DOI: 10.1111/cgf.70162
Eurographics Symposium on Rendering 2025
B. Wang and A. Wilkie
(Guest Editors)

COMPUTER GRAPHICS *forum*
Volume 44 (2025), Number 4

Perceived quality of BRDF models

Behnaz Kavoosighafi[1], Rafał K. Mantiuk[2], Saghi Hajisharif[1], Ehsan Miandji[1], and Jonas Unger[1]

[1]Linköping University, Sweden
[2]University of Cambridge, UK

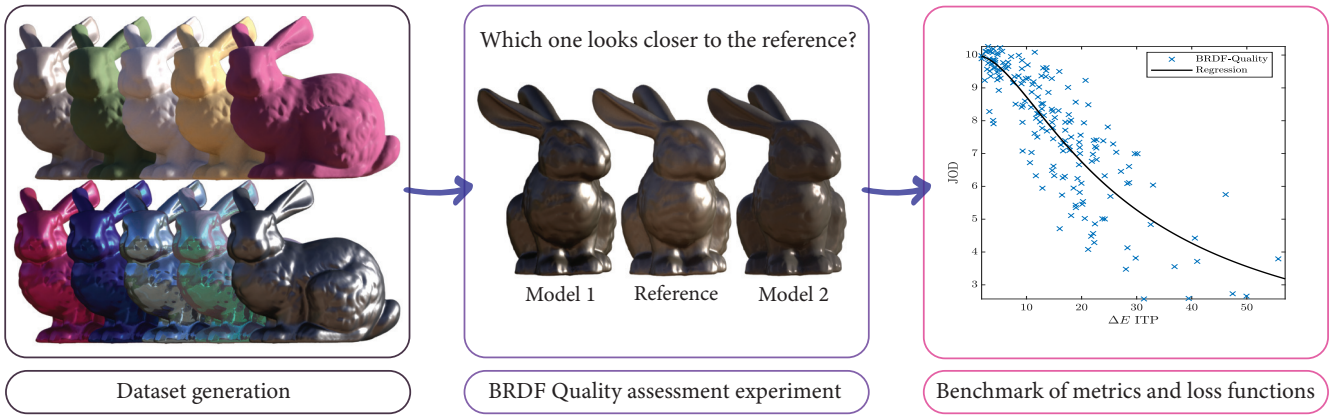


Figure 1: *Overview of our pipeline: (1) We generated a dataset of video sequences containing 1 431 BRDF models with complex geometry and environment map; (2) We then conducted a perceptual quality assessment using our multifocal HDR display; (3) Finally, we analyzed existing BRDF-space and image-space metrics as both quality metrics and loss functions.*

Abstract

Material appearance is commonly modeled with the Bidirectional Reflectance Distribution Functions (BRDFs), which need to trade accuracy for complexity and storage cost. To investigate the current practices of BRDF modeling, we collect the first high dynamic range stereoscopic video dataset that captures the perceived quality degradation with respect to a number of parametric and non-parametric BRDF models. Our dataset shows that the current loss functions used to fit BRDF models, such as mean-squared error of logarithmic reflectance values, correlate poorly with the perceived quality of materials in rendered videos. We further show that quality metrics that compare rendered material samples give a significantly higher correlation with subjective quality judgments, and a simple Euclidean distance in the ITP color space (ΔE_{ITP}) shows the highest correlation. Additionally, we investigate the use of different BRDF-space metrics as loss functions for fitting BRDF models and find that logarithmic mapping is the most effective approach for BRDF-space loss functions.

CCS Concepts

• ***Computing methodologies*** → ***Perception; Reflectance modeling; Rendering;***

1. Introduction

Accurate capture, representation, and reproduction of the appearance of real-world materials is fundamental to photorealistic rendering in computer graphics. The bidirectional reflectance distribution function (BRDF) [Nic65] describes how light interacts with and is scattered at surfaces in a scene, determining its appearance under varying illumination and viewing conditions. Although techniques for measuring BRDFs, spatially varying BRDFs and bidirectional texture functions (BTFs) from real materials have advanced significantly [DJ18; KHM*24; MTH*24; HHN*17], fitting these measurements to concise and efficient models still remains a crucial and challenging step.

Traditional BRDF fitting methods, [NDM05; KSK10; LKYU12], primarily focus on minimizing cost functions based on L_2 errors in linear or logarithmic space between the measured and modeled reflectance values. However, these metrics often fail to

correlate well with both numerical errors, e.g. PSNR and MSE, as well as the perceptual quality of rendered images. In other words, there is a clear difference between the BRDF model fitting errors and the corresponding errors/quality obtained when the fitted models are used in renderings. This disconnect stems from several factors, including the non-linear nature of human visual perception, approximations inherent in the rendering process, and the influence of the characteristics of the scene such as illumination, shape, normal directions, and view direction to name a few.

To address this limitation, we conduct a perceptual BRDF quality assessment experiment, collecting subjective data for a diverse set of complex BRDFs, modeled using nine different BRDF models with varying levels of approximation error and sourced from multiple BRDF datasets. We investigate the effectiveness of existing image-space and BRDF-space metrics in evaluating the quality of BRDF fits, using the collected subjective data as a benchmark. Through evaluation, we identify the best-performing image-space and BRDF-space metrics based on their correlation with subjective quality judgments. Additionally, we benchmark existing BRDF fitting loss functions to find solutions that yield more visually accurate and compelling renderings of real-world materials. Our study focuses on isotropic BRDFs, due to the limited availability of anisotropic datasets and the limitations of current BRDF models—particularly learning-based ones—which are not well-suited for modeling anisotropy.

Our key contributions are:

1. New BRDF model quality dataset measured on an HDR stereoscopic display for a range of classical parametric and new neural BRDF models with a complex animated object. The dataset can be accessed at `https://doi.org/10.17863/CAM.118756` and also at the project web page.
2. A benchmark of both image- and BRDF-space quality metrics on our new dataset.
3. A benchmark of loss functions for fitting BRDF models.

2. Background and related work

Modeling and representation of scattering properties, as described by the BRDF, is fundamental to computer graphics and vision. A large body of work has been directed towards the development of models that can describe BRDFs and SVBRDFs efficiently. These models can be divided into two main categories, *parametric models* and *non-parametric* data-driven models.

Parametric models are typically derived either from empirical observations of scattering characteristics [Pho75; War92; LKYU12; AS00; Bli77] (phenomenological), or through theoretical analysis based on the physics of light-surface interactions, such as microfacet theory [CT81; WMLT07; Bur12; HP17; TS67]. However, for accurate modeling and simulation of real-world materials, it is often necessary to measure the scattering properties from physical samples. This is typically done using gonioreflectometers [Foo97; WSB*98; ELU11; DJ18], light stage setups [DHT*00; MGW01; MMS*05; RMS*08; BWW*08], or catadioptric systems [MDL*98; Dan01; KN06; NZV*11; GAHO07; YSX*24]. Such measurements can be used to fit parametric models and to build data-driven models.

A challenge inherent to high-quality BRDF acquisition is the combination of the high sampling density required to obtain accurate model fits and the dimensionality (4D) of the BRDF space. To address this problem, it is common to employ nonparametric techniques that project the data onto a low-dimensional space to enable more efficient rendering and reduce storage requirements. Approaches in this domain include matrix factorization [LRR04; SSN18; KM99], tensor decomposition [SZC*07; BÖK11; TUKK20], analytical methods [SJR18; BSN16; CBP22; PSS*12], and machine learning-based techniques [FR22; LSZ*24; TUGM22; GSZ*25; ZSR*24; FWH*22; CNN20; HGC*20; SRRW21]. While these non-parametric techniques yield more precise BRDF representations due to their data-driven modeling, they generally come with higher storage and computation costs compared to parametric models.

Central to all previous and current efforts towards the development of new BRDF models intended for modeling real-world materials and techniques for BRDF acquisition is the need to accurately measure and characterize the error introduced using some metric. BRDF metrics are typically computed in one of two domains: in *image-space*, where the BRDF accuracy is evaluated as rendering error, and in *BRDF-space*, where the focus is on measurement and representation error. The rendering error is typically evaluated using the mean squared error (MSE), the peak signal-to-noise ratio (PSNR), or using some of the commonly used perceptually based metrics, including the structural similarity index measure (SSIM) [WBSS04a] or the visual difference predictor for HDR images (HDR-VDP) [MDMS05; MKRH11].

Image-based metrics are by nature not suitable for applications like BRDF fitting and measurements as they require the image to be rendered in each iteration. Instead, the cost functions used for the BRDF fitting and optimization tasks are typically based on the ℓ_1 or ℓ_2 norms, which measure the error distance between the original and approximated reflectance in the linear space. When applied to BRDF modeling, these functions are often weighted by the cosine of the incoming and outgoing directions to balance the contributions of reflectance values [LFTG97; NDM05; LKYU12; PL07; WLT04; HP17]. To better align the ℓ_2 distance with the human perceptual system, Forés et al. [FFG12] explored various functions based on the root mean square (RMS) and concluded that a cube-root cosine-weighted metric provides a better perceptual approximation. This was further confirmed by Lavoué et al. [LBFS21], who demonstrated that a cubic root transformation prior to computing RMS-based cost functions improves the quality of the approximations when used for BRDF fitting purposes. Based on the established relationship between the logarithm of intensity and the human perceptual system [VS13], Löw et al. [LKYU12] proposed a logarithmic cosine-weighted ℓ_2 metric, which was later refined through the incorporation of an exponent factor to further mitigate the impact of grazing angles [CMF18]. In related studies, log-relative mapping was introduced [SJR18; NJR15], demonstrating that logarithmic mapping performs similarly to cubic root mapping, while ℓ_2 log mapping produces a shorter tail in the specular component. Although these representation-based techniques offer a good approximation of the tabulated data without dependence on specific scene configurations (e.g., illumination type, or the positions of light sources and cameras), their focus on minimizing numerical

errors often overlooks the complex details of reflectance behavior, which can result in renderings that lack perceptual accuracy.

A key problem is the weak correlation between the error measured in BRDF space and the rendering error, i.e. a low fitting error may still yield a high rendering error. This problem can be observed both in our experiments and in previous work [NDM05; BP20; HFM16], and points to the need for accurate metrics to bridge this gap.

