EventHands: Real-Time Neural 3D Hand Pose Estimation from an Event Stream

V. Rudnev[1] V. Golyanik[1] J. Wang[1] H.-P. Seidel[1] F. Mueller[2] M. Elgharib[1] C. Theobalt[1]

[1]MPI Informatics, Saarland Informatics Campus [2]Google Inc.

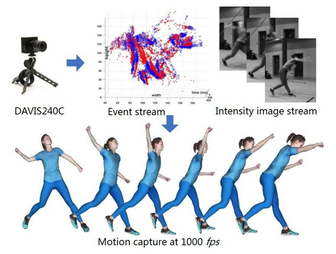
max planck institut informatik
SIC Saarland Informatics Campus
4D and Quantum Vision Group
Visual Computing and AI Department
Google

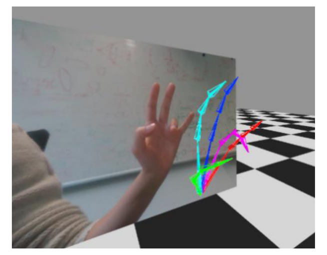
2021 ICCV OCTOBER 11-17 VIRTUAL

Motivation

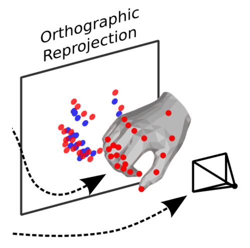
- Hand pose estimation is important for Interations in Virtual Environments, Gesture Recognition, Gaming and more.
- Current RGB(D) methods have their limitations:
  - Fail on fast moving hands due to motion blur,
  - Fail in low-light conditions due to sensor sensitivity,
  - Consume large data bandwidth

Related Work

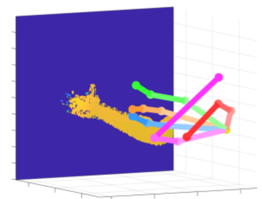

EventCap


Mueller et al.

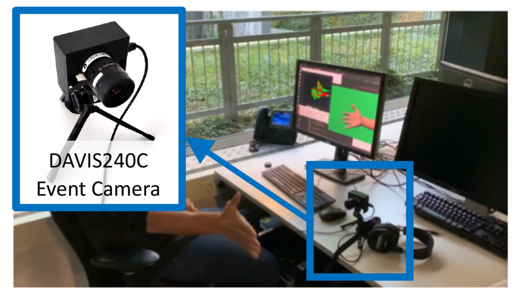

Zhou et al.



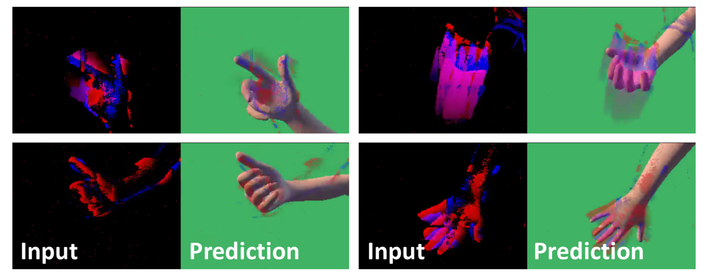
Boukhayma et al.


Depth and Output Pose
Moon et al. (Depth-based)

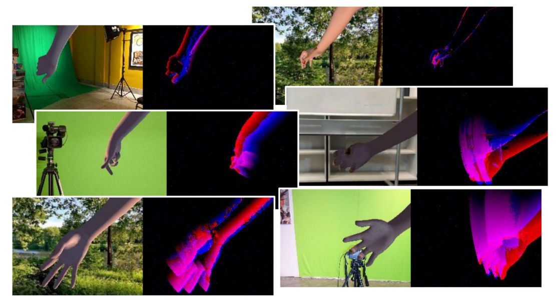
Contributions



Live Demo



Hand Pose Prediction

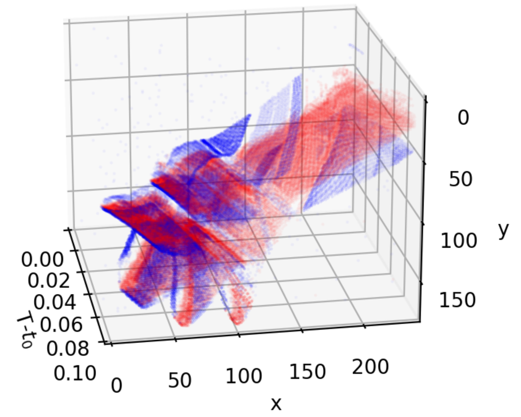

Large-Scale Dataset

- Live 3D Hand Pose Estimation at 1 KHz from a Single Event Stream
- New High-Throughput Event Stream Simulator for Hands
- Large-Scale Annotated Dataset of Synthetic Hand Event Streams

