

☞ Piazza SS. Annunziata, Florence. The image represents a study of the Italian piazza created using Lightscape and AutoCAD.

Courtesy of Andrezej Zarzycki, Sommerville, Mass., 1998.



[Donald P. Greenberg]

*HOW TO GENERATE SYNTHETIC IMAGES
WITH ENOUGH FIDELITY TO BE TRULY ACCURATE
REPRESENTATIONS OF REAL-WORLD SCENES.
NOT JUST AMAZINGLY APPEALING IMAGERY.*

A FRAMEWORK FOR REALISTIC IMAGE SYNTHESIS

Our goal at the Cornell Program of Computer Graphics is to develop physically based lighting models and perceptually based rendering procedures that produce synthetic images visually and measurably indistinguishable from real-world images. Fidelity of the physical simulation is the primary concern.

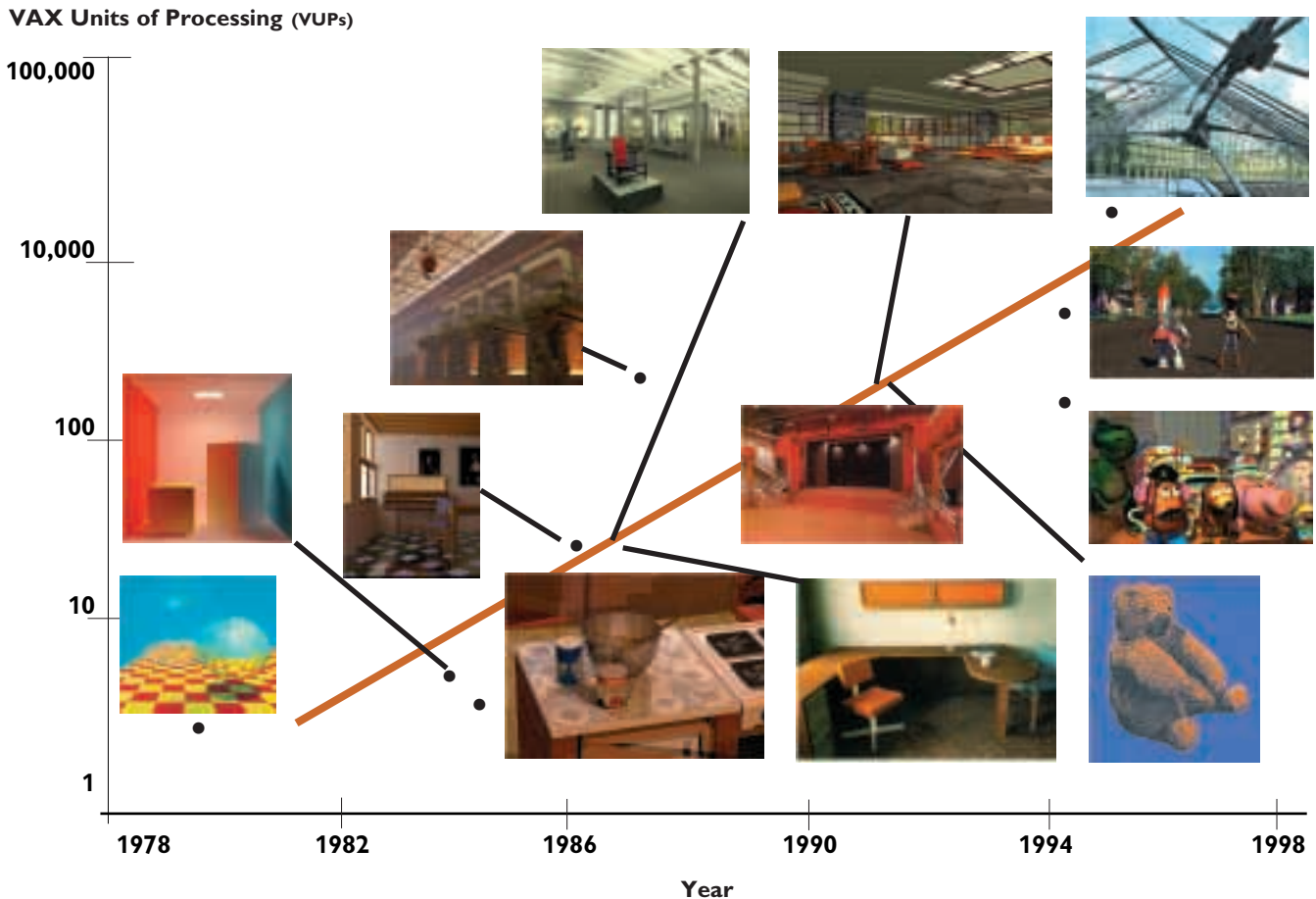
Here, I emphasize the formal comparisons between simulations and actual measurements, the difficulties algorithm designers and scientists encounter building light-reflection and light-transport models, and the need to tap the vast amount of psychophysical research conducted over the past 50 years, as well as future research directions. We

