

Testing tone mapping operators with human-perceived reality

Akiko Yoshida

Volker Blanz

Karol Myszkowski

Hans-Peter Seidel

Max-Planck-Institut für Informatik
Saarbrücken, Germany

Abstract. A number of successful tone mapping operators for contrast compression have been proposed due to the need to visualize high dynamic range (HDR) images on low dynamic range (LDR) devices. They were inspired by fields as diverse as image processing, photographic practice, and modeling of the human visual systems (HVS). The variety of approaches calls for a systematic perceptual evaluation of their performance. We conduct a psychophysical experiment based on a direct comparison between the appearance of real-world scenes and HDR images of these scenes displayed on an LDR monitor. In our experiment, HDR images are tone mapped by seven existing tone mapping operators. The primary interest of this psychophysical experiment is to assess the differences in how tone mapped images are perceived by human observers and to find out which attributes of image appearance account for these differences when tone mapped images are compared directly with their corresponding real-world scenes rather than with each other. The human subjects rate image naturalness, overall contrast, overall brightness, and detail reproduction in dark and bright image regions with respect to the corresponding real-world scene. The results indicate substantial differences in the perception of images produced by individual tone mapping operators. We observe a clear distinction between global and local operators—in favor of the latter—and we classify the tone mapping operators according to naturalness and appearance attributes. © 2007 SPIE and IS&T. [DOI: 10.1117/1.2711822]

1 Introduction

The need for high dynamic range (HDR) images, containing a broader dynamic range than most of today's display devices, has greatly increased. Therefore, how to produce visually compelling HDR images has been one of the most important discussions in computer graphics, and a number of techniques have been introduced. To represent HDR images on low dynamic range (LDR) devices, a number of successful tone mapping operators have been presented. They are useful not only for HDR photography but also for lighting simulation in realistic rendering and global illumination techniques, which provide real-world luminance ranges.

Because a variety of tone mapping operators has been proposed, only a systematic perceptual evaluation can reveal the strengths and weaknesses of the wide range of

approaches presented in recent years. We conducted a psychophysical experiment of a direct comparison between the appearance of real-world scenes and tone mapped images of these scenes. The primary interest of this experiment is to investigate the differences in perception of tone mapped images when they are directly compared with their corresponding real-world views and indirectly compared with each other. To our knowledge, this work is the first direct comparison of tone mapped images with corresponding real-world scenes.

In Section 2, we briefly describe the seven tone mapping operators selected for our experiment. Section 3 gives an outline of our perceptual evaluation, and Section 4 discusses the obtained results. Finally, Section 5 concludes the article.

2 Tone Mapping Operators

Several early research papers appeared on photographic tone reproduction.^{2,12} Bartleson and Breneman showed theoretical predictions of the optimum tone reproduction functions for prints and transparencies and proved it by experimental data. Jones and Condit researched the relationship of brightness and camera exposure with a number of exterior pictures and proposed prediction methods of a *meter constant* of an image. In addition, Miller *et al.* proposed an idea of tone mapping in the field of lighting engineering.¹⁸ Their method deals with brightness as a power function of luminance (Steven's law) and preserves the brightness of an image before and after reducing the dynamic range. In 1993, the idea of tone reproduction was first introduced by Tumblin and Rushmeier³⁵ into the computer graphics community. The main goal of tone reproduction is to adjust the dynamic range of an image to the range that can be displayed on physical devices when the luminance range of the images does not fit that of the physical devices. The dynamic range must be compressed for very bright images and expanded for very dark images. However, not only should a tone mapping operator adjust the dynamic range, but it should also preserve the image's details and contrast. Usually, if the overall contrast is simply preserved, the details of an image may be lost, and vice versa. How to preserve them at the same time is always a difficult problem to conquer for tone mapping operators. A number of tone mapping techniques have been presented, and most of them can be categorized into two groups: global and local opera-

Paper 05205RR received Nov. 24, 2005; revised manuscript received Oct. 16, 2006; accepted for publication Nov. 8, 2006; published online Mar. 15, 2007. This paper is a revision of a paper presented at the SPIE conference on Human Vision and Electronic Imaging X, San Jose, California, Jan. 2005. The paper presented there appears (unrefereed) in SPIE Proceedings Vol. 5666.

1017-9909/2007/16(1)/013004/14/\$25.00 © 2007 SPIE and IS&T.

tors. *Global operators* apply the same transformation to every pixel of an image,^{7,10,40,30,36} while *local* ones adapt their scales to different areas of an image.^{1,4,8,9,22,27} The existing tone mapping operators are summarized in a recent review by Devlin⁶ and in a book by Reinhard *et al.*²⁸

For our perceptual evaluation, seven commonly used tone mapping operators were selected. The global operators are represented by the methods developed by Pattanaik *et al.*,²³ Ward *et al.*,⁴⁰ and Drago *et al.*⁷ and by a simple log-linear scaling. The local operators are the fast bilateral filtering presented by Durand and Dorsey,⁸ the Ashikhmin operator,¹ and the photographic tone reproduction by Reinhard.²⁷ In the following sections, we briefly discuss each of those seven operators.

2.1 Global Operators

The simplest tone reproduction is a linear mapping, which scales the radiances to the range between 0 and 255 on a gamma-corrected LDR monitor whose input values between 0 and 255 produce linear intensities for all but the smallest input values. Ambient lighting in the display environment prevents total darkness. If the logarithm of the radiances is taken and linearly scaled to [0, 255], it is called a *logarithmic linear mapping*.²⁸ Without gamma correction, many commonly available displays (LCDs, CRTs) follow an approximately log-linear response.

The histogram adjustment tone mapping operator developed by Ward Larson *et al.*⁴⁰ built on earlier work.^{38,10} The algorithm features strong contrast compression for pixels belonging to sparsely populated regions in the image histogram, which helps to overcome the problem of dynamic range shortage. This method leads to a monotonic tone reconstruction curve, which is applied globally to all pixels in the image. The slope of the curve is constrained by considering the human contrast sensitivity to guarantee that the displayed image does not exhibit more contrast than what is perceived in the real scene.

Pattanaik *et al.* extended the perceptual models framework by Tumblin and Rushmeier³⁵ and presented a new time-dependent tone mapping operator, which is based on psychophysical experiments and a photoreceptor model for luminance values.²³ This algorithm deals with the changes of threshold visibility, color appearance, visual acuity, and sensitivity over time for average luminance. This algorithm briefly can be decomposed into two models: the visual adaptation model and the visual appearance model. The signals that simulate the adaptation measured in the retina are used to set the half-saturation. These pixel values are computed as the logarithmic average of all pixels. Dark and bright adaptation effects are taken into account to decide upon the luminance weight value for the current frame based on the previous frame history. In our experiment, we ignore those temporal responses of human adaptation. To reproduce visual appearance, it is assumed that a time-dependent adaptation value determines reference white and reference black colors. Then, the visual appearance model recalculates the visual appearance values with those reference points. By assembling those visual adaptation and appearance models, the scene appearance is reproduced with changes to visual adaptation depending on time. This

method is also useful to predict the visibility and appearance of scene features because it deals with reference white and black points.

Drago *et al.* introduced *adaptive logarithmic mapping*,⁷ which addresses the need for a fast algorithm suitable for interactive applications, which automatically produces realistic-looking images for a wide variation of scenes exhibiting a high dynamic range of luminance. This global tone mapping function is based on logarithmic compression of luminance. To preserve details while providing high contrast compression, we use a family of logarithmic functions ranging from \log_2 to \log_{10} with increasing compressive power. The \log_{10} is applied for the brightest image pixel, and for the remaining pixels the logarithm base is smoothly interpolated between the values of 2–10 as a function of their luminance. The Perlin bias power function²⁵ is used for interpolation between the logarithm bases to provide better steepness control of the resulting tone mapping curve.

