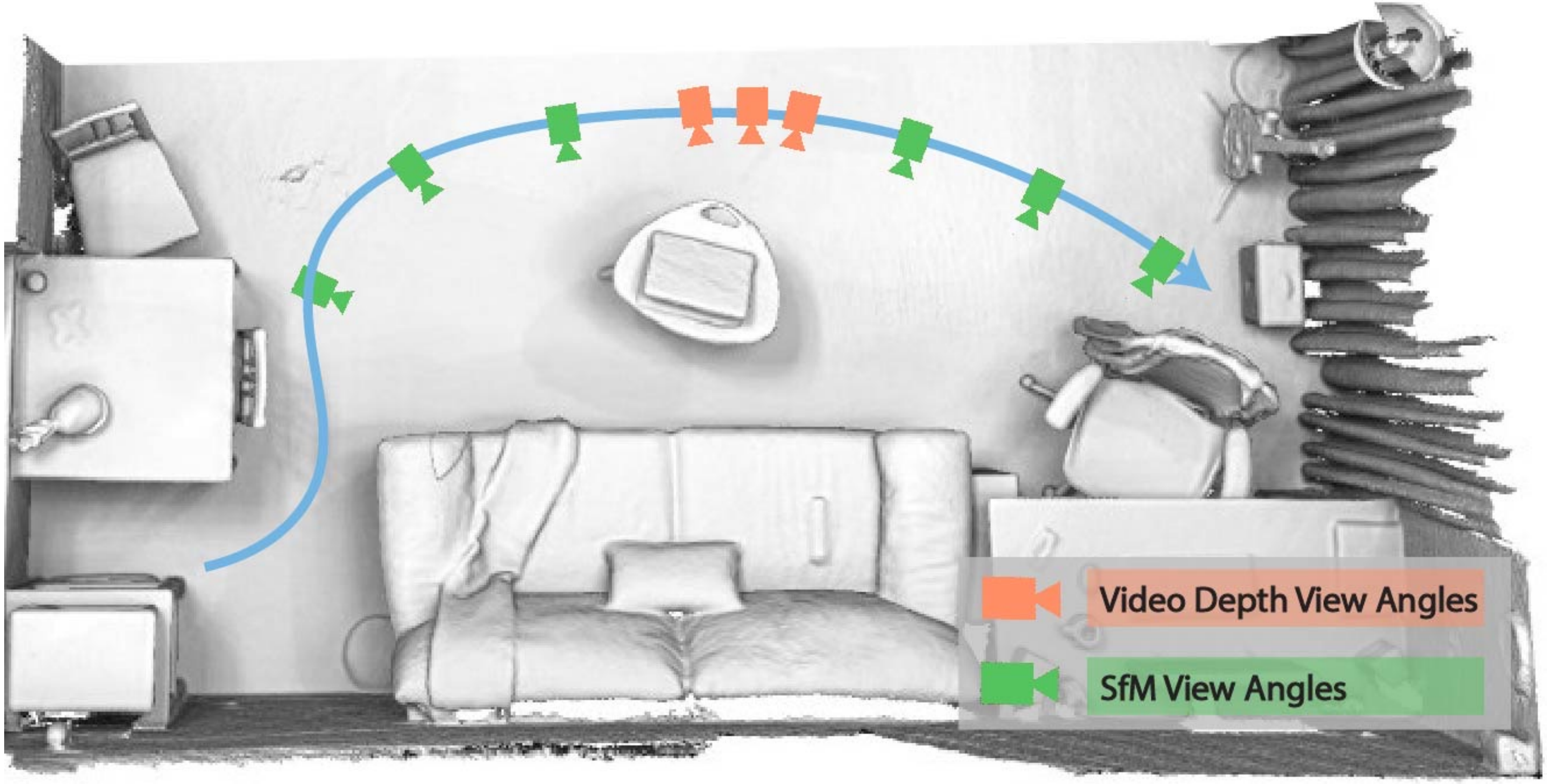
LightedDepth:
Video Depth Estimation in Light of Limited Inference View Angles

Shengjie Zhu and Xiaoming Liu

Michigan State University

Paper ID:8492
Paper Session: TUE-PM-082

LightedDepth:
Video Depth Estimation in Light of Limited Inference View Angles



- Observation:
  Limited Camera View
- Challenge:
  Camera Pose is Needed
- Solution:

  1. Optimize in 2D Correspondence rather than 3D

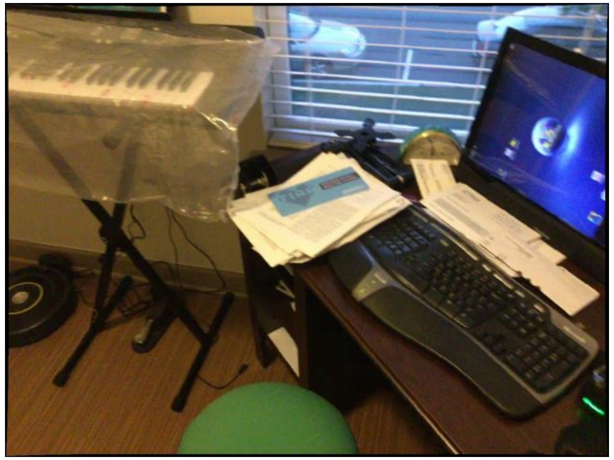
  2. Rely on Depth Prior Learning

  3. Connect Two with Efficient Camera Scale Estimation

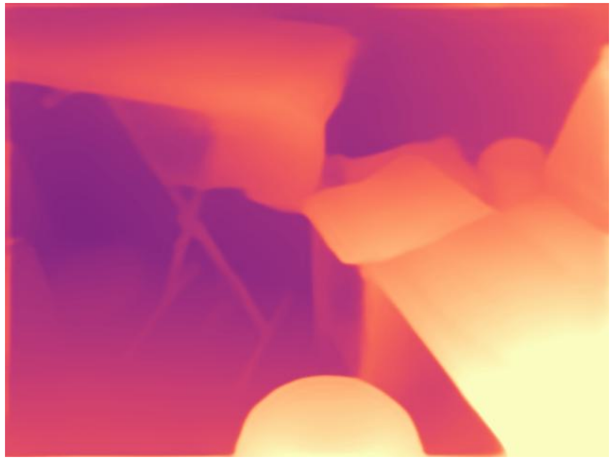
Problem Introduction

Two-View Video Depth Estimation or SfM

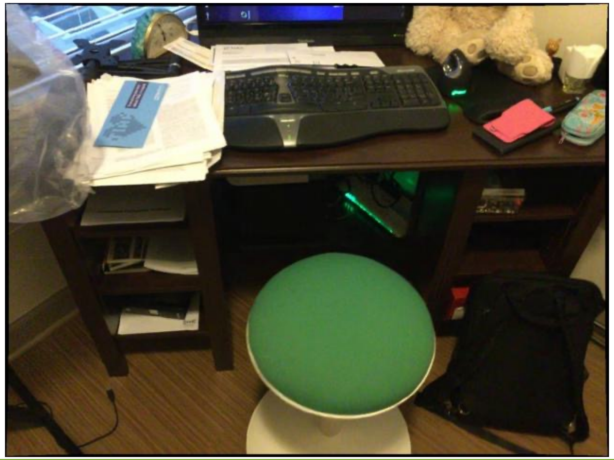
Source



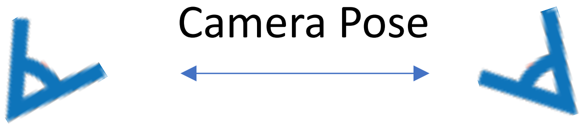
Depth



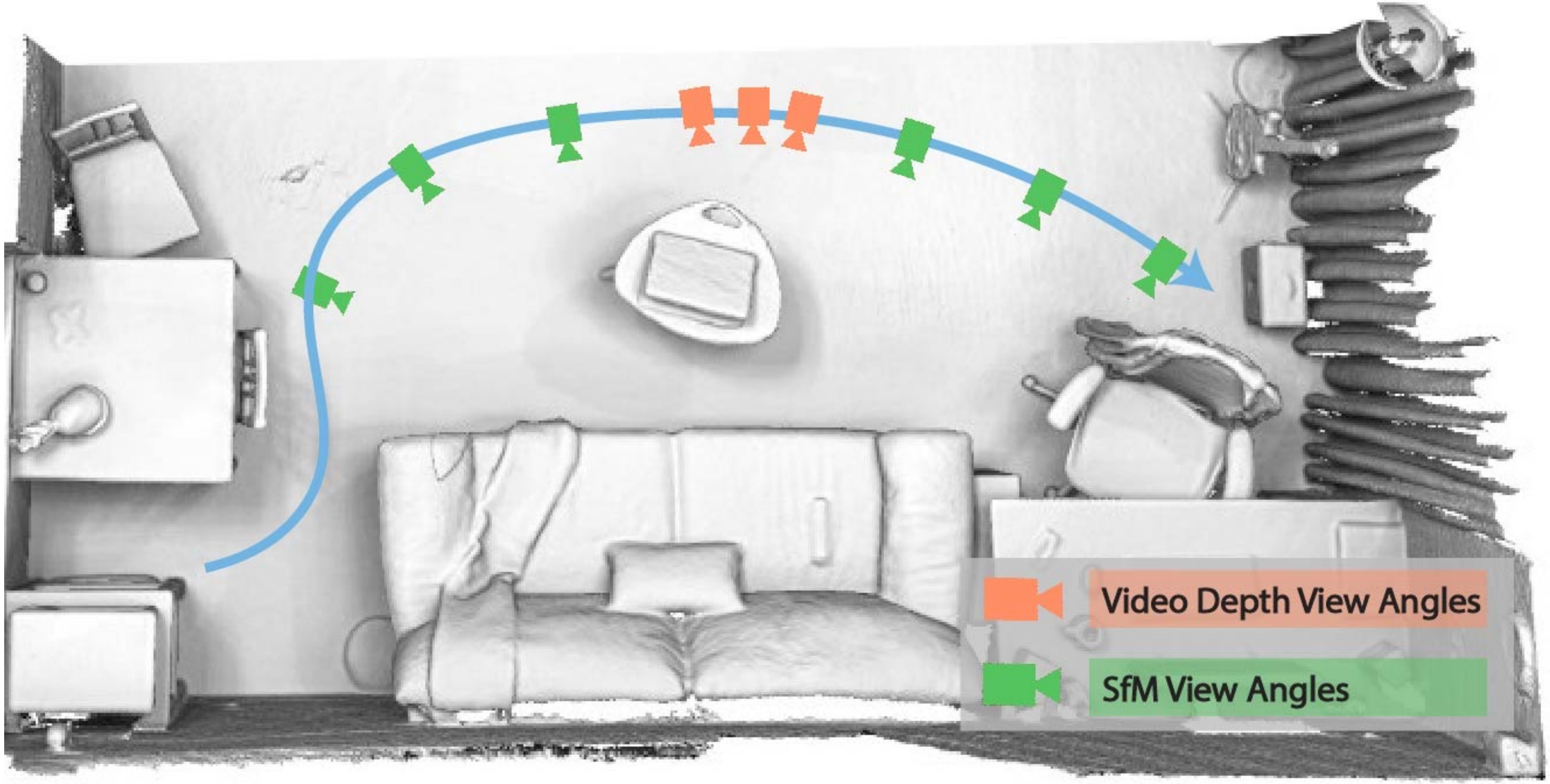
Support



Correspondence



Comparison to Classic SfM



- Two-View SfM Conducts in Limited Views
- Compared to Simplified SfM, Resemble more to Video-Depth Estimation

