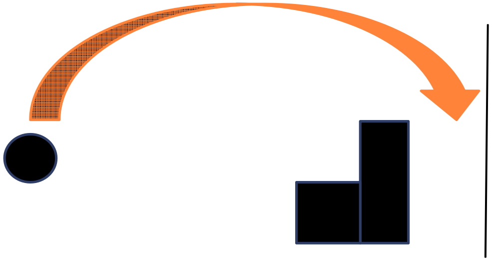
PlatoNeRF: 3D Reconstruction in Plat’s Cave via Single-View Two-Bounce Lidar

Tzofi Klinghoffer, Xiaoyu Xiang, Siddharth Somasundaram, Yuchen Fan, Christian Richardt, Ramesh Raskar, Rakesh Ranjan

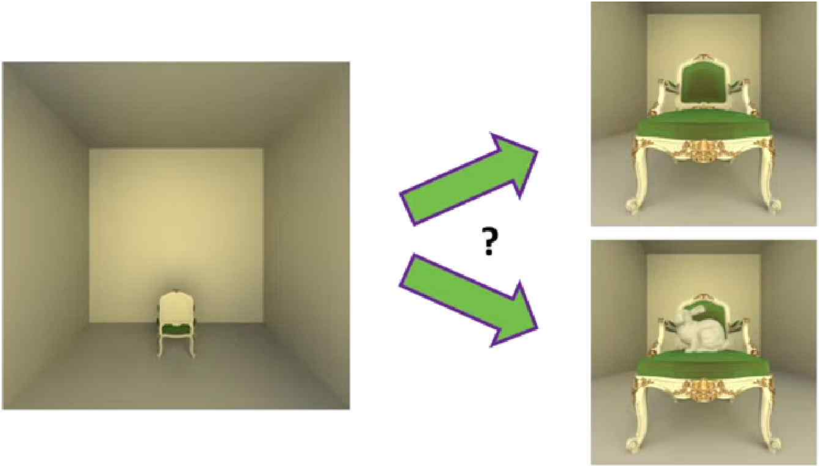
CVPR2024

Recovering 3D scene geometry from a single-view is critical

- for many applications, ranging from autonomous vehicles(AV) to extended reality(XR).
  - ex) Imagine the ball drops and bounces behind your couch in XR



Single-View 3D Reconstruction is Challenging



Existing Methods

- Single-view 3D reconstruction with NERF
  - challening, either rely on data priors or use visual cues(shadows) to infer occluded geometry from a single view
- Diffusion, Transformers,..
  - rely on data priors
  - hallucinate content which may not be physically accurate

