Appearance-from-Motion: Recovering Spatially Varying Surface Reflectance under Unknown Lighting

Yue Dong[1] Guojun Chen[2] Pieter Peers[3] Jiawan Zhang[2] Xin Tong[1,2]
[1]Microsoft Research [2]Tianjin University [3]College of William & Mary

Figure 1: ***Appearance-from-Motion***. *Renderings under the "Uffizi Gallery" light probe of surface reflectance recovered under unknown illumination. (a) "beer bottle", (b) "coffee mug", (c) "rusted copper", and (d) "Starbucks can".*

Abstract

We present "appearance-from-motion", a novel method for recovering the spatially varying isotropic surface reflectance from a video of a rotating subject, with known geometry, under unknown natural illumination. We formulate the appearance recovery as an iterative process that alternates between estimating surface reflectance and estimating incident lighting. We characterize the surface reflectance by a data-driven microfacet model, and recover the microfacet normal distribution for each surface point separately from temporal changes in the observed radiance. To regularize the recovery of the incident lighting, we rely on the observation that natural lighting is sparse in the gradient domain. Furthermore, we exploit the sparsity of strong edges in the incident lighting to improve the robustness of the surface reflectance estimation. We demonstrate robust recovery of spatially varying isotropic reflectance from captured video as well as an internet video sequence for a wide variety of materials and natural lighting conditions.

CR Categories: I.3.7 [Computer Graphics]: Three-Dimensional Graphics and Realism—Rendering

Keywords: Appearance Modeling, Unknown Lighting

Links: DL PDF

1 Introduction

Accurately reproducing the rich and varied appearance of real-world materials is critical for photo-realistic rendering. Some of the most successful appearance modeling techniques today are measurement-based, where observations of a material's appearance are used to drive or fit a digital appearance model suitable for photo-realistic rendering applications. Despite recent advances in making acquisition techniques and acquisition devices more accurate, more robust, and more easily accessible, capturing high-quality surface reflectance in general environments remains a challenging task.

The surface reflectance of an opaque object is formally described by the spatially varying bidirectional reflectance distribution function (svBRDF) that relates the incident irradiance to outgoing radiance at every surface point. The appearance at each surface point is the inner product between its BRDF and the incident lighting at that point. Many existing appearance modeling methods probe the svBRDF by carefully controlling the incident lighting (e.g., moving a point light or directional light source [Holroyd et al. 2010; Lensch et al. 2003], sweeping a linear light source [Gardner et al. 2003], or using environmental structured illumination [Aittala et al. 2013; Tunwattanapong et al. 2013; Wang et al. 2011]). Unfortunately, the use of active illumination also implies full control over the incident lighting, thus binding these methods to laboratory-like settings. Recently, researchers have started to take appearance acquisition outside the lab and under uncontrolled lighting. These methods typically require a calibration object in the scene to infer the lighting or material properties [Hertzmann and Seitz 2003; Ren et al. 2011], or are limited to Lambertian [Shan et al. 2013] or homogeneous objects [Romeiro and Zickler 2010] only.

In this paper we introduce "appearance-from-motion", a novel method for recovering the *spatially varying* isotropic surface reflectance under unknown natural illumination from a video sequence, and registered geometry, of a rotating object. We model the surface reflectance by a data-driven microfacet model with a 1D isotropic microfacet distribution, and formulate appearance recovery as an iterative estimation procedure that alternates between estimating surface reflectance and estimating incident lighting. We solve for the surface reflectance on temporal gradients of the obser-

vations of each surface point. Surface reflectance recovery in the temporal gradient domain has the advantage that the influence of diffuse surface reflectance is mostly mitigated, allowing us to resolve the microfacet normal distribution in isolation from diffuse lighting effects. A key observation is that natural incident lighting is sparse in the (angular) gradient domain, with few dominant edges (e.g., at the skyline, around windows or light sources, etc.). By directing the computations toward the effects of these salient edges, we can regularize the otherwise ill-conditioned recovery of the surface reflectance. In addition, we also rely on the sparsity of the incident illumination gradients to break the inherent ambiguity in the joint recovery of lighting and reflectance (i.e., a mirror-like material under blurred lighting is indistinguishable from a rough specular material under appropriate high frequency lighting).

The proposed appearance-from-motion framework greatly simplifies acquisition, making data-driven appearance modeling more accessible to non-expert users. Acquisition is as simple as capturing a video while rotating the object, either by using a turntable or simply by hand. Our free-form acquisition is perfectly suited for in-situ acquisition and is not bound by laboratory-like conditions. We demonstrate the effectiveness of our method on several captured datasets and illustrate the robustness of our method by recovering surface reflectance from an internet video sequence.

2 Related Work

There exists a rich body of prior work on appearance acquisition. The majority of these rely on active illumination to probe the surface reflectance of a subject (e.g., [Aittala et al. 2013; Holroyd et al. 2010; Gardner et al. 2003; Tunwattanapong et al. 2013]). While accurate and robust, active illumination methods are limited to acquisition environments with full control over the incident lighting. Passive appearance acquisition methods, on the other hand, infer surface reflectance from the unknown lighting present during acquisition. We focus this overview of related work on passive appearance acquisition methods – a comprehensive survey on active appearance modeling techniques can be found in [Dorsey et al. 2008].

The first class of passive appearance recovery methods place one or more reference samples with known BRDF in the scene. Hertzmann and Seitz [2003] recover surface normals and reflectance by capturing multiple homogeneous reference objects, and model the surface reflectance at each surface point on the target subject as a linear combination of the reference materials. Treuille et al. [2004] extend this method to multi-view data. Recently, Ren et al. [2011] employed a BRDF checker card of 24 reference materials to record the light response for a handheld linear light source waved over the subject, and reconstruct the BRDF at each surface point as a linear combination of the 24 reference materials. A disadvantage shared by these methods is that a reference sample needs to be placed in the scene. Furthermore, the recovered surface reflectance is limited to the appearance subspace spanned by the reference materials. In contrast, our appearance-from-motion method does not require a reference sample, and it employs a more flexible data-driven reflectance model.

A second class of passive appearance modeling techniques employs inverse rendering to jointly recover surface reflectance and lighting from a single/few viewpoint observations [Marschner 1998]. An extensive mathematical framework for inverse rendering based on spherical harmonics was proposed by Ramamoorthi and Hanrahan [2001]. Using this framework, they demonstrate accurate recovery of lighting and surface reflectance from a small set of views of a homogeneous object. Romeiro et al. [2008] introduce passive reflectometry for inferring surface reflectance from a single image of a homogeneous sphere with *known* natural incident lighting. Romeiro and Zickler [2010] further extend this framework to handle unknown lighting. They do not explicitly reconstruct the incident lighting, but instead rely on the statistics of real-world illumination to constrain the recovery of the surface reflectance. Lombardi and Nishino [2012] estimate both lighting and homogeneous surface reflectance from an object with known shape. They model the surface reflectance as a log-linear combination of data-driven basis functions, and employ a heavy tail prior and low entropy prior to constrain the recovery of the incident lighting. All of these methods exploit the rich variation in surface normals to infer *homogeneous* surface reflectance. An exception is the work by Barron and Malik [2013] that infers shape, spatially varying reflectance and lighting from a single photograph. However, the spatially varying reflectance is restricted to diffuse reflectance with spatially varying albedo only. In contrast, our method relies on a dense *temporal* sampling to estimate both diffuse and specular isotropic *spatially varying* surface reflectance.

A third class of passive methods infers spatially varying surface reflectance from multi-view observations of a subject assuming some underlying reflectance representation. Nishino et al. [2001] model surface reflectance from a sparse set of views under static unknown lighting using a diffuse texture and a homogeneous parametric specular component. Yu et al. [2006] use tensor factorization to reconstruct the svBRDF and illumination from a sparse set of images under static, unknown lighting. They model the incident lighting using a low-frequency Spherical Harmonics representation, and represent the svBRDF by a texture-modulated base material. Li et al. [2013] recover geometry, reflectance, and lighting from a human performance. They model surface reflectance by a per-surface point diffuse albedo, and a per-region specular component represented by the Phong BRDF. Haber et al. [2009] reconstruct the appearance of a single object from internet photo collections. They model incident lighting using an all-frequency wavelet representation, and represent the BRDF as a linear combination of 10 basis materials. Shan et al. [2013] reconstruct a simplified model (i.e., diffuse reflectance and uniform lighting with a single directional light source) to estimate the appearance of very large-scale scenes from internet image collections. All these methods either ignore specular surface reflectance, or assume a simplified model. In contrast, the proposed method uses a flexible data-driven specular surface reflectance model that can accurately model a wide range of isotropic materials.

