TSPE-GS: Probabilistic Depth Extraction for Semi-Transparent Surface Reconstruction via 3D Gaussian Splatting

Zhiyuan Xu[1,2], **Min Nan**[1], **Yuhang Guo**[1], **Tong Wei**[1,2*]
[1]School of Computer Science and Engineering, Southeast University, Nanjing, China
[2]Key Laboratory of Computer Network and Information Integration (Southeast University),
Ministry of Education, China
{xuzhiy, mn614, gyh_seu, weit}@seu.edu.cn

Abstract

3D Gaussian Splatting-based geometry reconstruction is regarded as an excellent paradigm due to its favorable tradeoff between speed and reconstruction quality. However, such 3D Gaussian-based reconstruction pipelines often face challenges when reconstructing semi-transparent surfaces, hindering their broader application in real-world scenes. Mainstream methods assume one depth per pixel; under semi-transparency this causes depth ambiguity and reconstruction failure. To address these challenges, we propose TSPE-GS (Transparent Surface Probabilistic Extraction for Gaussian Splatting), a novel probabilistic depth extraction approach that uniformly samples transmittance to model the multimodal distribution of opacity and depth per pixel, replacing the previous single-peak distribution that caused depth confusion across surfaces. We fuse TSDFs progressively to recover separate outer and inner surfaces. Our method can be easily generalized to other Gaussian-based reconstruction pipelines, effectively extracting semi-transparent surfaces without requiring additional training overhead. Extensive experiments on both public and self-collected semi-transparent datasets, as well as opaque object datasets, demonstrate that TSPE-GS significantly enhances reconstruction accuracy for semi-transparent surfaces while maintaining reconstruction quality in opaque scenes.

Code — https://github.com/nortonii/TSPE-GS
Extended version — https://arxiv.org/abs/2511.09944

Introduction

Geometric reconstruction represent fundamental challenges in graphics, essential for advancing applications such as augmented reality (AR), virtual reality (VR), autonomous driving (Zhou et al. 2024; Bruno et al. 2010). Recent research has seen the rise of Neural Radiance Fields (NeRF) as a leading approach for achieving precise geometric reconstruction (Mildenhall et al. 2021). Despite their effectiveness, NeRF models rely on volume rendering techniques that necessitate dense point sampling along rays, resulting in significant computational overhead and limiting real-time applications.

Recently, 3D Gaussian Splatting (3DGS) (Kerbl et al. 2023) has emerged as a promising method for surface reconstruction, representing complex scenes as a set of 3D Gaussians. This technique has rapidly advanced to include surface reconstruction (Guédon and Lepetit 2024; Huang et al. 2024; Yu, Sattler, and Geiger 2024; Chen et al. 2024). Notably, the SuGaR method (Guédon and Lepetit 2024) aligns 3D Gaussians with surfaces and employs Poisson surface reconstruction (Kazhdan and Hoppe 2013) to extract meshes from depth maps. However, Poisson reconstruction has significant limitations, including the neglect of opacity and scale of Gaussian primitives, as well as unreliable depth maps that hinder real-time processing. To overcome these challenges, truncated signed distance functions (TSDF) (Curless and Levoy 1996) have been utilized for mesh reconstruction, extracting depth information by assessing the visibility of different Gaussian primitives. This approach encodes the distance to the nearest surface at localized points, allowing for integration of depth data from multiple viewpoints and facilitating continuous, smooth surface representations.

Despite the advancements in surface reconstruction using 3DGS, Reconstructing semi-transparent scenes remains challenging, both for the semi-transparent surfaces and the occluded objects behind them (Li et al. 2025). These limitations are crucial for applications like robotic laboratory systems that require millimeter-precise manipulation of glassware, such as beakers and test tubes or tasks involving the retrieval of items from semi-transparent containers, such as food wrapped in cling film or objects inside plastic bags (Sajjan et al. 2020). The first challenge in reconstructing semi-transparent surfaces arises from the common assumption in most existing depth extraction methods that there is only a single visible surface per ray. Consequently, these methods approximate the surface depth either as the weighted average (expected depth) (Chen et al. 2024) or a representative value (such as the median depth) (Zhang et al. 2024) of the visible Gaussian primitives' depths. However, when multiple surfaces are visible simultaneously, the weighted average tends to fall between the actual surfaces, causing ambiguity. The second challenge arises from surface extraction methods utilizing TSDF depth fusion. If an object is sealed inside a plastic bag, even when employing current semi-transparent surface reconstruction techniques

---

[*]Corresponding author.

to extract the outermost depth from all viewpoints for TSDF fusion, the internal object remains unreconstructable, despite being clearly visible, as it does not exhibit depth from any viewpoint.

However, a simple yet effective approach exists that can simultaneously address the aforementioned two challenges. We argue that 3DGS can represent semi-transparent and occluded surfaces accurately, but current depth extraction methods underutilize this capability. The limitation lies primarily in the depth extraction methods, which have yet to fully exploit this potential. Specifically, the geometric expressiveness of 3DGS can be described as a mapping between visibility and depth values observed from multiple viewpoints. Existing depth extraction techniques have typically considered only the integral value (expected depth) or the median value of this mapping, overlooking the richer information contained within the full distribution. To address this, we propose an efficient method dubbed Transparent Surface Probabilistic Extraction for Gaussian Splatting (TSPE-GS). TSPE-GS captures the full visibility-depth mapping and applies a tailored analysis and processing strategy to simultaneously identify multiple visible surface depths. By leveraging tailored TSDF fusion across these depths, TSPE-GS enables concurrent and accurate reconstruction of both semi-transparent surfaces and occluded internal objects, as illustrated in Figure 1. We evaluate TSPE-GS on the BMVS, DTU, and αSurf datasets (Yao et al. 2020; Aanæs et al. 2016; Wu et al. 2025), along with our custom dataset designed specifically to include objects occluded by semi-transparent surfaces. Results show that TSPE-GS maintains reconstruction accuracy on general scenes (BMVS and DTU) without any degradation, while significantly improving reconstruction quality in semi-transparent surface scenarios (αSurf and our dataset). Our method is simple and efficient, imposing negligible computational overhead relative to the highly optimized 3DGS rendering process. In summary, our contributions are as follows:

- We are the first to identify the challenge of robustly extracting objects occluded by semi-transparent surfaces within the 3DGS depth estimation reconstruction framework.
- We propose a method to address this challenge by introducing a probabilistic depth extraction technique combined with a progressive TSDF fusion strategy that effectively reconstructs multi-layer surfaces.
- Extensive experiments show that our method significantly outperforms state-of-the-art approaches in accuracy and completeness on both semi-transparent and opaque datasets, demonstrating its effectiveness and robustness.