Single-Photon Lidar

Single Photon Avalanche Diode

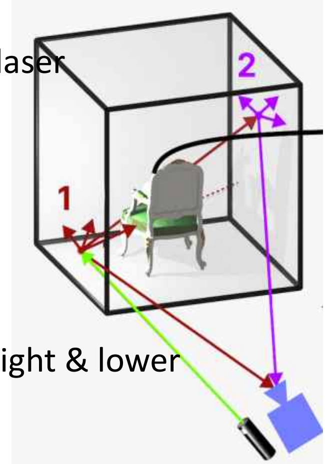
PlatoNeRF : NeRF + Single-Photon Lidar

NeRF ⟺ Single-Photon Lidar

- GOAL: To reconstruct visible & occluded geometry from a single view

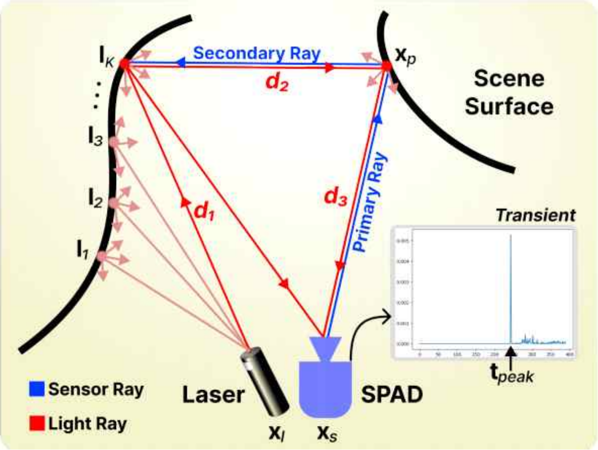
Method Outline

- illuminate individual points in the scene with a laser
- two-bounce light contains infos:
  - scene depth
  - presence of shadows created by the laser
- train NeRF to reconstruct the two-bounce ToF



=> higher accuracy, operate with higher ambient light & lower scene albedo, better generalization

Problem Definition



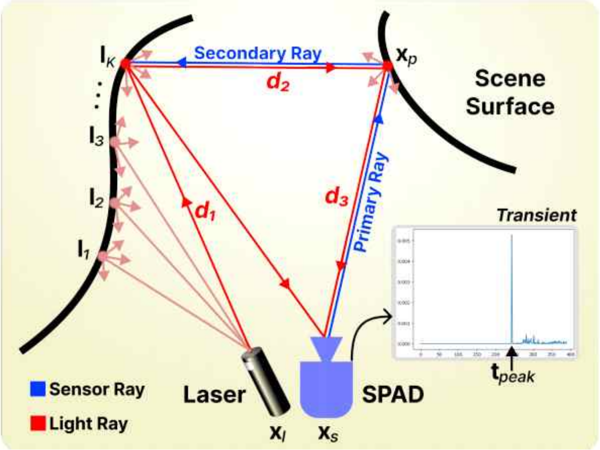
<Experimental Setup>

-Lidar system = SPAD sensor(x_s) + Laser(x_l)

-Laser sequentially points at K different points :$\{l_1, .., l_K\}$

-For each illumination point, an image is captured

Problem Definition



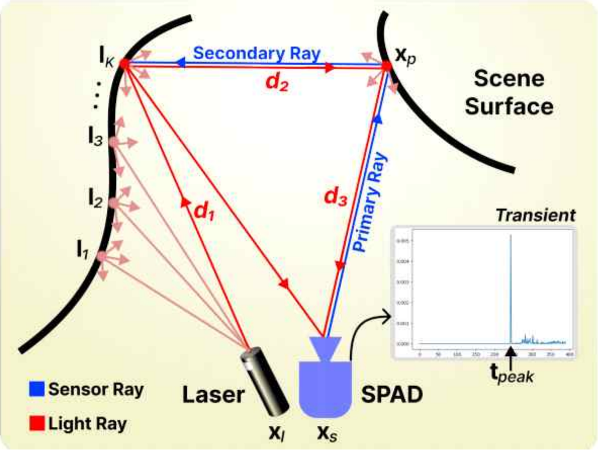
<One Bounce vs Two Bounce>

-x_l -> l -> x_s : one-bounce

-x_l -> l -> x_p -> x_s : two-bounce

-For each image: the pixel observing scene point l measures one-bounce signal, and all other pixels measure two-bounce signals or shadows

Problem Definition



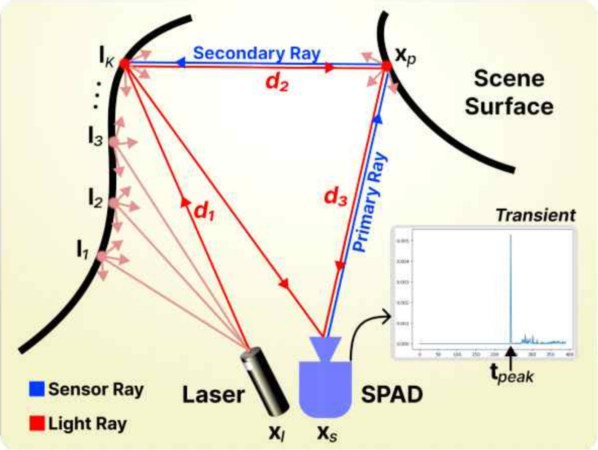
<Transient Measurement>

- Each image is a transient measurement
- measures the amount of light arriving at every pixel at a given time
- histogram of light intensity

