Practical SVBRDF Acquisition of 3D Objects with Unstructured Flash Photography

GILJOO NAM and JOO HO LEE, KAIST
DIEGO GUTIERREZ, Universidad de Zaragoza, I3A
MIN H. KIM, KAIST

Fig. 1. We propose a practical method that enables us to capture spatially-varying BRDFs from *unstructured flash photographs*. It yields high-quality SVBRDFs, as well as detailed geometry of 3D objects, without relying on any expensive supporting hardware nor controlled illumination, using instead any hand-held conventional digital camera with a built-in flash. No input geometry is needed in our algorithm. Image (a) shows our acquisition setup, while (b) – (e) show example results of our reconstructions.

Capturing spatially-varying bidirectional reflectance distribution functions (SVBRDFs) of 3D objects with just a single, hand-held camera (such as an off-the-shelf smartphone or a DSLR camera) is a difficult, open problem. Previous works are either limited to planar geometry, or rely on previously scanned 3D geometry, thus limiting their practicality. There are several technical challenges that need to be overcome: First, the built-in flash of a camera is almost colocated with the lens, and at a fixed position; this severely hampers sampling procedures in the light-view space. Moreover, the near-field flash lights the object partially and unevenly. In terms of geometry, existing multiview stereo techniques assume diffuse reflectance only, which leads to overly smoothed 3D reconstructions, as we show in this paper. We present a simple yet powerful framework that removes the need for expensive, dedicated hardware, enabling practical acquisition of SVBRDF information from real-world, 3D objects with a single, off-the-shelf camera with a built-in flash. In addition, by removing the diffuse reflection assumption and leveraging instead such SVBRDF information, our method outputs high-quality 3D geometry reconstructions, including more accurate high-frequency details than state-of-the-art multiview stereo techniques. We formulate the *joint* reconstruction of SVBRDFs, shading normals, and 3D geometry as a multi-stage, iterative inverse-rendering reconstruction pipeline. Our method is also directly applicable to any existing multiview 3D reconstruction technique. We present results of captured objects with complex geometry and reflectance; we also validate our method numerically against other existing approaches that rely on dedicated hardware, additional sources of information, or both.

Authors' addresses: Giljoo Nam; Joo Ho Lee, KAIST, School of Computing, Daejeon, South Korea, 34141; Diego Gutierrez, Universidad de Zaragoza, I3A, Zaragoza, Spain, 50018; Min H. Kim, KAIST, School of Computing, Daejeon, South Korea, 34141, minhkim@kaist.ac.kr(corresponding-author).

1 INTRODUCTION

Acquiring and reproducing the appearance of real-world objects is one of the main goals of computer graphics. Many different advances have been presented recently, from methods relying on specialized hardware (e.g., [Ghosh et al. 2008; Holroyd et al. 2010; Nam et al. 2016; Schwartz et al. 2013; Tunwattanapong et al. 2013]), to mobile setups (e.g., [Aittala et al. 2015; Hui et al. 2017; Riviere et al. 2015, 2017]). However, there is an inevitable tradeoff between the capabilities of these methods and their cost. For instance, while mobile methods are cheap but limited to *near-planar* geometries, prices for professional acquisition systems, such as Otoy LightStage, X-Rite TAC7 and Lumio3D, start at $200,000.

Despite the obvious practical advantages of simpler acquisition setups, capturing the SVBRDFs of a full, nonplanar 3D object with an off-the-shelf camera, such as a smartphone camera, has not yet been demonstrated, due to its technical challenges. First, capturing SVBRDF information requires carefully controlled, dense sampling of the light-view space; this is usually achieved with professional supporting structures, such as light domes, or four-dimensional gantries. Since the built-in flash of a conventional camera is almost colocated with the lens, this leads to a severely limited sampling of

the light-view space; moreover, the near-field flash light illuminates the object unevenly. Second, since the vast majority of conventional cameras lack a depth sensor, recovering geometry must rely on passive multiview stereo techniques that assume diffuse reflectance only; however, in the presence of spatially-varying bidirectional reflectance, this leads to inaccurate reconstructions. Third, due to the unstructured nature of hand-held acquisition, pixel correspondences between 3D points in the object and 2D pixels in multiple images are not guaranteed. These correspondences are guaranteed in a photometric stereo setup, where a fixed camera and varying light sources are used, or when the cameras and lights are locked in a physical structure, providing structured input. In a hand-held approach, both the intrinsic/extrinsic parameters as well as the 3D geometry of the object should be given, which is information that we do not have. The resulting misalignments are the main hindrance of high-quality reconstruction of 3D geometry and SVBRDF with an unstructured capture setup.

In this work, we present a compact, practical capture method using multiple unstructured flash photographs as input (see Figure 1(a)). It requires just a single conventional camera (including smartphone cameras) with a built-in flash. Key to our method is a novel *joint* reconstruction of SVBRDFs, shading normals, and 3D geometry. Such joint reconstruction is performed on a multi-stage, iterative and alternating optimization inverse-rendering pipeline, which progressively improves 3D-to-2D correspondences, leading to high-quality reconstruction of both SVBRDFs and 3D geometry.

Our work significantly advances the state of art of appearance acquisition setups. It allows the practical acquisition of SVBRDF information of 3D objects using a single camera, not limited to planar surfaces, and not requiring a commercial 3D scanner to accurately capture input geometry. In addition, our method yields significantly better 3D geometry reconstructions than state-of-the-art multiview stereo techniques, since we take SVBRDF information into account in the reconstruction instead of assuming diffuse surfaces; as our results show, we acquire well-defined, high-frequency details that get lost to over-smoothing in existing methods. Last, we validate our method including side-by-side photographs of real objects, novel renderings under different illuminations, objective error against measured BRDFs, and objective comparisons against state-of-the-art methods that use more sophisticated hardware.

2 RELATED WORK

Most acquisition methods using a conventional camera focus *either* on geometry *or* reflectance capture exclusively, while simultaneous acquisition of both usually requires specialized hardware (such as a mechanized gantry, a light stage, or a commercial 3D scanner). Previous works can be classified as follows: (a) Reflectance capture from known 3D geometry; (b) Reflectance capture limited to 2D planar geometry; (c) 3D reconstruction assuming diffuse reflection only; and (d) Simultaneous acquisition of reflectance and 3D geometry; Our work falls in this last category, removing the need for any specialized hardware. For a more in-depth discussion on acquisition, we refer the reader to recent review works on the topic [Guarnera et al. 2016; Weinmann and Klein 2015; Weyrich et al. 2009].

Reflectance from known 3D geometry. Lensch et al. [2001; 2003] introduced a pioneering method to capture SVBRDFs of known 3D objects, comprising clustered basis reflectances. Per-texel reflectance is progressively refined with linear blending. Zhou et al. [2016] proposed an SVBRDF acquisition method that jointly optimizes reflectance bases and blending weights on known 3D geometry. This method aims at finding the smallest number of basis reflectances, then blending them smoothly. Different from our method, these techniques require a commercial 3D scanner to accurately capture the input 3D geometry.

Reflectance capture of 2D planar geometry. Conventional cameras such as a smartphone camera have also been used to capture reflectance information of *planar* surfaces. Using an LCD screen and a camera, Aittala et al. [2013] propose an efficient SVBRDF capture method restricting the range of angular reflectance samples, as well as a two-shot, flash/no-flash reflectance acquisition method for the particular case of stationary materials [Aittala et al. 2015], for which larger areas can be synthesized from small reconstructions. Riviere et al. [2015] and Hui et al. [2017] capture the appearance of near-flat objects using a smartphone camera from varying viewpoints. The light source provides active illumination, from which normals and reflectance are estimated. Ren et al. [2011] propose a portable system consisting of a smartphone camera, a hand-held linear light source, and a custom-built BRDF chart. They take a short video of a target object along with the BRDF chart while moving the hand-held light tube, and recover SVBRDFs from a linear combination of the reference BRDFs. Last, Thanikachalam et al. [2017] present an acquisition setup similar to Won et al.'s [2012], focusing on capturing reflectance of planar art paintings.

Other acquisition systems that capture high-quality SVBRDFs on planar surfaces rely on more sophisticated hardware. Nam et al. [2016] offer simultaneous acquisition of reflectance and normals at microscopic scale, placing the sample in a small dome with computer-controlled LED illumination. Other systems include four-axis spherical gantries to sample many different light-camera combinations [Lawrence et al. 2006], linear light source reflectometry [Chen et al. 2014; Gardner et al. 2003; Wang et al. 2008], or condenser lenses [Dong et al. 2010]. While these acquisition methods are also limited to near-flat objects, we estimate SVBRDFs and geometry for full 3D objects.

3D Reconstruction with diffuse assumption. Shading normals are often used to enhance geometric details, assuming *diffuse-only* reflectance for the whole object. Higo et al. [2009] and Park et al. [2016] first get a base geometry using structure-from-motion (SfM) and multiview stereo (MVS), then update the geometry using estimated surface normals assuming diffuse reflectance. Hernandez et al. [2008] use shape-from-silhouette instead for the first step. Zollhofer et al. [2015] run Kinect Fusion [Newcombe et al. 2011] to get a signed distance function of a surface. They further refine it using surface shading cues. Won et al. [2012] use two mobile devices, as a camera and light source respectively. They take multiple images from a fixed viewpoint under varying light directions, and reconstruct the surface from photometric stereo, similar to [Nam and Kim 2014]. Other recent methods have further demonstrated the use of smartphone cameras, to capture the 3D shape of objects [Kolev et al.

2014; Ondruska et al. 2015], or even large scenes [Kahler et al. 2015; Schöps et al. 2015], again based on both SfM and MVS techniques. All these methods assume that the surface reflectance of the object being reconstructed is diffuse only; as such, they cannot recover SVBRDF information.

Simultaneous acquisition of reflectance and 3D geometry. Previous works that capture reflectance and 3D geometry simultaneously have relied on specialized hardware setups, whose prices may be as high as several hundred thousand dollars.

