Do HDR Displays Support LDR Content?
A Psychophysical Evaluation

Ahmet Oğuz Akyüz[1,2] Roland Fleming[2] Bernhard E. Riecke[2] Erik Reinhard[3,1] Heinrich H. Bülthoff[2]

[1]University of Central Florida [2]Max Planck Institute for Biological Cybernetics [3]University of Bristol

Abstract

The development of high dynamic range (HDR) imagery has brought us to the verge of arguably the largest change in image display technologies since the transition from black-and-white to color television. Novel capture and display hardware will soon enable consumers to enjoy the HDR experience in their own homes. The question remains, however, of what to do with existing images and movies, which are intrinsically low dynamic range (LDR). Can this enormous volume of legacy content also be displayed effectively on HDR displays? We have carried out a series of rigorous psychophysical investigations to determine how LDR images are best displayed on a state-of-the-art HDR monitor, and to identify which stages of the HDR imaging pipeline are perceptually most critical. Our main findings are: (1) As expected, HDR displays outperform LDR ones. (2) Surprisingly, HDR images that are tone-mapped for display on standard monitors are often no better than the best single LDR exposure from a bracketed sequence. (3) Most importantly of all, LDR data does not necessarily require sophisticated treatment to produce a compelling HDR experience. Simply boosting the range of an LDR image linearly to fit the HDR display can equal or even surpass the appearance of a true HDR image. Thus the potentially tricky process of *inverse tone mapping* can be largely circumvented.

CR Categories: I.4.0 [Image Processing and Computer Vision]: General—Image Displays; I.3.3 [Computer Graphics]: Picture/Image Generation—Display Algorithms; H.1.2 [Models and Principles]: User/Machine SystemsHuman factors—Human Information Processing

Keywords: High dynamic range (HDR) imaging, high dynamic range display devices, tone mapping, psychophysics

1 Introduction

High dynamic range imaging (HDRI) is currently receiving considerable attention from both academia and industry. For instance, the movie and games industries are rapidly switching to an HDR rendering pipeline, aided by floating point support on graphics cards, while Hollywood routinely employs HDR cameras and image based lighting for special effects. It seems reasonable to assume that end-users will be able to afford HDR-enabled technologies within just a few years.

What makes HDRI attractive to such a diverse set of groups in both industry and academia is that it allows capture, storage, and processing of photometrically correct information, independent of artificial limits imposed by traditional imaging and display devices. This brings about the notion of scene referred data capture, closing the gap between the camera and the photometer.

Despite its advantages, there remain several open problems in the field of HDR imaging. One of the most important of these is the question of how to display HDR content directly. Conventional monitors are not tailored to display floating point data, let alone provide the necessary luminance range to convey a true HDR experience. Therefore, dynamic range reduction (i.e. tone mapping) is commonly employed to prepare HDR imagery for display on conventional screens.

This problem is significantly reduced with the recent advent of HDR display devices [Seetzen et al. 2004]. In the near future, we anticipate that the price of high dynamic range display devices will reach consumer levels. The implication would be that HDRI will find wide-spread use. During this short transition time, there will be a significant need to display conventional images on HDR display devices. Although this need may reduce over time, eight-bit photography will be with us for a considerable length of time.

This means that an enormous body of existing images will need to be displayed on HDR display hardware. Display algorithms will typically have to scale up the luminance range, rather than compress it. Thus, we are faced with the problem of inverse tone reproduction, to which currently only few solutions exist [Banterle et al. 2006; Meylan et al. 2006; Rempel et al. 2007].

Before a good inverse tone reproduction operator can be developed, though, it is desirable to determine the boundaries and limitations to which such an algorithm should adhere. As currently very little is known about HDR display hardware, and even less about inverse tone mapping, we have carried out several studies to determine the circumstances under which conventional images may be displayed on HDR displays. We first determine whether the visual experience offered by such displays is indeed superior to conventional displays. In a second experiment, we test whether the visual experience can be predominantly attributed to their improved contrast, or to their higher absolute luminance levels. Finally, we present an experiment designed to test whether LDR images can be displayed on an HDR display, after having been appropriately inverse tone-mapped, and whether this procedure matches the visual quality afforded by the direct display of HDR images.

2 Previous Studies

In industry and academia, many groups are addressing a range of problems associated with capture, storage, and display of HDR images. An overview of the software developments in these areas is given by Reinhard et al. [2005], whereas HDRI hardware is discussed in detail by Hoefflinger [2007].

High dynamic range display devices are a relatively new development. The models currently known are all based on an LCD screen where the uniform backlight is replaced by a spatially varying backlight. Early prototypes used a projector and Fresnel lens assembly to backlight the LCD screen [Seetzen and Whitehead 2003]. The intensity range of a high dynamic range image is then split into two, yielding separate data driving the LCD and the back-projector. Difficulties aligning the projector with the LCD display, as well as limitations in the black-level that can be achieved with this set-up, make this approach commercially impractical.

