Analysis of Reproducing Real-World Appearance on Displays of Varying Dynamic Range

Akiko Yoshida, Rafał Mantiuk, Karol Myszkowski and Hans-Peter Seidel[†]

MPI Informatik, Saarbrücken, Germany

Abstract

We conduct a series of experiments to investigate the desired properties of a tone mapping operator (TMO) and to design such an operator based on subjective data. We propose a novel approach to the tone mapping problem, in which the tone mapping parameters are determined based on the data from subjective experiments, rather than an image processing algorithm or a visual model. To collect this data, a series of experiments are conducted in which the subjects adjust three generic TMO parameters: brightness, contrast and color saturation. In two experiments, the subjects are to find a) the most preferred image without a reference image (preference task) and b) the closest image to the real-world scene which the subjects are confronted with (fidelity task). We analyze subjects' choice of parameters to provide more intuitive control over the parameters of a tone mapping operator. Unlike most of the researched TMOs that focus on rendering for standard low dynamic range monitors, we consider a broad range of potential displays, each offering different dynamic range and brightness. We simulate capabilities of such displays on a high dynamic range (HDR) display. This allows us to address the question of how tone mapping needs to be adjusted to accommodate displays with drastically different dynamic ranges.

Categories and Subject Descriptors (according to ACM CCS): I.3.8 [Computer Graphics]: High dynamic range images, Visual perception, Tone mapping

1. Introduction

A great variety of tone mapping operators have been developed in recent years (refer to [RWPD05] for a detailed survey) in response to accessible and simple high dynamic range (HDR) image acquisition technology. A majority of existing operators are designed to produce images that just "look good". Some operators, especially those designed specifically for realistic image synthesis applications, use models of brightness or contrast perception to achieve a good match between the image's appearance and the corresponding real-world scene. In practice, each operator boils down to an image processing algorithm that transforms HDR pixels into their LDR counterparts using either a monotonic function with respect to the HDR pixel intensity (global operators) or a more complex relation that involves local pixel neighborhood considerations (local operators). While new tone mapping operators are proposed, there is little under-

standing whether their improvements and additional complexity really lead to better images. It turns out that it is difficult to select one existing operator that consistently performs the best in terms of user preferences or fidelity to the original scene appearance for all HDR images [RWPD05].

Evaluation of tone mapping operators (TMO) is an active research area [DMMS02, KYJF04, LCTS05, YBMS05], which at the current stage is more focused on choosing correct psychophysical techniques than on providing any clear guidance as to how existing operators should be improved to produce consistently high quality images. All existing evaluation methods treat each tested TMO as a "black box" and its performance is compared with respect to other operators, without explaining the reasons underlying human judgments. While some evaluation methods go one step further and attempt to analyze the reproduction quality of overall brightness, global contrast, and details (in dark and bright image regions) [LCTS05, YBMS05], but again they are focused on comparing which operator is better for each of these tasks. Those studies do not provide any deeper analy-

---

† [yoshida,mantiuk,karol,hpseidel]@mpi-inf.mpg.de

sis as to how pixels of an HDR image have been transformed and what the impact of such a transformation is on desired tone mapped image characteristics [DZB05]. Another important question is how the outcome of the transformation depends on the particular HDR image content.

In all discussed evaluation experiments only one set of parameters per TMO and per HDR image is considered in order to reduce the number of images that must be compared by subjects. The choice of the parameters may dramatically affect the appearance of tone mapped images and thus the performance of a TMO. Experimenters commonly set such parameters based on their choice or a small pilot study, which may lead to the results that are biased by the choices of a limited number of subjects. Sometimes the original authors of TMOs are asked to prepare images according to their preference, since "they should be the best qualified to get the best results". However, different people may have different ideas concerning the preferred image appearance and the meaning of "the best results" can be fuzzy. Even the calibration of the display used for image tuning and the actual experiment can affect the results. A limited number of TMOs offer a method of an automatic parameter estimation (e.g. [Rei02, RD05]). However, these estimation methods rely mostly on the authors observation for a small set of images or the practices borrowed from photography, rather than a study with a large number of images and subjects.

A standard pool of HDR images for tone mapping evaluation has been recently proposed [Joh05], but a common practice for every experimenter is to rely on his own version of tone mapped images, which makes cross-comparison of the results for independent evaluations difficult.

Another common problem is averaging the experimental results across subjects based on low-cross subject variability. This lack of variability can often be caused by the choices imposed on the subjects by the experiment design, which does not offer any possibility of adjusting the image appearance to the subject's real preferences within available range of parameters of the tested TMOs. The net result of published studies is that they often present contradictory results even if the same HDR images are used. Some operators shown as performing the worst in one experiment obtain the top scores in another experiment. This suggests that the TMO evaluation methodology should be improved.

In this work, instead of proposing a new TMO and then running the subjective evaluation to show that it performs better than the other operators, we take the opposite approach. We want to first identify the output tone characteristics that lead to perceptually attractive images. Therefore, we start from measuring the subjective preference and the perception of fidelity for images produced by a generic TMO, whose characteristic and parameters are well understood. Our goal is to find some universal rules that facilitate a design of the TMO that consistently produces preferred image appearance.

In this respect, there are some similarities of our approach goals with more fundamental research in psychophysics, which raises the issue of image appearance preferences as a function of various image characteristics. For example, Fedorovskaya et al. [FdB97] report that the relation between preference and colorfulness has a shape of inverted "∪" with the maximum preference achieved for color saturation increased by 10%–20% in respect to the original image. Similar results are obtained for contrast and brightness manipulation [JRW02]. Higher color saturation is needed to compensate for reduced brightness of a display in order to achieve more natural image appearance [deR96]. The preferred mean luminance levels are found for images that contain human faces [DZB05]. Image preferences with respect to colorfulness, contrast and brightness are studied in digital photography [SEL00].

What makes our study different from this fundamental research, which is motivated by the applications of color reproduction in television and photography, is that we focus on the particular problem of tone mapping HDR images for a broad range of display devices. For this purpose we use in our experiments HDR images, which are displayed on an HDR display with fully controllable minimum and maximum display luminance values. Therefore, we can investigate much wider dynamic range than is possible using traditional LDR displays and neutral density filters.

2. Experimental Design

We conducted two experiments on an HDR display to assess how people adjust the settings of a TMO. In Experiment 1, the subjects were given the task of adjusting an HDR image shown on the HDR display so that it looked the best in their opinion. In Experiment 2, the subjects sat in front of an HDR display showing an image and the corresponding real-world scene. Their task was to adjust the image to achieve the closest reproduction of the real-world scene on the display. Additionally, in Experiment 2 we simulated several potential display devices by limiting the lowest and highest luminance outputs of the HDR display.