Related Work

Multi-view Stereo Surface Reconstruction Multi-view stereo (MVS) depth estimation is a classical problem in computer vision and graphics (Furukawa and Ponce 2009; Barnes et al. 2009; Bleyer, Rhemann, and Rother 2011; Schonberger and Frahm 2016; Goesele et al. 2007; Hirschmuller 2007). Traditional methods extract dense point clouds from multi-view images through patch-based matching (Barnes et al. 2009) and reconstruct surfaces via triangulation (Boissonnat and Teillaud 2006). While achieving high geometric accuracy, they often fail in weakly textured regions. Recent end-to-end neural networks (Yao et al. 2018; Kendall et al. 2017; Luo, Schwing, and Urtasun 2016; Gu et al. 2020; Ding et al. 2022) improve MVS by learning feature representations, but require costly volumetric representations and training with ground truth depth, limiting practical deployment in applications like robotics, AR/VR, and large-scale 3D reconstruction.

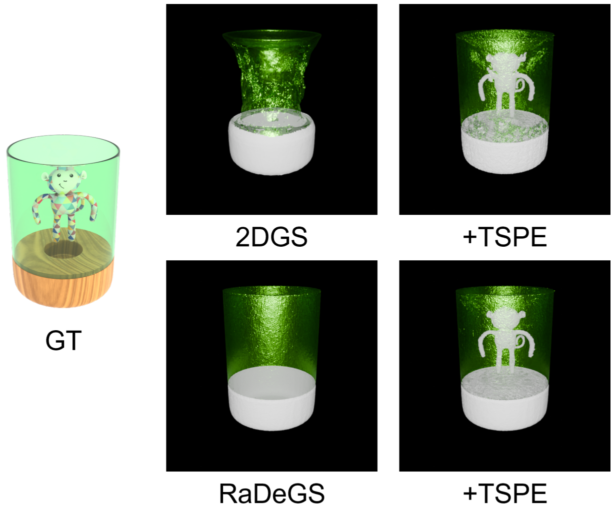

Figure 1: Visual comparison of reconstruction on a semi-transparent object in αSurf dataset with and without TSPE.

Neural Surface Reconstruction Unlike implicit methods requiring depth supervision, NeRF (Mildenhall et al. 2021) models scenes via MLPs mapping 5D coordinates to color and density, enabling photorealistic novel view synthesis. Extensions integrate signed distance functions (SDFs) (Hart, Sandin, and Kauffman 1989) for mesh extraction via marching cubes (Li et al. 2023; Yu et al. 2021; Wang et al. 2021, 2022; Yariv et al. 2021), and NeuS (Wang et al. 2021) bridged volume rendering with SDFs for unbiased surfaces. Neuralangelo (Li et al. 2023) further improved large-scale reconstruction using multiresolution hash grids. To represent open surfaces, unsigned distance fields (UDFs) were proposed (Liu et al. 2023; Long et al. 2023). Other works incorporate multi-view stereo principles by adding photometric consistency losses to enhance geometry (Ding et al. 2022; Fu et al. 2022), but NeRF-based methods remain computationally expensive, limiting real-world use.

Gaussian-Splatting based Surface Reconstruction 3DGS represents scenes with Gaussian primitives, but extracting geometry is challenging due to its discrete and unstructured nature (Kerbl et al. 2023). SuGaR (Guédon and Lepetit 2024) first extracted meshes by encouraging Gaussian fitting and using Poisson reconstruction, but achieved limited accuracy lacking geometric regularization. The idea of approximating 3D Gaussians as planes (shortest axis as normal) was extended in 2D Gaussian Splatting (2DGS) (Huang et al. 2024) and PGSR (Chen et al. 2024)—the former compressing 3DGS into 2D planes, the latter emphasizing cross-view geometric consistency, though compressing reduces rendering quality. GOF (Yu, Sattler, and Geiger 2024) defines a Gaussian opacity field using maximum Gaussian-ray intersection depth, but pixel-wise depth calculation adds heavy computation. RaDeGS (Zhang et al. 2024) improved on this by proving the level set normal is planar, reducing training time and improving geometry, yet introduced new approximations and bias. Inspired by these, we propose a geometry extraction method without approximations and additionally incorporate multi-view geometric consistency to optimize depth and normals, achieving high-precision reconstruction with comparable computational resources.

Preliminaries

Render process: 3DGS fits a scene using a set of Gaussian primitives $\mathcal{G} = \{g_i | i = 1, .., N\}$. Each Gaussian primitive $\mathcal{G}_i$ is parameterized by a full 3D covariance matrix $\Sigma_i \in \mathbb{R}^{3\times 3}$, which is centered at the world space point $\mu_i \in \mathbb{R}^3$:

$$g_i(x) = e^{-\frac{1}{2}(x-\mu_i)^T(\Sigma_i)^{-1}(x-\mu_i)} \tag{1}$$

Next, spherical harmonics are assigned to each Gaussian to model view-dependent colors and employ α-blending to integrate colors in sequence:

$$c = \sum_{i=1}^{N} \alpha_i \prod_{j=1}^{i-1} (1 - \alpha_j) c_i \tag{2}$$

where $\alpha_i = g_i \cdot o_i$ is parameterized by the covariance matrix multiplied by a learnable opacity o_i for each Gaussian.

Depth extraction: In depth extraction, based on different 3DGS methods, depth extraction can be classified into two forms: Expected depth and Median depth.

