Multiview Shape and Reflectance from Natural Illumination

Geoffrey Oxholm and Ko Nishino
Department of Computing
Drexel University, Philadelphia, PA 19104, USA
{gao25,kon}@drexel.edu

Abstract

The world is full of objects with complex reflectances, situated in complex illumination environments. Past work on full 3D geometry recovery, however, has tried to handle this complexity by framing it into simplistic models of reflectance (Lambetian, mirrored, or diffuse plus specular) or illumination (one or more point light sources). Though there has been some recent progress in directly utilizing such complexities for recovering a single view geometry, it is not clear how such single-view methods can be extended to reconstruct the full geometry. To this end, we derive a probabilistic geometry estimation method that fully exploits the rich signal embedded in complex appearance. Though each observation provides partial and unreliable information, we show how to estimate the reflectance responsible for the diverse appearance, and unite the orientation cues embedded in each observation to reconstruct the underlying geometry. We demonstrate the effectiveness of our method on synthetic and real-world objects. The results show that our method performs accurately across a wide range of real-world environments and reflectances that lies between the extremes that have been the focus of past work.

1. Introduction

The beauty of our world is due in part to the interaction of complex reflectances and rich illumination environments. Past work on full 3D geometry estimation has focused on compressing this complexity to fit simplistic models. Reflectance is often assumed to be Lambertian, mirrored, or diffuse plus specular; and illumination is assumed to be simple, often a single point light source. Recently there has been some progress in directly exploiting the complexity of real-world appearance to recover a single-view geometry estimate [16]. It is not clear, however, how such single-view methods can be extended to reconstruct the full geometry.

The foundation of multi-view stereo is the notion of *photometric consistency*. For Lambertion objects, for example, multiple observations enable us to divide up the object space and test each voxel for consistency with the observation images [20]. That is, due to the viewpoint independent appearance of Lambertian materials, a voxel that truly contains the surface must have the same appearance when projected into each of the images. If the object has a more complex reflectance, by controlling the illumination we may deduce per-pixel orientations (surface normals) for each of the observations. We can then impose an analogous *orientation consistency* metric [5]. Though there have been many other ways to express the notion of consistency (some of which are discussed below) each is made possible through a simplistic model of illumination or reflectance that excludes many real-world settings.

In this paper, we introduce a method that uses what we can directly (and passively) acquire (a dozen calibrated images of the object and a panorama of the illumination environment) to estimate what we cannot (the reflectance properties and full 3D geometry of the object). The main contribution of this work is a probabilistic 3D geometry and reflectance estimation method that fully exploits the complexity of non-trivial reflectance and non-trivial illumination. We use a triangular mesh model to link the observations, so that orientation cues embedded in each observation can be combined to constrain the position of each facet.

The appearance of a pixel in an image provides a multimodal distribution of possible orientations, the shape of which depends on the illumination environment and the reflectance properties of the object. Those pixels reflecting unique scene components (like the sun) provide stronger constraints, while those reflecting less descriptive components (like the sky or a tree) provide weaker constraints. Such weak constraints, however, become strong when the orientation distributions of multiple observations corroborate a tighter range of orientations. Our overall method is to jointly optimize reflectance and shape by keeping one fixed as the other is optimized. To model the reflectance we use the Directional Statistics BRDF model [12, 14]. We use silhouette intersection as a starting point.

We test our method quantitatively on synthetic data and real-world objects. The synthetic experiments show that our

method performs well in a wide range of real-world illumination environments and for a wide range of real-world reflectances. To perform quantitative real-world evaluation, we introduce a new data-set containing four objects of which we have acquired ground-truth geometry and images in three diverse indoor and outdoor locations.

2. Related work

The field of multi-view geometry estimation has received a great deal of attention over the years [19]. Here we will focus only on those methods that have some important similarity to our method.

The concept of photometric consistency (discussed above) has played a role in many methods [10, 20, 22]. Pons *et al*. [17] extended the notion using patch-based comparisons instead of a simple single-voxel intensity comparison giving flexibility to changes in radiance. Jin *et al*. [6, 7] moved the concept beyond Lambertian reflectances to a diffuse plus specular model by measuring the rank of the radiance tensor field across many (~ 40) images. In this work we go further by working under complex natural illumination, with the more general DSBRDF reflectance model [12, 14], and with a more sparse (~ 12) set of images.

