Learning Depth with Event Cameras

Prof. Dr. Guillermo Gallego

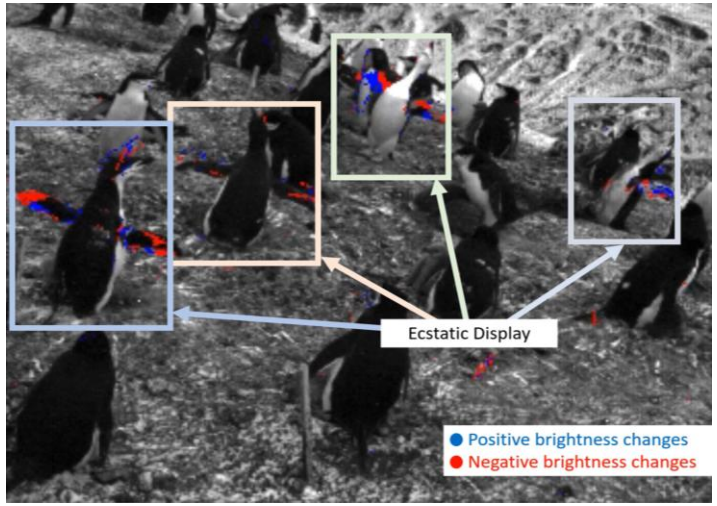
MLaftermath Workshop @ ECDF

with contributions from Suman Ghosh and Diego Hitzges

09.03.2026

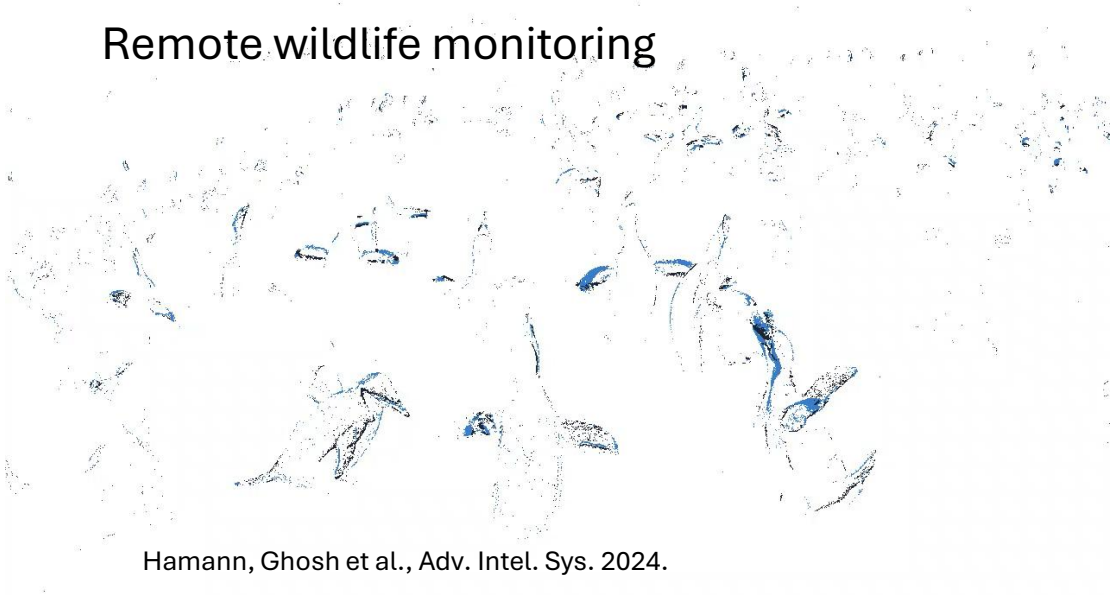
Live Demo

Motion and scene perception with event cameras



Hamann, Ghosh et al., CVPR 2024.

Remote wildlife monitoring

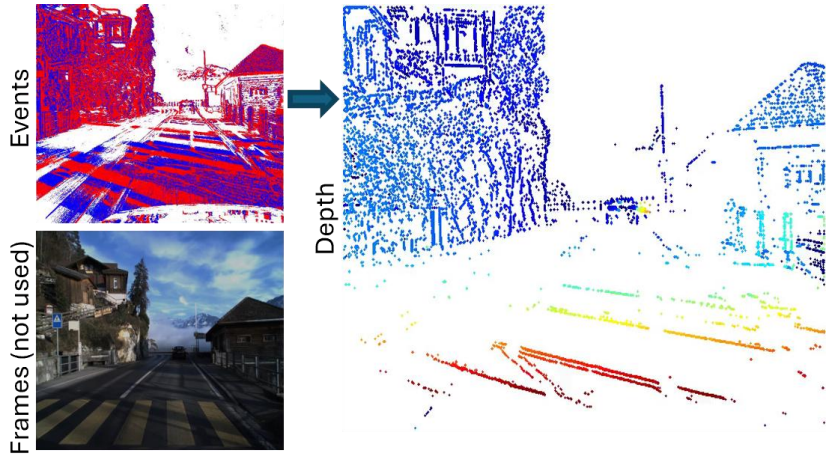


Hamann, Ghosh et al., Adv. Intel. Sys. 2024.

Slip detection during manipulation

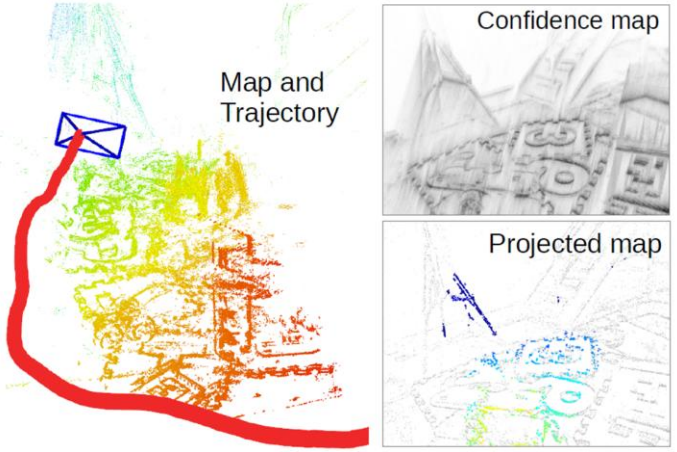
Reinold, Ghosh, Gallego, WACVW 2025.

Stereo depth estimation



Ghosh, Gallego, Adv. Intel. Sys. 2022.
Hitzges*, Ghosh*, Gallego, NeurIPS 2025.

SLAM



Ghosh, Cavinato, Gallego, ECCVW 2024.

Event-Based Stereo Depth Estimation: A Survey

Suman Ghosh and Guillermo Gallego

***Abstract*—Stereopsis has widespread appeal in computer vision and robotics as it is the predominant way by which we perceive depth to navigate our 3D world. Event cameras are novel bio-inspired sensors that detect per-pixel brightness changes asynchronously, with very high temporal resolution and high dynamic range, enabling machine perception in high-speed motion and broad illumination conditions. The high temporal precision also benefits stereo matching, making disparity (depth) estimation a popular research area for event cameras ever since their inception. Over the last 30 years, the field has evolved rapidly, from low-latency, low-power circuit design to current deep learning (DL) approaches driven by the computer vision community. The bibliography is vast and difficult to navigate for non-experts due its highly interdisciplinary nature. Past surveys have addressed distinct aspects of this topic, in the context of applications, or focusing only on a specific class of techniques, but have overlooked stereo datasets. This survey provides a comprehensive overview, covering both instantaneous stereo and long-term methods suitable for simultaneous localization and mapping (SLAM), along with theoretical and empirical comparisons. It is the first to extensively**

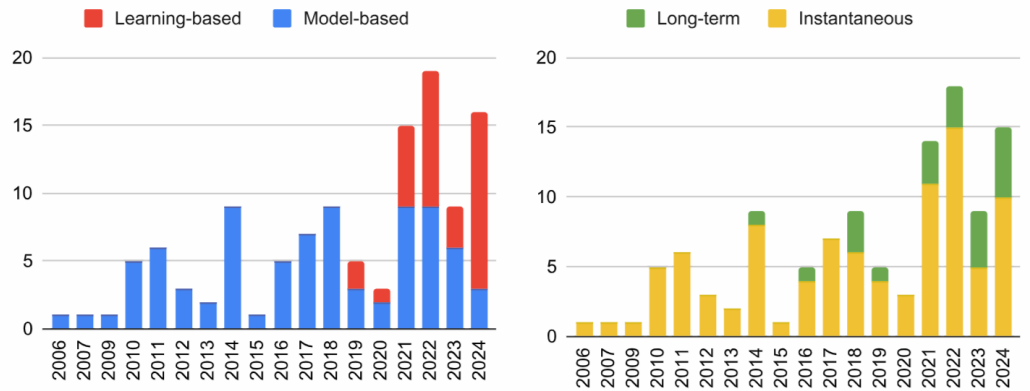
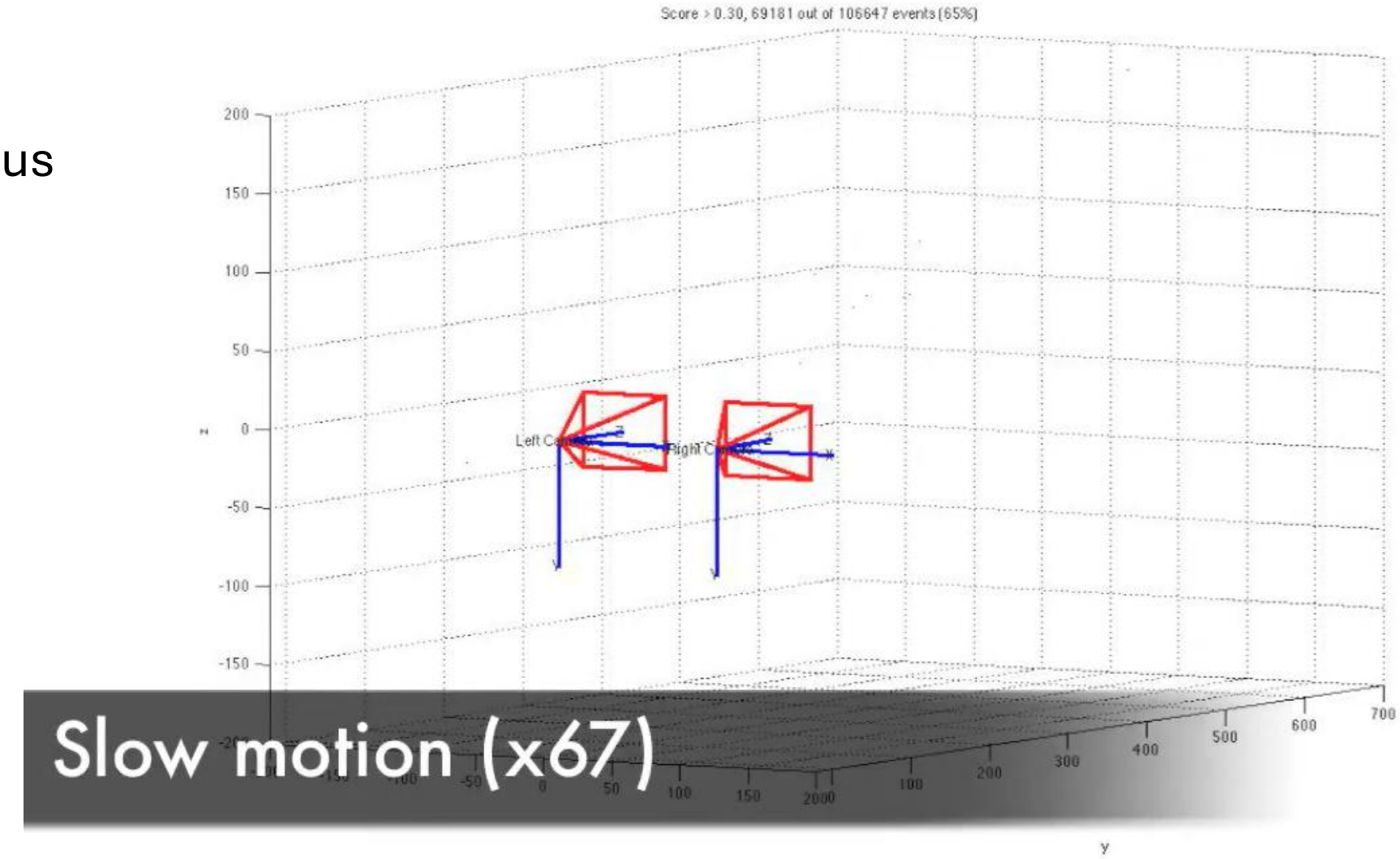


