

Blurred LiDAR for Sharper 3D: Robust Handheld 3D Scanning with Diffuse LiDAR and RGB

Nikhil Behari¹, Aaron Young¹, Siddharth Somasundaram¹,
 Tzofi Klinghoffer¹, Akshat Dave¹, Ramesh Raskar¹
¹ Massachusetts Institute of Technology

{nbehari, aryoung, sidsoma, tzofi, ad74, raskar}@mit.edu

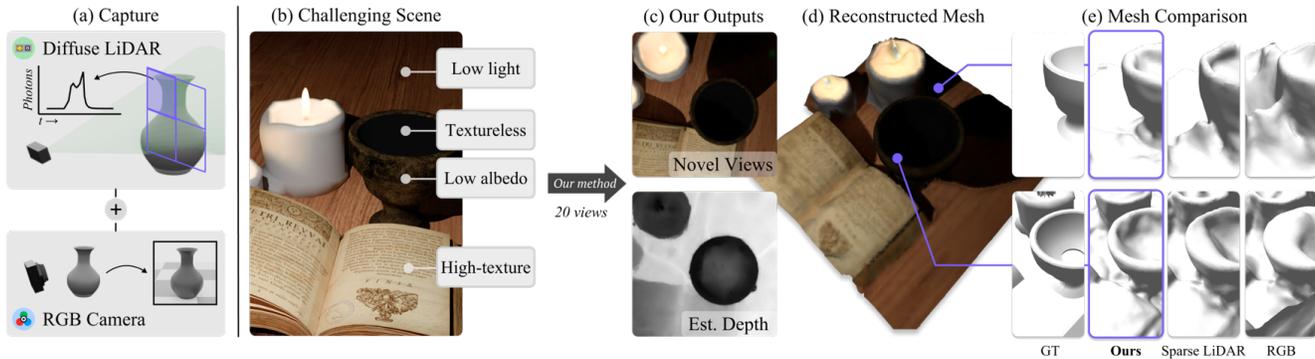


Figure 1. **Blurred LiDAR for Sharper 3D:** We propose leveraging (a) the complementary strengths of RGB with a diffuse (blurred) LiDAR for robust handheld scanning (b) on limited views in challenging low-texture, low-light, low-albedo scenes. Our method dynamically balances RGB and diffuse LiDAR sensor inputs to estimate (c) precise color, depth, and normals, from which we can (d) reconstruct accurate color-3D meshes. Our approach leveraging blurred LiDAR, counterintuitively, can (e) improve 3D reconstruction over conventional LiDAR, which uses sparse spot illumination.

Abstract

3D surface reconstruction is essential across applications of virtual reality, robotics, and mobile scanning. However, RGB-based reconstruction often fails in low-texture, low-light, and low-albedo scenes. Handheld LiDARs, now common on mobile devices, aim to address these challenges by capturing depth information from time-of-flight measurements of a coarse grid of projected dots. Yet, these sparse LiDARs struggle with scene coverage on limited input views, leaving large gaps in depth information. In this work, we propose using an alternative class of “blurred” LiDAR that emits a diffuse flash, greatly improving scene coverage but introducing spatial ambiguity from mixed time-of-flight measurements across a wide field of view. To handle these ambiguities, we propose leveraging the complementary strengths of diffuse LiDAR with RGB. We introduce a Gaussian surfel-based rendering framework with a scene-adaptive loss function that dynamically balances RGB and diffuse LiDAR signals. We demonstrate that, surprisingly, diffuse LiDAR can outperform traditional sparse LiDAR, enabling robust 3D scanning with accurate color and geometry estimation in challenging environments.

Project page (with code+data): nikhilbehari.github.io/bls3d-web/

1. Introduction

3D reconstruction has become increasingly important in applications such as virtual reality, mobile scanning, and robotics. In each such field, it is desirable to have compact, low-bandwidth, and low-cost capture hardware. Yet, simultaneously, 3D reconstruction must be robust to many challenging conditions, such as low-texture and low-albedo objects, and low-lighting scenes. Such qualities are common in real-world settings—for example, when a robot navigates an indoor space with textureless walls or when the Mars Helicopter scans dark, featureless sand dunes [24]. Thus, an ideal setup combines lightweight, compact hardware with accurate, robust reconstruction capabilities.

While recent work has extensively explored 3D reconstruction from RGB, these methods struggle to achieve robust performance in these challenging low texture, albedo, and lighting settings. For instance, recent work in Neural Radiance Fields (NeRF) [29] has enabled high-fidelity novel view synthesis in ideal settings; extensions have also demonstrated accurate reconstruction using signed distance fields [51] and Gaussian Surfels [7] for precise depth, normal and mesh estimation. However, NeRFs and Gaussian Splatting [18] primarily rely on multi-view appearance vari-

ations (e.g., texture) from RGB images; however, these images lack depth information in scenes with low-texture, low-lighting, or low-albedo. In this paper, we propose fusing RGB with an unconventional but widely-available sensor modality—diffuse LiDAR—for robust handheld 3D scanning in these challenging scene conditions.

To improve reconstruction in these challenging settings, LiDAR is often used with RGB to enhance depth estimation in NeRF and other 3D reconstruction techniques. In handheld scanning, this is achieved by projecting a coarse grid of points into the scene and estimating precise depth at these locations. These depth values can directly supervise reconstruction models [8]. As these *sparse* LiDARs are coupled with active illumination, they provide high signal-to-noise ratio (SNR) even in low-SNR conditions, making them effective for compact mobile and robotic scanning setups [5, 35, 43]. However, sparse LiDAR has a key limitation: each measurement captures only a coarse grid of individual depth points, resulting in poor scene coverage and therefore requiring many captures for adequate depth understanding. This trade-off between depth accuracy and sparse coverage can restrict its utility in settings where extensive multi-view captures are impractical.

In this work, we explore the benefits of an alternative diffuse LiDAR that, when paired with RGB, can surprisingly *enhance* 3D reconstruction over sparse LiDAR in challenging conditions such as low-texture, low-light, and low-albedo. Unlike sparse LiDARs, which project individual dots to measure point depth, these LiDARs emit a *diffuse* flash; each diffuse LiDAR pixel then captures a wide field-of-view, resulting in spatially blurred measurements. However, this lower depth precision comes with the benefit of much higher scene coverage. Our key insight is that, despite their spatially blurred measurements, these diffuse LiDARs can be leveraged in an analysis-by-synthesis framework to recover depth, enabling significantly improved scene reconstruction with limited views. A novel supervision strategy is needed, however, to integrate these diffuse LiDARs with RGB for reconstruction; unlike sparse LiDAR, direct point-wise depth supervision is infeasible from spatially blurred depth signals. Thus, we propose a scene-adaptive loss that dynamically balances RGB and diffuse LiDAR signals, prioritizing LiDAR in regions where RGB offers fewer multi-view cues.

Our contributions in this work are as follows:

- We propose leveraging diffuse LiDAR sensors with RGB for robust handheld 3D scanning in challenging low-texture, low-light, low-albedo scenarios.
- We demonstrate that 1) diffuse LiDAR improves spatial coverage over conventional sparse lidar at the cost of spatial ambiguity, and 2) this spatial ambiguity can be resolved with RGB information (Sec. 3).
- We propose a Gaussian-surfel 3D reconstruction tech-

nique and a scene-adaptive loss for balancing the complementary strengths of RGB and diffuse LiDAR (Sec. 4).

- We demonstrate the benefits of diffuse LiDAR over conventional sparse LiDAR through recoverability analysis (Fig. 3), quantitative empirical evaluations (Tab. 1, Tab. 2, Fig. 6), and qualitative real world experiments (Fig. 7).

Scope of this Work. We focus on commercial-grade LiDARs used in handheld settings, which have low spatial and temporal resolution. We consider 3D scanning of static, non-specular objects; future works could explore robustness to different materials and dynamic scenes. We also assume pose estimated from RGB – future work could explore joint pose estimation with LiDAR.

