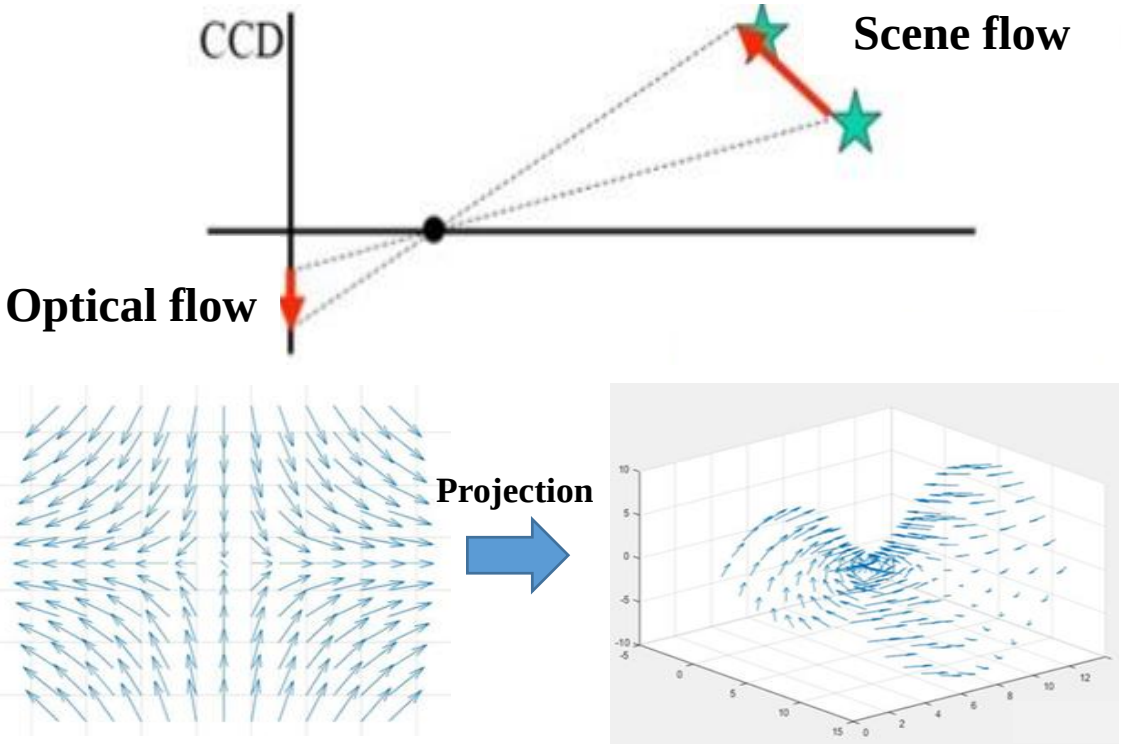
Bring Event into RGB and LiDAR: Hierarchical Visual-Motion Fusion for Scene Flow

Hanyu Zhou[1], Yi Chang[1*], Zhiwei Shi[1], Luxin Yan[1]

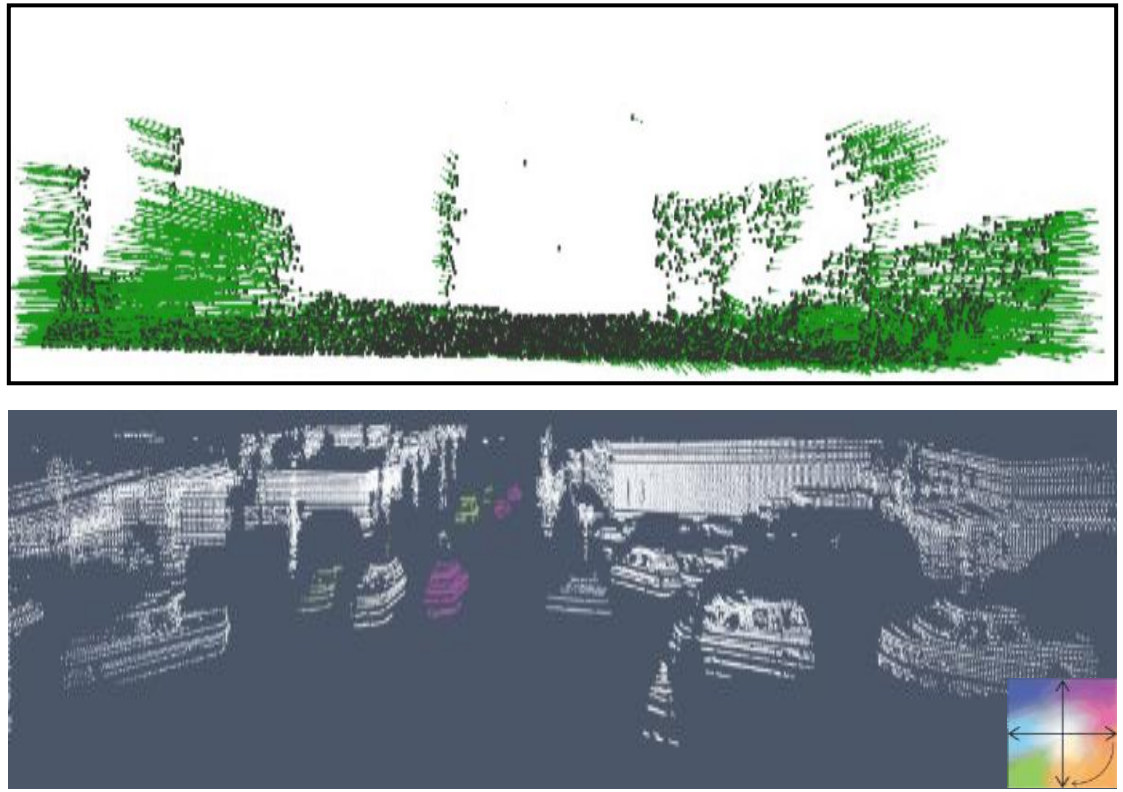
[1] Huazhong University of Science and Technology

