How bright should a virtual object be to appear opaque in optical see-through AR?

Jingyu Liu[†*] Akshay Jindal[‡] Claire Mantel[†] Søren Forchhammer[†] Rafal K. Mantiuk[‡]

Department of Photonics Engineering
Technical University of Denmark[†]

Department of Computer Science and Technology
University of Cambridge[‡]

ABSTRACT

Reproduction of occlusions and opaque surfaces are the major challenges of additive optical see-through (OST) displays. This is because the user of an OST display sees a linear mixture of display and environment light, which creates an impression of transparency unless the displayed color is sufficiently bright. The primary goal of this work is to determine how bright a displayed surface needs to be in relation to environment light to be perceived as opaque. We test multiple factors that could affect the perception of opacity: background luminance, contrast, spatial frequency, and accommodation depth in foveal vision. The subjective results, collected on a high-dynamic-range multi-focal stereo display, indicate that a virtual object needs to be, on average, 60 times brighter than the background environment light to be perceived as opaque. A higher contrast of the texture of the virtual object and a background that is out of focus can reduce the required luminance ratio. We demonstrate that a model of visual perception based on Weber's law and accounting for contrast masking and defocus blur can predict the experimental data with an averaged prediction error of 8.29%. Existing perceptual image difference metrics (PSNR, FovVideoVDP and HDR-VDP-3) can also predict the effect of major factors, but with lower accuracy (e.g. prediction error of 34% for PSNR with PU21 encoding).

Index Terms: Human-centered computing—HCI theory, concepts and models

1 INTRODUCTION

One of the main challenges of optical see-through augmented reality (OST-AR) is reproducing virtual objects that appear opaque rather than transparent. OST-AR displays use beam splitters or waveguides to add the light from the display to the light coming from the environment, so the resulting color is always a mixture of the displayed and environment light. Blocking a light field from the environment in a spatially-varying manner is challenging and it is not done in practice. However, it is possible to utilize the limitations of visual perception and make virtual objects appear opaque. The main mean of achieving that is by increasing the luminance of the virtual object and reducing the luminance of the environment light (for example, using neutral-density filters) [23]. It is also possible to use high contrast textures to mask the background. It is, however, unclear how large should the luminance ratio be (between the object and the background) and what the texture contrast should be to make the virtual objects opaque.

In this work, we measure how much brighter than the background (in terms of luminance) a virtual object needs to be to appear opaque. We also want to know how this value depends on the texture of the object — its contrast and spatial frequency. Finally, we quantify the effect of defocus blur on transparency. Defocus blur can be expected in AR as the virtual object is likely to be displayed at a different focal depth than the background, which will result in blurring the background and making it less visible. We conduct a subjective study on an high-dynamic-range multi-focal stereo (HDR-MFS) display [38] to measure the aforementioned effects. We added a small angular velocity to our virtual stimuli to simulate head motion in typical AR scenarios. However, we do not account for motion in our visual model. The high dynamic range (HDR) displays let us test a much larger luminance range, up to 4000 cd/m^2. The two focal planes of the display let us test for natural defocus blur. Our findings should extend to stereoscopic displays in general, which are typically used in OST-AR.

The secondary goal of this work is to model those effects in foveal vision so that the findings can be generalized to different conditions. We find that a simple visual model involving luminance masking, contrast masking, and the modulation transfer function (MTF) of the eye's optics can well explain our data. The data can also be partially explained by the image difference metrics that model relevant aspects of visual perception, such as PU21 luminance coding [20], HDR-VDP-3 [19], and FovVideoVDP [21].

The main contributions of our work include:

- The measurement of the smallest luminance required to achieve perceptual opacity for a set of background luminance levels contrasts and spatial frequency of virtual object texture, and defocus blur.
- A model predicting the luminance required for a virtual object to be perceived as opaque, depending on the background luminance, the virtual object contrast, and the accommodation depth. The model is derived from psychophysical contrast detection and discrimination models and is calibrated on our dataset.
- Demonstration how existing perceptual image difference metrics can be used to estimate the luminance required to make virtual objects opaque.

All our experimental data and the source code of the model are publicly available[1].

2 RELATED WORK

The relevant previous work is organized in two parts. Firstly, Sect. 2.1 introduces existing optical see-through (OST) display systems, laboratory prototypes, and commercial products, focusing mainly on the approaches developed to achieve opacity of the virtual object, either physically or perceptually. Then the key features of the human visual system (HVS) at play for the perception of opacity and the existing opacity models are presented in Sect. 2.2.

2.1 OST displays and opacity

A notable side effect of OST-AR is the inability to darken the environment using the additive OST displays. Depending on the transparency setup and the properties of the optical combiner, the background always adds a certain amount of intensity to the light path

*e-mail: jing@fotonik.dtu.dk

[1] https://github.com/gfxdisp/opaque-ar

(for example, the background always contributes 50% when the optical combiner is 50R-50T). Nevertheless, the existing OST display systems have experimented with different methods to improve the opacity perception for the additive feature of OST-AR.

2.1.1 Background modulation approaches

The environment (background) light in an OST-AR headset can be reduced globally by a fixed percentage using neutral density filters. Such an approach is used in Microsoft Hololens 2 [22]. The global reduction of background light can also be controlled to better adapt to brighter or darker environments [23]. Global methods, however, cannot be applied to individual objects. Without increasing the luminance difference between the virtual and real environments, the light must be blocked locally to make virtual objects opaque. Cakmakci et al. [7] proposed one of the early local methods that can modulate a part of the background in the final view with the help of an LCoS spatial light modulator (SLM). Other reflective SLM such as DMD can also modulate the background selectively [16]. Their solution utilized a binary image factorization to optimize the result. This paper focuses on the approaches that do not involve local background modulation.

2.1.2 Radiometric compensation

Without the help of modulation layers, the appearance of color in OST displays can be improved using radiometric compensation techniques. Hincapié-Ramo et al. [12] proposed a SmartColor system that can enhance the contrast of the projected virtual object to preserve as much of the original hue. Langlotz et al. [17] proposed a method to neutralize the background based on a pixel-wise radiometric compensation. Zhang et al. [37] also make compensation in the color space based on opponent color theory (maximization of the color difference of virtual object color and background color with a set of constraints). Unlike local or global modulation layers, radiometric compensation methods cannot create a physical opaqueness, yet they attempt to trick our eyes to achieve *perceived opaqueness*. In the next section, we will introduce opacity (or transparency) perception based on the knowledge and features of HVS.

