DOI: 10.1111/cgf.14054
Eurographics Symposium on Rendering 2020
C. Dachsbacher and M. Pharr
(Guest Editors)

Volume 39 (2020), Number 4

An Adaptive BRDF Fitting Metric

J. Bieron P. Peers

College of William & Mary

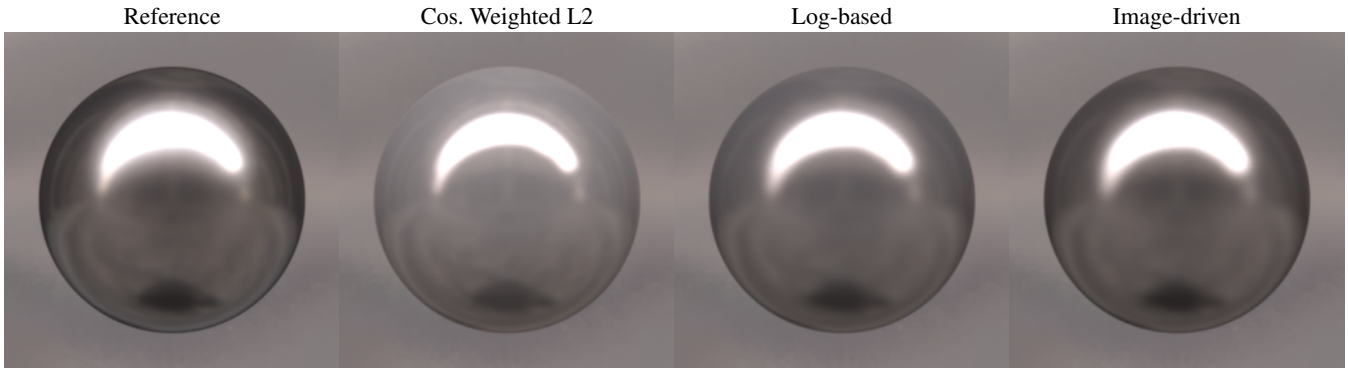

Figure 1: *Comparison of visualization of a Cook-Torrance BRDF model [CT82] fitted to the* Nickel *material from the MERL-MIT BRDF database [MPBM03] using three different BRDF fitting metrics: classic cosine weight error [LFTG97], squared error of the logarithm of the BRDF [LKYU12], and our novel image-driven BRDF fitting strategy.*

Abstract

We propose a novel image-driven fitting strategy for isotropic BRDFs. Whereas existing BRDF fitting methods minimize a cost function directly on the error between the fitted analytical BRDF and the measured isotropic BRDF samples, we also take into account the resulting material appearance in visualizations of the BRDF. This change of fitting paradigm improves the appearance reproduction fidelity, especially for analytical BRDF models that lack the expressiveness to reproduce the measured surface reflectance. We formulate BRDF fitting as a two-stage process that first generates a series of candidate BRDF fits based only on the BRDF error with measured BRDF samples. Next, from these candidates, we select the BRDF fit that minimizes the visual error. We demonstrate qualitatively and quantitatively improved fits for the Cook-Torrance and GGX microfacet BRDF models. Furthermore, we present an analysis of the BRDF fitting results, and show that the image-driven isotropic BRDF fits generalize well to other light conditions, and that depending on the measured material, a different weighting of errors with respect to the measured BRDF is necessary.

Keywords: Isotropic BRDF, fitting, image-driven

1. Introduction

Bidirectional reflectance distribution functions (BRDFs) characterize the scattering of incident irradiance to outgoing radiance at a surface point. Accurately mimicking the reflectance behavior of real-world materials is crucial for photorealistic light transport simulations. While directly using the measured surface reflectance of a material guarantees accurate results, it is not a universally practical solution due to the potential storage requirements, the lack of efficient sampling strategies, and/or the difficulty of altering the material properties. While there exist partial solutions to these shortcomings [DJ18], currently the standard procedure to overcome these limitations is to fit an analytical BRDF model to the measurements such that the resulting BRDF best reflects the reflectance behavior of the measured exemplar material.

BRDF fitting is typically formulated as an optimization for the set of BRDF parameters that minimizes a cost function that numerically quantifies the difference between measured surface reflectance samples and the corresponding samples from the analytical BRDF model. The most commonly used cost function is the cosine weighted square distance [LFTG97, NDM05, WLT04]. How-

ever, when visualized these BRDF fits tend to exhibit a too strong diffuse component because the squared error is dominated by the much larger specular reflectance. To alleviate this issue, a log-based cost function that non-linearly compresses the dynamic range has been proposed [LKYU12, SJR18]. While the visualizations of log-based BRDF fits have improved, it is not universally the case (Figure 1). Based on these observations, we can draw two conclusions: First, the visual appearance of the BRDF is not taken into account in existing BRDF fitting cost functions, yet it is one of the primary ways we evaluate the quality of the BRDF fit. Second, the approximation error is not uniformly distributed, and highly dependent on the analytical BRDF model, BRDF parameters, and measured surface reflectance. Yet, existing cost functions assume a fixed error distribution independent of these factors.

In this paper we propose an easy to implement and novel fitting metric for isotropic BRDFs that adapts to the underlying material as well as the analytical BRDF model, and that takes into account the fidelity of the visual appearance of the material. To achieve this, we reformulate isotropic BRDF fitting as a two stage optimization. In a first stage we generate a set of candidate BRDF fits based on a novel family of cost functions characterized by a single free parameter that (de)emphasizes low versus high reflectance values. In the second stage, we render a sphere under the *Eucalyptus Grove* [Deb98] light probe for each candidate BRDF fit and select the fit that is visually most similar to a rendering of the reference measured isotropic BRDF using a perceptually-based image similarity metric. In contrast to a weighted sum of both costs, a two stage approach avoids the need for delicate balancing of both terms. Our two stage approach can be viewed as optimizing data fidelity in the first stage, and optimizing visual fidelity in the second stage (constrained to the subspace of high data fidelity BRDFs). While our method is easy to implement, and greatly improves visual accuracy as well as robustness of the isotropic BRDF fitting process, it is computationally more expensive compared to classic BRDF fitting methods. Therefore, we additionally introduce a light-weight alternative in which the optimal free-parameter of the BRDF fitting metric is precomputed (i.e., averaged over many materials) for each BRDF model, and which can be implemented with only minimal adjustments to any existing cosine-weighted BRDF fitting framework.

We thoroughly analyze various aspects of our fitting method, and show that it consistently outperforms existing cosine weighted and log-based fitting metrics on the isotropic MERL-MIT BRDF database [MPBM03] in terms of visual fidelity, and validate our numerical conclusions through a user-study. In addition, we investigate two different perceptual image metrics [LPU*13, ZIE*18] for selecting the best BRDF fit. Whereas, both perform similarly on average, the former places greater emphasis on color fidelity and the latter on accuracy of the specular sharpness.

In summary, our contributions are:

1. A novel two stage image-driven isotropic BRDF fitting method that outperforms existing BRDF fitting metrics in terms of visual appearance fidelity;
2. A light-weight BRDF fitting strategy that outperforms existing BRDF fitting metrics at the same computational cost;
3. Fitted BRDF parameters for the Cook-Torrance and GGX microfacet BRDF models for all materials in the MERL-MIT BRDF database [MPBM03] using our novel image-driven BRDF fitting metric.

2. Related Work

The bidirectional reflectance distribution function (BRDF) describes the scattering of incident lighting on an opaque surface [NRH*77]. The BRDF is a 4D function that relates incident irradiance at a surface point to outgoing radiance, and thus it is an essential component in any global illumination rendering system. Over the past decades, numerous BRDF models (e.g., [AS00, CT82, WMLT07, Bli77, HHP*92, HTSG91, LFTG97, War92, HHdD16, HP17]) have been proposed that are increasingly more accurate or more efficient to evaluate.

The availability of high-resolution measured surface reflectance datasets [MPBM03] has enabled experimental validation of existing BRDF models [NDM05], and stimulated the development of new BRDF models that are partially physically based (e.g., based on microfacet theory) augmented with empirical components to obtain better matches to measured data (e.g., [LKYU12, BSH12]) and even automated methods for searching the space of analytical BRDF models [BLPW14]. A key component in such an empirical data-driven BRDF analysis and development process is the ability to fit analytical BRDF models to measured data, i.e., finding the optimal BRDF parameters such that the evaluation of the BRDF model best matches the measured data.

BRDF Fitting Ward [War92] and He et al. [HHP*92] validate their BRDF model by fitting it to a few measured materials. However, they do not specify the exact fitting metric or fitting strategy. Lafortune et al. [LFTG97] employ a squared error of the BRDF times the cosine of both the incident and outgoing direction. Westin et al. [WLT04] compare the accuracy of four analytical BRDF models [Pho75, LFTG97, War92, HTSG91] by fitting them to five measured materials using a similar metric as Lafortune et al., but without the weighting by the cosine of the outgoing direction. Similarly, Ngan et al. [NDM05] compare seven BRDF models [AS00, CT82, War92, Bli77, HTSG91, Dö5, LFTG97] to a large set of measured materials [MPBM03], limiting incident angles to 80° to avoid unreliable measurements at grazing angles, and using a similar fitting metric as Westin et al. : a squared difference of the BRDF weighted by the cosine of the incident direction. Furthermore, they also experimented with a cubic and logarithmic error metric, but conclude that these metrics produce too "blurry" fits for the former, and numerical instabilities around zero with the latter. Recently, Holzschuch and Pacanowski [HP17] proposed a novel highly accurate two-scale microfacet BRDF model. To fit their BRDF model, Holzschuch and Pacanowksi use a variant of the square of the cosine weighted BRDF metric, and include a compressive weighting function [BSN16] and account for errors introduced at grazing angles in the MERL dataset.

