Article:
Wanaset, Rapee, Guarnera, Giuseppe Claudio and Smith, William Alfred Peter (2025) Neural Field Multi-view Shape-from-polarisation. Computer graphics forum. ISSN: 0167-7055

https://doi.org/10.1111/cgf.70177

Eurographics Symposium on Rendering 2025
B. Wang and A. Wilkie
(Guest Editors)

COMPUTER GRAPHICS *forum*
Volume 44 (2025), Number 4

Neural field multi-view shape-from-polarisation

R. Wanaset, G. C. Guarnera and W. A. P. Smith

Department of Computer Science, University of York, UK

Figure 1: *From multi-view* raw *division-of-focal-plane colour-polarisation images (one input view in Col. 1), we reconstruct a neural SDF (surface normals, Col. 2) and neural fields for diffuse and specular radiance (Cols. 3–4). Col. 1 includes a zoom-in inset that enlarges a* 16×16 *patch of the* Colour Polarisation Filter Array *(CPFA), revealing its mosaic pattern. Combining the two radiance fields yields unpolarised radiance (Col. 5). From this representation we can render colour, polarised images from any view; the RGB image rendered at a* 0° *polariser angle appears in Col. 6. All images are rescaled and tone-mapped for display.*

Abstract

We tackle the problem of multi-view shape-from-polarisation using a neural implicit surface representation and volume rendering of a polarised neural radiance field (P-NeRF). The P-NeRF predicts the parameters of a mixed diffuse/specular polarisation model. This directly relates polarisation behaviour to the surface normal without explicitly modelling illumination or BRDF. Via the implicit surface representation, this allows polarisation to directly inform the estimated geometry. This improves shape estimation and also allows separation of diffuse and specular radiance. For polarimetric images from division-of-focal-plane sensors, we fit directly to the raw data without first demosaicing. This avoids fitting to demosaicing artefacts and we propose losses and saturation masking specifically to handle HDR measurements. Our method achieves state-of-the-art performance on the PANDORA benchmark. We apply our method in a lightstage setting, providing single-shot face capture.

CCS Concepts

• ***Computing methodologies*** → ***Reconstruction;***

1. Introduction

When unpolarised light reflects from a surface it becomes partially polarised. This applies to both specular reflection [RC01] and diffuse reflection [AH06] that arises from transmission out of the surface after subsurface scattering. The degree and angle of polarisation are related to the local surface normal direction and view vector and, hence, their measurement provides constraints for the reconstruction of surface geometry. This cue has been exploited by a wide array of shape-from-polarisation methods [MMSG05, AH06, MSB*12, MSB*17, MKI04].

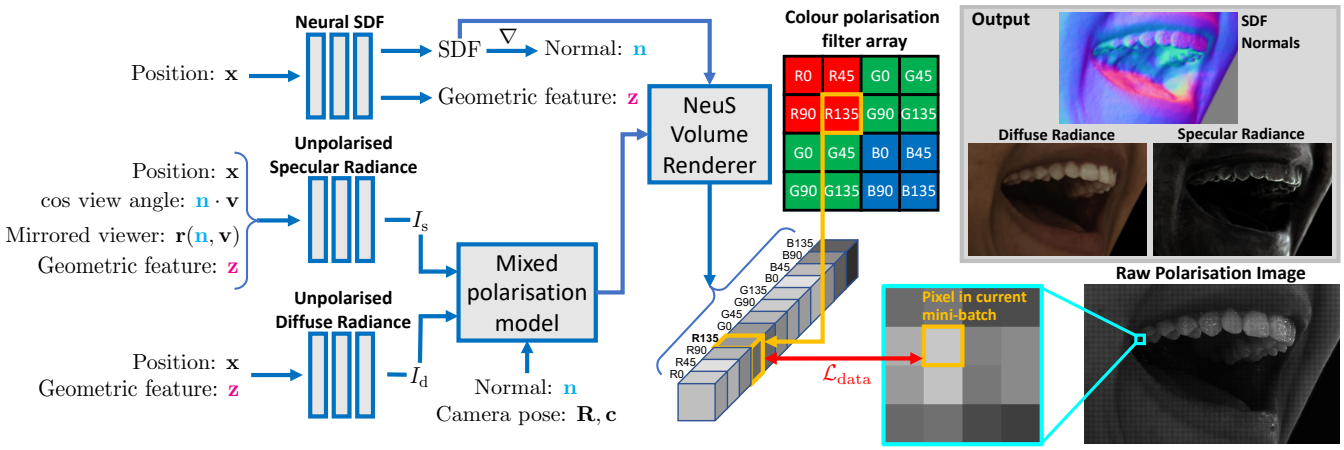

Figure 2: *Neural shape-from-polarisation. We use a neural SDF to represent the surface and whose gradient provides the surface normal. Two other MLPs learn unpolarised diffuse and specular radiance, with diffuse radiance being conditioned on position and geometric features, assuming Lambertian reflectance, and specular radiance additionally on the cosine of the view angle and reflection direction. Via a mixed polarisation model we capture the dependence between surface normal, camera pose and unpolarised radiances to predict polarised radiance. This is volume rendered according to the NeuS [WLL*21] model for any combination of colour channels and polariser angles. We select the appropriate prediction from the channel dimension for each rendered pixel and compute a data loss to the raw measurement.*

However, these methods have seen limited adoption, partly due to the challenges of capturing polarimetric images. Recent advancements, such as commodity division-of-focal-plane (DoFP) sensors that capture polarisation images in a single shot [pol], have mitigated this issue. Nonetheless, polarisation alone is a weak shape cue, providing strong signals only at occluding boundaries for diffuse regions or within sparse specularities.

Multi-view polarisation measurements potentially overcome this restriction. As few as two multi-view measurements of the same point uniquely determine the surface normal direction from polarisation constraints alone (Sec. 3.2). The challenge is to choose a representation that is amenable to optimisation while integrating information from multiple views. The recent rise of neural fields [XTS*22] and their use for implicit surface representation [YKM*20, ZLW*21, LZP*20, KJJ*21] provides a compact and adaptive parameterisation that can be rendered differentiably, e.g. NeRF [MST*20] and NeuS [WLL*21].

Recent works have begun to explore the factorisation, i.e. inverse rendering, of neurally-modelled radiance into underlying physical quantities, including illumination, geometry and material properties via the bidirectional reflectance distribution function (BRDF). Capturing and modelling *polarised* radiance offers the potential for higher accuracy, requirement for fewer input views and the resolution of ambiguities that arise when decomposing RGB radiance alone. A recent line of work integrates polarisation into neural radiance models [DZV22, LOU*24]. However, they require the full Stokes vector at each pixel and perform a full inverse rendering, entailing estimation of the incident illumination and modelling of a polarised BRDF. Instead, we directly exploit the shape-from-polarisation cue in a way that is independent of the illumination environment and make very limited assumptions about material reflectance models. Moreover, we fit our model directly to raw polarisation sensor data, bypassing the need for demosaicing, which is more complex for a Colour Polarisation Filter Array (CPFA) compared to conventional RGB demosaicing (the mosaic pattern has size 4×4 versus 2×2 for RGB). Our main contributions are:

- Neural shape-from-polarisation: we propose a framework to extract geometry and appearance from multi-view polarisation images without inverse rendering illumination or requiring an explicit BRDF model;
- 3D reconstruction: linking surface geometry with polarised radiance through a multiview mixed polarisation model and the surface normals of a neural SDF, we reconstruct object geometry with improved accuracy over the state-of-the-art;
- Passive capture: unlike photometric stereo methods, we estimate object geometry without relying on known enviromental lighting, only assuming unpolarised incident light which is a good approximation for natural light;
- Radiance decomposition: we provide a method to separate diffuse and specular reflections from the object surface;
- Demosaicing: the first neural shape reconstruction method that does not require demosaiced measurements, fitting instead directly to the raw sensor data;
- HDR handling: working with saturated HDR data, we propose a new loss calculated in logarithmic space and apply a conditional mask that ignores saturated information during training.

The intuition behind our approach is that specular and diffuse reflectance produce different, out-of-phase degrees of polarisation. Observing the same point from multiple views allows us to see the same diffuse radiance (Lambertian assumption) but varying specular radiances (dependent on view direction), with degrees of polarisation that follow different models. Together, this provides a cue for extracting specular and diffuse components from polarised image and constraining the surface normal direction and hence the SDF.

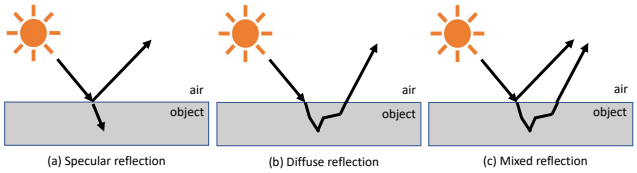

Figure 3: *Unpolarised light incident on a surface can be specularly reflected (a), acquiring partial polarisation as described by the Fresnel equations; (b) diffusely reflected, where light scatters multiple times inside the object before being refracted at the surface, also acquiring partial polarisation due to transmission, out of phase with specular polarisation; (c) mixed reflection, the model used in this work, is the combination of both types of reflection.*

2. Related work

Shape from Polarisation (SfP) The Fresnel equations describe the link between surface normal orientation and polarisation properties of light reflected off a surface. Such link is exploited by SfP techniques, aimed at estimating surface normal from polarimetric measurements. While multi-view stereo typically does not work well with smooth, featureless and glossy surfaces, polarisation can be used on a wide range of materials, such as metals [MMSG05], dielectrics [AH06, GPDG12], dark and shiny surfaces [MSB*12, MSB*17], as well as transparent ones [MKI04, MCZ*22]. Furthermore, *polarisation cameras* are able to record the polarisation state in a single shot, thus providing a dense cue, only limited by camera resolution, and enabling pixel-wise surface normal estimation.

