Single-Shot Reflectance Measurement from Polarized Color Gradient Illumination

Graham Fyffe Paul Debevec
USC Institute for Creative Technologies
12015 Waterfront Drive, Los Angeles, CA 90094
fyffe@ict.usc.edu

Abstract

We present a method for acquiring the per-pixel diffuse albedo, specular albedo, and surface normal maps of a subject at a single instant in time. The method is single-shot, requiring no optical flow, and per-pixel, making no assumptions regarding albedo statistics or surface connectivity. We photograph the subject inside a spherical illumination device emitting a static lighting pattern of vertically polarized RGB color gradients aligned with the XYZ axes, and horizontally polarized RGB color gradients inversely aligned with the XYZ axes. We capture simultaneous photographs using one of two possible setups: a single-view setup using a coaxially aligned camera pair with a polarizing beam splitter, and a multi-view stereo setup with different orientations of linear polarizing filters placed on the cameras, enabling high-quality geometry reconstruction. From this lighting we derive full-color diffuse albedo, single-channel specular albedo suitable for dielectric materials, and polarization-preserving surface normals which are free of corruption from subsurface scattering. We provide simple formulae to estimate the diffuse albedo, specular albedo, and surface normal maps in the single-view and multi-view cases and show error bounds which are small for many common subjects including faces.

1. Introduction

As the physical and digital worlds converge, there is an increasing need for creating digital models of people and objects. A digital model typically consists of geometric information, indicating the shape of the object's surfaces, and reflectance information, indicating how each part of the object reflects light. Acquiring the geometry of an object can be done in many ways, from laser scanning, to structured light, to passive photogrammetry, the latter of which can be performed from photos all shot at the same instant. Acquiring the reflectance of an object – the coloration of each surface, which parts are diffuse, which parts are shiny, and high-resolution surface detail – typically involves photographing the object under a series of lighting conditions and fitting the observations to a reflectance model. With the reflectance captured, the digital model can be digitally rendered as if lit by the light of any desired environment, making the object a useful digital asset.

Single-shot scanning techniques, where the acquisition takes place at a single brief moment of time, make model acquisition more efficient and much easier to apply to dynamic subjects such as facial performances. However, since the subject is lit by just one lighting condition, not much about the object's reflectance can be captured beyond its appearance under diffuse illumination. Far more useful would be to have, at each surface point, a measurement of the subject's diffuse color, its specular component, and a high-resolution measurement of its surface normal.

In this paper, we present a novel single-shot scanning technique which uses a color polarized illumination setup to record precisely these measurements. The subject is placed in a sphere of red, green, and blue LEDS, with horizontally-oriented and vertically-oriented linear polarizers distributed throughout. With this setup, different gradient directions of light are produced on the different polarizations of the color channels, and the subject is photographed with a set of cameras some of which are polarized horizontally and some of which are polarized vertically. We leverage the fact that for dielectric materials including skin, the specular reflection component preserves the polarization, and the diffuse component depolarizes the light. These lighting conditions and cameras allow the diffuse color, specular intensity, and photometric surface orientation to be estimated at each pixel location on the object, yielding a single-shot scanning technique for both geometry and reflectance. We demonstrate the technique with two experimental setups: one using two DSLR cameras and a polarizing beam splitter to record reflectance from a single viewpoint, and a multi-view setup with a set of DSLR cameras placed around the subject with differently oriented polarization filters.

978-1-4799-8667-5/15/$31.00 ©2015 IEEE

2. Related Work

Reflectance measurement is a well-studied topic, and many methods have been proposed to obtain meaningful reflectance information for a subject within a small budget of photographs. One hugely successful method is *photometric stereo*, introduced by Woodham [14], wherein multiple observations of a subject are made from a fixed viewpoint while varying the illumination. Using as few as three incident illumination directions, the surface normal and albedo of Lambertian surfaces may be recovered. Woodham also describes the use of three colored lights to perform photometric stereo in a single photograph, which is possible under the *constant chromaticity* assumption that all scene materials vary only in brightness, and not in hue or saturation. Scenes with multi-color materials violate this assumption. Still within the realm of single-shot methods, the constant chromaticity assumption may be relaxed, so that surfaces can be categorized into a few regions of constant chromaticity [2]. Yet, subjects having subtle chromaticity variations violate this assumption, so while surface normal estimates may be accurate, the recovered surface albedo maps appear quantized. Alternatively, the constant chromaticity assumption may be lifted altogether, by introducing the assumption that optical flow can align photographs sharing common illumination in some color channels while varying in others [10], or photographs taken under complementary illumination conditions [13, 9]. Scenes with complex motion violate this assumption, such as the facial motions of speech. Optical flow may be avoided using dichroic color filters to separate the color signal into six simultaneously photographed color channels [8], using an apparatus with complementary light sources that appear white to the naked eye, however this introduces the assumption that the reflectance spectra of the scene materials exist in a low-dimensional linear basis. Our proposed approach uses a polarizing beam splitter or polarizing filters to separate the reflected light into six simultaneously photographed channels. We employ complementary color spherical gradients to encode the surface normal and specular albedo in the polarization-preserving signal. While previous methods employing color photometric stereo are adversely affected by subsurface scattering in the diffuse component, our method is unaffected by subsurface scattering as the polarization-preserving signal consists primarily of achromatic single scattering and specular reflection. Further, the diffuse albedo is encoded in the depolarizing signal, enabling polarization-based diffuse-specular separation in a single shot where previous methods require multiple photographs or color-space heuristics. Our approach is a form of polarization multiplexing [5] where the object is simultaneously illuminated by different polarizations, but we use area light sources and multiple color channels to enable multi-view single-shot photometric stereo from the single lighting condition.

Another category of single-shot works employs multi-view stereo reconstruction to estimate scene geometry, followed by further analysis of the corresponded photographs to estimate surface detail. Surface details may be estimated without photometric stereo, based on an assumed relationship between dark texture features and surface relief [3], but this assumption is violated by dark convex features or light concave features. A shape-from-shading approach may be employed that estimates the relationship between surface normal and observed colors using the coarse stereo base mesh [15], however the assumption that this relationship holds over sufficiently large regions of the object is violated in general scenes. We extend our proposed polarization-encoded photometric stereo approach to the multi-view case, where inexpensive linear polarizing filters are used in place of a polarizing beam splitter. Thus with a single shot we perform multi-view stereo reconstruction, photometric stereo, and polarization-based diffuse-specular separation enabling highly detailed geometry reconstruction and reflectance estimation.