2.2 Local Operators

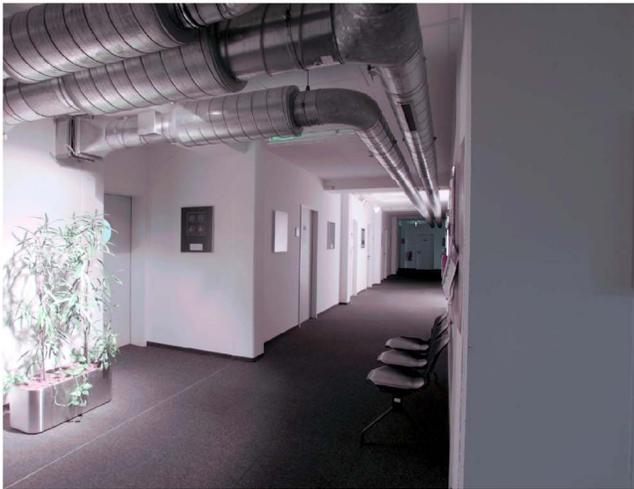
Reinhard *et al.* presented a photographic tone mapping operator inspired by photographic film development and the printing process.²⁷ The luminance of an image is initially mapped by using a global tone mapping function to compress the range of luminance into the displayable range. To enhance the quality of an image, a local adaptation is based on a photographic “dodging-and-burning” technique, which allows a different exposure for each part of the applied image. The most recent version of this method operates automatically, freeing the user from setting parameters.²⁶ To automate processes, low-contrast regions are found by a center-surround function at different scales. Then, a tone mapping function is locally applied. The automatic dodging-and-burning method enhances contrast and details in an image while preserving the overall luminance characteristics.

Ashikhmin presented a new tone mapping method, which works on a multipass approach.¹ The method takes into account two basic characteristics of the human visual systems (HVS): signaling absolute brightness and local contrast. First this method calculates the local adaptation luminance by calculating an average luminance of neighboring pixels fitting in a bound-limited contrast range (similar to Reinhard *et al.*²⁷). Then it applies the capacity function, which is based on the linearly approximated threshold vs. the intensity function, and calculates the final pixel values. The final calculation restores the details, which may be lost during compression. A tone mapped pixel value is obtained by multiplying a detail image, which is given by the ratio of pixel luminance to the corresponding local-world adaptation.

A fast bilateral filtering method was presented by Durand and Dorsey.⁸ This method considers two different spatial frequency layers: a base layer and a detail layer. The base layer preserves high-contrast edges and removes high-spatial frequency details of lower contrast. The detail layer is created as the difference of the original image and the base layer in logarithmic scale. After contrast compression in the base layer, both layers are summed up to create a tone mapped image.



(a)



(b)

Fig. 1 View of scenes selected for our tone mapping evaluation. Both images are tone mapped by the Drago TMO.⁷ (a) Scene 1. Maximum pixel luminance: 4,848.9506 cd/m²; minimum luminance: 0.0076 cd/m²; dynamic range: 638,019:1. (b) Scene 2. Maximum pixel luminance: 159.697 cd/m²; minimum luminance: 0.006 cd/m²; dynamic range: 26,616:1.

3 Perceptual Evaluation

Image comparison techniques can be roughly classified into two major categories: subjective and objective methods. Subjective methods obtain data from human observers, and the data are usually analyzed by statistical techniques^{20,21,41} while objective methods are based on theoretical models. Work with goals similar to ours about perceptual evaluation of tone reproductions has been recently published by Kuang *et al.*^{15,14} and Ledda *et al.*¹⁶

Kuang *et al.* conducted two paired comparison experiments of eight tone mapping operators (TMOs) in 2004.¹⁵ Tone mapped images were displayed on an LDR monitor without reference for observing overall rendering performance and grayscale tone mapping performance, respectively.

In 2005, Kuang *et al.* conducted a similar experiment with six TMOs (two poorly scored ones in the 2004 work¹⁵ were eliminated).¹⁴ They asked subjects to examine overall preference and several image attributes (highlight details, shadow details, overall contrast, sharpness, colorfulness,

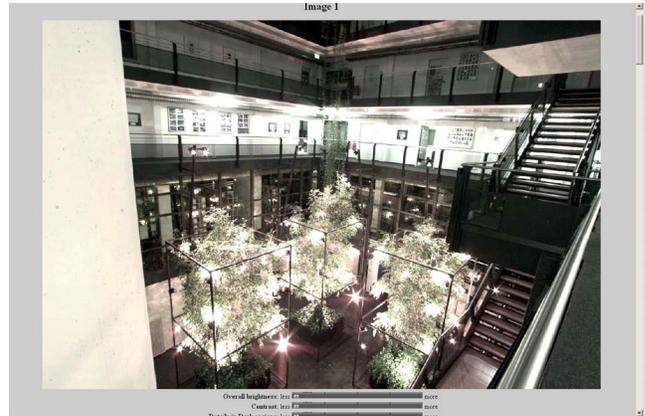


Fig. 2 A screenshot of HTML pages presented to the participants.

and artifacts) over tone mapped color images displayed on an LCD monitor. As in the previous work in 2004, they did not provide references for subjects. For both works in 2004 and 2005, they used a pairwise comparison.

In 2005, Ledda *et al.* also conducted psychophysical experiments of six tone mapping operators. They asked subjects to make paired comparison of tone mapped images displayed on an LDR monitor with an HDR image as their reference¹⁶ on a Brightside HDR display.³¹ In their experiment, the subjects actually never saw the real-world scenes measured by an HDR display. They measured overall similarity and detail reproduction of tone mapped images. First, they studied overall similarity of TMOs and clarified that their selected TMOs can be ranked. They asked subjects to compare a pair of tone mapped images and choose the one that appeared closer to the reference displayed on an HDR monitor. Following the first experiment, detailed reproductions of TMOs in both bright and dark regions of an image were studied in the second experiment. It showed a different result from that of Kuang *et al.*¹⁵ for the fast bilateral filtering.⁸ As Ledda *et al.* wrote, the fast bilateral filtering generates images with higher contrast and more detail visibility than in the reference images. Therefore, it had poor scores in the experiment with reference on an HDR monitor¹⁶ while it performed quite well without references.¹⁵ This observation is also confirmed by Yoshida *et al.*⁴² The experiments over the same HDR monitor as that of Ledda *et al.* were conducted with and without references of the corresponding real-world view. The result shows that subjects behaved differently with and without reference. They enhanced contrast proportionally to the dynamic range of an HDR display, even more than that of an original image, if they had no reference. However, if they had reference, they adjusted the contrast almost the same as that of an original image and kept it approximately on the same level even for different dynamic ranges of a display.⁴²

Our experiment described in this paper differs from the above in many ways. First, we provide corresponding real-world views of a tone mapped image for subjects as the reference. HDR monitors can produce a much wider dynamic range of luminance than that of today's common display devices, but they are still limited in terms of the representation of high contrast in high-frequency areas of an HDR image, and their luminance range cannot match

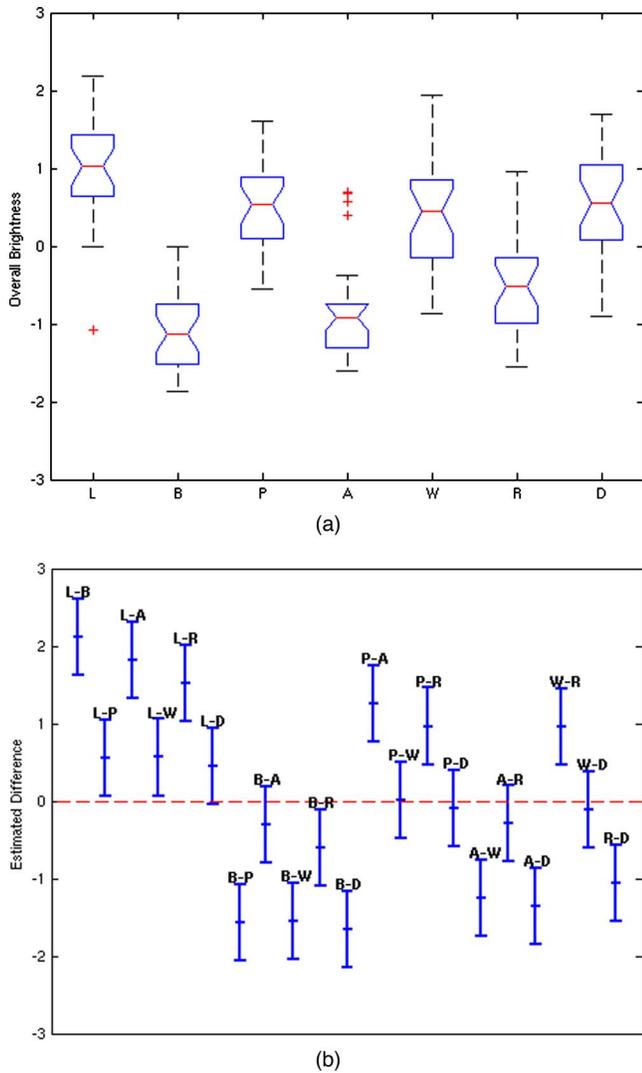


Fig. 3 Distributions and F and p values of overall brightness for the main effect of the tone mapping operators. The tone mapping operators are labeled as linear (L), bilateral filtering (B), Pattanaik (P), Ashikhmin (A), Ward (W), Reinhard (R), and Drago (D) methods. (a) Overall brightness. $F=46.14$, $p=0.0$. A box shows the lower quartile, median, and upper quartile values. The whiskers are lines extending from each end of the box to show the extent of the rest of the data. (b) Estimated mean difference. The middle bar shows the mean value of estimated difference. The top and bottom bars show the 95% confidence interval.

