

# Shape and Reflectance Estimation in the Wild

Geoffrey Oxholm, *Member, IEEE*, Ko Nishino *Member, IEEE*

**Abstract**—Our world is full of objects with complex reflectances situated in rich illumination environments. Though stunning, the diversity of appearance that arises from this complexity is also daunting. For this reason, past work on geometry recovery has tried to frame the problem into simplistic models of reflectance (such as Lambertian, mirrored, or dichromatic) or illumination (one or more distant point light sources). In this work, we directly tackle the problem of joint reflectance and geometry estimation under known but uncontrolled natural illumination by fully exploiting the surface orientation cues that become embedded in the appearance of the object. Intuitively, salient scene features (such as the sun or stained glass windows) act analogously to the point light sources of traditional geometry estimation frameworks by strongly constraining the possible orientations of the surface patches reflecting them. By jointly estimating the reflectance of the object, which modulates the illumination, the appearance of a surface patch can be used to derive a nonparametric distribution of its possible orientations. If only a single image exists, these strongly constrained surface patches may then be used to anchor the geometry estimation and give context to the less-descriptive regions. When multiple images exist, the distribution of possible surface orientations becomes tighter as additional context is given, though integrating the separate views poses additional challenges. In this paper we introduce two methods, one for the single image case, and another for the case of multiple images. The effectiveness of our methods is evaluated extensively on synthetic and real-world data sets that span the wide range of real-world environments and reflectances that lies between the extremes that have been the focus of past work.

**Index Terms**—shape from shading, multiview stereo, shape estimation, reflectance estimation

## 1 INTRODUCTION

WHEN we look at the world around us, we do not see a series of 2D images. We see geometry, and photometry. By perceiving geometry, we enable ourselves to interact with the world, and by perceiving photometry, we can infer physical properties.

Our world is full of objects made of different materials situated in complex environments. The combined effect of this interaction, though beautiful, makes decoupling appearance into geometry and photometry extremely challenging. Past work on geometry estimation has therefore focused on compressing this complexity to fit simple models. Illumination is often controlled and assumed to be point light sources, and reflectance is often assumed to be Lambertian, mirrored or dichromatic.

Though such assumptions give rise to beneficial properties that aid geometry estimation, they limit us to a small slice of real-world materials or force us into the darkroom where we can control lighting. In this paper we relax both material and illumination assumptions to fully bring shape (and reflectance) estimation into “the wild”. Our goal is to estimate what we cannot directly (and passively) acquire (the shape and reflectance of the object) from what we can (one or more images of the object and a panorama of the illumination environment).

Certainly it is true that when an object with a non-trivial reflectance is situated in a non-trivial illumination environment we can expect complex appearance. Within this complex appearance, however, there is structure that gives strong surface orientation cues. Unique portions of the illumination environment (such as the sun or stained glass) act analogously to the point light sources of past work in

that they constrain the space of possible surface orientations to those that reflect them. The strength of this constraint, however, depends on the reflectance, the illumination, and what part of the scene is being reflected. Points reflecting unique regions of the scene are tightly constrained while those reflecting less salient regions may be quite weakly constrained.

If only a single image is available, we address the ambiguity in weakly constrained regions with careful priors designed to propagate the information from tightly constrained regions. If multiple images are available then such weak constraints become strong when additional viewpoints corroborate a tighter range of possible surface orientations.

In either case, an inherent ambiguity exists tightly coupling shape and reflectance. The shape of the object cannot be deduced without knowledge of the reflectance which modulates the illumination. On the other hand, the reflectance itself cannot be guessed without understanding what scene components are being reflected. Our overall approach is therefore to estimate the two jointly.

We leverage the Directional Statistics BRDF model for reflectance, and utilize a normal field or triangle mesh to represent geometry in the single and multiview cases, respectively. By keeping one fixed as the other is estimated, and utilizing novel constraints, we may recover both without enabling either to absorb the error of the other.

We evaluate our methods extensively on synthetic and real-world data for which we have acquired ground truth geometry. The results show that our approaches enable accurate shape and real-world reflectance estimation under complex natural illumination, effectively moving shape estimation out of the darkroom and into the real-world.

## 2 RELATED WORK

**G**EOMETRY estimation, whether from a single image, or multiple images, is a cornerstone of computer vision research. Instead of giving a full survey of these longstanding problems, we will focus on those methods that have important similarity to our own. Interested readers can refer to surveys by Durou et al. [1], Zhang et al. [2] and Seitz et al. [3] for more context.

### 2.1 Single viewpoint geometry estimation

The problem of single view geometry estimation has traditionally taken place in the darkroom where the illumination conditions can be controlled. Goldman et al. [4], for example, moved beyond the Lambertian reflectance assumption to the more general parametric Ward model. By capturing many images as a light source is moved around, sufficient information can be gathered to estimate the reflectance parameters and surface orientations. For more general, isotropic reflectances, Alldrin et al. [5], [6] showed that by structuring the illumination densely around the viewing direction, isocontours can be extracted that give depth information regardless of the actual (isotropic) BRDFs of the object. Unfortunately, no such structure can be assumed about natural illumination. Indeed, appearance is tightly coupled to the combined role of reflectance and shape. We therefore seek to jointly estimate the reflectance along with the object geometry.

Hertzmann and Seitz instead assume that they have a sphere of the same material as the target object [7]. They can then work with much more general reflectance types as the sphere serves as a measured reflectance map. By comparing the sequence of appearances of a surface patch on the target object with those of all the points on the reference sphere, the orientation for the surface patch can be deduced. The downside, however, is that a sphere of the same material needs to be acquired (or painted). By jointly estimating the reflectance, we move away from requiring a known reflectance.

There has been some work on shape estimation in natural illumination though the focus has been on Lambertian reflectance. Huang and Smith [8], and Johnson and Adelson [9] observe that under natural illumination, the appearance of a Lambertian object takes a parametric form despite the inherently nonparametric nature of the illumination environment. By representing the illumination using the spherical harmonics model [10], a parametric reflectance map can be computed that directly links appearance to surface orientation. Huang and Smith [8] also showed that the illumination parameters themselves can be estimated. Barron and Malik [11], [12] went further by detecting highlights as outliers. When the reflectance is not assumed to be Lambertian, however, the reflectance map stays nonparametric, making such approaches inapplicable.

Attention has also been paid to the other extreme form of reflectance—the purely specular. Adato et al. [13] observe the flow of the reflected, yet unknown, illumination environment for a known relative movement of the environment.

In order to effectively rotate the environment, however, this method requires that the relationship between the camera and the object be fixed, and be able to move together. Tappen [14] observes that the local curvature of mirrored surfaces result in characteristic patterns (such as the curved lines of distorted buildings and trees). Without knowing the illumination, such patterns give cues about the curvature that can then be used to recover the shape. For general BRDFs, however, such detailed analysis of the appearance is not feasible as the reflectance will smooth away the edges of the scene which would otherwise encode curvature.

Between the extremes of Lambertian and purely specular, we have the wide range of real-world reflectances that surround us. Here, neither a direct nor parametric link between illumination and surface orientation exists. In this work we show that by jointly estimating the reflectance, nonparametric distributions of likely surface orientations can be extracted for each pixel. In the single image case, we show how this complex problem becomes tractable through priors that extend the influence of the sparse but salient surface orientation cues that are reflected from the lighting.

### 2.2 Multiple viewpoint geometry estimation

A single image is inherently limited in its ability to capture the geometry of an object. As such, the problem of multiple viewpoint geometry estimation has also received considerable attention. Although the connection to the single viewpoint case is intuitive, the problem itself is not, and there are many ways to formalize the relationship.

Lambertian objects exhibit another helpful property, that of viewpoint independent appearance. A surface patch that has a certain appearance when viewed from one direction will appear the same when viewed from another direction. This simple observation gives way to the notion of *photometric consistency*. By dividing up the 3D space that the object sits in into voxels, we can then test if a voxel contains the actual object surface by checking if its appearance is consistent throughout the various observation images [15], [16], [17]. In order to build robustness to changes in radiance this notion was later extended to patch-based comparisons by Pons et al. [18]. Jin et al. [19], [20] moved the concept beyond Lambertian reflectance to the Ward model. They note that this more general reflectance still exhibits strongly constrained appearance variation. They then exploit this by measuring (and constraining) the radiance tensor field across many ( $\sim 40$ ) images. In this paper we go further by working with arbitrary isotropic BRDFs, and by working under natural illumination.

The problem of arbitrary reflectance has been approached by extending the notion of consistency to *orientation consistency*. By leveraging past work on single-viewpoint geometry estimation, each viewing direction can be converted into an analogous geometry observation. Hernández et al. [21], for example, use controlled lighting to convert each observation location into a reliable geometry observation in the form of a surface normal field. Similarly, Treuille et al. [22] use a reference sphere of the same material to extract

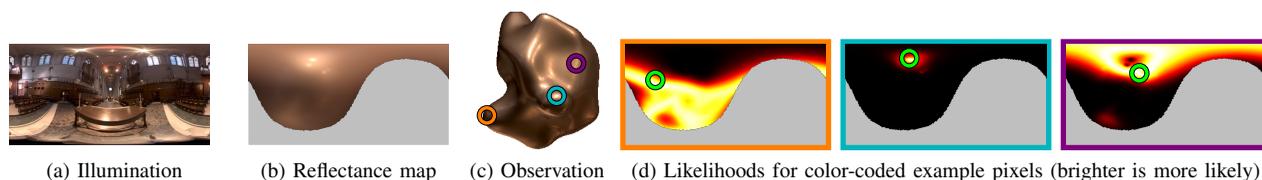


Fig. 1: **Surface orientation likelihood spherical panoramas.** The illumination (a) and reflectance combine to form a complex reflectance map (b). When this is compared with the observed appearance of a pixel in an observation (c), we arrive at a nonparametric distribution of surface orientations (d). The green circles denote the true orientation of the corresponding pixels (sorted left to right, and color coded).

normal fields for each observation location. The goal for these methods is then to estimate a full 3D model that is consistent with the newly created geometry observations. Our multiview framework has a similar motivation, but operates with the highly ambiguous, nonparametric surface orientation distributions that result from natural illumination for non-trivial reflectances. In other words, instead of having reliable geometry observations in the form of normal fields we have nonparametric multi-modal distributions of possible surface orientations. We introduce a framework for compactly representing this complexity to perform full shape and reflectance estimation.