2. Related Work

2.1. 3D Neural Reconstruction

NeRF has emerged as an effective approach for learning volumetric scene representations from 2D images [29]. Building on NeRF, methods for learning implicit surfaces enable surface geometry and normals to be learned [46]. Finally, 3D Gaussian Splatting [18] has been proposed for real-time rendering and, built on top of it, 2D “Surfels” have enabled estimation of high-fidelity 3D surfaces [7].

Neural reconstruction has been widely applied beyond RGB cameras, e.g. sonar [34], CT [11], radar [4], and LiDAR [13, 48, 52]. However methods that leverage time-of-flight sensors [25–27] primarily focus on high-end, scanning laser setups, rather than emerging low-cost, small-form-factor, single-photon sensors. In addition, while fusing RGB with depth when training NeRF has been proposed [8, 41], these methods rely on RGB-based cues from COLMAP [37] or monocular depth estimation, which are limited in the low-texture, albedo, and illumination scenarios that we focus on.

2.2. Time-of-Flight Imaging

LiDARs emit short pulses of light and measure the *time of flight* (ToF) of incident photons. While traditional LiDARs are used to measure only point depth, single-photon avalanche diodes (SPADs) are an increasingly common LiDAR sensor that record *transient* images of photon intensity over time [33]. In addition to depth estimation [14, 16], SPADs have been used for a myriad of applications, such as measuring fluorescence lifetimes [21], non-line-of-sight (NLOS) imaging [12, 19, 36], and seeing through scattering media [9, 44, 49]. Recently, NeRF-based methods have been developed for 3D reconstruction [2, 15, 20, 27, 32] and NLOS imaging [10, 30, 38] from SPADs. Whereas most methods rely on high-cost or laboratory-grade hardware setups, the focus of our work is to demonstrate the utility of low-cost SPADs that use diffuse, rather than point-based, illumination. While recent work has shown promising results

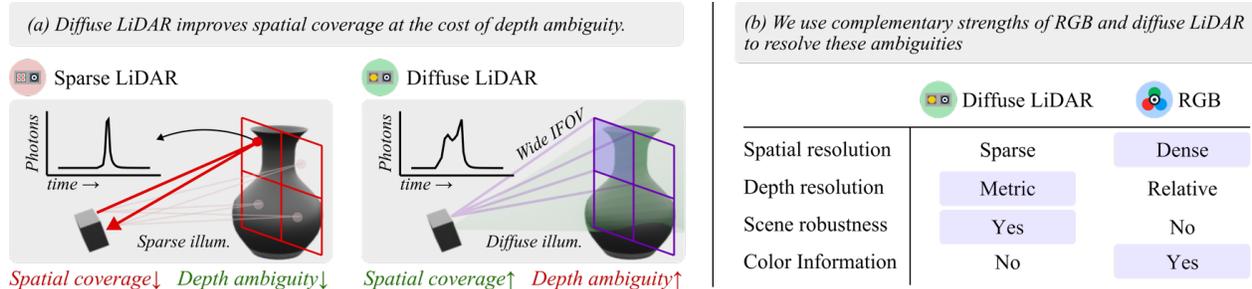


Figure 2. **Diffuse LiDAR + RGB for surface reconstruction.** (a) Sparse (conventional) LiDAR vs Diffuse LiDAR. Sparse LiDAR projects a grid of points which enable precise timing returns corresponding to individual depths; alternatively, diffuse LiDAR projects a diffuse flash illumination and measures the returns over a wide per-pixel instantaneous field-of-view (IFOV), increasing spatial coverage but also ambiguity in inferred depth. (b) Diffuse LiDAR and RGB have complementary strengths; RGB provides dense spatial and color information, while diffuse LiDAR provides coarse, metric depth even under challenging scenarios.

with these sensors [31, 39], none have explored the complementary strengths of low-cost SPADs with RGB sensors. While SPADs and RGB have been used together for single-view depth estimation [22, 28, 42], we focus on multi-view 3D reconstruction across challenging scenes.

3. Leveraging Diffuse LiDAR and RGB

LiDAR is widely used with RGB for surface reconstruction in robotics and handheld scanning. Although RGB provides dense spatial and color information, it struggles with low-light, low-texture, and low-albedo scenes. In these challenging settings, LiDAR can improve reconstruction by providing metric depth at discrete sparse locations; however, with limited input views, conventional LiDAR may lack the scene coverage needed for accurate reconstruction. In this section, we discuss depth supervision using an alternative class of LiDAR, diffuse LiDARs. These LiDARs greatly improve spatial coverage over conventional LiDARs, but also introduce spatial blur into the imaging process. We analytically show how diffuse LiDARs, counterintuitively, can improve recoverability over sparse LiDAR despite their spatial blur. We then offer insights for leveraging the combined strengths of diffuse LiDAR and RGB, particularly for challenging settings where RGB alone may not be sufficient for accurate 3D reconstruction (Fig. 2).

3.1. LiDAR for Improved Handheld 3D Scanning

LiDARs are widely used to improve 3D reconstruction by providing point depth supervision. They operate with a co-located illumination source and a detector, as shown in Fig. 2(a). The laser emits a pulse of light towards a scene point \mathbf{x} . The time it takes for that pulse of light to travel along the path from the laser, to \mathbf{x} , to detector is $t_{\mathbf{x}} = 2 \|\mathbf{x}\| / c$, where c is the speed of light. The ToF measurement can be expressed as

$$i_{\mathbf{x}}(t) = \delta(t - t_{\mathbf{x}}), \quad (1)$$

where $\delta(\cdot)$ is the delta function modeling the time delay

based on the camera distance to \mathbf{x} . From a ToF measurement, we can estimate the $t_{\mathbf{x}}$, from which we can directly estimate the depth of a scene point.

Although conventional LiDARs provide benefits for 3D scanning, they have several limitations, particularly when RGB is also insufficient for reconstruction. First, each pixel images exactly one scene point, which means that a large number of pixels are required to obtain sufficient spatial coverage of the scene. Second, most solid-state LiDAR systems are based on avalanche diodes, which have substantial power requirements. Third, LiDARs based on time-resolved imaging output high-dimensional data structures, which incurs a bandwidth cost. Taken together, these limitations impose a tradeoff for lightweight mobile applications: LiDARs require a large number of pixels to obtain sufficient spatial coverage, but increasing the number of pixels induces a power and bandwidth penalty [53].

3.2. Diffuse LiDAR Enables Better Coverage

We aim to leverage the 3D capabilities of LiDARs, without making large concessions on spatial information, bandwidth, or power. Our key insight is that we can utilize *diffuse LiDARs*, which encode more information into a single pixel measurement by (1) increasing the pixel’s instantaneous field of view (IFOV) and (2) using a diffused light source, as shown in Fig. 2(b). In this case, the pixel transient measurement can be expressed as

$$i(t) = \int_{\mathbf{x} \in \Omega} i_{\mathbf{x}}(t) d\mathbf{x}, \quad (2)$$

where Ω is the set of 3D surface points \mathbf{x} in the pixel’s IFOV.

The key benefit of this capture setup is that each sensor pixel can capture information from multiple scene points *simultaneously*. Therefore, fewer pixels are needed to capture information from all scene points. The challenge now lies in separating light coming from different scene points. In a conventional camera, large IFOV typically results in pixel blur that cannot be recovered. However, a LiDAR can mea-

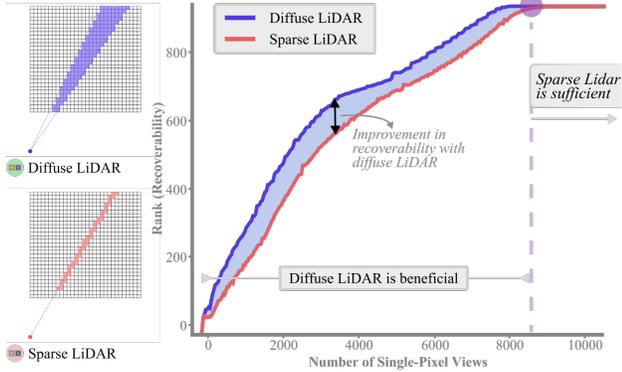


Figure 3. **Improved recoverability with diffuse LiDAR when input views are limited.** Full rank is 900 in our analysis simulation. Diffuse LiDAR has greater voxel coverage than conventional sparse LiDAR; this greater coverage can improve rank, and thereby recoverability, when using a limited number of input views. We consider this specific limited-view domain in this work. As the number of input views increases, sparse LiDAR can eventually provide sufficient coverage for scene recoverability.

sure the ToF of incident light, providing a key source of information in disambiguating light contributions from different scene points. A diffuse LiDAR’s ability to decompose light contributions from different scene points is partially determined by the number of views [40], which motivates an analysis on when diffuse LiDAR is beneficial.

