High-Fidelity Polarimetric Implicit 3D Reconstruction with View-Dependent Physical Representation

Yu Qiu, Sijia Wen*, Hainan Zhang, Zhiming Zheng

Beijing Advanced Innovation Center for Future Blockchain and Privacy Computing, School of Aritificial Intelligence, Beihang University, Beijing, China
{qiuyu_13, sijiawen, zhanghainan}@buaa.edu.cn, zzheng@pku.edu.cn

Abstract

Neural implicit methods have made remarkable progress in 3D reconstruction. However, previous methods often assume view-independent properties of target objects, which fails to accurately reconstruct objects with challenging characteristics, such as transparency and high reflectivity. To address this limitation, we propose a polarimetric implicit 3D reconstruction method that integrates geometric and polarization information, enabling the production of high-quality meshes in complex scenes. For high-fidelity surface reconstruction, we introduce a view-dependent physical representation that thoroughly analyzes the subtle physical properties of reflections. The reconstruction process is further enhanced by a simple yet effective view-dependent detection algorithm and optimized using the principles of ray tracing and polarization. Experimental results demonstrate the superior performance of the proposed method in both real and synthetic scenarios.

Introduction

High-fidelity 3D reconstruction remains a significant challenge across numerous fields, including virtual reality, autonomous driving, robotics, and medical visualization. With the development of implicit surface representation (Park et al. 2019) and neural radiance fileds (Mildenhall et al. 2021), neural implicit methods (Yariv et al. 2020; Wang et al. 2021; Yariv et al. 2021; Zhang et al. 2021) have made significant progress recently. Despite the remarkable shape recovery results, most methods still rely on photometric consistency, which presents significant challenges for high-fidelity reconstruction. The high complexity of light paths in these objects entangles the surface color with their geometry, environmental lighting, and viewing direction, making it difficult to achieve pleasing neural inverse rendering.

To overcome the view-dependent issues caused by surface characteristics, Ref-NeuS (Ge et al. 2023) separates the view-dependent and view-independent components of the surface reflection to reduce the impact of reflective areas. NeRO (Liu et al. 2023) assumes opaque surfaces with BRDF to recover the shape of the reflective objects. PANDORA (Dave, Zhao, and Veeraraghavan 2022) and NeRSP (Han et al. 2024) decompose specular and diffuse reflections and generate meshes using polarization information. However, these methods only consider reflection and cannot address other complex surface characteristics of objects, such as transmission.

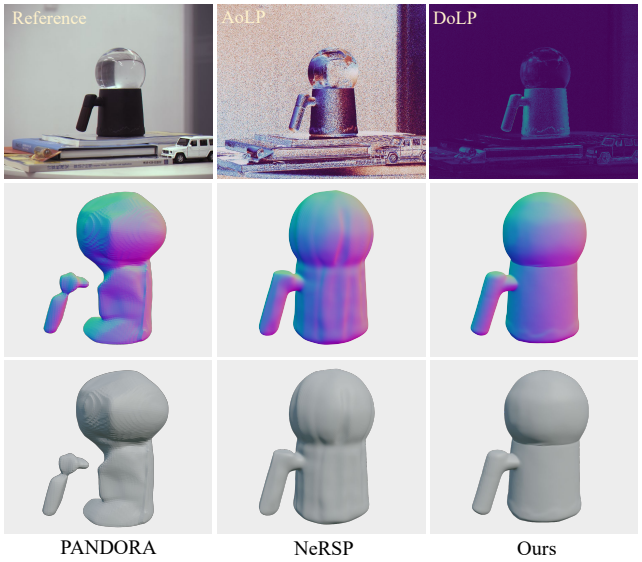
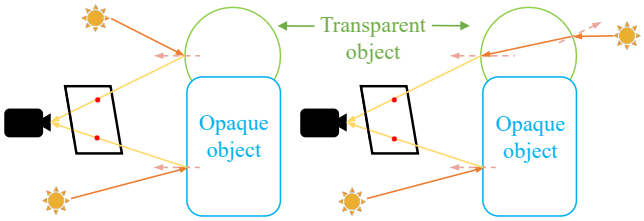


Figure 1: Compared to the state-of-the-art methods, PANDORA (first column) and NeRSP (second column), the proposed method delivers superior 3D reconstruction results, including high-quality meshes and normal maps. The reference image, along with AoLP and DoLP, are displayed in the top row.

For complex materials and textures, some methods employ silhouette constraints for shape recovery, but this often results in overly smooth surfaces. Li et al. (Li, Yeh, and Chandraker 2020) modeled complex light paths to reconstruct transparent object with sparse images, yet it still suffers from excessive smoothness. Other researchers (Tsai, Veeraraghavan, and Sankaranarayanan 2015; Lyu et al. 2020; Xu et al. 2022) utilize a static background with a coded pattern to establish a mapping between the camera's view rays and corresponding locations. This mapping constrains the light path, aiding in the detailed reconstruction of an object's shape. However, these methods often require

*Corresponding author.



(a) Neural implicit fields (b) Ours

Figure 2: Previous 3D reconstruction methods often assume there is only one propagation behavior of light in a scene, such as reflection (a). In contrast, our method (b) takes into account the propagation behavior of light more comprehensively, thereby achieving high-fidelity reconstruction.

strict calibration and precise control, which can be cumbersome and time-consuming in practical applications. Some researchers (Shao et al. 2024) employ polarization information to recover the shape of transparent objects. But the assumption of unpolarized incident light makes it difficult to apply it to natural light scenarios.