### 3 BAYESIAN SHAPE AND REFLECTANCE ESTIMATION

THE appearance of an object is due to the illumination, viewing conditions, shape of the object, and its reflectance. We assume the illumination is known but uncontrolled natural illumination  $\mathbf{L}$ , the object material has an isotropic reflectance function, and that it has been segmented from the background. We also assume that we have one or more images  $\mathcal{I} = \{\mathbf{I}^1, \dots, \mathbf{I}^M\}$ ,  $M \geq 1$ , from a calibrated camera. These assumptions can be met using existing work (SfM may be sufficient if the surrounding environment is feature-rich).

Our primary contribution is a probabilistic framework for estimating the remaining components—the geometry  $\mathbf{G}$ , and reflectance  $\Psi$ . We formulate this as the maximum a posteriori (MAP) estimate of the posterior distribution

$$p(\mathbf{G}, \Psi | \mathcal{I}) \propto p(\mathcal{I} | \mathbf{G}, \Psi) p(\mathbf{G}) p(\Psi), \quad (1)$$

where the likelihood  $p(\mathcal{I} | \mathbf{G}, \Psi)$  quantifies how consistent the geometry and reflectance are with the observation(s), and the priors  $p(\mathbf{G})$  and  $p(\Psi)$  encode practical constraints.

Our overall algorithm is modeled after expectation maximization. After initialization, an outer loop alternates between updating the likelihood variance (discussed below) and updating the geometry and reflectance. When updating these two components one is kept fixed while the other is updated, alternating back and forth until convergence.

### 4 A SINGLE IMAGE

LET us first fully inspect what can be learned from a single image. We will begin by investigating how shape can be deduced from reflectance, then we will discuss

how reflectance can be deduced from shape. We will then show how to soundly initialize this iterative process.

#### 4.1 Shape from reflectance

##### 4.1.1 Image formation likelihood

Consider a single pixel  $\mathbf{I}_i$  in the image  $\mathbf{I}$ . For example, the pixel of Fig. 1c circled in orange. The appearance of this pixel is due to the reflectance  $\Psi$ , the illumination environment  $\mathbf{L}$ , and the underlying orientation  $\mathbf{N}_i$  of the corresponding surface point, with some added Gaussian noise of uniform variance,

$$\mathbf{I}_i = \mathbf{E}(\Psi, \mathbf{L}, \mathbf{N}_i) + \mathcal{N}(\mathbf{0}, \sigma^2). \quad (2)$$

The likelihood thus takes the form of a Gaussian centered on the predicted irradiance  $\mathbf{E}_i$ . Here we use the log-intensities to remain sensitive to subtle detail,

$$p(\mathbf{I}_i | \mathbf{N}_i) = \mathcal{N}(\ln(\mathbf{I}_i) | \ln(\mathbf{E}_i), \sigma^2). \quad (3)$$

The surface orientation  $\mathbf{N}_i$  determines what hemisphere of light will be modulated by the reflectance and integrated to form the appearance. As shown in Fig. 1b, since the predicted irradiance is a function of  $\mathbf{N}_i$  (which itself can be expressed in 2D spherical coordinates  $\mathbf{N}_i = (\theta_i, \phi_i)$ ), we can visualize it as a 2D spherical panorama. Note that only the half of the image corresponding to the camera-facing hemisphere is filled in, while the self-occluded half is shown in light gray.

The likelihood may be visualized similarly, by computing Eq. 3 for each orientation of a spherical panorama. Three examples are shown in Fig. 1d where brighter values correspond to higher-probability orientations. Note how examples 1 and 3 have no clear minimum; the true orientation of the underlying surface point (which is indicated with a green circle) can seldom be directly inferred by appearance.

The overall likelihood can then be expressed over all pixels of the image

$$p(\mathcal{I} | \mathbf{G}) = p(\mathbf{I} | \mathbf{G}) = \prod_{i \in \Omega} p(\mathbf{I}_i | \mathbf{N}_i), \quad (4)$$

where, in the single image case, the overall shape of the object  $\mathbf{G}$  is expressed as a set of surface normals  $\mathbf{G} = \{\mathbf{N}_i\}$  for each pixel  $i$  of the object  $\Omega$ .

##### 4.1.2 Surface orientation constraints

In the center of Fig. 1d we see a single bright spot in the distribution indicating that the appearance of this pixel,

which lies in a highlight, can only be explained by a small range of possible surface orientations. This occurrence, though rare, gives valuable context to the area surrounding the well-determined pixel. We introduce two spatial priors to help propagate this information. We also utilize the occluding boundary as a strong unary prior  $p_b$  on pixels at the edge of the object. The spatial priors  $p_g$  and  $p_s$  are formulated pairwise,

$$p(\mathbf{G}) = \prod_{i \in \Omega} p_b(\mathbf{N}_i) \cdot \prod_{i \in \Omega} \prod_{j \in \text{ne}(i)} p_g(\mathbf{N}_i, \mathbf{N}_j) \cdot p_s(\mathbf{N}_i, \mathbf{N}_j). \quad (5)$$

**Occluding boundary prior** As first explored by Ikeuchi and Horn [23], surface patches on the occluding boundary of smooth objects must be oriented orthogonally to the viewing direction. Since the area subtended by each boundary pixel may include a range of orientations, not all of which will be orthogonal to the viewing direction, we formulate this prior as a tight, but non-singular, distribution,

$$p_b(\mathbf{N}_i) \propto \begin{cases} \exp\{-\beta_b \arccos^2(\mathbf{N}_i \cdot \mathbf{B}_i)\} & \text{if } i \in \mathbf{B} \\ 1 & \text{otherwise,} \end{cases} \quad (6)$$

where  $\mathbf{B}$  is a list of the orthogonal vectors for pixels on the occluding boundary of the object, and  $\beta_b$  controls the strength of the prior.

**Reflected gradient prior** Our first spatial prior ensures that the resulting gradient is the same as the observed image gradient

$$p_g(\mathbf{N}_i, \mathbf{N}_j) \propto \exp\{-\beta_g \|\mathbf{E}_i - \mathbf{E}_j - (\mathbf{I}_i - \mathbf{I}_j)\|^2\}, \quad (7)$$

where  $\|\cdot\|$  is the Euclidean norm, and  $\beta_g$  controls the strength of the prior.

**Smoothness prior** Although the gradient prior gives sufficient context to pixels reflecting texture-rich regions of the illumination, in most cases, the reflectance map contains many areas with little or no appearance variation. Our smoothness prior is designed to allow small changes in surface orientation (angles less than  $\pi/3$ ) while strongly penalizing sharp changes in orientation. To do so, we formulate this as a logistic function,

$$p_s(\mathbf{N}_i, \mathbf{N}_j) \propto \left(1 + \exp\{-s(\arccos(\mathbf{N}_i \cdot \mathbf{N}_j) + t)\}\right)^{-1}, \quad (8)$$

where  $t = \pi/3$  is the threshold between weak and strong penalization and  $s = 10$  is the speed of the transition. Early experiments with more simple priors (such as Gaussian and Laplacian) resulted in geometry estimates that exhibited unnatural flatness. Higher-order priors that operate on larger blocks of pixels are too computationally complex, since a globally optimal solution needs to be found. We found that this formulation nicely accounts for both concerns.

## 4.2 Reflectance from shape

Now we will describe our method for estimating the reflectance using the image of the object, the illumination environment, and the current geometry estimate  $\mathbf{G}$  as input.

### 4.2.1 The directional statistics BRDF model

To model the reflectance function, we adopt the Directional Statistics Bidirectional Reflectance Distribution Function (DSBRF) model, introduced by Nishino [24], [25] and later extended by Lombardi and Nishino [26] to estimate reflectance in natural illumination. The model offers a compact representation of isotropic BRDFs and is naturally paired with a simple, data-driven prior.

Using a linear camera, the irradiance  $\mathbf{E}(\Psi, \mathbf{L}, \mathbf{N}_j)$  (of Eq. 2) is

$$\mathbf{E}_j = \int \varrho(t(\omega_i, \omega_o); \Psi) \mathbf{L}(\omega_i) \max(0, \mathbf{N}_j \cdot \omega_i) d\omega_i, \quad (9)$$

where  $t$  is a function that transforms the incoming  $\omega_i$  and outgoing  $\omega_o$  angles into the alternate BRDF parameterization variables  $\theta_d$  and  $\theta_h$ . The reflectance function is expressed as a sum of lobes

$$\varrho^{(\lambda)}(\theta_d, \theta_h; \kappa^{(\lambda)}, \gamma^{(\lambda)}) = \sum_r \exp\left\{\kappa^{(r,\lambda)}(\theta_d) \cos^{\gamma^{(r,\lambda)}(\theta_d)}(\theta_h)\right\} - 1, \quad (10)$$

where the halfway vector parameterization (i.e.,  $(\theta_h, \phi_h)$  for the halfway vector and  $(\theta_d, \phi_d)$  for the difference vector) [27] is used.  $\kappa^{(\lambda)}$  and  $\gamma^{(\lambda)}$  are functions that encode the magnitude and acuteness of the reflectance, respectively, of lobe  $r$  along the span of  $\theta_d$  for a particular color channel  $\lambda$ . These curves are modeled as a log-linear combination of data-driven basis functions,

$$\begin{aligned} \kappa^{(r,\lambda)}(\theta_d) &= \exp\left\{b_\mu(\theta_d; \kappa, r, \lambda) + \sum_k \psi_k b_k(\theta_d; \kappa, r, \lambda)\right\}, \\ \gamma^{(r,\lambda)}(\theta_d) &= \exp\left\{b_\mu(\theta_d; \gamma, r, \lambda) + \sum_k \psi_k b_k(\theta_d; \gamma, r, \lambda)\right\}, \end{aligned}$$

where  $b_\mu$  is the mean basis function,  $b_k$  is the  $k^{\text{th}}$  basis function, and  $\psi_k$  are the DSBRDF coefficients. We may compute these basis functions from a set of measured reflectance functions using functional principal component analysis.

