The Limits of Human Vision

Michael F. Deering

Sun Microsystems

ABSTRACT

A model of the perception limits of the human visual system is presented, resulting in an estimate of approximately 15 million variable resolution pixels per eye. Assuming a 60 Hz stereo display with a depth complexity of 6, we make the prediction that a rendering rate of approximately ten billion triangles per second is sufficient to saturate the human visual system. 17 different physically realizable computer display configurations are analyzed to understand their visual perceptions limits. The displays include direct view CRTs, stereo projection displays, multi-walled immersive stereo projection displays, head-mounted displays, as well as standard TV and movie displays for comparison. A theoretical maximum triangle per second rate is also computed for each of these display configurations.

Keywords: Visual Perception, Image quality, virtual reality, stereo displays, immersive projection displays, fishtank stereo.

1 INTRODUCTION

With improvements in 3D graphics technology we are on the brink of producing hardware that matches or exceeds the needs of the human visual system. This fact must now be taken into account when designing 3D graphics hardware. As part of an effort to design future hardware and display systems, a study was made of the rendering impact of matching 3D rendering capability to the known limits of the human visual system.

While real-time 3D computer graphics has historically traded off image quality and resolution to meet frame rate and cost constraints, this is becoming less and less the case. The ultimate limits of human visual perception must now be included in hardware trade-offs. A model of such visual limits will be developed, and illustrated in terms of several common display devices. Combined with other rendering assumptions, an estimated theoretical bounds on the maximum triangle rendering rate needed to saturate the human visual system will be made.

2 Limits of Human Vision

The eventual consumer of all 3D rendering is the human visual system. With display technology and real-time hardware rendering speeds ever increasing, we are on the threshold of a generation of machines that will surpass the visual system's input capabilities. On a machine with a single user and a sustained render frame rate of 60 Hz, even present day CRTs exceed the maximum spatial frequency detection capability of the visual system, in regions away from where the fovea is looking. To take advantage of this situation, a hardware rendering architecture could implement some form of *variable resolution frame buffer*. In such a frame buffer, the spatial resolution is not fixed, but must be per-frame programmable to match the variable-resolution nature of human vision. It is assumed that such pixels must be antialiased, and that the antialiasing filter's frequency cut-off must also vary dynamically to match the local effective pixel density. An important question is to understand the precise quantitative details of the variable resolution frame buffer configurations necessary to match the variable resolution nature of the human visual system. This section will provide this information in familiar computer

graphics terms, but first starts with some details of human spatial vision limitations. ([2] is a good reference for much of this material.)

Highest resolution perceivable pixels: 28 seconds of arc.

Several physical factors limit the highest spatial frequencies that can be perceived by the human eye. The diffraction limit of the pupil, the foveal cone spacing, neural trace and physiological tests all agree on a maximum perceived frequency of approximately one cycle per arc-minute (half arc-minute pixels). This is under optimal (but non-vernier[†]) conditions, including 100% contrast. While not quite directly comparable, so-called "20/20" vision represents detecting image features twice as large.

Variable resolution: **1/2@±1°, 1/4@±2°, 1/8@±5°, 1/16@±12°**

This high resolution, however, applies only to the central 2° of vision. Outside of this, the cone spacing and measured preceptorial acuity drop off even faster than the optical limits. In most textbooks (see [2], page 60), this drop off is plotted as a sharp cusp. However, this representation does not do justice to how small the high spatial frequency perception region of the visual field is. Figure 1a plots an alternate visualization of this data onto the surface of a unit sphere: which portions of the 4π steradian field of view are perceived at what resolution. There are 5 false color bands, each corresponding to a factor of two less perceptorial resolution. Figure 1b is a zoom into the central region of the sphere. The center most green region corresponds to the central ±1° of the fovea. The purple from there to ±2°, red to ±5°, orange to ±12°, and yellow to the optical edge caused by the human face. The white represents the non-visible regions. This optical edge has a complex shape, and varies both in the individual and the literature. For our calculations, we used the data of [3], where the maximum field of view varied horizontally from -59° to +110°, and vertically from -70° to +56°. To show both sides of this more than 180° field, two unit spheres are shown, one for a right eye and one for a symmetrically reversed left eye

Thus if the direction of gaze is known, across the entire visual field, the human visual system can perceive approximately only one fifteenth the visual detail that would be discernible if foveal resolutions were available for the entire field

To understand the possible impact on 3D graphics systems, Table 1 presents a comparison of visual and display parameters for several representative display devices.

---

†. A common hyperacuity example is when one can detect a shifts as small as three seconds of arc in the angular position of a large visual object. Here the visual system is reconstructing higher spatial frequency information from a large number of lower frequency samples. However, the visual system can do the same for lower frequency rendered 3D graphics images so long as the higher spatial frequencies were present during the antialiasing process.

