DEPTH PERCEPTION IN REAL AND PICTORIAL SPACES

A Computational Framework to Represent and Simulate the Built Environment

NAN-CHING TAI, MEHLIKA INANICI
University of Washington, Department of Architecture
Seattle, WA, 98195, USA.
tai@u.washington.edu, inanici@u.washington.edu

Abstract. Architectural design is often mediated on two-dimensional representation systems and envisioned three-dimensionally in the pictorial space. The developments of advanced digital technologies have enabled us to create the pictorial representations of un-built design projects that can appear as real as photographs. The visually appealing pictures produced by photorealistic rendering tools are useful for visualizing the form and the spatial layout of the proposed architectural design; but they may be inadequate and misleading for simulating the perceptual qualities of space. This paper draws from the recent developments in computer graphics (physically based renderings and perceptually based tone mapping techniques) and proposes a computational framework to faithfully represent and simulate pictorial spaces. Guidelines are provided for generating images with appropriate representation and simulation techniques so that architects can make informed design decisions about the perceptual qualities of their designs and researchers can study depth perception in computer environments.

Keywords. Depth Perception: Pictorial Cue; Physically based Rendering; Tone Mapping; High Dynamic Range Imagery.

1. Introduction

Depth perception is the process of acquiring three dimensional understanding of an environment. The third dimension of depth is lost when the three dimensional (3D) world is optically projected as two dimensional (2D) images on the retina. Depth cues gathered from the 3D scene are processed to converge in a depth interpretation. Among various dynamic and static depth cues, the pictorial cue is considered as the most significant one in the field of architecture, where design projects are often mediated on the two-dimensional drawings and envisioned three dimensionally on the pictorial space. Pictorial cue is a collection of cues (such as occlusion, familiar size, relative size, aerial perspective, and texture gradients) that can be reproduced on the planar media to create the illusion of depth.

Architects have been relying on the pictorial space to visualize the spatial qualities of their designs for centuries. In the past few decades, computer technology has been incorporated into the different phases of design process; and the computer generated pictorial space has become an environment for architects to envision the spatial characteristics of their designs. Advanced digital technologies have further enabled architects to create the visual representations of un-built design projects that can appear as real as photographs. Photorealistic rendering tools produce 'visually appealing' pictures, where accuracy is not the main objective. Freehand sketches and photorealistic renderings are useful for visualizing the form and the spatial layout of the proposed architectural design; they may be inadequate and misleading for studying the perceptual qualities of a space (Fig. 1). Perceptual qualities are dependent on attributes such as the lighting patterns and distributions, absence or presence of shadow, material opacity, transparency, translucency, and views through windows.

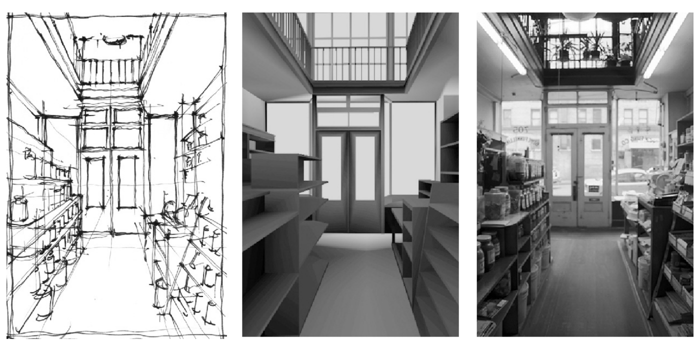


Figure 1. Representation of a space through a freehand sketch, computer rendering and photograph.

This paper draws from the recent developments in physically based rendering and perceptually based tone mapping techniques and proposes a computational framework to faithfully represent and simulate pictorial spaces. The objective is to present a framework for generating images to envision and analyze depth perception in built environments. The fidelity of the proposed framework is demonstrated through the repetition of a classic experiment that investigates the pictorial depth cue of size perspective in the computer environment.

2. Accurate representation in a computer environment

Vision is the result of the interaction of the light source, environment and visual system. Perceptual theorist J.J. Gibson defines a faithful picture as "a delimited surface so processed that it yields a sheaf of light rays to a given point which is the same as would be the sheaf of rays from the original scene to a given point" (Gibson, 1954). This statement defines the faithful representation of the perceived scene; it also illustrates the difficulty of the reproducing the visual perception.

