DOI: 10.1111/cgf.14344
Eurographics Symposium on Rendering 2021
A. Bousseau and M. McGuire
(Guest Editors)

Volume 40 (2021), Number 4

Unified Shape and SVBRDF Recovery using Differentiable Monte Carlo Rendering

Fujun Luan[1,3], Shuang Zhao[2], Kavita Bala[1], Zhao Dong[3]

[1]Cornell University
[2]University of California, Irvine
[3]Facebook Reality Labs

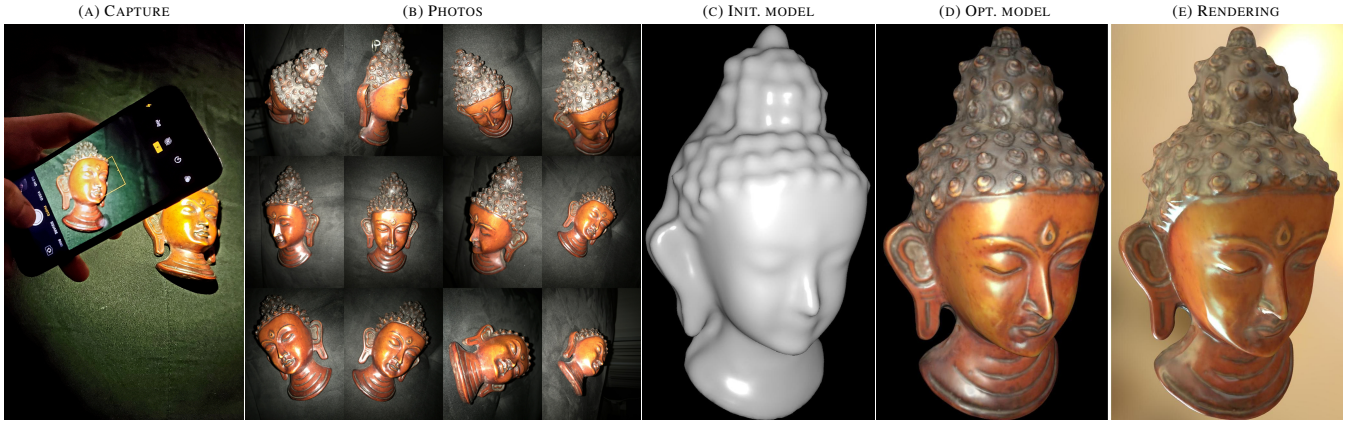


Figure 1: *(a) Capture setup. Our method takes as input: multi-view photos (100 captured and 12 shown for this example) of an object (b) and a rough initial model with its geometry obtained using standard methods (c). Then, a novel analysis-by-synthesis optimization is performed to refine the model's shape and reflectance in a unified fashion, yielding a high-quality 3D model (d). We show in (e) a re-rendering of the result under environmental lighting.*

Abstract

Reconstructing the shape and appearance of real-world objects using measured 2D images has been a long-standing inverse rendering problem. In this paper, we introduce a new analysis-by-synthesis technique capable of producing high-quality reconstructions through robust coarse-to-fine optimization and physics-based differentiable rendering.

Unlike most previous methods that handle geometry and reflectance largely separately, our method unifies the optimization of both by leveraging image gradients with respect to both object reflectance and *geometry. To obtain physically accurate gradient estimates, we develop a new GPU-based Monte Carlo differentiable renderer leveraging recent advances in differentiable rendering theory to offer unbiased gradients while enjoying better performance than existing tools like PyTorch3D [RRN*20] and redner [LADL18]. To further improve robustness, we utilize several shape and material priors as well as a coarse-to-fine optimization strategy to reconstruct geometry. Using both synthetic and real input images, we demonstrate that our technique can produce reconstructions with higher quality than previous methods.*

1. Introduction

Reconstructing the shape and appearance of real-world objects from 2D images has been a long-standing problem in computer vision and graphics. Previously, the acquisition of object geometry and (spatially varying) reflectance has been studied largely independently. For instance, many techniques based on multiview-stereo (MVS) [GHP*08, SSWK13, TFG*13, NLW*16, AWL*15, HSL*17, RPG16, RRFG17] and time-of-flight imaging [NIH*11, IKH*11] have been introduced for the reconstruction of 3D shapes. Although these methods can also provide rough estimations of surface reflectance, they usually rely on the assumption of simple

(e.g., diffuse-dominated) reflectance and can produce unsatisfactory results for glossy objects. On the other hand, previous approaches that specialized at recovering an object's spatially varying reflectance [ZCD*16, GLD*19, GSH*20] typically require object geometries to be predetermined, limiting their practical usage for many applications where such information is unavailable.

Recently, great progress has been made in the area of Monte Carlo differentiable rendering. On the other hand, how this powerful tool can be applied to solve practical 3D reconstruction problems—a main application area of differentiable rendering—has remained largely overlooked. Prior works (e.g., [NLGK18]) have mostly relied on alternative Poisson reconstruction steps during the optimization, leading to suboptimal geometry quality. Instead, by leveraging edge sampling that provides unbiased gradients of mesh vertex positions, we optimize object shape and SVBRDF in a unified fashion, achieving state-of-the-art reconstruction quality.

In this paper, we demonstrate that detailed geometry and spatially varying reflectance of a real-world object can be recovered using a unified *analysis-by-synthesis* framework. To this end, we apply gradient-based optimization of the rendering loss (i.e., the difference between rendered and target images) that are affected by both object geometry and reflectance. Although such gradients with respect to appearance are relatively easy to compute, the geometric gradients are known to be much more challenging to compute and, therefore, have been mostly approximated in the past using techniques like soft rasterization [LLCL19] in computer vision. We, on the other hand, leverage recent advances in physics-based differentiable rendering to obtain *unbiased* and *consistent* geometric gradients that are crucial for obtaining high-quality reconstructions.

Concretely, our contributions include:

- A Monte Carlo differentiable renderer specialized for collocated configurations. Utilizing edge sampling [LADL18], our renderer produces unbiased and consistent gradient estimates.
- A new analysis-by-synthesis pipeline that enables high-quality reconstruction of spatially varying reflectance and, more importantly, mesh-based object geometry.
- A coarse-to-fine scheme as well as geometry and reflectance priors for ensuring robust reconstructions.
- Thorough validations and evaluations of individual steps that come together allowing practical and high-quality 3D reconstruction using inexpensive handheld acquisition setups, which benefits applications in many areas like graphics and AR/VR.

We demonstrate the effectiveness of our technique via several synthetic and real examples.

2. Related work

Shape reconstruction. Reconstructing object geometry has been a long-standing problem in computer vision.

Multi-view Stereo (MVS) recovers the 3D geometry of sufficiently textured objects using multiple images of an object by matching feature correspondences across views and optimizing photo-consistency (e.g., [SD99, VTC05, FP09, SZPF16]).

Shape from Shading (SfS) relates surface normals to image intensities [Hor70, IH81, QMC*17, QMDC17, MKC*17, HQMC18]. Unfortunately, these methods have difficulties handling illumination changes, non-diffuse reflectance, and textureless surfaces.

Photometric Stereo (PS) takes three or more images captured with a static camera and varying illumination or object pose, and directly estimate surface normals from measurements [Woo80, HLHZ08, ZT10, TFG*13, PF14, QLD15, QMD16]. These methods typically do not recover reflectance properties beyond diffuse albedo.

Reflectance reconstruction. Real-world objects exhibit richly diverse reflectance that can be described with spatially-varying bidirectional reflectance distribution functions (SVBRDFs).

Traditional SVBRDF acquisition techniques rely on dense input images measured using light stages or gantry (e.g, [Mat03, LKG*03, HLZ10, DWT*10, CDP*14, DWMG15, KCW*18]). To democratize the acquisition, some recent works exploit the structure (e.g., sparsity) of SVBRDF parameter spaces to allow reconstructions using fewer input images (e.g., [YDMH99, DCP*14, WWZ15, ZCD*16, KGT*17, PNS18]). Additionally, a few recent works have been introduced to produce plausible SVBRDF estimations for flat objects using a small number of input images (e.g., [AWL*15, AAL16, HSL*17, GLD*19, DAD*19, GSH*20]). Despite their ease of use, these techniques cannot be easily generalized to handle more complex shapes.

Joint estimation of shape and reflectance. Several prior works jointly estimate object shape and reflectance. Higo et al. [HMJI09] presented a plane-sweeping method for albedo, normal and depth estimation. Xia et al. [XDPT16] optimized an apparent normal field with corresponding reflectance. Nam et al. [NLGK18] proposed a technique that alternates between material-, normal-, and geometry-optimization stages. Schmitt et al. [SDR*20] perform joint estimation using a hand-held sensor rig with and 12 point light sources. Bi et al. [BXS*20] use six images and optimize object geometry and reflectance in two separate stages.

All these methods either rely on MVS for geometry reconstruction or perform alternative optimization of shape and reflectance, offering little to no guarantee on qualities of the reconstruction results. We, in contrast, formulate the problem as a unified analysis-by-synthesis optimization, ensuring locally optimal results.

Differentiable rendering of meshes. We now briefly review differentiable rendering techniques closely related to our work. For a more comprehensive summary, please see the survey by Kato et al. [KBM*20].

Specialized differentiable renderers have long existed in computer graphics and vision [GZB*13, GLZ16, TSG19, ALKN19, CLZ*20]. Recently, several general-purpose ones [LADL18, NDVZJ19] have been developed.

A key technical challenge in differentiable rendering is to estimate gradients with respect to object geometry (e.g., positions of mesh vertices). To this end, several approximated methods (e.g., [LHJ19, LLCL19, RRN*20]) have been proposed. Unfortunately, inaccuracies introduced by these techniques can lead to

degraded result quality. On the contrary, Monte Carlo edge sampling [LADL18,ZWZ*19], which we use for our differentiable renderer, provides unbiased gradient estimates capable of producing higher-quality reconstructions.

