Blind Recovery of Spatially Varying Reflectance from a Single Image

Kevin Karsch David Forsyth
University of Illinois

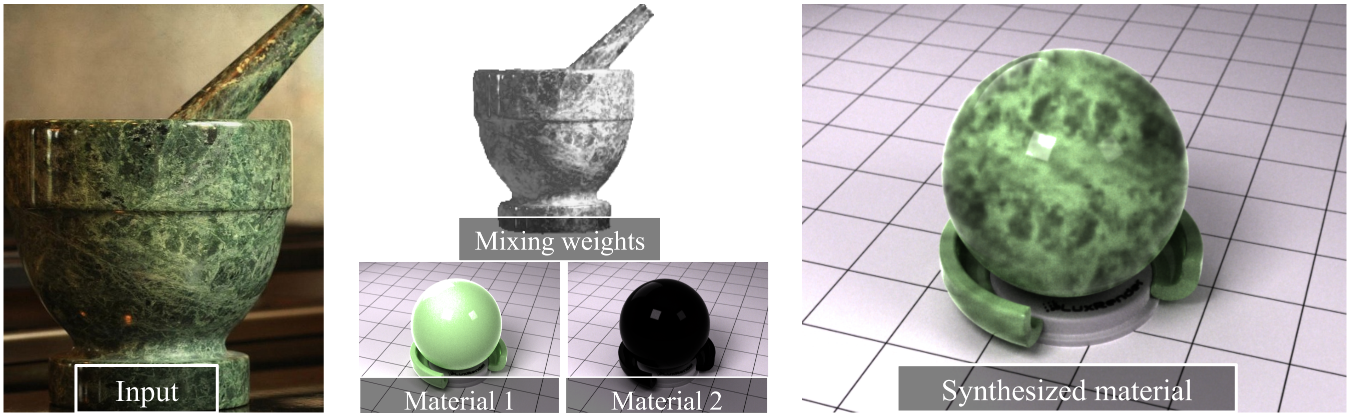

Figure 1: *From a single photograph, our method estimates spatially varying materials (diffuse reflectance and specular parameters). The input image is decomposed into k low-order, parametric materials (Material 1 and 2) and a set of per-pixel material mixing coefficients (Mixing weights); shape and illumination is jointly inferred. This decomposition can be transferred to new shapes (Synthesized material) and also used to generate new materials.*

Abstract

We propose a new technique for estimating spatially varying parametric materials from a single image of an object with unknown shape in unknown illumination. Our method uses a low-order parametric reflectance model, and incorporates strong assumptions about lighting and shape. We develop new priors about how materials mix over space, and jointly infer all of these properties from a single image. This produces a decomposition of an image which corresponds, in one sense, to microscopic features (material reflectance) and macroscopic features (weights defining the mixing properties of materials over space). We have built a large dataset of real objects rendered with different material models under different illumination fields for training and ground truth evaluation. Extensive experiments on both our synthetic dataset images as well as real images show that (a) our method recovers parameters with reasonable accuracy; (b) material parameters recovered by our method give accurate predictions of new renderings of the object; and (c) our low-order reflectance model still provides a good fit to many real-world reflectances.

CR Categories: I.2.10.d [Artificial Intelligence]: Vision and Scene Understanding—Modeling and recovery of physical attributes; I.3.8 [Computer Graphics]: Applications; I.4.8.c [Image Processing and Computer Vision]: Image Models

Keywords: reflectance estimation, shape from shading, material transfer, material modeling

1 Introduction

Humans are quite good at guessing an object's material based on appearance alone [Adelson 2000]. However, material[1] estimation from a single photograph remains a challenging and unsolved problem in computer vision. Appearance is often considered a function of object shape, incident illumination, and surface reflectance, and many solutions have been proposed addressing the problem of material estimation from a single image if shape and/or illumination are known precisely.

Romeiro and Zickler first showed how to estimate reflectance under known shape and illumination [Romeiro et al. 2008], and Romeiro et al. later extended this work by marginalizing over illumination [Romeiro and Zickler 2010]. Generalizing further, Lombardi and Nishino [Lombardi and Nishino 2012] recover reflectance and illumination from an image assuming only that the object's shape is known, and Oxholm and Nishino [Oxholm and Nishino 2012] estimate reflectance and shape under exact lighting. If multiple images are available, it is also possible to recover shape and spatially varying reflectance [Alldrin et al. 2008; Goldman et al. 2010]. These techniques provide valuable intuition for moving forward, yet they hinge on knowing *exact* shape or *exact* illumination, or have strict setup requirements (directional light, multiple photos, etc), and require a fundamentally different approach when additional information is not available.

Such approaches have been proposed by Barron and Malik [Barron and Malik 2012a; Barron and Malik 2012b], who use strict priors to jointly recover shape, diffuse albedo, and illumination. However, as in many shape-from-shading algorithms, all surfaces are assumed to be Lambertian. Glossy surfaces are thus impossible to recover and may cause errors in estimation. Furthermore, Lambertian models of material are not suitable for describing a large percentage of real-world surfaces, limiting the applicability of these techniques.

A major concern of prior work is in recovering real-world BRDFs and high-frequency illumination [Lombardi and Nishino 2012; Oxholm and Nishino 2012; Romeiro and Zickler 2010], or that recovered shapes are integrable and reconstructions are exact (image

[1] We abbreviate "material reflectance" with "material."

and re-rendered image match exactly) [Barron and Malik 2012a; Barron and Malik 2012b]. However, it is well known that recovering these high-parametric solutions is ill-posed in many cases, and precise conditions must be met to estimate these robustly. For example, real-world BRDFs can be extracted from a curved shape and single directional light (known a priori) [Chandraker and Ramamoorthi 2011], and surface normals can be found given enough pictures with particular lighting conditions and isotropic (yet unknown) BRDFs [Alldrin and Kriegman 2007; Shi et al. 2012].

We opt for lower-order representations of reflectance and illumination. Our idea is to reduce the number of parameters we recover, relax the constraints imposed by prior methods, and attempt to recover materials from a more practical perspective. Our main goal is material inference, but we must jointly optimize over shape and illumination since these are unknown to our algorithm. We consider simple models of reflectance and lighting (models often used by artists, where perception is the only metric that matters), and impose only soft constraints on shape reconstruction. Our material model is low-order (only five parameters), allowing us to tease good estimates of materials from images, even if our recovered shape and illumination estimates are somewhat inaccurate. We also show that our model can be extended to spatially varying materials by inferring mixture coefficients (linearly combining 5-parameter materials) at each pixel. Figure 1 demonstrates the results of our estimation technique on a marble mortar and pestle.

Most similar to our work is the method of Goldman et al. [Goldman et al. 2010]. They estimate per-pixel mixture weights and a set of parametric materials, but require multiple HDR images under *known* lighting and impose limiting priors on mixture weights. Our method is applicable to single, LDR images (lighting is jointly estimated), and we also develop new priors for better mixture weights.

Contributions. Our primary contribution is a technique for extracting spatially varying material reflectance (beyond diffuse parameters) directly from an object's appearance in a single photograph *without requiring any knowledge of the object's shape or scene illumination*. We use a low-order parameterization of material and develop a new model of illumination that can be described also with only a few parameters, allowing for efficient rendering. *Because our model has few parameters, we tend to get low variance and thus robustness in our material estimates (e.g. bias-variance tradeoff).* By design, our material model is the same that is used throughout the 3D art/design community, and describes a large class of real-world materials (Sec 2). We show how to efficiently estimate materials from plausible initializations of lighting and shape, and propose novel priors that are crucial in estimating material robustly (Sec 3). We extend this formulation to spatially varying materials in Section 4. Our material estimates perform favorably to baseline methods and measure well with ground truth, and we demonstrate results for both synthetic and real images (Sec 5). We show applications in relighting (Figs 6, 8), material transfer (Fig 1), and material generation (Fig 12).