Hernàndez *et al*. [5] use photometric stereo to convert each observation location into a reliable geometry observation in the form of a surface normal field by taking several pictures under a moving point light source. Treuille *et al*. [21] allow for arbitrary BRDFs, but assume that the reflectance is known by observing the appearance of a sphere of the same material. By comparing the change in appearance of each point on the object with that of each point on the sphere, they also arrive at a reliable surface normal field for each observation location. Both methods then estimate the 3D geometry that is consistent with the newly created geometry observations. Though our overall approach is to also use orientation clues from the observations, we do so without controlling the illumination, and without such limiting assumptions about the reflectance properties of the object. In order to extract orientation cues in this setting, we turn to recent work in shape from shading.

In the Lambertian case, significant progress has been made due to the observation that natural illumination may be sufficiently approximated with the parametric spherical harmonics representation. Work by Johnson and Adelson [8] uses this to deduce the surface orientations, and albedo of an object. Barron and Malik [2, 3] showed further that the illumination parameters and specular highlights of the object can also estimated. For complex reflectance, however, no such approximation of the illumination can be made. The extreme case of mirrored reflectance, for example, directly reflects the inherently non-parametric environment surrounding the object.

In the case of such mirrored reflectance, Adato *et al*. [1] observe the flow of the reflected, yet unknown, illumination environment for a known relative movement of the environment to deduce the shape of the object. 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.

For the non-Lambertian case, in recent work we showed how the appearance of a point on the object can give only a multi-modal, non-parametric distribution of possible orientations [16]. Strong priors can then be imposed to reduce the ambiguity, and estimate the single-view geometry. Our primary contribution in this work is a canonical way to combine these complex distributions from separate observations to extract the geometry (and reflectance) of the object.

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 $\boldsymbol{L}$, the object material has an isotropic reflectance function, and that it has been segmented from the background. We also assume that we have multiple images $\boldsymbol{\mathcal{I}} = \{\boldsymbol{I}^1, \ldots, \boldsymbol{I}^M\}$ from calibrated cameras. 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 $\boldsymbol{G}$, and reflectance $\boldsymbol{\Psi}$. We formulate this as the maximum a posteriori (MAP) estimate of the posterior distribution

$$p(\boldsymbol{G}, \boldsymbol{\Psi}|\boldsymbol{\mathcal{I}}) \propto p(\boldsymbol{\mathcal{I}}|\boldsymbol{G}, \boldsymbol{\Psi})p(\boldsymbol{G})p(\boldsymbol{\Psi}) \,, \tag{1}$$

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

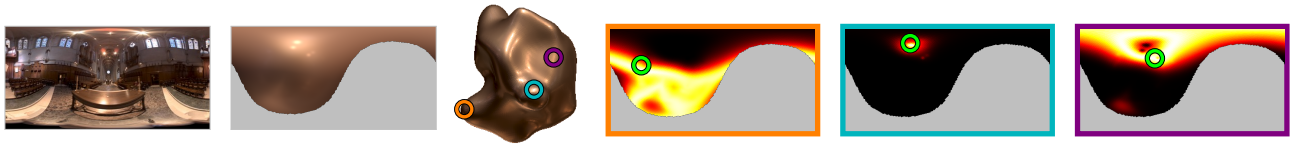
In order to find the geometry and reflectance that optimize the posterior, we adopt an iterative approach, keeping one fixed while estimating the other. In the next sections we describe precisely how we represent the geometry $\boldsymbol{G}$, and the reflectance $\boldsymbol{\Psi}$ and how we estimate them, we then discuss the overall optimization strategy and some implementation considerations.

3.1. Shape from reflectance

First we will describe our method for estimating the shape of the object using images of the object, the illumination environment, and the current reflectance $\boldsymbol{\Psi}$ as input.

3.1.1 What each image tells us

Let us begin with a single pixel $\boldsymbol{I}_x$ in a single image $\boldsymbol{I} \in \boldsymbol{\mathcal{I}}$. For example, consider the pixel of Fig. 1c circled in orange. The appearance of this pixel is due to the reflectance $\boldsymbol{\Psi}$, the



(a) Illumination (b) Reflectance map (c) Observation (d) Likelihoods for color-coded example pixels (brighter is more likely)

Figure 1. **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 non-parametric distribution of orientations (d). The green circles denote the true orientation of the corresponding pixels (sorted left to right, and color coded).

