# PlatoNeRF: 3D Reconstruction in Plato's Cave via Single-View Two-Bounce Lidar

## Supplementary Material

## A. Time & Wavelength Gating in Lidar

As described in the main text, PlatoNeRF (and lidar-based methods) offer fundamental advantages over RGB-based methods in practical scenarios with uncontrolled scene albedos and ambient illumination. Lidars can leverage their picosecond timing resolution for *time gating* to enhance signal-to-background ratio (SBR) of measured shadow images. In addition, unlike RGB sensors, lidar sensors do not require wideband spectral sensitivity. Therefore, ambient illumination that has different wavelength than that of the laser's can be suppressed using *wavelength gating*.

The principle of time gating is illustrated in Fig. 1. A measured lidar signal $i(t)$ can be decomposed into the pulse signal $s(t)$ and (roughly) constant ambient background noise $n(t) = N$. An RGB sensor would integrate over this timing information and measure

$$i = \int_0^T i(t)dt = \int_0^T s(t) + n(t)dt \tag{1}$$

$$= \int_0^T s(t)dt + NT, \tag{2}$$

where $T$ is the length of the transient signal. The measurement $i$ results in a SBR of

$$\text{SBR} = \frac{\int_0^T s^2(t)dt}{N^2 T} \tag{3}$$

On the other hand, a lidar sensor would only use relevant parts of the transient, i.e., around the signal peak. A time-gated lidar would therefore measure

$$i = \int_{T_1}^{T_2} s(t)dt + NT, \tag{4}$$

$$\text{with SBR}_{\text{gated}} = \frac{\int_{T_1}^{T_2} s^2(t)dt}{N^2 W}, \tag{5}$$

where $T_1$ and $T_2$ determine the gated window in the transient signal and $W = T_2 - T_1$ is the window size. Note that the numerator of Eq. (3) is roughly the same as the SBR in Eq. (5) because $s(t) \approx 0$ for $t < T_1$ and $t > T_2$, as shown in Fig. 1(a). Therefore, time gating offers an SNR improvement of $\frac{T}{W}$ over techniques that leverage RGB or intensity signals. Note that the SBR enhancement is inversely proportional to the gated window. We do not account for Poisson noise effects, which, in practice, would introduce trade-offs in determining the window size. Empirical results are plotted in Fig. 1(b)-(d) on the effects of time gating on enhancing contrast in shadow images.

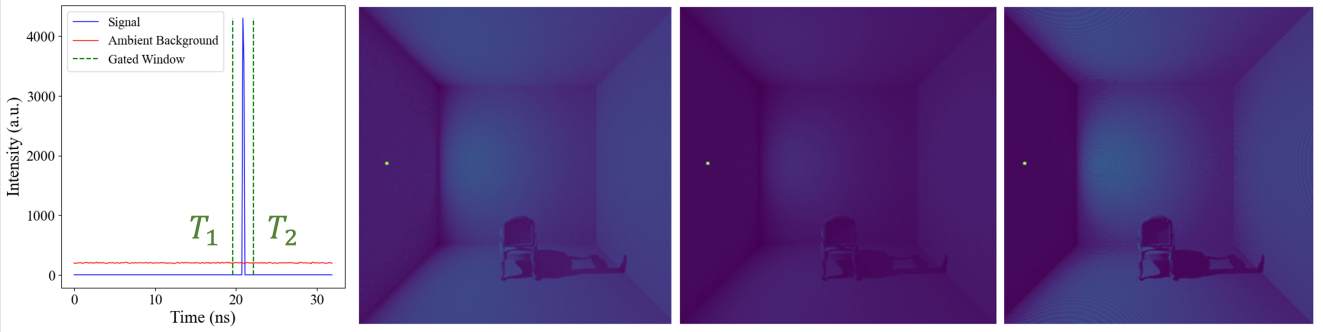
A similar idea can be applied to gate wavelengths. Most of the signal will be concentrated within a narrow spectral range, and all other intensities can be gated out with a narrow-band pass filter, as shown in Fig. 2. This figure plots the emission spectra of an LED light [4] and the gating profile is determined by a 685 nm PicoQuant pulsed laser [3].

