# Luminance-based specular gloss characterization

Frédéric B. Leloup, 1,2,\* Michael R. Pointer, Philip Dutré, and Peter Hanselaer, Philip Dutré, and Peter Hanselaer,

<sup>1</sup>Light & Lighting Laboratory, Catholic University College Sint-Lieven, B-9000, Gent, Belgium 
<sup>2</sup>Computer Graphics Research Group, K. U. Leuven, B-3001, Leuven, Belgium 
<sup>3</sup>Department of Colour Science, University of Leeds, Leeds LS2 9JT, UK 
<sup>4</sup>Department of Electrical Engineering (ESAT-ELECTA), K. U. Leuven, B-3001, Leuven, Belgium 
\*Corresponding author: Frederic.Leloup@kahosl.be

Received January 11, 2011; revised April 11, 2011; accepted April 18, 2011; posted April 19, 2011 (Doc. ID 140994); published May 31, 2011

Gloss is a feature of visual appearance that arises from the directionally selective reflection of light incident on a surface. Especially when a distinct reflected image is perceptible, the luminance distribution of the illumination scene above the sample can strongly influence the gloss perception. For this reason, industrial glossmeters do not provide a satisfactory gloss estimation of high-gloss surfaces. In this study, the influence of the conditions of illumination on specular gloss perception was examined through a magnitude estimation experiment in which 10 observers took part. A light booth with two light sources was utilized: the mirror image of only one source being visible in reflection by the observer. The luminance of both the reflected image and the adjacent sample surface could be independently varied by separate adjustment of the intensity of the two light sources. A psychophysical scaling function was derived, relating the visual gloss estimations to the measured luminance of both the reflected image and the off-specular sample background. The generalization error of the model was estimated through a validation experiment performed by 10 other observers. In result, a metric including both surface and illumination properties is provided. Based on this metric, improved gloss evaluation methods and instruments could be developed. © 2011 Optical Society of America

OCIS codes: 100.2960, 290.1483, 330.5020, 330.5510.

#### 1. INTRODUCTION

Besides color, gloss is a major characteristic that determines the appearance of an object. Physically, gloss results from the directionally selective reflectance properties of a surface, with a preference toward the specular reflection direction [1]. Although alternative measurement methods have been introduced [2,3], instrumental gloss evaluation is usually restricted to the measurement of this specular reflection by use of a glossmeter. This standardized instrument [4,5] measures the specularly reflected flux from the surface in a predefined geometry, typically at three angles of incidence (20°, 60°, and 85°). The apertures of the source and receiver are well defined. The measured flux is compared to the reflected flux from a working standard in the same geometry that in turn is calibrated against a primary standard. The primary standard consists of a polished black glass with refractive index n of 1.567 at a wavelength of 589.3 nm (sodium D line), and it is assigned a value of 100 gloss units irrespective of the angle of incidence.

The correlation between visual gloss (VG) appraisal and instrumental gloss characterization has been examined before. In an investigation of the perceptual gloss dimensions of a series of painted specimens, Billmeyer and O'Donnell reported that no simple linear relation existed between the obtained visual data and instrumental gloss measurements [6]. Two other studies confirmed this finding. Obein *et al.* compared observer judgments of the glossiness of 10 black, coated samples with instrumental gloss readings [7]. Ji *et al.* related visually scaled gloss of a set of 84 neutral and colored samples to measurement results obtained with a standard glossmeter [8]. Both found a similar nonlinear relationship, with higher

response at the extremes of the scale, i.e., for matte and high-gloss surfaces.

These findings suggest that VG perception is not only determined by the amount of specular reflection from the surface. Indeed, according to Hunter there are in fact six visual criteria that can be used to rank surfaces based on their gloss [9]. These are identified as

- Specular gloss—related to the amount of specular reflection from the surface
- Contrast gloss (CG)—associated with the observed contrast between the specularly reflecting areas and adjacent surface areas
- Distinctness-of-image (DOI) gloss—perceived distinctness and sharpness of the observed image in the specular reflection direction of the surface
- Absence-of-bloom gloss—related to the absence of haze or milky appearance adjacent to reflected highlights
- Sheen—gloss appraisal of low gloss or matte surfaces at grazing angles of incidence and viewing
- Surface-uniformity gloss—associated with the perceived surface smoothness and freedom from texture and markings

Hunter's conclusion formed the basis for an in-depth investigation of the influence of surface properties on gloss perception. In this respect, research results were recently reported on the interaction of surface gloss with texture [10], three-dimensional (3D) shape [11–14], and color [14–16].

An essential component in gloss formation that has received less attention is the geometry of illumination. Gloss investigations reported in the literature have generally been performed with samples being illuminated by only a single, fixed light source. On high-gloss surfaces, a highlight corresponding to the image of the source emerges in the specular reflection direction. The luminance of the source, the specular reflectance value of the material, and the image distortion (DOI) caused by the surface characteristics determine the luminance of this highlight. The luminance of the sample region that surrounds the specular highlight is determined by the illumination of the same source and the off-specular reflection characteristics of the sample. In a recent paper Leloup et al. [17] showed that for high-gloss samples, the contrast between the reflected image of the source and the sample background affects gloss perception, in accordance with Hunter's CG. Consequently, the glossiness of a sample could be drastically influenced by introducing a light source that is positioned out of the specular field of view of the observer.

A configuration with more than just one light source better corresponds to the complex illumination conditions that are encountered in daily life, with both direct and indirect indoor as well as outdoor contributions. With the aim of accurately rendering photorealistic images on displays, the computer graphics community was the first to examine the influence of a complex illumination geometry on the appearance properties of a material. Complex illumination scenes have been implemented by use of illumination maps that are acquired by photographically capturing the incident illumination from every direction at a specific point in the scene [18]. Fleming et al. demonstrated that observers estimate the glossiness of glossy surfaces more reliably under complex realistic illumination than under simple synthetic illumination, and that surfaces appear glossier under complex conditions [19]. Furthermore, Doerschner et al. introduced a glossiness transfer function that estimates the change in perceived gloss when moving from one illumination field to another [20]. Finally, in agreement with Doerschner et al., Olkkonen and Brainard measured a substantial effect on gloss perception when introducing geometric changes in the light field [21].

