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

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

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

---

∗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
To appear in ACM TOG 34(6).

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 cd/m² to 5000 cd/m². 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 10⁶ 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 10⁸ 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; Jisexpeet 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 postreceptorial 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; Iwama et al. 2005; Pajak et al. 2010], while in color appearance models for HDR images [Kim et al. 2009] even larger ranges 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
We use this paradigm to investigate how the mismatch between background and adaptation luminance affects the visual performance across luminance levels. We start the discovery of the new local adaptation model with a simple detection model. The model contains an unknown spatial 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 setup 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 design as in [Seezen 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 1.32m). The backlight was produced by a 3500 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 cd/m². For conditions involving luminances below 0.5 cd/m² 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 $L_p$ most of the time. A detection target was only briefly flashed for 200 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 adaption mechanism can respond in less than 50 ms, we cannot guarantee that the observers remained fully adapted to the uniform field $L_p$ 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 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 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.

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
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.

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:

\[
S = \frac{L}{\Delta L} = \text{CSF}(\rho, L) \quad \Rightarrow \quad \Delta L = \frac{L}{\text{CSF}(\rho, L)} = \text{tvi}(L)
\]

where \( \rho \) is spatial frequency and \( L \) is the luminance of the uniform background. For simplicity, we fix the spatial frequency \( \rho \) to the frequency of the largest amplitude in our stimulus (3.5 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 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 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 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

\[\text{logarithmic contrast is the log difference between two luminance levels: } G = \log_{10}(L + \Delta L) - \log_{10}(L).\]

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) \( O \):

\[
L_O = I \ast O.
\]

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

\[
R = k \frac{L^n}{L^n + \sigma(L_a)^n},
\]

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

\[^2\text{We do not consider here retinal illuminance in trolands as both CSF and } tvi \text{ functions account for the effect of pupil size.}\]
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 \( \Delta L \):

\[
\frac{dR}{dL} \propto \frac{1}{\Delta L}.
\]

(4)

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

\[
\Delta L(L, L_a) \propto \frac{(L^n + L_a^n)^2}{n L^\alpha L_a^\alpha}.
\]

(5)

For modeling purposes we are more interested in the threshold elevation that is relative to the point of complete adaptation: \( L = L_a \). Such threshold elevation due to maladaptation is given by:

\[
T_e(L, L_a) = \frac{\Delta L(L, L_a)}{\Delta L(L_a, L_a)} = \frac{(L^n + L_a^n)^2}{4 L^n L_a^\alpha}
\]

(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 tvit 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:

\[
\Delta L_{det}(L, L_a) = tvit(L_O) T_e(L_O, L_a),
\]

(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 cd/m². The background was a uniform field of 5 cd/m². 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.

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 cd/m² of a variable diameter, on a background of 5 cd/m². 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 ms. The results indicate that the size of the adapting pattern has a significant effect on the state of adaptation.

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 cd/m² was surrounded by a concentric ring of the same luminance and varying inner and outer diameters on a background of 5 cd/m². 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.

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 cd/m². 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 cd/m².

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.

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 cd/m², the dark squares in the checkerboards are 1 cd/m², and the background is 5 cd/m². 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).

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 cd/m², roughly corresponding to a uniform distribution of perceived brightness. The detection target was placed at 9 different positions on a central patch of 2500 cd/m², 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.

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 model.
luminance $L_{O}$, shown in green in Figure 6. In this section we use our experimental data to find a model capable of predicting $L_{O}$.

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 [Nos 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 of four exponential functions, similar to the OTF by 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 (receptor response), a power function with an exponent as a free parameter, or a custom non-linearity designed.

The results of our fitting procedure are shown in Table 1. The goodness-of-fit is reported as the reduced $\chi^2$ statistic:

$$\chi^2_{red} = \frac{1}{N - d - 1} \sum_{i=1}^{N} \frac{(o_i - m_i)^2}{\sigma^2},$$  \hspace{1cm} (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. $\sigma$ is the standard error, which is due to both within- and between-observer variations. A value of $\chi^2_{red}$ 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 $\chi^2$ statistics penalizes models with large number of parameters ($d$).

