Comparing Real & Synthetic Scenes using Human Judgements of Lightness

Ann McNamara Alan Chalmers
Department of Computer Science

Tom Troscianko Iain Gilchrist
Department of Experimental Psychology
University of Bristol Bristol
mcnamara@cs.bris.ac.uk

Abstract. Increased application of computer graphics in areas which demand high levels of realism has made it necessary to examine the manner in which images are evaluated and validated. In this paper, we explore the need for including the human observer in any process which attempts to quantify the level of realism achieved by the rendering process, from measurement to display. We introduce a framework for measuring the perceptual equivalence (from a lightness perception point of view) between a real scene and a computer simulation of the same scene. Because this framework is based on psychophysical experiments, results are produced through study of vision from a *human* rather than a *machine* vision point of view. This framework can then be used to evaluate, validate and compare rendering techniques.

1 Introduction

The aim of realistic image synthesis is the creation of accurate, high quality imagery which faithfully represents a physical environment, the ultimate goal being to create images which are perceptually indistinguishable from an actual scene. Rendering systems are now capable of accurately simulating the distribution of light in an environment. However, physical accuracy does not ensure that the displayed images will have authentic visual appearance. Reliable image quality assessments are necessary for the evaluation of realistic images synthesis algorithms. Typically the quality of an image synthesis method is evaluated using numerical techniques which attempt to quantify fidelity using image to image comparisons (often comparisons are made with a photograph of the scene that the image is intended to depict).

Several image quality metrics have been developed whose goals are to predict the *visible* differences between a pair of images. It is well established that simple approaches, such as mean squared error (MSE), do not provide meaningful measures of image fidelity, more sophisticated techniques are necessary. As image quality assessments should correspond to assessments made by humans, a better understanding of features of the **H**uman **V**isual **S**ystem (HVS) should lead to more effective comparisons, which in turn will steer image synthesis algorithms to produce more realistic, reliable images. Any feature of an image not visible to a human is not worth computing. Results from psychophysical experiments can reveal limitations of the HVS. However, problems arise when trying to incorporate such results into computer graphics algorithms. This is due to the fact that, often, experiments are designed to explore a single dimension of the HVS at a time under laboratory conditions. The HVS comprises many complex mechanisms, which rather than function independently, often work in conjunction with each other, making it more sensible to examine the HVS as a whole. Rather than attempting to reuse results from previous psychophysical experiments, new experiments are needed which examine the complex response HVS as a *whole* instead of than trying to isolate features for individual investigations. In this work we study the ability of the HVS to perceive albedo and the impact of rendering quality on *this task*. Rather than deal with atomic aspects of perception, this study examines a complete task in a more realistic setting.

Human judgements of lightness are compared in real scenes, and synthetic images. Correspondence between these judgements is then used as an indication of the fidelity of the synthetic image.

1.1 Lightness Perception

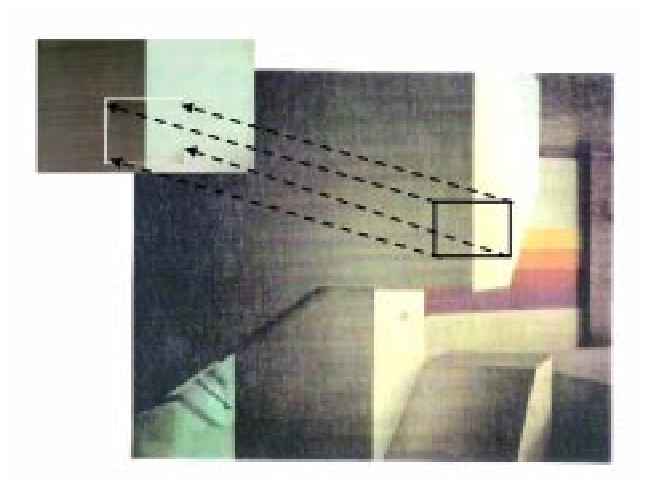

Fig. 1. Importance of depth perception for lightness constancy

Lightness is apparent reflectance, brightness is apparent intensity of the illuminant. Reflectance is the proportion of light falling on an object that is reflected to the eye of the observer. Reflectance (albedo) is constant, the perception of lightness depends of reflectance [1]. Gilchrist [8] showed that the perception of the degree of “lightness” of a surface patch (i.e. whether it is white, gray or black) is greatly affected by the perceived distance and orientation of the surface in question, as well as the perceived illumination falling on the surface - where the latter were experimentally manipulated through a variety of cues such as occlusion, or perspective.