3. Apparatus

Our apparatus consists of a 2.7m spherical geodesic structure outfitted with 2,040 Luxeon Rebel LEDs arranged into 680 evenly distributed clusters each having one red, one green and one blue LED, half of which are vertically polarized and half horizontally polarized. The LEDs are controlled by a driver to emit color gradient illumination that is a function of the lighting direction Θ relative to the center of the sphere, which we break into two functions: $L_v(\Theta)$ for vertically polarized LEDs, and $L_h(\Theta)$ for horizontally polarized LEDs. Writing RGB color values as (R, G, B), the illumination functions are:

$$L_v(\Theta) = \tfrac{1}{2}(1 + \Theta_x, 1 + \Theta_y, 1 + \Theta_z); \quad (1)$$

$$L_h(\Theta) = \tfrac{1}{2}(1 - \Theta_x, 1 - \Theta_y, 1 - \Theta_z). \quad (2)$$

We place subjects at the center of the sphere and photograph them using one or more pairs of cameras, each pair having one camera with vertically polarizing optics and the other with horizontally polarizing optics. We explore two types of polarizing optics: a polarizing beam splitter, allowing the two cameras to be coaxially aligned producing an image pair with no parallax; and linear polarizing filters, which do not allow for coaxial images but are inexpensive and suitable for multi-view stereo capture. We calibrate the exposure and white balance of the cameras and image processing software such that photographs of a 50% reflective diffuse grey card at the center of the sphere produce pixel values of (0.25, 0.25, 0.25) in all camera views (for both horizontally and vertically polarized cameras).

4. Lighting Model

For a given pixel, consider the 6-channel intensity vector $I = (r_v, g_v, b_v, r_h, g_h, b_h)$, where $\mathbf{i}_v = (r_v, g_v, b_v)$ are the vertically polarized color intensities and $\mathbf{i}_h = (r_h, g_h, b_h)$ are the horizontally polarized color intensities. We model the observed reflectance function as a sum of three components: a depolarizing RGB multiple-scattering component $\mathbf{m}(\Theta)$, a polarization-preserving achromatic single-scattering component $t(\Theta)$, and a polarization-preserving achromatic specular component $s(\Theta)$, which is modulated by the Fresnel equations depending on the polarization axis according to the Mueller calculus [4]. Thus we have:

$$\mathbf{i}_* = \int_\Omega \mathbf{m}(\Theta) \circ \tfrac{1}{2}\ L_v(\Theta) + L_h(\Theta)\ + t(\Theta) L_*(\Theta) + s(\Theta)\ {}^{\mathsf{T}}_{*}\mathbf{R}_\phi \mathbf{M} \mathbf{R}_\psi [\mathbf{1}, L_h(\Theta) - L_v(\Theta), \mathbf{0}, \mathbf{0}]^{\mathsf{T}}\ d\omega_\Theta, \quad (3)$$

where $*$ denotes either the subscript v or h, $\circ$ is the element-wise product, $\ _h$ and $\ _v$ are the Stokes vectors $[\frac{1}{2}, \frac{1}{2}, 0, 0]^{\mathsf{T}}$ and $[\frac{1}{2}, -\frac{1}{2}, 0, 0]^{\mathsf{T}}$ for horizontal and vertical linear polarization, $\mathbf{R}_\phi$ is the Stokes rotation matrix from the surface frame to the camera frame, $\mathbf{R}_\psi$ is the Stokes rotation matrix from the light source frame to the surface frame, $\mathbf{M}$ is the Mueller reflection matrix, $\mathbf{1} = (1, 1, 1)$ and $\mathbf{0} = (0, 0, 0)$. For dielectric materials, $\mathbf{M}$ takes the form of a polarizer:

$$\mathbf{M} = \begin{bmatrix} \frac{1}{2}(p_\perp^2 + p_\parallel^2) & \frac{1}{2}(p_\perp^2 - p_\parallel^2) & 0 & 0 \\ \frac{1}{2}(p_\perp^2 - p_\parallel^2) & \frac{1}{2}(p_\perp^2 + p_\parallel^2) & 0 & 0 \\ 0 & 0 & p_\perp p_\parallel & 0 \\ 0 & 0 & 0 & p_\perp p_\parallel \end{bmatrix}, \quad (4)$$

where $p_\parallel$ and $p_\perp$ are the Fresnel factors:

$$p_\parallel = \frac{\cos\theta_t - \eta\cos\theta_i}{\cos\theta_t + \eta\cos\theta_i}; \quad p_\perp = \frac{\cos\theta_i - \eta\cos\theta_t}{\cos\theta_i + \eta\cos\theta_t}, \quad (5)$$

with θ_i the angle of incidence and θ_t the angle of transmission (related by $\sin\theta_i = \eta\sin\theta_t$), and η the relative index of refraction of the surface material in air. The rotation matrix $\mathbf{R}_\phi$ (and likewise $\mathbf{R}_\psi$) has the form:

$$\mathbf{R}_\phi = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & \cos 2\phi & -\sin 2\phi & 0 \\ 0 & \sin 2\phi & \cos 2\phi & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}. \quad (6)$$

Substituting $L_v(\Theta)$ and $L_h(\Theta)$ with (1) and (2) and simplifying, we consider the following quantities:

$$\mathbf{i}_v + \mathbf{i}_h = \int_\Omega \mathbf{m}(\Theta) + \ fs(\Theta) + t(\Theta)\ \mathbf{1}\ d\omega_\Theta; \quad (7)$$

$$\mathbf{i}_v - \mathbf{i}_h = \int_\Omega t(\Theta) + k_1 s(\Theta)\ \mathbf{C}\Theta + k_2 s(\Theta)\mathbf{1}\ d\omega_\Theta, \quad (8)$$

where $k_1 = \cos 2\phi \cos 2\psi f + \sin 2\phi \sin 2\psi p_\perp p_\parallel$, $k_2 = \frac{1}{2}\cos 2\phi(p_\perp^2 - p_\parallel^2)$, $f = \frac{1}{2}(p_\perp^2 + p_\parallel^2)$, and $\mathbf{C}$ is the color crosstalk matrix between illumination primaries and camera primaries, obtained via a simple calibration process. We refer to these quantities as $\mathbf{i}^+$ and $\mathbf{i}^-$. Defining the overall energy, or *albedo*, of the multiple-scattering component $\boldsymbol{\alpha}_m$, and likewise the albedos of the specular and single scattering components α_s and α_t (absorbing the Fresnel factor f into α_s), we find:

$$\mathbf{i}^+ = \boldsymbol{\alpha}_m + (\alpha_s + \alpha_t)\mathbf{1}. \quad (9)$$

In the following, we rely on the idea that $\mathbf{i}^-$ encodes Θ without knowing the surface normal in advance, which requires $k_2 \approx 0$. We note this is satisfied when $p_\perp \approx p_\parallel$, which is true at normal incidence and less true towards the Brewster's angle. It is also satisfied when ϕ is a multiple of $45°$, regardless of incidence. We also require $k_1 > 0$ and ideally $k_1 \approx f$, which is possible using the tuned polarization arrangement in [11] where $\phi = -\psi$. Henceforth we substitute $k_1 = f$ and $k_2 = 0$, and discuss the impact of this simplification later on. Thus:

$$\mathbf{i}^- \approx \int_\Omega t(\Theta) + fs(\Theta)\ \mathbf{C}\Theta d\omega_\Theta. \quad (10)$$

