

Polarized 3D: High-Quality Depth Sensing with Polarization Cues

Achuta Kadambi^{*1}, Vage Taamazyan^{1,2}, Boxin Shi^{1,3}, and Ramesh Raskar¹

¹MIT Media Lab ²Skoltech ³SUTD

Abstract

Coarse depth maps can be enhanced by using the shape information from polarization cues. We propose a framework to combine surface normals from polarization (hereafter polarization normals) with an aligned depth map. Polarization normals have not been used for depth enhancement before. This is because polarization normals suffer from physics-based artifacts, such as azimuthal ambiguity, refractive distortion and fronto-parallel signal degradation. We propose a framework to overcome these key challenges, allowing the benefits of polarization to be used to enhance depth maps. Our results demonstrate improvement with respect to state-of-the-art 3D reconstruction techniques.

1. Introduction

Today, consumer 3D cameras produce depth maps that are often noisy and lack sufficient detail. Enhancing 3D depth maps obtained from compact sensors such as the Kinect is therefore an increasingly popular research area. 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. There are dozens of papers that combine low-quality depth maps with surface normal maps obtained from SfS or PS. Well-regarded papers include [44, 14, 43] using SfS, and [31, 15] using PS. As a complementary technique, we propose the first use of surface normals from polarization to enhance depth maps.

The shape of an object causes small changes in the polarization of reflected light, best visualized by rotating a polarizing filter in front of a digital camera. Obtaining surface normals 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 PS, 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, obtaining surface normals through polarization is not yet a mature technique. The obtained normals are drastically distorted. Specific open problems [24, 3] 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.

In this paper, we address each of these challenges 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 use the polarization normals to enhance the depth map. An overview of our approach is summarized in Figure 1.

1.1. Contributions

Conceptually, we propose the only technique that exploits normals from polarization cues to enhance the quality of a coarse depth map. We devise a physics-based framework, wherein the coarse depth map is used to resolve azimuthal ambiguity (addressing problem 1) and correct for

*achoo@mit.edu

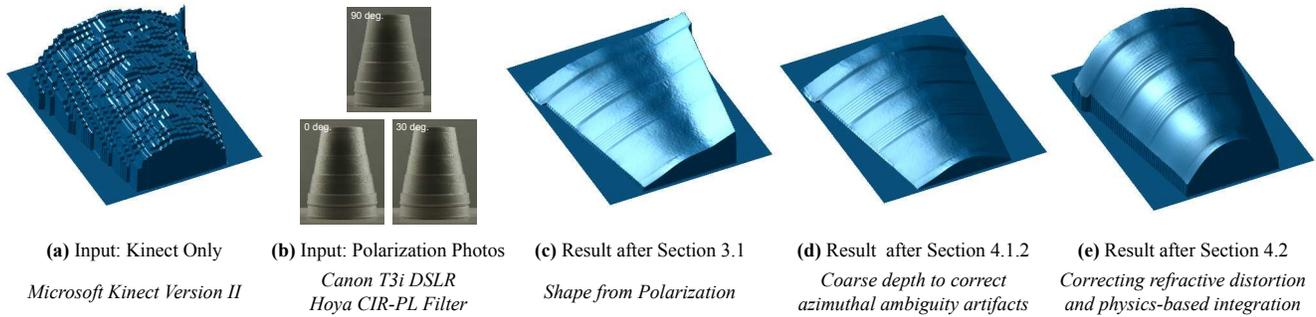


Figure 1. Outline of proposed technique. (a) The Kinect depth of an object is combined with (b) three photos at different rotations of a polarizing filter. (c) Integration of surface normals obtained from Fresnel equations. Note the azimuthal ambiguity (observed as a flip in the shape) and distortion 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..

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, the proposed technique is benchmarked against ground truth data and state-of-the-art 3D enhancement techniques [43]. The proposed technique demonstrates clear improvement on a wide variety of scenes.

2. Related Work

Shape from Polarization (SfP) estimates surface normals by analyzing the polarization properties of reflected light. An overview can be found in [36], 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 [37, 25]. On the other hand it is also possible to estimate the shape of dielectric objects using the cues from diffusely polarized reflections [27, 3]. Taken together these papers illustrate the benefit of polarization in a controlled, research setting, but—regardless of which polarization technique is used—SfP cues alone remains an ill-posed problem due to several ambiguities in shape. 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, [2] use two viewpoints to obtain polarization measurements. The work by [27] 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 [26]. In comparison, we use the additional measurement of coarse depth to sufficiently address major artifacts in classic SfP.

Combining depth and normal cues is, by now, a popular technique to obtain 3D information. Generally speaking, prior art combines a geometric-based technique to obtain rough depth with a photometric-based technique to obtain surface normals. This fusion is very 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 [31], multi-view stereo with SfS [42] or PS [47, 19, 7], consumer depth sensing with SfS [44, 14, 43], and consumer depth sensing with PS [48, 15, 40]. 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 [17] or large-scale, real-time surface reconstruction [32]. Tab. 1 summarizes the benefits and limitations of our proposed approach.