- Expected depth D_p^{exp} is defined as:

$$D_p^{\text{exp}} = \sum_{g \in \mathcal{G}_p} T_{g-1} \alpha_g d_{p,g}. \tag{3}$$

- Median depth D_p^{med} is defined as:

$$D_p^{\text{med}} = \max_{g \in \{g \mid T_{p,g} < 0.5\}} d_{p,g}. \tag{4}$$

where $\mathcal{G}_p$ represents the set of Gaussian primitives associated with pixel p, $T_g = \prod_{j=1}^{g-1}(1 - \alpha_j)$ denotes the accumulated transmittance for Gaussian g, and $d_{p,g}$ is the depth of Gaussian primitive g at pixel p.

The Proposed Method

We next detail the core components of TSPE-GS, starting with probabilistic depth modeling.

Probabilistic Depth Modeling

First, based on the previous definitions, we can derive the following formula that represent a relationship between T_p and d_p:

$$d_p \to T_p : \quad T_p(d) = \min_{g \in \{g \mid d_{p,g} > d\}} T_{p,g} \tag{5}$$

Next, we can extend this mapping to a continuous form. For a specified pixel p, we assign an opacity α_d to each infinitesimal depth element within the depth range d. Under the constraints of

$$\int_{T_{g-1}}^{T_g} d_T \mathrm{d}T = \Delta T d_g \tag{6}$$

and

$$\alpha_g = 1 - e^{\int_{d \in g} \ln(1-\alpha_d)\, \mathrm{d}d}, \tag{7}$$

we can express the transmittance T_d corresponding to each depth d for pixel p as: $T_d = \prod_0^d (1-\alpha_d)$. It is straightforward to verify that, under this definition, the following properties hold: $T_{p,i} \geq T_{p,j}$ only if $d_i \leq d_j$, the transmittance T approaches 0 as depth tends to infinity, and approaches 1 as depth approaches zero. These conditions naturally suggest modeling the relationship between T and d using a cumulative distribution function (CDF), which can be expressed as:

$$P(d_{p,g} < d) = 1 - T_d \tag{8}$$

when defining this probabilistic framework, we also obtain a definition for the probability density function (PDF):

$$\begin{aligned} p(d) &= -\frac{\mathrm{d}T_d}{\mathrm{d}d} = -\frac{\mathrm{d}e^{\int_0^d \ln(1-\alpha_d)\mathrm{d}d}}{\mathrm{d}d} \\ &= -\ln(1 - \alpha_d)T \approx \alpha_d T_d \end{aligned} \tag{9}$$

Noting that each infinitesimal α_d is sufficiently small, so $\lim_{x\to 0} \ln(1 - x) \to -x$. This approximation shows that $p(d)$ matches precisely the weighting used in the computation of expected depth, ensuring consistency between the probabilistic framework and depth expectation. According to experiments in Appendix A[1], $p(d)$ exhibits a unimodal distribution very close to a Gaussian when no transparent surfaces are present, while the CDF displays a classic S-shaped curve, as illustrated in Figure 2.

The median depth clearly corresponds to the depth at which the CDF equals 0.5. However, for expected depth, some transformations are required: Referring to the above definitions, the relationship can be expressed as follows:

$$T_g = T_{g-1} \cdot (1 - \alpha_g) \tag{10}$$

Consequently, $T_{g-1}\alpha_g$ can be reformulated as:

$$T_{g-1}\alpha_g = T_{g-1} - T_g \tag{11}$$

This leads to a reformulation of expected depth. By substituting the constraint given in Eq. 6, we obtain:

$$D_p^{\text{exp}} = \sum_{g \in \mathcal{G}_p} \Delta T_g d_{p,g} = \sum_{g \in \mathcal{G}_p} \int_{T_{g-1}}^{T_g} d_T \mathrm{d}T = \int_0^1 d_T \mathrm{d}T \tag{12}$$

---

[1] Appendix available on arXiv

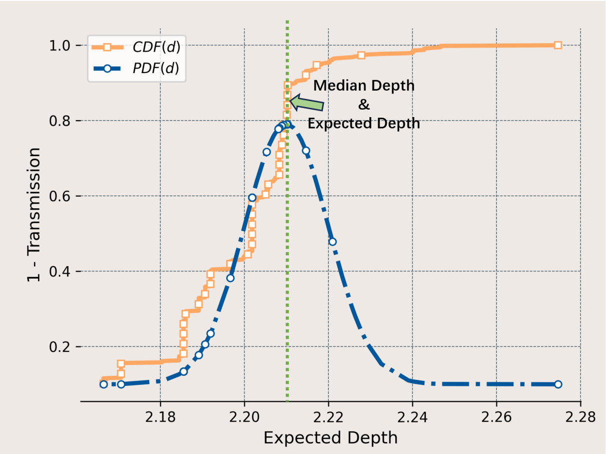

Figure 2: Depth PDF and CDF for opaque surfaces.

Subsequently, we perform a change of variables to obtain the following form:

$$D_p^{\text{exp}} = \int_0^{d_{\max}} d\frac{\mathrm{d}T_d}{\mathrm{d}d}\mathrm{d}d = \int_0^{d_{\max}} dp(d)\mathrm{d}d = E(d) \quad (13)$$

Thus, the expected depth is the expected value of the previously defined depth distribution. Since it is a unimodal distribution similar to a Gaussian distribution, the point where the CDF equals 0.5 and the expected value both clearly lie at the peak of the unimodal distribution, which corresponds to the mean of the Gaussian distribution, as shown in Figure 2. Currently, many 3D surface reconstruction methods are based on these two types of depth, considering that both depths closely align with the target surface height. Therefore, it can be inferred that in non-transparent scenarios, the position of peak points represents the location of surfaces.

This corresponds to the situation in the absence of semi-transparent surfaces. Surprisingly, even in the presence of multiple visible surfaces, as shown in Figure 3, the depth distribution exhibits multiple peaks resembling a mixture of Gaussian distributions. These peaks reliably correspond to visible surfaces, as consistently demonstrated by empirical experiments. Further details of these experiments can be found in Appendix B.

