

MIT Open Access Articles

*Depth Sensing Using Geometrically
Constrained Polarization Normals*

The MIT Faculty has made this article openly available. *Please share* how this access benefits you. Your story matters.

Citation: Kadambi, Achuta, Vage Taamazyan, Boxin Shi, and Ramesh Raskar. "Depth Sensing Using Geometrically Constrained Polarization Normals." *International Journal of Computer Vision* 125, no. 1–3 (June 22, 2017): 34–51.

As Published: <http://dx.doi.org/10.1007/s11263-017-1025-7>

Publisher: Springer-Verlag

Persistent URL: <http://hdl.handle.net/1721.1/116370>

Version: Author's final manuscript: final author's manuscript post peer review, without publisher's formatting or copy editing

Terms of use: Creative Commons Attribution-Noncommercial-Share Alike



Depth Sensing using Geometrically Constrained Polarization Normals

Achuta Kadambi · Vage Taamazyan · Boxin Shi · Ramesh Raskar

Received: date / Accepted: date

Abstract Analyzing the polarimetric properties of reflected light is a potential source of shape information. However, it is well-known that polarimetric information contains fundamental shape ambiguities, leading to an underconstrained problem of recovering 3D geometry. To address this problem, we use additional geometric information, from coarse depth maps, to constrain the shape information from polarization cues. Our main contribution is a framework that combines surface normals from polarization (hereafter polarization normals) with an aligned depth map. The additional geometric constraints are used to mitigate physics-based artifacts, such as azimuthal ambiguity, refractive distortion and fronto-parallel signal degradation. We believe our work may have practical implications for optical engineering, demonstrating a new option for state-of-the-art 3D reconstruction.

Keywords Computational Photography · Light Transport · Depth Sensing · Shape from Polarization

A. Kadambi, V. Taamazyan, B. Shi, R. Raskar
MIT Media Lab
75 Amherst Street, Cambridge MA 02139, USA
Tel.: +1 617-715-4683
E-mail: †achoo@mit.edu, vaheta@gmail.com, †shiboxin@media.mit.edu, raskar@media.mit.edu

V. Taamazyan
Skoltech-MIT Initiative, Skolkovo Institute of Technology

V. Taamazyan
Tardis 3D Technologies LLC
E-mail: vaheta@tardis3d.ru

B. Shi
Artificial Intelligence Research Center, National Institute of Advanced Industrial Science and Technology (AIST)
E-mail: boxin.shi@aist.go.jp

1 Introduction

Today, photographers use polarizing filters on 2D cameras to create aesthetic photographs. But what if a polarizing filter is used in the context of 3D photography? In the vein of computational imaging, we present a co-design of polarized optics and post-capture processing for the task of 3D imaging.

Currently, consumer 3D cameras like the Microsoft Kinect produce depth maps that are often noisy and lack sufficient detail. Using computational processing alone is unlikely to mitigate the problem (e.g., if the noise is filtered, so is the detail). As such, there is recent interest in using a *joint* optical and computational approach to both enhance detail and remove noise. One of the most promising solutions is to combine the captured, coarse depth map with surface normals obtained from photometric stereo (PS) or shape-from-shading (SfS). This depth-normal fusion is logical—the coarse depth map provides the geometric structure and the surface normals capture fine detail to be fused. Encouraging results have been shown by several papers that combine low-quality depth maps with surface normal maps obtained from SfS or PS.¹ Unfortunately, SfS and PS are limited by similar scene assumptions, e.g., restrictive lighting and material assumptions. As a complementary technique, this paper proposes the first use of surface normals from polarization to enhance depth maps. While our proposed technique also has assumptions on lighting and material properties, these assumptions work on new types of scenes that PS and SfS cannot handle.

For over a century it has been known that the shape of an object causes small changes in the polarization of reflected light, best visualized by rotating a polarizing filter in front

† Correspondence addressed to A.K. and B.S.

¹ For example, [62, 16, 60] using SfS, and [38, 17] using PS.

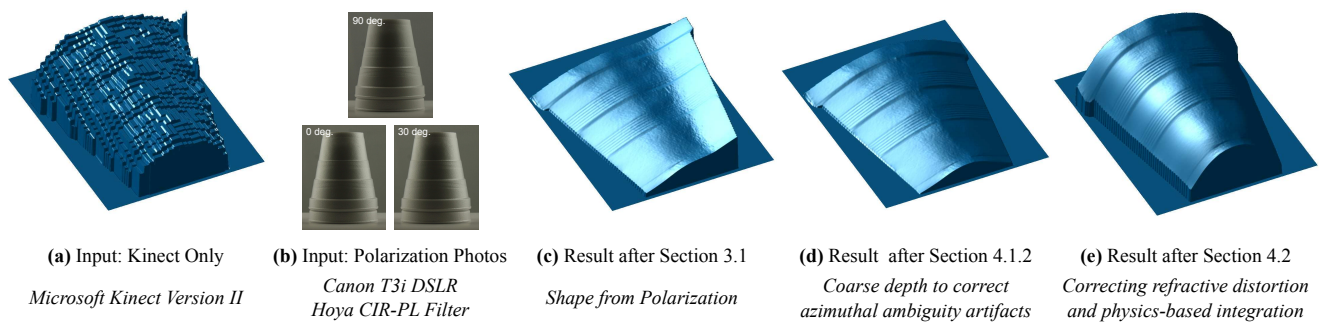


Fig. 1 Outline of proposed technique. (a) The Kinect depth of an object is combined with (b) three photos at different rotations of a polarizing filter. (c) Integration of surface normals obtained from Fresnel equations. Note the azimuthal ambiguity (observed as a flip in the shape) and distortion of the zenith angle (observed as flatness in the shape). (d) Integration of surface normals after correcting for azimuthal ambiguity removes the flip, and the final result is shown in (e) after correcting for zenith distortion and using physics-based integration.

of a digital camera.² If properly constrained, obtaining shape through polarization has potential advantages over SfS and PS, including:

- Passive capture: assuming light incident on an object is unpolarized, the surface normals can be obtained by rotating a polarizer at the imaging sensor.
- Robustness to diffuse interreflections: unlike SfS and P-S, diffuse interreflections do not significantly corrupt the estimated shape.
- Material invariant capture: the physics of the shape from polarization problem hold for materials ranging from dielectrics to metals to translucent objects.
- Lighting robust capture: if the incident light is unpolarized shape estimation is robust and can be conducted indoors, outdoors, or under patterned illumination.

However, there appears to be untapped potential as obtaining surface normals through polarization is not yet a mature technique. The obtained normals are drastically distorted. Specific open problems [31,4] include:

1. **Ambiguity:** The azimuth component of the surface normal contains an ambiguity of π radians, which leads to ambiguous flips in the 3D shape.
2. **Refractive distortion:** Obtaining the zenith component of the surface normal requires knowledge of the refractive index to estimate accurate 3D shape.
3. **Fronto-parallel surfaces:** When the zenith angle is close to zero, the obtained normals are noisy.
4. **Depth discontinuities:** Even if the normals are obtained correctly, integration of gradients must be performed to recover the 3D shape.
5. **Relative depth:** Integrating surface normals obtains only relative 3D shape, up to offset and scaling constants.

We believe that a solution to the above 5 problems is a significant step toward practical depth sensing. Although there are partial solutions in prior art, that address restricted cases, a comprehensive solution to the above 5 problems has

² Formalized by Augustin-Jean Fresnel (1788-1827) as the famous “Fresnel Equations”.

not been presented. In this paper, the 5 challenges are addressed by starting with a coarse depth map as a constraint to correct the normals obtained from polarization. While we do not solve all open problems, our correction is sufficient to realize early-stage practical results and hopefully spur future work. An overview of our approach is summarized in Figure 1.

A preliminary version of this work was demonstrated live at SIGGRAPH in August 2015 [26] and appeared at ICCV in December 2015 [25]. This paper includes additional material pertaining to: a theoretical analysis of mixed reflections based on wave interference (Section 5.1), a practical analysis of the unpolarized world assumption (Section 5.2), additional results including a failure case and demonstration using a laser scanner, and finally expansions to the prose and discussion of related works.

1.1 Contributions

Conceptually, we propose a technique that exploits surface normals from polarization cues to enhance the quality of a coarse depth map. This is the first technique that uses a depth map to address the fundamental ambiguities in SfP. Very specifically, we devise a physics-based framework, wherein the coarse depth map is used to resolve azimuthal ambiguity (addressing problem 1) and correct for refractive distortion (solving problem 2). To recover 3D shape, we propose a spanning tree integration scheme that uses the degree of polarization as a weighting parameter. This approach, specifically designed for polarization normals, addresses problem 3. As is well-known, the general fusion of depth and normals solves problems 4 and 5.

Taken together, we then analyze our technique for suitability on real-world, mixed surfaces using a wave-based analysis of light transport. Practical results are benchmarked against ground truth data, Kinect data and state-of-the-art 3D enhancement techniques [60], demonstrating clear improvement on a wide variety of scenes.

Table 1 The fusion of polarization and depth works on complex scenes, with shiny objects, interreflections, and uncontrolled lighting. Single-shot capture is possible using a polarization camera. These cameras are sold with a sensor mosaic for multiple polarization channels.⁴

Depth Prior +X	Material Assumption	Lighting Assumption	Lighting Type	Diffuse Interreflections	Specular Interreflections	Minimum Images	Compact Sensor
Photometric Stereo [67,17]	Lambertian	Distant Lighting	Active and Controlled	Not robust	Not robust	3	No
Shape From Shading [59,62,16,60,42]	Lambertian	Spherical Harmonic See Basri [5].	Passive	Not robust	Not robust	1 [Regular Camera]	Yes
Shape from Polarization [Proposed Combination]	Dielectrics or Low-frequency Transition	Unpolarized	Passive	Robust	Not robust	3 [Camera + Polz. Filt.] 1 [Polz. Camera]	Yes

2 Related Work

Shape from Polarization (SfP) estimates surface normals by analyzing the polarization properties of reflected light. An overview can be found in [44], which describes how the degree of polarization and orientation of specular reflections can be used to obtain surface normals. The information in specularly polarized light can also be extended to transparent objects [46,32]. On the other hand it is also possible to estimate the shape of dielectric objects using the cues from diffusely polarized reflections [34,4]. A few notable extensions have been proposed to enhance SfP. For instance, Zhang et al. propose a method whereby incident, polarized information is used to improve the SNR characteristics of shape from diffuse polarization [66]. In another example, Huynh et al. combine spectral (wavelength) information with polarimetric measurements to recover surface normals and refractive index [20]. Recently, Smith et al. proposed a very interesting approach that allows linear depth estimation in uncalibrated scenes with assumptions that appear more general to previous work, but still require scene conditions like the refractive index to be known [53]. Taken together these papers illustrate the benefit of using polarization, but—regardless of which polarization technique is used—the five problems we illustrated in Section 1 are not handled by a single paper. These include, for example, a lack of unicity when solving for the azimuth and zenith components of the estimated surface normal. To solve such ambiguities, [3] uses two viewpoints to obtain polarization measurements. The work by [34] instead opts to use priors on the distribution of surface normals, which was extended to obtain rough shape from space carving on multi-view data [33]. For a more complete picture, the reader is directed to a survey of polarimetric shape techniques, published recently by Stolz et al. [54] as well as chapter 10 of Robles-Kelly and Huynh [45]. Our work is closely related to that of [39], which uses information from shading to constrain SfP. In comparison, we use the additional measurement of coarse depth to sufficiently address major artifacts in classic SfP, which presents an additional option that is complementary to previous works.