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 [22], estimation of specular roughness and anisotropy [9], inference of per-pixel surface reflectance parameters through circular polarization cues [10, 12], and for multi-view facial performance capture [11]. Polarization cues are also widely used in computational imaging applications, such as separation of diffuse and specular reflections [30, 49], dehazing of images [38], image mosaicing and panoramic stitching [39], illumination multiplexing [6] and camera [23, 18] or display hardware [21]. In addition, polarization cues can be used to recover shape of translucent objects [5], shape of the ocean surface [46], or address scattering underwater [41].

3. Basics of shape from polarization

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

Table 1. Polarization allows depth enhancement 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.

Depth Prior +X	Material Assumption	Lighting Assumption	Lighting Type	Diffuse Interreflections	Specular Interreflections	Minimum Images	Compact Sensor
Photometric Stereo [48, 15]	Lambertian	Distant Lighting	Active and Controlled	Not robust	Not robust	3	No
Shape From Shading [42, 44, 14, 43, 34]	Lambertian	Spherical Harmonic See Basri [4].	Passive	Not robust	Not robust	1 [Regular Camera]	Yes
Shape from Polarization [Proposed Technique]	Dielectrics or Low-frequency Transition	Unpolarized	Passive	Robust	Not robust	3 [Camera + Polz. Filt.] 1 [Polz. Camera]	Yes

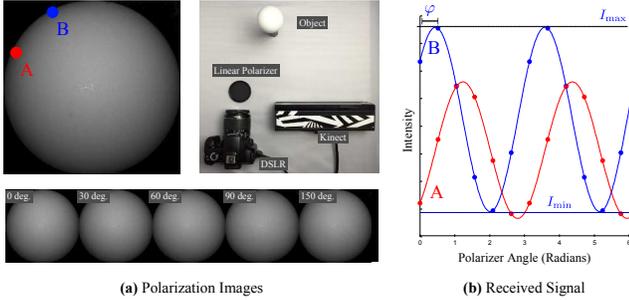


Figure 2. Capture setup. In (a) a standard camera with a polarizing filter is used to photograph a diffuse sphere under different filter rotations. The captured photographs in the bottom row look similar, but in (b), a sinusoidal pattern is observed when a single pixel is plotted against filter angle. The phase encodes azimuth angle and the amplitude and offset encode zenith angle.

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

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

where the three unknown variables in this equation are I_{max} , I_{min} , and φ , shown in Fig. 2. Sampling different values on the sinusoid amounts to taking pictures with different rotations of the polarizer angle.

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 and $\varphi + \pi$ 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 [16]) 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. Assuming the refractive index is known, the zenith angle can be estimated either in closed-form, or by numerical optimization.

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

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

where $|n^*|^2 = n^2(1 + \kappa^2)$ and κ is the attenuation index of the material, allows the zenith angle to be found [28]. 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. Variants of the method thus described are implemented in previous SfP work [2, 25, 27]. Due to the limitations of SfP (see bullets 1-5 from Section 1), SfP has never been considered as a robust alternative to SfS.

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.¹ Refer to the supplement for assumption details.

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 $\mathbf{N}^{\text{depth}} \in \mathbb{R}^{M \times N \times 3}$. The coarse depth map contains quantization errors and noise, so

¹At first glance, assumption 4 may seem limiting, however practical results are obtained on scenes with varying surface reflectivity (Fig. 9c and 9d). By analyzing the interference of polarized light, we show in the supplement that assumption 4 need not be strictly met to obtain results.

