Rendering complex scenes for psychophysics using RADIANCE: How accurate can you get?

Alexa I. Ruppertsberg and Marina Bloj

Department of Optometry, University of Bradford, Bradford BD7 1DP, UK

Rendering packages are used by visual psychophysicists to produce complex stimuli for their experiments, tacitly assuming that the simulation results accurately reflect the light–surface interactions of a real scene. RADIANCE is a physically based, freely available, and commonly used rendering software. We validated the calculation accuracy of this package by comparing simulation results with measurements from real scenes. RADIANCE recovers color gradients well but the results are shifted in color space. Currently, there is no better simulation alternative for achieving physical accuracy than by combining a spectral rendering method with RADIANCE. © 2006 Optical Society of America

OCIS codes: 330.0330, 330.1690, 330.1710, 330.5510, 120.2040.

1. INTRODUCTION

Visual psychophysicists are starting to use computer graphics tools to create more complex stimuli, such as entire scenes incorporating different objects.[1–9] Although this development is welcome, diversifying the standard stimulus set of gratings, Gaussians and Mondrians, use of rendered images makes a tacit assumption: that the elaborate computer graphics image reflects the true physical properties of such a complex stimulus. Even if the rendered image looks convincingly realistic, it may nevertheless depict a situation that is physically inaccurate. However, if we knew that the rendered image reflected the same physical properties as its real counterpart, then we would have a very valuable tool at hand, as we could study human vision with naturalistic stimuli under very, close parametrical control.[10]

There are two issues, however. One is the accuracy with which rendering packages model light transport and material properties; the other is the issue of displaying such an image. The luminance and chromaticity output of a display device, e.g., a computer monitor, is limited. Thus, if the rendered image is supposed to show a bright patch of more than 500 cd/m^2, because its real counterpart would, then a standard computer monitor will not show such brightness. To overcome these brightness display limitations, tone-mapping operators have been developed[11,12] that map the brightness range of the image to the available brightness range of the display. In the best case this will result in a displayed image that *appears* to have a larger brightness range than the display is actually capable of producing. In the future, displays with increased dynamic ranges will become commercially available (high-dynamic-range displays[13]). Until then, simulation results should stay within the capabilities of the display at hand. Similarly, the number of different chromaticity values for a display is limited by the spectra of the phosphors (gamut) and by the resolution of the graphics card, leaving a large number of chromaticity values within the gamut undisplayable.

The calculation accuracy of a rendering package depends on how well the laws of physics are comprised by the simulation program. Within the computer graphics community, extensive validation of new rendering algorithms is a topic of current interest.[14–16] The assessment of the calculation accuracy varies. For simple scenes the results of a simulation can be compared with the analytically determined solution of the lighting equations.[17] Other possibilities are either to compare behavior differences for rendered images versus real objects[10] or to compare the results of a simulation with measurements from a real scene.[15,16,18–21] There are three major problems when comparisons are made between simulation results and measured values from a real scene: (a) inaccuracies in measurements (surface properties, photometric distributions of luminaries); (b) violations of assumptions (e.g., real surfaces do not have Lambertian surface properties but are modeled as such); and (c) the real scene is of limited scope when validation results are to be generalized.

RADIANCE,[22,23] a physically based, freely available rendering package, has been used in the computer graphics,[24–26] architectural, and lighting communities for some time, and is quickly becoming the image generation package of choice for the visual perception community to simulate experimental scenes for psychophysical and behavioral studies.[1–9] RADIANCE has been subjected to validation studies from within the architectural and lighting communities.[17,20,27] These studies focus on luminance accuracy and work with error margins of ±20%, the standard in the industry.[28] The software is predominantly utilized to decide between different design options, i.e., in a comparative mode rather than to predict absolute values.

In this study we validated RADIANCE by comparing the results of simulations with measurements from real scenes, with an emphasis on color and luminance accuracy, and tested whether the results met the accuracy standard for visual psychophysics. We set up real scenes that fit within the gamut of a standard computer monitor (CRT). We used a spectroradiometer for measuring color

signals, ensured that surfaces had Lambertain surface properties, and measured the photometric distribution of the luminaries instead of relying on manufacturer's data. We first describe the setup of our physical scenes with two different illumination modes: simple direct illumination and complex mutual illumination. We then describe the simulation of these scenes in RADIANCE for two color-coding methods: (a) RGB triplets (RGB and sRGB) and (b) spectral data. For the latter we implemented an *N*-step algorithm.[6] We compare the calculated results for both color-coding methods and evaluate the effect of the color accuracy improvement for the human observer in a psychophysical experiment. We conclude with a discussion about the use of RADIANCE as a stimulus generator and exploratory tool in vision research.

2. PHYSICAL SETUP OF SCENES

A. Lighting Booth
The scenes were set up in a lighting booth consisting of a 2.2 m by 2.2 m by 2.2 m frame built from aluminum scaffold tubing (see Fig. 1, left). Three walls were covered with black wool cloth to provide nonreflective surfaces and the floor was covered with dark blue carpet. Low-voltage spotlights (Altman MR 16 Micro Ellipse, with a 75 W, 36 deg reflector) could be hung from the scaffold at any location and angle. Two tungsten-to-daylight conversion filters (LEE Filters, full and half CTB) were placed in front of the spotlights to adjust the intensity and color of the light.

All color signal measurements (radiance in $[W/(sr\ m^2)]$) were taken with a spectroradiometer (Photo Research PR650; measuring aperture 1 deg, from 380 to 780 nm in 5 nm steps) controlled by a standard PC with MATLAB (MathWorks) and were the average of five consecutively measured spectra.

B. Simple Illumination Scene
The simple illumination scene was an example of direct illumination only. Our object was a Macbeth ColorChecker, a card with 24 distinctive color patches, mounted in front of the central wall of the lighting booth at 45 deg in relation to the floor. One spotlight illuminated the card at an angle of 0 deg (illumination angle in relation to the surface normal of the card). We measured the geometric layout of the scene and the color signal of three patches (8, purplish-blue; 14, red; and 15, green) and a white reflectance standard (Spectralon). The measured patch or standard was always in the center of the spotlight beam (see Fig. 1, middle). We will refer to these four scenes as BLUE, RED, GREEN, and WHITE. Our measured luminance values were between 1.7 and 34 cd/m² and all measured colors fell within the gamut of our monitor.