3. Our method

We formulate the problem of joint estimation of object geometry and reflectance as an *analysis-by-synthesis* (aka. inverse-rendering) optimization. Let $\boldsymbol{\xi}$ be some vector that depicts both the geometry and the reflectance of a real-world object. Taking as input a set of images $\tilde{\mathcal{I}}$ of this object, we estimate $\boldsymbol{\xi}$ by minimizing a predefined loss $\mathcal{L}$:

$$\boldsymbol{\xi}^* = \arg\min_{\boldsymbol{\xi}} \mathcal{L}(\mathcal{I}(\boldsymbol{\xi}), \boldsymbol{\xi}; \tilde{\mathcal{I}}), \tag{1}$$

where $\mathcal{I}(\boldsymbol{\xi})$ are a set of renderings of the object generated using the geometry and reflectance provided by $\boldsymbol{\xi}$. We further allow the loss $\mathcal{L}$ to directly depend on the parameters $\boldsymbol{\xi}$ for regularization.

Acquisition setup. Similar to recent works on reflectance capture [AWL*15, ACGO18, HSL*17, RPG16, AAL16, GLD*19, DAD*19, RWS*11], we utilize an acquisition setup where the object is illuminated with a point light collocated with the camera. This collocated configuration significantly simplifies both forward and differentiable rendering processes, allowing the analysis-by-synthesis problem to be solved efficiently. Common collocated configurations include a smartphone's flash and camera as well as a consumer-grade RGBD sensor mounted with an LED light.

Overview of our method. Efficiently solving the optimization of Eq. (1) requires computing gradient $\mathrm{d}\mathcal{L}/\mathrm{d}\boldsymbol{\xi}$ of the loss $\mathcal{L}$ with respect to the geometry and reflectance parameters $\boldsymbol{\xi}$. According to the chain rule, we know that

$$\frac{\mathrm{d}\mathcal{L}}{\mathrm{d}\boldsymbol{\xi}} = \frac{\partial\mathcal{L}}{\partial\mathcal{I}}\frac{\mathrm{d}\mathcal{I}}{\mathrm{d}\boldsymbol{\xi}} + \frac{\partial\mathcal{L}}{\partial\boldsymbol{\xi}}, \tag{2}$$

where $\partial\mathcal{L}/\partial\mathcal{I}$ and $\partial\mathcal{L}/\partial\boldsymbol{\xi}$ can be computed using automatic differentiation [PGC*17]. Further, estimating gradients $\mathrm{d}\mathcal{I}/\mathrm{d}\boldsymbol{\xi}$ of rendered images requires performing differentiable rendering. Despite being relatively easy when the parameters $\boldsymbol{\xi}$ only capture reflectance, differentiating the rendering function $\mathcal{I}$ becomes much more challenging when $\boldsymbol{\xi}$ also controls object geometry [LADL18]. To this end, we develop a new differentiable renderer that is specific to our acquisition setup and provides unbiased gradient estimates.

In the rest of this section, we provide a detailed description of our technique that solves the analysis-by-synthesis optimization (1) in an efficient and robust fashion. In §3.1, we detail our forward-rendering model and explain how it can be differentiated. In §3.2, we discuss our choice of the loss $\mathcal{L}$ and optimization strategy.

3.1. Forward and differentiable rendering

In what follows, we describe (i) our representation of object geometry and reflectance; and (ii) how we render these representations in a differentiable fashion.

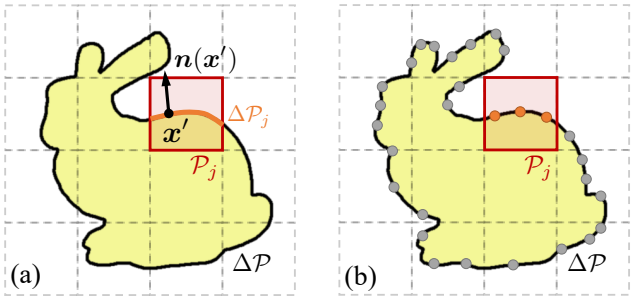


Figure 2: ***Differentiable rendering:*** *(a) To properly differentiate the intensity I_j of a pixel $\mathcal{P}_j$ (illustrated as red squares) with respect to object geometry, a boundary integral over $\Delta\mathcal{P}_j$ (illustrated as the orange curve) needs to be calculated. (b) We perform Monte Carlo edge sampling [LADL18, ZWZ*19] by (i) sampling points (illustrated as small discs) from pre-computed discontinuity curves $\Delta\mathcal{P}$, and (ii) accumulating their contributions in the corresponding pixels (e.g., the orange samples contribute to the pixel $\mathcal{P}_j$).*

Object geometry and reflectance. We express object geometries using standard triangle meshes. Compared to other representations that are popular in 3D reconstruction, such as SDF volumes [PFS*19,JJHZ20,ZLW*21], occupancy networks [MON*19] or sphere-based clouds [Las20], triangle meshes can be efficiently rendered and edited with many 3D digital content creation tools. Further, as a widely adopted format, triangle meshes can be easily imported into numerous applications in computer graphics, vision, and augmented/virtual reality (AR/VR).

One of the biggest challenges when using meshes for 3D reconstruction is that topological changes are difficult. We show in the following sections that this can be addressed by using reasonable initial geometries and a coarse-to-fine optimization process.

To depict an object's spatially varying reflectance, we use the Disney BRDF [KG13], a parametric model offering a good balance between simplicity and flexibility. This model has also been used by many prior works (e.g., [LSC18, LSR*20, BXS*20]). Using this BRDF model, the spatially varying reflectance of an object is described using three 2D texture maps specifying, respectively, diffuse albedo a_d, specular albedo a_s, surface roughness α. And surface normals $\boldsymbol{n}$ are computed from updated mesh vertex positions at every step. Thanks to the efficiency of our system (which we will present in the following), we directly use fine meshes to express detailed geometries and do not rely on approximations like bump/normal mapping.

Forward rendering. Given a virtual object depicted using parameters $\boldsymbol{\xi}$, we render one-bounce reflection (aka. direct illumination) of the object. Specifically, assume the point light and the camera are collocated at some $\boldsymbol{o} \in \mathbb{R}^3$. Then, the intensity I_j of the j-pixel is given by an area integral over the pixel's footprint $\mathcal{P}_j$, which is typically a square on the image plane:

$$I_j = \frac{1}{|\mathcal{P}_j|}\int_{\mathcal{P}_j} \underbrace{I_\mathrm{e}\frac{f_\mathrm{r}(\boldsymbol{o}\to\boldsymbol{y}\to\boldsymbol{o})}{\|\boldsymbol{y}-\boldsymbol{o}\|^2}}_{=:I(\boldsymbol{x})}\,\mathrm{d}A(\boldsymbol{x}), \tag{3}$$

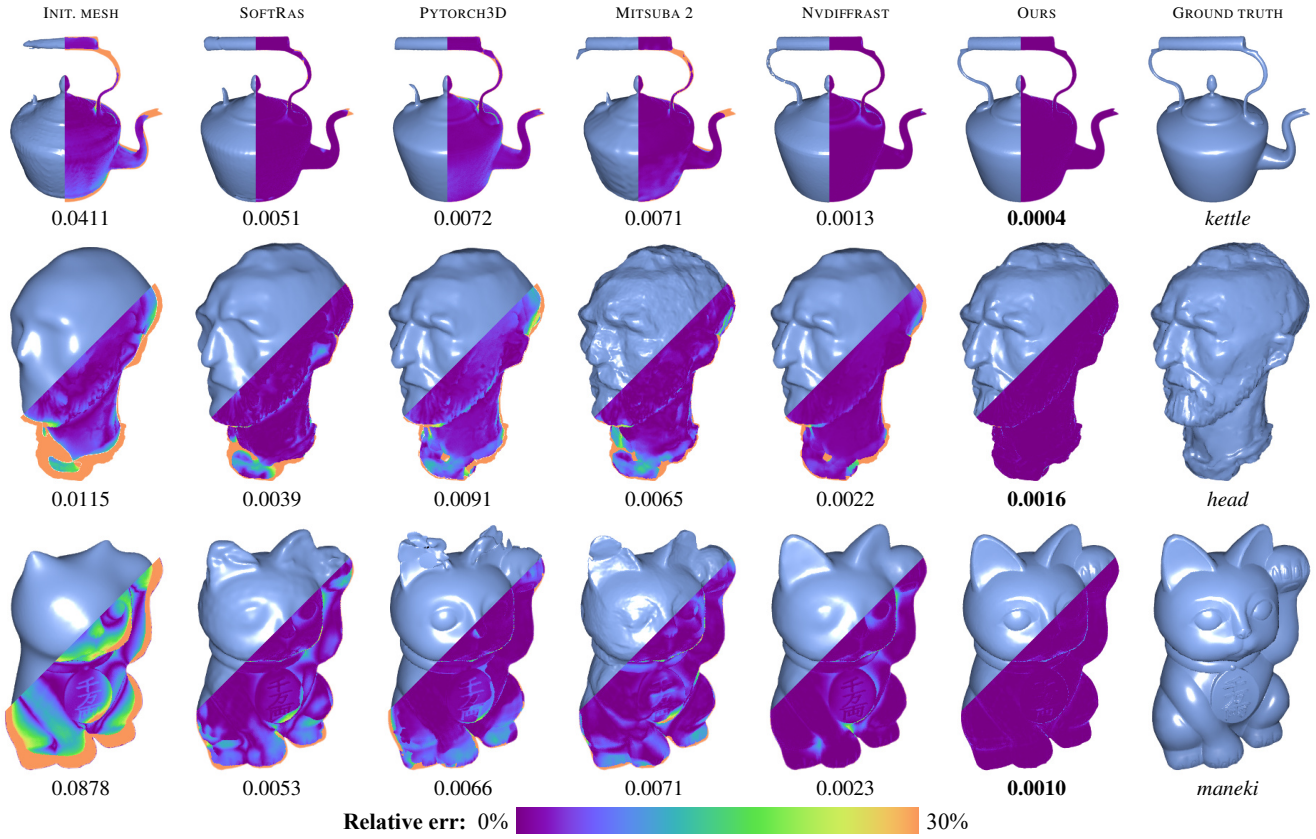


Figure 3: ***Comparison*** *with SoftRas [LLCL19], PyTorch3D [RRN*20], Mitsuba 2 [NDVZJ19] and Nvdiffrast [LHK*20]. We render all reconstructed geometries using Phong shading and visualize depth errors (wrt. the ground-truth geometry). Initialized with the same mesh (shown in the left column), optimizations using gradients obtained with SoftRas and PyTorch3D tend to converge to low-quality results due to gradient inaccuracies caused by soft rasterization. Mitsuba 2, a ray-tracing-based system, also produces visible artifacts due to biased gradients resulting from an approximated reparameterization [LHJ19]. Nvdiffrast is using multisample analytic antialiasing method to provide reliable visibility gradients, which yields better optimization result overall. When using gradients generated with our differentiable renderer, optimizations under identical configurations produce results closely resembling the targets. The number below each result indicates the average point-to-mesh distance capturing the Euclidean accuracy [JDV*14] of the reconstructed geometry (normalized to have a unit bounding box).*

