



Aalborg Universitet

AALBORG UNIVERSITY
DENMARK

Why a Single Measure of Photorealism Is Unrealistic

Højilind, Simon Xavier Duran; Schnatterbeck, Camilla Gisela Hansen; Klein, Jákup; Østergaard, Lars; Arberg, Roman; Nadarajah, Stephanie Githa; Hansen, Martin Weiss; Kraus, Martin

Published in:

Proceedings of the Spring Conference on Computer Graphics (SCCG 2014)

Publication date:

2014

Document Version

Early version, also known as pre-print

[Link to publication from Aalborg University](#)

Citation for published version (APA):

Højilind, S. X. D., Schnatterbeck, C. G. H., Klein, J., Østergaard, L. L., Arberg, R., Nadarajah, S. G., ... Kraus, M. (2014). Why a Single Measure of Photorealism Is Unrealistic. In Proceedings of the Spring Conference on Computer Graphics (SCCG 2014) Comenius University, Bratislava. (Spring conference on Computer graphics).

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 ?

Take down policy

If you believe that this document breaches copyright please contact us at vbn@aub.aau.dk providing details, and we will remove access to the work immediately and investigate your claim.

Why a Single Measure of Photorealism Is Unrealistic

Simon Højland*
Aalborg University

Camilla Schnatterbeck†
Aalborg University
Stephanie Nadarajah‡
Aalborg University

Jákup Klein‡
Aalborg University
Martin W. Hansen**
Aalborg University

Lars Østergaard§
Aalborg University
Martin Kraus††
Aalborg University

Roman Arberg¶
Aalborg University



Figure 1: Frames of three time-lapse videos. From left to right: virtual reconstruction of a Viking building using virtual light sources; virtual reconstruction using image-based lighting; physical reconstruction of a similar building.

Abstract

While the concept of photorealism has important applications in computer graphics, the research community has not agreed on a definition of photorealism that specifies how to measure it. We employed two different test procedures, which correspond to different use scenarios, in order to determine the photorealism of a virtual reconstruction of a historic Viking building using two different lighting techniques. Even in this limited case, the measured degree of photorealism appears to depend on both the test procedure as well as the tested imagery; therefore, we conclude that there is no single measure of photorealism that is appropriate in all situations. Instead, photorealism appears to be a multifaceted phenomenon that requires different measurement procedures for different use scenarios.

CR Categories: I.3.7 [Computer Graphics]: Three-Dimensional Graphics and Realism—Color, shading, shadowing, and texture; H.1.2 [Models and Principles]: User/Machine Systems—Human factors

Keywords: photorealism, rendering, perception

*e-mail: shajli11@student.aau.dk

†e-mail: cschna11@student.aau.dk

‡e-mail: jklein11@student.aau.dk

§e-mail: llas11@student.aau.dk

¶e-mail: rarber11@student.aau.dk

||e-mail: snadar11@student.aau.dk

**e-mail: mwh@create.aau.dk

††e-mail: martin@create.aau.dk

1 Introduction

Photorealism is an important goal in many applications of computer graphics, e.g., computer-generated visual effects in movies and TV productions or advertisements in print media. An image is usually defined to be photorealistic if it is perceptually indistinguishable from a photograph of a scene [Rademacher et al. 2001], but such a definition does not specify how to measure photorealism. For example, this definition neither specifies whether a reference photograph has to be provided for comparison nor how much time is available for inspection — to name just two unknowns of a potential test procedure. Thus, many different procedures have been suggested [Meyer et al. 1986; Rademacher et al. 2001; Longhurst et al. 2003; McNamara 2005; Sundstedt et al. 2005; Meseth et al. 2006; Kozłowski and Kautz 2007; Sundstedt et al. 2007; Glencross et al. 2008; Ward and Glencross 2009; Stich et al. 2011; Fan et al. 2012; Kryven and Cowan 2013] with few attempts to compare the results of different procedures [Longhurst et al. 2003; Kozłowski and Kautz 2007; Sundstedt et al. 2007]. To our knowledge, no systematic comparison of different tests including photographs has been published.