2.1. PHYSICALLY BASED RENDERING

Rendering is the process of making a 2D image of a 3D scene. Physically based rendering tools use advance algorithms to simulate the complex physical processes between the light source and the environment to generate digital images that mimic the real world. It is possible to numerically approximate the luminance values and distribution patterns in simulated spaces through faithful modeling of geometry, material, and light source properties (Figure 2).

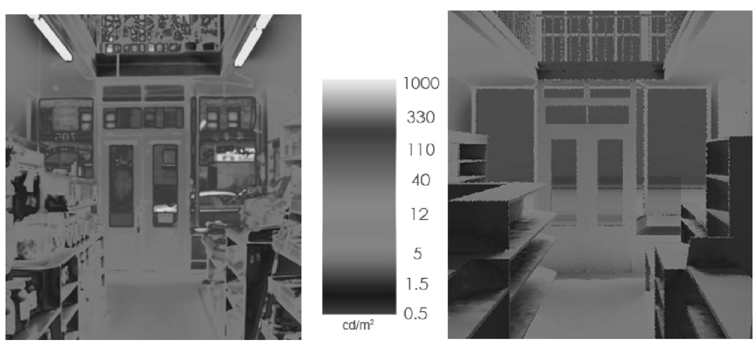


Figure 2. Luminance distributions in the real and simulated scenes.

Luminance values in a natural scene can range as much as 14 logarithmic units from direct sunlight to dark shadows. In a physically based rendering system, lighting values are internally calculated as floating point numbers in absolute units, but the data is usually clipped into 24 bit/pixel integer values during the image storage process. The clipped data spans a dynamic range of about 2 logarithmic units. This solution is efficient in terms of file size, but the clipped luminance information is lost, and therefore, the image does not contain the full range of the lighting information anymore.

There are image formats that enable the storage of high dynamic range values. RGBE (HDR) format uses a 32 bit encoding, where an 8 bit mantissa is allocated for each channel of primary R, G, B that share a common 8 bit exponent

(Ward, 1991). This format can encompass the complete range of the human visual system. Therefore, a physically based rendering stored in a High Dynamic Range (HDR) image format provides lighting data from a simulated scene that can numerically encompass the sheaf of light-rays found in a real scene.

The renderings used in this paper are generated with the Radiance Lighting Simulation and Rendering System (Ward and Shakespeare, 1997) and they are stored in RGBE (HDR) image format. The numerical accuracy of Radiance software has been validated (Mardaljevic, 2001).

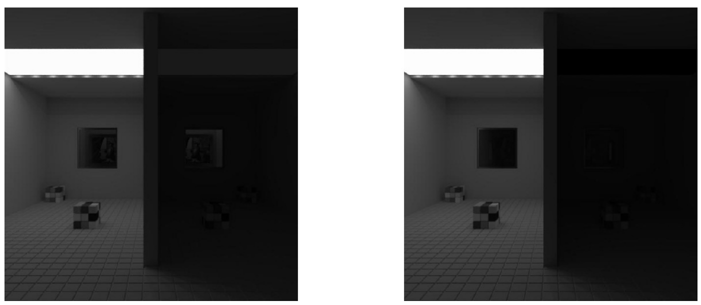
2.2 PERCEPTUALLY BASED TONE MAPPING

The output media, either a hardcopy print or a computer display can reflect or emit a limited portion (only 2 logarithmic units) of the luminance range visible to human beings. Therefore, even with a physically accurate HDR image, the luminance range of the sheaf of rays that is reflected or emitted from a conventional media may not match to the original scene. Tone mapping techniques are used to compress the HDR data into displayable range in a meaningful manner such that visual appearance can be reproduced (Reinhard et al., 2006).

