Precomputed Radiance Transfer for Reflectance and Lighting Estimation

Daniel Thul
ETH Zurich
daniel.thul@inf.ethz.ch

Vagia Tsiminaki*
IBM Research Zurich
tsi@zurich.ibm.com

Ľubor Ladický
ETH Zurich
lubor.ladicky@inf.ethz.ch

Marc Pollefeys
ETH Zurich / Microsoft
marc.pollefeys@inf.ethz.ch

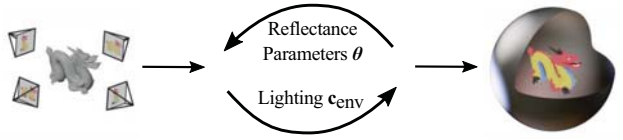
2020 International Conference on 3D Vision (3DV) | 978-1-7281-8128-8/20/$31.00 ©2020 IEEE | DOI: 10.1109/3DV50981.2020.00125

Abstract

Decomposing scenes into reflectance and lighting is an important task for applications such as relighting, image matching or content creation. Advanced light transport effects like occlusion and indirect lighting are often ignored, leading to subpar decompositions in which the albedo needs to compensate for insufficiencies in the estimated shading. We show how to account for these advanced lighting effects by utilizing precomputed radiance transfer to estimate reflectance and lighting. Given the geometry of an object and one or multiple images, our method reconstructs the object's surface reflectance properties—such as its albedo and glossiness—as well as a colored lighting environment map. Evaluation on synthetic and real data shows that incorporation of indirect light leads to qualitatively and quantitatively improved results.

1. Introduction

Estimating the surface properties of objects and the lighting of the scene through multi-camera observations is of particular interest for many computer graphics applications such as shadow or highlight removal [20], image matching [54], material or image editing [31, 24, 46, 5], image relighting [47], or realistic insertion of virtual content in augmented reality [12, 23, 36]. In such scenarios, images need to be decomposed into their intrinsic components: *surface properties* such as shape, diffuse albedo, specular albedo, and roughness; and *illumination* such as a distant lighting environment map or spatially varying lighting. This is a challenging task since images are the result of a complex interaction of all of these properties. According to Adelson and Pentland [1], if we observe an object in an image, we cannot certainly say whether the color of that object is due to the albedo, the illumination, or a combination of both. Similar ambiguity appears between specularity and sharpness of illumination. To solve this inverse problem, the majority of intrinsic decomposition methods introduce strong assumptions over the unknown properties or make simplifications of the interaction of these components.



(a) Input: Geometry and Images (b) Optimization (d) Output: Albedo and Lighting

Relighting

(a) Input (b) Relightings

Image Editing

(a) Original Image (b) Overlay (c) Relit overlay

Figure 1. *Top:* Overview of our method's pipeline. *Middle:* Relightings based on our method's outputs. *Bottom:* Image edit based on our method's outputs.

We argue that the ambiguity between object and light color can, in certain cases, be resolved. For example: in the presence of interreflections, white light illuminating a green object produces a different outcome than green light illuminating a white object of the same shape. Convex objects are a degenerate case since they do not give rise to interreflections and thus object and light color become interchangeable.

Here, we explore the possibility of accurately modeling the light transport in the scene, capturing effects such as self-shadowing and indirect lighting, in order to estimate reflectance and illumination. These are our contributions:

- We present a novel decomposition framework, in a multi-view setting, using multi-bounce precomputed radiance transfer to accurately model reflectance and lighting

* This work has been done while the author was at ETH Zurich.

2475-7888/20/$31.00 ©2020 IEEE
DOI 10.1109/3DV50981.2020.00125

Algorithm 1: Pseudocode for our method. This figure shows the diffuse case. The non-diffuse version is analogous but allows for several oberserved radiance values per surface point.

Data: Mesh $\mathcal{M}$, views $\mathcal{V}$, and SH degree d
Result: Mesh surface properties $\boldsymbol{\theta}$ and lighting coefficients $\boldsymbol{c}_{\text{env}}$
$\mathbf{M} \longleftarrow$ ConstructMeasurementMatrix(d)
`// Initialize parameters (gray albedo)`
$\boldsymbol{\theta} \longleftarrow$ InitialSurfaceParameters($\mathcal{M}$)
`// Find the observed radiance per surface point`
$\boldsymbol{\iota} \longleftarrow$ RadiancePerPoint($\mathcal{M}$, $\mathcal{V}$)
repeat
  $\boldsymbol{\mathcal{T}} \longleftarrow$ ComputeRadianceTransfer($\mathcal{M}$, $\boldsymbol{\theta}$, d)
  `// Optimize lighting under constraint M*x >= 0`
  $\boldsymbol{c}_{\text{env}} \longleftarrow$ ConstrainedLinearLeastSquares($\boldsymbol{\mathcal{T}}$, $\boldsymbol{\iota}$, $\mathbf{M}$)
  `// Optimize albedo`
  $\Delta\boldsymbol{\theta} \longleftarrow$ ApproximateDerivative($\boldsymbol{\mathcal{T}}$, $\boldsymbol{\iota}$, $\boldsymbol{\theta}$, $\boldsymbol{c}_{\text{env}}$)
  $\boldsymbol{\theta} \longleftarrow$ GradientDescentStep($\boldsymbol{\theta}$, $\Delta\boldsymbol{\theta}$)
until *converged*

- Our method can estimate colored, spatially-varying object albedo, object glossiness, and colored environment lighting maps without the need for priors such as a preference for white or monochromatic lighting
- We formulate a *measurement matrix* that gives rise to constraints which ensure that the estimated spherical harmonics lighting stays physically plausible

2. Related work

Decomposing an image into albedo and shading, often referred to as intrinsic image decomposition, has been extensively studied in the literature [16]. To resolve the ill-posed behavior of the problem most of the existing works build on the "retinex theory" [26] by setting the albedo to be piecewise smooth. The vast majority of intrinsic decomposition methods impose priors in the log-domain [53, 16, 14, 27, 3, 58, 25], emphasizing pairwise smoothness in the color space. Bell *et al*. [4] state that pixels far apart in the image can have the same reflectance and thus, introduce constraints to all $\mathcal{O}(n^2)$ pairs of pixels in the image and not just to pairs of neighboring pixels. Considering these long-range interactions, they decompose real-world photographs accurately. To disambiguate the problem Chen and Koltun [9] use RGB-D. All these methods rely on strong 2D priors and do not explicitly model the indirect lighting and thus can not describe complex lighting conditions.

To model the scene lighting, a spherical harmonics approximation [43] is often used in the literature. Zollhöfer *et al*. [60] use one constant set of spherical harmonics for the entire scene. Maier *et al*. [35] introduce local spherical harmonics coefficients to model spatially-varying illumination. They use RGB-D input data and recover greyscale albedo and illumination while refining the geometry. Similar to our setup, Mélou *et al*. [38] use RGB input images and a reconstructed 3D mesh to estimate a global set of spherical harmonics coefficients. In contrast to our method, these methods do not take self-shadowing and indirect light into account.

The rendering equation [21] describes effects such as soft shadows, caustics, ambient occlusion and indirect lighting. Carroll *et al*. [7] model indirect lighting using a single light color and one indirect bounce and require user input. Li *et al*. [30] use a series of CNNs to predict individual bounces. Lombardi and Nishino [33] build on this idea and use the Monte Carlo path tracing formulation to model the indirect lighting of the scene. They compute gradients with respect to the lighting environment numerically, making their method computationally expensive. Zhang *et al*. [56] use a radiosity method to solve for the indirect lighting. They use RGBD scans to reconstruct rooms as a set of diffuse, non-spatially varying surfaces lit by point, line, area, and environment lights, the number and type of which the user specifies. Recent methods for differentiable path tracing make gradient computation more tractable, but require the light sources to be modeled explicitly [29, 2, 39].

