The effect of display capabilities on the gloss consistency between real and virtual objects

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

Akshay Jindal*
akshay.jindal@intel.com
Intel Corporation
Bellevue, United States of America
University of Cambridge
Cambridge, United Kingdom

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-inf.mpg.de
Max-Planck-Institut für Informatik
Saarbrücken, Germany

Piotr Didyk
piotr.didyk@usi.ch
Università 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

Rafał K. Mantiuk
rafal.mantiuk@cl.cam.ac.uk
University of Cambridge
Cambridge, United Kingdom
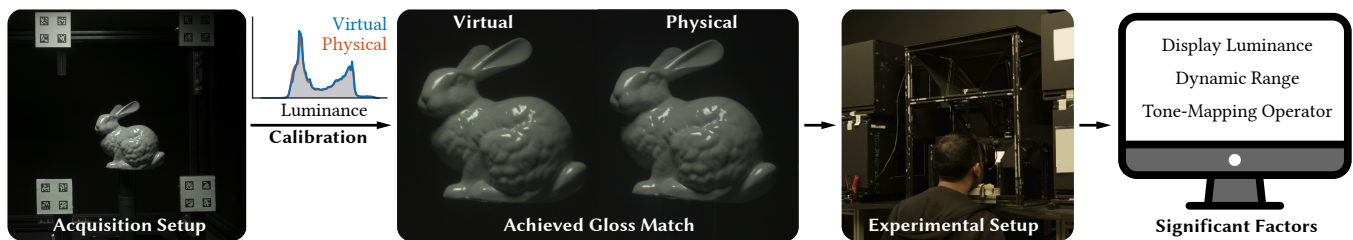
Figure 1: We study the reason behind the perceived gloss discrepancy between physical and virtual stimuli. To this end, we enhance an HDR stereoscopic display with an improved signal correction pipeline. Our setup allows for the first time to accurately match the level of perceived gloss between virtual and physical stimuli. We leverage the display to identify the key display properties for faithful gloss reproduction.

ABSTRACT

A faithful reproduction of gloss is inherently difficult because of the limited dynamic range, peak luminance, and 3D capabilities of display devices. This work investigates how the display capabilities affect gloss appearance with respect to a real-world reference object. To this end, we employ an accurate imaging pipeline to achieve a perceptual gloss match between a virtual and real object presented side-by-side on an augmented-reality high-dynamic-range (HDR) stereoscopic display, which has not been previously attained to this extent. Based on this precise gloss reproduction, we conduct a series of gloss matching experiments to study how gloss perception

degrades based on individual factors: object albedo, display luminance, dynamic range, stereopsis, and tone mapping. We support the study with a detailed analysis of individual factors, followed by an in-depth discussion on the observed perceptual effects. Our experiments demonstrate that stereoscopic presentation has a limited effect on the gloss matching task on our HDR display. However, both reduced luminance and dynamic range of the display reduce the perceived gloss. This means that the visual system cannot compensate for the changes in gloss appearance across luminance (lack of gloss constancy), and the tone mapping operator should be carefully selected when reproducing gloss on a low dynamic range (LDR) display.

CCS CONCEPTS

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

KEYWORDS

gloss perception, gloss reproduction, tone mapping, reproducing reality, HDR, computational display

1 INTRODUCTION

Gloss is an important appearance cue vital to material perception. It often conveys information about the object's functionality, quality, or value. Therefore, accurate modeling, reproduction, and visualization of gloss properties are critical in many applications, such as product design, manufacturing, or any applications benefiting from augmented reality, where seamless integration of the real and the virtual worlds is critical. Consequently, a faithful gloss reproduction on a display device is essential for all these applications.

Despite continuous improvements in display technologies, it has been shown that there is still a significant discrepancy in gloss perception between real-world and rendered stimuli [Fores et al. 2013; Xiao et al. 2023]. Recently, Chen et al. [2022] conducted an in-depth investigation of this effect and discovered that the difference in perceived gloss is relatively stable across various geometries and illumination conditions. Unfortunately, while they were able to investigate and model the magnitude of the effect for a specific display-object-illumination combination, the exact cause remains unknown. Understanding the source of this mismatch remains an open research problem critical for future works aiming to propose scene-agnostic gloss compensation or novel screen designs with the correct gloss display.

In this work, we build upon the prior studies and aim to find the core factors influencing the mismatch in perceived gloss between virtual and real stimuli from the display point of view. To achieve this goal, a faithful reproduction of gloss on display, as the first step, is an essential prerequisite to further study on other individual factors. We employ an accurate imaging pipeline, MTF correction, and a camera-guided tone correction step to bring gloss perception in the display exceedingly close to the real world (Sec. 3), which has not been achieved by previous work. Based on this, we provide a systematic study on the effects of the object albedo, display luminance, dynamic range, stereoscopic capabilities on perceived gloss. In contrast to prior work, our experiment directly compares the perceived gloss between real-world objects and virtual stimuli displayed side-by-side on an HDR stereoscopic display.

The analysis of the collected data reveals that stereopsis has limited influence on the perceived gloss in our gloss-matching experiment despite reports from users expressing a noticeable decrease in realism in the monoscopic condition compared to the stereoscopic condition. The peak luminance and dynamic range of the display are two significant factors driving the mismatch in perceived gloss. We observe a tight correlation between the variation of perceived gloss and perceived contrast across scenes shown at varying luminance levels. In order to delve deeper into the effect of dynamic range, we conduct another user study using different tone mapping operators with varied cut-off luminance and steepness. Interestingly, we observe that the over-exposed/cut-off effect caused by a highly steep tone mapping operator increases the perceived

gloss, indicating that a carefully selected tone mapping operator can be used to compensate the gloss degradation due to contrast compression on LDR display. We confirm this finding by a pairwise comparison study on images compressed by 15 different tone mapping operators generated from assuming three different cut-off luminance levels and five steepness (Sec. 6). The data and code are available at https://stereohdrgloss.mpi-inf.mpg.de.

2 RELATED WORK

We begin with a brief discussion on factors affecting gloss perception and display capabilities required for faithful reproduction of gloss. We then summarize prior studies on gloss perception that involved a direct comparison of displayed images to physical objects.

Gloss perception. Gloss is one of the essential visual attributes that contribute to the overall look and feel of a material, alongside other factors such as color, texture, and transparency [Anderson 2011; Fleming 2017]. Previous studies have shown that the material's reflectance properties are not the sole factor that influences gloss perception; surface geometry and the illumination of the scene can also affect the perception of gloss. Geometry with higher surface curvature tends to increase the highlight intensity, thereby boosting the perceived gloss. However, the opposite effect can also be observed — bumpy surface with overly high curvature results in reduced contrast and sharpness of reflection [Ho et al. 2008; Marlow et al. 2012]. Furthermore, the presence of strong directional lighting or high-contrast patterns in the environment can also amplify the perceived gloss of a material [Adams et al. 2018; Dror et al. 2004; Pont and te Pas 2006; Zhang et al. 2020]. The interaction between geometry and illumination leads to complex modifications to contrast, coverage, distinctiveness of highlights, and image statistics on material surface [Marlow and Anderson 2013; Marlow et al. 2012; Motoyoshi et al. 2007].