where $\boldsymbol{y}$ is the intersection between the object geometry $\mathcal{M}$ and a ray that originates at $\boldsymbol{o}$ and passes through $\boldsymbol{x}$ on the image plane. Further, I_e denotes the intensity of the point light; $f_\mathrm{r}(\boldsymbol{o} \to \boldsymbol{y} \to \boldsymbol{o})$ indicates the cosine-weighted BRDF at $\boldsymbol{y}$ (evaluated with both the incident and the outgoing directions pointing toward $\boldsymbol{o}$); and A is the surface-area measure. We note that no visibility check is needed in Eq. (3) since, under the collocated configuration, any point $\boldsymbol{y} \in \mathcal{M}$ visible to the camera must be also visible to the light source.

We estimate Eq. (3) using Monte Carlo integration by uniformly sampling N locations $\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_N \in \mathcal{P}_j$ and computing $I_j \approx \frac{1}{N} \sum_{i=1}^{N} I(\boldsymbol{x}_i)$ where I is the integrand defined in Eq. (3).

Differentiable rendering. Computing image gradients $\mathrm{d}\mathcal{I}/\mathrm{d}\boldsymbol{\xi}$ in Eq. (2) largely boils down to differentiating pixel intensities Eq. (3) with respect to $\boldsymbol{\xi}$. Although this can sometimes be done by differentiating the integrand I—that is, by estimating $\int_{\mathcal{P}_j} (\mathrm{d}I/\mathrm{d}\boldsymbol{\xi})\, \mathrm{d}A$—doing so is insufficient when computing gradients with respect to object geometry (e.g., vertex positions). Consequently, the gradient $\mathrm{d}\mathcal{I}/\mathrm{d}\boldsymbol{\xi}$ has usually been approximated using soft rasterization [LLCL19, RRN*20] or reparameterized integrals [LHJ19]. Biased gradient estimates, unfortunately, can reduce the quality of optimization results, which we will demonstrate in §4.

On the other hand, a few general-purpose unbiased techniques [LADL18, ZWZ*19] have been introduced recently. Unfortunately, these methods focus on configurations without point light sources—which is not the case under our collocated configuration. We, therefore, derive the gradient $\mathrm{d}I_j/\mathrm{d}\boldsymbol{\xi}$ utilizing mathematical tools used by these works. Specifically, according to Reynolds transport theorem [Rey03], the gradient involves an *interior* and a *boundary*

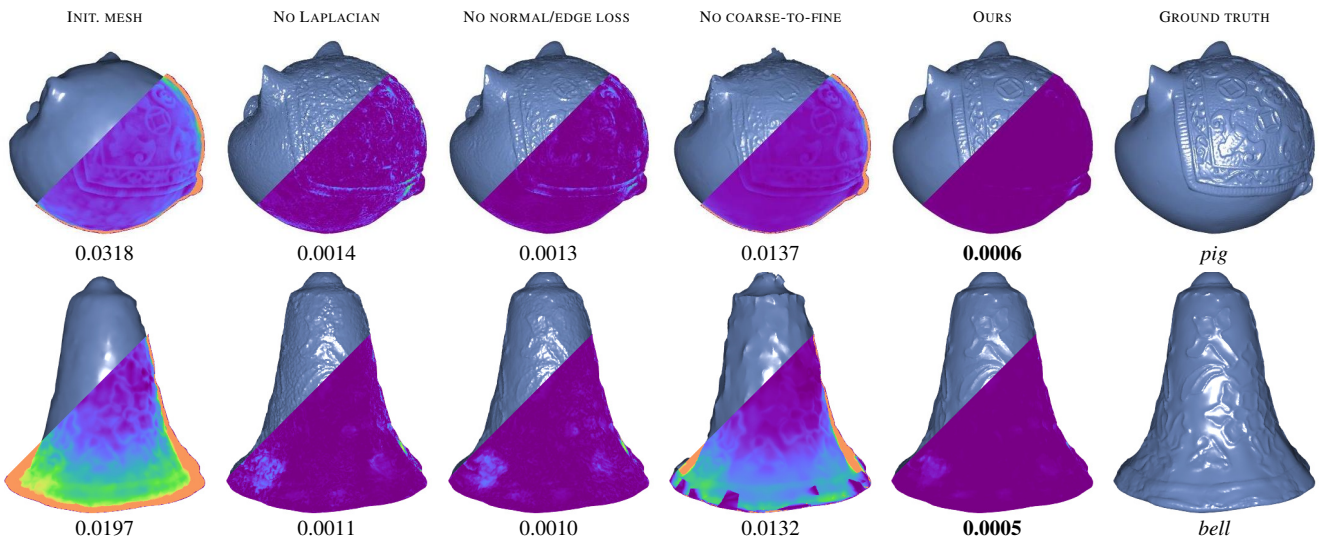


Figure 4: ***Ablation study*** *on our mesh loss of Eq. (9) and coarse-to-fine framework. Using the identical initializations and optimization settings, we show geometries (rendered under a novel view) optimized with (i) various components of the mesh loss; and (ii) the coarse-to-fine process disabled. Similar to Figure 3, the number below each result indicates the average point-to-mesh distance.*

integrals:

$$\frac{\mathrm{d}I_j}{\mathrm{d}\boldsymbol{\xi}} = \frac{1}{|\mathcal{P}_j|}\left[\overbrace{\int_{\mathcal{P}_j} \frac{\mathrm{d}I}{\mathrm{d}\boldsymbol{\xi}}(\boldsymbol{x})\,\mathrm{d}A(\boldsymbol{x})}^{\text{interior}} + \overbrace{\int_{\Delta\mathcal{P}_j} \left(\boldsymbol{n}(\boldsymbol{x}')\cdot\frac{\mathrm{d}\boldsymbol{x}'}{\mathrm{d}\boldsymbol{\xi}}\right)\Delta I(\boldsymbol{x}')\,\mathrm{d}\ell(\boldsymbol{x}')}^{\text{boundary}}\right], \tag{4}$$

where the interior term is simply Eq. (3) with its integrand I differentiated. The boundary one, on the contrary, is over curves $\Delta\mathcal{P}_j := \Delta\mathcal{P} \cap \mathcal{P}_j$ with $\Delta\mathcal{P}$ comprised of jump discontinuity points of I. In practice, $\Delta\mathcal{P}$ consists of image-plane projections of the object's silhouettes. Further, $\boldsymbol{n}(\boldsymbol{x})$ is the curve normal within the image plane, $\Delta I(\boldsymbol{x})$ denotes the difference in I across discontinuity boundaries, and ℓ is the curve-length measure (see Figure 2-a).

Similar to the Monte Carlo estimation of Eq. (3), we estimate the interior integral in Eq. (4) by uniformly sampling $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_N \in \mathcal{P}_j$. To handle the boundary integral, we precompute the discontinuity curves $\Delta\mathcal{P}$ (as polylines) by projecting the object's silhouette onto the image plane at each iteration. At runtime, we draw $\boldsymbol{x}'_1, \ldots, \boldsymbol{x}'_M \in \Delta\mathcal{P}$ uniformly. Then,

$$\frac{\mathrm{d}I_j}{\mathrm{d}\boldsymbol{\xi}} \approx \overbrace{\frac{1}{N}\sum_{i=1}^{N} \frac{\mathrm{d}I}{\mathrm{d}\boldsymbol{\xi}}(\boldsymbol{x}_i)}^{\text{interior}} + \overbrace{\frac{1}{M}\frac{|\Delta\mathcal{P}|}{|\mathcal{P}_j|}\sum_{i=1}^{M} \mathbb{1}[\boldsymbol{x}'_i \in \mathcal{P}_j]\left(\boldsymbol{n}(\boldsymbol{x}'_i)\cdot\frac{\mathrm{d}\boldsymbol{x}'_i}{\mathrm{d}\boldsymbol{\xi}}\right)\Delta I(\boldsymbol{x}'_i)}^{\text{boundary}}, \tag{5}$$

where $|\Delta\mathcal{P}|$ denotes the total length of the discontinuity curves $\Delta\mathcal{P}$, and $\mathbb{1}[\cdot]$ is the indicator function.

In practice, we estimate gradients of pixel intensities via Eq. (5) in two rendering passes. In the first pass, we evaluate the interior component independently for each pixel. In the second pass, we evaluate the boundary component for each $\boldsymbol{x}'_i$ in parallel and accumulate the results in the corresponding pixel (see Figure 2-b).