Limitations. Since we are using low-order material models that are isotropic and have monochromatic specular components, we cannot hope to estimate BRDFs of arbitrary shape (e.g. as measured by a gonioreflectometer), and there are some materials not encoded by our representation. Our recovered lighting and shape are not necessarily correct with respect to the true lighting/shape, although they are consistent with one another and sometimes give good estimates; as such, we only make claims about the accuracy of our material estimates. We use infinitely distant spherical lighting without considering interreflections, and we do not attempt to solve color constancy issues; lighting environments in our dataset integrate to the same value (per channel). Since we only have a single view of the object, certain material properties (e.g. specularities) may not be visible (depending especially on the coverage of the normals, i.e. flat surfaces provide much less information than curved surfaces).

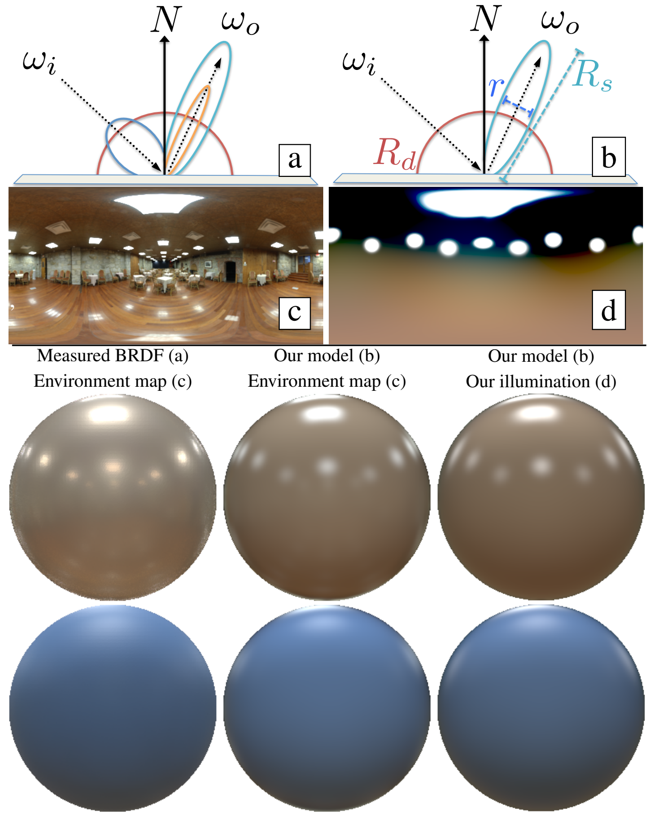

Figure 2: *A general BRDF can be made up of numerous reflection "lobes" (a), but in practice (e.g. surface modeling), a simple BRDF with one diffuse and one specular lobe tends to suffice (b). We use this representation as well as a low-order parameterization of illumination. Our illumination considers a real-world, omnidirectional lighting environment (c), and approximates it with a mixture of Gaussians and spherical harmonics; (d) shows our model fit to (c). We observe only slight perceptual differences when rendering with different combinations of the high- and low-order parameterizations (bottom rows).*

Due to our low-order model and perhaps mixture priors, shading effects can sometimes manifest in the spatial mixture map (Fig 11).

2 Low-order reflectance and illumination

Many previous methods have attempted to use high-order models of material (e.g. a linear combination of basis functions learned from measured BRDFs [Romeiro and Zickler 2010]) and illumination (parameterized with wavelets [Romeiro and Zickler 2010] or even on an image grid [Lombardi and Nishino 2012; Oxholm and Nishino 2012], consisting of hundreds to thousands of parameters or more). We propose the use of more rigid models of shape and illumination, which still can describe the appearance of most real world objects, and provide necessary rigidness to estimate materials when neither shape or illumination are exactly known.

Representing material. We represent materials using an isotropic diffuse and specular BRDF model consisting of only five parameters: diffuse albedo in the red, green, and blue channels (R_d), monochromatic specular albedo (R_s) and the isotropic "roughness" value (r), which is the concentration of light scattered in the specular direction, and can be considered (roughly) to be the size of the specular "lobe" (a smaller roughness value indicates a smaller specular lobe, where $r = 0$ encodes a perfect specular reflector).

This type of material model is surprisingly general for its low number of parameters. Ngan et al. have previously shown that such parameterizations provide very good fits to real, measured BRDFs [Ngan et al. 2005]. Perceptually, this model can also encode a family of isotropic, dielectric surfaces (mattes, plastics, and many other materials in the continuum of perfectly diffuse to near-perfect specular) [Pharr and Humphreys 2010]. There is also compelling evidence that such a material model suffices for photorealistic rendering, as this is the same material parameterization found most commonly in 3D modeling and rendering packages (such as Blender[2], which only considers diffuse and specular reflection for opaque objects), and used extensively throughout the 3D artist/designer community[3].

We write our BRDF following the isotropic substrate model as described in *Physically Based Rendering* [Pharr and Humphreys 2010], which uses a microfacet reflectance model and assumes the Schlick approximation to the Fresnel effect [Schlick 1994].

Figure 2 shows a comparison of what a measured BRDF (a) might look like in comparison to our material parameterization (b). We compare measured BRDFs to our material model (fit using the procedure described in Sec 5) rendered in natural illumination in the bottom row (left two columns).

Representing illumination. Consider a single point within a scene and the omnidirectional light incident to that point. This incident illumination can be conceptually decomposed into luminaires (light-emitters) and non-emitting objects. We consider these two separately, since the two tend to produce visually distinct patterns in object appearance (depending of course on the material). Luminaires will generally cause large, high-frequency changes in appearance (e.g. specular highlights), and non-emitters usually produce small, low-frequency changes.

Using this intuition, we parameterize each luminaire as a two dimensional Gaussian in the 2-sphere domain (sometimes known as the Kent distribution), and approximate any other incident light (non-emitted) as low-order functions on the sphere using 2^{nd} order spherical harmonics[4].

Such a parameterization has very few parameters relative to a full illumination environment (or environment map): six per light source (two for direction $L^{(d)}$, one for each intensity $L^{(I)}$, concentration κ, ellipticalness β, and rotation about the direction γ) and 27 spherical harmonic coefficients (nine per color channel), but more importantly, this parameterization still enables realistic rendering at much higher efficiency. Rendering efficiency is crucial to our procedure, as each function evaluation in our optimization method (Sec 3) requires rendering.

Our lighting environments maintain only high frequencies in regions of emitting sources, and is encoded by low-frequency spherical harmonics everywhere else. However, rendering with full versus approximate (our) lighting produces similar results (bottom middle vs bottom right). For additional discussion, see Sec 3.

3 Estimating specular reflectance

Our idea is to jointly recover material, shape, and illumination, such that a rendering of these estimates produces an image that is similar to the input image. Following Barron and Malik [Barron and Malik 2012a], we also enforce a strong set of priors to bias our material estimate towards plausible results.

Our goal is to recover a five dimensional set of material parameters $\mathbf{M} = (R_d^{(r)}, R_d^{(g)}, R_d^{(b)}, R_s, r)$, while jointly optimizing over illumination $\mathbf{L}$ and surface normals $\mathbf{N}$. Following notation in section 2, we denote R_d as RGB diffuse reflectance, R_s as monochromatic specular reflectance, and r as the roughness coefficient (smaller $r \Rightarrow$ narrower lobe $\Rightarrow$ shinier material). Illumination $\mathbf{L} = \{L, s\}$ is parameterized as a mixture of m Gaussians in the 2-sphere domain $L = \{L_1, \dots L_k\}$ with direction $L_i^{(d)}$, intensity $L_i^{(I)}$, concentration $L_i^{(\kappa)}$, ellipticalness $L_i^{(\beta)}$, and rotation $L_i^{(\gamma)}$ $(i \in \{1, \dots, k\})$. Statistics of real illumination environments are nonstationary and can contain concentrated bright points [Dror et al. 2004]; the Gaussian mixture aim to represent these peaks. Indirect light s is represented as a 9×3 matrix of 2^{nd} order spherical harmonic coefficients (9 per color channel). $\mathbf{N}$ is simply a vector of per-pixel surface normals parameterized by azimuth and elevation directions.

