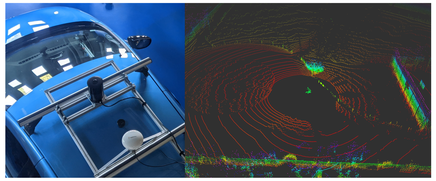
DELTA: Dense Depth from Events and LiDAR using Transformer's Attention

Vincent Brebion[1,2,3] Julien Moreau[1] Franck Davoine[2]

[1]Université de technologie de Compiègne, CNRS, Heudiasyc Lab., France [2]CNRS, INSA Lyon, UCBL, LIRIS Lab., France [3]Centre for Environmental and Climate Science, Lund University, Sweden

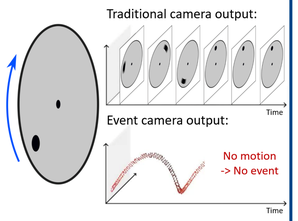
LiDAR sensor

- Essential sensor in numerous robotic applications
- Provides **accurate depth** information
- But in a **very sparse** way
- And at a **very low rate** (10–20Hz)



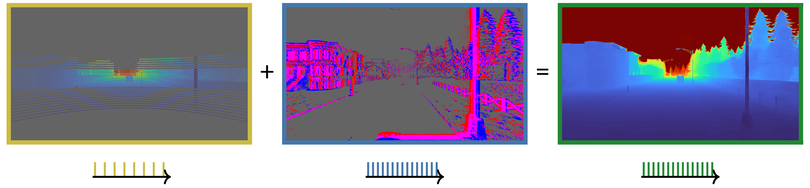
Event camera

- Neuromorphic sensor (inspired by human vision)
- Only reacts to **brightness changes**
- Produces **very sparse** data, highlighting only **edges** and **textures**
- Provides data at a **very high rate** with a **low latency**



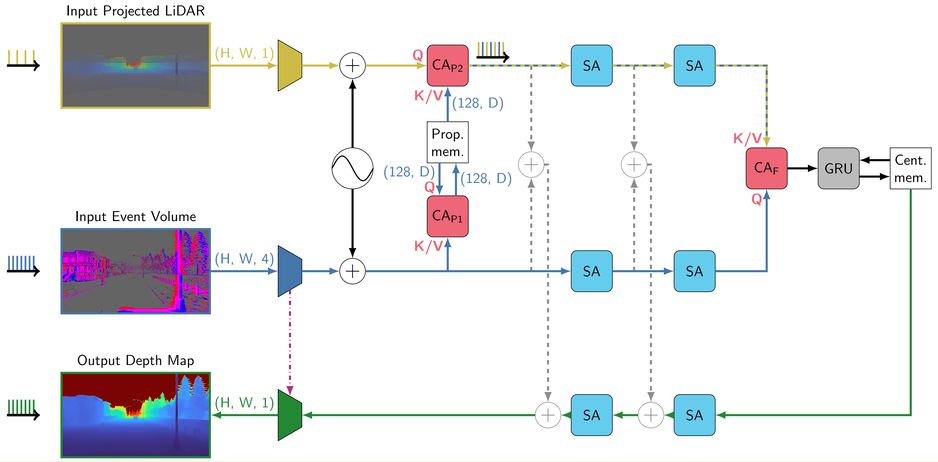
Motivation & Objectives

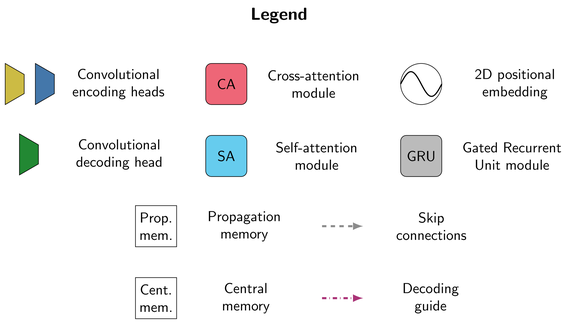
- A LiDAR and an event camera offer **complementary** information
- Can we combine the **projected LiDAR point clouds** with **high-rate temporal windows of event data** to estimate **accurate dense depth maps** at a **higher rate** than the LiDAR input?



