Three Varieties of Realism in Computer Graphics

James A. Ferwerda
Program of Computer Graphics
Cornell University

ABSTRACT

This paper describes three varieties of realism that need to be considered in evaluating computer graphics images and defines the criteria that need to be met if each kind of realism is to be achieved. The paper introduces a conceptual framework for thinking about realism in images, and describes a set of research tools for measuring image realism and assessing its value in graphics applications.

Keywords: computer graphics, realism, human vision, visual perception

Figure 1: Homage to Magritte: "The Treason of Image Synthesis". *This* is not a pipe either.[6]

1 INTRODUCTION

In 1922 Rene Magritte created a painting titled "The Treason of Images" that showed a wooden pipe with the caption "Ceci n'est pas une pipe." (This is not a pipe.). Aside from being provocative, the point that Magritte was trying to make is that an image of a thing is not the same as the thing itself, and we should be careful not to confuse the two. Making an image involves processes of selection, approximation, and abstraction, and this is as true of computer-generated images as it is of paintings, so if Magritte were alive today he might have made the image shown in Figure 1 to remind us that *this* is not a pipe either. The key insight, is that an image is a visual *representation* of a scene, in that it "re-presents" selected properties of the scene to the viewer with varying degrees of realism.

Realistic image synthesis has been one of the major research directions in computer graphics since its inception. The pursuit of realism has been both a seductive goal and an important driving problem for the field, inspiring many significant breakthroughs in modeling, rendering, and display algorithms. Yet even with these advances, making compelling realistic images is still more of an art than a science, and the process is computationally expensive, so some researchers have questioned both the need for and the value of realism in many graphics applications.[5,7,24]

A key part of this controversy is that there are no agreed-upon standards for measuring the realism of computer-generated images. Sometimes physical accuracy is used as a criterion; at other times perceptual criteria are applied; and finally, under many conditions an ad-hoc “looks-good” standard is used. To be able to measure the realism of images produced by different graphics techniques, and to be able to evaluate whether realism adds value in different application domains, objective standards are needed.

This paper describes three standards of realism that need to be considered in evaluating computer graphics images, and defines criteria that need to be met if each kind of realism is to be achieved. The goal of this paper is to provide a conceptual framework for thinking about realism in images, and to describe a set of research tools that can be used to measure image realism and assess its value in graphics applications.

2 THREE VARIETIES OF REALISM IN COMPUTER GRAPHICS

In her groundbreaking book, Margaret Hagen[9] introduced the concept that different methods of depiction produce different *varieties of realism* in which certain properties of a scene are accurately represented, and others are approximated, abstracted, or omitted. A key part of this concept is the idea that representational pictures can be realistic in some respects and not in others. For example, Magritte’s painting might realistically represent the pipe’s true shape, yet it is likely that it only approximately represents the pipe’s true material properties. Although Hagen focused primarily on how the geometric aspects of scenes are represented by images, it is possible to extend her basic concept and identify three varieties of realism in computer graphics that differ in the level of visual coding at which realism is defined. They are:

physical realism - in which the image provides the same **visual stimulation** as the scene:
photo-realism - in which the image produces the same **visual response** as the scene: and
functional realism - in which the image provides the same **visual information** as the scene.

Notice that each of these standards uses different criteria to determine if an image is realistic, and therefore each places different demands on the image generation process. Together they provide a set of objective benchmarks that can be used to evaluate the realism of a wide range of computer graphics techniques. In the following sections the fundamental concepts behind each of these standards will be described in greater detail, and their implications for computer graphics will be explored.

3 PHYSICAL REALISM

The first standard listed above is *physical realism*. Here the criterion for realism is that the image has to provide the same *visual stimulation* as the scene. If we neglect optical filtering and scattering in the eye, this means that the image has to be an accurate point-by-point representation of the spectral irradiance values at a particular viewpoint in the scene. This places strict demands on the image generation process. First, the model must contain accurate descriptions of the shapes, materials, and illumination properties of the scene. Next, the renderer must be able to accurately simulate the spectral and intensive properties of the light energy arriving at the observer’s viewpoint. Finally, the display device must be able to accurately reproduce these energies. Although physically-based image synthesis methods can achieve the first two goals, conventional displays cannot, in general, reproduce the rendered light energies, so creating physically realistic *images*, is currently impossible except under restricted conditions.

