Practical SVBRDF Capture In The Frequency Domain

Miika Aittala

Aalto University
NVIDIA Research

Tim Weyrich

University College London

Jaakko Lehtinen
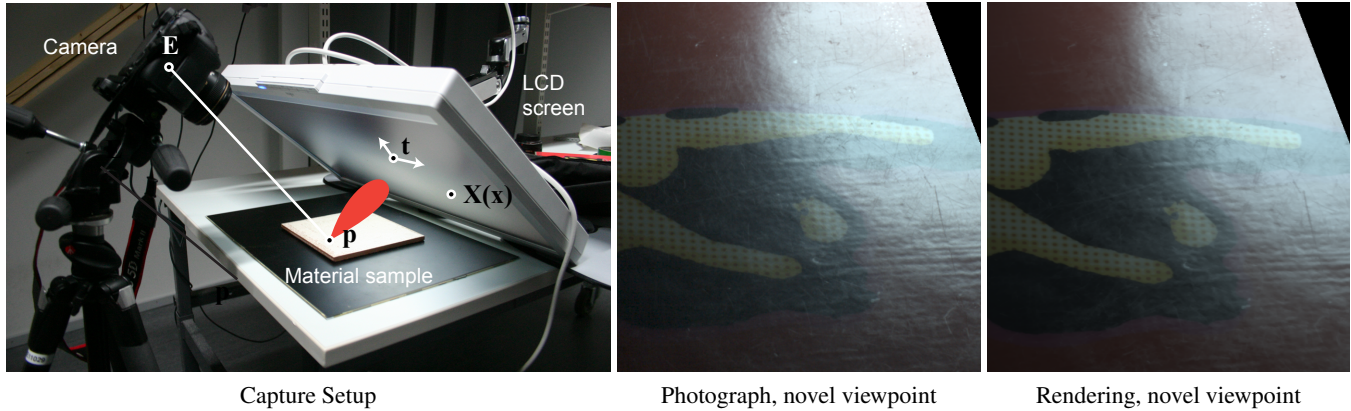
Aalto University
NVIDIA Research

Figure 1: *Left: our measurement setup consists of a screen and a camera. Middle: a photograph of a material sample, taken under a novel viewpoint and illumination not used in the capture. Right: our rendering with matching lighting and viewing conditions.*

Abstract

Spatially-varying reflectance and small geometric variations play a vital role in the appearance of real-world surfaces. Consequently, robust, automatic capture of such models is highly desirable; however, current systems require either specialized hardware, long capture times, user intervention, or rely heavily on heuristics. We describe an acquisition setup that utilizes only portable commodity hardware (an LCD display, an SLR camera) and contains no moving parts. In particular, a laptop screen can be used for illumination. Our setup, aided by a carefully constructed image formation model, automatically produces realistic spatially-varying reflectance parameters over a wide range of materials from diffuse to almost mirror-like specular surfaces, while requiring relatively few photographs. We believe our system is the first to offer such generality, while requiring only standard office equipment and no user intervention or parameter tuning. Our results exhibit a good qualitative match to photographs taken under novel viewing and lighting conditions for a range of materials.

CR Categories: I.4.1 [Image Processing and Computer Vision]: Digitization and Image Capture—Reflectance;

Keywords: appearance capture, reflectance, SVBRDF, Fourier analysis

Links: ◈DL 📄PDF 🌐WEB

1 Introduction

Most natural materials exhibit spatially-varying surface reflectance properties. Even if perhaps mostly flat and homogeneous over large scales, they still exhibit scratches, scuffing and other local material variations that greatly contribute to their look. Accordingly, most digital assets in games and films are nowadays assigned normal maps and spatially-varying reflectance parameters. Nevertheless, the *acquisition* of spatially varying BRDF (SVBRDF) parameters from real surfaces remains cumbersome.

The SVBRDF is a six-dimensional function of space and angles, which complicates its acquisition. Exhaustive sampling of the six-dimensional space leads to prohibitive acquisition times [Dana and Wang 2004; Holroyd et al. 2010], or, if the samples are too sparsely distributed, incurs aliasing, for instance of narrow specular reflectance lobes. More recent work devised devices for smart capture of representative samples [Lensch et al. 2003; Dong et al. 2010; Ren et al. 2011], or aggressively reduced the amount of input data, making strong assumptions on the spatial material distribution and relying on user interaction to touch up the data [Clark 2010; Dong et al. 2011] which generally sacrifices accuracy.

A recent trend recognizes the need for practical high-resolution SVBRDF capture in an informal setting, devising simple hardware to capture representative reflectance properties in subspaces of the SVBRDF and inferring the full function through data amplification [Dong et al. 2010; Ren et al. 2011]. Our work takes this further, offering *independent per-point BRDF reconstructions* with hardware already at most artists' desks. To this end we follow two key design decisions toward a practical SVBRDF acquisition system.

First, we confine observations to a smaller range of the angular domain, using a single viewpoint and illuminating the sample using a planar light source significantly smaller than a full spherical lighting environment. To still capture the most prominent features of the reflectance lobes, we concentrate this sampling on the mirror direction as seen from the fixed viewpoint. In practice, this restricts applicability to near-planar surfaces, but we argue that this case is common enough to cover a majority of real-world scenarios.

Second, we use fully automated commodity hardware only, allowing for capture and processing with no intermittent user intervention.

Our concrete setup consists of a single LCD screen and one camera, facing a near-planar material sample from opposite sides, see Figure 1, left. The static camera takes images of the sample illuminated by intensity patterns displayed on the screen.

In principle, this setup could emulate a set of spatially distributed point-light sources, to measure individual directions of incidence. To keep the number of input photographs manageable, however, the number of discrete point lights would have to be limited, incurring aliasing for specular highlights. Instead, we accomplish aliasing-free reconstruction of very narrow, even mirror-like reflectance lobes by illuminating with continuous, band-limited illumination patterns.

Having restricted the range of angular reflectance samples, extrapolating the reconstruction to the full range of views and illumination requires fitting a reflectance model to the data. Any such (non-linear) fit requires repeated model evaluation to compare the current prediction to measurements, which for our continuous illumination requires integrating the product of the BRDF with the illumination pattern. For traditional reflectance models, however, this would generally imply expensive numerical integration in every iteration of the optimizer. We solve this problem by jointly designing reflectance model and illumination patterns, so that the respective integrals can be computed analytically, making an iterative model fit computationally feasible.

In this image formation model, we approximate reflectance lobes as Mixtures of Gaussians (MoG) in the illuminating screen's domain and use Fourier basis patterns, windowed by a Gaussian, for illumination. As we will show, typical reflectance functions are well-behaved in our chosen basis: they require only few basis functions for reconstruction and separate well into diffuse and specular components in that domain.

In summary, we believe our system is the first to automatically capture normal maps and a wide range of reflectance properties, from diffuse to almost perfectly specular, using a static system built from low-cost commodity parts, and requiring no user intervention or parameter tuning. Our key technical contribution is the jointly-designed combination of continuous illumination patterns and a carefully crafted image formation model that can predict the values of observations analytically, and captures the essential features of reflectance using relatively few measurements. We evaluate our approach using ground-truth comparisons for material samples exhibiting a wide range of reflectance properties.

2   Related Work

Previous work in appearance capture has acquired image-based representations, as well as explicit surface reflectance representations that range from pure albedo reconstruction to BRDF acquisition and even reconstruction of local subsurface scattering properties at every surface point of an object [Weyrich et al. 2008]. In this section, we will focus on per-point surface reflectance reconstruction; image-based object representations and subsurface scattering are considered outside our scope.

2.1   (SV-)BRDF Acquisition

Normal and Reflectance Capture   Jointly estimating reflectance functions and normals from reflectance observations has a long tradition, starting with photometric stereo [Woodham 1978; Ikeuchi 1981], moving on to more general BRDF models [Coleman and Jain 1982; Nayar et al. 1990; Weyrich et al. 2006; Holroyd et al. 2008; Ghosh et al. 2009], partially deriving normal orientations by

exploiting internal symmetries in physical BRDFs [Zickler et al. 2002; Holroyd et al. 2008]. In our approach, we obtain normal information as part of the model fit.

Model Assumptions   Acquisition of SVBRDFs usually aims at increasing efficiency by making simplifying world assumptions, for instance by assuming isotropic reflectance and by modeling appearance as a spatially varying linear combination of few basis BRDFs [Lensch et al. 2003; Goldman et al. 2005; Lawrence et al. 2006; Dong et al. 2010; Ren et al. 2011]. We, too, assume isotropic reflectance and use a few-parameter BRDF model, but we aim at deriving independent, per-point reflectance fits.

Strong Priors   Some approaches aggressively minimize the set of measurements required, while softening the requirement of faithful reconstructions. CrazyBump [Clark 2010] obtains shape and albedo from single photographs and allows manual editing of appearance. Their system uses a proprietary set of heuristics and heavily relies on user input. By adding a flash-exposed image, Glencross et al. [2008] allow for a more principled heuristic, deriving perceptually plausible elevation maps from single viewpoints. Dong et al. [2011] present the system AppGen, which decomposes a single image into shading and diffuse albedo components by a user-guided intrinsic-image approach. The normals are computed using shape from shading, and the specular BRDFs are assigned by the user. In all of these cases, the results often look surprisingly convincing but do not bear a strong quantitative relation to reality.

Linear Light Sources   Gardner et al. [2003] and Ren et al. [2011] use setups with a moving linear light source and capture the intensity profiles of the reflections as a function of time. Very roughly speaking, our data can be seen as Fourier transforms of such profiles. Our arrangement requires no mechanical movement. Our choice of sampling directly in the Fourier instead of primal basis allows sampling with less images.

Basis Illumination   One problem that traditionally drives up the number of light sources to sample illumination directions is aliasing: in order to be able to capture very narrow reflectance lobes, a dense angular sampling is required. Sato et al. [2003] hence propose the use of continuous, low-pass filtered illumination. A system that builds upon this principle is the work by Ghosh et al. [2007], who built a catadioptric system to illuminate a surface point with continuous, zonal basis functions over a hemisphere. Our work is strongly inspired by their approach; in comparison, while Ghosh et al. image a single surface point from various directions simultaneously, we trade the angular for the spatial domain, thus acquiring fully spatially-varying appearance from a single viewpoint only. Ghosh et al. [2007] further note that expressing reflectance directly using the basis used for acquisition leads to ringing in the presence of sharp specularities. Similar to them, we address this by fitting a parametric BRDF model to the measurements. Our model then extrapolates the data to allow for novel viewpoints.

