NRC Publications Archive Archives des publications du CNRC

Comparing lighting quality evaluations of real scenes with those from high dynamic range and conventional images

Newsham, G. R.; Cetegen, D.; Veitch, J. A.; Whitehead, L. A.

This publication could be one of several versions: author's original, accepted manuscript or the publisher's version. /
La version de cette publication peut être l'une des suivantes : la version prépublication de l'auteur, la version
acceptée du manuscrit ou la version de l'éditeur.

For the publisher's version, please access the DOI link below. / Pour consulter la version de l'éditeur, utilisez le lien
DOI ci-dessous.

Publisher's version / Version de l'éditeur:

<https://doi.org/10.1145/1670671.1670677>

ACM Transactions on Applied Perception, 7, 2, pp. 1-25, 2010-02-01

NRC Publications Record / Notice d'Archives des publications de CNRC:

<https://nrc-publications.canada.ca/eng/view/object?id=94338995-6db1-48b6-bb25-d321b4a0d2a>

<https://publications-cnrc.canada.ca/fra/voir/objet?id=94338995-6db1-48b6-bb25-d321b4a0d2a7>

Access and use of this website and the material on it are subject to the Terms and Conditions set forth at

<https://nrc-publications.canada.ca/eng/copyright>

READ THESE TERMS AND CONDITIONS CAREFULLY BEFORE USING THIS WEBSITE.

L'accès à ce site Web et l'utilisation de son contenu sont assujettis aux conditions présentées dans le site

<https://publications-cnrc.canada.ca/fra/droits>

LISEZ CES CONDITIONS ATTENTIVEMENT AVANT D'UTILISER CE SITE WEB.

Questions? Contact the NRC Publications Archive team at

PublicationsArchive-ArchivesPublications@nrc-cnrc.gc.ca. If you wish to email the authors directly, please see the
first page of the publication for their contact information.

Vous avez des questions? Nous pouvons vous aider. Pour communiquer directement avec un auteur, consultez la
première page de la revue dans laquelle son article a été publié afin de trouver ses coordonnées. Si vous n'arrivez
pas à les repérer, communiquez avec nous à PublicationsArchive-ArchivesPublications@nrc-cnrc.gc.ca.

National Research
Council Canada

Conseil national de
recherches Canada

Canada

<http://www.nrc-cnrc.gc.ca/irc>

Comparing lighting quality evaluations of real scenes with those from high dynamic range and conventional images

NRCC-53285

Newsham, G.R; Cetegen, D.; Veitch, J.A.;
Whitehead, L.

March 2010

A version of this document is published in / Une version de ce document se trouve dans:
ACM Transactions on Applied Perception, 7, (2), pp. 1-25, February 01, 2010,
DOI: [10.1145/1670671.1670677](http://dx.doi.org/10.1145/1670671.1670677)

The material in this document is covered by the provisions of the Copyright Act, by Canadian laws, policies, regulations and international agreements. Such provisions serve to identify the information source and, in specific instances, to prohibit reproduction of materials without written permission. For more information visit <http://laws.justice.gc.ca/en/showtdm/cs/C-42>

Les renseignements dans ce document sont protégés par la Loi sur le droit d'auteur, par les lois, les politiques et les règlements du Canada et des accords internationaux. Ces dispositions permettent d'identifier la source de l'information et, dans certains cas, d'interdire la copie de documents sans permission écrite. Pour obtenir de plus amples renseignements : <http://lois.justice.gc.ca/fr/showtdm/cs/C-42>

National Research
Council Canada

Conseil national
de recherches Canada

Canada

Comparing Lighting Quality Evaluations of Real Scenes with those from High Dynamic Range and Conventional Images

GUY R. NEWSHAM , DUYGU CETEGEN AND JENNIFER A. VEITCH
National Research Council Canada – Institute for Research in
Construction
and
LORNE WHITEHEAD
University of British Columbia

Thirty-nine participants viewed six interior scenes in an office/laboratory building and rated them for brightness, uniformity, pleasantness, and glare. The scenes were viewed in three presentation modes: participants saw the real space, and images of the spaces on a 17-inch computer monitor in both conventional and high dynamic range (HDR) mode. HDR mode allowed the high range of luminances in the real scene to be accurately reproduced, with maximum luminances more than ten times higher than those in the conventional images. For those participants who saw the images before the real spaces (the most relevant order for practical applications), the HDR images were rated as significantly more realistic than the conventional images. However, this effect was limited to scenes with relatively large areas of high luminance, which in this study was represented by scenes with windows and daylight. Ratings of the HDR images were significantly related to simple photometric descriptors of the images in the expected manner: brightness and glare ratings were positively correlated with overall and elevated luminance, and non-uniformity ratings were positively correlated with luminance variability. These results suggest that for evaluations of visual appearance of interior scenes featuring large areas of high luminance, the HDR method may be used as a surrogate for experiencing a real space both for lighting quality research, and in the design process.

Categories and Subject Descriptors: I.3.3 [Computer Graphics]: Picture/Image Generation; J.5 [Arts & Humanities]: Architecture; J.4 [Social & Behavioral Sciences]: Psychology; I.6.8 [Simulation & Modeling]: Types of Simulation --- Visual.

General Terms: Design, Experimentation, Human Factors, Measurement

Additional Key Words and Phrases: high dynamic range, lighting, luminance, virtual reality

1. INTRODUCTION

In the architectural decision-making process, using still images (i.e. graphic renderings or photographs) or videos to gather data on human (client) preferences for design purposes is widely accepted. Such simulations are much cheaper and more easily manipulated than building full-scale models, and can thus more easily facilitate planning for the future [Stamps 1990; Stamps 1998; Sheppard 2005; Dockerty et al. 2006; Kemeny and Panerai 2003].

Real environments can be simulated in many different ways, and it is important to choose the correct method depending on the purpose of the study and in which terms realism is framed [Bellman and Landauer 2000]. In lighting research, visual

¹ This work was sponsored by the National Research Council Canada (NRC), Istanbul Technical University, and the NSERC/3M Industrial Chair Program of the University of British Columbia (UBC). Authors' addresses: Newsham, Cetegen, and Veitch, National Research Council Canada – Institute for Research in Construction, Ottawa, Ontario, Canada, K1A 0R6; Whithead, Department of Physics and Astronomy, University of British Columbia, Vancouver, BC, V6T 1Z1 Canada.

representations of spaces and scenes, presented in a variety of ways, have been used for evaluation purposes. Hendrick et al. [1977] conducted early work using projected images from photographic slides. Students completed semantic differential ratings of six scenes (a non-daylit conference room lit in different ways) after viewing each one for 30 seconds; they did not observe the real space. By comparing the mean ratings to those made by a different group viewing the real spaces, the authors concluded that viewing the slides was a reasonable substitute for viewing the real space, for these types of ratings. Mahdavi and Eissa [2002] recorded semantic differential ratings of three non-daylit office spaces presented in two ways in a between-subjects design: viewing the real space, and viewing high-quality computer renderings on a conventional display. The authors concluded that overall the ratings of the renderings were adequately similar to those of the real spaces.

