

Visualizing lighting with images : converging between the predictive value of renderings and photographs

Citation for published version (APA):

Engelke, U., Stokkermans, M. G. M., & Murdoch, M. J. (2013). Visualizing lighting with images : converging between the predictive value of renderings and photographs. In B. E. Rogowitz, T. N. Pappas, & H. Ridder, de (Eds.), *Proceedings of Human Vision and Electronic Imaging XVIII, 86510L (March 14, 2013), Burlingame, California, USA* (pp. L1-L10). (Proceedings of SPIE; Vol. 8651). SPIE. <https://doi.org/10.1117/12.2008465>

DOI:

[10.1117/12.2008465](https://doi.org/10.1117/12.2008465)

Document status and date:

Published: 01/01/2013

Document Version:

Publisher's PDF, also known as Version of Record (includes final page, issue and volume numbers)

Please check the document version of this publication:

- A submitted manuscript is the version of the article upon submission and before peer-review. There can be important differences between the submitted version and the official published version of record. People interested in the research are advised to contact the author for the final version of the publication, or visit the DOI to the publisher's website.
- The final author version and the galley proof are versions of the publication after peer review.
- The final published version features the final layout of the paper including the volume, issue and page numbers.

[Link to publication](#)

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal.

If the publication is distributed under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license above, please follow below link for the End User Agreement:

www.tue.nl/taverne

Take down policy

If you believe that this document breaches copyright please contact us at:

openaccess@tue.nl

providing details and we will investigate your claim.

Visualizing Lighting with Images: Converging Between the Predictive Value of Renderings and Photographs

Ulrich Engelke^{*}, Mariska G. M. Stokkermans, and Michael J. Murdoch
Philips Research, High Tech Campus 34, Eindhoven, The Netherlands

Performing psychophysical experiments to investigate lighting perception can be expensive and time consuming if complex lighting systems need to be implemented. In this paper, display-based experiments are explored as a cost effective and less time consuming alternative to real-world experiments. The aim of this work is to better understand the upper limit of prediction accuracy that can be achieved when presenting an image on a display rather than the real-world scene. We compare the predictive value of photographs and physically-based renderings on a number of perceptual lighting attributes. It is shown that the photographs convey statistically the same lighting perception as in a real-world scenario. Initial renderings have an inferior performance, but are shown to converge towards the performance of the photographs through iterative improvements.

Keywords: Image-based lighting prediction, rendering, photograph, psychophysical experiments, visual perception.

1. INTRODUCTION

Lighting systems and luminaire concepts are typically designed and evaluated based on real-world prototypes. These prototypes are usually expensive and time consuming to build. Often, they cannot be installed because the building that the system is intended for is not finished yet. Computer-based renderings are typically used to circumvent these shortcomings. The resulting renderings are used for technology development and concept communication of real-world lighting systems and luminaire designs. They can also be used for perceptual testing, meaning that renderings are used in psychophysical experiments instead of real-world lighting systems. In this way, experiments can be easily set up and reproduced across laboratories worldwide.

Used for this purpose, the success of a rendering depends on its ability to predict human perception and experience of the real-world scene. In previous work we showed that the predictive value of renderings depends on a multitude of factors related to the 3D modeling, tone-mapping, and display of the rendered images.¹ In the context of indoor office lighting, we showed that the renderings predict a real lighting system with reasonable accuracy on a number of perceptual lighting attributes. Still, a considerable margin of prediction error remained even after identifying the most suitable rendering pipeline. One could argue that the remaining prediction error is due to certain factors not being fully understood and validated yet. However, the error could also simply be due to the fact that we are representing a 3D real-world scene on a 2D display, thus losing one dimension that might be instrumental in judging lighting perception.

The aim of this work is to better understand the upper limit of prediction accuracy of lighting perception that can be achieved with images on a display compared to the real-world scene. We consider here photographs to be a 2D ground truth due to their geometrical accuracy. We therefore took photographs and created renderings of our lighting laboratory. We then performed psychophysical experiments in which people rated their lighting perception in the real room as well as on a display viewing the photographs and the renderings. The outcomes of the experiments are used to iteratively improve the renderings and validate them until converged to the predictive value of the photographs.

This paper is organized as follows. In Section 2, we will explain the lighting conditions and the real-world experiment that we performed in our lighting laboratory. In Section 3 we then discuss and analyze in detail three display experiments in which people rated both photographs and several iterations of renderings of the same conditions as in the real laboratory. In Section 4 we discuss the outcomes and in Section 5 we draw conclusions.