Equally related to our work are the monitor-based capture setups presented in more recent papers by Ghosh et al. [2009; 2010]. They both display polarized, low-order spherical harmonic patterns on a monitor, observing a surface with a polarized camera under mirror configuration. They exploit polarization-dependence of light reflectance to extract specular parameters and (in [Ghosh et al. 2009]) normals from small sets of images. In contrast to our work, however, their methods require the reflectance lobe to be fully contained in the monitor; as soon as a lobe overlaps with the monitor boundary (as for instance the case with diffuse or glossy reflectance), the reconstruction breaks down. A key property of our method is that it models this case directly and thus enables reconstruction of wide-lobed reflectances as well. Also, their method requires (mechanically)

changing polarization throughout the measurements, increasing the risk of image misalignment due to mechanical movement, which is an error source avoided by our solid-state design.

2.2 Similar Hardware Setups

Wang et al. [2011] present a setup similar to ours that recovers statistical surface normal and reflectance from a very low number of photos under step-edge illumination. Like ours, their model uses Gaussians to approximate the reflectance, but the range of materials that conform to their model assumptions is very limited and, most critically, they do not reconstruct spatially-varying reflectance.

Screens have also been used to illuminate objects in other areas of appearance capture. Environment matting aims at tabulating light transport from a background scene through a (typically half-transparent or refractive) foreground object [Chuang et al. 2000; Zhu and Yang 2004]. The focus of such work, however, is more in finding memory-efficient encodings of the transfer function and on finding minimal pattern sequences that allow identifying the origin of a ray path.

Also, there are methods that used screen illumination to reconstruct surface normals only. Francken et al. [2008] use a setup very similar to ours, imaging Gray codes displayed by a monitor reflected in a surface sample. They need many images to reduce aliasing and only reconstruct normals, not BRDFs. Similar to our work, their reconstruction assumes near-planar overall geometry. Other approaches lift this restriction by placing the Gray-code displaying screen at two subsequent planar positions [Bonfort et al. 2006], or by using a second camera [Nehab et al. 2008], thus allowing for triangulation of general 3-D points in space. Chen et al. [2006] use a static camera and moving light source to observe highlights and derive normal directions from these observations. They need very many images and still suffer from aliasing.

3 Overview

Our capture setup observes the reflections of the patterns emitted by the screen on the material sample. By reversing the direction of light transport, the observations can be seen as viewing rays leaving the camera and reflecting into the screen in the form of a reflectance lobe (see Figure 1, left): each pixel measures the integral of the product of this projected lobe and the illumination pattern.

In the following sections, we describe how we model this light transport in our measurement setup: by choosing (windowed) Fourier basis patterns (Section 4) and by modeling reflectance as a Mixture of Gaussians (Section 5), we obtain a complete image formation model that still allows for analytic integration, see Section 6. Our derivation involves careful approximations, using a planar domain for our basis patterns and to reparameterize reflectance lobes. This parameterization further allows us to locally approximate perspective relations within the setup by affine projections into that plane. This consideration of the geometric terms introduced by the local lighting lifts restrictions on the measurement setup commonly present in previous work that assumes distant lighting.

Section 4.4 will take a closer look at the properties of our measurements: through the use of a Fourier basis, our observations are essentially spectral samples of the reflectance convolved by the Fourier-transformed windowing function, and we will show that reflectance lobes behave and separate well within that domain.

Our model's analytical integrability allows for efficient SVBRDF reconstruction through Bayesian inference (Section 7). The use of such a flexible optimization framework enables us to introduce carefully chosen priors that improve stability of the reconstruction

$\mathcal{A}(\mathbf{x})$	affine part of diffuse term	$\mathcal{N}^E(\mathbf{x})$	monitor emittance Gaussian
$\mathbf{a}_d$	diffuse albedo RGB vector	$\mathcal{N}^{S,l}(\mathbf{x})$	specular Gaussians
$\mathbf{a}_s$	specular albedo RGB vector	$\mathcal{N}^W(\mathbf{x})$	window Gaussian
$d(\mathbf{p})$	orthogonal distance from screen	ω	2D frequency
$D(\mathbf{x};\mathbf{p})$	distance between points	$\mathbf{p}$	point on sample surface (world)
$\mathbf{E}$	camera position (world)	ρ	full BRDF (screen)
$E_s(\mathbf{X}\to\mathbf{p})$	monitor emittance distribution	ρ_d,ρ_s	diffuse and specular BRDF (screen)
$\mathbf{f}(\mathbf{x})$	the function that is measured	σ	specular glossiness
γ	specular kurtosis	$\mathbf{T}$	matrix of screen to world transform
$g^j(\mathbf{x})$	jth pattern emitted from screen	$\mathbf{t}$	screen origin (world)
i	pixel index	$w(\mathbf{x})$	window function
j	frequency index	$\mathbf{X}$	point on screen (world)
$\mathbf{n}$	surface normal	$\mathbf{x}$	point on screen (local)
$\mathcal{N}^{D,k}(\mathbf{x})$	diffuse Gaussians	$\mathbf{Z}$	captured data

Table 1: *List of symbols.*

(Section 7). Results are presented in Section 8. See Table 1 for the list of symbols.

4 Reflectance Measurements

This section describes our light transport model, the geometric configuration of our setup, and how we account for the radiometric properties of the screen used for illumination. In addition, we motivate our choice of spectral emission patterns.

4.1 Light Transport Model

We capture J photos, each with a different pattern displayed on the screen. The image data $\mathbf{Z}_{i,j}$ recorded by the camera is the exitant radiance from the point $\mathbf{p}^i$ seen in the ith pixel towards the camera under the jth illumination condition. Let the camera be located in position $\mathbf{E}$ in world coordinates, and parameterize the screen plane by $\mathbf{x}$ in its local 2D coordinate system, with corresponding world coordinates $\mathbf{X}(\mathbf{x})$ (cf. Figure 1, left). Since we assume the screen to be the sole, unoccluded light source, the values of the observations can be modeled by the reflectance equation written as an integral over the screen:

$$\mathbf{L}^j(\mathbf{p}\to\mathbf{E}) = \int_{\mathbb{R}^2} \rho(\mathbf{x};\mathbf{p}) L_e^j(\mathbf{X}(\mathbf{x})\to\mathbf{p})\,d\mathbf{x}. \tag{1}$$

Here $\rho(\mathbf{x};\mathbf{p})$ is the apparent reflectance function of point $\mathbf{p}$ evaluated in the direction of the point $\mathbf{X}$. It captures the joint effects of the compound (diffuse+specular) reflectance lobe, cosine term, solid angle-to-area variable change, the position of $\mathbf{p}$ with respect to the screen, and any possible local interreflection and shadowing effects. L_e^j is the emitted radiance from the jth emission pattern. Our concrete goal is to formulate an analytic model that describes the observed pictures in terms of the known geometric setup and intuitive parameters, i.e., albedos, normals, etc., and for each $\mathbf{p}^i$, find the set of parameters that minimize deviation from the observations.

Coordinate Systems　Because the screen is close to the sample, we must explicitly account for local illumination effects, i.e., the variable relative positioning of the sample and screen. The world and screen coordinates systems are related by

$$\mathbf{X}(\mathbf{x}) = \mathbf{Tx} + \mathbf{t}, \tag{2}$$

where $\mathbf{T}\in\mathbb{R}^{3\times 2}$ is an orthogonal matrix and $\mathbf{t}$ is the origin of the screen in world coordinates. These are determined by a simple calibration procedure (Appendix A). To enable the use of analytic Fourier transforms defined over the entire plane, we treat the screen plane as having an infinite extent and restrict to the physical screen by multiplication by a window function. We parameterize the plane such that the physical screen rectangle is $\mathbf{x}\in[-\pi,\pi]\times[-\pi/a,\pi/a]$, where a is the aspect ratio (e.g., $a = 16/9$).
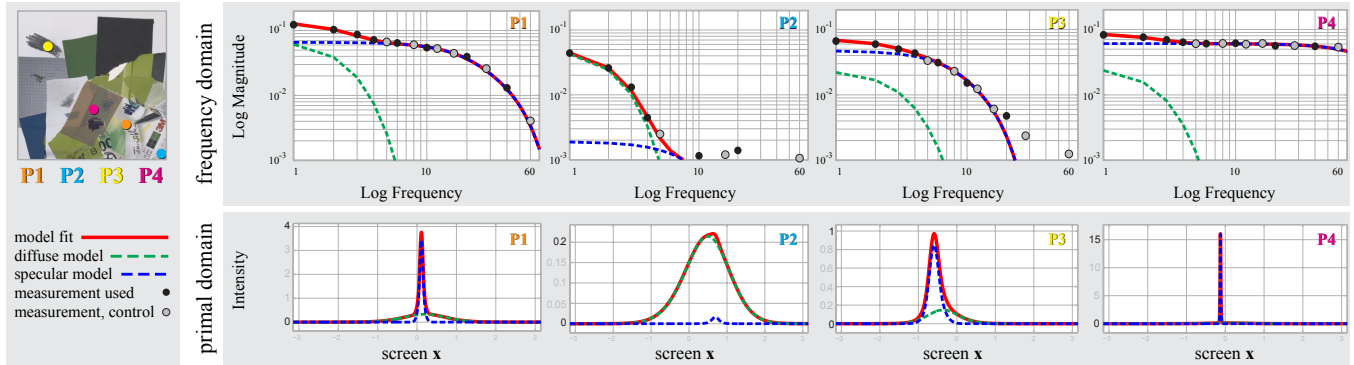
Figure 2: *A material sample, with four separate points marked.* Top row: *a slice through the horizontal frequency axis.* Bottom row: *primal-domain reflectance lobes, parameterized over the screen.* $\mathbf{P}_1$: *moderate gloss.* $\mathbf{P}_2$: *diffuse.* $\mathbf{P}_3$: *dull gloss.* $\mathbf{P}_4$: *highly specular. Note the different vertical axis scales in the primal domain plots.*

4.2 Monitor Properties

Emission Distribution Function   To account for the non-uniform emission distribution of typical screens, we expand the emission pattern as

$$L_e^j(\mathbf{X} \to \mathbf{p}) = g^j(\mathbf{x})\, E_s(\mathbf{X}(\mathbf{x}) \to \mathbf{p}). \qquad (3)$$

Here $g^j(\mathbf{x})$ is a 2D pattern to be displayed on the screen, and E_s is the emission distribution function.

