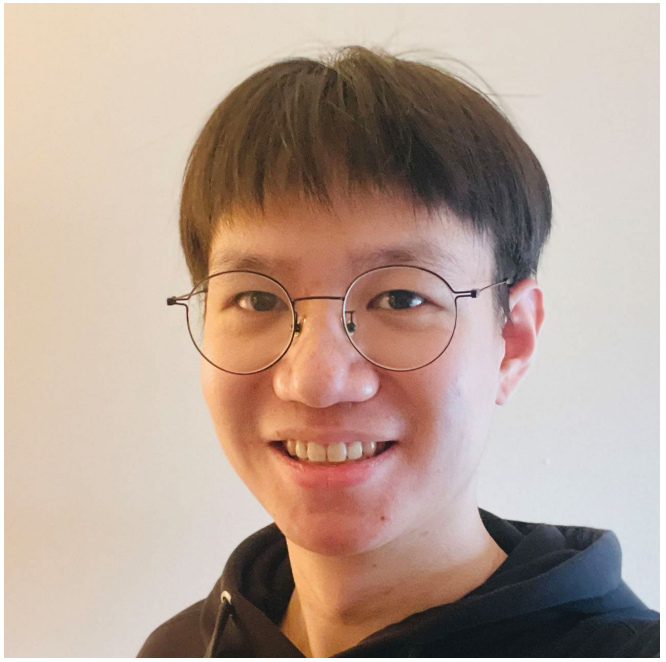
Unbounded Scene Modeling

Lingjie Liu

Sept 23, 2024

Announcements for Next Week's Classes



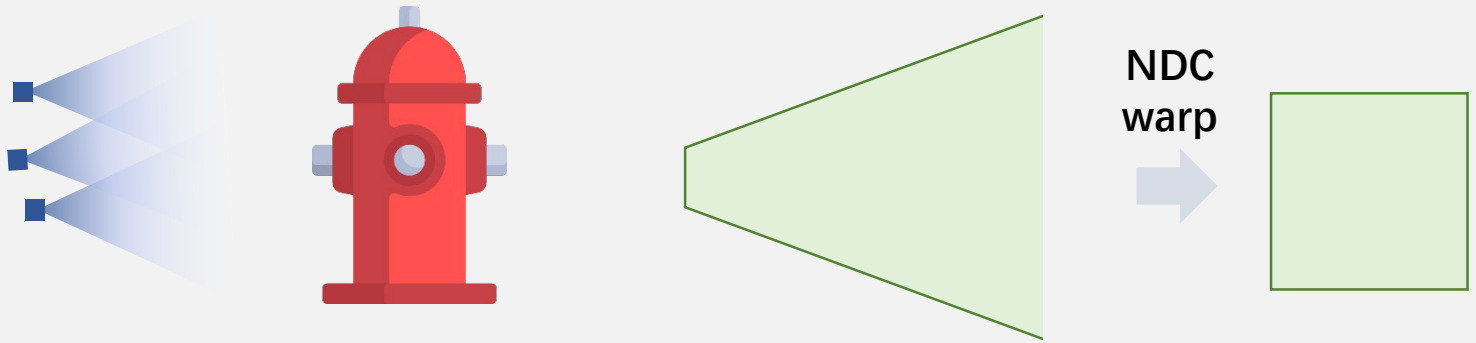
Jiahui Lei (Senior PhD at Penn)
Sept 30 (Monday) 3:30-5PM

Oct 2 (Wed) 3:30-5PM

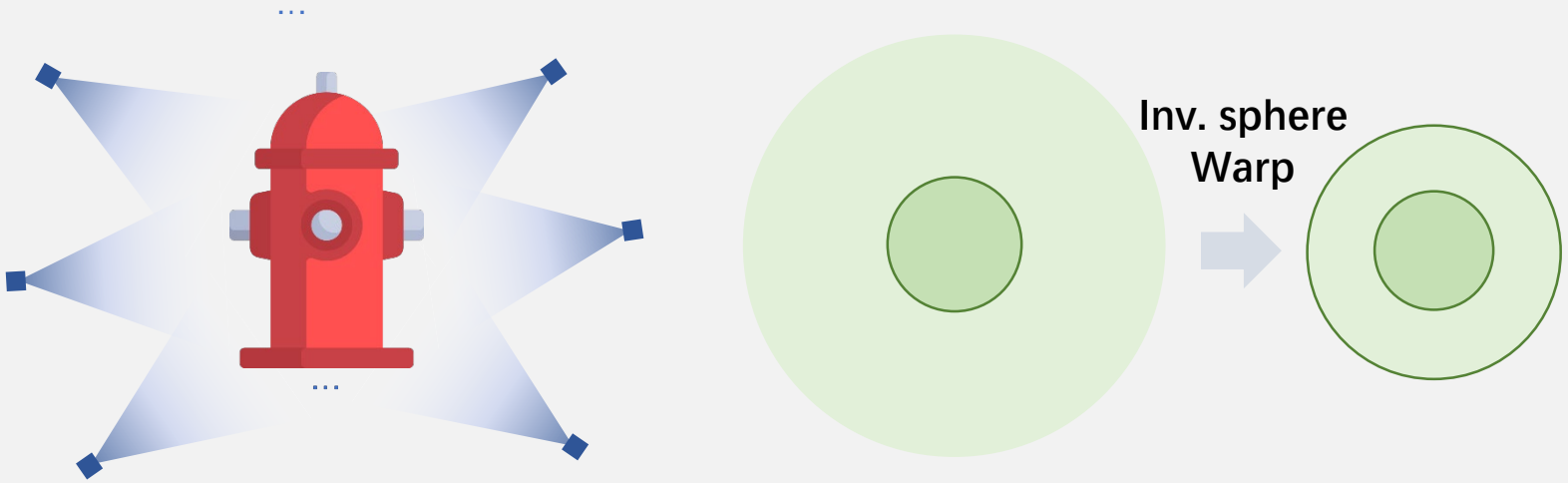
Background: NeRF rendering of unbounded scene/open scene

