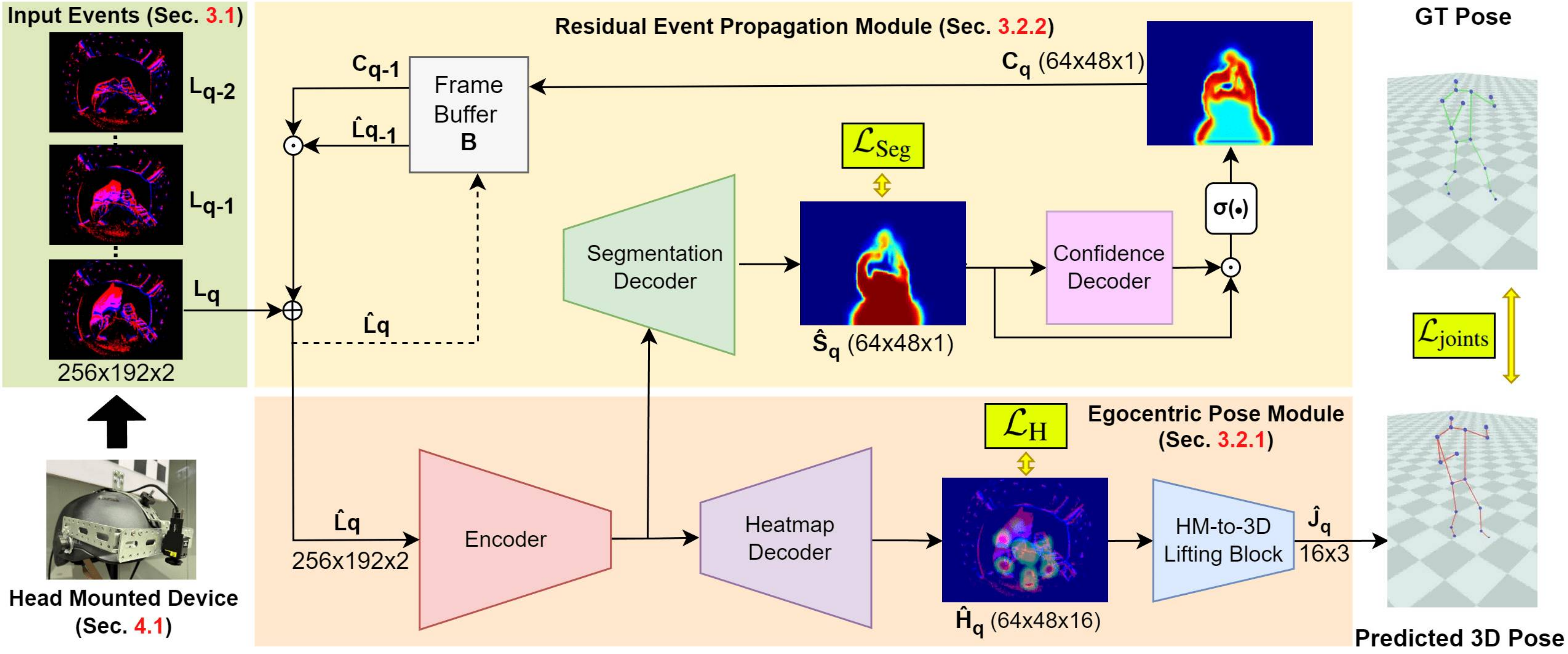


3D Pose Estimation from an HMD

Millerdurai, Akada, Wang, Luvizon, Pagani, Stricker, Theobalt, Golyanik. IJCV, 2025.

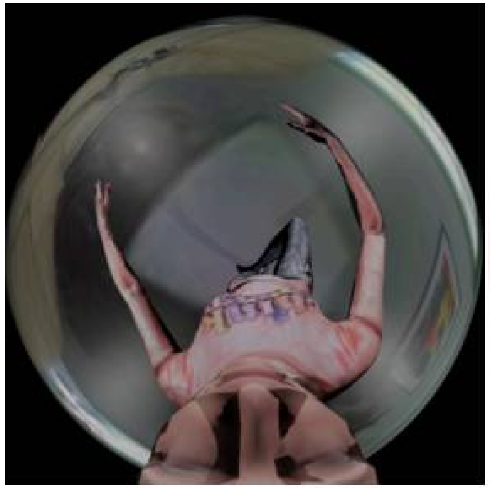
Overview of EventEgo3D++

MAX-PLANCK-INSTITUT
FÜR INFORMATIK

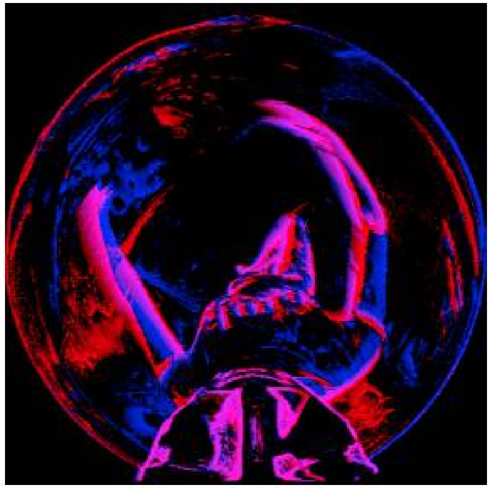


Millerdurai, Akada, Wang, Luvizon, Pagani, Stricker, Theobalt, Golyanik. IJCV, 2025.

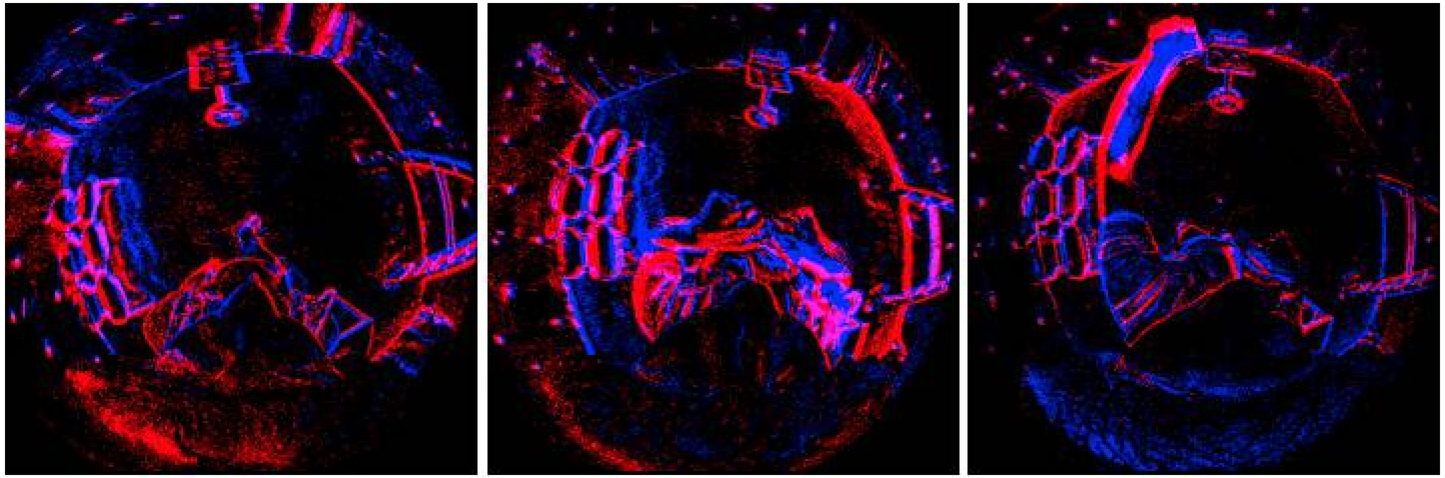
EventEgo3D++



Synthetic RGB Frame

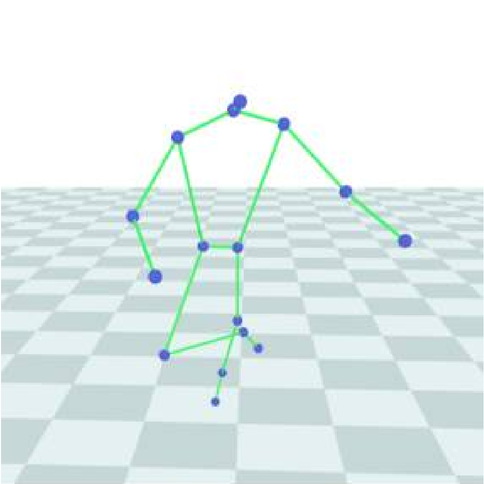


Simulated Event Stream

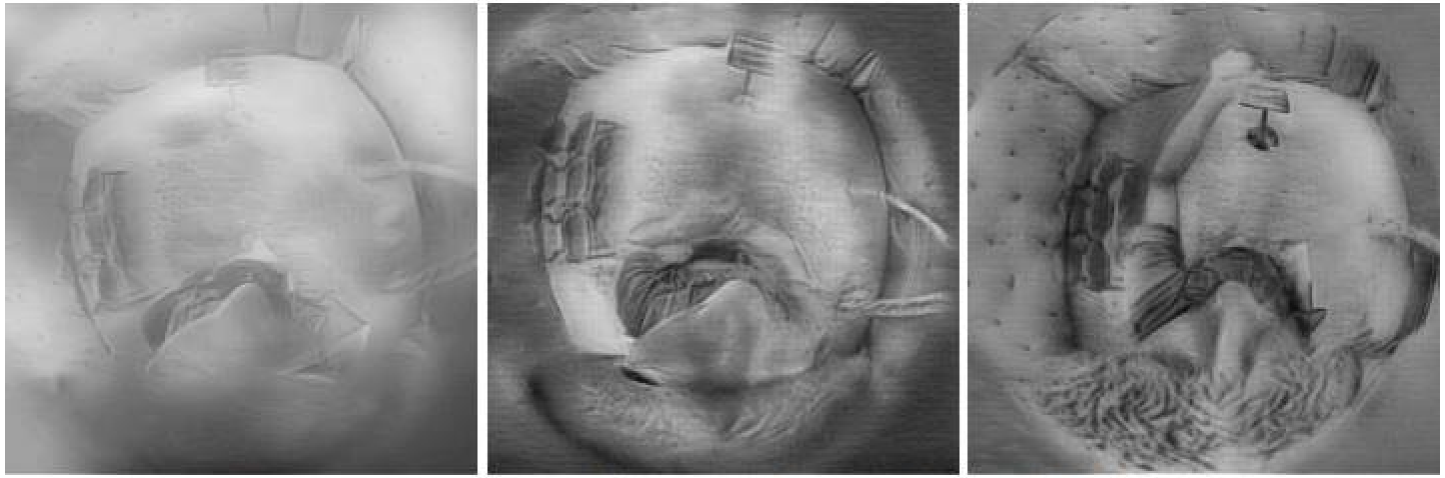


Event Stream

Human Body Mask



3D Body Joints



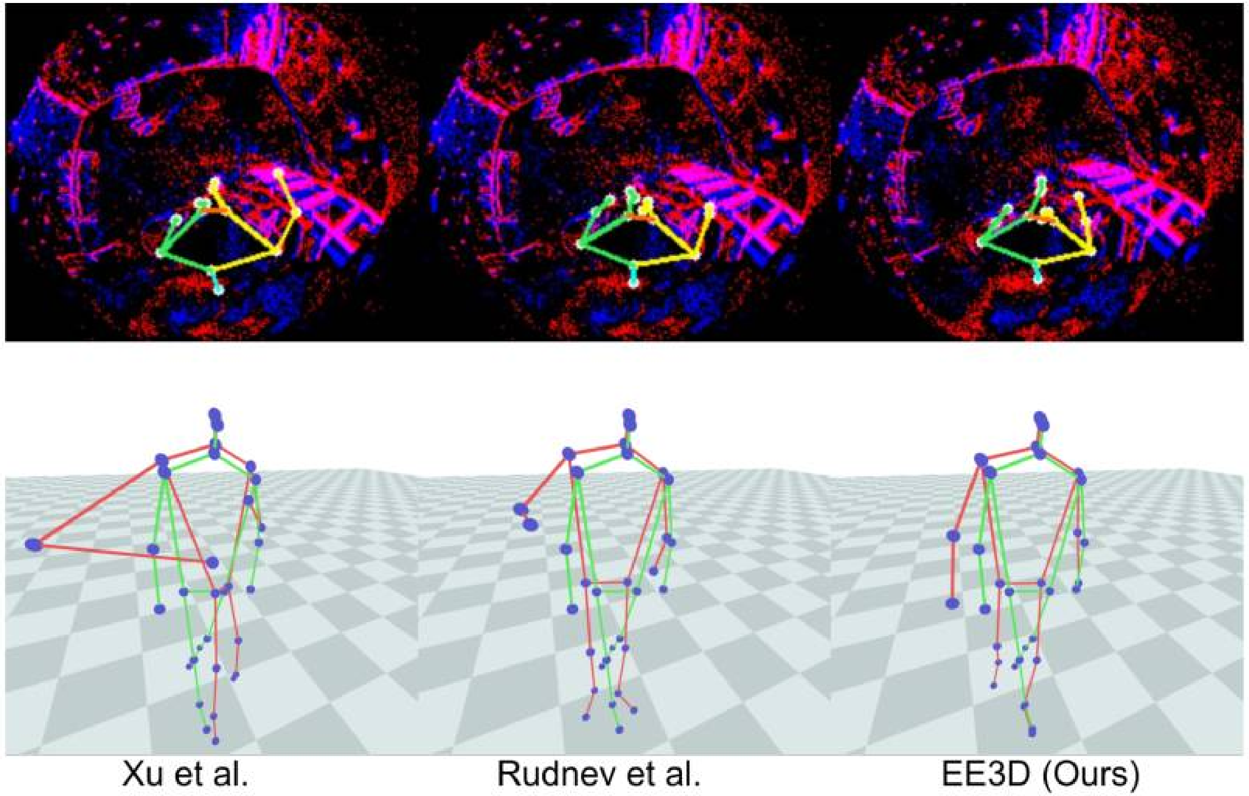
Reconstructed Images from Event Stream

Millerdurai, Akada, Wang, Luvizon, Pagani, Stricker, Theobalt, Golyanik. IJCV, 2025.