We presume the specular component is statistically similar (in the 0^{th} and 1^{st} statistical moment) to a Phong cosine lobe with some exponent n_s, and the single-scattering component is statistically similar to a *spherical cosine lobe* [7] with some exponent n_t. Then:

$$\mathbf{i}^- = \mathbf{C}\ \tfrac{n_s+1}{n_s+2}\alpha_s \mathbf{r}_s + \tfrac{n_t}{n_t+2}\alpha_t \mathbf{r}_t\ , \quad (11)$$

where $\mathbf{r}_s$ is the mean specular reflection direction (presumed coincident with the ideal specular reflection direction), and $\mathbf{r}_t$ is the mean single-scattering reflection direction (presumed roughly coincident to the surface normal) We may estimate the specular albedo as:

$$\hat{\alpha}_s = \ \mathbf{C}^{-1}\mathbf{i}^- \ = \ \tfrac{n_s+1}{n_s+2}\alpha_s \mathbf{r}_s + \tfrac{n_t}{n_t+2}\alpha_t \mathbf{r}_t\ , \quad (12)$$

where the error $|\hat{\alpha}_s - \alpha_s|$ is bounded by $\frac{\alpha_s}{n_s+2} + \frac{\alpha_t n_t}{n_t+2}$ via the triangle inequality. We note that many subjects (such as human faces) have $n_s \quad 1$, $n_t \approx 1$ and $\alpha_t < \alpha_s$, and hence the error is small (a few percent). We estimate the *diffuse albedo* $\boldsymbol{\alpha}_d = \boldsymbol{\alpha}_m + \alpha_t \mathbf{1}$ as:

$$\hat{\boldsymbol{\alpha}}_d = \mathbf{i}^+ - \hat{\alpha}_s \mathbf{1} = \boldsymbol{\alpha}_d + (\alpha_s - \hat{\alpha}_s)\mathbf{1}, \quad (13)$$

noting that the error in each diffuse albedo color channel is the same as the error in the specular albedo. Finally we estimate the surface normal via:

$$\hat{\mathbf{n}} \propto \mathbf{C}^{-1}\mathbf{i}^- + \hat{\alpha}_s \mathbf{v} \propto \mathbf{n} + (\mathbf{n}\cdot\mathbf{v})^{-1}\ \tfrac{1}{2}\ \tfrac{(n_s+2)\hat{\alpha}_s}{(n_s+1)\alpha_s} - 1\ \mathbf{v} + \tfrac{(n_s+2)n_t\alpha_t}{2(n_s+1)(n_t+2)\alpha_s}\mathbf{r}_t\ , \quad (14)$$

normalizing to obtain $\hat{\mathbf{n}}$, where $\mathbf{v}$ is the view vector and $\propto$ denotes proportionality. Again, for many subjects the error is small, since $n_s \gg 1$, $\alpha_s \approx \hat{\alpha}_s$ $n_t \approx 1$, $\alpha_t < \alpha_s$, and $\mathbf{r}_t \approx \mathbf{n}$. Note that as the surface normal is derived only from polarization-preserving reflectance, it is uncorrupted by variations in the diffuse color or by blurring due to subsurface scattering. Yet, the surface normal estimate does suffer from artifacts near the Brewster's angle, and where $k_1 < 0$ in (8). We simulated the technique with the full Fresnel equations for a black specular sphere. Fig. 1 shows the angular error in the surface normal estimate for various parameters. While the polarizer arrangement of [11] produces superior results for a single view, it quickly falls apart if the camera is repositioned. The polarizer arrangement of [9] has more artifacts at the top and bottom of the sphere, but remains stable as the camera moves to different viewpoints around the equator. We employ the latter arrangement, so that we may capture multi-view data.

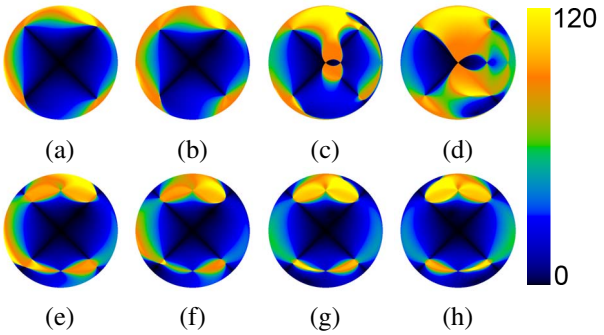


Figure 1. Simulation results for a black specular sphere, showing angular error of surface normal estimate in false color from 0° (blue-black) to 120° (orange-yellow). (a-d) Simulations using the polarizer arrangement of [11]; (e-h) Simulations using the arrangement of [9]. The first column uses $\eta = 2$. The other columns use $\eta = 1.4$. The first two columns show the frontal view. The third column has the viewpoint rotated by 30° in yaw, and the fourth column rotated by 60°. The first row improves with greater index of refraction (a), yet degrades rapidly as the viewpoint changes (c, d). The second row is stable across viewpoints, with good surface normal estimates in the central region.

4.1. Non-Coaxial Camera Pairs

We also consider the case where polarizing beam splitters are not employed, and instead each camera pair captures two independent, but close, views of the subject, with one camera having a vertical linear polarizing filter on the lens and the other horizontal. In this case, reflectance estimation is still possible if a correspondence between the images can be established, for example through stereo photogrammetry. Multiple camera pairs may be efficiently distributed around the subject, with lens polarization orientations alternating in a checkerboard fashion, so that polarization separation may be performed for any vertically or horizontally neighboring cameras (see Fig. 2). We find that

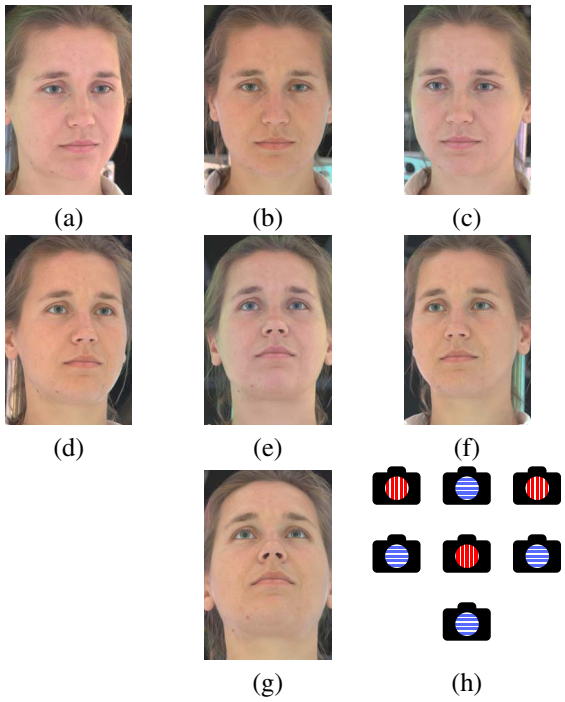


Figure 2. Photographs from our multi-view setup, with 7 cameras. Views (a, c, e) have vertical polarizing filters over the lens, while (b, d, f, g) have horizontal filters. (h) Schematic view of camera arrangement, with red vertical lines indicating vertical polarizing filters and blue horizontal lines indicating horizontal polarizing filters. The cameras are spaced approximately 7° apart. Any horizontally or vertically neighboring cameras may be used as a camera pair for reflectance estimation. More cameras could be added, expanding the checkerboard polarization pattern.