2.2 Opacity perception

In real scenes, the opacity of an object is defined by its capacity not to transmit light. Objects that transmit light are either translucent or transparent depending on whether they are composed of materials with a varying or uniform refraction index. However, those physical properties do not directly translate into an impression of transparency for a human observer. As neurons in the HVS do not represent information as absolute luminance levels but rather as image contrast. The contrast threshold can be defined as the minimum contrast that can be resolved by an observer. Threshold contrast can increase when another masker signal is superimposed on it due to the phenomenon known as *contrast masking* [18]. Contrast masking is an indicator in HVS that can be used to model if a signal would be visible through a mask (perceived as transparent). A thorough review of the perception of transparency can be found in [10]. To characterize the perception of transparency, [5, 15, 25, 28] proposed a set of experiments under different stimuli conditions, such as lightness, contrast, and frequency, and derived models of perceived transparency. Further, [9] proposed a chromatic-dependent model based on a subtractive color mixture and concluded that the subjective results of transparent perception are closer to a subtractive than an additive color mixture model.

Virtual objects have no actual volume or specific opacity in additive-only OST-AR systems. The light entering the human eye is a linear combination of the light coming from the real scene (background) and the light from the virtual object. Because the background light is always present in OST-AR, it is necessary to generate perceptual opaqueness when it does not occur physically. Zhang et al. investigated the influence of background luminance and contrast on the perception of transparency in OST-AR in [34]. They set-up a scaling experiment in which they ask participants to rate the transparency of a uniform virtual object. They present observers with combination of background at various luminance and contrast levels and virtual object at varied luminance levels (thus obtaining different contrast levels within the virtual object + background area). They analyze the obtained ratings as a function of the ratio between the contrast within the virtual object + background area and the contrast of the background. One of their finding is that the ratings of perceived transparency depend on the contrast of the background and on the ratio between the luminance of the background and that of the virtual object.

A relevant aspect for transparency perception in OST-AR is that the depth of the different elements seen can differ. If the depth difference between a virtual object in focus and the background is large enough, the background will appear blurred on the retina. The defocus blur can be characterized by the circle of confusion (COC) (radius of the disc corresponding to the projection of an out of focus point on the retina). This can be accounted for when simulating the MTF to model contrast gain for the perception of transparency.

3 Experiment

This section describes the experiment conducted to measure the perception of opacity in an additive OST display.

3.1 Design

To set a clear goal for the participants, we define *opaque threshold* as the threshold of luminance for a virtual object to appear opaque over the real-scene background. The objective for the participants was to adjust the luminance of the virtual object so that it started to appear opaque.

3.2 Apparatus

We use a prototype of an HDR-MFS display explained elsewhere [38]. Here, we give a brief overview of the relevant components of the display. Fig. 1 illustrates the displays and a physical checkerboard background on which virtual objects were shown. A pair of HDR displays (per eye) forms two planes at which the image is shown in the 3D space in front of the observer (dashed purple and green lines in Fig. 1). The far display plane shows images at a similar distance as the background (814 mm or 1.23 D[2]), and the near plane is in front of the background (519 mm or 1.93 D). We use those two distances to measure the effect of defocus blur on the perception of transparency. Each HDR display can produce luminance up to 4000 cd/m^2 with the black level of less than 0.01 cd/m^2, which let us test a wide range of luminance levels for the target virtual object. The HDR displays use a linear tone curve with no normalization. The other details of the display setup can be found in the caption of Fig. 1 and in [38].

To ensure that the scene is rendered from the correct eye position and to account for each observer's inter-ocular distance, we estimated each eye's position in a calibration procedure, illustrated in Fig. 2. The participants were asked to align crosses displayed on each of the four display planes with 4 LEDs of known positions located near the background. The correspondences let us find the projection matrices from the 3D word space to the pixel coordinates of each display [11, 24].

3.3 Stimuli

The virtual object and the background in our experiment were shown in grayscale. The virtual object was rendered on the near or far focal

[2] where D denotes Diopter

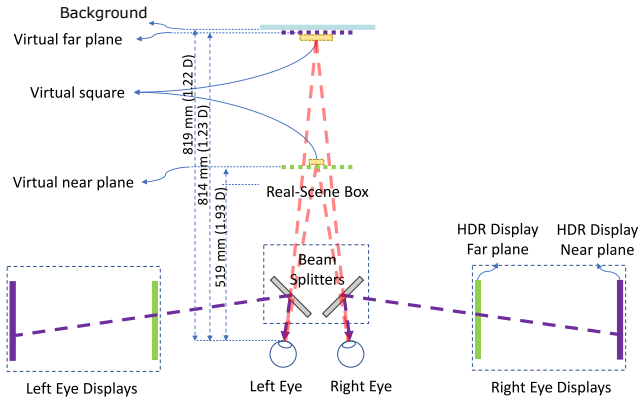


Figure 1: The schematic of the high-dynamic-range stereo multi-focal display, as seen from the top. The image for each eye is formed by combining the light from two HDR displays using 50R-50T beamsplitters (not shown). The real scene image is seen through smaller beamsplitters (70R-30T) in front of the eyes (shown in the diagram). The dashed lines represent virtual focal planes of the corresponding displays (matching inline color) as seen by the participant.

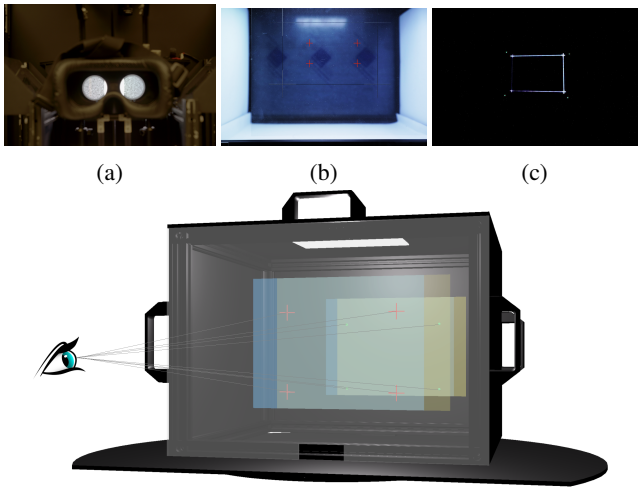


Figure 2: Calibration process. 2a shows a VR view-piece used to view the displayed stimuli, 2b and 2c shows near and far focal plane calibration markers. Participants need to align the virtual cross with the real cross 2b and LEDs 2c. 2d is an overview illustration depicting the relative position of the real scene box, calibration focal planes, and observers' eyes.

plane, with a fixed projected size of $6.67^\circ \times 6.67^\circ$ (the object is larger when shown on the far plane). Its texture contained a band-limited white noise with a mean frequency of f_v, contrast C_v, and mean luminance L_v. The noise samples were drawn from the uniform distribution and then filtered in the Fourier domain using the filter formed by a Gaussian function of logarithmic frequency with the standard deviation of one octave. The maximum amplitude of the noise was scaled to correspond to contrast C_v. The background was printed, attached to a foam board, and positioned at the distance of ≈ 1.22 D from the participants' eyes (the value differed slightly between participants due to the calibration for the eye position). The background was rectangular with an angular size of $25.46^\circ \times 18.04^\circ$.

