Deep Polarization Imaging for 3D Shape and SVBRDF Acquisition

Valentin Deschaintre
Imperial College London
v.deschaintre@ic.ac.uk

Yiming Lin
Imperial College London
yiming.lin11@ic.ac.uk

Abhijeet Ghosh
Imperial College London
abhijeet.ghosh@ic.ac.uk

Abstract

We present a novel method for efficient acquisition of shape and spatially varying reflectance of 3D objects using polarization cues. Unlike previous works that have exploited polarization to estimate material or object appearance under certain constraints (known shape or multiview acquisition), we lift such restrictions by coupling polarization imaging with deep learning to achieve high quality estimate of 3D object shape (surface normals and depth) and SVBRDF using single-view polarization imaging under frontal flash illumination. In addition to acquired polarization images, we provide our deep network with strong novel cues related to shape and reflectance, in the form of a normalized Stokes map and an estimate of diffuse color. We additionally describe modifications to network architecture and training loss which provide further qualitative improvements. We demonstrate our approach to achieve superior results compared to recent works employing deep learning in conjunction with flash illumination.

1. Introduction

In this work, we extend practical acquisition of shape and spatially varying reflectance of 3D objects. Accurately acquiring the shape and appearance of real-world objects and materials has been an active area of research in vision and graphics with a wide range of applications including analysis/recognition, and digitization for visual effects, games, virtual reality, cultural heritage, advertising or design for example. Advances in digital imaging over the last two decades has resulted in image-based acquisition techniques becoming an integral component of appearance modeling and 3D reconstruction. A recent trend has been towards making acquisition more practical, employing commodity off-the-shelf equipment and more recently relying on minimalistic capture coupled with advances in deep learning techniques.

Our method allows to better recover appearance by leveraging deep learning and additional photometric cues. Similar to the recent works that employ deep learning for 3D

object appearance capture [25, 7], we employ acquisition under frontal flash illumination. However, compared to previous work, we additionally leverage polarization imaging and exploit polarization cues in conjunction with deep learning.

The recent work of Ba et al. [4] proposed a similar approach to estimate surface normals of homogeneous 3D objects under uncontrolled environment illumination. We show that arbitrary environment illumination can limit the quality of polarization signal and describe how flash illumination solves this problem, resulting in high quality surface normal estimate. Furthermore, our method can process non homogeneous objects and estimates a physically based spatially-varying BRDF of higher quality than previous methods [25, 7].

We create a new HDR synthetic dataset simulating polarization behaviour on different geometries and SVBRDFs and train an improved deep network inspired by Deschaintre et al. and Li et al. [11, 25]. Observing the polarization information of a 3D object from a single view direction under frontal flash illumination, our method estimates the 3D shape as surface normal and depth maps and spatially varying reflectance properties as diffuse and specular albedo maps and specular roughness map which enable high quality renderings of acquired objects under novel lighting conditions.

To summarize, in this paper we make the following contributions:

- We propose the first method for joint 3D shape and SVBRDF estimation combining polarization cues (under flash illumination) and deep learning.

- We publish a new synthetic dataset with diffuse polarization effects for supervised learning.

- Analysis of polarization cues under different lighting and practical acquisition protocol for high quality results.

- An improved deep network architecture and training loss for 3D object appearance acquisition.

2. Related work

There exist a significant body of prior work on reflectance capture [42, 18], with a primary focus on accuracy of measurements and reduction of the time-complexity of the acquisition. Traditionally, reflectometry setups have been complex and suited only for laboratory-like settings. More recent work has however investigated practical acquisition techniques employing off-the-shelf equipment, as well as non-laboratory environments. In the following, we discuss these latter approaches and review related work on polarization imaging.

2.1. Practical appearance Acquisition

2.1.1 Commodity hardware

Advances in mobile technology have recently given rise to more compact and portable designs for reflectance measurements. Wu & Zhou [45] have proposed an integrated system for hand-held acquisition of shape and reflectance of objects with a Kinect sensor. Aittala et al. [2] have proposed a two-shot method for acquisition of stationary materials using a mobile phone. They employ a pair of flash-no flash observations (in general indoor environment) of the sample coupled with statistical analysis to extract reflectance maps. The method has been extended to a single flash image for stationary materials using neural synthesis [1]. Riviere et al. [34] proposed two mobile acquisition setups for acquisition of more general spatially varying planar surfaces. Free-form acquisition with flash illumination has also been employed for acquiring SVBRDFs of planar surfaces [20], and non-planar 3D objects [31]. However, these approaches rely on a large number of pictures of a sample and/or strong prior such as self-repetitive materials or planar geometry.

2.1.2 Exploiting deep learning

Several recent methods have been proposed for surface reflectance and shape estimation from sparse observations, including as few as a single observation by exploiting deep learning techniques. Many works have focused on SVBRDF estimation of planar samples under unknown environmental illumination [23, 46], or uncalibrated flash illumination [11, 24] or both [13], with further improvements using multiple flash measurements [12, 14]. Deep learning has also been employed to estimate reflectance properties of objects of unknown shape under unknown illumination [15, 27]. While powerful, these methods are limited to smooth convex shapes and estimate homogeneous reflectance but do not recover object shape. Closer to our work, deep learning has been recently employed for joint shape and spatially varying reflectance estimation of non-planar objects from observations under flash illumination [25] or combination of flash and ambient illumination [7]. Compared to these works, we demonstrate our approach to achieve higher quality results by combining deep learning with polarization cues.

2.2. Exploiting polarization

Polarization has been extensively studied in both vision and graphics, but mainly in strictly controlled environments where the polarization state of the incident light can be fine tuned by an operator. It has proved to be a useful channel of information for shape estimation, material classification and reflectance components separation. Most recently Baek et al. [6] conducted an extensive study on polarization and its effect on different BRDFs.
The vast majority of previous work have studied the polarization resulting from reflection under unpolarized or linearly polarized incident light, with two notable exceptions [22, 16] using circularly polarized illumination. Our method falls in this first category, using unpolarized flash illumination.

2.2.1 Reflectance separation/estimation

Appearance modeling methods strongly rely on the accurate separation of surface reflectance into its diffuse and specular components. Here, researchers have looked at polarization imaging, both exclusively [44, 30, 10, 26, 17] as well as in conjunction with color space methods [32, 41], for diffuse-specular separation.