EventEgo3D++

Method	Metric	Avg. (σ)
Tome *et al.* [36]	MPJPE	173.01 (23.62)
	PA-MPJPE	113.67 (12.76)
Xu *et al.* [44]	MPJPE	133.53 (36.42)
	PA-MPJPE	100.47 (26.52)
Rudnev *et al.* [33]	MPJPE	114.52 (26.54)
	PA-MPJPE	84.87 (14.08)
EventEgo3D (Ours)	MPJPE	**107.30** (25.78)
	PA-MPJPE	**79.66** (14.83)

Numerical comparisons on the EE3D-R dataset (in mm).



Qualitative Comparisons

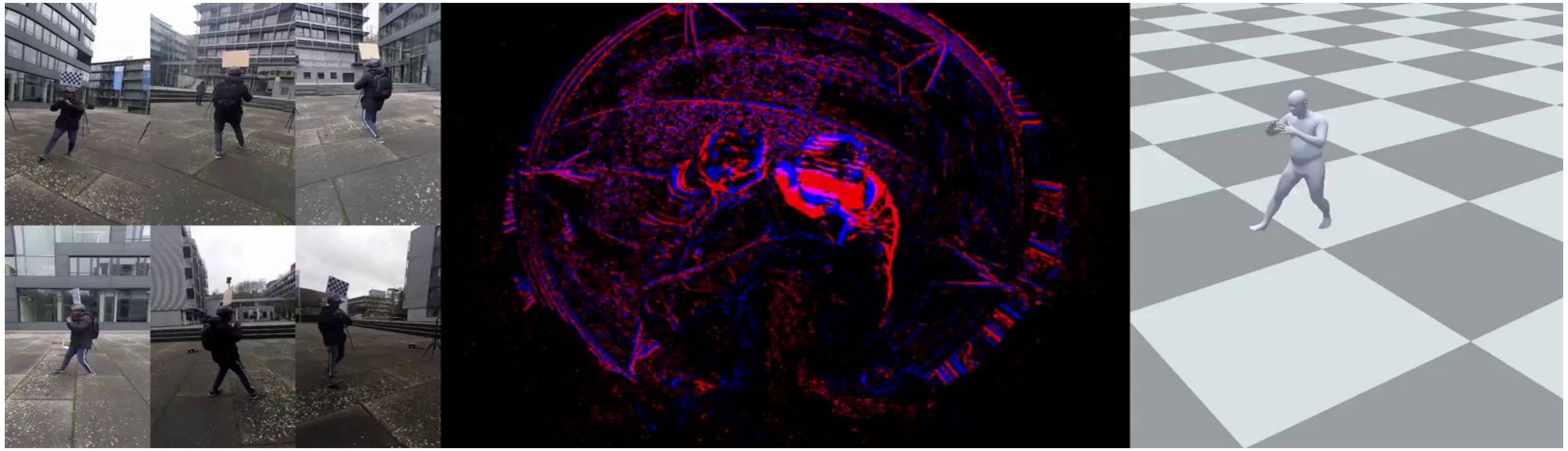
	MPJPE ↓	PA-MPJPE ↓
Baseline (EPM only)	111.01	85.58
Baseline with segmentation decoder	108.85	84.98
Baseline with REPM w/o Confidence decoder	107.58	83.95
EventEgo3D (Ours)	**107.30**	**79.66**

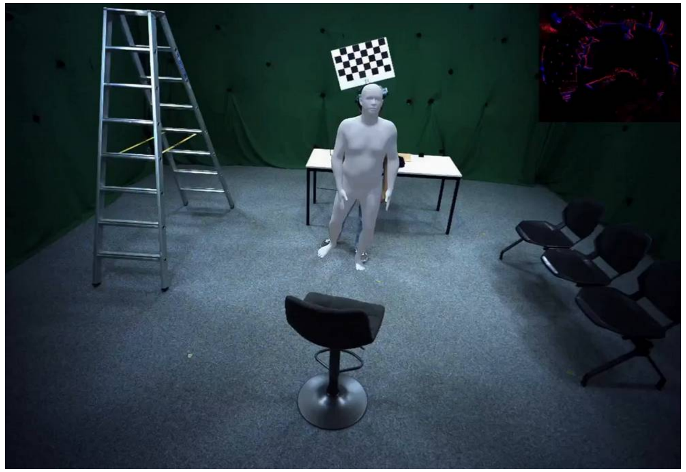
Ablative study on the EventEgo-R dataset.

	Params ↓	FLOPs ↓	Pose Update Rate ↑
Tome *et al.* [36]	77.01M	11.46G	77.07
Rudnev *et al.* [33]	11.2M	3.58G	**489.56**
Xu *et al.* [44]	82.18M	44.06G	68.65
EventEgo3D (Ours)	**1.25M**	**416.84M**	139.88

Comparison between the tested methods in terms of the number of parameters, required performance and the supported pose update rates.

EventEgo3D++





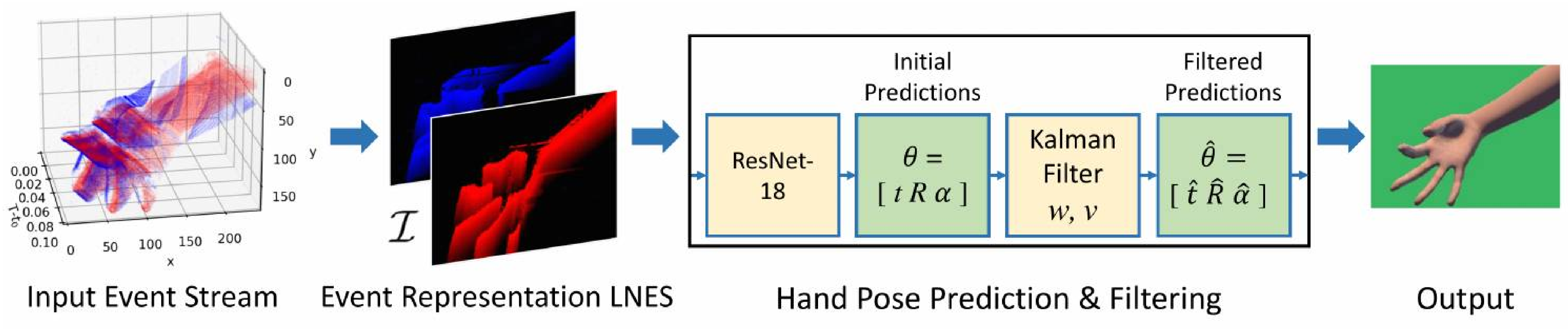
SMPL [Loper et al., 2015] annotations for EE3D-Wild and EE3D-Real

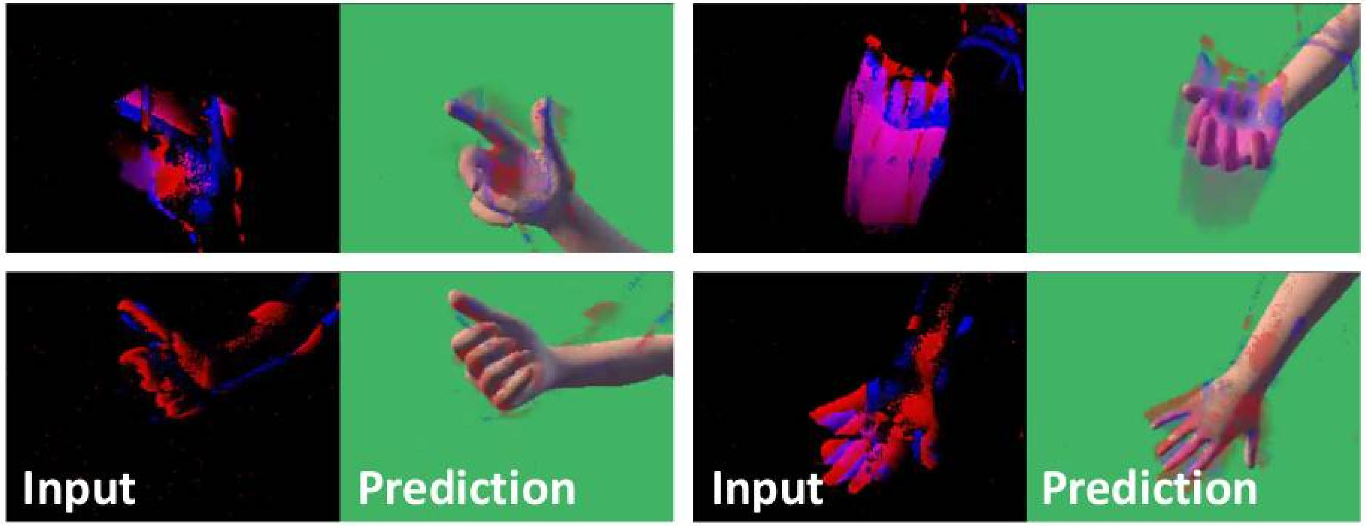
- 12 sequences, 155 minutes (indoors); 9 sequences, 116 minutes (outdoors)
- 30 cameras, 50 fps (indoors); outside: 6 cameras, 60 fps (outdoors)
- Ground-truth 3D annotations using a multi-view camera set-up

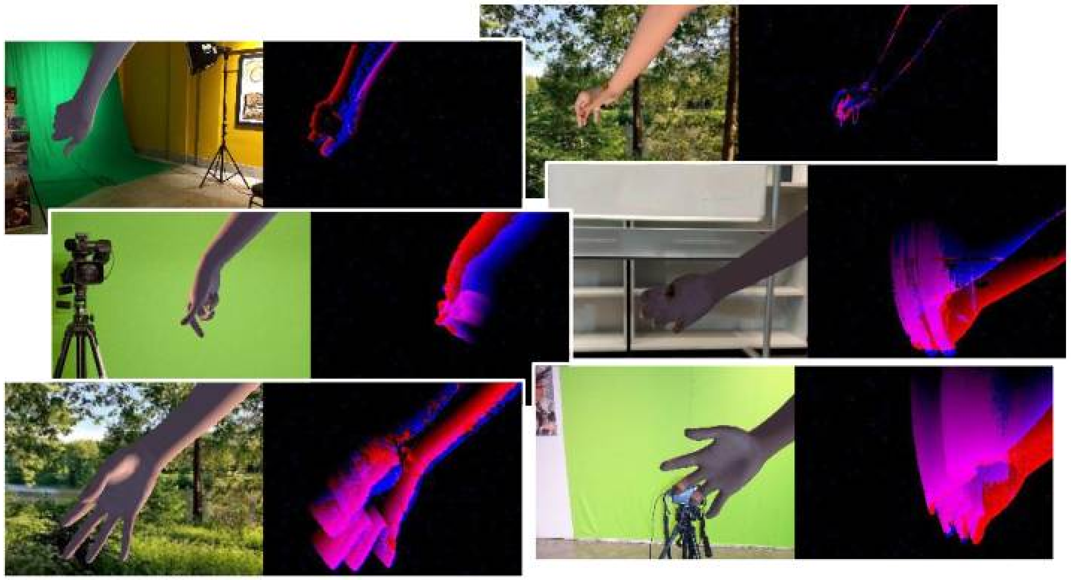
Millerdurai, Akada, Wang, Luvizon, Pagani, Stricker, Theobalt, Golyanik. IJCV, 2025.