Many light stage designs exist, relying on discrete spherical illumination with polarized light sources (e.g., [Ghosh et al. 2010, 2011, 2008; Graham et al. 2013; Nagano et al. 2015]). Tunwattanapong et al. [2013] built a similar structure with an LED arm that orbits rapidly to create a continuous spherical illumination with harmonics patterns. Holroyd et al. [2010] built a spherical gantry equipped with a projector-camera pair on two mechanical arms, using phase-shift patterns for 3D geometry. Other similar dome structures of multiple cameras have been presented [Schwartz et al. 2013], using structured light patterns for 3D geometry and representing reflectance as bidirectional texture functions (BTF).

There are more approachable methods that require less expensive hardware, such as a light probe, a multi-light structure, or an RGB-D camera. Zhou et al. [2013] built a multi-light device consisting of 72 LED lights on a circle board, which allows them to combine SfM and photometric stereo to get 3D geometry. They further estimate SVBRDF based on the 3D geometry, but do not use this information to refine the geometry or surface normals. Oxholm et al. [2014] utilize an environment map for estimating spherical illumination by capturing a light probe, then solve an inverse rendering problem. Xia et al. [2016] capture a video sequence of more than a thousand frames using a mechanical rotary stage. The method requires per-vertex dense sampling with at least two clear changes of illumination per vertex. Last, Wu et al. [2016; 2015] rely on depth information using Kinect Fusion [Newcombe et al. 2011] as well as spherical illumination or IR illumination from the depth camera. Recently, Baek et al. [2018] proposed an SVBRDF acquisition method that can capture polarimetric appearance of both diffuse and specular reflection with high-resolution normals, relying on input geometry from structured lighting.

These systems either rely on specialized hardware (which may be bulky, expensive to build, or hard to get), or alternatively make use of additional sources of information like multiple lights, spherical illumination, or depth. In addition, they usually require time-consuming calibration processes, while acquisition times are often in the order of several hours; our method is free from all these restrictions.

3 OVERVIEW

We provide an overview of our method in Figure 2. Our input consists of a set of K unstructured flash photographs $\mathbf{I} = \{I_k\}$ taken with two different exposures to extend the dynamic range. We convert I_k into linear radiance L_k by accounting for exposure variation (see Section 4). We then formalize our problem as obtaining SVBRDF information $\mathbf{F}$ (described as a set of basis BRDFs $\mathbf{F}_b$ and corresponding weight maps $\mathbf{W}$), shading normals $\mathbf{N}$, and 3D geometry $\mathbf{X}$. First, in our initialization step, we obtain a set of camera extrinsic parameters

and a *rough base geometry* using a conventional 3D reconstruction technique that includes SfM, MVS and mesh reconstruction (refer to Section 6 for the impact of the initial process). From this initial geometry, our method then simultaneously reconstructs SVBRDF information while improving the recovered 3D geometry, using an iterative process. It starts with an inverse rendering stage (Section 5), whose goal is to obtain a first approximation for $\mathbf{W}$, $\mathbf{F}_b$, and $\mathbf{N}$. We first reconstruct SVBRDF information ($\mathbf{F}_b$ and spatial weights $\mathbf{W}$, Section 5.1). We then estimate shading normals $\mathbf{N}$ relying on SVBRDF information (Section 5.2). After obtaining $\mathbf{W}$, $\mathbf{F}_b$, and $\mathbf{N}$, we update details of the 3D geometry $\mathbf{X}$ with variant input of shading normals $\mathbf{N}$ by means of Poisson surface reconstruction, which allows us to additionally force photometric consistency in the reconstruction (Section 6). We repeat this iterative optimization of inverse rendering and geometric reconstruction until the error converges.

We only need to estimate the position of the flash with respect to the camera, and the camera optical parameters. This is a process that needs to be performed only once; we use an image of multiple chrome balls and checkerboard images as described in Lensch's work [2003].
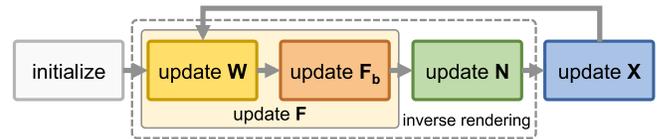

Fig. 2. Overview of our algorithm. We iterate estimations of reflectance $\mathbf{F}$ (defined as basis functions $\mathbf{F}_b$ with corresponding weight maps $\mathbf{W}$), shading normals $\mathbf{N}$, and 3D geometry $\mathbf{X}$. Section 5 describes the estimation of $\mathbf{F}$ and $\mathbf{N}$, while Section 6 describes the estimation of $\mathbf{X}$.

4 IMAGE FORMATION MODEL

We use an off-the-shelf conventional camera with a built-in flash to capture a set of unstructured flash photographs as input. Since the dynamic range of a conventional camera is insufficient to capture detailed specular reflections under the flash illumination, we rely on multiple exposures either varying the exposure time Δt (for cell phones with a fixed flash intensity), or varying the flash intensity Δg (often described as the EV number in DSLR cameras). Our image formation model for pixel position $\mathbf{u}$ can be formulated as [Debevec and Malik 1997]:

$$I(\mathbf{u}) = L(\mathbf{o}; \mathbf{x})\Delta t \Delta g, \qquad (1)$$

where $I(\mathbf{u})$ is the captured image, and $L(\mathbf{o}; \mathbf{x})$ is the outgoing radiance from point $\mathbf{x}$ on the 3D geometry in the view direction $\mathbf{o}$. The captured radiance at point $\mathbf{x}$ can be formulated as the reflection equation:

$$L(\mathbf{o}; \mathbf{x}) = f(\mathbf{i}, \mathbf{o}; \mathbf{x}, \mathbf{n}) L(-\mathbf{i}; \mathbf{x})(\mathbf{n} \cdot \mathbf{i}), \qquad (2)$$

where $f(\mathbf{i}, \mathbf{o}; \mathbf{x}, \mathbf{n})$ is the reflectance function at point $\mathbf{x}$, $\mathbf{n}$ is the normal vector, and $L(-\mathbf{i}; \mathbf{x})$ represents incident light at $\mathbf{x}$ from light vector $\mathbf{i}$.

Finding correspondences. In our hand-held setup, the information about multiple exposures is stored per point $\mathbf{x}$ in 3D, rather than per pixel $\mathbf{u}$ in 2D. We thus need to obtain the geometric relationship between $\mathbf{x}$ and $\mathbf{u}$. We first obtain the camera's intrinsic
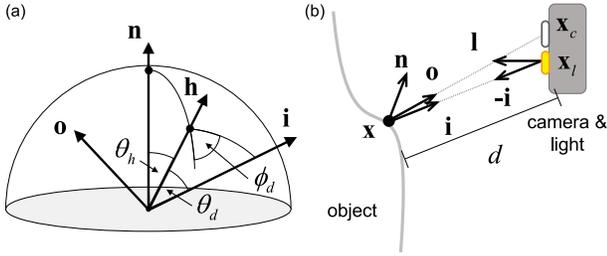
Fig. 3. (a) The Rusinkiewicz parameterization of the light and camera. (b) Geometry of our setup with a smartphone. Due to the position of the flash light and the camera, we are limited to a single sample in θ_d.

Table 1. List of variables used in the paper

$\mathbf{x}_p$	3D position of p-th vertex, $p = 1, \ldots, P$
$\mathbf{n}_p$	geometric normal of $\mathbf{x}_p$
$\tilde{\mathbf{n}}_p$	shading normal of $\mathbf{x}_p$
$\mathbf{i}_{p,k}$	light direction of $\mathbf{x}_p$ at k-th camera
$\mathbf{o}_{p,k}$	view direction of $\mathbf{x}_p$ at k-th camera
$L(\mathbf{o}_{p,k}; \mathbf{x}_p)$	outgoing radiance from $\mathbf{x}_p$ towards $\mathbf{o}_{p,k}$
$L(-\mathbf{i}_{p,k}; \mathbf{x}_p)$	incoming radiance at $\mathbf{x}_p$ from $-\mathbf{i}_{p,k}$
$f(\mathbf{i}, \mathbf{o}; \mathbf{x}, \mathbf{n})$	BRDF at $\mathbf{x}$
$f_b(\mathbf{i}, \mathbf{o})$	b-th basis BRDF, $b = 1, \ldots, B$
$\omega_{p,b}$	blending weight of f_b at $\mathbf{x}_p$
$v_{p,k}$	visibility of $\mathbf{x}_p$ at I_k
$\mathbf{F}_b$	set of basis BRDFs, $\{\mathbf{f}_b\}$
$\mathbf{W}$	set of blending weights, $\{\omega_{p,b}\}$
$\mathbf{N}$	set of shading normals, $\{\tilde{\mathbf{n}}_p\}$
$\mathbf{X}$	set of 3D points, $\{\mathbf{x}_p\}$

parameters using Zhang's method [2000]; this yields the initial relationship between a camera pixel $\mathbf{u} \in \mathbb{R}^2$ and its corresponding captured surface point $\mathbf{x} \in \mathbb{R}^3$ as the perspective projection matrix $\pi \in \mathbb{R}^{3 \times 3}$. Since photographs are captured without any supporting structure, we obtain the rotation/translation transformation matrix $[R|t] \in \mathbb{R}^{3 \times 4}$ defining the extrinsic relationship between the camera and the surface point for each photograph using SfM. In addition, we incrementally update the intrinsic parameters such as focal length for each picture using SfM for more accurate geometric correspondences by compensating the focus breathing effect due to auto-focusing. The resulting intrinsic and extrinsic parameters define the geometric relationship between pixel $\mathbf{u}$ and point $\mathbf{x}$ in Equation (1), as $[\mathbf{u}, 1]^\mathsf{T} = \pi [R|t] [\mathbf{x}, 1]^\mathsf{T}$.

Initial geometry. From the captured images and camera poses, we first obtain a dense 3D point cloud using MVS [Schönberger et al. 2016]. However, this initial point cloud usually suffers from severe high-frequency noise due to specular reflections created by the flash illumination, which violates the underlying diffuse texture assumption of the MVS method. To mitigate this noise, we first create a low-resolution (2^7 voxel grid) mesh using the screened Poisson surface reconstruction [Kazhdan and Hoppe 2013]. We then subdivide this low-resolution mesh to obtain a finer mesh (2^{10}), which is used as the initial geometry of our method. Despite the high resolution of this initial geometry, fine geometric details are missing since they have been removed together with noise during the Poisson reconstruction step. Our iterative geometry update algorithm (see Section 6) recovers these details.