We phrase our problem as a continuous optimization by solving the parameters which minimize the following:

$$\underset{\mathbf{M},\mathbf{N},\mathbf{L}}{\text{argmin}}\ E_{\text{rend}}(\mathbf{M},\mathbf{N},\mathbf{L}) + E_{\text{mat}}(\mathbf{M}) + E_{\text{illum}}(\mathbf{L}) + E_{\text{shape}}(\mathbf{N})$$

$$\text{subject to } 0 \le \mathbf{M}^{(i)} \le 1,\ i \in \{1, \dots, 5\}, \tag{1}$$

where E_{rend} is the error between a rendering of our estimates and the input image, and E_{mat}, E_{illum}, and E_{shape} are priors that we place on material, illumination, and shape respectively. In the remainder of this section, we discuss the rendering term and the prior terms. Figure 3 shows the result of our optimization technique at various stages for a given input.

Rendering error. Our optimization is guided primarily by a term that penalizes pixel error between the input image and a rendering of our estimates. The term itself is quite simple, but efficiently optimizing an objective function (which includes the rendering equation) can be challenging. Writing I as the input, we define the term as the average squared error for each pixel:

$$E_{\text{rend}}(\mathbf{M},\mathbf{N},\mathbf{L}) = \sum_{i \in \text{pixels}} \sigma_i^{rend} ||I_i - f(\mathbf{M},\mathbf{N}_i,\mathbf{L})||^2, \tag{2}$$

where $f(\mathbf{M},\mathbf{N},\mathbf{L})$ is our rendering function, and $\sigma_i^{rend} = I_i^2$ reweights the error to place more importance on brighter points (primarily specularities).

Notice that we do not strictly enforce equality (as in [Barron and Malik 2012a; Barron and Malik 2012b]), as this soft constraint allows more flexibility during the optimization, and because our parameterizations are too stiff for equality to hold. This has the added benefit of reducing variance in our estimates.

As in any iterative optimization scheme, each iteration requires a function evaluation (and most likely a gradient or even hessian depending on the method). If chosen naïvely, f can take hours or longer to evaluate and differentiate, and here we describe how to construct f so that this optimization becomes computationally tractable.

The key to efficiency is in our low-order parameterization of illumination. By considering emitting and non-emitting sources separately, we treat our rendering as two sub-renders: one is a “full” render using the emitting luminaires (which are purely directional since we assume the light is at infinity), and the other is a diffuse-only render using all other incident light (reflected by non-emitters).

[2]http://wiki.blender.org/index.php/Doc:2.6/Manual/Materials

[3]http://www.luxrender.net/forum

[4]We assume all lighting comes from an infinite-radius sphere surrounding the object, as done in previous methods [Barron and Malik 2012a; Lombardi and Nishino 2012; Oxholm and Nishino 2012; Romeiro and Zickler 2010])

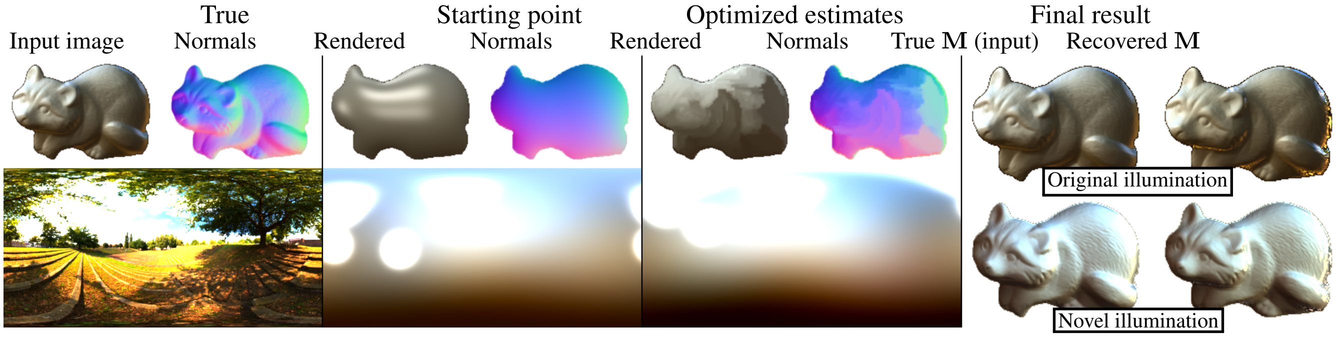

Figure 3: *Results from our optimization procedure. On the left is the input image (top left), true surface normals (top right), and true illumination (below), followed to the right by estimates that we begin our optimization with. The estimated rendering (using Eq 3), estimated surface normals, and estimated illumination are displayed in the third column. The rightmost column shows the true material (left) and our estimated material (right) rendered onto the true shape in the original lighting environment (top), and rendered in novel lighting (bottom). Our initialization is described in Sec 3.1, and uses no prior information about the input.*

Denoting Ω_e and Ω_n as set of "emitting" and "non-emitting" light directions respectively, $l(\omega)$ as the light traveling along direction ω, and f_M as the BRDF defined by material M, we write our rendering function as

$$f_i^{(e)} = \int_{\Omega_e} f_M(\omega, v) l(\omega) \max(\omega \cdot \mathbf{N}_i, 0) d\omega$$
$$f_i^{(n)} = \int_{\Omega_n} l(\omega) \max(\omega \cdot \mathbf{N}_i, 0) d\omega$$
$$f(\mathbf{M}, \mathbf{N}, \mathbf{L})_i = f_i^{(e)} + R_d f_i^{(n)}, \quad (3)$$

for the i^{th} image pixel and a particular view direction v. Notice that $f_i^{(n)}$ is simply irradiance over the non-emitting regions of the sphere, and is modulated by diffuse reflectance (R_d) since Lambertian BRDFs are constant.

We can compute both of these efficiently, because Ω_e is typically small (most lighting directions are occupied by negligible Gaussian components), and it is well known that diffuse objects can be efficiently rendered through spherical harmonic projection [Ramamoorthi and Hanrahan 2001a]. In terms of previous notation, directional sources L are used in the full render ($f_i^{(e)}$), and s is used for the diffuse-only render ($f_i^{(n)}$).

The intuition behind such a model is that indirect light contributes relatively low-frequency effects to an object's appearance, and approximating these effects leads to only slight perceivable differences [Ramamoorthi and Hanrahan 2001b]. A variation of this intuition is used for efficiently choosing samples in Monte Carlo ray tracing (e.g. importance sampling [Pharr and Humphreys 2010]), which causes problems in continuous optimization techniques since rendering is then non-deterministic.

Material prior. The rigidity of our material model (5 parameters to describe the entire surface), is a strong implicit prior in itself, but we also must deal with the ambiguity that can exist between diffuse and specular terms. For example, if a specular lobe (r) is large enough, then the specular albedo and diffuse albedo can be confused (e.g. dark specular albedo/bright diffuse albedo may look the same as bright specular albedo/dark diffuse albedo). Thus, we add a simple term to discourage large specular lobes, persuading the diffuse component to pick up any ambiguity between it and the specular terms:

$$E_{\text{mat}}(\mathbf{M}) = \lambda_m r^2. \quad (4)$$

The only material parameter that is constrained is specular lobe size, and $\lambda_m = 1$ in our work.

Illumination prior. We develop our illumination prior by collecting statistics from spherical HDR imagery found across the web (more details in Sec 5). Each image gives us a sample of real-world illumination, and to see how each sample relates to our illumination parameters, we fit our lighting model (SO(2) Gaussians + 2^{nd} order spherical harmonics) to each spherical image. Fitting is done using constrained, non-linear least squares, and the number of Gaussians (corresponding roughly to luminaires) is determined by the number of peaks in the HDR image (smoothed to suppress noise). Priors are developed by clustering the Gaussian parameters, and through principal component analysis on the spherical harmonic coefficients.

Denote $\bar{\kappa}_j$, $\bar{\beta}_j$ as the means of the j^{th} clusters (clustered independently using k-means) for the concentration, and ellipticalness of Gaussian parameters from our fitting process. Intuitively, these cluster centers give a reasonable sense of the shape of luminaires found in typical lighting environments, and we enforce our estimated sources to have shape parameters similar to these:

$$E_{\text{illum}}^{\text{means}}(L_i) = \mathcal{S}(\{|L_i^{(\kappa)} - \bar{\kappa}_j|\}_{i=1}^k) + \mathcal{S}(\{|L_i^{(\beta)} - \bar{\beta}_j|\}_{i=1}^k), \quad (5)$$

where $\mathcal{S}$ is the softmin function (differentiable min approximation) and $|\cdot|$ is a differentiable approximation to the absolute value (e.g. $\sqrt{x^2+\epsilon}$).