### 4.2.2 Probabilistic reflectance estimation

In order to estimate the parameters  $\Psi$  we continue with our probabilistic formulation of Eq. 1. Here, the likelihood is the same as above, though the geometry, and hence the per pixel normals  $\{\mathbf{N}_i\}$ , are kept fixed

$$p(\mathcal{I}|\Psi) = p(\mathbf{I}|\Psi) = \prod_{i \in \Omega} \mathcal{N}(\ln(\mathbf{I}_i) | \ln(\mathbf{E}_i), \sigma^2), \quad (11)$$

where, again,  $\Omega$  is the set of all pixels  $i$  of the object,  $\mathbf{I}_i$  refers to the appearance of the pixel, and  $\mathbf{E}_i$  refers to its predicted irradiance.

We utilize the prior by Lombardi and Nishino [26], which encourages the coefficients  $\psi_k \in \Psi$  of the eigen-functions to be within the distribution of observed reflectances,

$$p(\Psi) = \mathcal{N}(\Psi | \mathbf{0}, \beta_\Psi \Sigma_\Psi), \quad (12)$$

where the covariance  $\Sigma_\Psi$  is computed from the MERL database [28], and the scalar  $\beta_\Psi$  controls the prior strength.

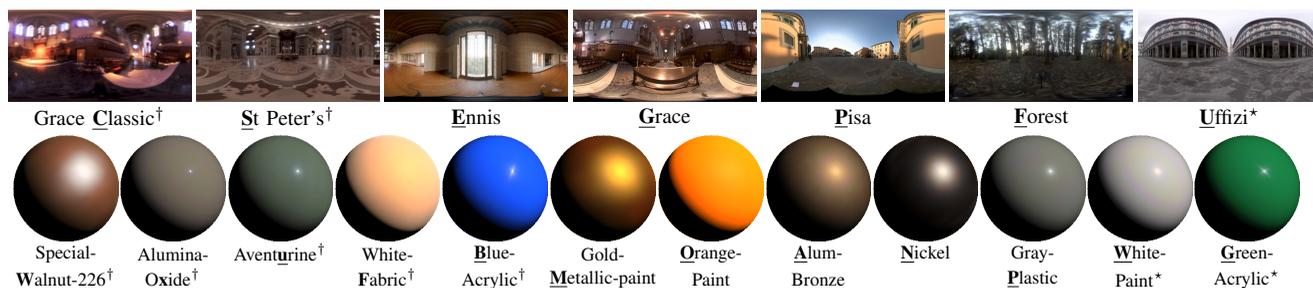


Fig. 2: **Synthetic data.** Our synthetic data are formed by rendering 10 shapes [9] with real-world BRDFs [28] under real-world illumination environments [29]. As indicated by † and \*, a subset of the shown reflectances and environments are used only in the single image case, or multiple image case, respectively, while all others are used in both.

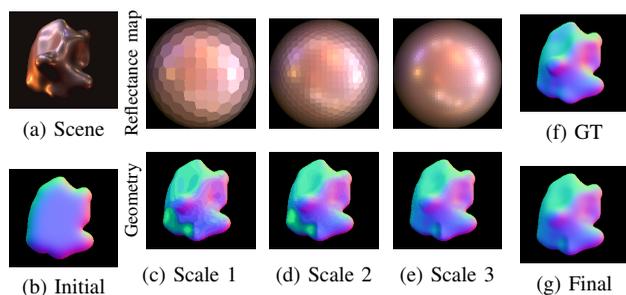


Fig. 3: **Multi-scale geometry estimation.** We partition the possible surface orientations into a geodesic hemisphere. By incrementally increasing the resolution of the hemisphere we avoid local minima. Our final step is a gradient-descent based minimization to fine-tune the result.

### 4.3 Implementation and optimization

#### 4.3.1 Initial estimation

We begin our optimization with a naive geometry estimation. Without an estimate for the reflectance, the likelihood of Eq. 3 and the reflectance gradient prior of Eq. 7 are meaningless. We may still, however, use the smoothness and occluding boundary priors. As shown in Fig. 3b, the result is a normal field that resembles a soap-bubble. Although this is a crude estimate of the object geometry, it serves as a reasonable starting point and allows us to bootstrap the iterative optimization.

#### 4.3.2 Multi-scale geometry estimation

As we have noted, the primary challenge in this problem stems from the nonparametric nature of the reflectance map. Even if a pixel has a unique appearance, the reflectance map is rarely smooth enough that a simple iterative optimization scheme will find this optimal value. To address this, we introduce a novel global optimization algorithm. As shown in Fig. 3, we limit the space of possible surface orientations to a set of evenly-distributed orientations on the geodesic hemisphere. In this figure the reflectance map is rendered as a hemisphere instead of a half-filled spherical panorama to conserve space.

There are several benefits to this formulation. First, by limiting the surface orientations to a finite set, we enable the use of rapid global optimization approximation methods.

And second, the effect of an inaccurate reflectance estimation is minimized. An inaccurate reflectance estimate may cause the likelihood of Eq. 3 to have a unique minimum that is incorrect. By limiting the possible surface orientations, we force each pixel to align with the *region* it is most similar to, while obeying the constraints of our priors. The size of these regions is gradually decreased allowing the priors to play a more subtle role at finer scales.

For a given material estimate, the surface orientations are estimated at each of three scales by incrementally dividing the geodesic hemisphere. To approximate the optimal assignment of surface orientations, we use the alpha-beta swap method of Boykov et al. [30]. To help interpolate from one scale to the next, the orientation estimate from the previous scale is first smoothed with a Gaussian kernel before being re-discretized at the next finer resolution. The progression of this optimization scheme is shown in Fig. 3.

#### 4.3.3 Refinement and integrability

At this point we have a good estimate for the object geometry. In order to ensure that the underlying surface is integrable [31], and to refine the still somewhat coarse surface orientation estimate, we apply a gradient descent optimization. Similar to Ikeuchi and Horn [23], we perform this optimization in gradient space. The primary difference in our work is due to the DSBRDF. The final shape estimation for our running example is shown in Fig. 3g. Note that this refinement is only possible because we already have a good estimate through discrete global optimization.

### 4.4 Experimental evaluation

#### 4.5 Synthetic images

Just as real-world materials lie somewhere between the Lambertian and mirrored materials that have been the focus of past work, real-world illumination environments lie somewhere between the point lights and infinitely complex lighting that are respectively ideal for geometry recovery of these materials. In order to better understand what properties of reflectance and illumination are conducive to real-world geometry estimation we generated a large synthetic database. As shown in Fig. 2, we utilized 6 different publicly-available illumination environments [29] and 10 different measured BRDFs [28]. These sets were

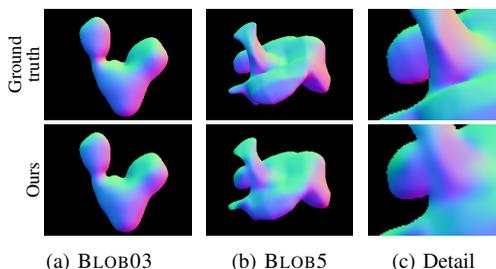
		Reflectances										mean
		A	W	P	N	M	F	O	X	U	B	
Environments	C	12.2°	11.3°	20.0°	19.1°	15.9°	12.1°	27.9°	29.8°	14.8°	34.4°	19.8°
	E	11.9°	17.3°	15.4°	23.2°	17.4°	35.2°	22.4°	19.6°	28.4°	38.2°	22.9°
	P	19.2°	21.1°	22.1°	26.9°	24.4°	27.8°	26.8°	27.5°	28.9°	36.0°	26.1°
	G	22.8°	22.4°	28.8°	25.4°	27.3°	25.4°	37.6°	29.0°	28.2°	32.3°	27.9°
	S	19.4°	23.4°	27.1°	25.3°	31.4°	35.9°	31.7°	34.5°	42.6°	35.8°	30.1°
	F	25.1°	26.4°	32.3°	30.5°	34.4°	37.2°	29.2°	35.5°	36.5°	37.4°	32.4°
mean		18.4°	20.1°	24.3°	25.0°	25.1°	28.9°	29.3°	29.3°	35.7°	26.6°	

(a) Geometry errors

		Reflectances										mean
		A	W	P	N	M	F	O	X	U	B	
Environments	C	0.57	0.95	0.91	1.10	1.29	0.98	1.10	1.50	1.54	2.05	1.20
	P	0.54	0.71	1.06	1.90	1.39	0.89	0.90	1.54	1.50	1.98	1.24
	E	0.41	0.60	0.58	1.98	1.06	0.81	0.98	1.37	1.75	1.77	1.13
	G	0.44	1.05	0.93	1.87	1.81	1.08	1.09	0.87	1.98	3.02	1.41
	S	0.65	1.10	0.89	1.75	1.99	1.11	1.70	1.27	1.50	2.01	1.40
	F	0.62	0.98	1.11	1.91	1.83	0.97	1.19	1.38	1.60	1.68	1.33
mean		0.54	0.90	0.91	1.75	1.56	0.97	1.16	1.32	1.65	2.09	1.29

(b) Reflectance errors

**TABLE 1: Synthetic results summary.** Each cell shows the average RMS geometry or reflectance error across the 10 blobs for an illumination (row) and reflectance (column) combination. The headers correspond to the bold letters in Fig. 2. For quick inspection, lower errors are given a brighter background coloring. The last row and column are means.

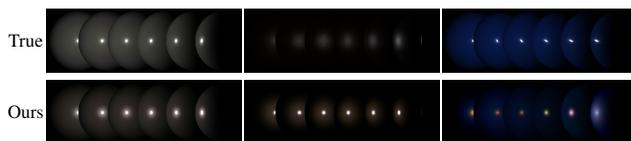


**Fig. 4: Shape accuracy.** Shapes with self occlusions, like BLOB05 (b) and BLOB01 present challenges to our algorithm, while simple shapes like BLOB03 (a) are more reliably estimated.

chosen to span a wide range of real-world illuminations and reflectances. In each of these 60 scenarios, we rendered the 10 shapes of the Blobby Shapes database [9], for a total of 600 different experiments. Note that when computing the DSBRDF prior, we omit the ground-truth BRDF. In other words, the ground-truth BRDF is not a part of the reflectance training data for the object being analyzed.

