Gloss management for consistent reproduction of real and virtual objects

Bin Chen
binchen@mpi-in.mpg.de
Max-Planck-Institut für Informatik
Saarbrücken, Germany

Michal Piovarči
michal.piovarci@gmail.com
Institute of Science and Technology Austria
Klosterneuburg, Austria

Chao Wang
chaowang@mpi-inf.mpg.de
Max-Planck-Institut für Informatik
Saarbrücken, Germany

Hans-Peter Seidel
hpseidel@mpi-sb.mpg.de
Max-Planck-Institut für Informatik
Saarbrücken, Germany

Piotr Didyk
piotr.didyk@usi.ch
Universitá della Svizzera Italiana
Lugano, Switzerland

Karol Myszkowski
karol@mpi-inf.mpg.de
Max-Planck-Institut für Informatik
Saarbrücken, Germany

Ana Serrano
anase@unizar.es
Universidad de Zaragoza, I3A
Zaragoza, Spain

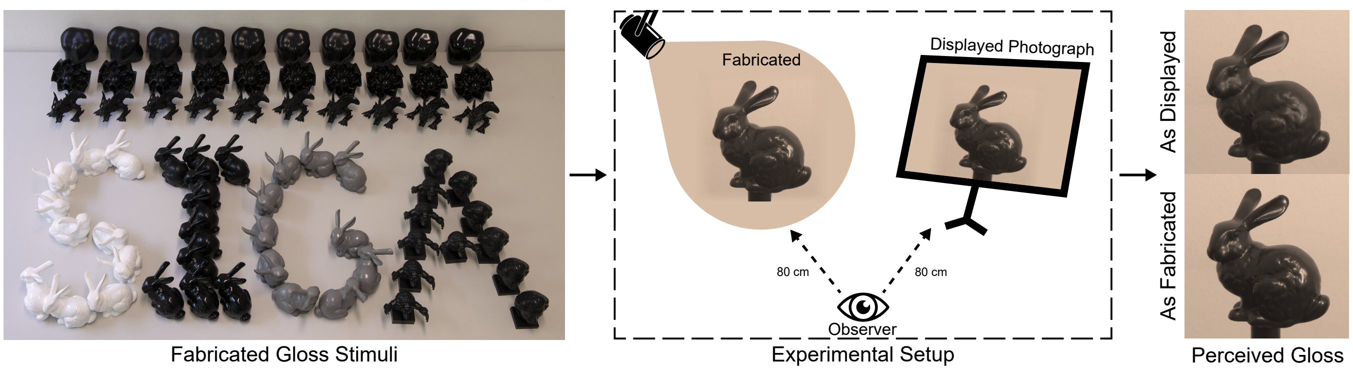

Figure 1: We study differences in gloss perception between physical and displayed stimuli. We start by fabricating an extensive set of geometries with different glossy finishes (left). We then use these models in a psychophysical experiment that compares them to their digital counterparts (center). The results demonstrate significant differences in perceived gloss between the two presentation methods (right). We address this gap with a new gloss management system that compensates for the observed differences.

ABSTRACT

A good match of material appearance between real-world objects and their digital on-screen representations is critical for many applications such as fabrication, design, and e-commerce. However, faithful appearance reproduction is challenging, especially for complex phenomena, such as gloss. In most cases, the view-dependent nature of gloss and the range of luminance values required for reproducing glossy materials exceeds the current capabilities of display devices. As a result, appearance reproduction poses significant problems even with accurately rendered images. This paper studies the gap between the gloss perceived from real-world objects and their digital counterparts. Based on our psychophysical experiments on a wide range of 3D printed samples and their corresponding photographs, we derive insights on the influence of geometry, illumination, and the display's brightness and measure the change in gloss appearance due to the display limitations. Our evaluation experiments demonstrate that using the prediction to correct material parameters in a rendering system improves the match of gloss appearance between real objects and their visualization on a display device.

CCS CONCEPTS

• **Computing methodologies** → *Appearance and texture representations*; *Perception.*

KEYWORDS

glossiness perception, glossiness reproduction

1 INTRODUCTION

Appearance capture, modeling, and reproduction are essential steps enabling computer graphics to simulate the appearance of the physical world and render visual representations of real objects.

Unfortunately, due to the complexity of the physical world and processes governing human perception, it is often challenging to create images that faithfully reproduce real-world appearance. The limitations come primarily from display devices. Limited spatial resolution, brightness, contrast, and color are key factors that lead to the inaccurate presentation of rendered images. Additionally, standard two-dimensional displays cannot reproduce many visual cues, such as binocular disparity, motion parallax, and accommodation, which contribute to the perceived appearance. Several existing works suggest that in the presence of such limitations, an accurate rendering of real-world objects is insufficient to reproduce their appearance on standard screens faithfully [Tanaka and Horiuchi 2015; Zhong et al. 2021]. The gap between the real-world and displayed images leads to challenges in applications where an accurate preview of material properties is essential. An example of such an application, and the focus of our work, is appearance fabrication; a problem that is important for many disciplines such as product design, cultural heritage preservation, or prosthetic fabrication.

One of the predominant appearance attributes is color, for which appearance models, management strategies, and fabrication techniques are well studied in the literature [Brunton et al. 2015; Fairchild 2013; Sumin et al. 2019]. Surface gloss is also a vital aspect of material appearance and recognition [Chadwick and Kentridge 2015; Fleming 2017; Ged et al. 2010]. Despite its importance, gloss fabrication and a faithful preview of designs and real objects on a standard 2D displays is a challenging task. One of the reasons for this is that gloss is highly view and illumination-dependent. The specular reflection is slightly different in each eye, creating binocular cues [Obein et al. 2004], and the dynamic range of specular reflection is significantly larger than achievable on common display devices [Phillips et al. 2009]. Additionally, surface gloss is modulated by the structure of the object's geometry. The discrepancies between the complexity of the gloss effect and the limitations of display devices become especially important when preparing designs for fabrication. While the importance of gloss in material perception is evident, obtaining an accurate match between the displayed and fabricated gloss is a little-studied but complex problem. Furthermore, studying the differences between physical samples and digital reproduction remains challenging as the need to obtain ground truth physical stimuli often leads to unfeasible sample complexity.

In this work, we investigate how accurately a display can reproduce the appearance of gloss. To keep the study design feasible, we rely on the properties of the human visual system. More specifically, we assume that observers can still make consistent gloss judgments despite the manifold nature of gloss cues [Chadwick and Kentridge 2015; Doerschner et al. 2010]. We have designed a psychophysical experiment to investigate how the perception of gloss changes between physical objects and their displayed counterparts. We measure the effect of the glossy finish, scene illumination, sample geometry, and display brightness. Then, we perform an in-depth statistical analysis of our collected data and show that displaying an object does indeed have a significant effect on the perceived gloss. Based on the data collected in the experiment, we propose an analytical correction that compensates for the gloss differences between real and displayed materials. We demonstrate that the correction can be directly integrated into gloss management workflows in two applications. First, we show that the correction is needed to design objects with desired gloss. Second, we demonstrate that our correction improves the visualization of digitized artifacts. All data is available at: https://glossmanager.mpi-inf.mpg.de.