Dong *et al*. [10] use a video of a rotating object with known geometry to recover surface reflectance parameters and detailed cube environment maps with a focus on specular objects. Using very similar input, Park *et al*. [40] recover highly detailed environment maps from mirror-like reflectance, taking specular interreflection and the Fresnel effect into account.

A multitude of deep learning-based methods have been proposed lately. Some of those methods stay in the image domain and directly produce an output image without recovering intermediate geometry, reflectance or illumination. This approach has been used for relighting [51, 41] and separation of distinct illuminants in the scene [19]. Other methods use the input image to directly predict the environment lighting, going from a low dynamic range input to a high dynamic range environment map [18, 6, 28, 57, 17] or predicting a set of spatially varying lighting coefficients to capture indoor lighting environments [13, 50]. Lastly, some deep learning methods predict reflectance parameters as well as illumination. Meka *et al*. [37] and Georgoulis *et al*. [15] train an encoder-decoder architecture to produce diffuse albedo, specular albedo, and shininess for a single, uniformly colored object in the input image. Yu and Smith[55] train a similar network to turn an RGB image into diffuse albedo and normal maps and then recover a low frequent lighting model by solving a linear system. Wang *et al*. [52] combine neural networks and traditional optimization techniques to estimate diffuse and specular material properties and higher frequent illumination. In contrast to the previously mentioned works, Kanamori and Endo [22] take self occlusion into account and predict normals, diffuse albedo, and illumination for full-body human images.

We use an image formation model which accounts for both direct and indirect lighting without the need of heavy 2D priors. Contrary to Mélou *et al*. [38], we define the albedo on the surface without enforcing additional multi-view constraints. We approximate the lighting with spherical harmonics of rela-

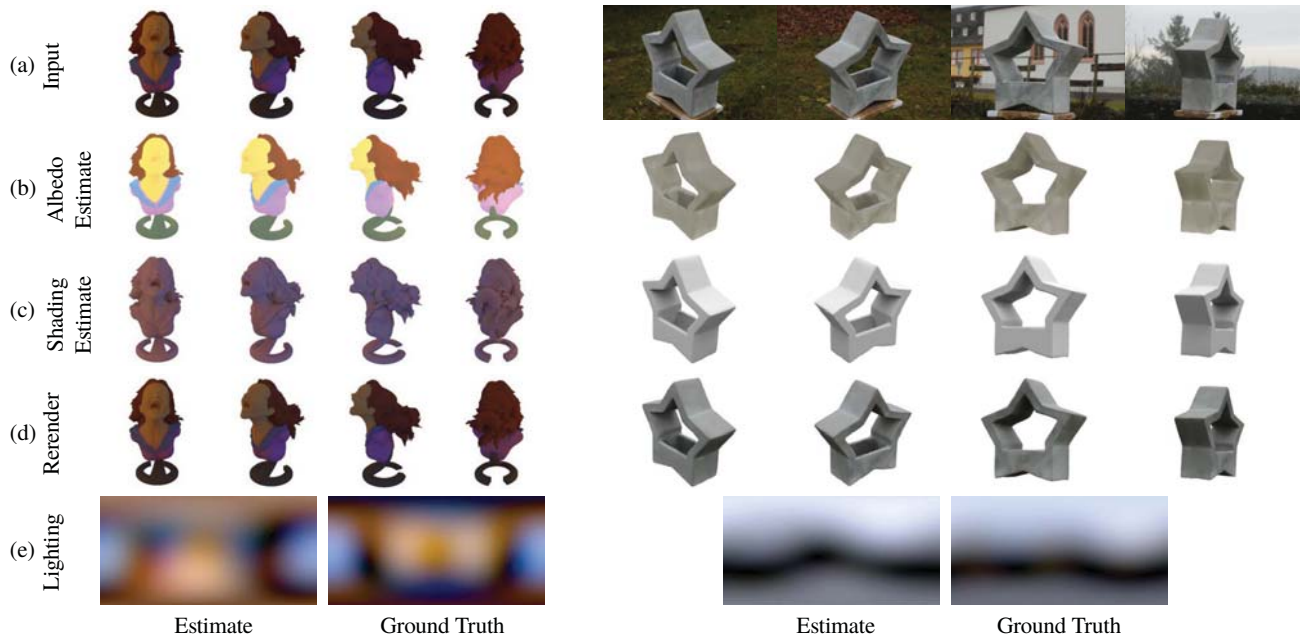


Figure 2. Diffuse datasets and results, "Joyful Yell" (synthetic, *left*) and "Star" (real-world, *right*). **(a)** Subset of the input images. **(b)** Estimated albedo. **(c)** Estimated shading shown on the object (rerendering divided by albedo). **(d)** Rerendering of the input based on the estimated albedo and lighting. **(e)** Estimated lighting environment *(left respectively)* and ground truth lighting environment projected onto same degree positive spherical harmonics (right respectively). More results can be found in the supplemental.

tively high order. As opposed to Lombardi and Nishino [33], we precompute the radiance transfer (PRT) [48, 49] to factor the lighting out of the image formation model, making optimization straight forward. To our knowledge, ours is the first reflectance and illumination estimation method that uses multi-bounce precomputed radiance transfer to address the problem of modeling indirect lighting.

3. Foundations

Before we formulate our method we take a look at the rendering equation and a special formulation of it known as *precomputed radiance transfer* (PRT). We use the notation of Dutré *et al*. [11] (see notation overview in the supplemental).

The rendering equation describes the radiance distribution in the scene. For radiance L at point x leaving towards direction Θ:

$$L(x \rightarrow \Theta) = \int_{\Omega_x} f_r(x, \Psi \rightarrow \Theta) L(x \leftarrow \Psi) \cos(N_x, \Psi) \, \mathrm{d}\omega_\Psi \quad (1)$$

Here, f_r is the bidirectional reflectance distribution function (BRDF), N_x the surface normal at point x, and $\cos(N_x, \Psi)$ the cosine of the angle between N_x and Ψ.

This integral equation depends on scene geometry, surface reflectance properties and light emission. Given scene geometry and radiance samples in the form of camera images, we show how to estimate surface reflectance parameters and illumination in Section 4. We assume that the light source is a distant environment L_{env}, independent of position x. To describe this directional quantity, we use a set of spherical harmonics (SH) basis functions $\mathbf{b}(\Theta)$ and associated coefficients $\mathbf{c}_{\text{env}}$:

$$L_{\text{env}}(\Theta) = \mathbf{b}(\Theta)^\intercal \cdot \mathbf{c}_{\text{env}} \quad (2)$$

Light transport is linear w.r.t. the illumination and PRT allows us to find solutions to the rendering equation which make this linearity apparent. PRT lets us map the environment radiance $\mathbf{c}_{\text{env}}$ to the SH basis coefficients $\mathbf{c}(x)$ of the omnidirectional radiance incoming at point x using a precomputed matrix $\mathbf{T}(x)$, which we call *transfer matrix*:

$$\mathbf{c}(x) = L(x \leftarrow \square) = \mathbf{T}(x) \cdot \mathbf{c}_{\text{env}} \quad (3)$$

This linearity holds even in the presence of shadowing and indirect lighting. Based on this formulation we build our method in the next section. Please refer to the supplemental for a more in-depth derivation of the version of PRT we use here.