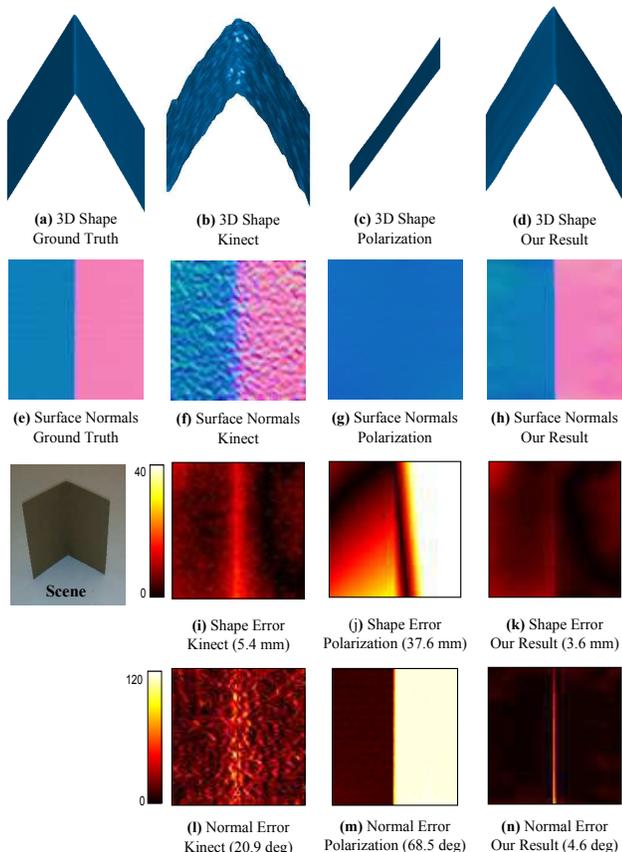


Figure 3. A commonly used benchmark scene [13, 29]. Combining polarization with Kinect results in improved performance. The top row shows the 3D shape of a corner. The second row shows the surface normals. The third row plots the estimated surface error in millimeters and the fourth row depicts the estimated angular error of surface normals in degrees w.r.t. the ground truth.

a robust method such as [24, 20] should be used to obtain normals. Specifically, we choose the plane principal component analysis technique introduced in [20] for its robustness (see supplement for technical details).

4.1.1 Removing low-frequency azimuthal ambiguity

Consider the corner scene in Fig. 3. Using a coarse depth sensor, a low-frequency version of the surface is acquired (note the smoothness in the 3D shape in Fig. 3b). 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.

Let $\mathbf{N}^{\text{polar}}$ denote the normal map obtained from polarization cues. 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 $\hat{\mathcal{A}} = \arg \min_{\mathcal{A}} \|\mathbf{N}^{\text{depth}} - \mathcal{A}(\mathbf{N}^{\text{polar}})\|_2^2$. 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, linear operator. The two states

correspond to rotating the azimuth angle by π , or not. Since the goal is to solve low-frequency ambiguity, we impose an additional constraint that \mathcal{A} is a smooth operator in the sense of total variation. Taken together, this can be expressed as a total variation minimization problem:

$$\hat{\mathcal{A}} = \arg \min_{\mathcal{A}} \|\mathbf{N}^{\text{depth}} - \mathcal{A}(\mathbf{N}^{\text{polar}})\|_2^2 + \gamma \|\nabla \mathcal{A}\|_1 \quad (5)$$

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

where the parameter γ controls the (piecewise) smoothness of the solution. Many well-known solvers exist to solve this optimization program. Since the decision variable is binary, we use graph-cuts, which is often used to segment an image into foreground and background patches. After obtaining $\hat{\mathcal{A}}$ we can correct low-frequency changes in the ambiguity by applying the operator to the polarization normal:

$$\mathbf{N}^{\text{corr}} = \hat{\mathcal{A}}(\mathbf{N}^{\text{polar}}). \quad (6)$$

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. 3g to the correctly flipped normals in Fig. 3h. 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 high-frequency ambiguity. To address this challenge we force these regions of the surface to be closed.

Fig. 4a 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 a fronto-parallel part of the surface, i.e., where the zenith angle is close to zero.

Given the anchor and pivot points, we define a facet as the set of points between the anchor and pivot points (see Fig. 4b). 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. This surface has two facets and two closed 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

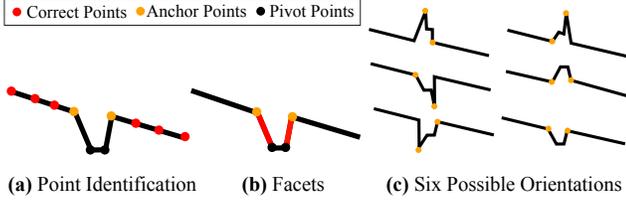


Figure 4. Addressing high-frequency ambiguity. Consider a planar surface with a high-frequency pit. (a) Anchor and pivot points are identified to group points on the ambiguity region into (b) facets. (c) Each facet can be rotated by π radians, creating ambiguities.

the high frequency detail has such a discontinuity exactly at the anchor point. Therefore, we assume the high-frequency surface is closed.

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 [33] and uncalibrated PS [45].

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. 5). However, for non-dielectrics, the zenith angle surface normal will be distorted, which when integrated, causes distortions to the 3D shape.²

To undistort the zenith angle, we first find the regions of the depth map that provide a good estimate of the coarse object shape. Specifically, we define a binary mask as

$$\mathbf{M} = 1 \text{ if } \nabla^T \mathbf{N}^{\text{depth}} \leq \epsilon \text{ and } \nabla^T \mathbf{N}^{\text{corr}} \leq \epsilon, \quad \mathbf{M} = 0 \text{ o.w.,} \quad (7)$$

where ϵ is a smoothness threshold. Intuitively, the mask takes the value of 1 in confident regions, where the object lacks high-frequency detail (as determined by the polarization normals) and zero otherwise. For the corner in Fig. 3, observe that the sharp point of the corner—where the Kinect data is inaccurate due to multipath—is masked out since the divergence in \mathbf{N}^{corr} is high.