hope our research helps establish a more fundamental, scientific approach toward developing rendering algorithms.

Although the earliest computer graphics renderings, in the late 1960s, involved simple environments with direct lighting, the graphics community today generates pictures of complex

Figure 1. Runtime (in VAX Units of processing)

The estimated polygonal complexity of several familiar models plotted against the year the images were generated. Each scene can be considered complex for the time it was rendered. The number of polygons in complex scenes is increasing exponentially with time.



scenes with shadows, shading, textures, and inter-reflections. For decades, high-quality simulations have been used for a number of industrial tasks, including pilot training, automotive design, and architectural walkthroughs. The entertainment industry has concentrated on developing techniques for creating startling special effects and realistic simulations with dramatic results. Even today's low-cost virtual reality games use amazingly convincing imagery.

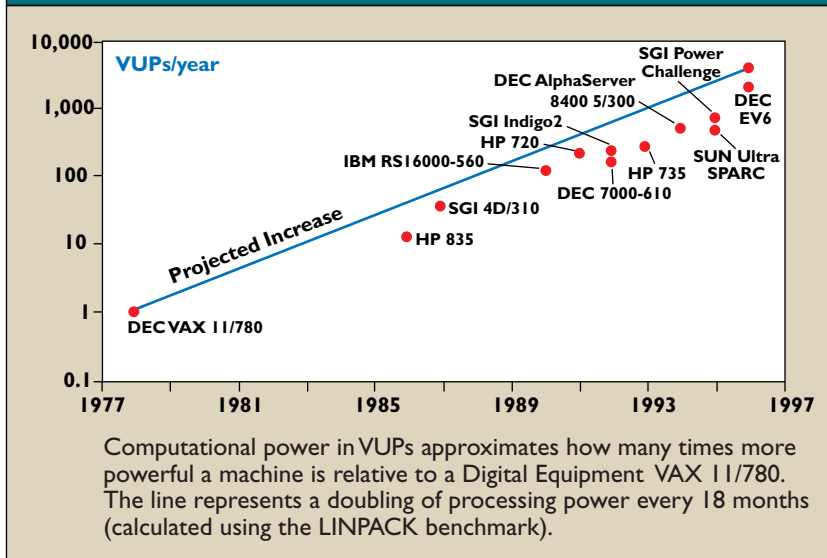
But are these images correct? Would they represent the scene accurately if the environment actually existed? In general, the answer is no; yet the imagery is convincing and appealing because the resulting pictures are believable.

However, if we could generate simulations guaranteed to be correct—where the algorithms and resulting images were truly accurate representations—the graphics simulations could be used *predictively*. Using

such simulations to predict reality is the holy grail of computer graphics. It also represents a major paradigm shift for the computer graphics industry, as such ability will have much broader applicability than just picture making.

A look at how accurate simulations are used in other areas might clarify this hypothesis. The entire electronics industry now depends on simulations for chip design, especially for testing and design modifications prior to fabrication. The same is true for vehicle design, engine performance, and crash-worthiness evaluations in the automotive industry. Why not also use computer graphics algorithms for the testing and development of printing technologies, photographic image capture, and display devices? Why not use these accurate but artificial scenes for developing algorithms in image processing, robotics, and machine vision? If we knew the simulated images were correct, we could

Figure 2. Approximate computational power of common computers plotted against the year they were introduced.



readily control and isolate design variables, obtain any desired precision or resolution, and avoid the tedious, time-consuming, expensive nature and constraints of experimental measurements.