We found that the emission distribution of a high-quality panel can be sufficiently well modeled as follows. We only consider the angular distribution of emittance, independent of the screen position. A convenient parameterization is given by letting the viewer be located at unit distance perpendicularly away from the screen origin $\mathbf{t}$, with the screen emitting full-intensity white. The radiance observed by the viewer from each screen position is a function of $\mathbf{x}$. We approximate this function as a Gaussian $\mathcal{N}^{\tilde{E}}(\mathbf{x})$ with a diagonal covariance and a mean that is zero on the horizontal axis, corresponding to typically observed emittance from LCD monitors. The emission distribution function E_s towards a general position $\mathbf{p}$ is a Gaussian $\mathcal{N}^E(\mathbf{x}; \mathbf{p})$ obtained by suitable scaling and translation of $\mathcal{N}^{\tilde{E}}$. Appendix A describes a simple calibration method for obtaining the parameters.

Unknown Response Curve   We eliminate gamma and other nonlinearities in the monitor emission as follows. Let $h(\mathbf{x})$ be the grayscale pattern to be displayed. We use an exposure time of three seconds, and display the time-varying, thresholded, pure black and white pattern $\tilde{h}(\mathbf{x};s) = H(h(\mathbf{x}) - s)$, where s is the elapsed exposure time scaled to $[0,1]$, and H is the step function. The effect is the same as if we had shown the original pattern using linear gray values.

4.3 Illumination Patterns

Our emission patterns are

$$g^j(\mathbf{x}) = w(\mathbf{x})\, \mathrm{e}^{-\sqrt{-1}\,\omega^{j\top}\mathbf{x}}, \qquad (4)$$

where $w(\mathbf{x})$ is a smooth window function. With this choice, the measurements $\mathbf{Z}_{i,j}$ give, for each point $\mathbf{p}$ on the sample, the 2D Fourier transform of the product

$$\mathbf{f}(\mathbf{x}; \mathbf{p}) := \rho(\mathbf{x}; \mathbf{p})\, E_s(\mathbf{X}(\mathbf{x}) \to \mathbf{p})\, w(\mathbf{x}) \qquad (5)$$

with respect to the primal variable $\mathbf{x}$, evaluated at the 2D frequency ω^j. Since the functions being transformed are real-valued and consequently their Fourier transforms are conjugate symmetric, each measurement determines the value of the Fourier transform at two points in the frequency plane.

Window Function   We modulate the complex exponentials by a smooth window function that falls off towards the the edges of the screen. If we did not window the illumination, we would be implicitly using a sharp rectangle window ($w(\mathbf{x}) = 1$ on the screen and $w(\mathbf{x}) = 0$ outside its extents). Its effect in the Fourier domain is a convolution of the rest of the terms by a sinc, which we could not handle analytically. We use a Gaussian window, so that the corresponding convolution can be easily computed in closed form because we approximate the rest of the terms by simple functions (Gaussian, affine). Specifically, our window $w(\mathbf{x})$ is a zero-mean Gaussian $\mathcal{N}^W(\mathbf{x})$ with covariance $\Sigma_W = \mathrm{diag}(1, (1/a)^2)$.

Complex Values   As the patterns contain both negative and imaginary values, they cannot be directly realized on a monitor. Thanks to linearity, we obtain a corresponding measurement by appropriately summing the four photos $g^{j,k}(\mathbf{x}) = w(\mathbf{x}) \max(0, \sin(\omega^{j\top}\mathbf{x} + k\pi/2))$ for $k = \{0,1,2,3\}$. The subtraction of opposite-phase images also cancels any common ambient illumination.

4.4 Properties of Fourier Measurements

Our measurements are point samples of the spectrum of $\mathbf{f}(\mathbf{x};\mathbf{p})$ (Equation 5). Clearly, if $\mathbf{f}$ was an arbitrary function with an arbitrary spectrum, a few samples of the spectrum would not enable us to deduce much about the reflectance. Our crucial observation is *that the spectra of typical reflectance functions are simple, smooth functions, and that glossiness, albedos, and relative strengths of diffuse and specular components map to directly observable properties of the spectra.* For instance, decay rate directly corresponds to glossiness, as is well known [Ramamoorthi and Hanrahan 2001]. The simple shapes enable our analytic model, which is carefully crafted to have an analytic Fourier transform, to be fit directly in the spectral domain, and to capture the shapes of the reflectance lobes using only a limited number of measurements.

To illustrate, let us examine the data visually. Figure 2 shows a photograph of a material sample, and four points $\mathbf{P}_1, \dots, \mathbf{P}_4$ with different reflectance properties highlighted. The plots show the magnitudes of Fourier coefficients of $\mathbf{f}(\mathbf{x}; \mathbf{P}_1), \dots, \mathbf{f}(\mathbf{x}, \mathbf{P}_4)$ on a slice through the horizontal frequency axis, i.e., with $g(\mathbf{x}) = w(\mathbf{x}) \exp\{-\sqrt{-1}\,\Omega_x x\}$ with varying Ω_x, on a log-log scale. The far right end of the range corresponds to 60 wavelengths over the screen. The measurements are shown as black dots. The solid lines denote our model fit after optimization, and they illustrate that the shapes are simple enough to be successfully captured from the relatively few, low-frequency samples only. Dashed lines illustrate the diffuse and specular components of the model. Gray dots denote points that were captured for verification purposes, and not used in the model fit.
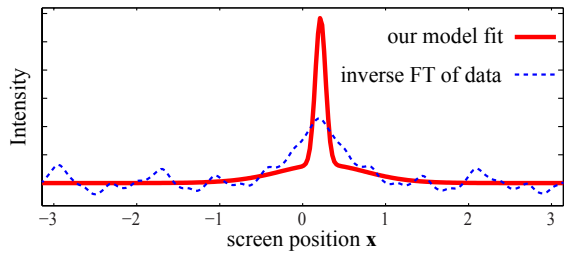
Figure 3: *A slice through the reflectance function over the screen plane (primal domain) for a single surface point.* Blue curve: *inverse Fourier transform of the data.* Red curve: *our model fit.*

The plots illustrate how different reflectance properties are evident in the spectrum. Point $\mathbf{P}_1$ is glossy, but contains a noticeable diffuse component. We observe that the spectrum inflects at around frequency 6, where the diffuse power trails off, leaving the glossy component to decay around our maximum captured frequency. Point $\mathbf{P}_2$ is mostly diffuse. We observe that its spectrum is essentially contained in the frequencies 0–5, with the rest of the observations well below the noise floor. Point $\mathbf{P}_3$ has a diffuse and dull glossy component. The overall shape of the spectrum resembles that of the diffuse point — the spectrum of $\mathbf{P}_3$ has decayed to the noise floor by frequency 13, in comparison to frequency 6 for the diffuse point $\mathbf{P}_2$. The closely related shapes of diffuse and dull glossy lobes would make heuristic diffuse-specular separation difficult, but our model and optimizer, explained in the following sections, correctly identify them. Point $\mathbf{P}_4$ is highly specular, which shows up in the spectrum of its narrow lobe: the magnitude hasn't noticeably decayed by the highest frequency we capture, indicating there is significant energy left in the unmeasured part of the spectrum. The shape, however, remains simple and smooth.

The fact that we only take a few point samples of the continuous, infinite spectrum also indicates that the measurements are, in themselves, not a good representation for reflectance functions even for the single outgoing slice for which the measurements were taken: we cannot interpret the data as a complete Fourier transform of **f**. Figure 3 shows the reflectance function of a point with both diffuse and relatively sharp specular contributions. The blue curve is obtained by interpreting the spectral samples as basis coefficients and evaluating the corresponding truncated Fourier sum. This results in severe ringing due to the energy in the frequencies not captured by the measurements. Furthermore, the coefficients themselves do not directly determine the normals, glossiness, and so on. In contrast, a model fit to the same data yields much more meaningful parameters. This is consistent with the findings reported by Ghosh et al. [2007].

Position and Phase   The position where the reflectance lobe of a surface point **p** intersects the screen is mostly determined by the unknown local surface normal and the known relative position of the point and the screen. The position is consequently a strong clue that can be used to infer normals. Fourier measurements have the desirable property that translation and shape are separated into magnitude and phase. As shown above, the shape of a narrow lobe is captured well in the magnitude plot. In addition, however, also its location can be resolved very well from the phase information, which enables us to resolve both position and shape accurately.

Dynamic Range   Our lighting patterns are effectively area sources with different spatial textures. It is not hard to see that the dynamic range of a scene under area lighting is reduced compared to point sources: as reflectance is qualitatively a convolution of lighting and BRDF [Ramamoorthi and Hanrahan 2001], the energy of narrow reflectance lobes is spread out by low-frequency illumination. This

enables us to take single exposures rather than use bracketing.

5 Reflectance Model

This section describes our analytic model for the apparent reflectance function ρ from Equation 1. We develop a model in the primal domain, because the necessary geometric considerations are simple, but carefully construct the model so that it is easy to Fourier transform analytically, as required by the subsequent optimization. We start by expanding the apparent reflectance function into diffuse and specular components as

$$\rho(\mathbf{x};\mathbf{p}) = \mathbf{a}_d(\mathbf{p})\rho_d(\mathbf{x};\mathbf{p},\mathbf{n}) + \mathbf{a}_s(\mathbf{p})\rho_s(\mathbf{x};\mathbf{p},\mathbf{E},\sigma,\gamma,\mathbf{n}), \quad (6)$$

where $\mathbf{a}_d$, $\mathbf{a}_s$ are the diffuse and specular albedos, and ρ_d, ρ_s are diffuse and specular reflectance lobes that depend on the relative positions of the camera, the surface point, the screen point, local surface normal **n**, specular roughness σ and kurtosis γ, which intuitively controls the pointiness of the specular lobe.