Analysis of Scene Recoverability. To quantify the benefits of diffuse LiDAR, we perform an analysis to compare diffuse and sparse LiDARs. For computational and conceptual simplicity, we consider a 2D scene without loss of generality. Our analysis consists of an approximate linear model (details in Supplementary) to map voxelized scene geometry \mathbf{x} to LiDAR measurements \mathbf{y} as described by: $\mathbf{y} = \mathbf{A}\mathbf{x}$. The size and structure of \mathbf{A} is a function of the number of views and the IFOV of the LiDAR. As a result, by computing matrix properties (i.e. rank) of \mathbf{A} , we can quantify the recoverability of a scene under different LiDAR capture configurations. In Fig. 3, we plot the rank of \mathbf{A} for a diffuse LiDAR (high IFOV) and a sparse LiDAR (low IFOV) as we increase the number of views of the scene. We observe that the rank increases with number of views for both configurations, but the diffuse LiDAR offers an improvement in recoverability due to its spatial coverage of the scene. In spite of the improvement over sparse LiDAR, diffuse LiDAR alone is still insufficient for scene recovery when using fewer views and low spatial resolution, which motivates our RGB-SPAD fusion in Sec. 4.

3.3. Balancing Diffuse LiDAR & RGB

As discussed above, diffuse LiDAR has potential to improve depth recoverability over sparse LiDAR in limited-view scenarios; however, it also introduces ambiguity due

to spatially mixed measurements encoded in time. To address these ambiguities, we leverage the complementary strengths of RGB sensors, which provide dense spatial and color data but struggle in low-light, low-texture conditions (see Fig. 2b). Our method aims to fuse both sensors to handle scenes ranging from ideal to challenging. Yet, unlike sparse LiDAR, diffuse LiDAR captures depth across a wide field of view, making direct point-wise depth supervision used in sparse LiDAR supervision infeasible. Thus, a new rendering approach is required to effectively combine these complementary signals for accurate 3D reconstruction.

4. Reconstruction with Diffuse LiDAR & RGB

In this section, we describe an inverse rendering procedure that recovers surface geometry and texture from multi-view RGB and diffuse LiDAR measurements. The RGB and LiDAR sensors are rigidly mounted together with known relative pose between the two sensors. We use COLMAP [37] to obtain the camera intrinsics and per-frame extrinsics.

Our analysis-by-synthesis reconstruction method is based on a Gaussian surfel scene representation [7] and a differentiable RGB-transient rendering algorithm. By using a scene-adaptive loss function, we can dynamically leverage RGB signal in high-texture, high SNR settings, and gradually de-emphasize RGB and prioritize diffuse LiDAR signals in low-texture and low-SNR regions.

4.1. Surfel-Based RGB-SPAD Rendering

Gaussian Surfels. We represent the 3D scene as a composition of Gaussian surfel primitives [7]. Surfels are similar to the 3D Gaussian representation used in GS [18], with the key difference being that the 3D Gaussian is flattened into a 2D Gaussian. These 2D Gaussians can be oriented to align with the surface of a 3D object.

The shape and location of the Gaussian can be parametrized by a covariance matrix $\Sigma \in \mathbb{R}^{3 \times 3}$ and mean vector $\mathbf{x}_i \in \mathbb{R}^3$ respectively. The resulting 3D Gaussian can be expressed as

$$G(\mathbf{x}; \mathbf{x}_i, \Sigma_i) = e^{-\frac{1}{2}(\mathbf{x}-\mathbf{x}_i)^\top \Sigma_i^{-1}(\mathbf{x}-\mathbf{x}_i)}. \quad (3)$$

Covariance matrices, by definition, are positive semi-definite and can be interpreted as describing the orientation of an ellipsoid. To enforce both of these properties, the covariance matrix can be decomposed as

$$\Sigma = (\mathbf{R}\mathbf{S})(\mathbf{R}\mathbf{S})^\top, \quad (4)$$

where $\mathbf{R} \in \mathbb{R}^{3 \times 3}$ is a rotation matrix and $\mathbf{S} \in \mathbb{R}^{3 \times 3}$ is a diagonal scaling matrix describing the scaling of the principal axes. \mathbf{R} can be analytically expressed in terms of a quaternion $\mathbf{r} \in \mathbb{R}^4$ and the scaling matrix can be expressed as $\mathbf{S} = \text{diag}(\mathbf{s})$, where $\mathbf{s} \in \mathbb{R}^3$ is the scaling vector. To flatten the 3D Gaussian to a 2D surfel, we set $\mathbf{s} = [s_1, s_2, 0]^\top$.

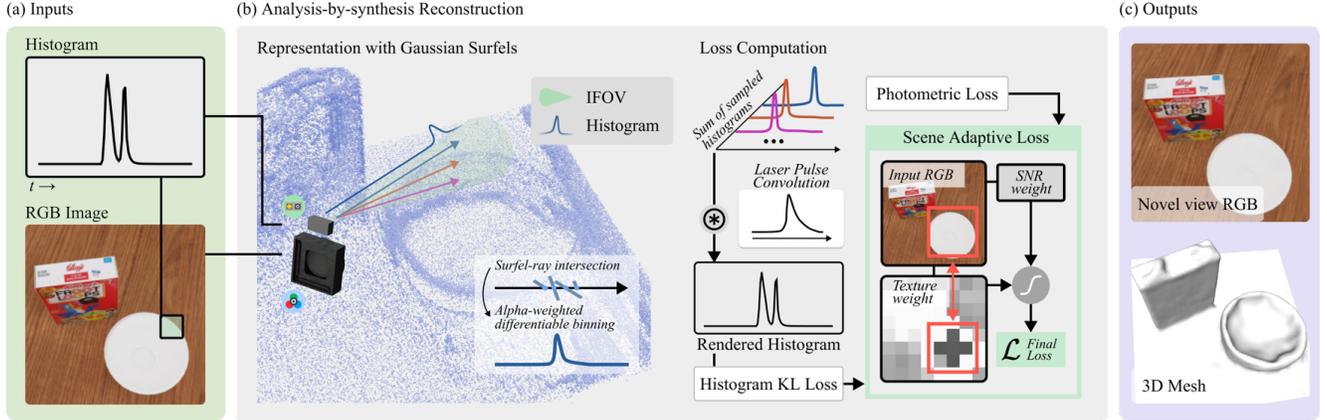


Figure 4. **Inverse rendering with RGB and diffuse LiDAR.** We consider a compact hardware setting with co-located RGB camera and diffuse LiDAR. At each view, we capture an (a) RGB image and coarse (8×8) histograms, for which each pixel contains mixed signal from a wide IFOV ω . We perform (b) analysis-by-synthesis reconstruction using Gaussian surfels, sampling rays within each pixel IFOV and rendering transients with alpha-weighted differentiable binning. Loss signal from RGB and transient inputs are balanced dynamically with a scene-adaptive loss producing (c) high-fidelity RGB and depth/normals for accurate mesh reconstruction.

