Modeling Surround-aware Contrast Sensitivity

Shinyoung Yi[†]   Daniel S. Jeon[†]   Ana Serrano[‡]
Se-Yoon Jeong[§]   Hui-Yong Kim[§]   Diego Gutierrez[‡]   Min H. Kim[†]

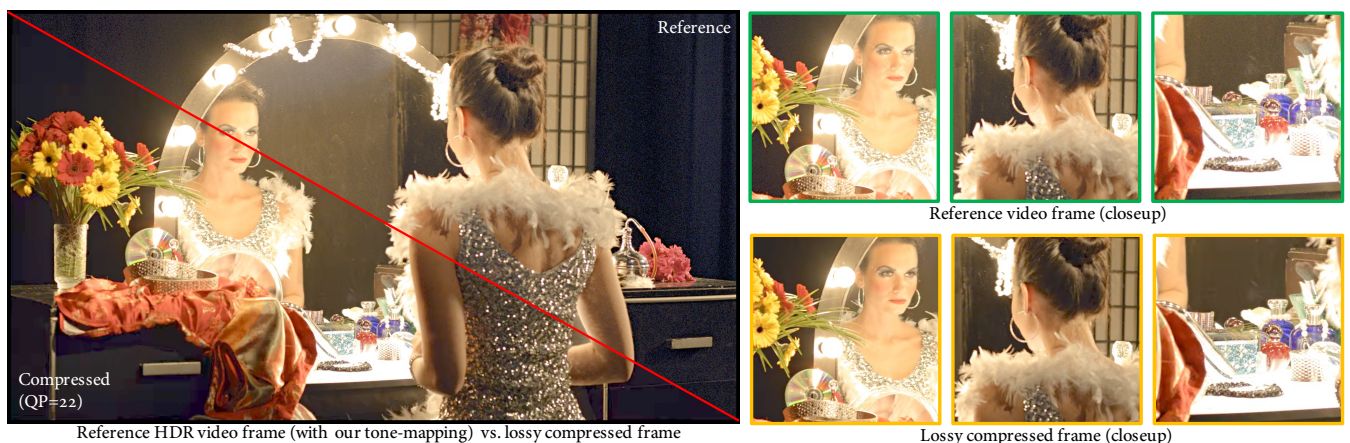[†] KAIST   [‡] Universidad de Zaragoza, I3A   [§] ETRI

Figure 1: *We compare a reference HDR video frame and a lossy compressed frame using our surround-aware CSF model. We compress the original video by three orders of magnitude without perceivable artifacts.*

Abstract
Despite advances in display technology, many existing applications rely on psychophysical datasets of human perception gathered using older, sometimes outdated displays. As a result, there exists the underlying assumption that such measurements can be carried over to the new viewing conditions of more modern technology. We have conducted a series of psychophysical experiments to explore contrast sensitivity using a state-of-the-art HDR display, taking into account not only the spatial frequency and luminance of the stimuli but also their surrounding luminance levels. From our data, we have derived a novel surround-aware contrast sensitivity function (CSF), which predicts human contrast sensitivity more accurately. We additionally provide a practical version that retains the benefits of our full model, while enabling easy backward compatibility and consistently producing good results across many existing applications that make use of CSF models. We show examples of effective HDR video compression using a transfer function derived from our CSF, tone-mapping, and improved accuracy in visual difference prediction.

CCS Concepts
• *Computing methodologies → Perception; Image compression;*

1. Introduction

Display technology has advanced rapidly, making the viewing experience progressively more realistic. In particular, the dynamic range and luminance levels of modern displays have been significantly expanded. Many perceptual studies have been conducted on many different types of visual media, including monitor displays, projectors, printing materials, etc. These perceptual studies are bounded by the characteristics of the experimental medium, a dependency that is well understood [CIE04].

Despite the rapid development of display technology, gathering psychophysical measurements using such modern displays has barely kept up due mainly to limited accessibility to expensive display technology and calibration devices. Often new methods and applications rely on ad-hoc modifications or simple adoption of existing data and models which were obtained with different media, with the underlying strong assumption that they can be carried over to the new viewing conditions.

Contrast sensitivity is a critical aspect of human vision, and has

been the subject of many research studies over decades. Many psychophysical studies have been conducted to discover the response of the human visual system to patterns of varying spatial frequencies and luminance levels; these studies, as argued before, are bounded by the display technology used in their experiments. Most of them report measurements using conventional displays ranging from 0.1 to 100 cd/m². To the best of our knowledge, there is no reliable data of surround-aware contrast sensitivity on an HDR display, which takes into account both the luminance of the stimuli and its surroundings.

We have conducted a series of psychophysical experiments to explore the perceptual impact in contrast sensitivity under these conditions, up to 1000 cd/m². In particular, we are interested in the impact of the surrounding luminance on the perceptibility of contrast for patterns of different spatial frequencies. We have measured perceptual thresholds of spatial frequency under different combinations of stimuli and surround luminance levels. From this, we have developed a novel surround-aware contrast sensitivity function (CSF) that predicts human contrast sensitivity better than existing models. In addition, we provide a practical version of our CSF, which retains the benefits of our complete model while facilitating the prediction of CSF values beyond our measurement range.

Our practical surround-aware CSF model can be plugged into many applications that make use of existing CSF models. This leads to more accurate results especially in images with large luminance contrast, given the extended luminance and contrast range explored in our experiments. We first demonstrate effective HDR video compression by coupling our CSF model to a transfer function that converts HDR signals to the conventional video compression framework. We then present the results of HDR tone-mapping by simply substituting the embedded traditional CSF model in a state-of-the-art tone mapper with our surround-aware CSF. Last, we demonstrate improved accuracy in visual difference prediction using the HDR-VDP-2 framework [MKRH11]. Our dataset and implementation of both surround-aware CSF models will be available publicly.

2. Related Work

Measured CSFs. Experiments to determine contrast thresholds of the HVS start at least as early as 1946 [Bla46]. Schade measured and characterized the dependency of the CSF with respect to spatial frequency by asking observers to judge the visibility of sinusoidal patterns, discovering the now well-known effect of sensitivity dropping as the spatial frequency of the judged stimulus increases [Sch56]. Later, various studies have demonstrated that several factors and luminance conditions can influence the shape of the CSF. Blakemore and Campbell discovered that after adapting for several minutes to a sinusoidal contrast pattern of certain frequency, sensitivity at similar frequencies decreases significantly, while sensitivity to other frequencies remains unaffected [BC69].

In addition to achromatic stimuli, there is a body of work devoted to study the chromatic CSF [KML13, Kel83, HMTN10], however, these studies are mostly restricted to very limited conditions and luminance ranges. Wuerger et al. recently studied the chromatic CSF at a wide range of luminance, but they still treat the cases that

luminance of the sinusoidal pattern and outside of the pattern has the same value [WAK*20, KAPO*20]. Other studies include the analysis of different visual conditions, such as visualization under different eccentricities [RVN78], or neurological conditions, such as arousal [LBLM14], in the shape of the CSF.