5.1 Diffuse BRDF

The diffuse BRDF is assumed to be Lambertian. Absorbing the incident normal cosine and the Jacobian of the reparameterization of the incoming solid angle onto the emitter screen area yields

$$\rho_d(\mathbf{x};\mathbf{p},\mathbf{n}) = \frac{d(\mathbf{p})}{\pi D(\mathbf{x};\mathbf{p})^4}[\mathbf{n}^\top \mathbf{T}\mathbf{x} + \mathbf{n}^\top(\mathbf{t}-\mathbf{p})], \quad (7)$$

where $d(\mathbf{p})$ is the orthogonal distance from **p** to the screen, and $D(\mathbf{x};\mathbf{p})$ is the distance from **p** to $\mathbf{X}(\mathbf{x})$. The bell-shaped function $1/D(\mathbf{x};\mathbf{p})^4$ originates from the inverse square distance and the normalization of the point-to-screen vectors in the two cosines. We approximate it with a mixture of three Gaussians $\mathcal{N}^{D,k}(\mathbf{x};\mathbf{p})$ as described in Appendix B, yielding

$$\rho_d'(\mathbf{x};\mathbf{p},\mathbf{n}) = \underbrace{\frac{d(\mathbf{p})}{\pi}\left[\mathbf{n}^\top \mathbf{T}\mathbf{x} + \mathbf{n}^\top(\mathbf{t}-\mathbf{p})\right]}_{:=\mathcal{A}(\mathbf{x};\mathbf{p},\mathbf{n})}\sum_{k=1}^{3}\mathcal{N}^{D,k}(\mathbf{x};\mathbf{p}). \quad (8)$$

The leading affine function $\mathcal{A}(\mathbf{x};\mathbf{p},\mathbf{n})$ encodes the effect of the local surface normal. The Gaussians do not depend on any of the unknown parameters, only the geometric configuration.

5.2 Specular BRDF

Our specular model is an approximation to a microfacet BRDF that is symmetric around the halfway vector. It is constructed such that the projection of the lobe onto the screen is a sum of Gaussians. This is vital for our ability to analytically compute its integral with the illumination patterns. Our model captures the stretching and scaling effects related to the screen projection, as well as the incidence angle-dependent narrowing caused by the half-angle parameterization. At moderate to high gloss, it closely approximates microfacet models such as the Blinn model [Blinn 1977].
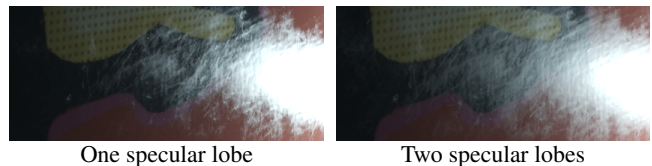

One specular lobe   Two specular lobes

Figure 4: *A single specular lobe (left) cannot capture heavy-tailed reflectance lobes of e.g. layered materials. A two-lobe model (right) matches observed behavior much better. All results use two lobes.*
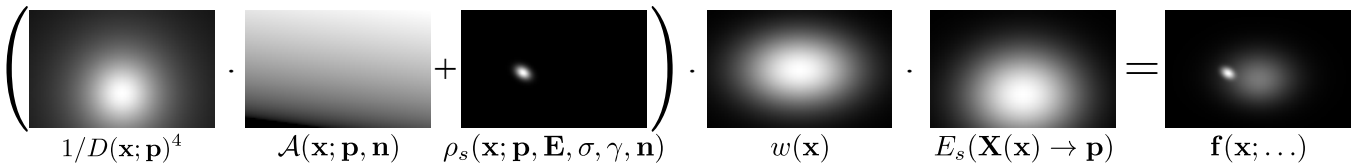
Figure 5: *The primal domain reflectance components visualized in screen coordinates. The albedos $\mathbf{a}_d, \mathbf{a}_s$ have been omitted for clarity.*

We use two lobes because materials often exhibit layers and other effects that result in pointiness and heavy tails of the specular lobe, which a single, simple lobe is unable to capture [Ngan et al. 2005]. This is evident in our data, where the magnitude spectra of the specular lobes rarely show the log-quadratic decay that simple parametric models predict. (Pointier lobes with heavier tails exhibit more log-linear behavior.) The importance of modeling this behavior has been recognized in recent work (e.g. [Löw et al. 2012]) on BRDF models aimed at data fitting. We introduce a kurtosis parameter γ that controls the pointiness of the lobe by adjusting the relative widths of the two lobes while keeping its overall glossiness fixed. Specifically, our specular lobe model is a mixture of two Gaussians $\rho_s = \sum_l \mathcal{N}^{S,l}(\mathbf{x};\mathbf{p},\mathbf{E},\sigma,\gamma,\mathbf{n})$ on the screen plane. The means and standard deviations of the Gaussians are derived from σ, γ, $\mathbf{n}$ and the geometric configuration as detailed in Appendix C. Our lobes approximate the Blinn model closely, with exponent $\alpha = \sigma^{-2}$. The difference is small at typical exponents, as demonstrated by our results. We use the Blinn model in all our renderings. The difference between one and two lobes is illustrated in Figure 4.

Fresnel Term The apparent brightness of specular reflection depends on the angles of incidence and exitance through Fresnel effects. The lobe we observe contains these effects baked in. To model the apparent brightening and dimming with different elevation angles, we assume that the surface is a dielectric with IOR of roughly 1.65, and approximately cancel the Fresnel effect by dividing the specular albedo by the corresponding Schlick Fresnel prediction with $F_0 = 0.06$. In renderings, we use the same assumption to reintroduce the effect for novel viewing and incidence angles. Validation results are shown for our datasets in Figure 10.

5.3 Rendering

The model data (albedos, normals, glossiness, kurtosis) is extremely easy to use in a renderer. The diffuse albedo is standard, and the specular model is a sum of two Fresnel-modulated Blinn lobes.

6 Full Image Formation Model

Combining everything in the preceding sections, our full image formation model is

$$\mathbf{L}^j(\mathbf{p} \to \mathbf{E}) = \int_{\mathbb{R}^2} \mathbf{f}'(\mathbf{x};\mathbf{p},\mathbf{a}_d,\mathbf{a}_s,\sigma,\gamma,\mathbf{n})\, g^j(\mathbf{x})\, d\mathbf{x}, \qquad (9)$$

where

$$\mathbf{f}'(\mathbf{x};\mathbf{p},\mathbf{a}_d,\mathbf{a}_s,\sigma,\gamma,\mathbf{n}) =$$
$$\left[\mathbf{a}_d \mathcal{A}(\mathbf{x};\mathbf{p},\mathbf{n}) \sum_{k=1}^{3} \mathcal{N}^{D,k}(\mathbf{x};\mathbf{p}) + \mathbf{a}_s \sum_{l=1}^{2} \mathcal{N}^{S,l}(\mathbf{x};\mathbf{p},\mathbf{E},\sigma,\gamma,\mathbf{n}) \right]$$
$$\mathcal{N}^{W}(\mathbf{x}) \mathcal{N}^{E}(\mathbf{x};\mathbf{p}) \quad (10)$$

is our approximation to $\mathbf{f}(\mathbf{x};\mathbf{p})$. The observed radiance is an inner product, taken over the screen, between the displayed 2D pattern g^j and the function $\mathbf{f}'$ that encodes the unknown reflectance function, known geometric quantities and screen emission, using Gaussian

approximations to several of the terms. Figure 5 illustrates the components.

The terms related to the diffuse Gaussians, the window function and emission distribution only depend on the geometric configuration and not any of the unknown parameters, and we can therefore analytically compute the multiplications $\mathcal{N}^{DWE,k}(\mathbf{x};\mathbf{p}) := \mathcal{N}^{D,k}\mathcal{N}^{W}\mathcal{N}^{E}$ and $\mathcal{N}^{WE}(\mathbf{x};\mathbf{p}) := \mathcal{N}^{W}\mathcal{N}^{E}$ in a pre-processing pass.

Fourier Transform We need the Fourier transform of the model to predict the values of the measurements $\mathbf{Z}_{i,j}$ by the unknowns. This is relatively simple to obtain due to the closedness of Gaussians under multiplication, convolution, and Fourier transform, and a simple formula we've derived for the Fourier transform of the affine function times a Gaussian. Denoting the Fourier transformed functions by a hat, the result is

$$\hat{\mathbf{f}}'(\omega;\mathbf{p},\mathbf{a}_d,\mathbf{a}_s,\sigma,\gamma,\mathbf{n}) =$$
$$\mathbf{a}_d \sum_{k=1}^{3} \hat{\mathcal{A}}^{DWE,k}(\omega;\mathbf{p},\mathbf{n})\hat{\mathcal{N}}^{DWE,k}(\omega;\mathbf{p})$$
$$+ \mathbf{a}_s \sum_{l=1}^{2} \hat{\mathcal{N}}^{S,l}(\omega;\mathbf{p},\mathbf{E},\sigma,\gamma,\mathbf{n})\hat{\mathcal{N}}^{WE}(\omega;\mathbf{p}), \qquad (11)$$

where

$$\hat{\mathcal{A}}^{DWE,k}(\omega;\mathbf{p},\mathbf{n}) = \frac{d(\mathbf{p})}{\pi}\left(-\sqrt{-1}\Sigma_{DWE,k}\mathbf{T}^{\top}\mathbf{n}\right)^{\top}\omega$$
$$+ \frac{d(\mathbf{p})}{\pi}\left(\mu_{DWE,k}^{\top}\mathbf{T}^{\top}\mathbf{n} + \mathbf{n}^{\top}(\mathbf{t} - \mathbf{p})\right), \quad (12)$$

and $\Sigma_{DWE,k}$ and $\mu_{DWE,k}$ are the covariances and means of the respective Gaussians. Despite an uninviting initial appearance, the result is simple to evaluate for any measured position $\mathbf{p}^i$ at any 2D frequency ω^j. In particular, we estimate that evaluating this formula is several orders of magnitude faster than brute force numerical integration of Equation 9 using a traditional BRDF model.

7 Optimization

We pose the problem of finding reflectance parameters given the data as a Maximum A Posteriori (MAP) problem, and solve it using nonlinear optimization (Levenberg-Marquardt; derivatives are computed using finite differences). We minimize the sum of squared differences between the model predictions and the measurements at every surface point, frequency, and color channel simultaneously; real and imaginary components are stacked as real-valued pairs. The specular glossiness parameter σ can reach values that are very close to zero, but must not be negative. To improve the scaling and to prevent negative values, we optimize $\log\sigma$ instead. Similarly, we optimize $\log\gamma - 1$ to ensure $\gamma > 1$.