Perception of the lightness of patches varying in reflectance may thus be a suitable candidate for the choice of visual task. It is simple to perform, and it is known that lightness constancy depends on the successful perception of lighting and the 3D structure of a scene, for example figure 1. When viewed in isolation the patches on the top left hand corner appear to be of different luminance. However, when examined in the context of the entire scene, it can be seen that the patches have been cut from the edge of the stairwell, and is perceived as an edge where the entire stairwell has the same luminance. Eliminating the depth cues means the patches are perceived as different, demonstrating the dependency of lightness perception on the correct perception of three dimensional structure, [10]. As the key features of any scene are illumination, geometry and depth, the task of lightness matching encapsulates all three key characteristics into one task. This task is particularly suited to this experimental framework, apart from being simple to perform it also allows excellent control over experimental stimuli. Subsequent sections describe an experimental framework, with such a lightness matching task at the core, to allow human observers to compare real and synthetic scenes.

The remainder of this paper is divided into the following sections. In Section 2, we describe previous research. In Section 3, we describe the steps taken to build the experiment in order to facilitate easy human comparison between real and synthetic scene, we also discuss the actual organisation of participants in terms of scheduling. Section 4 describes the experiment, the results are presented in section 5 and finally, conclusions are drawn in section 6.

2 Previous Work

Models of visual processing enable the development of perceptually based error metrics for rendering algorithms that will reduce the computational demands of rendering while preserving the visual fidelity of the rendered images. Much research investigating this issue is under way.

Using a simple five sided cube as their test environment Meyer et al [13] presented an approach to image synthesis comprising separate physical and perceptual modules. They chose diffusely reflecting materials to built a physical test environment. Each module is verified using experimental techniques. The test environment was placed in a small dark room. Radiometric values predicted using a radiosity lighting simulation of a basic environment are compared to physical measurements of radiant flux densities in the real environment. Then the results of the radiosity calculations are transformed to the RGB values for display, following the principles of colour science.

Measurements of irradiation were made at 25 locations in the plane of the open face for comparison with the simulations. Results show that irradiation is greatest near the centre of the open side of the cube. This area provides the best view of the light source and other walls. The calculated values are much higher than the measurements. In summary, there is good agreement between the radiometric measurements and the predictions of the lighting model. Meyer et al. then proceeded by transforming the validated simulated value to values displayable on a television monitor. A group of twenty experimental participants were asked to differentiate between real environment and the displayed image, both of which were viewed through the back of a view camera. They were asked which of the images was the real scene. Nine out of the twenty participants (45%) indicated that the simulated image was actually the real scene, i.e. selected the wrong answer, revealing that observers were simply guessing. Although participants considered the overall match and colour match to be good, some weaknesses were cited in the sharpness of the shadows (a consequence of the discretisation in the simulation) and in the brightness of the ceiling panel (a consequence of the directional characteristics of the light source). The overall agreement lends strong support to the perceptual validity of the simulation and display process.

Rushmeier et al. [15] used perceptually based metrics to compare image quality to a captured image of the scene being represented. The image comparison metrics were derived from [4],[6], [11]. Each is based on ideas taken from image compression techniques. The goal of this work was to obtain results from comparing two images using these models that were large if large differences between the images exist, and small when they are almost the same. These suggested metrics include some basic characteristics of human vision described in image compression literature. First, within a broad band of luminance, the visual system senses relative rather than absolute luminances. For this reason a metric should account for luminance variations, not absolute values. Second, the response of the visual system is non-linear. The perceived "brightness" or "lightness" is a non-linear function of luminance. The particular non-linear relationship is not well established and is likely to depend on complex issues such as perceived lighting and 3-D geometry. Third, the sensitivity of the eye depends on the spatial frequency of luminance variations. The perceptual metrics derived were used to compare images in a manner that roughly corresponds to subjective human vision, in particular the Daly model performed very well.