Despite this limitation, developing physically accurate image synthesis techniques has been a popular goal in computer graphics research for at least the past 20 years. Although part of the appeal is probably the intellectual challenge of “doing it right”, physically-based methods do have some concrete applications. Because the digital images can be accurate numerical simulations of light reflection and transport, they can be used for quantitative analysis in a wide range of design and engineering applications. Example application domains include illumination engineering, material science, and manufacturing.[12,26]

However, adopting physical realism as the standard for generating *observable* realistic images has a number of drawbacks. First, as mentioned above, in most cases the images are not realizable on existing displays. Second, physically-based image synthesis is extremely computationally expensive, which limits its applicability in interactive graphics applications. And finally, physical realism is overkill if one’s job is to create images for human observers, since the standard doesn't take the limitations or capabilities of vision into account. These considerations have led to the second standard for realism in computer graphics: photo-realism.

4 PHOTO-REALISM

When we speak of photo-realism in computer graphics, we usually mean that we want to create an image that is indistinguishable from a photograph of a scene. This is a fine goal but it begs the question of realism since it doesn't explain why a photograph is realistic. Although this is largely an unanswered question that has puzzled psychologists for more than a century, at least one way to move toward a concrete definition of photo-realism is to say that the image has to be *photo-metrically* realistic. Photometry is the measure of the eye's response to light energy, so this definition requires that the image has to produce the same *visual response* as the scene even though the physical energy coming off the image may be different than the scene.

Adopting this criterion allows us to take the observer's visual system into account in the image generation process, and in particular it allows us to take advantage of the limitations of vision to simplify the task of making realistic images. This standard for realism is not new, and in fact it is the tacit assumption behind color imaging technology, that takes advantage of the trichromatic nature of vision to reduce the requirements for describing colors from their full spectral representations to their metameric RGB or CMYK equivalents. Recently, graphics researchers have started to exploit other aspects of vision to create images that are photo-realistic according to the visually-based definition given above.

One major problem in realistic imaging is tone-reproduction. The problem is that existing displays often cannot reproduce the vast ranges of light energy found in different scenes. By developing models of how the visual system adapts to these ranges, researchers have been able to design algorithms that reproduce the appearance of these scenes within the limitations of the display devices (see Reinhard et al.[19] for a recent review). A potential benefit of this visually-based approach is that with proper calibration, the images can be predictive visual simulations that accurately show what an observer would see if they were in the scene. If the visual models can be validated, then the images could be used could be used for quantitative *visual* analysis in a wide range of design and engineering applications.

Another problem that has benefited from a visually-based definition of photo-realism is the issue of how to increase the efficiency of the image synthesis process. Standard physically-based rendering algorithms can be inefficient because they may spend time computing image features that will be invisible. This has limited the utility of these algorithms for interactive graphics applications since it may take minutes or hours to render a single image. However, by taking advantage of the limits of contrast sensitivity in complex scenes, researchers have been able to develop more efficient perceptually-based rendering algorithms that only compute image features precisely enough to make them indistinguishable from physically-correct solutions (Myszkowski et al.[16] and Yee et al.[27] provide recent reviews of this work). The hope is that this approach will eventually allow photo-realistic rendering techniques to be used in a wide range of applications.

Although adopting photo-realism as the standard for realism in computer graphics has advantages over using purely physical metrics, it also has a number of limitations. First, even though perceptually-based algorithms can be faster than their physical counterparts, in their complete forms they are still too slow for interactive applications and it is unclear how to take further advantage of the low-level visual models on which they are based. The approach of reducing the complexity of the visual models they may increase performance somewhat, but it does so at the cost of undermining the image's value as a predictive simulation, and it may also reduce image quality by allowing artifacts to become visible. While this is an important issue, there are two even larger problems with photo-realism. First, it is unclear that photo-realism is necessary or even desirable in a wide range of graphics applications, and second, adopting photo-realism as a standard for visual realism in computer graphics, classifies most renderings as failures, yet says nothing about their obvious utility in many application domains.