Further author information: (Send correspondence to U.E.)
U.E.: E-mail: ulrich.engelke@philips.com, Telephone: +31-40-27 48160

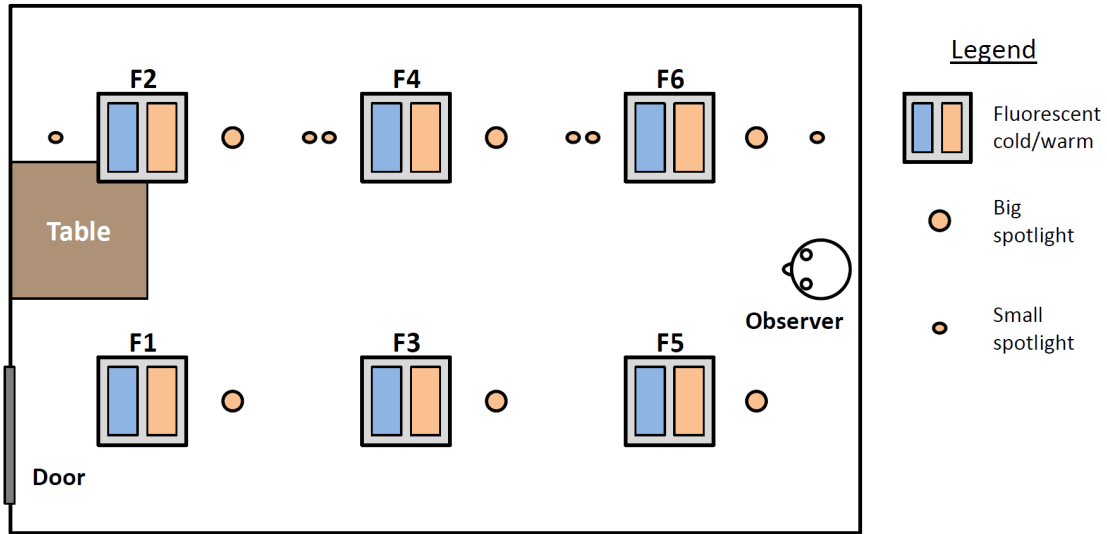


Figure 1. Overview of the laboratory layout.

Table 1. Overview of the lighting conditions used in the psychophysical experiments (black dash: luminaire turned off).

Condition number	Intensity of enabled luminaires							Color temperature	
	F1	F2	F3	F4	F5	F6	Small spots	Big spots	
1	-	-	Full	Full	-	-	-	-	2700K
2	Full	-	-	Full	Full	-	-	-	2700K
3	Full	Full	-	-	Full	Full	-	-	2700K
4	Full	Full	Full	Full	Full	Full	-	-	2700K
5	-	-	Full	Full	-	-	-	-	6500K
6	Full	-	-	Full	Full	-	-	-	6500K
7	Full	Full	-	-	Full	Full	-	-	6500K
8	Full	Full	Full	Full	Full	Full	-	-	6500K
9	Half	Half	Half	Half	Half	Half	-	-	2700K
10	Min	Min	Min	Min	Min	Min	-	-	2700K
11	Half	Half	Half	Half	Half	Half	-	-	6500K
12	Min	Min	Min	Min	Min	Min	-	-	6500K
13	-	-	-	-	-	-	Full	-	3000K
14	-	-	-	-	-	-	-	Full	3000K
15	-	-	-	-	-	-	-	Min	3000K

2. BASELINE EXPERIMENT

We performed a real-world experiment to obtain a ground truth on lighting perception with respect to several lighting attributes. The experiment is in the remainder of the paper referred to as the ‘baseline’ experiment. In the following, we explain the lighting conditions in the laboratory, the experiment procedures, and the results of the experiment.

2.1 Office Lighting Conditions

The baseline experiment took place in one of our lighting laboratories at Philips Research. An overview of the laboratory layout is shown in Fig. 1. The laboratory was empty apart from one table and a seat for the participant to sit on. The observer was sitting at one end of the room with the back close to the wall, viewing