Combining depth and normal cues is, by now, a popular technique to obtain 3D information. Generally speaking, pri-

or art combines a geometric-based technique to obtain rough depth with a photometric-based technique to obtain surface normals (one of the early works in this line is Helmholtz Stereopsis [70]). Such a fusion is well-motivated: (1) The geometric approach helps to remove the ambiguities in photometric techniques, such as SfS or uncalibrated PS; (2) The photometric approach helps in adding surface details to the coarse depth map from the geometric data; and (3) the rough depth map provides anchor points for the surface-from-gradient problem, addressing the challenge of non-integrable surfaces at depth discontinuities. There are numerous existing works that partially or completely reflect these three aspects. Combinations that have been explored previously include: combining a laser scan with PS [38], multi-view stereo with SfS [59] or PS [65,23,9], consumer depth sensing with SfS [62,16,60], and consumer depth sensing with PS [67,17,50]. If high-quality surface normals are not available, fusing a sequence of overlapping depth maps is a popular approach to produce a smooth surface for various interactive applications [21] or large-scale, real-time surface reconstruction [40]. Tab. 1 summarizes the benefits and limitations of our proposed approach, while Fig. 2 outlines the logic of using polarization instead of SfS and PS.

Polarization in computational imaging: Some researchers have exploited polarized spherical gradient illumination patterns coupled with a polarizer in front of a camera to capture the behavior of polarized light transport for high-resolution facial scanning of static expressions [29], estimation of specular roughness and anisotropy [11], inference of per-pixel surface reflectance parameters through circular polarization cues [12,14], and for multi-view facial performance capture [13]. Polarization cues are also widely used in computational imaging applications, such as separation of multipath information in ToF imaging [56], separation of diffuse and specular reflections [37,69], dehazing of images [48], image mosaicing and panoramic stitching [49], illumination multiplexing [7], vehicular detection [8] and camera [30,22] or display hardware [28]. In addition, polarization cues can be used to recover shape of translucent objects [6], shape of the ocean surface [64], or address scattering underwater [55]. For a more complete overview of

⁴ A specific example of a polarization camera can be found at this URL: <http://www.4dtechnology.com/products/polarcam.php>

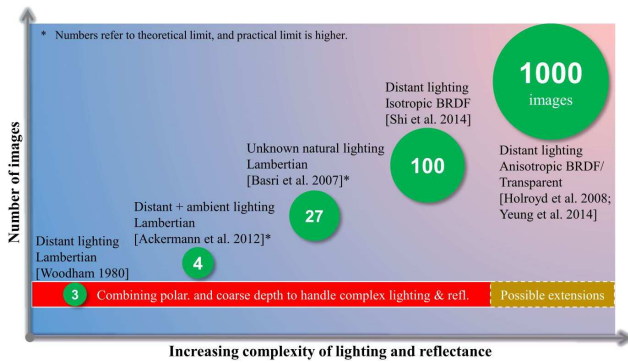


Fig. 2 Why use polarization for obtaining surface normals? When lighting conditions become less controlled or the material properties increase in complexity, alternative techniques may require more images. By imposing geometric constraints on polarization cues, it may be possible to span a large range of material complexities and lighting conditions using only 3 photographs. For a more complete picture, the figure references are, from left-to-right: [58],[1],[5],[51],[19],[61]. The proposed technique is shown in the red bar, with a possible extension, shown in orange.

polarization imaging and its applications the reader is directed to [57] and [68].

3 Basics of shape from polarization

To provide a self-contained overview, we review the shape from polarization problem in condensed form.

3.1 Surface normals from polarization cues

A photograph is captured with a polarizer at an angle ϕ_{pol} . At a single image point, the intensity can be written as

$$I(\phi_{\text{pol}}) = \frac{I_{\text{max}} + I_{\text{min}}}{2} + \frac{I_{\text{max}} - I_{\text{min}}}{2} \cos(2(\phi_{\text{pol}} - \phi)), \quad (1)$$

where the three unknown variables in this equation are I_{max} , I_{min} , and ϕ , shown in Fig. 3. Sampling different values on the sinusoid amounts to taking pictures with different rotations of the polarizer angle.⁵ We use standard spherical and cartesian coordinates to represent the normal vectors. In the former, we are reliant on the zenith angle and azimuth angle to define the surface normal.

Obtaining the azimuth of surface normal: By sampling three values of ϕ_{pol} it is sufficient to characterize the amplitude, phase, and offset of the received signal. The azimuth angle, ϕ is encoded as the phase of the received signal. However, note that the solution is not unique: *two azimuth angles, shifted apart by π radians cannot be distinguished in the polarized images.* Concretely, note that an azimuth angle of ϕ

⁵ For this paper, we rotate the polarizer to the desired angle. However, the mechanical process of rotation could lead to small errors, for which Schechner has devised a self-calibrating solution [47].

and $\phi + \pi$ return the same value for Equation 1. In practice, this leads to disappointing results when using shape from polarization. Solving this ambiguity is one focus of this paper.

Obtaining the zenith of surface normal: The degree of polarization is based on the amplitude and offset of Equation 1 and can be written as

$$\rho = \frac{I_{\text{max}} - I_{\text{min}}}{I_{\text{max}} + I_{\text{min}}}. \quad (2)$$

Substituting the Fresnel equations (see [18]) into Equation 2 allows the degree of polarization to be written as

$$\rho = \frac{(n - \frac{1}{n})^2 \sin^2 \theta}{2 + 2n^2 - (n + \frac{1}{n})^2 \sin^2 \theta + 4 \cos \theta \sqrt{n^2 - \sin^2 \theta}}, \quad (3)$$

where n denotes the refractive index and θ the zenith angle. If the refractive index is known, the zenith angle can be estimated either in closed-form, or by numerical optimization. Unfortunately, it is difficult to know the refractive index at each pixel, particularly in a scene with mixed materials; this is one of the challenges with SfP that we address.

Specular vs diffuse polarization: Equation 3 is robust for dielectric surfaces, but cannot be used on non-dielectric surfaces, such as mirrors or metals. These materials do not reflect back any diffuse light, but the relation

$$\rho = \frac{2n \tan \theta \sin \theta}{\tan^2 \theta \sin^2 \theta + |n^*|^2}, \quad (\text{Specular Model}) \quad (4)$$

where $|n^*|^2 = n^2(1 + \kappa^2)$ and κ is the attenuation index of the material, allows the zenith angle to be found [35]. It is possible to identify whether to use Equation 3 or 4 to obtain the zenith angle based on the degree of polarization at a single pixel.⁶ Variants of the method thus described are implemented in previous SfP work [3, 32, 34]. Due to the limitations of SfP (see bullets 1-3 from Section 1), SfP has never been considered as a robust alternative to SfS or PS.

Implementing shape from polarization: Although SfP has many ambiguities, as described above, our method uses standard SfP to recover an *initial* normal map from polarization. This will have all the aforementioned ambiguities, but our contribution is to correct it with geometric constraints. To obtain the initial normal map, we choose a refractive index to 1.5 (this value is obviously wrong, but is close to the real-world value of many common objects). To decide whether Equation 3 or 4 is more germane, we make the observation that if ρ is high, the reflection is specular, and the latter equation is more relevant. The result from SfP on the coffee cup scene is shown in Figure 1c, where the challenges with SfP can be readily observed.

⁶ In practice, the degree of polarization is generally an order of magnitude larger for specular dominant reflections.

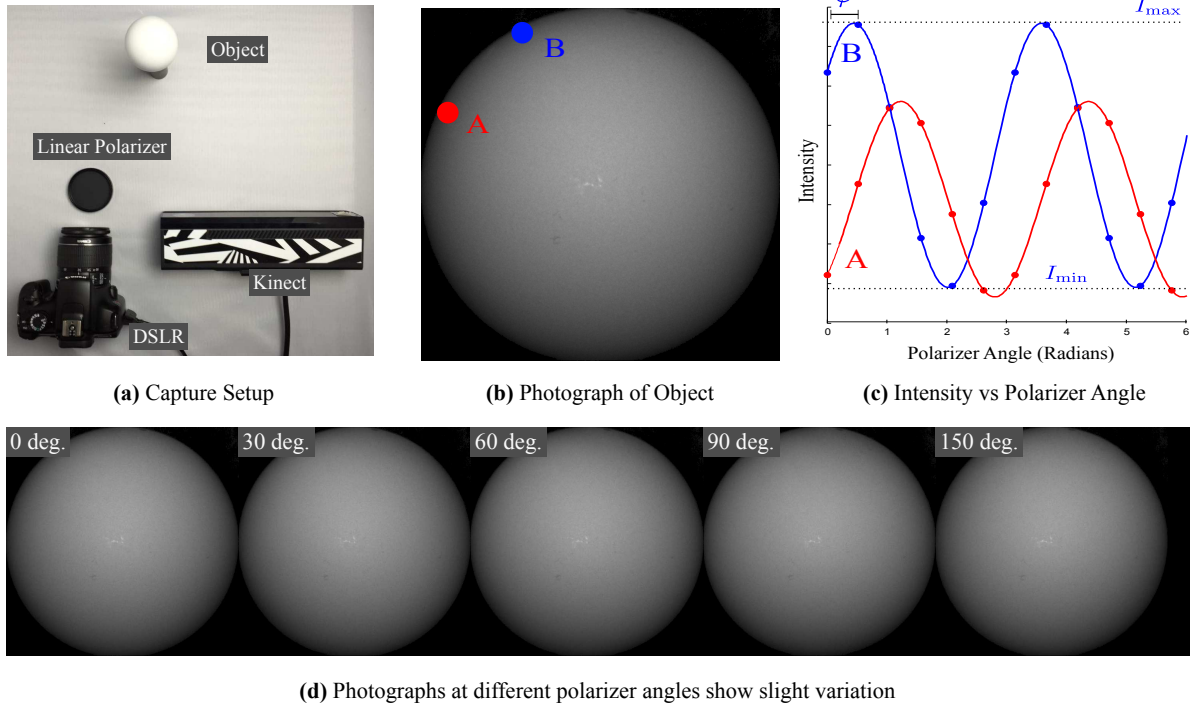


Fig. 3 Capture setup. In (a) a standard camera with a polarizing filter is used to photograph a diffuse sphere under different filter rotations. The Kinect is not used in this figure, but shown for context (as the rest of the paper uses the Kinect). (b) A photograph of the sphere with two points on the geometry labeled. (c) Plotting intensity at different polarizer angles at the labelled points on the geometry. Observe the variation in the sinusoid that is due to the geometry. (d) Selected photos of the object at different polarizer angles show a slight variation (figure must be viewed digitally).