3 Assumptions

Recovering spatially surface reflectance under unknown incident lighting is a difficult and ill-conditioned problem. A single observation of a surface point under unknown lighting provides insufficient constraints to recover both the incident lighting as well as detailed surface reflectance. Prior work on homogeneous reflectance recovery under unknown lighting exploits observations from multiple surface points to reconstruct both the lighting and surface reflectance. However, in the case of spatially varying surface reflectance each surface can potentially consist of a different material. To overcome this, we propose to reconstruct the surface reflectance for each surface point separately from multiple observations from different viewpoints. In particular, we rotate the subject under fixed but unknown lighting, and record a video sequence from a fixed vantage point. Similarly to prior work, we make a number of assumptions to make robust and practical recovery possible:

Geometry: We rely on the availability of registered geometry to track the motion of surface points and record respective changes in the observed reflectance for each surface point in the video sequence. While a seemingly restrictive requirement, it should be

noted that the majority of appearance acquisition methods (even under controlled lighting) also assume known geometry or employ a separate acquisition phase to recover the geometry (e.g., svBRDF acquisition methods assume a planar surface; single view methods require known normals; etc.).

Lighting: Similarly to the majority of prior work on reflectance recovery, we assume incident lighting is distant (i.e., only has a directional dependence), and is on average color-neutral (i.e., gray world assumption). We ignore interreflections and self-occlusions, and assume that direct lighting is the dominant factor. Finally, we assume the lighting does not vary for the duration of the acquisition.

We store the incident lighting in the "cross" or "cube map" parameterization at a resolution of $6 \times 128 \times 128$, and ensure that any operation is correctly performed on the spherical domain (e.g., a spherical blur would result in a "spatially-varying" kernel in the "cross" parameterization).

Surface Reflectance: We assume that the surface reflectance is isotropic, and that it can be accurately characterized by a microfacet reflectance model. Formally, the surface reflectance at a point x is described by the Bidirectional Reflectance Distribution Function (BRDF) [Nicodemus et al. 1977], a 4D function that relates incident irradiance to outgoing radiance. A commonly used approximation to general surface reflectance is the dichromatic BRDF:

$$f_r(\omega_i, \omega_o; x) = \frac{\rho_d(x)}{\pi} + \rho_s(x) f_s(\omega_i, \omega_o; x), \tag{1}$$

where ω_i and ω_o are the incident and outgoing directions respectively, ρ_d and ρ_s are the total diffuse and specular reflectivity or albedo, and f_s is the specular surface reflectance. We model the specular surface reflectance using the generalized (isotropic) microfacet model [Ashikhmin et al. 2000]:

$$f_s(\omega_i, \omega_o; x) = \frac{D(\omega_h; x) G(\omega_i, \omega_o; x) F(\omega_i, \omega_o)}{4(\omega_i \cdot n(x))(\omega_o \cdot n(x))}, \tag{2}$$

where $D(\omega_h)$ is the microfacet normal distribution function (NDF) of the halfway direction ω_h, $F(\omega_i, \omega_o)$ is the Fresnel reflectance (we assume a fixed index of refraction of 1.3 for all materials), and $G(\omega_i, \omega_o)$ is the shadowing and masking term.

We store the NDF as a 1D tabulated function that monotonically decreases with $(\omega_h \cdot n)$ at a resolution of 32 samples uniformly distributed in $[0, \Lambda]$ (where Λ is the size of the footprint of the NDF, and $\Lambda \leq 0.5\pi$), and linearly interpolate in between. We will adaptively choose Λ to maximize the sample rate without cutting off the NDF. Furthermore, we follow Ashihkmin et al. [2000] and compute the shadowing and masking term directly from the NDF as:

$$G(\omega_i, \omega_i; x) = G'(\omega_i, \omega_h; x) G'(\omega_o, \omega_h; x), \tag{3}$$

with

$$G'(\omega; x) = \frac{(\omega \cdot n(x))}{\int (\omega_h \cdot \omega) D(\omega_h; x) d\omega_h}. \tag{4}$$

While other surface reflectance representations are possible (e.g., analytic models or spherical harmonics based representations), the proposed data-driven microfacet representation has several advantages: it offers a fair degree of freedom, it exhibits a wide frequency response, and it naturally enforces a physically plausible form without overly constraining the results. However, this comes at the cost of requiring a non-linear fitting procedure.

4 Appearance-from-Motion

The goal of our appearance-from-motion framework is to recover, for every surface point x, the surface reflectance f_r defined by diffuse albedo $\rho_d(x)$, specular albedo $\rho_s(x)$, and NDF $D(\omega_h; x)$, that best match the observations $I(\omega_o'; x, t)$ (a video of an rotating object) under unknown natural lighting $E(\omega_i')$:

$$\underset{\{\rho_d, \rho_s, D\}_x, E(\omega_i')}{\operatorname{argmin}} \sum_t \sum_x ||I(\omega_o'; x, t) - L(\omega_o'; x, t)||^2, \tag{5}$$

where $L(\omega_o'; x, t)$ is the outgoing radiance at a surface point x at time t:

$$L(\omega_o'; x, t) = \int_\Omega f_r(\omega_i(x,t), \omega_o(x,t); x) E(\omega_i')(n(x,t) \cdot \omega_i') d\omega_i'. \tag{6}$$

We denote directions in the *global* coordinate frame by the prime symbol, and (time-varying) directions in the *local* frame of a point x defined by its surface normal $n(x,t)$ at time t. Both are related via a rotation defined by the surface normal: $\omega' = R_{n(x,t)}(\omega(x,t))$.

To recover the surface reflectance, we also need to recover the incident lighting $E(\omega_i')$ in order to evaluate Equation (6). A fundamental ambiguity in the joint recovery of lighting and reflectance is that (angular) sharpness in BRDF can be traded-off for blurriness in the lighting while matching the observations. Our solution to conquering this ambiguity is based on two key observations. First, we note that our prime concern is to accurately recover the spatially varying surface reflectance, and the recovery of the incident lighting is incidental. Second, we observe that natural incident lighting is sparse in the gradient domain, and that the lighting has few dominant discontinuities. The effects of such dominant discontinuities in the incident lighting are easier to detect in the observations, and thus provide a more stable cue to robustly recover the surface reflectance. We will therefore employ a sparse gradient prior in the recovery of the lighting, effectively placing greater emphasis on the recovery of dominant discontinuities, and focus the recovery of the surface reflectance around these discontinuities.

The minimization in Equation (5) is non-convex in the unknowns. To facilitate the computation of this minimization, we iteratively alternate between estimating the normal distribution $D(\omega_h; x)$ (Section 4.1), diffuse and specular albedo $\rho_d(x)$ and $\rho_s(x)$ respectively (Section 4.2), and the incident lighting $E(\omega_i')$) (Section 4.3) while keeping the others fixed until convergence.

4.1 NDF Recovery

Given an estimate of the incident lighting, we can recover each surface point's BRDF separately by considering slices through the observations $I(\omega_o'; x, t)$ with constant surface location x: $T_x(\omega_o', t) = I(\omega_o'; x, t)$. Such a slice forms a 1D *temporal trace* that records the temporal variation in the appearance of a surface point x, and thus the effects of the incident lighting on a single BRDF. The recovery of a surface point's BRDF can now be formulated as:

$$\underset{\rho_{d_x}, \rho_{s_x}, D_x}{\operatorname{argmin}} \sum_t ||T_x(\omega_o', t) - L_x(\omega_o'; t)||^2. \tag{7}$$

Note, we use the subscript x to denote that the surface location is fixed, and that the minimization in Equation (7) is performed for each surface point *separately*.

The accuracy of the result from the minimization in Equation (7) greatly depends on the accuracy of the provided incident lighting and the signal-to-noise ratio of the observations. To improve the robustness of the recovery of the surface reflectance, we focus on matching *changes* in the appearance due to salient discontinuities in

the lighting. To achieve this goal, we reformulate the minimization in Equation (7) in terms of the *temporal* gradient, and employ a robust detection method to locate the discontinuities in the lighting.

Gradient Domain Recovery The temporal gradient of the 1D trace can be expressed as:

$$\begin{aligned}\nabla_t L_x(\omega_o';t) &= \nabla_t \int_\Omega f_r(\omega_{ix}(t),\omega_{ox}(t))E(\omega_i')(n_x(t)\cdot\omega_i')d\omega_i',\\ &= \nabla_t \int_\Omega \frac{\rho_{dx}}{\pi}E(\omega_i')(n_x(t)\cdot\omega_i')d\omega_i'\\ &+ \nabla_t \int_\Omega \rho_{sx} f_{sx}(\omega_{ix}(t),\omega_{ox}(t))E(\omega_i')(n_x(t)\cdot\omega_i')d\omega_i',\\ &= \frac{\rho_{dx}}{\pi}\int_\Omega E(\omega_i')\nabla_t(n_x(t)\cdot\omega_i')d\omega_i',\\ &+ \rho_{sx}\int_\Omega \nabla_t\left(f_{sx}(\omega_i,\omega_{ox}(t))E(R_{n_x(t)}(\omega_i))\omega_{iz}\right)d\omega_i.\end{aligned} \tag{8}$$

In the last term, the incident lighting is expressed in the local frame at surface point x. Observe that while the incident lighting $E(\omega_i')$ expressed in the global frame is temporally invariant, the incident lighting $E(R_{n(x,t)}(\omega_i))$ expressed in the local frame is not.