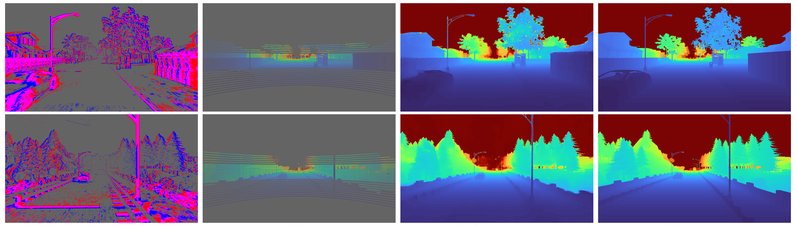
- Other notable event-based depth estimation methods from the SOTA:
  - RAMNet [1]: events + frames, CNN, limited accuracy
  - StereoSpike [2]: events in stereo, SNN, good accuracy
  - ALED [3]: events + LiDAR, our previous work, CNN, SOTA accuracy except for close objects
- How can we improve the SOTA?
  - Compared to CNNs, **attention-based models** do not operate on neighborhoods of limited size
  - This property could be valuable for estimating the depth of objects with few LiDAR points close to them (*e.g.*, close objects)

Our model: DELTA, exploiting self- and cross-attention to model the spatial and temporal relations within and between the event and LiDAR data





Qualitative results on the synthetic SLED dataset [3]

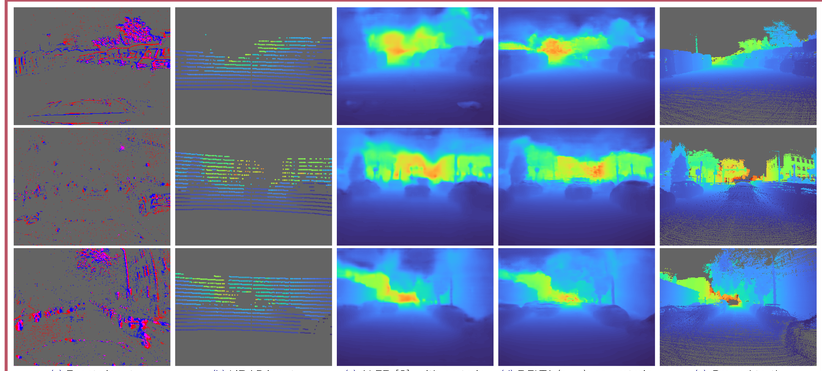


(a) Events input (b) LiDAR input (c) DELTA (ours) (d) Ground truth

Quantitative results on the synthetic SLED dataset [3]

Map	Cutoff	ALED [3], previous SOTA							DELTA (ours)						
		Mean↓	AbsRel↓	RMS↓	RMSlog↓	δ^1↑	δ^2↑	δ^3↑	Mean↓	AbsRel↓	RMS↓	RMSlog↓	δ^1↑	δ^2↑	δ^3↑
Town01	10m	1.24	0.211	8.022	0.478	0.890	0.962	0.978	**0.66**	**0.106**	**6.981**	**0.282**	**0.963**	**0.981**	**0.989**
	20m	2.10	0.232	11.973	0.489	0.885	0.952	0.970	**1.33**	**0.133**	**10.601**	**0.311**	**0.948**	**0.973**	**0.983**
	30m	2.74	0.239	13.831	0.481	0.875	0.945	0.966	**1.91**	**0.147**	**12.437**	**0.317**	**0.932**	**0.965**	**0.979**
	100m	4.26	0.241	17.182	0.468	0.860	0.935	0.959	**3.22**	**0.157**	**15.177**	**0.317**	**0.909**	**0.954**	**0.973**
	200m	**4.53**	0.173	**18.775**	0.405	0.892	0.947	0.966	4.54	**0.119**	19.606	**0.311**	**0.921**	**0.957**	**0.973**
Town03	10m	2.01	0.290	14.900	0.456	0.904	0.952	0.966	**0.54**	**0.082**	**5.656**	**0.184**	**0.958**	**0.973**	**0.986**
	20m	2.87	0.301	17.134	0.516	0.883	0.939	0.961	**1.31**	**0.117**	**9.958**	**0.234**	**0.934**	**0.962**	**0.980**
	30m	3.35	0.292	17.706	0.507	0.874	0.935	0.959	**1.93**	**0.133**	**11.831**	**0.251**	**0.919**	**0.952**	**0.974**
	100m	4.62	0.275	18.840	0.484	0.860	0.928	0.955	**3.40**	**0.146**	**15.119**	**0.265**	**0.897**	**0.941**	**0.968**
	200m	4.87	0.216	20.059	0.438	0.882	0.937	0.960	**4.63**	**0.122**	**19.363**	**0.271**	**0.906**	**0.944**	**0.967**

Qualitative results on the MVSEC dataset



(a) Events input (b) LiDAR input (c) ALED [3], with pretrain (d) DELTA (ours), no pretrain (e) Ground truth

References

[1] 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.

[2] Ulysse Rançon, Javier Cuadrado-Anibarro, Benoit R. Cottereau, and Timothée Masquelier. Stereospike: Depth learning with a spiking neural network. *IEEE Access*, 10:127428–127439, 2021.

[3] Vincent Brebion, Julien Moreau, and Franck Davoine. Learning to estimate two dense depths from LiDAR and event data. In *Image Analysis - 22nd Scandinavian Conference, SCIA 2023, Sirkka, Finland, April 18-21, 2023, Proceedings, Part II*, volume 13886 of *Lecture Notes in Computer Science*, pages 517–533. Springer, 2023.