DOI: 10.1111/cgf.13470
Eurographics Symposium on Rendering 2018
T. Hachisuka and W. Jakob
(Guest Editors)

Volume 37 (2018), Number 4

Acquisition and Validation of Spectral Ground Truth Data for Predictive Rendering of Rough Surfaces

O. Clausen[1,2] R. Marroquim[2] and A. Fuhrmann[1]

[1]TH Köln, Germany
{olaf.clausen, arnulph.fuhrmann}@th-koeln.de
[2]Federal University of Rio de Janeiro, Brazil
marroquim@lcg.ufrj.br

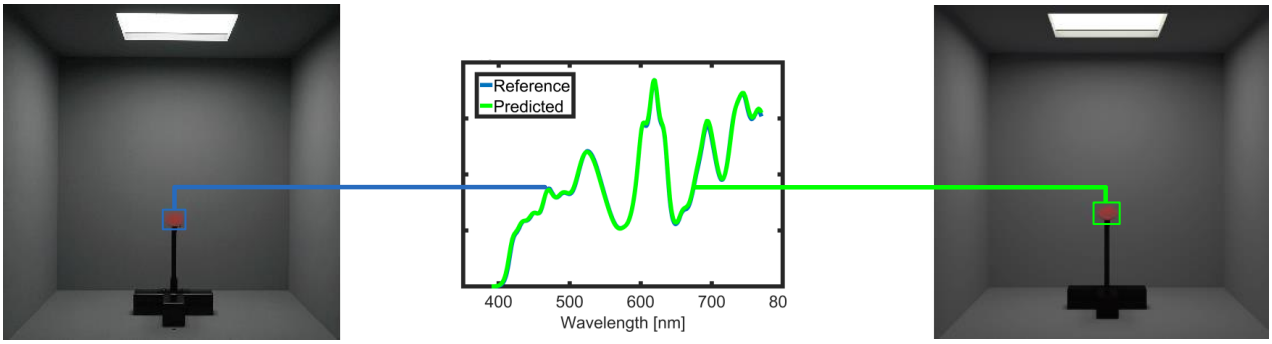

Figure 1: *Spectral radiance comparison of a real scene (left) with a predicted image based on our ground truth data (right).*

Abstract

Physically based rendering uses principles of physics to model the interaction of light with matter. Even though it is possible to achieve photorealistic renderings, it often fails to be predictive. There are two major issues: first, there is no analytic material model that considers all appearance critical characteristics; second, light is in many cases described by only 3 RGB-samples. This leads to the problem that there are different models for different material types and that wavelength dependent phenomena are only approximated. In order to be able to analyze the influence of both problems on the appearance of real world materials, an accurate comparison between rendering and reality is necessary. Therefore, in this work, we acquired a set of precisely and spectrally resolved ground truth data. It consists of the precise description of a new developed reference scene including isotropic BRDFs of 24 color patches, as well as the reference measurements of all patches under 13 different angles inside the reference scene. Our reference data covers rough materials with many different spectral distributions and various illumination situations, from direct light to indirect light dominated situations.

CCS Concepts
•***Computing methodologies*** → *Reflectance modeling;*

1. Introduction

Physically based rendering has become a popular technique for offline and real-time rendering, and has been adopted in many industrial applications. It is based on physical principles to model the interaction of light with matter, and its primary goal is usually realism [PH10, p.1]. The tremendous advancements of physically based rendering has lead to photorealistic images, but often fails to be predictive [WWMC09, p.11].

According to Ulbricht et al. [UWP06], the field of predictive rendering deals with methods that yield radiometrically correct renditions of nature. Thus, to be fully predictive all physical properties of light and its interaction with matter have to be taken into account. Although, they are well described in the physics literature,

current hardware is incapable of computing all properties in an acceptable time. Therefore, in computer graphics many approximations are still used.

There are two major issues related to these approximations: first, there is no analytic material model that considers all appearance critical characteristics; second, light is in many cases described by only 3 RGB-samples. This leads to the problem that there are different models for different material types and that wavelength dependent phenomena are only approximated.

To determine the influence of these issues on the appearance of real materials in a real lighting environment, the comparison of rendered images against reference data captured in a real world scene is mandatory. This comparison method is called experimental verification method. It requires the physical description of the scene as well as precisely acquired spectrally resolved reference data. In the following we summarized both and refer to it as ground truth data. An important observation is that the fidelity of the experimental method strongly depends on the accuracy of the ground truth data, hence its validation is necessary.

In this paper we introduce a new reference scene (Normbox) similar to the Cornell box, which we used to acquire a precisely and spectrally resolved set of ground truth data. It consists of the accurate description of the Normbox including isotropic BRDFs of 24 color patches, as well as the reference measurements of all patches under 13 different angles inside the Normbox. Our reference data covers rough materials with many different spectral distributions and various illumination situations, from direct light to indirect light dominated situations. All data and scripts can be downloaded from http://cg.web.th-koeln.de/brdf-gtd/.

We registered all measurement errors during the acquisition of the ground truth data. However, it is challenging or even impossible to estimate the influence of the measurement error on the appearance of the reference target within the box. Therefore we conducted the experimental verification method, where we used the reference renderer Mitsuba [Jak10] to predict the spectral radiance of the target patch. We determined the spectral error and the perceived color differences between the reference data and the predicted data to figure out the sources of error and the fidelity of the ground truth data.

Hereby is the spectral high resolved radiance a key advantage compared to previous approaches, where only colormetric quantities or single light measurements are used. It allows for a spectral localization of the errors and, consequently, a more founded allocation of their sources.

Briefly, the major contributions of this paper are:

- we developed a new reference scene similar to the Cornell Box;
- we precisely acquired a set of spectral ground truth data;
- we conducted, for the first time, a spectral radiance comparison of rendered images predicted by a reference renderer with the real-world;
- we determined the fidelity of our ground truth data.

2. Previous work

Greenberg et al. [GTS*97] proposed a framework for verifying photorealistic renderers. This framework is divided into three steps: light reflection model, global light transport simulation, and image display. According to the authors, each step must be verified individually by comparing it with the real world. The first two steps may be verified by comparing physical parameters, while the third step requires a perceptual comparison.

Our work deals with the first two steps. We use the experimental verification method to investigate the influence of approximated light reflection models and the global light transportation on the appearance of real world materials. In the following, we present some related works, which conducted the experimental verification method.

In 1984, the first experimental verification was reported by Coral et al. [CTGB84]. They showed that the radiosity algorithm behaves similar to light propagation in a real setup of a scene. Their work also introduced the well known Cornell Box, a cube consisting of fiber board panels painted with flat latex paints to minimize specular reflection. It has five walls and one open side in order to take pictures and illuminate the scene by a diffuse light. Three sides are painted white, while the left and right side are painted red and blue to investigate the effect of color bleeding. Unfortunately, they did not perform a comparison to quantify the physical or perceptual differences. This approach only allowed to state that the color bleeding effect is visible in both the rendered image and the reference scene. There was no quantification of the color differences.

Meyer et al. [MRC*86] followed up on the approach of the Cornell Box. They changed the Cornell Box slightly (e.g. they mounted a diffuse light source in the ceiling of the cube and placed two small boxes within the cube) and extended the approach with a radiometric comparison. They took measurements of irradiance at 25 locations on the plane of the open face of the cube, and compared them with rendered images using a radiosity method. The resulting differences are less than 4%. A major drawback of this approach is the limited spatial and spectral resolution of the irradiance measurements. It is difficult to identify the measured points in the rendered image, and due to the missing spectral resolution it is not possible to evaluate the color differences.

Greenberg et al. [GTS*97] presented a framework for realistic image synthesis. The framework extends the work of Meyer et al. by increasing the spatial and spectral resolution of the reference data. They acquired the reference data using a camera with a 1,280 x 1,024 CCD array. To achieve a spectral measurement, they used eight band pass filters. However, they did not provide results of the described experimental verification. Pattanaik et al. [PFTG97] implemented the presented framework, but only provide a reference image and predicted image, as well as the resulting difference image, without communicating more detailed results. However, the presented reference and predicted image already show clearly visible differences.