We further simplify this expression by observing that the gradient of the diffuse component is approximately zero ($\nabla_t(n_x(t))\cdot\omega_i') \approx 0$), when the time granularity is small compared to the rotation. Thus the effects of the diffuse component on the temporal gradient are negligible. Similarly, for small time-steps, changes in ω_o are negligible; a surface point generally moves less than a pixel per frame. Hence, the effects of $f_s(\omega_i,\omega_o)$ on the gradient are approximately constant (i.e., the gradient of the lighting is the dominant factor):

$$\nabla_t L_x(\omega_o';t) \approx \rho_s(x)\int_\Omega f_s(\omega_i,\omega_{ox}(t))\omega_{iz}\nabla_t E(R_{n_x(t)}(\omega_i))d\omega_i. \tag{9}$$

Following these approximations, we can then directly recover the shape of the NDF from the temporal gradient of the 1D trace and from the temporal gradient of the incident lighting expressed in the local coordinate frame:

$$\underset{\overline{D}_x}{\operatorname{argmin}}\sum_t ||\nabla_t T_x(\omega_o';t) - \int_\Omega f_{sx}(\omega_i,\omega_{ox}(t))\omega_{iz}\nabla_t E(R_{n_x(t)}(\omega_i))d\omega_i||^2, \tag{10}$$

where $\overline{D}_x$ is the unnormalized NDF that includes the specular albedo ρ_{sx}.

Robust Discontinuity Detection Given the gradient domain minimization formulation (Equation (10)), we now want to further regularize this minimization by only considering the dominant discontinuities in the incident lighting. Due to the bandwidth limiting behavior of surface reflectance, the recovered incident lighting is likely more blurry than the ground truth incident lighting. Consequently, simply thresholding the temporal gradients of the estimated incident lighting to extract the dominant discontinuities is unlikely to provide robust and reliable results. Instead, we take inspiration from motion deblurring and follow a strategy similar to that of Xu and Jia [2010]. First we apply a spherical Gaussian blur (with a kernel size of $0.01sr$) to the incident lighting to remove any noise and ringing artifacts in the estimated lighting. Next, we sharpen the edges along local *angular* gradients (denoted by $\nabla_\omega E$) using a shock filter [Osher and Rudin 1990], resulting in an enhanced incident lighting $E_s(\omega_i)$. The shock filter is computed by repeatedly applying a morphological operator based on the sign of the Laplacian ΔE:

$$E_{i+1} = E_i - \text{sign}(\Delta E_i)||\nabla_\omega E_i||di, \tag{11}$$

(a) Measurement (b) After shock filter (c) Selected edge (c) Ground truth lighting

Figure 2: ***Robust Detection of Dominant Discontinuities****. Given a possibly blurred estimate of the incident lighting (a), our method applies a shock filter (b) to detect dominant edges (highlighted in blue) (c) to aid in the recovery of surface reflectance. The detected edges are similar to those found in the ground truth incident lighting (d).*

until convergence (i.e., $(E_{i+1}-E_i)/E_i < 0.01$). Finally, we compute the temporal gradient on $E_s(\omega_i)$ (by finite differencing) and only keep dominant edges:

$$\nabla_t E(\omega_i) = \begin{cases}\nabla_t E_s(\omega_i) & r(\omega_i) > \tau_r\\ 0 & otherwise,\end{cases} \tag{12}$$

where $r(\omega_i)$ is defined as:

$$r(\omega_i) = \frac{||\sum_{\omega\in\mathcal{N}(\omega_i)}\nabla_t E_s(\omega)||}{\sum_{\omega\in\mathcal{N}(\omega_i)}||\nabla_t E_s(\omega_i)|| + 0.5}, \tag{13}$$

and where $\mathcal{N}(\omega_i)$ is the footprint of the NDF D_x from the previous iteration, i.e., a circle with radius Λ centered around ω_i. τ_r is initially set to $0.1\max(r(\omega_i))$, and gradually decreased every iteration by 10% [Cho and Lee 2009]. Figure 2 illustrates this process of detecting dominant edges.

We can now solve Equation (10) using a non-negative least squares minimization to ensure non-negative probabilities in the NDF. Furthermore, we constrain the NDF to be monotonic to further regularize the recovery. Finally, the NDF D_x can then be recovered from $\overline{D}_x$ via normalization (i.e., unit integration).

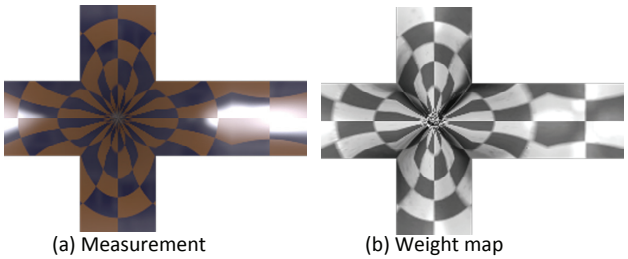
4.2 Albedo Recovery

Given the incident lighting $E(\omega_i')$ and the recovered NDF D_x, we would like to recover both the diffuse and specular albedo. For this we express the 1D trace $T_x(\omega_o',t)$ as the weighted sum of a *normalized* diffuse trace $T_{dx}(\omega_o',t)$ and *normalized* specular trace $T_{sx}(\omega_o',t)$:

$$\begin{aligned}T_x(\omega_o',t) &= \int_\Omega(\frac{\rho_{dx}}{\pi}+\rho_{sx}f_s(\omega_i,\omega_o))E(\omega_i')(n_x(t)\cdot\omega_i')d\omega_i',\\ &= \frac{\rho_{dx}}{\pi}\int_\Omega E(\omega_i')(n_x(t)\cdot\omega_i')d\omega_i'\\ &+ \rho_{sx}\int_\Omega f_s(\omega_i,\omega_o)E(\omega_i')(n_x(t)\cdot\omega_i')d\omega_i',\\ &= \rho_{dx}T_{dx}(\omega_o',t)+\rho_{sx}T_{sx}(\omega_o',t).\end{aligned} \tag{14}$$

$T_{dx}(\omega_o',t)$ and $T_{sx}(\omega_o',t)$ are fully determined by known components. As a result, the recovery of the diffuse albedo ρ_{dx} and specular albedo ρ_{sx} can now be formulated as a robust non-negative linear least squares minimization:

$$\underset{\rho_{sx},\rho_{dx}}{\operatorname{argmin}}||T_x(\omega_o',t) - \left(\rho_{dx}T_{dx}(\omega_o',t)+\rho_{sx}T_{sx}(\omega_o',t)\right)||^2 \quad \text{s.t. } \rho_{dx},\rho_{sx}\geq 0 \tag{15}$$



(a) Measurement (b) Weight map

Figure 3: ***Confidence Weighting*** *for the recovery of the incident lighting based on the accuracy of the recovered BRDF and on the bandpass behavior of the BRDF. (a) the observed reflectance from a synthetic checkerboard svBRDF with specular blue tiles and diffuse yellow tiles, illuminated by the* Uffizi *light probe. (b) the resulting confidence weighting that places higher weight on regions that provide more information on the incident illumination.*

4.3 Recovering Incident Lighting

As noted before, a key problem in the joint recovery of incident lighting and surface reflectance is the fundamental ambiguity that sharpness of the BRDF can be traded-off for blurriness in the incident lighting and vise versa. We break this ambiguity, by exploiting the observation that natural incident lighting is sparse in the gradient domain. Similar to Levin et al. [2007] we include a $||\nabla_\omega E(\omega_i')||^{0.8}$ sparsity prior to regularize the recovery of incident lighting:

$$\underset{E(\omega_i')}{\operatorname{argmin}} \sum_t \sum_x w(x)||I(\omega_o';x,t) - L(\omega_o';x,t)||^2 + \lambda||\nabla_\omega E(\omega_i')||^{0.8}, \tag{16}$$

where $\lambda = 0.01$, and the weighting function is the product of three terms: $w(x) = w_c(x)w_d(x)w_s(x)$, where each terms encodes:

- **Confidence in the Recovered BRDF at x.** The stronger the salient edges in the incident lighting over the path of the 1D temporal traces, the more confidence we have in the recovery of the BRDF. We estimate confidence by the total temporal gradient across the (shock-filtered) incident lighting recovered in the previous iteration: $w_c(x) = \sum_t |\nabla_t E_s(\omega_i(x,t))|$.
- **Bandpass Behavior of the BRDF**: All BRDFs act as a bandpass filter over the incident lighting. To improve robustness, we give lower weight to BRDFs that aggressively filter the incident lighting: $w_d(x) = \frac{1}{\sigma}$, where σ is the standard deviation of the NDF.
- **Specular Signal-to-Noise Ratio**: A strong diffuse reflection can mask the effects of the specular reflections, aversely affecting the robustness of the estimate. We therefore give high weight to surface points with a high *relative* specular albedo: $w_s(x) = \frac{\rho_s}{\rho_s+\rho_d}$.

