

# Modelling Surround-aware Contrast Sensitivity for HDR Displays

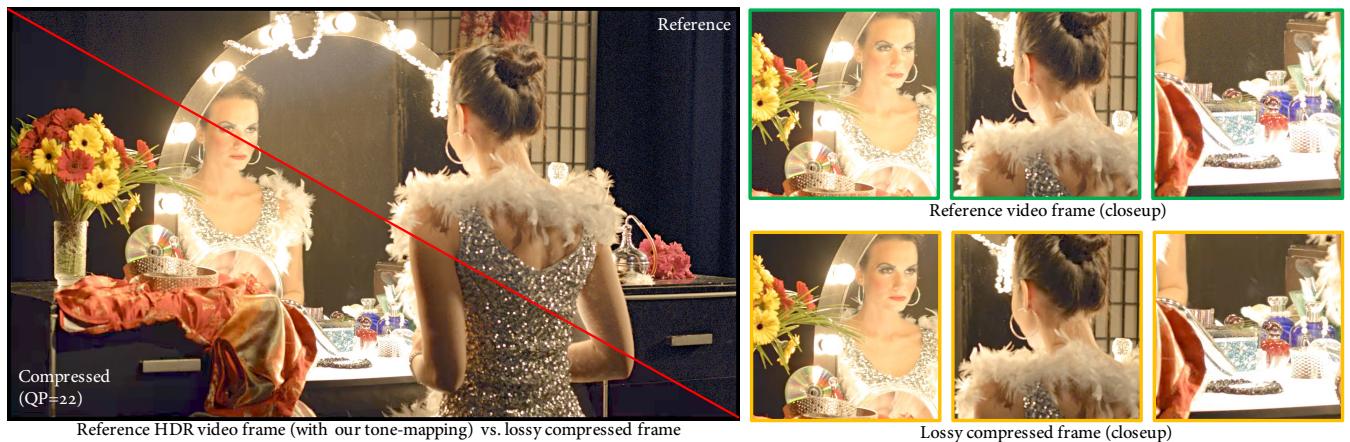
Shinyoung Yi<sup>†</sup> Daniel S. Jeon<sup>†</sup> Ana Serrano<sup>‡</sup>  
 Se-Yoon Jeong<sup>§</sup> Hui-Yong Kim<sup>§¶</sup> Diego Gutierrez<sup>‡</sup> Min H. Kim<sup>†</sup>

<sup>†</sup> KAIST

<sup>‡</sup> Universidad de Zaragoza, I3A

<sup>§</sup> ETRI

<sup>¶</sup> Kyung Hee University



**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many existing applications rely on psychophysical datasets of human perception gathered using older, sometimes outdated displays.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 [?].

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/m<sup>2</sup>. 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/m<sup>2</sup>. 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 [?]. 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 [?]. 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 [?]. 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 [?].

In addition to achromatic stimuli, there is a body of work devoted to study the chromatic CSF [?, ?, ?], 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 [?, ?]. Other studies include the analysis of different visual conditions, such as visualization under different eccentricities [?], or neurological conditions, such as arousal [?], in the shape of the CSF.

**Analytical CSFs.** Daly's model is one of the most popular [?], and has been used in many applications; unfortunately details of the derivation of the model are not provided. Barten developed a physical model [?] to serve as background to his previously proposed formula to evaluate image quality [?]. 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 [?]. 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 [?]. 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 [?]. Recently, Mantiuk et al. [?] proposed a chromatic CSF model with consideration of a wide range of luminance based on recent perceptual measurement datasets including [?, ?]. 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 [?, ?]. 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 [?, ?]; 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 [?]. 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.

In addition to studies about detection thresholds, such as contrast sensitivity and TVI, there are some works that study the effect of background or surround luminance on various factors, such as the effect on perceived brightness [?, ?], perceived dynamic range [?], and chromatic adaptation [?]. Background or surround luminance may also influence several aspects when perceiving natural images, such as perceived contrast of images measured by adjusting gamma [?], subjective quality of tone mapping operators [?], and user preferences for HDR displays [?].

It is important to note that the CIECAM97 model [?] defines stimuli, background, and surround of the adapting field separately. It defines stimuli as a 2° region in the center of view, background

as a  $10^\circ$  to  $12^\circ$  region around the stimuli, and surround as the remainder of the visual field (ambient illumination). The *surrounding luminance* taken into account for our measurements and model (described in Sections 3 and 4 respectively) then covers the luminance of the entire *background* and a part of *surround* as defined in CIECAM97. Although for smaller display devices, the *surrounding luminance* of our model can correspond fully to ambient illumination, our model does not focus on modeling the reflection of ambient illumination on the display screen, which has been previously studied in several works [?, ?].

Recently, Daly et al. [?] presented a study about the effect of surround luminance on contrast sensitivity. However, the surround luminance described in their experiment refers to the luminance level outside a display that covers a field of view of  $33^\circ \times 20^\circ$ , so the variables of their study are substantially different from those of our experiment, as described in Section 3. Additionally, their work does not provide a full CSF model since only a single level of spatial frequency was tested.