Second generation high dynamic range display devices are therefore constructed by placing an LCD screen in front of a 2D array of ultra-bright LEDs, which can be individually modulated [Seetzen et al. 2004]. Such an assembly overcomes the disadvantages of a projector-based system, but the cost of LEDs, as well as current limitations in the manufacturing process, place practical limitations on the resolution of the backlight array. For instance, 18" displays were manufactured containing 760 LEDs placed behind an LCD screen with a resolution of 1280 by 1024 pixels. The black level of this system is 0.03 cd/m^2 and its peak luminance is 8500 cd/m^2.

The latest model is the BrightSide DR37-P, which is the one used for our experiments. Its display area measures 32.26" by 18.15" with a resolution of 1920 by 1080 pixels for the TFT active matrix LCD screen, which is illuminated by 1380 LEDs. This configuration is capable of a contrast ratio in excess of 200,000:1, with a black level of 0.015 cd/m^2. The peak luminance is rated to be higher than 3000 cd/m^2.

The progress in HDR display technology led to various experiments aimed at understanding how to best utilize the emerging HDR display devices. To this end two main types of studies have appeared. The first type investigates the advantages of HDR displays in conveying enhanced realism and visual quality. In a set of experiments carried out on the aforementioned 18" HDR displays, Seetzen et al. [2006] investigated the effects of peak luminance, contrast, and amplitude resolution on user preference for the purpose of arriving at appropriate design criteria for HDR display devices. These experiments revealed that for a given contrast ratio, the perceived image quality increases with peak luminance up to a certain value, after which it decreases again. Thus, there exists an optimal peak luminance level, which depends on the chosen contrast ratio.

In another perceptual evaluation, Yoshida et al. [2006] used an HDR display to simulate several displays of different dynamic ranges. Participants were asked to adjust the parameters of a simple tone mapping operator based on their preference and the fidelity of the renderings with respect to real scenes. The main focus of this study was to discover the desired properties of a tone mapping operator. However, the experiments also revealed that participants had a tendency toward brighter images. Boosting contrast by lowering the black level was found to be of secondary importance.

The second type of study focuses on what to do with the enormous volume of existing LDR material. The problem here is to discover what sort of image processing best prepares LDR content for display on an HDR display device. Algorithms solving this general problem are known as inverse tone reproduction operators. As an example, one may invert the photographic tone reproduction operator and combine its output with a density map of the light sources estimated from the input image [Banterle et al. 2006]. The density map allows for a greater increase of dynamic range, as naïve inversion of the tone mapping operator can only yield a moderate dynamic range without causing blocking artifacts. Although this approach is found to work well for static images, the density map gives rise to flickering artifacts for video sequences limiting the amount of dynamic range enhancement.

Alternatively, one may focus on highlights, and vary the display range allocated to specular highlights separately from the remaining diffuse image content [Meylan et al. 2006]. For each image, a pixel intensity is selected manually which serves as a threshold between the diffuse regions and the specular highlights. These two regions are then linearly scaled using functions of different slopes. In a sense, this approach is justified by theories of lightness perception [Gilchrist et al. 1999] which require an anchor point between the luminance values in an image and the lightness values they represent. This approach is found to outperform simple linear scaling for dark images, although the naïve linear scaling works just as well for brighter images.

Real-time inverse tone mapping is desirable, especially if it can be included in the control hardware built into HDR display devices [Rempel et al. 2007]. Such an algorithm may be constructed by first linearizing the image using a gamma function followed by a linear expansion of the luminance range. If necessary, the output is "cleaned-up" to reduce noise and quantization artifacts. The final HDR image is obtained by further smoothly amplifying the contrast of the brightest regions to prevent discontinuity artifacts.

Previous psychophysical studies were intended to establish criteria either for the design of HDR displays or for determining the parameters of an inverse tone mapping operator. In contrast, our work is designed to establish criteria for the display of both LDR and HDR images, given a specific HDR display device. Hence, we approach the problem from the viewpoint of a user of HDR display devices, rather than a designer of such devices. In addition, we investigate how LDR, HDR, and mixed imaging pipelines compare with respect to each other.

3 Experiment One: HDR vs. LDR

It is generally taken for granted that HDR images look better than LDR ones. We put this to the test. In particular, we investigated the relative ordering of the following three imaging pipelines in terms of subjective preference of the observer:

1. HDR capture, HDR display (full HDR pipeline)
2. HDR capture, LDR display (tone mapping pipeline)
3. LDR capture, LDR display (conventional pipeline)

Although it is generally assumed that visual preference is given to the full HDR pipeline, with the tone mapping pipeline rated as second-best and the conventional imagining pipeline least desirable, we are not aware of any validation study that upholds this ordering. In fact, our results show that this ordering does not necessarily hold in reality.