that of the realworld.³¹ Additionally, as shown by Yoshida *et al.*,⁴² subjects may behave differently when observing their preference or archiving the fidelity of an image, and an HDR monitor can even provide more contrast than an original image. Because, in this paper, we want to measure how close tone mapped images are to the real-world view, we do not select an HDR monitor as the reference. Second, we select the ranking method using a slider. The paired comparison method is very simple and powerful for observing interval scales of given algorithms (TMOs in this case) along a given dimension. Thurstone's law of comparative judgment³⁴ is the most common used analysis for paired comparison. However, the paired comparison analyzed by Thurstone's law has two problems; A paired comparison

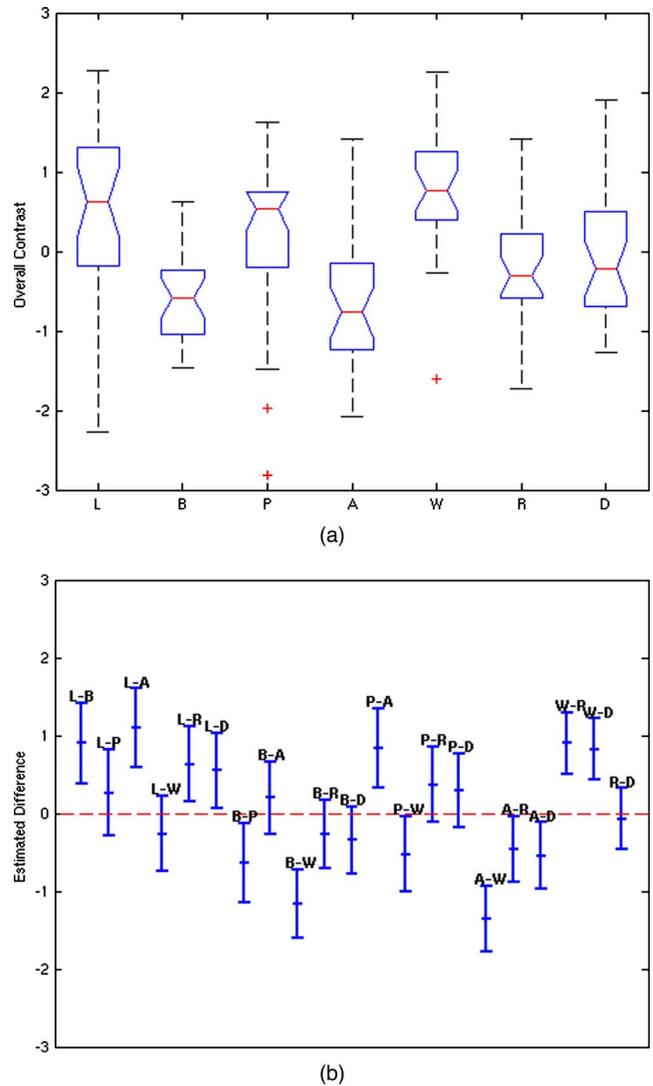


Fig. 4 Distributions and F and p values of overall contrast for the main effect of the tone mapping operators. The tone mapping operators are labeled as linear (L), bilateral filtering (B), Pattanaik (P), Ashikhmin (A), Ward (W), Reinhard (R), and Drago (D) methods. (a) Overall contrast. $F=8.74$, $p=2.1058E-08$. (b) Estimated mean difference.

experiment needs $n(n-1)/2$ experimental trials for n stimuli. The total number of trials increases very rapidly as the number of stimuli-increases. Additionally, Thurstone's law relies on the assumption of unidimensional scaling. It is quite useful if we want to compare the performance of one specific attribute, but in our experiment, we want to have more insight into multidimensional scaling than into unidimensional scaling. Therefore, we select the rating-scale method instead of the paired comparison. The details for those methods and analysis can be found in a book by Bartleson and Grum.³

3.1 Experimental Design and Participants

We selected a subjective method for our experiment with seven tone mapping operators (see Section 2) because (1) our goal is a comparison of tone mapped images and their

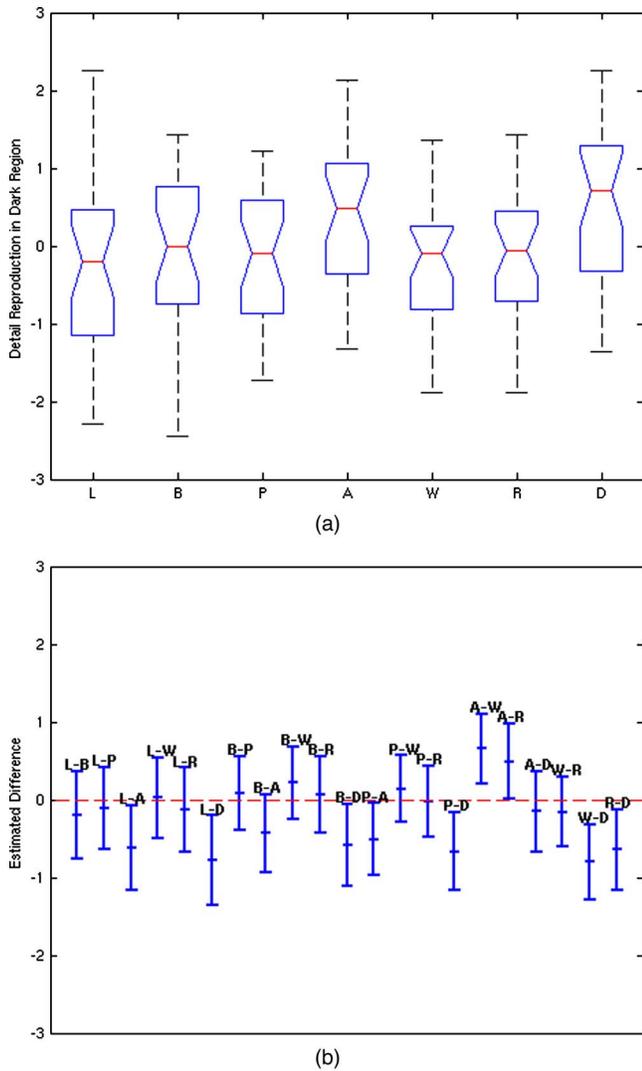


Fig. 5 Distributions and F and p values of detail reproductions in dark regions for the main effect of the tone mapping operators. The tone mapping operators are labeled as linear (L), bilateral filtering (B), Pattanaik (P), Ashikhmin (A), Ward (W), Reinhard (R), and Drago (D) methods. (a) Detail reproductions in dark regions. $F = 3.18$, $p = 0.0054$. (b) Estimated mean difference.

corresponding real-world views and (2) there exists no objective method to accommodate such a cross-media comparison.

The experiment was performed with the participation of 14 human observers. All 14 participants were graduate students and researchers of the Computer Graphics group in the Max-Planck-Institut für Informatik in Germany. Two were female, and the rest were male. They ranged in age between 24 and 34. All were unaware of both the goal of the experiment and the tone mapping operators. Additionally, every participant reported normal or corrected-to-normal vision.