CVPR 2024

Scene Flow: 3D Metric Motion



Scene Flow Schematic

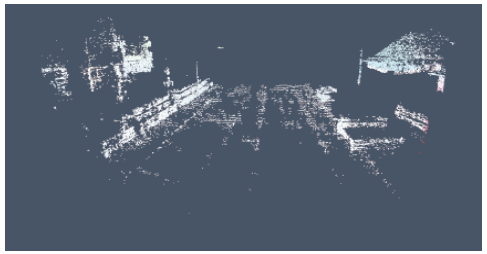


Scene Flow Visualization

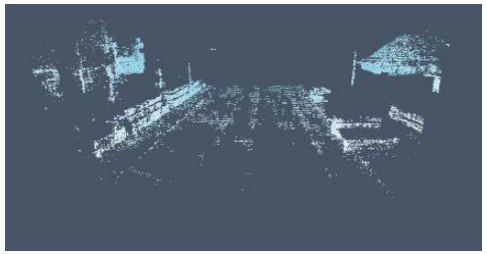
Related Methods

RGB

2D Dense

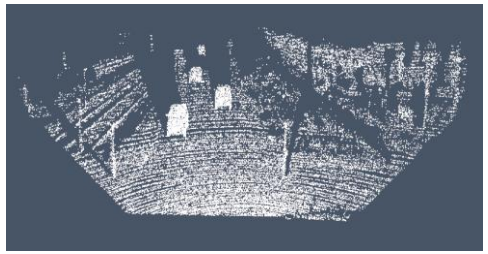


LiDAR



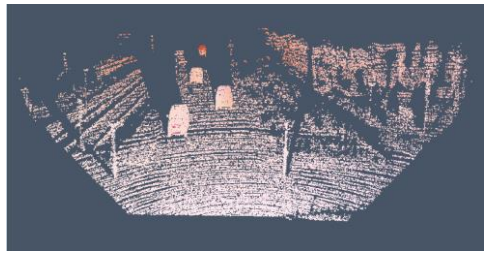
3D Sparse