Event Cameras

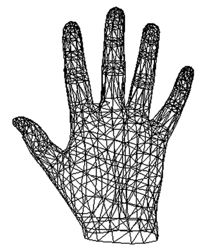
- React to brightness changes (events) asynchronously per-pixel, instead of shooting full frames,
- Use abstract data representation useful for generalization,
- Have low data bandwidth, essentially infinite temporal resolution and high dynamic range

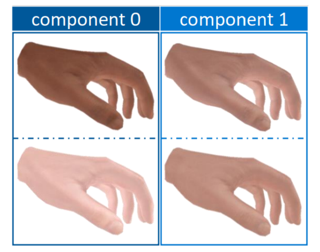

Event camera


Example event stream
(red is positive, blue is negative)

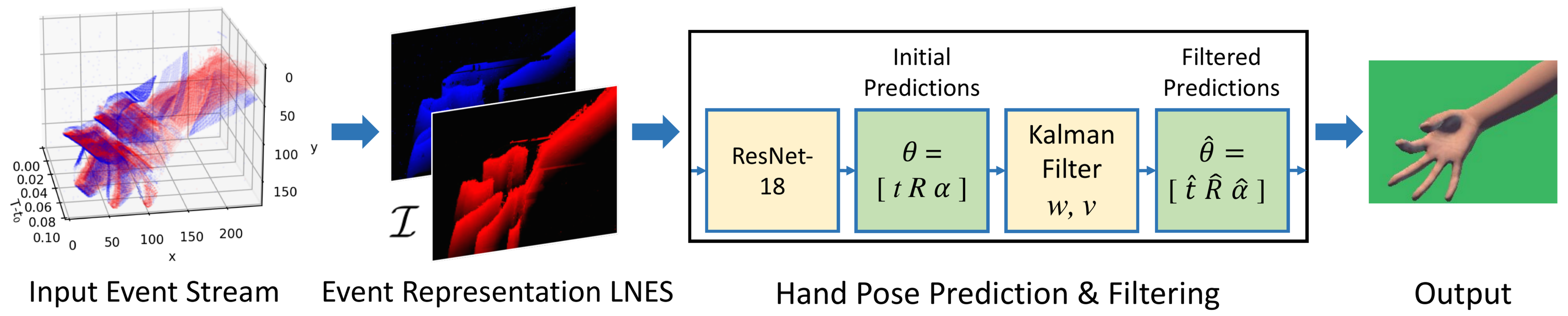
$$\begin{cases} \mathcal{L}(u_i,t_i) - \mathcal{L}(u_i,t_p) = C & (p = 1) \\ \mathcal{L}(u_i,t_i) - \mathcal{L}(u_i,t_p) = -C & (p = -1) \end{cases}$$

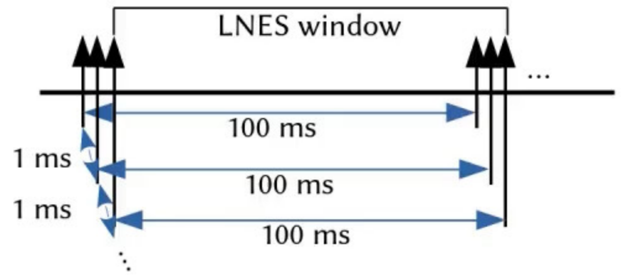
MANO and HTML models





Approach





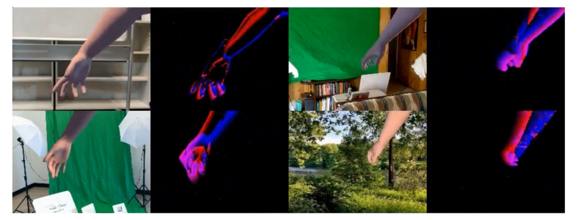
LNES Definition: $\mathcal{I}(x_i, y_i, p_i) = \frac{t_i - t_0}{L}$

α hand pose (six PCA coeff.)