We also find the principal components (per channel) of the spherical harmonic coefficients fit to our data. During estimation, we reparameterize the estimated SH coefficients using weight vectors $w_{\{r,g,b\}}$, principal component matrices $S_{\{r,g,b\}}$, and means of all fit SH components $\mu_{\{r,g,b\}}$: $s(w) = [\mu_r + S_r w_r, \mu_g + S_g w_g, \mu_b + S_b w_b]$. We impose a Laplacian prior on the weight vector:

$$E_{\text{illum}}^{\text{pca}}(w) = \sum_{i \in \text{weights}} |w|. \quad (6)$$

This coerces the recovered SH components to lie near the dataset mean, and slide along prominent directions in the data. We found that seven principal components (per channel) roughly explained over 95% of our data (eigenvalue sums contain >95% of the mass), and we discard the two components corresponding to the smallest eigenvalues (then $w_{\{r,g,b\}} \in \mathbb{R}^7$ and $S_{\{r,g,b\}} \in \mathbb{R}^{9\times 7}$).

We also impose a gray world assumption, namely that each color channel should integrate (over the sphere of directions) to roughly the same value. Because we only have a single view of an object, some portions of the lighting sphere have significantly more influence than others; e.g. the hemisphere behind the object is mostly unseen and has smaller influence than the hemisphere in front. We weight the integration appropriately so that the dominant hemisphere has more influence (using $W_\theta = \cos(\frac{\theta-\pi}{2})$, where θ is the angle between the view direction and the direction of integration). This integration translates to a simple inner product (due to the nice properties of spherical harmonics), making the prior easy to compute:

$$E^{\text{gray}}_{\text{illum}}(s) = \quad ||G^T s_r - G^T s_g|| + ||G^T s_g - G^T s_b|| \\ + ||G^T s_r - G^T s_b||, \tag{7}$$

where G is the pre-computed integral of 2^{nd} order spherical harmonic basis functions (weighted by W_θ), and $s_{\{r,g,b\}}$ are the current estimates of spherical harmonic coefficients.

Our prior is then a weighted sum of these three terms:

$$E_{\text{illum}}(\mathbf{L}) = \lambda_m \sum_{i=1}^{m} E^{\text{means}}_{\text{illum}}(L_i) + \lambda_p E^{\text{pca}}_{\text{illum}}(w) + \lambda_g E^{\text{gray}}_{\text{illum}}(s), \tag{8}$$

keeping in mind $\mathbf{L} = \{L, s(w)\}$, and w as PCA weights described above. We set $\lambda_m = \lambda_p = \lambda_g = 0.1$.

Shape prior. We also optimize over a grid of surface normals, and impose typical shape-from-contour constraints: smoothness, integrable shape, and boundary normals are assumed perpendicular to the view direction. Let $N_i = (N^x_i, N^y_i, N^z_i)$, and $\hat{N}$ be the set of normals perpendicular to the occluding contour and view direction. We write the prior as:

$$E_{\text{shape}}(\mathbf{N}) = \quad \sum_{i \in \text{pixels}} \lambda_s \eta^s_i ||\nabla N_i|| + \lambda_I ||\nabla_y \frac{N^x_i}{N^z_i} - \nabla_x \frac{N^y_i}{N^z_i}|| \\ + \lambda_c \textstyle\sum_{c \in \text{boundary pixels}} ||N_c - \hat{N}_c||, \tag{9}$$

where the first term encodes smoothness where the input is also smooth (modulated by image dependent weights η), the second term enforces integrability, and the third ensures that boundary normals are perpendicular to the viewing direction. We set the weights as $\lambda_s = 1$, $\lambda_I = \lambda_c = 0.1$.

3.1 Initialization

Initial estimates of shape come from a naïve shape-from-contour algorithm (surface assumed tangent to view direction at the silhouette, and smooth elsewhere), and light is initialized with the mean of our dataset (if applicable; leaving out the illumination that generated the input image). We estimate an initial R_d by rendering irradiance with initial estimates of shape and lighting, dividing by the input image to get per-pixel albedo estimates, and averaging the RGB channels; R_s, r are set as small constants (0.01 for our results). Full details of our initialization procedure can be found in supplemental material.

3.2 Undoing estimation bias

Our low-parametric models tend to introduce bias into our estimates, but at the same time reduce estimation variance; e.g. bias-variance tradeoff). However, we have found that our priors produce consistent estimation bias: we typically see a smaller specular lobe and specular albedo due most likely to our material prior (Eq 4). We may also observe omitted-variable bias for images with materials not encoded by our model, but we do not address here.

Past methods point out that there are clear visual distinctions between different types and levels of gloss [Fleming et al. 2003; Sharan et al. 2008; Wills et al. 2009], and we use the input image coupled with our estimates to develop an "un-biasing" function. We develop a simple regression method (simple methods should suffice since the bias appears to be consistent) which works well for removing bias and produces improved results. Our goal is to find a linear prediction function that takes a vector of features to unbiased estimates of R_d, R_s and r. Our features consist of our estimates of specular albedo and specular lobe size, as well as histogram features computed on the resulting rendered image, normal map, input image, and the error image (rendered minus input); features are computed for both raw and gradient images. Given a set of results from our optimization technique with ground truth material parameter (obtained, e.g., from our dataset in Sec 5), we compute a bias prediction function by solving an L_1 regression problem (with L_2 regularization). For more details, see the supplemental material.

4 Recovering spatially varying reflectance

We propose an extension of Eq 1 for estimating spatial mixtures of materials. First, we define our appearance model simply as a spatially varying linear combination of renderings. Radiance at pixel i is defined as:

$$\sum_{j \in \text{materials}} m_{i,j} f(\mathbf{M}_j, \mathbf{N}_i, \mathbf{L}), \tag{10}$$

where $m_{i,j}$ is the j^{th} mixture weight at pixel i, and $\mathbf{M}_j$ is the j^{th} material. The rendering error term for estimating spatial materials then becomes:

$$E^{\text{mix}}_{\text{rend}}(\mathbf{M}_1, \ldots, \mathbf{M}_k, \mathbf{N}, \mathbf{L}, m) = \\ \sum_{i \in \text{pixels}} \sigma^{rend}_i \Big|\Big| I_i - \sum_{j \in \text{materials}} m_{i,j} \quad f(\mathbf{M}_j, \mathbf{N}_i, \mathbf{L}) \Big|\Big|^2. \tag{11}$$

We define three properties that the spatial maps (m) must adhere to: unity, firmness[5], and smoothness. First, the unity prior ensures that the mixture weights must be nonnegative and sum to one at every pixel:

$$\forall i, j \;\; m_{i,j} > 0, \qquad \forall i \;\; \sum_j m_{i,j} = 1. \tag{12}$$

As noted by Goldman et al. [Goldman et al. 2010], this prevents overfitting and removes certain ambiguities during estimation.

We place another prior on the "firmness" of our mixture maps. For certain objects, many patches on the surface are dominated by a single material (e.g. checkerboard); for others, the surface is roughly uniform over space (e.g. soap can be made of a diffuse layer and a glossy film which are both present over the whole surface); there are even materials ranging in between (e.g. marble). We would like a structured way of controlling which type of spatial mixture we produce, and we do so by imposing an exponential prior on each mixture element:

$$E^{\text{mix}}_{\text{firm}}(m) = \sum_{i,j} m^{\alpha}_{i,j}, \tag{13}$$

where $\alpha > 0$ controls how firm a mixture will be. For example, with the unity constraint, $\alpha > 1$ encourages uniform mixture weights (not firm, e.g. soap), and $\alpha < 1$ encourages mixture weights to be near zero or one (firm, e.g. checkerboard). For results in this paper, we use $\alpha = 0.5$. Notice that for $\alpha < 1$ this function is no longer convex, although in practice our optimization still seems to fare well. Our prior is more general (and controllable) than the method of Goldman et al. [Goldman et al. 2010], which assumes that each pixel is the linear combination of *at most* two materials.