Fig. 1. Publications on event-based stereo depth estimation in the last two decades, classified according to criteria #1: whether they are model-based (i.e., hand-crafted) or learning-based (i.e., data-driven) methods (left), and #2: whether they produce instantaneous depth outputs or have long-term motion consistency (right). Plots created from a compiled spreadsheet of growing number of papers.

interest in computer vision and robotics because it tries to mimic the same functionality of the human brain (i.e., inverting

Stereo Depth Estimation

Methods
- Instantaneous
- Long-term



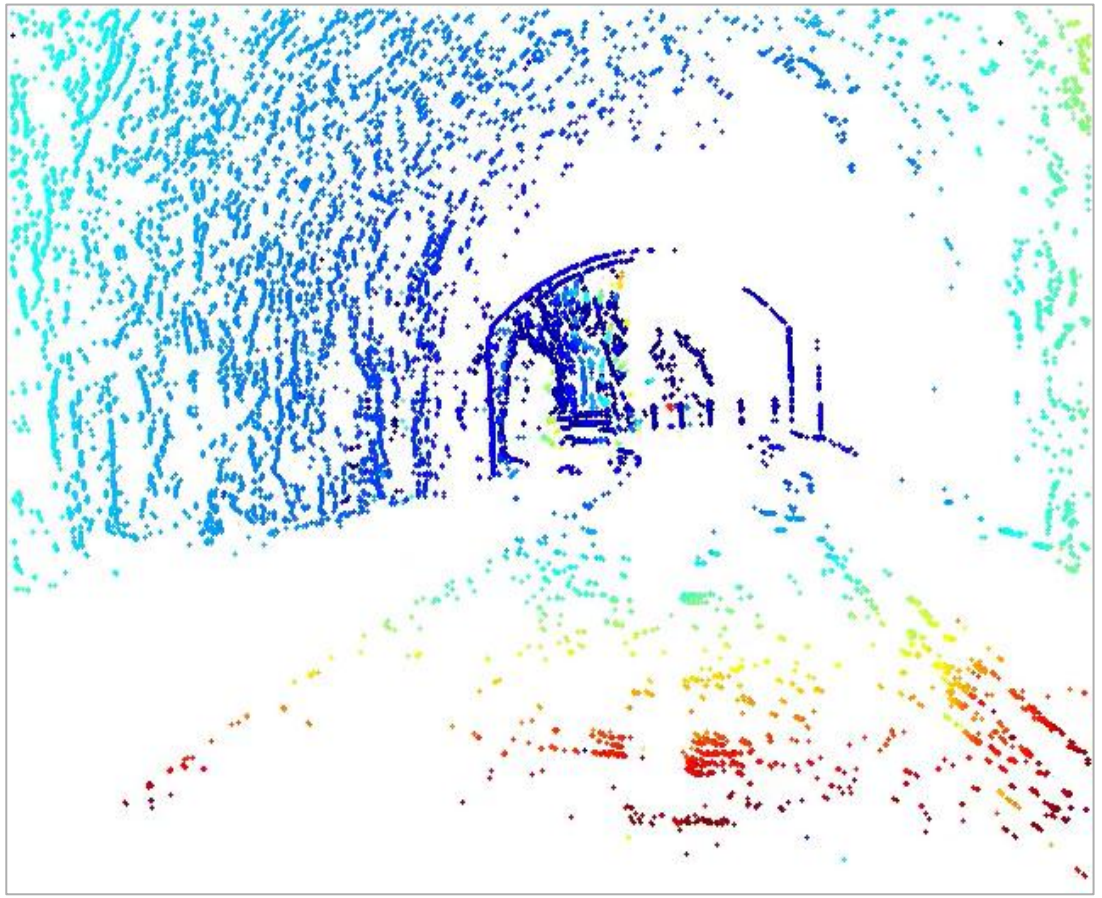
Carneiro et al., Event-based 3D reconstruction from neuromorphic retinas, Neural Networks 2013.

Stereo Depth Estimation

Methods
- Instantaneous
- Long-term

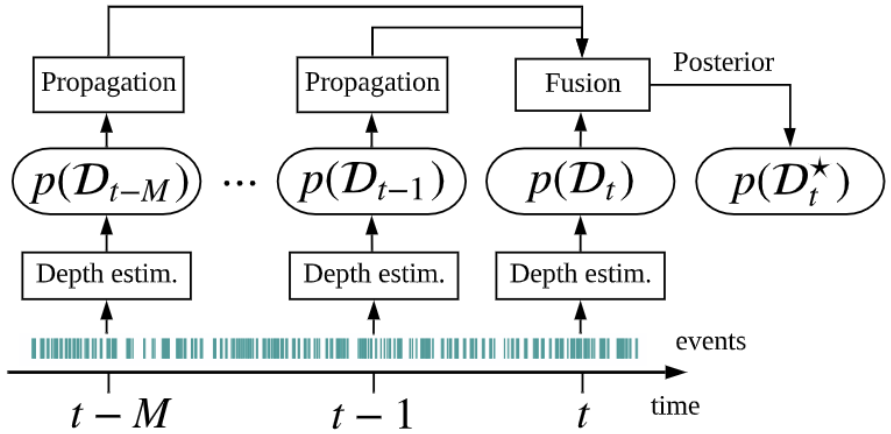
Stereo Events

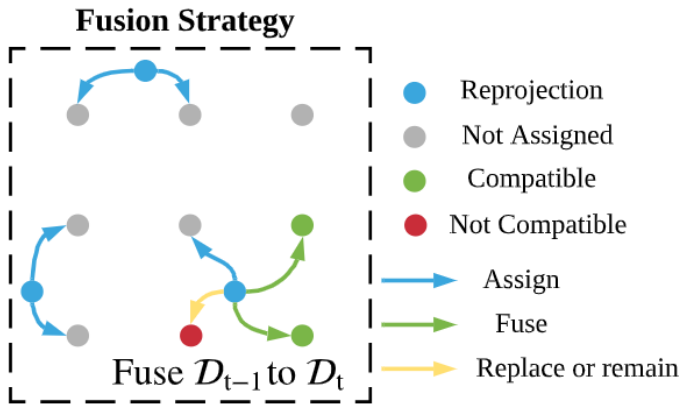
Depth



Stereo Depth Estimation

- Methods
  - Instantaneous
  - Long-term
    - Late fusion
    - Early fusion



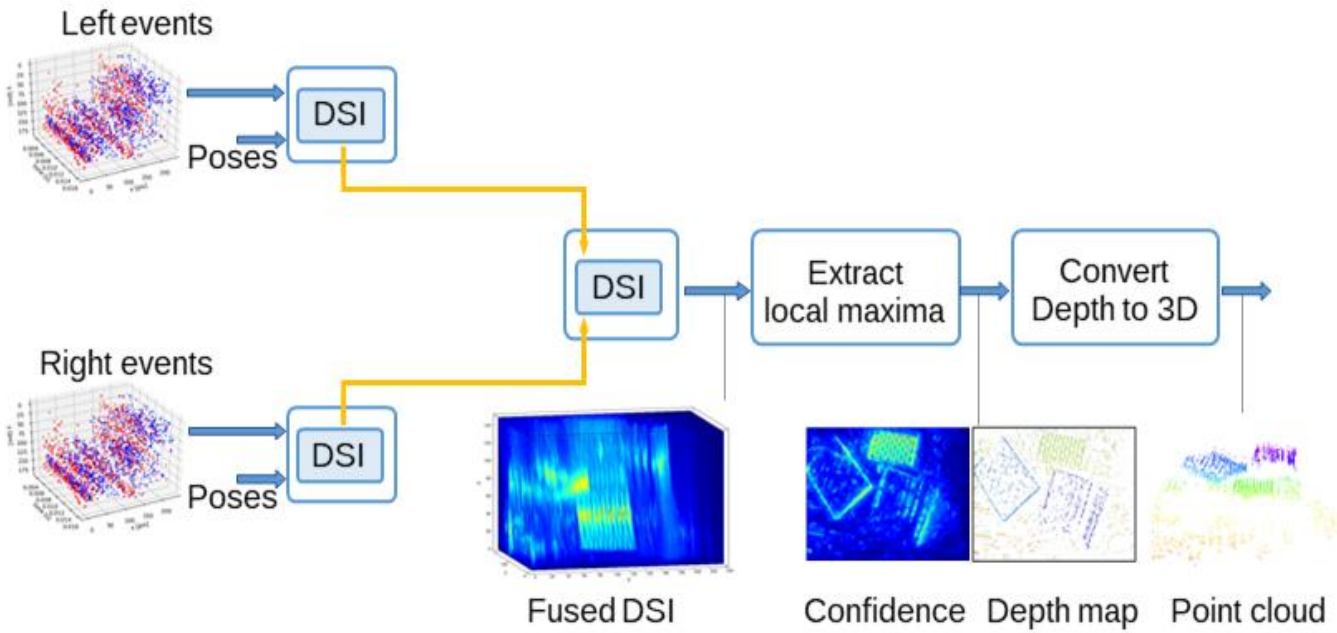


Zhou, Y., Gallego, G., Shen S., *Event-based Stereo Visual Odometry*, IEEE T-RO. 2021.

Stereo Depth Estimation

- Methods
  - Instantaneous
  - Long-term
    - Late fusion
    - Early fusion

No explicit event matching across cameras is needed!

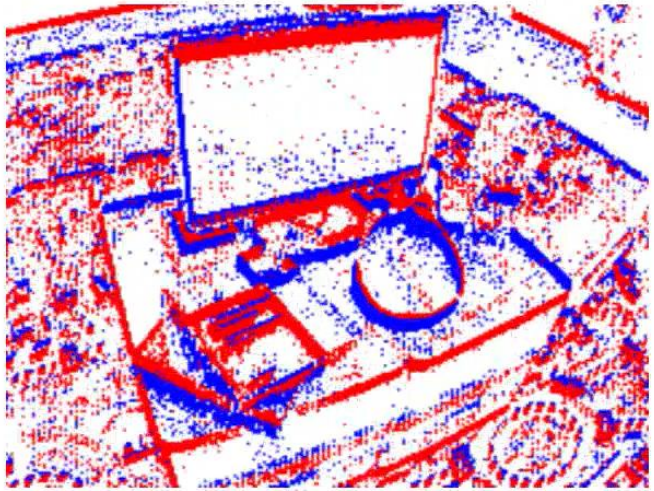


Ghosh, S., Gallego, G., *Multi-Event-Camera Depth Estimation and Outlier Rejection by Refocused Events Fusion*, Adv. Intell. Sys. 2022.