Ghosh et al. [16] proposed measurement of the complete Stokes parameters of reflected circularly polarized spherical illumination to recover detailed reflectance parameters. The method assumes known or acquired shape information and controlled polarized illumination. Also related is the work of Miyazaki et al. [29], who employ linear polarization imaging under unpolarized illumination from multiple calibrated point light sources, coupled with inverse rendering in order to estimate shape, albedo and specular roughness of a homogeneous convex object. Baek et al. [5] take this further with detailed Muller matrix measurements under a co-incident point source illumination to jointly estimate shape and polarimetric SVBRDF of convex objects. Riviere et al. [35] employ linear polarization imaging in uncontrolled outdoor environments to estimate shape and spatially varying reflectance of planar surfaces. They however require acquisition from two near-orthogonal, oblique viewpoints, and further require known environmental illumination to solve for specular roughness. In this work, we aim at acquisition of 3D objects using polarization imaging while simplifying data capture complexity compared to prior works.

2.2.2 Surface normals estimation

Shape from polarization has been extensively studied in the vision literature. Two strategies are typically employed to infer orientation.
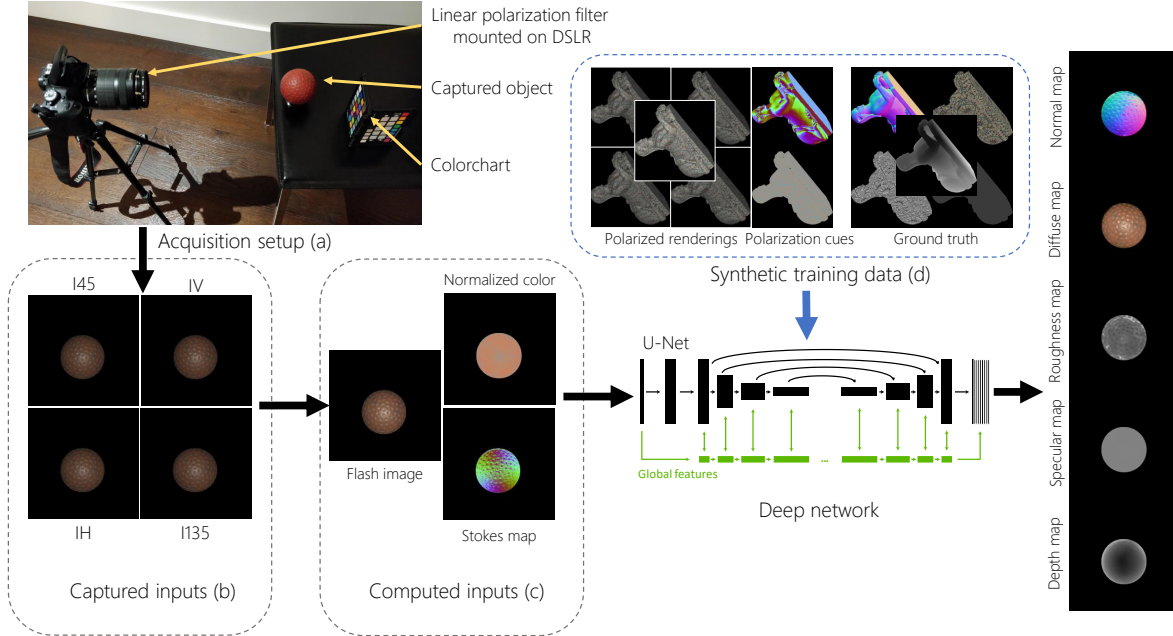
Figure 1. Pipeline of our method for estimation of object shape and SVBDRF from polarization cues. We use a linear polarization filter mounted on DSLR and a color chart (a) to capture the polarized images (b), from which we compute the remaining explicit cues (c). We use a deep network trained on synthetic data (d) generated with 20 randomly rotated complex meshes and 1200 SVBRDFs. Our method estimates the shape and SVBRDF of an object as normal, diffuse, specular, roughness and depth maps.

The first approach relies on the degree of polarization and inverting the Fresnel equations from a single view. Here, most prior work has focused on shape from specular reflection with the degree of polarization reaching an extremum at Brewster angle [40, 38, 19]. Atkinson and Hancock [3], however, measure the degree of polarization due to *diffuse* reflection for shape estimation. Kadambi et al. [21] proposed a method to enhance coarse depth maps by fusing shape from polarization cues with the output of a depth sensor. They employ the coarse 3D geometry to resolve the well known azimuthal ambiguity in polarization normals. Instead of relying on normals, Smith et al. [39] have proposed direct inference of surface depth by combining specular and diffuse polarization cues with a linear depth constraint formulation. They demonstrate depth recovery under uncalibrated (unpolarized) point source as well as low order spherical harmonic illumination. However these approaches rely on traditional optimization and a single view, limiting the quality of the recovered shapes in the absence of additional depth or shading cues. Our work aims to lift the restrictions of classical optimization to recover high quality 3D shape from a single viewpoint.

A second approach consists of observing the sample from multiple viewpoints [43, 33, 28, 37, 9, 47]. The key idea is then that one view constrains the surface normal to one plane and in theory only one additional view (and at most two [43]) are necessary to fully determine the normal to the surface.

The recent work of Ba et al. [4] leverages polarization imaging in conjunction with deep learning to estimate 3D shape (normals) of non-planar objects exhibiting homogeneous reflectance with single view acquisition under general environmental illumination. Besides the acquired polarization images, they provide their network with a physics based normals estimate (modulo azimuthal ambiguity) computed from the imaged polarization. They however do not recover spatially varying reflectance information and rely on environment illumination which we found to provide noisy polarization information –see Fig. 6– limiting acquisition accuracy. As opposed to these work our method relies on deep learning and flash polarization imaging from a single viewpoint to recover both high quality object shape *and* detailed spatially varying reflectance.

3. Overview

Our method aims at acquiring both the 3D shape (surface normals and depth) and spatially varying reflectance of an object using practical acquisition involving frontal flash illumination and single view acquisition. To tackle this highly ill-posed problem, we propose to leverage linear polarization cues in surface reflectance, providing strong initial cues to our deep network as shown in Fig. 1.

