

Display Adaptive Tone Mapping

Rafał Mantiuk*

MPI Informatik, Germany
Sharp Laboratories of America

Scott Daly

Sharp Laboratories of America

Louis Kerofsky

Sharp Laboratories of America

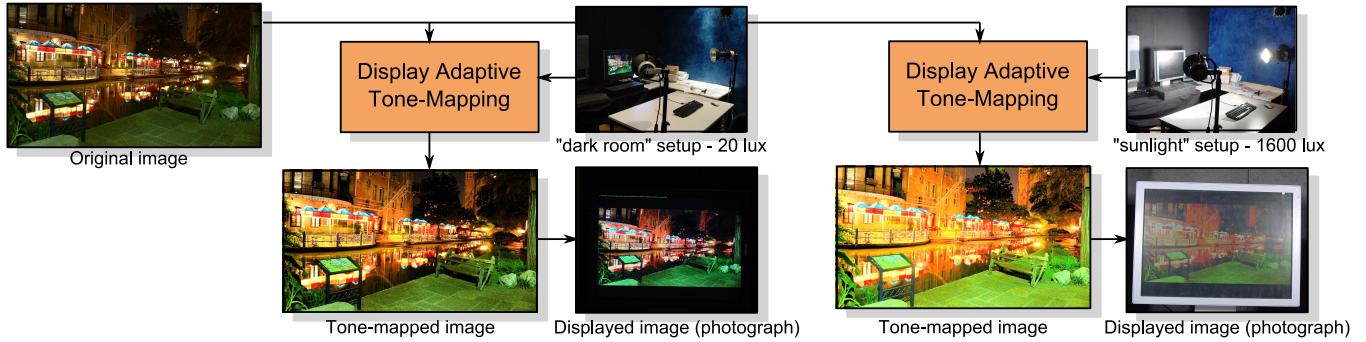


Figure 1: Image reproduced adaptively for low ambient light (dark room scenario – left) and high ambient light (sunlight scenario – right). The display adaptive tone mapping can account for screen reflections when generating images that optimize visible contrast.

Abstract

We propose a tone mapping operator that can minimize visible contrast distortions for a range of output devices, ranging from e-paper to HDR displays. The operator weights contrast distortions according to their visibility predicted by the model of the human visual system. The distortions are minimized given a display model that enforces constraints on the solution. We show that the problem can be solved very efficiently by employing higher order image statistics and quadratic programming. Our tone mapping technique can adjust image or video content for optimum contrast visibility taking into account ambient illumination and display characteristics. We discuss the differences between our method and previous approaches to the tone mapping problem.

CR Categories: I.3.3 [Computer Graphics]: Picture/Image Generation—Display algorithms; I.4.2 [Image Processing and Computer Vision]: Enhancement—Greyscale manipulation, sharpening and deblurring

Keywords: high dynamic range, tone mapping, image reproduction, visual perception, optimization, display-adaptive, viewing conditions

*e-mail:mantiuk@gmail.com

ACM Reference Format
Mantiuk, R., Daly, S., Kerofsky, L. 2008. Display Adaptive Tone Mapping. *ACM Trans. Graph.* 27, 3, Article 68 (August 2008), 10 pages. DOI = 10.1145/1360612.1360667 <http://doi.acm.org/10.1145/1360612.1360667>.

Copyright Notice
Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or direct commercial advantage and that copies show this notice on the first page or initial screen of a display along with the full citation. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, to republish, to post on servers, to redistribute to lists, or to use any component of this work in other works requires prior specific permission and/or a fee. Permissions may be requested from Publications Dept., ACM, Inc., 2 Penn Plaza, Suite 701, New York, NY 10121-0701, fax +1 (212) 869-0481, or permissions@acm.org.
© 2008 ACM 0730-0301/2008/03-ART68 \$5.00 DOI 10.1145/1360612.1360667
<http://doi.acm.org/10.1145/1360612.1360667>

1 Introduction

Reproducing natural and artificial scenes on display devices of a limited dynamic range (contrast) is a challenging problem in photography, cinematography, printing and visualization. So far, the best results are achieved when each image is manually adjusted on the target display. This, however, is a tedious task that often requires expert skills. The question arises whether the manual adjustments can be replaced with a computational algorithm. We address this question by demonstrating that the image reproduction tasks can be formulated as an optimization problem, in which the best tradeoff between preserving contrast in all ranges of a tone-scale is found. Such optimization is driven by a perceptual metric that weights contrast distortions according to their visibility and importance.

Tone mapping should not only ensure that the resulting pixel values are in the range 0–255, but also that the actual tones shown on a particular display of certain capabilities will convey the desired image content. This is especially important with the recent developments in display technologies (LCD, LCoS, PDP, DLP, OLED, e-paper, backlight modulation [Seetzen et al. 2004], rear-projection), and the variety of applications in which they are employed (home entertainment, mobile displays, electronic books, cockpit displays, etc.). All these display devices can differ dramatically in their peak brightness, contrast (dynamic range) and black level, and can change their characteristic with the viewing conditions (sunlight vs. office light, as shown in Figure 1). Therefore, it cannot be expected that the same image shown on different devices will produce the desired appearance. Tone mapping with no knowledge of the target display is not a fully defined problem, similar to gamut mapping with no knowledge of the target gamut.

We propose a tone mapping operator that produces the least distorted image, in terms of visible contrast distortions, given the characteristic of a particular display device. The distortions are weighted using the human visual system (HVS) model, which accounts for all major effects, including luminance masking, spatial contrast sensitivity and contrast masking (Section 3.4). Such a tone mapping operator is naturally formulated as an optimization prob-

lem (Section 3), where the error function is weighted by the HVS model and constraints are dictated by the display limitations (Section 3.3). We demonstrate that the problem can be solved very efficiently if the error function is computed from higher order image statistics (Section 4.1) and the non-linear optimization problem is reduced to a small-size quadratic programming task (Section 4.3). A straightforward extension of our method ensures temporal coherence and makes it suitable for video sequences (Section 5). The performance of our technique is validated in several less studied scenarios of tone reproduction when the viewing conditions and display capabilities vary (Sections 6.1 and 6.2). Our experimental study shows that images that are adaptively tone mapped to illumination conditions are preferred in terms of contrast reproduction (Section 6.3). Finally, the method is compared with other tone mapping operators (Section 6.4).

2 Previous Work

The problem of tone-reproduction was recognized by early painters, who struggled with the limited contrast range of available pigments. Since the pigments did not offer much contrast for dark colors, Leonardo da Vinci tended to use midrange colors for all objects, so that he could achieve desired contrast and strong depth-from-shading effects even if the actual brightness levels were distorted [Livingstone 2002, p. 115]. Livingstone in her book [2002, pp. 109–125] gives a good overview of other techniques painters used to overcome limited dynamic range. Much later when film-based **photography** was invented, the first practitioners of this new technique found that capturing the enormous dynamic range of luminance in the real world on a chemically limited negative was difficult. Film manufacturers tried to reduce this problem by designing film stocks and print development systems that gave a desired S-shaped tone curve with slightly enhanced contrast (about 15%) in the middle range and gradually compressed highlights and shadows [Hunt 2004, Ch. 6]. To overcome the limitations of print, photographers also locally modified image exposure with the **dodging and burning** technique [Adams 1981].