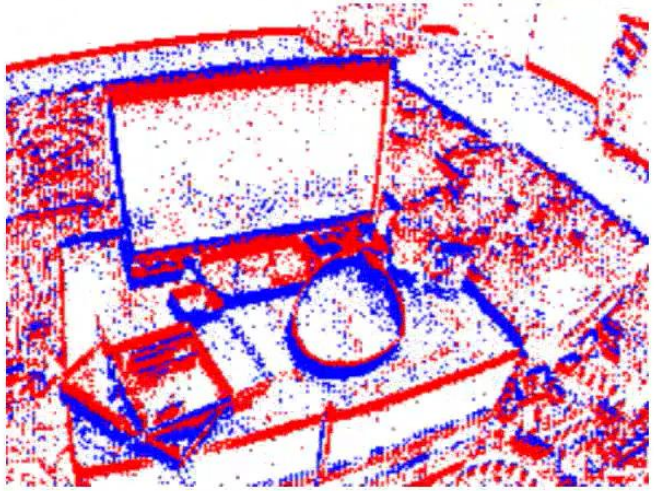
Multi-Camera Event-based Multi-View Stereo (MC-EMVS)

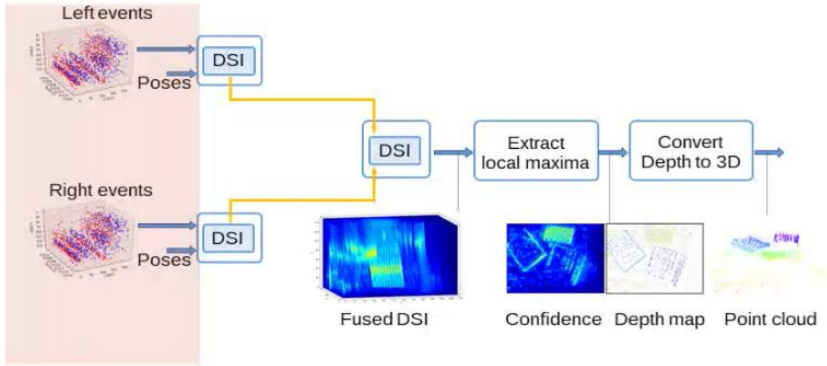
Input: Events & Camera poses

Left events



Right events

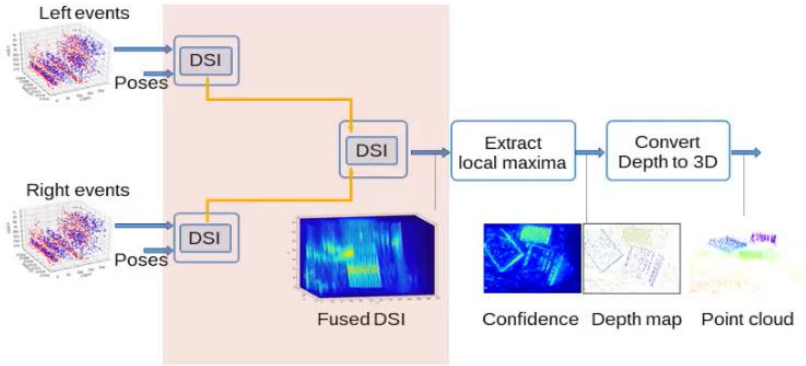




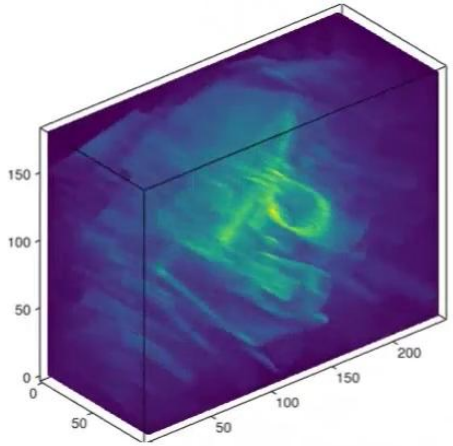
Ghosh, S., Gallego, G., *Multi-Event-Camera Depth Estimation and Outlier Rejection by Refocused Events Fusion*, Adv. Intell. Sys. 2022.

Multi-Camera Event-based Multi-View Stereo (MC-EMVS)

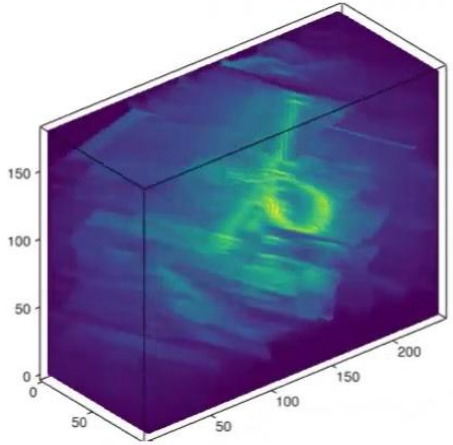
We build and fuse
event-ray densities (DSIs)



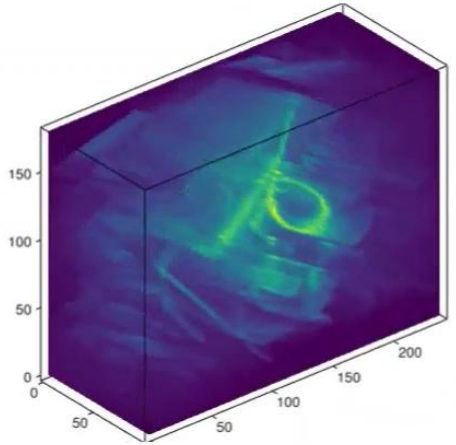
Left
camera



Right
camera



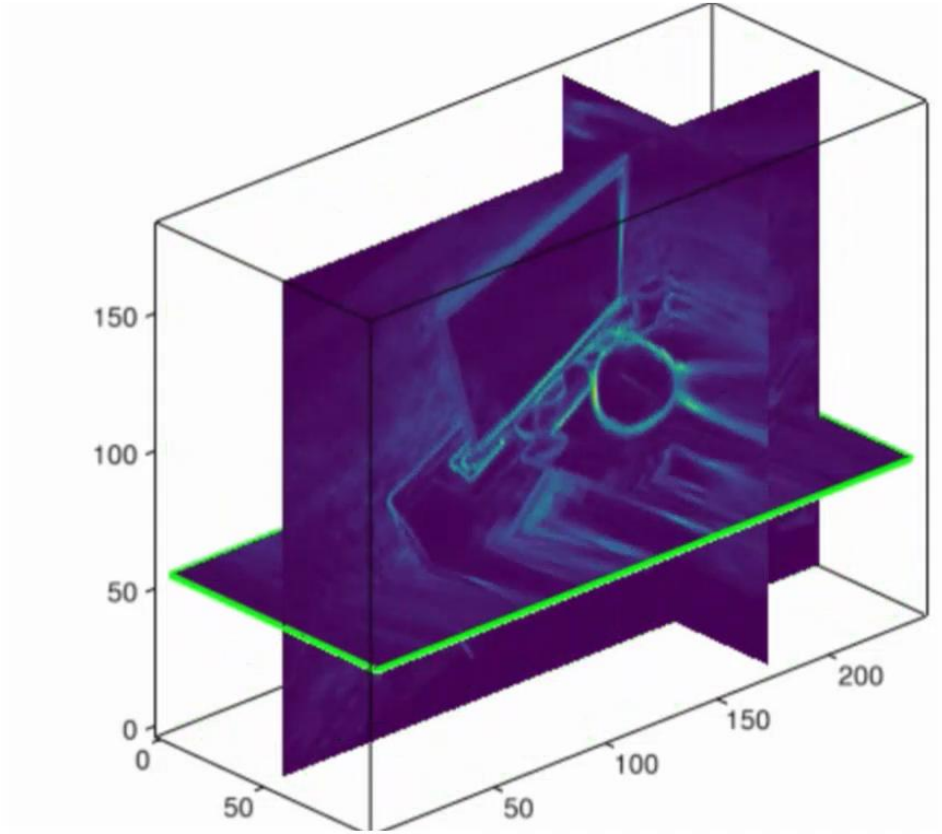
Fused
DSI

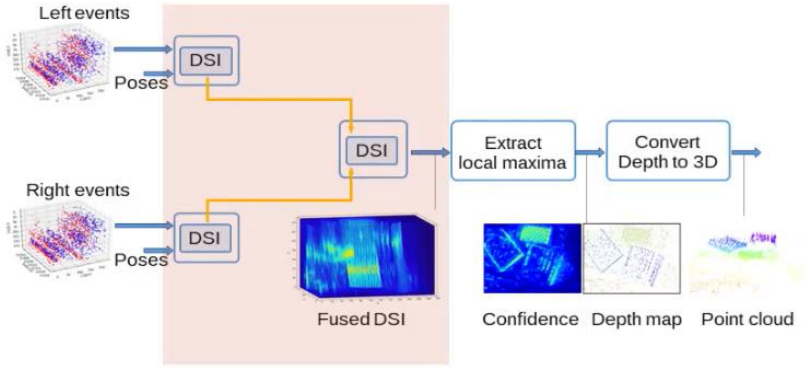


Ghosh, S., Gallego, G., *Multi-Event-Camera Depth Estimation and Outlier Rejection by Refocused Events Fusion*, Adv. Intell. Sys. 2022.

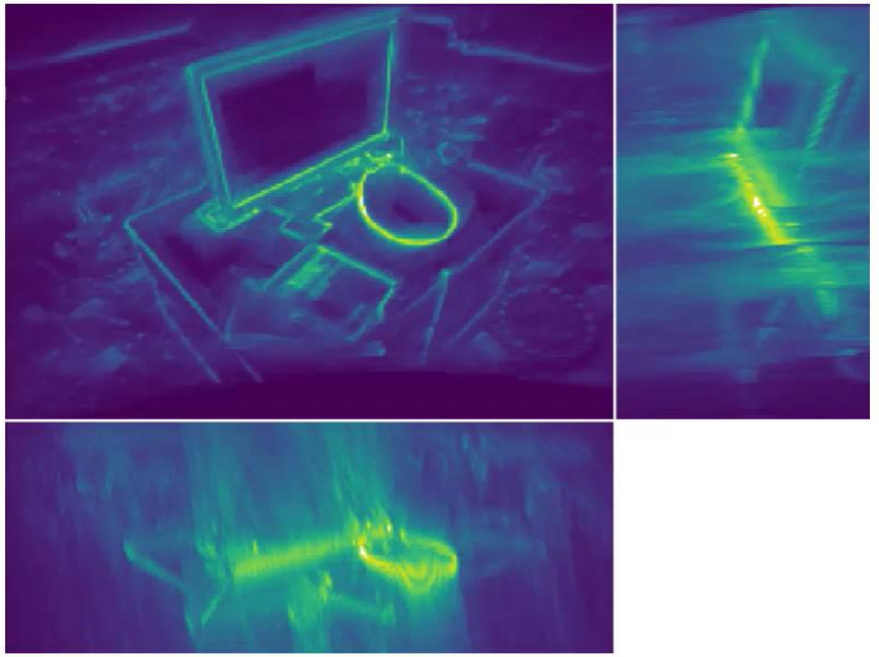
Multi-Camera Event-based Multi-View Stereo (MC-EMVS)

Fused DSI (combined ray density)





max along each axis



Ghosh, S., Gallego, G., *Multi-Event-Camera Depth Estimation and Outlier Rejection by Refocused Events Fusion*, Adv. Intell. Sys. 2022.