The background texture was a checkerboard of a fixed contrast of 0.6 and a frequency of 1 cpd. The reported luminance of the background, L_b, corresponds to the average luminance measured in the spot of 1 degrees diameter as seen through the display (measured with Jeti Specbos 1211 spectroradiometer). The examples of the background and virtual object can be found in Fig. 3.

In the experiment, the background was static while the virtual object made a repetitive horizontal motion with an angular speed modulated by a sinusoidal function of the frequency of 0.5 Hz and amplitude of 0.5°. We added such a relative movement to simulate a more realistic scenario — in a typical AR scenario, the background moves with the head motion. We also noted that the transparency was easier to detect when the object was in motion; therefore, the moving square provides a more conservative estimate. However, the motion effect is not accounted in our visual model.

The tested conditions for the stimuli included two levels of background luminance L_b, texture contrast of the virtual object C_v, texture frequency of the virtual object f_v, and the virtual object depth D_v. The values used are listed in Table 1. The participants adjusted the luminance of the virtual object (square), L_v, to make it appear opaque. L_b, C_v, f_v, D_v alternately become independent factor and dependent factor throughout the experiment.

Table 1: Parameters of the tested conditions. L_b — background luminance, C_v — texture contrast of the virtual object, f_v — texture frequency of the virtual object, D_v — the depth of the virtual object.

stimuli	L_b(cd/m^2)	C_v	f_v(cpd)	D_v(D)	conditions
levels	2 20	0.1 0.5	1 6	1.93 1.23	16

3.4 Procedure

The experiment was implemented in Matlab with the help of Psych-Toolbox [6] and custom code controlling the display. In the test, the participant's task was to adjust the luminance of the moving virtual object with a scroll wheel on the trackball until it appeared opaque. The luminance was adjusted on the logarithmic scale to account for Weber's law. The exact wording used in the briefing form was: "*Make the adjustment to the smallest brightness at which the virtual object is opaque so that the background does not show up on the virtual object.*" Note that the participants were asked to always look at the virtual object during the experiment.

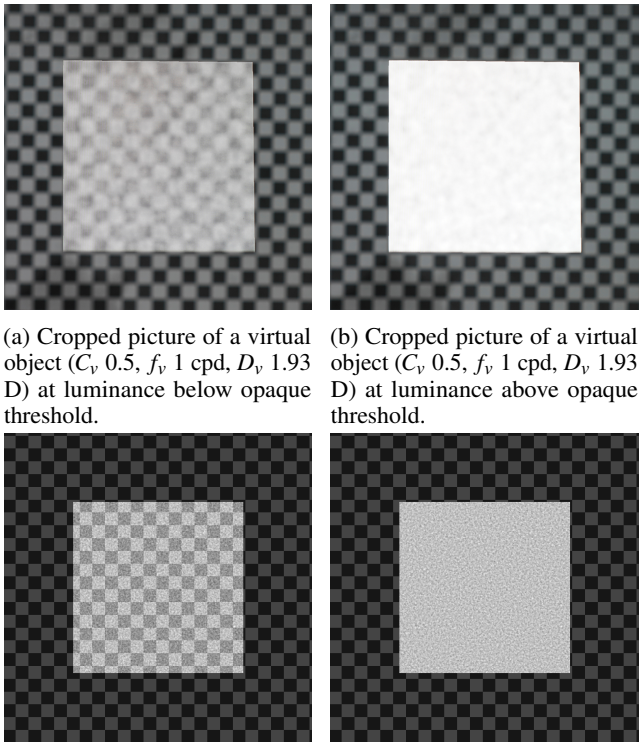
The experiment consisted of 80 trials. The stimuli for each trial were randomly selected from the 16 conditions of factors listed in Table 1. Each condition was measured five times. In each trial, the initial value of L_v was randomized within a uniform distribution of the range of [0.5 500] cd/m^2 to motivate the participant to both increase and decrease the luminance during the test. The whole procedure took around 45 minutes, including individual calibration for each participant.

3.5 Participants

We invited 11 participants for our experiment. The participants are aged 24 to 41 years, with eight men and three women. All had self-reported normal or corrected-to-normal vision acuity (with corrective lenses for five of them). They were naive regarding the conditions used in the experiment. During the experiment, every participant was guided through the calibration. Fig. 4 shows a participant in the experiment.

3.6 Results and ANOVA analysis

An outlier analysis was performed on the data points (using a log scale for the luminance) using the Matlab toolbox LIBRA [30]. The



(a) Cropped picture of a virtual object (C_v 0.5, f_v 1 cpd, D_v 1.93 D) at luminance below opaque threshold.

(b) Cropped picture of a virtual object (C_v 0.5, f_v 1 cpd, D_v 1.93 D) at luminance above opaque threshold.

(c) Stimulus of a virtual object generated with physically accurate transparency at L_b 2 cd/m^2, L_v 3.7 cd/m^2, C_v 0.5, f_v 6 cpd, D_v 1.23 D

(d) Stimulus of an opaque virtual object generated at L_b 2 cd/m^2, L_v 3.7 cd/m^2, C_v 0.5, f_v 6 cpd, D_v 1.23 D

Figure 3: Examples of the visual stimuli (virtual object + background) were presented in the experiment. The top row (3a and 3b) are photographs of the setup, while the bottom row (3c and 3d) consists of simulations. As the photographs (top row - 3a and 3b) are taken with the virtual object shown at D_v 1.93 D and in focus, the background is out-of-focus and therefore presents blur.

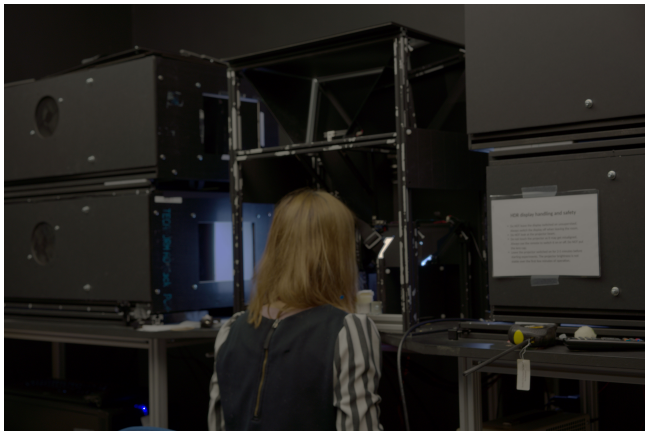


Figure 4: A participant in the experiment

outlier analysis is applied with two variables — the opacity threshold and subject. Hubert et al. [14] is the underlying algorithm applied in the LIBRA for outlier analysis: The method calculates an outlyingness value for each data point with a skewness adjustment based on robust measures of location and scale of the data points, assuming the underlying distribution of the data points is not symmetric. The data points whose skewness-adjusted outlyingness exceeds the preset threshold are eliminated. After outlier removal, we get the filtered data points for the follow-up analysis. The average threshold luminance of the filtered data points for the 16 conditions we tested is shown in Fig. 5.

