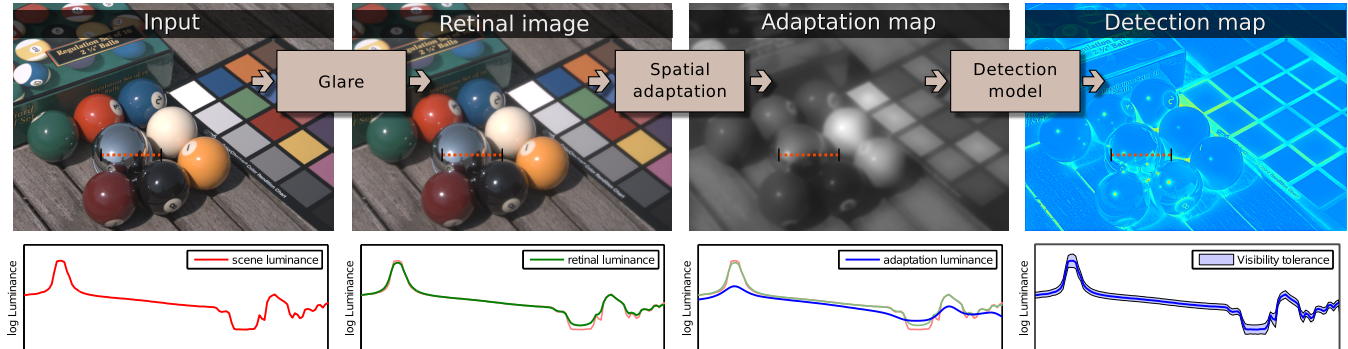# A Model of Local Adaptation

Peter Vangorp[*]
Bangor University, UK &
MPI Informatik,
Germany

Karol Myszkowski[†]
MPI Informatik,
Germany

Erich W. Graf[‡]
University of Southampton,
United Kingdom

Rafał K. Mantiuk[§]
Bangor University, UK &
The Computer Laboratory,
University of Cambridge, UK

Figure 1: *Processing steps of our spatial adaptation model. First, optical glare is simulated to produce a retinal image. Then, the local luminance adaptation map is computed using our novel adaptation model. The plots below show the luminance profile for the pixels marked with the dashed-orange line. Note that the eye cannot adapt to small highlights as shown by the flattened blue curve in the "adaptation luminance" plot. As one of the applications, the adaptation map can be used to estimate the smallest visible contrast in complex images (detection map) and therefore represents a visibility tolerance for each pixel.*

Abstract

The visual system constantly adapts to different luminance levels when viewing natural scenes. The state of visual adaptation is the key parameter in many visual models. While the time-course of such adaptation is well understood, there is little known about the spatial pooling that drives the adaptation signal. In this work we propose a new empirical model of local adaptation, that predicts how the adaptation signal is integrated in the retina. The model is based on psychophysical measurements on a high dynamic range (HDR) display. We employ a novel approach to model discovery, in which the experimental stimuli are optimized to find the most predictive model. The model can be used to predict the steady state of adaptation, but also conservative estimates of the visibility (detection) thresholds in complex images. We demonstrate the utility of the model in several applications, such as perceptual error bounds for physically based rendering, determining the backlight resolution for HDR displays, measuring the maximum visible dynamic range in natural scenes, simulation of afterimages, and gaze-dependent tone mapping.

CR Categories: I.3.0 [Computer Graphics]: General—;

Keywords: perception, local adaptation, tone mapping, visual metric, high dynamic range, glare

This version of the paper contains an amendment in Section 4, Equation 6, which clarifies the derivation of the threshold elevation function. This amendment is not included in the paper

[*]e-mail: p.vangorp@bangor.ac.uk
[†]e-mail: karol@mpi-inf.mpg.de
[‡]e-mail: e.w.graf@soton.ac.uk
[§]e-mail: mantiuk@bangor.ac.uk

1 Introduction

Luminance adaptation is a fundamental mechanism of the visual system, that enables us to see in drastically varying illumination conditions. The mechanism is so crucial that most visual models must be provided with the actual value of adapting luminance to produce correct predictions. Examples of such models are contrast sensitivity functions (CSFs), appearance models (e.g., CIECAM02), and many perception-inspired tone mapping operators [Ferwerda et al. 1996; Pattanaik et al. 2000; Irawan et al. 2005].

In classical psychophysical experiments the state of adaptation is controlled by displaying stimuli on a uniform adapting field. Whilst such a simplified stimulus is effective in enabling the isolation and discounting of adaptation effects, it does not reflect the complex spatial light distribution of real-world scenes or images shown on high contrast (HDR) displays. In such complex scenes the state of adaptation is generally unknown. This obviously hinders the application of visual models to complex images and necessitates ad hoc assumptions about the adaptation state.

There are a number of elaborate cone and retina adaptation models [Finkelstein et al. 1990; Wilson 1997; van Hateren 2005] that are based on neurological and psychophysical measurements of the retina. Such models have been adapted, often in simplified form, in graphics for tone mapping [Pattanaik et al. 2000; Irawan et al. 2005; van Hateren 2006] or for simulating afterimages [Ritschel and Eisemann 2012; Jacobs et al. 2015]. These models, however, capture mostly temporal aspects of adaptation and are not capable of predicting how the state of adaptation varies when the gaze moves from one part of the scene to another. In contrast to the previous work, we study the effect of spatial pooling on adaptation, assuming that the adaptation mechanism is in a steady state. There is little work

on the nature of spatial pooling on local adaptation. Even though some models account for pooling in horizontal and amacrine cells [Wilson 1997], they make assumptions about the spatial interactions of the cells and they have not been validated against psychophysical data.

In this work we propose a novel model of local adaptation, based on new psychophysical measurements on a high dynamic range (HDR) display that was specifically designed for this purpose. The best-fitting model is found by exhaustive search of the space of possible models and then by cross-validation on an independent dataset that was generated to maximally differentiate between the models. The local adaptation model leads to a simple and efficient predictor of the smallest noticeable differences in images. We show the application of our local adaptation model and detection threshold predictor on several examples, including deriving error bounds for physically based rendering, determining the backlight resolution for HDR displays, measuring the maximum visible dynamic range in complex natural scenes, simulating afterimages, and gaze-dependent tone mapping.

To derive our model of local adaptation, we make several assumptions. Our goal is to build an empirical model capable of explaining our psychophysical data, rather than trying to model the underlying biological mechanisms. To simplify our task, the adaptation state is predicted only for the central part of the fovea. The adaptation pools are likely to be larger for parafoveal vision, where the signal is pooled from a number of photoreceptors. Then, we assume that the eye is fixated on a target and it reaches a steady adaptation state. We do not model time-course of adaptation as there is a number of existing models, which can be combined with ours. Finally, we consider mostly photopic luminance in the range from $1\,\mathrm{cd/m^2}$ to $5000\,\mathrm{cd/m^2}$.

The main contributions of this work are:

- novel measurements of the spatial extent of local adaptation on a high-resolution HDR display;

- a new efficient method for selecting the optimal model from an exhaustive set of potential complex non-linear models;

- a novel model of local adaptation that explains how the extent of local adaptation varies with absolute luminance levels;

- a simple and efficient predictor of detection thresholds, providing conservative error bounds on distortions in complex images.

2 Related Work

Sites of adaptation in the eye and the retina   The visual system operates in an environment where light intensity can vary enormously. The 10^9 change in illuminant between night and day is striking, but the visual system must also cope with changes up to 10,000 fold between the dark shadows and bright highlights of a single scene. The output of the retina is about 100 fold less than this, and the effectiveness of our visual system in daylight conditions relies upon pre-retinal processes, as well as adaptive mechanisms within photoreceptors and post-receptor mechanisms. The pupil contracts to increases in illumination, which helps to reduce retinal luminance by a moderate factor of 8 (often much less). Further, the light entering the eye is scattered in the optics (cornea and lens) and on the retina causing glare [Vos and van den Berg 1999; IJspeert et al. 1993; Deeley et al. 1991]. Glare has an important effect on adaptation as it serves to elevate the luminance of dark parts of a scene relative to its bright neighborhoods.

Retinal light adaptation amplifies weak sensory signals and prevents strong signals from saturating neural responses. In cone-mediated pathways, adaptation at lower light levels occurs beyond the photoreceptors at the point where cone bipolar cells synapse on to ganglion cells. At higher light levels, where the benefits of spatial pooling are less evident, adaptation shifts to the cones themselves [Dunn et al. 2007]. Adaptation in the rod-mediated (scotopic) pathway occurs mainly through postreceptoral mechanisms, perhaps at the level of bipolar cells [Dunn and Rieke 2008]. These factors, along with consideration of the differences in the distribution of rods and cones across the retina, yield a picture of the spatial extent of light adaptation in different lighting conditions. In the central part of the fovea each cone is wired to an individual midget bipolar cell and one foveal midget ganglion cell for each bipolar cell [Ahmad et al. 2003], whereas the spatial adaptation pool for rods in the human retina can be 10 minutes of arc in diameter [Hess et al. 1990, p. 82].

Psychophysics of luminance adaptation   Light and dark adaptation have been extensively studied using psychophysical methods — for a review, refer to [Barlow 1972]. In a study with similar goals to ours, Westheimer [1967] investigated the size effect of a disk-shaped patch on the increment threshold for a small, briefly flashing stimulus that is placed in the disk center. He demonstrated that illumination of retinal regions in the immediate neighborhood of the stimulus acts to raise the adaptation level, and different intensity surround beyond the disk boundary acts to lower it. This effect stabilizes beyond the disk diameter over $0.25°$ [Westheimer 1967, Figure 3]. Our experiments are inspired by this work, but we consider much more complex backgrounds and both increments and decrements. In follow-up work [McKee and Westheimer 1970] similar considerations have been performed for chromatic channels in the stabilized fovea.

Ad hoc local adaptation models   Many tone mapping operators and color and image appearance models assume that the eye is adapted to a "global" luminance level regardless of the gaze position [Ward 1994; Ferwerda et al. 1996; Fairchild 1998]. Such a global adaptation state is computed as an arithmetic or geometric average luminance to partially account for the non-linear response of the HVS to light [Ward 1994; Reinhard et al. 2002]. To find a local adaptation state, Chiu et al. [1993] and Jobson et al. [1997] compute low-pass filtered images. However, the spatial support of such filters is chosen ad hoc. When such adaptation maps are used for tone mapping, they may result in halo artifacts. Edge-stopping filters limit such spatial processing to the regions with homogeneous pixel intensity, which greatly reduces the halo artifacts, but also ignores glare due to bright pixels in the proximity of high-contrast edges. The spatial support of such filters is typically fixed [Ledda et al. 2004; Kuang et al. 2007] (e.g., to 2% of the image size) or adaptively expanded as a function of the local variability in pixel intensities [Reinhard et al. 2002].