We employ off-the-shelf equipment for polarization measurements of an object: a DSLR camera fitted with a

linear polarizing filter on the lens, a tripod for stable mounting of the camera, and a color checker chart for white balacing and radiometric calibration of the observed reflectance.

While polarization imaging close to the Brewster angle allows extraction of many appearance cues directly [35], this can only be reliably done for planar surfaces. Hence, we use deep learning to compensate for the limitations of the polarization signal over the surface of a 3D object.

We train an improved U-Net, inspired from Deschaintre et al and Li et al. [11, 25], to employ polarization images of an object as input alongside some explicit cues provided by the polarization signal, and output five maps related to appearance and shape: diffuse and specular albedo, specular roughness, surface normal and depth. From the acquired polarization information, we compute two specific cues to provide as additional input to the deep network. The first is a shape cue in the form of a normalized *Stokes map*: it computes the normalized variation in the reflectance under different polarization filter orientations, providing a π ambiguous initialization for normals. The second is a reflectance cue in the form of normalized *diffuse color* computed by normalizing the reflectance minima obtained (through sinusoidal fitting) from the acquired polarized images.

To train our deep network, we create a synthetic dataset consisting of 20 complex 3D geometries of realistic objects mapped with procedurally and artistically generated SVBRDFs (based on the dataset published by Deschaintre et al. [13]). We employ specialised decoder branches in the network to output high quality shape and reflectance parameter maps and use a mix of L1 and rendering loss to train our network. We further improve the rendering loss by developing a differentiable polarized renderer, providing better gradients on the diffuse and specular behaviours.

4. Method

4.1. Data generation

To leverage polarization cues with a deep network, we need a large dataset of objects captured with different polarizer orientations along with ground truth SVBRDF. Measuring such a large dataset would require advanced, expensive equipments and colossal time. Instead, we leverage synthetic data to create a dataset of over 100000 sets of images.

Our training dataset is generated using 20 complex meshes of realistic objects and 2000 different materials (SVBRDFs) and our test dataset uses 6 and 30 unique meshes and materials respectively. We use a material model similar to Deschaintre et al. [11] using the full Fresnel equations to enable the polarization contribution computation (detailed equations are available in supplemental). For each set of polarization images, a mesh and material are selected and randomly rotated to augment diversity. We generate a rendering for four polarization filter angles ($0°, 45°, 90°,$ and $135°$) alongside the ground truth SVBRDF and depth
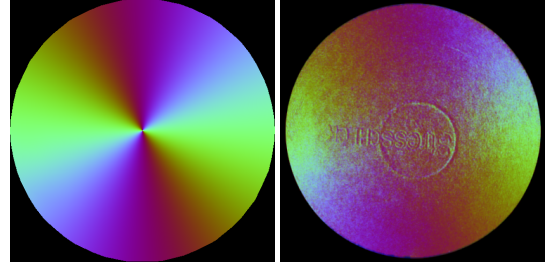

Figure 2. (a) On the left is an ideal normalized Stokes map for a sphere under frontal flash illumination: RGB color coding for Stokes vectors, R (s_0) is set to 0.5, G (s_1) and B (s_2) are normalised and mapped to $0 - 1$ range for visualisation. (b) On the right, we illustrate the signal captured in practice with a measured Stokes map of a rubber ball with embossed text under flash illumination.

maps. We further augment the dataset with a normalized Stokes map and diffuse color, computed from the different polarized renderings. As perfect Stokes map do not occur in real acquisition, see Fig. 2, we augment our synthetic generation with Gaussian noise to mimic the perturbation in the acquisition process. To better benefit from polarization cues, we simulate HDR data capture and use 16 bits PNGs. See Fig. 1 for an example of our synthetic dataset.

4.2. Polarization information

Stokes parameters. The polarization state (Stokes parameters [8]) of a reflected light gives useful cues about the surface normal. Indeed, the transformation of the Stokes parameters upon reflection largely depends on the normal of the surface. Measuring the reflected Stokes parameters under unpolarized light (e.g. flash illumination) is a relatively simple procedure, requiring only three observations with linear polarizing filter set to e.g., $0°$, $45°$ and $90°$. These three images, named I_H, I_{45} and I_V, can be used to calculate the Stokes parameters of linear polarization per pixel with the following equations:

$$\begin{aligned} s_0 &= I_h + I_v \\ s_1 &= I_h - I_v \\ s_2 &= 2 * I_{45} - s_0 \end{aligned} \tag{1}$$

Here, s_0 represents the unfiltered reflectance, s_1 represents the horizontally polarized reflectance, and s_2 represents the $45°$ polarization reflectance.

Directly measured Stokes parameters still depend on the BRDF of the surface and the lighting conditions. We normalise s_1 and s_2 with respect to each other to extract the directional information about the surface normal up to a π ambiguity (see Fig. 2). We use this normalized Stokes parameters as an additional cue for the network, helping to disambiguate the shape from the reflectance.

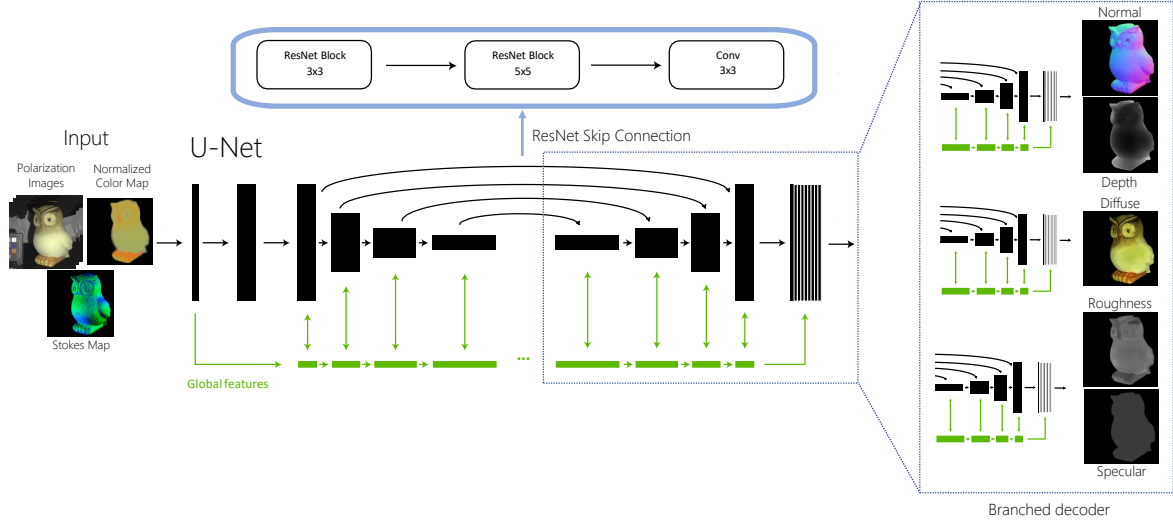In the general case, measured Stokes parameters consist