To achieve high-fidelity 3D reconstruction, we propose a polarimetric implicit 3D reconstruction framework that integrates geometric and polarization information, enabling the production of high-quality 3D surfaces of objects with complex characteristics. First, we use silhouette and color constraints to obtain an initial shape. Unlike the former polarimetric methods, which assume unpolarized incident light, we employ three MLPs to directly predict the Stokes vectors of polarized incident light. By applying Fresnel equations and Muller matrices (Clarke 2009), we decompose specular and diffuse reflections to compute the outgoing Stokes vectors. Due to the more realistic modeling analysis, the proposed method can perform more accurate reconstruction results in real-world scenarios. For considering the complex properties of object surfaces, we introduce a view-dependent physical representation that thoroughly analyzes the subtle physical properties of light propagation(Figure 2). In detail, we introduce a simple yet effective confidence map for regions with view-dependent characteristics. In order to estimate the shape of the objects accurately, given the strong relationship between polarization cues and surface normals, we apply a fusion algorithm to optimize the real AoLP φ captured by a polarization camera, using this fused AoLP map $\widetilde{\varphi}$ to constrain the physically rendered AoLP. Inspired by PIR (Shao et al. 2024), we employ ray tracing techniques to generate reflection confidence maps w_p, accurately constraining the rendered AoLP. Figure 1 demonstrates the superior performance of our method in reconstructing objects with complex characteristics compared to previous state-of-the-art methods.

To summarize, our main contributions are as follows:

- We propose a polarimetric implicit 3D reconstruction framework that integrates geometric and polarization information, enabling the production of high-quality meshes in complex scenes.
- For high-fidelity surface reconstruction, we introduce a view-dependent physical representation that thoroughly analyzes the subtle physical properties of reflections.
- Extensive experimental results demonstrate that the proposed method outperforms the most advanced techniques without any restrictions on ambient lighting conditions.

Related work

Neural Implicit Fields in 3D Reconstruction

Implicit representations based on neural networks have achieved promising results in novel view synthesis and 3D reconstruction. Surface rendering based on the differentiable ray casting is applied for surface reconstruction with different forms of implicit shape representations, such as the occupancy function (Peng et al. 2020) and signed distance function(SDF) (Park et al. 2019; Yariv et al. 2020). IDR (Yariv et al. 2020) extracts surface points with the zero-level set of SDF representations, and utilizes neural network gradients to solve a differentiable rendering formulation. VolSDF (Peng et al. 2020) and NeuS (Wang et al. 2021) learn the implicit surface function by introducing the volume rendering (Mildenhall et al. 2021), to improve the surface reconstruction quality from captured images. However, the assumption of view-independent for the reconstructed target object imposes limitations on the robustness and stability of the aforementioned methods.

Polarization in High-fidelity 3D Reconstruction

As a passive imaging principle with weak lighting conditions, polarization cues perform well in shape estimating tasks (Ba et al. 2020; Mingqi et al. 2022). Recently, many works have emerged that combine polarization information for high-fidelity mesh reconstruction of object appearance. PANDORA (Dave, Zhao, and Veeraraghavan 2022) is the first work that combines NeRF (Mildenhall et al. 2021) and polarization cues for inverse rendering. They train coordinate-based multi-layer perceptrons (MLPs) to estimate normals, diffuse radiance and specular radiance. After that, the estimated information are combined by a simplified model to generate the outgoing Stokes vectors. MVAS (Cao et al. 2023) makes full use of the physical characteristics of polarization, using the relationship between tangent vectors and normal vectors, and using azimuth maps to constrain the geometry of the reconstruction, achieving excellent results. NeRSP (Han et al. 2024) further utilizes polarization information based on MVAS to achieve sparse-view 3D reconstruction. However, all of the above methods have strict limitations on the scene of data acquisition, and it is necessary to ensure that the specular reflection component or diffuse reflection component dominates in the AoLP maps captured by the camera as much as possible, which is rare in real lighting scenes. NeISF (Li et al. 2024) considers the above issues and uses three MLPs networks to directly predict the incident Stokes vectors of the last bounce, achieving excellent results. However, the physical rendering component of NeISF is designed only for opaque materials, which results in limitations when applied to transparent scenarios.

Complex Characteristics in 3D Reconstruction

Reconstructing objects with complex characteristics has always been a fundamental computer vision problem. Many researchers have analyzed the interaction of light on such surfaces exhaustively.

For transmitted rays, according to the Fresnel equation, the propagation process of light on the surface of an object is affected by the refractive index of different media. This refraction phenomenon occurring in special media such as glass can affect the results of 3D reconstruction. Many traditional methods use some special hardware setups to reconstruct such objects, including light field probes (Wetzstein et al. 2011), scatter-trace photography (Morris and Kutulakos 2007), tomography (Trifonov, Bradley, and Heidrich 2006) or time-of-flight imaging (Tanaka et al. 2016). Meanwhile, some researchers (Kutulakos and Steger 2008) have established a mapping by setting a static background with a coded pattern. Differentiable tracing of refractive ray paths is then used to directly optimize a 3D mesh approximation of the object, ensuring silhouette consistency and smoothness simultaneously. PIR (Shao et al. 2024) applies polarization information to transparent objects and has achieved exciting results. However, PIR assumes that the incident light is unpolarized. When the incident light is polarized, the reconstruction quality will significantly decrease.

For reflected rays, Ref-NeuS (Ge et al. 2023) separates the view-dependent and view-independent components of the surface reflection to reduce the impact of reflective areas on reconstruction results. Kim et al. (Kim et al. 2017) model reflections by using BRDF, have also achieved impressive results. In recent years, some works (Dave, Zhao, and Veeraraghavan 2022; Li et al. 2024; Han et al. 2024) have modeled specular and diffuse reflections using polarimetric cues to recover accurate shapes.