MC-EMVS

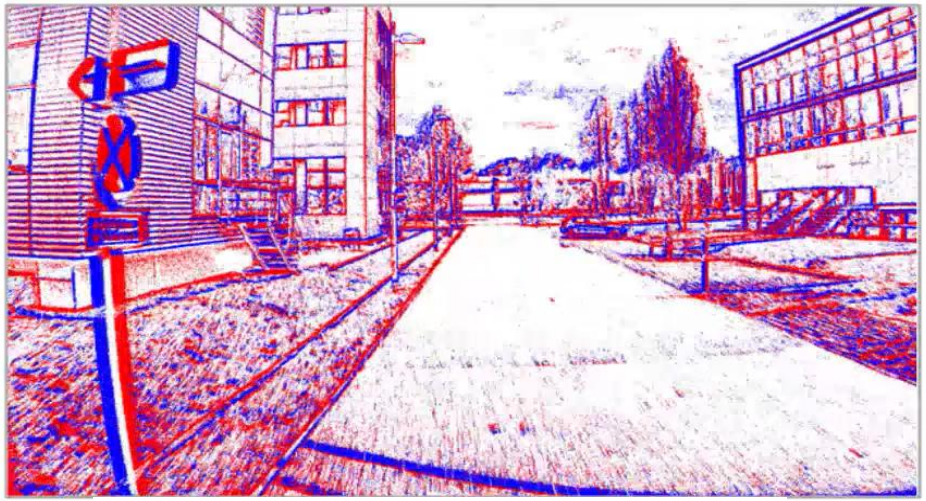
Results on 1Mpx (Prophesee Gen 4 HD) stereo dataset TUM-VIE

Klenk et al., TUM-VIE: The TUM Stereo Visual-Inertial Event Dataset, IROS 2021.

Ghosh, S., Gallego, G., *Multi-Event-Camera Depth Estimation and Outlier Rejection by Refocused Events Fusion*, Adv. Intell. Sys. 2022.

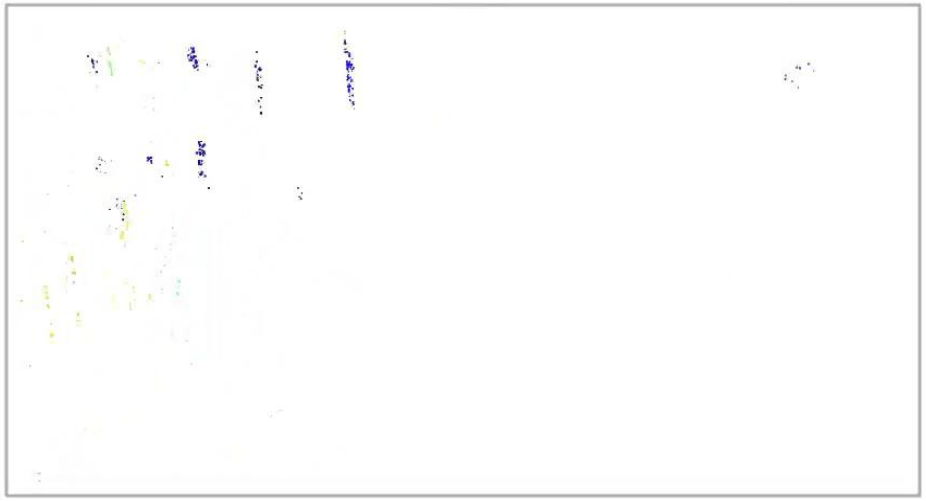
Biking (Pose from VIO system)

Left events

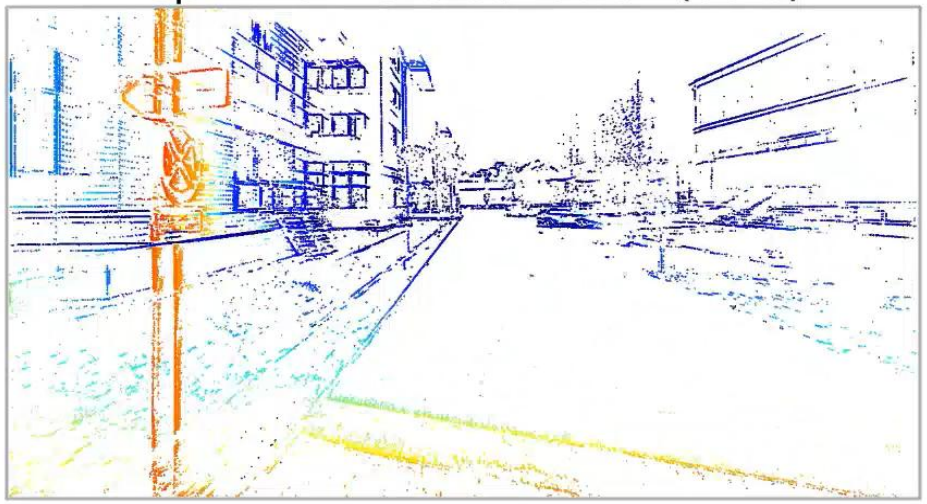


Confidence map (ours)

ESVO [Zhou et al, *TRO* 2021]

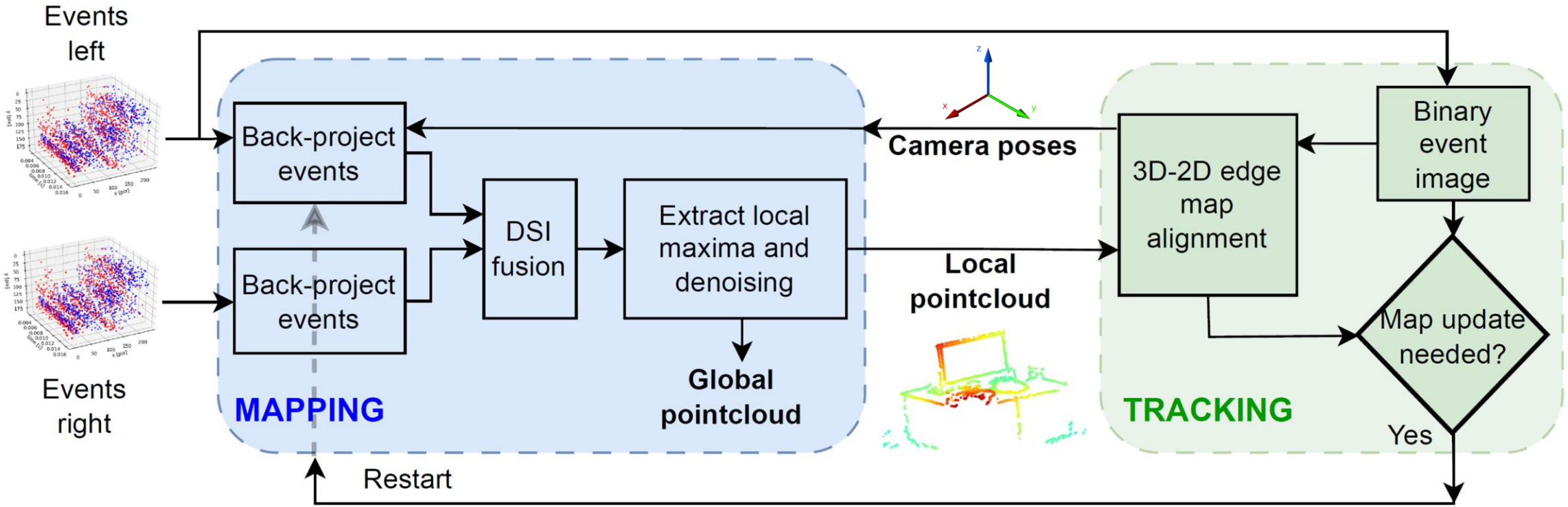


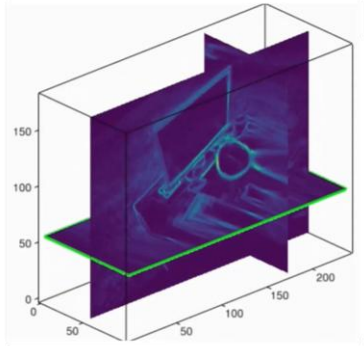
Depth from Stereo fusion (ours)

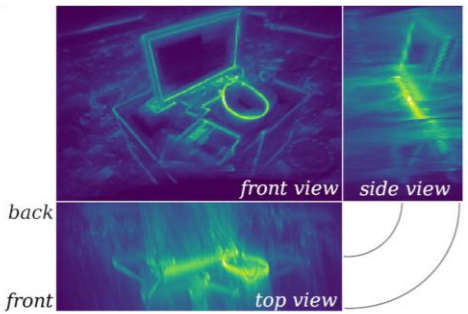


Ghosh, S., Gallego, G., *Multi-Event-Camera Depth Estimation and Outlier Rejection by Refocused Events Fusion*, Adv. Intell. Sys. 2022.