RGB



LiDAR

Motion Fusion



3D Dense

Problem

Daytime Scene

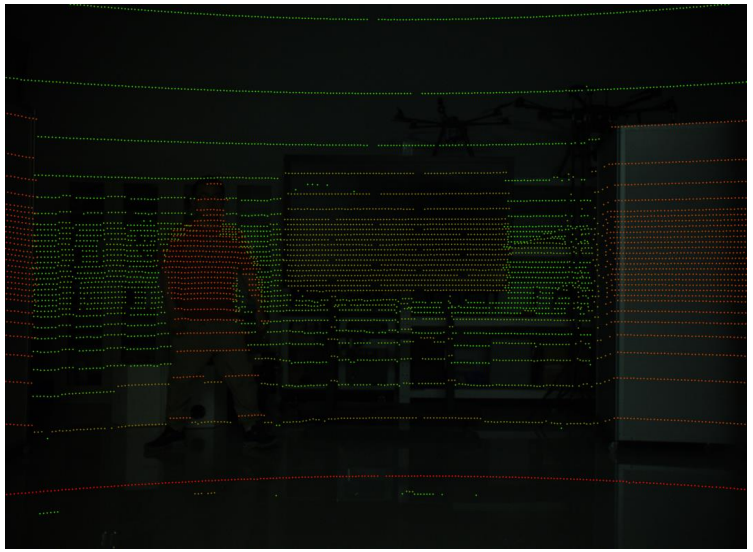
Nighttime Scene

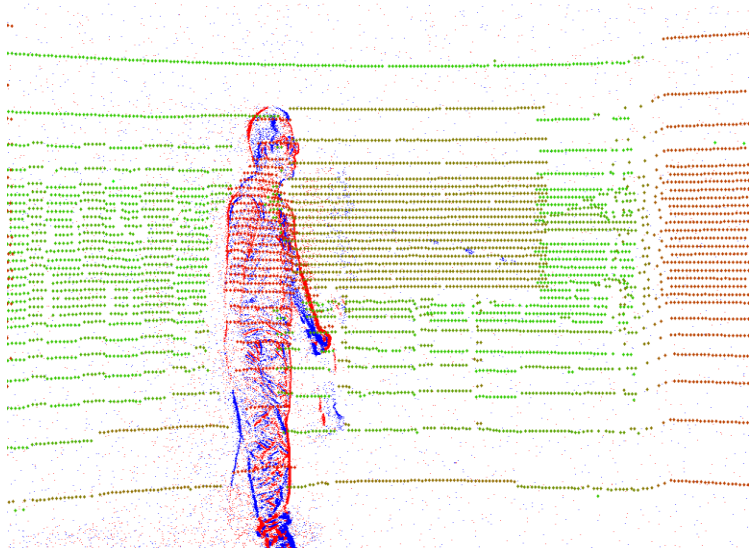
Visual heterogeneous

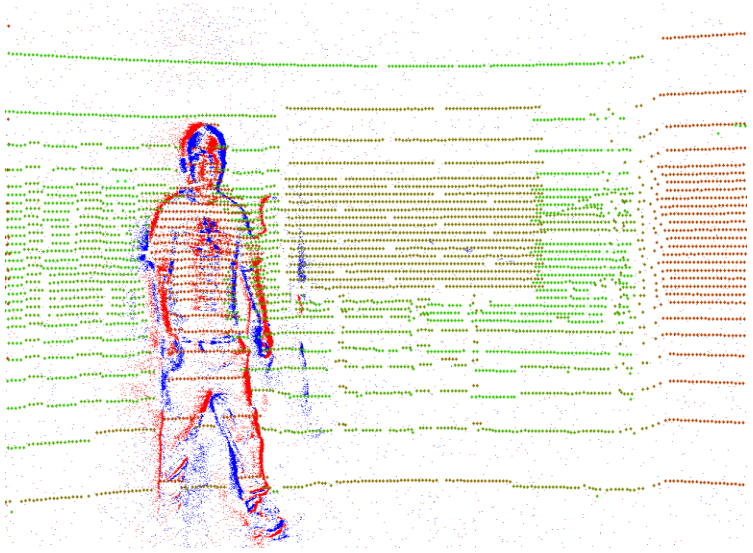
RGB: low dynamic range

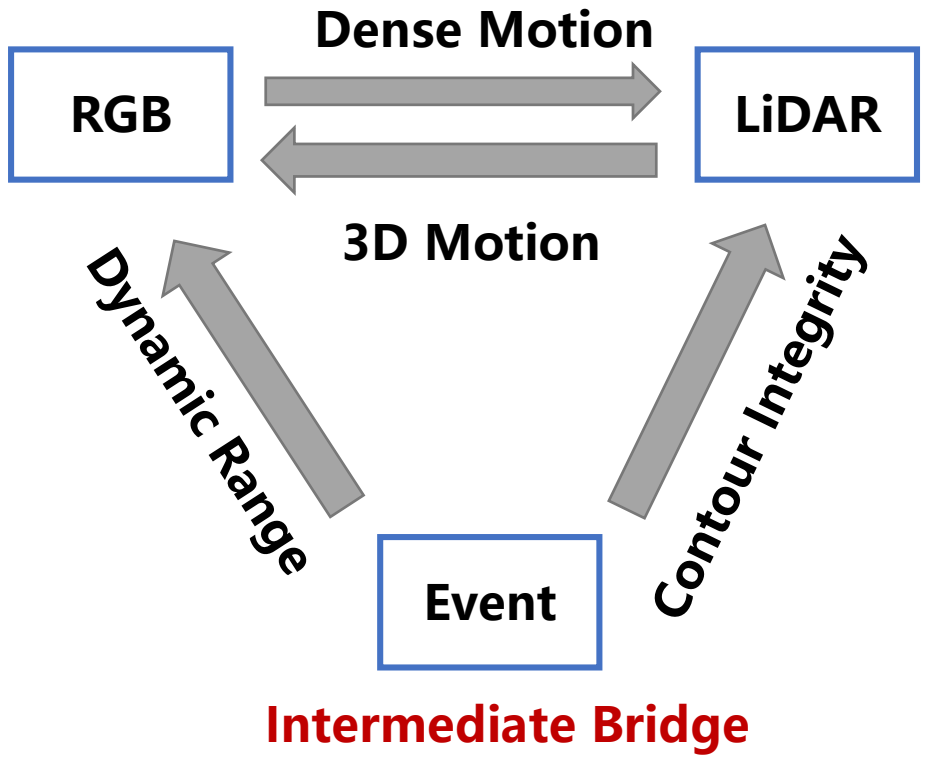
LiDAR: incomplete contour