But to be predictive, the simulations must first be proved to be correct. This difficult task requires a major multidisciplinary effort among physicists, computer scientists, and perception psychologists, as well as experimental measurements and comparisons. Unfortunately, relatively little work has sought to correlate the results of computer graphics simulations with real scenes. However, with more accurate image acquisition and measurement devices today, these comparisons are now achievable—if we can generate physically accurate computer simulations (see Fitzmaurice et al.’s “Sampling, Synthesis, and Input Devices” in this issue).

From early computer-generated images, such as the Phong goblet in 1975, to today’s synthesized pictures, the complexity of the visual environments rendered by computer scientists, entertainment and advertising producers, and architecture firms has grown exponentially (see Figure 1). Increased complexity has been accompanied by exponential growth in computational costs for realistic rendering. Fortunately, the available processing power has also increased exponentially.

According to Moore’s Law, with a doubling of chip density every 18 months, we now have approximately 10,000 times the processing power available when the first graphics algorithms were implemented (see Figure 2). There has also been a concomitant increase in memory capacity, offsetting constraints on environ-

ment complexity, as well as significant reduction in cost per compute cycle. A look ahead promises a combination of increasing computational power and algorithmic improvements that will allow us to compute images that are physically and perceptually correct.

How do we get there? I start by describing in general terms the field’s long-term development efforts, particularly those at Cornell, in attempting to achieve these results. I also want to encourage the computer graphics community to develop physically based algorithms of great realism and fidelity. Although there are many frontiers for future research in computer graphics, for physically based realistic image synthesis, three areas are

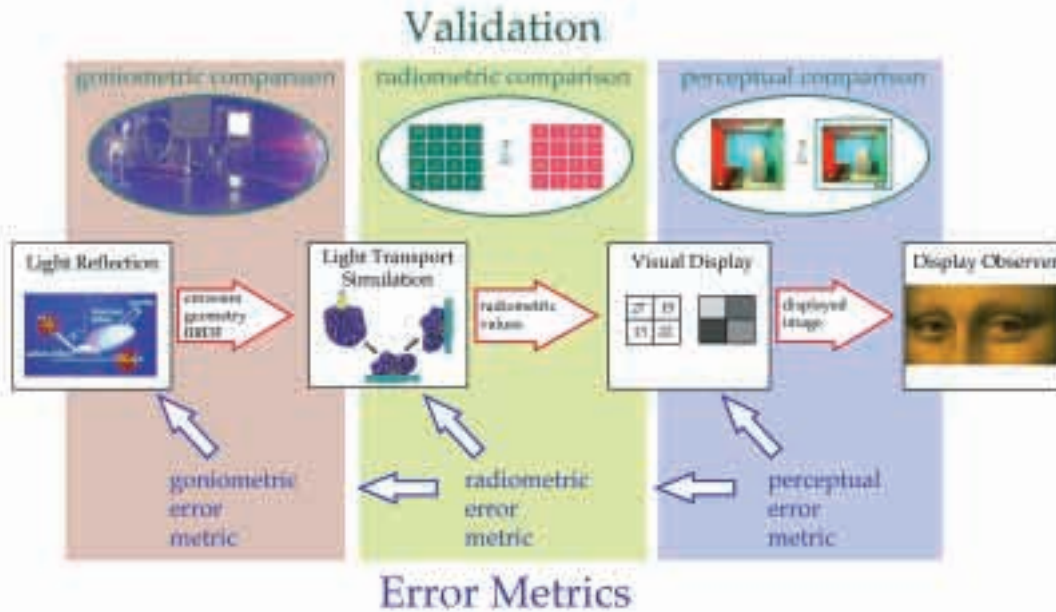
critical: local light reflection models, light transport simulation, and perceptually based issues.

For more than a decade at Cornell, we have been developing a system to test, validate, and improve the fidelity and efficiency of computer graphics algorithms (see Figure 3). The system is organized into three subsections, or stages, dealing with the local light reflection model, the global light transport simulation, and the image display. Of paramount importance is that at each stage, simulations are compared with measured experiments.

