Neural Rendering and Secondary Cues: Learning Hidden Neural Radiance Fields

Kushagra Tiwary

PhD Student, Camera Culture, MIT Media Lab

Dream: Can we figure out *everything* that makes up this room?

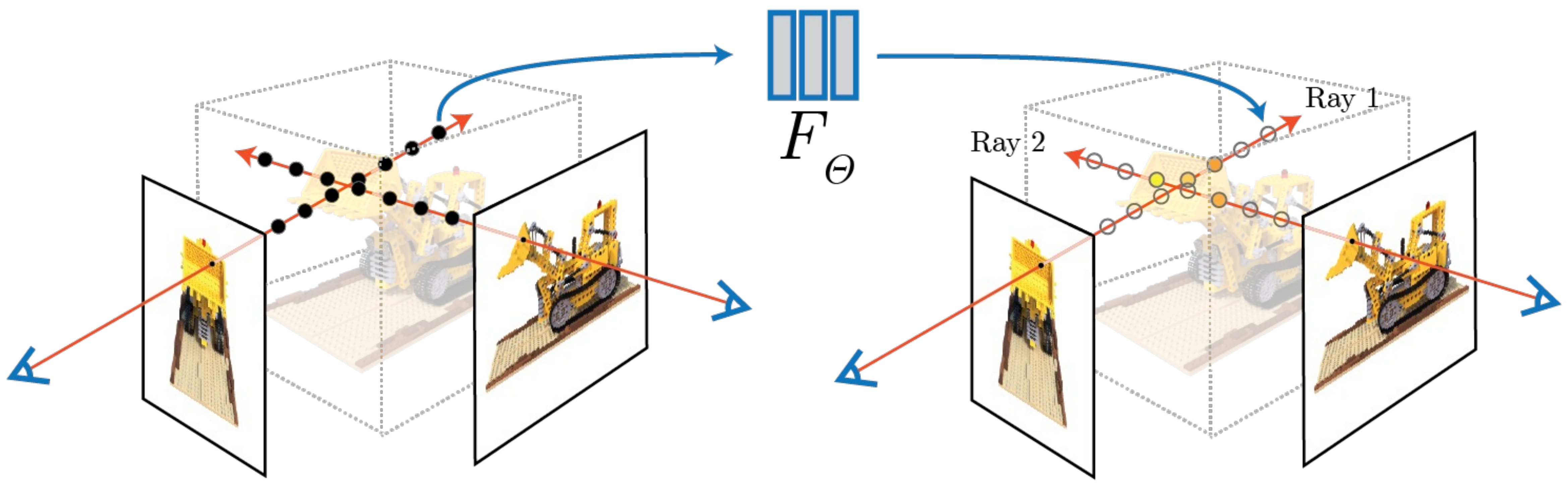
"Enhance (34, 46)": *Infinite zoom, Super resolution, Virtual Camera from Mirror's Perspective*

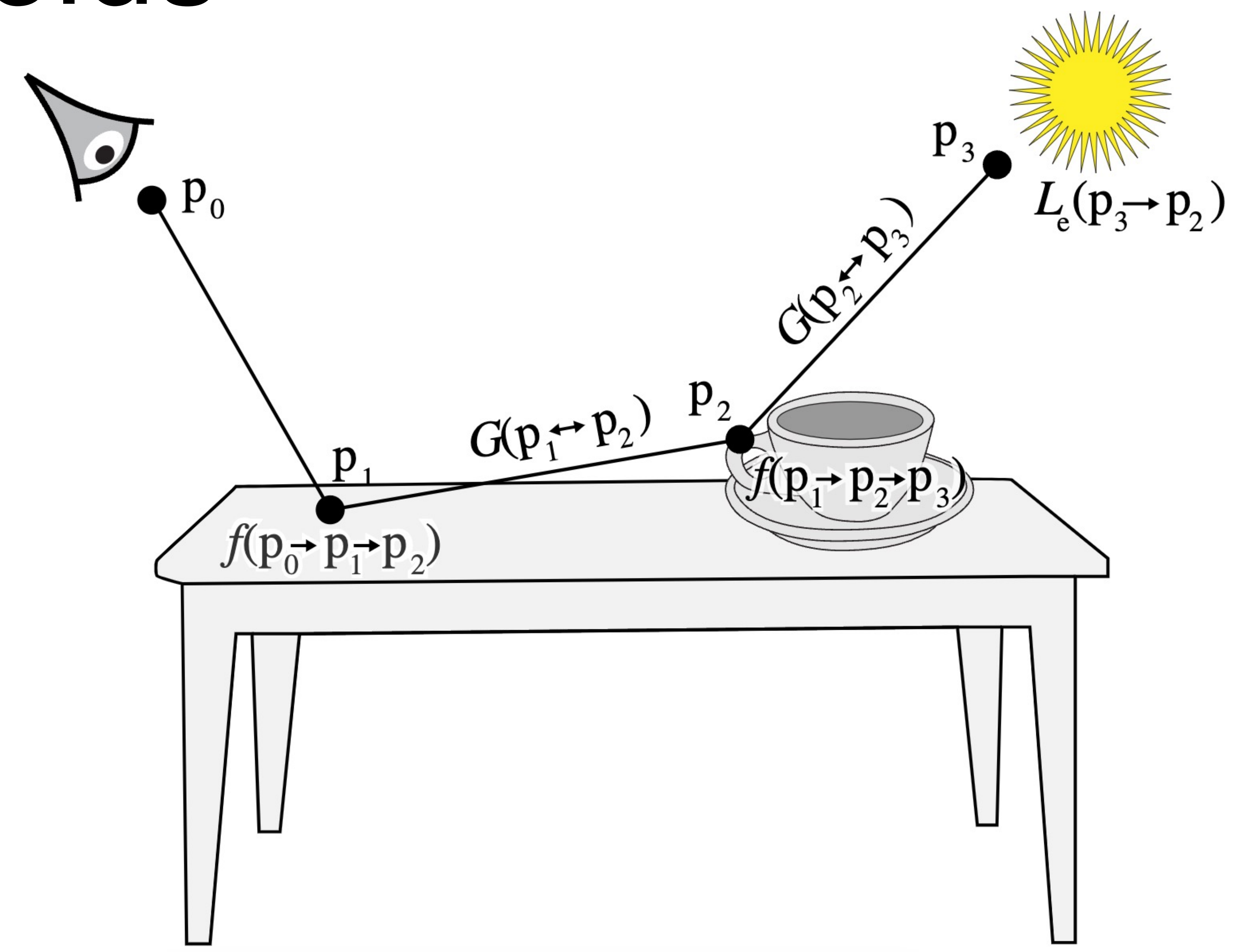
"Go right": *Parallax, Perspective Change, Occlusion-Aware*

"Enhance (57, 19)": *Perspective Change, Parallax, Occlusion-Aware, Super Resolution*

Making Esper Possible..

Insight: Modelling complex light transport enables learning of *hidden* neural fields





Neural Radiance Fields:

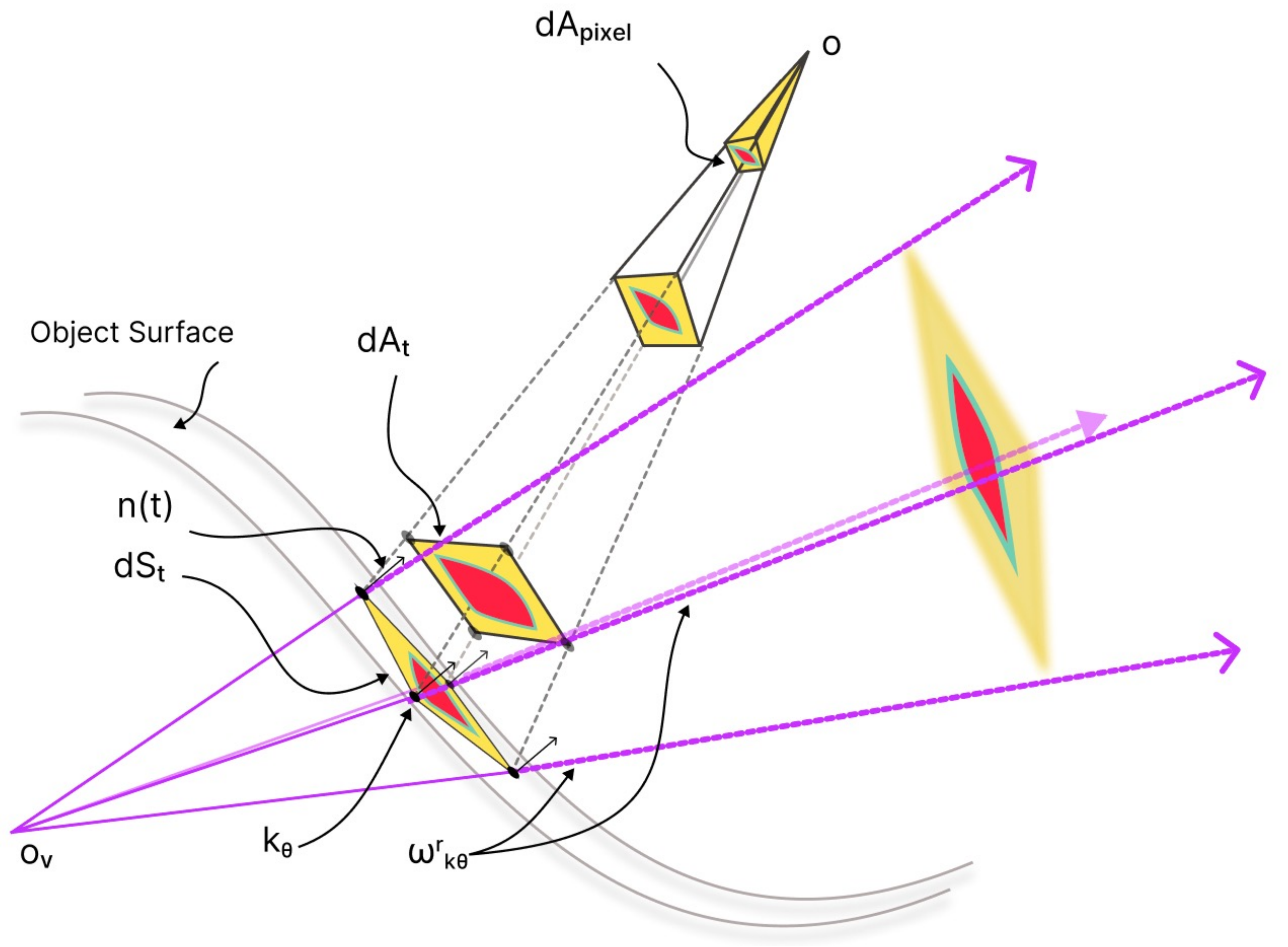
Enables learning a 5D world from pixel data

Light Transport:

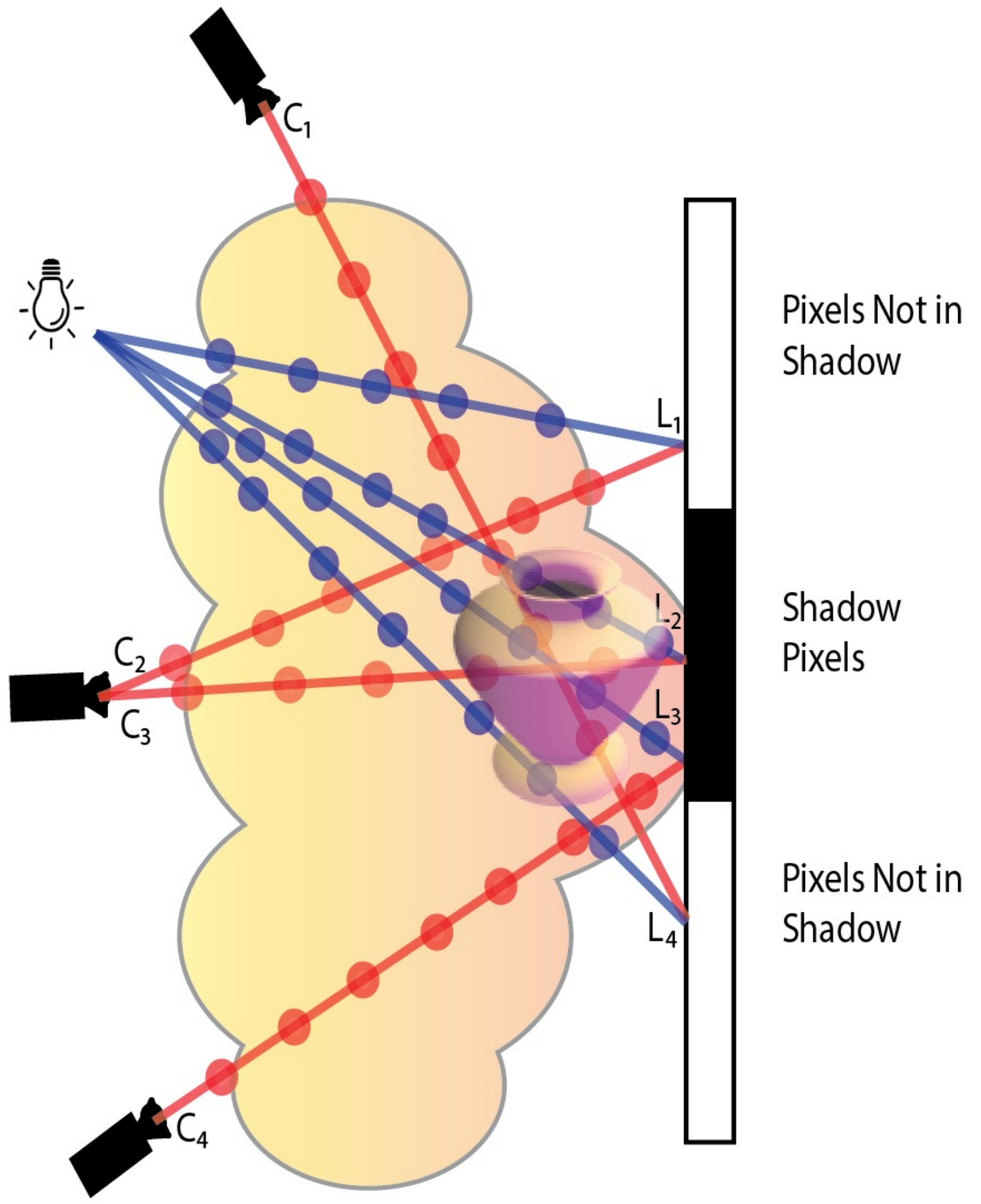
Models the distribution of radiance in a scene, enables inference of properties through secondary cues

NeRF (original paper)
PBR (book): Physically-based rendering

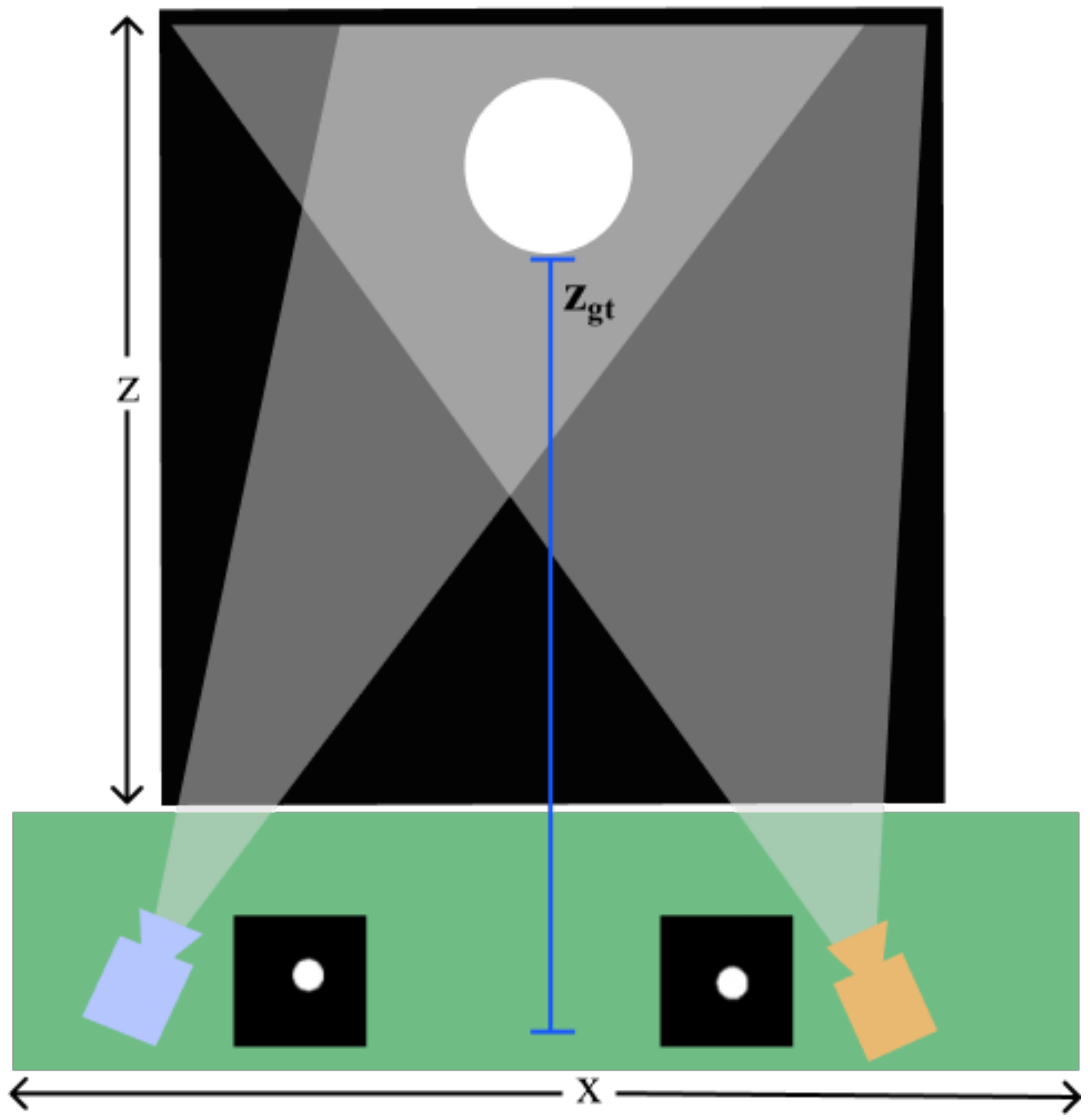
Secondary Cues: Reflections, Shadows, Triangulation



Reflections



Shadows

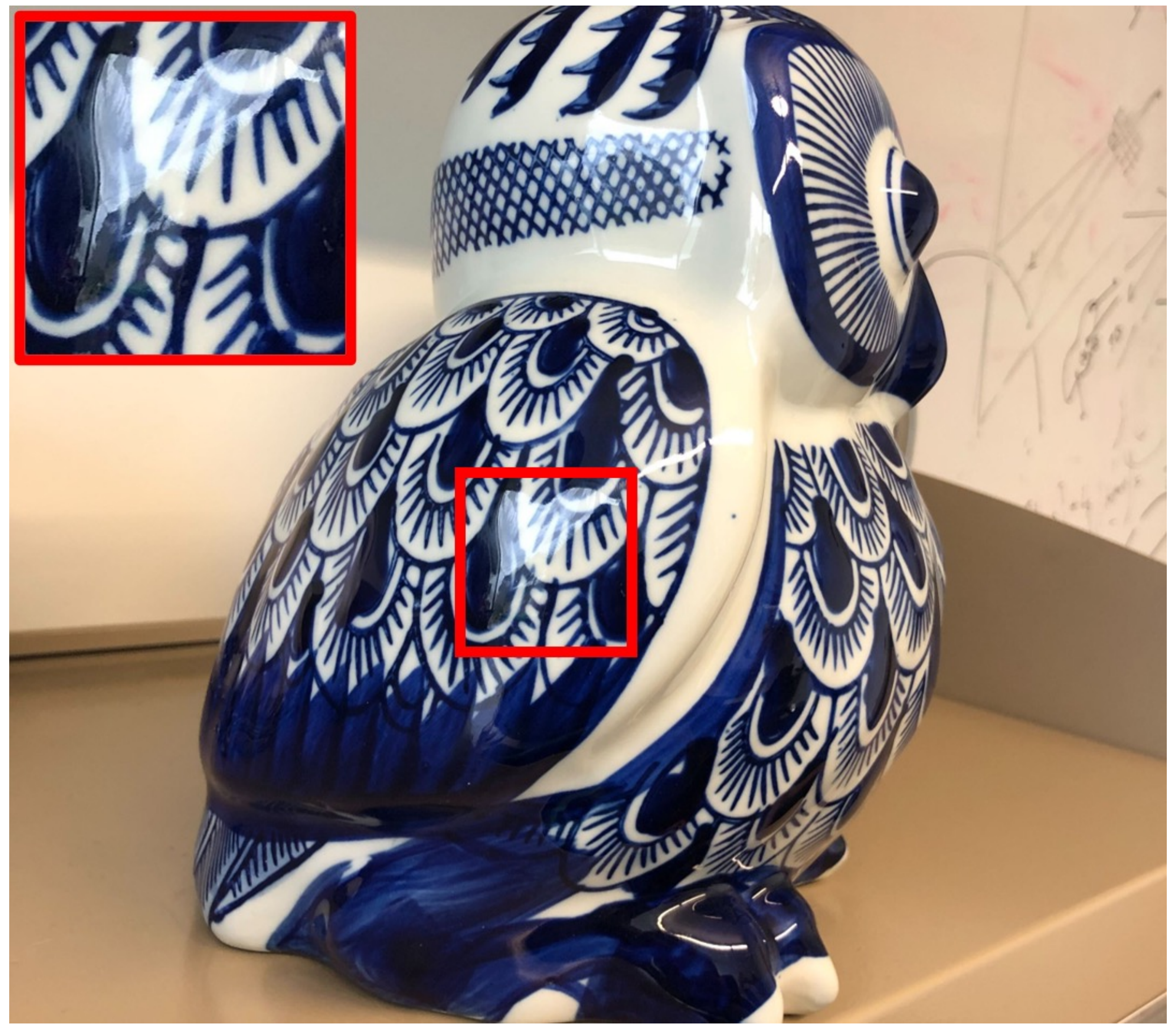


Triangulation

Exploiting reflections is challenging

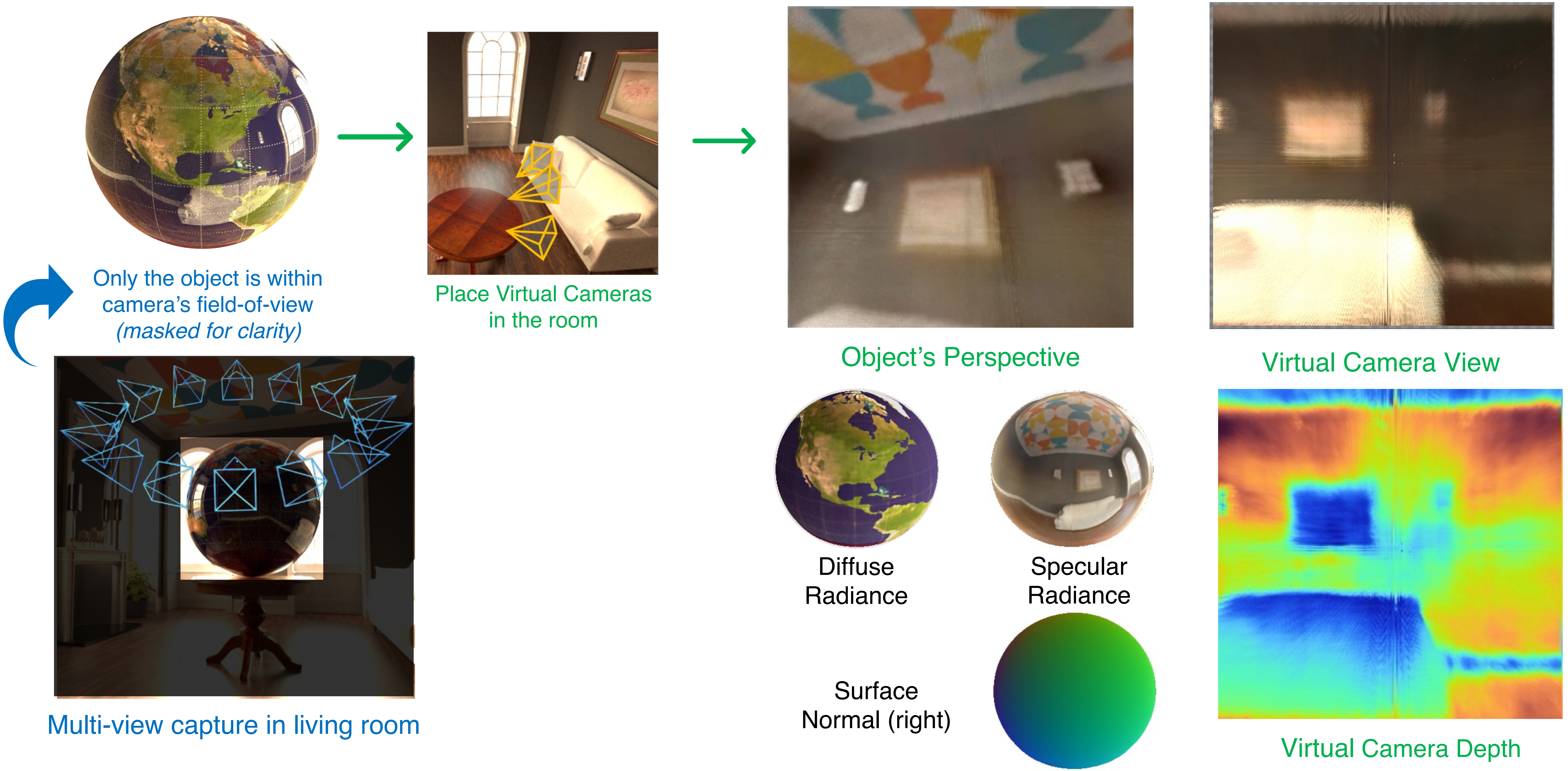
Distorted by reflector's geometry

Mixed with reflector's texture

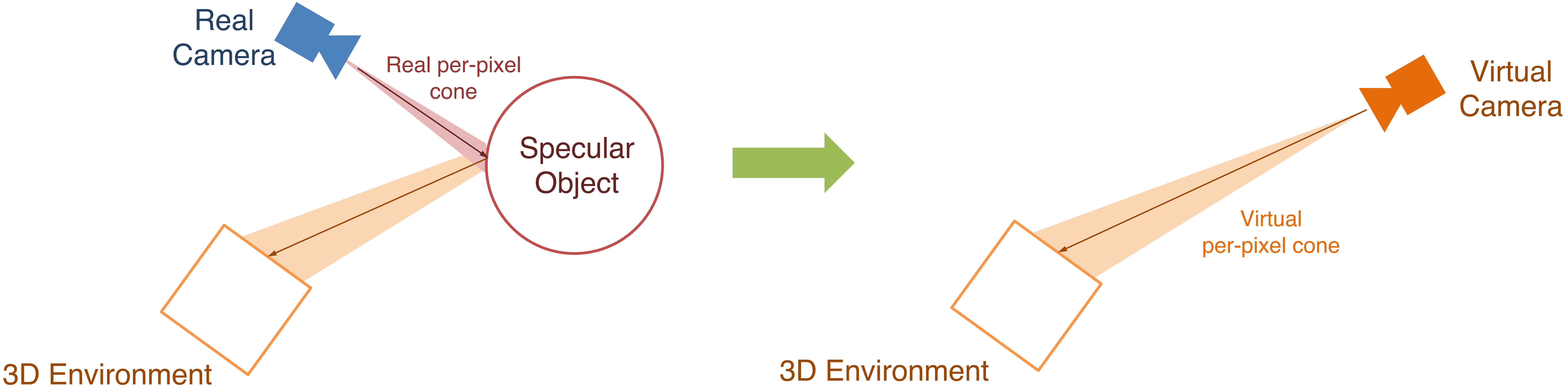


2D projection of the 3D environment

ORCa: Turning Objects into Radiance-field Cameras



Reflections can be modelled as radiance fields captured by virtual camera



ORCa recovers fine environment details

Sampling 2D Environment Map*

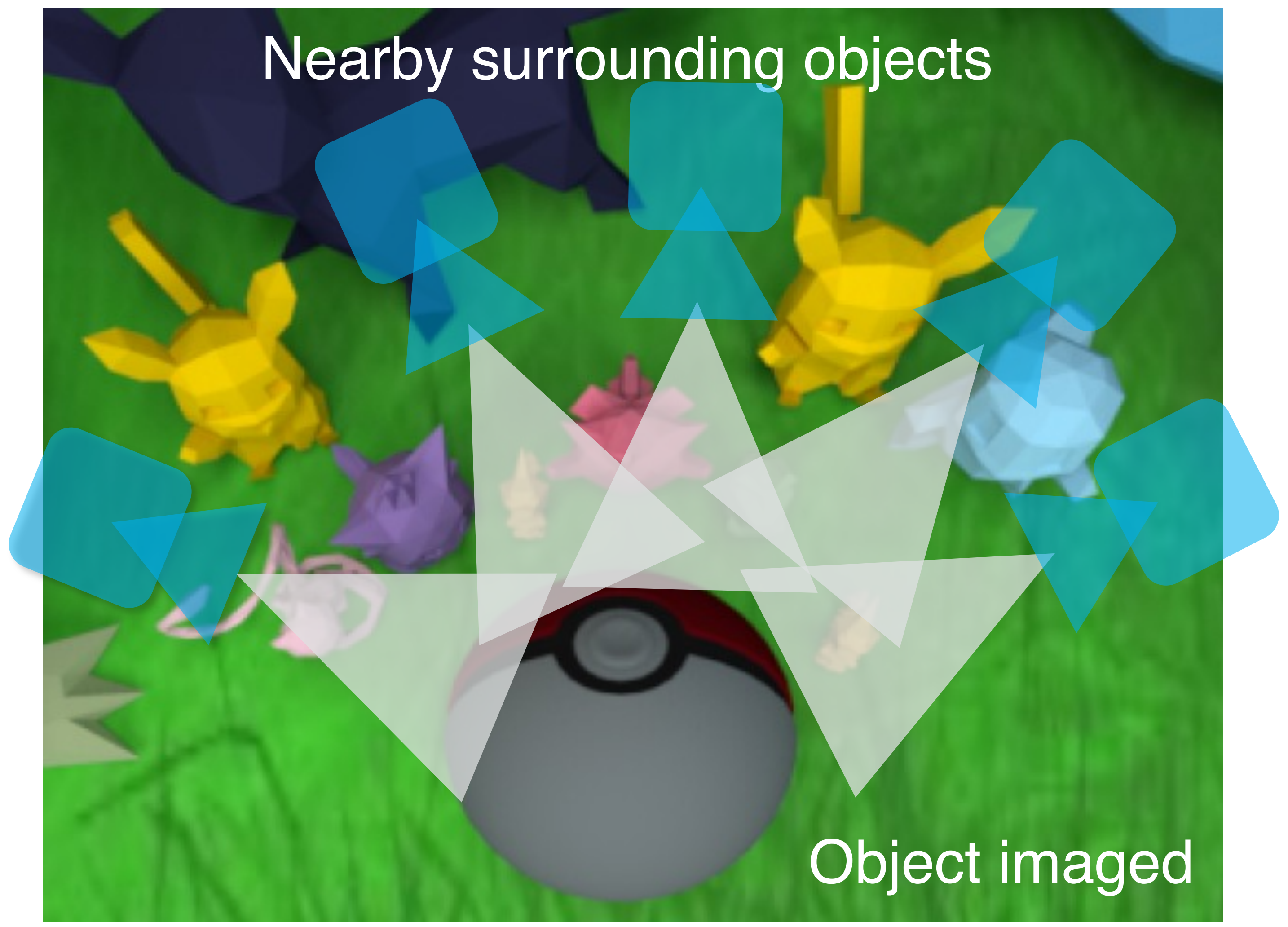
Sampling 5D Environment Radiance Field

Environment Radiance Fields enable Virtual View synthesis for viewpoints that are beyond field-of-view of the original camera