Figure 3. Our network architecture has a general U-Net [36] shape. We divide the decoders into three different branches, each handling a related set of output map(s), specifically: normal and depth, diffuse albedo, roughness and specular albedo. We introduce res-blocks on the skip connections between the encoder and the different branches of the decoder, allowing the network to adapt the information forwarded to the different branches of the decoder.

of a mix of contributions from specular and diffuse polarization caused by their respective reflectance. These two types of polarization are captured by the Fresnel equations on surface reflectance and transmission for specular and diffuse [3] polarization respectively. The magnitude of specular polarization usually dominates under direct area illumination, which is why previous studies on polarization under controlled spherical illumination have modelled only specular polarization [16, 19]. On the other hand, given our setting of frontal flash illumination, the direct specular reflection is limited to a very small frontal patch, and most of the object surface instead exhibits *diffuse polarization*. We therefore model the normalized Stokes map as the result of diffuse polarization in our synthetic training data. Under more complex environmental illumination, arbitrary mixture of specular and diffuse polarization can be observed (Sec. 5.3), which is not currently modelled synthetically.

Diffuse color. We also employ the polarization measurements to compute an estimate of normalized diffuse color. Rotating a linear polariser in front of the camera lens will change the observed intensity, as the specular reflection reaches its minimum when the polariser axis is parallel to the plane of incidence. As our flash light is white and the residual specular signal is weak, we are able to extract an estimate of the normalized diffuse color.

In practice, the minimum intensity information does not necessarily fall exactly at the three polarization angles we capture. We therefore perform a sinusoidal fitting per pixel for each observation (I_h, I_v and I_{45}) to fit the minimum value as proposed by [35]. We normalize the minimum reflectance values to extract the normalized diffuse color which we provide to our network as a reflectance cue. This

color information can however be lost in some over saturated pixels caused by extreme dynamic range of flash illumination (despite HDR imaging).

4.3. Network architecture

To estimate the shape and spatially varying reflectance of an object with our acquisition method, we train our deep network to output diffuse and specular albedos, specular roughness, normal map and depth map of the input object. We employ an Encoder-Decoder architecture inspired by U-Net [36], similar to previous works [11, 24, 12]. We specialize our decoder architecture and split it into three branches, each specialized in an aspect of shape or appearance as shown in Fig. 3. We group the specular albedo and roughness maps in one branch and the normal and depth maps in another as they are closely related. Finally, a third branch handles the diffuse albedo. All three branches of the decoder receive the same inputs from the encoder, but we propose to make the skip connections more flexible. Specifically, we add two res-blocks and a convolution layer to the skip connections, allowing the training process to adjust the information transferred to each decoder branch from the encoder. We train our network on 512x512 images.

Polarization rendering loss. We train our network using two losses: a L1 loss to regularize the training, computing an absolute difference between the output maps and the targets, and a novel polarized rendering loss. Rendering losses have shown to be efficient in training reflectance acquisition methods [11, 24]. We improve on this approach by simulating the polarization behaviour of surface reflectance in a differentiable fashion, allowing to take gradients of rendering effects from diffuse and specular polarization into account in the training process.

Code and data are available on the project page[1].

4.4. Acquisition procedure

Our acquisition process involves capturing an object under flash illumination with three polarization filter orientations: $0°$, $45°$, and $90°$. We employ a DSLR camera, a tripod and a linear polarizing filter mounted on the lens and manually rotate the polarizer on the lens to acquire the data. However, recently available polarization sensors (e.g. Sony polarsens) could also be used for this purpose, allowing rapid capture of this information in a single shot. We add a small color checker next to the captured object for white balancing and employ HDR capture (using auto-exposure bracketing on the camera) to better extract the polarization information and match the object appearance as closely as possible. The acquisition proces takes around a minute. We illustrate a typical acquisition scene in Fig. 1.

5. Evaluation

To the best of our knowledge our method is the first to leverage polarization imaging and flash illumination to recover 3D objects shape and SVBRDF. To provide a point of reference, we compare our results to Li et al. [25] and Boss et al. [7] as these methods target similar outputs. An important distinction is that these method do not benefit from the polarization information and rely on a single flash picture [25] or a flash picture and an environmentally lit picture [7]. Furthermore, the method by Boss et al. [7] focuses on lightweight architecture and fast inference and therefore uses a shallower network. We capture or generate the required inputs to these methods to provide as fair a comparison as possible despite the difference in inputs.

More results and comparisons are available in the supplemental materials, alongside moving light animations.

5.1. Comparisons

5.1.1 Quantitative comparisons

We first compare quantitatively to Li et al. [25] and Boss et al. [7] using L1 distance. We evaluate the error on the normal maps, depth and directly on renderings as these are not affected by the different BRDF models chosen by the different methods. This numerical evaluation is performed on 250 combinations of 6 randomly rotated meshes and 30 SVBRDF. The rendering error is computed over 20 renderings for each results with varying light properties. We show in Tab. 1 that our method strongly benefits from the polarization cues, white balancing and HDR imaging with significantly lower error on depth, normal and renderings.