C. Complex Illumination Scene
The complex illumination scene was set up to produce a strong case of indirect (interreflected) mutual illumination. A white cylinder with a diameter of 25.5 cm lay on a green card on top of a table. Two spotlights, mounted 20 cm apart, illuminated the green card at 0 deg (with respect to the surface normal of the card; see Fig. 1, right). This created a color gradient on the bottom half of the cylinder. The green card and the white cylinder had Lambertian surface properties (verified by measurements). We measured the color signal of this gradient by taking 21 samples along a virtual line from where the cylinder touched the green card to 21 cm above the green card in 17 cm steps. The measuring angle with respect to the surface normal of the cylinder changed from around 90 to 0 to approximately −54 deg. Luminance values were between 8.8 and 123.5 cd/m².

Lighting booth   Simple scene setup   Complex scene setup
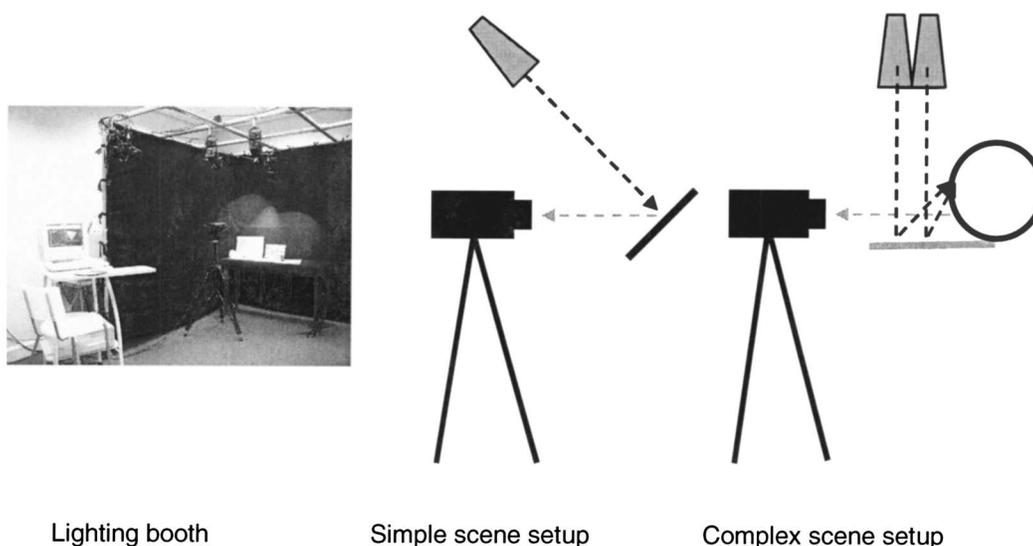
Fig. 1.   Left: Photograph of the lighting booth in which the simple and complex scenes were set up. The measurements were taken with a spectroradiometer. Middle and right: Sketch of the simple and complex illumination scenes. In the simple illumination scene the surface of the Macbeth ColorChecker was illuminated by a single spotlight at 0 deg and the color signal was measured at 45 deg. In the complex illumination scene the surface of a green card was illuminated by two spotlights at 0 deg. The light bouncing off the green card created a green gradient on the white cylinder. The color signal of the cylinder was measured.

3. SIMULATING SCENES IN RADIANCE

We simulated the setup of our lighting booth in RADI-ANCE for both illumination conditions. To model our particular spotlights, we measured the illuminance distribution in 2 deg steps in the horizontal plane using an illuminance meter (Minolta T-10) for both spotlights with the filters in place and then used RADIANCE's ies2rad function.[29] The two spotlights differed only in their lens settings. To establish the color of the illuminants, we used the measured color signals of the WHITE scene for each spotlight. All images were rendered with the default values of RADIANCE except for the number of light bounces, which was set to 5 ($-ab$ 5).[30]

A. Materials

As we had ensured that our objects in the complex illumination scene had Lambertian surface properties, we could use RADIANCE's material type of plastic to simulate our objects without violating this material type's surface properties. This material type was also used for the simple illumination scenes though the Macbeth ColorChecker does not have Lambertian surface properties (own measurements) because we measured only one point in these scenes and the surface normal of the Macbeth ColorChecker does not change.

B. Color in RADIANCE

In RADIANCE the colors of objects and light sources are specified by RGB values. It is straightforward to calculate RGB values from a surface reflectance function $S(\lambda)$, which we derived from our measurements (see Appendix A). Since RADIANCE computes the physical interaction of light and surfaces in radiance [W/(sr m^2)], we followed Yang and Maloney[6] and used radiance values as alternative color descriptors.

When interreflections are present in a scene, the representation of an object's color with three or any number of discrete samples will lead to an underestimation of the intensity of mutual illumination.[31] For both illumination conditions we evaluated RADIANCE for RGB, sRGB,[32] and radiance values.

C. N-step Rendering

For the color descriptors based on radiance values, we implemented an N-step algorithm.[6] For $N=3$ the spectrum is divided into three consecutive, equally spaced wavebands ([380–510 nm], [515–645 nm], and [650–780 nm] in our case) and the average value for each of the wavebands is calculated and used as the corresponding color descriptor. Figure 2 (left) shows a surface reflectance function (dashed curve) and a three-step approximation in black. For $N=9$ the spectrum is divided into nine wavebands and nine average values are calculated (dark gray curve in Fig. 2). To implement a nine-step approximation in RADIANCE, three images are rendered, each image accounting for a different part of the spectrum. Given our measurements, $N=81$ is the best possible approximation and involves rendering 27 images. To make a single displayable image, the information of N images has to be collapsed into a standard three-channel RGB image (method is outlined in Appendix A).
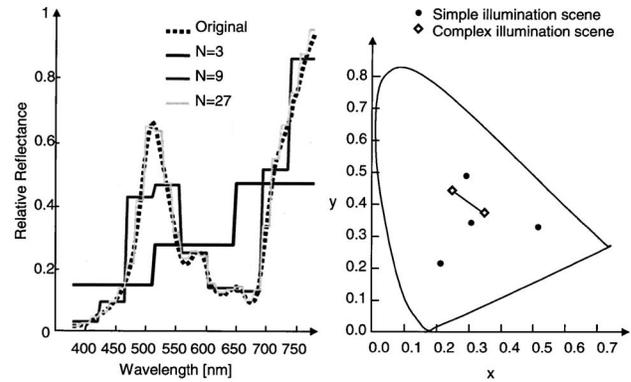
Fig. 2. Left: Example of the N-step algorithm. The reflectance spectrum (Original, dashed curve) is approximated either by 3 (black curve), 9 (dark gray curve), or 27 (light gray curve) steps. Right: CIE x,y chromaticity measurement of the simple (filled symbols) and the complex illumination (open symbols) scenes. The line connecting the open symbols represents all the chromaticity values that were measured along the gradient.

