

When 3D Gaussian Splatting Recovers Real Surfaces

Songhe Wang¹ and David Miller¹

School of EECS, Penn State, University Park, PA, USA
{sxw5765,djm25}@psu.edu

Abstract. When does 3D Gaussian Splatting (3DGS) recover the true scene surface rather than just overfitting view-dependent appearance? We answer this by developing a mathematical framework—based on a first-hit rendering abstraction—that cleanly isolates geometry from appearance. We prove that geometric misalignment forcefully converts spatial textures into high-frequency angular signals via parallax. This establishes a strict identifiability window: if angular capacity is bounded, surface-consistent solutions are mathematically preferred; if unrestricted, the same images can be perfectly explained by an incorrect, opaque “billboard” geometry. Experiments on synthetic stress tests confirm this prediction, showing billboard failures emerge precisely at high angular capacities. Conversely, in the real-world datasets we evaluate under standard capture protocols, reconstructions remain surface-consistent even at high SH degrees, which is consistent with the prediction that rich spatial texture can push billboard solutions outside the tested angular-capacity range.

Keywords: 3D Gaussian Splatting · Geometric Identifiability · Spherical Harmonics · Light Fields

1 Introduction

3D Gaussian Splatting (3DGS) renders high-quality views in real time and has rapidly become a default primitive for 3D reconstruction and editing [20]. Crucially, modern pipelines no longer treat 3DGS as a hidden renderer that only produces images: they increasingly export the learned Gaussians as geometry and build downstream systems on top of them. On the one hand, a growing line of work treats 3DGS as a surrogate surface representation and demonstrates that high-quality triangle meshes can be extracted from learned Gaussian fields, making 3DGS a central building block for geometry processing and reconstruction [2, 7, 15, 25–27, 35, 39–43]. Across these lines of work, the learned Gaussian field is implicitly treated as if it were the true scene geometry: meshes are carved out of it, and downstream edits are applied on top of it. Yet all of this evidence is essentially *empirical*: to our knowledge there is no theoretical guarantee that the standard 3DGS objective actually recovers the underlying surface. Without such a guarantee, treating 3DGS as ground-truth geometry is methodologically fragile: a model that fits all training views while drifting off the true surface could silently

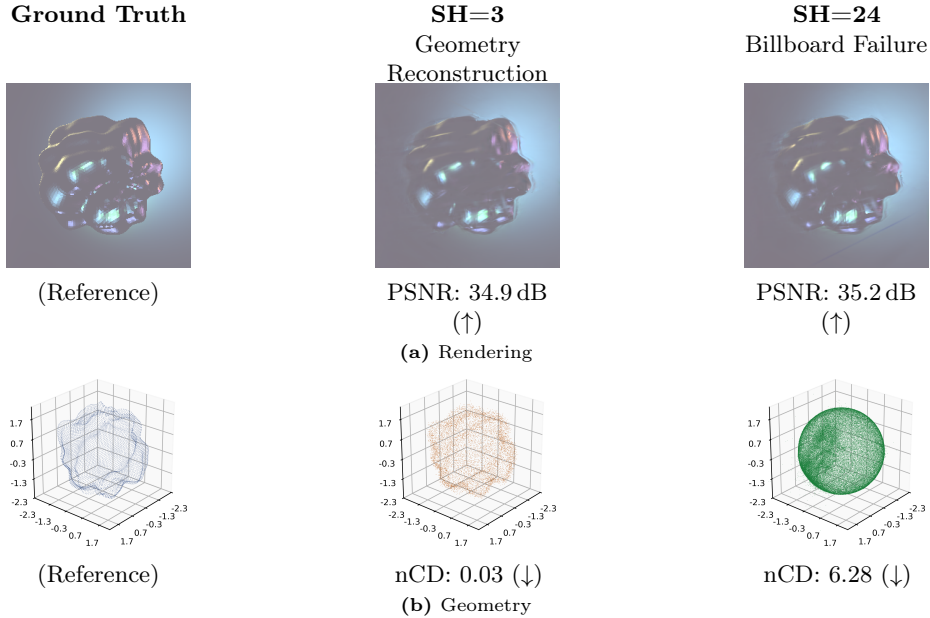


Fig. 1: Billboard failure in 3D Gaussian Splatting. Top: a glinty ground-truth object (left) and two 3DGS reconstructions with low spherical-harmonic order (SH= 3, middle) and high order (SH= 24, right). Both reconstructions achieve high image PSNR, but the underlying 3D shape is completely different. Bottom: visualizing Gaussian centers as point clouds shows that SH= 3 recovers the detailed surface with a small nCD distance, whereas SH= 24 collapses into an off-surface opaque billboard-like layer whose nCD error is much larger. This paradox—good appearance yet catastrophic geometry as radiance capacity increases—is the central phenomenon that our theory explains and our experiments systematically characterize in this paper.

corrupt downstream meshes and contact manifolds. This raises a basic question of *geometric trustworthiness*: under what conditions are the learned Gaussians truly surface-consistent, so that reconstruction and editing built on top of them rest on a sound geometric foundation? Here, *surface-consistent* means geometrically close to the true surface under nCD, not merely photometrically accurate.

In this work, we address this question from a theoretical point of view. We build a mathematical framework that isolates the interaction between a scene’s spatial geometry and its view-dependent appearance. Specifically, we abstract the rendering process to analyze the best-achievable population loss for any candidate surface under a bounded *angular capacity*—a strict limit on how complex the view-dependent colors are allowed to be. Within this framework, we demonstrate that optimizing appearance essentially reduces to a mathematical projection problem. This allows us to rigorously compare the approximation error of the true geometry against any misaligned geometry by evaluating their required angular frequencies.

Through this theoretical analysis, we predict a severe and counterintuitive vulnerability, which we refer to as the *opaque billboard* failure mode. Our proofs reveal that if the angular capacity is allowed to be excessively large, a geometrically misaligned surface can exploit high-frequency view-dependent colors to perfectly mimic the spatial details of the true background. In this regime, an artificial interior sheet can block the rays and paint the correct colors using high angular capacity, so the images look perfectly right even though the geometry is completely wrong (see Fig. 1). Conversely, our theory guarantees that if the angular capacity is appropriately restricted, the model is mathematically incapable of sustaining this illusion, ensuring the true geometry remains strictly easier to fit.

On the empirical side, we instantiate these theoretical predictions in practice. We construct a controlled synthetic stress-test dataset with tunable object complexity to deliberately probe this vulnerability. As predicted by our theoretical framework, training 3DGS with large angular capacity reliably triggers opaque billboard solutions: rendering quality remains near-optimal, while the geometry drifts completely off the true surface. In our real-scene benchmarks, the observed spatial texture complexity appears to push the induced frequency ω_{fake} beyond the SH degrees we test (e.g., up to $L = 24$). Under these standard training setups and moderate capacities, we do not observe opaque billboard failures.