Gloss perception on display. While modern desktop displays can reproduce real-world colours and textures convincingly, for more subtle tasks, such as gloss reproduction investigated in this work, they are typically limited by their dynamic range and the lack of binocular depth cues. Reflections of light sources appear as glossy highlights that exhibit binocular disparity with respect to the reflecting surface. Such disparity might lead to difficulties in binocular fusion or even uncomfortable binocular rivalry [Muryy et al. 2016; Templin et al. 2012], in particular for strongly glossy materials, where disparity matching improves with the distinctiveness of reflection patterns [Hess et al. 1999]. While it is well-accepted that binocular cues enhances glossy surface appearance [Hurlbert et al. 1991; Mausfeld et al. 2014; Wendt and Faul 2019; Wendt et al. 2010, 2008; Yamamoto et al. 2012], there is evidence that reproduction of physically-correct disparity is not necessary for gloss realism [Blake and Brelstaff 1988; Blake and Bülthoff 1990; Kerrigan and Adams 2013]. Our work offers further insights into these observations as we can better isolate these factors using a high-fidelity HDR stereoscopic display, similar to the one that passed the visual Turing test [Zhong et al. 2021]. In this work, we adopt a similar display specifically for the gloss matching task listed as a limitation in the original work.

The majority of commercial displays fall short of the real-world dynamic range, especially when it comes to the reproduction of high-gloss surfaces. Because of that, (HDR) images of real scenes need to be tone mapped before they can be shown on a display. However, this operation influences gloss reproduction, which is a well-known and unresolved problem in gloss perception research [Adams et al. 2018; Doerschner et al. 2010; Ferwerda et al. 2001; Fleming et al. 2003; Fores et al. 2014; Pellacini et al. 2000; Wills et al. 2009]. Phillips et al. [2009] found that images displayed on HDR displays appear glossier than those on LDR displays and that differences in gloss are more easily perceivable on HDR displays. Adams et al. [2018] show that a tone mapping operator has a substantial effect on the perceived gloss, especially when the object is isolated from the background. In this work, we first perfectly reproduce real object's dynamic range and then further investigate the role of the display's peak luminance and tone mapping operators on gloss perception.

While movement of a specular reflection across the surface is also reported as a convincing cue that enhances gloss perception [Doerschner et al. 2011; Hartung and Kersten 2003; Lichtenauer et al. 2013; Wendt et al. 2010] and can be recreated on head tracking-based stereoscopic [Sakano and Ando 2010] and multi-view autostereoscopic [Sakano and Ando 2012, 2021] displays, we do not consider multi-view displays and motion parallax effects in this work.

Reality vs. displayed images. Faithful reproduction of reality is one of the "holy grails" in computer graphics and display technologies. While previous works have compared real-world scenes and displayed version [Drago and Myszkowski 2001; Masaoka et al. 2013; McNamara 2006; Meyer et al. 1986; Zhong et al. 2021], we focus on works related to gloss reproduction. Prior work has reported gloss mismatch between real-world materials and their photographs [Tanaka and Horiuchi 2015] or rendered images [Filip et al. 2018]. Xiao et al. [2023] extended this exploration to VR and established a relationship between parameters of the Cook-Torrance BRDF model and perceived gloss. Fores et al. [2013] reveal that the gloss sensitivity of human vision is different between real objects and displayed images through a side-by-side comparison. Van Assen et al. [2016] investigated the effect of the spatial structure of illumination on gloss perception using a real sphere and its digital photograph. They found a significant effect of highlights shape on perceived gloss in both real and display settings and noted a mismatch between perceived gloss between the two settings. More recently, Chen et al. [2022] evaluated the effect of geometry, illumination, and display luminance on the gloss discrepancy between real-world objects and their displayed counterpart on LDR display. A lookup table has been proposed to compensate for the gloss discrepancies in the considered conditions. However, the reasons behind these discrepancies are still unknown. All of these works studied the gloss of real and virtual objects either in two independent experiments or on LDR display, which also varied in geometries, materials, dynamic range, or possibly viewing conditions without accurate gloss reproduction. In contrast, our experimental setup facilitates a side-by-side comparison of a real-world scene and a displayed HDR stereoscopic image with highly accurate gloss reproduction, allowing for more accurate psychophysical measurements.

3 CAPTURE AND REPRODUCTION OF GLOSS

The main goal of this work is to identify which display parameters are responsible for the faithful reproduction of gloss. An essential prerequisite to such an investigation is an electronic display that can match the perceived gloss of real-world objects. As a starting point, we used a custom-built HDR stereoscopic display, similar to the one described in [Zhong et al. 2021]. The display allows a single observer to view stereoscopic images through an optical setup comparable to the Wheatstone mirror stereoscope (Fig. 1 right). This display could reach a peak luminance greater than $3\,000\,\text{cd/m}^2$ and a black level around $0.01\,\text{cd/m}^2$, which far exceeds the range found in most commercial displays. The display from the work of Zhong et al. [2021] enabled them to pass the visual Turing test and create virtual depictions of real objects indistinguishable from their real-world counterparts. For more details about the display please refer to the supplemental material. Unfortunately, the Turing capabilities of the display are limited to matte or semi-matte surfaces. The authors attribute this to two effects: (1) optical blur, which is especially prevalent in gloss reflections, and (2) inaccuracy in lumigraph synthesis across the entire dynamic range of the display. In this section, we elaborate an accurate image acquisition and correction pipeline that addresses both issues. In the next section, we validate that our improvements provide a faithful perceived gloss reproduction. We start by briefly summarizing our image acquisition and display pipeline, followed by our improvements towards gloss reproduction.

3.1 Light field capture

We capture a dense HDR light field of the real object with a 2D gantry mounted with a SONY A7R III camera ($f11$, ISO100) fitted with a 55 mm prime lens (Sony SEL55F18Z). The HDR exposure stack was demosaiced [Menon et al. 2006], merged into a 16-bit HDR image considering noise reduction [Hanji et al. 2020] and color correction to the Rec.709 RGB space as used by our display [Finlayson et al. 2015]. The camera pose of each light field image is recovered using removable AprilTag markers [Olson 2011] placed in the scene box, (Fig. 1 left). Finally, we render the light field using the recovered camera poses and a homography-based light field rendering algorithm [Isaksen et al. 2000] that can deliver accurate binocular disparity to each participant.

3.2 Reproduction of gloss

While our capture and calibration setup can acquire high-fidelity images, it is still prone to inaccuracies arising from optical limitations that can have a significant effect on gloss perception. Next, we discuss how these errors can be mitigated using MTF-based correction and camera-guided tone correction.

MTF correction. Every camera system, even with high-quality lens, suffers from optical aberrations. We found that such aberrations have a strong effect on the reproduced gloss and must be corrected. A first-order approximation that models such aberrations is the shift-invariant modulation transfer function (MTF) of the lens, which describes the attenuation of spatial frequencies. Here, we use an MTF measurement to reduce the effect of such aberrations. We

start by capturing the Modulated Sinusoidal Siemens Star chart[1], (Fig. 7 left). The edge at the center of the photograph is used to calculate the modulation transfer function (MTF) of the camera [Burns and Williams 2018]. We fit a Gaussian function to the data to alleviate the measurement errors in the estimated MTF. Before fitting a smooth curve to the MTF data, we restrict the small values so that the modulation does not drop below 0.3. This ensures that the noise present at high frequencies is not amplified. The fitted and modified MTF for our camera is shown in Fig. 7 middle. Finally, to obtain a sharp image of the specular reflection we perform deconvolution on the captured images using the smooth MTF filter. More details on this procedure are available in the supplementary. An example result is shown in Fig. 7 right. The blurry appearance of the highlights in the processed image has been largely reduced compared to the original image. We found the simple Fourier space deconvolution to achieve similar quality to more complex Lucy-Richardson and Wiener deconvolution methods [Campisi and Egiazarian 2017]. This was likely due to the high dynamic range and relatively low noise in our images.