In contrast to previous methods, the proposed method analyzes the physical properties of both transmitted and reflected light, and employs polarization information to guide the shape recovery, ultimately obtaining high-fidelity 3D reconstruction results.

Method

Leveraging the position of the camera and the refractive index of the reconstructed object, the proposed method aims to integrate polarization information to achieve high-fidelity mesh generation for objects with complex characteristics. We first introduce the polarimetric implicit 3D reconstruction strategy. Subsequently, we present the view-dependent physical representation, which is based on 3D visual and physical properties. The presentation concludes with a discussion on the optimization and loss function. Figure 3 illustrates the overview of the proposed method.

Polarimetric Implicit 3D Reconstruction

Implicit-Surface Representation. We use a signed distance field net f_{sdf} to represent a 3D scene. f_{sdf} is an implicit function that takes a 3D coordinate point $x_p \in \mathbb{R}^3$ as input and outputs the corresponding sdf value for that point:

$$f_{sdf}(x_p) = d_{sdf}, z_p, \tag{1}$$

where the d_{sdf} represents the distance d from the coordinate point to the surface of the object, and z_p represents a global geometry feature vector. The normal vector of the object surface is obtained by the following formula:

$$\hat{n}_p(x) = \nabla_x f_{sdf}(x) / \|\nabla_x f_{sdf}(x)\|_2, \tag{2}$$

The color c of the object surface is related to the viewing direction v_p and surface normal vector $\hat{n}_p$. We designed a network f_{color} and obtained more refined texture information. f_{color} is an MLP, which the forward model is therefore:

$$f_{color}(x_p, \hat{n}_p, v_p, z_p) = c. \tag{3}$$

Physics-Based Polarimetric Rendering. In order to accurately model the interaction process of incident light on the surface of an object, we use the Mueller matrix to calculate the outgoing Stokes vector. In Mueller calculus (Clarke 2009), a 4D Stokes vector s can be written as s=[s_0,s_1,s_2,s_3]. Where s_0 represents light intensity, s_1 and s_2 represent linear polarization components of the x-axis and $45°$ direction, s_3 represents circular polarization, respectively. In this work, we assume there is no circular polarization and thus neglect s_3.

In real scenes, the light shining on the surface of the objects usually includes ambient light, which is polarized. Neglecting such polarized lights will affect the accuracy of the reconstruction. Inspired by NeISF (Li et al. 2024), we use three MLP networks to predict the incident Stokes vector. Specifically, we use the $f_L \in \mathbb{R}^2$ network to predict specular reflection intensity $s_{spec}[0]$ and diffuse reflection intensity $s_{diff}[0]$. The f_{spec} network aims to predict the $s_{spec}[1]$ and $s_{spec}[2]$ components of the incident Stokes vector for specular reflection. The f_{diff} network to predict the $s_{diff}[1]$ and $s_{diff}[2]$ components of the incident Stokes vector for diffuse reflection. In contrast to NeISF, which predicts only $s_{diff}[1]$, we utilize the Mueller matrix for physical modeling of the object surface and apply f_{diff} to simultaneously predict both $s_{diff}[1]$ and $s_{diff}[2]$, thereby fitting the input of the physical model. The formulas for three MLPs are as follows:

$$f_L(\mathbf{x}, \omega_i) = s_{spec}[0], s_{diff}[0], \tag{4}$$

$$f_{spec}(\mathbf{x}, \omega_i) = s_{spec}[1, 2], \tag{5}$$

$$f_{diff}(\mathbf{x}, \omega_i) = s_{diff}[1, 2], \tag{6}$$

where [n] denotes the n^{th} element in Stokes vector. $\mathbf{x}$ represents the surface points of the object predicted by the f_{sdf}, and ω_i represents the incident direction of the ambient light interacting with the surface points.

According to the Fresnel equation and PIR (Shao et al. 2024), the amplitude reflection coefficients r_s and r_p can be written as follows:

$$\begin{aligned} r_s &= \frac{\eta_i \cos\theta_i - \eta_t \cos\theta_t}{\eta_i \cos\theta_i + \eta_t \cos\theta_t} = -\frac{\sin(\theta_i - \theta_t)}{\sin(\theta_i + \theta_t)}, \\ r_p &= \frac{\eta_i \cos\theta_t - \eta_t \cos\theta_i}{\eta_t \cos\theta_i + \eta_i \cos\theta_t} = -\frac{\tan(\theta_i - \theta_t)}{\tan(\theta_i + \theta_t)}. \end{aligned} \tag{7}$$

Where θ_i represents the angle between the incident ray and the normal, and θ_t corresponds to the angle of refracted ray. η_i and η_t are the refractive indices of the incident medium