for many subjects, including human faces, standard stereo photogrammetry algorithms are able to reconstruct reasonably good geometry from photographs captured under the described conditions, despite the view dependence of the specular color gradient reflections, due to the low angular variation in the illumination. (As in [11, 9], moderately to highly specular surfaces will not be well reconstructed, as the discrete light sources will become apparent in the specular reflection.) With pixel correspondence established between views, we consider the case where the vertically and horizontally polarized pixel values may be sourced from different views, finding:

$$\mathbf{i}^+ = \boldsymbol{\alpha}_m + \tfrac{n_s+1}{2(n_s+2)}\mathbf{C}(\alpha_{sv}\mathbf{r}_{sv} - \alpha_{sh}\mathbf{r}_{sh}) + \tfrac{1}{2}(\alpha_{sv} + \alpha_{sh}) + \alpha_t \mathbf{1}; \quad (15)$$

$$\mathbf{i}^- = \mathbf{C}\ \tfrac{n_s+1}{2(n_s+2)}(\alpha_{sv}\mathbf{r}_{sv} + \alpha_{sh}\mathbf{r}_{sh}) + \tfrac{n_t}{n_t+2}\alpha_t\mathbf{r}_t + \tfrac{1}{2}(\alpha_{sv} - \alpha_{sh})\mathbf{1}, \quad (16)$$

where $\mathbf{r}_{sv}$ and $\mathbf{r}_{sh}$ are the mean specular reflection vector associated with the vertically and horizontally polarized source pixels, which differ as they are sourced from different views, and α_{sv} and α_{sh} are the specular albedos associated with the vertically and horizontally polarized source pixels, which may also differ due to the view-dependence of Fresnel reflection. We estimate the specular albedo as:

$$\hat{\alpha}_s = \frac{\mathbf{C}^{-1}\mathbf{i}^-}{\frac{1}{2}(\mathbf{v}_v + \mathbf{v}_h)} = \frac{n_s+1}{n_s+2}(2\mathbf{n}\mathbf{n}^\mathsf{T} - \mathbf{I})\frac{\alpha_{sv}\mathbf{v}_v + \alpha_{sh}\mathbf{v}_h}{2(\mathbf{v}_v + \mathbf{v}_h)} + \frac{\frac{2n_t}{n_t+2}\alpha_t\mathbf{r}_t + (\alpha_{sv} - \alpha_{sh})\mathbf{1}}{\frac{1}{2}(\mathbf{v}_v + \mathbf{v}_h)} \quad , \quad (17)$$

where $\mathbf{v}_v$ and $\mathbf{v}_h$ are the view vectors associated with the vertically and horizontally polarized source pixels, and $\mathbf{I}$ is the identity matrix. Defining $\alpha_s = \frac{1}{2}(\alpha_{sv} + \alpha_{sh})$, the error $|\hat{\alpha}_s - \alpha_s|$ is bounded by $\frac{n_s+1}{n_s+2}\left|\frac{1}{2}(\alpha_{sv}^2 + \alpha_{sh}^2) - \alpha_s\right| + \frac{2n_t\alpha_t}{n_t+2} + \sqrt{3}|\alpha_{sv} - \alpha_{sh}|\left|\frac{1}{2}(\mathbf{v}_v + \mathbf{v}_h)\right|^{-1}$, which is small as $n_s \gg 1$, $\alpha_{sv} \approx \alpha_{sh} \approx \alpha_s$, $n_t \approx 1$, and $\alpha_t < \alpha_s$. We estimate the diffuse albedo as:

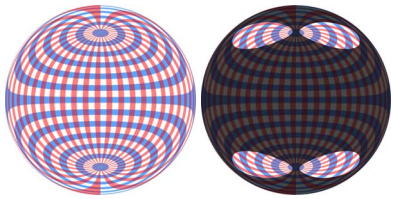
$$\begin{aligned}\hat{\boldsymbol{\alpha}}_d &= \mathbf{i}^+ - \hat{\alpha}_s\,\mathbf{1} + \tfrac{1}{2}\mathbf{C}(2\hat{\mathbf{n}}\hat{\mathbf{n}}^\mathsf{T} - \mathbf{I})(\mathbf{v}_v - \mathbf{v}_h) \\ &= \boldsymbol{\alpha}_d + (\alpha_s - \hat{\alpha}_s)\mathbf{1} \\ &\quad + \tfrac{1}{2}\mathbf{C}\left[\frac{n_s+1}{n_s+2}(2\mathbf{n}\mathbf{n}^\mathsf{T} - \mathbf{I})(\alpha_{sv}\mathbf{v}_v - \alpha_{sh}\mathbf{v}_h)\right. \\ &\quad \left. - (2\hat{\mathbf{n}}\hat{\mathbf{n}}^\mathsf{T} - \mathbf{I})\hat{\alpha}_s(\mathbf{v}_v - \mathbf{v}_h)\right] \quad , \quad (18)\end{aligned}$$

which shares the same error as the specular albedo (in each color channel), plus an additional error term that is small since $n_s \gg 1$, $\mathbf{n} \approx \hat{\mathbf{n}}$, and $\alpha_{sv} \approx \alpha_{sh} \approx \hat{\alpha}_s$. We estimate the surface normal via:

$$\begin{aligned}\hat{\mathbf{n}} &\propto \mathbf{C}^{-1}\mathbf{i}^- + \tfrac{1}{2}\hat{\alpha}_s(\mathbf{v}_v + \mathbf{v}_h) \\ &\propto \mathbf{n} + (\mathbf{n}\cdot(\tfrac{\alpha_{sv}}{\alpha_s}\mathbf{v}_v + \tfrac{\alpha_{sh}}{\alpha_s}\mathbf{v}_h))^{-1}\left[\tfrac{1}{2}\left(\tfrac{(n_s+2)\hat{\alpha}_s}{(n_s+1)\alpha_s} - \tfrac{\alpha_{sv}}{\alpha_s}\right)\mathbf{v}_v\right. \\ &\quad \left. + \tfrac{1}{2}\left(\tfrac{(n_s+2)\hat{\alpha}_s}{(n_s+1)\alpha_s} - \tfrac{\alpha_{sh}}{\alpha_s}\right)\mathbf{v}_h + \tfrac{(n_s+2)n_t\alpha_t}{(n_s+1)(n_t+2)\alpha_s}\mathbf{r}_t\right. \\ &\quad \left. + \tfrac{(n_s+2)(\alpha_{sv} - \alpha_{sh})}{2(n_s+1)\alpha_s}\mathbf{1}\right] \quad , \quad (19)\end{aligned}$$