For the first stage, we want to derive an accurate, physically based local light reflection model for arbitrary reflectance functions. For the past eight years, we have assembled a measurement laboratory to “goniometrically” measure and compare the local reflection model with a large number of samples. If the simulation model is correct, accurate data can be sent—in terms of geometry, emission, and reflectance functions—to the next stage.

With this information, rendering algorithms still have to accurately simulate the physical propagation of light energy throughout the modeled environment, or digitally encoded geometry, including cubes, spheres, and higher-order curved surfaces. This model of the physical world is sometimes simple, but often very complex, with millions of polygonal surfaces, or facets, to evaluate for lighting. For arbitrary reflectance functions and complex geometries, current simulation procedures are computationally excessive. Most global illumination algorithms use simplifying assumptions; although we create images of spectacular quality, none guarantees physical accuracy. However,

Figure 3. The system's three stages: local light reflection model, global light transport simulations, and image display procedures.



The squares represent processes and simulations and the ovals comparison measurements; the arrows contain only data. Each stage is related to itself and its predecessor stage by providing a feedback loop to reduce computations, resolution, and representations when below a given error threshold.

if it were feasible to simulate these reflection and transport processes, we could measure and compare the resulting radiometric scene values.

Two factors in the process of rendering accurate realistic images deserve special emphasis: That the first two of our stages deal only with physically based simulations, and that at the end of these two stages, we have not yet created a “picture.” We are still only comparing measured and simulated radiant energy on an image plane with full dynamic range and infinite resolution.

If the results of the first two physical stages are accurate, we can proceed to the third stage of creating and comparing images *perceptually*. Since any comparison has to incorporate the human vision system, this stage occurs entirely in the perceptual domain. The computational processes has to account for the limited dynamic range, limited spatial resolution, and limited color gamut of the display or printing devices. But the “display mappings” of calculated luminance (as a wavelength value to RGB values in the range supported by the output device) must also account for the viewer’s position and focus, state of adaptation, and the vast, complex, and mostly unknown relationships among the scene’s spatial, temporal, and chromatic attributes.

A major benefit of this research will be to reduce the computational expense of global illumination algorithms, thus improving their efficiency. An inherent cause of the slowness of these algorithms is that they spend too much time computing scene features that are measurably unimportant and perceptually below the visible thresholds of the average human observer. Algorithms could be substantially accelerated if we develop error metrics that correctly predict the perceptual thresholds of scene features. These techniques would allow not only realistic visual display, but a feedback loop for reducing the magnitude of the physical computations.

Light Reflectance Models

Light reflectance models have always been of great interest to the computer graphics community. The most commonly used model was derived about 25 years ago at the University of Utah [9]. Designed at a time when processing was expensive, the Phong direct lighting model is a clever scheme but was an erroneous representation; it is neither accurate in representing the true reflection behavior of surfaces nor is it energy consistent. In fact, it most closely represents the behavior of one material—hard plas-

tic. The arbitrary nature in which the specular (mirror-like) and diffuse (rough) reflection coefficients are assigned and their associated energy is not physically correct. Yet the entire graphics industry is based on these early formulations, and all major graphics hardware manufacturers today use the same computationally efficient but overly simplified shading model.

What really happens when light of a certain wavelength, coming from a certain incoming direction, strikes a surface? How much is absorbed? How much is reflected? How much energy is sent in each of the many scattering directions? The attribute describing this physical behavior is called the “bidirectional reflectance distribution function,” or BRDF, and is for any material a function of five parameters: the

In complex scenes, the computation of the BRDF function and the visibility along the hemispherical directions are computationally expensive. A more formidable task is computing the solutions for complex environments with realistic reflection functions and accurate visibility along all incoming and outgoing directions (for the outgoing radiances at all points in the scene). Any point in a scene can potentially receive energy from any other point, directly or indirectly. Most usable algorithms make simplifying assumptions for the BRDF function, as well as for the visibility computation and for the solution of the integral equation representing the light transport. Yet they still produce images of startling quality and realism.

