

# Evaluation of Reverse Tone Mapping Through Varying Exposure Conditions

Belen Masia<sup>1</sup>

Sandra Agustin<sup>1</sup>

Roland W. Fleming<sup>2</sup>

Olga Sorkine<sup>3</sup>

Diego Gutierrez<sup>1,4</sup>

<sup>1</sup>Universidad de Zaragoza

<sup>2</sup>Max Planck Institute for Biological Cybernetics

<sup>3</sup>New York University

<sup>4</sup>Instituto de Investigación en Ingeniería de Aragón (I3A)

## Abstract

Most existing image content has low dynamic range (LDR), which necessitates effective methods to display such legacy content on high dynamic range (HDR) devices. Reverse tone mapping operators (rTMOs) aim to take LDR content as input and adjust the contrast intelligently to yield output that recreates the HDR experience. In this paper we show that current rTMO approaches fall short when the input image is not exposed properly. More specifically, we report a series of perceptual experiments using a Bright-side HDR display and show that, while existing rTMOs perform well for under-exposed input data, the perceived quality degrades substantially with over-exposure, to the extent that in some cases subjects prefer the LDR originals to images that have been treated with rTMOs. We show that, in these cases, a simple rTMO based on gamma expansion avoids the errors introduced by other methods, and propose a method to automatically set a suitable gamma value for each image, based on the image key and empirical data. We validate the results both by means of perceptual experiments and using a recent image quality metric, and show that this approach enhances visible details without causing artifacts in incorrectly-exposed regions. Additionally, we perform another set of experiments which suggest that spatial artifacts introduced by rTMOs are more disturbing than inaccuracies in the expanded intensities. Together, these findings suggest that when the quality of the input data is unknown, reverse tone mapping should be handled with simple, non-aggressive methods to achieve the desired effect.

**CR Categories:** I.3.7 [Computing Methodologies]: Computer Graphics—; I.4.10 [Computing Methodologies]: Image Processing and Computer Vision—Image Representation

**Keywords:** Tone management, high dynamic range imaging, image processing, perception, psychophysics, human visual system

## 1 Introduction

High dynamic range display devices are becoming increasingly common [Seetzen et al. 2004], yet very large amount of existing low dynamic range legacy content and prevalence of 8-bit photography persist. This presents us with the problem of reverse tone mapping. The aim of reverse tone mapping operators (rTMOs) is to endow low dynamic range (LDR) imagery with the appearance of a higher dynamic range without introducing objectionable artifacts. Ideally, an rTMO should take a standard LDR image as input and reconstruct as accurately as possible the true luminance values of the original scene. As depicted in Figure 1, this is an ill-posed problem. For most scenes and imaging devices, the image data

is irreversibly distorted by unknown nonlinearities, sensor noise, lens flare, blooming, and perhaps most importantly, sensor saturation, which clips high intensities to a constant value. Reverse tone mappers must somehow reconstruct the missing data, or boost the contrast in a way that does not cause the clipped regions to appear visually unpleasant.

Existing rTMOs tackle this ill-posed problem in different ways, leading them to succeed and fail in different conditions. For example, some reverse tone mapping strategies may handle small clipped highlights well, but cause large saturated regions to appear unnatural. Conversely, other rTMOs may avoid introducing artifacts in over-exposed conditions, but fail to enhance under-exposed images sufficiently. The key is to understand which strategies produce the best possible visual experience, for which a number of user studies have recently been conducted [Yoshida et al. 2006; Seetzen et al. 2006; Akyüz et al. 2007; Banterle et al. 2009b]. These experiments have yielded many valuable insights which may guide future rTMO and even HDR display design. However, they have been applied only to subjectively *correctly exposed* images, usually with knowledge of the dynamic range of the original, real-world scene. A key challenge in rTMO design is how to handle non-optimal LDR content, particularly images that are incorrectly exposed.

Our research is dedicated to finding non-intrusive ways to take advantage of the higher dynamic range of the display medium, irrespective of the dynamic range of the original image. Reverse tone mapping also sheds light on a general problem in signal processing: taking partial, distorted or corrupted data and reconstructing the original as faithfully as possible. Here our quality criterion is perceptual faithfulness rather than physical accuracy.

The vast amount of LDR legacy content spans a large range of exposures. Under- or over-exposure may be due to different reasons, including bad choices by the photographer or pure artistic intentions. Legacy professional material may have been shot to make the most appropriate use of the dynamic range available at the time, very different from what is currently available. Additionally, the information about the dynamic range of the real scene is typically not recorded. It is therefore crucial to extend previous studies by taking into consideration varying exposure conditions for a set of images without additional information.

We have performed a series of psychophysical studies assessing how rTMOs handle images across a wide range of exposure levels. We have found that, while existing rTMOs perform sufficiently well for dimmer (under-exposed) images, their performance systematically decreases for brighter (over-exposed) input images. This suggests that there is a need for an rTM method that effectively deals with over-exposed content. We show that simply boosting the dynamic range by means of an adaptive  $\gamma$  curve achieves good results that outperform the current rTMOs, and propose a simple method to obtain a suitable value of  $\gamma$  for each image.

We additionally observe that artifacts produced by some rTMOs are also visible in low dynamic range renditions of the images. This is because many artifacts are not simply due to inappropriate intensity levels, but also have a spatial component. We perform a second user study to shed light on which type of inaccuracies introduced

by reverse tone mapping most hamper our perception of the final image. This information can further help future rTMO design.

## 2 Previous Work

**Reverse tone mapping.** Dynamic range expansion, along with related subsequent problems such as contour artifacts, has been initially addressed by bit-depth extension techniques [Daly and Feng 2003] and decontouring methods [Daly and Feng 2004]. However, these techniques are designed for extension to bit-depths much lower than that of HDR displays. More recently, a few works have looked at the problem of reverse tone mapping for the display of LDR images and videos on HDR displays. The general approach of these reverse tone mapping techniques has been to identify the bright areas within the image, and in particular areas that have been clamped due to sensor saturation, such as light sources. Those areas are typically significantly expanded, while the rest is left unchanged or mildly expanded, to prevent noise amplification. We offer here a brief discussion on reverse tone mapping techniques, and refer the reader to the work by Banterle and colleagues [2009a] for a comprehensive review on the topic.