3.2 Acquiring HDR Images

Figure 1 shows the acquired images for our perceptual evaluation. We selected indoor architectural scenes because the scene setup conditions are easily reproducible for an experiment. As seen in the figure, the dynamic ranges of

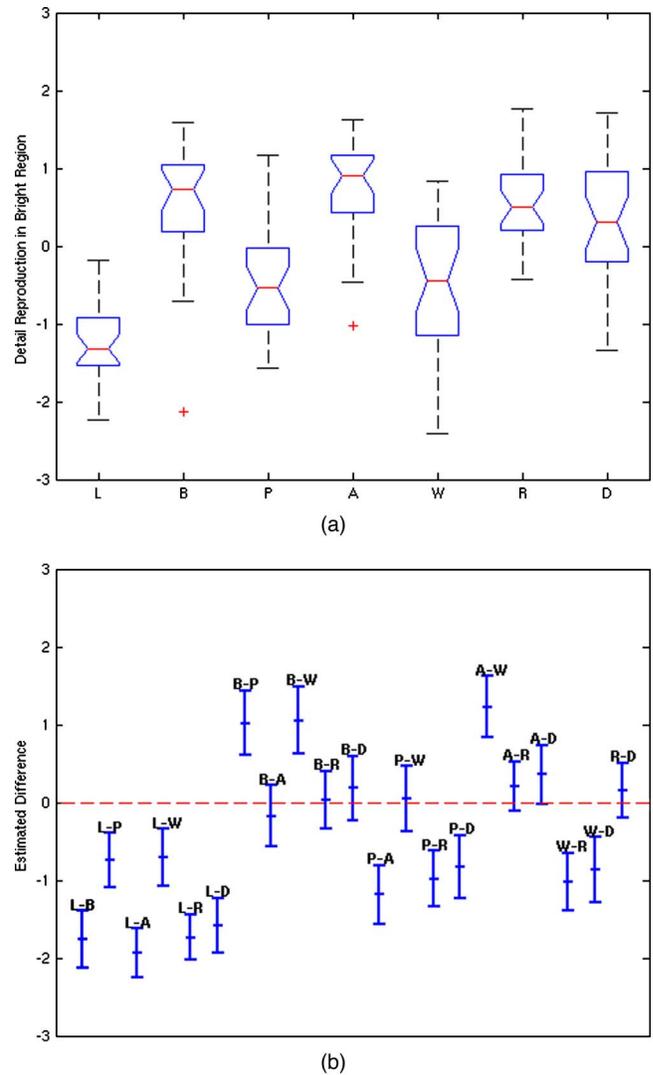


Fig. 6 Distributions and F and p values of detail reproductions in bright regions for the main effect of the tone mapping operators. The tone mapping operators are labeled as linear (L), bilateral filtering (B), Pattanaik (P), Ashikhmin (A), Ward (W), Reinhard (R), and Drago (D) methods. (a) Detail reproductions in bright regions. $F = 30.355$, $p = 0.0$. (b) Estimated mean difference.

both scenes are large enough to require HDR. Scene 1 has bright spotlights around the trees and quite dark areas behind the glass. Scene 2 also has highly bright and dark areas, and, in addition, it has some gray areas (on pipes), as pointed out by Gilchrist, result in a scaling problem.¹¹ The scaling problem concerns how the range of luminances in an image is mapped onto a range of perceived grays. The absolute luminances in both scenes were measured by a MINOLTA light meter LS-100.¹³ In Scene 1, the brightest area is 13,630 cd/m² and the darkest area is 0.021 cd/m². In Scene 2, the brightest area is 506.2 cd/m² and the darkest area is 0.019 cd/m².

To acquire HDR images for our perceptual evaluation, a Kodak Professional DCS560 digital camera and CANON EF lenses of 24 mm and 14 mm were used to take pictures with different shutter speeds ranging from 1/2000 to 8.0 seconds. Both lenses have a large enough

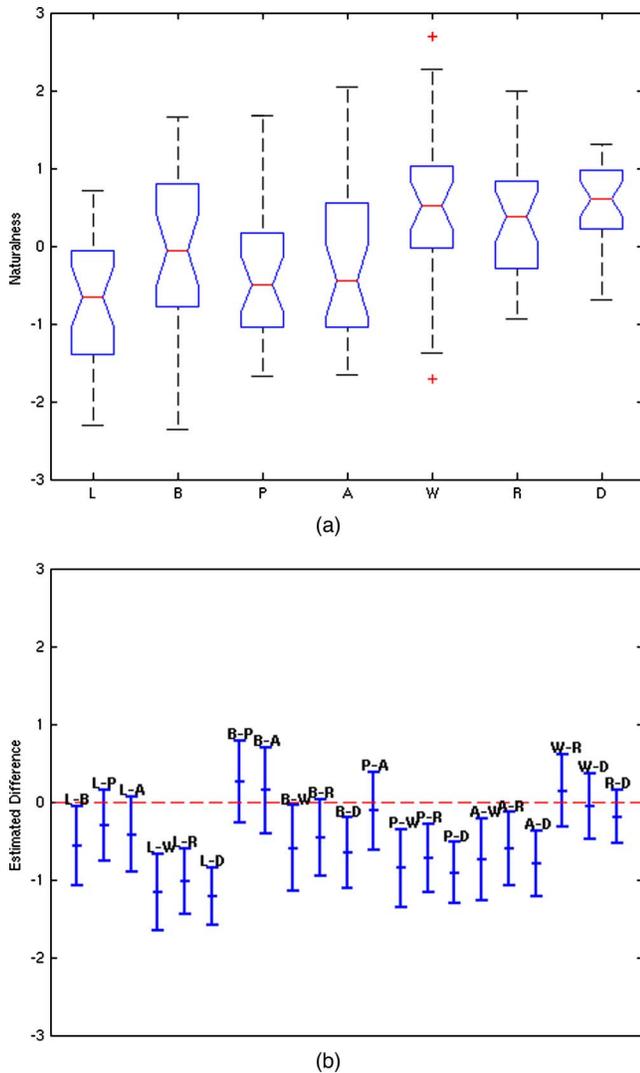


Fig. 7 Distributions and F and p values of naturalness for the main effect of the tone mapping operators. The tone mapping operators are labeled as linear (L), bilateral filtering (B), Pattanaik (P), Ashikhmin (A), Ward (W), Reinhard (R), and Drago (D) methods. (a) Naturalness. $F=8.11$, $p=8.3877E-08$. (b) Estimated mean difference.

field of view to cover the view of human eyes. Because those images were saved in Kodak's raw format, they were converted to 36-bit TIFF format using the program raw2image included in the Kodak DCS Photo Desk Package. Several techniques to recover the camera response curve function.^{5,19,29,39} We used one of the newest methods, the Robertson *et al.* method,²⁹ for static images to recover the camera response curve. The HDR images were constructed with the recovered response curve and saved in the Radiance RGBE format.³⁷ Each HDR image was created using 15 static images.

3.3 Experimental Procedure

Prior to the main experiment, we conducted a pilot study with experienced subjects to fine-tune the parameter settings for each tone mapping operator. We asked subjects to choose the best image for each tone reproduction method from a selection of multiple tone mapped images. Addi-

tional postprocessing, such as gamma correction, was performed according to the suggestions in each respective paper of the tone mapping operators. All images used in our experiment are shown in Appendix B. All HDR images as well as our implementations of the tone reproductions used in the experiment are available on our website.*

In the main experiment, the participants were asked to stay in the same position as that in which each of the HDR images had been taken (Fig. 1) and view seven images, one after another, for each of the two scenes. An sRGB-calibrated monitor (Dell UltraSharp 1800FP) displaying images of 1280×1024 resolution at 60 Hz was used. The subjects compared each image with its corresponding real-world view and gave ratings for image appearance and realism. The image appearance attributes judged in our experiment are overall brightness, contrast, detail reproductions in dark regions, and detail reproductions in bright regions. The subjects rated how well each of those attributes was reproduced in tone mapped images with respect to the real-world view. Additionally, the subjects were asked to rate image naturalness in terms of reproducing the overall appearance of the real-world views. The subjects moved scrollbars to give all of the ratings. The subjects were allowed to move back and forth among images for one scene (see Fig. 2 for an example screenshot). The whole procedure for each subject took approximately 20 to 30 minutes.

4 Results

Our experiment is a seven (tone mapping operators) \times two (scenes) within-subjects design. This experiment has two independent variables (IVs: the tone mapping operators and the scenes) and five dependent variables (DVs: overall brightness, overall contrast, detail reproductions in dark regions, detail reproductions in bright regions, and naturalness).[†] Our primary interest is whether the images produced by different tone mapping operators are perceived differently when they are compared to their corresponding real-world views. To analyze a set of data, we used MATLAB's Statistics Toolbox.¹⁷ As preliminary data processing, all obtained scores were normalized over each of the attributes and each of the subjects in order to scale the standard deviations 1.0 as $x_i \rightarrow (x_i - \mu_x) / \sigma_x$, where x_i is a score and μ_x and σ_x are, respectively, the mean and the standard deviation over an attribute of a subject.