Human's visual system consistently deals with dynamic range reduction. Although human beings can operate over the 14 logarithmic units in a given time, they can only adapt to 4 logarithmic units simultaneously. In addition, the dynamic range of the perceived luminance is reduced to approximately to 32 to 1 in the visual pathway from the retina to the brain (Reinhard et al., 2006). The underlying idea in perceptually based tone mapping techniques is to create a compressed visual stimulus that may not be an equivalent of the light pattern received in the retina from the depicted physical scene, but an equivalent of the stimuli in the later stages of visual processing, which will create a similar visual perception.

Figure 3 illustrates a scene where two identical paintings in the back room are equally illuminated. Each of the paintings is viewed through a window, one of them from a relatively brighter room and the other from a darker room. Figure 3a illustrates a scene that has a low dynamic luminance range: the brightest part of image is 90 cd/m^2 and the darkest part is 0.01 cd/m^2. Figure 3b illustrates a scene that has a wider dynamic range. The intensity of the light source on the bright room is increased. The highest and lowest luminance values are 9,000 and 0.6 cd/m^2, respectively. Both images are linearly mapped to be displayed in a conventional media. It should be obvious that this technique is not a good representation for the HDR scenes. The image in Figure 3b does not represent a brighter environment than Figure 3a, and it would not stimulate the same visual perception as the real scene would.

Several human vision based tone mapping operators have been developed in the past two decades. They are mainly categorized into two as global and local operators. In general, human visual system is modeled through non-linear mathematical functions. In global operators, dynamic range of each pixel is compressed by the same nonlinear function. Global operators are based on the premise that human beings adapt to the overall light levels and see the environment and objects based on that adaptation level. On the other hand, local operators compress each pixel by a different nonlinear curve based on each pixel's luminance value and its relation to the neighboring pixels (Reinhard et al., 2006). Local adaptors are based on the premise that the human eye adapts locally as it scans the environment.



(a) Low dynamic range *(b) High dynamic range*

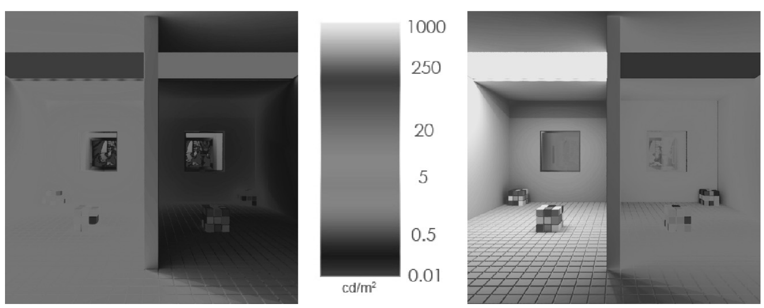
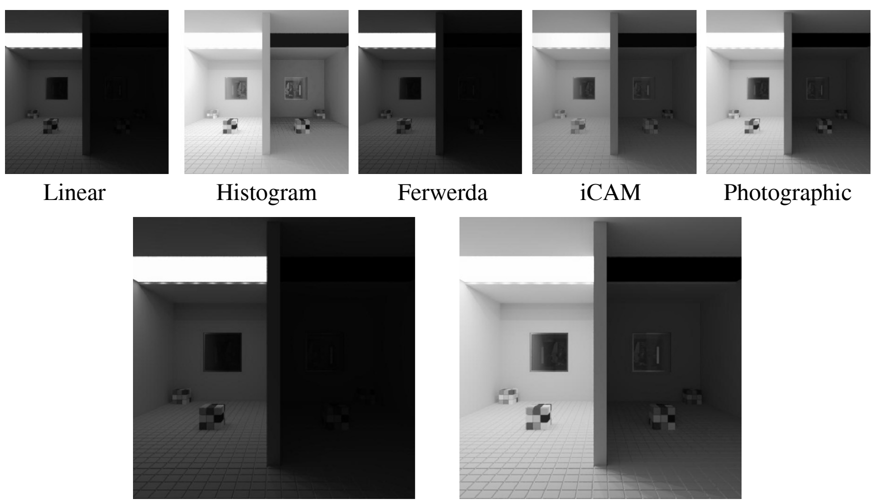


Figure 3. Linearly mapped images for two scenes that have different dynamic ranges; false color images demonstrate the luminance distributions for each scene.