*Dave et. al, Pandora (RGB only)

Advantages of Environment Radiance Fields

Reconstruction of Captured Images *(masked for clarity)*



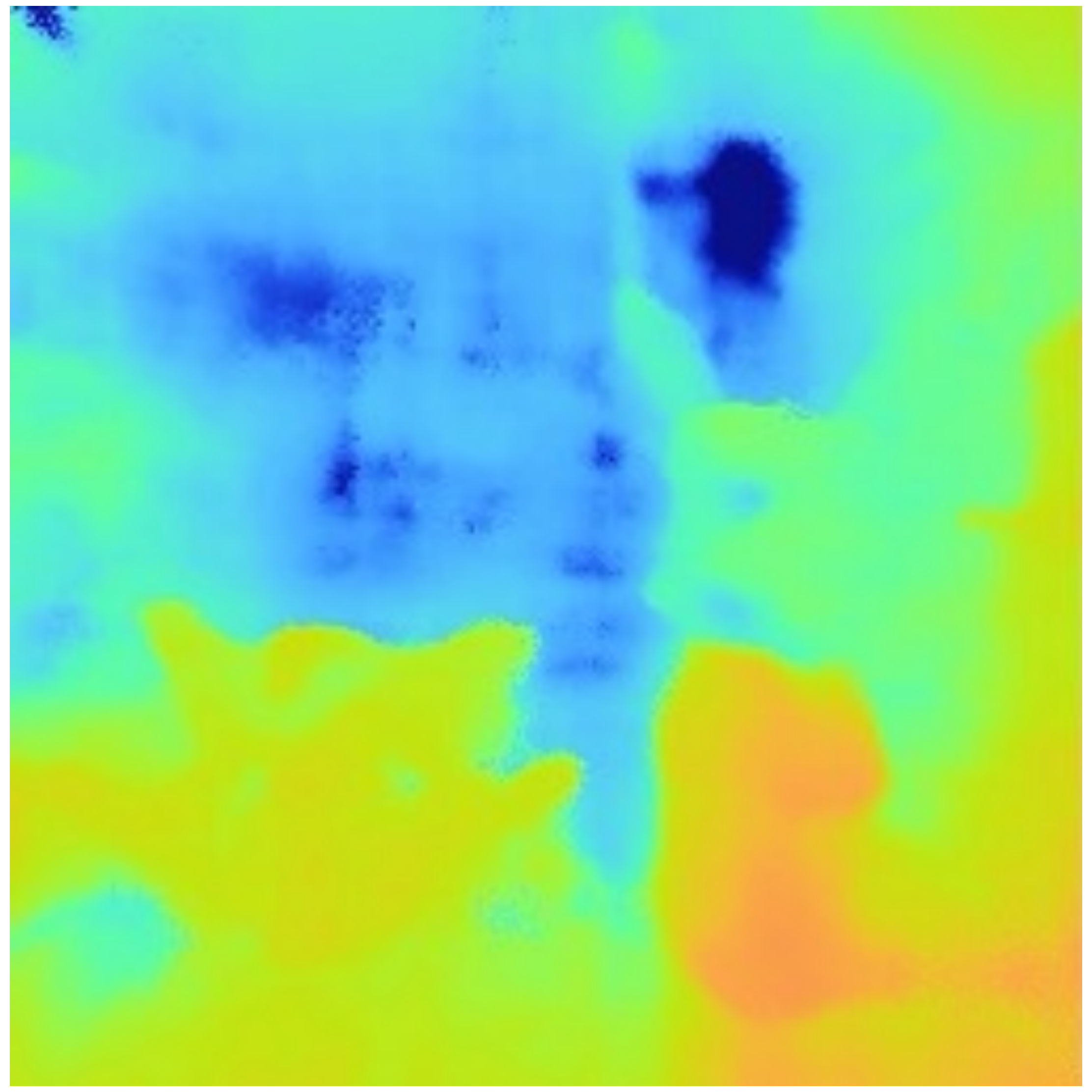
Scene

2D Environment Map*

Cannot model parallax or depth

5D Environment Radiance Field

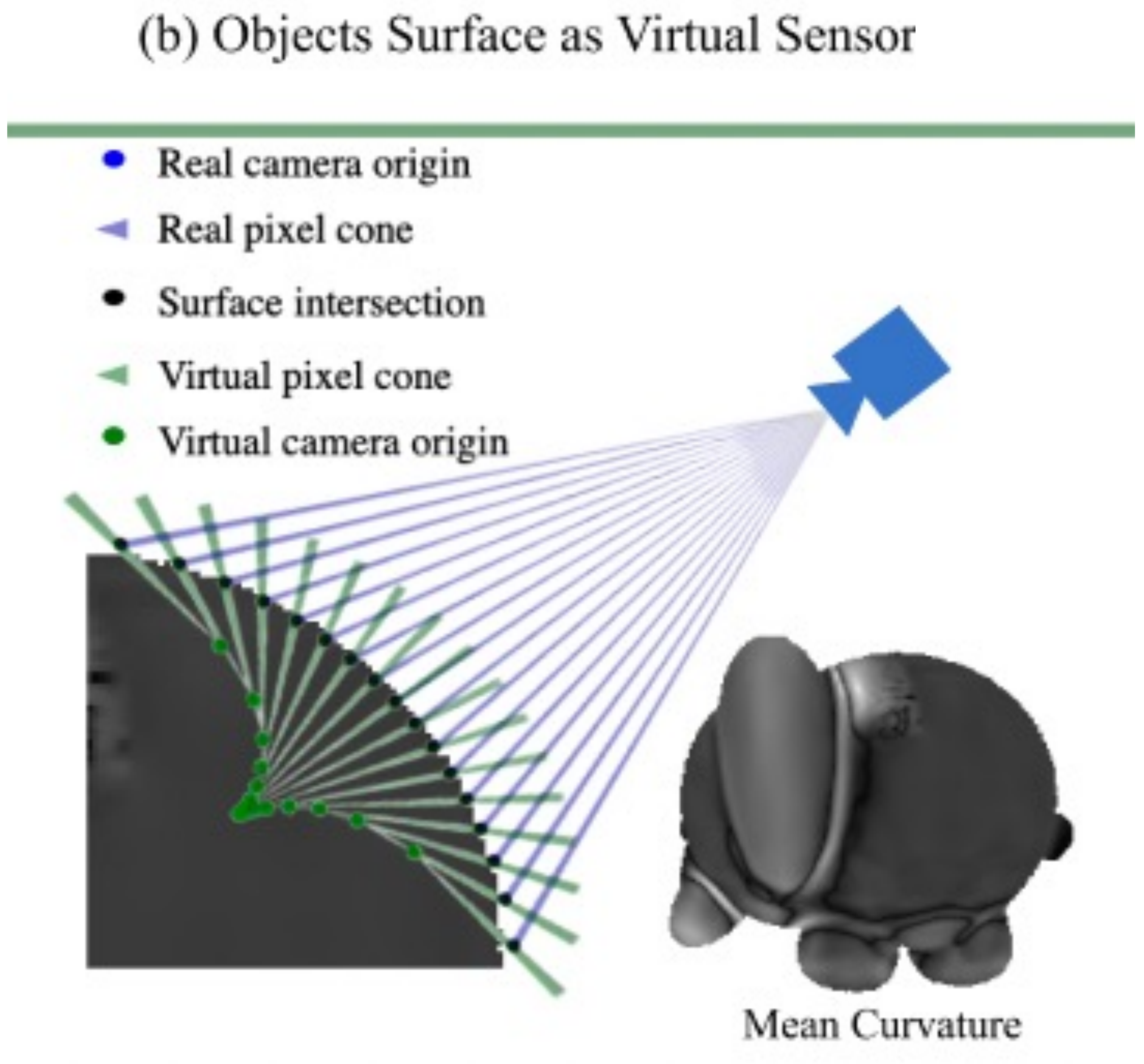
Parallax effect in translated views



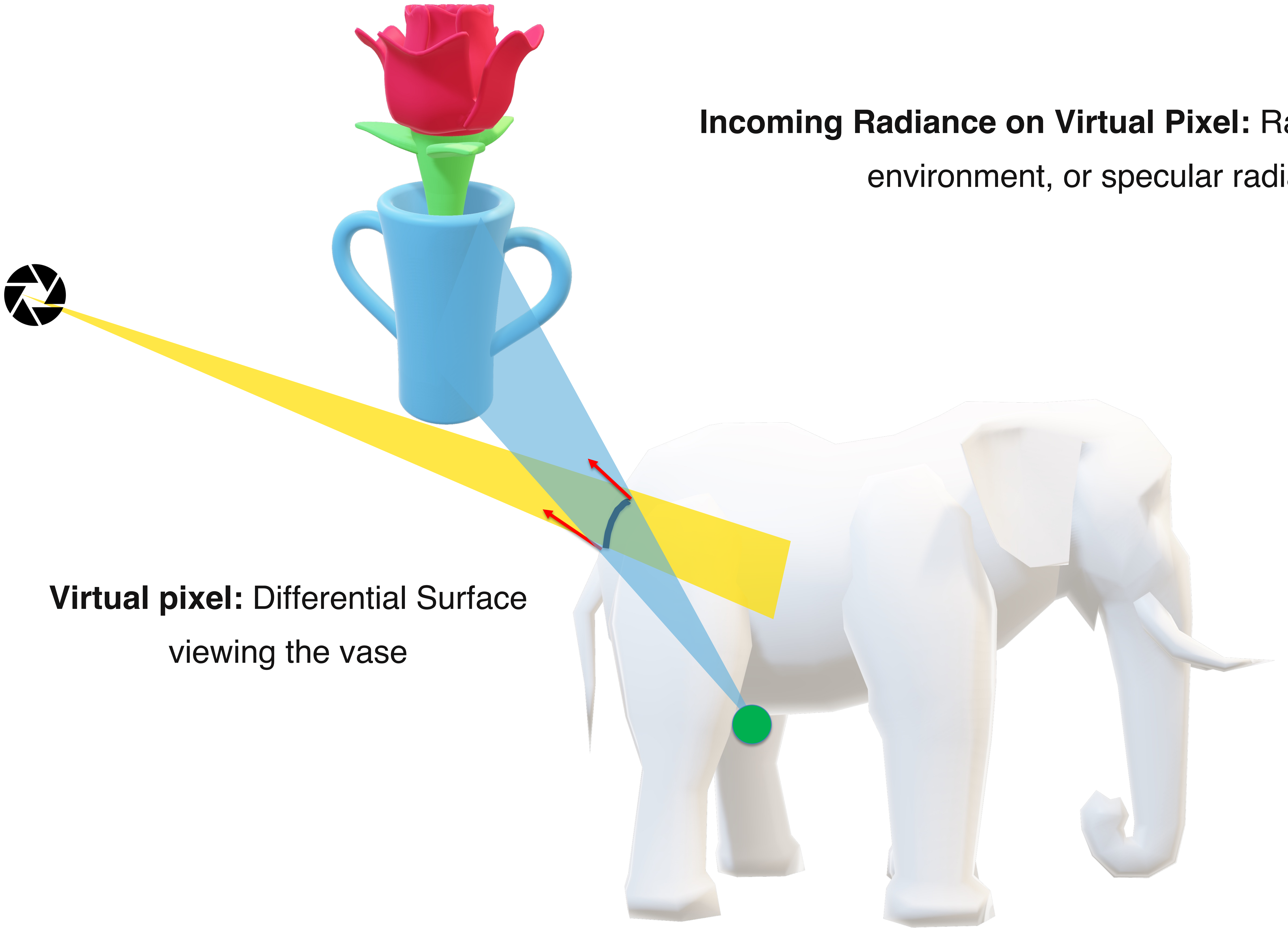
Depth map of the environment

*Dave et. al, Pandora (RGB only)

ORCa: Three step approach



Virtual Pixel Radiance depend on Pixel, Size, Local Geometry and Camera Pose

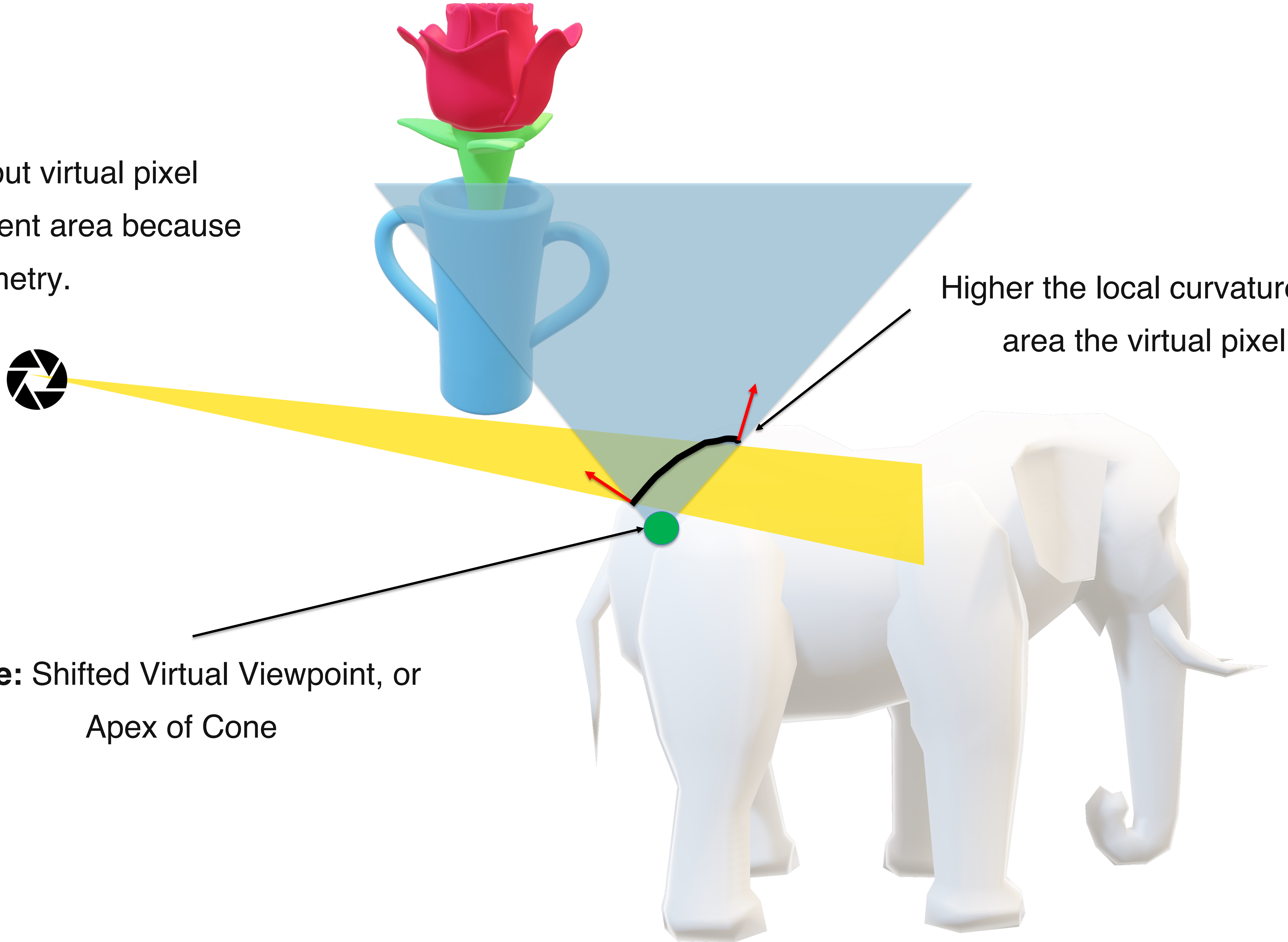


Incoming Radiance on Virtual Pixel: Radiance from the environment, or specular radiance

Virtual pixel: Differential Surface viewing the vase

Virtual Pixel Radiance depend on Real-Pixel Size, Local Geometry and Camera Pose

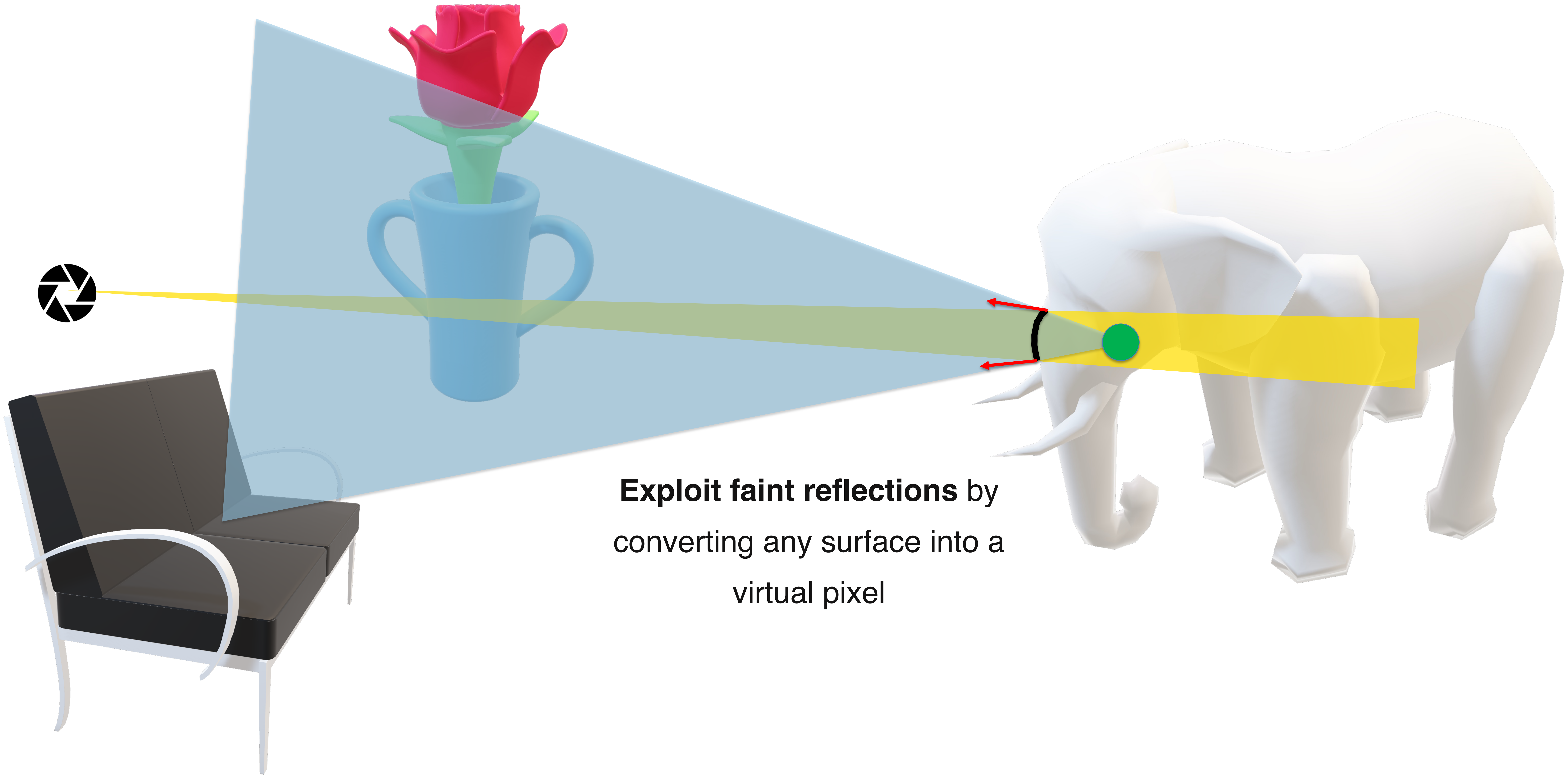
Same camera pose, but virtual pixel views a completely different area because of local geometry.



Higher the local curvature, the more area the virtual pixel views!

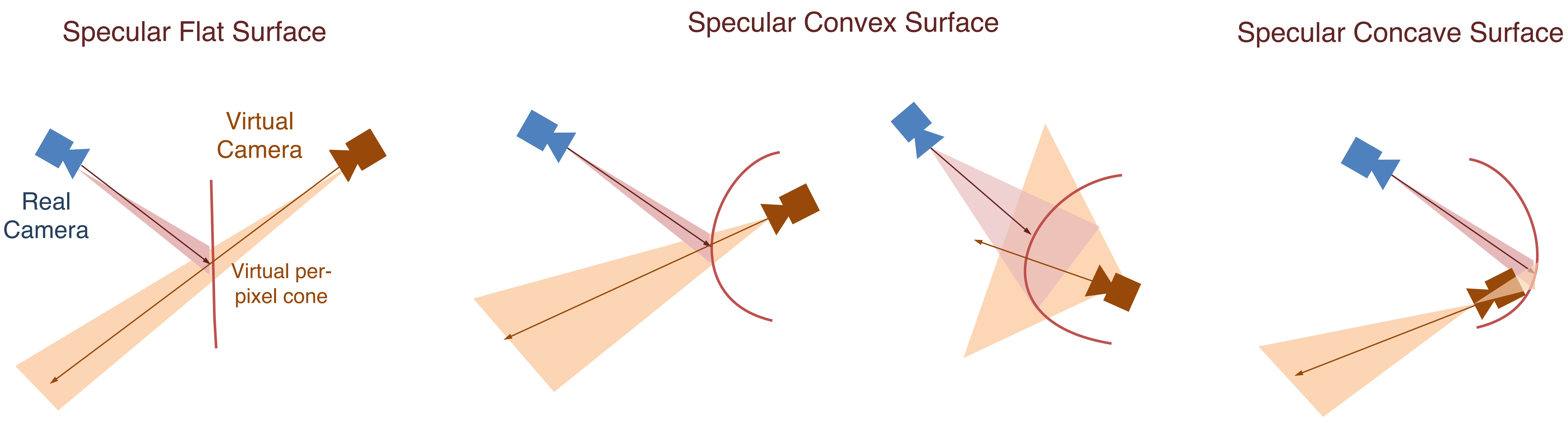
Note: Shifted Virtual Viewpoint, or Apex of Cone

Convert any surface into a Virtual Pixel if you know local surface geometry



Exploit faint reflections by converting any surface into a virtual pixel

Summary: Virtual Pixel Radiance depend on Pixel, Size, Local Geometry and Camera Pose



- Virtual cone same size as real cone

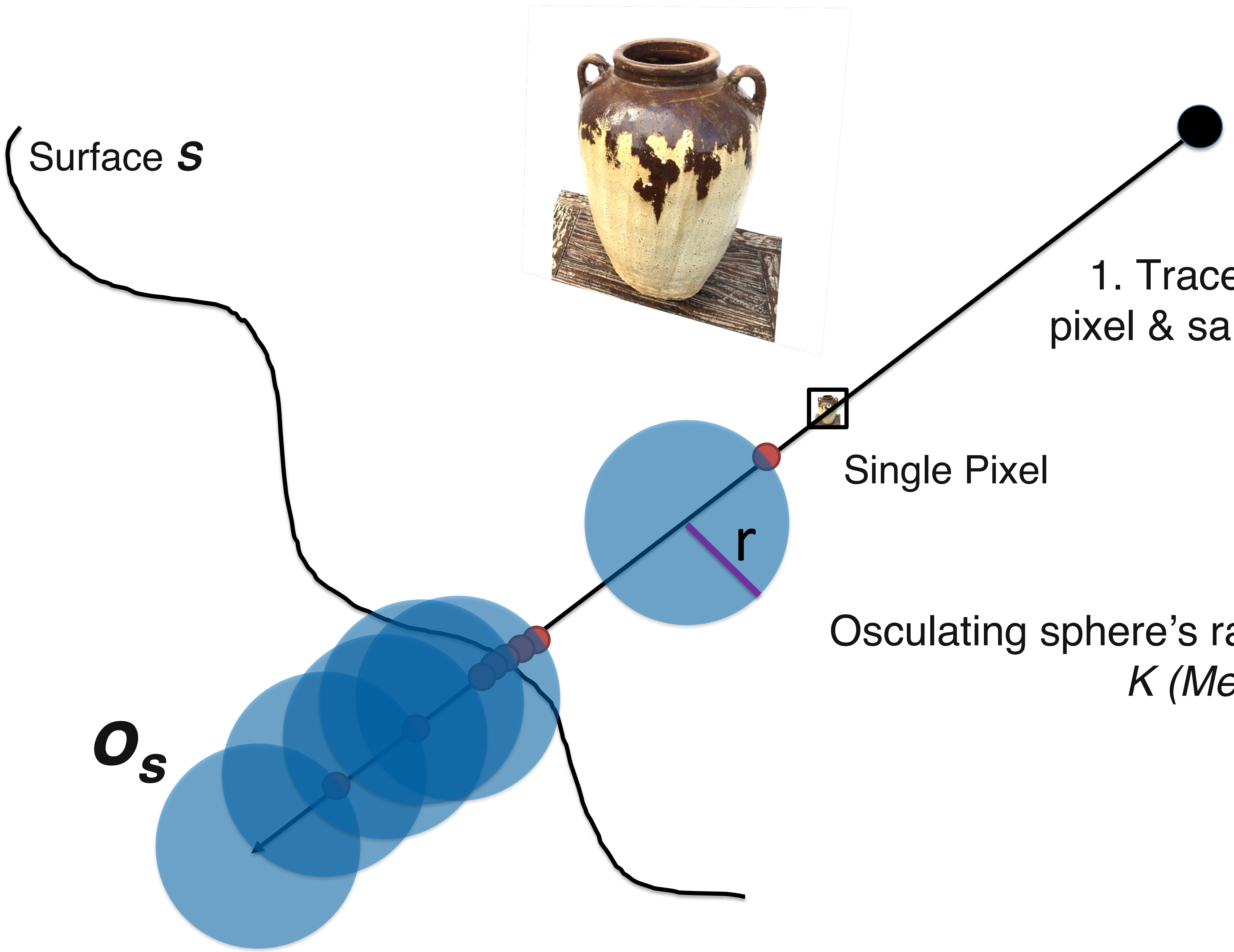
- Low Curvature samples smaller area
- Virtual viewpoint further from surface