	Li et al.	Boss et al.	Ours
Normal	42.23°	47.69°	**12.00°**
Depth	0.339	0.327	**0.174**
Renderings	0.103	0.185	**0.024**

Table 1. We evaluate our method against Li et al. [25] and Boss et al. [7] over our synthetic test set. The normal error is reported in degrees, while the rest is reported as RMSE distance. For all, lower is better. We compare 20 renderings with different illumination for each result rather than the parameters maps as the material model used by these methods vary. We see that our method, leveraging white balance, HDR inputs and polarization cues produces significantly better results on the complex shapes of our dataset. A qualitative comparison is available in Fig. 4

5.1.2 Qualitative comparisons

For qualitative comparison, we evaluate our method against Li et al. [25] and Boss et al. [7] on synthetic data and on real data (see supplemental material for comparison to [7] on real data).

In Fig. 4 we show a comparison on synthetic test data. We can see that thanks to the polarization cues, our method captures the global 3D shape of the object much better than single image methods. An important distinction is that our method does not correlate the SVBRDF variation in the input to normal variation in the output as the Stokes map disambiguate these information.

In Fig. 5 we show results on real objects. Our method better recovers the global shape of the object as well as its appearance showing that it generalizes well to real acquisition. This is particularly seen in the rendering under a novel flash lighting direction where our results demonstrate appropriate shading variation due to the estimated surface normal and reflectance maps.

More real object and synthetic results are available in supplemental material alongside light animated renderings.

5.2. Ablation study

We evaluate our method components by removing them one at a time. We evaluate quantitatively the error and report it in Tab. 2.

Improved skip connections. The first column of the table (a) evaluates our method with standard skip connections. The res-block on the skip connections allows the network to forward the most relevant information to each separate decoder branch, helping to decorrelate diffuse response from the other parameters – such correlation effect is visible in Fig. 4 in Li et al.'s result for example.

Polarized rendering loss. The second column (b) evaluates our method with a rendering loss similar to Deschaintre et al. [11]. The differentiable polarized renderings we implement helps the network to better separate the diffuse and specular signal with small improvement in the roughness and specular, but mostly in de-lighting the diffuse albedo.

Polarization cues. The third column(c) evaluates our

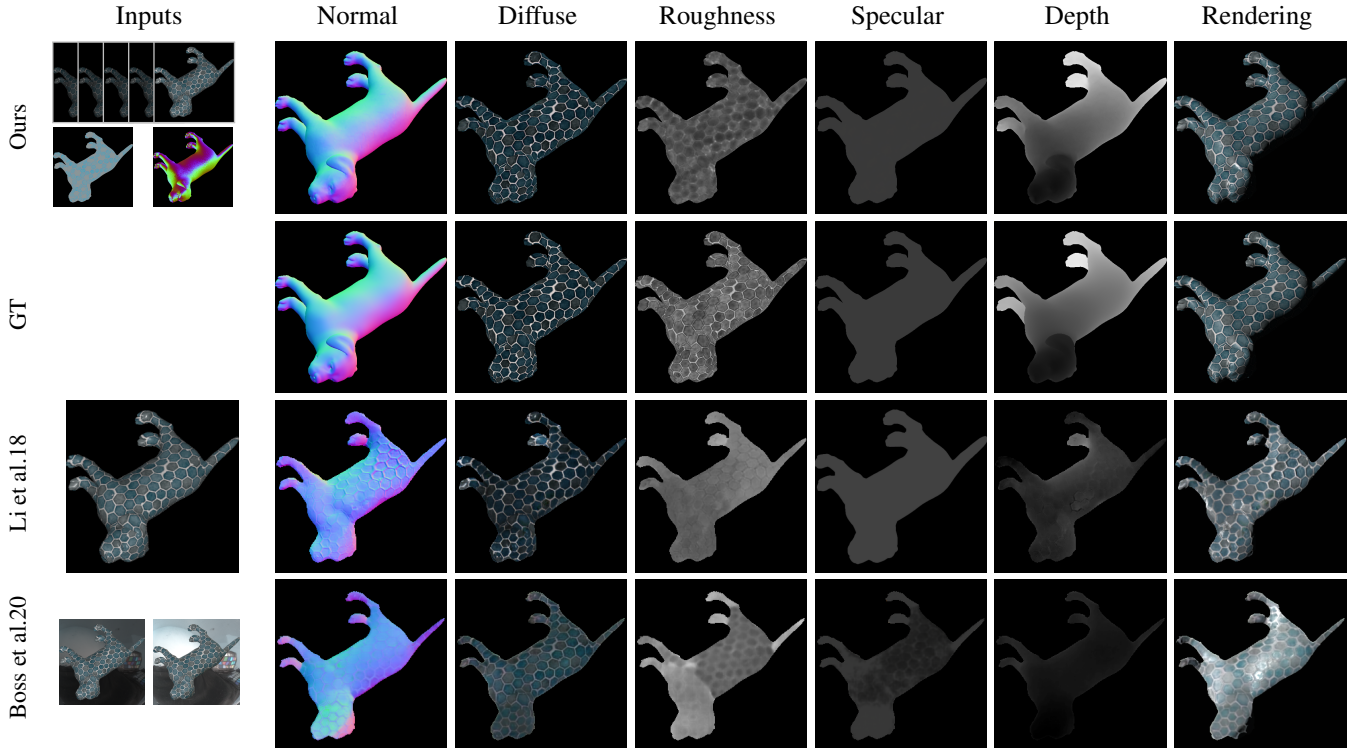| | Inputs | Normal | Diffuse | Roughness | Specular | Depth | Rendering |

Figure 4. We evaluate our method and compare to Li et al. [25] and Boss et al. [7] on sythetic data. By leveraging polarization information, our method produces more plausible results and better captures the appearance of the input. While the re-renderings (far right column) and shape can be directly compared, the BRDF parameters maps are provided for qualitative evaluation as different BRDF models are used by the different methods (Ours and GT use the same model). The inputs are adapted to each method and we use Li et al.'s and Boss et al.'s published code to generate their results. More synthetic comparisons are provided in the supplemental material.

method with a single HDR, white balanced flash input without any polarization information. All the recovered parameters significantly suffer from the absence of polarization cues. We find our single image method rendering error to be lower than compared methods, which we attribute to our use of a white balanced, HDR input and training on complex meshes, helping to recover the global curvature.

We provide qualitative comparison showing these improvements and results of our single image network in the supplemental material.