2.1. Subjects

There were in total 24 individual participants in two experiments. Four of them were female and the rest were male. The range of their age was 24 – 46 years and the average was 28. All of them reported normal or corrected to normal vision. All but two subjects were not aware of the purpose of the experiments. Eight subjects took part in both experiments, the others in only one of them. Experiment 1 was completed by 15 subjects. Experiment 2 involved a separate setup for each of three real-word scenes, therefore 13, 7, and 6 subjects completed Experiment 2 for each scene respectively. A single session took approximately 20 – 30 minutes for both experiments.

2.2. Stimuli

Experiment 1 employed 25 HDR images commonly used for testing TMOs (see Figure 1). There were 14 outdoor scenes, 9 indoor scenes, and 2 CG rendered scenes. We did not include images of people or animals in the test set, since these are rare for HDR images. Some of the images depicted daytime scenes, the others night or evening scenes. The images were displayed in their original resolution or scaled to the resolution of the HDR display if they were too large.

#1 (3.08) #2 (3.93) #3 (2.57) #4 (6.21) #5 (4.06)

#6 (4.51) #7 (2.85) #8 (4.88) #9 (3.50) #10 (5.36)

#11 (2.69) #12 (3.46) #13 (3.70) #14 (3.96) #15 (3.43)

#16 (2.99) #17 (3.03) #18 (4.68) #19 (3.54) #20 (2.60)

#21 (2.98) #22 (4.25) #23 (3.84) #24 (3.16) #25 (3.48)

Figure 1: *25 HDR images used in Experiment 1. They are numbered 1–25 and their dynamic ranges are shown in parentheses in decimal-logarithmic units. Their dynamic ranges are computed using the* pfsstat *command from pfstools. The Drago TMO [DMAC03] is applied for the convenience of view.*

Figure 2 shows the HDR images used for Experiment 2. These are the HDR photographs of our experimental scenes that we set up next to the HDR display. Each of the three HDR images was created using the multiple exposure technique from 15 low dynamic range images taken with a Kodak Professional DCS560 mounted on a tripod. We used Robertson's method [RBS99] implemented in the *PFScalibration* software to calibrate a camera and create the HDR images. We selected the lens (CANON EF 50mm) and the

position of the camera, so that the image displayed on the monitor closely matched the real scene. The subject's viewpoint was not restricted and the setup allowed them to have comfortable viewing of both the real scene and the display from the distance of about 1.5 times screen height from the HDR display. Images 26 and 27 (the left and the middle images in Figure 2) contain the same object layout but differ in the lighting condition. Both scenes were lit with the 800 Watt HMI lamp (JOKER-BUG 800), which gave approximately daylight illumination. For Image 27, the lamp was covered by the diffuser (Lightbank) to decrease the intensity of the light source. As shown in Figure 2, the absolute luminance values were very different with or without a diffuser for the HMI lamp. The table setup in Images 26 and 27 included a MacBeth Color Checker^{TM}, an 18% reflective gray card and several objects ranging in their reflectance from black to white. The experimental sessions for all images except Image 28 were conducted in the room whose lighting condition could be fully controlled and was set to a typical dark office illumination (64 lux). In the pilot study we verified that the level of ambient light does not have a significant influence on the results. For Image 28, the experiment was conducted with natural light and completed within two hours in the afternoon under stable weather conditions.

#26 (2.62) #27 (2.68) #28 (2.19)
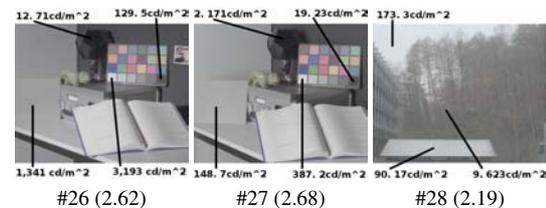
Figure 2: *Three HDR images and several measurements of luminance of the real scenes. Their dynamic ranges in* $\log_{10}$ *unit are also shown after each number of the images. These images were shown to the subjects with their corresponding real-world views in Experiment 2 and without reference as done in Experiment 1. The Drago TMO [DMAC03] is applied for the convenience of view.*

The images were shown on the Brightside 18″ LED-based HDR display [SHS*04] which consists of an LCD panel ($1,280 \times 1,024$ pixels) and a matrix of 760 separately controlled white LEDs, acting as a back light. The minimum and maximum luminance levels of the display we used for the experiments were 0.2 and 3,000 cd/m^2, which gave the maximum dynamic range of 4.18 log-10 units. The HDR monitor was calibrated by measuring its luminance response for a range of input values using the MINOLTA LS-100 luminance meter. Then, the measured values were used to create an inverse lookup table, which was used by the display driver. The display driver was implemented in graphics hardware as a fragment program to allow for real-time interaction with images.

Although tone mapping of images for the HDR display
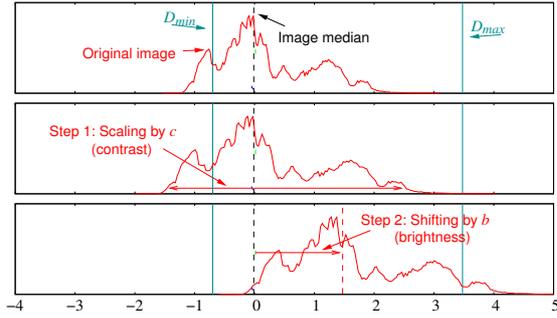
Figure 3: *Illustration how the generic tone mapping modifies the image histogram (see Equation 1).*

may seem like a futile exercise, we found several reasons for this approach. Firstly, we wanted to check if tone mapping is necessary for HDR displays, and if it is so, what kind of tone mapping. For this reason, we included several images whose dynamic range exceeded the dynamic range of the display (refer to Figure 1). Secondly, we in fact artificially limited minimum and maximum luminance of the display in the experiments as described later (although most of the data has been collected for the full dynamic range). We also did not want to use different displays of different color characteristic in the experiments.

2.3. Generic Tone Mapping

The purpose of our psychophysical experiments is to collect data from human observers to determine what are the desired or important properties of a TMO. This knowledge should help in the design of new TMOs or automatic parameter estimation for the existing TMOs. Since examining all possible TMOs is not feasible in an experimental setup, we consider only a global TMO that involves linear scaling and shifting of color values in the logarithmic domain. Even though this is probably the simplest TMO that is practically used, it can mimic the behavior of many global TMOs, such as [Tum99](Appendix A) or [FPSG96], and is in fact a part of any TMO that requires "gamma correction", such as [RSSF02] (since a power function that is used in gamma correction corresponds to linear scaling in the logarithmic domain). Many TMOs produce output pixel values in an arbitrary range, which must be linearly scaled or shifted to fit the dynamic range of a display (e.g. [FLW02,DD02,RD05]). For these and other TMOs such scaling (contrast adjustment) and shifting (brightness adjustment) operations are essential for the final appearance of a tone mapped image and are therefore analyzed in this work in more detail.