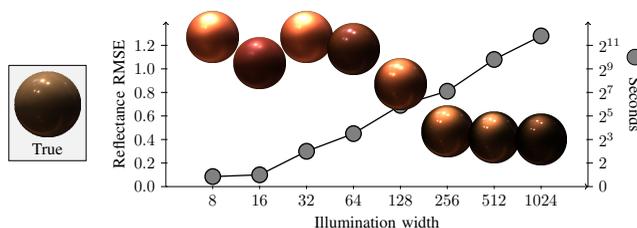
In order to directly compare with the work of Johnson and Adelson [9], we have also briefly tested our method assuming the ground-truth reflectance map is known. In this case we achieve median angular errors consistently below 15°. This shows the strength of our model to overcome the inherent ambiguities of a nonparametric reflectance map. The increased error when the reflectance is not known is due to the added challenge. Since it is more practical, we will focus the rest of this discussion on that case.

In Table 1a we show the average root-mean-squared



(a) Ennis - Gray Plastic (b) Pisa - Special Walnut (c) Forest - Blue Acrylic

**Fig. 5: Reflectance accuracy.** Some reflectances such as Gray Plastic (a) are consistently well recovered, while others such as acrylic paints (c) have challenging complexity.



**Fig. 6: Reflectance accuracy over illumination size.** For a specular material, the accuracy of the reflectance map increases as detail is added to the illumination map. Data points are shown as rendered spheres. The ground truth rendering is inset.

error of the geometry estimates for all of the 60 different scenarios. Each number is an average over the 10 different shapes. Note that, as shown in Fig. 4, some shapes are more challenging to estimate than others. In particular, the most challenging shapes are those with multiple self occlusions, an occurrence that violates our assumptions about smoothness and integrability. Since the presented errors are averaged over all 10 shapes, they encompass a wide range of complexity and difficulty.

The entries of Table 1 are shaded brighter to indicate increased accuracy, and the column and row headers correspond to the underlined characters in Fig. 2. Some reflectances, such as Alum-Bronze (A) and Gray-Plastic (P) consistently yield more accurate results. These reflectances all have a strong diffuse lobe, mixed with a moderate to strong specular lobe. This means that the amount of light being integrated to form the appearance is of two scales—large for the diffuse lobe, and small for the specular lobe. Since the environments themselves are comprised with both high frequency and low frequency details, this mixture gives powerful cues about the surface orientations.

This trend holds for the majority of the illumination environments. The St. Peter's (S) basilica and Forrest (F) environments, however, are more challenging. They exhibit a dominance of high frequency components and repetition, making shape estimation very challenging.

Table 1b shows the RMS error of the reflectance estimates, and Fig. 5 shows some example results. Though the results show a similar trend to the geometry error, there are a couple reflectances that proved especially challenging to estimate. Nickel (N) has a very weak diffuse component, making it especially hard to estimate. Blue Acrylic (B) is challenging due to its monochromatic reflectance. Since no green or red light is reflected we see colorful highlights.

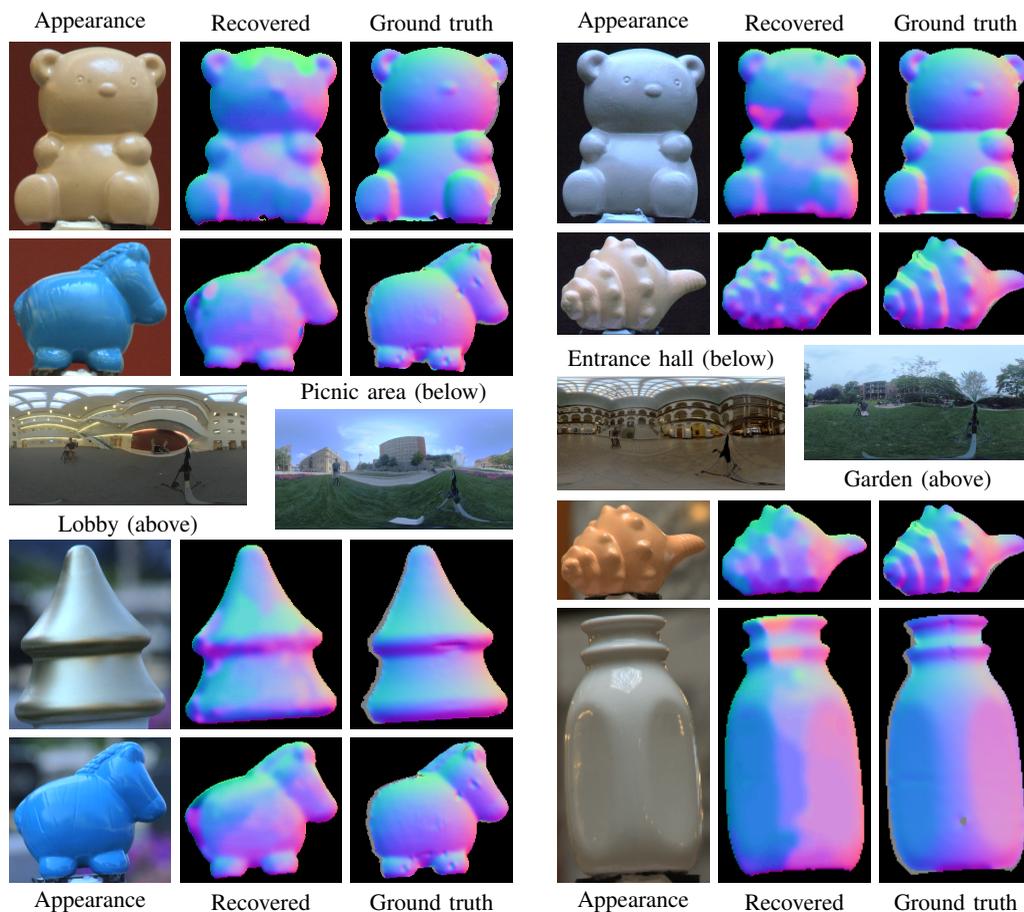


Fig. 7: **Real-world results.** We captured several objects in four different natural lighting environments, and aligned ground-truth normal maps. As we discuss in the text, differences in the lighting environments have a clear impact on the accuracy of the recovered geometry.

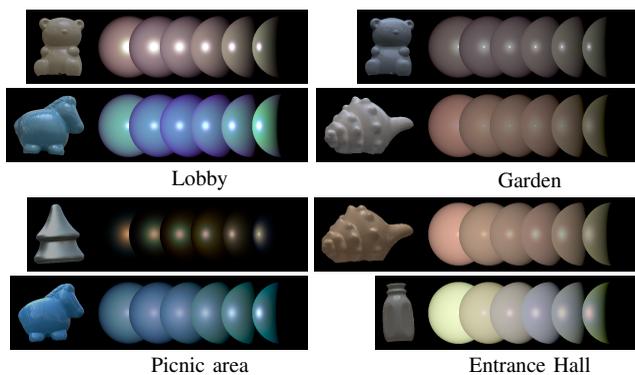


Fig. 8: **Real-world reflectance results.** Each entry in this table corresponds to the results shown in Fig. 7.

As shown in Fig. 6, the size of the illumination map plays an important role in recovering reflectances with a strong specular component. The left vertical axis shows the average RMS error of the recovered reflectance. Data points are displayed as spheres rendered under the full illumination environment. Accuracy of the recovered reflectance naturally increases as the illumination map grows, while running time grows linearly. We chose to use illumination maps of  $256 \times 128$  for our experiments.

#### 4.6 Real-world images

As shown in Fig. 7, we have also acquired images, and aligned ground-truth geometry for several objects in both outdoor and indoor real-world scenes.<sup>1</sup> The ground-truth geometry was acquired using a Canon VIVID 910 light-stripe range-finder. Illumination environments were acquired using a reflective sphere. Although real-world data comes with added sources of noise and inaccuracy, our method is able to recover the shape of the objects quite well in each of the environments.

In each of the four sections of the figure we show the illumination environment along with the image, recovered normal field and ground-truth normals for each of the objects in the scene. The top left section shows a bear figure and a horse figure in a large lobby. The dominant illumination from the overhead skylights is relatively uniform. This leads to some inaccuracy along the top of the objects. The walls of the lobby, however, contain many different points of illumination that contribute to the smoothly varying geometry recovered along the middle of the objects. In this scene the bear has a mean angular error of  $24^\circ$  and the horse has a mean angular error of  $17^\circ$ .

The garden scene in the top right is illuminated primarily

1. Available at: <http://cs.drexel.edu/~kon/natgeom>

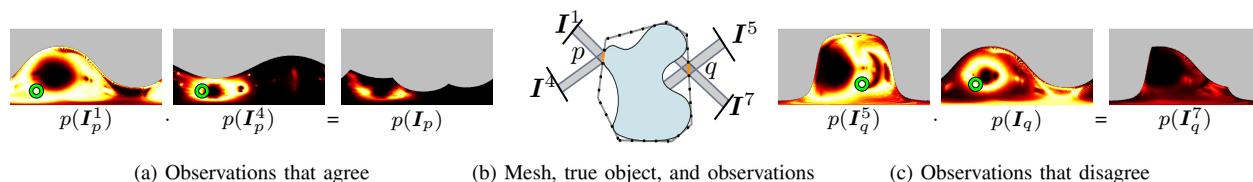


Fig. 9: **Nonparametric orientation consistency.** When a point on the mesh (dashed) is close to the true surface (solid), the observations will agree (a), resulting in a dense orientation distribution with a clear peak (bright region). When the point is not yet well aligned, the observations will disagree (c), resulting in a flat, near-zero distribution.

by a cloudy sky. This smoothly varying illumination gives rise to more accurate estimates in the top portions of the objects. In the picnic area scene of the lower left we have a similar environment. This one, however, contains flowers and has more texture in the grass at the bottom of the scene. These two features give extra surface orientation cues that the garden scene does not have. The effect can be seen most dramatically in the cheek of the bear. Since the lower portion of the garden scene is essentially uniform, points oriented to the left and right have very similar reflected appearances. The objects in the garden scene have mean errors of  $19^\circ$  and  $26^\circ$  whereas the objects in the picnic area scene both have mean angular errors of  $17^\circ$ .