An accurate verification of the global light transport simulation was conducted by Schregle and Wienold [SW04]. They broke the verification process into simple component studies which can be checked individually, but also combined to build more complex

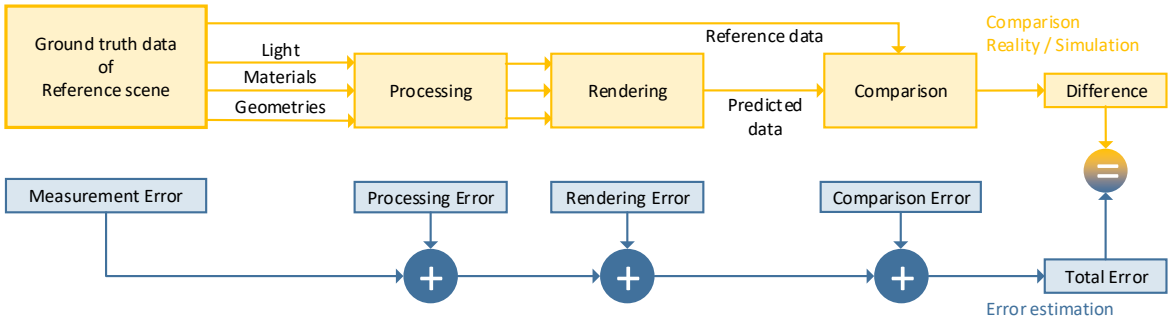

Figure 2: *Overview of our validation framework.*

compound case studies. A simple setup, consisting of a box and an area light source, was used as a reference scene. Depending on the study case, diffuse patches or a brushed aluminum plate were integrated. Flexible sensors in the ceiling measured the direct illuminance at different positions, which were used as reference data. The authors compared the reference data to a forward and backward raytracing solution: Photon Map [Jen01] and Radiance [LS04, War94]. The differences between the reference illuminance and the predicted illuminance is in all conducted case studies below 3% on average.

A more recent work by Bärz et al. [BHM10] followed a similar approach. Instead of a light source on the ceiling of the cube, they placed a monitor at the left side. On the opposite side, they mounted a Macbeth ColorChecker. As reference data, they measured all patches of the ColorChecker within the cube with a colorimeter. They conducted two (perceptual) comparisons; the comparison of the measured reference data with, first, the predicted data itself and second, with the predicted data displayed on a monitor model, identical to the used light source. In both comparisons they achieve for most patches color differences below the just noticeable difference (JND). However, there are also several patches with extraordinary high color differences.

The works described so far use a reference scene similar to the Cornell Box, which provides a simple and constant environment. One major advantage is that the ground truth data can be easily and accurately determined. Nonetheless, natural scenes consist of more complex geometries, materials and light conditions, which cannot be validated by such simple setups. Some works have proposed more realistic reference scenes like a car, a car dashboard, an office or an atrium [TTOO90, KP96, Mar99, DM01, MMK*06]. To verify the quality of the rendered images, they compare the rendered images with photos or several luminance measurements. Notwithstanding, there remains the major problem that it is impossible to determine the ground truth data of such complex scenes accurately. In particular, scenes where sunlight is present are not useful for an accurate comparison, since the illumination conditions are constantly changing.

3. Validation framework

Our work has two major aims. First, the acquisition of precisely and spectrally resolved ground truth data of 24 rough patches under 13 different angles. And second, the validation of the acquired ground truth data. For the acquisition of the ground truth data we created a reference scene which provides very controlled conditions. For the validation, all measurement errors of the ground truth data, as well as their influence on the appearance of the predicted image based on the ground truth data, have to be known.

The latter leads to a fundamental issue. It is challenging, or even impossible, to estimate the influence of each source of error on the appearance of the predicted image. Depending on the state of the reference scene the influence of each source of error may change, hence, the measurement errors alone are not very meaningful.

To overcome this issue we implemented the framework as shown in Figure 2. Our framework is divided into two paths: the comparison path (yellow) and the error path (blue). The comparison path describes the comparison between the real reference scene and its synthetically rendered image. In its first phase the ground truth data of the reference scene is acquired, which consists of the light and material distributions, scene geometries, and reference data. In the second phase this data is processed in order to be used as input data for the rendering phase, and, consequently, to generate a predicted image of the reference scene (predicted data). Next, in the comparison step, the predicted data is compared to the reference data to determine the deviations from reality. The resulting differences are composed of various error sources, which are considered in the error path.

The error path describes the sources of error introduced at every phase of the comparison path. Each source of error contributes to the total error, which corresponds to the resulting difference of the comparison path. As already stated, the individual contribution of each source of error to the total error is unknown.

To be able to quantify the contributions to the total error and, thus, the influence on the appearance of the predicted data, we implemented our described framework with Mitsuba [Jak10] as the reference renderer. We assumed that the rendering error is negligible for the simple reference scene we use. Thus, the resulting difference from the comparison path consists of the measurement error, processing error and comparison error. To further restrict the influence of the measurement error a spectral error analysis is conducted. It allows for a spectral localization of the errors and, consequently, a more substantiated attribution of their sources.

4. Implementation comparison path

4.1. Reference scene

As reference scene, we propose the Normbox, a modified Cornell box. As displayed in Figure 3, the Normbox consists of a wooden box, a sample holder, an integrating sphere combined with an opal glass panel for the light source, and a portable spectroradiometer mounted on an aluminum profile base to acquire the reference data. Each component is described in more details below.

4.1.1. Wooden box

It is a cube with sides made of 15 mm thick multiplex panels and a 19 mm panel for the ceiling. The inner dimensions of the cube are 600x600x600 mm. One side of the cube is open to allow measurements. On the ceiling there is an aperture of dimensions 220x200 mm, where the light of the integrating sphere enters the box. The inner sides of the box are painted with a matte, gray dispersion paint (RAL 7004) that provides a flat spectrum with $a^*, b^* < 1$ and $L \approx 65$. According to EN ISO 3668:2001 [ISO01], this is sufficient for visual comparison purposes. In the middle of the box, a black painted sample holder is placed. The sample is mounted at the flexible head of the holder, which enables its rotation around one axis.

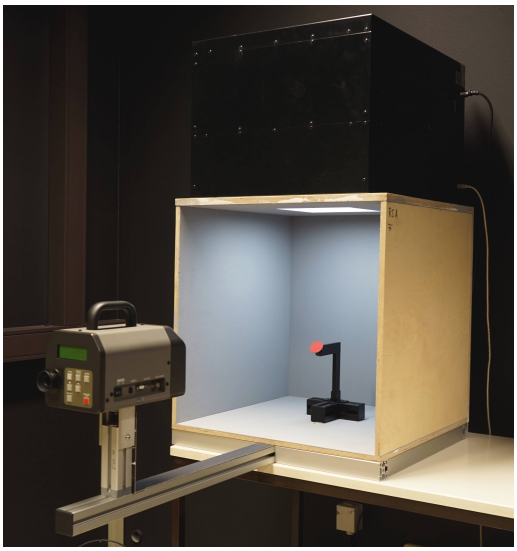

Figure 3: *The Normbox consists of a wooden box and the spectroradiometer CS-1000A by Konica Minolta. As light source, an integrating sphere combined with an opal panel is used.*

4.1.2. Area light source

Measuring the emission behavior of arbitrary light sources is a complex procedure requiring special equipment. A common way to simplify the simulation of the emission behavior is to approximate it by a homogeneous panel radiator with a Lambertian behavior. To minimize the difference between the assumed light source in the simulation and the used light source in the Normbox, we use an integrating sphere LE7-4x by Image Engineering combined with an opal panel. The homogeneity of the area light source deviates from the assumption by $\pm 3\%$ and the diffusivity by $\pm 2.6\%$.