3.1 Stimuli

We used images taken from 10 different scenes that represent a broad range of typical environments including outdoors day and night, landscapes, indoors, and close-up objects (Figure 1). All images were captured at a resolution of 2592 by 1944 and then down-sampled to 1296 by 972 to fit to the display resolution. This corresponds to 90% of the available display resolution; the extra space is left black due to the power consumption limits of the HDR display.

For each scene, we generated an HDR image to represent the first pipeline using the multiple exposures technique [Debevec and Malik 1997]. This involves capturing a bracketed exposure sequence where each exposure is separated by one f-stop, linearizing the exposures using the inverse of the camera response, and combining them into a single radiance map [Debevec and Malik 1997; Mitsunaga and Nayar 1999]. All exposures were captured by a Nikon

Figure 1: *The test scenes used in our experiment.*

Scene	**DR** ($\log_{10}$)	**Scene**	**DR** ($\log_{10}$)
Apple	4.64	Shop	5.78
Mexican_mug	5.36	Neckar_island	4.84
Rooftop	3.61	Rooftop2	3.62
Room	4.06	Room2	5.64
Street_lamp	6.83	Valley	3.61

Table 1: *The dynamic ranges (DR) of the HDR images (see Figure 1) used in our experiments. All images are carefully selected to have a dynamic range around 5 orders of magnitude, as this corresponds to the dynamic range of the HDR display device.*

E5400 digital camera. The scene luminances were measured by using an 18% gray card and a Photo Research PR-650 colorimeter. This data was used to reproduce the original scenes with physically correct luminances. All HDR images were created from 10 exposures and distributed around 5 orders of dynamic range (Table 1).

Images that represent the second pipeline were obtained by tone mapping the HDR images. To ensure our results generalize beyond the specific features of any single algorithm, we compared three different algorithms that performed well in previous tone mapping validation studies [Drago et al. 2002; Kuang et al. 2004; Ledda et al. 2005; Yoshida et al. 2005]. Specifically, we used the histogram adjustment technique [Ward et al. 1997], bilateral filtering [Durand and Dorsey 2002], and the photographic tone mapping operator [Reinhard et al. 2002].

For the third pipeline, we used two individual exposures from the original bracketed sequences. One was the *objective best* exposure, in the sense that it contained the smallest number of under- and over-exposed pixels. The other was the *subjective best* exposure as indicated by 20 participants in a pilot study. Specifically, the participants were asked to choose which image they preferred from the middle 5 exposures (out of 10; the other exposures were clearly under- or over-exposed). When the objective-best and the subjective-best were the same image, we included the subjective second-best to maintain the same number of images for all scenes.

3.2 Experimental Design

The experimental design consisted of a ranking study where the participants' task was to order the following six images for each scene according to their preference (the instructions given were "Please indicate which image looks best to you"):

- The HDR image
- The images tone-mapped with the histogram adjustment technique (referred to as *hist*), the bilateral filter (*bila*), and the photographic tone mapping operator (*phot*)
- Objective- (*obje*) and subjective-best (*subj*) exposures

The HDR image was reproduced with physically correct luminances by matching the gray card values measured in the original scene to that of the displayed image. Our choice of scenes was to a large extent motivated by this requirement. However, some highlights still exceeded the peak luminance of the HDR display, and thus were clipped around $3000\,cd/m^2$ (e.g. the highlights on the *apple* and *mexican_mug* images). Also, three of the scenes (*rooftop*, *rooftop2*, and *valley*) were too light and could not be reproduced with physical accuracy. These scenes were displayed with the highest possible average luminance without causing burnout in large image areas.

The LDR images were displayed such that their appearance on the HDR display is matched to their appearance on a Dell UltraSharp 2007FP 20.1" LCD monitor. To this end, the calibration properties such as peak luminance, black level, gamma, primaries and the white point of the Dell monitor were measured and simulated on the HDR monitor. In this process, we have not set the LEDs to a constant value but compressed their luminance range. This allows us to accurately simulate the Dell display, even though the characteristics of the LCD panels may be different. The simulation ensured that the same monitor is used to display all stimuli to eliminate negative effects that may occur by switching between different displays.

Each trial started with a consecutive presentation of all six images in random order, where each image was shown for 2 seconds. This was followed by a simultaneous presentation of all images on a 2×3 montage grid, which remained on the screen until the ranking of the six images was completed. The ranking was indicated by pressing the corresponding keys on the keyboard; the participants first chose the image they preferred the most, and then proceeded in order of descending preference. Selected images were grayed out for clarity. In case of a mistake, participants could reset their decisions for the current trial.