## B. Simulated Dataset Details

In this section, we describe the simulated datasets that we render and use to compare our method to past work in more detail. We render four simulated scenes, as described in the main text, with both a lidar and RGB camera in Mitsuba [2]. The lidar data is used to run PlatoNeRF and Bounce Flash (BF) Lidar [1] and the RGB data is used to run S$^3$-NeRF [6]. The same sixteen scene points are illuminated in both the lidar and RGB data. In the lidar data, the sixteen points are illuminated with a laser and, in the RGB data, point light sources are placed at each of the sixteen points. A camera to world transform from OpenGL (x right, y up, z back) to Mitsuba (x left, y up, and z forward) is used to train each method with this data. Ground truth depth for both the train view and 120 test views are provided. A subset of the test view frames are shown in the video results on the project page. All data has been released for use in future work.

**Lidar Data.** The lidar (direct time of flight) data is rendered at $512 \times 512$ spatial resolution with a temporal resolution (bin size) of 128 ps. We simulate a laser by using a spot light source and setting the cutoff angle as $0.2$ and the beam width as $0.1$. To choose the illumination points, we randomly illuminate twenty four points in the scene and then heuristically choose sixteen that maximize diversity.

**RGB Data.** To compare with S$^3$-NeRF, we render each scene with both lidar (to run our method) and RGB (to run S$^3$-NeRF) in Mitsuba. When rendering with RGB, we compute the location of the scene point where the laser first hits the scene and place a point light source at this location. By placing point light sources at the same location as where the laser hits the scene, we ensure the same shadows are cast in the scene in both the lidar and RGB data. RGB images are rendered with max depth to set to 2, ensuring only first-bounce light is rendered, as required by S$^3$-NeRF. Rendered images are gamma corrected prior to training.




(a) Histogram (b) Noiseless (c) Ambient Noise (d) Ambient Noise (Time Gated)

Figure 1. **Time Gating with Lidar.** (a) A transient is plotted at a single pixel. Note that most of the signal (blue) is concentrated within a few timing bins $\sim$ 20 ns. By only gating a window (green) around the signal, most of the noise profile (red) can be suppressed. (b)-(d) Measured intensity images without time gating (b, c) and with time gating (d).

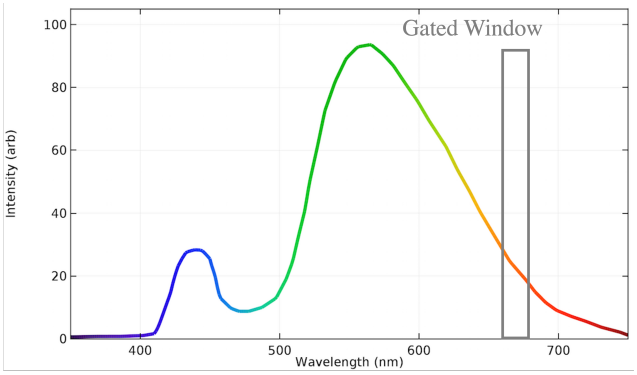



Figure 2. **Wavelength Gating.** Ambient illumination under an LED light is compared to the spectral gating window needed for a spectral window centered at 685 nm. Figure adapted from [4].

## C. Training Details

**Reproducibility.** We have released all data, code, and model checkpoints, along with documentation, to ensure our work is fully reproducible by others. These can be accessed from our project webpage. Simulated data is rendered in Mitsuba world coordinates and PlatoNeRF uses OpenGL camera coordinates. Code for simulating time of flight measurements in Mitsuba is also provided.