As a summary, our contributions are:

- We provide a mathematical framework for 3D Gaussian Splatting that cleanly isolates geometry from appearance, turning the geometric identifiability problem into a rigorous frequency comparison.
- Within this framework, we mathematically predict the *opaque billboard* failure mode. We prove that given unbounded or large angular capacity, an incorrect surface can perfectly match training images by memorizing spatial textures as view-dependent signals.
- We establish a theoretical *identifiability window*, proving that strictly bounding angular capacity is mathematically necessary to reject billboards and guarantee accurate surface recovery.
- We complement the theory with controlled diagnostics, demonstrating that opaque billboard failures can be deterministically forced in synthetic stress tests, while no opaque-billboard cases appear in our real-scene diagnostics.

Scope. Although framed around 3DGS, the mechanism applies to any proxy or surface representation with a finite-degree SH radiance head; 3DGS is our motivating instance.

2 Related Work

Meshes and downstream interactions in 3DGS. Many methods extract surfaces from 3DGS using surface alignment [15, 41], 2D Gaussians or surfels [9, 18], SDFs [26, 42], geometric priors [7, 39], sparse adaptations [35, 40], or hybrid optimization [27], as surveyed in [2, 25, 43]. Others use Gaussians for dynamics, contact, and deformation [3, 13, 16, 19, 29, 33, 37, 45]. These applications rely on

learned Gaussians as geometry, yet standard 3DGS [20] lacks surface-recovery guarantees.

Plenoptic bandwidth and spherical harmonics. Classical studies in image-based rendering showed that view synthesis with imperfect proxy geometry converts spatial texture variation into angular variation, tying sampling requirements to depth and disparity [4, 5, 10, 11, 14, 22, 24, 30]. To model angular radiance efficiently, low-order spherical harmonics (SH) have long been used [32, 36]. In 3DGS, SH coefficients serve as a finite-dimensional view-dependent radiance head. Our contribution is not to rederive plenoptic bandwidth or SH capacity from scratch, but to reinterpret this frequency transport as a geometric-identifiability condition. We study how a finite SH head can prevent or permit a wrong proxy surface to explain parallax-induced color variation.

Shape–radiance ambiguity in neural fields. Neural radiance-field methods also expose ambiguities between geometry and view-dependent appearance [28, 38, 47]. Related analyses have observed that color estimation and geometry can be coupled in ways that obscure shape recovery [12]. Our analysis gives a complementary parallax-frequency explanation: bounded angular capacity creates a regime where the true surface is easier to fit than a misaligned proxy, whereas excessive capacity can remove that preference.

3 Background and Motivation

To formally analyze the relationship between angular capacity and geometric correctness, we must first establish how 3D Gaussian Splatting (3DGS) couples geometry with appearance, and how we can mathematically abstract this coupling to expose its core vulnerabilities.

3.1 3DGS, View-Dependence, and Angular Capacity

In 3DGS [20], a scene is not represented by continuous meshes or implicit neural fields, but by an unstructured collection of 3D Gaussian primitives. Each primitive is defined by explicit geometric parameters—a 3D center, a covariance matrix dictating its scale and rotation, and an opacity value. To render a novel view, these primitives are splatted onto the image plane and alpha-composited along camera rays.

Crucially, real-world surfaces exhibit view-dependent appearance, such as specular highlights that shift as the observer moves. To model this, 3DGS does not assign a single static RGB value to a Gaussian. Instead, the emitted color of each primitive is a function of the viewing direction $\hat{v} \in \mathbb{S}^2$. This view-dependence is parameterized using Spherical Harmonics (SH). For each color channel, a Gaussian stores a set of coefficients $\{c_{\ell m}\}$, and the color is evaluated as a linear combination of basis functions:

$$\text{color}(\hat{v}) = \sum_{\ell=0}^L \sum_{m=-\ell}^{\ell} c_{\ell m} Y_{\ell m}(\hat{v}).$$

Here, $Y_{\ell m}$ denotes the real Spherical Harmonic basis function of degree ℓ and order m . Conceptually, these basis functions serve as the angular equivalent of a Fourier series on the sphere. The degree ℓ dictates the spatial frequency of the signal: $\ell = 0$ represents a purely diffuse, view-independent color, while higher values of ℓ allow the function to represent sharper, more rapid changes in color as the camera angle shifts.

By truncating this summation at a maximum degree L , the architecture enforces a strict low-pass filter on the primitive’s angular capacity. This truncation defines a specific function space, which we denote as:

$$\mathcal{V}_L := \text{span}\{Y_{\ell m} : 0 \leq \ell \leq L, -\ell \leq m \leq \ell\}.$$

Here, we only allow SH terms up to degree $\ell \leq L$. Consequently, the view-dependent color function of each Gaussian is strictly restricted to a finite-dimensional subspace. A larger L simply expands this space, allowing the model to represent more complex angular dependencies.

3.2 Isolating the Phenomenon: The First-Hit Abstraction

Standard 3DGS rendering relies on volumetric alpha-blending, where the final pixel color is accumulated from multiple semi-transparent Gaussians along a ray. While necessary for gradient flow during early optimization, volumetric compositing obscures the mathematical limits of geometric identifiability by entangling visibility, occlusion, and depth-ordering.

However, the specific failure mode we aim to explain—where a grossly misaligned “billboard” surface fits the training images by painting fake, view-dependent textures—is fundamentally a local, surface-level deception. It occurs because the optimizer places dense, opaque geometry in the wrong physical location, and relies on the SH coefficients of that specific geometry to perfectly mimic the background rays passing through it.

To rigorously study this geometry-appearance tradeoff, we simplify the rendering process to a **first-hit abstraction**. We assume the optimized geometry forms an opaque surface, meaning the color of any ray is entirely determined by the appearance function of the very first point it intersects. To ground our first-hit abstraction in empirical reality, we conducted a pilot experiment profiling the actual per-ray alpha-compositing weights of a trained 3DGS model. This pilot study reveals an extreme depth-wise concentration: for 92.6% of foreground rays, the front-most 3 contributing Gaussians systematically capture over 90.2% of compositing mass. This statistic should not be interpreted as proving that 3DGS rendering is generally first-hit or opaque. Rather, it provides empirical motivation that, in the mostly opaque scenes used in our controlled experiments, the learned compositing weights are often highly concentrated near the front-most contributors. The first-hit model is therefore best understood as an analytically tractable abstraction of one important regime, not as a faithful model of all 3DGS behavior. For this pilot, we sort contributing Gaussians by depth along each sampled foreground ray and measure the cumulative alpha-compositing mass of the first three contributors.

This abstraction is analytically highly self-consistent because it preserves the exact structural mechanism that enables billboards. It cleanly isolates the core interaction into two distinct roles:

- **Geometry:** Which 3D point does the model claim the ray hits first?
- **Appearance:** What color must the angular function $\mathcal{V}_L$ emit at that specific point to satisfy the training data?

