CVPR JUNE 3-7, 2026 DENVER COLORADO

SEA-Flow3D: Simplified, Efficient, and Accurate Scene Flow via Spatial Vector Sampling and Multi-scale Refinement

Han Ling[1,3], Quansen Sun[1], Yinghua Yao[3], Ivor Tsang[2,3], Yinghui Sun[4]

[1]Nanjing University of Science and Technology [2]Nanyang Technological University
[3]IHPC & CFAR A*STAR, Singapore [4]Southeast University

南京理工大学 NANJING UNIVERSITY OF SCIENCE & TECHNOLOGY

NANYANG TECHNOLOGICAL UNIVERSITY SINGAPORE

Centre for Frontier AI Research A*STAR CFAR

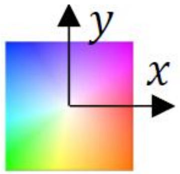
東南大學 SOUTHEAST UNIVERSITY

Introduction

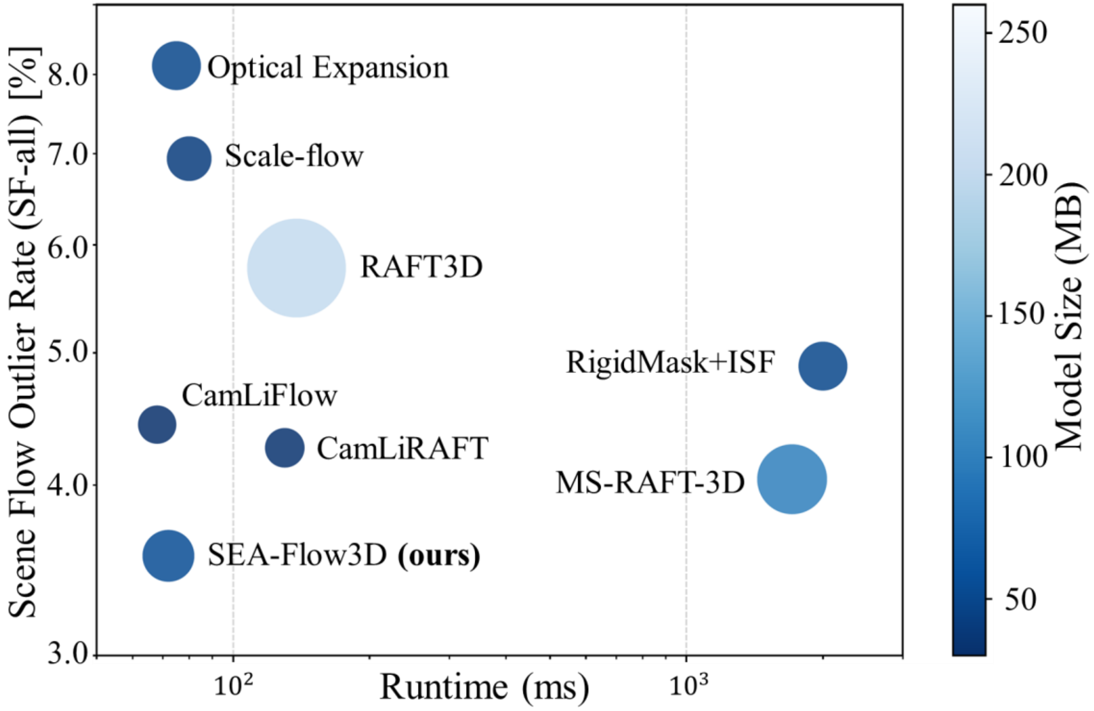
Frame 1