normalizing to obtain the estimate. The error is again small, as $n_s \gg 1$, $\alpha_{sv} \approx \alpha_{sh} \approx \alpha_s \approx \hat{\alpha}_s$, $n_t \approx 1$, $\alpha_t < \alpha_s$, and $\mathbf{r}_t \approx \mathbf{n}$. If a point on the surface is associated with multiple camera pairs, we obtain multiple estimates (one per camera pair). The error of each estimate is approximately inversely proportional to the dot product term $\mathbf{n}\cdot(\frac{\alpha_{sv}}{\alpha_s}\mathbf{v}_v + \frac{\alpha_{sh}}{\alpha_s}\mathbf{v}_h)$, so we may optimally blend the estimates as a weighted sum, weighting each estimate by its associated dot product term. As the dot product term contributes most of the variation in magnitude of each unnormalized estimate, we obtain the optimal blend simply by summing the unnormalized estimates, and then normalizing the sum. We note that this scheme automatically favors data from front-facing views via the dot product term, which is common practise in multi-view texture blending. We correspondingly blend $\hat{\alpha}_s$ and $\hat{\boldsymbol{\alpha}}_d$ weighted by the magnitude of the unnormalized surface normal estimate. We also found that weighting by $\hat{\boldsymbol{\alpha}}_d$ produces good blending, which we employ in Fig. 13. Furthermore, for all quantities we modulate the weight for each view by a visibility estimate obtained via rasterized depth images, eroded and blurred by a 5-pixel radius to suppress seams due to discontinuities.

Detecting polarization reversal The polarization pattern of our multi-view setup, like that in [9], violates our assumption of good polarization separation as the surface normal points more upwards or downwards, which contributes to the artifacts in the surface normal estimates visible in *e.g.* Fig. 5. We note, however, that the polarization in these regions is approximately reversed (with $k_1 < 0$ in (8)), with vertical polarization exchanged with horizontal, as visualized in Fig. 3. Thus, if coarse geometry is available, for example from multi-view stereo, it may be utilized to distinguish reversed polarization from non-reversed polarization by computing the surface normal estimate for both cases, and retaining the one that is closer to the coarse geometric normal. We apply this strategy before blending the surface normal estimates from multiple camera pairs.



(a) (b)

Figure 3. Polarization reversal. A synthetic rendering (a) of a pattern of blue lines of latitude and red lines of longitude reflected on a mirror ball exhibits regions where the horizontal and vertical axes appear exchanged (highlighted in (b)).

Contrast with previous work We compare our method to [11, 9] who estimate diffuse and specular albedo maps using polarization difference imaging under white spherical illumination. A significant advantage of our method is that it employs only a single photograph per camera, shot simultaneously, allowing for rapid capture. While most results shown made use of a prototype apparatus with longer exposure times, we show results in Fig. 13 for a performance captured at 8 fps with an exposure time of 12.5 milliseconds per frame. In contrast, the previous works [11, 9] require 14 photographs per camera, making dynamic capture difficult. Under our lighting model, the previous works estimate $\hat{\boldsymbol{\alpha}}_d = \boldsymbol{\alpha}_m$ and $\hat{\alpha}_s = \alpha_s + \alpha_t$, with the single-scattering component being wholly included in the *specular*

albedo. (The surface normal estimates are equivalent.) It may be more advantageous to include the single-scattering component in the *diffuse* albedo, as single-scattering reflectance behaves more similarly to diffuse than to specular. Of course, if all the photographs from [11, 9] were available, one could estimate α_t as $\hat{\alpha}_t = \hat{\alpha}_s - \hat{\alpha}_s$, yielding separate estimates of all three components. An advantage of the previous works over our proposed method is that the polarization differencing of each independent gradient direction entirely avoids the artifacts in the surface normal estimates shown in Fig. 1.

5. Results and Discussion

We show reflectance estimation results using the proposed method for a variety of subjects and materials, first using the polarizing beam splitter arrangement and then using the multi-view stereo arrangement with 7 cameras and linear polarizing filters. Fig. 4 shows the two input photographs for a plastic orange subject, which has surface bumps and moderately glossy specular reflectance, along with the estimated diffuse albedo, specular albedo, and surface normal. We compare to the results from [9], which makes use of a similar arrangement of polarizing filters on the light sources, but has either only vertical (preferred) or only horizontal (non-preferred) polarizing filters in front of the camera lens. Real materials tend to specularly reflect polarized light to a greater degree along the axis parallel to the surface, reaching an extreme at the Brewster's angle where *only* light polarized along this axis is reflected. As we ignored this effect in our lighting model, we observe the artifacts predicted in Fig. 1(f). Despite using a combination of the preferred and non-preferred filters via the polarizing beam splitter, the diffuse-specular separation produced by our proposed method does not appear to suffer from any of the artifacts observed in the non-preferred arrangement in [9], though a small amount of specular pollution is visible in the four corners of the diffuse map where polarization separation is weakest, also faintly visible in the preferred arrangement in [9]. The surface normal estimates exhibit artifacts in areas with polarization reversal, where the reflection vector is essentially flipped.

Results on subjects with saturated colors Fig. 5 illustrates the robustness of the proposed surface normal estimates to strongly saturated diffuse material colors. Unlike methods that rely on Lambertian reflection, our method estimates surface normals from achromatic polarization-preserving reflectance, remaining accurate within the region of angles near the equator for all surface colors. Surface normals outside this region are biased towards the diffuse color, as the Fresnel term attenuating the diffuse reflectance differs in the two input photographs.