4 Setup and the Geometry-Fixed Projection Principle

To formally evaluate whether the optimization landscape prefers the true geometry over a misaligned one, we abstract the 3DGS rendering process into a minimal *first-hit* model. This isolates the core geometry-appearance interaction without the confounding effects of volumetric alpha-blending.

4.1 Minimal First-Hit Model and Competing Error Curves

We parameterize a camera ray by its origin $y \in \mathbb{R}^3$ and viewing direction $\hat{v} \in \mathbb{S}^2$. Let $S, U \subset \mathbb{R}^3$ denote the true scene geometry and a candidate (potentially misaligned) surface, with respective first-hit points $x_S(y, \hat{v})$ and $x_U(y, \hat{v})$. The ground-truth pixel observation is driven by the true physical appearance $C(x, \hat{v})$ defined on the true surface manifold S , yielding $I^*(y, \hat{v}) := C(x_S(y, \hat{v}), \hat{v})$. If our model commits to the candidate geometry U , it must assign an appearance function $g(x, \hat{v})$ strictly constrained by the angular capacity L (i.e., $g(x, \cdot) \in \mathcal{V}_L$). The model’s prediction is thus $I_{U,g}(y, \hat{v}) := g(x_U(y, \hat{v}), \hat{v})$.

Let $(Y, \hat{V}) \sim \mathbb{P}$ be a random training ray sampled from the distribution, and define $X_U := x_U(Y, \hat{V})$. We call this expected image-space MSE over the ray distribution the *population loss*: $\mathcal{J}(U, g) := \mathbb{E}_{(Y, \hat{V}) \sim \mathbb{P}}[|I_{U,g}(Y, \hat{V}) - I^*(Y, \hat{V})|^2]$. To determine geometric identifiability, we abstract away specific optimizer dynamics and directly evaluate the *best-achievable error* for both the true and candidate surfaces under the same angular capacity L :

$$A(L) := \inf_{g: g(x, \cdot) \in \mathcal{V}_L} \mathcal{J}(S, g), \quad (1)$$

$$B_U(L) := \inf_{g: g(x, \cdot) \in \mathcal{V}_L} \mathcal{J}(U, g). \quad (2)$$

In plain terms, $A(L)$ measures how well degree L fits the real appearance given perfect geometry, while $B_U(L)$ measures how well degree L can fake the appearance to compensate for wrong geometry. The entire identifiability narrative reduces to identifying a mathematical regime where $A(L) \ll B_U(L)$.

4.2 Geometry-Fixed Optimality as Pointwise Projection

To evaluate $A(L)$ and $B_U(L)$ without explicitly solving the global optimization, we must first identify the mathematically optimal appearance function for a fixed

geometry. Conditioned on hitting a specific physical point $x \in U$, we define the through-seen target $F_U(x, \cdot)$ as the geometry-induced angular target: the colors a candidate point must emit across views to reproduce the true scene. Formally, it is the conditional expectation of the ground-truth image color over the camera origins Y , given the specific hit point and direction $\hat{V}$:

$$F_U(x, \cdot) := \mathbb{E}[I^*(Y, \hat{V}) \mid X_U = x, \hat{V}] \quad (3)$$

which yields $F_U(x, \hat{v})$ when evaluated at a specific direction $\hat{v}$. This defines the exact regression function in the angular domain that the training data demands at that location.

Framing the target this way reduces (in our abstraction) the daunting global optimization to independent, local geometric projections.

Proposition 1 (Pointwise projection). *For a fixed surface U and SH degree L , the global minimizer of $\mathcal{J}(U, g)$ is given pointwise by $g^*(x, \cdot) = \Pi_{\mathcal{V}_L}^{(x)}(F_U(x, \cdot))$, where $\Pi_{\mathcal{V}_L}^{(x)}$ is the orthogonal projection onto $\mathcal{V}_L$ under the L^2 inner product weighted by the conditional viewing distribution at x . Consequently, the optimal loss is exactly the expected projection distance:*

$$B_U(L) = \mathbb{E}_{x \sim \mu_U} [\text{dist}(F_U(x, \cdot), \mathcal{V}_L)^2].$$

Because the appearance parameters at point x only affect rays hitting x , the global integral cleanly decouples into independent least-squares approximations (a rigorous measure-theoretic proof via disintegration is provided in the supplementary material). This principle is crucial: it transforms the geometric identifiability problem into a pure signal-processing question. We now only need to quantify how well the induced target $F_U(x, \cdot)$ can be approximated by the available angular space $\mathcal{V}_L$. That is, we measure the weighted projection error $\text{dist}(F_U(x, \cdot), \mathcal{V}_L)$, which precisely corresponds to the residual energy of F_U in the high-frequency SH degrees $\ell > L$.

5 The True Surface: Why Smooth Residuals Yield Rapid Decay

To evaluate the best-achievable loss $A(L)$ on the correct geometry, we must formalize the physical appearance of the true scene. We adopt a decomposition where the ground-truth appearance $C(x, \hat{v})$ consists of a view-independent diffuse texture $T(x)$ and a view-dependent specular residual $R(x, \hat{v})$.

Assumption 1 (Diffuse + Sobolev-smooth residual). *The ground-truth appearance on S decomposes as $C = T + R$. Moreover, there exist $s_{\text{true}} > 0$ and $M_{\text{true}} < \infty$ such that for almost every $x \in S$, the view-dependent residual is mathematically smooth:*

$$\|R(x, \cdot)\|_{H^{s_{\text{true}}}(\mathbb{S}^2)} \leq M_{\text{true}}.$$

This assumption mathematically enforces that while view-dependent effects like specular highlights exist, they are not arbitrarily sharp or infinitely spiky in direction. The Sobolev norm $H^s(\mathbb{S}^2)$ controls the angular high-frequency energy, meaning a larger s ensures a stronger decay of SH coefficients. This realistically models most non-mirror materials and is precisely what makes the true surface straightforward to fit.

When the model geometry is perfectly correct, the through-seen target aligns flawlessly with the physical appearance: $F_S(x, \hat{v}) = T(x) + R(x, \hat{v})$. Because the diffuse term $T(x)$ does not change with the viewing direction $\hat{v}$, it resides entirely within the SH degree-0 subspace ($\mathcal{V}_0 \subset \mathcal{V}_L$).

Consequently, the spatial complexity of the true texture—no matter how high-frequency or intricate—consumes absolutely zero angular capacity. The approximation difficulty stems exclusively from the residual R :

$$\text{dist}(F_S(x, \cdot), \mathcal{V}_L) = \text{dist}(R(x, \cdot), \mathcal{V}_L).$$

Because R is assumed to be Sobolev-smooth, its projection error onto a bandlimited SH subspace naturally decays. We formalize this using standard spectral approximation on the sphere (full derivation provided in the supplementary material):

Lemma 1 (Sobolev smoothness implies fast SH truncation). *Let $P_{\leq L}$ be the $L^2(\mathbb{S}^2)$ orthogonal projection onto $\mathcal{V}_L$. If $f \in H^s(\mathbb{S}^2)$, then its truncation error is bounded by:*

$$\|f - P_{\leq L}f\|_{L^2(\mathbb{S}^2)}^2 \leq (1 + L(L+1))^{-s} \|f\|_{H^s(\mathbb{S}^2)}^2.$$

(For simplicity, we state the truncation bound under the uniform sphere measure; extending this to pointwise weighted norms requires mild density bounds on $\nu_U(\cdot | x)$, as detailed in the supplementary material).