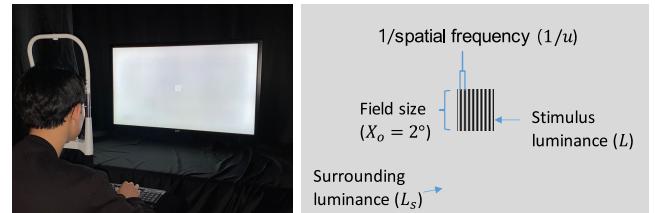
Barten's [?] is the only existing CSF model that takes into account the influence of the surrounding luminance from the display and its interplay with stimuli luminance. This model is an extension to a previous model [?]. The extended model takes into account the measurements obtained by Rogers et al. [?], who measured contrast sensitivity only for three subjects, while analyzing the visibility of *airplane dashboards under very 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 [?, ?], 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  $1920 \times 1080$  pixels, a peak luminance of over  $4000\text{cd}/\text{m}^2$ , and a maximum luminance for a full-white screen of approximately  $1600\text{cd}/\text{m}^2$ . 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



**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 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.

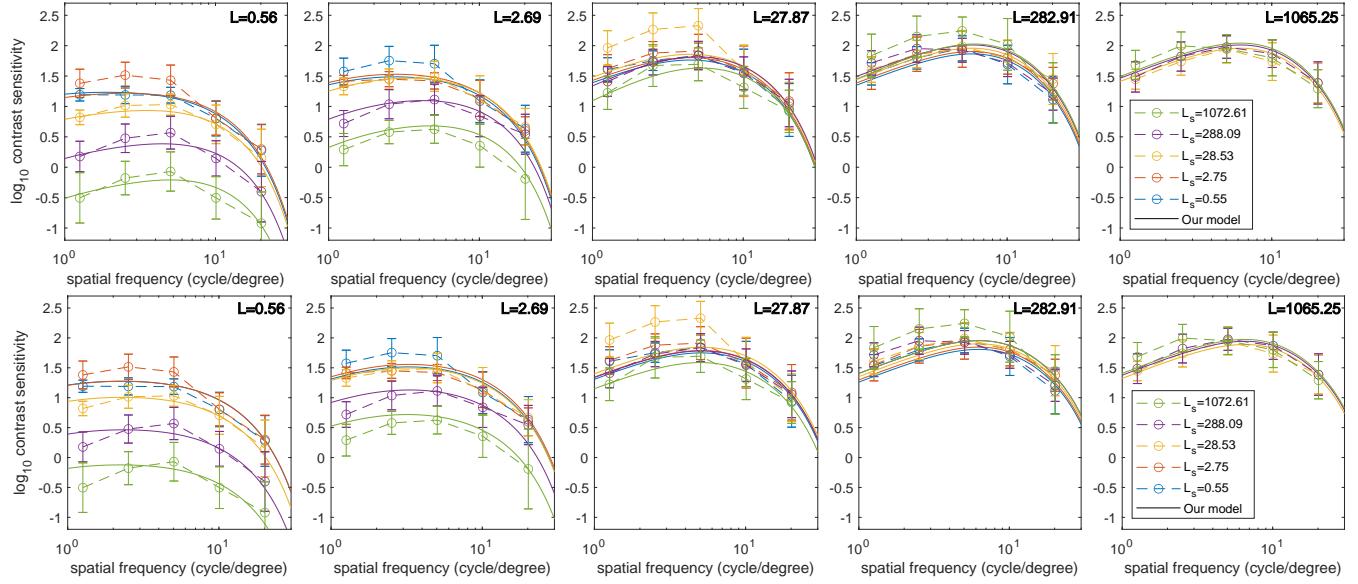
**Calibration.** We found that the luminance of the center region where the stimuli are shown is affected by the luminance of the surrounding area of the display due to the characteristics of the HDR display technology, but not the other way around. Thus, we measured calibration functions, which maps 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  $X_o = 2^\circ$  of visual angle that covers the fovea region on the retina, which corresponds to  $81 \times 81$  pixels from a viewing distance of  $d = 1.25\text{m}$ . The visual angle subtended by the screen at that distance is  $45.14^\circ \times 26.32^\circ$ , which is used to control the adaptation level while focusing on the frequency stimuli.

The spatial modulation of the sinusoidal pattern is the cosine function of the pixel position, with a random offset  $\phi$ . 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  $u = \{1.26, 2.52, 5.04, 10.08, 20.16\} \text{ cpd}$ . 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



**Figure 3:** Our surround-aware CSF models (solid line) compared with our measurement data (dashed line with a circle marker), for different surround luminances  $L_s$ . Top: Our full CSF model. Bottom: Our practical CSF model. Each of the five colors for both dashed and solid lines indicates the five levels of surrounding luminance  $L_s$  as shown in the legend. Refer to Figure 8 to compare the predicted contrast sensitivity by HDR-VDP-2.2 with our perceptual measurements.

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.