Figure 3 illustrates how the weighting function $w(x)$, for a synthetic svBRDF with dark blue specular tiles and bright yellow diffuse tiles on a spherical object, favors surface points that provide more reliable information on the incident illumination (i.e., a high weight is assigned to the dark blue specular tiles, and conversely, a low weight is given to the bright yellow diffuse tiles).

We solve Equation (16) using a coarse-to-fine dyadic multi-resolution method, and apply at each scale a modified version of the stochastic random walk algorithm of Gregson et al. [2013]. The key differences are that our method uses a cost function over multiple observations (i.e., video-frames) instead of the standard deconvolution cost function over a single image, and that the recovered incident lighting is a spherical function as opposed to an image. While originally intended for image deconvolution, we found that stochastic deconvolution [Gregson et al. 2013] is especially well-suited for our problem as it tends to focus on areas of high gradient. Algorithm 1 lists the relevant parameters and summarizes the recovery algorithm of incident lighting at a single scale.

input : Observations $I(\omega_o';x,t)$, estimated BRDF $f_r(\omega_i,\omega_o;x)$, geometry $n(x,t)$, number of iterations $N = 100$ and $K = TotalPixelNumber$, reset probability $\Xi = 0.0005$, mutation strength $\sigma = \Lambda/2$ (Λ is the footprint size of the NDF), minimum acceptance rate $A = 0.4$, lighting quantum $\psi = 0.05$, quantum modifier $\gamma = 0.5$
output: Incident Lighting $E(\omega_i')$

```
initialization: E = I;
for k=1 to N do
    α = 0;
    ω_c = random direction;
    Δε_c = change in error (Eq. 16) if one adds (subtracts) ψ to E(ω_c);
    for j=1 to K do
        // Random Walk
        if (random() ≥ Ξ) then
            mutate: ω_m = sample normal (mean=ω_c, std.dev.=σ);
        else reset: ω_m = random direction;
        ;
        Δε_m = change in error (Eq. 16) if one adds (subtracts) ψ to E(ω_m);
        if (Δε_m improvement) then
            α ← α + 1/K;
            permanently apply change to E(ω_m) (add/subtract ψ)
        // Move center if ω_m is less worse than ω_c,
           otherwise random accept
        if (Δε_c ≤ 0 and Δε_m ≥ Δε_c) or (random() < Δε_m/Δε_c) then
            Δε_c = Δε_m;
            ω_c = ω_m;
    // Reduce ψ if acceptance rate is too low
    if (α < A) then ψ ← ψ * γ;
    ;
```

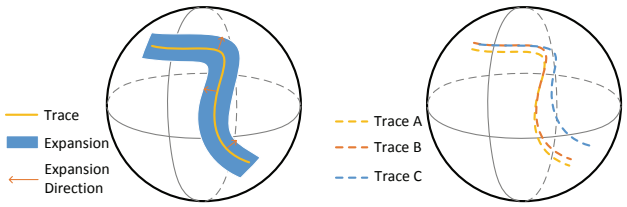
Algorithm 1: Incident Lighting Recovery. Summary of parameters and the algorithm for recovering incident lighting. This algorithm is further embedded in a coarse-to-fine dyadic multi-resolution scheme.

We further improve performance by removing the effects of diffuse reflectance (computed and up-scaled from the prior scale) from the observations, as the diffuse reflectance contributes little to the recovery of the incident lighting due to its low-pass behavior. We halt the multi-resolution refinement when the bandwidth of the incident lighting exceeds that of the BRDFs. Instead of using the bandwidth of the recovered BRDFs directly, we use a cut-off frequency that maintains 95% of the energy (in lower frequencies) of the NDFs to minimize the impact of noise.

4.4 Initialization

Our reflectance recovery algorithm iteratively alternates between solving for the NDFs (Section 4.1), diffuse and specular albedo (Section 4.2), and for lighting (Section 4.3), while keeping the other components fixed. To bootstrap this iterative process, we follow the process outlined below:

NDF D_x Of the three aforementioned stages, estimation of the NDF is most robust – thresholding on the shock filtered temporal gradient is robust to some degree of blurring in the incident lighting. Furthermore, the temporal trace $T_x(\omega_o',t)$ represents a 1D slice of the blurred incident lighting (filtered by the BRDF). We therefore propose to utilize the 1D trace directly as initial incident lighting to recover the NDF. However, the trace only provides a 1D slice,



(a) Expanding the incident lighting (b) Intersect trace region

Figure 4: ***Initialization****. (a) we expand the temporal trace to create a rough approximation of the incident lighting for each surface point. (b) the specular surface albedos of all surface points are brought to the same global scale based on the relative ratio of the intersecting points of all pair-wise combinations of temporal traces.*

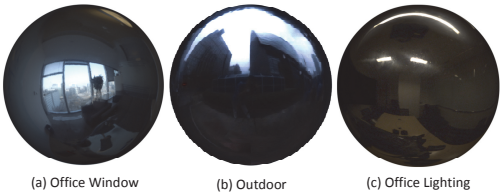
whereas incident lighting is a 2D function. To circumvent this problem, we hallucinate a rough approximation of the incident lighting by copying the closest point on the trace (i.e., an expansion of the trace in a direction orthogonal to the trace direction; see Figure 4 (a)). Note that each surface point has its own trace, and hence we create a different incident lighting approximation per surface point for the initial step – subsequent iterations use a single incident lighting estimate for all surface points. Given this rough approximation, we can then compute the NDF using the minimization in Equation (10).

Initially we allow the NDF to cover the full range (i.e., $\Lambda = 0.5\pi$). In subsequent iterations, we speed up computations by limiting the range of the NDF to contain 95% of the energy of the initial estimated NDF. We assume the values outside this range are due to negligible noise, and set the NDF to zero outside this range.

Shadowing & Masking We initialize the shadowing and masking term $G(\omega_i, \omega_o)$ to 1, and update it after each computation of a new NDF (Equations (3) and (4)). Consequently, the shadowing and masking term is based on the previous iteration's NDF. However, this is still a good approximation due to the low frequency nature of the shadowing and masking term.

Specular Albedo ρ_s While using the 1D trace directly as a rough approximation of the incident lighting suffices to recover an initial NDF, it does not help in the estimation of the specular albedo. The gradient of the shock-filtered 1D trace is proportional to the ground truth gradient *times the specular albedo*. Hence, naively applying Equation (15) will always produce a specular albedo ≈ 1 (since the specular albedo is present in both the trace and the approximation of incident lighting). However, we observe that for 1D traces that intersect, we can compute their relative ratio, and thus the relative ratio of their respective specular albedos (Figure 4 (b)). By performing a global optimization on all pairs, we can determine the relative specular albedos up to a global scale factor. This global scale factor is an inherent ambiguity shared by all methods that jointly recover BRDFs and lighting. We keep this global scale factor consistent over the iterations by normalizing the incident lighting for each iteration – without observations of a known reference material, we cannot resolve for this global scale exactly.

Diffuse Albedo ρ_d We initialize the diffuse albedo to zero. Consequently, when recovering the incident lighting (Section 4.3), the diffuse component gets baked into the lighting such that the specular reflection mimics the diffuse and specular reflection under the real incident lighting. We therefore add a correction step after the first iteration that subtracts the maximal diffuse reflectance from the



(a) Office Window (b) Outdoor (c) Office Lighting

Figure 5: ***Incident Lighting.*** *The three lighting conditions under which (a) the* "beer bottle" *(Office Window), (b) the* "coffee mug" *(Outdoor), and (c) the* "rusted copper" *and* "toy duck" *(Office Lighting) were captured.*

incident lighting:

$$E'(\omega_i) = E(\omega_i) - \frac{\rho_d{}'}{\pi} \int E(\omega)(\omega_i \cdot \omega) d\omega, \qquad (17)$$

where $\rho_d{}'$ is set such that $\min E'(\omega_i) = 0$. While this is likely an overestimation, we found that the diffuse and specular albedo converge to the ground truth after a few iterations.