2 RELATED WORK

In this section, we provide an overview of previous work on the perception of glossy materials and related perceptual experiments that directly compare physical material samples to their displayed depictions. We also discuss glossy appearance fabrication techniques.

Gloss perception. Glossiness is arguably one of the most important material appearance attributes [Anderson 2011; Fleming 2017; Marlow et al. 2012]. The gloss perception depends not only on the material reflectance but also on the surface geometry, and scene illumination [Serrano et al. 2021; Zhang et al. 2020a]. Interactions of these factors can change highlight coverage, brightness, and contrast, as well as modify the distinctiveness of the reflected images on objects' surface. Each of these phenomena has a direct effect on the perceived gloss [Marlow and Anderson 2013; Marlow et al. 2012]. High-contrast environment patterns, and strong directional lights increase the perceived gloss [Adams et al. 2018; Dror et al. 2004; Pont and te Pas 2006; Zhang et al. 2020b]. Likewise, by increasing surface curvature, the highlight intensity increases as well [Kim et al. 2012]. Increasing surface bumpiness typically increases the highlight coverage leading to stronger gloss perception. However, too fine bump structure has the opposite effect due to reduced distinctness of reflections and reduced contrast [Ho et al. 2008; Marlow et al. 2012]. Our work considers perceived gloss differences between the real world and displayed images. The inclusion of a display adds additional complexity to gloss perception. To handle this additional challenge, we follow the guidelines for geometry [Havran et al. 2016; Serrano et al. 2021; Vangorp et al. 2007] and illumination [Adams et al. 2018; Fleming et al. 2003; Leloup et al. 2010; Pont and te Pas 2006] selection in our psychophysical studies.

Reality vs. displayed images. High fidelity reproduction of real-world appearance in displayed images is an important goal in computer graphics. Perceptual experiments have been performed to formally evaluate perceived differences between directly seen real-world scenes and their displayed depictions. This way interesting insights have been gained on rendering algorithms [Drago and Myszkowski 2001; McNamara 2006; Meyer et al. 1986], tone mapping [Ashikhmin and Goyal 2006; Yoshida et al. 2005], reproduction of contrast [Yoshida et al. 2006], brightness [McNamara 2006], spatial scene details [Masaoka et al. 2013], binocular disparity [Vangorp et al. 2014], and lightfields [Zhong et al. 2021]. The most relevant for our work are efforts in material appearance evaluation on 2D displays. Predominantly independent glossiness rating sessions are performed for the real-world materials and either their photographs [Tanaka and Horiuchi 2015; van Assen et al. 2016], or rendered images [Filip et al. 2018]. In Sec. 4 we discuss these papers in more detail in the context of our results. The key difference here is that we present the real and displayed material sample pairs in a side-by-side manner. This is arguably the most direct method for measuring perceived gloss differences that not only removes scale

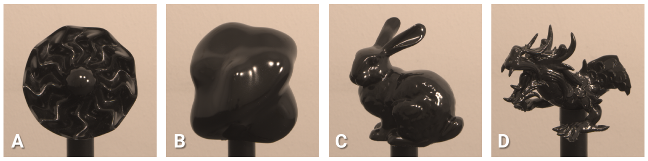

Figure 2: Selected geometries, from left to right: (A) *ghost*, (B) *blob*, (C) *bunny*, and (D) *dragon*. Photos are captured in our three spotlights illumination, followed by a chromatic adaptation to 4000 K.

biases [Pérez-Ortiz et al. 2020], but also assures a more consistent luminance adaptation. Moreover, we investigate diverse illumination and display conditions strongly connected to predictive rendering and 3D manufacturing, where we fabricate multiple shapes with a significant variation of curvatures and physical gloss levels. For the first time, we can systematically and quantitatively measure the perceived gloss differences across a wide range of dimensions.

Gloss fabrication. Accurate gloss reproduction is an active area of research. Achieving spatially varying gloss is possible through microstructure modification [Elkhuizen et al. 2019; Piovarči et al. 2017], careful material distribution [Baar et al. 2014; Matusik et al. 2009], or a combination of the two approaches [Lan et al. 2013; Malzbender et al. 2012]. Recently Piovarči et al. [2020] proposed novel printing hardware for gloss fabrication. Their hardware can reproduce spatially-varying gloss by dithering a discrete set of varnishes. Our fabrication technique takes inspiration from this work. To achieve a different glossy finish, we too rely on a set of varnishes. However, since we aim for a uniform gloss on the entire object, we pre-mix the varnishes before manually applying them.

3 EXPERIMENTAL DESIGN

The goal of this work is to study differences in perceived gloss between real objects and their corresponding virtual counterparts visualized on a display. To this end, we create a large set of physical stimuli where we vary relevant aspects of gloss perception: objects' geometry, objects' gloss, objects' lightness, environment illumination, and display brightness.

3.1 Stimuli

We briefly describe here our stimuli design. We encourage the reader to consult Sec. S1 in the supplemental for comprehensive implementation details and discussions on the selected stimuli.

3.1.1 Geometry and illumination. We select the *dragon* and the *ghost* models because they lead to strong differences in gloss judgements [Havran et al. 2016; Serrano et al. 2021], the *blob* model that shows a good performance in gloss discrimination experiments [Vangorp et al. 2007], and the *bunny* model that features more high-frequency details. As shown in Fig. 2 with this selection we achieve variation of surface curvatures, which is instrumental in glossiness studies [Faul 2019; Ho et al. 2008; Kim et al. 2012; Marlow et al. 2012].

We aim to investigate gloss perception during natural viewing conditions where the observer is adapted to the ambient illumination. However, this is difficult to achieve with traditional experiment setups where participants are limited to a confined photo box. We create room-scale controlled illumination and focus on three classes of viewing conditions: *diffuse illumination*, *one spotlight*, and more *complex illumination* (three or five spotlights). In Fig. 3 we show panoramas corresponding to each of our illuminations.