3.2. Analysis-by-synthesis optimization

We now present our analysis-by-synthesis optimization pipeline that minimizes Eq. (1).

Object parameters. As stated in §3.1, we depict object geometry using a triangle mesh (which is comprised of per-vertex positions $\boldsymbol{p}$ and UV coordinates $\boldsymbol{u}$ as well as per-triangle vertex indices) and reflectance using three 2D texture maps specifying the object's spatially varying diffuse albedo a_d, specular albedo a_s, and surface roughness α, respectively. In this way, our combined geometry and reflectance parameters are given by $\boldsymbol{\xi} = (\boldsymbol{p}, \boldsymbol{u}, a_\mathrm{d}, a_\mathrm{s}, \alpha)$. Note, we do not modify the connectivity of the triangle vertices and rely on additional re-meshing steps, which we will discuss in §3.4, to improve mesh topology.

Loss. A key ingredient in our analysis-by-synthesis optimization is the loss $\mathcal{L}$. Let $\tilde{\mathcal{I}} := (\tilde{\boldsymbol{I}}_1, \tilde{\boldsymbol{I}}_2, \ldots)$ be a set of images of some object (with camera location and pose calibrated for each image $\tilde{\boldsymbol{I}}_k$). Then, our loss takes the form:

$$\mathcal{L}(\mathcal{I}(\boldsymbol{\xi}), \boldsymbol{\xi}; \tilde{\mathcal{I}}) := \mathcal{L}_\mathrm{rend}(\mathcal{I}(\boldsymbol{\xi}); \tilde{\mathcal{I}}) + \mathcal{L}_\mathrm{reg}(\boldsymbol{\xi}), \tag{6}$$

where $\mathcal{L}_\mathrm{rend}$ is the *rendering* loss that measures the difference between rendered and target object appearances. Specifically, we set

$$\mathcal{L}_\mathrm{rend}(\mathcal{I}(\boldsymbol{\xi}); \tilde{\mathcal{I}}) := \lambda_\mathrm{rend} \sum_k \left\| \Phi_k(\boldsymbol{I}_k(\boldsymbol{\xi})) - \Phi_k(\tilde{\boldsymbol{I}}_k) \right\|_1, \tag{7}$$

where $\lambda_\mathrm{rend} > 0$ is a user-specified weight, $\mathcal{I}(\boldsymbol{\xi}) := (\boldsymbol{I}_1(\boldsymbol{\xi}), \boldsymbol{I}_2(\boldsymbol{\xi}), \ldots)$ denotes images rendered using our forward-rendering model of Eq. (3) with object geometry and reflectance

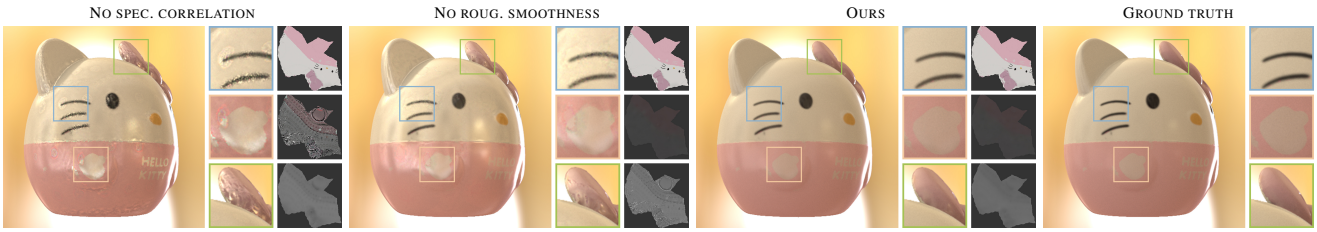


Figure 5: ***Ablation study*** *on our material loss of Eq. (10). Using identical initial reflectance maps and optimization configurations, models optimized with various components of the material loss neglected are rendered under a novel environmental illumination. On the right of each reconstruction result, we show the optimized reflectance maps (from top to bottom: diffuse albedo, specular albedo, and roughness).*

Figure 6: ***Ablation study*** *on number of input images. For each object, we show a novel-view rendering of our reconstruction under environmental lighting, given varying number of input images. The input images have viewing/lighting positions uniformly sampled around the object.*

specified by $\boldsymbol{\xi}$ (under identical camera configurations as the input images), and Φ_k captures pixel-wise post-processing operations such as tone-mapping and background-removing masking.

We estimate gradients of the rendering loss of Eq. (7) with respect to the object parameters $\boldsymbol{\xi}$ using our differentiable rendering method described in §3.1. We will demonstrate in §4.1 that accurate gradients are crucial to obtain high-quality optimization results.

In Eq. (6), $\mathcal{L}_{\text{reg}}(\boldsymbol{\xi})$ is a *regularization* term for improving the robustness of the optimization, which we will discuss in §3.3. Gradients of this term can be obtained easily using automatic differentiation.

Optimization process. Like any other analysis-by-synthesis method, our technique takes as input an initial configuration of an object's geometry and reflectance. In practice, we initialize object geometry using MVS or Kinect Fusion. Our technique is capable of producing high-quality reconstructions using crude initializations (obtained using low-resolution and noisy inputs). For the reflectance maps, we simply initialize them as constant-valued textures.

Provided an initial configuration of the object's geometry and reflectance, we minimize the loss of Eq. (6) using the Adam algorithm [KB14].

Further, to make the optimization more robust, we leverage a coarse-to-fine approach that periodically performs remeshing and upsamples the reflectance-describing textures. We will provide more details on this process in §3.4.

3.3. Regularization

Using only the rendering loss $\mathcal{L}_{\text{rend}}$ expressed in Eq. (7) can make the optimization unstable and/or converge to local minima. To address this problem, we regularize the optimization by introducing another loss $\mathcal{L}_{\text{reg}}$ that in turn consists of a *material* loss $\mathcal{L}_{\text{mat}}$ and a *mesh* one $\mathcal{L}_{\text{mesh}}$:

$$\mathcal{L}_{\text{reg}}(\boldsymbol{\xi}) := \mathcal{L}_{\text{mesh}}(\mathcal{M}) + \mathcal{L}_{\text{mat}}(a_{\text{d}}, a_{\text{s}}, \alpha), \tag{8}$$

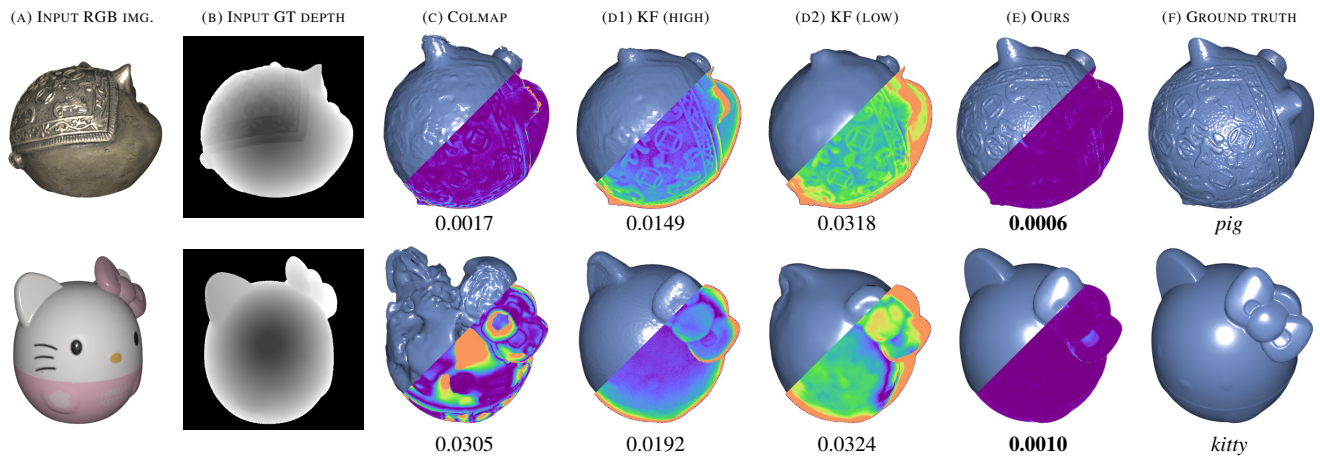


Figure 7: ***Comparison*** *with COLMAP [SZPF16] and Kinect Fusion [NIH*11] using synthetic inputs. The* COLMAP *results (c) are generated using 50 RGB images (a) with exact camera poses; the* KF-High *(d1) and* KF-Low *(d2) results are created using 50 ground-truth depth images (b) and low-resolution noisy ones, respectively. Our method (e), when initialized with* KF-Low *(d2) and using RGB inputs (a), produces much more accurate geometries than* COLMAP *and* KF-High*. Similar to Figure 3, the numbers indicate average point-to-mesh distances.*

which we will discuss in the following.

Mesh loss. We encourage our optimization to return "smooth" object geometry by introducing a *mesh loss*:

$$\mathcal{L}_{\text{mesh}}(\mathcal{M}) := \mathcal{L}_{\text{lap}}(\mathcal{M}) + \mathcal{L}_{\text{normal}}(\mathcal{M}) + \mathcal{L}_{\text{edge}}(\mathcal{M}), \quad (9)$$

where the *mesh-Laplacian loss* $\mathcal{L}_{\text{lap}}$ of a mesh with n vertices is given by $\mathcal{L}_{\text{lap}}(\mathcal{M}) := \lambda_{\text{lap}} \|\boldsymbol{L}\boldsymbol{V}\|^2$ where $\boldsymbol{V}$ is an $n \times 3$ matrix with its i-th row storing coordinates of the i-th vertex, and $\boldsymbol{L} \in \mathbb{R}^{n \times n}$ is the mesh's Laplacian matrix [NISA06].