5 Results

Acquisition & Calibration The proposed appearance-from-motion method only requires a video sequence of a rotating target subject, and registered geometry of the subject for each frame in the video. Our method estimates the surface reflectance at every point on the geometry *separately*; we make no assumptions on the underlying geometry or on the spatial distribution (e.g., smoothness) of the surface reflectance or texture.

Unless noted, the results in this paper are computed from video sequences captured with a Canon EOS 5D Mark II equipped with an EF 70-200 F4L lens. We directly use single-exposure radiometrically linear RAW images, and calibrate the intrinsic camera parameters using the method of Zhang et al. [2000]. In addition, we obtain the subject's shape using an Artec 3D scanner, or by hand-modeling it for very simple scenes. The subject is rotated using a Directed Perception Pan-Tilt Unit-D46. Care was taken to ensure that the surrounding environment was distant enough compared to the relative size of the subject. We register the geometry using ICP on manually marked correspondences for the first frame that are then tracked to subsequent frames via Lucas-Kanade [1981] optical flow using a Shi-Tomasi [1994] corner detector; we did not perform any further fine-tuning to improve the registration. White balancing is performed directly on the recovered incident lighting based on the gray world assumption, i.e., the average intensity among the color channels should be equal.

Results We recover the surface reflectance for five objects exhibiting a wide range of materials with rich spatially-varying details (e.g., *"rusted copper"*), sharp edges (e.g., the label on a *"beer bottle"*), and with surface reflectance ranging from highly specular (e.g., *"coffee mug"*) to near diffuse (e.g., center of the label on a *"beer bottle"*). We record a input video sequence of these objects under different types of illumination (shown in Figure 5) ranging from an office setting with a bright window (e.g., used for the *"beer bottle"* scene), artificial office lighting (used for the *"copper"* and *"duck"* scenes) to outdoor environments (i.e., used for the *"coffee mug"* scene).

Figures 1 and 17 show visualizations under novel incident lighting of the recovered surface reflectance. Note that for the *"coffee mug"* video sequence, the object was manually rotated without using a

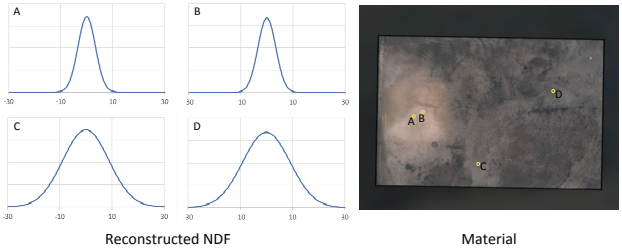


Figure 6: ***Recovered NDFs****. Four selected reconstructed NDFs from the* "rusted copper" *sample. Note that our method can reconstructed consistent results (a & b) and (c & d) for similar materials, without explicitely enforcing spatial coherence.*

rotation stage. Also note that the *"Starbucks can"* sequence was obtained from YouTube[1], where we assumed a gamma of 2.2 to radiometrically compensate the video, and where we hand-modeled the geometry as a cylinder. The number of required frames in the videos varies depending on the object shape and illumination. A necessary condition for a successful surface reflectance recovery is that each pixel's trace crosses a dominant edge in the lighting. For the acquired video sequences, the number of frames ranged from 300 for the *"beer bottle"* to 500 for the *"coffee mug"* to 1000 for the *"rusted copper"* example. The *"Starbucks can"* video sequence contained 1200 frames. All example, except the *"Starbucks can"* and the *"rusted copper"* examples, were rotated over two axes. The *"Starbucks can"* was only rotated around one axis, and the *"rusted copper"* was rotated over 4 axes. In addition, Figure 17 also shows visualizations of the geometry, recovered reflectance components, and recovered incident lighting for each of the example scenes. Even though recovering incident lighting is not the main focus, the resulting inferred lighting is still plausible; note that some of the dark regions in the recovered lighting are never "seen" in the reflections, and hence cannot be recovered. Reflectance recovery took approximately 8 to 10 hours on a dual Intel Xeon E5-2690 processor with 64GB memory, where the majority of the computational effort was spent on the NDF recovery. We recovered the reflectance properties for 1024×1024 surface points distributed in the UV space of the input geometry. On average, 8% of the time was spent on preprocessing the data, 70% on the NDF recovery, 12% on estimating the diffuse and specular albedo, and only 10% was spent on recovering the incident lighting. Due to the various non-linear steps (e.g., shock-filter), no theoretical guarantees exist on the convergence rate, or that the proposed optimization converges to a global minimum. However, practically we found that method produces good results, and we did not encounter vastly different convergences rates under different lighting conditions or for different ranges of material properties.

Figure 6 shows four selected recovered NDF from the *"rusted copper"* sample. As can be seen, our method can recover consistent NDFs for surface points that exhibit similar reflectance without explicitely enforcing spatial coherence. Note that the estimated lighting is shared among all surface points, and hence implicitly acts as a spatial coherence regularizer. In addition, we show in Figure 7 that the recovered surface reflectance is not over-fitted to the incident lighting by rendering the results under different incident illumination from the lighting present during capture. The *"rusted copper"* and *"toy duck"* samples are visualized under natural illumination from a window (while captured under office lighting), the *"coffee mug"* was captured in outdoor lighting and is visualized under indoor lighting, and the *"beer bottle"* is also lit by office lighting (while captured under natural lighting from a window). As can be seen, no discernable patterns related to the capture lighting

[1] `http://www.youtube.com/watch?v=PYU1zgSjws0`

Figure 7: ***Visual Comparison under Novel Lighting****. A comparison of visualizations of the recovered surface reflectance and a reference photograph under a similar lighting condition that differs from the acquisition environment. The surface region for the* "toy duck" *for which we reconstructed surface reflectance is marked in the reference photograph.*

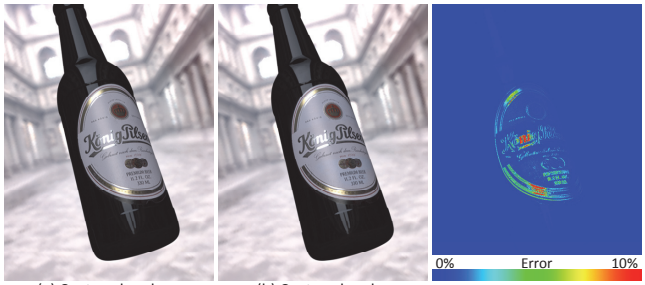


Figure 8: ***Repeatability under Different Acquisition Conditions****. Visualizations of the reconstructed surface reflectance of the* "beer bottle"*, acquired from two separate sequence under (a) the* "window lighting" *and (b) the* "office lighting"*. A false color difference image, normalized by the peak value, is shown in (c).*

are revealed, indicating that no overfitting occurs. Note, that the reference images are not a pixel-perfect match as it is difficult to place the object in exactly the same location viewed from the same viewpoint with the exact same camera. Furthermore, the handheld *"coffee mug"* was computed from linearized H.264 video frames (instead of RAW images to ensure a sufficient frame rate) which introduces some additional minor color differences.

Figure 8 demonstrates the robustness and repeatability of the reflectance estimation under different acquisition conditions (i.e., different motion and lighting). We captured two separate sequences of the *"beer bottle"* under the *"window lighting"* and *"office lighting"* conditions respectively, and rerendered the recovered surface reflectance for both under the *Uffizi Gallery* environment.

6 Discussion

The previous results demonstrate the versatility and robustness of our method on real-world scenes. While instructive, it only paints a partial image due to the difficulty of obtaining accurate ground truth surface reflectance and lighting. In this section we carefully validate the impact of each of the four variables that influence the performance of our method: lighting, material, geometry, and motion. We perform this validation on simulated measurements of synthetic and measured materials in order to study the impact of each variable in isolation.

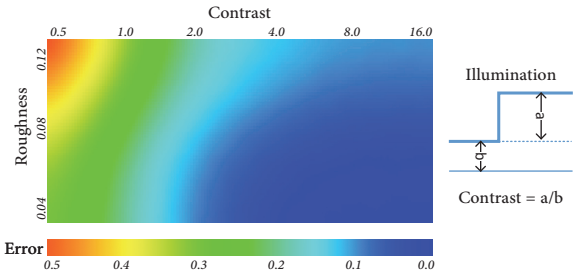


Figure 9: ***Contrast of Lighting Discontinuities.*** *A false color plot of the reconstruction error on Gaussian NDFs with varying roughness (0.04 to 0.12). A synthetic incident lighting condition was used that contains a single step-edge with varying contrast. A contrast level of* 4 *yields an error of less than* 12%.

6.1 Impact of Lighting

Discontinuity We rely on strong discontinuities in the incident lighting to robustly recover surface reflectance. A discontinuity can be characterized by its contrast (i.e., the difference between the bottom and the peak), and by its smoothness (i.e., rate of change).