Applications Two-View SfM

AR Rendering

NerF Rendering

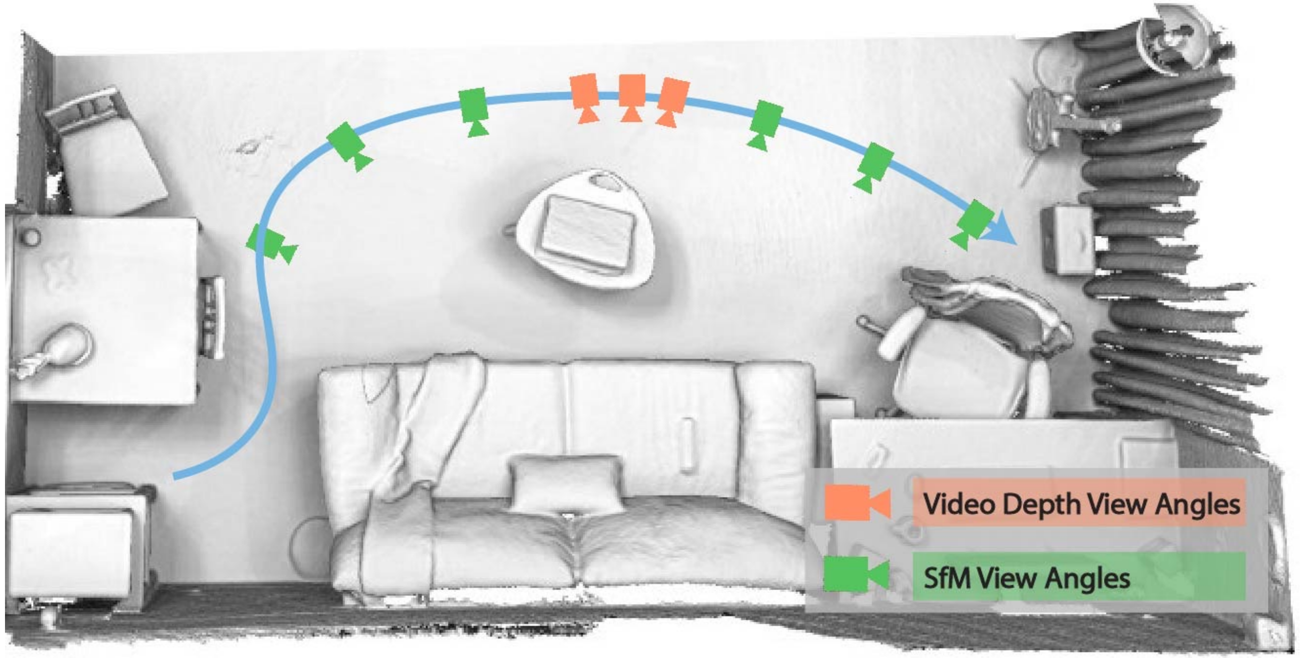
- View Angles are limited
- Challenge for COLMAP

Image Courtesy:

- https://www.youtube.com/watch?v=RRBpz2zaA-w
- https://www.matthewtancik.com/nerf

Two-View SfM
as
Video Depth Estimation

Comparison to Classic SfM



Two-View SfM

Monocular Depth

Video Depth

Correspondence

Normalized Pose

Pose Scale

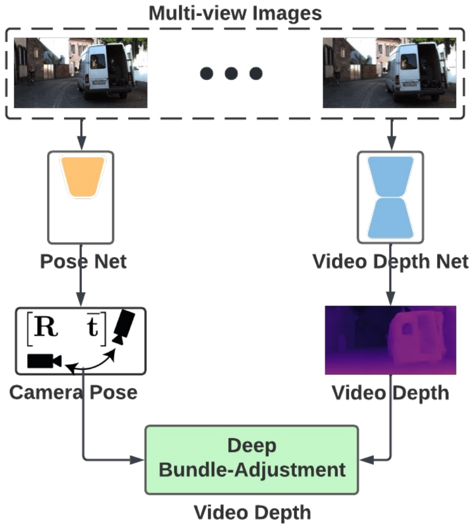
Observation:

- Two-View SfM Conducts in Limited Views
- Compared to Simplified SfM, Resemble more to Video-Depth Estimation

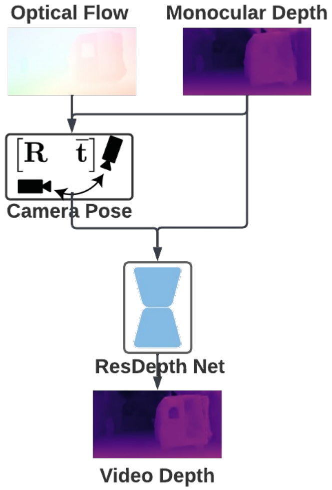
Solution:

- Rely on 3D prior
- Rely on 2D Constraint

- Prior Works

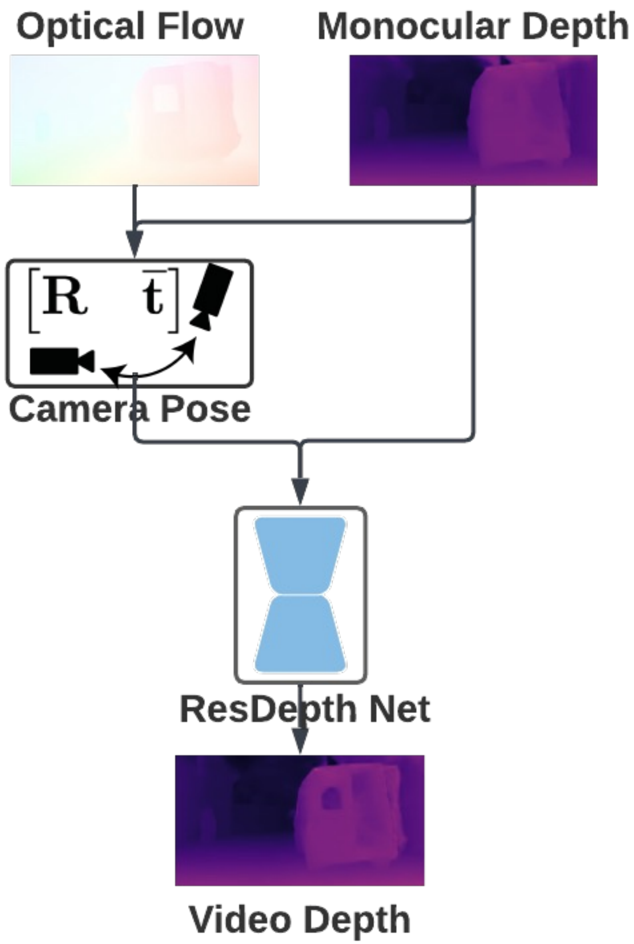


- Ours



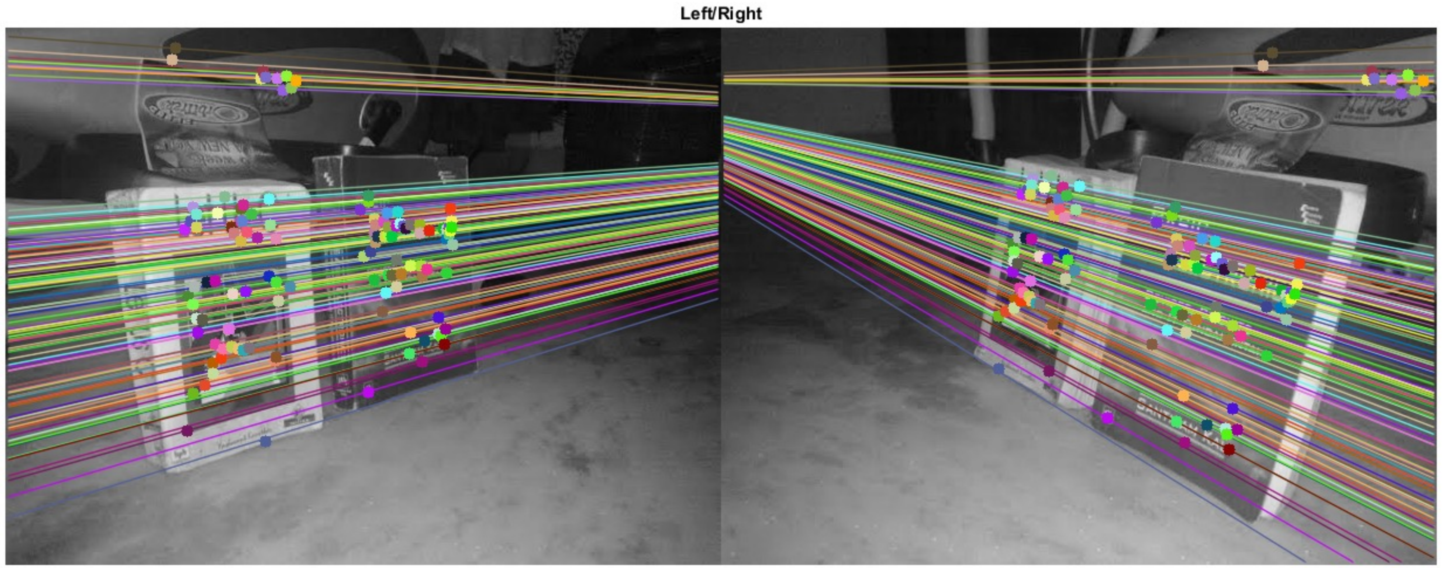
Normalized Pose from Correspondence

Our Pose Estimation



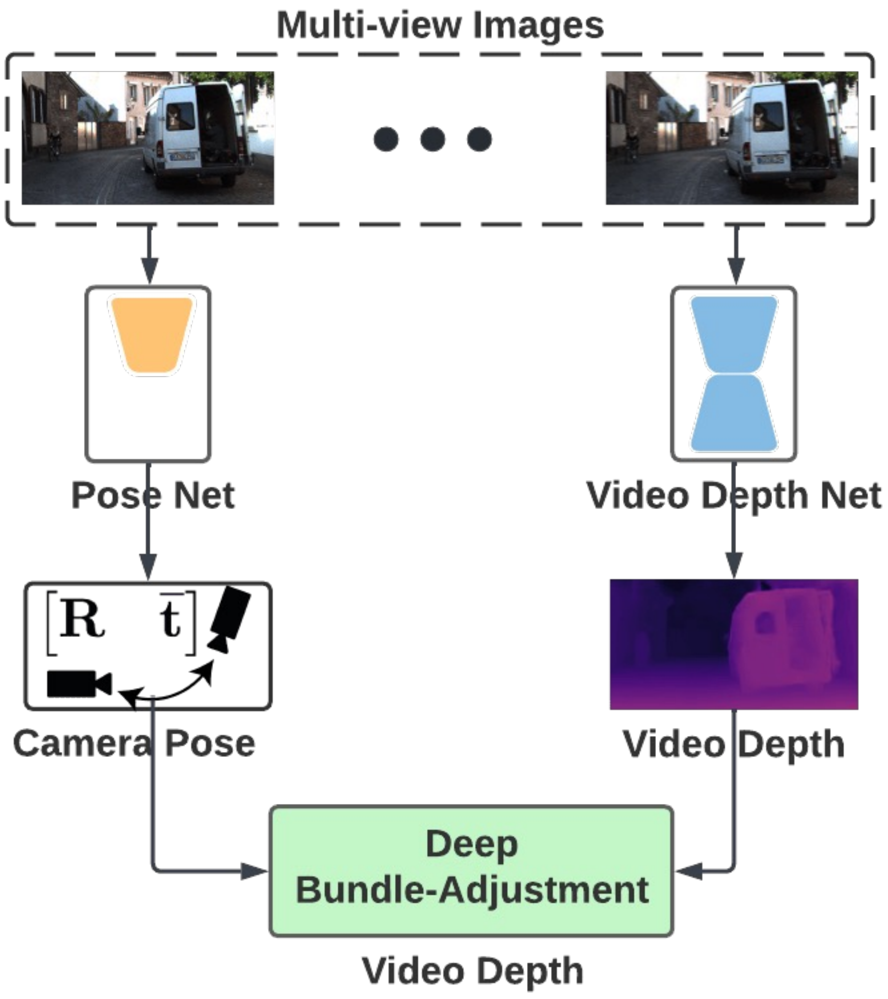
- **Normalized Pose Estimation** with **RANSAC** on **2D Correspondence**

- **Epipolar Geometry:** $\xrightarrow{\text{Unknown Scale}}$ **Normalized Pose**

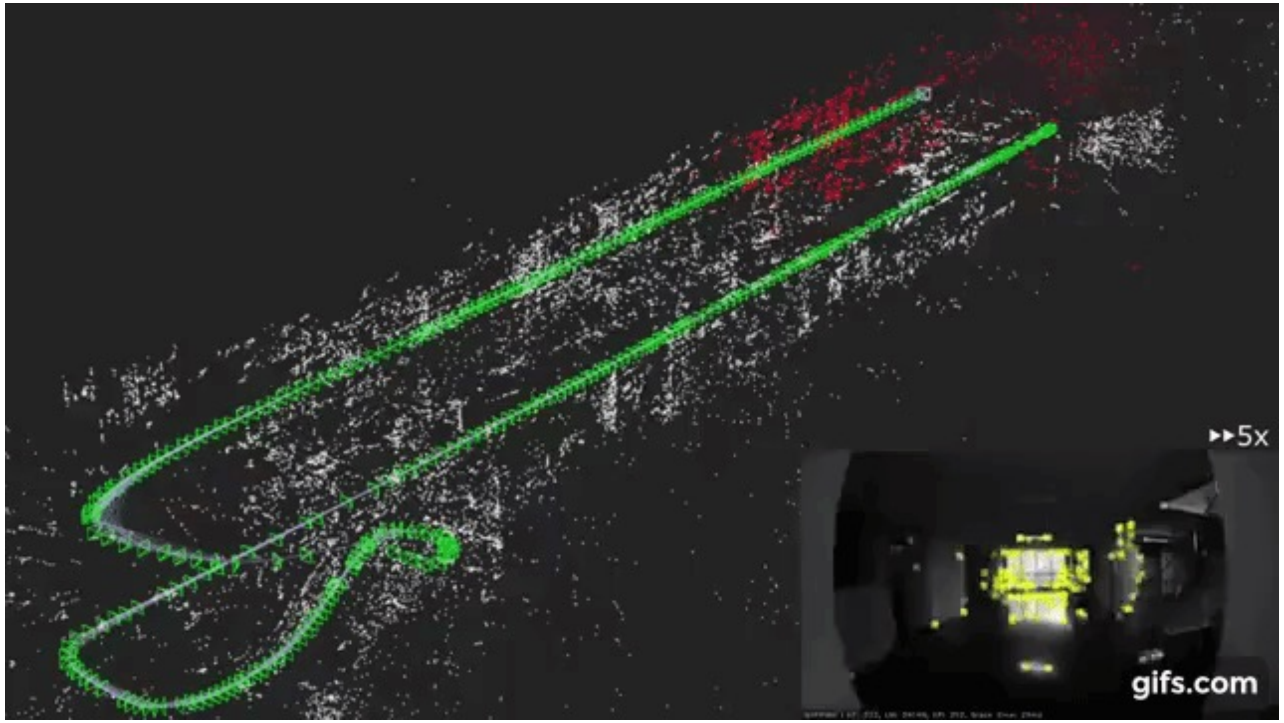


- Image Courtesy: https://docs.opencv.org/3.4/da/de9/tutorial_py_epipolar_geometry.html

Prior Pose Estimation



- Jointly Estimate Pose & Depth with **Deep-Bundle Adjustment** in **3D Space**



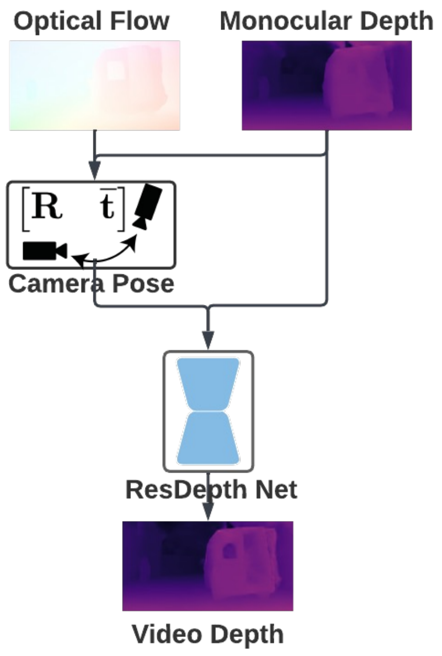
- Image Courtesy: https://www.google.com/imgres?imgurl=https%3A%2F%2Fi.ytimg.com

Our Pose Estimation

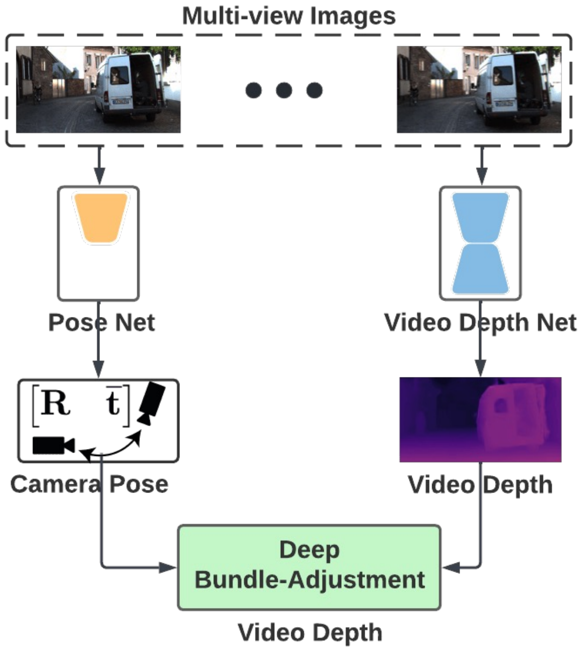
Type	Mehod	All F1-epe	All F1-a1	Background F1-epe	Background F1-a1
Type1	RAFT	1.284	4.539	1.238	4.759
Type2	DeepV2D	9.957	22.610	2.180	9.789
	Ours	9.321	20.723	1.631	7.692

- Type 1: Correspondence from 2D Matching
- Type 2: Correspondence from 3D Bundle-Adjustment