4 Framework for Depth-Polarization Fusion

Scenes are assumed to have the following properties: (1) unpolarized ambient light; (2) no specular interreflections; (3) only dielectric materials or low-frequency changes in materials; and (4) diffuse-dominant or specular-dominant surfaces.⁷

4.1 Correcting normals from polarization

We use the obtained depth map to correct systematic distortions in the normals from polarization. Let $\mathbf{D} \in \mathbb{R}^{M \times N}$ denote the obtained depth map. Our correction scheme operates in the normal domain, so we find the surface normals from the depth map, denoted as $\mathbf{N}^{\text{depth}} \in \mathbb{R}^{M \times N \times 3}$. Following the work of [27], we express the depth map $\mathbf{D} \in \mathbb{R}^{M \times N}$ as a point cloud of real-world coordinates as $\mathbf{P}_{x,y} = \left[-\frac{u}{f_x} \mathbf{D}_{x,y} - \frac{v}{f_x} \mathbf{D}_{x,y} \ \mathbf{D}_{x,y} \right]^T$, where u and v denote pixel coordinates and f_x and f_y denote the focal length in units of pixels. For each scene point $\mathbf{P}_{x,y}$, we find J points whose Euclidean distance away from $\mathbf{P}_{x,y}$ is less than a fixed absolute distance.⁸ Thus, for a given point, $\mathbf{P}_{x,y}$, we have a set that in-

cludes this point and its neighbors, $\{\mathbf{P}_{x,y}, \mathbf{P}_{x_1,y_1}, \dots, \mathbf{P}_{x_J,y_J}\}$. By stacking all of these points into a matrix,

$$\mathbf{Q} = \begin{bmatrix} - & \mathbf{P}_{x,y}^T & - \\ - & \mathbf{P}_{x_1,y_1}^T & - \\ & \vdots & \\ - & \mathbf{P}_{x_J,y_J}^T & - \end{bmatrix}, \quad (5)$$

we find the normal by solving

$$\mathbf{N}_{x,y}^{\text{depth}} = \arg \min_{\mathbf{n}} \left\| \left(\mathbf{Q} - \tilde{\mathbf{Q}} \right) \mathbf{n} \right\|_2^2,$$

where each identical row of the matrix $\tilde{\mathbf{Q}} \in \mathbb{R}^{(J+1) \times 3}$ contains the mean across the first dimension of \mathbf{Q} (i.e., a point that is the average of all the points in \mathbf{Q}). The smoothness of $\mathbf{N}^{\text{depth}}$ can be changed by controlling the size of the neighborhood that the search algorithm returns.

4.1.1 Removing low-frequency azimuthal ambiguity

Consider the corner scene in Fig. 4. Using a coarse depth sensor, a low-frequency version of the surface is acquired (note the smoothness in the 3D shape in Fig. 4b). On the other hand, the shape from polarized normals is very inaccurate due to the azimuthal flip, but the high-frequency detail can be recovered.

⁷ Refer to Section 5 for a detailed analysis of key assumptions.

⁸ For all experiments, this distance was set to 20 millimeters. To find the neighborhood, we use the kd-tree search algorithm, which can be implemented by the `RANGESEARCH` command in MATLAB.

Let $\mathbf{N}^{\text{polar}} \in \mathbb{R}^{M \times N \times 3}$ denote the normal map obtained from polarization cues.⁹ The goal is to find an operator \mathcal{A} that relates $\mathbf{N}^{\text{polar}}$ and $\mathbf{N}^{\text{depth}}$, which can be expressed numerically as

$$\widehat{\mathcal{A}} = \arg \min_{\mathcal{A}} \left\| \mathbf{N}^{\text{depth}} - \mathcal{A} \left(\mathbf{N}^{\text{polar}} \right) \right\|_2^2.$$

Without any additional constraints, this optimization is ill-posed. However, to resolve polarization ambiguity we are only interested in representing \mathcal{A} as a binary operator, i.e., the operator \mathcal{A} has a matrix representation \mathbf{A} of dimension $M \times N$, where $\mathbf{A} \in \{0, 1\}$. Here, the binary state of each element, \mathbf{A}_{mn} , corresponds to whether a normal at a specific pixel location will undergo an azimuthal rotation or not. The operator's function follows suit: at each pixel, the operator \mathcal{A} either rotates the azimuth angle by π or does nothing depending on the binary value of \mathbf{A}_{mn} . Since the goal is to solve low-frequency ambiguity, we impose an additional constraint that the matrix representation of \mathcal{A} is smooth in the sense of total variation. Taken together, this can be expressed as a minimization problem:

$$\widehat{\mathcal{A}} = \arg \min_{\mathcal{A}} \left\| \mathbf{N}^{\text{depth}} - \mathcal{A} \left(\mathbf{N}^{\text{polar}} \right) \right\|_2^2 + \gamma \mathcal{F}_2(\mathbf{A}) \quad (6)$$

subject to $\mathbf{A} \in \{0, 1\}$,

where the parameter γ controls the (piecewise) smoothness of the solution and \mathcal{F}_2 is the standard 2D TV regularizer, such that $\mathcal{F}_2(\mathbf{A}) = \sum_m \sum_i |\nabla_x \mathbf{A}_{mi}| + |\nabla_y \mathbf{A}_{mi}|$, where ∇_x and ∇_y are horizontal and vertical gradients. In particular, since the decision variable is binary, we use a modified version of graph-cuts, which is often used to segment an image into foreground and background patches.¹⁰ After obtaining $\widehat{\mathcal{A}}$ it is possible to correct low-frequency changes in the ambiguity by applying the operator to the polarization normal:

$$\mathbf{N}^{\text{corr}} = \widehat{\mathcal{A}} \left(\mathbf{N}^{\text{polar}} \right). \quad (7)$$

After correcting for low-frequency ambiguity, we can return to the physical experiment on the corner. By applying the techniques introduced in this section we have traversed from the ambiguous normals in Fig. 4g to the correctly flipped normals in Fig. 4h. For this example, the ambiguity was low-frequency in nature, so the coarse depth map was sufficient.

4.1.2 Removing high-frequency azimuthal ambiguity

If the depth map is coarse, consisting of low-frequency information, then it cannot be used to resolve regions with

⁹ $\mathbf{N}^{\text{polar}}$ is obtained through shape from polarization and this normal map will suffer from the physics-based artifacts described previously. This can be seen visually in Figure 1c.

¹⁰ The 2D-TV implementation parallels the optimization program from prior work [24].

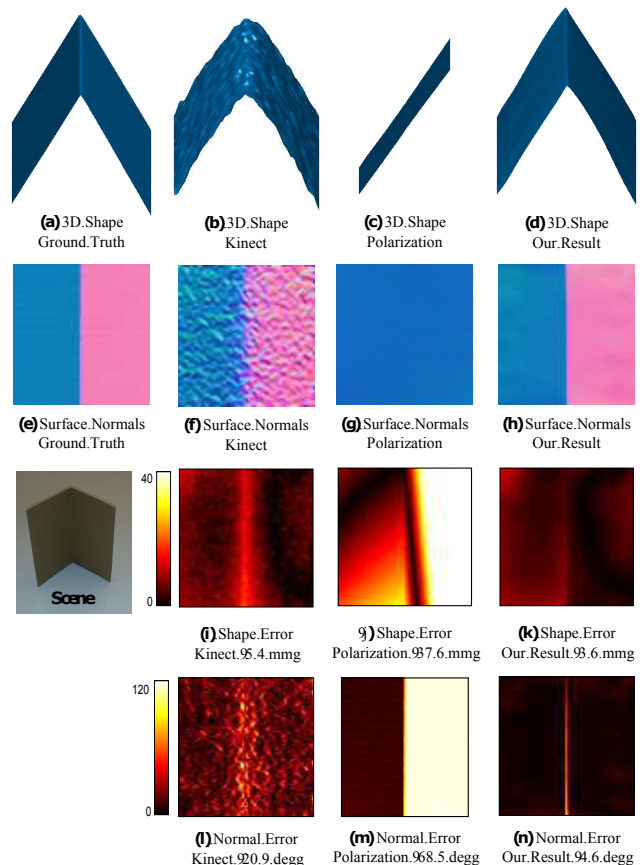


Fig. 4 A commonly used benchmark scene [15,36]. Combining polarization with Kinect results in improved performance. The top row shows the 3D shape of a corner. The second row shows the surface normals. The third row plots the estimated surface error in millimeters and the fourth row depicts the estimated angular error of surface normals in degrees w.r.t. the ground truth.

high-frequency ambiguity. To address this challenge we force these regions of the surface to be closed.

Fig. 5a illustrates a conceptual example with a high-frequency V-groove on a plane. The normals are disambiguated correctly on the plane, but the ridge cannot be disambiguated using the method from Section 4.1.1. In particular, observe that the high-frequency ridge can take one of six forms. To constrain the problem, we define an anchor point at the start of the high frequency region and a pivot point at the center of the ridge. The anchor point represents the boundary condition for the high-frequency ridge and the pivot point occurs on an edge not on the boundary.

Given the anchor and pivot points, we define a facet as the set of points between the anchor and pivot points (see Fig. 5b). A facet can form a planar or nonplanar surface. Assuming there are K facets, there are $2 \times 2^K - V$ possible surface configurations, where V is the number of possible closed surfaces.¹¹ This surface has two facets and two closed

¹¹ Calculation of $2 \times 2^K - V$: each facet can have 2 possible normal orientations due to the azimuthal ambiguity, leading to 2^K possible sur-

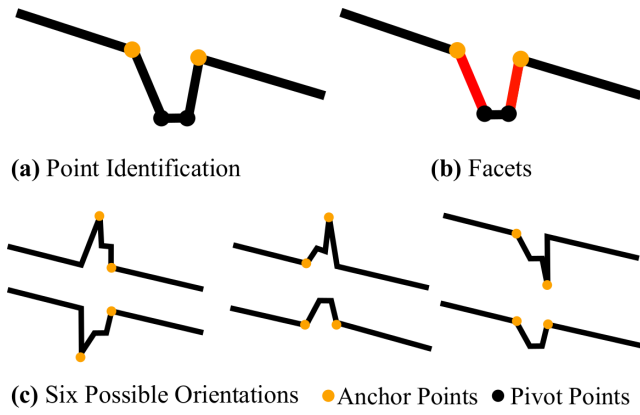


Fig. 5 Addressing high-frequency ambiguity. Consider a planar surface with a high-frequency pit. (a) Anchor and pivot points are identified to group points on the ambiguity region into (b) facets. (c) Each facet can be rotated by π radians, creating ambiguities.

configurations, and therefore six possible surface configurations. Four of these are not closed, i.e., the high-frequency region has a discontinuity at an anchor point. The discontinuity is physically possible—i.e., the V-groove could actually be a ramp in the real world—but it is less likely that the high frequency detail has such a discontinuity exactly at the anchor point. Therefore, we assume the high-frequency surface is closed.¹² Of the two closed surfaces, one is concave and the other is convex. There is no way to distinguish between these surfaces using polarization cues. This is not unique to polarization enhancement: the convex/concave ambiguity applies to the entire surface from SfS [41] and uncalibrated PS [63].

4.1.3 Correcting for refractive distortion

Recall that estimation of the zenith angle requires knowledge of the refractive index. For materials within the dielectric range, deviation in the estimated zenith angle is only a minor source of error (Fig. 6). However, for non-dielectrics, the zenith angle surface normal will be distorted, which when integrated, causes distortions to the 3D shape.

To undistort the zenith angle, we first find the regions of the depth map that provide a good estimate of the coarse

face configurations due only to the degrees of freedom of the facets. By constraining the rest of the surface using the anchor point as a boundary condition, this leads to 2×2^K surface configurations. Since we assume that the facet is continuous with the anchor point, the overall dimensionality is reduced to $2 \times 2^K - V$.

¹² Implementing high-frequency ambiguity correction. First, a difference image is formed of the depth normals and polarization normals. The difference image only contains detailed features, as it would not show up in the former (otherwise the method from Section 4.1.1 would have been sufficient). From the difference image, the pixel at an edge from 0 to a non-zero value represents an anchor point. An edge corresponding to a change in sign (e.g., at the V-groove of a corner) is a pivot point. A greedy approach is used to flip surface facets, which enforces a closed surface constraint. Details about the closed surface constraint can be found in [34].

object shape. Specifically, we define a binary mask as

$$\mathbf{M}_{x,y} = 1 \text{ if } |\nabla^T \mathbf{N}_{x,y}^{\text{depth}}| \leq \varepsilon \text{ and } |\nabla^T \mathbf{N}_{x,y}^{\text{corr}}| \leq \varepsilon, \quad \mathbf{M}_{x,y} = 0 \text{ o.w.,} \quad (8)$$

where ε is a smoothness threshold. Note that the depth map has been smoothed, such that it is not noisy, but lacks fine detail. Intuitively, Equation 8 is saying that if the depth map has low divergence (the depth map predicts no detail) and the polarization data has low divergence (it is either noisy or predicts no detail), then we should use the depth normal. For the corner in Fig. 4, observe that the sharp point of the corner—where the Kinect data is inaccurate due to multipath—has a mask value of 0 since the divergence in \mathbf{N}^{corr} is high.