Based on these observations, we propose a unified hypothesis that accounts for surface depth in both semi-transparent and non-transparent scenarios:

Unified surface depth hypothesis The depth probability distribution for a given pixel can be modeled as a (uni- or multi-)modal distribution whose peaks correspond directly to the physical surface depths along the viewing ray. Formally, whether the distribution is unimodal (absence of semi-transparent surfaces) or multimodal (presence of semi-transparent surfaces), the set of surface depth locations $\{\mu_k\}$ are given by the set of local maxima of the distribution:

$$\{\mu_k\} = \{d \mid p'(d) = 0, \quad p''(d) < 0\} \quad (14)$$

where each μ_k corresponds to one visible surface depth. This hypothesis generalizes the classical unimodal assumption to handle complex scenes involving semi-transparent layers.

However, in the presence of semi-transparent surfaces, both the expected depth and the median depth fail to accurately represent the depths of the multiple visible layers. This is mainly because when a pixel corresponds to multiple surfaces, these single-value statistics can only reflect one layer at most. Consider a bimodal depth distribution modeled as a mixture of Gaussians:$p(d) = w_1 \cdot \mathcal{N}(d; \mu_1, \sigma_1^2) + w_2 \cdot \mathcal{N}(d; \mu_2, \sigma_2^2)$, where $\mathcal{N}(d; \mu, \sigma^2)$ is a Gaussian with mean μ and variance σ^2, and $w_1 + w_2 = 1$ are the mixing weights.

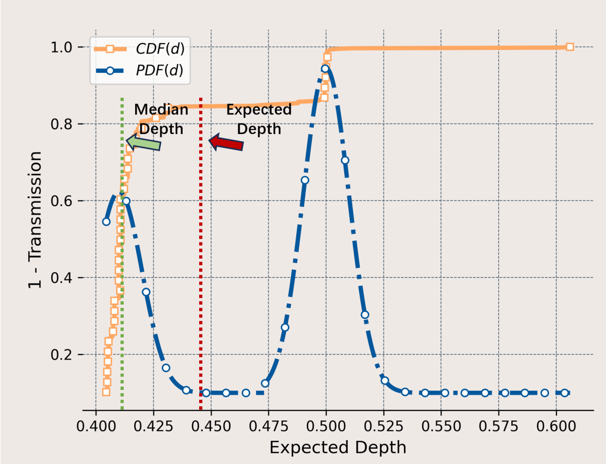

Figure 3: Depth PDF and CDF for semi-transparent surfaces.

Assuming that the two components are well separated with large variances, the peaks naturally correspond to μ_1 and μ_2, representing the visible surface depths. The expected depth, calculated as $D_{exp} = w_1\mu_1 + w_2\mu_2$, lies between these peaks, losing direct physical interpretability. Similarly, the median depth also falls between the two peaks, complicating the interpretation of depth for semi-transparent surfaces (see Figure 3).

To overcome these challenges, we propose TSPE-GS, a novel method based on the key observation that the peaks of the depth probability distribution correspond directly to physical surface locations. Our approach for reconstructing scenes with semi-transparent surfaces involves three key steps: 1) obtain the probability distribution; 2) calculate the locations of the peaks; and 3) fuse peak depths.

Obtaining the Probability Distribution To compute the probability distribution $P(d)$, our objective is to establish the relationship between depth d and quantity $1 - T$, while ensuring that the two constraints are satisfied. We can define $\alpha_d = 1 - e^{\frac{\ln(1-\alpha_g)}{d_g - d_{g-1}}(d - d_{g-1})}$, which holds for any d such that $d_{g-1} < d < d_g$. This definition of α_d simultaneously satisfies constraints in Eq. 6 and Eq. 7. Because we need to sample many values of T with corresponding depths d and then use density estimation to calculate the PDF, high-precision depth values are not required. Therefore, Instead of deriving d from α_d, we set $d = d_g$ whenever $T_g < T_d < T_{g-1}$. Experiments show that the PDF obtained by this simplification closely matches the strict PDF. More detailed experiments

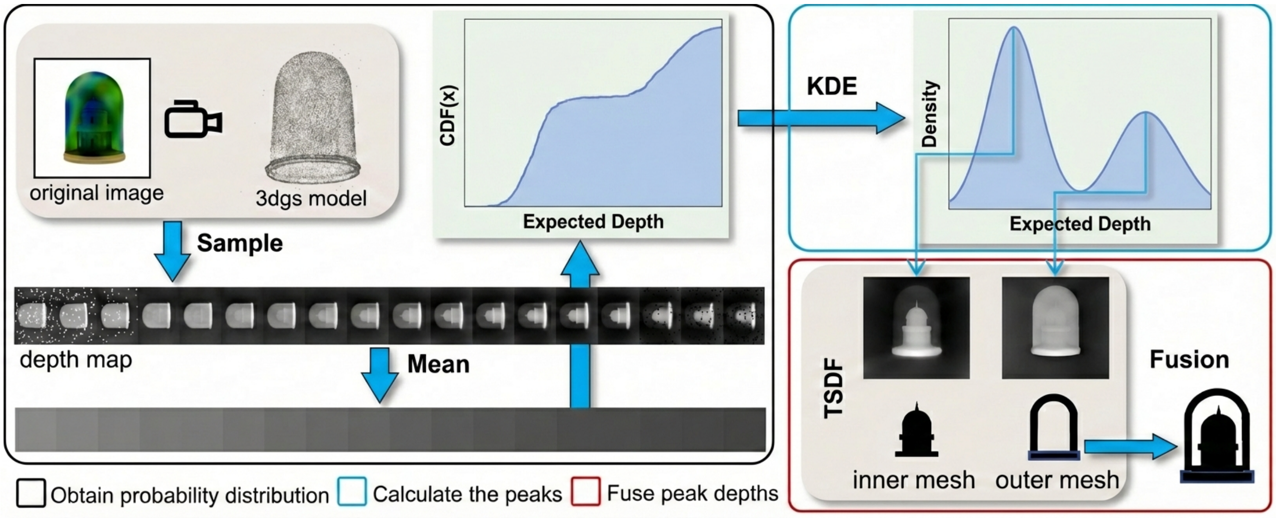

Figure 4: Workflow of the TSPE-GS pipeline for multi-layer depth estimation and reconstruction.

and derivations can be found in Appendix C.