In this work, we present two experiments involving human test participants to measure photorealism: one test procedure lacked a reference while the other test procedure required test participants to compare the computer-generated imagery to photographed imagery. The two procedures correspond to two use scenarios, namely computer-generated imagery in a movie that doesn't resemble anything in the filmed footage and computer-generated imagery that resembles filmed objects such that a direct comparison is possible. Due to this difference, we assumed that the two test procedures would always result in different measurements.

The computer-generated imagery in our experiments were virtual reconstructions (using two different lighting techniques) of a historic Viking building that were compared to a time-lapse video of a physical reconstruction of a similar Viking building. As expected, the experiments showed very different results for the two test procedures; however, this was only true for one of the employed lighting techniques while the results of the two test procedures were very similar for the other lighting technique. This suggests that the two test procedures result in fundamentally different measurements. We

conclude that a single test procedure for photorealism is very unlikely to appropriately cover all use scenarios; instead, different test procedures have to be employed depending on the specific application.

2 Previous Work

In the last ten years visual perception has received increasing attention in the computer graphics community as demonstrated by the ACM Symposium on Applied Perception, the ACM Transactions on Applied Perception, Eurographics state of the art reports [OSullivan et al. 2004; Bartz et al. 2008], SIGGRAPH courses [McNamara et al. 2011], and the publication of books on the topic [Thompson et al. 2011; Cunningham and Wallraven 2011], to name just a few examples.

One recurring issue is the measurement of photorealism of computer-generated imagery. It is widely accepted that an image should be considered photorealistic if it is perceptually indistinguishable from a photograph of a scene [Rademacher et al. 2001]. This concept suggests that photorealistic computer-generated imagery should be compared with photographs by human test subjects. In particular, automatic methods to identify photorealistic renderings have to be validated with tests that employ human test subjects.

However, not all studies that use human test subjects to evaluate the degree of photorealism also include photographs [Sundstedt et al. 2005; Sundstedt et al. 2007; Kozlowski and Kautz 2007]. Rademacher et al. [2001] have shown that human test subjects can be strongly influenced in their perception of photorealism by certain visual cues; therefore, the results of tests with and without comparisons to photographs are likely to be significantly different, and it is unclear in how far the results are comparable.

Longhurst et al. [2003] compared a real scene with a photograph and a rendering. The results showed that human test subjects can easily identify the real scene. This is not surprising since the test setup did not take the changing optical accommodation of the test subjects' eyes into account. Meyer et al. [1986] avoided this problem by using a projection onto Fresnel lenses using view cameras, which strongly decreased the brightness and sharpness of the projected images. Supposedly, this has contributed to the result that test subjects were not able to determine which of the dark and blurry images was real and which was computer-generated.

McNamara [2005] suggested evaluating photorealism by matching lightness in renderings and photos to a real scene. While this is an interesting approach, it is unlikely to be a sufficient test of photorealism since other visual features of an image could prevent the impression of photorealism.

Rendering images that are perceptually indistinguishable from photographs is notoriously difficult. In some studies, the bar is lowered by limiting the time for inspection of the images, e.g. to 3 seconds [Glencross et al. 2008; Ward and Glencross 2009] or even 150 milliseconds [Kryven and Cowan 2013].

One common observation in tests of photorealism with comparisons to photographs is the strong dependency on the specific scenes [Meseth et al. 2006; Ward and Glencross 2009; Stich et al. 2011]; i.e., it is often observed that a rendering technique produces photorealistic results for some scenes but fails to produce photorealistic results in other cases.

This indicates that tests of photorealism without a comparison to photographs, e.g. Sundstedt et al. [2007] and Kozlowski and Kautz

[2007], are problematic. On the other hand, comparisons of different measures of photorealism appear to be limited to tests without photographs — with the exception of Longhurst et al. [2003], which is limited in other ways as discussed above.

Therefore, this work attempts to provide a comparison of different measures of photorealism in tests that include photographs.

3 Experiments and Results

Figure 1 presents frames of the three videos that were used in the experiments. The two computer-generated virtual reconstructions of a historic Viking building were composited onto a photographed time-lapse video. One of the renderings employed virtual light sources while the other one employed image-based lighting using a recorded environment map. For comparison, another time-lapse video of a physical reconstruction of a similar Viking building was recorded.