Tracking and Mapping with Stereo Events

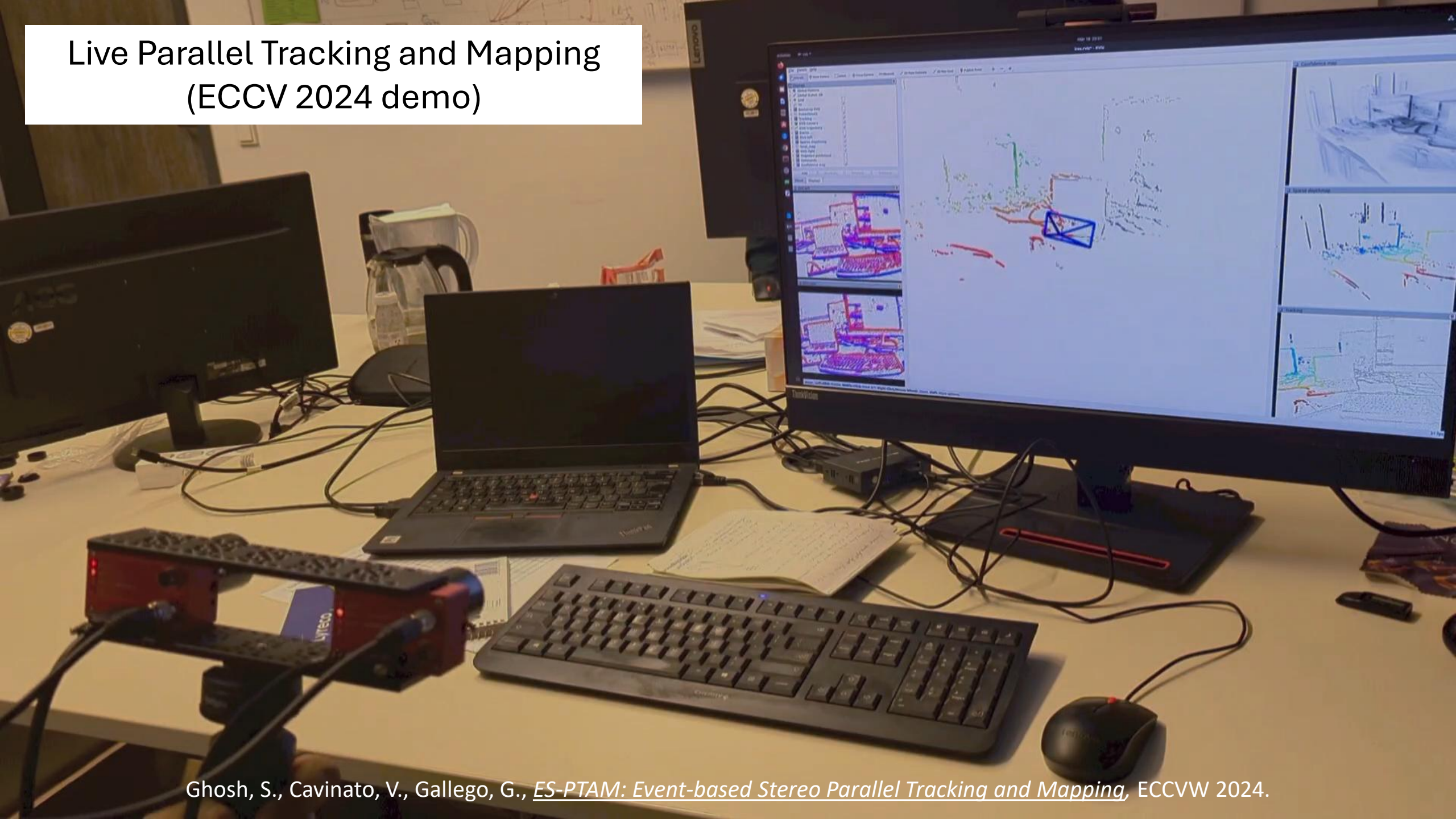






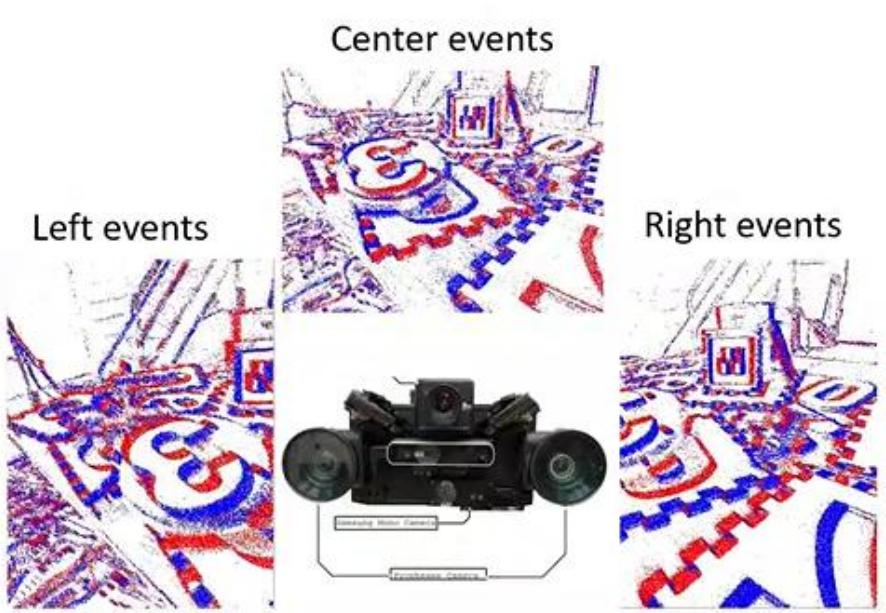
Ghosh, S., Cavinato, V., Gallego, G., *ES-PTAM: Event-based Stereo Parallel Tracking and Mapping*, ECCVW 2024.

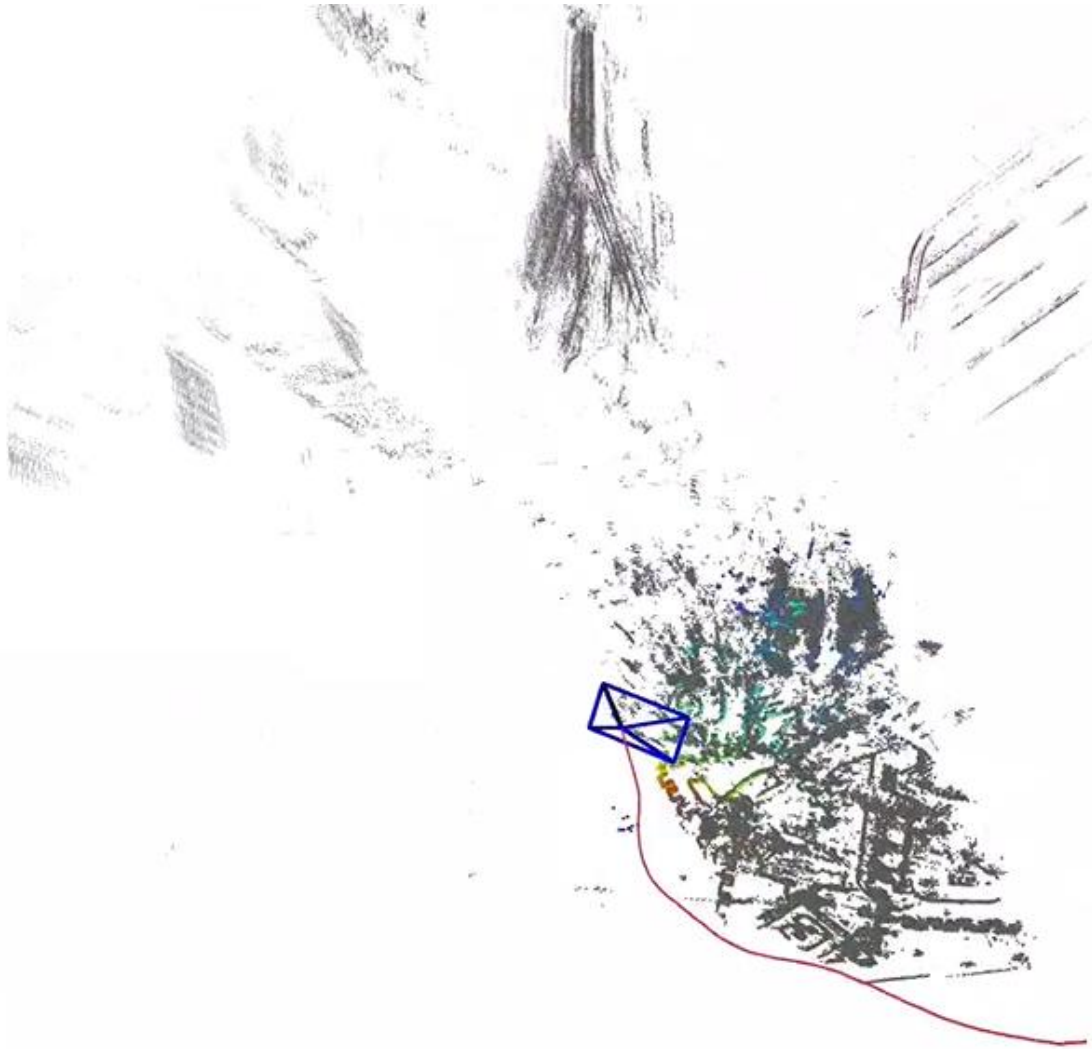
Live Parallel Tracking and Mapping (ECCV 2024 demo)

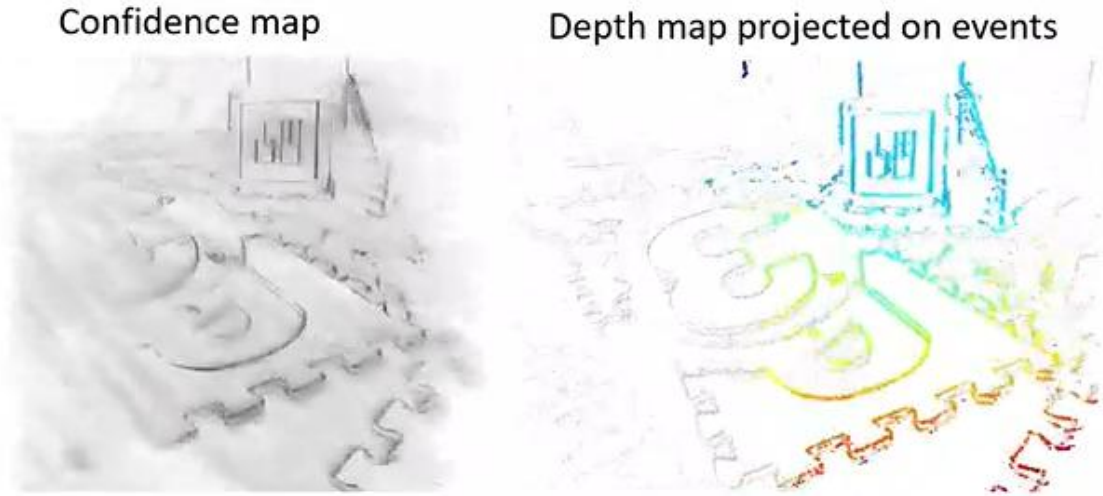


Ghosh, S., Cavinato, V., Gallego, G., *ES-PTAM: Event-based Stereo Parallel Tracking and Mapping*, ECCVW 2024.

Results on Trinocular setup



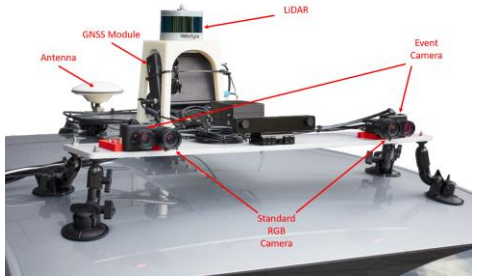




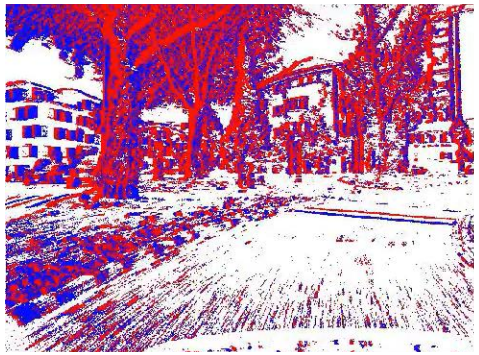
Camera trajectory + Local (colored) and Global Point-cloud

Ghosh, S., Cavinato, V., Gallego, G., *ES-PTAM: Event-based Stereo Parallel Tracking and Mapping*, ECCVW 2024.