Let θ^{depth} and θ^{corr} denote the zenith components of $\mathbf{N}^{\text{depth}}$ and \mathbf{N}^{corr} from section 4.1.1. Within each patch, we rotate the corrected normals, i.e., $\hat{\mathcal{R}} = \arg \min_{\mathcal{R}} \left\| \mathbf{M} \odot \theta^{\text{depth}} - \mathcal{R}(\theta^{\text{corr}}) \right\|_2^2$. To correct for refractive index, the normals are updated by applying the rotation operator

$$\mathbf{N}^{\text{corr}} := \hat{\mathcal{R}}(\mathbf{N}^{\text{corr}}). \quad (8)$$

²Zenith distortion could also occur when assumption 4 is violated. In such a case, model mismatch occurs because a mixed diffuse and specular surface does not conform to either Equation 3 or 4. Because a deviation is observed *only* in the zenith angle, Equation 8 serves a dual purpose of correcting both refractive distortion and model mismatch error.

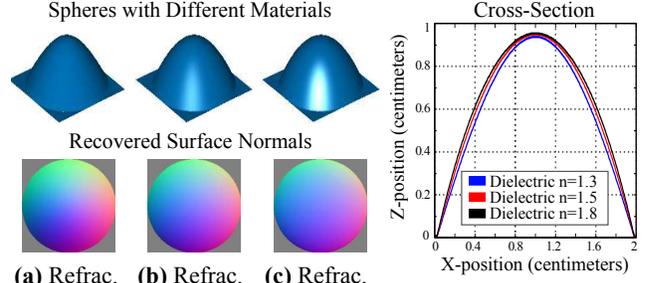


Figure 5. Within the dielectric range ($n=1.3$ to 1.8), refractive distortion has little effect on shape reconstruction (simulated example). We simulate a scene with three spheres, each having different material properties but geometrically identical. If the refractive index is unknown—and a hard-coded 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).

(a) Refrac. Index=1.3 (b) Refrac. Index=1.5 (c) Refrac. Index=1.8 (d) Recovered Surface

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 [1, 48]. To recover plausible 3D shape, we develop an integration scheme that incorporates the input depth map (D) and physical intuition from polarization (\mathbf{N}^{corr}) to recover the depth coordinates of the surface $\hat{\mathbf{D}} \in \mathbb{R}^{M \times N}$.

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. 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

$$\nabla_S^2 \hat{\mathbf{D}} = \nabla_S^T \mathbf{N}^{\text{corr}}, \quad (9)$$

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 [1, 8].

The physics of polarization are used to motivate the selection of graph weights. Specifically, the polarization nor-

mals are considered to be noisy when the degree of polarization ρ is low.³ 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

$$\mathbf{W} = \rho \text{ if } \rho > \tau \text{ and } \mathbf{M} = 0, \quad \mathbf{W} = \tau \text{ otherwise,} \quad (10)$$

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 [1], with the additional inclusion of both \mathbf{N}^{corr} and $\mathbf{N}^{\text{depth}}$ in the update process. Finally, we update the corrected normals as

$$\mathbf{N}_{x,y}^{\text{corr}} := \mathbf{N}_{x,y}^{\text{depth}} \text{ if } \mathbf{W}_{x,y} \leq \tau. \quad (11)$$

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

$$\left\| \mathbf{M} \odot (\hat{\mathbf{D}} - \mathbf{D}) \right\|_2^2, \quad (12)$$

where we have used element-wise 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

$$\begin{bmatrix} \lambda \mathbf{M} \odot \mathbf{I} \\ \nabla_S^2 \end{bmatrix} \text{VEC}(\hat{\mathbf{D}}) = \begin{bmatrix} \lambda \text{VEC}(\mathbf{M} \odot \mathbf{D}) \\ \nabla_S^T(\mathbf{N}^{\text{corr}}) \end{bmatrix}, \quad (13)$$

where VEC denotes the vectorization operator, \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. Refer to the supplement for solver details.

5. Assessment and Results

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

5.1. Robustness in the wild

Robustness to lighting conditions: Assuming unpolarized incident light, the proposed technique is robust to varying lighting conditions. As shown in Fig. 6, depth enhancement is shown to be near-identical for three lighting conditions: (Fig. 6b) indoor lighting; (Fig. 6c) under interfering

³Estimation of the sinusoidal parameters from Equation 1 becomes unstable when there is little contrast between I_{\min} and I_{\max} .