The generic TMO we use in the experiments is described with three parameters: brightness, contrast, and saturation of color. Brightness and contrast parameters are considered as an offset of luminance and as a difference between the

maximum and minimum luminance values, respectively. To adjust color saturation, color coordinates are interpolated or extrapolated between the original pixel color and its corresponding luminance value for the D65 white point. All adjustments are performed in the logarithmic domain to approximate non-linear response of the human visual system to light. Formally, the TMO can be modeled as

$$\log_{10} R' = c \cdot \log_{10} R + b, \tag{1}$$

$$\log_{10} Y' = 0.2126 \log_{10} R' + 0.7152 \log_{10} G' + 0.0722 \log_{10} B', \tag{2}$$

$$\log_{10} R'' = \log_{10} Y' + s(\log_{10} R' - \log_{10} Y') \tag{3}$$

where b, c, and s are brightness, contrast, and color saturation parameters respectively, Y' is the new luminance value, and R'' is the output red channel value. Equations (1) and (3) are applied for green and blue channels in the same way as for the red channel. Note that Y' is an approximation of luminance, which is used for our convenience (luminance should be a weighted sum of linear instead of logarithmic R, G and B coordinates). To assure that the adjustment of contrast has a minimum impact on the perceived brightness of a scene, the pixels of each HDR image were multiplied by a constant factor, so that the median luminance value of each image was $\bar{Y} = \log_{10}(1) = 0$. This way the multiplication by the contrast parameter in Equation 1 "stretched", but did not shift image histogram. This is illustrated in Figure 3.

2.4. Experimental Procedure

The two psychophysical experiments were conducted with and without reference scenes, respectively. For both experiments, each HDR image was shown on the HDR monitor one after another with a user interface that allowed the subjects to interactively adjust parameters of brightness, contrast, and color saturation using a mouse. The first two parameters were adjusted using a 2D slider interface and color saturation was adjusted using a 1D slider as shown in Figure 4. Since we found in the pilot study that brightness and contrast are difficult to control separately, we decided to use a 2D slider that would allow adjustment of both parameters at the same time. In Experiment 1, the subjects were asked to adjust these parameters until the most preferred reproduction of each HDR image was achieved in their own opinion without reference images (preference). In Experiment 2, their task was to achieve the closest reproduction of the real-world view (fidelity). They were asked to reproduce the details of all objects in an HDR image as seen in the real scene and, if possible, to adjust the HDR image brightness to match the real scene.

The parameters of brightness, contrast, and color saturation of the generic TMO (refer to Section 2.3) were allowed to be adjusted within the range of -3.0 – 5.0, 0.1 – 4.0, and 0.1 – 4.0 respectively. Before starting the experiments, the ranges were checked to be large enough to reproduce ev-
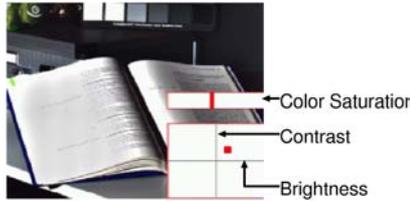
Figure 4: *A screenshot of the user interface used in our experiments. The horizontal and vertical axes of the 2D slider (bottom right) adjust brightness and contrast, respectively. The 1D slider above changes color saturation.*

erything from very dark/low contrast images to extremely bright/high contrast images with color settings ranging from grayscale to color-saturated image.

While Experiment 1 was conducted using the full dynamic range of the HDR monitor ($0.2 - 3,000$ cd/m^2), in Experiment 2 we restricted the minimum and the maximum luminance of the HDR display to simulate a range of potential display devices as listed in Table 1. Additionally, for each simulated display, the subjects were given a questionnaire sheet to mark the score of its reproduction, which could be "good" (3), "average" (2), or "poor" (1). The subjects were naïve as to what technically differs in each of the 14 dynamic range and brightness settings.

#	D_{min}	-	D_{max}	#	D_{min}	-	D_{max}
1	0.2	-	3,000	8	1.0	-	3,000
2	0.2	-	80	9	80.0	-	1,000
3	0.2	-	200	10	80.0	-	3,000
4	0.2	-	1,000	11	200.0	-	1,000
5	1.0	-	80	12	200.0	-	3,000
6	1.0	-	200	13	1,000.0	-	3,000
7	1.0	-	1,000	14	0.2	-	3,000

Table 1: *The range of the minimum and maximum luminance values of the HDR display that simulates potential displays (given in cd/m^2). Note that the dynamic range between 0.2 and 3,000 cd/m^2 was used twice in the test to validate consistency of the results.*

Finally, another experiment was conducted with Images 26 – 28 and four subjects using the same procedure as in Experiment 2 but the task was to adjust parameters to their preference (as in Experiment 1) and no reference scene was given. The details and the supplemental materials on the experiment are available at [web].

3. Results and Discussion

The results for both experiments are summarized in Figure 5. The plot shows a large variance in the preferred TMO parameters, which indicates that the subjects used a broad

range of possible parameters. There is also a strong correlation between brightness and contrast parameters. The contrast parameter is biased toward an enhanced contrast ($c > 1$ in Equation (1) indicates that the contrast was higher in a tone mapped image than in an original image).

Preliminary screening indicates that the results for Subject 22 are significantly different than for the other participants (probably due to improper use of the user interface) and therefore this data is removed from the further analysis (marked as blue '×'es in Figure 5).

We ran the multivariate analysis of variance (MANOVA) to test *main effects* of subjects' gender and expertise on measured parameters. The F distribution and a probability value p, which is derived from F, are used to determine whether there is a statistically significant difference between populations of samples. The higher p value, the more we can believe that the populations of samples are not statistically different. In our experiment, the gender difference is not significant ($F(3, 496) = 1.187$, $p \gg 0.05$ and $F(3, 360) = 1.970$, $p \gg 0.05$ for Experiments 1 and 2, respectively) as $p > 0.05$ shows that the difference between populations of samples (male and female in this case) is not statistically significant. Two people were aware of the experiment purpose, and therefore they were considered as experts. The population means of experts and non-experts are however not significantly different ($F(3, 496) = 0.3237$, $p \gg 0.05$ and $F(3, 360) = 2.2304$, $p \gg 0.05$ for Experiments 1 and 2, respectively). Therefore, we analyze all collected data together in the following sections.