To consider visual appearance in the evaluation, Brady et al. [BLPW14] and Havran et al. [HFM16] analyzed the impact of different perceptual metrics, including ΔE_{ITP}, C-SSIM [LPU*13], and HDR-VDP2 [MKRH11], to compare anisotropic BRDFs with optimized shapes. Building on the strengths of image similarity metrics, Bieron et al. [BP20] proposed the use of LPIPS [ZIE*18a] and C-SSIM on rendered images to select the best BRDF fit from candidates obtained through numerical error minimization. To develop this further, user studies have been conducted to derive appearance metrics, exploring factors such as variations in lighting and geometry [LMS*19], and to analyze the correlation between analytical BRDF models and perceptual scores [LBFS21]. More recently, Filip et al. [FDS*24] conducted a user study on videos of flat surfaces to gather ratings on 16 perceptual attributes. Their findings demonstrate that a weighted combination of Pearson's correlation and the ℓ_1 norm captures perceived material similarity. Although significant progress has been made, this is still an open research problem.

While most prior work has focused on analytic BRDF models, we present the first comprehensive evaluation of the emerging class of machine learning–based BRDF models, which introduce a broader range of rendering artifacts. Our study also considers a significantly larger and more diverse set of evaluation metrics, including several that have not previously been applied in BRDF research. To ensure a wide variety of material appearances, our dataset incorporates complex materials from the DTU and RGL-EPFL datasets, expanding beyond the simpler examples found in the MERL database. In contrast to previous work, we conduct a subjective perceptual study using a stereoscopic HDR display to gather high-quality labels, enabling an in-depth analysis of both BRDF-space and image-space metrics. Our primary goal is to identify and benchmark the metrics that best correlate with human perceptual judgments, with the aim of informing quality assessment and improving loss functions for BRDF fitting.

3. Dataset

One of our key contributions is a new dataset of rendered BRDF samples designed for quality assessment experiments. We employed isotropic measured BRDFs samples from the MERL [MPBM03], DTU [NJR15], and RGL-EPFL [DJ18] datasets. The samples were used to fit one of the popular BRDF models and then render an animation of a rotating object. In total, 159 material samples were used. To ensure a consistent treatment of all materials, we transformed the RGL-EPFL dataset to align with the MERL coordinate system, following the Rusinkiewicz coordinate framework [Rus98].

Table 1: *BRDF models used in our dataset. Three values listed for* SPARSEBRDF *and* HYPERBRDF *correspond to three different compression/encoding levels. The parameters listed for* HYPERBRDF *correspond to the number of query samples used.*

BRDF model	#parameters
WARD [War92]	7
GGX [WMLT07]	8
NEURALBRDF [SRRW21]	675
SPARSEBRDF [TUGM22]	256, 512, 2 048
HYPERBRDF [GSZ*25]	40, 400, 4 000

3.1. BRDF models

To study the relationship between the choice of metric used in BRDF modeling/fitting and the visual quality obtained in the resulting BRDF approximations, we utilized five different BRDF models, as listed in Table 1. The WARD [War92] and GGX [WMLT07] models are selected as representatives of parametric approaches, with WARD embodying an empirical model and GGX a microfacet-based model. NEURALBRDF [SRRW21] offers a simple MLP architecture that is overfitted to each material. In contrast, SPARSEBRDF [TUGM22], and HYPERBRDF [GSZ*25] provide representations that can learn multiple materials, either through sparse encoding or a hypernetwork.

We used the logarithmic cosine-weighted mapping error metric introduced by Löw et al. [LKYU12] to fit the tabulated data to parametric models:

$$\epsilon = \| f(\rho_{\text{model}}(\omega_i, \omega_o; \mathrm{p})) - f(\rho_{\text{ref}}(\omega_i, \omega_o) \|^2, \text{ where} \quad (1)$$

$$f(\rho) = \log\left(1 + \rho \cdot \cos(\theta_i) \cdot \cos(\theta_o)\right). \quad (2)$$

Here, ρ_{ref} represents the measured BRDF and ρ_{model} is the prediction of the model. The terms $\omega_i = (\theta_i, \phi_i)$, and $\omega_o = (\theta_o, \phi_o)$ denote the incident and outgoing directions, with θ_i and θ_o as the incident and outgoing elevation angles, and ϕ_i and ϕ_o as the corresponding azimuth angles. The vector of parameters is denoted as p, depending on the model. The fitting process was carried out using the `lsqcurvefit` function in MATLAB, which solves the non-linear least squares problem from Eq. (1).

For the data-driven models, we followed the original methodologies. Specifically, we adopted the same architecture proposed in the original NEURALBRDF paper, which employs an auto-encoder with two hidden layers and 675 parameters, training a unique latent space for each material in the dataset. In contrast, for SPARSEBRDF and HYPERBRDF, the dataset was randomly split into training and testing sets, with 75% of the data allocated for training and 25% for testing set. When training and testing SPARSEBRDF, we applied the *log-plus* transformation used in their work, which is a simplified version of Eq. (2) that lacks cosine weight mapping. The number of BRDF parameters required for the SPARSEBRDF model is defined by sparsity. Therefore, to cover a diverse range of BRDF approximations, we produced three models from SPARSEBRDF with training sparsity values of 256, 512, and 2 048. The number of dictionaries trained for each sparsity was set to 4. Dur-

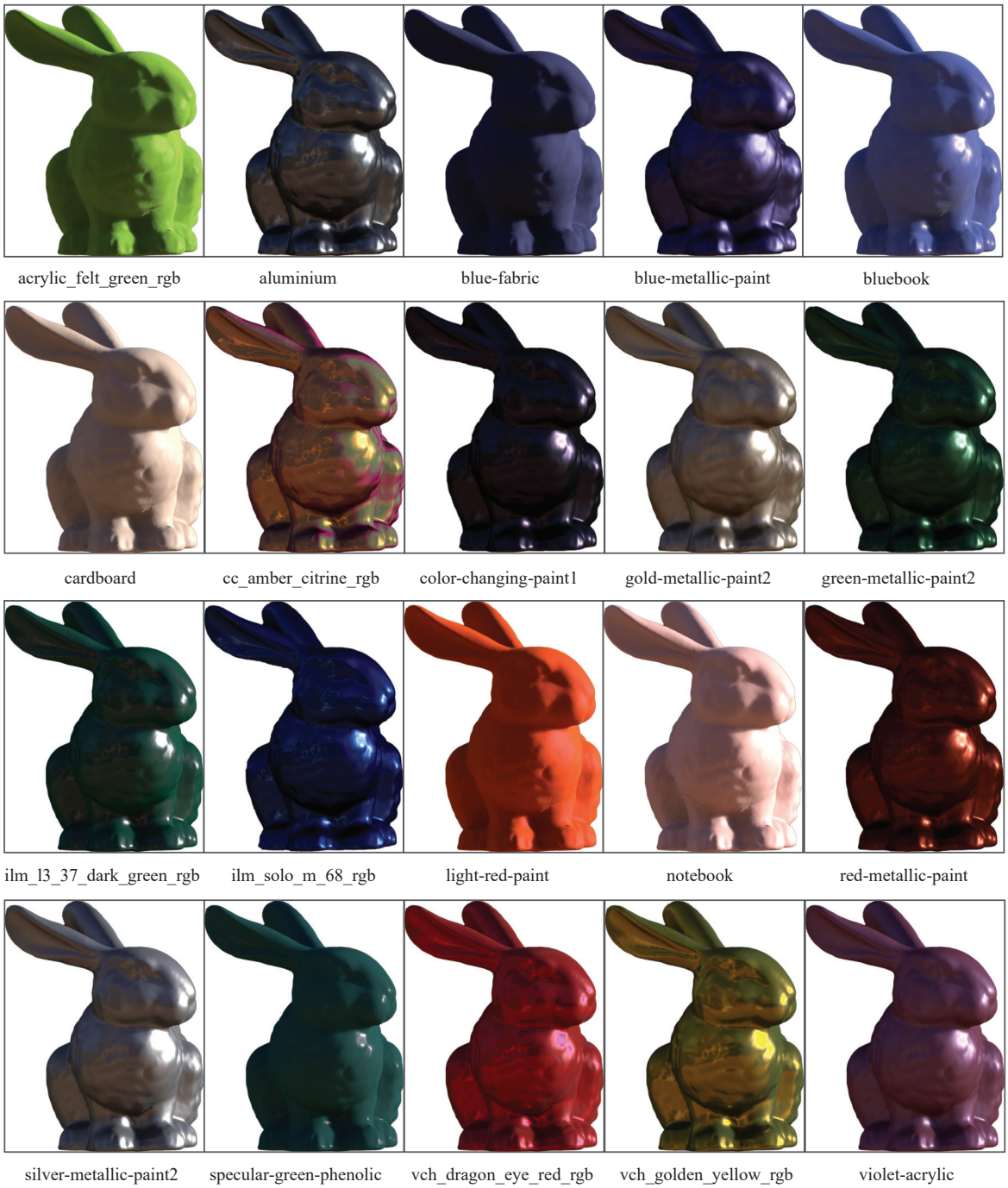


Figure 2: ***Materials used in our experiment.*** *The images show the first frame of HDR video sequences, for which we manually adjusted the exposure and converted the image to the sRGB color space. Note that the materials may appear dim due to the tone mapping applied.*

ing inference, the sparsity was set to the corresponding sparsity of the trained model.

For HYPERBRDF, the encoder and hypernetwork decoder were trained on the full set of available reflectance measurements. During the inference phase, we set the query sample sizes to 40, 400, and 4 000, respectively. These configurations yielded three distinct variants of HYPERBRDF. In total, we considered nine BRDF models in our experiments.

3.2. Rendering

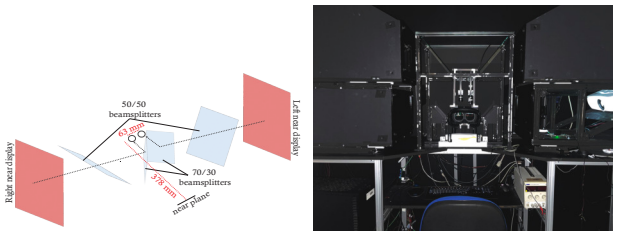
Existing BRDF studies often involve inspecting 2D images on low-dynamic range monitors. However, it has been shown that gloss perception is strongly affected by tone mapping and dynamic range reproduction [AKLM18; CJP*23], as well as the lack of 3D disparity depth cues [WF19]. In the real world, we typically interact with objects to assess material properties; we may rotate an object to examine its appearance from varying angles and geometrical perspectives. To simulate this natural interaction, we generated a set of realistic reproductions of the materials, where each material was reproduced as a stereoscopic HDR video clip, each showing a complex object (*Stanford bunny*, see Figure 2) rotating about the vertical axis so that the video can be looped. We kept the illumination (image-based lighting) and camera position fixed to resemble a natural interaction with an object.

