PERGAMON

Computers & Graphics 25 (2001) 511–518

COMPUTERS & GRAPHICS

www.elsevier.com/locate/cag

Technical Section

Validation proposal for global illumination and rendering techniques

Frédéric Drago[a], Karol Myszkowski[b,*]

[a] *University of Aizu, Aizu Wakamatsu, 965-8580 Japan*
[b] *Max-Planck-Institute for Computer Science, Stuhlsatzenhausweg 85, Im Stadtwald, 66-123 Saarbrucken, Germany*

Abstract

The goal of this study is to develop a complete set of data characterizing geometry, luminaires, and surfaces of a non-trivial existing environment for testing global illumination and rendering techniques. This paper briefly discusses the process of data acquisition. Also, the results of experiments on evaluating lighting simulation accuracy, and rendering fidelity for a Density Estimation Particle Tracing algorithm are presented. The importance of using the BRDF of surfaces in place of the more commonly used specular and diffuse reflectance coefficients is investigated for the test scene. The results obtained are contrasted with an "artistic approach" in which a skilled artist manually sets all reflectance characteristics to obtain a visually pleasant appearance that corresponds to the existing environment.

Keywords: Global illumination; BRDF; Density Estimation Particle Tracing; Lighting simulation; Realism; Psychophysics

1. Introduction

It is relatively easy to use commodity rendering techniques to create great looking images; however, it is much more difficult to create images that match the appearance of a real world environment as perceived by the human observer. The basic precondition to achieving this goal is physically-based lighting simulation, which is computationally a demanding problem. To make computation tractable in practical applications, many simplifying assumptions are usually introduced to underlying physical models. Because analytic evaluation of such simplifications and interactions between them is generally impractical, the correctness of a given technique must be checked experimentally by comparison of simulation results to some reference data [10]. For example, distribution of illumination at some predefined points derived analytically (available only for very simple scenes [13]) or measured experimentally can be used to validate the lighting simulation part of a rendering algorithm. However, an accurate solution of the global illumination problem does not guarantee the correct appearance of the resulting images if a proper tone mapping operator (TMO) is not applied to display those images on the CRT device [17]. On the other hand, the solution accuracy can be relaxed in many cases without noticeable differences in image quality as perceived by the human observer. Clearly, some metrics of the image quality are needed to compare the appearance of virtual and real-world images. Although, some attempts of using the objective metrics based on the computer modeling of the Human Visual System have been recently reported [15], still the most robust results can be obtained using the traditional approach based on subjective metrics involving experiments with the human subjects [11]. Unfortunately, such experimental validation was almost never performed for existing global illumination solutions, which makes it difficult to compare their efficiency, or even test their implementation correctness.

One of the reasons such validation experiments are seldom performed in practice is lack of standardized, robust, non-trivial, and easily accessible test data.

*Corresponding author. Tel.: +49-681-9325-429; fax: +49-681-9325-499.

E-mail address: myszkowski@mpi-sb.mpg.de (K. Myszkowski).

0097-8493/01/$ - see front matter
PII: S0097-8493(01)00072-3

Ideally, a standard set of diversified tests should be available, such that the performance of global illumination solutions could be measured in terms of the achieved lighting simulation accuracy, image fidelity and computation time [13]. The existence of such tests would make possible the comparison of particular elements of rendering algorithms as well. Other areas of research such as computer vision, and augmented reality could also benefit from such models in their studies.

There have been some successful attempts to develop such standard tests. A well-known example is the Cornell box [3], which was used to validate radiosity solutions [8,12,16], and more recently density estimation particle tracing algorithm [19]. Rushmeier et al. [15] developed a model of a conference room with four different lighting systems and corresponding photographs of its real world counterpart. Recently, McNamara et al. [11] built a low scale environment, and requested the human observers to compare its appearance against images generated by the Radiance rendering system [20]. Mardaljevic [10] used Radiance to predict illumination in a simple room under real sky conditions, and reported a good agreement of the simulation results with the illumination measurement. These scenes are of relatively low complexity in terms of geometry and lighting. The scene models used in the latter two cases are not publicly available, so their reproduction is difficult. For the Cornell box and conference room only the Lambertian and specular reflectance data are posted on the corresponding Web pages [3,15]. A more complex scene of a conference room (see in [7] plate III.19) was considered by Ward-Larson to test his Radiance package [20]. However, the model does not correspond well to its counterpart in reality. A similar problem concerns the early model of an atrium at the University of Aizu [13], in which the geometry data were excessively simplified (less than 30,000 polygons were used), and the reflectance properties tuned in an ad hoc manner.