Therefore, to obtain the required depths for different values of T', we only need to compute $D'_p = \max_{g \in \{g | T_{p,g} < T'\}} d_{p,g}$, which represents the maximum depth $d_{p,g}$ among all indices where the transmittance $T_{p,g}$ is less than the threshold T'. This method is easy to incorporate into existing 3DGS frameworks. Below is an example pseudocode illustrating how to compute D'_p for a given T':

In this work, the sampling of T' values is performed using uniform sampling, defined as $T' \in \left\{ \frac{k}{N} \mid k = 1, 2, \ldots, N \right\}$. Thanks to the extremely fast rendering speed of 3DGS, the sampling process for all viewpoints can be completed within approximately $1 \sim 2$ minutes. Finally, The depth distribution $p(d)$ is obtained by applying kernel density estimation to the sampled depth values.

Algorithm 1: 3D Gaussian Splatting Depth Extraction

Input: For each pixel p, lists of transmittance values $T_{p,g}$, depths $d_{p,g}$, colors c_g, opacities $\alpha_{p,g}$ for each Gaussian g
Parameter: Opacity threshold τ, transmittance threshold T'
Output: Color and depth C_p, D'_p

```
for each pixel $p$ do
    Initialize $C_p \leftarrow 0$, $\alpha_p \leftarrow 0$, $D'_p \leftarrow 0$
    for each Gaussian $g$ intersecting pixel $p$ do
        Compute transmittance $T_{p,g} = \prod_{k=1}^{g-1}(1 - \alpha_{p,k})$
        Compute opacity contribution $\alpha_{p,g}$
        Update opacity: $\alpha_p \leftarrow \alpha_p + (1 - \alpha_p)\alpha_{p,g}$
        Update color: $C_p \leftarrow C_p + (1 - \alpha_p)\alpha_{p,g}c_g$
        if $T_{p,g} < T'$ then
            Assign depth: $D'_p \leftarrow d_{p,g}$
            break
        end if
    end for
    Output $C_p$, $D'_p$
end for
```

Calculating the Peak Locations

Having obtained the probability distribution $p(d)$ of depth d for each pixel across all viewpoints, surface depth can be estimated by detecting peaks in this distribution. However, this approach faces two main challenges: the presence of abnormal Gaussian primitives causing spurious multiple peaks, and the high computational cost of performing peak detection on every pixel at all viewpoints.

However, we found a useful property to address these challenges. By averaging depth values d across all pixels at each of the N transmittance thresholds T' for a viewpoint, we get a depth sequence of length N that exhibits similar multi-peak characteristics as the per-pixel distributions, conforming to the unified surface depth hypothesis (see Figure 4). To mathematically explain this, we categorize all visible surfaces into $\mathcal{S}$ types, where each type corresponds to surfaces with distinct materials or occlusion relationships. For a surface type $s \in \mathcal{S}$, the depth distributions at different pixels belonging to that surface form a family of functions $p_p(d) = p(d + t)$, where t denotes the spatial shift due to different positions of the surface in the scene.

For any two different spatial shifts t_0 and t_1, corresponding to depth distributions $p_{p,0}$ and $p_{p,1}$, let their extremum points be d_0 and d_1. The $\text{CDF}_{p,0}(d_0)$ and $\text{CDF}_{p,1}(d_1)$ both equal the same transmittance threshold T'.

Now, define an interpolated inverse cumulative distribution function (ICDF) as: $\text{ICDF}(T) = a \cdot \text{ICDF}_0(T) + (1 - a) \cdot \text{ICDF}_1(T)$, for $a \in [0, 1]$. The probability distribution corresponding to this ICDF satisfies $\text{CDF}(d') = T'$, meaning that d' is also an extremum point of this combined distribution. The proof is as follows: from the definition of the inverse CDF, ICDF(T) is the quantile such that $F(\text{ICDF}(T)) = T$. Differentiating this w.r.t. T gives:

$$p(d) = \frac{1}{\text{ICDF}} = \frac{1}{a \cdot \text{ICDF}'_0 + (1 - a) \cdot \text{ICDF}'_1} \tag{15}$$

To identify extrema of $p(d)$, we examine where the derivative $\frac{\mathrm{d}}{\mathrm{d}d}p(d)$ vanishes. Using the chain rule and noting $d =$

$\mathrm{ICDF}(T)$, this corresponds to the condition:

$$p(d)\frac{\mathrm{d}}{\mathrm{d}T}\left(a\cdot\mathrm{ICDF}_0'(T)+(1-a)\cdot\mathrm{ICDF}_1'(T)\right)=0. \quad (16)$$

The second derivative of the ICDF is related to the derivative of the PDF by:

$$\mathrm{ICDF}''(T)=-\frac{p'(\mathrm{ICDF}(T))\cdot\mathrm{ICDF}'(T)}{p(\mathrm{ICDF}(T))^2}. \quad (17)$$

Evaluating at $T=T'$, and noting that $p'_{p,0}(d_0)=0$ and $p'_{p,1}(d_1)=0$, it follows that

$$\frac{d}{dT}\left(a\cdot\mathrm{ICDF}_0''(T)+(1-a)\cdot\mathrm{ICDF}_1''(T)\right)\bigg|_{T=T'}=0. \quad (18)$$

Therefore, the derivative of the denominator of $p(d)$ is zero at $d'=\mathrm{ICDF}(T')$, hence $p(d)$ has a stationary point at d'. This implies that for the same surface type, there exists a transmittance threshold T' such that the aggregated depth value $d'=\mathrm{ICDF}(T')$, which corresponds to the mean over pixels, is an extremum point of the combined distribution. Moreover, the depth at each pixel, $d_p=\mathrm{ICDF}_p(T')$, corresponds to the extremum point of its individual distribution. As a result, using the mean depth values computed from each depth map at a given transmittance T is sufficient to identify all surface depth. Following the same approach, we construct a PDF for the combined depth distribution.

After obtaining the new PDF, we use a local maxima detection algorithm to estimate a peak score for each sampled point and define peaks with scores above a certain threshold as target peaks. Selecting an appropriate threshold is crucial: too high a value may miss true visible surface depths, while too low a value can introduce floating artifacts by including spurious peaks. Thus, the threshold should balance capturing genuine surface depths and minimizing false positives for robust extraction. We further analyze this influence in the experiments section.