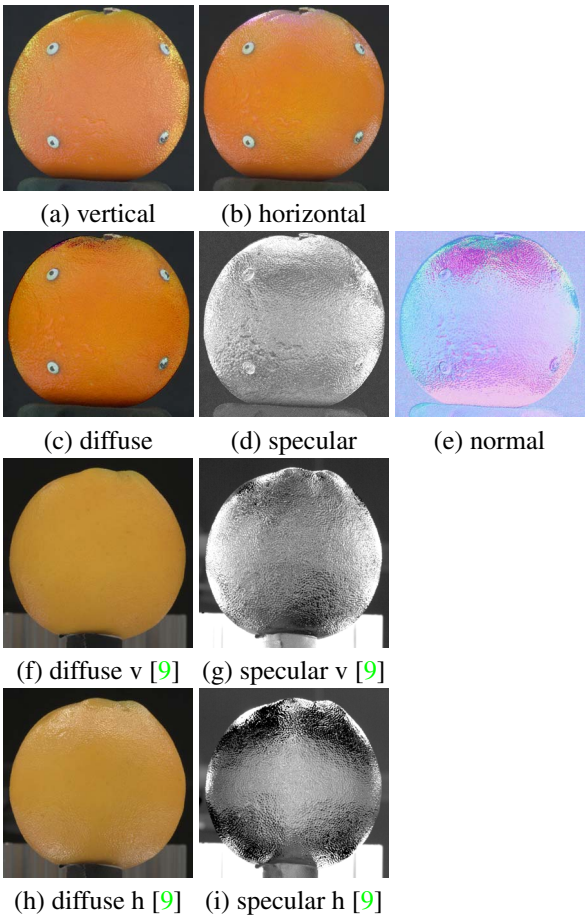
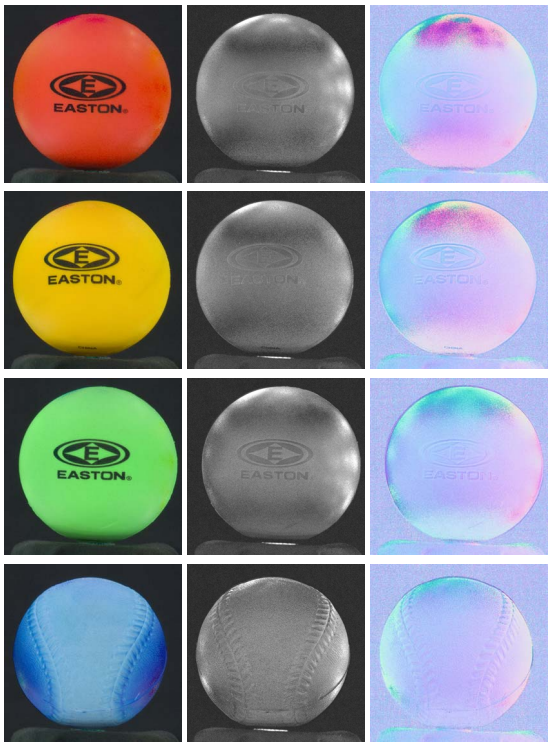


Figure 4. Reflectance estimation results using a polarizing beam-splitter for a plastic orange. (Specular albedo shown $\times 10$.) (a, b): the simultaneous photographs under differing linear polarization; (c, d, e): the estimated reflectance maps; (f, g): result reproduced from [9], showing diffuse and specular albedo for a similar plastic orange with vertical polarizing filter in front of lens; (h, i): result from [9], with horizontal polarizing filter in front of lens. The proposed method appears to produce a diffuse-specular separation (c, d) comparable in quality to the preferred method (f, g), though the surface normals suffer from artifacts in the regions where the non-preferred method (h, i) is poor.

Results on subjects with varying specularity The surface normal estimate for the green statue with rough specularity in Fig. 6 appears somewhat flattened, and interestingly does not appear to suffer from inconsistent surface normals. We suppose both of these phenomena may be explained by the polarization-preserving reflection being dominated by single scattering. There is also a low-frequency bias appearing as a vertical gradient over the image, which we attribute to view-dependence of the polarization separation behavior of the polarizing beam splitter cube. The surface normal estimate for the moderately rough specular red statue appears consistent, except for largely up-



(a) diffuse (b) specular (c) normal

Figure 5. Reflectance estimation results using a polarizing beamsplitter for spherical objects having various diffuse colors and moderately rough specularity. (Specular albedo shown $\times 5$.) As observed in Fig. 4, the surface normal estimate is consistent across materials within the central band around the equator, and becomes inconsistent above and below this region, accompanied by two dark bands in the specular albedo estimate.

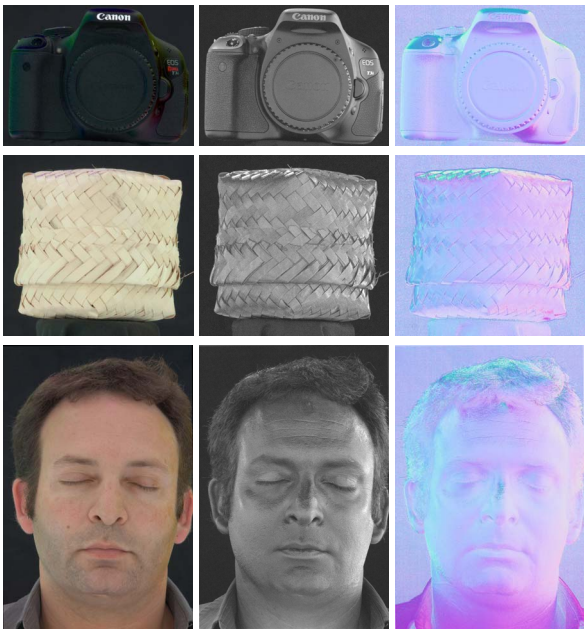
ward or downward surface normals. The surface normal estimate for the glossy specular statue is similarly consistent, but has additional noise wherever the highlights from individual lights break apart.

Results on additional subjects Fig. 7 shows results for subjects with a variety of materials. The dark plastic body of the camera is a good example illustrating the small amount of specular pollution in the estimated diffuse albedo, due to violations of the ideal polarization-preserving assumption. This pollution only occurs towards grazing angles. The woven basket illustrates that spatially-varying specular albedo is recovered even when the subject has significant diffuse reflectance. The human face result shows good diffuse-specular separation, and good surface normal estimates except for largely upward or downward pointing normals. (The purple haze towards the bottom of the surface normal image is the same defect in our polarizing beam splitter that causes the gradient haze in Fig. 6(c).)

(a) diffuse (b) specular (c) normal

Figure 6. Reflectance estimation results using a polarizing beamsplitter for objects with relief details, having materials with varying specular properties. (Specular albedo shown $\times 5$.) Top row: rough specularity produces a surface normal estimate with some flattening bias; Middle row: moderately rough specularity produces a good surface normal estimate except for upward or downward normals; Bottom row: high-gloss specularity produces a surface normal estimate broken up by discrete specular highlights.

Results with non-coaxial cameras and multi-view stereo We tested the PVMS2 software from [6] and the commercial software Agisoft PhotoScan [1] on photographs captured using our multi-view setup for four human subjects and a small statue, producing the meshes shown in Fig. 9. Despite the view-dependence of the colored specular reflections, reasonably good meshes are obtained. We then employed the Agisoft PhotoScan meshes to reproject vertical and horizontal polarized images into neighboring views, producing the reflectance estimates shown in Fig. 8. Estimates from all pairs of neighboring cameras were blended using the method in Section 4.1. The surface normal disambiguation guided by the estimated geometry improves



(a) diffuse (b) specular (c) normal

Figure 7. Reflectance estimation results using a polarizing beamsplitter for additional subjects: a camera, a woven basket, and a human face. (Specular albedo shown $\times 5$.)

the estimate in upward-pointing and downward-pointing regions. We finally refined the meshes with the surface normal estimates using a method similar to [12], producing the detailed meshes shown in Figs. 10 and 11. These meshes and reflectance maps are suitable for rendering the subject under novel views and illumination, shown in Fig. 12. Finally, we show selected frames in Fig. 13 from a short performance captured at 8 fps with an exposure time of 12.5 milliseconds per frame, using six cameras. Diffuse albedo, specular albedo, and surface normal are all estimated at a single point in time, without any optical flow.