Additionally, we use a *normal-consistency loss* $\mathcal{L}_{\text{normal}}$ to encourage normals of adjacent faces to vary slowly by setting $\mathcal{L}_{\text{normal}}(\mathcal{M}) := \lambda_{\text{normal}} \sum_{i,j} [1 - (\boldsymbol{n}_i \cdot \boldsymbol{n}_j)]^2$, where the sum is over all pairs (i, j) such that the i-th and the j-th triangles share a common edge, and $\boldsymbol{n}_i$ and $\boldsymbol{n}_j$ denote the normals of these triangles.

Lastly, we penalize the mesh for having long edges, which usually yield ill-shaped triangles, by utilizing an *edge-length loss* $\mathcal{L}_{\text{edge}} := \lambda_{\text{edge}} \left(\sum_i e_i^2\right)^{1/2}$, where e_i denotes the length of the i-th face edge.

Material loss. Our material loss $\mathcal{L}_{\text{mat}}$ regularizes the reflectance maps representing diffuse albedo a_{d}, specular albedo a_{s}, and surface roughness α. Specifically, we set

$$\mathcal{L}_{\text{mat}}(a_{\text{d}}, a_{\text{s}}, \alpha) := \mathcal{L}_{\text{spec}}(a_{\text{d}}, a_{\text{s}}) + \mathcal{L}_{\text{roug}}(\alpha), \quad (10)$$

where $\mathcal{L}_{\text{spec}}$ correlates diffuse and specular albedos [SDR*20]: assuming nearby pixels with similar diffuse albedos to have similar specular ones, we set $\mathcal{L}_{\text{spec}}(a_{\text{s}}, a_{\text{d}}) := \lambda_{\text{spec}} \sum_{\boldsymbol{p}} \|a_{\text{s}}[\boldsymbol{p}] - (\sum_{\boldsymbol{q}} a_{\text{s}}[\boldsymbol{q}] \mu_{\boldsymbol{p},\boldsymbol{q}}) / (\sum_{\boldsymbol{q}} \mu_{\boldsymbol{p},\boldsymbol{q}})\|_1$, where $\mu_{\boldsymbol{p},\boldsymbol{q}} := \exp\left(- \frac{\|\boldsymbol{p}-\boldsymbol{q}\|_2^2}{2\sigma_1^2} - \frac{(a_{\text{d}}[\boldsymbol{p}] - a_{\text{d}}[\boldsymbol{q}])^2}{2\sigma_2^2} \right)$ is the bilateral weight between pixels with indices $\boldsymbol{p}, \boldsymbol{q} \in \mathbb{Z}^2$.

Spatially varying surface roughness is known to be challenging to optimize even when the object geometry is known [GLD*19]. To regularize our optimization of surface roughness, we introduce a smoothness term that measures its total variation: $\mathcal{L}_{\text{roug}}(\alpha) := \lambda_{\text{roug}} \sum_{i,j} (|\alpha[i+1, j] - \alpha[i, j]| + |\alpha[i, j+1] - \alpha[i, j]|)$, where $\alpha[i, j]$ indicate the value of the (i, j)-th pixel in the roughness map.

3.4. Improving robustness

As described in §3.2, when minimizing the loss of Eq. (6), we keep the mesh topology unchanged. This, unfortunately, can severely limit the flexibility of our optimization of object geometry, making the result highly sensitive to the quality of the initial mesh. Additionally, without taking precautions, updating vertex positions can introduce artifacts (e.g., self intersections) to the mesh that cannot be easily fixed by later iterations.

To address these problems, we utilize a few extra steps.

Coarse-to-fine optimization. Instead of performing the entire optimization at a single resolution, we utilize a coarse-to-fine process for improved robustness. Similar steps have been taken in several prior works, although typically limited to either geometry [SLS*06,SY10,KH13,TSG19] or reflectance [DCP*14,RPG16, HSL*17].

Specifically, we start the optimization by using low-resolution meshes and reflectance maps. If the input already has high resolutions, we simplify them via remeshing and image downsampling. Our optimization process then involves multiple stages. During each stage, we iteratively refine the object geometry and reflectance with fixed mesh topology. After each stage (except the final one), we upsample the mesh (using instant meshes [JTPSH15]) and the texture maps (using simple bilinear interpolation).

Robust surface evolution. During optimization, if the vertex positions are updated naïvely (i.e., using simple gradient-based up-

Table 1: ***Rendering performance***. *We report the rendering cost in seconds (averaged across 100 times) of each differentiable renderer in resolution* 512×512 *and* 4 *samples per pixel on a Titan RTX graphics card.*

	SoftRas	PyTorch3D	Mitsuba 2	Nvdiffrast	Redner	Ours
Kettle	0.0184	0.0202	0.0877	0.0013	0.1196	0.0143
Maneki	0.0192	0.0224	0.0863	0.0010	0.1029	0.0146
Pig	0.0971	0.0772	0.0913	0.0014	0.1336	0.0263
Kitty	0.0249	0.0334	0.0889	0.0010	0.1190	0.0225

dates with no validation checks), the mesh can have degraded quality and even become non-manifold (i.e., with artifacts like holes and self-intersections). Motivated by other optimization-driven mesh editing algorithms [SVJ15, LTJ18], we evolve a mesh using a pipeline implemented in the El Topo library [B*09]: Given the initial positions of a set of vertices with associated displacements, El Topo moves each vertex along its displacement vector while ensuring no self-intersection is generated.

4. Results

We implement our differentiable renderer (§3.1) in C++ with CUDA 11 and OptiX 7.1. For vectorized and differentiable computations, we utilize the Enoki library [Jak19], which has been demonstrated to be more efficient than generic ones like Tensorflow and PyTorch for rendering [NDVZJ19].

We implement the rest of our optimization pipeline, including the loss computations, in PyTorch. We use one set of weights for our optimizations: $\lambda_{\text{rend}} = 1$ for the rendering loss of Eq. (7); $(\lambda_{\text{lap}}, \lambda_{\text{edge}}, \lambda_{\text{normal}}) = (0.1, 1, 0.01)$ for the mesh loss of Eq. (9) and $(\lambda_{\text{spec}}, \lambda_{\text{roug}}) = (0.01, 0.001)$ for the material loss of Eq. (10).

In practice, our optimization involves 500–1000 iterations (for all coarse-to-fine stages) and takes 0.5–2 hours per example (see the supplement for more details).

4.1. Evaluations and comparisons

Please see the supplemental material for more results.

Comparison with differentiable renderers. Our renderer enjoys high performance, thanks to its specialized nature (i.e., focused on the collocated configuration) and the combined efficiency of RTX ray tracing offered by OptiX and GPU-based differentiable computations by Enoki. As demonstrated in Table 1, our renderer is faster than SoftRas [LLCL19], PyTorch3D [RRN*20], and Mitsuba 2 [NDVZJ19] without the need to introduce bias to the gradient estimates. Nvdiffrast [LHK*20] is faster than our system but produces approximated gradients. Lastly, compared to Redner [LADL18], another differentiable renderer that produces unbiased gradients, our renderer offers better performance.

To further demonstrate the importance for having accurate gradients, we conduct a synthetic experiment where the shape of an object is optimized (with known diffuse reflectance). Using identical input images, initial configurations, losses, and optimization settings (e.g., learning rate), we ran multiple optimizations using Adam [KB14] with gradients produced by SoftRas, PyTorch3D, Mitsuba 2, Nvdiffrast, and our technique, respectively. As shown in Figure 3, using biased gradients yields various artifacts or blurry geometries in the optimized results.

Effectiveness of shape optimization. To ensure robustness when optimizing the shape of an object, our technique utilizes a mesh loss (§3.3) as well as a coarse-to-fine framework (§3.4). We conduct another experiment to evaluate the effectiveness of these steps. Specifically, we optimize the shape of the *pig* model using identical optimization configurations except for (i) having various components of the mesh loss turned off; and (ii) not using the coarse-to-fine framework. As shown in Figure 4, with the mesh Laplacian loss $\mathcal{L}_{\text{lap}}$ neglected (by setting $\lambda_{\text{lap}} = 0$), the resulting geometry becomes "bumpy"; without the normal and edge-length losses $\mathcal{L}_{\text{normal}}$ and $\mathcal{L}_{\text{edge}}$, the optimized geometry also has artifacts due to sharp normal changes and ill-shaped triangles. Additionally, without performing the optimization in a coarse-to-fine fashion (by directly starting with a subdivided version of the initial mesh), the optimization gets stuck in a local optimum (with all losses enabled).

