LayeredFlow: A Real-World Benchmark for Non-Lambertian Multi-Layer Optical Flow

Hongyu Wen, Erich Liang, Jia Deng

PRINCETON UNIVERSITY

PRINCETON VISION & LEARNING LAB

ECCV

Motivation

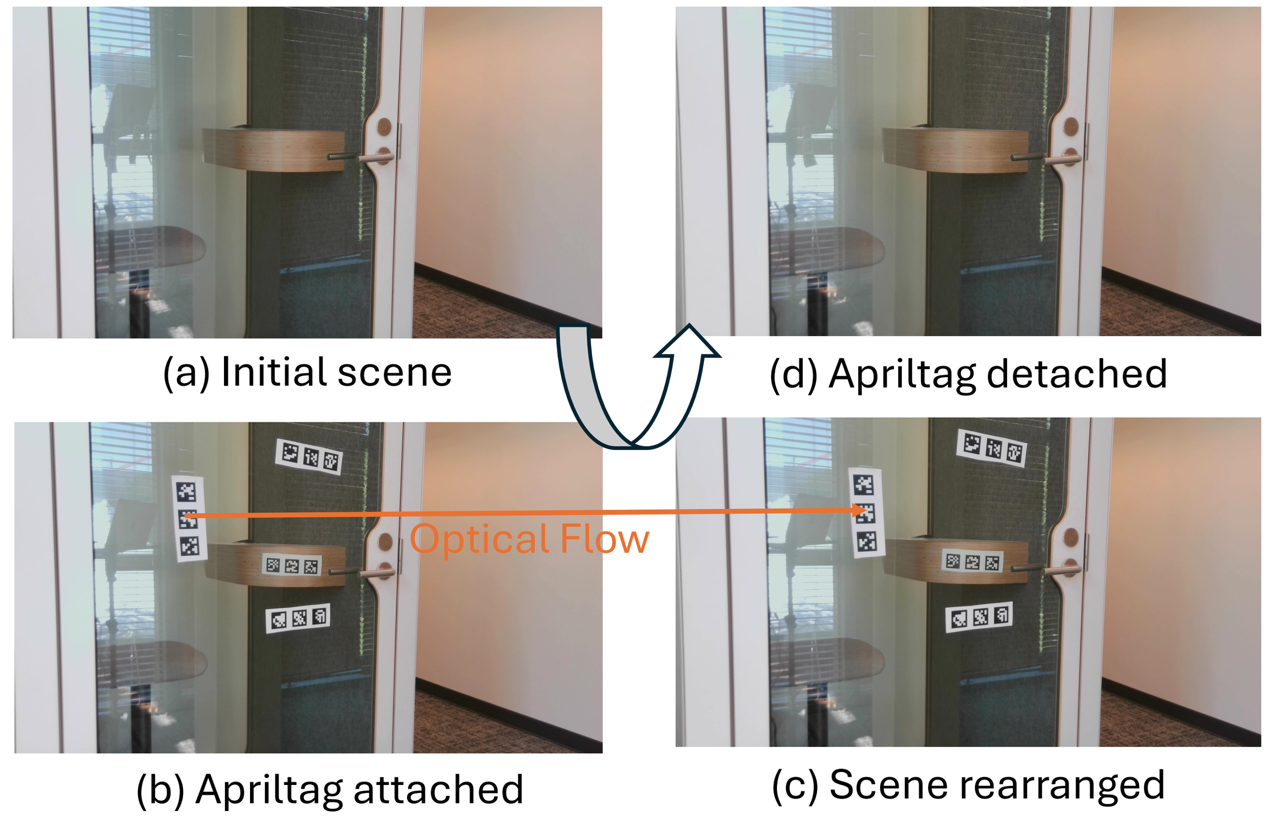
- Non-Lambertian objects, such as transparent and reflective objects, are common in daily life.
- Most existing benchmarks rarely include them. Even those focused on non-Lambertian objects offer limited diversity.
- No existing dataset provides the multi-layer information that human can infer from transparent surfaces.

Contribution

- We focus on optical flow prediction, a fundamental task in 3D vision.
- We propose a real-world, diverse benchmark with accurate, multi-layer ground-truth for evaluation.
- We propose a large-scale synthetic dataset with multi-layer dense annotation for training.
- We propose a baseline model for multi-layer optical flow prediction.