- Correspondence from Type1 outperforms Type2

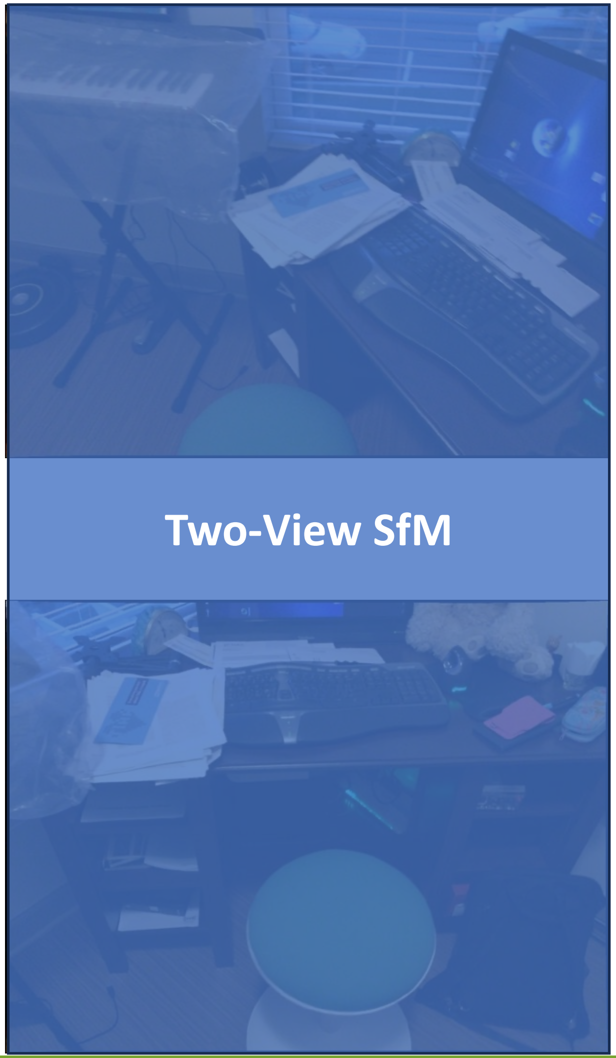


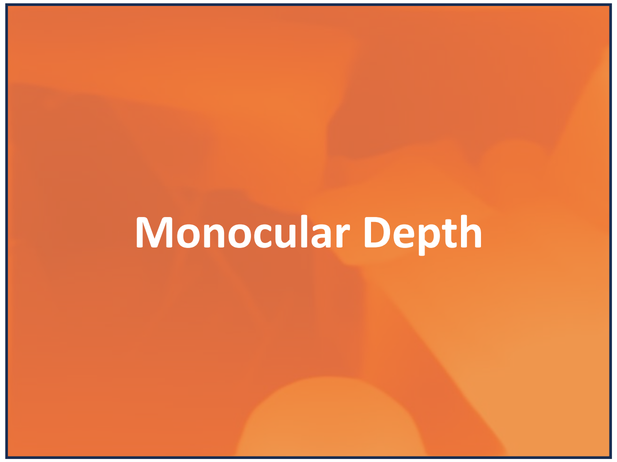
Type1:
- Two View Image Matching



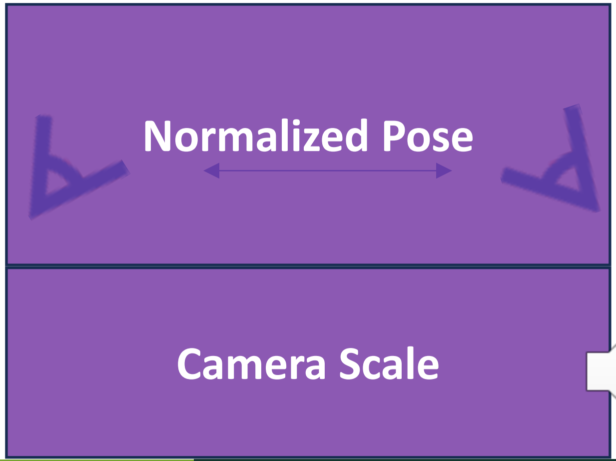
Type2:
- Multiview
- Apply Bundle Adjustment

Two-View SfM

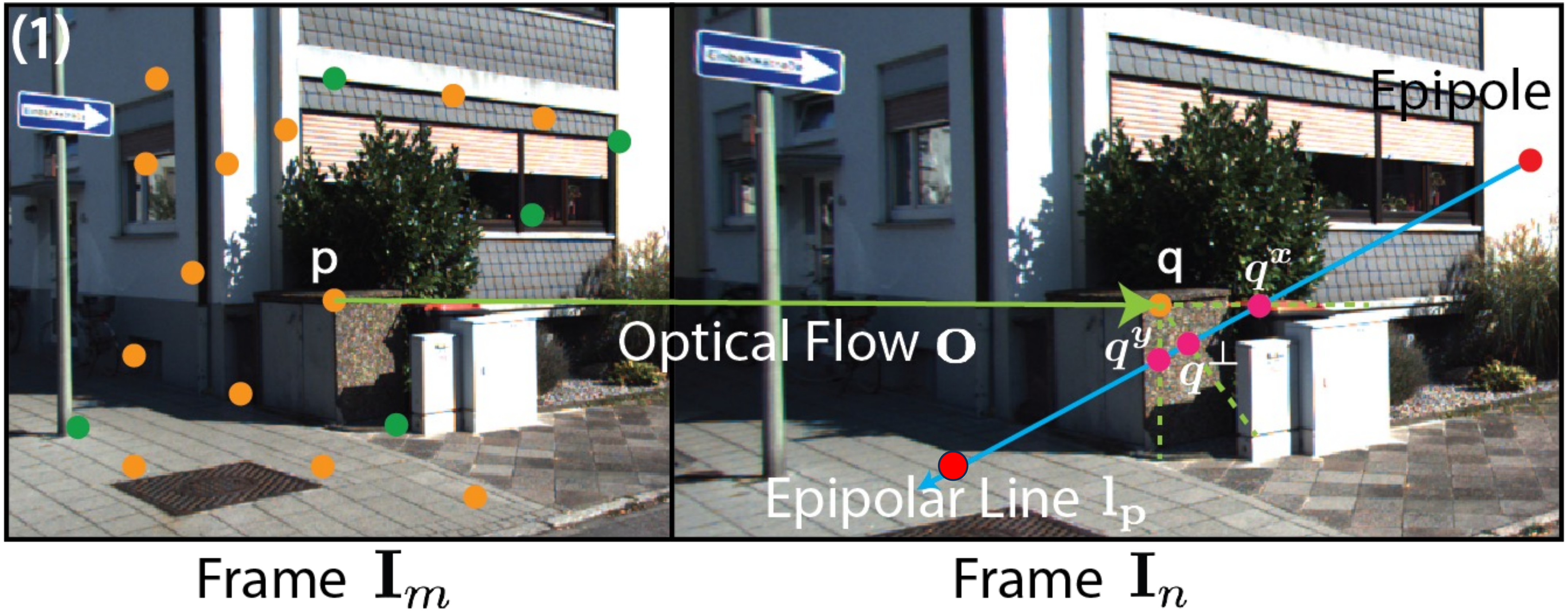

Two-View SfM


Monocular Depth


Correspondence

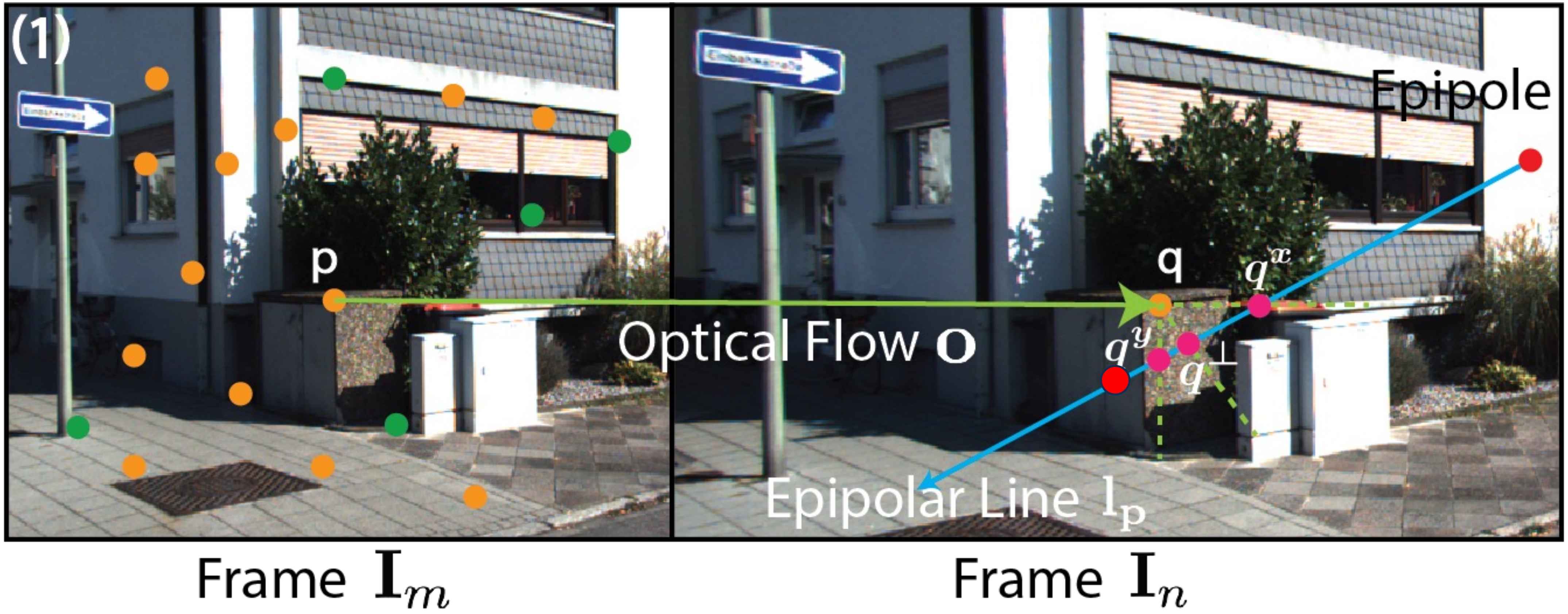

Normalized Pose
Camera Scale

Two-View SfM, Pose Scale Estimation



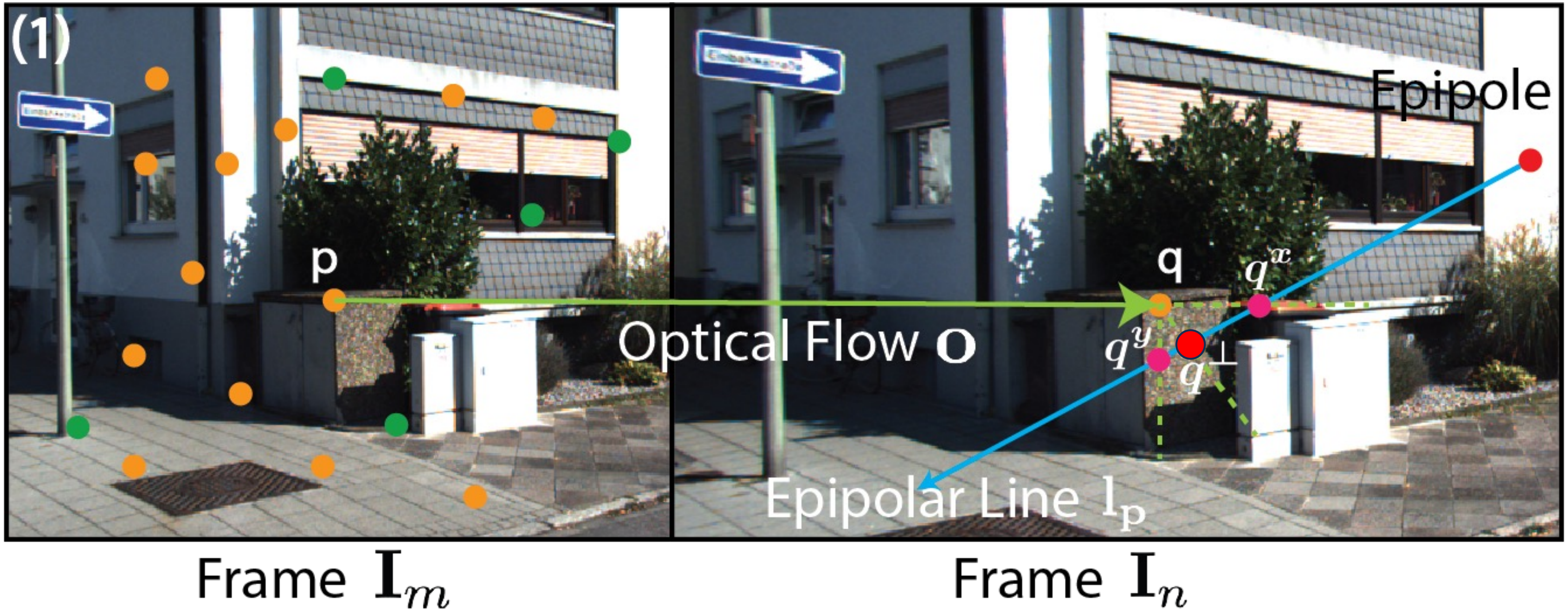
- Pose Scale Adjust Correspondence in Epipolar Line
- Adjust Scale to Align towards Correspondence (Optical Flow)

