Polarization Wavefront Lidar: Learning Large Scene Reconstruction from Polarized Wavefronts (Supplementary Information)

Dominik Scheuble[1,2*] Chenyang Lei [5*] Seung-Hwan Baek[4] Mario Bijelic[3,5] Felix Heide[3,5]
[1]Mercedes-Benz AG [2]TU Darmstadt [3]Torc Robotics [4]POSTECH [5]Princeton University

In this supplemental document, we present additional details on the PolLidar prototype, the simulation and reconstruction method, the testing and training datasets, and the training process. We also provide additional quantitative and qualitative results in support of the findings from the main manuscript.

Contents

1. PolLidar Prototype

As the PolLidar is an entirely novel prototype, we provide additional details about its construction (Sec. 1.1); subsequently, we share insight into key characteristics of the sensor (Sec. 1.2); then, we validate that the surface-induced polarization cues can be captured with the proposed prototype (Sec. 1.3); finally, we discuss improvements that will be made for future prototypes (Sec. 1.4).

[*]These authors contributed equally to this work.

1.1. Prototype Construction

The PolLidar extends the concept from Beamagines L3CAM lidar [1] starting from traditional ToF and adding polarization capabilities. However, the emitter and receiver are a redesign with custom opto-mechanics to fit waveplates and polarizer necessary to modulate the polarization. Fig. 1 illustrates a sectioned view of emitter and receiver to showcase the polarization optics. All used parts are listed in Tab. 1. The optics are designed to minimize the angle of incidence (AoI) onto the waveplates and polarizer. The maximum AoI on the emitter side is 2°, whereas on the receiving side, the maximum AoI is 8°. The optics on the emitting side allow for beam divergence of 0.364°. On the receiving side, an entrance pupil of 6.8 mm is realized. An overview of sensor key specifications is provided by Tab. 2.

The raw wavefront is captured with a National Instruments PCIe-5764 FlexRIO-Digitizer ADC, sampling at 1 Gs/s. The sampling frequency allows for 1 ns wide bins resulting in a 15cm range resolution. The ADC is interfaced from a LabView application which triggers the acquisition after the piezoelectric motors are finished rotating the waveplates or linear polarizer to their respective rotation angle θ_i. Raw wavefronts for a single frame with 150 rows and 236 columns amount to approximately 100 MB of raw binary data where two bytes per bin are used to encode the measured intensity. For the acquisition, a Python API interfacing the LabView application has been developed that can be leveraged in future works for e.g. online optimization as discussed in [6].

Item#	Part Description	Quantity	Model Name
1	Silicon-based Avalanche Photo Diode (APD)	1	Proprietary
2	Pulsed Laser source	1	Proprietary
3	Polarization-maintaining (PM) Fiber 10.5 µm	1	Coherent PM1060L
4	Emitter fiber Colimator	1	Thorlabs F230APC-1064
5	Emitter Lens 1	1	Edmund Optics 67-537
6	Emitter Lens 2	1	Edmund Optics 67-998
7	Emitter Lens 3	1	Edmund Optics 68-001
8	Receiver Lens 1	1	Edmund Optics 68-001
10	Receiver Optics	3	Proprietary
12	Band-pass Filter 1064nm	1	Proprietary
13	Quarter-wave Plate	2	Thorlabs WPQ10M-1064
14	Half-wave Plate	1	Thorlabs WPH10M-1064
15	Linear Polarizer	1	Throlabs LPNIRB100
16	Piezoelectric Motors	4	Thorlabs ELL14K
17	MEMS Micro Mirror Scanning System	1	Proprietary
18	Micro Mirror Receiving System	1	Proprietary

Table 1. List of used parts for receiver and emitter.

Specification	Value
Vertical resolution	150 rows over 23.95°
Horizontal resolution	236 columns over 31.05°
Temporal resolution	1488 bins over 223 m
Rotation angle θ_i resolution	0.01 °
Wavelength	1064 nm
Beam divergence	0.326°
Max. AoI on polarizing optics	2°(emitter), 8°(receiver)

Table 2. Key specification of constructed PolLidar prototype.

1.2. System Response and Noise Characteristics

The polarization cues necessary for Shape-from-Polarization (SfP) approaches are reflected in the intensity readings of the ADC. Depending on the degree-of-polarization (DoP), the intensity change for different polarization states can be, as function

[1] https://beamagine.com/product.

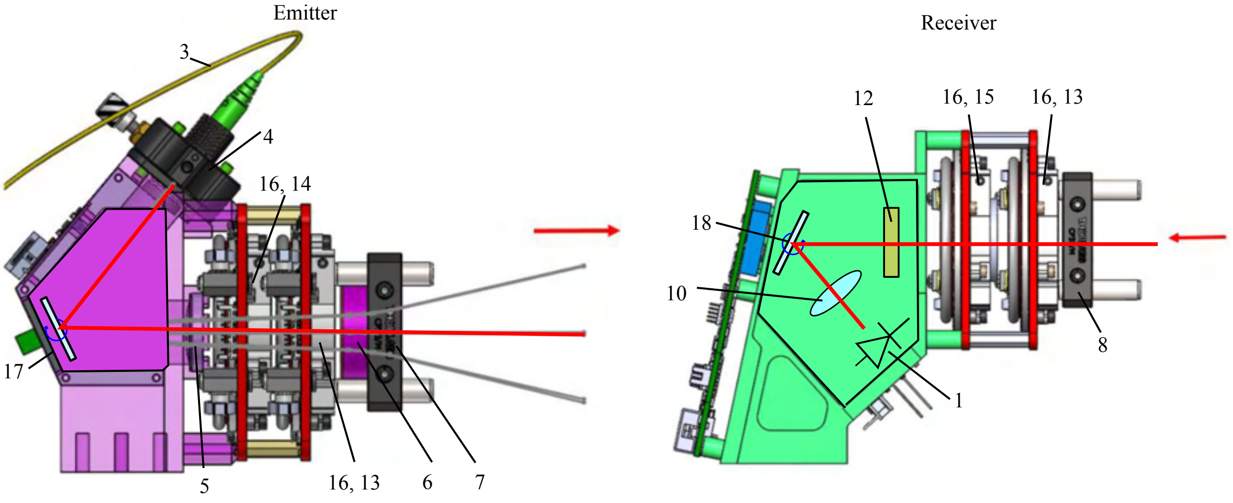


Figure 1. Sectioned view of emitter (left) and receiver (right). Parts numbers can be referenced from Tab. 1. For the protection of intellectual property, parts of the sectioned view are covered and replaced by schematic placeholders.

of geometry, material and viewing direction, very subtle. Thus, accurate intensity readings are crucial for a successful reconstruction. Hence, we provide inside in the system response of the PolLidar towards the key parameters of laser power and bias voltage of the APD. Furthermore, we investigate the noise characteristics of the sensor to better distinguish polarization-induced intensity change from sensor noise.

System Response To obtain meaningful intensity readings, laser power and bias need to be adjusted according to the scene. In contrast to conventional lidar sensors, we allowed access to these low-level parameters. We investigate the system response to laser power and bias in a controlled environment as shown in Fig. 2. Here, we repeatedly measure a single pixel on a calibrated 90% reflection target for different biases and laser powers. We average 50 frames per laser power/bias to limit the effect of noise. We find that large bias values quickly lead to saturation at around 0.4V. This will render the intensities meaningless for SfP as all polarization states will likely be saturated making reconstruction unfeasible. In contrast, small biases result in low intensity readings below the noise floor and subsequently dropped points preventing reconstruction altogether. As indicated by Fig. 2, we observe an exponential increase in intensity depending on the bias. Thus, the bias operating point needs to be selected with care to allow for successful reconstruction. On the other hand, we find that the laser power is well behaved in this regard and find an approximately linear relationship between laser power and intensity. Motivated by the findings in Fig. 2, we opt to select one laser power per captured frame and but always collect frames with 3 different biases. Details on acquisition are provided in Sec. 3.2.

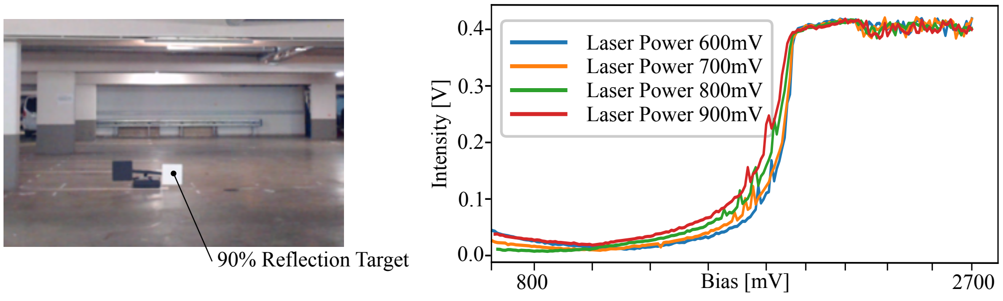


Figure 2. System response towards bias and laser power. We observe an approximately exponential relationship between bias and intensity but only a linear one between laser power and intensity for a single pixel on a 90% reflection target. We average 50 frames per laser power, bias pair to reduce the effect of noise.