The **V**isible **D**ifference **P**redictor (VDP) is a perceptually based image quality metric proposed by Daly [4]. Myskowski [14] realised the VDP had many potential applications in realistic image synthesis. He completed a comprehensive validation and calibration of VDP response via human psychophysical experiments. Then, he used the VDP local error metric to steer decision making in adaptive mesh subdivision, and isolated regions of interest for more intensive global illumination computations. The VDP was tested to determine how close VDP predictions come to subjective reports of visible differences between images by designing two human psychophysical experiments. Results from these experiments showed a good correspondence with VDP results for shadow and lighting pattern masking and in comparison of the perceived quality of images generated as subsequent stages of indirect lighting solutions.

McNamara et al [12] built an experimental framework to facilitate human comparison between real and synthetic scene. They ran a series of psychophysical experiments in which human observers were asked to compare regions of a real physical scene with regions of the computer generated representation of that scene. The comparison involved lightness judgements in both the generated image and the real scene. Results from these experiments showed that the visual response to the real scene and a high fidelity rendered image was similar. The work presented in this paper extends this work to investigate comparisons using three dimensional objects as targets, rather than simple regions. This allows us to examine scene characteristics such as shadow, object occlusion and depth perception.

3 Experimental Design

This section outlines the steps involved in building a well articulated scene containing three dimensional objects placed within a custom built environment to evoke certain perceptual cues such as lightness constancy, depth perception and the perception of shadows. Measurements of virtual environments are often inaccurate. For some applications[1] such estimation of input may be appropriate. However, for these purposes an accurate description of the environment is essential to avoid introducing errors at such an early stage. Also, once the global illumination calculations have been computed, it is important to display the resulting image in the correct manner while taking into account the limitations of the display device. As we are interested in comparing different rendering engines, it is vital that we minimise errors in the model and display stages, this means then that any errors arising can be attributed to the rendering technique employed to calculate the image. This study required an experimental set-up comprised of a real

[1]The level of realism required is generally application dependent. In some situations a high level of realism is not required, for example games, educational techniques and graphics for web design.

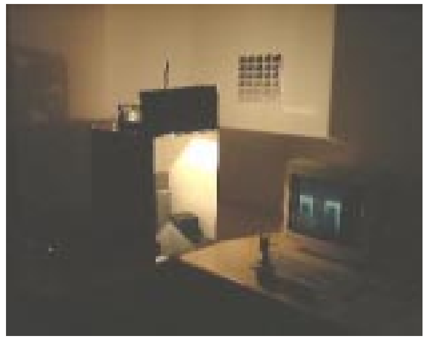

Fig. 2. The test environment showing real environment and computer image.

environment and a computer representation of that three dimensional environment. The measurements required for this study, the equipment used to record them are described herein, along with the rendering process employed to generate the physical stimuli.

3.1 The Real Scene

The test environment was a five sided box shown in figure 2. Several objects that were placed within the box for examination. All interior surfaces of the box were painted with white matt house paint. To accommodate the three dimensional objects, custom paints were mixed, using precise ratios to serve as the basis for materials in the scene. To ensure correct, accurate ratios were achieved, 30ml syringes were used to mix paint in parts as shown in Table 1. The spectral reflectance of the paints were measured using a TOPCON-100 spectroradiometer, these values were transformed to RGB tristimulus values following [16].