Let $\theta^{\text{depth}} \in \mathbb{R}^{M \times N}$ and $\theta^{\text{corr}} \in \mathbb{R}^{M \times N}$ denote the zenith components of $\mathbf{N}^{\text{depth}}$ and \mathbf{N}^{corr} from section 4.1.1. Over the pixel grid, we use a patch based optimization. A patch is a rectangular group of P pixel coordinates with associated, scalar angles from depth, $\{\theta_{x_1,y_1}^{\text{depth}}, \dots, \theta_{x_P,y_P}^{\text{depth}}\}$ and polarization, $\{\theta_{x_1,y_1}^{\text{corr}}, \dots, \theta_{x_P,y_P}^{\text{corr}}\}$.¹³ For the p -th patch, the goal is to find a rotation operator $\hat{\mathcal{R}}_p$, such that

$$\hat{\mathcal{R}}_p = \arg \min_{\mathcal{R}} \sum_{i=1}^P \mathbf{M}_{x_i,y_i} |\theta_{x_i,y_i}^{\text{depth}} - \mathcal{R}(\theta_{x_i,y_i}^{\text{corr}})|^2. \quad (9)$$

To correct for refractive index, the normals are updated by applying the rotation operator to pixels within this patch, such that

$$\mathbf{N}_{x_i,y_i}^{\text{corr}} := \text{SPH2CART}(1, \mathcal{R}_p(\theta_{x_i,y_i}^{\text{corr}}), \varphi_{x_i,y_i}), \quad (10)$$

where φ_{x_i,y_i} denotes the azimuth angle of $\mathbf{N}_{x_i,y_i}^{\text{corr}}$, SPH2CART represents an operator that converts spherical coordinates to cartesian, and the first argument of SPH2CART is the radial distance, typically equal to 1 for surface normals. It is important to note that $\hat{\mathcal{R}}_p$ is applied on a patch-wise basis as it will be different at different patches (due to the spatially-varying nature of the problem).

4.2 Corrected normals from polarization to enhance the coarse depth map

Given the corrected normals, it is possible to integrate to obtain the 3D shape. Unfortunately, surface normal integration is known to be a challenging task due to depth discontinuities [2,67]. To recover plausible 3D shape, we develop an integration scheme that incorporates the input depth map (\mathbf{D}) and physical intuition from corrected polarization normals (\mathbf{N}^{corr}) to recover the depth coordinates of the surface $\hat{\mathbf{D}} \in \mathbb{R}^{M \times N}$.

¹³ Note: this notion of a patch operates on the pixel grid and is thus different from our convention of defining a point cloud neighborhood (cf. Section 4.1). For our paper, we use a 7x7 patch.

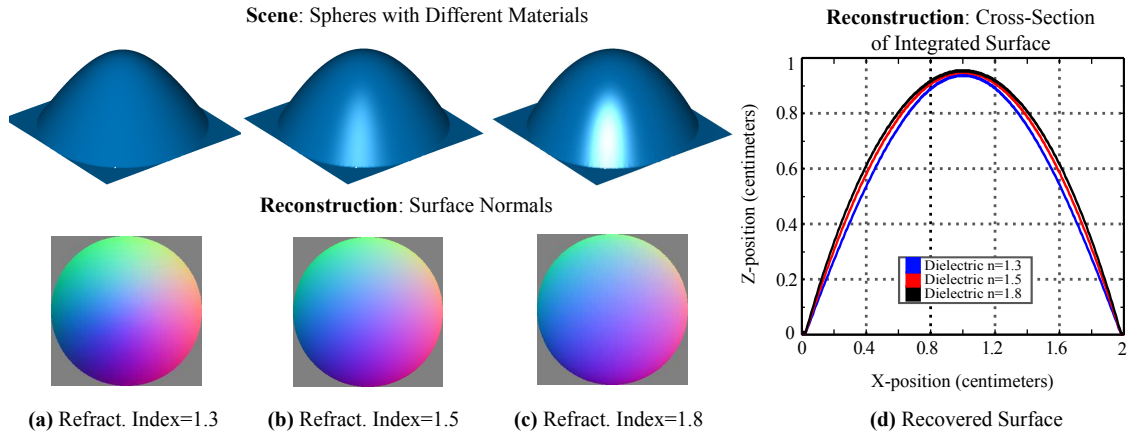


Fig. 6 Within the dielectric range ($n=1.3$ to 1.8), refractive distortion has little effect on shape reconstruction (simulated example). We simulate a scene with three spheres, each having different material properties but geometrically identical. If the refractive index is unknown—and a hard-coded threshold is used—the estimated surface normals shown in the bottom row of (a)-(c) exhibit slight distortion. When the surfaces are integrated, shown in the upper row of (a)-(c), the shape changes slightly, shown in (d).

4.2.1 Spanning tree constraint

The standard way to integrate surface normals uses the well-known Poisson equation, written as $\nabla^2 \hat{\mathbf{D}} = \nabla^T \mathbf{N}^{\text{corr}}$ for our problem. This is the optimal solution in the sense of least squares and works well when the noise model is asystematic.

For the polarization problem, the surface normals have systematic error. Intuitively, it is desirable to avoid integration using unreliable surface normals (see Fig. 7). In particular, the surface can be recovered in closed form by using only the minimum spanning tree over a weighted, 2D graph (the spanning tree is found using Kruskal’s algorithm). The optimal solution is written as

$$\nabla_S^2 \hat{\mathbf{D}} = \nabla_S^T \mathbf{N}^{\text{corr}}, \quad (11)$$

where S denotes the set of gradients used in the reconstruction and ∇_S^2 and ∇_S^T represent Laplace and divergence operators computed over S . For accurate integration, the set S includes a spanning tree of the graph. Let $\mathbf{W}_{x,y}$ denote the weights of the 2D grid. To find the weights, most previous work uses either random sampling, gradient magnitudes, or constraints on integrability [2, 10].

The physics of polarization are used to motivate the selection of graph weights. Specifically, the polarization normals are considered to be noisy when the degree of polarization ρ is low.¹⁴ A low degree of polarization most commonly occurs when the zenith angle is close to zero (i.e. fronto-parallel surfaces). For the depth map, the mask operator \mathbf{M} , defined in section 4.1.3, provides a weight of confidence.

We initialize S , the set of gradients used in the integration, as the empty set. The first gradients that are added to S are those that lie on the minimum spanning tree of the

weighted graph with weights

$$\mathbf{W}_{x,y} = \rho_{x,y} \text{ if } \rho_{x,y} > \tau \text{ and } \mathbf{M}_{x,y} = 0, \quad \mathbf{W}_{x,y} = \tau \text{ otherwise,} \quad (12)$$

where τ reflects the level of confidence in the polarization vs depth normals. We then update S by using the iterative α -approach described in [2] by using \mathbf{N}^{corr} in the update process. To incorporate $\mathbf{N}^{\text{depth}}$ into the approach, we update the corrected normals using the output from the α -approach, i.e.,

$$\mathbf{N}_{x,y}^{\text{corr}} := \mathbf{N}_{x,y}^{\text{depth}} \text{ if } \mathbf{W}_{x,y} \leq \tau. \quad (13)$$

4.2.2 Depth fidelity constraint

When integrating surface normals, only a relative 3D shape up to an unknown offset and scaling is obtained. Here, the depth fidelity constraint serves to preserve the global coordinate system and enforce consistency between the integrated surface and accurate regions of the depth map. Specifically, the depth constraint takes the form of

$$\left\| \text{VEC} \left(\mathbf{M} \odot (\hat{\mathbf{D}} - \mathbf{D}) \right) \right\|_2^2, \quad (14)$$

where VEC denotes the vectorization operator and \odot represents Hadamard multiplication. Here, we have used Hadamard multiplication with the mask to enforce fidelity only where the depth map is reliable. Both the depth fidelity and spanning tree constraints are incorporated into a sparse linear system

$$\begin{bmatrix} \lambda \mathbf{M} \odot \mathbf{I} \\ \nabla_S^2 \end{bmatrix} \text{VEC}(\hat{\mathbf{D}}) = \begin{bmatrix} \lambda \text{VEC}(\mathbf{M} \odot \mathbf{D}) \\ \nabla_S^T(\mathbf{N}^{\text{corr}}) \end{bmatrix}, \quad (15)$$

¹⁴ Estimation of the sinusoidal parameters from Equation 1 becomes unstable when there is little contrast between I_{\min} and I_{\max} .

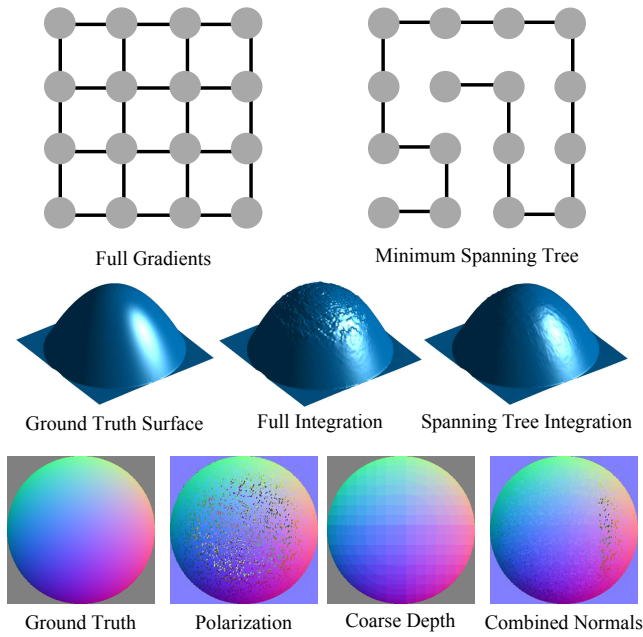


Fig. 7 Using polarization cues to find the best gradients to integrate the surface (simulated example). Represent the image points as nodes and gradients as edges. The full gradients are shown in the upper-left. We break the edges that have a low degree of polarization to find a minimum spanning tree for the grid (upper right). We validate this in simulation, shown at bottom, where the normals from polarization have error in fronto-parallel regions. Full integration results in a noisy depth map, while the spanning tree integration preserves the shape of the contour.

where \mathbf{I} is the identity matrix of size $MN \times MN$ and λ is a scalar parameter to adjust the tradeoff between spanning tree and depth fidelity constraints.¹⁵

5 Reflection and Lighting Assumptions

This section analytically validates the practicality of our technique by justifying the reflectance and lighting assumptions, first described in Section 4. Specifically, Section 5.1 includes two propositions, one for the azimuth angle, the other for the zenith angle, to explain why the proposed method will work on certain classes of mixed surfaces but not others (e.g., if a reflection is 75 percent specular and 25 percent specular, the method can work, but not if the components are equal). In Section 5.2 the unpolarized world assumption is described in more detail.