Bounded

Unbounded

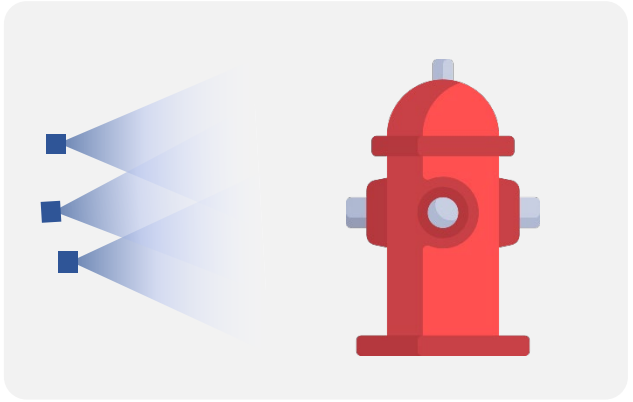


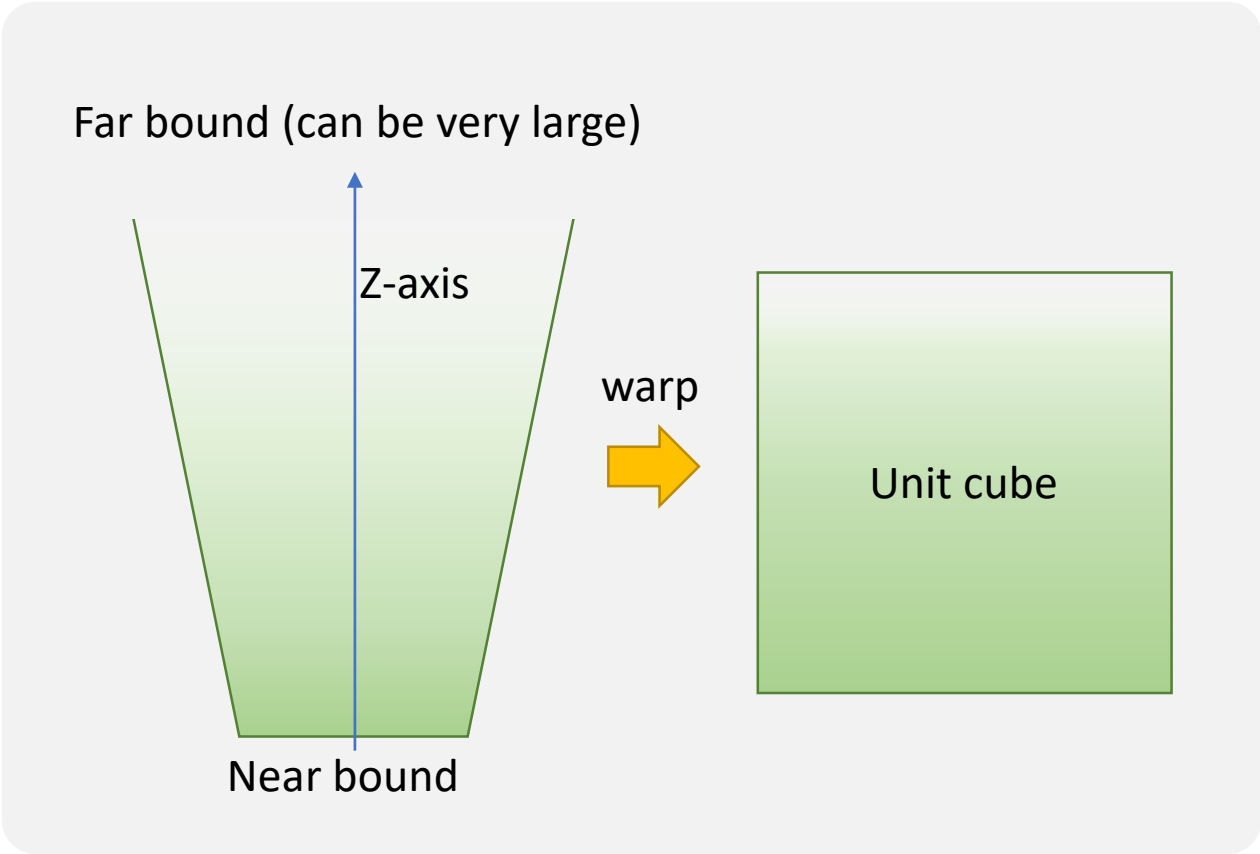
(a) Forward-facing (LLFF)



(b) 360°, object-centric (NeRF-360)