This rate comes from the Laplace–Beltrami eigenvalues $\ell(\ell+1)$, which define the Sobolev weights $(1 + \ell(\ell+1))^s$. Truncating the expansion at the maximum degree L then yields the specific factor $(1 + L(L+1))^{-s}$ observed in the error bound. Bounding the Sobolev norm directly suppresses the energy in the tail $\ell > L$. By applying this lemma to our residual, we immediately establish the upper bound for the true surface.

Proposition 2 (True-surface upper bound). *Under Assumption 1, the best achievable loss on the true surface satisfies:*

$$A(L) \leq (1 + L(L+1))^{-s_{\text{true}}} M_{\text{true}}^2,$$

up to a multiplicative constant depending only on view-direction sampling.

The proof follows directly from our pointwise projection principle (Proposition 1). Since $T(x)$ projects with zero error, the squared distance

$$\text{dist}(F_S(x, \cdot), \mathcal{V}_L)^2$$

is simply the truncation error of $R(x, \cdot)$. Applying Lemma 1 and averaging over the surface points yields the bound.

This result establishes a critical baseline: $A(L)$ drops rapidly with L . A moderate SH degree is more than sufficient to make the true surface an attractive minimum for the optimizer. If this were the whole story, one might assume that simply increasing L is harmless. However, as we demonstrate next, a misaligned geometry interacts with diffuse texture in a fundamentally different way, forcefully converting spatial details into high-frequency angular signals.

6 The Misaligned Surface: Parallax Forges High Angular Frequencies

Having established that the true surface inherently requires very little angular capacity, we now turn to what happens when the model proposes a geometrically misaligned candidate. We aim to show that wrong geometry is penalized heavily because it forces spatial complexity to masquerade as angular complexity.

6.1 Formalizing the Billboard Scenario

To analyze the “opaque billboard” failure mode, where incorrect geometry achieves optimal appearance, we establish three minimal conditions.

Assumption 2 (Visible texture mode). *There exists a visible surface patch and a local tangential coordinate u such that the true diffuse texture $T(u)$ contains a sinusoidal component $a \sin(ku)$ with amplitude $a > 0$ and spatial frequency $k > 0$.*

Assumption 3 (Visible misalignment magnitude). *On a visible set of rays, the candidate surface U is misaligned from the true surface S by a non-negligible depth gap $\Delta z > 0$.*

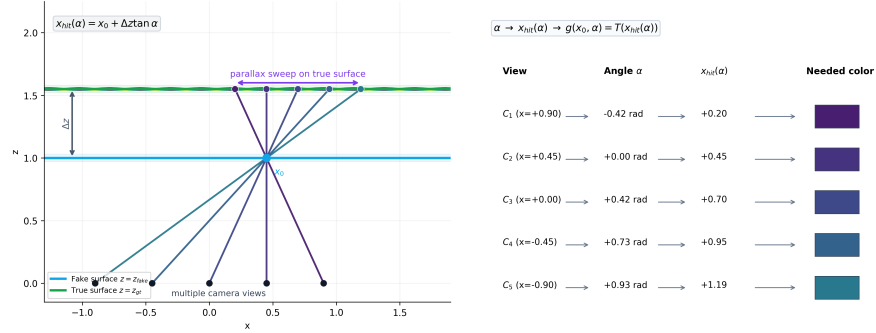
Assumption 4 (Residual cannot erase texture). *The true specular residual R is not strong enough to adversarially cancel the spatial texture: there exists $0 \leq \rho < 1$ such that $|R(x, \hat{v})| \leq \rho a$.*

These assumptions collectively define the basic requirements for geometry to be visually identifiable from RGB alone. If the scene lacks texture entirely (Assumption 2 fails) or the residual completely washes out the diffuse color (Assumption 4 fails), the optimization lacks the necessary signal to distinguish depth. If the depth gap is negligible (Assumption 3 fails), the candidate surface effectively behaves like the true surface.

Material-model limitation. Assumption 4 excludes view-dependent effects that can cancel or dominate diffuse texture, including mirrors, inter-reflections, transparency, exposure/white-balance changes, and nonlinear tone mapping. Then the residual may itself carry high angular frequency, breaking our spatial-texture/angular-residual separation.

6.2 The Frequency Transport Law via Parallax

To see exactly how angular complexity is generated by geometric error, consider a standard 2D cross-section for parallax analysis. Place the camera at the origin, let the true surface be the line $z = z_{\text{gt}}$, and let the candidate surface be a billboard at $z = z_{\text{fake}}$, yielding a depth gap $\Delta z := z_{\text{gt}} - z_{\text{fake}} > 0$. We parameterize the camera rays by a sweeping angle α . Equivalently, assume a moving camera observing a fixed physical point x_0 on the fake surface from different view angles α . See Fig. 2 for a visual illustration of both the geometric transport and the induced view-to-color mapping.



(a) Parallax geometry under a misaligned surface. Rays from different camera centers all intersect the same fake point x_0 on $z = z_{\text{fake}}$, but correspond to different true hit points on $z = z_{\text{gt}}$. The transported hit location follows $x_{\text{hit}}(\alpha) = x_0 + \Delta z \tan \alpha$, producing a sweep over the true surface. (b) View-to-color mapping at a fixed fake point. For each camera/view angle α , the required color at x_0 is the true texture evaluated at the transported point, i.e., $g(x_0, \alpha) = T(x_{\text{hit}}(\alpha))$. Thus, wrong geometry forces one 3D point to encode different colors across views.

Fig. 2: Two-panel visualization of parallax-induced frequency transport under misaligned geometry.

A direct geometric computation (supplementary material) reveals that if a ray first hits the wrong surface at lateral coordinate x_0 , the physical hit point on the true background surface moves according to:

$$x_{\text{hit}}(\alpha) = x_0 + \Delta z \tan \alpha. \quad (4)$$

Consequently, when the model evaluates the loss on the wrong surface, its through-seen target function F_U must pull the color from the background texture $T(x_{\text{hit}}(\alpha))$ (the residual term R is handled via Assumption 4 and the supplementary material). Using our assumed texture component and approximating $\tan \alpha \approx \alpha$ for moderate viewing angles, the target signal on the wrong surface becomes:

$$a \sin(k(x_0 + \Delta z \alpha)) = a \sin(kx_0 + (k\Delta z)\alpha).$$

This reveals a profound physical transformation: the *spatial* frequency k has been forcefully multiplied by the depth error Δz to generate a new *angular* frequency:

$$\omega_{\text{fake}} \approx k\Delta z. \quad (5)$$

This scale law dictates the appearance burden placed on wrong geometry: finer textures or larger depth errors demand increasingly rapid angular color shifts.