5 SPATIALLY-VARYING REFLECTANCE AND NORMALS

We first aim to obtain SVBRDF and normal information ($\mathbf{W}$, $\mathbf{F}_b$, and $\mathbf{N}$ in Figure 2) from our input photographs. Given a set of P surface (vertex) points $\mathbf{X} = \{\mathbf{x}_p\}$, captured from different light/view directions $\mathbf{i}_k$ and $\mathbf{o}_k$ in K photographs, we can express the captured radiance as $\mathbf{L} = \{L(\mathbf{o}_k; \mathbf{x}_p)\}$. We then formulate the inverse rendering problem that satisfies the image formation model in Equation (2) as finding the set of two unknowns $\{f(\mathbf{i}_{p,k}, \mathbf{o}_{p,k}; \mathbf{x}_p, \mathbf{n}_p), \mathbf{n}_p\}$ that minimizes the following objective function:

$$O = \sum_{p=1}^{P} \sum_{k=1}^{K} v_{p,k} \big(L(\mathbf{o}_{p,k}; \mathbf{x}_p) - f(\mathbf{i}_{p,k}, \mathbf{o}_{p,k}; \mathbf{x}_p, \mathbf{n}_p) L(-\mathbf{i}_{p,k}; \mathbf{x}_p)(\mathbf{n}_p \cdot \mathbf{i}_{p,k}) \big)^2 ,$$

(3)

where $v_{p,k}$ is the visibility function of vertex $\mathbf{x}_p$ in image k. This inverse rendering problem factorizing reflectance and shading is

a severely ill-posed, underdetermined problem. We develop an iterative alternating optimization approach, which updates the four unknown elements $\mathbf{W}$, $\mathbf{F}_b$, $\mathbf{N}$, and $\mathbf{X}$ until the rendering results satisfy the input images. To avoid overfitting, we test the optimized parameters with unused datasets by separating input photographs into different training and test datasets. We first reconstruct the full space of SVBRDFs, and use this information to obtain normals, as explained in the rest of this section.

5.1 Reconstructing the SVBRDF ($\mathbf{F}$)

To obtain the SVBRDF $\mathbf{F}$, we first estimate a set $\mathbf{F}_b = \{f_b\}$ of basis BRDFs, then blend them with spatially-varying weight maps $\mathbf{W}$. Similar to other approaches [Alldrin et al. 2008; Chen et al. 2014; Lawrence et al. 2006; Nam et al. 2016; Wu and Zhou 2015; Zhou et al. 2016], our reconstructed SVBRDF $\mathbf{F}$ in Equation (3) can be formulated as:

$$\mathbf{F} = \big\{ f(\mathbf{i}, \mathbf{o}; \mathbf{x}_p) \big\} = \left\{ \sum_{b=1}^{B} \omega_{p,b} f_b(\mathbf{i}, \mathbf{o}) \right\},$$

(4)

with $\mathbf{W} = \{\omega_{p,b}\}$ the set of per-point blending weights.

Flash photography setup. BRDF acquisition requires dense sampling in θ_h, θ_d, and ϕ_d [Rusinkiewicz 1998] (see Figure 3(a)), which is usually achieved using additional supporting hardware. The specular reflectance changes rapidly as a function of $\theta_h = \cos^{-1}(\mathbf{h} \cdot \mathbf{n})$, where $\mathbf{h}$ is the halfway vector, while Fresnel effects strongly depend on $\theta_d = \cos^{-1}(\mathbf{h} \cdot \mathbf{i})$. On the other hand, reflectance remains almost constant along ϕ_d, which is the azimuth angle of light $\mathbf{i}$ around $\mathbf{h}$. In our setup, the captured datasets include dense sampling along the θ_h and ϕ_d dimensions. However, since the light and the camera are fixed and very close together in our setup (Figure 3(b)), this leads to a *single* sample for θ_d at about $\sim5°$. Coincidentally, note that the angle $\theta_d = 5°$ is the optimal angle for the one-shot capture, as recently shown by Nielsen et al. [2015].

Reflectance model. We use the Cook-Torrance (CT) model [1982] with a non-parametric normal distribution function (NDF) term for better presentations of specular reflection (following a recent evaluation on BRDF models [Bagher et al. 2016]; we do not rely on

analytic functions such as Beckmann [Cook and Torrance 1982] or GGX [Walter et al. 2007] given our dense sampling along the θ_h and ϕ_d angles as explained above). Our basis reflectance model f_b can be expressed as follows:

$$f_b(\mathbf{i}, \mathbf{o}) = \frac{\rho_d}{\pi} + \rho_s \frac{D(\mathbf{h})G(\mathbf{n}, \mathbf{i}, \mathbf{o})F(\mathbf{h}, \mathbf{i})}{4(\mathbf{n} \cdot \mathbf{i})(\mathbf{n} \cdot \mathbf{o})}, \quad (5)$$

where ρ_d and ρ_s are the diffuse and specular albedos, D is the univariate NDF term for specularity, G is the geometric term, and F is the Fresnel term. Our NDF is represented as a non-parametric tabulated function $D(\theta_h) \in \mathbb{R}^M$. We set $M=90$, so that each element stores the BRDF value at corresponding θ_h with the square root mapping of angles, following Matusik's work [2003]. We use the V-groove cavity model for the shadowing/masking term G. Since there is no observation of the Fresnel effect in our setup, we set the term F as a constant [Nam et al. 2016]; in practice, this helps reduce complexity during the optimization. We found that this approximation performs better than using a constant index of refraction (IOR) value for the Fresnel term [Aittala et al. 2013; Dong et al. 2010; Xia et al. 2016], as Figure 8 will show.

Reconstructing SVBRDFs. We can represent the basis BRDF in Equation (5) as a coefficient vector $f_b = [\rho_d, \rho_s FD(\theta_h)]^\mathsf{T} \in \mathbb{R}^{M+1}$. Note that the geometric factor G is excluded from the coefficient vector. From Equation (2), we first convert captured radiance into captured reflectance $f'_{p,k}$. Given this captured reflectance $f'_{p,k}$ and a measurement vector $\Phi_{p,k} \in \mathbb{R}^{M+1}$ specifying the sampled θ_h angles and the geometric factor $G/(4(\mathbf{n} \cdot \mathbf{i})(\mathbf{n} \cdot \mathbf{o}))$ per pixel observation, we have $f'_{p,k} = \Phi_{p,k}^\mathsf{T} f_p$. We can blend the basis BRDFs f_b and spatial weights $\omega_{p,b}$ to approximate $\sum_{b=1}^{B} \omega_{p,b} f_b$. We can then reformulate Equation (3) as an objective function to reconstruct the basis BRDFs and its corresponding weights as:

$$\underset{\mathbf{F}_b, \mathbf{W}}{\text{minimize}} \sum_{p=1}^{P} \sum_{k=1}^{K} v_{p,k} \left(f'_{p,k} - \Phi_{p,k}^\mathsf{T} \sum_{b=1}^{B} \omega_{p,b} f_b \right)^2. \quad (6)$$

For a more stable separation of diffuse and specular reflection, we clamp $D(\theta_h > 60)$ to zero, following existing distribution functions such as Beckmann and GGX used in recent rendering works [Lee et al. 2018]. In addition, we impose non-negative and monotonicity constraints on $D(\theta_h)$ (which should monotonically decrease as θ_h increases). We do not impose a smoothness constraint in order to reproduce strong specular peaks more accurately. To update $\mathbf{F}_b$, we minimize Equation (6) while keeping $\mathbf{W}$ fixed. It becomes a quadratic programming problem on $\mathbf{F}_b$ with sparse input data. We use a commercial sparse quadratic programming solver (NAG e04nkc) [NAG 2015].

Reconstructing $\mathbf{W}$. We first estimate the set of diffuse components of the basis BRDF $\mathbf{F}_b$ by averaging color observations around the median brightness per vertex, clustering materials using K-mean in CIELAB space: this leads to an initial binary-labeled set of weights $\mathbf{W}$. Finding the optimal number of basis BRDFs is still an open challenge. Zhou et al. [2016] determine the number of basis progressively in an ad-hoc manner, increasing it until the optimization converges. We instead follow other prior works [Alldrin et al. 2008; Lawrence et al. 2006; Lensch et al. 2003; Nam et al. 2016; Ren

et al. 2011], and set this number empirically. In Figure 13, we explore the impact of this basis number in our results.

In subsequent iterations, we update $\mathbf{W}$ using the estimated $\mathbf{F}_b$ from the previous optimization. Updating $\mathbf{W}$ with fixed $\mathbf{F}_b$ in Equation (6) is equivalent to minimizing the following objective function on each point $\mathbf{x}_p$:

$$\underset{\boldsymbol{\omega}_p}{\text{minimize}} \frac{1}{2}\|\mathbf{Q}\boldsymbol{\omega}_p - \mathbf{r}\|^2 \quad \text{s.t.} \quad \omega_{p,b} > 0, \quad \sum_{b=1}^{B} \omega_{p,b} = 1. \quad (7)$$

The k-th row in $\mathbf{Q} \in \mathbb{R}^{K \times B}$ is $\left[f_1(\mathbf{i}_{p,k}, \mathbf{o}_{p,k}), \dots, f_B(\mathbf{i}_{p,k}, \mathbf{o}_{p,k}) \right]$, and and the k-th element of $\mathbf{r} \in \mathbb{R}^K$ is $L(\mathbf{o}_{p,k}; \mathbf{x}_p)/L(-\mathbf{i}_{p,k}; \mathbf{x}_p)/(\mathbf{n}_p \cdot \mathbf{i}_{p,k})$, where we apply the visibilities $v_{p,k}$ and additional weights $w_{p,k} = cos(\theta_i)sin(\theta_h)$ (where $\theta_i = cos^{-1}(\mathbf{n}_p \cdot \mathbf{i}_{p,k})$ and $\theta_h = cos^{-1}(\mathbf{n}_p \cdot \mathbf{h}_{p,k})$) to each element of both $\mathbf{Q}$ and $\mathbf{r}$. The cosine terms compensate for the unstable observations at grazing angles, while the sine terms prevent bias towards specular observations. Minimizing Equation (7) is a standard quadratic programming problem, which we solve using the convex quadratic programming solver (NAG e04ncc).