The principle of this algorithm is related to hyperspectral imaging where one image is taken for each waveband, resulting in N images for a scene. Together these N images carry the full spectral information for each single pixel in the image.

Depending on how we code color, the output of RADIANCE has to be interpreted accordingly. If we code color in RGB triplets, RADIANCE's three-value per pixel output corresponds to the RGB color signal of that pixel. If we code color by radiance values, then the resulting three values per pixel output from RADIANCE is interpreted as a three-waveband approximation of the color signal in radiance [W/(sr m^2)]. When more than one image is rendered, each image is an approximation for a particular section of the spectrum, and to recover the entire spectrum, the three values of each consecutive rendered image for that pixel need to be combined.

4. CALCULATION ACCURACY RESULTS

To compare the simulation results with the measured data, we converted all of them into CIE x,y chromaticity and luminance values[33,34] and into CIE L, a, and b values[33] to compute the color difference ΔE. For the simulation results based on RGB triplets, the resulting RGB values were converted to X,Y,Z tristimulus values (with the inverse of matrix T, Appendix A) from which CIE x,y chromaticity values were computed (Appendix B); for the sRGB simulation results a different matrix T was used.[32] The luminance value is identical to the Y value of the XYZ tristimulus value. For the simulation results based on the N-step algorithm, we used approximated matching functions $x'(\lambda)$, $y'(\lambda)$, and $z'(\lambda)$ to yield XYZ tristimulus values and similarly converted them to CIE x,y chromaticity values. The approximated matching functions were derived by applying the N-step algorithm to the original matching functions downloaded from Ref. 35.

Note that this section deals with calculated results only. So far we have made no attempt to display the rendered images on a monitor.

A. Simple Illumination Scene

We simulated three individual color patches (RED, GREEN, and BLUE) from the Macbeth ColorChecker and a white reflectance standard (WHITE). Color was coded either with an RGB triplet, sRGB triplet, or with radiance values. For the radiance values, we implemented the N-step algorithm for four different N's ($N = 3, 9, 27$, and 81) to study the improvement in recovering the original color signal. To evaluate the simulation results of the simple illumination scenes, we used four measures: (1) relative root-mean-square (rms_{rel}) luminance error, measuring the relative deviation from the measured luminance value [Eq. (1), $n = 4$]; (2) the color difference ΔE in CIELAB; (3) the human sensitivity d' to color differences in a psychophysical experiment (see Subsection 4.B); and (4) the graphical representation of the results in CIE x, y chromaticity space. For the simulation results, we averaged across a 5 by 5 pixel central region of the color patch
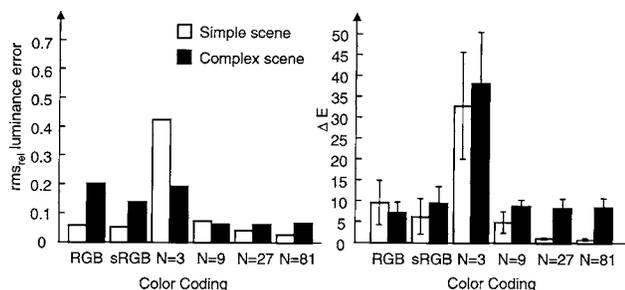
Fig. 3. Left: Relative root-mean-square (rms_{rel}) luminance error for the simple and complex illumination scenes for different color-coding schemes. Right: Mean CIELAB ΔE for the simple ($n = 4$) and complex illumination scenes ($n = 21$) for different color-coding schemes. Bars indicate ± 1 standard deviation.
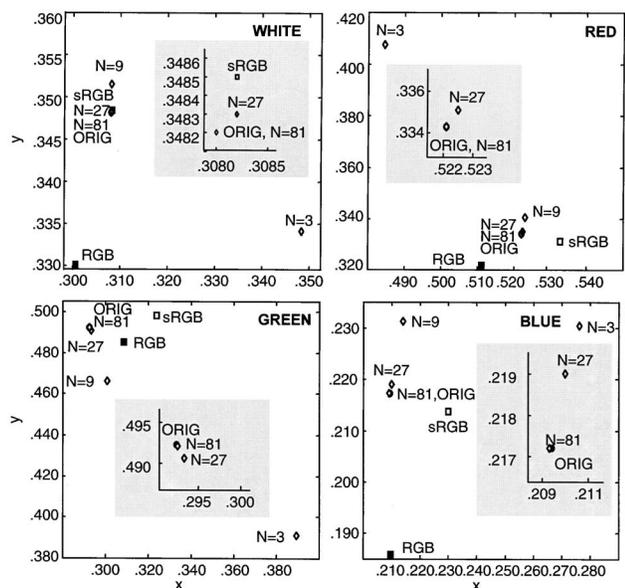
Fig. 4. Top panels: CIE x, y chromaticity results for the WHITE and RED scenes with RGB (filled squares), sRGB (open squares), and N-step coding (open diamonds). The measured value is labeled with ORIG (filled circles). The inset enlarges the area close to the measured value. Bottom panels: CIE x, y chromaticity results for the GREEN and BLUE scenes. Symbols are the same as in the top panels.

that corresponded to the measurement area of 1 deg visual angle of the spectroradiometer.

$$\text{rms}_{\text{rel}} = \sqrt{\frac{\sum_{i=1}^{n} \left(1 - \frac{S_i}{M_i}\right)^2}{n}}. \tag{1}$$

The rms error used here is a relative measure for the deviation between the simulation result S_i and the measured data M_i.

We found that, for all color-coding schemes, except the three-step approximation, the relative rms luminance error was less than 10% (Fig. 3, left, white bars). If the relative rms luminance error were the only measure, then RGB and sRGB coding yield satisfactory results with a 5% deviation. However, the color difference ΔE (Fig. 3, right, white bars) revealed that only spectral rendering with nine or more steps yielded ΔE values of less than 5. By definition, a ΔE value of 1 is a just noticeable difference, and images with an average $\Delta E < 3$ are not discriminable from each other.[36] The 27-step approximation reduced ΔE to 1, whereas the ΔE value for the RGB coding was 9.5. From the three color-coding schemes that rendered only one image (RGB, sRGB, and $N = 3$), sRGB coding yielded the most accurate results with a ΔE value of 6; the three-step approximation result was inadequate ($\Delta E = 32.8$).