Yet, computer-based gloss research also suffers from some restrictions. The limited dynamic range of a display constitutes the main drawback. Pellacini et al. [22] stressed that digital imagery representations are not likely to allow for the investigation of any gloss attribute related to the absolute intensity of a highlight, even though methods have been introduced to map high dynamic range (HDR) images on standard dynamic range (SDR) displays [23,24]. With the introduction of HDR displays, Phillips et al. investigated the effect of image dynamic range on gloss perception [25,26]. In a gloss discrimination task, apparent gloss was judged both in HDR and tone-mapped SDR images. Important differences were observed; objects in HDR images were perceived as glossier and more discriminating than in their SDR counterparts. Doerschner et al. surmise that there are cues to surface material signaling glossiness present in HDR scenes that are absent or weak in scenes presented on SDR displays [27].

Industrial glossmeters do not provide proper estimations of the perceived gloss on high-gloss surfaces, because the geometry and intensity of the illumination are not taken into account. On the exception of the study reported by Leloup *et al.* [17], the influence of the applied illumination on gloss perception by use of realistic illumination scenes and real objects has not yet been addressed. In this paper, complex real-world

illumination conditions were simulated by use of an adjustable off-specular light source in addition to the specular source. Three achromatic glass samples with an identical DOI but a different diffuse backreflection were used in combination with 29 illumination conditions determined by the intensity of both light sources. A psychophysical experiment was conducted with 65 stimuli and 10 observers. The technique of magnitude estimation was used to derive a perceptual gloss scale that was related to the measured luminance of both the reflected image and the off-specular sample background. The performance of the psychophysical function was evaluated by a validation experiment with 22 new stimuli and 10 further observers. From the results, improved gloss evaluation methods and instruments could be developed based on luminance measurements in specular and off-specular directions.

# 2. EXPERIMENTAL SETUP

#### A. Light Booth

The psychophysical experiments were conducted in a light booth presented in Fig. 1, identical to the booth described by Leloup *et al.* [17]. A uniform rectangular light source is positioned 60 cm from the sample holder, with an incidence angle of 60° toward the sample normal. This source will be denoted as the specular light source. An additional ambient light source is positioned perpendicular to the sample, at a distance of 60 cm, and will further be referred to as the background light source. To obtain different illumination settings, the luminance of both light sources can be separately adjusted.

Samples are mounted on a fixed sample holder and observed in the mirror reflection direction of the specular light source. The observer's head is fixed by a chin rest, which guarantees both a well-defined viewing direction and a fixed viewing distance of 100 cm. Illumination and viewing distances are chosen such that the sample surface and the reflected image of the specular light source are both within the observer's depth of focus. A baffle prevents the observers from directly viewing the light sources.

# **B. Samples**

The sample set consists of three flat glass samples of  $10\,\mathrm{cm} \times 10\,\mathrm{cm}$  that are  $3\,\mathrm{mm}$  thick. The rear side of the samples is painted white, gray, and black, respectively. The samples were characterized with a HunterLab UltraScan PRO

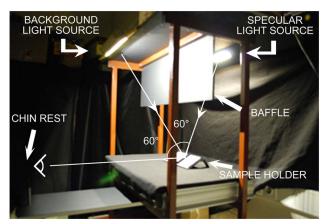


Fig. 1. (Color online) Side view of the test booth with the specular and background light sources.

Table 1. CIELAB Values (L\*a\*b\*) and Average Specular Gloss Values of the Three Samples<sup>a</sup>

| Sample Color | $\mathbf{L}^*$ | $\mathbf{a}^*$ | $\mathbf{b}^*$ | Gloss $20^{\circ}$ (SGU) | Gloss $60^{\circ}$ (SGU) | Gloss 85° (SGU) |
|--------------|----------------|----------------|----------------|--------------------------|--------------------------|-----------------|
| White        | 82.7           | -1.76          | 1.19           | $86.6 \pm 0.3$           | $93.4 \pm 0.2$           | $99.4 \pm 0.1$  |
| Gray         | 46.1           | -2.22          | -6.35          | $86.9 \pm 0.1$           | $92.5 \pm 0.1$           | $99.3 \pm 0.2$  |
| Black        | 3.2            | -0.05          | 0.08           | $85.2 \pm 1.0$           | $91.2 \pm 0.6$           | $99.2 \pm 0.3$  |

"CIELAB values (L\*a\*b\*) were obtained in SPEX mode and calculated under CIE Illuminant D65 in combination with the CIE 1964 standard colorimetric observer. Average specular gloss values were obtained in three basic geometries (20°, 60°, and 85°) and are expressed in SGUs.

sphere-based spectrophotometer, measuring the d: 8° spectral reflectance of the front side of the samples. This instrument offers the facility either to include or to exclude the light reflected in the specular direction. Two sets of measurements were performed: with the exclusion port opened [specular excluded (SPEX)] and closed [specular included (SPIN)], respectively. When subtracting the SPEX from the SPIN readout, a value of 4% was obtained for each sample, originating from the identical front surface reflection.

CIELAB colorimetric coordinates were calculated under CIE standard illuminant D65, in combination with the CIE 1964 standard colorimetric observer. SPEX mode results are presented in Table 1. Standard specular gloss measurements were performed at the three basic geometries (20°, 60°, and 85° angle of incidence) with a BYK-Gardner micro-TRI-gloss-S glossmeter. Measurements were repeated at five different positions on the surface to check uniformity. Mean values and variances, expressed in specular gloss units (SGUs), are reported in Table 1. Almost no differences are observed among the samples within one and the same geometry, indicating that front surface reflection completely dominates the measurement.

