I^2-SLAM: Inverting Imaging Process for Robust Photorealistic Dense SLAM


Gwangtak Bae*


Changwoon Choi*


Hyeongjun Heo


Sang Min Kim


Young Min Kim

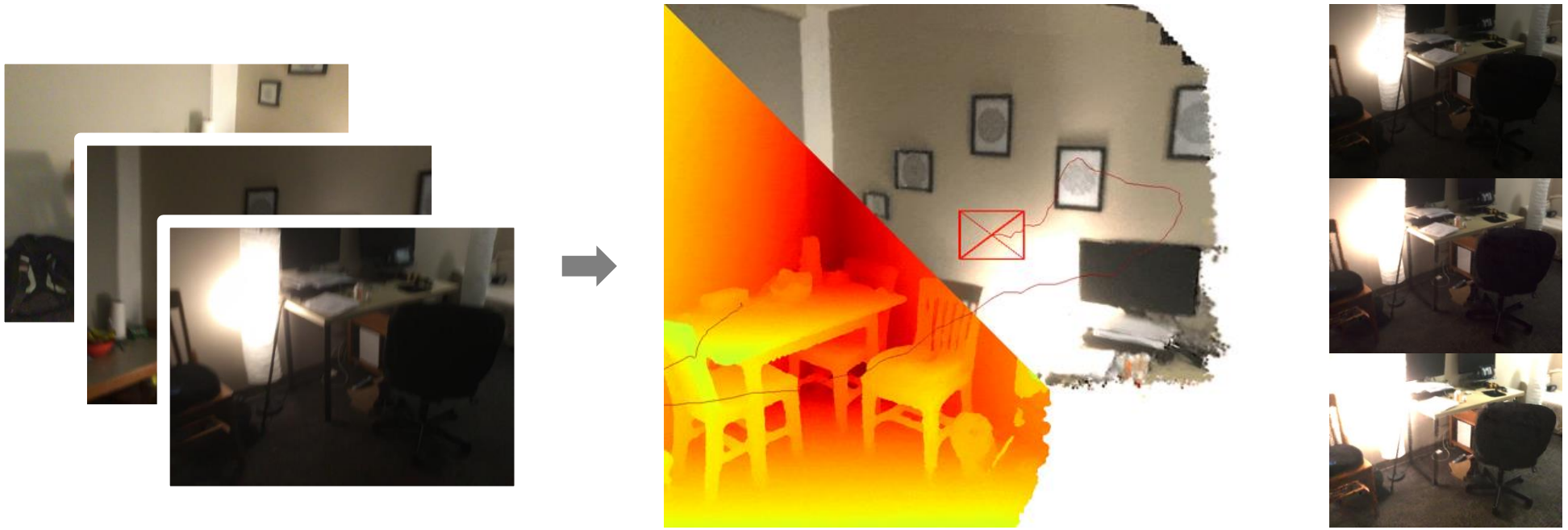
3D Vision Lab, Seoul National University
(* equal contribution)

Summary

From casually captured video,

I²-SLAM *reconstructs photorealistic and sharp HDR maps*

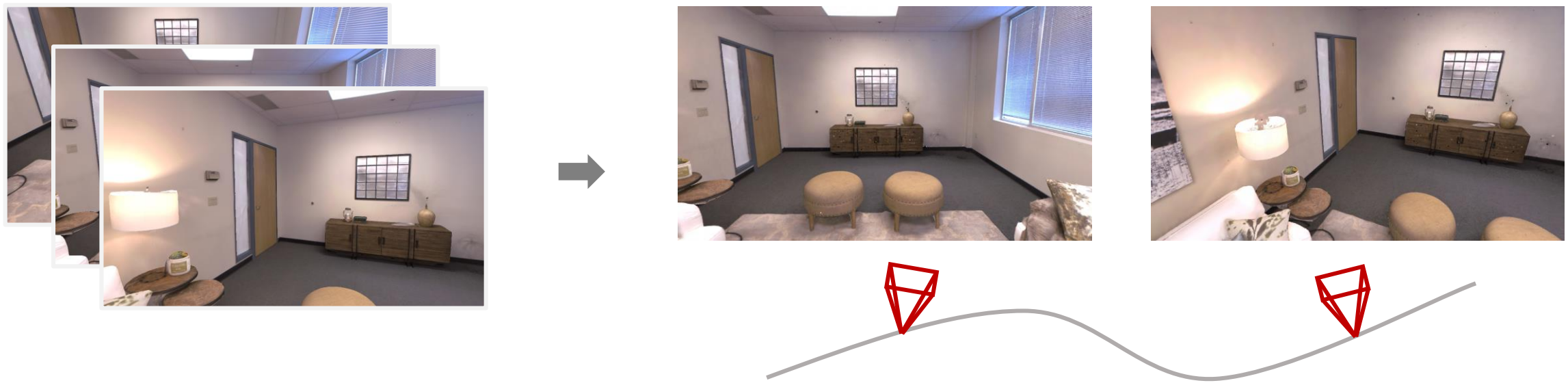
by inverting imaging process



Problems of previous works

✓ Photorealistic Dense SLAM

- Use NeRF or 3DGS for map representation
- Strength : Photorealistic / dense / detailed maps