Frame 2

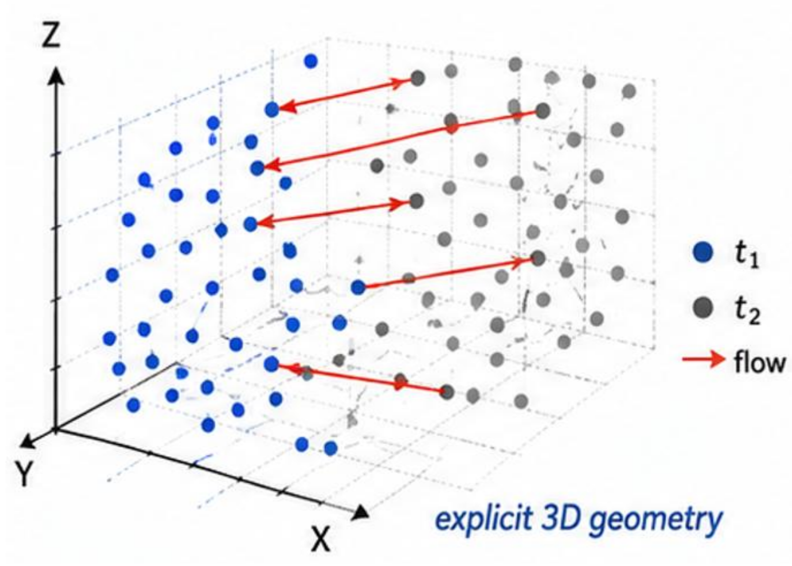
Optical Flow



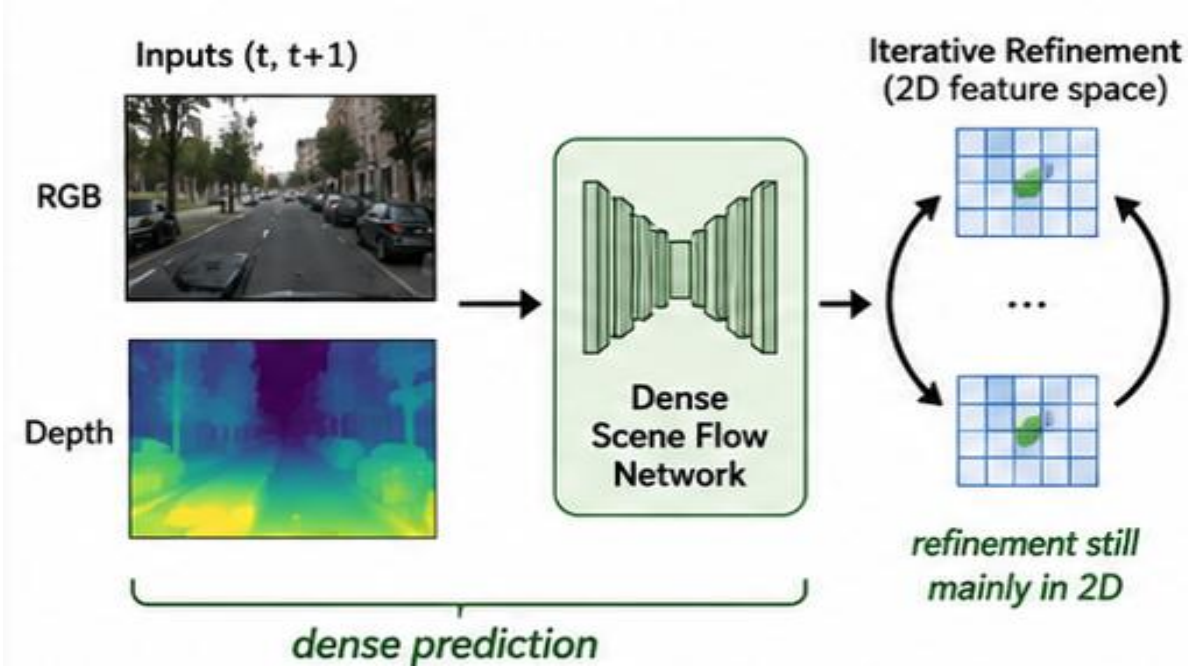
Depth change



Introduction



Point-cloud Scene Flow

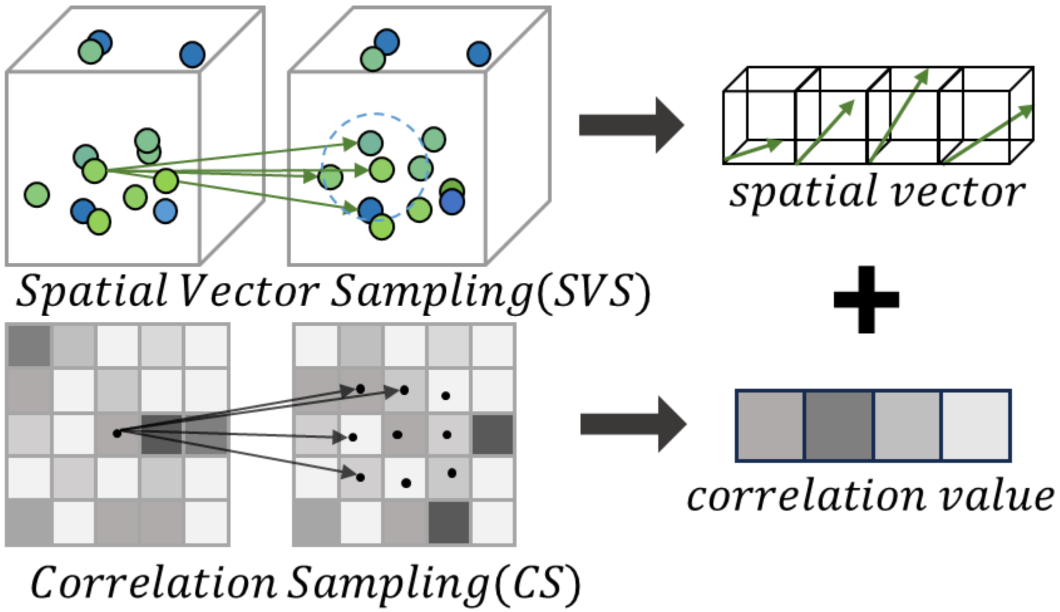


RGB-D Dense Scene Flow

Problems:

- Dense 3D motion estimation is **challenging in complex dynamic scenes**.
- Point-based methods exploit geometry but suffer from **sparse and low-resolution outputs**.
- Dense RGB-D methods still optimize mainly in **2D correlation space**.
- Depth cues are not continuously used during **iterative motion refinement.**
- Rigid optimization is fragile in **non-rigid scenes and increases computation.**