In the entrance hall scene (bottom right) we see the highest accuracy along the top of the objects. This is due to variation in the overhead lighting. On the other hand, the pattern in the walls is relatively uniform throughout, giving rise to less accurate detail around the middle of the objects. In this scene the mean errors for the objects are  $20^\circ$  for the shell and  $16^\circ$  for the milk bottle.

Fig. 8 shows the recovered reflectances for the real-world results of Fig. 7. The slight color shifting that occurs in some results is due to the relatively small amount of reliable information present in the grazing angles of the illumination environments. Despite this, the results are plausible as can be confirmed in the consistency of the estimates and the overall diffuseness or glossiness that intuitively agree with those of the observations. The ceramic finish of the milk bottle proves to be the most challenging. Its highly specular finish results in appearances that cannot be described without extremely high resolution, high accuracy illumination representation, which was not available in our testing environment.

## 5 MULTIPLE IMAGES

SINCE a single image is necessarily limited in its expression of geometry, we turn our attention now to the case when multiple images are available.

### 5.1 Shape from reflectance

#### 5.1.1 A unified coordinate frame

As we saw, when only a single image is available, pixels with ambiguous surface orientation distributions must rely on their neighbors to reduce the ambiguity. When multiple images are available, each observation serves as a separate

constraint on the distribution of possible surface orientations. In order to compare observations from different images, however, we must first provide a means to link regions from different images to the same physical location on the object surface.

Fig. 9 illustrates how a single geometry model can be used to coordinate the observations. In the middle of the figure we see the ground-truth object (with a solid boundary) circumscribed by a coarse geometry estimate (dashed). If we take a single point  $p$  on this geometry estimate, and project it into each of the observations images  $I^m \in \mathcal{I}$  we may then compute the likelihood density for that point as a product of the separate observations

$$p(\mathcal{I}_p | \mathcal{N}_p) = \prod_{m \in \Omega_p} p(I_p^m | \mathcal{N}_p), \quad (13)$$

where  $I_p^m$  is the appearance of the projected point  $p$  in image  $m$ , and  $\Omega_p$  is the set of images that can view the point. Note that  $p(I_p^m | \mathcal{N}_p)$  is identical to the single image Eq. 3, but for a back-projected surface point.

Two examples are shown in Fig. 9. In the case on the left (a), the imaged point  $p$  is quite close to the true geometry. A direct consequence of that is that the actual imaged appearance is of the same surface point in both  $I^1$  and  $I^4$ . The surface orientation distributions for these two observations therefore overlap nicely, and the resulting distribution for the point is concentrated, with a small (bright) region.

On the right (c), we see the projection of a point  $q$  that is far removed from the true surface. The consequence of this is that the two imaged appearances attributed to this point are actually of different points of the real object. Since the imaged geometry for image  $I^5$  is oriented upwards, and the imaged geometry in  $I^7$  is oriented downwards, their surface orientation distributions are unlikely to overlap. In this case we can see that the resulting distribution exhibits no clear orientation for the point  $q$ .

Point  $q$ , whose distribution is flat and near-zero, has no true orientation, it is far removed from the surface at this stage; as the mesh evolves, it may end up somewhere quite different from its current location.  $p$ , however, has a clear target orientation, and is likely near the correct location.

#### 5.1.2 Surface patches

Now that we have seen how to unite multiple observations to derive tighter surface orientation distributions, we may turn our attention to recovering a full 3D model. By

focusing on the facets of the model  $f \in \mathcal{G}$ , we provide a way to use surface orientation cues to deduce geometry. The goal then is to morph the points of the mesh so that the facet orientations are consistent with the observations.

In order to take full advantage of higher-resolution observation images, we take  $J$  uniformly distributed samples from a facet and average them (in our case  $J = 6$ ),

$$p(\mathbf{I}_f^m | \mathcal{N}_f) \propto \sum_{j=1}^J w_j \cdot p(\mathbf{I}_{f,j}^m | \mathcal{N}_f), \quad (14)$$

where the weights  $w_j$  ( $\sum w_j = 1$ ) are higher for samples near the center of the facet. This sampling is done by division of the barycentric parametrization of the facet, and the weights are set equal to the minimal barycentric coordinate value. Here,  $\mathbf{I}_{f,j}^m$  indicates a specific pixel—the  $j^{\text{th}}$  sample of facet  $f$  in observation  $m$ . The final likelihood for the facet is then the product of the per-image distributions (as in Eq. 13),

$$p(\mathcal{I} | \mathcal{N}_f) = \prod_{m \in \Omega_f} p(\mathbf{I}_f^m | \mathcal{N}_f). \quad (15)$$

### 5.1.3 Probabilistic shape estimation

Now that we have described how to form the likelihood for a single facet, we may express the full likelihood of Eq. 1, as the product over all facets,

$$p(\mathcal{I} | \mathcal{G}) = \prod_{f \in \mathcal{G}} p(\mathcal{I} | \mathcal{N}_f). \quad (16)$$

Finally, we place three priors on the mesh itself,  $p(\mathcal{G}) = p_c(\mathcal{G})p_h(\mathcal{G})p_e(\mathcal{G})$ . The first prior is inspired by recent work on minimal surface constraints [12]. To propagate the shape of accurate regions to those far removed from the true surface (as in Fig. 9c), we encourage the local curvature to be constant. This is approximated as the variance of the angles between the normal of each facet  $\mathcal{N}_f$  and those of the facets that lie on a ring around it

$$p_c(\mathcal{G}) \propto \prod_{f \in \mathcal{G}} \exp \left\{ -\beta_c \left( \frac{1}{|\mathbf{r}(f)|} \sum_{h \in \mathbf{r}(f)} \arccos^2(\mathcal{N}_f \cdot \mathcal{N}_h) - \left[ \frac{1}{|\mathbf{r}(f)|} \sum_{h \in \mathbf{r}(f)} \arccos(\mathcal{N}_f \cdot \mathcal{N}_h) \right]^2 \right) \right\}, \quad (17)$$

where  $\mathbf{r}(f)$  denotes the set of facets that surround  $f$ ,  $|\cdot|$  denotes the cardinality of that set, and  $\beta_c$  controls the strength of the prior. To impose the prior only on the immediate neighborhood, for example, facets that share a single point in common with  $f$  can be used. To impose the prior more globally, facets that lie on the subsequent rings surrounding  $f$  may be used. By operating on the variance of these angles, we are able to encourage uniform curvature while not imposing any first-order constraint. In section 5.3, we describe our use of this prior.

The next prior  $p_h(\mathcal{G})$  ensures that the mesh does not grow outside of the visual hull. It is designed to give no penalty to any point  $v \in \mathcal{G}$  inside the hull, while

penalizing points that leave the hull. Since computing point-to-mesh distances is expensive, we first detect the point on the hull  $\mathbf{h}_v$  closest to each vertex  $v$ . We also compute the surface orientation of the hull at that point  $\mathbf{n}_v$ . The (signed) distance between each vertex  $v$  of our geometry estimate and the hull can then be estimated using a simple dot product,

$$p_h(\mathcal{G}) \propto \prod_{v \in \mathcal{G}} \exp \left\{ -\beta_h \max(0, (v - \mathbf{h}_v) \cdot \mathbf{n}_v) \right\}, \quad (18)$$

where  $\beta_h$  controls the strength of the prior. If the segmentations used to make the visual hull cannot be trusted, for example, this weight can be set to zero.

The last prior  $p_e(\mathcal{G})$  is due to our implicit assumption about the triangles that make up the mesh. It helps ensure that the triangles are roughly equilateral so that samples within each triangle may be assumed to be relatively nearby on the actual surface. It is formulated in terms of the variance of edge lengths of each facet,

$$p_e(\mathcal{G}) \propto \prod_{f \in \mathcal{G}} \exp \left\{ -\beta_e \left( \frac{1}{3} \sum f_e^2 - \left[ \frac{1}{3} \sum f_e \right]^2 \right) \right\}, \quad (19)$$

where  $f_e$  is the length of an edge  $e$  of the facet, and again  $\beta_e$  controls the strength of the prior.

### 5.1.4 Parameterizing the distribution

Recall that the facet likelihoods  $p(\mathcal{I}_f | \mathcal{N}_f)$  are nonparametric in that they depend on the inherently nonparametric illumination environment. Because of this, a direct optimization is intractable (the visualizations in Fig. 9 are themselves discrete approximations). In order to optimize without performing an exhaustive search, we need a way to faithfully parametrize the distribution while providing a way to avoid local minima.

To do so, we first pick a finite set of  $L$  orientations  $\{\mathcal{N}^l\}$  by uniformly sampling the unit sphere. We then encode the distribution as a mixture of Von Mises-Fisher distributions centered at these orientations. The concentration (spread) of each distribution  $\kappa_l$  is proportional to the probability of the corresponding surface orientation  $\mathcal{N}^l$  as computed by Eq. 15 (in our case  $\kappa_l = 200 \cdot p(\mathcal{N}_f)$ ),

$$p_{\text{approx}}(\mathcal{I}_f | \mathcal{N}_f) \propto \sum_{l=1}^L C(\kappa_l) \exp \left\{ \kappa_l \mathcal{N}^l \cdot \mathcal{N}_f \right\} \quad (20)$$

where  $C(\kappa_l)$  is a normalization constant.

This formulation gives a continuous expression that is differentiable everywhere. The original distribution may have large areas with the same probability due to textureless regions of the illumination environment leading to ambiguous gradients. The parameterized distribution, on the other hand, will have a zero gradient only at local maxima and minima. In our case we set  $L = 1024$ .

