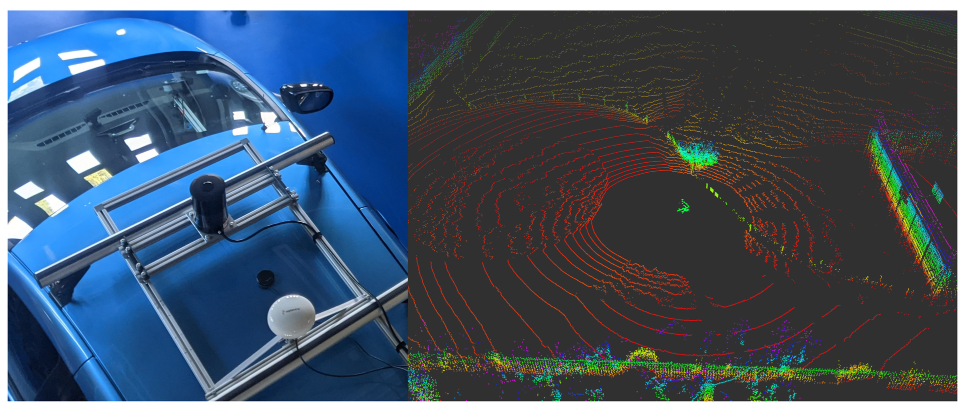
Learning to Estimate Two Dense Depths from LiDAR and Event Data

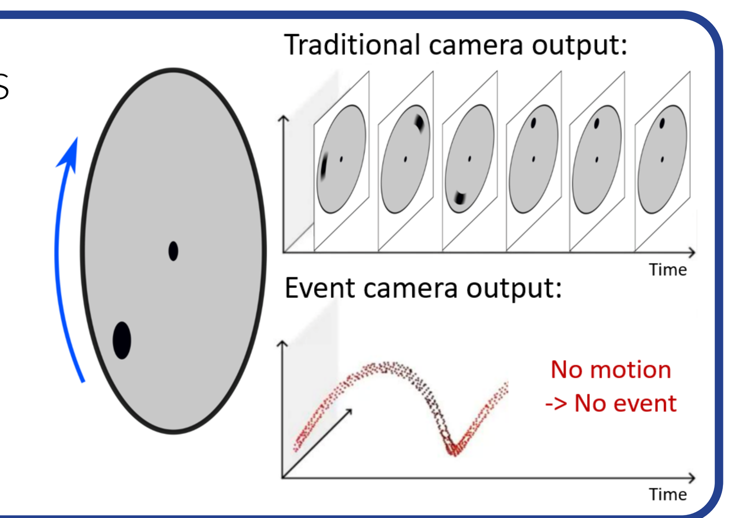
Vincent Brebion Julien Moreau Franck Davoine

Heudiasyc Lab, CNRS, Université de technologie de Compiègne (UTC), France
firstname.lastname@hds.utc.fr

LiDAR sensors provide accurate but sparse 3D information about their surrounding environment. While they are a key component for autonomous navigation, their sparsity often constitutes a limiting factor