When the montage view was active the participants could recall any image and observe it in isolation at a higher resolution. This helped to identify some details which may be lost in the montage view due to down-sampling. It was also possible to rapidly switch back and forth between any two images to compare them more easily. The participants could return to the montage view at any time and continue ranking. Although there was no time limit, a trial usually did

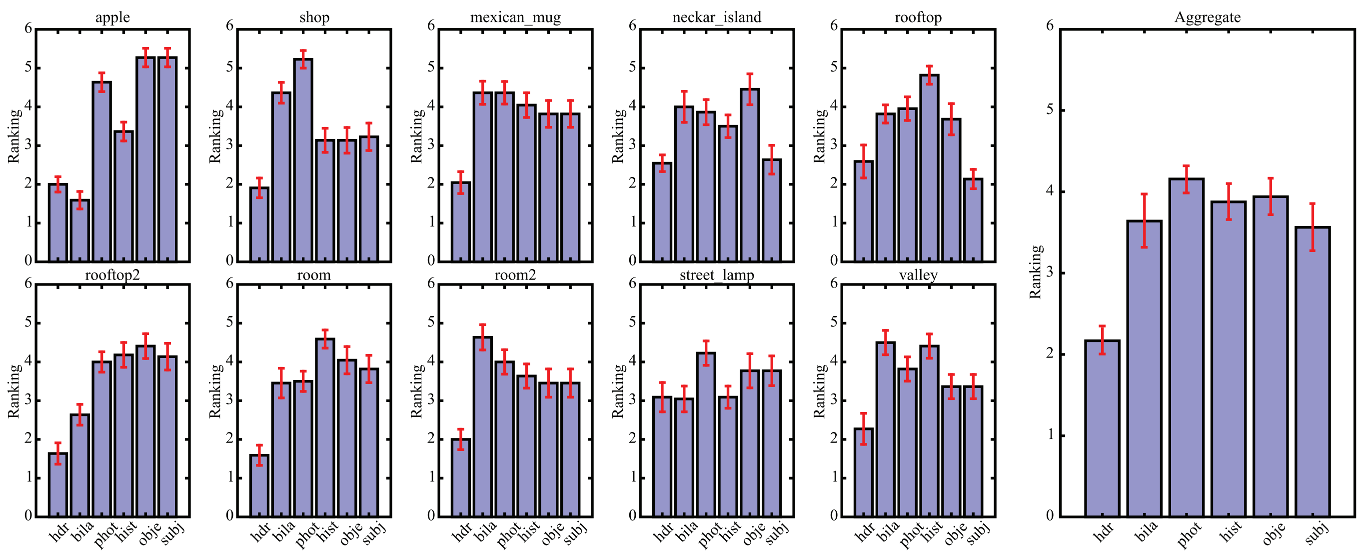

Figure 2: *Mean rankings from Experiment One.* hdr *represents the physically correct HDR image;* bila*,* phot*,* hist *represent the bilateral filtering, photographic, and histogram adjustment operators respectively, and* obje *and* subj *represent the objective- and subjective-best images. Left: rankings for the individual scenes. Right: Aggregate result across all scenes. Note that a lower ranking indicates a higher preference. Error bars denote* ± 1 *standard error.*

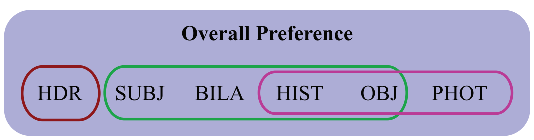

Figure 3: *The similarity groups of the first experiment as revealed by Tukey's HSD. Items in the same set are statistically indistinguishable.*

not exceed two minutes. A gender-balanced group of 22 naïve participants between 20 and 40 years old took part in the experiment.

3.3 Results

The mean rankings of all participants are shown in Figure 2. Note that a lower ranking indicates a higher preference.

There were significant overall differences between the 6 image processing pipelines, as revealed by a two-way repeated measures analysis of variance (ANOVA)[1]: $F(5,105) = 20.132, p < 0.001$. To determine which algorithms are statistically different from each other, we performed post-hoc tests using Tukey's HSD. The resulting similarity groups at the 95% significance level are shown in Figure 3.

These tests show that participants generally preferred the true HDR presentation (see also Figure 2), although there are exceptions. For instance, for two of the scenes (*neckar_island* and *rooftop*) the subjective-best single exposure performed at least as well as the full HDR image (and outperformed all the tone-mapped images). Also for the *apple* and *street_lamp* scenes the tone-mapped images rival the HDR image.

In our experiment we found that bilateral filtering in general performed significantly better than the photographic tone mapping operator (see Figure 3). However, perhaps surprisingly, there is no clear advantage of tone-mapped HDR images over the best single exposures.

[1]The dependent variable was ranking, with presentation type (6 levels) and the scene (10 levels) as two within-subject independent variables.