4. Method

Our method takes as input the object's geometry as well as extrinsically and intrinsically calibrated images of the object. Such data can be acquired for example by means of multi-view stereo methods. Further, all input images need to show the object under the same lighting condition.

4.1. Lighting environment estimation

In the previous section we saw how the incoming radiance at surface point x can be expressed as a linear operator with

(a) (b) (c) (d)

Figure 3. Qualitative comparison between our method and Mélou *et al*. [38] as well as Chen and Koltun [8]. **(a)** The input image emulating the synthetic setup of [38]. **(b)** The result of [38]. Like us they use a mesh and multiple views as input. **(c)** The result of [8]. Their method takes a single RGB-D view as input. **(d)** The result obtained by our method.

respect to the lighting (Eq. 3). Cameras measure a quantity which is proportional to the outgoing radiance $L(x \rightarrow \Theta)$. This outgoing radiance can be found by integrating the incoming radiance reflected towards viewing direction Θ:

$$\begin{aligned} & L(x \rightarrow \Theta) \qquad (4) \\ \overset{(1,3)}{=} & \int_{\Omega_x} f_r(x, \Psi \rightarrow \Theta) \mathbf{b}(\Psi)^\intercal \cos(N_x, \Psi) \, \mathrm{d}\omega_\Psi \cdot \mathbf{T}(x) \cdot \mathbf{c}_{\text{env}} \\ = & \ \boldsymbol{\tau}(x, \Theta)^\intercal \cdot \mathbf{c}_{\text{env}} \end{aligned}$$

where

$$\boldsymbol{\tau}(x, \Theta)^\intercal = \int_{\Omega_x} f_r(x, \Psi \rightarrow \Theta) \mathbf{b}(\Psi)^\intercal \cos(N_x, \Psi) \, \mathrm{d}\omega_\Psi \cdot \mathbf{T}(x) \qquad (5)$$

Thus we can describe the radiance leaving a surface point x towards Θ by the scalar product of a *transfer vector* $\boldsymbol{\tau}(x, \Theta)$ and the environment light coefficient vector $\mathbf{c}_{\text{env}}$. This outgoing radiance is supposed to match the observed radiance values found in the input images.

Let $\boldsymbol{\iota} = (\iota_1, \ldots, \iota_m)^\intercal$ be the linear intensity values observed in the input images and $\mathcal{X} = (x_1, \ldots, x_m)$ the corresponding surface points. $(\Theta_1, \ldots, \Theta_m)$ are the directions pointing from surface point x_i to the image containing observation ι_i. Note that surface points can appear in $\mathcal{X}$ several times if they have been observed in multiple images. We formulate an error measure between the observed intensities $\boldsymbol{\iota}$ and the estimated radiances $L(x_i \rightarrow \Theta_i)$:

$$\begin{aligned} E(\mathbf{c}_{\text{env}}) = \sum_{i=1}^{m} \left(\boldsymbol{\tau}(x_i, \Theta_i)^\intercal \cdot \boldsymbol{c}_{\text{env}} - \iota_i\right)^2 = \|\boldsymbol{\mathcal{T}} \cdot \mathbf{c}_{\text{env}} - \boldsymbol{\iota}\|_2^2 \\ \text{with} \quad \boldsymbol{\mathcal{T}} := (\boldsymbol{\tau}(x_1, \Theta_1), \ldots, \boldsymbol{\tau}(x_m, \Theta_m))^\intercal \end{aligned} \qquad (6)$$

Attention needs to be paid to the fact that the coefficients $\mathbf{c}_{\text{env}}$ with respect to the spherical harmonics basis can give rise to negative values in the estimated lighting environment. To rule out such physically impossible solutions we define a measurement matrix $\mathbf{M}$ that, when multiplied with the coefficient vector $\mathbf{c}_{\text{env}}$, returns a vector of values sampled from the reconstructed lighting environment. Each row $\mathbf{M}_l$ of this matrix consists of the basis functions $b_j(\cdot)$ evaluated in a random direction Φ_l:

$$\mathbf{M}_l = (b_1(\Phi_l), \ldots, b_k(\Phi_l)) \qquad (7)$$

We require all of the environment lighting samples obtained in this way to be positive, effectively constraining the parameter space to a convex polytope of physically plausible solutions. This gives rise to a constrained optimization problem:

$$\begin{aligned} \arg\min_{\mathbf{c}_{\text{env}}} \ E(\mathbf{c}_{\text{env}}) = \arg\min_{\mathbf{c}_{\text{env}}} \|\boldsymbol{\mathcal{T}} \cdot \mathbf{c}_{\text{env}} - \boldsymbol{\iota}\|_2^2 \\ \text{s.t.} \quad \mathbf{0} \leq \mathbf{M} \cdot \mathbf{c}_{\text{env}} \end{aligned} \qquad (8)$$

In practice we independently sample 1,000 uniformly distributed directions Φ_l to construct $\mathbf{M}$.

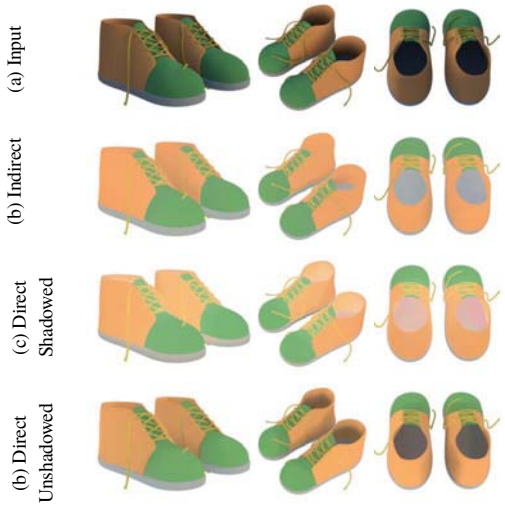


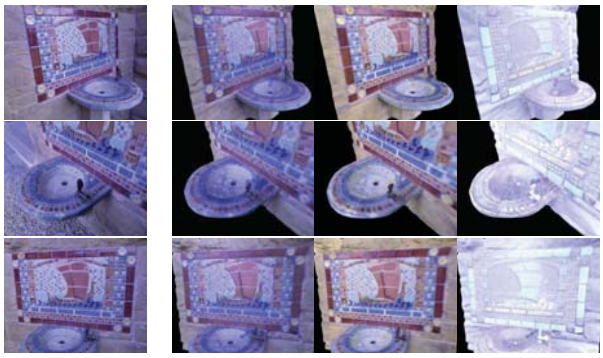
Figure 4. Albedo estimates produced by our method under different lighting model assumptions. **(a)** A subset of the synthetic input images. **(b)** Albedo estimates obtained using the proposed multi-bounce light transport that takes occlusion and indirect light into account. **(c)** Albedo estimates obtained when taking only direct, occluded lighting into account, without consideration for indirect lighting. **(d)** Albedo estimates when taking only direct, unoccluded lighting into account, without consideration for indirect lighting or self-shadowing.