5.1 Analyzing mixed reflections

Reflections from real-world objects are mixed as they include both specular and diffuse components.¹⁶ In this paper, we are able to recover 3D geometry for mixed surfaces, subject to the constraint that the reflection is composed of an unequal mixture of diffuse and specular components. We term this the “diffuse or specular dominant” assumption, first listed in Section 4.

One of the reasons for this assumption is that when the measured irradiance from diffuse and specular reflection components is small wrt. noise, then the azimuthal angle encounters a stochastic *additional* ambiguity of $\pi/2$ radians. We refer to this error as *stochastic azimuthal ambiguity*. In comparison, when the measured irradiance from the reflection components is large wrt. noise, then the azimuthal angle predictably aligns to either diffuse or specular reflection components.

Proposition 1: *Stochastic ambiguities in the azimuthal angle are avoided when polarized reflections are either diffuse or specular dominant.*

Proof: *Consider a single scene point where the image intensity of a mixed reflection can be written as*

$$I = I_S + I_D, \quad (16)$$

where I_S is the intensity due to a specular reflection, while I_D is due to a diffuse reflection. Both specular and diffuse reflections will have polarized (denoted by $I_{\tilde{S}}$) and unpolarized components (denoted by $I_{\bar{S}}$). Then, $I_S = I_{\tilde{S}} + I_{\bar{S}}$ and $I_D = I_{\tilde{D}} + I_{\bar{D}}$. The total image intensity can now be expressed as

$$I = I_{\tilde{S}} + I_{\bar{S}} + I_{\tilde{D}} + I_{\bar{D}}. \quad (17)$$

Writing this in terms of electric field vectors,

$$\mathbf{E} = \mathbf{E}_0 + \mathbf{E}_{\tilde{D}} + \mathbf{E}_{\tilde{S}}, \quad (18)$$

where the unpolarized electric field vector \mathbf{E}_0 can have any direction. By noting that the diffuse and specular components differ in phase by a factor of $\pi/2$, we write

$$\mathbf{E} = \mathbf{E}_0 + \mathbf{E}_{\tilde{D}} \sin(\phi_{pol} - \varphi) + \mathbf{E}_{\tilde{S}} \cos(\phi_{pol} - \varphi), \quad (19)$$

where ϕ_{pol} is the polarizer angle, and φ the azimuth angle at this scene point. Since the light source is incoherent, upon taking the magnitude, we obtain

$$I = |\mathbf{E}|^2 = I_0 + I_{\tilde{D}} \sin^2(\phi_{pol} - \varphi) + I_{\tilde{S}} \cos^2(\phi_{pol} - \varphi), \quad (20)$$

¹⁵ A value of $\lambda = 0.02$ is recommended.

¹⁶ In this paper, we use the term “reflection components” following an optical imaging convention. This is identical to the term “reflectance components”, which may be more familiar to some readers.

which is rearranged as

$$I = I_0 + \frac{I_{\bar{S}} + I_{\bar{D}}}{2} + \frac{I_{\bar{S}} - I_{\bar{D}}}{2} \cos(2(\phi_{\text{pol}} - \phi)). \quad (21)$$

Clearly if the surface is specular dominant, then $I_{\bar{S}} \gg I_{\bar{D}}$ and Equation 21 can be approximated as

$$I = I_0 + \frac{I_{\bar{S}}}{2} + \frac{I_{\bar{S}}}{2} \cos(2(\phi_{\text{pol}} - \phi)). \quad (22)$$

Similarly, if the surface is diffuse dominant, then $I_{\bar{S}} \ll I_{\bar{D}}$ and Equation 21 can be approximated by neglecting $I_{\bar{S}}$. Note that either $I_{\bar{S}} \geq I_{\bar{D}}$ or $I_{\bar{S}} < I_{\bar{D}}$. Represent $\Delta = |I_{\bar{S}} - I_{\bar{D}}|$. Then, depending on whether the diffuse or specular polarized component is greater, Equation 21 can be written as

$$I = I_0 + \frac{I_{\bar{S}} + I_{\bar{D}}}{2} + \frac{\Delta}{2} \cos(2(\phi_{\text{pol}} - \phi)) \quad \text{if } I_{\bar{S}} \geq I_{\bar{D}}$$

$$I = I_0 + \frac{I_{\bar{S}} + I_{\bar{D}}}{2} + \frac{\Delta}{2} \cos(2(\phi_{\text{pol}} - (\phi - \pi/2))) \quad \text{if } I_{\bar{S}} < I_{\bar{D}}, \quad (23)$$

where now the azimuth angle has shifted by $\pi/2$ radians. Only for small values of Δ will the flip in azimuthal angle be randomly dependent on noise. Specifically, if Δ is large, the model will consistently conform to one of the two cases in Equation 23, which is the desired result. ■

The key implication of *Proposition 1* is that azimuthal angle is robust to mixed surfaces as long as the assumption of diffuse/specular dominance is met and that the correct model, either diffuse/specular, is identified.

We now seek to show that any perturbation in zenith angle as a function of mixed reflectance can be corrected for with depth constraints.

Proposition 2: Assuming the conditions of *Proposition 1* hold, perturbations in zenith angle due to mixed reflections can be corrected by applying the rotation operator $\hat{\mathcal{R}}$, as described in Equation 10.

Proof: Under the conditions of *Proposition 1*, it is assumed that the reflection component is either (1) specular dominant, or (2) diffuse dominant. First, consider case (1). Denote \hat{I}_{\min} , \hat{I}_{\max} , and $\hat{\rho}$ as the corresponding variables of I_{\min} , I_{\max} , and ρ for mixed reflections. Since, for case (1), $I = I_{\bar{S}} + I_{\bar{D}} \approx I_{\bar{S}}$, it follows from Equations 1 and 21 that:

$$\hat{I}_{\min} = I_{\bar{D}},$$

$$\hat{I}_{\max} = I_{\bar{S}},$$

$$\hat{\rho} = \frac{I_{\bar{S}} - I_{\bar{D}}}{2I_0 + I_{\bar{S}} + I_{\bar{D}}} = \frac{I_{\bar{S}} - I_{\bar{D}}}{2(I_{\bar{S}} + I_{\bar{D}}) + I_{\bar{S}} + I_{\bar{D}}}. \quad (24)$$

Now, $\hat{\rho}$ is dependent on both diffuse and specular components. Under case (1), the polarization of diffuse component

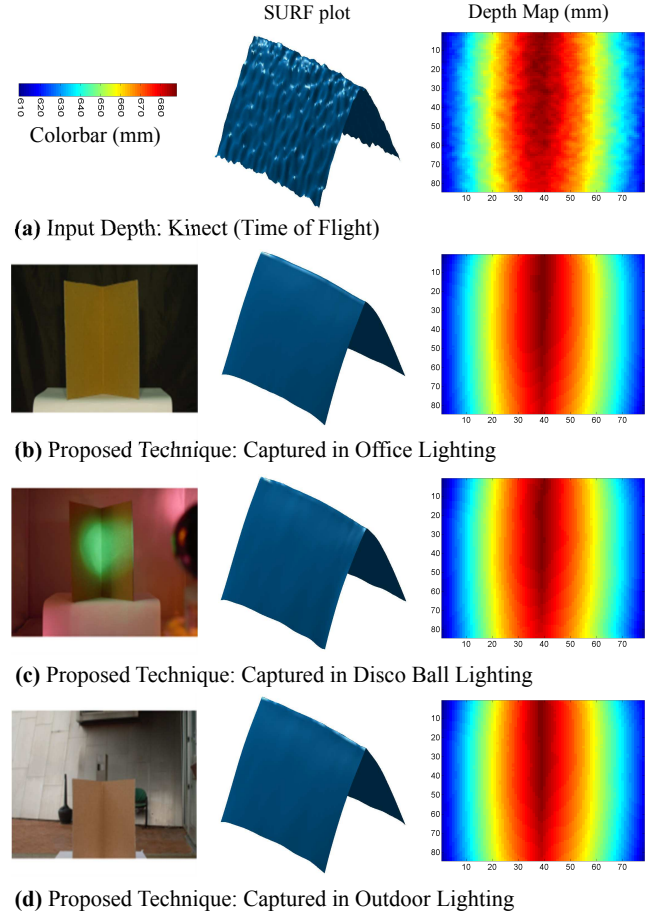


Fig. 8 Polarization enhancement works in a range of lighting conditions (real experiment). (a) ToF Kinect, due to multipath, fails to capture an accurate corner. (b) Polarization enhancement indoors. (c) Polarization enhancement under disco lighting. The disco ball casts directional uneven lighting into the corner and introduces caustic effects. (d) Polarization enhancement outdoors on a partly sunny, winter day.

is assumed to be negligible wrt. the polarization of the specular component. Therefore, the relation in Equation 16 allows us to express the measured degree of polarization, after simplification, as:

$$\hat{\rho} = \rho \frac{I - I_{\bar{D}}}{I}. \quad (25)$$

Since $\frac{I - I_{\bar{D}}}{I}$ acts as a scaling factor on the degree of polarization, and since the degree of polarization has a monotonic relationship with zenith angle, it follows that \exists a rotation operator that aligns the mixed zenith angle with the true zenith angle. For case (2), the opposite is assumed, such that the measured degree of polarization is expressed as:

$$\hat{\rho} = \rho \frac{I - I_{\bar{S}}}{I}. \quad (26)$$

Analogous to case (1), $\frac{I - I_{\bar{S}}}{I}$ acts as a scaling factor on the degree of polarization, which is the desired result. ■

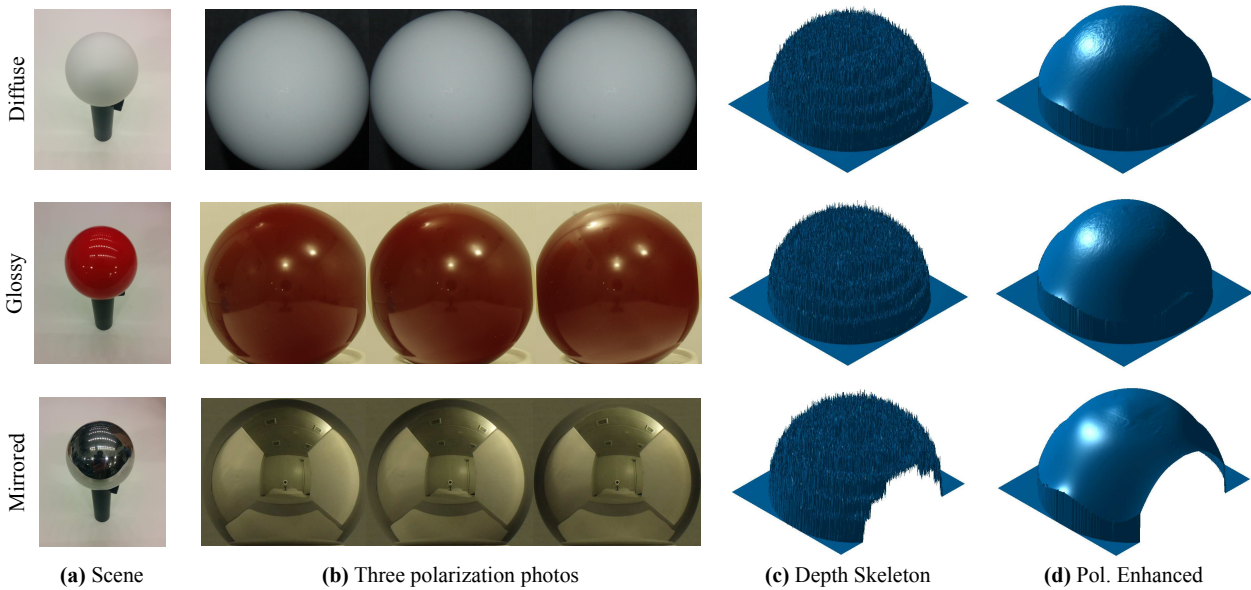


Fig. 9 Polarization enhancement works for varied material conditions. Since one of the objects is shiny, a simulated depth skeleton with additive Gaussian noise is used as the depth template; this demonstrates that our key contribution of combining a general depth map with polarization normals can work for a range of material conditions. A challenging object can preclude recovery of a coarse depth map, leading to instances like the failure case in Fig. 14.