6.3 The Immutable Lower Bound for Wrong Geometry

This induced angular frequency creates a mathematical bottleneck. As established earlier, any SH basis of degree L acts as a low-pass filter along viewing arcs.

Lemma 2 (Bandlimit along great circles). *If $h \in \mathcal{V}_L$, restricting h to any great circle yields a 1D trigonometric polynomial with maximum frequency L .*

When the induced frequency exceeds the model’s capacity ($L < \omega_{\text{fake}}$), the model is mathematically incapable of fitting the signal. For analytical clarity, we use a global Fourier bound rather than relying only on local linearization. This is governed by a fundamental property of Fourier approximation (proven in the supplementary material):

Lemma 3 (High frequency cannot be fit by low frequency). *Suppose the frequency ω is an integer. Consider $f(\phi) = a \sin(\omega\phi)$ on $[0, 2\pi]$. The best L^2 approximation of f by any trigonometric polynomial of degree $L < \omega$ leaves an unavoidable residual error strictly bounded away from zero.*

Because the specular residual cannot completely erase the texture (Assumption 4), we can invoke the triangle inequality on the function spaces to guarantee that the projection error of the combined signal remains large. This culminates in our strict lower bound for misaligned geometry.

Here $a \asymp b$ denotes equality up to constants independent of L .

Proposition 3 (Misaligned-surface lower bound). *Under Assumption 2, Assumption 3, and Assumption 4, there exist $\omega_{\text{fake}} \asymp k\Delta z$ and a constant $c_0 > 0$ such that for all angular capacities $L < \omega_{\text{fake}}$,*

$$B_U(L) \geq c_0.$$

The proof mechanically links our findings: the through-seen function on a misaligned patch contains an angular sinusoid of frequency $\omega_{\text{fake}} \approx k\Delta z$. By Lemma 2, the $\mathcal{V}_L$ appearance function cannot exceed frequency L along this viewing circle. Thus, whenever $L < \omega_{\text{fake}}$, Lemma 3 ensures a massive, unoptimizable constant error on this patch, which lower-bounds the global integral $B_U(L)$. We now possess the two essential halves of the identifiability story: the true surface benefits from a fast-decaying upper bound $A(L)$, while the wrong surface suffers from a stubborn lower bound $B_U(L)$. By synthesizing these bounds, we can explicitly characterize the intermediate capacity window that successfully recovers accurate 3D geometry.

7 Putting It Together: The Identifiability Window

We are finally positioned to synthesize the two halves of our analysis. By juxtaposing the rapid decay of the true surface’s approximation error with the stubborn constant lower bound of the misaligned surface, the theoretical mechanism behind 3DGS geometric identifiability emerges clearly (see the supplementary material for a numeric example).

Theorem 1 (Intermediate- L identifiability window). *Under Assumption 2, Assumption 3, and Assumption 4, there exist constants $c_0 > 0$ and $\omega_{\text{fake}} \asymp k\Delta z$ such that: (i) for all $L \geq 0$, $A(L) \leq (1+L(L+1))^{-s_{\text{true}}} M_{\text{true}}^2$ (up to view-sampling constants); and (ii) for all $L < \omega_{\text{fake}}$, $B_U(L) \geq c_0$. Therefore, any L satisfying $(1+L(L+1))^{-s_{\text{true}}} M_{\text{true}}^2 < c_0$ and $L < \omega_{\text{fake}}$ guarantees a strict separation $A(L) < B_U(L)$: the true surface is strictly easier to fit than the misaligned surface U under angular capacity L .*

This theorem rigorously partitions the angular capacity of 3DGS into three distinct regimes. When L is too small, the model cannot fit the true specular residual R , yielding high error for all candidate geometries. In the *intermediate regime*, $A(L)$ collapses as the true smooth residual is approximated, yet $B_U(L)$ remains trapped by the $L < \omega_{\text{fake}}$ barrier, creating the optimal geometry-identifiable window. However, once $L \gtrsim \omega_{\text{fake}}$, this lower-bound mechanism vanishes, clarifying the opaque billboard paradox introduced earlier. As proven in the supplementary material, the SH basis becomes expressive enough to memorize the parallax-induced false texture, allowing a misaligned surface to achieve exactly zero population loss. In this over-parameterized regime, the optimizer is mathematically permitted to “cheat”: it can erect an arbitrary, vision-blocking sheet in empty space and task the high-degree Spherical Harmonics with projecting the missing background spatial frequencies as rapid, view-dependent color shifts.

8 Experiment

Datasets. We construct a synthetic dataset of 100 shapes (25 base geometries $\times$ 4 variants) to systematically trigger and evaluate the failure modes predicted in Sec. 5–7. Real-world benchmarks typically feature rich, high-frequency spatial textures (large k), which naturally drive the required angular frequency $\omega_{\text{fake}} \approx k\Delta z$ beyond typical SH capacities. By using synthetic shapes with tunable complexity under controlled lighting, we can deliberately evaluate the vulnerability to billboard collapse.

We additionally evaluate the same diagnostic on ten real-scene datasets with available geometry where 3DGS can be trained (Section 8.2). These real-world experiments probe whether opaque billboard failures arise in practice when using high SH degrees under standard multi-view capture protocols. Under our labeling criterion, we find no opaque-billboard cases: across these datasets the learned

Gaussians remain surface-consistent even at large angular capacity, in line with the “safe” regime predicted by our theory.

We train 3DGS with default hyperparameters and vary only the SH degree L . Geometry is evaluated by normalized symmetric Chamfer distance (nCD) between Gaussian centers and GT surfaces; appearance by PSNR. We define “opaque billboard” as runs with near-best PSNR but severely degraded nCD; full protocol and thresholds and the definitions of billboard failure, billboard onset, outlier point removal are in the supplementary material.

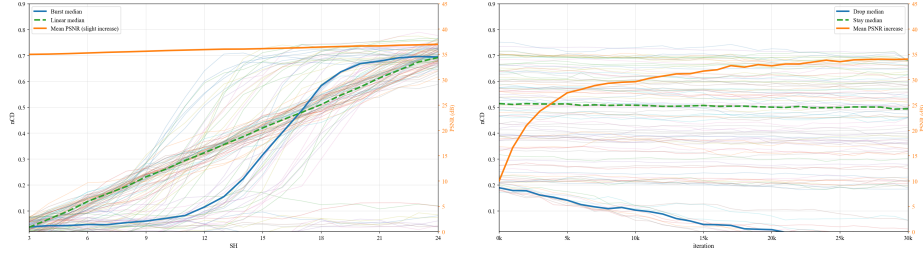


Fig. 3: Left: nCD vs. SH degree. Thin lines: per-shape nCD (left y-axis, lower is better); orange: mean PSNR (right y-axis). Solid blue: “burst” median; dashed green: “linear” median. **Right: nCD vs. iteration (SH=24).** Thin lines: per-shape nCD (left y-axis, lower is better); orange: mean PSNR (right y-axis). Solid blue: “drop” median; dashed green: “stay” median.