4.2. Surface property estimation

In addition to the lighting environment we also estimate the object's surface properties such as diffuse albedo, specular albedo and roughness, depending on the choice of BRDF. Let $\boldsymbol{\theta}_i$ be the vector of parameters at surface point x_i and let $\boldsymbol{\theta} = (\boldsymbol{\theta}_1^\intercal, \ldots, \boldsymbol{\theta}_m^\intercal)^\intercal$ be the vector of all parameters to be estimated. Since the transfer vectors $\boldsymbol{\tau}$ depend on those parameters we will now write $\boldsymbol{\tau}(x_i, \Theta_i, \boldsymbol{\theta})$ and $\boldsymbol{\mathcal{T}}(\boldsymbol{\theta})$ respectively.

The optimization problem then becomes:

$$\begin{aligned} \arg\min_{\mathbf{c}_{\text{env}}, \boldsymbol{\theta}} \ E(\mathbf{c}_{\text{env}}, \boldsymbol{\theta}) = \arg\min_{\mathbf{c}_{\text{env}}, \boldsymbol{\theta}} \|\boldsymbol{\mathcal{T}}(\boldsymbol{\theta}) \cdot \mathbf{c}_{\text{env}} - \boldsymbol{\iota}\|_2^2 \\ \text{s.t.} \quad \mathbf{0} \leq \mathbf{M} \cdot \mathbf{c}_{\text{env}} \quad \text{and} \quad \mathbf{l} \leq \boldsymbol{\theta} \leq \mathbf{u} \end{aligned} \qquad (9)$$



(a) Input images (b) Estimates

Figure 5. Results on the "Fountain" [59] multi-view stereo dataset picturing a non-diffuse real-world scene with spatially varying albedo and glossiness. **(a)** Subset of the input images. **(b)** Our albedo and shading estimates using the Cook-Torrance reflectance model.

where $\mathbf{l}$ and $\mathbf{u}$ model possible lower and upper bounds on the parameters. Constrained optimization of this objective with respect to the lighting environment coefficients $\mathbf{c}_{\text{env}}$ is done via gradient descent. The partial derivatives of $E(\mathbf{c}_{\text{env}}, \boldsymbol{\theta})$ can be expressed analytically:

$$\frac{\partial\, E(\mathbf{c}_{\text{env}}, \boldsymbol{\theta})}{\partial\, \mathbf{c}_{\text{env}}} = 2\, \boldsymbol{\mathcal{T}}(\boldsymbol{\theta})^{\intercal} \cdot (\boldsymbol{\mathcal{T}}(\boldsymbol{\theta}) \cdot \mathbf{c}_{\text{env}} - \boldsymbol{\iota}) \tag{10}$$

Finding a tractable analytic expression for the partial derivatives with respect to the BRDF parameters $\boldsymbol{\theta}$ is on the other hand not possible in general.

In order to solve this problem we take a closer look at the structure of the terms $\boldsymbol{\tau}(x_i, \Theta_i, \boldsymbol{\theta})^{\intercal} \cdot \boldsymbol{c}_{\text{env}}$, hiding parts that are not dependent on $\boldsymbol{\theta}$ (see Eq. 4):

$$\boldsymbol{\tau}(x_i, \Theta_i, \boldsymbol{\theta})^{\intercal} \cdot \boldsymbol{c}_{\text{env}} = \underbrace{\int_{\Omega_{x_i}} f_r(x_i, \boldsymbol{\theta}_i, \dots) \dots \, \mathrm{d}\omega_{\Psi}}_{=:\, \boldsymbol{\sigma}(x_i, \boldsymbol{\theta}_i)^{\intercal}} \cdot \underbrace{\mathbf{T}(x_i, \boldsymbol{\theta}) \cdot \boldsymbol{c}_{\text{env}}}_{\mathbf{c}(x_i, \boldsymbol{\theta})} \tag{11}$$

Here, $\mathbf{c}(x_i, \boldsymbol{\theta})$ are the spherical basis coefficients describing the total incoming radiance at point x_i (as previously defined in Eq. 3) and $\boldsymbol{\sigma}(x_i, \boldsymbol{\theta}_i)$ describes the last bounce towards the camera.

The exact partial derivatives of Eq. 11 with respect to $\boldsymbol{\theta}_j$ are given by:

$$\frac{\partial\, \boldsymbol{\tau}(x_i, \Theta_i, \boldsymbol{\theta})^{\intercal} \cdot \boldsymbol{c}_{\text{env}}}{\partial\, \boldsymbol{\theta}_j} = \begin{cases} \frac{\partial\, \boldsymbol{\sigma}(x_i, \boldsymbol{\theta}_i)^{\intercal}}{\partial\, \boldsymbol{\theta}_i} \cdot \mathbf{c}(x_i, \boldsymbol{\theta}) + \boldsymbol{\sigma}(x_i, \boldsymbol{\theta}_i)^{\intercal} \cdot \frac{\partial\, \mathbf{c}(x_i, \boldsymbol{\theta})}{\partial\, \boldsymbol{\theta}_i} & \text{if } i = j \\ \boldsymbol{\sigma}(x_i, \boldsymbol{\theta}_i)^{\intercal} \cdot \frac{\partial\, \mathbf{c}(x_i, \boldsymbol{\theta})}{\partial\, \boldsymbol{\theta}_j} & \text{otherwise} \end{cases} \tag{12}$$

While $\boldsymbol{\sigma}(x_i, \boldsymbol{\theta}_i)$—the last bounce towards the camera—is only dependent on the parameters $\boldsymbol{\theta}_i$ at point x_i, the incoming radiance $\mathbf{c}(x_i, \boldsymbol{\theta})$ can depend on parameters from anywhere on the surface dut to indirect lighting.

Lighting	Method	MSE ($\pm\sigma$) $\cdot 10^2$
Indoor	Ours (full)	$\mathbf{0.2916 \pm 0.1425}$
	Ours (single bounce)	0.3796 ± 0.1341
	Ours (unregularized)	0.6850 ± 0.2236
	2D intrinsic	7.6986 ± 3.8926
Sky	Ours (GT)	$\mathbf{0.2150 \pm 0.1383}$
	Ours (smooth)	0.3483 ± 0.1640
	Ours (smooth+noise)	0.3929 ± 0.1815
	2D intrinsic	8.0255 ± 3.0056

Table 1. Quantitative evaluation of our method based on the synthetic "Joyful Yell" dataset (Figure 2, *left*) with known ground truth. The table shows scale invariant MSE values between estimated and ground truth albedo averaged across eight views. The different experimental setups are explained in Section 5.3. For comparison we also report numbers for a 2D intrinsic decomposition [4].

In order to make the derivatives tractable we approximate them by assuming that the last bounce is the dominant term and that the change in incoming radiance due to other surface points' parameters is thus negligible:

$$\frac{\partial\, \boldsymbol{\tau}(x_i, \Theta_i, \boldsymbol{\theta})^{\intercal} \cdot \boldsymbol{c}_{\text{env}}}{\partial\, \boldsymbol{\theta}_j} \approx \begin{cases} \frac{\partial\, \boldsymbol{\sigma}(x_i, \boldsymbol{\theta}_i)^{\intercal}}{\partial\, \boldsymbol{\theta}_i} \cdot \mathbf{c}(x_i, \boldsymbol{\theta}) & \text{if } i = j \\ \mathbf{0} & \text{otherwise} \end{cases} \tag{13}$$

As evidenced by the results shown in later sections, this is a useful approximation to the exact partial derivatives and allows a gradient descent optimization to find good solutions.