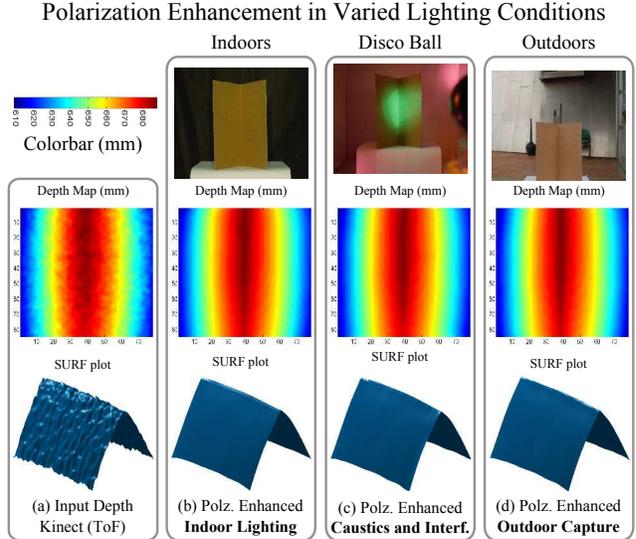


Figure 6. Polarization enhancement works in a range of lighting conditions (real experiment). (a) ToF Kinect, due to multipath, fails to capture an accurate corner. (b) Polarization enhancement indoors. (c) Polarization enhancement under disco lighting. The disco ball casts directional uneven lighting into the corner and introduces caustic effects. (d) Polarization enhancement outdoors on a partly sunny, winter day.

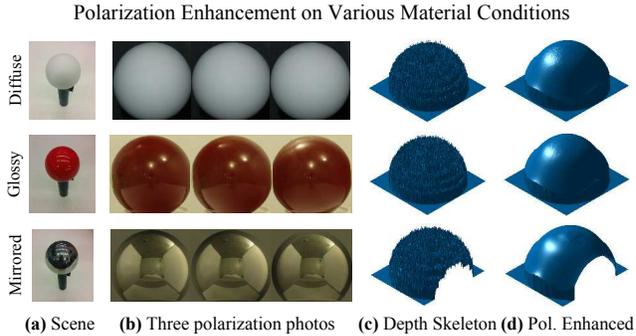


Figure 7. Polarization enhancement works for varied material conditions. A noisy depth skeleton is used as the depth template and the refinement technique recovers the sphere for all materials.

illumination from a disco ball; and (Fig. 6d) even outdoors. The last two conditions violate lighting assumptions of SfS.

Robustness to material properties: As shown in Fig. 7 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 [44, 14, 43, 34].

Robustness to diffuse multipath: Diffuse multipath has been an active challenge in the ToF community [35, 13, 29]. The proposed technique of polarization enhancement drastically outperforms a state-of-the-art technique for multipath correction, while using fewer images [29]. Refer to the cap-

tion of Fig. 8 for details.

5.2. Results on various scenes

Additional results are shown in Fig. 9, along with the qualitative comparisons to shading refinement, directly performed by Wu et al [43].

Diffuse face scene: The mannequin scene, shown in Fig. 9a, 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.

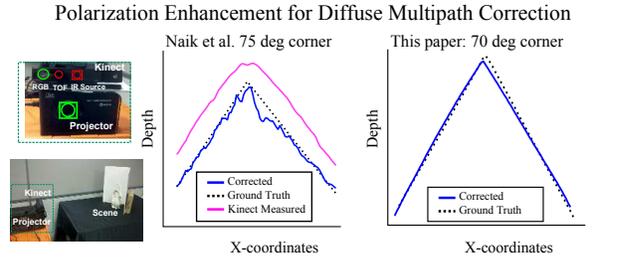
Coffee cup scene: Fig. 9b shows depth reconstruction for a coffee cup made of Styrofoam. Such a surface is not Lambertian, and causes artifacts in shading refinement. The proposed technique is dramatically better than shading refinement, and as shown in the close-ups, is able to cleanly recover the grooves (300 micron feature size). For this scene, the proposed technique outperforms a laser scan of the object (see supplement for comparison).

Two-face scene: To illustrate robustness to mixed-materials, Fig. 9c 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. 9d 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.

5.3. 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



(a) Kinect Multipath Correction Naik CVPR 2015 (b) This Paper: Slice through Fig 4

Figure 8. The proposed technique can correct multipath interference in ToF sensors. Comparing the proposed technique against Naik et al. [29], 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.

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

	Init. Depth	Shading [43]	Proposed
Corner, Fig. 3	5.39	4.78	3.63
Mirror Ball, Fig. 7	8.50	17.58	8.25
Diffuse Face, Fig. 9a	18.58	18.30	18.28
Coffee Cup, Fig. 9b	3.79	3.84	3.48

techniques [43] 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. 3 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. 9c at 300 microns. The proposed technique can resolve finer detail than some laser scanners.

5.4. Implementation details

As shown in Fig. 2, 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 camera. Ground truth is obtained using a multi-stripe, triangulation, laser scanner and benchmarks are obtained through ICP alignment.⁴ Source code, datasets and runtime details can be found on the project webpage (www.media.mit.edu/~achoo/polar3D/).