Fig. 6 presents the measurements (outlier removed) for each of the factors. The background luminance has the largest effect on the threshold luminance, followed by the contrast of the virtual object. The frequency and depth have much smaller effects. The virtual object with a low-contrast texture appears opaque when its luminance is around 60 times higher than that of the in-focus background. The luminance can be reduced to 40 times that of the background luminance if we use a higher contrast texture for the virtual object and when the background is out-of-focus.

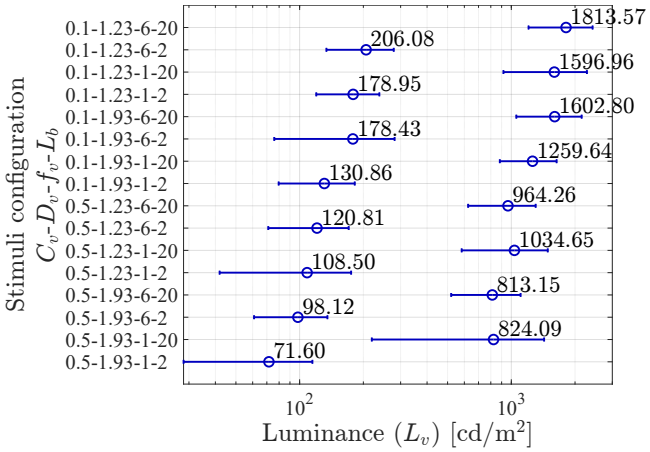


Figure 5: Average and standard deviation across all participants of the luminance threshold of opacity for the virtual object (on the filtered data points). The threshold is highly correlated with background luminance, followed by contrast and accommodation depth. In contrast, the effect of frequency does not show a consistent trend (for the configurations pairs 0.5-1.93-1-20/ 0.5-1.93-6-20 and 0.5-1.23-1-20/ 0.5-1.23-6-20, the effect of frequency is inverse of other frequency stimuli pairs).

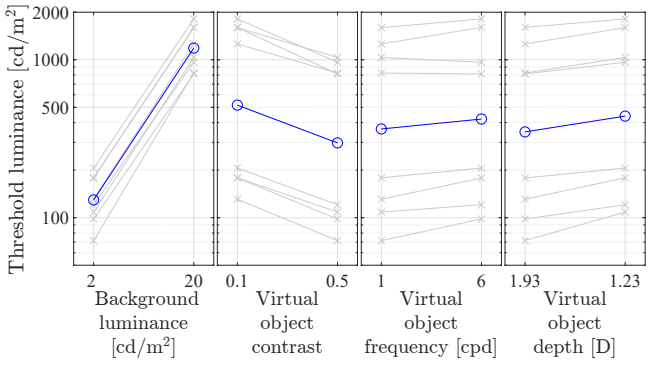


Figure 6: Effects of the four factors tested in the experiment. Each subplot shows one independent factor and three dependent factors. The crosses show the mean luminance for each dependent factor across all participants. The circles show mean luminance averaged across dependent factors. The background luminance has the most substantial effect, followed by contrast, accommodation depth, and frequency.

Table 2: Results of N-way ANOVA on the main effects and interactions for factors with a significant effect (threshold $p < 0.05$). Columns correspond to degree of freedom (d.f.), F-score, p-value and size of effect. All factors are fixed except the subjects who are considered drawn randomly. Factors are ordered based on their size of effect. The greyed factors (above the double horizontal line) are those selected to be modeled in Sect. 4.1.

Factor	d.f.	F	*p*	Size of effect
L_b	1	1687,35	1,75E-12	8,59E-01
C_v	1	183,48	9,28E-08	5,28E-02
Subj.	10	5,05	6,52E-04	5,12E-02
D_v	1	50,26	3,34E-05	9,34E-03
L_b*Subj.	10	6,68	4,10E-08	5,09E-03
f_v	1	14,27	3,62E-03	3,72E-03
C_v*Subj.	10	3,78	2,03E-04	2,88E-03
f_v*Subj.	10	3,42	5,92E-04	2,61E-03
D_v*Subj.	10	2,44	1,13E-02	1,86E-03
L_b*f_v	1	12,08	7,19E-04	9,21E-04
D_v*f_v	1	10,39	1,65E-03	7,92E-04
C_v*f_v	1	7,98	5,58E-03	6,08E-04
Error	115		Sum Sq.	4,16E-01
Total	175		Sum Sq.	4,74E+01

The results are then analyzed using an ANOVA to evaluate which independent variables influence the opacity threshold chosen by the participants [13]. The analysis is conducted using a logarithmic scale for the luminance relating to Weber's law. The independence, normality, and homoscedasticity assumptions necessary for the validity of ANOVA, were evaluated. The independence assumption is fulfilled by participant selection and experiment procedure control. The normality assumption was validated through *Anderson-Darling* test [1], with an averaged p-value over all the 16 configuration of 0.5339. The homogeneity of variances was verified with a *Bartlett* test [4].

Two ANOVA models were used: a repeated measures ANOVA applied to all repetitions and an N-way ANOVA applied on results averaged over all repetitions. As the repetition factor is not found significant and significant factors are identical in both cases, the repeated measure ANOVA does not present additional information and is not presented here.

The statistically significant ($p < 0.05$) factors of an N-way ANOVA on the main effects and second-order interactions are presented in Table 2. The size of effects indicates the relative importance of those factors on the dependent variable (threshold of perceived opaqueness). Based on this, the three most significant factors are considered for modeling the perceived opacity in Sect. 4.1: L_b, C_v, and D_v. As indicated by the variation among their size of effect, their respective impacts on the opacity threshold differ significantly. The $Subj$ and $Subj.*L_b$ factors are disregarded as the focus of this study is not to model personal preferences but a standard observer.

4 Model

The measured luminance threshold for opacity can be predicted by either a simple visual model or perceptual image quality/ difference metrics. We demonstrate both approaches in the following subsections.