**PlatoNeRF.** We train our model for 200k iterations. For the first 25k iterations, only the distance loss is applied, while both the distance and shadow losses are applied thereafter. We use a threshold of 15% on the shadow confidence map (computed as the maximum of the cross-correlation described in Sec 3.3 of the main text) when extracting ground truth shadow masks from the raw lidar measurements. This threshold is used across all experiments, except the ambient light experiment, where we further tune it.

Table 1. **Ablations on Number of Illumination Points.** We study how varying the number of illumination points between two and sixteen impacts PlatoNeRF reconstruction quality.

| Illumination Spots | PlatoNeRF | |
|---|---|---|
| # Spots | L1 (m) | PSNR (dB) |
| 16 | **0.0862** | **26.58** |
| 8 | 0.0912 | 26.33 |
| 4 | 0.1347 | 25.15 |
| 2 | 0.2147 | 21.61 |

**Bounce Flash Lidar.** Bounce Flash (BF) Lidar consists of two steps: (1) estimating visible geometry via constraints on ellipsoidal geometry, and (2) estimating occluded geometry with shadow carving. For each scene, we run a grid search over thresholds for shadow extraction and occupancy probability (applied to the occupancy probabilities predicted from shadow carving) to maximize BF Lidar accuracy.

**$S^3$-NeRF.** We found the default training parameters provided for $S^3$-NeRF work the best on our data. We only modify the light intensity parameter to match our rendered data when training. When training with ambient light, we run a grid search over the ambient light intensity ($amb\_i$) parameter to maximize $S^3$-NeRF reconstruction quality, but find that under a reasonably high ambient area light, $S^3$-NeRF is not able to reconstruct the scene regardless of this parameter.

## D. Extended Ablations

In this section, we add further detail and discussion on the results of our ablations, quantitatively reported in the main text. In addition, we provide further ablation on the impact of non-planar background geometry (Fig. 5), the number of illumination points (Tab. 1), and the shadow mask threshold (Fig. 7) on PlatoNeRF reconstruction.

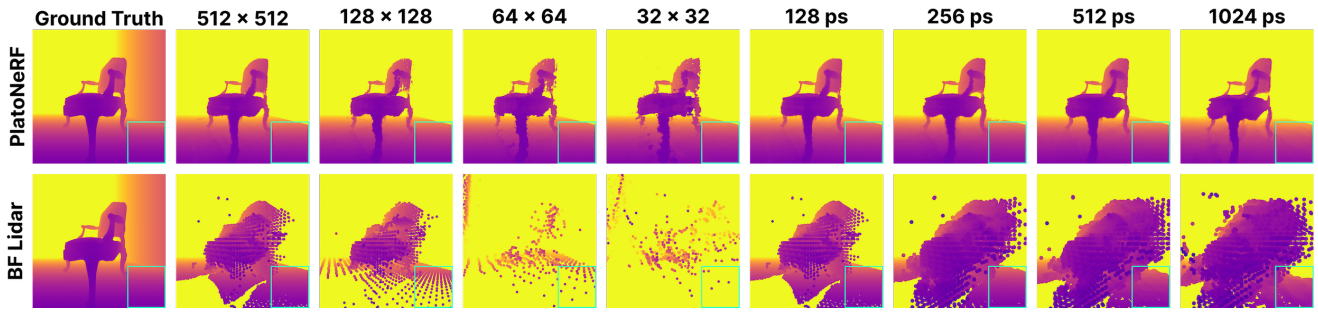



Figure 3. **Spatial- and Temporal-Resolution Ablation.** We compare PlatoNeRF and Bounce Flash (BF) Lidar as spatial- and temporal-resolution is reduced. PlatoNeRF continues to produce smooth geometry in both cases, whereas BF Lidar produces sparse geometry when spatial resolution is reduced and bumpy geometry when temporal resolution is reduced, as highlighted in the area in the green boxes.