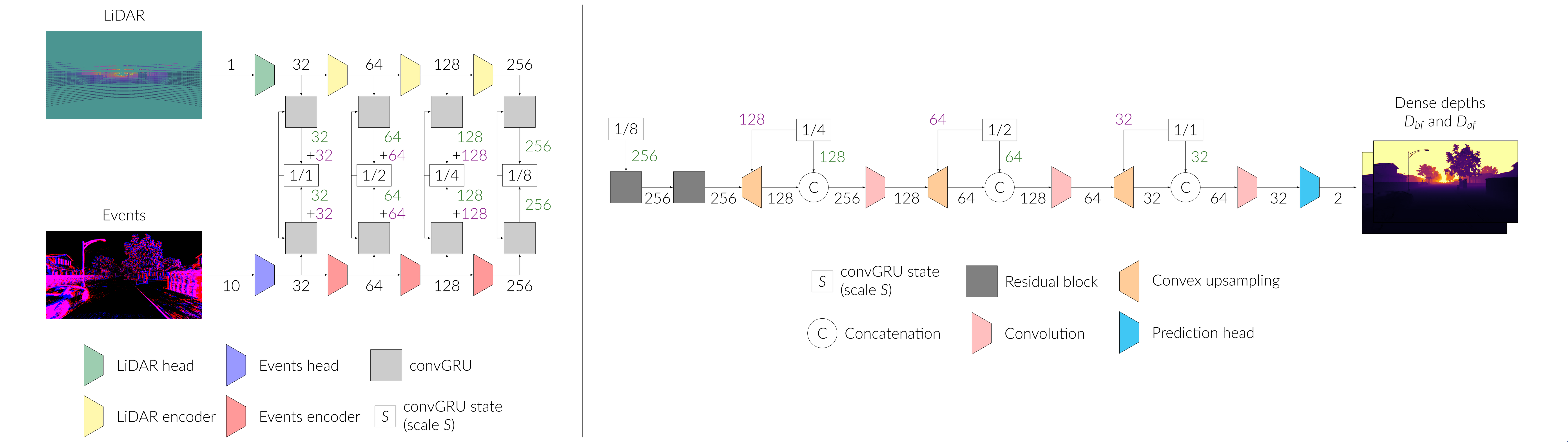
Event cameras are emerging sensors which only react to brightness changes, and output them as a fully asynchronous stream of data. They offer many advantages: high dynamic range, low latency, and no motion blur



Motivation and Goals

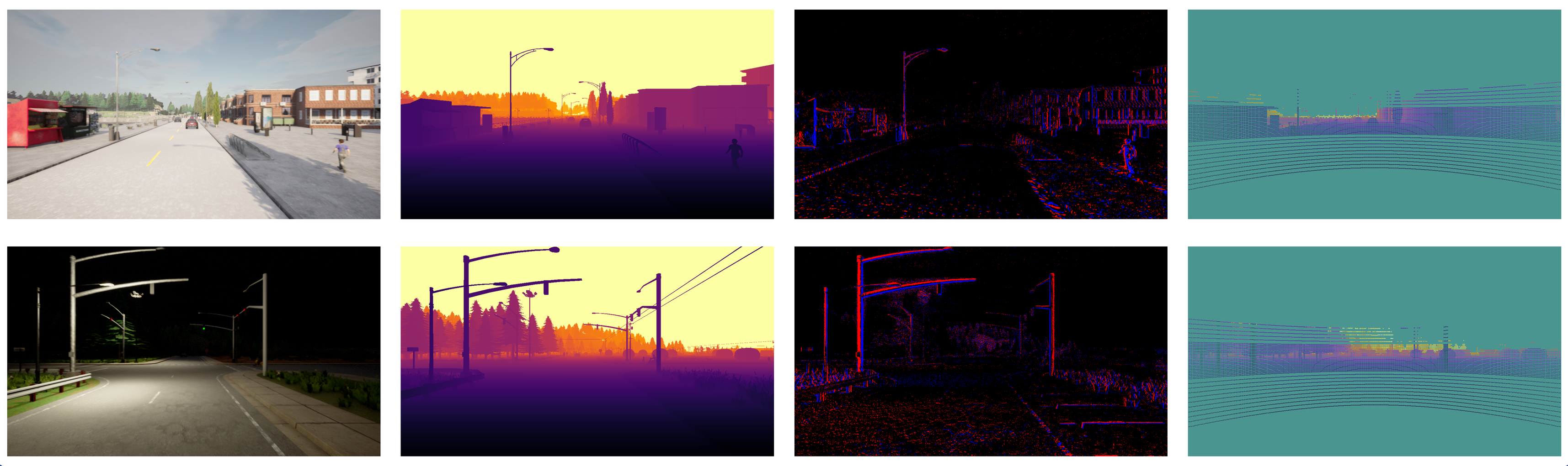
- Main objective: fusing asynchronous LiDAR and event data to estimate dense depth maps
- By definition, events represent **a change in illumination** ⇒ They might also represent **a change of depth** ⇒ Estimating a single depth per event is **erroneous**
- Our solution: computing **two** dense depth maps at each time step, one **before** the events (D_{bf}), and one **after** the events (D_{af})