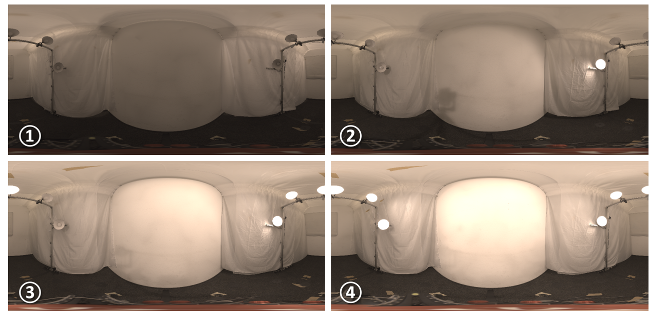

Figure 3: Selected illuminations captured in panoramas for visualization. ① *diffuse* illumination. ② *one spotlight.* ③ *three spotlights.* ④ *five spotlights.*

3.1.2 Gloss Fabrication. To physically realize the selected geometries we fabricate them on a stereolithography printer. The geometries are hand-polished and evenly coated with matte black (Pantone Hexachrome Black U) lacquer. We opt for uniform achromatic color based on recent studies that show no significant effect of sample color on gloss perception [Tanaka and Horiuchi 2015; van Assen et al. 2016]. We investigate black surfaces that appear more glossy than their brighter analogues [Hunter and Harold 1987; Motoyoshi and Matoba 2012; Pellacini et al. 2000; Wills et al. 2009]. In a more limited scope, we investigate the lightness and gloss interactions with gray and white samples (Sec. 4.1). In order to produce gloss variations, we rely on off-the-shelf varnishes. We mix Schmincke 610 glossy varnish and Schmincke 611 matte varnish in different proportions. For clarity, we refer to the mixtures by the percentage of glossy varnish. We include in our experiment *ten* samples with the percentage of glossy varnish ranging from 10% to 100% in 10% increments. An example of manufactured objects can be seen in Fig. 4, and the complete set can be found in Sec. S2 in the supplemental.

3.1.3 Display settings. To display the virtual samples we use an EIZO CS2420 24.1 inch monitor with a matte ISP LCD panel, 100% sRGB coverage, 1000:1 contrast ratio, and 1920×1200 resolution which is accurately calibrated. These specifications match well with most common displays used for professional appearance reproduction workflows. As an additional factor in our experiments we vary the display brightness. We test for three representative luminance levels: 220 cd/m^2 as an example of typical office monitor in daylight conditions, 110 cd/m^2 the suggested luminance level in nighttime conditions, and 55 cd/m^2 that approximates the cinema luminance level [SMPTE-196M 2003]. We term these three levels *bright*, *medium* and *dark* respectively. To create a virtual depiction of our fabricated samples we capture HDR photographs [Hanji et al. 2020], color adapted to match our 4000 K light sources. To match the luminance range of the HDR images to our scene we use the

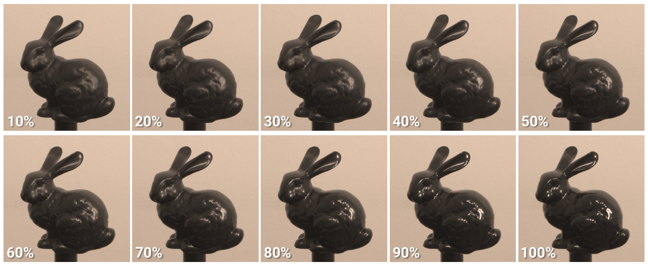

Figure 4: Varnishing results for one of our geometries. The percentage of glossy varnish ranges from 10% to 100%. Photos are captured in our three spotlights illumination, followed by a chromatic adaptation to 4000 K.

gamma-offset-gain display model [Berns 1996]. We measure the peak brightness, black level, and surface reflection of the display to restore the luminance of the real world as much as possible (see Fig. 5 for an example).

3.2 Main Experiment: Measuring the gloss mismatch

We aim to investigate which factors influence the gloss mismatch between real objects and their virtual counterparts. We have a total of 10 gloss samples × 4 geometries × 4 illuminations × 3 display brightness, which yields a total of 480 conditions. The additional experiment for analyzing objects' lightness is described in Sec. 4.1.

Validating the samples. We carry out a preliminary study and statistical analysis to validate our selected varnish mixtures (please refer to Sec. S3 in the supplemental). In this study, we confirm that (i) the perceived glossiness of our samples increases monotonically and is uniformly distributed based on perceptual interval scaling analysis following Thurstone's law [Thurstone 1927], and that (ii) users are able to distinguish all the samples both in the real setup and in the display. Our data also shows that the participants' ordering is not always perfect, suggesting that adjacent samples are close to the discrimination threshold while still being distinguishable.

Participants. A total of 42 participants (19 to 34 years old, 20 females, 22 males, no intersex/others) completed the experiment, two of them were the authors and the rest were graduate and undergraduate students from a local campus. All participants had normal or corrected-to-normal vision and were naïve to the goal of the experiment except for the authors. The experiment was approved by the department ethic board and the participants provided written consent and were economically compensated.

Procedure. The participants' task during each trial was to match the image on the display (one of the ten photos corresponding to our ten real samples) to that of the real object in terms of perceived gloss. On the left of the participant, we positioned the display, directly facing the observer and positioned to avoid direct reflection from the light sources. In front of the participant we showed the physical sample (one at a time), which was attached to a holder that maintained a consistent position and orientation of each sample with respect to the participant. The glossiness starting point in the display was randomly reset after each trial to avoid cross-impact between different trials. Each participant saw a total of 160 trials out of the 480 conditions (10 gloss samples × 4 geometries × 4 illuminations under a random display brightness) distributed in 4 sessions, one for each illumination. Please, refer to Sec. S4 in the supplemental for extended details about the procedure.

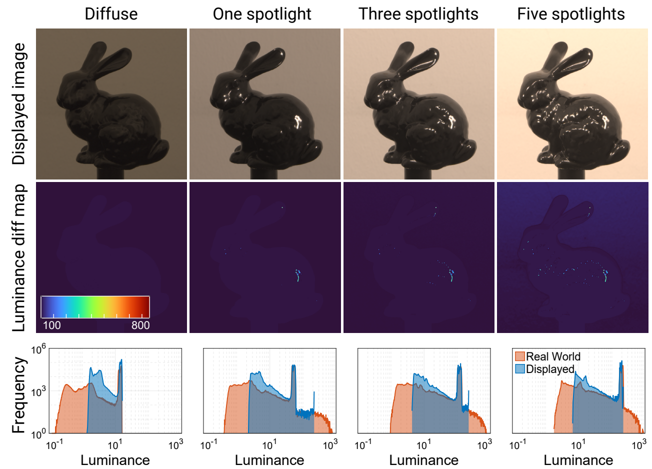

Figure 5: Luminance coverage of our four illuminations. First row: images of the bunny with 100% glossiness level in 4 illuminations. Second row: map of luminance difference between displayed image and real world. Third row: luminance distribution of displayed image and the real world. An exact peak luminance match is possible only for diffuse illumination, while for the remaining illuminations the main discrepancy lies over-saturated highlights. Additionally, the display is not able to accurately reproduce the dark regions for any of the illuminations.