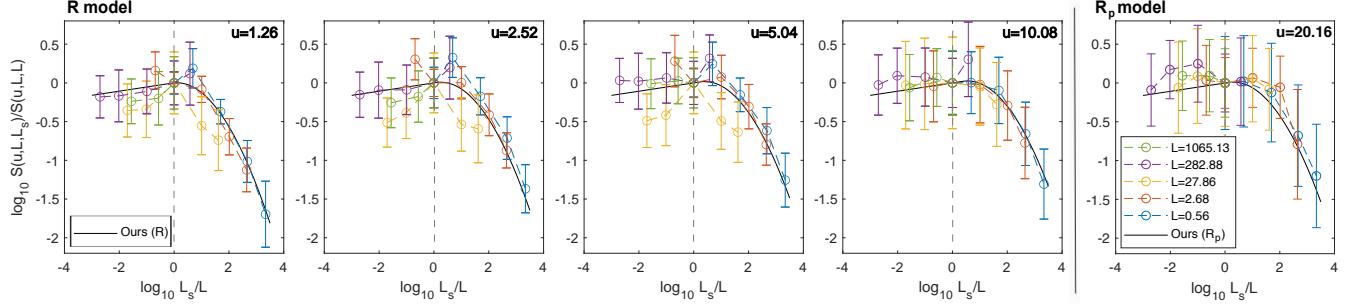
**Validation.** Fifteen random samples of the 250 stimuli were measured twice without any notification to participants. We compute the root-mean-square-error (RMSE) of the two measurements in the decibel contrast unit, following Watson and Ahumada Jr. [?] and Mantiuk et al. [?]. The average of differences 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.

#### 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 [?]. 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 ( $L = 1065.25 \text{ cd/m}^2$ ) when displayed against the darkest surround levels ( $L_s = 0.55, 2.75 \text{ cd/m}^2$ ). 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



**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  $u$ .

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 [?], which is one of the common CSF models, widely used in several applications such as video coding [?], medical imaging [?], and tone-mapping [?]. This will allow us to provide backward compatibility for many existing applications. However, Barten’s model [?] 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:

$$S(u, L, L_s) = R(u, L^*) S_B(u, L), \quad (1)$$

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.

The function  $R$  can be regressed from our measurements as:

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

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:

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

with  $r$  given by the following expression:

$$\begin{aligned} r(u, l^*; a, b, c, d) = & -a(l^*)^2 + bl^* - a(l^* + c)\sqrt{(l^* + c)^2 + d} \\ & - ad \ln(\sqrt{(l^* + c)^2 + d} + l^* + c) \\ & + a[c\sqrt{c^2 + d} + d \ln(\sqrt{c^2 + d} + c)], \end{aligned} \quad (4)$$

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

**Table 1:** Optimized parameters for our full surround-aware  $R$  (Section 4.1), and our practical  $R_p$  (Section 4.2). Here,  $\sigma_0$ ,  $\eta$ , and  $k$  are parameters from the original Barten’s 1992 model. Refer to Section 3 in the supplementary material for more details.

$R$ model	$a$	$p_1$	$p_2$	$q_1$	$q_2$	$q_3$	$\sigma_0$	$\eta$	$k$
$R$ model	0.07935	-0.6363	0.2157	2246	0.65	-15.56	0.0103	0.0148	10.1826
$R_p$ model	$a$	$b$	$c$	$\lambda$					
	0.076	0.073	-0.13	0.24					

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  $b' = \lim_{l^* \rightarrow -\infty} \frac{dr}{dl^*}(u, l^*)$ . We then model only  $b'$  and  $c$  as functions of  $u$ , and fit  $a$  and  $d$  as constants:

$$\begin{aligned} b'(u; q_1, q_2, q_3) &= \frac{q_1}{1 + e^{q_2(\log_{10} u - q_3)}}, \\ c(u; p_1, p_2) &= p_1 \log_{10} u + p_2, \end{aligned} \quad (5)$$

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 slope of the sensitivity flattens as frequency increases, for negative values of  $\log_{10}(L_s/L)$ .

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

$$r(u, l^*; a, b, c) = -a(l^*)^2 + bl^* - a(l^* + c)|l^* + c| + ac|c|. \quad (6)$$

The optimized parameters are presented in Table 1.

#### 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 [?]

or Daly's [?] model, are used commonly in various applications [?, ?, ?]. 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  $\lambda$  to adapt existing models to our measurements.

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

$$S_p(u, L, L_s) = \lambda R_p(L^*; a, b, c) S_B^o(u, L), \quad (7)$$

where  $R_p(L^*; a, b, c)$  is the  $u$ -independent, practical relative sensitivity function,  $S_B^o(u, L)$  refers to Barten's original CSF model [?] without optimization to our measured data, and  $\lambda$  is a multiplicative scalar parameter that adjusts the scale of  $S_B^o(u, L)$ . 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  $\lambda$  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. Summary of CSF models

Here, we provide a brief summary of our CSF models for the convenience of reproduction. Barten [?]’s CSF model  $S_B^o(u, L)$  is:

$$S_B^o(u, L) = \frac{M_{opt}(u)/k}{\sqrt{\frac{2}{T} \left( \frac{1}{X_o^2} + \frac{1}{X_{max}^2} + \frac{u^2}{N_{max}^2} \right) \left( \frac{1}{\eta p E} + \frac{\Phi_0}{1 - e^{-(u/u_0)^2}} \right)}}, \quad (8)$$

where

$$\begin{aligned} M_{opt}(u) &= e^{-2\pi^2\sigma^2 u^2}, \\ \sigma &= \sqrt{\sigma_0^2 + (C_{ab}d)^2}, \\ d &= 5 - 3 \tanh(0.4 \log_{10} L), \\ E &= \frac{\pi d^2}{4} L \left( 1 - \left( \frac{d}{9.7} \right)^2 - \left( \frac{d}{12.4} \right)^4 \right), \end{aligned} \quad (9)$$

and each parameter takes values specified in Table 2.