To visualize these results we plotted the CIE x, y chromaticity values for all four color-patch simulations together with the original measurements (Fig. 4). For clarification we have inserted enlarged views of the CIE x, y chromaticity space close to the measured values (ORIG).

B. Perceptual Accuracy for Simple Scenes

A first step in establishing whether colors yielded by N-step rendering can actually be distinguished from the color measured in the real scene is to display them on a monitor, converting radiance values into RGB frame buffer values. For a calibrated standard monitor with 24 bit resolution (three guns with 8 bits each), the resulting RGB frame buffer values are identical for $N = 27$, $N = 81$, and the original values measured in the real scene. The only approximations that produced frame buffer values different from the ones corresponding to the original scene were $N = 3$ and $N = 9$.

In a perceptual experiment we tested observers' ability to distinguish between the originally measured color and the three-step and nine-step approximations for each of the four scenes (WHITE, RED, GREEN, and BLUE).

1. Design of Psychophysical Experiment

In an oddity paradigm[37] the task of the observer was to spot the odd stimulus embedded in a series of identical stimuli. In our case, we simultaneously presented stimuli in a pielike arrangement consisting of three 120-deg pieces. The pie display had a diameter of 5.7 deg of visual angle and was based on a similar display configuration used by Wyszecki and Fielder.[38] Two of the three pieces had the same color (see Fig. 5, left). The third piece was the odd stimulus, and the observers had to indicate its position by a keyboard response (left, top, or right). We always compared the original color O with one of the wave-
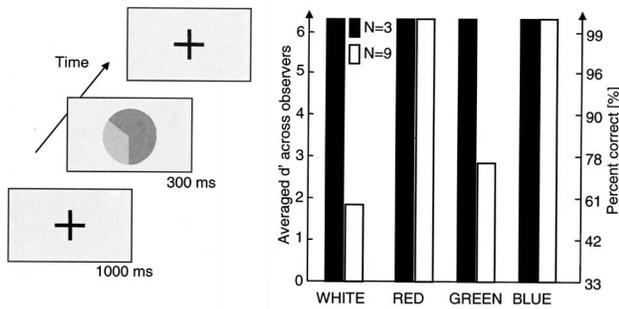
Fig. 5.   Left: Temporal composition of a single trial in the perceptual experiment, starting with a fixation display for 1000 ms, followed by the pie display for 300 ms, and terminated with another fixation display until the observer responded. The odd stimulus in this example is the left pie piece. Right: Results of the perceptual experiment expressed as d' values (left y axis) and percent correct (right y axis) averaged across all observers. Each bar represents 144 trials (36 trials $\times$ 4 observers).

band approximations N, yielding six possible display configurations: $\langle O,O,N \rangle$, $\langle O,N,O \rangle$, $\langle N,O,O \rangle$, $\langle N,N,O \rangle$, $\langle N,O,N \rangle$, and $\langle O,N,N \rangle$. Hence in three trials the approximation was the odd stimulus and in three trials the original was the odd stimulus. Each configuration was repeated six times. We computed the percentage correct from observers' responses and converted it to d' (see Appendix 5 in Ref. 37).

In Fig. 5 (left) the layout of a single trial is shown. A trial started with the presentation of a fixation cross for 1000 ms, followed by the pie display presented for 300 ms, after which the fixation cross reappeared. The observer then reported the position of the odd stimulus, with the next trial starting once a response had been collected.

The mean luminance values of the four scenes ranged between 4.3 and 34.6 cd/m^2. To ensure adequate observer adaptation, only approximations of one particular scene (WHITE, RED, GREEN, or BLUE) were presented in a block. All backgrounds within a block were gray with the same mean luminance as the colors presented. An initial adaptation period of 2 min preceded each block.

The perceptual experiment was run on a standard PC under MATLAB (Mathworks) including use of the Psychophysics Toolbox,[39] and the stimuli were presented on a calibrated CRT monitor. The observer was seated 1 m away from the screen in a dark room. All four observers (the authors and two naïve males) had normal color vision as tested by the Farnsworth–Munsell 100-Hue Test, having made four or less permutation errors.

2. Results
Figure 5 (right) shows the averaged d' values and corresponding percent correct levels for four observers (36 trials per observer and condition). Chance level in this task was at 33%. Observers performed at the 99% correct level for all four scenes when discriminating between the three-step approximation and the original color. However, for the nine-step approximation, 99% correct was achieved only with the RED and the BLUE scenes. Performance dropped for the WHITE (58%) and the GREEN scenes (77%), although it still remained above the chance level. Thus observers could discriminate between all

tested approximations and the original color in this task, confirming the results of the color difference values ΔE.

C. Complex Illumination Scene
In the cylinder scene, color was also coded with either an RGB triplet, sRGB triplet, or with radiance values. We implemented the N-step algorithm for four different N's ($N = 3$, 9, 27, and 81). For the simulation results we averaged across 7 by 5 pixel areas in the rendered image that corresponded to the measurement locations of the cylinder gradient, yielding 21 simulation sample points to compare with 21 measurement points. To evaluate the simulation results of the complex illumination scene, we used three different measures: (1) relative root-mean-square (rms$_{rel}$) luminance error [Eq. (1), $n = 21$]; (2) the color difference ΔE in CIELAB averaged across all 21 sample points; and (3) the graphical representation of the chromaticity and luminance profile of the cylinder gradient in CIE x,y,Y space.

We found the overall relative rms luminance error was increased in comparison with the simple illumination scenes (excluding the nine-step approximation from the simple scenes). Only spectral rendering with nine or more steps led to a luminance error of less than 7% (Fig. 3, left, black bars). However, the error did not improve as more steps were used in the approximations. sRGB coding yielded a luminance error of 14% and RGB coding of 20%. The color difference values ΔE were also increased for the complex illumination scene (Fig. 3, right, black bars). All coding schemes except the three-step approximation yielded a ΔE value of around 8. The lowest ΔE value obtained was for the RGB coding ($\Delta E = 7$). For spectral rendering, increasing the number of steps beyond nine did not lead to an improvement in accuracy.