NDC space warping

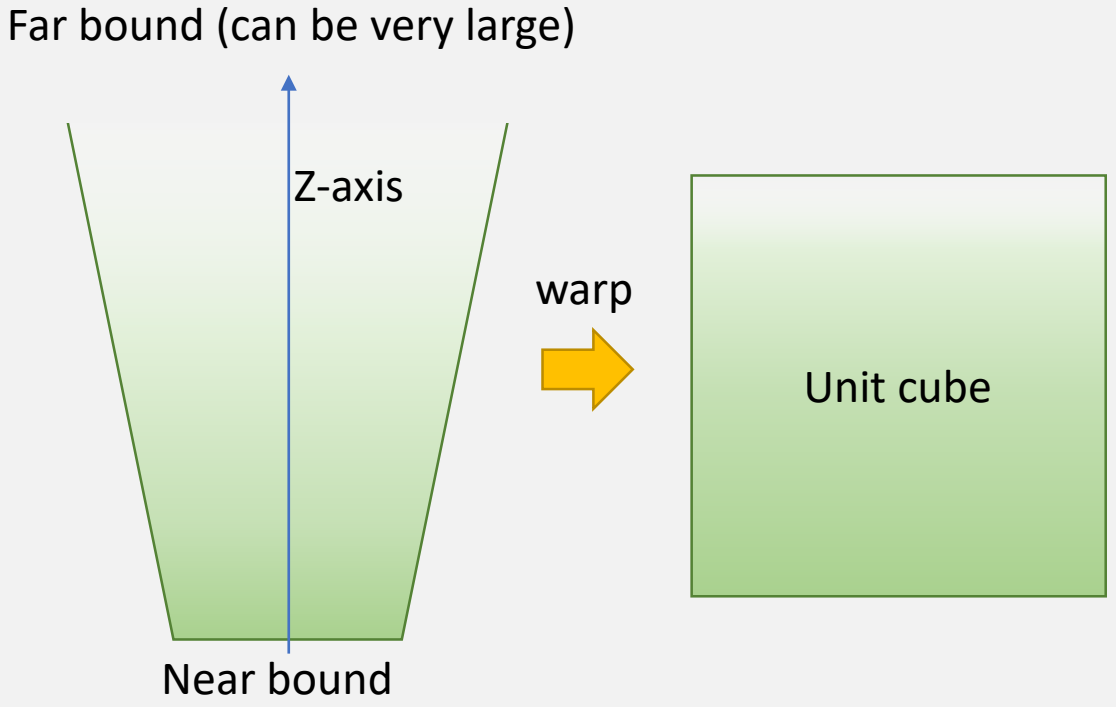




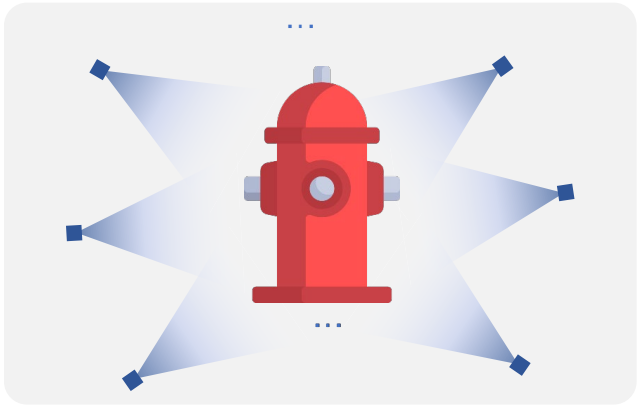
NDC space warping

$$\begin{pmatrix} \frac{n}{r} & 0 & 0 & 0 \\ 0 & \frac{n}{t} & 0 & 0 \\ 0 & 0 & \frac{-(f+n)}{f-n} & \frac{-2fn}{f-n} \\ 0 & 0 & -1 & 0 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \\ 1 \end{pmatrix} = \begin{pmatrix} \frac{n}{r}x \\ \frac{n}{t}y \\ \frac{-(f+n)}{f-n}z - \frac{-2fn}{f-n} \\ -z \end{pmatrix}$$

$$\text{project} \rightarrow \begin{pmatrix} \frac{n}{r}\frac{x}{-z} \\ \frac{n}{t}\frac{y}{-z} \\ \frac{(f+n)}{f-n} - \frac{2fn}{f-n}\frac{1}{-z} \end{pmatrix}$$



Inverse sphere space warping



Outer circle (can be very large)

Inner circle

warp

Inverse sphere space warping (contraction)

$$\text{contract}(\mathbf{x}) = \begin{cases} \mathbf{x} & \|\mathbf{x}\| \leq 1 \\ \left(2 - \frac{1}{\|\mathbf{x}\|}\right)\left(\frac{\mathbf{x}}{\|\mathbf{x}\|}\right) & \|\mathbf{x}\| > 1 \end{cases}$$

Outer circle (can be very large)

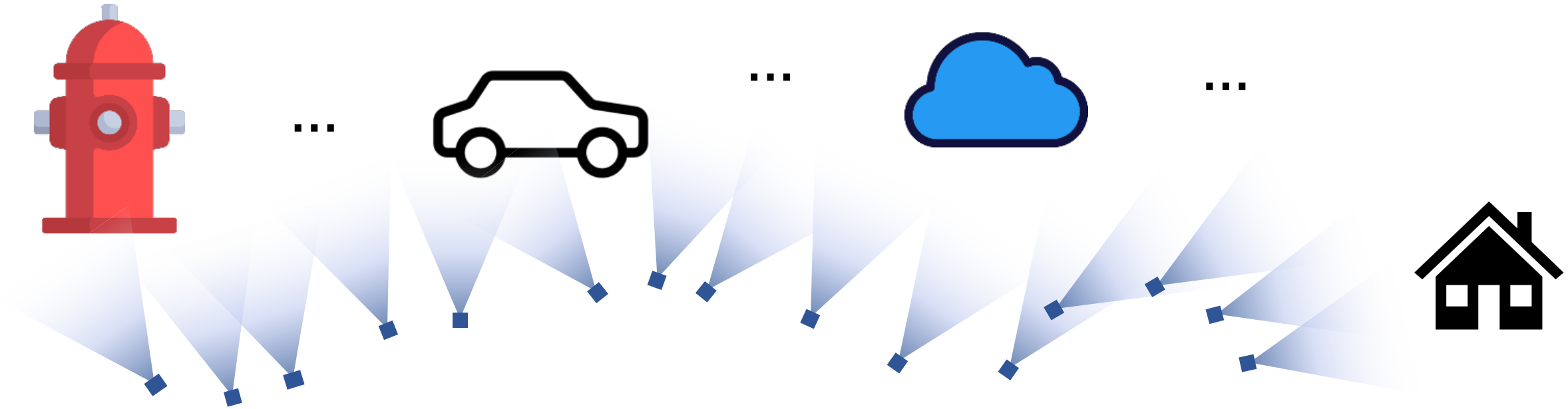
F^2-NeRF: Fast Neural Radiance Field Training with Free Camera Trajectories

Peng Wang[1,2*], Yuan Liu[1*], Zhaoxi Chen[2], Lingjie Liu[3], Ziwei Liu[2], Taku Komura[1], Christian Theobalt[3], Wenping Wang[4]

[1]The University of Hong Kong
[2]S-Lab, Nanyang Technological University
[3]MPI Informatics, Saarland Informatics Campus
[4]Texas A&M University

Problem

What if the input camera trajectory is very irregular?