This research expands upon the atrium data by providing a detailed geometry model, considering calibrated luminaires, measuring BRDF of the most important construction materials, and measuring illumination at multiple sample points. The data[1] is used to validate the lighting simulation accuracy and rendering quality of the Density Estimation Particle Tracing (DEPT) algorithm [18].

The following sections briefly describe the most significant elements of the atrium data, the apparatus used to obtain experimental data, and the results of the experiments. The main focus is the problem of lighting simulation accuracy, and visual fidelity of the resulting images. The conclusion also proposes some directions for future experiments.

[1] For the convenience of the research community, the complete data related to this research can be downloaded from a dedicated Web page [2], which makes possible to perform a similar validation of third party algorithms and software.

2. Characteristics of the data set

When preparing test data for validation of global illumination algorithms the question arises what should be the basic characteristics of such "a good test data set". Simple test scenes are useful to test selected algorithm features and give a rapid idea of the solution accuracy [13]. However, it is important to know how a global illumination algorithm behaves in a really complex environment where most of its features might be solicited at once. We think that a good test data which is designed specifically for this purpose should feature the following characteristics:

- Significant geometrical complexity to test the algorithm behavior for scenes typical in practical applications. In particular, the scaling of computation load and memory requirements as a function of the scene complexity are of primary importance.
- Significant lighting complexity involving luminaires with different types of goniometric diagrams such as spot light sources with limited spatial effect or global lights affecting more significant regions of the scene must be considered. Some scene regions should be illuminated exclusively by indirect lighting, while in other regions direct lighting should dominate to test the accuracy of lighting simulation and the quality of resulting shading for each of those lighting components.
- Surfaces with various reflectance characteristics, including those with BRDF that are difficult to approximate by pure diffuse or specular models, should be considered to test the performance of various mechanisms used to transfer lighting energy between scene surfaces.
- The environment must exist to do measurements and allow a perceptual validation of the obtained results.

These conditions are met by the model of atrium at the University of Aizu. Besides obvious location reasons, the atrium was chosen for its challenging lighting model, the richness and diversity of the materials employed in the construction, and the geometrical purity of its architectural structure.

2.1. Geometry

Since our goal was to develop the reference validation data, which requires strict accuracy, we did not use semi-automatic approaches based on photogrametric methods [4] which could greatly simplify the atrium modeling.[2] Instead, we chose a traditional modeling approach using the 3D Studio Max package. Although, it was a long and tedious process, we were able to reconstruct the atrium to a satisfying level of precision and detail. The atrium substructure was modeled based on the blueprints, while details were measured directly from the actual scene. For better control over the number of polygons, automatic modeling procedures such as lofting or boolean compound objects were avoided. Almost every hidden part was removed, and the pieces beveled with a minimum number of edges necessary for 45° smoothing. Every planar surface is represented by a minimal number of triangles, which makes it possible to use the model "as is" by global illumination solutions that do not require an additional mesh tessellation. The final model is composed of almost 700,000 triangles of greatly varying sizes, which makes it attractive for testing many hierarchical solutions and clustering scenarios.

2.2. Surface reflectance

The atrium is built with common construction materials, but their reflectance characteristics are sufficiently different to make this study challenging in terms of light scattering modeling. Although, for exact modeling the BRDF of all construction materials is required, for experimental purposes six materials representing more than 80% of the total surface area in the atrium were selected. The reflectance attributes for the remaining surfaces were estimated based on available literature, e.g., [14]. To derive the color of painted surfaces, which were not selected for the BRDF measurement, we used the Standard Paint Color charts edited by the Japan Paint Manufacturers Association, and derived the corresponding RGB values from a Toyo 88 RGB color finder. The BRDF measurements were performed at the Keldysh Institute of Applied Mathematics (Moscow, Russia); more details concerning the measurement procedure are given in [9]. To reduce the random errors in the measured BRDF a fairing procedure was performed to remove the outlier BRDF samples.