Propositions 1 and 2 provide analytical justification for the practical results of this paper on real-world, mixed surfaces. In the case of unstable azimuth perturbation, this unique effect will only happen when diffuse and specular components are nearly equal (following from *Proposition 1*). This will not happen in many scenes, and in particular it will not happen at many scene points. In the case of zenith perturbation, *Proposition 2* shows that the perturbation amounts to a scaling on the ground truth degree of polarization. Therefore, the correction applied to the zenith angle motivated first by refractive error (i.e., rotating the zenith angle to match the depth normal) also serves to correct for the diffuse/specular perturbation in zenith angle. Nevertheless, more study is required to assess performance limits of mixed surfaces.

5.2 Unpolarized world assumption

This paper follows the standard SfP assumption that the light incident on the object is unpolarized, such that any polarization imparted to the reflected light is due to the shape. For this to hold, the light source must be unpolarized and the scene must not consist of specular interreflections.

Incident lighting is unpolarized: Natural light sources, like the sun, are unpolarized. Most types of indoor lighting are also unpolarized, like incandescent or halogen bulbs. Sometimes, solid-state light sources are used indoors, like LED lamps. However, in almost all cases, a diffusing sheet is placed over the light source, which acts to depolarize the

light. We believe that the unpolarized source assumption is valid.

No specular interreflections assumption Assuming the light source is unpolarized, after a specular reflection, the light is polarized. If this is incident on the scene point of interest, the polarimetric properties of the reflected light will not be due solely to the shape. In practice, the presence of specular interreflections is a limiting factor for most of the popular shape depth sensing technologies (e.g., structured light, time of flight, photometric stereo, shape-from-shading). We feel that this is likely the most serious limitation for SfP and, by extension, our proposed technique. For example, even though the sun is unpolarized, due to air scattering, the polarization incident on the surface-level will vary based on factors like the weather and time of day. An open problem will be how to make polarization based techniques more viable in outdoor environments.

6 Assessment and Results

Previous techniques in shading enhancement have had limited success under challenging material or lighting conditions. The proposed technique, using polarization, is able to handle more complicated scenes.

6.1 Robustness in the wild

Robustness to lighting conditions: Assuming unpolarized incident light, the proposed technique is robust to varying

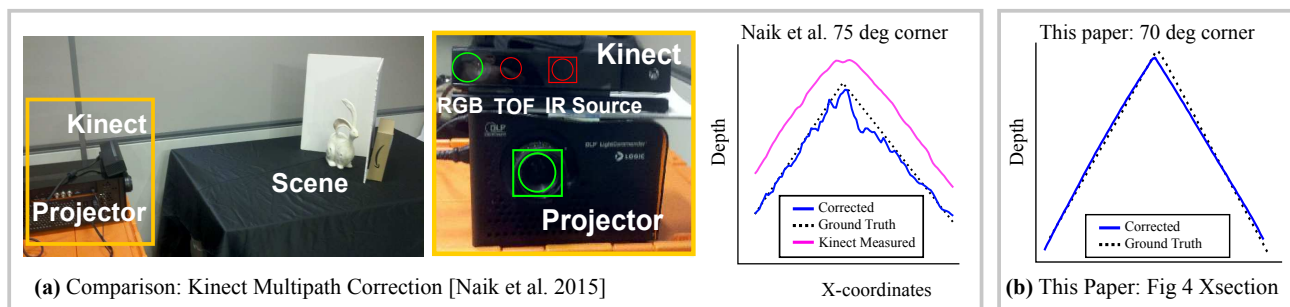


Fig. 10 The proposed technique can correct multipath interference in ToF sensors. Comparing the proposed technique against Naik et al. [36], which combines ToF with structured illumination patterns from a projector. The technique by Naik et al. uses 25 coded illumination photographs. With 3 photographs from a polarizer and the Kinect depth map, the proposed technique preserves the sharp edge of ground truth.

lighting conditions. As shown in Fig. 8, depth enhancement is shown to be near-identical for three lighting conditions: (Fig. 8b) indoor lighting; (Fig. 8c) under interfering illumination from a disco ball; and (Fig. 8d) even outdoors. The last two conditions violate lighting assumptions of SfS.

Robustness to material properties: As shown in Fig. 9 the proposed technique is evaluated on three materials: (1) diffuse; (2) glossy; and (3) mirror-like. Polarization enhancement is consistent for each material, though slightly worse for the mirror-like object. Comparison papers that use shading enhancement can only work on Lambertian surfaces [62, 16, 60, 42].

Robustness to diffuse multipath: Diffuse multipath has been an active challenge in the ToF community [43, 15, 36]. The proposed technique of polarization enhancement drastically outperforms a state-of-the-art technique for multipath correction, while using fewer images [36]. Refer to the caption of Fig. 10 for details.

6.2 Results on various scenes

Fig. 11 shows results on various scenes along with qualitative comparisons to shading refinement, directly performed by Wu et al [60].

Diffuse face scene: The mannequin scene, shown in Fig. 11a, was selected to compare the best-case performance of shading enhancement with our proposed technique of polarization enhancement. Specifically, the mannequin is coated with diffuse paint and lit by distant lighting to conform to SfS assumptions. Even under ideal conditions for shading refinement, the proposed technique using polarization leads to slightly improved 3D reconstruction. As shown in the close-up, the concave eye socket causes challenges for shading refinement due to diffuse interreflections.

Coffee cup scene: Fig. 11b shows depth reconstruction for a coffee cup made of Styrofoam. Such a surface is not Lambertian, and causes artifacts in shading refinement. The proposed technique is dramatically better than shading refinement, and as shown in Fig. 12, is able to cleanly recover the grooves (300 micron feature size). For this scene, the proposed technique outperforms a laser scan of the object (refer to Figure 12 for the comparison).

Two-face scene: To illustrate robustness to mixed-materials, Fig. 11c shows a mannequin, painted with two paints of different pigments and specularities. Shading enhancement cannot handle the shininess of the face, so the entire reconstruction is poor. Moreover, at the point of material transition, local artifacts are visible (best seen in the close-up). In comparison, the proposed technique of polarization enhancement recovers the surface well, and is robust to material change (see close-up). Note that the lack of artifacts at the point of material transition verifies that assumption 4 need not be strict (since the paints have different proportions of diffuse and specular reflectivity).

Trash can scene: Fig. 11d depicts a scene for everyday objects under natural lighting. The scene consists of a hard, plastic trash can with a shiny, plastic liner in a well-illuminated machine shop with windows. This is a challenging scene for depth enhancement, with uncontrolled lighting, mixed materials and specular objects. The proposed technique performs drastically better than shading refinement. In particular, the reconstruction from shading refinement contains holes in the recovered surface that correspond to specular highlights in the image. Furthermore, since the liner is highly specular, shading refinement cannot resolve the ridges. In comparison, the proposed technique reconstructs many of the ridges in the liner.

Enhancing a laser scanner: Fig. 13 shows that polarization enhancement can be a beneficial additional to even higher-quality depth sensors. As shown in Fig. 13c, given a laser scan of the cup, we corrupt it with quantization and additive

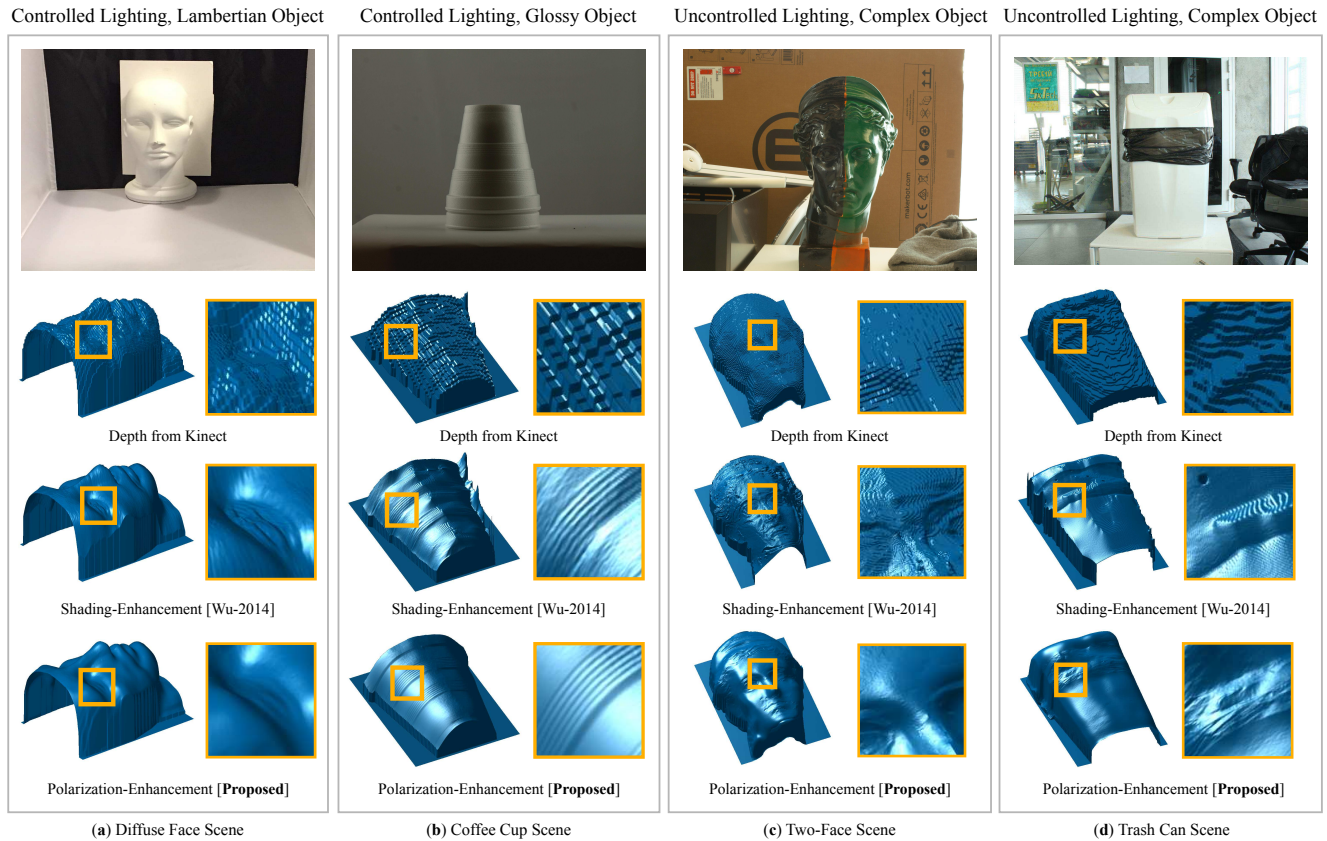


Fig. 11 Various captures, ranging from controlled scenes to complex scenes. Please zoom in using your PDF viewer.

Table 2 Mean absolute error (mm) with respect to a laser scanner.