5 FUNCTIONAL REALISM

These considerations suggest a third standard for realism in computer graphics and that is *functional realism*. Here the criterion for realism is that the image has to provide the same *visual information* as the scene. Information here means knowledge about the meaningful properties of objects in a scene, such as their shapes, sizes, positions, motions and

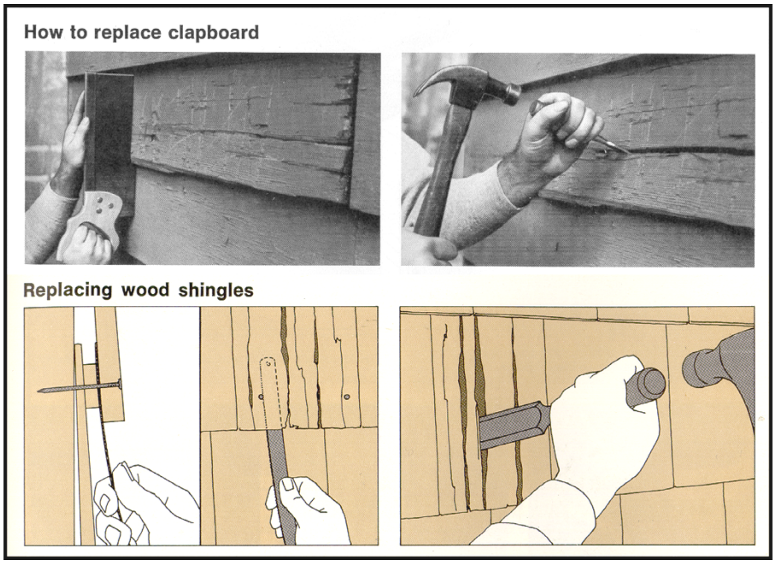

Figure 2: Functional realism in technical illustrations. Adapted from [18].

materials, that allows an observer to make reliable visual judgments and to perform useful visual tasks. Realism is defined in terms of the *fidelity* of the information the image provides. If an image lets you do the task you need to do, and allows you to perform the task as well as you could in the real world, then for that task, the image is realistic. The beauty of a functional definition of realism is that it admits a wide range of rendering styles from physically-based simulation through photo-realism, to more abstract approaches such as non-photorealistic rendering.[8,21,23,24] Some examples may make these concepts clearer.

One example of functional realism in computer graphics is the images used in flight simulators. These images aren't realistic by either of the standards described earlier. Typically, they aren't physically accurate simulations, nor are they photo-realistic renderings, but they are functionally realistic in that they provide the observer with much of the same visual information that they would receive if they were flying a real plane. The proof of the realism of these images is that they allow the observer to learn skills that then transfer into the real world. Although this is a good example of functional realism in computer graphics, it could be argued that there might be some advantage to using photo-realistic images in this setting, however the next example illustrates that this is not always the case.

Figure 2 presents two illustrations from the Reader's Digest Complete Do-It-Yourself Manual[18] that show how to replace the siding on a house. It is interesting to compare these two renderings and ask which provides better information about how to do the job. The photos are certainly clear enough, but are they preferable to the drawings? In terms of providing information, the drawings offer a number of benefits over the photos. First, the drawings can eliminate irrelevant details produced (in this case) by shading, shadows, and surface texture. Second, the drawings can facilitate visual segmentation and grouping by color (e.g. hand white, tools gray, wood tan). Third, the drawings make it possible to show viewpoints that would be difficult or impossible in a photograph, such as the edge shot shown in the first panel or the point-of-view shot shown in the second. Finally, the rendering of the saw blade shows that the drawings can make use of "special effects" like artificial transparency to depict important features that would be hidden in photographs.

What these examples show is that there are potentially many different rendering styles that can produce images that provide useful information to human observers. In part this is why the development of technologies like photography have not eliminated the art of illustration. A good illustration may actually be better at conveying information to an observer than a physically accurate or photo-realistic image. The challenge this observation raises however, is how to develop metrics of functional realism that can be used to evaluate the effectiveness of different rendering styles.

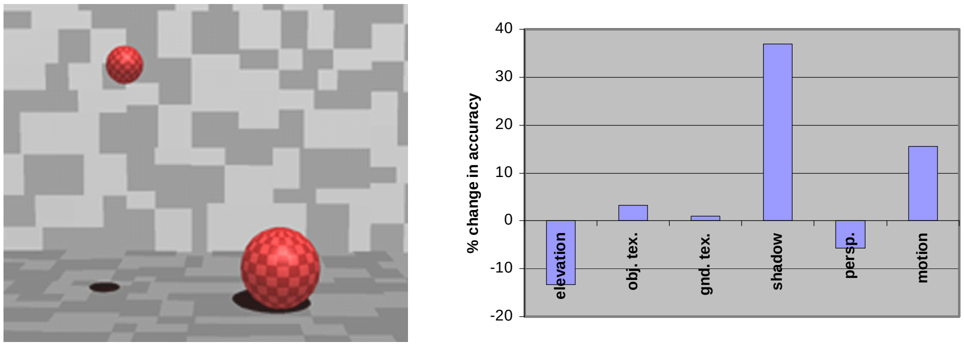

Figure 3: Measuring functional realism in computer graphics. Adapted from [25].