Banterle et al. [2006; 2007] apply the inverse of Reinhard’s tone mapping operator [Reinhard et al. 2002] to the LDR image and detect areas of high luminance in the resultant HDR image. They then produce a so-called *expand-map* by density estimation of the bright areas, and use this map to interpolate between the LDR image and the initial inverse tone mapped HDR image, thus modulating the expansion range. This framework has been extended to video by designing a temporally-coherent version of the expand-map [Banterle et al. 2008]. The LDR2HDR framework of Rempel et al. [2007] is similar in spirit, but their expand-map (which they term *brightness enhancement function*) can be computed in real time using the GPU. The image intensity is first linearized, and a binary mask is computed by thresholding the saturated pixels; the brightness enhancement map is computed as a blurred version of the binary mask, combined with an edge stopping function to retain contrast of prominent edges. The contrast of the LDR image is then scaled according to the enhancement map. Note that the expansion is affected by the size of the bright objects: larger objects may receive more brightness boost. Recently, Kovaleski and Oliveira [2009] presented a reverse tone mapping technique which is also based on real-time computation of a brightness enhancement function, but substitutes a bilateral filter for the combination of a Gaussian blur and an edge stopping function used by Rempel et al. [2007].

Meylan et al. [2006; 2007] explicitly focus on specular highlight detection and apply a steep linear tone mapping curve to the presumably clamped areas, whereas the rest of the image is expanded by a mild linear curve. A more sophisticated segmentation and classification of bright areas in the image is done in the work of Didyk and colleagues [2008]: they segment the bright image areas and label them as diffuse surfaces, light sources, specular highlights and reflections using a trained classifier. Different expansion functions are designed for each class to reproduce the dynamic range more accurately (in particular, the luminance of light sources and highlights is expanded more than that of reflections, while bright diffuse surfaces are not expanded). The method is suitable for high-quality video enhancement thanks to the temporal coherence of the segmentation and the expansion function. Finally, Wang et al. [2007] propose to fill in the texture information of the clamped bright areas by transferring texture from other (well exposed) areas, although the method may not be viable if a suitable region for transferring detail is not found elsewhere. Both methods [Didyk et al. 2008; Wang et al. 2007] rely on user assistance to guide the process, whereas we are interested in more automatic approaches.

**User studies.** It is now generally accepted that HDR displays provide a richer visual experience than their LDR counterparts. However, different parameters such as luminance, contrast or spatial resolution influence our visual experience, which makes it difficult to come up with an ideal combination. Additionally, image content probably also affects our preferences. In computer graphics, several researchers have performed a series of user studies, the findings of which may even influence future hardware development.

Yoshida et al. [2006] judged subjective preference (without a reference image) and fidelity (by comparing to a real world scene) for a series of tone mapped images. Users could adjust brightness, contrast and saturation for each individual image. Although their work was geared towards the design of a forward tone mapping operator, their conclusions are also useful for rTMO development: they found that, in general, brighter images were preferred over dimmer ones. Interestingly, however, in certain cases users would break this tendency and keep a significant portion of the image dark, reducing overall brightness and giving more importance to contrast.

Seetzen et al. [2006] analyzed the influence of luminance, contrast and amplitude resolution of HDR displays, to guide future display designs. Their studies show that the preferred luminance and contrast levels are related: for a given contrast, perceived image quality increases with peak luminance, reaches a maximum and then slowly decreases.

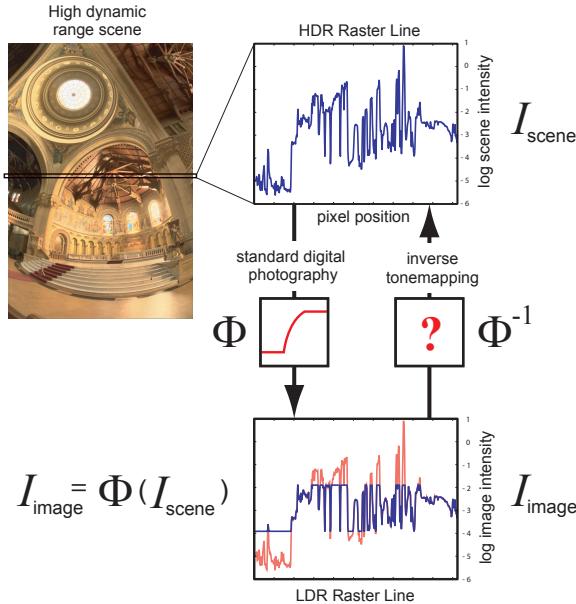
Akyüz and colleagues [2007] performed a series of psychophysical studies which revealed that a linear range expansion of the LDR image could surpass the appearance of a true HDR image, suggesting that simple solutions may suffice for reverse tone mapping. Recently, Banterle et al. [2009b] have presented a psychophysical evaluation of existing reverse tone mapping techniques, the results of which indicate that nonlinear contrast enhancement may yield better results overall.

These previous studies provide useful insight into the desirable behavior of tone mapping operators. A key difference with our work is that they were performed on correctly exposed images, whereas we are interested in analyzing reverse tone mapping across varying exposure conditions. In this work, we define over-exposed pixels as those with values  $\geq 254$ , and under-exposed pixels as those with null values [Rempel et al. 2007; Martin et al. 2008].