Noise Model Furthermore, we investigate the noise characteristics of the sensor. To this end, we continuously capture the same scene 800 times as shown by Fig. 3. On the right of Fig. 3, we show intensity distributions for four selected pixels on interesting targets. We observe that the noise can be described as approximately Gaussian centered around a single mean. Thereby, we ensure to operate the laser diode and detector with sufficient warm up, such that over the measurements no temperature drift is possible. However, there is a substantial deviation of the intensity with standard deviations up to 0.03 mV which is coherent with the specifications of the installed APD. Therefore for some regions of a captured frame, the intensity deviation from noise will likely overshadow the desirable deviation observable for different polarization states. As a result, we opt to average 10 frames to increase resilience from sensor noise, as further discussed in Sec. 3.2. In addition, we test the sensor for repeatability by returning to the same rotation angles θ_i after setting a series of different rotation angles. We observe no drift in this regard.

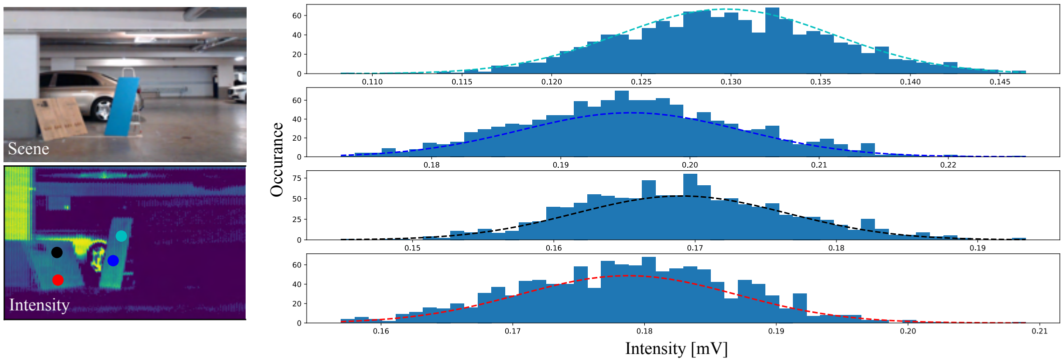


Figure 3. Noise characteristics of the PolLidar. We repeatedly measure the same scene 800 times. On the right, we show intensity distributions for selected pixels. The intensities are centered around a mean indicating no e.g. temperature-dependent drift. However, we observe substantial deviations around the mean, indicating that multiple wavefronts must be averaged to reduce the sensor noise.

1.3. Assessment of Polarization Cues

Our proposed SfP reconstruction method relies on rotating ellipsometry for acquisition. However, the results from rotating ellipsometry are not directly interpretable and do not allow an intuitive assessment weather the PolLidar is able to capture scene-dependent polarization cues. To this end, we follow the experiments from [1]. We fix the rotation angles of HWP and QWP θ^1 and θ^2 to $\theta^1 = \theta^2 = 0$. When assuming the scene to be solely diffusive, the azimuth angle of the surface normal can be directly found using the LP. Intuitively, the observed intensity I after the LP will have maxima where the rotation angle θ^4 of the LP and the angle of polarization of the reflected light align. In theory, the intensity varies sinusoidally with respect to θ^4. The phase-shift ϕ or position of the maxima translates directly to the azimuth angle of the surface normal up to an ambiguity of 180°. The measured intensity I at the APD is given as

$$I(\phi, \theta^4) = \frac{I_{\max} + I_{\min}}{2} + \frac{I_{\max} - I_{\min}}{2} \cos\left(2\theta^4 - 2\phi\right), \tag{1}$$

where $I_{\max}$ and $I_{\min}$ are the maximal and minimal observed intensities, respectively. Note that for our setup, the rotation angle θ^3 must be set to $\theta^3 = \theta^4$ to eliminate the effect of the QWP in the receiver.

We use this acquisition approach as an intuitive assessment of the PolLidar. To this end, we setup scenes with flat targets placed with different azimuth angles in front of the sensor. An example is shown in Fig. 4, where the object in focus is the blue metal plate oriented to the right and left, respectively. Next, we increase the rotation angle θ^4 of the LP in steps of 2°and average 10 wavefronts per rotation angle to limit the effect of sensor noise. As shown on the right of Fig. 4, we observe that the intensity for both scenes can be described as a sinusoid. Furthermore, the phase-shift between left- and right-oriented target is clearly visible and is close to the difference in the azimuth angle of the two metal plates. As such, we find that PolLidar is able to scan a scene accurately enough to capture polarization cues.

1.4. Real-Time Capability

By design, the PolLidar prototype is not real-time capable. To allow the largest possible flexibility we opted for finely controllable orientations of the wave plates, allowing us to adjust the polarization states instead and capturing all possible

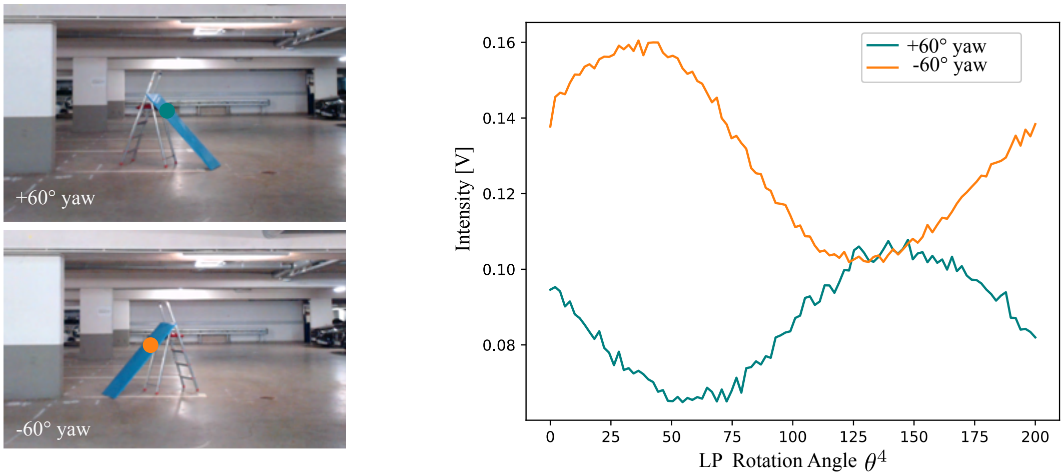


Figure 4. Assessment of Polarization Cues. We increase the rotation angle of the LP θ^4 in steps of 2°and observe a single pixel on the right- and leftwards oriented blue metal plate. We find that the intensity varies sinusoidally with respect to θ^4 and we observe a phase shift depending on the normal, adhering well with the theory and experiments from [1].

combinations of emitted and captured polarization states. As a consequence setting the polarization state requires considerable time. However, we imagine running the measurement in parallel as in passive cameras which employ four constant states.[2]. In detail, this would require multiple linear polarizers at fixed rotation angles, which are placed in front of an array of APDs similar to a Bayer-Pattern for RGB cameras. In summary, this work contributes a first proof showcasing polarization lidar bridging previous application domains allowing required distances and resolutions for autonmous driving vehicles especially enhancing the geometric representation for high distant objects.
Future works may optimize the number of required rotation angles for real-time capable reconstruction in automotive scenarios.

2. Lidar Forward Model and Simulator

In the following, we formalize the components of the polarimetric lidar forward model in Section 2.1. Then, we discuss the necessary changes made to the CARLA simulator and lastly show how we tune noise characteristics.

2.1. Temporal-Polarimetric Lidar Forward Model

Following the formulation from Baek et al. [2], we provide an overview of the individual components of the lidar forward model in the following.

Linear Polarizer A linear polarizer at a rotation angle θ with respect to a reference axis has a Mueller matrix given as

$$\mathbf{L} = \frac{1}{2}\begin{bmatrix} 1 & \cos 2\theta & \sin 2\theta & 0 \\ \cos 2\theta & \cos^2 2\theta & \cos 2\theta \sin 2\theta & 0 \\ \sin 2\theta & \cos 2\theta \sin 2\theta & \sin^2 2\theta & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}. \tag{2}$$

[2] see for example the Sony IMX253MZR image, https://www.sony-semicon.com/files/62/flyer_industry/IMX250_264_253MZR_MYR_Flyer_en.pdf.