illumination environment $\boldsymbol{L}$, and the underlying orientation $\boldsymbol{N}_x$ of the corresponding surface point, with some added Gaussian noise of uniform variance,

$$\boldsymbol{I}_x = \boldsymbol{E}(\boldsymbol{\Psi}, \boldsymbol{L}, \boldsymbol{N}_x) + \mathcal{N}(\boldsymbol{0}, \sigma^2) \; . \tag{2}$$

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

$$p(\boldsymbol{I}_x|\boldsymbol{N}_x) = \mathcal{N}\big(\ln(\boldsymbol{I}_x)|\ln(\boldsymbol{E}_x), \sigma^2\big) \; . \tag{3}$$

The orientation $\boldsymbol{N}_x$ (the only free parameter) 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 $\boldsymbol{N}_x$ (which itself can be expressed in 2D spherical coordinates $\boldsymbol{N}_x = (\theta_x, \phi_x)$), 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.

3.1.2 A unified coordinate frame

In the single image (SfS) case, the next step would be to determine the most likely orientations for each pixel. Ambiguous distributions like the ones in Fig. 1d make this problem under-constrained. Adding a smoothing constraint [16] can help, but will fail when whole regions have ambiguous appearances. Learned constraints on surface curvature [3] address this, but cannot account for depth-discontinuities.

It is here that multiple observations become essential since each observation serves as separate constraint on the distribution of possible 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. 2 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 circumscribed by a coarse geometry estimate. If we take a single point p, and project it into each of the observations images $\boldsymbol{I}^m \in \boldsymbol{\mathcal{I}}$ we may then compute the likelihood density for that point as a product of the separate observations

$$p(\boldsymbol{\mathcal{I}}_p|\boldsymbol{N}_p) = \prod_{m \in \Omega_p} p(\boldsymbol{I}_p^m|\boldsymbol{N}_p) \; , \tag{4}$$

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

Two examples are shown in Fig. 2. 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 $\boldsymbol{I}^1$ and $\boldsymbol{I}^4$. The orientation distributions for these two observations therefore overlap nicely, and the resulting distribution for the point is concentrated, with a small (bright) region.

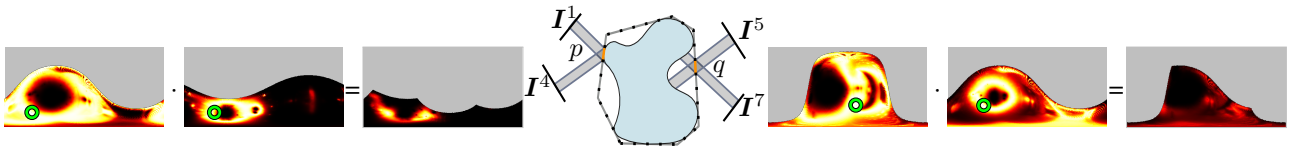
On the right side of the example mesh, we see the projection of a point q that is far removed from the true surface. The consequence of this is that the two imaged appearances attributed to this point are actually of different points of the real object. Since the imaged geometry for image $\boldsymbol{I}^5$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

3.1.3 Surface patches

Now that we have seen how to unite multiple observations to derive tighter orientation distributions, we may finally turn our attention to recovering a full 3D model. By focusing on the facets of the model $f \in \boldsymbol{G}$, we provide a way to use orientation cues to deduce the full 3D 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 $J = 6$),

$$p(\boldsymbol{I}_f^m|\boldsymbol{N}_f) = c \sum_{j=1}^{J} w_j \cdot p(\boldsymbol{I}_{f,j}^m|\boldsymbol{N}_f) \; , \tag{5}$$



(a) Observations that agree (b) Mesh, true object, and observations (c) Observations that disagree

Figure 2. **Nonparametric orientation consistency** When a point on the mesh (dashed) is accurate, 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.

where the weights w_j ($\sum w_j = 1$) are higher for samples near the center of the facet, and c is to ensure the distribution integrates to one. 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, $\boldsymbol{I}^m_{f,j}$ indicates a specific pixel—the j^{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. 4),

$$p(\boldsymbol{\mathcal{I}}|\boldsymbol{N}_f) = \prod_{m \in \Omega_f} p(\boldsymbol{I}^m_f|\boldsymbol{N}_f) \,. \tag{6}$$

3.1.4 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,

$$p(\boldsymbol{\mathcal{I}}|\boldsymbol{G}) = \prod_{f \in \boldsymbol{G}} p(\boldsymbol{\mathcal{I}}|\boldsymbol{N}_f) \,. \tag{7}$$