Visual perception is a complex process that is yet to be fully understood. Therefore, different tone mapping algorithms are built upon different approaches, approximations and simplifications. To identify the appropriate tone mapping algorithm for the purposes of this study, it is necessary to review the comparison studies that evaluate various tone mapping operators. These studies can be grouped into three: Different tone-mapped images are compared i) with each other (Drago et al., 2003; Kuang et al., 2004) ii) with HDR images displayed on an HDR display device (Ledda et al, 2005); and iii) with real scenes (Yoshida et al., 2005; Kuang et al., 2006; Cadik et al., 2008).

Although results of these comparisons reveal that there is no single operator that works for all types of scenes, the photographic tone mapping operator (Reinhard et al., 2002) consistently provide good results (Yoshida et al., 2005; Ledda et al., 2005; Kuang et al., 2007; Akyuz et al., 2008; Cadik et al., 2008). Photographic tone mapping operator globally compresses the dynamic range, and then locally adjusts the pixel values to enhance the local contrast using a technique similar to the photographic dogging and burning.

The tone mapped images displayed in Figure 4 are the results of the default settings of the linear, histogram (Ward et al., 1997), Ferwerda (Ferwerda et al., 1996), iCAM (Fairchild et al., 2002) and Photographic (Reinhard at al., 2002) operators. As seen in the image, the photographic tone mapping operator preserves both the overall contrast of the entire scene and the detail visibility. For the purpose of compressing a HDR image that can revoke the same visual perception as the original scene, the Photographic tone mapping operator is identified as a viable tone mapping solution for this study.



Linear Histogram Ferwerda iCAM Photographic

Linear Photographic

Figure 4. Comparisons of different tone-mapped images of an HDR scene.

3. Framework to envision spatial depth

The combined utilization of physically based rendering and perceptually based tone mapping techniques provide a practical solution to represent and simulate pictorial spaces. This approach allows architects to evaluate the impact of design alternatives on the perceived spatial depth, and researchers to study depth perception in computer environments. The next section discusses methods used to estimate depth in real and pictorial spaces.

3.1. METHODS FOR ESTIMATING DEPTH PERCEPTION

Several methods have been used in the perceptual studies to estimate the perceived depth in a real environment. Visual matching is the most straightforward method. In this method, observers are asked to view a target in a distance and prompted to set another target to match that distance. This method works well in physical environments to determine the perceived distance. In the pictorial space, since the full cue condition is deduced from a 2D media, the estimation of the third dimension may not be as intuitive as it is in a physical environment.

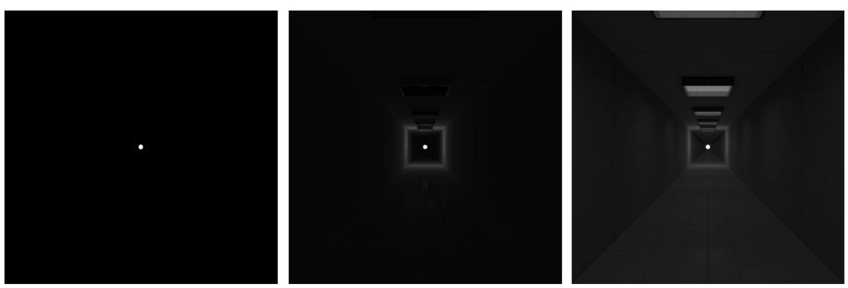
Holway and Boring (1941) suggested another method of visual matching, which utilizes size perspective as an indirect measurement of the perceived distance. Size perspective (also known as the pictorial cue of relative size and familiar size) indicates that the size of the retinal image of an object reduces in proportion to its distance from the observer. By comparing the sizes of the retinal images of the same objects or different objects with well-known sizes, our visual system can gauge the depth. The size-distance relation observed in real and pictorial spaces is theorized as the size-distance invariance hypothesis (Kilpatrick and Ittleson, 1953), and it suggests that perceived sizes of objects are derived from their perceived distances.