Wave Plate The Mueller matrix of a wave plate with retardance ϕ at angle θ to the horizontal is defined as

$$
\begin{aligned}
\mathbf{R} &= \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & \mathbf{R}_{11} & \mathbf{R}_{12} & \mathbf{R}_{13} \\ 0 & \mathbf{R}_{21} & \mathbf{R}_{22} & \mathbf{R}_{23} \\ 0 & \mathbf{R}_{31} & \mathbf{R}_{32} & \mathbf{R}_{33} \end{bmatrix}, \\
\mathbf{R}_{11} &= \cos^2 2\theta + \sin^2 2\theta \cos\phi, \\
\mathbf{R}_{12} &= \sin 2\theta \cos 2\theta (1 - \cos\phi), \\
\mathbf{R}_{13} &= \sin 2\theta \sin\phi, \\
\mathbf{R}_{21} &= \sin 2\theta \cos 2\theta (1 - \cos\phi), \\
\mathbf{R}_{22} &= \cos^2 2\theta \cos\phi + \sin^2 2\theta, \\
\mathbf{R}_{23} &= -\cos 2\theta \sin\phi, \\
\mathbf{R}_{31} &= -\sin 2\theta \sin\phi, \\
\mathbf{R}_{32} &= \cos 2\theta \sin\phi, \\
\mathbf{R}_{33} &= \cos\phi.
\end{aligned} \tag{3}
$$

Ideal half/quarter-wave plates have a retardance values of π and $\pi/2$, respectively, resulting in the Mueller matrices

$$\mathbf{W}(\theta) = \mathbf{R}(\theta, \phi = \pi), \tag{4}$$

$$\mathbf{Q}(\theta) = \mathbf{R}(\theta, \phi = \pi/2). \tag{5}$$

Surface Fresnel Reflection and Transmission The transmitted and reflected components are represented as the Fresnel Mueller matrices

$$
\mathbf{F}_{T,R} = \begin{bmatrix} \frac{\mathrm{F}^{\perp}+\mathrm{F}^{\parallel}}{2} & \frac{\mathrm{F}^{\perp}-\mathrm{F}^{\parallel}}{2} & 0 & 0 \\ \frac{\mathrm{F}^{\perp}-\mathrm{F}^{\parallel}}{2} & \frac{\mathrm{F}^{\perp}+\mathrm{F}^{\parallel}}{2} & 0 & 0 \\ 0 & 0 & \sqrt{\mathrm{F}^{\perp}\mathrm{F}^{\parallel}}\cos\delta & \sqrt{\mathrm{F}^{\perp}\mathrm{F}^{\parallel}}\sin\delta \\ 0 & 0 & -\sqrt{\mathrm{F}^{\perp}\mathrm{F}^{\parallel}}\sin\delta & \sqrt{\mathrm{F}^{\perp}\mathrm{F}^{\parallel}}\cos\delta \end{bmatrix}. \tag{6}
$$

We compute the perpendicular and parallel Fresnel coefficients $F^{\perp,\parallel}$ for reflection and transmission [4], respectively. δ is the phase shift that has the value of ϕ or 0 for the dielectric component.

Coordinate Conversion A coordinate-conversion Mueller matrix has the following form

$$
\mathbf{C} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & \cos 2\theta & \sin 2\theta & 0 \\ 0 & -\sin 2\theta & \cos 2\theta & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}, \tag{7}
$$

where θ is the rotation angle of the input x basis vector.

Depolarization Matrices The depolarization matrices for diffuse $\mathbf{D}^d$ and specular reflection $\mathbf{D}^s$ are diagonal matrices where the j-th diagonal entry is given by

$$\mathbf{D}_j^{\{d,s\}}(\tau) = |\mathbf{D}|^{\{d,s\}} \exp\left(\frac{(\tau - t_{\text{peak}}^2)}{2\sigma^2}\right). \tag{8}$$

As denoted in the main paper, $|\mathbf{D}^d|$ and $|\mathbf{D}^s|$ describe the amplitude of the depolarization matrices and t_{peak} is the temporal index of the wavefront peak. The parameter σ defines the pulse width which we tune such that the pulse width of downsampled synthetic rays resemble the pulse width of the real device.

Micro-facet Distribution We use the Smith Shading and masking term G from [7] and the GGX facet distribution term D from [11] defined as follows

$$\begin{aligned} D(\theta_h; m) &= \frac{m^2}{\pi \cos^4 \theta_h (m^2 + \tan^2 \theta_h)^2}, \\ G(\theta_i, \theta_o; m) &= \frac{2}{1 + \sqrt{1 + m^2 \tan^2 \theta_i}} \frac{2}{1 + \sqrt{1 + m^2 \tan^2 \theta_o}}, \end{aligned} \tag{9}$$

where θ_h is the half-way angle, θ_i and θ_o are the incident and outgoing angles. m is the surface roughness.

Lidar Forward Model With all individual components of the lidar forward model in hand, the Mueller matrix $\mathbf{M}$ of the scene is given as sum of specular and diffuse reflection as

$$\mathbf{M}\,(\tau) = \underbrace{\frac{D(\theta_h; m) G(\theta_i, \theta_o; m)}{4 \cos \theta_i \cos \theta_o} \mathbf{D}^s(\tau) \mathbf{F}_R}_{\text{specular}} + \underbrace{\mathbf{C}_{\mathrm{n} \to \omega} \mathbf{F}_T^o \mathbf{D}^d(\tau) \mathbf{F}_T^i \mathbf{C}_{\omega \to \mathrm{n}}}_{\text{diffuse}}, \tag{10}$$

where $\theta_h = \cos^{-1}(\mathbf{h} \cdot \mathbf{n})$, $\theta_i = \cos^{-1}(\mathbf{n} \cdot \boldsymbol{\omega})$, $\theta_o = \cos^{-1}(\mathbf{n} \cdot \boldsymbol{\omega})$, $\mathbf{n}$ is the surface normal and $\boldsymbol{\omega}$ is the viewing direction. After applying distance dependent attenuation and the cosine shading, the temporal-polarimetric Mueller matrix of the scene can be written as

$$\mathbf{H} = \frac{\mathbf{n} \cdot \omega}{d^2} \mathbf{M}. \tag{11}$$

The Mueller matrix of the emitter $\mathbf{P}_i$, consisting of HWP and QWP, is defined as

$$\mathbf{P}_i = \mathbf{Q}(\theta_i^2) \mathbf{W}(\theta_i^1). \tag{12}$$

The Mueller matrix of the receiver $\mathbf{A}_i$, consisting of QWP and LP, is defined as

$$\mathbf{A}_i = \mathbf{L}(\theta_i^4) \mathbf{Q}(\theta_i^3). \tag{13}$$

The laser outputs horizontally polarized light defined by the Stokes vector $\mathbf{s}_{\text{laser}} = [1, 1, 0, 0]^T$. Hence, the Stokes vector of the light received at the APD can be modeled as

$$\mathbf{s}_{\text{APD}} = \mathbf{P}_i(\theta_i) \mathbf{H} \mathbf{P}_i(\theta_i) \mathbf{s}_{\text{laser}} \tag{14}$$

where the first element of $\mathbf{s}_{\text{APD}}$ equals the intensity measured by the APD.

2.2. Implementation Details of Polarization CARLA Simulator

We implement the previously discussed polarimetric lidar forward in CARLA. To this end, we rely on the full-wavefront lidar simulator presented by [6]. We extract the necessary normals $\mathbf{n}$, index of refraction μ, diffuse and specular depolarization amplitude $|\mathbf{D}^d|$, $|\mathbf{D}^s|$ and roughness m from CARLA. Analogous to [6], we use custom material cameras for diffuse amplitude $|\mathbf{D}^d|$, specular amplitude $|\mathbf{D}^s|$, roughness m. When following the notation of [6], the diffuse amplitude equals d, the specular amplitude equals s and roughness translates to α.

However, the index of refraction μ is not assigned to materials in CARLA by default. To this end, we modify the CARLA ray-tracer to return a material ID for each hit point. When the ray-tracer returns a hit point, we query the face of the mesh at this hit point for the name of the assigned material. Worlds in CARLA have more than 5000 assigned materials. We cluster the materials based on their name and their respective parent material. In total, we define 10 clusters and assign an index of refraction to each cluster, which we subsequently look-up during rendering. This method is suitable for static objects like buildings, infrastructure, or vegetation. However, moving objects such as vehicles in CARLA are considered only with a simplified mesh to reduce the computational cost for the ray-tracing. This simplified mesh does not have any materials assigned to it as it was not intended for rendering but only for extracting distance information with the ray-tracer. We circumvent this problem by manually assigning materials to each face of a vehicle mesh as shown in Fig. 5.