Motivation



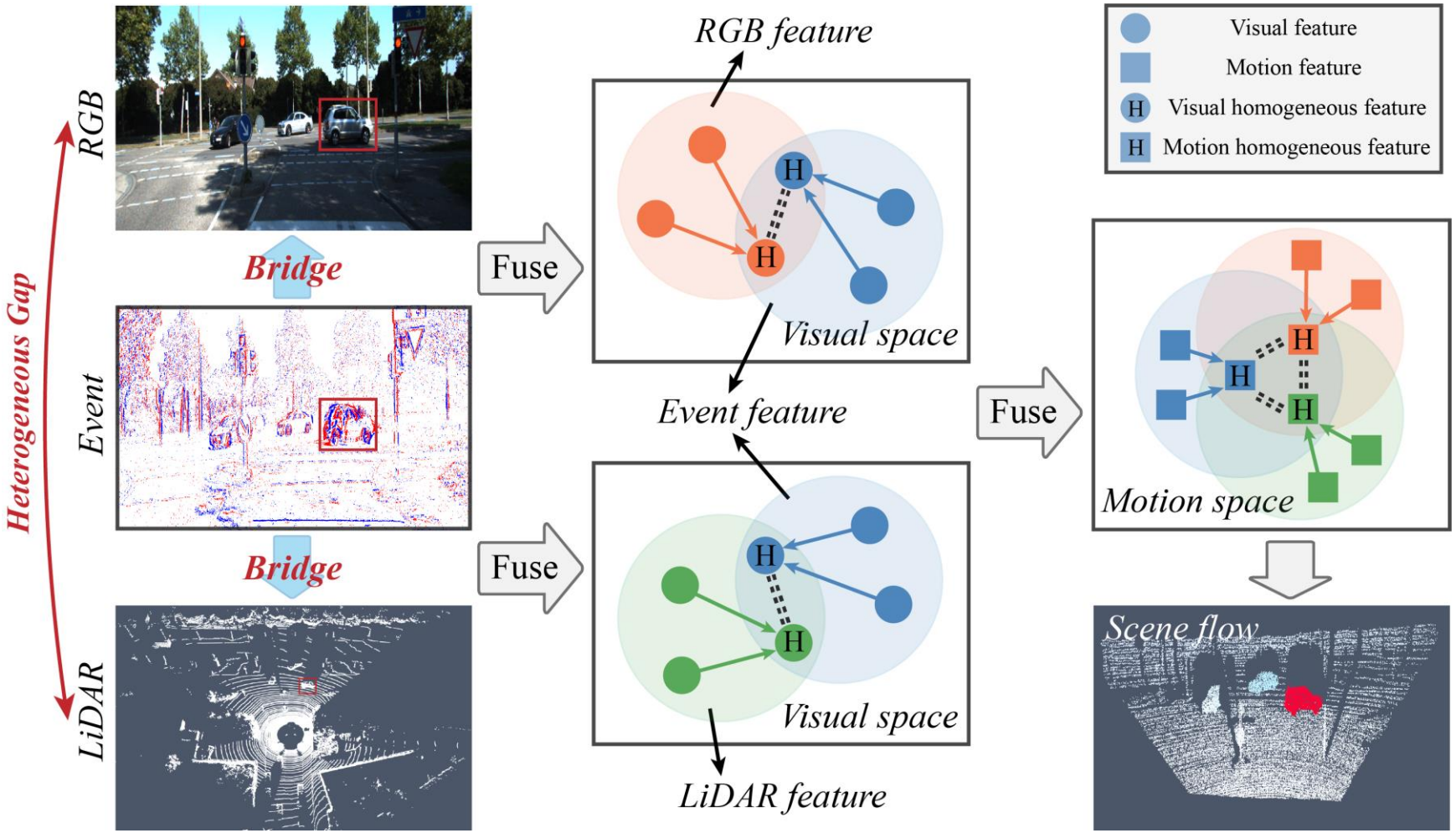






Intermediate Bridge

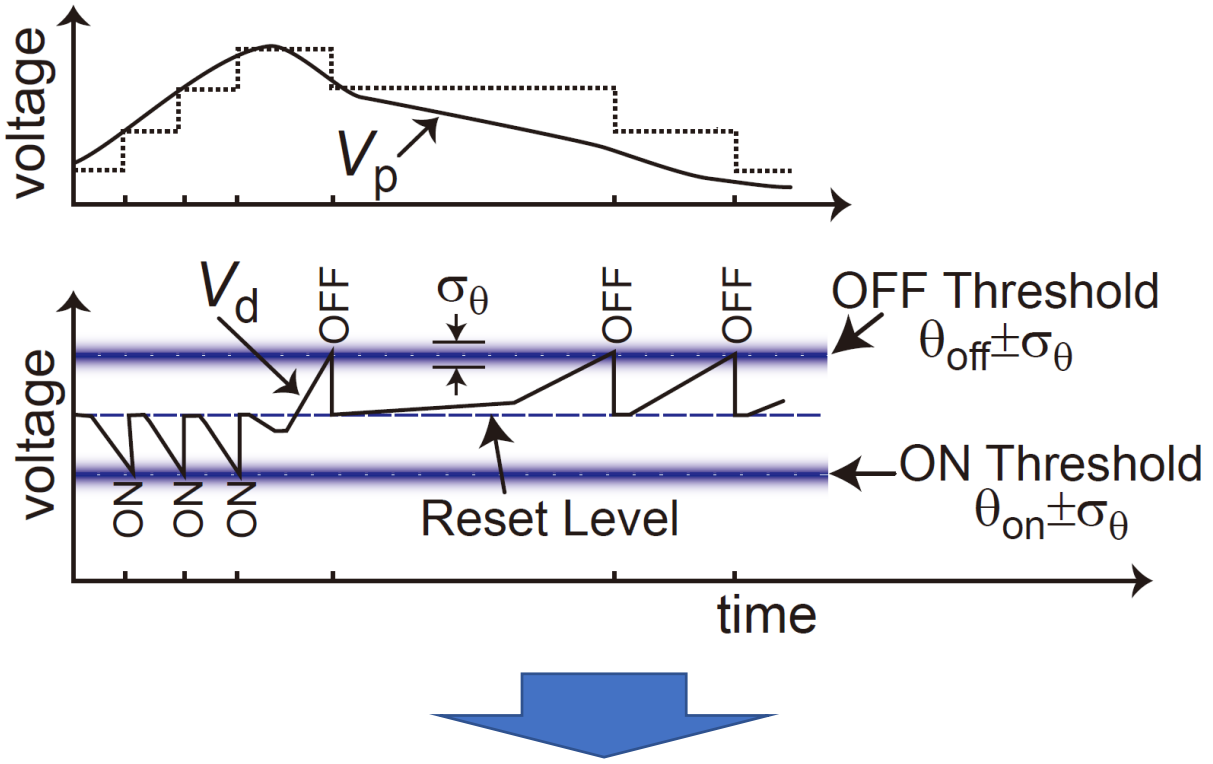
Hierarchical Visual-Motion Fusion

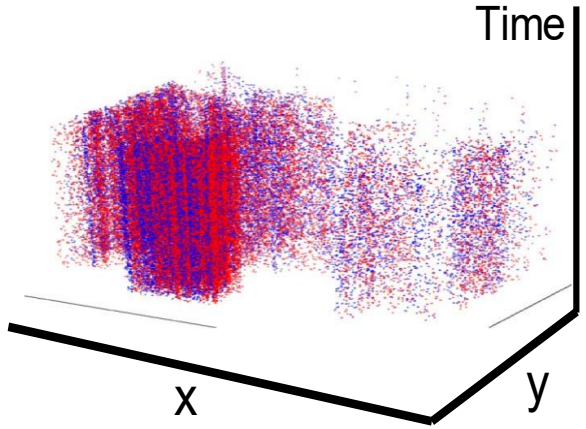


Exploring a homogeneous space to fuse the cross-modal complementary knowledge

Homogeneous Event-RGB Luminance