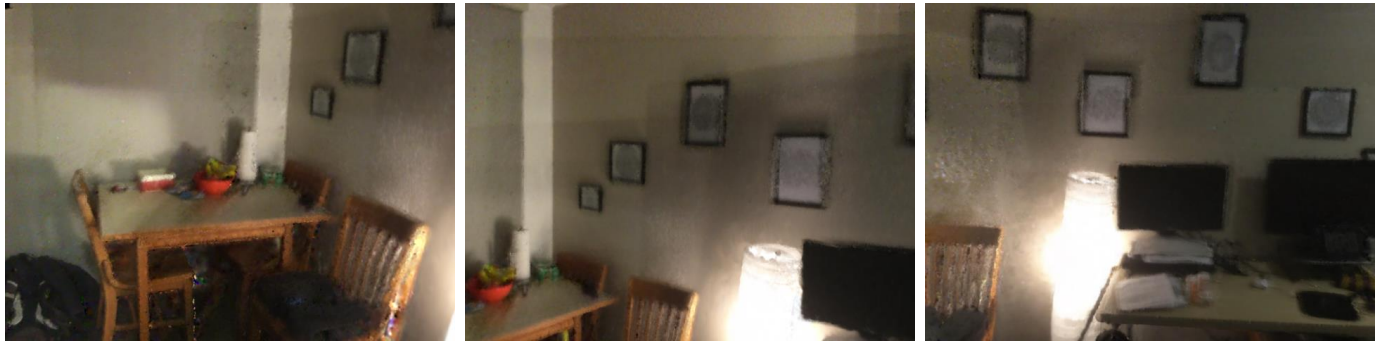


Problems of previous works

✓ Photorealistic Dense SLAM

- Weakness : Do not maintain their performances in real-world data

SplaTAM[1] in ***synthetic data***



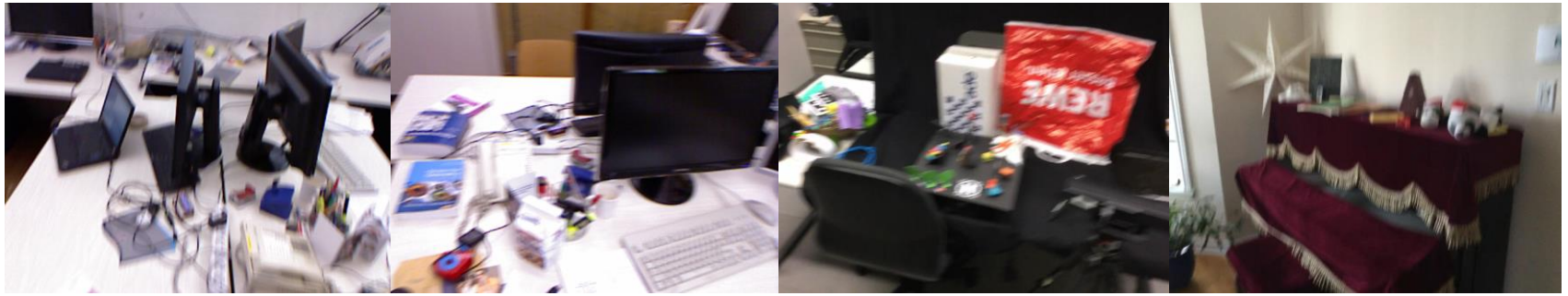
SplaTAM[1] in ***real-world data***

[1] Keetha et al., SplaTAM: Splat Track & Map 3D Gaussians for Dense RGB-D SLAM, CVPR 2024

Challenges in real-world data

✓ Motion blur

- Camera movement during exposure



✓ Varying appearances

- Auto exposure and white balancing

Method overview

✓ Key contribution

HDR radiance field

***I*²-SLAM**

simulates motion blur and tone mapping,

compares with input frames,

optimizes HDR map and camera poses

Input frame

Blurry LDR

Image formation process

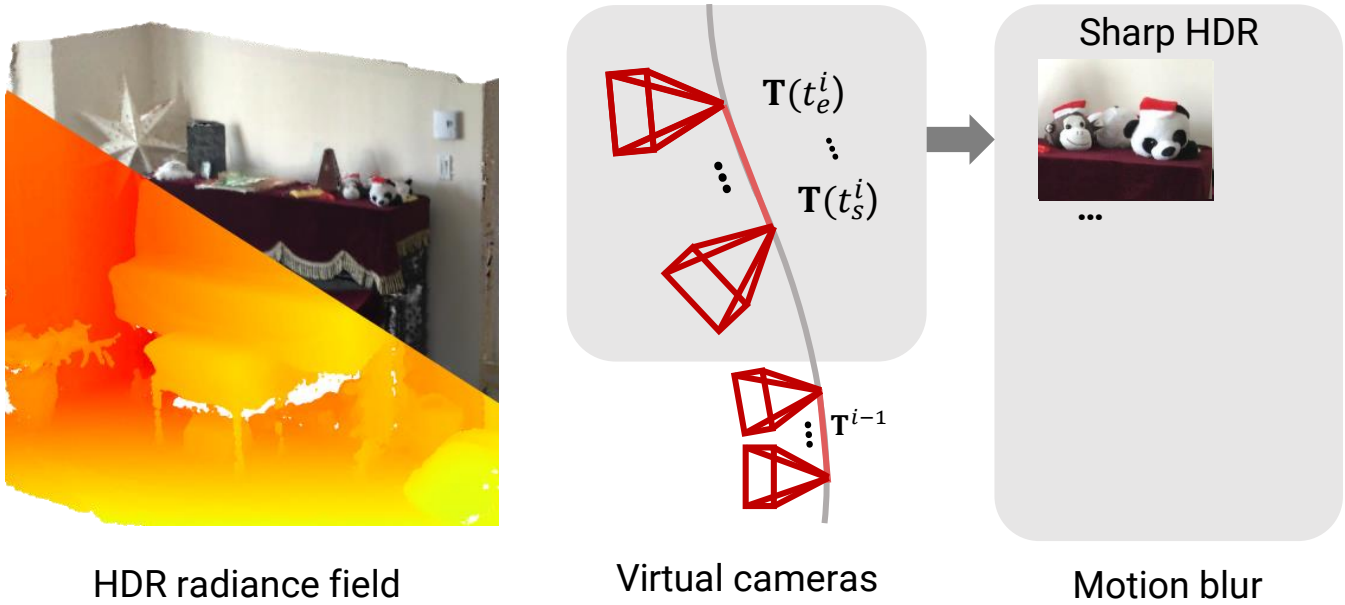
✓ Motion blur

- Camera trajectory during exposure

$$C^i_{\mathrm{HDR}}(\mathbf{p}) = \int_{t_s^i}^{t_e^i} \mathbf{c}(\mathbf{T}(t), \mathbf{p}) dt$$