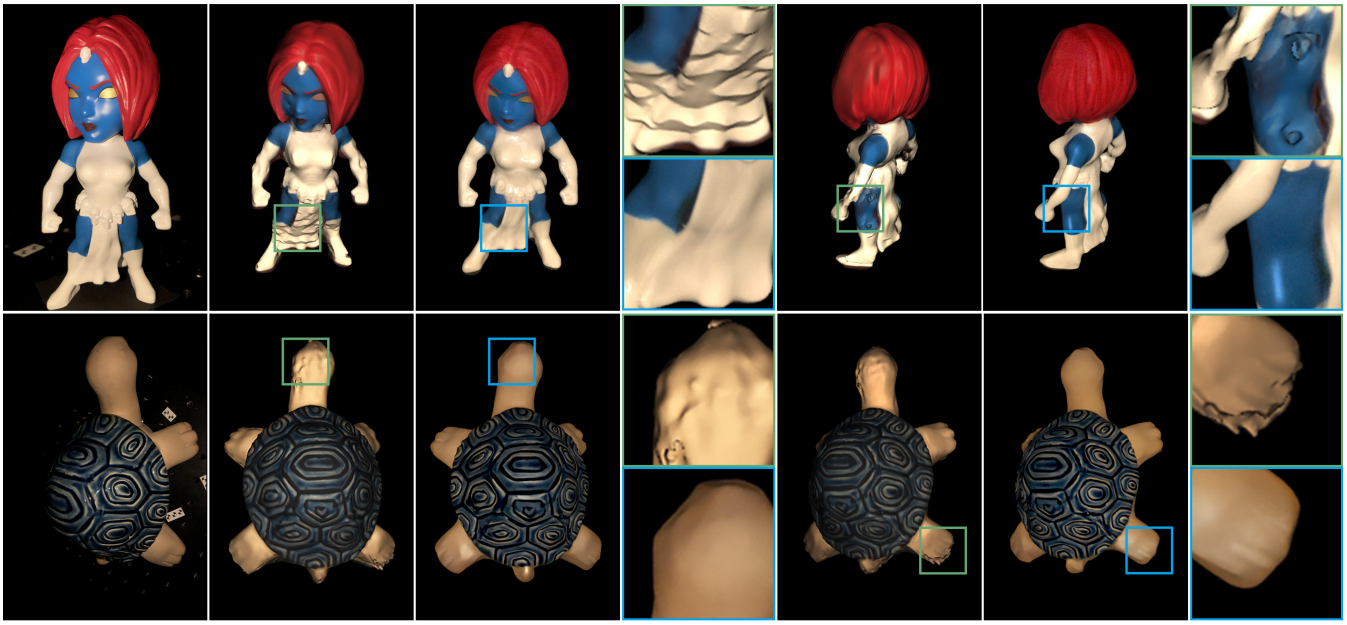
Effectiveness of material loss. Our material loss of Eq. (10) is important for obtaining clean reflectance maps that generalize well to novel settings. As shown in Figure 5, without correlating diffuse and specular albedos (by having $\lambda_{\text{spec}} = 0$), diffuse colors are "baked" into specular albedo, leading to heavy artifacts under novel environmental illumination. With the roughness smoothness $\mathcal{L}_{\text{roug}}$ disabled, the resulting roughness map contains high-frequency noise that leads to artifacts in rendered specular highlights.

Number of input images. We evaluate how the number of input images affects the reconstruction quality of our method in Figure 6. Using a small number (e.g., 10) of images, the optimization becomes highly under-constrained, making it difficult for our model to produce accurate appearance under novel viewing conditions. The accuracy our novel-view renderings improves quickly as the number of input images increases: With 50 or more input images, our renderings closely match the groundtruth.

Comparison with previous methods. To further evaluate the effectiveness of our technique for recovering object geometry, we compare with several previous methods [SZPF16, NIH*11, NLGK18].

Figure 7 shows comparisons with COLMAP [SZPF16] and Kinect Fusion [NIH*11] using synthetic inputs. Our technique, when using crude initial geometries (obtained using Kinect Fusion with low-resolution and noisy depth images), produces results with much higher quality than the baselines. COLMAP fails badly for the *kitty* example since the object contains insufficient textures for correspondences to be reliable established.

Additionally, we compare our technique with the work from Nam et al. [NLGK18] using real inputs (i.e., photographs). As demonstrated in Figure 8, Nam et al.'s method does not explicitly optimizes object geometry based on image losses and returns geometries with heavy artifacts. Our technique, on the other hand, is



(a) Photo (b) Nam et al. (c) Ours (d) Insets (e) Nam et al. (f) Ours (g) Insets

Figure 8: ***Comparison*** *with Nam et al. [NLGK18]: We render the reconstructed object under novel view. Nam et al.'s method produces bumpy geometry and inaccurate highlights. In contrast, our method produces much cleaner results that closely resemble the real object.*

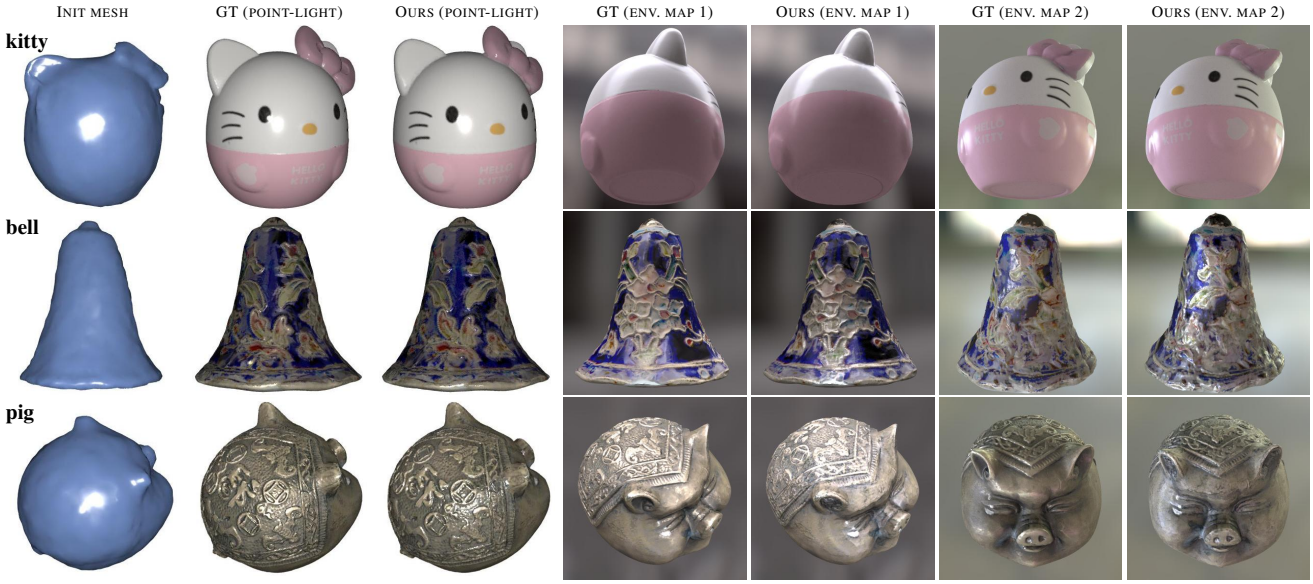


Figure 9: ***Reconstruction results*** *using synthetic inputs: We obtain the initial meshes (in the left column) using Kinect Fusion [NIH* 11] with low-resolution (48 × 48) and noisy depths. Our analysis-by-synthesis pipeline successfully recovers both geometric and reflectance details, producing high-fidelity results under novel viewing and illumination conditions.*

much more robust and capable of reproducing the clean geometry and appearance of the physical model.

4.2. Reconstruction results

Figure 9 shows reconstruction results obtained using synthetic input images and rendered under novel views and illuminations. The initial geometries are obtained using Kinect Fusion with low-resolution noisy depth inputs. Using 50 input images, our tech-

nique offers the robustness for recovering both smooth (e.g., the *kitty* example) and detailed (e.g., the *pig* example) geometries and reflectance.

We show in Figure 10 reconstruction results using as input 100 real photographs per example. The initial geometries are obtained using COLMAP. Our analysis-by-synthesis technique manages to accurately recover the detailed geometry and reflectance of each model.

Please note, in Figures 9 and 10, the detailed geometric structures (e.g., those in the *bell*, *pig*, *chime*, and *buddha* examples) fully emerge from the mesh-based object geometries: no normal or displacement mapping is used.

Lastly, since our reconstructed models use standard mesh-based representations, they can be used in a broad range of applications (see Figure 11).

4.3. Discussion and Analysis

We believe the quality gain to be obtained for three main reasons: First, we use Monte Carlo edge sampling [LADL18] that provides accurate gradients with respect to vertex positions, allowing our method to provide more accurate reconstructions of object geometries (cf. Figure 8 against [NLGK18]); Second, we exploit robust surface evolution, e.g., elTopo, on top of gradient descent, which ensures a manifold mesh (i.e., without self-intersections or other degenerated cases) after every iteration; Third, our coarse-to-fine strategy and the other regularization terms have come together to make our pipeline more robust in practice.

Failure cases. Despite our pipeline being robust for most cases in synthetic/real-world settings, failure cases still exist. Firstly, our method has difficulties handling complex changes of the mesh topology—which is a well-known limitation for mesh-based representations. Secondly, when modeling object appearance, our method relies on a simplified version of the Disney BRDF model only dealing with opaque materials, and thus is limited at reconstructing sophisticated surface appearances, such as anisotropic reflection or subsurface scattering.

5. Conclusion

We introduce a new approach to jointly recover the shape and reflectance of real-world objects. At the core of our technique is a unified analysis-by-synthesis pipeline that iteratively refines object geometry and reflectance. Our custom Monte Carlo differentiable renderer enjoys higher performance than many existing tools (such as SoftRas, PyTorch3D, and Mitsuba 2). More importantly, our renderer produces unbiased geometric gradients that are crucial for obtaining high-quality reconstructions. To further improve the robustness of our optimization, we leverage a coarse-to-fine framework regularized using a few geometric and reflectance priors. We conduct several ablation studies to evaluate the effectiveness of our differentiable renderer, losses, and optimization strategies.

Limitations and future work. Our technique is specialized to using a collocated camera and point light. This configuration can have difficulties in capturing materials exhibiting strong retroreflection. Generalization to other configurations would be useful in the future.

To refine mesh topology, our technique relies on remeshing steps (between coarse-to-fine stages). How topology can be optimized in a robust and flexible fashion is an important problem for future investigation.

Lastly, more advanced regularizations of geometry and/or appearance may enable high-quality reconstructions with fewer input images.

Acknowledgements. We thank Chenglei Wu, Yujia Chen, Christoph lassner, Sai Bi, Zhengqin Li, Giljoo Nam, Yue Dong, Hongzhi Wu, Zhongshi Jiang as well as the anonymous reviewers for their valuable discussions. We thank the digital artist James Warren from Facebook Reality Labs for modeling and rendering the two table scenes, and Inseung Hwang from KAIST for making comparisons with Nam et al. [NLGK18]. This work was supported in part by NSF grants 1900783 and 1900927.