To better understand the source of large parameter variations, we plot brightness and contrast parameter settings separately for several selected subjects and images in Figure 6. Similar plots for all subjects and images can be found at [web]. The left pane of Figure 6 shows that the settings can significantly differ between subjects, ranging from the preference for high contrast and low brightness (Subject 6) to the opposite preference for low contrast but bright images (Subject 2). The significant difference of subjects' settings is statistically established by MANOVA ($F(52, 1872.8) = 10.7864$, $p < 0.05$ and $F(64, 1349.0) = 7.6678$, $p < 0.05$ for Experiments 1 and 2). We can expect that two different individuals have different notions of a perfect image, therefore the TMO settings must be affected by the subject's tastes. This is an important observation with several consequences. Firstly, a TMO designed to render the best looking images should account for the user's tastes, for example by offering user adjustable parameters. Secondly, when ranking or assessing performance of TMOs in subjective experiments (e.g., [LCTS05]), the subjective influence should be taken as a factor in the analysis since two different subjects are likely to propose two different TMO rankings if they differ in their tastes. Finally, we cannot average parameter settings across all subjects for further analysis, since those parameters significantly differ across individuals.
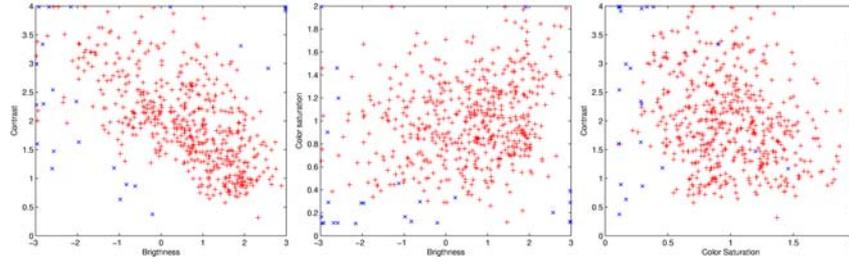
Figure 5: *The results for all images, all subjects and for the full dynamic range of the display (red '+'). One subject regarded as an outlier is marked with blue '×'es.*

More consistency in the parameter settings can be observed across images. The center and right panes of Figure 6 show that both parameters follow a similar line of decreasing contrast and increasing brightness. While the images follow the similar pattern of parameter settings, the population means of the parameters are significantly different ($F(72,1414.4) = 7.6420$, $p < 0.05$ and $F(6,718) = 17.1307$, $p < 0.05$ for Experiments 1 and 2). This indicates that the TMO settings are affected by image characteristics.

3.1. Contrast and Brightness Preference

To understand what the subjects' motivation for the choice of contrast and brightness parameters was, we plot the histograms of the resulting images in Figure 7 (the histograms for more subjects can be found at [web]). Although each subject adjusted the same image in different ways, he or she also followed a similar scheme when choosing TMO parameters for all the images. For example, the histograms resulting from the adjustments of Subject 6 are always more spread out than for the other subjects. This would indicate that the magnitude of contrast enhancement is correlated with the tastes of an individual. Additionally, there is another interesting observation which seems to be consistent across both all images and all subjects: the display maximum luminance, above which pixels are clipped, falls into approximately the same part of the histogram (see the blue vertical lines on the right of each plot in Figure 7). This indicates that people tend to "anchor" the brighter part of an image to the display maximum luminance, and then they extend or compress contrast in the direction of lower luminance to get the best looking image.

It is interesting to see whether the same observation can be generalized to a broad range of displays or if it is applicable only to an HDR display. We plot histograms in Figure 8 for a single subject and single tone mapped image but for several simulated displays of different brightness and dynamic ranges. The figure clearly indicates that subjects adjust images for the capabilities of a display, but they also follow the same scheme as for the HDR display (0.2–$3,000\ cd/m^2$) — they map approximately the same part of the histogram

to the maximum luminance of the display and then adjust contrast.

3.2. Improved Tone Mapping Algorithm

The motivation for remodelling a TMO is to provide new parameters that would be more intuitive to use. As mentioned earlier, the settings for contrast and brightness are strongly correlated. An average correlation coefficient for all images and both experiments is $\bar{R} = -0.7217 \pm 0.1622$. This suggests that both contrast and brightness could be replaced with parameters that do not exhibit such strong correlation and are therefore easier to control by the users. In the case of contrast and brightness, the 2D slider is usually adjusted along a slanted line (refer to Figure 6 center and right), which is neither intuitive nor convenient. A better user interface would use decorrelated parameters, so that the subjects could either use a simpler 1D sliders or move the 2D slider along the axes instead of a slanted line.

In Section 3.1, we analyzed and identified the strategy that the subjects use for adjusting TMO settings. Now, we show that this strategy can be modelled. We rewrite Equation (1) as

$$\log_{10} R' = c \cdot \log_{10}(R/Y_{max}) + \log_{10}(D_{max}) \qquad (4)$$

where Y_{max} is the maximum luminance value in an image that we want to reproduce on a display, which we call "anchor white". The same formula is used for the blue and green channels. The above equation mimics the operation performed by the subjects in our experiments. Firstly, the formula extends or compresses the image histogram by the scale factor c to the left side of the anchor white Y_{max}. Then, the anchor white is shifted to the display's maximum luminance D_{max}. Note that we use the same contrast parameter c as in Equation (1), but we replace the brightness parameter b with the anchor white Y_{max}.

To better understand how Equation (4) relates to the original contrast and brightness parameters, we plot a function of c assuming constant Y_{max}. Firstly, we find the relation between b and D_{max} from Equations (1) and (4) as

$$b = \log_{10}(D_{max}) - c \cdot \log_{10}(Y_{max}). \qquad (5)$$
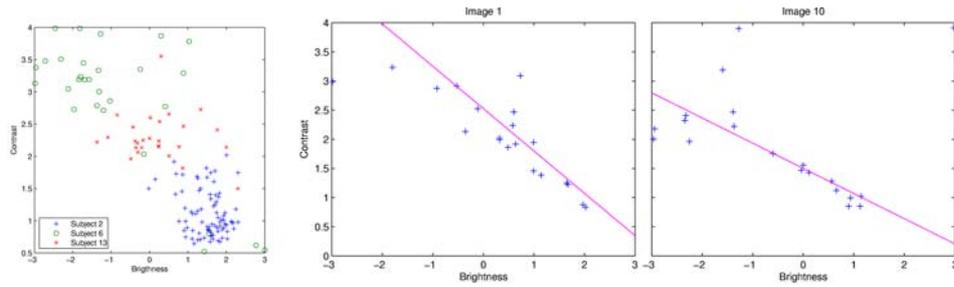
Figure 6: *The contrast–brightness relation for three selected subjects (left) and images (center – Image 1; right – Image 13). Both contrast and brightness settings differ significantly from subject to subject and from image to image.*
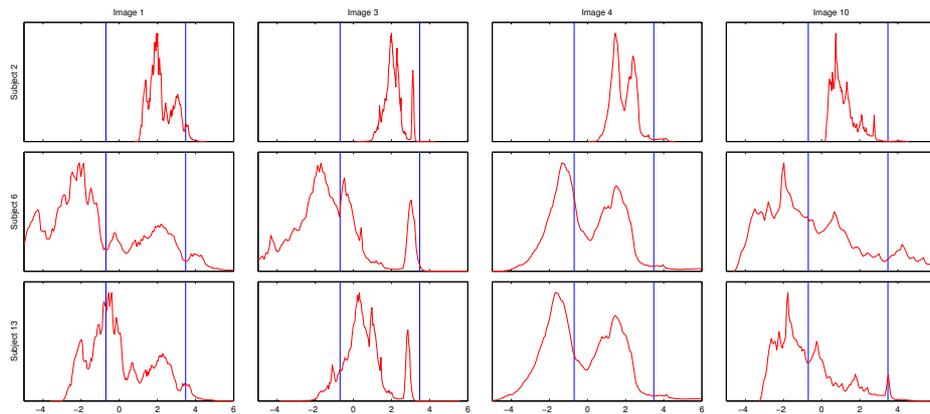
Figure 7: *The histograms of Images after the subjects' brightness and contrast adjustments – data from Experiment 1 (the full dynamic range of a display: 0.2–3,000 cd/m^2). The blue vertical lines denote display minimum and maximum luminance. The horizontal axis is scaled in log luminance units. The subjects differ in their adjustments between each other, but each subject follows his or her tastes consistently for all images.*