pixel location

Interpolated camera poses



HDR radiance field

Virtual cameras

Motion blur

Image formation process

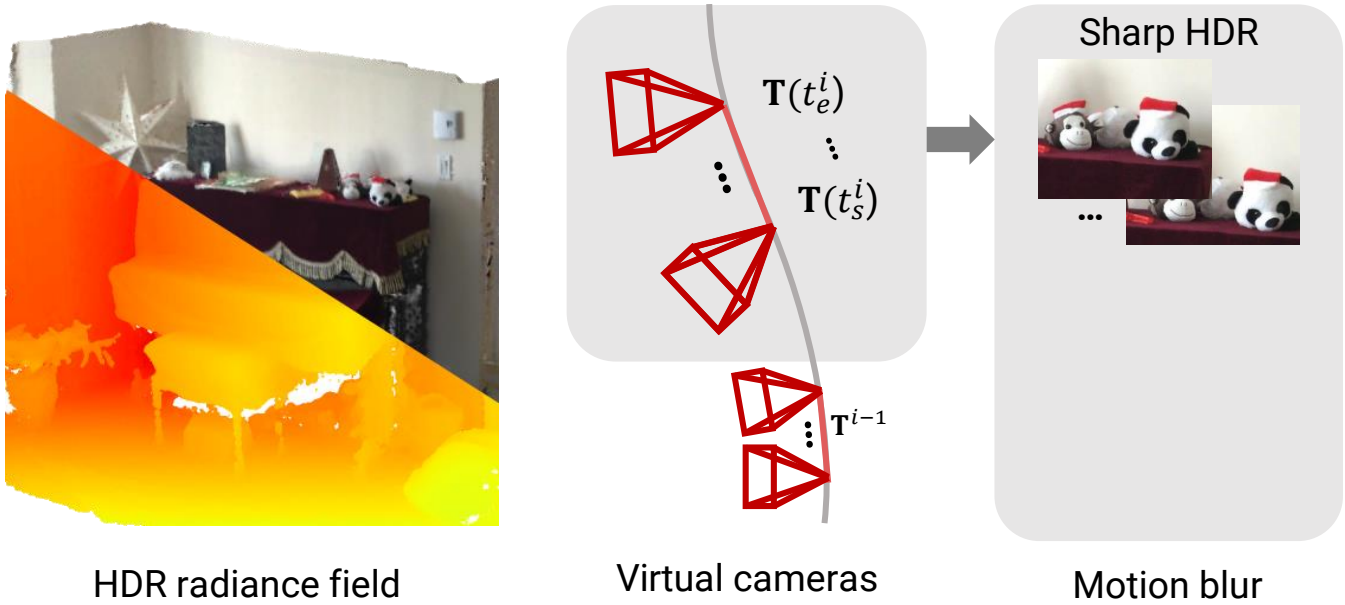
✓ Motion blur

- Camera trajectory during exposure

$$C^i_{\mathrm{HDR}}(\mathbf{p}) = \int_{t_s^i}^{t_e^i} \mathbf{c}(\mathbf{T}(t), \mathbf{p})\, dt$$

pixel location

Interpolated camera poses



HDR radiance field

Virtual cameras

Motion blur

Image formation process

✓ Motion blur

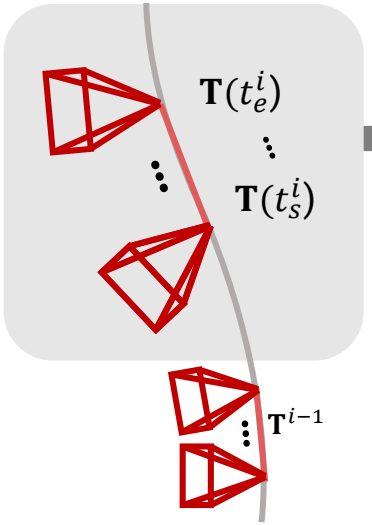
- Camera trajectory during exposure

$$C^i_{\text{HDR}}(\mathbf{p}) = \int_{t_s^i}^{t_e^i} \mathbf{c}(\mathbf{T}(t), \mathbf{p}) dt$$

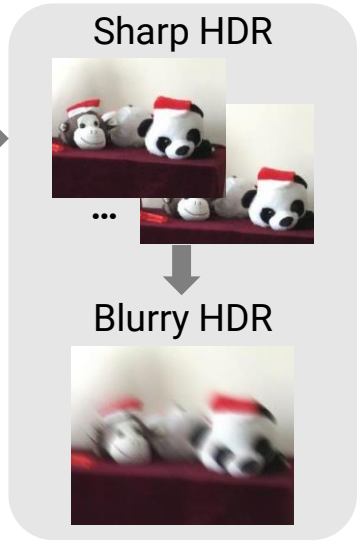
pixel location

Interpolated camera poses

HDR radiance field



Virtual cameras



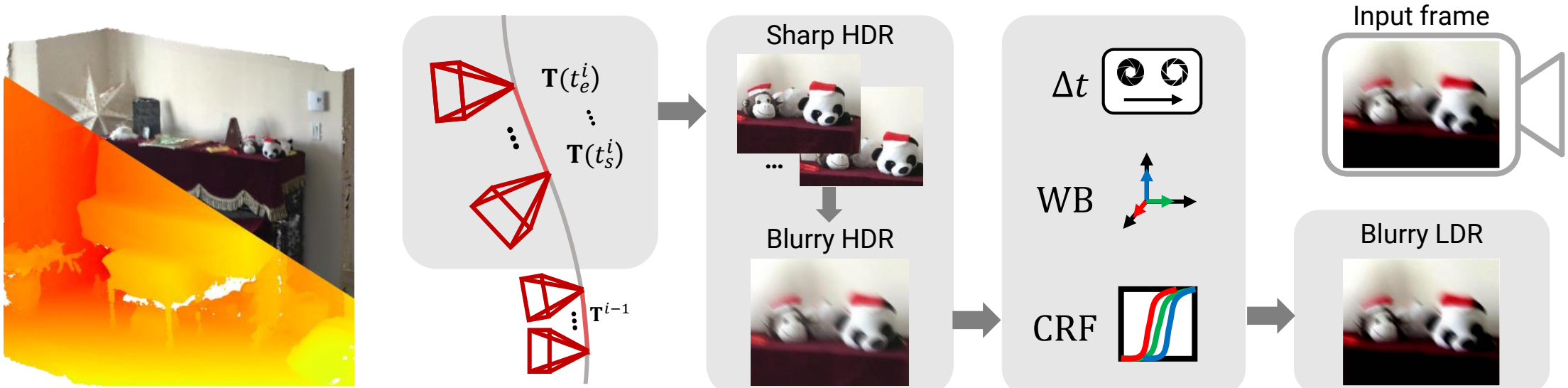
Motion blur

Image formation process

✓ Tone mapping

- Exposure time, white balance, camera response function

$$C^i_{\mathrm{LDR}}(\mathbf{p}) = \psi^i(C^i_{\mathrm{HDR}}(\mathbf{p})) \qquad \psi^i(\Delta t^i \cdot \mathbf{c}) = \mathrm{CRF}^i\left(\mathrm{WB}^i(\Delta t^i \cdot \mathbf{c})\right)$$