Figure 9 shows a false color plot summarizing the impact of the discontinuity contrast. We vary incident lighting contrast (of a synthetic lighting environment that contains a single step-edge) and surface reflectance roughness (of an Gaussian NDF) and compute the reconstruction error. From this plot we can see that contrast is the dominating factor, and that roughness only plays a minor role – there is roughly a factor of two difference in contrast between rough and sharp BRDFs for a similar error. A modest contrast ratio of 4 already yields an error of less than 12%.

The impact of the smoothness on the accuracy of the results depends on the specific bandpass behavior of the surface reflectance. A low frequency BRDF can be accurately recovered from a smooth edge, whereas a high-frequency BRDF requires a sharp discontinuity. Figure 10 shows the impact of smoothing the incident lighting (*Uffizi Gallery*) on the "Wall Paper" svBRDF dataset [Dong et al. 2010] (mapped to a sphere). The difference images in Figure 10 and other figures follow the same false color coding and scale as in Figure 9. Sharp specular reflections become blurred when recovered from blurred incident lighting, and resulting error is approximately proportional to the size of the applied blur. This is not unexpected, and similar conclusions were made in prior work on inverse rendering [Ramamoorthi and Hanrahan 2001].

Consistency While the previous quantifies how the characteristics of the lighting impact the error, it does not indicate how consistent the method is under "good" incident lighting conditions. We validate that recovering surface reflectance is consistent over different incident lighting conditions using two synthesized input sequences of a measured "Wall Paper" svBRDF dataset [Dong et al. 2010] under different lighting conditions, and visualized under a third lighting condition. Figure 11 shows the resulting visualizations as well as the RMSE with a ground truth visualization. For both cases, the recovered surface reflectance produces qualitatively and quantitatively similar results.

6.2 Impact of Material

Accuracy To validate the accuracy on real-world materials, we ran our method on the 100 materials from the MERL-MIT dataset [Matusik et al. 2003] which spans a large section of the

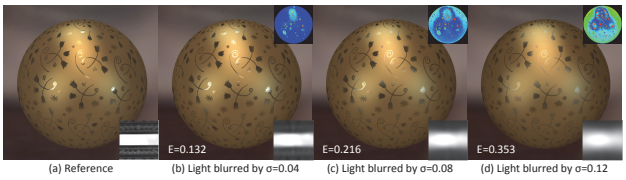


Figure 10: ***Impact of Smoothness of the Incident Lighting.*** *Recovered surface reflectance captured under incident lighting (*Uffizi Gallery*) smoothed with a Gaussian kernel, and rerendered under the* St. Peter's Basilica *light probe. The frequency bandwidth of the incident lighting is directly related to the maximal recoverable bandwidth of the surface reflectance.*

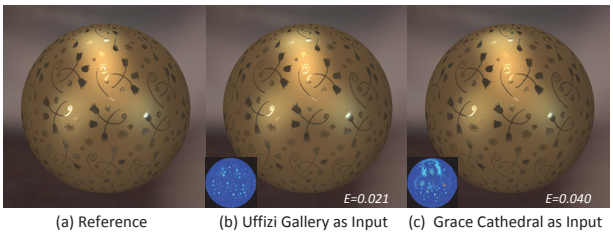


Figure 11: ***Robustness of Reflectance Recovery under Different Lighting Conditions****. Ground truth visualization lit by the* St. Peter *light probe (a). Rerendering and respective error of the surface reflectance recovered under the* Uffizi Gallery *light probe (b) and under the* Grace Cathedral *light probe (c).*

space of isotropic surface reflectance, and thus provides a good insight on how our method performs on arbitrary physical materials. We recovered surface reflectance for each material applied to homogeneous spheres under the *Uffizi Gallery* light probe, and compared the error on visualizations under *Grace Cathedral*, *St. Peter's Basilica*, and the *Eucalyptus Grove* light probe. Figure 12 plots the sorted average error on the *per-surface point recovered* surface reflectance. The errors range from 1% to 14%, with an average error of approximately 7%. We also show visualizations of the ground truth and the recovered surface reflectance for four selected materials.

Note that homogeneous materials (as in Figure 12) represent a more difficult case for our method than heterogeneous materials, because the estimation of albedo and lighting becomes ambiguous – the diffuse reflectance can be "baked into" the incident lighting (and have an effective zero diffuse albedo) while producing the same observations[2]. Note, this is not true for the spatially varying case, where the ratio between diffuse and specular albedo varies per-surface point. We resolve this ambiguity by assuming that the minimum intensity in the incident lighting is 0, and add a $\lambda_{reg}||\min(L(\omega_i))||$ term, with $\lambda_{reg} = 0.01$, to the recovery of the incident lighting. While this assumption is unlikely to hold, and thus overestimate the diffuse albedo, the overestimation is approximately proportional to the ratio between $\min(L(\omega_i))$ and $\int L(\omega_i)d\omega_i$, and thus in practice the degree of over-estimation is small.

Diffuse-only Our method relies mostly on specular reflectance cues to estimate the incident lighting. The sharper the specular reflections, the better the estimate of the incident lighting, and thus the more accurate we can detect salient edges. However, as shown

[2]This ignores the Fresnel and the geometry terms, a reasonable approximation since for many materials these do not sufficiently constrain the minimization to break the ambiguity.

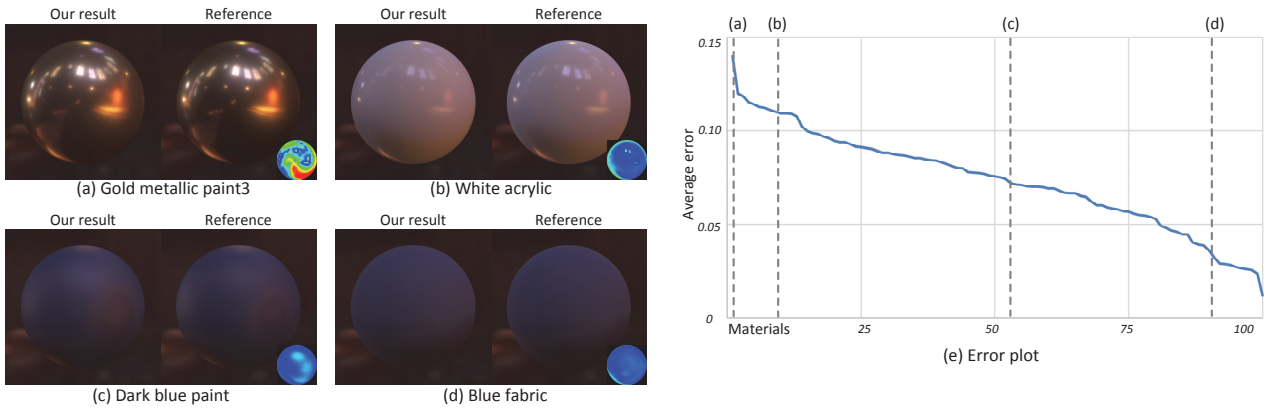


Figure 12: ***Accuracy Validation****. A visualization of error in the recovered surface reflectance of homogeneous spheres captured under the* Uffizi Gallery *light probe. A comparison to ground truth for four seclected materials is shown on the left.*

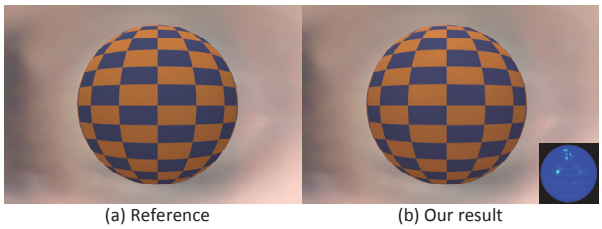


Figure 13: ***Diffuse Surface Reflectance****. In the absence of specular surface reflectance, no salient edges are detected in the temporal traces. Yet, our method is still able to recover the diffuse surface reflectance accurately.*

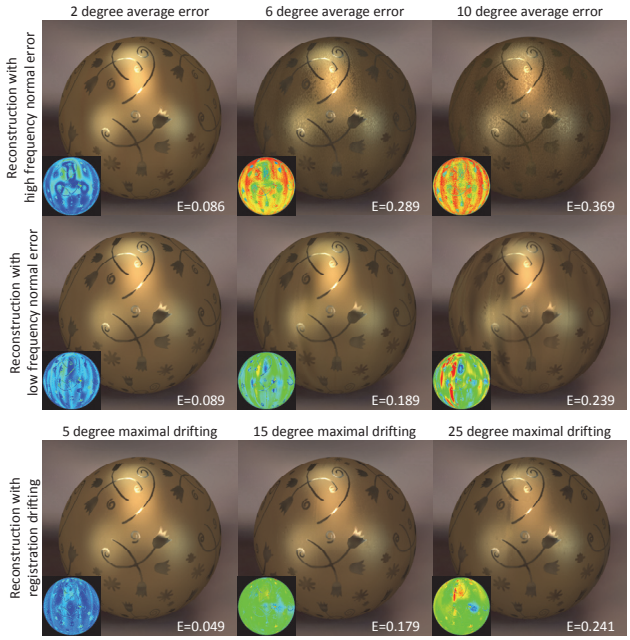


Figure 14: ***Impact of Normal & Registration Errors.*** *(top-middle) Visualizations of the impact (and reconstruction errors) of incorrect surface normals by applying low frequency and high frequency normal perturbations with varying maximum deviations. (bottom) Visualizations of the impact and reconstruction errors of (constant cumulative) drift in the geometry registration. Please refer to Figure 15 for a reference visualization of the subject illuminated by the* St. Peter's Basilica *lighting environment.*

in Figure 12(c), our method also works well on rough specular materials – the accuracy requirements for the detection of salient edges decreases for rough specular materials. The proposed method also works well in the limit case of diffuse-only surface reflectance. Figure 13 shows the correct reconstruction of a spatially varying diffuse albedo on a synthetic test scene. While our method also produces some specular NDFs, their corresponding specular albedo is too low to visually impact the results.