Finally, we encourage spatial smoothness of the mixtures, as nearly all mixed-materials contain spatial structure:

$$E^{\text{mix}}_{\text{smooth}}(m) = \sum_{i,j} ||\nabla_x m_{i,j}|| + ||\nabla_y m_{i,j}||, \tag{14}$$

[5] We define the firmness prior as the decisiveness of mixture weights to snap to 0 or 1, and in this sense it has no relation to tactile properties.

where ∇_x and ∇_y are spatial gradient operators in the image domain.

By inserting our new rendering term and mixture priors into the objective function for single materials (Eq 1), we define a new optimization problem for estimating spatially varying materials:

$$\begin{aligned}
\underset{\mathbf{M}_1,\ldots,\mathbf{M}_k,\mathbf{N},\mathbf{L}}{\operatorname{argmin}} \quad & E_{\text{rend}}^{\text{mix}}(\mathbf{M}_1,\ldots,\mathbf{M}_k,\mathbf{N},\mathbf{L},m) + E_{\text{mat}}(\mathbf{M}) + \\
& E_{\text{illum}}(\mathbf{L}) + E_{\text{shape}}(\mathbf{N}) + E_{\text{firm}}^{\text{mix}}(m) + E_{\text{smooth}}^{\text{mix}}(m), \\
& \text{subject to:} \quad 0 \le \mathbf{M}_j^{(i)} \le 1,\ i \in \{1,\ldots,5\},\ \forall j, \\
& \qquad \forall i,j\ \ m_{i,j} > 0, \\
& \qquad \forall i\ \ \sum_j m_{i,j} = 1.
\end{aligned} \tag{15}$$

Solving this objective function can be difficult, but we have had success using constrained quasi-Newton methods (L-BFGS Hessian).

Our optimization results in a decomposition of the input image into k materials $\mathbf{M}$, a set of per-pixel weights for each material m, per-pixel surface normals $\mathbf{N}$, and illumination parameters $\mathbf{L}$. In this work, we focus on the correctness of our mixture materials and their applications.

5 Experiments

We evaluate the results of our method for objects with homogeneous (spatially uniform) materials in Section 5.1, as well as our inhomogeneous (spatially varying) material results in Section 5.2. We report results for both a dataset we collected containing ground truth material information, as well as for the Drexel Natural Illumination dataset.

5.1 Homogeneous materials

For evaluation and training our bias predictors (Sec 3.2), we have collected a dataset consisting of 400 images rendered with real world shapes, materials, and illumination environments (all chosen from well-established benchmark datasets). We use the 20 ground truth shapes available in the MIT Intrinsic Image dataset [Grosse et al. 2009], and render each of these objects with 20 of the materials approximated from the MERL BRDF dataset [Matusik et al. 2003], for a total of 400 images. We use 100 different illumination environments (50 indoor, 50 outdoor) found across the web, primarily from the well known ICT light probe gallery[6] and the sIBL archive[7]. We ensure that each object is rendered in 10 unique indoor and 10 unique outdoor lighting environments, permuted such that each illumination environment is used exactly four times throughout the dataset. Each lighting environment is white balanced and has the same mean (per channel).

Our dataset has two "versions." The first version of our dataset (**fit dataset**) is rendered using our low-order reflectance model (we approximate MERL BRDFs by fitting our own 5-parameter material model to the measured data, and render using our fits). The resulting images are highly realistic, and allow us to both compare our material estimates with ground truth, and regress bias prediction functions (as in Sec 3.2).

The second version (**measured dataset**) is rendered using only *measured* BRDFs (from the MERL dataset); these images are truly realistic as the shape, material, and lighting are all sampled directly from real-world data. Furthermore, these images are synthesized using a physical renderer and thus include shadows and bounced light. This dataset gauges how well our method can generalize to real images and reflectances not encoded by our model.

Results. We generate results using the optimization procedure described in Sec 3, followed by our bias regression method as in Sec 3.2. Bias prediction functions are learned through leave-one-out cross validation.

In this section, we report results from our optimization technique (**blind optimized**), and after bias regression (**optimized+regressed**). For comparison, we compute a **baseline** material estimate which computes the R_d by averaging the image pixels in each channel, and R_s and r from the average found in our material dataset, and then regress and apply bias predictors to the baseline estimates (**baseline+regressed**). We also compare to materials achieved by our optimization assuming the shape and illumination are known[8] and fixed (**known shape+light**); hence only the material is optimized. Results using this procedure gauge the difficulty of our optimization problem, and shows how much our optimization can improve with more sophisticated initialization procedures.

On our fit dataset, our full method (optimized + regressed) is capable of recovering highly accurate material parameters. Figure 4 plots the true material from our "fit" dataset against our estimated parameters for each of the 400 images. A perfect material estimate would lie along the diagonal (dashed line). Overall, we see a linear trend in our diffuse results, and that our bias regression can significantly improve our optimized estimates of specular reflectance and specular lobe size (and even better than shape+light).

We also develop two ways of measuring visual error in our materials. We define **original illumination** as the average pixel error from comparing the input image with the image produced by rendering our *estimated* material onto the *true* shape in the *true* lighting (which are known for our all images in our dataset). This is a harsh test, as any errors in material must manifest themselves once rendered with the true shape and light. The second metric (**cross rendered**) is even more telling: we compare renderings of the input object with the a) true material and b) our estimated material in six novel illumination environments not present in our dataset and compute average pixel error. This measure exposes material errors across unique, unseen illumination.

Using these measures, our full method achieves low error for both versions of our dataset. Figure 5 shows these error measures for three different metrics (per-pixel L2 and L1 norms, and absolute log difference), and optimized+regressed performs the best overall for both datasets. This indicates that both our optimization and regression are crucial components, and one is not dominating inference since optimized+regressed consistently outperforms baseline+regressed. Known+shape light also performs well, indicating that our optimization procedure might improve if better initializations are available.

We demonstrate that in many cases our method can do very well at visually reproducing both measured and fit reflectances, even in novel illuminations. We show qualitative results for both versions of our dataset in Figs 6 and 7 – these are some of our best and median results. Our material estimates are typically visually accurate in original and novel illumination, even for many of the measured BRDFs in our measured dataset. We also observe that our regression generally helps for both datasets, indicating that our learned bias predictors may generalize to complex materials and real-world images. However, it is clear that our results degrade for complex reflectance functions that lie well outside our model (Fig 7, measured dataset columns 1+4).

Finally, we demonstrate our method's capability on *real images* from the Drexel Natural Illumination dataset in Fig 8. Our model

[6]http://gl.ict.usc.edu/Data/HighResProbes

[7]http://www.hdrlabs.com/sibl/archive.html

[8]Known lighting is fit to our parameterization and may still be some distance from ground truth

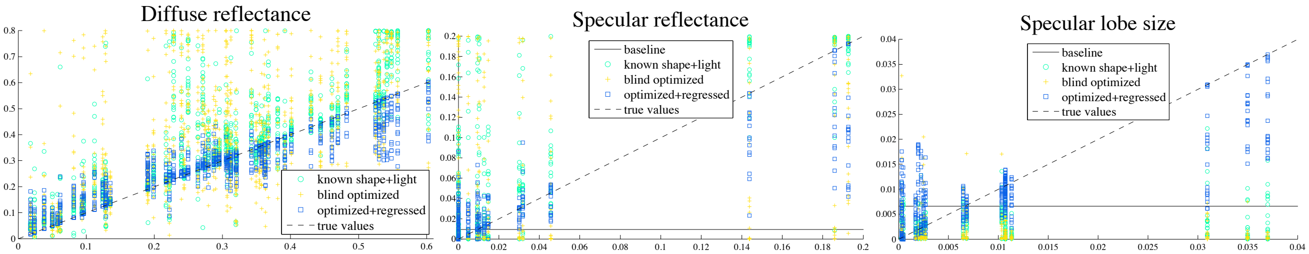

Figure 4: *Errors in material estimates for each image in our dataset. Each plot shows the true material value on the horizontal axis plotted against our estimate of diffuse reflectance (R_d), specular reflectance R_s, and specular lobe size r (left to right). We show the results for our baseline, the material produced given accurate initial shape and lighting, our blind optimization technique (blind optimized), and the material regressed by un-biasing our optimization results (blind regressed); details in Sec 5.*