- High Curvature samples larger area
- Virtual viewpoint closer to surface

- Virtual camera outside the surface

Object Surface as Virtual Sensors & Pixels

2. For each point (p_i) on the ray, estimate local geometry by fitting an osculating sphere to p_i

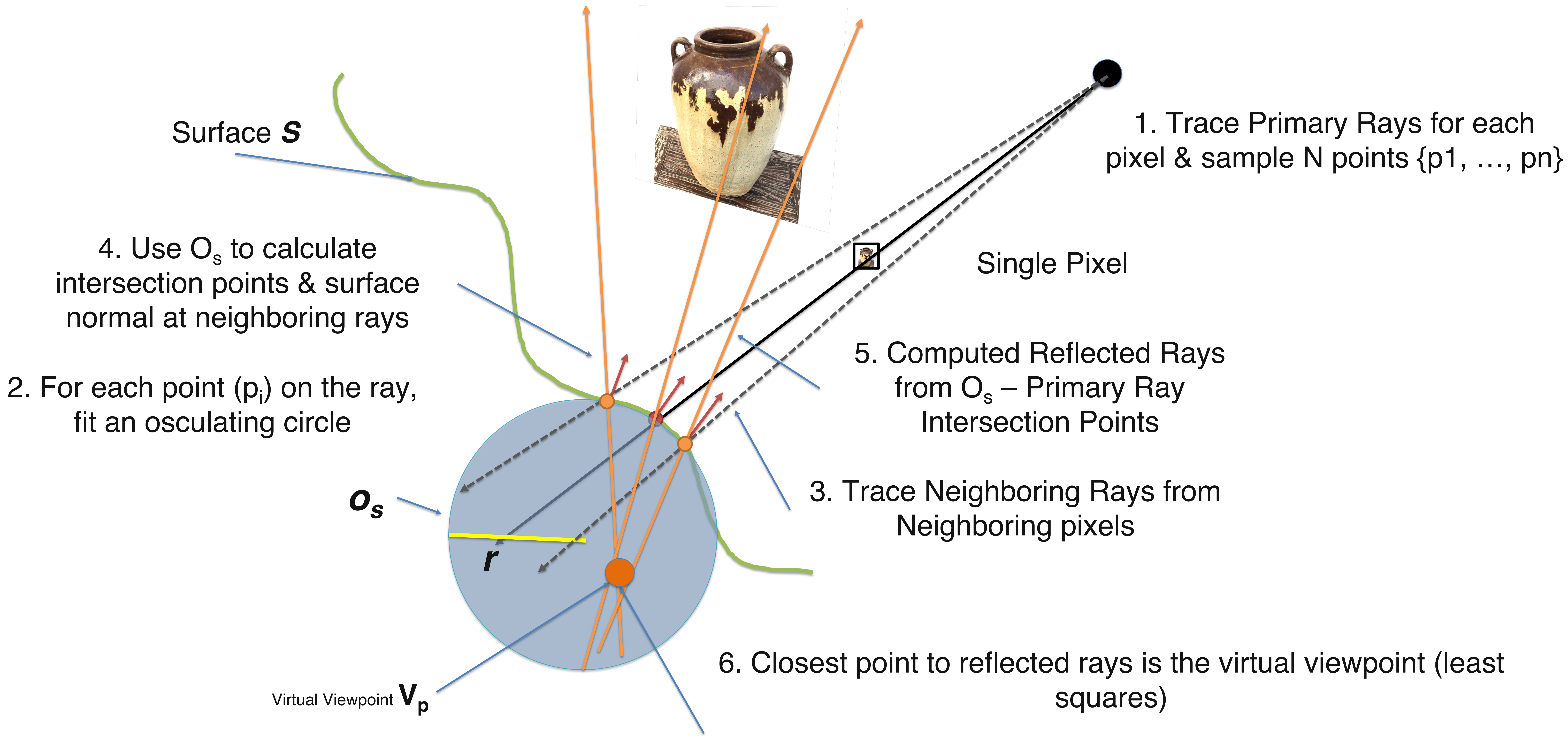


1. Trace Primary Rays for each pixel & sample N points {p1, …, pn}

Osculating sphere’s radius is related to the mean curvature!
K (Mean Curvature) = 2/∇n*

[* Exploring Differential Geometry in Neural Implicits, Novello et. Al]

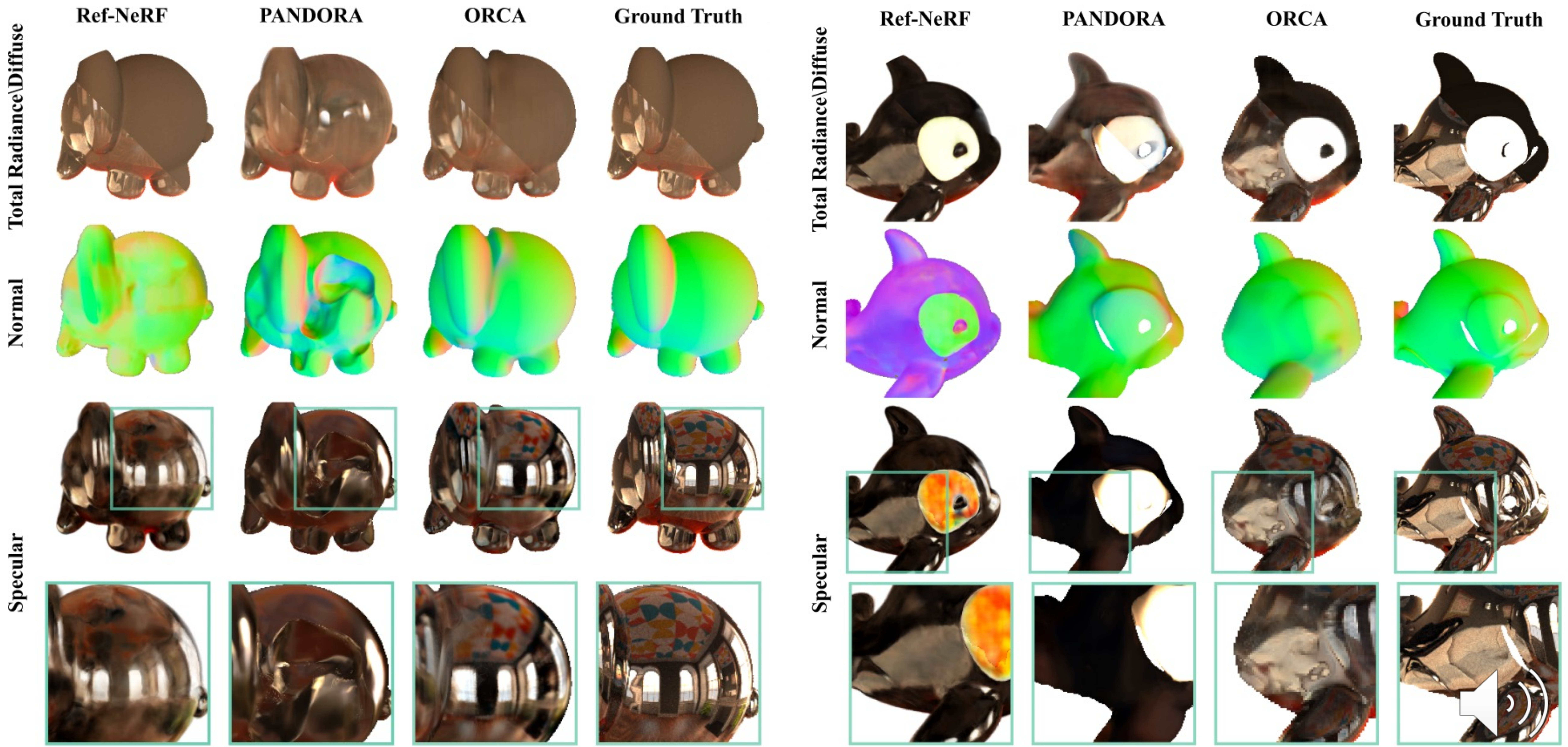
Consider one such osculating sphere…



(this heuristic holds true in the limit and approximates the true virtual viewpoint as dS -> 0)

Virtual Sensors Sample using Virtual Cones…

Accurate diffuse-specular separation and smoother geometry



Tradeoffs in Resolution

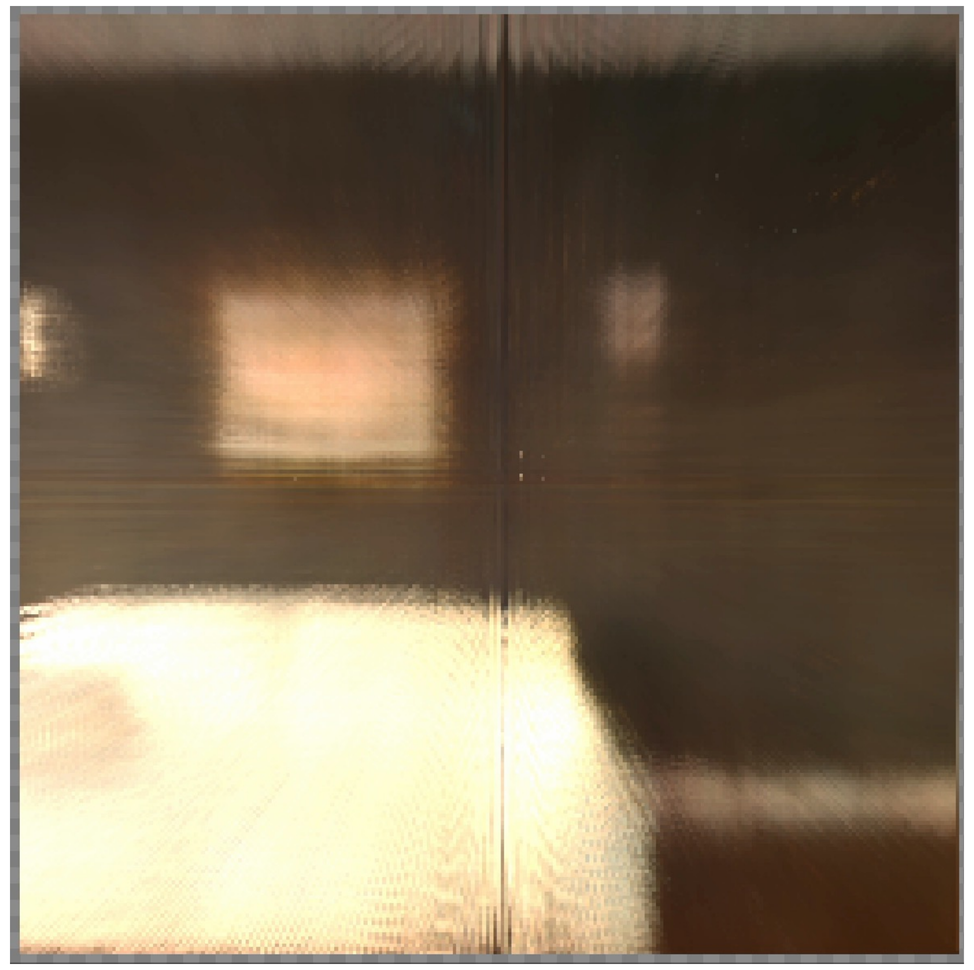
Multi-view Images of a 35cm cup in 10m wide hallway

Extracted *Hidden Radiance Field* of the environment

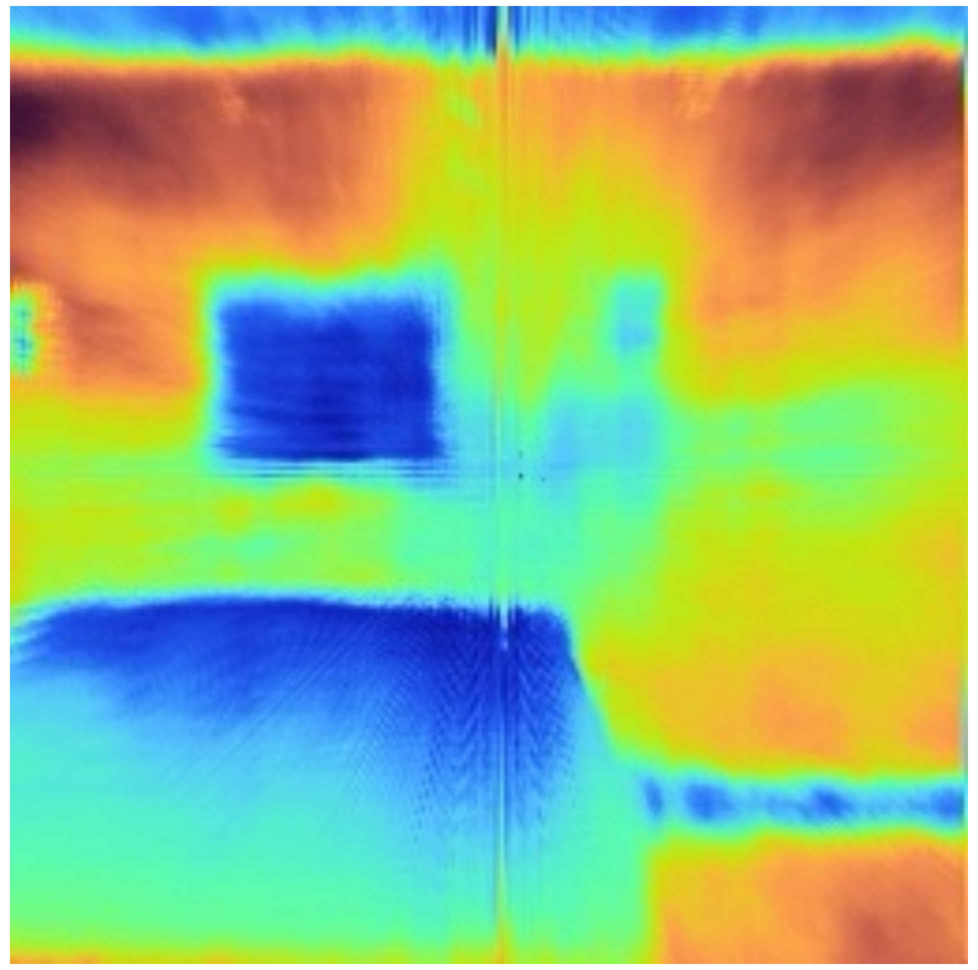
Data Priors, Environment Priors could fix this!

Summary: Hidden Radiance Field Cameras enable *finer recovery*, *parallax*, and *depth estimation*

Object's Perspective

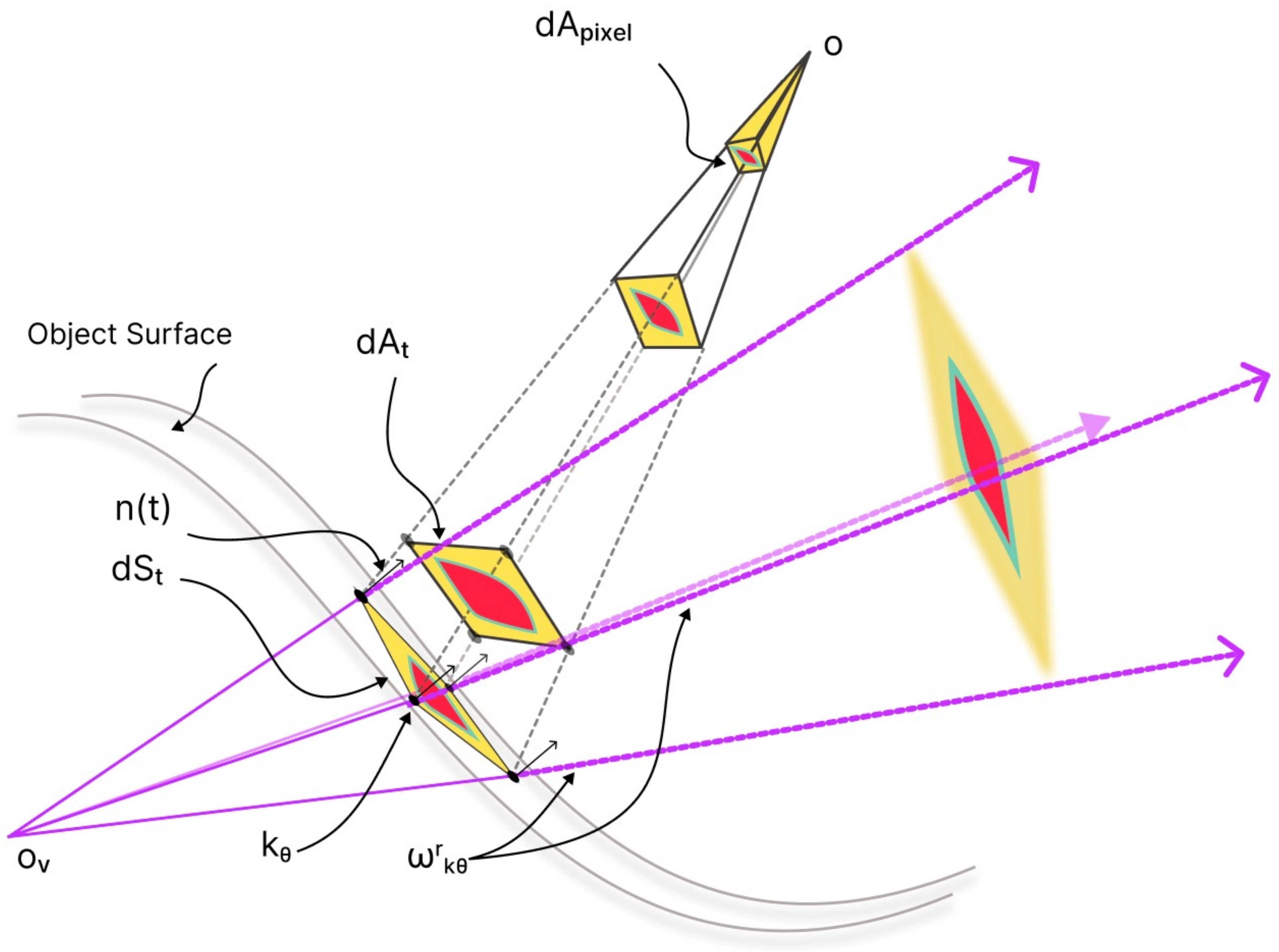


Virtual Camera

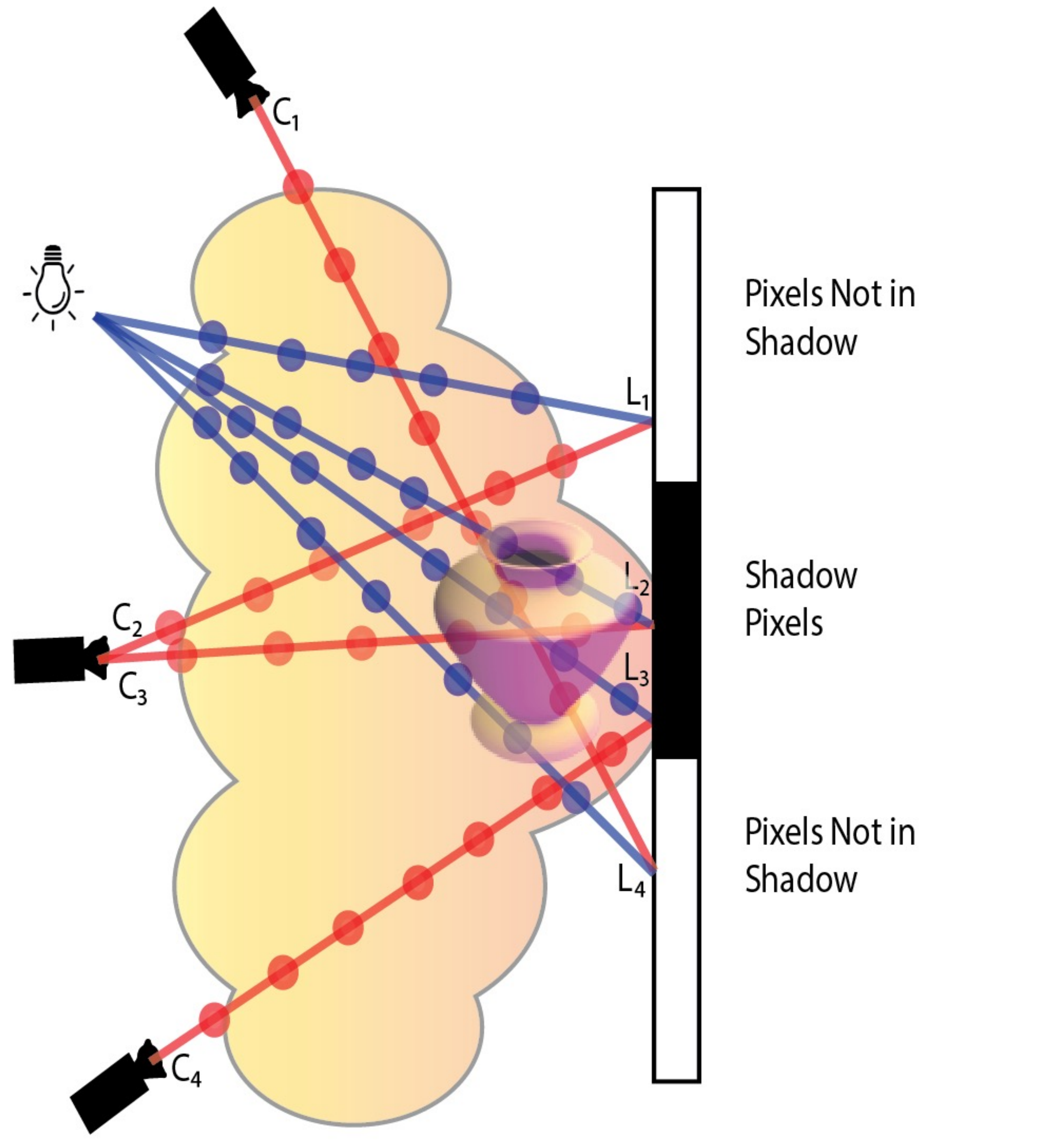


Virtual Depth

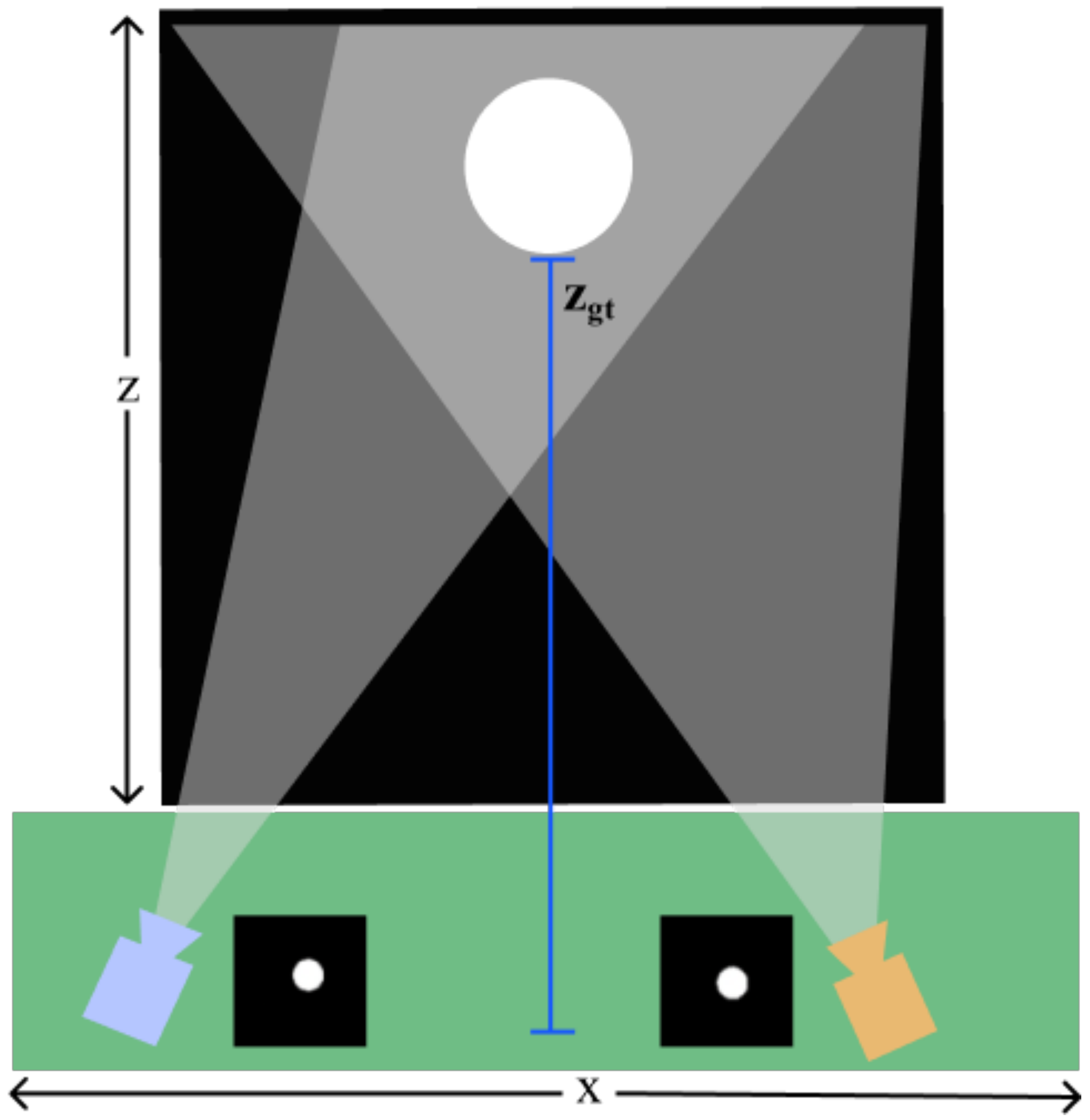
Secondary Cues: Reflections, Shadows, Triangulation