As confidence in image evaluation has grown, several research studies in lighting have used images as their only stimuli to derive information for lighting practice. For example, Oi [2005] studied the differences among generations in evaluating an interior lighting environment by projecting rendered images on a screen with a liquid crystal projector in a dark room. Tuaycharoen and Tregenza [2005] studied the relationship between image content and discomfort glare. In this case, the images were from the viewpoint of a building occupant looking out of a window with both urban and natural views, and a projector was used to present them to participants, potentially at very high luminances. Newsham et al. [2005] used projection to present images of non-daylit office interiors in a manner that enabled participants to optimize the lit scene to their preference. Although there was no explicit comparison to real scenes, they reported good agreement between ratings of the images in this experiment, and those of similar real scenes made by different participants in another experiment.

Although these studies all provide useful information, all can be faulted in that a key physical parameter in visual perception, luminance, was not satisfactorily addressed. Some researchers appear to have given no consideration to the importance of presenting realistic luminances. In some investigations, luminances were matched to those naturally occurring but, because of the limitations of display technology, the authors limited the scenes presented. For example, Newsham et al. [2005] limited luminance to 140 cd/m^2 , adequate for scenes of non-daylit offices where no electric light sources are visible. Conventional display technologies have also limited the dynamic range of luminances and contrast ratios that could be presented: Tuaycharoen and Tregenza [2005] did reproduce realistic luminances for potentially glary daylit situations, but this involved bulky digital projectors. These projectors delivered luminances up to $150,000 \text{ cd/m}^2$, but at the low dynamic range (distinct luminance levels) of conventional equipment.

Conventional digital cameras cannot capture the full dynamic range (luminance range, and contrast ratios) of a scene in the pixels of a single exposure image. Although specific high dynamic range (HDR) cameras do exist, these cameras are very expensive. To overcome these drawbacks, multiple exposures from a conventional digital camera can be combined with software to generate an image representing a broad range of luminances [Reinhard et al. 2006; Spheron]. There are several techniques to combine the multiple exposures with the help of a specific camera response function [Mann and Picard 1995; Debevec and Malik 1997; Mitsunaga and Nayar 1999]. Spot luminance measurements can be made at the time of image capture to better calibrate the pixel luminance values. Inanici [2006] demonstrated that this technique can, if carefully deployed, produce HDR images which record pixel luminances with an average error up to 12% compared to measurements made with a particular luminance meter, over luminances up to $13,000 \text{ cd/m}^2$.

However the information in HDR images cannot be reproduced visually on current conventional computer monitors. A conventional LCD monitor can present luminances

up to $\sim 200 \text{ cd/m}^2$, with only 256 distinct luminance levels. Fortunately, there exists an HDR display technology to overcome these limitations [Seetzen et al. 2003; Whitehead et al. 2005]. In this device the uniform backlight of a conventional LCD screen is replaced by an array of white light emitting diodes (LEDs), each of which is approximately 5 mm in diameter. The image is shown on the foreground colour LCD screen as in a conventional display, but it is backlit by a low-resolution version of the same image formed on the LED array. The LED array can attain very high luminances, if required. When these two layers are combined, with correct settings of the parameters of the display, an image is presented with high luminance values and contrast ratios similar to the levels in the real environment. The HDR display in our current study could display luminances up to 4000 cd/m^2 .

Newsham et al. [2002] used an early version of this HDR display, which could present images with a maximum luminance of 1800 cd/m^2 (using a digital projector as a backlight), to present photographs of non-daylit offices. This pilot study had many limitations, but produced results that suggested that HDR images were viewed in a different way than conventional images, and justified further study. Very recently, other authors have explored how images on an LED-backlit HDR display are perceived compared to other types of presentation. Akyüz et al. [2007] presented participants with HDR images (with peak luminance of 3000 cd/m^2) and variously formulated conventional images of the same scenes on the same HDR display; participants were asked to rank-order the images for preference. HDR images had significantly higher preference overall, however, for some scenes, variations of the conventional image rivalled the HDR image. Yoshida et al. [2006] presented participants with HDR images (with peak luminance of 3000 cd/m^2) and the corresponding real scenes simultaneously; participants were asked to use brightness, contrast, and colour saturation controls to match the HDR image to the real scene. After making these adjustments, participants rated the accuracy of reproduction. Several trials were conducted in which the luminance range of the HDR display was restricted in various ways, to simulate displays with different properties, including a conventional LCD screen. Three scenes were used, one of an office desk under two different electric lighting conditions, and the third was a view to the outdoors under naturally occurring daylight. Results showed that the preferred combination of control settings was significantly affected by scene. HDR images did demonstrate significantly better perceived reproduction than conventional images.

The current study set out to test the following hypothesis: *The visual appearance of HDR images of office building interiors will be judged as more realistic than that of conventional images.*

This study complements those of Akyüz et al. [2007] and Yoshida et al. [2006], in using a broader range of outcome variables than the former, and a broader range of images than the latter. Further, whereas both of these studies compared HDR to conventional images, only the latter compared HDR directly to reality; the current study compares HDR to both conventional images and reality. Support for the hypothesis will demonstrate that the use of HDR imagery may go beyond simply producing more appealing images. If the results show that viewing HDR images is perceptually similar to viewing the equivalent real space, then HDR imaging may find application as a tool in the architectural decision-making process, in vision and lighting quality research, and in applied perception research more generally.

2. METHODS & PROCEDURES

2.1 Participants

Thirty-nine participants (19 women and 20 men, ages 18-55 years) were recruited from a temporary staffing agency. Participation was voluntary and the reward for participation was the staffing agency's normal hourly rate for a clerical job. All participants had

English as their first language and self-reported normal or corrected-to-normal vision. Two participants were recruited each day, although attendance was not perfect and nine participated alone.

2.2 Stimuli

2.2.1 Real Spaces. Six spaces were chosen in a research facility building located in Ottawa, Canada. We chose the spaces to be representative of those found in a typical office building, and to contain illumination from a variety of electric and daylight sources. Figures A1-A6 in Appendix A show images of the spaces.

2.2.2 Conventional and HDR Images. The participants saw two images of the chosen spaces, one in conventional and the other in HDR mode. All images in this experiment were taken with a Canon EOS Digital Rebel XT camera mounted on a tripod at seated head height and fitted with a wide-angle eye lens (Canon EFS10-22mm). The angle of view was 74° vertical and 97° horizontal; camera capture size was 1728 x 1152 pixels.

The conventional images were taken with the aperture priority exposure program, with a sensitivity of ISO-100. These images were saved in JPEG format. The conventional images were displayed on the HDR display but in a conventional mode, and display brightness and gamma adjustments were set individually so that luminance on the display did not exceed 200 cd/m², typical of conventional LCD displays, and that the overall appearance was a reasonable for a single conventional image, as judged by the experimenter; no standardized tone mapping was employed. We chose to do this, rather than use a conventional display for the conventional images, so that all images were the same size, any colour differences between displays would be eliminated, and to exclude any other external cues that might bias the subjects' ratings; a similar choice, for similar reasons, was made by Akyüz et al. [2007] and Yoshida et al. [2006].

To make HDR images, sets of images were captured in each space from the same location as the conventional images. The camera was connected to a computer and all image capture settings were controlled using remote shooting software. The camera was set to full manual mode and no auto-exposure features were used to ensure that the only difference between the multiple images was the exposure time; the aperture size was set to f/4. Images with exposure times from 4 to 1/4000s were taken sequentially.