References

[AAL16] AITTALA M., AILA T., LEHTINEN J.: Reflectance modeling by neural texture synthesis. *ACM Trans. Graph. 35*, 4 (2016), 1–13. 2, 3

[ACGO18] ALBERT R. A., CHAN D. Y., GOLDMAN D. B., O'BRIEN J. F.: Approximate svBRDF estimation from mobile phone video. In *Proc. EGSR: Experimental Ideas & Implementations* (2018), Eurographics Association, pp. 11–22. 3

[ALKN19] AZINOVIC D., LI T.-M., KAPLANYAN A., NIESSNER M.: Inverse path tracing for joint material and lighting estimation. In *Proc. IEEE/CVF CVPR* (2019), pp. 2447–2456. 2

[AWL*15] AITTALA M., WEYRICH T., LEHTINEN J., ET AL.: Two-shot SVBRDF capture for stationary materials. *ACM Trans. Graph. 34*, 4 (2015), 110–1. 1, 2, 3

[B*09] BROCHU T., ET AL.: El Topo: Robust topological operations for dynamic explicit surfaces, 2009. https://github.com/tysonbrochu/eltopo. 8

[BXS*20] BI S., XU Z., SUNKAVALLI K., KRIEGMAN D., RAMAMOORTHI R.: Deep 3D capture: Geometry and reflectance from sparse multi-view images. In *Proc. IEEE/CVF CVPR* (2020), pp. 5960–5969. 2, 3

[CDP*14] CHEN G., DONG Y., PEERS P., ZHANG J., TONG X.: Reflectance scanning: estimating shading frame and BRDF with generalized linear light sources. *ACM Trans. Graph. 33*, 4 (2014), 1–11. 2

[CLZ*20] CHE C., LUAN F., ZHAO S., BALA K., GKIOULEKAS I.: Towards learning-based inverse subsurface scattering. *ICCP* (2020), 1–12. 2

[DAD*19] DESCHAINTRE V., AITTALA M., DURAND F., DRETTAKIS G., BOUSSEAU A.: Flexible SVBRDF capture with a multi-image deep network. In *Computer Graphics Forum* (2019), vol. 38, Wiley Online Library, pp. 1–13. 2, 3

[DCP*14] DONG Y., CHEN G., PEERS P., ZHANG J., TONG X.: Appearance-from-motion: Recovering spatially varying surface reflectance under unknown lighting. *ACM Trans. Graph. 33*, 6 (2014), 1–12. 2, 7

[DWMG15] DONG Z., WALTER B., MARSCHNER S., GREENBERG D. P.: Predicting appearance from measured microgeometry of metal surfaces. *ACM Trans. Graph. 35*, 1 (2015), 1–13. 2

[DWT*10] DONG Y., WANG J., TONG X., SNYDER J., LAN Y., BEN-EZRA M., GUO B.: Manifold bootstrapping for SVBRDF capture. *ACM Trans. Graph. 29*, 4 (2010), 1–10. 2

Figure 10: ***Reconstruction results*** *of real-world objects: Similar to the synthetic examples in Figure 9, our technique recovers detailed object geometries and reflectance details, producing high-quality results under novel viewing and illumination conditions.*

Figure 11: ***Applications:*** *The high-quality models generated by our technique can be used for 3D digital modeling (top) and object insertion in augmented reality (bottom).*

[FP09] FURUKAWA Y., PONCE J.: Accurate, dense, and robust multiview stereopsis. *IEEE Trans. Pattern Anal. Mach. Intell. 32*, 8 (2009), 1362–1376. 2

[GHP*08] GHOSH A., HAWKINS T., PEERS P., FREDERIKSEN S., DEBEVEC P.: Practical modeling and acquisition of layered facial reflectance. *ACM Trans. Graph. 27*, 5 (Dec. 2008). 1

[GLD*19] GAO D., LI X., DONG Y., PEERS P., XU K., TONG X.: Deep inverse rendering for high-resolution SVBRDF estimation from an arbitrary number of images. *ACM Trans. Graph. 38*, 4 (2019), 134–1. 2, 3, 7

[GLZ16] GKIOULEKAS I., LEVIN A., ZICKLER T.: An evaluation of computational imaging techniques for heterogeneous inverse scattering. In *ECCV* (2016), Springer, pp. 685–701. 2

[GSH*20] GUO Y., SMITH C., HAŠAN M., SUNKAVALLI K., ZHAO S.: MaterialGAN: Reflectance capture using a generative SVBRDF model. *ACM Trans. Graph. 39*, 6 (2020), 254:1–254:13. 2

[GZB*13] GKIOULEKAS I., ZHAO S., BALA K., ZICKLER T., LEVIN A.: Inverse volume rendering with material dictionaries. *ACM Trans. Graph. 32*, 6 (2013), 1–13. 2

[HLHZ08] HOLROYD M., LAWRENCE J., HUMPHREYS G., ZICKLER T.: A photometric approach for estimating normals and tangents. *ACM Trans. Graph. 27*, 5 (2008), 1–9. 2

[HLZ10] HOLROYD M., LAWRENCE J., ZICKLER T.: A coaxial optical scanner for synchronous acquisition of 3D geometry and surface reflectance. *ACM Trans. Graph. 29*, 4 (2010), 1–12. 2

[HMJI09] HIGO T., MATSUSHITA Y., JOSHI N., IKEUCHI K.: A hand-held photometric stereo camera for 3-d modeling. In *Proc. ICCV* (2009), IEEE, pp. 1234–1241. 2

[Hor70] HORN B. K.: Shape from shading: A method for obtaining the shape of a smooth opaque object from one view. 2

[HQMC18] HAEFNER B., QUÉAU Y., MÖLLENHOFF T., CREMERS D.: Fight ill-posedness with ill-posedness: Single-shot variational depth super-resolution from shading. In *Proc. IEEE/CVF CVPR* (2018), pp. 164–174. 2

[HSL*17] HUI Z., SUNKAVALLI K., LEE J.-Y., HADAP S., WANG J., SANKARANARAYANAN A. C.: Reflectance capture using univariate sampling of BRDFs. In *ICCV* (2017), IEEE, pp. 5362–5370. 1, 2, 3, 7

[IH81] IKEUCHI K., HORN B. K.: Numerical shape from shading and occluding boundaries. *Artificial intelligence 17*, 1-3 (1981), 141–184. 2

[IKH*11] IZADI S., KIM D., HILLIGES O., MOLYNEAUX D., NEWCOMBE R., KOHLI P., SHOTTON J., HODGES S., FREEMAN D., DAVISON A., FITZGIBBON A.: Kinectfusion: Real-time 3D reconstruction and interaction using a moving depth camera. In *Proc. ACM UIST* (2011), pp. 559–568. 1

[Jak19] JAKOB W.: Enoki: structured vectorization and differentiation on modern processor architectures, 2019. https://github.com/mitsuba-renderer/enoki. 8

[JDV*14] JENSEN R., DAHL A., VOGIATZIS G., TOLA E., AANÆS H.: Large scale multi-view stereopsis evaluation. In *Proc. IEEE CVPR* (2014), pp. 406–413. 4

[JJHZ20] JIANG Y., JI D., HAN Z., ZWICKER M.: Sdfdiff: Differentiable rendering of signed distance fields for 3D shape optimization. In *Proc. IEEE/CVF CVPR* (2020), pp. 1251–1261. 3

[JTPSH15] JAKOB W., TARINI M., PANOZZO D., SORKINE-HORNUNG O.: Instant field-aligned meshes. *ACM Trans. Graph. 34*, 6 (2015). 7

[KB14] KINGMA D. P., BA J.: Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980* (2014). 6, 8

[KBM*20] KATO H., BEKER D., MORARIU M., ANDO T., MATSUOKA T., KEHL W., GAIDON A.: Differentiable rendering: A survey, 2020. arXiv:2006.12057. 2

[KCW*18] KANG K., CHEN Z., WANG J., ZHOU K., WU H.: Efficient reflectance capture using an autoencoder. *ACM Trans. Graph. 37*, 4 (2018), 127–1. 2

[KG13] KARIS B., GAMES E.: Real shading in unreal engine 4. *Proc. Physically Based Shading Theory Practice 4* (2013), 3. 3

[KGT*17] KIM K., GU J., TYREE S., MOLCHANOV P., NIESSNER M., KAUTZ J.: A lightweight approach for on-the-fly reflectance estimation. In *ICCV* (2017), IEEE, pp. 20–28. 2

[KH13] KAZHDAN M., HOPPE H.: Screened poisson surface reconstruction. *ACM Trans. Graph. 32*, 3 (2013), 1–13. 7

[LADL18] LI T.-M., AITTALA M., DURAND F., LEHTINEN J.: Differentiable Monte Carlo ray tracing through edge sampling. *ACM Trans. Graph. 37*, 6 (2018), 1–11. 1, 2, 3, 4, 8, 10

[Las20] LASSNER C.: Fast differentiable raycasting for neural rendering using sphere-based representations. *arXiv preprint arXiv:2004.07484* (2020). 3

[LHJ19] LOUBET G., HOLZSCHUCH N., JAKOB W.: Reparameterizing discontinuous integrands for differentiable rendering. *ACM Trans. Graph. 38*, 6 (2019), 1–14. 2, 4

[LHK*20] LAINE S., HELLSTEN J., KARRAS T., SEOL Y., LEHTINEN J., AILA T.: Modular primitives for high-performance differentiable rendering. *ACM Trans. Graph. 39*, 6 (2020). 4, 8

[LKG*03] LENSCH H. P., KAUTZ J., GOESELE M., HEIDRICH W., SEIDEL H.-P.: Image-based reconstruction of spatial appearance and geometric detail. *ACM Trans. Graph. 22*, 2 (2003), 234–257. 2

[LLCL19] LIU S., LI T., CHEN W., LI H.: Soft rasterizer: A differentiable renderer for image-based 3D reasoning. In *ICCV* (2019), IEEE, pp. 7708–7717. 2, 4, 8