The two most common methods are ray tracing, introduced to the graphics community in 1979 [12],

ALTHOUGH WE CREATE IMAGES OF SPECTACULAR QUALITY, NONE GUARANTEES PHYSICAL ACCURACY.

incoming wavelength of light, two surface roughness properties, and the incoming and outgoing directions.

For the past 30 years, much research has focused on developing more accurate light reflection models [1, 3]. Today it is possible to accurately simulate the reflection phenomena with a physically based optics model yielding all three of a surface’s major reflectivity components: specular, directional diffuse, and uniform diffuse reflections. Although the model is computationally expensive, new compact representational schemes have been derived to accurately describe the BRDF’s dominant behavior. The functions capture the diffuse, directional diffuse, and specular characteristics, including off-specular peaks; they are also energy consistent and obey the laws of reciprocity. Furthermore, the representation method is suitable for progressive algorithms, monotonically converging to a correct solution [8].

Light Transport

Once the emission, geometry, and reflection functions are known, we can simulate the light transport. The general equations are well known [7], but until recently, neither the processing power nor the physically based reflection models were available for performing accurate simulations.

and radiosity, introduced in 1984 [5]. Although the past 15 years have seen many improvements, neither of these commonly used algorithms is sufficiently accurate, as each neglects various and significant mechanisms of light transport.

Ray tracing. View-dependent ray tracing methods originally computed only some of the transport paths, but accounted for specular transport in a visually compelling manner. In essence, ray tracing reduced the BRDF expression to include only the transport path in the specular direction, thus simplifying the computations but ignoring diffuse-diffuse and specular-diffuse interactions. Although subsequent modifications can account for additional physical effects, in practice, the simulations still require a great deal of computation time. And since the algorithms are view-dependent, every time the observer (camera) moves, the full computational cycle has to be repeated.

Radiosity methods. View-independent radiosity-type solutions are traditionally computed by boundary element methods, interleaving the computation of the global light transport and the local lighting representation. In essence, these approaches model the transport processes by determining the “form-factor,” or percentage of illumination leaving one surface element and reaching another. Using typical assumption

Figure 4. The goal of realistic image synthesis (example from photography) [10].



of diffuse (Lambertian) reflection, computations are based on geometric relationships, including shape, size, orientation, distance, and occlusion.

Although computationally expensive, due to the complex visibility problem in real scenes, simultaneous equations can obtain surface radiosities. For complex environments, it is not practical to explicitly solve the full set of energy equations. Most solutions compute a portion of the global transport iteratively and update local reflection values until reaching some convergence criteria [2]. To create high-quality images, the requirement of very fine local representations, particularly in areas of high illumination gradients, such as shadow boundaries, gives rise to an exponential increase in elements. This combination of operations involving high global and high local complexity generally causes an explosion in resource consumption in terms of memory and time.

Much research has gone into improving the basic boundary element method and reducing the time

required at the expense of additional data structures and more memory use [6]. Consequently, computer memory and display technology usually limits the maximum input size and solution quality.

A major advantage of these radiosity approaches is that once the illumination of the scene is computed, the results are independent of the observer's position. Realistic simulations can then be displayed using standard graphics hardware accelerators in real time since the lighting calculations do not have to be repeated. However, no one has yet produced a closed-form solution; accurate solutions are computed only through Monte Carlo particle tracing methods for statistically simulating the propagation of light energy through millions of samples. Still, by using a sufficient number of particles, albeit at major computational expense, these statistical simulations of the reflectance function and the global light transport approach physical accuracy.

Despite these impressive advances in reducing com-

putational tasks, most still involve the inherent constraint of being applicable in only diffuse environments and static scenes. More general solutions are needed for complex geometric environments with arbitrary reflectance functions.

Perception

In addition to physical accuracy, a major goal of realistic image synthesis is to create an image that is perceptually indistinguishable from an actual scene. In Figure 4, which is not a trick photograph, the man is holding a real physical image generated by the rules of photographic tone reproduction.

Generating a visual image is the final stage of realistic image synthesis. At the end of the light transport process, we have a global illumination solution representing the radiometric values at every point in a 3D scene. The final stage in image synthesis involves mapping these simulated scene radiances to produce a visual image. This mapping process is an underappreciated yet important part of the image synthesis process; it has to account for the physical characteristics of the display device, the perceptual characteristics of the observer, and the conditions under which the image will be viewed.