Finally, we place three priors on the mesh itself, $p(\boldsymbol{G}) = p_c(\boldsymbol{G})p_a(\boldsymbol{G})p_e(\boldsymbol{G})$. The first prior is inspired by recent work on minimal surface constraints [3]. To propagate the shape of accurate regions to those far removed from the true surface (as in Fig. 2c), we encourage the local curvature to be constant. This is approximated as the variance of the angles between the normal of each facet and those of the facets a certain distance away

$$p_c(\boldsymbol{G}) \propto \prod_{f \in \boldsymbol{G}} \exp\left\{\beta_c \mathrm{Var}\left[\arccos(\boldsymbol{N}_f \cdot \boldsymbol{N}_{f,n})\right]\right\} , \tag{8}$$

where β_c controls the strength of the prior. The set of normals $\{\boldsymbol{N}_{f,n}\}$ over which the variance is computed contains those facets that are a certain distance from the facet itself. 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. In section 4, we describe our use of this prior.

The remaining two priors are due to our implicit assumptions about the triangles that make up the mesh. Our area prior $p_a(\boldsymbol{G})$ helps ensure that the triangles all have roughly the same size by comparing their area $A(f)$ to their initial mean a and variance σ_a^2,

$$p_a(\boldsymbol{G}) = \prod_{f \in \boldsymbol{G}} \mathcal{N}(A(f)|a, \beta_a \sigma_a^2) \,, \tag{9}$$

where β_a controls the strength of the prior. Our final prior $p_e(\boldsymbol{G})$ additionally 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,

$$p_e(\boldsymbol{G}) \propto \prod_{f \in \boldsymbol{G}} \exp\left\{\beta_e \mathrm{Var}\left[f_e\right]\right\} \,, \tag{10}$$

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

3.1.5 Parameterizing the distribution

Recall that the facet likelihoods $p(\boldsymbol{\mathcal{I}}_f|\boldsymbol{N}_f)$ are nonparametric in that they depend on the inherently nonparametric illumination environment. Because of this, a direct optimization is intractable (the visualizations in Fig. 2 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 $\{\boldsymbol{N}^l\}$ 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 κ_l is proportional to the probability of the corresponding orientation $\boldsymbol{N}^l$ as computed by Eq. 6 (in our case $\kappa_l = 200 \cdot p(\boldsymbol{N}_f)$),

$$p_{\text{approx}}(\boldsymbol{\mathcal{I}}_f|\boldsymbol{N}_f) \propto \sum_{l=1}^{L} C(\kappa_l) \exp\left\{\kappa_l \boldsymbol{N}^l \cdot \boldsymbol{N}_f\right\} \tag{11}$$

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

This formulation has several benefits. 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. Additionally, the value of L may be adjusted to avoid local maxima. In our case we set $L = 1024$ and increase it by 128 each iteration.

3.2. Reflectance from shape

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

3.2.1 The directional statistics BRDF model

To model the reflectance function, we adopt the Directional Statistics Bidirectional Reflectance Distribution Function (DSBRF) model, introduced by Nishino [14, 15] and later extended by Lombardi and Nishino [12] 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 $\boldsymbol{E}(\boldsymbol{\Psi}, \boldsymbol{L}, \boldsymbol{N}_x)$ is

$$\boldsymbol{E}_x = \int \varrho(t(\omega_i, \omega_o); \boldsymbol{\Psi}) \mathbf{L}(\omega_i) \max(0, \boldsymbol{N}_{\mathbf{x}} \cdot \omega_i) d\omega_i, \quad (12)$$

where t is a function that transforms the incoming ω_i and outgoing ω_o angles into the alternate BRDF parameterization variables θ_d and θ_h. The reflectance function is expressed as a sum of lobes

$$\varrho^{(\lambda)}(\theta_d, \theta_h; \kappa^{(\lambda)}, \gamma^{(\lambda)}) = \sum_r \exp\left\{\kappa^{(r,\lambda)}(\theta_d) \cos^{\gamma^{(r,\lambda)}(\theta_d)}(\theta_h)\right\} - 1\,, \quad (13)$$

where the halfway vector parameterization (i.e., (θ_h, ϕ_h) for the halfway vector and (θ_d, ϕ_d) for the difference vector) [18] 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 θ_d for a particular color channel λ. These curves are modeled as a log-linear combination of data-driven basis functions,

$$\kappa^{(r,\lambda)}(\theta_d) = \exp\left\{b_\mu(\theta_d; \kappa, r, \lambda) + \sum_i \psi_i b_i(\theta_d; \kappa, r, \lambda)\right\},$$