5.3. Limitations

Our method is currently limited to flash illumination where the polarization signal is dominated by diffuse polarization. The more general case of acquisition in arbitrary environmental illumination including outdoor illumination remains an open challenge due to the potentially complex mixing of specular and diffuse polarization signal. In our experiments, we found this to result in inconsistent cues with strong discontinuities in the Stokes map as shown in Fig. 6 (a). This inconsistency comes from the different light source and inter-reflection composing the illumination on a 3D object in the wild. We can retrieve interesting information in some cases where specular polarization dominates

	(a) Skip	(b) Loss	(c) Polarization	Ours
Normal	14.17°	12.38°	24.14°	**12.00°**
Diffuse	0.05	0.087	0.0783	**0.0382**
Roughness	0.115	0.131	0.155	**0.113**
Specular	0.07	0.0339	0.0555	**0.0293**
Depth	0.176	0.196	0.223	**0.174**
Rendering	0.0285	0.0351	0.0485	**0.024**

Table 2. We evaluate the contribution of our different technical components computed over our test set. For each column, we trained without this component (a) Improved skip connections (b) Polarized rendering loss (c) Polarization cues. The normal error is reported in degrees, while the rest is reported as RMSE distance. For all, lower is better. Improved skip connections and polarized rendering loss improve our results, but most importantly the polarization cues significantly improves the results on all properties.

providing us a cleaner signal similar to the flash illumination case (see Fig. 6, b). However, an accurate simulation of mixed polarization under general complex lighting environment remains an open challenge.

We are also limited in principle to acquiring dielectric objects as the information extracted through polarization cues is strictly valid for dielectrics, Baek et al. [6] show that metals polarize light elliptically. Note that dielectric assumption can still hold in practice for some metallic surfaces in
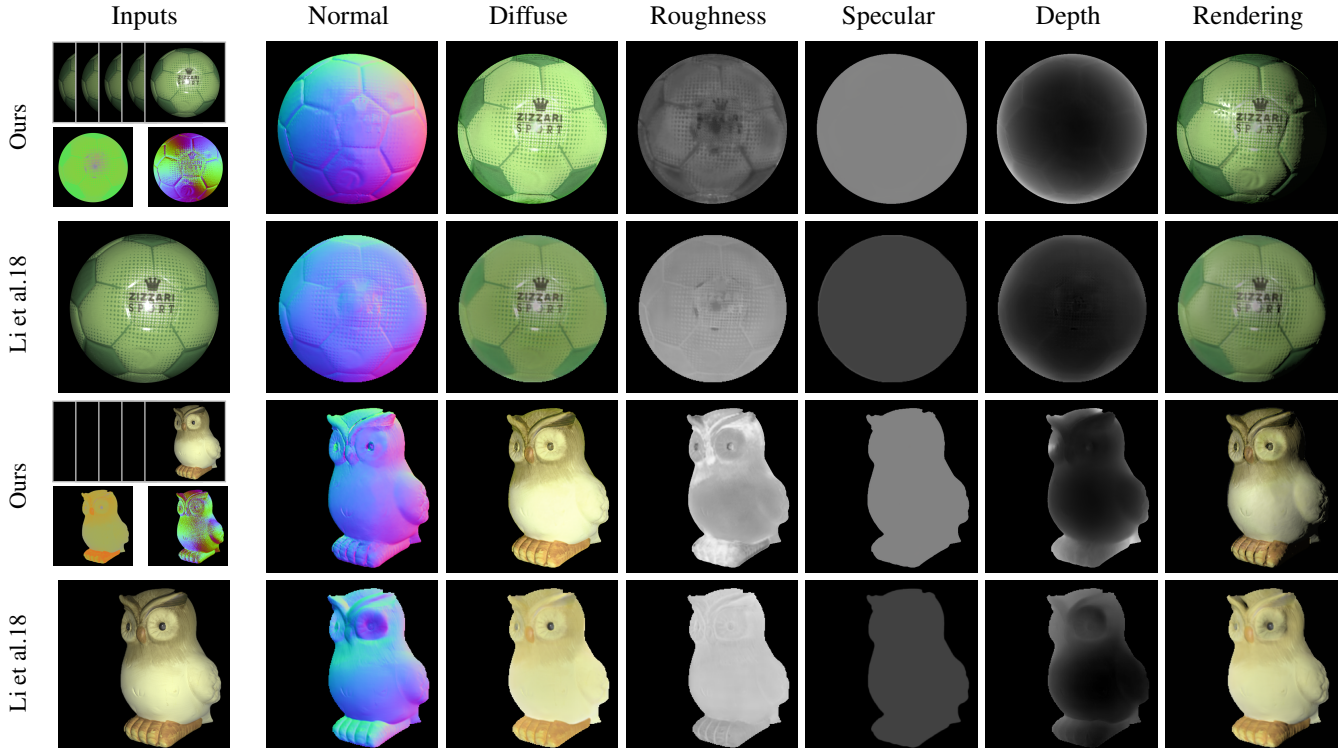
Figure 5. A comparison of our results and Li et al. [25] on real objects. We better capture the global shape and appearance of the object as can be seen in the normals and depth (normalized to be comparable) and especially in the renderings. The SVBRDF maps are displayed in the methods' respective models which are used to create renderings that can be directly compared. The polarization cues allow us to retrieve details that are difficult to extract from a single image such as the small notch on the top right of the green ball.

the real world (metal-dielectric composite, weathering effects) [35], and our acquisition approach should apply in such cases. Our method is able to provide high quality estimate of surface normal and depth, as well as specular roughness. However, the diffuse albedo estimates, in some cases, have a few specular highlights baked-in due to saturation of the flash illumination during data capture. Similarly, our approach does not fully resolve the roughness/specular ambiguity existing in single light and view acquisition methods.

6. Conclusion

In this paper, we propose the first method that combines polarization information in surface reflectance under flash illumination with recent deep learning techniques to acquire shape and SVBRDF of 3D objects. While most recent methods relied only on the RGB information of photograph(s) in a similar capture setting, our method allows to better capture shape and SVBRDF by taking linear polarization information into account. We propose a novel synthetic training dataset simulating diffuse polarization as well as improvements to network architecture and training loss for this purpose. We believe our method provides a significant improvement in quality of acquired objects while maintaining the acquisition process to be accessible and efficient.