	Init. Depth	Shading [60]	Proposed
Corner, Fig. 4	5.39	4.78	3.63
Mirror Ball, Fig. 9	8.50	17.58	8.25
Diffuse Face, Fig. 11a	18.58	18.30	18.28
Coffee Cup, Fig. 11b	3.79	3.84	3.48

noise to form the depth skeleton. All details are lost in this skeleton, but after polarization enhancement, details down to the 300 micron grooves are made visible. Adding a polarizer to a mid-range laser scanner, might be a low-cost solution to computationally replicate the performance of a higher-quality laser scanner.

Failure case: Failure cases occur in the context of all limitations detailed in the paper. We illustrate one such failure case in Fig. 14. The depth map obtained from the chrome sphere contains significant error in low-frequency content. Our “enhanced” shape tries to preserve low-frequency fidelity with the noisy depth map instead of recovering the correct shape.

6.3 Quantitative analysis of enhancement

Tab. 2 shows the mean absolute error wrt. a laser scan for a sampling of scenes from this paper. Since shading-based

techniques [60] cannot handle shiny objects like the chrome sphere or glossy coffee cup, the error actually increases wrt. the input depth. In contrast, the proposed technique of polarization reduces error for all scenes. Because polarization can handle interreflections (which the Kinect cannot), polarization shows the most improvement on the corner scene. Refer to Fig. 4 for additional metrics.

To verify the resolution enhancement of the proposed approach, we used a precision caliper to measure the grooves of the cup in Fig. 11c at 300 microns. The proposed technique can resolve finer detail than some laser scanners.

6.4 Implementation details

As shown in Fig. 3, the capture setup includes the following: a Canon Rebel T3i DSLR camera with standard Canon EF-S 18-55mm f/3.5-5.6 IS II SLR lens, a linear polarizer with quarter-wave plate, model Hoya CIR-PL. Calibration is performed on the polarizer’s transmission axis. Values for τ and ϵ are the same for all scenes. The latest model of Microsoft Kinect is used to obtain most depth maps. Normal maps and depth maps are registered using the intrinsic parameters of the Kinect and relative pose (translation only). To measure polarization cues the sensor response must be linear, enforced by preprocessing CR2 raw files from the

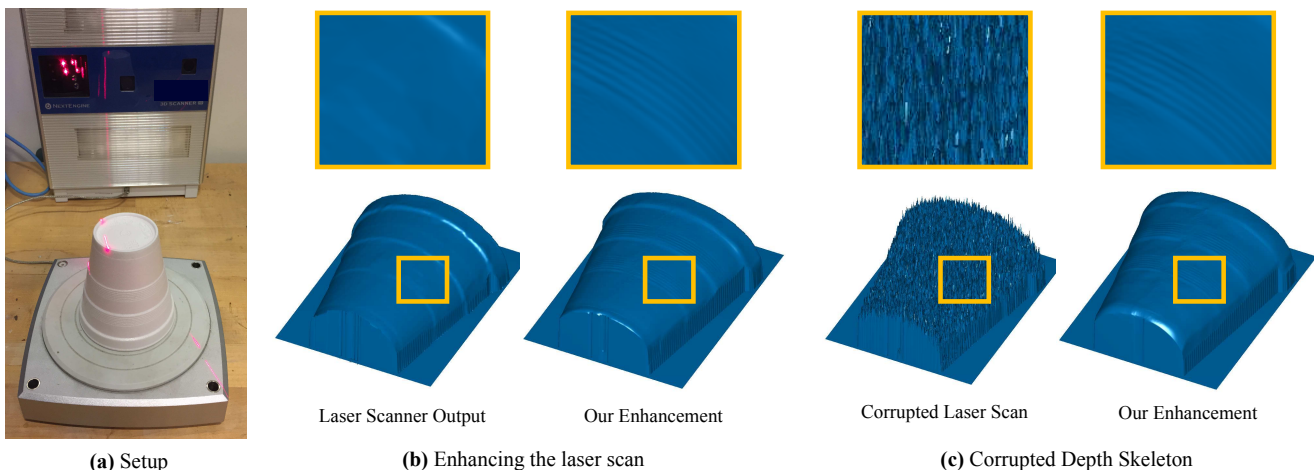


Fig. 13 Our technique can be used to augment even high quality depth maps (physical experiment). (a) The Styrofoam cup being laser scanned using multistripe technology. (b) We can combine the direct output from the laser scan with polarization cues to recover even more detail. (c) The laser scanner output is corrupted with additive Gaussian and quantization noise. Polarization cues are still able to recover the ridges on the cup. The ridges are approximately 300 microns in depth. This figure is best viewed digitally.

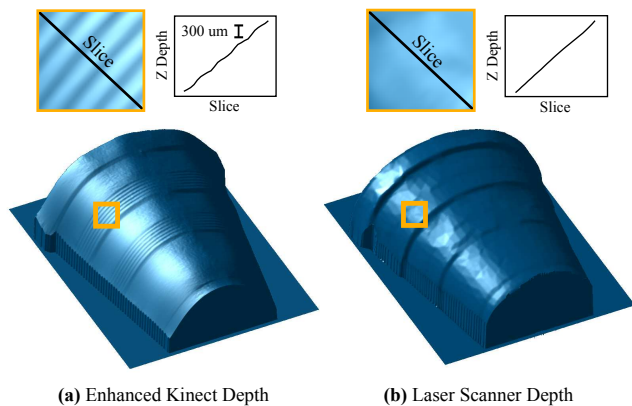


Fig. 12 Qualitative comparison to a multistripe laser scanner. (a) Our result, using a Kinect and three polarized photographs. (b) The surface from the laser scanner. Notice that the Kinect has been enhanced to the point of capturing the grooves in the cup, approximately 300 microns in depth.

camera. Ground truth is obtained using a multi-stripe, triangulation, laser scanner and benchmarks are obtained through ICP alignment.¹⁷

7 Discussion

In summary, we have proposed the first technique of depth enhancement using polarization normals. Although shading refinement is an established area, with incremental progress each year, the proposed technique leverages different physics to demonstrate complementary advantages.

Benefits: By using the depth map to place numerous constraints on the shape-from-polarization problem, this paper

resolves many of the ambiguities in shape-from-polarization research while demonstrating compelling advantages over alternative techniques (SfS and PS). In particular, SfS and PS assume Lambertian objects and distant/controlled lighting, while the proposed technique has demonstrated results on diffuse to mirror-like objects in controlled and uncontrolled settings.¹⁸ Moreover, the proposed technique can be made passive, can be implemented in a single-shot, and requires no baseline (Tab. 1). While not specific to multipath correction, the proposed technique, while using fewer images, can outperform a paper entirely dedicated to ToF multipath correction (Fig. 10).

Material assumptions Although we have presented practical results on a range of materials, the technique is still sensitive to certain material configurations. For example, if the material change is rapid in a scene (like glitter on an object), the proposed method would fail (as it relies on patches with fixed material properties). An engineering improvement would be to use higher spatial resolution sensors, as sensors continue to increase in resolution. On a more fundamental level, inspired by the work of Huynh et al. [20], it may be possible to use additional information from different spectral channels to recover the refractive index. Another area of future work would be to explore how layered or transparent materials would affect the polarimetric approach. A clear advantage of approaches that rely on polarization is that they generalize to non-Lambertian surfaces, like metals. For the proposed work to generalize to metallic surfaces, the key challenge lies in obtaining the depth of a metallic object.

¹⁸ While there are variants of generalized photometric stereo for non-Lambertian objects and natural environment lighting, they usually require about 50-100 images according to a recent survey in [52].

¹⁷ Laser Scanner: nextengine.com/assets/pdf/scanner-techspecs.pdf

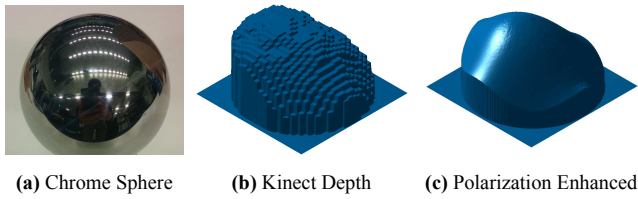


Fig. 14 A failure case due to lack of low-frequency shape information in the depth skeleton (physical experiment). (a) The scene is a chrome sphere. (b) The Kinect depth has an incorrect shape due to the specularities. (c) Our polarization enhanced algorithm takes the characteristics of the distorted depth skeleton.

Shape from mixed polarization: As of today, the community has developed models for diffuse or specular polarization (Equation 3 and 4, respectively). But a “mixed reflection” will consist of light that is both diffusely and specularly reflected. This causes model mismatch (as the model does not conform to either Equation 3 and 4). Mixed reflections have been a problem for prior work that uses only polarimetric information. However, in this paper, we have shown that the depth constraints we impose offer a degree of robustness to mixed reflections. This has been shown analytically in Section 5.1 of this paper and verified with practical results. We must emphasize that our solution is not perfect – we still require diffuse-dominant or specular-dominant surfaces (so that we can pick whether to use Equation 3 and 4 for the initial normal map from polarization).

Practical applications: The proposed work offers complementary benefits to shading or photometric solutions, offering 3D scans of objects that may be shiny or lit by area sources. Although we used a capture setup that was not real-time (since we took three photos at three different polarizer rotations), there are off-the-shelf camera systems that allow single-shot capture. With a production budget, we have little doubt that the implementation could be a 2-frame, real-time technique (the 2 frames would be a depth frame, and the single-shot frame from the polarization array camera). However, the method has value even as a static technique (as demonstrated in this research paper) where the obtained results can be seen as compelling; for example, in Figure 12, the proposed technique outperforms a scan of the coffee cup scene captured with a multi-thousand dollar professional-grade 3D scanner (that is also not real-time). Moreover, the proposed technique (depth and polarization) is a new direction to probe, which could be combined with other sources of information.

Limitations: The proposed technique requires 3 images for capturing the polarimetric information; however, off-the-shelf solutions allow single-shot polarimetric capture.¹⁹ For ro-

¹⁹ Polarization mosaic: moxtek.com/optics-product/pixelated-polarizer

bust performance, the assumptions described in Section 4 and Tab. 1 must be met. Note that some of these limitations are also present in SfS and PS contexts. For example, the proposed technique cannot handle specular interreflections, but SfS or PS methods cannot handle any interreflections, whether diffuse or specular.

Open challenges: While the proposed technique is capable of obtaining encouraging results (e.g. Fig. 11d), several scientific challenges remain, including: (1) better methods to handle mixed reflections, where the camera measures a mixture of diffuse and specular reflections at a single scene point, (2) whether there is a way to correctly resolve high-frequency detail without resorting to the closed surface heuristic (Sec. 4.1.2), and (3) alternate ways to circumvent a low degree of polarization at fronto-parallel facets (Sec. 4.2.1). Additional information, e.g., from multi-view data, circular polarization, or shading, might be a way to improve on our technique. In conclusion, we hope our practical results spur continued interest in using polarization for 3D sensing.

Acknowledgements The authors thank G. Atkinson, T. Boulton, S. Izadi, D. Miyazaki, G. Satat, N. Naik, I.K. Park, H. Zhao for valuable feedback. Achuta Kadambi is supported by the Charles Draper Doctoral Fellowship and the Qualcomm Innovation Fellowship. Boxin Shi is supported by a project commissioned by the New Energy and Industrial Technology Development Organization (NEDO). The work of the MIT-affiliated coauthors was supported by the Media Lab Consortium members.