Our full surround-aware CSF model  $S$  is:

$$S(u, L, L_s) = R(L^*; a, b(u), c(u)) S_B(u, L), \quad (10)$$

where  $L^* = L_s/L$  and  $S_B$  indicates Barten’s CSF described in Equation (8) with three fitted parameters  $\sigma_0$ ,  $\eta$ , and  $k$  shown in the first row of Table 1. Our log-log scaled relative sensitivity model  $r(l^* = \log_{10} L^*) = \log_{10} R(L^*)$  is:

$$r(l^*; a, b, c) = -a(l^*)^2 + bl^* - a(l^* + c)|l^* + c| + ac|c|. \quad (11)$$

Here,  $b$  and  $c$  are functions of  $u$ :

$$\begin{aligned} c(u; p_1, p_2) &= p_1 \log_{10} u + p_2, \\ b(u; q_1, q_2, q_3) &= \frac{q_1}{1 + e^{q_2(\log_{10} u - q_3)}} - 2ac(u). \end{aligned} \quad (12)$$

**Table 2:** Typical parameters values for Barten [?]’s CSF model  $S_B^o$ , which is used in previous work such as [?] and [?]. However, the value of  $X_o = 2^\circ$  is taken from our experimental setup.

$S_B^o$ model	$\Phi_0$	$C_{ab}$	$T$	$X_{max}$	$N_{max}$	$p$
	$3 \times 10^{-8}$ sec deg <sup>2</sup>	0.08 arc min/mm	0.1sec	$12^\circ$	15cycles	$1.2 \times 10^6$ /sec <sup>2</sup> /Td
	$u_0$ 7cycles/deg	$\sigma_0$ 0.5 arc min	$\eta$ 0.03	$k$ 3.0	$X_o$ $2^\circ$	

Our practical surround-aware CSF model  $S_p$  is:

$$S_p(u, L, L_s) = \lambda R_p(L^*; a, b, c) S_B^o(u, L), \quad (13)$$

where the log-log scale of  $R_p$  uses the same formula as Equation (11), but  $a$ ,  $b$ , and  $c$  are constant shown in the second row of Table 1.

### 4.4. 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.

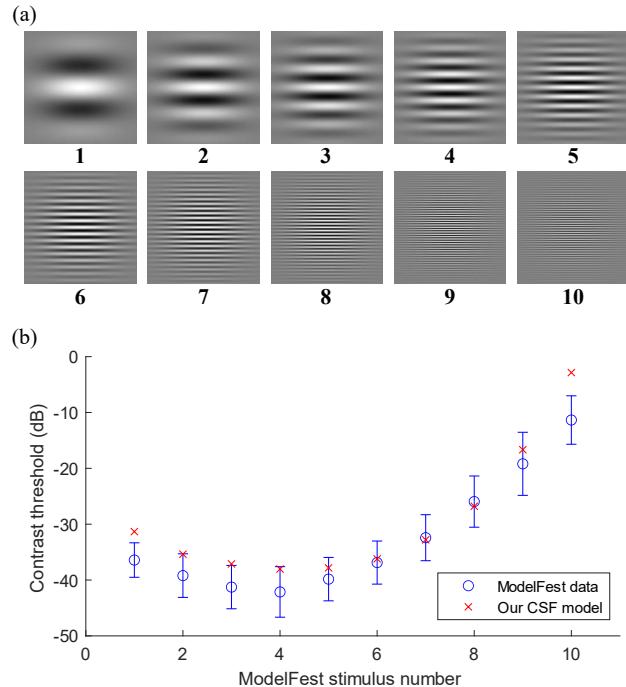
To avoid the risk of overfitting, we carefully designed our model with combinations of monotonically increasing or decreasing functions for each parameter, rather than using a high-order polynomial regression. For instance, we model our relative sensitivity function Equation (4) as a linear function when the luminance ratio decreases in the log-log domain and model it as a simple quadratic function when the luminance ratio increases. 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 ( $\Delta$ RMSE=RMSE<sub>test</sub> - RMSE<sub>train</sub>) is smaller for the practical model ( $\Delta$ RMSE=0.95 dB), compared with the full model ( $\Delta$ RMSE=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 with ModelFest dataset.** We compare our CSF model with an existing dataset that contains contrast sensitivity thresholds for different stimuli to validate our experiment and model. We use the ModelFest dataset [?], which contains 43 stimuli with various sizes (at most  $2.133^\circ$ ) and different shapes. We only use the first ten stimuli since, similarly to the ones in our experiment, they are composed of sinusoidal shapes with fixed sizes and a single frequency, while stimuli 11–43 have varying sizes of stimuli, mixed frequencies, or complex images, which cannot be directly predicted by our CSF model. As shown in Figure 5, our CSF prediction qualitatively follows the trend of the ModelFest dataset. The RMS prediction error of our model to the ten stimuli is 3.98 dB,