The resulting parametrization of the Gaussian kernels $\{\mathbf{x}_i, \mathbf{r}_i, \mathbf{s}_i, o_i, \mathcal{C}_i\}$ has five degrees of freedom: the mean $\mathbf{x}_i \in \mathbb{R}^3$ is the 3D location, $\mathbf{r}_i \in \mathbb{R}^4$ is the rotation expressed as a quaternion, $\mathbf{s}_i \in \mathbb{R}^3$ is scaling, $o_i \in \mathbb{R}$ is the opacity, and $\mathcal{C}_i \in \mathbb{R}^k$ are spherical harmonic coefficients.

Surfel Rasterization. We follow prior works in volumetric rendering to rasterize surfel primitives to the image plane [7, 18, 54]. The projection of a 3D Gaussian onto a 2D image plane can be approximated as a 2D Gaussian with mean $\mathbf{u}_i \in \mathbb{R}^2$ and covariance $\Sigma'_i \in \mathbb{R}^{2 \times 2}$

$$G'(\mathbf{u}; \mathbf{u}_i, \Sigma'_i) = e^{-\frac{1}{2}(\mathbf{u}-\mathbf{u}_i)^\top \Sigma'_i{}^{-1}(\mathbf{u}-\mathbf{u}_i)}. \quad (5)$$

Given a view transformation $\mathbf{W} \in \mathbb{R}^{3 \times 3}$ and the Jacobian of the affine approximation of the projective transformation $\mathbf{J} \in \mathbb{R}^{3 \times 3}$, the covariance in image space is approximately

$$\Sigma'_i = (\mathbf{J}\mathbf{W}\Sigma_i\mathbf{W}^\top\mathbf{J}^\top)[:, :2]. \quad (6)$$

RGB Rendering. The color $\tilde{\mathbf{C}}$ at a pixel \mathbf{u} can be computed by sampling all n surfels along the camera ray corresponding to the pixel. These samples can then be integrated to render a color using the alpha compositing equation

$$\tilde{\mathbf{C}} = \sum_{i=0}^n T_i \alpha_i \mathbf{c}_i, \quad T_i = \prod_{j=0}^{i-1} (1 - \alpha_j), \quad (7)$$

where $\alpha_i = G'(\mathbf{u}; \mathbf{u}_i, \Sigma'_i) o_i$ is the alpha-blending weight and \mathbf{c}_i is the color computed using \mathcal{C}_i [18].

Surface Rendering. The depth \tilde{D} at a pixel can be estimated by computing the expected distance along the ray

$$\tilde{D} = \frac{1}{1 - T_{n+1}} \sum_{i=0}^n T_i \alpha_i d_i(\mathbf{u}), \quad (8)$$

where $d_i(\mathbf{u})$ is the distance of the i th Gaussian from the camera along pixel direction \mathbf{u} . A key benefit of the surfel representation is that depth can be estimated precisely for any 2D surface in 3D space [7]. The resulting intersection between the pixel ray and i th surfel can be estimated as

$$d_i(\mathbf{u}) = d_i(\mathbf{u}_i) + (\mathbf{W}\mathbf{R}_i)[2, :] \mathbf{J}_{pr}^{-1}(\mathbf{u} - \mathbf{u}_i), \quad (9)$$

where \mathbf{J}_{pr}^{-1} is the Jacobian of inverse mapping, and $(\mathbf{W}\mathbf{R}_i)$ transforms the rotation matrix to the camera space. The surface normal $\tilde{\mathbf{N}}$ can be computed by replacing $d_i(\mathbf{u})$ in Eq. (8) with the z direction of the rotation matrix $\mathbf{R}_i[:, 2]$.

Histogram Rasterization. We now synthesize ToF measurements from surfel primitives. A ToF measurement $i[x, y, t] \in \mathbb{R}^{N_x \times N_y \times N_t}$ is a 3D data structure, where N_x and N_y are the number of pixels in the x and y direction, and N_t is the number of timing bins. Each pixel measurement is discretized into a histogram with N_t bins, where the bin width is the timing resolution Δt of the LiDAR. In our problem setting, we use larger pixels, which introduces spatial blur. As a result, each pixel measures radiance along a cone, rather than along an individual ray.

In order to render the histogram measurement $i[t]$ at a pixel $\mathbf{u} = [x, y]$, we first sample a set of rays \mathcal{R} within the pixel cone, and compute the set of all surfels \mathcal{S}_r intersecting each ray $\mathbf{r} \in \mathcal{R}$. For each surfel $s \in \mathcal{S}_r$, we assign its radiance contribution to a temporal histogram bin index β_s

$$\beta_s = \left\lfloor \frac{2d_s}{\Delta t} \right\rfloor, \quad (10)$$

where d_s is the depth to the surfel s that can be computed using Eq. (9). Computation of the bin index can be interpreted as a temporal quantization of Eq. (1). In practice,

Table 1. **Quantitative evaluation of low-texture scene reconstruction on rendered scenes.** We compare our method to: 1) Gaussian surfels (RGB Only), 2) Surfels with RGB and monocular depth, and 3) Surfels with RGB and sparse LiDAR. We evaluate for 10 input images on four scenes, each with four texture variations. We obtain consistent improvement in depth (D.MAE ↓) and normal (N.MAE ↓) estimation across texture variations, improving robustness over conventional LiDAR in low-texture, limited-view settings.

Method	Blender				Chair				Hotdog				Lego			
	PSNR ↑	SSIM ↑	D.MAE ↓	N.MAE ↓	PSNR ↑	SSIM ↑	D.MAE ↓	N.MAE ↓	PSNR ↑	SSIM ↑	D.MAE ↓	N.MAE ↓	PSNR ↑	SSIM ↑	D.MAE ↓	N.MAE ↓
(a) Full Texture Datasets: Scenes with full texture on object and ground plane.																
Surfels RGB Only	28.92	0.848	0.081	34.27	28.68	0.852	0.062	28.95	30.85	0.859	0.068	36.12	27.67	0.794	0.074	38.40
Surfels RGB w/ Mono. Depth	27.77	0.876	0.033	17.50	27.94	0.866	0.028	16.06	28.97	0.886	0.031	20.75	27.42	0.818	0.032	21.64
Surfels RGB w/ Sparse LiDAR	24.68	0.839	0.058	24.74	26.06	0.840	0.057	23.44	27.85	0.856	0.054	29.49	26.12	0.801	0.058	30.63
Ours (RGB + Diffuse LiDAR)	30.67	0.881	0.025	19.17	30.25	0.891	0.017	18.05	30.14	0.894	0.016	24.62	28.39	0.845	0.025	25.27
(b) Textured Object Datasets: Scenes with textured objects on completely textureless ground planes.																
Surfels RGB Only	25.44	0.944	0.226	50.88	22.89	0.928	0.474	59.75	37.02	0.962	0.104	41.49	26.06	0.878	0.176	54.25
Surfels RGB w/ Mono. Depth	27.66	0.952	0.129	28.91	21.99	0.932	0.452	55.76	36.04	0.961	0.065	26.33	27.65	0.888	0.052	23.67
Surfels RGB w/ Sparse LiDAR	26.96	0.937	0.105	36.91	25.64	0.936	0.115	38.68	27.03	0.932	0.080	34.98	25.77	0.879	0.093	39.99
Ours (RGB + Diffuse LiDAR)	25.08	0.945	0.033	10.47	32.18	0.970	0.030	7.33	35.62	0.963	0.024	16.34	30.00	0.924	0.024	16.94
(c) Textured Plane Datasets: Scenes with completely textureless objects on textured ground planes.																
Surfels RGB Only	26.34	0.848	0.090	37.53	23.69	0.841	0.089	36.90	29.30	0.879	0.101	44.77	25.86	0.858	0.091	42.49
Surfels RGB w/ Mono. Depth	26.34	0.883	0.036	19.27	23.76	0.867	0.049	20.61	30.41	0.921	0.050	27.08	24.95	0.889	0.055	28.24
Surfels RGB w/ Sparse LiDAR	25.36	0.867	0.057	23.95	24.92	0.842	0.063	31.37	28.33	0.894	0.045	29.28	24.69	0.861	0.067	31.06
Ours (RGB + Diffuse LiDAR)	23.72	0.820	0.045	21.13	25.70	0.841	0.037	22.80	28.97	0.858	0.034	22.32	26.60	0.850	0.046	25.69
(d) No Texture Datasets: Completely textureless scenes.																
Surfels RGB Only	-	-	0.885	36.74	-	-	0.904	42.14	-	-	0.850	63.31	-	-	0.898	51.27
Surfels RGB w/ Mono. Depth	-	-	0.716	67.01	-	-	0.793	60.23	-	-	0.911	46.24	-	-	0.835	63.31
Surfels RGB w/ Sparse LiDAR	-	-	0.125	40.86	-	-	0.107	37.29	-	-	0.111	43.54	-	-	0.106	41.63
Ours (RGB + Diffuse LiDAR)	-	-	0.045	15.68	-	-	0.045	13.55	-	-	0.041	16.35	-	-	0.042	14.54