Table 1: Display and Human Vision Limits

Display Device	Display Physical Size (per plate)	Display pixel resolution	Display pixel sz in min	Display FOV in steradians	Pixels: 0.47 min limit	Pixels: 1.5 min limit	Pixels: eye limit	Pixels: display limit	Pixels: eye & display limit	Triangles: eye & display limit
18" CRT at 24"	14"×11.2"	1280×1024	1.6	0.25	13.8M	1.34M	1.57M	1.31M	1.08M	0.39B
23" CRT at 24"	19"×11.9"	1920×1200	1.4	0.35	19.2M	1.86M	1.91M	2.30M	1.57M	0.57B
34" CRT at 24"	26.25"×21"	1280×1024	3.0	0.77	42.0M	4.07M	3.33M	1.31M	1.31M	0.47B
65 °HMD*	na	1280×1024	2.9	0.75	40.8M	3.95M	3.26M	1.31M	1.31M	0.94B
95° HMD*	na	1280×1024	4.0	1.23	66.5M	6.44M	4.86M	1.31M	1.31M	0.94B
31" TV at 11.1'	24.8"×18.6"	640×480	1.0	0.03	1.4M	0.13M	0.39M	0.31M	0.28M	0.10B
60" HDTV @ 8'	52.3"×29.4"	1920×1080	1.0	0.16	8.6M	0.84M	1.17M	2.07M	1.06M	0.38B
50' Movie @ 65'	46'×19.5'	2350×1000	1.0	0.19	10.4M	1.01M	1.27M	2.35M	1.14M	0.41B
137' IMAX™ @ 78'	110'×80.62'	4096×3002	1.2	1.07	58.1M	5.63M	4.34M	12.29M	4.14M	1.49B
26'×7' wall @ 8'*	26'×6.93'	3840×1024	2.9	1.38	74.8M	7.24M	5.37M	3.94M	3.2M	2.30B
10' table @ 3'*	7.5'×6'	1280×1024	9.5	2.55	138.1M	13.37M	8.18M	1.31M	1.19M	0.86B
3 Wall Virtual Portal*	6'×7.5'	3×1024×1280	6.7	6.73	365.2M	35.34M	12.98M	3.92M	1.95M	1.40B
4 Wall Virtual Portal*	6'×7.5'	4×1024×1280	6.7	9.25	502.2M	48.61M	14.83M	5.25M	2.28M	1.64B
5 Wall Virtual Portal*	6'×7.5'	5×1024×1280	6.7	10.33	560.3M	54.23M	14.83M	6.58M	2.28M	1.64B
6 Wall Virtual Portal*	6'×7.5'	6×1024×1280	6.7	12.57	682.1M	66.02M	14.83M	7.91M	2.28M	1.64B
3 Wall Virtual Portal*	6'×7.5'	3×1448×1810	4.8	6.73	365.2M	35.34M	12.98M	7.84M	3.88M	2.79B
3 Wall Virtual Portal*	6'×7.5'	3×2048×2560	3.4	6.73	365.2M	35.34M	12.98M	15.69M	7.24M	5.21B
Human eye FOV*	na	na	na	4.17	226.1M	21.88M	14.83M	na	na	10.68B
Full Sphere FOV*	na	na	na	12.57	682.0M	66.0M	na	na	na	491.00B

*stereo.   M: All pixel numbers are in units of millions of pixels.   B: Triangles are in units of billions of triangles per second.

Display Devices: The rectangular ones are characterized by their diagonal measurement and typical user viewing distance. The bottom two entries are the pure limits of the visual system, and a non-tracked visual system (Full Sphere). The table represents a stereo display table top, either tilted at 45°, or with the users head and body tilted 45° relative to a flat table. The 3 wall Virtual Portal is a left, center, and right wall projective immersive display. The video resolution numbers are reversed to indicate that this display is taller than wide; in actuality the video projectors are turned sideways, and the computer generated video format is more traditionally wider than tall. The 4 wall case adds a ceiling display; the 5 wall case adds the floor; the 6 wall case is a complete cube for purposes of comparison. The two higher resolution 3 wall configurations assume tiled 2 then 4 video projectors per wall, for a total of 6 then 9 video projectors. Notice that even the 9 projector system only brings the pixel visual angle to 3.4 minutes of arc, still more than four times lower resolution than that of a standard 1280×1024 desktop CRT.

Display Physical Size (per plate): Because diagonal measurements can be misleading, the full rectangular dimensions of each display device (or each screen for the multiple screen Virtual Portal) is given. The size numbers were adjusted to match both physical display Bezels (where appropriate) and indicated video aspect ratio (typically 1.25 or 1.333).

Display Pixel Resolution: The pixel resolution of the displays. (The movie resolution is an empirical number for 35 mm production film. The IMAX number is the preferred digital source format from their web site for 15/70 70 mm 15 perf film, and is approximately the same physical pixel density on the film as the 35 mm number.) The aspect ratio of the device is also determined by these numbers.