## 5.2 Reflectance from shape

In order to estimate the parameters  $\Psi$  we continue with our probabilistic formulation of Eq. 1. Here, the likelihood

		Reflectances							mean
		M	O	A	N	W	G	P	
Environments	P	0.44%	0.46%	0.53%	0.59%	0.52%	0.47%	0.49%	0.50%
	G	0.49%	0.51%	0.57%	0.53%	0.67%	0.57%	0.52%	0.55%
	F	0.50%	0.51%	0.61%	0.59%	0.60%	0.59%	0.59%	0.58%
	E	0.52%	0.60%	0.57%	0.56%	0.60%	0.98%	0.68%	0.64%
	U	0.65%	0.53%	0.66%	0.65%	0.74%	0.71%	0.95%	0.70%
mean		0.52%	0.54%	0.59%	0.58%	0.63%	0.66%	0.64%	0.60%

(a) Geometry errors

		Reflectances							mean
		M	O	A	N	W	G	P	
Environments	P	0.90	0.27	0.61	0.92	0.56	0.21	0.37	0.56
	G	0.57	0.22	1.20	1.08	0.55	0.24	0.32	0.55
	F	0.67	0.26	0.75	1.19	0.50	0.20	0.32	0.50
	E	0.82	0.22	1.17	1.13	0.47	0.26	0.48	0.48
	U	0.75	0.25	1.72	0.92	0.60	0.23	0.36	0.55
mean		0.75	0.25	1.17	1.08	0.55	0.23	0.36	0.55

(b) Reflectance errors

Observations	Initial	Final	Reduction
	5	2.57%	1.20%
7	1.67%	0.91%	46%
9	1.21%	0.50%	59%
11	1.06%	0.62%	41%
13	0.98%	0.57%	41%

(c) Error over number of views

TABLE 2: **Synthetic results summary.** Each cell in (a) and (b) shows the average RMS geometry or reflectance error across the 10 blobs for an illumination (row) and reflectance (column) combination. The headers correspond to the bold letters in Fig. 2. For quick inspection, lower errors are given a brighter background coloring. The last row and column are means. 9 images are used in each scenario. In (c) we see the geometry error decrease as more views are added.

is the same as above, though the geometry, and hence the surface orientations of the facets  $N_f$ , are kept fixed,

$$p(\mathcal{I}|\Psi) = \prod_{f \in \mathcal{G}} \prod_{m \in \Omega_f} \mathcal{N}(\ln(I_f^m) | \ln(E_f^m), \sigma^2), \quad (21)$$

where  $\Omega_f$  is again the set of images in which facet  $f$  appears, and  $I_f^m$  refers to the appearance at the center of the facet in image  $m$ , and  $E_f^m$  refers to its predicted irradiance.

We use the same learned prior as before

$$p(\Psi) = \mathcal{N}(\Psi | \mathbf{0}, \beta_\Psi \Sigma_\Psi), \quad (22)$$

where the covariance  $\Sigma_\Psi$  is computed from the MERL database [28], and the scalar  $\beta_\Psi$  controls the prior strength.

### 5.3 Implementation and optimization

Our overall optimization scheme alternates between computing the Gaussian noise variance  $\sigma^2$ , and estimating the maximum a posteriori (MAP) estimate of the reflectance parameters  $\Psi$  and then geometry  $\mathcal{G}$ . This three-step optimization framework is iterated until convergence, typically around six iterations. To find the MAP estimate of the reflectance parameters  $\Psi$  and geometry  $\mathcal{G}$  we maximize the corresponding log-posteriors using gradient descent. In the case of the reflectance, this corresponds to finding reflectance coefficients  $\Psi = \{\psi_k\}$ . In the case of the geometry, this corresponds to finding the 3D locations for each of the vertices of the mesh. Since the likelihood, and two of our priors are expressed in terms of the facet normals, it is important to note that these normals are themselves functions of the point locations (specifically, the normalized cross product of two facet edge vectors).

In order to use a single set of prior weights across all environments, all input images are scaled by a constant factor so that the mean intensity of the illumination environment is 1. Both the geometry and reflectance estimation components run on the GPU for a combined running time of  $\sim 20$  min. per top-level iteration.

To bootstrap the process we first extract a rough estimate for the object geometry. As many other authors have done, we assume that the objects have been segmented from the background, enabling us to leverage the visual hull work of Laurentini [32] to initialize our geometry estimate. The mesh is then re-triangulated using the Poisson reconstruction [33], and small triangles are collapsed to

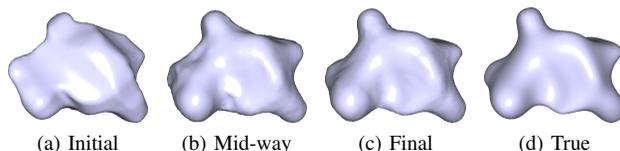


Fig. 10: **Shape optimization iterations.** The geometry estimate at several stages shows how inaccurate regions are carved away as the estimate tightens around the true shape.

help standardize the area of the triangles. The result of this step is shown in Fig. 10a. With an initial geometry estimate in place, we then perform the first reflectance estimation iteration. The prior weight  $\beta_\Psi$  is set to  $2^{-3}$ .

When optimizing the geometry we adopt one additional time-saving approximation. We assume that each camera is far enough away from the object that the mean viewing direction is sufficiently close to the actual per-pixel viewing direction (i.e., orthographic camera). This assumption allows us to pre-compute a single reflectance map for the camera pose that applies to every point in the image.

The curvature-based smoothing prior of Eq. 17 is controlled based on the number of facets in the object. Since all of our meshes have approximately 10,000 facets, we set the ring at which the smoothness is computed to be 2. For objects with more, or fewer facets the ring can be increased or decreased accordingly. The prior weights are set to  $\beta_c = 2$ ,  $\beta_h = 16$  and  $\beta_e = 0.5$ .

Throughout the optimization process we take into account occlusion when computing which images  $\Omega_f$  contain a facet. We do not, however, model any global light transport effects such as shadows or interreflection. Additionally, samples that are observed at grazing angles (an angle greater than  $75^\circ$  from the viewing direction) are discarded. This threshold was chosen to avoid overly constrained likelihood distributions in the case of the geometry refinement, and misleading grazing angle reflectance properties in the case of reflectance estimation.

As we showed in Fig. 6, the size of the illumination map impacts the running time. In the case with multiple images, the effect is compounded. For this reason, in our experiments we use an illumination map of size  $64 \times 128$ . This size is sufficient for all but the most specular reflectances.

	Horse			Pig			Shell			Milk Bottle		
	Initial	Final	Delta	Initial	Final	Delta	Initial	Final	Delta	Initial	Final	Delta
<b>Hall</b>	1.7%	1.1%	35%	0.8%	0.7%	13%	0.8%	0.7%	13%	1.0%	0.7%	30%
<b>Indoor</b>	1.3%	1.0%	23%	1.1%	0.9%	18%	0.8%	0.5%	38%	1.0%	0.7%	30%
<b>Outdoor</b>	1.4 %	1.1%	21%	0.8%	0.5%	38%	1.2%	1.0%	17%	0.6%	0.7%	+17%

TABLE 3: **Overview of real world results.** For each of our three illumination environments (rows) we evaluate each of the four objects (sections). The initial error and final error are shown, along with the relative change. The overall performance is a reduction of error by 23%.

## 5.4 Experimental evaluation

We evaluate our method quantitatively on two databases: a synthetic database, and a new real-world data set with ground-truth geometry. Since there are no other methods that recover full 3D shape with arbitrary reflectance in natural illumination, we cannot include any direct comparison.

To quantify the accuracy of our geometry estimates we compute the distance of each point on the estimated geometry to the ground-truth object. We then compute the root-mean-squared (RMS) error as a percentage of the bounding box diagonal length of the ground truth object. If, for example, the true object fits in a box with a meter diagonal, an error of 1.0% indicates a RMS error of 1cm.

## 5.5 Synthetic data evaluation

As with the single image case, we again performed many synthetic experiments in order to explore the space of real-world illumination and reflectance. The 5 illumination environments, and renderings of the 7 materials we used are shown in Fig. 2. As before, when training the reflectance model and prior, the ground-truth BRDF is omitted to ensure a fair evaluation.

Table 2 gives an overview of our results when 9 images are used. Each of the rows and columns correspond to the environments and reflectances shown in Fig. 2, respectively. The last row, and column show averages.

The consistency of results within each column of Table 2b shows clearly that certain reflectances are harder to accurately estimate than others. Most notably, the two metals Alum-Bronze (A) and Nickel (N) show the highest errors. These materials exhibit some uncommon grazing angle reflectance properties that are difficult to recover. Other reflectances such as Orange-Paint (O) and Green-Acrylic (G), however, are consistently more accurately estimated.

Table 2a shows the geometry results. As a baseline, these numbers should be compared with the mean initial RMS error of 1.00%, so even in the worst case the error is being reduced significantly. The worst geometry estimation result, with a RMS error of 0.87%, comes from the Green-Acrylic (G) reflectance in the Ennis (E) illumination environment. This is likely due to the lack of green in the scene, making the appearance due primarily to the light coming from the doorway in the center. Due to the diverse, and smoothly varying color, intensity, and texture of the scene, the Pisa (P) illumination environment gives the best performance overall with a mean RMS of 0.45%. Only one reflectance is challenging in this environment—Nickel (N), which has only a weak diffuse component. The best reflectance, Gold-

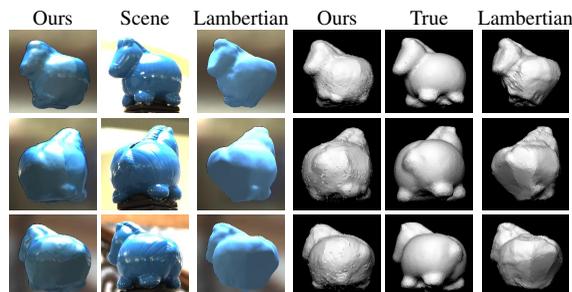


Fig. 11: **Lambertian assumption comparison.** Results of our full pipeline with the use of the DSBRDF model (our proposed method) are compared with results assuming the object has a Lambertian reflectance. Each row corresponds to a different view point. Re-renderings using the reflectance and geometry estimates are shown, as are pure white Lambertian shadings for better geometry comparison.