$$t_{\text{peak}} = \frac{d}{c} = \frac{d_1 + d_2 + d_3}{c} \quad (1)$$

$$= \frac{\|\mathbf{x}_l - \mathbf{l}\|_2 + \|\mathbf{l} - \mathbf{x}_p\|_2 + \|\mathbf{x}_p - \mathbf{x}_s\|_2}{c}, \quad (2)$$

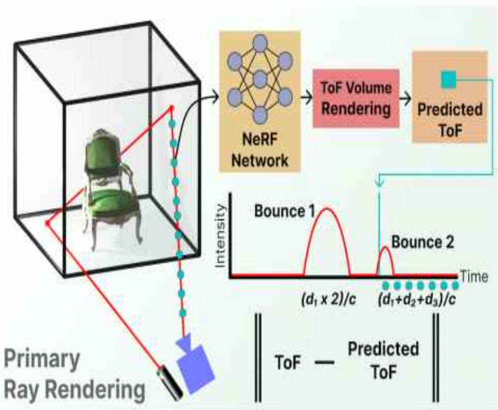
Problem Definition



<Shadow Measurement>

- If x_p lies in shadow, no two-bounce signal will be measured.

Two-Bounce Volumetric Lidar Rendering
<Rendering Primary Rays>

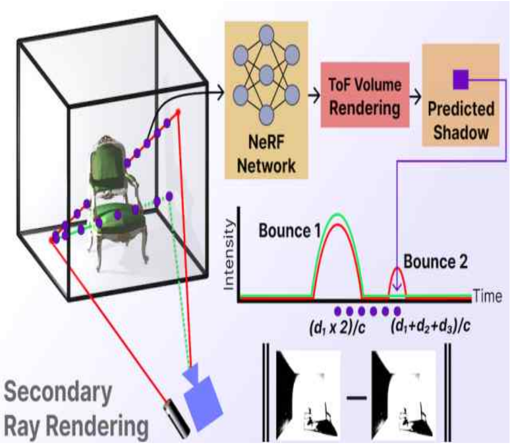


- Goal of rendering along the primary ray : to compute the two bounce tof t_{peak} by determining the depth d3

$$\hat{d}_3(\mathbf{r}_p) = \sum_{i=1}^{N} T_i \alpha_i t_i$$

$$\mathcal{L}_{\text{primary}} = \|t_{\text{peak}} - \hat{t}_{\text{peak}}\|_2^2$$

Two-Bounce Volumetric Lidar Rendering
<Rendering Secondary Rays>



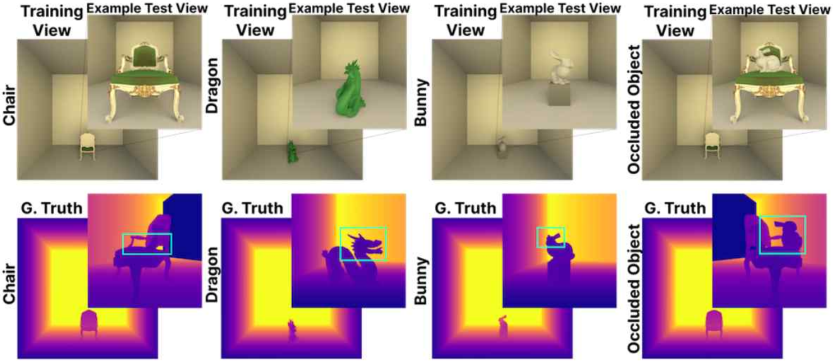
- Goal of rendering secondary ray : to determine if xp lies in shadow or not.