4.1 Model of perceived opacity

A virtual object shown on the OST display is judged as transparent if one can see a pattern (contrast) of the background through that object. Using Weber's definition of contrast, the contrast of the background pattern can be expressed as

$$c_\mathrm{b} = \frac{m_\mathrm{b}}{L_\mathrm{b}} \tag{1}$$

where m_b is the amplitude of luminance variations (modulation) and L_b is the mean luminance of the background. When the background is mixed with a virtual object of the mean luminance L_t, its contrast is reduced to:

$$c_\mathrm{b} = \frac{m_\mathrm{b}}{L_\mathrm{b} + L_\mathrm{t}} \tag{2}$$

We can assume that the background will be detected if its contrast exceeds a certain threshold (noise of the visual system). We can model the signal (contrast of the background) to noise ratio as:

$$S = \frac{c_\mathrm{b}}{c_\mathrm{det} + a\,c_\mathrm{t}} \tag{3}$$

where c_det is the minimum detectable contrast (typically around 1%), c_t is the contrast of the target virtual object (masker), and a is the multiplier of the contrast of the virtual object. Both c_det and a are the parameters of the model. We assume that the background is detected and the target object is seen as transparent when S is above 1. The equation above is similar to models of contrast masking of the visual system [8, 19, 32], which contain a signal-independent (c_det) and a signal-dependent ($a\,c_\mathrm{t}$) component. Here, however, we do not consider the frequency of the masker as we found the frequency-dependent effect to be small (refer to Fig. 6 and Table 2).

Defocus Blur We can observe in Fig. 6 that the threshold luminance is lower when the virtual object is shown at a different depth than the background (at 1.93 D). The defocus blur can explain this effect — the eye was accommodated to a different distance than the background pattern, making the pattern appear blurry. To take defocus blur into consideration for our model, we first need to find the diameter of the COC — the convolution kernel used to simulate blur due to the finite pupil size. The diameter of the kernel can be approximated (in degrees) as [26]:

$$d_\mathrm{COC} = \frac{180}{\pi}\,10^{-3}\,p_e\,|F - D_o| \tag{4}$$

where F is the depth of the focal plane in diopters and D_o is the depth of the object out-of-focus, also in diopters [26]. As the virtual object was always in focus in our experiment, F was set to $1/D_v\,\mathrm{m}^{-1}$. D_o was the same for all tested conditions (1.23 D). p_e is the diameter of the observer's pupil in mm, which depends on the size and luminance of the adapting field. We use the unified formula from [33], assuming the field size of 1 deg^2 and the luminance of the virtual object.

We want to find how much the contrast of the background pattern is reduced because of the defocus blur. To find the contrast reduction, we transform the COC kernel into the frequency domain to obtain an MTF of the eye. Because COC is radially symmetric, its 1D projection is a rectangular function. The Fourier transform of the rectangular function is a sinc function:

$$\mathrm{MTF}(\omega) = |\mathrm{sinc}\,(\omega\,d_\mathrm{COC})| \tag{5}$$

where ω is the spatial frequency and d_COC is given in Equation 4.

To model our data, we consider only the fundamental frequency, which is 1 cpd for the checkerboard background. Therefore, the defocus blur will cause the drop of contrast equal to $\mathrm{MTF}(1)$, so that the signal-to-noise ratio becomes:

$$S = \frac{c_\mathrm{b} \cdot \mathrm{MTF}(1)}{c_\mathrm{det} + a\,c_\mathrm{t}} \tag{6}$$

For more complex background patterns, the modulated contrast can be computed numerically using Fourier transformation.

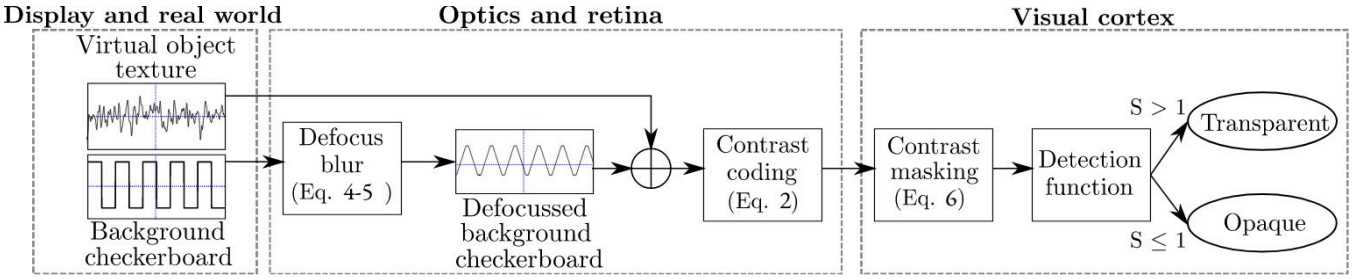


Figure 7: Block diagram for our visual model. We apply MTF based on diopter distance to simulate defocus blur on the real-scene background. The blurred background and the noise signal of the virtual object are then taken into the contrast coding. We conduct a contrast masking algorithm and define whether the model output takes a virtual object as opaque based on the detection function.

Model fitting We fit the parameters of the model to the measurements. The loss function for fitting is:

$$L = \sqrt{\frac{\sum_{i=1}^{8}(\lg d_i - \lg m_i)^2}{8}} \tag{7}$$

where d_i and m_i are the subjective data and model predictions for L_v in different stimuli configuration respectively. lg stands for $\log_{10}$. We drop the spatial frequency factor f and only perform fitting over remaining 3 factors L_b, C_v, D_v (8 conditions in total) as frequency has little effect on threshold luminance (Sect. 3) The model predictions for the range of factors used in our experiment can be found in Fig. 8. To find the detection threshold for each condition, we assume $S = 1$ and solve Equation 6 for L_b. Note that L_b is a part of c_{b} in Equation 6 defined by Equation 2. The model can explain the data well, especially considering the variability in subjective measurements shown in Fig. 5. The fitted parameters of the model with a smallest value of the loss function are $a = 0.0122$, and $c_{\mathrm{det}} = 0.0057$. We measured the fitting by the averaged prediction error in the same form of Equation 7 (i.e. RMSE on $\lg L_v$). The result is 0.0346 or 8.29% in percentage. We also measure the goodness of fit over all 16 conditions through reduced chi squared analysis on $\lg L_v$:

$$\chi_\nu = \sum_{i=1}^{16} \frac{(\lg d_i - \lg m_i)^2}{\nu \cdot \sigma_i^2} \tag{8}$$

where ν is the degree of freedom of the measurement. In our analysis ν is 14 (16 stimuli configurations minus 2 parameters, a and c_{det}, in the prediction model). σ_i^2 is the variance of each observation $\lg d_i$. We got a χ_ν of 0.15 for the fitting of the model, which indicates an over-fitting ($\chi_\nu < 1$). However, according to [2] the reduced chi squared analysis may have limitation in testing a non-linear model like ours with respect to fitted parameters, because the degree of freedom is not well-defined.