In the first experiment, 60 test participants watched (in random order) the video of the physical reconstruction and a computer-generated video — 30 participants watched the one rendered with virtual light sources and the other 30 the one using image-based lighting. All test participants were first-year university students. The following manuscript was used to ask them whether they believed that they had seen a photographed or a computer-generated video:

Welcome and thank you for participating. If you would please take a seat in front of the computer, then we will begin. On the screen in front of you, you will be shown some small videos and it is then up to you to tell if the video is real or if a 3D object has been placed in the video. You will be shown two random videos, out of several different ones, where some are more realistic than others. When you have cast your vote the next video will be shown.

Test participants were allowed to watch each video as often as they wished. Figure 2 shows the percentages of test participants answering that the computer-generated video was “real.” For the physical reconstruction, 78% of the test participants answered correctly that it was “real.”

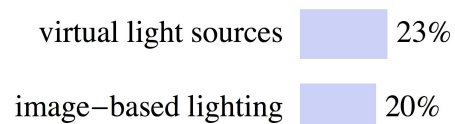


Figure 2: Percentages of 30 test participants answering that the computer-generated rendering was “real” in the tests without comparison.

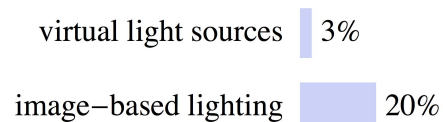


Figure 3: Percentages of 30 test participants answering that the computer-generated rendering was the “real” one in the tests with a comparison to a photographed video.

In the second experiment, the same 60 test participants were asked with the following manuscript which of the two videos they had watched was photographed:

Now you have seen two videos, one of them contained a 3D house and one contained a real house, can you tell me which video contained the real house?

Test participants were again given the opportunity to watch the videos arbitrarily often before they answered the question. Figure 3 shows the percentages of test participants answering that the computer-generated house was real.

4 Discussion and Conclusion

We used a one-tailed Fisher's exact test with a 5% significance level for the statistical analysis [Agresti 1992].

If considered isolated, the results for the rendering with virtual light sources confirmed our assumption that it would be significantly ($p = 0.026$) easier for test participants to correctly identify computer-generated imagery in the second test procedure. However, we were surprised that both test procedures resulted in the same outcome for the rendering with image-based lighting.

If the results for both rendering methods are taken into account, there is no significant difference between the two test procedures ($p = 0.11$). We assume that this difference fails to be significant because of the small number of test participants. If this assumption is correct, then the experiments suggest that there is not only a difference between the results of the two test procedures but that this difference also depends strongly on the specific imagery. Therefore, we conclude that it is impossible to predict the outcome of one of the test procedures from the outcome of the other one. Since the test procedures correspond to different use scenarios, this means that it is unlikely that any single test procedure can measure photorealism for all use scenarios.

If there is no single measurement procedure, one might wonder whether photorealism is actually a useful scientific concept. We believe that it is provided that it is understood as a multifaceted phenomenon that requires different measurement procedures in different situations — similarly to other scientific phenomena, e.g., electromagnetic radiation in its many forms (radio waves, visible light, X-rays, etc.).

Acknowledgements

The authors would like to thank the test participants for their time, the staff of Lindolm Høje Museum and Vikingecenter Fyrkat for their support, Claus B. Madsen for his valuable input, Otto Mønstedts Fond for the financial support of the publication at SCCG 2014, and the anonymous reviewers for their comments.