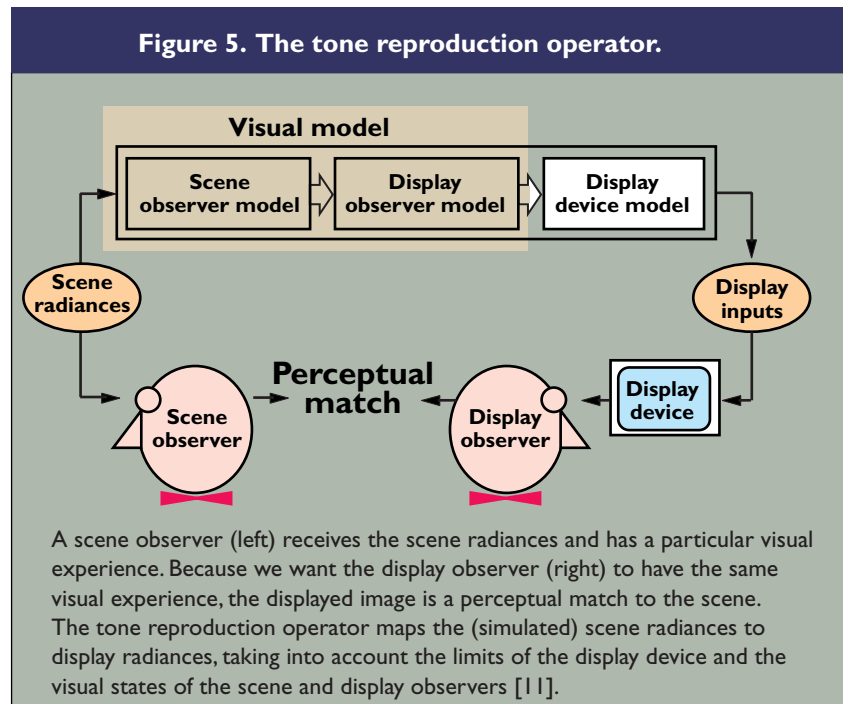
While the physically based rendering methods described earlier make it possible to accurately simulate the radiometric properties of visible scenes, such physical accuracy does not guarantee the images displayed at the end of the process will have a realistic visual appearance. There are two reasons for these inadequate results. Display devices are limited in a number of ways, including spatial resolution, temporal resolution, absolute and dynamic luminance range, and color gamuts. Moreover, the scene's observer and the display observer may be in very different visual states, influencing how they perceive the displayed visual information.

Display technologies place fundamental limits on the fidelity of the display process. In the spatial domain, displays have fixed addressability and resolution and are bounded in extent. In the temporal domain, they have fixed refresh rates and discrete update intervals. In luminance, the absolute and dynamic ranges producible on displays are both small relative to the ranges that can be measured in real

scenes. Finally, with respect to color, the displays are trichromatic with limited color gamuts. The fact that display devices work as well as they do in creating acceptable visual representations of scenes is due to the fact that the human visual system is as limited as it is.

Human Visual Function

For the past 150 years, psychophysicists have measured the characteristics of human visual function for the average human observer. The contrast sensitivity function, which plots the spatial transfer properties of vision, and the temporal response properties



of the visual system are now well known.


Temporal response properties indicate that at high illumination levels, the limit of flicker sensitivity is approximately 75–80Hz. In the luminance domain, the threshold-vs.-intensity functions show the relationship between just noticeable differences (JND) in intensity and the background illumination level. Over a wide range of illumination levels, the visual system obeys Weber's Law, which holds that the size of the JND is a constant proportion of the background level. In color vision, the shapes and sizes of the MacAdam ellipses on the standard chromaticity diagram (as established by the Commission Internationale d'Éclairage) indicate that color discrimination is not uniform within the spectral gamut, but varies with chromaticity and with the direction of the chromatic difference. Lastly, our visual acuity decreases dramatically with the distance from the fovea, or central visual

field. All of these factors are important criteria for determining the resolution needed for the computation and display of synthetic images.