Analytical CSFs. Daly's model is one of the most popular [Dal92], and has been used in many applications; unfortunately details of the derivation of the model are not provided. Barten developed a physical model [Bar92] to serve as background to his previously proposed formula to evaluate image quality [Bar89]. This model provides a good fit for many historical CSF measurements; however, later works have suggested that its validity may decrease out of photopic (cone-mediated) vision [KML13]. Barten also proposed a more sophisticated model taking into account a larger number of physical quantities, which yielded a better fit with measured data at the cost of becoming cumbersome to use as a standard [Bar99]. Mantiuk et al. developed a custom CSF model for predicting visible differences in images. They observed that the models proposed by Barten and Daly did not yield accurate fits to their experimental data, and hypothesized that these functions may capture conditions that are different from visual inspection of static images [MKRH11]. Recently, Mantiuk et al. [MKA*20] proposed a chromatic CSF model with consideration of a wide range of luminance based on recent perceptual measurement datasets including [WAK*20, KAPO*20]. Although all these models treat the influence of spatial frequency and stimulus luminance, they do not focus on the joint influence of stimulus and surrounding luminance.

Background/Surround Impact. The effect of background or surrounding luminance was studied for the first time in the 60s [Wes60, VNB67]. These first studies discovered that the contrast sensitivity function changes when it is measured at different mean background intensities. This effect is particularly relevant for modern display technology, which usually allows for high dynamic range. Some recent works have focused on characterizing the effect of surrounding luminance [KK10, BKP14]; however, this effect was studied on isolation and for limited luminance ranges. In general, little is known about the interplay between background luminance and stimuli luminance in contrast sensitivity. Vangorp et al. studied the effect of background or surrounding luminance [VMGM15]. They measured and modeled the threshold-versus-intensity (TVI), which denotes just-noticeable-difference of luminance between two small uniform-luminance patch, taking background luminance into account. They treated the two different luminance values as variables of their model. However, their work is based on TVI, which does not have spatial frequency as a variable unlike CSF.

Barten's [Bar03] is the only existing CSF model that takes into account the influence of the surrounding luminance and its interplay with stimuli luminance. The model is an extension to his previous one [Bar92], taking into account measurements by Rogers et al. [RC73], who measured contrast sensitivity on just three subjects, while analyzing the visibility of *airplane dashboards under specific viewing conditions.* Such measurements are therefore not representative of an average user under typical viewing conditions on a modern display. In contrast to this work, our measurements

cover a large range of background and stimuli luminances, which allows us to model contrast sensitivity both in mesopic and scotopic viewing conditions. Our experiments are performed with a state-of-the-art HDR display with an array of independently controlled high power LEDs as back lighting system, which has been rigorously calibrated.

3. Measurement of Contrast Sensitivity

In order to measure a surround-aware contrast sensitivity function for wide luminance ranges, we have conducted a perceptual experiment using a state-of-the-art HDR display. Our experiment follows *the method of adjustment*, in which subjects are presented with sinusoidal patterns at different luminance levels, and they have to adjust the contrast until they start recognizing the spatial modulation patterns. This method has been used extensively in the literature [GM68, BB71], and allows us to obtain absolute contrast discrimination thresholds for each of our different stimuli. Figure 2 shows our experimental setup, and the stimulus for an example trial.

3.1. Experimental Setup

HDR Display. Our input images consist of grayscale HDR images with linear intensity encoded. We used a 47-inch SIM2 display (HDR47ES4MB) with a resolution of 1920×1080 pixels, a peak luminance of over $4000 \, \text{cd/m}^2$, and a maximum luminance for a full-white screen of approximately $1600 \, \text{cd/m}^2$. Note that the spatial resolution of the backlight LED unit in the display is lower than that of the front LCD unit. However, the visual stimuli within the field of view require the contrast levels achievable by the high-resolution LCD unit alone; therefore, the resolution of the backlight unit is not a limiting factor in our experiments. Subjects were presented with sinusoidal patterns of luminance L in a center region of the screen, surrounded by a background of luminance L_s. Note that L is the average luminance over the sinusoidal pattern.

Calibration. We found that the luminance of the center region where the stimuli are shown is affected by the luminance of the surrounding area of the display due to the characteristics of the HDR display technology, but not the other way around. Thus, we measured calibration functions, which maps control signals at the center region luminance to output luminance values at that region, for each surrounding luminance levels. We used a Specbos Jeti 1200 spectroradiometer, and measured the calibration functions $c(L, L_s)$ for more than 150 fine levels of L for each of 5 levels of L_s, to accurately produce all our combinations of stimuli luminance and surrounding luminance. Then we have generated stimuli image files with taking the inverse functions of these calibration functions to produce intended pixel luminance. Refer to Section 1.1 in the supplementary material for more details.

Stimuli. The area subtended by the sinusoidal patterns of the stimuli is set to $X_o = 2°$ of visual angle that covers the fovea region on the retina, which corresponds to 81×81 pixels from a viewing distance of $d = 1.25$m. The visual angle subtended by the screen at that distance is $45.14° \times 26.32°$, which is used to control the adaptation level while focusing on the frequency stimuli.
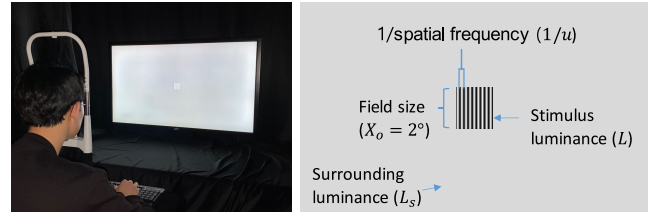

Figure 2: *(Left) Our experimental setup. (Right) The main variables of the stimuli are the stimulus luminance L, the surrounding luminance L_s, the spatial frequency u, and the direction of the pattern.*

The spatial modulation of the sinusoidal pattern is the cosine function of the pixel position, with a random offset ϕ. We sample two directions of the stimulus $D = \{\text{horizontal}, \text{vertical}\}$, five stimulus luminances $L = \{0.56, 2.69, 27.87, 282.91, 1065.25\} \, \text{cd/m}^2$, five surrounding luminances $L_s = \{0.55, 2.75, 28.53, 288.09, 1072.61\} \, \text{cd/m}^2$, and five spatial frequencies $u = \{1.26, 2.52, 5.04, 10.08, 20.16\}$ cpd. This yields a total of 250 ($2 \times 5 \times 5 \times 5$) different stimuli to exploit the entire available combinations of luminance ranges on the display. To avoid the undersampling problem by the display resolution, the values of u were chosen so that the periods of the sinusoidal function are exactly 32, 16, 8, 4, and 2 pixels, respectively.