we find that performing soft histogramming similar to prior work [27] improves gradient flow during rendering. The idea is to distribute the surfel contribution over bins β_s and $\beta_s + 1$, with respective weights

$$w_{\beta_{s,1}} = 1 - \left(\frac{d_s}{\Delta t} - \beta_s \right), w_{\beta_{s,2}} = 1 - w_{\beta_{s,1}}. \quad (11)$$

The final rendered histogram $i[t]$, for each bin b at a pixel, is constructed by summing the opacity-weighted contributions across all surfels:

$$i[t] = \sum_{\mathbf{r} \in \mathcal{R}} \sum_{s \in \mathcal{S}_r} o_s \cdot (w_{\beta_{s,1}} \delta[t - \beta_s] + w_{\beta_{s,2}} \delta[t - \beta_s - 1]). \quad (12)$$

4.2. Scene-Adaptive Loss Function

The key benefit of sensor fusion in this work is that the shortcomings of one sensor can be overcome by relying on information from the other (Fig. 2). In order to adaptively do so during optimization, we quantify the “usefulness” of information available in the RGB image. The loss function will use this quantification of usefulness to adaptively determine whether to rely on the RGB or LiDAR image more.

Quantifying Usefulness of RGB. There are two key aspects of usefulness that we consider in the RGB images: texture and signal-to-noise ratio (SNR). Intuitively, RGB images contain more information when they contain textured scenes with sufficient SNR. We divide the RGB images into a set of patches \mathcal{P} such that each patch p corresponds to the IFOV of exactly one diffuse LiDAR pixel. We then compute per-patch values for SNR and texture. The SNR of an image patch p can be computed as $w_{snr} = \mu_p / \sigma_p^2$, where μ_p and σ_p^2 are the mean and variance of the pixel intensities in the patch. The amount of texture within a patch can be quantified as the variance $w_{texture} = \sigma_p^2$. We use these weights as input to a sigmoid function

$$w_p(x, \epsilon, k) = \frac{1}{1 + e^{-k(x-\epsilon)}}, \quad (13)$$

to determine the usefulness of RGB. We set input $x = w_{texture}$, translation parameter to be $\epsilon = aw_{snr} + b$, and a , b , and steepness parameter k to be hyperparameters. This sigmoid weight helps us determine the weighting of the RGB patch, where high variance leads to higher weighting and low SNR leads to lower weighting.

RGB Loss. The RGB loss consists of two terms: data fidelity (L1) and perceptual similarity (SSIM) [47]

$$\begin{aligned} \mathcal{L}_{\text{RGB}} = & (1 - \lambda_{\text{SSIM}}) \cdot \sum_{p \in \mathcal{P}} w_p \|\tilde{\mathbf{C}}_p - \mathbf{C}_{p,gt}\|_1 \\ & + \lambda_{\text{SSIM}} \cdot (1 - \text{SSIM}(\tilde{\mathbf{C}}, \mathbf{C}_{gt})) \cdot \frac{1}{|\mathcal{P}|} \sum_{p \in \mathcal{P}} w_p. \end{aligned} \quad (14)$$

\mathbf{C} is the full image, \mathbf{C}_p is the patch image, and λ_{SSIM} is a hyper-parameter weight [7]. The L1 term is computed patch-wise, where each patch is weighted differently. The SSIM term is computed for the entire image and is weighted by the average weight across all image patches.

Histogram Loss. The histogram loss is the KL divergence between the normalized rendered histogram and ground truth (gt) histogram

$$\mathcal{L}_{\text{hist}} = \sum_{p \in \mathcal{P}} (1 - w_p) \cdot \text{KL}(i_p[t] \parallel i_{p,gt}[t]), \quad (15)$$

where $i_p[t]$ is the histogram corresponding to RGB patch \mathcal{P} .

Combined Loss Function. The three terms in the final loss function are the RGB loss \mathcal{L}_{rgb} , the LiDAR loss $\mathcal{L}_{\text{hist}}$, and a depth-normal consistency regularization \mathcal{L}_{reg} [7]. Putting these terms together results in the combined loss

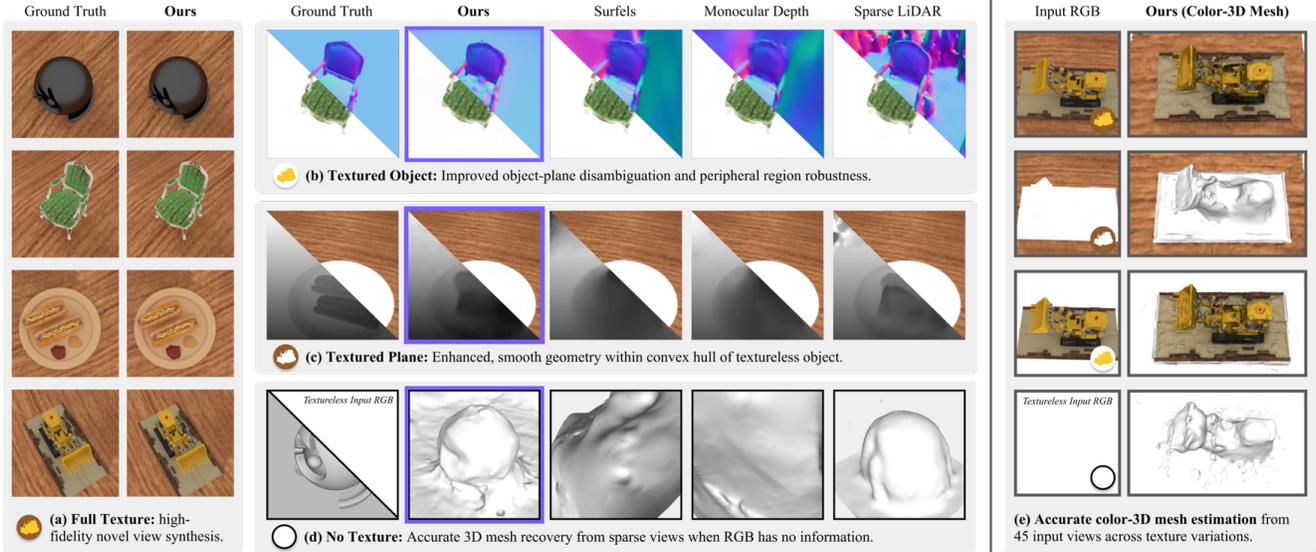


Figure 5. **Qualitative comparisons on rendered scenes with varying texture.** We enable (a) accurate RGB novel-view synthesis on full-texture scenes, where RGB may be prioritized by our adaptive loss. When (b) textured objects are on textureless planes, we enable greater object-plane separation and peripheral region robustness; on textureless objects, we (c) enhance geometry estimation within the object convex hull visible from RGB. We also (d) improve over sparse LiDAR in scenes without RGB signal, where our adaptive loss may weight diffuse LiDAR more heavily. We enable (e) accurate color-mesh estimation across a wide range of texture and object variations.