8.1 Opaque billboard failure at large angular capacity

Section 7 predicts that when the angular capacity L exceeds the induced frequency threshold ω_{fake} , opaque billboard solutions can perfectly minimize the training objective. Our synthetic experiments strongly support this prediction: when L is made sufficiently large, we frequently observe opaque billboard failures, where 3DGS no longer reconstructs the true surface but still produces high-quality renderings. The observed dense layers are consistent with our first-hit regime, but do not imply all 3DGS rendering behaves this way. As shown in Fig. 3, we observe three consistent patterns. (a) As the SH degree increases, approximately 89% of shapes generally increase in nCD between the Gaussian centers and the ground-truth surface, while the average PSNR slightly increases. For these shapes, the high-SH runs are labeled as either billboard-onset or opaque billboard under our criterion: nCD grows to several times nCD_{best} even though PSNR stays within 2 dB of $\text{PSNR}_{\text{best}}$. This indicates that as the angular capacity grows, the Gaussians gain enough representational power that they no longer need to stay on the surface to provide good renderings. (b) Among the shapes that drift away from the surface, we see two dominant behaviors, which correspond to the two medians in Fig. 3. For roughly 42% of cases, nCD grows almost linearly with SH degree (the dashed green “linear” median), leading to billboard-onset

Table 1: Geometry and appearance metrics across real-scene datasets We report $\text{nCD} := \text{CD}/\text{diag}(\text{bbox}(\text{GT}))$ (lower is better) and PSNR (higher is better). All values are our measurements; citations point to dataset sources.

Dataset	nCD@SH=3	PSNR@SH=3	nCD@SH=24	PSNR@SH=24
DTU [1]	0.0090	31.6	0.0100	32.3
MobileBrick [23]	0.0600	28.9	0.0700	29.5
YCB-Video [44]	0.0420	28.1	0.0500	28.8
T-LESS [17]	0.0650	27.6	0.0750	28.1
ScanNet [8]	0.0110	29.3	0.0120	29.9
ScanNet++ [46]	0.0070	29.8	0.0080	30.3
Matterport3D [6]	0.0045	30.4	0.0050	31.0
HM3D [31]	0.0047	30.8	0.0053	31.4
Tanks & Temples [21]	0.0080	29.0	0.0090	29.7
ETH3D [34]	0.0090	30.1	0.0100	30.6
Average	0.0220	29.6	0.0254	30.2

or opaque-billboard labels only at the highest SH degrees. For roughly 53% of cases, nCD remains small up to a moderate SH degree and then undergoes a sharp increase once SH enters the range 6–12 (the solid blue “burst” median). In these “burst” cases the high-SH runs are almost always labeled as opaque billboard: beyond a certain angular capacity, the Gaussians can fit the appearance almost entirely by adjusting their SH coefficients without moving in space. (c) Finally, Fig. 3 also shows that, for a subset of objects, no opaque-billboard regime appears: all runs for these shapes remain surface-consistent under our labeling criterion. Even at high SH degrees, nCD stays within $2 \times \text{nCD}_{\text{best}}$ and the learned Gaussians stay close to the ground-truth surface. These cases match the “safe” identifiability window predicted by our theory, where higher spatial complexity (large k) pushes the required angular frequency beyond our tested capacities, which effectively mitigates this failure mode in practical settings.

8.2 Real-scene experiments

Do opaque billboard failures occur on real objects under standard capture protocols? To answer this question, we train 3DGS with varying SH degrees on ten real-scene datasets with ground-truth geometry where 3DGS can be applied: DTU [1], MobileBrick [23], YCB-Video [44], T-LESS [17], ScanNet [8], ScanNet++ [46], Matterport3D [6], HM3D [31], Tanks & Temples [21], and ETH3D [34] (Table 1). For each dataset we apply the same billboard labeling criterion as in the synthetic experiments.

Across all objects and SH degrees, we do not observe any run that satisfies the opaque-billboard condition: no run simultaneously has PSNR within 2 dB of the per-object $\text{PSNR}_{\text{best}}$ and nCD at least $5 \times \text{nCD}_{\text{best}}$. Under this PSNR band, nCD remains tightly concentrated: as summarized in Table 1, for every dataset the nCD at $\text{SH} = 24$ differs from that at $\text{SH} = 3$ by at most a factor of about

1.2 $\times$. In other words, increasing angular capacity in these real-scene settings does not lead to the large geometric drifts that characterize opaque billboard failures.

Thus, within standard multi-view capture protocols on these diverse real-world benchmarks, the learned Gaussians remain surface-consistent according to our criterion. This empirical finding is consistent with our theoretical frequency scale law: typical real-scene setups are characterized by rich spatial textures (very large k). This naturally drives the parallax-induced frequency $\omega_{\text{fake}} \approx k\Delta z$ so high that even large capacities like $\text{SH} = 24$ cannot bridge the gap ($L \ll \omega_{\text{fake}}$). Consequently, the model remains within the geometry-identifiable window in our tested settings, favoring surface-consistent solutions over billboard collapse.

9 Conclusion & Discussion

Increasing angular capacity is a double-edged sword: it helps fit specularities, but can let misaligned geometry explain away parallax-induced texture. This yields an intermediate- L identifiability window and supports capacity scheduling, e.g., starting with low L . Our guarantees apply to mostly opaque scenes with smooth view-dependent residuals; mirror-like, transparent, or strongly camera-processed scenes may require geometry or material priors.

Effect of geometric regularization. Our analysis studies the image reconstruction objective with bounded angular radiance capacity. Geometry-aware variants such as surface-aligned Gaussian methods, 2D Gaussian formulations, SDF-coupled methods, or normal/depth regularization modify this problem by either restricting the admissible geometry class or adding explicit penalties against off-surface solutions. Such priors can shrink or remove the billboard-favorable regime predicted by our theorem. Thus, our results should not be read as a claim that geometry-aware methods fail in the same way; rather, they explain why such geometric priors are valuable when angular appearance capacity is large.