Ashikhmin and Premoze [AP07] propose a novel data-driven microfacet BRDF model, and fit the microfacet distribution directly from dense backscatter observations. A similar strategy is employed by Bagher et al. [BSH12] to fit the normal facet distribution from the data slice at $\theta_h = 0$, and estimate Fresnel reflectance

from the data slice at $\theta_d = 0$. Bagher et al. perform: 1) a visual comparison on rendered images similar to Ngan et al. [NDM05], but provide false color difference images, and 2) a quantitative comparison of the squared difference of the BRDF times the cosine of the incident and excitant directions. Both Ashikhmin and Premoze as well as Bagher et al. assume that the BRDF model accurately characterizes the relation between backscatter and non-backscatter reflections. Consequently, when this assumption is not met, the fitted BRDF parameters are suboptimal for non-backscatter directions. In contrast, the proposed image-driven fitting method takes into account all measurements and does not make assumptions on the predictive nature of the underlying BRDF model.

Löw et al. [LKYU12] propose two novel BRDF models and validate their accuracy by fitting to measured data [MPBM03] using two different fitting metrics E1 and E2. The E1 fitting metric is identical to the cosine weighted fitting metric used in prior work, and the E2 metric is the squared error on the logarithm of the cosine weighted BRDF (plus one to avoid numerical issues around zero). Löw et al. report that the logarithmic fitting metric emphasizes wide-angle scattering errors better, and that it produces visually superior results. Clausen et al. [CMF18] extend Löw et al.'s logarithmic metric by raising the foreshortening cosine to a hand-tuned exponent to further reduce the impact of samples at grazing angles. Sun et al. [SJR18] propose a data-driven diffuse-specular separation method (that partially relies on an image-driven metric for computing the diffuse and specular color of the separated measured BRDF), and fit one and two-lobe analytical BRDF models to the specular lobe. For the one-lobe case, the authors note that it is essentially similar to fitting with the logarithmic metric. In this paper we only consider single-lobe models, and based on Sun et al.'s observation will assume that a logarithmic fitting matches the results of Sun et al. in visual quality.

Comparing Visual Material Appearance Ngan et al. [NDM05] argue that the visual quality of the BRDF fits is important, and in addition to comparisons based on numerical error, they also supply visualizations of the different fitted BRDFs on a canonical spherical object under the *Grace Cathedral* light probe [Deb98]. Havran et al. [HFM16] propose a BRDF similarity metric that uses a perceptual image similarity metric (i.e., CSSIM [LPU*13]) to compare visualizations of anisotropic BRDFs applied to a specially crafted geometry and lit by a directional light source. Similar to Havran et al., we will also use CSSIM to judge the visual similarity of BRDFs.

Fores et al. [FFG12] investigate the perceptual qualities of three different BRDF fitting metrics: the regular squared BRDF difference, squared difference of cosine weighted BRDFs, and a cube-root cosine weighted difference of BRDFs. Fores et al. conclude from a perceptual study on renderings of fitted measurements to three BRDF models [War92, CT82, AS00] on a blob-shaped object [VLD07] under the Eucalyptus Grove light probe [Deb98] that the cubic metric produces perceptually better fits for all models. Similarly, Brady et al. [BLPW14] rely on image difference metrics (i.e., a regular squared difference metric and SSIM [WBSS04]) on visualizations of fitted BRDF models on a set of measured materials to identify which BRDF model performs best. Our image-driven method is similar in spirit to Fores et al. and Brady et al. in the sense that we also select the best visual match from a set of candidates, but instead of selecting a BRDF model, we select the best BRDF fit.

In recent concurrent work, Lagunas et al. [LMS*19] introduce a learned material appearance similarity metric that maps an image to a feature space that better correlates with the perception of material appearance. Our BRDF fitting strategy is orthogonal to this work, and the learned metric, as well as any other perceptually based metric, can be easily used in our framework.

Inverse Rendering Our method bears some similarity to inverse rendering [Mar98,RH01,WK15] where the appearance of a scene is estimated such that visualizations of the scene best match a set of reference photographs. A key difference is that inverse rendering typically matches the appearance to a relatively small number of images, and often leverage priors to guide the appearance estimation to a plausible solution. Furthermore, inverse rendering matches the appearance in relation to a whole scene, and thus includes shape variations as well as lighting. In contrast, BRDF fitting starts from an exhaustive set of measurements, and is independent of the shape of the material sample as well as lighting. Our adaptive BRDF fitting metric borrows from both, by taking data accuracy into account as in classic BRDF fitting, as well as visual fidelity as in inverse rendering.

Recently, combinations of model parameter accuracy and visual fidelity has been explored for training deep networks for inferring spatially-vary material properties [LSC18,GLD*19,DAD*19]. However, these methods fix the BRDF model a-priori. Furthermore, these learning techniques assume the ground truth BRDF model parameters are known at training time (and use model parameter accuracy as a loss function), whereas in BRDF fitting we aim to recover the model parameters.

3. Image-driven Isotropic BRDF Fitting

We desire a BRDF fitting metric that fulfills the following two goals: (1) a metric that takes the visual fidelity of the fitted BRDF compared to the measured BRDF into account, and (2) a metric on the surface reflectance values that adapts to the characteristics of the analytical BRDF model and the measured BRDF. Combining two metrics that quantify each goal in a single cost function is non-trivial. Both errors will most likely have an incompatible range, and thus would need appropriate weighting if combined together. Moreover, the scale of both metrics might also be dependent on various other factors, resulting in a potentially different weight for each measured material. Finally, optimizing such a combined metric is likely to be non-trivial as the combined cost function can result in a complex error landscape with many local minima. Instead of jointly optimizing both the visual appearance and surface reflectance metrics, we propose to perform a two stage optimization. In a first step, we generate a number of candidate BRDF fits based on an adaptive BRDF metric (subsection 3.2). Next, from the set of candidate fits, we select the one most optimal according to a visual appearance similarity metric (subsection 3.1).

3.1. Visual Appearance Metric

The visual appearance of a material depends on the shape and lighting under which the material is viewed. We desire an estimate of vi-

sual similarity of material appearance that generalizes well to any shape and natural lighting condition. Hence, we need to establish a reference geometry, reference lighting, and an image similarity metric to compare visualizations of the reference scene with the fitted and measured BRDF.

Image Similarity Our goal is to obtain BRDF fits that produce a similar appearance as the measured material. While simple image metrics such as Mean Square Root Error (MSRE) or Peak Signal to Noise Ratio (PSNR) are easy to implement and quick to evaluate, such metrics fail to characterize our perception of the differences and similarities. Instead, we opt to rely on a perceptually-based image similarity metric. In particular, we will consider two such metrics: CSSIM [LPU*13] and the Learned Perceptual Image Patch Similarity (LPIPS) [ZIE*18]. CSSIM has been successfully used before to characterize appearance similarity of BRDFs [HFM16]. LPIPS has not been used in the context of characterizing appearance similarity. Both CSSIM and LPIPS operate on "low" dynamic range images. We therefore tonemap all rendered images using a gamma 2.2 correction.

Reference Lighting Perceptual studies have shown that humans can best judge material appearance under natural lighting [FDA03, FFG12]. Empirically, we observe that perceptually-based metrics like LPIPS and CSSIM also work more consistently under natural lighting. We therefore follow Fleming et al.'s recommendation of natural lighting for appearance evaluation, and use the *Eucalyptus Grove* light probe [Deb98] to illuminate the reference scene.

Reference Shape Prior research has indicated that a blob aids human viewer in the perception of material reflectance [VLD07]. However, in our case, we do not rely on a human viewer to judge similarity. This allows us more freedom in the choice of shape. This was also recognized by Harvan et al. who optimized a shape for judging reflectance similarity under *directional* lighting. In contrast to Havran et al. we have opted to measure appearance similarity under natural lighting. In such a case, the appearance at each surface point is the integral of the lighting times the BRDF over the visible hemisphere. We therefore desire a shape that offers an unoccluded view of the sphere of incident directions for a wide sampling of surface normal directions. Following Occam's razor, we opt for using a sphere as it (1) meets all the requirements, (2) is easier to render, and (3) is rotationally invariant.

3.2. Adaptive BRDF Metric

Inspired by the dynamic range compression behavior of the log-based metric, we introduce a compression function Λ over the cosine weighted BRDF and that takes an additional parameter γ as input that determines the degree of compression:

$$\varepsilon_\gamma = \sum_{\theta_i,\theta_o,\phi_o} |\Lambda(f_r(\omega_i,\omega_o;p)\cos\theta_i,\gamma) - \Lambda(\hat{f}_r(\omega_i,\omega_o)\cos\theta_i,\gamma)|^2 \sin\theta_i \sin\theta_o \cos\theta_o, \quad (1)$$

where $f_r(\omega_i,\omega_o;p)$ is the analytical BRDF model characterized by the BRDF parameters p and defined over incident and outgoing directions ω_i and ω_o respectively. $\hat{f}_r$ is the measured BRDF. The above metric is similar to the classic cosine-weighted BRDF fitting metrics [LFTG97, NDM05, WLT04], with exception of the inclusion of the compression function Λ.

We follow Löw et al. [LKYU12] and sample ϕ_o and θ_o in 1 degree increments, and θ_i in 10 degree increments. Unlike Löw et al., we sample the full 90° for ϕ_o.

When rendering the reference scene, we also observe that the distribution of evaluated incident and outgoing directions differs. Incident directions are typically integrated over the sphere of directions, and hence its Jacobian for solid angle to spherical coordinates is $\sin\theta_i$. However, the proportion of outgoing directions is closely related to the occurrence of *visible* surface normals. Surface normals at grazing angle occupy relatively less pixels due to view-foreshortening. In case of our spherical geometry, this corresponds to a projection of the visible hemisphere of direction to a disc, and the corresponding Jacobian is $\sin\theta_o\cos\theta_o$.