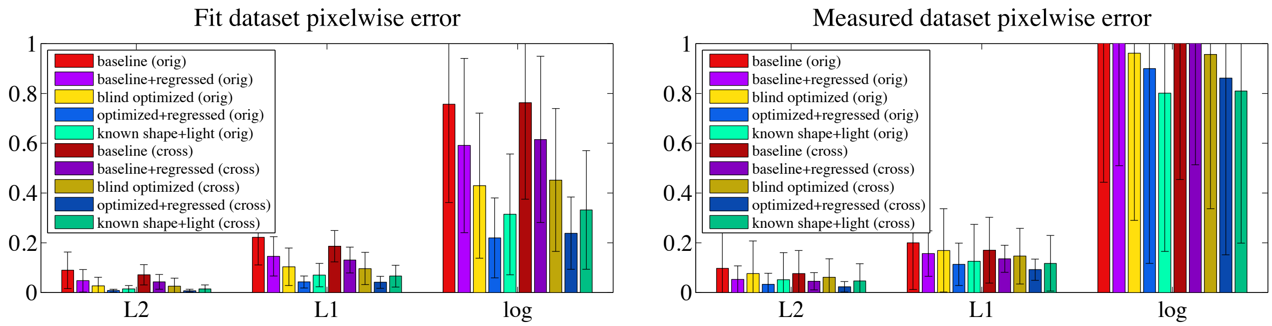

Figure 5: *We compare the average per-pixel error of the input image and a re-rendered image with estimated material (but with the true shape and true lighting that produced the input image) for various techniques and for both versions of our dataset; see Sec 5 for details. We compute errors in the original illumination (orig), and averaged over six novel illumination environments (cross), for three different metrics: L2 and L1 norm, and the absolute log difference, and show the mean over the dataset (error bars indicate one standard deviation). Our full method (optimized+regressed) achieves low error relative to others. We also observe similar (yet slightly worse) error on our measured dataset, indicating that, for a variety of cases, our a) our method can handle real-world materials, and b) that our material model is capable of visually reproducing complex reflectance functions.*

appears somewhat robust to spatially varying reflectance in these images, but suffers from the complexity of the imaged reflectances and because we assume only a single material is present; this suggests ideas for future work.

5.2 Inhomogeneous materials

For ground truth evaluation, we again use the **measured dataset**. We use our mixture estimation procedure to estimate $k = \{2, 3\}$ materials[9] for each dataset image, and compare to the results in of our method for $k = 1$. For additional comparison, we compute a baseline material estimate by clustering the image into k components (using k-means); computing diffuse albedo (per component) by averaging the image pixels in each channel, and the specular components are fixed to a small yet reasonable value.

We measure error by rendering our *estimated* material onto the *true* shape in the *true* lighting (which are known for our all images in the dataset), and compare this to the input image. We do the same test, but for six novel lighting environments not found in the dataset (e.g. estimated material versus true material in novel light). We denote these as “orig” and “cross” lighting respectively. These are harsh tests of generalization, as any errors in material must manifest themselves once rendered with the true shape and light, and the “cross” measure exposes material errors across unique and unseen illumination.

Fig 10 shows quantitative results averaged over the entire dataset for L2, L1, and absolute log difference error metrics. Our mixture materials (optimized-$\{2, 3\}$) consistently outperform single material estimation (optimized-1), and are always better than the baseline estimates.

We observe a similar trend in our qualitative results (Fig 9). Because we are attempting to estimate true, measured BRDFs which may lie outside of our 5-parameter material model, estimation may not work well with a single material. However, by adding multiple materials, we typically get improved results, even in novel illumination. This indicates that our mixture weights are typically robust to shading artifacts such as shadows and specularities. It is clear that adding more components helps, although the distinction between $k = 2, 3$ is subtle (both qualitatively and quantitatively).

6 Applications

Once we have decomposed an image into its materials and spatial mixing weights, we can apply this intrinsic material information to new surfaces as in Fig 1. Applying the materials (microstructure) to a novel object is straightforward, but transferring the mixture weights (macrostructure) can be challenging in certain cases (e.g. when a mapping from one surface to another is not easily computed).

We propose a straightforward solution: choose a small patch of the image defined by the mixture weights that is nearly fronto-parallel (determined from our predicted surface normals; to avoid foreshortening), and synthesize a larger texture (seeded with the small patch) using existing methods; e.g. [Efros and Leung 1999]. Then, map the surface of the object that the material will be transferred to onto a plane (also using existing methods; e.g. [Sheffer et al.

[9] We use a spatial mixture for homogeneous materials as our mixture maps generalize current literature. They capture spatial variation in material (as in [Goldman et al. 2010]), but we use them to also encode any kind of surface variation not well-captured due to long-standing SFS assumptions.

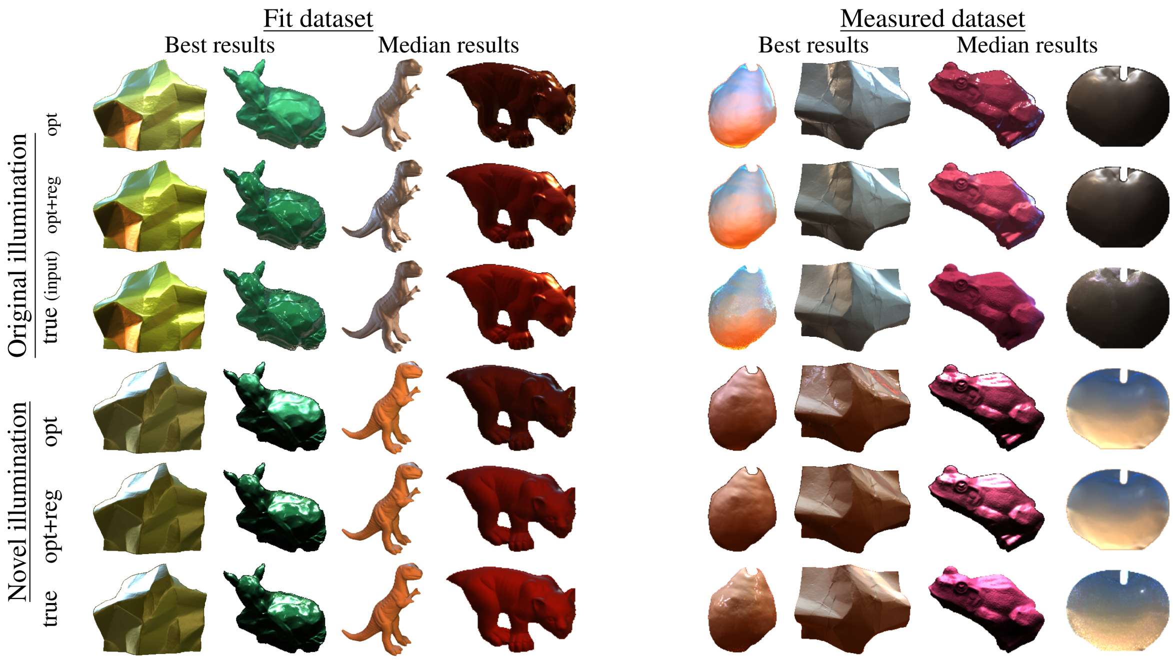

Figure 6: *Qualitative results on both versions of our dataset Materials are estimated using our blind optimized (opt) and optimized+regressed (reg) methods, and compared to ground truth (true). The true original illumination image is also the input for estimating material. Notice that our technique can recover both glossy and matte materials, performs well even for these complex shapes. Our method attains visually pleasing results even for complex reflectance functions not encoded by our model (e.g. measured dataset) even in new lighting conditions.*