Based on the work of Moon and Spencer [1945], Larson et al. [1997] propose a foveal adaptation with a spatial extent of $1°$ (visual degree), computed as the arithmetic mean of luminance values within that extent. They build a histogram of the logarithm of such foveal adaptation values for all image pixels, and through its integration they derive a halo-free tone mapping curve that is local to a particular adaptation level, but does not account for spatial configurations of pixel intensities as in local tone mapping. The foveal adaptation by Larson et al. has been adopted in many follow-up works [Pattanaik et al. 2000; Irawan et al. 2005; Pajak et al. 2010], while in color appearance models for HDR images [Kim et al. 2009] even larger regions of $10°$ have been considered. In contrast to those, Pattanaik et al. consider one-fifth of the white-point luminance (as originally proposed in [Hunt 1995]), which they determine using the paper-white reflectance patch in the Macbeth ColorChecker chart. Tumblin et al. [1999] interactively select the local adaptation region around the fixation point so that an S-shaped global tone mapping

curve produces well-exposed images. Also, per-pixel adaptation is assumed in tone mapping [Schlick 1995] and in detecting perceivable differences between images [Mantiuk et al. 2011], which is overly conservative and overestimates sensitivity to contrast details. Reinhard and Devlin [2005] found that a linear combination of per-pixel and global adaptation leads to high-quality tone mapped images.

The existing approaches to local adaptation in tone mapping, color appearance, and image quality evaluation are clearly ad hoc, even if they do refer to perceptual findings. For example, the Naka–Rushton response was originally derived as a function of adaptation state for single receptors [Naka and Rushton 1966; Valeton 1983], but its variants are commonly used for larger adaptation areas, sometimes the entire image, without any in-depth justification [Tumblin et al. 1999; Pattanaik et al. 2000; Irawan et al. 2005; Ledda et al. 2004; Reinhard et al. 2002; Kim et al. 2009; Reinhard and Devlin 2005]. In this work, we propose a perceptually grounded model of local adaptation that accounts for the spatial configuration of HDR pixels, and we show that its use can be beneficial in many different applications (Section 7).

Impact of surround on lightness perception   Although lightness perception is not the focus of this work, the stimuli and methods used in lightness experiments share many similarities to our experiments. Radonjic et al. [2011] investigate the luminance-to-lightness mapping for a test patch surrounded by Mondrian-like checkerboard stimuli, and postulate that the Naka–Rushton-like model [Naka and Rushton 1966; Valeton 1983] might explain the collected data but do not provide any specific model parameters as a function of the surround configuration. Allred et al. [2012] extend this work by considering the influence of two rings of such Mondrian patches. They found that a darker surround ring makes the test patch brighter, further surround influences brightness less, and consistently lower (or higher) luminance in the surround affects the test patch lightness stronger. They also observe rotational symmetry, i.e., for a given set of surrounding patches with different luminance, their particular layout does not influence the test patch lightness. Our luminance adaptation experiments employ similar stimuli (Experiments 5 and 6), but aim at measuring the influence of surround on the detection thresholds rather than lightness.

3   Experiment 1: Probe-on-Flash

We start the discovery of the new local adaptation model with a probe-on-flash experiment, which will introduce our experiment set-up and demonstrate how maladaptation affects visual performance. In Section 4 we use the results of this experiment to motivate a simple detection model. The model contains an unknown spatial adaptation component, which we discover by first collecting data in a series of experiments (Section 5) and then fitting an exhaustive set of candidate models (Section 6). The best performing models are then validated and discussed. For brevity, the experiments are only briefly described and discussed. Refer to the supplementary materials for more details on the experiments and the discussion of the results.

A classical Probe-on-Flash psychophysical paradigm offers a method for measuring visual system performance when the eye is adapted to different luminance than the luminance of the background [Geisler 1978; Hood et al. 1979; Craik 1938]. We use this paradigm to investigate how the mismatch between background and adaptation luminance affects the visual performance across luminance levels.

Apparatus   To achieve the high brightness and local contrast level required for the measurements, we built a custom high dynamic range display with a projector backlight (Figure 2) using a similar
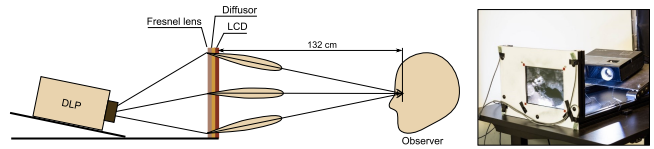

Figure 2: *The schematics and a photograph of the HDR display used in the experiments.*

design as in [Seetzen et al. 2004] but with a number of improvements. A 9.7" Apple iPad "retina" LCD panel with a resolution of 2048×1536 served as a front modulator. It ensured that the angular resolution surpassed the maximum resolvable resolution of the eye (in excess of 240 pixels per visual degree for the viewing distance of $1.32\,\mathrm{m}$). The backlight was produced by a $3500\,\mathrm{lm}$ Acer P1267 DLP projector with a resolution of 1024×768, from which the color wheel was removed to increase brightness. To maximize display efficiency, the light coming from the projector was directed towards the observer using a Fresnel lens. The projector was focused on the LCD panel to maximize backlight resolution. However, a diffuser with a custom adjusted spacer was introduced to eliminate diffraction and Moiré patterns coming from the matrices of pixels in the DLP and LCD. The contrast of each of these two components was approximately $1000{:}1$; their combined contrast was measured to be in excess of $750\,000{:}1$ and the maximum luminance was above $5000\,\mathrm{cd/m^2}$. For conditions involving luminances below $0.5\,\mathrm{cd/m^2}$ the display luminance was boosted by a factor of 100 and observers wore ND 2.0 glasses (1% transmittance) to compensate. The display was calibrated for accurate absolute luminance reproduction using custom software and a JETI Specbos 1211 spectroradiometer. A custom display algorithm was implemented using OpenGL and GLSL to enable real-time display of arbitrary images.

Stimuli   This experiment is meant to induce a state of as strong a maladaptation as possible. To achieve this goal, the observers adapted to a uniform adaptation field of luminance L_f most of the time. A detection target was only briefly flashed for $200\,\mathrm{ms}$ in a probe-on-flash experiment (Figure 3 right) similar to [Westheimer 1967]. Such a short presentation time and relatively small detection target prevent the eye from adapting to the pedestal luminance L_p. However, since the neural adaptation mechanism can respond in less than $50\,\mathrm{ms}$, we cannot guarantee that the observers remained fully adapted to the uniform field L_f during the flash.

The detection target (Figure 3 top-right) was a horizontal or vertical step edge modulated by a Gaussian envelope and shown on a pedestal of $0.2°$ diameter with luminance L_p. The polarity of the edge (the order of the dark and bright side) was randomized between trials. The advantage of an edge over other detection targets is that it can be made very small and, unlike Gabor patches, an edge consists of similar frequency spectra as edges in natural images.

Procedure   Observers were asked to look at the center of the screen, where the detection target would appear, during the whole experiment to maintain adaptation. A faint fixation circle of $0.2°$ diameter appeared briefly before the onset of the edge. The 2-alternative forced-choice (2AFC) task of the observer was to determine whether the edge (detection target) was oriented horizontally or vertically. The detection threshold was found using the QUEST procedure [Watson and Pelli 1983]. At least 40 QUEST trials per observer were used to determine each threshold. Before proceeding to the next stimulus, the observer adapt for 1 to $3\,\mathrm{min}$, depending on whether bright- or dark-adaptation was required. Each threshold was measured for at least 5 observers and averaged. Observers were 20–40 years old and had normal or corrected-to-normal visual acuity.
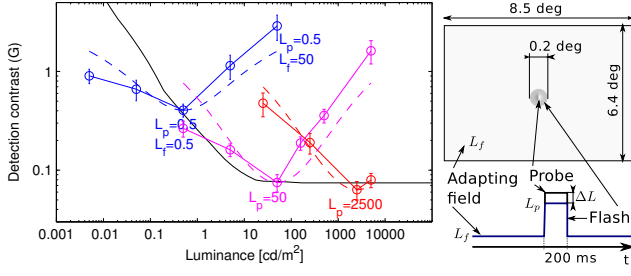
Figure 3: *The results (left) of the controlled maladaptation experiment, the stimulus (top-right) and its presentation (bottom-right). The experiment results (solid lines) are plotted as a function of varying adaptation luminance $L_a \approx L_f$, and one of three fixed levels of pedestal luminance L_p. Error bars represent the within-observer standard error of the mean (SEM). The black line is the tvi function (plotted as logarithmic contrast). The dashed lines show the threshold elevation predicted by the Naka–Rushton photoreceptor response model.*

Results   The results of the experiment are shown in Figure 3 (left). The threshold curves (solid lines) for each pedestal luminance L_p have their minimum at the fully adapted state ($L_p = L_f$). The predictions of the classical Naka–Rushton [1966] model, which will be discussed in Section 4, are applied directly to the incoming luminance field (dashed lines). One salient difference is that our measurements form asymmetric curves, while the Naka–Rushton model predicts symmetric elevation of the thresholds. In the next section, we discuss how this asymmetry can be explained by glare.

4   Detection Model

To explain the results from our probe-on-flash experiment, we design a simple detection model. The detection model will be the basis for finding the model of local adaptation.

The three parabolas in Figure 3 for three pedestal levels (L_p) are shifted vertically to each other. This shift is caused by the loss of sensitivity of photoreceptors at low luminance levels. Since the trough of each parabola represents detection on a uniform field (no flash), it should be possible to predict that case with an ordinary sensitivity model, such as a contrast sensitivity function (CSF). A CSF predicts sensitivity S, which is the inverse of the detection contrast $\Delta L / L$; hence:

$$S = \frac{L}{\Delta L} = CSF(\rho, L) \quad \Rightarrow \quad \Delta L = \frac{L}{CSF(\rho, L)} = tvi(L) \tag{1}$$

where ρ is spatial frequency and L is the luminance of the uniform background. For simplicity, we fix the spatial frequency ρ to the frequency of the largest amplitude in our stimulus (5.5 cpd) so that we can define the threshold-versus-intensity (tvi) curve as a function of luminance. Such a function is plotted a black line in Figure 3. Note that for better presentation, the figure plots logarithmic contrast G [1] instead of luminance increments. The curve represents the smallest detectable difference in luminance when the eye is fully adapted to the luminance level L. We use the CSF from [Mantiuk et al. 2011].