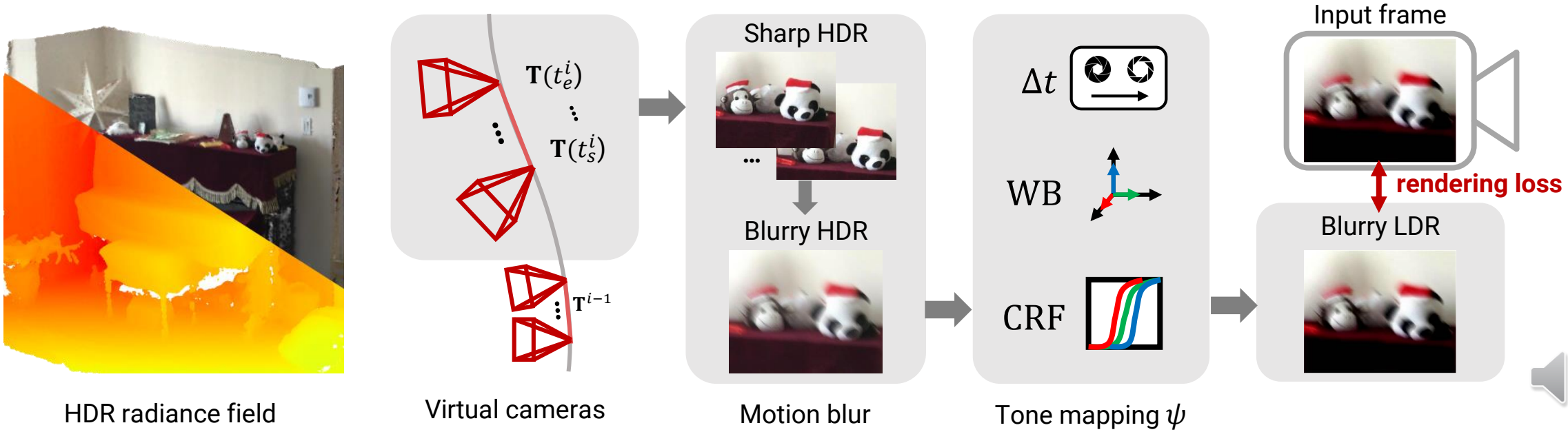
HDR radiance field | Virtual cameras | Motion blur | Tone mapping ψ

Image formation process

✓ Tone mapping

- Exposure time, white balance, camera response function

$$C^i_{\mathrm{LDR}}(\mathbf{p}) = \psi^i(C^i_{\mathrm{HDR}}(\mathbf{p})) \qquad \psi^i(\Delta t^i \cdot \mathbf{c}) = \mathrm{CRF}^i\left(\mathrm{WB}^i(\Delta t^i \cdot \mathbf{c})\right)$$

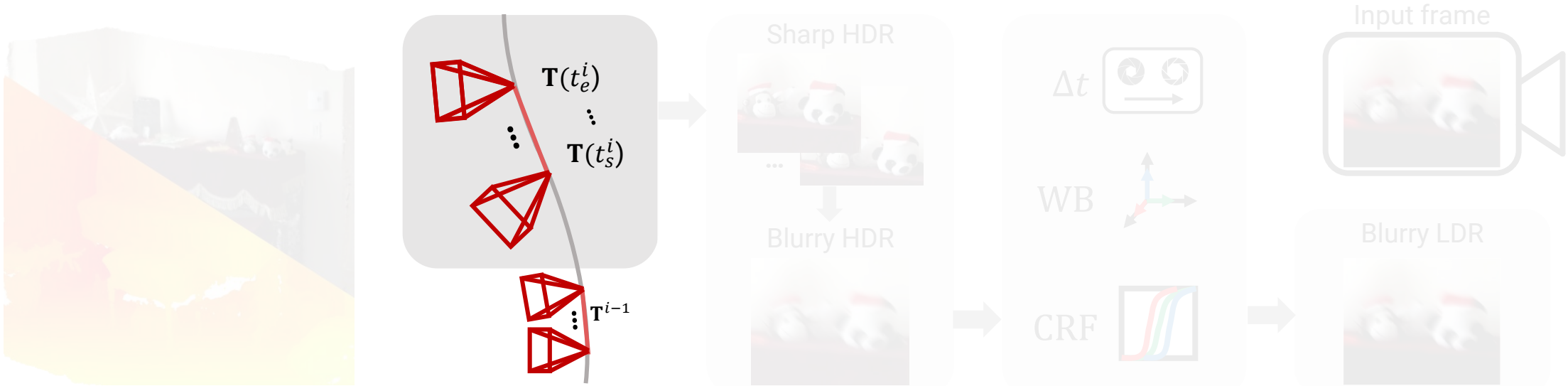


Trajectory regularization

✓ **Regularize per-frame camera trajectory**

1) Alignment with the global trajectory

2) Motion blur kernel size $\propto$ velocity $\cdot$ exposure time

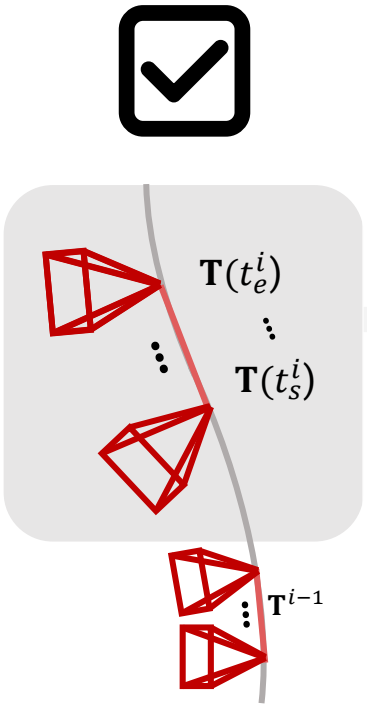


Trajectory regularization

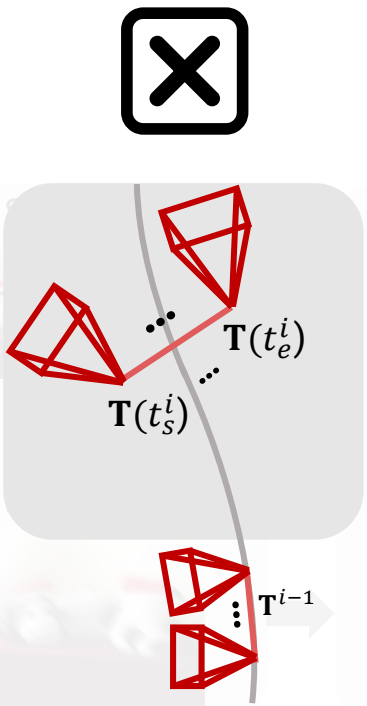
✓ **Regularize per-frame camera trajectory**