## 3 Experiment One: rTMO Evaluation

To assess the overall performance of an rTMO, it is important to evaluate it across a range of different imaging conditions. To this end, we have performed a user study in which subjects directly compared the output of three reverse tone mapping schemes (plus standard LDR visualization) across a range of exposures, from clearly under-exposed to clearly over-exposed images. We asked subjects to rate the appearance of the reverse tone mapped images on a calibrated Brightside DR37-P monitor (32.26" wide and 18.15" high), with a black level of  $0.015 \text{ cd/m}^2$  and a peak luminance of over  $3000 \text{ cd/m}^2$ . Calibration of the Brightside monitor was performed to confirm linearity and stable performance during the experiment and to enable comparison to specific intensities in  $\text{cd/m}^2$  should the need have arisen in the analysis, as per standard practice in psychophysics. Temperature compensation was turned off to avoid changes in intensity (this was possible thanks to the air conditioning in the room). The LDR versions of the images were displayed by approximately matching the contrast to a typical desktop TFT (Dell).

Ambient luminance was kept at about  $20 \text{ cd/m}^2$ , and the participants were seated approximately one meter away from the monitor. Based on the subjects’ ratings, we can infer which rTMOs are most



**Figure 1:** The reverse tone mapping problem. Standard imaging loses data by transforming the raw scene intensities  $I_{\text{scene}}$  through some unknown function  $\Phi$ , which clips and distorts the original scene values to create the  $I_{\text{image}}$ , shown in the bottom panel (values clipped from the original are shown in red). The goal of an rTMO is to invert  $\Phi$  to reconstruct the original scene data, or to convincingly “fake” it.

effective at recreating the experience of an HDR scene without visually objectionable side-effects. As opposed to other studies, we do not provide a ground truth HDR image for direct comparison, since it is almost always unavailable in the case of legacy content.

**Stimuli:** The stimuli consist of photographs of nine scenes with different lighting conditions, captured with a Nikon D200 at an original resolution of 3872 by 2592 (down-sampled for visualization purposes on the Brightside monitor, which has a 1920 by 1080 pixel resolution). Each scene was captured with four different exposure times. Five scenes were made up of bright images (from approximately correct exposure to clearly over-exposed), and the remaining four were made up of dark images (from clearly under-exposed to approximately correct). Figure 2 shows a representative image of each scene, while Figure 3 shows the four exposures for two example scenes. The stimuli (please refer to the supplementary material for the complete series of all the scenes) have been obtained from a previous study on exposure perception [Martin et al. 2008], where the authors analyze basic image data to try to obtain a correlation between image statistics and the perception of under- and over-exposure.

From each exposure in the bracketed sequence, we obtained three candidate renditions for display on the HDR monitor using a representative subset of reverse tone mapping algorithms: LDR2HDR [Rempel et al. 2007], Banterle’s operator [Banterle et al. 2006] and linear contrast scaling [Akyüz et al. 2007]. Except for the straightforward linear scaling (in Yxy color space, and thus performed on linearized values) we obtained the images from the authors of the original algorithms, in order to ensure accuracy in the implementation. For the LDR2HDR algorithm the parameters used were 150 pixels for the standard deviation of the large Gaussian blur applied to the mask, a brightness amplification factor  $\alpha = 4$  and a gradient image baseline width for divided differences of 5 pixels,

plus a  $9 \times 9$ -pixel kernel for the antialiasing blur and a 4-pixel radius for the *open* operator used to clean up the final edge stopping function (please refer to the original paper for a detailed explanation of these parameters). In the case of Banterle’s operator, when generating the expand-map, the parameters of the density estimation were a radius ranging from 16 to 42 pixels (smaller radius for lower exposures) and a threshold of 1 to 4 light sources (lower threshold for higher exposures), being 2048 the number of generated light sources for Median Cut sampling. In both cases, Banterle’s operator and LDR2HDR, images were linearized using gamma correction ( $\gamma = 2.2$ ). We also added a fourth LDR rendition in which the original images are presented within a luminance range matched to a typical desktop TFT monitor. The goal of this fourth image is to study whether the established assumption that visual preference is given to HDR holds over a range of exposures.

**Subjects:** A gender-balanced set of twelve subjects with normal or corrected-to-normal acuity and normal color vision were recruited to participate in the experiment. All subjects were unaware of the purpose of the study, and were unfamiliar with HDR imaging.

**Procedure:** Participants viewed the stimuli on the Brightside HDR display in a dark room. On each trial, subjects were presented with all four renditions of a given exposure of a given scene in a  $2 \times 2$  array (a *stimulus quadruple*). The positions of the four renditions within the array were random across trials, and the order of the trials was random with the constraint that consecutive trials did not present the same scene. The subjects’ task was to rate the quality of the four renditions on a scale from 1 to 7, according to how accurately the images depicted how the scene would appear to the subject if they were actually present in the scene. Thus the key criterion for comparison was the subjective fidelity of the renditions. Subjects were given unlimited time for each trial and could modify their rating of any of the renditions on a given trial before proceeding to the next trial. Additionally, they were free to assign the same values to all four renditions on a given trial, although they were instructed to try to use as much of the 1-7 scale as possible within the experiment as a whole. To aid them in setting their scale, and to accustom them to the experimental procedure, the subjects were presented with a number of practice trials before the start of the experiment.

**Results:** Several conclusions can be drawn from this test. First, for our images, there was a clear difference in perceived quality between the bright and the dark series: subjects clearly preferred the reverse tone mapped depictions of darker images over brighter ones. This can be seen in Figure 4: not only is the overall mean value significantly higher in the former case, but it is relatively stable across exposure as well. In contrast, for the bright images, there is a general downward trend in ratings across the four exposure levels.

Note that this gradual decrease in performance does not correlate with the subjective perception of quality of the original LDR image: in a previous pilot study, users picked different exposures for each series as the subjective best, not necessarily the same as the objective best (defined as the one with the smallest proportion of under- and over-exposed pixels [Akyüz et al. 2007]). The trend instead correlates with the proportion of over-exposed pixels and the mean luminance, which do increase with exposure.

Secondly, we can observe *systematic* differences between the rTMOs. On average, subjects rated the LDR2HDR and the Linear rTMOs best (the difference between the two failed to reach statistical significance), followed by the LDR images, and finally the output of the Banterle’s rTMO (see Figure 4). Pairwise Wilcoxon



**Figure 2:** Representative samples of the stimuli used in our tests. Top: bright images (Building, Lake, Graffiti, Strawberries, Sunset), each showing a certain degree of over-exposure. Bottom: dark images (Car, Flowers, Crayons, Pencils), with varying degrees of under-exposure.