The CSF predicts the detection threshold for the case when the stimulus is presented on a uniform field of certain luminance L. However, our patterns (Figure 3 top-right) are mostly non-uniform so there is no easy way to find L. In some literature L is said to be

---

[1] Logarithmic contrast is the log difference between two luminance levels: $G = \log_{10}(L + \Delta L) - \log_{10}(L)$.
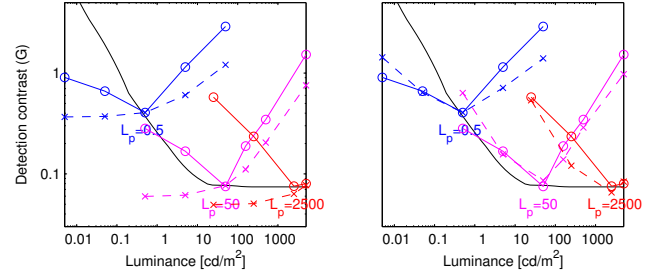
Figure 4: *Prediction of detection thresholds (dashed-lines) from Figure 3 using the model combining tvi and glare (left); and maladaptation in addition to those two components (right).*
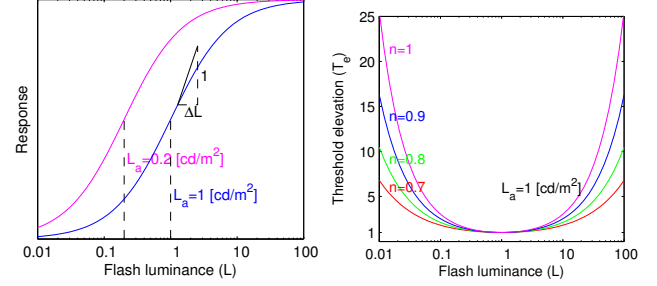
Figure 5: *Left: Naka–Rushton photoreceptor response model. The slope of the curve is inversely proportional to the detection thresholds. Right: Threshold elevation due to maladaptation for flash luminance L (x-axis) and fixed adaptation luminance ($L_a = 1$), as predicted by the Naka–Rushton model (Equation 6).*

adaptation luminance. However, the results in Figure 3 demonstrate that it cannot be the case. Take for example the left-most blue point in Figure 3, where the observer adapted to $L_a = 0.005 \, \text{cd/m}^2$ and was presented with the pedestal of a detection target with the pedestal of $L_p = 0.5 \, \text{cd/m}^2$. If the adaptation luminance (close to $0.005 \, \text{cd/m}^2$) was to predict the threshold, the measurement should be on or above the black tvi line. However, it is well below the line, as if the detection threshold would be influenced more by the pedestal luminance L_p than by the adapting field luminance L_f.

Based on the above observations, we propose that the tvi is *not* a function of adaptation luminance, but instead the tvi is a function of retinal luminance [2]. The retinal luminance is the measure of light reaching the retina and it accounts for the light that is scattered in the optics and on the retina (glare). To compute retinal luminance L_O, we need to convolve the incoming luminance image I with a point spread function (PSF) due to the glare effect, which in this paper we call the glare spread function (GSF) O:

$$L_O = I * O. \tag{2}$$

Figure 4 (left) shows the prediction of the combined effect of glare and tvi: $\Delta L = tvi(L_O)$, using the GSF from CIE Recommendation 135/1-6 [Vos and van den Berg 1999]. The right part of each curve matches the measurements better, demonstrating that the model of glare can explain the elevated threshold when a dark detection target is placed on bright surround. But it cannot explain the opposite situation, when $L_f < L_p$. To account for that case, we need a model of maladaptation.

Maladaptation   There is much evidence suggesting that the response of the receptoral mechanism can be explained by the Naka–

---

[2] We do not consider here retinal illuminance in trolands as both CSF and tvi functions account for the effect of pupil size.
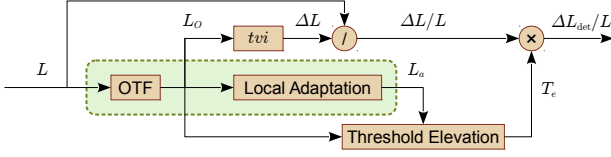
Figure 6: *The diagram of the detection model. The dashed-green area marks the unknown components that are discovered in Section 6.*

Rushton equation [Valeton 1983; Naka and Rushton 1966]:

$$R = k \frac{L^n}{L^n + \sigma(L_a)^n}, \qquad (3)$$

where L is the flash luminance, n is a constant between 0.7 and 1, and k is a scaling constant. $\sigma(L_a)$ is a semi-saturation which controls the translation of the response curve along the log-luminance axis as shown in Figure 5 (left). Electro-physiological readings of single receptors in the rhesus monkeys indicate that the semi-saturation constant is higher than background luminance at lower luminance levels and is approximately equal to background luminance at higher luminance levels [Valeton 1983]. Although these measurements are often cited in the graphics literature [Irawan et al. 2005; Pattanaik et al. 2000], they apply only to single cones and ignore any spatial and postreceptoral effects. Our psychophysical data depicted in Figure 3 reveals a very different characteristic: the response is strongest and the thresholds smallest when the eye is fully adapted to the background luminance. To position the Naka–Rushton curve to match this observation, the half-saturation constant must be equal to adaptation luminance: $\sigma(L_a) = L_a$.

The photoreceptor has the strongest response when the detection thresholds are the smallest, therefore the change in response (derivative of R) is inversely proportional to the detection threshold ΔL:

$$\frac{dR}{dL} \propto \frac{1}{\Delta L}. \qquad (4)$$

After differentiating Equation 3 and introducing into Equation 4, we get:

$$\Delta L(L, L_a) \propto \frac{(L^n + L_a^n)^2}{n\, L^{n-1}\, L_a^n}. \qquad (5)$$

For modeling purposes we are interested in the elevation of detection contrast ($\Delta L/L$) rather than absolute increments (ΔL). Furthermore, threshold elevation should be relative to the point of complete adaptation: $L = L_a$. Such threshold elevation due to maladaptation is given by:

$$T_e(L, L_a) = \frac{\Delta L(L, L_a)\, L_a}{\Delta L(L_a, L_a)\, L} = \frac{(L^n + L_a^n)^2}{4\, L^n\, L_a^n} \qquad (6)$$

and is shown in Figure 5 (right) for several values of the parameter n. Similar principles were used by Irawan et al. [2005] to derive a *tvi* function that accounts for adaptation (TVIA). The difference is that their approach required numerical computation while we derive an analytic solution.

Figure 4 (right) shows the predictions when the threshold elevation model (Equation 6) is introduced into our detection model, so that:

$$\Delta L_{\text{det}}(L, L_a) = tvi(L_O)\, T_e(L_O, L_a), \qquad (7)$$

where L_O is the retinal image from Equation 2. The complete detection model is illustrated in Figure 6. The model predicts reasonably well our simple experiment in which the adaptation luminance

L_a is controlled and thus approximately known (ignoring partial adaptation to the flash). However, in complex images L_a is usually unknown. In the following sections we address the problem of computing L_a for arbitrary complex images.
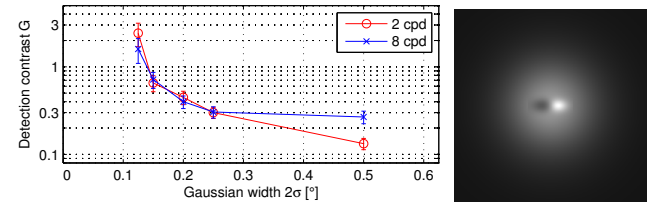
5 Local adaptation experiments

To find a model of local adaptation we conduct a series of experiments, each measuring a different aspect of the adaptation field. The experimental procedure was similar to Experiment 1, however, the pedestals remained visible the whole time and only the detection target (the same edge or a Gabor patch) was briefly displayed for 200 ms.

5.1 Experiment 2: Frequency selectivity

Some sources postulate that local luminance adaptation is pooled within a receptive field of a visual channel, which is tuned to a band of spatial frequencies [Shapley and Enroth-Cugell 1984]. If this is the case, the eye should adapt to differently sized pools of local adaptation when detecting targets of different spatial frequencies. To test this hypothesis, we investigate the detection of Gabor patches of two frequencies (2 and 8 cpd) on a Gaussian pedestal of varying size of the fixed maximum luminance of $500\,\text{cd/m}^2$. The background was a uniform field of $5\,\text{cd/m}^2$. An example stimulus and the measured thresholds are shown in Figure 7.

If the adaptation mechanism was tuned to different spatial frequencies depending on the detected target, we would expect to see vastly different shapes of the two curves. But given only subtle differences, we have no evidence for spatial selectivity of the adaptation mechanism. This result also supports the choice of an edge as a detection target for the following experiments, as it is not restricted to a single frequency and is a more representative stimulus for complex scenes.