The results show that there are several models with comparable $\chi^2$ 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 $6^\circ$ 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 cd/m$^2$ to 5000 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:

$$\arg\max_S [M_i(S) - M_j(S)]^2$$

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 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

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

<table>
<thead>
<tr>
<th>$#_i$</th>
<th>$#_j$</th>
<th>Model</th>
<th>$d$</th>
<th>$\chi^2_{red}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2</td>
<td>$\left( n_{w1} \circ g \circ \eta_{rain} + n_{w2} \circ g \circ \eta_{rain} \right) \circ \Phi_{Deeley}$</td>
<td>11</td>
<td>1.36</td>
</tr>
<tr>
<td>2</td>
<td>3</td>
<td>$\left( \exp \circ g \circ \log + \exp \circ g \circ \log \right) \circ \Phi_{Deeley}$</td>
<td>8</td>
<td>1.54</td>
</tr>
<tr>
<td>3</td>
<td>4</td>
<td>$\left( \exp \circ g \circ \log + \exp \circ g \circ \log \right) \circ \Phi_{Deeley}$</td>
<td>4</td>
<td>1.63</td>
</tr>
<tr>
<td>4</td>
<td>5</td>
<td>$\left( \exp \circ g \circ \log + \exp \circ g \circ \log \right) \circ \Phi_{Deeley}$</td>
<td>5</td>
<td>1.68</td>
</tr>
<tr>
<td>5</td>
<td>6</td>
<td>$\left( \exp \circ g \circ \log + \exp \circ g \circ \log \right) \circ \Phi_{Deeley}$</td>
<td>5</td>
<td>1.72</td>
</tr>
<tr>
<td>6</td>
<td>7</td>
<td>$\left( \exp \circ g \circ \log + \exp \circ g \circ \log \right) \circ \Phi_{Deeley}$</td>
<td>3</td>
<td>1.75</td>
</tr>
<tr>
<td>7</td>
<td>8</td>
<td>$\left( \exp \circ g \circ \log + \exp \circ g \circ \log \right) \circ \Phi_{Deeley}$</td>
<td>1</td>
<td>1.8</td>
</tr>
<tr>
<td>8</td>
<td>9</td>
<td>$\left( \exp \circ g \circ \log + \exp \circ g \circ \log \right) \circ \Phi_{Deeley}$</td>
<td>2</td>
<td>1.83</td>
</tr>
<tr>
<td>9</td>
<td>10</td>
<td>$\left( \exp \circ g \circ \log + \exp \circ g \circ \log \right) \circ \Phi_{Deeley}$</td>
<td>5</td>
<td>1.93</td>
</tr>
<tr>
<td>10</td>
<td>11</td>
<td>$\left( \exp \circ g \circ \log + \exp \circ g \circ \log \right) \circ \Phi_{Deeley}$</td>
<td>8</td>
<td>1.94</td>
</tr>
</tbody>
</table>

### 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.

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:

$$L_0 = \alpha n_1^{-1} (n_1(L_0) * g_{\alpha}) + (1 - \alpha) n_2^{-1} (n_2(L_0) * g_{\sigma_2})$$

(10)

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

$$n(L_0) = a \frac{L_0 - b}{L_0 + L_0^c},$$

(11)

with the parameters $a$–$d$ listed in Figure 17. $L_0$ is the retinal luminance given by the convolution with the Deeley et al. [1991] OTF with assumed pupil diameter 4 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 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 $\chi^2_{red}$ 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 (σ = 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 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.

Figure 18: Response of the three selected models from Table 2 to stimuli containing a 0.2° disk at luminance ranging from 1 cd/m² (black) to 10 000 cd/m² (magenta). The σ-values are the standard deviations of Gaussian convolution used by the models (in visual degrees).

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. 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 ∆Ldet thresholds per pixel
3. for all pixels do
4. while confidence interval > ∆Ldet
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 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 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.

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
To appear in ACM TOG 34(6).

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.

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 \( \Delta_{\text{det}} L/L \leq 0.25 \). 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° × 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. 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].
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 cd/m² 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 τ01 used in our model already incorporates the effect of pupil size and it is a function of luminance, in cd/m², 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
Cerebral Blood Flow and Oxygenation in Awake Macaque Monkeys: A Pilot Study

1. Introduction
2. Methods
3. Results
4. Discussion
5. Conclusion

References

Acknowledgments

To appear in ACM TOG 34(6).
To appear in ACM TOG 34(6).