4.1.3. Portable spectroradiometer

To acquire the reference data, as described in the validation framework, we use the portable high-precision spectroradiometer CS-1000A by Konica Minolta. This device accurately measures the spectral radiance in a wavelength range from 380 to 780 nm in 0.9 nm intervals. The device has an acceptance angle of 1 degree and is used with its 50 mm standard lens. Furthermore, it provides a radiometric accuracy of $\pm 2\%$ in a luminance range of 1 to 8000 cd/m^2, and a spectral accuracy of ± 0.3 nm. The CS-1000A is calibrated on the standard illuminant A.

4.2. Step 1: Acquisition of Ground Truth Data

The ground truth data of the Normbox consists of the reflection behavior of the gray painted walls and the 24 ColorChecker patches, the emission behavior of the area light source in the ceiling, the dimensions and positions of all involved objects, and finally, the reference data of all patches.

4.2.1. Reflection behavior

The X-Rite ColorChecker patches uniformly cover the sRGB color gamut including 6 neutral patches. The patches and the gray paint have a rough, isotropic, and uniform surface characteristic with a noticeable backscattering depending on the patch. As noted by Rump et al. [RZK11], the ColorChecker patches are far from being ideally diffuse. Hence, we decided to measure the BRDFs using a gonioreflectometer designed by the optics department of the TH Köln. The device allows the measurement of the in-plane BRDF with an error of approximately $\pm 0.7\%$. Each measured in-plane BRDF consists of 153 measurements with a spectral resolution of 1nm. The BRDF is measured for the incident angles $\theta_i = -80^\circ, -70^\circ, ..., 0^\circ$ and the reflection angles $\theta_o = -80^\circ, -70^\circ, ..., 80^\circ$. The incident angle varies only from $-80^\circ - 0^\circ$ since for isotropic materials the remaining range would be redundant, and angles larger than 80° cannot be measured by the device. To capture the backscattering the lightsource and spectrometer of the gonioreflectometer are tilted out of plane by 4° and -4°, respectively.

4.2.2. Emission behavior

As aforementioned, the light source of the Normbox is assumed to be ideally homogeneous and Lambertian. To minimize the difference between the assumed and real light source the spectral radiance is sampled at 25 points. Test renderings show that there are no noticeable differences between the representation of the light source with 25 spectral radiance values and the representation with one averaged value. Nevertheless, we used in this work the representation with 25 spectra.

4.2.3. Geometrical measurement

To describe the geometry of the Normbox the positions and dimensions of the wooden box, the area light source, the sample holder, and the spectroradiometer are acquired. They are measured manually by using a yardstick, which has a measurement error of approx-

imately ±1 mm. Furthermore, the angle of the sample in regards to the ground is measured using an Apple iPhone 5 with an accuracy of approximately ±1°.

4.2.4. Reference data

The reference data covers measurements of all 24 ColorChecker patches under 13 different sample angles inside our Normbox. The angles cover various illumination situations, where at 30° the direct light and at 90° the indirect light dominates. The reference data are measured by the spectroradiometer CS-1000A that points towards the center of the sample holder. All patches are placed one after the other on the sample holder, and their spectral radiance is measured for the 13 different sample angles. The sample angles vary from 30° to 90° in steps of 5°, where the sample is parallel to the light source for 0° and perpendicular for 90°. In regards to the measurement device, the angles vary from 60° to 0°, where the device is perpendicular to the sample at 0°. To guarantee the same sample angle for all patches, first the sample angle is adjusted and then the spectral radiance for all patches is measured.

4.3. Step 2: Processing

The acquired ground truth data must be processed in order to be properly imported by the Mitsuba renderer. There are three processing steps necessary. First, the measured geometries are converted into the widespread used OBJ file format. Second, the spectral radiance of the area light source is interpolated to the Mitsuba spectral resolution of 200 samples between 360nm and 830nm. And third, an analytical BRDF model is fitted to the BRDF measurements, as described in the next section.

4.3.1. BRDF model

We use the Cook-Torrance BRDF model [CT82] as it is a physically based reflection model for isotropic materials which has been successfully used for fitting measured BRDFs (see e.g. [NDM05, KSKK10]). This model describes the reflection of light from a material as a combination of a specular and a diffuse component, as defined in the following equation

$$f_r(\lambda,\omega_o,\omega_i) = \frac{k_{d,\lambda}}{\pi} + k_{s,\lambda}\frac{D(\omega_h)F(\omega_i,\omega_h)G(\omega_o,\omega_i,\omega_h)}{4\cos(\theta_o)\cos(\theta_i)}. \quad (1)$$

The first term represents the diffuse component while the second is the specular component which is based on a microfacet model, hence taking into account masking and shadowing effects, as well as Fresnel's law. The microfacet model is composed of a specular albedo $k_{s,\lambda}$, the distribution function $D(\omega_h)$, the Fresnel factor $F(\omega_i,\omega_h)$, and the geometric attenuation factor $G(\omega_o,\omega_i,\omega_h)$.

The geometric attenuation factor can be approximated as the separable product of two monodirectional shadowing-masking functions G_1:

$$G(\omega_o,\omega_i,\omega_h) \approx G_1(\omega_o,\omega_h)G_1(\omega_i,\omega_h). \quad (2)$$

For the distribution function we use the GGX distribution along with the Smith shadowing-masking term G_1 which was derived from the GGX distribution [WMLT07, Smi67]:

$$D_{GGX}(\omega_h) = \frac{\alpha_g^2}{\pi\cos^4\theta_h(\alpha_g^2+tan^2\theta_h)^2}, \quad (3)$$

$$G_1(\omega_o,\omega_h) = \frac{2}{1+\sqrt{1+\alpha_g^2\,tan^2\theta_o}}. \quad (4)$$

According to Heitz [Hei14], the Smith G_1 matches measured data better than the classical V-cavities. The parameter α defines the roughness of the material.

Under the assumption of unpolarized light the Fresnel factor for dielectrics can be computed as [CT82]:

$$F_{Cook-Torrance}(\omega_i,\omega_h) = \frac{1}{2}\frac{(g-c)^2}{(g+c)^2}\left\{1+\frac{[c(g+1)-1]^2}{[c(g-c)+1]^2}\right\}, \quad (5)$$

where $g = \sqrt{\eta-1+c^2}$, $c = \cos(\beta)$, β is the angle between ω_h and ω_i, and η is the ratio between the two refractive indexes of the materials that build the interface. In graphics, most objects are surrounded by air ($\eta_{air} \approx 1$), so that η is equal to the refractive index of the object material.

In the conducted fitting of the BRDF model to the measured BRDFs the roughness α_g, one global refractive index η and the wavelength dependent albedos k_s and k_d are fitted.

4.3.2. BRDF fitting

We fit the BRDF model to the measured in-plane BRDFs. The native model is slightly modified by subtracting a correction value c:

$$f_c(\lambda,\omega_o,\omega_i) = f_r(\lambda,\omega_o,\omega_i) - c. \quad (6)$$

For some materials the correction value minimizes the fitting error, while for others it assumes value zero and thus is not taken into account. The fitting process is divided into a non linear and a linear optimization process, where the latter is a subprocess of the non linear optimization process. The model parameters α_g, η, and the correction value c are determined in the nonlinear optimization. For the linear optimization, a linear regression is implemented to determine the wavelength dependent specular albedo k_s and the diffuse albedo k_d.

The fitted $BRDF_{fit}$ and the measured $BRDF_{ref}$ are the input values for the cost function. Our cost function is based on the one by Löw et al. [LKYU12], and is defined as:

$$g(\lambda,\omega_o,\omega_i;p) = ln(1+\cos^w\theta_i\, BRDF_{fit}(\lambda,\omega_i,\omega_o;p)), \quad (7)$$

$$\hat{g}(\lambda,\omega_o,\omega_i) = ln(1+\cos^w\theta_i\, BRDF_{ref}(\lambda,\omega_i,\omega_o)), \quad (8)$$

$$E(p) = \sum_k\sum_l (g^{kl}(p)-\hat{g}^{kl})^2, \quad (9)$$

where k represents the incident directions and l the reflection directions.