Figure 7: *Experiment 1 stimulus (right) and results (left). The spatial pooling of the adaptation mechanism appears to be similar for Gabor targets of different spatial frequencies.*

5.2 Experiment 3: Extent

To measure the extent of the visual area that influences the adaptation luminance, edge targets were displayed on a disk-shaped pedestal of 2500 or $50\,\text{cd/m}^2$ of a variable diameter, on a background of $5\,\text{cd/m}^2$. Figure 8 shows that the detection threshold, and hence the adaptation luminance, levels off around a diameter of $0.5°$ of visual angle, which is smaller than the $1°+$ extent used in most ad hoc models but larger than the extent of about $0.1°$ proposed by Wilson [1997] based purely on retinal physiology. Similar leveling off around a diameter over $0.25°$ was found by Westheimer [1967], but he considered a much smaller flashing stimulus of $0.017°$ that was presented every second for $10\,\text{ms}$. The results indicate that the size of the adapting pattern has a significant effect on the state of adaptation.
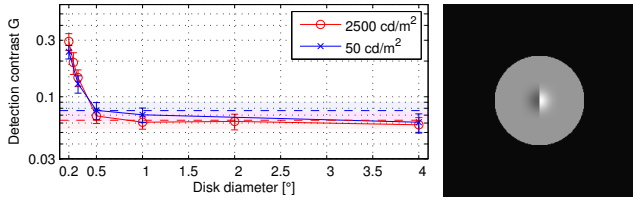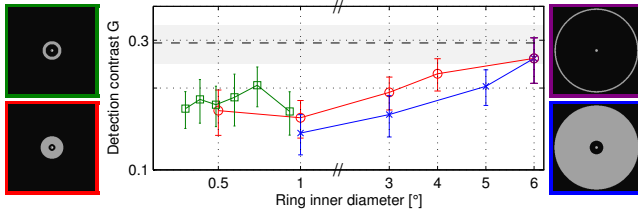
Figure 8: *The detection thresholds for targets on pedestals of different diameters levels off around $0.5°$ of visual angle. The horizontal dashed lines indicate the detection threshold for a pedestal that covers the entire screen (from Experiment 1).*

5.3 Experiment 4: Long-range effects

The disk pedestal in Experiment 3 can capture the extent of the pooling, but as the central part of the pedestal dominates the adaptation state, the measurements are not sensitive to the weak influence of luminance further away from the fixation point. To measure such likely, long-range effects, a $0.2°$ diameter pedestal of $2500\,\mathrm{cd/m^2}$ was surrounded by a concentric ring of the same luminance and varying inner and outer diameters on a background of $5\,\mathrm{cd/m^2}$. Three different groups of rings were tested: a) rings with a fixed area; b) rings with a fixed outer diameter; and c) rings of which the area increased with the inner diameter to compensate for the weaker effect of more distant regions. Refer to the supplementary materials for the exact specification of the stimuli.

Figure 9 shows that the long-range effect of rings of $6°$ diameter or more is negligible compared to no ring at all. The short-range effects are clearly dominant.
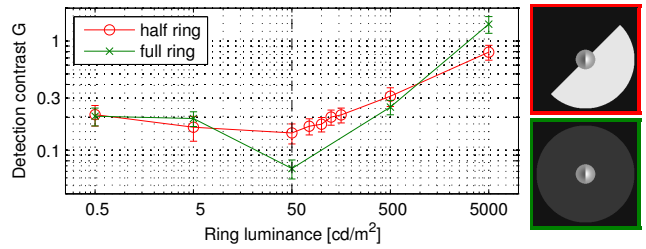

Figure 9: *The detection thresholds for targets surrounded by a ring with different inner diameters. The horizontal dashed black line and gray band indicate the detection threshold and SEM for a pedestal without a ring (from Experiment 3). Example stimuli from each curve are zoomed out to demonstrate the long-range extent of the adaptation. The small dot in the center of each stimulus is the $0.2°$ diameter pedestal with the target edge.*

5.4 Experiment 5: Non-linear pooling

The two previous experiments measured pooling as a function of distance from the fixation point. However, they cannot explain what kind of non-linearity is involved: pooling might occur in linear (luminance) space, in logarithmic space, or in any other non-linear space. To determine this non-linearity, the stimulus was flanked by a concentric half or full ring of $1°$ outer diameter. The luminance of this ring varied from 0.5 to $5000\,\mathrm{cd/m^2}$. The half ring was cut diagonally to reduce any possible interference with the vertical or horizontal detection target. The background was fixed at $0.5\,\mathrm{cd/m^2}$.

One salient feature of the results shown in Figure 10 is that the effect of adaptation is asymmetric for lower and higher luminance of the half-ring. This is further evidence of the strong effect of glare. However, the exact form of the non-linearity is difficult to determine without considering other elements of the adaptation model.
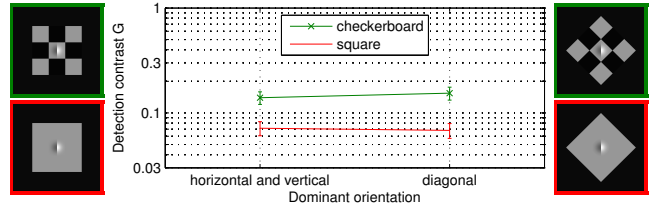

Figure 10: *The detection thresholds for targets flanked by a half or full ring with different luminances. The pedestal luminance is indicated with a vertical dashed line.*

5.5 Experiment 6: Orientation and contrast masking

The sharp contrast edge between the pedestal and background not only elevates adaptation luminance, but it also creates a strong masking signal. Contrast masking mostly affects the detection of signals of similar spatial frequency and orientation. To vary the amount of contrast masking, we created stimuli that had edges aligned with the detection target, or slanted $45°$ to the target. As shown in Figure 11 we used two pedestal patterns in two orientations. The bright squares are $2500\,\mathrm{cd/m^2}$, the dark squares in the checkerboards are $1\,\mathrm{cd/m^2}$, and the background is $5\,\mathrm{cd/m^2}$. The squares of the checkerboards have side length $0.2°$. This experiment was also meant to confirm the radially symmetric characteristic of the pooling we assumed in all other experiments.

We would expect the effect of masking to be much stronger when the edge of the pedestal is aligned with the detected edge. However, the results in Figure 11 indicate little difference between the two orientations of the pedestal. This shows no evidence to support the hypothesis that the elevated thresholds are caused by contrast masking. The results are also consistent with radially symmetric pooling (refer also to similar results in the context of lightness perception [Allred et al. 2012] as discussed in Section 2).
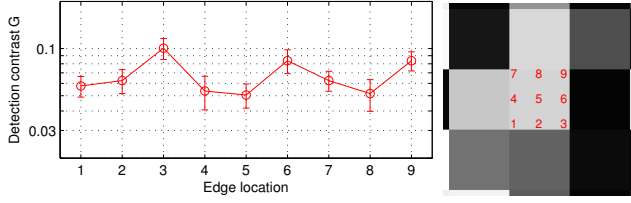

Figure 11: *The detection thresholds for targets embedded in squares or checkerboards with different orientations and luminances.*

### 5.6 Experiment 7: Mondrian and complex images

To enrich the dataset with more real-life adaptation patterns, we also measured detection thresholds for more complex scenarios in which we did not try to isolate any effects.

The first set of images contained a Mondrian-style pattern of square patches of side length $2°$ with exponentially distributed luminances from 0.25 to $5000\,\mathrm{cd/m^2}$, roughly corresponding to a uniform distribution of perceived brightness. The detection target was placed at 9 different positions on a central patch of $2500\,\mathrm{cd/m^2}$, numbered in Figure 12.

The second set consisted of 4 natural images from the HDR Photographic Survey [Fairchild 2008] in which the detection target was positioned to maximize maladaptation. The images and the experiment results are shown in Figure 13.

Figure 12: *The detection thresholds for targets placed at different positions on the central patch in a Mondrian-style pattern as shown on the right.*
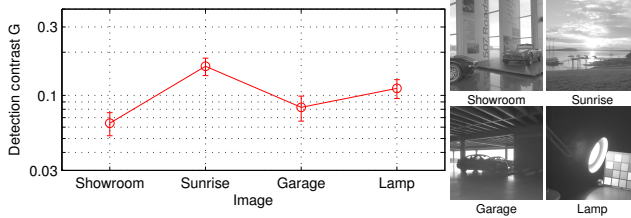

Figure 13: *The detection thresholds for targets placed in various natural images.*
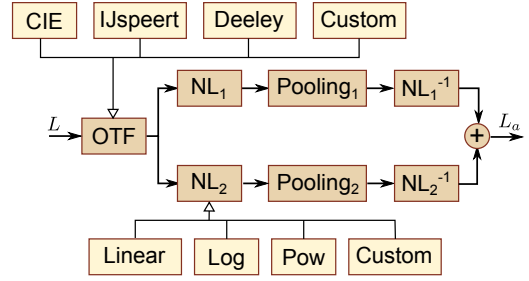
6 Local adaptation model

The detection model introduced in Section 4 should in principle predict the results of our spatial adaptation experiments from Section 5. The missing element, however, is the computation of the adaptation luminance L_a, shown in green in Figure 6. In this section we use our experimental data to find a model capable of predicting L_a.

Visual models are either built upon known physiological constraints or simply designed in an ad hoc manner (Section 2). The former approach often leads to models of excessive complexity, which may fail to generalize. The latter approach may result in a simpler model, but it cannot ensure that the choice of model is adequate and optimal. Here we take a different approach and search a large space of candidate models, each composed of a combination of likely components.

The input signal to the adaptation mechanism must be retinal luminance and hence the first stage of our model is the optics of the eye modeled as a glare spread function (GSF, in the spatial domain) or an optical transfer function (OTF, in the Fourier domain). We started with 8 candidate models from the literature, shown in Figure 14, and selected three that were the most distinct: the GSF from CIE Recommendation 135/1-6 [Vos and van den Berg 1999], the OTF by Deeley et al. [1991] and the OTF by IJspeert et al. [1993]. We also included a custom parametric OTF, which was a linear combination
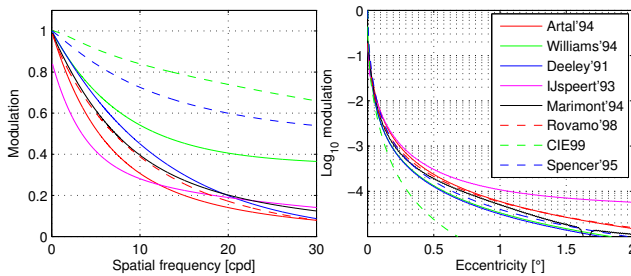
Figure 14: *Comparison of the optical transfer functions (OTFs, left) and the corresponding glare spread functions (GSFs, right) considered in the model.*

Figure 15: *Local adaptation model and the explored combinations of its components. OTF – optical transfer function; NL – non-linearity; Pooling – spatial summation (Gaussian convolution); NL^{-1} – inverse non-linearity. The open arrows indicate alternative model components. This local adaptation model details the green area of the detection model in Figure 6.*

Table 1: *Ranking of the models after fitting to the data from Experiments 2–7. The "∘" symbol denotes function composition. n_{cust} is a custom non-linearity. g is a Gaussian convolution. $p_{...}$ is an OTF. The df column is the number of degrees of freedom (free parameters).*