Absolute Luminance

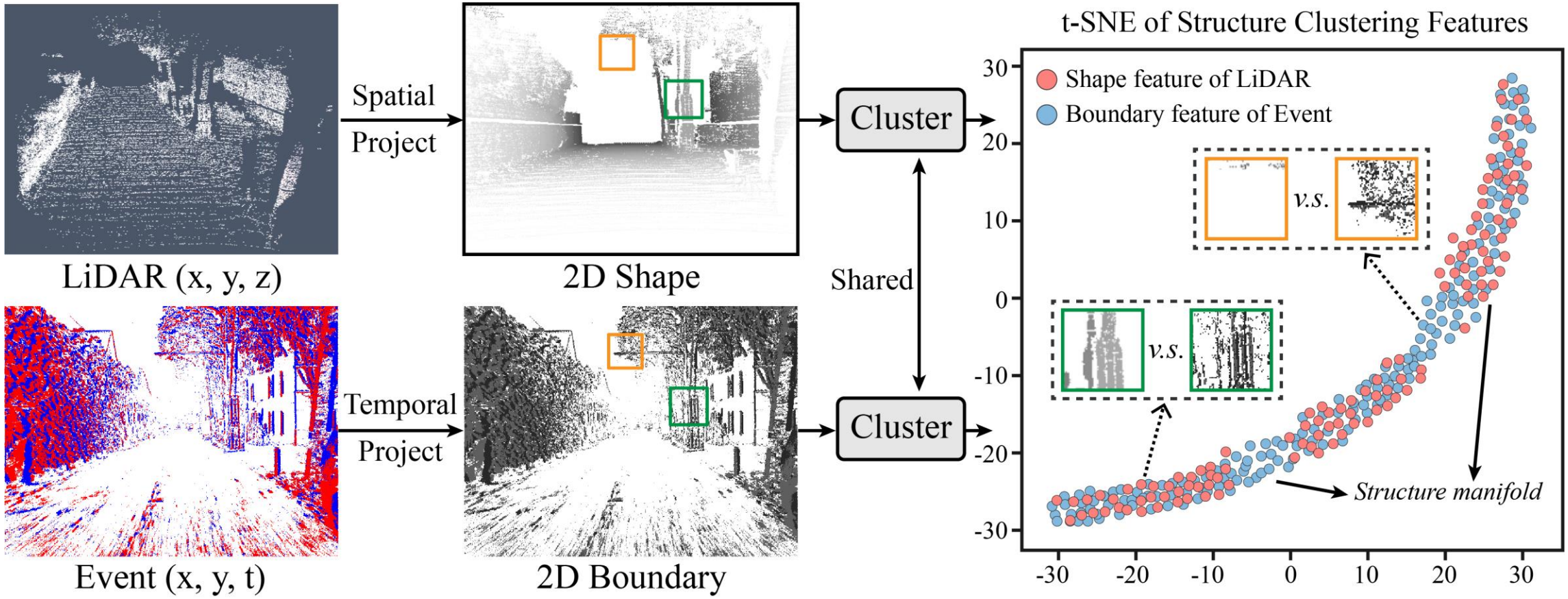




Relative Luminance

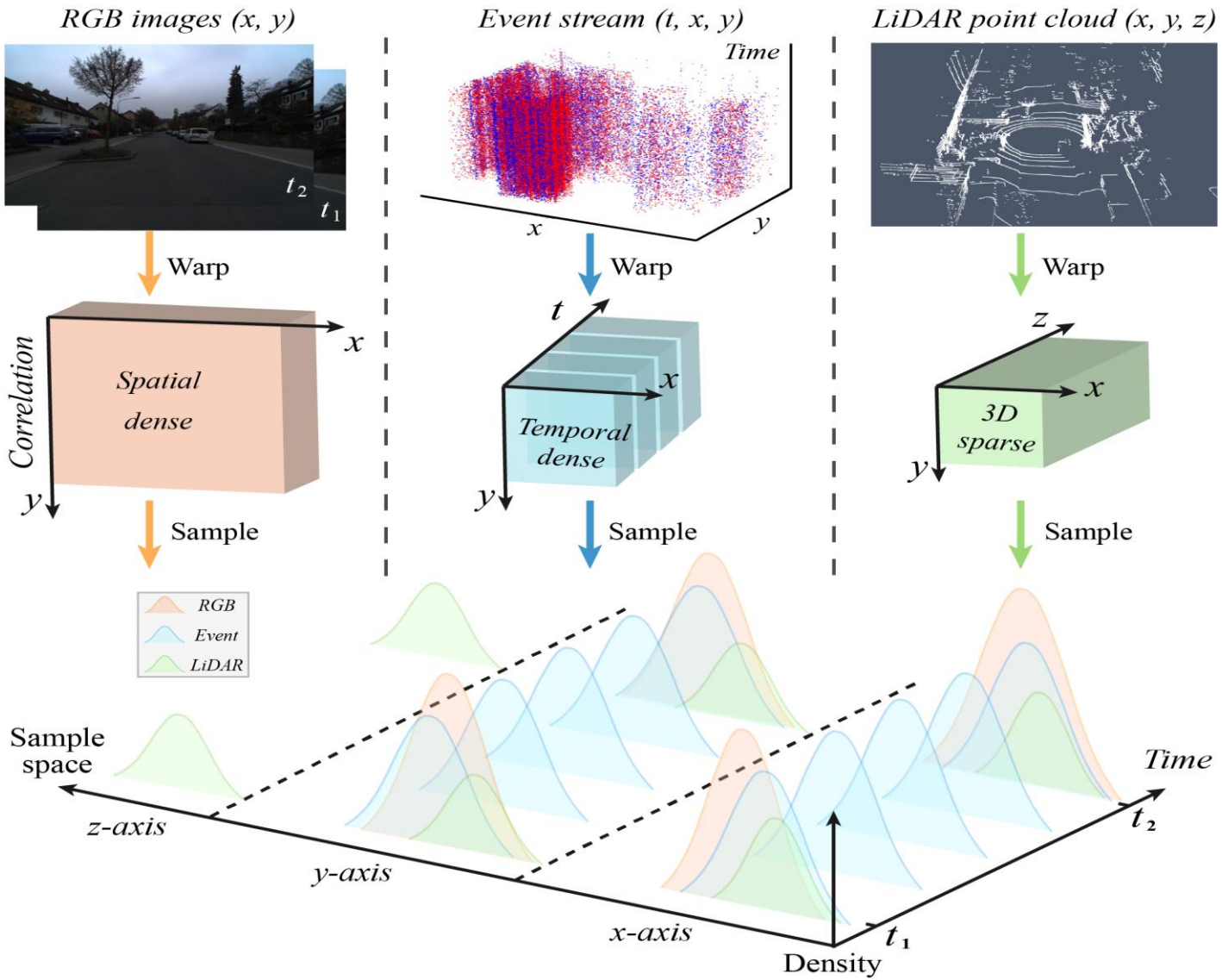
Event can improve the dynamic range of RGB

Homogeneous Event-LiDAR Structure



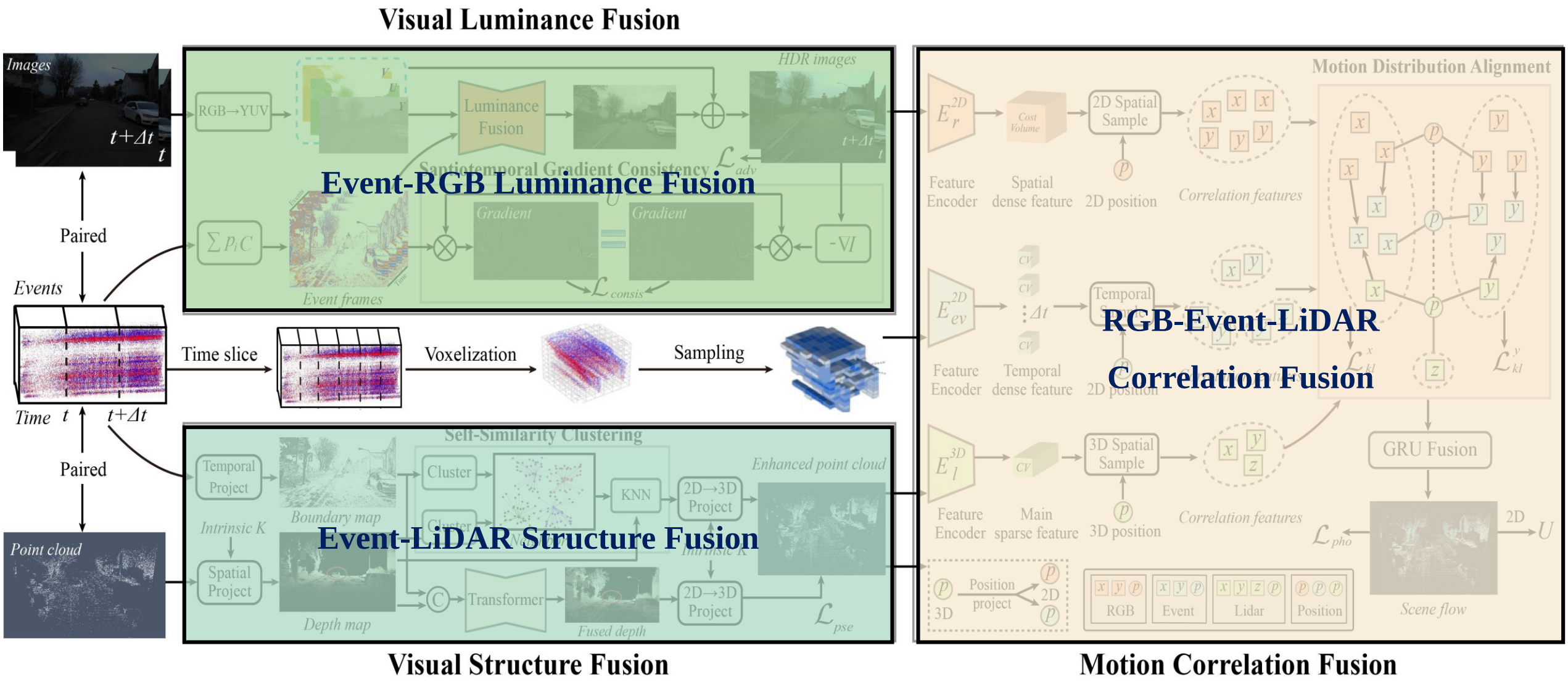
Event benefits to compensating the structure of LiDAR