1) Alignment with the global trajectory

2) Motion blur kernel size $\propto$ velocity · exposure time



Virtual cameras



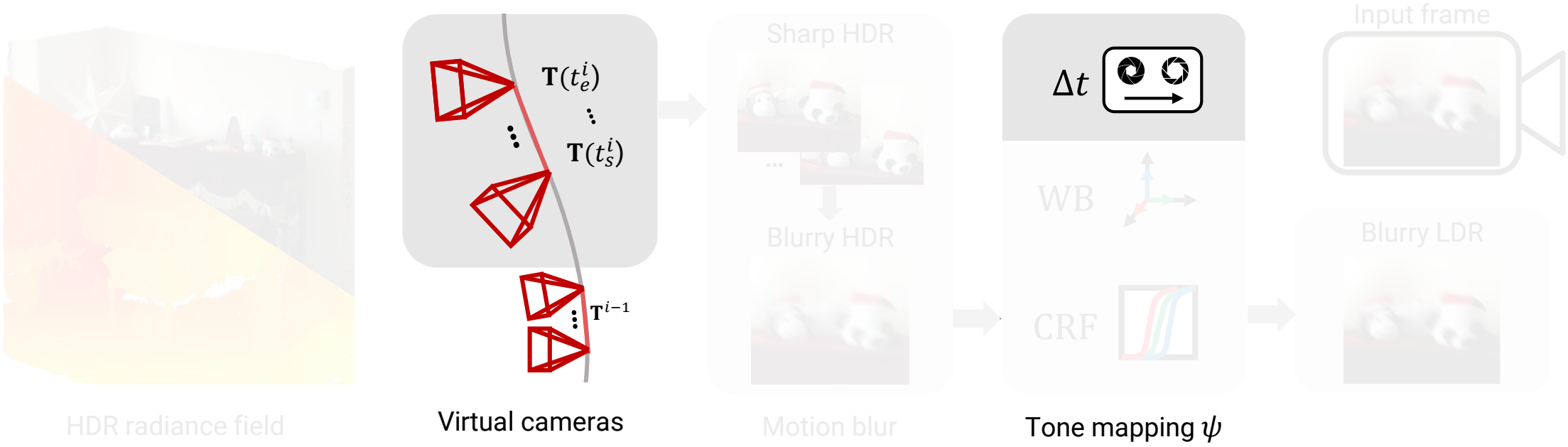
Virtual cameras

Trajectory regularization

✓ **Regularize per-frame camera trajectory**

1) Alignment with the global trajectory

2) Motion blur kernel size $\propto$ velocity · exposure time

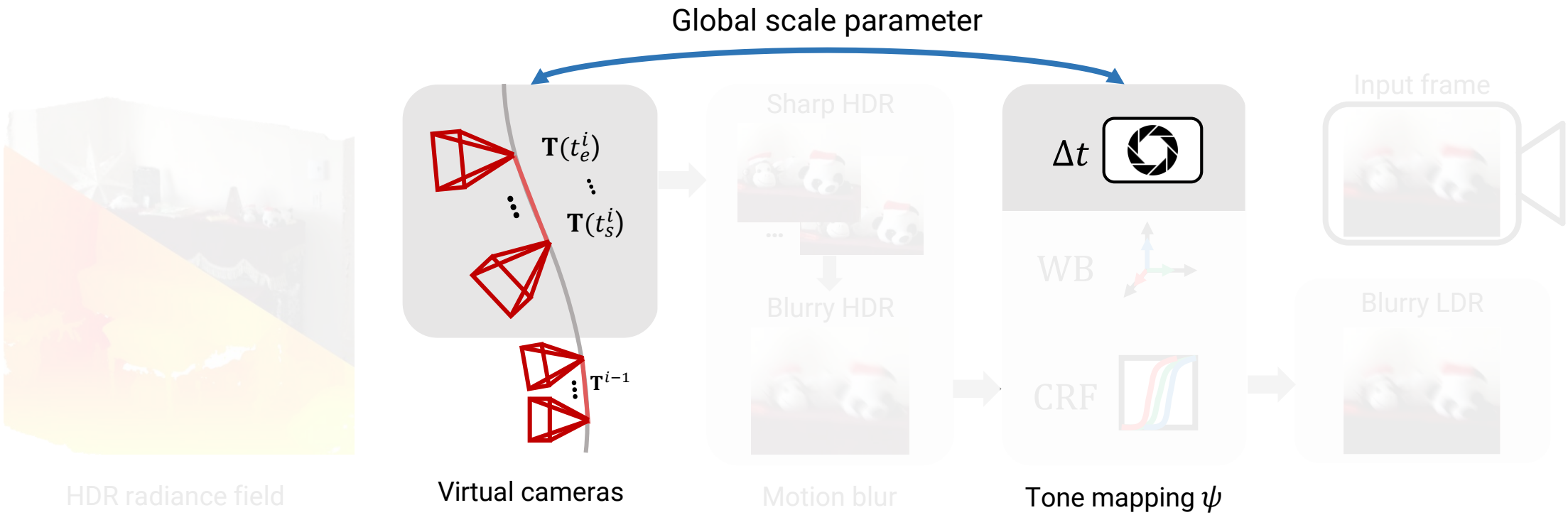


Trajectory regularization

✓ **Regularize per-frame camera trajectory**

1) Alignment with the global trajectory

2) Motion blur kernel size $\propto$ velocity · exposure time

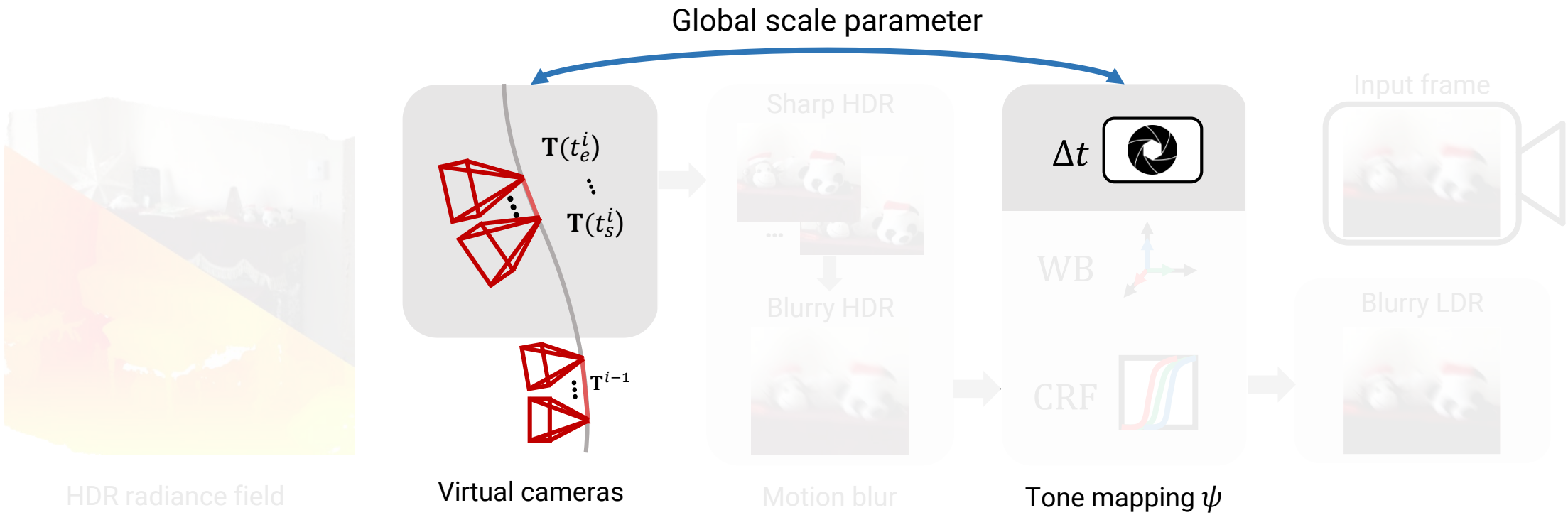