Finally, we need to define the compression function $\Lambda(\cdot,\gamma)$, where the parameter γ controls the degree of compression. A simple function that fulfills this goal is the power function:

$$\Lambda(r,\gamma) = r^{\frac{1}{\gamma}}. \quad (2)$$

We have opted to make the power inversely proportional to γ as this yields a more uniform change in visual qualities of the fitted BRDFs when sampling the parameter γ uniformly. Note that when $\gamma = 1$, the metric is equivalent to the classic cosine weighted metric.

While our BRDF distance metric shares high-level similarities with the cosine-weighted p-norm used in prior work [PR12], it differs conceptually significantly. Whereas the p-norm compresses the *error*, we compress the range of the *BRDFs* as we desire the metric to adapt to the properties of the BRDF, not to the distribution of the error.

3.3. Implementation

To facilitate ease of implementation, we have designed our fitting strategy to leverage existing components (e.g., numerical optimization frameworks, render systems, etc...) as much as possible. We therefore, separate the optimization in two stages: (1) fitting the BRDF solely based on measured reflectance, and (2) fitting based on visual similarity.

In the first stage of our two stage image-driven BRDF fitting method, we generate 21 candidate BRDF fits, for γ uniformly sampled in $[1,3]$. We minimize the adaptive cost function using a robust direct search method (*patternsearch* in Matlab). However, our method is not married to this particular non-linear optimization method. Since the adaptive metric (Equation 1) is differentiable, faster gradient based method can also be used. To accelerate the fitting process, we observe that a small change in γ only yields modest changes in the fitted BRDF parameters. Hence, we can use the solution of a nearby γ as the starting point for the optimization. In our implementation we first do a full optimization from a neutral starting point for $\gamma = 1$ (i.e., classic least squares BRDF fitting). We then use this solution as the starting point for the next γ value, and iteratively work our way out to $\gamma = 3$. Empirically, we found that a full optimization requires on average $1,000$ iterations, whereas using a neighboring solution as a start point requires on average

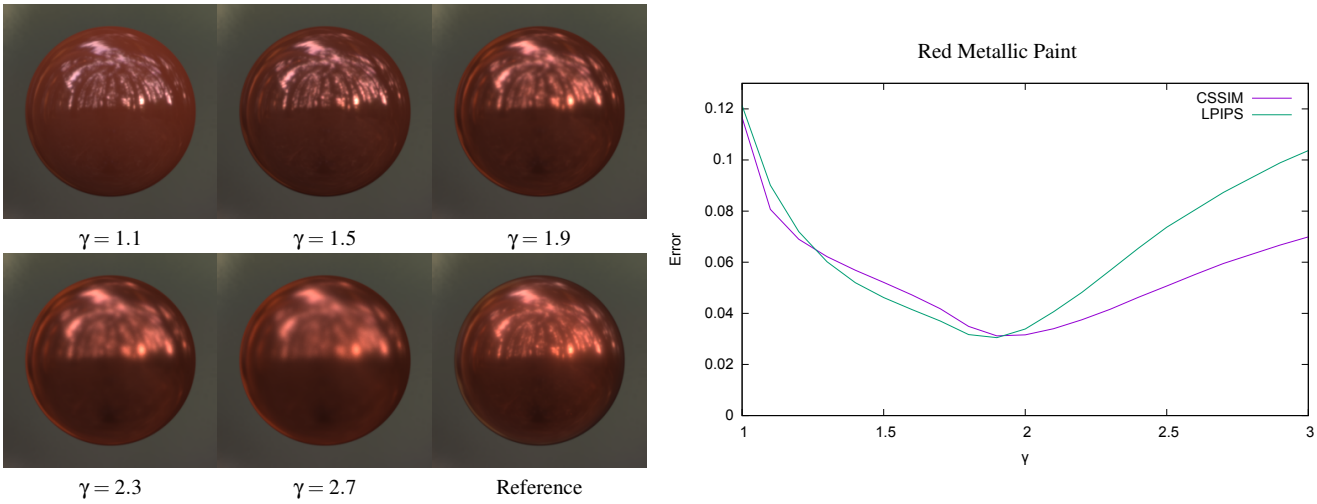

Figure 2: *The effect of different values of* γ *on the BRDF fits illustrated on* Red Metallic Paint. ***Left:*** *Low* γ *values produce sharper BRDF fits, and conversely, high* γ *values produce more blurry BRDF fits.* ***Right:*** *The CSSIM and LPIPS errors plotted (with respect to a visualization of the reference measured material) for each* γ *value; a minimum is reached at* $\gamma = 1.9$ *for both CSSIM and LPIPS.*

250 iterations. Hence, compared to a classic BRDF fitting metric, the computational overhead is approximately a factor 6 (1000 iterations versus $1000 + 20 \times 250 = 6000$ iterations). Figure 2 shows BRDF fits for a selection of different γ values for the *Red Metallic Paint* material. High γ values tend to produce more diffuse fits, and conversely low γ values place more emphasis on the specular component. We do not explore γ values below 1, as the BRDF fits obtained with a regular cosine weighted L^2 are already overly sharp. We also do not explore γ values over 3, as the resulting BRDF fits are very blurred and the minimal visual improvement does not justify the additional computational expense. Note however that using a larger range for γ can only improve the quality of the results, albeit at significant additional computational costs.

In the second stage, we render the BRDF fits obtained for the different γ values for the reference scene, and compute and find the rendering, and thus corresponding BRDF fit, with the lowest CSSIM or LPIPS error. In our implementation we use Mitsuba [Jak10] to render the BRDF fits. However, any other rendering system capable of correctly integrating light probe lighting over the BRDF can be used instead.

4. Results

Figure 3 and Figure 4 show selections of 4 materials from the MERL-MIT BRDF database [MPBM03], fit to the Cook-Torrance [CT82] and GGX [WMLT07] BRDF models respectively. The 1st column shows a reference rendering of the material under the *Uffizi Gallery* (odd rows) and the *St. Peter's Basilica* (even rows) light probes [Deb98]. Note that these lighting conditions are different from the lighting condition used for selection in the second stage of our image-driven fitting method (i.e., *Eucalyptus Grove*). For each analytical BRDF model we show a cosine-weighted L^2 fit (2nd column), Löw et al.'s log-based fit (3rd column), and our image-driven BRDF fit with the CSSIM and LPIPS metrics (4th and 5th column respectively), and a direct CSSIM fit (6th column); we used the same Jacobian and sampling pattern for all BRDF fitting metrics. We have empirically verified that our L^2 and log-based fits for the Cook-Torrance BRDF model are visually similar to or better than the BRDF fits in prior work [NDM05, LKYU12]. The direct CSSIM fit is obtained by optimizing the BRDF parameters such that the CSSIM error on the rendering of the BRDF model under the target illumination (i.e., *Uffizi Gallery*, or *St. Peter's Basilica*) is minimized; the cost function does not rely on surface reflectance measurements, and directly optimizes the image error. SSIM (and thus CSSIM) has been shown to be ill-suited for optimization [BVW12], requiring us to use a robust, but time-consuming, direct-search non-linear optimization with a well-chosen starting point; we use the L^2 BRDF fit as the starting point. This direct CSSIM BRDF fit is the best possible BRDF fit that approximates the visual appearance under the *target* lighting. Hence it presents a lower bound on the capabilities of how well the BRDF model can mimic the appearance of the measured material according to CSSIM similarity. Due to the many local minima, we were not able to obtain a robust direct LPIPS fit. We list the CSSIM and LPIPS errors between the rendering and the reference image for each BRDF fit.

From these results (over all MERL-MIT materials) we can draw the following conclusions:

1. The cosine weighted BRDF fits are visually the most dissimilar. This is confirmed by the CSSIM and LPIPS errors which are significantly larger than for the other BRDF fits.
2. The log-based BRDF fits are visually a better match than the L^2 BRDF fits for these materials. Nevertheless, in general we observe a mismatch in brightness for both the specular (i.e., too dark) and diffuse (i.e., too bright) components.
3. In 50% (according to CSSIM) or 47% (according to LPIPS) of the materials in the MERL-MIT BRDF database evaluated under four light probes (*Grace Cathedral*, *Uffizi Gallery*, *St. Pe-*

Figure 3: *Image-driven BRDF fitting on three selected materials with the Cook-Torrance microfacet BRDF model compared to the classic cosine weighted L^2 BRDF fitting metric and Löw et al.'s log-based fitting metric. The CSSIM and LPIPS error is shown below each BRDF fit visualization. Note that the "Direct CSSIM" results are optimized to produce an as low as possible CSSIM error under the lighting shown. Hence is represents a lower bound on the CSSIM error for each material under the shown lighting.*

Figure 4: *Image-driven BRDF fitting on three selected materials with the GGX microfacet BRDF model compared to the classic cosine weighted L^2 BRDF fitting metric and Löw et al.'s log-based fitting metric. The CSSIM and LPIPS error is shown below each BRDF fit visualization. Note that the "Direct CSSIM" results are optimized to produce an as low as possible CSSIM error under the lighting shown. Hence is represents a lower bound on the CSSIM error for each material under the shown lighting.*

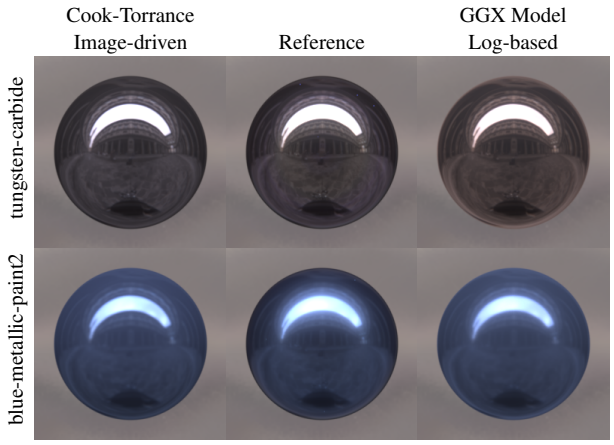

Figure 5: *A comparison of our image-driven Cook-Torrance BRDF fits to log-based GGX BRDF fits for the* Tungsten Carbide *and* Blue Metallic Paint 2.

ter's Basilica, *Beach*), the visual fidelity of the image-driven Cook-Torrance BRDF fits is equal to or better than the log-based GGX microfacet BRDF fits. This illustrates that a good BRDF fitting metric on a suboptimal BRDF can outperform a suboptimal BRDF fitting metric on a superior BRDF model. Examples of this are shown in Figure 5.