⁴Laser Scanner: nextengine.com/assets/pdf/scanner-techspecs.pdf

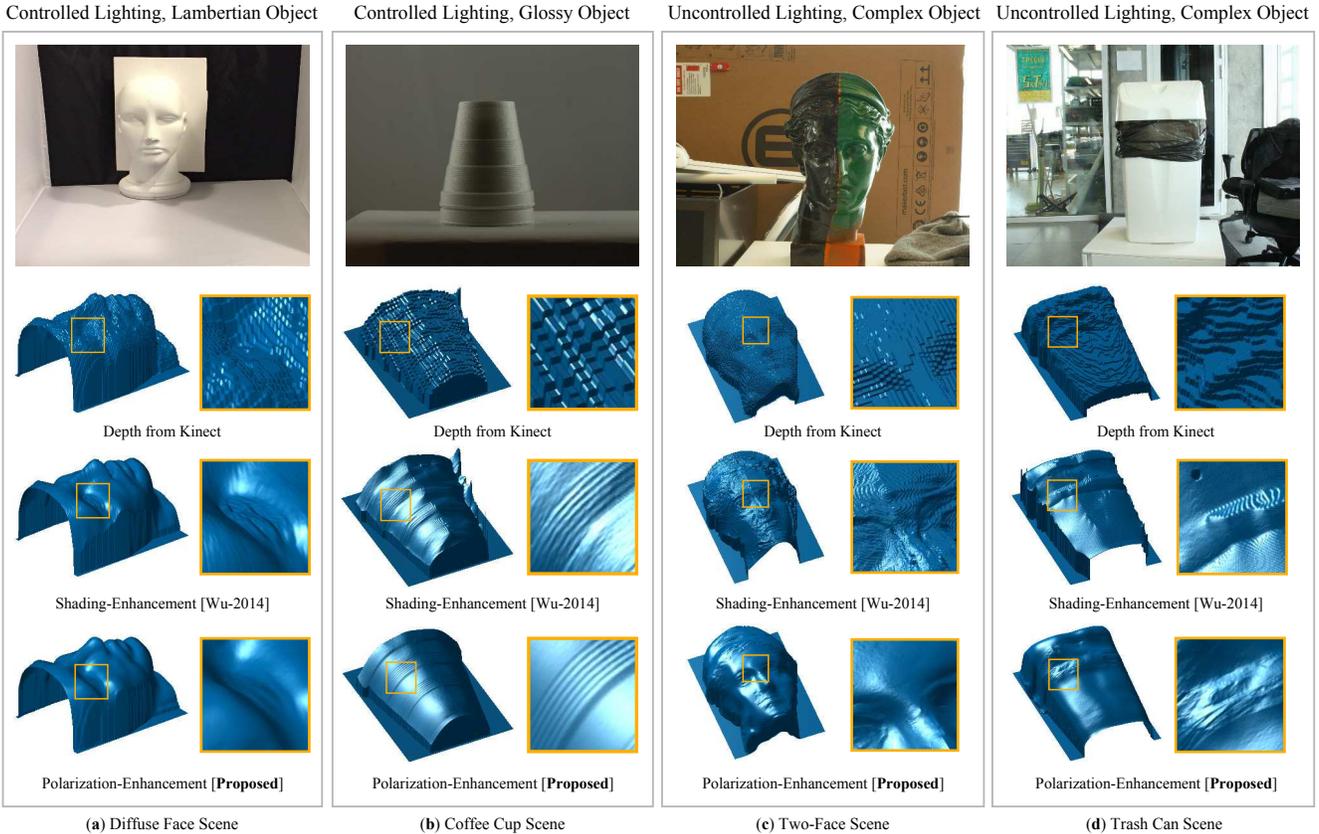


Figure 9. Various captures, ranging from controlled scenes to complex scenes. Please zoom in using PDF viewer.

6. 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 prior 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. 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. 8).

Limitations: The proposed technique requires 3 images for capture; however, off-the-shelf solutions allow single-shot capture.⁵ For robust performance, the assumptions de-

⁵Polarization mosaic: moxtek.com/optics-product/pixelated-polarizer

scribed 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. 9d), several scientific challenges remain, including: (1) better methods to compute polarization normals on scene facets containing mixtures of diffuse and specular reflections (see footnotes^{1,2}), (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 interest in using polarization for 3D sensing.

Acknowledgments: The authors thank Gary Atkinson, Terrance Boulton, Daisuke Miyazaki, Shahram Izadi, Guy Satat, Nikhil Naik, In Kyu Park, Hang Zhao and the anonymous reviewers for valuable feedback. Achuta Kadambi is supported by a Charles Draper Doctoral Fellowship. Boxin Shi is partially supported by the Singapore MOE Academic Research Fund MOE2013-T2-1-159 and the SUTD Digital Manufacturing and Design (DManD) Centre, supported by the Singapore National Research Foundation.