Metallic-Paint (M), has the best of both worlds—strong diffuse with moderate glossiness. This enables the appearance to capture both low-frequency and high-frequency detail of the illumination.

Table 2c shows the impact of additional views for a subset of our synthetic data (Blob01, Pisa, Gold Metallic Paint). In each case the views are distributed evenly around the object, that is, a new set of images is used for each case. Though it is clear that additional views improve not only the initial estimate and the final result, the viewing directions themselves plays an important role. This can be seen by the lower error when 9 views are used compared to 11 or even 13 views. Also note that due to the rich information provided by natural illumination, even in the case of 13 views, significantly fewer images are being used than in past work with controlled illumination.

Though our method assumes the object is made out of a single material, we believe that by using an increased number of images, this constraint can be relaxed. Past work on multiple complex material estimation assumes known geometry [34]. By utilizing a crude geometry estimate, as we have done, it would be possible to incorporate their findings into a framework that enables multiple material estimation. A simple extension of our framework would involve a separate optimization step in which the object is partitioned into discrete blocks of separate materials.

## 5.6 Real-world data evaluation

To quantitatively evaluate our method on real-world objects we acquired a new data set.<sup>2</sup> The data set contains four

2. Available at: <http://cs.drexel.edu/~kon/multinatgeom>

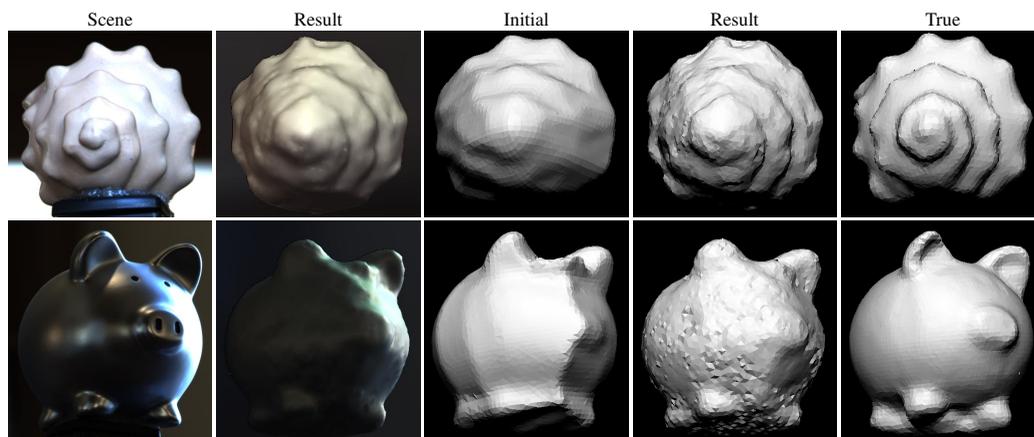


Fig. 12: **Real-world results - indoor.** Results for the indoor environment. The first two columns show the captured scene and the rendered result. The next three columns compare the initial geometry, the recovered result, and the true geometry. These are rendered with a diffuse reflectance to highlight the geometric differences.

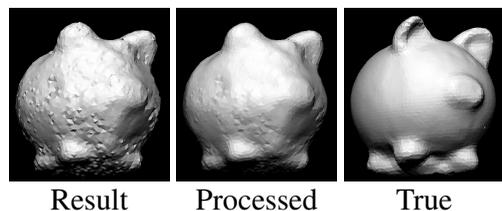


Fig. 13: Optimization artifacts can be post-processed

rendered diffusely. Though the object material has a strong Lambertian component, the purely Lambertian model is naturally unable to handle the wide range of intensity variation. Bright highlights increase the albedo estimate, causing the darkest regions of the image to be only explainable by a downward facing facet orientation. Consequently the back side of the toy horse, which has a dark appearance, collapses as its facets bend towards the ground.

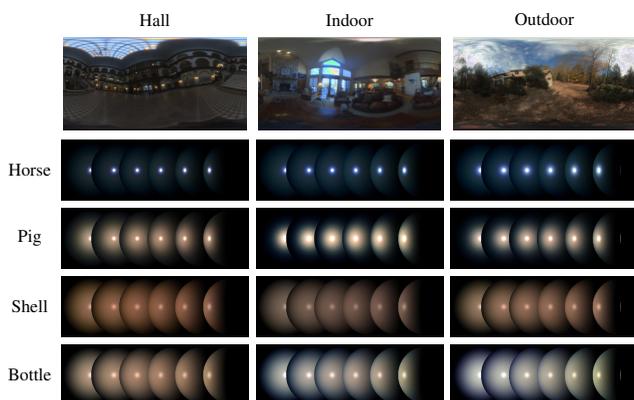


Fig. 14: **Real-world reflectance results.** All 12 of the estimated reflectances are shown. Each row corresponds to an object, and each column to an illumination environment.

Table 3 summarizes the geometry error across all of our data set. On average our method removes 21% of the original error. Figs. 12, 15 and 16 show the results for a selection of these real-world experiments. The first two columns compare an image of the scene with the corresponding final rendering of our result using the estimated geometry and reflectance. To give additional visibility into the geometry, the last three columns show diffuse renderings of the initial, final, and true geometry. Note that the bottom side of the objects is never visible to the camera due to the support structure. As a direct consequence of this, objects with complex bases result in higher error. Note also that imaging the illumination environment necessarily results in a low-pass filter of the true illumination environment as fine detail is compressed into coarse pixels. This decreases the sharpness of highlights in the rendered results.

objects imaged in three different indoor and outdoor environments from multiple angles (approximately 18) using a tripod at two different heights. Along with the high-dynamic-range (HDR) images, the data set contains HDR illumination maps acquired using multiple images of a steel ball, and ground-truth 3D models of the objects acquired using a laser light-stripe range finder and manually finished.

Our reflectance estimates are all shown in Fig. 14. Each row shows the results for one object, while the columns correspond to the (pictured) illumination environment. Note that the results are quite consistent across the environments though there exists a slight white-balancing issue for the hall scene leading to some color shifting of the reflectance estimates. Also note that the pig object was purely specular when the hall environment images were taken. It was then sprayed with a diffuser. The change in its reflectance can be seen in the estimates as well. As discussed in Section 5.3, we use a relatively small ( $128 \times 64$ ) illumination environment in our implementation to accelerate running times. For more diffuse objects, this size is sufficient, but in the case of the highly specular objects, necessary detail is lost. As a consequence, the reflectance has absorbed

Fig. 11 compares one of our results with the result if a Lambertian reflectance model is used in place of the DSBRDF model described in Section 4.2. Each row of the figure is an additional viewpoint. On the left we show the rendered results of both methods compared with the observed scene. On the right we show the geometry

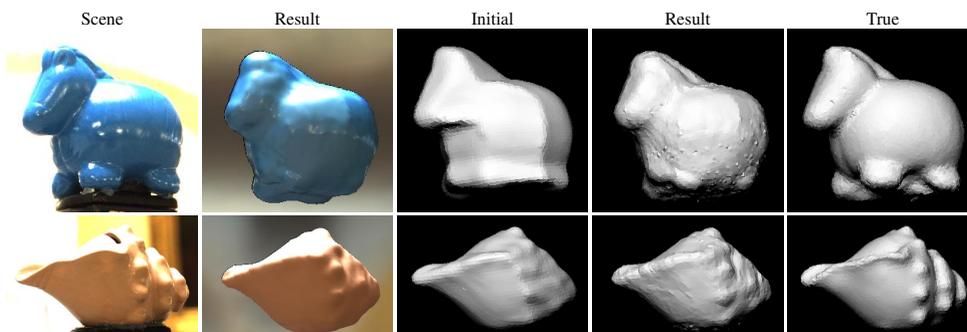


Fig. 15: **Real-world results - hall.** results for the entrance hall environment.

this missing data as error. The shell and horse reflectances are quite plausible and very consistent across the different environments, as can be seen in the re-rendered results, which we discuss next.

The first environment, shown in Fig. 12, is inside a home. Though the windows provide the dominant source of light, within the home are several additional light sources. Despite its complexity, the rich detail of the shell geometry is nicely carved away. Although the pig object frequently exhibits global illumination effects (shadows and interreflections), its 13 views give sufficient context to carve out the concave ears, and feet quite well. The bumpy texture is an artifact of our optimization formulation. The error function is in terms of the facet orientations which are expressed as a complex function of the actual free parameters, the vertex locations. As shown in Fig. 13, these perturbations can be removed by post-processing [35]. Future work could also investigate additional smoothing constraints weighted by the variation in local appearance.

The second illumination environment, shown in Fig. 15, is a large hall. It has a modest amount of natural light coming from the top, but is primarily illuminated by several lights placed evenly throughout the environment. The neck, mane and feet of the horse object are carved away nicely. Similarly, the bumps, and tight ridge of the shell have become much more detailed and accurate.

The final environment, shown in Fig. 16, is outside a wooded home. Overall the results in this environment are quite strong. One noteworthy exception is the milk bottle. As mentioned above, the estimated reflectance has absorbed the inaccuracies of the illumination map coming from acquisition and shrinking. This error leads to a lower likelihood, effectively increasing the prior weights, which is then realized as an overly smooth final result. The toy horse result shows detail in the top of the head not captured, even by the laser scanner. The feet, and neck concavities are carved away nicely as well.

## 6 CONCLUSION

**I**N this work we have presented two probabilistic methods to jointly estimate the geometry and non-trivial reflectance of an object situated in complex, natural illumination. Instead of making simplifying assumptions about the illumination or reflectance we have shown how to use

the complexity to our advantage. In the case where we have only a single image, our first method showed how unique scene regions, when reflected off the object, give strong cues about the local geometry, while careful priors can be used to reduce the varying degrees of ambiguity elsewhere. Our second method showed how to combine multiple observations, enabling additional views to act as further constraints on the local geometry. By fully exploiting the rich signal of real-world illumination reflected by real-world reflectance we have derived methods that recover geometry in the wild, without any expensive equipment.