6.3 Impact of Geometry

Surface Normal Errors While many 3D shape acquisition methods offer very good depth-accuracy, they often do not make guarantees with respect to surface normal accuracy, and typically lack surface normal detail. We therefore focus on the impact of errors on the surface normals on the accuracy of the reflectance recovery. In particular we consider two cases: low frequence normal errors (i.e., spatially coherent and slowly varying error) and high frequency normal errors (i.e., incoherent spatially varying surface detail). We recover the surface reflectance of a spherical object with the measured "Wall paper" material [Dong et al. 2010] illuminated by the *Uffizi Gallery* light probe. In all cases, we use the exact geometry (a sphere) to recover the surface reflectance, but apply a perturbation ranging from 2 degrees to 10 degrees to the surface normals. Figure 14 shows visualizations of the ground truth and recovered surface reflectance under the *St. Peter's Basilica* light probe, as well as the average BRDF reconstruction error. Both low and high frequency perturbation are robust to perturbations of 2 degrees or less. While the error on high frequency perturbations grows larger, it is visually less noticeable (incoherent noise) than the error due to low frequency perturbations (spatially coherent but noticeable artifacts).

Registration Errors Our technique also requires accurate registration of the geometry to each frame in the video sequence. We currently rely on optical flow to track correspondences which is known to suffer from drift over long sequences. Figure 14 (bottom) illustrates the impact of drift on the rotation estimation where a cumulative systematic error was introduced to the rotation of the registered object in each frame, reaching a maximum error of 5, 15, and 25 degrees, respectively, at the end of a 1440-frame sequence. As can be seen, our method is more robust to registration errors than to surface normal errors, and only at a substantial 25 degree error, the differences become noticeable. It should be noted that our method is not married to a specific registration method, and (future) more sophisticated registration algorithms can easily be swapped in. In

Figure 15: ***Insufficient Rotation.*** *A common type of natural lighting environment is one where the horizon is the only source of a strong discontinuity. Only rotating the object horizontally does not provide sufficient information to recover the surface reflectance (b). However, an additional vertical rotation (c) produces results close to the reference (a).*

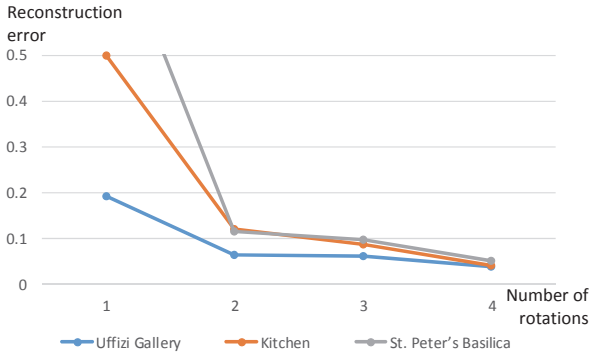


Figure 16: ***Impact of Motion.*** *We quantify the impact of motion by plotting the reconstruction error under three different lighting environments for a varying number of rotation axes. Two rotation axes already provide good results for lighting environments with large area light sources with sharp edges (e.g.,* Uffizi Gallery*). For more complex lighting environments (e.g.,* St. Peter's Basilica*) four or more rotations are needed.*

practice, we found that our naive registration algorithm provided sufficiently accurate results for our examples (including the hand-rotated *coffee mug*) without any hand-tuning.

6.4 Impact of Motion

A final variable that impacts the accuracy of the reflectance recovery is the motion applied during acquisition. Since we do *not* exploit spatial coherence and process each surface point separately, a salient edge needs to be "scanned" over each surface point.

In general, our method will fail when the rotations are too fast (e.g., causing motion blur), too limited rotations (e.g., near static object rotation), or exhibit too limited *local* motions (e.g., near static points). Figure 15 illustrates the effects of ill-chosen or insufficient rotations. In this figure the recovered surface reflectance of a horizontally rotated object in an environment with a single horizontal salient edge (i.e., horizon) clearly exhibits artifacts. However, an additional vertical rotation of the object allows us to accurately recover the surface reflectance.

The ideal number of rotation axes depends on the nature of the incident lighting. If the incident lighting contains a long salient edge, then a rotation around two different axis provides sufficient information to accurately recover the surface reflectance – a single axis suffices if reflectance recovery close to the pole is not necessary. However, if the incident lighting is a collection of directional light sources, then a more elaborate acquisition might be required to ensure that every surface point "sees" at least one such directional light source. Figure 16 plots the reconstruction error for a varying number of rotation axes on three types of lighting environments: large area light sources with sharp edges (*Uffizi Gallery*), medium sized area light sources (*Kitchen*), and point light sources (*St. Peter's Basilica*). The rotation axes are uniformly distributed in the X-Y plane: $(cos(\pi i/n), sin(\pi i/n), 0)$. For the *Uffizi Gallery* lighting environment, a two axis rotation already provides low reconstruction error. For the more challenging *St. Peter's Basilica* at least 4 rotation axes are needed.

7 Conclusion

We presented "appearance-from-motion", a novel method for recovering spatially varying isotropic surface reflectance under unknown natural illumination. Our method significantly simplifies acquisition; only a video of a rotating object and its geometry are needed. We demonstrated the effectiveness of our method on a variety of objects with a wide range of reflectance properties, and under different types of incident lighting.

We currently ignore self-occlusions and interreflections, limiting our method to mostly convex object shapes. Taking occlusions into account would require reformulating the incident lighting to include an occlusion term: $L(\omega_i) \times G(x,t)$, which can be directly derived from the available geometry. Including the self-occlusion term is an interesting avenue for future research, as this can introduce additional salient edges, and thus regularize the recovery further. Another fruitful direction for future research is the joint recovery of shape and high-quality appearance especially on subjects that exhibit high frequency surface normal variation. We believe our appearance-from-motion framework can serve as an integral component of such a system.

Acknowledgements

We wish to thank the reviewers for their constuctive feedback. Pieter Peers was partially funded by NSF grants: IIS-1217765, IIS-1350323, and a gift from Google.

References

AITTALA, M., WEYRICH, T., AND LEHTINEN, J. 2013. Practical svbrdf capture in the frequency domain. *ACM Trans. Graph. 32*, 4, 110:1–110:12.

ASHIKHMIN, M., PREMOZE, S., AND SHIRLEY, P. 2000. A microfacet-based BRDF generator. In *Proceedings of the 27th annual conference on Computer graphics and interactive techniques*, 65–74.

BARRON, J. T., AND MALIK, J. 2013. Shape, illumination, and reflectance from shading. Tech. Rep. UCB/EECS-2013-117, EECS, UC Berkeley, May.

CHO, S., AND LEE, S. 2009. Fast motion deblurring. *ACM Trans. Graph. 28*, 5, 145:1–145:8.

DONG, Y., WANG, J., TONG, X., SNYDER, J., LAN, Y., BEN-EZRA, M., AND GUO, B. 2010. Manifold bootstrapping for SVBRDF capture. *ACM Trans. Graph. 29*, 4, 98:1–98:10.

DORSEY, J., RUSHMEIER, H., AND SILLION, F. 2008. *Digital Modeling of Material Appearance*. Morgan Kaufmann Publishers Inc.