2.3. Luminaires

Two types of luminaires were considered. Goniometric diagrams with spatial candle power distribution for these luminaires were received from their manufacturer Matsushita Electric. While these diagrams are highly accurate, a practical problem arises with estimation of the maintenance factor for every luminaire to account for its utilization level, accumulated dirt and so on. Some atrium luminaires are used on a daily basis, while others are usually switched off to cut down electricity costs. The illumination due to every luminaire was measured trying to minimize the influence of indirect lighting. A luxmeter probe was placed in a deep box in such a way that the probe was directly illuminated, while indirect lighting influence was suppressed. To reduce interreflections within the box all its sides were covered with a black fabric. Measurements to determine a reference were taken from newly installed bulbs in a dirt free fixture. The maintenance factor for this reference luminaire was assumed to be 100%. The maintenance factors of the other luminaires were scaled in respect to the reference, based on measurements of illumination at similarly selected sample points. The lowest maintenance factor was 62%, which clearly shows that the catalogue data should be properly interpreted to account for characteristics of the real world environment.

[2] Actually, the reconstruction of the geometry using photogrametric methods, and evaluating their accuracy using our reference model could be an interesting topic of further research.

3. Results

In the following two sections we describe our experiments with the atrium scene, which were focused on two main issues:

- Estimation of the lighting simulation accuracy through comparison with measurement data;
- Evaluation of computer images fidelity in respect to the real world environment.

The DEPT technique [18], which is a part of the commercial package Inspirer (Integra, Inc.) was used for lighting simulation, and image rendering.

3.1. Measuring lighting simulation accuracy

Our primary goal was to validate the DEPT technique in terms of lighting simulation accuracy for a non-trivial environment using the most accurate input data that were available to us. Also, we evaluated the difference in lighting simulation accuracy for the measured BRDF, and less rigorous approaches to light scattering such as Lambertian and mirror reflections. In the latter case, the reflectance coefficients were derived by hemispherical integration of the measured BRDF for various incident angles, and averaging of the values obtained. This corresponds to the situation when the approximated reflectance can be acquired somehow, e.g., by using a simple measurement device, or by taking standard values from textbooks for the most common materials. This scenario is also useful when the global illumination and rendering software does not support BRDF. We also investigated the impact of ignoring directional light transfer completely, which is typical for simple radiosity techniques.

The results of lighting simulation for the three approaches to the surface reflectance estimation were compared at 84 sample points against the measurement data. Although, all measurement points were directly illuminated, indirect lighting was also significant accounting for about 30–55% of total illumination. The graphs in Fig. 1 show the distribution of measured and simulated illumination at selected sample points, which are located at the atrium floor along two different lines (refer to the Web page [2] for detailed simulation results and error analysis for all sample points). The best results were obtained from the measured BRDF, in which case the average simulation error in respect to the measured illumination was 10.5%. For the approach based on the averaged diffuse and specular reflectance, the error increased to 16.8%, and for the radiosity-like solution the error amounted to 18.2%. The errors of less than 5%, 10%, and 20% were obtained for about 40%, 75%, and 93% of the sample points when the measured BRDFs were used, and 8%, 31%, and 68% of the sample points when only the Lambertian reflection was considered. These results show that the Lambertian approximation of light reflection leads to significant errors for the atrium scene. These errors can be attributed to highly reflective glossy materials from the side walls of the atrium, which directionally reflect significant lighting energy toward the floor.

The results obtained may appear pessimistic in terms of the simulation accuracy achieved, however, the following tolerances for lighting design applications are proposed in the guidelines issued by Commission International de l'Éclairage [6]: 10% for average illuminance calculations, and 20% for measured point values. Also, according to Fisher [6] it is difficult to design a complex situation to a 5% accuracy at present. Such a high tolerance is a result of the realistic evaluation of the accuracy of input data for lighting simulation such as the BRDF of materials, description of luminaires, and simplifications of geometrical models. In this study, the DEPT technique meets most of the requirements imposed on the simulation accuracy in lighting engineering applications.