Many polarisation-based methods either deal with specular or diffuse materials, due to the different reflection phenomena (see Fig. 3). Atkinson and Hancock [AH06] assumed diffuse reflectance to estimate the depth map of an object. Miyazaki et al. [MTHI03] proposed a framework [IMTI08] to separate reflection components, from which the objects shape can be inferred. Morel et al. [MMSG05] developed a SfP method aimed at metals, using a specular polariastion model. However, most real-world objects exhibit a mixture of both diffuse and specular polarisation, causing model mismatch [TKR16]. Smith et al. [SRT16] relaxed the classic assumption and classify each pixel as diffuse dominant or specular dominant. Taamazyan et al. [TKR16] used both viewpoint and polarisation information to recover shape of an object, relying on a mixed polarisation model. Polarisation data from at least 2 viewpoints constrains surface normal estimation, posed as a non-linear least square problem. Cui et al. [CGS*17] used polarimetric multi-view stereo to handle a variety of objects with mixed polarisation reflection, using iso-depth contours to propagate depth.

An additional challenge for SfP methods is the azimuthal ambiguity, i.e. two azimuthal angles shifted by π radians cannot be distinguished. Miyazaki et al. [MSB*17] used space carving to estimate the rough structure of an object, before integrating priors to the SfP pipeline. Similarly, Zhu and Smith [ZS19] used multi-view information and a coarse depth map obtained from stereo cues as a guide surface for disambiguation. Kadambi et al. [KTSR15] combined a single polarisation image with the depth map obtained from a RGBD camera, the latter used to disambiguate normal direction. Tian et al. [TPW*23] propose a depth estimation from polarimetric stereo. The polarimetric ambiguities are resolved by iso-depth constraints. The dual-GRU (Gated Recurrent Unit) architecture is employed to utilise multi-domain similarity - the similarity of RGB and polarisation images.

By establishing consistency of tangent space among multiple viewpoints, MVAS [CSOM23] manages to reconstruct textureless 3D objects which have been challenging for conventional stereo methods. PolarPMS [ZOMO24] exploits photometric and polarimetric consistencies. While being able to reconstruct 3D object up to pixel-level resolution, in contrast to ours which employs SDF to implicitly infer geometry, PolarPMS iteratively generates several pairs of depth and normal hypotheses and picks the one that minimises inconsistency between views. PMVIR [ZMO20, ZMO23] address the multi-view SfP problem using a mesh-based representation and do not fully resolve the ambiguity, relying instead on the most plausible azimuth angle. Moreover, being a refinement method, PMVIR requires a reasonable initial shape whereas ours learns the geometry from scratch. Ba et al. [BGW*20] and Lei et al. [LQX*22] tackle the monocular SfP problem, recovering only a single normal map from one view. Both methods train networks using data from many scenes, learning general priors that generalise to unseen novel input. In contrast to our work, we fit a neural representation to a single scene/object using multi-view input. This provides much higher quality reconstruction and a complete representation of object geometry as opposed to per-view normal maps.

Furthermore, in our work, we explicitly include a mixed polarisation model and let our system learn an implicit neural representation, encoding 3D geometry and appearance from multi-view 2D polarisation images via volume rendering techniques. Our networks map viewing direction and surface location to normal direction, diffuse and specular reflections, then render polarisation images for comparison with the raw images (see Fig. 2). Table 1 gives an overview of methods' attributes. Among SfP methods, P-NeRF shares the largest commonality to PANDORA and PMVIR. For a dedicated review, see Tan et al. [TJK*23].

Specular-diffuse separation Reflected radiance can be accurately modelled as a combination of specular and diffuse reflections. Specular reflection occurs when light is reflected off a smooth surface, whereas diffuse reflection could be due to either subsurface scattering or reflection from a rough surface (see Fig. 3).

While some methods rely on pixel intensity [Sha92, MZBK06, SMNO18] to separate reflectance components, polarisation-based reflectance separation has been widely exploited. Riviere et al. [RRFG17] proposed a passive method for uncontrolled environments, that estimates the reflectance of a planar surface using polarisation measurements from 3 views, one along the normal and two from viewpoints near the Brewster angle. Nogué et al. [NLG22] proposed planar surface reflectometry using a near-field display, which requires 3 linear polarisation measurements.

Several recent techniques employ neural networks for radiance separation. Inspired by real-time graphics, Boss et al. [BJB*21] presented a pre-integrated lighting network that converts illumination integration process into a query network, the latter resulting in efficient rendering and radiance decomposition. PhySG [ZLW*21]

tackles radiance separation by using Spherical Gaussians and data-driven embedding to model reflectance and lighting respectively. Dave et al. [DHGH*22] proposed a 2-step method to decompose specular and diffuse reflectances from a single polarimetric image. The initial separation is done by analysing the relationship among polarisation cues and reflected radiance, then refined by the network trained on synthetic scenes.

Neural field SfP Closest to our work, PANDORA [DZV22] incorporates polarisation properties into a neural inverse rendering pipeline. Both approaches use multi-view polarisation images, COLMAP [SZPF16, SF16] camera poses, and binary masks. However, PANDORA performs full inverse rendering, modelling the specular BRDF and incident illumination environment, passing this through a polarised BRDF model and rendering Stokes vectors that are compared to those recorded by the camera. This requires first demosaicing the raw images (a challenging task in itself) to provide full Stokes vectors at every pixel. In contrast, we use black box neural fields to learn diffuse and specular radiances and use only a mixed polarisation model relating surface normal direction to degree and angle of polarisation. This allows us to directly fit to the raw data without demosaicing. GNeRP [LWLC24] uses a Gaussian representation for surface normals, with the mean indicating overall orientation and covariance capturing high-frequency variation. It employs DoLP reweighting to balance higher DoLP in specular regions. Unlike GNeRP, our work volume-renders radiance which is linked with surface normal through our mixed polarisation model. Both methods, though, use implicit neural BRDFs which can handle high-frequency details. Han et al. [HGF*24] focus on polarimetric cues, splitting them into geometric and photometric cues derived from Stokes vectors. While many works assume single-bounced illumination, polarised rays contain rich information (e.g., as modeled by [KJCB23]). NeISF [LOU*24] relaxes this assumption, using coordinate-based MLPs to capture the Stokes field of the second-last bounce. Diffuse and specular reflections are modelled separately: diffuse Mueller matrices depend on surface normals, while specular ones depend on microfacet normals.

Lightstage-based capture In general, lightstage-based methods use spherical illumination patterns to perform a version of photometric stereo. Many light-stage appearance-capture pipelines aim to recover not only geometry but also spatially varying material parameters (see [GGH*17] for a survey). For example, Ma et al. [MHP*07] introduced the spherical gradient illumination patterns while Lattas et al. [LWZG19] use spherical binary patterns. This provides a per-pixel estimate of the surface normal and material properties including diffuse albedo and specular intensity. This cue can be integrated across views using a multiview stereo type approach [GFT*11]. Lattas et al. [LLK*22] use less-constrained, desk-based illumination constructed from a panel of LCDs.

Some of these methods use properties of polarisation for the purposes of separating diffuse and specular reflection. This is based on a simplistic model in which specular reflection is assumed to perfectly preserve the plane of linear polarisation while diffuse reflection completely depolarises the reflected light. In fact, diffuse reflection caused by subsurface scattering (as in human skin) partially polarises the light that is transmitted out of the surface (regardless of whether the incident light was polarised or not) and specular reflection similarly partially polarises unpolarised incident light when it is reflected. It is this shape-from-polarisation cue that our method exploits, negating the need for varying illumination patterns.

Method	Viewpoints	One-shot possible	Polarisation and X	Polarisation model (diffuse/specular)	Primary input
P-NeRF (ours)	dense	yes	no X	both	raw polarisation images
[ZS19]	pair	yes	2-view stereo	either	polarisation and RGB images
PANDORA [DZV22]	dense	yes	no X	both	polarisation images
PMVIR [ZMO20]	dense	yes	no X	both	polarisation images
NeISF [LOU*24]	dense	no	no X	both	bracketed images
[SRT16]	single	yes	shading cues	either	polarisation image
[LQX*22]	single	yes	no X	either	polarisation image
[BGW*20]	single	yes	no X	both	polarisation image
NeRSP [HGF*24]	sparse	yes	photometric cues	both	polarisation images
[AH06]	triplet	yes	no X	diffuse	polarisation images
[MMSG05]	dense	no	no X	specular	polarisation images
[KTSR15]	single	yes	coarse depth map	either	coarse depth map and polarisation image

Table 1: ***Method attributes.*** *Modern methods usually consider both types of polarisation models, either one at a time or co-existed models, whereas the traditional methods (e.g. [AH06, MMSG05]) solely focus on specific model. While most methods allow single-shot capture, some methods (e.g. [LOU*24, MMSG05] rely on special setup, prohibiting simultaneously capturing. Because of polarimetric ambiguities, single-view methods struggle to reconstruct 3D geometry on its own and have to consider extra input such as [SRT16] using shading cues or [LQX*22, BGW*20] with pre-tranied networks). Among mixed-model methods with dense input, there are only 3 of them that allow single-shot capturing: P-NeRF, PANDORA and Polarimetric MVIR. The result comparison for this method category is shown in section 7.*

In addition, the diffuse/specular separation technique used by [MHP*07] requires each light source in the lightstage to be polarised with a particular plane of polarisation. The orientation of the polarisation filters can only be tuned for a single viewpoint. This means that such a design cannot be used to provide multiview information. This reduces coverage of the face but also means that geometric, multiview shape cues cannot be combined with the photometric cues. [GFT*11] proposed an alternative in which polariser orientations on the lights were arranged such that an approximate separation was possible from any viewpoint on the equator of the dome. This allows the use of multiview information but at the cost of losing exact diffuse/specular separation.