4 Experiment Two: Dynamic Range vs. Luminance

The results of the first experiment suggest that HDR presentation is preferred to LDR presentation in general. It also indicates that the source of LDR data, be it a tone-mapped image or a carefully selected exposure, plays a minor role in participants' preference.

One of the important questions that naturally follows is, what makes the HDR experience superior? Is it the higher dynamic range or the higher peak luminance that is simultaneously achievable by the HDR display? The answers are crucial to the design of future algorithms and display technologies because they indicate which factors are most worth investing research resources in.

Another important question is, can we rival the visual experience afforded by HDR images displayed on an HDR monitor by using LDR images? In other words, can amplifying the dynamic range of a conventional image rival the visual sensation associated with a real HDR image? Our second experiment targets these questions.

4.1 Stimuli

To evaluate the effect of dynamic range and luminance, and understand which of them plays a more important role in our appreciation of an image we used three stimuli for each scene:

1. The HDR image with physically correct luminances (referred to as *HDR*)

2. The subjective-best exposure with the same average luminance[2] (*SUBJ_AVG*)

[2]Created by matching its gray card luminance to that of the HDR image.

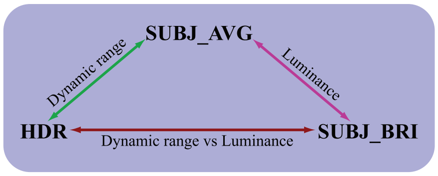

Figure 4: *Various comparisons between the three types of images are shown by the labeled arrows. SUBJ_AVG and SUBJ_BRI have the same dynamic range, SUBJ_AVG and HDR have the same mean luminance, HDR has higher dynamic range than SUBJ_BRI but has lower mean luminance.*

3. The subjective-best exposure with a greater average luminance[3] (*SUBJ_BRI*)

Note that *SUBJ_AVG* and *SUBJ_BRI* maintain their input dynamic range. This set allows various comparisons between the effects of dynamic range and luminance as shown in Figure 4. For instance, we can compare *HDR* and *SUBJ_AVG* to examine the effect of dynamic range; *SUBJ_AVG* and *SUBJ_BRI* to examine the effect of luminance; *HDR* and *SUBJ_BRI* to examine the effect of dynamic range versus luminance.

To evaluate the feasibility of inverse tone mapping, we expanded the dynamic range of the subjective-best exposure to the dynamic range of the HDR display using:

$$L' = k\left(\frac{L - L_{min}}{L_{max} - L_{min}}\right)^{\gamma}$$

where L is the luminance of the pixel being scaled, L_{min} and L_{max} are the minimum and maximum luminances of the image, k is the maximum input intensity of the HDR display, and γ determines the non-linearity of the scaling. This operation was applied to all pixels individually. Note that this operation was performed in the Yxy color space and therefore leaved the chromaticities intact.

The exponent γ determines how the mean luminance of the image will change relative to other pixels. For $\gamma = 1$ all pixels will be scaled equally, whereas for $\gamma > 1$ the mean luminance will be relatively darker and for $\gamma < 1$ it will be relatively lighter. As we did not have *a-priori* knowledge on what the correct value should be we included three natural alternatives:

4. $\gamma = 1$, i.e. linear scaling (referred to as *SUBJ_LIN*)
5. $\gamma = 0.45$, i.e. non-linear scaling with $\gamma < 1$ (*SUBJ_0.45*)
6. $\gamma = 2.2$, i.e. non-linear scaling with $\gamma > 1$ (*SUBJ_2.2*)

Figure 5 illustrates how the mean luminance is mapped by each method. In total, we had 6 stimuli per each scene.

4.2 Experimental Design

The overall design of the experiment was identical to the first experiment; a ranking procedure with 6 stimuli per scene was used. However, in addition to evaluating general preferences, we also evaluated several important visual attributes such as *naturalness*, *visual appeal*, *spaciousness*, and *visibility*. This gives us more in-depth information as to how each attribute is affected by the different presentation types. It also helps discover the relative significance of each attribute in participants' overall preference.

[3] We use four times the average luminance of the HDR image because it corresponds approximately to doubled brightness. A smaller factor induces a subtle change in brightness and a larger factor goes beyond the luminance range of the HDR display.