4 DATA ANALYSIS

In this section we perform a thorough analysis of the data collected in the main experiment to study the influence of geometry, illumination and display brightness in perceived glossiness differences between real objects and their displayed counterparts. In our alternative forced choice experiment our dependent variables (selected matching image in the display) are categorical variables. Therefore, we use multinomial logistic regression [Böhning 1992] for our analysis. We include the *illumination* (diffuse, one spotlight, three spotlights, five spotlights), *geometry* (dragon, bunny, ghost, blob) and *display brightness* (dark, medium, bright) as factors, as well as their interactions. For each factor one level is selected as baseline for computing this regression. The model coefficients (and their corresponding p-values computed using Wald tests) describe how different levels of each factor influence the dependent variable with respect to the baseline levels. We select as baseline the *bunny* under *five spotlights* with a *dark* display brightness. In all our tests we fix the significance level to $\alpha = 0.01$.

For computing the goodness of fit of our multinomial logistic model we use Nagelkerke pseudo R-square [Nagelkerke et al. 1991], which ranges from 0 (bad fit) to 1 (perfect fit) and can be seen as a

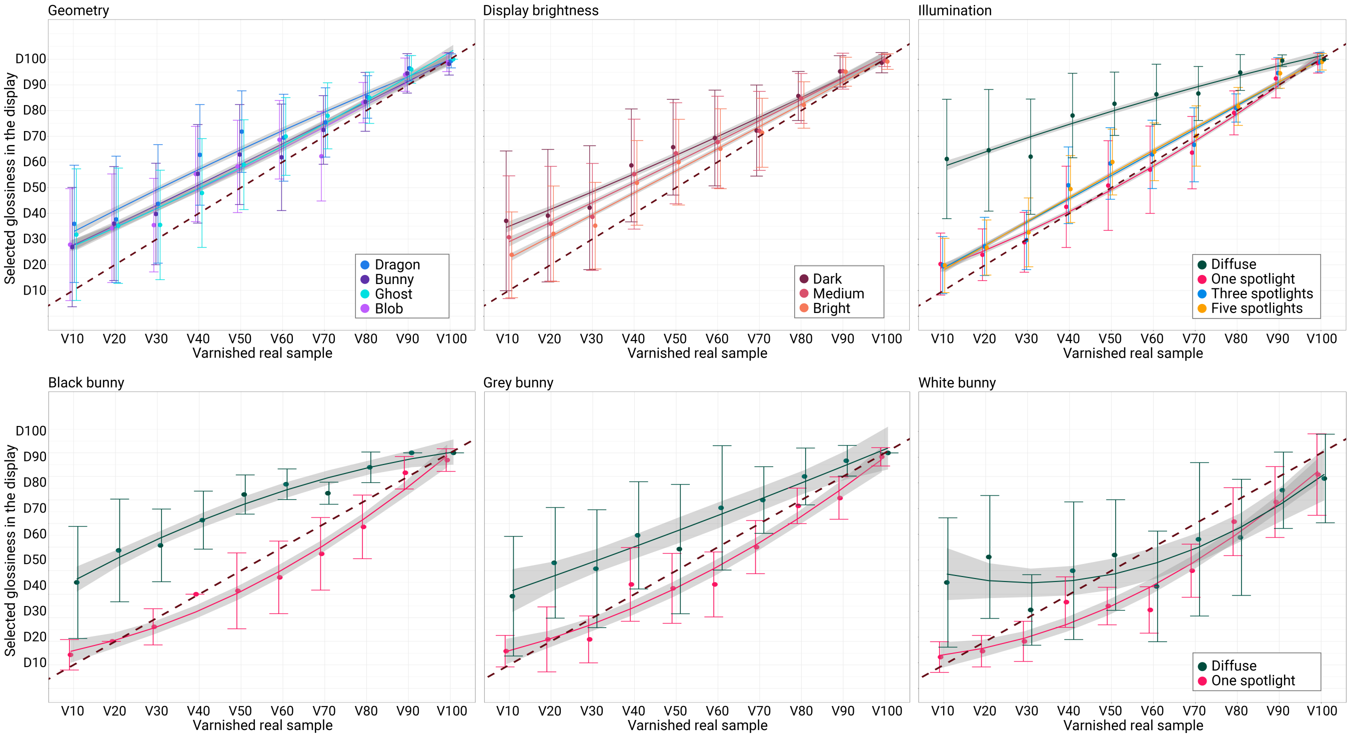

Figure 6: Main results of our statistical analysis. For each plot we aggregate all the data grouped by the levels of the corresponding factor. The dashed black line illustrates the ground-truth selection, i.e., when the image selected on the display matches that of the corresponding varnished real sample. Filled points represent the mean for each level and errorbars represent the standard deviation. We fit a second-order polynomial function to our points represented by the associated colored line and accompanying grey shadow (95% confidence interval). First row: Effect of the geometry, display brightness and illumination. Second row: Effect of the object's surface lightness for the particular case of the *bunny* geometry visualized in the *bright* display condition as described in our additional experiment in Sec. 4.1.

reliable measure of the percentage of the variation in the dependent variable explained by the model. The complete model achieves $R^2 = 0.87$ with a significant effect for all tested factors as well as their interactions ($p < 0.001$). A model without interactions achieves $R^2 = 0.85$, and the data reveals that, although these interactions are significant, they do not have a strong effect. Therefore, in this section we focus on discussing our main factors. We further confirm the significance of these factors by computing likelihood ratio tests between the final model and different models removing each of the factors one at a time. We summarize the main results in Fig. 6 and discuss the observed effects for each of the factors in this section. Note that all our observations are based on statistically significant effects. Please refer to Sec. S5 in the supplemental for additional details and figures of our collected data.

Effect of geometry. For the tested geometries, only the *dragon* differs consistently from the baseline for different glossiness levels, while the *blob* and the *ghost* are statistically indistinguishable from the baseline *bunny* for almost all glossiness levels. This can be seen in Fig. 6 (top row-left). For the *dragon*, participants are more likely to select in the display an image with higher glossiness in order to match each particular glossiness level of the physical object. This insight is in accordance to the findings of Serrano et al. [2021]. In their experiments, carried out using traditional displays, they found that the *dragon* geometry was rated significantly less glossy. The cause might be that complex, strongly tessellated geometries make the discrimination of subtle glossy effects more difficult than smooth surfaces [Vangorp et al. 2007]. Likewise, stronger gloss corrections might be required to reduce the perceptual gap to the real samples. Nevertheless, although this is a significant and interesting effect, its strength is limited. Therefore, for our applications in Sec. 5, we choose to average the collected data across geometries and correct equally for all geometries.