Event-based 3D Hand Tracking

EventHands

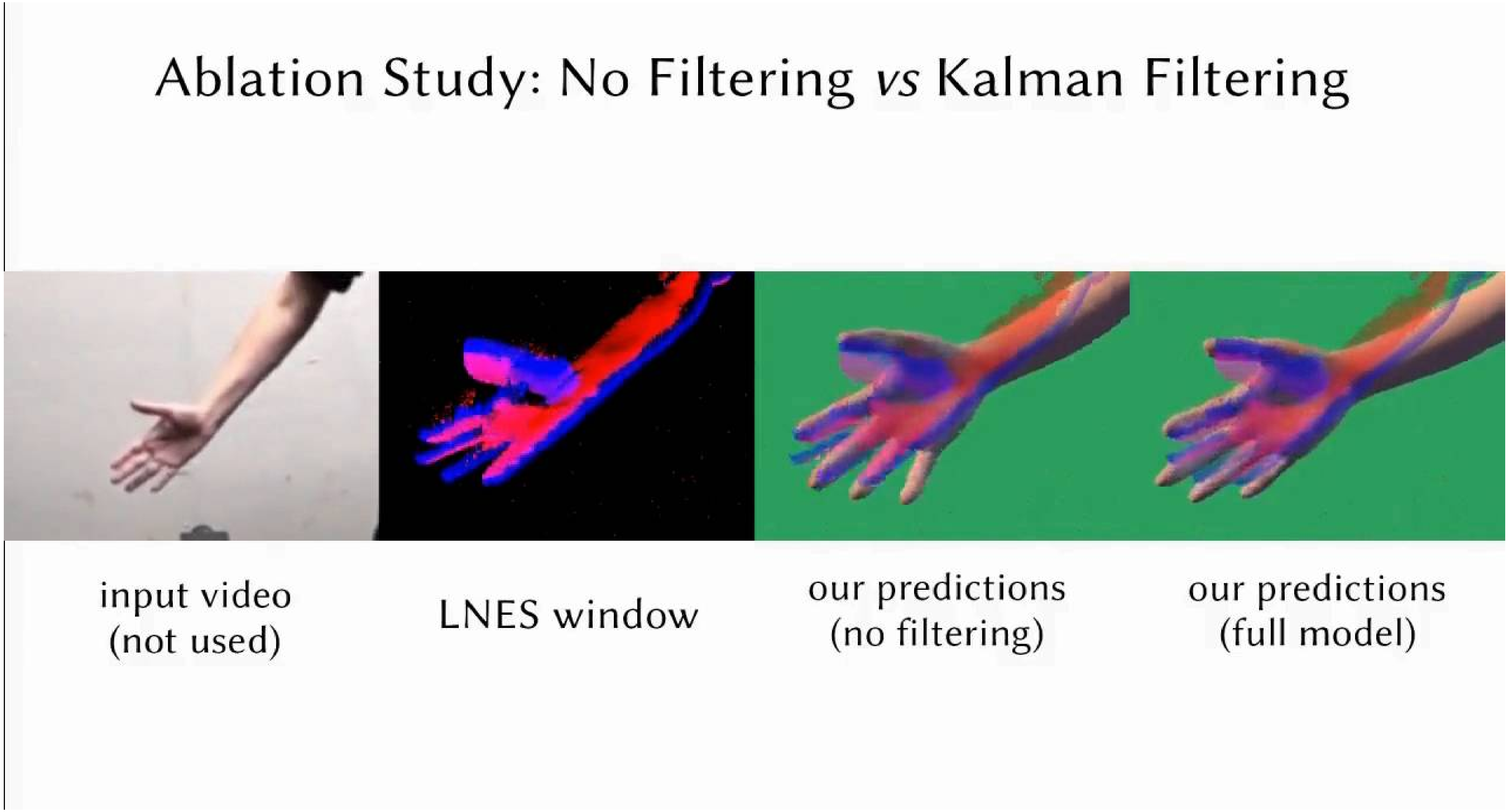






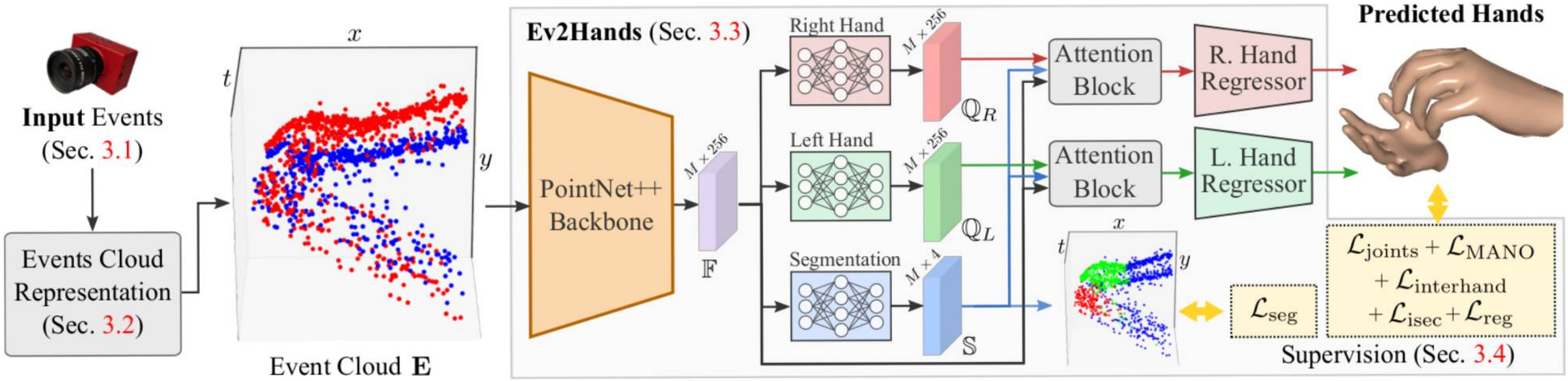
Rudnev, Golyanik, Wang, Seidel, Müller, Elgharib, Theobalt. ICCV, 2021.

EventHands



Rudnev, Golyanik, Wang, Seidel, Müller, Elgharib, Theobalt. ICCV, 2021.

Ev2Hands



Workflow of the approach for tracking two interacting hands in 3D from a single event camera.

Event cloud $\mathbf{E} \in \mathbb{R}^{M\times 5}$, with $\mathbf{E}_k = (x_k, y_k, t_k, P_k, N_k)$

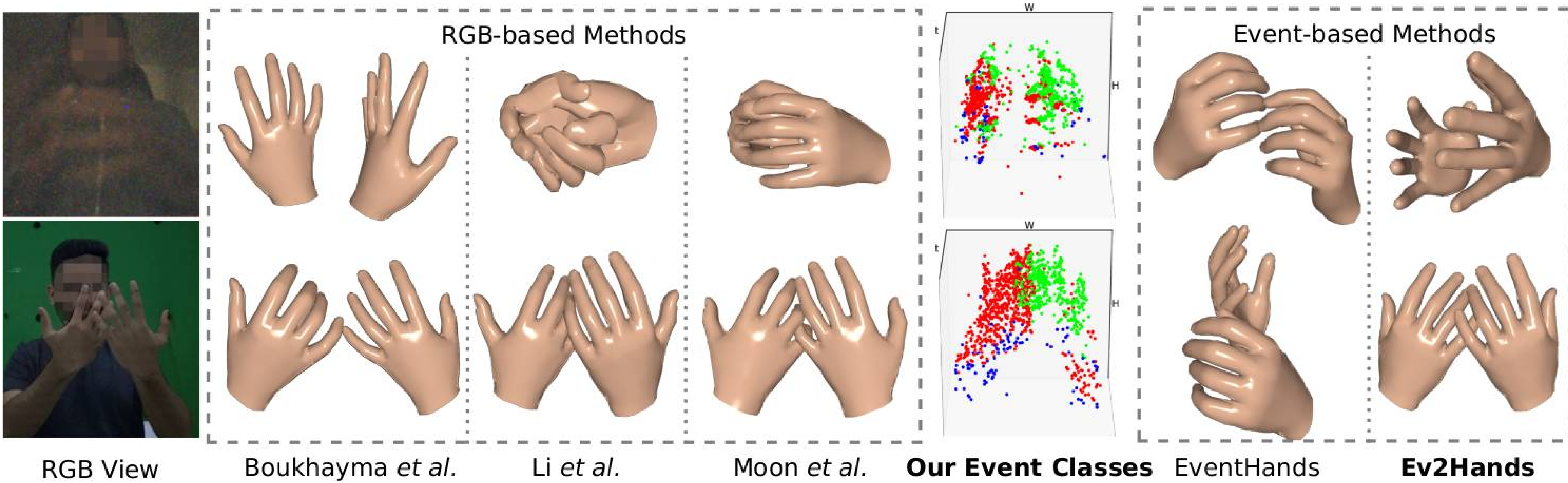
M: the total number of events

t_k: average time of the combined events

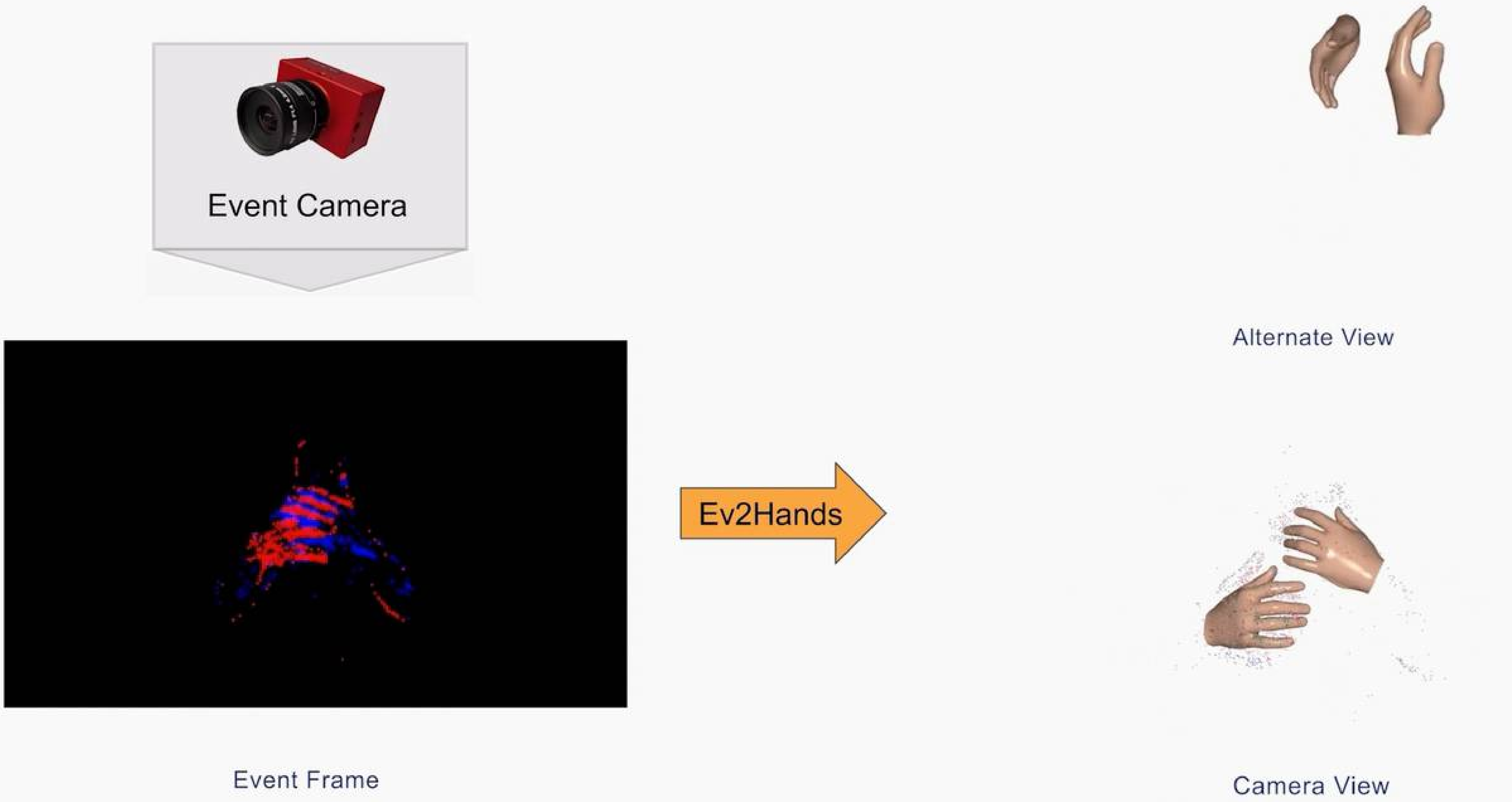
P_k, N_k: Normalised numbers of positive and negative events

Millerdurai, Luvizon, Rudnev, Jonas, Wang, Theobalt, Golyanik. 3DV, 2024.