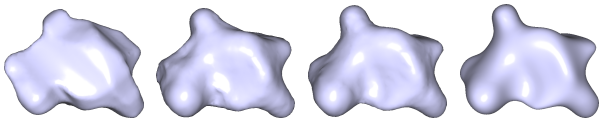
$$\gamma^{(r,\lambda)}(\theta_d) = \exp\left\{b_\mu(\theta_d; \gamma, r, \lambda) + \sum_i \psi_i b_i(\theta_d; \gamma, r, \lambda)\right\},$$

where b_μ is the mean basis function, b_i is the i^{th} basis function, and ψ_i are the DSBRDF coefficients. We may compute these basis functions from a set of measured reflectance functions using functional principal component analysis (FPCA).

3.2.2 Probabilistic reflectace estimation

In order to estimate the parameters $\boldsymbol{\Psi}$ we continue with our probabilistic formulation of Eq. 1. Here, the likelihood is the same as above, though the geometry, and hence the surface orientations of the facets $\boldsymbol{N}_f$, are kept fixed,

$$p(\boldsymbol{\mathcal{I}}|\boldsymbol{\Psi}) = \prod_{f \in \boldsymbol{G}} \prod_{m \in \Omega_f} \mathcal{N}\big(\ln(\boldsymbol{I}_f^m) | \ln(\boldsymbol{E}_f^m), \sigma^2\big)\,, \quad (14)$$



(a) Initial (b) Mid-way (c) Final (d) True

Figure 3. **Shape optimization iterations**

where Ω_f is again the set of images in which facet f appears, and $\boldsymbol{I}_f^m$ refers to the appearance at the center of the facet in image m, and $\boldsymbol{E}_f^m$ refers to its predicted irradiance.

We utilize the prior by Lombardi and Nishino [12], which encourages the coefficients $\psi_i \in \boldsymbol{\Psi}$ of the eigenfunctions to be within the distribution of observed reflectances,

$$p(\boldsymbol{\Psi}) \sim \mathcal{N}(0, \beta_\Psi \boldsymbol{\Sigma}_\Psi), \quad (15)$$

where the covariance $\boldsymbol{\Sigma}_\Psi$ is computed from the MERL database [13], and the scalar β_Ψ controls the prior strength.

4. Optimization and implementation

Our overall optimization scheme alternates between computing the Gaussian noise variance σ^2, and estimating the the maximum a posteriori (MAP) estimate of the reflectance parameters $\boldsymbol{\Psi}$ and then geometry $\boldsymbol{G}$. This three-step optimization framework is iterated until convergence, typically around 6 iterations. To find the MAP estimate of the reflectance parameters $\boldsymbol{\Psi}$ and geometry $\boldsymbol{G}$ we minimize the corresponding log-posteriors using gradient descent. So that a single set of prior weights can be used, all images are scaled by a constant factor so that the mean intensity of the illumination environment is 1.

To bootstrap the process, the first step is to 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 [11] to initialize our geometry estimate. The mesh is then re-triangulated using the Poisson reconstruction [9], and small triangles are collapsed to help standardize the area of the triangles. The result of this step is shown in Fig. 3a. With an initial geometry estimate in place, we then perform the first reflectance estimation iteration. The prior weight β_Ψ is set to 2^{-3}.

When refining 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, accelerating the computation of Eq. 3.

The geometry estimation iterations follow a coarse-to-fine approach in that the scale of curvature-based smoothing prior defined in Eq. 8 is iteratively reduced by picking the set of facets over which the variance is computed. Initially,

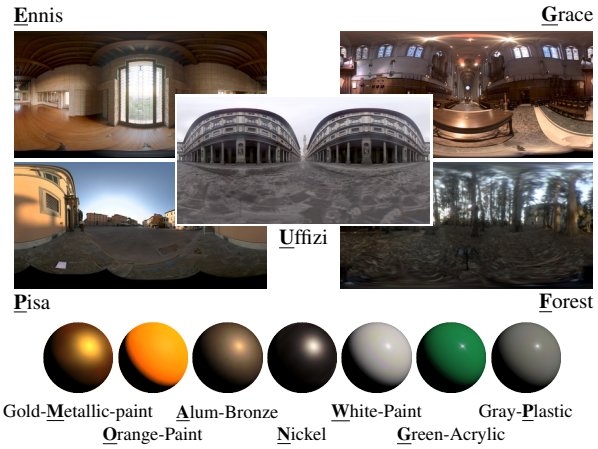


Figure 4. **Synthetic data.** Our synthetic data are formed by rendering 10 shapes [8] with 7 real-world BRDFs (bottom) [13] under 5 real-world natural illumination environments (top) [4].