3.2. Experimental Procedure

Thirteen subjects (10 males and 3 females) with an age range of 20 to 45 years took part in the study. They all reported trichromatic normal or corrected-to-normal vision. Subjects performed the experiment in a dark room. There was no illumination except for the display device The experiment was divided in five sessions; in each one, a single surrounding luminance value L_s was tested. The order of the five sessions was randomized for each participant. Before starting the experiment, subjects spent five minutes adapting to the dark room viewing conditions. Then, for each surrounding luminance value tested, subjects adapted for two additional minutes. During each session subjects were presented with a total of 50 ($2 \times 5 \times 5$) combinations of (D, u, L) in random order to avoid ordering effects. In each trial the stimulus was initially fixed to zero contrast. Subjects were asked to adjust the contrast of the stimuli until they could barely perceive the displayed pattern. They could use the right or left keyboard arrows to increase or decrease contrast in steps. In the case when the given stimuli has a zero contrast, when the participants press the right arrow key, the first contrast change is set to a precomputed extremely small value, which is designed to be invisible ranging from 0.0003 to 0.0066. In subsequent key presses, the right arrow increases the stimulus contrast with $1.3 \times$ contrast intervals. They then had to register whether the shown pattern was horizontal or vertical by pushing the *h* or *v* keys, respectively. To increase reliability of our measured values for sensitivity, if a user identified a wrong direction, the same stimuli was displayed again at a random order until the answer was correct.

Validation. Fifteen random samples of the 250 stimuli were measured twice without any notification to participants. We compute the root-mean-square-error (RMSE) of the two measure-
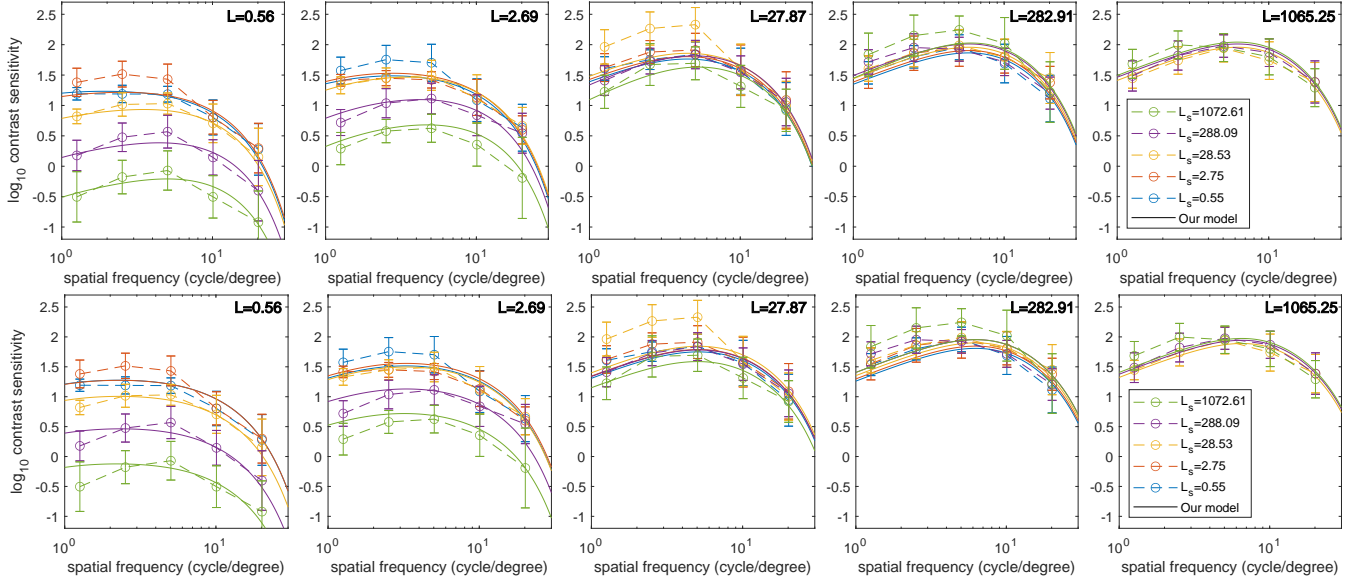
Figure 3: *Our surround-aware CSF models (solid line) compared with our measurement data (dashed line), for different surround luminances L_s. Top: Our full CSF model. Bottom: Our practical CSF model.*

ments in the decibel contrast unit, following Watson and Ahumada Jr. [WA05] and Mantiuk et al. [MKRH11]. The average of differences of short-term repeatability over the entire pool of participants is 4.70 dB. The inter-participant difference is 5.37 dB, i.e., the average of 1-sigma errors of each CSF measurement. This means that the short-term repeatability error for each participant is less than inter-participant error of our experiment.

Discussion of the number of participants. With the consideration of the length of the experiments, we had to limit the number of participants. However, note that similar experiments of color perception have used even less participants in the past, for instance, five and six in [MKRH11] and [KWK09], respectively. As stated in a previous study [AGFC12], there are perceivable differences in contrast vision. Despite the fact, our measurement dataset include an aforementioned gender bias due to the hiring limitation of volunteering participants.

4. Surround-aware Contrast Sensitivity

Our surround-aware CSF depends mainly on three variables: spatial frequency u, stimulus luminance L, and surrounding luminance L_s, and it can be expressed as $S(u, L, L_s)$. We have not observed a significant trend for the horizontal and vertical directions of the sinusoidal pattern; this is in accordance with previous work [Bar03]. Figure 3 shows our CSF measurements for the rest of conditions, with the vertical and horizontal directions averaged, together with the fitted models that we will describe in this section. Refer to Section 2 in the supplementary material for the individual measured data for the two directions. We found that, due to a hardware limitation in the SIM2 display, quantization artifacts appear on the brightest sinusoidal patterns ($L = 1065.25$ cd/m^2) when displayed against the darkest surround levels ($L_s = 0.55$, 2.75 cd/m^2). We therefore discard these two cases for the modeling.

We have discovered a strong dependency of contrast sensitivity on the surrounding luminance. As Figure 3 shows, when a darker stimulus is surrounded by brighter luminance levels ($L < L_s$), contrast sensitivity drops significantly across all spatial frequency bands. This can be clearly appreciated specially in the first two plots. In particular, the measured CSF of the darkest stimuli surrounded by the brightest luminance level presents the lowest sensitivity level. On the other hand, when the presented stimulus is brighter than the surrounding luminance level ($L > L_s$), sensitivity is barely affected by the surrounding luminance level. This can be seen specially in the last three plots in the figure. These are the key insights that we take into account in this section for modeling our CSF.