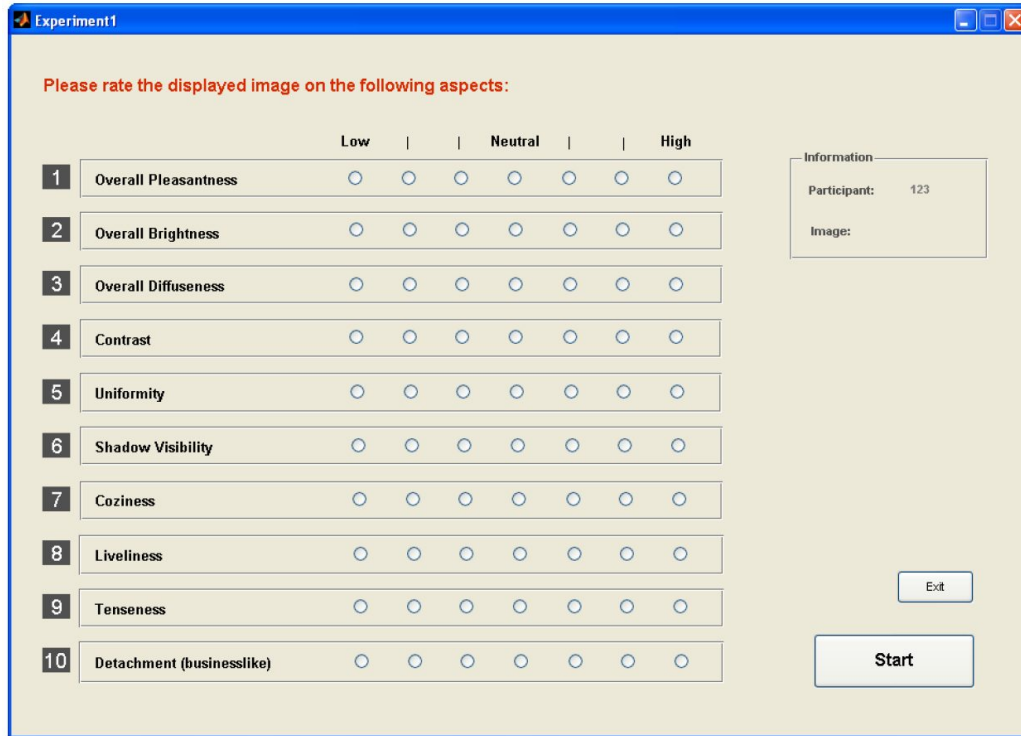


Figure 2. Graphical user interface used in the real room experiments as well as in the display experiments.

into the room towards the door and table. The ceiling was equipped with 6 fluorescent luminaires (60×60 cm), 6 big spotlights facing downwards, and 6 small spotlights facing the right wall from the observer's perspective.

The luminaires were individually addressable using DALI lighting control.² The fluorescent luminaires were changed between three intensities (Full, Half, Min) and two color temperatures (2700K or 6500K). The big spotlights were altered between two intensities (Full, Min) and the small spotlights were operated at full intensity. All spotlights were operated at a color temperature of 3000K. Given the large number of possible conditions that could be created using these settings, we necessarily defined an incomplete design with a total of 15 different lighting conditions, which is summarized in Table 1.

2.2 Experiment Procedures

The participants were presented the 15 lighting conditions in random order. During the transition between any two conditions, the room turned fully dark. The participants were asked to rate each of the conditions with respect to 10 attributes on the atmosphere and appearance of the light in the room: 'Pleasantness', 'Brightness', 'Diffuseness', 'Contrast', 'Uniformity', 'Shadow Visibility', 'Coziness', 'Liveliness', 'Tenseness', and 'Detachment'. Each attribute was rated on a 7-point scale ranging from 'Low' to 'High' using the interface shown in Fig. 2. No time limitations were imposed on the rating process.

2.3 Experiment Results

A total of 28 participants took part in the baseline experiment. In fact, the experiment consists of the data of two separate experiments with 16 and 12 participants, respectively. As we found no statistically significant differences between these experiments, we combine the outcomes into one baseline experiment. We compute the mean opinion scores (MOS) for all $I = 15$ conditions and $J = 10$ lighting attributes as follows

$$\text{MOS}_{i,j} = \frac{1}{N} \sum_{n=1}^N r_{n,i,j}, \quad i \in 1 \dots 15, \quad j \in 1 \dots 10 \quad (1)$$

