Dynamic 3D Scene Analysis by Point Cloud Accumulation



Shengyu Huang[1] , Zan Gojcic[2] , Jiahui Huang[3], Andreas Wieser[1], Konrad Schindler[1]

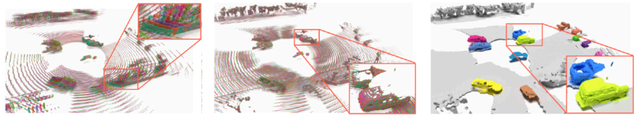
ETH Zürich[1] NVIDIA[2] BRCist[3]


ECCV TEL AVIV 2022

Project page

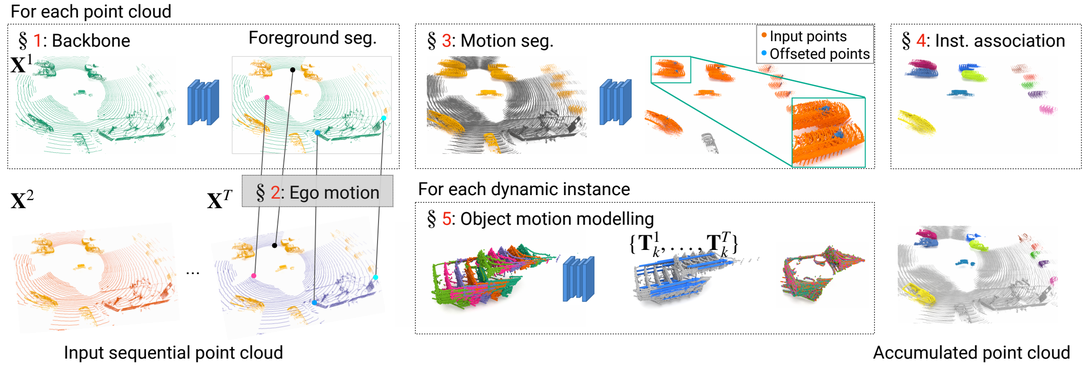
Problem statement

- Given an *ordered* sequence of T point cloud frames (a), the goal is to estimate the flow vectors that align each of the *source* frames to the *target* frame, and hence accumulate the point clouds (b). Such accumulation can serve as a pre-processing step to benefit downstream tasks like surface reconstruction (c).



(a) (b) (c)

Methodology



- We first estimate the ego motion (§2) using predicted *background* points. Then we segment the foreground points into *static* and *dynamic* parts through motion segmentation (§3). The instance IDs of the dynamic parts are obtained from spatio-temporal association (§4). Built upon the scene *rigidity*, we explain the flow vectors of the static parts by estimated ego-motion, and that of the dynamic instances by regressed per-instance rigid motion (§5).

- Loss = $L_{\text{ego}} + L_{\text{FG}} + L_{\text{motion}} + L_{\text{offset}} + L_{\text{obj}}$

Experimental results

- SoTA results on multi-frame scene flow evaluation.

Dataset	Method	Static part EPE avg.↓	Static part AccS↑	Static part AccR↑	Static part ROutlier↓	Dynamic foreground EPE avg. ↓	Dynamic foreground EPE med.↓	Dynamic foreground AccS↑	Dynamic foreground AccR↑	Dynamic foreground ROutliers ↓
Waymo	PPWC-Net [63]	0.414	17.6	40.2	12.1	0.475	0.201	9.0	29.3	22.4
	FLOT [44]	0.129	65.2	85.0	2.8	0.625	0.231	9.8	27.4	33.8
	WsRSF [19]	0.063	87.3	96.6	0.6	0.381	0.094	31.3	64.0	10.1
	NSFPrior [34]	0.187	79.8	89.1	4.7	0.237	0.077	44.7	68.6	11.5
	Ours	0.028	97.5	99.5	0.1	0.197	0.062	53.3	77.5	5.9
	Ours+	**0.018**	**99.0**	**99.7**	**0.1**	**0.173**	**0.043**	**69.1**	**86.9**	**5.1**
	Ours (w. ground)	0.042	91.9	98.8	0.1	0.219	0.071	47.1	72.8	8.5
nuScenes	PPWC-Net [63]	0.316	16.1	37.0	8.7	0.661	0.307	7.6	24.2	31.9
	FLOT [44]	0.153	51.7	78.3	4.3	1.216	0.710	3.0	10.3	63.9
	WsRSF [19]	0.195	57.4	82.6	4.8	0.539	0.204	17.9	37.4	22.9
	NSFPrior [34]	0.584	38.9	56.7	26.9	0.707	0.222	19.3	37.8	32.0
	Ours	0.111	65.4	88.6	1.1	0.301	0.146	26.6	53.4	**12.1**
	Ours+	**0.091**	**72.8**	**91.9**	**0.9**	**0.301**	**0.135**	**32.7**	**56.7**	13.7
	Ours (w. ground)	0.134	55.3	83.8	1.9	0.37	0.182	18.2	43.8	17.5

- Generalisation across different numbers of input frames on *Waymo* dataset.