Benchmark

Data Acquisition Pipeline



We utilize AprilTag Markers to collect ground-truth. We manually annotate each marker with its material property (transparent, reflective or diffuse) and layer index to provide multi-layer ground-truth

Benchmark Gallery

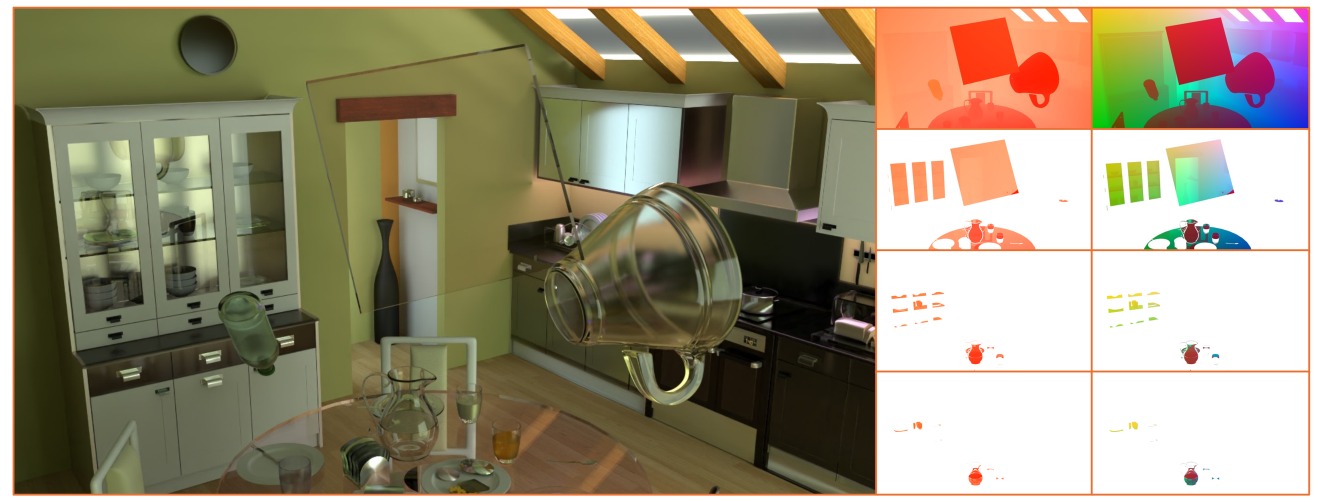
Our benchmark exhibits greater scene and object diversity than existing ones. We captured 2k images of 360 distinct objects in 155 indoor and 30 outdoor scenes under different lighting conditions.

Synthetic Dataset

Data Generation Pipeline

We use Blender to create a comprehensive synthetic dataset of 60k images of non-Lambertian objects from 30 high-quality scenes designed by artists.

Multi-layer Ground-truth



We modified Blender's ray tracing code to embed our ground truth collection in the rendering process. The images are annotated with dense multi-layer ground-truth that aligned with human perception.