To ensure the best correspondence between image luminance and real luminance, separate camera response functions were developed for spaces illuminated with daylight, and those without daylight. The number of images to combine for each HDR image was determined individually for each space, depending on the luminance variations of each scene [Debevec and Malik 1997]. A brightness histogram of each image was examined to ensure that the darkest images used had no RGB values greater than 200 and the lightest images had no RGB values less than 20, as recommended by Reinhard et al. [2006, p.146]. The individual exposures were combined into HDR images using the software Photosphere [Anywhere]. Over all images, the exposure times used in the final HDR images covered the range 2 to 1/500 seconds; the range for each individual image is shown in Appendix A. Shorter exposure times were not included because these generated Photosphere files with maximum luminances that may have been realistic, but were well beyond those that the HDR display could present, and this led to very poor luminance reproduction overall.

On the day the images were taken, luminance measurements at a number of points in the image were taken using a Gigahertz Optic LDM-9810 luminance meter. A single calibration point was chosen in each scene, for which we later would compare the HDR image luminance calculated by Photosphere to the real scene luminance. The calibration factor, the ratio of the real space luminance to the Photosphere luminance, was then applied to all pixel luminance values in the image. Then these images were displayed on the HDR display, and the display brightness and gamma adjustments were set to create the best overall match across all measurement points between the on-screen luminance

and the luminance in the real scene, while also preserving reasonable colour reproduction, as judged by the experimenter.

A comparison of the spot luminances measured in each real space at the time the images were taken, and the luminances of the HDR and conventional images as presented on-screen, is shown in Appendix A Figures A1-A6. These comparisons show clearly that in terms of luminance the HDR images very obviously come closer to the real space than the conventional image. Nevertheless, there are still differences between the real and HDR luminances, which are sometimes substantial in percentage terms. However, it is important to note that these differences might be more attributable to unavoidable measurement error in the comparisons than to differences in the luminance field itself, as detailed below:

- In spaces with daylight, luminance measurements were made both before and after taking the photographs at multiple exposures, and the average of these two measurements was used for comparisons. Although we strove to conduct this process with stable exterior luminance conditions, luminance values did change over the typical 20-minute period of these measurements.
- Repeatability in aiming the luminance meter was not perfect, as is typical in field measurements of luminance.
- The Gigahertz luminance meter yielded the mean luminance over the area of a 1° circle. When comparing to the luminances in the HDR image created by the Photosphere software, Photosphere calculated luminances over a rectangular area only, which may lead to calibration error.
- When measuring the luminance on the HDR display the 20' measurement circle of the luminance meter covered a larger representative area of the image than the 1° circle did in the real space, which could generate large comparison errors in regions of the image with large luminance variation over small areas (e.g. luminaire surfaces).
- While matching the luminance on the HDR display to real space luminance measurements, we had to compromise. If we adjusted the gamma and brightness levels to match the high luminance values, the error for points with low luminance was excessively high, and vice versa. We made a final choice based on average luminance errors, and the appearance of luminance gradients and colour, as judged by the experimenter.
- The size of an individual LED was relatively large compared to some of the features in the images, which could cause luminance “bleeding”. For example, if an LED was activated to a high level to provide a realistic luminance on a small, bright feature, this then also backlit a neighbouring darker feature. To first order this effect is counteracted in the image rendering software controlling the LCD panel, but this correction cannot be perfect due to the limited dynamic range of the LCD. This could cause local luminances to be in error, although perceptually such luminance gradients were not so apparent.

One must also be cognisant that for low luminances, small absolute differences can generate large percentage differences. For instance, for point 17 in the gym, the real space and HDR readings were 7 cd/m^2 and 10.3 cd/m^2 respectively, which gave a difference of 39%. The absolute difference of 3.3 cd/m^2 falls within the margin of error for many calibrated luminance instruments.

2.3 Dependent Measures

During the procedure described below, participants rated real scenes and images on semantic differential (adjective pair) scales. The scales rated were: dim-bright, nonuniform-uniform, unpleasant-pleasant, glaring-not glaring. These were derived from a larger set of 27 scales commonly used in lighting research [e.g. Hendrick et al. 1977], and which have often been successfully abridged [e.g. Mahdavi and Eissa 2002;

Newsham et al. 2005]. These are also similar to the rating scales used by Seetzen et al. [2006] when investigating perceived image quality on HDR displays of various properties. Prior to beginning their ratings, participants were given the dictionary definitions of these adjectives. The ratings were made on the pages of a booklet (one page per evaluation). Each scale showed the adjectives at either end of a 100 mm line. Participants placed a pencil mark on the line to indicate the strength of their opinion, and the distance along the line in mm formed the score on a scale of 0-100.

In a final set of ratings, pairs of images of the same space were displayed and the participants' task was simply to indicate which one of the images was more realistic.

2.4 Procedure

On the day of the experiment, participants were directed to the space where image evaluation would occur. They were given information about the experiment, and then asked sign a consent form.

According to random assignment, one participant evaluated the real spaces first, followed by the images, and the other evaluated the images first, followed by the real spaces. When participants evaluated the six real spaces they were viewed in a pre-determined random order. All spaces were checked before the participants visited them to ensure that the objects in the space were the same as when the images were taken. The participants sat in chairs that were located at the viewpoints from which the images had been taken, and were informed which direction to look towards, this direction being the centre of the corresponding image. The participants were asked to take the whole scene into consideration when making their judgements of the visual appearance of spaces. The participants were given 30 seconds to become accustomed to the scene and then were prompted to complete the rating scales. Meanwhile the experimenter took a luminance measurement for the calibration point in the space. When they finished, they proceeded to the next space.

Image evaluation took place at a desk in an open-plan office laboratory; Figure 1 shows this space. The experimenter ran the experiment from the other side of the partition behind the computer screen. Each participant saw all 12 images (six HDR and six conventional images) of the spaces in a random order. Conventional and HDR images of the same spaces were not shown sequentially. When rating an image, the participants sat in a chair facing the HDR display; the lighting in the laboratory provided a luminance on the furniture panel behind the display of around 100 cd/m^2 . Before the first image, and between the images, a blank white image (luminance 126 cd/m^2 , chosen to be similar to the average luminance of all images) was displayed for 30 seconds. When an image appeared on the screen, the participants had 30 seconds to become accustomed to the scene, then the experimenter invited the participants to complete the rating scales for that scene. When the participants finished the ratings, they informed the experimenter so that the experimenter could move on to the next blank screen and image. The participants were asked to keep their attention on the screen when it was blank.

Figure 1. Laboratory space for image evaluations

After these two sets of ratings, participants returned to the office laboratory for the final set of ratings. In this case, the six pairs of images, HDR and conventional, of the same space were displayed sequentially; which image of the pair was shown first was determined randomly. The participants' task was simply to indicate which one of the image pair was more realistic. Again, the screen was blank for 30 seconds before the first image was displayed. Then each image of the pair was displayed for 15 seconds, after which the judgment was made. While one participant completed this task, the other waited outside the laboratory space.

After this phase, participants were asked to provide any other comments they had on the experiment. The whole procedure took about 75 minutes.