Two-View SfM, Pose Scale Estimation



- Pose Scale Adjust Correspondence in Epipolar Line
- Adjust Scale to Align towards Correspondence (Optical Flow)

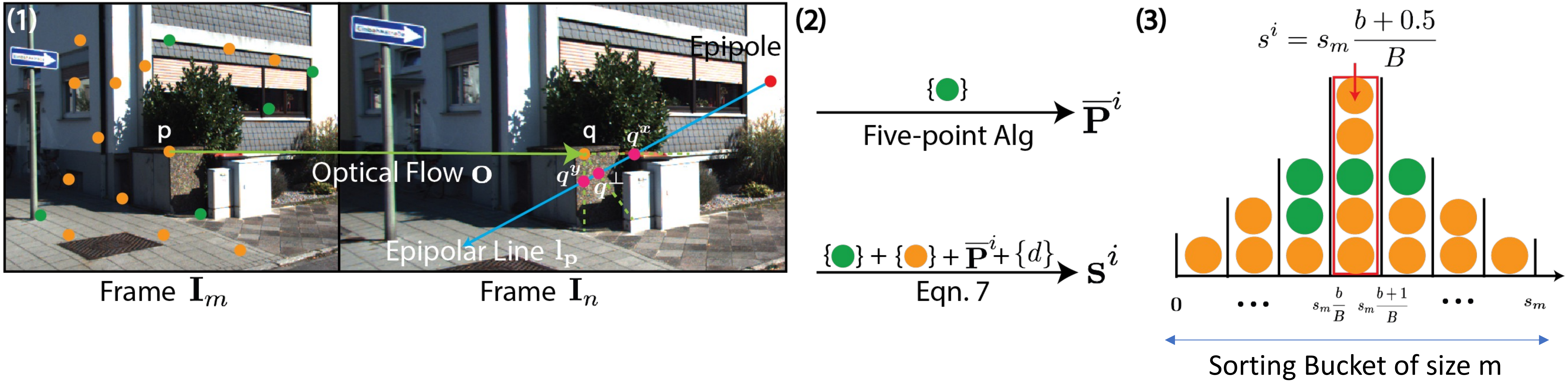
Two-View SfM, Pose Scale Estimation



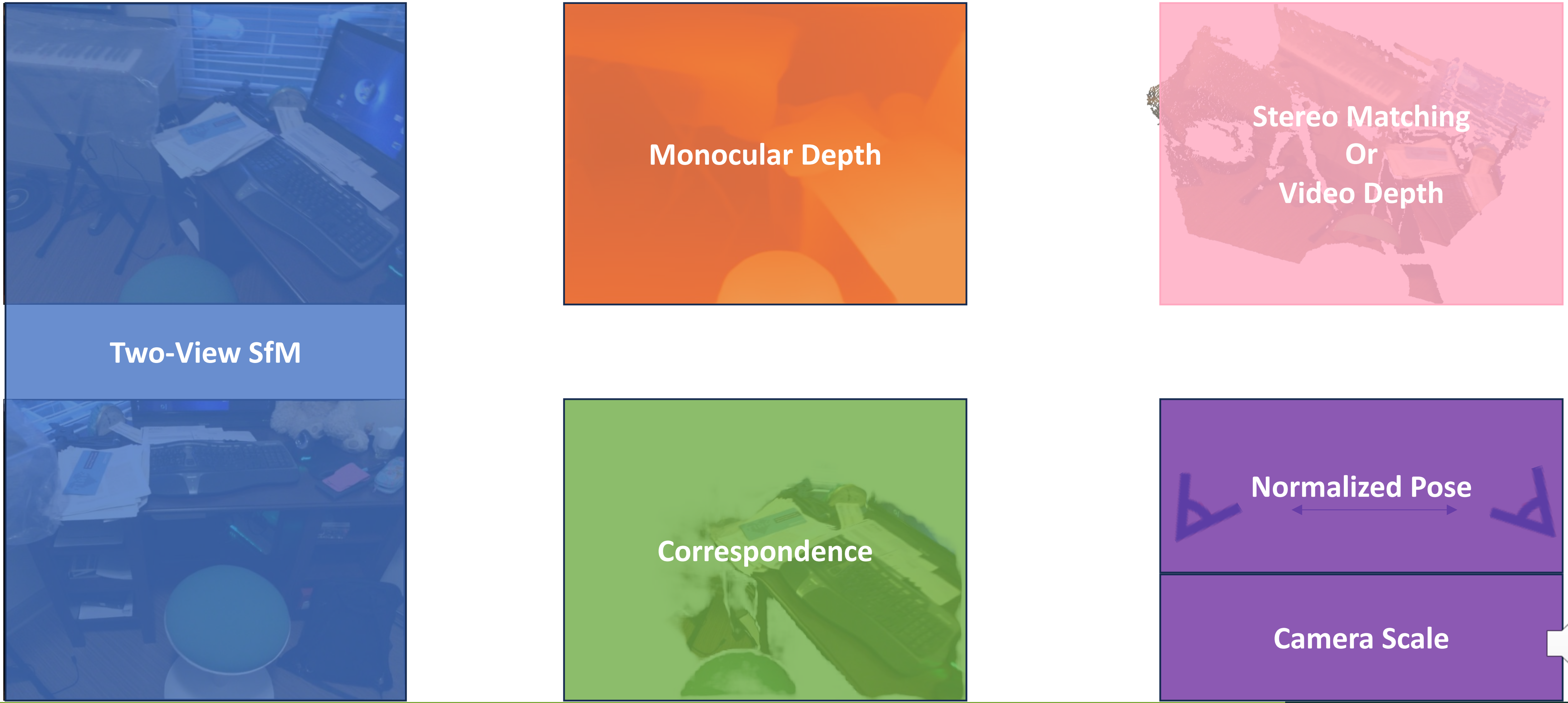
- Pose Scale Adjust Correspondence in Epipolar Line
- Adjust Scale to Align towards Correspondence (Optical Flow)

Two-View SfM, Pose Scale Estimation

- Pose Scale is a 1 DoF unknown scale
- Use Bucket Sorting to Select the Optimal Scale



Two-View SfM

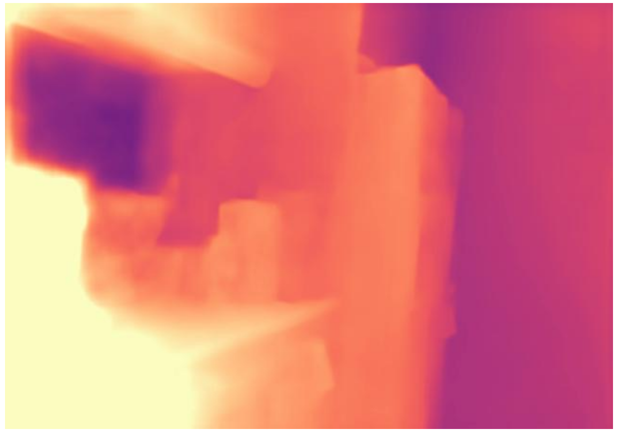


Two-View SfM
Monocular Depth
Stereo Matching
Or
Video Depth
Correspondence
Normalized Pose
Camera Scale

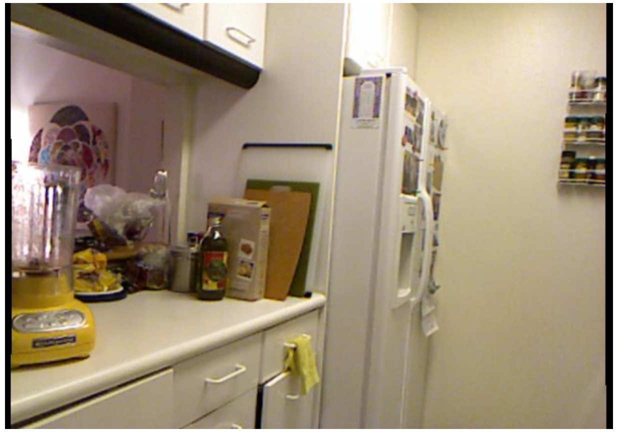
Video Depth as Residual Monocular Depth Estimation