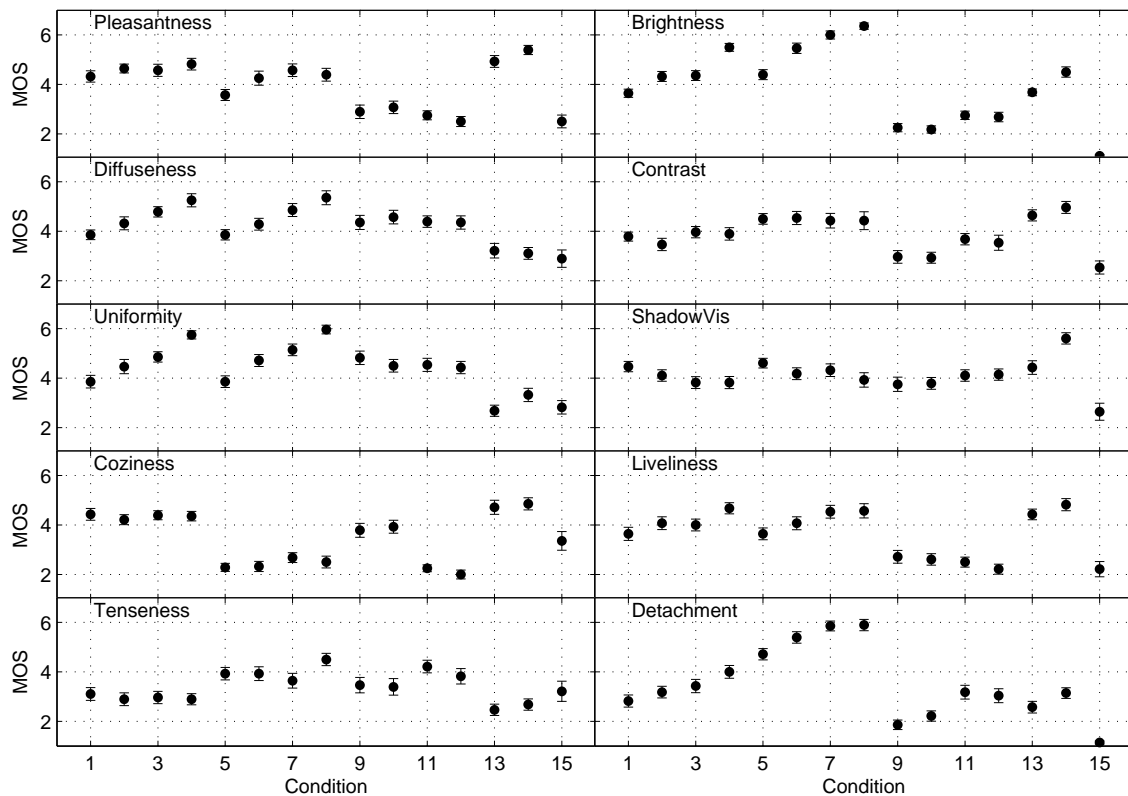


Figure 3. Mean opinion scores from the baseline experiment.

with $r_{n,i,j}$ being the rating of participant n on attribute j for condition i , and N being the total number of participants. The MOS and the respective standard errors for the baseline experiment are shown in Fig. 3. These values are considered to be the ground truth for lighting perception with respect to the 10 perceptual lighting attributes.

3. DISPLAY EXPERIMENTS

Analogous to the baseline experiment, we performed three display experiments to validate the predictive value of photographs and renderings of the lighting laboratory. In the following sections, we discuss the three display experiments and their results in relation to the baseline experiment.

3.1 Experiment 1: Renderings and Photographs

The goal of this first experiment, referred to as experiment E01, was to obtain an upper limit for the predictive value of an image on a 2D display by means of photographs as well as to verify the relative predictive accuracy of the respective renderings.

3.1.1 Stimuli

We took photographs of the lighting laboratory for each of the 15 lighting conditions (see Table 1). The photographs were taken with a Nikon D200 and a fisheye lens of 10.5mm focal length. A fixed aperture of f/8 and exposures of 1/5s and 1/10s were chosen. The photographs were de-fished, cropped, and brightness adjusted using Adobe Lightroom 3. Example photographs for three light conditions are shown in the top row of Fig. 4.



Figure 4. Photographs (top) and renderings (bottom) of the lighting laboratory.

We further created renderings for all 15 lighting conditions. The renderings were established using a 3-step process. Firstly, the room was modeled using 3dsMax Design.³ The model was then rendered into high dynamic range (HDR) images using Indigo renderer,⁴ an unbiased and physically-based renderer which achieves photorealistic images with accurate light transport simulation. Finally, the HDR images were tone-mapped into low dynamic range images using the Reinhard's photographic tone reproduction algorithm,⁵ which was shown to be superior in realism compared to other tone-mapping algorithms.⁶ Example renderings for three light conditions are shown in the bottom row of Fig. 4. The renderings created for this first experiment are in the following referred to as Rv01.

3.1.2 Experiment Procedures

In experiment E01, we presented the photographs and renderings in random order on a NEC P462 46" full-HD display. The participants were intentionally not made aware of the fact that both photographs and renderings were shown; they were simply told that a number of images are shown. The display was placed in a laboratory with low ambient lighting (approx. $20 \frac{cd}{m^2}$) behind the display. The participants were seated at a distance of about 1 m to the display. Seventeen people took part in this experiment, who rated their lighting perception on the same attributes as in the baseline experiment (see Fig. 2).

3.1.3 Experiment Results

To validate the predictive value of the photographs and renderings, we compute the difference mean opinion scores (DMOS) between the MOS of the display experiments (D) and the baseline experiment (B) as follows

$$DMOS_{i,j} = MOS_{i,j}^{(D)} - MOS_{i,j}^{(B)}. \quad (2)$$

The DMOS for the photographs and the renderings are shown for each condition in Fig. 5. The absolute DMOS averaged over all conditions are presented in Fig. 6.