Our Asynchronous LiDAR and Events Depths (ALED) Densification Network

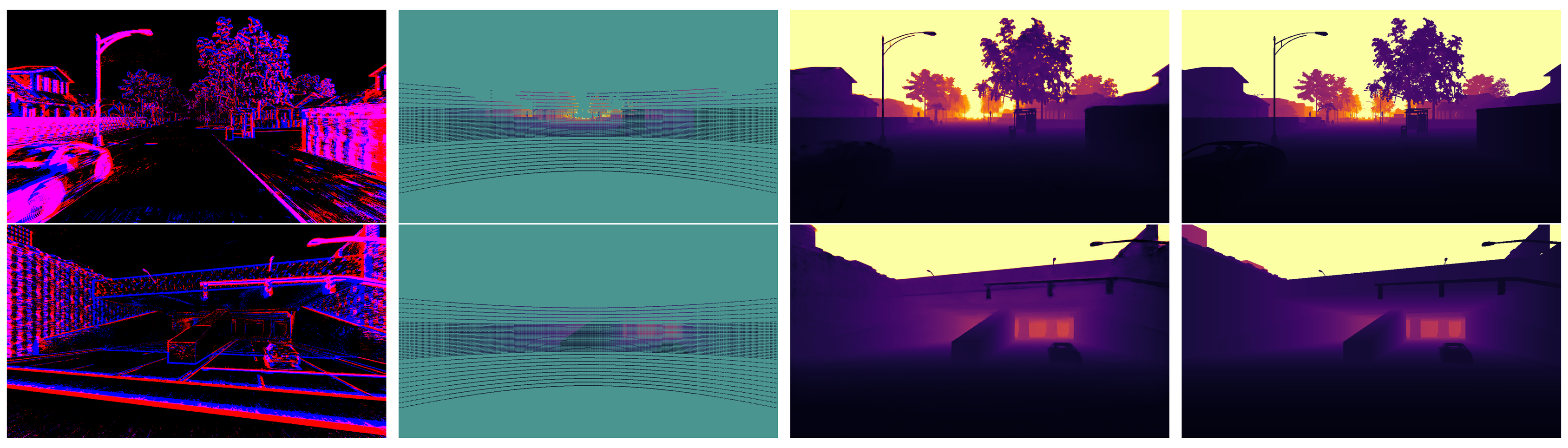


Our Synthetic LiDAR Events Depths (SLED) Dataset

- Only one dataset available: MVSEC [1]
  - Low-resolution (346×260)
  - Approximate ground truth (no synchronization, errors for moving objects)
- SLED is a novel synthetic dataset, recorded in CARLA [2]
  - High-resolution (1280×720)
  - Perfect ground truth
  - ~30 minutes of data, recorded in a wide range of environments (urban, suburban, highway, countryside) and of atmospheric conditions (day/night, sunny/overcast)