Ev2Hands

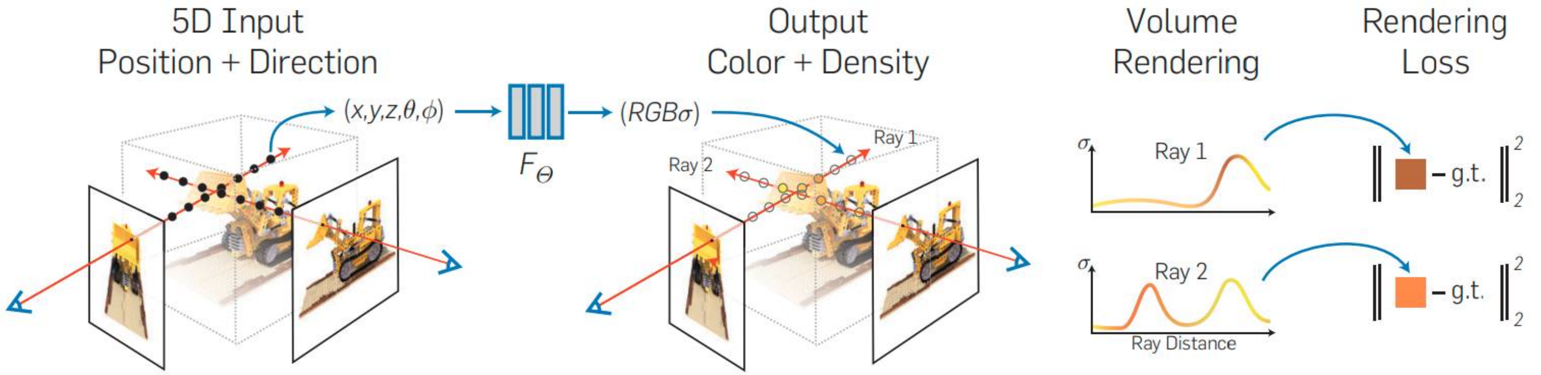


Qualitative comparisons

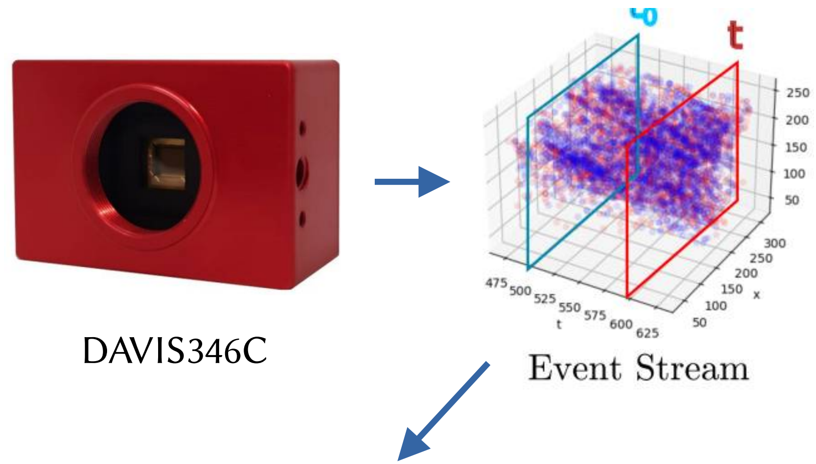

Event Camera
Event Frame
Ev2Hands
Alternate View
Camera View

Event-based Novel-view Rendering

RGB- *vs* Event-based Neural Rendering



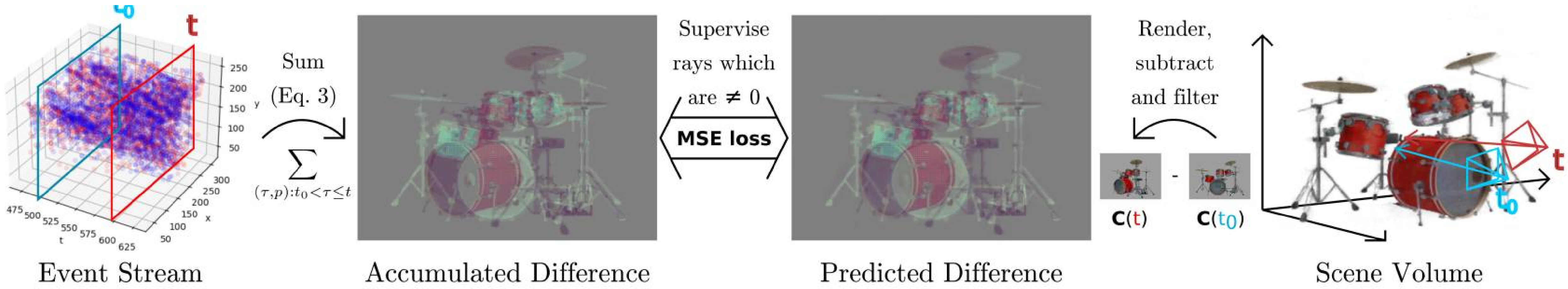
Mildenhall et al. NeRF. Communications of the ACM, 2021.



?

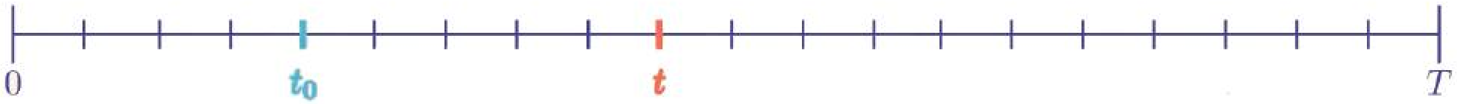
Thought Experiment

EventNeRF

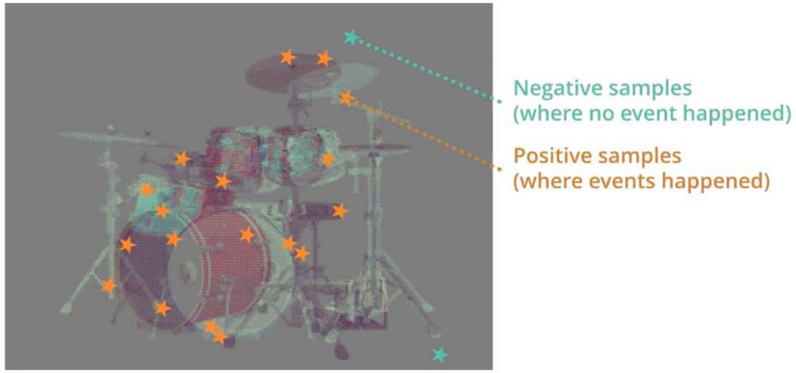


Learning a NeRF volume of a static object using a moving colour event camera

$$\mathcal{L}_{\text{recon},xy}(t_0, t) = \mathbf{MSE}(\hat{\mathbf{L}}_{xy}(t) - \hat{\mathbf{L}}_{xy}(t_0), E_{xy}(t_0, t))$$



$$\mathcal{L} = \frac{1}{N_{\text{windows}}} \sum_{i=1}^{N_{\text{windows}}} \mathcal{L}_{\text{recon}}(t_0, t) \quad t = \frac{i}{N_{\text{windows}}}, t_0 \sim U[t - L_{\max}, t)$$

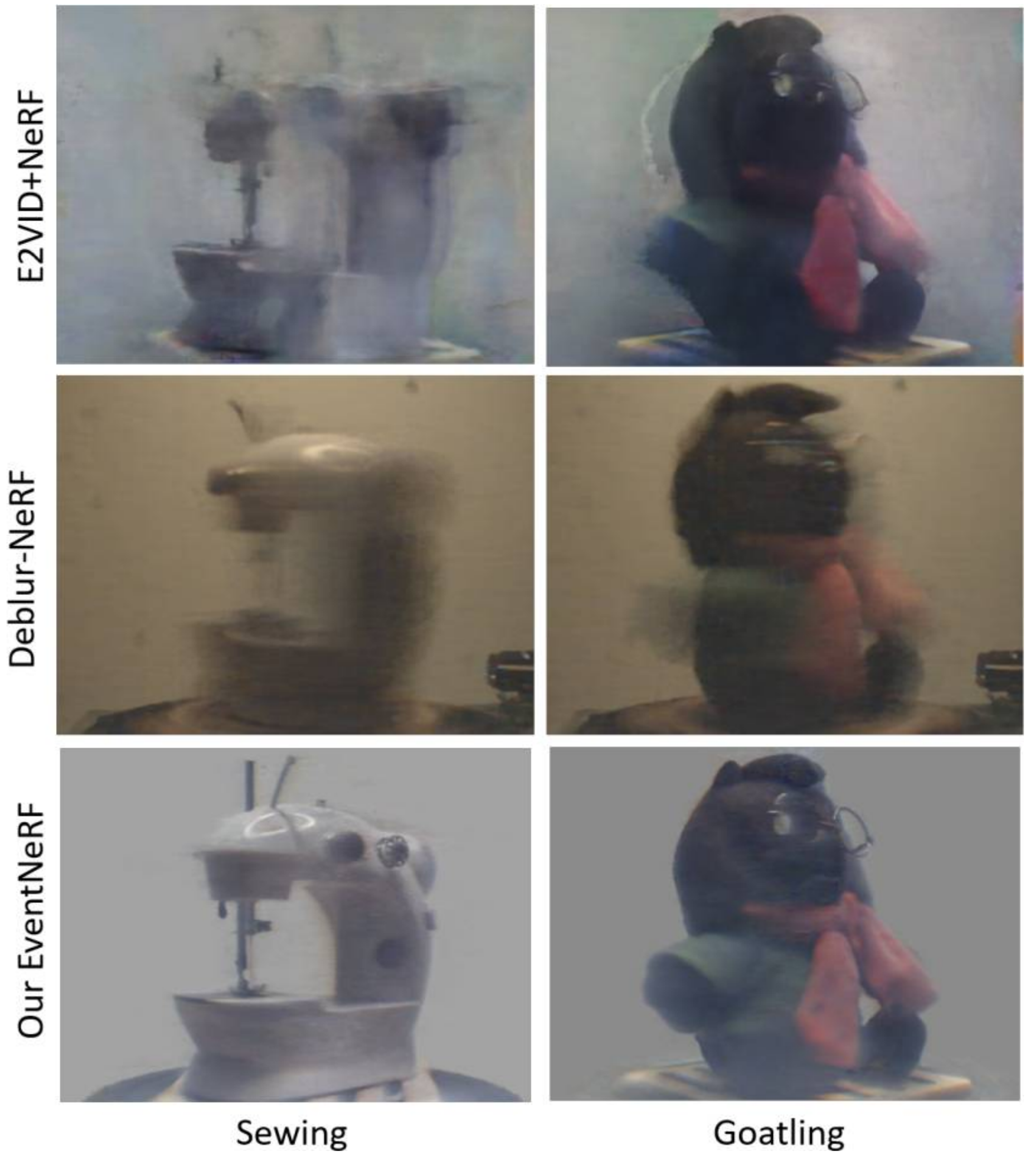


Rudnev, Elgharib, Theobalt, Golyanik. CVPR, 2023.