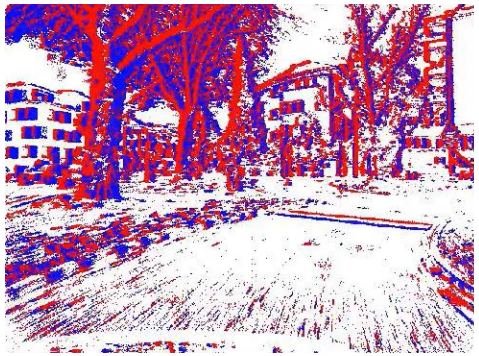
Results on Driving (DSEC)



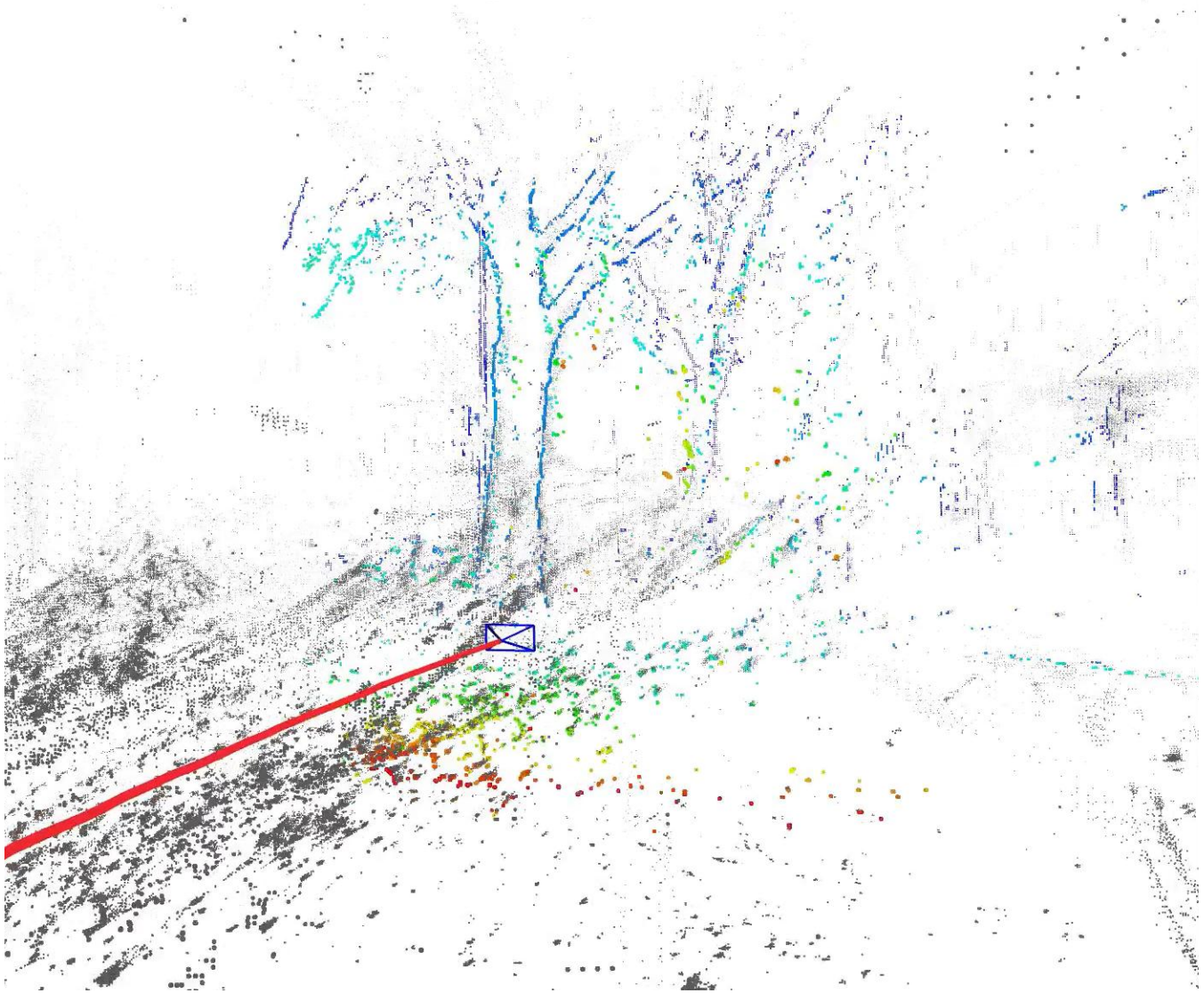
Left events



Right events



Camera trajectory + Local (colored) and Global Point-cloud



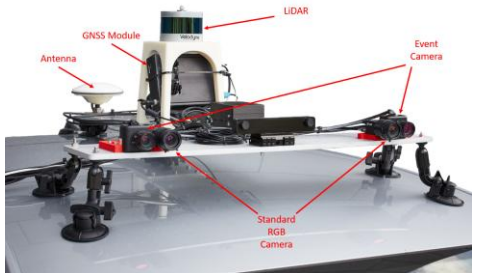
Confidence map

Depth map

Map projected on events

Ghosh, S., Cavinato, V., Gallego, G., *ES-PTAM: Event-based Stereo Parallel Tracking and Mapping*, ECCVW 2024.

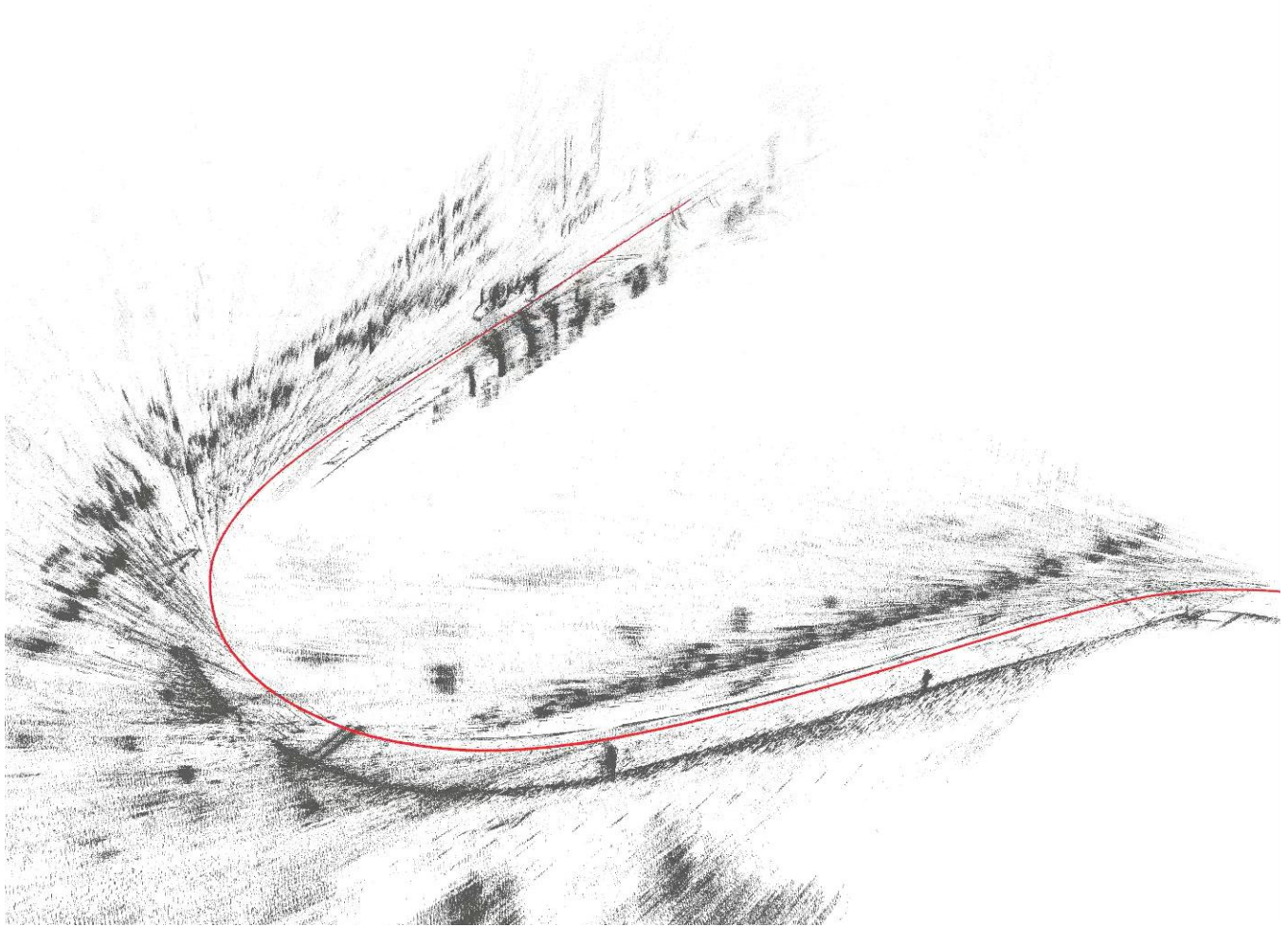
Results on Driving (DSEC)