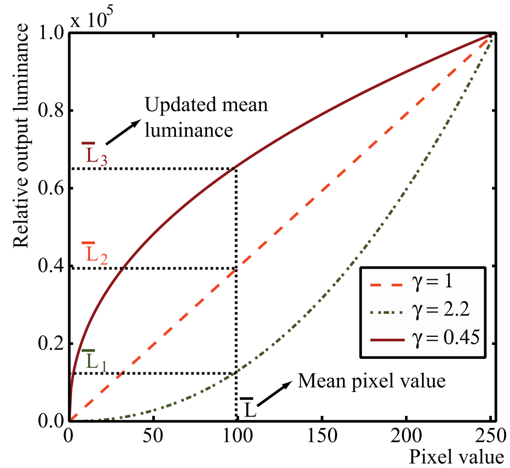

Figure 5: *Three alternatives for expanding the dynamic range of an LDR image. Each method sets the mean luminance to a different value in relation to the other pixel values.*

Attribute	F-Number
Naturalness	$F(5,75) = 40.208, p < 0.001$
Visual Appeal	$F(5,75) = 12.246, p < 0.001$
Spaciousness	$F(5,75) = 21.467, p < 0.001$
Visibility	$F(5,75) = 18.201, p < 0.001$
Overall	$F(5,75) = 13.717, p < 0.001$

Table 2: *The analysis results for the second experiment. The F-statistics show that different presentation types induce significantly different rankings.*

Thus, the second experiment was composed of five independent ranking tasks. First the individual attributes were tested in the order mentioned above, followed by an overall preference test. The experiment involved 16 participants who did not take part in Experiment 1. Gender was evenly distributed.

4.3 Results

The mean rankings for each attribute aggregated over all scenes are shown in Figure 6. The standard error of the mean is depicted by the error bars. Similar to the first experiment a shorter bar indicates a higher ranking for the corresponding display condition. As shown in the figure, the subjective-best exposure linearly scaled to the display range (*SUBJ_LIN*) is the most favored (or visually more appealing) for all attributes except naturalness. On the other hand, both of the non-linearly expanded images are favored the least for all attributes.

A two-way repeated measures ANOVA confirms that significant differences exist between different presentation types (see Table 2). The post-hoc analysis carried out with Tukey's HSD reveals the 95% statistical similarity groups as shown in Figure 7. The items enclosed in the same set are statistically indistinguishable.

Comparing the first three bars in each plot (Figure 6) shows the effect of dynamic range and luminance on visual preference. The general pattern indicates that the brighter subjective-best exposure is preferred to the HDR image, and the HDR image in turn is preferred to the subjective-best exposure with the same mean luminance. This result suggests that brightness comes first for most participants, and only when two images have the same brightness the higher dynamic range one is preferred.

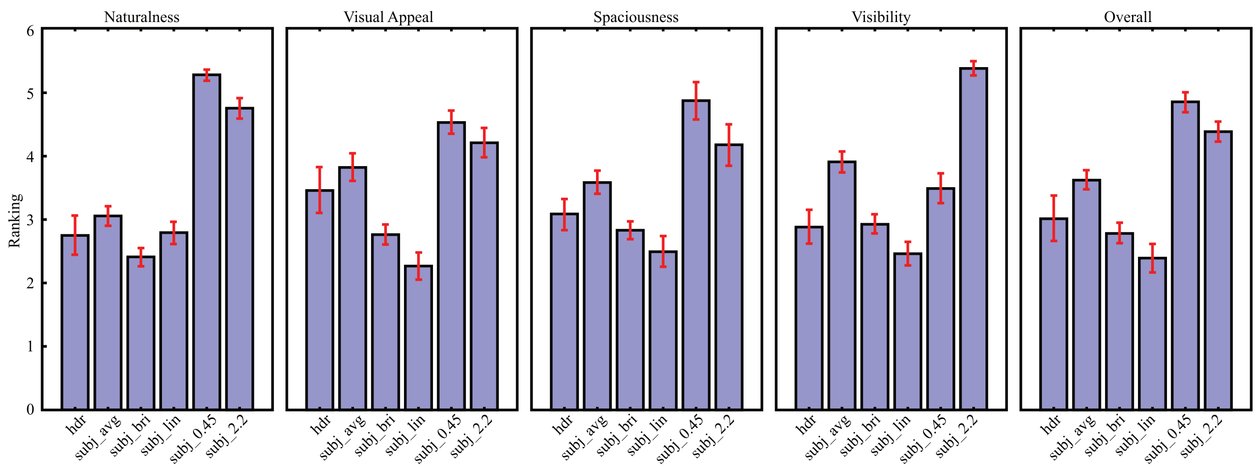

Figure 6: *Experiment Two: mean rankings for each image class, separated by visual attribute. Error bars denote standard errors.*

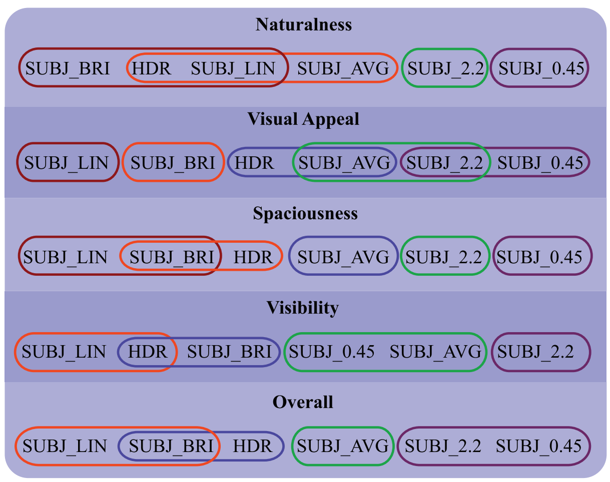

Figure 7: *The* 95% *statistical similarity groups for each attribute.*

In the same figure, the results of different ways of inverse tone mapping are shown by the three right-most bars in each plot. For all attributes, simple linear scaling produces a more favorable image compared to either of the non-linear approaches.