**Figure 3:** The complete bracketed sequence for the Building and Flowers scenes.

(i-j)	$p_b(i,j)$	$p_d(i,j)$
LDR2HDR - Banterle's	2.0532e-21	2.8633e-7
LDR2HDR - Linear	0.5734	0.0283
LDR2HDR - LDR	1.7762e-6	1.4976e-11
Banterle's - Linear	1.1739e-22	0.0013
Banterle's - LDR	4.4489e-11	0.1938
Linear - LDR	1.4697e-7	2.0538e-6

**Table 1:** Results of the Wilcoxon rank sum tests for the bright and dark series (denoted by subindices  $b$  and  $d$  respectively). Values of  $p < 0.05$  are considered to indicate statistically significant differences between rTMOs. Thus, all differences were significant except for LDR2HDR vs. Linear in the bright series and Banterle vs. LDR in the dark series.

rank sum tests (similar to a non-parametric version of the t-test) reveal that these differences were significant to  $p < 0.05$ , except for LDR2HDR vs. Linear in the bright series and Banterle's operator vs. the LDR depiction in the dark series (see Table 1 for the complete results).

It is important to note, however, that this ordering does not hold for all conditions. For instance, the LDR depiction was systematically ranked lower than two of the rTMOs, suggesting that indeed HDR visualization is still preferred over LDR, even for under- and over-exposed images. Surprisingly, though, it ranked higher on average than Banterle's rTMO for bright images.

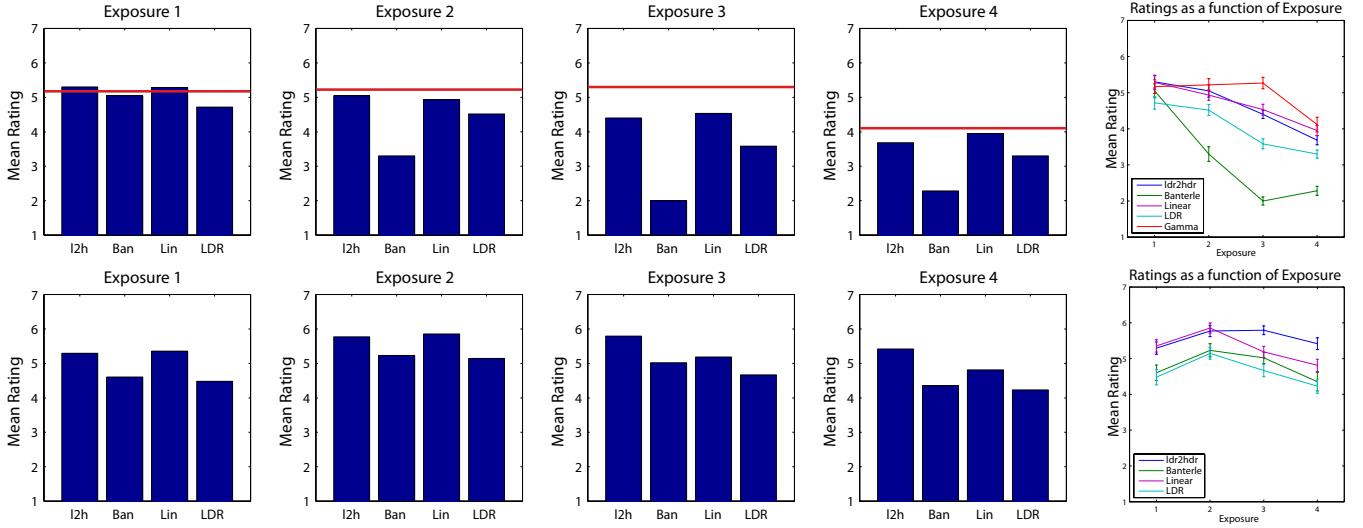
The poor overall performance of Banterle's rTMO with this data set is probably due to the fact that it often exaggerates the errors in poorly exposed images, resulting in intrusive artifacts. This becomes clear when we measure the extent to which each rTMO yields outlier rating values for each image. We calculate the median rating for each image across rTMOs. We then obtain the *outlier index* as the difference in rating for each rTMO relative to this median

value. When an rTMO is *neutral*, simply reflecting the overall quality of the exposure of the image, then the outlier index tends to be close to zero. However, when an rTMO stands out relative to the others (for example due to the introduction of artifacts), then the outlier index tends to deviate from zero. In Figure 5, we plot the histogram of the outlier index values for the three rTMOs and the LDR depiction. It is notable that for LDR2HDR, Linear and LDR, the distribution tends to be relatively tightly tuned, while for Banterle's the spread is much broader. This means that on the one hand, when it performs well, it tends to equal or exceed the others. However, it sometimes introduces substantial artifacts that cause the images to look worse than if they were not reverse tone mapped at all.

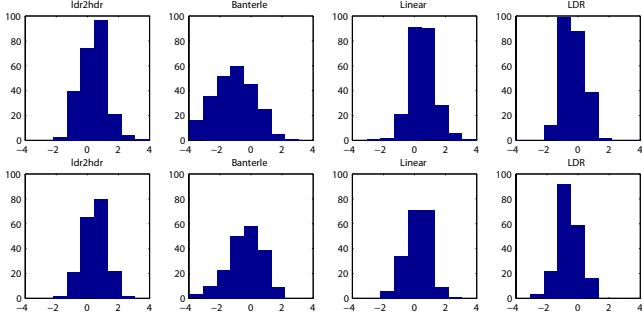
Although this seems to contradict a recent study where Banterle's operator actually outperformed other rTMOs [Banterle et al. 2009b], it is important to note that the experiments carried out in both cases differ significantly: first of all, in the work by Banterle et al. [2009b] the LDR source images were again well exposed, which is the regime within which Banterle's rTMO performs well, as we also found. However, when the source material is less flattering, we found that the algorithm sometimes produces clearly visible artifacts, which leads to lower ratings. Second, in [Banterle et al. 2009b] the authors used a 2AFC paradigm with direct ground truth comparison, whereas we propose a rating approach, which allows users to report their relative subjective preferences. Both tasks are valid ways of assessing fidelity. However, ours has the advantage that it is closer to the real usage scenario: in general the ground truth is unknown and is not presented for comparison.