Camera-guided tone correction. The MTF correction successfully recovers the high-frequency details in the image. However, the slanted-edge technique we used cannot accurately capture the low-frequency portion of the MTF. To compensate for the changes in low frequencies (mostly due to glare), we apply a camera-guided tone correction step to further improve the match of the captured gloss. 1) We display the light field at 100% gloss level, which spans the widest dynamic range, with its real object side-by-side. 2) Then, we capture the image from the position of one of the viewer's eyes. 3) Next, we calculate the histograms of luminance distributions for the real and displayed objects, which are shown in Fig. 8 bottom left, where a small but significant mismatch can be observed between the two distributions. 4) Finally, we match the distributions and improving the appearance of the displayed object using a histogram matching method. To preserve the calibrated color, we perform the procedure for one color channel (green, as it covers the largest intensity range in our case) and use the same mapping for other channels. The same procedure can be repeated to optimize the accuracy of histogram matching function (curve in Fig. 8 bottom right). We found one iteration is enough to achieve a significant improvement in the overall luminance match (Fig. 8 bottom middle). The mapping function enhanced contrast in the dark areas, indicating that the mismatch was indeed caused by glare. We utilized the same matching function for light fields at all the gloss levels. Please refer to the supplemental for more details about the used histogram matching algorithm.

4 MAIN EXPERIMENT: FACTORS INFLUENCING GLOSS PERCEPTION

Our improved signal correction pipeline allows us to capture the full dynamic range of our scenes, reaches satisfactory peak luminance, and provides binocular cues. Moreover, the individual display properties can be toggled on/off. For the first time, these capabilities allow us to investigate which display factors drive the perceived gloss discrepancy between displayed and real stimuli. To this end,

we conduct a psychophysical experiment where we investigate the following hypotheses: *Disabling individual display capabilities has an effect on the magnitude of perceived gloss estimated during a gloss matching task.*

4.1 Stimuli

Previous studies [Chen et al. 2022, 2021; Serrano et al. 2021] suggest that the geometry does not typically contribute to a significant discrepancy between the real and displayed stimuli. The notable exception are highly complex geometries with micro-scale geometry variations at the scale of surface roughness. Therefore, for our studies, we opted for the Stanford *bunny* geometry and produced *10* gloss levels (cover approximately $8.7 \sim 87$ gloss units measured with a glossmeter at $60°$) following a similar fabrication procedure of Chen et al. [2022] (Fig. 2 bottom). We apply 3D printed objects with gray painting for our experiment. However, we also verify the influence of albedo by adding 10 black objects. To reduce the number of times we needed to manually swap real objects during the experiment, we used a turntable holding four objects at a time, controlled from a PC.

The type of illumination plays a critical role in amplifying perceived gloss discrepancy. In general, diffuse illumination creates a greater discrepancy than complex illumination [Chen et al. 2022; Tanaka and Horiuchi 2015]. However, pure diffuse light provides a limited gloss range and is not commonly used in our daily life, while complex light sources potentially decrease the perceived gloss [van Assen et al. 2016]. Therefore, in our experiment, we selected a more representative *area light* that provides a good trade-off between producing glossy reflections and manifesting a perceived gloss difference. We lit the real-scene-box with a LitraStudio 3000 Lumen RGBWW programmable LED light. Refer to the supplemental for a demo of our illumination.

4.2 Display Factors

The production of various visual cues by a display has a direct influence on its ability to reproduce gloss. Nowadays, despite display resolution and refresh rate surpassing the thresholds of human visual perception, the dynamic range and support for binocular vision are still in a continuous process of development. This paper specifically examines the impact of a display's dynamic range, stereoscopic capabilities, and luminance on its ability to reproduce gloss accurately. Here, we provide a detailed motivation for different display factors in our main experiment. For a quick overview, see Tab. 1.

Baseline. In order to demonstrate the perceptually accurate gloss reproduction, we conduct a gloss matching experiment that incorporates all available visual cues provided by our customized HDR stereoscopic display, denoted as $C_{baseline}$ in Tab. 1. By varying the conditions of various factors in the baseline experiment, we created experiment conditions to investigate the effects of each factor individually.

Albedo. To validate if our display's ability to reproduce gloss generalizes well to different material albedo, we repeat the $C_{baseline}$ condition with black-colored bunnies denoted as C_{black}.

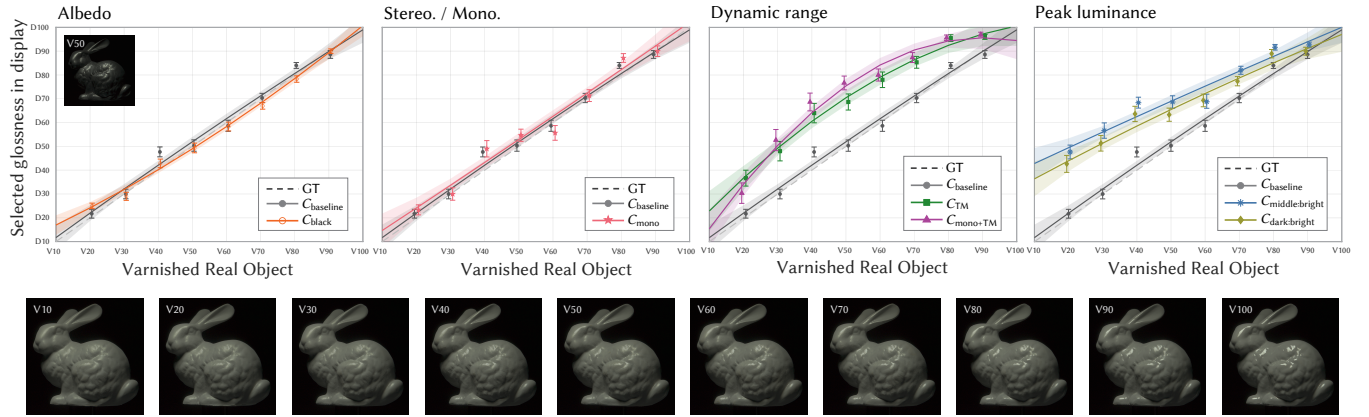[1]TE253 chart by Image Engineering

Figure 2: To create each plot, we grouped the data according to the levels of the corresponding factor. The dashed black line on each plot indicates the ground-truth selection, which occurs when the displayed image matches the corresponding varnished real sample. The filled points represent the mean matching gloss level, while the error bars indicate the standard error. To represent the data, we used a second-order polynomial function and plotted the associated colored line with a light-colored shadow representing the 95% confidence interval. The bottom row shows the HDR images of the objects corresponding to each gloss level in the plots. Note that for better visualization, all stimuli images are enhanced by γ equal to $\frac{1}{3.6}$.