4.1. A Full Surround-aware CSF Model

We base our derivation of a surround-aware CSF on the formulation of Barten's CSF model [Bar92], which is of one of the common CSF models, widely used in several applications such as video coding [MND13], medical imaging [DIC04], and tone-mapping [FPSG96]. This will allow us to provide backward compatibility for many existing applications. However, Barten's model [Bar92] is independent of the surround luminance L_s. To take L_s into account, we introduce a relative scaling function $R = S(u, L, L_s)/S(u, L, L_s = L)$, which we term the *relative contrast sensitivity*. (In the following, we refer to $S(u, L, L_s = L)$ as $S(u, L, L)$ for convenience.) Our surround-aware CSF model can then be expressed as:

$$S(u, L, L_s) = R(u, L^*) S_B(u, L), \tag{1}$$

where $L^* = L_s/L$, and S_B corresponds to Barten's CSF model, optimized to our measured data. Refer to Section 3.1 in the supplementary material for more details about Barten's original CSF model.
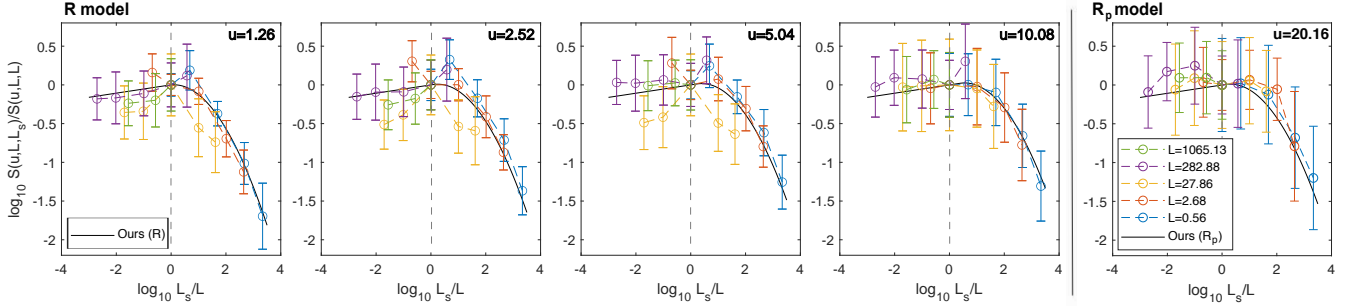
Figure 4: *The first four plots show our relative sensitivity $R(u,L^*)$ as a function of the ratio between surround luminance L_s and stimulus luminance L. From left to right (increasing spatial frequency u), it can be seen how the slope flattens for negative values of L_s/L. The rightmost plot shows our practical relative sensitivity $R_p(L^*)$, which does not depend on u.*

The function R can be regressed from our measurements as:

$$R\left(u,L^*\right) = \frac{S_{\text{data}}\left(u,L,L_s\right)}{S_{\text{data}}\left(u,L,L\right)}. \tag{2}$$

We observe a nonlinear trend of the relative factor R with respect to the luminance ratio L^* as shown in Figure 4. Choosing adequate modeling functions to describe our observed behavior and fitting them to our data (refer to Section 3.2 in the supplementary material for complete model derivation details), our resulting model for the relative contrast sensitivity becomes:

$$R\left(u,L^*\right) = 10^{r(u,\log_{10}L^*)}, \tag{3}$$

with r given by the following expression:

$$r\left(u,l^*;a,b,c,d\right) = -a\left(l^*\right)^2 + bl^* - a\left(l^*+c\right)\sqrt{\left(l^*+c\right)^2+d}$$
$$- ad\ln\left(\sqrt{\left(l^*+c\right)+d}+l^*+c\right)$$
$$+ a\left[c\sqrt{c^2+d}+d\ln\left(\sqrt{c^2+d}+c\right)\right], \tag{4}$$

where $l^* = \log_{10}L^*$. Defining all parameters a, b, c, and d as functions of the spatial frequency u would provide the most accurate results, at the risk of overfitting our measurements. To avoid this overfitting, we first define $b' := b+2ac$, which represents the partial derivative of r with respect to l^* so that $b' = \lim_{l^* \to -\infty}\frac{\mathrm{d}r}{\mathrm{d}l^*}\left(u,l^*\right)$. We then model only b' and c as functions of u, and fit a and d as constants:

$$b'\left(u;q_1,q_2,q_3\right) = \frac{q_1}{1+e^{q_2(\log_{10}u-q_3)}},$$
$$c\left(u;p_1,p_2\right) = p_1\log_{10}u + p_2, \tag{5}$$

where $q_{1,2,3}$ and $p_{1,2}$ are model parameters for b' and c, respectively. The first four plots in Figure 4 show how the the slope of the sensitivity flattens as frequency increases, for negative values of $\log_{10}\left(L_s/L\right)$.

In our regression results, the parameter d is always close to zero (approximately $d \doteq 2 \times 10^{-14}$); we thus set $d = 0$ so that r becomes:

$$r\left(u,l^*;a,b,c\right) = -a\left(l^*\right)^2 + bl^* - a\left(l^*+c\right)\left|l^*+c\right| + ac\left|c\right|. \tag{6}$$

The optimized parameters are presented in Table 1.

R model	a	$p1$	$p2$	$q1$	$q2$	$q3$	σ_0	η	k
	0.07935	-0.6363	0.2157	2246	0.65	-15.56	0.0103	0.0148	10.1826
R_p model	a	b	c	λ					
	0.076	0.073	-0.13	0.24					

Table 1: *Optimized parameters for our full surround-aware R (Section 4.1), and our practical R_p (Section 4.2). Here, σ_0, η, and k are parameters from the original Barten's 1992 model. Refer to Section 3 in the supplementary material for more details.*

4.2. A Practical Surround-aware CSF Model

For many applications (e.g., HDR video coding, tone mapping, or visual difference predictors), an univariate CSF model $S(L)$, dropping dependency of u, is typically used to adjust the perceived luminance level L for a specific frequency band u or a maximum argument of u. In our full model $S(u,L,L_s)$, both $R(u,L^*)$ and $S_B(u,L)$ does depend on the frequency u. However, in such practical applications, a simpler model facilitates the prediction of CSF values beyond the original measurement range, with less potential risks of overfitting than more complex models. We thus propose a *practical relative contrast sensitivity* model $S_p(u,L,L_s)$ by dropping the dependency with the frequency u in $R(u,L^*)$, yielding a simpler $R_p(L^*)$.

Moreover several existing CSF models, such as Barten's [Bar92] or Daly's [Dal92] model, are used commonly in various applications [MKRH11, MDK08, MND13]. However, these models do not take into account the effect of the surrounding luminance L_s. A desirable property of our model is backward compatibility, so that it can be easily incorporated into such existing perception-based models and applications. We can achieve this compatibility by designing new relative contrast sensitivity that does only require a scaling factor λ to adapt existing models to our measurements.