## 4 Experiment Two: HDR vs. LDR Monitor

We notice that artifacts produced by LDR2HDR and Banterle's rTMOs are typically visible in low dynamic range renditions of the images. This is because they generally have a spatial component:



**Figure 4:** Top: bright images series. The blue bars represent the mean ratings across subjects for the four rTMOs (LDR2HDR, Banterle's, Linear and LDR) with increasing exposure levels (see Section 3). The last chart clearly shows the downward trend in perceived image quality. Error bars represent standard errors on the mean. The red line in the first four charts represents the mean ratings for our proposed  $\gamma$ -curve expansion (see Section 5). It can be seen that it rates generally higher and is more stable. Bottom: same information for the dark images series, showing higher overall means and a more stable perceived quality across exposures.

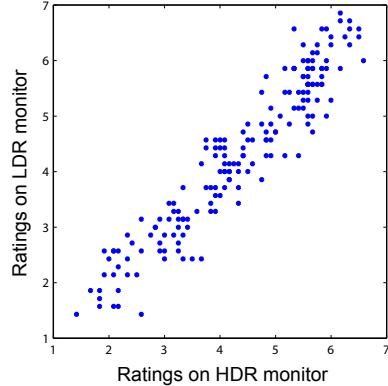


**Figure 5:** Distribution of outlier indices for all four rTMOs. Top: bright series. Bottom: dark series.

they are not simply due to inappropriate intensity levels for certain features, but they also include fringes, visibly boosted noise and other artifacts. To analyze this, we performed a second experiment with seven new subjects, which was identical to the first experiment, except that on each trial, the  $2 \times 2$  stimulus array was tone mapped using histogram adjustment [Ward et al. 1997]<sup>1</sup>. The array was then presented on a standard TFT monitor (note that this means that the LDR control condition now appears much darker than on a normal TFT).

In Figure 6, we plot the average ratings for each image in the LDR control condition against the average ratings in the HDR condition. As can be seen from the scatter plot, the ratings in the LDR control condition correlated extremely strongly with the ratings in the original experiment on the HDR monitor ( $r^2 = 0.9018$ ). We found no significant difference between bright and dark images.

This result does not imply that the images look *the same* in LDR as in HDR: the subjects were not asked to compare these conditions directly, and previous studies have confirmed that HDR depictions are preferred over LDR [Akyüz et al. 2007]. Indeed, none of the



**Figure 6:** Scatter plot showing a strong correlation between ratings on an HDR monitor and ratings when the images were tone mapped back down to LDR and presented on a standard TFT monitor.

subjects saw both renditions. However, it does demonstrate that the pattern of preferences is extremely well conserved. In other words, the images that were less preferred on the HDR monitor were also less preferred when tone mapped back down to LDR. This has two important implications. First, the strong correlation found suggests that a reasonably predictive evaluation of a rTMO could be made without directly testing on an HDR monitor. Second, as noted, the subjective ratings of HDR images that have been generated from LDR images seem to depend more on the presence or absence of disturbing spatial artifacts than on the exact intensities of different features. A similar observation (confirmed by our test) was made by Aydin et al. [2008]: they noted that the key issue in image reproduction is to accurately maintain the important features while preserving overall structure, whereas achieving an optical match becomes relatively less important. This becomes even more salient given that the dark-adaptation state of the observer is typically unknown, making absolute intensities meaningless to the user.

The design philosophy that emerges from these considerations is

<sup>1</sup>We have used the *pcond* program in *Radiance* to tone map the stimuli.

that it is generally better to apply simpler, less-aggressive rTMO schemes if the original image is imperfect. Failing to fully recreate the HDR experience is less disturbing to users than unintended artifacts that can occur when poorly-exposed images are adjusted too aggressively. In the following section we present a simple and robust approach to boosting the dynamic range of over-exposed images, and show that it is less prone to artifacts than other rTMOs.

## 5 Expanding over-exposed content

Our experiments have shown that the danger with computationally sophisticated reverse tone mapping schemes is the potential to make the image appear worse than before processing, through the introduction of objectionable artifacts. However, the goal of a rTMO is to make the image content look better in general and avoid, under any circumstances, making it look worse. Simple global reverse tone mappers, such as linear scaling and gamma boosting, never cause polarity reversals, ringing artifacts or spuriously boost regions well beyond their context. Our first experiment clearly indicates that there is room for improvement in devising an rTMO for bright input images with large saturated areas, whilst darker images turn out much better. We thus focus on the former in this section.

Examining the bright sequence in Figure 3 we observe that as exposure increases, more detail is lost as pixel values become saturated, and colors fade to white. It thus seems reasonable to attempt to depict the image in a way that the remaining details become more prominent, as opposed to boosting saturated areas as existing rTMOs do. Note that we do not aim to recover information lost to over-exposure, for which existing hallucination techniques may work [Wang et al. 2007], but rather to increase perceived quality.

We make the following key observations, which have been confirmed by previous studies on reverse tone mapping: on the one hand, darker HDR depictions are usually preferred for bright input LDR images [Meylan et al. 2006]; on the other hand, in many cases contrast enhancements improve perceived image quality [Rempel et al. 2007]. These suggest expansion of the linearized luminance values following a simple  $\gamma$  curve, which has the desired effect of darkening the overall appearance of the images while increasing contrast. Linearization of the luminance values prior to the dynamic range expansion was done with a gamma curve ( $\gamma = 2.2$ ), following the findings by Rempel et al. [2007] which note that simple gamma correction can be used for linearization instead of the inverse of the camera response without producing visible artifacts. To avoid amplifying noise, a bilateral filter [Tomasi and Manduchi 1998] can be used prior to expansion [Rempel et al. 2007]. Gamma expansion may potentially boost noise; however, over-exposed images tend to be significantly less noisy than under-exposed ones. Our psychophysical tests confirmed that noise amplification did not affect the final perceived quality.