Table 1: The conditions in our main experiment. We use bold to represent changed factors relative to C_{baseline}. Note that the tone mapping operator used in C_{TM} and $C_{\text{mono+TM}}$ is the same as $TM_{\text{low-gradual}}$ used in Sec. 6

Conditions	Lightness	Stereo./Mono.	Dynamic range	Lumi. (display:real)
C_{baseline}	gray	stereoscopic	HDR	bright : bright
C_{black}	**black**	stereoscopic	HDR	bright : bright
C_{mono}	gray	**monoscopic**	HDR	bright : bright
C_{TM}	gray	stereoscopic	**TM**	bright : bright
$C_{\text{mono+TM}}$	gray	**monoscopic**	**TM**	bright : bright
$C_{\text{middle : bright}}$	gray	stereoscopic	HDR	**middle : bright**
$C_{\text{dark : bright}}$	gray	stereoscopic	HDR	**dark : bright**

Stereoscopic. To examine the effect of stereoscopic presentation on perceived gloss, we switch to monoscopic presentation by showing the same image to both eyes, denoted as C_{mono}. Crossed fusion stereoscopic pairs for both stereoscopic and monoscopic views using 50% gloss level stimuli are shown in Fig. 9 left and right, respectively.

Dynamic Range. To evaluate the effect of dynamic range, we include a condition with light fields that were tone mapped, denoted as C_{TM} (the same as $TM_{\text{low-gradual}}$ used in Sec. 6). We also included a condition that combined monoscopic presentation and tone mapping, as it is the most common scenario in which gloss is presented on a display. Thus, we also include another condition, $C_{\text{mono+TM}}$, to examine the interaction of these two factors.

Display Peak Luminance. The display brightness or luminance is a significant aspect that impacts the gloss consistency between the real world and displayed image [Chen et al. 2022]. To study this effect, we created two conditions in which a real object at original *bright* illuminance level was matched to its depiction on a dimmer display. We dimmed the display luminance 10× (denoted as *middle*)

or 100× (denoted as *dark*) relative to the real object, corresponding to two experiment conditions: $C_{\text{middle:bright}}$, and $C_{\text{dark:bright}}$.

4.3 Procedure

Ten volunteers participated in this experiment, aged from 22 to 31, 3 females and 7 males, recruited from research students. All participants had normal or corrected-to-normal vision, normal stereo acuity (tested with the Titmus stereo "fly test"), and were naïve to the experiment's goal. The experiment was approved by the department's ethics board. The participants provided written consent and were compensated for their time.

Before the experiment, each participant completed a calibration procedure to ensure the light field is rendered at the correct location relative to their eyes. In the experiment, participants were shown a real object and its rendering side-by-side and could scroll between the ten light fields corresponding to different gloss levels of that object. The rendering was initialized randomly from one of the ten possible gloss levels. The participant's task was to: *"Adjust the gloss level of the rendering so that it closely resembles the real object"*.

We excluded the most glossy and the least glossy objects from the set of real objects to prevent data saturation in extreme cases, such that participants always had an option to increase/decrease the gloss of the rendering relative to the real object. We also added a small random ±5° offset to the turntable holding the stimuli, to prevent participants from comparing textural imperfections on objects' surface. This experiment consisted of 192 trials in total, divided into two 1 hour sessions. All conditions were mixed together and each condition was repeated 3 times. All trials were randomly ordered to enhance the experiment robustness. Participants were allowed to take breaks at any point in the experiment.

5 FACTORS AFFECTING PERCEIVED GLOSS

In this section, we report an analysis of the data collected in the main experiment. During the experiment, we asked participants to

pick one out of ten images that most closely match the gloss of the real object. Since our dependent variables (matching gloss levels) are categorical ordinal variables, we use ordinal logistic regression for data analysis. We select one level as the baseline for each factor and calculate the regression of the remaining levels with respect to it. Please refer to the supplementary material for the complete statistical analysis including Nagelkerke pseudo R-squares for the goodness of fit of the models, estimates of the regression coefficients, confidence intervals (2.5% and 97.5%), t-values (t), and p-values (p) for each factor.

Accurate reproduction of gloss. We report our results in Fig. 2. As seen in the first plot, participants can almost perfectly match the correct gloss level (mean square error of $1.02e{-}3$ assuming gloss of V100 to be 1) in C_{baseline} condition, indicating a highly accurate reproduction of gloss. We further validated the efficacy of our MTF correction and camera-guided tone correction (Sec. 3.2) on gloss reproduction by repeating C_{baseline} without these corrections. The results present a large mismatch between the perceived gloss of virtual and real stimuli, indicating adding our correction produces a significant improvement in gloss perception, and it brings gloss perception in the display closer to that of the real world. More details on this additional study can be found in the supplementary.

Effect of albedo. We plot the collected data of condition C_{black} in the first graph of Fig. 2. Similar as C_{baseline}, the fitted curve from participants' selections aligned with a near-perfect gloss match. The statistical analysis did not reveal significant differences between the gray and black objects ($p = 0.188$), which supports and complements the findings of Chen et al. [2022]. Our and their results suggest that albedo is not a significant factor for perceived gloss differences between real and displayed objects in most types of illuminations, except for pure diffuse light, which is uncommon in practical scenarios.

Effect of stereo cues. As shown in the second plot in Fig. 2, the results suggest a slight drop in perceived gloss in the monoscopic condition compared to the stereoscopic (baseline) condition. However, this effect was small, and we have no evidence of statistical significance ($p = 0.904$). It is noteworthy that participants informally reported a decrease in the realism of the monoscopic condition after finishing the experiment. We encourage the reader to compare the monoscopic and stereoscopic conditions in Fig. 9 and more in supplemental. The lack of statistical difference could be due to factors such as the viewing distance and small curvature of the object's surface, which could limit the amount of binocular disparity. However, the lack of the effect agrees with previous findings that suggest that while disparity is a critical cue for improving gloss constancy performance [Hurlbert et al. 1991; Wendt et al. 2010, 2008], it may not always be necessary for accurate gloss matching in monocular condition [Fores et al. 2013; Obein et al. 2004], particularly when all other visual cues are faithfully reproduced.

Effect of display peak luminance. When the participants were matching a bright real object to the gloss reproduced on a dim display, they selected much higher gloss levels than that of the real object (the rightmost panel in Fig. 2). This effect was significant when the display luminance was reduced 10× (middle:bright condition $p < 0.001$), and 100× (dark:bright condition $p < 0.001$). This
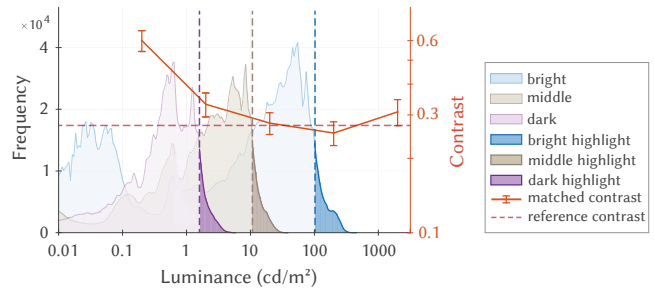

Figure 3: The luminance histograms for the bunny object (50% gloss) shown at three display luminance levels is compared with the matching contrast data from [Ashraf et al. 2022] (orange line). To maintain the perceived magnitude of contrast, physical contrast must be increased as the luminance is reduced. The same effect may apply to perceived gloss.

effect can be explained by the changes in perceived contrast with luminance. When contrast (of a sinusoidal grating) is matched across luminance levels, higher contrast is needed at lower luminance to match the reference contrast at higher luminance [Kulikowski 1976]. The exact mechanism behind such a failure of contrast constancy is still disputed, but some authors note the similarity between supra-threshold contrast matching and detection thresholds [Kulikowski 1976; Peli et al. 1991]. In the case of our experiment, the observers most likely increased the gloss level to compensate for the loss of perceived contrast on the dark display. This is illustrated in Fig. 3, in which we plot the histograms of absolute luminance at three display luminance levels together with a line of matching contrast from a recent study [Ashraf et al. 2022]. We can presume that gloss matching follows a similar characteristic as matching physical contrast.