Numerical Comparisons

Scene	E2VID [45] + NeRF [37] PSNR↑	SSIM↑	LPIPS↓	Our EventNeRF PSNR↑	SSIM↑	LPIPS↓
Drums	19.71	0.85	0.22	**27.43**	**0.91**	**0.07**
Lego	20.17	0.82	0.24	**25.84**	**0.89**	**0.13**
Chair	24.12	0.92	0.12	**30.62**	**0.94**	**0.05**
Ficus	24.97	0.92	0.10	**31.94**	**0.94**	**0.05**
Mic	23.08	0.94	0.09	**31.78**	**0.96**	**0.03**
Hotdog	24.38	0.93	0.12	**30.26**	**0.94**	**0.04**
Materials	22.01	0.92	0.13	**24.10**	**0.94**	**0.07**
Average	22.64	0.90	0.15	**28.85**	**0.93**	**0.06**

Quantitative comparisons against E2VID+NeRF

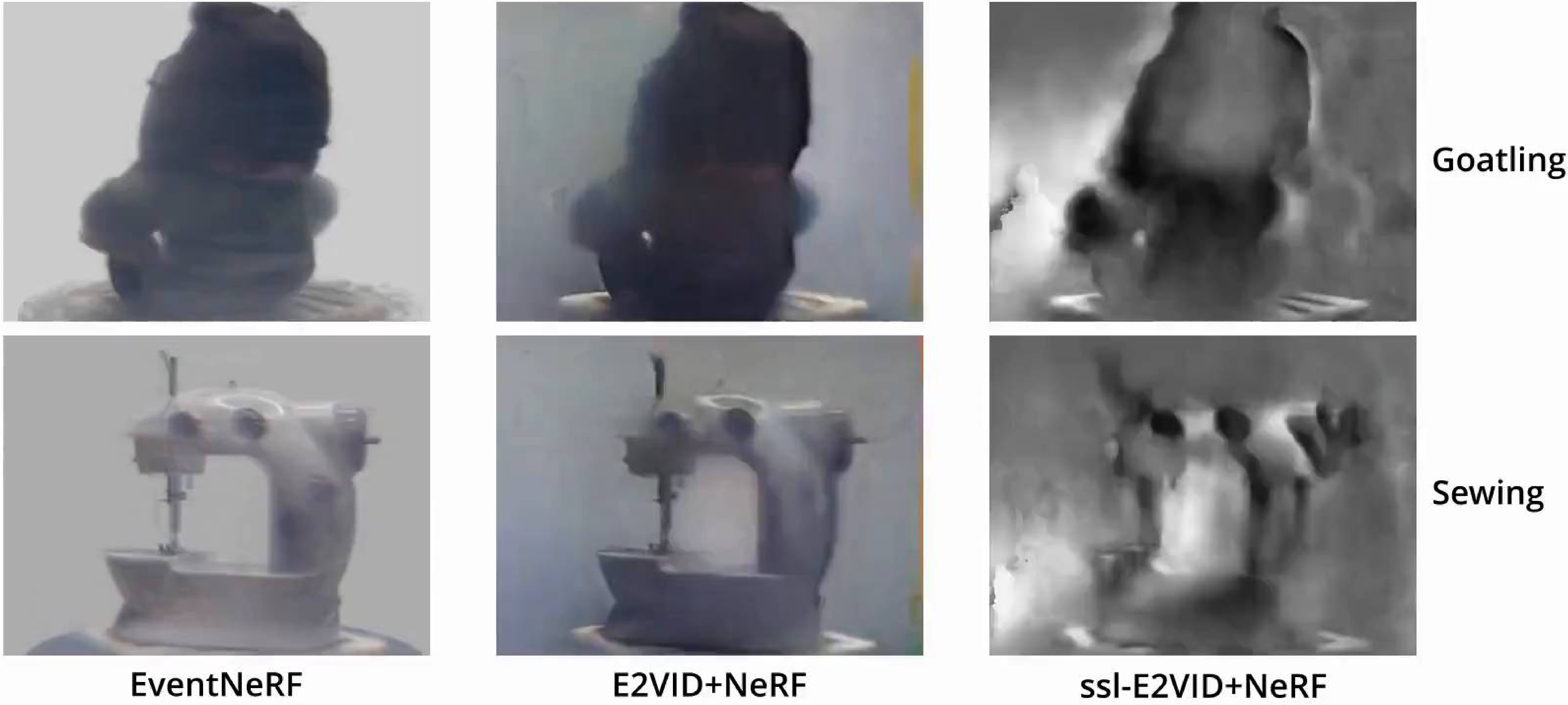


- Reconstructing RGB frames first is sub-optimal!
- Deblur-NeRF's assumptions are not satisfied

E2VID: Rebecq et al. TPAMI, 2019.
NeRF: Mildenhall et al., ECCV 2020.
Deblur-NeRF: Ma et al., CVPR 2022.

Rudnev, Elgharib, Theobalt, Golyanik. CVPR, 2023.

EventNeRF



Rudnev, Elgharib, Theobalt, Golyanik. CVPR, 2023.

EventNeRF

EventNeRF
(real events)

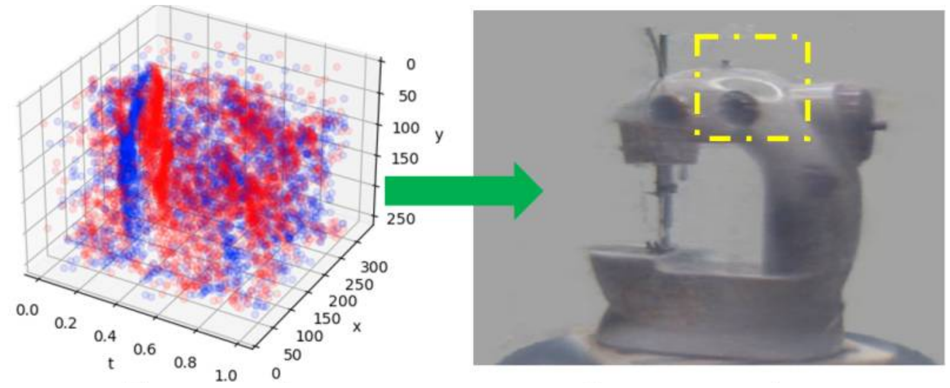
EventNeRF
(synthetic events)

Deblur-NeRF
(real frames)

Deblur-NeRF
(synthetic frames)

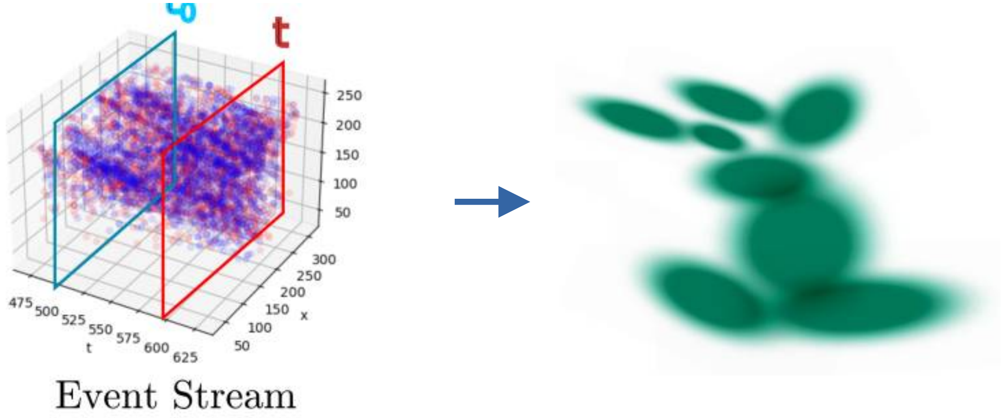
Rudnev, Elgharib, Theobalt, Golyanik. CVPR, 2023.

What's Next?



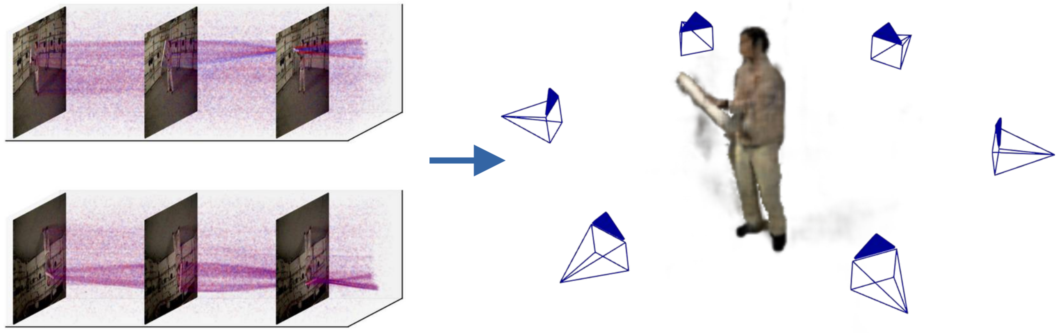
EventNeRF

i/



a **new 3D representation** for rigid scenes
Zahid et al., 3DV 2025

ii/

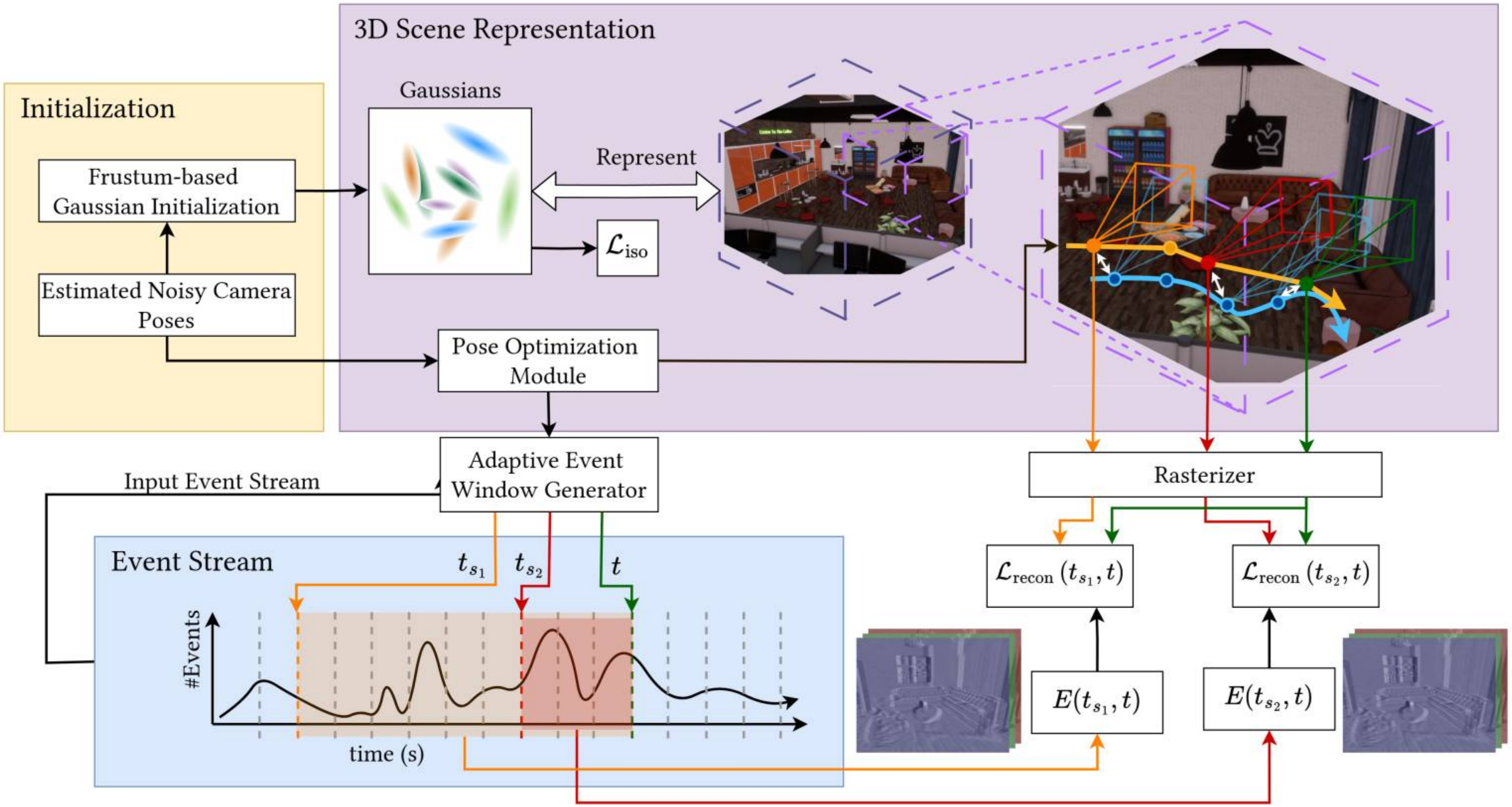


multiple event cameras, a **dynamic** scene
Rudnev et al., CVPRW 2025

iii/

single event camera, a **dynamic** scene (WiP)