In Fig. 6 (left) we have plotted the chromaticity profile of the measured cylinder gradient (filled circles with ring) together with the simulation results in CIE x,y chromaticity space. The profile shows the 21 data points corresponding to different heights along the cylinder. The arrows indicate the first data point (0 cm) where the cylinder touched the green card. The chromaticity profile of the measured color signals goes from green to white in a linear fashion. Both RGB coding schemes (open and filled triangles) lead to chromaticity profiles that cross the
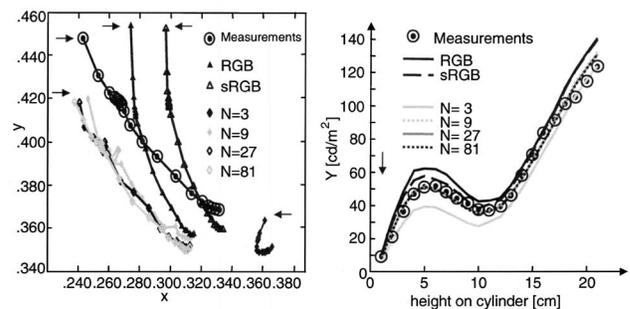
Fig. 6.   Left: CIE x,y chromaticity profile recovery of the complex illumination scene for different color-coding schemes. Arrows indicate the chromaticity where the cylinder touches the green card (0 cm). Right: Luminance profile recovery of the complex illumination scene for different color-coding schemes. Arrow indicates the luminance where the cylinder touches the green card (0 cm).

true profile, with the sRGB profile pushed toward red. The three-step approximation profile (filled diamonds) was further toward red. The nine-step and higher approximation profiles lay together parallel to the true profile, shifted toward blue but in close proximity to the locus of the measurements. The luminance profiles (luminance values over height on cylinder) are shown in Fig. 6 (right) with the measured values as filled circles with rings. The arrow indicates where the cylinder touched the green card. The original measurements showed that the luminance profile of the cylinder from 0 cm first increased (up to 5 cm) as a result of mutual illumination, then dropped (11 cm) as the mutual illumination contribution faded and the shade due to the shape of the cylinder from direct illumination became more relevant. The final increase in luminance was due to direct illumination from the spotlights above. This qualitative behavior was well captured in all simulations. Both RGB coding schemes led to overestimation of the entire profile, whereas the three-step approximation led to underestimation. All nine-step and higher approximation profiles lay together, in extremely good agreement with the measured values.

For our complex illumination scene, accuracy dropped and highly accurate luminance results were obtained only for nine-step or higher approximations (error <7%). Use of more than nine steps did not result in an improvement in luminance accuracy. For color accuracy, all color-coding schemes, except $N=3$, led to low ΔE values (around 8). However, there was a distinct difference in the results. For RGB- and sRGB-based coding, the resulting profile was rotated in CIE x,y space in comparison with the true profile, whereas for the 9-, 27-, and 81-step approximations the profile was merely shifted in comparison with the true profile.

5. DISCUSSION

Complex naturalistic images rendered with RADIANCE are increasingly used by visual scientists to study the human perception of color, surface orientation, illuminant estimation, and surface reflectance properties. While some studies used RGB coding,[1–5] others employed spectral rendering as their color-coding method.[6–9] RGB coding does give correct results if all participating surfaces and lights have flat spectra, or if only one of them is not flat (see Ref. 3 for a more detailed explanation). One group has also employed a tone-mapping algorithm and the addition of artificial glare.[4,5] However, this group is not interested in the spectral accuracy of their stimuli. None of these studies have addressed the question of how well real-world properties were reflected in the rendered scenes. An underlying assumption in these studies is that the human visual system deals with these complex stimuli in just the same way as it would with the real world, because they both depict similar things and are complex. As long as we do not know how accurate the complex stimulus is and how well it reflects real-world properties, use of computer graphics rendered images is not different from using photographs or simple stimuli, such as gratings.

The photometric and colorimetric accuracy is only part of the relationship between the stimulus and the real world. With a physical rendering computer graphics package it is possible to assess this relationship in terms of the photometric and colorimetric accuracy by comparing the simulation results with measurements from real scenes; this approach was pioneered by Meyer and colleagues in 1986.[18] Evaluation studies for RADIANCE[17,20,21,27] have used different validation approaches, which reflected the needs of the architectural community and for which this software was originally developed. In computer graphics the evaluation of RADIANCE reappears as a future goal in Drago and Myszkowski's work[16] when they proposed general validation schemes for global illumination algorithms. The major goal with global illumination algorithms is to attain perceptual accuracy, not necessarily photometric accuracy. From a visual psychophysicist's point of view, a rigorous experimental validation study of RADIANCE with regard to its photometric and colorimetric accuracy was needed. Our results are also informative to the architectural and computer graphics community as this study made few assumptions about material and luminaire properties.

We tested two different illumination conditions and two different color-coding methods. Color descriptors were either RGB triplets or radiance values; the latter were implemented with an N-step algorithm.[6] The main difference from Yang and Maloney's work was that we used measurements of a real scene to compare with the simulation results. A similar method of hyperspectral imaging implemented in RADIANCE was used by Delahunt and Brainard,[8] where a monochromatic image was rendered for each waveband. This approach, however, is less efficient as three times as many images need to be rendered in comparison with the N-step algorithm. Spectral rendering is the chosen technique when accurate renderings are required,[40] and an entire area of research is devoted to finding accurate solutions while reducing the number of calculations.[40–45]

A. Simple Illumination Scenes
With the simple illumination scenes we found that the measured luminance values were well matched by almost all color-coding methods. The lowest relative rms luminance error of 2% was obtained for the 81-step approximation, increasing to 4%, 7%, and 42% for $N=27$, $N=9$, and $N=3$, respectively. RGB and sRGB coding yielded a relative rms luminance error of 5–6%. This is close to the measuring accuracy of the spectroradiometer (4%). Although the luminance accuracy results across different coding schemes seemed fairly homogeneous (except for $N=3$), the chromaticity accuracy showed a different picture. The ΔE values for RGB and sRGB were larger than 5 (9 and 6, respectively), whereas the ΔE values for the nine-step approximation were just below 5, decreasing to 1 for the 27- and 81-step approximations. These low ΔE values hint at the problem we had when trying to display the simulations on a standard computer monitor in the perceptual experiment. The deviations between the simulation results and the measurements were so small that a

standard graphics card (8 bits/channel) was not able to resolve them as different frame buffer values. As color vision scientists we would need at least a resolution of 12 bits/channel to yield different frame buffer values. Even then, the measuring accuracy of the spectroradiometer of ±0.006 in CIE x and y for common CRT phosphors (Manual of PR650) would limit the necessary measuring precision, which also explains why dithering was not a viable alternative.