3. RESULTS

3.1 Analysis Strategy

For the semantic differential ratings, the general experimental design was 3 (presentation type: real, HDR image, conventional image) x 6 (scenes) within-subjects, with four outcome variables. An additional between-subjects variable, presentation order (images or real scenes rated first) was also considered. Therefore, we conducted a series of multivariate analyses of variance (MANOVAs). Each analysis looked at specific contrasts on the presentation type variables, specified *a priori*: HDR vs. real; and, HDR vs. conventional. Full support for our hypothesis would be demonstrated by no significant differences in the former contrast, and significant differences in the latter, with the mean values of ratings of the HDR images falling closer to the ratings of the real spaces than the ratings of the conventional images.

For ratings of realism in the direct comparison of HDR and conventional images, we used univariate Chi-squared tests for each of the six scenes.

We also sought to correlate the ratings of the HDR images with photometric descriptors. ANOVA was used for this analysis, with the semantic differential ratings as independent outcomes, and various luminance-based measures as predictors.

3.2 Data Cleaning

Data preparation and screening was conducted using the procedures recommended by Kline [1998]. According to Kline, absolute skewness values greater than 3 and absolute

kurtosis values greater 8 indicate univariate normality problems. Only one outcome for one presentation/scene combination had a kurtosis greater than 8. Univariate outliers were identified by examining frequency distributions of standardised scores, and scores greater than 3 standard deviations from the mean were excluded from analysis. This process reduced the final sample size for analysis to 33.

3.3 Semantic Differential Ratings

The results of the initial fully-factorial MANOVA are shown in Table I, and Figure 2 shows the graph of mean ratings. Table I is divided into three sub-sections, covering the main effects of presentation type and scene, and their interaction, respectively. The first row in each sub-section shows the statistics for the overall effect over all outcomes and levels of the independent variable(s); η^2_{partial} is the proportion of variance in the outcome(s) explained by the effect. Each sub-section is further divided, showing the effect on each individual outcome measure. Finally, under each outcome in the presentation type sub-section, we show the effect of the specific presentation type contrasts (for clarity the statistics are indented). For example, in the first sub-section we see that presentation type has a significant effect on the outcome variables overall. Specifically, there are significant effects for the dim-bright, non-uniform-uniform, and unpleasant-pleasant outcomes. Focussing on the dim-bright outcome, there are significant differences between the HDR and real presentations, and between the HDR and conventional presentations, and so on.

Table I. Results of the initial fully-factorial MANOVA, empty cells indicate non-significant tests.

Effect	Outcome / Contrast	Statistical test	η^2_{partial}
Presentation		Wilks' $\lambda = .295$, $F_{8,118} = 12.42^{***}$	.46
	dim-bright	$F_{2,62} = 61.21^{***}$	.66
	HDR vs. real	$F_{1,31} = 29.55^{***}$	.49
	HDR vs. conv.	$F_{1,31} = 53.20^{***}$	.63
	non-uniform-uniform	$F_{2,62} = 11.67^{***}$	.27
	HDR vs. real	$F_{1,31} = 14.42^{***}$	.32
Presentation	unpleasant-pleasant	$F_{2,62} = 22.78^{***}$	.42
	HDR vs. real	$F_{1,31} = 36.88^{***}$	.54
	HDR vs. conv.		
	glaring-not glaring	$F_{1,31} = 6.37^*$	.17
	HDR vs. real		
	HDR vs. conv.		
Scene		Wilks' $\lambda = .216$, $F_{20,505} = 14.85^{***}$	.32
	dim-bright	$F_{5,155} = 45.34^{***}$	.59
	non-uniform-uniform	$F_{5,155} = 5.40^{***}$	.15
	unpleasant-pleasant	$F_{5,155} = 9.77^{***}$	.24
	glaring-not glaring	$F_{5,155} = 26.48^{***}$	.46
Presentation x Scene		Wilks' $\lambda = .435$, $F_{40,1166} = 7.15^{***}$	.18
	dim-bright	$F_{10,310} = 21.84^{***}$	.41
	non-uniform-uniform	$F_{10,310} = 4.95^{***}$	.14
	unpleasant-pleasant	$F_{10,310} = 7.38^{***}$	.19
	glaring-not glaring	$F_{10,310} = 7.22^{***}$	.19

* = $p \leq 0.05$, ** = $p \leq 0.01$, *** = $p \leq 0.001$

Figure 2. Mean ratings on the four semantic differential scales, by presentation type. (Note for clarity, error bars are not shown on these graphs, but standard deviations were typically half the mean values, individual standard deviation values are available from the authors on request.)

This initial analysis appears to provide general support for the hypotheses. HDR image ratings are closer to those of the real space than the conventional image ratings. However, there was a significant presentation type x scene interaction, suggesting that the effect of presentation type varies according to the scene content. Further, an exploratory three-way MANOVA with presentation order as an additional between-subjects variable indicated a significant effect of presentation type x presentation order (Wilks' $\lambda = .766$, $F_{8,114} = 2.03$, $p = 0.03$, $\eta^2_{\text{partial}} = .13$). We addressed these issues in subsequent analyses.

We repeated the fully-factorial MANOVAs with the sample split into two equal groups according to presentation order. For brevity, we do not show the statistical tables of results for this stage, but only the graphs of mean ratings, shown in Figure 3. It is clear from Figure 3 that the differences between ratings of the different presentation types was substantially suppressed for those who saw the real spaces first. One can easily imagine that seeing the real space first might create different expectations regarding the appearance of the subsequent images, and vice versa. However, it is important to recognize that in the practical application of HDR images and displays for design decision-making, observers would be evaluating images prior to the existence of the real space. Similarly, in research applications, observers would be viewing the HDR images as a surrogate for viewing a real space that they would not experience directly. Therefore, in further analysis of these results we chose to focus on the sub-sample of ratings from people who rated the images prior to seeing the real spaces as being more relevant to the application of the HDR technology.

Figure 3. Mean ratings on the four semantic differential scales, by presentation type. (a) shows data for those who saw and rated the real spaces before the images (N=16); (b) shows data for those who saw and rated the images before the real spaces (N=16).

For this sub-sample, we repeated the MANOVA analysis for each of the six scenes separately, to allow us to interpret the presentation type x scene interaction (which was also significant for this sub-sample). Results are shown in Tables II(a)-(f) and Figures 4(a)-(f).

Table II(a). Results of the MANOVA analysis, for the sub-sample who saw and rated the images first, and for Scene 1 (corridor); empty cells indicate non-significant tests.

Effect	Outcome / Contrast	Statistical test	η^2_{partial}
Wilks' $\lambda = .347$, $F_{8,58} = 5.05^{***}$	.41
	dim-bright HDR vs. real HDR vs. conv.		
	non-uniform-uniform HDR vs. real HDR vs. conv.	$F_{2,32} = 14.37^{***}$ $F_{1,16} = 23.67^{***}$	.47 .60
	unpleasant-pleasant HDR vs. real HDR vs. conv.	$F_{2,32} = 5.85^{**}$	.27
	glaring-not glaring HDR vs. real HDR vs. conv.	$F_{2,32} = 3.67^*$ $F_{1,16} = 9.98^{**}$	.19 .38

* = $p \leq 0.05$, ** = $p \leq 0.01$, *** = $p \leq 0.001$

Table II(b). Results of the MANOVA analysis, for the sub-sample who saw and rated the images first, and for Scene 2 (gym); empty cells indicate non-significant tests.