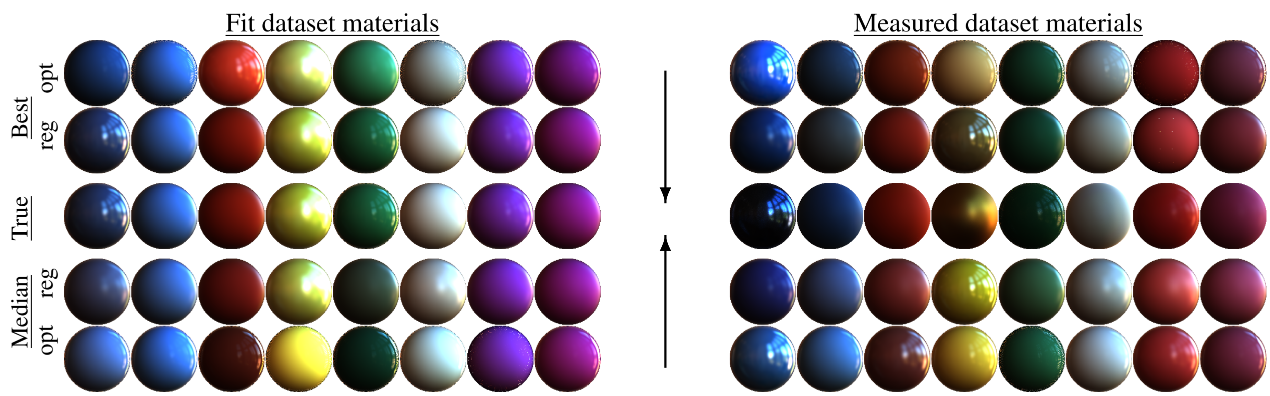

Figure 7: *Comparison of estimated materials rendered in novel lighting. The true materials lie on the middle row alongside our per-material best and median optimized (opt) and regressed (reg); arrows indicate the direction in which materials should improve. We achieve very good results for input images that are well described by our model in the fit dataset (rows 2 and 4 generally look like row 3), and even in many cases for measured BRDFs. However, low-order model bias prevents our method from capturing certain materials well (e.g. column 4; measured dataset).*

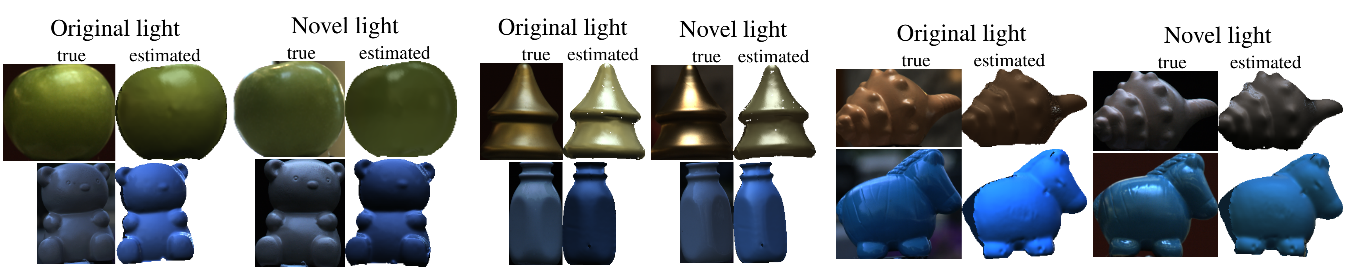

Figure 8: *Results on real data from the Drexel Natural Illumination dataset. This dataset contains real images and corresponding ground truth shape and lighting information. We estimate materials from one picture, and render the material using the true shape and light for the original illumination and another illumination from the dataset (novel light); we compare to the real picture of the object in both scenes (original and novel). Even in the presence of slight spatial variation (e.g. top left; apple) and complex reflectance (top middle) our method can still recover decent estimates. Still, addressing these issues is key to generalizing our method's applicability.*

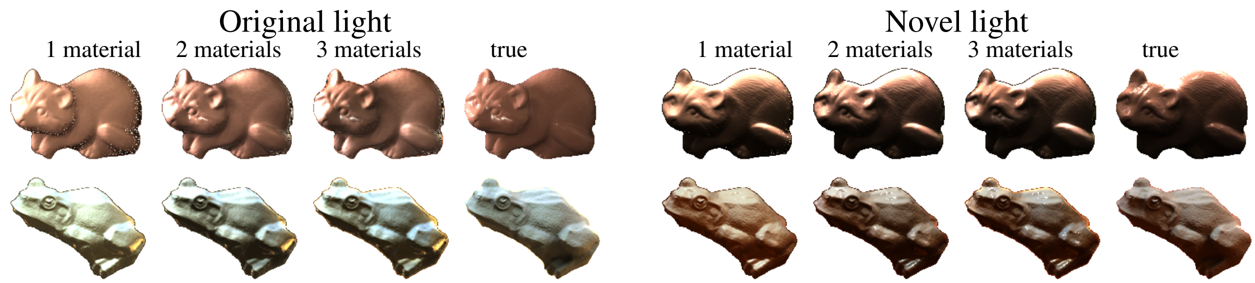

Figure 9: *Best (top row) and median (bottom row) results from the "measured dataset" which contains physically rendered objects with measured BRDFs. Typically these materials are not well encoded by our low-order material model with 1 mixture component, but increasing the number of mixture components improves re-rendering error. We show our estimated materials for one, two, and three mixture components, and compare these to the ground truth result (also the input image) in both the original and novel illumination environments.*

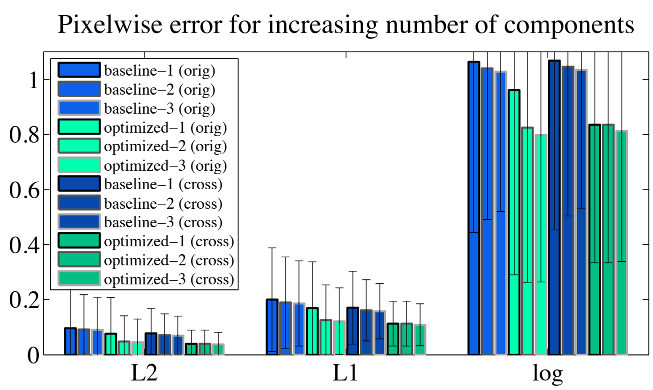

Figure 10: *Quantiative results on our "measured" dataset. Our mixture materials (optimized-$\{2, 3\}$) consistently outperform single material estimation (optimized-1); see text for details.*

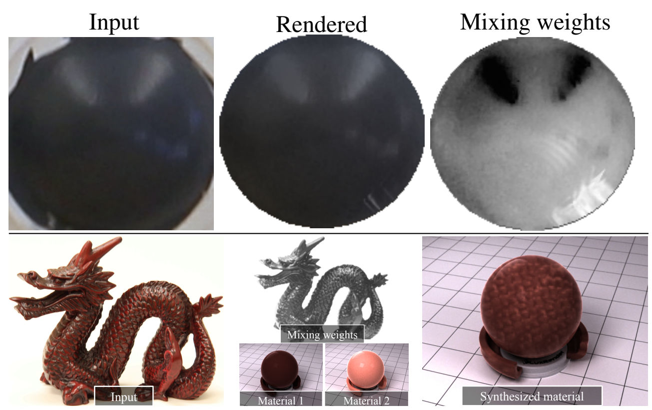

Figure 11: *Failure examples. The top row demonstrates an incorrect mixture map estimate: specularities have been detected as a separate material. A material transfer result is shown on bottom, but our material model contains no mesostructure and appears flat.*

2006]); this mapping defines correspondences between the synthesized mixture weights and the new mesh. We generate all of our transfer/generation results using this technique, and more sophisticated methods are clear directions for future work.

We also propose a generative material modeling strategy: besides transferring a complete mixture material, we can combine estimates from multiple images to create new materials (e.g. materials from one and mixture weights from another, and so on).

Generative results (as well as direct transfer results) are shown in Fig 12. We have decomposed four swatches from our dataset (all unique colors and mediums and spanning the three illumination environments in our dataset) using $k = 2$ mixture components. We apply each set of materials to each synthesized mixture, and render the result onto spheres. We assert that our estimated materials correspond to microstructure and mixing weights correspond to macrostructure, which appears correct for these results (microstructure varies vertically, macrostructure horizontally).