Effect of display brightness. All three brightness levels present statistically significant differences for most glossiness values. We can observe in Fig. 6 (top row-middle) that for lower display brightness participants more underestimate the perceived glossiness in the display and select higher glossiness values to match their real counterparts, while for higher display brightness participants select glossiness levels closer to their matching real counterpart. This might indicate that the larger the gap between the luminance levels reproduced by the display and the luminance in the physical world, the larger gloss correction is required. For our applications in Sec. 5, we apply our correction taking into account this effect as a function of display brightness.

Effect of illumination. We find three main clusters of effects for the tested illuminations. The more complex illuminations (*five spotlights* and *three spotlights*) are statistically indistinguishable for all glossiness levels, while the *diffuse* and *one spotlight* are significantly different from all others. For the *diffuse* illumination we can see in Fig. 6 (top row-right) that participants are more likely to select in the display an image with higher glossiness to match that of the printed object (i.e., participants underestimate the glossiness perceived in the display). This is in accordance with the work of Tanaka and Horiuchi [2015]. In this work, the authors consider a D65 diffuse illumination similar to our *diffuse* setup and report that the glossiness ratings for flat plastic exemplars observed in the real-world are on average significantly higher than when their photographs are displayed. A similar effect is also observed by Filip et al. [2018] for a variety of diverse flat material exemplars. Interestingly, our analysis reveals that this previously observed effect heavily depends on the type of illumination and is not observed for other illuminations.

For the complex illuminations (*five spotlights* and *three spotlights*, participants also tend to underestimate the glossiness perceived in the display. However, this effect is less pronounced than for the *diffuse* illumination, and participants' selections are closer to their real counterpart (dashed black line). In this condition, the key difference with respect to the *diffuse* illumination is the presence of pronounced and more complex highlight patterns which, together with the increasing surface glossiness, appear to facilitate the task of matching with the real world, reducing the required gloss correction. This presence of highlights, even in their clamped version due to the dynamic range limitations of the display (Fig. 5), seems to be a more important gloss cue for such matching than the faithful reconstruction of higher luminance levels in the scene, as achieved for the *diffuse* illumination. Nevertheless, such highlight clamping still seems to impact the matching task, since in the *one spotlight* illumination, where clamping is more limited (Fig. 5), less gloss correction is required.

In particular, for the *one spotlight* illumination and the mid-glossy samples we can find the closest match between the real objects and the displayed ones. Interestingly, in these conditions participants are more likely to select in the display an image with lower glossiness to match that of the printed object. This effect has been also observed by van Assen et al. [2016], where they consider a painted sphere that is illuminated by a single spotlight with varying shapes. Their findings show that perceived glossiness decreases for physical objects and, similarly to our results, this trend is reduced with increasing object glossiness. For our applications in Sec. 5 we consider these three clusters according to our insights: diffuse illumination, spotlight illumination and complex illumination.

4.1 Additional experiment: Effect of surface lightness

We perform an additional experiment to analyze the effect of the object's surface lightness in our task. For this, we select a subset of stimuli based on our previous observations. We select the *bunny* under *diffuse* and *one spotlight* illumination visualized on the bright display condition. In addition to our already collected data for the *black* bunny, we repeat our experiment for a *grey* (Pantone Cool Gray 8U) and a *white* (Pantone White U) geometry. Please, refer to Sec. S6 in the supplemental for more details. Results for this experiment are shown in Fig. 6 (bottom row). On one hand, we can observe that the trends for the *grey* surface are very similar to those of the *black* surface. Existing works show the human ability to discount specular highlights from lightness under moderate lightness changes [Olkkonen and Brainard 2010; Todd et al. 2004; Toscani et al. 2017], which might support the small differences in required gloss corrections between the *black* and *grey* samples for the *one spotlight* condition. On the other hand, the *white bunny* requires more correction. The previous observation might not hold for such strong lightness changes, in this case, our white surface overrides most specular highlights. This is aligned with other studies that still observe some gloss and lightness interactions, so that bright surfaces are perceived less glossy [Chadwick and Kentridge 2015; Hunter and Harold 1987; Motoyoshi et al. 2007; Pellacini et al. 2000; Wills et al. 2009]. For the *diffuse* illumination, higher lightness tends to dominate on the surface appearance and further dilutes the gloss cues by reducing the contrast of the reflections. In such conditions, gloss cues are hard to perceive both in the real world and displayed scenes, and this may contribute to the overall reduced magnitude of gloss correction. We believe this aspect may require further investigation due to the intricate changes of the observed correction, especially for the *white bunny* model.

5 APPLICATIONS

Our perceptual experiment shows differences in perceived gloss between physical and virtual stimuli. We propose a gloss management system to achieve control of gloss representation across the two modalities. During our perceptual studies, we ask the participants to match the gloss of real-world samples with their displayed counterparts. Such a design directly leads to correcting differences in perceived gloss via a look-up table (LUT). We perform several generalizations of the LUT based on insights from our statistical analysis. First, we cluster the corrections across all geometries. Next, for each illumination, we fit an analytical correction based on display brightness and sample varnish. Finally, for each varnish mixture, we measure the reflectance and fit a Cook-Torrance model with GGX distribution [Trowbridge and Reitz 1975]. The roughness α of the fitted model is then mapped to the varnish mixture via an invertible function $\alpha = \mathcal{M}(V)$, where V is the varnish mixture. The final correction model is then:

$$\alpha' = w_0(\mathbf{B}) + w_1(\mathbf{B})\alpha + w_3(\mathbf{B})\alpha^2, \tag{1}$$

where α' is the corrected roughness that can be converted into a varnish mixture through $\mathcal{M}^{-1}$, and $w_i(\mathbf{B})$ are the per-illumination fitting weights expressed as a function of display brightness $\mathbf{B}$. For more details about the measurements and fitting please refer to Sec. S7 in the supplemental.

5.1 Digital Product Design

Digital product design relies on careful calibration to match the appearance of a virtual scene with the manufactured object. A mismatch in desired gloss leads to expensive design iterations that involve the fabrication process. In this application we demonstrate that our correction improves the gloss match which could lead to faster design iterations.

Procedure. We consider two novel geometries not used during our main study, a bust of *Einstein* and *armadillo*. We select two levels of desired gloss equivalent to 20% and 60% varnish mixture and two target illuminations, *diffuse* and *one spotlight*. We then physically realize two sets of objects. One set is manufactured directly with the selected varnish mixture. For the second set we update the varnish mixture using our correction from Equation 1. An example of *Einstein* in *one spotlight* is shown in Fig. 7 (top), where 60% glossiness needs to be increased to 65.6% to compensate the discrepancy. In total, 18 volunteers (21 to 32 years old, 6 females, 12 males, and no intersex/others) with normal or corrected to normal vision participated in this experiment. Participants are presented with a two-alternative forced choice experiment (2AFC). We use the digital design as a reference shown on the display. Their task is to select the closest match in terms of gloss between a corrected and non-corrected fabricated sample.