the set is chosen to contain the facets that are a distance of 3 triangles away. In subsequent iterations, this number is decreased, until only the facets sharing a point with the triangle influence its local curvature constraint. The prior weights are set to $\beta_c = 2$, $\beta_a = 1$ and $\beta_e = 0.5$.

Throughout the optimization process we take into account occlusion when computing which images Ω_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° 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. Finally, in order to keep volume and centering near constant, we center and resize the result after each iteration.

5. 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.

5.1. Synthetic data evaluation

The ideal illumination environment in which to estimate the geometry of a Lambertian object is one with a few, differently colored point light sources. The resulting appearance will depend exclusively on the relative angles between the surface point orientations and the lights. Such an illumination environment would, however, provide a hopelessly sparse set of orientation cues for a mirrored object, as only a few points on the object would reflect any light at all. For a mirrored object, on the other hand, the ideal illumination environment is one in which each orientation reflects a unique illumination value. Lambertian reflectance, in this case, would hopelessly blur such detail. Real-word illuminations and reflectances, however, lie in between these extremes. In order to test the role of reflectance and illumination in shape estimation, we have performed hundreds of experiments with a wide range of real-world environments and BRDFs.

Fig. 4 gives an overview of our synthetic data. Note the bold letters, as they are used as indexes in Table 1. Each of the 10 blobs from the Blobby Shapes database [8] is rendered in 5 publicly available illumination environments [4] with 7 different measured BRDFs from the MERL database [13]. When training the reflectance model and prior, the ground-truth BRDF is omitted to ensure a fair evaluation.

Table 1 gives an overview of our geometry and reflectance estimates when 9 images are used. On the left we analyze the geometry error for each of the 35 different illumination and reflectance combinations (averaged over the 10 shapes). Each of the rows and columns correspond to the illumination environments and reflectances shown in Fig. 4, respectively. The last row, and column show averages for the corresponding illumination or reflectance.

Reflectances

Environments	M	O	A	N	W	G	P	mean
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

Environments	M	O	A	N	W	G	P	mean
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

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. 4. 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.