Obviously, the problem with the proposed expansion lies in automatically obtaining an image-dependent suitable  $\gamma$  value, to avoid the cumbersome manual readjustment of the display settings for each individual image to be shown. For this, we first obtain a mea-

sure of image brightness, for which we compute its key value; this key acts as an indicator of whether the scene is subjectively dark or light. Since overall brightness can be approximated with log-luminance [Tumblin and Rushmeier 1993; Reinhard et al. 2002], we estimate the key of an image as [Akyüz and Reinhard 2006]:

$$k = \frac{\log L_{avg} - \log L_m}{\log L_M - \log L_m} \quad (1)$$

where  $\log L_{avg} = (\sum_{x,y} \log(L(x,y) + \delta))/n$ .  $L_m$  and  $L_M$  are the minimum and maximum image luminances respectively,  $n$  is the number of pixels and  $L(x,y)$  is the pixel luminance. The small offset  $\delta$  prevents singularities when  $L(x,y) = 0$ . We exclude 1% of the highest and lowest pixel values following the suggestion in [Akyüz and Reinhard 2006], to make the estimation less sensitive to outliers. We asked users in a pilot study to manually adjust the value of  $\gamma$  in a set of images, and fitted empirical data with a linear regression  $\gamma = a \cdot k + b$  (with  $a = 10.44$  and  $b = -6.282$ ), which relates  $\gamma$  as a function of the image key ( $r^2 = 0.82$ ). We have used this expression in this work to compute the reverse tone mapped results in this paper. Table 2 shows the key and  $\gamma$  values used for all the stimuli.

To provide a subjective evaluation of the performance of this strategy, we repeated Experiment One (Section 3), substituting the LDR depiction with our  $\gamma$ -expanded versions in order to maintain the  $2 \times 2$  stimulus array. The red line in Figure 4 shows the results.

Experiment One provides useful information about the *subjective* perception of image quality. However, we are also interested in evaluating our approach from an *objective* point of view. The problem is the fact that the intended comparison needs to be performed between an LDR and an HDR image. Recently, Aydin and colleagues [2008] have presented a novel image quality metric which identifies visible distortions between two images, independently of their respective dynamic ranges. The metric uses a model of the human visual system, and classifies visible changes between a reference and a test image. The authors identify three types of structural changes: loss of visible contrast (when contrast visible in the reference image becomes invisible in the second one), amplification of invisible contrast (when invisible contrast in the reference image becomes visible in the second one), and reversal of visible contrast (when contrast polarity is reversed in the second image with respect to the reference). It is important to remember that, as Rempel and colleagues noted [2007], contrast enhancement tends to increase perceived quality, and therefore is a desired outcome of the rTMO.

Figure 7 shows the results of this metric<sup>2</sup> comparing two of the original LDR images (reference images) with the corresponding outputs using linear expansion, LDR2HDR, Banterle's operator and our proposed  $\gamma$  curve. Our method reveals more detail, shows no loss of contrast and minimizes gradient reversals. Note that while our approach may fail to utilize the dynamic range to its full extent in some cases, it has the important and experimentally validated advantage of avoiding objectionable and unpredictable artifacts.

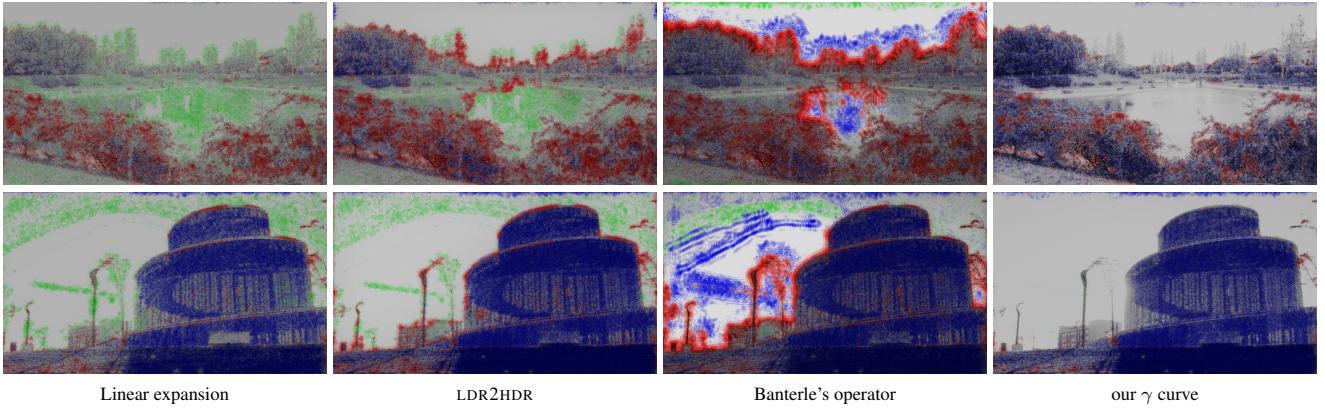
## 6 Discussion and Conclusions

Previous works on the perception of HDR images and rTM design have assumed that the input images were, in general, correctly exposed. While these provide valuable knowledge that could guide the development of both HDR display hardware and reverse tone mapping algorithms, existing LDR legacy content actually covers a wide range of exposures, including material that suffers from bad exposure. As currently designed, existing rTMOs tend to boost