Spatial Vector Sampling

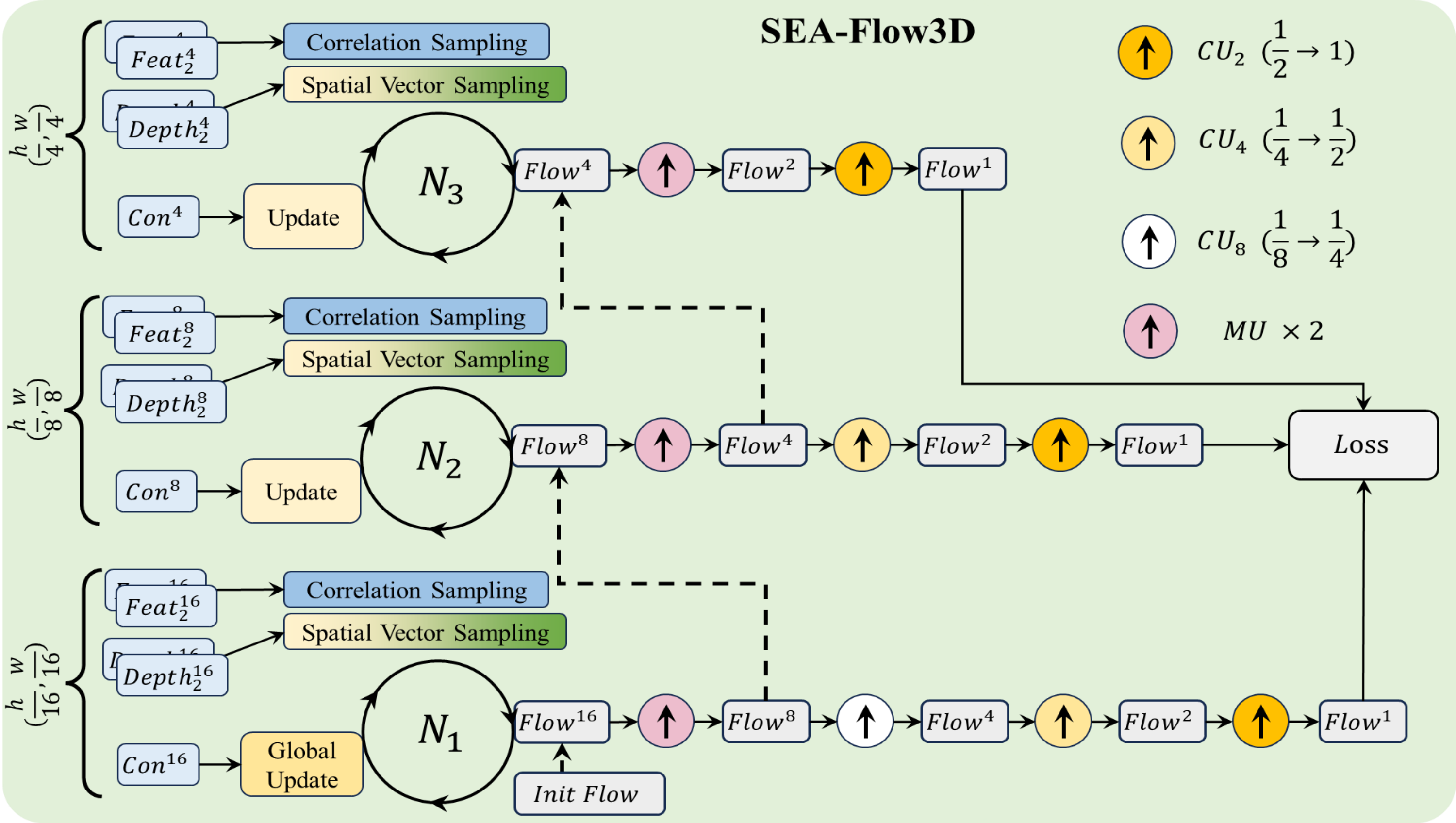


Projection Field: We first construct a disparity-augmented projection field for the second frame, where each pixel is represented as (u, v, D_2)**.**

Local Sampling: Given the current optical flow estimate, each source pixel is projected to the target frame, and SVS samples a local neighborhood around this projected position.

Spatial Vector: For each sampled candidate, SVS computes a **3D spatial vector**. which is **paired with the corresponding 2D correlation cost** for scene flow refinement.

Multi-scale Refinement Framework



Multi-scale Inference: SEA-Flow3D starts refinement from the coarse **1/16 scale**, then progressively updates scene flow at **1/8 and 1/4 scales**, resolving global motion first and local details later.

Context Upsampling: We use context-guided upsampling masks to propagate predictions across scales and recover full-resolution motion fields with minimal extra cost.

Results: Comparison on the KITTI Scene Flow Benchmark.