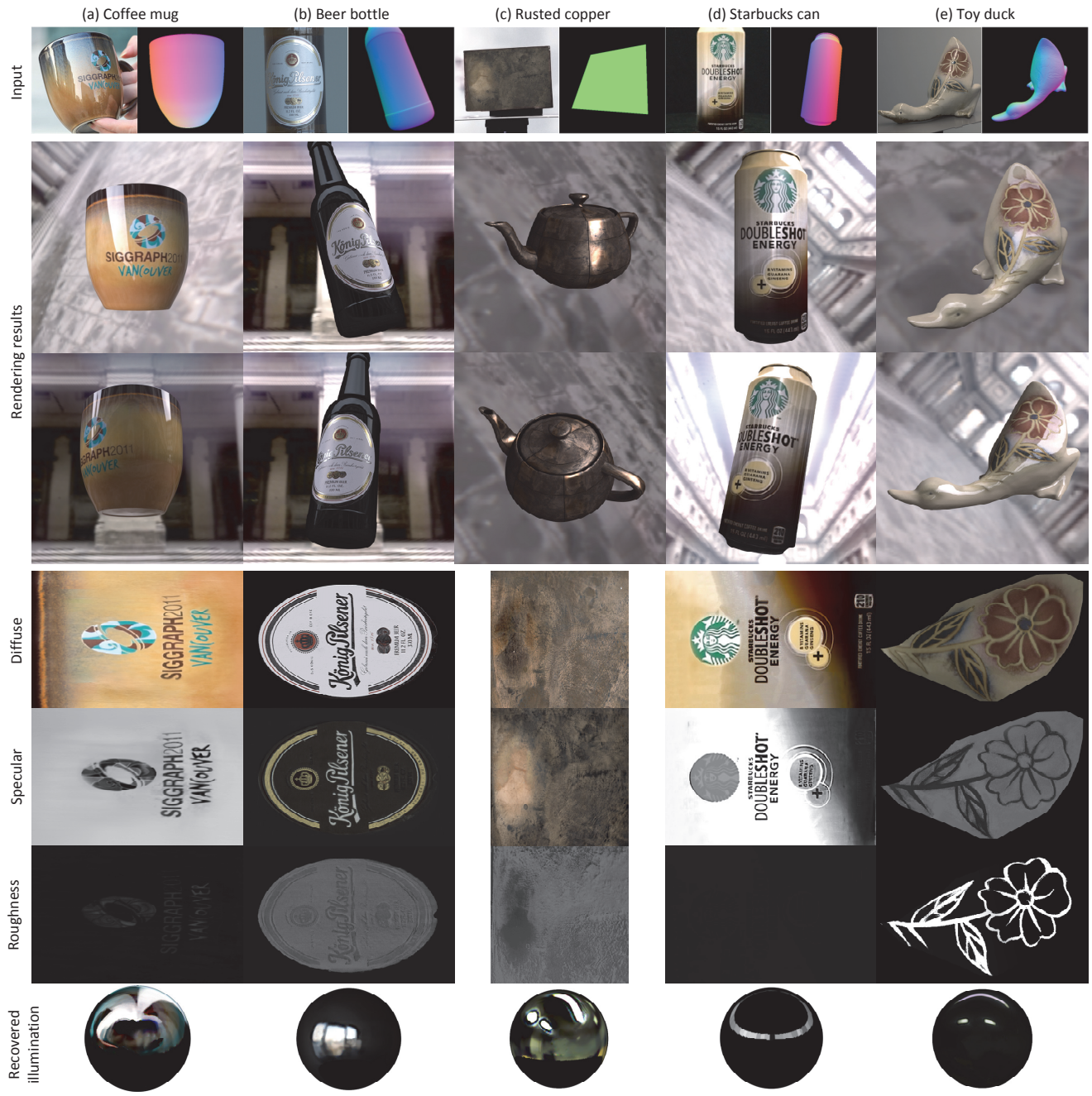


Figure 17: ***Additional Results.*** *(a) "coffee mug", (b) "beer bottle", (c) "rusted copper", (d) "Starbucks can", and (e) "toy duck". One frame of the input sequence, as well as visualizations of the geometry, the estimated lighting, and the recovered reflectance properties are shown for each example. For visualization purposes, the recovered tabulated NDF is summarized by the roughness of a Cook-Torrance BRDF fitted to the recovered reflectance.*

GARDNER, A., TCHOU, C., HAWKINS, T., AND DEBEVEC, P. 2003. Linear light source reflectometry. *ACM Trans. Graph. 22*, 3, 749–758.

GREGSON, J., HEIDE, F., HULLIN, M. B., ROUF, M., AND HEIDRICH, W. 2013. Stochastic Deconvolution. In *CVPR*, 1043–1050.

HABER, T., FUCHS, C., BEKAER, P., SEIDEL, H. P., GOESELE, M., AND LENSCH, H. 2009. Relighting objects from image collections. In *CVPR*, 627–634.

HERTZMANN, A., AND SEITZ, S. M. 2003. Shape and materials by example: A photometric stereo approach. In *CVPR*, 533–540.

HOLROYD, M., LAWRENCE, J., AND ZICKLER, T. 2010. A coaxial optical scanner for synchronous acquisition of 3D geometry and surface reflectance. *ACM Trans. Graph. 29*, 4, 99:1—-99:12.

LENSCH, H. P. A., KAUTZ, J., GOESELE, M., HEIDRICH, W., AND SEIDEL, H.-P. 2003. Image-based reconstruction of spatial appearance and geometric detail. *ACM Trans. Graph. 22*, 2, 234–257.

LEVIN, A., FERGUS, R., DURAND, F., AND FREEMAN, W. T. 2007. Image and depth from a conventional camera with a coded aperture. *ACM Trans. Graph. 26*, 3, 70:1–70:9.

LI, G., WU, C., STOLL, C., LIU, Y., VARANASI, K., DAI, Q., AND THEOBALT, C. 2013. Capturing relightable human performances under general uncontrolled illumination. *Comput. Graph. Forum 32*, 2, 275–284.

LOMBARDI, S., AND NISHINO, K. 2012. Reflectance and natural illumination from a single image. In *ECCV*, 582–595.

LUCAS, B., AND KANADE, T. 1981. An iterative image registration technique with an application to stereo vision. In *Proc. Int. Joint Conf. on Artificial Intelligence*, 674–679.

MARSCHNER, S. R. 1998. *Inverse Rendering for Computer Graphics*. PhD thesis, Cornell University.

MATUSIK, W., PFISTER, H., BRAND, M., AND MCMILLAN, L. 2003. A data-driven reflectance model. *ACM Trans. Graph. 22*, 3, 759–769.

NICODEMUS, F. E., RICHMOND, J. C., HSIA, J. J., GINSBERG, I. W., AND LIMPERIS, T. 1977. Geometric considerations and nomenclature for reflectance. *Monograph 161,National Bureau of Standards (US)*.

NISHINO, K., ZHANG, Z., AND IKEUCHI, K. 2001. Determining reflectance parameters and illumination distribution from a sparse set of images for view-dependent image synthesis. In *ICCV*, 599–606.

OSHER, S., AND RUDIN, L. I. 1990. Feature-oriented image enhancement using shock filters. *SIAM J. Numer. Anal. 27*, 4, 919–940.

RAMAMOORTHI, R., AND HANRAHAN, P. 2001. A signal-processing framework for inverse rendering. In *Proceedings of the 28th Annual Conference on Computer Graphics and Interactive Techniques*, SIGGRAPH '01, 117–128.

REN, P., WANG, J., SNYDER, J., TONG, X., AND GUO, B. 2011. Pocket reflectometry. *ACM Trans. Graph. 30*, 4, 45:1–45:10.

ROMEIRO, F., AND ZICKLER, T. 2010. Blind reflectometry. In *ECCV*, 45–58.

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

SHAN, Q., ADAMS, R., CURLESS, B., FURUKAWA, Y., AND SEITZ, S. M. 2013. The visual turing test for scene reconstruction. In *3DV*, 25–32.

SHI, J., AND TOMASI, C. 1994. Good features to track. In *CVPR*, 593–600.

TREUILLE, A., HERTZMANN, A., AND SEITZ, S. M. 2004. Example-based stereo with general BRDFs. In *ECCV*, 457–469.

TUNWATTANAPONG, B., FYFFE, G., GRAHAM, P., BUSCH, J., YU, X., GHOSH, A., AND DEBEVEC, P. E. 2013. Acquiring reflectance and shape from continuous spherical harmonic illumination. *ACM Trans. Graph. 32*, 4, 109.

WANG, C.-P., SNAVELY, N., AND MARSCHNER, S. 2011. Estimating dual-scale properties of glossy surfaces from step-edge lighting. *ACM Trans. Graph. 30*, 6, 172:1–172:12.

XU, L., AND JIA, J. 2010. Two-phase kernel estimation for robust motion deblurring. In *ECCV*, 157–170.

YU, T., WANG, H., AHUJA, N., AND CHEN, W.-C. 2006. Sparse lumigraph relighting by illumination and reflectance estimation from multi-view images. In *Rendering Techniques*, 41–50.

ZHANG, Z. 2000. A flexible new technique for camera calibration. In *IEEE PAMI*, vol. 22, 1330– 1334.