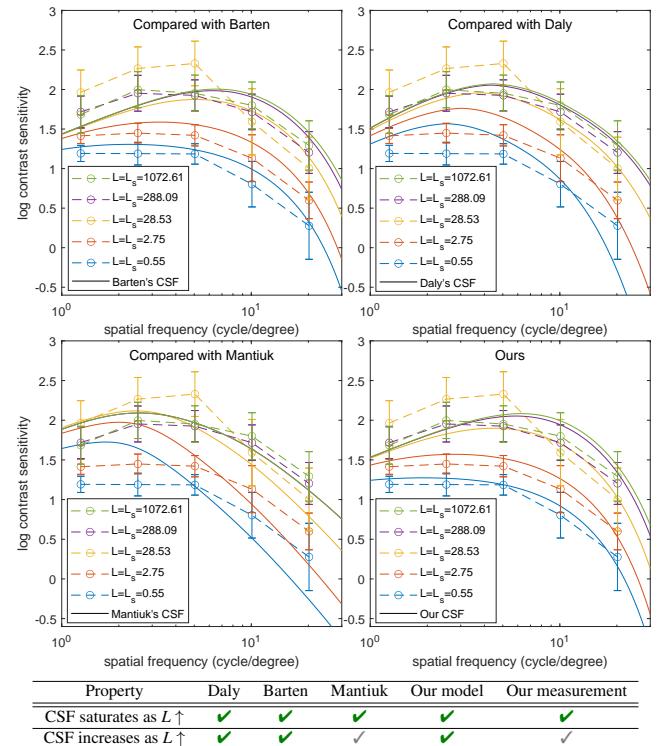
which is smaller than the standard deviation of the ModelFest measurements (4.18 dB), which intuitively indicates a good fit to our experimental data.

**Comparison with other CSF models.** We show in Figure 6 our measured CSF at  $L = L_s$  compared with Barten [?], Daly [?]’s, and Mantiuk et al. [?]’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 [?, ?, ?]. Second, we also find that the measured sensitivity tends to increase as the luminance level of the stimuli rises, with the exception of  $L = 28.53 \text{ cd/m}^2$ . This luminance level produces the highest sensitivities for low spatial frequencies ( $u=1.26, 2.52, 5.04 \text{ cpd}$ ), as also observed in the data collected by Mantiuk et al. [?] (Figure 6, bottom). Both Mantiuk et al.’s and our observations found the highest sensitivity levels of low frequency bands to span luminance levels between  $20 \text{ cd/m}^2$  and  $28.53 \text{ cd/m}^2$ . This may be related to the interplay between cones and rods at mesopic vision levels ( $10^{-1}$  to  $10^1 \text{ cd/m}^2$  [?]).



**Figure 5:** (a) First ten stimuli of the ModelFest dataset which consist of Gabor patterns of fixed sizes, and (b) comparison of our full CSF model’s prediction with the ModelFest dataset.

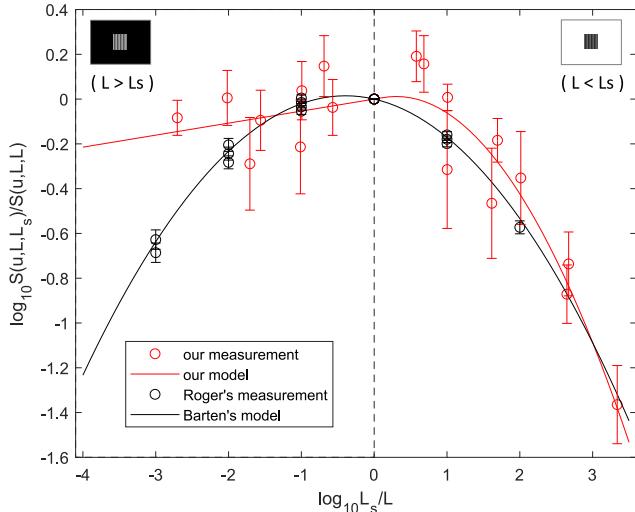
To our knowledge, there is only one other CSF model that takes into account the influence of the surrounding luminance  $L_s$  [?]. Barten’s surround-aware CSF is based on measurements from a technical report by Rogers and Carel [?], who measured just three subjects while analyzing dashboard visibility in airplanes. Both Barten’s 2003 model and our practical model rely on Barten’s 1992 model, and in particular on the function  $S_B^0$ . The key difference is the inclusion in our model of the relative contrast sensitivity term



**Figure 6:** 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 [?], Daly [?], Mantiuk et al. [?], and our full CSF models. Dashed lines with circles markers represent our measurement and solid lines represent the CSF models. Each of five colors of both dashed and solid lines indicates five levels of surrounding luminance  $L_s$  as shown in the legend. 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. [?]’s CSF model and our measured CSF increase as  $L$  increases except for  $L = 20 - 30 \text{ cd/m}^2$ , while “✓” indicates consistent agreements.