E-3DGS: A New Scene Representation



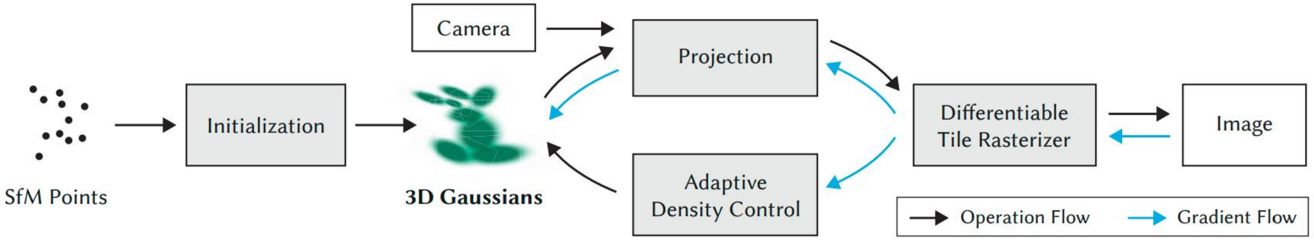
Overview of E-3DGS

synthetic dataset

real dataset

E-3DGS: A New Scene Representation

MAX-PLANCK-INSTITUT
FÜR INFORMATIK



Kerbl *et al.* 3DGS. SIGGRAPH, 2023.

$$G_i(\boldsymbol{x}) = \exp\left(-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu}_i)^T \boldsymbol{\Sigma}_i^{-1}(\boldsymbol{x}-\boldsymbol{\mu}_i)\right)$$

scene location — 3D covariance matrix — Gaussian centroid

$$\boldsymbol{\Sigma}_i' = \boldsymbol{J}\boldsymbol{W}\boldsymbol{\Sigma}_i\boldsymbol{W}^T\boldsymbol{J}^T$$

2D projection — viewing direction — the Jacobian of the affine approx. of the projective transformation

Specific to E-3DGS:

$$\mathcal{L}_{\text{recon}}(t_s, t) = \frac{\alpha}{|\mathcal{X}_{\text{noevs}}|} \cdot \left(\sum_{\boldsymbol{x}\in\mathcal{X}_{\text{noevs}}} \mathcal{L}_{\boldsymbol{x}}(t_s, t)\right) + \frac{1-\alpha}{|\mathcal{X}_{\text{evs}}|} \cdot \left(\sum_{\boldsymbol{x}\in\mathcal{X}_{\text{evs}}} \mathcal{L}_{\boldsymbol{x}}(t_s, t)\right)$$

Event and **no-event** parts of the reconstruction loss

$$\mathcal{L}_{\text{iso}} = \frac{1}{|\mathcal{G}|}\sum_{g\in\mathcal{G}} \left\|S_g - \bar{S}_g\right\|_1$$

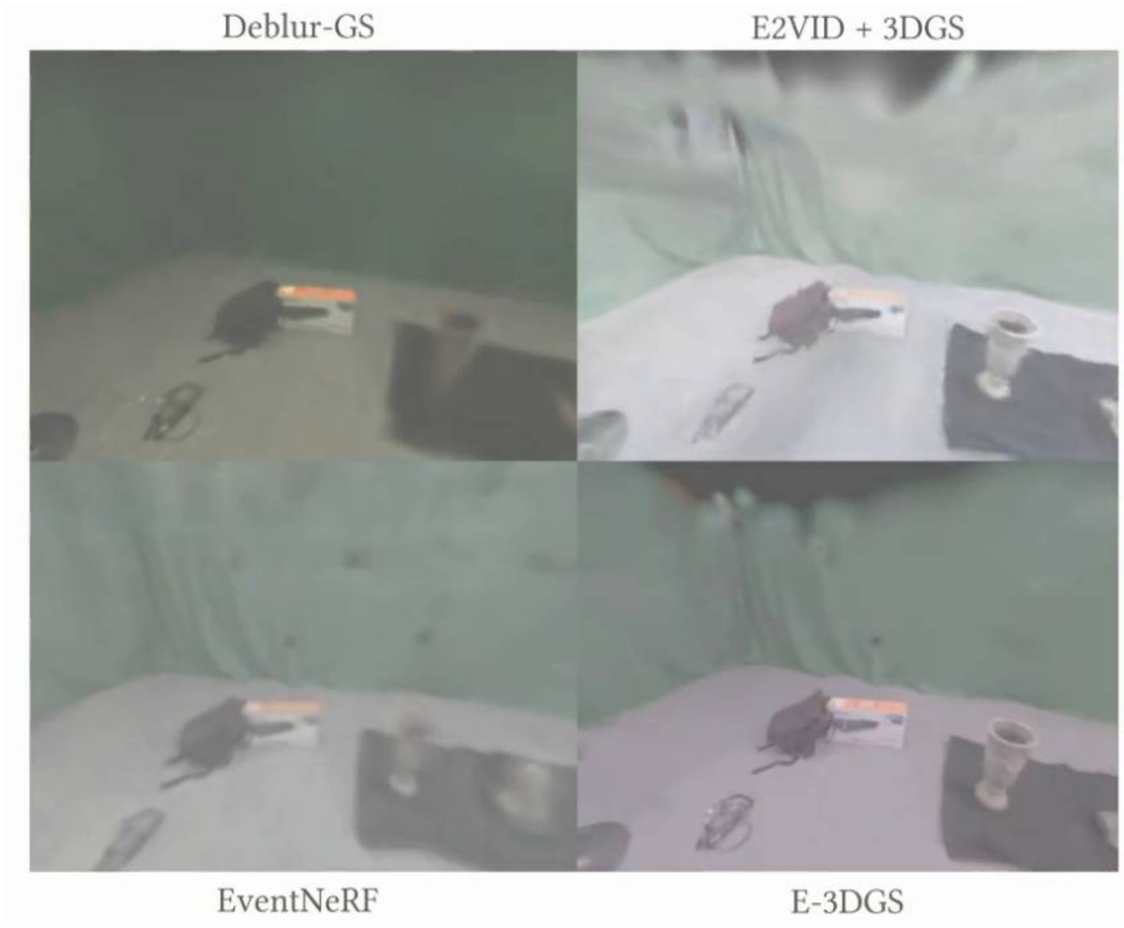
As-isotropic-as-possible regularisation

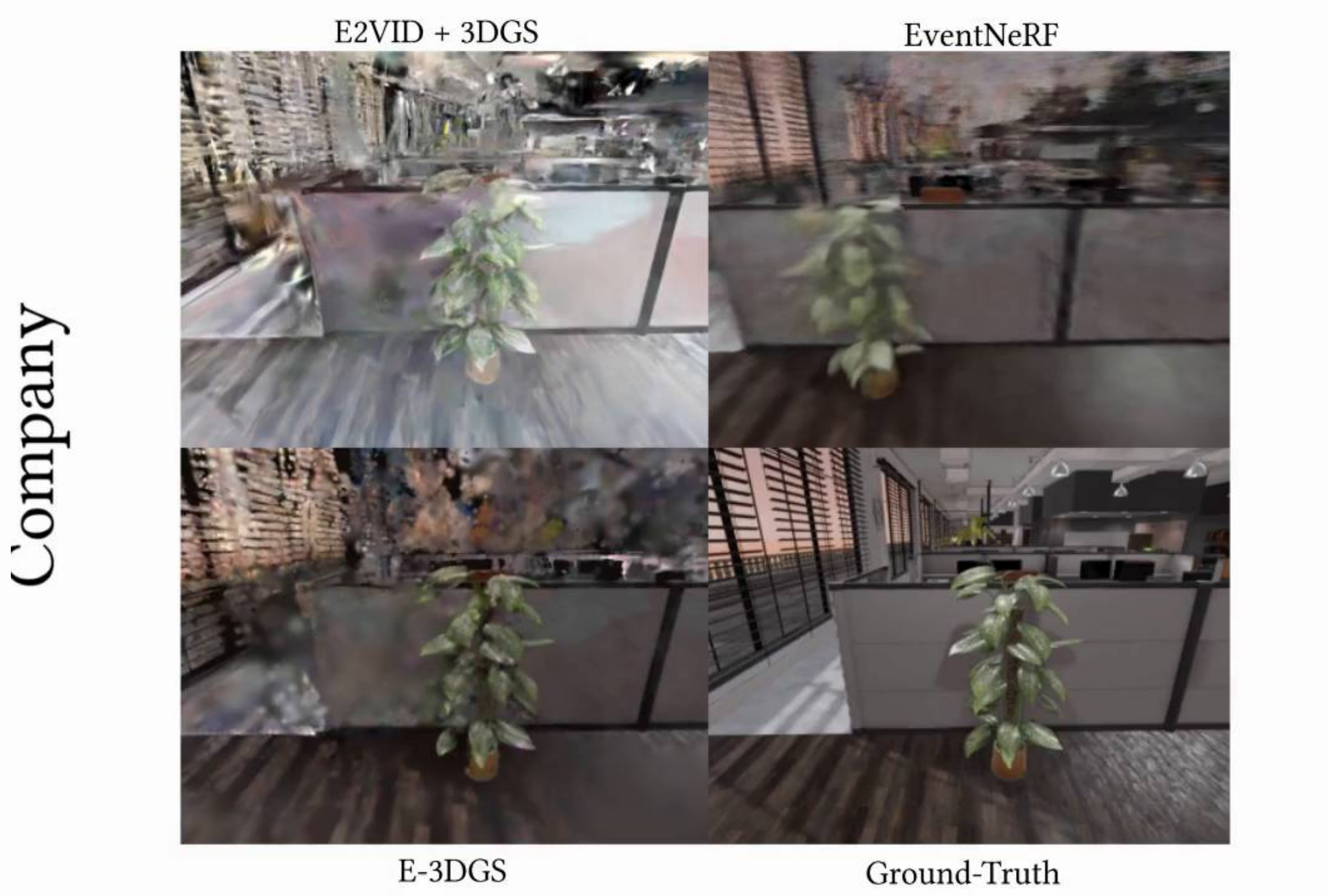
$$\mathcal{L} = \lambda_1 \mathcal{L}_{\text{recon}}(t_{s_1}, t) + \lambda_2 \mathcal{L}_{\text{recon}}(t_{s_2}, t) + \lambda_{\text{iso}}\mathcal{L}_{\text{iso}} + \lambda_{\text{pose}}\mathcal{L}_{\text{pose}}$$

The final loss (two different event windows)

E-3DGS: A New Scene Representation

MAX-PLANCK-INSTITUT
FÜR INFORMATIK

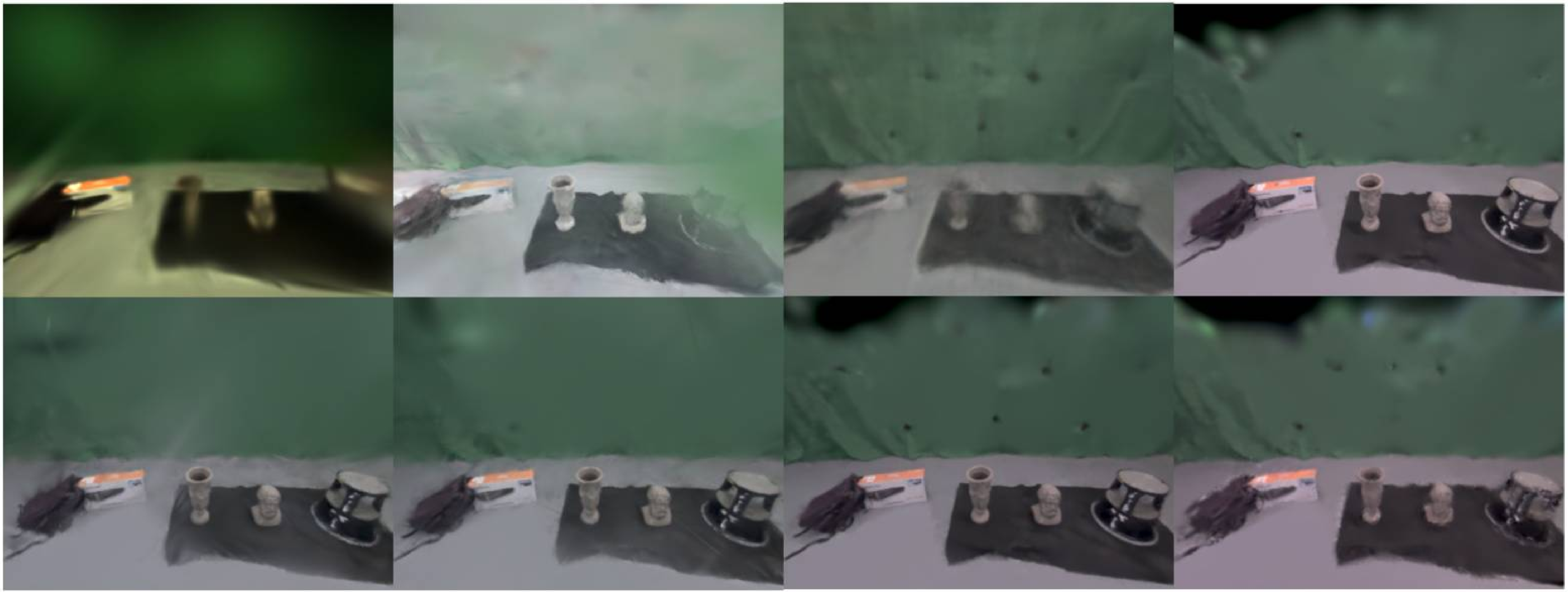




Zahid, Rudnev, Ilg, Golyanik. 3DV, 2025.