Effect	Outcome / Contrast	Statistical test	η^2_{partial}
Wilks' $\lambda = .260$, $F_{8,58} = 6.96^{***}$	.49
	dim-bright HDR vs. real HDR vs. conv.	$F_{2,32} = 28.13^{***}$ $F_{1,16} = 30.57^{***}$	.64 .66
	non-uniform-uniform HDR vs. real HDR vs. conv.	$F_{2,32} = 9.96^{***}$ $F_{1,16} = 10.40^{**}$	.38 .39
	unpleasant-pleasant HDR vs. real HDR vs. conv.	$F_{2,32} = 24.03^{***}$ $F_{1,16} = 20.54^{***}$	.60 .56
	glaring-not glaring HDR vs. real HDR vs. conv.	$F_{1,16} = 4.54^*$	.22

* = $p \leq 0.05$, ** = $p \leq 0.01$, *** = $p \leq 0.001$

Table II(c). Results of the MANOVA analysis, for the sub-sample who saw and rated the images first, and for Scene 3 (mezzanine); empty cells indicate non-significant tests.

Effect	Outcome / Contrast	Statistical test	η^2_{partial}
Presentation		Wilks' $\lambda = .340$, $F_{8,58} = 5.19^{***}$	.42
	dim-bright HDR vs. real HDR vs. conv.	$F_{2,32} = 14.85^{***}$ $F_{1,16} = 19.30^{***}$	.48 .55
	non-uniform-uniform HDR vs. real HDR vs. conv.	$F_{2,32} = 3.64^*$	.19
	unpleasant-pleasant HDR vs. real HDR vs. conv.	$F_{2,32} = 5.86^{**}$	.27
	glaring-not glaring HDR vs. real HDR vs. conv.	$F_{2,32} = 7.19^{**}$ $F_{1,16} = 5.06^*$ $F_{1,16} = 12.66^{**}$	.31 .24 .44

* = $p \leq 0.05$, ** = $p \leq 0.01$, *** = $p \leq 0.001$

Table II(d). Results of the MANOVA analysis, for the sub-sample who saw and rated the images first, and for Scene 4 (lobby); empty cells indicate non-significant tests.

Effect	Outcome / Contrast	Statistical test	η^2_{partial}
Presentation		Wilks' $\lambda = .119$, $F_{8,58} = 13.78^{***}$	.66
	dim-bright HDR vs. real HDR vs. conv.	$F_{2,32} = 50.30^{***}$ $F_{1,16} = 56.07^{***}$	.76 .78
	non-uniform-uniform HDR vs. real HDR vs. conv.	$F_{2,32} = 18.36^{***}$ $F_{1,16} = 24.16^{***}$	.53 .60
	unpleasant-pleasant HDR vs. real HDR vs. conv.	$F_{2,32} = 13.04^{***}$ $F_{1,16} = 16.90^{***}$	.45 .51
	glaring-not glaring HDR vs. real HDR vs. conv.	$F_{2,32} = 4.86^*$	.23

* = $p \leq 0.05$, ** = $p \leq 0.01$, *** = $p \leq 0.001$

Table II(e). Results of the MANOVA analysis, for the sub-sample who saw and rated the images first, and for Scene 5 (open-plan office); empty cells indicate non-significant tests.

Effect	Outcome / Contrast	Statistical test	η^2_{partial}
Wilks' $\lambda = .300$, $F_{8,54} = 5.57^{***}$	.45
	dim-bright	$F_{2,30} = 14.77^{***}$	.50
	HDR vs. real	$F_{1,15} = 23.93^{***}$	.62
	HDR vs. conv.		
	non-uniform-uniform	$F_{2,30} = 5.25^*$	.26
	HDR vs. real	$F_{1,15} = 6.91^*$	.32
	HDR vs. conv.		
	unpleasant-pleasant	$F_{2,30} = 4.97^*$	.25
	HDR vs. real	$F_{1,15} = 10.64^{**}$	.42
	HDR vs. conv.		
	glaring-not glaring	$F_{2,30} = 7.89^{**}$	.35
	HDR vs. real		
	HDR vs. conv.	$F_{1,15} = 9.65^{**}$	.39

* = $p \leq 0.05$, ** = $p \leq 0.01$, *** = $p \leq 0.001$

Table II(f). Results of the MANOVA analysis, for the sub-sample who saw and rated the images first, and for Scene 6 (staircase); empty cells indicate non-significant tests.

Effect	Outcome / Contrast	Statistical test	η^2_{partial}
Wilks' $\lambda = .113$, $F_{8,58} = 14.34^{***}$	.66
	dim-bright	$F_{2,32} = 112.4^{***}$	.88
	HDR vs. real	$F_{1,16} = 5.44^*$	.25
	HDR vs. conv.	$F_{1,16} = 157.7^{***}$	.91
	non-uniform-uniform	$F_{2,32} = 11.76^{***}$	.42
	HDR vs. real		
	HDR vs. conv.	$F_{1,16} = 19.96^{***}$	.56
	unpleasant-pleasant	$F_{2,32} = 15.12^{***}$	.49
	HDR vs. real	$F_{1,16} = 5.37^*$	.25
	HDR vs. conv.	$F_{1,16} = 28.20^{***}$	.64
	glaring-not glaring	$F_{2,32} = 5.70^{**}$	.26
	HDR vs. real		
	HDR vs. conv.	$F_{1,16} = 7.97^*$	.33

* = $p \leq 0.05$, ** = $p \leq 0.01$, *** = $p \leq 0.001$

(a)

(b)

Figure 4. Mean ratings on the four semantic differential scales, by presentation type, and for each scene. Graphs show data for those who saw and rated the images before the real spaces. Scenes: (a) corridor; (b) gym; (c) mezzanine; (d) lobby; (e) open-plan office; (f) staircase.

This analysis clearly shows that our hypothesis is strongly supported for two of the six scenes, the lobby and the staircase, but finds little support in the data from the other four scenes.

3.4 Realism Ratings

The result of the Chi-squared analysis is shown in Table III. At this point in the experiment, all participants had seen both the real scenes and the images, therefore all participants are included in this analysis. The HDR image was rated as significantly more realistic in four of the six scenes, whereas the difference was not significant for the other two scenes.

Table III. Results of the Chi-squared analysis on realism ratings, by scene.

Scene	More realistic?		χ^2
	HDR	conventional	
Corridor	21	12	2.46
Gym	21	12	2.46
Mezzanine	28	5	16.03***
Lobby	31	2	25.49***
Open-plan office	24	9	6.82**
Staircase	26	7	10.94***

* = $p \leq 0.05$, ** = $p \leq 0.01$, *** = $p \leq 0.001$

3.5 Relating Photometric Descriptors to Semantic Differential Ratings

As in the lighting quality using images research of Newsham et al. [2005], we sought to relate the ratings of the HDR images to photometric descriptors. The primary motivation was a validity check on the method: did the ratings correspond to photometric properties in an expected manner? The analysis may also provide direction to future, more specific investigations into physical correlates for lighting quality ratings. Given the considerations above, we again focussed on data from participants who saw the images before the real spaces. We conducted within-subjects analysis of variance (ANOVA) in which the ratings of brightness, uniformity, pleasantness and glare were dependent variables, and luminance-based photometric values from each of the six HDR images were the independent variables. In these analyses we scaled the independent variables according to the differences in the photometric value.