6 MEASURING FUNCTIONAL REALISM

So how can we develop metrics of functional realism in computer graphics? There are probably many answers to this question, but one concrete approach is to explore the relationship between *accuracy* and *fidelity* in computer graphics images. Here accuracy refers to the correctness of the image with respect to some physically measurable property of the scene such as radiance. Fidelity on the other hand means: does the image tell the truth? Does it allow the observer to perceive important properties of the scene with the same certainty that they could in the real world? Although it is possible to measure accuracy with instruments, the only way to measure fidelity is to see how well observers are able to perform meaningful visual tasks using different kinds of images.

Although the work has not been explicitly directed toward measuring functional realism, there have been a number of studies that have explored the relationships between the characteristics of rendered images and the abilities of observers to perform visual tasks.[1,10,11,20,25] Typically these studies take the form of factorial experiments, where choices that can be made in the image generation process such as setting the geometric level-of-detail, selecting a shading model, or deciding to render shadows are manipulated, and the subject's ability to perform a task is assessed.

Figure 3 summarizes a study by Wanger et al.[25] that illustrates this approach. The goal of the study was to explore how different rendering styles affect subjects' abilities to perform 3d spatial manipulation tasks. The top image shows the scaling experiment in which the subject's task was to adjust the size of one ball until it matched the other. The 3d spatial locations of the balls were varied from trial to trial and the six visual cues indicated in the graph were either included or omitted from the rendered images. The bars in the graph show how including each cue affected how accurately the subjects were able to perform the task. The first thing to notice is that not all the cues had equal effects on performance. Some, such as texture had only small effects, while others such as shadow had major effects. The second thing to notice is that in some cases the effects were in opposite directions, so while including motion parallax improved performance, including perspective cues actually degraded performance on this task. Studies like this one point to an adaptive approach to image generation, where the visual information in the image changes in concert with the task at hand.

Although this study and others like it provide a methodology for measuring the relationship between accuracy and fidelity in computer graphics images, and the results could be used as the basis of metrics for functional realism, what has been missing is a mathematical framework to integrate these results into a coherent model that can be used to predict what kinds of images provide the best visual information for different tasks.

7 MODELING FUNCTIONAL REALISM

Vision researchers refer to the problem of perceiving the properties of scenes on the basis of the information available in images as the *cue combination problem.*[14] The problem is that while images provide many visual cues for scene properties, not all of these cues are available or reliable at all times. For example cast shadows are usually a valuable cue for the spatial relations between objects, but under diffuse illumination conditions this cue is not available, or is so weak as to be unreliable, and other cues like occlusion or motion parallax must be used. How the visual system decides what it is seeing on the basis of the information provided by image cues is one of the central problems of visual perception, and the answer is of great importance to the field of Computer Graphics since its mission is to create these images.

Recently researchers have developed a comprehensive mathematical framework for addressing this problem that characterizes visual perception as a process of *probabilistic inference.*[13] Within this framework perceiving involves making inferences about the properties of a scene *(S)*, on the basis of the information contained in images *(I)*. The quality of the inferences we can make depends in part on the reliability of the image information. Both the characteristics of the imaging process (for example whether it is noisy) and any prior knowledge the perceiver may have about the scene make the information more or less reliable. These ideas are formalized in Bayes' formula shown in Equation 1.

$$p(S \mid I) = p(I / S)\, p(S) / p(I) \qquad (1)$$

Roughly this equation reads: the reliability of the information provided about some scene property given an image *p(S|I)* is equal to the likelihood of obtaining that image given the scene *p(I|S)* scaled by a measure of how often that scene property occurs *p(S)*. The denominator *p(I)* is a normalizing constant.

Some of the benefits of this Bayesian framework are that: 1) it allows imaging processes and perceptual processes to be described in a common mathematical language; 2) it allows the reliability (fidelity) of the visual information provided by images to be quantified; and 3) it provides a set of tools (ideal observer analysis) that can assess whether an observer's performance in a visual task is constrained by the fidelity of the image information or by the limitations of human visual processing. Recently Schrater and Kersten[22] have significantly extended the utility of this framework by formalizing the influence that the perceiver's task has on how image-based sources of information are integrated to estimate scene properties.