4.1 Main Effects

Because two scenes were used in the experiment, we examined how much influence comes from the difference between two scenes. The *main effect* of the scenes was first tested by using analysis of variance (ANOVA). The F distribution and a probability value p , which is derived from F , were used to determine whether there is a statistically significant difference between populations of samples. The higher the p value is, the more we can believe that the populations of samples are not statistically different (refer to Tabachnick³³ for more details of the main effect and ANOVA). In our experiment, the scene difference is not

*http://www.mpi-sb.mpg.de/resources/tmo_and_real/

[†]IVs are the variables, which do not depend on participants, while DVs depend on participants' ratings.

Table 1 p values computed by t -test for overall brightness.

	Bilateral	Pattanaik	Ashikhmin	Ward	Reinhard	Drago
Linear	0.0000	0.0012	0.0000	0.0037	0.0000	0.0140
Bilateral		0.0000	0.0546	0.0000	0.0002	0.0000
Pattanaik			0.0000	0.9507	0.0000	0.5444
Ashikhmin				0.0000	0.0873	0.0000
Ward					0.0000	0.5580
Reinhard						0.0000

statistically significant and small enough to be ignored ($p \gg 0.05$ for all attributes). To achieve our goal, we needed to investigate the tone mapping performance for indoor architectural scenes.

The main effects of tone mapping operators were tested as shown in Figs. 3–7. Figure 3(a) shows the main effect of the tone mapping operators for overall brightness. According to the significance values, the overall brightness is the most differently perceived attribute. As seen in the figure, it is manifested that images produced by the linear, Pattanaik, Ward, and Drago methods (i.e., global methods) have substantially higher overall brightness than the others. These tone mapping operators are perceived the most differently when compared to their corresponding real-world views. Figure 3(b) shows the estimated differences of means. The top and bottom bars show a 95% confidence interval, and the middle bar shows the mean value of the differences. Note that a pair containing 0 difference (the red line in plots) in its 95% confidence interval indicates that they are not significantly different. The same result is shown in Table 1 as significance values calculated by a t -test. As shown in Fig. 3(b) and Table 1, the pair of the linear TMO and bilateral filtering has the biggest population difference. On the other hand, the pairs of bilateral filtering and Ashikhmin, of Pattanaik and Ward, of Pattanaik and Drago, of Ashikhmin and Reinhard, and of Ward and Drago are not significantly different for overall brightness.

The main effect for overall contrast is shown in Fig. 4. The linear, Pattanaik, and Ward methods have a higher overall contrast than the others. Global operators have a

stronger contrast than local ones, as shown in Fig. 4(a). It corresponds to the expectations because global methods require a high overall contrast to retain details in low-contrast regions. The estimated mean difference and t -test result are shown in Fig. 4(b) and Table 2. The ANOVA result [Fig. 4(a)] shows a significant difference among all TMOs; however, if we make pairwise comparisons for each TMO, a number of pairs are not considered significantly different [Fig. 4(b) and Table 2].

Detail reproduction in dark regions (Fig. 5 and Table 6) shows the least significance among the attributes, but it is still highly significant ($p=0.0054$). The Ashikhmin and Drago methods are perceived to have the most details reproduced in dark regions. The linear, Pattanaik, Ward, and Reinhard methods have almost equal scores, and the bilateral filtering has slightly less detail reproductions than those four. Although ANOVA shows that TMOs are significantly different for detail reproduction in dark regions: [$p=0.0054$ in Fig. 5(a)], the estimated difference of means and a t -test show that a number of TMOs are not significantly different if they are compared pairwise.

Detail reproduction in bright regions is the second most differently perceived attribute, as shown in Fig. 6. The bilateral filtering, Ashikhmin, and Reinhard methods provide significantly more detail reproductions in bright regions than the others. According to Fig. 6(a), all of the local operators are perceived with more detail reproductions than global ones. This result comes from the fact that local operators use different scales for small regions of an image

Table 2 p values computed by t -test for overall contrast.

	Bilateral	Pattanaik	Ashikhmin	Ward	Reinhard	Drago
Linear	0.0010	0.3531	0.0001	0.2713	0.0121	0.0260
Bilateral		0.0161	0.4003	0.0000	0.2197	0.1132
Pattanaik			0.0017	0.0326	0.1323	0.2305
Ashikhmin				0.0000	0.0322	0.0125
Ward					0.0000	0.0001
Reinhard						0.6942

Table 3 p values computed by t -test for detail reproduction in bright regions.

	Bilateral	Pattanaik	Ashikhmin	Ward	Reinhard	Drago
Linear	0.0000	0.0001	0.0000	0.0003	0.0000	0.0000
Bilateral		0.0000	0.3665	0.0000	0.8794	0.4011
Pattanaik			0.0000	0.8670	0.0000	0.0001
Ashikhmin				0.0000	0.2044	0.0653
Ward					0.0000	0.0001
Reinhard						0.4053

while global operators use only one scale for the whole image and tend to saturate bright parts of an image. Figure 6(b) and Table 3 also show the distances between local and global TMOs.

Figure 7 and Table 7 show the main effect for naturalness. As we can see in the figure, the Ward, Reinhard, and Drago methods are perceived to have the most natural appearance. As with detail reproduction in dark regions, ANOVA shows that all TMOs are significantly different ($p=8.3877E-08$); however, each pairwise comparison does not show a significant difference for almost half of the pairs for naturalness.

4.2 Correlations

The correlations between all attributes were tested using the Pearson r correlation coefficient.²⁴ All of the Pearson r values between each pair of the attributes are shown in Table 4. As the table shows, pairs of the naturalness and each of the overall brightness and detail reproductions in dark regions have very weak correlations. On the other hand, the naturalness and each of the overall contrast and detail reproductions in bright regions have stronger correlations, but they are still small. It can be concluded from this result that none of those image appearance attributes has a strong enough influence to determine naturalness by itself, and naturalness can be considered by a combination of those attributes multidimensionally.

The pair of the overall brightness and detail reproductions in bright regions has the biggest negative correlation,

Table 4 Correlations with the Pearson r values of all pairs of overall brightness (O.B.), overall contrast (O.C.), detail reproductions in dark regions (D.D.), detail reproductions in bright regions (D.B.), and naturalness (N.).

	O.C.	D.D.	D.B.	N.
O.B.	0.3321	-0.1023	-0.55880	-0.0844
O.C.		-0.1114	-0.28220	0.2440
D.D.			0.26340	0.0729
D.B.				0.3028

with a Pearson coefficient of $r=-0.55880$. This was expected, because as overall brightness decreases, bright parts are less saturated and more visible.

4.3 Dimension Estimate and Mahalanobis Distances

In addition, the data were analyzed by using multivariate analysis of variance (MANOVA). MANOVA is an extension of ANOVA to analyze multiple IVs and/or multiple DVs (see Tabachnick³³ for details). MANOVA in MATLAB's Statistics Toolbox provides an estimate of the dimension d of the space containing the group means and the significance values for each of the dimensions. If $d=0$, then the means are at the same point. If $d=1$, the means are different but along a line; if $d=2$, the means are on a plane but not along a line, and similarly for higher values of d . The null hypotheses are tested by calculating the significant values (p values) in each of the dimensions such as the means are in N -dimensional space, where N is the number

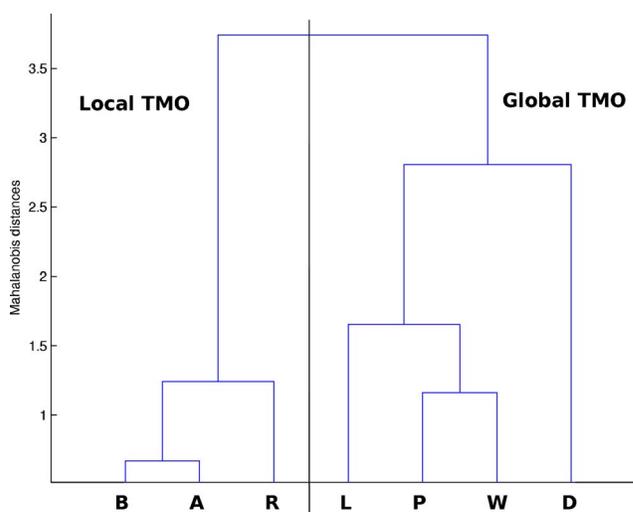


Fig. 8 Tree-structured Mahalanobis distances to determine similarity among the tone mapping operators given by MANOVA. As in Figs. 3–7, the tone mapping operators are labeled as linear (L), bilateral filtering (B), Pattanaik (P), Ashikhmin (A), Ward (W), Reinhard (R), and Drago (D) methods. Note that those tone mapping operators are divided into global and local methods by Mahalanobis distances.