4.2 Perceptual image difference metrics

Perceptual image difference metrics can also be used to predict the luminance required for opacity. Two of those metrics are photometric, i.e., they can take input pixel values in SI units ($\mathrm{cd/m^2}$): HDR-VDP-3 [19] and FovVideoVDP [21]. Both metrics model the early stages of human vision (optical and retinal pathways, spatio-temporal contrast sensitivity, cortical magnification, and contrast masking) to predict visual perception. FovVideoVDP is an extension of HDR-VDP3 toward displays with a large field of view (particularly used for AR/VR) that models the variation of spatio-temporal sensitivity of the HVS across the visual field and extends the metric to the temporal dimension. The other metrics applied are PSNR, SSIM, MS-SSIM [31], VSI [35], and FSIM [36]. Since the metrics were all designed for standard dynamic range (SDR) images, they require a mapping from high dynamic range (HDR) physical luminance to perceptually uniform units [3]. The PU21 mapping described in [20] is applied to the stimuli.

All those metrics are full reference ones, and in this context, they evaluate the difference between the visual stimuli (background + virtual object) synthesized to be physically identical to what was presented to the participants of the experiment (Sect. 3) and a reference that is identical except for an opaque virtual object. When simulating the reference with an opaque virtual object, the brightness of the virtual object is increased by the median value of the background to obtain a virtual object with similar luminance in both transparent and opaque cases, as depicted for L_b=2 $\mathrm{cd/m^2}$, L_v=3.7 $\mathrm{cd/m^2}$, C_v=0.5, f_v=6 cpd, D_v=1.23 D in Fig. 3c and Fig. 3d respectively. We assume the in-focus condition as none of the metrics can model defocus blur (D_v=1.23 D).

Examples of the metric predictions are depicted in Fig. 9. To predict the luminance threshold for opacity with each metric, we find an optimal metric value, *OptM*, for which the luminance of the virtual object is the closest to the subjective measurement. *OptM* represents the metric value above which stimuli are evaluated as opaque. It was computed as the average over all configurations of the metric values for the stimuli at the subjective luminance threshold:

$$OptM = \overline{Metric(\lg d_i)}\Big|_1^8 \tag{9}$$

where i is the configuration index. As in Equation 7 d_i is subjective data for L_v. For the simulation with metrics, we do not consider the depth factor D_v since the metrics we applied do no model defocus blur; Thus the average of the *OptM* is over the 8 configurations of L_b, C_v, and f. The prediction error is then computed as the RMSE between the predicted and measured $\lg L_v$ luminance for all conditions. The prediction errors for the objective metrics and the perceptual model from Sect. 4.1 are presented in Table 3. The table shows that our custom model from Sect. 4.1 results in the smallest prediction error of 8.29%, but the effect can also be well predicted by the PU21 transform used in PSNR (error of 34%). The more complex metrics, such as HDR-VDP-3 and FovVideoVDP, account for many more effects, but they do not improve prediction performance. We have also tested the combination of PU21 with SSIM, MS-SSIM, VSI, and FSIM, but we found all those metrics to be too sensitive to predict the luminance threshold of opacity (*OptM* was 1 for all those metrics).

5 Discussion

During the setting up of the experiment described in Sect. 3, both band-limited noise and checkerboard were tried as background patterns. Preliminary results indicated stronger inter-observer agreement when using the checkerboard pattern, and it was selected for the final set-up. The likely reason is that the sharp edges due to the harmonics with higher frequencies act as a stronger transparency cue.

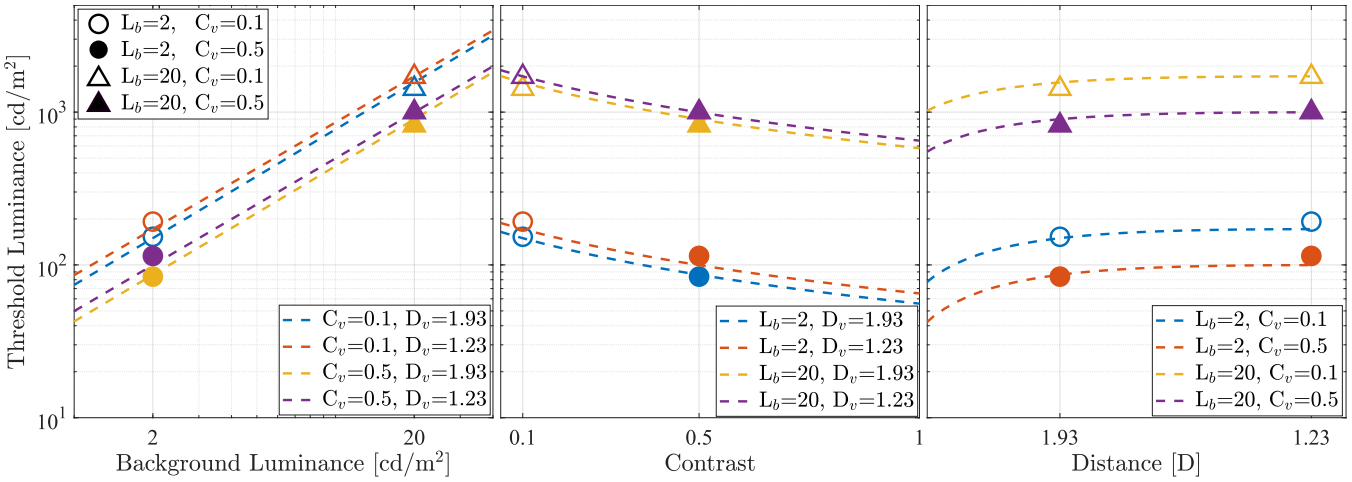
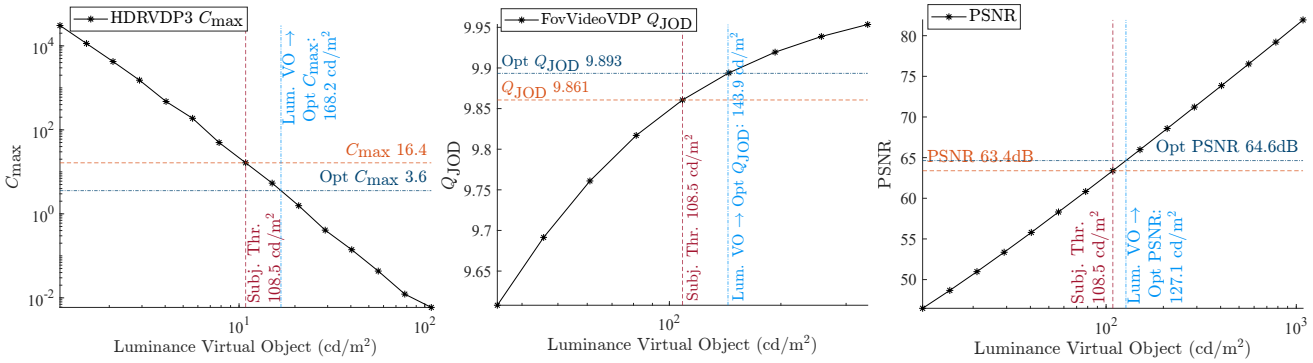


Figure 8: The threshold luminance required for the object to appear opaque, as predicted by the proposed model. The data marks represent the measurements, and the dashed lines are the model predictions. Refer to Table 1 for the explanation of the variables.