	<b>1</b>	<b>2</b>	<b>3</b>	<b>4</b>
<i>Building</i>	0.697 / 1.22	0.762 / 1.5	0.816 / 1.75	0.845 / 2.6
<i>Lake</i>	0.7714 / 1.1	0.7453 / 1.2	0.7487 / 1.5	0.7830 / 2.25
<i>Graffiti</i>	0.7666 / 1.2	0.8193 / 1.35	0.8738 / 1.5	0.9184 / 1.75
<i>Strawberries</i>	0.6696 / 1.22	0.7218 / 1.35	0.7218 / 1.55	0.8479 / 1.9
<i>Sunset</i>	0.7022 / 1.1	0.8103 / 1.35	0.8016 / 1.4	0.8713 / 1.75

**Table 2:** Key and  $\gamma$  values for the five scenes and the four exposure levels.

<sup>2</sup>We have used the online implementation provided by the original authors of the paper: <http://drim.mpi-inf.mpg.de/generator.php>



**Figure 7:** Comparing the results of several rTMOs with the image quality metric from Aydin et al.[2008]. The reference LDR images are Lake (top) and Building (bottom) as depicted in Figure 2 (which correspond to the third and second exposure levels in the series. Please refer to the supplementary material for all the exposures in all the scenes). Green, blue and red identify loss of visible contrast, amplification of invisible contrast and contrast reversal respectively. Our  $\gamma$  expansion does not lose any contrast, while minimizing gradient reversals. More importantly, it reveals more detail in the most significant areas of the images (trees, grass, bushes and buildings in the images shown).

over-exposed areas more than the rest of the image. The strategy works well for small areas such as light sources or highlights if the rest of the image is correctly exposed, but no performance evaluation on generally over-exposed imagery had been performed.

Experiment One shows that performance of rTMOs decreases for input images containing a large number of over-exposed pixels, while they seem to perform significantly better for darker images. This suggests that for bright images the consensual approach of boosting bright areas could be improved. We have shown that a simple rTMO based on  $\gamma$  expansion, without the need for explicitly detecting saturated areas, outperforms existing rTMOs in these cases, and propose an empirical expression to automatically find a suitable  $\gamma$  as a function of the image’s key, without user interaction. This rTMO has the desired properties of boosting contrast and detail in non-saturated areas of the image, visually compensating for the lack of information in the saturated ones.

We have performed two validation studies, both subjective and objective. The first one has confirmed that our approach increases the perceived image quality for these kind of images. Pairwise Wilcoxon rank sum tests revealed that the differences in rating were statistically significant with respect to all other rTMOs tested. Given that it produces darker overall images with increased contrast, this result is in accordance with previous suggestions [Meylan et al. 2006; Rempel et al. 2007]. The second evaluation uses a recently published image quality metric which operates with arbitrary dynamic ranges [Aydin et al. 2008]. The metric concludes that our method reveals more detail in non-saturated areas, does not reduce contrast and shows less gradient reversals than the other rTMOs tested. Thus, the artists’ original intentions are better preserved.

In both experiments we used typical numbers of subjects for a within-subject design in psychophysics, and the results were highly coherent across subjects. In Experiment One the reported results are statistically significant to the  $p < 0.05$  level, meaning that the chances that the outcome of the pairwise comparisons would change after running more subjects from the same population is less than 5%. Indeed, for many of the results, the probability is many orders of magnitude lower than this, which implies that the qualitative pattern of the results is well conserved across subjects. Likewise, data from Experiment Two exhibit a correlation coefficient of 0.9018, notably conclusive in statistical terms.

Our findings seem to indicate that superior rTMOs should take into

account global statistics about the image, and not just individual pixel values. We have derived a simple strategy based on the key value of the images, but more sophisticated strategies could also be devised, possibly including high-level semantics.

We also ran the same expansion on the images from the dark series: as expected, we found no significant improvements over the tested rTMOs, given that our expansion is designed for bright images (please refer to the supplementary material for the complete data).

The results from our second experiment confirm that spatial artifacts are more disturbing than inaccuracy in reproduced intensity levels [Aydin et al. 2008]. We found a very strong correlation in the pattern of preferences when viewing images on HDR and LDR displays. This does not mean that the images looked the same, but it does suggest that the artifacts that emerge with poorly-exposed input images are spatial in nature and severe enough that HDR evaluation is not necessary: they can also be clearly seen in LDR.

We do not aim to create new depictions of LDR material, which would potentially interfere with the original intentions and artistic vision. Our goal is much like that of an audio mastering engineer: we wish to increase the illusion of power, presence and fidelity in the final display medium, while preserving the author’s original vision of the content. Our results complement those in the work by Akyüz et al. [2007], where the authors show that, for *correctly exposed* imagery, a simple linear expansion works well and suggest that sophisticated treatment of LDR data may not be necessary. In fact, our work is consistent with that of Akyüz et al. [2007] in the sense that our proposed  $\gamma$  curves approach linear scaling when the image is approximately correctly exposed. Together, both studies suggest that potentially complex operators might not be needed.

The conclusions drawn aim to be valuable for further development of HDR display technology, HDR imaging in general and the development of future LDR expansion algorithms in particular. However, further tests on LDR expansion are desirable. As the community investigates this issue further, this and similar studies will surely be extended and updated. Future reverse tone mapping strategies could allow the user to control dynamic range expansion based on her own preferences or intended goal. Aspects such as tonal balance or mood could potentially be independently controlled by the user, in a similar manner to existing tools for LDR or HDR images [Bae et al. 2006; Lischinski et al. 2006; Farbman et al.

2008]. Similarly, reverse tone mapping for video content is a key challenge in this field. In order to develop operators that gracefully handle changes in exposure over time, it is crucial to first understand how they fail in the static case, for which we hope this work stimulates future research.

## 7 Acknowledgements

We would like to express our gratitude to the anonymous reviewers for their valuable comments. We would also like to thank Miguel Martin for contributing the original images, the authors of the other rTMOs, especially Francesco Banterle, Allan Rempel and Wolfgang Heidrich, for running their algorithms on our images, Aydin et al. for making their quality metric publicly available, and the participants in the experiments for their time. This research has been partially funded by the Spanish Ministry of Science and Technology (TIN2007-63025) and the Gobierno de Aragón (project OTRI 2009/0411). Roland Fleming was funded by DFG FL624/1-2, by BW-FIT grant “Information at your Fingertips” and by the Max Planck Society. Olga Sorkine’s research is supported in part by an NYU URDF grant.