Appearance	% White	Reflectance	Patch#	Patch Reflectance
Black	0	0.0471	0	.0494
Dark Gray	10	0.0483	0	.0494
Dark Gray	20	0.0635	2	.0668
Dark Gray	30	0.0779	4	.0832
Dark Gray	40	0.0962	6	.1012
Dark Gray	50	0.1133	7	.1120
Gray	60	0.1383	9	.1224
Gray	70	0.1611	14	.1680
Light Gray	80	0.2002	15	.2259
Light Gray	90	0.3286	19	.3392
Light Gray	95	0.4202	23	.4349
Almost White	97.5	0.5292	26	.5512
Almost White	98.25	0.5312	26	.5512
White	100	0.8795	29	.8795

Table 1. Paint Reflectance along with Reflectance of Corresponding Patch

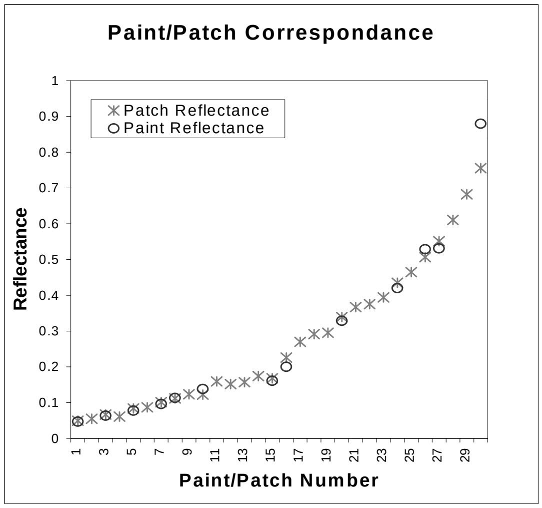

Fig. 3. Correspondence of Patches to Paints

As in [12] a small, front-silvered, high quality mirror was incorporated into the set up to allow the viewing conditions to facilitate alternation between the two settings, viewing of the original scene or viewing of the modelled scene on the computer monitor. When the optical mirror was in position, subjects viewed the original scene. In the absence of the optical mirror the computer representation of the original scene was viewed. The angular sub-tenses of the two displays were equalised, and the fact that the display monitor had to be closer to the subject for this to occur, was allowed for by the inclusion of a +2 diopter lens in its optical path; the lens equated the optical distances of the two displays.

3.2 Illumination

The light source consisted of a 24 volt quartz halogen bulb mounted on optical bench fittings at the top of the test environment. This was supplied by a stabilised 10 amp DC power supply, stable to 30 parts per million in current. The light shone through a 70 mm by 115 mm opening at the top of the enclosure. Black masks, constructed of matt cardboard sheets, were placed framing the screen and the open wall of the enclosure, a separate black cardboard sheet was used to define the eye position. An aperture in this mask was used to enforce monocular vision, since the VDU display did not permit stereoscopic viewing.

3.3 The Graphical Representations

Ten images were considered for comparison to the real scene, they are listed here along with the aims that we hoped to achieve from the comparison.

1. **Photograph**: Comparison to a photograph is needed to enable us to evaluate our method to more traditional image comparison metrics. The reasoning behind this is that most current techniques compare to “reality” by comparing to a captured

image. We wanted to see if this is equivalent to comparing to a real physical environment and so included a photograph, taken with a digital camera, as one of our test images.