4. The results for our image-driven BRDF fitting metric exhibit a CSSIM error close to the optimal direct CSSIM fits. Note that our image-driven BRDF fitting metric is disadvantaged compared to the direct CSSIM fit as it performs the second stage (i.e., selection of the best candidate) under a different lighting, while the direct CSSIM fit is optimized against the target lighting.

Our image-driven BRDF fitting method is particularly effective when the underlying analytical BRDF is unable to characterize the measured surface reflectance exactly. In such as case, measured reflectance reproduction is mostly irrelevant and the visually best match is selected.

Finally, the supplemental materials include a complete listing of the fitted BRDF parameters for the Cook-Torrance and GGX microfacet BRDF models for all MERL-MIT materials fitted with the different metrics, as well as visualizations and CSSIM/LPIPS errors of the fitted BRDFs under four different light probes. From these supplemental results, one can observe that our image-driven metric does not necessarily produce the most visually accurate results under all lighting conditions for a single material. This suggest that there might not exist a single unique BRDF fit that produces the visually most accurate material appearance uniformly. However, as we will show in section 5, our image-driven fitting metric performs best on average over all tested materials and lighting conditions.

5. Analysis & Discussion

5.1. Analysis

The results in section 4 qualitatively indicate that our image-driven metric is able to produce visually more accurate BRDF fits. This is further qualitatively backed up by Table 1 that lists the percentage of the MERL-MIT materials for which each BRDF fitting metric produces the lowest CSSIM/LPIPS error on a sphere rendered under four different lighting conditions (i.e., two indoor light probes (*Grace Cathedral* and *St. Peter's Basilica*) and two outdoor light probes (*Uffizi Gallery* and *Beach*)); we use the same image similarity metric for error comparison as for selection in the second stage of the image-based fitting process. If the optimal γ value equals 1, then the cosine weighted L^2 metric and our image-driven metric are identical; we assign half a win to each metric in case that such a BRDF fit produces the lowest error. As can be seen, our image-driven BRDF fitting metric produces visually more accurate fits than the competing BRDF fitting metrics, and that both CSSIM and LPIPS perform comparable. Furthermore, as illustrated in Figure 6 in cases where other BRDF metrics provide a visually better fit, the differences in CSSIM/LPIPS errors are small. In contrast, when our image-driven metric provides the best result, the differences can be significant. This is further corroborated by looking at the total CSSIM and LPIPS error under different light probes over the MERL-MIT BRDF database (Table 2; columns 2 & 3 vs. 5 for the Cook-Torrance BRDF model, and columns 7 & 8 vs. 10 for the GGX microfacet BRDF model) where our image-driven BRDF fitting metric achieves the overall lowest cumulative visual error. We also observe that as the analytical BRDF model becomes more expressive (i.e., GGX vs. Cook-Torrance), that classic BRDF fitting metrics produce visually more accurate results; when the BRDF model can better explain the measurements, visual accuracy follows from data fidelity. However, even in the case of GGX, our image-driven fitting method still offers a significant advantage.

5.2. Ablation Study

Image-driven vs. Direct CSSIM Table 2 (columns 6 and 11) also lists the cumulative error for the direct CSSIM BRDF fit over different light probes. Unlike previous experiments, these are the BRDF fits obtained under a single fixed lighting conditions (i.e., *Eucalyptus Grove*) and then visualized under the other light probes. Interestingly, the cumulative errors for the direct CSSIM fits are larger than the cumulative CSSIM and LPIPS errors for our image-driven metric. This indicates that the direct CSSIM BRDF fits are potentially overfitted to the lighting condition. Our image-driven BRDF fitting method short-circuits this issue by creating a set of candidate fits solely based on their sampled reflectance values independently of their visual accuracy, reducing the likelihood of overfitting. Empirically, we found that overfitting to the lighting occurs more often for advanced BRDF models such as GGX. Potentially, lighting overfitting could be combatted by directly optimizing over an ensemble of light probes. However, this would pose a number of significant problems. First, to avoid biasing, the ensemble of light probes would need to be representative of the distribution of real world light probes. Second, the computational costs would be orders of magnitude higher without additional optimizations. Recent advances in differential rendering systems could reduce the computational burden. Finally, optimizing in the image domain is more complex and would likely result in a less robust optimization. For example, multiple restarts where needed for stable direct CSSIM BRDF fitting (with a single light probe) whereas our image-driven BRDF fitting did not require a restart.

Table 1: *Percentage of MERL-MIT materials for which the Cook-Torrance/GGX BRDF fit computed with the respective fitting metric has a lower CSSIM/LPIPS error than all other metrics (for the same BRDF model) for four different lighting conditions.*

CSSIM

Model:	Cook-Torrance			GGX		
Metric:	Cos. L^2	Log-based	Image-driven	Cos. L^2	Log-based	Image-driven
Grace	6.50%	35.00%	58.50%	21.50%	29.00%	49.50%
Uffizi	5.00%	11.00%	84.00%	8.00%	12.00%	80.00%
St. Peters	9.00%	24.00%	67.00%	21.50%	15.00%	63.50%
Beach	4.50%	16.00%	79.50%	22.50%	14.00%	63.50%
Total	6.25%	21.50%	72.25%	18.38%	17.50%	64.12%

LPIPS

Model:	Cook-Torrance			GGX		
Metric:	Cos. L^2	Log-based	Image-driven	Cos. L^2	Log-based	Image-driven
Grace	4.00%	29.00%	67.00%	19.50%	23.00%	57.50%
Uffizi	8.00%	22.00%	70.00%	13.00%	16.00%	71.00%
St. Peters	4.50%	32.00%	63.50%	16.50%	24.00%	59.50%
Beach	5.00%	23.00%	72.00%	17.00%	20.00%	63.00%
Total	5.38%	26.50%	68.12%	16.50%	20.75%	62.75%

Table 2: *Comparison of the cumulative CSSIM and LPIPS errors over all materials in the MERL-MIT BRDF database for four different light probes and for both the Cook-Torrance and GGX microfacet BRDF models. The optimal γ parameter selection, the image-driven BRDF fitting, as well as the Direct CSSIM fit are performed under the* Eucalyptus Grove *light probe.*

CSSIM Average Errors over All Materials

Model:	Cook-Torrance					GGX				
Metric:	Cos. L^2	Log-based	Optimal $\gamma = 2.1$	Image-driven	Direct CSSIM	Cos. L^2	Log-based	Optimal $\gamma = 2.5$	Image-driven	Direct CSSIM
Grace Cathedral	0.04215	0.02355	0.02271	0.02216	0.02671	0.03593	0.01866	0.01596	0.01747	0.02594
Uffizi Gallery	0.05779	0.03160	0.02098	0.02081	0.01989	0.05201	0.02532	0.01722	0.01688	0.02073
St. Peters	0.03105	0.02221	0.01938	0.01775	0.01954	0.02744	0.01994	0.01563	0.01507	0.01690
Beach	0.02310	0.01679	0.01326	0.01266	0.01233	0.02060	0.01515	0.01114	0.01118	0.01294
All Lightings	0.03852	0.02354	0.01908	0.01834	0.01962	0.03399	0.01977	0.01499	0.01515	0.01913

LPIPS Average Errors over All Materials

Model:	Cook-Torrance					GGX				
Metric:	Cos. L^2	Log-based	Optimal $\gamma = 1.9$	Image-driven	Direct CSSIM	Cos. L^2	Log-based	Optimal $\gamma = 2.0$	Image-driven	Direct CSSIM
Grace Cathedral	0.05086	0.03094	0.03325	0.02845	0.03518	0.04224	0.02412	0.02442	0.02167	0.04158
Uffizi Gallery	0.05440	0.03413	0.02733	0.02664	0.02483	0.04812	0.02923	0.02381	0.02333	0.02596
St. Peters	0.03474	0.02627	0.02661	0.02184	0.02712	0.03021	0.02212	0.02072	0.01923	0.02669
Beach	0.03772	0.02461	0.02081	0.01947	0.01777	0.03226	0.02330	0.01864	0.01805	0.02117
All Lightings	0.04443	0.02899	0.02687	0.02410	0.02622	0.03821	0.02469	0.02190	0.02057	0.02885

Table 3: *Cumulative error over 7 rotations of the blob lit by the* Eucalyptus Grove *light probe for image-driven BRDF fits selected on a sphere, a different rotation of the blob, Havran et al. [HFM16]'s shape lit by the aforementioned light probe, plus Havran et al. 's scene as designed (point lighting).*