## ACKNOWLEDGMENTS

This work was supported by the Office of Naval Research grants N00014-11-1-0099 and N00014-14-1-0316, and the National Science Foundation awards IIS-0746717, IIS-0964420, IIS-1353235 and IIS-1421094.

## REFERENCES

- [1] J.-D. Durou, M. Falcone, and M. Sagona, "Numerical Methods for Shape-from-Shading: A New Survey with Benchmarks," *Computer Vision and Image Understanding*, vol. 109, no. 1, pp. 22–43, Jan. 2008.
- [2] R. Zhang, P.-S. Tsai, J. E. Cryer, and M. Shah, "Shape from Shading: A Survey," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 21, no. 8, pp. 690–706, Aug. 1999.
- [3] S. Seitz, J. Diebel, D. Scharstein, R. Szeliski, and B. Curless, "A Comparison and Evaluation of Multi-View Stereo Reconstruction Algorithms," in *IEEE Int'l Conf. on Computer Vision and Pattern Recognition*, 2006, pp. 1–8.
- [4] D. B. Goldman, B. Curless, A. Hertzmann, and S. M. Seitz, "Shape and Spatially-Varying BRDFs from Photometric Stereo," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 32, no. 6, pp. 1060–71, Jun. 2010.
- [5] N. G. Alldrin, T. Zickler, and D. J. Kriegman, "Photometric stereo with non-parametric and spatially-varying reflectance," in *IEEE Int'l Conf. on Computer Vision and Pattern Recognition*, 2008, pp. 1–8.
- [6] N. G. Alldrin and D. J. Kriegman, "Toward Reconstructing Surfaces With Arbitrary Isotropic Reflectance : A Stratified Photometric Stereo Approach," in *IEEE Int'l Conf. on Computer Vision*, 2007, pp. 1–8.
- [7] A. Hertzmann and S. M. Seitz, "Example-Based Photometric Stereo: Shape Reconstruction with General, Varying BRDFs," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 27, no. 8, pp. 1254–64, Aug. 2005.
- [8] R. Huang and W. Smith, "Shape-from-Shading Under Complex Natural Illumination," in *IEEE Int'l Conf. on Image Processing*, 2011, pp. 13–16.
- [9] M. K. Johnson and E. H. Adelson, "Shape Estimation in Natural Illumination," in *IEEE Int'l Conf. on Computer Vision and Pattern Recognition*, 2011, pp. 1–8.

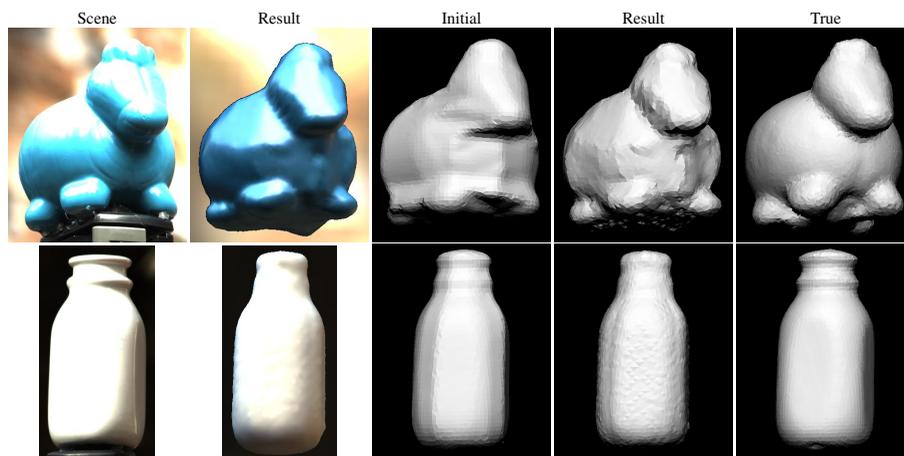


Fig. 16: **Real-world results - outdoor.** Results for the outdoor environment.

[10] R. Basri, D. Jacobs, and I. Kemelmacher, "Photometric Stereo with General, Unknown Lighting," *Int'l Journal of Computer Vision*, vol. 72, no. 3, pp. 239–257, Jun. 2006.

[11] J. Barron and J. Malik, "Color Constancy, Intrinsic Images, and Shape Estimation," in *European Conf. on Computer Vision*, 2012.

[12] —, "Shape, Albedo, and Illumination from a Single Image of an Unknown Object," in *IEEE Int'l Conf. on Computer Vision and Pattern Recognition*, 2012, pp. 334–341.

[13] Y. Adato, Y. Vasilyev, T. Zickler, and O. Ben-Shahar, "Shape from specular flow," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 32, no. 11, pp. 2054–70, Nov. 2010.

[14] M. F. Tappen, "Recovering Shape from a Single Image of a Mirrored Surface from Curvature Constraints," *IEEE Int'l Conf. on Computer Vision and Pattern Recognition*, pp. 2545–2552, Jun. 2011.

[15] S. Seitz and C. Dyer, "Photorealistic Scene Reconstruction by Voxel Coloring," in *IEEE Int'l Conf. on Computer Vision and Pattern Recognition*, 1997, pp. 1067–1073.

[16] K. Kutulakos and S. Seitz, "A Theory of Shape by Space Carving," *Int'l Journal of Computer Vision*, vol. 38, no. 3, pp. 199–218, 2000.

[17] G. Vogiatzis, P. Torr, and R. Cipolla, "Multi-View Stereo via Volumetric Graph-Cuts," in *IEEE Int'l Conf. on Computer Vision and Pattern Recognition*, 2005, pp. 391–398.

[18] J. Pons, R. Keriven, and O. Faugeras, "Multi-view Stereo Reconstruction and Scene Flow Estimation with a Global Image-Based Matching Score," *Int'l Journal of Computer Vision*, pp. 1–24, 2007.

[19] H. Jin, D. Cremers, A. Yezzi, and S. Soatto, "Shedding Light on Stereoscopic Segmentation," in *IEEE Int'l Conf. on Computer Vision and Pattern Recognition*, 2004, pp. 36–42.

[20] H. Jin, S. Soatto, and A. Yezzi, "Multi-View Stereo Reconstruction of Dense Shape and Complex Appearance," *Int'l Journal of Computer Vision*, vol. 63, no. 3, pp. 175–189, Apr. 2005.

[21] C. Hernandez, G. Vogiatzis, and R. Cipolla, "Multiview Photometric Stereo," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 30, no. 3, pp. 548–554, 2008.

[22] A. Treuille, A. Hertzmann, and S. Seitz, "Example-Based Stereo with General BRDFs," in *European Conf. on Computer Vision*, 2004, pp. 457–469.

[23] K. Ikeuchi and B. K. P. Horn, "Numerical Shape from Shading and Occluding Boundaries," *Artificial Intelligence*, 1981.

[24] K. Nishino, "Directional Statistics BRDF Model," in *IEEE Int'l Conf. on Computer Vision*, 2009, pp. 476–483.

[25] K. Nishino and S. Lombardi, "Directional Statistics-based Reflectance Model for Isotropic Bidirectional Reflectance Distribution Functions," *Journal of Optical Society America, A*, vol. 28, no. 1, pp. 8–18, Jan. 2011.

[26] S. Lombardi and K. Nishino, "Reflectance and Natural Illumination from a Single Image," in *European Conf. on Computer Vision*, Oct. 2012, pp. 582–595.

[27] S. Rusinkiewicz, "A New Change of Variables for Efficient BRDF Representation," in *Eurographics Workshop on Rendering*, 1998, pp. 11–22.

[28] W. Matusik, H. Pfister, M. Brand, and L. McMillan, "A data-driven reflectance model," *ACM Trans. on Graphics*, vol. 22, no. 3, pp. 759–769, Jul. 2003.

[29] Debevec, Paul, "Light Probe Image Gallery," <http://www.pauldebevec.com/Probes/>, 2012.

[30] Y. Boykov, O. Veksler, and R. Zabih, "Fast Approximate Energy Minimization via Graph Cuts," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 23, no. 11, pp. 1222–1239, Nov. 2001.

[31] R. Frankot and R. Chellappa, "A Method for Enforcing Integrability in Shape from Shading Algorithms," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 10, no. 4, pp. 439–451, Jul. 1988.

[32] A. Laurentini, "The Visual Hull Concept for Silhouette-Based Image Understanding," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 16, no. 2, 1994.

[33] M. Kazhdan, M. Bolitho, and H. Hoppe, "Poisson Surface Reconstruction," in *Eurographics Symposium on Geometry Processing*, 2006, pp. 61–70.

[34] S. Lombardi and K. Nishino, "Single Image Multimaterial Estimation," in *IEEE Int'l Conf. on Computer Vision and Pattern Recognition*, 2011, pp. 238–245.

[35] M. Desbrun, M. Meyer, P. Schröder, and A. H. Barr, "Implicit Fairing of Irregular Meshes Using Diffusion and Curvature Flow," in *ACM SIGGRAPH*, 1999, pp. 317–324.



**Geoffrey Oxholm** received his Bachelors of Science in 2004 from Lafayette college. He received his Masters and PhD from the Department of Computing in the College of Computing and Informatics at Drexel University in 2009 and 2014, respectively. He is now a member of the Creative Technologies Lab of Adobe Research.



**Ko Nishino** is an associate professor in the Department of Computing in the College of Computing and Informatics at Drexel University. He is also an adjunct associate professor in the Computer and Information Science Department of the University of Pennsylvania and a visiting associate professor of Osaka University. He received a B.E. and an M.E. in Information and Communication Engineering in 1997 and 1999, respectively, and a PhD in Computer Science in 2002, all from The

University of Tokyo. Before joining Drexel University in 2005, he was a Postdoctoral Research Scientist in the Computer Science Department at Columbia University. His primary research interests lie in computer vision and include appearance modeling and synthesis, geometry processing, and video analysis. His work on modeling eye reflections received considerable media attention including articles in New York Times, Newsweek, and NewScientist. He received the NSF CAREER award in 2008.