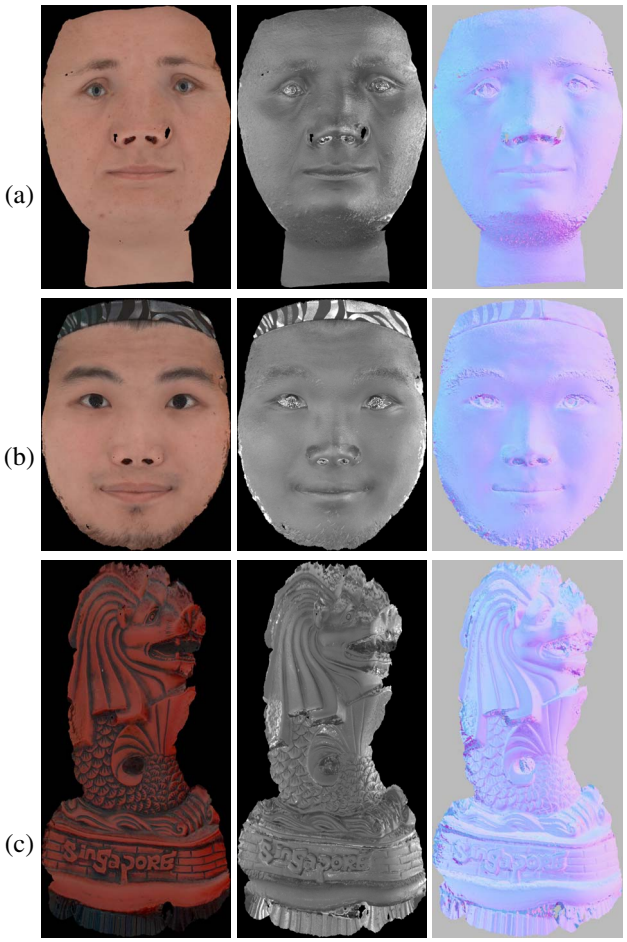


Figure 8. Reflectance estimates from multi-view setup, shown in unwrapped UV texture space. (Specular albedo shown $\times 5$.) Black regions in albedo maps were not visible in at least two views. Row (a) is the subject in Fig. 9(c); row (b) is the subject in Fig. 9(e); row (c) is the subject in Fig. 9(g).

6. Limitations and Future Work

Our method shares some limitations with other polarization separation methods for surface normal acquisition. If a latitude-longitude polarization arrangement is used (as in [9], and in all our examples), surface normals pointing significantly upward or downward are estimated incorrectly. This can be partially mitigated using the polarization arrangement from [11] in the single-view setup, and by detecting polarization reversals in the multi-view setup. It would be interesting to see if adding more camera viewpoints and perhaps further varying the angle of polarization on the cameras could provide better coverage of surface normal orientations. Since denser camera arrays are getting more popular for stereo reconstruction (50 to 100 cameras is not uncommon) this would be a natural extension. The surface normal estimate relies on the polarization-preserving signal in the photographs, so very diffuse materials yield noisy surface normal estimates, and the surface normal estimates are very sensitive to noise ("digital grain") in the photographs. Additionally, as our lighting apparatus uses discrete LED light sources, the specular signal breaks apart into individual highlights for high-gloss (mirror-like) surfaces. See Fig. 6 for examples. The method in [11, 9] shares similar limitations, however an alternative surface normal estimate (the so-called "diffuse normal") is available for regions with low polarization-preserving signal, and the use of multiple photographs in the process helps to mitigate noise. Nevertheless, our method is suitable for many subjects, including human faces.

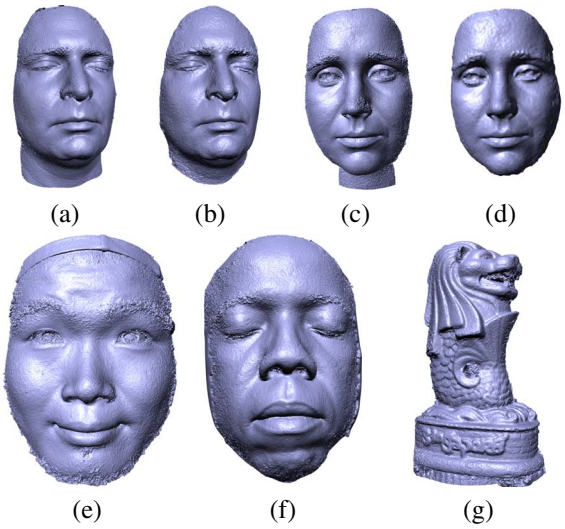


Figure 9. Reconstructed geometry using standard multi-view stereo techniques on a mix of vertically and horizontally polarized views under polarized spherical RGB gradient illumination. (a,c,e,f,g) Subjects processed using Agisoft PhotoScan. (b, d) Same subjects as (a, c) processed using PMVS2 software.

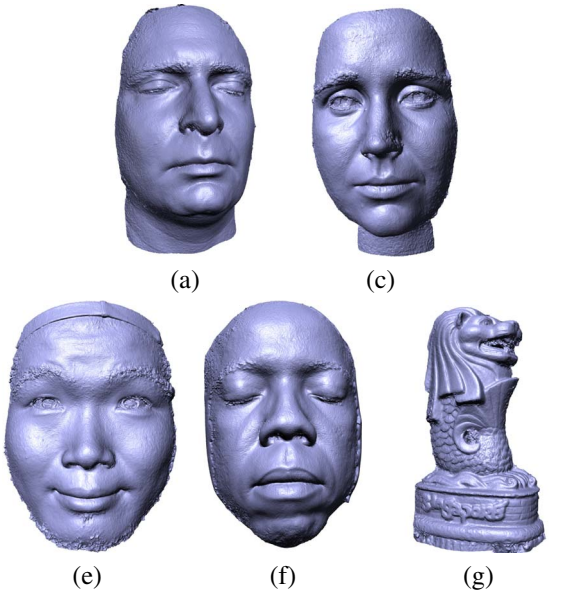


Figure 10. Refined geometry results corresponding to the base geometry in Fig. 9, using the photometric surface normal estimates in Fig. 8. Note the skin pores and lip wrinkles visible in (a,c,e,f), the fine forehead wrinkles in (e), and the scale and brick detail in (g), which were not discernible in the base geometry.

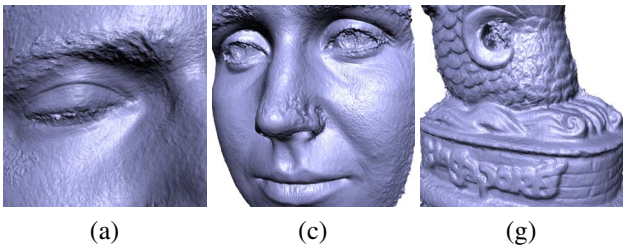


Figure 11. Zoomed views of the refined geometry in Fig. 10.