## References

- AKYÜZ, A. O., AND REINHARD, E. 2006. Color appearance in high dynamic range imaging. *SPIE Journal of Electronic Imaging* 15, 3.
- AKYÜZ, A. O., FLEMING, R., RIECKE, B. E., REINHARD, E., AND BÜLTHOFF, H. H. 2007. Do HDR displays support LDR content?: a psychophysical evaluation. *ACM Trans. Graph.* 26, 3, 38.
- AYDIN, T. O., MANTIUK, R., MYSZKOWSKI, K., AND SEIDEL, H.-P. 2008. Dynamic range independent image quality assessment. *ACM Trans. Graph.* 27, 3, 69.
- BAE, S., PARIS, S., AND DURAND, F. 2006. Two-scale tone management for photographic look. *ACM Trans. Graph.* 25, 3, 637–645.
- BANTERLE, F., LEDDA, P., DEBATTISTA, K., AND CHALMERS, A. 2006. Inverse tone mapping. In *GRAPHITE ’06*, ACM, New York, NY, USA, 349–356.
- BANTERLE, F., LEDDA, P., DEBATTISTA, K., CHALMERS, A., AND BLOJ, M. 2007. A framework for inverse tone mapping. *Vis. Comput.* 23, 7, 467–478.
- BANTERLE, F., LEDDA, P., DEBATTISTA, K., AND CHALMERS, A. 2008. Expanding low dynamic range videos for high dynamic range applications. In *Proceedings of the Spring Conference on Computer Graphics*, ACM, New York, NY, USA.
- BANTERLE, F., DEBATTISTA, K., ARTUSI, A., PATTANAIK, S., MYSZKOWSKI, K., LEDDA, P., BLOJ, M., AND CHALMERS, A., 2009. High dynamic range imaging and LDR expansion for generating HDR content. Annex Eurographics 2009, April.
- BANTERLE, F., LEDDA, P., DEBATTISTA, K., BLOJ, M., ARTUSI, A., AND CHALMERS, A. 2009. A psychophysical evaluation of inverse tone mapping techniques. *Computer Graphics Forum* 28, 1, 13–25.
- DALY, S., AND FENG, X. 2003. Bit-depth extension using spatiotemporal microdither based on models of the equivalent input noise of the visual system. In *Proceedings of Color Imaging VIII: Processing, Hardcopy, and Applications*, SPIE, vol. 5008, 455.
- DALY, S., AND FENG, X. 2004. Decontouring: prevention and removal of false contour artifacts. In *Proceedings of Human Vision and Electronic Imaging IX*, SPIE, vol. 5292, 130.
- DIDYK, P., MANTIUK, R., HEIN, M., AND SEIDEL, H.-P. 2008. Enhancement of bright video features for HDR displays. *Computer Graphics Forum* 27, 4, 1265–1274.
- FARBMAN, Z., FATTAL, R., LISCHINSKI, D., AND SZELISKI, R. 2008. Edge-preserving decompositions for multi-scale tone and detail manipulation. *ACM Trans. Graph.* 27, 3, 67.
- KOVALESKI, R. P., AND OLIVEIRA, M. M. 2009. High-quality brightness enhancement functions for real-time reverse tone mapping. *The Visual Computer* 25, 5-7 (April), 539–547.
- LISCHINSKI, D., FARBMAN, Z., UYTTENDAELE, M., AND SZELISKI, R. 2006. Interactive local adjustment of tonal values. *ACM Trans. Graph.* 25, 3, 646–653.
- MARTIN, M., FLEMING, R., SORKINE, O., AND GUTIERREZ, D. 2008. Understanding exposure for reverse tone mapping. In *Congreso Español de Informática Gráfica*, 189–198.
- 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*.
- MEYLAN, L., DALY, S., AND SÜSSTRUNK, S. 2007. Tone mapping for high dynamic range displays. In *Proc. IS&T/SPIE Electronic Imaging: Human Vision and Electronic Imaging XII*, vol. 6492.
- REINHARD, E., STARK, M., SHIRLEY, P., AND FERWERDA, J. 2002. Photographic tone reproduction for digital images. *ACM Trans. Graph.* 21, 3, 267–276.
- REMPEL, A. G., TRENTACOSTE, M., SEETZEN, H., YOUNG, H. D., HEIDRICH, W., WHITEHEAD, L., AND WARD, G. 2007. Ldr2Hdr: on-the-fly reverse tone mapping of legacy video and photographs. *ACM Trans. Graph.* 26, 3, 39.
- SEETZEN, H., HEIDRICH, W., STUERZLINGER, W., WARD, G., WHITEHEAD, L., TRENTACOSTE, M., GHOSH, A., AND VOROZCOVS, A. 2004. High dynamic range display systems. *ACM Trans. Graph.* 23, 3, 760–768.
- SEETZEN, H., LI, H., YE, L., WARD, G., WHITEHEAD, L., AND HEIDRICH, W. 2006. Guidelines for contrast, brightness, and amplitude resolution of displays. In *Society for Information Display (SID) Digest*, 1229–1233.
- TOMASI, C., AND MANDUCHI, R. 1998. Bilateral filtering for gray and color images. In *ICCV*, 839–846.
- TUMBLIN, J., AND RUSHMEIER, H. E. 1993. Tone reproduction for realistic images. *IEEE Computer Graphics and Applications* 13, 6, 42–48.
- WANG, L., WEI, L.-Y., ZHOU, K., GUO, B., AND SHUM, H.-Y. 2007. High dynamic range image hallucination. In *Eurographics Symposium on Rendering*, 321–326.
- WARD, G., RUSHMEIER, H., AND PIATKO, C. 1997. A visibility matching tone reproduction operator for high dynamic range scenes. *IEEE Trans. on Visualization and Computer Graphics* 3, 4, 291–306.
- 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.