Fuse Peak Depths

After extracting peak depths from probabilistic distributions, we fuse the resulting multi-layer depth maps into a volumetric grid using TSDF. Each peak depth map is integrated into the TSDF volume according to camera poses, with TSDF values and weights updated via weighted averaging. However, directly fusing all peak depth maps can cause problems such as surface roughness and holes due to interference between layers. This occurs because inner layer depths—corresponding to occluded surfaces—spatially overlap with outer layer depths covering the same regions. Since inner layer depths are generally less accurate, their interference degrades the quality of outer surface reconstruction. Further analysis are provided in Appendix D. To mitigate this, we propose a progressive reconstruction strategy: first reconstruct the outermost surfaces and freeze their associated voxels in the TSDF volume; then reconstruct inner surfaces while preventing interference from the frozen outer layers. The pseudocode is detailed in Appendix E. In summary, our method combines probabilistic depth modeling, multi-peak detection, and progressive TSDF fusion to reconstruct multi-layer surfaces in scenes with transparency and occlusion. Figure 4 illustrates the overall workflow.

Method	Cas	Tab	Cof	Kit	Bot	Mon	Dou	Vas	**Avg.** ↓
Mip360	4.01	5.22	2.10	2.38	2.32	4.39	2.90	1.19	3.13
NeuS	5.09	1.19	1.07	0.39	2.27	5.92	0.87	1.95	2.34
HFS	5.09	0.85	2.84	N/A	1.49	3.22	N/A	8.68	N/A
angelo	2.07	0.48	0.80	0.58	0.40	2.46	0.51	1.70	1.12
NeRRF	1.27	0.57	0.82	3.09	3.72	2.90	0.83	3.49	2.11
PGSR	1.44	0.60	3.60	2.12	4.83	12.20	2.24	0.86	3.48
GOF	0.64	0.85	1.08	0.96	0.84	2.80	1.19	0.71	1.13
GSF	1.52	1.74	0.61	0.82	0.62	3.00	0.87	1.63	1.39
2DGS	1.21	0.88	1.14	1.00	2.05	7.70	1.38	1.13	2.06
+TSPE	0.72	1.45	0.99	0.70	1.45	1.47	1.12	0.72	**1.08**
RaDe	0.49	0.57	0.98	0.44	0.80	3.00	0.63	0.85	0.97
+TSPE	0.48	0.51	1.04	0.51	0.76	1.66	0.56	0.76	**0.82**

Table 1: Quantitative comparison on the αSurf dataset (Wu et al. 2025) for translucent object reconstruction evaluated by Chamfer distance ($\times 10^{-2}$). Note that HFS (Wang, Skorokhodov, and Wonka 2022) fails to learn any surface on "kitchen table" and "double table" scenes.

Method	BLO	BOS	CAP	COM	FTR	GUM	HFD	SBL	**Avg.**↓
2DGS	2.35	6.23	1.03	3.68	2.28	1.53	3.42	1.66	2.65
PGSR	2.41	0.43	0.71	3.01	1.28	1.37	2.76	0.70	1.58
GOF	2.40	1.20	0.50	1.99	1.22	1.56	3.31	1.10	1.66
TSGS	1.99	0.64	0.50	2.80	2.14	1.61	3.57	1.41	1.83
RaDe	2.43	0.49	0.48	1.98	1.23	1.07	4.79	1.19	1.71
+TSPE	0.96	0.59	0.69	1.51	0.74	0.58	2.51	0.48	**0.88**

Table 2: Quantitative results for various methods evaluated on the bottleship dataset. The Avg. column shows the mean number of valid entries per row.

Experiments

Datasets. To validate the effectiveness of our method in reconstructing semi-transparent surfaces, we conducted experiments on two semi-transparent datasets: the publicly available αsurf (Wu et al. 2025) dataset and our self-captured semi-transparent dataset — which contains objects occluded by semi-transparent surfaces and is casually named the "bottleship" dataset. In addition, we also evaluated the reconstruction performance of our method on opaque surface datasets, primarily focusing on two classic geometric reconstruction benchmarks: DTU (Aanæs et al. 2016) and BlendedMVS (Yao et al. 2020).

Baselines. To position our method within the field of semi-transparent surface reconstruction, we benchmarked our proposed approach against state-of-the-art (SOTA) methods, including 3D Gaussian-based reconstruction pipelines and NeRF-based geometry reconstruction pipelines.

Metrics. Our method aims to evaluate geometry reconstruction quality, with Chamfer Distance (CD) serving as the primary evaluation metric.

Implementation. During the training phase, we adopted the latest open-source implementations of the baselines and strictly adhered to their hyperparameter settings, including training epochs and optimizers. While in the geometry extraction stage, to ensure consistency, we uniformly employed the TSDF method for geometric reconstruction across all methods.

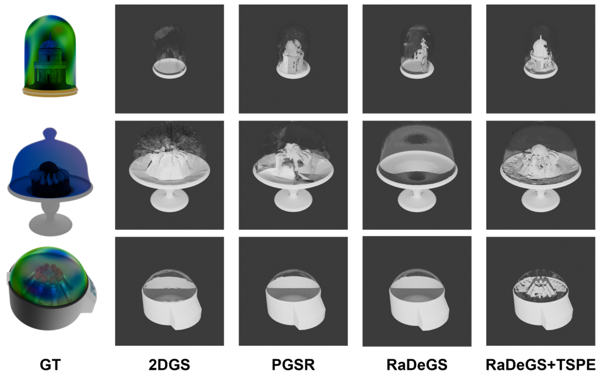

Figure 5: We visually compare our method with other Gaussian-based geometry reconstruction pipelines to demonstrate the effectiveness.