We selected photometric descriptors for each of the four dependent variables based on previous research [Newsham et al. 2005], these are defined in Table IV. For brightness and glare, we tested overall luminance and elevated luminance. For uniformity, we tested luminance variability. For pleasantness, we tested overall luminance and variability. Table IV shows the photometric descriptors for each image, derived from pictures of the screen using a calibrated Prometric video photometer; video photometer iso-plots of each image are shown in Appendix B, Figures B1-B6. Table V shows the results of the analyses, with examples shown graphically in Figure 5.

Table IV. Luminance-based photometric values from each of the six HDR images.

Predictor	Description	Concept	Corridor	Gym	Mezzanine	Lobby	Open-Plan Office	Staircase
LUM	Mean luminance (cd/m ²)	Overall luminance	46.2	39.2	59.3	202.2	41.6	458.3
MAXLUM	Maximum luminance (cd/m ²)	Elevated luminance	1040	1691	1513	2311	1094	3224
PIX>1000	Pixels in Prometric image (total 98431) >1000 cd/m ²	Elevated luminance	19	246	637	4850	26	5787
SD/LUM	Standard deviation of luminances, normalized by mean luminance	Variability	1.60	2.62	2.22	1.77	1.43	0.96
75P:25P/LUM	Ratio of luminances of 75 th to 25 th percentile pixels normalized by mean luminance	Variability	0.045	0.067	0.064	0.027	0.091	0.005

Table V. Results of the analysis on physical correlates to semantic differential ratings for the HDR images, for the sub-sample who saw and rated the images first. Polynomial contrasts are shown, empty cells indicate non-significant contrasts.

Rating	Predictor	Polynomial Component									
		Linear		Quadratic		Cubic		Quartic		Quintic	
		F-test	$\eta^2_{part.}$	F-test	$\eta^2_{part.}$	F-test	$\eta^2_{part.}$	F-test	$\eta^2_{part.}$	F-test	$\eta^2_{part.}$
dim-bright	LOG ₁₀ (LUM)	385***	.96			4.69*	.23				
	LOG ₁₀ (MAXLUM)	388***	.96			15.8***	.50	5.54*	.26		
	PIX>1000	430***	.96								
nonuni-uni	SD/LUM			7.45*	.32			8.65**	.35	22.6***	.59

	75P:25P/LUM	6.17*	.28	6.46*	.29			27.2***	.63	4.55*	.22
unple-ple	LOG ₁₀ (LUM)	17.4***	.52	9.18**	.36	7.66*	.32				
	SD/LUM	6.52*	.29			6.24*	.28	16.7***	.51	9.02**	.36
	75P:25P/LUM	16.1***	.50			5.80*	.27	11.0**	.41		
gla-notgla	LOG ₁₀ (LUM)	38.1***	.70							7.91*	.33
	LOG ₁₀ (MAXLUM)	36.5***	.70	24.2***	.60					8.60**	.35
	PIX>1000	34.9***	.69							8.83**	.36

* = $p \leq 0.05$, ** = $p \leq 0.01$, *** = $p \leq 0.001$

Figure 5. Example scatterplots of the relationships between mean ratings and photometric descriptors, for the HDR images. Graphs show data for those who saw and rated the images before the real spaces.

4. DISCUSSION

We had predicted that the visual appearance of HDR images of office building interiors would be judged as more realistic than that of conventional images, and expected that there would be no differences between ratings of real scenes and HDR images, but differences between HDR and conventional images. The analyses of the semantic differential ratings for the three presentation types suggest that our hypothesis is supported for certain scenes only. Only the results for the lobby and staircase scenes followed the expected pattern. The lobby and the staircase were also among the four scenes for which the HDR image was rated as significantly more realistic in side-by-side image comparisons, after the real spaces had been experienced. These two spaces had

substantially higher luminance than the other spaces (Appendix A and Table IV), and were the only two spaces with extensive window areas and daylight. In other words, the HDR images showed a significant reality improvement over conventional images for scenes with relatively large, high-luminance areas in which the conventional images could not reproduce the real scene luminance, the very situation for which the HDR device was designed. For scenes with only small areas of high luminance there was no apparent benefit for the HDR display. The mezzanine and open-plan office images were viewed as more realistic on the HDR display in the side-by-side comparison with the conventional display, but there were no obvious advantages in the semantic differential ratings. One explanation is that the HDR cannot faithfully reproduce small, high-luminance areas due to the size of the individual sources in the LED array. Future advances in HDR technology might mitigate this limitation. Akyüz et al. [2007] also found that some types of conventional images were rated as high or more highly than HDR images, for some scenes. However, one of the scenes for which this was true did appear to have (luminances were not reported) a large area of high luminance. Akyüz et al. [2007] also commented that some conventional images might be viewed more favourably simply because conventional images are currently more familiar to participants.

Even with present hardware, the results do show that although HDR images were not always better than conventional images when compared to reality, they were generally no worse. This is an important point when considering the use of images in the design process. One would like to think that decisions made on the basis of images prior to the existence of a real space would be the same as those that would have been made had it been possible to view the design alternatives in reality. In this context, if the prevailing (and presumably satisfactory) image presentation device is a conventional computer display, then HDR images will be at least as good, and in many important situations better, for the kind of visual judgements made in this study.

The findings related to ratings of brightness and uniformity give confidence in the overall validity of the ratings. As expected, ratings of brightness on the HDR images were strongly related to measures of overall luminance and elevated luminance. For example, there was a very strong linear effect of $\text{LOG}_{10}(\text{LUM})$, $\text{LOG}_{10}(\text{MAXLUM})$, and $\text{PIX} > 1000$, with some smaller, higher-order effects. This mirrored the findings of Newsham et al. [2005] and is logically to be expected. We tried the same predictors for ratings of glare, and again found strong linear effects along with smaller, higher-order effects.

Ratings of uniformity were related to physical measures of luminance variability. However, these relationships were complex and involved multiple polynomial effects of similar size. Newsham et al. [2005], found a significant linear effect for a measure similar to SD/LUM , whereas the linear effect in this study was not for SD/LUM , but for the related $75\text{P}:25\text{P}/\text{LUM}$. In both studies, the linear component suggested that ratings of uniformity increase as luminance variability decreases, as would be expected. It is worth noting from Figure 4 the surprising trend for the real spaces to be rated as more uniform than the images. Whereas the analysis with HDR images does suggest that perceived uniformity does relate to variability in luminance as expected, this might not be the case across presentations. Newsham et al. [2005], with projected conventional images, in addition to uniform-non-uniform, tried using two other scales that were expected to be synonyms (varied – unvaried; simple – complex), but ratings were very different. This observation raises legitimate questions about how participants interpret the semantic scales related to uniformity.