Taking the two previous considerations into account, this new practical model can be written as:

$$S_p\left(u,L,L_s\right) = \lambda R_P\left(L^*;a,b,c\right)S_B^o\left(u,L\right), \tag{7}$$

where $R_p\left(L^*;a,b,c\right)$ is the u-independent, practical relative sensitivity function, $S_B^o\left(u,L\right)$ refers to Barten's original CSF model [Bar92] without optimization to our measured data, and λ is a multiplicative scalar parameter that adjusts the scale of $S_B^o\left(u,L\right)$. Note that R_p is still L_s dependent, so it yields a surround-aware CSF, but has less potential risks of overfitting beyond the original measurement range of u. Also, S_B^o provides a backward compatibility to the conventional CSF model. The values of the parameters a,

b, c, and λ are written in the bottom row of Table 1. Note that, despite our simpler R_p function, frequency-dependent effects are still captured by S_B^o.

4.3. Validation of the Models

To validate the accuracy of our models, we evaluate generalization errors by randomly separating the entire measurements into training and test datasets. We have used 85% of our data for training, and 15% for testing.

Our resulting full CSF model is shown in Figure 3 (top). The training and test errors of our fitting are RMSE=2.69 dB and RMSE=3.93 dB, respectively. The fitting results of our practical CSF model are shown in Figure 3 (bottom). Its training and test errors are RMSE=3.16 dB and RMSE=4.11 dB, respectively.

To avoid the risk of overfitting, we carefully designed our model with combinations of monotonically increasing or decreasing functions for each parameter, rather than using a high-order polynomial regression. For instance, we model our relative sensitivity function Equation (4) as a linear function when the luminance ratio decreases in the log-log domain and model it as a simple quadratic function when the luminance ratio increseas. Moreover, in Equation (5) we include additional parameters to b and c using monotonically decreasing functions. The choice of our model components enforces the local smoothness in the predicted values, implicitly avoiding overfitting.

Note that, although our practical model is slightly less accurate than the full model according to the training error, the generalization error of the practical model (ΔRMSE=RMSE$_{\text{test}}$ - RMSE$_{\text{train}}$) is smaller for the practical model (ΔRMSE=0.95 dB), compared with the full model (ΔRMSE=1.34 dB). This means that our practical model can be better generalized and thus provides a good trade-off between accuracy and ease of use with less parameters.

Comparison. We show in Figure 5 our measured CSF at $L = L_s$ compared with Barten [Bar92], Daly [Dal92]'s, and Mantiuk et al. [MKRH11]'s CSF models. Note that these previous models do not account for surrounding luminance, therefore, we can only compare to them for the particular case $L = L_s$. First, we observe that the measured sensitivity saturates for each frequency as the luminance of the stimulus increases. This is in accordance with previous models [Dal92,Bar92,MKRH11]. Second, we also find that the measured sensitivity tends to increase as the luminance level of the stimuli rises, with the exception of $L = 28.53$ cd/m^2. This luminance level produces the highest sensitivities for low spatial frequencies ($u=$ 1.26, 2.52, 5.04 cpd), as also observed in the data collected by Mantiuk et al. [MKRH11] (Figure 5, bottom). Both Mantiuk et al.'s and our observations found the highest sensitivity levels of low frequency bands to span luminance levels between 20 cd/m^2 and 28.53 cd/m^2. This may be related to the interplay between cones and rods at mesopic vision levels (10^{-1} to 10^1 cd/m^2 [Ska16]).

To our knowledge, there is only one other CSF model that takes into account the influence of the surrounding luminance L_s [Bar03]. Barten's surround-aware CSF is based on measurements from a technical report by Rogers and Carel [RC73], who measured just three subjects while analyzing dashboard visibility in airplanes.
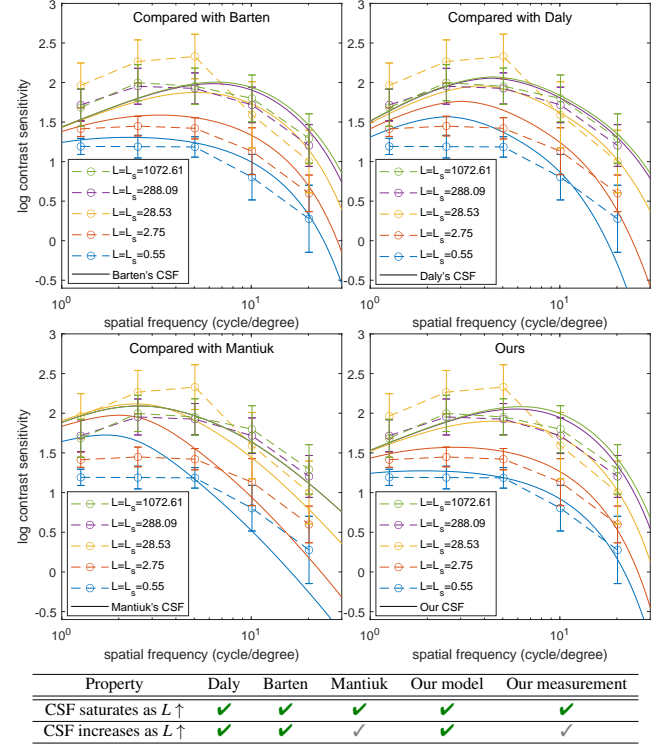

Property	Daly	Barten	Mantiuk	Our model	Our measurement
CSF saturates as $L\uparrow$	✔	✔	✔	✔	✔
CSF increases as $L\uparrow$	✔	✔	✓	✔	✓

Figure 5: *Our surround-aware CSF measurements from the lowest to brightest luminance levels of stimuli. The length of one side of the error bar indicates one standard deviation of participants' responses. Our measured CSFs at $L = L_s$ compared with Barten [Bar92]'s, Daly [Dal92]'s, Mantiuk et al. [MKRH11]'s, and our full CSF models. Dashed lines represent our measurement and solid lines represent the CSF models. The CSF scales of each model are different and thus are adjusted manually to compare both trends. The "✓" marks mean that such properties partially hold, explicitly, Mantiuk et al. [MKRH11]'s CSF model and our measured CSF increase as L increases except for $L = 20 - 30$ cd/m^2, while "✔" indicates consistent agreements.*

Both Barten's 2003 model and our practical model rely on Barten's 1992 model, and in particular on the function S_B^o. The key difference is the inclusion in our model of the relative contrast sensitivity term $R_p(L^*)$. As shown in Figure 6, as L_s/L decreases, our model decays linearly with a small slope in log-log scale, whereas Barten's model decays quadratically. The data from the airplane dashboard experiment [RC73] differs significantly from our data using a modern HDR display when the surround luminance L_s is low: we will show in the following section that this key difference has a strong impact in resulting applications.