We utilized Mitsuba 3 [JSR*22] to render 150 frames with 1 024 samples per pixel, illuminated using the *Pixar Campus* environment map. The videos were rendered at 30 frames per second, resulting in 5-second video sequences for each material. The 3D scene was configured to show the rotating bunny at about 470 mm from the observer's eyes. The distance was selected to coincide with the focal plane of the display we used in the experiment and reduce potential verge-vs-accommodation conflict. The *Stanford Bunny* was selected as our test geometry due to its optimal combination of surface bumpiness and moderate curvature. Excessive curvature, as indicated in previous studies [HLM08; MKA12], can reduce contrast and negatively impact perceptual evaluation. Moreover, the use of a video sequence, rather than a static image, enabled us to capture a wider range of surface curvatures than would be possible with simpler shapes like spheres or blobs. Additionally, we chose natural illumination using an environment map rather than artificial illumination sources, such as point lights, to reflect realistic viewing conditions. Prior work by Serrano et al. [SCW*21] indicated that varying environment maps minimally affect the perception of glossiness and metallicness—two essential attributes for material discrimination. Thus, we proceeded with the *Pixar Campus* environment map as the illumination source throughout our dataset.

To direct the observer's focus solely on the object, the environment map was hidden during rendering. The video for each eye was cropped to 880×970 px (off-center projection) and spanned approximately $6^\circ \times 6.6^\circ$ visual field. The rendered frames were initially stored in the OpenEXR format in absolute BT.709 RGB color space, then encoded into HDR video (BT.2020 with the PQ EOTF, SMPTE ST 2084). The frames were rendered on a cluster of NVIDIA A100 GPUs, and rendering each video sequence took approximately 130 minutes.

4. BRDF quality assessment experiment

In this section, we measure, in a quality assessment experiment, how well each BRDF model can approximate each material sample. Due to the large size of our dataset, the perceptual quality experiment was conducted on a subset of 20 materials selected from the entire set of 159. This subset will be then used in Section 5 to address the central question of our work: how well do error metrics in BRDF space correlate with human assessment of rendering quality? The complete set of 159 materials will be then used in Section 6 to assess the effectiveness of BRDF-space metrics as loss functions for BRDF fitting.



(a) (b)

Figure 3: *Haploscope used in the experiment. (a) Physical arrangement of mirrors, beamsplitters, and HDR displays. (b) Photograph of the display.*

4.1. Stimuli

Our goal was to create a dataset that is sufficient for the evaluation of quality metrics. Therefore, we prioritized the diversity of quality levels rather than coverage of materials and BRDF models. We selected 20 BRDFs for our experiments, including 11 from MERL, 6 from RGL-EPFL, and 3 from DTU. Figure 2 shows the first frame of the reference video sequences rendered for each of these materials. The selection was made to represent a diverse range of diffuse, glossy, and specular materials, which should ensure the generalization of our findings to other materials in the dataset.

4.2. Apparatus

The experiment was carried out on a custom-built multifocal HDR stereoscopic display, similar to the one described in [ZJY*21]. The observers could view a stereoscopic video through an optical setup similar to a Wheatstone mirror stereoscope (see Figure 3). The stereoscopic projection was calibrated for each participant by performing a grid alignment task. The procedure allowed us to determine the position of the observer's eyes and ensure correct geometric projection (e.g., account for the inter-pupillary distance). We used this display not only because of its accurate stereoscopic calibration but also for its HDR capabilities, which are essential for faithful reproduction of the gloss [CJP*23]. We could accurately reproduce color and luminance up to 4 000 cd/m^2, with the black level below 0.01 cd/m^2, offering a dynamic range far superior to that of most commercially available displays. The display was colorimetrically calibrated using a spectroradiometer (Specbos 1211).

4.3. Experimental procedure

We employed a pairwise comparison protocol in which an observer was presented with three stereoscopic videos played simultaneously: a reference video rendered using the original measured BRDF of the material and two test videos, each rendered using a different BRDF model. One of the test videos could be the reference. Observers were instructed to select the test video for which the material appearance was closer to the reference.

For better data efficiency of the experiment, we used an active sampling technique, specifically ASAP [MWP*21], to strategically select the pairs of conditions to compare. ASAP uses previous measurements to schedule pairs of conditions that maximize information gain. Each participant completed between two and three full ASAP batches of comparisons, corresponding to an average of 450 pairwise comparisons per participant.

Participants

We recruited 20 participants (7 female and 13 male) aged 14 to 46 years. Prior to the experiment, each participant received a briefing form and signed a consent form. The study protocol was approved by the departmental ethics committee. Participants were compensated for their involvement in the study. Each experimental session lasted between 30 and 60 minutes. The participants were encouraged to take short breaks during the study.

4.4. Results

To scale and analyze the results of our pairwise comparison experiment, we adopted the Bayesian scaling approach, which represents the quality in just-objectionable difference (JOD) units [PM17]. Before scaling, we used the outlier analysis from the *pwcmp* package to exclude two observers whose responses were substantially different from the others (according to the interquartile distance criterion) [PM17]. The reference condition was assigned a quality score of 10 JOD, and the quality values decreased for lower quality. 1 JOD difference between two conditions means that 75% of observers selected one condition as better than the other.

The subjective quality scores of the BRDF models for each material are shown in Figure 4 and scaled across all materials in Figure 5. The results indicate that some materials are much more difficult to model than others. For example, none of the BRDF models could match the quality of the reference for `bluebook`, `specular-green-phenolic`, or `violet-acrylic`, indicating that these materials exhibit complex reflectance properties—such as multilayered surfaces or strong specular highlights—that are not well captured by current models. The learning-based models performed generally better than the parametric models. GGX typically provided better approximations than WARD's model due to its greater flexibility in modeling glossy surfaces. With a few exceptions, NEURALBRDF provided the most accurate approximation of the material samples, consistently ranked at the top for challenging materials such as `silver-metallic-paint2`, `color-changing-paint1`, and all materials in the RGL-EPFL dataset but at the cost of requiring the largest number of parameters. The HYPERBRDF variants also outperformed the SPARSEBRDF family, especially on complex or specular materials. Among them, HYPERBRDF-L0, which employs the highest number of parameters (4000), achieved the best performance. It is important to note that direct comparisons between HYPERBRDF and SPARSEBRDF are complicated due to differences in their model capacities. For simpler, more diffuse materials (`acrylic_felt_green_rgb`, `blue-fabric`, and `notebook`), almost all models, including the learning-based and parametric ones, achieved high subjective score. These findings support an adaptive approach to BRDF modeling: lightweight parametric models are well-suited for diffuse surfaces with minimal perceptual sacrifice, while high-capacity learning-based models are necessary to accurately capture the appearance of visually complex or highly specular materials. The overall results in Figure 5 show that NEURALBRDF is the best data-driven model, while GGX is the best parametric model.

Table 2: *State-of-the-art BRDF-space metrics. Here, ρ_{ref} is the reference and ρ_{model} is the approximated BRDF, $C = \cos(\theta_i) \cdot \cos(\theta_o)$, $\epsilon = 10^{-\lambda}$, with λ ranging from 0 to 6, and N is the number of reflectance measurements. The abbreviations stand for RMS: root-mean-squared, E: error, Log: logarithmic, W: weighted, CR: cubic root, and MA: mean absolute. We do not list the remaining MA-variants as these are analogous to the RMS- variants.*

	BRDF-space metric
RMS-E	$\sqrt{\frac{1}{N} \cdot \sum (\rho_{\text{model}} - \rho_{\text{ref}})^2}$
RMS-LogE	$\sqrt{\frac{1}{N} \cdot \sum (\log(\rho_{\text{model}} + \epsilon) - \log(\rho_{\text{ref}} + \epsilon))^2}$
RMS-LogWE	$\sqrt{\frac{1}{N} \cdot \sum (\log(\rho_{\text{model}} \cdot C + \epsilon) - \log(\rho_{\text{ref}} \cdot C + \epsilon))^2}$
RMS-CRWE	$\sqrt{\frac{1}{N} \cdot \sum \sqrt[3]{(\rho_{\text{model}} \cdot C - \rho_{\text{ref}} \cdot C)^2}}$
MA-LogWE	$\frac{1}{N} \cdot \sum \lvert \log(\rho_{\text{model}} \cdot C + \epsilon) - \log(\rho_{\text{ref}} \cdot C + \epsilon) \rvert$

5. Quality metrics for BRDF

The results of our study presented in the previous sections are limited to a single geometry, a single illumination map, and just 20 materials. However, the measured subjective scores are sufficient to evaluate automated quality metrics for the BRDF materials. If we can find a reliable metric, it can be used to label much larger datasets. Therefore, in this section, we assess the performance of existing metrics in both BRDF-space and image-space and compute the correlation of their results with the collected subjective data. We report the Spearman rank-order correlation coefficients (SROCC), where a higher SROCC value indicates stronger alignment between the metric outputs and the subjective JOD values. Other performance metrics (RMSE, PLCC, KROCC) and detailed results can be found in the supplementary HTML report.

5.1. BRDF-space metrics

We evaluated eight different BRDF-space metrics and loss functions, five of which are listed in Table 2. For RMS-LOGE, RMS-LOGWE, MA-LOGE, and MA-LOGWE, we tested the $\epsilon = 10^{-\lambda}$ value from $\lambda = 0$ to $\lambda = 6$ to examine how the scalar added for numerical stability and robustness to noise affects metric performance.