Homogeneous RGB-Event-LiDAR Correlation



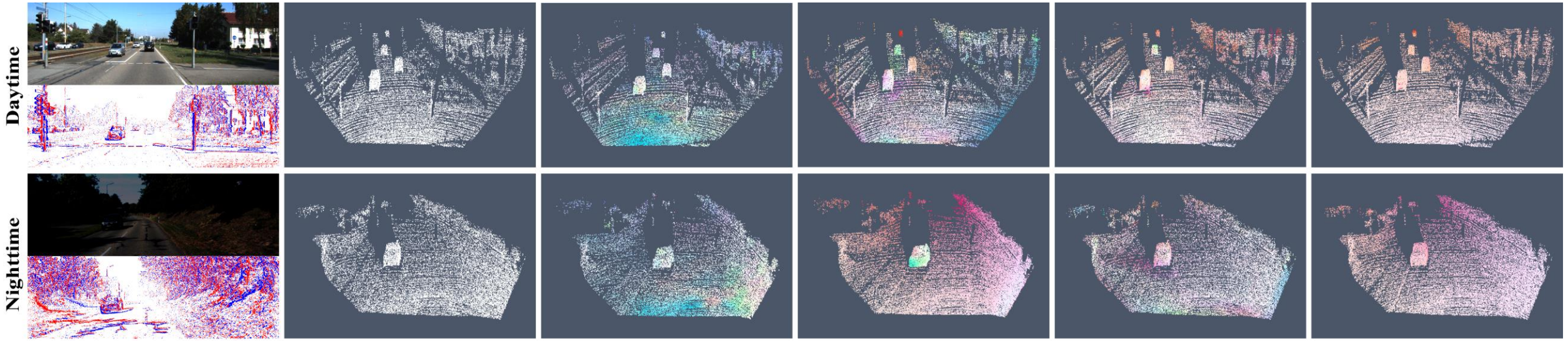
Modality	RGB	Event	LiDAR
t-axis	Sparse	Dense	Sparse
x-axis	Dense	Sparse	Sparse
y-axis	Dense	Sparse	Sparse
z-axis	×	×	√

Fusion Framework



Comparison on Synthetic Event-KITTI

Method		Scene flow methods						Optical flow methods			
		RAFT-3D	RAFT-3D w/ e	PV-RAFT	CamLiFlow	RPEFlow	**VisMoFlow**	SMURF	FlowFormer	RPEFlow	**VisMoFlow**
Input		RGB	RGB	PC	RGB+PC	RGB+PC+EV	**RGB+PC+EV**	RGB	RGB	RGB+PC+EV	**RGB+PC+EV**
Day	EPE	0.095	–	0.055	0.033	0.060	**0.012**	2.01	0.607	0.556	**0.198**
	ACC	73.48%	–	79.96%	91.40%	81.73%	**98.55%**	83.34%	89.08%	88.98%	**97.11%**
Night	EPE	0.112	0.104	0.055	0.047	0.056	**0.027**	11.360	2.085	0.716	**0.353**
	ACC	65.65%	72.65%	79.97%	85.62%	81.47%	**95.62%**	55.12%	77.05%	78.85%	**96.28%**