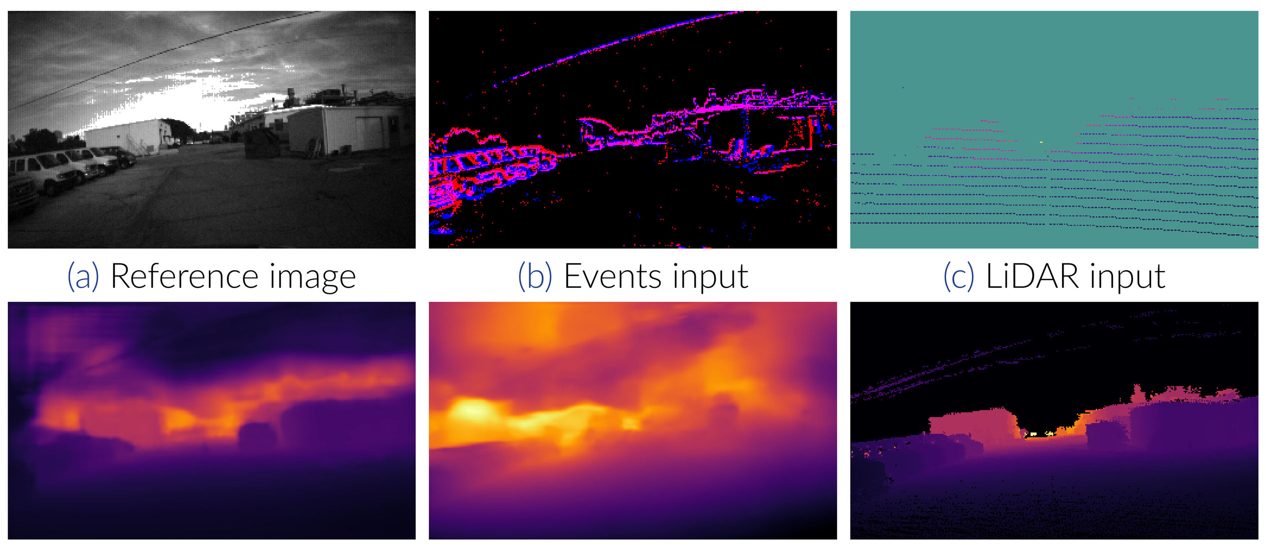
Results on our SLED Dataset



(a) Events input (b) LiDAR input (c) ALED (ours) (d) Ground truth

Results on the MVSEC Dataset

Recording	Cutoff	Event- and frame-based RAMNet [3]	Event- and frame-based EvT⁺ [4]	LiDAR- and event-based Cui et al. [5]	LiDAR- and event-based ALED
Outdoor day 1	10m	1.39	1.27	1.24	**0.50**
	30m	2.76	2.37	4.87	**1.02**
	100m	-	-	-	**1.60**
Outdoor night 1	10m	2.50	**1.48**	2.26	1.52
	30m	3.82	2.84	4.50	**1.95**
	100m	-	-	-	**2.54**
Outdoor night 2	10m	1.21	1.48	1.88	**1.09**
	30m	3.28	2.92	4.67	**1.64**
	100m	-	-	-	**1.97**
Outdoor night 3	10m	1.01	1.40	1.78	**0.81**
	30m	3.43	2.79	4.55	**1.33**
	100m	-	-	-	**1.66**



(a) Reference image (b) Events input (c) LiDAR input

(d) ALED (ours) (e) RAMNet [3] (f) Ground truth

References

[1] Alex Zihao Zhu, Dinesh Thakur, Tolga Özaslan, Bernd Pfrommer, Vijay R. Kumar, and Kostas Daniilidis. The multivehicle stereo event camera dataset: An event camera dataset for 3D perception. *IEEE Robotics and Automation Letters*, 3:2032–2039, 2018.

[2] Alexey Dosovitskiy, Germán Ros, Felipe Codevilla, Antonio López, and Vladlen Koltun. CARLA: An open urban driving simulator. In *Proceedings of the 1st Annual Conference on Robot Learning*, pages 1–16, 2017.

[3] Daniel Gehrig, Michelle Rüegg, Mathias Gehrig, Javier Hidalgo-Carrió, and Davide Scaramuzza. Combining events and frames using recurrent asynchronous multimodal networks for monocular depth prediction. *IEEE Robotics and Automation Letters*, 6:2822–2829, 2021.

[4] Alberto Sabater, Luis Montesano, and Ana Cristina Murillo. Event transformer+. A multi-purpose solution for efficient event data processing. *ArXiv*, abs/2211.12222, 2022.

[5] Mingyue Cui, Yuzhang Zhu, Yechang Liu, Yun-Meng Liu, Gang Chen, and Kai Huang. Dense depth-map estimation based on fusion of event camera and sparse LiDAR. *IEEE Transactions on Instrumentation and Measurement*, 71:1–11, 2022.

Région Hauts-de-France SIVALab/ Renault Group cnrs Alliance Sorbonne Université utc