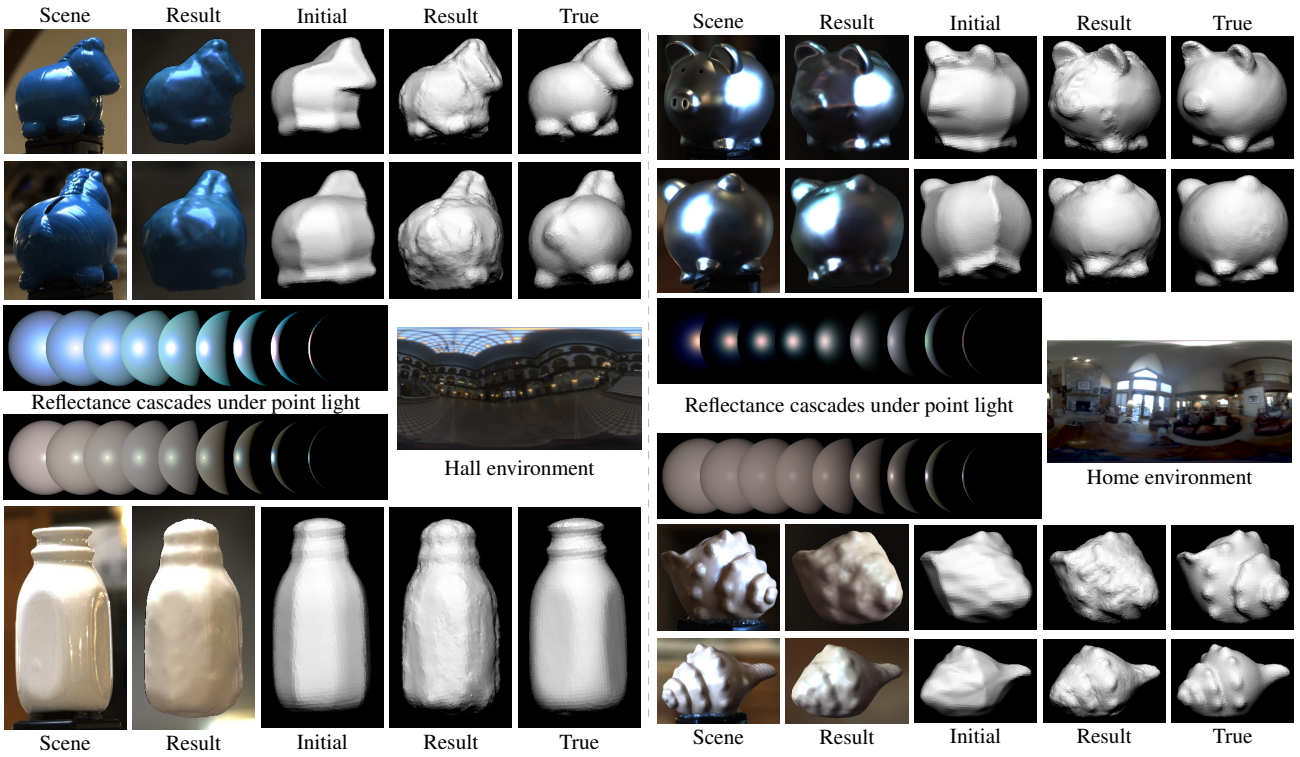


Figure 5. **Real-world novel-view results** The first two columns in each section compare full appearance, while the last three are rendered with a diffuse model to highlight geometric differences with the initial estimate and ground truth. The center images show the captured illumination environment and the recovered reflectances rendered on spheres with a moving point light.

The consistency of results within each column of Table 1b 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 consistenly more accurately estimated.

Table 1a shows the geometry results. As a baseline, these numbers should be compared with the mean initial RMS error of 1.19%, so even in the worst case the error is being reduced significantly. The worst geometry estimation result, with a RMS error of 0.98%, 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 0.50%. Only one reflectance is challenging in this environment—Nickel (N), which has only a weak diffuse component. The best reflectance, Gold-Metallic-Paint (M), has the best of both worlds—strong diffuse, and moderate specular components. This enables the appearance to capture both low-frequency and high-frequency detail of the illumination.

5.2. Real-world data evaluation

To quantitatively evaluate our method on real-world objects we introduce a new data-set.[1] The data-set contains four 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.

Fig. 5 shows several results. For each result one or two novel viewpoint renderings are shown, (*i.e.* viewpoints that were not used in the estimation process). Diffuse renderings of the initial, final, and true geometry are included for visual comparison. 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 has the impact of decreasing the sharpness of highlights in the rendered results. Recall from Section 4 that we ignore observations at grazing angles angles due to irregular reflectance properties, and po-

[1] Available at http://cs.drexel.edu/~kon/multinatgeom

tentially inaccurate object segmentation. Consequently, the reflectance estimates exhibit some inaccuracy at such angles. Overall the estimated reflectances are quite plausible, as can be seen in the cascades and the re-rendered results.

The first illumination environment, a large hall, has a modest amount of natural light coming from the top of the scene, but is primarily illuminated by several lights placed evenly throughout the environment. The milk bottle, though a fairly simple shape, has the most challenging reflectance due to its highly reflective glaze. Since the light at the top of the scene is fairly low-frequency, the effect of this is less dominant in the upward-facing portions of the object. Its initial RMS error is 0.92% and its final error is 0.76%, a 17% reduction. In the toy horse example it is clear that large regions included in the original visual hull have been nicely carved away. Note especially the fine detail in the mane and nose. Its initial error was 1.66% and its final is 1.20%, a 28% reduction.

The second environment is inside a home. Though the windows provide the dominant source of light, within the home are several additional light sources. The many peaks of the shell geometry are revealed in the result, though self-cast shadows (and some interreflection) reduce their clarity somewhat. Note also in the lower view the large strip of erroneous material carved out on the right hand side. Its initial error was 0.77% and its final is 0.73%, a 5% reduction. Despite its complex appearance, the reflectance of the piggy-bank is a nearly ideal one; it filters no colors and is neither too diffuse, nor too specular. Although it frequently exhibits global illumination effects (shadows and interreflections), its 13 views give sufficient context to carve out the concave ears, and feet. Its initial error was 1.60% and its final is 0.81%, a 49% reduction.