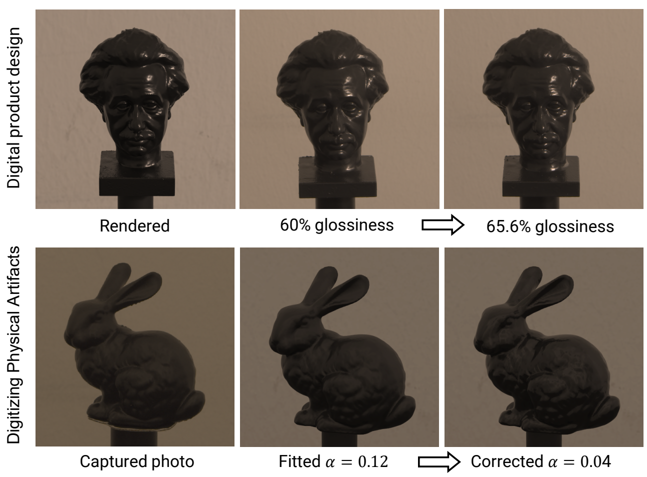

Figure 7: Examples of our two applications. Fist row: digital product design. The 60% glossiness under *one spotlight* illumination should be corrected to 65.6%. Second row: digitizing physical artifacts. The roughness parameter α in fitted Cook-Torrance model under *diffuse* illumination should be corrected from 0.12 to 0.04. This figure is only illustrative, to compare the perceived glossiness one would need to compare directly the displayed and real objects.

Results. In Fig. 8 (left) we show the results of the study for each illumination aggregated by geometry and glossiness levels (please refer to Sec. S8 in the supplemental for additional details). For both illuminations users consistently chose our corrected manufactured version as closer to the displayed image in terms of glossiness.

5.2 Digitizing Physical Artifacts

Faithful display of physical artifacts has a range of applications in fabrication from enabling quality inspection of fabricated designs, to cultural heritage preservation. However, the mismatch between captured and displayed gloss can significantly degrade the perceived realism. In this application, we show that our correction can be used to faithfully reproduce the gloss of digitized objects.

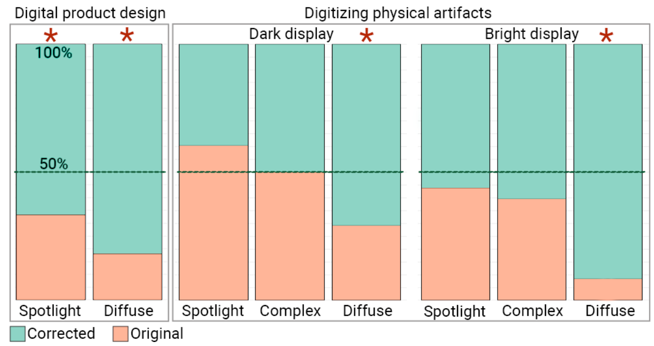

Figure 8: Results of the two-alternative forced choice study for evaluating our two applications: digital product design (left) and digitizing physical artifacts (right). Red stars mark significant differences (binomial test).

Procedure. To demonstrate the generalization of our correction we use a different fabrication technique and a different varnish manufacturer. Using the Formlabs Form 2 printer we fabricate two geometries: *bunny* and *ghost*. To achieve a different glossy finish we manually mix oil-based varnishes from Amsterdam. We digitize the objects by measuring the Cook-Torrance parameters of each varnish mixture. To display the objects we use physically correct renderer with two settings: (1) reflectance parameters as measured, and (2) roughness modified by our correction from Equation 1. Figure 7 (bottom) demonstrates roughness correction for the *bunny* under the *diffuse* illumination. Twelve volunteers (21 to 32 years old, 5 females, 7 males, and no intersex/others) with normal or corrected to normal vision participated in this experiment. Participants are presented with a two-alternative forced choice experiment. We use the fabricated artifact as a reference. Their task is to select the closest match in terms of gloss between a corrected and not-corrected displayed sample.

Results. In Fig. 8 (right) we show the results of the study for each illumination and display brightness aggregated by geometry and glossiness levels (please refer to Sec. S8 in the supplemental for additional details). For the diffuse illumination users consistently chose our corrected displayed version as closer to the presented real object in terms of glossiness. For the one spotlight and complex illuminations we do not observe statistically significant differences between our correction and the original image. As we discuss in Sec. 4, for these two illuminations less correction is needed, therefore for some conditions participants may not be able to consistently identify the differences in perceived gloss.

6 CONCLUSIONS, LIMITATIONS, AND FUTURE WORK

This work investigates the problem of gloss appearance matching between the real world and display depiction as a function of material glossiness, surface geometry, scene illumination, and display luminance. To this end, we fabricate a wide range of differently painted glossy objects and create a dataset of the corresponding HDR photographs and rendered images. We use the dataset in a large-scale perceptual experiment, where we systematically investigate the gloss matching task as a function of essential factors influencing gloss perception. The collected data enables deriving the quantitive measurements of gloss difference between the real and virtual worlds. We find a strong dependency of such corrections on the scene illumination, display brightness, and weaker influence of the object's geometry. Finally, we propose a model which predicts the correction to minimize the gap between real samples and displayed counterparts. We show that our gloss correction can significantly reduce the appearance gap in the digital product design, fabrication, and digitalization of physical artifacts.

We consider artifacts manufactured from plastic, a typical 3D printing material. However, discrepancies between the real-world and displayed gloss can be also observed for metals [Tanaka and Horiuchi 2015] and other materials [Filip et al. 2018]. Similarly, we consider easy-to-control office illumination, while brighter outdoors scenes may require different, potentially more extensive, corrections. The same applies to the display characteristics. While our experiments consider standard 2D displays, a different correction may be required depending on the display dynamic range, tone reproduction [Phillips et al. 2009], the ability to reproduce motion parallax [Sakano and Ando 2010] and binocular disparity [Obein et al. 2004]. In Sec. 4 we observed that complex and strongly tessellated geometries (*dragon*) seem to produce larger discrepancies. Revealing the reason behind this phenomenon requires investigating the interaction between fabricated micro-geometry and deposited varnish. Extending our investigation and model to different materials, illuminations, display setups, and micro-geometry is an exciting avenue for future work with promising and more general applications to gloss appearance reproduction and management with predictive rendering. We derived our gloss correction based on extensive experiments with water-based varnishes covering a wide gloss range from matte to high gloss. We generalize our findings by formulating the correction through the roughness factor of a fitted reflectance model for each varnish. Our application demonstrates that this strategy enables correction for other varnish families. An interesting direction for future work is to investigate the application of our correction to alternative techniques for modifying surface roughness, such as polishing or microstructure patterning.