Color. We use color basis BRDFs and monochromatic blending weights. When updating $\mathbf{F}_b$, we optimize for each color channel independently. When updating $\mathbf{W}$, we optimize blending weights of each vertex using all color channels.

5.2 Reconstructing Normals ($\mathbf{N}$)

Once we obtain a set of basis BRDFs $\mathbf{F}_b = \{f_b\}$ and blending weights $\mathbf{W} = \{\omega_{p,b}\}$ to yield SVBRDFs $\mathbf{F}$, we need to estimate a set of per-vertex shading normals $\mathbf{N} = \{\tilde{\mathbf{n}}_p\}$. Since estimating surface normals with BRDFs is again a severely ill-posed problem, we apply an iterative optimization. First, we feed initial surface normals $\mathbf{n}_p$ from the current geometry (updated in the previous iteration) as input variable for BRDF f at point $\mathbf{x}_p$ in Equation (3). Since incoming/outgoing radiance, incoming/outgoing directions, and the reflection of the vertex are all known, we can factorize shading normals $\tilde{\mathbf{n}}_p$ using a standard linear least squares regression with no constraints. Note that geometric normals $\mathbf{n}_p$ at point $\mathbf{x}_p$ are different from shading normal $\tilde{\mathbf{n}}_p$. As we iterate the optimizations, the difference between $\mathbf{n}_p$ and $\tilde{\mathbf{n}}_p$ gradually converges (see Figure 15). Since observations at grazing and mirror-reflection angles are less reliable, we introduce an additional weight taking into account illumination angles, as $w_p = cos(\theta_i) sin(\theta_h)$, where $\theta_i = cos^{-1}(\mathbf{n} \cdot \mathbf{i})$. The sine term prevents potential bias towards misaligned surface normals and specular reflectances from previous estimations.

6 GEOMETRY UPDATE WITH PHOTOMETRIC CONSISTENCY

After estimating weights $\mathbf{W}$, basis BRDFs $\mathbf{F}_b$ and shading normals $\mathbf{N}$, we now aim to reconstruct the geometry $\mathbf{X}$ that agrees with the shading observations. Figure 4 illustrates this process. From an initial point cloud, we first obtain a rough base geometry, which we subdivide into a finer mesh. We then update this geometry with estimated shading normals, for which two options exist: Nehab's method [2005], and the screened Poisson reconstruction method [Kazhdan and Hoppe 2013]. While in principle both methods can preserve both low- and high-frequency details, Nehab's method

transfers surface gradients to the target geometry directly, while the Poisson method employs a coarse-to-fine reconstruction. This is a crucial difference in our unstructured capture setup, since the shading normals $\mathbf{N}$ can be assumed to contain high frequency noise, and the direct transfer in Nehab's method would also transfer this noise. We thus choose the screened Poisson method, designed to reconstruct implicit surfaces using tri-quadratic B-spline basis functions in a voxel grid in the coarse-to-fine approach. This leads to a robust performance when integrating noisy surface normals to 3D geometry. Figure 5 shows a comparison between the two methods.

The screened Poisson method reconstructs an implicit surface $\hat{\chi}$ from an input point cloud as:

$$\underset{\chi}{\text{minimize}} \int \left\| \mathbf{V}(\mathbf{x}_p) - \nabla \chi(\mathbf{x}_p) \right\|^2 d\mathbf{x}_p + \alpha \sum_{\mathbf{x}_p \in \mathbf{X}} \chi^2(\mathbf{x}_p), \quad (8)$$

where $\mathbf{V} \colon \mathbb{R}^3 \to \mathbb{R}^3$ is a vector field derived from the set of shading normals $\mathbf{N}$, $\nabla \chi(\mathbf{x}_p)$ is the gradient of the implicit scalar surface function $\chi \colon \mathbb{R}^3 \to \mathbb{R}$, $\chi^2(\mathbf{x}_p)$ is the squared distance between a point $\mathbf{x}_p$ and the implicit surface χ, and α is the weight of the regularization term; we determine $\alpha \in [0.1, 4.0]$ depending on the confidence of the initial geometry. For discretizing the implicit surface function χ, we set the resolution of our voxel grid to 2^9 or 2^{10} for each dimension, which roughly corresponds to $0.1 - 0.2$ mm for the physical objects captured in this paper.

While the original algorithm employs geometric normals, we leverage our shading normals $\tilde{\mathbf{n}}_p$, and aim to find an implicit surface $\hat{\chi}$ whose gradients match $\tilde{\mathbf{n}}_p$ instead (i.e., each vertex should present consistent shading given different view and light directions). Once the implicit function is determined, we apply marching cubes [Kazhdan and Hoppe 2013] to convert it to a polygonal mesh.

We iteratively update $\mathbf{W}$, $\mathbf{F}_b$, $\mathbf{N}$, and $\mathbf{X}$ until we find the optimal 3D geometry and SVBRDF. We evaluate the Hausdorff distance [Cignoni et al. 1998] between the previous mesh and the new $\mathbf{X}$. We repeat the whole process in Figure 2 until the test RMS error of the photometric difference in Equation (3) starts to increase. To avoid overfitting, we randomly separate captured images into training and testing groups with a 9:1 ratio.
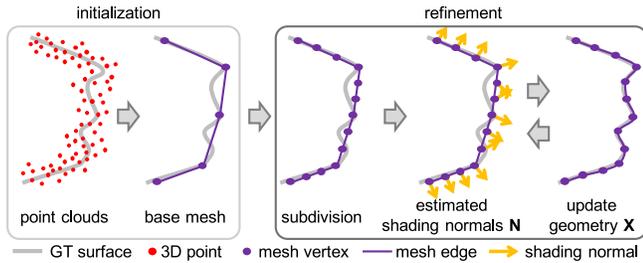

Fig. 4. Updating 3D geometry. Initialization yields a rough base geometry from an initial point cloud. After subdividing it into a finer mesh, we apply Poisson surface reconstruction with photometric consistency.

7 RESULTS AND EVALUATION

Captured results. We demonstrate our method with two different off-the-shelf cameras: a DSLR (Nikon D7000) and a mobile phone
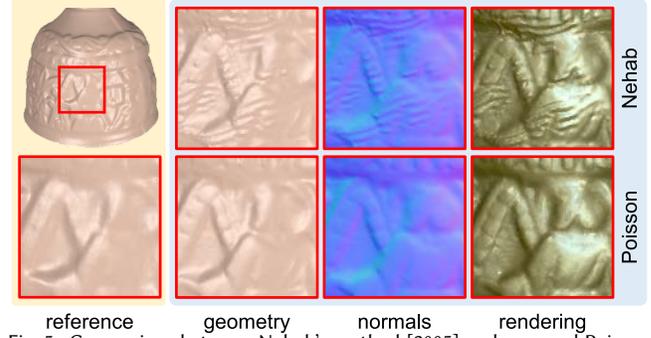
Fig. 5. Comparison between Nehab's method [2005] and screened Poisson surface reconstruction [Kazhdan and Hoppe 2013] for our unstructured capture setup. The result of Nehab method contains high-frequency noise from the estimated shading normals in our unstructured capture setup. In contrast, the geometry from the screened Poisson method presents cleaner details, handling the input noise robustly through its coarse-to-fine reconstruction.

(Nexus 5X). When using a mobile phone, we use the official Android API which supports RAW data, and allows to adjust camera parameters such as exposure time, f-stop, or ISO. We capture between $100 - 400$ images per object, which takes approximately $10 - 20$ minutes.

Figure 6 shows some of our reconstructed results, including side-by-side comparisons with a photograph, novel renderings under environment maps, and the recovered normal maps. These objects span a wide range of geometries and materials, including metal, wood, plastic, ceramic, resin, and paper. Our method yields good results in all cases, including complex shapes like the genus-one shape of the ceramic frog. Please refer to the supplemental material for videos.

Influence of ambient light. Using a mobile phone, we capture the input images under dark illumination to minimize the impact of ambient light; however, this is not necessary using a DLSR, which has a stronger built-in flash. This is shown in Figure 7. We capture one image I_1 under both flash light and indoor lighting (Figure 7(b)), and a second one I_2 without the flash (Figure 7(c)), using the same capture settings (1/250 sec., $f/16$, ISO 100) for both images. We calculate the signal-to-noise ratio (SNR) from those images as $SNR = 20\log_{10}\left(\|I_1 - I_2\|_F / \|I_2\|_F\right)$, where $\|\cdot\|$ is the Frobenius norm. The flash image shows a significantly higher SNR, 85.22 dB, which confirms that the built-in flash in DSLRs is bright enough to remove the need for a darkroom.