	CSSIM Metric				LPIPS Metric			
Model	Sphere	Blob	Havran shape	Havran scene	Sphere	Blob	Havran shape	Havran scene
Cook-Torrance	14.95872	15.04680	15.54202	24.87558	15.56582	15.42327	15.88830	30.99339
GGX Microfacet	11.84850	11.90496	12.44426	14.19923	14.17241	14.06424	14.25998	17.37047

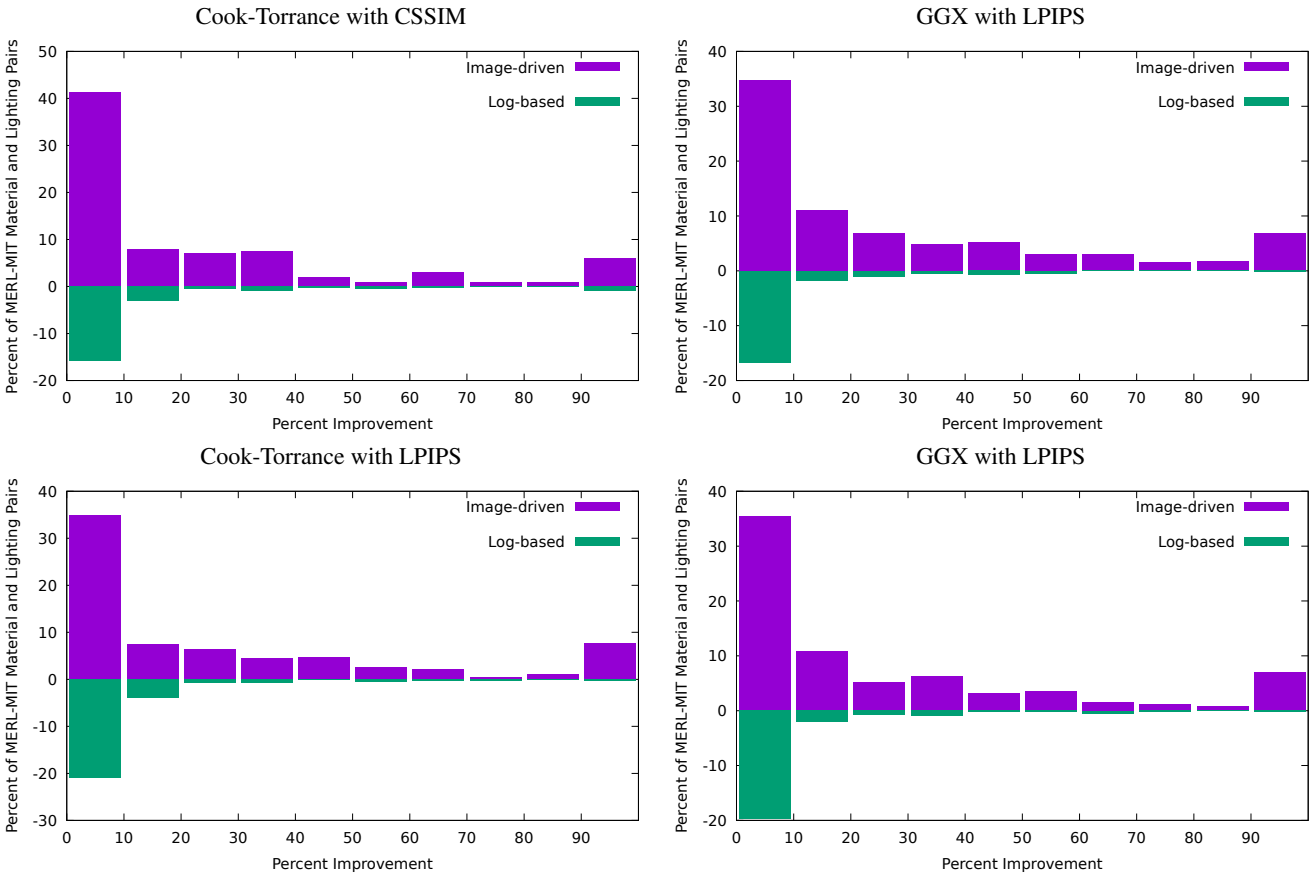

Figure 6: *Histogram of relative error for the Cook-Torrance BRDF model and the GGX microfacet BRDF model. The relative errors are accumulated under four different lighting conditions (*Grace Cathedral*,* Uffizi Gallery*,* St. Peter's Basilica*, and* Beach *light probes), and does not include the light probe under which the BRDF fit was selected. The* purple *histogram includes the materials where our image-driven metric outperforms the log-based metric. The* green *histogram includes the materials for which the log-based metric outperforms our image-driven metric.*

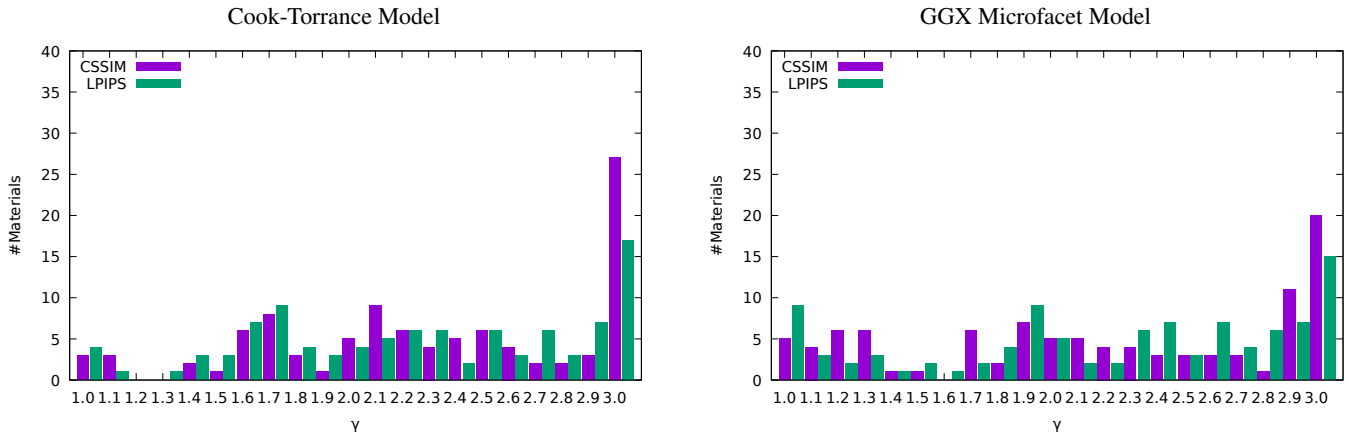

Figure 7: *Histogram of the number of MERL-MIT materials per selected γ value by CSSIM and LPIPS, for both the Cook-Torrance BRDF model and the GGX Microfacet BRDF model.*

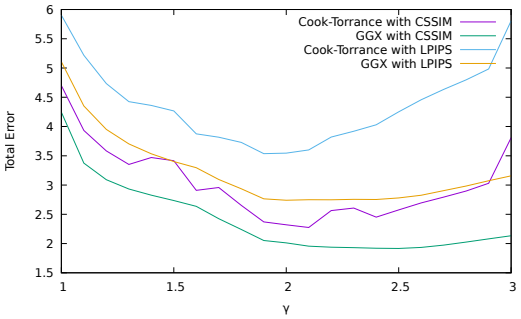

Figure 8: *Cumulative CSSIM and LPIPS errors over all MERL-MIT materials visualized under the* Eucalyptus Grove *lighting for varying γ values. The cumulative CSSIM errors reach a minimum for $\gamma = 2.1$ and $\gamma = 2.5$ for the Cook-Torrance and GGX microfacet BRDF model respectively. For the LPIPS metric, the minimum is found at $\gamma = 1.9$ for Cook-Torrance and $\gamma = 2.0$ for the GGX microfacet model. Note that the CSSIM and LPIPS error values are not compatible and their relative order does not indicate that one is better than the other.*

Adaptive γ vs. Fixed γ The previous experiment indicates BRDF fitting based solely on visual similarity is not ideal. However, it does not show the necessity of an adaptive metric. Indeed, it is a valid question whether there exists a single optimal γ value. Figure 7 plots the histogram of number of MERl-MIT materials per selected γ value. From this we can draw two conclusions:

1. There does not exist a single "optimal" γ value, and the distribution is fairly uniform.
2. The last histogram bin contains a significantly larger number of materials. This is an artifact of the fact that we do not consider γ values larger than 3 because of the minimal improvement in visual BRDF appearance outside this range.

While modest, the adaptive two stage optimization incurs a computational overhead ($6\times$) compared to a classic cosine-weighted or log-based BRDF fitting metrics. BRDF fitting is typically performed once in a precomputation step, and thus computational cost is not a significant concern for most applications. However, to accommodate applications for which fitting cost is of importance, we propose a *light weight* alternative, by precomputing the optimal *fixed* γ value that produces the overall best BRDF fits. Once this γ value is precomputed, it can then be used, at the same computational cost as classic fitting metrics, to fit isotropic BRDF models. Figure 8 plots the total CSSIM and LPIPS error over the MERL-MIT BRDF database for the Cook-Torrance and GGX microfacet BRDF models. Each of the plots follow a slightly different trajectory, and reach a minimum at a different γ value: when using CSSIM as the selection metric, the optimal fixed γ values are 2.1 for the Cook-Torrance BRDF model and 2.5 for the GGX BRDF model, and when using LPIPS as selection metric, the optimal γ value is 1.9 for the Cook-Torrance model and 2.0 for the GGX microfacet model. Table 2 (4th and 9th column) shows the cumulative CSSIM/LPIPS errors for the fixed γ BRDF fitting (with $\gamma = 2.1/1.9$ and $\gamma = 2.5/2.0$ for the Cook-Torrance and GGX BRDF models respectively). For both BRDF models, the optimal fixed γ solution outperforms prior fitting metrics in terms of CSSIM and LPIPS error. Hence, it offers an attractive alternative when reduced computational cost is essential. Interestingly, the fixed $\gamma = 2.5$ metric marginally outperforms, in terms of total CSSIM error, the image-driven metric for the GGX model. This is mainly due to the significantly better performance under the *Grace Cathedral* light probe. When the target lighting (i.e., *Grace Cathedral*) deviates significantly from the selection lighting (i.e., *Eucalyptus Grove*) then it is possible, as is the case here, that a fixed γ can outperform, on average over the MERL-MIT BRDF database, the adaptive image-driven metric.