In Equation 7 and 8 the measured and fitted BRDFs are weighted by $\cos^w\theta_i$. Consequently, BRDF values for incident angles around 0° are weighted more heavily than for incident angles close to 90°. The optional exponent $w > 1$ further enhances this weighting effect. The function then weights the diffuse and specular component at

small incident angles stronger than the comparatively high Fresnel component at grazing angles. In this work the exponent w was set to values between 1 and 4, and in two cases to 6. These choices lead to better fits of the reflection at small incident angles, which in our work is of more importance than the Fresnel reflection at grazing angles. Furthermore, we observed that in some cases the exponent also improves the overall fitting error (cf. Figure 11).

4.4. Step 3: Rendering

In this work we use the Mitsuba renderer [Jak10] as reference renderer. We built it on a 64bit Windows platform and configured it as spectral renderer with 200 wavelength samples. The predicted images are generated on basis of the acquired and processed ground truth data as described in Sections 4.2 and 4.3.

To integrate the BRDF of Equation 1 we had to slightly extend the Mitsuba renderer by combining two material models to generate a new one. Specifically, we additively combined the integrated smooth diffuse material with a rough conductor material. The first describes an ideally diffuse material, while the second is defined by the microfacet model of the Cook-Torrance model. To integrate our fitted specular albedo we changed the "specular Reflectance" of the rough conductor material respectively. Mitsuba's native Fresnel calculation was replaced by the Cook-Torrance calculation defined in Equation 5. Finally, the predicted images of all ColorChecker patches under 13 sample angles were generated. The predicted image has a resolution of 10 x 10 pixels, where each pixel stores a whole radiance spectrum. All radiance spectra are averaged and, henceforth, called predicted data.

4.5. Step 4: Comparison

There are different methods to compare the predicted and reference data. In this work a physical and a perceptual comparison method are chosen. The physical method determines the radiance differences between the predicted and the reference data. The perceptual comparison provides information about the perceived color differences.

4.5.1. Physical comparison

The spectral radiance difference can be easily calculated by subtracting the predicted radiance spectrum from the reference spectrum. To allow the comparison between the results of different sample angles and different color patches, the spectral difference is normalized by the reference spectrum. This error metric is called normalized spectral error **(NSE)**. It has the advantage that it is spectrally resolved and thus allows for an accurate analysis of occurring errors. However, it is not appropriate to determine the total error, or to compare the total errors of different results. Therefore, the root mean square error **(RMSE)** between the predicted and reference spectrum is additionally calculated. Again, to allow for comparability, the RMSE is normalized by the averaged reference spectrum. Hereafter, this error metric is called normalized root mean square error **(NRMSE)**.

4.5.2. Perceptual comparison

A widespread metric for perceived color differences is the CIE DeltaE metric. For industrial color difference evaluation, the CIE recommends the latest version of the CIE color difference metric, the CIE DeltaE 2000. To calculate the *dE00* the predicted and reference spectrum are converted to the XYZ space. For the necessary normalization the reference spectrum of the ColorChecker white patch under a sample angle of 30° is used. This ensures that all Y values are between 0 and 1. For the the XYZ to Lab space transformation the standard illuminant D65 is use. The *dE00* are calculated as described in [CIE01], where the weighting factors K_L, K_C and K_H are set to the default value 1. The *dE00* values are classified by the rating scale given in the book Color Imaging [RKAJ08, p.461]. The rating scale is actually defined for the CIE DeltaE 1994 metric, but it can also be used for the *dE00* metric since they are very similar. In the book, a DeltaE of 1 is considered as the just noticeable color difference (JND). A value of 2 leads to discernible color differences for patches that are next to each other. Color differences larger than 5 are easily perceived in a side-by-side comparison.

5. Implementation error path

The error path is divided into the categories measurement error, processing error, rendering error and comparison error, as previously shown in Figure 2. Each category is further subdivided into various possible sources of errors as shown in Figure 4, as detailed next.

5.1. Measurement error

The Measurement error can be divided into four different errors introduced during the ground truth data acquisition. The first two regards the geometric measurements of the Normbox. As mentioned in Section 4.2.3 the measurement error of a yardstick is roughly ± 1 mm and the angle accuracy of an iPhone 5 is roughly $\pm 1^\circ$. The third measurement error is introduced by the measurement of the in-plane BRDF. This error can be divided into the measurement error of the used spectrometer Ocean Optics QE Pro, and the calibration error. The QE Pro has a signal-to-noise ratio of 1000:1 and a maximal non-linearity of 0.5%. The calibration is done with a relative calibration method using a Spectralon diffuse reflectance standard in combination with reference BRDF measurements by Durell et al. [DSM*15]. The BRDF acquired by the calibrated spectrometer deviates from the reference BRDF in averaged by 0.7%. Finally, the fourth error source is the spectroradiometer CS1000A, which is used to measure the reference data and the light source emission. According to the manufacturer, the accuracy is $\pm 2\%$.

5.2. Processing error

Processing the acquired ground truth data introduces again errors which should be considered. The conversion of the measured geometries into a digital format introduces a negligible error due to the geometrical simplicity of the Normbox, and the precision limitations of our measurement device. On the other hand, fitting the BRDF model and processing the light source emission lead to considerable higher errors. The fitting error varies in the range of 5%-

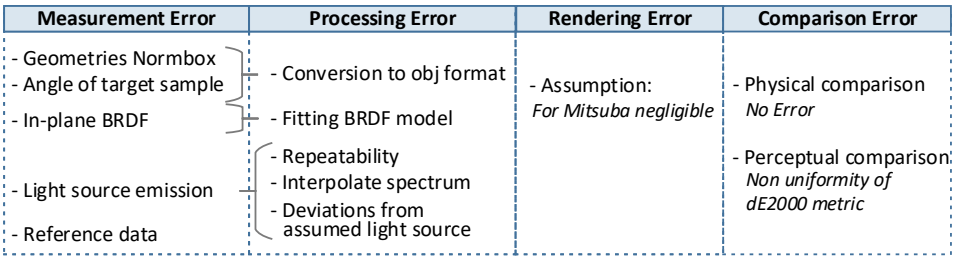

Figure 4: *Error sources that occurs during the implementation of the comparison path of the validation framework*

10% depending on the color patches, but patch 18 has a conspicuous error of 18%. The processing error of the light source emission is again divided into the errors repeatability, spectrum interpolation and the deviations from the assumed light source. The repeatability error is $< 1\%$ and the spectrum interpolation error is negligible, due to the densely sampled spectrum with 200 samples. The deviation from the assumptions are defined by $\pm 3\%$ for the homogeneity and by $\pm 2.6\%$ for the diffusivity. The intensity of the light source is kept constant during all measurements. Notwithstanding, we registered a maximal *dE00* of 0.28.

5.3. Rendering error & Comparison error

Finally we have to consider the rendering and comparison error. The rendering error is certainly unknown, but we make the assumption that it is negligible for Mitsuba. The physical comparison and the perceptual comparison errors have to be considered separately. The actual physical comparison does not introduce errors, since we compare the raw radiance spectra, and they are not processed in any way. Conversely, the perceptual comparison is error prone, because the raw data is processed and the error metric *dE00* is based on the perceptual non uniform CIELAB color space. The latter leads to distinct perceived color differences of the same DeltaE depending on the location in the CIE 1931 xy chromaticity space.

6. Validation of the ground truth data

In this section we validate the ground truth data by investigating the influence of the measurement error on the appearance of the predicted data. First, the physical and perceptual differences are presented to quantify the maximal influence of the measurement error. The resulting differences also include the processing and comparison error. Second, our results are compared to the one of Schregle and Wienold [SW04] and by Bärz et al. [BHM10]. And finally, a spectral error analysis is conducted to further restrict the influence of the measurement error.