Reflections

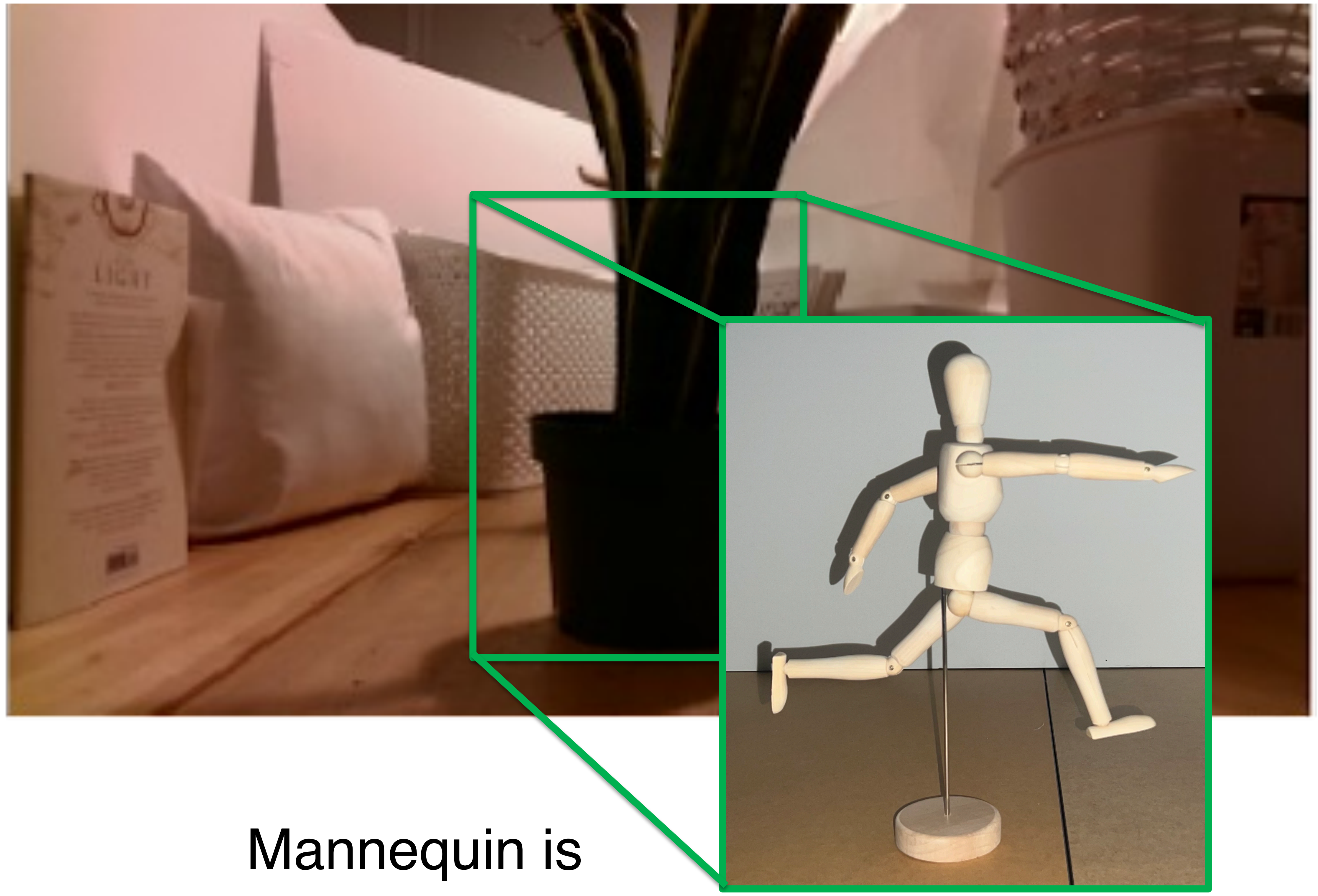


Shadows

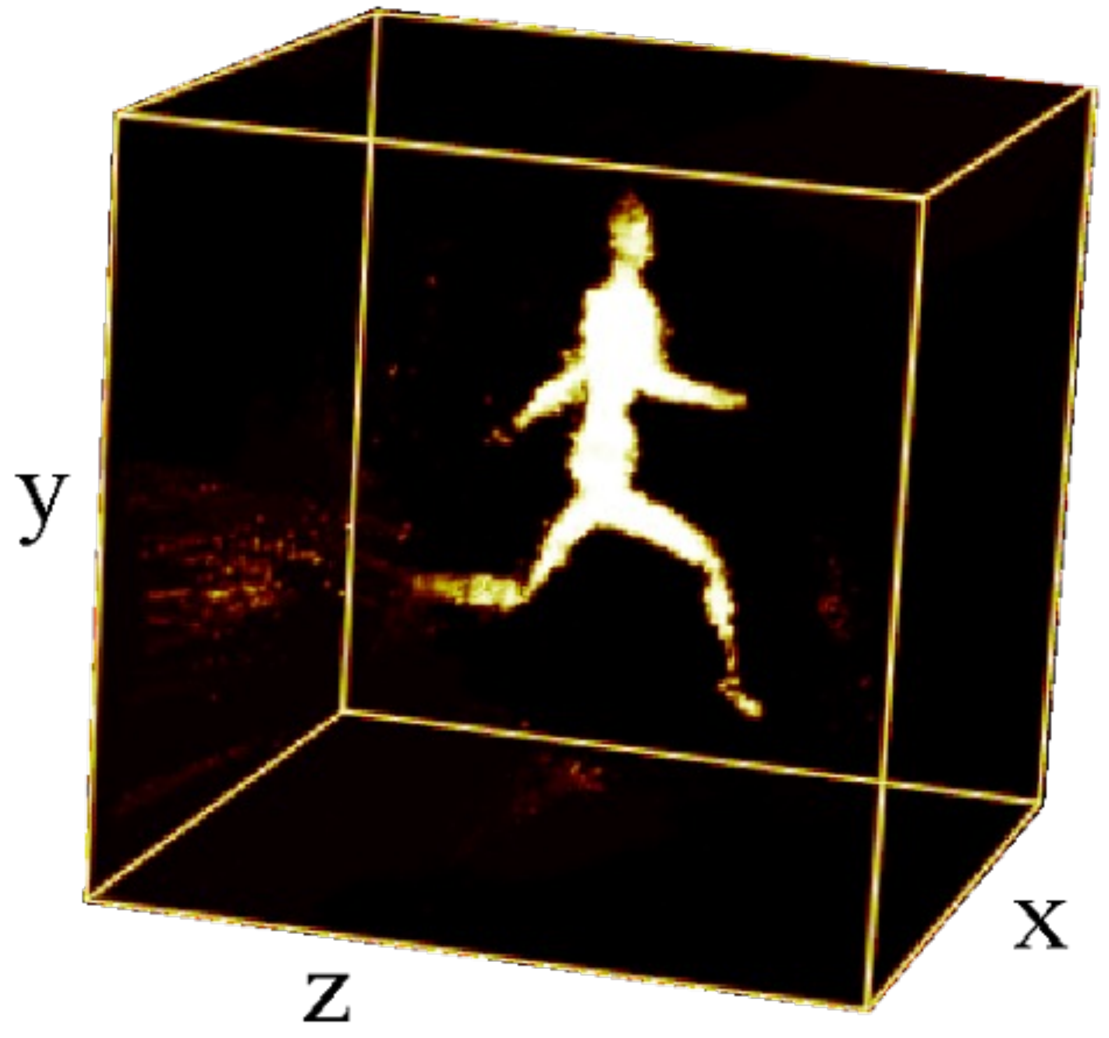


Triangulation

Imaging Behind Occluders using Shadows



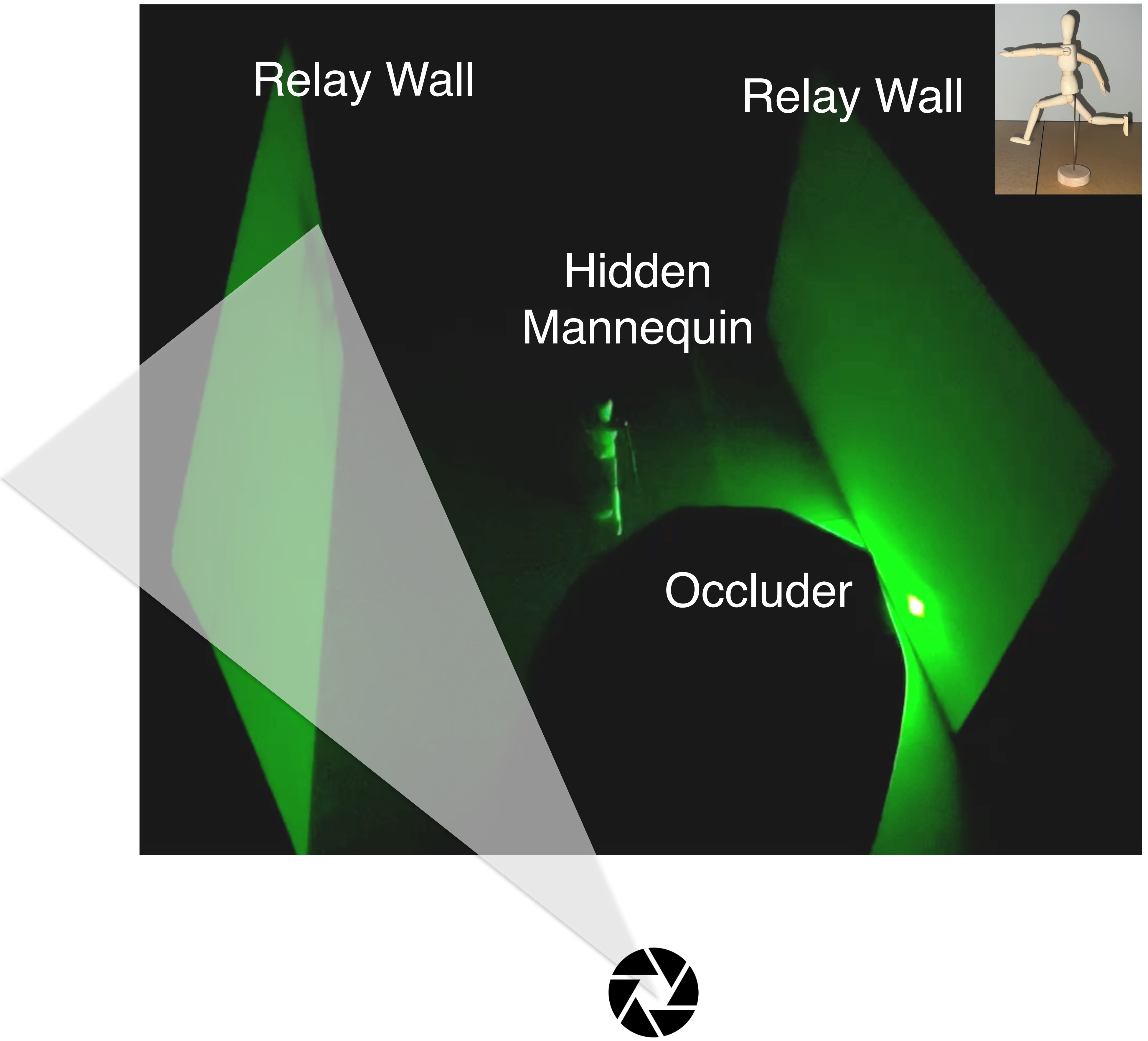
Mannequin is occluded



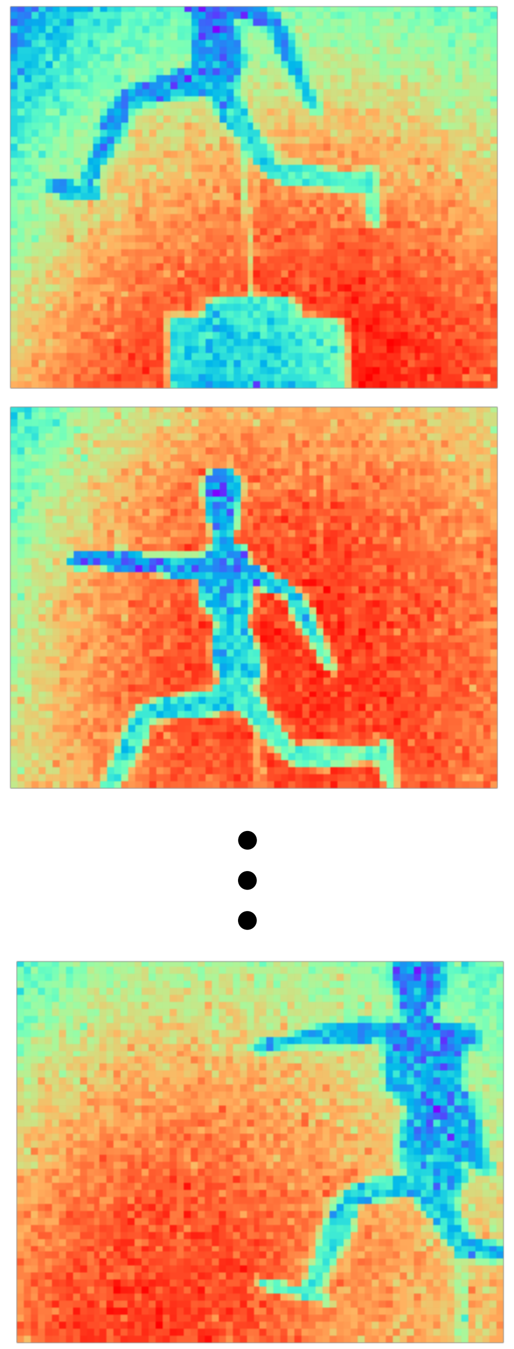
3D Reconstruction
Of Hidden Mannequin

C. Henley et al., ***"Imaging behind occluders using two-bounce light"***, ECCV 2020

Imaging Behind Occluders Pipeline

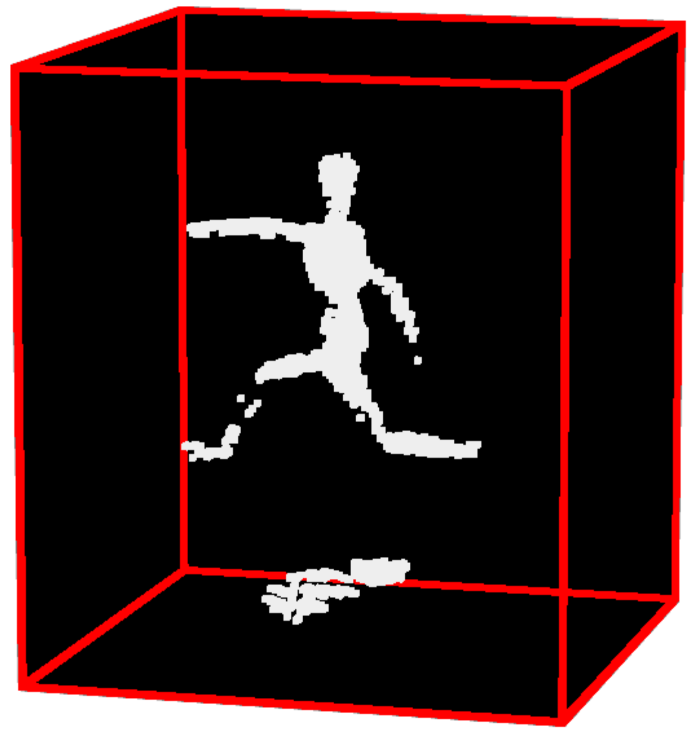


Experimental Setup



60 Shadow
Measurements

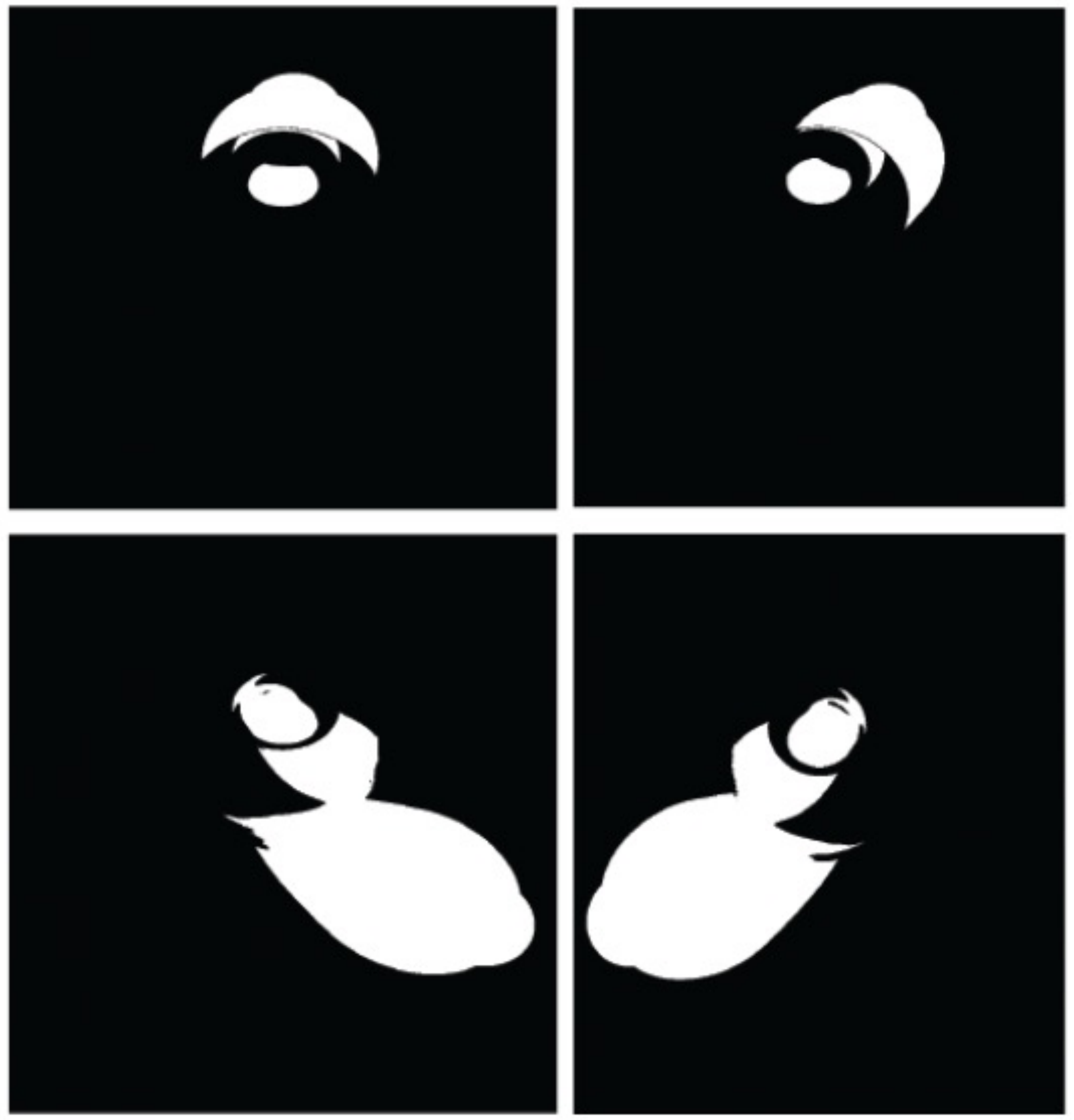
Shadow Carving



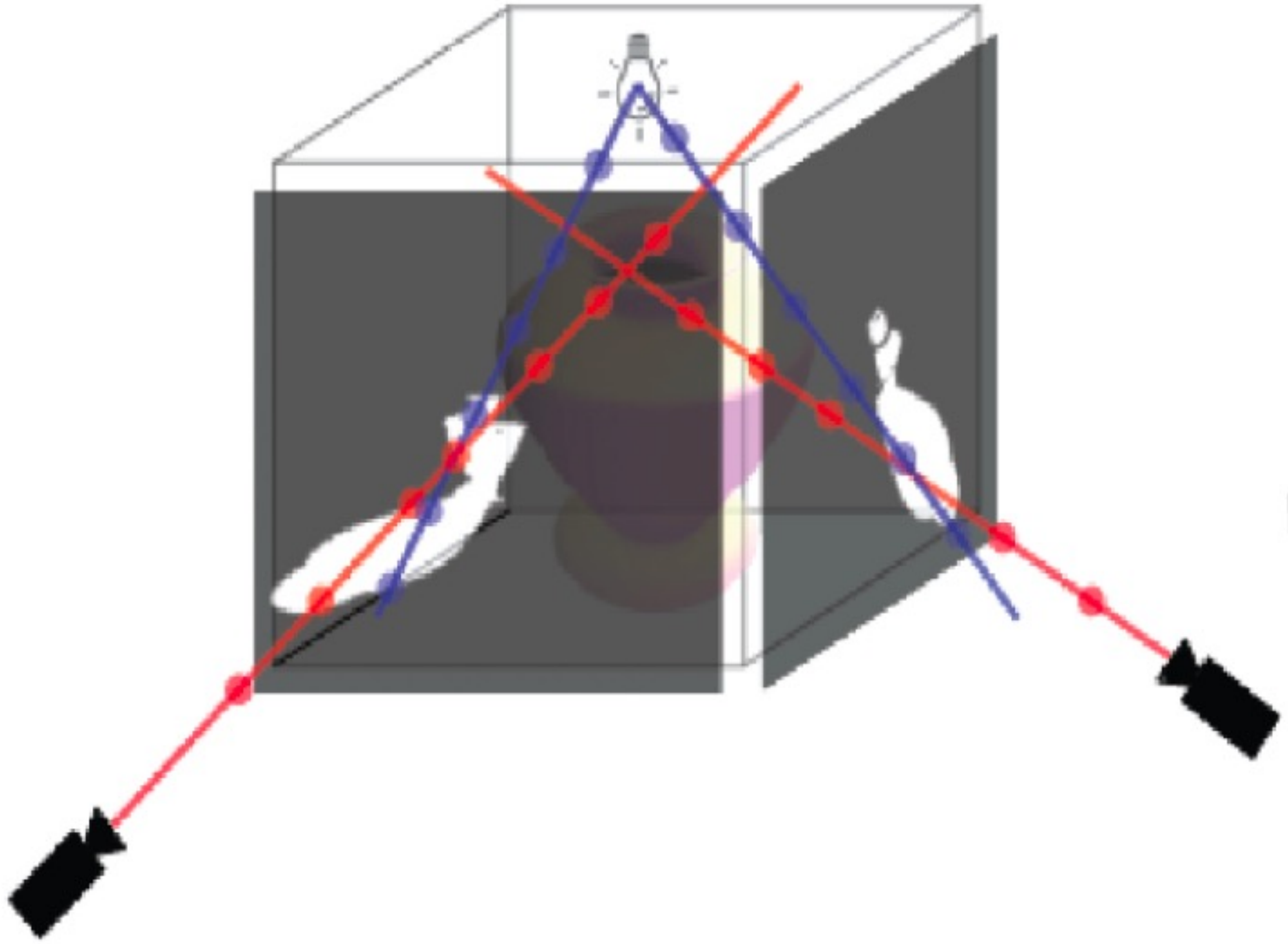
C. Henley et al., ***"Imaging behind occluders using two-bounce light"***, **ECCV 2020**

Learning Neural Fields from Shadow Measurements

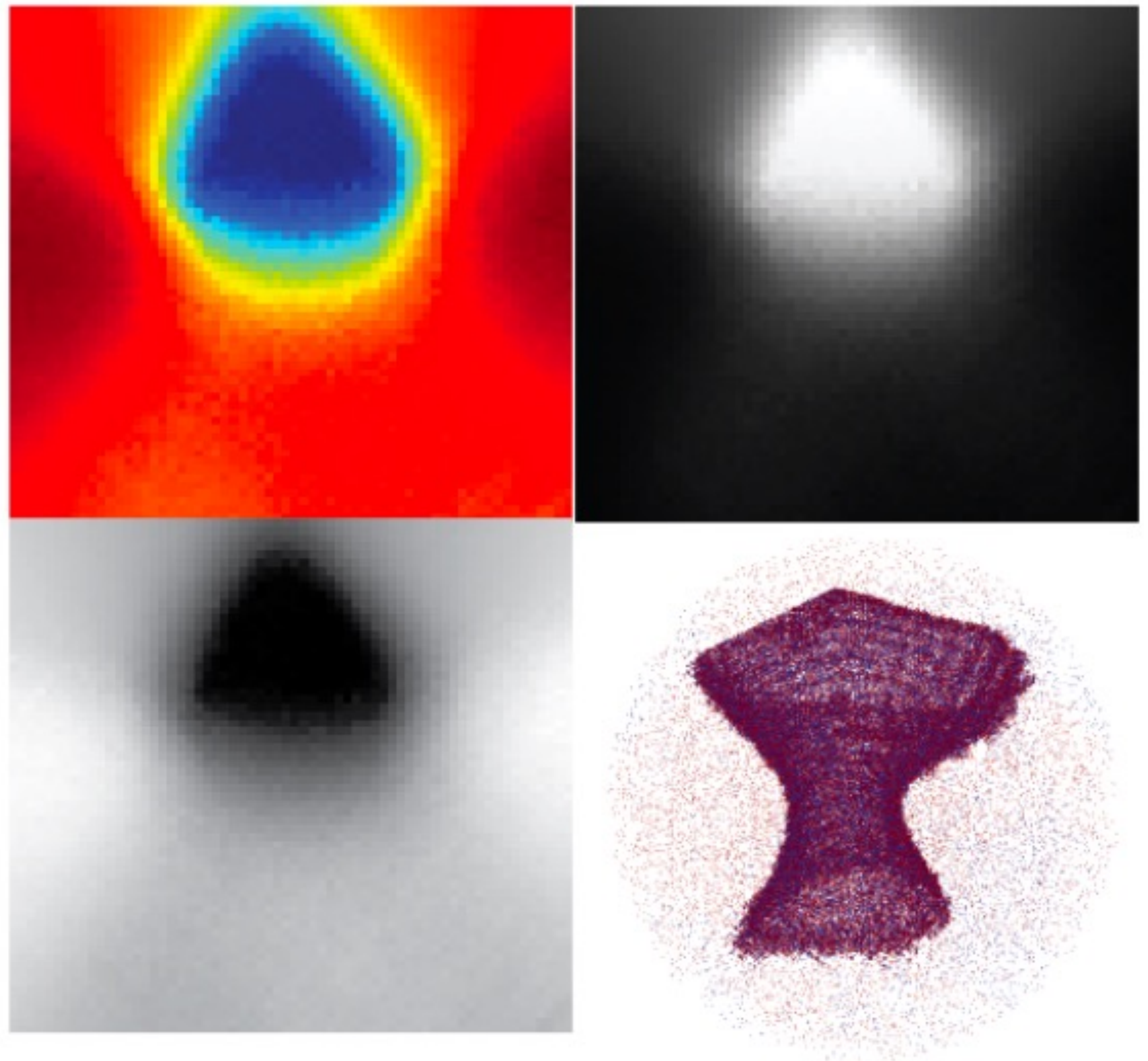
Binary Shadow Masks captured with varying camera position and fixed lighting



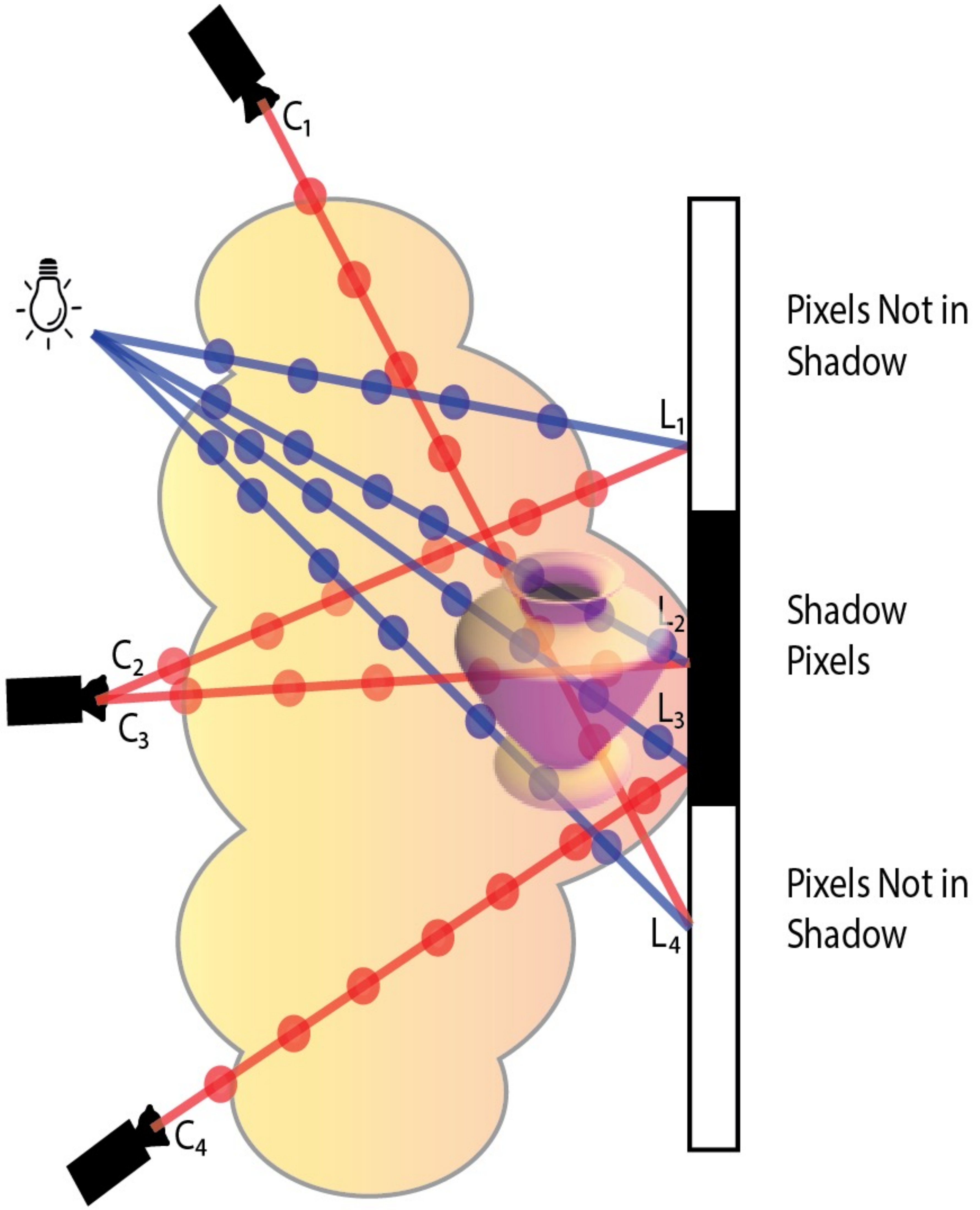
Proposed approach to exploit shadow cues in the scene



Estimated Depth, Shadow Mask, Disparity Map, and Mesh only through binary shadows



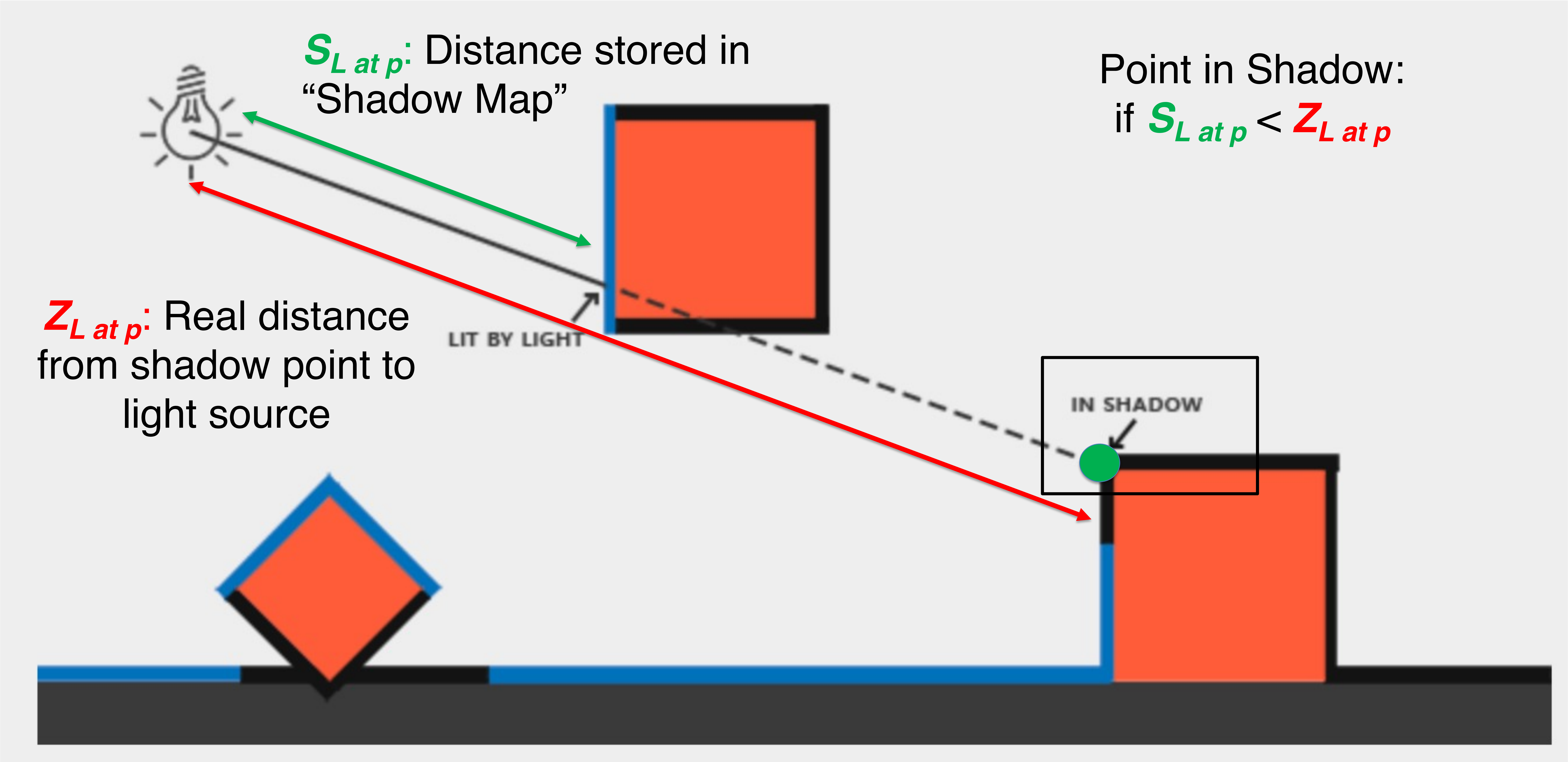
What are Shadows?