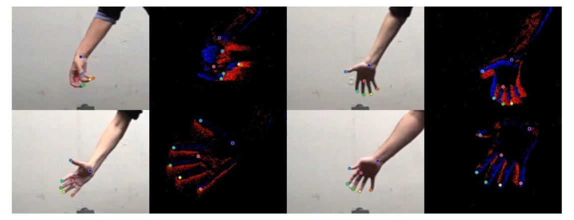
t, R translation and rotation (axis-angle)

Datasets

synthetic training set

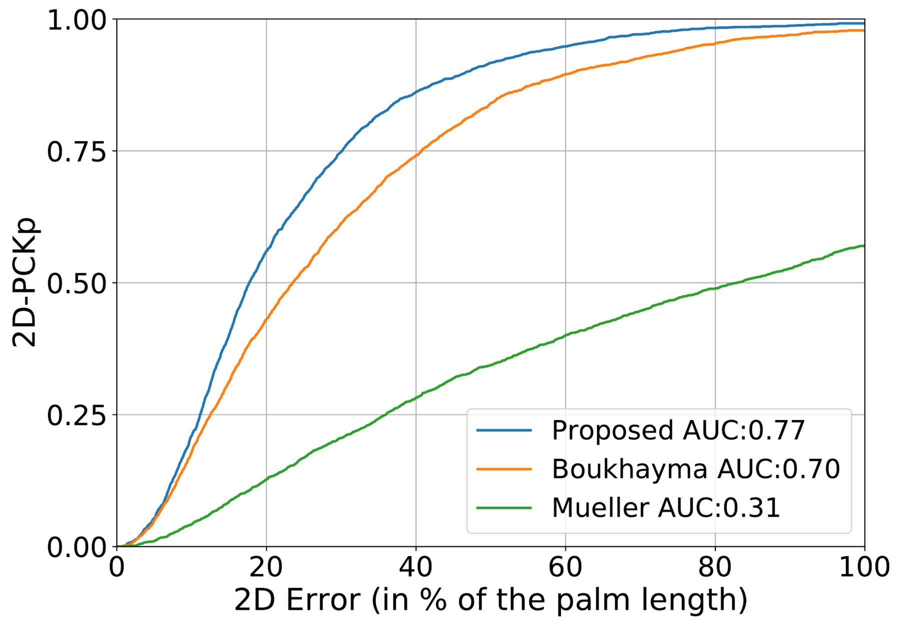


real test set



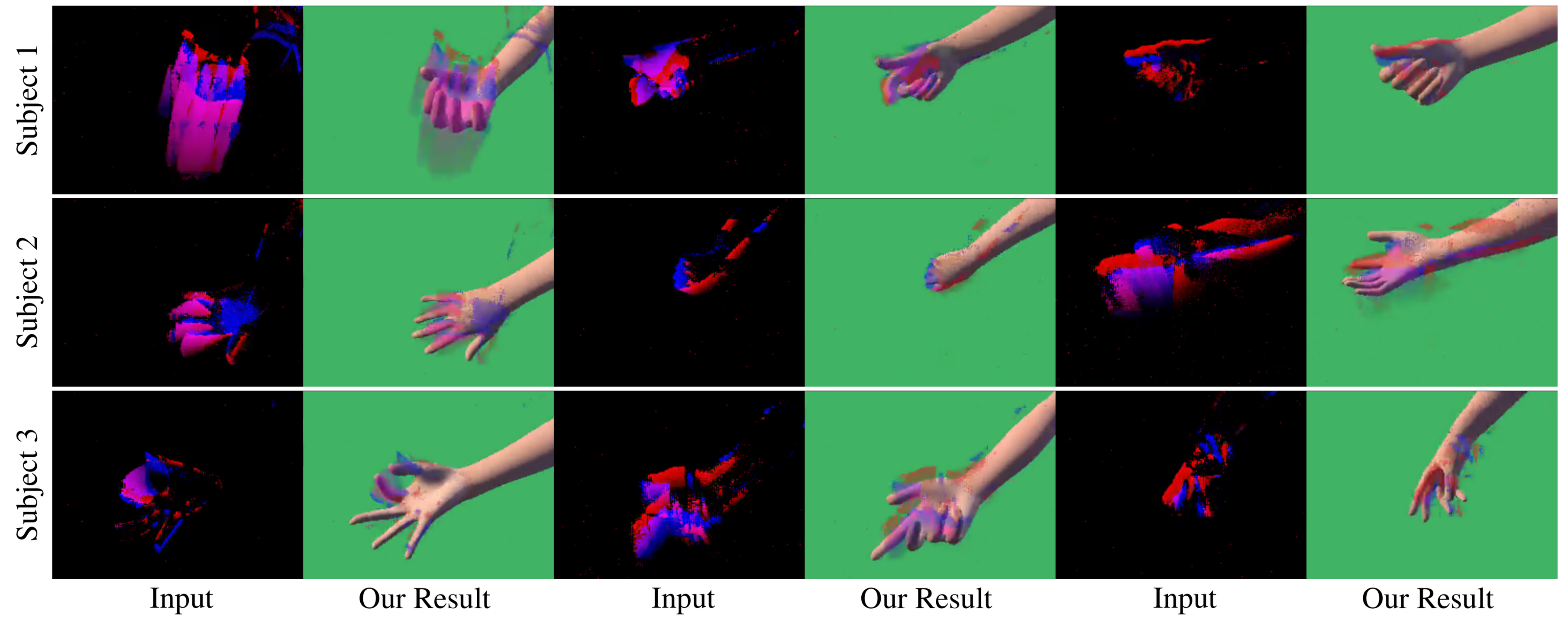
Comparison to RGB Methods and Ablation Study

		synthetic		real
		2D-AUCp	**3D-AUC**	**2D-AUCp**
	no filtering	**0.89**	***0.85***	***0.75***
	no aug.	***0.88***	**0.86**	0.70
EOI	33ms	0.86	***0.85***	0.70