7 Conclusion

We have demonstrated a new technique for estimating spatially varying parametric materials from an image of a single object of unknown shape in unknown illumination, going beyond the typical Lambertian assumptions made by existing shape-from-shading techniques. Strong priors and low-order parameterizations of lighting and material are key in providing enough constraints to make this inference tractable. Such rigid parameterizations often lead to estimation bias, and we also present a simple yet powerful technique for removing this bias.

Our results suggest that material recovery is not necessarily dependent upon the joint recovery of accurate shape and illumination; *as long as the shape and illumination are consistent with each other, materials can still be robustly estimated.* This is encouraging from a material inference standpoint, as even the best shape-from-shading algorithms still produce flawed estimates in many scenarios.

As far as we know, our method is the first to estimate parametric material models without assuming shape or illumination is known a priori. We believe that our method provides good initial evidence that solving this problem is in fact feasible, and provides a foundation for estimating materials from photographs alone.

Our decompositions can be transferred to new shapes, imbuing them with similar appearance as the input image. Furthermore, our decompositions are also generative, and can be used to create new materials by simultaneously transferring decompositions from multiple objects (e.g. mixing weights from one, materials from another). Our re-rendering results do not incorporate any information from our estimated surface normals, and the spatial frequency of our mixture weights are defined by the input image resolution (some artifacts visible in Fig 1); intelligently incorporating and up sampling these estimates are reasonable directions for future work.

References

ADELSON, E. H. 2000. Lightness perception and lightness illusion. In *New Cognitive Neurosciences*, 339–351.

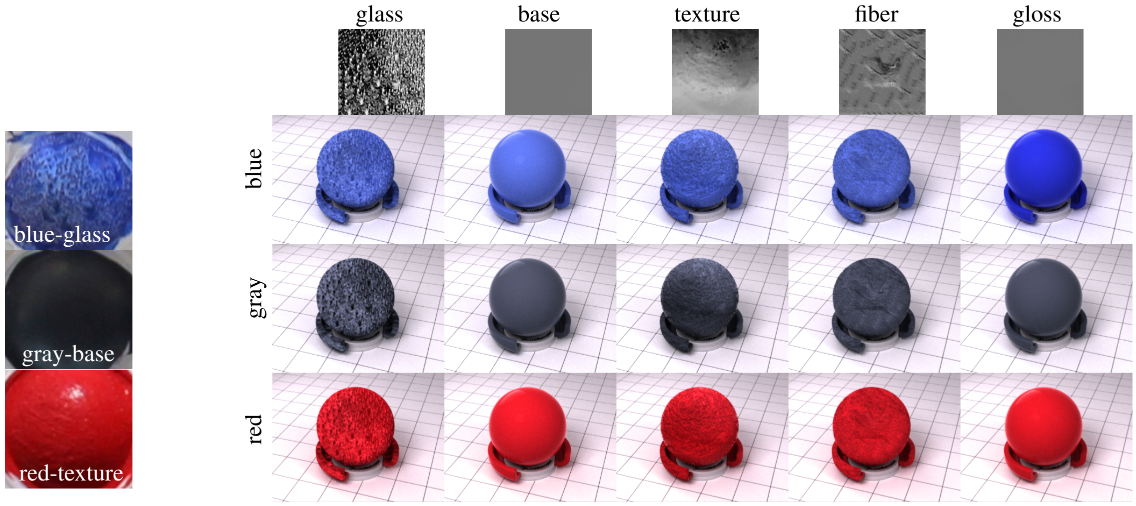

Figure 12: *Material transfer and generation for several material "swatches" (hemispheres painted with different colors/mediums/coats). We decompose single images (on left) into two material components and a spatial mixture map. Then, we synthesize new materials by taking all combinations of the inferred materials and the derived mixture weights, and render these combinations onto spheres in novel illumination (using LuxRender: http://luxrender.net). Images along the diagonal show a transfer material result for a given picture on the left. The off-diagonals show the generative capabilities of our algorithm: by combining multiple decompositions (materials + mixing weights), we can generate new, unseen materials. We expect that full 3D textures will give better results, but it is currently impossible to estimate 3D textures from a single picture. Best viewed in color at high resolution.*

ALLDRIN, N., AND KRIEGMAN, D. 2007. Toward reconstructing surfaces with arbitrary isotropic reflectance : A stratified photometric stereo approach. In *ICCV*.

ALLDRIN, N. G., ZICKLER, T., AND KRIEGMAN, D. 2008. Photometric stereo with non-parametric and spatially-varying reflectance. In *CVPR*.

BARRON, J. T., AND MALIK, J. 2012. Color constancy, intrinsic images, and shape estimation. In *ECCV*.

BARRON, J. T., AND MALIK, J. 2012. Shape, albedo, and illumination from a single image of an unknown object. In *CVPR*.

CHANDRAKER, M., AND RAMAMOORTHI, R. 2011. What An Image Reveals About Material Reflectance. In *ICCV*.

DROR, R. O., WILLSKY, A. S., AND ADELSON, E. H. 2004. Statistical characterization of real-world illumination. *J Vis 4*, 9, 821–837.

EFROS, A. A., AND LEUNG, T. K. 1999. Texture synthesis by non-parametric sampling. In *ICCV*.

FLEMING, R. W. W., DROR, R. O., AND ADELSON, E. H. 2003. Real-world illumination and the perception of surface reflectance properties. *J Vis 3*, 5, 347–368.

GOLDMAN, D., CURLESS, B., HERTZMANN, A., AND SEITZ, S. 2010. Shape and spatially-varying brdfs from photometric stereo. *IEEE TPAMI 32*, 6, 1060–1071.

GROSSE, R., JOHNSON, M. K., ADELSON, E. H., AND FREEMAN, W. T. 2009. Ground-truth dataset and baseline evaluations for intrinsic image algorithms. *ICCV*.

LOMBARDI, S., AND NISHINO, K. 2012. Reflectance and Natural Illumination from a Single Image. In *ECCV*.

MATUSIK, W., PFISTER, H., BRAND, M., AND MCMILLAN, L. 2003. A data-driven reflectance model. *ACM Transactions on Graphics 22*, 3 (July), 759–769.

NGAN, A., DURAND, F., AND MATUSIK, W. 2005. Experimental analysis of brdf models. In *Proceedings of the Eurographics Symposium on Rendering*, 117–226.

OXHOLM, G., AND NISHINO, K. 2012. Shape and Reflectance from Natural Illumination. In *ECCV*.

PHARR, M., AND HUMPHREYS, G. 2010. *Physically Based Rendering, Second Edition: From Theory To Implementation*, 2nd ed. Morgan Kaufmann Publishers Inc., San Francisco, CA, USA.

RAMAMOORTHI, R., AND HANRAHAN, P. 2001. An efficient representation for irradiance environment maps. In *ACM SIGGRAPH*.

RAMAMOORTHI, R., AND HANRAHAN, P. 2001. A signal-processing framework for inverse rendering. In *ACM SIGGRAPH*.

ROMEIRO, F., AND ZICKLER, T. 2010. Blind reflectometry. In *ECCV*.

ROMEIRO, F., VASILYEV, Y., AND ZICKLER, T. 2008. Passive reflectometry. In *ECCV*, 859–872.

SCHLICK, C. 1994. An Inexpensive BRDF Model for Physically-based Rendering. *Computer Graphics Forum 13*, 233–246.

SHARAN, L., LI, Y., MOTOYOSHI, I., NISHIDA, S., AND ADELSON, E. H. 2008. Image statistics for surface reflectance perception. *J. Opt. Soc. Am. A 25*, 4 (Apr), 846–865.

SHEFFER, A., PRAUN, E., AND ROSE, K. 2006. Mesh parameterization methods and their applications. *Found. Trends. Comput. Graph. Vis. 2*, 2 (Jan.), 105–171.

SHI, B., TAN, P., MATSUSHITA, Y., AND IKEUCHI, K. 2012. Elevation angle from reflectance monotonicity: photometric stereo for general isotropic reflectances. In *ECCV*, 455–468.

WILLS, J., AGARWAL, S., KRIEGMAN, D., AND BELONGIE, S. 2009. Toward a perceptual space for gloss. *ACM Transactions on Graphics 28*, 4, 1–15.