3. Model mechanism

We begin by describing our mixed polarisation model and then extending it to the multi-view case.

3.1. A mixed polarisation model

When (partially) polarised light passes through a linear polarising filter and is then measured by a camera, the recorded intensity varies sinusoidally with the angle of rotation (ϑ) of the linear polariser. For subsurface diffuse reflectance that arises from internal scattering followed by transmission out of the surface, this is described by the diffuse polarisation model [AH06]:

$$I_{\vartheta}^{\text{diffuse}} = I_d(1 + \rho_d(\theta)\cos(2\vartheta - 2\phi(\varphi))), \tag{1}$$

where $I_d \in \mathbb{R}^3_{\geq 0}$ is the unpolarised RGB diffuse radiance, $\rho_d \in [0,1]$ is the diffuse degree of linear polarisation (DoLP) which depends on the zenith angle $\theta \in [0, 90^\circ]$ of the surface normal in a viewer-centred coordinate system, i.e. the angle between surface normal

and view vectors, and $\phi(\varphi) = \varphi \bmod \pi \in [0, 180^\circ)$ is the phase angle which depends on the azimuth angle $\varphi \in [0, 360^\circ)$ of the surface normal in a viewer-centred coordinate system. In general, the DoLP describes the proportion of a wave that is linearly polarised. Specifically, ρ_d characterises the proportion of diffusely reflected light that is linearly polarised. There exists a similar expression for polarised specular reflectance:

$$I_\vartheta^{\text{specular}} = I_s(1 + \rho_s(\theta)\cos(2\vartheta - 2\phi(\varphi) + \pi)), \tag{2}$$

where $I_s \in \mathbb{R}^3_{\geq 0}$ is the unpolarised RGB specular radiance in the direction of the viewer and $\rho_s \in [0,1]$ is the specular degree of polarisation. Note that both ρ_d and ρ_s have fixed expressions derived from physical laws (Fresnel theory) and also depend on the refractive index of the surface which can either be assumed known or estimated [SRT16].

For a glossy surface exhibiting a superposition of specular and (subsurface) diffuse reflection, we can write a mixed polarisation model as the sum of $I_\vartheta^{\text{diffuse}}$ and $I_\vartheta^{\text{specular}}$:

$$\begin{aligned} I_\vartheta^{\text{mixed}} &= I_\vartheta^{\text{diffuse}} + I_\vartheta^{\text{specular}} \\ &= (I_d + I_s) + (I_d\rho_d(\theta) - I_s\rho_s(\theta))\cos(2\vartheta - 2\phi(\varphi)). \end{aligned} \tag{3}$$

3.2. Multi-view mixed polarisation model

We now rewrite the mixed polarisation model to explicitly account for the transformation from a world coordinate system to the coordinate system of a given camera.

We denote a 3D point in world coordinates as $\mathbf{x} = (x, y, z)$ and the surface normal in world coordinates at that point as $\mathbf{n}(\mathbf{x}) = [n_x(\mathbf{x}), n_y(\mathbf{x}), n_z(\mathbf{x})]^T$ with $\|\mathbf{n}(\mathbf{x})\| = 1$. We define camera pose by a rotation matrix $\mathbf{R} \in SO(3)$ that rotates world to camera coordinates and the position of the camera centre by $\mathbf{c} \in \mathbb{R}^3$. Hence, the view direction from which a camera with centre $\mathbf{c}$ observes point $\mathbf{x}$ is given by: $\mathbf{v}(\mathbf{c}, \mathbf{x}) = \mathbf{c} - \mathbf{x}/\|\mathbf{c} - \mathbf{x}\|$. The spherical coordinates of the surface normal at $\mathbf{x}$ in the camera coordinate system can be obtained as follows. The zenith angle is given by the angle between $\mathbf{n}(\mathbf{x})$ and $\mathbf{v}(\mathbf{c}, \mathbf{x})$: $\theta(\mathbf{c}, \mathbf{x}, \mathbf{n}) = \arccos(\mathbf{n}(\mathbf{x}) \cdot \mathbf{v}(\mathbf{c}, \mathbf{x}))$, while the azimuth angle is given by rotating the surface normal to camera coordinates: $\mathbf{n}_c(\mathbf{R}, \mathbf{x}, \mathbf{n}) = \mathbf{R}\,\mathbf{n}(\mathbf{x})$, then converting the Cartesian representation to the spherical azimuth angle:

$$\varphi(\mathbf{R}, \mathbf{x}, \mathbf{n}) = \text{atan2}\left(\frac{n_{c,y}(\mathbf{R}, \mathbf{x}, \mathbf{n})}{n_{c,x}(\mathbf{R}, \mathbf{x}, \mathbf{n})}\right). \tag{4}$$

We can now rewrite our multi-view mixed model as a function of all free parameters: position ($\mathbf{x}$), the camera pose ($\mathbf{R}, \mathbf{c}$), the surface normal ($\mathbf{n}$) and diffuse and specular unpolarised radiances (I_d, I_s):

$$\begin{aligned} &I_\vartheta^{\text{mixed}}(\mathbf{x}, \mathbf{R}, \mathbf{c}, \mathbf{n}, I_d, I_s) = (I_d + I_s) + \\ &[I_d\rho_d(\theta(\mathbf{c}, \mathbf{x}, \mathbf{n})) - I_s\rho_s(\theta(\mathbf{c}, \mathbf{x}, \mathbf{n}))]\cos[2\vartheta - 2\phi(\varphi(\mathbf{R}, \mathbf{x}, \mathbf{n}))]. \end{aligned} \tag{5}$$

For a single view, there are four unknowns in this expression: two components of the orientation of the surface normal and I_d and I_s (assuming grayscale radiance). A single observation with a demosaiced polarisation camera provides three observations (the three parameters of a sinusoid: unpolarised intensity, degree of polarisation and angle of polarisation). Hence, inverting the mixed model is ill-posed for one view. The surface normal is independent of the viewpoint and, under the Lambertian assumption, so is I_d. If we add a second view, this means we only add one additional unknown but gain three more observations at which point the problem becomes well-posed. This provides the motivation for multi-view shape-from-polarisation, though note in practice the correspondences between pixels in different views are unknown making the problem harder than this theoretical well-posed version.

4. Image formation

We now combine our multi-view mixed polarisation model with an image formation model and state our simplifying assumptions. This leads us to our neural formulation.

We assume single-bounce reflectance (i.e. no interreflections), diffuse reflectance (from subsurface scattering) independent of view direction (i.e. Lambertian model), unpolarised incident illumination that is arbitrarily distributed but fixed across images, a refractive index fixed to 1.5, isotropic specular BRDF. Thus the diffuse term is view-independent, while the specular term remains view-dependent, as made explicit in Eq. (6). Under these assumptions, we can write the rendering equation for the outgoing radiance, L_o, from position $\mathbf{x}$ towards the viewer, $\boldsymbol{\omega}_\text{o} = \mathbf{v}$, as the sum of diffuse and specular radiance:

$$L_\text{o}(\mathbf{x}, \boldsymbol{\omega}_\text{o}) = \overbrace{\frac{\mathbf{b}_\text{d}}{\pi}\int_\Omega L_\text{i}(\mathbf{x}, \boldsymbol{\omega}_\text{i})(\boldsymbol{\omega}_\text{i} \cdot \mathbf{n})d\boldsymbol{\omega}_\text{i}}^{\text{diffuse}} + \underbrace{\int_\Omega f_\text{s}(\mathbf{x}, \boldsymbol{\omega}_\text{i}, \boldsymbol{\omega}_\text{o}; b_\text{r})(\boldsymbol{\omega}_\text{i} \cdot \mathbf{n})L_\text{i}(\mathbf{x}, \boldsymbol{\omega}_\text{i})d\boldsymbol{\omega}_\text{i}}_{\text{specular}}, \tag{6}$$

where f_s denotes the specular portion of the BRDF, $\mathbf{b}_\text{d} \in \mathbb{R}^3$ is the diffuse albedo, b_r is the roughness and $L_\text{i}(\mathbf{x}, \boldsymbol{\omega}_\text{i})$ is the incident radiance from direction $\boldsymbol{\omega}_\text{i}$. For fixed illumination, since both diffuse parameters $\mathbf{b}_\text{d}$ and surface normal $\mathbf{n}$ depend on position, the diffuse radiance depends only on position $\mathbf{x}$, which we denote $F_{I_d}(\mathbf{x})$. Note that this function learns the combined effect of diffuse shading (including baked self-occlusion of illumination) and texture (i.e. spatially varying diffuse albedo).