Secondly, we choose two images (Images 1 and 10 in this example) and find the median percentage of the clipped pixels $\bar{C}$ (see the third row of Figure 9) in order to compute Y_{max}:

$$Y_{max} = percentile(Y, 100 - \bar{C}) \qquad (6)$$

where Y is a set of luminance (or luminance factor) values of the pixels in an image. Note that the above formula gives the location on the histogram for a given percentage of clipped pixels $\bar{C}$. We use the computed Y_{max}, the maximum luminance of the display $D_{max} = 3,000 \ cd/m^2$ and Equation (5) to plot the function of c as a continuous magenta line in Figure 6 (center and right). The important observation is that the plotted functions for both images approximate well the correlation between contrast and brightness parameters. This indicates that the largest variations between subjects in the resulting images are due to different selections of contrast parameter c while the anchor white Y_{max} does not vary much between subjects.

We intentionally named the parameter Y_{max} as "anchor white" to refer to the lightness perception theory [GKB*99].

According to this theory, the human visual system assesses the lightness of an object based on the anchor luminance value, which acts as a reference for a white reflective surface. Such anchor luminance does not need to be the highest luminance in an image. This is especially true for the scenes that contain self-luminous surfaces, such as lights or the sun. The theory postulates that a "common denominator" for lightness estimation is a white reflectance, instead of gray, often used in photography. Our experiment confirms this since "anchoring" reflectance white to the maximum luminance of a display was a dominant strategy for adjusting the TMO settings.

The linear TMO we obtain in Equation (4) is easier to control than our original one, since both parameters of the contrast c and the clipping level Y_{max} modify independent aspects of image appearance. Moreover, if we transform Equation (4) from the logarithmic to the linear domain, we have the following formula:

$$R' = D_{max} \cdot (R/Y_{max})^c. \qquad (7)$$
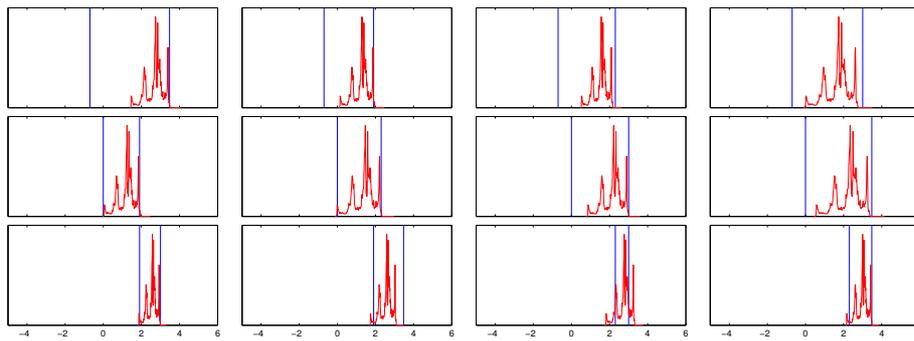
Figure 8: *The histograms of Image 1 after Subject 1 brightness and contrast adjustments – data from Experiment 2. The notation is the same as in Figure 7.*

This re-parameterized form of the original TMO formula from Equation (1) is similar to a global contrast adjustment operation [†], employed as a final-cut in many TMOs and as enhancement operation in image editing software. The importance of Equation (7) comes from the fact that we derived this formula based only on the analysis of the data we collected in our experiment without any prior assumptions on the parameters of the tone reproduction function. We have shown that the users try to adjust the TMO parameters along the parameter c, even if they have a non-standard user interface as used in our experiments. Moreover, we have shown that the same formula is valid for a broad range of display devices, ranging from dark CRT monitors to HDR displays.

3.3. Image and Subject Influence on TMO Parameters

We analyze how contrast, color saturation and the percentage of clipped pixels in dark and bright regions (dependent variables – DV) differ between subjects and images (independent variables – IV). We want to find out which of the two DVs is responsible for the large variance in the IV.
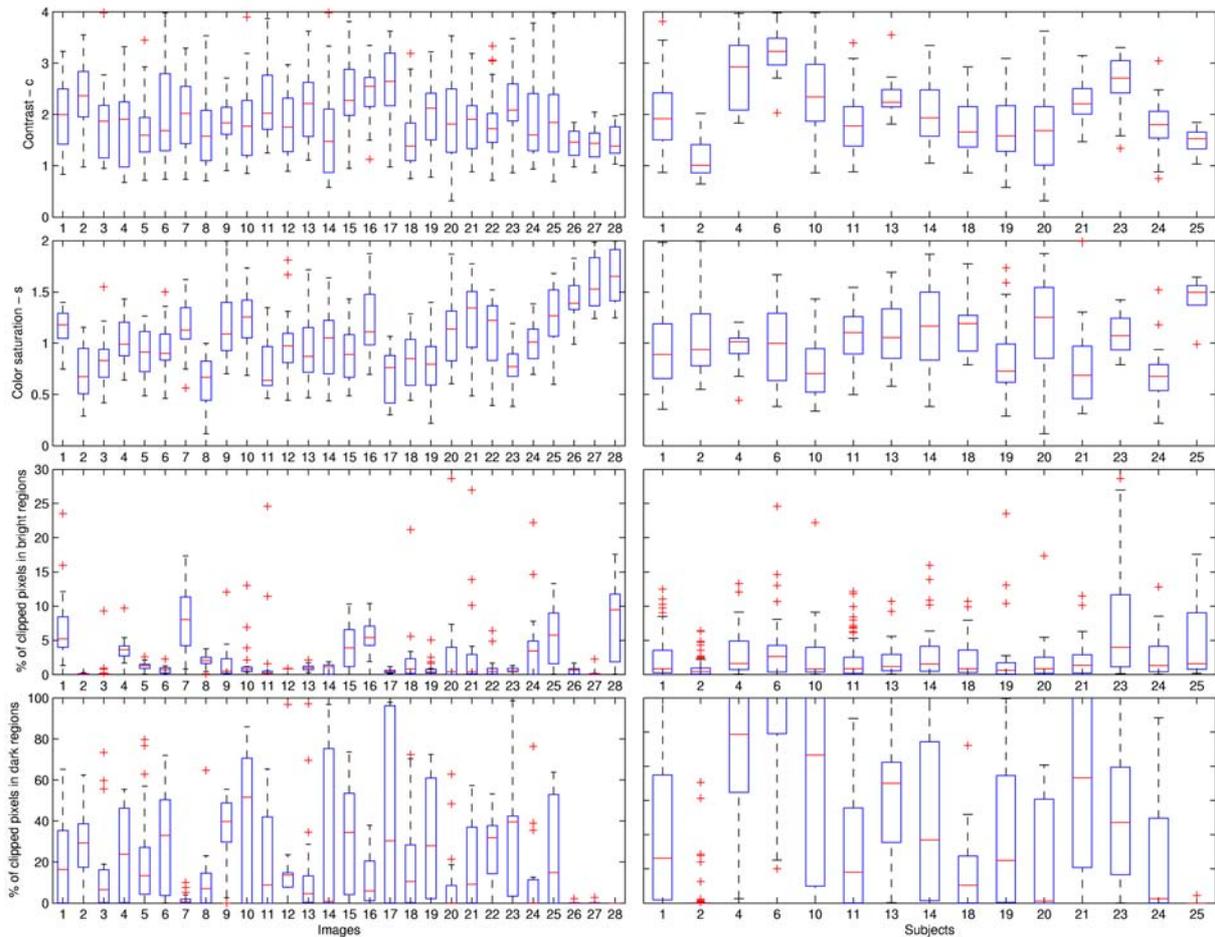
The variations of the DVs with respect to the IVs are summarized in Figure 9. Note that we do not include the brightness parameter in this figure. This is because brightness is strongly correlated with contrast and it is also fully determined by contrast and anchor white as described in Section 3.2. Anchor white, on the other hand, is related to the percentage of pixels clipped in bright regions (refer to Equation (6)).