References

1. Aanæs, H., Jensen, R.R., Vogiatzis, G., Tola, E., Dahl, A.B.: Large-scale data for multiple-view stereopsis. *International Journal of Computer Vision* **120**(2), 153–168 (2016). <https://doi.org/10.1007/s11263-016-0902-9>
2. Bao, Y., Ding, T., Huo, J., Liu, Y., Li, Y., Li, W., Gao, Y., Luo, J.: 3d gaussian splatting: Survey, technologies, challenges, and opportunities. *IEEE Transactions on Circuits and Systems for Video Technology* **35**(7), 6832–6852 (2025). <https://doi.org/10.1109/TCSVT.2025.3538684>
3. Borycki, P., Smolak-Dyżewska, W., Waczyńska, J., Mazur, M., Tadeja, S., Spurek, P.: GASP: Gaussian splatting for physics-based simulations. *Computer Vision and Image Understanding* **265**, 104699 (2026). <https://doi.org/10.1016/j.cviu.2026.104699>
4. Buehler, C., Bosse, M., McMillan, L., Gortler, S., Cohen, M.: Unstructured lumigraph rendering. In: *Proceedings of the 28th Annual Conference on Computer Graphics and Interactive Techniques*. pp. 425–432 (2001). <https://doi.org/10.1145/383259.383309>

5. Chai, J.X., Tong, X., Chan, S.C., Shum, H.Y.: Plenoptic sampling. In: Proceedings of the 27th Annual Conference on Computer Graphics and Interactive Techniques. pp. 307–318 (2000). <https://doi.org/10.1145/344779.344932>
6. Chang, A., Dai, A., Funkhouser, T., Halber, M., Nießner, M., Savva, M., Song, S., Zeng, A., Zhang, Y.: Matterport3D: Learning from RGB-D data in indoor environments. In: 2017 International Conference on 3D Vision (3DV). pp. 667–676 (2017). <https://doi.org/10.1109/3DV.2017.00081>
7. Chen, D., Li, H., Ye, W., Wang, Y., Xie, W., Zhai, S., Wang, N., Liu, H., Bao, H., Zhang, G.: PGSR: Planar-based gaussian splatting for efficient and high-fidelity surface reconstruction. *IEEE Transactions on Visualization and Computer Graphics* **31**(9), 6100–6111 (2025). <https://doi.org/10.1109/TVCG.2024.3494046>
8. Dai, A., Chang, A.X., Savva, M., Halber, M., Funkhouser, T., Nießner, M.: ScanNet: Richly-annotated 3d reconstructions of indoor scenes. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). pp. 5828–5839 (2017). <https://doi.org/10.1109/CVPR.2017.261>
9. Dai, P., Xu, J., Xie, W., Liu, X., Wang, H., Xu, W.: High-quality surface reconstruction using gaussian surfels. In: ACM SIGGRAPH 2024 Conference Papers. pp. 1–11 (2024). <https://doi.org/10.1145/3641519.3657441>
10. Do, M.N., Marchand-Maillet, S., Vetterli, M.: On the bandwidth of the plenoptic function. *IEEE Transactions on Image Processing* **21**(2), 708–717 (2012). <https://doi.org/10.1109/TIP.2011.2163895>
11. Durand, F., Holzschuch, N., Soler, C., Chan, E., Sillion, F.X.: A frequency analysis of light transport. *ACM Transactions on Graphics* **24**(3), 1115–1126 (2005). <https://doi.org/10.1145/1073204.1073320>
12. Fang, Q., Song, Y., Li, K., Bo, L.: Reducing shape-radiance ambiguity in radiance fields with a closed-form color estimation method. In: Advances in Neural Information Processing Systems (NeurIPS). pp. 58902–58920 (2023). <https://doi.org/10.52202/075280-2570>
13. Feng, Y., Feng, X., Shang, Y., Jiang, Y., Yu, C., Zong, Z., Shao, T., Wu, H., Zhou, K., Jiang, C., Yang, Y.: Gaussian Splashing: Unified particles for versatile motion synthesis and rendering. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). pp. 518–529 (2025). <https://doi.org/10.1109/CVPR52734.2025.00057>
14. Gortler, S.J., Grzeszczuk, R., Szeliski, R., Cohen, M.F.: The lumigraph. In: Proceedings of the 23rd Annual Conference on Computer Graphics and Interactive Techniques. pp. 43–54 (1996). <https://doi.org/10.1145/237170.237200>
15. Guédon, A., Lepetit, V.: SuGaR: Surface-aligned gaussian splatting for efficient 3d mesh reconstruction and high-quality mesh rendering. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). pp. 5354–5363 (2024). <https://doi.org/10.1109/CVPR52733.2024.00512>
16. Guo, M., Chiang, M.J.Y., Santesteban, I., Sarafianos, N., Chen, H.Y., Halimi, O., Božič, A., Saito, S., Wu, J., Liu, C.K., Stuyck, T., Larionov, E.: PGC: Physics-based gaussian cloth from a single pose. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). pp. 21215–21225 (2025). <https://doi.org/10.1109/CVPR52734.2025.01976>
17. Hodaň, T., Haluza, P., Obdržálek, Š., Matas, J., Lourakis, M., Zabulis, X.: T-LESS: An RGB-D dataset for 6d pose estimation of texture-less objects. In: 2017 IEEE Winter Conference on Applications of Computer Vision (WACV). pp. 880–888 (2017). <https://doi.org/10.1109/WACV.2017.103>

18. Huang, B., Yu, Z., Chen, A., Geiger, A., Gao, S.: 2d gaussian splatting for geometrically accurate radiance fields. In: ACM SIGGRAPH 2024 Conference Papers. pp. 1–11 (2024). <https://doi.org/10.1145/3641519.3657428>
19. Jiang, Y., Yu, C., Xie, T., Li, X., Feng, Y., Wang, H., Li, M., Lau, H., Gao, F., Yang, Y., Jiang, C.: VR-GS: A physical dynamics-aware interactive gaussian splatting system in virtual reality. In: ACM SIGGRAPH 2024 Conference Papers. pp. 1–11 (2024). <https://doi.org/10.1145/3641519.3657448>
20. Kerbl, B., Kopanas, G., Leimkühler, T., Drettakis, G.: 3d gaussian splatting for real-time radiance field rendering. *ACM Transactions on Graphics* **42**(4), 1–14 (2023). <https://doi.org/10.1145/3592433>
21. Knapitsch, A., Park, J., Zhou, Q.Y., Koltun, V.: Tanks and temples: Benchmarking large-scale scene reconstruction. *ACM Transactions on Graphics* **36**(4), 78:1–78:13 (2017). <https://doi.org/10.1145/3072959.3073599>
22. Levoy, M., Hanrahan, P.: Light field rendering. In: Proceedings of the 23rd Annual Conference on Computer Graphics and Interactive Techniques. pp. 31–42 (1996). <https://doi.org/10.1145/237170.237199>
23. Li, K., Bian, J.W., Castle, R., Torr, P.H.S., Prisacariu, V.A.: MobileBrick: Building LEGO for 3d reconstruction on mobile devices. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). pp. 4892–4901 (2023). <https://doi.org/10.1109/CVPR52729.2023.00474>
24. Lin, Z., Shum, H.Y.: A geometric analysis of light field rendering. *International Journal of Computer Vision* **58**(2), 121–138 (2004). <https://doi.org/10.1023/B:VISI.0000015916.91741.27>
25. Luo, J., Huang, T., Wang, W., Feng, W.: A review of recent advances in 3d gaussian splatting for optimization and reconstruction. *Image and Vision Computing* **151**, 105304 (2024). <https://doi.org/10.1016/j.imavis.2024.105304>
26. Lyu, X., Sun, Y.T., Huang, Y.H., Wu, X., Yang, Z., Chen, Y., Pang, J., Qi, X.: 3DGSR: Implicit surface reconstruction with 3d gaussian splatting. *ACM Transactions on Graphics* **43**(6), 1–12 (2024). <https://doi.org/10.1145/3687952>
27. Ma, S., Luo, Y., Yang, W., Yang, Y.: MaGS: Reconstructing and simulating dynamic 3d objects with mesh-adsorbed gaussian splatting. In: Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV). pp. 8745–8755 (2025). <https://doi.org/10.1109/ICCV51701.2025.00818>
28. Mildenhall, B., Srinivasan, P.P., Tancik, M., Barron, J.T., Ramamoorthi, R., Ng, R.: NeRF: Representing scenes as neural radiance fields for view synthesis. In: *Computer Vision – ECCV 2020*. pp. 405–421. Springer (2020). https://doi.org/10.1007/978-3-030-58452-8_24
29. Müller, M., Heidelberg, B., Hennix, M., Ratcliff, J.: Position based dynamics. *Journal of Visual Communication and Image Representation* **18**(2), 109–118 (2007). <https://doi.org/10.1016/j.jvcir.2007.01.005>
30. Nguyen, H.T., Do, M.N.: Error analysis for image-based rendering with depth information. *IEEE Transactions on Image Processing* **18**(4), 703–716 (2009). <https://doi.org/10.1109/TIP.2009.2012884>
31. Ramakrishnan, S.K., Gokaslan, A., Wijmans, E., Maksymets, O., Clegg, A., Turner, J., Undersander, E., Galuba, W., Westbury, A., Chang, A.X., Savva, M., Zhao, Y., Batra, D.: Habitat-Matterport 3d dataset (HM3D): 1000 large-scale 3d environments for embodied AI. In: *Advances in Neural Information Processing Systems (NeurIPS) Datasets and Benchmarks Track* (2021)
32. Ramamoorthi, R., Hanrahan, P.: An efficient representation for irradiance environment maps. In: Proceedings of the 28th Annual Conference on Computer