References

- [1] A. Agrawal, R. Raskar, and R. Chellappa. What is the range of surface reconstructions from a gradient field. *ECCV*, 2006. 5, 6
- [2] G. A. Atkinson and E. R. Hancock. Multi-view surface reconstruction using polarization. *ICCV*, 2005. 2, 3
- [3] G. A. Atkinson and E. R. Hancock. Recovery of surface orientation from diffuse polarization. *IEEE TIP*, 2006. 1, 2
- [4] R. Basri, D. Jacobs, and I. Kemelmacher. Photometric stereo with general, unknown lighting. *IJCV*, 2007. 3
- [5] T. Chen, H. P. A. Lensch, C. Fuchs, and H.-P. Seidel. Polarization and phase-shifting for 3d scanning of translucent objects. *CVPR*, 2007. 2
- [6] O. G. Cula, K. J. Dana, D. K. Pai, and D. Wang. Polarization multiplexing and demultiplexing for appearance-based modeling. *IEEE TPAMI*, 2007. 2
- [7] C. Esteban, G. Vogiatzis, and R. Cipolla. Multiview photometric stereo. *IEEE TPAMI*, 2008. 2
- [8] R. Fraile and E. R. Hancock. Combinatorial surface integration. *ICPR*, 2006. 5
- [9] A. Ghosh, T. Chen, P. Peers, C. A. Wilson, and P. Debevec. Estimating specular roughness and anisotropy from second order spherical gradient illumination. *EGSR*, 2009. 2
- [10] A. Ghosh, T. Chen, P. Peers, C. A. Wilson, and P. Debevec. Circularly polarized spherical illumination reflectometry. *SIGGRAPH Asia*, 2010. 2
- [11] A. Ghosh, G. Fyffe, B. Tunwattanapong, J. Busch, X. Yu, and P. Debevec. Multiview face capture using polarized spherical gradient illumination. *SIGGRAPH Asia*, 2011. 2
- [12] G. C. Guarnera, P. Peers, P. Debevec, and A. Ghosh. Estimating surface normals from spherical stokes reflectance fields. *ECCV Workshops*, 2012. 2
- [13] M. Gupta, S. K. Nayar, M. B. Hullin, and J. Martin. Phasor imaging: A generalization of correlation-based time-of-flight imaging. *ACM Transactions on Graphics (TOG)*. 4, 6
- [14] Y. Han, J. Lee, and I. Kweon. High quality shape from a single RGBD image under uncalibrated natural illumination. *ICCV*, 2013. 1, 2, 3, 6
- [15] S. M. Haque, A. Chatterjee, and V. M. Govindu. High quality photometric reconstruction using a depth camera. *CVPR*, 2014. 1, 2, 3
- [16] E. Hecht. Optics, 4th. *International edition, Addison-Wesley, San Francisco*, 2002. 3
- [17] S. Izadi, D. Kim, O. Hilliges, D. Molyneaux, R. Newcombe, P. Kohli, J. Shotton, S. Hodges, D. Freeman, A. Davison, and A. Fitzgibbon. Kinectfusion: Real-time 3D reconstruction and interaction using a moving depth camera. *ACM UIST*, 2011. 2
- [18] S. Jayasuriya, S. Sivaramakrishnan, E. Chuang, D. Gurusaribam, A. Wang, and A. Molnar. Dual light field and polarization imaging using cmos diffractive image sensors. *Optics Letters*, 40(10):2433–2436, 2015. 2
- [19] N. Joshi and D. Kriegman. Shape from varying illumination and viewpoint. *ICCV*, 2007. 2
- [20] K. Klasing, D. Althoff, D. Wollherr, and M. Buss. Comparison of surface normal estimation methods for range sensing applications. *ICRA*, 2011. 4
- [21] D. Lanman, G. Wetzstein, M. Hirsch, W. Heidrich, and R. Raskar. Polarization fields: Dynamic light field display using multi-layer lcds. *SIGGRAPH Asia*, 2011. 2
- [22] W.-C. Ma, T. Hawkins, P. Peers, C.-F. Chabert, M. Weiss, and P. Debevec. Rapid acquisition of specular and diffuse normal maps from polarized spherical gradient illumination. *Eurographics*, 2007. 2
- [23] A. Manakov, J. F. Restrepo, O. Klehm, R. Hegedüs, E. Eisemann, H.-P. Seidel, and I. Ihrke. A reconfigurable camera add-on for high dynamic range, multispectral, polarization, and light-field imaging. *SIGGRAPH*, 2013. 2
- [24] N. J. Mitra and A. Nguyen. Estimating surface normals in noisy point cloud data. *Eurographics Symp. on Comp. Geom.*, 2003. 1, 4