The introduction of digital photography and image processing opened new possibilities for image reproduction. Stockham [1972] introduced a computational visual model capable of reducing dynamic range, which consisted of a logarithmic function followed by suppression of low frequencies. One of the most notable early algorithms employed for image rendering was the **retinex** [Land and McCann 1971], inspired by the theories of lightness perception. The problem of limited color gamut has been extensively studied in the context of color printing, resulting in a range of **gamut-mapping** algorithms [Morovic and Luo 2001]. Since colors change their appearance with viewing conditions, reproducing an image involves reproducing its **color appearance**. Although color appearance is a complex phenomenon, it can be predicted by computational models, such as CIECAM02 [Moroney et al. 2002] or iCAM [Kuang et al. 2007]. When automatic algorithms are not sufficient and high quality results are required, mapping tones and colors must be performed manually by a skilled artist. This is the case in the cinematographic movie post-processing process, called **color-grading**.

Computer graphics techniques, capable of rendering high-contrast scenes, shifted focus from color to luminance as the main limiting factor of display devices. A recent book [Reinhard et al. 2005] gives a good review of a number of **tone mapping operators (TMOs)**, intended to map high dynamic range (HDR) images to standard displays. More recent work on tone mapping shows a trend towards user-assisted image reproduction [Lischinski et al. 2006], stylized rendering [Bae et al. 2006] and finding other means than luminance to extend image contrast [Smith et al. 2006]. A large number of

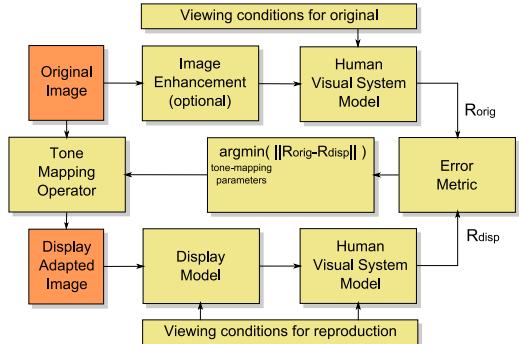


Figure 2: The proposed formulation of the tone mapping problem.

the proposed operators, and no proven method to validate them, inspired work on **subjective comparison** [Ledda et al. 2005]. Recently, the problem of image reproduction has gradually shifted towards displays, as they are equipped with more advanced image processing and display algorithms, which not only enhance the TV signal, but also adapt rendering to viewing conditions (ambient light sensor), save power (backlight dimming), **up-scale color gamut** [Muijs et al. 2006] and increase **dynamic range** [Meylan et al. 2006; Rempel et al. 2007].

3 Tone mapping as the minimum visible distortion problem

The original goal of the tone mapping problem, as formulated by Tumblin and Rushmeier [1991], is to reproduce a scene on a display so that the brightness sensation of a displayed image is equal to or closely matches real-world brightness sensation. The perfect match between the original and its rendering on a display or in a hard-copy format is almost never possible, as an output medium is hardly ever bright enough, though it offers sufficient dynamic range (contrast) and color gamut. Therefore, the rendering on an output device is a tradeoff between preserving certain image features at the cost of the others. For example, high contrast and brightness of an image can often be preserved only at the cost of clipping (saturating) a certain amount of pixels in bright or dark regions. The choice of which features are more important should be driven by a particular application, for which an appropriate metric could be designed, possibly involving some aspects of the visual perception. These considerations lead us to a general tone mapping framework, illustrated in Figure 2, which is formulated as an optimization problem.

Having an original image as input, which can be in HDR or any high quality format, we want to generate a display-adapted image that will be the best possible rendering of an original scene. We assume that this goal is achieved if the response of the HVS for an image shown on the display, R_{disp} , is as close as possible to the response evoked by the original scene, R_{orig} (refer to Figure 2). Both responses can almost never be identical as a display can only show limited dynamic range and color gamut. Also the viewing conditions, such as ambient light or luminance adaptation, differ between the original scene and its rendering, making a match even more difficult. The solution to the optimization problem is a set of tone mapping parameters that minimize the difference between R_{orig} and R_{disp} . Figure 2 contains also a processing block for image enhancement, as many applications require to reproduce images that are sharper and more colorful than the originals. The display model introduces physical constraints on devices' color and luminance reproduction.

The framework shares some similarities with the TMO originally proposed by Tumblin and Rushmeier over 15 years ago [1991, Fig. 2b], and in the latter work of Pattanaik et al. [2000]. The difference is that these approaches assume $R_{disp} = R_{orig}$ and then invert the HVS and display models to compute a tone mapped image. If we follow this approach and compute desired display luminance that evokes the same sensation as a real world scene ($R_{disp} = R_{orig}$), we can end up with an image that is too bright or has too much contrast for a given display. In such a situation, if we apply the limitations of a display and clamp luminance values, we get R_{disp} significantly different from R_{orig} , which is unlikely to be the global minimum of our optimization problem. Furthermore, our framework can be used with arbitrary HVS and display models, while previous approaches required the models to be invertible. More differences are discussed in Section 6.4.

The problem formulation above has been used before in the context of digital halftoning for printing [Pappas et al. 2003], but it has not been employed to derive a tone mapping operator. The major difficulty lies in the fact that even simplified models of a display, the HVS and a tone mapping operator lead to a complex non-linear optimization problem, which may exhibit local minima, or be too complex to solve in reasonable time. In the following subsections we will present a combination of such models, which are sufficiently complete to realize the goals outlined above, and at the same time lead to a standard optimization problem, which can be solved efficiently.

3.1 Tone Mapping

To make the problem solvable by reducing degrees of freedom of the optimized system, a tone mapping operator with a set of adjustable parameters must be introduced. To retain maximum flexibility, we employ a piece-wise linear tone-curve. Although the high contrast scenes seem to require local, detail-enhancing operators, we demonstrate that a well designed tone-curve can produce good results even for such scenes. Akyüz et al. [2007] also confirm the importance of a global tone-curve by showing that the results of sophisticated TMO are no better than the best single exposure. In this study we do not consider the color appearance issues, as we did not find the color appearance models robust enough for our purpose. Such models cope well with uniform color fields, however they usually do not consider the influence of spatial luminance modulation on color appearance, which we can observe when tone mapping high contrast images. To retain color information from the reference image, we employ the desaturated color-to-luminance ratios [Schlick 1994]:

$$R' = \left(\frac{R}{L} \right)^s L', \quad (1)$$

where L is the luminance, R the trichromatic value, L' is the tone-mapped luma and R' is the tone-mapped color channel. For our results $s = 0.6$.

3.2 Image Enhancement

Image enhancement modifies the original image to improve its appearance. Such enhancement is a common practice in many imaging products, such as cameras and advanced TV displays, as people tend to prefer images that are sharper, and have higher contrast and more saturated colors than the original scenes. Image enhancement is often used for stylization; for example higher contrast to give a desired harsh look, soft focus for making actors appear younger, or color shift to create a surrealistic mood. Colors can be altered to be closer to so-called *memory colors*, which are the colors that we remember seeing rather than the colors that can be measured in the

actual scene [Bodrogi and Tarczali 2002], which are usually more colorful than the colors of the original scene.

Recent studies [Yoshida et al. 2006] show that contrast enhancement as high as 100% can be desired if a display offers a sufficient dynamic range, however actual strength of the enhancement strongly depends on subjective preference. To avoid over-enhancement, we follow a common practice in producing reflective prints for the consumer market and enhance the contrast of a reference image by 15% [Hunt 2004, p. 55] ($e = 1.15$, as discussed in Section 4.2).

3.3 Display Model

The display model primarily accounts for the limited capabilities of a display device, such as maximum brightness, or dynamic range (contrast ratio). These are affected by the technical aspects of a display, as well as viewing conditions, such as ambient light that is reflected from a screen. Such reflected light elevates luminance of the darkest pixels shown on a display, thus reducing available dynamic range.