References

- AGRESTI, A. 1992. A survey of exact inference for contingency tables. *Statistical Science* 7, 1 (February), 131–153.
- BARTZ, D., CUNNINGHAM, D., FISCHER, J., AND WALLRAVEN, C. 2008. The role of perception for computer graphics. *Eurographics state-of-the-art-reports*, 65–86.
- CUNNINGHAM, D., AND WALLRAVEN, C. 2011. *Experimental design: From user studies to psychophysics*. AK Peters, Ltd.
- FAN, S., NG, T.-T., HERBERG, J. S., KOENIG, B. L., AND XIN, S. 2012. Real or fake?: Human judgments about photographs and computer-generated images of faces. In *SIGGRAPH Asia 2012 Technical Briefs*, ACM, New York, NY, USA, SA '12, 17:1–17:4.
- GLENCROSS, M., WARD, G. J., MELENDEZ, F., JAY, C., LIU, J., AND HUBBOLD, R. 2008. A perceptually validated model for surface depth hallucination. *ACM Trans. Graph.* 27, 3 (Aug.), 59:1–59:8.
- KOZŁOWSKI, O., AND KAUTZ, J. 2007. Is accurate occlusion of glossy reflections necessary? In *Proceedings of the 4th Symposium on Applied Perception in Graphics and Visualization*, ACM, New York, NY, USA, APGV '07, 91–98.
- KRYVEN, M., AND COWAN, W. 2013. Modelling perceptually efficient aquatic environments. In *Proceedings of the ACM Symposium on Applied Perception*, ACM, New York, NY, USA, SAP '13, 131–131.
- LONGHURST, P., LEDDA, P., AND CHALMERS, A. 2003. Psychophysically based artistic techniques for increased perceived realism of virtual environments. In *Proceedings of the 2nd International Conference on Computer Graphics, Virtual Reality, Visualisation and Interaction in Africa*, ACM, New York, NY, USA, AFRIGRAPH '03, 123–132.
- MCMANARA, A., MANIA, K., AND GUTIERREZ, D. 2011. Perception in graphics, visualization, virtual environments and animation. In *SIGGRAPH Asia 2011 Courses*, ACM, 17.
- MCMANARA, A. M. 2005. Exploring perceptual equivalence between real and simulated imagery. In *Proceedings of the 2nd Symposium on Applied Perception in Graphics and Visualization*, ACM, New York, NY, USA, APGV '05, 123–128.
- MESETH, J., MÜLLER, G., KLEIN, R., RÖDER, F., AND ARNOLD, M. 2006. Verification of rendering quality from measured btfs. In *Proceedings of the 3rd Symposium on Applied Perception in Graphics and Visualization*, ACM, New York, NY, USA, APGV '06, 127–134.
- MEYER, G. W., RUSHMEIER, H. E., COHEN, M. F., GREENBERG, D. P., AND TORRANCE, K. E. 1986. An experimental evaluation of computer graphics imagery. *ACM Trans. Graph.* 5, 1 (Jan.), 30–50.
- OSULLIVAN, C., HOWLETT, S., MORVAN, Y., MCDONNELL, R., AND OCONOR, K. 2004. Perceptually adaptive graphics. *Eurographics state of the art reports* 4, 141–164.
- RADEMACHER, P., LENGYEL, J., CUTRELL, E., AND WHITTED, T. 2001. Measuring the perception of visual realism in images. In *Proceedings of the 12th Eurographics Conference on Rendering*, Eurographics Association, Aire-la-Ville, Switzerland, Switzerland, EGWR'01, 235–248.
- STICH, T., LINZ, C., WALLRAVEN, C., CUNNINGHAM, D., AND MAGNOR, M. 2011. Perception-motivated interpolation of image sequences. *ACM Trans. Appl. Percept.* 8, 2 (Feb.), 11:1–11:25.
- SUNDSTEDT, V., DEBATTISTA, K., AND CHALMERS, A. 2005. Perceived aliasing thresholds in high-fidelity rendering. In *Proceedings of the 2nd Symposium on Applied Perception in Graphics and Visualization*, ACM, New York, NY, USA, APGV '05, 166–166.

- SUNDSTEDT, V., GUTIERREZ, D., ANSON, O., BANTERLE, F., AND CHALMERS, A. 2007. Perceptual rendering of participating media. *ACM Trans. Appl. Percept.* 4, 3 (Nov.).
- THOMPSON, W., FLEMING, R., CREEM-REGEHR, S., AND STEFANUCCI, J. K. 2011. *Visual perception from a computer graphics perspective*. CRC Press.
- WARD, G., AND GLENCROSS, M. 2009. A case study evaluation: Perceptually accurate textured surface models. In *Proceedings of the 6th Symposium on Applied Perception in Graphics and Visualization*, ACM, New York, NY, USA, APGV '09, 109–115.