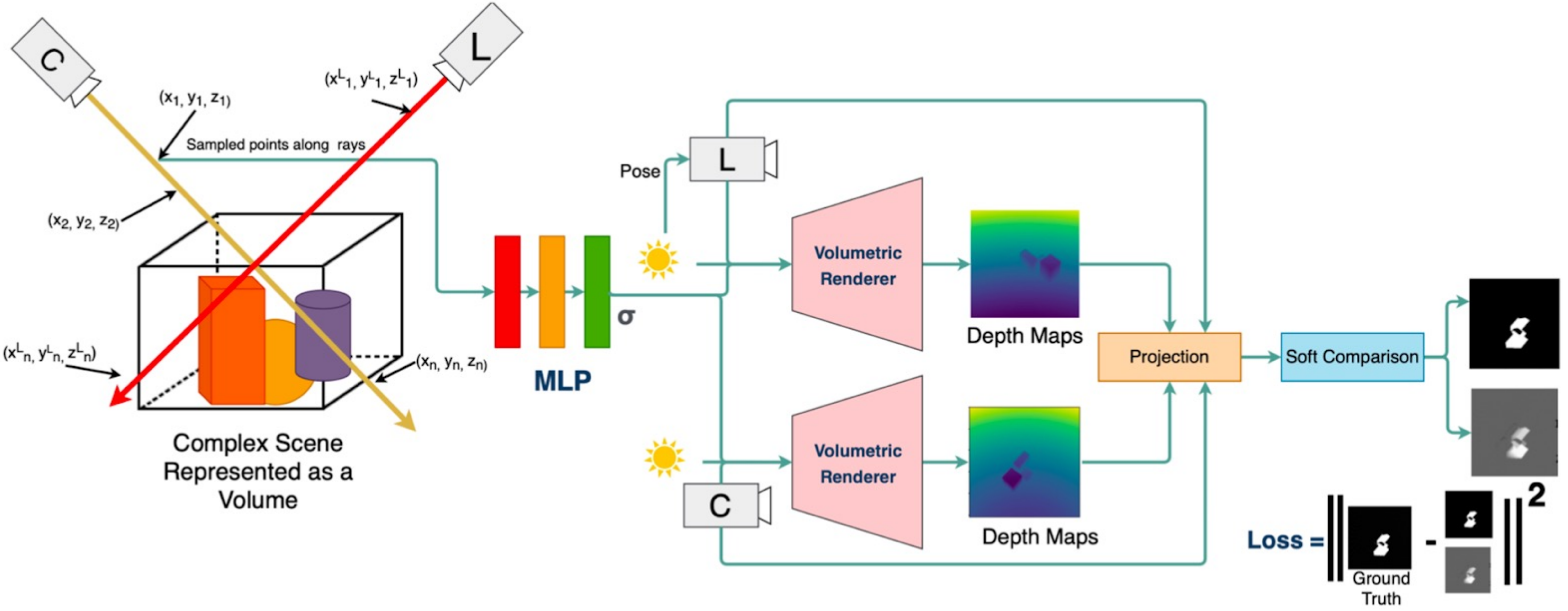
All points in the world <u>without</u> a direct path to the *light source* are defined to be in **shadow**.

Quick Primer on Shadow Mapping

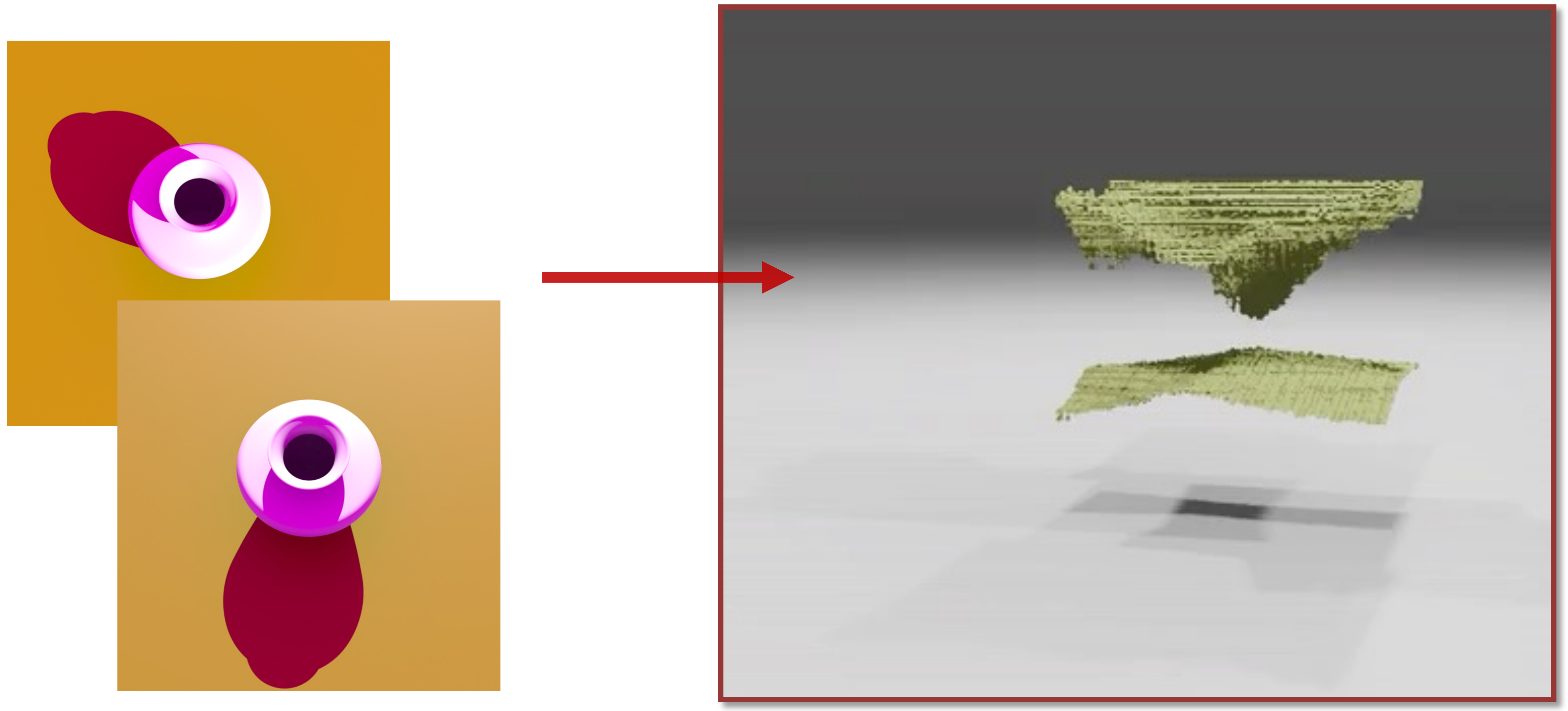
Shadow Map: Distance to the scene from the light's perspective



Differentiable Shadows Forward Model

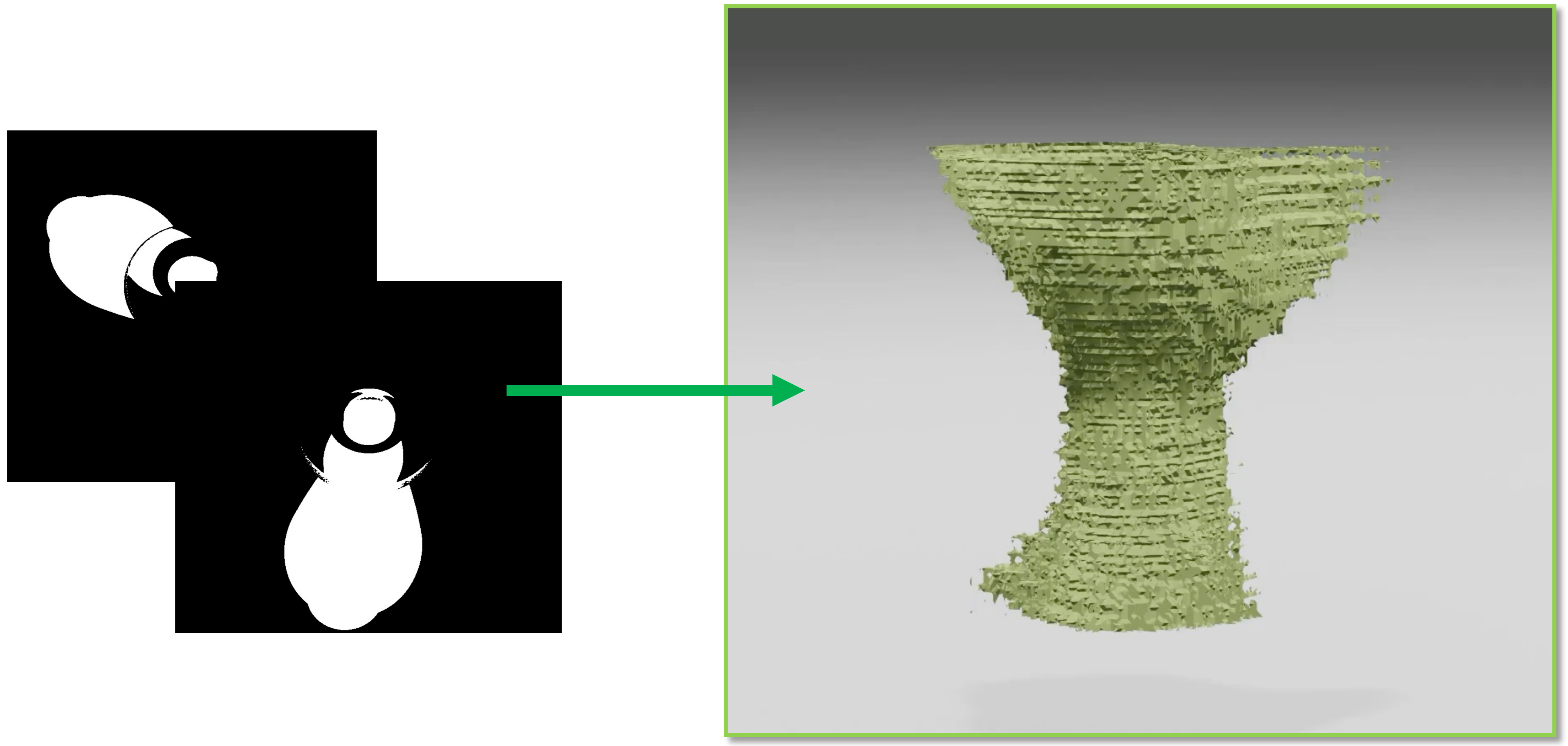


Recovering Hidden Geometry using Shadows



Vase Dataset: Poorly Sampled Vertical Faces, Oblique Lighting exposes Vase Geometry, Texture Less

Photometric Consistency: Changing viewpoints to top-down leads to poor 3D reconstruction



Neural Fields from Shadow Constraints: Learns Vase is Hollow, forced to exploit hidden geometry!

What cue comes r … is

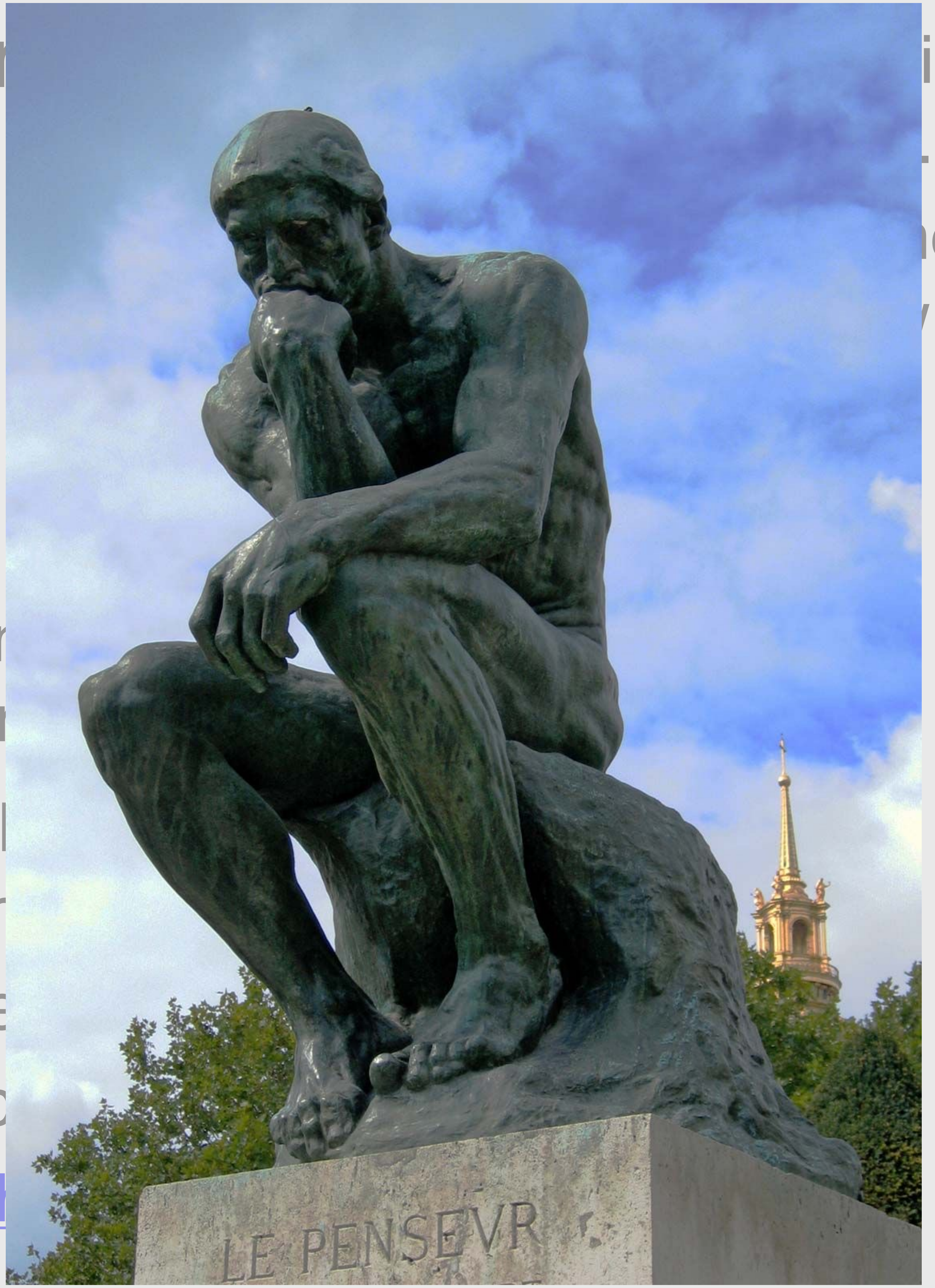
…

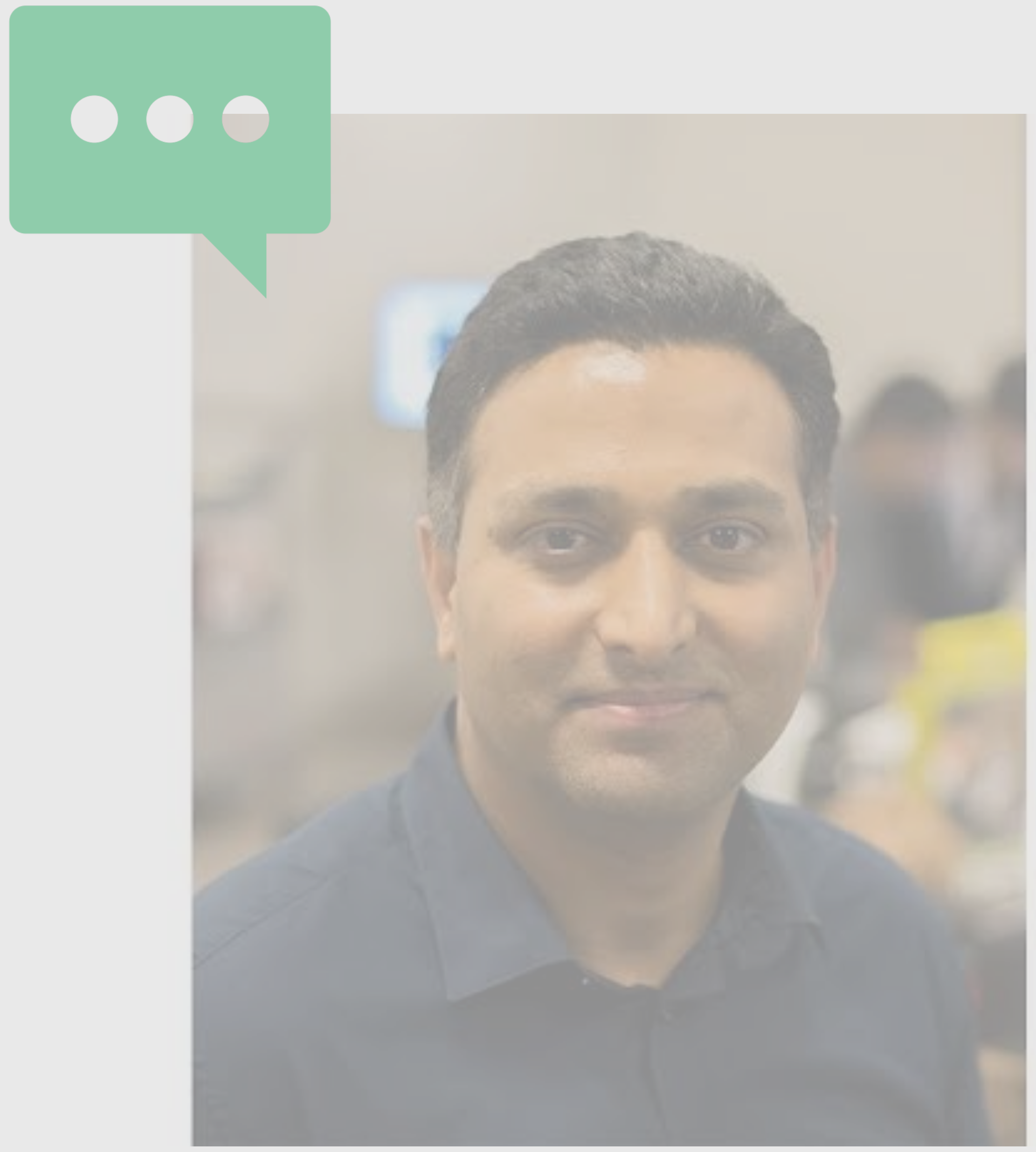
… ow

Rar …

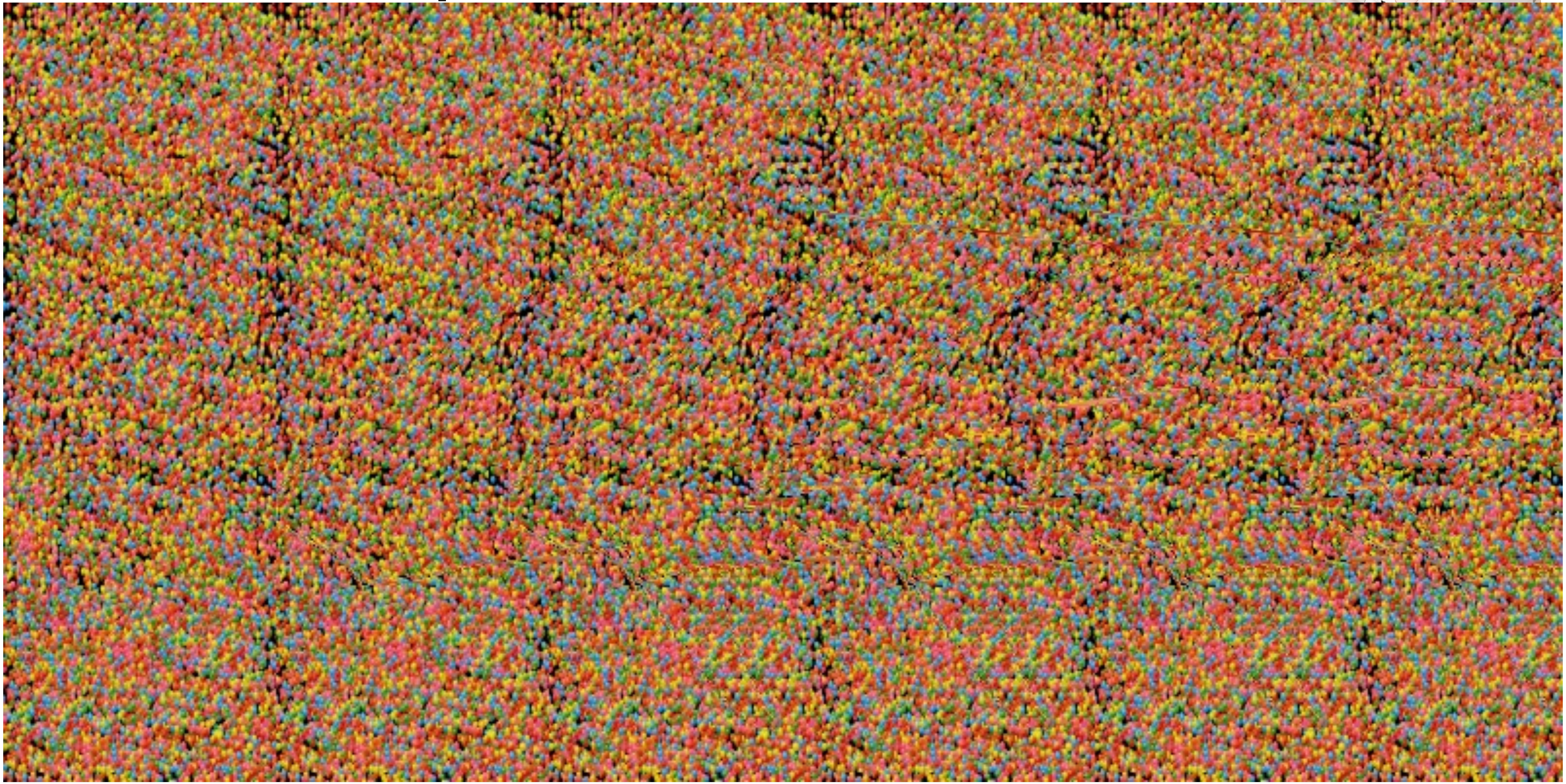
extr …

- Op …
- Im …
- de …
- sc …
- wh …

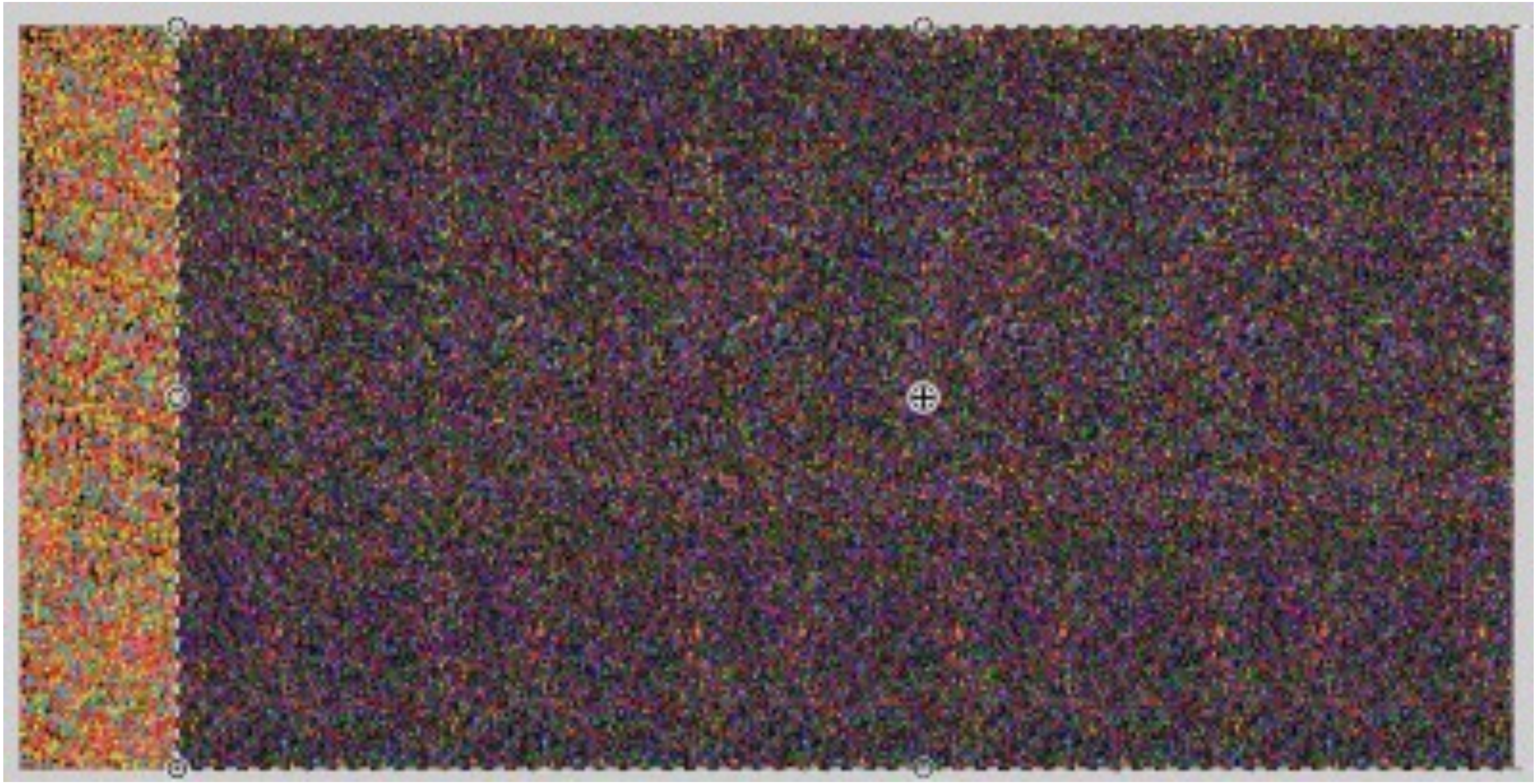




We learn Stereopsis



We learn Stereopsis

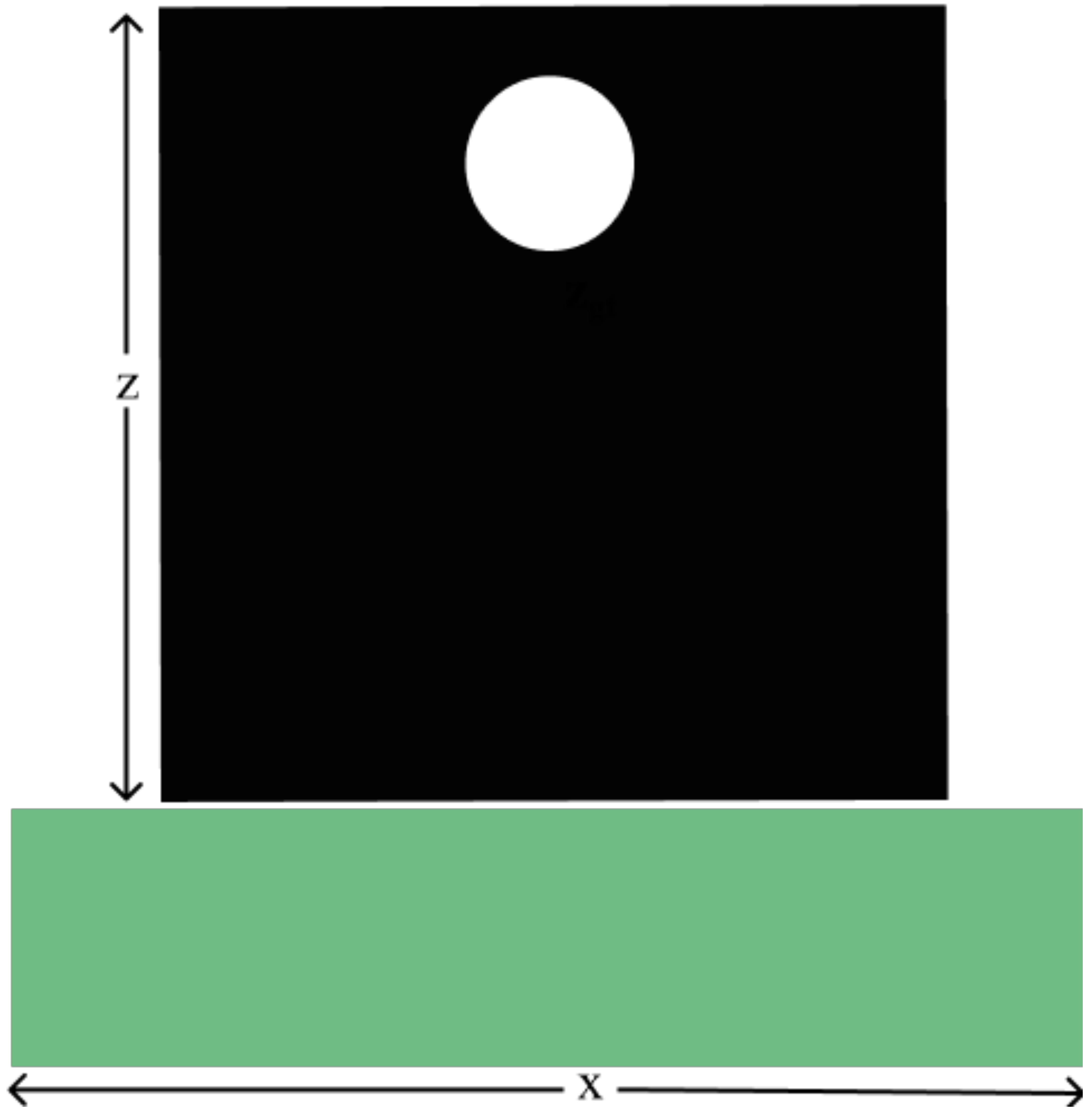


Can we automatically learn Stereopsis?