Spatial and spectral sample reflection characteristics were measured with a 3D bidirectional scatter distribution function instrument [28]. The angle of incidence with respect to the sample normal was fixed at 60°, while the viewing angle ranged from 0° to 85° in the opposite half-plane, 60° corresponding to the specular reflection direction (denoted as  $-60^{\circ}:60^{\circ}$  geometry). In Fig. 2, spectral bidirectional reflectance distribution function (BRDF) values  $q_{e,\lambda}$  are presented for viewing angles ranging from 35° to 85°. Solid curves represent the BRDF functions of the three glass samples, while the dashed curve represents the BRDF function of the black glass standard of the industrial glossmeter. All samples clearly show

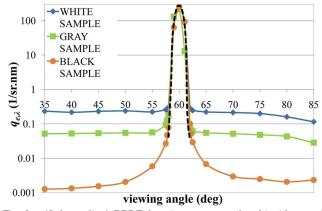


Fig. 2. (Color online) BRDF functions at an angle of incidence of  $-60^\circ$  and at wavelength  $589.3\,\mathrm{nm}$ . The viewing angle ranges from  $35^\circ$  to  $85^\circ$ . Solid curves represent the results of the three test samples, while the dashed curve represents the results of the black glass gloss reference sample.

a specular peak similar to the instrument signature, indicating a distortion-free reflected image and perfect DOI. At nonspecular viewing angles, BRDF values tend to a constant value originating from the background reflection of the samples.

# 3. PSYCHOPHYSICAL EXPERIMENT

In order to derive a perceptual gloss scale that takes into account the influence of the applied illumination, a psychophysical experiment was conducted using the method of magnitude estimation. Magnitude estimation is one of the most frequently used psychophysical scaling techniques [29], and it has already been successfully applied in previous research on surface gloss perception [8,22,30]. The technique is designed to discover functional relationships between the physical properties of a presented stimulus and its perceptual attributes [31].

Combinations of three samples and 29 illumination settings resulted in 65 stimuli, which were presented to 10 observers. The illumination settings were generated by separate adjustment of the intensity of both the specular and background light sources. Pictures of the gray sample under four illumination settings are presented in Fig. 3. In marked contrast with the off-specular sample background, the reflection image of the specular light source can be clearly discerned.

For all stimuli, luminance measurements were performed from the viewing position of the observer with a two-dimensional luminance camera (MURATest by Eldim). From these measurements, the average luminance of the image of the specular light source, denoted by the image luminance  $L_{\rm im}$ , and of both sample regions next to this image, denoted by the background luminance  $L_b$ , were calculated.  $L_{\rm im}$  ranged from 0 to  $300~{\rm cd.m^{-2}}$ , while  $L_b$  ranged from 0 to  $55~{\rm cd.m^{-2}}$ . A graphic representation of the stimuli in an  $L_b$  versus  $L_{\rm im}$  graph is presented in Fig. 4.

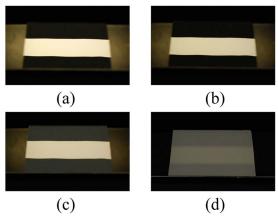


Fig. 3. (Color online) Pictures of the gray sample under four illumination settings. The image of the specular light source can be clearly discerned.

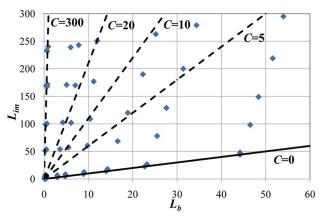


Fig. 4. (Color online) Graphic representation of the stimuli in an  $L_b$  versus  $L_{\rm im}$  graph. All data points lie above the straight line on which  $L_{\rm im}$  and  $L_b$  are equal (zero contrast). Additional dashed lines of constant psychometric contrast C are also represented.

 $L_{\rm im}$  is made up of two contributions: a first contribution originating from specular reflection at the front surface and a second contribution due to diffuse reflection from the rear side of the sample. Assuming a uniform illumination of the sample,  $L_{\rm im}$  can be written as

$$L_{\rm im} = \rho_s L_{sp} + L_b, \tag{1}$$

where  $\rho_s$  and  $L_{sp}$ , respectively, denote the specular reflectance of the sample and the luminance of the specular light source. From Eq. (1), one can conclude that  $L_{\rm im}$  is always larger than or equal to  $L_b$ . This can also be observed in Fig. 4; all data points lie above the straight line on which  $L_{\rm im}$  and  $L_b$  are equal.

Observers were presented with a standard stimulus at the beginning of the experiment and told to assign a modulus value of 100 to it. The choice of the standard stimulus can have a marked influence on the outcome of a magnitude estimation experiment. Preferentially, the modulus should lie in the middle of the range, such that the slope for variables larger than the standard is the same as the slope for variables lower than the standard [32,33]. Former results have demonstrated that within one illumination condition the psychometric contrast C, defined by CIE [34] as

$$C = \frac{|L_{\rm im} - L_b|}{L_b},\tag{2}$$

roughly correlates with VG appraisal [17]. Following this indication, C was calculated for every stimulus. Neglecting the eight stimuli for which C equals 0 (specular light source turned off),  $\log C$  values of the presented stimuli ranged from -1.07 to 2.59, and the stimulus with a  $\log C$  value of 0.88 was

chosen as the standard. Corresponding values of  $L_{\rm im}$  and  $L_b$  were 190 cd.m<sup>-2</sup> and 22 cd.m<sup>-2</sup>, respectively. Some lines of constant psychometric contrast C are also represented in Fig. 4.

The remaining stimuli were presented in sequence to the observers. They were asked to rate the stimuli on a scale where zero represents no gloss and such that the ratio of ratings corresponds to the ratio of perceived glossiness to the standard stimulus. As the observers' glossiness appraisal cannot be predicted in advance, observers were not given an upper limit (open-ended scale) and were allowed to use any number they thought to be appropriate. The standard stimulus was presented for a second time at the end of each session, and the ratings for this second presentation were compared to the modulus. All 10 observers had normal or corrected-to-normal vision, and were naive with respect to the purpose of the experiment. They performed the test, which lasted approximately 20 min, in two different random orders of presentation.