Of more interest to the lighting quality research community are ratings of pleasantness, which were related to mean luminance and luminance variability. Higher luminance was generally related to higher pleasantness, similar to Akyüz et al. [2007],

Seetzen et al. [2006], and Newsham et al. [2005]. However, both Newsham et al. and Seetzen et al. (when contrast ratios were below 7500) also found a negative quadratic component, suggesting that if luminance was too high pleasantness might suffer, an effect we did not see in the present study. The negative quadratic effect in Newsham et al. and Seetzen et al. might have been a reaction to glare, which was not separately rated in these studies. The range of luminances used in Newsham et al. [2005] was 0-140 cd/m², far below that used in the present study, but Seetzen et al., covered luminances of >6000 cd/m². Interestingly, the negative quadratic effect in Seetzen et al. was absent for a contrast ratio of 10000. Newsham et al. [2005] found that attraction increased as variability increased, whereas the present study found that pleasantness decreased as variability increased. There may be a confound here, in that in this study the scenes with highest luminance (also a predictor of pleasantness), tended to be the most photometrically uniform. Such explorations are at an early stage, and invite more dedicated work.

This study suggests that the HDR device is a promising tool for lighting research, and adds to the body of knowledge of how well HDR displays can simulate reality. Nevertheless, like most early applications of a technology, there is room for improvement in future studies that may seek to replicate and extend these findings. There were shortcomings in the current experiment that should be considered when interpreting the results. First, our results were drawn from a small set of interior scenes and should be confirmed for a larger set of interior and exterior scenes with a wider variety of luminance profiles.

As well as the potential photometry errors described above, there was one source of error that was particular to the scenes featuring daylight. The images were all taken on sunny afternoons several weeks prior to data collection with participants. The experiment was conducted during the afternoon, and the most common sky type at this location and time of year was a clear sky. Nevertheless, there were inevitable differences between the luminance conditions experienced by the participants in the real spaces, and those prevailing when the images were taken. This might serve to suppress the perceived reality of the HDR images featuring daylight.

Another issue is the exposure time to each image. In this study exposure time was limited 30 seconds due to unavoidable scheduling issues associated with other experiments that the participants were involved in on the same day. Such short exposures are common in other published research in this field, but it is possible that longer adaptation times might have yielded different results.

It is also obvious that the potential field of view for the images and the real spaces they represented was substantially different. This was a limitation of the HDR hardware available at the time of the study. Larger HDR displays are under development that might address this issue. Alternatively, future studies could exclude the laboratory surround with baffles, and position the participant's head closer to the screen, although this might be perceived as a very artificial situation, and could bias responses for other reasons.

The current HDR hardware, although capable of much higher luminances than conventional displays, was not capable of reproducing the maximum luminances present in the scenes we studied. Again, future development might raise this maximum luminance, and one would expect this would lead to greater perceived realism.

5. CONCLUSIONS

For a range of typical office interior scenes, this study demonstrated that HDR images presented on a HDR display are rated as significantly more realistic than conventional computer images. However, this was true only for images that contained relatively high areas of high luminance, which in this study was the result of windows and daylight in the scene. This suggests that for evaluations of visual appearance of this type of scene,

the HDR method may be used as a surrogate for experiencing a real space both for research, and in the design process.

ACKNOWLEDGEMENTS

We are particularly grateful for the technical support of Michelle Mossman (UBC), Helge Seetzen and David Tan (Brightside Technologies), and Greg Ward (Anywhere Software). We also thank our NRC colleague Sandra Mancini, for her help in running the experiment.

REFERENCES

- 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 Transactions on Graphics*. 26(3), Article No. 38. ANYWHERE. URL: <http://www.anywhere.com>; last accessed 2007-07-18.
- BELLMAN, K., AND LANDAUER, C. 2000. Playing in the mud: Virtual worlds are real places. *Applied Artificial Intelligence* 14, 93-123.
- DEBEVEC, P., AND MALIK, J. 1997. Recovering high dynamic range radiance maps from photographs. In *Proceedings of ACM SIGGRAPH*, Los Angeles, CA, USA, 369-378.
- DOCKERTY, T., LOVETT, A., APPLETON, K., BONE, A., AND SUENNENBERG G. 2006. Developing scenarios and visualisations to illustrate potential policy and climatic influences on future agricultural landscapes. *Agriculture, Ecosystems & Environment* 114(1), 103-120.
- HENDRICK, C., MARTYNIUK, O., SPENCER, T.J., AND FLYNN, J.E. 1977. Procedures for investigating the effect of light on impression: simulation of a real space by slides. *Environment and Behavior* 9(4), 491-510.
- INANICI, M.N. 2006. Evaluation of high dynamic range photography as a luminance data acquisition system. *Lighting Research and Technology* 38(2), 123-136.
- KEMENY, A., AND PANERAI, F. 2003. Evaluating perception in driving simulation experiments. *Trends in Cognitive Sciences* 7(1), 31-37.
- KLINE, R.B. 1998. *Principles and Practice of Structural Equation Modeling*. Guilford Press, New York, USA.
- MAHDAVI, A., AND EISSA, H. 2002. Subjective evaluation of architectural lighting via computationally rendered images. *Journal of the Illuminating Engineering Society*, 31(2), 11-20.
- MANN, S., AND PICARD, R.W. 1995. On being 'undigital' with digital cameras: extending dynamic range by combining differently exposed pictures. In *Proceedings of IS&T (Imaging Science & Technology) 46th annual conference*, Cambridge, MA, USA, 422-428.
- MITSUNAGA, T., AND NAYAR, S.K. 1999. Radiometric self calibration. In *Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, Fort Collins, CO, USA, 1, 374-380.
- NEWSHAM G.R., SEETZEN, H., VEITCH, J.A., CHAUDHURI, A., AND WHITEHEAD, L.A. 2002. Lighting quality evaluations using images on a high dynamic range display. In *Proceedings of the ARCC/EAAE Conference on Architectural Research*, Montreal, Canada, 1-9.
- NEWSHAM, G.R., RICHARDSON, C., BLANCHET, C., AND VEITCH, J.A. 2005. Lighting quality research using rendered images of offices. *Lighting Research and Technology* 37(2), 93-115.
- OI, N. 2005. The difference among generations in evaluating interior lighting environment. *Journal of Physiological Anthropology and Applied Human Science* 24(1), 87-91.
- REINHARD, E., WARD, G., PATTANAIK, S., AND DEBEVEC P. 2006. *High Dynamic Range Imaging: Acquisition, display and Image-Based Lighting*. Morgan Kaufman, Amsterdam, the Netherlands.
- SEETZEN, H., WHITEHEAD, L.A., AND WARD, G. 2003. High dynamic range display using low and high resolution modulators. In *Proceedings of the Society for Information Display International Symposium*. Baltimore, MD, USA.
- SEETZEN, H., LI, H., YE, L., HEIDRICH, W., WHITEHEAD, L.A. AND WARD, G. 2006. Observations of luminance, contrast and amplitude resolution of displays. *SID Symposium Digest of Technical Papers*, 37(2), 1229-1233.
- SHEPPARD S.R.J. 2005. Landscape visualization and climate change: the potential for influencing perceptions and behaviour. *Environmental Science and Policy* 8(6), 637-654.
- SPHERON. URL: http://www.spheron.com/en/PI_spheroCamHDR/PI_spheroCamHDR.php; last accessed 2007-07-18.
- STAMPS, A.E. 1990. *Use of photographs to simulate environments: a meta-analysis*. *Perceptual and Motor Skills* 71, 907-913.
- STAMPS, A.E. 1998. Measures of architectural mass: from vague impressions to definite design features. *Environment and Planning B: Planning and Design* 25(6), 825-836.
- TUAYCHAROEN, N., AND TREGENZA, P.R. 2005. Discomfort glare from interesting images. *Lighting Research and Technology* 37(4), 329-341.
- WHITEHEAD, L., WARD, G., STUERZLINGER, W., AND SEETZEN, H. 2005. High dynamic range display devices. *U.S. Patent* No. 6,891,672.
- YOSHIDA, A., MANTIUK, R., MYSZKOWSKI, K., AND SEIDEL, H.-P. 2006. Analysis of reproducing real-world appearance on displays of varying dynamic range. *Proceedings of Eurographics 2006*, Vienna, Austria. 25(3), 415-426.