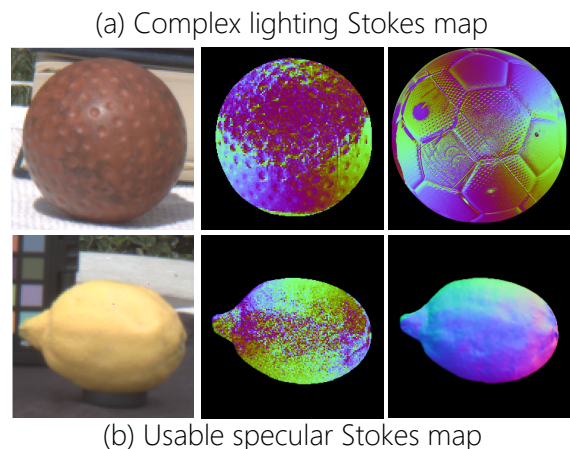(a) Complex lighting Stokes map

(b) Usable specular Stokes map

Figure 6. Polarization cues under environmental illumination. (a) An example of mixed Stokes map where we can see strong discontinuities between diffuse and specular polarization. (b) An exploitable pure specular Stokes map of a lemon under environment lighting that we used to infer shape. We include the input pictures for the Stokes maps (left) which show the capture conditions.

Acknowledgments

References

[1] Miika Aittala, Timo Aila, and Jaakko Lehtinen. Reflectance modeling by neural texture synthesis. *ACM Trans. Graph.*, 35(4):65:1–65:13, July 2016. 2

[2] Miika Aittala, Tim Weyrich, and Jaakko Lehtinen. Two-shot svbrdf capture for stationary materials. *ACM Transactions on Graphics*, 34(4):110, 2015. 2

[3] Gary Atkinson and Edwin R Hancock. Recovery of surface orientation from diffuse polarization. *Image Processing, IEEE Transactions on*, 15(6):1653–1664, 2006. 3, 5

[4] Yunhao Ba, Alex Gilbert, Franklin Wang, Jinfa Yang, Rui Chen, Yiqin Wang, Lei Yan, Boxin Shi, and Achuta Kadambi. Deep shape from polarization. In *European Conference on Computer Vision (ECCV)*, 2020. 1, 3

[5] Seung-Hwan Baek, Daniel S. Jeon, Xin Tong, and Min H. Kim. Simultaneous acquisition of polarimetric svbrdf and normals. *ACM Transactions on Graphics (Proc. SIGGRAPH Asia 2018)*, 36(6):268:1–15, 2018. 2

[6] Seung-Hwan Baek, Tizian Zeltner, Hyun Jin Ku, Inseung Hwang, Xin Tong, Wenzel Jakob, and Min H. Kim. Image-based acquisition and modeling of polarimetric reflectance. *ACM Transactions on Graphics (Proc. SIGGRAPH 2020)*, 39(4), 2020. 2, 7

[7] Mark Boss, Varun Jampani, Kihwan Kim, Hendrik P.A. Lensch, and Jan Kautz. Two-shot spatially-varying brdf and shape estimation. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2020. 1, 2, 6, 7

[8] E. Collett. *Field Guide to Polarization, SPIE Field Guides vol. FG05*. SPIE, 2005. 4

[9] Zhaopeng Cui, Jinwei Gu, Boxin Shi, Ping Tan, and Jan Kautz. Polarimetric multi-view stereo. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, July 2017. 3

[10] Paul Debevec, Tim Hawkins, Chris Tchou, Haarm-Pieter Duiker, Westley Sarokin, and Mark Sagar. Acquiring the reflectance field of a human face. In *Proceedings of ACM SIGGRAPH 2000*, pages 145–156, 2000. 2

[11] Valentin Deschaintre, Miika Aittala, Frédo Durand, George Drettakis, and Adrien Bousseau. Single-image svbrdf capture with a rendering-aware deep network. *ACM Trans. Graph.*, 37(128):15, Aug. 2018. 1, 2, 4, 5, 6

[12] Valentin Deschaintre, Miika Aittala, Frédo Durand, George Drettakis, and Adrien Bousseau. Flexible svbrdf capture with a multi-image deep network. *Computer Graphics Forum (Proceedings of the Eurographics Symposium on Rendering)*, 38(4), July 2019. 2, 5

[13] Valentin Deschaintre, George Drettakis, and Adrien Bousseau. Guided fine-tuning for large-scale material transfer. *Computer Graphics Forum (Proceedings of the Eurographics Symposium on Rendering)*, 39(4), 2020. 2, 4

[14] Duan Gao, Xiao Li, Yue Dong, Pieter Peers, Kun Xu, and Xin Tong. Deep inverse rendering for high resolution SVBRDF estimation from an arbitrary number of images. *ACM Transactions on Graphics*, 37(4), July 2019. 2

[15] Stamatios Georgoulis, Konstantinos Rematas, Tobias Ritschel, Efstratios Gavves, Mario Fritz, Luc Van Gool, and Tinne Tuytelaars. Reflectance and natural illumination from single-material specular objects using deep learning. *PAMI*, 2017. 2

[16] Abhijeet Ghosh, Tongbo Chen, Pieter Peers, Cyrus A. Wilson, and Paul Debevec. Circularly polarized spherical illumination reflectometry. *ACM Trans. Graph. (Proc. SIGGRAPH Asia)*, 29:162:1–162:12, December 2010. 2, 5

[17] Abhijeet Ghosh, Graham Fyffe, Borom Tunwattanapong, Jay Busch, Xueming Yu, and Paul Debevec. Multiview face capture using polarized spherical gradient illumination. *ACM Trans. Graphics (Proc. SIGGRAPH Asia)*, pages 129:1–129:10, 2011. 2

[18] D. Guarnera, G.C. Guarnera, A. Ghosh, C. Denk, and M. Glencross. Brdf representation and acquisition. *Computer Graphics Forum*, 35(2):625–650, 2016. 2

[19] Giuseppe Claudio Guarnera, Pieter Peers, Paul E. Debevec, and Abhijeet Ghosh. Estimating surface normals from spherical stokes reflectance fields. In *ECCV Workshop on Color and Photometry in Computer Vision*, pages 340–349, 2012. 3, 5