Figure 6 illustrates the SROCC results for BRDF-space metrics. Before computing the metrics, grazing angles were discarded for all models. The results show a poor correlation between BRDF-space metrics and the subjective data, with the maximum correlation reaching only 0.553. The choice of the ϵ constant had a notable impact on performance, as it controls the level of noise suppression. When set too high (e.g., 1), both signal and noise are overly

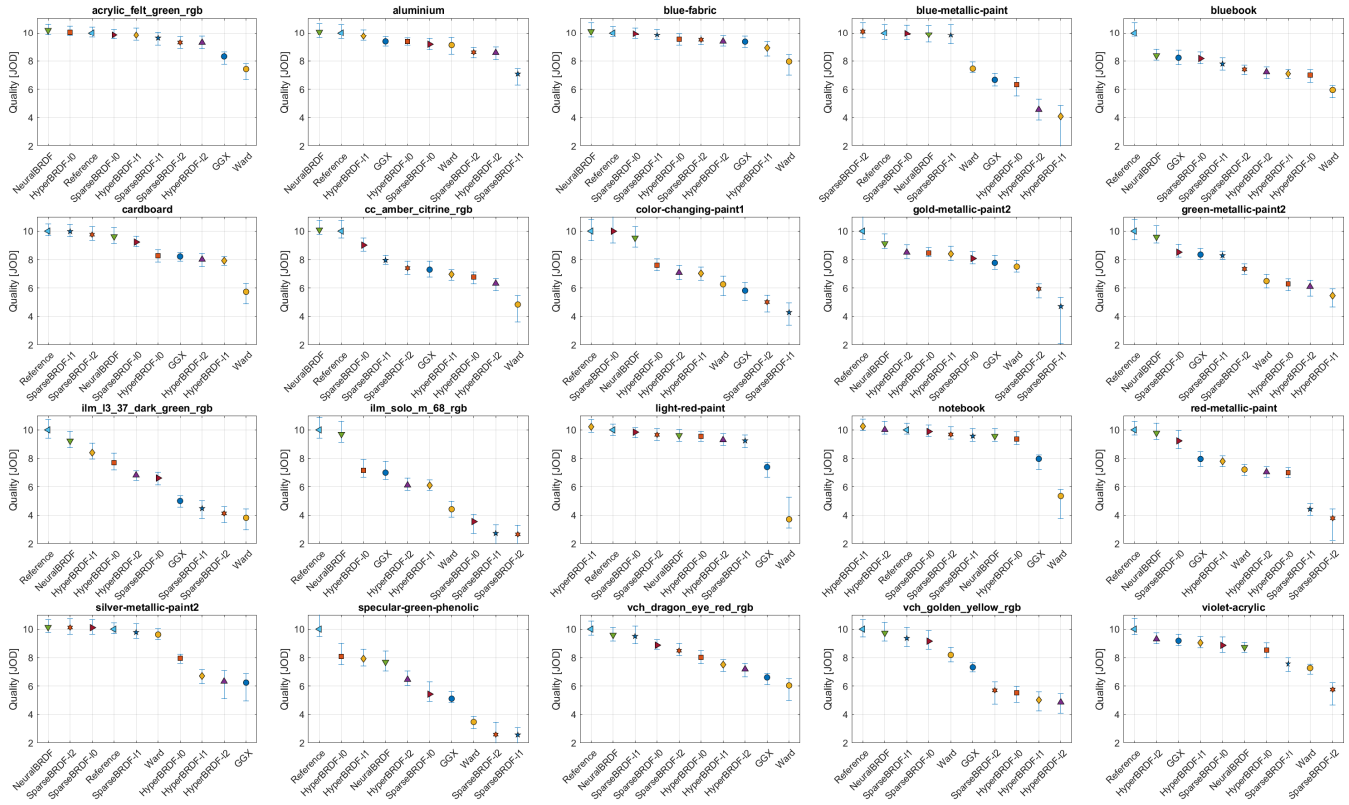


Figure 4: ***The results of the subjective BRDF quality experiment.*** *The y-axis is the quality in JOD units (the higher, the better). Each plot shows results for a single material. The models are ordered from the best to the worst. The error bars denote 95% confidence intervals. The order of the plots is the same as for the material samples shown in Figure 2.*

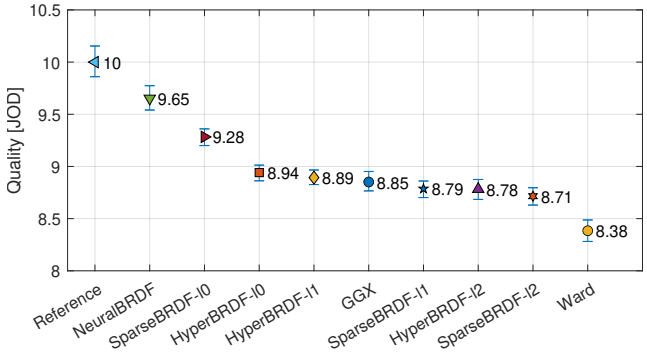


Figure 5: *The subjective experiment results scaled across all materials. The notation is the same as in Figure 4.*

attenuated, reducing prediction accuracy. In contrast, values that are too small allow measurement noise to dominate, leading to degraded quality estimates. The highest correlation was achieved with $\epsilon = 0.0001$. Further gains were achieved by using absolute differences (MA variants) and cosine weighting (WE variants), helping in better alignment of BRDF-space metrics with subjective quality scores.

Lavoué et al. [LBFS21] identified RMS-LOGWE and MA-LOGWE with $\epsilon = 0.001$ as the best L_p metrics in their study. We found that both metrics offer better performance, when $\epsilon = 0.0001$ with MA-LOGWE outperforming RMS-LOGWE (SROCC 0.553 vs. 0.518). However, our obtained correlations are much smaller than those reported in their work (∼0.8). These discrepancies are likely due to the differences in the datasets. Their dataset was measured for the MERL BRDF materials and mostly parametric BRDF models, while we used materials from three different datasets and included modern learning-based BRDF models. More importantly, we could accurately reproduce material properties on our stereoscopic HDR display, while no such guarantees were possible for the stimuli shown in their crowd-sourced experiment.

5.2. Image-space metrics

Next, we tested the performance of 33 state-of-the-art image/video-space metrics, listed in Table 3. The correlation values for those metrics are shown in the bottom part of Figure 6.

Because our test and reference BRDF video clips were in HDR format, we used PU21 [MA21] encoding to make the content suitable for processing with non-HDR metrics. The metrics that could operate on HDR content were supplied with video frames in the appropriate format, e.g., linear BT.709, or linear BT.2020, color

Table 3: *A list of image-space metrics used in the evaluation. The "color" column states if the metric accounts for colors. The "temporal" column states if the metric considers temporal information, and the "HDR" column indicates whether the metric natively supports HDR or requires a PU21 encoding [MA21].*

Metric	Color	Temporal	HDR	Details
PSNR-Y	No	No	PU21	Mathematical measure of the pixel-wise difference.
SSIM [WBSS04b]	No	No	PU21	Measure of difference in luminance, contrast, and structural information.
MS-SSIM [WSB03]	No	No	PU21	Multi-scale variant of SSIM.
IW-SSIM [WL11]	No	No	PU21	SSIM weighted by local information content.
GMSD [XZMB13]	No	No	PU21	Gradient magnitude similarity measure.
MS-GMSD [ZSB17]	No	No	PU21	Multi-scale variant of GMSD.
DSS [BSMP15]	No	No	PU21	DCT subbands' similarity measure.
NLPD [LBBS16]	No	No	PU21	Distance measure of the Laplacian pyramid's normalized contrast.
FSIMc [ZZMZ11]	Yes	No	PU21	A phase congruency and gradients' magnitude similarity measure.
VSI [ZSL14]	Yes	No	PU21	Visual saliency-based weighting of features' similarity.
HaarPSI [RBKW18]	Yes	No	PU21	A Haar wavelet coefficients' similarity measure.
sCIELab [ZW*96]	Yes	No	native	A CSF-based difference measure in the CIELab color space.
HDR-FLIP [ANA*20]	Yes	No	native	Fusion of color and feature differences in a perceptual uniform space.
CIEDE2000 [SWD05]	Yes	No	native	A perceptual color measure in the CIELab color space.
Hybrid Delta E L*a*b* [AAF20]	Yes	No	native	A taxicab distance between the lightness and chroma in the CIELab color space.
DeltaE ITP [ITU19]	Yes	No	native	A color measure in the ICtCp color space.
LPIPS [ZIE*18b]	Yes	No	PU21	A difference measure of the activations of a deep neural network (Alex and VGG Nets).
DISTS [DMWS20]	Yes	No	PU21	A DNN-based model tolerant to texture resampling.
AHIQ [LGS*22]	Yes	No	PU21	A hybrid model of CNN and ViT for quality assessment.
TOPIQ [CMH*24]	Yes	No	PU21	A CNN-based model for high-level semantic information extraction for quality assessment.
PieApp [PCMS18]	Yes	No	PU21	A CNN-based metric trained on pairwise comparison data.
STRRED [SB13]	No	Yes	PU21	Entropy differences in wavelet subbands.
VMAF [LBN*18]	No	Yes	PU21	A fusion of VIF, DLM, and temporal information metrics.
HDR-VDP-3 [MHH23]	No	No	native	A visual different predictor based on low-level human vision for image distortions.
ColorVideoVDP [MHA*24]	Yes	Yes	native	A visual different predictor based on spatial, temporal, and chromatic low-level human vision.

spaces depending on the metric. We also provided the experimental display specification to the metrics that require such information (HDR-FLIP, HDR-VDP-3, ColorVideoVDP). For image-based metrics, the average of the metric output over all frames was used. The left-eye videos were used for evaluation.

The results in Figure 6 indicate that the best prediction is achieved by one of the simplest metrics — ΔE color difference in the ITP color space [LPY*16]. Even its spatial extension that incorporates the contrast sensitivity function (ΔE ITP spatial) resulted in much smaller correlations. We can deduce that color information is crucial, as the PSNR computed on RGB values (PSNR-RGB) resulted in a much higher correlation than the one computed on luminance alone (PSNR-Y). The performance of LPIPS varies with the backbone network, and the variant based on AlexNet performs better than VGG for this task. The detailed results (see the supplementary HTML) indicate that the AlexNet variant can better capture quality degradations for complex specular materials, such as `ilm_solo_m_68_rgb` or `vch_golden_yellow_rgb`. The metrics intended for the evaluation of video compression, such as VMAF, STRRED, or VIF, performed poorly. Not surprisingly, blind quality metrics such as BRISQUE and NIQUE failed at the task—those metrics have never been trained on the distortions found in our dataset.

6. BRDF fitting

Our results from the previous section show that the BRDF-space metrics are much less accurate than the image-space metrics when predicting perceived BRDF quality. However, the main advantage of BRDF-space metrics is that they can be directly used as a loss function when fitting BRDF models. Here, we test whether good BRDF-space quality metrics also make good loss functions for fitting BRDF models.