3.2. EXPERIMENTS & RESULTS

Holway and Boring (1941) conducted psychophysical experiments to examine two hypotheses of size perception. Law of size constancy indicates that the perception of the size of an object is relatively constant regardless of its variable distance to the observer. Law of visual angle indicates that the size of object should be determined by the size it projects on the retina. The psychophysical experiments examined the validity of these two hypotheses and concluded that in a perfectly isolated environment, the perceived size of the object follows the law of visual angle. However, with the presence of depth cues (such as vanishing lines, shading, diminishing context patterns), the size constancy takes place and allows the observer to correctly perceive the intrinsic size of the object. The results confirm to the size-distance invariance hypothesis, and provide a practical way to measure the effect of scene context on depth perception.

The classic Holway and Boring experiment conducted in a physical space is repeated in a computer generated pictorial space. The comparisons between the experiment results are used to investigate the credibility of the suggested computational framework to study depth perception through indirect visual matching method. In this experiment, observers are presented with a scene that includes a luminous disk with unknown size located in a long dark corridor. Observers are asked to make a judgment on the size of the disk. The experiment manipulates different conditions by restricting the availability of the light to reveal the contextual information of depth cue within the scene. Figure 5 illustrates the same target scene under different conditions. Condition 1 does not provide any contextual information. The size of the disk is predicted based on retinal projection. In condition 2, partial contextual information is available; observers tend to perceive the disk further away and bigger than the retinal projection. In condition 3, with full contextual information, the disk appears further away than the retinal projection and observers are able to provide a good estimation of its actual size. Note that the computer-based experiment is conducted in stable and relatively dark environment to limit the impact of the peripheral vision. Subjects were given sufficient time to adapt to the lighting conditions in the room.

The results of the computer-based experiment agree with the results of the classic experiments: The more the contextual depth information is presented, the better the observer estimates the actual size of the object based on its located distance.



Condition 1 *Condition 2* *Condition 3*

Figure 5. Examples of target scenes used in the computer based experiment.

4. Conclusion

This paper describes a computational framework where spatial depth in pictorial spaces can be estimated using the indirect measurement method. The research has two main foci. The first focus is to discuss whether a computer generated pictorial space can be a faithful representation of the corresponding real scene. Image generation and display guidelines are suggested based on the latest developments in lighting simulation and tone reproduction techniques. Radiance software is used to generate numerically accurate digital images. For digital images that have displayable range, linear tone mapping is used; for scenes that have wider dynamic range, photographic tone mapping operator is suggested to reproduce the visual appearance.

The second focus is to develop a practical methodology to study spatial depth in pictorial spaces. Size-distance invariance hypothesis is repeated to determine whether the size-distance relation of a perceived object and its perceived distance can be observed in a computer generated pictorial space. The results from computer based experiment is compared with the original classic experiment; and it is concluded that although some of the depth information of full cue condition is lost in the transition from the real space to the computer-generated pictorial space, interrelationship of the perceived size and perceived distance can be observed and studied both in physical and computer environments.

The computational framework allows users to systematically manipulate the depth cues and examine their effect. Figure 6 illustrates one of the applications of this approach. The lighting patterns on the original indirect visual matching experiment setup are varied with the addition of a skylight, and the settings are compared to measure the effect of varying lighting patterns on perceived size of the target (disk). In this approach, the effect of contextual depth information can be studied by comparing the measurements of the perceived size of the same object under different pictorial cues as a result of direct manipulation of various architectural configurations.

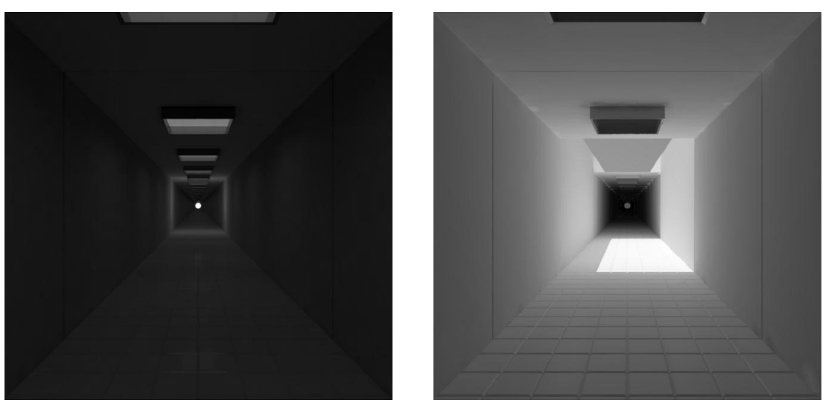


Figure 6. Investigation of different architectural configurations and their impacts on the perception of spatial depth.