Finally, we extend the ray-tracer to return the surface normal $\mathbf{n}$ for each hit point. The Unreal Engine underlying CARLA provides a straightforward interface for querying the mesh for normals at a hit point. For extracting the normal, we add an extra channel for the normals to the lidar model in the C++ code and adapt the auxiliary Python interface accordingly. As the Unreal Engine returns normals in world-coordinates, we transform the normals into the local sensor frame.

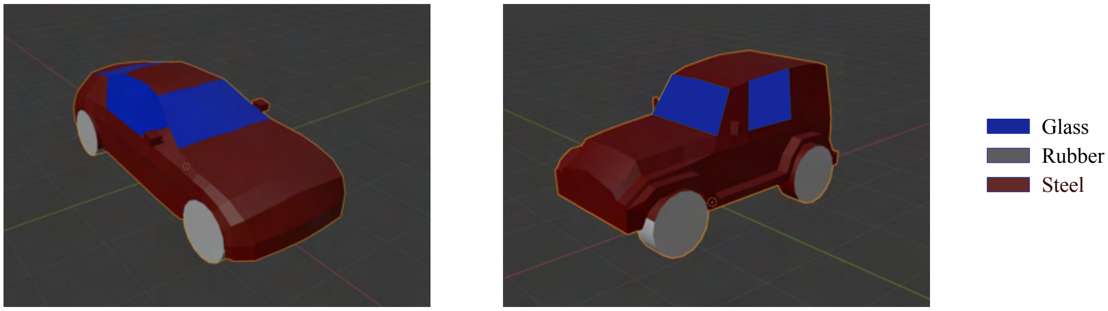


Figure 5. Assigned material to car meshes in CARLA. The ray-tracer in CARLA uses simplified meshes for cars to reduce rendering costs. These meshes do not have a material assigned to them. We manually assign 3 materials to each face of the car meshes, e.g., glass, steel and rubber.

2.3. Simulating Sensor Noise

The wavefronts $\tilde{\mathbf{I}}$ obtained from our Carla simulator are noise-free, and we corrupt the synthetic images by adding simulated sensor noise. The raw wavefronts are added by a noise $\eta_{\text{SENSOR}}(\tilde{\mathbf{I}})$, which consists of a Poissonian signal-dependent noise and a Gaussian signal-independent component

$$\eta_{\text{SENSOR}}(\tilde{\mathbf{I}}) = \eta_p(\tilde{\mathbf{I}}, a_p) + \eta_g(\sigma_g), \tag{15}$$

where η_p a Poissonian signal-dependent component, and η_g a Gaussian signal-independent component. To choose the correct parameters for a_p and σ_g, we rely on the noise characteristics presented in Sec. 1.2 and set a_p and σ_g to 1e-3 and 1e-4, respectively.

3. Additional Detail on Dataset

In the following, we provide additional details about ground truth generation and acquisition of the real-world large-scene polarimetric lidar dataset.

3.1. Ground Truth

As shown in Fig. 6, for generating ground truth distance and normals, we pair the PolLidar sensor with a Velodyne VLS-128 reference lidar. Both lidars are mounted on a movable platform allowing to scan a scene from different positions. After PolLidar acquisition is completed, we move the reference lidar through the scene to obtain multiple reference point clouds from different positions. We accumulate the reference point clouds after previous registration with the Iterative-Closest-Point (ICP) algorithm presented in [10] to obtain a dense accumulated lidar map.

Distance The lidar map is then used to generate ground truth distances $\mathbf{d}_{\text{gt}}$. To this end, we first transform the lidar map to PolLidar coordinates using the transformation $\mathbf{T} \in \mathbb{R}^{4x4}$. We obtain $\mathbf{T}$ by iteratively aligning reference and PolLidar point clouds for different scenes by means of the ICP algorithm as implemented by [12]. Next, we can extract the ground truth distance by calculating azimuth and elevation angle for each point of the lidar map and then applying bilinear interpolation with the view encoding $\mathbf{V}$ of the PolLidar. As we move the prototype through the scene, the lidar map might have several valid distances per viewing direction that are occluded from one viewpoint but visible from another. This would severely distort the ground truth when applying interpolation to the full lidar map. Thus, before interpolation, we remove hidden points that are invisible from the PolLidar viewpoint by means of [9]. Eventually, we output a ground truth distance map $\mathbf{d}_{\text{gt}} \in \mathbb{R}^{H \times W}$ encoding the ground truth distance for every viewing direction.

Normals For generating ground truth normals, we first mesh the lidar map using [8] and then query the mesh at the ground truth point locations for their normals. Compared to ray-tracing with the viewing direction, querying is beneficial as the meshing method often introduces artifacts around object edges enlarging the object. Ray-tracing would produce erroneous normals as the true object is possibly occluded by an enlarged edge. Finally, we output a ground truth normal map $\mathbf{n}_{\text{gt}} \in \mathbb{R}^{H \times W \times 3}$ encoding the ground truth normal for every viewing direction.

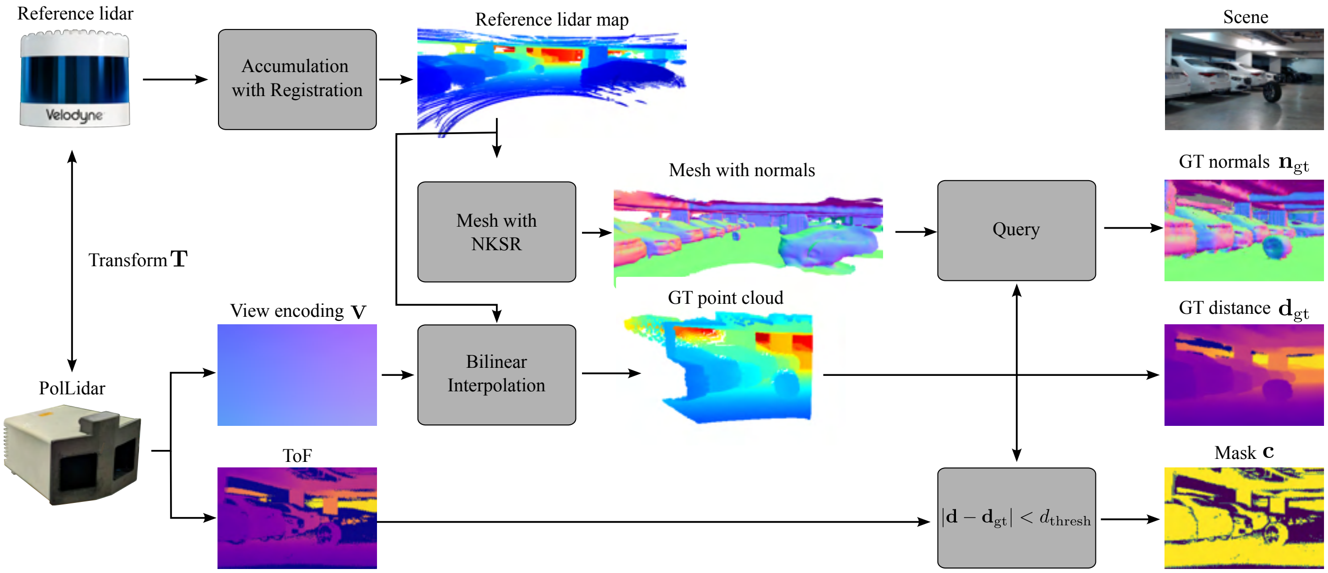

Reference lidar
Velodyne
Accumulation with Registration
Reference lidar map
Scene
Transform $\mathbf{T}$
Mesh with NKSR
Mesh with normals
Query
GT normals $\mathbf{n}_{\text{gt}}$
View encoding $\mathbf{V}$
GT point cloud
GT distance $\mathbf{d}_{\text{gt}}$
PolLidar
Bilinear Interpolation
ToF
Mask $\mathbf{c}$
$|\mathbf{d} - \mathbf{d}_{\text{gt}}| < d_{\text{thresh}}$

Figure 6. Acquisition of Ground Truth Data. We move PolLidar and reference lidar through a scene to capture dense reference lidar maps. We interpolate the lidar maps with the viewing direction of the PolLidar to extract GT distance $\mathbf{d}_{\text{gt}}$ information. For GT normals $\mathbf{n}_{\text{gt}}$, we first mesh the lidar map and extract normals from the mesh by querying with the GT point cloud.