	100ms	0.78	0.80	0.56
ECI-S	33ms	0.83	0.81	0.66
	100ms	0.69	0.76	0.56
ECI	33ms	0.86	0.83	0.69
	100ms	0.76	0.79	0.52
LNES	33ms	***0.88***	***0.85***	0.72
	300ms	0.87	0.84	0.72
	proposed	***0.88***	***0.85***	**0.77**

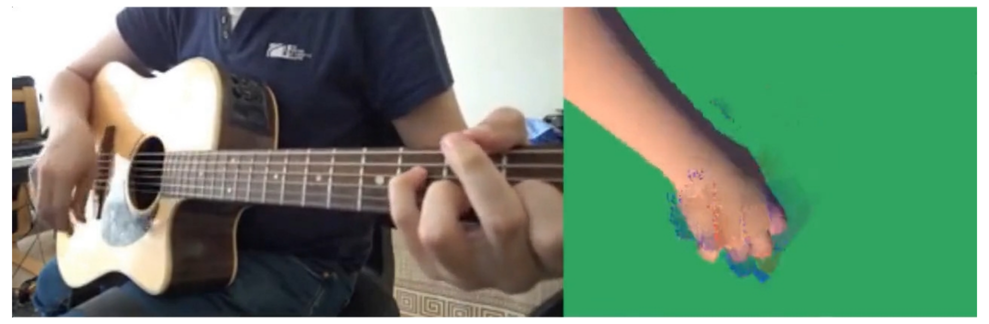


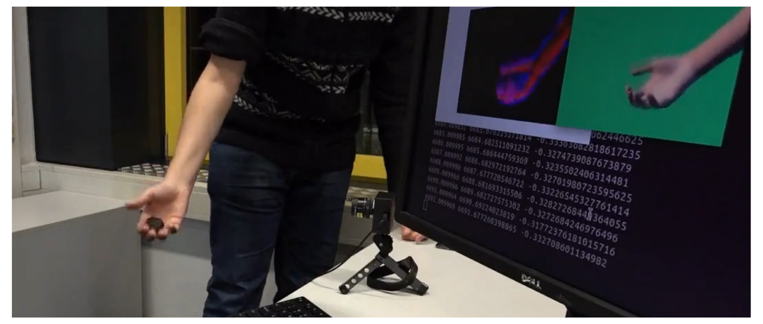

Reference (500 fps) RGB Input (30 fps) Mueller *et al.* [36] Boukhayma *et al.* [10] Zhou et al. [71] *EventHands* (Ours)