Left events

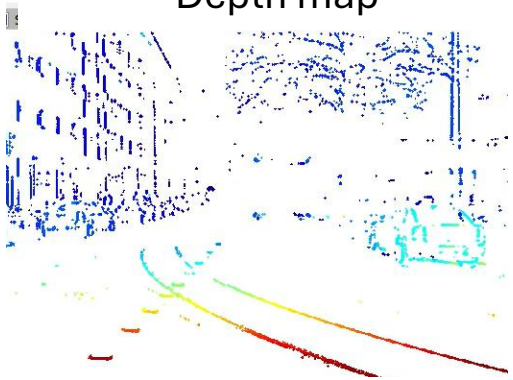
Right events

Camera trajectory + Global Point-cloud

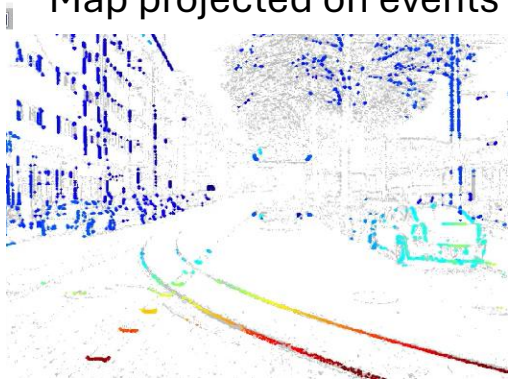


Confidence map

Depth map

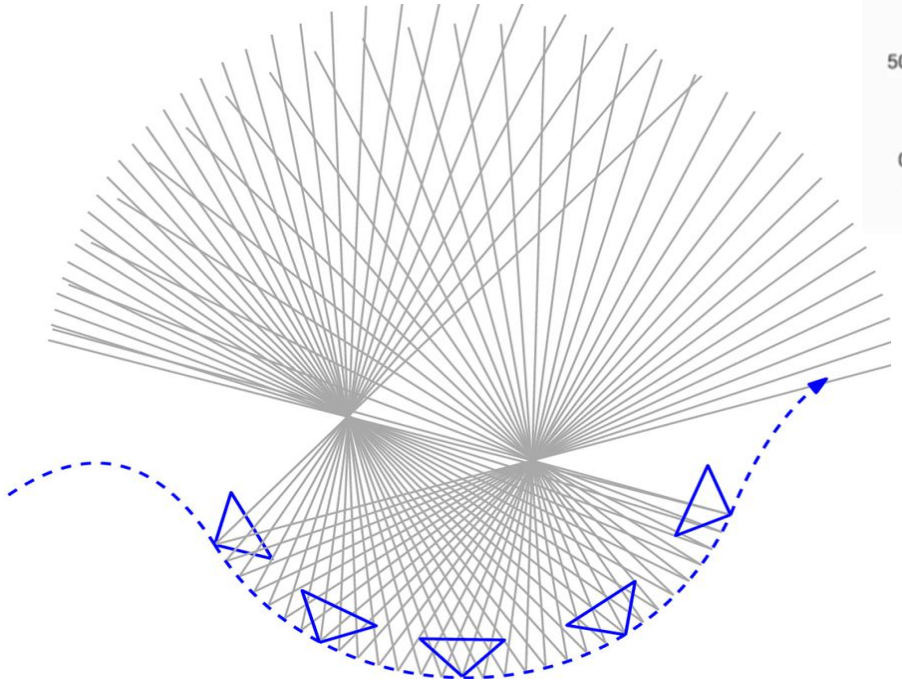


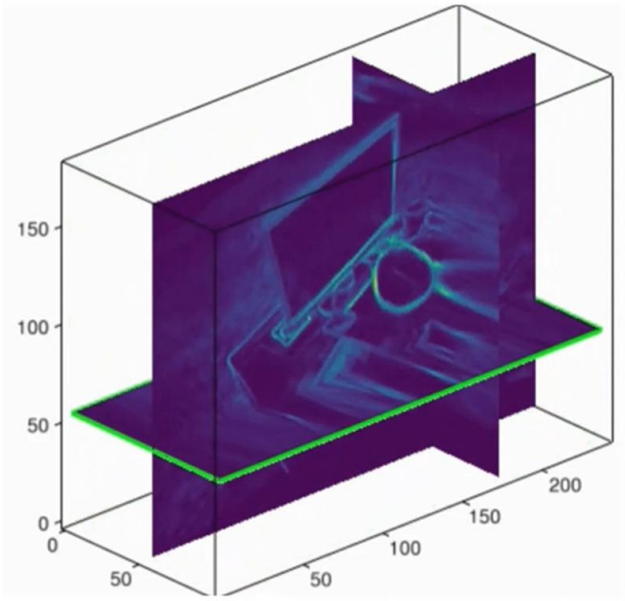
Map projected on events

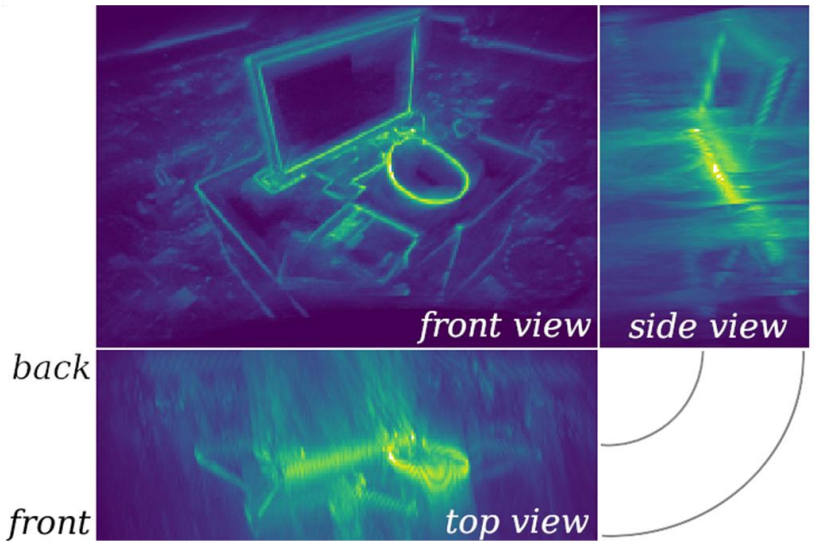


Ghosh, S., Cavinato, V., Gallego, G., *ES-PTAM: Event-based Stereo Parallel Tracking and Mapping*, ECCVW 2024.

Can we do better than argmax (counting rays)?

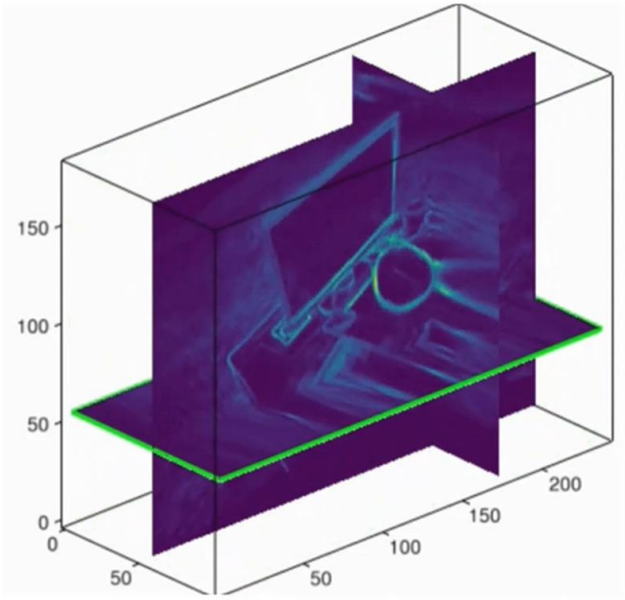


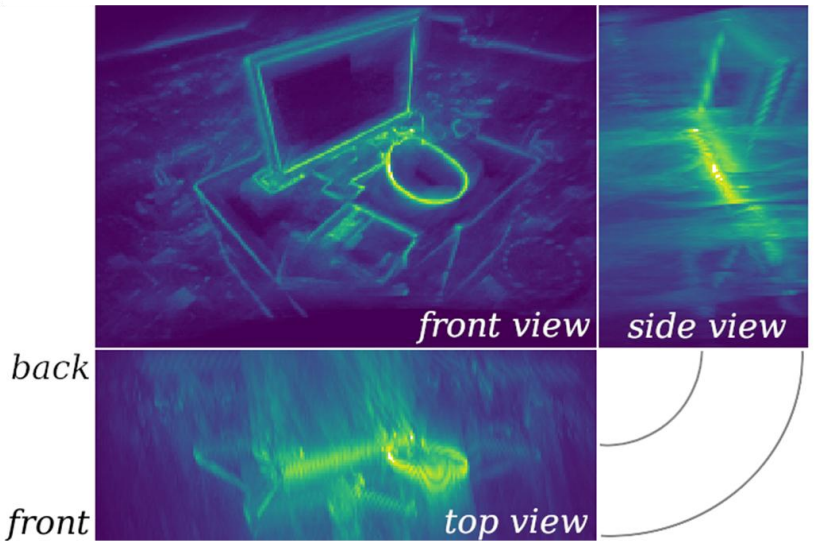


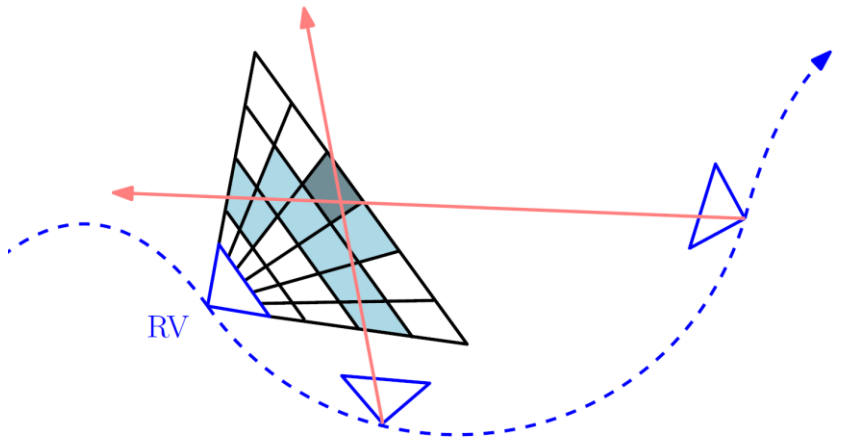


Disparity Space Image (DSI)

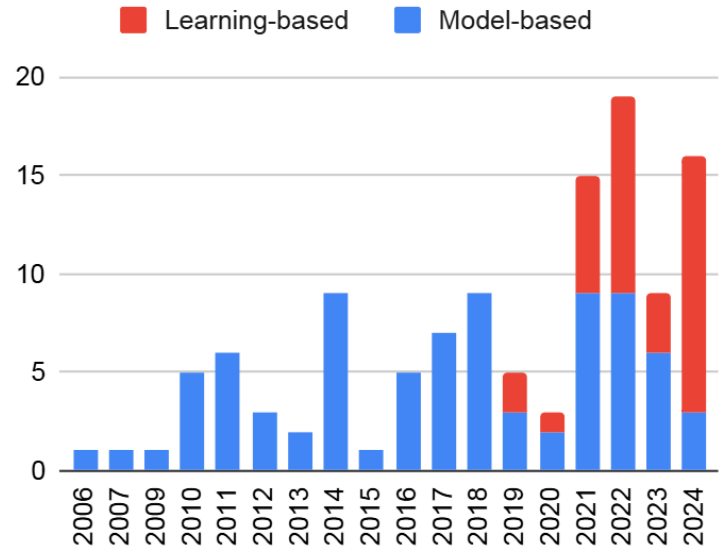
Can we do better than argmax (counting rays)?



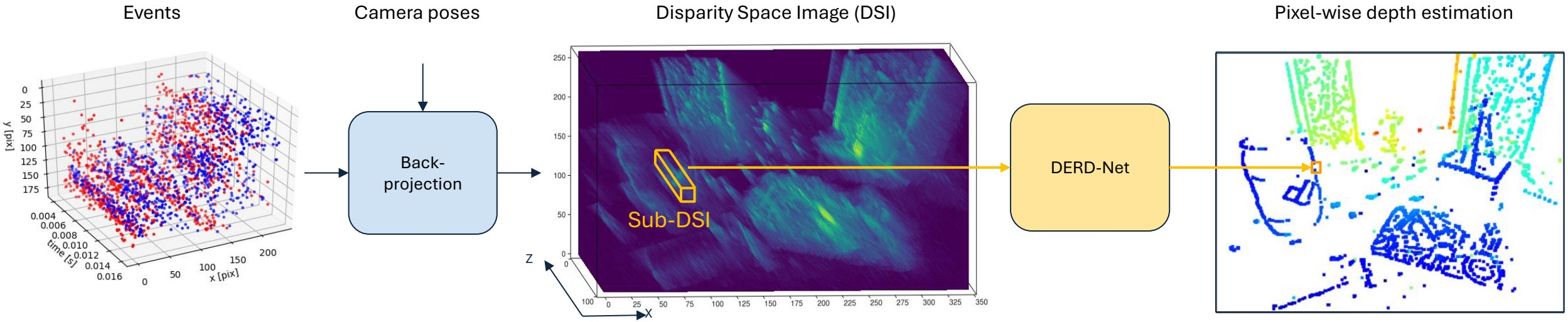




Disparity Space Image (DSI)