References

1. Ackermann, J., Langguth, F., Fuhrmann, S., Goesele, M.: Photometric stereo for outdoor webcams. *Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on* (2012)
2. Agrawal, A., Raskar, R., Chellappa, R.: What is the range of surface reconstructions from a gradient field. *ECCV* (2006)
3. Atkinson, G.A., Hancock, E.R.: Multi-view surface reconstruction using polarization. *ICCV* (2005)
4. Atkinson, G.A., Hancock, E.R.: Recovery of surface orientation from diffuse polarization. *IEEE TIP* (2006)
5. Basri, R., Jacobs, D., Kemelmacher, I.: Photometric stereo with general, unknown lighting. *IJCV* (2007)
6. Chen, T., Lensch, H.P.A., Fuchs, C., Seidel, H.P.: Polarization and phase-shifting for 3d scanning of translucent objects. *CVPR* (2007)
7. Cula, O.G., Dana, K.J., Pai, D.K., Wang, D.: Polarization multiplexing and demultiplexing for appearance-based modeling. *IEEE TPAMI* (2007)
8. Dickson, C.N., Wallace, A.M., Kitchin, M., Connor, B.: Long-wave infrared polarimetric cluster-based vehicle detection. *JOSA A* **32**(12), 2307–2315 (2015)
9. Esteban, C., Vogiatzis, G., Cipolla, R.: Multiview photometric stereo. *IEEE TPAMI* (2008)
10. Fraile, R., Hancock, E.R.: Combinatorial surface integration. *ICPR* (2006)
11. Ghosh, A., Chen, T., Peers, P., Wilson, C.A., Debevec, P.: Estimating specular roughness and anisotropy from second order spherical gradient illumination. *EGSR* (2009)

12. Ghosh, A., Chen, T., Peers, P., Wilson, C.A., Debevec, P.: Circularly polarized spherical illumination reflectometry. SIGGRAPH Asia (2010)
13. Ghosh, A., Fyffe, G., Tunwattanapong, B., Busch, J., Yu, X., Debevec, P.: Multiview face capture using polarized spherical gradient illumination. SIGGRAPH Asia (2011)
14. Guarnera, G.C., Peers, P., Debevec, P., Ghosh, A.: Estimating surface normals from spherical stokes reflectance fields. ECCV Workshops (2012)
15. Gupta, M., Nayar, S.K., Hullin, M.B., Martin, J.: Phasor imaging: A generalization of correlation-based time-of-flight imaging. ACM Transactions on Graphics (TOG)
16. Han, Y., Lee, J., Kweon, I.: High quality shape from a single RGB-D image under uncalibrated natural illumination. ICCV (2013)
17. Haque, S.M., Chatterjee, A., Govindu, V.M.: High quality photometric reconstruction using a depth camera. CVPR (2014)
18. Hecht, E.: Optics, 4th. International edition, Addison-Wesley, San Francisco (2002)
19. Holroyd, M., Lawrence, J., Humphreys, G., Zickler, T.: A photometric approach for estimating normals and tangents. ACM Transactions on Graphics (TOG) (2008)
20. Huynh, C.P., Robles-Kelly, A., Hancock, E.R.: Shape and refractive index from single-view spectro-polarimetric images. International journal of computer vision **101**(1), 64–94 (2013)
21. Izadi, S., Kim, D., Hilliges, O., Molyneaux, D., Newcombe, R., Kohli, P., Shotton, J., Hodges, S., Freeman, D., Davison, A., Fitzgibbon, A.: Kinectfusion: Real-time 3D reconstruction and interaction using a moving depth camera. ACM UIST (2011)
22. Jayasuriya, S., Sivaramakrishnan, S., Chuang, E., Guruaribam, D., Wang, A., Molnar, A.: Dual light field and polarization imaging using cmos diffractive image sensors. Optics Letters **40**(10), 2433–2436 (2015)
23. Joshi, N., Kriegman, D.: Shape from varying illumination and viewpoint. ICCV (2007)
24. Kadambi, A., Boufounos, P.T.: Coded aperture compressive 3-d lidar. In: 2015 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), pp. 1166–1170. IEEE (2015)
25. Kadambi, A., Taamazyan, V., Shi, B., Raskar, R.: Polarized 3d: High-quality depth sensing with polarization cues. ICCV (2015)
26. Kadambi, A., Taamazyan, V., Shi, B., Raskar, R.: Polarized 3d: synthesis of polarization and depth cues for enhanced 3d sensing. In: SIGGRAPH 2015: Studio, p. 23. ACM (2015)
27. Klasing, K., Althoff, D., D.Wollherr, Buss, M.: Comparison of surface normal estimation methods for range sensing applications. I-CRA (2011)
28. Lanman, D., Wetzstein, G., Hirsch, M., Heidrich, W., Raskar, R.: Polarization fields: Dynamic light field display using multi-layer leds. SIGGRAPH Asia (2011)
29. 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. Eurographics (2007)
30. Manakov, A., Restrepo, J.F., Klehm, O., Hegedüs, R., Eisemann, E., Seidel, H.P., Ihrke, I.: A reconfigurable camera add-on for high dynamic range, multispectral, polarization, and light-field imaging. SIGGRAPH (2013)
31. Mitra, N.J., Nguyen, A.: Estimating surface normals in noisy point cloud data. Eurographics Symp. on Comp. Geom. (2003)
32. Miyazaki, D., Kagesawa, M., Ikeuchi, K.: Transparent surface modeling from a pair of polarization images. TPAMI (2004)
33. Miyazaki, D., Shigetomi, T., Baba, M., Furukawa, R., Hiura, S., Asada, N.: Polarization-based surface normal estimation of black specular objects from multiple viewpoints. 3DIMPVT (2012)
34. Miyazaki, D., Tan, R.T., Hara, K., Ikeuchi, K.: Polarization-based inverse rendering from a single view. ICCV (2003)
35. Morel, O., Meriaudeau, F., Stolz, C., Gorria, P.: Polarization imaging applied to 3d reconstruction of specular metallic surfaces. Electronic Imaging (2005)
36. Naik, N., Kadambi, A., Rhemann, C., Izadi, S., Raskar, R., Kang, S.B.: A light transport model for mitigating multipath interference in tof sensors. CVPR (2015)
37. Nayar, S.K., Fang, X.S., Boulton, T.: Separation of reflection components using color and polarization. IJCV (1997)
38. Nehab, D., Rusinkiewicz, S., Davis, J., Ramamoorthi, R.: Efficiently combining positions and normals for precise 3d geometry. SIGGRAPH (2005)
39. Ngo Thanh, T., Nagahara, H., Taniguchi, R.i.: Shape and light directions from shading and polarization. CVPR (2015)
40. Nießner, M., Zollhöfer, M., Izadi, S., Stamminger, M.: Real-time 3d reconstruction at scale using voxel hashing. SIGGRAPH Asia (2013)
41. Oliensis, J.: Uniqueness in shape from shading. IJCV (1991)
42. Or-el, R., Rosman, G., Wetzler, A., Kimmel, R., Bruckstein, A.: Rgb-d-fusion: Real-time high precision depth recovery. CVPR (2015)
43. O'Toole, M., Heide, F., Xiao, L., Hullin, M.B., Heidrich, W., Kutulakos, K.N.: Temporal frequency probing for 5d transient analysis of global light transport. SIGGRAPH (2014)
44. Rahmann, S., Canterakis, N.: Reconstruction of specular surfaces using polarization imaging. CVPR (2001)
45. Robles-Kelly, A., Huynh, C.P.: Imaging spectroscopy for scene analysis. Springer Science & Business Media (2012)
46. Saito, M., Sato, Y., Ikeuchi, K., Kashiwagi, H.: Measurement of surface orientations of transparent objects using polarization in highlight. CVPR (1999)
47. Schechner, Y.Y.: Self-calibrating imaging polarimetry. IEEE ICCP (2015)
48. Schechner, Y.Y., Narasimhan, S.G., Nayar, S.K.: Instant dehazing of images using polarization. CVPR (2001)
49. Schechner, Y.Y., Nayar, S.K.: Generalized mosaicing: Polarization panorama. IEEE TPAMI (2005)
50. Shi, B., Inose, K., Matsushita, Y., Tan, P., Yeung, S.K., Ikeuchi, K.: Photometric stereo using internet images. IEEE 3DV (2014)
51. Shi, B., Tan, P., Matsushita, Y., Ikeuchi, K.: Bi-polynomial modeling of low-frequency reflectances. TPAMI (2014)
52. Shi, B., Wu, Z., Mo, Z., Duan, D., Yeung, S.K., Tan, P.: A benchmark dataset and evaluation for non-lambertian and uncalibrated photometric stereo. CVPR (2016)
53. Smith, W.A., Ramamoorthi, R., Tozza, S.: Linear depth estimation from an uncalibrated, monocular polarisation image (2016)
54. Stolz, C., Kechiche, A.Z., Aubreton, O.: Short review of polarimetric imaging based method for 3d measurements. In: SPIE Photonics Europe, pp. 98,960P–98,960P. International Society for Optics and Photonics (2016)
55. Treibitz, T., Schechner, Y.Y.: Active polarization descattering. IEEE TPAMI (2009)
56. Wallace, A.M., Liang, B., Trucco, E., Clark, J.: Improving depth image acquisition using polarized light. International Journal of Computer Vision **32**(2), 87–109 (1999)
57. Wolff, L.B.: Polarization vision: a new sensory approach to image understanding. Image and Vision computing **15**(2), 81–93 (1997)
58. Woodham, R.J.: Photometric method for determining surface orientation from multiple images. Optical engineering (1980)
59. Wu, C., Wilburn, B., Matsushita, Y., Theobalt, C.: High-quality shape from multi-view stereo and shading under general illumination. CVPR (2011)
60. Wu, C., Zollhöfer, M., Nießner, M., Stamminger, M., Izadi, S., Theobalt, C.: Real-time shading-based refinement for consumer depth cameras. SIGGRAPH Asia (2014)
61. Yeung, S.K., Wu, T.P., Tang, C.K., Chan, T.F., Osher, S.J.: Normal estimation of a transparent object using a video. TPAMI (2014)
62. Yu, L.F., Yeung, S.K., Tai, Y.W., Lin, S.: Shading-based shape refinement of RGB-D images. CVPR (2013)
63. Yuille, A., Snow, D.: Shape and albedo from multiple images using integrability. CVPR (1997)

64. Zappa, C.J., Banner, M.L., Schultz, H., Corrada-Emmanuel, A., Wolff, L.B., Yalcin, J.: Retrieval of short ocean wave slope using polarimetric imaging. *Measurement Science and Technology* **19**(5), 055,503 (2008)
65. Zhang, L., Curless, B., Hertzmann, A., Seitz, S.: Shape and motion under varying illumination: Unifying structure from motion, photometric stereo, and multi-view stereo. *ICCV* (2003)
66. Zhang, L., Hancock, E.R.: A comprehensive polarisation model for surface orientation recovery. In: *Pattern Recognition (ICPR), 2012 21st International Conference on*, pp. 3791–3794. IEEE (2012)
67. Zhang, Q., Ye, M., Yang, R., Matsushita, Y., Wilburn, B., Yu, H.: Edge-preserving photometric stereo via depth fusion. *CVPR* (2012)
68. Zhao, Y., Yi, C., Kong, S.G., Pan, Q., Cheng, Y.: *Multi-band polarization imaging and applications* (2015)
69. Zickler, T., Ramamoorthi, R., Enrique, S., Belhumeur, P.N.: Reflectance sharing: Predicting appearance from a sparse set of images of a known shape. *IEEE TPAMI* (2006)
70. Zickler, T.E., Belhumeur, P.N., Kriegman, D.J.: Helmholtz stereopsis: Exploiting reciprocity for surface reconstruction. *International Journal of Computer Vision* **49**(2-3), 215–227 (2002)