[LSC18] LI Z., SUNKAVALLI K., CHANDRAKER M.: Materials for masses: SVBRDF acquisition with a single mobile phone image. In *ECCV* (2018), Springer, pp. 72–87. 3

[LSR*20] LI Z., SHAFIEI M., RAMAMOORTHI R., SUNKAVALLI K., CHANDRAKER M.: Inverse rendering for complex indoor scenes: Shape, spatially-varying lighting and SVBRDF from a single image. In *Proc. IEEE/CVF CVPR* (2020), pp. 2475–2484. 3

[LTJ18] LIU H.-T. D., TAO M., JACOBSON A.: Paparazzi: surface editing by way of multi-view image processing. *ACM Trans. Graph. 37*, 6 (2018), 221–1. 8

[Mat03] MATUSIK W.: *A data-driven reflectance model*. PhD thesis, Massachusetts Institute of Technology, 2003. 2

[MKC*17] MAIER R., KIM K., CREMERS D., KAUTZ J., NIESSNER M.: Intrinsic3d: High-quality 3D reconstruction by joint appearance and geometry optimization with spatially-varying lighting. In *ICCV* (2017), IEEE, pp. 3114–3122. 2

[MON*19] MESCHEDER L., OECHSLE M., NIEMEYER M., NOWOZIN S., GEIGER A.: Occupancy networks: Learning 3D reconstruction in function space. In *Proc. IEEE/CVF CVPR* (2019), pp. 4460–4470. 3

[NDVZJ19] NIMIER-DAVID M., VICINI D., ZELTNER T., JAKOB W.: Mitsuba 2: A retargetable forward and inverse renderer. *ACM Trans. Graph. 38*, 6 (2019), 1–17. 2, 4, 8

[NIH*11] NEWCOMBE R. A., IZADI S., HILLIGES O., MOLYNEAUX D., KIM D., DAVISON A. J., KOHI P., SHOTTON J., HODGES S., FITZGIBBON A.: Kinectfusion: Real-time dense surface mapping and tracking. In *ISMAR* (2011), IEEE, pp. 127–136. 1, 7, 8, 9

[NISA06] NEALEN A., IGARASHI T., SORKINE O., ALEXA M.: Laplacian mesh optimization. In *Proc. the 4th international conference on Computer graphics and interactive techniques in Australasia and Southeast Asia* (2006), pp. 381–389. 7

[NLGK18] NAM G., LEE J. H., GUTIERREZ D., KIM M. H.: Practical SVBRDF acquisition of 3D objects with unstructured flash photography. *ACM Trans. Graph. 37*, 6 (2018), 1–12. 2, 8, 9, 10

[NLW*16] NAM G., LEE J. H., WU H., GUTIERREZ D., KIM M. H.: Simultaneous acquisition of microscale reflectance and normals. *ACM Trans. Graph. 35*, 6 (2016), 185–1. 1

[PF14] PAPADHIMITRI T., FAVARO P.: Uncalibrated near-light photometric stereo. 2

[PFS*19] PARK J. J., FLORENCE P., STRAUB J., NEWCOMBE R., LOVEGROVE S.: Deepsdf: Learning continuous signed distance functions for shape representation. In *Proc. IEEE/CVF CVPR* (2019), pp. 165–174. 3

[PGC*17] PASZKE A., GROSS S., CHINTALA S., CHANAN G., YANG E., DEVITO Z., LIN Z., DESMAISON A., ANTIGA L., LERER A.: Automatic differentiation in pytorch. 3

[PNS18] PARK J. J., NEWCOMBE R., SEITZ S.: Surface light field fusion. In *3DV* (2018), IEEE, pp. 12–21. 2

[QLD15] QUÉAU Y., LAUZE F., DUROU J.-D.: Solving uncalibrated photometric stereo using total variation. *Journal of Mathematical Imaging and Vision 52*, 1 (2015), 87–107. 2

[QMC*17] QUÉAU Y., MÉLOU J., CASTAN F., CREMERS D., DUROU J.-D.: A variational approach to shape-from-shading under natural illumination. In *International Workshop on Energy Minimization Methods in Computer Vision and Pattern Recognition* (2017), Springer, pp. 342–357. 2

[QMD16] QUÉAU Y., MECCA R., DUROU J.-D.: Unbiased photometric stereo for colored surfaces: A variational approach. In *Proc. IEEE CVPR* (2016), pp. 4359–4368. 2

[QMDC17] QUÉAU Y., MÉLOU J., DUROU J.-D., CREMERS D.: Dense multi-view 3D-reconstruction without dense correspondences. *arXiv preprint arXiv:1704.00337* (2017). 2

[Rey03] REYNOLDS O.: *Papers on mechanical and physical subjects: the sub-mechanics of the universe*, vol. 3. The University Press, 1903. 4

[RPG16] RIVIERE J., PEERS P., GHOSH A.: Mobile surface reflectometry. In *Computer Graphics Forum* (2016), vol. 35, Wiley Online Library, pp. 191–202. 1, 3, 7

[RRFG17] RIVIERE J., RESHETOUSKI I., FILIPI L., GHOSH A.: Polarization imaging reflectometry in the wild. *ACM Trans. Graph. 36*, 6 (2017), 1–14. 1

[RRN*20] RAVI N., REIZENSTEIN J., NOVOTNY D., GORDON T., LO W.-Y., JOHNSON J., GKIOXARI G.: Accelerating 3D deep learning with PyTorch3D. *arXiv preprint arXiv:2007.08501* (2020). 1, 2, 4, 8

[RWS*11] REN P., WANG J., SNYDER J., TONG X., GUO B.: Pocket reflectometry. *ACM Trans. Graph. 30*, 4 (2011), 1–10. 3

[SD99] SEITZ S. M., DYER C. R.: Photorealistic scene reconstruction by voxel coloring. *International Journal of Computer Vision 35*, 2 (1999), 151–173. 2

[SDR*20] SCHMITT C., DONNE S., RIEGLER G., KOLTUN V., GEIGER A.: On joint estimation of pose, geometry and svBRDF from a handheld scanner. In *Proc. IEEE/CVF CVPR* (2020), pp. 3493–3503. 2, 7

[SLS*06] SHARF A., LEWINER T., SHAMIR A., KOBBELT L., COHEN-OR D.: Competing fronts for coarse–to–fine surface reconstruction. In *Computer Graphics Forum* (2006), vol. 25, Wiley Online Library, pp. 389–398. 7

[SSWK13] SCHWARTZ C., SARLETTE R., WEINMANN M., KLEIN R.: Dome ii: A parallelized btf acquisition system. In *Material Appearance Modeling* (2013), pp. 25–31. 1

[SVJ15] SACHT L., VOUGA E., JACOBSON A.: Nested cages. *ACM Trans. Graph. 34*, 6 (2015), 1–14. 8

[SY10] SAHILLIOĞLU Y., YEMEZ Y.: Coarse-to-fine surface reconstruction from silhouettes and range data using mesh deformation. *Computer Vision and Image Understanding 114*, 3 (2010), 334–348. 7

[SZPF16] SCHÖNBERGER J. L., ZHENG E., POLLEFEYS M., FRAHM J.-M.: Pixelwise view selection for unstructured multi-view stereo. In *ECCV* (2016), Springer. 2, 7, 8

[TFG*13] TUNWATTANAPONG B., FYFFE G., GRAHAM P., BUSCH J., YU X., GHOSH A., DEBEVEC P.: Acquiring reflectance and shape from continuous spherical harmonic illumination. *ACM Trans. Graph. 32*, 4 (2013), 1–12. 1, 2

[TSG19] TSAI C.-Y., SANKARANARAYANAN A. C., GKIOULEKAS I.: Beyond volumetric albedo–a surface optimization framework for non-line-of-sight imaging. In *Proc. IEEE/CVF CVPR* (2019), pp. 1545–1555. 2, 7

[VTC05] VOGIATZIS G., TORR P. H., CIPOLLA R.: Multi-view stereo via volumetric graph-cuts. In *Proc. IEEE CVPR* (2005), vol. 2, pp. 391–398. 2

[Woo80] WOODHAM R. J.: Photometric method for determining surface orientation from multiple images. *Optical engineering 19*, 1 (1980), 191139. 2

[WWZ15] WU H., WANG Z., ZHOU K.: Simultaneous localization and appearance estimation with a consumer rgb-d camera. *IEEE TVCG 22*, 8 (2015), 2012–2023. 2

[XDPT16] XIA R., DONG Y., PEERS P., TONG X.: Recovering shape and spatially-varying surface reflectance under unknown illumination. *ACM Trans. Graph. 35*, 6 (2016), 1–12. 2

[YDMH99] YU Y., DEBEVEC P., MALIK J., HAWKINS T.: Inverse global illumination: Recovering reflectance models of real scenes from photographs. In *Proc. Computer graphics and interactive techniques* (1999), pp. 215–224. 2

[ZCD*16] ZHOU Z., CHEN G., DONG Y., WIPF D., YU Y., SNYDER J., TONG X.: Sparse-as-possible SVBRDF acquisition. *ACM Trans. Graph. 35*, 6 (2016), 1–12. 2

[ZLW*21] ZHANG K., LUAN F., WANG Q., BALA K., SNAVELY N.: Physg: Inverse rendering with spherical gaussians for physics-based material editing and relighting. In *Proc. IEEE CVPR* (2021). 3

[ZT10] ZHOU Z., TAN P.: Ring-light photometric stereo. In *ECCV* (2010), Springer, pp. 265–279. 2

[ZWZ*19] ZHANG C., WU L., ZHENG C., GKIOULEKAS I., RAMAMOORTHI R., ZHAO S.: A differential theory of radiative transfer. *ACM Trans. Graph. 38*, 6 (2019), 227:1–227:16. 3, 4