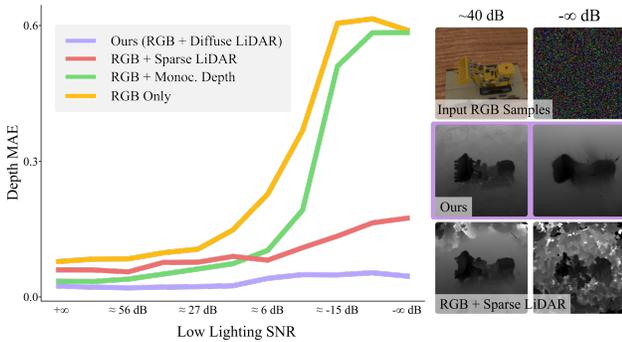


Figure 6. **Robust depth (MAE ↓) of our method in low lighting.** We simulate low lighting with added Gaussian noise; our *scene-adaptive loss weighting* can be used to rely on diffuse LiDAR inputs more heavily as RGB input SNR increases, enabling robust depth estimation across a wide scale of low lighting noise.

$$\mathcal{L} = \mathcal{L}_{rgb} + \mathcal{L}_{hist} + \mathcal{L}_{reg}. \quad (16)$$

5. Experiments

Comparisons. We compare our method against three baseline techniques: 1) Gaussian surfels rendering [7] using only RGB inputs; 2) Gaussian surfels with monocular depth priors [50], which [7] suggest can aid in surface reconstruction but that we expect to fail on low-texture scenes; 3) Gaussian surfels with sparse LiDAR depth loss. For the sparse LiDAR baseline, we apply the standard L1 depth loss commonly used in LiDAR-supervised NeRF methods ($\mathcal{L}_{depth}(\mathbf{r}) = \sum \|D_{est}(\mathbf{r}) - D_{true}(\mathbf{r})\|_1$).

5.1. Simulated Experiments

Experimental Setup. We simulate a diffuse LiDAR with an 8×8 pixel array, using approximate horizontal/vertical IFOV of 4.9° , maximum measurement distance 1.5 meters, and 40 picosecond bin resolution. We compare to sparse LiDAR-supervised Surfels using 8×8 points located at the center of each diffuse LiDAR pixel zone. We render RGB and 8×8 diffuse LiDAR histograms for four scenes following the rendering procedure described in Sec. 4: Blender Ball, Chair, Hotdog, Lego. To evaluate robustness to low and no-texture scenes, we consider four dataset variations: 1) Full Texture; 2) Textured Object; 3) Textured Plane; and 4) No Texture. Examples of these variations are shown for Lego in Fig. 5e. We consider a view-limited regime with 10 training and 10 test captures.

Results. We present quantitative and qualitative results on simulated varying-texture scenes in Tab. 1 and Fig. 5. Our technique effectively combines diffuse LiDAR and RGB to enhance color and geometry compared to baselines, in particular over RGB with sparse LiDAR. In full-texture scenes (Tab. 1a & Fig. 5a), our scene-adaptive loss enables higher reliance on RGB, with diffuse LiDAR aiding depth estimation in peripheral regions with limited views (e.g. distant planes at oblique angles). In mixed-texture (Tab. 1b-c & Fig. 5b-c) scenes, our scene-adaptive loss effectively prioritizes diffuse LiDAR in patches where RGB has less signal, yielding both improved ground plane separation and detail within textureless object convex hulls. In no-texture scenes (Tab. 1d & Fig. 5d), we significantly outperform baselines, and enable sharper reconstructions and object-plane separa-

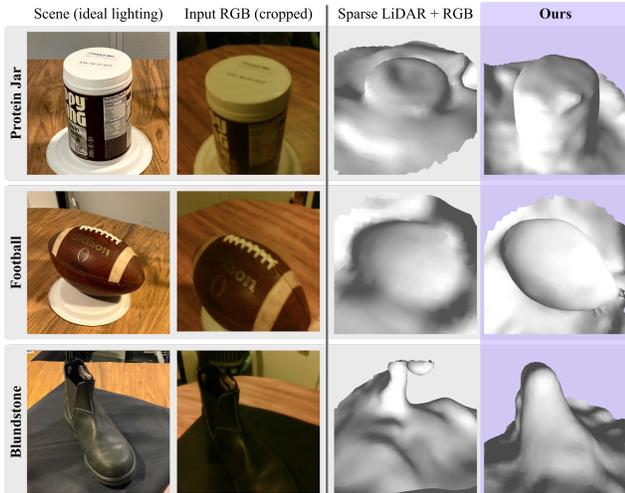


Figure 7. **Qualitative comparisons of real data captures.** We improve mesh reconstruction in challenging real-world scenes on few (90) inputs. RGB with sparse LiDAR fails to separate object and plane due to low albedo and poor spatial coverage, while our diffuse LiDAR improves boundary and geometry estimation. In very challenging settings (Blundstone), our method fails to separate object and plane, but nevertheless provides improved shape.

tion over RGB with sparse LiDAR. Additional quantitative baselines are provided in the supplement.

We present simulated results for low-lighting robustness in Fig. 6. We simulate low lighting with added Gaussian noise from $+\infty$ dB (no noise) to $-\infty$ dB (complete noise). At high SNR, our scene-adaptive loss can rely largely on RGB, resulting in only subtle benefits over baselines. As SNR decreases, we find our scene-adaptive loss can de-emphasize RGB and rely on diffuse LiDAR, enabling robust depth estimation even when noise fully obscures RGB cues. In these cases, RGB with sparse LiDAR struggles to reconstruct smooth geometry due to poor spatial coverage in LiDAR signals and the absence of multi-view RGB cues.

5.2. Real-world Experiments

Experimental Setup. We capture diffuse LiDAR histograms using a low-cost AMS TMF8828 [1] ToF sensor comprised of 18×12 SPAD pixels which are aggregated into 8×8 measurement zones on-device. The sensor is set to short-range, high-accuracy mode, increasing the temporal resolution but reducing the max measurement distance. RGB is captured with a co-located, rigidly mounted RealSense D435i module. We simulate a spot LiDAR by subsampling depth measurements from the D435i stereo IR depth map. We restrict input views to 90 captures uniformly sampled at a single elevation angle along the hemisphere. We describe this capture setup and sensor calibration in more detail in the supplement.

Experimental Results. We provide a qualitative comparison of our technique against RGB and sparse LiDAR-

Table 2. **LiDAR-only depth MAE (\downarrow) ablation.** Our reconstruction using diffuse LiDAR outperforms, on 10 training views, both sparse point LiDAR and sparse histogram LiDAR reconstruction. Additional baseline comparisons are provided in the supplement.

Method	Blender	Chair	Hotdog	Lego
Surfels - Sparse LiDAR Only	0.078	0.099	0.084	0.074
TransientNeRF [27] - Sparse Histograms	0.281	0.221	0.222	0.228
Ours - Diffuse LiDAR Only	0.041	0.040	0.036	0.037

guided reconstruction for in challenging, real-world low-texture, low-lighting, and low-albedo scenes, shown in Fig. 7. We find that our technique is effective in improving robustness of surface reconstruction across these challenging settings. As in synthetic experiments, we observe improved ground-plane separation, better geometry estimation for textureless concavities not visible from the RGB hull, and the ability to overcome complete failure cases of RGB methods in limited-view low-lighting conditions. We note that in very challenging conditions with very low lighting, albedo, and texture (“Blundstone”), we observe poor object-plane separation; yet, even in this case, we estimate improved geometry over RGB with sparse LiDAR.

Ablations. We compare our diffuse LiDAR approach to existing sparse LiDAR-based techniques for *LiDAR-only* reconstruction. We compare depth MAE using LiDAR-only loss for 1) Surfels-based sparse LiDAR, 2) TransientNeRF [27], which uses sparse histograms instead of point depths, and 3) ours with diffuse LiDAR loss only. We find that our approach consistently outperforms these sparse LiDAR baselines (shown in Tab. 2). Importantly, we achieve depth MAE values close to those in No Texture reconstruction (Fig. 5d), suggesting that our scene-adaptive loss is effective in dynamically using LiDAR in low-texture scenes. We provide an additional ablation of our scene-adaptive loss in the supplement, where we demonstrate that removing this scene-adaptive loss degrades reconstruction quality.