Most of the displays, both CRT and LCD, can be modelled with the formula:

$$L_d(L') = (L')^\gamma \cdot (L_{max} - L_{black}) + L_{black} + L_{refl} \quad (2)$$

where L_d is displayed luminance or radiance (as measured coming from the display surface), L' is luma or a pixel value (0–1), γ is a display gamma (usually close to 2.2), L_{max} is the peak display luminance (about 100 cd/m² for office displays, and about 500 cd/m² for LCD TV). L_{black} is the display black level, which is the luminance of the black pixel displayed in a perfectly dark room (usually from 0.1 to 0.8 cd/m² for LCD displays). L_{refl} is ambient light reflected from a display surface and it can be approximated in case of non-glossy screens with:

$$L_{refl} = \frac{k}{\pi} E_{amb} \quad (3)$$

where E_{amb} is ambient illuminance given in lux and k is the reflectivity for a display panel (about 1% for LCD displays, larger for CRT displays). Although the model from Equation 2 is usually employed separately for each trichromatic primary (red, green, blue), we use this model for luminance values only since the color issues are not in the scope of this work yet.

3.4 Human Visual System Model

The model of the human visual system (HVS) processes input luminance and chrominance data to produce the estimated response. Such estimate should scale image features relative to their visibility or importance. There are many choices of such models, ranging from the mean square difference, to complex appearance models. We decided to employ a model that estimates perceived contrast distortions, as contrast is one of the most important factors that affect overall image quality [Cadic et al. 2006]. In fact most of the visual models, employed to estimate perceived image difference [Lubin and Pica 1991; Daly 1993] or to drive a tone mapping operator [Pattanaik et al. 1998], operate on image contrast.

To estimate the response of the HVS to a contrast stimulus, we use a classical transducer function introduced by Wilson [1980], which is a function of contrast $W = \Delta L/L$ and sensitivity S : $R = T(W, S)$. The resulting value R is a hypothetical HVS response given in JND (Just Noticeable Difference) units. The original formula is included in the supplementary materials. Figure 3 shows two desirable properties of the transducer function: a) the contrast is attenuated below

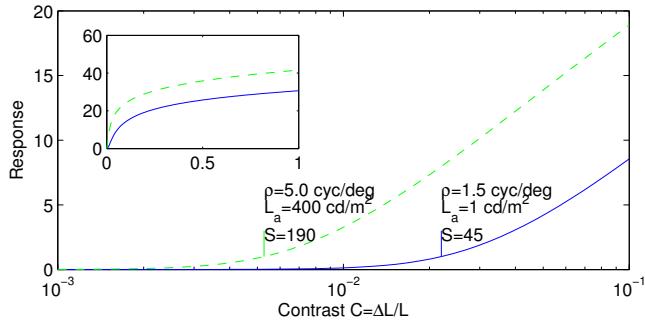


Figure 3: A contrast transducer function from [Wilson 1980]. The vertical lines show the detection thresholds and constants above indicate at which conditions they can be expected (ρ – spatial frequency, L_a – adapting luminance, S - sensitivity). The inset shows a compressive character of the transducer function for the larger contrast range.

the detection threshold, which prevents invisible noise from being considered as important and thus preserved by a tone mapper; and b) the contrast above the detection threshold is firstly enhanced and then compressed, since the visual system is the most sensitive for the contrast changes near the threshold and less sensitive for the contrast differences at high contrast levels (facilitation and contrast masking).

The sensitivity S is the inverse of the detection threshold and is modelled with the Contrast Sensitivity Function (CSF):

$$S = CSF(\rho, L_a, v_d), \quad (4)$$

where ρ is the frequency given in cycles per degree, L_a is the adapting luminance, given in cd/m^2 and v_d is the viewing distance. The luminance masking is modelled by assuming local adaptation, so that L_a is equal to local background luminance. We use in our system the CSF from [Daly 1993], which we include in the supplementary materials.

The models above let us find the response of the visual system given a contrast value W . To find a set of contrast values in a complex image, we employ the Laplacian pyramid. First, we compute a logarithm of image luminance values, $I = \log_{10}(L)$. Then, we use an efficient algorithm to compute the Gaussian pyramid [Burt and Adelson 1983] I_l for the image I . The contrast in the logarithmic ratio units for the l -th frequency band is then equal to $G_l = I_l - I_{l+1}$, where I_1 is the original image and larger l values indicate coarser levels of the Gaussian pyramid. The pyramid is contracted up to the band that has its medium frequency lower than 3 cycles per visual degree [cpd]. The sensitivity for the luminance patterns of frequency lower than 3 cpd drops rapidly and therefore they have little influence on our contrast metric. The formulas for computing the medium frequency of a band are included in the supplementary materials. Since the Laplacian pyramid represents contrast as the logarithmic ratios G and the transducer function was modelled for Weber contrast W , we need to convert between these units using the formula:

$$W = 10^{|G|} - 1 \quad (5)$$

4 Efficient solution

In this section we explain how the optimization problem can be efficiently solved for the display and the HVS models introduced in the previous section.

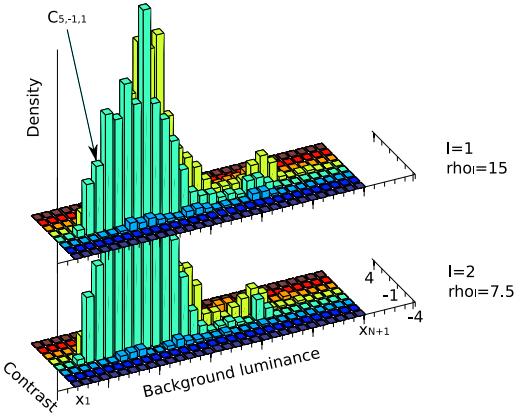


Figure 4: The conditional probability density function $c_{i,m,l}$ for the Memorial church image, the first two contrast pyramid levels.

4.1 Conditional contrast probability

Computing a response of the visual system on the entire image, which can easily exceed several mega-pixels, is a prohibitively expensive step, especially as a numerical solver for our optimization problem (refer to Figure 2) needs to execute it several hundred times. However, our particular problem allows us to run the visual model on a custom-designed higher order image statistic, which summarizes a common behavior of multiple pixels. Our image statistic can be understood as a conditional histogram of contrast values (G_l) that depends on logarithmic luminance of the local background (I_{l+1}) and a pyramid level l . Such a statistic, which we refer to as a conditional probability density function, lets us predict how many pixels will be affected by a particular tone mapping operation.

The visual illustration of the conditional probability density function for the *Memorial church* image is shown in Figure 4. To compute the function, we divide the dynamic range of an input image I into N bins of equal size and denote centers of these bins as $x_{i=1,\dots,N}$. The N is selected so that the difference between bins is about 0.1 (\log_{10} units, there are on average about 30 bins). Then, for each l -th level of the Laplacian pyramid and for each i -th bin, we compute the histogram of all contrast values G_l , whose corresponding local background I_{l+1} belongs to the i -th bin. The collection of such histograms gives us a conditional probability density function:

$$c_{i,m,l} = P\left(m \delta - \frac{\delta}{2} \leq G_l < m \delta + \frac{\delta}{2} \mid x_i - \frac{\delta}{2} \leq I_{l+1} < x_i + \frac{\delta}{2}\right), \quad (6)$$

where $\delta = x_{i+1} - x_i$ is the distance between bin centers, and $m = -M, \dots, -1, 1, \dots, M$. The use of contrast bins that have the same size as the x_i bins simplifies our further computations. The value of M is chosen such that $M \delta < 0.7$, which gives a good trade-off between the number of bins and the maximum contrast that is captured in the structure. The marked density value $c_{5,-1,1}$ in Figure 4 represents the relative count of the contrast values of spatial frequency ≈ 15 cycles per degree, of the background log-luminance $\approx x_5$ and the contrast magnitude $\approx -\delta$.