2. **Radiance: 2 Ambient Bounces**: A Radiance [17] image generated using 2 ambient bounces is generally considered to be a high quality image. Here we wanted to determine if 2 ambient bounces gives a similar perceptual impression to an 8 ambient bounce image which is more compute intensive.
3. **Radiance: 8 Ambient Bounces**: We wanted to investigate if there was a marked difference using a Radiance image generated using 8 ambient bounces, as this involves considerably more compute time, and might not be necessary i.e. may not provide any more perceptual information than an image rendered using 2 ambient bounces.
4. **Radiance: 8 Ambient Bounces BRIGHT**: This image had its brightness increased manually to see if this affected perception. The brightness was doubled (i.e. the intensity of each pixel was multiplied by 2) to see what, if any effect this had on the perception of the image.
5. **Radiance: Default**: Image generated with the default Radiance parameters. This would determine whether extra compute time makes a significant difference. The default image renders in a very short time, however ambient bounces of light are absent, we wanted to compare this to imagery where interreflections were catered for.
6. **Radiance: Controlled Errors in Estimate Reflectance Values**: The RGB values for the materials were set to equal values to see what difference, if any, this made compared to using measured values. A poor perceptual response to this image would confirm our suspicion that material properties must be carefully quantified if an accurate result is required. This comparison, and the next, was to demonstrate the importance of using exact measurements rather than estimations for material values.
7. **Radiance: Controlled Errors in Estimate of Light Source**: The RGB values for the light source were set to equal values to see what difference this made compared to using measured values. This experiment will show the necessity of measuring emission properties of sources in an environment if an accuracy is the aim.
8. **Radiance: Tone Mapped**: We wanted to investigate the difference tone mapping would make to our test image. Tone mapping transforms the radiance values computed by the rendering engine to values displayable on a display device in a manner that preserves the *subjective* impression of the scene. The Tone Mapping Operator (TMO) used here was introduced by Ferwerda et al. [5]. Although the image examined does not have a very high dynamic range, we were interested to see the effects tone mapping would have on image perception.
9. **Renderpark: Raytraced**: This was a very noisy image generated using stochastic raytracing. This experiment was designed to see how under-sampling would affect perception. Here the effect of under-sampling is exaggerated but might give insights in to how much undersampling a rendering engine can ”get away with” without affecting perceptual performance.
10. **Renderpark: Radiosity**: Finally,to investigate the effects of meshing in a radiosity solution, a poorly meshed radiosity image was used. We wanted to demonstrate the importance of using an accurage meshing strategy when employing radiosity techniques.

These images are shown in the accompanying colour plate.

The media used for stimulus presentation was a gamma corrected 20-inch monitor with the following phosphor chromaticity coordinates:

$$\begin{array}{llll} x_r = 0.6044 & x_g = 0.2808 & x_b = .1520 & x_w = 0.2786 \\ y_r = 0.3434 & y_g = 0.6016 & y_b = .0660 & y_w = 0.3020 \end{array}$$

4 Experiment

Eighteen observers participated in the experiment, and were naive of the purpose of the experiment. All had normal or corrected-to-normal vision. Both condition order and trial order were fully randomised across subjects and conditions. Participants were given clear instructions.

4.1 Training on Munsell Chips

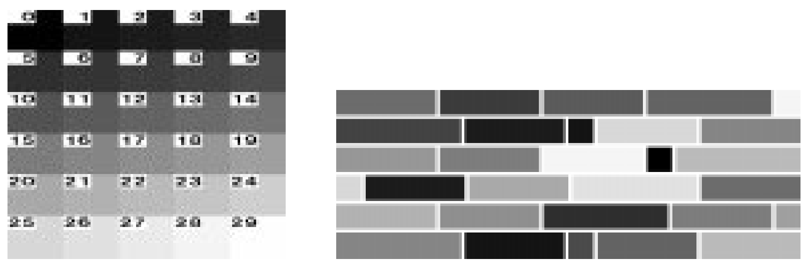

Fig. 4. Patch arrangement used to train participants with Reference Chart)

In [12], the task involved matching regions to a control chart which meant observers had to look away from the scene under examination to choose a match. Moving between scene and chart may affect adaptation to the scene in question, also the view point is not fixed, for this reason we decided to *train* participants on the control patches first. Once trained on the patches participants could then recall the match from memory. Training was conducted as follows. Observers were asked to select, from a numbered grid of 30 achromatic Munsell chips presented on a white background, a sample to match a second unnumbered grid (figure 4) simultaneously displayed on the same background, under constant illumination. The unnumbered grid comprised 60 chips. At the start of each experiment participants were presented with two grids, one an ordered numbered regular grid the other an unordered unnumbered irregular grid comprising one or more of the chips from the numbered grid. Both charts were hung on the wall approximately one meter from the participant. Each participant was asked to match the chips on the unnumbered grid to one of the chips on the numbered grid on the left. In other words they were to pick a numbered square on the left and place it right next to the grid on the right which in the grid would match it exactly. This is done in a random manner, a laser pointer [2] was used to point to the unnumbered chip under examination. Then the numbered chart was removed, and the unnumbered chart replaced by a similar chart but one where the chips had a different order. Participants repeated the task, this time working from memory to recall the number each chip would match to. The results of this training exercise are graphed in figure 5. The graph on the left shows the average