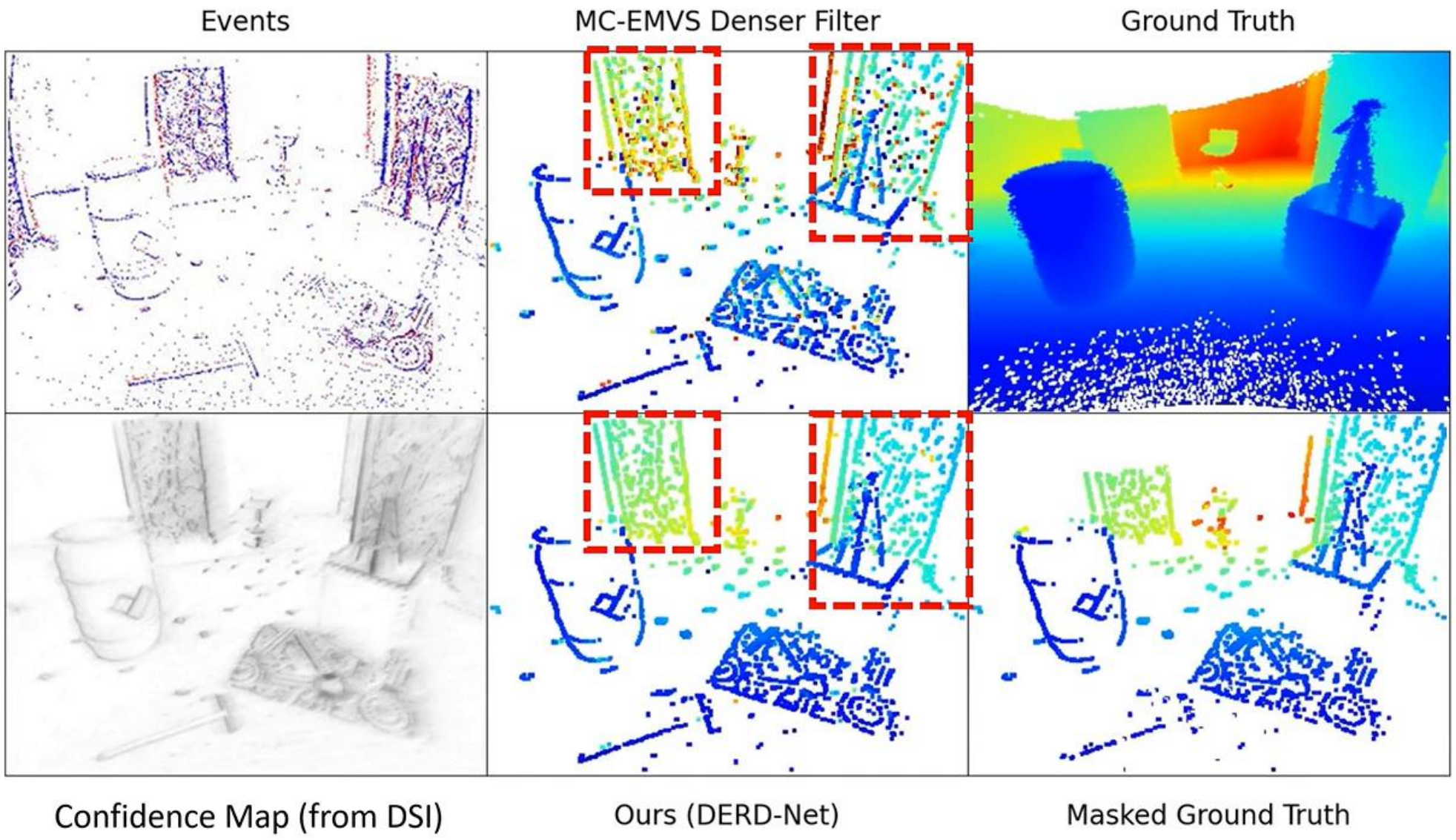
DERD-Net: Learning Depth from DSIs



Inference time: 0.37 ms
Model size: < 1 MB

Hitzges, D.*, Ghosh, S.*, Gallego, G., *DERD-Net: Learning Depth from Event-based Ray Densities*, NeurIPS 2025.

DERD-Net Results



Hitzges, D.*, Ghosh, S.*, Gallego, G., *DERD-Net: Learning Depth from Event-based Ray Densities*, NeurIPS 2025.

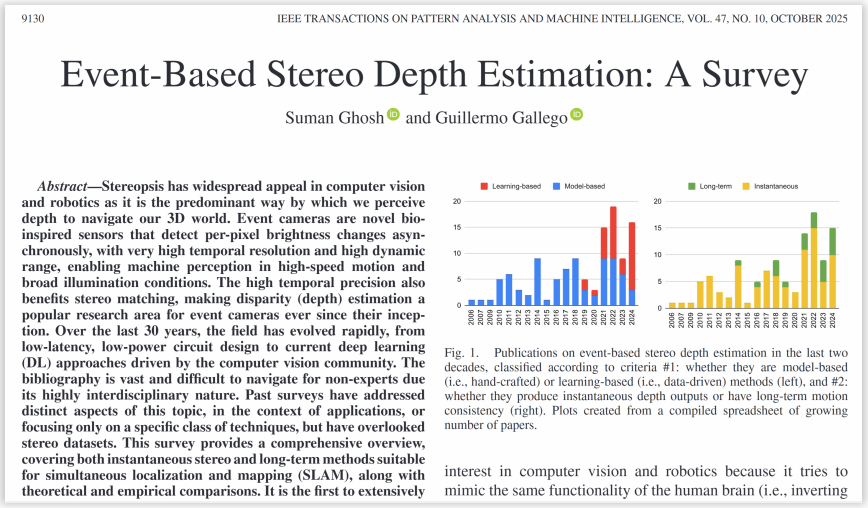
Conclusion

- Including camera motion information improves accuracy.
- DSIs are powerful intermediate representations that forego explicit stereo event matching.
- Sharp semi-dense depth maps can be extracted by simply finding the local maxima.
- Depth can be used for visual odometry via simple edge alignment.
- Replacing local maxima detection with a DNN further improves performance.
- Harness sparsity: Processing sub-DSIs independently enables efficient and robust estimation with a lightweight network.

What’s next?

- Efficient on-device processing with modern HD event cameras.
- Foundation Models for event-based stereo.
- Night-time perception and highly dynamic environments.
- Active mapping and control.
- Better benchmarking.
  - Dynamic scenes with independently moving objects.
  - Moving beyond synchronous frame-based depth evaluations.

Summary



9130 IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE, VOL. 47, NO. 10, OCTOBER 2025

Event-Based Stereo Depth Estimation: A Survey

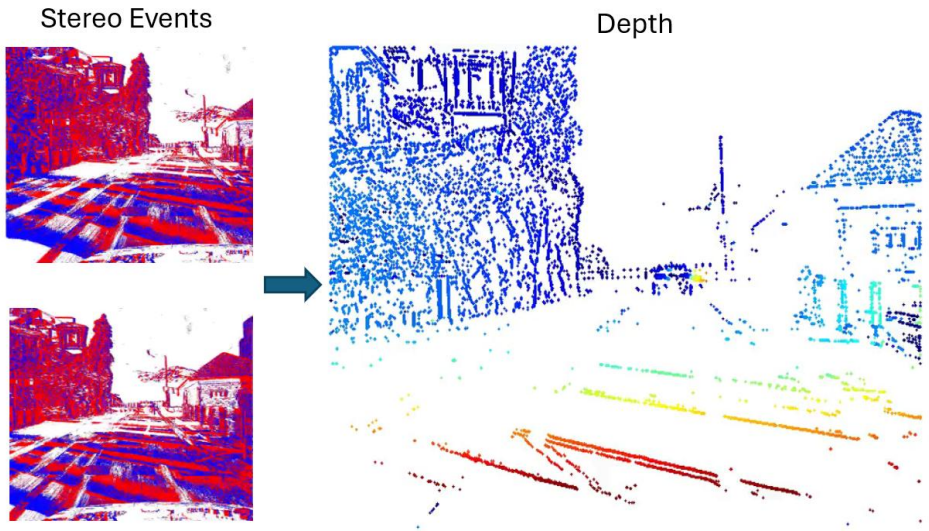
Suman Ghosh and Guillermo Gallego

Abstract—Stereopsis has widespread appeal in computer vision and robotics as it is the predominant way by which we perceive depth to navigate our 3D world. Event cameras are novel bio-inspired sensors that detect per-pixel brightness changes asynchronously, with very high temporal resolution and high dynamic range, enabling machine perception in high-speed motion and broad illumination conditions. The high temporal precision also benefits stereo matching, making disparity (depth) estimation a popular research area for event cameras ever since their inception. Over the last 30 years, the field has evolved rapidly, from low-latency, low-power circuit design to current deep learning (DL) approaches driven by the computer vision community. The bibliography is vast and difficult to navigate for non-experts due its highly interdisciplinary nature. Past surveys have addressed distinct aspects of this topic, in the context of applications, or focusing only on a specific class of techniques, but have overlooked stereo datasets. This survey provides a comprehensive overview, covering both instantaneous stereo and long-term methods suitable for simultaneous localization and mapping (SLAM), along with theoretical and empirical comparisons. It is the first to extensively

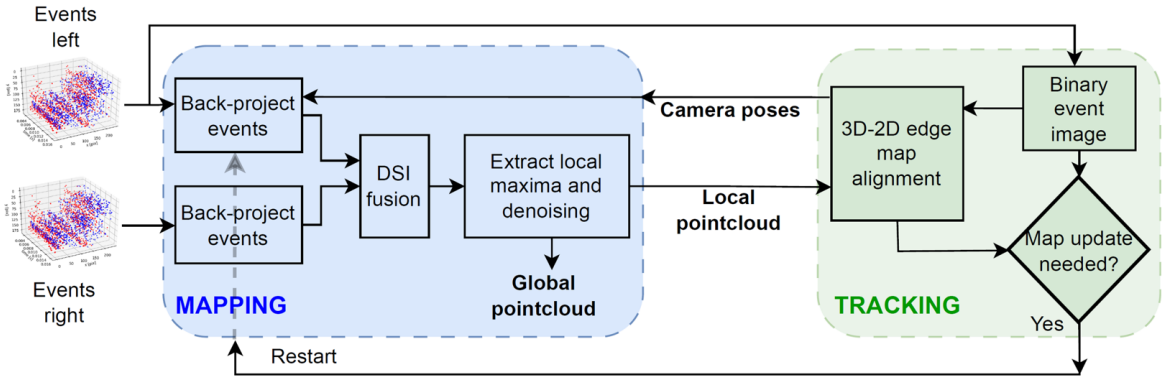
Fig. 1. Publications on event-based stereo depth estimation in the last two decades, classified according to criteria #1: whether they are model-based (i.e., hand-crafted) or learning-based (i.e., data-driven) methods (left), and #2: whether they produce instantaneous depth outputs or have long-term motion consistency (right). Plots created from a compiled spreadsheet of growing number of papers.

interest in computer vision and robotics because it tries to mimic the same functionality of the human brain (i.e., inverting

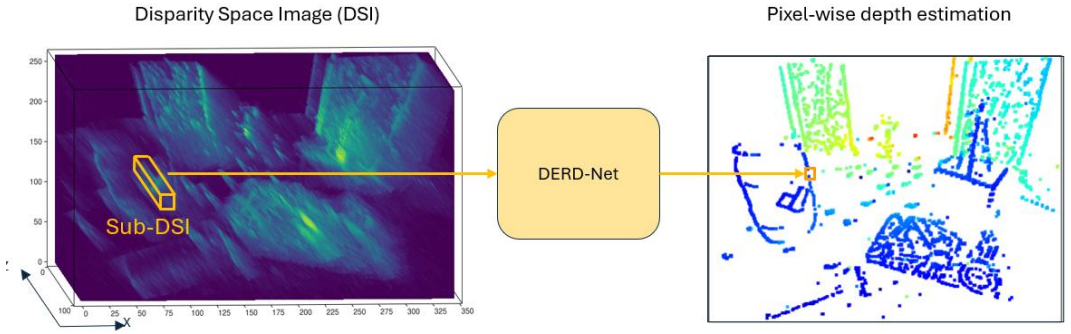
Event-based Stereo Survey, T-PAMI 2025.



Multi-Camera EMVS, AISY 2022.



ES-PTAM, ECCVW 2024.



DERD-Net, NeurIPS 2025.