3.2. Estimating fidelity of rendered images

A robust prediction of real-world scene appearance based on valid input data opens many important applications for realistic rendering. We designed a number of psychophysical experiments involving the human subjects to check how different are rendered images in respect to their real-world counterpart. In this study we considered material reflectance based on the measured BRDF, which gave the best accuracy of lighting simulation. Also, we considered the "artistic approach" in which Lambertian and mirror reflection coefficients are intuitively tuned by a skilled artist to get the best match of image appearance in respect to the

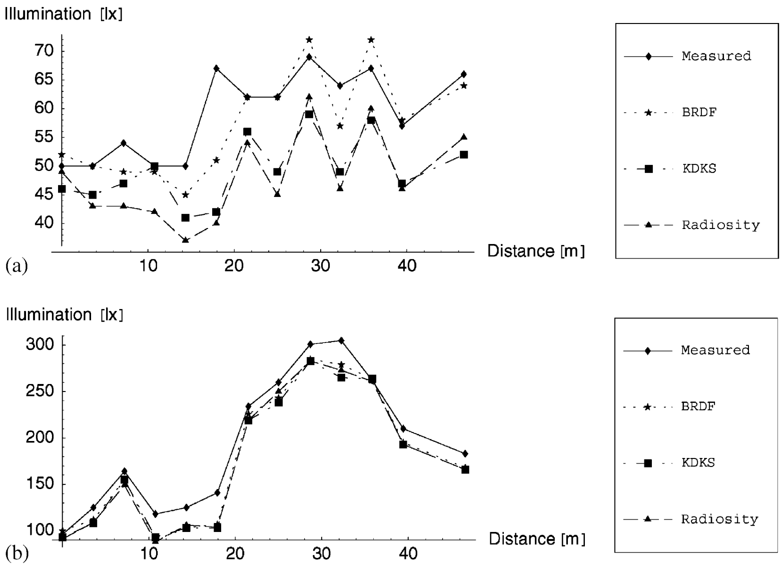

Fig. 1. Illumination values [lx] distribution on the atrium floor along (a) the left-hand side corridor (near the wall), and (b) the central part of the atrium. Measured illumination is compared against the results of simulation in which the BRDF, averaged diffuse and specular (KdKs), and averaged diffuse (radiosity) reflectance coefficients are considered.

real-world atrium. Such an artistic approximation is common in many practical applications [1]. Fig. 2 shows images obtained using the measured BRDF and manually tuned reflectance.

We intended to compare the rendered images to photographs of the atrium displayed on the CRT device, or printed using a high quality color printer. Unfortunately, because of mixed fluorescent and incandescent lighting in the atrium, the standard photographic techniques failed to produce satisfactory results. Different color reversal films were tested such as daylight and tungsten films designed for the color temperature of 55 00 K (Kelvin) and 3100 K, respectively. Photographs taken using films with color balanced for the daylight conditions had a yellowish tone because of the low color temperature of the lighting in the scene, compared to typical daylight conditions. The use of films with color balanced for tungsten lighting resulted in the opposite phenomenon; since the overall color temperature of the lighting is more than 3100 K due to the fluorescent lighting, photographs were cast in a blue hue. If lighting accuracy would not be the objective we could have used special compensating filters or a flash to reduce this color problem. Much better color fidelity was obtained using digital cameras because in this case the white-point was adjusted for the environment. However, due to the high dynamic range of luminances in the scene, the images lack of details in highlight and shadow areas. An example of such an image is shown in Fig. 3, and samples of photographs obtained using all discussed techniques as well as more technical details on the films and cameras used in our experiments are provided on the Web page [2]. While the photographs were not perfect in reconstructing the overall atrium appearance (the question arises whether the term "photo-realistic rendering" is adequate, or rather just "realistic rendering" should be used instead), they were useful in evaluating some of the more detailed aspects of rendering such as the fidelity of geometry and light distribution modeling.