#### A. Data Analysis

The absolute magnitude estimation results were analyzed using a method that was previously employed by Ji et~al. [8] and Luo et~al. [35]. For each stimulus s, the geometric mean  $\hat{S}_s$  is computed from the raw observer data  $S_{s,o}$  of all observers o. The results of each individual observer are normalized against the geometric mean by calculating the scaling factor  $a_o$  and line intercept  $b_o$  from Eq. (3) using the least-squares fitting method:

$$\log \hat{S}_s = a_o \log S_{s,o} + b_o. \tag{3}$$

Afterward, both coefficients  $a_o$  and  $b_o$  are used to calculate the normalized estimation values  $S_{s,o,\mathrm{norm}}$  of each observer for each stimulus:

$$S_{s.o.\text{norm}} = 10^{(\log \hat{S}_s - b_o)/a_o}.$$
 (4)

The observer repeatability and agreement are evaluated with the coefficient of variation CV, defined as

$$CV = \frac{100}{\bar{y}} \sqrt{\frac{\sum_{s=1}^{n} (x_s - y_s)^2}{n}}.$$
 (5)

 $x_s$  and  $y_s$  both represent a dataset,  $\bar{y}$  being the mean value of dataset  $y_s$ . n is the number of evaluated stimuli, i.e., 65.

# **B.** Results

Values of both the scaling factor  $a_o$  and the line intercept  $b_o$  are reported for each individual observer in Table 2. For a perfect agreement between a set of observer data and the

Table 2. Values of Both the Scaling Factor  $a_o$  and the Line Intercept  $b_o$  for All Sets of Observer Data

| Scaling Factor $a_o$ Observer | 1     | 2     | 3      | 4     | 5      | 6     | 7     | 8     | 9     | 10    |
|-------------------------------|-------|-------|--------|-------|--------|-------|-------|-------|-------|-------|
| Presentation order 1          | 0.925 | 0.920 | 1.065  | 0.885 | 1.190  | 0.900 | 0.941 | 0.863 | 0.900 | 0.869 |
| Presentation order 2          | 0.967 | 0.900 | 0.922  | 0.880 | 1.064  | 0.890 | 0.945 | 0.794 | 0.907 | 0.904 |
| Line Intercept $b_o$ Observer | 1     | 2     | 3      | 4     | 5      | 6     | 7     | 8     | 9     | 10    |
| Presentation order 1          | 0.148 | 0.135 | -0.316 | 0.124 | -0.400 | 0.030 | 0.081 | 0.461 | 0.250 | 0.240 |
| Presentation order 2          | 0.102 | 0.230 | 0.023  | 0.120 | -0.130 | 0.080 | 0.053 | 0.503 | 0.157 | 0.161 |

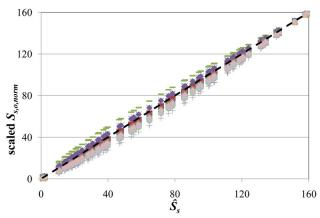


Fig. 5. (Color online) 20 sets of normalized observer data  $S_{s,o,nom}$ , scaled to the maximum geometric mean value, and plotted against the geometric mean data  $\hat{S}_s$ .

set of geometric mean data,  $a_o$  would equal unity and  $b_o$  zero. Values of the scaling factor  $a_o$  range from 0.794 to 1.190, meaning that a variation in magnitude of almost 40% is encountered. All values of the line intercept  $b_o$  are close to zero. The presentation order does not have a significant influence.

The 20 sets of normalized observer data  $S_{s,o,\text{norm}}$  have been scaled to the maximum geometric mean value. These scaled values are plotted against the geometric mean data  $\hat{S}_s$  in Fig. 5. The geometric mean values of the 65 stimuli range from 1 to 160, with the modulus numbering 100.

The observed scatter in Fig. 5 provides a first visual impression of the agreement among observers. Yet, for a quantitative estimation of the interobserver agreement, CV as defined in Eq. (5) is utilized. Each normalized set of observer data  $S_{s.o.norm}$  is compared to the set of geometric mean data  $\hat{S}_s$ . CV values for each observer are reported in Table 3, and range from 4% to 13% with an average of 7%.

The values of CV for two sets of data of each individual observer (observer repeatability) are also gathered in Table 3. Again, values are small and range from 0% to 9%, with an average of 3%. This value is comparable to the value of 4%

reported by Ji  $et\ al.$  for seven observers scaling surface gloss [8].

# 4. PSYCHOPHYSICAL SCALING FUNCTION

Several psychophysical scaling functions relating the geometric mean data  $\hat{S}_s$  to the measured luminance of both the reflected image  $L_{\rm im}$  and the off-specular sample background  $L_b$  were tested with least-squares fitting techniques. Both multiple linear and nonlinear regression analyses were performed. For each solution, the goodness of fit was evaluated with the adjusted coefficient of determination  $R_{\rm adj}^2$ . Normality and constant variance assumptions were always tested and solutions were rejected if the probability value p for one or both assumptions was inferior to 5%.

An overview of the principal hypothetic functions and corresponding regression results is presented in Table 4. Psychometric contrast as defined in Eq. (2) provides a poor fit ( $R_{\rm adj}^2=0.523$ ), in accordance with the conclusions of Pellacini and co-workers [22,30]. Moreover the constant variance assumption is not met. Formulas based on alternative contrast definitions, such as the Michelson contrast, were also tested. Again, the results show a very poor fit ( $R_{\rm adj}^2=0.45$ ), and both the normality and constant variance assumption are not met.

A better VG prediction is obtained with a formula based on the difference between  $L_{\text{im}}$  and  $L_b$ :

$$VG = 15(L_{im} - L_b)^{0.4}. (6)$$

The adjusted coefficient of variation  $R_{\rm adj}^2$  numbers 0.945. However, the constant variance assumption is again not met. The importance of satisfying this assumption becomes evident from Fig. 6, where the VG predictions as defined in Eq. (6) are plotted against the corresponding geometric mean data  $\hat{S}_s$  for all presented stimuli. For perfect agreement, all data points should lie on a straight line. It is clear that stimuli with identical differences between the luminance of the image and the background are however assessed as having a different gloss.