Trajectory regularization

✓ **Regularize per-frame camera trajectory**

1) Alignment with the global trajectory

2) Motion blur kernel size $\propto$ velocity · exposure time



Results

✓ Baselines

I^2-SLAM is a generic method that can be combined

with any existing photorealistic dense SLAM methods

		Modality	Representation
***I*^{2}-SLAM**	+	RGB-D SLAM	3DGS[1]
		RGB SLAM	NeRF[2]

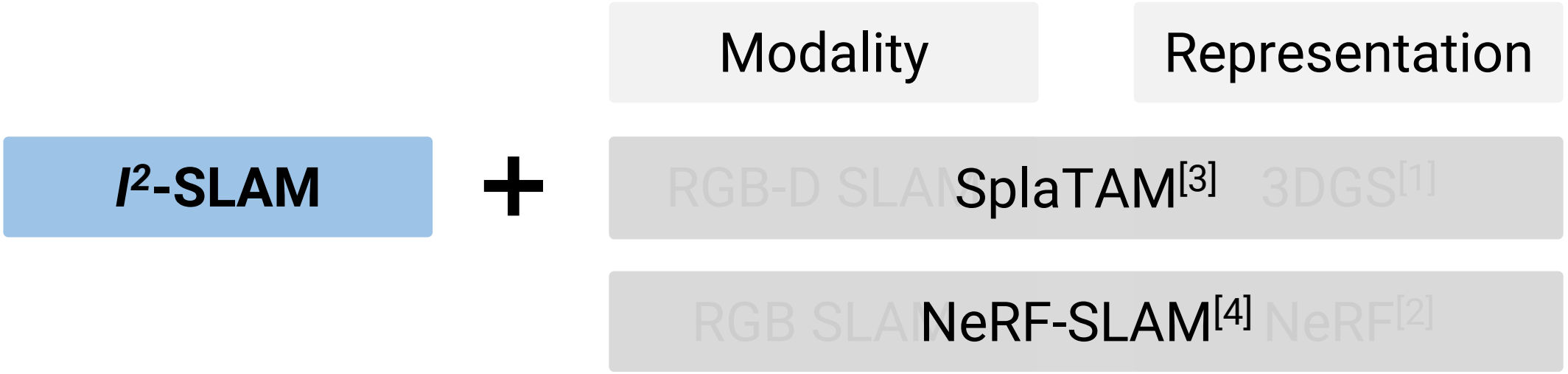
[1] Kerbl et al., 3D Gaussian Splatting for Real-Time Radiance Field Rendering, SIGGRAPH 2023
[2] Mildenhall et al., NeRF: Representing scenes as neural radiance fields for view synthesis, ECCV 2020