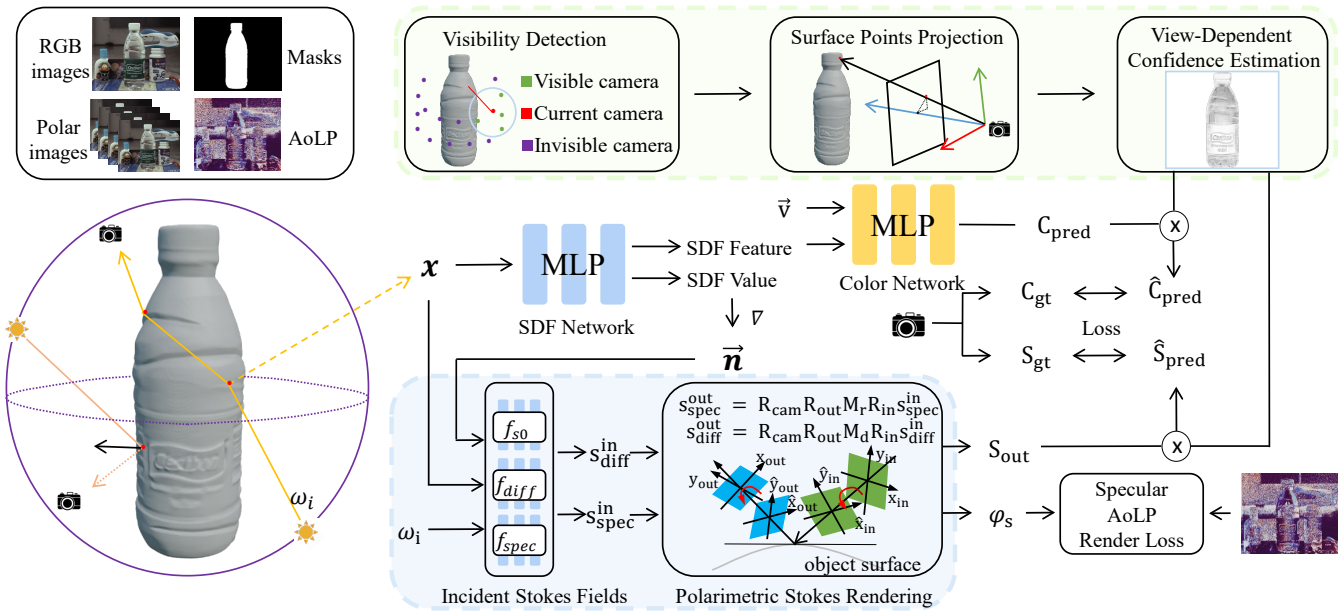


Figure 3: Overall architecture of the proposed method. The green area and the blue area represent the view-dependent physical representation and the polarimetric implicit 3D reconstruction, respectively.

and refract medium. The specular Stokes component s^i_{spec} and diffuse Stokes component s^i_{diff} of the incident light after interaction on the surface of the object can be mathematically modeled using Mueller matrix M_r and M_d:

$$\begin{aligned} s^o_{spec} &= R_c R_o M_r R_i s^i_{spec}, \\ s^o_{diff} &= R_c R_o M_d R_i s^i_{diff}, \end{aligned} \tag{8}$$

where the s^o_{spec} and s^o_{diff} denote outgoing specular Stokes vectors and outgoing diffuse Stokes vectors in camera coordinate system, respectively. R_c,R_o,R_i are coordinate system transformation matrices used to unify the reference frame to obtain the correct outgoing tokens. The formulas for M_r and M_d are as follows:

$$M_r = \frac{1}{2}\begin{bmatrix} (r_s^2+r_p^2) & (r_s^2-r_p^2) & 0 & 0 \\ (r_s^2-r_p^2) & (r_s^2+r_p^2) & 0 & 0 \\ 0 & 0 & 2r_s r_p & 0 \\ 0 & 0 & 0 & 2r_s r_p \end{bmatrix}, \tag{9}$$

$$M_d = \frac{f_T}{2}\begin{bmatrix} R_s+R_p & R_s-R_p & 0 & 0 \\ R_s-R_p & R_s+R_p & 0 & 0 \\ 0 & 0 & 2\sqrt{R_s R_p} & 0 \\ 0 & 0 & 0 & 2\sqrt{R_s R_p} \end{bmatrix}, \tag{10}$$

where $R_s = (1+r_s)^2$, $R_p = (1+r_p)^2\eta^2$, $f_T = -\frac{\eta_i \cos\theta_t}{\eta_t \cos\theta_i}$.

Furthermore, we utilize the relationship between the Stokes vector components of specular reflection and the surface normal vectors to accurately reconstruct areas of transparent objects. Specifically, for the specular Stokes vectors obtained through physical rendering in Eq.8, we can obtain the rendered AoLP φ_{spec}:

$$\varphi_{spec} = \frac{1}{2}\tan^{-1}\left(\frac{s^o_{spec}[2]}{s^o_{spec}[1]}\right). \tag{11}$$

To reduce the impact of inaccurate polarization information, inspired by PIR (Shao et al. 2024), we calculate light trace within transparent areas based on the Fresnel equations and the reflection percentage estimation w_p. w_p is defined as the ratio of the reflection intensity to total intensity, which can be finally obtained by the Fresnel term.

View-Dependent Physical Representation

For multi-view reconstruction, view consistency is essential for accurate surface reconstruction. However, due to special properties, some objects with complex medium exhibit view-dependent surface colors. To address this issue, we fully leverage 3D information to efficiently detect regions with view dependency by exploring color information. Furthermore, we present an AoLP fusion algorithm tailored for polarized incident light, which can more accurately constrain the surface normal vectors of the reconstructed mesh.

View-Dependent Confidence We jointly consider surface information and color information to estimate the view-dependent confidence map. Specifically, for a certain surface point $\mathbf{x_p}$ on the object's surface and the current camera $\mathbf{k}$, we calculate the distance d, which is from the surface point to the camera. Taking the current camera $\mathbf{k}$'s position in the world coordinate system as the center, we draw a circle with a radius r, and then obtain n camera coordinates and their corresponding images contained within the circle. As shown in Figure 4, by checking the sdf value, we can estimate the visibility of n cameras at that surface point and obtain n' valid cameras and their corresponding images.