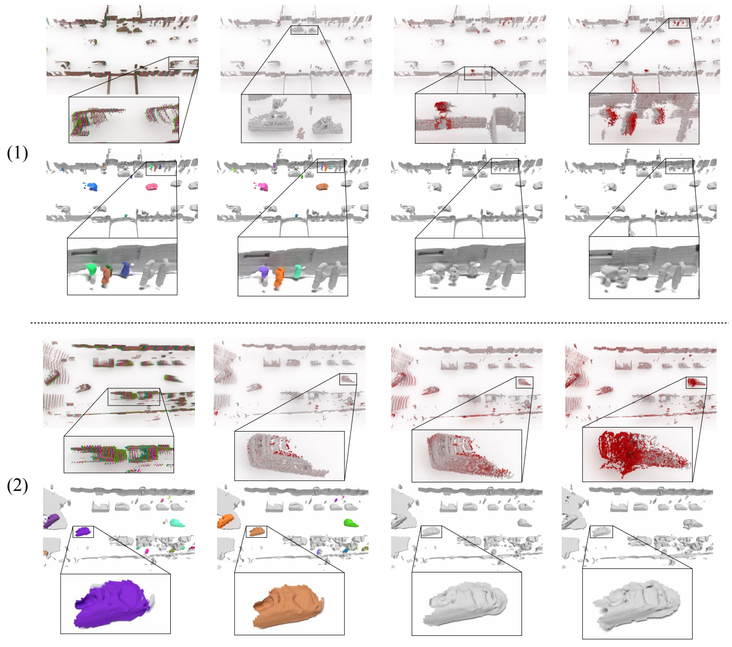
	3	4	5	6	7	8	9	10
static EPE avg.	**0.022**	0.025	0.028	0.032	0.037	0.044	0.054	0.066
dynamic EPE avg.	0.199	**0.168**	0.190	0.218	0.250	0.294	0.348	0.412

- Superior runtime per 5-frame (*Waymo*) and 11-frame (*nuScenes*) sample.

	Waymo	*nuScenes*
PPWC-Net [63]	0.608	0.990
FLOT [44]	1.028	2.010
WsRSF [19]	1.252	1.460
NSFPrior [34]	212.256	63.460
Ours	**0.174**	**0.250**

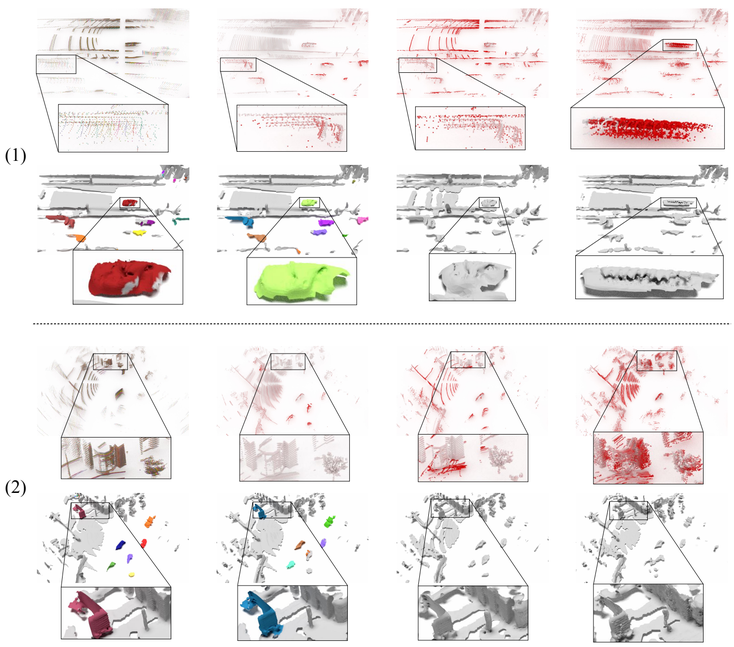
	Waymo	*nuScenes*
ego-motion estimation	0.100	0.188
motion segmentation	0.024	0.040
instance association	0.036	0.009
TubeNet	0.014	0.013

- Surface reconstruction on *Waymo* dataset.



(a) Input & GT reconstruction (b) Ours (c) NSFPrior (d) FLOT

- Surface reconstruction on *nuScenes* dataset.



(a) Input & GT reconstruction (b) Ours (c) NSFPrior (d) FLOT