The condition DMOS in Fig. 6 are consistently lower for the photographs compared to the renderings Rv01, indicating that the predictive value of the renderings does not yet match the predictive value of the photographs. This observation holds for a wide range of lighting attributes. The results in Fig. 5 show that the difference to the baseline MOS varies with the lighting attribute. For instance, the attributes 'Pleasantness' and 'Brightness' exhibit particularly large DMOS whereas the attribute 'Diffuseness' has the lowest DMOS.

The DMOS show a trend on the predictive value but do not identify whether or not the differences between the real room and the images are statistically significant. We therefore perform a statistical significance analysis using Linear Mixed Models (LMM).⁷ We chose LMM over the more widely used analysis of variance (ANOVA)

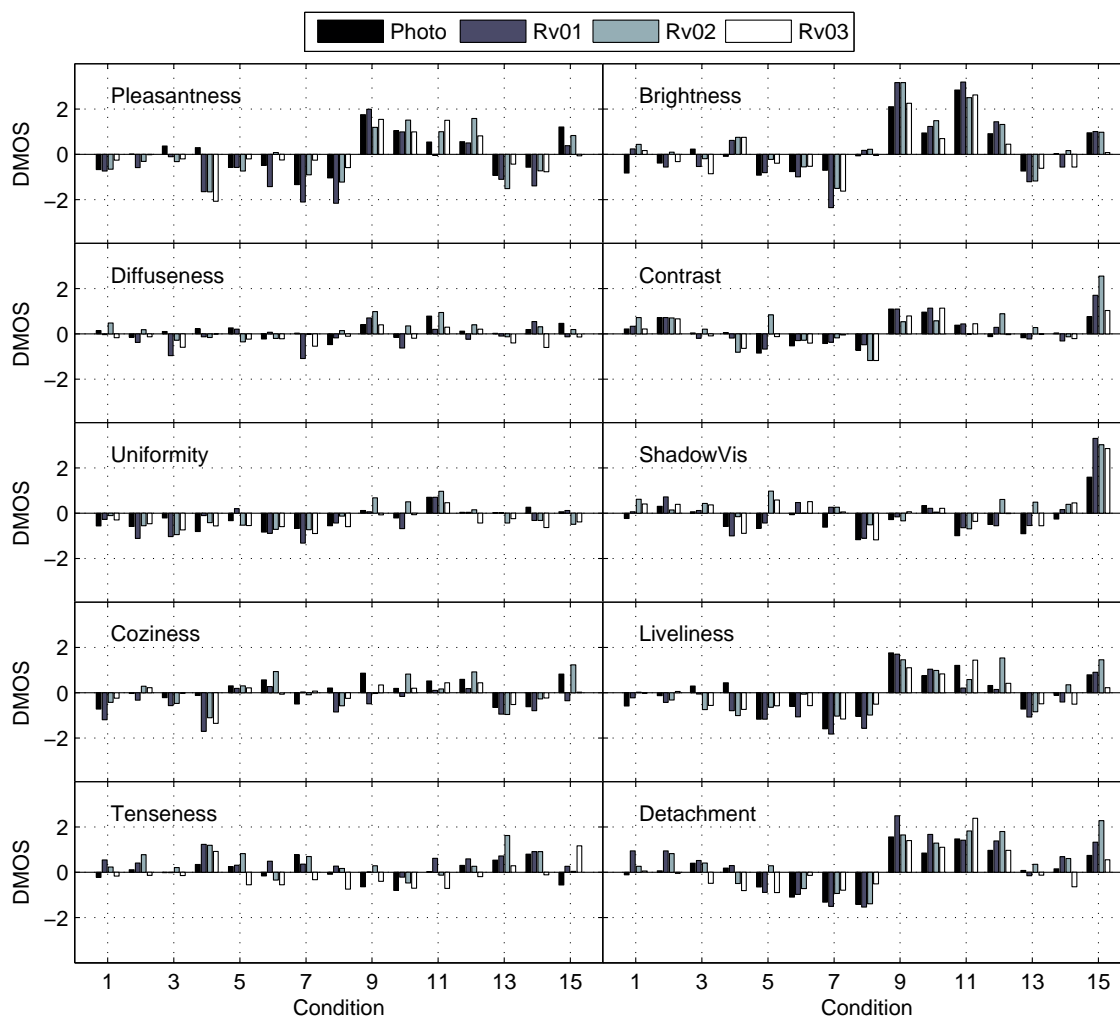


Figure 5. Difference mean opinion scores (DMOS) between the MOS from the baseline experiment and the MOS from the display experiments.

for several reasons. Firstly, LMM are robust to unbalanced data sets, which we are dealing with due to the different numbers of participants in the experiments. Secondly, LMM are able to cope with data that contains responses from both within and between observers. This is the case in our experiments as, for instance, in experiment E01, we have within responses for the photographs and renderings. The responses to the baseline experiment, however, are between observers. Finally, LMM are robust to data that is not normal distributed.