The psychophysical experiment was an additional test to determine whether any of the N-step approximations was indistinguishable from the original color for human observers. Because of the resolution of the graphics card, we could test only the three- and nine-step approximations. Our perceptual task was deliberately designed to make small color differences visible to the human eye. If observers were not able to distinguish colors under these conditions, then it would be senseless to strive for more accurate calculations. We do not exclude the possibility that there might be other conditions in which these color differences become indistinguishable. In the perceptual experiment we found perfect discrimination for all four scenes when the approximation was based on three steps. This is in line with the computed ΔE values (WHITE, 28.1; RED, 30.5; GREEN, 51.2; BLUE, 21.4). Also, for the nine-step approximations, the computed ΔE values and the results from our perceptual experiment agreed (WHITE, 2.4; RED, 3.6; GREEN, 8.5; BLUE, 5.1). In common with other studies,[36,46] a threshold value of $\Delta E = 2$ is more appropriate for our task. Hence we can predict that observers would have been able to distinguish all simulations based on RGB coding from the original (WHITE, 9.8; RED, 5.5; GREEN, 5.9; BLUE, 17.0). The computed ΔE values for simulations based on sRGB were smaller than those based on RGB coding, and we can predict that observers would have been able to distinguish three scenes based on sRGB coding from the original (RED, 3.7; GREEN, 10.3; BLUE, 9.4), although potentially not the WHITE scene ($\Delta E = 1.6$).

Taken together, for the simple illumination scenes, we have shown that calculation accuracy improves with the number of steps in the approximations and outperforms RGB and sRGB coding. Conversely, this increased accuracy cannot be displayed using standard 8 bit/channel graphics cards.

B. Complex Illumination Scenes
In the complex illumination scene we were interested in how an entire gradient, which has a luminance and a chromaticity profile, was simulated in contrast to a single 1 deg patch in the simple illumination scene. This complex scene is still simple in several ways. The number of involved surfaces is very small, only two artificial light sources were used, no use of RADIANCE's ability to simulate daylight was made, and all involved materials were tested and had diffuse surface properties. To simulate this scene does not test the rendering software to any significant extent, but it is comparable to the kind of renderings used in psychophysical studies employing RADIANCE.[1–9]

We found that the measured luminance values were extremely well matched by the simulation results. The lowest relative rms luminance errors of 6% were obtained for
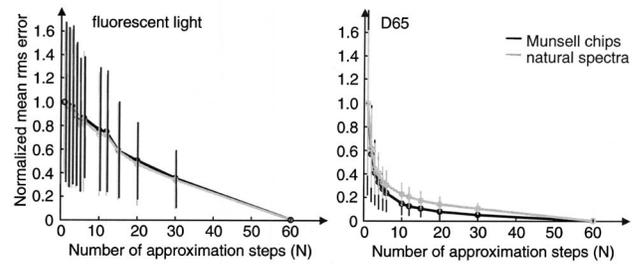
Fig. 7.   Left: Mean rms recovery error for 219 color signals from Munsell matt chips and natural spectra with a fluorescent light source. The mean rms error is normalized to 1 for the one-step approximation. Right: Accordingly for a D65 light source.

the 9-, 27-, and 81-step approximations, showing no further improvement for more than nine steps, whereas sRGB coding yielded a relative rms luminance error of 14%, and RGB and the three-step approximation yielded 20%. Even though the scene complexity is not demanding, the rms error for the sRGB and RGB coding increased by factors of 3 and 4, respectively, compared with the simple scenes.

The color accuracy for the complex scenes was worse than for the simple scenes. None of the color-coding schemes yielded ΔE values of less than 5. All, except $N = 3$, yielded ΔE values of around 8. Similar to the relative rms luminance error, no improvement of color accuracy was found by increasing the number of steps used in the approximations. It is difficult to assess which reconstruction led to the best result, even when the resulting chromaticity profiles in CIE x, y space are taken into account. The simulated gradients based on the N-step approximations ($N = 9$, 27, and 81) lie parallel to the original measurements, whereas the RGB-based simulated gradients cross the original gradient. We were intrigued that no further improvement was seen when more steps were used in the approximations.

In general, one would assume that the rms error for reconstructing a color signal decreases as a function of the number of steps used in the approximation (see Fig. 7). To quantify this assumption, we used two sets of reflectance spectra: Munsell matt chips and natural spectra,[47] each containing 219 spectra and two different light sources—a fluorescent light source and the CIE D65—and we calculated the corresponding color signals. We then reconstructed the color signals with different numbers of steps used in the approximation. Figure 7 shows the mean of the reconstruction error [rms error, Eq. (2), $n = 61$] for the color signals of the reflectance spectra and two light source spectra (CIE D65 and fluorescent light source) as a function of steps used in the approximation. The mean rms error (averaged across 219 color signals) was normalized to 1 for the one-step approximation (which corresponds to the average), as this was the highest expected error. The error rate decrease is mainly dependent on the smoothness of the signal; we see the slowest decrease for both spectra sets under fluorescent lighting, which is known to contain spikes. The error decreases quicker for color signals from the Munsell chips under daylight (D65).

$$\text{rms} = \sqrt{\sum_{i=1}^{n} (S_i - M_i)^2/n};$$

$$S_i = i\text{th simulation result,}$$

$$M_i = i\text{th measurement.} \qquad (2)$$

Given that our light sources were tungstenlike, which are similar in smoothness to CIE D65, and our surfaces diffuse, we would have expected a sharper decrease in the relative rms error for more steps used in the approximations. The observed stagnation of the error might be due to intrinsic limitations of RADIANCE's hybrid approach of Monte Carlo simulation-based and deterministic ray tracing.

6. SUMMARY

We set out to evaluate the physical rendering package RADIANCE[23] in terms of its luminance and color accuracy. To limit the number of potential sources of errors, we set up simple scenes that were under our control and we made sure that diffuse surface properties were not assumed, but actually true. Having taken a set of measurements from our real scenes, we simulated these scenes based on different color-coding schemes to compare the influence of the coding method on accuracy. In particular, we used RGB, sRGB,[32] and N-step coding for radiance values.[6] We evaluated the results in terms of relative rms luminance error [Eq. (1)], the color difference measure ΔE, and with a perceptual experiment.