6. Conclusion

In this work, we demonstrate how to leverage the complementary strengths of RGB and blurred LiDAR sensors for robust handheld 3D scanning in low-texture, low-light, and low-albedo environments. We show through recoverability analysis, quantitative evaluation, and qualitative real world experiments that diffuse LiDAR can, counterintuitively, enable more robust 3D scanning in these challenging settings. Future work in this direction could explore the role of diffuse LiDAR for joint ego pose and 3D estimation, and analyze robustness to other scenarios such as moving objects and challenging material types. The proposed reconstruction technique also holds potential for robust 3D scanning in other domains that require robust, mobile 3D scanning such as AR, VR, and robotics. We believe that this work can unlock the potential for combining these unconventional yet widely available sensors with RGB for robust 3D vision.

Acknowledgements. Nikhil Behari is supported by the NASA Space Technology Graduate Research Opportunity Fellowship. Aaron Young and Siddharth Somasundaram are funded by the National Science Foundation (NSF) Graduate Research Fellowship Program (Grant #2141064). Tzofi Klinghoffer is supported by the Department of Defense (DoD) National Defense Science and Engineering Graduate (NDSEG) Fellowship Program.

References

- [1] AMS OSRAM AG. TMF882X Datasheet. <https://ams-osram.com/>, n.d. 8, 14
- [2] Benjamin Attal, Eliot Laidlaw, Aaron Gokaslan, Changil Kim, Christian Richardt, James Tompkin, and Matthew O’Toole. Törf: Time-of-flight radiance fields for dynamic scene view synthesis. *Advances in neural information processing systems*, 34:26289–26301, 2021. 2
- [3] Gilad Baruch, Zhuoyuan Chen, Afshin Dehghan, Tal Dimry, Yuri Feigin, Peter Fu, Thomas Gebauer, Brandon Joffe, Daniel Kurz, Arik Schwartz, et al. Arkitscenes: A diverse real-world dataset for 3d indoor scene understanding using mobile rgb-d data. *arXiv preprint arXiv:2111.08897*, 2021. 13
- [4] David Borts, Erich Liang, Tim Broedermann, Andrea Ramazzina, Stefanie Walz, Edoardo Palladin, Jipeng Sun, David Brueggemann, Christos Sakaridis, Luc Van Gool, et al. Radar fields: Frequency-space neural scene representations for fmcw radar. In *ACM SIGGRAPH 2024 Conference Papers*, pages 1–10, 2024. 2
- [5] Sony Semiconductor Solutions Corporation. Sony semiconductor solutions to release spad depth sensor for smartphones with high-accuracy, low-power distance measurement performance, powered by the industry’s highest photon detection efficiency, 2023. <https://www.sony-semicon.com/en/news/2023/2023030601.html>. 2
- [6] Angela Dai, Angel X Chang, Manolis Savva, Maciej Halber, Thomas Funkhouser, and Matthias Nießner. Scannet: Richly-annotated 3d reconstructions of indoor scenes. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 5828–5839, 2017. 13
- [7] Pinxuan Dai, Jiamin Xu, Wenxiang Xie, Xinguo Liu, Huamin Wang, and Weiwei Xu. High-quality surface reconstruction using gaussian surfels. In *ACM SIGGRAPH 2024 Conference Papers*, pages 1–11, 2024. 1, 2, 4, 5, 6, 7
- [8] Kangle Deng, Andrew Liu, Jun-Yan Zhu, and Deva Ramanan. Depth-supervised nerf: Fewer views and faster training for free. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 12882–12891, 2022. 2
- [9] Daniele Faccio, Andreas Velten, and Gordon Wetzstein. Non-line-of-sight imaging. *Nature Reviews Physics*, 2(6): 318–327, 2020. 2
- [10] Yuki Fujimura, Takahiro Kushida, Takuya Funatomi, and Yasuhiro Mukaigawa. Nlos-neus: Non-line-of-sight neural implicit surface. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 10532–10541, 2023. 2
- [11] Daniel Gilo, Tzofi Klinghoffer, and Or Litany. Epi-naf: Enhancing neural attenuation fields for limited-angle ct with epipolar consistency conditions, 2024. 2
- [12] Felix Heide, Matthew O’Toole, Kai Zang, David B Lindell, Steven Diamond, and Gordon Wetzstein. Non-line-of-sight imaging with partial occluders and surface normals. *ACM Transactions on Graphics (ToG)*, 38(3):1–10, 2019. 2
- [13] Shengyu Huang, Zan Gojcic, Zian Wang, Francis Williams, Yoni Kasten, Sanja Fidler, Konrad Schindler, and Or Litany. Neural lidar fields for novel view synthesis. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 18236–18246, 2023. 2
- [14] Suren Jayasuriya, Adithya Pediredla, Sriram Sivaramakrishnan, Alyosha Molnar, and Ashok Veeraraghavan. Depth fields: Extending light field techniques to time-of-flight imaging. In *2015 International Conference on 3D Vision*, pages 1–9. IEEE, 2015. 2
- [15] Sacha Jungerman and Mohit Gupta. Radiance fields from photons. *arXiv preprint arXiv:2407.09386*, 2024. 2
- [16] Sacha Jungerman, Atul Ingle, Yin Li, and Mohit Gupta. 3d scene inference from transient histograms. In *European Conference on Computer Vision*, pages 401–417. Springer, 2022. 2
- [17] Jaewon Kam, Jungeon Kim, Soongjin Kim, Jaesik Park, and Seungyong Lee. Costdcnet: Cost volume based depth completion for a single rgb-d image. In *European Conference on Computer Vision*, pages 257–274. Springer, 2022. 14
- [18] Bernhard Kerbl, Georgios Kopanas, Thomas Leimkühler, and George Drettakis. 3d gaussian splatting for real-time radiance field rendering. *ACM Trans. Graph.*, 42(4):139–1, 2023. 1, 2, 4, 5
- [19] Ahmed Kirmani, Tyler Hutchison, James Davis, and Ramesh Raskar. Looking around the corner using transient imaging. In *2009 IEEE 12th International Conference on Computer Vision*, pages 159–166. IEEE, 2009. 2
- [20] Tzofi Klinghoffer, Xiaoyu Xiang, Siddharth Somasundaram, Yuchen Fan, Christian Richardt, Ramesh Raskar, and Rakesh Ranjan. Platonerf: 3d reconstruction in plato’s cave via single-view two-bounce lidar. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 14565–14574, 2024. 2
- [21] Jongho Lee, Atul Ingle, Jenu V Chacko, Kevin W Eliceiri, and Mohit Gupta. Caspi: collaborative photon processing for active single-photon imaging. *Nature Communications*, 14(1):3158, 2023. 2
- [22] Yijin Li, Xinyang Liu, Wenqi Dong, Han Zhou, Hujun Bao, Guofeng Zhang, Yinda Zhang, and Zhaopeng Cui. Deltar: Depth estimation from a light-weight tof sensor and rgb image. In *European conference on computer vision*, pages 619–636. Springer, 2022. 3
- [23] Tsung-Han Lin, Connor Henley, Siddharth Somasundaram, Akshat Dave, Moshe Laifenfeld, and Ramesh Raskar. Hand-held mapping of specular surfaces using consumer-grade flash lidar. In *2024 IEEE International Conference on Computational Photography (ICCP)*, pages 1–12. IEEE, 2024. 14
- [24] Ralph D Lorenz, Sylvestre Maurice, Baptiste Chide, David Mimoun, Alexander Stott, Naomi Murdoch, Martin Giller,