5. Applications of the CSF Model

Our practical CSF can be plugged into many HDR display applications that rely on a CSF model. We illustrate examples of HDR
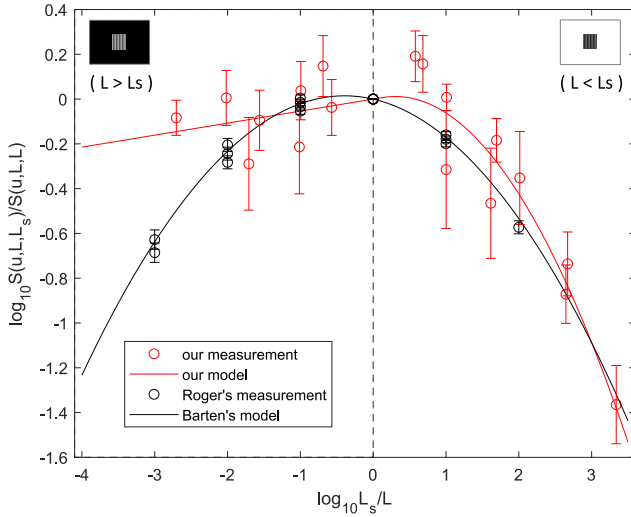
Figure 6: *Comparison of our relative sensitivity R_p of our practical model and Barten's [Bar03], as function of L_s/L. As the luminance ratio decreases, our model decays linearly with a small slope, whereas Barten's model decays quadratically. We speculate that this is due to the two different viewing conditions used to gather data in both models: HDR display (ours) vs. airplane dashboard (Barten).*

Figure 7: *HDR video compression results, using Barten's models ([Bar92] and [Bar03]), and our surround-aware CSF. Our novel CSF allows to preserve details better, thus reducing visible artifacts. Refer to the supplementary video.*

video compression, tone-mapping and prediction of visual differences using our practical CSF[†].

5.1. HDR Video Compression

Current compression methods rely on the traditional integer-based framework; this requires that float-based HDR video content be converted to integers before compression. Quantization artifacts are thus inevitable in the existing video compression workflow described by the ITU-R recommendation [ITU17]. The transfer function of the ITU-R standard compression pipeline is derived from Barten's CSF model [Bar92], and determines which luminance levels are allocated larger bit rates to avoid visible quantization artifacts. Miller et al. [MND13] later presented the perceptual quantizer (PQ), which has been used as a standard transfer function in HDR video coding. The values of the PQ transfer function are again derived from Barten's CSF model [Bar92]. Here, we demonstrate how HDR video compression can also benefit from our new CSF model.

We work on YCbCr space for color space conversion, 10-bit quantization, 4:2:0 chroma subsampling, and HEVC encoding. For determining the compression level, we adjust the quantization parameter (QP) in HEVC encoding, which takes integer values 0-51. For more details, including the derivation of a transfer function from our CSF, please refer to Section 4.1 in the supplementary material. Figure 7 compares the results of our video compression with the standard video compression (ITU-R), and using Barten's subsequent model [Bar03]. Using our CSF allows to preserve details
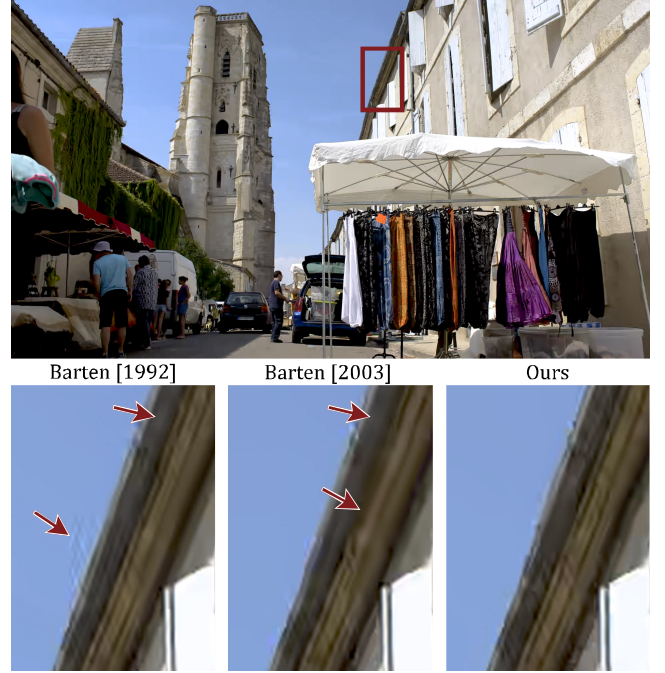
better, thus reducing visible artifacts. Another example frame of our compression is shown in Figure 1. We compress the original video by three orders of magnitude without perceivable artifacts (from 2,491,838 to 2,099 kbps, for QP=22). Refer to the supplementary material for the video examples.

5.2. HDR Tone Mapping

Tone mapping is related to human contrast sensitivity as it tries to preserve the perception of HDR content after remapping to a low-luminance display [RWPD05]. Many tone-mapping operators adopt existing CSFs such as Daly's [Dal92] or Barten's [Bar92]. As a proof-of-concept application, we apply our practical CSF model to Mantiuk's operator [MDK08], which relies on Daly's CSF to estimate contrast sensitivity. This is one of the most widely used tone-mapping operators, and has been ranked by a recent survey as one of the best performing algorithms [EMU17]. Additionally, we also apply Barten's CSF model to this tone-mapper to show that our surround-aware CSF is more suitable for this task. Figure 8 shows how Barten's model [Bar03] fails to produce good tone mapping results, due to its excessive drop in contrast sensitivity for scenes with dark surround luminance (shown in Figure 6). Figure 9 shows the results, compared to the latest implementation of Mantiuk's operator [Man20]. As discussed in Section 3, for dark surrounding environments our CSF presents a lower sensitivity than Daly's CSF, therefore it preserves contrasts better, resulting in cleaner images with less residual haze.