Acknowledgements

The authors would like to thank to Dr. John Palmer at the University of Washington, Department of Psychology for his suggestions and discussions on conducting the experiment. The authors also thank the students and staff of the College of Architecture and Urban Planning at the University of Washington who have participated in the experiments.

References

Akyuz, A. and Reinhard, E.: 2008, Perceptual Evaluation of Tone-Reproduction Operators Using the Cornsweet-Craik-O'Brien Illusion, *ACM Transactions on Applied Perception*, **4** (4), Article 20.

Cadik, M., Wimmer, M., Neumann, L. and Artusi, A.: 2008, Evaluation of HDR tone mapping methods using essential perceptual attributes, *Computer & Graphics*, **32**, 35– 44.

Drago, F., Martens, W. L., Myszkowsk, K., and Seidel, H.P.: 2003, Perceptual evaluation of tone mapping operators, *ACM SIGGRAPH 2003 Sketches & Applications*, ACM Press, New York, pp. 1-1.

Fairchild, M. and Johnson, M.: 2002, Meet iCAM: An Image Color Appearance Model, *10th Color Imaging Conference*, pp. 33-38.

Ferwerda, J. A., Pattanaik, S.N., Shirley, P. and Greenberg, D.P.: 1996, An Adaptation Model for Realistic Image Synthesis, *Proceeding of SIGGRAPH*, pp. 249-258.

Gibson, J.J.: 1954, A Theory of Pictorial Perception, *Audio-visual Communications Review*, **1**, 3-23.

Holway, A. H. and Boring, E.G.: 1941, Determinants of Apparent Visual Size with Distance Variant, *American Journal of Psychology*, **54**, 21-37.

Kilpartrick, F.P. and Ittleson, W.H.: 1953, The size-distance invariance hypothesis, *Psychological Review*, **60**, 223-231.

Kuang, J., Yamacuchi, H., Johnson, G., and Fairchild, M.: 2004, Testing HDR image rendering algorithms, *12th Color Imaging Conference*, pp. 315-320.

Kuang, J., Liu, C., Johnson, G., and Fairchild, M.: 2006, Evaluation of HDR image rendering algorithms using real-world scenes, *International Congress of Imaging Science*.

Ledda, P., Chalmers, A., Troscianko, T., and Seetzen, H.: 2005, Evaluation of tone mapping operators using a high dynamic range display, *ACM SIGGRAPH '05*, pp. 249–258.

Mardaljevic, J.: 2001, The BRE-IDMP dataset: a new benchmark for the validation of illuminance prediction techniques, *Lighting Research and Technology*, 33(2), pp. 117–136.

Reinhard, E., Stark, M, Shirley, P. and Ferwerda, J.: 2002, Photographic Tone Reproduction for Digital Images, *Proceedings of SIGGRAPH.*

Reinhard, E., Ward, G., Pattanaik, S. and Debevec, P.: 2006, *High dynamic range imaging: acquisition, display, and image-based lighting*, Morgan Kaufman Publisher.

Ward, G.: 1991, Real Pixels, *Graphics Gems II*, Academic Press, pp. 80-83.

Ward, G., Rushmeier, H. and Piatko, C.: 1997, A Visibility Matching Tone Reproduction Operator for High Dynamic Range Scenes, *IEEE Transactions on Visualization and Computer Graphics*, **3**(4), 291-306.

Ward, G. and Shakespeare, R: 1997, *Rendering with Radiance*, Morgan Kaufman Publisher, San Francisco.

Yoshida, A., Blanz, V., Myskowski, K., and Seidel, H. P.: 2005, Perceptual evaluation of tone mapping operators with real-world scenes, *Human Vision and Electronic Imaging X, IS&T/SPIE's 17th Annual Symposium on Electronic Imaging*, pp. 756-768.