[2] non-invasive medium

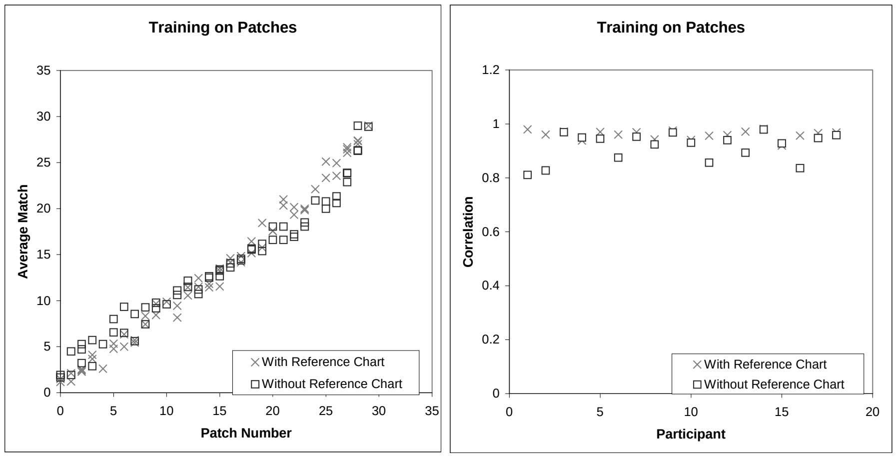

Fig. 5. Average of Matching to Training Patches with and without the reference chart shown on the right along with the Average Correlation for both cases on the left

match across 18 subjects, both with the reference chart and without the reference chart. The graph on the right shows the average correlation. This correlation gives an indication of the extent to which two sets of data are linearly related. A values close to 1 indicates a strong relationship, while a value of 0 signifies there is no linear relationship. A correlation of 1 would result if the participant matched each unnumbered patch to its corresponding numbered patch, in reality this is not the case and some small errors are made, what we need to determine is if the errors made when matching from memory i.e. without the chart are about the same size as the errors made with the reference chart in place. The correlation value when matching the patches with the chart in place is 0.96, and when matching from memory the result is 0.92, indicating a very small difference of 0.04 between the two conditions. From this small difference we can conclude that participants are *just as good* at matching the patches without the reference chart in place. Thus, this training paradigm proved to be reliable and stable. This has the dual benefit of speeding up the time taken per condition, as well as ensuring participants do not need to move their gaze from image to chart, thus eliminating any influence due to adaptation.

4.2 Matching to Images

Each participant was presented with a series of images, in a random order, one of which was the real environment. Participants were not explicitly informed which image was the physical environment. The images presented were the real scene, the photograph and the 9 rendered images. There were 17 different objects in the test environment, subjects were also asked to match the 5 sides of the environment (floor, ceiling, left wall, back wall and right wall) giving a total of 21 matches. The paints used on the objects match to the training patches as shown in graph 3, and detailed in table 3.1. Participants were asked to judge the lightness of target objects in a random manner.