Method	Depth Error				Optical Flow Error			Scene Flow Error		
	D1-all	D2-bg	D2-fg	D2-all	Fl-bg	Fl-fg	Fl-all	SF-bg	SF-fg	SF-all
DRISF [17]	2.55	2.90	9.73	4.04	3.59	10.40	4.73	4.39	15.94	6.31
ISF [2]	4.46	4.88	11.34	5.95	5.40	10.29	6.22	6.58	15.63	8.08
OpticalExp [33]	1.81	3.39	8.54	4.25	5.83	8.66	6.30	7.06	13.44	8.12
BinaryTTC [1]	1.81	3.84	9.39	4.76	5.84	8.67	6.31	7.45	13.74	8.50
Scale-Flow [12]	1.81	2.55	8.24	3.50	5.24	5.71	5.32	6.06	11.32	6.94
RigidMask [34]	1.89	2.09	8.92	3.23	2.63	7.85	3.50	3.25	13.08	4.89
CamLiRAFT [14]	1.81	1.91	8.11	2.94	2.08	7.37	2.96	2.68	12.16	4.26
CamLiFlow [13]	1.81	1.92	8.14	2.95	2.31	7.04	3.10	2.87	12.23	4.43
RAFT-3D [26]	1.81	2.51	9.46	3.67	3.39	8.79	4.29	4.27	13.27	5.77
MS-RAFT-3D [21]	1.59	1.91	6.17	2.68	2.43	5.75	2.98	2.95	9.47	4.04
SF2SE3 [22]	1.65	2.20	7.66	3.11	3.17	8.79	4.11	3.75	13.15	5.32
M-FUSE [19]	1.65	2.14	8.10	3.13	2.66	7.47	3.46	3.43	11.84	4.83
SEA-Flow3D+G (Ours)	1.81	1.90	7.94	2.91	2.08	6.95	2.89	2.63	11.91	4.17
SEA-Flow3D+M (Ours)	**1.42**	**1.50**	**5.59**	**2.18**	**1.98**	**5.30**	**2.53**	**2.43**	**9.14**	**3.55**