Validation of our BRDF model. We have validated our reconstructions against the full MERL BRDF dataset, following the process shown in Figure 8(a). From a full 3D BRDF, we subsample a (θ_d, ϕ_d) plane at $\theta_d = 5°$, which corresponds to the BRDF sampling angles in our flash photography setup. We add random noise to the subsampled data, whose average intensity is set to one tenth of the original signal intensity. We then fit the noisy samples to our BRDF model and reconstruct the full 3D BRDF using the fitted parameters. As opposed to previous works that model the Fresnel function with a fixed IOR [Aittala et al. 2013; Dong et al. 2010; Xia et al. 2016], we combine a fixed Fresnel term ($F(\mathbf{h}, \mathbf{i}) = 1.0$) with specular albedo ρ_s as a Fresnel color vector instead. This is motivated by the fact that
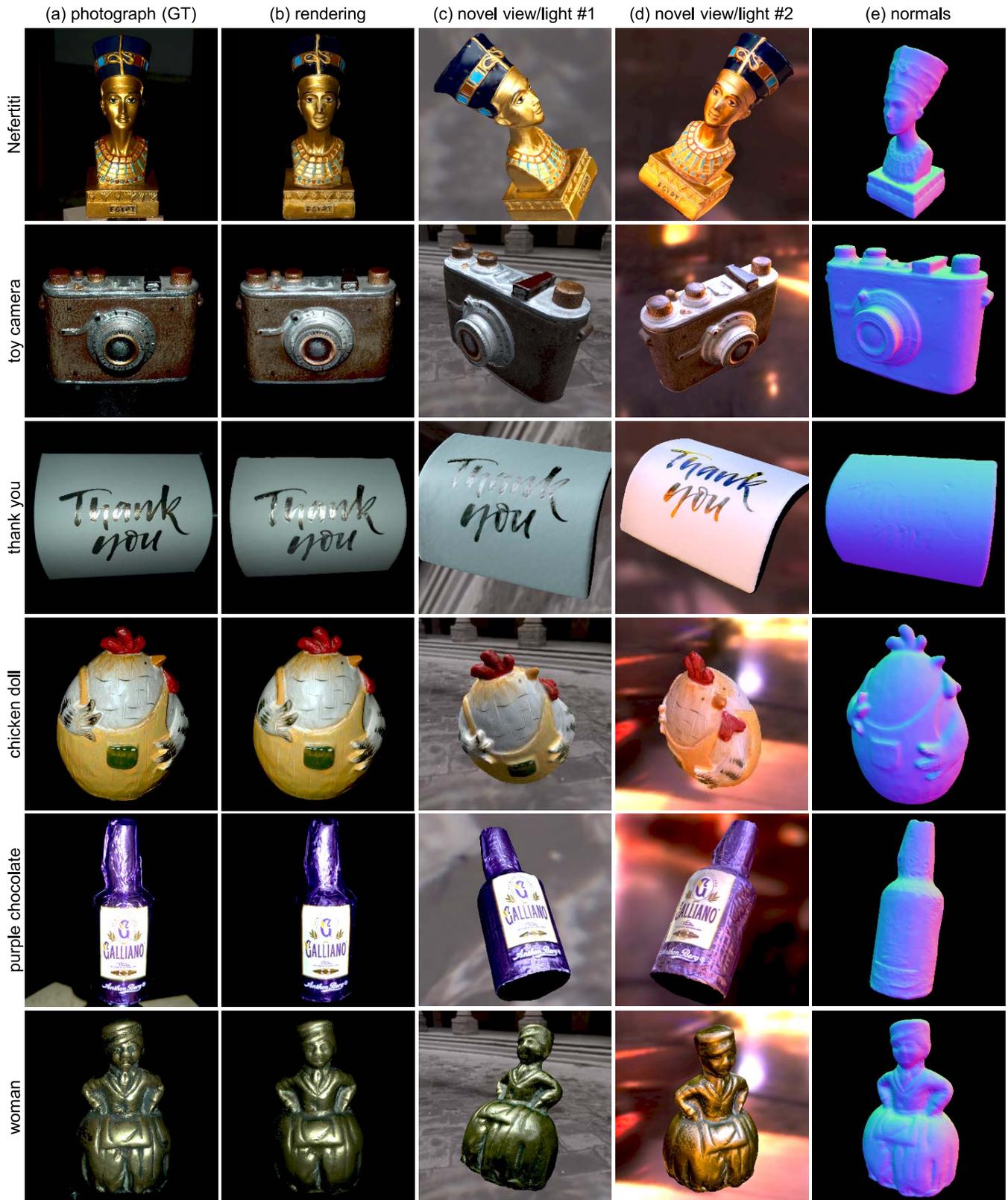
Fig. 6. Results of our acquisition. For each object, we present a photograph (a), our reconstruction (b), two novel renderings (c) and(d), and surface normals (e). Please refer to the supplemental materials for the video versions.

Fig. 7. (a) Capture setup using a DLSR camera. (b) Captured image using the flash light (1/250 sec., f/16, ISO 100). Its high SNR of 85.22 dB, confirms that the built-in flash is bright enough to avoid the need for a darkroom. (c) Captured image (amplified x16 for visualization) with only ambient illumination without flash using the same settings.
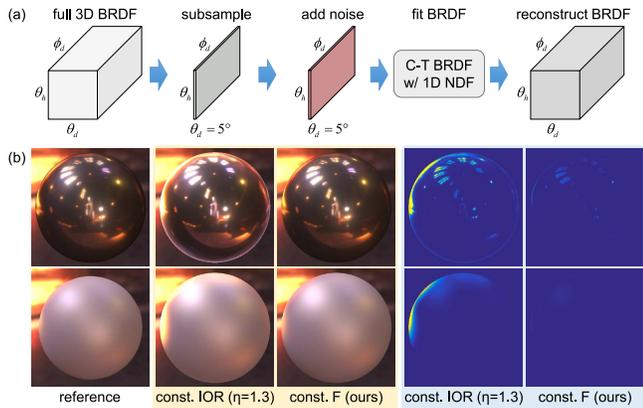
Fig. 8. (a) Validation experiment design of our BRDF reconstruction method. (b) Comparison of our reconstructed BRDFs with two measured BRDFs from the MERL dataset (top: brass; bottom: pearl paint). The second and fourth columns show the result using a constant index of refraction (IOR). The third and fifth show our results with a constant Fresnel term F, leading to a more accurate match. The average PSNR for the 100 MERL BRDFs are 12.98 dB and 34.50 dB respectively, for the two approaches. Please refer to the supplemental material for results on the full MERL dataset.

when the Fresnel effect increases rapidly as θ_d approaches $\pi/2$, the denominator $4(\mathbf{n} \cdot \mathbf{i})(\mathbf{n} \cdot \mathbf{o})$ in Equation (5) decreases significantly, leading to poor reflectance estimations from inaccurate F.

Figure 8(b) shows rendering results of the reconstructed BRDFs for two examples, (top) brass and (bottom) pearl paint. The second column shows the estimated appearance with a constant IOR (as previous works), while in the third we show the result of our constant Fresnel approximation; as the false color maps indicate, a constant Fresnel term leads to more accurate reconstructions. The average PSNR of rendering results for the MERL dataset is 12.98 dB and 34.50 dB, respectively, for the two different Fresnel terms. Please refer to the supplemental material for all the results.

Geometric accuracy. We evaluate our geometry reconstruction accuracy, compared against a commercial 3D desktop scanner (NextEngine), and the state of the art 3D reconstruction method COLMAP [Schönberger 2016]. Since the quality of the output geometry from COLMAP depends on the voxel resolution, we include both a high-and a low-resolution result (2^{10}, and 2^7 voxel grids, respectively). As Figure 11 shows, the high-resolution reconstruction contains high-frequency errors, while the low resolution leads to over-smooth surfaces. Starting from this low resolution result, our method yields
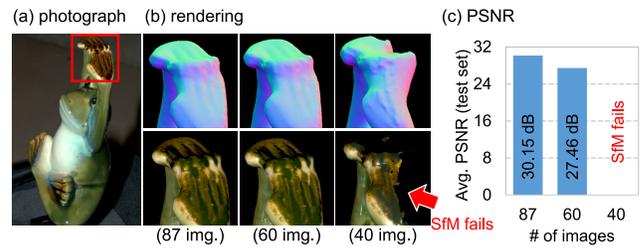

Fig. 9. Impact of the number of input images. (a) Captured photograph of the ceramic frog. (b) Rendering results using 87, 60, and 40 input images. (c) Corresponding PSNR values. Even with only 60 images, our method yields a good reconstruction quality, with a plausible PSNR value (27.46 dB). Using 50 images or less, SfM/MVS fails to produce a valid initial geometry, leading to poor results.

a final reconstruction on-par with the professional 3D scanner, while also providing spatially-varying reflectance information. The average Hausdorff differences [Cignoni et al. 1998] from the reference scanned geometry for the three methods are 0.1063, 0.0917, and 0.0593 mm, respectively. The input images were taken with a hand-held smartphone camera.

Impact of initial geometry. Figure 12 shows the impact of the input geometry on the final results. We evaluate our geometric reconstruction from two different input geometries: one from visual hull [Matusik et al. 2000] (Figure 12, top), and one from COLMAP (bottom). For both, structural details and reflectance information are gradually improved in each iteration by our method (notice how the stair-like artifacts in the initial visual hull reconstruction disappear). While both input geometries lead to good results, the COLMAP input geometry yields sharper details. The methods converged after nine and seven iterations, respectively.

Impact of the number of basis BRDFs. Figure 13 evaluates the impact of the number of basis BRDFs B on the final reflectance. The RMSE between reconstructions and photographs in the test dataset rapidly decreases as B increases. For the object shown in the figure, we choose $B = 9$; for the other objects shown in this paper, Figure 14 shows the basis BRDFs (upper right spheres) and their associated blending weight maps. Specular reflections and diffuse albedos are well separated by the weight maps.

Impact of the number of input images. We have evaluated the impact of the number of input images in the final result (shown in Figure 9 for the ceramic frog). Our method degrades gracefully as this number decreases, until it reaches a threshold where MVS fails to yield a complete 3D point cloud, resulting in only a partial geometry of the object, or SfM fails to find enough camera poses to reconstruct 3D geometry. For the object shown, MVS fails with less than 50 images, while SfM fails with less than 40. Using as few as 60 images still produces good reconstruction results, with a high PSNR. This behavior is similar across different objects: our method is robust and yields good reconstructions even with a reduced set of images, provided that SfM and MVS can generate an initial geometry.