For luminance accuracy, trichromatic coding schemes (RGB and sRGB) yielded a relative rms luminance error of 20% or less, which is an acceptable value for the practice of the lighting and architectural community. For higher accuracy levels (5%–6%), spectral rendering is recommended. In terms of color accuracy, the calculation accuracy results were all perceptually noticeable ($\Delta E > 2$), even though our complex scene was fairly simple and did not test RADIANCE to its full extent. No large differences between the RGB and sRGB and the spectral rendering simulation results were found for the complex illumination scene. Unlike the simple scenes, calculation accuracy did not improve when more than nine steps were used in the approximation. Therefore we recommend spectral rendering with at least nine steps, but this may need adjustment for different kinds of scenes. An additional benefit of using only spectral information is that all calculations are independent of devices; device-dependent calculations take effect only at the stage of actually displaying an image.

Given our scene, with only a few objects and with verified Lambertian surface properties, RADIANCE's calculation results are shifted in color space. While this may not be a problem for an architect, it could be for a visual psychophysicist. Other alternatives for physically accurate stimuli include use of real objects and lights[48] or hyperspectral images.[49] Each of these approaches has its own merits but also technical complications. For greater flexibility, a simulation package is the best choice. Currently, there is no better simulation alternative for a visual psy-

chophysicist to achieve physical realism (i.e., the simulation result providing the same visual stimulation as a real scene[50]) than by combining a spectral rendering method with RADIANCE.

APPENDIX A: CALCULATING RGB VALUES FROM SURFACE REFLECTANCE FUNCTIONS

We converted a surface reflectance function $S(\lambda)$ to X, Y, Z tristimulus values by using $x(\lambda)$, $y(\lambda)$, and $z(\lambda)$ matching functions (we used the Vos-adjusted CIE 1931 matching functions[34]) and Eqs. (A1)–(A3)[33]:

$$X = k \int S(\lambda)x(\lambda)\mathrm{d}\lambda, \qquad (A1)$$

$$Y = k \int S(\lambda)y(\lambda)\mathrm{d}\lambda, \qquad (A2)$$

$$Z = k \int S(\lambda)z(\lambda)\mathrm{d}\lambda. \qquad (A3)$$

The X, Y, Z tristimulus values can then be converted to RGB with a conversion matrix T, which is based on the primaries of the monitor. T is a 3 by 3 matrix (for a description on how to obtain this matrix, see Refs. 51 and 52 among others). We used RADIANCE's inbuilt monitor primaries to derive the conversion matrix T. Multiplying the XYZ tristimulus values with T yields an RGB triplet, which is then used as the color descriptor in the material file. The same procedure is used to code the color of the illuminant.

APPENDIX B: CIE x, y CHROMATICITY VALUES

From XYZ tristimulus values, CIE x, y chromaticity values are obtained as follows:

$$x = X/(X + Y + Z), \qquad (B1)$$

$$y = Y/(X + Y + Z). \qquad (B2)$$

ACKNOWLEDGMENTS

REFERENCES AND NOTES

1. H. Boyaci, L. T. Maloney, and S. Hersh, "The effect of perceived surface orientation on perceived surface albedo in binocularly viewed scenes," J. Vision **3**, 541–553 (2003).

2. H. Boyaci, K. Doerschner, and L. T. Maloney, "Perceived surface color in binocularly viewed scenes with two light sources differing in chromaticity," J. Vision **4**, 664–679 (2004).

3. K. Doerschner, H. Boyaci, and L. T. Maloney, "Human observers compensate for secondary illumination originating in nearby chromatic surfaces," J. Vision **4**, 92–105 (2004).

4. R. W. Fleming, R. O. Dror, and E. H. Adelson, "Real-world illumination and the perception of surface reflectance properties," J. Vision **3**, 347–368 (2003).

5. R. W. Fleming, A. Torralba, and E. H. Adelson, "Specular reflections and the perception of shape," J. Vision **4**, 798–820 (2004).

6. J. N. Yang and L. T. Maloney, "Illuminant cues in surface color perception: tests of three candidate cues," Vision Res. **41**, 2581–2600 (2001).

7. J. Yang and S. K. Shevell, "Surface color perception under two illuminants: the second illuminant reduces color constancy," J. Vision **3**, 369–379 (2003).

8. P. B. Delahunt and D. H. Brainard, "Does human color constancy incorporate the statistical regularity of natural daylight?" J. Vision **4**, 57–81 (2004).

9. P. B. Delahunt and D. H. Brainard, "Color constancy under changes in reflected illumination," J. Vision **4**, 764–778 (2004).

10. A. Johnston and W. Curran, "Investigating shape-from-shading illusions using solid objects," Vision Res. **36**, 2827–2835 (1996).

11. G. Ward Larson, H. Rushmeier, and C. Piatko, "A visibility matching tone reproduction operator for high dynamic range scenes," IEEE Trans. Vis. Comput. Graph. **3**, 291–306 (1997).

12. J. Tumblin and H. Rushmeier, "Tone reproduction for realistic images," IEEE Comput. Graphics Appl. **13**, 42–48 (1993).

13. H. Seetzen, W. Heidrich, W. Stuerzlinger, G. Ward, L. Whitehead, M. Trentacoste, A. Ghosh, and A. Vorozcovs, "High dynamic range display systems," ACM Trans. Graphics **23**, 760–768 (2004).

14. D. P. Greenberg, S.-C. Foo, K. E. Torrance, P. Shirley, J. Arvo, E. Lafortune, J. A. Ferwerda, B. Walter, B. Trumbore, and S. Pattanaik, "A framework for realistic image synthesis," in *Proceedings of SIGGRAPH 97* (Association for Computing Machinery, 1997), pp. 477–494.

15. K. Myszkowski and T. L. Kunii, "A case study towards validation of global illumination algorithms: progressive hierarchical radiosity with clustering," Visual Comput. **16**, 271–288 (2000).

16. F. Drago and K. Myszkowski, "Validation proposal for global illumination and rendering techniques," Comput. Graph. **25**, 511–518 (2001).