Effect of dynamic range. We simulate the condition of 3D (stereo) and 2D LDR display by applying a simple tone mapping operator to our HDR stimulus (same as $TM_{\text{low-gradual}}$ in Sec. 6). Our results, as displayed in the third plot of Fig. 2 show that the dynamic range has a strong effect on the perceived gloss ($p < 0.001$), with the C_{TM} condition being perceived to have a lower gloss than the HDR condition. This outcome aligns with prior research on traditional displays [Phillips et al. 2009]. The statistical analysis did not yield significant evidence for an interaction effect ($p = 0.463$) between the tone mapped stereo condition (C_{TM}) and the tone mapped mono condition ($C_{\text{mono+TM}}$). This finding, together with the previously discussed results of our study, emphasizes that the primary factor influencing matching gloss between real and displayed objects is the dynamic range compressed by tone mapping.

6 THE EFFECT OF TONE MAPPING ON PERCEIVED GLOSS DIFFERENCES

In this section, we conduct additional series of experiments specifically for tone mapping, in which we keep all other factors the same as C_{baseline} in Tab. 1 except the dynamic range.
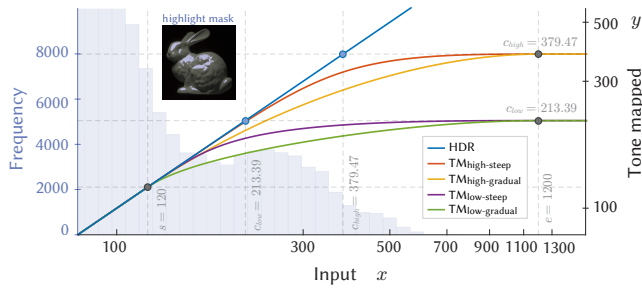
Figure 4: Four tone mapping operators used in our experiments. We draw the intensity distribution (green channel) of our 50% gloss stimulus in highlight region with y-axis on the left, and tone mapped luminance with y-axis on the right. The presented range of distribution corresponds to the masked highlight regions shown in inset.

Figure 5: Results of tone mapping experiment for both stereoscopic (left) and monoscopic (right) presentation. The overlap of $TM_{\text{high-gradual}}$ and $TM_{\text{low-steep}}$ suggests complex effect of tone mapping on gloss perception.

6.1 Stimuli

As a representative tone-mapping operator, we opted for Bézier curve, which can be represented as $TM(s, e, c, \rho, \tau)$, where s represents the compression starting point, e is the compression ending point, c is the clipping point, and ρ and τ can be used to control the steepness of a tone curve by either adjusting ρ with pre-fixed τ, or adjusting τ with a fixed ρ. The details of tone curve generation are presented in the supplemental. In this experiment, we utilize four tone curves: $TM_{\text{high-steep}}$, $TM_{\text{high-gradual}}$, $TM_{\text{low-steep}}$, and $TM_{\text{low-gradual}}$, in which *high* and *low* represent clipping points at 379.47 cd/m^2 and 213.39 cd/m^2, *steep* and *gradual* correspond to $\rho = 0$ and $\rho = 0.3$ respectively.

6.2 Results and analysis

We applied the same ordinal logistic regression as the main experiment. For the effect of the tone mapping operators, we first select HDR as a baseline and compare all tone mapping operators with respect to HDR. Then, for comparing the tone mapping operators with each other, we select each as a baseline and observe the effects of the others. As shown in Fig. 5, there are statistically significant differences between different tone mapping operators: HDR and $TM_{\text{high-steep}}$ form a cluster; $TM_{\text{high-gradual}}$ and $TM_{\text{low-steep}}$ form another cluster, while $TM_{\text{low-gradual}}$ is significantly different from all others. The gloss difference is statistically indistinguishable within each cluster, and this conclusion holds for both stereoscopic (left) and monoscopic (right), which can be explained by similar reasons reported in Sec. 5. Please refer to supplemental for details of data analysis. The most interesting phenomenon happens on $TM_{\text{high-gradual}}$ and $TM_{\text{low-steep}}$, where even though the clipped points (effectively, the highlight intensity is clipped as seen in Fig. 4) are very different (379.47 cd/m^2 *vs.* 213.39 cd/m^2), they form a cluster with fitted curve almost overlapped Fig. 5), especially for experiments with stereopsis. The effect of different tone curves can be better observed in Fig. 10, where we select the stimulus with 50% gloss level and visualize the 1D scanline across the highlight regions of the HDR image and its four tone mapped versions. The luminance level of $TM_{\text{high-gradual}}$ is clearly much higher than that of $TM_{\text{low-steep}}$, and $TM_{\text{high-gradual}}$ presents much higher variance
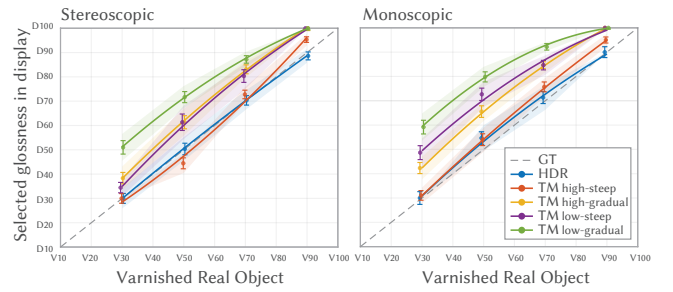
in highlight region than $TM_{\text{low-steep}}$, in which texture details have been smoothed out. This might indicate the gloss perception is preserved when using suitable tone mapping operator even in lower dynamic range, which is attractive for reproducing perceptually accurate gloss on LDR displays. To validate this phenomenon, we conduct another experiment in Sec. 6.3.

6.3 Over-exposed/cut-off effect on gloss perception

Based on the analysis in Sec. 5 and Sec. 6.2, stereopsis has a limited effect on the perceived gloss. Thus, we conduct this experiment on a 2D HDR display qualified by *Display HDR 1000* standard (Dell UltraSharp 32 4K HDR Monitor (UP3221Q)). Stimulus with middle gloss level 50% is selected as a representative in this experiment. We choose 3 clipping luminance levels typical of consumer displays: 160 cd/m^2, 250 cd/m^2, 500 cd/m^2. For each clipping level, five tone mapping operators with different steepness (controlled by τ) are generated, shown in Fig. 12 bottom, together with the compressed images in the top three rows. We conduct a pairwise comparison experiment on all 15 tone-mapped images and the original HDR image. Please refer to the supplemental document for more details about this experiment.