- [25] D. Miyazaki, M. Kagesawa, and K. Ikeuchi. Transparent surface modeling from a pair of polarization images. *TPAMI*, 2004. 2, 3
- [26] D. Miyazaki, T. Shigetomi, M. Baba, R. Furukawa, S. Hiura, and N. Asada. Polarization-based surface normal estimation of black specular objects from multiple viewpoints. *3DIMPVT*, 2012. 2
- [27] D. Miyazaki, R. T. Tan, K. Hara, and K. Ikeuchi. Polarization-based inverse rendering from a single view. *ICCV*, 2003. 2, 3
- [28] O. Morel, F. Meriaudeau, C. Stolz, and P. Gorria. Polarization imaging applied to 3d reconstruction of specular metallic surfaces. *Electronic Imaging*, 2005. 3
- [29] N. Naik, A. Kadambi, C. Rhemann, S. Izadi, R. Raskar, and S. B. Kang. A light transport model for mitigating multipath interference in tof sensors. *CVPR*, 2015. 4, 6, 7
- [30] S. K. Nayar, X.-S. Fang, and T. Boulton. Separation of reflection components using color and polarization. *IJCV*, 1997. 2
- [31] D. Nehab, S. Rusinkiewicz, J. Davis, and R. Ramamoorthi. Efficiently combining positions and normals for precise 3d geometry. *SIGGRAPH*, 2005. 1, 2
- [32] M. Nießner, M. Zollhöfer, S. Izadi, and M. Stamminger. Real-time 3d reconstruction at scale using voxel hashing. *SIGGRAPH Asia*, 2013. 2
- [33] J. Oliensis. Uniqueness in shape from shading. *IJCV*, 1991. 5
- [34] R. Or-el, G. Rosman, A. Wetzler, R. Kimmel, and A. Bruckstein. Rgb-d-fusion: Real-time high precision depth recovery. *CVPR*, 2015. 3, 6
- [35] M. O’Toole, F. Heide, L. Xiao, M. B. Hullin, W. Heidrich, and K. N. Kutulakos. Temporal frequency probing for 5d transient analysis of global light transport. *SIGGRAPH*, 2014. 6
- [36] S. Rahmann and N. Canterakis. Reconstruction of specular surfaces using polarization imaging. *CVPR*, 2001. 2
- [37] M. Saito, Y. Sato, K. Ikeuchi, and H. Kashiwagi. Measurement of surface orientations of transparent objects using polarization in high-light. *CVPR*, 1999. 2
- [38] Y. Y. Schechner, S. G. Narasimhan, and S. K. Nayar. Instant dehazing of images using polarization. *CVPR*, 2001. 2
- [39] Y. Y. Schechner and S. K. Nayar. Generalized mosaicing: Polarization panorama. *IEEE TPAMI*, 2005. 2
- [40] B. Shi, K. Inose, Y. Matsushita, P. Tan, S.-K. Yeung, and K. Ikeuchi. Photometric stereo using internet images. In *IEEE 3DV*, 2014. 2
- [41] T. Treibitz and Y. Y. Schechner. Active polarization descattering. *IEEE TPAMI*, 2009. 2
- [42] C. Wu, B. Wilburn, Y. Matsushita, and C. Theobalt. High-quality shape from multi-view stereo and shading under general illumination. *CVPR*, 2011. 2, 3
- [43] C. Wu, M. Zollhöfer, M. Nießner, M. Stamminger, S. Izadi, and C. Theobalt. Real-time shading-based refinement for consumer depth cameras. *SIGGRAPH Asia*, 2014. 1, 2, 3, 6, 7
- [44] L.-F. Yu, S.-K. Yeung, Y.-W. Tai, and S. Lin. Shading-based shape refinement of RGB-D images. *CVPR*, 2013. 1, 2, 3, 6
- [45] A. Yuille and D. Snow. Shape and albedo from multiple images using integrability. *CVPR*, 1997. 5
- [46] C. J. Zappa, M. L. Banner, H. Schultz, A. Corrada-Emmanuel, L. B. Wolff, and J. Yalcin. Retrieval of short ocean wave slope using polarimetric imaging. *Measurement Science and Technology*, 19(5):055503, 2008. 2
- [47] L. Zhang, B. Curless, A. Hertzmann, and S. Seitz. Shape and motion under varying illumination: Unifying structure from motion, photometric stereo, and multi-view stereo. *ICCV*, 2003. 2
- [48] Q. Zhang, M. Ye, R. Yang, Y. Matsushita, B. Wilburn, and H. Yu. Edge-preserving photometric stereo via depth fusion. *CVPR*, 2012. 2, 3, 5
- [49] T. Zickler, R. Ramamoorthi, S. Enrique, and P. N. Belhumeur. Reflectance sharing: Predicting appearance from a sparse set of images of a known shape. *IEEE TPAMI*, 2006. 2