From a first look at Figure 9 we can conclude that all four DVs are significantly different between images and subjects. This is confirmed by the two-way analysis of variance (ANOVA) for the main effects of the subjects and the images, which are run separately for each DV ($7.07 < F < 74.21$, $p < 0.001$). Contrast varies more between subjects than images (see the first row if Figure 9) and is probably determined mostly by subjects' personal tastes as discussed in Section 3.1. Color saturation and the percentage of clipped pixels in the dark regions (rows 2 and 4 in Figure 9) do not show any consistency between the subjects and the images and therefore it is not possible to draw any conclusion for these parameters. The third and the fourth rows of Figure 9 (note the difference in the scale used for these plots) show that there are significantly more pixels saturated in dark regions than in bright regions. This suggests that the subjects prefer sacrificing a significant portion of the dark part of an HDR image, probably in order to improve contrast. The same tendency can be observed in Figure 7, which shows that Subjects 6 and 13 decided to push a large part of the histogram below the minimum luminance of the display, while preserving the brightest pixels. This observation suggests that TMOs should follow a similar pattern and saturate more pixels in the dark regions. This is contrary to the most common approach employed in many TMOs where the same number of the darkest and brightest pixels are clipped. Such TMOs do not produce the best results if they do not provide an adjustment for the number of pixels clipped in dark regions.

Conclusions on the measured values of clipped pixels in bright regions can be drawn directly from the actual images. We observed that the most pixels are clipped for the images that contain large bright objects which should appear self-luminous in the reproduction, like the sky in Images 1, 7, 16, 25 and 28, or the sun in Image 24 (refer to Figure 1 for images and Figure 9 for the magnitude of clipping). Then, follow the images that contain small self-luminous objects, such as Christmas lights in Image 15 and the images that depict dark scenes without self-luminous objects (Images 12 and 27). There is also less clipping for the images of low dynamic range (Images 3, 11 and 20).


[†] The operation from Equation (7) is sometimes confusingly called *gamma-correction*. However, since the original meaning of *gamma-correction* denotes compensating the non-linary of CRT monitors, using this term in the context of image enhancement may not be appropriate.

Figure 9: *The variation of the DVs (contrast, color saturation and the percentage of clipped pixels in bright and dark regions) with respect to the IVs (images and subjects) – data from the experiments in which full dynamic range of the HDR display was used. Only 15 subjects participated in these experiments. Images are numbered as in Figure 1 and Figure 2. The Notation: red lines – median; blue boxes – spanning from 25th to 75th percentile; whiskers – minimum and maximum values without outliers; red crosses – outliers.*

3.4. Choosing Default TMO Parameters

User adjusted TMO parameters are not desirable in many applications and it would be helpful if their values could be automatically found at least to render a "best guess" image. We want to check if there is any correlation between the DVs (the TMO parameters) and the IVs, so that, for example, an image characteristic can predict the values of contrast and the percentage of clipped pixels in the bright regions (needed to compute anchor white Y_{max}). If we find such a correlation, we can propose a method to automatically choose TMO parameters.

Although an algorithm cannot predict a user's tastes, it may be possible to guess some TMO parameters based on the characteristics of an image. To verify this hypothesis, we compute a set of variables characterizing each HDR image; the dynamic range of an image, which is a difference between the logarithm of the highest and the lowest luminance in an image; the key value, α; and L_{white}. α and L_{white} are used for the automatic parameter estimation in the photographic TMO [Rei02]. All these variables require the value of the maximum and the minimum luminance in an image, which can be calculated in a variety of ways. We compute the minimum and the maximum as percentiles: 0.01, 0.1, 1, 10, 20, 30 (of brightest and darkest pixels), and as the minimum or maximum value of a low-pass filtered image, where the filter is the Gaussian with different values of σ (1, 2, 5, 10, and 20). This gives in total 30 different variables that could characterize an image (3 variables times 10 estimates of the minimum and the maximum luminance).
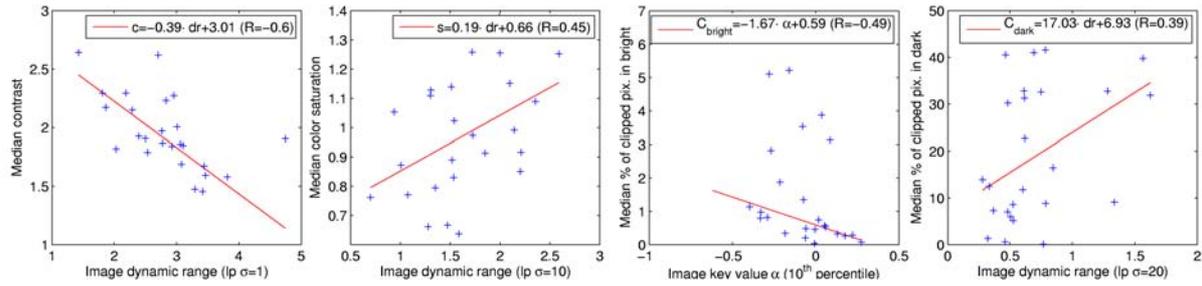
Figure 10: *The prediction of tone mapping parameters (contrast, color saturation, the percent of clipped pixels in bright and dark regions) based on image statistics. Plot legends: robust fits of the linear model and correlation coefficients.*