$R_p(L^*)$ . As shown in Figure 7, as  $L_s/L$  decreases, our model decays linearly with a small slope in log-log scale, whereas Barten’s model decays quadratically. The dataset acquired by Roger and Carel [?] differs from our dataset significantly when the surround luminance  $L_s$  is low. A potential explanation is the large difference in experimental environments. While we use for our measurements a modern HDR display, their measurements were taken using an airplane dashboard with a function generator and white masks to produce their stimuli [?]. Additionally, their measurements may be affected by phosphor persistence and video bandwidth of CRTs [?]. We show in the following section that this key difference has a strong impact in resulting applications.

**Comparison with HDR-VDP-2.2.** Whereas our CSF model directly handles surrounding luminance as a variable, other works such as HDR-VDP-2.2 [?, ?] and the work of Vangorp et al. [?]



**Figure 7:** Comparison of our relative sensitivity  $R_p$  of our practical model and Barten’s [?], 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).

have handled the effect of surround using models of glare, which is caused by intra-ocular light scattering. In this context, in order to show the usefulness of our model, we show predictions of our measured stimuli using HDR-VDP-2.2, which contains both models of glare and (not surround-aware) CSF.

HDR-VDP predicts visibility differences between test and reference images, which are the inputs to the method. The method then outputs detection maps consisting of the prediction probability of each pixel  $P_{map}$  and single scalar detection probabilities  $P_{det}$ , which are computed by pooling the detection maps. In order to run HDR-VDP-2.2 with our measured data, we assign the stimuli described in Figure 2 with zero contrast of sinusoidal patterns as the reference images and the stimuli with non-zero contrast of sinusoidal patterns as the test images. We adjust the contrast of the test images until we reach a probability of detection  $P_{det} = 0.5$ . Then the adjusted contrast becomes a contrast threshold, which is the inverse of contrast sensitivity. The predicted contrast sensitivities of our experimental variables are reported as solid lines in Figure 8 (a). Figure 8 (b) shows several examples of detection maps. While HDR-VDP-2.2, which implements the glare model, roughly predicts the trend of our measured surround-aware CSF, the prediction errors with our train and test sets ( $RMSE = 7.36\text{dB}$  and  $RMSE = 7.90\text{dB}$ ) are considerably higher to those of our CSF models. Additionally, our models provide analytic surround-aware CSF functions which can be easily computed, as opposed to predicting contrast sensitivity by iteratively running HDR-VDP-2.2 as described above.

## 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

video compression, tone-mapping and prediction of visual differences using our practical CSF<sup>†</sup>.

### 5.1. HDR Video Compression

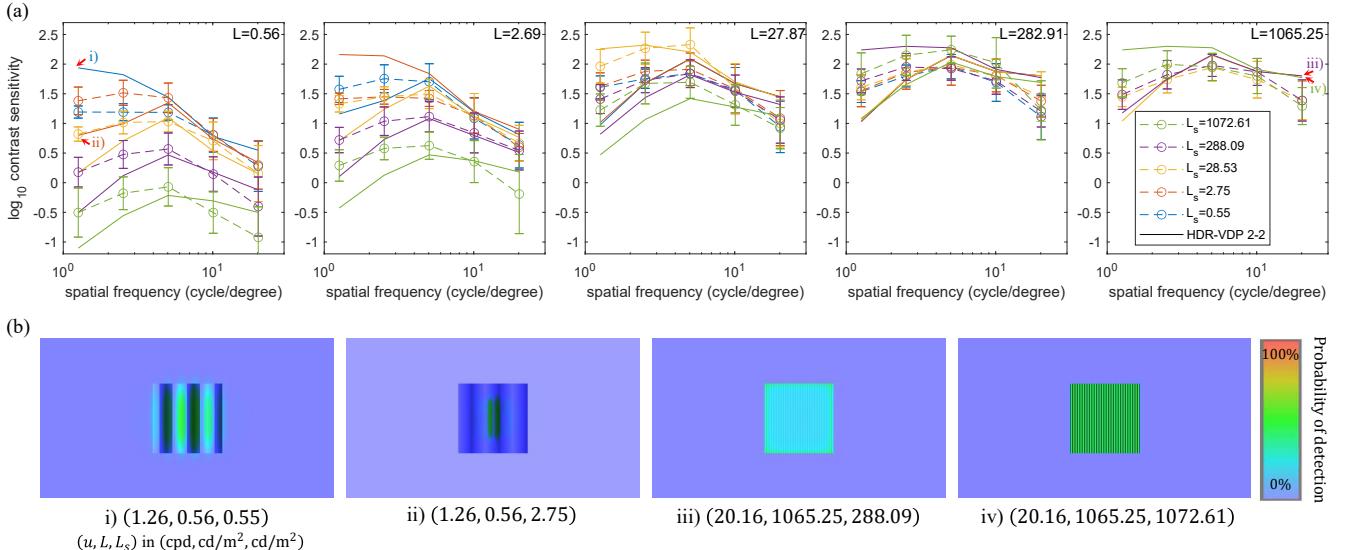
Current compression methods rely on the traditional integer-based framework; this requires that float-based HDR video content be converted to integers before compression. Quantization artifacts are thus inevitable in the existing video compression workflow described by the ITU-R recommendation [?]. The transfer function of the ITU-R standard compression pipeline is derived from Barten’s CSF model [?], and determines which luminance levels are allocated larger bit rates to avoid visible quantization artifacts. Miller et al. [?] 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 [?]. 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 9 compares the results of our video compression with the standard video compression (ITU-R), and using Barten’s subsequent model [?]. 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.