E-3DGS: A New Scene Representation

Deblur-GS | E2VID + 3DGS | EventNeRF | E-3DGS



w/o $\mathcal{L}_{\text{iso}}$ | w/o $\mathcal{L}_{\text{iso}}$ & $\mathcal{L}_{\text{pose}}$ | w/o AW (no adaptive event window) | w/o PR (no pose refinement)

Zahid, Rudnev, Ilg, Golyanik. 3DV, 2025.

Dynamic EventNeRF (This Workshop)

4D and Quantum Vision Group | SAARLAND UNIVERSITY | MAX-PLANCK-INSTITUT FÜR INFORMATIK

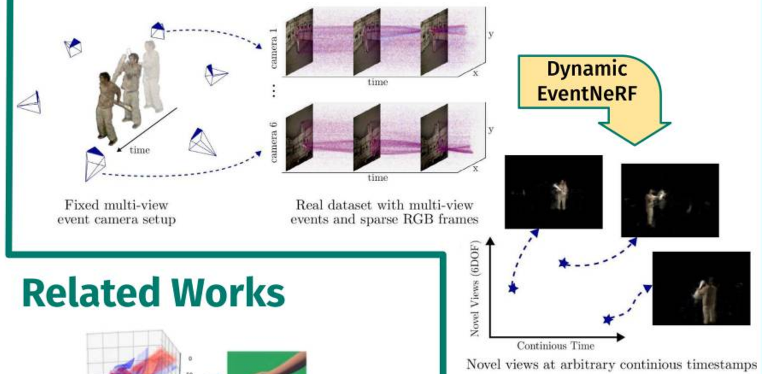
Dynamic EventNeRF: Reconstructing General Dynamic Scenes from Multi-view RGB and Event Streams

Viktor Rudnev Gereon Fox Mohamed Elgharib Christian Theobalt Vladislav Golyanik

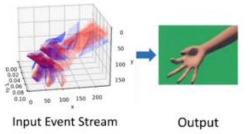
Event-based Vision 2025 CVPR Nashville June 11-15, 2025

Project Page, Code & Data
https://4dqv.mpi-inf.mpg.de/DynEventNeRF/

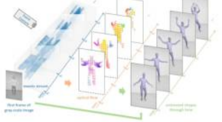
Using *multi-view event cameras* to reconstruct *general* dynamic scenes at *arbitrary time and viewpoints*



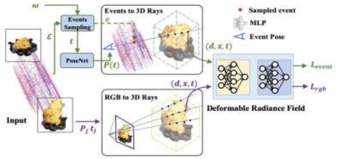
Related Works



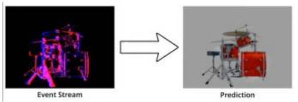
EventHands (Rudnev et al., 2021)



EventHPE (Zou et al., 2021)



DE-NeRF (Ma et al., 2023)

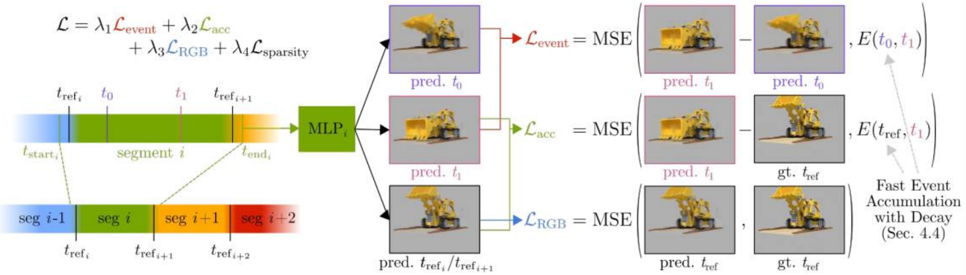


EventNeRF (Rudnev et al., 2023)

References: **[25]** Deformable Neural Radiance Fields using RGB and Event Cameras (Ma et al., 2023); **[37]** Fast Event-based Double Integral for Real-time Robotics (Lin et al., 2023); **[46]** High Speed and High Dynamic Range Video with an Event Camera (Rebecq et al., 2019); **[48]** EventHands: Real-time Neural 3D Hand Pose Estimation from an Event Stream (Rudnev et al., 2021); **[49]** EventNeRF: Neural Radiance Fields from a Single Colour Event Camera (Rudnev et al., 2023); **[52]** Non-rigid Neural Radiance Fields: Reconstruction and Novel View Synthesis of a Dynamic Scene from Monocular Video (Tretschk et al., 2021); **[61]** FreeNeRF: Improving Few-shot Neural Rendering with Free Frequency Regularization (Yang et al., 2023); **[65]** EventHPE: Event-based 3d human pose and shape estimation (Zou et al., 2021)

Method

$$\mathcal{L} = \lambda_1 \mathcal{L}_{\text{event}} + \lambda_2 \mathcal{L}_{\text{acc}} + \lambda_3 \mathcal{L}_{\text{RGB}} + \lambda_4 \mathcal{L}_{\text{sparsity}}$$



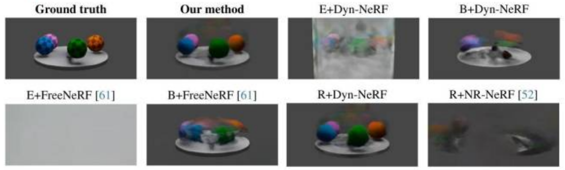
$$\mathcal{L}_{\text{event}} = \text{MSE}\left(\text{pred. } t_1 - \text{pred. } t_0,\ E(t_0, t_1)\right)$$

$$\mathcal{L}_{\text{acc}} = \text{MSE}\left(\text{pred. } t_1 - \text{gt. } t_{\text{ref}},\ E(t_{\text{ref}}, t_1)\right)$$

$$\mathcal{L}_{\text{RGB}} = \text{MSE}\left(\text{pred. } t_{\text{ref}},\ \text{gt. } t_{\text{ref}}\right)$$

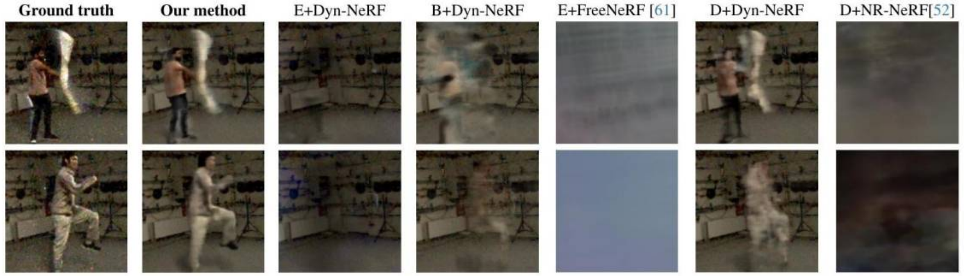
$$\mathcal{L}_{\text{sparsity}}(n) = \gamma_n \cdot \frac{1}{WH} \sum_{x,y=1}^{WH} \left(1 - T^{(x,y)}_{N_{\text{samples}}+1}\right) \qquad E_{x,y}(t_0, t_1) = \sum_{i=i_{\text{start}}}^{i_{\text{end}}} \delta^{x,y}_{x_i,y_i} p_i C_{p_i} b^{i_{\text{end}}-i_{\text{start}}-i},\ b = 0.93$$

Synthetic Results (E - E2VID [46], B - Blurry RGB, R - GT RGB [37])

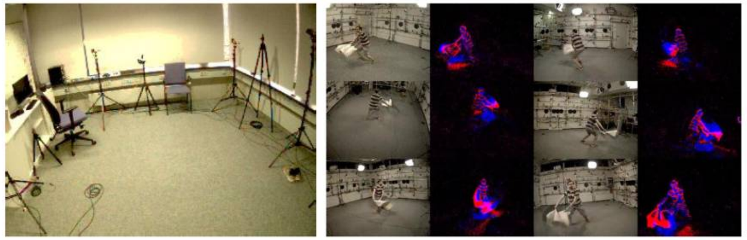


Method	PSNR↑ (Average)	SSIM↑ (Average)	LPIPS↓ (Average)
Our Dyn-EventNeRF	**26.99**	**0.89**	**0.15**
E2VID [46] + Dyn-NeRF	17.49	0.77	0.44
Blurry RGB + Dyn-NeRF	22.88	0.85	0.22
E2VID [46] + FreeNeRF [61]	15.87	0.77	0.40
Blurry RGB + FreeNeRF [61]	23.32	0.86	0.18
GT RGB [37] + Dyn-NeRF	25.80	0.88	0.18
GT RGB [37] + NR-NeRF [52]	18.89	0.79	0.42

Real Results (E - E2VID [46], B - Blurry RGB, D - EDI [37])

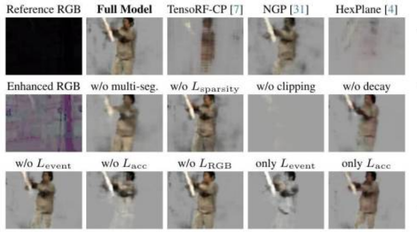


Real Multi-View Event Dataset

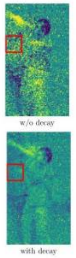


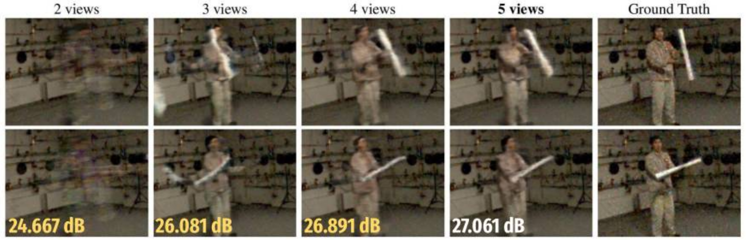
Over 18 minutes of 6 simultaneous event+RGB streams (DAVIS 346C), calibrated and deblurred at 5 FPS

