Perceptually-Driven Graphics

Ellen J. Scher Zagier*

Abstract

The purpose of this paper is to explicitly raise a topic that has either been on the back burner, has found itself a subtopic of other areas, or has been traditionally embedded in a tangential field. The secondary purpose is to explore a topic which is of interest to many facets of computer graphics and is relatively new enough to inspire very lively discussion and creative thinking. *Perceptually-driven graphics*, as I present it here, has the primary purpose of computing only that information which is ultimately usable by the human visual system. Its purpose is pure efficiency. In all but a few very experimental systems, the amount of computation done to produce a final image is many orders of magnitude greater than the amount of information the human visual system can possibly process. Perceptually-driven graphics ideally takes into account every piece of usable, existing knowledge about what the eye actually "sees" to drive computation. It directly exploits the logarithmic falloff of resolution from foveal center to extreme periphery.

This paper will also stimulate questions such as: How does a perceptually-driven graphics system fit in with current image-based rendering strategies and with Frameless Rendering? How does it work with new research in multiuser displays? And finally, how crucial is effective eye tracking technology?

1 Introduction

The idea of presenting graphics which models what the eye sees is not new. It permeates the very essence of all computer graphics. The main purpose of this paper is to bring these issues together, but go a step further and assume the visual representation, and final interpretation, of the individual user is known. This representation, with no more information and no less, is ideally what is provided by the user's personal display.

A personalized display which captures an individual user's view is not a new idea. This is the basis for Head Mounted Display devices. Its technological advancement, understandably, falls quite short of a true user's perceptual view. Many of the breakthroughs along the ultimate lines of perceptually-driven graphics lie in the field of Human Vision and Visual Perception. A cortical (the part of the brain receiving the retinal encoding) description of the complex logarithmic mapping of spatial resolution has been rigorously researched mathematically [7]. This is colloquially understood as the falloff of resolution from fovea to periphery. This knowledge has been used in Computer Graphics for ideas for sampling in volume rendering [4] and for research into new displays, a wavelet approach is used to foveate images, that is, match retinal resolution based on an assumed eyepoint and point of attention [2].

*Department of Computer Science, University of North Carolina, Chapel Hill, NC 27599-3175.
scher@cs.unc.edu

2 Some Features of the Human Visual System

I describe some of the many features of the human visual system that ultimately must be addressed in creating a system that mimics what the eye perceives. Most systems provide the necessary information for one feature with overkill, such as retinal resolution, and other features inadequately, such as brightness dynamic range. Some of these items have been raised before in developing a future imaging system [8].

Image contrast, the ratio of image intensities, is of more importance to the human visual system than absolute intensity. So, although the **brightness dynamic range** of the total human visual system (rod system and cone system) is 10^6, the range of surface colors 10^2 is *usually* sufficient in computer graphics displays.

The primary feature of human vision that is rarely exploited now, and is of primary concern in this paper, is the **logarithmic falloff of resolution from foveal center**. The highest visual acuity is only found within a half-minute of visual arc. Other static spatial feature of the human visual system include **multiscale processing**, we process images at different spatial scales, **steropsis**, one of many depth cues and **edge detection**, lateral inhibitors allow us to detect edges and filling in allows us to infer boundaries with missing information.

The eye does not have full spatial resolution when information is shown temporally. The blur in an image sequence can conform to the **motion smear**, or final blur on the retina [6]. **Critical Fusion Frequency**, or the eye's "frame rate" has always been a factor in designing displays. The fact that the critical fusion frequency is lower as intensity decreases can be exploited with dark scenes.

Image compression techniques take advantage of our sensitivity to **lightness-darkness over opponent colors**. Fewer bits of information are required for encoding color adequately because of **non-uniformly distributed cones** in the retina.

3 The Pipeline

The samples computed from the synthetic environment should match the sample size and distribution of photoreceptors on the retina. In order to optimize computational efficiency, a graphics system should take advantage of the variance in size, color resolution, etc. among retinal samples. This necessarily requires a sample-based system. Thus the pipeline for Raytracing [9] is the model I use. It is likely that ultimately all stages of the pipeline will be driven by the way the human visual system functions. The first two stages, **Display Traversal** and **Modeling Transformation** are beyond the scope of this paper. The last two stages, **Tracing Rays** and **Display** are discussed below.