By projecting surface points onto different coordinate systems, we can project the surface points onto n' different images and obtain n' color information. Due to the strong view-dependent surface points in transparent areas, we uti-

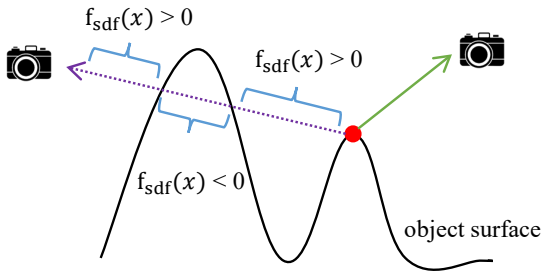


Figure 4: Visibility Detection. The solid green line corresponds to the visible camera, while the dashed purple line corresponds to the invisible camera. By comparing the values of sdf at different sampling points, we can easily determine the visibility of the camera to a certain surface point.

lize this property to calculate the variance of n' colors and use it as our view-dependent confidence map.

$$VDC = clip(1 - \frac{\sqrt{\sum_{i=1}^{n'}(c_i - c_k)^2}}{n'}, 0, 1). \quad (12)$$

In this way, we can effectively detect regions with view-dependent and add it as a weight to the loss, thereby reducing their impact on network training.

AoLP Fused. AoLP represents the angle of linearly polarized light, which can help us predict the normal vectors of an object's surface and guide multi-view 3D reconstruction. Unfortunately, due to the ambiguity between π and $\pi/2$, we do not know which component of specular reflection and diffuse reflection dominates on the object's surface. Especially in natural light environments, light interacts with the surface of objects, resulting in a mixture of specular and diffuse reflection components in the AoLP and DoLP captured by the camera. This makes it challenging to directly utilize AoLP and DoLP information to guide 3D reconstruction.

We utilize 3D information to obtain more accurate AoLP maps. Inspired by Ding et al. (Ding, Ji, and Ye 2022), we note that at a certain surface point of an object, the AoLP values obtained by different cameras are not the same. Due to surface interaction and the influence of incident light direction, the AoLP values obtained from certain views have a larger component dominated by diffuse reflection or specular reflection. By utilizing this property, we not only predict view-dependent regions but also compare the intensity of the two polarization images, using the corresponding AoLP with higher polarization intensity values.

Optimization

The network parameters are optimized using a collection of polarized images, each containing color, Stokes vector, AoLP, DoLP, mask, and camera pose data. Our loss function consists of four primary components. The first component is the silhouette loss function. Following IDR, we constrain the geometric shape of the reconstructed object by inputting masks. The initial geometric shape can provide a rough surface normal vector, which greatly helps our Stokes loss.

The second component is color loss. Simply using masks for reconstruction may result in excessively smooth results. Therefore, we also utilize the RGB information captured by the polarization camera to further assist in reconstructing more accurate geometric shapes. It can be represented as:

$$\mathcal{L}_{RGB} = \frac{1}{N}\|VDC * (C - \widetilde{C})\|_2^2, \quad (13)$$

Here, N represents the number of sampled light rays, $\widetilde{C}$ denotes the ground truth color, C represents the predicted color, VDC is the view-dependent confidence map.

The third component is the weighted Stokes loss. To further obtain better reconstruction results, we use the MLP networks and physical rendering module to calculate rendering Stokes vectors, and then use the real Strokes vectors($\widetilde{S}$) captured by a polarization camera to constrain the rendering Strokes vectors (S):

$$\mathcal{L}_{stokes} = \frac{1}{N}\|VDC * (S - \widetilde{S})\|_2^2. \quad (14)$$

The fourth component is the AoLP rendering loss. For transparent object areas, we use the polarimetric camera to capture the real AoLP maps, and then fuse them to obtain a more accurate fused AoLP maps. We use fused AoLP map φ_{cam} to constrain the rendered AoLP φ_{spec} calculated by the specular Stokes vectors, then calculate the $\mathcal{L}_{aolp}$:

$$\mathcal{L}_{aolp} = \frac{1}{N}\|(1 - VDC) * w_p * (\varphi_{spec} - \widetilde{\varphi_{cam}})\|_2^2. \quad (15)$$

In summary, the composite loss function for our model, denoted by $\mathcal{L}$, is articulated as:

$$\mathcal{L} = \mathcal{L}_{sihouette} + \lambda_c\mathcal{L}_{RGB} + \lambda_s\mathcal{L}_{stokes} + \lambda_{aolp}\mathcal{L}_{aolp}, \quad (16)$$

where λ_c, λ_s and λ_{aolp} are the coefficients for the corresponding loss terms.