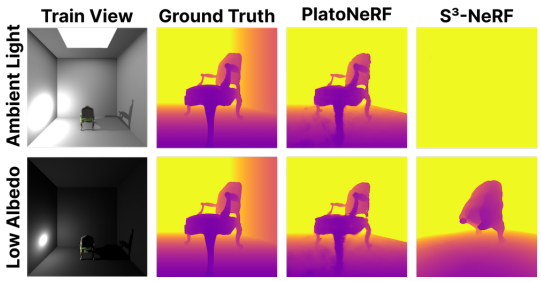



Figure 4. **Ambient Light and Low Albedo Background Ablation.** We compare PlatoNeRF and $S^3$-NeRF when trained on scenes with ambient light or low albedo background. PlatoNeRF is robust to both, whereas the performance of $S^3$-NeRF degrades.

### D.1. Spatial- and Temporal-Resolution

Qualitative results comparing PlatoNeRF and Bounce Flash (BF) Lidar under varying spatial- and temporal-resolutions are shown in Fig. 3. This ablation is important because lidars on consumer devices are often constrained to much lower resolutions than research-grade lidars. Spatial resolution is varied by downsampling the number of pixels, while keeping the field of view of the lidar the same. As spatial resolution is reduced, geometry predicted by BF Lidar becomes sparser. The depth estimation of visible points in the scene remains accurate, but there is no interpolation between these points. The sparsity in visible depth information negatively impacts the shadow carving step of BF Lidar, leading to poor reconstruction of the chair in lower spatial resolution regimes. On the other hand, because PlatoNeRF is able to smoothly interpolate across missing pixels, the resulting reconstruction is significantly more accurate.

Temporal resolution is related to the bin size of the transient (i.e. the amount of time between each lidar measurement). To increase the bin size and thus reduce the temporal resolution of the lidar, we integrate intensities within the bins. For example, when increasing bin size from 128 to 256 ps, we sum intensities for over every two bins. BF Lidar results maintain the shape of the chair (since shadow carving is not significantly affected), but the visible geometry becomes rough and bumpy since the supervision for the depth of each visible pixel is less precise. On the other hand, PlatoNeRF maintains smooth reconstructions.

### D.2. Ambient Light

Qualitative results comparing PlatoNeRF and $S^3$-NeRF reconstructions under ambient light are shown in Fig. 4 (top row). While $S^3$-NeRF is able to model small amounts of ambient light, it fails under realistic amounts of ambient light, in this case, from an added area light. On the other hand, PlatoNeRF is still able to accurately reconstruct the scene with the same ambient light added.

### D.3. Low-Albedo Backgrounds

Qualitative results comparing PlatoNeRF and $S^3$-NeRF reconstructions with a low albedo background are shown in Fig. 4 (bottom row). $S^3$-NeRF is able to accurately reconstruct the visible portion of the scene, but is unable to recover occluded geometry due to worse contrast in the shadow (though it is still discernible to the human eye, as shown in Fig. 4). On the other hand, PlatoNeRF is not significantly affected by scene albedo due to its use of a lidar rather than RGB sensor.

### D.4. Non-Planar Background Geometry

We study the impact of non-planar background geometry on PlatoNeRF. In the main text, we show results on a scene with curved walls, resulting in similar depth L1 and PSNR scores as the same scene with planar walls. This result indicates that PlatoNeRF is robust to non-planar foreground and background geometry. In Fig. 5, we further increase the complexity of the background geometry by adding two objects: a couch and painting. As shown by the extracted shadow in Fig. 5, the additional background objects cause the shadow to be contorted based on the geometry its cast on. PlatoNeRF is still able to accurately reconstruct the full

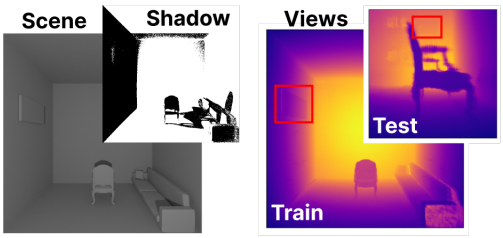



Figure 5. **Non-Planar Background Geometry Ablation.** We provide an additional example of PlatoNeRF reconstruction of a scene with complex, non-planar background geometry. PlatoNeRF accurately reconstructs both the non-planar foreground and background geometery from a single view.