Table 5 Mahalanobis distances among the tone mapping operators provided by MANOVA. The three biggest distances are written in a bold font, while the three smallest distances are underlined. All the biggest differences are from the linear tone mapping. Those Mahalanobis distances are visualized in Fig. 8.

	Bilateral	Pattanaik	Ashikhmin	Ward	Reinhard	Drago
Linear	15.4530	1.6541	14.2361	2.7122	10.6089	6.6940
Bilateral		7.4749	<u>0.6674</u>	9.2726	1.3353	8.8120
Pattanaik			6.4395	<u>1.1613</u>	3.9887	2.8066
Ashikhmin				8.9709	1.2405	6.2989
Ward					4.5301	2.9536
Reinhard						3.7406

Table 6 p values computed by t -test for detail reproduction in dark regions.

	Bilateral	Pattanaik	Ashikhmin	Ward	Reinhard	Drago
Linear	0.4869	0.6613	0.0252	0.9355	0.6131	0.0087
Bilateral		0.7400	0.0884	0.3572	0.8153	0.0311
Pattanaik			0.0308	0.5240	0.9247	0.0096
Ashikhmin				0.0057	0.0453	0.5508
Ward					0.4793	0.0017
Reinhard						0.0147

Table 7 p values computed by t -test for naturalness.

	Bilateral	Pattanaik	Ashikhmin	Ward	Reinhard	Drago
Linear	0.0305	0.1845	0.0864	0.0000	0.0000	0.0000
Bilateral		0.3282	0.6094	0.0329	0.0644	0.0049
Pattanaik			0.6420	0.0012	0.0017	0.0000
Ashikhmin				0.0069	0.0124	0.0004
Ward					0.5603	0.7800
Reinhard						0.2608

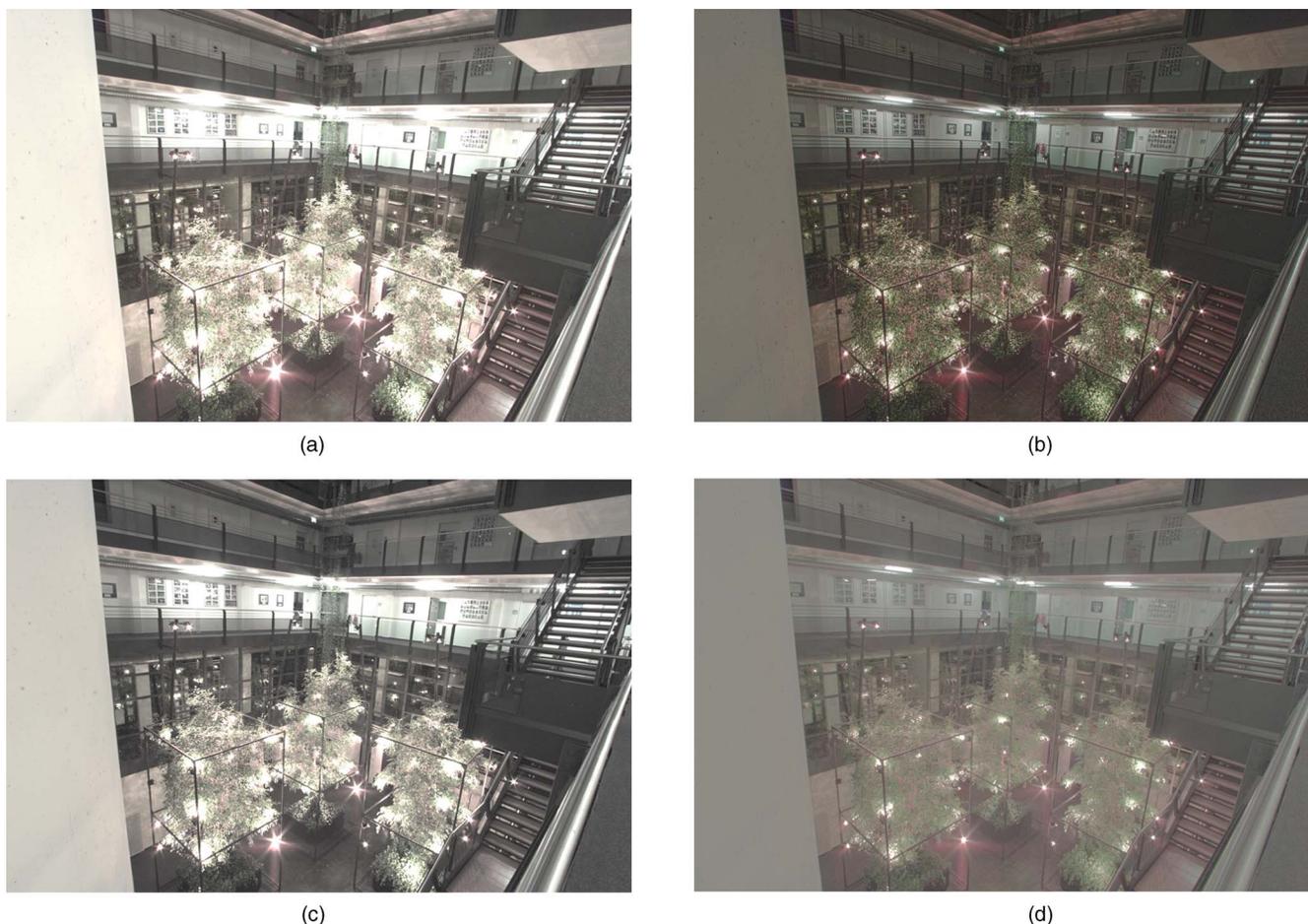


Fig. 9 The tone mapped images for Scene 1: (a) linear, (b) bilateral filtering, (c) Pattanaik, and (d) Ashikhmin.

of dimensions. From our set of data, MANOVA returns $d = 3$, which indicates that the means are in a space with at least three dimensions. The p values for each of the dimensions in our perceptual experiment are

$$p(d = 0) = 0.0000,$$

$$p(d = 1) = 0.0000,$$

$$p(d = 2) = 0.0017,$$

$$p(d = 3) = 0.9397,$$

$$p(d = 4) = 0.8833.$$

For $d = 3$, the possibility is the largest and above the significance level; therefore, the means are located in an at least three-dimensional space or even in a higher-dimensional space, but it is not possible to locate them either on a line or on a surface.

Additionally, the Mahalanobis distances of IVs are calculated. The Mahalanobis distance is calculated as

$$d(x_1, x_2) = \langle x_1 - x_2, \mathbf{S}^{-1}(x_1, x_2) \rangle \quad (1)$$

where $\mathbf{S}(x_1, x_2) = 1/(n-1) \sum_{i=1}^n (x_1 - x_2)^i (x_1 - x_2)^i$, where \mathbf{X} is a data matrix, \mathbf{x}_i is the i th row of \mathbf{X} , $\bar{\mathbf{X}}$ is a row vector of means, and n is the number of rows. The Mahalanobis distance is a measurement of the similarity of data points relative to the probability distribution of data, which has different variances along different dimensions. Table 5 shows the Mahalanobis distances among the tone mapping operators given by MANOVA. According to Table 5, the linear tone mapping and bilateral filtering are perceived the most differently when compared with their corresponding real-world views. The second and the third most different combinations come from the combination of the linear tone mapping and Ashikhmin method and of the linear tone mapping and Reinhard method. All three biggest differences are found with respect to the linear tone mapping. On the other hand, the least difference is provided between bilateral filtering and the Ashikhmin method. This result is visualized in Fig. 8 as a dendrogram plot of a hierarchical binary tree. An interesting result shown in Fig. 8 is that those seven tone mapping reproductions are divided into global and local methods by Mahalanobis distances. Three local operators (bilateral, Ashikhmin, and Reinhard) are



(a)



(b)



(c)

Fig. 10 The tone mapped images for Scene 1: (a) Ward, (b) Reinhard, and (c) Drago.

similar to each other, and four global operators (linear, Pattanaik, Ward, and Drago) are similar to each other, but both categories of global and local operators are separated by a large distance.

5 Discussion

In this concluding section, we discuss the results obtained in our experiments and, whenever possible, compare them with the results obtained in other independent studies.^{14,16} It

is worth noting that Kuang *et al.*^{14,15} did not use any reference scenes in their experiment. Ledda *et al.*¹⁶ showed HDR images to their subjects as the reference, but the subjects never actually saw the real-world scenes captured in the HDR images (see Figs. 9–12).