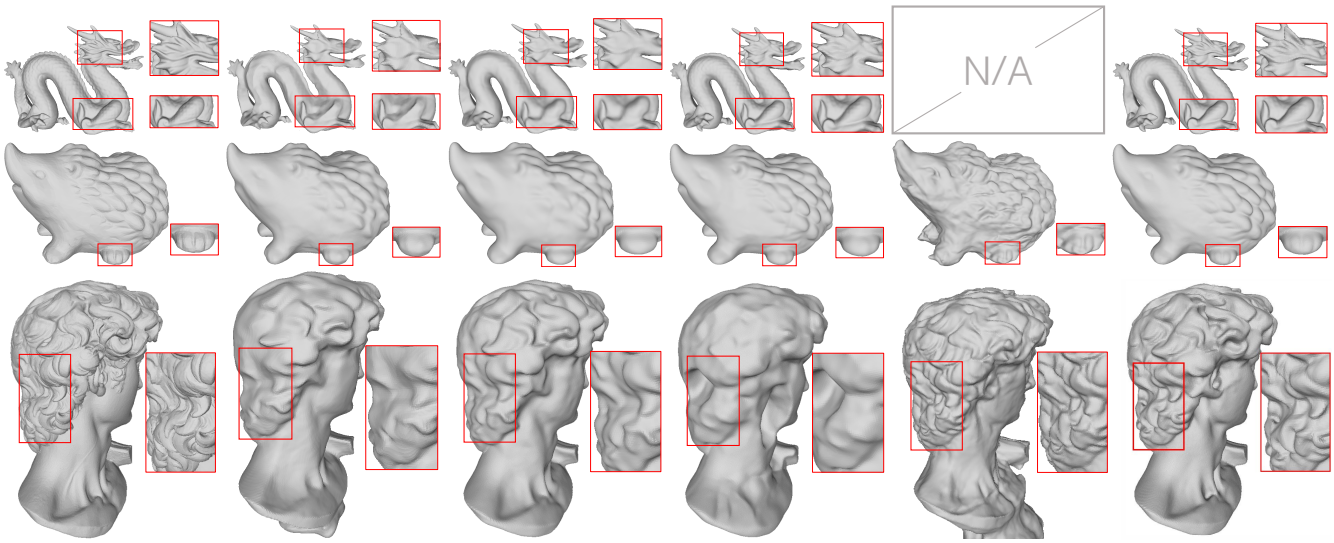
Experiments

Dataset

The experimental datasets include both a collection of objects with complex characteristics from real-world scenes and a publicly accessible synthetic dataset. For each object, we captured between 40 and 60 polarized images using a polarized color camera, FLIR BFS-U3-51S5PC-C. The dataset comprises original images, Stokes vectors, AoLP maps, DoLP maps, masks, and camera pose data. We use the EiSeg software (Hao et al. 2022) of the PaddlePaddle platform to label the target object's mask and use Colmap (Schonberger and Frahm 2016) to extract the pose information of the camera. We utilize SMVP3D Dataset as the synthetic dataset, which includes reflective objects and was published by NeRSP (Han et al. 2024).

Implementation Details

Our implementation was carried out in PyTorch (Paszke et al. 2019), excluding any customized CUDA kernels. The model was optimized for 3000 epochs, with a batch size of 2048 pixel rays. The learning rate we set was 1e-4, and all methods were executed on a single NVIDIA A100 GPU.



(a) GT (b) PANDORA (c) NeRSP (d) PIR (e) NeISF (f) Ours

Figure 5: Visualization comparison of recovered geometry on SMVP3D dataset. Our approach can preserve clearer details.

Method	Dragon	Hedgehog	David	Average
PANDORA	1.23	2.81	3.81	2.62
NeRSP	0.98	**1.38**	<u>2.21</u>	<u>1.52</u>
PIR	<u>0.95</u>	1.56	2.40	1.63
NeISF	N/A	19.72	3.15	11.44
Ours	**0.94**	<u>1.46</u>	**1.88**	**1.43**

Table 1: The table shows the Chamfer Distance (CD) (↓) on ground truth mesh and extracted meshes on the synthetic dataset as a quantitative measure for reconstructed geometry quality. The smallest and second smallest errors are labeled in bold and underlined. "N/A" denotes the experiment where a specific method cannot output reasonable shape estimation results.

Method	Dragon	Hedgehog	David	Average
PANDORA	5.47	8.48	6.27	6.74
NeRSP	4.00	2.16	<u>3.95</u>	3.37
PIR	<u>2.68</u>	<u>1.63</u>	8.14	4.15
Ours-no-pol	<u>2.68</u>	1.68	4.74	<u>2.94</u>
Ours	**2.60**	**1.42**	**3.74**	**2.59**

Table 2: The table shows the mean angular error (MAE) (↓) on ground truth normal maps and predicted normal maps on the synthetic dataset.

Our training is divided into three stages. In the first stage, the algorithm constrains the f_{sdf} using RGB and silhouette to obtain the initial shape, which requires 1500 epochs. In the second stage, we freeze the parameters of the f_{sdf} and train three incident Stokes networks using real Stokes information captured by the camera, which requires 1000 epochs. In the third stage, we unfreeze the f_{sdf} and perform joint optimization on the f_{sdf} and three incident Stokes MLPs, which require 500 epochs. λ_c is set to 1 by default, λ_s is set to 1 in the second and third stages, and λ_{aolp}, like PIR, is set to 0.02 by default. Further implementation details are in the supplement.

Baselines

We compare with several state-of-the-art models, including PANDORA (Dave, Zhao, and Veeraraghavan 2022), PIR (Shao et al. 2024), NeRSP (Han et al. 2024), NeISF (Li et al. 2024). We also designed a method labeled 'Ours-no-pol', which represents a version of our method without the polarization module. It should be noted that PIR is the only method specifically for reconstructing transparent objects, while all other methods are capable of decomposing specular and diffuse reflections, making them suitable for reconstructing reflective objects.

Comparisons

We conducted experiments on both synthetic and real datasets to validate the effectiveness of our method. More comparison results on synthetic and real data can be found in the supplementary material.