What if the input camera trajectory is very irregular? – We call that a “free” trajectory

When the trajectory is irregular…

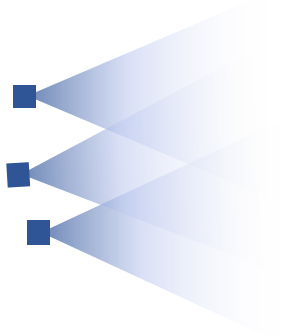
DVGO | Plenoxels | Instant-NGP | F^2-NeRF | Ground truth

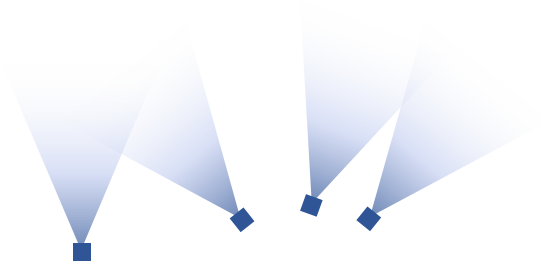
Our idea

Adaptive warping method from input trajectories

Fixed warping

requires





determines

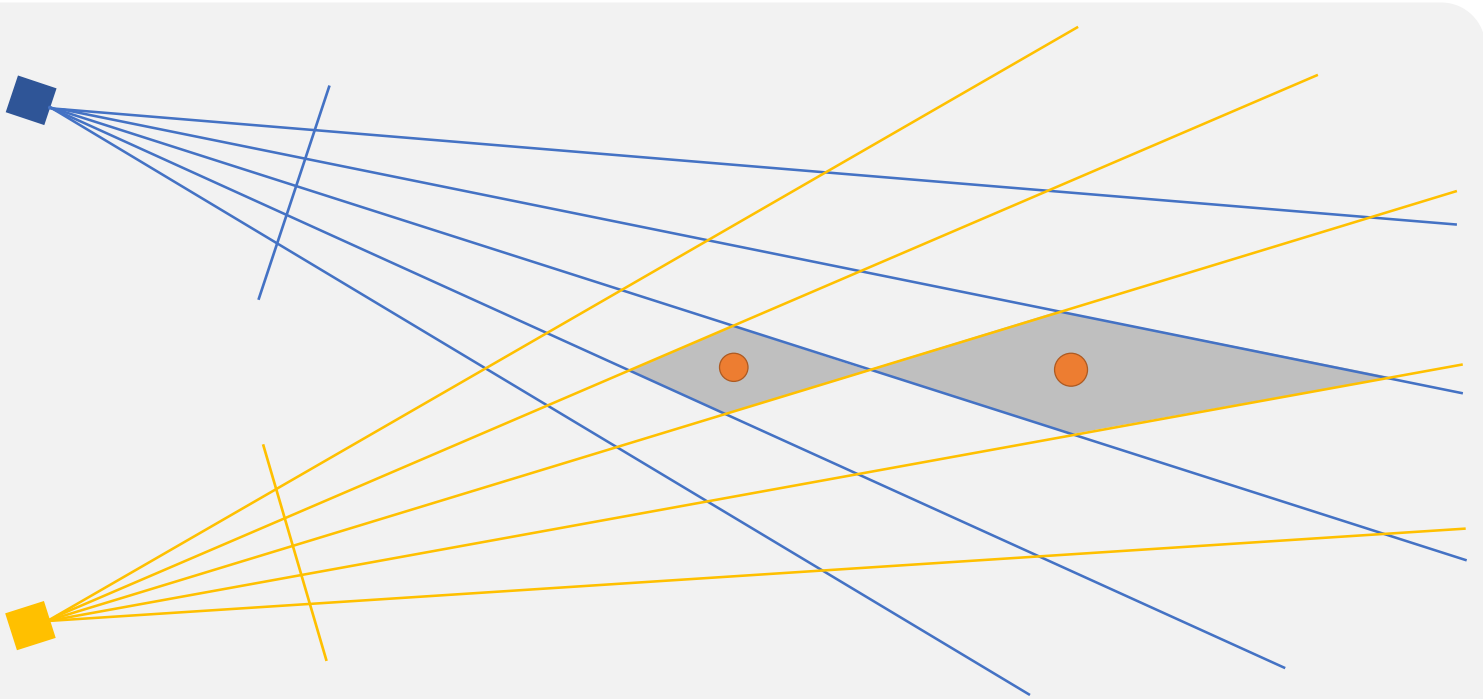
Adaptive warping

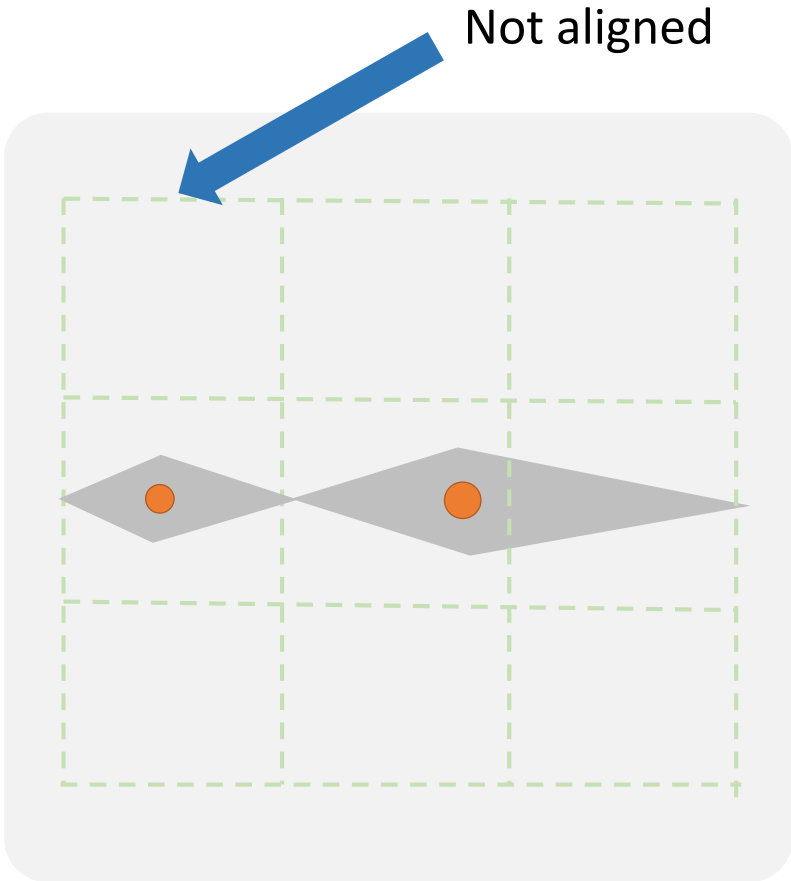
Why space warping

Very far region

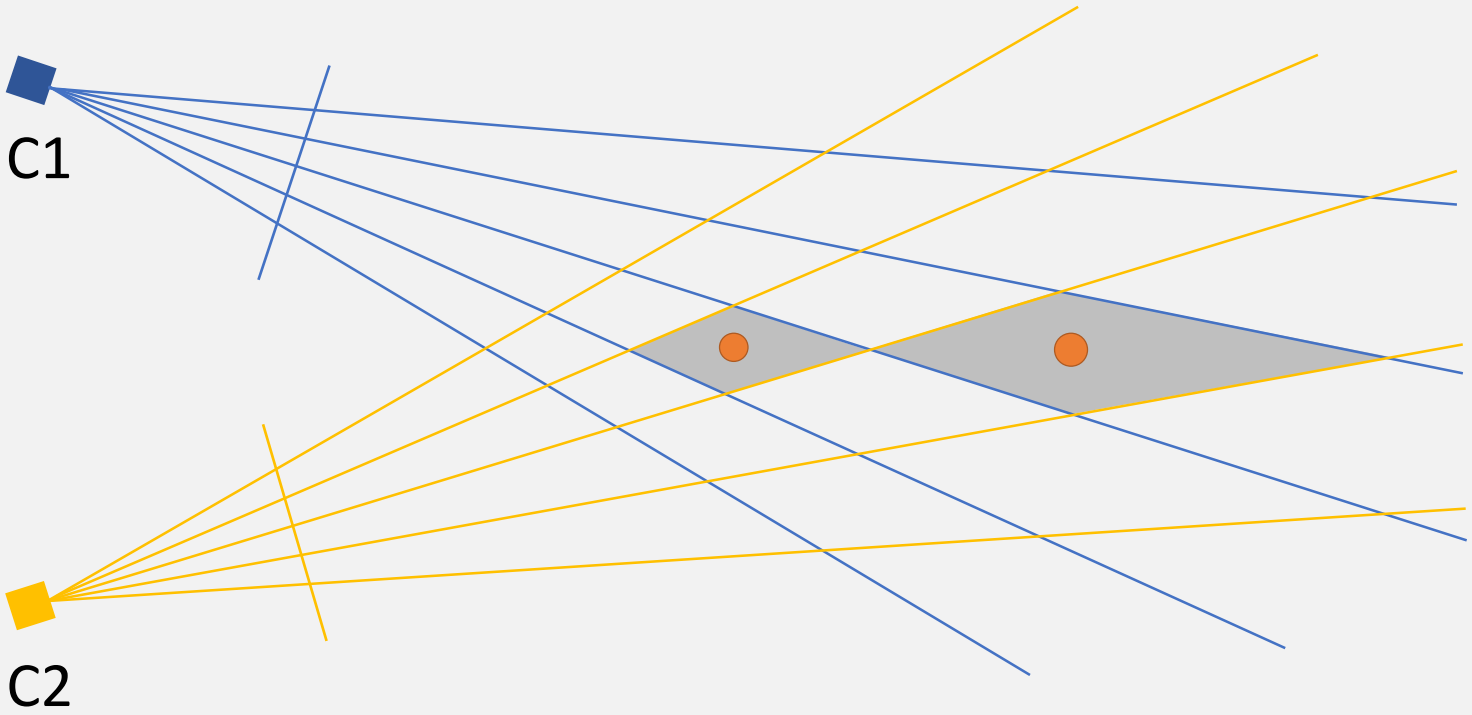
The “resolution” of the scene is spatial-variant