Mask For certain pixels/view directions, no information can be extracted from the raw wavefront, when e.g. no object was hit, or the measured intensity falls below the noisefloor. We use the ToF map $\mathbf{d}$ extracted from the sensor with conventional peak-finding to exclude these points from training. To this end, we compare if the ToF map and ground truth distance are within a certain bound and define the mask of valid pixels / view directions as

$$\mathbf{c} = \begin{cases} 1 & \text{if } |\mathbf{d} - \mathbf{d}_{\text{GT}}| < d_{\text{thresh}} \\ 0 & \text{otherwise} \end{cases}, \tag{16}$$

where $\mathbf{c} \in \mathbb{R}^{H \times W}$ denotes the binary mask of valid view directions and d_{thresh} is a threshold we define as 0.8m. Furthermore, the mask $\mathbf{c}$ is helpful for eliminating erroneous ground truth. For instance, erroneous ground truth distance are likely to appear around object edges due to accumulation errors and resolution limits of the reference lidar, resulting in a widening of object edges. The mask excludes these points as shown in Fig. 6 for the silhouette of the car.

3.2. Acquisition

In the following, we provide further details on how we acquire frames with the PolLidar and describe the settings used. When acquiring a frame, we perform rotating ellipsometry as described in e.g. [4]. To this end, we measure 36 different rotation angle combinations θ_i, where the subscript $i \in \{0, 1, ..., 35\}$ is used to distinguish the different combinations. The resulting stokes vector of the emitted light is visualized in the Poincaré sphere in Fig. 7. As shown, the polarization state of the emitted light is uniformly distributed along the sphere. For the emitter, the HWP is set to $\theta_i^1 = 0$ and the QWP to $\theta_i^2 = 5° \cdot i$. For the receiver, the QWP is rotated to $\theta_i^3 = 25° \cdot i$ and the LP to $\theta_i^4 = 0$.

We opt to capture 10 frames for aggregation per rotation angle θ_i, as we find that it is a good compromise between acquisition time, amount of generated data, and denoising.

Furthermore, we choose a laser power of 600 mV and 900 mV for indoor and outdoor scenes respectively. Due to the high sensitivity of intensity towards the bias voltage, we capture the same scenario with the bias voltages $\{1980, 2000, 2020\}$ mV.

After acquisition with the PolLidar is complete, we capture a reference lidar map to extract ground truth as discussed in Sec. 3.1. To this end, we move the setup through the scenery until we cover a similar area visible from the PolLidar in the initial position.

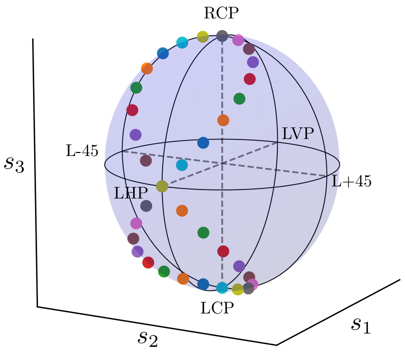


Figure 7. Poincaré sphere visualizing the Stokes vector of the 36 kinds of differently polarized emitted light.

4. Reconstruction

The proposed reconstruction approach is a two-step approach. First, we apply classical ellipsometric reconstruction to disentangle the scene from the polarizing optics of the emitter and receiver. From this, we obtain the Mueller matrix $\mathbf{H}$ of the scene. Additional details on this, are provided in Sec. 4.1. Next, we feed this as an input to a neural network that predicts distance offsets and normals. Additional details on the network are provided in Sec. 4.3.

4.1. Ellipsometric Reconstruction

Ellipsometric reconstruction is used to disentangle the Mueller matrix of the scene from the Mueller matrices of emitter $\mathbf{P}_i$ and receiver $\mathbf{A}_i$. As discussed in the paper, the additional ellipsometric reconstruction helps the network to learn a better scene reconstruction. In this section, we provide some additional ellipsometric reconstruction results.

In order to recover the Mueller matrix of the scene $\mathbf{H}_{\text{meas}}$, we solve a least-squares optimization problem as defined by

$$\underset{\mathbf{H}_{\text{meas}}}{\text{minimize}} \sum_{i=1}^{N} \left(I_i - [\mathbf{A}_i, \mathbf{H}_{\text{meas}}, \mathbf{P}_i \mathbf{s}_{\text{laser}}]_0\right)^2 , \tag{17}$$

see [2]. We visualize the reconstruction approach in Fig. 8, where the individual elements of the Mueller matrix $\mathbf{H}_{\text{meas}}$ are shown on the right. In order to validate the correctness of the reconstructed Mueller matrix, we then render the intensity image using the lidar forward model and compare it with the measured intensities for the 36 different rotation angles θ_i. This is visualized for the left of Fig. 4.1. On the top, we show the measured intensity for θ_0. On the bottom, we show for two selected pixels, the re-rendered / reconstructed intensities over the 36 different measurements using the polarimetric lidar forward model. We find that the reconstructed intensities are in good agreement with the measured ones. The small deviations are likely due to the inherently noisy measurements. Aside from the disentanglement from the polarization optics, we thus believe that the ellipsometric reconstruction provides additional benefit as an additional denoising step.

4.2. Network Details

We present the details of our network architecture in Tab. 3. Specifically, we first use two convolution layers to process the input features. We then use 4 encoder layers to encode the features, each layer consists of a max-pooling and two convolution layers. At the bottleneck, we use 8 transformer layers [5]. At last, we use 4 decoder layers with skip-connection to the prior layer. Finally, we use a 1×1 convolution layer to get a four-channel output, which is the normals and distance, respectively.

4.3. Training Details

We evaluate our method on both the simulation data and experimental data. Our simulated Carla dataset consists of 62 different scenes with different IDs. The contents of each scene are generally different. We select 44 scenes and 18 scenes for training and evaluation respectively, leading to 1430 and 539 test frames. We apply different laser biases during training. Specifically, we random sample the bias of our simulated PolLidar from 10 to 900. The intensities and distances for different biases are shown in Fig. 9.

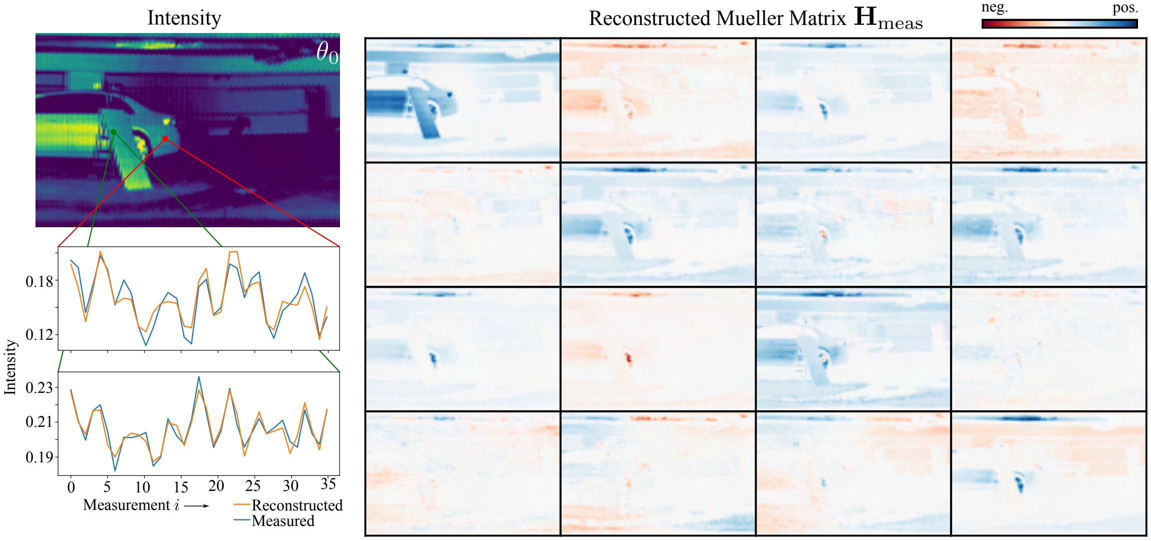


Figure 8. Ellipsometric Reconstruction. We show the individual elements of the reconstructed Mueller matrix on the right. To validate the correctness of the reconstruction, we render the intensities using the discussed polarimetric lidar forward model, as shown on the left. We find agreement between re-rendered / reconstructed and measured intensities. Note that different color scales are applied to each element for better visualization.