Finally, we decided on the cross-media comparison [5] in which images displayed on the CRT screen were compared directly against the real-world atrium. The set-up used in the first part of the experiment is shown in Fig. 4. We asked 25 subjects (12 of them got basic training in realistic image synthesis and could be considered as experts, others were students in computer science thus accustomed to watch computer generated imagery) to rate how different are rendered images in respect to their real-world counterpart. We conducted two sets of experiments. The goal of the first one was to grade the overall realism of the images. The objective of the second experiment was to obtain detailed comments about specific features of the images, which makes possible evaluation of selected algorithms in our rendering solution.

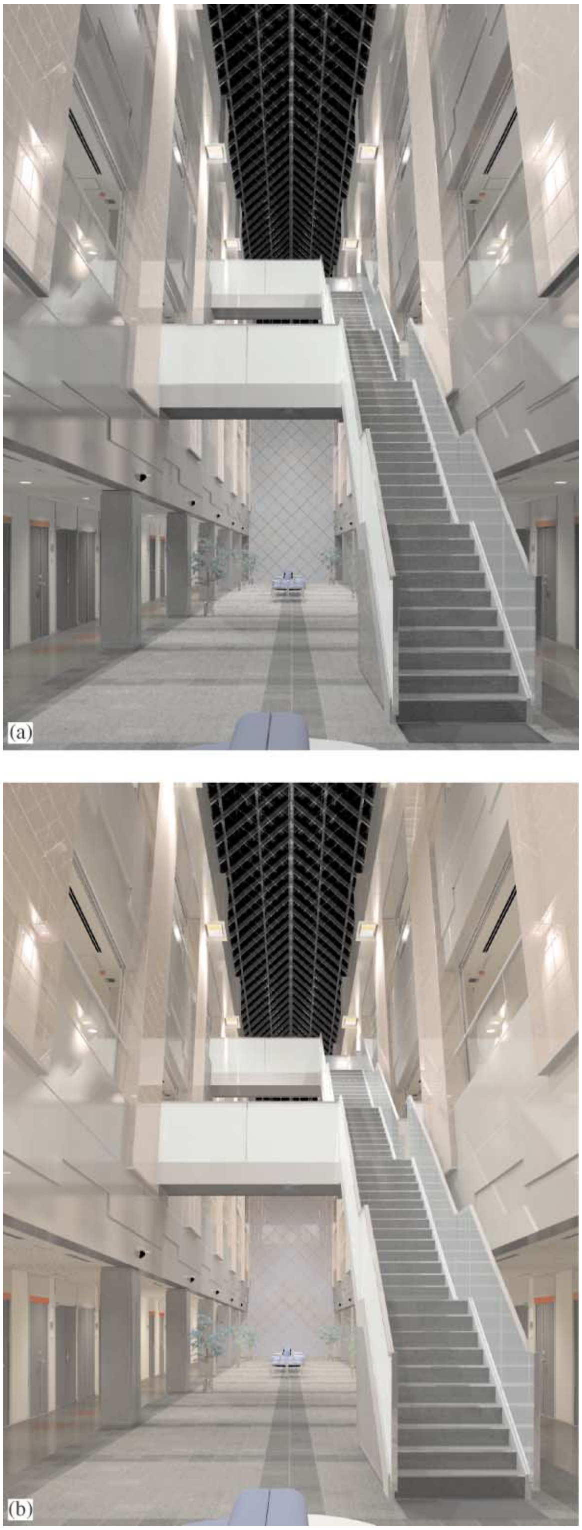

Fig. 2. Rendered images: (a) the artistic approach, (b) using measured BRDF.

Fig. 3. Photograph obtained by a digital camera.

Fig. 4. Set-up used for experimental comparison of the atrium images displayed on a CRT device against the real-world view. The subject observed the atrium through an aperture which limited the field of view to the one similar to the displayed images.