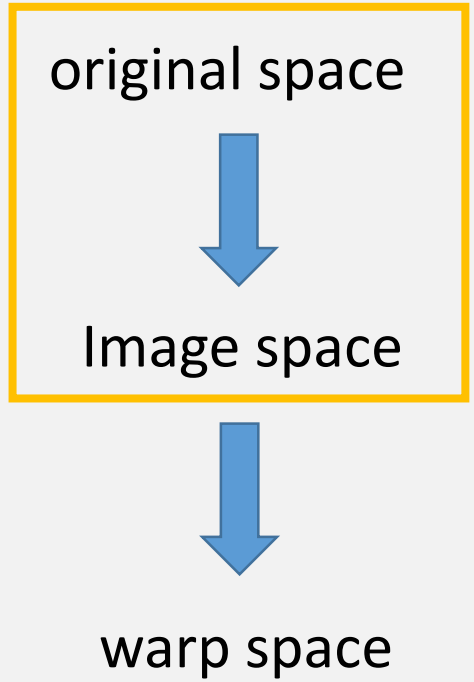
Why space warping





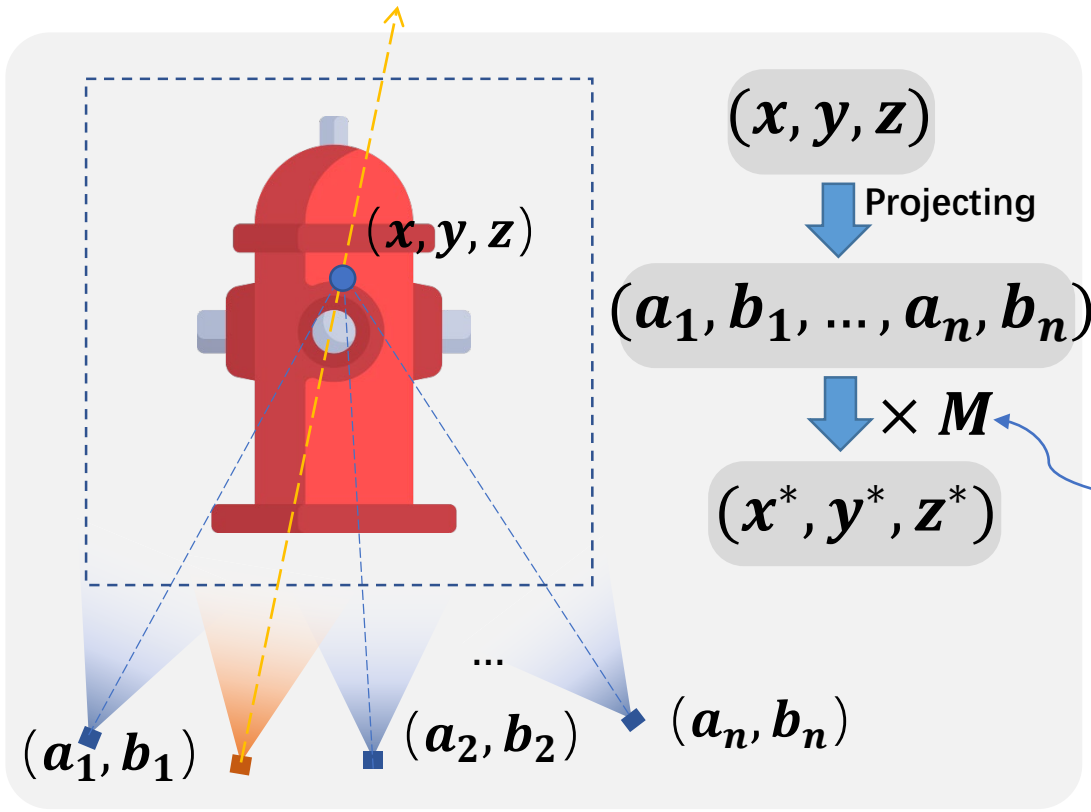
Key idea: Perspective warping (2D)

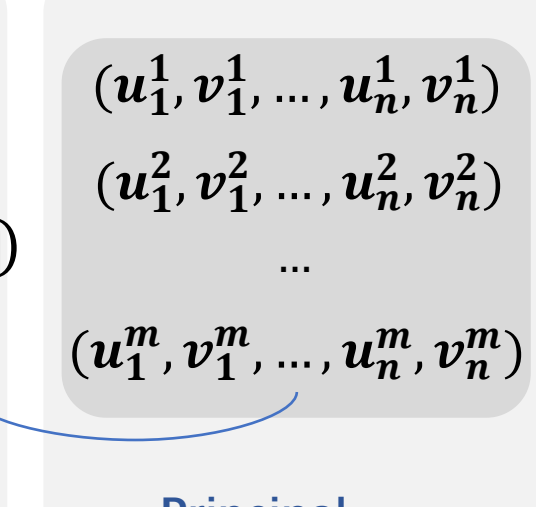




$$F(\boldsymbol{x}) = (C_1(\boldsymbol{x}), C_2(\boldsymbol{x}))$$

Key idea: Perspective warping (3D)