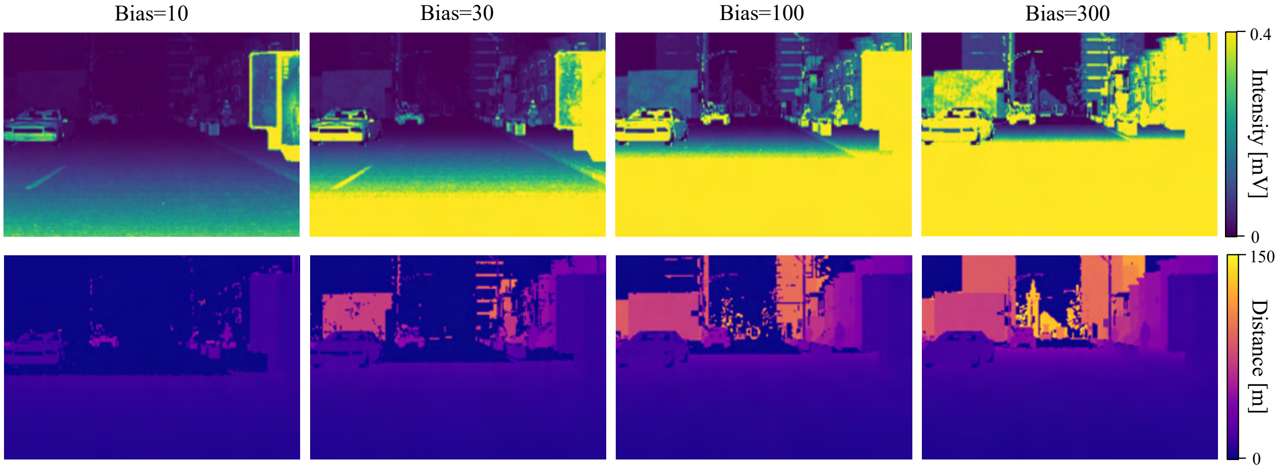


Figure 9. Frames with different biases. When the bias is low, a limited number of points are detected as the intensity falls below the noise floor. When the bias is large, we see saturation effects in regions close to the sensor or in regions of high reflectivity. We use different biases at training time to increase the robustness against saturation and low-intensity readings.

5. Additional Quantitative and Qualitative Results

In the following, we provide additional quantitative and qualitative results on real and synthetic data to validate the proposed method further.

5.1. Additional Synthetic Results

In Fig. 11, we provide additional synthetic results. Consistent with the main paper, we find that existing SfP methods underperform in low DoP regions. To further validate this, we show the DoP defined as

$$\text{DoP} = \frac{\sqrt{\mathbf{H}_{0,1}^2 + \mathbf{H}_{0,2}^2}}{\mathbf{H}_{0,0}}, \tag{18}$$

Name	Layer setting	Output dimension
input conv	Conv $[3 \times 3, 64]$, Instance Normalization, ReLU Conv $[3 \times 3, 64]$, Instance Normalization, ReLU	$H \times W \times 64$
encoder1	MaxPooling, stride 2 Conv $[3 \times 3, 128]$, Instance Normalization, ReLU Conv $[3 \times 3, 128]$, Instance Normalization, ReLU	$\frac{1}{2}H \times \frac{1}{2}W \times 128$
encoder2	MaxPooling, stride 2 Conv $[3 \times 3, 256]$, Instance Normalization, ReLU Conv $[3 \times 3, 256]$, Instance Normalization, ReLU	$\frac{1}{4}H \times \frac{1}{4}W \times 256$
encoder3	MaxPooling, stride 2 Conv $[3 \times 3, 512]$, Instance Normalization, ReLU Conv $[3 \times 3, 512]$, Instance Normalization, ReLU	$\frac{1}{8}H \times \frac{1}{8}W \times 512$
encoder4	MaxPooling, stride 2 Conv $[3 \times 3, 512]$, Instance Normalization, ReLU Conv $[3 \times 3, 512]$, Instance Normalization, ReLU	$\frac{1}{16}H \times \frac{1}{16}W \times 512$
attention	Transformer Block $\times 8$	$\frac{1}{16}H \times \frac{1}{16}W \times 512$
decoder4	Concat [encoder4 outputs, attention outputs] Bilinear Upsampling with scale 2 Conv $[3 \times 3, 512]$, Instance Normalization, ReLU Conv $[3 \times 3, 512]$, Instance Normalization, ReLU	$\frac{1}{8}H \times \frac{1}{8}W \times 512$
decoder3	Concat [encoder3 outputs, decoder4 outputs] Bilinear Upsampling with scale 2 Conv $[3 \times 3, 256]$, Instance Normalization, ReLU Conv $[3 \times 3, 256]$, Instance Normalization, ReLU	$\frac{1}{4}H \times \frac{1}{4}W \times 256$
decoder2	Concat [encoder2 outputs, decoder3 outputs] Bilinear Upsampling with scale 2 Conv $[3 \times 3, 128]$, Instance Normalization, ReLU Conv $[3 \times 3, 128]$, Instance Normalization, ReLU	$\frac{1}{2}H \times \frac{1}{2}W \times 128$
decoder1	Concat [encoder1 outputs, decoder2 outputs] Bilinear Upsampling with scale 2 Conv $[3 \times 3, 64]$, Instance Normalization, ReLU Conv $[3 \times 3, 64]$, Instance Normalization, ReLU	$H \times W \times 64$
output conv	Conv $[1 \times 1, 4]$	$H \times W \times 4$

Table 3. Details of our network architecture. “Concat” operation means that we concatenate two elements along the channel dimension.

where the subscript denotes the respective index of the Mueller matrix $\mathbf{H}$. We show the DoP for the two selected scenes of Fig. 5 in the main paper in Fig. 10. When normals and viewing direction are aligned, as visualized in the two right columns, the DoP is low. This is true for e.g. buildings and parts of the car that face the sensor. Contrarily, the side or hood of the car is for instance a high DoP region as viewing direction and normal are almost perpendicular. Consequently, Baek et al. [2] are unable to reconstruct normals in these low DoP regions, as shown in Fig. 11. For high DoP regions, however, satisfying performance is achieved.

We also compare the normal reconstruction to PCA as a point-cloud based method that considers a neighborhood of points. This method performs well in areas with flat geometry and high point density but degrades significantly at long ranges, e.g., cars in far distances, and geometry transition regions, e.g., the area between road and car. The proposed method leverages the additional cues from polarization to resolve normals in regions with sparse points and in transition regions. Furthermore, the proposed method achieves satisfying reconstruction results for regions with little polarization information by taking a local neighborhood into account.

5.2. Additional Real-World Results

We provide additional qualitative results from the real-world dataset in Fig. 12. We find similar trends as in the main paper. The proposed method consistently outperforms PCA in areas where the point cloud is sparse. This is visible in the zoom-ins on the fine structures, e.g., car roofs, where the neighborhood is too sparse for PCA to reconstruct correct surface normals. This is visible for the roof structure in the fourth row or the car on the left in the fifth row. The point cloud visualization also

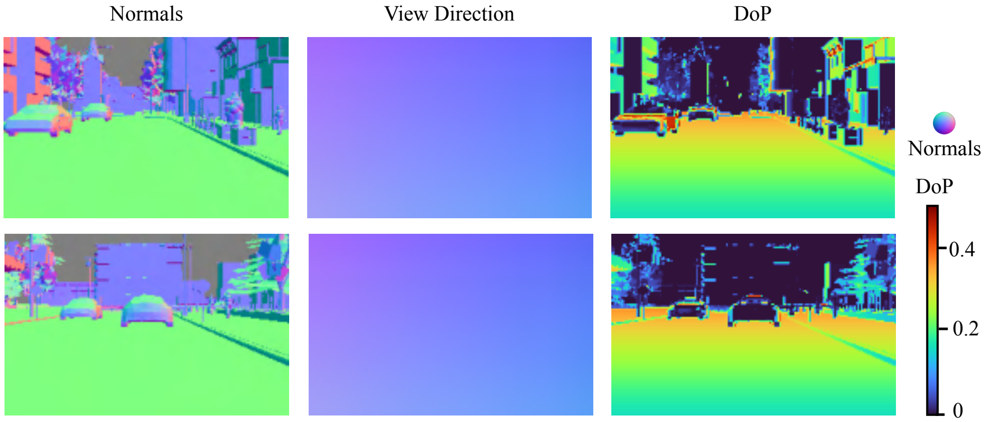


Figure 10. Degree-of-Polarization (DoP): On the right, we show the DoP for the two scenes of Fig. 5 in the main paper. When comparing with normals and viewing direction, we find low DoP regions on objects where normals and viewing direction align. This is problematic for other SfP reconstruction approaches as little polarization cues are present in these areas for a successful reconstruction.

reveals the benefit of the distance reconstruction compared to conventional argmax-methods in the context of reconstructing fine details. This is pronounced in the fourth row of Fig. 12 for the bumper/headlight area of the car on the left. The proposed method is able to reconstruct the pocket in the chassis where the headlight is located correctly. This detail is lost in the PCA results, as here, conventional peak-finding is applied. Furthermore, our method provides high-quality reconstruction results in the area around the grill of the car. Additionally, we show data in night-time conditions where the PolLidar expectedly shows similar reconstruction results as during day-time conditions.