In the first set of experiments, the subjects were allowed to watch the atrium scene for 10 s, and then they were exposed randomly for the same period of time to the BRDF-based image, artistic image, or just the digital photograph. The subjects were asked how well the real-world atrium is reproduced by every presented image. Also, more specific questions concerning lighting reconstruction and tone reproduction on the CRT device were posed. In all cases the photographs got the highest scores with the mean ranking falling into the range 7.36–8.48 (for the scale spanning the range 0–10). The artistic and BRDF-based approaches got lower scores falling into the ranges 6.88–7.24 and 6.56–6.96, respectively. What is remarkable is that the differences in the mean ranking of photograph and computer images were rather small, below the value of standard deviation error. The artistic approach which is much easier and cheaper (no expensive BRDF measurement is required) was slightly higher ranked than the BRDF-based rendering, which to a certain extent explains why always pragmatic industry chooses tweaking rendering parameters instead of performing fully-fledged physically-based computations. Obviously, such an approach will fail for rendering the scenes whose appearance is unknown and cannot be reconstructed even by a skilled artist because of the lack of reference views.

In the second set of experiments unlimited observation time was allowed, and the photographs were not considered anymore. The detailed questions posed to the subjects concerned the quality of lighting and shadows reconstruction, texture and luminaire rendering, highlights and reflections appearance, and contrast reproduction. Two sets composed of four images rendered from different viewpoints were displayed, and the subjects were allowed to walk around the atrium and look every feature in more detail. The artistic approach was found to be more realistic in terms of overall lighting and tone. It is not surprising since this approach is based on careful observation of the environment and a complete freedom in tweaking surface reflectance parameters until satisfying realism was reached. The appearance of highlights, reflections, and shadows were almost equally ranked for the BRDF-based and artistic approaches.

The statistical analysis of the results shows that in the first set of experiments in which more general questions were posed, the subjects answered with a certain confidence. The results obtained are predictable, and might be used as a good statistical base. In the second set of experiments the variance is substantially higher, and a lack of consistency in answers was observed. The ratings showed that the subjects could not reliably differentiate between stimuli in terms of particular details. This suggests that the results obtained should not be used as a solid base for conclusions on the respective merits of the BRDF and artistic approaches.

Detailed comments provided by the subjects revealed that lack of contrast in highlight and shadow areas is especially annoying. This could be improved by more careful selection of the TMO. The architectural perfection of the model does not account for variations in construction and disparities of materials. This becomes especially visible near highlights and reflection regions, which are distorted because of inaccuracies in tile positioning. Man made defects are not reproduced in the simulation, which affects the fidelity and realism of the resulting images. Also, in the real-world scene subjectively stronger specularity can be observed for pink tiled columns and brown painted panels in the central part of the atrium. This disagreement in appearance can be explained by inaccuracies in capturing rapidly changing specular reflectance during the BRDF measurement. The detailed report and statistical analysis of the psychophysical experiments are available on the Web page [2].

The atrium test is a good tool to evaluate overall rendering performance and quality. However, it is too complex to validate isolated components of the global illumination algorithms. This requires some step-by-step validation procedure [13].

4. Conclusions

Rendering of realistic images is still a very involved process. When a level of detail similar to the one provided by the atrium is required, geometric modeling is very laborious. Luminaire characteristics are usually acquired directly from the manufacturers, however, for some decorative luminaires this may be difficult. Experience shows that estimation of the maintenance factor for luminaires is required to achieve reliable input data. A difficult problem is measurement of surface scattering characteristics, which requires costly equipment. Hopefully, in the future, it will be common practice for manufacturers to provide reflectance characteristics of construction and finishing materials.

Comparisons of lighting simulation accuracy, and rendering fidelity for various models of light scattering by the atrium surfaces were performed. The best lighting simulation accuracy was obtained when the measured BRDF was considered for the most exposed atrium surfaces. The best image fidelity was obtained from the manually adjusted reflectance characteristics based on real-world atrium observations. This unexpected result can be explained by the efficiency of tweaking reflectance parameters to improve image appearance, and reducing perceivable differences in respect to the reference view. The images obtained for the measured BRDF were also of good quality, which indicates that the DEPT technique is a suitable tool for predictive rendering based on valid input data.

An important future objective is to establish an experimental framework for testing and comparison of global illumination algorithms. The atrium data posted on the Web, and the proposed set of validation tests in this paper are the initial steps toward this goal. As future work we plan to evaluate other global illumination packages such as Radiance, and make the atrium data available in a greater variety of commonly used file formats.