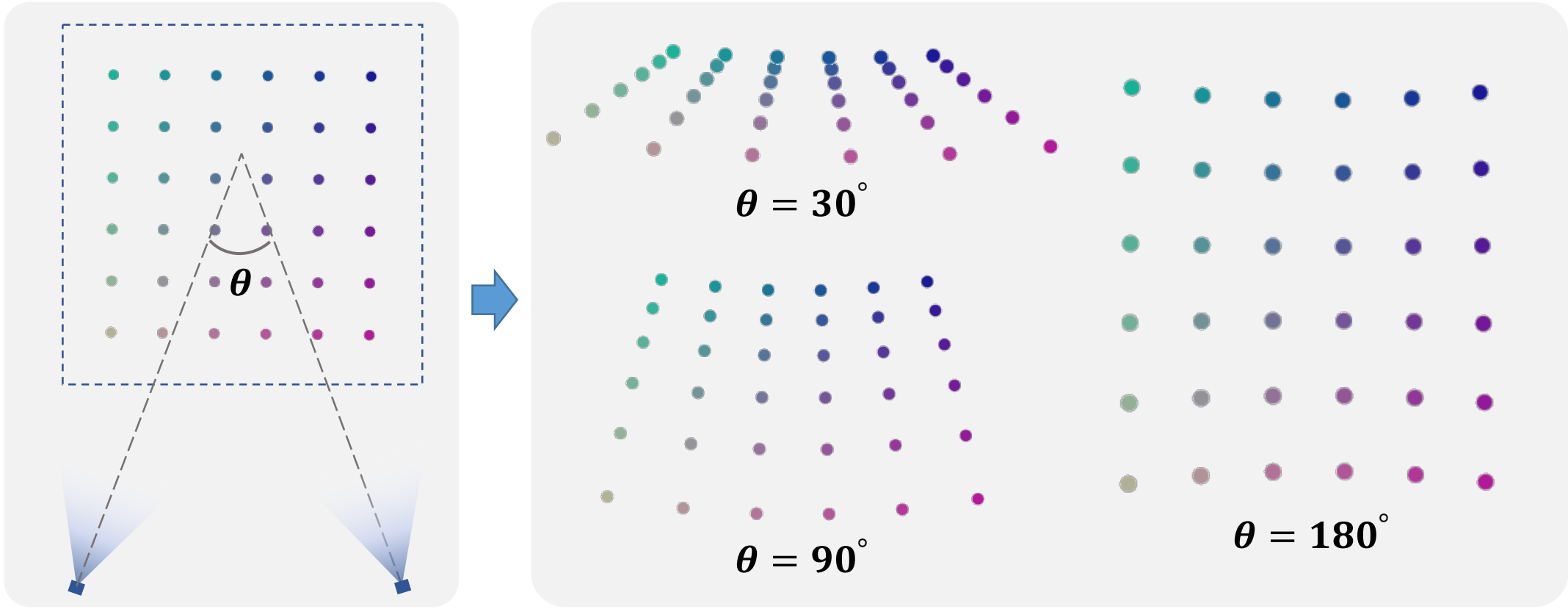




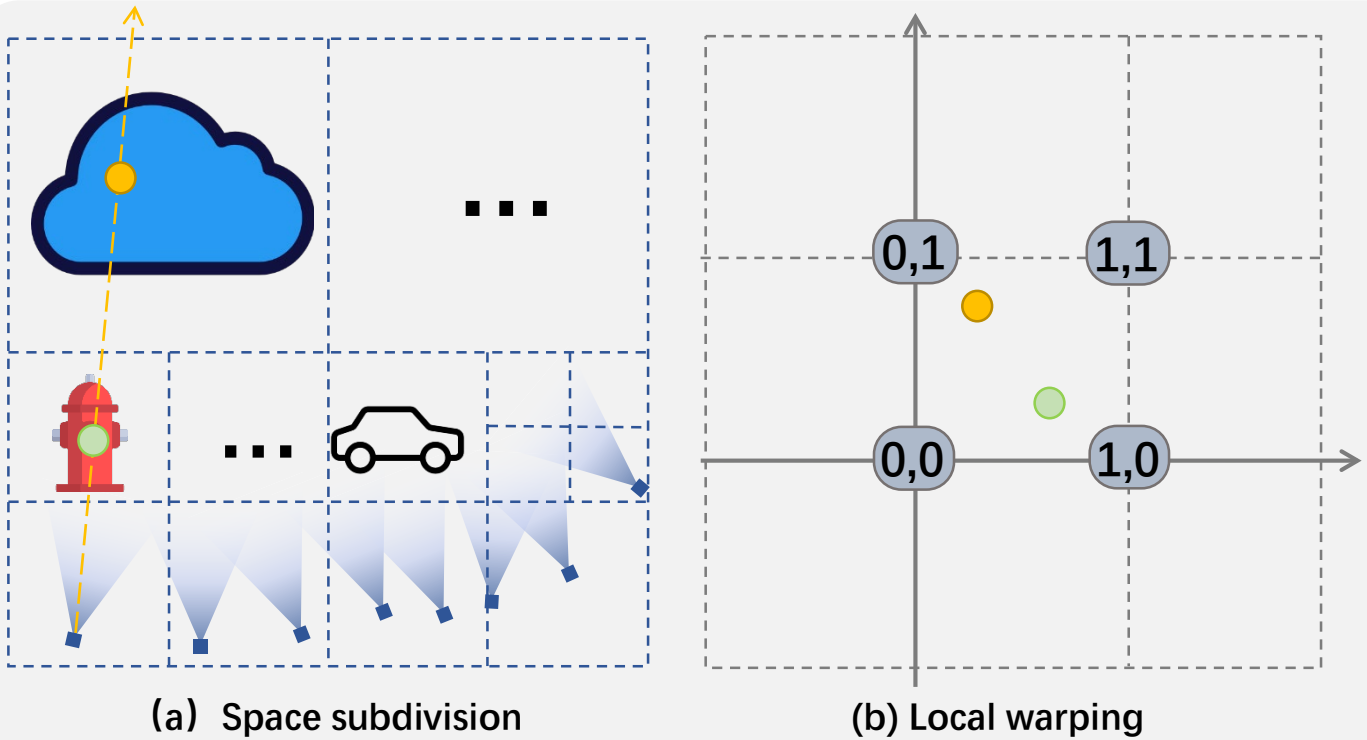
(a) Perspective warping

(b) Matrix construction

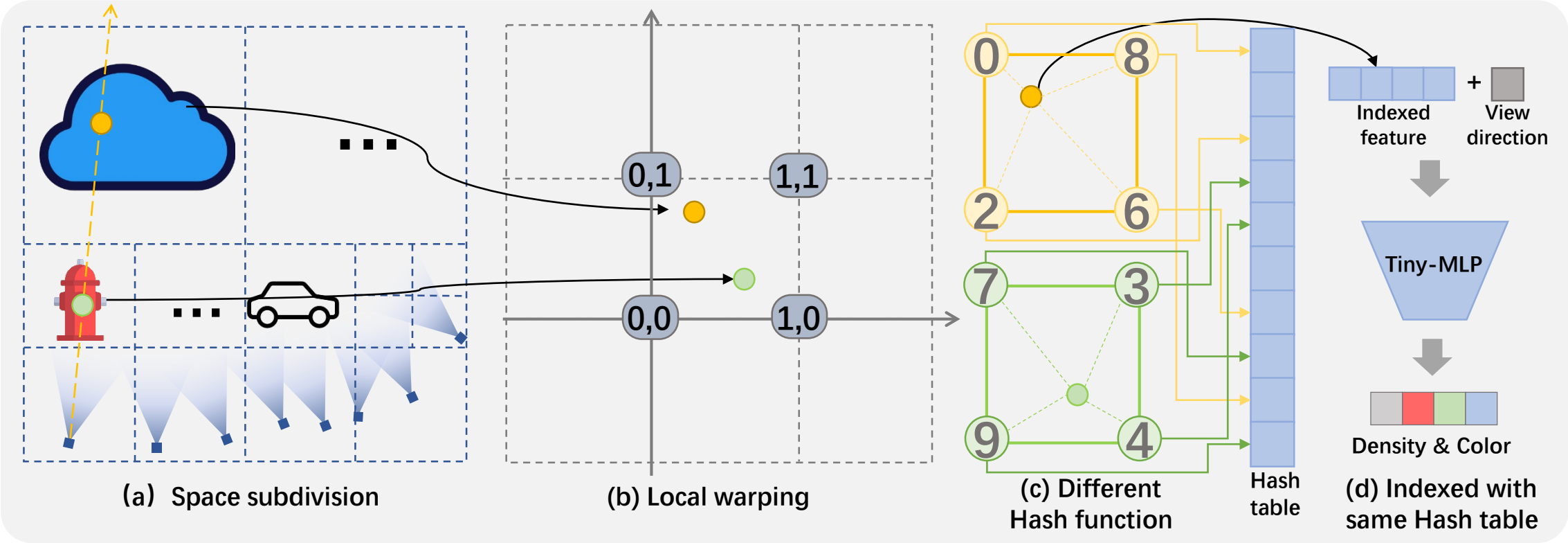
Visualization of perspective warping



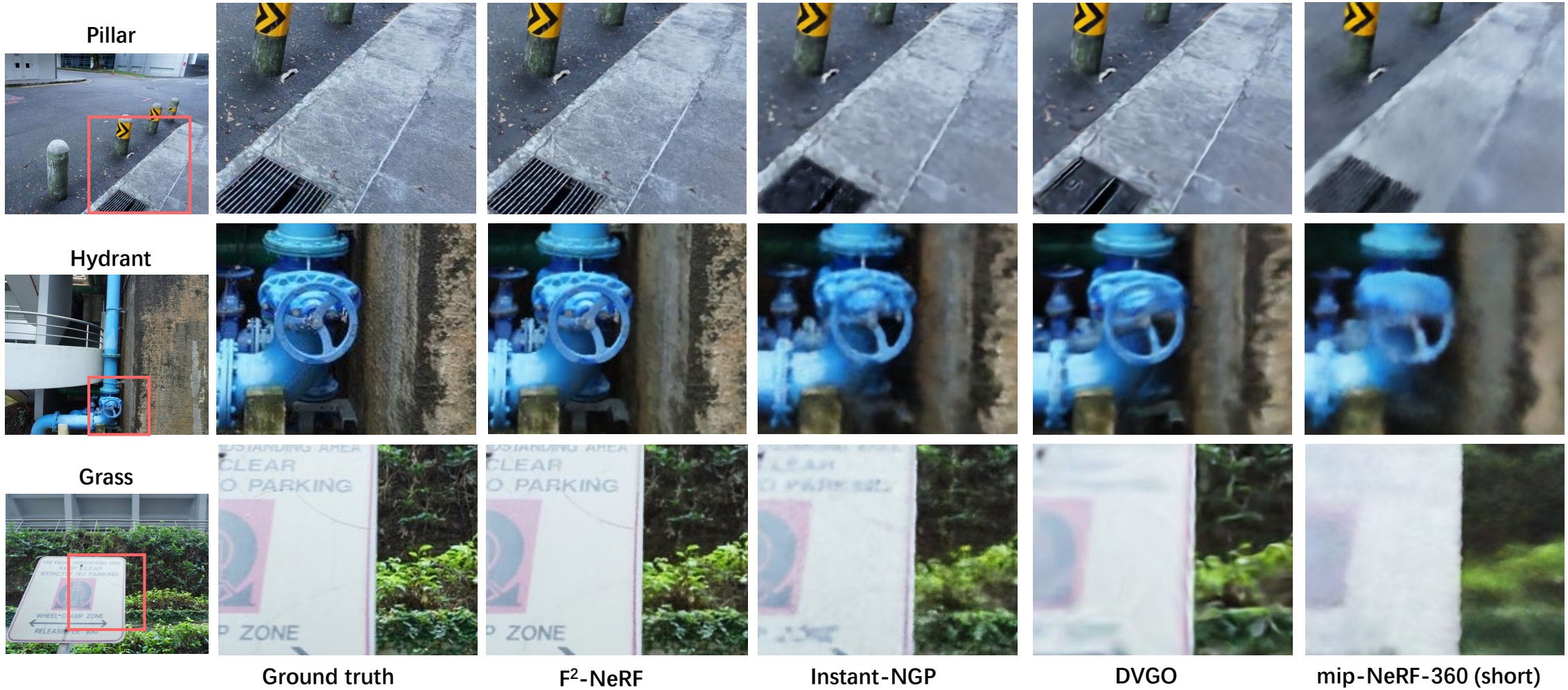
Pipeline



Pipeline


0,1
1,1
0,0
1,0
0
8
2
6
7
3
9
4
Hash table
Indexed feature
+
View direction
Tiny-MLP
Density & Color
(a) Space subdivision
(b) Local warping
(c) Different Hash function
(d) Indexed with same Hash table

Visual comparisons



Hydrant

Mip-NeRF-360 (short)

DVGO

Instant-NGP

F^2-NeRF

Pillar

Mip-NeRF-360 (short)

DVGO

Instant-NGP

F^2-NeRF

Quantitative results

Method	Tr. time	PSNR↑	SSIM↑	LPIPS(VGG)↓
NeRF++ [62]	hours	23.47	0.603	0.499
mip-NeRF-360 [3]	hours	**27.01**	**0.766**	**0.295**
mip-NeRF-360$_{\text{short}}$	30m	22.04	0.537	0.586
Plenoxels [58]	25m	19.13	0.507	0.543
DVGO [39]	21m	23.90	0.651	0.455
Instant-NGP [26]	6m	24.43	0.677	0.413
F^2-NeRF	12m	**26.32**	**0.779**	**0.276**

Table 1. **Results on the Free dataset**. In mip-NeRF-360$_{\text{short}}$, we early stop the training to make them finished in 30 minutes. Training times are evaluated on a 2080ti GPU.