Table 4: *The mean standard deviation of each model's parameters over the MERL-MIT BRDF database for 5 random starting points for the first stage of our BRDF fitting process with $\gamma = 1$. All BRDF parameters are in the 0 to 1 range. For the Cook-Torrance BRDF model, Fresnel reflectance is determined by the reflectance at normal incidence. For the GGX microfacet BRDF model, Fresnel is parameterized by the index of refraction (η) which was rescaled to $[0..1]$ via $\frac{\eta-1}{4}$.*

Parameter	Cook-Torrance	GGX
Diffuse	0.000428	0.003035
Specular	0.003230	0.013166
Roughness	0.002168	0.010775
Fresnel	0.005846	0.077113

Impact of Reference Shape In subsection 3.1 we opted for using sphere as the reference shape by virtue of its simplicity. To better understand the implications of this choice, we compute (Table 3) the cumulative CSSIM and LPIPS errors for the sphere, blob, and Havran et al. [HFM16]'s shape as the reference shape under the *Eucalyptus Grove* light probe, computed over visualization of the blob under 7 different rotations (excluding the selection orientation) for different lighting conditions over all MERL-MIT materials. For completeness, we also include Havran et al. [HFM16]'s metric with directional lighting as the selection metric (instead of the reference shape under natural lighting).

From this can can draw the following conclusions:

1. The sphere and the blob reference shape perform very similar under both image similarity metrics; the blob slightly outperforming the sphere for LPIPS, and the sphere outperforming the blob for CSSIM.
2. Havran et al. 's shape performs well, albeit slightly less than the sphere and the blob. We believe this is mainly due to the self-occlusion of the ground plane cutting of some of the important grazing angle cues. In general, most reasonable convex (i.e., sphere-like) shapes are acceptable as the reference shape.
3. Havran et al.'s shape under directional lighting does not perform well under either perceptual image similarity metric for isotropic BRDFs. However, it should be noted that Havran et al. 's BRDF similarity metric is designed for *anisotropic* BRDFs instead of isotropic BRDF as in our case.

Repeatability of Fitting The repeatability of our fitting strategy is in most part determined by the stability of the non-linear parameter optimization for $\gamma = 1$ (i.e., classic least squares BRDF fitting),

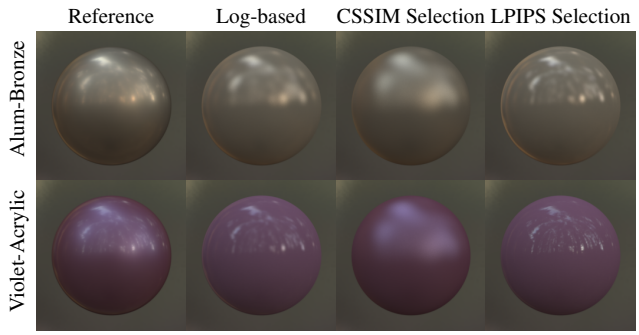

Figure 9: *Two materials that cannot be well expressed by the Cook-Torrance BRDF model. Without a more expressive model, some aspect of the analytical fit will be visibly wrong (i.e. specular sharpness or diffuse color/intensity).*

since subsequent optimizations start from the solution of the previous fits and the second stage is deterministic given the outcome of the first stage. The stability of this $\gamma = 1$ optimization depends on two factors: the choice of optimization algorithm, and the complexity of the BRDF model. We opted for the *patternsearch* non-linear optimization algorithm which is robust to the choice of the starting point. To validate, we compute the standard deviation over 5 BRDF fits with random starting point for the Cook-Torrance and GGX microfacet BRDF models over the MERL-MIT BRDF database. The results (Table 4) show, indeed, that for these BRDF models and using *patternsearch*, the fitting is stable and repeatable. It is advisable to validate the stability of the BRDF fitting when changing the BRDF model or non-linear optimization algorithm, and if needed improve the stability by taking the best BRDF fit from multiple attempts with random starting points for $\gamma = 1$.

CSSIM vs. LPIPS In this paper, we have used two different image similarity metrics: CSSIM [LPU*13] and LPIPS [ZIE*18]. Both image similarity metrics capture perceptual appearance similarity well. However, both have slightly different 'preferences'. CSSIM tends to prioritize color fidelity over sharpness, and hence tends to selected slightly more blurred BRDFs. LPIPS gives more weight to sharpness, and tends to produce images with better matching highlight shapes. While often subtle, the difference in preference is most obvious when the analytical BRDF cannot fit the measured reflectance well (e.g., as exemplified in Figure 9 for the *Violet-Acrylic* and *Alum-Bronze* material).

We note that our image-driven BRDF fitting metric is not married to CSSIM or LPIPS, and any other image and/or appearance similarity metrics such as the learned metric of Lagunas et al. [LMS*19] can be used. Conversely, regular image difference metrics can also be used. For example, we found that a squared image difference on tone-mapped images performs reasonably well in many cases, albeit not as robustly (with respect to perceptual similarity) as CSSIM or LPIPS.

Complex Shapes and Advanced BRDFs In the above in-depth analyses, we have focused on a limited number of BRDF models (i.e., the Cook-Torrance and GGX microfacet BRDF models) and relatively simple shapes. For completeness, we empirically investigate whether our BRDF fits perform well on more complex shapes with indirect light transport and advanced BRDF models.

In Figure 10 we show visualizations of the *tungsten-carbide* material fitted to the GGX microfacet BRDF model using both the log-based and our image-driven metric on the *Buddha* geometry model which exhibits some modest amounts of indirect lighting and high frequency surface normal variations under two different light probes (i.e., *Uffizi Gallery* and *St. Peter's Cathedral*). While differences in highlights are less visible due to the high-frequency normal variations, the overall appearance of the image-driven BRDF fit is visually closer to the reference visualization.

More advanced BRDF models are capable of more accurately modeling surface reflectance, and thus better fit reflectance measurements, thereby implicitly guaranteeing good visual fidelity. Yet, such advanced BRDF models can also benefit from our adaptive BRDF fitting metric. Figure 11 shows BRDF fits of four selected materials from the MERL-MIT BRDF dataset using the advanced two-scale microfacet BRDF model of Holzschuch and Pacanowski [HP17]. We compare, using the author's BRDF implementation, the fitted BRDF parameters reported by Holzschuch and Pacanowski with image-driven BRDF fits under two different light probes: the light probe used by Holzschuch and Pacanowski (at a slightly lower exposure to reduce oversaturation of specular highlights) and the *Uffizi Gallery* light probe. These results show that even on such a complex BRDF model, our adaptive image-driven BRDF fitting metric can produce visually more accurate BRDF fits.

5.3. User-study

In the above analyses we have taken the error values of the perceptual metrics (i.e., CSSIM and LPIPS) at face value. However, we do notice that for some materials (e.g., such as *Violet-Acrylic* and *Alum-Bronze* shown in Figure 9), each metric picks a suboptimal γ value. Hence, neither metric fully captures the human perception of material appearance. To validate the quality of our BRDF fits we perform an additional large-scale user study on Amazon's Mechanical Turk.

Experiment Description & Setup Our user-study takes the form of a 2 alternative forced choice (2AFC) experiment where the user is presented with a reference image and two different BRDF fits. The user then has to select the visualization that "best matches the material appearance of the reference visualization". In our user-study we compare our image-driven BRDF fitting result with a log-based BRDF fit; we randomly change the order of which fit is shown on which side of the reference rendering. We ran our user study for both BRDF models (Cook-Torrance and GGX), and for both selection metrics (CSSIM and LPIPS), as well as the light-weight fixed γ version computed with CSSIM and LPIPS. Hence, we ran 8 different user-studies that share all experiment parameters; we did not mix the different image-driven fits in the same study.