We compute the correlation matrix to check if there is a correlation between any of the computed 30 variables and the median values of the TMO parameters for the subjects: contrast, color saturation and the percentage of clipped pixels. We use this matrix to find the variable that is the most correlating with each TMO parameter. The relations between the most correlated variables and the TMO parameters, together with the results of the robust linear regression, are plotted in Figure 10. The highest correlation is found for the dynamic range of an image computed using low-pass filtered images ($\sigma = 1$) and the contrast parameter (the first plot in Figure 10). The negative slope of this relationship is intuitive — the images of higher dynamic range must be stronger compressed to fit into the dynamic range of a display. A weaker correlation and less intuitive relation is found for color saturation and the percentage of clipped pixels in the dark regions. These TMO parameters probably cannot be predicted using the given set of image characteristic variables. The prediction of the percentage of pixels clipped in bright regions is more reliable. It correlates with the image key coefficient α (computed using the 10-th percentile for the minimum and the maximum values in an HDR image). We observed that this prediction is less accurate for the images that contain large self-luminous objects

The plots in Figure 10 show that both the contrast parameter and the number of clipped pixels in the bright regions are correlated with image content, and therefore they can be predicted. Such predictions can be used for parameter estimation in TMOs. Although the predicted values will not be optimal for many images and subjects, they could be used as the "best guess" for the TMO parameter setting. Our experiments did not include a sufficient number of images and subjects to build a reliable model for such a parameter estimation, but they proved that such estimation is possible and can be an interesting direction for further research.

3.5. Influence of a Display

It is interesting to know how the dynamic range and brightness of a display influences the parameters of a TMO. Fig-

ure 11 illustrates how the contrast setting increases as the dynamic range of a display increases. If the dynamic range of a display is too low, the subjects compress contrast. On the other hand, they expand contrast even above the contrast of an original image ($c > 1$) when a display offers higher dynamic range. However, this behavior differs slightly between both experiments: if the subjects adjust the HDR images to their preference, they enhance contrast proportionally to the dynamic range of a display (Figure 11 top), but if their goal is to achieve the fidelity to the real-world scene, they adjust contrast slightly above 1.0 and keep it approximately on the same level even for the HDR displays (Figure 11 bottom). This suggests that the TMO profiled for fidelity should not enhance contrast above the contrast of an original scene, and the TMO profiled for preference should take full advantage of the display dynamic range.
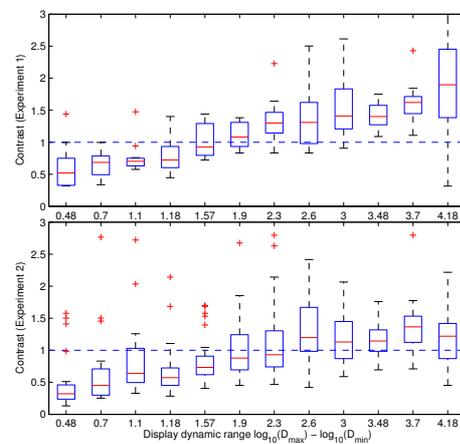
Figure 11: *The relation between the dynamic range of a simulated display and the contrast parameters. The relation is plotted for the data from Experiment 1 (preference) at the top and Experiment 2 (fidelity) at the bottom. The subjects tend to enhance contrast more if their goal is the most preferred image. Notation is the same as in Figure 9.*

3.6. Preferred Display Device

We examine how the minimum and maximum luminance values of a display can affect subjective preferences for displayed images. We use the data from the questionnaire used in Experiment 2. The preference scores for each simulated display are averaged over three scenes and over all subjects. Figure 12 illustrates the ranking of potential display devices simulated on the HDR display. The figure clearly shows that the subjects prefer brighter displays of higher dynamic range. A typical LCD display ($1 - 200\ cd/m^2$) is in the middle of the preference scale. Interestingly, the brighter display but of lower dynamic range ($80 - 3,000\ cd/m^2$) has higher preference score than the typical LCD. The displays of the broadest dynamic range top the ranking, but the broadest dynamic range display ($0.1 - 3,000\ cd/m^2$) comes unexpectedly lower than the $1 - 3,000\ cd/m^2$ model. However, the rankings in the top group ($1 - 3,000$, $0.2 - 1,000$, $1 - 1,000$, and $0.2 - 3,000\ cd/m^2$ models) are not significantly different from each other ($F(3,126) = 0.82$, $p > 0.05$). The high scores for the brightest displays of the highest dynamic range indicate that both high luminance and high contrast are important for reproducing digital images.
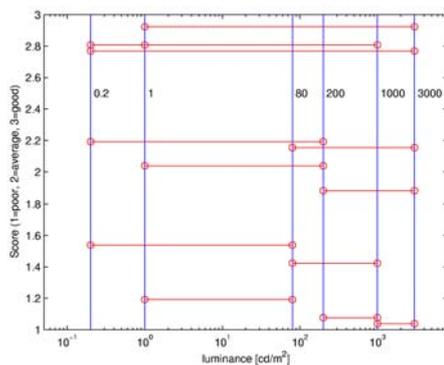
Figure 12: *The preference of several simulated monitors, which differ in the minimum and the maximum luminance they can display. The preference score is indicated on the y-axis, and the left and right end points of the horizontal lines represent the minimum and maximum luminance of the display. The scores of the displays are shifted by a small random offset to avoid overlapping of the lines.*

To better understand the relation between the minimum and the maximum luminance of a display and the preference score, we fit the data to the linear model using multiple linear regression. The averaged preference score S is given by

$$S = \alpha \cdot log_{10}(D_{min}) + \beta \cdot log_{10}(D_{max}) - \gamma \qquad (8)$$

where D_{min} and D_{max} are the display minimum and maximum luminance values, $\alpha = -0.47$ (± 0.05), $\beta = 0.87$ (± 0.11) and $\gamma = 0.25$ (± 0.31). The model accounts for nearly 60% of the data ($R^2 = 0.57$). The negative α indicated

that "darker" displays are more preferred (i.e., lower minimum luminance) and positive β indicates that also "brighter" displays (i.e., higher maximum luminance) are preferred. However, since the trend is stronger for β, we can assume that the maximum luminance is more important than the minimum luminance of a display. The percentages of oversaturated pixels are far smaller than that of undersaturated pixels (see the third and fourth rows in Figure 9). This indicates that people are more sensitive for oversaturation than undersaturation since they carefully avoided oversaturated pixels but did not pay much attention to undersaturated pixels compared to oversaturation.