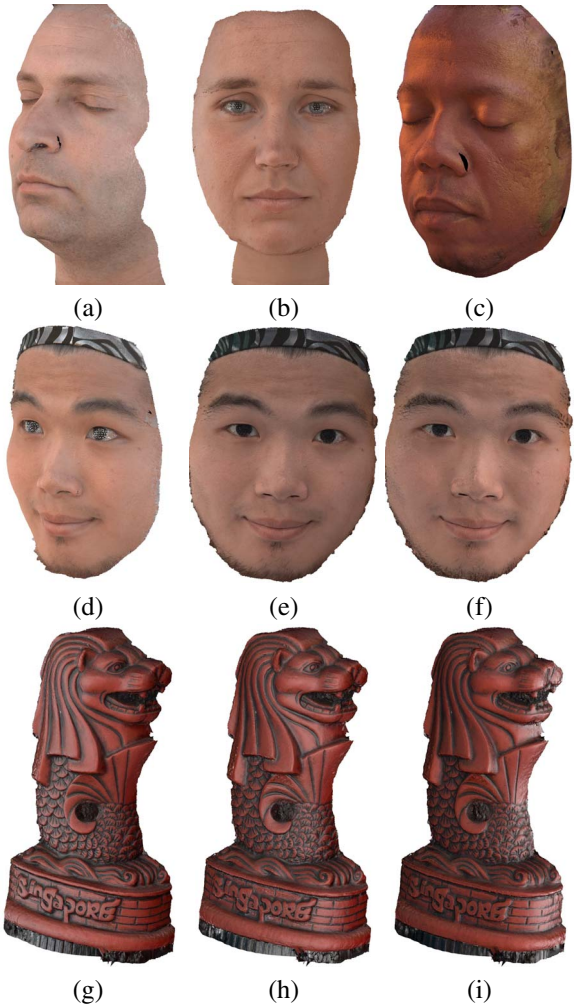


Figure 12. Renderings of recovered geometry and reflectance with (a-d) novel viewpoint and HDRI lighting environments; and (e-i) novel point-light illumination. All renderings use a modified Blinn-Phong specular lobe with exponent 64 and measured albedo.

7. Conclusions

We have presented a method for single-shot capture of geometry and reflectance. We showed that using complementary colored spherical gradients to encode direction information into the polarization-preserving signal enables polarization-based diffuse-specular separation, and simultaneous photometric surface normal estimation. We demonstrated the technique using two setups: one with a polarizing beam splitter to coaxially align two cameras,

(a) diffuse (b) specular (c) normal

Figure 13. Recovered reflectance for three frames from a short performance captured at 8 fps. (Specular albedo shown $\times 5$.)

and another with linear polarizing filters on the cameras, using multi-view stereo reconstruction to bring disparate views into correspondence. We show high-quality diffuse-specular separation and surface normal estimates with both setups, and detailed geometry reconstruction in the multi-view case, including dynamic capture. To our knowledge, this is the first technique for estimating geometry and reflectance in a single shot, without overly restrictive BRDF assumptions, material clustering heuristics, or optical flow.

Acknowledgements

The authors thank the following for their support and assistance: Ari Shapiro, Maria Mikhisor, Antionne Scott, Shanhe Wang, Koki Nagano, Xueming Yu, Jay Busch, Kathleen Haase, Bill Swartout, Cheryl Birch, Randall Hill and Randolph Hall. This work was sponsored by the University of Southern California Office of the Provost and the U.S. Army Research, Development, and Engineering Command (RDECOM). The content of the information does not necessarily reflect the position or the policy of the US Government, and no official endorsement should be inferred.

References

[1] Agisoft photoscan. http://www.agisoft.com/, 2014. 7

[2] R. Anderson, B. Stenger, and R. Cipolla. Color photometric stereo for multicolored surfaces. In D. N. Metaxas, L. Quan, A. Sanfeliu, and L. J. V. Gool, editors, *ICCV*, pages 2182–2189. IEEE, 2011. 2

[3] T. Beeler, B. Bickel, P. Beardsley, B. Sumner, and M. Gross. High-quality single-shot capture of facial geometry. *ACM Transactions on Graphics (Proc. SIGGRAPH)*, 29(4):40:1–40:9, 2010. 2

[4] E. Collett. *Field guide to polarization*. SPIE Press, 2005. 3

[5] O. G. Cula, K. J. Dana, D. K. Pai, and D. Wang. Polarization multiplexing for bidirectional imaging. In *Proc. of the 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition - vol. 2*, CVPR'05, pages 1116–1123, Washington, DC, USA, 2005. IEEE Computer Society. 2

[6] Y. Furukawa and J. Ponce. Accurate, dense, and robust multi-view stereopsis. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 32(8):1362–1376, 2010. 7

[7] G. Fyffe and P. Debevec. Cosine lobe based relighting from gradient illumination photographs. *JVRB - Journal of Virtual Reality and Broadcasting*, 9(2012)(2), 2012. 3

[8] G. Fyffe, X. Yu, and P. Debevec. Single-shot photometric stereo by spectral multiplexing. In *IEEE International Conference on Computational Photography*, 2011. 2

[9] A. Ghosh, G. Fyffe, B. Tunwattanapong, J. Busch, X. Yu, and P. Debevec. Multiview face capture using polarized spherical gradient illumination. *ACM Transactions on Graphics*, pages 1–10, 2011. 2, 4, 5, 6, 8

[10] H. Kim, B. Wilburn, and M. Ben-Ezra. Photometric stereo for dynamic surface orientations. In *Proc. of the 11th European Conference on Computer Vision: Part I*, ECCV'10, pages 59–72, Berlin, Heidelberg, 2010. Springer-Verlag. 2

[11] 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. In J. Kautz and Pattanaik, editors, *Eurographics Symposium on Rendering*, 2007. 3, 4, 5, 6, 8

[12] D. Nehab, S. Rusinkiewicz, J. Davis, and R. Ramamoorthi. Efficiently combining positions and normals for precise 3D geometry. *ACM Transactions on Graphics (Proc. of ACM SIGGRAPH 2005)*, 24(3), Aug. 2005. 8

[13] C. A. Wilson, A. Ghosh, P. Peers, J.-Y. Chiang, J. Busch, and P. Debevec. Temporal upsampling of performance geometry using photometric alignment. *ACM Transactions on Graphics*, 29(2), Mar. 2010. 2

[14] R. J. Woodham. Photometric method for determining surface orientation from multiple images. *Optical Engineering*, 19(1):191139, 1980. 2

[15] C. Wu, B. Wilburn, Y. Matsushita, and C. Theobalt. High-quality shape from multi-view stereo and shading under general illumination. In *Proc. 2011 IEEE Conf. on Computer Vision and Pattern Recognition*, CVPR'11, pages 969–976, Washington, DC, USA, 2011. IEEE Computer Society. 2