For a number of materials in the MERL-MIT database, the BRDF fits of both image-driven and log-based BRDF fits are very similar. Showing such scenes to the user will be counterproductive as the user might get frustrated and loose concentration. We therefore select a subset of 50 materials that are more likely to show differences. To select this subset, we sort the materials according to

Figure 10: *Comparison of log-based and image-driven BRDF fits for* tungsten-carbide *with the GGX microfacet BRDF model visualized on the* Buddha *model under the* Uffizi Gallery *and* St. Peter's Cathedral *light probes.*

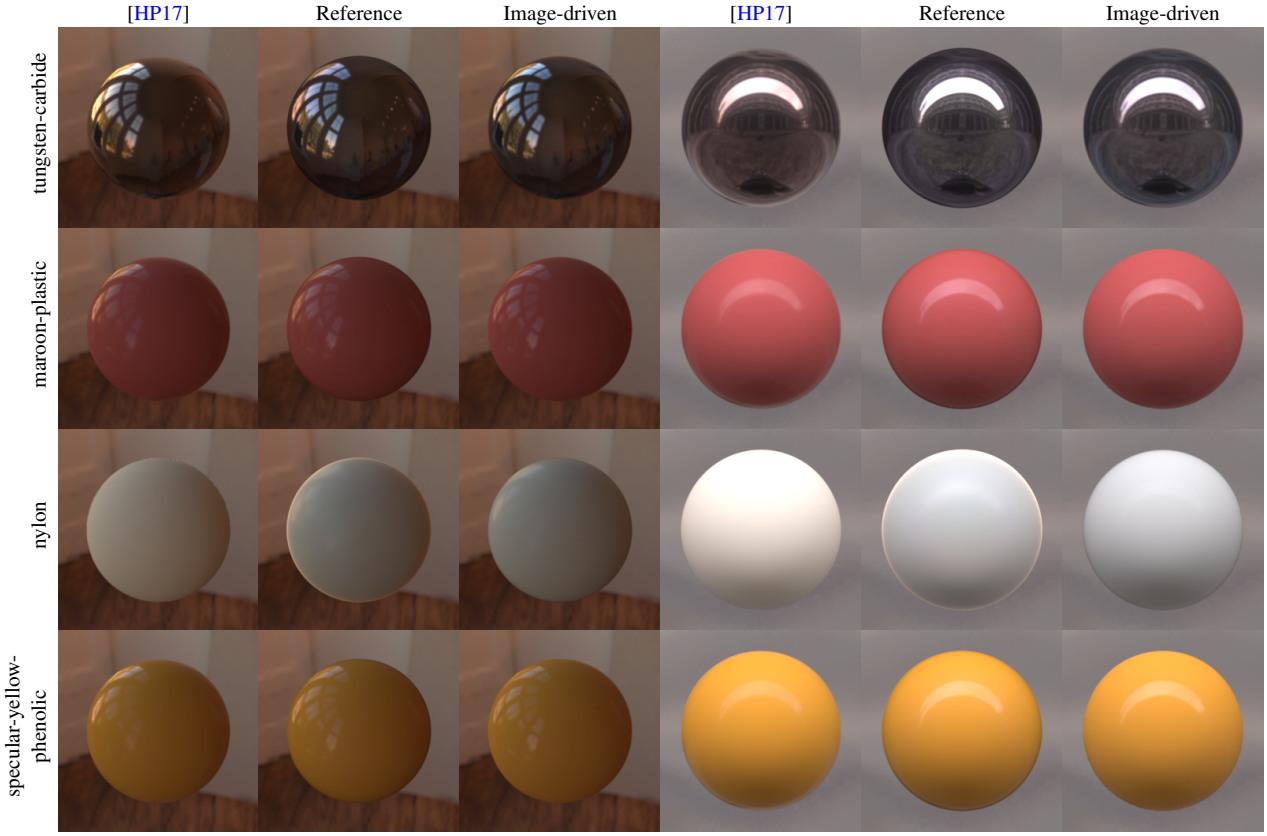

Figure 11: *Comparison of previously published fits to image-driven fits with an advanced BRDF Model [HP17] for four materials under two different lighting conditions.*

the LPIPS 'error' (image difference) between the log-based Cook-Torrance BRDF fit and our image-driven BRDF fit, and retain the 50 with the highest error. To compare the results from the different user-studies, we use the same subset for each study, including the studies that feature the GGX BRDF model. Note that this is a very conservative threshold. Figure 12 shows the last included BRDF from this list; both BRDF fits are visually very similar. Using CSSIM instead of LPIPS to sort the materials yielded a similar subset of 50 materials despite slight differences in the sort.

To avoid bias in the reference shape or lighting in the visualizations presented to the user, we use a different set of conditions than we use in the image-driven selection step. Following Fleming et al. [FDA03] we present a scene under natural lighting. We opt for the *Uffizi Gallery* light probe because it is color neutral which helps

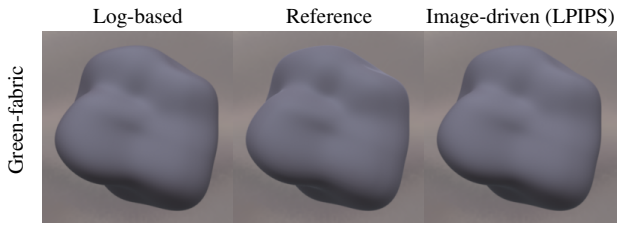

Figure 12: *Cook-Torrance BRDF fits for the last material included in our user study. This was the fiftieth material as sorted by LPIPS difference between log-based and the corresponding (LPIPS) image-driven BRDF fits on our reference scene.*

Figure 13: *Reference images for* Aluminium *shown for the five orientations of the blob used in the user study.*

the users in detecting material color differences. Furthermore, we follow Vangorp et al. [VLD07] and use a blob shape. To avoid conditioning of the observer to the scene, we randomly show 5 different orientations of the blob (Figure 13). These 5 orientations are a subset from the 8 uniformly sampled rotations along the up axis of the object, and which exhibit sufficient specular cues for the user to judge.

For each of the 5 shapes and 50 materials, we obtained 5 judgments, yielding 25 judgments per material over 8 experiments. In total we collected $10,000$ user judgments.

Results Table 5 summarizes the result from the user study. We report the number of 'wins' for each tested MERL-MIT material, where a 'win' is counted as obtaining more than a certain percentage of the votes (i.e., 50%, 66%, and 75%, corresponding to receiving at least 13, 17, and 19 votes out of 25 respectively). For all models, thresholds, and similarity metrics our image-driven BRDF fit scores more wins than a log-based BRDF fit. On average, image-driven BRDF fitting results are preferred in 74% of the tested materials for a threshold of 50% (a simply majority of user judgments). Increasing the threshold to 66%, image-driven BRDF fitting results are preferred in 36% of the tested materials ($\sim$ 18% of the MERL-MIT database). Note that this does not imply that the users preferred the log-based fit for 82% of the MERL-MIT BRDFs; log-based BRDF fits were preferred only in $\sim$ 2% of the MERL-MIT materials, and they were unable to decide in 80% of the materials (including the majority of very diffuse materials) at this threshold. For the threshold of 75%, users preferred image-driven BRDF fits were preferred for 21% of tested materials, corresponding to 10% of the MIT-MERL database, while no log-based fits were preferred (all other materials were undecided).

We observe there are slightly more 'undecided' materials for the GGX BRDF fits than for the Cook-Torrance BRDF fits. This is not an unexpected results as the GGX model can better fit the measured materials. We expect that the better the analytical BRDF model

Table 5: *Summary of the user-study results on image-driven BRDF fitting (using CSSIM and LPIPS as selection metric) compared to log-based fitting for the Cook-Torrance BRDF model and the GGX microfacet BRDF model. Each column lists the number of tested material (out of* 50*) preferred by a certain minimum percentage (i.e., threshold) of users.*

Threshold:	50%		66%		75%	
Winner:	Us	Log	Us	Log	Us	Log
CT with CSSIM	39	11	18	0	11	0
CT with LPIPS	39	11	22	0	13	0
GGX with CSSIM	36	14	18	4	9	0
GGX with LPIPS	34	15	17	6	8	0
Total with CSSIM	75	25	36	4	20	0
Total with LPIPS	73	26	39	6	21	0

can replicate the measured reflectance, the more similar the BRDF fits for the different γ values will be. In the limit, if the analytical BRDF model can perfectly fit the measurements, then any BRDF fitting metric will produce the same result. However, as demonstrated in Figure 11, fitting current state-of-the-art models still benefits from our adaptive BRDF metric.

We also ran the user-study for the lightweight fixed γ value BRDF fits. For a 66% threshold, we found that for the Cook-Torrance BRDF model with $\gamma = 2.1$ (CSSIM computed) the number of "wins" are $16-0$ (image-driven fit versus log-based fit, respectively), and for $\gamma = 1.9$ (LPIPS computed) $18-1$. For the GGX BRDF model we counted $11-1$ for $\gamma = 2.5$ (CSSIM computed), and $17-6$ for $\gamma = 2.0$ (LPIPS computed). For the lightweight fixed γ value BRDF fits we observe that the CSSIM computed γ values perform better in general. We posit that this is related to the fact that CSSIM tends to prefer BRDF fits which, while more blurry (as evidenced by the higher average γ value), exhibit higher color fidelity. An interesting avenue for future research would be to investigate the relative weight users give to color fidelity versus specular highlight appearance.

Material Class Analysis To better understand the results from the user-study, we manually categorize the 50 materials in different material classes: *metals* (13 materials), *phenolics* (9 materials), *paints* (11 materials), *plastics* (8 material), and *others* (9 materials). Table 6 summarizes the results. Image-driven BRDF fitting significantly outperforms log-based fitting for *metals*, and performs well for *paints* and *plastics*. For the other materials classes its performance is on par with log-based fitting. We argue that, in particular for *phenolics*, that both BRDF models can accurately model such types of materials, and thus the fitting metric matters less.

6. Conclusions

We presented a novel BRDF fitting method that takes both the accuracy of the surface reflectance as well as the fidelity of the visual appearance into account. A key advantage of our method is that it is straightforward to implement, thereby allowing for easy adaption in existing frameworks. We demonstrated that our image-driven method produces BRDF fits of better visual quality than existing cosine-weighted or log-based fitting metrics. For roughly half

Table 6: *Summary of user-study results per material class for a win-threshold of* 50% *and* 66%.