4.2.1 BRDF models and parameter regularization

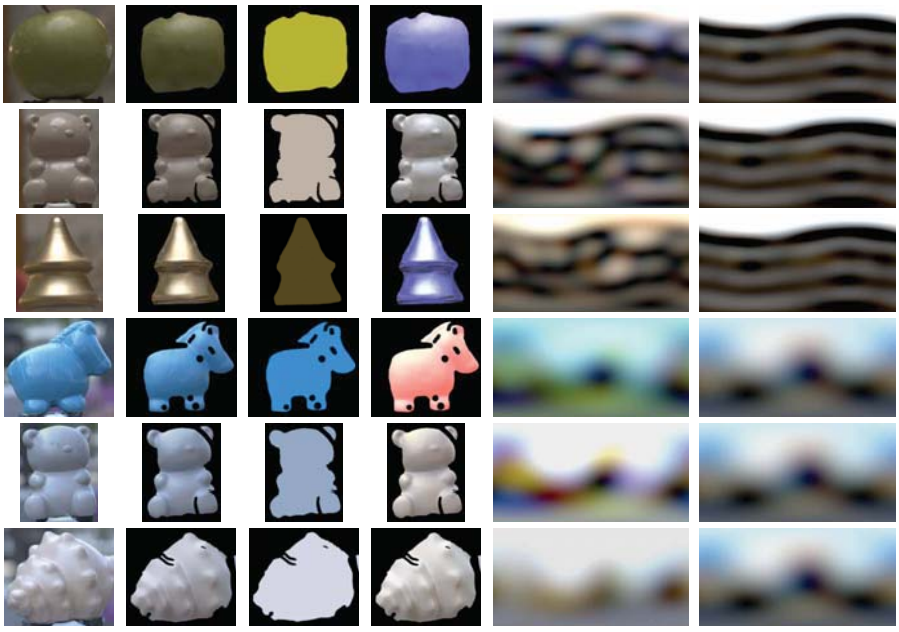
We implemented our method for a diffuse Lambertian BRDF and a Cook-Torrance BRDF for glossy reflectance. The spatially varying BRDF parameters are regularized by pairwise quadratic loss terms between neighboring surface points. Please refer to the supplemental for more details.

4.3. Optimization

In order to optimize the constrained objective $E(\mathbf{c}_{\text{env}}, \theta)$ (Eq. 9) we perform optimization in alternating steps, optimizing for the environment lighting coefficients $\mathbf{c}_{\text{env}}$ every even and for the surface parameters $\boldsymbol{\theta}$ every odd step. Algorithm 1 describes this process.

4.4. Implementation

We implemented our method as a multi-threaded CPU application. For the results we show here we sample the surface points x_i at the vertex positions of the respective meshes. This allows us to store the estimated parameters as vertex properties of the mesh. Sampling at vertex positions is not a requirement of our method though. One could also sample the surface points x_i at texel positions of a mapped texture, obtaining parameters in texture space instead.



(a) Input (b) Rerender (c) Albedo (d) Shading (e) Estimated Environment (f) "GT" Environment

Figure 6. Results of our method on a real-world dataset of glossy objects [32] under two different illuminations. For the results shown here we estimate a single set of BRDF parameters for all of the object's surface. **(a)** The single-view input images. **(b+c+d)** Outputs of our method. The shading picture is computed by dividing the rerendering by the albedo. **(e+f)** The estimated lighting environment as well as the "ground truth" lighting environment. The latter shows the actual lighting environment projected onto a positive spherical harmonics basis of the same degree as used for estimation. This allows to compare the theoretical best possible approximation *(f)* with the actually achieved approximation *(e)*.

5. Results and experiments

In this section we take a look at results achieved on a variety of datasets, theoretical aspects of our method, and practical applications. If not specified otherwise, we report results using three bounces of light (two indirect) and 32 alternating optimization steps for albedo and light respectively.

5.1. Results on diffuse data

First we look at two diffuse datasets: a synthetically generated one that uses the well-known "Joyful Yell" model rendered under an environment with strong yellow and blue hues; and a real-world dataset of a star shaped object captured under natural outdoor illumination. The geometry for the latter was recovered using multi-view stereo [44, 45] and ground truth lighting was captured with a Ricoh Theta Z1 360° camera.

Figure 2 shows a subset of the input images from both datasets and albedo and lighting estimates obtained by our method. Even in the case of the strongly hued lighting environment, our method successfully separated albedo and lighting. Additional results on these datasets using different lighting environments can be found in the supplemental.

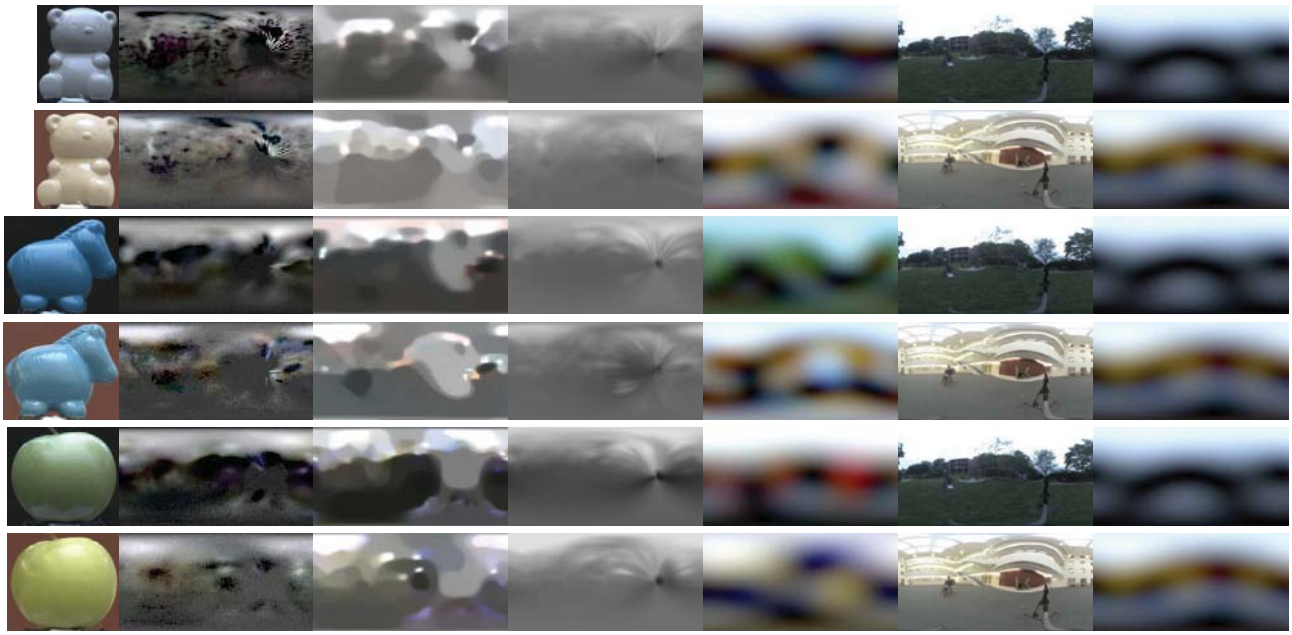
We use the "Joyful Yell" model to qualitatively compare our results to the work of Mélou *et al*. [38] who also use multiple views as well as geometry for reflectance estimation. We also show the result of Chen and Koltun [8] who use single view RGB-D images as input. We matched the experimental setup described by Mélou *et al*. [38] and show the results in Figure 3. Our method manages to predict the influence of the indirect lighting mostly present in the region between the neck and hair and the mouth which in turn leads to an improved albedo estimate.