Ablations and Design Choice Studies



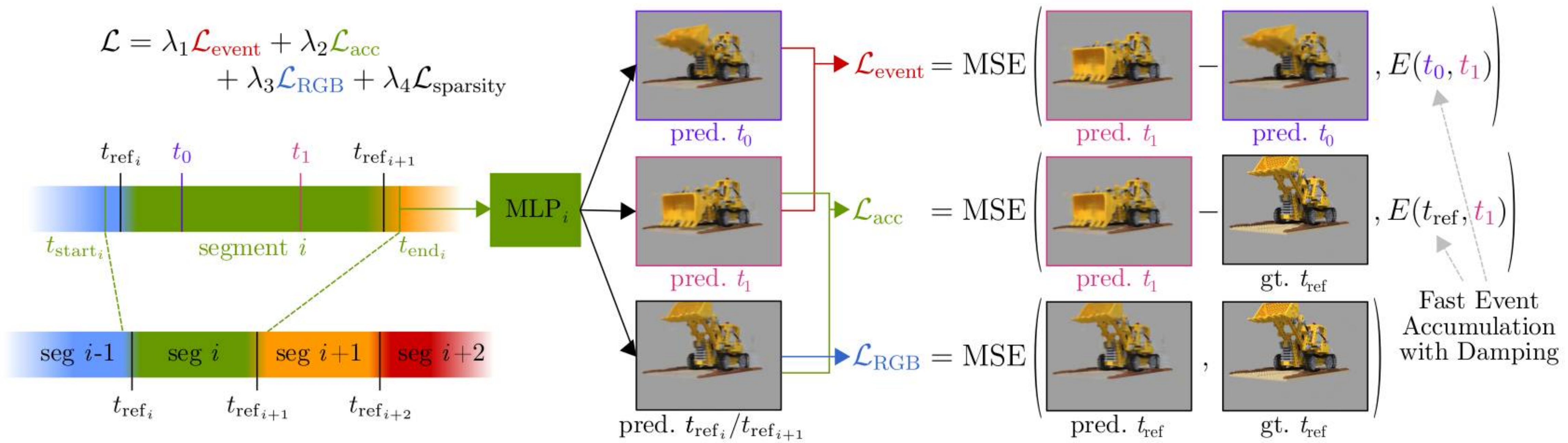
Method	PSNR↑
Ours (TensoRF-CP [7])	24.167
Ours (NGP [31])	23.239
Ours (HexPlane [4])	21.694
w/o clipping	24.471
w/o L_{event}	24.511
w/o L_{acc}	25.113
w/o L_{RGB}	26.593
only L_{event}	25.525
only L_{acc}	25.795
w/o L_{sparsity}	26.853
Our Full Model (Final)	**27.097**





Dynamic EventNeRF (This Workshop)

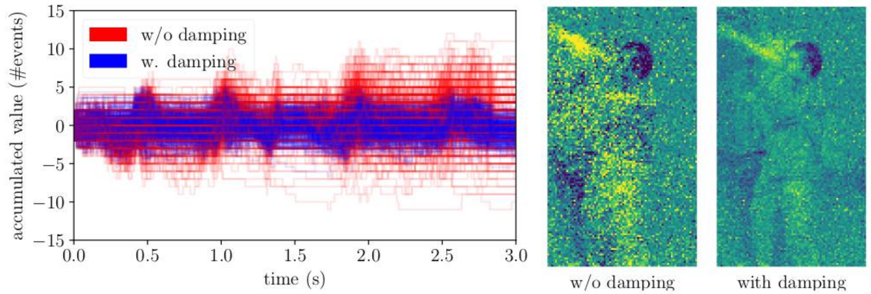
MAX-PLANCK-INSTITUT
FÜR INFORMATIK



Overview of Dynamic EventNeRF

$$\mathcal{L} = \lambda_1 \mathcal{L}_{\text{event}} + \lambda_2 \mathcal{L}_{\text{acc}} + \lambda_3 \mathcal{L}_{\text{RGB}} + \lambda_4 \mathcal{L}_{\text{sparsity}}$$

- $\mathcal{L}_{\text{event}}$: data term (event space)
- $\mathcal{L}_{\text{acc}}$: propagate RGB to all timestamps
- $\mathcal{L}_{\text{RGB}}$: model matches GT RGB
- $\mathcal{L}_{\text{sparsity}}$: reconstruct only foreground

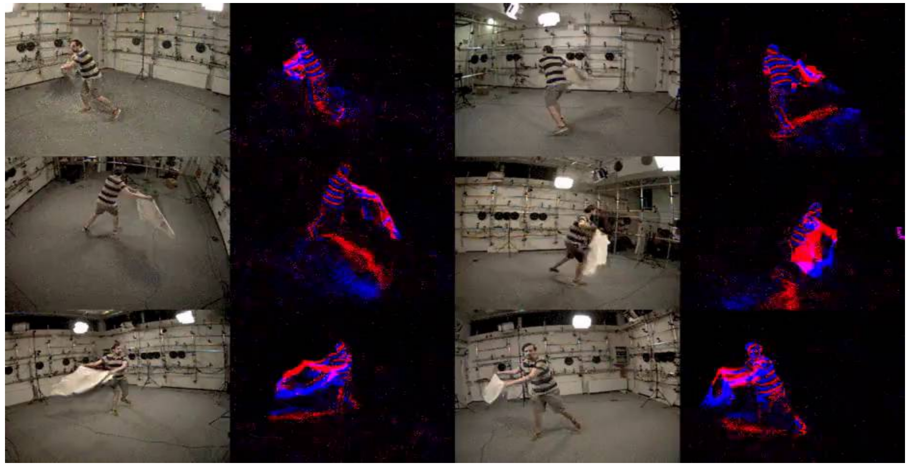


Demonstration of accumulation damping

Rudnev, Fox, Elgharib, Theobalt, Golyanik. CVPRW, 2025.

Dynamic EventNeRF

MAX-PLANCK-INSTITUT
FÜR INFORMATIK

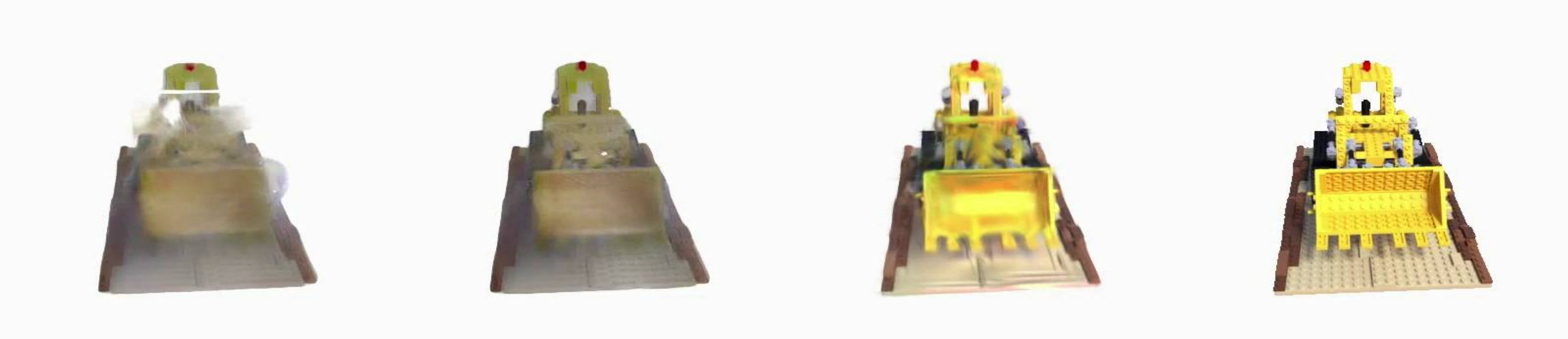


input multi-view event stream

novel-view renderings (different sequences; with zoom-in)

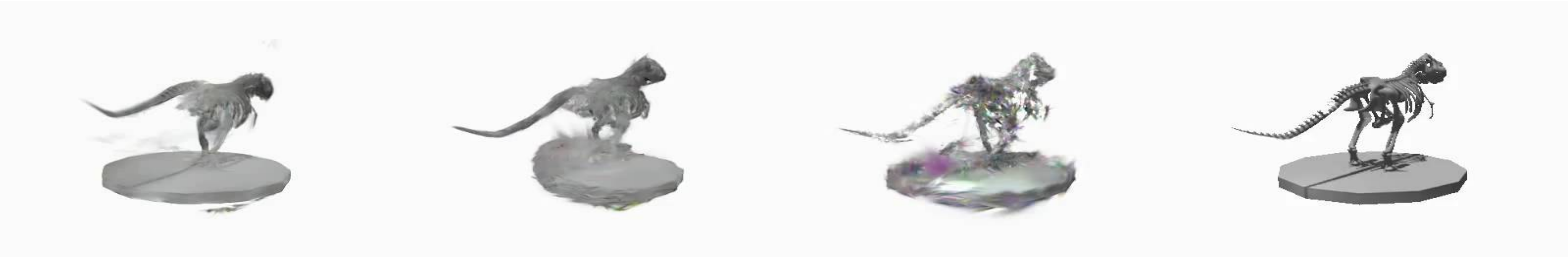
Rudnev, Fox, Elgharib, Theobalt, Golyanik. CVPRW, 2025.

Single Event Camera, A Dynamic Object



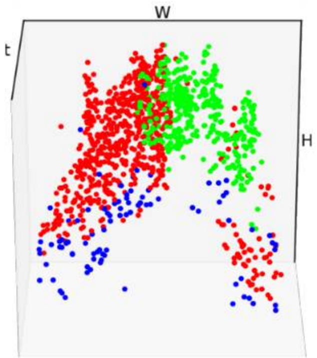
3DGS | D3DGS | ours | ground truth

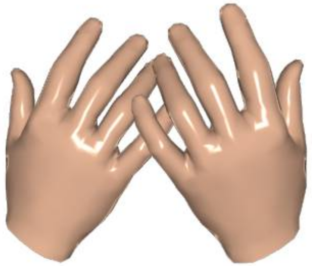
Single Event Camera, A Dynamic Object

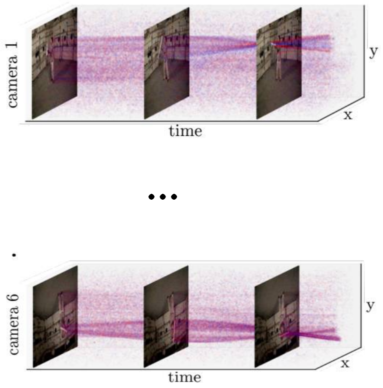


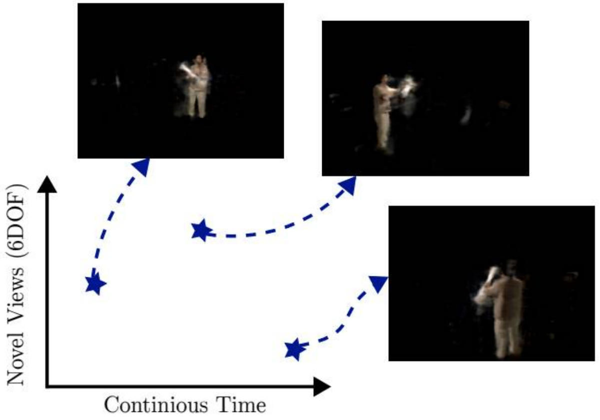
3DGS D3DGS ours ground truth

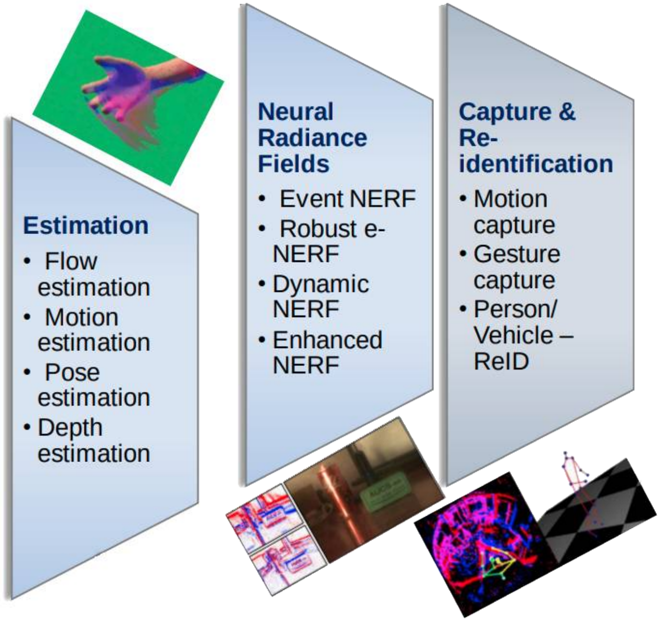
Take-home Messages











Images: Chakravarthi et al., ECCVW, 2024.

- Event-based vision has a strong potential in 4D tracking and reconstruction
  - **Fast-moving objects, fast capture, low lighting, mobile devices, etc.**
- Existing formulations cannot just be “borrowed” from RGB-based vision
  - The problems and approaches **have to be rethought**
- Generalisable techniques require **new datasets for training**
  - Examples: **3D human pose estimation and 3D hand tracking**

Thanks! Questions?