#	Model	df	χ^2_{red}
1	$(\exp \circ g \circ \log + n_{cust}^{-1} \circ g \circ n_{cust}) \circ p_{IJspeert}$	7	1.26
2	$(n_{cust}^{-1} \circ g \circ n_{cust} + n_{cust}^{-1} \circ g \circ n_{cust}) \circ p_{Deeley}$	11	1.33
3	$(g + n_{cust}^{-1} \circ g \circ n_{cust}) \circ p_{cust}$	11	1.35
4	$(\text{pow}^{-1} \circ g \circ \text{pow} + n_{cust}^{-1} \circ g \circ n_{cust}) \circ p_{Deeley}$	8	1.45
5	$(\exp \circ g \circ \log + n_{cust}^{-1} \circ g \circ n_{cust}) \circ p_{CIE}$	7	1.46
6	$(n_{cust}^{-1} \circ g \circ n_{cust} + n_{cust}^{-1} \circ g \circ n_{cust}) \circ p_{CIE}$	11	1.48
7	$(g + n_{cust}^{-1} \circ g \circ n_{cust}) \circ p_{IJspeert}$	7	1.51
8	$(\text{pow}^{-1} \circ g \circ \text{pow} + n_{cust}^{-1} \circ g \circ n_{cust}) \circ p_{cust}$	12	1.53
9	$(n_{cust}^{-1} \circ g \circ n_{cust} + n_{cust}^{-1} \circ g \circ n_{cust}) \circ p_{cust}$	15	1.55
10	$(n_{cust}^{-1} \circ g \circ n_{cust} + n_{cust}^{-1} \circ g \circ n_{cust}) \circ p_{IJspeert}$	11	1.55

of four exponential functions, similar to the OTF of IJspeert et al. [1993]. The parameters of that OTF were free parameters of the model. Pupil diameter affects the shape of the OTF, especially at higher frequencies. Accordingly, we added pupil size changes to our modeling, but we did not observe any improvement due to pupil changes in our model predictions.

Spatial pooling may take different forms but we restricted our search to the convolution with a mixture of Gaussian functions. However, we allowed each term of the Gaussian mixture to be optionally preceded by one of several non-linearities: logarithmic (a common approximation of the receptor response), a power function with an exponent as a free parameter, or a custom non-linearity designed as a monotonic, C_1-continuous function created from a cubic interpolation of four nodes, where the position of each node was a free parameter. Each non-linearity was paired with its inverse applied after Gaussian convolution. The schematic diagram of possible model combinations is shown in Figure 15.

Given all unique combinations of model components, we generated 56 candidate local adaptation models and fitted each separately to the results of Section 5. We used a genetic optimization algorithm [Vidal et al. 2012] with a constrained range of plausible parameter values, which we extended to support parallel computation of the fitness function with the help of the Message Passing Interface (MPI). To fit all 56 models with up to 15 free parameters each in reasonable time we used an HPC cluster.

The results of our fitting procedure are shown in Table 1. The
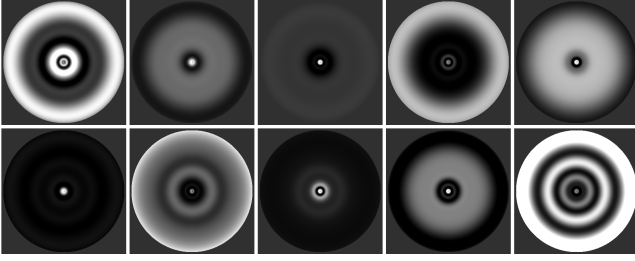
Figure 16: *Ten stimuli optimized to maximize differences between the best performing models.*

goodness-of-fit is reported as the reduced χ^2 statistic:

$$\chi^2_{\text{red}} = \frac{1}{N-d-1} \sum_{i=1}^{N} \frac{(o_i - m_i)^2}{\sigma^2}, \qquad (8)$$

where N is the number of fitted stimuli, d is the number of degrees of freedom (free parameters), o_i is the measurement and m_i is the model prediction. σ is the standard error, which is due to both within- and between-observer variations. A value of χ^2_{red} close to 1 indicates that the model error is close to the variance in the measurements and provides a good fit. Values below 1 could indicate over-fitting. Note that χ^2 statistics penalizes models with large number of parameters (d).

The results show that there are several models with comparable χ^2_{red} value. The best-fitting models, however, are relatively complex with many degrees of freedom and we have no guarantee that they do not over-fit our data.

6.1 Experiment 8: Model-driven stimuli

To discriminate between similarly performing models and to avoid the risk of over-fitting, we generated yet another set of 10 stimuli for the experiment. The novelty here was that each stimulus was automatically generated by an optimization process to maximize the difference in predicted detection thresholds between models. This maximizes the likelihood that the newly collected data will discover the most generalizable model that is robust to a wide range of possible stimuli.

The automatically generated stimuli were concentric patterns of $6°$ diameter, shown in Figure 16. The profile of each pattern was generated as a cubic interpolation between 10 nodes distributed according to the square of the radius (more nodes near the center). The luminance values for the nodes were the parameters of the optimization, which could vary from $0.1\,\text{cd/m}^2$ to $5000\,\text{cd/m}^2$. To find a stimulus S that results in the largest difference in prediction of two models #i and #j, we solved the following optimization problem:

$$\underset{S}{\arg\max}[M_i(S) - M_j(S)]^2 \qquad (9)$$

where M_i and M_j are the predicted detection thresholds (Equation 7) for the first and second models for stimulus S. We found S for all pairs of the 20 best performing models and then selected 10 (shown in Figure 16) that resulted in the largest value of the objective function.

The detection thresholds were measured using the same experimental procedure as in Section 5. The recomputed goodness-of-fit errors, this time including the newly generated stimuli, are listed in Table 2. The first and third best performing models from Table 1 were not

Table 2: *Ranking of the models after cross-validation with the model-driven dataset. The #$_1$ column is the model rank in Table 1.*

#	#$_1$	Model	df	χ^2_{red}
1	2	$(n_{\text{cust}}^{-1} \circ g \circ n_{\text{cust}} + n_{\text{cust}}^{-1} \circ g \circ n_{\text{cust}}) \circ p_{\text{Deeley}}$	11	1.36
2	4	$(\text{pow}^{-1} \circ g \circ \text{pow} + n_{\text{cust}}^{-1} \circ g \circ n_{\text{cust}}) \circ p_{\text{Deeley}}$	8	1.54
3	12	$(\exp \circ g \circ \log + \text{pow}^{-1} \circ g \circ \text{pow}) \circ p_{\text{Deeley}}$	4	1.63
4	14	$(\text{pow}^{-1} \circ g \circ \text{pow} + \text{pow}^{-1} \circ g \circ \text{pow}) \circ p_{\text{Deeley}}$	5	1.68
5	24	$n_{\text{cust}}^{-1} \circ g \circ n_{\text{cust}} \circ p_{\text{Deeley}}$	5	1.72
6	25	$(g + \exp \circ g \circ \log) \circ p_{\text{Deeley}}$	3	1.75
7	29	$\exp \circ g \circ \log \circ p_{\text{Deeley}}$	1	1.8
8	31	$\text{pow}^{-1} \circ g \circ \text{pow} \circ p_{\text{Deeley}}$	2	1.83
9	32	$\exp \circ g \circ \log \circ p_{\text{cust}}$	5	1.93
10	41	$(g + \text{pow}^{-1} \circ g \circ \text{pow}) \circ p_{\text{cust}}$	8	1.94
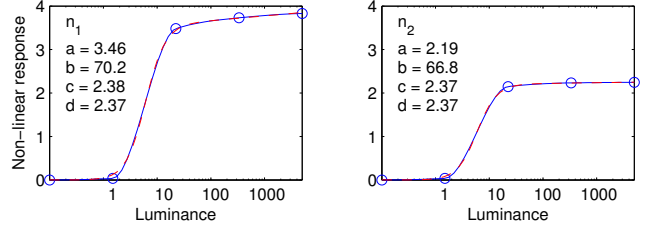
Figure 17: *The two non-linear functions used in model #1. The dashed red lines are the fits of the sigmoidal functions (Equation 11). The parameters of those fits are listed in the plot.*

robust to the new cross-validation stimuli and dropped out of the top 10. Instead, a number of simpler models moved up in the ranking, and the difference between the best performing models after this procedure is still relatively small. Clearly, this approach is much more effective than a random or arbitrary selection of additional stimuli.

The approach serves as a cross-validation for our model fits, but also as a way to introduce some elements of sparse sensing into psychophysical measurements. In contrast to most sparse sensing methods, here we are working with strongly non-linear models and cannot find the linear basis that could be easily measured. However, we can find the inputs that maximize the chance of differentiating between alternative models.

6.2 Best model

The best performing model #1 from Table 2 can be formally expressed as:

$$L_a = \alpha\, n_1^{-1}\left(n_1(L_O) * g_{\sigma_1}\right) + (1 - \alpha)\, n_2^{-1}\left(n_2(L_O) * g_{\sigma_2}\right) \quad (10)$$

where $\alpha = 0.654$, $*$ is the convolution operator and the parameters for the Gaussian kernels g are $\sigma_1 = 0.428°$ and $\sigma_2 = 0.0824°$. n_1 and n_2 are custom non-linearities plotted in Figure 17. For ease of use, we approximate these non-linearities with sigmoidal functions:

$$n(L_O) = a\,\frac{L_O^c}{b + L_O^d}, \qquad (11)$$

with the parameters a–d listed in Figure 17. L_O is the retinal luminance given by the convolution with the Deeley et al. [1991] OTF with assumed pupil diameter $4\,\text{mm}$. The detail on the OTF can be found in the supplementary materials.