17. A. Khodulev and E. Kopylov, "Physically accurate lighting simulation in computer graphics software," presented at the Sixth International Conference on Computer Graphics and Visualization, St. Petersburg, Russia, 1–5 July 1996; http://www.keldysh.ru/pages/cgraph/articles/pals/.

18. G. W. Meyer, H. E. Rushmeier, M. F. Cohen, D. P. Greenberg, and K. E. Torrance, "An experimental evaluation of computer graphics imagery," ACM Trans. Graphics **5**, 30–50 (1986).

19. A. Takagi, H. Takaoka, T. Oshima, and Y. Ogata, "Accurate rendering technique based on colorimetric conception," Comput. Graph. **24**, 263–272 (1990).

20. J. Mardaljevic, "Validation of a lighting simulation program under real sky conditions," Light. Res. Technol. **27**, 181–188 (1995).

21. G. G. Roy, "A comparative study of lighting simulation packages suitable for use in architectural design," School of Engineering, Murdoch University, Perth, Australia (2000); http://eng.murdoch.edu.au/FTPsite/LightSim.pdf.

22. We will refer to the software package RADIANCE in capital letters and to the physical quantity radiance [W/(sr m2)] in small letters.

23. G. J. Ward, "The RADIANCE lighting simulation and rendering system," in *Proceedings of SIGGRAPH 94* (Association for Computing Machinery, 1994), pp. 459–472.

24. H. Rushmeier, G. Ward, C. Piatko, P. Sanders, and B. Rust, "Comparing real and synthetic images: some ideas about metrics," in *Rendering Techniques '95: Proceedings of the Eurographics Workshop*, P. Hanrahan and W. Purgathofer, eds. (Springer-Verlag, 1995), pp. 82–91.

25. A. McNamara, A. Chalmers, T. Troscianko, and E. Reinhard, "Fidelity of graphics reconstructions: a psychophysical investigation," in *Rendering Techniques '98: Proceedings of the Eurographics Workshop*, G. Drettakis and N. L. Max, eds. (Springer-Verlag, 1998), pp. 237–246.

26. K. Mania, T. Troscianko, R. Hawkes, and A. Chalmers, "Fidelity metrics for virtual environment simulations based on spatial memory awareness states," Presence, Teleoperators Virtual Environ. 12, 296–310 (2003).

27. K. W. Houser, D. K. Tiller, and I. C. Pasini, "Toward the accuracy of lighting simulations in physically based computer graphics software," J. Illum. Eng. Soc. **28**, 117–129 (1999).

28. M. S. Rea, *The IESNA Lighting Handbook*, 9th ed. (Illuminating Engineering Society of North America, 2000).

29. G. Ward Larson and R. Shakespeare, *Rendering with Radiance: The Art and Science of Lighting Visualization* (Morgan Kaufmann, 1998).

30. We compared simulation results for different numbers of light bounces, i.e., changing the parameter $-ab$, and found no difference between three and more bounces.

31. S. A. Shafer, "Shape recovery from interreflection," in *Physics-Based Vision: Principles and Practice: Shape Recovery*, L. B. Wolff, S. A. Shafer, and G. E. Healey, eds. (Jones and Bartlett, 1992), pp. 303–304.

32. M. Stokes, M. Anderson, S. Chandrasekar, and R. Motta, "A standard default color space for the internet—sRGB," Version 1.10 (1996); http://www.w3.org/Graphics/Color/sRGB.

33. G. Wyszecki and W. S. Stiles, *Color Science*, 2nd ed. (Wiley, 2000).

34. J. J. Vos, "Colorimetric and photometric properties of a 2-deg fundamental observer," Color Res. Appl. **3**, 125–128 (1978).

35. See http://www.cvrl.org/database/text/cmfs/ciexyzjv.htm.

36. M. Stokes, M. D. Fairchild, and R. S. Berns, "Precision requirements for digital color reproduction," ACM Trans. Graphics **11**, 406–422 (1992).

37. N. A. Macmillan and C. D. Creelman, *Detection Theory: A User's Guide* (Cambridge U. Press, 1991).

38. G. Wyszecki and G. H. Fielder, "Color-difference matches," J. Opt. Soc. Am. **61**, 1501–1513 (1971).

39. D. H. Brainard, "The Psychophysics Toolbox," Spatial Vis. **10**, 433–436 (1997).

40. R. Hall, "Comparing spectral color computation methods," IEEE Comput. Graphics Appl. **19**, 36–45 (1999).

41. G. W. Meyer, "Wavelength selection for synthetic image generation," Comput. Vis. Graph. Image Process. **41**, 57–79 (1988).

42. C. F. Borges, "Trichromatic approximation for computer graphics illumination models," Comput. Graph. **25**, 101–104 (1991).

43. M. S. Peercy, "Linear color representation for full spectral rendering," in *Proceedings of SIGGRAPH 93* (Association for Computing Machinery, 1993), pp. 191–198.

44. G. M. Johnson and M. D. Fairchild, "Full-spectral color calculations in realistic image synthesis," IEEE Comput. Graphics Appl. **19**, 47–53 (1999).

45. M. S. Drew and G. D. Finlayson, "Multispectral processing without spectra," J. Opt. Soc. Am. A **20**, 1181–1193 (2003).

46. D. H. Brainard, "Color appearance and color difference

specification," in *The Science of Color*, 2nd ed., S. K. Shevell, ed. (OSA and Elsevier Science, 2003), pp. 191–216.

47. See http://spectral.joensuu.fi/databases/index.html.

48. J. M. Kraft and D. H. Brainard, "Mechanisms of color constancy under nearly natural viewing," Proc. Natl. Acad. Sci. U.S.A. **96**, 307–312 (1999).

49. S. M. Nascimento, D. H. Foster, and K. Amano, "Psychophysical estimates of the number of spectral-reflectance basis functions needed to reproduce natural scenes," J. Opt. Soc. Am. A **22**, 1017–1022 (2005).

50. J. A. Ferwerda, "Hi-Fi rendering," presented at the Perceptually Adaptive Graphics Preconference Proceedings, Snowbird, Utah, 26–29 May 2001; http://isg.cs.tcd.ie/campfire/jimferwerda2.html.

51. D. Travis, *Effective Color Displays* (Academic, 1991).

52. J. D. Foley, A. van Dam, S. K. Feiner, and J. F. Hughes, *Computer Graphics: Principles and Practice*, 2nd ed. (Addison-Wesley, 1992).