4.2 Objective function

Our tone curve is a piece-wise linear function with the nodes at the points (x_i, y_i) , as shown in Figure 5, where the values x_i are the same as for the conditional probability density function from Equation 6. The constraints on feasible tone curves can be more

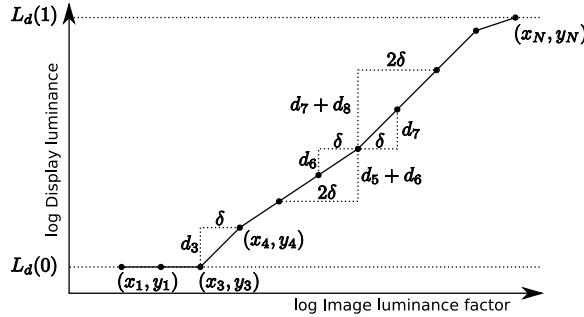


Figure 5: Example of a piece-wise linear tone curve and the notation used to specify it. Four contrast samples are shown for the node (x_7, y_7) . The samples are multiples of δ in an input image, and the sums of d_m in a tone-mapped image.

conveniently specified, if instead of actual values, we operate on differences $d_i = y_{i+1} - y_i$ for $i = 1..N - 1$. Then our goal is to find a visual error due to a tone curve given by x_i , d_i , minimum $L_d(0)$ and maximum $L_d(1)$ logarithmic luminance of a display (refer to Equation 2).

The minimum contrast distortion can be expected if the tone curve has the slope close to 1 (log-log plot), and thus the contrast responses for the original and displayed images are the closest to each other. If we consider for each tone level x_i contrast magnitudes that are multiples of δ , our goal is to minimize all squared differences between the contrast response given by the tone curve, and the contrast response of the original image:

$$\arg \min_{d_1, \dots, d_{N-1}} = \sum_{l=1}^{N-1} \sum_{i=1}^M \sum_{\substack{m=-M \\ m \neq 0}}^M \left[T(\sum_{k \in \phi} d_k, S_d) - T(e \sum_{k \in \phi} \delta, S_r) \right]^2 \cdot c_{i,m,l} \quad (7)$$

such that

$$d_i \geq 0 \quad \text{for } i = 1..N-1$$

$$\sum_{i=1}^{N-1} d_i \leq L_d(1) - L_d(0) \quad \text{for } i = 1..N-1 \quad (8)$$

where $\phi = i+m, \dots, i-1$ for $m < 0$ and $\phi = i, \dots, i+m-1$ for $m \geq 0$ is the range of differences d_k that form a contrast magnitude between tone level x_i and x_m . The constant e is the constant enhancement factor, which multiplies the reference image contrast as discussed in Section 3.2. The first constraint ensures that the tone curve has a non negative slope and the second that we do not exceed the display's dynamic range. T is the transducer function, introduced in Section 3.4.

Due to the difference in viewing conditions, sensitivity can be in fact different for a displayed image and a reference image ($S_d \neq S_r$). For example, we often want to see images of higher perceivable contrast, as if our eyes were adapted to high luminance levels and thus very sensitive. We can achieve this if we assume that the luminance of adaptation $L_a = 1000 \text{ cd/m}^2$ (or any large value) and thus $S_r = \text{CSF}(\rho_l, 1000, d)$. However, when an image is displayed, the HVS is much less sensitive for darker pixels shown on a display, and the sensitivity equals to $S_d = \text{CSF}(\rho_l, 10^{y_i}, d)$, where 10^{y_i} is the display luminance for the tones i . This step introduces dependence of the display model on the result of the optimization problem by enforcing larger contrast on darker displays to compensate for the loss of sensitivity.

Multiplication by a probability density function $c_{i,m,l}$ lets us relate the error to the amount of contrast values of particular background

luminance (i), contrast magnitude (m) and spatial frequency (l) in an image. This gives a major performance improvement, as multiple contrast instances in an image are summarized in one term. The use of the probability density function also makes the optimization problem independent of the image size, thus making it suitable also for high resolution images.

4.3 Fast quadratic solver

Equation 7 represents a non-linear optimization problem, which cannot be efficiently solved using standard methods. We can, however, reduce it to a standard problem, if we approximate the non-linear transducer function T with a linear scaling constant. The left transducer term from Equation 7 can be written in a matrix notation as:

$$T(A d, S_d) \approx K A d \quad (9)$$

where d is a column vector of d_1, \dots, d_{N-1} , A is a $0/1$ matrix with $N-1$ columns, where each row j represents one term of the three sums (over l , i and m) from Equation 7 and $A_{jk} = 1$ for $k \in \phi$. K is a diagonal matrix that contains the scaling factors approximating the non-linear transducer:

$$K_{jj} = \frac{T([A d]_j, S_d)}{[A d]_j} \quad (10)$$

To avoid singularities, $K_{jj} = 0$ for $[A d]_j = 0$ (response for no contrast). Since the right transducer term in Equation 7 does not depend on the vector d , it can be precomputed and stored as a column vector B . Then, the objective function from Equation 7 can be written in the matrix notation as:

$$[K A d - B]^T C [K A d - B] = \quad (11)$$

$$d^T A^T K^T C K A d - 2 B^T C K A d^T + B^T C B$$

where C is a diagonal matrix with the density function values $c_{i,m,l}$. Equation 11 represents a quadratic programming problem, of the dimensionality $N-1$, which, given that $N \approx 30$, can be solved very efficiently using the standard method [Gill et al. 1981]. Still, the solution is valid only for the scaling factors K , and not for the transducer function T . To find the result for the transducer function T , we solve the quadratic problem iteratively, each time computing a new scaling factor K and using the result d from the previous iteration. For our transducer function, the optimization usually converges ($\Delta d_i < 0.1 \delta$) in 3-7 iterations.

The problem can be ill defined if some luminance levels x_i are not linked with other luminance levels by any contrast in an image ($c_{i,\dots} = 0$). To make such a problem solvable, the corresponding columns should be eliminated from the matrix A and rows from the vector d . Since the matrix $A^T K^T C K A$ (after removing columns) is positive definite, the quadratic program has a global minimizer, which is unique.

4.4 Final tone curve and inverse display model

In the last two steps, we recover the final position of the tone-curve nodes:

$$y_i = L_d(0) + \sum_{k=1}^{i-1} d_k + \alpha (L_d(1) - L_d(0) - \sum_{k=1}^{N-1} d_k) \quad (12)$$

The last term shifts overall image brightness according to the factor $\alpha \in <0; 1>$, in case the displayed image dynamic range is lower than the dynamic range of a display (e.g. when displaying low dynamic range images on an HDR display). We set $\alpha = 1$ to display

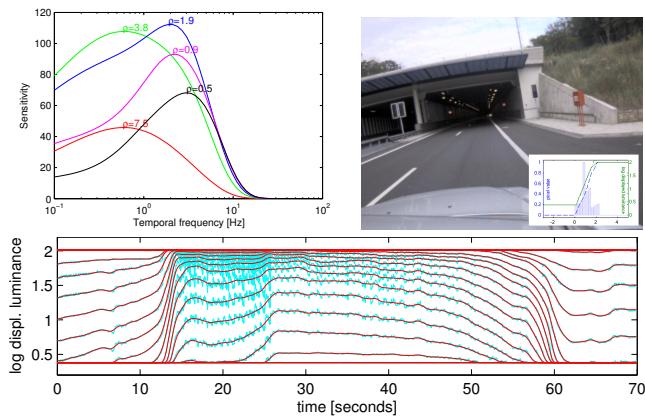


Figure 6: Top-left: spatio-temporal contrast sensitivity function for several spatial frequencies ρ from [Kelly 1979]. Top-right: a frame from the Tunnel sequence. Bottom: coordinates of the tone curve nodes y_i before (cyan) and after (red) temporal filtering for the Tunnel sequence.

possibly bright images, but the coefficient α can be also related to the scene brightness to distinguish between low-key and high-key scenes. Finally, display luminance values are transformed to pixel values using the inverse of Equation 2 (inverse display model).