(a) C_max indicator from HDR-VDP-3 [19] (static virtual object)

(b) FovVideoVDP [21] (dynamic virtual object)

(c) PSNR (static version of the virtual object)

Figure 9: Examples of objective measures evaluations for the configuration L_b 2 cd/m^2, C_v 0.1, F_v 1cpd and F_v of 1.23 D. The red dash curve indicates the subjective threshold (Subj. Thr.) for opacity from the experiment. The average metric value over all configurations for a virtual object at the subjective threshold is indicated in dark blue. The luminance of the virtual object (Lum. VO) yields that value in light blue.

Table 3: Prediction error for objective metrics and the model proposed in Sect. 4.1

Metric / model	**Avg. metric value of subj. thresholds**	**RMSE on** $\lg L_v$ **/ %**
HDR-VDP-3 C_max	3.600	0.26862 / 85%
FovVideoVDP Q_JOD	9.893	0.1983 / 58%
PU21 + PSNR	64.6 dB	0.1276 / 34%
Our model from Sect. 4.1	—	**0.0346 / 8.29%**

However, the agreement among observers still presented variability in the final set-up. The screening for outliers yielded a higher concentration of outlying data points for two of the eleven participants. As those outliers are located towards higher luminance values than the average, it indicates a more conservative understanding of what constitutes opacity for those participants.

The effect of defocus blur in our prediction model is smaller than the subjective data. The reason can be that we simplified the MTF in accounting for the effect of defocus blur as a sinc function. Applying a comprehensive multi-factor MTF parameterized by Zernike polynomial [27, 29] can increase the precision of the modeling and consider the difference among individuals of the population.

The performance comparison in Sect. 4.2 between the perceptual model from Sect. 4.1 and objective metrics shows that the objective metrics achieve higher prediction errors. When fitting the prediction model to the subjective data, there are two parameters, a and c_{det}, tuned in Equation 6. While for objective metrics, *OptM* is the only parameter adapted to the data. Most likely, the additional degree of freedom allows for a better fit.

Extending the scope of our results towards more realistic stimuli is necessary for a straightforward application. The two main directions are: an increased range of characteristics for the background and virtual objects and a more complex 3D real scene. Our model was derived with a limited level of factors in general. However, since we rely on principled models, contrast detection and MTF, there is a good chance that the predictions can generalize outside the measured parameter space. One possible direction for future work is to set up a more comprehensive test that takes more levels of factors and

a broader range of factors into consideration, which will enable us to develop a prediction model with more complex psychophysical models such as the ones introduced in [25]. Another direction is introducing a 3D virtual object into the experiment with a more complex background scene that provides a more realistic scenario with other visual cues for opacity perception, such as occlusion concerning disparity for the stereoscopic display.

As for the applications of our work, our model can help in setting display design guidelines for AR hardware developers. The findings are independent of an AR display or technology used. We report the results in terms of physical units (cd/m^2, visual degrees) so that they can generalize to any OST display. It can also be used in real-time graphics to optimize the opacity of the rendered content w.r.t. the parameters of our model (virtual content luminance and contrast, background luminance, and distance to background).

6 Conclusion

Making virtual objects appear opaque on an OST-AR display is an important and challenging problem. Our results show that an opaque appearance can be achieved mainly by increasing the luminance of the virtual object and, to a lesser degree, by using high contrast textures, which could mask the background. Defocus blur can also improve opacity, depending on the content, the background, and depth separation between the virtual object and real background. While those findings are not surprising, it was important to quantify the strength of these effects. We found that the virtual object's luminance needs to be 40 to 60 times higher than the luminance of the real environment to avoid a transparent appearance. The high luminance requirement imposes a significant limitation for OST AR displays — they cannot reproduce virtual opaque objects that match the brightness of a real scene, making them unsuitable for practical AR applications, in which virtual objects should seamlessly blend with the environment.

We proposed a simple model to explain our experimental data and predict the luminance required to achieve opacity. It relies on Weber's law for luminance, a simple masking model, and the model of defocus blur. Because the model was fitted to limited data, it may not account for the full range of conditions, such as much lower luminance levels (at which Weber's law does not hold). More complex visual metrics can be accounted for, such as PU21 encoding, FovVideoVDP, and HDR-VDP-3, but at the cost of much higher complexity and less accurate predictions for the range of factors tested in our experiment. Those metrics also do not account for defocus blur, which must be modeled separately.

We hope that our findings and the proposed model can help make better decisions regarding the luminance levels used in OST-AR displays. For example, the model could be used to control the peak luminance of the display, and if that is possible, the transmittance of the filter is used to dim the real environment [23]. To achieve that, an AR headset would need to measure the luminance of the real environment and use our model to control the display and dimming parameters.

Acknowledgments

This project has received funding from the European Union's Horizon 2020 research and innovation programme under Marie Skłodowska-Curie Grant Agreement No. 765911 (RealVision).

References

[1] T. W. Anderson and D. A. Darling. Asymptotic theory of certain" goodness of fit" criteria based on stochastic processes. *The annals of mathematical statistics*, pp. 193–212, 1952.

[2] R. Andrae, T. Schulze-Hartung, and P. Melchior. Dos and don'ts of reduced chi-squared. *arXiv preprint arXiv:1012.3754*, 2010.

[3] T. Aydin, R. Mantiuk, and H.-P. Seidel. Extending quality metrics to full luminance range images. In *Human Vision and Electronic Imaging XIII*, vol. 6806, pp. 109 – 118. International Society for Optics and Photonics, SPIE, 2008. doi: 10.1117/12.765095

[4] M. S. Bartlett. Properties of sufficiency and statistical tests. *Proceedings of the Royal Society of London. Series A-Mathematical and Physical Sciences*, 160(901):268–282, 1937.

[5] J. Beck, K. Prazdny, and R. Ivry. The perception of transparency with achromatic colors. *Perception & psychophysics*, 35(5):407–422, 1984.