The result of the multivariate analysis (see Section 4) shows that the seven tone mapping operators were perceived very differently in terms of all of the attributes when compared to their corresponding real-world views. Observer assessments of overall brightness show the most significant differences among the tone reproductions, and global operators (the linear, Pattanaik, Ward, and Drago methods) have more brightness than local ones (the bilateral filtering, Ashikhmin, and Reinhard methods). Regarding overall contrast, global operators have more contrast than local ones, but the difference is less pronounced than for overall brightness. Since overall brightness and contrast are correlated,⁴² this result is expected. The linear, Pattanaik, and Ward methods show more overall contrast reproduction than the others. However, it was shown that the bilateral filtering and photographic TMOs (i.e., local operators) had the highest rating scales in overall contrast when no reference was provided for subjects.¹⁴

Detail reproduction in dark regions is the least significant of the attributes, but it is still highly significant. The Ashikhmin and Drago methods are perceived as providing the most details in dark regions. The bilateral filtering exhibits slightly better detail reproduction than the linear, Pattanaik, Ward, and Reinhard methods. In another independent study (performed without the reference scenes),¹⁴ the bilateral filtering and Reinhard methods were reported to have high scores in detail reproductions in dark regions. It is similar to our findings. On the other hand, if an HDR image was provided as the reference on an HDR monitor (Ledda *et al.*¹⁶), the bilateral filtering reproduced the poorest accuracy of details in dark regions.

Perceptual variation is the second highest for detail reproduction in bright regions. Counter to overall brightness, local operators are perceived with significantly more detail reproductions in bright regions than global ones. Even when no reference was provided, the local TMOs were considered to be better operators for detail reproduction in bright regions.¹⁴ In the study with the HDR references,¹⁶ it can also be seen that the Reinhard TMO reproduced details in bright regions quite well. However, the fast bilateral filtering, Ward, and Drago TMOs were reported as having the opposite effect compared to our results.

Similar results to those of our experiments with respect to detail reproduction in dark and bright regions have been also recently reported by Smith *et al.*,³² who proposed objective local and global contrast metrics. They considered 18 HDR images, which were compared to their tone mapped counterparts using their objective contrast distortion metrics.

Concerning the naturalness attribute, the Ward, Reinhard, and Drago methods obtained the highest scores. The Ward and Reinhard methods are also ranked as the second and the third preferred TMOs, respectively, in the research of Kuang *et al.* However, the fast bilateral filtering is not ranked in our experiments as highly as in the experiments without reference.¹⁵

While the results of our experiments show some simi-



Fig. 11 The tone mapped images for Scene 2: (a) linear, (b) bilateral filtering, (c) Pattanaik, and (d) Ashikhmin.

larities to the results of other studies,^{14,16} a number of differences are observed as well. This may come from the difference between an experiment with or without reference. Even when a reference is provided to subjects, the results differ with an HDR image reference or with the corresponding real-world scene.

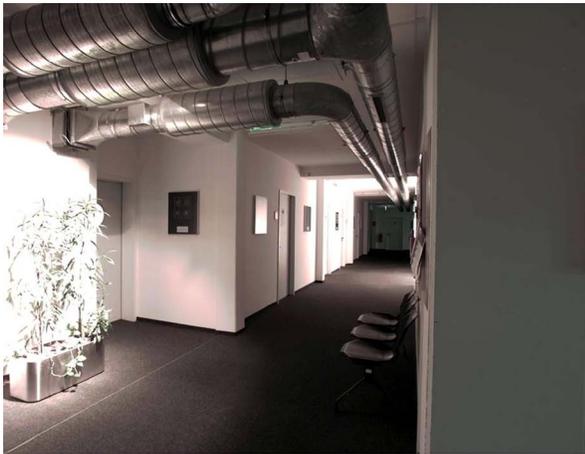
Correlations between naturalness and each of the image appearance attributes were tested (see Section 4.2). The obtained result shows that none of the image appearance attributes has a strong influence on the perception of naturalness by itself. This may suggest that naturalness depends on a combination of the other attributes with different weights. All the other possible pairs between attributes were also tested. The biggest negative correlation happens between overall brightness and detail reproductions in bright regions, which may be due to saturation or to contrast compression in bright regions.

The multivariate analysis of variance (MANOVA) shows that the means of the set of data are embedded in an at least three-dimensional space but neither on a line nor on a two-dimensional surface. In terms of the similarity of the

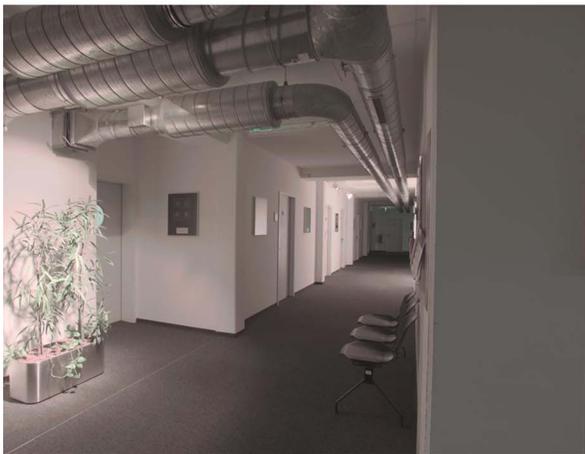
tone mapping operators computed by Mahalanobis distances, the biggest differences are between the linear tone mapping and each of the fast bilateral filtering, the Ashikhmin method, and the photographic tone reproduction by Reinhard *et al.* (i.e., local methods). The least differences are between fast bilateral filtering and the Ashikhmin method, between the methods of Pattanaik and Ward, and between the Ashikhmin method and the photographic reproduction. The Mahalanobis distances are visualized in a dendrogram plot (Fig. 8), which shows that all studied tone mapping operators are divided into global and local categories in terms of similarity.

6 Conclusions

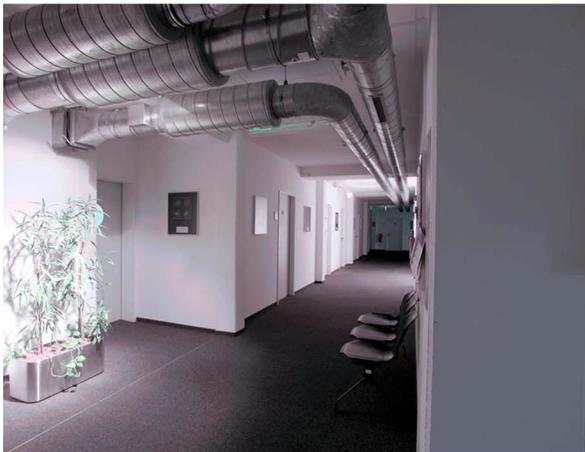
We conducted a psychophysical experiment over seven tone mapping operators and two scenes with 14 human observers. We asked subjects to stay at the point where an HDR image was taken by camera and compare tone mapped images and their corresponding real-world views. Tone mapped images were shown one after another, and



(a)



(b)



(c)

Fig. 12 The tone mapped images for Scene 2: (a) Ward, (b) Reinhard, and (c) Drago.

subjects used a slider to rate overall brightness, overall contrast, detail reproduction in bright and dark regions, and naturalness for each tone mapped image. Our principal goal of this work was to study how people perceive tone mapped images when they are compared with the real-world views.

Our results demonstrate that qualitative differences in

tone mapping operators have a systematic effect on the perception of scenes by human observers. They provide a solid basis for selecting the appropriate algorithm for a given application and for assessment of new approaches to tone mapping techniques.

Acknowledgments

We wish to express our warmest gratitude to the 14 people who participated in our psychophysical experiment. Additionally, we would like to thank Grzegorz Krawczyk and Rafał Mantiuk for providing some tone mapping operators. We are also grateful to Frédéric Drago for providing his software package and tone mapped images by his tone reproduction.

This work was supported in part by the European Community within the scope of the RealReflect project IST-2001-34744, “realtime visualization of complex reflectance behavior in virtual prototyping.”