Table 3. Evaluation of Both the Observer Agreement and Repeatability by Calculation of the Coefficient of Variation  $\mathrm{CV}^a$ 

|  | Observer         | 1   | 2      | 3      | 4 | 5      | 6      | 7   | 8       | 9      | 10  | Mean   |
|--|------------------|-----|--------|--------|---|--------|--------|-----|---------|--------|-----|--------|
| Observer agreement<br>Observer repeatability | CV (%)<br>CV (%) | 4 3 | 6<br>2 | 4<br>9 | 8 | 7<br>6 | 7<br>1 | 4 0 | 13<br>6 | 7<br>1 | 8 3 | 7<br>3 |

<sup>&</sup>quot;For the observer agreement, each individual set of data  $S_{s,o,\text{norm}}$  was compared to the set of geometric mean data  $\hat{S}_s$ . For observer repeatability, two sets of data of the same observer were compared.

Table 4. Overview of Hypothetic Functions Describing the Relationship between the Geometric Mean Data  $\hat{S}_s$  and Both the Luminance  $L_{\rm im}$  and  $L_b{}^a$ 

|   |                | Regi  | ression Re | sults |                | Normality | Assumption | Constant Variance Assumption |          |  |  |
|---|----------------|-------|------------|-------|----------------|-----------|------------|------------------------------|----------|--|--|
| Function  | $\overline{a}$ | b     | e1         | e2    | $R_{ m adj}^2$ |           | p          |                              | p        |  |  |
| $a(rac{L_{	ext{im}}-L_b}{L_b})^{e1} \ a(L_{	ext{im}}-L_b)^{e1}$  | 45.89          | _     | 0.180      | _     | 0.523          | Pass      | 0.0676     | Fail                         | 0.0144   |  |  |
| $a(L_{\mathrm{im}}^{L_b}-L_b)^{e1}$                               | 15.09          | _     | 0.404      | _     | 0.945          | Pass      | 0.2883     | Fail                         | < 0.0001 |  |  |
| $aL_{\rm im}^{e1} - bL_{b}^{e2}$                                  | 21.66          | 17.18 | 0.379      | 0.354 | 0.963          | Pass      | 0.6400     | Pass                         | 0.3053   |  |  |
| $aL_{ m im}^{e1}$ – $bL_b^{e2}$ $aL_{ m im}^{1/3}$ – $bL_b^{1/3}$ | 28.11          | 20.59 | _          | _     | 0.958          | Pass      | 0.1635     | Pass                         | 0.6057   |  |  |

 $<sup>^{</sup>a}$ The values of the regression coefficients and the adjusted coefficient of determination  $R_{\rm adj}^{2}$  are presented. Results of the normality and constant variance assumptions are indicated based on the reported probability value v.

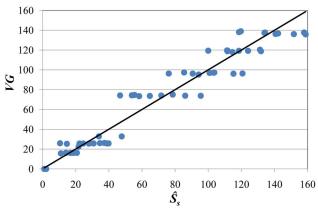


Fig. 6. (Color online) VG predictions as defined in Eq. (6), plotted against the corresponding geometric mean data  $\hat{S}_s$ .

The highest adjusted coefficient of determination  $R_{\rm adj}^2$  (0.963) was obtained with the following relationship:

$$VG = 22L_{\rm im}^{0.379} - 17L_b^{0.354}. (7)$$

Both regression assumptions are now met. The graphic representation of this function, plotted against the corresponding geometric mean data  $\hat{S}_s$ , is presented in Fig. 7.

Interestingly, VG as defined in Eq. (7) is similar to the brightness function put forward by Bodmann *et al.* [36], defined as

$$B = aL^p - B_0. (8)$$

B denotes the estimated brightness of an achromatic stimulus of luminance L. The term  $B_0$  accounts for simultaneous contrast effects when the test stimulus is viewed against a background of nonzero luminance, and was previously introduced by Jameson and Hurvich [37] and Hurvich [38]. The exponent p has a value of approximately 1/3, which is very close to the exponents reported in Eq. (7).

A power function of luminance with an exponent of 1/3 is also encountered in the CIELAB lightness L\*. Moreover, Pellacini and co-workers [22,30] introduced a CG function, formulated in terms of a cubic root function of both the specular reflectance  $\rho_s$  and the diffuse reflectance  $\rho_d$ :

$$CG = \left(\rho_s + \frac{\rho_d}{2}\right)^{1/3} - \left(\frac{\rho_d}{2}\right)^{1/3}.$$
 (9)

These findings suggest that in Eq. (7) a fixed value of 1/3 for both exponents could also involve a good solution. The result for this regression analysis is also included in Table 4:

$$VG = 28L_{\rm im}^{1/3} - 21L_b^{1/3}.$$
 (10)

The value of  $R_{\rm adj}^2$  numbers 0.958, and both regression assumptions are met.

By substitution of Eq. (1) in Eq. (10), VG can be rewritten as

$$VG = 28 \left( \rho_s L_{sp} + L_b \right)^{1/3} - 21 L_b^{1/3}.$$
 (11)

The background luminance  ${\cal L}_b$  can be correlated to the sample illuminance  ${\cal E}$ :

$$L_b = \frac{\rho_d}{\pi} E. \tag{12}$$

If the background light source is turned off, the illuminance E is proportional to the luminance of the specular light source  $L_{sp}$ . Equation (12) can then be rewritten as

$$L_b = \frac{\rho_d}{\pi} \left( \alpha L_{sp} \right), \tag{13}$$

with  $\alpha$  being a proportionality factor determined by the mutual position of the source and the sample.

By substitution of Eq. (13) into Eq. (11), VG can finally be written as

$$VG = 28L_{sp}^{1/3} [(\rho_s + c_1 \rho_d)^{1/3} - c_2 \rho_d^{1/3}], \tag{14}$$

with  $c_1$  and  $c_2$  being two constants.

The resemblance between Eq. (9) (CG) and Eq. (14) (VG) is striking. Moreover, it can be concluded from Eq. (14) that the gloss perception increases with the cubic root of the luminance of the light source and accordingly with the level of illumination, indicating an incomplete luminance adaptation. Equation (9) does not take into account this influence. Indeed, Pellacini and co-workers [22,30] applied only one level of illumination in their experiments.

Variations in the intensity of illumination also change the dynamic range of the scene. Doerschner *et al.* demonstrated that CG and specular gloss are affected by HDR cues such as the self-luminosity threshold, i.e., the degree to which highlights are seen as self-luminous [27]. While Doerschner *et al.* modified the dynamic range by successively introducing a black and a white background, in this study, the scene dynamic range was adapted by changing the intensity of both the specular and background light source.