To test the BRDF-space metrics (see Table 2) as loss functions, we fitted the GGX parameters to our dataset using the gradient-free `patternsearch` global optimization method in MATLAB. This method was selected because it is robust to local minima and the selection of a starting point. The fitted parameters were used to render material samples, and the quality of the fit was evaluated on rendered images with ΔE_{ITP}.

Figure 7 illustrates the relationship between the mean SROCC for quality predictions (see Figure 6) and the average ΔE_{ITP} values across all 159 materials in our dataset. The metrics with adjustable ϵ values are connected by dashed lines. Note that a higher SROCC value indicates more accurate quality prediction, and a lower ΔE_{ITP} indicates a better BRDF fit.

Our results highlight the significant impact of logarithmic transformation on the effectiveness of loss functions compared to their linear counterparts (RMS-E and MA-E), consistent with observations reported in prior works [LKYU12; LBFS21]. We found out that RMS-LogE with $\epsilon = 0.001$ is the most effective loss function, achieving the lowest ΔE_{ITP} error while maintaining relatively high SROCC values of 0.501. Consistent with our findings for BRDF-space quality metrics, the choice of ϵ plays a critical role in the performance of the loss function. Specifically, we observed

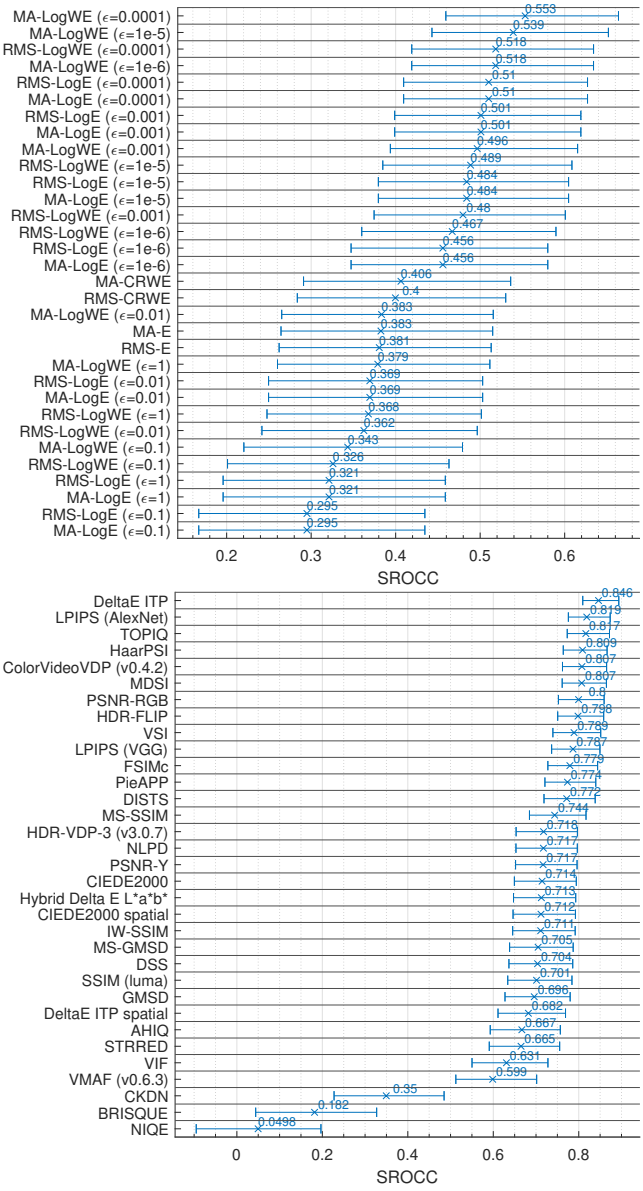


Figure 6: *The Spearman rank-order correlation coefficients for BRDF-space metrics (top) image-space metrics (bottom) tested on our BRDF dataset.*

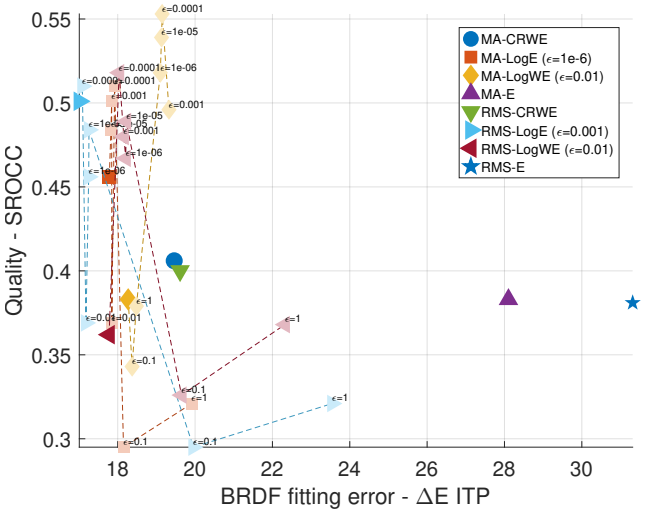


Figure 7: *Relation between the ΔE_{ITP} values from renderings of materials fitted with different BRDF-space metrics as loss functions and their corresponding mean SROCC values as quality metrics. The markers with saturated colors indicate the variant with the ϵ constant that results in the smallest ΔE_{ITP}.*

Table 4: *Comparison of BRDF fitting using the image-driven metric from [BP20] and* RMS-LogE *($\epsilon = 0.001$).*

Method	$\Delta E_{ITP} \downarrow$
[BP20]	16.56
RMS-LogE ($\epsilon = 0.001$)	17.07

that $\epsilon = 1$ results in the poorest performance for RMS-LogE and RMS-LogWE, likely due to the distortion of relative differences for small BRDF values. The mean absolute (MA) variants, however, show less sensitivity to changes in ϵ value.

We also observed that applying cosine weighting slightly reduced the performance of loss functions, despite its clear benefits in metric-based evaluations. This difference may be due to several factors. During optimization, the optimizer appears to prioritize relative differences, and since the logarithmic transformation already compresses the dynamic range, additional weighting has limited impact. Cosine weighting may also emphasize angular regions that, while perceptually important, contribute less to the gradient signal, or may even highlight high-variance regions (such as specular peaks), introducing noise that hinders convergence. Additionally, by suppressing contributions from less-emphasized regions, cosine weighting can reduce gradient diversity, which limits generalization. In contrast, when used as a metric, cosine weighting better reflects the perceptual importance of angular regions in rendered appearance, which explains its improved alignment with subjective assessments.

Additionally, we see a weak correlation between the fitting error and the quality of the metric predictions. For example, RMS-LogE with $\epsilon = 0.001$ is a reasonably good quality predictor with (SROCC=0.496) and one of the best loss functions. The same metric with $\epsilon = 0.01$ is an equally good loss function, but a much worse quality metric.

We also compared the performance of the best-performing fitting metric with the image-driven metric proposed by Bieron et al. [BP20]. To do so, we sweeped the γ parameter in their method from 1 to 3 and rendered the fitted BRDFs using the same illumination setup (*Pixar Campus*) and geometry (*Stanford Bunny*). The best fit was selected based on the lowest ΔE_{ITP} value. Table 4 presents the average ΔE_{ITP} scores across all materials for both approaches, showing that the image-driven metric outperforms the

simpler RMS-LOGE with $\epsilon = 0.001$. This result reinforces the advantage of image-space metrics over BRDF-space metrics in capturing perceptual quality, albeit with significantly higher computational cost—each material required 21 separate optimizations and renderings to determine the best fit.

7. Conclusions

In this paper, we conducted a comprehensive analysis of the impact of BRDF modeling error on perceived quality. To achieve this, we generated the first HDR stereoscopic video dataset of 1 431 different BRDF models and performed a psychophysical experiment using our multifocal HDR display. Our analysis of the correlation between the collected subjective data with both image-space and BRDF-space metrics shows a significantly stronger alignment with image-space metrics, with the Euclidean distance in the ITP color space exhibiting the highest correlation. Additionally, we investigated the use of BRDF-space metrics as loss functions for BRDF fitting. Our results demonstrate that logarithmic transformation is highly effective, both in terms of its correlation with subjective data and its alignment with image-space metrics.

Our experiment included five BRDF models spanning both analytical and machine learning-based approaches, selected to cover a wide range of complexities and representational capacities. It is important to note that these models were not chosen to identify the best-performing BRDF model, but rather to introduce a diverse set of reconstruction errors that could be used to evaluate quality metrics. Due to significant differences in memory usage and computational cost (Table 1), direct comparisons between models fall outside the intended scope of this study.

Acknowledgments

We would like to thank Joseph March and Dongyeon Kim for their help in calibrating the stereoscopic display. The renderings were partially done on the Berzelius supercomputing resource, provided by the National Supercomputer Center at Linköping University and funded by the Knut and Alice Wallenberg Foundation. This work is a part of PRIME, which is funded by the European Union's Horizon 2020 research and innovation program under the Marie Skłodowska Curie grant agreement No 956585. This research was also supported by the Wallenberg Autonomous Systems, AI and Software Program (WASP) and the WASP NEST __main__.

References

[AAF20] ABASI, SAEEDEH, AMANI TEHRAN, MOHAMMAD, and FAIRCHILD, MARK D. "Distance metrics for very large color differences". *Color Research & Application* 45.2 (2020), 208–223 8.

[AKLM18] ADAMS, WENDY J., KUCUKOGLU, GIZEM, LANDY, MICHAEL S., and MANTIUK, RAFAL K. "Naturally glossy: Gloss perception, illumination statistics, and tone mapping". *Journal of Vision* 18.13 (Dec. 2018), 4. ISSN: 1534-7362. DOI: 10.1167/18.13.4 5.

[ANA*20] ANDERSSON, PONTUS, NILSSON, JIM, AKENINE-MÖLLER, TOMAS, et al. "FLIP: A Difference Evaluator for Alternating Images". *Proc. ACM Comput. Graph. Interact. Tech.* 3.2 (2020), 15–1 8.

[AS00] ASHIKHMIN, MICHAEL and SHIRLEY, PETER. "An Anisotropic Phong BRDF Model". *Journal of Graphics Tools* 5.2 (2000), 25–32. DOI: 10.1080/10867651.2000.10487522 2.

[Bli77] BLINN, JAMES F. "Models of light reflection for computer synthesized pictures". *SIGGRAPH Comput. Graph.* 11.2 (July 1977), 192–198. ISSN: 0097-8930. DOI: 10.1145/965141.563893 2.