References

1. M. Ashikhmin, “A tone mapping algorithm for high contrast images,” in P. Debevec and S. Gibson (eds.), *13th Eurographics Workshop on Rendering*, pp. 145–155, Pisa, Italy (2002).
2. C. J. Bartleson and E. J. Breneman, “Brightness reproduction in the photographic process,” *Photograph. Sci. Eng.* **11**, 254–262 (1967).
3. C. J. Bartleson and F. Grum, *Optical Radiation Measurements*, Vol. 5 of *Visual Measurements*, pp. 467–471, Academic Press, Burlington, MA (1984).
4. P. Choudhury and J. Tumblin, “The trilateral filter for high contrast images and meshes,” In *Proc. 13th Eurographics Workshop on Rendering*, ACM International Conference Proceeding Series, pp. 186–196, Eurographics Association (2003).
5. P. E. Debevec and J. Malik, “Recovering high dynamic range radiance maps from photographs,” in Turner Whitted (ed.), *SIGGRAPH 97 Conf. Proc.*, Annual Conference Series, pp. 369–378, ACM SIGGRAPH, Addison-Wesley, Los Angeles, CA (1997).
6. K. Devlin, A. Chalmers, A. Wilkie, and W. Purgathofer, “Tone reproduction and physically based spectral rendering,” in Dieter Fellner and Roberto Scopigno (eds.), *State of the Art Reports, Eurographics*, pp. 101–123, Eurographics Association (2002).
7. F. Drago, K. Myszkowski, T. Annen, and N. Chiba, “Adaptive logarithmic mapping for displaying high contrast scenes,” in P. Brunet and D. Fellner (eds.), *Proc. Eurographics*, pp. 419–426 (2003).
8. F. Durand and J. Dorsey, “Fast bilateral filtering for the display of high-dynamic-range images,” in *Proc. ACM SIGGRAPH 2002*, Computer Graphics Proceedings, Annual Conference Series (2002).
9. R. Fattal, D. Lischinski, and M. Werman, “Gradient domain high dynamic range compression,” in *Proc. ACM SIGGRAPH 2002*, pp. 249–256 (2002).
10. J. A. Ferwerda, S. Pattanaik, P. Shirley, and D. P. Greenberg, “A model of visual adaptation for realistic image synthesis,” in H. Rushmeier (ed.), *SIGGRAPH 96 Conf. Proc.*, Annual Conference Series, pp. 249–258, ACM SIGGRAPH, Addison-Wesley, New Orleans, LA (1996).
11. A. Gilchrist, C. Kossyfidis, F. Bonato, T. Agostini, J. Cataliotti, X. Li, B. Spehar, V. Annan, and E. Economou, “An anchoring theory of lightness perception,” *Psychol. Rev.* **106**(4), 795–834 (1999).
12. L. A. Jones and H. R. Condit, “The brightness scale of exterior scenes and the computation of correct photographic exposure,” *J. Opt. Soc. Am.* **31**, 651 (1941).
13. Konica Minolta, Inc., <http://konicaminolta.com>.
14. J. Kuang, G. M. Johnson, and M. D. Fairchild, “Image preference scaling for HDR image rendering,” in *Proc. 13th Color Imaging Conf.*, pp. 8–13 (2005).
15. J. Kuang, H. Yamaguchi, G. M. Johnson, and M. D. Fairchild, “Testing HDR image rendering algorithms,” in *Proc. IS&T/SID 12th Color Imaging Conf.*, pp. 315–320 (2004).
16. P. Ledda, A. Chalmers, T. Troscianko, and H. Seetzen, “Evaluation of tone mapping operators using a high dynamic range display,” in *Proc. ACM SIGGRAPH 2005* (2005).
17. MathWorks, Inc., MATLAB, <http://www.mathworks.com>.
18. G. S. Miller and C. R. Hoffman, “Illumination and reflection maps: Simulated objects in simulated and real environments,” in *SIGGRAPH 84 Course Notes for Advanced Computer Graphics Animation* (1984).
19. T. Mitsunaga and S. K. Nayar, “Radiometric self calibration,” in *Proc. IEEE Conf. on Computer Vision and Pattern Recognition* (1999).

20. G. R. Newsham, H. Seetzen, and J. A. Veitch, "Lighting quality evaluations using images on a high dynamic range display," in *Proc. ARCC/EAAE Conf. on Architectural Research*, pp. 1–9 (2002).
21. M. Nijenhuis, R. Hamberg, C. Teunissen, S. Bech, H. Looren de Jong, P. Houben, and S. K. Pramanik, "Sharpness, sharpness related attributes, and their physical correlates," *Proc. SPIE* **3025**, 173–184 (1997).
22. S. N. Pattanaik, J. A. Ferwerda, M. Fairchild, and D. P. Greenberg, "A multiscale model of adaptation and spatial vision for realistic image display," in *SIGGRAPH '98 Proc.*, pp. 287–298 (1998).
23. S. N. Pattanaik, J. E. Tumblin, H. Yee, and D. P. Greenberg, "Time-dependent visual adaptation for fast realistic image display," in *Proc. of ACM SIGGRAPH 2000*, Annual Conference Series, pp. 47–54, ACM Press/ACM SIGGRAPH/Addison-Wesley Longman, New Orleans, LA (2000).
24. K. Pearson, "Regression, heredity, and panmixia," *Philos. Trans. R. Soc. London, Ser. A* **187**, 253–318 (1896).
25. K. Perlin and E. M. Hoffert, "Hypertexture," in *Proc. ACM SIGGRAPH 89* **23**, 253–262 (1989).
26. E. Reinhard, "Parameter estimation for photographic tone reproduction," *J. Graphics Tools* **7**(1), 45–51 (2003).
27. E. Reinhard, M. Stark, P. Shirley, and J. Ferwerda, "Photographic tone reproduction for digital images," in *SIGGRAPH 2002 Conf. Proc.*, ACM SIGGRAPH, Addison-Wesley, San Antonio, CA (2002).
28. E. Reinhard, G. Ward, S. Pattanaik, and P. Debevec, *High Dynamic Range Imaging: Acquisition, Display, and Image-Based Lighting*, Morgan Kaufmann, San Francisco, CA (2005).
29. M. A. Robertson, S. Borman, and R. L. Stevenson, "Dynamic range improvement through multiple exposures," *IEEE Int. Conf. on Image Processing* (1999).
30. A. Scheel, M. Stamminger, and H.-P. Seidel, "Tone reproduction for interactive walkthroughs," in *Proc. Eurographics Conf.*, pp. 301–312 (2000).
31. H. Seetzen, W. Heidrich, W. Stuerzlinger, G. Ward, L. Whitehead, M. Trentacoste, A. Ghosh, and A. Vorozcovs, "High dynamic range display systems," in *Proc. of ACM SIGGRAPH 2004* (2004).
32. K. Smith, G. Krawczyk, K. Myszkowski, and H.-P. Seidel, "Beyond tone mapping: Enhanced depiction of tone mapped HDR images," in *Proc. of Eurographics 2006* (2006).
33. B. G. Tabachnick, *Using Multivariate Statistics*, 2nd ed., Harper Collins, New York (1989).
34. L. L. Thurstone, "A law of comparative judgment," *Psychol. Rev.* **34**, 273–286 (1927).
35. J. Tumblin and H. E. Rushmeier, "Tone reproduction for realistic images," *IEEE Comput. Graphics Appl.* **13**(6), 42–48 (1993).
36. J. Tumblin and G. Turk, "LCIS: A boundary hierarchy for detail-preserving contrast reduction," in Alyn Rockwood (ed.), *Siggraph 1999, Computer Graphics Proc.*, Annual Conference Series, pp. 83–90 (1999).
37. G. Ward, *Graphics Gems II*, pp. 80–83, Academic Press, Burlington, MA (1991).
38. G. Ward, "A contrast-based scalefactor for luminance display," *Graphics Gems IV*, pp. 415–421 (1994).
39. G. Ward, "Fast, robust image registration for compositing high dynamic range photographs from hand-held exposures," *J. Graphics Tools* **8**(2), 17–30 (2003).
40. G. W. Larson, H. Rushmeier, and C. Piatko, "A visibility matching tone reproduction operator for high dynamic range scenes," *IEEE Trans. Vis. Comput. Graph.* **3**(4), 291–306 (1997).
41. B. Watson, A. Friedman, and A. McGaffey, "Measuring and predicting visual fidelity," in *Proc. ACM SIGGRAPH 2001*, Computer Graphics Proceedings, Annual Conference Series, pp. 213–220, ACM Press/ACM SIGGRAPH/Addison-Wesley, Los Angeles, CA, (2001).
42. A. Yoshida, R. Mantiuk, K. Myszkowski, and H.-P. Seidel, "Analysis of reproducing real-world appearance on displays of varying dynamic range," in *Proc. of Eurographics 2006* (2006).

Biographies and photographs of the authors not available.