Following the rationale of pre-integrated lighting [MS16, BJB*21], the specular radiance can be approximated by a sum of two integrals, both dependent on position $\mathbf{x}$. One of the integrals is independent of the incident radiance, and models the geometric attenuation and Fresnel reflectance for cosine viewer angle $\cos\theta_v = \boldsymbol{\omega}_\text{o} \cdot \mathbf{n}$ of a microfacet specular BRDF. The second integral involves the inner product between the microfacet distribution D and the lighting, thus leading to the following expression for pre-integrated lighting, that depends only on the mirrored viewer direction $\mathbf{r}(\mathbf{n}, \mathbf{v}) = \mathbf{v} - 2(\mathbf{v} \cdot \mathbf{n})\mathbf{n}$ and surface characteristic, roughness b_r: $\tilde{L}_\text{i}(\mathbf{r}, b_\text{r}) = \int_\Omega D(b_\text{r}, \boldsymbol{\omega}_\text{i}, \mathbf{r})L_i(\mathbf{x}, \boldsymbol{\omega}_\text{i})d\boldsymbol{\omega}_\text{i}$. This split does not change the underlying physics: it merely re-parameterises the integral into smoother, lower-dimensional factors that our specular network can learn more easily [VHM*22]. Hence, we can reduce specular radiance to a function $F_{I_s}(\mathbf{x}, \mathbf{r}, \cos\theta_v, b_\text{r})$ of position, mirrored viewer, view angle in the local reference frame and roughness.

We do not attempt to estimate the illumination environment, L_i,

nor do we explicitly represent or estimate the specular BRDF, diffuse albedo or self-occlusion. Instead, we model diffuse and specular radiance as black boxes dependent on position, viewing direction and surface parameters. We then endow these unpolarised radiances with DoLP and AoLP according to the model in Eq. (3) which in turn depends on the surface normal direction.

5. Polarised Neural Radiance Fields

To implement P-NeRF, we jointly train 3 networks: one representing geometry via a signed distance function and two radiance networks (specular and diffuse). The radiance networks output unpolarised radiance that we use in Eq. (5) to compute polarised radiance and then volume render.

5.1. Scene representation

We represent surface geometry with a neural signed distance function [WLL*21, ZLW*21]: $F_{\text{SDF}} : \mathbb{R}^3 \rightarrow \mathbb{R}$. We can define the surface normal in world coordinates, $\mathbf{n}(\mathbf{x})$, at any position $\mathbf{x}$ where $F_{\text{SDF}}(\mathbf{x}) \approx 0$ as: $\mathbf{n}_{\text{SDF}}(\mathbf{x}) = \nabla F_{\text{SDF}}(\mathbf{x}) / \|\nabla F_{\text{SDF}}(\mathbf{x})\|$. We implement our method in Nerfstudio [TWN*23], building on top of NeuS-Facto [YCA*22]. For efficiency, this includes a proposal network [BMV*22] and hashgrid [MESK22] for learnt network input features. The input to the SDF network is therefore the learnt features from the hashgrid. The SDF network also outputs a geometric feature vector $\mathbf{z} \in \mathbb{R}^G$ which is used to pass features derived from the hashgrid features to the radiance networks. We assume that $\mathbf{z}$ implicitly encodes surface roughness b_r.

We use two neural fields to store unpolarised diffuse and specular radiance as described in Sec. 4. Combined with the surface normal provided by the SDF and the camera parameters for a training or validation view, these provide the inputs to our multi-view mixed polarisation model (Eq. (5)). Since diffuse radiance does not depend on viewer direction we use simply $F_{I_d} : (\mathbf{x}, \mathbf{z}) \mapsto I_d$ to map a 3D position to the RGB unpolarised diffuse radiance at that point. Specular reflectance does depend on viewing geometry, parameterised as discussed above: $F_{I_s} : (\mathbf{x}, \max(\mathbf{n} \cdot \mathbf{v}, 0), \mathbf{r}(\mathbf{n}, \mathbf{v}), \mathbf{z}) \mapsto I_s$. Again, the output is RGB unpolarised specular radiance at that point, but this time in the direction, $\mathbf{v}$, of the viewer. Since viewing rays may not be restricted to the upper hemisphere about the normal, we clamp the cosine of the view angle to be non-negative. We apply positional encoding to $\mathbf{x}$, $\mathbf{r}$ and $\cos\theta_v$ with 6, 4 and 4 frequencies respectively in order to allow the radiance networks to learn sharp details in surface albedo and specular reflections.

5.2. Rendering

We now show how to perform polarised volume rendering of the scene representation in Sec. 5.1 using the multi-view mixed polarisation model described in Sec. 3.2.

Since polariser angle is defined with respect to the camera coordinate frame, we cannot define polarised radiance along a ray with unspecified camera coordinate system. So we define the ray that we render in terms of a camera pose $(\mathbf{R}, \mathbf{c})$ and a pixel with normalised coordinates $\mathbf{u} = [u, v, 1]^\top$. The ray for this pixel is given by $\{\mathbf{p}(\mathbf{R}, \mathbf{c}, \mathbf{u}, t) = \mathbf{c} + t\mathbf{R}^\top \mathbf{u}, t \geq 0\}$.

We accumulate polariser-angle-dependent colours along the ray using NeRF [MST*20] time-discrete volume rendering:

$$I(\mathbf{R}, \mathbf{c}, \mathbf{u}, \vartheta) = \sum_{i=1}^{S} w(t_i) I_\vartheta^{\text{mixed}}(\mathbf{x}_i, \mathbf{R}, \mathbf{c}, \mathbf{n}_i, F_{I_d}(\mathbf{x}_i, \mathbf{z}_i), F_{I_s}[\mathbf{x}_i, \max(\mathbf{n}_i \cdot \mathbf{v}_i, 0), \mathbf{r}(\mathbf{n}_i, \mathbf{v}_i), \mathbf{z}_i]),$$

where $\mathbf{x}_i = \mathbf{p}(\mathbf{R}, \mathbf{c}, \mathbf{u}, t_i)$, $\mathbf{n}_i = \mathbf{n}_{\text{SDF}}(\mathbf{x}_i)$, $\mathbf{v}_i = \mathbf{v}(\mathbf{c}, \mathbf{x}_i)$, $\mathbf{z}_i$ are the geometric features from F_{SDF} at $\mathbf{x}_i$, the t_i are the S sample points along the ray and $w(t_i)$ the volume rendering weight given by the density derived from the SDF value, as in NeuS [WLL*21].

5.3. Training

We use a colour polarisation camera with DoFP sensor. This means that each pixel records only one of the 12 combinations of colour channel and polariser angle recorded by the camera using a 4×4 colour polarised mosaic (see Fig. 2). Demosaicing these measurements is still an open problem [MLB18] and introduces unnecessary approximation error. Instead, we train directly on the raw measurements with a single channel per pixel, achieving a sharp reconstruction demonstrated in Sec. 7. Fitting to raw measurement was done in RawNeRF [MHMB*21] for RGB images but we are the first to do it for a CPFA sensor.

We randomly sample a batch of K pixels from an image at every iteration. A sampled pixel is represented by $P = (I_k, c_k, \vartheta_k, M_k, \mathbf{u}_k, \mathbf{R}_k, \mathbf{c}_k)$ where I_k is the measured scalar value at that pixel, $c_k \in \{R, G, B\}$ and $\vartheta_k \in \{0^\circ, 45^\circ, 90^\circ, 135^\circ\}$ respectively denotes the colour channel and polariser angle recorded at that pixel, $M_k \in \{0, 1\}$ is a mask, and the undistorted pixel coordinate $\mathbf{u}_k$ and camera pose $(\mathbf{R}_k, \mathbf{c}_k)$ define the ray. Here, $\mathbf{u}_k$ is the location that the same light ray would hit on an ideal pinhole (lens-distortion-free) image plane, obtained by applying the inverse distortion model to the integer pixel coordinate. Note that because we do not demosaic *we cannot undistort* the images. Instead, we precompute the (non-integer) undistorted $\mathbf{u}_k$, for every integer pixel coordinate and use this to define the ray.

We train our networks to minimise the following loss: $\mathcal{L} = \mathcal{L}_{\text{recon}} + \lambda_{\text{mask}}\mathcal{L}_{\text{mask}} + \lambda_{\text{theta}}\mathcal{L}_{\text{theta}} + \lambda_{\text{smooth}}\mathcal{L}_{\text{smooth}}$. Due to our unique setting, the reconstruction loss has to adapt, which will be discussed in section 5.4. The mask loss encourages the silhouette of the SDF surface to match the binary object masks:

$$\mathcal{L}_{\text{mask}} = \frac{1}{K} \sum_k |(M_k - O_k)|, \tag{7}$$

where O_k is the accumulated density along the ray. The DoLP depends on the zenith angle between view vector and surface normal. For a visible surface point (whose volume rendering weight is > 0), this means that the zenith angle must be $< 90^\circ$ and the DoLP is only defined in this range. For this reason, we add a theta loss encouraging the SDF network to produce a reasonable zenith angle:

$$\mathcal{L}_{\text{theta}} = \frac{1}{K} \sum_k \sum_{i=1}^{S} -w(t_i) \min(\mathbf{v}_k \cdot \mathbf{n}_k, 0). \tag{8}$$

Finally, in the presence of noise we can optionally introduce a surface smoothness prior that penalises sharp changes in surface nor-

mal direction [ZMO20]:

$$\mathcal{L}_{\text{smooth}} = \frac{1}{K}\sum_k \arccos \mathbf{n}_k(\mathbf{x}) \cdot \mathbf{n}_k(\mathbf{x}+\boldsymbol{\epsilon}), \tag{9}$$

where $\boldsymbol{\epsilon} \sim \mathcal{N}(\mathbf{0}_3, \sigma \mathbf{I}_3)$ is a random displacement with σ set in the range of 0 to 0.001. Lastly, following Mip-NeRF 360 [BMV*22] we add a proposal loss and, following NeuS [WLL*21], an Eikonal loss to encourage the neural field F_{SDF} to correspond to a valid SDF.