Goal: Figure out the distance to the sphere.

Environment: no monocular cues.

- Spheres at varying distances
- Spheres with varying sizes
- No Shading & Lighting Cues



Can we automatically learn Stereopsis?

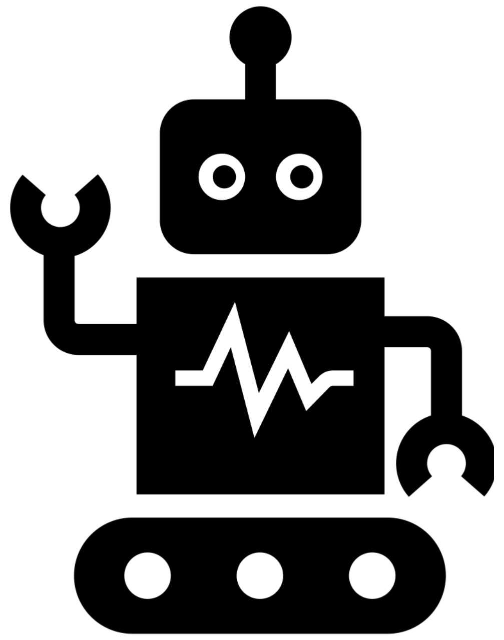
Goal: Figure out the distance to the sphere.

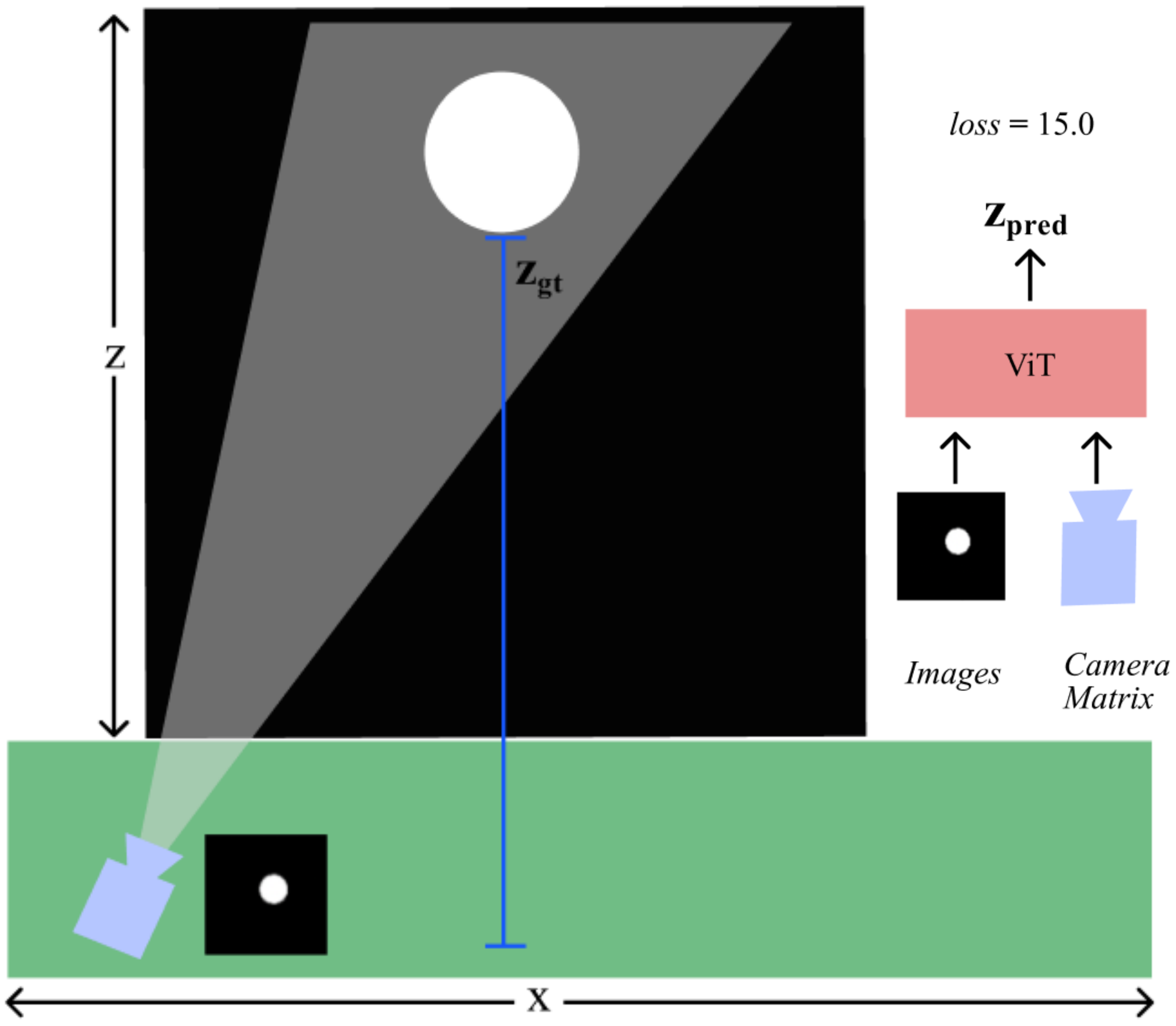
Environment: no monocular cues.
- Spheres at varying distances
- Spheres with varying sizes
- No Shading & Lighting Cues

Actions: Place Cameras
- Choose Positions
- Choose Yaw

Reward: Depth Estimation
- Neural network trained from scratch estimates depth

reward = -0.5





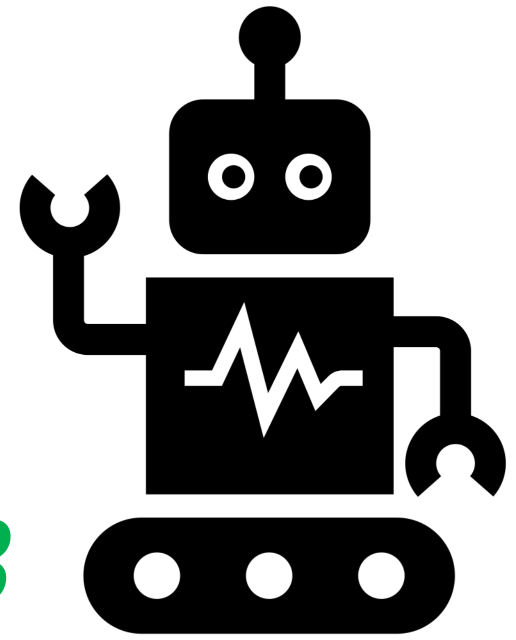
Can we automatically learn Stereopsis?

Goal: Figure out the distance to the sphere.

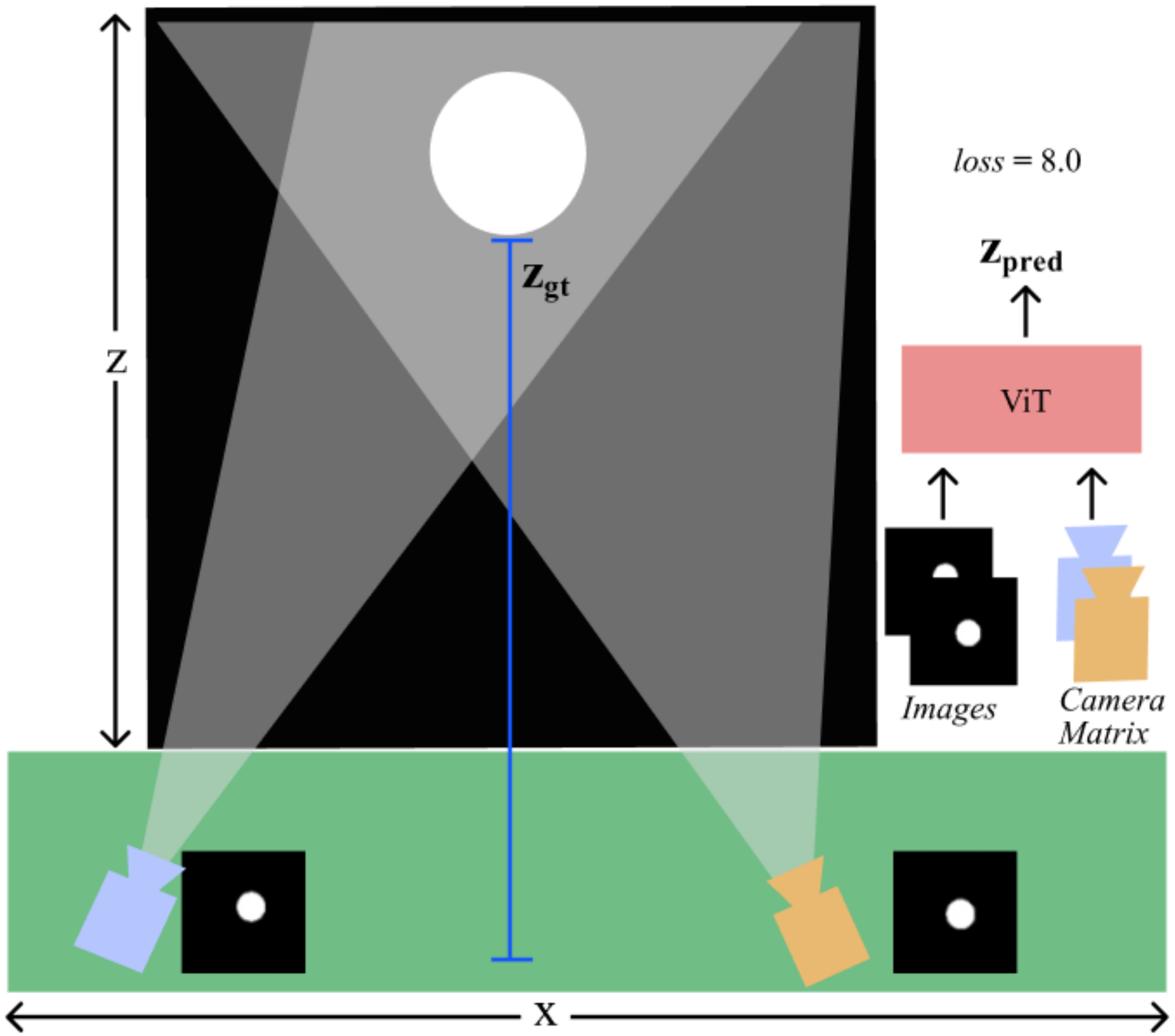
Environment: no monocular cues.
- Spheres at varying distances
- Spheres with varying sizes
- No Shading & Lighting Cues

Actions: Place Cameras
- Choose Positions
- Choose Yaw

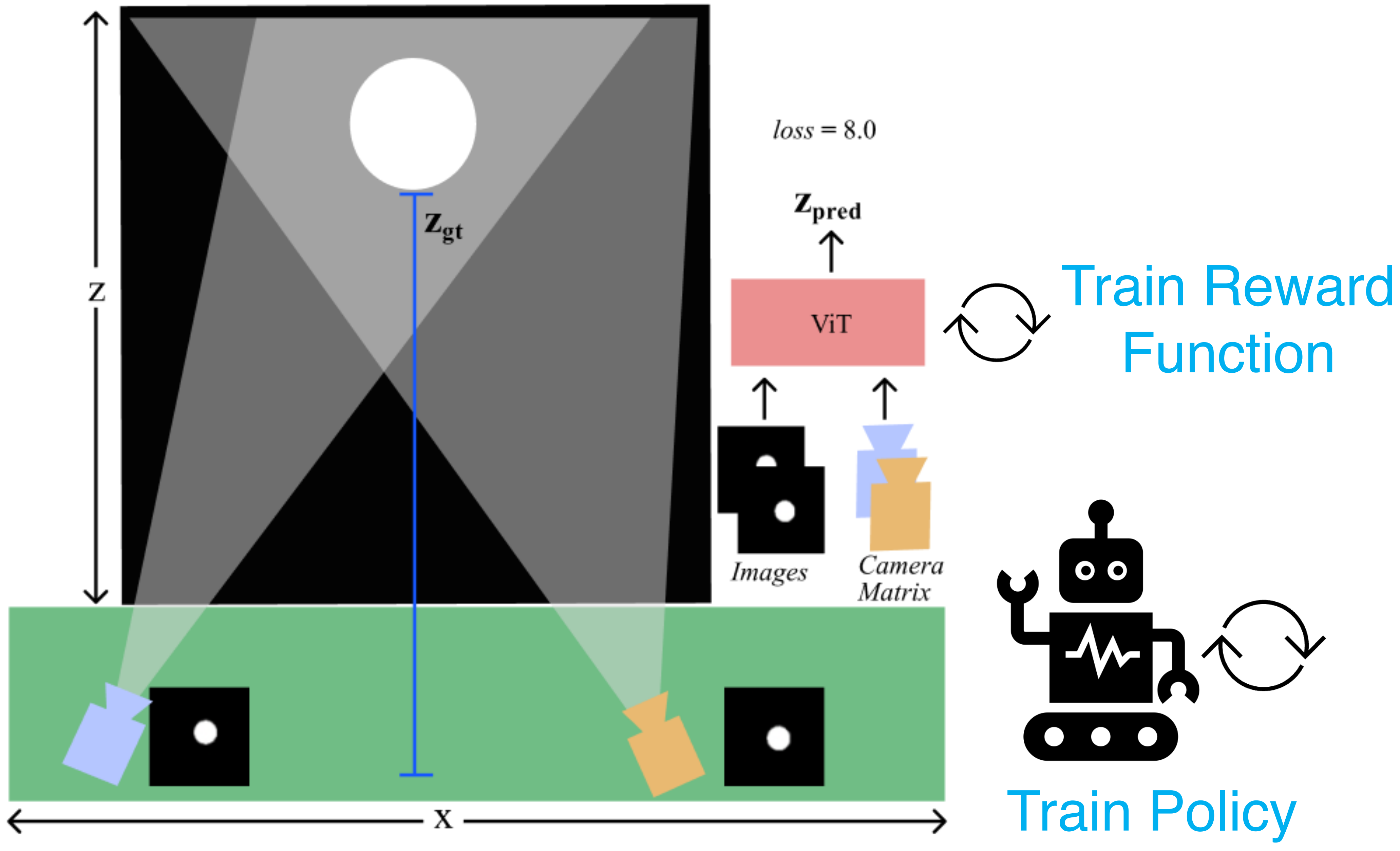
Reward: Depth Estimation
- Neural network trained from scratch estimates depth

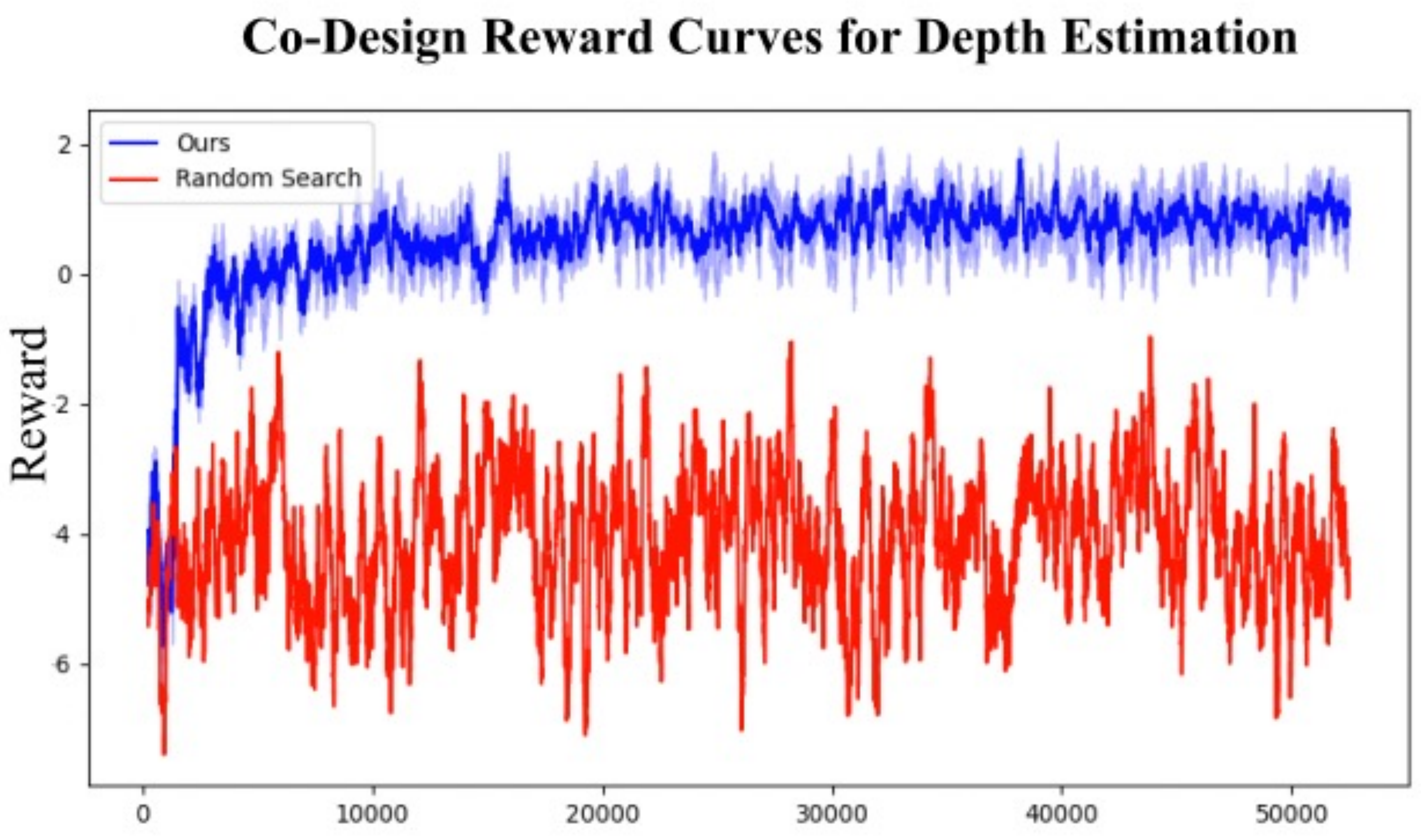


reward = +0.3



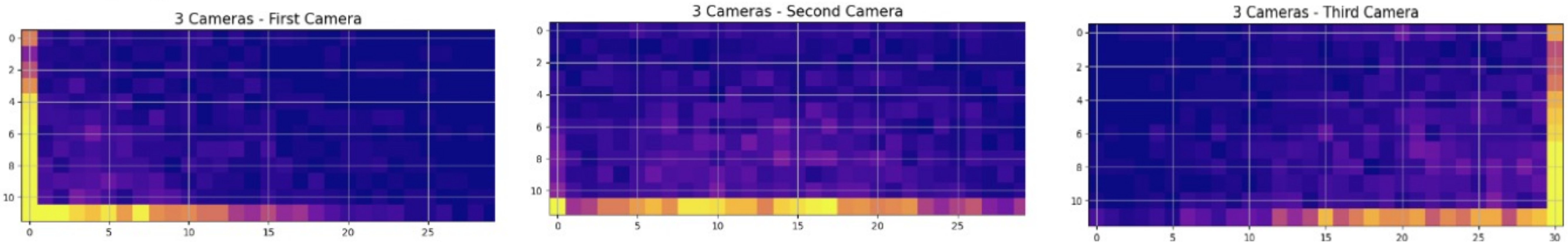
Testing the agent if it has learned Stereopsis





Jointly optimize depth perception (neural network) & place cameras (policy)

Camera Placement Heatmap when Agent Places 3 Cameras (N=7000)



Testing the agent if it has learned Stereopsis

Evaluating the Policy

Coverage	L1 Loss
0	14.0
1	9.2
2	7.2
3	5.7

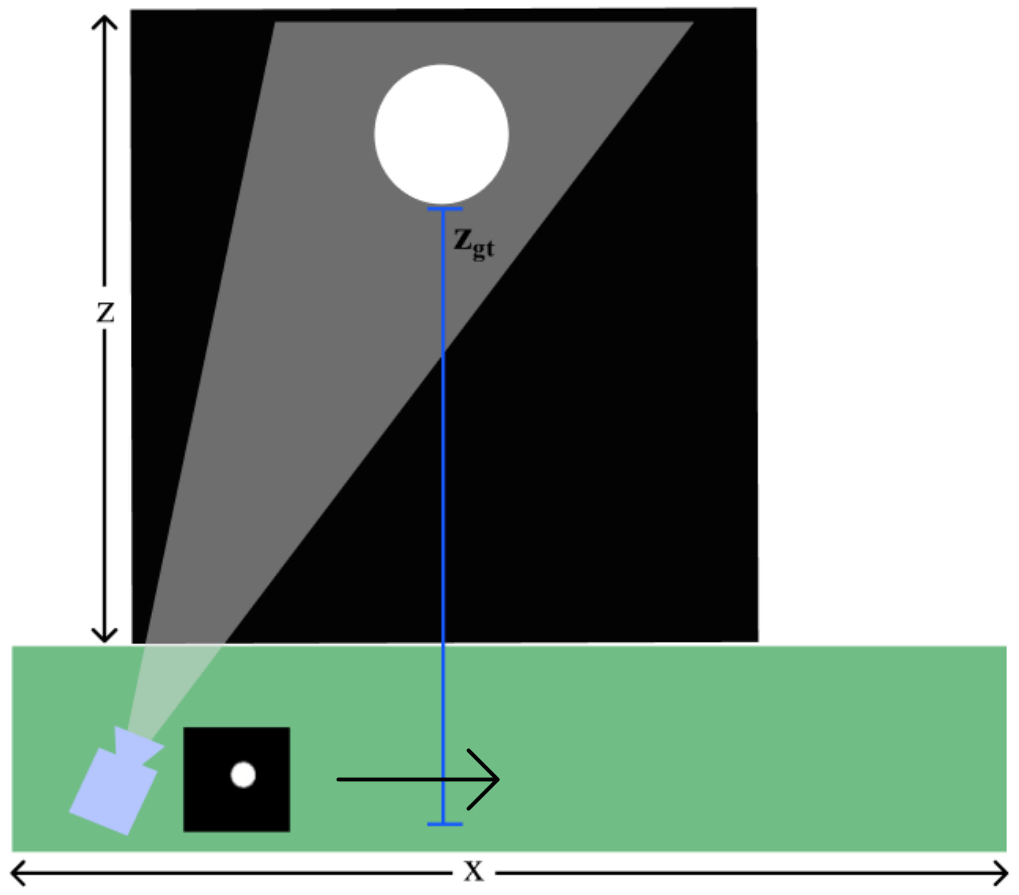
Table 1: Increases *"coverage"* leads to better depth estimation shows reliance on multi-view cues

Cam Config	Mean (x,z)	Std (x,z)	Mean Yaw	Std Yaw
1	$(-4.6, 79.2)$	$(10.0, 1.9)$	-15.7	39.8
2	$(-8.3, 78.3)$	$(7.8, 2.7)$	-3.6	43.3
	$(4.6, 77.7)$	$(9.1, 3.2)$	8.8	43.7
3	$(-10.4, 77.8)$	$(6.4, 2.9)$	-0.6	43.7
	$(-1.1, 77.6)$	$(8.6, 3.1)$	9.3	43.1
	$(8.5, 77.3)$	$(7.2, 3.3)$	15.4	41.2
4	$(-11.4, 77.7)$	$(5.4, 3.0)$	3.2	45.1
	$(-4.3, 77.6)$	$(7.5, 3.2)$	11.4	43.5
	$(3.5, 77.2)$	$(7.5, 3.2)$	15.1	41.7
	$(10.9, 77.4)$	$(5.5, 3.2)$	17.0	40.7
5	$(-12.1, 77.7)$	$(4.6, 3.1)$	5.4	43.4
	$(-6.5, 77.9)$	$(6.7, 3.0)$	8.0	44.2
	$(-0.17, 77.4)$	$(7.3, 3.3)$	14.0	41.5
	$(6.6, 77.1)$	$(6.8, 3.4)$	17.9	41.7
	$(12.2, 77.2)$	$(4.5, 3.3)$	18.7	40.5

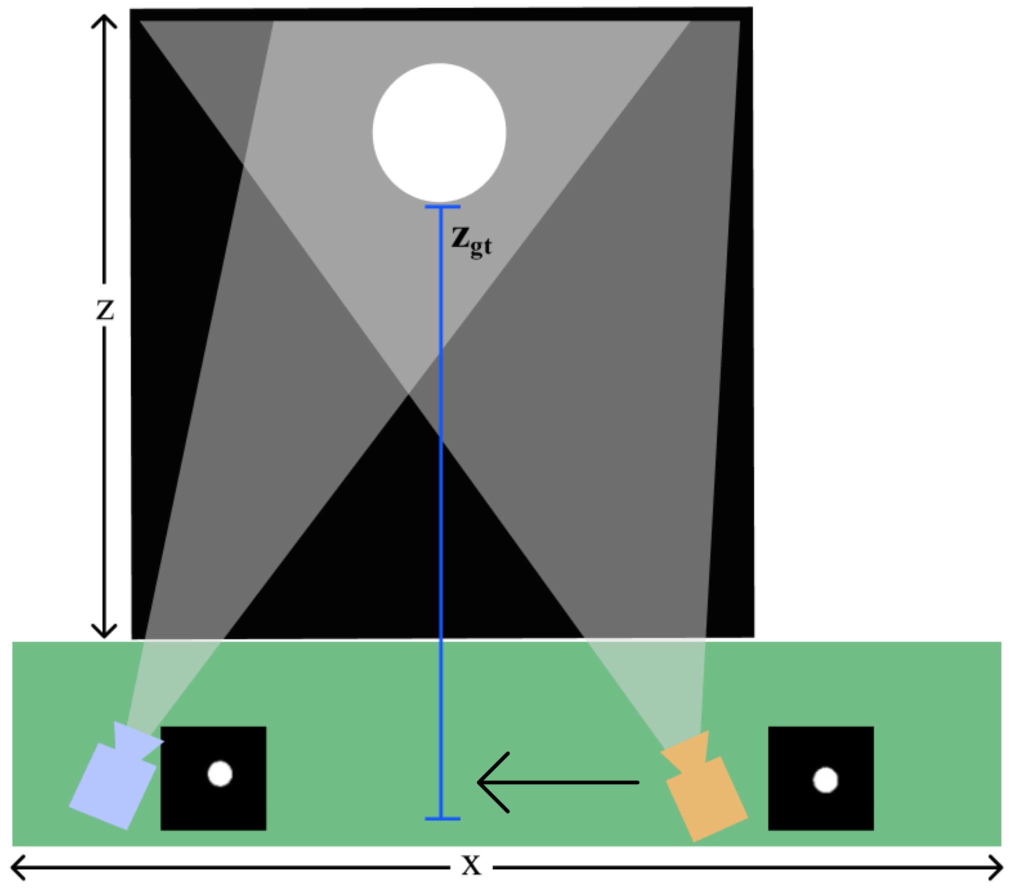
Table 2: Distribution of Actions by the camera placement policy:

1. Maximize Coverage
2. Maximize Baseline

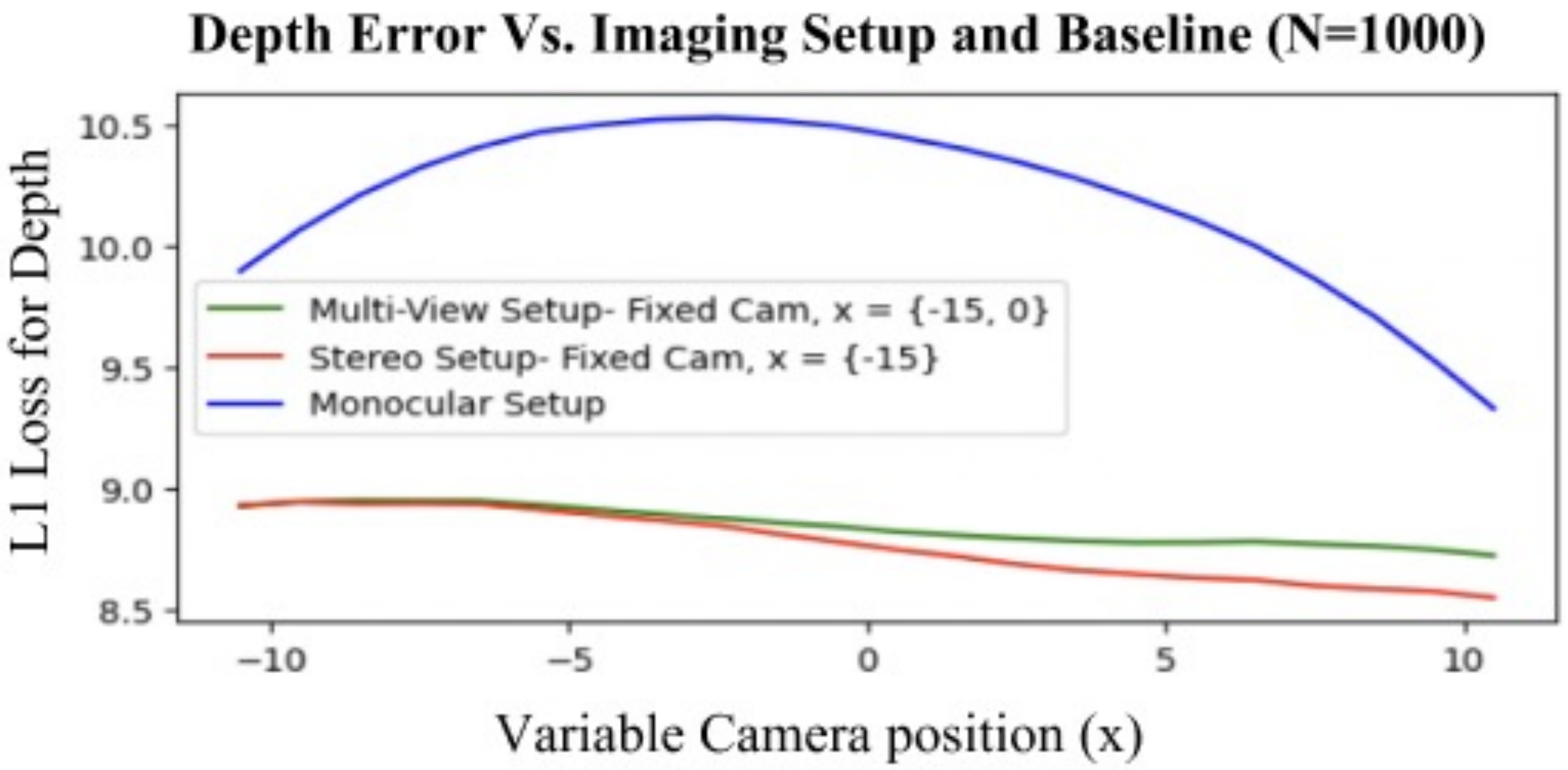
Test the Depth Estimation Network in isolation



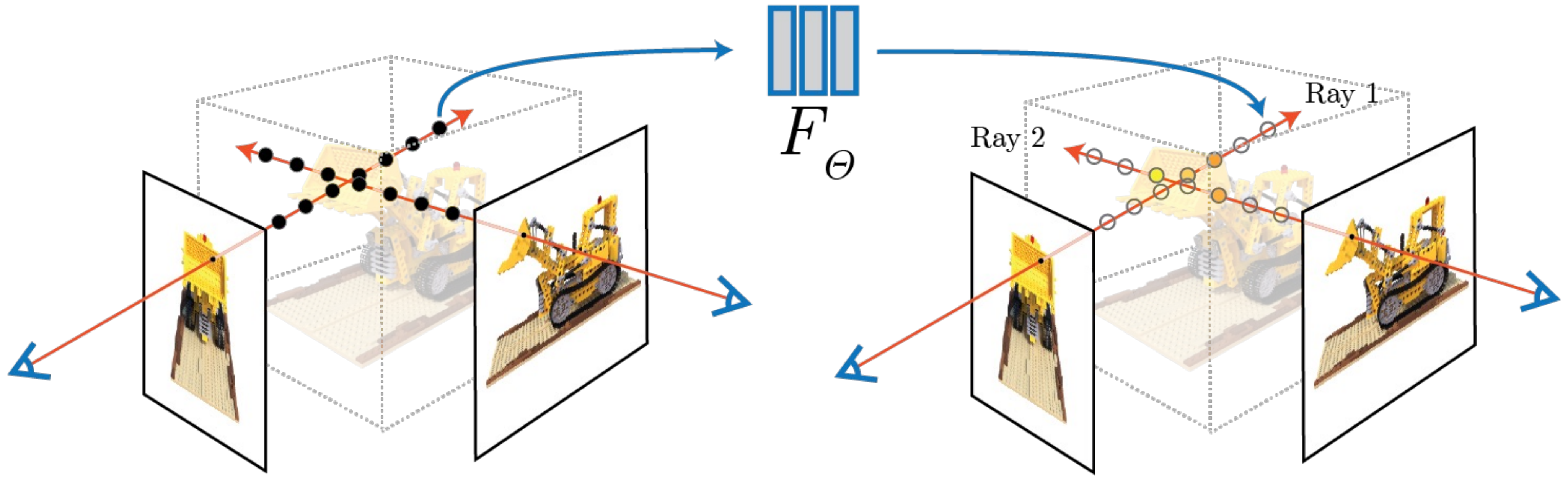
Sweep the Monocular Case

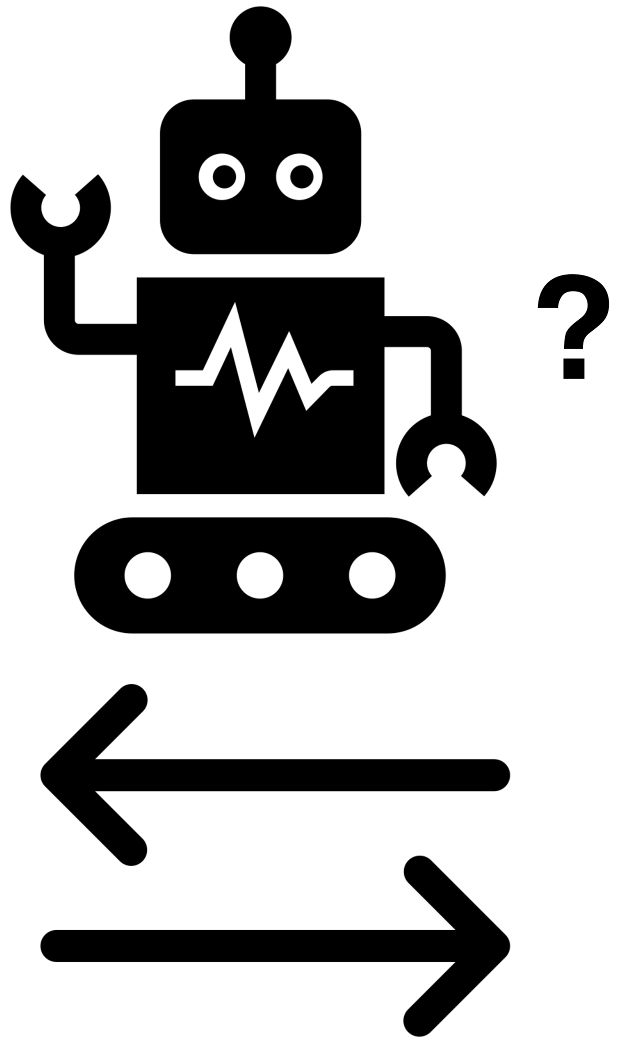


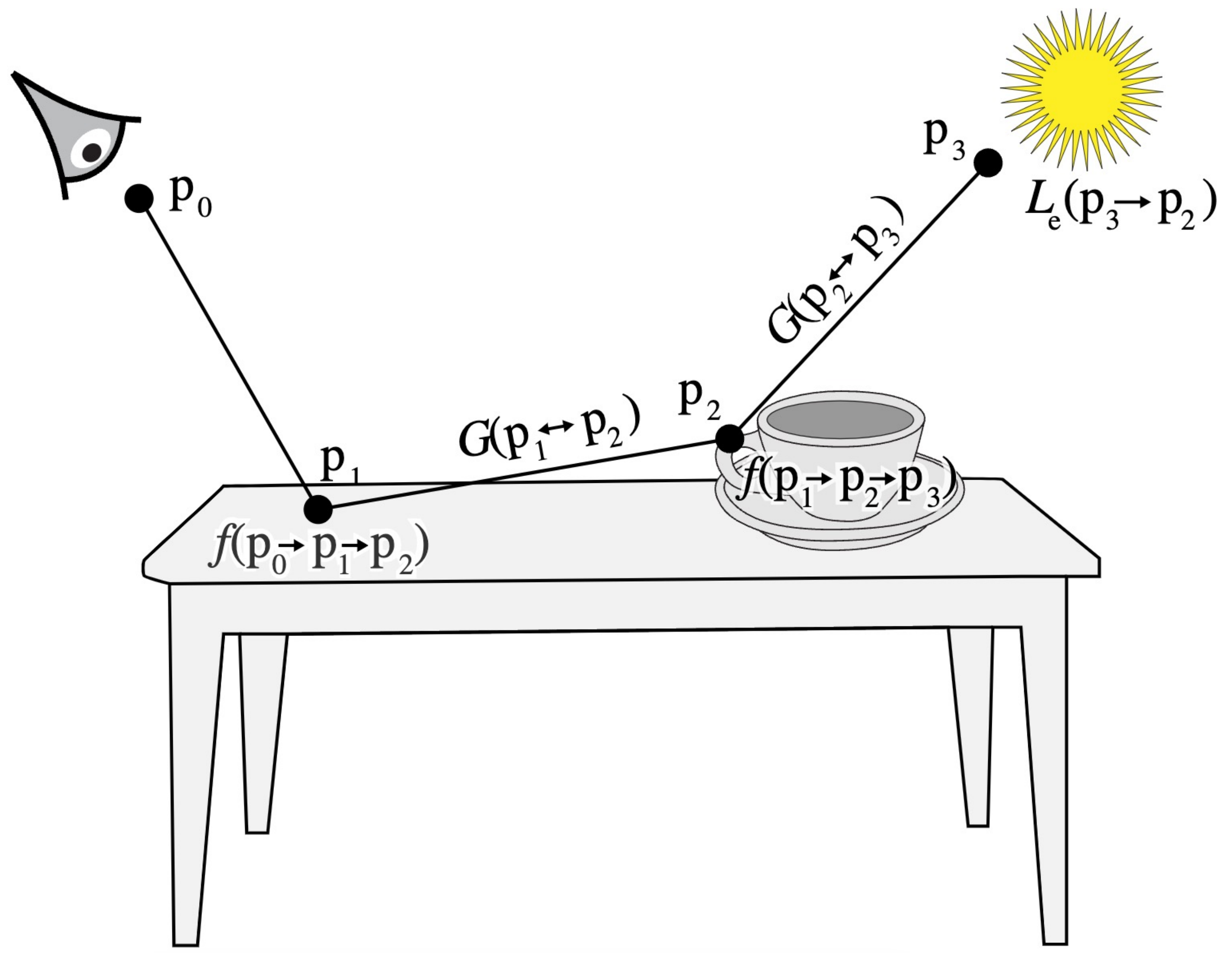
Sweep the Stereo Case



Making Esper Possible..



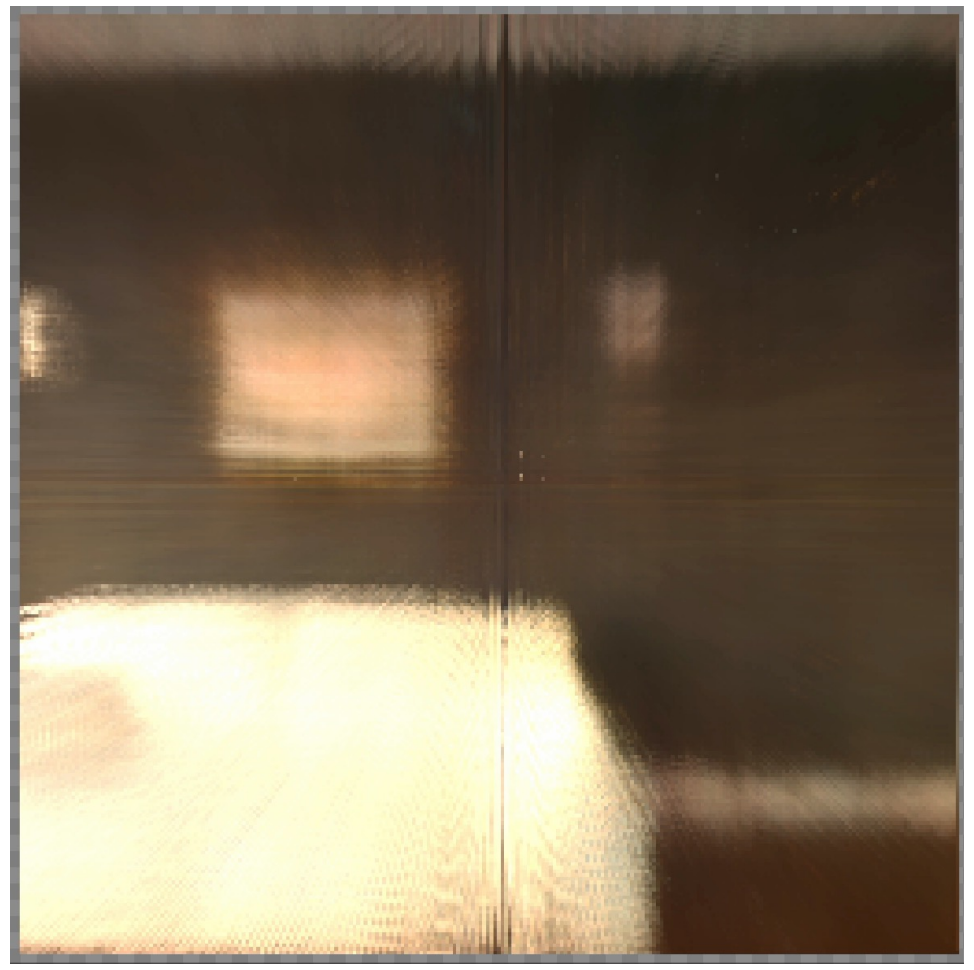




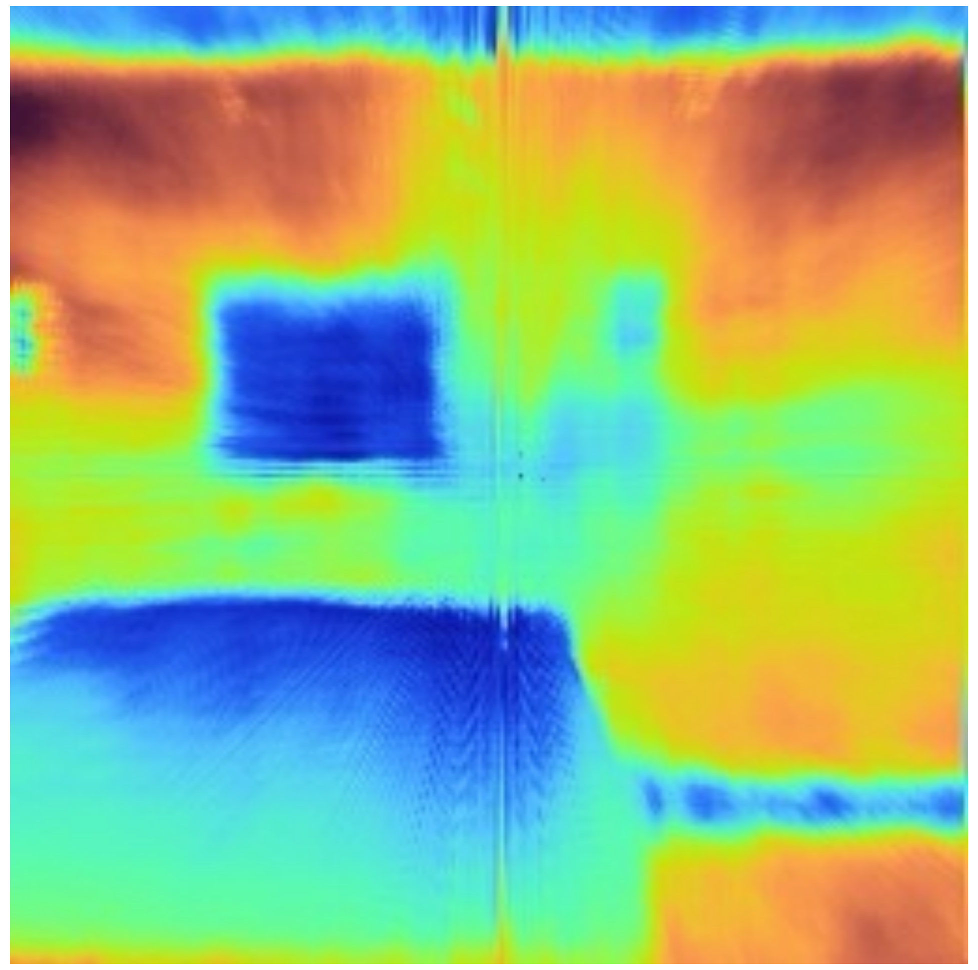
Neural Radiance Fields **Light Transport**

[NeRF (original paper)]
[PBR (book): Physically-based rendering]

Secondary Cues: Reflections, Shadows, Triangulation

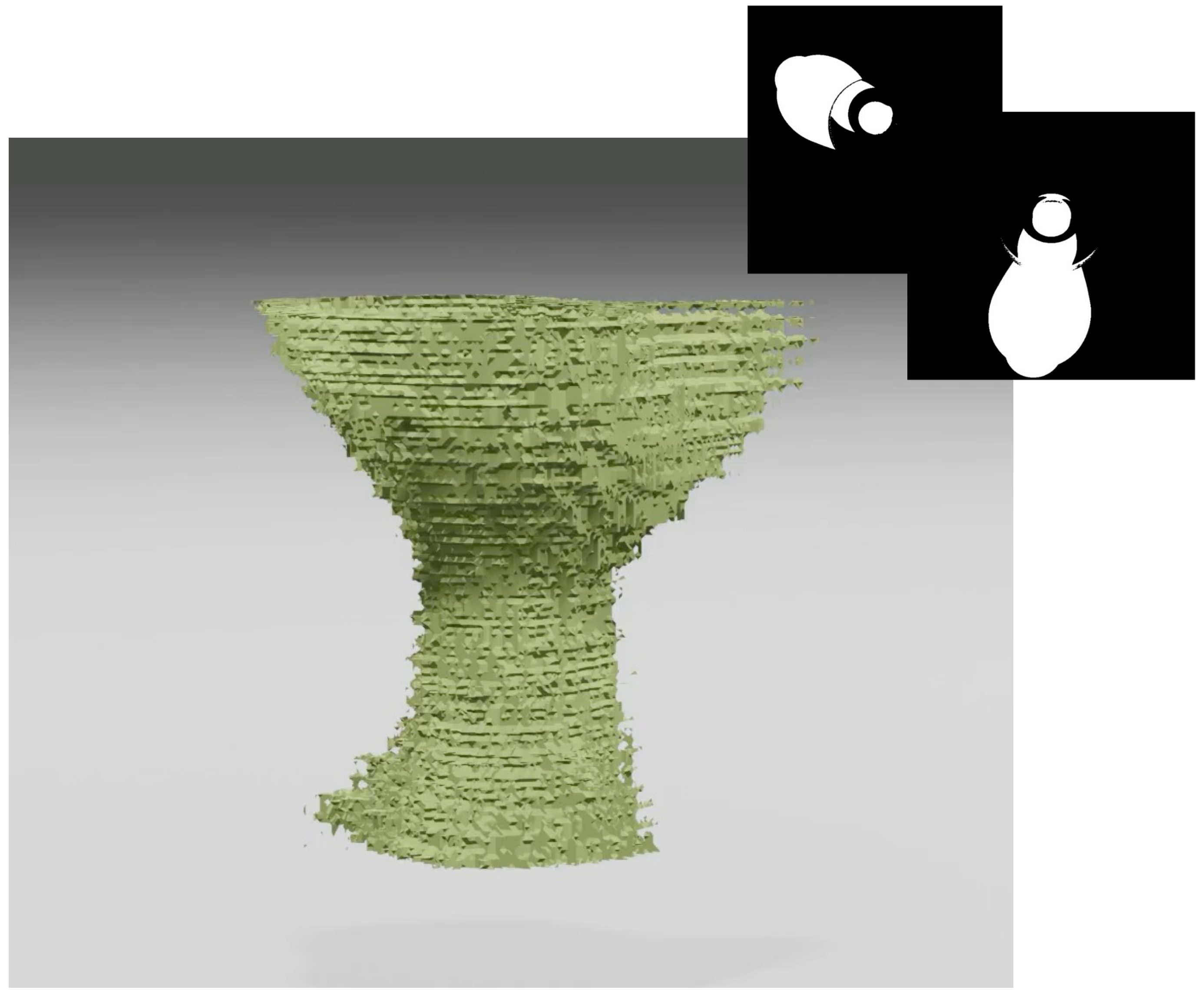


Virtual Camera

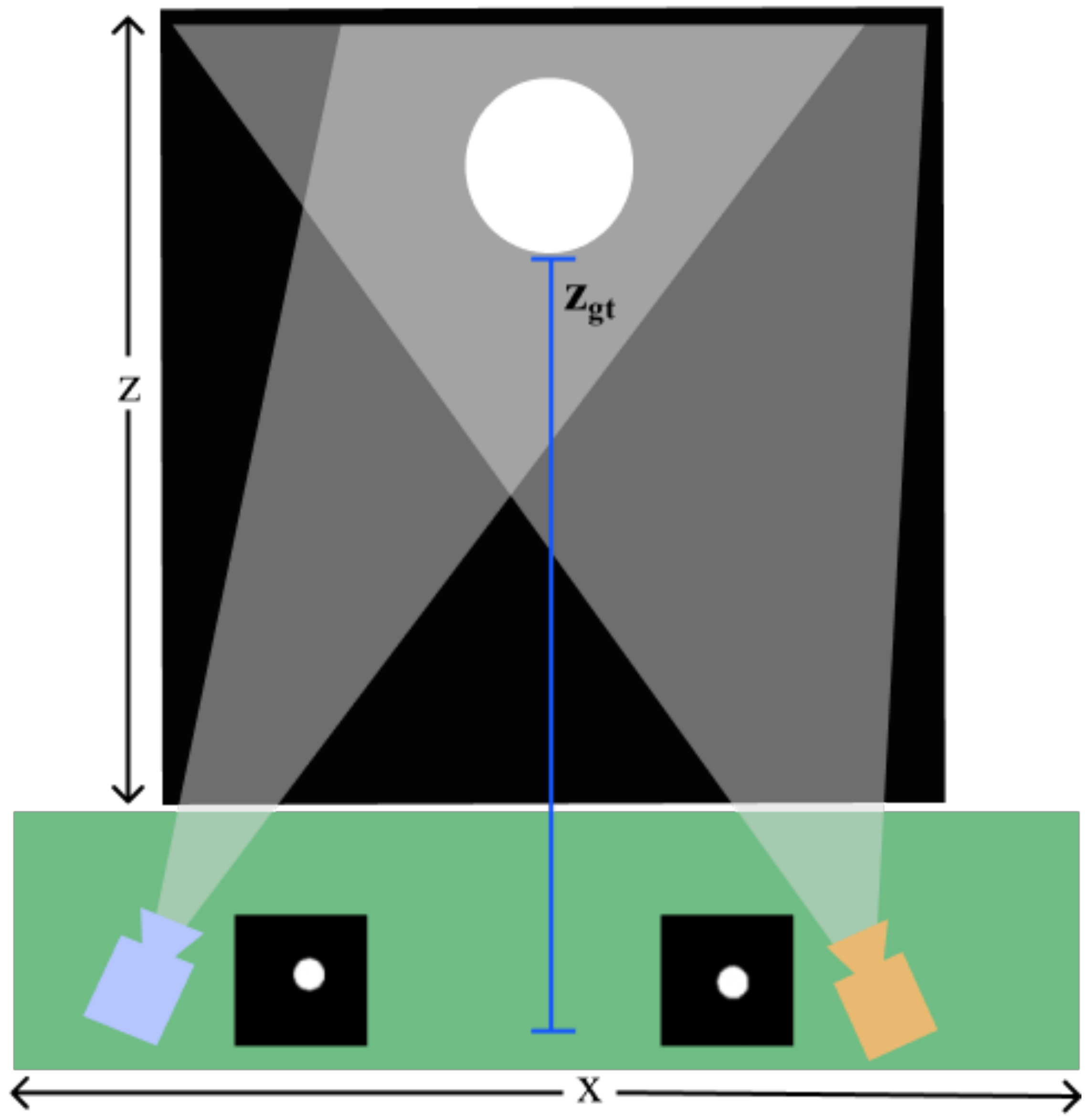


Virtual Depth

Reflections



Shadows



Triangulation

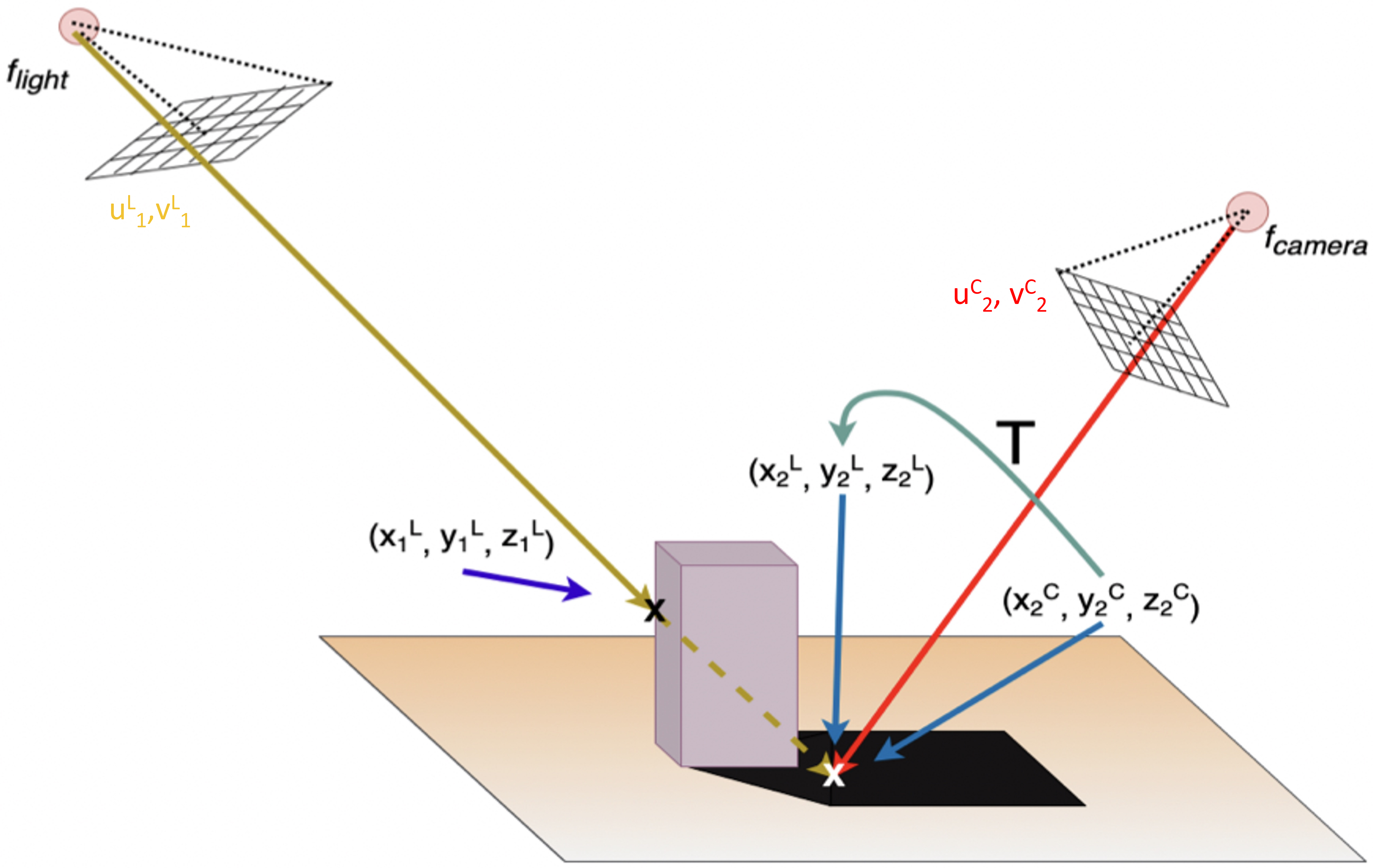
Thank you to all the collaborators!