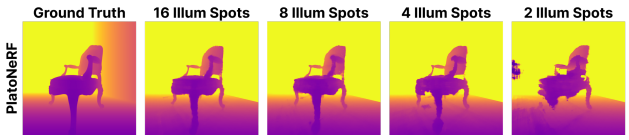



Figure 6. **Illumination Point Ablation.** We ablate the impact of varying the number of illumination points between two and sixteen on PlatoNeRF. While more illumination points improves reconstruction quality, the chair's geometry is still coarsely reconstructed with just two illumination points.

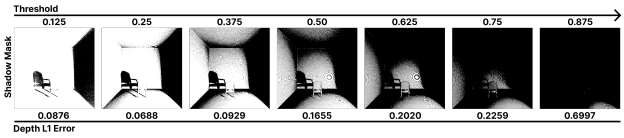



Figure 7. **Shadow Mask Ablation.** We vary the threshold used when creating shadow masks and report the change in L1 depth error across all test viewpoints. Ablation is done on the chair scene.

scene geometry. Certain parts of the scene, such as the wall behind the chair and the self-occluded face of the armrest, will not have two-bounce ToF measurements, resulting in PlatoNeRF interpolating the geometry in those areas.

### D.5. Number of Illumination Points

We further ablate the impact of reducing the number of illumination points used to train PlatoNeRF. In our main experiments, we use sixteen illumination points. We reduce that number to eight, four, and two and report the results in Tab. 1. Qualitative results are shown in Fig. 6. The scene is reconstructed for each number of illumination points, however, as the number is reduced, quality also decreases, as there is less information about occluded regions. When there are only two illumination points, the occluded chair legs are not reconstructed. We note that while we study the number of illumination points, their location is also an important factor in reconstruction quality. As the number of illumination points is reduced, the location of the remaining illumination points becomes increasingly important, i.e. casting shadows with the most relevance and diversity. In these experiments, we randomly choose which illumination points to use.

### D.6. Shadow Mask Threshold

We ablate the impact of the shadow mask threshold on PlatoNeRF reconstruction quality. Shadow masks are generated from the raw time-of-flight data, as described in Sec 3.3 of the main text. To ablate the impact of the probability threshold used to extract shadow masks, we vary it between zero and one in increments of $0.125$. Fig. 7 shows the resulting shadow masks and depth L1 error across test views at each threshold. While a shadow probability threshold of $0.15$ was used to generate the results in the main text, ablation results indicate that a threshold of $0.25$ leads to even better performance. While the approach employed by PlatoNeRF for extracting shadow masks is common in past work, such as BF Lidar, PlatoNeRF is agnostic to the shadow segmentation approach and more sophisticated methods [5] can be extended to transient data and employed in the future.

## References

[1] Connor Henley, Joseph Hollmann, and Ramesh Raskar. Bounce-flash lidar. *IEEE Transactions on Computational Imaging*, 8:411–424, 2022. 1

[2] Wenzel Jakob. Mitsuba renderer, 2010. http://www.mitsuba-renderer.org. 1

[3] PicoQuant. LDH series picosecond pulsed diode laser heads, 2023. 1

[4] Daniel Smith. Calculating the emission spectra from common light sources, 2016. 1, 2

[5] Florin-Alexandru Vasluianu, Tim Seizinger, Radu Timofte, Shuhao Cui, Junshi Huang, Shuman Tian, Mingyuan Fan, Jiaqi Zhang, Li Zhu, Xiaoming Wei, et al. Ntire 2023 image shadow removal challenge report. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 1788–1807, 2023. 4

[6] Wenqi Yang, Guanying Chen, Chaofeng Chen, Zhenfang Chen, and Kwan-Yee K. Wong. $S^3$-NeRF: Neural reflectance field from shading and shadow under a single viewpoint. In *NeurIPS*, 2022. 1