5.4. High dynamic range handling

Reflected radiance from glossy surfaces exhibits a very high dynamic range between diffuse and specular reflections. This is captured to some extent by the raw data from the camera sensor; however, strong specularities can still result in saturated pixels. We therefore handle the HDR nature of the signal in three ways.

First, we propose a novel masking strategy for handling saturations. For a sensor with bit depth d, the largest value that can be recorded ($I_{\max}$) is $2^d - 1$. Either the observed or predicted intensity can be saturated ($\geq I_{\max}$). We note that the only scenario in which the gradient of the reconstruction loss provides no useful information is when both observed and predicted values are saturated. In that case, the model lacks useful information on its distance from ground truth data. Thus, we define a saturation mask as:

$$M_k^{\text{sat}} = (I_k < I_{\max}) \vee (I_{c_k}(\mathbf{R}_k, \mathbf{c}_k, \mathbf{u}_k, \vartheta_k) < I_{\max}). \tag{10}$$

The mask therefore *never* removes a pixel from optimisation when either the measurement or the prediction is below saturation; it only zeroes the loss when both are saturated, where the gradient is uninformative anyway.

Second, we find that typical LDR reconstruction losses such as L1 and even losses designed specifically for HDR [MHMB*21] did not perform well on our HDR data. Instead, we found the best performance was obtained using a combination of L1 in a scaled and shifted log space and an L2 loss to encourage reconstruction of specularities, giving the following reconstruction loss: $\mathcal{L}_{\text{recon}} = \lambda_{\text{HDR}}\mathcal{L}_{\text{HDR}} + \lambda_{\text{spec}}\mathcal{L}_{\text{spec}}$. The HDR loss, including saturation and foreground masking, is defined as:

$$\mathcal{L}_{\text{HDR}} = \frac{1}{K}\sum_k M_k M_k^{\text{sat}} \left|\log(\alpha I_k + \beta) - \log(\alpha I_{c_k}(\mathbf{R}_k, \mathbf{c}_k, \mathbf{u}_k, \vartheta_k) + \beta)\right|, \tag{11}$$

where I_{c_k} denotes selecting colour channel c_k from the rendered RGB quantity. We set $\alpha = 60$ and $\beta = 3$ so that the measured intensities follow an approximately normal distribution. To promote reconstruction of specular regions, we define the specular loss as:

$$\mathcal{L}_{\text{spec}} = \frac{1}{K}\sum_k M_k M_k^{\text{sat}} \left(I_k - I_{c_k}(\mathbf{R}_k, \mathbf{c}_k, \mathbf{u}_k, \vartheta_k)\right)^2. \tag{12}$$

Finally, we use exponential activation for the outputs of the radiance networks to help them generate full dynamic range.

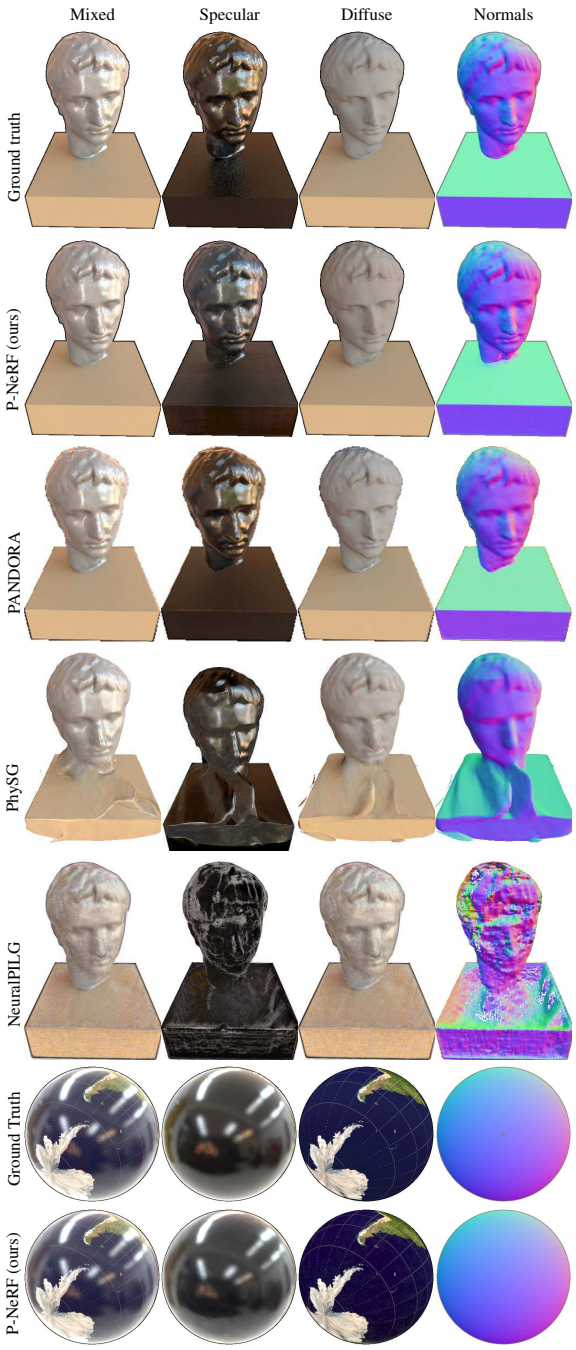

Figure 4: ***Reflectance decomposition and geometry estimation against ground truth synthetic data.*** *The 1st and the 6rd row respectively show the ground truth for the bust and the globe datasets.*

6. A shape-from-polarisation single-shot lightstage

We now describe the application of our method in a specific lighting condition. While more constrained than the general approach, this variant provides additional information about surface material properties and offers an attractive alternative to previous lightstage-based acquisition techniques. Specifically, we apply our method in spherically-uniform, unpolarised illumination. In contrast to previous methods [MHP*07, GFT*11, LWZG19, LLK*22], we do not modulate individual light sources to create varying illumination

Figure 5: ***Reflectance decomposition and geometry estimation on real data.*** *From left to right, unpolarised radiance, specular reflection, diffuse reflection and normal map. (Zoom for detail).*

patterns, we only require a single image from each camera under a fixed, uniform illumination. The method is therefore single-shot and suitable for dynamic objects such as faces, potentially running at full frame rate. Because the illumination is unpolarised, we make no assumption about its plane of polarisation relative to the camera. Rather than using polarisation to approximate diffuse/specular separation, we exploit it purely as a shape cue.

Under uniform illumination, $L_{\mathrm{i}}(\mathbf{x},\boldsymbol{\omega}_{\mathrm{i}}) = k, \forall \boldsymbol{\omega}_{\mathrm{i}} \in \Omega$, for an arbitrary constant k. The diffuse term in Eq. (6) then reduces to:

$$L_{\mathrm{diffuse}}(\mathbf{x},\boldsymbol{\omega}_{\mathrm{o}}) = \frac{\mathbf{b}_{\mathrm{d}}}{\pi} k \int_{\Omega} (\boldsymbol{\omega}_{\mathrm{i}} \cdot \mathbf{n}) d\boldsymbol{\omega}_{\mathrm{i}} = \mathbf{b}_{\mathrm{d}} k \tag{13}$$

In other words, the diffuse radiance estimated by fitting our model directly provides the diffuse albedo map (up to a scaling). Under the same lighting, the specular radiance becomes a view-dependent reflection coefficient that, after Fresnel compensation, yields the specular albedo (R_0) map [GFT*11]. In practice, we capture multi-view face images under uniform, unpolarised light, run our pipeline (Sec. 5), take F_{I_d} as diffuse albedo, and evaluate F_{I_s} at $\mathbf{v} = \mathbf{n}$ to obtain the specular albedo.

7. Experiments

7.1. Under spatially-varying illumination

Real data is captured using a Lucid 5.0MP colour polarisation camera with a $d = 12$ bit Sony IMX250MYR CMOS sensor with 35-60 images per object. The capture environment varies from object

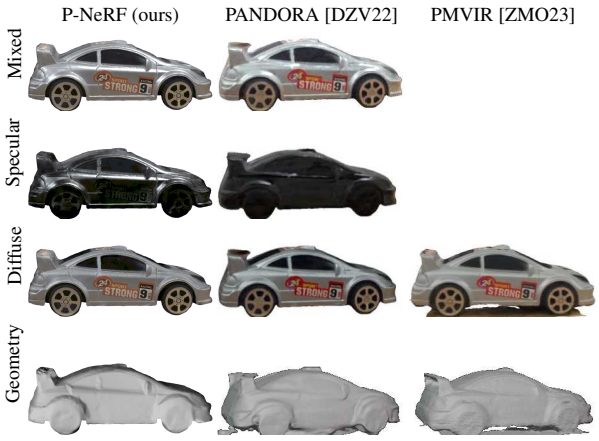

Figure 6: ***Reflectance decomposition and geometry estimation on car dataset.*** *From top to bottom, unpolarised radiance, specular reflection, diffuse reflection, and geometry.*

to object but exposure was set to aim for limited saturated pixels though some do occur and depth of field effects mean some regions in some images are blurred. We estimate camera poses by applying bilinear demosaicing and nonlinear gamma to the raw images and providing the RGB, 0° polariser angle images to the COLMAP structure-from-motion pipeline [SZPF16, SF16]. The demosaiced images are not used anywhere else in our pipeline. In all of our experiments we use loss weights $\lambda_{\mathrm{mask}} = 0.1$, $\lambda_{\mathrm{theta}} = 100$, $\lambda_{\mathrm{HDR}} = 2$ and $\lambda_{\mathrm{spec}} = 0.03$. In terms of geometric network architecture, we use a MLP of 2 layers with the size of 256 to encode SDF; the feature vector is set to have the size of $G = 256$; softplus with $\beta = 100$ is used as the activation function. Specular and diffuse reflections are modelled with 2 separated networks. They are identical in shape with 6 layers of size 256 for real objects and 2 layers of size 256 for synthetic objects, reflecting the image resolution of each object type. Real captures (2448×2048) contain markedly more pixels and fine-scale structure than the 512×512 synthetic renders, so larger MLPs are required to represent the additional detail.