6.3 Discussion

Our model fitting procedure brings several interesting insights. First, the top of the ranking is dominated by models that employ the OTF
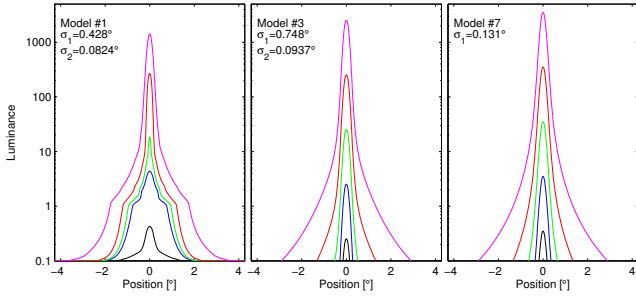
Figure 18: *Response of the three selected models from Table 2 to stimuli containing a* $0.2°$ *disk at luminance ranging from* $1\,\mathrm{cd/m^2}$ *(black) to* $10\,000\,\mathrm{cd/m^2}$ *(magenta). The* σ*-values are the standard deviations of Gaussian convolution used by the models (in visual degrees).*

of Deeley et al. [1991]. Second, all well-performing models involve pooling in a non-linear domain. The simplest model #7 pools the values in the logarithmic domain, while the best performing model #1 employs two custom non-linear functions. Finally, fairly complex models are required to substantially reduce the χ^2_{red} value. Despite its 11 free parameters, model #1 is robust to the cross-validation dataset.

The differences between the models are best visible in Figure 18, in which we plot the response of three selected models to a disk-shaped stimulus of $0.4°$ diameter and of varying luminance. Note that the shape of the response varies depending on the luminance of the disk, indicating the non-linear character of the models. Even though the overall shape of the response is similar, there are substantial differences, especially at lower luminance levels, where the support of the response gets substantially wider for model #1. This is in line with recent findings showing that the site of adaptation shifts from receptors (with small spatial extent) to postreceptoral mechanisms (with larger extent) as light levels are reduced [Dunn et al. 2007]. Model #7 is the easiest to compare with ad hoc adaptation models because of its simplicity: Gaussian blurring in the log-domain. The spatial extent of such blurring ($\sigma = 0.131°$) is clearly much smaller than $1°+$ assumed in most ad hoc models. This demonstrates that of the two extreme ad hoc options for the spatial extent of adaptation, adaptation to a single pixel may be a better approximation than adaptation to a heavily blurred image.

7 Applications

In this section we demonstrate how our local adaptation and detection models can be used in practice.

7.1 Error bounds for physically based rendering

Stochastic ray tracing methods tend to suffer from pixel noise for low sample counts. Adaptive sampling techniques increase the number of samples in a pixel until a convergence criterion is met. Typically that criterion is that the expected range of the true pixel value falls within the tolerance limits of the human visual system based on the current estimated pixel value. These tolerance limits are usually a constant Weber fraction, assuming photopic luminance.

Adaptive sampling techniques based on more principled perceptual criteria have been proposed [Ferwerda et al. 1997; Bolin and Meyer 1998; Ramasubramanian et al. 1999] but are rarely used in practice because of their implementation complexity and computational overhead. If such perceptual models account for contrast masking, they require high spatial frequency information, which is
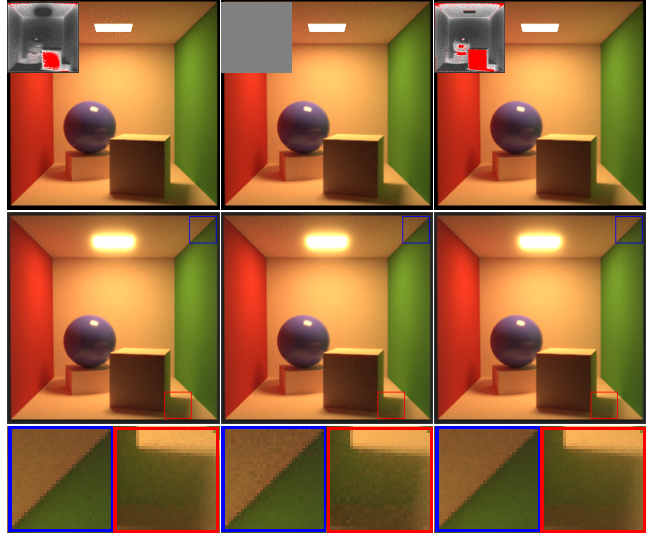
Figure 19: *(a) Basic path tracing with adaptive sampling using our detection model as a convergence threshold. Unconverged pixels are marked in red in the inset sample density map. When shown on an HDR display, glare (simulated in the bottom row) will cover most of the noise around bright light sources and highlights. Local adaptation (not simulated) will hide any remaining noise. (b) Equal-time comparison with non-adaptive sampling. (c) Typical adaptive sampling with a constant Weber fraction criterion.*

unavailable until a large number of samples is collected. In contrast to those methods, our visual model can work with noisy images generated after computing just a few samples per pixel. This is because the model is in fact a cascade of low-pass filters, which will eliminate high frequency noise in an incomplete solution. Moreover, our model does not require an expensive multi-scale decomposition and its computational complexity is much lower. Our predictions, however, are more conservative as we do not account for contrast masking.

As a proof of concept we extended the adaptive sampling implementation of the Mitsuba renderer [Jakob 2014] with our simple visual model:

1: render the image with n initial samples per pixel
2: evaluate detection model to obtain ΔL_{det} thresholds per pixel
3: **for all** pixels **do**
4:     **while** confidence interval $> \Delta L_{\mathrm{det}}$
5:         **and** sample count $< N$ **do**
6:     render n more samples

The only difference compared to a typical adaptive sampling algorithm is that thresholds are obtained by evaluating our detection model instead of using a Weber fraction. Thresholds need to be computed only once, so the computational overhead is negligible. Figure 19 shows an image rendered with our adaptive sampling criterion with $n = 64$ initial samples per pixel and $N = 2048\,\mathrm{spp}$ maximum. Our model predicted convergence after 469.7 samples per pixel (spp) on average. The method using a constant Weber fraction equal to the peak sensitivity of our model, was overly conservative and required $623.5\,\mathrm{spp}$ on average. The Weber criterion wasted samples in areas near bright light sources and highlights that would be covered by glare, and in dark areas where the human visual system can tolerate more noise. Non-adaptive sampling with a fixed number of samples per pixel resulted in much more visible noise for the same number of samples as our method.
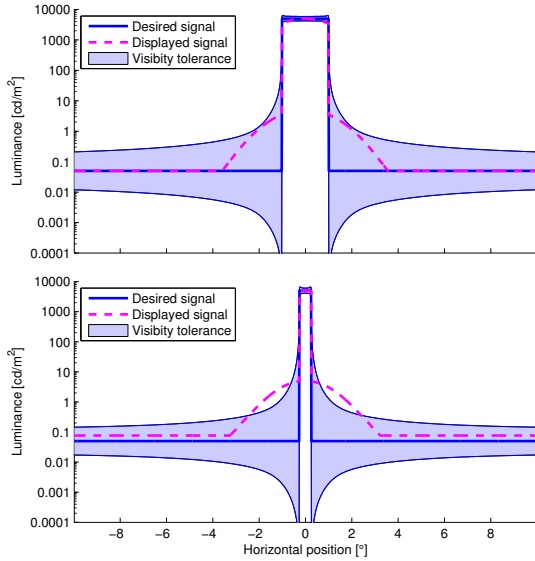
Figure 20: *The visibility of distortions on an HDR display caused by limited backlight resolution. The desired signal is a white square of $5000\,\mathrm{cd/m^2}$ on a background of $0.05\,\mathrm{cd/m^2}$. The plot shows a luminance profile of such a square as desired (solid blue line) and the one that is actually displayed due to limited resolution of the backlight (dashed magenta line). The backlight blur has a Gaussian profile with standard deviation $1°$ (the result depends on the viewing distance). The visibility bounds predicted by our model (blue) indicate that the display distortions are invisible when the square has a width of $2°$ (top) but they become visible when the square size is reduced to $0.5°$ (bottom).*

7.2 Optimal HDR display backlight resolution

Most available HDR displays achieve very high contrast by combining two light modulators, such as an LCD and a projector, or an LCD and an array of LEDs [Seetzen et al. 2004]. The projector or LED backlight modulator usually has a lower resolution because of physical constraints (e.g., limited number of LEDs), but also to reduce the effect of parallax due to both images being produced at slightly different depths. The result of such reduced resolution of one modulator is reduced local contrast. The problem is visually illustrated in Figure 20, in which we predict[3] when the distortions due to the backlight resolution will become visible. Seetzen et al. [2004] conducted a similar analysis, however their bounds were based on the glare amount alone, without considering local adaptation or luminance-dependent elevation of the detection thresholds. Our model can give more accurate predictions of display distortions.

7.3 Visible vs. physical dynamic range

Real-world scenes can potentially span an extremely high physical dynamic range. However, the simultaneously visible dynamic range is much more limited, mostly due to glare [McCann and Rizzi 2007], but also due to local spatial adaptation. Since both of these effects depend on the light distribution in the scene, the maximum simultaneously visible dynamic range is scene-dependent and therefore it is impossible to define it with a single number. However, since our model can predict both glare and local adaptation, it can also determine the maximum visible dynamic range for any given scene.

---

[3]Using model #3 from Table 2 to avoid extrapolating the custom non-linearity n_{cust}, used in higher-ranked models, below the minimum experimental luminance level of $0.1\,\mathrm{cd/m^2}$.
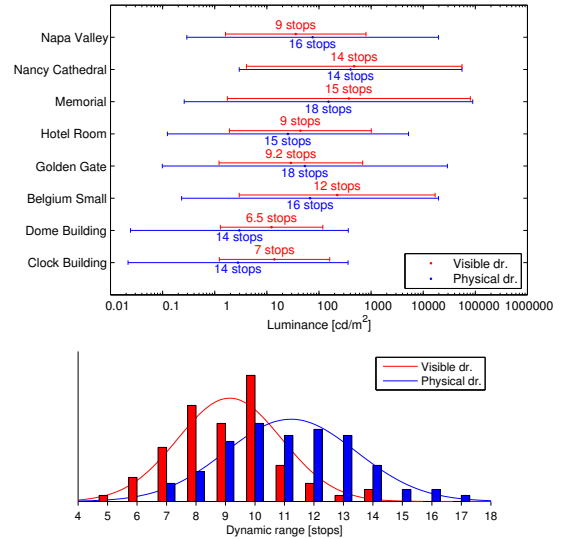
Figure 21: *Comparison of physical and visible dynamic range for (top) a few selected scenes and for (bottom) a collection of 76 images from the Southampton-York Natural Scenes (SYNS) dataset.*

To determine a dynamic range of any sensor, it is necessary to select the minimum signal-to-noise ratio (SNR) level that is considered as "usable". This is a typical assumption when measuring the dynamic range of digital cameras. In the case of the visual system, the signal is physical luminance and the noise is the amount of contrast that remains undetectable. For our experiments, we selected the SNR to be at least 4:1, and therefore we require the predicted[3] $\Delta_{\mathrm{det}}L/L \leq 0.25$. To determine the maximum visible dynamic range in a scene, we find the maximum and minimum luminance of all the pixels that meet this criterion.