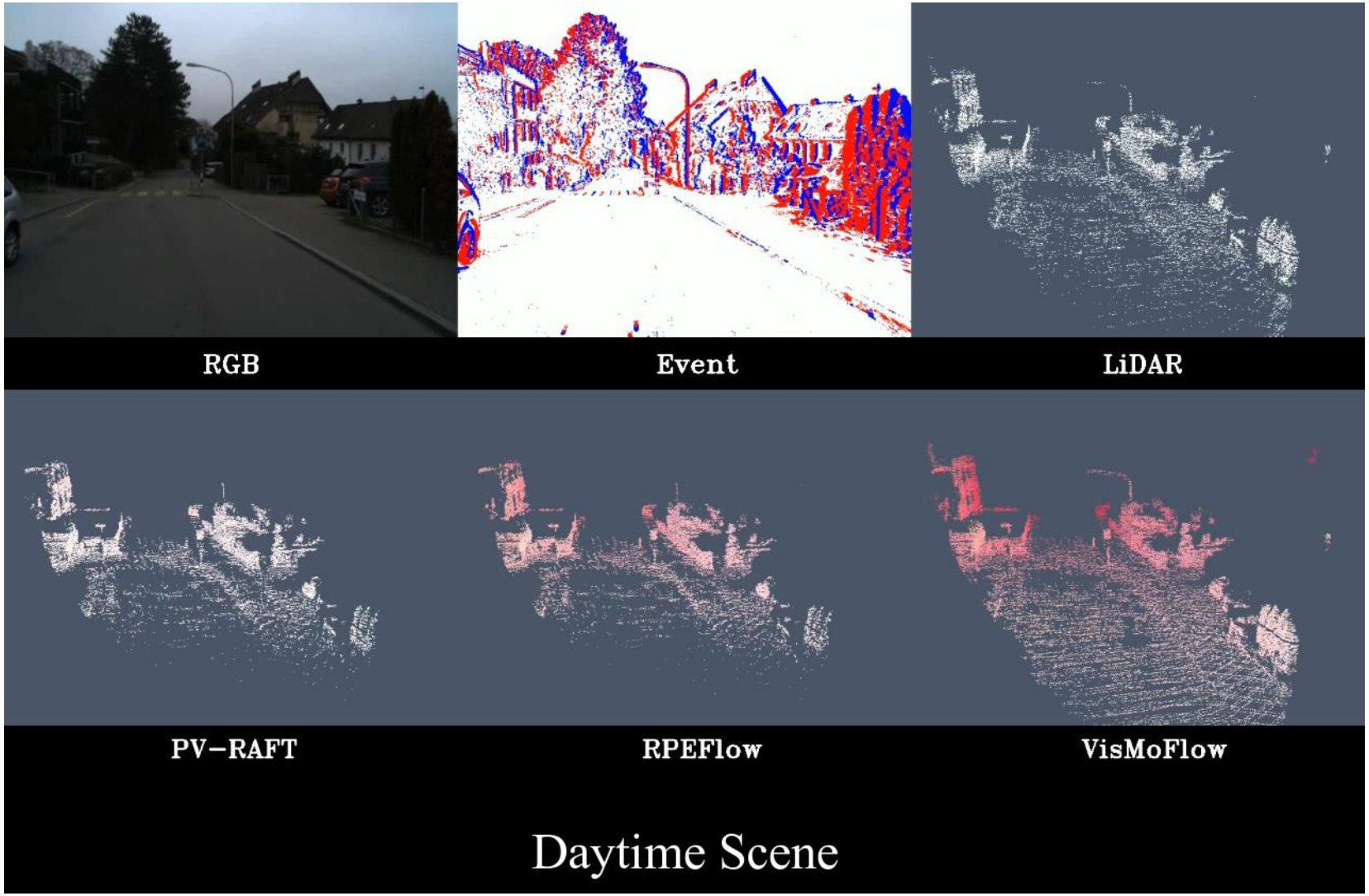


(a) RGB / Event (b) LiDAR (c) RAFT-3D (d) PV-RAFT (e) CamLiFlow (f) VisMoFlow

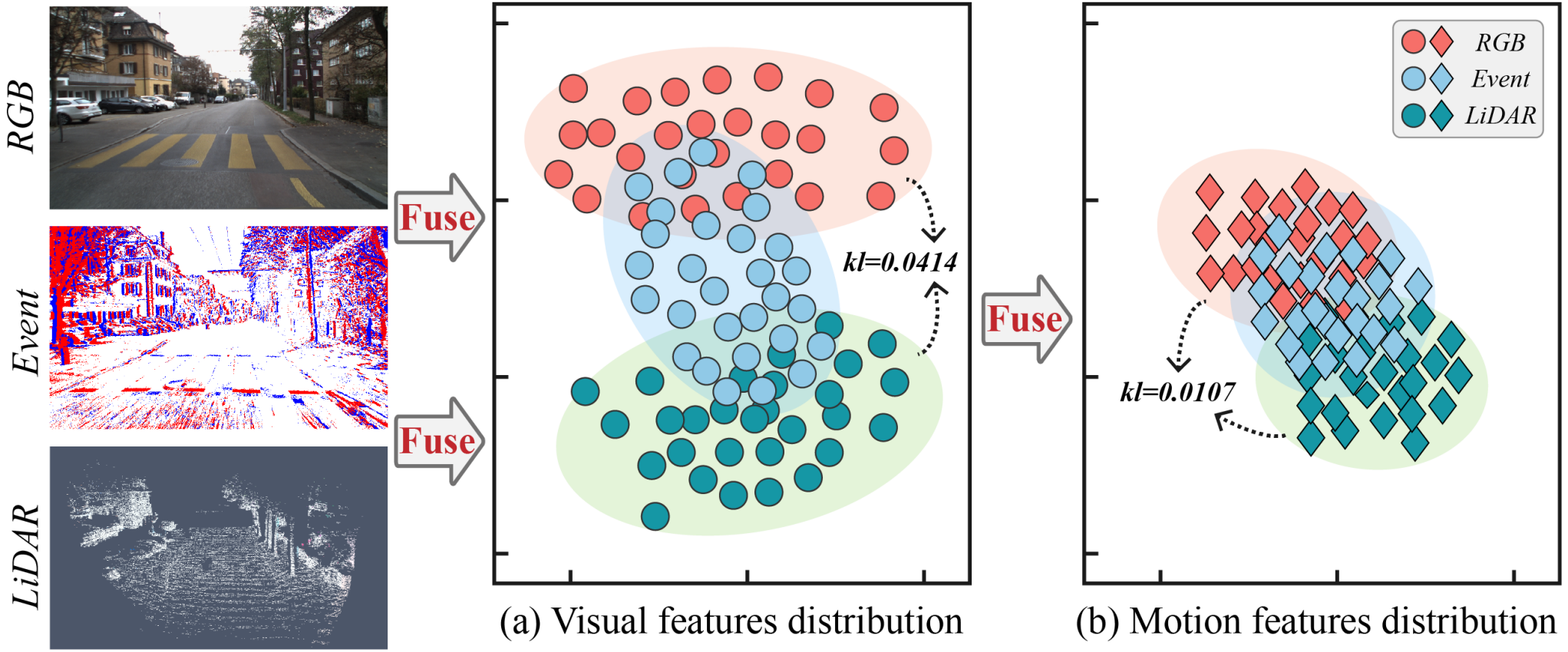
Comparison on Real DSEC

Method		RAFT-3D	PV-RAFT	CamLiFlow	RPEFlow	**VisMoFlow**
Input		RGB	PC	RGB+PC	RGB+PC+EV	**RGB+PC+EV**
Day	EPE	0.167	0.183	0.113	0.103	**0.084**
	ACC	13.16%	37.28%	55.69%	60.81%	**70.34%**
Night	EPE	0.359	0.190	0.125	0.094	**0.090**
	ACC	5.04%	40.98%	53.10%	66.49%	**68.31%**