5.2. Results on glossy data

We also tested our method on two glossy real-world datasets. For these datasets we do not only estimate a diffuse albedo, but also glossiness parameters for the Cook-Torrance BRDF.

"Objects under natural illumination" This dataset by Lombardi and Nishino [32] contains single-view images of glossy objects under different lighting conditions and surface normal maps for each view. We reconstructed partial geometry for each object from the provided surface normals and used this as input to our method. Since these objects have a uniform surface appearance, we estimate a single set of Cook-Torrance BRDF parameters shared by all surface points. Figure 6 visualizes the results. This dataset is challenging since the available geometry is of low quality and the objects are almost convex, exhibiting little to no self-occlusion and indirect lighting. Still, our method was able to recover sensible BRDF parameters and lighting environment estimates.

Fountain The second glossy dataset is the "Fountain" by Zhou and Koltun [59]. This multi-view stereo dataset features partially diffuse, partially glossy, spatially varying surface



(a) Input | (b) Lombardi and Nishino [32] | (c) Lombardi and Nishino [34] | (d) Meka *et al.* [37]† | (e) Our Result | (f) Ground Truth | (g) Ground Truth as Spherical Harmonics

Figure 7. Comparison between the methods of Lombardi and Nishino [32, 34], the method of Meka *et al*. [37] and ours. Results show the environment maps obtained given a single-view input image. **(a)** The input image from the "Objects Under Natural Illumination" dataset [32]. **(b+c)** The results of Lombardi and Nishino **(d)** The result of Meka *et al*. † **(e)** Our result **(f)** The ground truth environment. **(g)** The ground truth projected onto third degree positive spherical harmonics (the best outcome we could possibly achieve).
† Results shown here are adapted from [37] to match exposure and gamma of the other images and were converted from mirror ball to latitude-longitude mapping.

appearance. As such we estimate Cook-Torrance BRDF parameters at each surface point. Figure 5 shows the results we obtained. Our method correctly identified the mosaic tiles as being glossy, while keeping the sand colored wall surrounding the fountain diffuse. For comparative results of a non-glossy (Lambertian) estimation on this dataset, please refer to the supplemental.

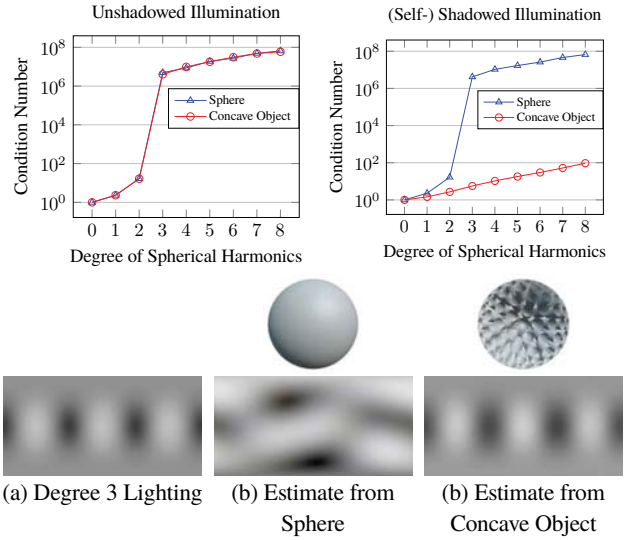
Comparison to related work We use the "Objects Under Natural Illumination" dataset [32] to compare the results of our method to three related works. Lombardi and Nishino [32] compute the maximum a posteriori solution to a Bayesian formulation to jointly solve for reflectance and illumination. Later, the same authors improve upon this method by proposing an enhanced reflectance model [34]. Lastly, Meka *et al*. [37] train a deep network to infer material and illumination information. Figure 7 shows the results of these three methods next to ours.

5.3. Experiments

Sensitivity to inaccurate geometry We tested the degradation in performance due to imperfections in the geometry. Based on the geometry of the "Joyful Yell" model we created two new versions, one which was strongly smoothed and another which was smoothed and then had additional noise applied to it. We estimated the albedo for the ground truth geometry and the degraded versions and computed the difference to the ground truth albedo. Rows "Ours (GT)", "Ours (smooth)", and "Ours (smooth+noise)" in Table 1 report the quantitative results for this experiment. While there is a measurable performance degradation, results are still superior to a 2D intrinsic decomposition which does not make use of geometric information. This means that we can use imperfect geometry obtained from single or multi-view methods and still expect an improved albedo and lighting estimation.

Influence of indirect light An important question is whether accounting for indirect light actually makes a difference. Figure 4 shows the difference between using multiple bounces and using only a single direct bounce for albedo estimation. One can see that results improve significantly in concave, recessed areas when taking indirect lighting into account. The contribution of indirect light is significant in those areas when compared to direct light alone. The rows labeled "ours (full)" and "ours (single bounce)" in Table 1 show that this difference can also be measured quantitatively.

Influence of occluded lighting A question that arises when estimating a lighting environment from indirect observations is what kind of fidelity one can expect from the result. Given



(a) Degree 3 Lighting (b) Estimate from Sphere (b) Estimate from Concave Object

Figure 8. Condition for unshadowed illumination *(top left)* and shadowed illumination *(top right)*. The different lines correspond to different meshes and the number of lighting coefficients increases from left to right. The *bottom* of the figure shows a third degree lighting environment **(a)** and the estimates that could be recovered of it, using a diffuse sphere **(b)** and a diffuse concave object **(c)**.

images of diffuse objects, one certainly does not expect to recover high frequency details in the lighting environment. Ramamoorthi and Hanrahan [42] show that the first 9 spherical harmonics coefficients are sufficient to describe diffuse illumination to an average error of below 1%. Does this mean we cannot hope to recover higher frequency coefficients?

To answer this question we look at our optimization objective (Eq. 9). The condition of $\mathcal{T}^\intercal\mathcal{T}$ gives us insight into the stability of the solution, with a lower condition number indicating a better behaved system. We analyze the condition for different degrees of spherical harmonics, with higher degrees corresponding to a higher number of coefficients, describing higher frequency lighting components.

Figure 8 shows the results of this analysis. We use two different geometries: a convex sphere and a non-convex object. The results for unshadowed transfer (Figure 8, *top left*) show that the condition number increases with increasing lighting environment degree. For both geometries, there is a very sharp rise in condition number by over five orders of magnitude between degrees 2 and 3 (9 and 16 coefficients respectively). This result confirms the observation of Ramamoorthi and Hanrahan [42] which state that 9 coefficients are sufficient to describe unshadowed diffuse illumination.

An interesting thing happens when considering self-occluded radiance transfer (Figure 8, *top right*). While the results for the sphere do not change—it does not exhibit any self-occlusion after all—the system becomes much better conditioned for the concave object. The concave object's surface points now "see" a smaller part of the lighting environment, giving us more information about its higher frequencies.

This indicates that it is possible to recover these higher frequency components from diffuse observations, given that the object's geometry is amenable to the task. We confirm this assumption by rendering the sphere and the concave object under a third degree spherical harmonics lighting environment and reconstructing the lighting environment from those observations respectively (Figure 8, *bottom*). One can see that observing the sphere is not sufficient to recover the original lighting environment, while observing the concave object gives rise to an estimate that faithfully captures the higher frequent lighting environment.

5.4. Application examples: relighting and image editing

Separating the surface reflectance properties from the lighting enables an array of applications, two of which we show here. Figure 1 *(middle)* shows relightings, that is synthesized images which keep the object's original albedo but place it in a different lighting environment.