As an example, we process a set of high quality HDR images from publicly available databases. First, we simulated viewing 8 standard images on a 40" HDR display of unrestricted brightness and dynamic range, from the viewing distance of 3 image heights (recommended for an HD resolution). When reporting physical dynamic range, we ignore the optical glare of the camera that took the images because it is usually much lower than the glare of the eye. Both physical and visible dynamic range is shown in Figure 21 (top). The graphs indicate the largest loss of visibility occurs in darker scene regions due to glare, but there is also a significant loss of visibility in brighter parts due to local adaptation. For some scenes, the physical and the visible dynamic range are almost identical, but for other scenes the visible dynamic range is just half of the physical range.

The bottom plot in Figure 21 shows the second experiment in which we measured the distribution of dynamic range in natural scenes using the Southampton-York Natural Scenes (SYNS) database [Adams et al. 2015]. The 19 source images were $360°$ panoramas captured with a SpheroCam HDR camera, from which we extracted a total of 76 wide-angle views with approximately $180° \times 90°$ field of view and computed the dynamic range. The histogram shows that simultaneously visible dynamic range in this sample of natural scenes varies between 5 and 14 stops with the peak at 10 stops.

7.4 Simulation of afterimages

Our model can easily be combined with a set of temporal filters to predict the time-course of adaptation. Because we can predict the adaptation per each spatial location, we can simulate the afterimage patterns seen when a luminance pattern changes abruptly over time.

$t = 0\,\mathrm{s}$	$t = 1\,\mathrm{s}$	$t = 2\,\mathrm{s}$	$t = 3\,\mathrm{s}$	$t = \infty$
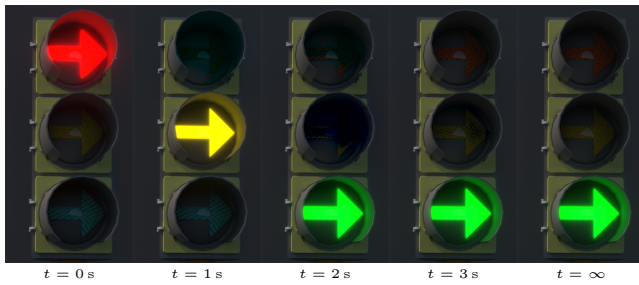
Figure 22: *Simulation of afterimages of traffic lights. The red light leaves a greenish afterimage, and the amber light leaves a bluish afterimage. Both afterimages last for a long time while the green light is active.*

In contrast to previous work [Ritschel and Eisemann 2012; Jacobs et al. 2015], we can accurately predict the blurriness instead of presenting sharp or arbitrarily blurred afterimages. As examples, we reproduce the traffic lights example from [Ritschel and Eisemann 2012] and simulate the appearance of an afterimage illusion in Figure 23. The hue and color saturation in our simulation is computed according to Jacobs et al. [2015]

7.5 Gaze-dependent tone mapping

When viewing natural scenes our gaze moves between areas of different brightness. This causes the visual system to constantly re-adapt to different luminance levels. When viewing images on a regular (LDR) display, the adaptation changes to a lesser degree as the luminance range reproduced on the display is much smaller. However, if real-time information about gaze position is available from an eye-tracker, the real-world adaptation process can be simulated on a regular display [Mantiuk and Markowski 2013].

We reproduced a gaze-dependent tone mapping system similar to the one presented in [Mantiuk and Markowski 2013]. Given an HDR image as input, our model predicts the spatial map of adaptation luminance levels that the eye would arrive at in the real-world scene. The effective state of adaptation follows the temporal process modeled as an exponential decay function from [Durand and Dorsey 2000]. The effective adaptation state was then used to tone map the image using the Naka–Rushton photoreceptor response, similar to [Reinhard and Devlin 2005]. Figure 24 shows two animation frames from a video capturing the session in which an observer scanned an image using gaze-dependent tone mapping. The frames demonstrate how the entire image changed in perceived brightness after the gaze moved from dark to bright image parts, delivering a better impression of the high dynamic range that could be found in the actual scene.

8 Limitations

Our experiments were limited to achromatic luminance adaptation, but our model can be generalized to color images, as demonstrated in the applications, by assuming that the pooling processes have the same spatial characteristics for all photoreceptors. This is a reasonable assumption for pooling caused by eye movements, chemical diffusion, and laterally interconnecting retinal neurons. Our experiments were limited by the luminance range of our display system, but the maximum luminance of $5000\,\mathrm{cd/m^2}$ exceeds most commercially available HDR displays. The lower end of the display luminance range limits our model to photopic stimuli, which cover most practical application scenarios. We would like to extend the measurement to the mesopic range as a future work. The nature of our stimuli with a constant diameter pedestal makes it impossible to

model the effect of strong luminance variations less than $0.1°$ from the fixation point.

Our model does not consider the Stiles-Crawford effect [Stiles and Crawford 1933; Gutierrez et al. 2005] and pupil contractions, which might contribute to optical blurring at low luminance levels. However, we found that an OTF with a variable pupil did not improve the accuracy of our model predictions, as discussed in Section 6. There is also no need to model the loss of retinal luminance due to pupil contractions. This is because the tvi used in our model already incorporates the effect of pupil size and it is a function of luminance, in cd/m^2, rather than retinal illuminance, in trolands.

In this work we do not consider contrast adaptation effects [Greenlee and Heitger 1988], which lead to increasing contrast detection thresholds resulting from prolonged inspection of high contrast patterns of similar spatial frequencies and orientations. Since we assume that the eye is fixated on a target and reaches steady adaptation, we do not model such time-dependent effects that involve gaze direction changes and account for local characteristics of attended image regions. We have not found an effect of masking in our Experiment 6 but other stimuli should be considered to confirm our findings, which we relegate as future work.

9 Conclusions

We have presented a quantitative model of local adaptation. The model was trained on empirical data from many different types of stimuli with varying luminance patterns of over $6°$ of visual angle, covering the entire foveal field of view. Out of an exhaustive set of plausible candidate models, the best fitting models were selected and cross-validated with an additional set of maximally discriminating stimuli. This procedure ensures that the model not only explains the training data but is also predictive for any new input. The model is conceptually simple to implement and computationally inexpensive to evaluate, requiring only 3 moderate-sized convolution filters: the OTF filter in the linear luminance domain and two Gaussian pooling filters in different non-linear domains. Our model of the spatial characteristics of local adaptation can easily be combined with existing temporal models, as demonstrated in the simulation of afterimages, to predict the time-course of local adaptation. We have used our model in a wide range of application scenarios for predicting the technical limitations and requirements of HDR image synthesis, compression, tone mapping, and display.

References

ADAMS, W., ELDER, J., GRAF, E., MURYY, A., AND LUGTIGHEID, A. 2015. Perception of 3D structure and natural scene statistics: The Southampton-York Natural Scenes (SYNS) dataset. *Vision Sciences Society 2015 Poster*.

AHMAD, K. M., KLOG, K., HERR, S., STERLING, P., AND SCHEIN, S. 2003. Cell density ratios in a foveal patch in macaque retina. *Visual Neuroscience 20*, 2 (June), 189–209.

(a) *Original*   **(b)** *Inverse chromatic*   **(c)** *Luminance*   **(d)** *Prediction*
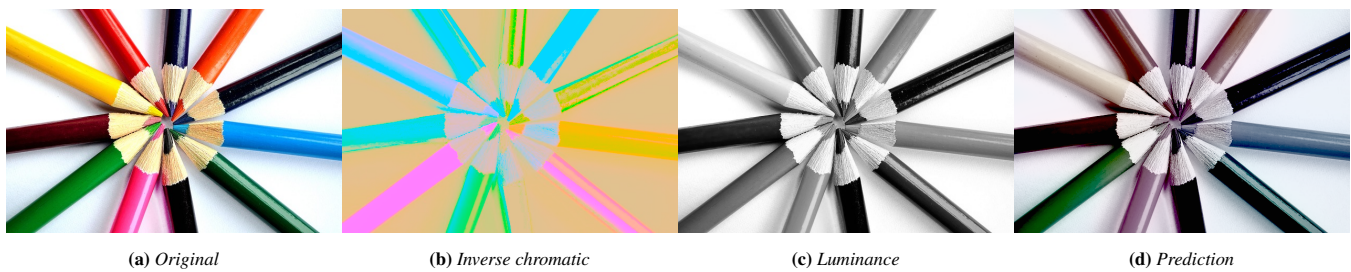
Figure 23: *Simulation of an afterimage illusion. The original image (a) is decomposed into the equiluminant inverse-chromatic image (b) and the luminance image (c). Stare at a point on (b) for at least* 10 s, *then look at the same point on (c). The chromatic information in the afterimage recombined with the luminance resembles the original image (a). Our model correctly predicts (d) the loss of chromatic saturation in this illusion. (For optimal results, try this on a standard sRGB display at a viewing distance of 8 image heights.)*

Figure 24: *Two frames from a session with gaze-dependent tone mapping, in which an observer shifted their gaze from a dark to a bright image region. The map in the middle shows the spatial adaptation map predicted by our model. The circles with numbers show corresponding gaze positions.*

ALLRED, S. R., RADONJIĆ, A., GILCHRIST, A. L., AND BRAINARD, D. H. 2012. Lightness perception in high dynamic range images: Local and remote luminance effects. *Journal of Vision 12*, 2, 7.

BARLOW, H. 1972. Dark and light adaptation: Psychophysics. In *Visual Psychophysics*, D. Jameson and L. Hurvich, Eds., vol. 7 / 4 of *Handbook of Sensory Physiology*. Springer Berlin Heidelberg, 1–28.

BOLIN, M. R., AND MEYER, G. W. 1998. A perceptually based adaptive sampling algorithm. In *Proc. of SIGGRAPH '98*, ACM Press, New York, New York, USA, ACM, 299–309.