ACKNOWLEDGMENTS

This work is supported by FWF Lise Meitner (Grant M 3319), European Research Council (project CHAMELEON, Grant no. 682080), Swiss National Science Foundation (Grant no. 200502), and academic gifts from Meta.

REFERENCES

Wendy J Adams, Gizem Kucukoglu, Michael S Landy, and Rafał K Mantiuk. 2018. Naturally glossy: Gloss perception, illumination statistics, and tone mapping. *J. Vis.* 18, 13 (2018), 4–4.

Barton L Anderson. 2011. Visual perception of materials and surfaces. *Curr. Biol.* 21, 24 (2011), R978–R983.

Michael Ashikhmin and Jay Goyal. 2006. A Reality Check for Tone-Mapping Operators. *ACM Trans. Appl. Percept.* 3, 4 (2006), 399–411.

Teun Baar, Sepideh Samadzadegan, Hans Brettel, Philipp Urban, and Maria V Ortiz Segovia. 2014. Printing gloss effects in a 2.5 D system. In *Measuring, Modeling, and Reproducing Material Appearance*, Vol. 9018. SPIE, 160–167.

Roy S Berns. 1996. Methods for characterizing CRT displays. *Displays* 16, 4 (1996), 173–182.

Dankmar Böhning. 1992. Multinomial logistic regression algorithm. *Ann. Inst. Stat. Math.* 44, 1 (1992), 197–200.

Alan Brunton, Can Ates Arikan, and Philipp Urban. 2015. Pushing the limits of 3D color printing: error diffusion with translucent materials. *ACM Trans. Graph.* 35, 1 (2015), 1–13.

Alice C Chadwick and RW Kentridge. 2015. The perception of gloss: A review. *Vis. Res.* 109 (2015), 221–235.

Katja Doerschner, Huseyin Boyaci, and Laurence T Maloney. 2010. Estimating the glossiness transfer function induced by illumination change and testing its transitivity. *J. Vis.* 10, 4 (2010), 8–8.

Frédéric Drago and Karol Myszkowski. 2001. Validation proposal for global illumination and rendering techniques. *Comput. Graph.* 25, 3 (2001), 511–518.

Ron O. Dror, Alan S. Willsky, and Edward H. Adelson. 2004. Statistical characterization of real-world illumination. *J. Vis.* 4 (2004), 821–837.

Willemijn Elkhuizen, Tessa Essers, Yu Song, Jo Geraedts, Clemens Weijkamp, Joris Dik, and Sylvia Pont. 2019. Gloss, Color, and Topography Scanning for Reproducing a Painting's Appearance Using 3D Printing. *ACM J. Comput. Cult. Herit.* 12, 4 (2019), 1–22.

Mark D Fairchild. 2013. *Color appearance models.* John Wiley & Sons.

Franz Faul. 2019. The influence of Fresnel effects on gloss perception. *J. Vis.* 19, 13 (2019), 1–39.

Jiří Filip, Martina Kolafová, Michal Havlíček, Radomír Vávra, Michal Haindl, and Holly Rushmeier. 2018. Evaluating physical and rendered material appearance. *Visual Comput.* 34, 6-8 (2018), 805–816.

Roland W Fleming. 2017. Material perception. *Annu. Rev. Vis. Sci.* 3 (2017), 365–388.

Roland W. Fleming, Ron O. Dror, and Edward H. Adelson. 2003. Real-world illumination and the perception of surface reflectance properties. *J. Vis.* 3, 5 (2003), 3–3.

Guillaume Ged, Gaël Obein, Zaccaria Silvestri, Jean Le Rohellec, and Françoise Viénot. 2010. Recognizing real materials from their glossy appearance. *J. Vis.* 10, 9 (2010), 18–18.

Param Hanji, Fangcheng Zhong, and Rafał K Mantiuk. 2020. Noise-aware merging of high dynamic range image stacks without camera calibration. In *European Conference on Computer Vision.* Springer, 376–391.

Vlastimil Havran, Jiri Filip, and Karol Myszkowski. 2016. Perceptually motivated BRDF comparison using single image. In *Comput. Graph. Forum*, Vol. 35. 1–12.

Y-X. Ho, M.S. Landy, and L.T. Maloney. 2008. Conjoint measurement of gloss and surface texture. *Psychol. Sci.* 19, 2 (2008), 196–204.

R.S. Hunter and R.W. Harold. 1987. *The measurement of appearance* (2nd ed.). Wiley, New York.

Juno Kim, Phillip J. Marlow, and Barton L. Anderson. 2012. The dark side of gloss. *Nature Neuroscience* 15, 11 (2012), 1590–1595.

Yanxiang Lan, Yue Dong, Fabio Pellacini, and Xin Tong. 2013. Bi-Scale Appearance Fabrication. *ACM Trans. Graph.* 32, 4, Article 145 (2013), 12 pages.

Frédéric B Leloup, Michael R Pointer, Philip Dutré, and Peter Hanselaer. 2010. Geometry of illumination, luminance contrast, and gloss perception. *J. Opt. Soc. Am. A* 27, 9 (2010), 2046–2054.

Tom Malzbender, Ramin Samadani, Steven Scher, Adam Crume, Douglas Dunn, and James Davis. 2012. Printing Reflectance Functions. *ACM Trans. Graph.* 31, 3, Article 20 (2012), 11 pages.

Phillip J Marlow and Barton L Anderson. 2013. Generative constraints on image cues for perceived gloss. *J. Vis.* 13, 14 (2013), 2–2.

Phillip J Marlow, Juno Kim, and Barton L Anderson. 2012. The perception and misperception of specular surface reflectance. *Curr. Biol.* 22, 20 (2012), 1909–1913.

Kenichiro Masaoka, Yukihiro Nishida, Masayuki Sugawara, Eisuke Nakasu, and Yuji Nojiri. 2013. Sensation of Realness From High-Resolution Images of Real Objects. *IEEE Trans. Broadcast.* 59, 1 (2013), 72–83.

Wojciech Matusik, Boris Ajdin, Jinwei Gu, Jason Lawrence, Hendrik P. A. Lensch, Fabio Pellacini, and Szymon Rusinkiewicz. 2009. Printing Spatially-Varying Reflectance. *ACM Trans. Graph.* 28, 5 (2009), 1–9.

Ann McNamara. 2006. Exploring Visual and Automatic Measures of Perceptual Fidelity in Real and Simulated Imagery. *ACM Trans. Appl. Percept.* 3, 3 (2006), 217–238.