We performed individual LMM significance tests for all 10 attributes taking presentation type (real room versus images), condition, and participants as factors into account. The latter two factors were found to always be significant. The results of the LMM-based significance analysis for presentation type are shown in Table 2. These results show that the ratings on the photographs are statistically the same as the ratings of the real room from the baseline experiment. This holds for all 10 lighting attributes. The renderings Rv01, on the other hand, have 4 attributes for which the difference to the baseline experiment is statistically different. These observations confirm that the predictive value of the renderings Rv01 does not yet match the predictive value of the photographs.

In a post-experiment survey we asked the participants if they noticed that both photographs and renderings were shown. All but three participants indeed noticed that this was the case, the others either thought that all

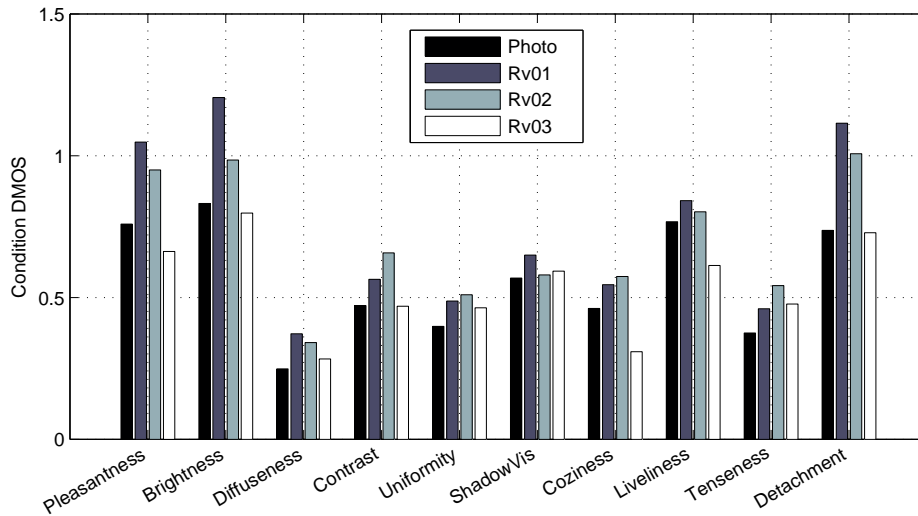


Figure 6. Absolute DMOS averaged over all 15 lighting conditions.

Table 2. Tests on statistically significant differences between the baseline experiment ratings and the respective display experiment ratings. The tests are performed using linear mixed models (××: significant difference with $p < 0.01$; ×: significant difference with $p < 0.05$; black dash: no significant difference).

Stimulus	Pleasantness	Brightness	Diffuseness	Contrast	Uniformity	Shadow Visibility	Coziness	Liveliness	Tenseness	Detachment
Photo	-	-	-	-	-	-	-	-	-	-
Rv01	××	×	-	-	-	-	××	-	-	×
Rv02	-	××	-	-	-	-	-	-	-	××
Rv03	-	-	-	-	-	-	-	-	-	-

images were renderings or photographs. Some participants reported that the renderings looked too ‘fake’ to be real. For instance, the carpet was reported to be too uniform compared to the carpet in the photographs. We used the participants’ feedback to improve the renderings and run a second iteration of our experiment.

3.2 Experiment 2: Renderings with Improved Materials and Geometry

3.2.1 Stimuli

Based on the results of experiment E01 and the feedback from the participants, we altered our renderings with the aim to improve the predictive value compared to the previous renderings. In particular, we assigned new, more realistic materials to the carpet as well as to the table top. Based on reflectivity measurements, we further lowered the reflectivity of the walls slightly. With regard to the geometry of the model, we improved the fluorescent luminaires as they exhibited an unwanted light leakage in the Rv01 renderings. We further added some LED spotlights that are present in the real room but were absent in the first iteration of renderings. We left them out in Rv01 because these LED spotlights were not turned on in any of the 15 conditions, however, they may still have an impact on the realism of the room and thus the predictive value of lighting perception.

Given the above changes to the 3D model, we created new renderings for the same 15 lighting conditions as in the baseline experiment. The renderings of this second iteration are in the following referred to as Rv02.

3.2.2 Experiment Procedures

The experiment procedures for the second display experiment, referred to as experiment E02, were the same as for the previous experiment (see Section 3.1.2), except that only the improved renderings Rv02 but no photographs were presented to the participants. Twelve people took part in this experiment, who rated their lighting perception on the same attributes as in the baseline experiment (see Fig. 2).