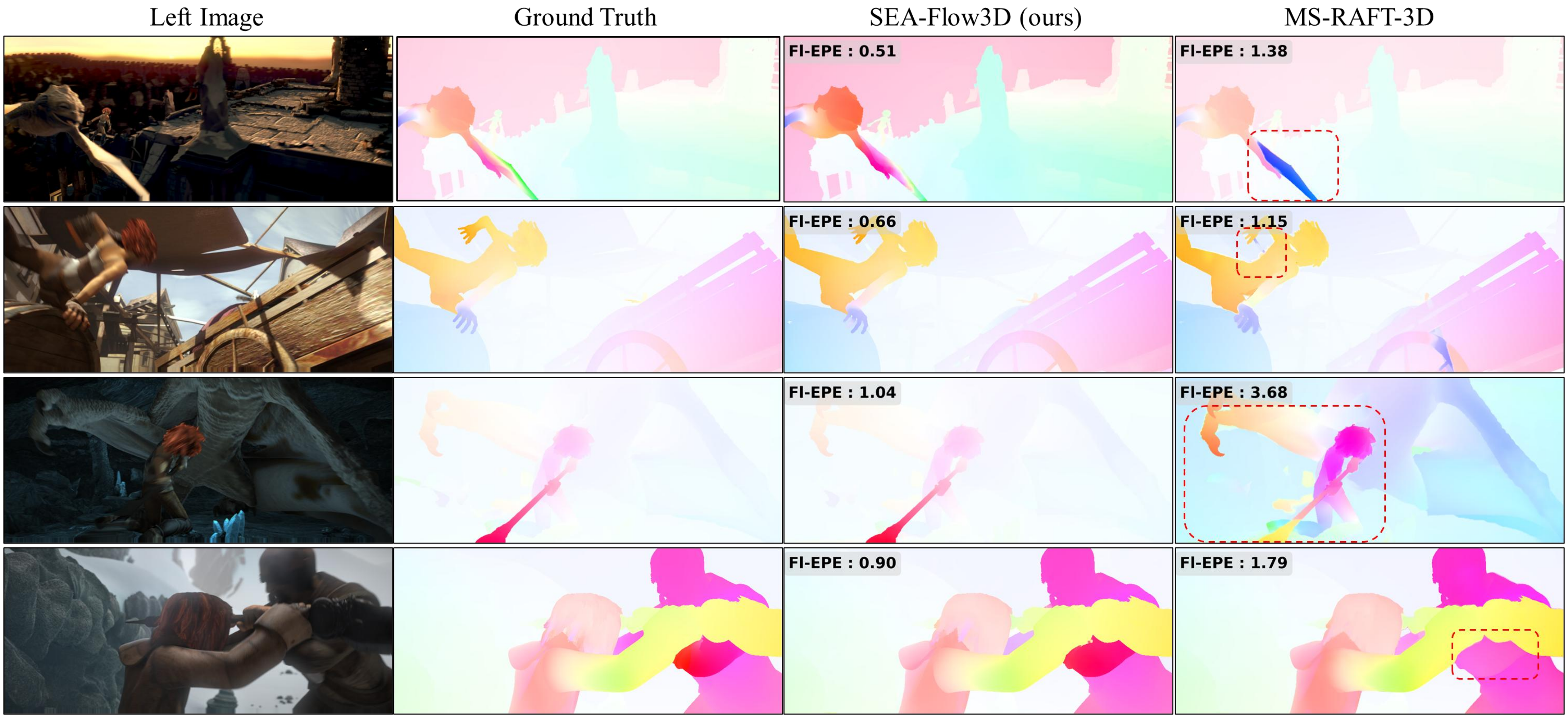
State of the art 3D Motion accuracy among all methods

Results : Sintel Benchmark.