3.1 Tracing Rays

We propose a front end renderer to the Frameless Rendering updating and display scheme back end. (Frameless Rendering is a rendering paradigm which performs stochastic temporal filtering by updating pixels in a random order, based on most recent available input data, and displaying them to the screen immediately [1].) The renderer is an antialiasing raytracer which handles diffuse reflections with a sample size proportional to the photoreceptor size.

Consider a camera collecting information from the real world. A digitizing frame-grabber can be used to extract discrete, typically uniform samples over space and time. It is not necessary, though, for the samples to be uniform in order to produce a perceptual accurate image. In fact, there is already non-uniformity inherent in data collection due to an aspect ratio that is not 1:1.

If the user's viewpoint is known, then samples could be extracted to map more closely to the photoreceptor distribution on the human retina. Samples near the foveal center would be more densely distributed and a sparse sampling would be tolerable in the periphery.

A combination of existing techniques can be combined to build a renderer that handles diffuse as well as specular lighting and computes the sample values independently by tracing independent rays [3]. A Monte Carlo method can be used for stochasticly sampling rays to approximate the surface reflectance function. A sufficient number of independent rays are cast nonuniformly to avoid aliasing artifacts [5].

3.2 Display

Sample size and distribution, based on a user's eyepoint, can address only one user's eyepoint per display. Currently, for many users to be accommodated, they each need to have their own visual interface into the synthetic world. One way to handle this would be to have a Head Mounted Display for each user.

The HMD design would be composed of stereo displays and optics. The HMD would be tracked in time to determine how the user's head moves. Finally the eyes relative to the head motion would be tracked. Although non-invasive and effective eye-tracking does not exist today, we can assume the availability of some sort of eye-tracking device.

The display device can take advantage of always knowing the user's eye position and sample appropriately in the foveal, macular and peripheral regions. The notion of uniform pixels is abandoned for a model that more closely resembles a retinal, filtered image.

The display would be sample addressable with programmable access to voltage level. This way a sample's value is maintained until changed.

4 Conclusion

Sample size, during the data collection stage, need only be proportional to the size of the photoreceptor in the human eye that ultimately receives the information. Computing the color of the sample takes into account retinal sensitivity to the color spectrum, again at the location ultimately responsible for processing the information. Each sample is computed independently of all other samples. Thus, computation is parallelizable, and can be computed and displayed framelessly. The eye-tracked display contains "pixels" which vary with size based on their final resting place on the human retina. Uniform computation and uniform display grids are obsoleted.

References

[1] Gary Bishop, Henry Fuchs, Leonard McMillan, and Ellen J. Scher Zagier. Frameless rendering: Double buffering considered harmful. In *Computer Graphics (SIGGRAPH '94 Proceedings)*, pages 175–176, July 1994.

[2] E.C. Chang and C.K. Yap. A wavelet approach to foveating images. In *13th ACM Symposium on Computational Geometry*, 1997.

[3] James T. Kajiya. The rendering equation. In David C. Evans and Russell J. Athay, editors, *Computer Graphics (SIGGRAPH '86 Proceedings)*, volume 20, pages 143–150, Aug 1986.

[4] Marc Levoy and Ross Whitaker. Gaze-directed volume rendering. *Computer Graphics (Proc. 1990 Symposium on Interactive 3D Graphics)*, 24(2):217–223, March 1990.

[5] Don P. Mitchell. The antialiasing problem in ray tracing. In *SIGGRAPH '90 Advanced Topics in Ray Tracing course notes*, August 1990.

[6] Ellen J. Scher Zagier. A human's eye view: Motion blur and frameless rendering. In *ACM Crossroads Magazine*, 1997.

[7] Eric L. Schwartz. Computational studies of the spatial architecture of primate visual cortex. In *Cerebral Cortex*, pages 359–410, 1994.

[8] Seth Shostak. The ultimate motion image system: What – and when. In *Advanced Imaging*, 1997.

[9] Turner Whitted. An improved illumination model for shaded display. *Communications of the ACM*, 23(6):343–349, June 1980.