The cubic root of luminance in Eqs. (10) and (14) predicts a compression at the high end of the scale. This compressive behavior was also reported by Phillips *et al.* [25,26] under the conditions of simple surface geometry and simple high-contrast illumination. Phillips *et al.* indicated Weber's law constraints on luminance just noticeable differences as a possible explanation and suggested that there may be limits to the effectiveness of specular intensity as a gloss cue [25,26].

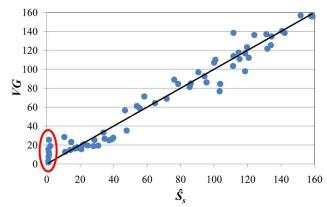


Fig. 7. (Color online) VG predictions as defined in Eq. (7), plotted against the corresponding geometric mean data  $\hat{S}_s$ . The specific test conditions for which the specular light source is turned off are indicated by the red ellipse.

| Scaling Factor $a_o$ Observer                | 1                | 2              | 3               | 4                | 5               | 6              | 7              | 8               | 9              | 10             |
|--|------------------|----------------|-----------------|------------------|-----------------|----------------|----------------|-----------------|----------------|----------------|
| Presentation order 1<br>Presentation order 2 | $0.560 \\ 0.721$ | 0.729<br>0.709 | 1.011<br>0.875  | 1.051<br>1.289   | 0.857<br>1.039  | 0.574<br>0.830 | 0.690<br>0.866 | 0.927 $1.072$   | 0.724<br>0.645 | 0.623<br>0.719 |
| Line Intercept                               | 1                | 2              | 3               | 4                | 5               | 6              | 7              | 8               | 9              | 10             |
| Presentation order 1 Presentation order 2    | 0.944<br>0.658   | 0.459<br>0.522 | -0.059<br>0.211 | -0.087<br>-0.561 | 0.128<br>-0.283 | 0.849<br>0.328 | 0.643<br>0.227 | 0.036<br>-0.213 | 0.545<br>0.728 | 0.707<br>0.554 |

Table 5. Values of Both the Scaling Factor  $a_o$  and the Line Intercept  $b_o$ , for the 20 Sets of Observer Data of the Validation Experiment

Finally, it is interesting to note that the experimental results are also in line with the variation of the perceived brightness Q with the luminance of an unrelated color as predicted by the Hunt color appearance model for unrelated colors [39]. Indeed, to a certain extent the specular image can be considered as an unrelated color surrounded by a rather dark background.

One has to be aware that the values of the numerical parameters occurring in Eqs. (7) and (10) are completely determined by the choice of the standard stimulus, which was given a value of 100. This illumination condition corresponds to an illuminance level of approximately 150 lux at the sample and the corresponding values of  $L_{\rm im}$  and  $L_b$  were  $190\,{\rm cd.m^{-2}}$ and 22 cd.m<sup>-2</sup>, respectively. As mentioned in the psychophysical experiment section, these values were chosen such that the standard stimulus would lie in the middle of the range of presented stimuli. The selection of a standard sample and a standard illumination condition for background illuminance and specular source luminance should be a future issue for debate. Indeed, the available standard practice for visual evaluation of gloss differences only defines two light sources that can be used to evaluate one or more available gloss attributes [40]. However, no recommendations on the relative distances between the light source, specimen, and observer are provided, and thus also not on the applicable illuminance level.

The VG prediction formulas as defined in Eqs. (7) and (10) suffer however from a major shortcoming. If the specular light source is turned off,  $L_{\rm im}$  equals  $L_b$  and no reflected image can be observed. Although looking at a mirrorlike surface, all observers reported gloss values close to 0 in this situation, for any value of  $L_b$  ranging between 0 to 55 cd.m<sup>-2</sup>. The results corresponding to this condition have been indicated in Fig. 7 by the red ellipse. As can be seen, VG predictions run up to 30.

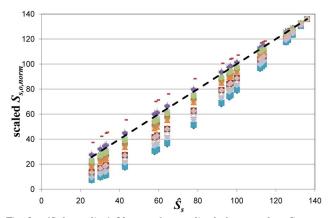


Fig. 8. (Color online) 20 sets of normalized observer data  $S_{s.o.\text{norm}}$ , scaled to the maximum geometric mean value, and plotted against the calculated geometric mean data  $\hat{S}_s$  of the validation experiment.

Equations (7) and (10) therefore only seem to be valid if the reflected image of the specular source is clearly visible. Under these circumstances, observers are judging the brightness of the image as an indication of gloss, while for very low psychometric contrast values, including C=0, they will probably judge the perceptibility of the reflected image. Because not enough experimental data are available for psychometric contrast values lower than 1, Eqs. (7) and (10) will not describe the gloss perception in this contrast region, which needs further investigation.

### 5. VALIDATION EXPERIMENT

The performance of a model is related to its prediction capability on independent test data. Assessment of this performance provides a measure of the quality of the chosen model [41]. To this end, a validation experiment was conducted with 10 new observers. They were asked to rate the glossiness of 22 new stimuli, made up of combinations of the same three samples with 22 other illumination settings than those previously applied in the main experiment. Again, luminance measurements were performed from the observer's viewing position, and both the average luminance  $L_{\rm im}$ and  $L_b$  were calculated.  $L_{\rm im}$  and  $L_b$  varied now from 0 to 240 cd.m<sup>-2</sup> and 0 to 40 cd.m<sup>-2</sup>, respectively, and the same standard stimulus was adopted ( $\log C = 0.88$ ). Observers were given the same instructions as in the main experiment, and the same procedure and data gathering technique were used. All observers had normal or corrected-to-normal vision and were again unaware of the purpose of the experiment. As in the main experiment, observers performed the test in two different orders of presentation.

For each of the 22 stimuli, VG predictions calculated from Eq. (10) were compared to the geometric mean  $\hat{S}_s$  determined from all reported estimations. The performance of the model was evaluated by calculation of the mean absolute error (MAE). The MAE determines the average magnitude of errors in a set of forecasts without considering their direction, and is defined as

$$MAE = \frac{1}{n} \sum_{s=1}^{n} |VG - S_s|, \qquad (15)$$

where n is the number of evaluated stimuli, i.e., 22.