Frame 1

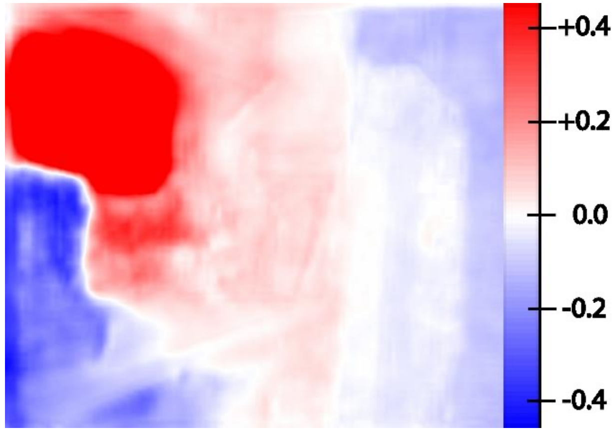
Monocular Depth



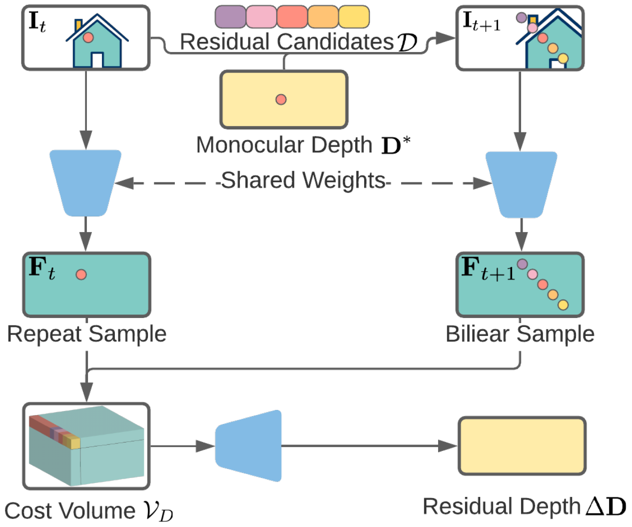
Video Depth



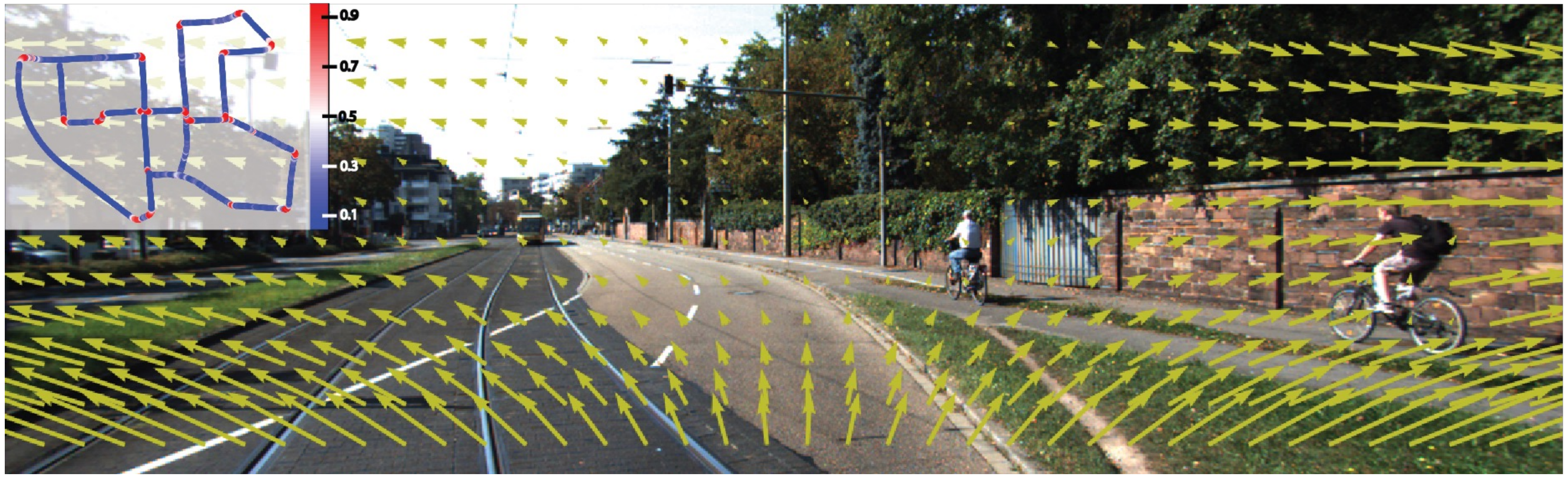
Frame 2

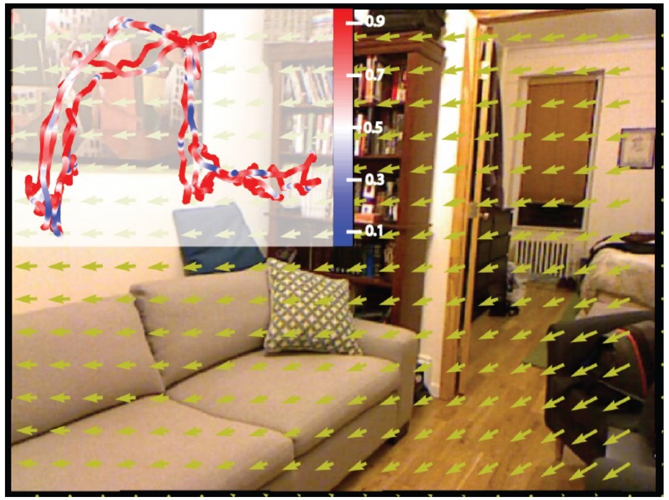


Residual Depth



Corner Cases



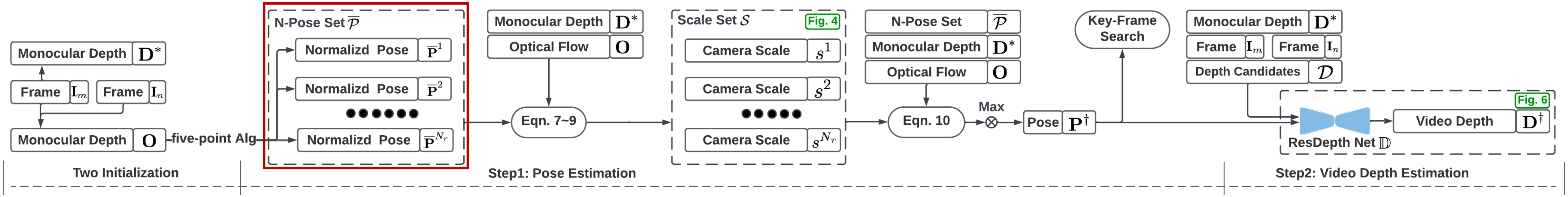


- Translation Dominant Case
- Rotation Dominant Case

Issue:
- Normalized Pose estimation degenerates on Rotation Dominant Cases.
- However, the latter is common in Indoor Setting

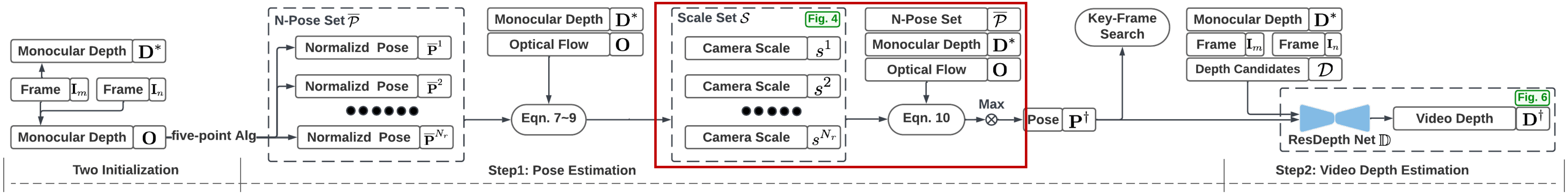
Solution:
- Actively Search KeyFrame with Sufficient Camera Scale
- Include Monocular Depthmap Projection as Additional Constraint

Entire Framework



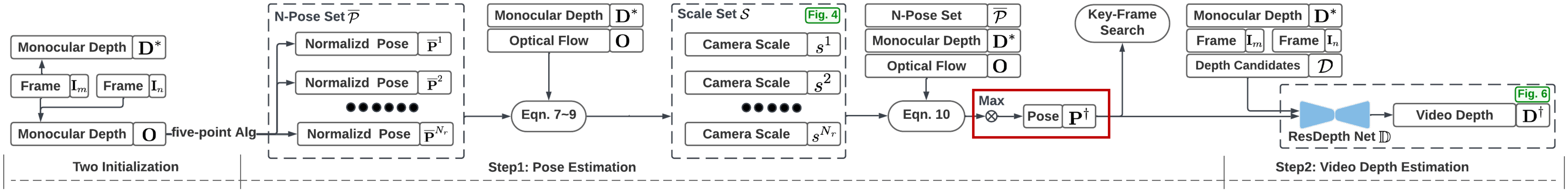
- Spawn Normalized Pose Candidates with Five-Point Algorithm

Entire Framework



- Spawn Normalized Pose Candidates with Five-Point Algorithm
- Compute Camera Scale, followed by Epipolar Constraint and Projection Constraint