Initial Guess The optimization starts from an initial guess: normals are initialized as perpendicular to the sample plane, and both glossiness and kurtosis are set to a low, constant value. Initial values for diffuse and specular albedos are heuristically determined from the data. As seen in Figure 2, the contribution of the diffuse component typically fades out around frequency 6. We use the magnitude of the measurement at this frequency to compute a specular albedo
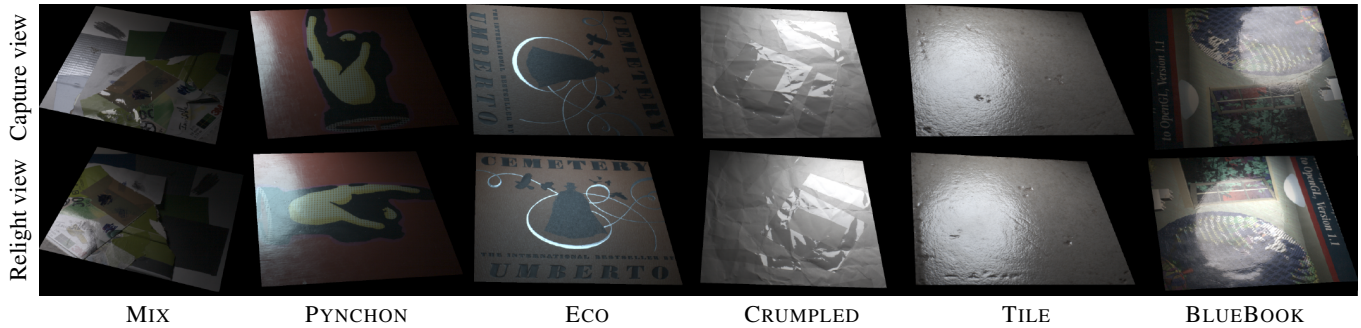
Figure 6: *Datasets captured.* Top row: *Viewpoints used for data capture.* Bottom row: *Viewpoints used for relighting for ground-truth comparison in Figure 9. All photographs are roughly clipped to the object region captured for reconstruction. The photos in the top row are for illustrative purposes only, and were not used as an input to the solver.*

initial guess. The diffuse albedo initial guess is obtained similarly by also considering the zero frequency data. See Appendix D for the detailed formulas.

Point-wise Priors   To avoid incorrect local minima, such as a very wide specular lobe that tries to fit to a diffuse observation, we state Gaussian-distribution priors that bias all model parameters toward plausible ranges. These point-wise priors are implemented as data-independent additional residuals in the error metric; see Appendix E for the implementation details and the numerical settings.

Multi-Resolution Scheme   We initially condition the optimization through data-dependent smoothness priors (Section 7.1) that we gradually lift in the course of a hybrid multi-resolution scheme. We start the model fit at the very low resolution of 32×32 pixels, strongly penalizing spatial variation of the reflectance variables; we also use an integrability prior that penalizes normal maps that cannot be formed as gradients of a height field, by penalizing the magnitude of the curl of the normal map (as normal maps originating from height fields have zero curl). We then sequentially upsample the solution by doubling its resolution, at each step relaxing the smoothness and integrability priors, and running a few[1] iterations of the optimizer, until we reach a globally consistent but low resolution (128×128) intermediate solution. We use this intermediate solution as initial guess to a high-resolution optimization at 512×512 pixels, where we drop the smoothness and integrability priors entirely and optimize the data fit independently at each point. For performance reasons, we do not optimize at the full data resolution of 1024×1024 but use a data-dependent upsampling scheme (Section 7.2) that preserves fine detail. The same upsampling scheme is also used in the multiresolution upsampling step.

7.1 Data-Dependent Smoothing

Our smoothness prior is implemented as a weighted sum of squared finite differences of the model parameters. The weights are chosen according to a heuristic that determines how plausible a smoothness assumption locally is: neighboring points with similar measurement data are expected to obtain similar values for the unknown reflectance parameters; however, when the data exhibits rapid spatial variation, the heuristic relaxes the smoothness constraint.

We compute approximate conservative upper bounds on the possible variation of the albedos using suitably scaled spatial derivatives of the measured data itself (Appendix E). These bounds are used as weights of the spatial smoothness priors. This prevents the solver from introducing rapid albedo variation in smooth regions, while al-

lowing sharp step edges where the data contains them. For glossiness and kurtosis maps, similar heuristics are more difficult to formulate, and we simply use a constant-weight penalty on spatial variation.

7.2 Upsampling

We upsample the intermediate and final solutions using the following scheme that, in spirit, resembles joint bilateral upsampling [Kopf et al. 2007]. Let I be the spatial index of a point in the high resolution data $\mathbf{Z}_I^{\text{hi}}$, and i an index in the low resolution data $\mathbf{Z}_i^{\text{lo}}$ and solution image $\mathbf{U}_i^{\text{lo}}$. We solve the following problem separately at each high-resolution spatial index I. Let $\mathbf{q}$ be a neighborhood of the corresponding pixel i in the low-resolution solution (we use a 5×5 window). We find a non-negative least squares approximation to $\mathbf{Z}_I^{\text{hi}}$ by solving the problem

$$\underset{\mathbf{w}}{\text{argmin}} \left\| \mathbf{Z}_I^{\text{hi}} - \sum_{m \in \mathbf{q}} \mathbf{w}_m \mathbf{Z}_m^{\text{lo}} \right\|^2 \quad \text{s.t.} \quad \mathbf{w} \geq 0 \qquad (13)$$

with an additional soft constraint $\sum \mathbf{w} = 1$, i.e., we find a local explanation of the high-resolution data as a convex combination of the low-resolution data. The upsampled solution is obtained simply as a linear combination of the low-resolution data using the same weights: $\mathbf{U}_I^{\text{hi}} = \sum_{m \in \mathbf{q}} \mathbf{w}_m \mathbf{U}_m^{\text{lo}}$.

8 Results

We demonstrate reconstructions for six datasets. We captured the input images using a Canon 5D Mark II and illuminated the scenes with a 24-inch IPS display tilted by roughly $45°$ (Figure 1, left), with the exception of the BLUEBOOK dataset, which was illuminated by a Lenovo laptop screen (Figure 8). The Fourier transform was sampled at frequencies $\{1, 2, 3, 4, 6, 10, 20, 40\}$ on the *x*- and *y*-axis of the frequency plane, and the corresponding integer frequencies $\{(1, 1), (2, 2), ...\}$ on the two diagonals, and at $(0, 0)$. In total, 131 photographs were captured per material sample (including the geometric calibration image).

The different datasets and the viewpoints used for the capture and for the ground truth photos are shown in Figure 6. The resulting model fits are shown in Figure 7, by individual component. All solutions have been computed with identical optimizer settings, i.e., with no scene-dependent parameter tuning, to resolution 1024×1024 as described in Section 7.

Figure 9 presents relighting results for novel views and illumination conditions for all datasets: we used the capture setup to simulate light sources at different positions, by displaying a white disk on black on the screen, and simulated the same conditions when rendering

---

[1]Iterations per resolution: 32×32: 12; 64×64: 3; 128×128: 3; 512×512: 4.
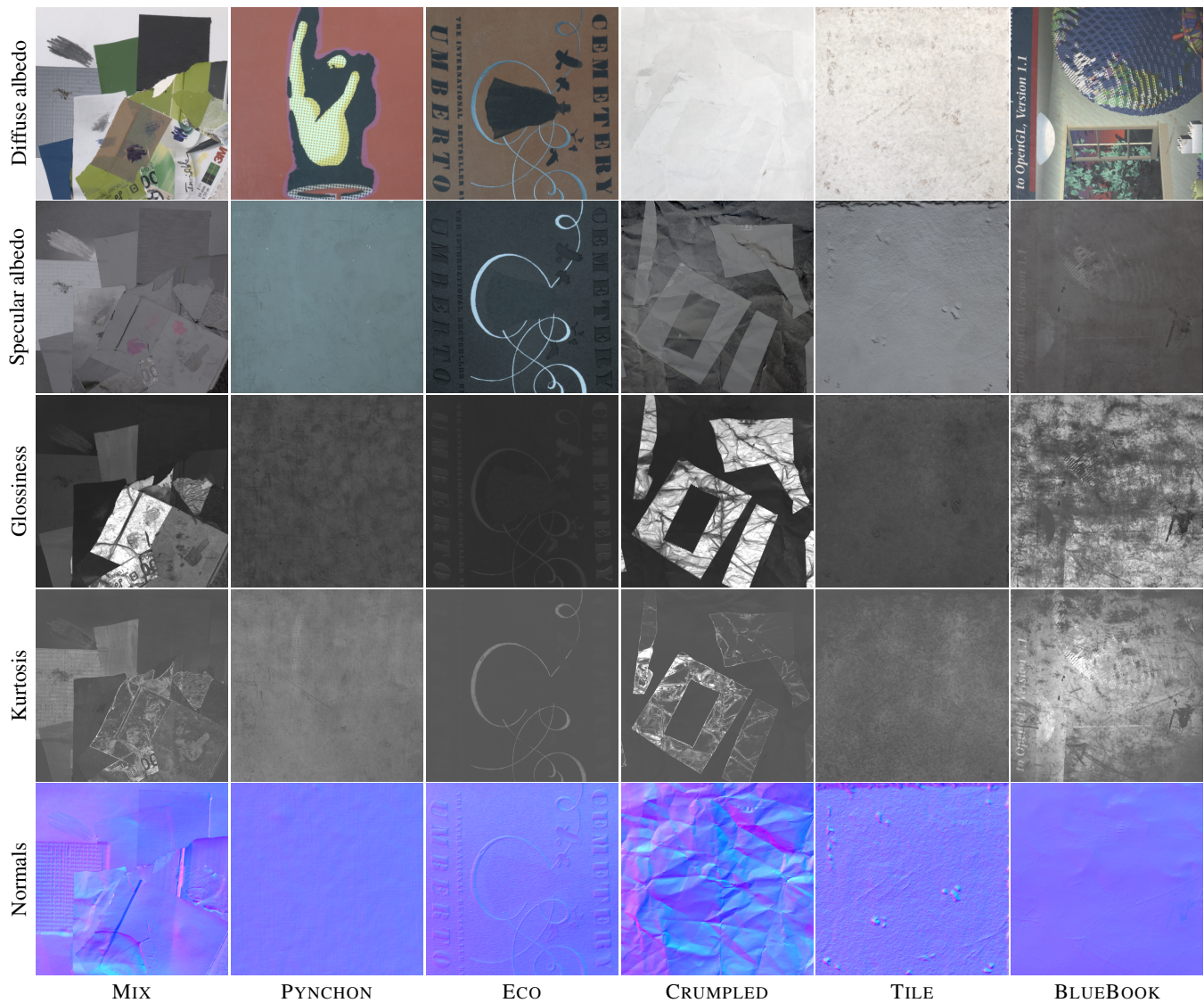
Figure 7: *Model fits. The normal map has been exaggerated by a factor of two for display purposes. The glossiness image shows the values* $1/\sigma$. *All images can be found on the project website for closer inspection.*

the model for comparison. (Note that there are slight differences in geometric alignment due to imperfect registration.) Many of the effects are more clearly seen under moving illumination and viewpoints. Please see the supplemental video for further results.