[20] Zhuo Hui, Kalyan Sunkavalli, Joon-Young Lee, Sunil Hadap, Jian Wang, and Aswin C. Sankaranarayanan. Reflectance capture using univariate sampling of brdfs. In *ICCV*, pages 5372–5380, 10 2017. 2

[21] Achuta Kadambi, Vage Taamazyan, Boxin Shi, and Ramesh Raskar. Polarized 3d: High-quality depth sensing with polarization cues. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 3370–3378, 2015. 3

[22] Kazutada Koshikawa. *A Polarimetric Approach to Shape Understanding of Glossy Objects*, page 190–192. Jones and Bartlett Publishers, Inc., USA, 1992. 2

[23] Xiao Li, Yue Dong, Pieter Peers, and Xin Tong. Modeling surface appearance from a single photograph using self-augmented convolutional neural networks. *ACM Trans. Graph.*, 36(4):45:1–45:11, July 2017. 2

[24] Zhengqin Li, Kalyan Sunkavalli, and Manmohan Krishna Chandraker. Materials for masses: Svbrdf acquisition with a single mobile phone image. In *ECCV*, 2018. 2, 5

[25] Zhengqin Li, Zexiang Xu, Ravi Ramamoorthi, Kalyan Sunkavalli, and Manmohan Chandraker. Learning to reconstruct shape and spatially-varying reflectance from a single image. *ACM Trans. Graph.*, 37(6), Dec. 2018. 1, 2, 4, 6, 7, 8

[26] Wan-Chun Ma, Tim Hawkins, Pieter Peers, Charles-Felix Chabert, Malte Weiss, and Paul Debevec. Rapid acquisition of specular and diffuse normal maps from polarized spherical gradient illumination. In *Rendering Techniques*, pages 183–194, 2007. 2

[27] Abhimitra Meka, Maxim Maximov, Michael Zollhoefer, Avishek Chatterjee, Hans-Peter Seidel, Christian Richardt, and Christian Theobalt. Lime: Live intrinsic material estimation. In *CVPR*, June 2018. 2

[28] Daisuke Miyazaki, Masataka Kagesawa, and Katsushi Ikeuchi. Polarization-based transparent surface modeling from two views. In *ICCV*, pages 1381–1386, 2003. 3

[29] Daisuke Miyazaki, Robby T. Tan, Kenji Hara, and Katsushi Ikeuchi. Polarization-based inverse rendering from a single view. In *ICCV*, pages 982–987, 2003. 2

[30] Volker Müller. Polarization-based separation of diffuse and specular surface-reflection. In *Mustererkennung 1995*, pages 202–209. Springer, 1995. 2

[31] Giljoo Nam, Joo Ho Lee, Diego Gutierrez, and Min H. Kim. Practical svbrdf acquisition of 3d objects with unstructured flash photography. *ACM Transactions on Graphics (Proc. SIGGRAPH Asia 2018)*, 36(6):267:1–12, 2018. 2

[32] Shree K. Nayar, Xi-Sheng Fang, and Terrance Boult. Separation of reflection components using color and polarization. *IJCV*, 21(3), 1997. 2

[33] Stefan Rahmann and Nikos Canterakis. Reconstruction of specular surfaces using polarization imaging. *CVPR*, 1:149, 2001. 3

[34] J. Riviere, P. Peers, and A. Ghosh. Mobile surface reflectometry. *Computer Graphics Forum*, 35(1):191–202, 2016. 2

[35] Jérémy Riviere, Ilya Reshetouski, Luka Filipi, and Abhijeet Ghosh. Polarization imaging reflectometry in the wild. *ACM Trans. Graph.*, 36(6), Nov. 2017. 2, 4, 5, 8

[36] O. Ronneberger, P.Fischer, and T. Brox. U-net: Convolutional networks for biomedical image segmentation. In *Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, volume 9351 of *LNCS*, pages 234–241, 2015. 5

[37] Firooz Sadjadiz and Farzad Sadjadi. Extraction of surface normal and index of refraction using a pair of passive infrared polarimetric sensors. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 1–5, 2007. 3

[38] Megumi Saito, Yoichi Sato, Katsushi Ikeuchi, and Hiroshi Kashiwagi. Measurement of surface orientations of transparent objects by use of polarization in highlight. *J. Opt. Soc. Am. A*, 16(9):2286–2293, 1999. 3

[39] W.A.P. Smith, R. Ramamoorthi, and S. Tozza. Linear depth estimation from an uncalibrated, monocular polarisation image. In *Proc. of European Conference on Computer Vision (ECCV)*, pages 517–526, 2016. 3

[40] Vimal Thilak, David G. Voelz, and Charles D. Creusere. Polarization-based index of refraction and reflection angle estimation for remote sensing applications. *Appl. Opt.*, 46(30):7527–7536, 2007. 3

[41] Shinji Umeyama and Guy Godin. Separation of diffuse and specular components of surface reflection by use of polarization and statistical analysis of images. *PAMI*, 26(5), 2004. 2

[42] Tim Weyrich, Jason Lawrence, Hendrik P. A. Lensch, Szymon Rusinkiewicz, and Todd Zickler. Principles of appearance acquisition and representation. *Found. Trends. Comput. Graph. Vis.*, 4(2):75–191, Feb. 2009. 2

[43] L. B. Wolff. Surface orientation from two camera stereo with polarizers. In *Proc. SPIE Conf. Optics, Illumination and Image Sensing for Machine Vision IV*, volume 1194, pages 287–297, 1989. 3

[44] Lawrence B. Wolff and Terrance E. Boult. Constraining object features using a polarization reflectance model. *PAMI*, 13(7):635–657, 1991. 2

[45] Hongzhi Wu and Kun Zhou. Appfusion: Interactive appearance acquisition using a kinect sensor. *Computer Graphics Forum*, 34(6):289–298, 2015. 2

[46] Wenjie Ye, Xiao Li, Yue Dong, Pieter Peers, and Xin Tong. Single photograph surface appearance modeling with self-augmented CNNs and inexact supervision. *Computer Graphics Forum*, 37(7), Oct 2018. 2

[47] Jinyu Zhao, Yusuke Monno, and Masatoshi Okutomi. Polarimetric multi-view inverse rendering. In *European Conference on Computer Vision (ECCV)*, 2020. 3