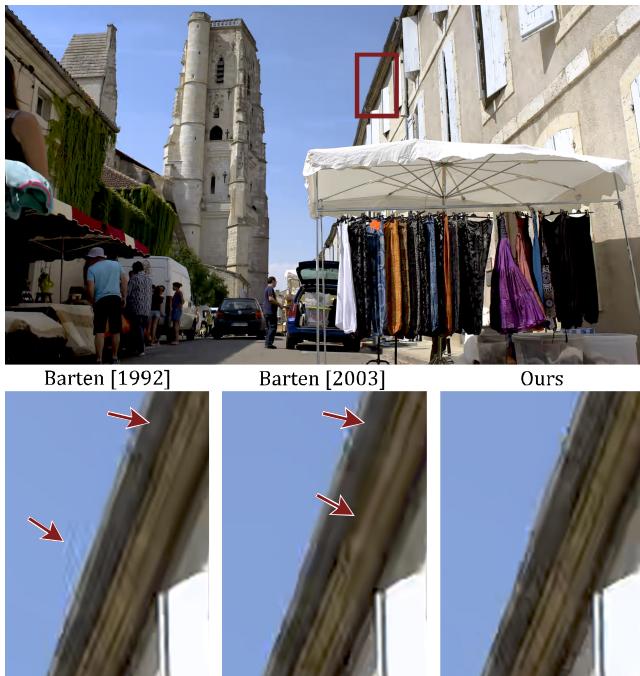
### 5.2. HDR Tone Mapping

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 [?, ?]. Many tone-mapping operators adopt existing CSFs such as Daly’s [?] or Barten’s [?]. As a proof-of-concept application, we apply our practical CSF model to Mantiuk’s operator [?], 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 recent surveys as one of the best performing algorithms [?, ?]. 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 10 shows how Barten’s model [?] fails to produce good tone mapping results, due to its excessive drop in contrast sensitivity for scenes with dark surround luminance (shown in Figure 7). Figure 11 shows the results, compared to the latest implementation of Mantiuk’s operator [?]. 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.

<sup>†</sup> 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:** Predicted contrast sensitivity of our experimental stimuli using HDR-VDP-2.2. The top row (a) shows our measurements (dashed lines with circle markers) and the predicted contrast sensitivity by HDR-VDP-2.2 (solid lines). Each of the five colors of both dashed and solid lines indicates each of the levels of surrounding luminance  $L_s$  as shown in the legend. The bottom row (b) shows several examples of detection maps obtained by HDR-VDP-2.2, which correspond to the red arrow markers in (a). It intuitively shows how HDR-VDP-2.2 detects contrast. Note that the detection maps are cropped at the center. Refer to Figure 3 for the predicted contrast sensitivity by our model.



**Figure 9:** HDR video compression results, using Barten’s models ([?] and [?]), and our surround-aware CSF for QP=37. Our novel CSF allows to preserve details better, thus reducing visible artifacts. Refer to the supplementary video.

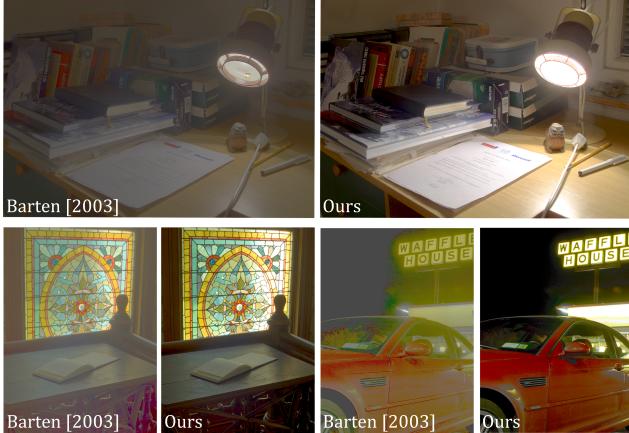
### 5.3. HDR Visual Difference Prediction

Many existing metrics to predict visible differences between images rely on a model of the human visual system (HVS), includ-

ing contrast sensitivity (e.g., [?, ?, ?, ?, ?]). According to a survey by Hanhart et al. [?], HDR-VDP-2 [?] is one of the most reliable metrics. However, it is based on a CSF which does not contain surrounding luminance as a control variable. We integrate our CSF model in the last version of this metric (HDR-VDP-2.2 [?]), substituting their original CSF with our surround-aware model. Figure 12 shows the result, using images from Cadík et al.’s dataset [?]. Although there are more recent user studies about visual difference of videos that measure the mean opinion scores (MOS) of scenes, such as [?], we use the dataset from [?] because it provides grid-based subjective maps for each scene. 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 12, 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 12, bottom). The average of the correlation coefficients of our method with six scenes is 0.5023 while that of HDR-VDP-2.2 is 0.3807. Additional results and correlation coefficients between the predicted probability maps and subjective maps are reported in Section 4.2. in the supplementary material.