Threshold of 50%

Class:	Metals		Phenolics		Paints		Plastics		Other	
Winner:	Us	Log	Us	Log	Us	Log	Us	Log	Us	Log
CT with CSSIM	11	2	4	5	9	2	7	1	8	1
CT with LPIPS	11	2	6	2	9	2	6	2	7	2
GGX with CSSIM	10	3	3	6	10	1	6	2	7	2
GGX with LPIPS	12	1	3	6	8	2	5	3	6	3
Total with CSSIM	21	5	7	11	19	3	13	3	15	3
Total with LPIPS	23	3	9	8	17	4	11	5	13	5

Threshold of 66%

Class:	Metals		Phenolics		Paints		Plastics		Other	
Winner:	Us	Log	Us	Log	Us	Log	Us	Log	Us	Log
CT with CSSIM	7	0	0	0	5	0	2	0	4	0
CT with LPIPS	10	0	1	0	4	0	2	0	5	0
GGX with CSSIM	8	0	0	3	5	0	3	0	2	1
GGX with LPIPS	8	0	0	2	2	0	1	0	2	0
Total with CSSIM	15	0	0	3	10	0	5	0	6	1
Total with LPIPS	18	0	1	2	6	0	3	0	7	0

of the MERL-MIT materials, image-driven Cook-Torrance BRDF fits outperform classic cosine-weighted and log-based BRDF fits on the superior GGX microfacet BRDF model. In addition, we also proposed a light-weight alternative with fixed γ parameter that outperforms existing metrics with the same computational cost.

Currently, our method relies on the CSSIM or LPIPS image similarity metrics to determine visual appearance similarity. However, neither CSSIM or LPIPS were designed to assess appearance similarity, and a dedicated appearance similarity metric could further improve our results. Finally, we only consider isotropic BRDFs; extensions to anisotropic materials would be an interesting avenue for future research.

Acknowledgments This work was partially supported by NSF grants IIS-1350323 and IIS-1909028.

References

[AP07] ASHIKHMIN M., PREMOZE S.: Distribution-based BRDFs, 2007. http://www.cs.utah.edu/~premoze/dbrdf/. 2

[AS00] ASHIKHMIN M., SHIRLEY P. S.: An anisotropic phong BRDF model. *Journal of Graphics Tools 5*, 2 (2000), 25–32. 2, 3

[Bli77] BLINN J.: Models of light reflection for computer synthesized pictures. *Computer Graphics 11*, Annual Conference Series (1977), 192–198. 2

[BLPW14] BRADY A., LAWRENCE J., PEERS P., WEIMER W.: genbrdf: Discovering new analytic brdfs with genetic programming. *ACM Transactions on Graphics 33*, 4 (August 2014). 2, 3

[BSH12] BAGHER M. M., SOLER C., HOLZSCHUCH N.: Accurate fitting of measured reflectances using a Shifted Gamma micro-facet distribution. *Computer Graphics Forum 31*, 4 (June 2012), 1509–1518. 2

[BSN16] BAGHER M. M., SNYDER J., NOWROUZEZAHRAI D.: A non-parametric factor microfacet model for isotropic brdfs. *ACM Trans. Graph. 35*, 5 (July 2016). 2

[BVW12] BRUNET D., VRSCAY E. R., WANG Z.: On the mathematical properties of the structural similarity index. *IEEE Transactions on Image Processing 21*, 4 (April 2012), 1488–1499. 5

[CMF18] CLAUSEN O., MARROQUIM R., FUHRMANN A.: Acquisition and validation of spectral ground truth data for predictive rendering of rough surfaces. *Computer Graphics Forum 37*, 4 (2018), 1–12. 3

[CT82] COOK R. L., TORRANCE K. E.: A reflectance model for computer graphics. *ACM Trans. Graph. 1*, 1 (1982), 7–24. 1, 2, 3, 5

[DÖ5] DÜR A.: On the ward model for global illumination. Unpublished manuscript: http://homepage.uibk.ac.at/~c70240/publications.html, 2005. 2

[DAD*19] DESCHAINTRE V., AITTALA M., DURAND F., DRETTAKIS G., BOUSSEAU A.: Flexible svbrdf capture with a multi-image deep network. *Comp. Graph. Forum 38*, 4 (July 2019). 3

[Deb98] DEBEVEC P.: Light probe gallery. http://www.pauldebevec.com/Probes/, 1998. 2, 3, 4, 5

[DJ18] DUPUY J., JAKOB W.: An adaptive parameterization for efficient material acquisition and rendering. *ACM Trans. Graph. 37*, 6 (2018), 274:1–274:14. 1

[FDA03] FLEMING R. W., DROR R. O., ADELSON E. H.: Real-world illumination and the perception of surface reflectance properties. *Journal of Vision 3*, 5 (2003). 4, 13

[FFG12] FORES A., FERWERDA J., GU J.: Towards a perceptually based metric for brdf modeling. In *Twentieth Color and Imaging Conference* (Nov. 2012), pp. 142–148. 3, 4

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

[HFM16] HAVRAN V., FILIP J., MYSZKOWSKI K.: Perceptually motivated brdf comparison using single image. *Computer Graphics Forum 35*, 4 (2016). 3, 4, 9, 11

[HHdD16] HEITZ E., HANIKA J., D'EON E., DACHSBACHER C.: Multiple-scattering microfacet bsdfs with the smith model. *ACM Trans. Graph. 35*, 4 (July 2016). 2

[HHP*92] HE X., HEYNEN P., PHILLIPS R., TORRANCE K., SALESIN D., GREENBERG D.: A fast and accurate light reflection model. *Computer Graphics 26*, Annual Conference Series (1992), 253–254. 2

[HP17] HOLZSCHUCH N., PACANOWSKI R.: A two-scale microfacet reflectance model combining reflection and diffraction. *ACM Trans. Graph. 36*, 4 (2017), 66:1–66:12. 2, 12, 13

[HTSG91] He X., Torrance K., Sillon F., Greenberg D.: A comprehensive physical model for light reflection. *Computer Graphics 25*, Annual Conference Series (1991), 175–186. 2

[Jak10] Jakob W.: Mitsuba: Physically based renderer. https://www.mitsuba-renderer.org, 2010. 5

[LFTG97] Lafortune E. P. F., Foo S.-C., Torrance K. E., Greenberg D. P.: Non-linear approximation of reflectance functions. In *SIGGRAPH '97: Proceedings of the 24th annual conference on Computer graphics and interactive techniques* (1997), pp. 117–126. 1, 2, 4

[LKYU12] Löw J., Kronander J., Ynnerman A., Unger J.: Brdf models for accurate and efficient rendering of glossy surfaces. *ACM Transactions on Graphics (TOG) 31*, 1 (January 2012). 1, 2, 3, 4, 5

[LMS*19] Lagunas M., Malpica S., Serrano A., Garces E., Gutierrez D., Masia B.: A similarity measure for material appearance. *ACM Trans. Graph. 38*, 4 (July 2019). 3, 12

[LPU*13] Lissner I., Preiss J., Urban P., Lichtenauer M. S., Zolliker P.: Image-difference prediction: From grayscale to color. *IEEE TIP 22*, 2 (2013), 435–446. 2, 3, 4, 12

[LSC18] Li Z., Sunkavalli K., Chandraker M. K.: Materials for masses: Svbrdf acquisition with a single mobile phone image. In *ECCV* (2018). 3

[Mar98] Marschner S. R.: *Inverse rendering for computer graphics*. PhD thesis, Cornell University, 1998. 3

[MPBM03] Matusik W., Pfister H., Brand M., McMillan L.: A data-driven reflectance model. *ACM Trans. Graph. 22*, 3 (2003), 759–769. 1, 2, 3, 5

[NDM05] Ngan A., Durand F., Matusik W.: Experimental analysis of BRDF models. *Eurographics Symposium on Rendering 2005* (2005), 117–226. 1, 2, 3, 4, 5

[NRH*77] Nicodemus F. E., Richmond J. C., Hsia J. J., Ginsberg I. W., Limperis T.: Geometric considerations and nomenclature for reflectance. *National Bureau of Standards Monograph 160* (1977). 2

[Pho75] Phong B. T.: Illumination for computer generated pictures. *Commun. ACM 18*, 6 (June 1975), 311–317. 2

[PR12] Pereira T., Rusinkiewicz S.: Gamut mapping spatially varying reflectance with an improved brdf similarity metric. *Comput. Graph. Forum 31*, 4 (2012), 1557–1566. 4

[RH01] Ramamoorthi R., Hanrahan P.: A signal-processing framework for inverse rendering. In *Proceedings of the 28th Annual Conference on Computer Graphics and Interactive Techniques* (2001), SIGGRAPH âĂŹ01, p. 117âĂŞ128. 3

[SJR18] Sun T., Jensen H. W., Ramamoorthi R.: Connecting measured brdfs to analytical brdfs by data-driven diffuse-specular separation. *ACM Trans. Graph. 37*, 6 (2018), In press. 2, 3

[VLD07] Vangorp P., Laurijssen J., Dutré P.: The influence of shape on the perception of material reflectance. *ACM Trans. Graph. 26*, 3 (July 2007). 3, 4, 14

[War92] Ward G. J.: Measuring and modeling anisotropic reflection. *SIGGRAPH Comput. Graph. 26*, 2 (1992), 265–272. 2, 3

[WBSS04] Wang Z., Bovik A. C., Sheikh H. R., Simoncelli E. P.: Image quality assessment: From error visibility to structural similarity. *IEEE Trans. Image Proc. 13*, 4 (2004), 600–612. 3

[WK15] Weinmann M., Klein R.: Advances in geometry and reflectance acquisition (course notes). In *SIGGRAPH Asia 2015 Courses* (2015), pp. 1:1–1:71. 3

[WLT04] Westin S. H., Li H., Torrance K. E.: *A comparison of four brdf models*. Tech. Rep. PCG-04-02, Cornell University, 2004. 1, 2, 4

[WMLT07] Walter B., Marschner S., Li H., Torrance K. E.: Microfacet models for refraction through rough surfaces. In *Rendering Techniques* (2007), pp. 195–206. 2, 5

[ZIE*18] Zhang R., Isola P., Efros A. A., Shechtman E., Wang O.: The unreasonable effectiveness of deep features as a perceptual metric. In *CVPR* (2018). 2, 4, 12