[BLPW14] BRADY, ADAM, LAWRENCE, JASON, PEERS, PIETER, and WEIMER, WESTLEY. "genBRDF: discovering new analytic BRDFs with genetic programming". *ACM Trans. Graph.* 33.4 (July 2014). ISSN: 0730-0301. DOI: 10.1145/2601097.2601193 3.

[BÖK11] BILGILI, AHMET, ÖZTÜRK, AYDN, and KURT, MURAT. "A General BRDF Representation Based on Tensor Decomposition". *Computer Graphics Forum* 30.8 (2011), 2427–2439. DOI: 10.1111/j.1467-8659.2011.02072.x 2.

[BP20] BIERON, J. and PEERS, P. "An Adaptive BRDF Fitting Metric". *Computer Graphics Forum* 39.4 (2020), 59–74. DOI: 10.1111/cgf.14054 3, 9.

[BSMP15] BALANOV, AMNON, SCHWARTZ, ARIK, MOSHE, YAIR, and PELEG, NIMROD. "Image quality assessment based on DCT subband similarity". IEEE. 2015, 2105–2109 8.

[BSN16] BAGHER, MAHDI M., SNYDER, JOHN, and NOWROUZEZAHRAI, DEREK. "A Non-Parametric Factor Microfacet Model for Isotropic BRDFs". *ACM Trans. Graph.* 35.5 (July 2016). ISSN: 0730-0301. DOI: 10.1145/2907941 2.

[Bur12] BURLEY, BRENT. "Physically based shading at Disney". *ACM SIGGRAPH 2012 Courses*. ACM. 2012, 1–7 2.

[BWW*08] BEN-EZRA, MOSHE, WANG, JIAPING, WILBURN, BENNETT, et al. "An LED-only BRDF measurement device". *2008 IEEE Conference on Computer Vision and Pattern Recognition*. 2008, 1–8. DOI: 10.1109/CVPR.2008.4587766 2.

[CBP22] COOPER, VICTORIA L., BIERON, JAMES C., and PEERS, PIETER. "Estimating Homogeneous Data-Driven BRDF Parameters From a Reflectance Map Under Known Natural Lighting". *IEEE Transactions on Visualization and Computer Graphics* 28.12 (2022), 4289–4303. DOI: 10.1109/TVCG.2021.3085560 2.

[CJP*23] CHEN, BIN, JINDAL, AKSHAY, PIOVARČI, MICHAL, et al. "The effect of display capabilities on the gloss consistency between real and virtual objects". en. *SIGGRAPH Asia 2023 Conference Papers*. Sydney NSW Australia: ACM, Dec. 2023, 1–11. ISBN: 9798400703157. DOI: 10.1145/3610548.3618226 5.

[CMF18] CLAUSEN, O., MARROQUIM, R., and FUHRMANN, A. "Acquisition and Validation of Spectral Ground Truth Data for Predictive Rendering of Rough Surfaces". *Computer Graphics Forum* 37.4 (2018), 1–12. DOI: 10.1111/cgf.13470 2.

[CMH*24] CHEN, CHAOFENG, MO, JIADI, HOU, JINGWEN, et al. "TOPIQ: A top-down approach from semantics to distortions for image quality assessment". (2024) 8.

[CNN20] CHEN, ZHE, NOBUHARA, SHOHEI, and NISHINO, KO. "Invertible Neural BRDF for Object Inverse Rendering". *Proceedings of the European Conference on Computer Vision (ECCV)*. 2020 2.

[CT81] COOK, ROBERT L. and TORRANCE, KENNETH E. "A reflectance model for computer graphics". *Proceedings of the 8th Annual Conference on Computer Graphics and Interactive Techniques*. SIGGRAPH '81. Dallas, Texas, USA: Association for Computing Machinery, 1981, 307–316. ISBN: 0897910451. DOI: 10.1145/800224.806819 2.

[Dan01] DANA, K.J. "BRDF/BTF measurement device". *Proceedings Eighth IEEE International Conference on Computer Vision. ICCV 2001*. Vol. 2. 2001, 460–466 vol.2. DOI: 10.1109/ICCV.2001.937661 2.

[DHT*00] DEBEVEC, PAUL, HAWKINS, TIM, TCHOU, CHRIS, et al. "Acquiring the Reflectance Field of a Human Face". *Proceedings of the 27th Annual Conference on Computer Graphics and Interactive Techniques*. SIGGRAPH '00. USA: ACM Press/Addison-Wesley Publishing Co., 2000, 145–156. ISBN: 1581132085. DOI: 10.1145/344779.344855 2.

[DJ18] DUPUY, JONATHAN and JAKOB, WENZEL. "An Adaptive Parameterization for Efficient Material Acquisition and Rendering". *Transactions on Graphics (Proceedings of SIGGRAPH Asia)* 37.6 (Nov. 2018), 274:1–274:18. DOI: 10.1145/3272127.3275059 1–3.

[DMWS20] DING, KEYAN, MA, KEDE, WANG, SHIQI, and SIMONCELLI, EERO P. "Image quality assessment: Unifying structure and texture similarity". 44.5 (2020), 2567–2581 8.

[ELU11] EILERTSEN, GABRIEL, LARSSON, PER, and UNGER, JONAS. "A versatile material reflectance measurement system for use in production". *Proceedings of SIGRAD*. Linköping University Press", 2011. ISBN: 978-91-7393-008-6 2.

[FDS*24] FILIP, JIRI, DECHTERENKO, FILIP, SCHMIDT, FILIPP, et al. *Material Fingerprinting: Identifying and Predicting Perceptual Attributes of Material Appearance*. 2024. arXiv: 2410.13615 [cs.CV] 3.

[FFG12] FORÉS, ADRIÀ, FERWERDA, JAMES, and GU, JINWEI. "Toward a Perceptually Based Metric for BRDF Modeling". *Final Program and Proceedings - IS and T/SID Color Imaging Conference* CIC'12 (Jan. 2012), 142–148. DOI: 10.2352/CIC.2012.20.1.art00025 2.

[Foo97] FOO, SING. "A Gonioreflectometer For Measuring The Bidirectional Reflectance Of Material For Use In Illumination Computation". (Feb. 1997) 2.

[FR22] FISCHER, MICHAEL and RITSCHEL, TOBIAS. "Metappearance: Meta-Learning for Visual Appearance Reproduction". *ACM Trans. Graph.* 41.6 (Nov. 2022). ISSN: 0730-0301. DOI: 10.1145/3550454.3555458 2.

[FWH*22] FAN, JIAHUI, WANG, BEIBEI, HASAN, MILOS, et al. "Neural Layered BRDFs". *ACM SIGGRAPH 2022 Conference Proceedings*. SIGGRAPH '22. Vancouver, BC, Canada: Association for Computing Machinery, 2022. ISBN: 9781450393379. DOI: 10.1145/3528233.3530732 2.

[GAHO07] GHOSH, ABHIJEET, ACHUTHA, SHRUTHI, HEIDRICH, WOLFGANG, and O'TOOLE, MATTHEW. "BRDF Acquisition with Basis Illumination". *2007 IEEE 11th International Conference on Computer Vision*. 2007, 1–8. DOI: 10.1109/ICCV.2007.4408935 2.

[GSZ*25] GOKBUDAK, FAZILET, SZTRAJMAN, ALEJANDRO, ZHOU, CHENLIANG, et al. "Hypernetworks for Generalizable BRDF Representation". *Computer Vision – ECCV 2024*. Ed. by LEONARDIS, ALEŠ, RICCI, ELISA, ROTH, STEFAN, et al. Cham: Springer Nature Switzerland, 2025, 73–89. ISBN: 978-3-031-73116-7 2, 3.

[HFM16] HAVRAN, V., FILIP, J., and MYSZKOWSKI, K. "Perceptually Motivated BRDF Comparison using Single Image". *Computer Graphics Forum* 35.4 (2016), 1–12. DOI: 10.1111/cgf.12944 3.

[HGC*20] HU, BINGYANG, GUO, JIE, CHEN, YANJUN, et al. "DeepBRDF: A Deep Representation for Manipulating Measured BRDF". *Computer Graphics Forum* 39.2 (2020), 157–166. DOI: 10.1111/cgf.13920 2.

[HHN*17] HAVRAN, VLASTIMIL, HOŠEK, JAN, NĚMCOVÁ, ŠÁRKA, et al. "Lightdrum-Portable Light Stage for Accurate BTF Measurement on Site". *Sensors (Basel)* 17.3 (Feb. 2017), 423. ISSN: 1424-8220. DOI: 10.3390/s17030423 1.

[HLM08] HO, YUN-XIAN, LANDY, MICHAEL S., and MALONEY, LAURENCE T. "Conjoint Measurement of Gloss and Surface Texture". *Psychological Science* 19.2 (2008), 196–204. DOI: 10.1111/j.1467-9280.2008.02067.x 5.

[HP17] HOLZSCHUCH, NICOLAS and PACANOWSKI, ROMAIN. "A two-scale microfacet reflectance model combining reflection and diffraction". *ACM Trans. Graph.* 36.4 (July 2017). ISSN: 0730-0301. DOI: 10.1145/3072959.3073621 2.

[ITU19] ITU-R. *BT.2124-0: Objective metric for the assessment of the potential visibility of colour differences in television*. 2019 8.

[JSR*22] JAKOB, WENZEL, SPEIERER, SÉBASTIEN, ROUSSEL, NICOLAS, et al. *Mitsuba 3 renderer*. Version 3.1.1. https://mitsuba-renderer.org. 2022 5.

[KHM*24] KAVOOSIGHAFI, BEHNAZ, HAJISHARIF, SAGHI, MIANDJI, EHSAN, et al. "Deep SVBRDF acquisition and modelling: a survey". *Computer Graphics Forum* 43.6 (2024), e15199. DOI: 10.1111/cgf.15199 1.

[KM99] KAUTZ, JAN and MCCOOL, MICHAEL D. "Interactive Rendering with Arbitrary BRDFs using Separable Approximations". *Rendering Techniques' 99*. Ed. by LISCHINSKI, DANI and LARSON, GREG WARD. Vienna: Springer Vienna, 1999, 247–260. ISBN: 978-3-7091-6809-7 2.