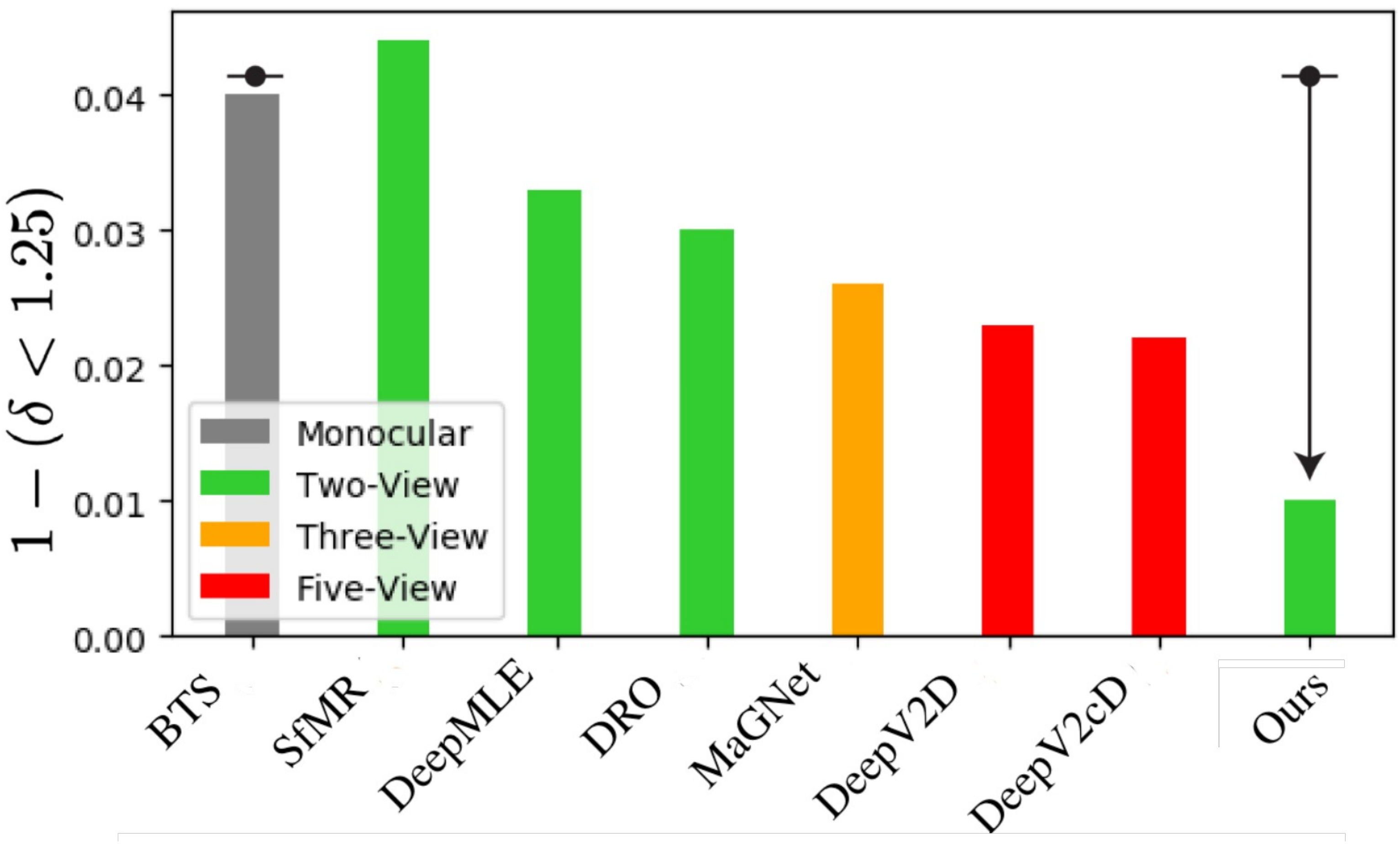
Entire Framework



- Spawn Normalized Pose Candidates with Five-Point Algorithm
- Compute Epipolar Constraint, Camera Scale, and Projection Constraint
- Acquire Best Pose