In a similar vein, we can keep the original lighting environment but change the object's surface properties instead (Figure 1, *bottom*). Here, we replace the object's original color with a cardboard appearance and apply the estimated lighting, resulting in a seamless image edit. Please refer to the supplemental for extended versions of these examples.

6. Conclusion

We presented a method for estimation of object reflectance and scene illumination from RGB images and associated geometry that accounts for both direct and indirect lighting. We showed that indirect lighting is an important factor in the process and proposed a novel solution using precomputed radiance transfer.

In the future we would like to explore the possibility of combining precomputed radiance transfer with a differentiable path tracer. PRT makes the formulation of the optimization problem with respect to the environment lighting very elegant and a differentiable renderer could be used to provide us with exact partial derivatives with respect to the surface parameters which we only approximate at the moment.

We would also like to extend our method to even more challenging scenarios such as very glossy reflectance and very high frequency lighting environments, both of which are not well represented by a spherical harmonics basis. Further it would be interesting to lift the restriction to distant lighting and allow for spatially varying illumination.

Other directions to explore include the joint optimization of the scene geometry, incorporating super-resolution techniques to increase the fidelity of the reconstructed albedo, and methods for illuminant estimation to achieve an even better separation of object albedo and lighting color, all of which pose interesting challenges for future research.

References

[1] E. Adelson and A. Pentland. *The perception of shading and reflectance*, page 409–424. Cambridge University Press, 1996.

[2] D. Azinović, T.-M. Li, A. Kaplanyan, and M. Nießner. Inverse path tracing for joint material and lighting estimation. In *Proc. Computer Vision and Pattern Recognition (CVPR), IEEE*, 2019.

[3] J. T. Barron and J. Malik. Intrinsic Scene Properties from a A Single RGB-D Image. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 17–24, 2013.

[4] S. Bell, K. Bala, and N. Snavely. Intrinsic Images in the Wild. *ACM Transactions on Graphics*, 33(4), 2014.

[5] N. Bonneel, B. Kovacs, S. Paris, and K. Bala. Intrinsic decompositions for image editing. In *Computer Graphics Forum*, volume 36, pages 593–609. Wiley Online Library, 2017.

[6] D. A. Calian, J.-F. Lalonde, P. Gotardo, T. Simon, I. Matthews, and K. Mitchell. From faces to outdoor light probes. In *Computer Graphics Forum*, volume 37, pages 51–61. Wiley Online Library, 2018.

[7] R. Carroll, R. Ramamoorthi, and M. Agrawala. Illumination decomposition for material recoloring with consistent interreflections. In *ACM Transactions on Graphics (TOG)*, volume 30, page 43. ACM, 2011.

[8] Q. Chen and V. Koltun. A Simple Model for Intrinsic Image Decomposition with Depth Cues. In *IEEE International Conference on Computer Vision*, pages 241–248, Dec 2013.

[9] Q. Chen and V. Koltun. A simple model for intrinsic image decomposition with depth cues. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 241–248, 2013.

[10] Y. Dong, G. Chen, P. Peers, J. Zhang, and X. Tong. Appearance-from-motion: Recovering spatially varying surface reflectance under unknown lighting. *ACM Transactions on Graphics (TOG)*, 33(6):1–12, 2014.

[11] P. Dutré, K. Bala, and P. Bekaert. *Advanced global illumination*. AK Peters/CRC Press, 2006.

[12] M.-A. Gardner, K. Sunkavalli, E. Yumer, X. Shen, E. Gambaretto, C. Gagné, and J.-F. Lalonde. Learning to predict indoor illumination from a single image. *ACM Trans. Graph.*, 36(6):176:1–176:14, Nov. 2017.

[13] M. Garon, K. Sunkavalli, S. Hadap, N. Carr, and J.-F. Lalonde. Fast spatially-varying indoor lighting estimation. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2019.

[14] P. V. Gehler, C. Rother, M. Kiefel, L. Zhang, and B. Schölkopf. Recovering intrinsic images with a global sparsity prior on reflectance. In *Advances in Neural Information Processing Systems 24: 25th Annual Conference on Neural Information Processing Systems 2011. Proceedings of a meeting held 12-14 December 2011, Granada, Spain.*, pages 765–773, 2011.

[15] S. Georgoulis, K. Rematas, T. Ritschel, E. Gavves, M. Fritz, L. Van Gool, and T. Tuytelaars. Reflectance and natural illumination from single-material specular objects using deep learning. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 40(8):1932–1947, Aug 2018.

[16] R. Grosse, M. K. Johnson, E. H. Adelson, and W. T. Freeman. Ground truth dataset and baseline evaluations for intrinsic image algorithms. In *IEEE International Conference on Computer Vision*, pages 2335–2342, 2009.

[17] Y. Hold-Geoffroy, A. Athawale, and J.-F. Lalonde. Deep sky modeling for single image outdoor lighting estimation. In *IEEE International Conference on Computer Vision and Pattern Recognition*, 2019.

[18] Y. Hold-Geoffroy, K. Sunkavalli, S. Hadap, E. Gambaretto, and J.-F. Lalonde. Deep outdoor illumination estimation. In *IEEE International Conference on Computer Vision and Pattern Recognition*, 2017.

[19] Z. Hui, A. Chakrabarti, K. Sunkavalli, and A. C. Sankaranarayanan. Learning to separate multiple illuminants in a single image. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 3780–3789, 2019.

[20] S. Jiddi, P. Robert, and E. Marchand. Reflectance and illumination estimation for realistic augmentations of real scenes. In *Mixed and Augmented Reality (ISMAR-Adjunct), 2016 IEEE International Symposium on*, pages 244–249. IEEE, 2016.

[21] J. T. Kajiya. The rendering equation. *SIGGRAPH Comput. Graph.*, 20(4):143–150, Aug. 1986.

[22] Y. Kanamori and Y. Endo. Relighting humans: Occlusion-aware inverse rendering for full-body human images. *ACM Trans. Graph.*, 37(6):270:1–270:11, Dec. 2018.

[23] K. Karsch, K. Sunkavalli, S. Hadap, N. Carr, H. Jin, R. Fonte, M. Sittig, and D. Forsyth. Automatic scene inference for 3d object compositing. *ACM Trans. Graph.*, 33(3):32:1–32:15, June 2014.

[24] E. A. Khan, E. Reinhard, R. W. Fleming, and H. H. Bülthoff. Image-based material editing. In *ACM Transactions on Graphics (TOG)*, volume 25, pages 654–663. ACM, 2006.

[25] N. Kong, P. V. Gehler, and M. J. Black. Intrinsic video. In *European Conference on Computer Vision*, pages 360–375, 2014.

[26] E. H. Land and J. J. McCann. Lightness and retinex theory. *Josa*, 61(1):1–11, 1971.

[27] K. J. Lee, Q. Zhao, X. Tong, M. Gong, S. Izadi, S. U. Lee, P. Tan, and S. Lin. Estimation of intrinsic image sequences from image+depth video. In *European Conference on Computer Vision*, pages 327–340, 2012.

[28] C. LeGendre, W.-C. Ma, G. Fyffe, J. Flynn, L. Charbonnel, J. Busch, and P. Debevec. Deeplight: Learning illumination for unconstrained mobile mixed reality. In *ACM SIGGRAPH 2019 Talks*, SIGGRAPH '19, pages 46:1–46:2. ACM, 2019.