Results. The pairwise comparison results from this experiment were mapped to a Just-Noticeable-Difference (JND) scale using Thurstone Case V assumptions and the 95% confidence intervals were calculated using bootstrapping [Perez-Ortiz and Mantiuk 2017]. As shown in Fig. 6, $[160, S_1]$ and $[160, S_2]$ are perceived glossier than $[250, S_5]$; $[250, S_1]$ is perceived glossier than $[500, S_5]$; $[250, S_1]$, $[500, S_1]$, and $[HDR]$ result in almost the same perceived gloss. This is strong evidence demonstrating that a tone curve with high steepness could produce similar or even higher gloss perception than a tone curve with higher cut-off luminance but lower steepness.

This phenomenon might be explained by a combination of two reasons. The first reason is that the tone mapping operator smooths out the details within the highlight region, which is equivalent to increasing the smoothness of geometry surface, creating an illusion of sharper highlight that further results in higher perceived gloss. This effect can be found in material surfaces with high smoothness
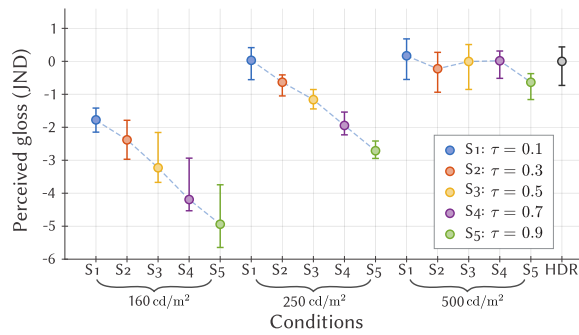
Figure 6: Perceived gloss in JND scale in our pair-wise comparison study. A difference of 1 JND between conditions A and B means that 75% of the population will pick condition A to be glossier than condition B. The HDR image was assigned as the 0 JND condition and other data points are presented relative to it. The luminance values denote the clipping luminance and S_1-S_5 denotes the steepness of the tone curve (in decreasing order). Refer to supplemental for pair-wise comparison matrix.

and low reflectivity like glossy ceramic [Schmid et al. 2020], corresponding to our high gloss stimuli cut-off by high-steep tone curve (first row of Fig. 11).

The second reason is that the tone mapping operator triggers the glare illusion effect. The glare illusion effect pertains to a phenomenon where a smooth gradient surrounding a bright local region evokes a perception of self-luminosity [Yoshida et al. 2008], increasing the sensation of luminance and further boosting the perceived gloss. This effect is more likely to happen on material surfaces with low smoothness and high reflectivity like metal [Schmid et al. 2020], corresponding to our low gloss stimuli cut-off by high-steep tone curve (second row of Fig. 11).

7 CONCLUSIONS

In this paper, we explore the reasons behind the discrepancy between the displayed image and its real-world reference from the perspectives of object albedo, display peak luminance, dynamic range, stereoscopic capabilities, and the effect of tone mapping. We use a signal correction pipeline that significantly improves the gloss reproduction on a custom-built stereoscopic HDR display, which is an essential prerequisite allowing us to measure the effect of other factors accurately. By analysis of the collected data, we reveal the reason behind the effect of each factor, providing insight and guidelines for future research. Besides, an interesting over-exposed/cut-off effect on specular highlight caused by high-steepness tone mapping operators has been revealed and validated in an additional experiment. Future work could focus on the effect of motion parallax, which has been reported as a strong factor that influences gloss perception. With the focus on the capabilities of the display, we selected representative geometry and illumination for our experiments. The finer effects of geometry, like complex surfaces, and illumination, like diffuse lighting, could be further studied in the future. Our current conclusion is limited to plastic and opaque objects; other materials like metal or transparent or

translucent material [Gigilashvili et al. 2021] could be interesting directions for future research.

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.

Maliha Ashraf, Rafal Mantiuk, Jasna Martinovic, and Sophie Wuerger. 2022. Suprathreshold contrast matching between different luminance levels. In *Color and Imaging Conference*, Vol. 30. Society for Imaging Science & Technology, 219–224.

A. Blake and G. Brelstaff. 1988. Geometry from Specularities. In *Proc. Int. Conf. on Computer Vision*. 394–403.

A. Blake and H. Bülthoff. 1990. Does the brain know the physics of specular reflection? *Nature* 343, 6254 (1990), 165–168.

Peter D Burns and Don Williams. 2018. Camera resolution and distortion: Advanced edge fitting. *Electronic Imaging* 30 (2018), 1–5.

Patrizio Campisi and Karen Egiazarian. 2017. *Blind image deconvolution: theory and applications*. CRC press.

Bin Chen, Michal Piovarči, Chao Wang, Hans-Peter Seidel, Piotr Didyk, Karol Myszkowski, and Ana Serrano. 2022. Gloss management for consistent reproduction of real and virtual objects. In *SIGGRAPH Asia 2022 Conference Papers*. 1–9.

Bin Chen, Chao Wang, Michal Piovarči, Hans-Peter Seidel, Piotr Didyk, Karol Myszkowski, and Ana Serrano. 2021. The effect of geometry and illumination on appearance perception of different material categories. *Visual Comput.* 37, 12 (2021), 2975–2987.

K. Doerschner, R.W. Fleming, O. Yilmaz, P.R. Schrater, B. Hartung, and D. Kersten. 2011. Visual motion and the perception of surface material. *Curr. Biol.* 21, 23 (2011), 2010–2016.

Katja Doerschner, Laurence T Maloney, and Huseyin Boyaci. 2010. Perceived glossiness in high dynamic range scenes. *Journal of Vision* 10, 9 (2010), 11–11.

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.

James A Ferwerda, Fabio Pellacini, and Donald P Greenberg. 2001. Psychophysically based model of surface gloss perception. In *Human vision and electronic imaging vi*, Vol. 4299. SPIE, 291–301.

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.

Graham D. Finlayson, Michal Mackiewicz, and Anya Hurlbert. 2015. Color Correction Using Root-Polynomial Regression. *IEEE Transactions on Image Processing* 24, 5 (may 2015), 1460–1470. https://doi.org/10.1109/TIP.2015.2405336

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.

Adria Fores, Mark D Fairchild, and Ingeborg Tastl. 2014. Improving the perceptual uniformity of a gloss space. In *Color and Imaging Conference*, Vol. 2014. Society for Imaging Science and Technology, 7–13.

Adria Fores, James Ferwerda, Ingeborg Tastl, and John Recker. 2013. Perceiving gloss in surfaces and images. In *Color and Imaging Conference*, Vol. 2013. Society for Imaging Science and Technology, 44–51.

Davit Gigilashvili, Weiqi Shi, Zeyu Wang, Marius Pedersen, Jon Yngve Hardeberg, and Holly Rushmeier. 2021. The role of subsurface scattering in glossiness perception. *ACM Transactions on Applied Perception (TAP)* 18, 3 (2021), 1–26.

Param Hanji, Fangcheng Zhong, and Rafał K Mantiuk. 2020. Noise-aware merging of high dynamic range image stacks without camera calibration. In *Computer Vision–ECCV 2020 Workshops: Glasgow, UK, August 23–28, 2020, Proceedings, Part III 16*. Springer, 376–391.