(a) Multimodal data (b) PV-RAFT (c) CamLiFlow (d) RPEFlow (e) VisMoFlow

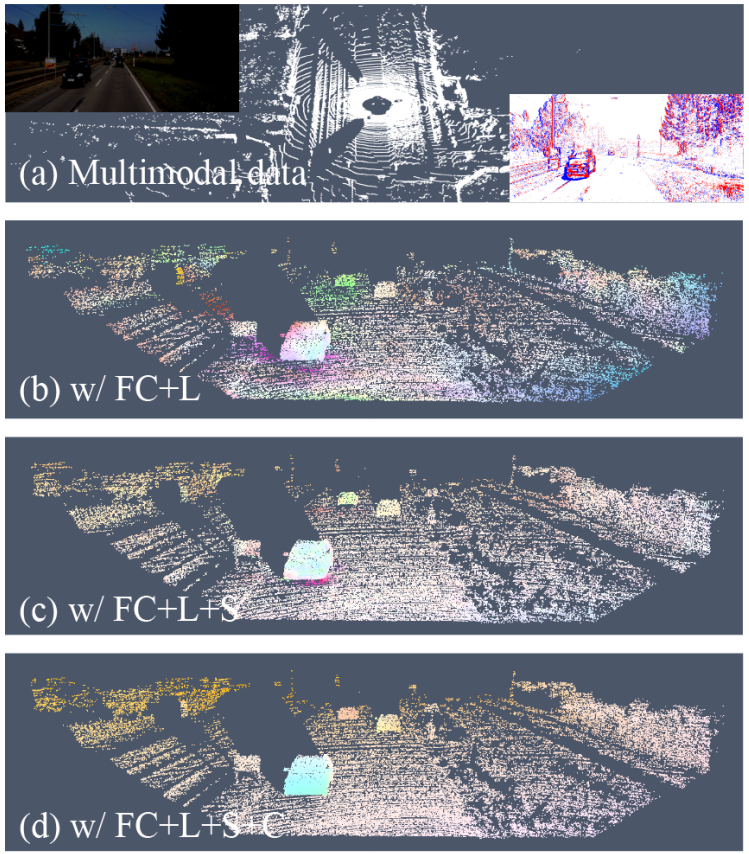

RGB
Event
LiDAR
PV-RAFT
RPEFlow
VisMoFlow
Daytime Scene

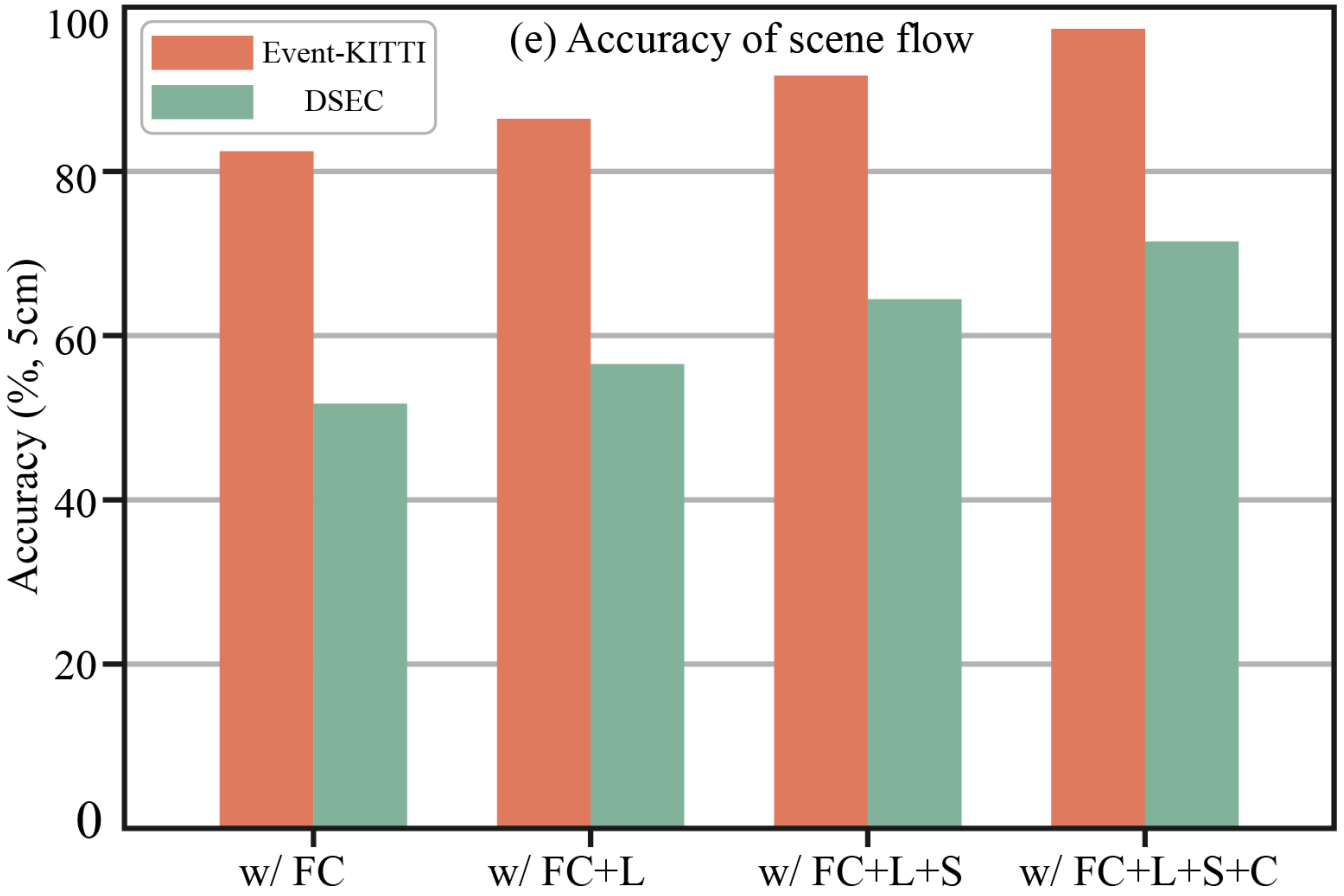
The Role of Event in Fusion



Event plays an RGB-LiDAR bridge, pulling their distributions together

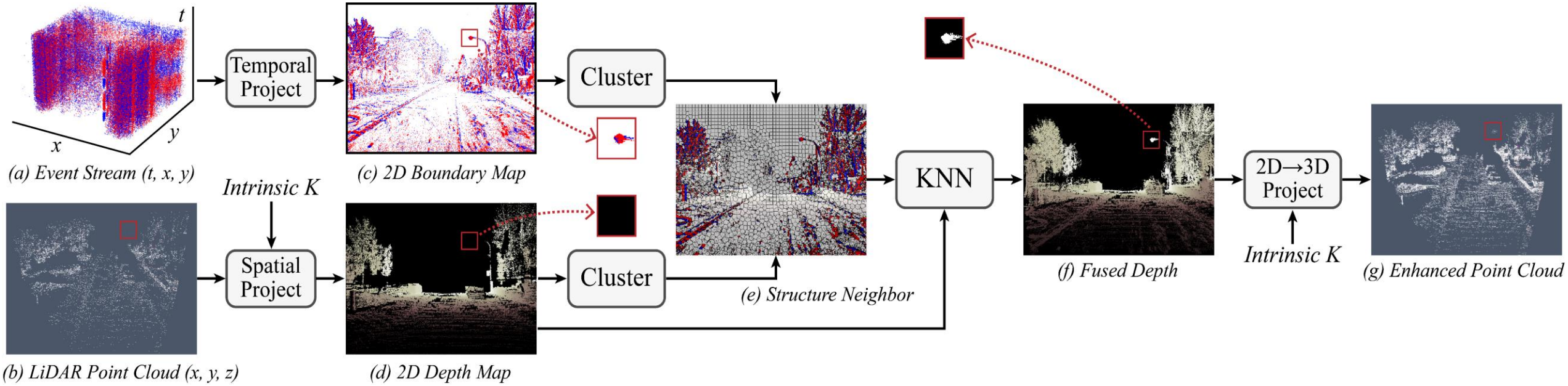
Effectiveness of Homogeneous Space





Homogeneous space can physically facilitate the fusion process

How does Visual Structure Fusion Work?



Ablation Study and Discussion

Fusion strategy	EPE	ACC
w/o any fusion	0.151	49.20%
w/ visual fusion	0.118	54.35%
w/ motion fusion	0.098	62.87%
w/ visual-motion fusion	**0.084**	**70.34%**

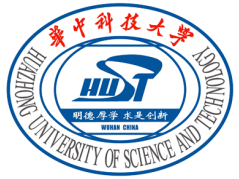
- Ablation study on visual-motion fusion.

Structure neighbor strategy	EPE	ACC
w/o any neighbor	0.095	64.70%
w/ grid-based neighbor	**0.088**	**69.05%**
w/ clustering-based neighbor	0.084	70.34%

- Choice of different structure neighbor strategies.

L_{consis}	L_{pse}	L_{corr}^{kl}	EPE	ACC
×	×	×	**0.122**	**53.17%**
√	×	×	**0.112**	**56.41%**
×	√	×	**0.107**	**58.25%**
×	×	√	**0.092**	**65.43%**
√	√	√	**0.084**	**70.34%**

- Ablation study on fusion losses.


HUAZHONG UNIVERSITY OF SCIENCE AND TECHNOLOGY
HUST
WUHAN CHINA


SEATTLE, WA
CVPR
JUNE 17-21, 2024

Thanks