- Xavier Jacob, Roger C Wiens, Franck Montmessin, et al. The sounds of a helicopter on mars. *Planetary and Space Science*, 230:105684, 2023. 1
- [25] Weihang Luo, Anagh Malik, and David B Lindell. Transientangelo: Few-viewpoint surface reconstruction using single-photon lidar. *arXiv preprint arXiv:2408.12191*, 2024. 2
- [26] Anagh Malik, Noah Juravsky, Ryan Po, Gordon Wetzstein, Kiriakos N Kutulakos, and David B Lindell. Flying with photons: Rendering novel views of propagating light. *arXiv preprint arXiv:2404.06493*, 2024.
- [27] Anagh Malik, Parsa Mirdehghan, Sotiris Nousias, Kyros Kutulakos, and David Lindell. Transient neural radiance fields for lidar view synthesis and 3d reconstruction. *Advances in Neural Information Processing Systems*, 36, 2024. 2, 6, 8, 13
- [28] Andreas Meuleman, Hakyong Kim, James Tompkin, and Min H Kim. Floatingfusion: Depth from tof and image-stabilized stereo cameras. In *European Conference on Computer Vision*, pages 602–618. Springer, 2022. 3
- [29] Ben Mildenhall, Pratul P Srinivasan, Matthew Tancik, Jonathan T Barron, Ravi Ramamoorthi, and Ren Ng. Nerf: Representing scenes as neural radiance fields for view synthesis. *Communications of the ACM*, 65(1):99–106, 2021. 1, 2
- [30] Fangzhou Mu, Sicheng Mo, Jiayong Peng, Xiaochun Liu, Ji Hyun Nam, Siddeshwar Raghavan, Andreas Velten, and Yin Li. Physics to the rescue: Deep non-line-of-sight reconstruction for high-speed imaging. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2022. 2
- [31] Fangzhou Mu, Carter Sifferrman, Sacha Jungerman, Yiquan Li, Mark Han, Michael Gleicher, Mohit Gupta, and Yin Li. Towards 3d vision with low-cost single-photon cameras. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 5302–5311, 2024. 3, 15
- [32] Mikhail Okunev, Marc Mapeke, Benjamin Attal, Christian Richardt, Matthew O’Toole, and James Tompkin. Flowed time of flight radiance fields. In *European Conference on Computer Vision*, pages 373–389. Springer, 2025. 2
- [33] Matthew O’Toole, Felix Heide, David B Lindell, Kai Zang, Steven Diamond, and Gordon Wetzstein. Reconstructing transient images from single-photon sensors. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1539–1547, 2017. 2
- [34] Mohamad Qadri, Kevin Zhang, Akshay Hinduja, Michael Kaess, Adithya Pediredla, and Christopher A Metzler. Aoneus: A neural rendering framework for acoustic-optical sensor fusion. In *ACM SIGGRAPH 2024 Conference Papers*, pages 1–12, 2024. 2
- [35] Sabbir Rangwala. The iPhone 12—LiDAR at your fingertips, 2020. <https://www.forbes.com/sites/sabbirrangwala/2020/11/12/the-iphone-12lidar-at-your-fingertips/>. 2
- [36] Ramesh Raskar, Andreas Velten, Sebastian Bauer, and Tristram Swedish. Seeing around corners using time of flight. In *ACM SIGGRAPH 2020 Courses*, pages 1–97. Association for Computing Machinery, 2020. 2
- [37] Johannes L Schonberger and Jan-Michael Frahm. Structure-from-motion revisited. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 4104–4113, 2016. 2, 4
- [38] Siyuan Shen, Zi Wang, Ping Liu, Zhengqing Pan, Ruiqian Li, Tian Gao, Shiyong Li, and Jingyi Yu. Non-line-of-sight imaging via neural transient fields. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 43(7):2257–2268, 2021. 2
- [39] Carter Sifferman, Yeping Wang, Mohit Gupta, and Michael Gleicher. Unlocking the performance of proximity sensors by utilizing transient histograms. *IEEE Robotics and Automation Letters*, 2023. 3
- [40] Siddharth Somasundaram, Akshat Dave, Connor Henley, Ashok Veeraraghavan, and Ramesh Raskar. Role of transients in two-bounce non-line-of-sight imaging. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 9192–9201, 2023. 4
- [41] Jiuhui Song, Seonghoon Park, Honggyu An, Seokju Cho, Min-Seop Kwak, Sungjin Cho, and Seungryong Kim. Därf: Boosting radiance fields from sparse input views with monocular depth adaptation. *Advances in Neural Information Processing Systems*, 36:68458–68470, 2023. 2
- [42] Zhanghao Sun, Wei Ye, Jinhui Xiong, Gyeongmin Choe, Jialiang Wang, Shuochen Su, and Rakesh Ranjan. Consistent direct time-of-flight video depth super-resolution. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 5075–5085, 2023. 3
- [43] TechInsights. LiDAR technology in Apple Vision Pro, 2024. <https://www.techinsights.com/blog/lidar-technology-apple-vision-pro>. 2
- [44] Andreas Velten, Thomas Willwacher, Otkrist Gupta, Ashok Veeraraghavan, Mounsi G Bawendi, and Ramesh Raskar. Recovering three-dimensional shape around a corner using ultrafast time-of-flight imaging. *Nature communications*, 3(1):745, 2012. 2
- [45] Oleg Voynov, Gleb Bobrovskikh, Pavel Karpyshev, Savely Galochkin, Andrei-Timotei Ardelean, Arseniy Bozhenko, Ekaterina Karmanova, Pavel Kopanov, Yaroslav Labutin-Rymsho, Ruslan Rakhimov, et al. Multi-sensor large-scale dataset for multi-view 3d reconstruction. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 21392–21403, 2023. 13
- [46] Peng Wang, Lingjie Liu, Yuan Liu, Christian Theobalt, Taku Komura, and Wenping Wang. Neus: Learning neural implicit surfaces by volume rendering for multi-view reconstruction. *arXiv preprint arXiv:2106.10689*, 2021. 2
- [47] Zhou Wang, Eero P Simoncelli, and Alan C Bovik. Multiscale structural similarity for image quality assessment. In *The Thirty-Seventh Asilomar Conference on Signals, Systems & Computers*, 2003, pages 1398–1402. Ieee, 2003. 6
- [48] Hanfeng Wu, Xingxing Zuo, Stefan Leutenegger, Or Litany, Konrad Schindler, and Shengyu Huang. Dynamic lidar resimulation using compositional neural fields. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 19988–19998, 2024. 2
- [49] Shumian Xin, Sotiris Nousias, Kiriakos N Kutulakos, Aswin C Sankaranarayanan, Srinivasa G Narasimhan, and

- Ioannis Gkioulekas. A theory of fermat paths for non-line-of-sight shape reconstruction. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 6800–6809, 2019. [2](#)
- [50] Lihe Yang, Bingyi Kang, Zilong Huang, Xiaogang Xu, Jiashi Feng, and Hengshuang Zhao. Depth anything: Unleashing the power of large-scale unlabeled data. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 10371–10381, 2024. [7](#)
- [51] Lior Yariv, Jiatao Gu, Yoni Kasten, and Yaron Lipman. Volume rendering of neural implicit surfaces. *Advances in Neural Information Processing Systems*, 34:4805–4815, 2021. [1](#)
- [52] Junge Zhang, Feihu Zhang, Shaochen Kuang, and Li Zhang. Nerf-lidar: Generating realistic lidar point clouds with neural radiance fields. In *Proceedings of the AAAI Conference on Artificial Intelligence*, pages 7178–7186, 2024. [2](#)
- [53] Tianyi Zhang, Mel J White, Akshat Dave, Shahaboddin Ghajari, Ankit Raghuram, Alyosha C Molnar, and Ashok Veeraraghavan. First arrival differential lidar. In *2022 IEEE International Conference on Computational Photography (ICCP)*, pages 1–12. IEEE, 2022. [3](#)
- [54] Matthias Zwicker, Hanspeter Pfister, Jeroen Van Baar, and Markus Gross. Ewa volume splatting. In *Proceedings Visualization, 2001. VIS'01.*, pages 29–538. IEEE, 2001. [5](#)