[29] T.-M. Li, M. Aittala, F. Durand, and J. Lehtinen. Differentiable monte carlo ray tracing through edge sampling. *ACM Trans. Graph. (Proc. SIGGRAPH Asia)*, 37(6):222:1–222:11, 2018.

[30] Z. Li, Z. Xu, R. Ramamoorthi, K. Sunkavalli, and M. Chandraker. Learning to reconstruct shape and spatially-varying reflectance from a single image. In *SIGGRAPH Asia 2018 Technical Papers*, page 269. ACM, 2018.

[31] G. Liu, D. Ceylan, E. Yumer, J. Yang, and J.-M. Lien. Material editing using a physically based rendering network. In *Computer Vision (ICCV), 2017 IEEE International Conference on*, pages 2280–2288. IEEE, 2017.

[32] S. Lombardi and K. Nishino. Reflectance and natural illumination from a single image. In *European Conference on Computer Vision*, pages 582–595. Springer, 2012.

[33] S. Lombardi and K. Nishino. Radiometric scene decomposition: Scene reflectance, illumination, and geometry from rgb-d images. In *3D Vision (3DV), 2016 Fourth International Conference on*, pages 305–313. IEEE, 2016.

[34] S. Lombardi and K. Nishino. Reflectance and illumination recovery in the wild. *IEEE transactions on pattern analysis and machine intelligence*, 38(1):129–141, 2016.

[35] R. Maier, K. Kim, D. Cremers, J. Kautz, and M. Nießner. Intrinsic3D: High-Quality 3D Reconstruction by Joint Appearance and Geometry Optimization with Spatially-Varying Lighting. In *IEEE International Conference on Computer Vision*, volume 4, 2017.

[36] D. Mandl, K. M. Yi, P. Mohr, P. M. Roth, P. Fua, V. Lepetit, D. Schmalstieg, and D. Kalkofen. Learning lightprobes for mixed reality illumination. In *Mixed and Augmented Reality (ISMAR), 2017 IEEE International Symposium on*, pages 82–89. IEEE, 2017.

[37] A. Meka, M. Maximov, M. Zollhoefer, A. Chatterjee, H.-P. Seidel, C. Richardt, and C. Theobalt. Lime: Live intrinsic material estimation. In *Proceedings of Computer Vision and Pattern Recognition (CVPR)*, June 2018.

[38] J. Mélou, Y. Quéau, J. Durou, F. Castan, and D. Cremers. Beyond multi-view stereo: Shading-reflectance decomposition. In *Scale Space and Variational Methods in Computer Vision - 6th International Conference, SSVM 2017, Kolding, Denmark, June 4-8, 2017, Proceedings*, pages 694–705, 2017.

[39] M. Nimier-David, D. Vicini, T. Zeltner, and W. Jakob. Mitsuba 2: A retargetable forward and inverse renderer. *Transactions on Graphics (Proceedings of SIGGRAPH Asia)*, 38(6), Dec. 2019.

[40] J. J. Park, A. Holynski, and S. M. Seitz. Seeing the world in a bag of chips. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 1417–1427, 2020.

[41] J. Philip, M. Gharbi, T. Zhou, A. Efros, and G. Drettakis. Multi-view relighting using a geometry-aware network. *ACM Transactions on Graphics (SIGGRAPH Conference Proceedings)*, 38(4), July 2019.

[42] R. Ramamoorthi and P. Hanrahan. An efficient representation for irradiance environment maps. In *Proceedings of the 28th Annual Conference on Computer Graphics and Interactive Techniques, SIGGRAPH 2001, Los Angeles, California, USA, August 12-17, 2001*, pages 497–500, 2001.

[43] R. Ramamoorthi and P. Hanrahan. A signal-processing framework for inverse rendering. In *Proceedings of the 28th annual conference on Computer graphics and interactive techniques*, pages 117–128. ACM, 2001.

[44] J. L. Schönberger and J.-M. Frahm. Structure-from-motion revisited. In *Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016.

[45] J. L. Schönberger, E. Zheng, M. Pollefeys, and J.-M. Frahm. Pixelwise view selection for unstructured multi-view stereo. In *European Conference on Computer Vision (ECCV)*, 2016.

[46] J. Shi, Y. Dong, H. Su, and X. Y. Stella. Learning non-lambertian object intrinsics across shapenet categories. In *Computer Vision and Pattern Recognition (CVPR), 2017 IEEE Conference on*, pages 5844–5853. IEEE, 2017.

[47] Z. Shu, E. Yumer, S. Hadap, K. Sunkavalli, E. Shechtman, and D. Samaras. Neural face editing with intrinsic image disentangling. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 5541–5550, 2017.

[48] P.-P. Sloan, J. Kautz, and J. Snyder. Precomputed radiance transfer for real-time rendering in dynamic, low-frequency lighting environments. In *ACM Transactions on Graphics (TOG)*, volume 21, pages 527–536. ACM, 2002.

[49] P.-P. J. Sloan, B. Luna, and J. Snyder. Local, deformable precomputed radiance transfer. In *SIGGRAPH '05*, 2005.

[50] S. Song and T. Funkhouser. Neural illumination: Lighting prediction for indoor environments. *Proceedings of 33th IEEE Conference on Computer Vision and Pattern Recognition*, 2019.

[51] T. Sun, J. T. Barron, Y.-T. Tsai, Z. Xu, X. Yu, G. Fyffe, C. Rhemann, J. Busch, P. Debevec, and R. Ramamoorthi. Single image portrait relighting. *ACM Trans. Graph.*, 38(4):79:1–79:12, July 2019.

[52] T. Y. Wang, T. Ritschel, and N. J. Mitra. Joint material and illumination estimation from photo sets in the wild. In *Proceedings of International Conference on 3DVision (3DV)*, 2018.

[53] Y. Weiss. Deriving Intrinsic Images from Image Sequences. In *IEEE International Conference on Computer Vision*, volume 2, pages 68–75, 2001.

[54] Y. Yu, K. Huang, W. Chen, and T. Tan. A novel algorithm for view and illumination invariant image matching. *IEEE transactions on image processing*, 21(1):229–240, 2012.

[55] Y. Yu and W. A. P. Smith. Inverserendernet: Learning single image inverse rendering. In *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 3150–3159, 2019.

[56] E. Zhang, M. F. Cohen, and B. Curless. Emptying, refurnishing, and relighting indoor spaces. *ACM Transactions on Graphics (TOG)*, 35(6):1–14, 2016.

[57] J. Zhang, K. Sunkavalli, Y. Hold-Geoffroy, S. Hadap, , J. Eisenmann, and J.-F. Lalonde. All-weather deep outdoor lighting estimation. In *IEEE International Conference on Computer Vision and Pattern Recognition*, 2019.

[58] Q. Zhao, P. Tan, Q. Dai, L. Shen, E. Wu, and S. Lin. A Closed-form Solution to Retinex with Nonlocal Texture Constraints. *IEEE transactions on Pattern Analysis and Machine Intelligence*, 34(7):1437–1444, 2012.

[59] Q. Zhou and V. Koltun. Color map optimization for 3d reconstruction with consumer depth cameras. *ACM Trans. Graph.*, 33(4):155:1–155:10, 2014.

[60] M. Zollhöfer, A. Dai, M. Innmann, C. Wu, M. Stamminger, C. Theobalt, and M. Nießner. Shading-based Refinement on Volumetric Signed Distance Functions. *ACM Transactions on Graphics*, 34(4):96, 2015.