3.2.3 Experiment Results

The DMOS for the renderings Rv02 are shown for all conditions in Fig. 5. The respective absolute DMOS averaged over all conditions are presented in Fig. 6. The results reveal that for 6 out of 10 attributes ('Pleasantness', 'Brightness', 'Diffuseness', 'Shadow Visibility', 'Liveliness', 'Detachment'), the DMOS for Rv02 could be reduced, and hence the predictive value improved, as compared to the renderings Rv01. For the other 4 attributes ('Contrast', 'Uniformity', 'Coziness', 'Tenseness'), however, the DMOS increased slightly. The LMM-based significance analysis shows that only 2 attributes exhibit statistically significant differences, an improvement over the 4 differences seen in Rv01.

Despite clear improvements of the predictive value on several attributes with the second iteration of renderings, we still see a margin for enhancing the renderings towards the performance of the photographs. Attributes like 'Pleasantness', 'Brightness', and 'Detachment' still exhibit a lower predictive value and are further away from the baseline experiment than the photographs. Given these observations, we decided to further amend the imaging pipeline and perform another iteration on the experiments.

3.3 Experiment 3: Renderings with Improved Tone Mapping

3.3.1 Stimuli

The previous experiment results show that some attributes continue to have a lower predictive value (larger DMOS) as compared to the photographs. The change in the 3D model seemed to have only a small positive impact on the predictive value of the renderings. The lumen output and color temperature of the luminaires in the model are in line with the specifications of the real luminaires and should therefore not be amended. What may have an impact on the lighting perception, however, is the tone-mapping algorithm. We therefore decided to investigate whether adapting the tone reproduction of the images may have a positive impact.

Given the above, we changed the key values of the tone-mapping algorithm⁵ based on insights described in⁸ to compensate for some of the most severe discrepancies, such as the 'Brightness' differences for conditions 9-15. These third generation renderings are in the following referred to as Rv03.

3.3.2 Experiment Procedures

In the third display experiment, referred to as experiment E03, the latest renderings Rv03 were again shown to the participants on the NEC P462 46" display in random order. The presentation on the display was part of a larger experiment in which the renderings were also presented with other display types. For comparison with the other experiments, we use only the data of the presentation on the NEC P462 46" display. A total of 16 people viewed the renderings on this display and rated their lighting perception on the same attributes as in the baseline experiment (see Fig. 2).

3.3.3 Experiment Results

The DMOS for the renderings Rv03 are presented for all conditions in Fig. 5. The respective condition DMOS are shown in Fig. 6.

The condition DMOS highlight that the latest iteration of renderings resulted in a clear improvement to the previous renderings. For 'Shadow Visibility' and 'Tenseness', the predictive value remained about the same. For all other attributes, however, the predictive value could be considerably improved. This applies especially for 'Pleasantness', 'Brightness', and 'Detachment', for which the DMOS were rather large for the previous renderings. The DMOS of the renderings Rv03 is in all cases close to the DMOS of the photographs. The observations on

Table 3. Tests on statistically significant differences between the photographs and the respective renderings. The tests are performed using linear mixed models (××: significant difference with $p < 0.01$; ×: significant difference with $p < 0.05$; black dash: no significant difference).

Stimulus	Pleasantness	Brightness	Diffuseness	Contrast	Uniformity	Shadow Visibility	Coziness	Liveliness	Tenseness	Detachment
Rv01	××	-	×	-	-	×	××	×	××	-
Rv02	-	-	-	-	-	×	-	-	-	-
Rv03	-	-	-	-	-	-	-	-	-	-

the DMOS are confirmed by the LMM-based significance analysis. We see from Table 2 that no statistically significant differences between the baseline experiment and the renderings Rv03 remain. The results show that the overall amendments to the renderings were sufficient to converge to a predictive value similar to that of the photographs.

4. DISCUSSION

The three iterations of experiments showed that it is possible to converge towards the predictive value of lighting perception by using physically-based renderings instead of photographs. Having a comparable predictive value to the baseline experiment does, however, not necessarily mean that the photographs and renderings Rv03 have the same visual properties. In an extreme case, it could be that the renderings overestimate all attributes whereas the photographs underestimate all attributes, or vice versa, as compared to the baseline experiment. We therefore also perform a LMM-based significance analysis for the differences between the photographs and the three iterations of renderings. The results are shown in Table 3.

The significance analysis reveals that indeed the renderings Rv03 are closest to the photographs for all 10 attributes. We observe the same trend as with the comparison to the baseline experiment; an improvement with the iterations from Rv01 through Rv02 to Rv03. It is noteworthy to highlight again that renderings Rv01 were shown in the same experiment as the photographs, whereas Rv02 and Rv03 were shown in later experiments and were thus seen by different observers. The significant differences between the photographs and Rv01 are therefore expected to be comparably larger to Rv02 and Rv03 because of a more critical within-observer analysis in the LMM.