6. Conclusion

In this paper we introduced a method to recover the full 3D geometry and reflectance of non-Lambertian objects situated in complex, natural illumination. Instead of framing reflectance and illumination into simplistic models, we have shown how to fully exploit the complexity. Though each observation provides unreliable and incomplete information, by carefully combining them we are able to accurately recover the geometry (and reflectance) of real-world objects. Since this method does not require any expensive equipment, it nicely fills the holes that have been left open by past work in structure from motion and multiview stereo.

Acknowledgements

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

References

[1] Y. Adato, Y. Vasilyev, T. Zickler, and O. Ben-Shahar. Shape from specular flow. *TPAMI*, 32(11):2054–70, Nov. 2010. 2

[2] J. Barron and J. Malik. Color Constancy, Intrinsic Images, and Shape Estimation. In *ECCV*, 2012. 2

[3] J. Barron and J. Malik. Shape, Albedo, and Illumination from a Single Image of an Unknown Object. In *CVPR*, pages 334–341, 2012. 2, 3, 4

[4] P. Debevec. Light Probe Image Gallery. http://www.pauldebevec.com/Probes/, 2012. 6

[5] C. Hernandez, G. Vogiatzis, and R. Cipolla. Multiview Photometric Stereo. *TPAMI*, 30(3):548–554, 2008. 1, 2

[6] H. Jin, D. Cremers, A. Yezzi, and S. Soatto. Shedding Light on Stereoscopic Segmentation. In *CVPR*, pages 36–42, 2004. 2

[7] H. Jin, S. Soatto, and A. Yezzi. Multi-View Stereo Reconstruction of Dense Shape and Complex Appearance. *IJCV*, 63(3):175–189, Apr. 2005. 2

[8] M. Johnson and E. Adelson. Shape Estimation in Natural Illumination. In *CVPR*, pages 1–8, 2011. 2, 6

[9] M. Kazhdan, M. Bolitho, and H. Hoppe. Poisson Surface Reconstruction. In *Eurographics Symposium on Geometry Processing*, pages 61–70, 2006. 5

[10] K. Kutulakos and S. Seitz. A Theory of Shape by Space Carving. *IJCV*, 38(3):199–218, 2000. 2

[11] A. Laurentini. The Visual Hull Concept for Silhouette-Based Image Understanding. *TPAMI*, 16(2), 1994. 5

[12] S. Lombardi and K. Nishino. Reflectance and Naturail ILlumination from a Single Image. In *ECCV*, pages 582–595, Oct. 2012. 1, 2, 5

[13] W. Matusik, H. Pfister, M. Brand, and L. McMillan. A data-driven reflectance model. *ACM Trans. on Graphics*, 22(3):759–769, July 2003. 5, 6

[14] K. Nishino. Directional Statistics BRDF Model. In *ICCV*, pages 476–483, 2009. 1, 2, 5

[15] K. Nishino and S. Lombardi. Directional Statistics-based Reflectance Model for Isotropic Bidirectional Reflectance Distribution Functions. *JOSA*, 28(1):8–18, Jan. 2011. 5

[16] G. Oxholm and K. Nishino. Shape and Reflectance from Natural Illumination. In *ECCV*, pages 528–5414, Oct. 2012. 1, 2, 3

[17] J. Pons, R. Keriven, and O. Faugeras. Multi-view Stereo Reconstruction and Scene Flow Estimation with a Global Image-Based Matching Score. *IJCV*, pages 1–24, 2007. 2

[18] S. Rusinkiewicz. A New Change of Variables for Efficient BRDF Representation. In *Eurographics Workshop on Rendering*, pages 11–22, 1998. 5

[19] S. Seitz, J. Diebel, D. Scharstein, R. Szeliski, and B. Curless. A Comparison and Evaluation of Multi-View Stereo Reconstruction Algorithms. In *CVPR*, pages 1–8, 2006. 2

[20] S. Seitz and C. Dyer. Photorealistic Scene Reconstruction by Voxel Coloring. In *CVPR*, pages 1067–1073, 1997. 1, 2

[21] A. Treuille, A. Hertzmann, and S. Seitz. Example-Based Stereo with General BRDFs. In *ECCV*, pages 457–469, 2004. 2

[22] G. Vogiatzis, P. Torr, and R. Cipolla. Multi-View Stereo via Volumetric Graph-Cuts. In *CVPR*, pages 391–398, 2005. 2