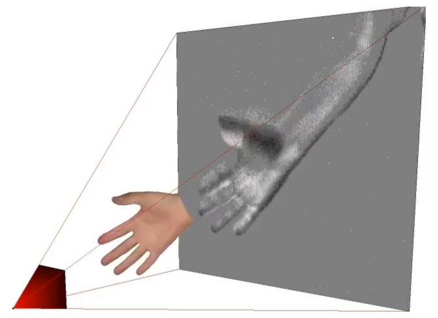
Various Results



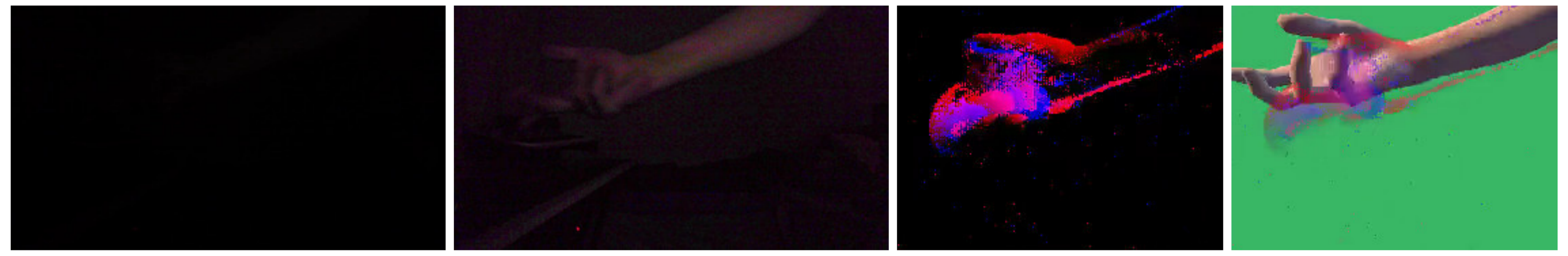
Input Our Result Input Our Result Input Our Result







Low-Light Performance (0.77 2D-PCKp AUC)


Original RGB (ISO 320) Brightened RGB LNES Input Result

References
1. Lan Xu, Weipeng Xu, Vladislav Golyanik, Marc Habermann, Lu Fang, and Christian Theobalt. EventCap: Monocular 3D capture of High-Speed Human Motions Using an Event Camera. *CVPR 2020*
2. Javier Romero, Dimitrios Tzionas, and Michael J. Black. Embodied Hands: Modeling and Capturing Hands and Bodies Together. *SIGGRAPH Asia 2017*
3. Neng Qian, Jiayi Wang, Franziska Mueller, Florian Bernard, Vladislav Golyanik, and Christian Theobalt. HTML: A Parametric Hand Texture Model for 3D Hand Reconstruction and Personalization. *ECCV 2020*
4. Jonathan Taylor, Vladimir Tankovich, Danhang Tang, Cem Keskin, David Kim, Philip Davidson, Adarsh Kowdle, and Shahram Izadi. Articulated Distance Fields for Ultra-Fast Tracking of Hands Interacting. *TOG 2017*
5. Guillermo Gallego, Tobi Delbruck, Garrick Michael Orchard, Chiara Bartolozzi, Brian Taba, Andrea Censi, Stefan Leutenegger, Andrew Davison, Jorg Conradt, Kostas Daniilidis, and Davide Scaramuzza. Event-Based Vision: A Survey. *IEEE PAMI 2020*
6. Adnane Boukhayma, Rodrigo de Bem, and Philip H.S. Torr. 3D Hand Shape and Pose from Images in the Wild. *CVPR 2019*
7. Franziska Mueller, Florian Bernard, Oleksandr Sotnychenko, Dushyant Mehta, Srinath Sridhar, Dan Casas, and Christian Theobalt. GANerated Hands for Real-Time 3D Hand Tracking from Monocular RGB. *CVPR 2018*

Acknowledgements This work was supported by the ERC Consolidator Grant 4DRepLy (770784). We thank Jalees Nahvi and Navami Kairanda for help with comparisons

4dqv.mpi-inf.mpg.de/EventHands