5 Discussion

Our first experiment confirms the general consensus that HDR monitors produce more appealing displays than conventional monitors. This may be attributed to their enhanced dynamic range and peak luminance. However, this general result is not observed for all scenes: for the *apple*, *rooftop*, and *street_lamp* the LDR versions challenge and even surpass the HDR images in participants' preference (see Figure 2).

The *rooftop* represents a particularly bright outdoors scene, whereas the *street_lamp* represents a night image with some artificial lights. As such, it is possible that the HDR versions of these images were seen as outliers by most participants and were therefore not preferred. This is important for designers of HDR content, as it suggests that the context in which a bright image or movie sequence occurs can affect its appeal.

Another interesting result of the first experiment is that tone-mapped HDR images are not found to be better than the best exposure of a bracketed sequence. It seems that although tone mapping operators preserve details and visibility in general, the overall visual quality may be compromised in the process. Some of this effect could also be due to familiarity: observers are used to seeing standard images with under- and over-exposed regions. The fact that tone mapping reduces their range significantly may cause tone-mapped images to look less natural than individual exposures. If this is true, general preference may change as HDR images become more widely experienced.

An issue that emerged in pilot work was that participants found it difficult to select the best exposure for scenes that contain several large regions of interest with very clearly different intensity levels[4]. For example, with the *rooftop2* scene, participants tended to try to optimize the appearance of either the shadowed table, or the sky and sunlit deck, even at the expense of quite unsatisfactory reproduction of the complementary region. By contrast, traditional tone mapping operators generally try to find a compromise that improves the overall visibility without favoring one region in particular. This observation could have consequences for the design of superior tone mapping algorithms.

The second experiment suggests that mean luminance generally plays a more important role than contrast in the quality of various visual attributes. This is extremely important for the design of novel displays as it means that brighter backlights may be sufficient to create a (literally) dazzling impression on consumers. However, it is clear that for many applications a combination of improved dynamic range and peak luminance is required. Further, our experiments were carried out over short exposure times. It is not clear if preference will alter over prolonged periods of exposure. This issue would require a separate set of experiments.

However, probably the most important finding from Experiment Two is how surprisingly simple it is to achieve perceptually acceptable inverse tone mapping. Despite the enormous engineering challenge posed by the problem, participants rate a trivial linear amplification of LDR pixel data as comparable with true HDR data, at least for the images we presented. This is particularly important for future HDR display designs, in that such an algorithm may be implemented in the display firmware and all LDR content is au-

[4]These regions are sometimes referred to as 'frameworks' in the psychophysics literature [Gilchrist et al. 1999].

tomatically enhanced in real time. Despite the advantages of linear scaling, however, we expect that different types of images may still require more sophisticated treatment; perhaps methods tuned to specific image content may yield superior results. It is also important to note that the LDR data used as input were high quality images: not only were they the best exposures from the bracketed sequence, they also contained no compression or visible quantization artifacts. It is likely that naïvely scaling would be less effective for highly compressed or non-optimal content, such as partially under- or over-exposed imagery.

Why do observers tolerate physically inaccurate inverse tone reproduction? One important factor that probably contributes is the fact that we are extremely insensitive to absolute units of intensity or contrast. Although we are exquisitely sensitive to intensity gradients, the overall scale of visual brightness changes by orders of magnitude depending on the prevailing illumination. Although HDR data is scene-referred, the human visual system is not. Thus, judgements of visual appeal are based on low-level attributes of an image, rather than the fidelity with which it recreates the original scene intensities.

6 Conclusions

When HDR displays reach the consumer market, two questions that will be asked are: "Do they really look any better than conventional displays?" and "What do I do with all my old (LDR) photos?" We performed two psychophysical experiments to address these pertinent questions. Our first experiment confirms that participants really do prefer HDR displays to LDR displays. We also find, perhaps surprisingly, that tone-mapped HDR images are often no better than the best single LDR exposure from a bracketed sequence. In other words, to truly benefit from the new technology you will probably need an HDR display, and not just an HDR camera.