We chose this particular task - that of matching materials in the scene against a display of originals - because the task has a number of attractive features. First, Gilchrist [9, 7] has shown that the perception of lightness (the perceptual correlate of reflectance) is strongly dependent on the human visual system's rendition of both illumination and 3-D geometry. These are key features of perception of any scene and are in themselves complex attributes. However, the simple matching procedure used here depends critically on the correct representation of the above parameters. Therefore, the task should be sensitive to any mismatch between the original and the rendered scene. Secondly, the matching procedure is a standard psychophysical task and allows excellent control over the stimulus and the subject's response. The task chosen here corresponds closely to the methodology of Gilchrist [2, 9, 7] which permits simple measures (of lightness) to be made at locations in complex scenes. Ultimately, the task was chosen to be simple while also being sensitive to perceptual distortions in the scene.

5 Results

Results for each participant were recorded and analysed independently. The value (or gray level) chosen by each participant in the real scene was compared with the values chosen in the rendered image. For a rendered image to be a faithful reproduction, the values in both cases should be closely related. To examine this relationship we carried out a linear correlation for each subject. This correlation gives an indication of the extent to which two sets of data are linearly related. A values close to 1 indicates a strong relationship, whilst a value of 0 signifies there is no linear relationship. A correlation of 1 would result if the participant chose exactly the same gray level for each object in the real scene and rendered image. Correlation values are shown in table 2, and graphed as shown in the colour plate, the graph on the right shows these values averaged.

To examine *the pattern* of these correlations across participants we carried out **AN**alysis **O**f **VA**riance (ANOVA). ANOVA is a powerful set of procedures used for testing significance where two or more conditions are used, here 10 conditions were examined [3]. A *repeated measures within subjects* ANOVA was used. There was a significant effect of condition:

$$F(9, 153) = 80.3; p < .001$$

This equation can be read as follows, the F statistic equals 80.3, with 9 degrees of fredom (10 images), 153 degrees of freedom for the error term (calculated as a function of image combinations). The P value indicates the probability that these differences occur by *chance*. This is a repeated measures within subjects analysis of variance as each subject performed each condition.

This means there are statistically reliable differences between the conditions. This is to be expected as some images were deliberately selected for variation in quality.

The ANOVA showed there are significant differences in perception across images. Further analyses were carried out to investigate where these differences occur. These analyses took the form of a paired comparison t-test. Here we took the correlation between the real scene and the photograph, and compared it to the correlation of the real scene to the other images. Results from the correlations are shown in the following table.

Image	Mean Correlation with REAL
Photograph	.8918
* 2 Ambient Bounces	.843
8 Ambient Bounces	.884
Brightened 8 Ambient Bounces	.865
* Default	.337
* Controlled Error Materials	.692
Tone Mapped	.879
Controlled Error Illumination	.862
* Raytraced	.505
* Radiosity	.830

Table 2. Comparison of Rendered Images to Real Environment

A star in the table indicates a statistically significant difference, reflecting a reliable decrement in quality when compared to the photograph. The significant t values were as follows:

Two Ambient Bounces: $(t(17) = 3.11; p < .01)$
Default Image: $(t(17) = 12.4; p < .001)$
Guessed Materials Image: $(t(17) = 10.7; p < .001)$
Raytraced Image: $(t(17) = 9.36; p < .001)$
Radiosity Image: $(t(17) = 3.00; p < .01)$

The t statistic equals (take Two Ambient Bounces as an example) 3.11, with 17 degrees of freedom (18 participants). The probability, p of this distribution happening by chance is less than 0.01. This means that while there are some small differences between the results of matching to the photograph and matching to other images, these differences are not significant.

In summary, our results show that there is evidence that the 2 Ambient Bounces image, the Default image, the Controlled Error Materials image, the Raytraced image and the Radiosity image are perceptually degraded compared to the photograph. However, there is no evidence that the others images in this study are perceptually inferior to the photograph. From this we can conclude that the 8 Ambient Bounces image, the Brightened 8 Ambient Bounces image, the Tone Mapped image and the Controlled Error Illumination image are of the same perceptual quality as a photograph of the real scene.

6 Conclusions

We have introduced a method for measuring the perceptual equivalence between a real scene and a computer simulation of the same scene, from a lightness matching point of view. Because this model is based on psychophysical experiments, results are produced through study of vision from a human rather than a machine vision point of view.