5.3. Additional Distance-Binned Evaluation of Normal Reconstruction

To further analyze the advantage of our approach over the existing baseline in regions of low point density, we analyze the mean angular error for normal reconstruction and bin the results per distance bin. Specifically, this analysis allows us to study further distances in more detail, where point distance due to constant angular sampling increases further, making PCA fail due to the missing neighborhood. Quantitative results are visualized in Fig. 13, showing in (a) an increased performance due to polarization in close distance by approx. factor of two compared to PCA. Thereby, for further distances PCA degenerates by 75% being outperformed by a factor of 2.5. Furthermore, our method is able to cope with sparser point clouds and shows a substantially lower dependency of reconstruction performance on distance. In Fig. 13(b), we plot the relative mean angular error between PCA and the proposed method. We see that the gain in reconstruction quality is closely related to the distance.

5.4. Additional Metrics for Distance Evaluation

To evaluate the distance estimation, we compare against the conventional argmax-peak-finding typically performed directly on the device by low-level electronics [3]. We list all evaluated metrics on the distance reconstruction in Tab. 4 and Tab. 5. In addition to the results in the main paper, we present here median and RMSE distance errors. We find that our approach outperforms conventional peak-finding on these metrics by large margins on both synthetic and real data, outperforming previous results by 41% on synthetic data and 17% on real-world data for mean absolute distance error. The comparatively lower gain in real-world data is likely related to the quality of the ground truth. For synthetic data, we have perfect ground truth as we have control over the entire rendering pipeline including the underlying geometry. However, for the real-world data, we are limited by our applied geometry reconstruction described in Sec. 3.1. This has inherent sensor noise obfuscating distance and inaccuracies of the accumulation of our ground-truth though pose estimation errors, beam divergence broadening object dimensions, and many more. Consequently, the median error is increased for both conventional and proposed distance estimation methods. In contradiction to that is the higher RMSE in the synthetic data which can be explained with outliers in the distance error. More specifically, if the conventional peak-finding method selects the wrong peak or misses the peak altogether, e.g., due to low intensities close to the noise floor, the distance error will be of many meters for both the conventional and proposed method, thus increasing the RMSE significantly. For the real data, the effect of increased RMSE is not as prominent, as we apply the mask dependent on d_{thresh} effectively suppressing these outliers.

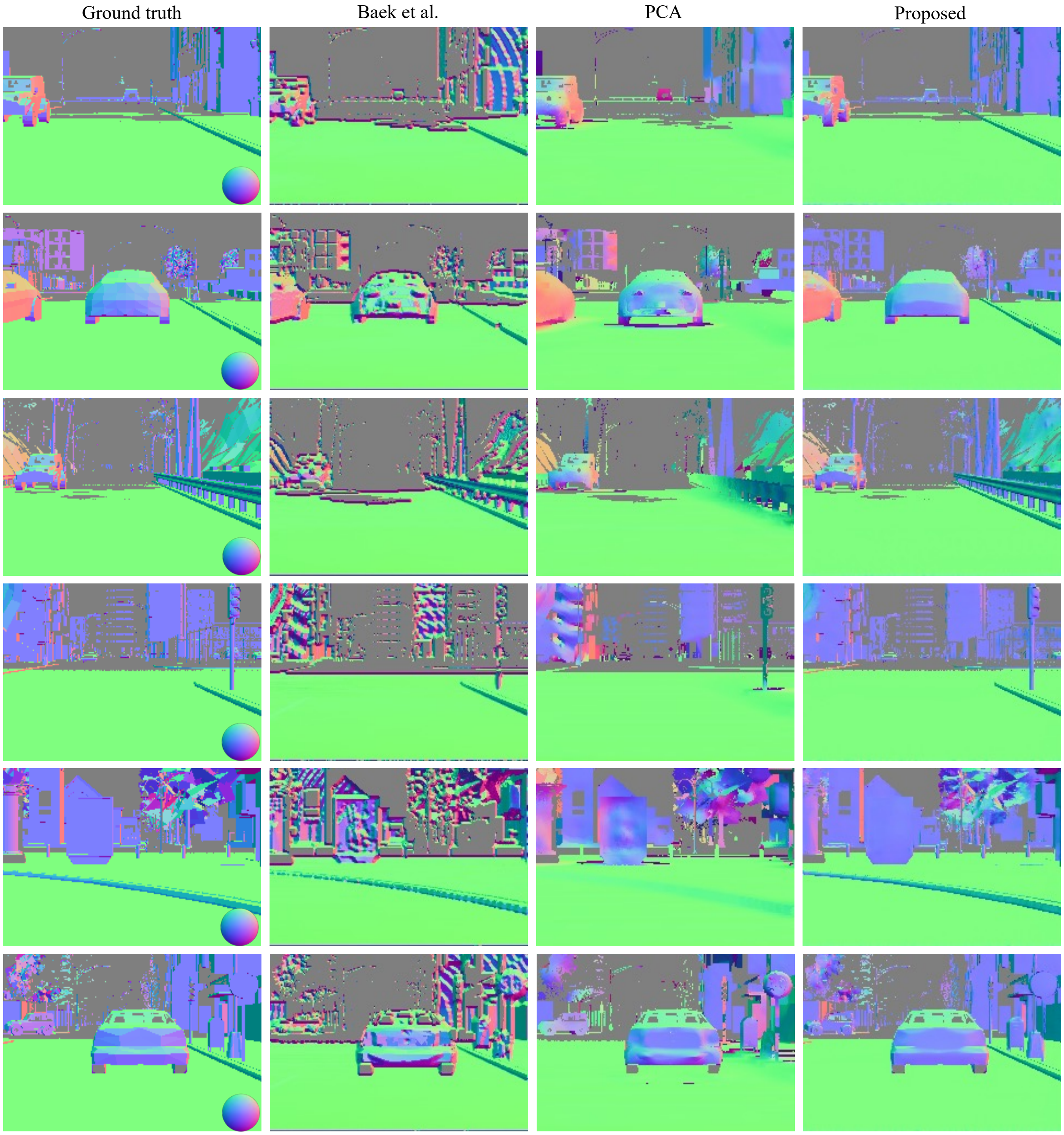


Figure 11. Additional Synthetic Results. Baek et al. [2] is unable to reconstruct normals in areas with low DoP, e.g., walls of buildings facing the sensor. PCA [13] applied in this setting is strongly dependent on point cloud density. This is visible for e.g. the poles in far distances. The proposed approach leverages polarization cues to reconstruct normals in sparse regions and is robust against low DoP areas.