Collaborators:
Akshat Dave
Nikhil Behari
Tzofi Klinghoffer
Connor Henley
Tristan Swedish
Siddharth Somasundaram
Bhavya Agarwalla

Mentors/Advisors:
Ashok Veeraghavan
Fadel Adib
Pulkit Agrawal
Ramesh Raskar

Backup Slides

Shadow Mapping



Let's consider pixels:

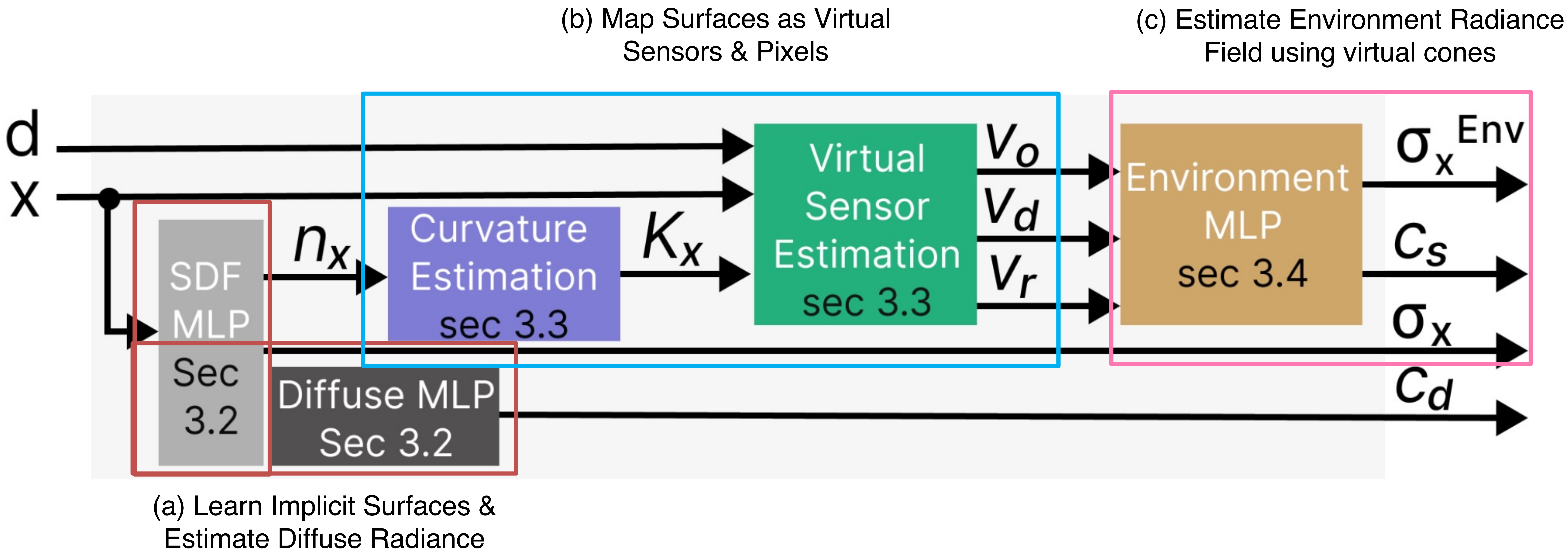
- $(u^C_2, v^C_2, 1) \rightarrow (x^C_2, y^C_2, z^C_2)$
- $(u^L_1, v^L_1, 1) \rightarrow (x^L_2, y^L_2, z^L_2)$
  - *Function **F**: pixel -> Depth at Pixel*
- *Transformation **T: from_camera_to_light***

Shadow Mapping:

1. $F_{camera}((u^C_2, v^C_2, 1)) = (x^C_2, y^C_2, z^C_2)$
2. $F_{light}((u^L_1, v^L_1, 1)) = (x^L_2, y^L_2, z^L_2)$
3. $T(x^C_2, y^C_2, z^C_2) = (x^L_2, y^L_2, z^L_2)$
4. If $z^L_1 < z^L_2$ then point (x^C_2, y^C_2, z^C_2) is <u>IN shadow.</u>

True Virtual Viewpoint Approximation

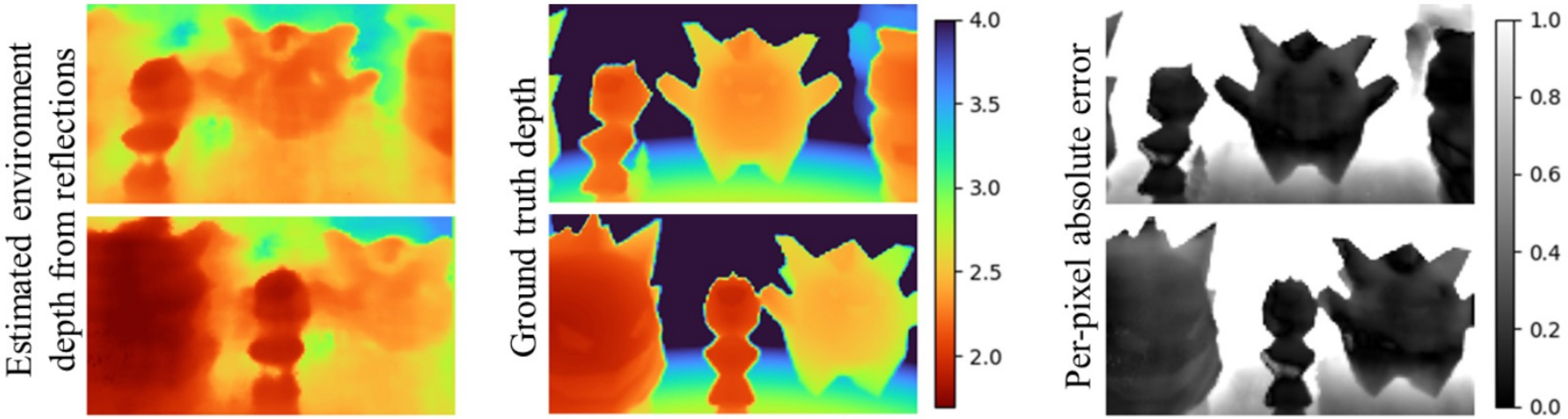
Putting it all together…



Make this slide better, put it inside as components & then show arrows

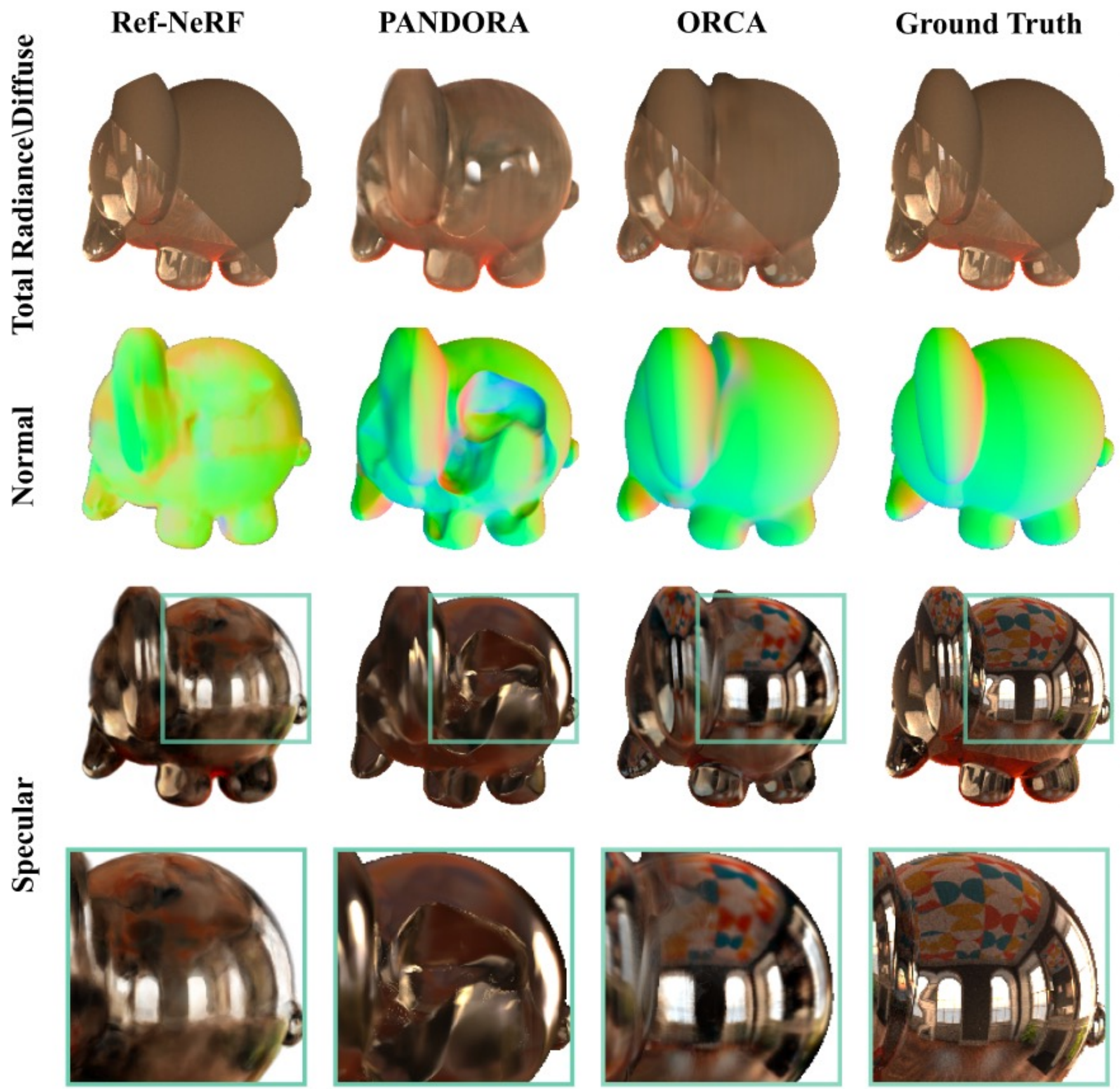
Quantitative results on depth estimation

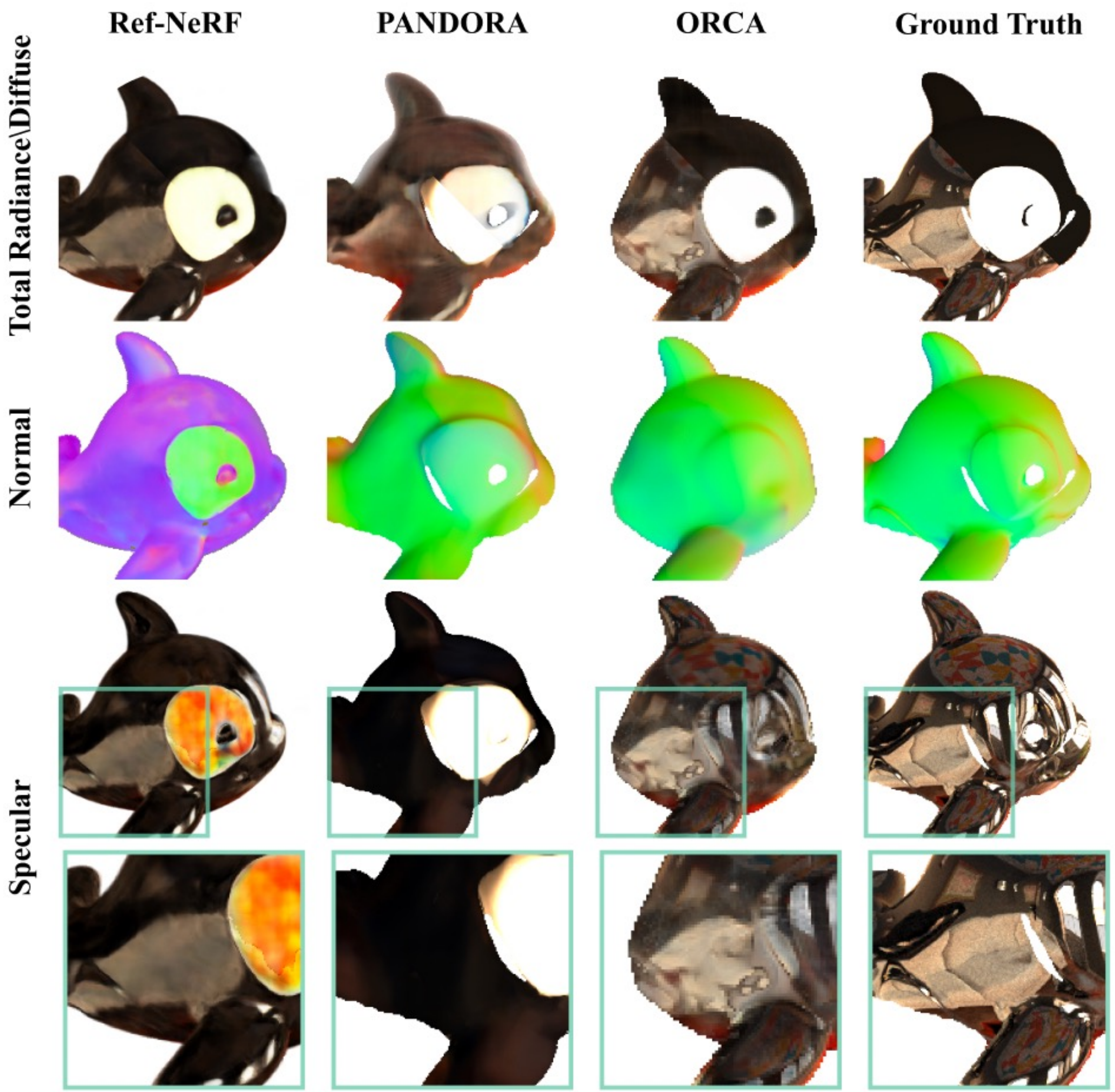
Example Captured Images



Per-Pixel Absolute Error increases with distance similar to most stereo setups

Accurate diffuse-specular separation and smoother geometry





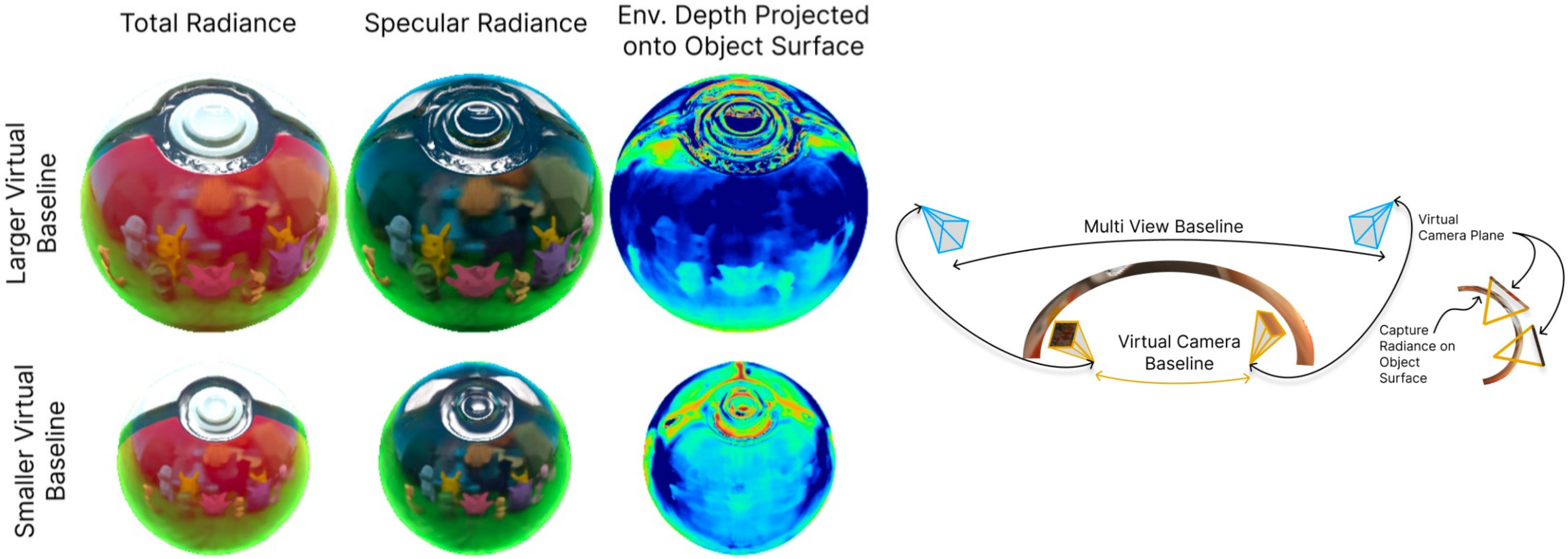
ORCa Applications

from learned environment radiance fields

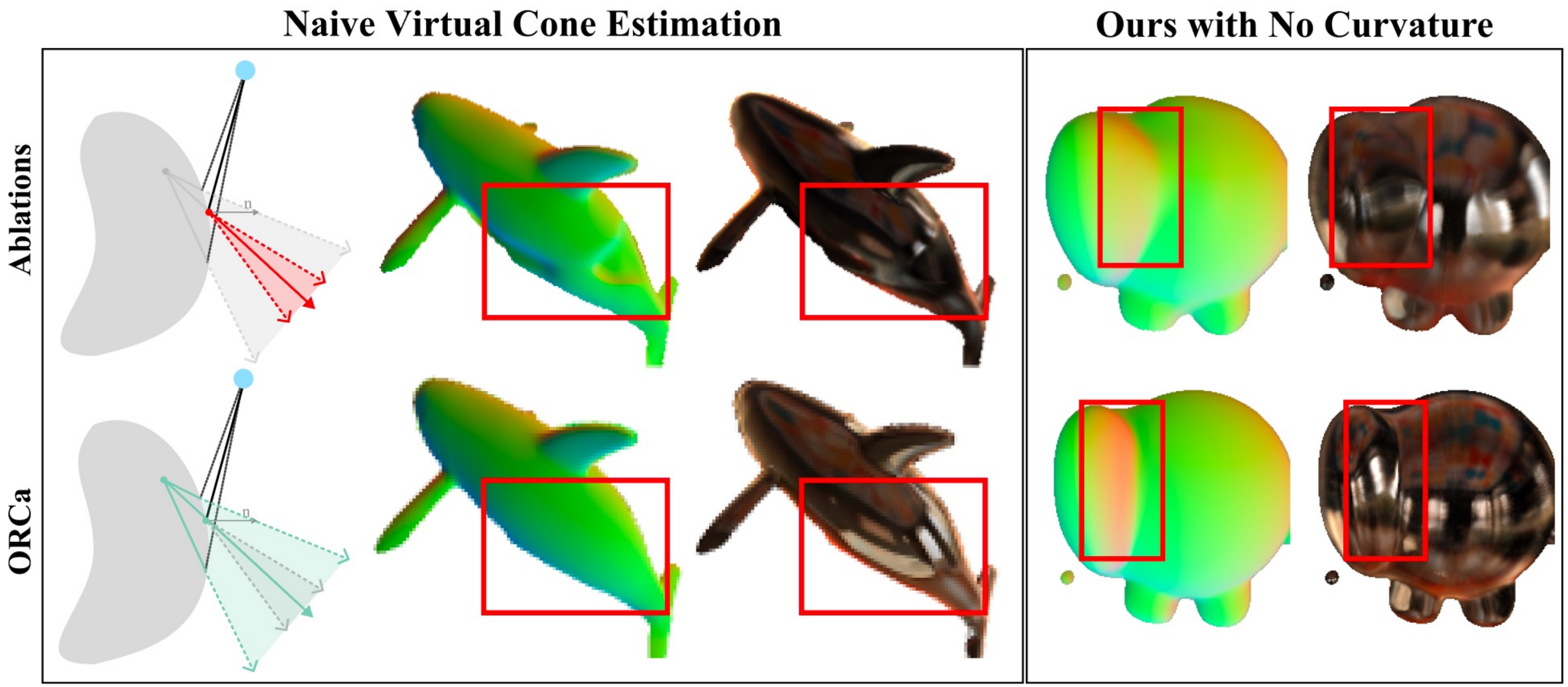
Virtual Object Insertion

Material Editing

Analysis: Object size as virtual baseline



Ablation: Effect of curvature estimation



Current Limitations

Approximations in our implementation

– Glossy surfaces with low roughness*(roughness not explicitly modeled)*

– Single reflecting object

– Inter-reflections not considered

– Mean Curvature approximation

Physical Constraints on exploiting reflections

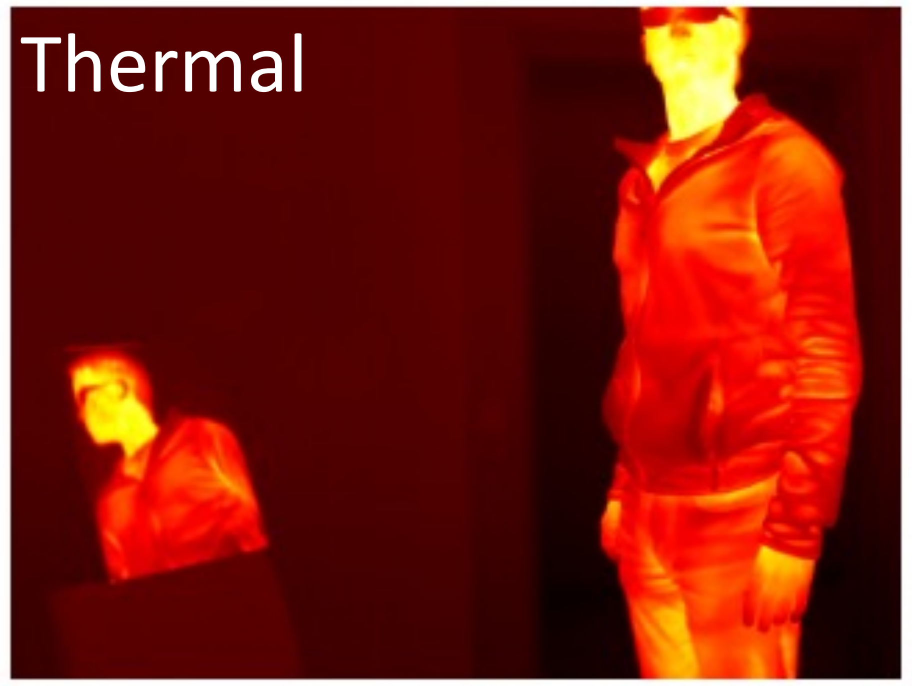
– Virtual Resolution

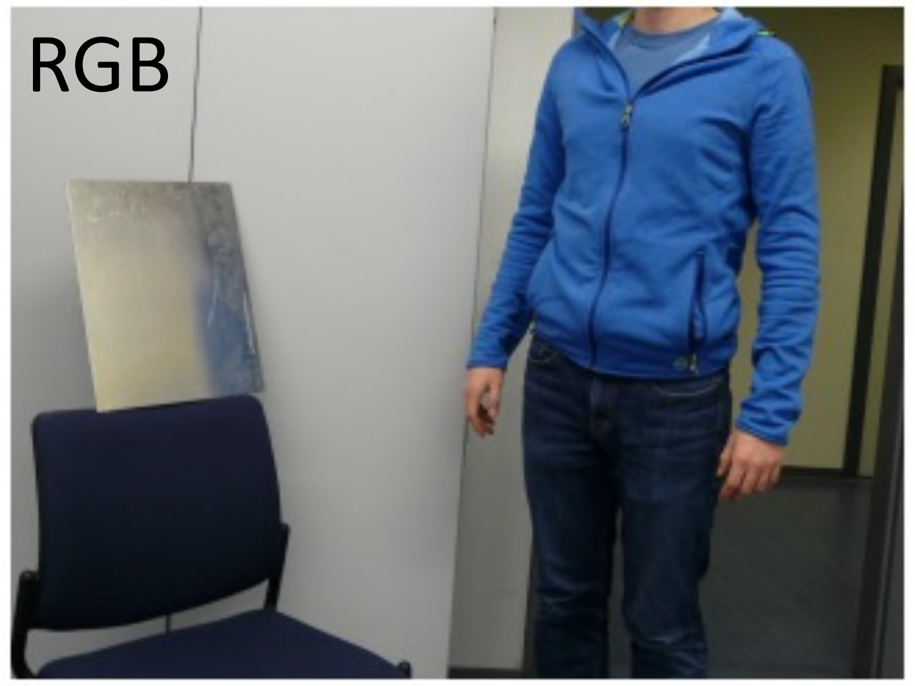
– Depth Estimation by Virtual Baseline

Opportunities with multi-view reflections

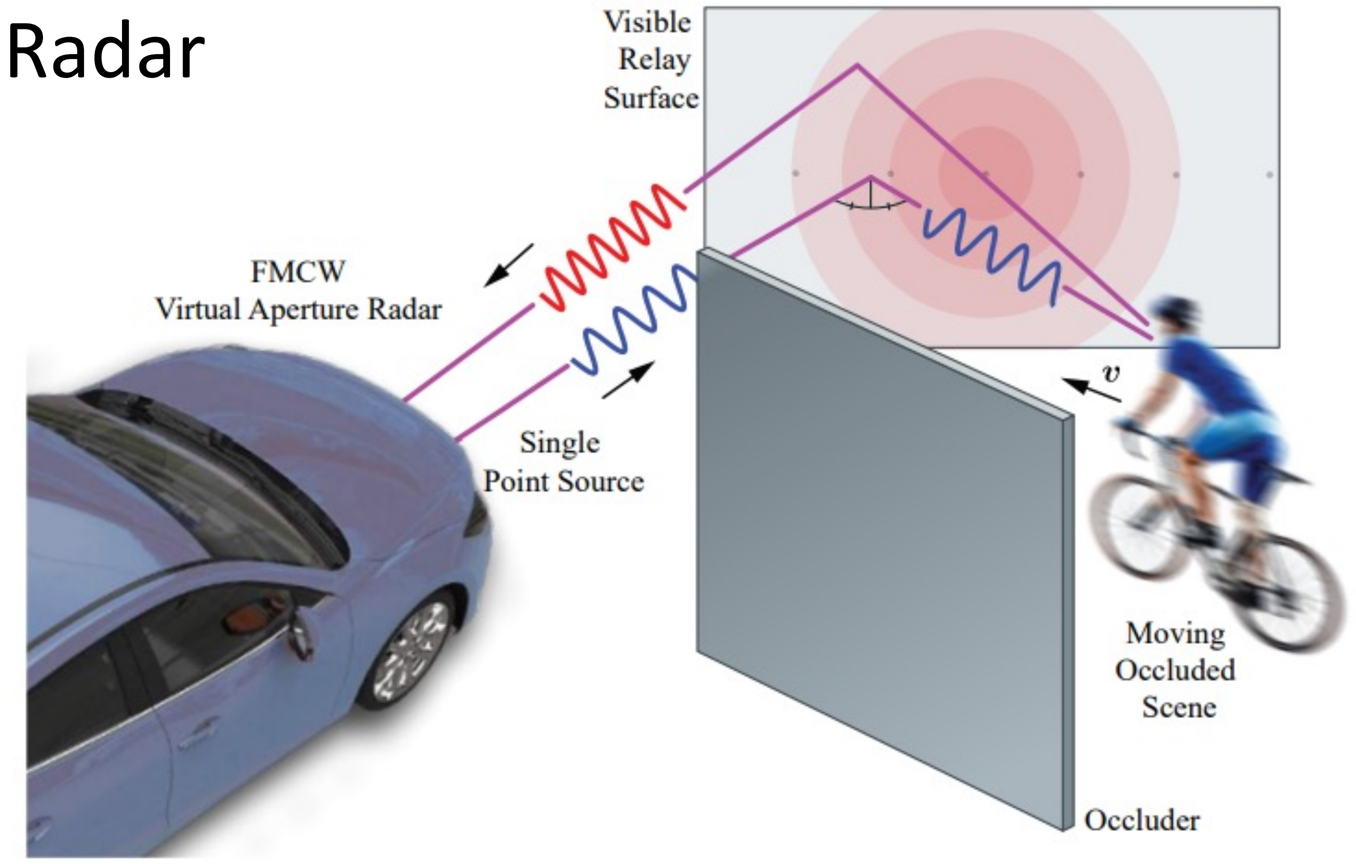
Objects as safety mirrors for navigation

Handling reflections in other imaging modalities





Radar



Zeise, Björn, and Bernardo Wagner. "Temperature Correction and Reflection Removal in Thermal Images using 3D Temperature Mapping." *ICINCO (2)*. 2016.
Scheiner, Nicolas, et al. "Seeing around street corners: Non-line-of-sight detection and tracking in-the-wild using doppler radar." *CVPR*. 2020.