Results

✓ Baselines

I^2-SLAM is a generic method that can be combined

with any existing photorealistic dense SLAM methods

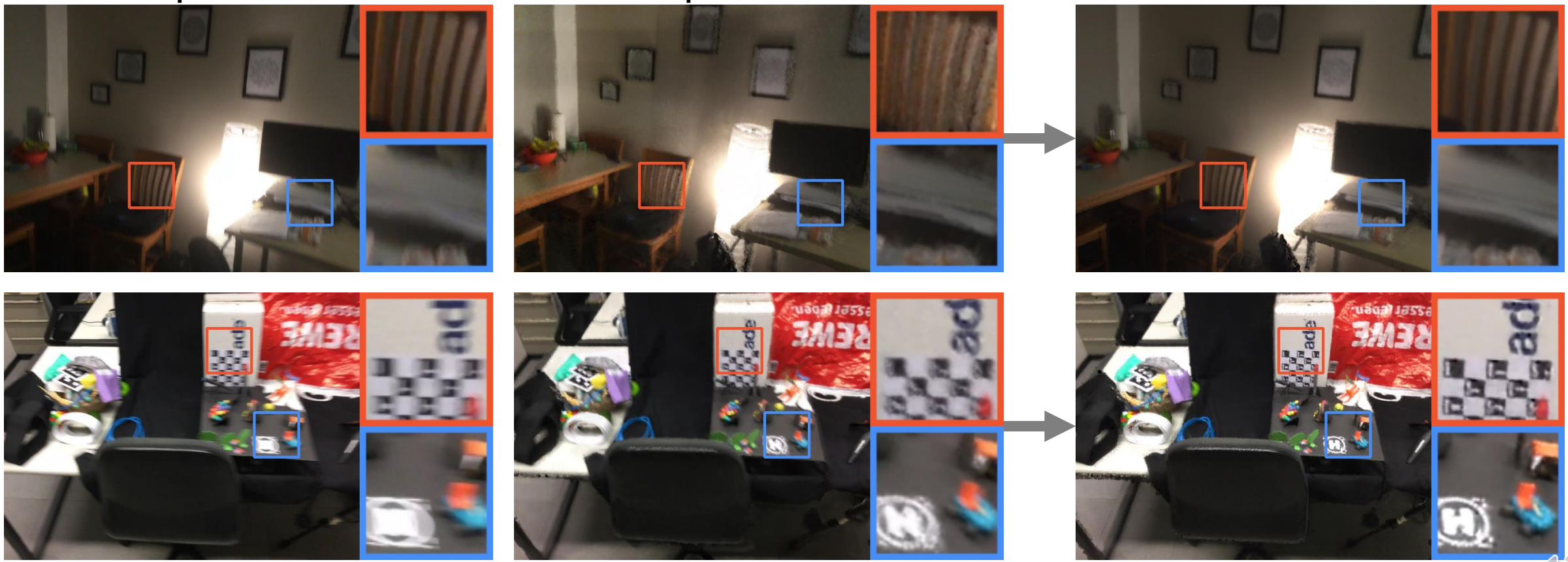


[1] Kerbl et al., 3D Gaussian Splatting for Real-Time Radiance Field Rendering, SIGGRAPH 2023
[2] Mildenhall et al., NeRF: Representing scenes as neural radiance fields for view synthesis, ECCV 2020
[3] Keetha et al., SplaTAM: Splat Track & Map 3D Gaussians for Dense RGB-D SLAM, CVPR 2024
[4] Rosinol et al., NeRF-SLAM: Real-Time Dense Monocular SLAM with Neural Radiance Fields, IROS 2023

Results

✓ RGB-D SLAM in ScanNet[1]

Input frame | SplaTAM[2] | *I*²-SLAM

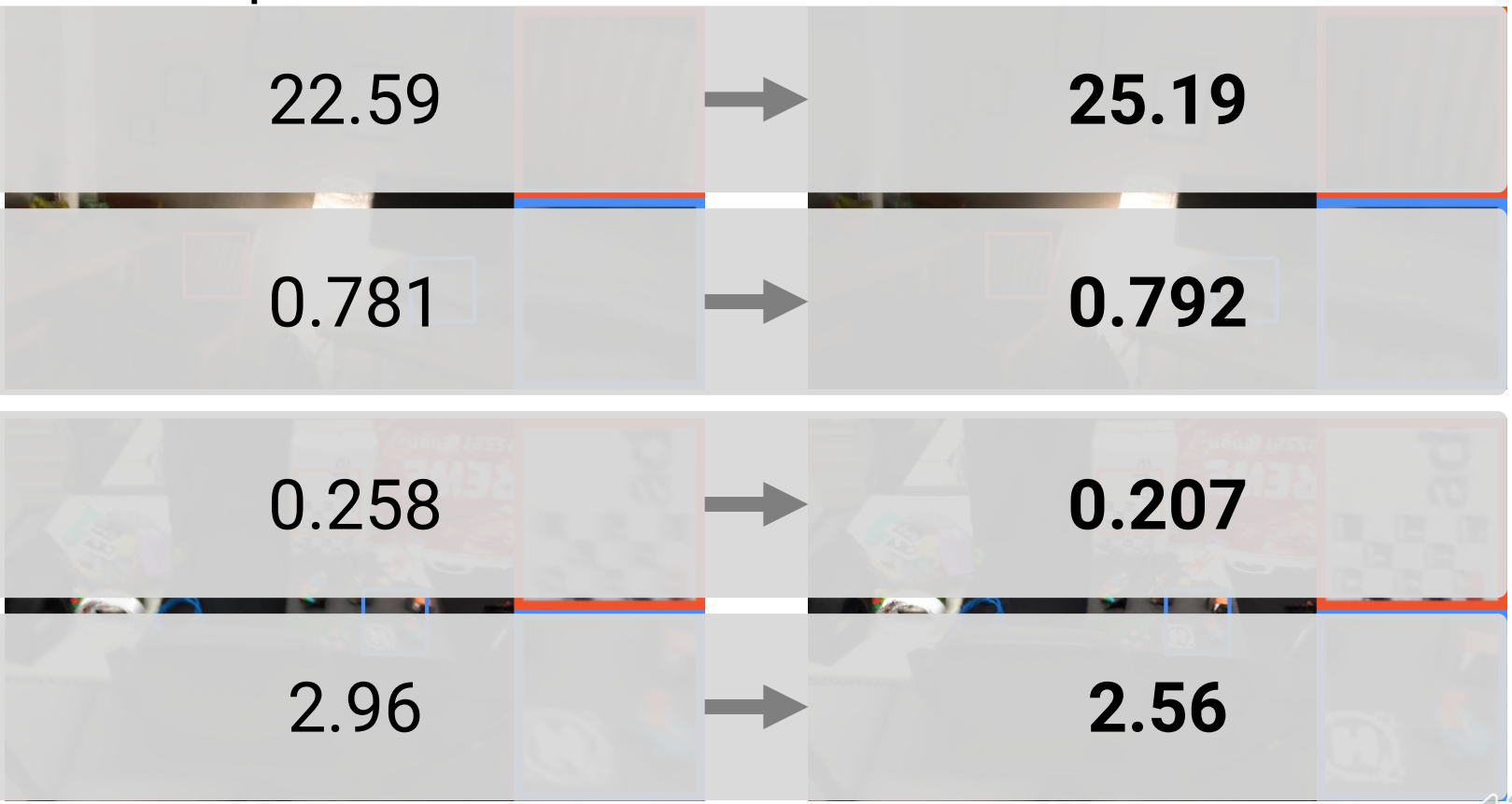


[1] Dai et al., Scannet: Richly-annotated 3d reconstructions of indoor scenes, CVPR 2017
[2] Keetha et al., SplaTAM: Splat Track & Map 3D Gaussians for Dense RGB-D SLAM, CVPR 2024