[†] For all applications, the surrounding luminance L_s of scenes is estimated as the geometric mean of the pixel values of the input HDR image.
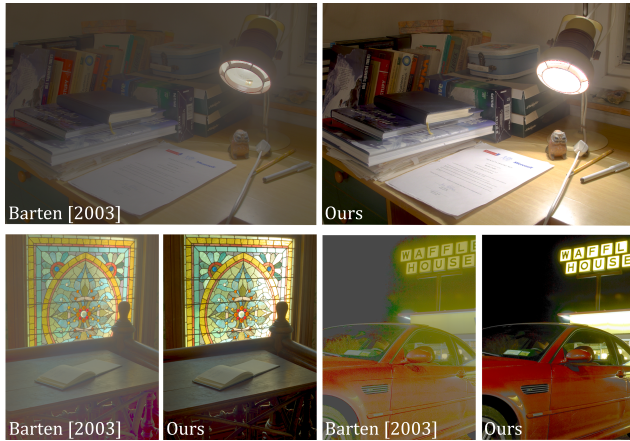
Figure 8: *Comparison of tone mapping results using Barten's surround-aware CSF [Bar03] and our practical model. Barten's CSF heavily drops for scenes with dark surround luminances (Fig. 6), yielding results with very low contrast.*

Figure 9: *Tone mapped images using Mantiuk's [MDK08] original tone mapping operator (left), and plugging our CSF model (right). For scenes with dark surround luminances, our CSF presents a low sensitivity, therefore it compresses contrast better leading to cleaner images with less residual haze (better seen in the digital version).*

5.3. HDR Visual Difference Prediction

Many existing metrics to predict visible differences between images rely on a model of the human visual system (HVS), including contrast sensitivity (e.g., [Dal93, Lub95, WA05, MDMS05]). According to a survey by Hanhart et al. [HBP*15], HDR-VDP-2 [MKRH11] is one of the most reliable metrics. However, it is based on a CSF which does not take into account the effect of surrounding luminance. We integrate our CSF model in the last version of this metric (HDR-VDP-2.2 [NMDSLC15]), substituting their original CSF with our surround-aware model. Figure 10 shows the result, using images from Cadík et al.'s dataset [CAMS11]. This dataset contains six images with distortion maps manually annotated by users, which allows to qualitatively compare the results predicted by the metrics. For images with high surround luminance, our predicted map approximates more closely the subjective map annotated by users (Figure 10, top). This is due to the wider luminance range of our experiments, which allows to model the CSF more precisely in those cases. In low-contrast images, our results largely converge with existing methods, as expected (Figure 10, bottom).
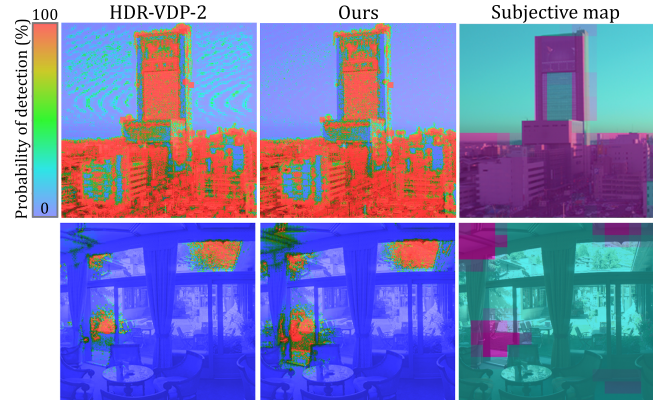

Figure 10: *Predicted probability map for detection of visual differences with HDR-VDP-2.2 (left) and our modified version using our CSF model (center). The rightmost image shows the subjective map of visual differences as annotated by users [CAMS11], where purple and cyan color indicates higher and lower perceived difference, respectively. Our results are more in accordance with users' annotations for high levels of surround luminance (top), while providing similar results to state of the art predictors in other cases.*

6. Conclusions

We have reported a series of psychophysical experiments using a state-of-the-art HDR display, from which we have derived a novel surround-aware contrast sensitivity function. Our work significantly updates existing old datasets (which involved only three users and very different viewing conditions [RC73]), and previous attempts to characterize the effect of surrounding luminance (which involved very limited luminance ranges [KK10, BKP14]). From our full CSF model, we have derived a second, more practical CSF which can be plugged in many existing HDR applications, and consistently provides good results across a wide range of such applications. We have shown examples of video compression, tone mapping, and prediction of visual differences. Our CSF leads to improved results in images with large luminance contrasts, which is a direct consequence of the extended luminance and contrast range covered in our perceptual experiments.

References

[AGFC12] ABRAMOV I., GORDON J., FELDMAN O., CHAVARGA A.: Sex & vision i: Spatio-temporal resolution. biology of sex differences. 4

[Bar89] BARTEN P. G.: The square root integral (sqri): A new metric to describe the effect of various display parameters on perceived image quality. In *Human Vision, Visual Processing, and Digital Display* (1989), vol. 1077, International Society for Optics and Photonics, pp. 73–82. 2

[Bar92] BARTEN P. G.: Physical model for the contrast sensitivity of the human eye. In *Human Vision, Visual Processing, and Digital Display III* (1992), vol. 1666, International Society for Optics and Photonics, pp. 57–72. 2, 4, 5, 6, 7

[Bar99] BARTEN P. G.: *Contrast sensitivity of the human eye and its effects on image quality*. SPIE press, 1999. 2

[Bar03] BARTEN P. G.: Formula for the contrast sensitivity of the human eye. In *Image Quality and System Performance* (2003), vol. 5294, International Society for Optics and Photonics, pp. 231–238. 2, 4, 6, 7, 8

[BB71] BLACKWELL O. M., BLACKWELL H. R.: Ieri: Visual performance data for 156 normal observers of various ages. *Journal of the Illuminating Engineering Society 1*, 1 (1971), 3–13. 3

[BC69] BLAKEMORE C., CAMPBELL F. W.: Adaptation to spatial stimuli. *The Journal of physiology 200*, 1 (1969), 11P–13P. 2

[BKP14] BAEK Y. S., KIM H.-S., PARK S.-O.: Determination of the perceived contrast compensation ratio for a wide range of surround luminance. *Journal of the Optical Society of Korea 18*, 1 (2014), 89–94. 2, 8

[Bla46] BLACKWELL H. R.: Contrast thresholds of the human eye. *JOSA 36*, 11 (1946), 624–643. 2

[CAMS11] CADÍK M., AYDIN T. O., MYSZKOWSKI K., SEIDEL H.-P.: On evaluation of video quality metrics: an hdr dataset for computer graphics applications. In *Human Vision and Electronic Imaging XVI* (2011), vol. 7865, International Society for Optics and Photonics, p. 78650R. 8

[CIE04] CIE: *CIE TC8-01 Technical Report, A Colour Apperance Model for Color Management System: CIECAM02*. Publication CIE 159-2004, Commission Internationale de l'Eclairage (CIE), Vienna, 2004. 1

[Dal92] DALY S. J.: Visible differences predictor: an algorithm for the assessment of image fidelity. In *Human Vision, Visual Processing, and Digital Display III* (1992), vol. 1666, International Society for Optics and Photonics, pp. 2–15. 2, 5, 6, 7

[Dal93] DALY S.: Digital images and human vision. In *The Visible Differences Predictor: An Algorithm for the Assessment of Image Fidelity*. MIT press, 1993, pp. 179–206. 8

[DIC04] DICOM, PS: PS 3-2004. 2004. Part 14: Grayscale standard display function. *Digital Imaging and Communications in Medicine (DI-COM). National Electrical Manufacturers Association* (2004). 4