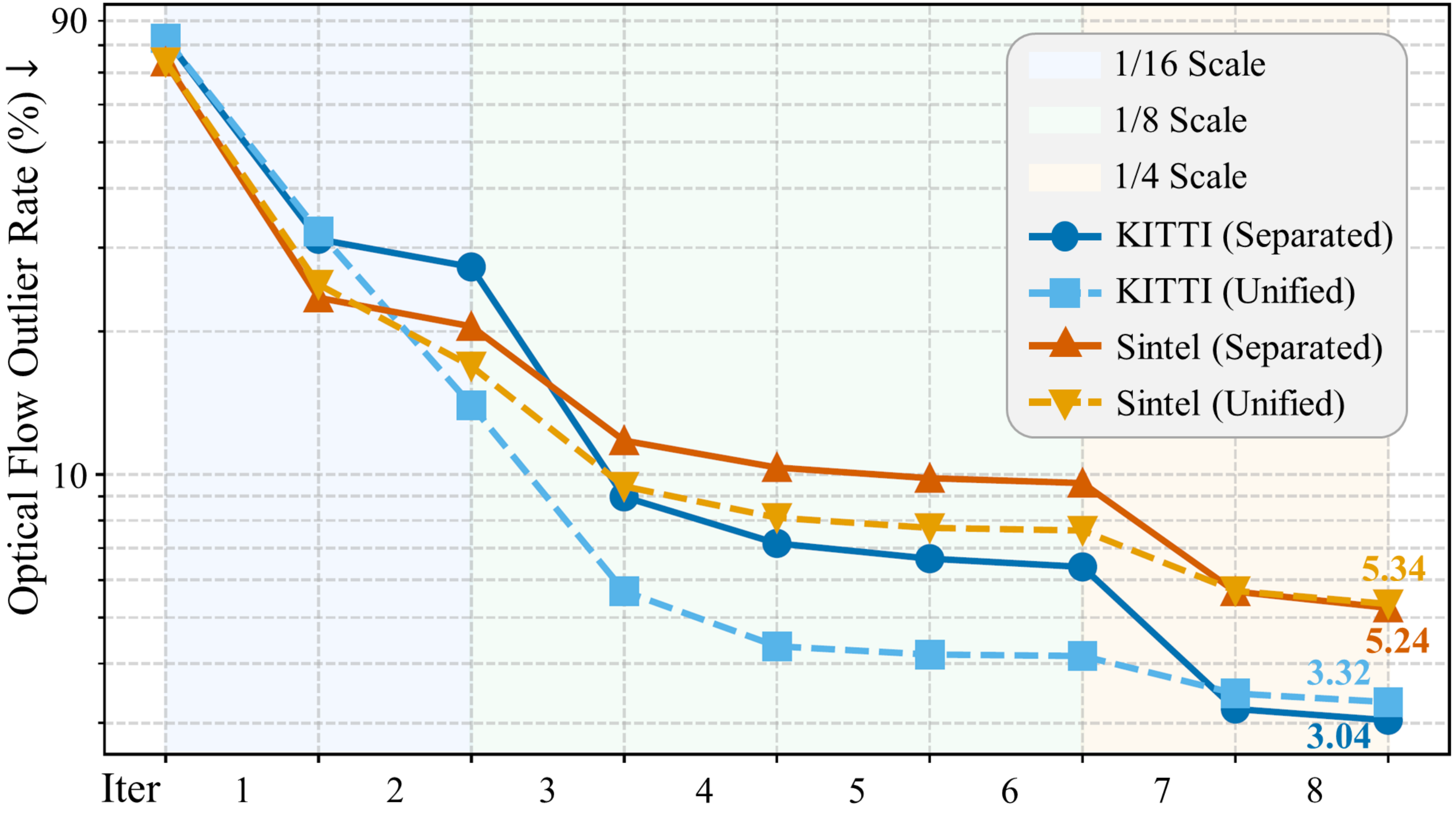
Method	Data	Input	Time/ms	**Sintel**	
				Clean	**Final**
PWC-Net [23]	C+T	RGB	15	2.55	3.93
FlowNet2 [8]	C+T	RGB	30	2.02	3.54
RAFT [25]	C+T	RGB	42	1.43	2.71
CamLiPWC [13]	T	RGB+P	68	1.79	2.55
CamLiRAFT [14]	T	RGB+P	130	1.27	2.38
RAFT-3D [26]	T	RGB+D	138	1.75	2.91
MS-RAFT-3D [21]	T	RGB+D	1710	1.06	2.22
SEA-Flow3D (ours)	T	RGB+D	72	**1.04**	**2.04**

20× faster inference than previous MS-RAFT-3D.