6.1. Results of the physical and perceptual comparison

For the physical comparison we first computed the RMSE between the reference and predicted data as described in Section 4.5.1. In Figure 5 (a) the RMSE for all patches and sample angles is shown. It can be remarked that there is a significant RMSE variation between different patches and sample angles. In particular, patch 19 has an prominent high RMSE. When comparing only the gray patches (patches 19-23, excluding patch 24), it is possible to observe that there exists a correlation between the radiance of the patch and the RMSE. This assumption is again confirmed by looking at each patch individually, where there is a decreasing trend of the RMSE for increasing sample angles (decreasing radiance).

The assumed correlation between the radiance and RMSE is further examined in the scatter plots shown in Figures 5 (b) and (c). The first plot shows the data points of all patches for small sample angles 30° - 60°, and the second for the remaining large sample angles 65° - 90°. While the first plot indicates a strong positive linear relationship between the radiance and RMSE, with a correlation coefficient of 0.84, the second plot shows only a moderate relationship, with a correlation coefficient of 0.62. In brief, it can be stated that there is a positive linear correlation between the radiance and the RMSE, in particularly for small sample angles.

Due to this correlation, it is not meaningful to compare the RMSE between different measurements. Therefore, we normalized the RMSE and analyzed the resulting NRMSE plots as shown in Figure 6 (a). The figure shows that the NRMSE of all patches (excluding patch 24) are nearly on the same level for small sample angles. But at larger sample angles there is an increasing trend of the NRMSE, which even varies between different patches. To examine this phenomena in more details the data points for small and large sample angles are again plotted separately as shown in Figure 6 (b) and (c). In the first plot, nearly all data points have an NRMSE in the range of $[0.02, 0.05]$, with the exception of patch 24 and two data points of patch 18. It shows that for small sample angles there is no correlation between radiance and NRMSE, which is in accordance with Figure 6 (b). For large sample angles, shown in the second plot, there is an exponential increase of NRMSE for decreasing radiance. The bluish and greenish patches tend to have higher NRMSE than reddish and yellowish patches. Summarizing, it can be stated that the NRMSE for small sample angles is nearly constant, while for large angles there is an exponential decrease.

The results of the physical comparison do not provide any information about the perceived color difference between the reference and predicted data. Nevertheless, this difference is of paramount importance for color critical applications using rendered images. Hence, we present in Figure 7 the CIE DeltaE 2000 for all patches and sample angles. In contrast to Figures 5 and 6, the perceived color differences do not show any trends or patterns. For most patches the differences are below the discernible threshold for a next to each other comparison. Only patches 10, 13, 19 and 24 have

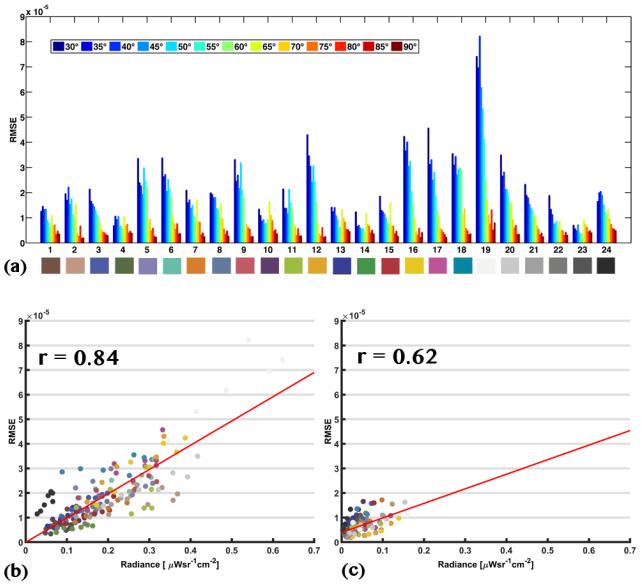

Figure 5: *(a) RMSE between the reference and predicted data of all patches and sample angles (bar color). Correlation between radiance and RMSE of (b) small sample angles (30° - 60°) and (c) large sample angles (65° - 90°).*

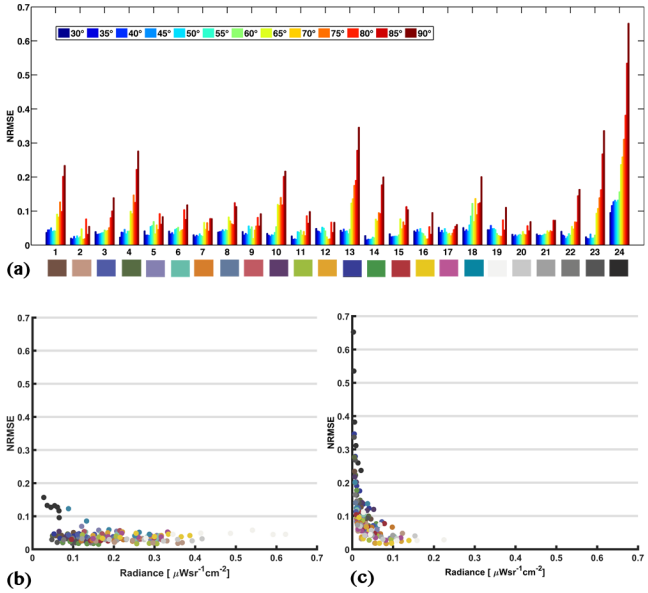

Figure 6: *(a) NRMSE between the reference and predicted data of all patches and sample angles (bar color). Correlation between radiance and NRMSE of small sample angles (b) and large sample angles (c)*

values for some angles slightly above the discernible color difference.

6.2. Comparison to previous work

We compare our results with those of the works by Schregle and Wienold [SW04] and by Bärz et al. [BHM10]. To the best of our knowledge, both works are the most recently published ones implementing a comparative validation of rendered images. Furthermore, their results show, up to this moment, the smallest differences between reference and predicted data.

Schregle and Wienold conducted a physical comparison between reference and predicted illuminance values, leading to a mean deviation of 2% in their case study for diffuse patch reflections. They measured the illuminance of the light reflected at a diffuse gray molleton patch and compared it to the prediction. In our results, most patches have deviations below 2%, with several among these achieving values considerably lower (cf. Figure 8). Considering patch 20, which has a similar reflection behavior as the gray molleton, it can be stated that our approach leads to clearly better results for this case.

Bärz et al. conducted a perceptual comparison. They computed the color difference between the reference and predicted data of all ColorChecker patches for one specific sample angle. We compared our results at the sample angle of 45° with their results, as shown in Figure 9. The figure demonstrates that the color differences of our approach are sometimes higher than the ones of Bärz et al., but we are mostly still below the JND. However, our results are more stable and do not contain extreme outliers as observed for patches 13 and 15 with Bärz et al. results.

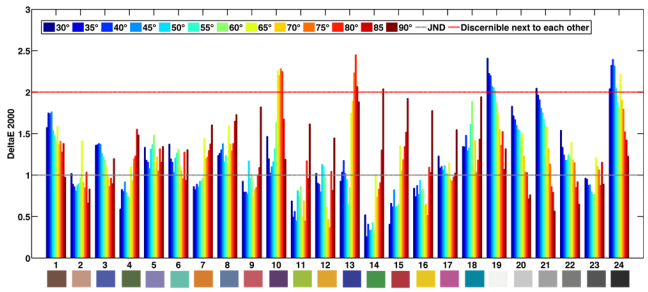

Figure 7: *Color difference (CIE DeltaE 2000) between reference and predicted data of all patches and sample angles.*

6.3. Spectral error analysis

6.3.1. Preliminary considerations

The observations in Section 6.1 lead to the assumption that the main error consists of two different components. The first error manifests itself as the correlation between the RMSE and the radiance of the patch, while the second error leads to the exponential increase of the NRMSE at large sample angles.