## 6. Discussion

In our work we measure and model the effect of *surrounding luminance* in the CSF. Following conventions from the CIECAM97 model, our experiment covers the luminance of the entire *background* and a part of *surround* (ambient illumination). Some existing works have previously modeled and taken into account this

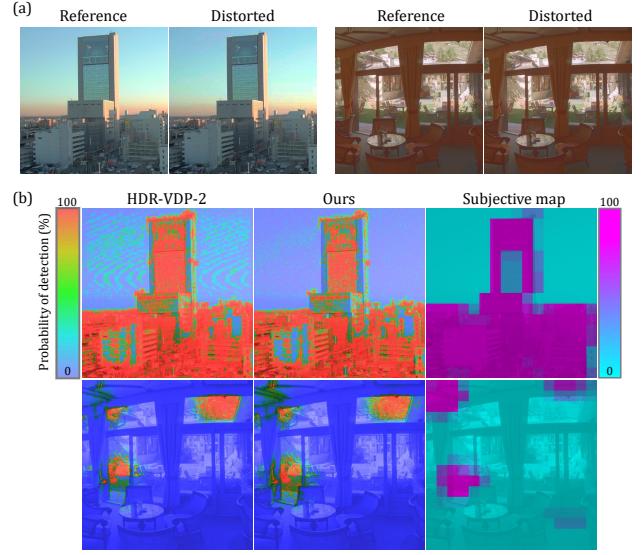


**Figure 10:** Comparison of tone mapping results using Barten’s surround-aware CSF [?] and our practical model. Barten’s CSF heavily drops for scenes with dark surround luminances (Fig. 7), yielding results with very low contrast.



**Figure 11:** Tone mapped images using Mantiuk’s [?] 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).

effect. Barten [?]’s surround-aware CSF is the only model that considers experimental measurements of the influence of surrounding luminance and its interplay with stimuli luminance [?]. This model has been used in several applications, for example, Kerofsky et al. [?] applies this CSF model in the context of image-quality metrics. HDR-VDP [?, ?] instead incorporates the effect of surrounding luminance by including a model of glare. We have shown in Sec. 4.4 that these two models cannot fully account for our collected measurements which are, to our knowledge, the first that cover a large range of background and stimuli luminances on a state-of-the-art HDR display rigorously calibrated. Other works in the field of image processing have also taken into account the perceptual effects of ambient illumination, for example on perceived quality of tone-mapped images [?] or perceived image quality on displays under energy-saving mode or high ambient illumination [?]. An interesting avenue of future work for improving image processing methods would be to take into account both the background and ambient illumination levels and their interplay.



**Figure 12:** (a) Scenes from [?], used for comparison of visible difference prediction. The images are HDR, so they are tonemapped for visualization. (b) 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 [?]. 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.

**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 [?] and [?], respectively. As stated in a previous study [?, ?, ?], 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.

**Impact of viewing distance.** The viewing distance for our experiment (1.25 m) is close to that of several previous works, such as 93 cm for Mantiuk et al. [?]’s CSF measurements and 1.32 m for Vangorp et al. [?]’s local adaptation experiment. In order to achieve a larger surrounding region in our stimuli, our experiment was designed with a relatively close distance. Far viewing distances, such as in a cinema, could affect the performance of our CSF due to accommodation effects [?, ?].

**Effect of the edge of the stimuli.** Recent experiments on CSFs [?, ?, ?] employ Gabor patches, which include sinusoidal patterns inside Gaussian envelopes, in order to avoid participants detecting the edges of the sinusoidal patterns rather than the patterns themselves. In our experimental setup, we assume that the potential influence of the edges of the frequency stimuli is negligible. Explicitly quantifying this effect on the surround-aware CSF using Gabor-based stimuli can be an interesting avenue for future work.

## **7. Conclusion**

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 [?]), and previous attempts to characterize the effect of surrounding luminance (which involved very limited luminance ranges [?, ?]). 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.

## **Acknowledgment**

Min H. Kim acknowledges the support of Korea NRF grant (2019R1A2C3007229) in addition to ETRI grant, Samsung Electronics, MSIT/IITP of Korea (2017-0-00072), National Research Institute of Cultural Heritage of Korea (2021A02P02-001), and Samsung Research Funding Center (SRFC-IT2001-04) for developing partial 3D imaging algorithms. Diego Gutierrez acknowledges funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation program (CHAMELEON project, grant agreement No 682080), and from the European Union's Horizon 2020 research and innovation program under the Marie Skłodowska-Curie grant agreements No 765121 and 956585.