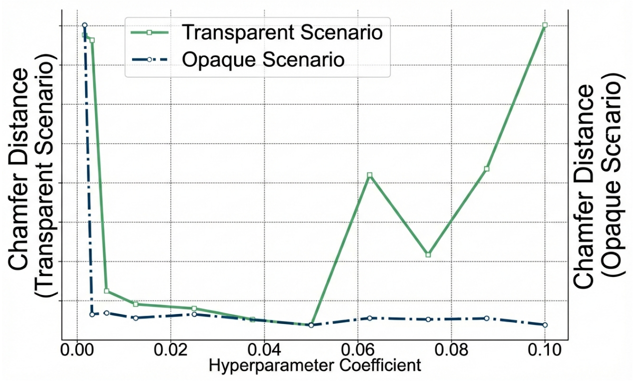

Figure 6: Impact of peak threshold on chamfer distance in semi-transparent and opaque scenarios.

Result Comparison

Table 1 and Table 2 present the reconstruction quality of various methods on semi-transparent datasets using NeRF-based and 3D Gaussian-based pipelines. Our approach demonstrates distinct advantages within the 3D Gaussian pipeline framework, achieving state-of-the-art results with RaDeGS as the backbone across both datasets. Additionally, our proposed method demonstrates competitive performance within the NeRF-based geometry reconstruction pipeline. Furthermore, to validate the generalization capability of our approach, Tables 3 present experimental results on opaque datasets using several backbone architectures. These results empirically demonstrate the effectiveness of our method in reconstructing opaque surfaces. However, we observe that not all 3D Gaussian-based methods are equally compatible with TSPE-GS; a fundamental requirement is that the underlying method must effectively represent the scene. For example, 2DGS performs poorly on our bottleship dataset, limiting TSPE-GS's applicability. See the extended version and Appendix F for detailed analyses.

Visual Comparison

In Figure 5, we compare our method with classical Gaussian-based geometric reconstruction methods. It can be observed that these SOTA geometric reconstruction methods tend to represent either a specific plane—either the outer or inner surface—failing to reconstruct both surfaces simultaneously. However, with the integration of our extraction module, we effectively extract both outer and inner surfaces.

Hyperparameter Analysis

We investigate the sensitivity of the peak threshold coefficient, a key hyperparameter in our probabilistic depth estimation process. As shown in Figure 6, both semi-transparent and opaque scenarios share a common range of values that achieve near-optimal Chamfer Distance results. This overlap indicates that our method is robust to the choice of this hyperparameter across diverse scene types.

Scene	**RaDe**	**+TSPE**	**2DGS**	**+TSPE**
Bea	0.50	0.56	1.08	0.66
Clo	0.87	0.87	0.35	0.26
Dog	1.07	1.06	0.64	0.33
Dur	0.49	0.49	5.72	2.91
Jad	0.14	0.14	0.29	0.14
Man	2.89	2.91	0.75	1.12
Scu	0.20	0.20	0.73	0.80
Sto	0.28	0.28	2.32	0.93
Avg↓	0.81	**0.81**	1.48	**0.89**

Table 3: BMVS dataset results: comparison of RaDeGS and 2DGS methods without and with TSPE.

Conclusion

We propose TSPE-GS, a probabilistic depth extraction method that reconstructs semi-transparent surfaces and occluded surfaces in Gaussian Splatting pipelines. By modeling depth distributions per pixel and progressively fusing TSDFs, our method provides a plug-and-play solution for modeling scenes with semi-transparent surfaces. Experiments show TSPE-GS outperforms previous Gaussian-based and NeRF-related methods on both semi-transparent and opaque scenes. **Limitation.** Our method depends on the underlying 3DGS's ability to fit scene geometry accurately; if the base 3DGS method fails to sufficiently model the geometry, the performance our approach will deteriorate.

Acknowledgements

This work was supported by the National Science Foundation of China (62576092) and the Big Data Computing Center of Southeast University. We would like to thank anonymous reviewers for their constructive suggestions.

References

Aanæs, H.; Jensen, R. R.; Vogiatzis, G.; Tola, E.; and Dahl, A. B. 2016. Large-scale data for multiple-view stereopsis. *International Journal of Computer Vision*, 120: 153–168.

Barnes, C.; Shechtman, E.; Finkelstein, A.; and Goldman, D. B. 2009. PatchMatch: A randomized correspondence algorithm for structural image editing. *ACM Trans. Graph.*, 28(3): 24.

Bleyer, M.; Rhemann, C.; and Rother, C. 2011. Patchmatch stereo-stereo matching with slanted support windows. In *Bmvc*, volume 11, 1–11.

Boissonnat, J.-D.; and Teillaud, M. 2006. *Effective Computational Geometry for Curves and Surfaces (Mathematics and Visualization)*. Mathematics and Visualization. Springer. ISBN 978-3-540-33258-9.

Bruno, F.; Bruno, S.; De Sensi, G.; Luchi, M.-L.; Mancuso, S.; and Muzzupappa, M. 2010. From 3D reconstruction to virtual reality: A complete methodology for digital archaeological exhibition. *Journal of Cultural Heritage*, 11(1): 42–49.

Chen, D.; Li, H.; Ye, W.; Wang, Y.; Xie, W.; Zhai, S.; Wang, N.; Liu, H.; Bao, H.; and Zhang, G. 2024. Pgsr: Planar-based gaussian splatting for efficient and high-fidelity surface reconstruction. *IEEE Transactions on Visualization and Computer Graphics*.

Curless, B.; and Levoy, M. 1996. A volumetric method for building complex models from range images. In *Proceedings of the 23rd annual conference on Computer graphics and interactive techniques*, 303–312.

Ding, Y.; Yuan, W.; Zhu, Q.; Zhang, H.; Liu, X.; Wang, Y.; and Liu, X. 2022. Transmvsnet: Global context-aware multi-view stereo network with transformers. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 8585–8594.

Fu, Q.; Xu, Q.; Ong, Y. S.; and Tao, W. 2022. Geo-neus: Geometry-consistent neural implicit surfaces learning for multi-view reconstruction. *Advances in Neural Information Processing Systems*, 35: 3403–3416.

Furukawa, Y.; and Ponce, J. 2009. Accurate, dense, and robust multiview stereopsis. *IEEE transactions on pattern analysis and machine intelligence*, 32(8): 1362–1376.

Goesele, M.; Snavely, N.; Curless, B.; Hoppe, H.; and Seitz, S. M. 2007. Multi-view stereo for community photo collections. In *2007 IEEE 11th international conference on computer vision*, 1–8. IEEE.