Comparisons on The Synthetic Dataset. The quantitative and qualitative comparisons with competing models are respectively showcased in the referenced Tab. 1 and Tab. 2. We also show the visualization results in Figure 5. PANDORA is unable to recover fine details on the surface of objects because it does not account for complex optical reflectance properties, and the applied BRDF model it used can not generalize well to these scenarios. In the dragon scene, NeISF failed to reconstruct. And in the hedgehog scene, due to the occurrence of cavities inside the generated model, the calculated CD value was relatively high. While NeRSP demonstrates promising performance in sparse scenes, it still strug-

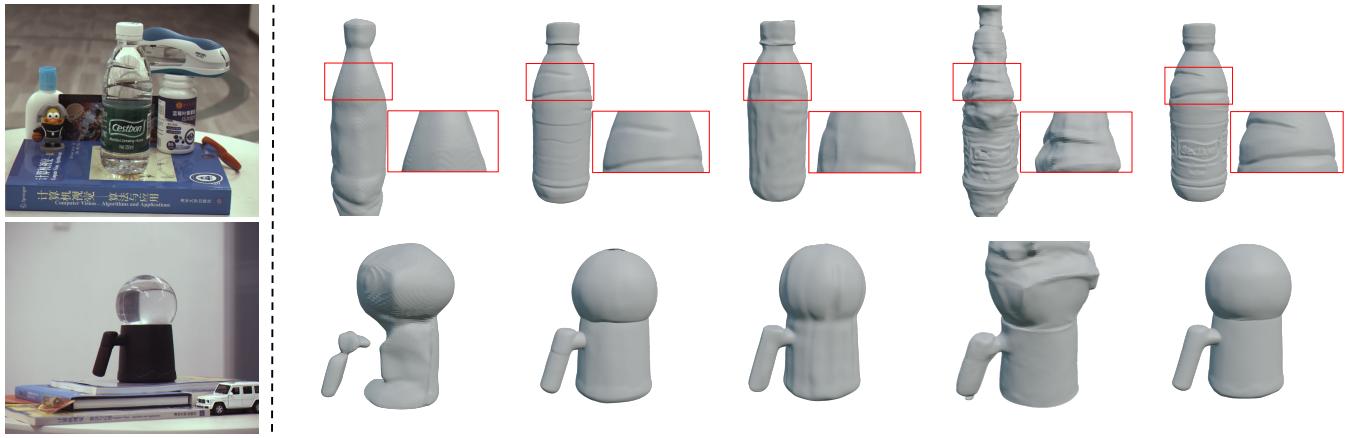


Figure 6: Visualization of the real scene object reconstruction results compared with baselines. Compared to existing methods, our method can preserve clearer texture information

gles to recover finer surface details when sufficient views are available. Since synthetic datasets include accurate AoLP information, PIR with AoLP constraints can perform well. However, the lack of color information as a constraint leads to a relatively smooth reconstructed surface. The experimental comparison results demonstrate that the proposed method can effectively integrate color and polarization information to achieve high-fidelity 3D reconstruction.

Comparisons on The Real Data. The comparative experiment on reconstruction with data collected under natural lighting conditions is presented in Figure 6. PANDORA and NeRSP are difficult to accurately reconstruct objects with complex characteristics such as transparency in natural light environments. Although NeISF can reconstruct objects in polarized incident light, it remains unable to address the challenge of transparency. PIR disregards color and relies solely on masks as constraints, but the assumption of an unpolarized incident light limits the performance to restore detailed 3D reconstruction. Experiment results on real scenarios demonstrate that the proposed method can deliver promising 3D reconstruction along with high-fidelity surface representation.

Ablation Study

In this section, we designed ablation experiments to demonstrate the effectiveness of our polarized cues. Taking the hedgehog object as an example, the results of our ablation experiments are shown in Tab. 3. To further illustrate our method, we present visualizations of the ablation experiment results of the real scene. As shown in Figure 7, our VDC module can alleviate the impact of view-dependent, and the polarization module can help the network obtain more accurate reconstruction results.

Conclusion

In this paper, we propose a polarimetric implicit 3D reconstruction method that integrates geometric and polarization information, enabling the generation of high-quality meshes in complex scenes. To address the challenges of 3D reconstruction in complex media, we introduce a view-dependent physical representation that thoroughly analyzes the subtle properties of reflections. Experimental results demonstrate the superior performance of the proposed method in both real and synthetic scenarios.

Method	VDC	polar	MAE (↓)
Baseline			1.68
w/ VDC	✔		1.55
w/ polar		✔	1.48
Full	✔	✔	**1.42**

Table 3: The table demonstrates the ablation experimental results.

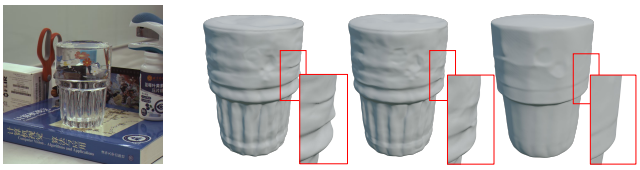


Figure 7: Ablation experiments were conducted on different modules. (a) shows the reference image. (b) shows the result of our baseline, while (c) presents the result after adding the view-dependent confidence map. (d) shows the result after adding the polarization module.

Limitations and Future Work

The view-dependent detection module enables the network to more effectively recover the shape of objects with complex medium. However, when the target object is fully transparent, this algorithm will degrade to some extent. Since transparent objects are common in daily life, our future work will focus on improving the reconstruction of such objects.

Acknowledgments

This work was supported in part by National Natural Science Foundation of China (NSFC) under Grant No. 62302024, in part by Beijing Advanced Innovation Center for Future Blockchain and Privacy Computing under Grant No. GJJ-23-003, and the Fundamental Research Funds for the Central Universities. Sijia Wen is the corresponding author.

References

Ba, Y.; Gilbert, A.; Wang, F.; Yang, J.; Chen, R.; Wang, Y.; Yan, L.; Shi, B.; and Kadambi, A. 2020. Deep shape from polarization. In *Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part XXIV 16*, 554–571. Springer.

Cao, X.; Santo, H.; Okura, F.; and Matsushita, Y. 2023. Multi-view azimuth stereo via tangent space consistency. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 825–834.

Clarke, D. 2009. *Stellar polarimetry*. John Wiley & Sons.