Gary W. Meyer, Holly E. Rushmeier, Michael F. Cohen, Donald P. Greenberg, and Kenneth E. Torrance. 1986. An experimental evaluation of computer graphics imagery. *ACM Trans. Graph.* 5, 1 (1986), 30–50.

Isamu Motoyoshi and Hiroaki Matoba. 2012. Variability in constancy of the perceived surface reflectance across different illumination statistics. *Vis. Res.* 53, 1 (2012), 30–39.

Isamu Motoyoshi, Shin'ya Nishida, Lavanya Sharan, and Edward H Adelson. 2007. Image statistics and the perception of surface qualities. *Nature* 447, 7141 (2007), 206–209.

Nico JD Nagelkerke et al. 1991. A note on a general definition of the coefficient of determination. *Biometrika* 78, 3 (1991), 691–692.

Gaë Obein, Kenneth Knoblauch, and Françise Viéot. 2004. Difference scaling of gloss: Nonlinearity, binocularity, and constancy. *J. Vis.* 4, 9 (2004), 4–4.

Maria Olkkonen and David H Brainard. 2010. Perceived glossiness and lightness under real-world illumination. *J. Vis.* 10, 9 (2010), 5–5.

Fabio Pellacini, James A Ferwerda, and Donald P Greenberg. 2000. Toward a psychophysically-based light reflection model for image synthesis. In *Proc. ACM SIGGRAPH.* 55–64.

Jonathan B Phillips, James A Ferwerda, and Stefan Luka. 2009. Effects of image dynamic range on apparent surface gloss. In *Color and imaging conference*, Vol. 2009. Society for Imaging Science and Technology, 193–197.

Michal Piovarči, Michael Wessely, Michał Jagielski, Marc Alexa, Wojciech Matusik, and Piotr Didyk. 2017. Directional screens. In *Proceedings of the 1st Annual ACM Symposium on Computational Fabrication*. ACM, 1.

Michal Piovarči, Michael Foshey, Vahid Babaei, Szymon Rusinkiewicz, Wojciech Matusik, and Piotr Didyk. 2020. Towards Spatially Varying Gloss Reproduction for 3D Printing. *ACM Trans. Graph.* 39, 6, Article 206 (2020), 13 pages.

Sylvia C Pont and Susan F te Pas. 2006. Material — Illumination Ambiguities and the Perception of Solid Objects. *Perception* 35, 10 (2006), 1331–1350.

María Pérez-Ortiz, Aliaksei Mikhailiuk, Emin Zerman, Vedad Hulusic, Giuseppe Valenzise, and Rafał K. Mantiuk. 2020. From Pairwise Comparisons and Rating to a Unified Quality Scale. *IEEE Trans. Image Process.* 29 (2020).

Yuichi Sakano and Hiroshi Ando. 2010. Effects of head motion and stereo viewing on perceived glossiness. *J. Vis.* 10, 9 (2010), 15–15.

Ana Serrano, Bin Chen, Chao Wang, Michal Piovarci, Hans-Peter Seidel, Piotr Didyk, and Karol Myszkowski. 2021. The effect of shape and illumination on material perception: model and applications. *ACM Trans. on Graph.* 40, 4 (2021).

SMPTE-196M. 2003. Motion-Picture Film - Indoor Theater and Review Room Projection - Screen Luminance and Viewing Conditions. *The Society of Motion Picture and Television Engineers* (2003).

Denis Sumin, Tobias Rittig, Vahid Babaei, Thomas Nindel, Alexander Wilkie, Piotr Didyk, Bernd Bickel, J KŖivánek, Karol Myszkowski, and Tim Weyrich. 2019. Geometry-aware scattering compensation for 3D printing. *ACM Trans. Graph.* 38, 4 (2019).

Midori Tanaka and Takahiko Horiuchi. 2015. Investigating perceptual qualities of static surface appearance using real materials and displayed images. *Vis. Res.* 115 (2015), 246–258.

Louis L Thurstone. 1927. A law of comparative judgment. *Psychol. Rev.* 34, 4 (1927), 273.

James T. Todd, J. Farley Norman, and Ennio Mingolla. 2004. Lightness Constancy in the Presence of Specular Highlights. *Psychol. Sci.* 15, 1 (2004), 33–39.

Matteo Toscani, Matteo Valsecchi, and Karl R Gegenfurtner. 2017. Lightness perception for matte and glossy complex shapes. *Vis. Res.* 131 (2017), 82–95.

TS Trowbridge and Karl P Reitz. 1975. Average irregularity representation of a rough surface for ray reflection. *J. Opt. Soc. Am. A* 65, 5 (1975), 531–536.

Jan Jaap R. van Assen, Maarten W. A. Wijntjes, and Sylvia C. Pont. 2016. Highlight shapes and perception of gloss for real and photographed objects. *J. Vis.* 16, 6 (2016), 1–14.

Peter Vangorp, Jurgen Laurijssen, and Philip Dutré. 2007. The influence of shape on the perception of material reflectance. In *Proc. ACM SIGGRAPH.* 77:1–77:9.

Peter Vangorp, RafałK̃. Mantiuk, Bartosz Bazyluk, Karol Myszkowski, Radosław Mantiuk, Simon J. Watt, and Hans-Peter Seidel. 2014. Depth from HDR: Depth Induction or Increased Realism?. In *Proc. ACM SAP 2014.* 71–78.

Josh Wills, Sameer Agarwal, David Kriegman, and Serge Belongie. 2009. Toward a perceptual space for gloss. *ACM Trans. Graph.* 28, 4 (2009), 1–15.

Akiko Yoshida, Volker Blanz, Karol Myszkowski, and Hans-Peter Seidel. 2005. Perceptual evaluation of tone mapping operators with real-world scenes. In *Human Vision and Electronic Imaging X*, Vol. 5666. SPIE, 192 – 203.

Akiko Yoshida, Rafał Mantiuk, Karol Myszkowski, and Hans-Peter Seidel. 2006. Analysis of Reproducing Real-World Appearance on Displays of Varying Dynamic Range. *Comput. Graph. Forum (Proc. of Eurographics)* 25, 3 (2006), 415–426.

Fan Zhang, Huib de Ridder, Pascal Barla, and Sylvia Pont. 2020a. Effects of light map orientation and shape on the visual perception of canonical materials. *J. Vis.* 20, 4, Article 13 (2020), 18 pages.

Fan Zhang, Huib de Ridder, Pascal Barla, and Sylvia Pont. 2020b. A systematic approach to testing and predicting light-material interactions. *J. Vis.* 19, 4, Article 11 (2020), 22 pages.

Fangcheng Zhong, Akshay Jindal, Özgür Yöntem, Param Hanji, Simon Watt, and Rafal Mantiuk. 2021. Reproducing reality with a high-dynamic-range multi-focal stereo display. *ACM Trans. Graph.* 40, 6 (2021), 241.