Figure 1 shows qualitative results on two real objects, from left to right, with a raw mosaiced input image, 0° re-rendering, normals and unpolarised diffuse, specular and mixed radiance. Figure 4 shows qualitative results for synthetic data from the PANDORA dataset [DZV22], with comparisons to the corresponding ground truth. Corresponding quantitative results are provided in Table 2 where we achieve state-of-the-art performance on all metrics. Across every object, the polarisation-based methods in the table achieve the lowest surface-normal MAE, whereas appearance-only baselines such as NeuS fall behind despite competitive radiance scores. Table 3 reports our ablation study for each loss element. Notably, our method uses only one measurement per pixel as from a DoFP polarisation camera, whereas other methods assume perfect demosaicing (requiring $12\times$ more data) or unpolarised RGB ($3\times$ more data). In Figures 5 and 6 we show further qualitative results on real objects with comparisons to PANDORA [DZV22] and PMVIR [ZMO23]. In all our results, we draw attention to the high frequency detail recovered in both geometry (c.f. the gnome) and

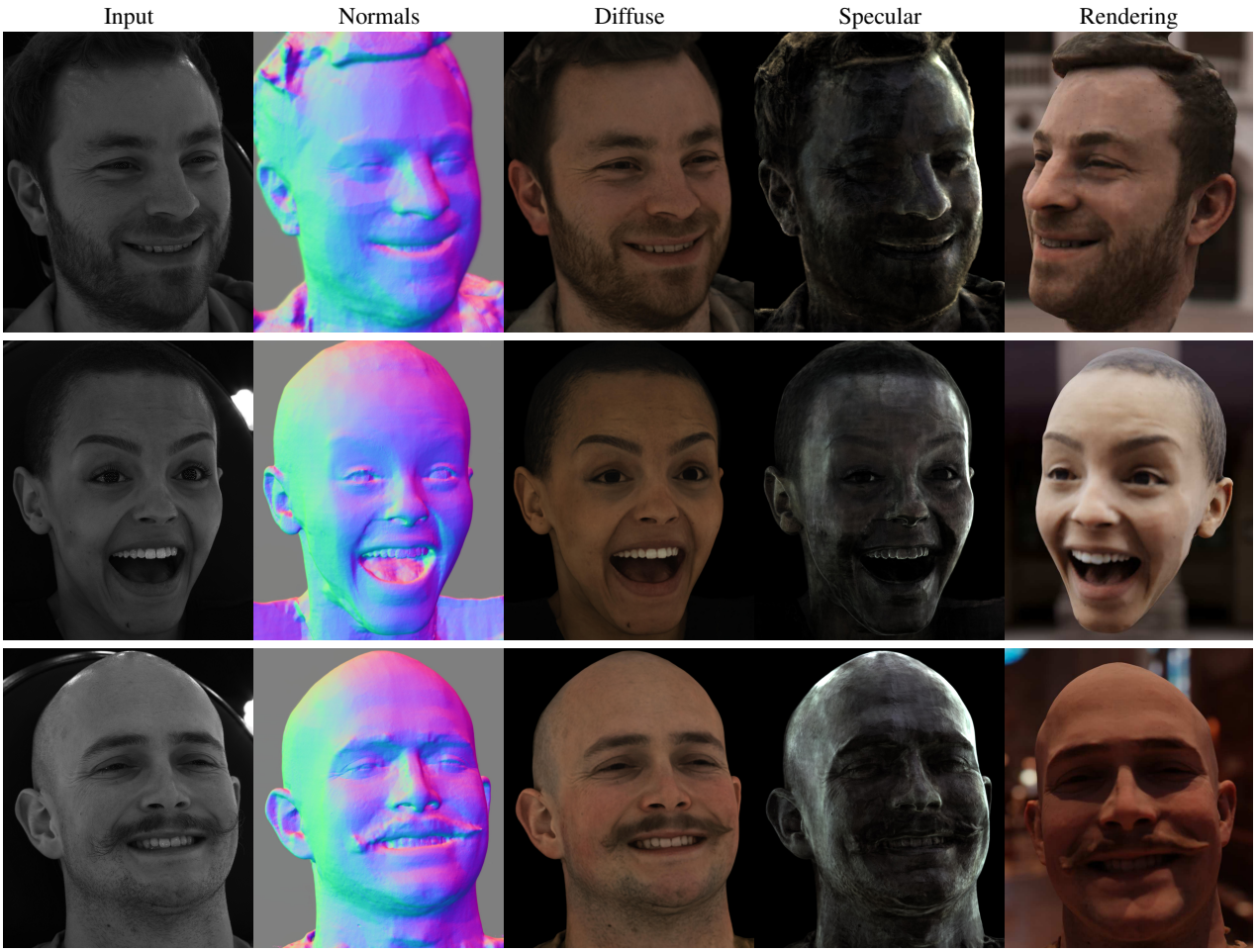

Figure 7: *Three facial subjects with different skin-tones. Columns: (1) raw CPFA view, (2) predicted surface normals, (3) diffuse radiance, (4) specular radiance, and (5) mesh re-rendered with the recovered material maps under a distinct HDR light probe for each subject [USC]. All images are tonemapped for display.*

diffuse and specular radiance maps (c.f. the lines of longitude and latitude - not recovered by PANDORA), the lack of texture transfer from diffuse to normals (c.f. the vase) and the fine HDR lighting detail recovered in the specular radiance (c.f. the car).

7.2. Under uniform illumination

For lightstage capture, we use the same CPFA cameras but this time from 15 viewpoints simultaneously providing a sparser input than for the static objects above. Uniform spherical illumination is provided by a geodesic dome comprising 160 nodes each supporting 9 LEDs set at full brightness. We extract meshes from the reconstructed SDF as follows. First, we construct an oriented point cloud by projecting all pixels by their expected termination depth [CL96] (filtering pixels with low accumulation values) and refining to ensure they lie on the zero level set. We evaluate the SDF gradient to determine normals. Second, we reconstruct a mesh using Poisson surface reconstruction [KBH06]. Finally, we transfer the diffuse and specular albedos to texture maps using xatlas.

Fig. 7 shows our results for three faces. Column 1 shows one raw CPFA view for each face. Columns 2 – 4 present the learned decomposition: surface normals, diffuse radiance, and specular radiance, respectively. In Column 5, we render the recovered meshes with their material maps under HDR light-probe illumination [USC] using Blender, demonstrating accurate geometry, albedo, and specular response. Figure 9 compares our reconstruction with that of a desktop-based facial-capture system [LLK*22]. Pose and grooming of the participant may vary due to different capture days. To make a fair comparison, we omit the photometric-normal refinement [LLK*22], and only render both meshes with their diffuse maps and geometric normals. Across the three images (raw mesh, textured mesh, and HDR-lit render), our method matches the competing setup in geometric details while requiring only a single-shot polarisation capture.

Benefiting from hash-grid [MESK22] and proposal network [BMV*22], our model converges rapidly within 30 minutes on NVIDIA A40 (compared to 15 hours of NeuS trained on NVIDIA

Scene	Method	Diffuse PSNR ↑ (dB)	Diffuse SSIM ↑	Specular PSNR ↑ (dB)	Specular SSIM ↑	Mixed PSNR ↑ (dB)	Mixed SSIM ↑	Normals MAE ↓ (°)
bust	NeuralPIL*	23.90	0.87	18.04	0.87	26.71	0.87	15.36
	PhySG*	22.64	0.94	23.00	0.94	19.94	0.72	9.81
	PANDORA†	25.82	0.81	22.96	0.75	22.79	0.79	3.91
	NeuS*	N/A	N/A	N/A	N/A	28.09	0.85	8.53
	P-NeRF	**37.59**	**0.999**	**32.01**	**0.983**	**32.72**	**0.962**	**0.4290**
globe	NeuralPIL*	13.09	0.55	12.92	0.55	20.04	0.66	38.73
	PhySG*	21.76	0.76	18.90	0.76	17.93	0.70	8.42
	PANDORA†	24.33	0.77	22.70	0.89	21.76	0.81	1.41
	NeuS*	N/A	N/A	N/A	N/A	23.57	0.81	3.72
	P-NeRF	**36.58**	**0.975**	**29.98**	**0.958**	**30.25**	**0.939**	**0.1144**

Table 2: *Quantitative evaluation on PANDORA [DZV22] synthetic image benchmark. * = method is given ground truth demosaiced RGB images. † = method is given ground truth demosaiced 12 channel CPFA images. No smoothness loss is applied to show performance without demosaicing benefits i.e. 12× more input.*