[EMU17] EILERTSEN G., MANTIUK R. K., UNGER J.: A comparative review of tone-mapping algorithms for high dynamic range video. In *Computer Graphics Forum* (2017), vol. 36, Wiley Online Library, pp. 565–592. 7

[FPSG96] FERWERDA J. A., PATTANAIK S. N., SHIRLEY P., GREENBERG D. P.: A model of visual adaptation for realistic image synthesis. In *Proc. ACM SIGGRAPH '96* (Aug. 1996), pp. 249–258. 4

[GM68] GUTH S., MCNELIS J.: Visual performance: a comparison in terms of detection of presence and discrimination of detail. *Illuminating Engineering 63*, 1 (1968), 32. 3

[HBP*15] HANHART P., BERNARDO M. V., PEREIRA M., PINHEIRO A. M., EBRAHIMI T.: Benchmarking of objective quality metrics for hdr image quality assessment. *EURASIP Journal on Image and Video Processing 2015*, 1 (2015), 1–18. 8

[HMTN10] HIRAI K., MIKAMI T., TSUMURA N., NAKAGUCHI T.: Measurement and modeling of chromatic spatio-velocity contrast sensitivity function and its application to video quality evaluation. In *Color and Imaging Conference* (2010), vol. 2010, Society for Imaging Science and Technology, pp. 86–91. 2

[ITU17] ITU-R: Itu recommendation sector, bt. 2100-1: Image parameter values for high dynamic range television for use in production and international program exchange. 7

[KAPO*20] KIM M., ASHRAF M., PÉREZ-ORTIZ M., MARTINOVIC J., WUERGER S., MANTIUK R. K.: Contrast sensitivity functions for hdr displays. In *London Imaging Meeting* (2020), vol. 2020, Society for Imaging Science and Technology, pp. 44–48. 2

[Kel83] KELLY D.: Spatiotemporal variation of chromatic and achromatic contrast thresholds. *JOSA 73*, 6 (1983), 742–750. 2

[KK10] KIM Y.-J., KIM H.-S.: Spatial luminance contrast sensitivity: Effects of surround. *Journal of the Optical Society of Korea 14*, 2 (2010), 152–162. 2, 8

[KML13] KIM K. J., MANTIUK R., LEE K. H.: Measurements of achromatic and chromatic contrast sensitivity functions for an extended range of adaptation luminance. In *Human Vision and Electronic Imaging XVIII* (2013), vol. 8651, International Society for Optics and Photonics, p. 86511A. 2

[KWK09] KIM M. H., WEYRICH T., KAUTZ J.: Modeling human color perception under extended luminance levels. *ACM Trans. Graph. (Proc. SIGGRAPH 2009) 28*, 3 (2009), 27:1–9. 4

[LBLM14] LEE T.-H., BAEK J., LU Z.-L., MATHER M.: How arousal modulates the visual contrast sensitivity function. *Emotion 14*, 5 (2014), 978. 2

[Lub95] LUBIN J.: A visual discrimination model for imaging system design and evaluation. In *Vision Models for Target Detection and Recognition: In Memory of Arthur Menendez*. World Scientific, 1995, pp. 245–283. 8

[Man20] MANTIUK R.: pfstools-2.1.0. http://pfstools.sourceforge.net/, 2020. URL: http://pfstools.sourceforge.net/. 7

[MDK08] MANTIUK R., DALY S., KEROFSKY L.: Display adaptive tone mapping. In *ACM SIGGRAPH 2008 papers*. 2008, pp. 1–10. 5, 7, 8

[MDMS05] MANTIUK R., DALY S. J., MYSZKOWSKI K., SEIDEL H.-P.: Predicting visible differences in high dynamic range images: model and its calibration. In *Human Vision and Electronic Imaging X* (2005), vol. 5666, International Society for Optics and Photonics, pp. 204–214. 8

[MKA*20] MANTIUK R. K., KIM M., ASHRAF M., XU Q., LUO M. R., MARTINOVIC J., WUERGER S.: Practical color contrast sensitivity functions for luminance levels up to 10 000 cd/m^ 2. 2

[MKRH11] MANTIUK R., KIM K. J., REMPEL A. G., HEIDRICH W.: Hdr-vdp-2: A calibrated visual metric for visibility and quality predictions in all luminance conditions. *ACM Transactions on graphics (TOG) 30*, 4 (2011), 1–14. 2, 4, 5, 6, 8

[MND13] MILLER S., NEZAMABADI M., DALY S.: Perceptual signal coding for more efficient usage of bit codes. *SMPTE Motion Imaging Journal 122*, 4 (2013), 52–59. 4, 5, 7

[NMDSLC15] NARWARIA M., MANTIUK R., DA SILVA M. P., LE CALLET P.: Hdr-vdp-2.2: a calibrated method for objective quality prediction of high-dynamic range and standard images. *Journal of Electronic Imaging 24*, 1 (2015), 010501. 8

[RC73] ROGERS J. G., CAREL W. L.: *Development of design criteria for sensor displays*. Tech. rep., HUGHES AIRCRAFT CO CULVER CITY CA DISPLAY SYSTEMS DEPT, 1973. 2, 6, 8

[RVN78] ROVAMO J., VIRSU V., NÄSÄNEN R.: Cortical magnification factor predicts the photopic contrast sensitivity of peripheral vision. *Nature 271*, 5640 (1978), 54–56. 2

[RWPD05] REINHARD E., WARD G., PATTANAIK S., DEBEVEC P.: *High Dynamic Range Imaging: Acquisition, Display and Image-Based Lighting*. Morgan Kaufmann, dec 2005. 7

[Sch56] SCHADE O. H.: Optical and photoelectric analog of the eye. *JoSA 46*, 9 (1956), 721–739. 2

[Ska16] SKALICKY S. E.: Luminance range for vision. In *Ocular and Visual Physiology*. Springer, 2016, pp. 299–312. 6

[VMGM15] VANGORP P., MYSZKOWSKI K., GRAF E. W., MANTIUK R. K.: A model of local adaptation. *ACM Transactions on Graphics (TOG) 34*, 6 (2015), 1–13. 2

[VNB67] VAN NES F. L., BOUMAN M. A.: Spatial modulation transfer in the human eye. *JOSA 57*, 3 (1967), 401–406. 2

[WA05] WATSON A. B., AHUMADA A. J.: A standard model for foveal detection of spatial contrast. *Journal of vision 5*, 9 (2005), 6–6. 4, 8

[WAK*20] WUERGER S., ASHRAF M., KIM M., MARTINOVIC J., PÉREZ-ORTIZ M., MANTIUK R. K.: Spatio-chromatic contrast sensitivity under mesopic and photopic light levels. *Journal of Vision 20*, 4 (2020), 23–23. 2

[Wes60] WESTHEIMER G.: Modulation thresholds for sinusoidal light distributions on the retina. *The Journal of Physiology 152*, 1 (1960), 67–74. 2