As stated in Section 4.2.4, the 13 sample angles describe a transition from direct light to a indirect light dominated situation. Moreover, we know that the light composition for sample angles up to 60° is roughly the same, while after 65° the proportion of indirect light increases exponentially.

The comparison of the NRMSE in Figure 6 and the increasing influence of the indirect light for increasing sample angles indicate

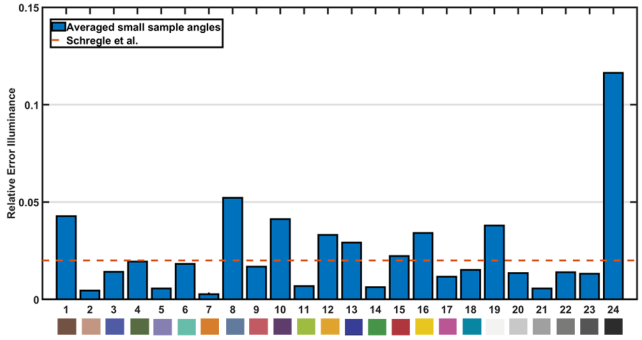

Figure 8: *Comparison of our approach with the one of Schregle and Wienold (red dashed line): Relative illuminance differences between reference and predicted data for diffuse reflection.*

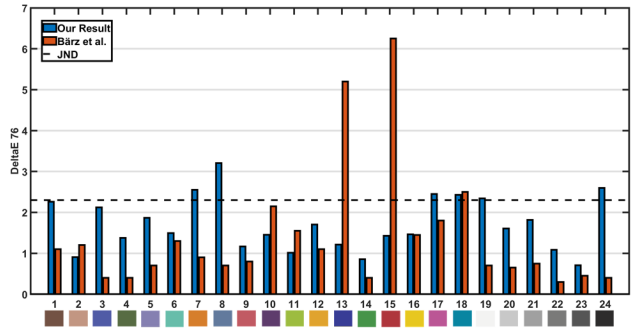

Figure 9: *Comparison of the color differences (DeltaE 76) between reference and predicted data with the approach of Bärz et al..*

a clear correlation. We conclude that the first error is related to the direct light (err_d), and the second error to indirect light (err_i).

6.3.2. Analysis of fitting error

As stated in Section 4.2.1 the 24 color patches and the gray wall are far from being diffuse and show a clear backscattering, as well as Fresnel reflection. Further investigations showed that the backscattering is dominated by blue frequencies, while the Fresnel reflection is clearly dominated by red frequencies.

Figure 10 shows the polar plots of the measured and fitted BRDF of patch 8 at a wavelength of 730 nm for incident directions $\theta_i = 0^\circ, 10^\circ, \ldots, 80^\circ$. The exponent w is set to 2. The figure illustrates that patch 8 has a diffuse reflection at small incident angles and a clear Fresnel reflection at grazing incident angles. Furthermore, a slight backscattering effect can be observed, which increases for small wavelengths. The good agreement of the measured and fitted curve demonstrates that the fitting in combination with the used BRDF model works well except for the backscattering which is not included in the model. At other wavelengths we observed similar good results.

However, the ColorChecker also contains some smoother patches, as demonstrated for patch 18 in Figure 11. The figure shows a comparison of three different fitted BRDFs, using the cost

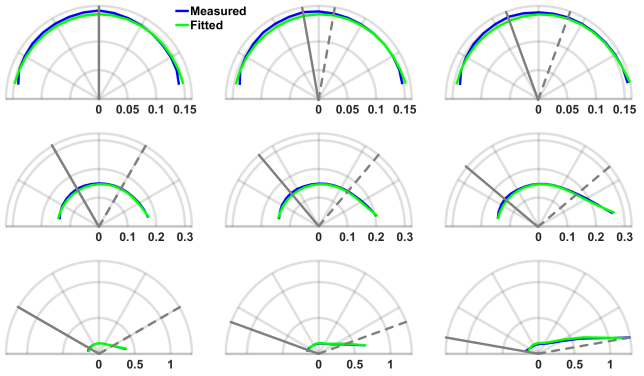

Figure 10: *Polar plots of the measured and fitted BRDF of patch 8 at wavelength 730 nm for different incident directions. The incident angle varies from $\theta_i = 0^\circ$ (top-left plot) to $\theta_i = 80^\circ$ (bottom-right plot).*

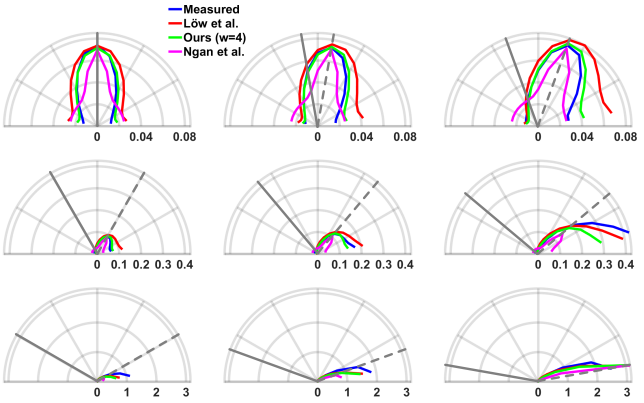

Figure 11: *Polar plots for patch 18 at a wavelength of 730 nm of the measured and three fitted BRDFs using the cost functions of Löw et al., Ngan et al., and ours with $w = 4$.*

functions of Löw et al. [LKYU12], Ngan et al. [NDM05], and ours. It is clear from the figure that our cost function leads to the best fitting especially at small incident angles, but it fails, as all other approaches, to precisely fit the Fresnel reflection at large incident angles.

The fitting of all ColorChecker patches with the Cook-Torrance BRDF model with GGX distribution in combination with our cost function lead to fitting errors in the range of $5\% - 10\%$ and 18% for patch 18. The analysis of the fitting errors showed that the fitting error for the diffuse and specular reflection at small incident angles is much smaller than for the backscattering and Fresnel reflection. Furthermore, it can be observed that the wavelength dependency of both phenomena is not fitted well. This results in a blue and red dominated fitting error for the backscattering and Fresnel reflection, respectively.

In short, it can be stated that the used BRDF model is not capable of fitting the observed wavelength dependent backscattering, leading to an overestimate of the diffuse term in the fitted BRDF of

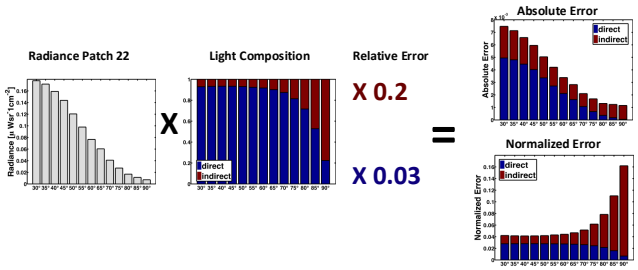

Figure 12: *Error model: The absolute error is simulated by multiplying the radiance with the light composition and the two errors, err_d (blue) and err_i (red). err_d is caused by a fitting error of the patch, err_i by a fitting error of the gray wall.*

nearly all patches. Besides, we observed that for some patches, especially for smoother materials, the BRDF model has problems to properly fit the Fresnel term, especially its wavelength dependency.

6.3.3. Error model

In our setup the used sample angles avoid that direct light is simultaneously incoming and measured from grazing angles. Thus, the measured direct light is not affected by the observed high fitting errors for the Fresnel term and backscattering. However, the indirect light is caused by reflections and interreflections at the gray wall. Hence, the indirect light is affected by all fitting errors of the gray wall including the high fitting errors at grazing angles. Therefore, we assume that err_d is much smaller as err_i.

Based on these assumptions we developed an error model as shown in Figure 12. In our error model the radiance of the measured reference data is weighted by the light composition to determine the proportion resulting from direct and indirect light. Both proportions are multiplied by the respective relative error (err_d, err_i) to compute the absolute error. To facilitate the comparison of the simulated error with the observed error in reality the absolute error is normalized by the reference radiance.