[6] D. H. Brainard and S. Vision. The psychophysics toolbox. *Spatial vision*, 10(4):433–436, 1997.

[7] O. Cakmakci, Y. Ha, and J. P. Rolland. A compact optical see-through head-worn display with occlusion support. In *Third IEEE and ACM International Symposium on Mixed and Augmented Reality*, pp. 16–25. IEEE, 2004.

[8] S. J. Daly. Application of a noise-adaptive contrast sensitivity function to image data compression. *Optical Engineering*, 29(8):977, 1990. doi: 10.1117/12.55666

[9] F. Faul and V. Ekroll. Psychophysical model of chromatic perceptual transparency based on substractive color mixture. *JOSA A*, 19(6):1084–1095, 2002.

[10] D. Gigilashvili, J.-B. Thomas, J. Y. Hardeberg, and M. Pedersen. Translucency perception: A review. *Journal of Vision*, 21(8):4–4, 2021.

[11] R. Hartley and A. Zisserman. *Multiple view geometry in computer vision*. Cambridge university press, 2003.

[12] J. D. Hincapié-Ramos, L. Ivanchuk, S. K. Sridharan, and P. Irani. Smartcolor: Real-time color correction and contrast for optical see-through head-mounted displays. In *2014 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*, pp. 187–194. IEEE, 2014.

[13] D. C. Howell. *Statistical methods for psychology, 7th Edition*. Wadsworth, 2010.

[14] M. Hubert and S. Van der Veeken. Outlier detection for skewed data. *Journal of Chemometrics: A Journal of the Chemometrics Society*, 22(3-4):235–246, 2008.

[15] F. A. Kingdom. Lightness, brightness and transparency: A quarter century of new ideas, captivating demonstrations and unrelenting controversy. *Vision Research*, 51(7):652–673, 2011.

[16] B. Krajancich, N. Padmanaban, and G. Wetzstein. Factored occlusion: Single spatial light modulator occlusion-capable optical see-through augmented reality display. *IEEE transactions on visualization and computer graphics*, 26(5):1871–1879, 2020.

[17] T. Langlotz, M. Cook, and H. Regenbrecht. Real-time radiometric compensation for optical see-through head-mounted displays. *IEEE transactions on visualization and computer graphics*, 22(11):2385–2394, 2016.

[18] G. E. Legge and J. M. Foley. Contrast masking in human vision. *Josa*, 70(12):1458–1471, 1980.

[19] R. Mantiuk, K. J. Kim, A. G. Rempel, and W. Heidrich. HDR-VDP-2: A calibrated visual metric for visibility and quality predictions in all luminance conditions. *ACM Transactions on Graphics*, 30(4):40:1–40:14, jul 2011. doi: 10.1145/2010324.1964935

[20] R. K. Mantiuk and M. Azimi. Pu21: A novel perceptually uniform encoding for adapting existing quality metrics for hdr. In *2021 Picture Coding Symposium (PCS)*, pp. 1–5, 2021. doi: 10.1109/PCS50896.2021.9477471

[21] R. K. Mantiuk, G. Denes, A. Chapiro, A. Kaplanyan, G. Rufo, R. Bachy, T. Lian, and A. Patney. Fovvideovdp: A visible difference predictor for wide field-of-view video. *ACM Trans. Graph.*, 40(4), jul 2021. doi: 10.1145/3450626.3459831

[22] Microsoft. Hololens 2, 2021.

[23] S. Mori, S. Ikeda, A. Plopski, and C. Sandor. Brightview: Increasing perceived brightness of optical see-through head-mounted displays through unnoticeable incident light reduction. In *2018 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*, pp. 251–258. IEEE, 2018.

[24] K. Simek. Dissecting the camera matrix, part 1: Extrinsic/intrinsic decomposition, 2012.

[25] M. Singh and B. L. Anderson. Toward a perceptual theory of transparency. *Psychological review*, 109(3):492, 2002.

[26] H. Strasburger, M. Bach, and S. P. Heinrich. Blur unblurred—a mini tutorial. *i-Perception*, 9(2):2041669518765850, 2018.

[27] L. N. Thibos, X. Hong, A. Bradley, and X. Cheng. Statistical variation of aberration structure and image quality in a normal population of healthy eyes. *JOSA A*, 19(12):2329–2348, 2002.

[28] M. Tommasi. A ratio model of perceptual transparency. *Perceptual and Motor Skills*, 89(3):891–897, 1999.

[29] A. Van Meeteren. Calculations on the optical modulation transfer function of the human eye for white light. *Optica Acta: International Journal of Optics*, 21(5):395–412, 1974.

[30] S. Verboven and M. Hubert. LIBRA: a MATLAB library for robust analysis. *Chemometrics and Intelligent Laboratory Systems*, 75(2):127–136, 2005. doi: 10.1016/j.chemolab.2004.06.003

[31] Z. Wang, A. Bovik, H. Sheikh, and E. Simoncelli. Image quality assessment: from error visibility to structural similarity. *IEEE Transactions on Image Processing*, 13(4):600–612, 2004. doi: 10.1109/TIP.2003.819861

[32] A. Watson and J. Solomon. Model of visual contrast gain control and pattern masking. *Journal of the Optical Society of America A*, 14(9):2379–2391, 1997.

[33] A. B. Watson and J. I. Yellott. A unified formula for light-adapted pupil size. *Journal of vision*, 12(10):12, jan 2012. doi: 10.1167/12.10.12

[34] L. Zhang and M. J. Murdoch. Perceived transparency in optical see-through augmented reality. In *2021 IEEE International Symposium on Mixed and Augmented Reality Adjunct (ISMAR-Adjunct)*, pp. 115–120. IEEE, 2021.

[35] L. Zhang, Y. Shen, and H. Li. Vsi: A visual saliency-induced index for perceptual image quality assessment. *IEEE Transactions on Image Processing*, 23(10):4270–4281, 2014. doi: 10.1109/TIP.2014.2346028

[36] L. Zhang, L. Zhang, X. Mou, and D. Zhang. Fsim: A feature similarity index for image quality assessment. *IEEE Transactions on Image Processing*, 20(8):2378–2386, 2011. doi: 10.1109/TIP.2011.2109730

[37] Y. Zhang, R. Wang, E. Y. Peng, W. Hua, and H. Bao. Color contrast enhanced rendering for optical see-through head-mounted displays. *IEEE Transactions on Visualization and Computer Graphics*, 2021.

[38] F. Zhong, A. Jindal, A. Ö. Yöntem, P. Hanji, S. J. Watt, and R. K. Mantiuk. Reproducing reality with a high-dynamic-range multi-focal stereo display. *ACM Transactions on Graphics*, 40(6):1–14, dec 2021. doi: 10.1145/3478513.3480513