8 FUNCTIONAL DIFFERENCE PREDICTORS (FDPs)

Realistic image synthesis is a computationally expensive process. Even with continual improvements in processing power it can still take minutes or hours to render a single image, and this limits the extent to which these methods can be incorporated into widely-used interactive graphics applications. To address this problem, graphics developers often use approximations to accurate light transport simulation in the rendering process. One example is the use of a simple ambient term to approximate diffuse interreflections among surfaces. Another is the substitution of environment maps for ray-traced specular reflections. Although these approximations are acceptable in many cases, under some conditions they can introduce visible errors that can undermine image realism.

As described in Section 4, graphics researchers have recently started to take a principled approach to the use of such approximations in realistic image synthesis with the development of perceptually-based rendering algorithms. These algorithms are grounded in the use of computational models of early visual processing known as *visible difference predictors* (VDPs).[4,15] Given two images (where often one is an approximation of the other) VDPs indicate which regions of the images are likely to be seen as different from one another. In perceptually-based rendering VDPs have typically been used to determine when an approximated rendering will be visually indistinguishable from a more accurate solution. Although these VDP-based techniques are well principled, they have two limitations. First, the threshold difference metrics in current VDPs are likely too strict to allow the performance gains necessary to produce realistic images at interactive rates. Second, current VDPs do not provide meaningful measures of many kinds of suprathreshold errors in rendered images.

These issues are illustrated in Figure 4. Here the image in panel a) was rendered using an accurate ray-tracing technique. The image in panel b) was rendered using an approximate, but much less computationally expensive environment-mapping technique. The image in panel c) shows the result of running these two images through a representative VDP.[2] Notice that the VDP accurately predicts that there are visible differences between the reflections in the teapots. But are these differences meaningful? In functional terms, the two images show many more similarities

Figure 4: Images rendered with a) ray-traced and b) environment-mapped reflections. c) VDP output predicting visible differences between the images. Adapted from [17].

than differences. In both the teapots and other objects appear to be made of the same materials; the shapes, sizes and locations of the objects appear similar; and both scenes seem to be illuminated in the same way. Arguably, for most intents and purposes, the images are *functionally equivalent.* We have recently started a program of research to develop a new class of perceptual metrics for realistic image synthesis that can predict whether visible image errors introduced by rendering approximation techniques produce differences in user performance on meaningful tasks.[17] These metrics will hopefully serve as the basis of a powerful new set of *functional difference predictors* (FDPs) for perceptually-based rendering that will allow dramatic improvements in rendering efficiency while providing assurance of the functional realism of the rendered images.

9 TOWARDS "HIGH-FIDELITY" COMPUTER GRAPHICS

The conceptual and methodological frameworks outlined in this paper provide a foundation on which predictive models of realism in computer graphics might be built. They suggest a program of research to map and model the relationships between the accuracy and fidelity of different approaches to realistic image synthesis. Key to this endeavor will be identifying the relevant tasks a graphics user might need to perform, because these will determine the kinds of visual information an image must faithfully represent. Advanced psychophysical models like these are at the heart of advances in computer graphics *and* vision science. To develop these models graphics and vision researchers will have to work more closely together, but fortunately there is great potential for a natural symbiosis between these communities. Though this collaboration we will hopefully be able to develop both a better understanding of the complex processes underlying visual perception, and to use this knowledge to develop fast, realistic, rendering algorithms that balance accuracy and efficiency, but always maintain visual fidelity.

REFERENCES

1. ARTHUR, W., BOOTH, K.S., AND WARE, C. (1993). Evaluating 3D task performance for fish tank virtual worlds. *ACM Transactions on Information Systems, 11(3),* 239-265.

2. BOLIN, M. AND MEYER, G. (1998). A perceptually-based adaptive sampling algorithm. *Proceedings of SIGGRAPH 1998,* 299-309.

3. DECARLO, D. AND SANTELLA, A. (2002). Stylization and abstraction of photographs. *Proceedings of SIGGRAPH 2002,* 769-776.

4. DALY, S. (1993). The visible differences predictor: an algorithm for the assessment of image fidelity. In (Watson, A.B., Ed.) *Digital Images and Human Vision*. MIT Press, Cambridge MA, 179-206.

5. FEINER, S, MACKINLAY, J., BLINN, J, GREENBERG, D., AND HAGEN, M. (1988). Designing effective pictures: Is photographic realism the only answer? Panel transcript. *Proceedings of SIGGRAPH 1988,* 1-24.