4.5 Timings

The non-optimized matlab implementation requires on average 1.7 seconds on 2.6GHz CPU to tone-map a 1M-pixel image. More than half of that time is spent on computing the probability density function (Equation 6), which is difficult to program efficiently in matlab. The quadratic programming solver accounts for only 9% of that time.

5 Extension to video

Our method should not be applied to a video sequence directly, since a tone-curve can change rapidly between consecutive frames, which can result in annoying flickering. It might be tempting to add an additional cost term to the objective function (Equation 7) to penalize temporal changes in the tone curve, but such an approach would not guarantee that the changes are not happening. Instead, the nodes of the tone-curve can be filtered in the time domain. The peak sensitivity of the HVS for temporal changes depends on the spatial frequency and varies from ≈ 0.5 to ≈ 4 Hz, as shown in Figure 6 top-left. To ensure that the frame-to-frame changes of a tone-curve are not salient, we do not allow for temporal variations above 0.5 Hz. We apply a windowed linear-phase FIR digital filter to the node coordinates y_i , assuming that x_i coordinates are the same for all frames. The filter is low-pass and has a cutoff frequency of 0.5 Hz. Figure 6 bottom shows the result of the filtering for the *Tunnel* sequence, which is included in the supplementary video.

6 Results

In this section we demonstrate the display-adaptive capabilities of our tone mapping method and validate them in a subjective study. Then we compare our technique with other popular methods. All low-dynamic range images shown in the results have been converted to the linear trichromatic values assuming the sRGB color space. Additional result images in full resolution can be found

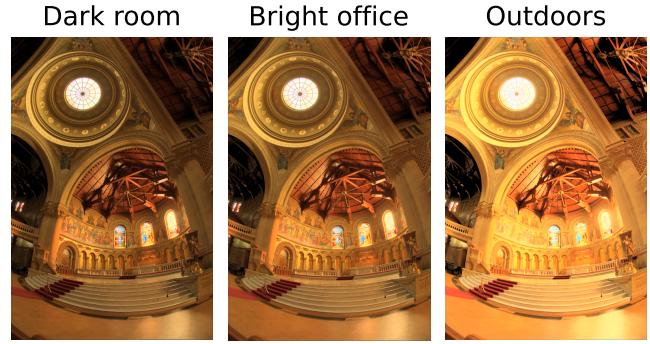


Figure 8: The Memorial church images tone-mapped for three different ambient illumination conditions. As ambient light increases, the images get brighter to avoid dark tones, which are the most affected by the screen reflections. For the outdoors illumination larger part of the bright pixels is clipped (saturated) and image contrast is increased. Note that these images are input to the display and do not depict actually displayed images.

on the project web page: <http://www.mpi-inf.mpg.de/resources/hdr/datmo/>.

6.1 Mobile display in the sunlight

A cell phone display is very hard to read in full sunlight, especially if the display is transmissive (modulates transparency), rather than transreflective (modulates both transparency and reflectivity). This common situation, in which the same mobile display is seen in three different illumination conditions, ranging from a dark room to outdoors on a sunny day, is simulated and shown in Figure 7. The top row shows how the effective dynamic range of a display gets compressed due to screen reflections, making lower tone values almost indistinguishable. Our tone mapping attempts to compensate for this by increasing the contrast of lower mid-tones (the lower part of the dashed-blue curves gets steeper with brighter ambient light). As the dynamic range of the display gets lower, an image is reproduced at lower contrast (compare solid-green tone curves). The tone-curve is also determined by image content and its dynamic range. A high dynamic range image, such as the *Tree* in the fourth row, is reproduced at lower contrast than the low dynamic range *Mantis* image in the third row (compare the tangents with the green dash-dot lines). Note that images shown in the figure give only a general impression how the images may look on a display, and they do not represent tone-mapped images. An example of the actual result of tone mapping, and how it differs for different ambient light conditions, is shown in Figure 8.

6.2 Display technologies

The diversity of the display technologies makes it very difficult to predict how the image will look to the user when it is displayed on a random device. The display adaptive tone mapping can compensate for the differences in display characteristic and additionally make the best possible use of the available display contrast and brightness. Figure 9 shows images displayed using three totally different display technologies: a hypothetical color e-paper display, based on the actual specifications of e-paper; a typical LCD display in a dim room; and a high-brightness HDR display, also in a dim room. The e-paper display offers the worst contrast, which enforces the use of very particular tone curves. The contrast in a part of mid-high tones is almost completely flattened for the *Napa valley* image and the e-paper display (1st column and 3rd row), since the contrast val-