In the example in Figure 12 err_d is determined by the averaged fitting error of the patch 22 at the same light and measurement angles as occurring at the direct light. The err_i can not be determined easily because all fitting errors of patch 22 contribute. We chose err_i in a way that the resulting normalized error match the NRMSE and is within the fitting error of the gray wall.

The trend of the simulated errors are in accordance with the ones of the RMSE and NRMSE in Figure 5 and 6. Furthermore, the values of the simulated normalized errors and the NRMSE agree as well.

6.3.4. Validation of error model

In the error model the fitting error was assumed as the only source of error. However, as pointed out in the error path (Section 5) there are many different sources of error which can lead to a similar NRMSE. Thus, it is necessary to validate the fitting error as the dominant source of error. The validation is divided into two parts. First, the composition of the simulated error is compared with the normalized spectral error between the reference data and predicted data (NSE). Second, the assumed spectral distribution of err_d and err_i is compared with the one of the fitting error.

The value and composition of the simulated normalized error as shown in Figure 12 are equal for small sample angles, where err_d clearly dominates. Starting from sample angle 65°, the normalized error in general and the influence of err_i increase. At sample angle 90° the error consists almost entirely of err_i.

These observations are in accordance with the NSE, as depicted in Figure 13. In each subplot the NSE for one sample angle is plotted. We can observe that the spectral distribution of the NSE is practically constant for small sample angles. Starting at sample angle 65 the NSE is constantly increasing, especially at wavelengths larger than 650nm, thus modifying its spectral distribution. Figure 13 illustrates well this transition from one error to another, which is also simulated by our error model.

From the simulated error we know that at sample angle of 30° the err_d dominates, while at sample angle of 90° the err_i clearly dominates. Based on this knowledge, it can be assumed that the NSE at a sample angle of 30° and 90° in Figure 13 match the spectral distribution of the err_d and the err_i, respectively.

To verify the fitting error as the dominant source of error, the spectral representations of the err_d and the err_i are compared to the fitting errors at the respective sample angles. Figure 14 (a) depicts the NSE of patch 19, 23 and 24 at a sample angle 30°, and the respective fitting error. The figure demonstrates well that the trend of the error strongly agrees for patches 19 and 24, and moderately for patch 23. However, it can be observed an offset of approximately 2%.

Figure 14 (b) depicts the NSE of patch 21 at sample angle 90°, and the fitting error of the gray wall at a grazing angle of $\theta_i = -70°$ and $\theta_o = 80°$. The NSE of patch 21 shows a clear increasing trend for larger wavelengths, which can be observed for all patches (see e.g. Figure 13 for sample angle 90°). The only source of error that shows a similar increasing trend is the fitting error of the gray patch at grazing angles as shown in plot. This is clear indication that at larger sample angles the fitting error of the gray wall dominates.

6.3.5. Other sources of errors

In Figure 13 we can see that the NSE of the sample angles 65° and 80° slightly deviates from the trend of the surrounding sample angles. The NSE at sample 65° is slightly higher and at 80° slightly lower. The same NSE behavior can be observed for all patches. This kind of error can be traced back to a faulty sample angle adjustment, because the sample angle is adjusted once for all patches. A further source of error, which directly influences the predicted data is the repeatability error of the light source of 1%. We also have to take into account that the reference data measured with the spectroradiometer from Konica Minolta contains a measurement error of ±2%.

7. Conclusions and future work

In this work we developed a novel reference scene (Normbox) which allows for the precise acquisition of ground truth data for the

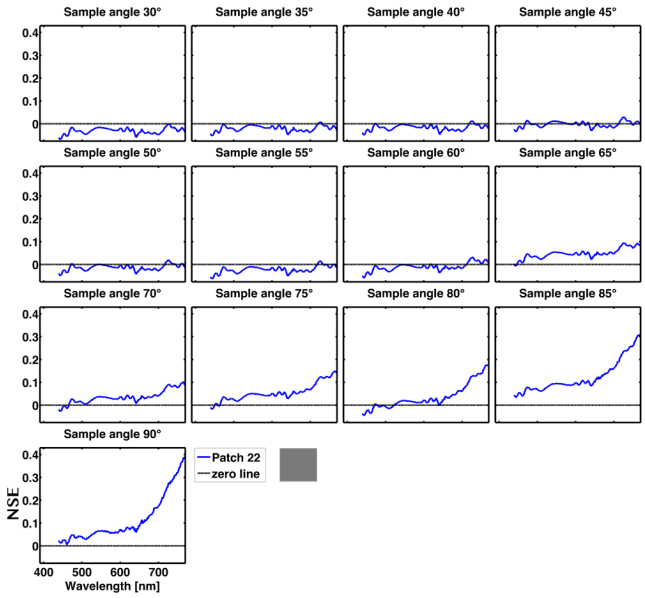

Figure 13: *NSE of patch 22: Each subplot depicts the NSE of one sample angle.*

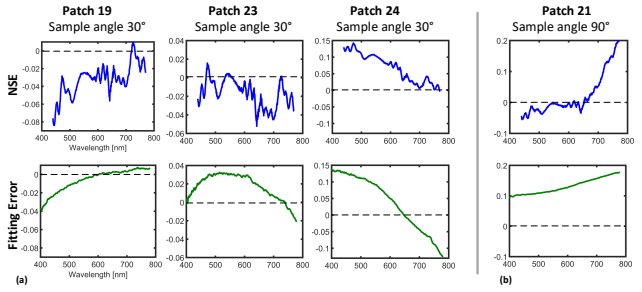

Figure 14: *Top row: NSE for (a) patches 19, 23 and 24 at sample angle* 30°*, (b) and for patch 21 at sample angle* 90°*. Bottom row: respective fitting error.*

experimental verification method. The key feature of our Normbox is the use of an integrating sphere which is controlled by an internal spectrometer. Combined with an opal panel it is a stable and nearly ideal homogeneous and diffuse area light source. We introduced the use of a spectroadiometer to measure the spectral radiance as reference data.

We used the Normbox to acquire a set of precisely and spectrally resolved ground truth data. It consists of the description of the reference scene including the BRDFs of the 24 ColorChecker patches as well as in total 312 reference spectra. The reference data covers rough materials with different spectral distributions and various illuminations situations, from direct light to indirect light dominated situations.

We implemented the experimental verification method, where we used the Cook-Torrance BRDF model with GGX distribution as fitting model, and the Mitsuba renderer as reference renderer to compute the predicted data. The comparison with the reference data leads to superior results when compared to previous work. The perceptual comparison further showed that the color difference of nearly all patches are below the discernible threshold in a next to each other comparison. To restrict the influence of the measurement error on the appearance of the predicted data, and thus to determine the fidelity of the ground truth data, a spectral error analysis is conducted.

The error analysis showed that the used Cook-Torrance BRDF model with GGX distribution is incapable to model all appearance critical scattering phenomena of rough surfaces, which lead, consequently, to deviations between reference and predicted data. The BRDF model fails to represent backscattering and has in several cases problems to properly fit both the large Fresnel reflection at grazing angles and the small reflection at lower incident angles at the same time. Furthermore, we observed that the backscattering as well as the Fresnel reflection are strongly wavelength dependent, where the backscattering is dominated by blue frequencies and the Fresnel reflection by red frequencies. Both wavelength dependencies are not fitted well by the used BRDF model.

Physically based rendering can achieve photorealism, but still fails to be fully predictive. Our acquired ground truth dataset is a starting point to further investigate the influence of the employed approximations on the appearance of real materials and, thus, a basis for future improved rendering techniques.

We plan to validate whether other published BRDF models (e.g. Holzschuch and Pacanowski [HP17], Dupuy et al. [DHI*15] or Bagher et al. [BSH12]) lead to better fitting results, especially of the wavelength dependency of the backscattering and the Fresnel reflection.