Scene	Loss Element	Diffuse PSNR ↑ (dB)	Diffuse SSIM ↑	Specular PSNR ↑ (dB)	Specular SSIM ↑	Mixed PSNR ↑ (dB)	Mixed SSIM ↑	Normals MAE ↓ (°)
bust	all	**37.95**	**0.999**	32.29	**0.984**	35.04	**0.964**	**0.3453**
	no smoothness loss	37.64	**0.999**	32.05	0.983	34.74	0.963	0.4322
	no theta loss	37.88	**0.999**	**32.39**	**0.984**	**35.14**	**0.964**	0.3460
	no specular loss	37.90	**0.999**	32.20	**0.984**	35.06	**0.964**	0.3562
	no HDR loss	29.20	0.974	24.80	0.972	24.90	0.925	2.2454
globe	all	**36.83**	**0.976**	**30.04**	**0.959**	**30.32**	**0.940**	0.1035
	no smoothness loss	36.64	0.975	30.03	0.958	30.27	0.939	0.1159
	no theta loss	36.61	0.975	30.00	0.958	30.23	0.938	0.1072
	no specular loss	36.74	**0.976**	29.99	**0.959**	30.25	0.939	0.1073
	no HDR loss	23.90	0.915	22.05	0.910	20.57	0.824	**0.04425**

Table 3: *Ablation study on PANDORA [DZV22] benchmark.*

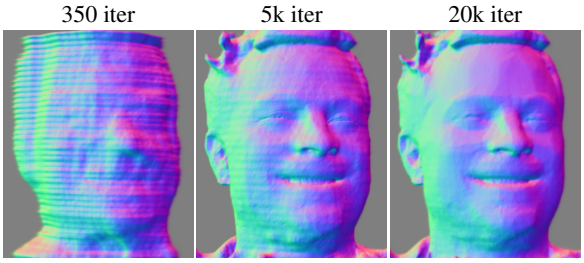

Figure 8: *Convergence: geometry after 350 iterations (∼2 minutes), 5k iterations (30 minutes) and 20k iterations (2 hours).*

RTX2080Ti). We show the convergence behavior in Fig. 8. A rough structure of the subject is formed within the first 350 iterations. The rendered normal map at 5k iterations is similar to that at 20k, implying the convergence point. To demonstrate the benefit of training on raw measurements, we conduct the ablation shown in Fig. 10. Demosaicing unintentionally mixes diffuse and specular radiances by spreading specularities into diffuse regions and vice versa. This leads to a specular artefact around the chin. Moreover, the model trained on raw measurement preserves high-frequency details which are blurred when compared to a counterpart trained on demosaiced images.

8. Conclusions

We have proposed P-NeRF, a multi-view shape-from-polarisation method based on neural representations. It maps measured polarisation to surface normals with a model that handles both specular and diffuse reflectance, without inverse rendering or an explicit lighting or BRDF model. This drives optimisation of the underlying neural SDF while the diffuse and specular radiance networks accurately separate the two sources of reflection. Many extensions are possible: polarisation could probe a wider range of materials, including ones poorly described by surface reflection, such as fabrics and transparent objects. A limitation and natural direction for future work would be to improve speed, for example by moving to a 3D Gaussian Splat representation [KKLD23].

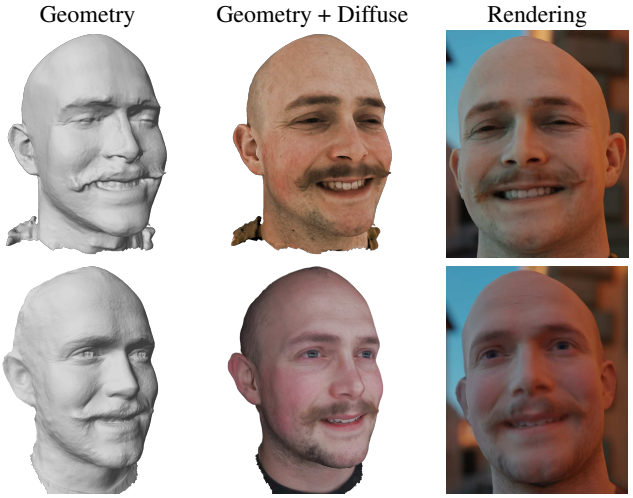

Figure 9: *Qualitative comparison of the same subject captured on two different days. Top row: our pipeline; bottom row: Lattas et al. [LLK*22] Column 1 shows the mesh geometry, Column 2 the mesh shaded with the estimated diffuse albedo, and Column 3 a Blender render under the Pisa Courtyard HDR probe [USC]. Both renders use geometric normals only; the photometric normals of [LLK*22] are disabled. Despite slight pose and appearance changes, the two reconstructions exhibit comparable geometric fidelity.*

Figure 10: *Demosaicing ablation: On the left we show diffuse and specular radiance for our method. On the right we show an ablation where we first demosaic and then train our method on all channels. Zoom to see blurring artefacts.*

Acknowledgements

The authors would like to thank Pete Cooper, John Mowbray and Danny Roberts for their amazing work designing, constructing and controlling our lightstage. This work was partly supported by the BBSRC grant BB/X01312X/1.

References

[AH06] ATKINSON G., HANCOCK E.: Recovery of surface orientation from diffuse polarization. *Image Processing, IEEE Transactions on 15* (07 2006), 1653 – 1664.

[BGW*20] BA Y., GILBERT A., WANG F., YANG J., CHEN R., WANG Y., YAN L., SHI B., KADAMBI A.: Deep shape from polarization. In *European Conference on Computer Vision* (2020), Springer, pp. 554–571.

[BJB*21] BOSS M., JAMPANI V., BRAUN R., LIU C., BARRON J. T., LENSCH H. P.: Neural-pil: Neural pre-integrated lighting for reflectance decomposition. In *Advances in Neural Information Processing Systems (NeurIPS)* (2021).

[BMV*22] BARRON J. T., MILDENHALL B., VERBIN D., SRINIVASAN P. P., HEDMAN P.: Mip-nerf 360: Unbounded anti-aliased neural radiance fields. *CVPR* (2022).

[CGS*17] CUI Z., GU J., SHI B., TAN P., KAUTZ J.: Polarimetric multi-view stereo. In *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (2017), pp. 369–378.

[CL96] CURLESS B., LEVOY M.: A volumetric method for building complex models from range images. In *Proceedings of the 23rd annual conference on Computer graphics and interactive techniques* (1996), pp. 303–312.

[CSOM23] CAO X., SANTO H., OKURA F., MATSUSHITA Y.: Multi-view azimuth stereo via tangent space consistency. In *Proc. CVPR* (2023).

[DHGH*22] DAVE A., HOLD-GEOFFROY Y., HAŠAN M., SUNKAVALLI K., VEERARAGHAVAN A.: Snapshot polarimetric diffuse-specular separation. *Opt. Express 30*, 19 (Sep 2022), 34239–34255.

[DZV22] DAVE A., ZHAO Y., VEERARAGHAVAN A.: Pandora: Polarization-aided neural decomposition of radiance. *arXiv preprint arXiv:2203.13458* (2022).

[GFT*11] GHOSH A., FYFFE G., TUNWATTANAPONG B., BUSCH J., YU X., DEBEVEC P.: Multiview face capture using polarized spherical gradient illumination. In *Proceedings of the 2011 SIGGRAPH Asia Conference* (2011), pp. 1–10.

[GGH*17] GUARNERA G. C., GHOSH A., HALL I., GLENCROSS M., GUARNERA D.: Material capture and representation with applications in virtual reality. In *ACM SIGGRAPH 2017 Courses* (New York, NY, USA, 2017), SIGGRAPH '17, Association for Computing Machinery.

[GPDG12] GUARNERA G. C., PEERS P., DEBEVEC P., GHOSH A.: Estimating surface normals from spherical stokes reflectance fields. In *Computer Vision – ECCV 2012. Workshops and Demonstrations* (2012), Springer Berlin Heidelberg, pp. 340–349.

[HGF*24] HAN Y., GUO H., FUKAI K., SANTO H., SHI B., OKURA F., MA Z., JIA Y.: NeRSP: Neural 3d reconstruction for reflective objects with sparse polarized images. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (2024), pp. 11821–11830.

[IMTI08] IKEUCHI K., MIYAZAKI D., TAN R. T., IKEUCHI K.: Separating reflection components of textured surfaces using a single image. *Digitally Archiving Cultural Objects* (2008), 353–384.

[KBH06] KAZHDAN M., BOLITHO M., HOPPE H.: Poisson surface reconstruction. In *Proceedings of the fourth Eurographics symposium on Geometry processing* (2006), vol. 7.

[KJCB23] KIM Y., JIN W., CHO S., BAEK S.-H.: Neural spectro-polarimetric fields. In *SIGGRAPH Asia 2023 Conference Papers* (2023), pp. 1–11.

[KJJ*21] KELLNHOFER P., JEBE L., JONES A., SPICER R., PULLI K., WETZSTEIN G.: Neural lumigraph rendering. In *CVPR* (2021).

[KKLD23] KERBL B., KOPANAS G., LEIMKÜHLER T., DRETTAKIS G.: 3d gaussian splatting for real-time radiance field rendering. *ACM Transactions on Graphics 42*, 4 (July 2023). URL: `https://repo-sam.inria.fr/fungraph/3d-gaussian-splatting/`.

[KTSR15] KADAMBI A., TAAMAZYAN V., SHI B., RASKAR R.: Polarized 3d: High-quality depth sensing with polarization cues. In *2015 IEEE International Conference on Computer Vision (ICCV)* (2015), pp. 3370–3378.

[LLK*22] LATTAS A., LIN Y., KANNAN J., OZTURK E., FILIPI L., GUARNERA G. C., CHAWLA G., GHOSH A.: Practical and scalable desktop-based high-quality facial capture. In *European Conference on Computer Vision* (2022), Springer, pp. 522–537.