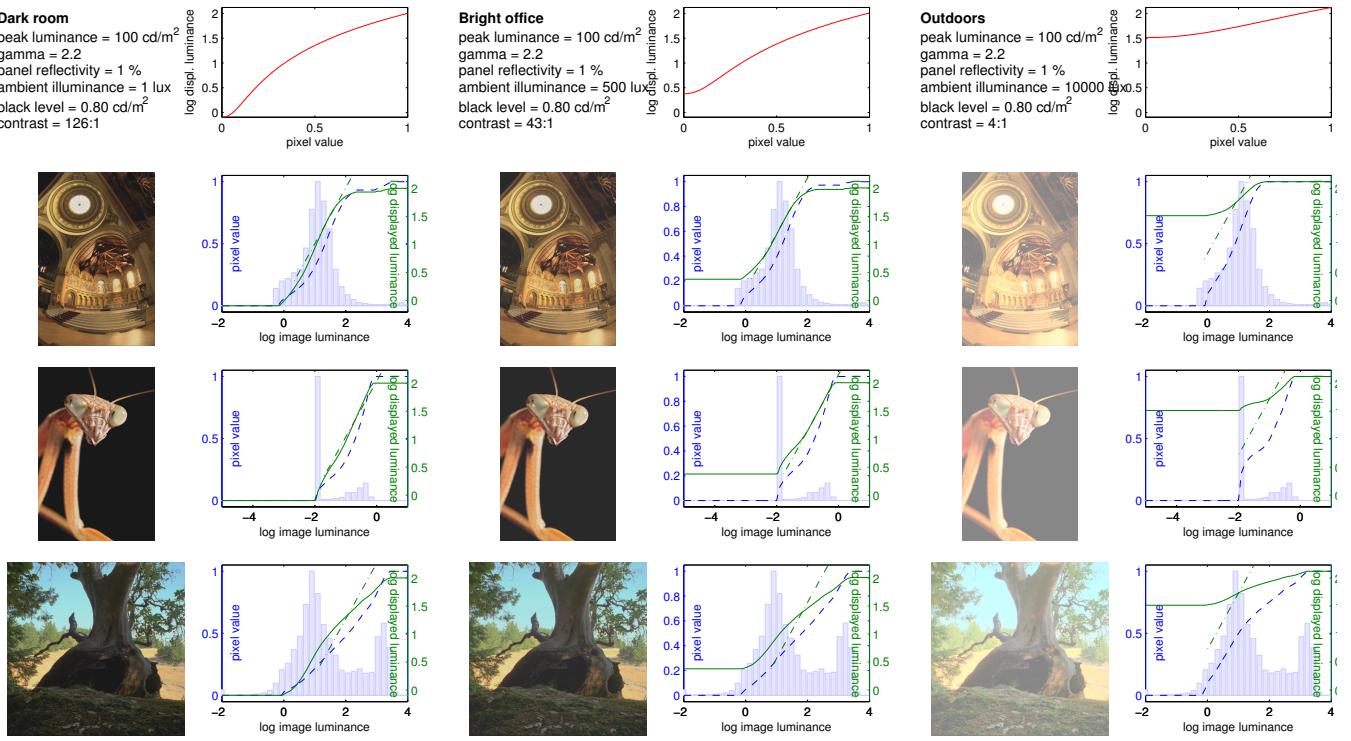


Figure 7: Images tone-mapped for a mobile display under different illumination conditions. The top row describes the display model parameters and its luminance response in each scenario (dark room, bright office, outdoors). Each plot in the rows below contains two tone curves that map image log luminance factor to either pixel values (blue) or display log luminance levels (green). The dash-dotted green line represents slope=1 (no contrast change). The light-blue bars show image histogram. The images depict simulated image appearance on a display, which however does not convey actual contrast or brightness due to print limitations. Please refer to the supplementary materials for the full resolution images and more examples.

ues in this range convey the least useful information (large contrast between the sky and the valley). The HDR display, on the other hand, can employ more regular tone curves, with profound contrast boost in lower mid-tones to compensate for lower sensitivity for dark pixel values.

6.3 Experimental validation

In the following experiment we validate our claim that the display adaptive tone mapping can improve overall image contrast under varying ambient illumination. The experiment involved a pair-wise comparison between an original standard dynamic range image, the result of our method at ambient illumination of 20 lux (dark room), and at 1600 lux (simulated sunlight). The images were displayed on the self-calibrating Barco Coronis Color 3MP display, which was set to the maximum luminance $L_{max} = 440 \text{ cd/m}^2$ and $\gamma = 2.2$. For the 20 lux scenario the lights in a room were dimmed. For the 1600 lux scenario we directed two photographic lights (K5600 Joker-Bug 800W) on the display, so that the light reflected from the screen was possibly uniform (the setup is shown in two small insets in Figure 1). We measured the display response with the Minolta LS-100 luminance meter for both scenarios and used it as a display model to generate images using our method. Nine participants, who were naive about the purpose of the experiment, took part. Each participant was asked the question “*Choose an image which has better overall contrast, looks sharper and reveals more details*”, before comparing 6 scenes \times 3 method combinations \times 3 repetitions = 54 pairs.

The average score (the number of times the image was selected) is

computed for both scenarios and summarized in the table below:

Dark room (20 lux)			Sunlight (1600 lux)		
DAT-20	DAT-1600	original	DAT-1600	DAT-20	original
1.56	0.99	0.46	1.96	0.90	0.14

where DAT-20 and DAT-1600 denotes display adaptive tone mapping for 20 lux and 1600 lux conditions. The χ^2 test on the Kendall coefficient of agreement u [Kendall and Babington-Smith 1940] indicated good consistency between participants for all image pairs, except one pair for the 20 lux scenario. The multiple comparison test indicated a statistically significant difference in overall scores ($p = 0.05$), although for some image pairs the difference between the second and the third ranked method was not significant. The results show that the images generated using our method were preferred to the original images, most probably because of better tone-scale allocation, which gave sharper results. The ranking of the DAT-20 and DAT-1600 images matched the ambient light level for which they were generated, which suggests that the display adaptive tone mapping can improve the contrast images shown on displays in bright environments.

6.4 Comparison with other methods

The purpose of this comparison is not ranking operators, since each operator has its own goals and merits, but rather showing differences in the underlying approaches. We choose for the comparison three popular global methods: the **photographic TMO** (only global algorithm) [Reinhard et al. 2002] using the implementation from the DVD attached to the book [Reinhard et al. 2005], the **his-**

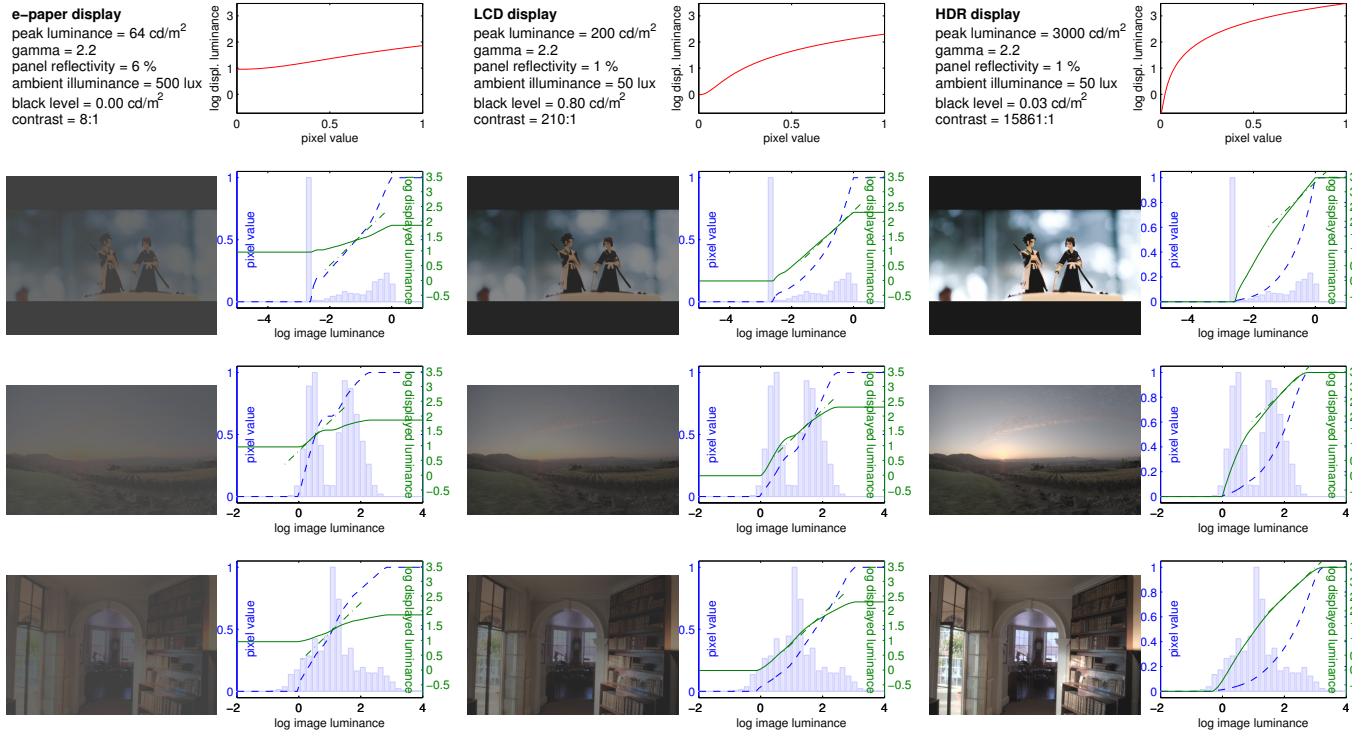


Figure 9: Images tone-mapped for different display technologies. The notation is the same as in Figure 7

togram adjustment TMO [Ward Larson et al. 1997] using the implementation from the Radiance package [Larson and Shakespeare 1998], and the **visual adaptation TMO** [Pattanaik et al. 2000] using the implementation from the *pfstmo* package [*pfstmo*]. Three images that we discuss below are shown in Figure 10 and the remaining images are included in the supplementary materials.