Bruce Hartung and Daniel Kersten. 2003. How does the perception of shape interact with the perception of shiny material? *Journal of Vision* 3, 9 (2003), 59–59.

R.F. Hess, F.A.A. Kingdom, and L.R. Ziegler. 1999. On the relationship between the spatial channels for luminance and disparity processing. *Vis. Res.* 39, 3 (1999), 559–568.

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.

AC Hurlbert, BG Cumming, and AJ Parker. 1991. Recognition and perceptual use of specular reflections. *Investigative Ophthalmology & Visual Science* 32 (1991), 105.

Aaron Isaksen, Leonard McMillan, and Steven J Gortler. 2000. Dynamically reparameterized light fields. In *Proceedings of the 27th annual conference on Computer graphics and interactive techniques.* 297–306.

Iona S Kerrigan and Wendy J Adams. 2013. Highlights, disparity, and perceived gloss with convex and concave surfaces. *Journal of Vision* 13, 1 (2013), 9–9.

J.J. Kulikowski. 1976. Effective contrast constancy and linearity of contrast sensation. *Vision Research* 16, 12 (jan 1976), 1419–1431. https://doi.org/10.1016/0042-6989(76)90161-9

Matthias Scheller Lichtenauer, Philipp Schuetz, and Peter Zolliker. 2013. Interaction improves perception of gloss. *Journal of Vision* 13, 14 (2013), 14–14.

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.

Rainer Mausfeld, Gunnar Wendt, and Jürgen Golz. 2014. Lustrous material appearances: Internal and external constraints on triggering conditions for binocular lustre. *i-Perception* 5, 1 (2014), 1–19.

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.

Daniele Menon, Stefano Andriani, and Giancarlo Calvagno. 2006. Demosaicing with directional filtering and a posteriori decision. *IEEE Transactions on Image Processing* 16, 1 (2006), 132–141.

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

Alexander A Muryy, Roland W Fleming, and Andrew E Welchman. 2016. 'Proto-rivalry': how the binocular brain identifies gloss. *Proceedings of the Royal Society B: Biological Sciences* 283, 1830 (2016), 20160383.

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

Edwin Olson. 2011. AprilTag: A robust and flexible visual fiducial system. In *2011 IEEE international conference on robotics and automation.* IEEE, 3400–3407.

Eli Peli, Jian Yang, Robert Goldstein, and Adam Reeves. 1991. Effect of luminance on suprathreshold contrast perception. *Journal of the Optical Society of America A* 8, 8 (aug 1991), 1352. https://doi.org/10.1364/JOSAA.8.001352

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.

Maria Perez-Ortiz and Rafal K Mantiuk. 2017. A practical guide and software for analysing pairwise comparison experiments. *arXiv preprint arXiv:1712.03686* (2017).

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.

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

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

Yuichi Sakano and Hiroshi Ando. 2012. Psychophysical evaluations of a current multi-view 3-D display: Its advantages in glossiness reproduction. *Journal of the Society for Information Display* 20, 5 (2012), 286–292.

Yuichi Sakano and Hiroshi Ando. 2021. Conditions of a Multi-View 3D Display for Accurate Reproduction of Perceived Glossiness. *IEEE Transactions on Visualization and Computer Graphics* 28, 10 (2021), 3336–3350.

A. C. Schmid, P. Barla, and K. Doerschner. 2020. Material category determined by specular reflection structure mediates the processing of image features for perceived gloss. *bioRxiv* (2020).

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).

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.

Krzysztof Templin, Piotr Didyk, Tobias Ritschel, Karol Myszkowski, and Hans-Peter Seidel. 2012. Highlight microdisparity for improved gloss depiction. *ACM Transactions on Graphics (TOG)* 31, 4 (2012), 1–5.

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.

Gunnar Wendt and Franz Faul. 2019. Differences in stereoscopic luster evoked by static and dynamic stimuli. *i-Perception* 10, 3 (2019), 2041669519846133.

Gunnar Wendt, Franz Faul, Vebjørn Ekroll, and Rainer Mausfeld. 2010. Disparity, motion, and color information improve gloss constancy performance. *Journal of vision* 10, 9 (2010), 7–7.

Gunnar Wendt, Franz Faul, and Rainer Mausfeld. 2008. Highlight disparity contributes to the authenticity and strength of perceived glossiness. *Journal of Vision* 8, 1 (2008), 14–14.

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

Weiye Xiao, Taoyong Tan, and Chengqi Xue. 2023. Gloss-Bridge: A Method to Reduce the Visual Perception Gap between Real-World Glossiness and PBR Reflectance Properties in Virtual Reality. *Applied Sciences* 13, 8 (2023), 4722.

Shoji Yamamoto, Masashi Sawabe, Mayu Yokoya, and Norimichi Tsumura. 2012. Enhancement of Gloss Perception by using Binocular Disparity. In *Conference on Colour in Graphics, Imaging, and Vision*, Vol. 2012. Society for Imaging Science and Technology, 226–230.

Akiko Yoshida, Matthias Ihrke, Rafał Mantiuk, and Hans-Peter Seidel. 2008. Brightness of the glare illusion. In *Proceedings of the 5th symposium on Applied perception in graphics and visualization.* 83–90.

Fan Zhang, Huib de Ridder, Pascal Barla, and Sylvia Pont. 2020. 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.
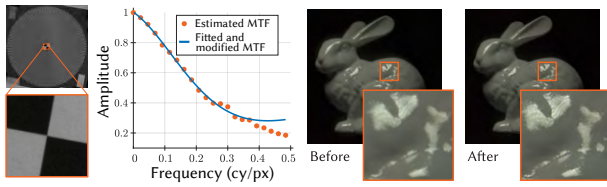
Figure 7: The deblurring procedure. The camera's MTF (in the plot) is determined using the photograph of the Siemens star chart (left) and the slanted edge technique. The MTF is used in deconvolution to restore higher frequencies, which are important for the gloss reproduction.
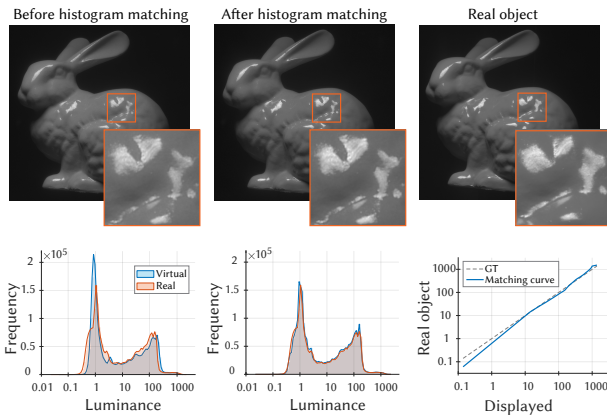