6. FU, S. (2001). Thanks to Suanne Fu for rendering "The Treason of Image Synthesis". Program of Computer Graphics, Cornell University.

7. GERSHON, N. BRAHAM, R., FRACCHIA, D., GLASSNER, A., MONES-HATTAL, B., AND ROSE, R. (1996). Breaking the myth: one picture is NOT (always) worth a thousand words. Panel transcript. *Proceedings of SIGGRAPH 1996*, 491.
8. GOOCH, B., SLOAN, P., GOOCH A., SHIRLEY, P., AND RIESENFELD, R. (1999). Interactive technical illustration. *Proceedings of the 1999 Symposium on Interactive 3D Graphics*, 31-38.
9. HAGEN, M. (1986). *Varieties of Realism*. Cambridge University Press.
10. HU, H., GOOCH, A.A., THOMPSON, W.B. SMITS, B.E. REISER J.J. AND SHIRLEY, P. (2000). Visual cues for imminent object contact in realistic virtual environments. *Proceedings IEEE Visualization 2000*.
11. HUBONA, G.S., WHEELER, P.N., SHIRAH, G.W., BRANDT, M. (1999). The relative contributions of stereo, lighting, and background scenes in promoting 3D depth visualization. . *ACM Transactions on Computer-Human Interaction, 6(3),* 214-242.
12. HUNT, F.Y., MARX, E., MEYER, G.W., VORBURGER, P.A. AND WESTLUND, H.B. (2001). First steps toward photorealistic rendering of coated surfaces and computer based standards of appearance. In *Service Life Methodology and Metrology* (J.W. Martin, Ed.). Oxford University Press. 1-10.
13. KNILL D.C. AND RICHARDS. (1996). *Perception as Bayesian Inference.* Cambridge University Press.
14. LANDY, M.S. AND MALONEY, L.T. (1990). A statistical framework for the combination of consonant depth cues. *Investigative Ophthalmology and Visual Science (Suppl.), 31,* 173.
15. LUBIN, J. (1993). A visual discrimination model for imaging system design and evaluation. In (Peli, E., Ed.) *Vision Models for Target Detection and Recognition*. World Scientific, New Jersey, 245-283.
16. MYSZKOWSKI, K., TAWARA, T, AKAMINE, H. AND SEIDEL, H.P (2001). Perception-guided global illumination solution for animation rendering. *Proceedings SIGGRAPH 2001,* 221-230.
17. PELLACINI, F. AND FERWERDA, J.A. (2002). FDPS: Functional difference predictors: measuring meaningful image differences. *Technical Report 2002-04,* Program of Computer Graphics, Cornell University.
18. READER'S DIGEST (1973).*The Reader's Digest Complete Do-It-Yourself Manual*. Reader's Digest Association.
19. REINHARD, E., STARK, M., SHIRLEY, P., AND FERWERDA, J.A. (2002). Photographic tone-reproduction for digital images. *Proceedings of SIGGRAPH 2002,* 267-276.
20. RODGER, J.C. AND BROWSE R.A. (2000). Choosing rendering parameters for the effective communication of 3D shape. *IEEE Computer Graphics and Applications, 20(2),* 20-28.
21. SAITO. T. AND TAKAHASHI, T. (1990). Comprehensible rendering of 3-D shapes. *Proceedings of SIGGRAPH 1990,* 197-206.
22. SCHRATER, P AND KERSTEN, D. (2000). How optimal depth cue integration depends on the task. *International Journal of Computer Vision*, 40, 73-91.
23. SELIGMANN, D. AND FEINER, S. (1991). Automated generation of intent-based 3D illustrations. *Proceedings of SIGGRAPH 1991,* 123-132.
24. STROTHOTTE, T. AND SCHLECTWEG, S. (2002). *Non-Photorealistic Computer Graphics: Modeling, Rendering, and Animation.* Morgan-Kaufman, San Francisco, USA.
25. WANGER, L.R., FERWERDA, J.A. AND GREENBERG. D.P. (1992). Perceiving spatial relationships in computer-generated images. *IEEE Computer Graphics and Applications, 12(3),* 44-58.
26. WARD. G. (1994). The RADIANCE lighting simulation and rendering system. *Proceedings of SIGGRAPH 1994,* 459-472.
27. YEE, H., PATTANAIK, S.N. AND GREENBERG, D.P. (2001). Spatiotemporal sensitivity and visual attention for efficient rendering of dynamic environments. *ACM Transactions on Graphics, 20(1),* 39-65.