Results

✓ RGB-D SLAM in ScanNet[1]

	SplaTAM[2]		*I*²-SLAM
PSNR ↑	22.59	→	**25.19**
SSIM ↑	0.781	→	**0.792**
LPIPS ↓	0.258	→	**0.207**
ATE-RMSE ↓	2.96	→	**2.56**



[1] Dai et al., Scannet: Richly-annotated 3d reconstructions of indoor scenes, CVPR 2017
[2] Keetha et al., SplaTAM: Splat Track & Map 3D Gaussians for Dense RGB-D SLAM, CVPR 2024

Results

✓ RGB SLAM in TUM[1]

Input frame | NeRF-SLAM[2] | *I*²-SLAM

[1] Sturm et al., A Benchmark for the Evaluation of RGB-D SLAM Systems, IROS 2012
[2] Rosinol et al., NeRF-SLAM: Real-Time Dense Monocular SLAM with Neural Radiance Fields, IROS 2023

Results

✓ RGB SLAM in TUM[1]

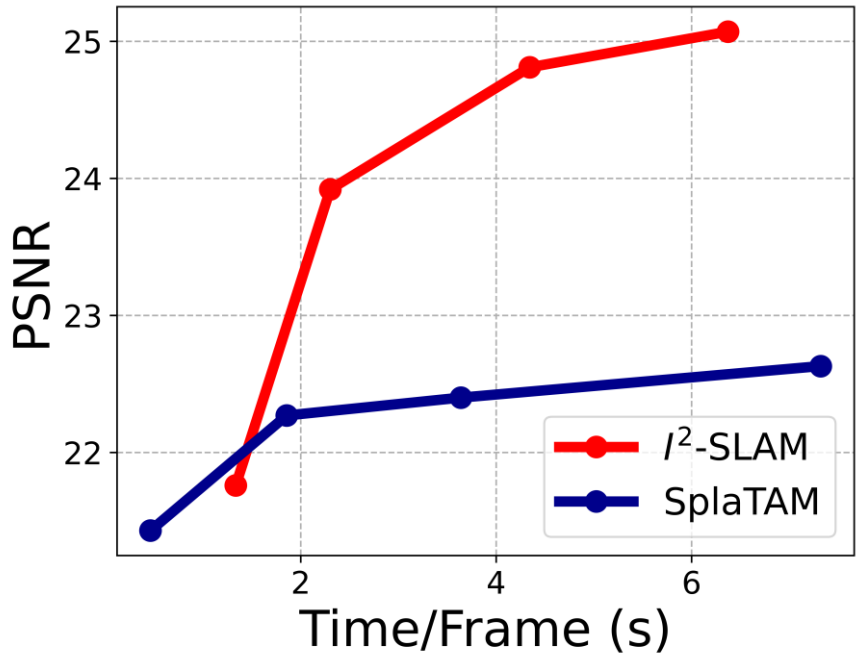
	NeRF-SLAM[2]		*I*²-SLAM
PSNR ↑	26.89	→	**29.40**
SSIM ↑	0.817	→	**0.861**
LPIPS ↓	0.227	→	**0.151**
Depth L1 ↓	19.91	→	**14.86**
ATE-RMSE ↓	3.21	→	**1.28**

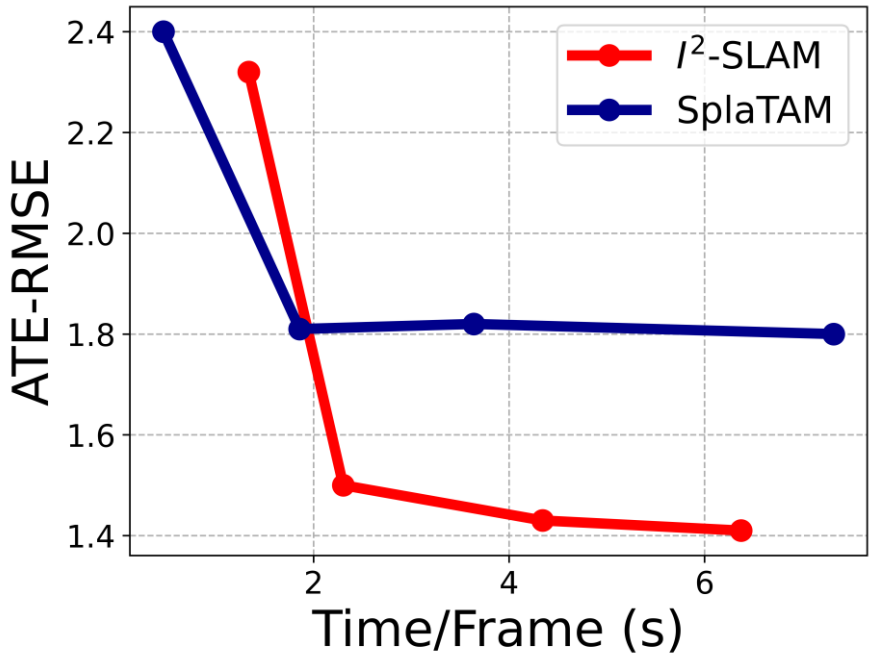
[1] Sturm et al., A Benchmark for the Evaluation of RGB-D SLAM Systems, IROS 2012
[2] Rosinol et al., NeRF-SLAM: Real-Time Dense Monocular SLAM with Neural Radiance Fields, IROS 2023

Runtime analysis

✓ Performance vs. runtime

- Varying number of iterations used for tracking and mapping
- Better performance when using a similar runtime





Conclusion

- ✓ ***I*²-SLAM** reconstructs sharp HDR maps

  from casually captured inputs

  by inverting imaging process

- ✓ ***I*²-SLAM** can be incorporated

  into existing SLAM methods using NeRF or 3DGS