Figure 8: We show the luminance channel of the displayed image before (top left) and after (top middle) histogram matching and the luminance map of the real object (top right). The corresponding histograms are shown in the bottom left and bottom middle. The histogram matching curve is shown in the bottom right. Note that this figure shows luminance maps of images that are re-captured from the viewing position for the purpose of histogram matching, different from the source RGB light field.
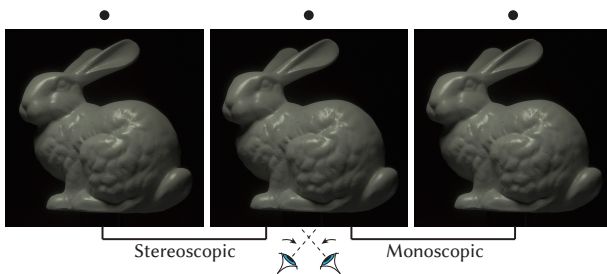
Figure 9: A demonstration of crossed fusion stereoscopic pairs of V50 stimulus, with stereoscopic on the left and monoscopic on the right. Refer to supplemental for more examples.
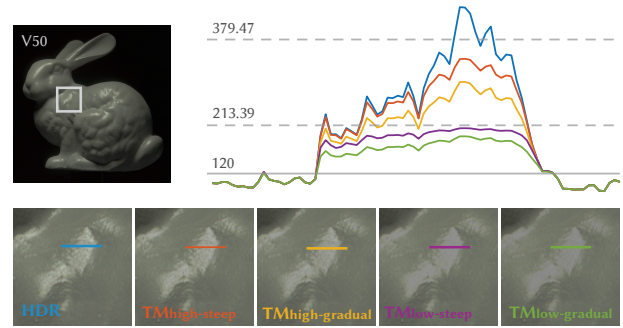
Figure 10: Insets of image in 50% gloss level (top left) after applying four tone mapping operators (bottom) and 1D scanlines in the highlight region (top right).
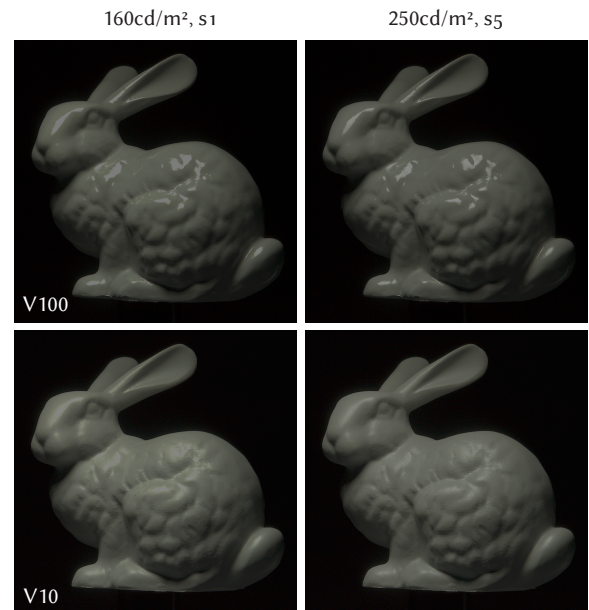
Figure 11: Demonstration of over-exposed/cut-off effect. We show tone mapped version of the most glossy stimulus V100 in the first row and the least glossy stimulus V10 in the second row. Similar gloss could be perceived from $[160 \text{ cd/m}^2, s_1]$ and $[250 \text{ cd/m}^2, s_5]$. Note that this figure is only for demonstration purposes without evidence from a user study. Zoom in for better comparison.
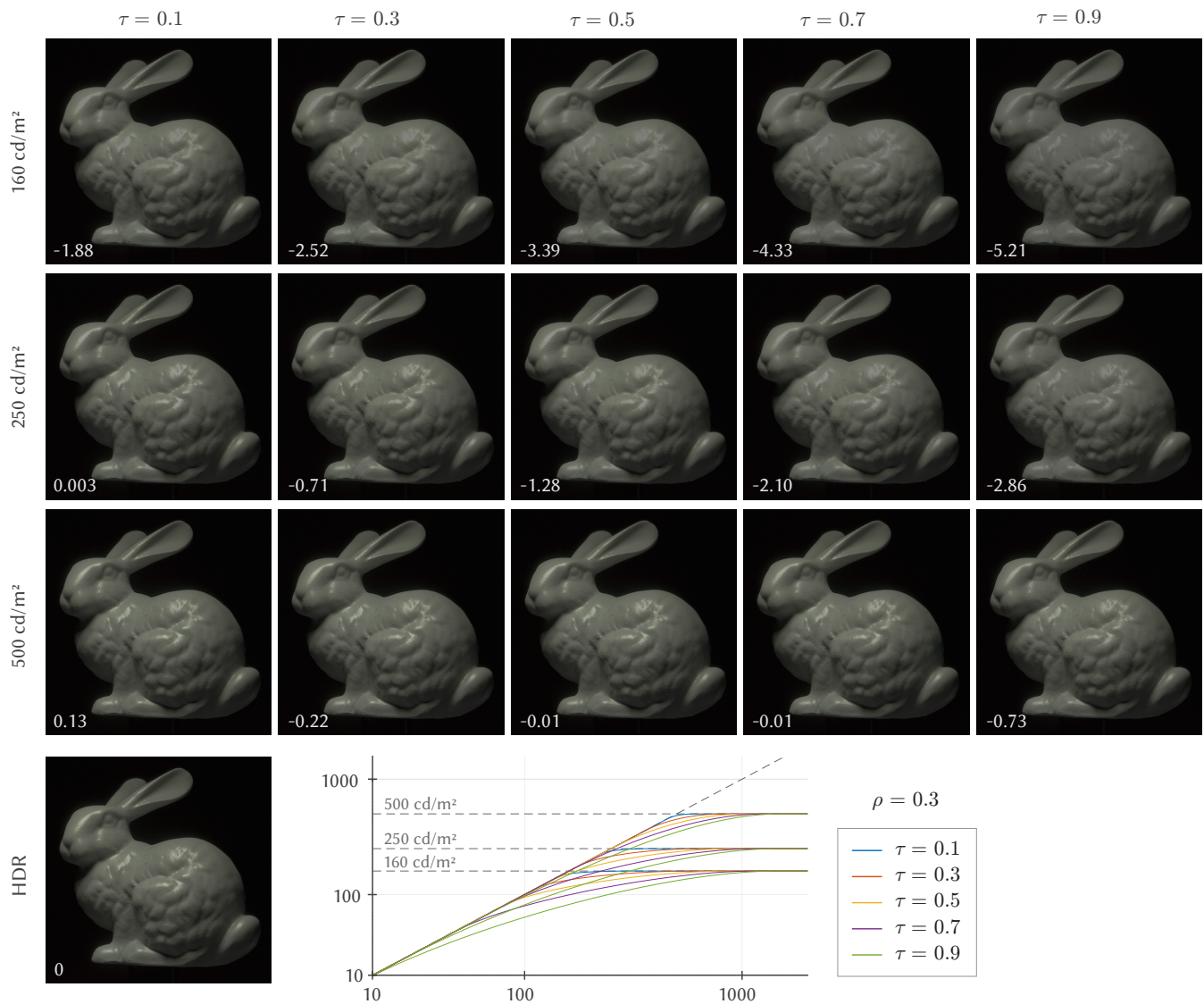
Figure 12: Stimuli used in our validation study. From first row to third row: 160 cd/m^2, 250 cd/m^2, 500 cd/m^2; from left to right: the steepness of tone mapping operators from high to low (τ = 0.1 to 0.9); The HDR image and all used tone mapping operators are shown in the last row. We show the perceived gloss difference related to HDR image in JND at bottom left of each stimulus (Fig. 6).