$$p_{\text{shadow}} = \prod_{j=1}^{N-1} (1 - \alpha_j)$$

$$\mathcal{L}_{\text{secondary}} = \| s - \hat{p}_{\text{shadow}} \|_2^2$$

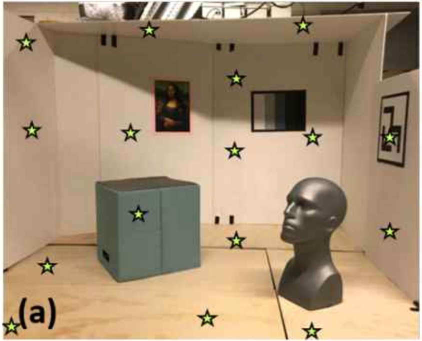
Datasets

- Simulated Datasets



Datasets

- Real Datasets (for validation)



Results

- Baselines
  - Bounce-Flash(BF) Lidar
  - s^3 - NeRF

- Metrics
  - L1 depth error
  - PSNR
  - Chamfer distance

Results

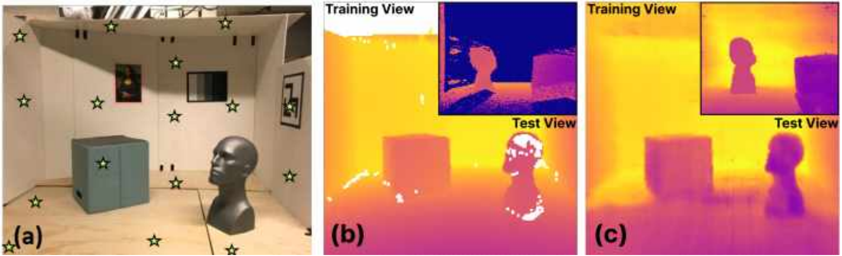
- Simulated Results

Approach	Chair Scene			Dragon Scene			Bunny Scene			Occlusion Scene		
	Train View	Test Views		Train View	Test Views		Train View	Test Views		Train View	Test Views	
	L1 Depth ↓	L1 Depth ↓	PSNR ↑	L1 Depth ↓	L1 Depth ↓	PSNR ↑	L1 Depth ↓	L1 Depth ↓	PSNR ↑	L1 Depth ↓	L1 Depth ↓	PSNR ↑
BF Lidar	0.0348	0.1837	19.63	0.0233	0.1049	22.58	0.0339	0.0660	25.16	0.0341	0.2151	18.96
S^3–NeRF	0.0602	0.1178	22.80	0.0619	0.1042	25.06	0.0633	0.0877	27.67	0.0682	0.1336	22.51
PlatoNeRF	**0.0222**	**0.0862**	**26.58**	**0.0186**	**0.0870**	**28.45**	**0.0191**	**0.0601**	**30.26**	**0.0185**	**0.0836**	**27.33**

Approach	Chamfer (Mean)↓	Std.↓
BF Lidar	0.0465	0.0014
S^3-NeRF	0.4129	0.0021
PlatoNeRF	**0.0280**	**0.0014**

Results

- Real-World Results



Ablation Study

- Ablations on Lidar Sensor

Spatial Resolution		
	L1 Depth (m)	
Downsample	Ours	BF Lidar
128×128	**0.0880**	0.1236
64×64	0.0932	0.1759
32×32	0.1070	0.1799

Temporal Resolution		
	L1 Depth (m)	
Upsample	Ours	BF Lidar
256 ps	**0.0965**	0.2802
512 ps	0.1210	0.3119
1024 ps	0.1833	0.3510

- Ablations on Scene Properties

Ambient Light		
	L1 Depth (m)	
Intensity	Ours	S^3–NeRF
0	0.0862	0.1178
4	**0.0794**	0.3080

Scene Albedo		
	L1 Depth (m)	
Albedo	Ours	S^3–NeRF
$0\times$ less	0.0862	0.1178
$4\times$ less	**0.0859**	0.2152