Method	Tr. time	PSNR↑	SSIM↑	LPIPS(VGG)↓
NeRF++ [62]	hours	26.21	0.729	0.348
mip-NeRF-360 [2]	hours	**28.94**	**0.837**	**0.208**
Plenoxels [58]	22m	23.35	0.651	0.471
DVGO [39]	16m	25.42	0.695	0.429
Instant-NGP [26]	6m	26.24	0.716	0.404
F^2-NeRF	14m	**26.39**	**0.746**	**0.361**

Table 2. **Results on the NeRF-360-V2 dataset.**

Method	Tr. time	PSNR↑	SSIM↑	LPIPS(VGG)↓
NeRF [25]	hours	26.50	0.811	0.250
mip-NeRF [2]	hours	**26.60**	**0.814**	**0.246**
Plenoxels [58]	17m	26.29	0.839	0.210
DVGO [39]	11m	26.34	0.838	0.197
TensoRF [6]	48m	**26.73**	0.839	0.204
Instant-NGP [26]	6m	25.09	0.758	0.267
F^2-NeRF	13m	26.54	**0.844**	**0.189**

Table 3. **Results on the LLFF dataset.**

Conclusion

- A complete pipeline for novel view synthesis with arbitrary camera trajectories
- An adaptive warping method for space compression of general scenes
- Fast to train with anchored neural hash grids

Limitations:

- Discontinuity on borders of sub-regions
- Aliasing
- Extremely large (>1k images) training data?

Aliasing

Thanks

Project page: http://totoro97.github.io/projects/f2-nerf