The progression of experiments showed that conveying lighting perception on a display using images is not a trivial task. Despite achieving a convergence of the images to the real-room experiment on all lighting attributes, it can still be seen in Fig. 5 that the accuracy between the images and the real room is dependent on the lighting attributes as well as the conditions. For instance, the ‘Pleasantness’ attribute is almost exclusively underestimated by the images for conditions 1-8 and consistently overestimated for conditions 9-12. Conditions 1-8 have luminaires with ‘Full’ intensity whereas conditions 9-12 have luminaires that are dimmed to ‘Half’ and ‘Min’ intensities. Despite having these reduced intensities being aligned with the real room, there appears to be a systematic error introduced by changing the overall light output. These errors have been considerably reduced by adapting the key value of the tone-mapping algorithm. Similar observations hold for other attributes, such as ‘Brightness’, ‘Liveliness’, and ‘Detachment’.

During the three experiment iterations discussed in this paper, we learned to better understand the effect of changes in the imaging pipeline on the resulting lighting perception. For instance, we found that the tone-

mapping algorithm has a large effect on lighting perception with regard to the 10 attributes we assessed. This confirms our earlier work,⁸ where we found that the key value of the Reinhard'02 algorithm has a significant effect on the visual performance of an image. In particular we found that the key of a scene (high-key versus low-key scene) as well as the surround illuminance in the room (viewing conditions) have strong effects on the preferred key values in the tone-mapping algorithm. Clearly the difference in preference of the key value is related to the predictive accuracy of lighting perception, but the details of this connection remain unknown.

Finally, it should be noted that the findings in this paper hold within the boundary conditions of the model used, our imaging pipeline, and the experiments performed. For instance, if a model of considerably different visual properties (e.g. higher complexity) or context would be created, or a gaming-like renderer instead of a physically-based renderer would be used, or the experiments would be performed under considerably different viewing conditions, the conclusions on the relative impact of certain parameters on the predictive value may be somewhat different. Obtaining a better understanding of the interaction between all parameters in the imaging pipeline for different contexts is key to making informed choices to maximize the predictive value of the renderings.

5. CONCLUSIONS

We presented a series of psychophysical experiments that we performed to evaluate the predictive value of images on a display compared to a real-world experiment with respect to a number of perceptual lighting attributes. We showed that the lighting perception of photographs is not significantly different to a respective real-world scenario. We further revealed that physically-based renderings can approach the predictive value of photographs when sufficient attention is paid to the 3D model, the materials, and the tone-mapping algorithm. Understanding the complex interaction between these different parameters in the imaging pipeline is key to optimizing the predictive value of the renderings in different contexts.

ACKNOWLEDGMENTS

The authors would like to thank Christian Vermorken, Zhi Kang Shao, Joris Albeda, and Irene Renkens for their help with performing the first experiment. We further extend our gratitude to all participants of the psychophysical experiments for sparing their valuable time.

REFERENCES

- [1] Salters, B., Murdoch, M. J., Sekulovski, D., Chen, S., and Seuntjens, P., "An evaluation of different setups for simulating lighting characteristics," in [*IS&T/SPIE Human Vision and Electronic Imaging XV*], **8291**, 1–13 (Jan. 2012).
- [2] Schneider Electric (Australia) Pty Ltd., "DALIcontrol." <http://www.clipsal.com/dalicontrol/home> (Aug. 2012).
- [3] Autodesk, "Autodesk 3ds Max Products." <http://usa.autodesk.com/3ds-max/> (Jan. 2013).
- [4] Glare Technologies Limited, "Indigo Renderer." <http://www.indigorenderer.com/> (Jan. 2013).
- [5] Reinhard, E., Stark, M., Shirley, P., and Ferwerda, J., "Photographic tone reproduction for digital images," in [*ACM Transactions on Graphics*], **21**, 267–276 (Aug. 2002).
- [6] Villa, C. and Labayrade, R., "Psychovisual assessment of tone-mapping operators for global appearance and colour reproduction," in [*IS&T European Conference on Colour in Graphics, Imaging, and Vision (CGIV)*], 189–196 (June 2010).
- [7] McCulloch, C. E. and Searle, S. R., [*Generalized, Linear, and Mixed Models*], John Wiley & Sons, Inc., New York, USA (2001).
- [8] Stokkermans, M. G. M., Murdoch, M. J., and Engelke, U., "Preference for key parameter of tone mapping operator in different viewing conditions," in [*Experiencing Light*], 1–4 (Nov. 2012).