[KN06] KUTHIRUMMAL, SUJIT and NAYAR, SHREE K. "Multiview radial catadioptric imaging for scene capture". *ACM Trans. Graph.* 25.3 (July 2006), 916–923. ISSN: 0730-0301. DOI: 10.1145/1141911.1141975 2.

[KSK10] KURT, MURAT, SZIRMAY-KALOS, LÁSZLÓ, and KŘIVÁNEK, JAROSLAV. "An anisotropic BRDF model for fitting and Monte Carlo rendering". *SIGGRAPH Comput. Graph.* 44.1 (Feb. 2010). ISSN: 0097-8930. DOI: 10.1145/1722991.1722996 1.

[LBBS16] LAPARRA, VALERO, BALLÉ, JOHANNES, BERARDINO, ALEXANDER, and SIMONCELLI, EERO P. "Perceptual image quality assessment using a normalized Laplacian pyramid". *Electronic Imaging* 28 (2016), 1–6 8.

[LBFS21] LAVOUÉ, GUILLAUME, BONNEEL, NICOLAS, FARRUGIA, JEAN-PHILIPPE, and SOLER, CYRIL. "Perceptual quality of BRDF approximations: dataset and metrics". *Computer Graphics Forum* 40.2 (2021), 327–338. DOI: 10.1111/cgf.142636 2, 3, 7, 8.

[LBN*18] LI, ZHI, BAMPIS, CHRISTOS, NOVAK, JULIE, et al. "VMAF: The journey continues". *Netflix Technology Blog* 25.1 (2018) 8.

[LFTG97] LAFORTUNE, ERIC P. F., FOO, SING-CHOONG, TORRANCE, KENNETH E., and GREENBERG, DONALD P. "Non-linear approximation of reflectance functions". *Proceedings of the 24th Annual Conference on Computer Graphics and Interactive Techniques*. SIGGRAPH '97. USA: ACM Press/Addison-Wesley Publishing Co., 1997, 117–126. ISBN: 0897918967. DOI: 10.1145/258734.258801 2.

[LGS*22] LAO, SHANSHAN, GONG, YUAN, SHI, SHUWEI, et al. "Attentions help CNNs see better: Attention-based hybrid image quality assessment network". 2022, 1140–1149 8.

[LKYU12] LÖW, JOAKIM, KRONANDER, JOEL, YNNERMAN, ANDERS, and UNGER, JONAS. "BRDF models for accurate and efficient rendering of glossy surfaces". *ACM Trans. Graph.* 31.1 (Feb. 2012). ISSN: 0730-0301. DOI: 10.1145/2077341.2077350 1–3, 8.

[LMS*19] LAGUNAS, MANUEL, MALPICA, SANDRA, SERRANO, ANA, et al. "A similarity measure for material appearance". *ACM Trans. Graph.* 38.4 (July 2019). ISSN: 0730-0301. DOI: 10.1145/3306346.3323036 3.

[LPU*13] LISSNER, INGMAR, PREISS, JENS, URBAN, PHILIPP, et al. "Image-Difference Prediction: From Grayscale to Color". *IEEE Transactions on Image Processing* 22.2 (2013), 435–446. DOI: 10.1109/TIP.2012.2216279 3.

[LPY*16] LU, TAORAN, PU, FANGJUN, YIN, PENG, et al. "ITP Colour Space and Its Compression Performance for High Dynamic Range and Wide Colour Gamut Video Distribution". *ZTE Communications* 1.1 (2016), 1–7 8.

[LRR04] LAWRENCE, JASON, RUSINKIEWICZ, SZYMON, and RAMAMOORTHI, RAVI. "Efficient BRDF importance sampling using a factored representation". *ACM Trans. Graph.* 23.3 (Aug. 2004), 496–505. ISSN: 0730-0301. DOI: 10.1145/1015706.1015751 2.

[LSZ*24] LI, ZHIQIANG, SHEN, XUKUN, ZHOU, XUEYANG, et al. "Hyper-SNBRDF: Hypernetwork for Neural BRDF Using Sinusoidal Activation". *2024 International Conference on 3D Vision (3DV)*. 2024, 965–974. DOI: 10.1109/3DV62453.2024.00068 2.

[MA21] MANTIUK, RAFAL K. and AZIMI, MARYAM. "PU21: A novel perceptually uniform encoding for adapting existing quality metrics for HDR". *2021 Picture Coding Symposium (PCS)*. Citation Key: Mantiuk2021. IEEE, June 2021, 1–5. ISBN: 978-1-66542-545-2. DOI: 10.1109/PCS50896.2021.9477471 7, 8.

[MDL*98] MATTISON, PHILLIP R., DOMBROWSKI, MARK S., LORENZ, JAMES M., et al. "Handheld directional reflectometer: an angular imaging device to measure BRDF and HDR in real time". *Scattering and Surface Roughness II*. Ed. by GU, ZU-HAN and MARADUDIN, ALEXEI A. Vol. 3426. International Society for Optics and Photonics. SPIE, 1998, 240–251. DOI: 10.1117/12.328461 2.

[MDMS05] MANTIUK, RAFAŁ, DALY, SCOTT J., MYSZKOWSKI, KAROL, and SEIDEL, HANS-PETER. "Predicting visible differences in high dynamic range images: model and its calibration". *Human Vision and Electronic Imaging X*. Ed. by ROGOWITZ, BERNICE E., PAPPAS, THRASYVOULOS N., and DALY, SCOTT J. Vol. 5666. Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series. Mar. 2005, 204–214. DOI: 10.1117/12.586757 2.

[MGW01] MALZBENDER, TOM, GELB, DAN, and WOLTERS, HANS. "Polynomial Texture Maps". *Proceedings of the 28th Annual Conference on Computer Graphics and Interactive Techniques*. SIGGRAPH '01. New York, NY, USA: Association for Computing Machinery, 2001, 519–528. ISBN: 158113374X. DOI: 10.1145/383259.383320 2.

[MHA*24] MANTIUK, RAFAL K., HANJI, PARAM, ASHRAF, MALIHA, et al. "ColorVideoVDP: A visual difference predictor for image, video and display distortions". 43.4 (2024). ISSN: 0730-0301. DOI: 10.1145/3658144 8.

[MHH23] MANTIUK, RAFAL K, HAMMOU, DOUNIA, and HANJI, PARAM. "HDR-VDP-3: A multi-metric for predicting image differences, quality and contrast distortions in high dynamic range and regular content". *arXiv preprint arXiv:2304.13625* (2023) 8.

[MKA12] MARLOW, PHILLIP J., KIM, JUNO, and ANDERSON, BARTON L. "The Perception and Misperception of Specular Surface Reflectance". *Current Biology* 22.20 (2012), 1909–1913. ISSN: 0960-9822. DOI: 10.1016/j.cub.2012.08.009 5.

[MKRH11] MANTIUK, RAFAŁ, KIM, KIL JOONG, REMPEL, ALLAN G., and HEIDRICH, WOLFGANG. "HDR-VDP-2: a calibrated visual metric for visibility and quality predictions in all luminance conditions". *ACM Trans. Graph.* 30.4 (July 2011). ISSN: 0730-0301. DOI: 10.1145/2010324.1964935 2, 3.

[MMS*05] MÜLLER, G., MESETH, J., SATTLER, M., et al. "Acquisition, Synthesis, and Rendering of Bidirectional Texture Functions". *Computer Graphics Forum* 24.1 (2005), 83–109. DOI: 10.1111/j.1467-8659.2005.00830.x 2.

[MPBM03] MATUSIK, WOJCIECH, PFISTER, HANSPETER, BRAND, MATT, and MCMILLAN, LEONARD. "A Data-Driven Reflectance Model". *ACM Trans. Graph.* 22.3 (July 2003), 759–769. ISSN: 0730-0301. DOI: 10.1145/882262.882343 3.

[MTH*24] MIANDJI, EHSAN, TONGBUASIRILAI, TANABOON, HAJISHARIF, SAGHI, et al. "FROST-BRDF: A Fast and Robust Optimal Sampling Technique for BRDF Acquisition". *IEEE Transactions on Visualization and Computer Graphics* 30.7 (Jan. 2024), 4390–4402. ISSN: 1077-2626. DOI: 10.1109/TVCG.2024.3355200 1.

[MWP*21] MIKHAILIUK, ALIAKSEI, WILMOT, CLIFFORD, PEREZ-ORTIZ, MARIA, et al. "Active Sampling for Pairwise Comparisons via Approximate Message Passing and Information Gain Maximization". *2020 IEEE International Conference on Pattern Recognition (ICPR)*. Jan. 2021 6.

[NDM05] NGAN, ADDY, DURAND, FRÉDO, and MATUSIK, WOJCIECH. "Experimental Analysis of BRDF Models". *Eurographics Symposium on Rendering (2005)*. Ed. by BALA, KAVITA and DUTRE, PHILIP. The Eurographics Association, 2005. ISBN: 3-905673-23-1. DOI: 10.2312/EGWR/EGSR05/117-126 1–3.

[Nic65] NICODEMUS, FRED E. "Directional Reflectance and Emissivity of an Opaque Surface". *Appl. Opt.* 4.7 (July 1965), 767–775. DOI: 10.1364/AO.4.000767 1.

[NJR15] NIELSEN, JANNIK BOLL, JENSEN, HENRIK WANN, and RAMAMOORTHI, RAVI. "On Optimal, Minimal BRDF Sampling for Reflectance Acquisition". *ACM Transactions on Graphics (TOG)* 34.6 (Nov. 2015), 186:1–186:11. DOI: 10.1145/2816795.2818085 2, 3.

[NZV*11] NAIK, NIKHIL, ZHAO, SHUANG, VELTEN, ANDREAS, et al. "Single view reflectance capture using multiplexed scattering and time-of-flight imaging". *ACM Trans. Graph.* 30.6 (Dec. 2011), 1–10. ISSN: 0730-0301. DOI: 10.1145/2070781.2024205 2.

[PCMS18] PRASHNANI, EKTA, CAI, HONG, MOSTOFI, YASAMIN, and SEN, PRADEEP. "Pieapp: Perceptual image-error assessment through pairwise preference". *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 2018, 1808–1817 8.

[Pho75] PHONG, BUI TUONG. "Illumination for computer generated pictures". *Commun. ACM* 18.6 (June 1975), 311–317. ISSN: 0001-0782. DOI: 10.1145/360825.360839 2.