**Validation Results:** Values of the normalization coefficients  $a_o$  and  $b_o$ , for each of the 20 sets of observer data, are reported in Table 5. In comparison to the main experiment (Table 2), values of both coefficients show larger deviations from 1 and 0, respectively. The larger variation leads to substantially more scatter in the normalized observer data  $S_{s,o,norm}$ , that are plotted against the calculated geometric

Table 6. Evaluation of the Observer Agreement and Repeatability in the Validation Experiment by Calculation of the Average Value of the Coefficient of Variation CV

|                        | Observer | 1  | 2  | 3 | 4 | 5 | 6  | 7  | 8 | 9  | 10 | Mean |
|------------------------|----------|----|----|---|---|---|----|----|---|----|----|------|
| Observer agreement     | CV (%)   | 21 | 15 | 3 | 6 | 4 | 17 | 12 | 3 | 17 | 19 | 12   |
| Observer repeatability | CV (%)   | 12 | 1  | 6 | 8 | 8 | 17 | 10 | 6 | 5  | 7  | 8    |

<sup>a</sup>For the observer agreement, each individual set of data  $S_{s.o.norm}$  was compared to the set of geometric mean data  $\hat{S}_s$ . For the observer repeatability, two sets of data of the same observer were compared.

mean data  $\hat{S}_s$  in Fig. 8. The geometric mean values of the 22 stimuli range from 25 to 140, the value of the common modulus numbering 100. CV values estimating both the observer agreement and repeatability are reported in Table 6 and confirm the augmented variation. For the observer agreement, average CV values range between 3% to 21%, the average numbering 12%. This value is larger than the average CV value reported in the main experiment (see Table 3). Yet, it is still significantly lower than the CV of 18%, found by Luo et al. for the agreement among observers who scaled colorfulness [35]. With respect to observer repeatability, CV values now range from 1% to 17%, with an average of 8%. The resulting MAE for VG as defined in Eq. (10) numbers 8.2, corresponding to 6.4% of the absolute range of estimation values. As a consequence, the VG function as defined in Eq. (10) can be considered as validated by this experiment.

#### 6. CONCLUSIONS

In this study, the influence of the geometry of illumination on the gloss perception of surfaces has been examined. Three flat glass samples having an identical specular reflectance and showing an undistorted reflected image were utilized. The reverse side of the samples was covered with a neutral paint, varying in lightness from white over gray to black. Realistic illumination conditions were simulated with a light booth containing two light sources. The specular light source induces a visible virtual image, while the background light source is used to change the overall sample luminance. The luminance of both light sources can be separately adjusted.

A visual experiment was conducted using the technique of absolute magnitude estimation. Ten observers rated the glossiness of 65 stimuli in two different orders of presentation. Both the observer agreement and repeatability were satisfactorily high: the average coefficient of variation CV, respectively, numbered 7% and 3%. A psychophysical scaling function was derived that predicts the perceived surface gloss from the measured luminance of both the reflected image and the adjacent off-specular sample background. A good VG prediction was obtained from the difference between the cubic root of both luminances, at least when the psychometric contrast values are larger than 1.

The performance of this formula was tested in an additional validation experiment. Ten new observers rated the glossiness of 22 stimuli, different from the applied stimuli in the main experiment, again in two different orders of presentation. Although both the observer agreement and repeatability were now lower, the MAE between the gloss predictions and the observer assessments was limited to 6.4% of the absolute

range of estimation values. This suggests that a measurement of the luminance of both the specular image and the surrounding background allows for the evaluation of the glossiness of the samples under different illumination conditions. Indeed, with a metric based on luminance measurements, both the reflection properties of the surface and the geometry and intensity of the illumination are taken into account. The gloss prediction formula can however not be used in the condition that no specular image can be observed (zero contrast). When the contrast between the specular image and background becomes very low (e.g., when a white sample is used in combination with an intense background illumination), observers are probably judging the perceptibility of the image as an indication of gloss, while for high contrast values, they will judge the brightness of the image.

This study is only a small step in the development of a new gloss perception correlate. In addition to investigating gloss perception at low contrast ranges, the impact of the distortion of the reflected image (DOI), the influence of the color of the sample, and the effect of the nonuniformity of the image light source must be included. Finally, investigations using real complex illumination and environment scenes combined with 3D samples could be performed, as the shape and curvature of 3D samples will transform the virtual image of the highlight and give rise to additional cues that observers could exploit when evaluating gloss [42,43].

# REFERENCES

- International Lighting Vocabulary CIE Publication 17.4, ISBN: 3900734070 (IEC/CIE, 1987).
- American Society for Testing and Materials "Standard test methods for measurement of gloss of high-gloss surfaces by goniophotometry," Test Method E430-05 (ASTM, 2005).
- American Society for Testing and Materials, "Standard test methods for instrumental measurement of distinctness-of-image gloss of coating surfaces," Test Method D5767 (ASTM, 2004).
- American Society for Testing and Materials, "Standard test method for specular gloss," Test Method D523 (ASTM, 2008).
- 5. International Organization for Standardization, "Paints and varnishes—determination of specular gloss of non-metallic paint films at 20 degrees, 60 degrees and 85 degrees," Standard 2813 (ISO, 1994).
- F. W. Billmeyer and F. X. D. O'Donnell, "Visual gloss scaling and multidimensional scaling analysis of painted specimens," Color Res. Appl. 12, 315–326 (1987).
- G. Obein, K. Knoblauch, and F. Viénot, "Difference scaling of gloss: nonlinearity, binocularity, and constancy," J. Vis. 4 (9), 711–720 (2004).
- 8. W. Ji, M. R. Pointer, R. M. Luo, and J. Dakin, "Gloss as an aspect of the measurement of appearance," J. Opt. Soc. Am. A 23, 22–33 (2006).
- R. S. Hunter, "Methods of determining gloss," National Bureau of Standards Research Paper RP958, J. Res. Natl. Bur. Stand. 18, 19–39 (1937).
- Y.-H. Ho, M. S. Landy, and L. T. Maloney, "Conjoint measurement of gloss and surface texture," Psychol. Sci. 19, 196–204 (2008).
- 11. R. W. Fleming, A. Torralba, and E. H. Adelson, "Specular reflection and the perception of shape," J. Vis. 4 (9), 790–820 (2004).
- J. F. Norman, J. T. Todd, and G. A. Orban, "Perception of three-dimensional shape from specular highlights, deformations of shading, and other types of visual information," Psychol. Sci. 15, 565–570 (2004).
- P. Vangorp, J. Laurijssen, and P. Dutré, "The influence of shape on the perception of material reflectance," ACM Trans. Graph. 26, 77 (2007).
- G. Wendt, F. Faul, V. Ekroll, and R. Mausfeld, "Disparity, motion, and color information improve gloss constancy performance," J. Vis. 10 (9), 7 (2010).