The first dataset MIX contains multiple layers of different types of cardboard, paper and tape, with various pencil markings, scratches, steep slopes and other features. Our algorithm has correctly resolved the wide range of reflectance from nearly diffuse paper to almost mirror-like packaging tape. The rich normal detail in various surface types, layerings, folds and scratches has been convincingly captured. Note also the interesting and correctly resolved specular color effect on the marker markings. The sample deviates from planarity by several millimeters and includes notable self-shadowing. The model gracefully absorbs these effects, and the result remains plausible.

PYNCHON is a book cover with a shiny, worn plastic finish that contains subtle variations in normals (scratches) and glossiness (scuffs). This dataset benefits particularly much from the two-lobe model, as the specular lobe has slowly falling of tails.

Figure 8: *A laptop screen was used for the* BLUEBOOK *set.*

ECO is a book jacket featuring debossed letters and a shiny swirl. The heavy paper has subtle normal variation, and a very wide specular component which we correctly capture.

The fourth dataset CRUMPLED is a crumpled paper with pieces of highly specular tape on it. It demonstrates the recovery of diffuse normals, and the continuity of the normal map across the diffuse-
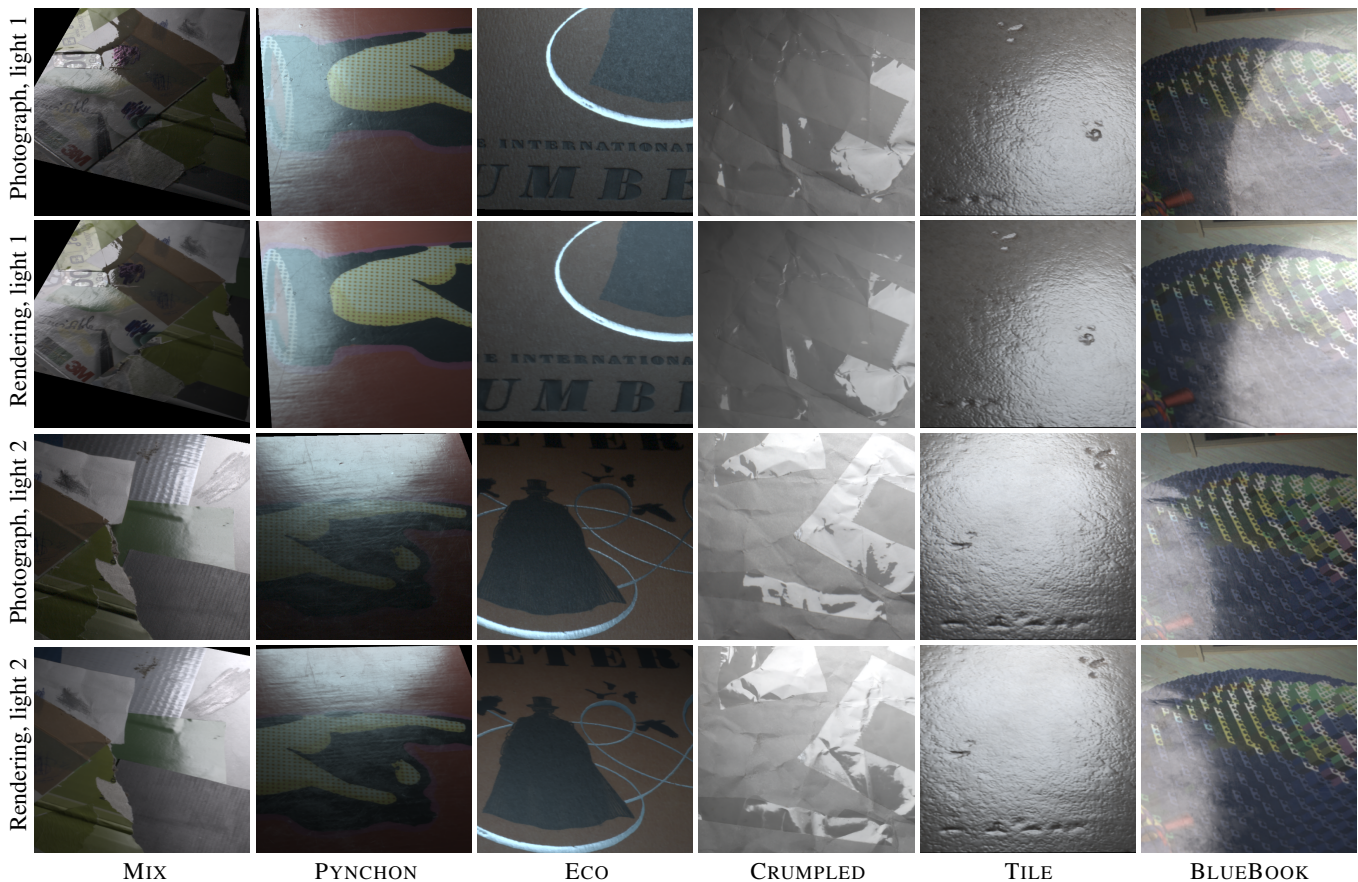
Figure 9: *Still frames comparing photographs and renderings of the datasets under novel lighting conditions. The viewing angle is the same for both lighting configurations, but the pictures are cropped to concentrate on different parts of the sample. The view has been rotated by 90 or 180 degrees from the capture angle (cf. Figure 6).*

specular edges. Some very smooth specularity has been resolved into the paper as well. The solution also shows extremely sharp specularity in the tape. The smallest resolved glossiness values are around 0.004, corresponding to a Blinn exponent of roughly 60 000. Accordingly, the sharp boundary of the illuminating disk becomes visible in the reflection, and is equally sharp in the photograph and in the reconstruction. Subtle differences in apparent shape of the highlights are at least partly due to misregistration of the two views and due to missing parallax (we do not reconstruct geometry).

TILE shows a moderately specular tile with both repeating texture as well as steep holes. Both features have been resolved well. Some cross-talk is observed in the deeper, possibly interreflecting and self-shadowing dents, but the errors are imperceptible in renderings.

BLUEBOOK is again a book cover with rich wear and tear, interesting grooves resulting from a heavy-handed person having written on a piece of paper on top of the book, and a near mirror-like lamination. In contrast to the previous datasets, it has been captured with the laptop-based setup. This could explain the small amount of cross-talk across reflectance components, unique to this dataset, suggesting that the laptop setup is of slightly lower quality than the system based on the desktop screen. In fact, upon visual inspection the spatial emission characteristic of the laptop screen used for capturing BLUEBOOK appears more complex than that of desktop screen, offering a probable error source. Nevertheless, interesting normal detail and scuffing are still captured.

Lastly, Figure 10 shows a novel, grazing viewpoint for each dataset,

illustrating how the post-fit addition of the Fresnel term leads to results consistent with the real reflectance.

8.1 Performance

We computed the results and measured the performance of our current Matlab implementation on a quad-core Intel Core i7-Q820 1.73 GHz laptop with 12 gigabytes of RAM. The full optimization procedure takes approximately two hours per dataset. The three significant stages of the computation are the regularized low-resolution phase (15 minutes), the high-resolution final solve (1.5 hours), and the final upsampling (10 minutes). Most of the computation time is spent evaluating the model and its derivatives.

8.2 Discussion

As the results show, our method convincingly captures a range of different materials on near-planar samples. Through the explicit consideration of geometric terms and the windowing of the Fourier basis, we even support a wide range of normal variations, as long as the reflectance lobe reaches the monitor and near-planarity is met on a coarse scale. (See, for instance, MIX and TILE that exhibit areas with steep slopes.)

While there is no fundamental reason why our model could not be extended to moderate anisotropy (by adjusting the half-vector lobes accordingly), we found that the current degrees of freedom are just in the range that allows a stable model fit with the given number of
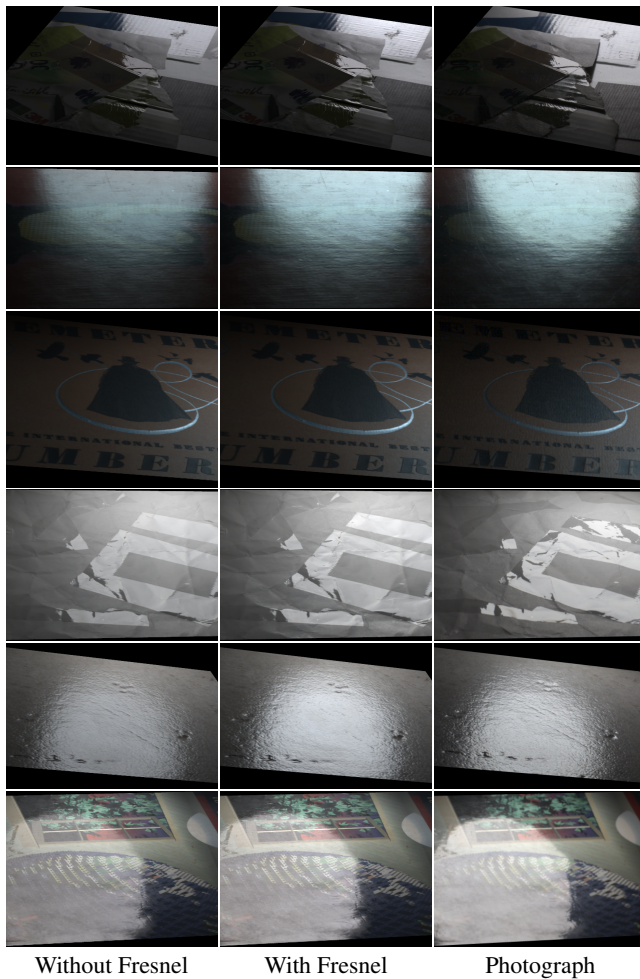
Without Fresnel   With Fresnel   Photograph

Figure 10: *Our simple approximation to Fresnel effects is sufficient to capture the most obvious effects of incidence and exitant angles on the strength of specular reflectance. The pictures shows each captured dataset rendered and photographed from a significantly lower viewing angle than the 45 degree angle used for capture.* Left: *our rendering, without accounting for changes in Fresnel reflectance.* Middle: *our result, with Fresnel approximation.* Right: *photograph.*

input photographs. Rather than increasing the number of measured basis functions we decided against modeling anisotropy, to keep the measurement times acceptable.