8. Acknowledgments

We would like to thank the anonymous reviewers for their very detailed and constructive reviews. We gratefully acknowledge Holger Weigand and Klaus Dollinger for invaluable discussions and feedback, and Aaron Finkenthei for precisely constructing all needed prototypes. This work was partially funded by the German Federal Ministry of Education and Research (BMBF) under the grant number 02K16C232 as part of the project “Retail 4.0”.

References

[BHM10] BÄRZ J., HENRICH N., MÜLLER S.: Validating photometric and colorimetric consistency of physically-based image synthesis. In *5th European Conference on Colour in Graphics, Imaging, and Vision and 12th International Symposium on Multispectral Colour Science, CGIV 2010/MCS'10, Joensuu, Finland, June 14-17, 2010* (2010), pp. 148–154. 3, 7, 8

[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. 11

[CIE01] CIE: *Improvement to industrial colour difference Evaluation*. Tech. rep., CIE 14x-2001, Commission internationale de l'éclairage, 2001. 6

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

[CTGB84] CORAL C. M., TORRANCE K. E., GRENNBERG D. P., BATTAILE B.: Modeling the interaction of light between diffuse surfaces. *Computer Graphics* (1984). 2

[DHI*15] DUPUY J., HEITZ E., IEHL J.-C., POULIN P., OSTROMOUKHOV V.: Extracting microfacet-based brdf parameters from arbitrary materials with power iterations. In *Proceedings of the 26th Eurographics Symposium on Rendering* (Aire-la-Ville, Switzerland, Switzerland, 2015), EGSR '15, Eurographics Association, pp. 21–30. 11

[DM01] DRAGO F., MYSZKOWSKI K.: Validation proposal for global illumination and rendering techniques. *Computers and Graphics 25*, 3 (2001), 511 – 518. 3

[DSM*15] DURELL C., SCHARPF D., MCKEE G., L'HEUREUX M., GEORGIEV G., OBEIN G., COOKSEY C.: Creation and validation of spectralon ptfe brdf targets and standards, 2015. 6

[GTS*97] GREENBERG D. P., TORRANCE K. E., SHIRLEY P., ARVO J., LAFORTUNE E., FERWERDA J. A., WALTER B., TRUMBORE B., PATTANAIK S., FOO S.-C.: A framework for realistic image synthesis. In *Proceedings of the 24th Annual Conference on Computer Graphics and Interactive Techniques* (New York, NY, USA, 1997), SIGGRAPH '97, ACM Press/Addison-Wesley Publishing Co., pp. 477–494. 2

[Hei14] HEITZ E.: Understanding the masking-shadowing function in microfacet-based brdfs. *Journal of Computer Graphics Techniques (JCGT) 3*, 2 (June 2014), 48–107. 5

[HP17] HOLZSCHUCH N., PACANOWSKI R.: A two-scale microfacet reflectance model combining reflection and diffraction. *ACM Trans. Graph. 36*, 4 (July 2017), 66:1–66:12. 11

[ISO01] *Paints and varnishes - Visual comparison of the colour of paints*. Standard, International Organization for Standardization, 2001. 4

[Jak10] JAKOB W.: Mitsuba renderer, 2010. http://www.mitsuba-renderer.org. 2, 3, 6

[Jen01] JENSEN H. W.: *Realistic Image Synthesis Using Photon Mapping*. A. K. Peters, Ltd., Natick, MA, USA, 2001. 3

[KP96] KARNER K. F., PRANTL M.: A concept for evaluating the accuracy of computer generated images. In *Proceedings of the Twelfth Spring Conference on Computer Graphics (SCCG' 96)* (August 1996). 3

[KSKK10] KURT M., SZIRMAY-KALOS L., KŘIVÁNEK J.: An anisotropic brdf model for fitting and monte carlo rendering. *SIGGRAPH Comput. Graph. 44*, 1 (Feb. 2010), 3:1–3:15. 5

[LKYU12] LÖW J., KRONANDER J., YNNERMAN A., UNGER J.: Brdf models for accurate and efficient rendering of glossy surfaces. *ACM Trans. Graph. 31*, 1 (Feb. 2012), 9:1–9:14. 5, 9

[LS04] LARSON G. W., SHAKESPEARE R.: *Rendering With Radiance: The Art And Science Of Lighting Visualization*. Booksurge Llc, 2004. 3

[Mar99] MARDALJEVIC J.: *Daylight Simulation: Validation, Sky Models and Daylight Coefficients*. PhD thesis, De Montfort University, Leicester, Institute of Energy and Sustainable Development, December 1999. 3

[MMK*06] MESETH J., MÜLLER G., KLEIN R., RÖDER F., ARNOLD M.: Verification of rendering quality from measured btfs. In *The 3rd Symposium on Applied Perception in Graphics and Visualization* (jul 2006). 3

[MRC*86] MEYER G. W., RUSHMEIER H. E., COHEN M. F., GREENBERG D. P., TORRANCE K. E.: An experimental evaluation of computer graphics imagery. *ACM Trans. Graph. 5*, 1 (Jan. 1986), 30–50. 2

[NDM05] NGAN A., DURAND F., MATUSIK W.: Experimental Analysis of BRDF Models. In *Eurographics Symposium on Rendering (2005)* (2005), Bala K., Dutre P., (Eds.), The Eurographics Association. 5, 9

[PFTG97] PATTANAIK. S. N., FERWERDA J. A., TORRANCE K. E., GREENBERG D. P.: Validation of global illumination solutions through ccd camera measurements. In *Proceedings of the Fifth Color Imaging Conference* (1997), no. 250–253, Society for Imaging Science and Technology. 2

[PH10] PHARR M., HUMPHREYS G.: *Physically Based Rendering, Second Edition: From Theory To Implementation*, 2nd ed. Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, 2010. 1

[RKAJ08] REINHARD E., KHAN E. A., AKYZ A. O., JOHNSON G. M.: *Color Imaging: Fundamentals and Applications*. A. K. Peters, Ltd., Natick, MA, USA, 2008. 6

[RZK11] RUMP M., ZINKE A., KLEIN R.: Practical spectral characterization of trichromatic cameras. *ACM Trans. Graph. 30*, 6 (Dec. 2011), 170:1–170:10. 4

[Smi67] SMITH B.: Geometrical shadowing of a random rough surface. *IEEE Transactions on Antennas and Propagation 15*, 5 (September 1967), 668–671. 5

[SW04] SCHREGLE R., WIENOLD J.: Physical Validation of Global Illumination Methods: Measurement and Error Analysis. *Computer Graphics Forum* (2004). 2, 7, 8

[TTOO90] TAKAGI A., TAKAOKA H., OSHIMA T., OGATA Y.: Accurate rendering technique based on colorimetric conception. In *Proceedings of the 17th Annual Conference on Computer Graphics and Interactive Techniques* (New York, NY, USA, 1990), SIGGRAPH '90, ACM, pp. 263–272. 3

[UWP06] ULBRICHT C., WILKIE A., PURGATHOFER W.: Verification of physically based rendering algorithms. *Computer Graphics Forum 25*, 2 (June 2006), 237–255. 1

[War94] WARD G. J.: The radiance lighting simulation and rendering system. In *Proceedings of the 21st Annual Conference on Computer Graphics and Interactive Techniques* (New York, NY, USA, 1994), SIGGRAPH '94, ACM, pp. 459–472. 3

[WMLT07] WALTER B., MARSCHNER S. R., LI H., TORRANCE K. E.: Microfacet Models for Refraction through Rough Surfaces. In *Rendering Techniques* (2007), Kautz J., Pattanaik S., (Eds.), The Eurographics Association. 5

[WWMC09] WILKIE A., WEIDLICH A., MAGNOR M., CHALMERS A.: Predictive rendering. In *ACM SIGGRAPH ASIA 2009 Courses* (New York, NY, USA, 2009), SIGGRAPH ASIA '09, ACM, pp. 12:1–12:428. 1