4. Conclusions

The major outcome of this work is a better understanding of how users adjust tone mapping operator (TMO) parameters to achieve either the best looking images (preference task) or the images that are the closest to real-world scenes (fidelity task). Based on this knowledge, we propose a better parameterization of a linear TMO in logarithmic domain, in which parameters are more intuitive and can be partly estimated from image characteristics. The TMO is controlled by two parameters: *anchor white* and *contrast*. The *anchor white* parameter is approximately consistent across subjects and depends on images — it is set to a lower value if an image contains large self-luminous objects. The *contrast* parameter is more subjective, and therefore users should be allowed to adjust it. We have shown that the parameters can be automatically estimated for a TMO based on an image characteristic to obtain a "best guess" result. The *contrast* parameter can be predicted from the dynamic range of an image (images of higher dynamic range must be reproduced with lower contrast), and the *anchor white* parameter is related to the image key value (although the prediction can be unreliable if an image contains large self-luminous objects). We believe that the results of our analysis are also applicable to complex TMO, which can benefit from both a better selection of user adjusted parameters and an automatic parameter estimation.

The second main subject of this work is an investigation how the dynamic range and brightness of a display affects the preference for tone reproduction. For 14 simulated monitors of varying brightness and dynamic range we do not find any major difference in the strategy the subjects use to adjust images for LDR and HDR displays. We notice however that the resulting images depend on a given task. If the goal is to find the best looking image (preference), subjects tend to strongly enhance contrast (up to four times that of the original image contrast), even at the cost of clipping a large portion of the darkest pixels. On the other hand, when the task is to achieve the best fidelity with a real-world scene, the subjects avoid clipping both in the dark and bright parts of an image and they do not extend contrast much above the contrast of an original image. In both tasks, there is a ten-

dency towards brighter images, which are achieved by over-saturating the brightest pixels belonging to self-luminous objects. The final investigation compares user's preference for displays of varying capabilities. The subjects prefer in the first order the displays that are bright, and in the second order, the displays that have low minimum luminance.

5. Acknowledgments

References

[DD02] DURAND F., DORSEY J.: Fast bilateral filtering for the display of high-dynamic-range images. *ACM Transactions on on Graphics 21*, 3 (2002), 257–266. 4

[deR96] DERIDDER H.: Naturalness and image quality: saturation and lightness variation in color images of natural scenes. *The Journal of Imaging Science and Technology 40*, 6 (1996), 487–493. 2

[DMAC03] DRAGO F., MYSZKOWSKI K., ANNEN T., CHIBA N.: Adaptive logarithmic mapping for displaying high contrast scenes. In *Proceedings of Eurographics* (2003), Brunet P., Fellner D., (Eds.), pp. 419–426. 3

[DMMS02] DRAGO F., MARTENS W., MYSZKOWSKI K., SEIDEL H.-P.: *Perceptual Evaluation of Tone Mapping Operators with Regard to Similarity and Preference*. Technical Report MPI-I-2002-4-002, Max-Planck-Institut fuer Informatik, 2002. 1

[DZB05] DELAHUNT P., ZHANG X., BRAINARD D.: Perceptual image quality: Effects of tone characteristics. *Journal of Electronic Imaging 14*, 2 (2005), 1–12. 2

[FdB97] FEDOROVSKAYA E., DERIDDER H., BLOMMAERT F.: Chroma variations and perceived quality of color images of natural scenes. *Color Research and Application 22*, 2 (1997), 96–110. 2

[FLW02] FATTAL R., LISCHINSKI D., WERMAN M.: Gradient domain high dynamic range compression. *ACM Transactions on Graphics 21*, 3 (2002), 249–256. 4

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

[GKB*99] GILCHRIST A., KOSSYFIDIS C., BONATO F., AGOSTINI T., CATALIOTTI J., LI X., SPEHAR B., ANNAN V., ECONOMOU E.: An anchoring theory of lightness perception. *Psychological Review 106*, 4 (1999), 795–834. 7

[Joh05] JOHNSON G. M.: Cares and concerns of CIE TC8-08: spatial appearance modeling and HDR rendering. In *Image Quality and System Performance II* (2005), SPIE, volume 5668, pp. 148–156. 2

[JRW02] JOBSON D., RAHMAN Z., WOODELL G.: The statistics of visual representation. In *Visual Information Processing XI* (2002), SPIE, volume 4736, pp. 25–35. 2

[KYJF04] KUANG J., YAMAGISHI H., JOHNSON G., FAIRCHILD M.: Testing HDR rendering algorithms. In *IS&T/SID Color Imaging Conference* (2004), pp. 315–320. 1

[LCTS05] LEDDA P., CHALMERS A., TROSCIANKO T., SEETZEN H.: Evaluation of tone mapping operators using a high dynamic range display. *ACM Transactions on Graphics 24*, 3 (2005), 640–648. 1, 5

[RBS99] ROBERTSON M. A., BORMAN S., STEVENSON R. L.: Dynamic range improvement through multiple exposures. *IEEE International Conference on Image Processing* (1999), 159 – 163. 3

[RD05] REINHARD E., DEVLIN K.: Dynamic range reduction inspired by photoreceptor physiology. *IEEE Transactions on Visualization and Computer Graphics 11*, 1 (2005), 13–24. 2, 4

[Rei02] REINHARD E.: Parameter estimation for photographic tone reproduction. *Journal of Graphics Tools 7*, 1 (2002), 45–52. 2, 9

[RSSF02] REINHARD E., STARK M., SHIRLEY P., FERWERDA J.: Photographic tone reproduction for digital images. *ACM Transactions on Graphics 21*, 3 (2002), 267–276. 4

[RWPD05] REINHARD E., WARD G., PATTANAIK S., DEBEVEC P.: *High Dynamic Range Imaging, First Edition : Acquisition, Display, and Image-Based Lighting*, 1st ed. Morgan Kaufmann, 2005. 1

[SEL00] SAVAKIS A., ETZ S., LOUI A.: Evaluation of image appeal in consumer photography. In *Human Vision and Electronic Imaging V* (2000), SPIE, volume 3959, pp. 111–120. 2

[SHS*04] SEETZEN H., HEIDRICH W., STUERZLINGER W., WARD G., WHITEHEAD L., TRENTACOSTE M., GHOSH A., VOROZCOVS A.: High dynamic range display systems. In *Proc. ACM SIGGRAPH 2004* (2004), pp. 760–768. 3

[Tum99] TUMBLIN J.: *Three Methods of Detail-Preserving Contrast Reduction for Displayed Images*. PhD thesis, Georgia Institute of Technology, 1999. 4

[web] http://www.mpi-inf.mpg.de/resources/hdr/subtmo/. 5, 6

[YBMS05] YOSHIDA A., BLANZ V., MYSZKOWSKI K., SEIDEL H.-P.: Perceptual evaluation of tone mapping operators with real-world scenes. In *Human Vision and Electronic Imaging X* (2005), SPIE, volume 5666, pp. 192–203. 1