Dave, A.; Zhao, Y.; and Veeraraghavan, A. 2022. Pandora: Polarization-aided neural decomposition of radiance. In *European Conference on Computer Vision*, 538–556. Springer.

Ding, Y.; Ji, Y.; and Ye, J. 2022. Polar-Photometric Stereo Under Natural Illumination. In *2022 International Conference on 3D Vision (3DV)*, 690–699. IEEE.

Ge, W.; Hu, T.; Zhao, H.; Liu, S.; and Chen, Y.-C. 2023. Refneus: Ambiguity-reduced neural implicit surface learning for multi-view reconstruction with reflection. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 4251–4260.

Han, Y.; Guo, H.; Fukai, K.; Santo, H.; Shi, B.; Okura, F.; Ma, Z.; and Jia, Y. 2024. NeRSP: Neural 3D Reconstruction for Reflective Objects with Sparse Polarized Images. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 11821–11830.

Hao, Y.; Liu, Y.; Chen, Y.; Han, L.; Peng, J.; Tang, S.; Chen, G.; Wu, Z.; Chen, Z.; and Lai, B. 2022. Eiseg: an efficient interactive segmentation tool based on paddlepaddle. *arXiv preprint arXiv:2210.08788*.

Kim, K.; Gu, J.; Tyree, S.; Molchanov, P.; Nießner, M.; and Kautz, J. 2017. A lightweight approach for on-the-fly reflectance estimation. In *Proceedings of the IEEE International Conference on Computer Vision*, 20–28.

Kutulakos, K. N.; and Steger, E. 2008. A theory of refractive and specular 3D shape by light-path triangulation. *International Journal of Computer Vision*, 76: 13–29.

Li, C.; Ono, T.; Uemori, T.; Mihara, H.; Gatto, A.; Nagahara, H.; and Moriuchi, Y. 2024. NeISF: Neural Incident Stokes Field for Geometry and Material Estimation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 21434–21445.

Li, Z.; Yeh, Y.-Y.; and Chandraker, M. 2020. Through the looking glass: Neural 3d reconstruction of transparent shapes. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 1262–1271.

Liu, Y.; Wang, P.; Lin, C.; Long, X.; Wang, J.; Liu, L.; Komura, T.; and Wang, W. 2023. Nero: Neural geometry and brdf reconstruction of reflective objects from multiview images. *ACM Transactions on Graphics (TOG)*, 42(4): 1–22.

Lyu, J.; Wu, B.; Lischinski, D.; Cohen-Or, D.; and Huang, H. 2020. Differentiable refraction-tracing for mesh reconstruction of transparent objects. *ACM Transactions on Graphics (TOG)*, 39(6): 1–13.

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.

Mingqi, S.; Chongkun, X.; Zhendong, Y.; Junnan, H.; and Xueqian, W. 2022. Transparent shape from single polarization images. *arXiv e-prints*, arXiv–2204.

Morris, N. J.; and Kutulakos, K. N. 2007. Reconstructing the surface of inhomogeneous transparent scenes by scatter-trace photography. In *2007 IEEE 11th International Conference on Computer Vision*, 1–8. IEEE.

Park, J. J.; Florence, P.; Straub, J.; Newcombe, R.; and Lovegrove, S. 2019. Deepsdf: Learning continuous signed distance functions for shape representation. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 165–174.

Paszke, A.; Gross, S.; Massa, F.; Lerer, A.; Bradbury, J.; Chanan, G.; Killeen, T.; Lin, Z.; Gimelshein, N.; Antiga, L.; et al. 2019. Pytorch: An imperative style, high-performance deep learning library. *Advances in neural information processing systems*, 32.

Peng, S.; Niemeyer, M.; Mescheder, L.; Pollefeys, M.; and Geiger, A. 2020. Convolutional occupancy networks. In *Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part III 16*, 523–540. Springer.

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.

Shao, M.; Xia, C.; Duan, D.; and Wang, X. 2024. Polarimetric inverse rendering for transparent shapes reconstruction. *IEEE Transactions on Multimedia*.

Tanaka, K.; Mukaigawa, Y.; Kubo, H.; Matsushita, Y.; and Yagi, Y. 2016. Recovering transparent shape from time-of-flight distortion. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 4387–4395.

Trifonov, B.; Bradley, D.; and Heidrich, W. 2006. Tomographic reconstruction of transparent objects. In *ACM SIGGRAPH 2006 Sketches*, SIGGRAPH '06, 55–es. New York, NY, USA: Association for Computing Machinery. ISBN 1595933646.

Tsai, C.-Y.; Veeraraghavan, A.; and Sankaranarayanan, A. C. 2015. What does a single light-ray reveal about a transparent object? In *2015 IEEE International Conference on Image Processing (ICIP)*, 606–610. IEEE.

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*.

Wetzstein, G.; Roodnick, D.; Heidrich, W.; and Raskar, R. 2011. Refractive shape from light field distortion. In *2011 International Conference on Computer Vision*, 1180–1186. IEEE.

Xu, J.; Zhu, Z.; Bao, H.; and Xu, W. 2022. Hybrid mesh-neural representation for 3D transparent object reconstruction. *arXiv preprint arXiv:2203.12613*.

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.

Yariv, L.; Kasten, Y.; Moran, D.; Galun, M.; Atzmon, M.; Ronen, B.; and Lipman, Y. 2020. Multiview neural surface reconstruction by disentangling geometry and appearance. *Advances in Neural Information Processing Systems*, 33: 2492–2502.

Zhang, K.; Luan, F.; Wang, Q.; Bala, K.; and Snavely, N. 2021. Physg: Inverse rendering with spherical gaussians for physics-based material editing and relighting. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 5453–5462.