Still images vs. video frames. We analyze the potential advantage of using video instead of still images, due to its simpler capture procedure and its potential to provide more input images. Figure 10
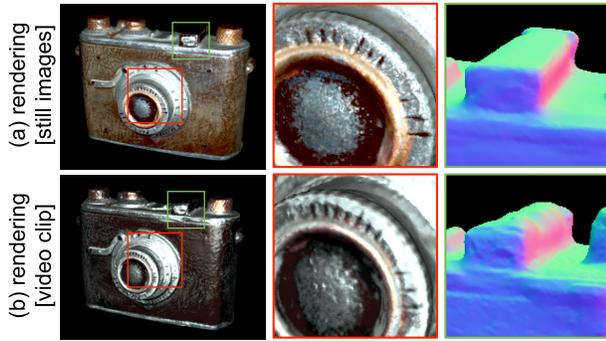
Fig. 10. Comparison using video frames as input. The top row shows a reconstructed 3D object from 195 still images, while the bottom row was generated from 1,009 video frames. Due to motion blur, inaccurate focus, and lower dynamic range, the reconstruction using video frames is less accurate (both in terms of color and geometry), especially in dark areas of the object.

compares the corresponding reconstructions. The top row shows the result from 195 still images, captured with a DSLR camera (Nikon D7000), while the bottom row was produced from 1,009 video frames captured by a mobile phone camera (Nexus 5X; we recovered linear signals of the video frames using existing methods [Aittala et al. 2015; Hui et al. 2017; Riviere et al. 2015]). Due to motion blur, inaccurate focus, and a lower dynamic range than the DSLR still images, the results from video frames show visible color shifts and geometric artifacts, especially in dark regions. Note that we use a mobile camera to capture video since the DSLR camera flash light cannot illuminate an object continuously.

Iterative optimization. Figure 15 analyzes the error evolution during the optimization for the Bell object in Figure 11. Figure 15(a) shows the photometric error, i.e., the difference between rendered and captured images, where the RMSE values for both training and test sets first decrease, then start to increase. We thus stop after the fifth iteration for this object. Refer to Table 2 for reconstruction details for every object.

Table 2. Reconstruction details for every object.

object	dominant material	camera type	# of img	# of BRDF	RMSE (train)	RMSE (test)
Nefertiti	painted resin	DSLR	202	13	0.5206	0.5619
toy camera	coated plastic	DSLR	195	9	0.4336	0.5130
thank you	paper	mobile	175	2	0.1401	0.1433
chicken doll	coated wood	DSLR	112	9	0.4677	0.4709
purple chocolate	alumn. foil	mobile	253	10	0.1090	0.1162
woman	bronze	mobile	183	9	0.1679	0.1870
bell	bronze	mobile	232	9	0.2175	0.2288
frog doll	ceramic	DSLR	87	5	0.2185	0.2299
metal cup	metal	mobile	416	7	0.1858	0.1975

Figure 15(b) shows the evolution of the geometry; our reconstruction pipeline is robust, and rapidly updates the bulk of the geometry in the first few iterations, then refines the geometry so that errors decrease more slowly. Last, Figure 15(c) plots the angle differences between the geometric normals $\tilde{\mathbf{n}}_p$ and the shading normals $\mathbf{n}_p$ for

every vertex, described in Section 5.2. The error starts to increase after the fifth iteration, possibly due to SVBRDF overfitting.

Comparison with a spherical illumination method. We compare in Figure 16 our method against another using spherical illumination [Xia et al. 2016], using the same physical object as in Xia's paper. Xia's method estimates the geometry, SVBRDF and incident illumination simultaneously; however, high-frequency details on the surface appear to be lost during the reconstruction. In contrast, our method with active local illumination preserves more geometric details, and leads to more accurate SVBRDFs. Note that we used a smartphone camera to obtain the input images.

Processing time. Each step of the initialization for obtaining the base rough geometry and extrinsic camera parameters takes: (SfM) ~5 minutes, (MVS) 2 – 4 hours, (meshing) ~1 minute. One iteration of our optimization takes about ~10 minutes, equally distributed for each step ($\mathbf{W}$, $\mathbf{F}_b$, $\mathbf{N}$, and $\mathbf{X}$). For the experiment, we used a desktop computer with Intel i7-3770 CPU 3.40 GHz and 32 GB of memory and an NVIDIA GTX1080 GPU.

8 DISCUSSION

We discuss and motivate in this section some key aspects of our method, as well as its limitations and potential avenues of future work.

Shadowing and masking. We adopt the V-groove cavity model for the shadowing/masking term G in our basis BRDF model for simplicity, given that G is generally smooth and can be simplified without significant loss of visual quality [Holroyd et al. 2008]. Although G can be derived from a data-driven normal distribution function D term (e.g., [Ashikhmin et al. 2000]), this requires an integration of D over the hemisphere for each observation, which is computationally expensive.

Sparsity of basis BRDFs. We do not impose any sparsity constraint when optimizing the blending weights of the basis BRDFs. Some previous works explicitly limit the number of non-zero blending weights per surface point to one or two; this is advantageous when the goal is material editing or material decomposition [Lawrence et al. 2006], or when there are just a few input images [Zhou et al. 2016]. In our case, this would unnecessarily limit the optimization process, leading to inaccurate reflectance reconstructions; as shown in Figure 14, accurate appearance is achieved when blending many basis BRDFs.

Types of cameras and flash. Our reconstruction algorithm does not depend on the type of cameras or the type of flash light; most off-the-shelf cameras in the market, e.g., DLSRs, mirrorless cameras, point-and-shoot cameras, and mobile phones, have built-in flash lights, and thus are suitable for our method. As we have shown in Section 7 with the built-in flash of an entry-level DSLR (e.g., Nikon D7000), a bright enough flash minimizes the impact of ambient illumination.

Limitations and future work. Our method is not free from limitations, which open up several avenues of interesting future work. Due to the hand-held nature of our capture setup, it may be that not all surface points are properly captured, missing specular highlights
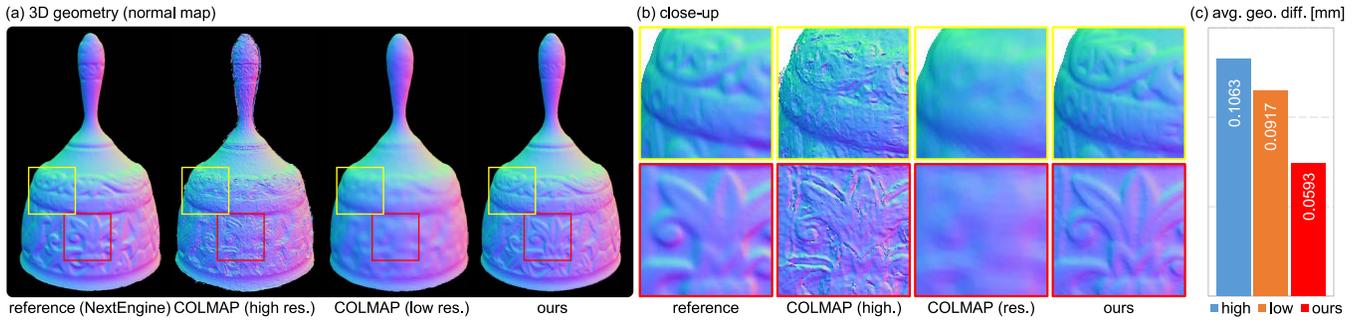
Fig. 11. We compare our geometric reconstruction with a reference geometry from a commercial 3D scanner (NextEngine), a high-resolution reconstruction using COLMAP (2^{10} voxel grid for each dimension, same resolution as our output), and a low-resolution reconstruction by COLMAP (2^7 voxel grid, used as input for our method). Our reconstruction leads to sharper and cleaner geometry, closer to the reference scanned geometry. The average differences from the reference geometry are shown on the right.
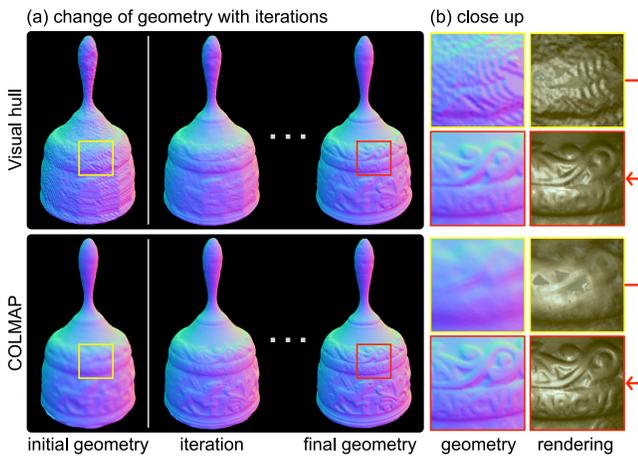

Fig. 12. Impact of the input geometry. We evaluate our geometric reconstruction results from two different input geometry: visual hull (top), and COLMAP (bottom). (a) Progressive updates from the initial geometry. (b) Close up results including SVBRDF rendering. Although in general an input COLMAP geometry leads to sharper results, note how our method is robust enough to eliminate the stair-like artifacts from the visual hull input geometry.

at mirror reflection angles. An alternative approach to represent SVBRDF is to estimate per-point BRDFs independently, as presented in [Hui et al. 2017; Riviere et al. 2015; Xia et al. 2016]. In this case, every surface point needs to be captured with at least one specular and one diffuse observation. This may be achieved if the target objects are 2D planes [Hui et al. 2017; Riviere et al. 2015], or if spherical illumination is used [Xia et al. 2016]. In our capture setup, we could increase significantly the number of input images. However, this becomes impractical and there is no obvious way to ensure that all the information has been captured. Instead, we have chosen a basis BRDF approach with their corresponding weight maps.

