MC-EMVS: Multi-Event-Camera Depth Estimation and Outlier Rejection by Refocused Events Fusion

Suman Ghosh and Guillermo Gallego

Paper

Summary:

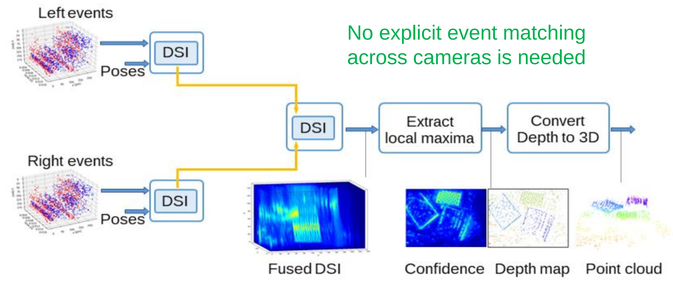
We propose a framework to fuse projected ray density volumes across cameras and time for stereo 3D reconstruction in SLAM.

Our stereo ray density fusion method:

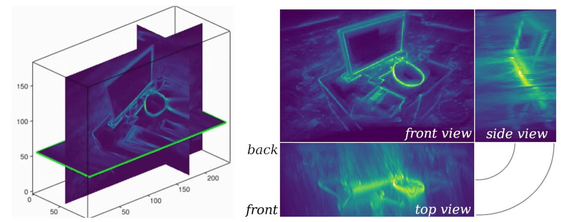
- Produces **SOTA results** (evaluated over five public datasets with diverse resolutions).
- **Speeds up convergence** of the 3D map compared to monocular.
- Scales well in **multi-camera (≥ 2) settings**.

Event Processing Pipeline

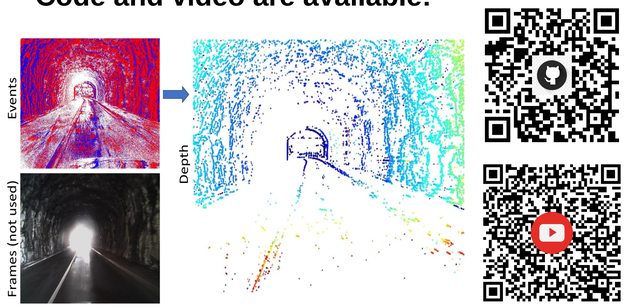
1. Shoot rays through events using known camera poses.
2. Collect ray densities into volumes called DSIs.
3. Fuse ray densities from different cameras and different times.
4. Extract depth and confidence maps from local maxima and noise filters.



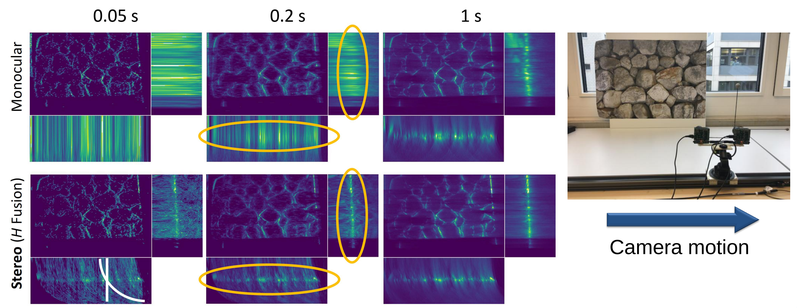
Disparity Space Image (DSI)



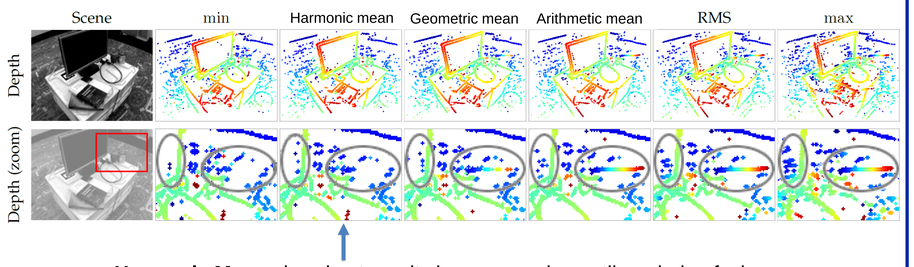
Code and video are available!



Monocular vs. Stereo

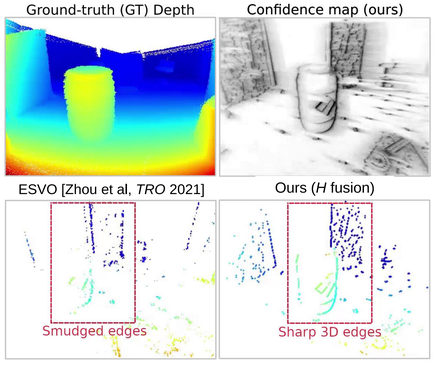


DSI Fusion Functions



Harmonic Mean gives best results by suppressing outliers during fusion.

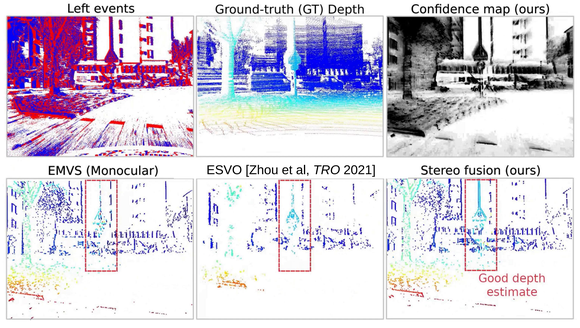
MVSEC dataset (346x260 px)



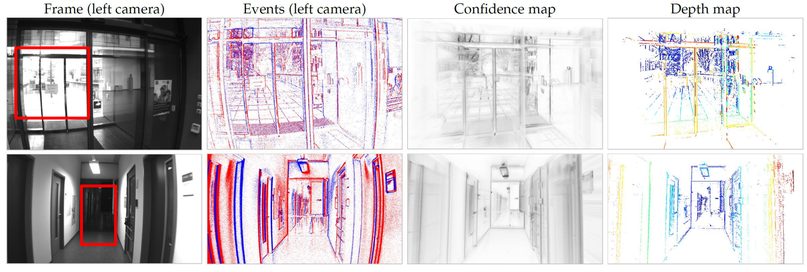
Alg	Mean Err [cm]	Median Err [cm]	Bad-pix [%]	SILog Err x100	Log RMSE x100
EMVS	33.78	14.35	3.84	4.2	20.72
ESVO	22.70	9.83	2.83	3.03	17.53
SGM	35.42	12.35	6.39	8.45	29.49
GTS	389.00	45.43	38.45	74.47	89.08
Ours (*H* fusion)	**20.07**	**9.53**	**1.35**	**1.72**	**13.24**

Evaluated over 200s (110 million events and 4000 ground truth depth maps) of MVSEC indoor sequences (flying 1, 2 and 3).

DSEC dataset (VGA resolution)



TUM-VIE dataset (1Mpx resolution)



We reconstruct HDR semi-dense depth maps.
Poses are generated using an offline VIO system.

Journal reference: Ghosh and Gallego, “Multi-Event-Camera Depth Estimation and Outlier Rejection by Refocused Events Fusion”, *Advanced Intelligent Systems* 2022.