Imaging system designers have used visual models for decades to improve the quality and reduce the bandwidth and computational load of imaging systems. In photography, subjective tone reproduction and preferred gamma curves incorporate Weber's Law and simultaneous contrast effects. In color printing, knowledge of the trichromatic nature of human vision leads to full-color reproduction from a limited number of inks, while awareness of spatial integration in vision has led to halftoning and color dithering techniques. Designers of simulation systems have taken advantage of differences in resolution across the visual field to reduce the level of detail for objects outside the focal region. And designers of image coding and compression systems, such as NTSC, JPEG, and MPEG, have used the

visual processing, we can develop mappings—from simulated scene radiances to display radiances—to produce images as realistic as possible. Our goal in realistic image synthesis is to show that these images can predict what an observer standing in the physical scene would see. Validation of the predictive aspect of the images is a key component of the framework. Models of visual processing would also allow us to create perceptually based error metrics for rendering algorithms that reduce the computational demands of rendering while preserving the visual fidelity of the rendered images.

In the Cornell research framework, we use the idea of a tone reproduction operator introduced in 1993 by Tumblin [11] (see Figure 5). The oval in the figure represents the scene radiances simulated by the light transport algorithms. A hypothetical scene observer receiving these radiances has a particular visual experience. On the right of the figure is a



THE FACT THAT DISPLAY DEVICES WORK AS WELL AS THEY DO IN CREATING ACCEPTABLE VISUAL REPRESENTATIONS OF SCENES IS DUE TO THE FACT THAT THE HUMAN VISUAL SYSTEM IS AS LIMITED AS IT IS.

spatial, temporal, and chromatic limits of human vision to determine bandwidths and quantization levels for visual features of different scales, choose refresh rates and motion prediction methods for image sequences, and guide the choice of color coding schemes. We are only beginning to take similar advantage of visual perception in realistic image synthesis [4, 11].

Improving the visual realism of synthetic images requires continued exploitation of the psychophysics literature for visual models that can be applied in computer graphics. A better understanding of the spatial, temporal, chromatic, and 3D properties of vision will certainly lead to more realistic and more efficient graphics algorithms.

To produce realistic images, we need to model not only the physical behavior of light, but the parameters of perceptual response as well. By modeling the transformations that occur in the brain during

display observer looking at a display device driven by a graphics frame buffer. Because the goal in realistic image synthesis is to give the display observer the same visual experience as the scene observer, the tone reproduction operator maps the simulated scene radiances to the display radiances with the goal of producing a perceptual match between the display and the scene. There are two major components to this mapping: The first is a model of the physical transfer properties of the display device, including information about the display's absolute and dynamic range limits, gamma correction factors, monitor white point, and color gamut. The second is a visual model of the scene and the display observers.

An accurate visual model is the essential component of a tone reproduction operator, allowing us to characterize the visual states of the scene and display observers and relate them to determine the mapping

from simulated scene radiances to display radiances.

Because the tone reproduction operator produces a perceptual match between the image and the scene, the images can be used predictively. Images produced this way can be used quantitatively in such industrial simulation applications as illumination engineering, transportation and safety design, and visual ergonomics.

To be able to claim that images generated by Cornell-developed visually based tone-reproduction operators are predictive, we still need comparison experiments as validation. The results of such experiments would allow us to tune the visual models so the images we create are truly predictive. Furthermore, an experimentally validated visual model would allow us to use the model in lieu of actual comparison experiments for developing perceptually based error metrics. These error metrics, along with the previously determined physically based error metrics, would allow us to create more realistic and efficient image synthesis algorithms. If the end product of a simulation is a visual image, an efficient “steepest ascent” path can be derived to obtain a high-fidelity visual solution with fewer computational demands.

We are just beginning this work, but predictive visual models of such phenomena are clearly at the heart of future advances in computer graphics. Our task is difficult; there are complex interactions between apparent reflectance, apparent illumination, and 3D spatial organization that dramatically affect our perceptions of identical visual stimulus. These interactions have implications for object recognition, color constancy, and other higher-order visual phenomena. The quantitative aspects of these relationships are still not well understood.