[LOU*24] LI C., ONO T., UEMORI T., MIHARA H., GATTO A., NAGAHARA H., MORIUCHI Y.: Neisf: Neural incident stokes field for geometry and material estimation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)* (June 2024), pp. 21434–21445.

[LQX*22] LEI C., QI C., XIE J., FAN N., KOLTUN V., CHEN Q.: Shape from polarization for complex scenes in the wild. In *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)* (2022), pp. 12622–12631.

[LWLC24] LI Y., WU R., LI J., CHEN Y.-C.: Gnerp: Gaussian-guided neural reconstruction of reflective objects with noisy polarization priors. *ICLR* (2024).

[LWZG19] LATTAS A., WANG M., ZAFEIRIOU S., GHOSH A.: Multi-view facial capture using binary spherical gradient illumination. In *ACM SIGGRAPH 2019 Posters*. 2019, pp. 1–2.

[LZP*20] LIU S., ZHANG Y., PENG S., SHI B., POLLEFEYS M., CUI Z.: Dist: Rendering deep implicit signed distance function with differentiable sphere tracing. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition* (2020), pp. 2019–2028.

[MCZ*22] MINGQI S., CHONGKUN X., ZHENDONG Y., JUNNAN H., XUEQIAN W.: Transparent shape from single polarization images. *arXiv preprint arXiv:2204.06331* (2022).

[MESK22] MÜLLER T., EVANS A., SCHIED C., KELLER A.: Instant neural graphics primitives with a multiresolution hash encoding. *ACM Trans. Graph. 41*, 4 (July 2022), 102:1–102:15.

[MHMB*21] MILDENHALL B., HEDMAN P., MARTIN-BRUALLA R., SRINIVASAN P. P., BARRON J. T.: NeRF in the dark: High dynamic range view synthesis from noisy raw images. *arXiv* (2021).

[MHP*07] MA W.-C., HAWKINS T., PEERS P., CHABERT C.-F., WEISS M., DEBEVEC P.: Rapid acquisition of specular and diffuse normal maps from polarized spherical gradient illumination. pp. 183–194.

[MKI04] MIYAZAKI D., KAGESAWA M., IKEUCHI K.: Transparent surface modeling from a pair of polarization images. *Pattern Analysis and Machine Intelligence, IEEE Transactions on 26* (02 2004), 73–82.

[MLB18] MIHOUBI S., LAPRAY P.-J., BIGUÉ L.: Survey of demosaicking methods for polarization filter array images. *Sensors 18*, 11 (2018).

[MMSG05] MOREL O., MERIAUDEAU F., STOLZ C., GORRIA P.: Polarization imaging applied to 3d reconstruction of specular metallic surfaces. *Proceedings of SPIE - The International Society for Optical Engineering 5679* (02 2005).

[MS16] MANSON J., SLOAN P.-P.: Fast Filtering of Reflection Probes. *Computer Graphics Forum* (2016).

[MSB*12] MIYAZAKI D., SHIGETOMI T., BABA M., FURUKAWA R., HIURA S., ASADA N.: Polarization-based surface normal estimation of black specular objects from multiple viewpoints. In *2012 Second International Conference on 3D Imaging, Modeling, Processing, Visualization & Transmission* (2012), pp. 104–111.

[MSB*17] MIYAZAKI D., SHIGETOMI T., BABA M., FURUKAWA R., HIURA S., ASADA N.: Surface normal estimation of black specular objects from multiview polarization images. *Optical Engineering 56* (04 2017).

[MST*20] 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 *ECCV* (2020).

[MTHI03] MIYAZAKI, TAN, HARA, IKEUCHI: Polarization-based inverse rendering from a single view. In *Proceedings Ninth IEEE International Conference on Computer Vision* (2003), pp. 982–987 vol.2.

[MZBK06] MALLICK S. P., ZICKLER T., BELHUMEUR P., KRIEGMAN D.: Dichromatic separation: specularity removal and editing. In *ACM SIGGRAPH 2006 Sketches* (2006).

[NLG22] NOGUE E., LIN Y., GHOSH A.: Polarization-imaging Surface Reflectometry using Near-field Display. In *Eurographics Symposium on Rendering* (2022), Ghosh A., Wei L.-Y., (Eds.), The Eurographics Association.

[pol] Sony's polarized sensor: beyond conventional imaging. https://thinklucid.com/tech-briefs/polarization-explained-sony-polarized-sensor/. Accessed: 2022-11-11.

[RC01] RAHMANN S., CANTERAKIS N.: Reconstruction of specular surfaces using polarization imaging. In *Proc. CVPR* (2001).

[RRFG17] RIVIERE J., RESHETOUSKI I., FILIPI L., GHOSH A.: Polarization imaging reflectometry in the wild. *ACM Transactions on Graphics (TOG) 36* (2017), 1 – 14.

[SF16] SCHÖNBERGER J. L., FRAHM J.-M.: Structure-from-motion revisited. In *Conference on Computer Vision and Pattern Recognition (CVPR)* (2016).

[Sha92] SHAFER S. A.: *Using Color to Separate Reflection Components*. Jones and Bartlett Publishers, Inc., USA, 1992, p. 43–51.

[SMNO18] SOUZA A. C., MACEDO M. C., NASCIMENTO V. P., OLIVEIRA B. S.: Real-time high-quality specular highlight removal using efficient pixel clustering. In *2018 31st SIBGRAPI Conference on Graphics, Patterns and Images (SIBGRAPI)* (2018), pp. 56–63.

[SRT16] SMITH W., RAMAMOORTHI R., TOZZA S.: Linear depth estimation from an uncalibrated, monocular polarisation image. In *Proc. ECCV* (10 2016), vol. 9912, pp. 109–125.

[SZPF16] SCHÖNBERGER J. L., ZHENG E., POLLEFEYS M., FRAHM J.-M.: Pixelwise view selection for unstructured multi-view stereo. In *European Conference on Computer Vision (ECCV)* (2016).

[TJK*23] TAN Z., JI Y., KONG W., TAO X., FAN W., ZHAO J., XU X., LUO M.: Shape recovery from polarization: A review. In *Proceedings of the 15th International Conference on Digital Image Processing* (2023), ICDIP '23.

[TKR16] TAAMAZYAN V., KADAMBI A., RASKAR R.: Shape from mixed polarization. *Computer Optics 42* (05 2016).

[TPW*23] TIAN C., PAN W., WANG Z., MAO M., ZHANG G., BAO H., TAN P., CUI Z.: Dps-net: Deep polarimetric stereo depth estimation. In *Proceedings of the IEEE/CVF International Conference on Computer Vision* (2023), pp. 3569–3579.

[TWN*23] TANCIK M., WEBER E., NG E., LI R., YI B., KERR J., WANG T., KRISTOFFERSEN A., AUSTIN J., SALAHI K., AHUJA A., MCALLISTER D., KANAZAWA A.: Nerfstudio: A Modular Framework for Neural Radiance Field Development. *arXiv preprint arXiv:2302.04264* (2023).

[USC] USC INSTITUTE FOR CREATIVE TECHNOLOGIES: High-resolution light probe image gallery. https://vgl.ict.usc.edu/Data/HighResProbes/. Accessed 5 Jun 2025.

[VHM*22] VERBIN D., HEDMAN P., MILDENHALL B., ZICKLER T., BARRON J. T., SRINIVASAN P. P.: Ref-NeRF: Structured view-dependent appearance for neural radiance fields. *CVPR* (2022).

[WLL*21] WANG P., LIU L., LIU Y., THEOBALT C., KOMURA T., WANG W.: NeuS: Learning neural implicit surfaces by volume rendering for multi-view reconstruction. In *Advances in Neural Information Processing Systems (NeurIPS* (2021), vol. 34, pp. 27171–27183.

[XTS*22] XIE Y., TAKIKAWA T., SAITO S., LITANY O., YAN S., KHAN N., TOMBARI F., TOMPKIN J., SITZMANN V., SRIDHAR S.: Neural fields in visual computing and beyond. *Computer Graphics Forum* (2022).

[YCA*22] YU Z., CHEN A., ANTIC B., PENG S. P., BHATTACHARYYA A., NIEMEYER M., TANG S., SATTLER T., GEIGER A.: SDFStudio: A Unified Framework for Surface Reconstruction, 2022. URL: https://github.com/autonomousvision/sdfstudio.

[YKM*20] YARIV L., KASTEN Y., MORAN D., GALUN M., ATZMON M., RONEN B., LIPMAN Y.: Multiview neural surface reconstruction by disentangling geometry and appearance. *Advances in Neural Information Processing Systems 33* (2020).

[ZLW*21] ZHANG K., LUAN F., WANG Q., BALA K., SNAVELY N.: PhySG: Inverse rendering with spherical gaussians for physics-based material editing and relighting. In *The IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)* (2021).

[ZMO20] ZHAO J., MONNO Y., OKUTOMI M.: Polarimetric multi-view inverse rendering. In *Computer Vision – ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part XXIV* (2020), p. 85–102.

[ZMO23] ZHAO J., MONNO Y., OKUTOMI M.: Polarimetric multi-view inverse rendering. *IEEE Transactions on Pattern Analysis and Machine Intelligence 45*, 7 (2023), 8798–8812.

[ZOMO24] ZHAO J., OISHI J., MONNO Y., OKUTOMI M.: Polarimetric patchmatch multi-view stereo. In *2024 IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)* (2024), pp. 3464–3472. doi:10.1109/WACV57701.2024.00344.

[ZS19] ZHU D., SMITH W.: Depth from a polarisation + rgb stereo pair. In *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)* (06 2019), pp. 7578–7587.