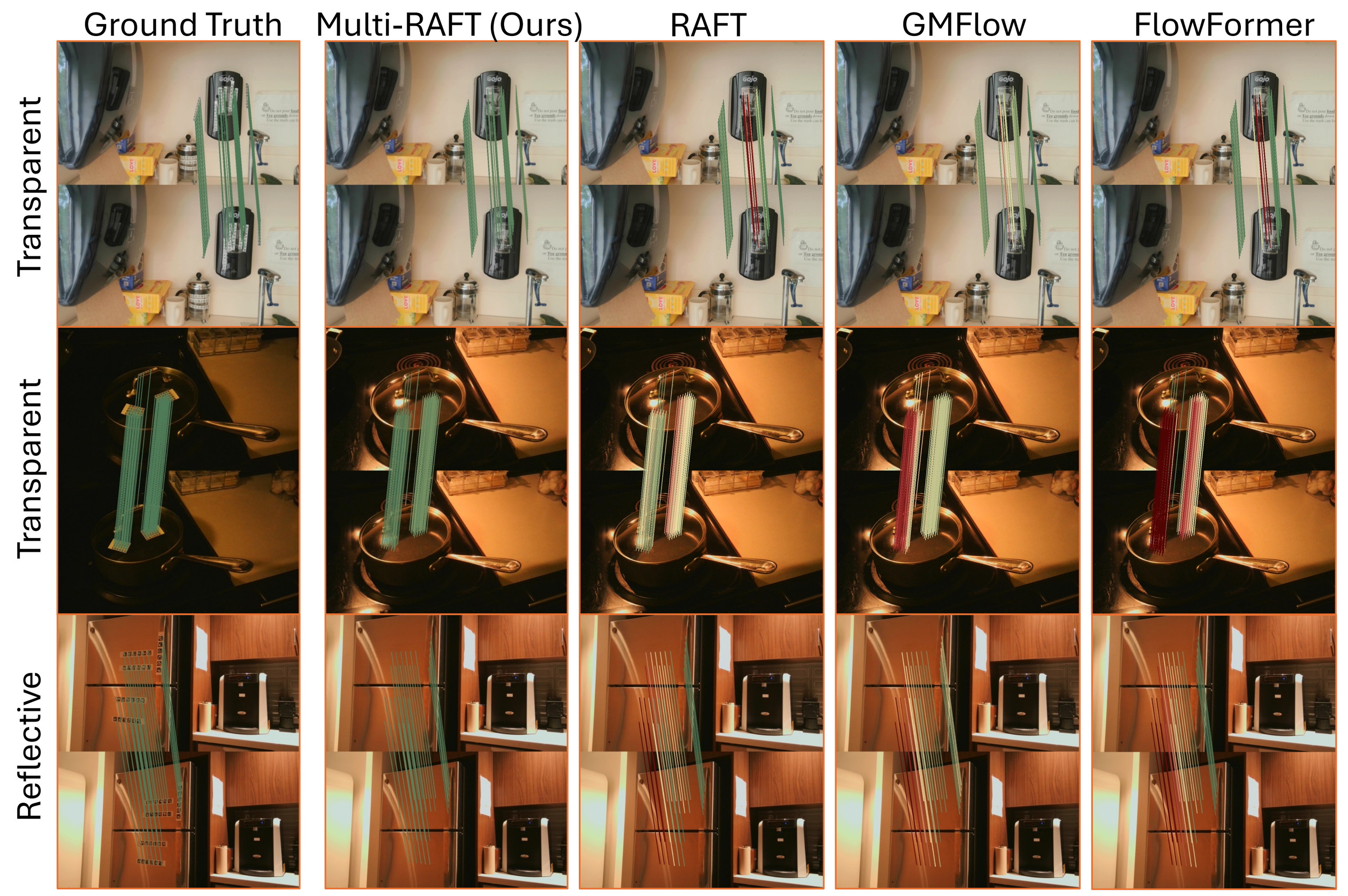
Experiments

Single Layer Optical Flow

Method	All				Transparent				Reflective				Diffuse			
	EPE↓	1px↓	3px↓	5px↓	EPE↓	1px↓	3px↓	5px↓	EPE↓	1px↓	3px↓	5px↓	EPE↓	1px↓	3px↓	5px↓
FlowNet-C	21.14	94.88	77.86	65.20	24.01	94.84	77.90	65.15	13.85	94.82	79.18	67.35	17.04	96.18	70.95	56.46
FlowNet2	20.67	86.42	66.54	56.66	23.54	87.19	67.61	57.55	13.52	84.57	63.45	53.97	15.42	76.30	54.82	47.54
PWC-Net	28.39	83.93	63.66	54.33	31.75	86.34	66.69	57.10	15.45	74.12	51.02	43.56	20.48	70.80	48.74	37.06
GMA	16.58	79.26	57.04	46.60	20.35	82.93	61.16	49.91	**8.18**	65.34	41.04	33.83	12.00	55.04	31.45	25.48
SKFlow	18.14	79.12	57.47	48.33	22.17	83.31	62.01	52.12	9.41	62.38	**38.95**	32.89	8.17	55.15	33.09	28.21
CRAFT	17.82	80.31	57.60	47.90	21.57	84.07	61.86	51.34	10.11	64.79	40.48	33.91	8.73	60.94	33.78	29.49
GMFlow	16.92	88.45	64.00	51.71	20.72	89.51	65.90	54.01	8.74	85.86	58.01	43.18	8.29	74.63	45.58	35.64
GMFlow+	17.62	89.83	67.21	54.29	21.36	90.36	68.83	56.45	9.68	88.65	61.91	45.80	10.06	82.53	52.35	41.21
FlowFormer	18.49	78.83	58.61	49.24	22.56	83.02	63.42	53.64	9.54	**61.73**	39.63	**32.24**	**5.01**	56.57	30.21	**21.33**
RAFT	16.49	78.45	55.64	45.78	20.11	82.72	59.69	49.06	8.51	62.05	40.49	33.21	10.76	**50.78**	**27.56**	24.39
RAFT-ft. (S)	17.94	79.53	59.47	49.69	21.96	82.94	63.15	52.85	8.89	66.41	45.21	37.11	9.07	57.70	36.44	31.34
RAFT-ft. (L)	17.46	78.13	53.12	43.33	18.54	82.15	**56.06**	45.73	17.30	62.60	41.75	33.89	14.69	52.60	34.73	28.87
RAFT-ft. (S+L)	**15.63**	**77.81**	**52.75**	**42.76**	**18.39**	**81.88**	56.17	**45.40**	11.73	61.93	39.48	32.97	6.95	52.75	31.23	24.24

Fine-tuning RAFT on the synthetic dataset boosts its performance on non-Lambertian objects without compromising the performance on diffuse objects.

Multiple Layer Optical Flow



Our baseline model, Multi-RAFT, takes inspiration from RAFT but utilizes multiple context encoders for multi-layer prediction. Trained on the synthetic dataset, Multi-RAFT has better understanding of non-Lambertian surfaces.