In general, reflectances that significantly break our model assumptions (anisotropy, shadowing, interreflections, etc.) will have an unexpected expression in the Fourier-space measurements and might misguide the model fit. Figure 11 shows our results for two such materials. While the optimizer has converged to a solution that plausibly captures many of the features in the materials, the model is unable to properly express the anisotropy and visibility effects. In particular, the optimizer has introduced spurious specularity to explain the self-shadowing and extreme normal variations of the towel. The highly anisotropic specular lobe of the brushed metal has been replaced with a moderately glossy isotropic lobe. Other unmodeled effects, such as back-scattering, simply do not manifest themselves under the chosen measurement geometry and would still allow the method to recover good normals and a plausible specular highlight. For multi-lobe reflectances, however, such as in fur or hair, we would not expect the system to produce plausible results.
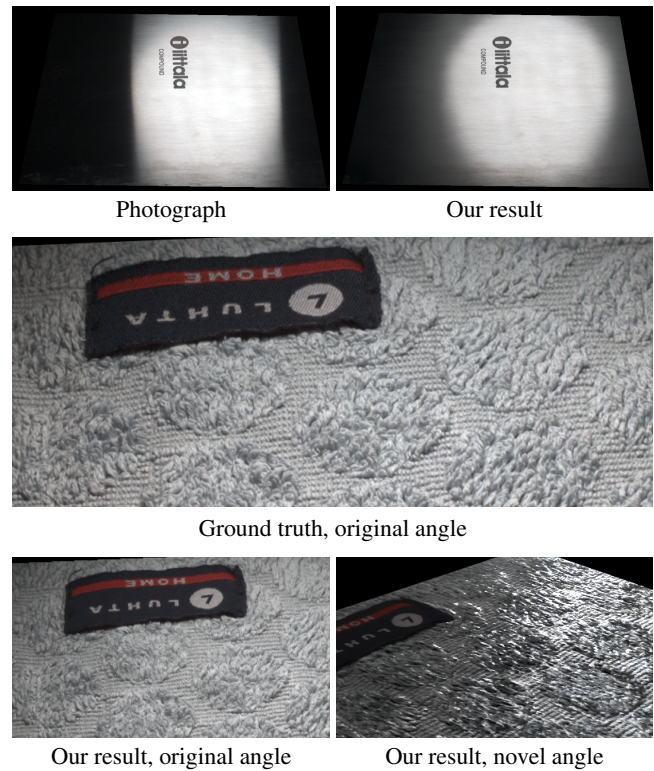
Photograph   Our result

Ground truth, original angle

Our result, original angle   Our result, novel angle

Figure 11: *Failure cases. Top left: a photograph of an anisotropic brushed metal surface. The reflectance lobe is very narrow in the horizontal direction, and wider in the vertical direction. Top right: our isotropic lobe model captures overall color, brightness, and texture, but is unable to reproduce the anisotropy. Middle: a towel (photograph). Bottom left: our model fit reproduces the appearance from the original capture direction, but does not generalize well to novel views (bottom right). The complex visibility caused by the fibers and bumps is mistaken for specularity in the model fit.*

Due to the lack of a built-in Fresnel model component, some polarization effects may become baked into our reconstructions. We consider this to be the biggest shortcoming of our current method. Nevertheless, our rendering heuristic reproduces the qualitative effects and yields good matches with photographs in our results. In particular, no parameters were tuned per-material.

9   Conclusion

We presented a novel method to acquire the SVBRDF of near-planar objects using commodity hardware. We believe to be the first to produce consistent normals and reflectance for both diffuse and specular surfaces using a pervasively available capture system. The simplicity of our setup and the observed robustness of our optimizer make us believe our system enables spatially varying reflectance capture for a wide audience.

Acknowledgments

A Calibration

Monitor calibration We obtain the monitor calibration (Section 4.2) as follows. This procedure only needs to be performed once for any given monitor. We capture a dataset from a sheet of plain white paper, which we assume to be roughly Lambertian. We then run a version of the optimization problem with the diffuse albedo fixed to constant one, specular albedo to zero, and the normals to $[0\ 0\ 1]^\top$, and the glossiness and kurtosis ignored. We choose an evenly spaced subset of a few hundred pixels and optimize to find the four monitor Gaussian parameters that give the best fit to the data there. This procedure also fixes the radiometric scale so that the albedo variables will be in proportion to the calibration paper albedo.

Geometric calibration We use simple geometric calibration method to measure the required screen transformation matrix $\mathbf{T}$, screen origin vector $\mathbf{t}$ and the camera position $\mathbf{E}$. This requires only a single photo taken from the same configuration as the data images. We require that a part of the screen must be visible in these photos, and display an image of a white square on the monitor. Additionally, we place a mirror and a square marker on the material sample plane. This image is enough to compute the calibration and to rectify the sample plane in the photos.

B Diffuse Model

We seek to approximate $1/D(\mathbf{x};\mathbf{p})^4$ by a mixture of Gaussians. We first approximate a prototype function $U(\mathbf{x}) = 1/(\mathbf{x}^\top\mathbf{x}+1)^2$ by three zero-mean Gaussians with mixing coefficients and co-variances $\alpha_{U,1} = 1.0391, \alpha_{U,2} = 1.5027, \alpha_{U,3} = 0.4583, \Sigma_{U,1} = 1.5617^2\mathbf{I}, \Sigma_{U,2} = 0.7138^2\mathbf{I}$ and $\Sigma_{U,3} = 0.3975^2\mathbf{I}$. Conveniently, $U(\mathbf{x})$ can be translated and uniformly scaled to yield $1/D(\mathbf{x};\mathbf{p})^4$ for any $\mathbf{p}$. The final approximation is the mixture $\sum_k \mathcal{N}^{D,k}(\mathbf{x};\mathbf{p})$ with mixing coefficients $\alpha_{D,k}(\mathbf{p}) = d(\mathbf{p})^{-2}\alpha_{U,k}$, means $\mu_{D,k}(\mathbf{p}) = \mathbf{T}^\top(\mathbf{p}-\mathbf{t})$ and covariances $\Sigma_{D,k}(\mathbf{p}) = d(\mathbf{p})^2\Sigma_{U,k}$.

C Specular Model

We now describe how the parameters of the two Gaussians $\mathcal{N}^{S,l}$ (Section 5.2) are obtained. Our goal is to formulate a Gaussian approximation to a microfacet distribution in a half-vector parameterization, and map the corresponding lobe onto the screen plane.

A general half-angle parameterized BRDF lobe has the form $f(\omega_i, \omega_o) = g_{\mathbf{n}}(\mathbf{h})$. For example, in the Blinn model, $g_{\mathbf{n}}(\mathbf{h}) = (\mathbf{h}^\top\mathbf{n})^\alpha$. We parameterize the half-vector on the plane $\Pi_{\mathbf{p},\mathbf{n}}$ orthogonal to $\mathbf{n}$ at unit distance from $\mathbf{p}$, with coordinates $\mathbf{y}$, such that $\mathbf{h} = P_Y(\mathbf{y})$ is the projection from $\Pi_{\mathbf{n},\mathbf{p}}$ onto the unit sphere centered at $\mathbf{p}$. With $\mathbf{p}$, $\mathbf{n}$, and ω_o fixed, $\mathbf{h}$ is a function of the incident angle ω_i only (and vice versa), and because the incident angle is determined by the screen point $\mathbf{x}$, the screen and the half-vectors are in 1-1 correspondence: $\mathbf{y} = T(\mathbf{x}) := P_Y^{-1}(N(H_{\omega_o}(P_\mathbf{X}(\mathbf{x}))))$, where $P_\mathbf{X}(\mathbf{x})$ determines ω_i from $\mathbf{x}$, H_{ω_o} averages a vector with ω_o, and N normalizes the result. The function T maps the screen point $\mathbf{x}$ onto a point $\mathbf{y}$ on $\Pi_{\mathbf{p},\mathbf{n}}$. The lobe, parameterized over the plane, is $g'_{\mathbf{n}}(\mathbf{y}) := g_{\mathbf{n}}(P_Y(\mathbf{y}))$, and it inherits its radially symmetric bell shape from $g_{\mathbf{n}}$.

We model a microfacet distribution $g'_{\mathbf{n}}$ on the normal plane by the average of two isotropic zero-mean Gaussians. Their standard deviations are determined by glossiness σ and the kurtosis parameter γ by $\sigma_1 = c\sigma/\gamma$ and $\sigma_2 = c\sigma\gamma$, where the correction factor $c = \sqrt{2/(\gamma^{-2}+\gamma^2)}$ ensures that the standard deviation of the summed lobe is σ. At $\gamma = 1$ the lobes coincide and reduce to the single-lobe case. Higher values give pointier lobes. The de-

sired BRDF lobe on the screen plane is the corresponding sum of Gaussians mapped back onto the screen through the first-order approximation of the mapping T^{-1} evaluated at $\mathbf{y} = (0,0)$, the center of the microfacet distribution. The affine approximation ensures the result is still a sum of Gaussians over the screen, with the main stretching and scaling effects, e.g., the narrowing of off-specular reflection, included.

D Initial guess

Let $\omega_6 = [6\ 6]^\top$, $\omega_0 = [0\ 0]^\top$, $\sigma_0 = 0.03$, $\gamma_0 = 1.03$, $\mathbf{n}_0 = [0\ 0\ 1]^\top$. Our specular albedo initial initial guess at point $\mathbf{p}_i$ is $\mathbf{a}_{s,0} := |\mathbf{Z}_{i,\omega_6}|/|\hat{\mathbf{f}}'(\omega_6;\mathbf{p}_i,\mathbf{0},\mathbf{1},\sigma_0,\gamma_0,\mathbf{n}_0)|$. Note that it reduces to the true albedo, assuming $\rho_d(\omega_6) = 0$ and σ_0, γ_0 and $\mathbf{n}_0$ are correct. Otherwise the value is a rough approximation to the specular albedo. By similar considerations, we use the initial guess $\mathbf{a}_{d,0} := (\mathbf{Z}_{i,\omega_0} - |\mathbf{Z}_{i,\omega_6}|)/\hat{\mathbf{f}}'(\omega_0;\mathbf{p}_i,\mathbf{1},\mathbf{0},\sigma_0,\gamma_0,\mathbf{n}_0)$ for the diffuse albedo.