[PL07] PELLACINI, FABIO and LAWRENCE, JASON. "AppWand: editing measured materials using appearance-driven optimization". *ACM Trans. Graph.* 26.3 (July 2007), 54–es. ISSN: 0730-0301. DOI: 10.1145/1276377.1276444 2.

[PM17] PEREZ-ORTIZ, MARIA and MANTIUK, RAFAL K. *A practical guide and software for analysing pairwise comparison experiments*. 2017. arXiv: 1712.03686 [stat.AP] 6.

[PSS*12] PACANOWSKI, ROMAIN, SALAZAR CELIS, OLIVER, SCHLICK, CHRISTOPHE, et al. "Rational BRDF". *IEEE Transactions on Visualization and Computer Graphics* 18.11 (2012), 1824–1835. DOI: 10.1109/TVCG.2012.73 2.

[RBKW18] REISENHOFER, RAFAEL, BOSSE, SEBASTIAN, KUTYNIOK, GITTA, and WIEGAND, THOMAS. "A Haar wavelet-based perceptual similarity index for image quality assessment". *Signal Processing: Image Communication* 61 (2018), 33–43 8.

[RMS*08] RUMP, MARTIN, MÜLLER, GERO, SARLETTE, RALF, et al. "Photo-realistic Rendering of Metallic Car Paint from Image-Based Measurements". *Computer Graphics Forum* 27.2 (2008), 527–536. DOI: 10.1111/j.1467-8659.2008.01150.x 2.

[Rus98] RUSINKIEWICZ, SZYMON M. "A New Change of Variables for Efficient BRDF Representation". *Rendering Techniques '98*. Ed. by DRETTAKIS, GEORGE and MAX, NELSON. Vienna: Springer Vienna, 1998, 11–22. ISBN: 978-3-7091-6453-2 3.

[SB13] SOUNDARARAJAN, RAJIV and BOVIK, ALAN C. "Video Quality Assessment by Reduced Reference Spatio-Temporal Entropic Differencing". *IEEE Transactions on Circuits and Systems for Video Technology* 23.4 (Apr. 2013), 684–694. ISSN: 1051-8215. DOI: 10.1109/TCSVT.2012.2214933 8.

[SCW*21] SERRANO, ANA, CHEN, BIN, WANG, CHAO, et al. "The effect of shape and illumination on material perception: model and applications". *ACM Trans. Graph.* 40.4 (July 2021). ISSN: 0730-0301. DOI: 10.1145/3450626.3459813 5.

[SJR18] SUN, TIANCHENG, JENSEN, HENRIK WANN, and RAMAMOORTHI, RAVI. "Connecting measured BRDFs to analytic BRDFs by data-driven diffuse-specular separation". *ACM Trans. Graph.* 37.6 (Dec. 2018). ISSN: 0730-0301. DOI: 10.1145/3272127.3275026 2.

[SRRW21] SZTRAJMAN, ALEJANDRO, RAINER, GILLES, RITSCHEL, TOBIAS, and WEYRICH, TIM. "Neural BRDF Representation and Importance Sampling". *Computer Graphics Forum* 40.6 (Sept. 2021), 332–346. DOI: 10.1111/cgf.14335 2, 3.

[SSN18] SOLER, CYRIL, SUBR, KARTIC, and NOWROUZEZAHRAI, DEREK. "A Versatile Parameterization for Measured Material Manifolds". *Computer Graphics Forum* 37.2 (2018), 135–144. DOI: 10.1111/cgf.13348 2.

[SWD05] SHARMA, GAURAV, WU, WENCHENG, and DALAL, EDUL N. "The CIEDE2000 color-difference formula: Implementation notes, supplementary test data, and mathematical observations". *Color Research Application* 30.1 (Feb. 2005). Citation Key: Sharma2005, 21–30. ISSN: 0361-2317. DOI: 10.1002/col.20070 8.

[SZC*07] SUN, XIN, ZHOU, KUN, CHEN, YANYUN, et al. "Interactive relighting with dynamic BRDFs". *ACM Trans. Graph.* 26.3 (July 2007), 27–es. ISSN: 0730-0301. DOI: 10.1145/1276377.1276411 2.

[TS67] TORRANCE, KENNETH E. and SPARROW, E. M. "Theory for off-specular reflection from roughened surfaces". *Journal of the Optical Society of America* 57.9 (1967), 1105–1114 2.

[TUGM22] TONGBUASIRILAI, TANABOON, UNGER, JONAS, GUILLEMOT, CHRISTINE, and MIANDJI, EHSAN. "A Sparse Non-parametric BRDF Model". *ACM Trans. Graph.* 41.5 (Oct. 2022). ISSN: 0730-0301. DOI: 10.1145/3533427 2, 3.

[TUKK20] TONGBUASIRILAI, TANABOON, UNGER, JONAS, KRONANDER, JOEL, and KURT, MURAT. "Compact and intuitive data-driven BRDF models". *The Visual Computer* 36.4 (Apr. 2020), 855–872. ISSN: 1432-2315. DOI: 10.1007/s00371-019-01664-z 2.

[VS13] VARSHNEY, LAV R. and SUN, JOHN Z. "Why do we perceive logarithmically?": *Significance* 10.1 (2013), 28–31. DOI: 10.1111/j.1740-9713.2013.00636.x 2.

[War92] WARD, GREGORY J. "Measuring and modeling anisotropic reflection". *Proceedings of the 19th Annual Conference on Computer Graphics and Interactive Techniques*. SIGGRAPH '92. New York, NY, USA: Association for Computing Machinery, 1992, 265–272. ISBN: 0897914791. DOI: 10.1145/133994.134078 2, 3.

[WBSS04a] WANG, ZHOU, BOVIK, A.C., SHEIKH, H.R., and SIMONCELLI, E.P. "Image quality assessment: from error visibility to structural similarity". *IEEE Transactions on Image Processing* 13.4 (2004), 600–612. DOI: 10.1109/TIP.2003.819861 2.

[WBSS04b] WANG, ZHOU, BOVIK, ALAN C, SHEIKH, HAMID R, and SIMONCELLI, EERO P. "Image quality assessment: from error visibility to structural similarity". 13.4 (2004), 600–612 8.

[WF19] WENDT, GUNNAR and FAUL, FRANZ. "Differences in Stereoscopic Luster Evoked by Static and Dynamic Stimuli". en. *i-Perception* 10.3 (May 2019), 2041669519846133. ISSN: 2041-6695, 2041-6695. DOI: 10.1177/2041669519846133 5.

[WL11] WANG, ZHOU and LI, QIANG. "Information Content Weighting for Perceptual Image Quality Assessment". *IEEE Transactions on Image Processing* 20.5 (May 2011), 1185–1198. ISSN: 1057-7149, 1941-0042. DOI: 10.1109/TIP.2010.2092435 8.

[WLT04] WESTIN, STEPHEN H., LI, HONGSONG, and TORRANCE, KENNETH E. *A Comparison of Four BRDF Models*. Technical Report PCG-04-02. Cornell University, 2004 2.

[WMLT07] WALTER, BRUCE, MARSCHNER, STEPHEN R., LI, HONGSONG, and TORRANCE, KENNETH E. "Microfacet Models for Refraction through Rough Surfaces". *Rendering Techniques*. Ed. by KAUTZ, JAN and PATTANAIK, SUMANTA. The Eurographics Association, 2007. ISBN: 978-3-905673-52-4. DOI: /10.2312/EGWR/EGSR07/195-206 2, 3.

[WSB*98] WHITE, DAVID, SAUNDERS, PETER, BONSEY, STUART, et al. "Reflectometer for Measuring the Bidirectional Reflectance of Rough Surfaces". *Applied optics* 37 (1998), 3450–4. DOI: 10.1364/AO.37.003450 2.

[WSB03] WANG, ZHOU, SIMONCELLI, EERO P, and BOVIK, ALAN C. "Multiscale structural similarity for image quality assessment". *Asilomar Conference on Signals, Systems & Computers*. Vol. 2. IEEE. 2003, 1398–1402 8.

[XZMB13] XUE, WUFENG, ZHANG, LEI, MOU, XUANQIN, and BOVIK, ALAN C. "Gradient magnitude similarity deviation: A highly efficient perceptual image quality index". 23.2 (2013), 684–695 8.

[YSX*24] YUAN, YUE, SUN, RUODUAN, XU, CHEN, et al. "Design of an image-based BRDF measurement method using a catadioptric multispectral capture and a real-time Lambert calibration". *Opt. Express* 32.1 (Jan. 2024), 425–443. DOI: 10.1364/OE.510627 2.

[ZIE*18a] ZHANG, RICHARD, ISOLA, PHILLIP, EFROS, ALEXEI A, et al. "The Unreasonable Effectiveness of Deep Features as a Perceptual Metric". *CVPR*. 2018 3.

[ZIE*18b] ZHANG, RICHARD, ISOLA, PHILLIP, EFROS, ALEXEI A, et al. "The Unreasonable Effectiveness of Deep Features as a Perceptual Metric". 2018 8.

[ZJY*21] ZHONG, FANGCHENG, JINDAL, AKSHAY, YÖNTEM, ALI ÖZGÜR, et al. "Reproducing reality with a high-dynamic-range multifocal stereo display". *ACM Trans. Graph.* 40.6 (Dec. 2021). ISSN: 0730-0301. DOI: 10.1145/3478513.3480513 5.

[ZSB17] ZHANG, BO, SANDER, PEDRO V, and BERMAK, AMINE. "Gradient magnitude similarity deviation on multiple scales for color image quality assessment". IEEE. 2017, 1253–1257 8.

[ZSL14] ZHANG, LIN, SHEN, YING, and LI, HONGYU. "VSI: A visual saliency-induced index for perceptual image quality assessment". 23.10 (2014), 4270–4281 8.

[ZSR*24] ZHOU, CHENLIANG, SZTRAJMAN, ALEJANDRO, RAINER, GILLES, et al. *Physically Based Neural Bidirectional Reflectance Distribution Function*. 2024. arXiv: 2411.02347 [cs.GR] 2.

[ZW*96] ZHANG, XUEMEI, WANDELL, BRIAN A, et al. "A spatial extension of CIELAB for digital color image reproduction". *SID international symposium digest of technical papers*. Vol. 27. Citeseer. 1996, 731–734 8.

[ZZMZ11] ZHANG, LIN, ZHANG, LEI, MOU, XUANQIN, and ZHANG, DAVID. "FSIM: A feature similarity index for image quality assessment". 20.8 (2011), 2378–2386 8.