In terms of geometry, our method inherits some of the fundamental limitations of image-based 3D modeling techniques: it is thus possible that overly complex geometries like pinecones cannot be reconstructed accurately.
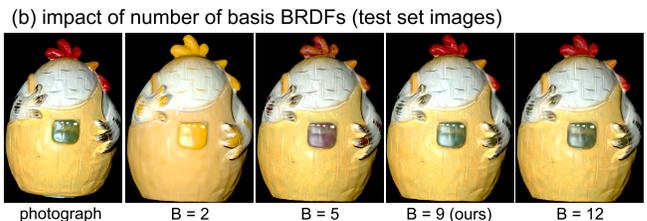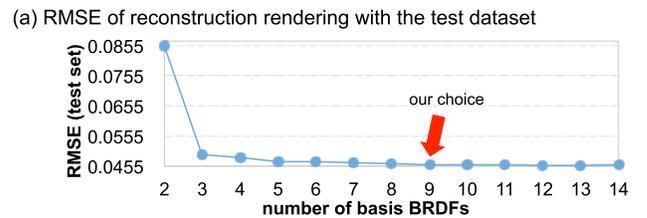

Fig. 13. Impact of the number of basis BRDFs. (a) Photograph in the test set. (b) Evolution of the photometric consistency error (in RMSE) with the number of basis B. (c) Rendering images with increasing basis BRDFs. Estimations with a small numbers ($B < 7$) tend to alter the colors in the reconstruction, while there are almost no visual differences when the number is large enough ($B > 7$).
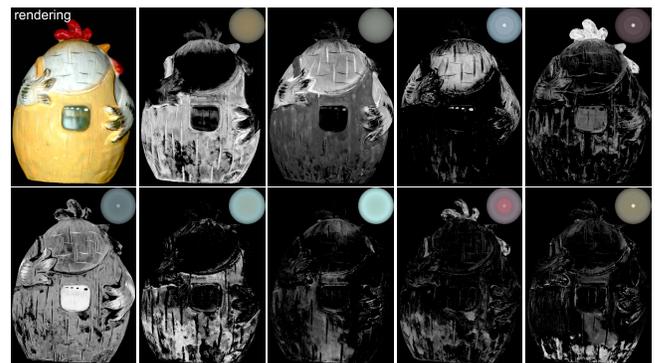

Fig. 14. Basis BRDFs (upper right sphere) and their associated blending weight maps. Specular reflections and diffuse albedos are well separated by the weight maps.
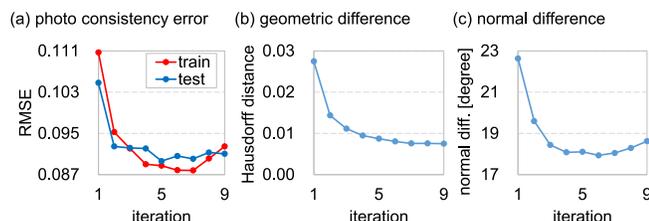
Fig. 15. (a) Photometric errors between reconstructions and photographs (object: Bell, shown in Figure 11). The RMSE values for both training and test sets decrease for several iterations, then start to increase at which point we stop the process (iteration #5). (b) Geometric errors. (c) Angle differences between the geometric normals and the estimated shading normals. The angle differences increase after the fifth iteration possibly due to SVBRDF overfitting.
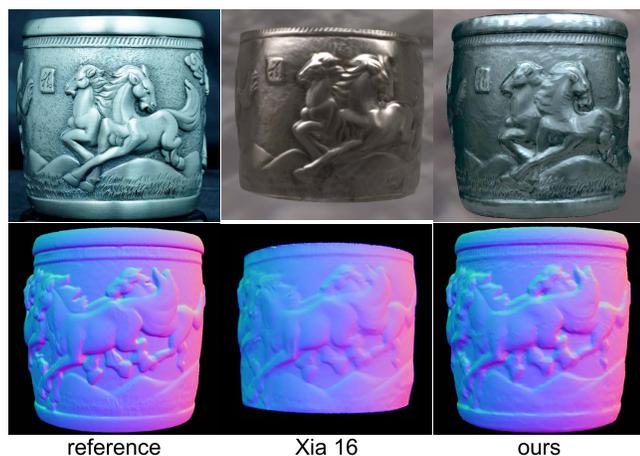

Fig. 16. We compare our method against the recent work by Xia et al., using spherical illumination [Xia et al. 2016]. (left) Reference photograph and geometry from the NextEngine 3D scanner. (middle) Results from Xia's method (images taken from the authors' paper). (right) Our results, showing more accurate geometry and reflectance information.

Last, we do not take into account interreflections, subsurface scattering, nor transparency in our light transport model. Although we show results with a wide range of materials, capturing objects with these characteristics with our unstructured approach remains an open challenge.

9 CONCLUSION

In summary, we have presented a novel solution to simultaneously reconstruct spatially-varying reflectance and 3D geometry using just an off-the-shelf camera, by jointly formulating three reconstruction problems (SVBRDF, shading normals and 3D geometry) as one. Current works aiming at obtaining similar information simultaneously usually require more complex hardware, limiting their applicability. We have shown the performance of our method over a wide range of materials and geometries, with results that are comparable or many times superior to state-of-the-art methods for capturing *only* reflectance or geometry; our geometric reconstructions are comparable to commercial 3D desktop scanning systems.

We believe that our work offers an attractive solution, which can facilitate in-the-wild geometry and reflectance acquisition for a wider public.

REFERENCES

Miika Aittala, Tim Weyrich, and Jaakko Lehtinen. 2013. Practical SVBRDF Capture in the Frequency Domain. *ACM Transactions on Graphics (TOG)* 32, 4, Article 110 (2013), 12 pages.

Miika Aittala, Tim Weyrich, and Jaakko Lehtinen. 2015. Two-shot SVBRDF capture for stationary materials. *ACM Transactions on Graphics (TOG)* 34, 4 (2015), 110:1–13.

Neil Alldrin, Todd Zickler, and David Kriegman. 2008. Photometric stereo with nonparametric and spatially-varying reflectance. In *Proc. IEEE Conf. Computer Vision and Pattern Recognition (CVPR) 2008*. 1–8.

Michael Ashikhmin, Simon Premoze, and Peter Shirley. 2000. A microfacet-based BRDF generator. In *Proc. ACM SIGGRAPH 2000*. 65–74. https://doi.org/10.1145/344779.344814

Seung-Hwan Baek, Daniel S. Jeon, Xin Tong, and Min H. Kim. 2018. Simultaneous Acquisition of Polarimetric SVBRDF and Normals. *ACM Transactions on Graphics (TOG)* 37, 6 (2018).

Mahdi M Bagher, John Snyder, and Derek Nowrouzezahrai. 2016. A Non-Parametric Factor Microfacet Model for Isotropic BRDFs. *ACM Transactions on Graphics (TOG)* 35, 5 (2016), 159.

Guojun Chen, Yue Dong, Pieter Peers, Jiawan Zhang, and Xin Tong. 2014. Reflectance scanning: estimating shading frame and BRDF with generalized linear light sources. *ACM Transactions on Graphics (TOG)* 33, 4 (2014), 117:1–11.

Paolo Cignoni, Claudio Rocchini, and Roberto Scopigno. 1998. Metro: Measuring error on simplified surfaces. In *Computer Graphics Forum*, Vol. 17. Wiley Online Library, 167–174.

Robert L. Cook and Kenneth E. Torrance. 1982. A Reflectance Model for Computer Graphics. *ACM Transactions on Graphics (TOG)* 1, 1 (1982), 7–24. https://doi.org/10.1145/357290.357293

Paul E. Debevec and Jitendra Malik. 1997. Recovering High Dynamic Range Radiance Maps from Photographs. In *Proc. ACM SIGGRAPH '97*. 369–378.

Yue Dong, Jiaping Wang, Xin Tong, John Snyder, Yanxiang Lan, Moshe Ben-Ezra, and Baining Guo. 2010. Manifold bootstrapping for SVBRDF capture. *ACM Transactions on Graphics (TOG)* 29, 4 (2010), 98.

Andrew Gardner, Chris Tchou, Tim Hawkins, and Paul Debevec. 2003. Linear Light Source Reflectometry. *ACM Transactions on Graphics (TOG)* 22, 3 (2003), 749–758.

Abhijeet Ghosh, Tongbo Chen, Pieter Peers, Cyrus A. Wilson, and Paul Debevec. 2010. Circularly Polarized Spherical Illumination Reflectometry. *ACM Transactions on Graphics (TOG)* (Dec. 2010).

Abhijeet Ghosh, Graham Fyffe, Borom Tunwattanapong, Jay Busch, Xueming Yu, and Paul Debevec. 2011. Multiview face capture using polarized spherical gradient illumination. *ACM Transactions on Graphics (TOG)* 30, 6 (2011), 129.

Abhijeet Ghosh, Tim Hawkins, Pieter Peers, Sune Frederiksen, and Paul Debevec. 2008. Practical modeling and acquisition of layered facial reflectance. *ACM Transactions on Graphics (TOG)* 27, 5 (2008), 139.

Paul Graham, Borom Tunwattanapong, Jay Busch, Xueming Yu, Andrew Jones, Paul Debevec, and Abhijeet Ghosh. 2013. Measurement-Based Synthesis of Facial Microgeometry. *Computer Graphics Forum* 32 (2013), 335–344.

Darya Guarnera, Giuseppe Guarnera, Abhijeet Ghosh, Cornelia Denk, and Mashhuda Glencross. 2016. BRDF Representation and Acquisition. *Computer Graphics Forum* 35 (2016), 625–650.

Carlos Hernandez, George Vogiatzis, and Roberto Cipolla. 2008. Multiview photometric stereo. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 30, 3 (2008), 548–554.

Tomoaki Higo, Yasuyuki Matsushita, Neel Joshi, and Katsushi Ikeuchi. 2009. A hand-held photometric stereo camera for 3-d modeling. In *Computer Vision, 2009 IEEE 12th International Conference on*. IEEE, 1234–1241.

Michael Holroyd, Jason Lawrence, Greg Humphreys, and Todd Zickler. 2008. A photometric approach for estimating normals and tangents. *ACM Transactions on Graphics (TOG)* 27, 5 (2008), 133.

Michael Holroyd, Jason Lawrence, and Todd Zickler. 2010. A coaxial optical scanner for synchronous acquisition of 3D geometry and surface reflectance. *ACM Transactions on Graphics (TOG)* 29, 3 (2010), 99.

Zhuo Hui, Kalyan Sunkavalli, Joon-Young Lee, Sunil Hadap, Jian Wang, and Aswin C Sankaranarayanan. 2017. Reflectance Capture Using Univariate Sampling of BRDFs. In *Proc. IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. 5362–5370.

Olaf Kahler, Victor Adrian Prisacariu, Carl Yuheng Ren, Xin Sun, Philip Torr, and David Murray. 2015. Very high frame rate volumetric integration of depth images on mobile devices. *IEEE Transactions on Visualization and Computer Graphics* 21, 11 (2015), 1241–1250.