Result

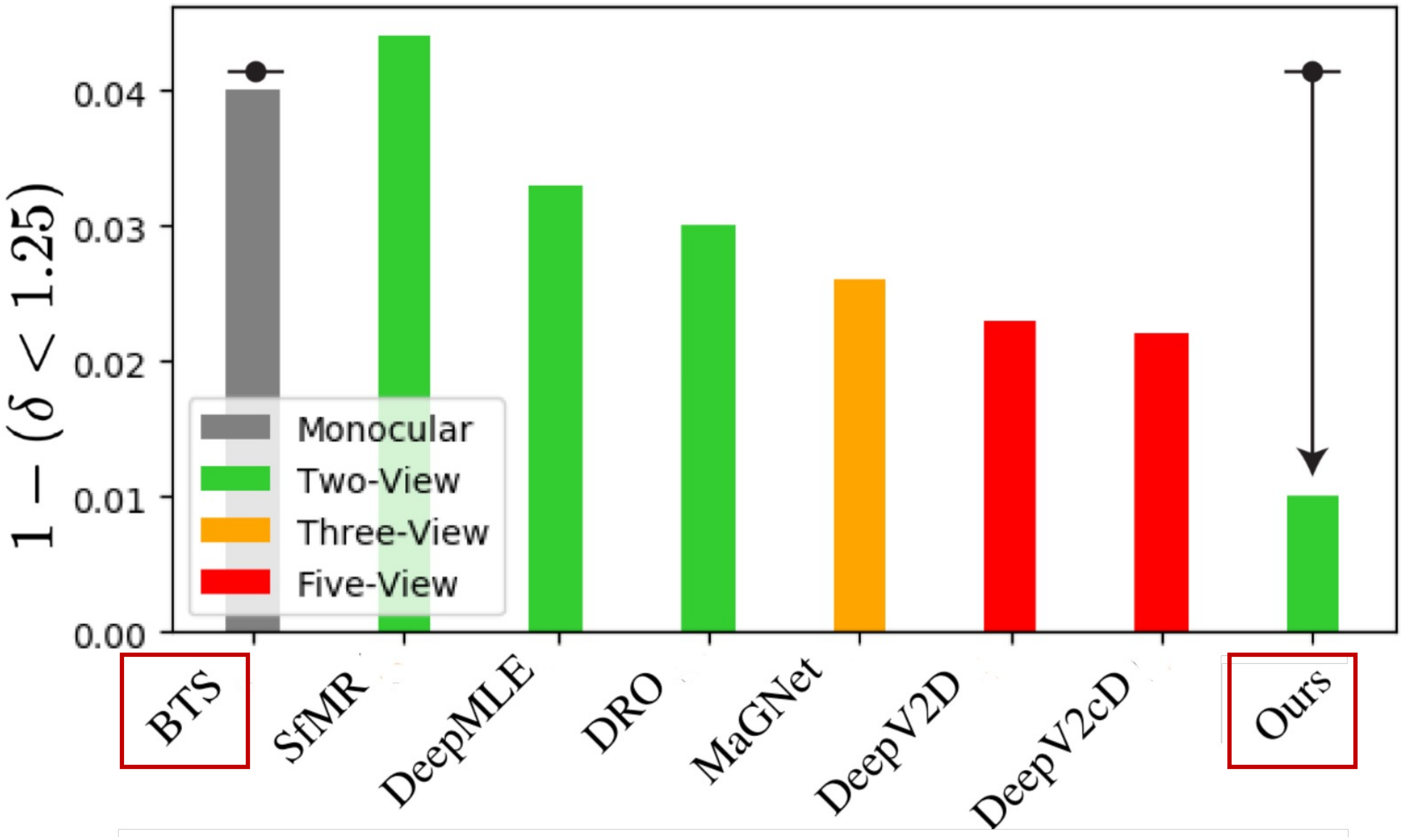
LightedDepth Performance



- Our method Significantly outperform even Prior work using Five Frames

- KITTI Dataset

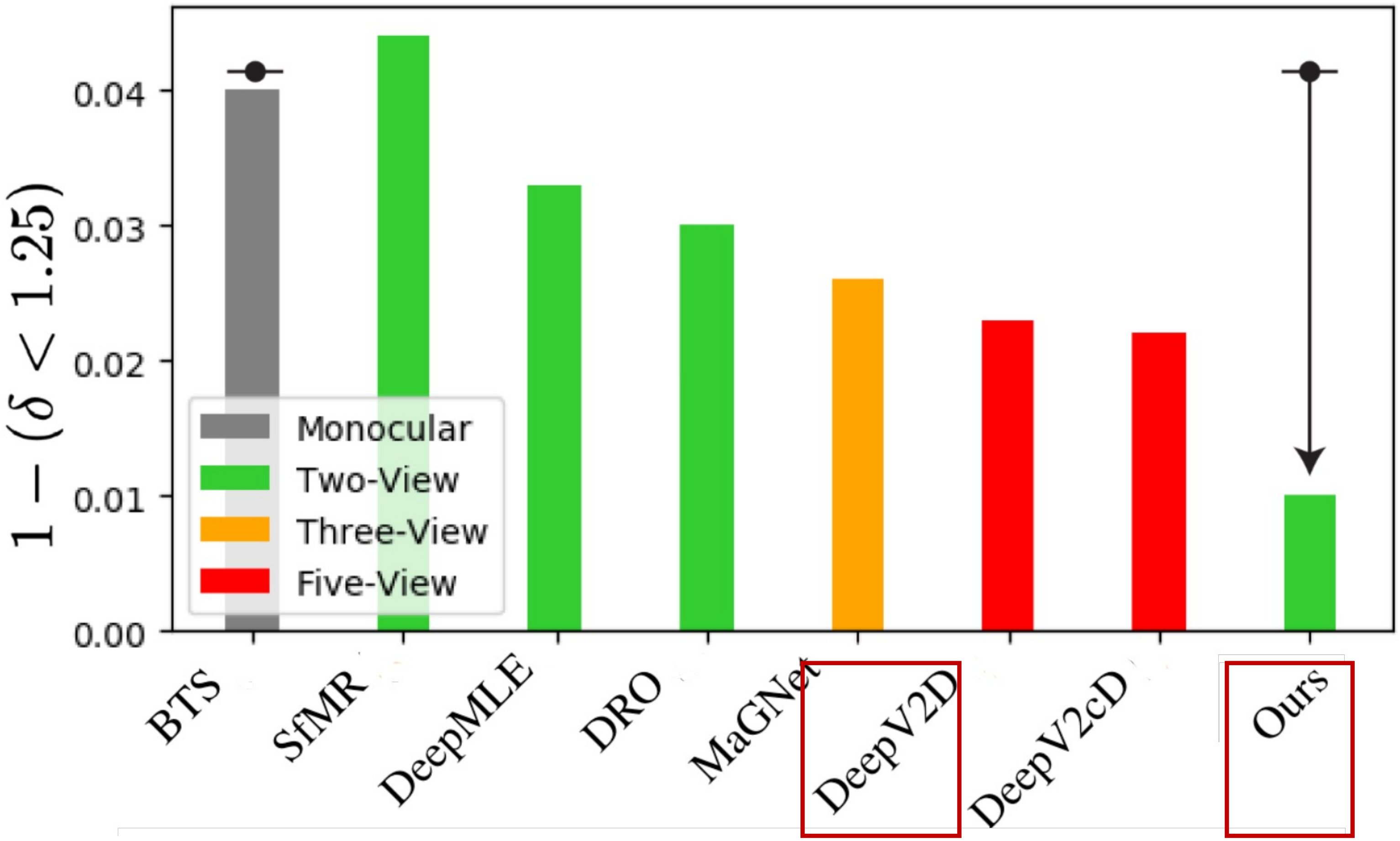
LightedDepth Performance



- Our method Significantly outperform even Prior work using Five Frames

- KITTI Dataset

LightedDepth Performance



- Our method Significantly outperform even Prior work using Five Frames

- KITTI Dataset

LightedDepth Performance

Method	Venue	Frame	Labels	Abs Rel	Sq Rel	RMSE	RMSE log	$\delta < 1.25$	$\delta < 1.25^2$	$\delta < 1.25^3$
DORN [14]	CVPR'18	1	D	0.069	0.300	2.857	0.112	0.945	0.998	0.996
BTS [27]	Arxiv'18	1	D	0.059	0.245	2.756	0.096	0.956	0.993	**0.998**
AdaBins [2]	CVPR'21	1	D	0.058	0.190	2.360	0.088	0.964	0.995	**0.999**
NeWCRFs [58]	CVPR'22	1	D	**0.052**	**0.155**	**2.129**	**0.079**	**0.974**	**0.997**	**0.999**
Ours + BTS [27]	CVPR'23	2	D+F	**0.037**	0.110	1.809	**0.059**	0.987	**0.998**	**0.999**
Ours + AdaBins [2]		2	D+F	0.045	0.108	1.817	0.064	0.987	**0.998**	**0.999**
Ours + NeWCRFs [58]		2	D+F	0.041	**0.107**	**1.748**	**0.059**	**0.989**	**0.998**	**0.999**
BA-Net [40]	ICLR'19	5	D+P	0.083	0.025	3.640	0.134	-	-	-
SfMR [50]	CVPR'21	2	D+F+P	0.055	0.224	2.273	0.091	0.956	0.984	0.993
DeepMLE [8]	Arxiv'22	2	D+F+P	0.060	0.203	2.257	0.089	0.967	**0.995**	**0.999**
DRO [20]	Arxiv'21	2	D+P	0.047	0.199	2.629	0.082	0.970	0.994	0.998
MaGNet [1]	CVPR'22	3	D	0.051	**0.160**	2.077	0.079	0.974	**0.995**	**0.999**
DeepV2D [41]	ICLR'20	2	D+P	0.064	0.350	2.964	0.120	0.946	0.982	0.991
		5	D+P	**0.037**	0.174	2.005	0.074	0.977	0.993	0.997
DeepV2cD [22]	ICPRAI'22	5	D+P	**0.037**	0.167	**1.984**	**0.073**	**0.978**	0.994	-
Ours + MonoDepth2 [18]	CVPR'23	2	D+F	0.032	0.106	1.889	0.057	0.986	**0.998**	**0.999**
Ours + BTS [27]		2	D+F	0.029	0.098	1.729	0.053	0.989	**0.998**	**0.999**
Ours + AdaBins [2]		2	D+F	0.030	0.089	1.655	0.052	0.989	**0.998**	**0.999**
Ours + NeWCRFs [58]		2	D+F	**0.028**	**0.087**	**1.597**	**0.049**	**0.991**	**0.998**	**0.999**

- On Outdoor KITTI

LightedDepth Performance

Method	Venue	Frame	Labels	Abs Rel	Sq Rel	RMSE	RMSE log	$\delta < 1.25$	$\delta < 1.25^2$	$\delta < 1.25^3$
DORN [14]	CVPR'18	1	D	0.069	0.300	2.857	0.112	0.945	0.998	0.996
BTS [27]	Arxiv'18	1	D	0.059	0.245	2.756	0.096	0.956	0.993	**0.998**
AdaBins [2]	CVPR'21	1	D	0.058	0.190	2.360	0.088	0.964	0.995	**0.999**
NeWCRFs [58]	CVPR'22	1	D	**0.052**	**0.155**	**2.129**	**0.079**	**0.974**	**0.997**	**0.999**
Ours + BTS [27]	CVPR'23	2	D+F	**0.037**	0.110	1.809	**0.059**	0.987	**0.998**	**0.999**
Ours + AdaBins [2]	CVPR'23	2	D+F	0.045	0.108	1.817	0.064	0.987	**0.998**	**0.999**
Ours + NeWCRFs [58]	CVPR'23	2	D+F	0.041	**0.107**	**1.748**	**0.059**	**0.989**	**0.998**	**0.999**
BA-Net [40]	ICLR'19	5	D+P	0.083	0.025	3.640	0.134	-	-	-
SfMR [50]	CVPR'21	2	D+F+P	0.055	0.224	2.273	0.091	0.956	0.984	0.993
DeepMLE [8]	Arxiv'22	2	D+F+P	0.060	0.203	2.257	0.089	0.967	**0.995**	**0.999**
DRO [20]	Arxiv'21	2	D+P	0.047	0.199	2.629	0.082	0.970	0.994	0.998
MaGNet [1]	CVPR'22	3	D	0.051	**0.160**	2.077	0.079	0.974	**0.995**	**0.999**
DeepV2D [41]	ICLR'20	2	D+P	0.064	0.350	2.964	0.120	0.946	0.982	0.991
DeepV2D [41]	ICLR'20	5	D+P	**0.037**	0.174	2.005	0.074	0.977	0.993	0.997
DeepV2cD [22]	ICPRAI'22	5	D+P	**0.037**	0.167	**1.984**	**0.073**	**0.978**	0.994	-
Ours + MonoDepth2 [18]	CVPR'23	2	D+F	0.032	0.106	1.889	0.057	0.986	**0.998**	**0.999**
Ours + BTS [27]	CVPR'23	2	D+F	0.029	0.098	1.729	0.053	0.989	**0.998**	**0.999**
Ours + AdaBins [2]	CVPR'23	2	D+F	0.030	0.089	1.655	0.052	0.989	**0.998**	**0.999**
Ours + NeWCRFs [58]	CVPR'23	2	D+F	**0.028**	**0.087**	**1.597**	**0.049**	**0.991**	**0.998**	**0.999**

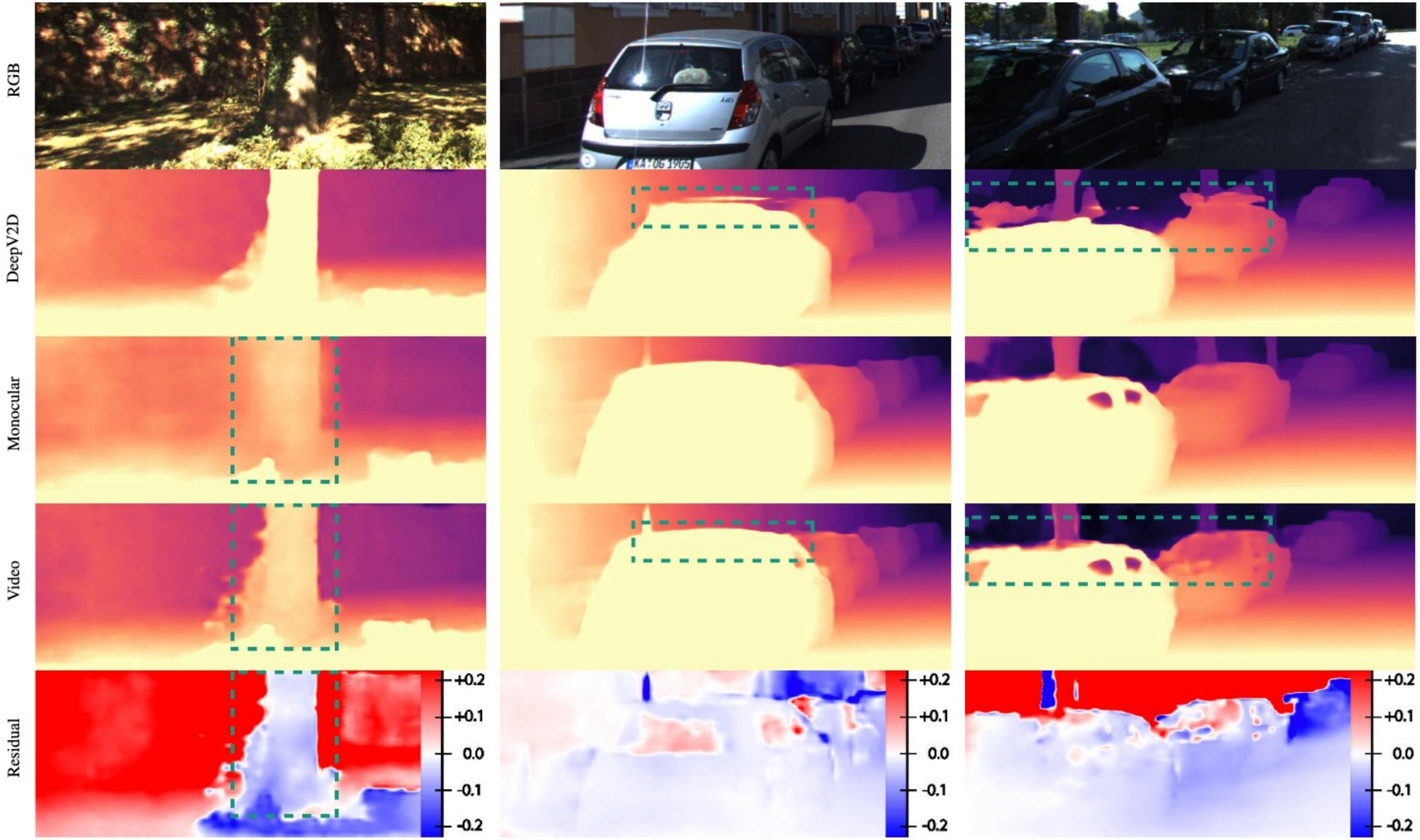
- On Outdoor KITTI

LightedDepth Performance

Method	Venue	Frame	Abs Rel	Sc Inv	RMSE	log10	$\delta < 1.25$	$\delta < 1.25^2$	$\delta < 1.25^3$
DORN [14]	CVPR'18	1	0.115	-	0.509	-	0.828	0.965	0.992
BTS [27]	Arxiv'18	1	0.108	0.115	0.404	0.047	0.885	0.978	0.994
AdaBins [2]	CVPR'21	1	0.103	0.106	0.370	0.044	0.903	0.983	0.997
NewCRFs [58]	CVPR'22	1	**0.095**	**0.090**	**0.334**	**0.041**	**0.922**	**0.992**	**0.998**
Ours + BTS [27]	CVPR'23	2	0.102	0.098	0.356	0.044	0.903	0.984	0.997
Ours + AdaBins [2]		2	0.095	0.089	0.326	0.040	0.923	0.990	0.998
Ours + NewCRFs [58]		2	**0.090**	**0.080**	**0.306**	**0.038**	**0.935**	**0.995**	**0.999**
DfUSMC [21]	CVPR'16	Multi	0.447	0.456	1.793	0.169	0.487	0.697	0.814
DeMoN [46]	CVPR'17	2	0.144	0.179	0.775	0.061	0.805	0.951	0.985
DeepV2D [41]	ICLR'20	2	0.094	0.133	0.521	0.403	0.905	0.975	0.992
		9	**0.061**	**0.094**	**0.403**	**0.026**	**0.956**	**0.989**	**0.996**
Ours + BTS [27]	CVPR'23	2	0.070	0.098	0.280	0.030	0.948	0.991	0.998
Ours + AdaBins [2]		2	0.064	0.089	0.255	0.027	0.961	0.994	0.999
Ours + NewCRFs [58]		2	**0.057**	**0.080**	**0.230**	**0.025**	**0.971**	**0.996**	**0.999**

- On Indoor NYUv2

Visual Result



Thanks For Watching!