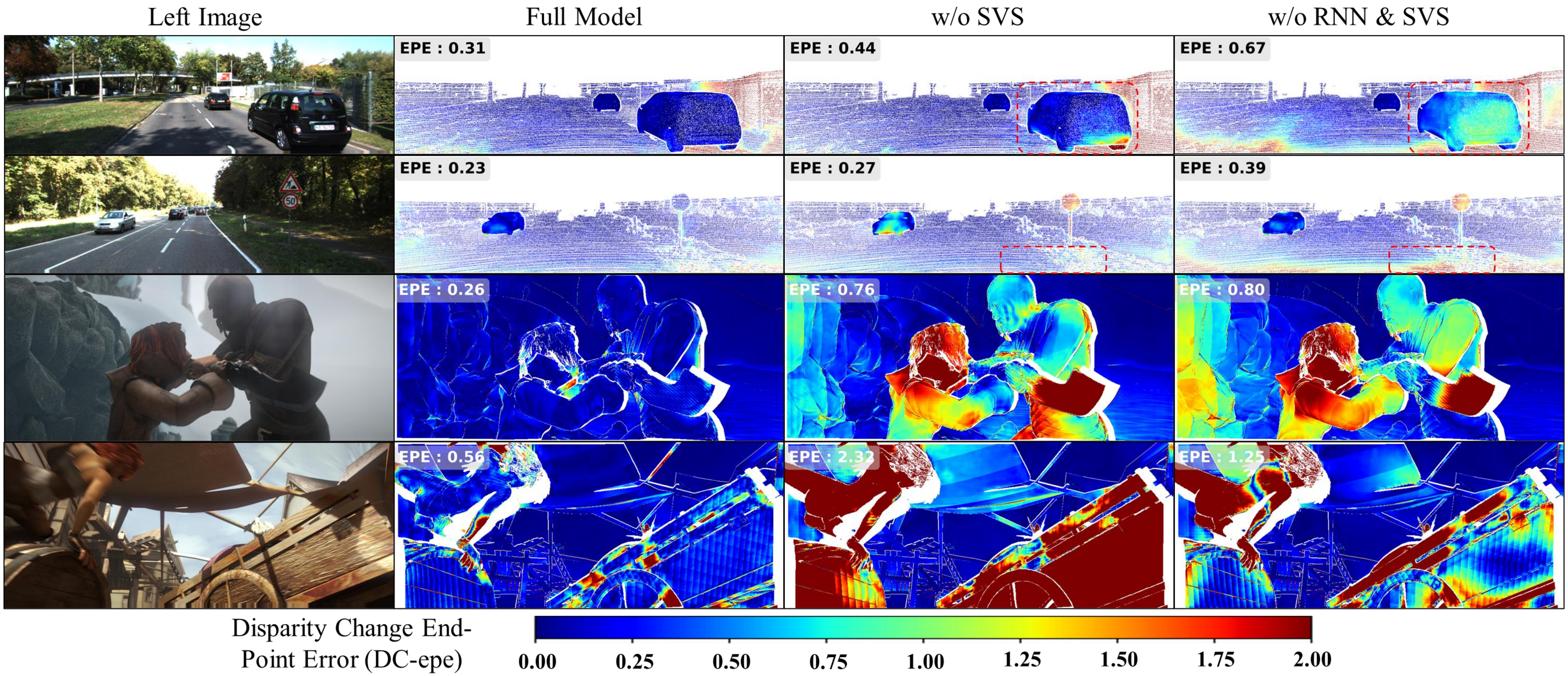
Results: Visualization comparison on the non-rigid Sintel scenes



Results : Ablation Study



Results : Ablation Study



With SVS introduced, the model significantly improves depth motion estimation.

Results:More experimental results

Left Image

Optical Flow

Depth change