Michael Kazhdan and Hugues Hoppe. 2013. Screened poisson surface reconstruction. *ACM Transactions on Graphics (TOG)* 32, 3 (2013), 29.

Kalin Kolev, Petri Tanskanen, Pablo Speciale, and Marc Pollefeys. 2014. Turning Mobile Phones into 3D Scanners. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*.

Jason Lawrence, Aner Ben-Artzi, Christopher DeCoro, Wojciech Matusik, Hanspeter Pfister, Ravi Ramamoorthi, and Szymon Rusinkiewicz. 2006. Inverse shade trees for non-parametric material representation and editing. *ACM Transactions on Graphics (TOG)* 25, 3 (2006), 735–745.

Joo Ho Lee, Adrian Jarabo, Daniel S. Jeon, Diego Gutierrez, and Min H. Kim. 2018. Practical Multiple Scattering for Rough Surfaces. *ACM Transactions on Graphics (TOG)* 37, 6 (2018).

Hendrik PA Lensch, Jan Kautz, Michael Goesele, Wolfgang Heidrich, and Hans-Peter Seidel. 2001. Image-based reconstruction of spatially varying materials. In *Rendering Techniques 2001*. Springer, 103–114.

Hendrik P. A. Lensch, Jan Kautz, Michael Goesele, Wolfgang Heidrich, and Hans-Peter Seidel. 2003. Image-based Reconstruction of Spatial Appearance and Geometric Detail. *ACM Transactions on Graphics (TOG)* 22, 2 (2003), 234–257.

Wojciech Matusik, Chris Buehler, Ramesh Raskar, Steven J Gortler, and Leonard McMillan. 2000. Image-based Visual Hulls. In *Proc. ACM SIGGRAPH '00*. 369–374.

Wojciech Matusik, Hanspeter Pfister, Matt Brand, and Leonard McMillan. 2003. A Data-Driven Reflectance Model. *ACM Transactions on Graphics* 22, 3 (July 2003), 759–769.

NAG. 2015. The NAG Library, Numerical Algorithms Group. http://www.nag.com/.

Koki Nagano, Graham Fyffe, Oleg Alexander, Jernej Barbic, Hao Li, Abhijeet Ghosh, and Paul E Debevec. 2015. Skin microstructure deformation with displacement map convolution. *ACM Transactions on Graphics (TOG)* 34, 4 (2015), 109–1.

Giljoo Nam and Min H. Kim. 2014. Multispectral Photometric Stereo for Acquiring High-Fidelity Surface Normals. *IEEE Computer Graphics and Applications* 34, 6 (2014), 57–68.

Giljoo Nam, Joo Ho Lee, Hongzhi Wu, Diego Gutierrez, and Min H. Kim. 2016. Simultaneous Acquisition of Microscale Reflectance and Normals. *ACM Transactions on Graphics (Proc. SIGGRAPH Asia 2016)* 35, 6 (2016). https://doi.org/10.1145/2980179.2980220

Diego Nehab, Szymon Rusinkiewicz, James Davis, and Ravi Ramamoorthi. 2005. Efficiently combining positions and normals for precise 3D geometry. *ACM Trans. Graph* 24, 3 (2005), 536–543. http://doi.acm.org/10.1145/1073204.1073226

Richard A Newcombe, Shahram Izadi, Otmar Hilliges, David Molyneaux, David Kim, Andrew J Davison, Pushmeet Kohi, Jamie Shotton, Steve Hodges, and Andrew Fitzgibbon. 2011. KinectFusion: Real-time dense surface mapping and tracking. In *Proc. Int. Sym. Mixed and augmented reality (ISMAR)*. 127–136.

Jannik Boll Nielsen, Henrik Wann Jensen, and Ravi Ramamoorthi. 2015. On optimal, minimal BRDF sampling for reflectance acquisition. *ACM Transactions on Graphics* 34, 6 (2015), 186.

Peter Ondruska, Pushmeet Kohli, and Shahram Izadi. 2015. Mobilefusion: Real-time volumetric surface reconstruction and dense tracking on mobile phones. *IEEE transactions on visualization and computer graphics* 21, 11 (2015), 1251–1258.

Geoffrey Oxholm and Ko Nishino. 2014. Multiview Shape and Reflectance from Natural Illumination. In *Computer Vision and Pattern Recognition (CVPR), 2014 IEEE Conference on*. IEEE, 2163–2170.

Jaesik Park, Sudipta Sinha, Yasuyuki Matsushita, Yu Wing Tai, and In So Kweon. 2016. Robust Multiview Photometric Stereo using Planar Mesh Parameterization. *IEEE Transactions on Pattern Analysis and Machine Intelligence* (2016).

Peiran Ren, Jiaping Wang, John Snyder, Xin Tong, and Baining Guo. 2011. Pocket reflectometry. *ACM Transactions on Graphics (TOG)* 30, 4 (2011), 45:1–10.

Jérémy Riviere, Pieter Peers, and Abhijeet Ghosh. 2015. Mobile Surface Reflectometry. *Computer Graphics Forum* (2015).

Jeremy Riviere, Ilya Reshetouski, Luka Filipi, and Abhijeet Ghosh. 2017. Polarization imaging reflectometry in the wild. *ACM Transactions on Graphics (TOG)* 36, 6 (2017), 206.

Szymon Rusinkiewicz. 1998. A New Change of Variables for Efficient BRDF Representation. In *Rendering Techniques '98 (Proceedings of Eurographics Rendering Workshop '98)*, G. Drettakis and N. Max (Eds.). Springer Wien, 11–22.

Johannes L. Schönberger. 2016. COLMAP. https://colmap.github.io.

Johannes L Schönberger, Enliang Zheng, Jan-Michael Frahm, and Marc Pollefeys. 2016. Pixelwise view selection for unstructured multi-view stereo. In *European Conference on Computer Vision*. Springer, 501–518.

Thomas Schöps, Torsten Sattler, Christian Häne, and Marc Pollefeys. 2015. 3d modeling on the go: Interactive 3d reconstruction of large-scale scenes on mobile devices. In *3D Vision (3DV), 2015 International Conference on*. IEEE, 291–299.

Christopher Schwartz, Ralf Sarlette, Michael Weinmann, and Reinhard Klein. 2013. Dome ii: A parallelized btf acquisition system. In *Proceedings of the Eurographics 2013 Workshop on Material Appearance Modeling: Issues and Acquisition*. Eurographics Association, 25–31.

Niranjan Thanikachalam, Loïc Baboulaz, Damien Firmenich, Sabine Süsstrunk, and Martin Vetterli. 2017. Handheld Reflectance Acquisition of Paintings. *IEEE Transactions on Computational Imaging* 3, 4 (2017), 580–591.

Borom Tunwattanapong, Graham Fyffe, Paul Graham, Jay Busch, Xueming Yu, Abhijeet Ghosh, and Paul Debevec. 2013. Acquiring Reflectance and Shape from Continuous Spherical Harmonic Illumination. *ACM Transactions on Graphics (TOG)* 32, 4 (2013), 109:1–12.

Bruce Walter, Stephen R Marschner, Hongsong Li, and Kenneth E Torrance. 2007. Microfacet models for refraction through rough surfaces. In *Proc. Eurographics*. 195–206.

Jiaping Wang, Shuang Zhao, Xin Tong, John Snyder, and Baining Guo. 2008. Modeling anisotropic surface reflectance with example-based microfacet synthesis. *ACM Transactions on Graphics (TOG)* 27, 3 (2008), 41:1–10.

Michael Weinmann and Reinhard Klein. 2015. Advances in Geometry and Reflectance Acquisition (Course Notes). In *SIGGRAPH Asia 2015 Courses*. ACM, Article 1, 71 pages.

Tim Weyrich, Jason Lawrence, Hendrik P.A. Lensch, Szymon Rusinkiewicz, and Todd Zickler. 2009. Principles of Appearance Acquisition and Representation. *Foundations and Trends in Computer Graphics and Vision* 4, 2 (2009), 75–191.

Jae Hyun Won, Man Hee Lee, and In Kyu Park. 2012. Active 3D shape acquisition using smartphones. In *Computer Vision and Pattern Recognition Workshops (CVPRW), 2012 IEEE Computer Society Conference on*. IEEE, 29–34.

Hongzhi Wu, Zhaotian Wang, and Kun Zhou. 2016. Simultaneous localization and appearance estimation with a consumer RGB-D camera. *IEEE transactions on visualization and computer graphics* 22, 8 (2016), 2012–2023.

Hongzhi Wu and Kun Zhou. 2015. Appfusion: Interactive appearance acquisition using a kinect sensor. In *Computer Graphics Forum*, Vol. 34. Wiley Online Library, 289–298.

Rui Xia, Yue Dong, Pieter Peers, and Xin Tong. 2016. Recovering Shape and Spatially-varying Surface Reflectance Under Unknown Illumination. *ACM Transactions on Graphics (TOG)* 35, 6, Article 187 (2016), 12 pages.

Z. Y. Zhang. 2000. A Flexible New Technique for Camera Calibration. *IEEE Trans. Pattern Analysis and Machine Intelligence* 22, 11 (Nov. 2000), 1330–1334.

Zhiming Zhou, Guojun Chen, Yue Dong, David Wipf, Yong Yu, John Snyder, and Xin Tong. 2016. Sparse-as-possible SVBRDF acquisition. *ACM Transactions on Graphics (TOG)* 35, 6 (2016), 189.

Zhenglong Zhou, Zhe Wu, and Ping Tan. 2013. Multi-view Photometric Stereo with Spatially Varying Isotropic Materials. In *Proc. IEEE Conf. Computer Vision and Pattern Recognition (CVPR) 2013*. 1482–1489.

Michael Zollhöfer, Angela Dai, Matthias Innmann, Chenglei Wu, Marc Stamminger, Christian Theobalt, and Matthias Nießner. 2015. Shading-based refinement on volumetric signed distance functions. *ACM Transactions on Graphics (TOG)* 34, 4 (2015), 96.