Both the histogram adjustment and our method can finely adjust tone-curve to image content, for example by compressing poorly represented mid-tones, as in the image *Bristol bridge* (1st row). Such non-trivial tone-curve results in better contrast in low and high tones than a pre-determined sigmoidal tone curve. The sigmoidal tone-curve used in the photographic TMO on the other hand results in better global contrast and is more consistent with the typical tone scale used in the photography. The next two low-dynamic range images (2nd and 3rd row) demonstrate how a low frequency background can affect the histogram adjustment TMO. Since the off-focus background plane occupies the larger portion of these images, the histogram adjustment TMO allocates for the bright background tones a larger portion of the dynamic range, but compresses darker tones, making the groom figure and goose's head too dark. The display adaptive TMO is less affected by the low-frequency background, which does not contain much contrast information.

The visual adaptation TMO (last column) is an example of a totally different approach to the tone mapping problem. The goal of this method is a possibly accurate simulation of the HVS adaptation processes and its limitations, including limited range of the photoreceptor response. The method faithfully preserves original scene contrast, but in a very small window of the scene dynamic range, and clips all tones that fall outside this window. Therefore, the method is not intended for producing visually attractive images from high contrast scenes, but rather for simulating contrast visibility in different states of luminance adaptation.

The example of the visual adaptation operator indicates the major

difference between our approach and most of the methods that employ a HVS model. These methods usually aim at producing images that are fully or partly processed by a HVS model, and account for such visual effects as loss of acuity at low light, visual glare, saturation of the photoreceptors, or local adaptation [Pattanaik et al. 1998; Thompson et al. 2002]. As discussed in Section 3, these methods compute visual response, which is then converted back to the luminance units using inverse models. In our approach we produce the results that are as close to the original (or enhanced) image as possible. Therefore, the HVS model is employed to penalize distortions, rather than to simulate perceptual effects. These two different approaches are not contradictory, and in fact a simulation of the perceptual effects could be a part of the *enhancement* block from the conceptual diagram in Figure 2.

7 Conclusions

This paper introduces a technique for reproducing scene-referred images on displays of limited contrast by minimizing visible distortions. The method finds a compromise between conflicting goals, such as preserving contrast and clipping the darkest and the brightest tones. The distortions are penalized using the HVS contrast perception model. The display model predicts display luminance response and imposes luminance limitations on the reproduced image. The algorithm leads to a unique, well defined solution, with no subjective parameters. The operator is used to tone-map images adaptively for display's dynamic range, brightness and ambient illumination.

Many recent studies on tone mapping undertake the difficult task of producing images that will be subjectively preferred. We avoid this challenging problem from the area of computational aesthetics by hiding it in the *image enhancement* step. Although many attempts to model subjective preference or quality of images have been made [Keelan 2002], there exist no reliable models that would

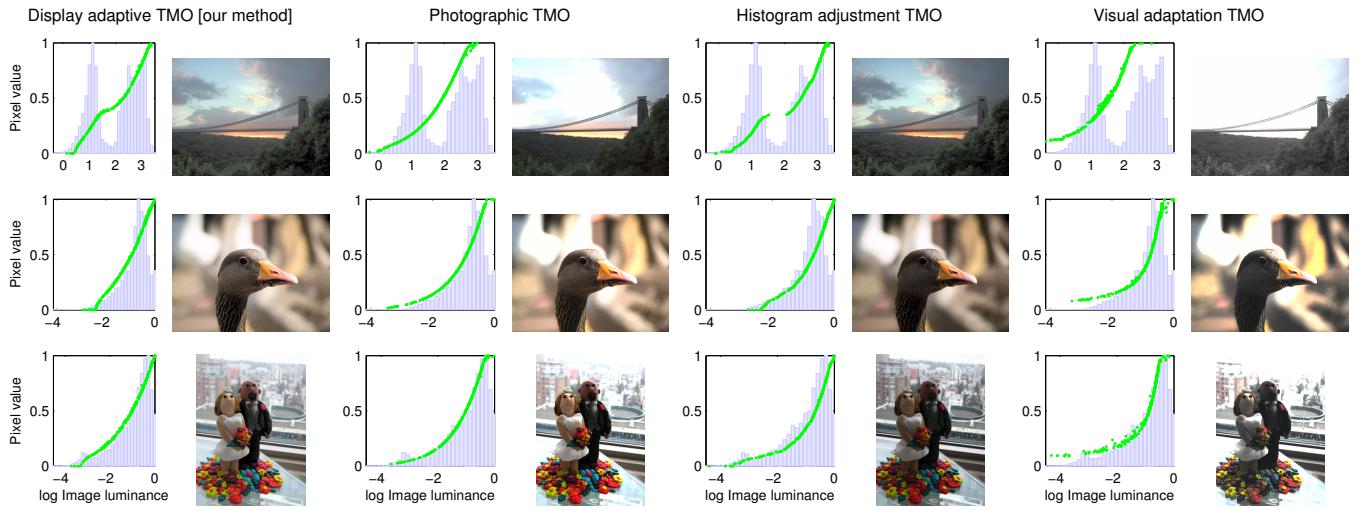


Figure 10: Comparison with other tone mapping methods. The plots show the distribution of image luminance (factor) and pixel values. Note, that even though the operators are global, the distributions do not form perfect curves due to non-linear color processing and clipping. For full size images and more examples, refer to the supplementary materials.

measure subjective image preference. For this reason we do not claim an operator that produces the best looking images (although in our opinion in many cases it does), but rather the operator that objectively solves the problem of reproducing large dynamic range on a displays of low contrast with the least visible contrast distortions.

In this paper we also propose the concept of a tone mapping closely coupled with a display device, which renders images optimized for a particular display and under the existing viewing conditions (ambient light). For example, a mobile phone should change its rendering algorithm when the backlight of its transreflective display is switched off to save power. Similarly, a TV display should adjust the display algorithm when light in a room is lit (simple dimming due to ambient illumination is already performed in some TV displays).

In future work we would like to address color issues and local tone mapping operations (e.g. sharpening). Our initial studies show promising results, although some fundamental problems need to be solved before these extensions are possible. For example, it is still not clear when a strong sharpening is perceived as a haloing artifact and when it can be considered as a desirable contrast boost (Cornsweet illusion), although some research on this problem has been done [Krawczyk et al. 2007].

Another extension of the algorithm can take advantage of the region of interest (ROI) information. The ROI information is already available in modern cameras in the form of face detection algorithms. Weighing ROI by the distance from the frame center or employing attention models [Le Meur et al. 2006] can be another choice. Such ROI weighting could be used to increase the importance of preserving contrast in certain parts of a scene, which can be included in the conditional probability density function (Equation 4).

Acknowledgments

We would like to thank Karol Myszkowski, Robert Herzog, Juliet O'Keefe and anonymous reviewers for their helpful comments. The images and video sequences used in this work are the courtesy of Greg Ward, Paul Debevec, Grzegorz Krawczyk, SpheronVR and the Flickr community (steena, ToastyKen, Gary VanDenBerg, scar-

latti2004, Cayusa, Destinys Agent, jystyn).