By conducting a series of experiments, based on the psychophysics of lightness perception, we can estimate how much alike a rendered image is to the original scene. Results show that given a real scene and a faithful representation of that scene, the visual response function in both cases is similar.

Because the complexity of human perception and the computational expensive rendering algorithms that exist today, future work should focus on developing efficient methods from which resultant graphical representations of scenes yield the same perceptual effects as the original scene. To achieve this the full gamut of colour perception, as opposed to simply lightness, must be considered by introducing scenes of increasing complexity.

References

1. E. H. Adelson, Lightness Perception and Lightness Illusions, 339–351, MIT Press, 1999, pp. 339–351.
2. J. Cataliotti and A. Gilchrist, *Local and global processes in lightness perception*, Perception and Psychophysics, vol. 57(2), 1995, pp. 125–135.
3. H. Coolican, *Research methods and statistics in psychology*, Hodder and Stoughton, Oxford, 1999.
4. S. Daly, *The visible difference predictor: an algorithm for the assessment of ima ge fidelity*, In A. B. Watson Editor, Digital Images and Human Vision, MIT Press, 1993, pp. 179–206.
5. J. A. Ferwerda, S.N. Pattanaik, P. Shirley, and D. P. Greenberg, *A model of visual adaptation for realistic image synthesis*, Computer Graphics **30** (1996), no. Annual Conference Series, 249–258.
6. J. Gervais, Jr. L.O. Harvey, and J.O. Roberts, *Identification confusions among letters of the alphabet*, Journal of Experimental Psychology: Human Perception and Perfor mance, vol. 10(5), 1984, pp. 655–666.
7. A. Gilchrist, *Lightness contrast and filures of lightness constancy: a common explanation*, Perception and Psychophysics, vol. 43(5), 1988, pp. 125–135.
8. A. Gilchrist, S. Delman, and A. Jacobsen, *The classification and integration of edges as critical to the perception of reflectance and illumination*, Perception and Psychophysics **33** (1983), no. 5, 425–436.
9. A. Gilchrist and A. Jacobsen, *Perception of lightness and illumination in a world of one reflectance*, Perception **13** (1984), 5–19.
10. A. L. Gilchrist, *The perception of surface blacks and whites*, Scientific American **240** (1979), no. 3, 88–97.
11. J. L. Mannos and D. J. Sakrison, *The effects of a visual criterion on the encoding of images*, IEEE Transactions on Information Theory **IT-20** (1974), no. 4, 525–536.
12. A. McNamara, A. Chalmers, T. Troscianko, and E. Reinhard, *Fidelity of graphics reconstructions: A psychophysical investigation*, Proceedings of the 9th Eurographics Rendering Workshop, Springer Verlag, June 1998, pp. 237–246.
13. G. W. Meyer, H. E. Rushmeier, M. F. Cohen, D. P. Greenberg, and K. E. Torrance, *An Experimental Evaluation of Computer Graphics Imagery*, ACM Transactions on Graphics **5** (1986), no. 1, 30–50.
14. K. Myszkowski, *The visible differences predictor: Applications to global illumination problems*, Rendering Techniques '98 (Proceedings of Eurographics Rendering Workshop '98) (New York, NY) (G. Drettakis and N. Max, eds.), Springer Wien, 1998, pp. 233–236.
15. H. Rushmeier, G. Ward, C. Piatko, P. Sanders, and B. Rust, *Comparing real and synthetic images: Some ideas about metrics*, Eurographics Rendering Workshop 1995, Eurographics, June 1995.
16. D. Travis, *Effective color displays*, Academic Press, 1991.
17. G. J. Ward, *The RADIANCE lighting simulation and rendering system*, Proceedings of SIGGRAPH '94 (Orlando, Florida, July 24–29, 1994) (Andrew Glassner, ed.), Computer Graphics Proceedings, Annual Conference Series, ACM SIGGRAPH, ACM Press, July 1994, ISBN 0-89791-667-0, pp. 459–472.