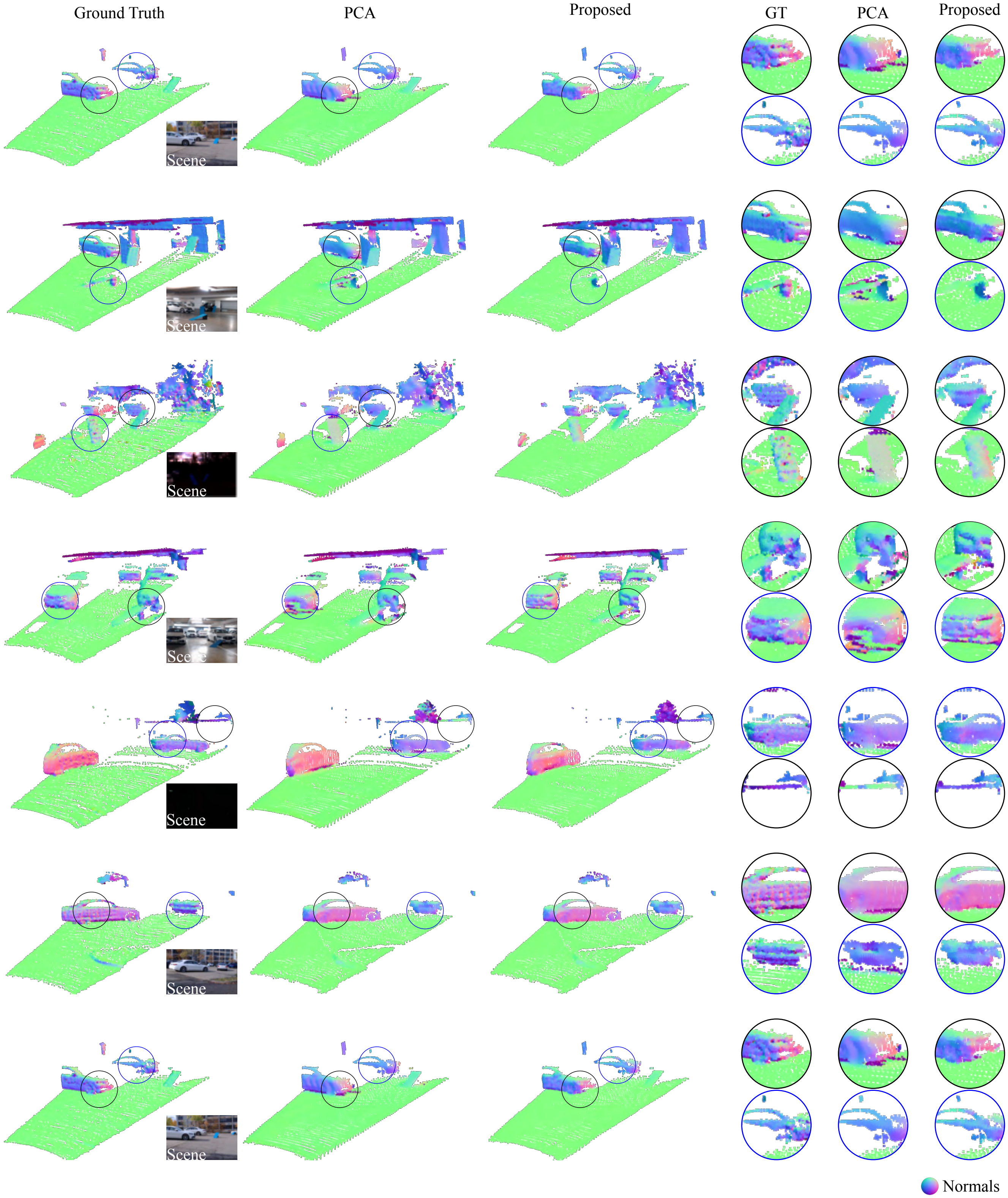
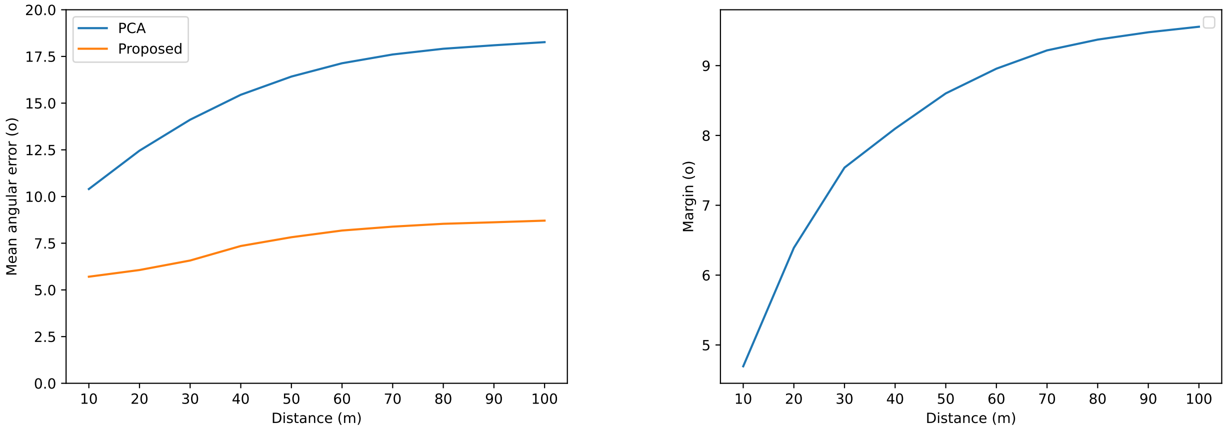


Figure 12. Additional Real-world Results. We show day- and night-time automotive scenes. PCA generates high-quality surfaces in areas with high point density. However, for fine details or objects far away, where the point cloud is sparse, the reconstruction is of low quality. The proposed method, however, leverages the polarization cues in these areas and thus outperforms PCA in regions of low-point density. This can be seen in the zoom-ins shown on the right, e.g. bumper area of the car in the fourth row or car roofs in the fifth and sixth rows.



(a) Absolute mean angular error

(b) Margin

Figure 13. Mean angular error dependent on distance. In (a), PCA achieves reasonable performance for close regions as the point clouds are more dense here. However, as the distance increases, the mean angular error of PCA increases rapidly. In (b), we plot of the margin between PCA and the proposed approach. We can see the performance gain is closely correlated with the distance.

Method	Distance Error [m] ↓		
	Mean	Median	RMSE
Conventional	0.32	0.12	3.25
Proposed	**0.19**	**0.03**	**3.06**

Table 4. Quantitative Distance Error on Synthetic Data. Our approach outperforms existing baselines on all evaluated distance metrics. Conventional distance estimation is limited by the temporal resolution of the sensor. Our approach leverages the wavefront information, thus outperforming the conventional distance estimation approach.

Method	Distance Error [m] ↓		
	Mean	Median	RMSE
Conventional	0.24	0.20	0.29
Proposed	**0.20**	**0.18**	**0.26**

Table 5. Quantitative Distance Error on Real-world Data. Due to noisier ground truth and sensor imperfections, the overall error is slightly larger compared to synthetic data, while confirming the trend from the synthetic evaluation. The proposed method relies on wavefront data and point neighborhoods, outperforming the conventional approach.

References

[1] Gary A. Atkinson and Edwin R. Hancock. Recovery of surface orientation from diffuse polarization. *IEEE Trans. Image Process.*, 15(6):1653–1664, 2006. 4, 5

[2] Seung-Hwan Baek and Felix Heide. All-photon polarimetric time-of-flight imaging. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2022. 5, 10, 12, 14

[3] Behnam Behroozpour, Phillip AM Sandborn, Ming C Wu, and Bernhard E Boser. Lidar system architectures and circuits. *IEEE Communications Magazine*, 55(10):135–142, 2017. 13

[4] Edward Collett. Field guide to polarization. Spie Bellingham, WA, 2005. 6, 9

[5] Alexey Dosovitskiy, Lucas Beyer, Alexander Kolesnikov, Dirk Weissenborn, Xiaohua Zhai, Thomas Unterthiner, Mostafa Dehghani, Matthias Minderer, Georg Heigold, Sylvain Gelly, et al. An image is worth 16x16 words: Transformers for image recognition at scale. *arXiv preprint arXiv:2010.11929*, 2020. 10

[6] Felix Goudreault, Dominik Scheuble, Mario Bijelic, Nicolas Robidoux, and Felix Heide. Lidar-in-the-loop hyperparameter optimization. 2023. 2, 7

[7] Eric Heitz. Understanding the masking-shadowing function in microfacet-based brdfs. *Journal of Computer Graphics Techniques*, 3(2):32–91, 2014. 7

[8] Jiahui Huang, Zan Gojcic, Matan Atzmon, Or Litany, Sanja Fidler, and Francis Williams. Neural kernel surface reconstruction. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 4369–4379, 2023. 8

[9] Sagi Katz, Ayellet Tal, and Ronen Basri. Direct visibility of point sets. In *ACM SIGGRAPH 2007 papers*, pages 24–es. 2007. 8

[10] Ignacio Vizzo, Tiziano Guadagnino, Benedikt Mersch, Louis Wiesmann, Jens Behley, and Cyrill Stachniss. Kiss-icp: In defense of point-to-point icp–simple, accurate, and robust registration if done the right way. *IEEE Robotics and Automation Letters*, 8(2):1029–1036, 2023. 8

[11] Bruce Walter, Stephen R Marschner, Hongsong Li, and Kenneth E Torrance. Microfacet models for refraction through rough surfaces. In *Proceedings of the 18th Eurographics conference on Rendering Techniques*, pages 195–206, 2007. 7

[12] Qian-Yi Zhou, Jaesik Park, and Vladlen Koltun. Open3D: A modern library for 3D data processing. *CoRR*, abs/1801.09847, 2018. 8

[13] Yufan Zhu, Weisheng Dong, Leida Li, Jinjian Wu, Xin Li, and Guangming Shi. Robust depth completion with uncertainty-driven loss functions. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 36, pages 3626–3634, 2022. 14