CHIU, K., HERF, M., SHIRLEY, P., SWAMY, S., WANG, C., AND ZIMMERMAN, K. 1993. Spatially nonuniform scaling functions for high contrast images. In *Proceedings of Graphics Interface '93*, 245–253.

CRAIK, K. J. W. 1938. The effect of adaptation on differential brightness discrimination. *The Journal of Physiology 92*, 4, 406–421.

DEELEY, R. J., DRASDO, N., AND CHARMAN, W. N. 1991. A simple parametric model of the human ocular modulation transfer function. *Ophthalmic and Physiological Optics 11*, 1 (Jan.), 91–93.

DUNN, F. A., AND RIEKE, F. 2008. Single-photon absorptions evoke synaptic depression in the retina to extend the operational range of rod vision. *Neuron 57*, 6, 894–904.

DUNN, F. A., LANKHEET, M. J., AND RIEKE, F. 2007. Light adaptation in cone vision involves switching between receptor and post-receptor sites. *Nature 449*, 7162 (Oct.), 603–6.

DURAND, F., AND DORSEY, J. 2000. Interactive tone mapping. *Eurographics Workshop on Rendering*.

FAIRCHILD, M. D. 1998. *Color Appearance Models*. Addison-Wesley. ISBN 0-201-63464-3.

FAIRCHILD, M. D. 2008. *The HDR Photographic Survey*. MDF Publications. http://rit-mcsl.org/fairchild/HDR.html.

FERWERDA, J. A., PATTANAIK, S., SHIRLEY, P., AND GREENBERG, D. P. 1996. A model of visual adaptation for realistic image synthesis. In *Proceedings of SIGGRAPH 96*, Annual Conference Series, ACM, 249–258.

FERWERDA, J. A., SHIRLEY, P., PATTANAIK, S. N., AND GREENBERG, D. P. 1997. A model of visual masking for computer graphics. In *Proc. of SIGGRAPH '97*, ACM Press, New York, New York, USA, ACM, 143–152.

FINKELSTEIN, M. A., HARRISON, M., AND HOOD, D. C. 1990. Sites of sensitivity control within a long-wavelength cone pathway. *Vision Research 30*, 8 (Jan.), 1145–1158.

GEISLER, W. S. 1978. Adaptation, afterimages and cone saturation. *Vision Research 18*, 3, 279 – 289.

GREENLEE, M. W., AND HEITGER, F. 1988. The functional role of contrast adaptation. *Vision Research 28*, 7, 791 – 797.

GUTIERREZ, D., ANSON, O., MUNOZ, A., AND SERON, F. 2005. Perception-based rendering: eyes wide bleached. In *Proc. Eurographics (Short Papers)*, 49–52.

HESS, R. F., SHARPE, L. T., AND NORDBY, K. 1990. *Night Vision: Basic, Clinical and Applied Aspects*. Cambridge University Press.

HOOD, D. C., FINKELSTEIN, M. A., AND BUCKINGHAM, E. 1979. Psychophysical tests of models of the response function. *Vision Research 19*, 401–406.

HUNT, R. W. G. 1995. *The Reproduction of Colour in Photography, Printing and Television: 5th Edition*. Fountain Press.

IJSPEERT, J. K., VAN DEN BERG, T. J., AND SPEKREIJSE, H. 1993. An improved mathematical description of the foveal visual point spread function with parameters for age, pupil size and pigmentation. *Vision Research 33*, 1 (Jan.), 15–20.

IRAWAN, P., FERWERDA, J. A., AND MARSCHNER, S. R. 2005. Perceptually based tone mapping of high dynamic range image streams. In *Eurographics Symposium on Rendering (2005)*, K. Bala and P. Dutré, Eds., Eurographics.

JACOBS, D. E., GALLO, O., COOPER, E. A., PULLI, K., AND LEVOY, M. 2015. Simulating the visual experience of very bright and very dark scenes. *ACM Trans. Graph. 34*, 3 (May), 25:1–25:15.

JAKOB, W., 2014. Mitsuba 0.5.0 Physically Based Renderer. http://www.mitsuba-renderer.org/.

JOBSON, D. J., RAHMAN, Z., AND WOODELL, G. A. 1997. A multi-scale retinex for bridging the gap between color images and the human observation of scenes. *IEEE Transactions on Image Processing: Special Issue on Color Processing 6*, 7, 965–976.

KIM, M. H., WEYRICH, T., AND KAUTZ, J. 2009. Modeling human color perception under extended luminance levels. *ACM Transactions on Graphics (Proc. SIGGRAPH 2009) 28*, 3, 27:1–9.

KUANG, J., JOHNSON, G. M., AND FAIRCHILD, M. D. 2007. iCAM06: A refined image appearance model for HDR image rendering. *Journal of Visual Communication and Image Representation 18*, 406–414.

LARSON, G. W., RUSHMEIER, H., AND PIATKO, C. 1997. A visibility matching tone reproduction operator for high dynamic range scenes. *IEEE Transactions on Visualization and Computer Graphics 3*, 4, 291–306.

LEDDA, P., SANTOS, L. P., AND CHALMERS, A. 2004. A local model of eye adaptation for high dynamic range images. In *Proceedings of AFRIGRAPH '04*, AFRIGRAPH, 151–160.

MANTIUK, R., AND MARKOWSKI, M. 2013. Gaze-dependent tone mapping. *Proc. of ICIAR 7950*, 426–433.

MANTIUK, R., KIM, K. J., REMPEL, A. G., AND HEIDRICH, W. 2011. HDR-VDP-2: A calibrated visual metric for visibility and quality predictions in all luminance conditions. *ACM Transactions on Graphics 30*, 4 (July), 40:1–40:13.

MCCANN, J. J., AND RIZZI, A. 2007. Camera and visual veiling glare in HDR images. *Journal of the Society for Information Display 15*, 9, 721.

MCKEE, S. P., AND WESTHEIMER, G. 1970. Specificity of cone mechanisms in lateral interaction. *The Journal of Physiology 206*, 1, 117–128.

MOON, P., AND SPENCER, D. E. 1945. The visual effect of non-uniform surrounds. *Journal of the Optical Society of America 35*, 3, 233–248.

NAKA, K. I., AND RUSHTON, W. A. H. 1966. S-potentials from luminosity units in the retina of fish (Cyprinidae). *Journal of Physiology 185*, 587–599.

PAJAK, D., CADIK, M., AYDIN, T. O., MYSZKOWSKI, K., AND SEIDEL, H.-P. 2010. Visual maladaptation in contrast domain. In *Human Vision and Electronic Imaging XV*, B. E. Rogowitz and T. N. Pappas, Eds., vol. 7527, Proc. SPIE, 752710–12.

PATTANAIK, S. N., TUMBLIN, J. E., YEE, H., AND GREENBERG, D. P. 2000. Time-dependent visual adaptation for fast realistic image display. In *Proc. of SIGGRAPH 2000*, ACM, 47–54.

RADONJIĆ, A., ALLRED, S. R., GILCHRIST, A. L., AND BRAINARD, D. H. 2011. The dynamic range of human lightness perception. *Current Biology 21*, 22, 1931 – 1936.

RAMASUBRAMANIAN, M., PATTANAIK, S. N., AND GREENBERG, D. P. 1999. A perceptually based physical error metric for realistic image synthesis. In *SIGGRAPH 99 Conference Proceedings*, A. Rockwood, Ed., Annual Conference Series, ACM, 73–82.

REINHARD, E., AND DEVLIN, K. 2005. Dynamic range reduction inspired by photoreceptor physiology. *IEEE Transactions on Visualization and Computer Graphics 11*, 1, 13–24.

REINHARD, E., STARK, M., SHIRLEY, P., AND FERWERDA, J. 2002. Photographic tone reproduction for digital images. *ACM Transactions on Graphics (Proc. SIGGRAPH) 21*, 3, 267–276.

RITSCHEL, T., AND EISEMANN, E. 2012. A Computational Model of Afterimages. *Computer Graphics Forum 31*, 2pt3 (May), 529–534.

SCHLICK, C. 1995. Quantization techniques for visualization of high dynamic range pictures. In *Photorealistic Rendering Techniques*, Eurographics, 7–20.

SEETZEN, H., HEIDRICH, W., STUERZLINGER, W., WARD, G., WHITEHEAD, L., TRENTACOSTE, M., GHOSH, A., AND VOROZCOVS, A. 2004. High dynamic range display systems. *ACM Transactions on Graphics 23*, 3, 760–768.

SHAPLEY, R., AND ENROTH-CUGELL, C. 1984. Chapter 9 Visual adaptation and retinal gain controls. *Progress in Retinal Research 3* (Jan.), 263–346.

STILES, W. S., AND CRAWFORD, B. H. 1933. The luminous efficiency of rays entering the eye pupil at different points. *Proceedings of the Royal Society of London B: Biological Sciences 112*, 778, 428–450.

TUMBLIN, J., HODGINS, J. K., AND GUENTER, B. K. 1999. Two methods for display of high contrast images. *ACM Transactions on Graphics 18*, 1, 56–94.

VALETON, J. M. 1983. Photoreceptor light adaptation models: An evaluation. *Vision Research 23*, 12, 1549–1554.

VAN HATEREN, H. 2005. A cellular and molecular model of response kinetics and adaptation in primate cones and horizontal cells. *Journal of Vision 5*, 4, 331–347.

VAN HATEREN, J. H. 2006. Encoding of high dynamic range video with a model of human cones. *ACM Transactions on Graphics 25*, 4, 1380–1399.

VIDAL, FRANCK, P., VILLARD, P.-F., AND LUTTON, E. 2012. Tuning of patient specific deformable models using an adaptive evolutionary optimization strategy. *IEEE Transactions on Biomedical Engineering 59*, 10, 2942–2949.

VOS, J. J., AND VAN DEN BERG, T. J. 1999. CIE 135/1-6 Disability Glare. Tech. rep., CIE.

WARD, G. 1994. A contrast-based scalefactor for luminance display. *Graphics Gems IV*, 415–421.

WATSON, A. B., AND PELLI, D. G. 1983. QUEST: a Bayesian adaptive psychometric method. *Perception & Psychophysics 33*, 2, 113–120.

WESTHEIMER, G. 1967. Spatial interaction in human cone vision. *Journal of Physiology 190*, 139–154.

WILSON, H. R. 1997. A neural model of foveal light adaptation and afterimage formation. *Visual Neuroscience 14*, 03 (June), 403–423.