Acknowledgements

We would like to thank William Martens for his advice concerning the organization of psychophysical experiments, and help in statistical analysis of the obtained results. We are grateful to Alan Chalmers for his valuable suggestions concerning this work. We thank to Vladimir Volevich for his help in using the Inspirer software, and providing clever scripts that simplified our tests. We thank Alexander Letunov, Sergey Ershov and Konstantin Kolchin of the Keldysh Institute of Applied Mathematics (Moscow, Russia) for measuring the BRDF of samples of the atrium construction materials. Finally, we thank Akira Fujimoto for his ongoing support for this project.

References

[1] Apodaca AA. Photosurrealism. Rendering Techniques '98 (Proceedings of the Ninth Eurographics Workshop on Rendering), Wien: Springer, 1998. p. 315–22.

[2] Aizu atrium. `http://www.mpi-sb.mpg.de/resources/atrium`. Max-Planck Institute for Computer Science, 2000.

[3] Cornell box. `http://www.graphics.cornell.edu/online/box/`. Cornell University, 1998.

[4] Debevec, PE, Taylor, CJ, Malik J. Modeling and rendering architecture from photographs: a hybrid geometry- and image-based approach. Proceedings of SIGGRAPH 96, August 1996. p. 11–20.

[5] Fairchild MD. Color appearance models. Reading, Massachusetts: Addison-Wesley, 1998.

[6] Fisher A. Tolerances in lighting design. In CIE Central Bureau, editor, Proceedings of the CIE Seminar on Computer Programs for light and lighting, 1992.

[7] Foley JD, van Dam A, Feiner SK, Hughes JF. Computer graphics, principles and practice, 2nd Edition. New York: Addison-Wesley, 1990.

[8] Goral CM, Torrance KE, Greenberg DP, Battaile B. Modelling the interaction of light between diffuse surfaces. Comput. Graphics (Proceedings of SIGGRAPH 84) 1984;18(3):213–22.

[9] Letunov AA, Barladyan BH, Zueva EY, Veznieviec V, Coldatov C. CCD-based device for BRDF measurements in computer graphics. Graphicon '99, 1999. p. 129–35.

[10] Mardaljevic J. Validation of a lighting simulation program under real sky conditions. Lighting Res. Technol. 1995;27(4):181–8.

[11] McNamara A, Chalmers A, Troscianko T, Gilchrist I. Comparing real and synthetic scenes using human judge-

ment of lightness.. In: Peroche B, Rushmeier H, editors. Rendering techniques '00. Berlin: Springer, 2000.
[12] Meyer GW, Rushmeier HE, Cohen MF, Greenberg DP, Torrance KE. An experimental evaluation of computer graphics imagery. ACM Trans. Graphics 1986;5(1):30–50.
[13] Myszkowski K, Kunii TL. A case study towards validation of global illumination algorithms: progressive hierarchical radiosity with clustering. The Visual Comput. 2000;16(5):271–88.
[14] Palik ED. Handbook of optical constants of solids. New York: Academic Press, 1985.
[15] Rushmeier H, Ward G, Piatko C, Sanders P, Rust B. Comparing real and synthetic images: some ideas about metrics. The accompanying test data are posted on the Web under the URL: `http://radsite.lbl.gov/mgf/compare.html`. In Sixth Eurographics Workshop on Rendering, p. 82–91. Eurographics, June 1995.
[16] Sillion F, Arvo JR, Westin SH, Greenberg DP. A global illumination solution for general reflectance distributions. In Computer Graphics (SIGGRAPH '91 Proceedings) 1991;25:187–96.
[17] Tumblin J, Rushmeier HE. Tone reproduction for realistic images. IEEE Comput. Graphics Appl. 1993;13(6):42–8.
[18] Volevich V, Myszkowski K, Khodulev A, Kopylov EA. Using the Visible Differences Predictor to improve performance of progressive global illumination computations. ACM Trans. Graphics 2000;19(2):122–61.
[19] Walter BJ. Density estimation techniques for global illumination. Ph.D. thesis, Cornell University, 1998.
[20] Ward GJ. The RADIANCE lighting simulation and rendering system. Proc. of Siggraph'94, 1994. p. 459–72.