- B. Xiao and D. H. Brainard, "Surface gloss and color perception of 3D objects," Vis. Neurosci. 25, 371–385 (2008).
- M. Mikula, M. Ceppan, and K. Vasco, "Gloss and goniocolorimetry of printed materials," Color Res. Appl. 28, 335–342 (2003).
- F. B. Leloup, M. R. Pointer, P. Dutré, and P. Hanselaer, "Geometry of illumination, luminance contrast, and gloss perception," J. Opt. Soc. Am. A 27, 2046–2054 (2010).
- R. O. Dror, A. S. Willsky, and E. H. Adelson, "Statistical characterization of real world illumination," J. Vis. 4 (9), 821–837 (2004).
- R. W. Fleming, R. O. Dror, and E. H. Adelson, "Real-world illumination and the perception of surface reflectance properties," J. Vis. 3 (5), 347–368 (2003).
- K. Doerschner, H. Boyaci, and L. T. Maloney, "Estimating the glossiness transfer function induced by illumination change and testing its transitivity," J. Vis. 10 (4), 8 (2010).
- 21. M. Olkkonen and D. H. Brainard, "Perceived glossiness and lightness under real-world illumination," J. Vis. 10 (9), 5 (2010).
- F. Pellacini, J. A. Ferwerda, and D. P. Greenberg, "Toward a psychophysically-based light reflection model for image synthesis," in *Proceedings of SIGGRAPH'00* (Association for Computing Machinery, 2000), pp. 55–64.
- J. Tumblin, J. K. Hodgins, and B. K. Guenter, "Two methods for display of high contrast images," ACM Trans. Graphics 18, 56– 94 (1999).
- G. W. Larson, H. Rushmeier, and C. Piatko, "A visibility matching tone reproduction operator for high dynamic range scenes," IEEE Trans. Vis. Comput. Graph. 3, 291–306 (1997).
- J. B. Phillips, J. A. Ferwerda, and S. Luka, "Effects of image dynamic range on apparent surface gloss," in *Proceedings of the 17th Color Imaging Conference: Color Science and Engineering Systems, Technologies, and Applications* (Society for Imaging Sciences and Technology, 2010), pp. 193–197.
- J. B. Phillips, J. A. Ferwerda, and A. Nunziata, "Gloss discrimination and eye movement," Proc. SPIE 7527, 75270Z (2010).
- K. Doerschner, L. T. Maloney, and H. Boyaci, "Perceived glossiness in high dynamic range scenes," J. Vis. 10 (9), 11 (2010).
- F. B. Leloup, S. Forment, P. Dutré, M. R. Pointer, and P. Hanselaer, "Design of an instrument for measuring the spectral

- bidirectional scatter distribution function," Appl. Opt. 47, 5454–5467 (2008).
- G. A. Gescheider, "Psychophysical scaling," Ann. Rev. Psychol. 39, 169–200 (1988).
- J. A. Ferwerda, F. Pellacini, and D. P. Greenberg, "A psychophysically-based model of surface gloss perception," Proc. SPIE 4299, 291–301 (2001).
- 31. W. S. Torgerson, Theory and Methods of Scaling (Wiley, 1958).
- 32. E. C. Poulton, "The new psychophysics: six models for magnitude estimation," Psychol. Bull. **69** (1), 1–19 (1968).
- 33. S. S. Stevens, "The direct estimation of sensory magnitudes—loudness," Am. J. Psychol. **69** (1), 1–25 (1956).
- 34. Commission Internationale de l'Eclairage, "Contrast and visibility" Technical Report 95:1992 (CIE, 1992).
- M. R. Luo, A. A. Clarke, P. A. Rhodes, A. Schappo, S. A. R. Scrivener, and C. J. Tait, "Quantifying colour appearance. part I. Lutchi colour appearance data," Col. Res. Appl. 16, 166–180 (1991).
- H. W. Bodmann, P. Haubner, and A. M. Marsden, "A unified relationship between brightness and luminance," in CIE Proceedings, Kyoto Session 1979 (CIE, 1980), pp. 99–102.
   D. Jameson and L. M. Hurvich, "Theory of brightness and color
- D. Jameson and L. M. Hurvich, "Theory of brightness and color contrast in human vision," Vision Res. 4, 135–154 (1964).
- L. M. Hurvich, "Color vision deficiencies," in *Handbook of Sen-sory Physiology*, Vol. VII/4: Visual Psychophysics, D. Jameson and L. M. Hurvich, eds. (Springer-Verlag, 1972), pp. 582–624.
- 39. R. W. G. Hunt, Measuring Colour (Fountain, 2001).
- American Society for Testing and Materials, "Standard test method for visual evaluation of gloss differences between surfaces of similar appearance," Test Method D4449 (ASTM, 2008).
- T. Hastie, R. Tibshirani, and J. Friedman, The Elements of Statistical Learning: Data Mining, Inference, and Prediction (Springer, 2009).
- J. Beck and K. Prazdny, "Highlights and the perception of glossiness," Percept. Psychophys. 30, 407–410 (1981).
- J. T. Todd, J. F. Norman, and E. Mingolla, "Lightness constancy in the presence of specular highlights," Psychol. Sci. 15, 33– 39 (2004).