- Graphics and Interactive Techniques. pp. 497–500 (2001). <https://doi.org/10.1145/383259.383317>
33. Rong, B., Grigorev, A., Wang, W., Black, M.J., Thomaszewski, B., Tsalicoglou, C., Hilliges, O.: Gaussian Garments: Reconstructing simulation-ready clothing with photorealistic appearance from multi-view video. In: International Conference on 3D Vision (3DV). pp. 1054–1063 (2025). <https://doi.org/10.1109/3DV66043.2025.00102>
 34. Schöps, T., Schönberger, J.L., Galliani, S., Sattler, T., Schindler, K., Pollefeys, M., Geiger, A.: A multi-view stereo benchmark with high-resolution images and multi-camera videos. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). pp. 3260–3269 (2017). <https://doi.org/10.1109/CVPR.2017.272>
 35. Shen, Z., Liu, Y., Chen, Z., Li, Z., Wang, J., Liang, Y., Yu, Z., Zhang, J., Xu, Y., Schaefer, S., Li, X., Wang, W.: SolidGS: Consolidating gaussian surfel splatting for sparse-view surface reconstruction. arXiv preprint arXiv:2412.15400 (2024), <https://arxiv.org/abs/2412.15400>, accessed 30 June 2026
 36. Sloan, P.P., Kautz, J., Snyder, J.: Precomputed radiance transfer for real-time rendering in dynamic, low-frequency lighting environments. In: Proceedings of the 29th Annual Conference on Computer Graphics and Interactive Techniques. pp. 527–536 (2002). <https://doi.org/10.1145/566570.566612>
 37. Vasile, F., Qiu, R.Z., Natale, L., Wang, X.: Gaussian-augmented physics simulation and system identification with complex colliders. arXiv preprint arXiv:2511.06846 (2025), <https://arxiv.org/abs/2511.06846>, accessed 30 June 2026
 38. Verbin, D., Hedman, P., Mildenhall, B., Zickler, T., Barron, J.T., Srinivasan, P.P.: Ref-NeRF: Structured view-dependent appearance for neural radiance fields. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). pp. 5481–5490 (2022). <https://doi.org/10.1109/CVPR52688.2022.00541>
 39. Wang, J., Liu, Y., Wang, P., Lin, C., Hou, J., Li, X., Komura, T., Wang, W.: GausSurf: Geometry-guided 3d gaussian splatting for surface reconstruction. arXiv preprint arXiv:2411.19454 (2024), <https://arxiv.org/abs/2411.19454>, accessed 30 June 2026
 40. Wang, L., Zhou, F., Yu, B., Cao, P., Yin, J.: OMEGAS: Object mesh extraction from large scenes guided by gaussian segmentation. IEEE Transactions on Circuits and Systems for Video Technology **36**(1), 792–802 (2026). <https://doi.org/10.1109/TCSVT.2025.3586755>
 41. Wolf, Y., Bracha, A., Kimmel, R.: GS2Mesh: Surface reconstruction from gaussian splatting via novel stereo views. In: Computer Vision – ECCV 2024. pp. 207–224. Springer (2025). https://doi.org/10.1007/978-3-031-73024-5_13
 42. Wu, Q., Zheng, J., Cai, J.: Surface reconstruction from 3d gaussian splatting via local structural hints. In: Computer Vision – ECCV 2024. Lecture Notes in Computer Science, vol. 15060, pp. 441–458. Springer (2025). https://doi.org/10.1007/978-3-031-72627-9_25
 43. Wu, T., Yuan, Y.J., Zhang, L.X., Yang, J., Cao, Y.P., Yan, L.Q., Gao, L.: Recent advances in 3d gaussian splatting. Computational Visual Media **10**(4), 613–642 (2024). <https://doi.org/10.1007/s41095-024-0436-y>
 44. Xiang, Y., Schmidt, T., Narayanan, V., Fox, D.: PoseCNN: A convolutional neural network for 6d object pose estimation in cluttered scenes. In: Robotics: Science and Systems (RSS) (2018). <https://doi.org/10.15607/RSS.2018.XIV.019>

- 45. Xie, T., Zong, Z., Qiu, Y., Li, X., Feng, Y., Yang, Y., Jiang, C.: PhysGaussian: Physics-integrated 3d gaussians for generative dynamics. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). pp. 4389–4398 (2024). <https://doi.org/10.1109/CVPR52733.2024.00420>
- 46. Yeshwanth, C., Liu, Y.C., Nießner, M., Dai, A.: ScanNet++: A high-fidelity dataset of 3d indoor scenes. In: Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV). pp. 12–22 (2023). <https://doi.org/10.1109/ICCV51070.2023.00008>
- 47. Zhang, K., Riegler, G., Snavely, N., Koltun, V.: NeRF++: Analyzing and improving neural radiance fields. arXiv preprint arXiv:2010.07492 (2020), <https://arxiv.org/abs/2010.07492>, accessed 30 June 2026