Conclusion

Global illumination approaches and algorithms are not yet practical, because they require excessive computational resources and hence excessive time. They are, however, yielding important scientific insights into the physical processes of light reflection and light transport, helping us pinpoint the related computational bottlenecks. And because they are physically correct, they have been used for simulating radiant heat exchange in turbine design, canopy detection in aerial reconnaissance, theater lighting, and architectural and automotive design. With computing power increasing exponentially, global illumination algorithms will eventually become the norm. We hope our research will provide a better scientific foundation for future rendering algorithms.

Although this work is derived primarily from research at one university—Cornell—the effort is

expanding to other universities, and there have been significant contributions by private and national laboratories. This effort should now be expanded to a greater portion of the computer graphics community. Only through greater focus on these issues can we hope to improve the fidelity of our rendering algorithms. **C**

REFERENCES

1. Blinn, J. Models of light reflection for computer synthesized pictures. In *Proceedings of ACM SIGGRAPH'77* (San Jose, Calif., July 20–22). ACM Press, New York, 1977, pp. 192–198.
2. Cohen, M., and Greenberg, D. The HEMI-CUBE, a radiosity solution for complex environments. In *Proceedings of ACM SIGGRAPH'85* (San Francisco, July 22–26). ACM Press, New York, 1985, pp. 31–40.
3. Cook, R., and Torrance, K. A reflectance model for computer graphics. In *Proceedings of ACM SIGGRAPH'81* (Dallas, Tex., Aug. 3–7). ACM Press, New York, 1981, pp. 307–316.
4. Ferwerda, J., Pattanaik, S., Shirley, P., and Greenberg, D. A model of visual adaptation for realistic image synthesis. In *Proceedings of ACM SIGGRAPH'96* (New Orleans, Aug. 4–9). ACM Press, New York, 1996, pp. 249–258.
5. Goral, C., Torrance, K., and Greenberg, D. Modeling the interaction of light between diffuse surfaces. In *Proceedings of ACM SIGGRAPH'84* (Minneapolis, July 23–27). ACM Press, New York, 1984, pp. 213–222.
6. Hanrahan, P., Salzman, D., and Aupperle, L. A rapid hierarchical radiosity algorithm. In *Proceedings of ACM SIGGRAPH'91* (Las Vegas, July 28–Aug. 2). ACM Press, New York, 1991, pp. 197–206.
7. Kajiya, J. The rendering equation. In *Proceedings of ACM SIGGRAPH'86* (Dallas, Tex., Aug. 18–22). ACM Press, New York, 1986, pp. 143–150.
8. LaFortune, E., Foo, S., Torrance, K., Greenberg, D. Non-linear approximation of reflectance functions. In *Proceedings of ACM SIGGRAPH'97* (Los Angeles, Aug. 3–8). ACM Press, New York, 1997, pp. 117–126.
9. Phong, B.-T. Illumination for computer-generated images. *Commun. ACM* 18, 6 (June 1975), 311–317.
10. Stroebel, L., Compton, J., Current, I., and Zakia, R. In *Photographic Materials and Processes*, L. Stroebel, J. Compton, I. Current, and R. Zakia, Eds. Focal Press, Boston, 1986, pp. 420.
11. Tumblin, J., and Rushmeier, H. Tone reproduction for realistic images. *IEEE Computer Graphics and Applications* 13, 6 (Nov. 1993), 42–48.
12. Whitted, T. An improved illumination model for shaded display. *Commun. ACM* 23, 6 (June 1980), 343–349.

DONALD P. GREENBERG (dpg@graphics.cornell.edu) is director of the Program of Computer Graphics at Cornell University, Ithaca, N.Y.

This work was supported by the National Science Foundation Science and Technology Center for Computer Graphics and Scientific Visualization (ASC-8920219), NSF ASC-9523483, and NSF CCR-9401961. These sources also supported the work that led to a special session and paper at SIGGRAPH 1997, with the same title as this article, by D. Greenberg, K. Torrance, P. Shirley, J. Arvo, J. Ferwerda, S. Pattanaik, E. LaFortune, B. Walter, S.-C. Foo, and B. Trumbore.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, to republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.