Gu, X.; Fan, Z.; Zhu, S.; Dai, Z.; Tan, F.; and Tan, P. 2020. Cascade cost volume for high-resolution multi-view stereo and stereo matching. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2495–2504.

Guédon, A.; and Lepetit, V. 2024. 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*, 5354–5363.

Hart, J. C.; Sandin, D. J.; and Kauffman, L. H. 1989. Ray tracing deterministic 3-D fractals. In *Proceedings of the 16th annual conference on Computer graphics and interactive techniques*, 289–296.

Hirschmuller, H. 2007. Stereo processing by semiglobal matching and mutual information. *IEEE Transactions on pattern analysis and machine intelligence*, 30(2): 328–341.

Huang, B.; Yu, Z.; Chen, A.; Geiger, A.; and Gao, S. 2024. 2d gaussian splatting for geometrically accurate radiance fields. In *ACM SIGGRAPH 2024 conference papers*, 1–11.

Kazhdan, M.; and Hoppe, H. 2013. Screened poisson surface reconstruction. *ACM Transactions on Graphics (ToG)*, 32(3): 1–13.

Kendall, A.; Martirosyan, H.; Dasgupta, S.; Henry, P.; Kennedy, R.; Bachrach, A.; and Bry, A. 2017. End-to-end learning of geometry and context for deep stereo regression. In *Proceedings of the IEEE international conference on computer vision*, 66–75.

Kerbl, B.; Kopanas, G.; Leimkühler, T.; and Drettakis, G. 2023. 3d gaussian splatting for real-time radiance field rendering. *ACM Trans. Graph.*, 42(4): 139–1.

Li, M.; Pang, P.; Fan, H.; Huang, H.; and Yang, Y. 2025. TSGS: Improving Gaussian Splatting for Transparent Surface Reconstruction via Normal and De-lighting Priors. *arXiv preprint arXiv:2504.12799*. Accepted by MM.

Li, Z.; Müller, T.; Evans, A.; Taylor, R. H.; Unberath, M.; Liu, M.-Y.; and Lin, C.-H. 2023. Neuralangelo: High-fidelity neural surface reconstruction. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 8456–8465.

Liu, Y.-T.; Wang, L.; Yang, J.; Chen, W.; Meng, X.; Yang, B.; and Gao, L. 2023. Neudf: Leaning neural unsigned distance fields with volume rendering. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 237–247.

Long, X.; Lin, C.; Liu, L.; Liu, Y.; Wang, P.; Theobalt, C.; Komura, T.; and Wang, W. 2023. Neuraludf: Learning unsigned distance fields for multi-view reconstruction of surfaces with arbitrary topologies. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 20834–20843.

Luo, W.; Schwing, A. G.; and Urtasun, R. 2016. Efficient deep learning for stereo matching. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, 5695–5703.

Mildenhall, B.; Srinivasan, P. P.; Tancik, M.; Barron, J. T.; Ramamoorthi, R.; and Ng, R. 2021. Nerf: Representing scenes as neural radiance fields for view synthesis. *Communications of the ACM*, 65(1): 99–106.

Sajjan, S.; Moore, M.; Pan, M.; Nagaraja, G.; Lee, J.; Zeng, A.; and Song, S. 2020. Clear grasp: 3d shape estimation of transparent objects for manipulation. In *2020 IEEE international conference on robotics and automation (ICRA)*, 3634–3642. IEEE.

Schonberger, J. L.; and Frahm, J.-M. 2016. Structure-from-motion revisited. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, 4104–4113.

Wang, J.; Wang, P.; Long, X.; Theobalt, C.; Komura, T.; Liu, L.; and Wang, W. 2022. Neuris: Neural reconstruction of indoor scenes using normal priors. In *European conference on computer vision*, 139–155. Springer.

Wang, P.; Liu, L.; Liu, Y.; Theobalt, C.; Komura, T.; and Wang, W. 2021. Neus: Learning neural implicit surfaces by volume rendering for multi-view reconstruction. *arXiv preprint arXiv:2106.10689*.

Wang, Y.; Skorokhodov, I.; and Wonka, P. 2022. Hf-neus: Improved surface reconstruction using high-frequency details. *Advances in Neural Information Processing Systems*, 35: 1966–1978.

Wu, T. W.; Zhong, F.; Riegler, G.; Vainer, S.; Deng, J.; Oztireli, C.; et al. 2025. αsurf: Implicit surface reconstruction for semi-transparent and thin objects with decoupled geometry and opacity. In *International Conference on 3D Vision 2025*.

Yao, Y.; Luo, Z.; Li, S.; Fang, T.; and Quan, L. 2018. Mvsnet: Depth inference for unstructured multi-view stereo. In *Proceedings of the European conference on computer vision (ECCV)*, 767–783.

Yao, Y.; Luo, Z.; Li, S.; Zhang, J.; Ren, Y.; Zhou, L.; Fang, T.; and Quan, L. 2020. Blendedmvs: A large-scale dataset for generalized multi-view stereo networks. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 1790–1799.

Yariv, L.; Gu, J.; Kasten, Y.; and Lipman, Y. 2021. Volume rendering of neural implicit surfaces. *Advances in Neural Information Processing Systems*, 34: 4805–4815.

Yu, A.; Li, R.; Tancik, M.; Li, H.; Ng, R.; and Kanazawa, A. 2021. Plenoctrees for real-time rendering of neural radiance fields. In *Proceedings of the IEEE/CVF international conference on computer vision*, 5752–5761.

Yu, Z.; Sattler, T.; and Geiger, A. 2024. Gaussian opacity fields: Efficient adaptive surface reconstruction in unbounded scenes. *ACM Transactions on Graphics (TOG)*, 43(6): 1–13.

Zhang, B.; Fang, C.; Shrestha, R.; Liang, Y.; Long, X.; and Tan, P. 2024. Rade-gs: Rasterizing depth in gaussian splatting. *arXiv preprint arXiv:2406.01467*.

Zhou, X.; Lin, Z.; Shan, X.; Wang, Y.; Sun, D.; and Yang, M.-H. 2024. Drivinggaussian: Composite gaussian splatting for surrounding dynamic autonomous driving scenes. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 21634–21643.