References

- ADAMS, A. 1981. *The Print, The Ansel Adams Photography Series* 3. New York Graphic Society.
- AKYÜZ, A. O., FLEMING, R., RIECKE, B. E., REINHARD, E., AND BÜLTHOFF, H. H. 2007. Do HDR displays support LDR content?: a psychophysical evaluation. *ACM Trans. Graph.* 26, 3, 38.
- BAE, S., PARIS, S., AND DURAND, F. 2006. Two-scale tone management for photographic look. *ACM Trans. Graph.* 25, 3, 637–645.
- BODROGI, P., AND TARCZALI, T. 2002. Investigation of colour memory. In *Colour Image Science*, L. MacDonald and M. R. Luo, Eds. John Wiley & Sons, ch. 2, 23–48.
- BURT, P., AND ADELSON, E. 1983. The laplacian pyramid as a compact image code. *Communications, IEEE Transactions on [legacy, pre-1988]* 31, 4, 532–540.
- CADIK, M., WIMMER, M., NEUMANN, L., AND ARTUSI, A. 2006. Image attributes and quality for evaluation of tone mapping operators. In *Proc. of the 14th Pacific Conf. on Comp. Graph. and Applications*, 35–44.
- DALY, S. 1993. The Visible Differences Predictor: An algorithm for the assessment of image fidelity. In *Digital Image and Human Vision*, MIT Press, A. Watson, Ed., 179–206.
- GILL, P., MURRAY, W., AND WRIGHT, M. 1981. *Practical optimization*. Academic Press.
- HUNT, R. 2004. *The Reproduction of Colour in Photography, Printing and Television: 6th Edition*. John Wiley & Sons.
- KEELAN, B. 2002. *Handbook of Image Quality: Characterization and Prediction*. Marcel Dekker.
- KELLY, D. 1979. Motion and vision. I. stabilized images of stationary gratings. *J. Opt. Soc. Am.* 69, 9, 1266–1274.

- KENDALL, M., AND BABINGTON-SMITH, B. 1940. On the method of paired comparisons. *Biometrika* 31, 3-4, 324–345.
- KRAWCZYK, G., MYSZKOWSKI, K., AND SEIDEL, H.-P. 2007. Contrast restoration by adaptive countershading. *Computer Graphics Forum* 26, 3, 581–590.
- 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, 5, 406–414.
- LAND, E. H., AND MCCANN, J. J. 1971. Lightness and the retinex theory. *Journal of the Optical Society of America* 61, 1, 1–11.
- LARSON, G., AND SHAKESPEARE, R. 1998. *Rendering with radiance: the art and science of lighting visualization*. Morgan Kaufmann Publishers Inc. San Francisco, CA, USA.
- LE MEUR, O., LE CALLET, P., BARBA, D., AND THOREAU, D. 2006. A coherent computational approach to model bottom-up visual attention. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 28, 5, 802–817.
- LEDDA, P., CHALMERS, A., TROSCIANKO, T., AND SEETZEN, H. 2005. Evaluation of tone mapping operators using a high dynamic range display. *ACM Trans. Graph.* 24, 3, 640–648.
- LISCHINSKI, D., FARBMAN, Z., UYTENDAELE, M., AND SZELISKI, R. 2006. Interactive local adjustment of tonal values. *ACM Trans. Graph.* 25, 3, 646–653.
- LIVINGSTONE, M. 2002. *Vision and Art: The Biology of Seeing*. Harry N. Abrams.
- LUBIN, J., AND PICA, A. 1991. A non-uniform quantizer matched to the human visual performance. *Society of Information Display Int. Symposium Technical Digest of Papers*, 22, 619–622.
- MEYLAN, L., DALY, S., AND SUSSTRUNK, S. 2006. The reproduction of specular highlights on high dynamic range displays. In *Proc. of the 14th Color Imaging Conference*.
- MORONEY, N., FAIRCHILD, M., HUNT, R., LI, C., LUO, M., AND NEWMAN, T. 2002. The CIECAM02 color appearance model. In *Proc. IS&T/SID 10th Color Imaging Conference*, 23–27.
- MOROVIC, J., AND LUO, M. 2001. The fundamentals of gamut mapping: A survey. *Journal of Imaging Science and Technology* 45, 3, 283–290.
- MUIJS, R., LAIRD, J., KUANG, J., AND SWINKELS, S. 2006. Subjective evaluation of gamut extension methods for wide-gamut displays. In *Proc. of the 13th International Display Workshop*, 1429–1432.
- PAPPAS, T., ALLEBACH, J., AND NEUHOFF, D. 2003. Model-based digital halftoning. *Signal Processing Magazine, IEEE* 20, 4, 14–27.
- PATTANAIK, S. N., FERWERDA, J. A., FAIRCHILD, M. D., AND GREENBERG, D. P. 1998. A multiscale model of adaptation and spatial vision for realistic image display. In *Proc. of SIGGRAPH'98*, 287–298.
- PATTANAIK, S., TUMBLIN, J., YEE, H., AND GREENBERG, D. 2000. Time-dependent visual adaptation for realistic image display. In *Proc. of SIGGRAPH 2000*, 47–54.
- PFSTMO. Library of tone-mapping operators. <http://www.mpi-inf.mpg.de/resources/tmo/>.
- REINHARD, E., STARK, M., SHIRLEY, P., AND FERWERDA, J. 2002. Photographic tone reproduction for digital images. *ACM Trans. Graph.* 21, 3, 267–276.
- REINHARD, E., WARD, G., PATTANAIK, S., AND DEBEVEC, P. 2005. *High Dynamic Range Imaging. Data Acquisition, Manipulation, and Display*. Morgan Kaufmann.
- SEETZEN, H., TRENTACOSTE, M., SEETZEN, H., YOUNG, H. D., HEIDRICH, W., WHITEHEAD, L., AND WARD, G. 2007. LDR2HDR: On-the-fly reverse tone mapping of legacy video and photographs. *ACM Trans. on Graph.* 26, 3, 39.
- SCHLICK, C. 1994. Quantization techniques for the visualization of high dynamic range pictures. In *Proc. 5th Eurographics Workshop on Rendering*, 7–18.
- SEETZEN, H., HEIDRICH, W., STUERZLINGER, W., WARD, G., WHITEHEAD, L., TRENTACOSTE, M., GHOSH, A., AND VOROZCOVS, A. 2004. High dynamic range display systems. *ACM Trans. Graph.* 23, 3, 757–765.
- SMITH, K., KRAWCZYK, G., MYSZKOWSKI, K., AND SEIDEL, H.-P. 2006. Beyond tone mapping: Enhanced depiction of tone mapped HDR images. *Computer Graphics Forum* 25, 3, 427–438.
- STOCKHAM, T. 1972. Image processing in the context of a visual model. *Proc. of the IEEE* 60, 7, 828–842.
- THOMSPON, W. B., SHIRLEY, P., AND FERWERDA, J. A. 2002. A spatial post-processing algorithm for images of night scenes. *Journal of Graphics Tools* 7, 1, 1–12.
- TUMBLIN, J., AND RUSHMEIER, H. E. 1991. Tone reproduction for realistic computer generated images. Technical Report GIT-GVU-91-13, Graphics, Visualization, and Usability Center, Georgia Institute of Technology.
- WARD LARSON, G., 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.
- WILSON, H. 1980. A transducer function for threshold and suprathreshold human vision. *Biological Cybernetics* 38, 171–178.
- YOSHIDA, A., MANTIUK, R., MYSZKOWSKI, K., AND SEIDEL, H.-P. 2006. Analysis of reproducing real-world appearance on displays of varying dynamic range. *Computer Graphics Forum* 25, 3, 415–426.