Appendix A

Figure A1. Measurement points in the corridor, and the associated luminance measurements in the real space and when displayed in HDR and conventional mode. Exposures used in HDR image were 1 – 1/500 s.

	cd/m2			% difference between real and .HDR
Point		on HDR display		
	real	.HDR	.JPEG	
1	725	665	157	8.3
2	680	730	161.5	-7.4
3	875	940	161	-7.4
4	673	700	157	-4.0
5	24.6	23	26	6.5
7	27	26.7	30	1.1
8	17.9	20	22.8	-11.7
12	54.9	31	37.7	43.5
13	24.5	29.8	32	-21.6
14	29.6	44.8	43.3	-51.4
15	25.2	27.4	32.3	-8.7
16	25.4	22.4	26	11.8
19	30	57.6	65	-89.5
20	31	29.6	35.4	4.5
21	24.5	23.4	29	4.5
22	41.7	41.6	49.2	0.2
23	34.4	24.9	30	27.6
24	33.1	21	24.9	36.6
25	12.3	9.1	9	26.0
27	22.90	22.7	26.5	0.9

Figure A2. Measurement points in the gym, and the associated luminance measurements in the real space and when displayed in HDR and conventional mode. Exposures used in HDR image were 1 – 1/500 s.

Point	cd/m2 on HDR display			% difference between real and .HDR
	real	.HDR	.JPEG	
2	28	31	33	-10.8
4	702	620	109	11.7
5	42	59	46.7	-39.9
6	15	17	14	-15.6
7	2075	991	148	52.2
8	2230	920	119	58.7
11	49	44.4	48.5	9.7
12	27	20.5	21	24.6
13	22	29	28	-31.2
14	44	47.9	48	-9.9
15	112	131	100	-17.0
16	54	55.8	56	-3.9
17	7	10.3	10	-38.8
18	7	9.2	8	-23.2
19	10	13.4	12.6	-40.1
20	10	12	11	-20.3
21	13	17	16	-28.2
22	9	7.8	7	16.3
24	509	486	154	4.4
25	463	445	133	3.8
26	104	137	109	-32.2

Figure A3. Measurement points in the mezzanine, and the associated luminance measurements in the real space and when displayed in HDR and conventional mode. Exposures used in HDR image were 1 – 1/500 s.

Point	cd/m2			% difference between real and .HDR
	real	.HDR	.JPEG	
1	200	196	120	2.0
2	1975	1650	174	16.5
3	381	393	144	-3.1
6	6430	1030	149	84.0
9	6200	1455	158	76.5
10	144	305	156	-111.8
11	76	130	111	-71.1
13	126	117	90	7.1
14	72	100	92	-38.9
15	133	174	105	-30.8
17	40.4	70	66	-73.3
18	21.7	35	45	-61.3
19	18.3	22	30	-20.2
20	13.85	18.7	23	-35.0
21	26.2	26.7	34	-1.9
22	15.85	24	28	-51.4

Figure A4. Measurement points in the lobby, and the associated luminance measurements in the real space and when displayed in HDR and conventional mode. Exposures used in HDR image were 1/4 – 1/500 s.

Point	cd/m2 on HDR display			% difference between real and .HDR
	real	.HDR	.JPEG	
1	1822	1200	140	34.1
2	2236	1546	127	30.9
3	2275	1503	130	33.9
4	2350	1598	167	32.0
5	2735	2465	164	9.9
6	1142	1165	134	-2.0
7	2575	2430	153	5.6
8	568	603	112	-6.2
9	363	369	85	-1.8
10	221	207	48	6.3
11	254	276	67	-8.9
12	47	67	21	-41.8
13	178	217.8	63.5	-22.6
14	61	85.2	20.8	-39.0
15	86	89	17	-3.2
16	65	96	31	-47.4
17	81	57	13	29.3
18	338	251	65	25.6
20	114	112	23	1.3
21	489	458	118	6.2
22	98	98	21	0.5

Figure A5. Measurement points in the open-plan office, and the associated luminance measurements in the real space and when displayed in HDR and conventional mode. Exposures used in HDR image were 2 – 1/500 s.

	cd/m2			% difference between real and .HDR
Point		on HDR display		
	real	.HDR	.JPEG	
1	65	70	65	-7.9
2	202	188	130	6.7
3	40	59	38	-48.4
4	1733	804	116	53.6
5	22	37	18.8	-66.6
6	24	27	18	-11.0
7	213	200	144	5.9
8	81	93	94	-14.3
9	26	24	17	7.2
11	8	9	3.8	-6.8
12	25	25	15.3	-0.5
13	26	20	14.6	24.0
14	31	26	20	16.4
15	7	10	3.4	-36.2
16	106	102	101	3.7
17	11	22	16	-109.5
18	54	56	57	-4.3
19	79	59	61	24.9
20	3	3.4	1.3	-18.9
21	12	8	2.2	30.4
22	4	4	3.8	7.0
25	180	180	133	0.1

Figure A6. Measurement points in the staircase, and the associated luminance measurements in the real space and when displayed in HDR and conventional mode. Exposures used in HDR image were 1/15 – 1/500 s.

Point	cd/m2			% difference between real and .HDR
	real	.HDR	.JPEG	
1	405	377	21	6.9
2	562	528	41	6.0
3	538	647	60	-20.4
4	519	522	55	-0.6
5	500	516	39	-3.2
6	933	964	96	-3.4
7	1275	1057	113	17.1
10	1300	923	100	29.0
11	698	640	74	8.2
12	9375	2571	180.2	72.6
13	1000	808	96	19.2
14	128	174	17	-36.5
15	117	127	12	-8.5
17	987	682	63	30.9
18	287	353	33	-23.2
19	164	190	15	-15.9
20	600	569	47	5.2
21	165	196	13	-19.1
22	707	597	49	15.6
26	1435	980	93	31.7

Appendix B

Figure B1. Luminance (cd/m^2) $\log_{10}$ iso-plot of the HDR image of the corridor, as displayed on the HDR display, taken with a video photometer.

Figure B2. Luminance (cd/m^2) $\log_{10}$ iso-plot of the HDR image of the gym, as displayed on the HDR display, taken with a video photometer.

Figure B3. Luminance (cd/m^2) $\log_{10}$ iso-plot of the HDR image of the mezzanine, as displayed on the HDR display, taken with a video photometer.

Figure B4. Luminance (cd/m²) log₁₀ iso-plot of the HDR image of the lobby, as displayed on the HDR display, taken with a video photometer.

Figure B5. Luminance (cd/m²) log₁₀ iso-plot of the HDR image of the open-plan office, as displayed on the HDR display, taken with a video photometer.

Figure B6. Luminance (cd/m²) log₁₀ iso-plot of the HDR image of the staircase, as displayed on the HDR display, taken with a video photometer.