From an engineering point of view, inverse tone mapping LDR images to recreate HDR images in scene-referred units is a difficult problem. However, our second experiment suggests that from a perceptual point of view, LDR data does not necessarily require sophisticated treatment to yield a compelling HDR experience. In fact, simple linear transformations seem to outperform other non-linear scalings that are not specifically tuned to the image content.

Acknowledgments

We'd like to thank to Reinhard Feiler for his help setting up the display; Harald Taufel for calibration, Martin Breidt, Franck Caniard, Thomas Tanner and Cengiz Terzibaş for valuable discussions; all our participants for allocating their time; and the reviewers for their valuable comments. The first author was supported in part by the National Science Foundation's funding of the "Water's Journey through the Everglades" project under grant number ESI0638977.

References

BANTERLE, F., LEDDA, P., DEBATTISTA, K., AND CHALMERS, A. 2006. Inverse tone mapping. In *GRAPHITE '06: Proc. of the 4th International Conf. on Computer Graphics and Interactive Techniques in Australasia and Southeast Asia*, 349–356.

DEBEVEC, P. E., AND MALIK, J. 1997. Recovering high dynamic range radiance maps from photographs. In *SIGGRAPH 97 Conference Proceedings*, Annual Conference Series, 369–378.

DRAGO, F., MARTENS, W. L., MYSZKOWSKI, K., AND SEIDEL, H.-P. 2002. Perceptual evaluation of tone mapping operators with regard to similarity and preference. Tech. Rep. MPI-I-2002-4-002, Max Plank Institut für Informatik.

DURAND, F., AND DORSEY, J. 2002. Fast bilateral filtering for the display of high-dynamic-range images. *ACM Transactions on Graphics 21*, 3, 257–266.

GILCHRIST, A. L., KOSSYFIDIS, C., BONATO, F., AGOSTINI, T., CATALIOTTI, J., LI, X., SPEHAR, B., ANNAN, V., AND ECONOMOU, E. 1999. An anchoring theory of lightness perception. *Psychological Review 106*, 795–834.

HOEFFLINGER, B., Ed. 2007. *High-dynamic-range (HDR) vision: Microelectronics, image processing, computer graphics*. Springer Series in Advanced Microelectronics. Springer, Berlin.

KUANG, J., YAMAGUCHI, H., JOHNSON, G. M., AND FAIRCHILD, M. D. 2004. Testing hdr image rendering algorithms. In *Proc. of IS&T/SID 12th Color Img. Conf.*, 315–320.

LEDDA, P., CHALMERS, A., TROSCIANKO, T., AND SEETZEN, H. 2005. Evaluation of tone mapping operators using a high dynamic range display. *ACM Trans. Graph. 24*, 3, 640–648.

MEYLAN, L., DALY, S., AND SÜSSTRUNK, S. 2006. The reproduction of specular highlights on high dynamic range displays. In *IS&T/SID 14th Color Imaging Conference*.

MITSUNAGA, T., AND NAYAR, S. K. 1999. Radiometric self calibration. In *Proceedings of CVPR*, vol. 2, 374–380.

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

REINHARD, E., WARD, G., PATTANAIK, S., AND DEBEVEC, P. 2005. *High Dynamic Range Imaging: Acquisition, Display and Image-Based Lighting*. Morgan Kaufmann, San Francisco.

REMPEL, A., TRENTACOSTE, M., SEETZEN, H., YOUNG, D., HEIDRICH, W., WHITEHEAD, L., AND WARD, G. 2007. Ldr2hdr: On-the-fly reverse tone mapping of legacy video and photographs. *ACM Trans. Graph. 26*, 3.

SEETZEN, H., AND WHITEHEAD, L. 2003. A high dynamic range display using low and high resolution modulators. *SID Digest*, 1450–1453.

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

SEETZEN, H., LI, H., YE, L., HEIDRICH, W., WHITEHEAD, L., AND WARD, G. 2006. Observations of luminance, contrast, and amplitude resolution of displays. In *Society for Information Display (SID)*, 1229 – 1233.

WARD, G., RUSHMEIER, H., AND PIATKO, C. 1997. A visibility matching tone reproduction operator for high dynamic range scenes. *IEEE Trans. on Visualization and Comp. Graphics 3*, 4.

YOSHIDA, A., BLANZ, V., MYSZKOWSKI, K., AND SEIDEL, H. 2005. Perceptual evaluation of tone mapping operators with real-world scenes. In *Stereoscopic Displays and Virtual Reality Systems XII. Edited by Woods, Andrew J.; Bolas, Mark T.; Merritt, John O.; McDowall, Ian E. Proceedings of the SPIE, Volume 5666, pp. 192-203 (2005).*, 192–203.

YOSHIDA, A., MANTIUK, R., MYSZKOWSKI, K., AND SEIDEL, H.-P. 2006. Analysis of reproducing real-world appearance on displays of varying dynamic range. *Computer Graphics Forum 25*, 3, 415–426.