E Priors

The residual we optimize is a stacked vector of all the data fit terms and the priors. The data fit vector contains the difference between the model prediction and the data, for every pixel, frequency, color channel and complex number component, weighted by the factor 160. All of our priors are specified as linear functions of the unknowns.

For every unknown $\mathbf{u}^k$, we specify the expected value μ^k and standard deviation v^k, reflecting our beliefs about plausible range of values, by adding a residual $(\mathbf{u}^k - \mu^k)/v^k$. The values we use are $\mu = 0, v = 1$ for diffuse albedo, $\mu = 0, v = 0.5$ for specular albedo, $\mu = -3, v = 1$ for $\log\sigma$, $\mu = -1, v = 0.5$ for $\log\gamma - 1$, and $\mu = 0, v = 0.1$ for normals.

In the low-resolution, spatially regularized stage, we construct the finite difference matrices $\mathbf{D}_x$ and $\mathbf{D}_y$. The normal map integrability prior is specified by the residuals $(-\mathbf{D}_y\mathbf{n}_x + \mathbf{D}_x\mathbf{n}_y)/0.01$, where $\mathbf{n}_x$ and $\mathbf{n}_y$ are the stacked vectors of all normals in tangent parameterization; this penalizes the finite difference curl of the normal map. For log glossiness and kurtosis, we add a residual consisting of finite differences in both directions, with unit weighting.

The spatially-varying albedo smoothness prior weights are computed as follows. We compute the model prediction of the DC term (zero frequency) for a perfectly diffuse material at every surface point. We divide the measured DC frequency photo by the diffuse prediction map and evaluate the spatial derivatives of the resulting image. The reciprocal of the derivative image, multiplied by $1/2$, is used as the diffuse smoothness prior weight map. The specular weights are computed similarly, using a perfectly specular material in the first step. To gradually down-weight the spatial priors, they are further weighted by the factor $32/w$ where w is the image dimension at the current multi-scale optimization stage.

References

BLINN, J. F. 1977. Models of light reflection for computer synthesized pictures. *Computer Graphics (Proc. SIGGRAPH) 11*, 2 (July), 192–198.

BONFORT, T., STURM, P., AND GARGALLO, P. 2006. General specular surface triangulation. In *Proc. Asian Conference on Computer Vision*, vol. II, 872–881.

CHEN, T., GOESELE, M., AND SEIDEL, H.-P. 2006. Mesostructure from specularity. In *Proc. IEEE Conference on Computer Vision and Pattern Recognition*, 1825–1832.

CHUANG, Y.-Y., ZONGKER, D. E., HINDORFF, J., CURLESS, B., SALESIN, D. H., AND SZELISKI, R. 2000. Environment matting extensions: towards higher accuracy and real-time capture. In *Proc. SIGGRAPH*, 121–130.

CLARK, R., 2010. Crazybump, www.crazybump.com.

COLEMAN, E., AND JAIN, R. 1982. Obtaining 3-dimensional shape of textured and specular surfaces using four-source photometry. *Computer Graphics and Image Processing 18*, 309–328.

DANA, K. J., AND WANG, J. 2004. Device for convenient measurement of spatially varying bidirectional reflectance. *Journal of the Optical Society of America A 21*, 1, 1–12.

DONG, Y., WANG, J., TONG, X., SNYDER, J., LAN, Y., BEN-EZRA, M., AND GUO, B. 2010. Manifold bootstrapping for SVBRDF capture. *ACM Transactions on Graphics (Proc. SIGGRAPH) 29*, 4 (July), 98:1–98:10.

DONG, Y., TONG, X., PELLACINI, F., AND GUO, B. 2011. Appgen: interactive material modeling from a single image. *ACM Transactions on Graphics (Proc. SIGGRAPH ASIA) 30*, 6 (Dec.), 146:1–146:10.

FRANCKEN, Y., CUYPERS, T., MERTENS, T., AND GIELIS, J. 2008. High quality mesostructure acquisition using specularities. In *Proc. IEEE Conference on Computer Vision and Pattern Recognition*, 1–7.

GARDNER, A., TCHOU, C., HAWKINS, T., AND DEBEVEC, P. 2003. Linear light source reflectometry. *ACM Transactions on Graphics (Proc. SIGGRAPH) 22*, 3 (July), 749–758.

GHOSH, A., ACHUTHA, S., HEIDRICH, W., AND O'TOOLE, M. 2007. BRDF acquisition with basis illumination. In *Proc. IEEE International Conference on Computer Vision*, 1–8.

GHOSH, A., CHEN, T., PEERS, P., WILSON, C. A., AND DEBEVEC, P. E. 2009. Estimating specular roughness and anisotropy from second order spherical gradient illumination. *Computer Graphics Forum (Proc. Eurographics Symposium on Rendering) 28*, 4, 1161–1170.

GHOSH, A., CHEN, T., PEERS, P., WILSON, C. A., AND DEBEVEC, P. 2010. Circularly polarized spherical illumination reflectometry. *ACM Transactions on Graphics (Proc. SIGGRAPH ASIA) 29*, 6 (Dec.), 162:1–162:12.

GLENCROSS, M., WARD, G., JAY, C., LIU, J., MELENDEZ, F., AND HUBBOLD, R. 2008. A perceptually validated model for surface depth hallucination. *ACM Transactions on Graphics (Proc. SIGGRAPH) 27*, 3 (Aug.), 59:1–59:8.

GOLDMAN, D., CURLESS, B., HERTZMANN, A., AND SEITZ, S. 2005. Shape and spatially-varying BRDFs from photometric stereo. In *Proc. IEEE International Conference on Computer Vision*, vol. 1, 341–348.

HOLROYD, M., LAWRENCE, J., HUMPHREYS, G., AND ZICKLER, T. 2008. A photometric approach for estimating normals and tangents. *ACM Transactions on Graphics (Proc. SIGGRAPH) 27*, 3 (Dec.), 133:1–133:9.

HOLROYD, M., LAWRENCE, J., AND ZICKLER, T. 2010. A coaxial optical scanner for synchronous acquisition of 3D geometry and surface reflectance. *ACM Transactions on Graphics (Proc. SIGGRAPH) 29*, 4 (July), 99:1–99:12.

IKEUCHI, K. 1981. Determining surface orientation of specular surfaces by using the photometric stereo method. *IEEE Transactions on Pattern Analysis and Machine Intelligence 3*, 6, 661–669.

KOPF, J., COHEN, M. F., LISCHINSKI, D., AND UYTTENDAELE, M. 2007. Joint bilateral upsampling. *ACM Transactions on Graphics (Proc. SIGGRAPH) 26*, 3 (July), 96:1–96:6.

LAWRENCE, J., BEN-ARTZI, A., DECORO, C., MATUSIK, W., PFISTER, H., RAMAMOORTHI, R., AND RUSINKIEWICZ, S. 2006. Inverse shade trees for non-parametric material representation and editing. *ACM Transactions on Graphics (Proc. SIGGRAPH) 25*, 3 (July), 735–745.

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 Transactions on Graphics 22*, 2 (Apr.), 234–257.

LÖW, J., KRONANDER, J., YNNERMAN, A., AND UNGER, J. 2012. BRDF models for accurate and efficient rendering of glossy surfaces. *ACM Transactions on Graphics 31*, 1 (Jan.), 9:1–9:14.

NAYAR, S., IKEUCHI, K., AND KANADE, T. 1990. Determining shape and reflectance of hybrid surfaces by photometric sampling. *IEEE Transactions on Robotics and Automation 6*, 4, 418–431.

NEHAB, D., WEYRICH, T., AND RUSINKIEWICZ, S. 2008. Dense 3D reconstruction from specular consistency. In *Proc. IEEE Conference on Computer Vision and Pattern Recognition*, 1–8.

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

RAMAMOORTHI, R., AND HANRAHAN, P. 2001. A signal processing framework for inverse rendering. In *Proc. SIGGRAPH*, 117–128.

REN, P., WANG, J., SNYDER, J., TONG, X., AND GUO, B. 2011. Pocket reflectometry. *ACM Transactions on Graphics (Proc. SIGGRAPH) 30*, 4 (July), 45:1–45:10.

SATO, I., OKABE, T., SATO, Y., AND IKEUCHI, K. 2003. Appearance sampling for obtaining a set of basis images for variable illumination. In *Proc. IEEE International Conference on Computer Vision*, 800–807 Vol. 2.

WANG, C.-P., SNAVELY, N., AND MARSCHNER, S. 2011. Estimating dual-scale properties of glossy surfaces from step-edge lighting. *ACM Transactions on Graphics (Proc. SIGGRAPH ASIA) 30*, 6 (Dec.), 172:1–172:12.

WEYRICH, T., MATUSIK, W., PFISTER, H., BICKEL, B., DONNER, C., TU, C., MCANDLESS, J., LEE, J., NGAN, A., JENSEN, H. W., AND GROSS, M. 2006. Analysis of human faces using a measurement-based skin reflectance model. *ACM Transactions on Graphics (Proc. SIGGRAPH) 25*, 3 (July), 1013–1024.

WEYRICH, T., LAWRENCE, J., LENSCH, H., RUSINKIEWICZ, S., AND ZICKLER, T. 2008. Principles of appearance acquisition and representation. *Foundations and Trends in Computer Graphics and Vision 4*, 2, 75–191.

WOODHAM, R. J. 1978. Photometric stereo: A reflectance map technique for determining surface orientation from image intensity. In *Proc. SPIE*, vol. 155, 136–143.

ZHU, J., AND YANG, Y.-H. 2004. Frequency-based environment matting. In *Computer Graphics and Applications (Proc. Pacific Graphics)*, 402–410.

ZICKLER, T., BELHUMEUR, P. N., AND KRIEGMAN, D. J. 2002. Helmholtz stereopsis: Exploiting reciprocity for surface reconstruction. *International Journal of Computer Vision 49*, 2/3, 215–227.