Display Pixel sz: The maximum angular size of a single display pixel in minutes of arc. For wide displays, outlying pixels can be quite a bit narrower in one dimension than this number.

Display FOV: The total solid angle visual field of view (FOV) of the devices, in units of steradians (independent of eye limits).

Pixels: 0.47 min limit: The maximum human perceivable pixels within the field of view, assuming uniform 28 second of arc perception. This is simply the number of 28 second of arc pixels that fit within the steradians of column 5.

Pixels: 1.5 min limit: The same information as the previous column, but for more practical 1.5 arc-minute perception pixels. (Practical in the sense that for many applications 20/30 visual aquity quality would be more than acceptable.)

Pixels: eye limit: The maximum human perceivable pixels within the field of view, assuming the variable resolution perception of Figure 4.

Pixels: display limit: The pixel limit of the display itself (multiplication of the numbers from column 3).

Pixels: eye & display limit: The number of perceivable pixels taking both the display and eye limits into account. This was computed by checking for each area within the display FOV which was the limit: the eye or the display, and counting only the lesser.

Triangles: eye & display limit: The limits of the previous column into maximum triangle rendering rates (in units of billions of triangles per second), using additional models developed in the next section.

To compute the tables numbers, each display was broken up into 102,400 smaller rectangular regions in space, and projected onto the surface of a sphere. Here the local maximum perceptible spatial frequency was computed and summed. Numerical integration was performed on

the intersection of these sections and the display FOV edges (or, in the case of the full eye, the edge of the visual field).

The angular size of uniform pixels on a physically flat display is not a constant; they will become smaller away from the axis. The effect is minor for most displays, but becomes quite significant for very large field of view displays. However, for simplicity this effect was not completely taken into account in the numbers in Table 1, as real display systems address this problem with multiple displays and/or optics.

All of these numbers represent a single snapshot in time, as they are computed for a single eye position relative to the display. All the numbers (other than the HMD, human eye, and full sphere) will change as a user leans in/back from the display and moves around. A more exacting calculation computes minimum / maximum numbers based on a presumed limit of user head / body motions.

There are several things to note from this table. The FOV of a single human eye is about one third of the entire 4π steradians FOV. A widescreen movie is only a twentieth of the eye's FOV, and normal television is less than a hundredth. A hypothetical spherical display about a non-tracked rotating (in place) observer would need over two thirds of a billion pixels to be rendered and displayed (per eye) every frame to guarantee full visual system fidelity. An eye tracked display would only require one forty-fifth as many rendered pixels, as the perception limit on the human eye is only about 15 million variable resolution pixels. This potential factor of forty-five performance requirement reduction was the motivating factor in investigating variable-resolution frame buffer architectures.

All the multi-screen configurations have a larger total FOV than the human visual system, though not necessarily of the same shape. These large screens pay the price in very large individual pixel angular sizes. The visual resolution for the standard resolution cases is equivalent to 20/210 vision, and the older 960x640 stereo video format was even worse.

The 34" direct view CRT (commercially advertised as a 37" display) at 24" viewing distance actually has a 65° field of view, and thus in many ways can be as immersive as the 65° HMD (and a lot less distorted).

3 The Limits of Rendering

The maximum rendering rate that may be needed for a real-time system can be estimated by the following simple model:

$$\Delta/\sec = \text{frame rate} \cdot \#\text{eyes} \cdot \text{screen pixels} \\ \cdot \text{depth complexity} \cdot \Delta/\text{pixel} \tag{1}$$

An empirical estimate of this last term as near unity is made.

The previous section developed estimates of screen pixels based on displays and perception.

Frame rate has not been extensively discussed within this paper, other than an assumption that it is at or above 60 Hz. Very little is known about the interaction of rapidly varying complex rendered images with the human visual system. The best we can do at this stage is to pick a number that is hoped to be high enough. Some have even speculated that very high rendering frame rates (in excess of 300 Hz) may interact more naturally with the human visual system to produce motion blur effects than the traditional computer graphics techniques. For the purposes of example in this paper we will use a 60 Hz rate, and hope that more complex visual experiments will be performed in the future to probe this question.

Spatial antialiasing must still take place, the pixels discussed in this paper are assumed to be antialiased with a high quality re-sampling filter, either based on super samples or area coverage techniques. The pixel counts in Table 1 must be multiplied by the supersampling density if counts of samples rather than pixels are desired.

Depth Complexity: assuming reasonable occlusion culling techniques, current experience is that depth complexity can be kept in a range of 3 to 6 in most (but by no mean all) cases. For purposes of example, a depth complexity of 6 is assumed.

These assumptions were used to compute the final column of Table 1 using equation (1), using two eyes for stereo displays. The only difference between displays marked as stereo is that the number in the finial column is doubled over the mono case. The numbers are in units of billions of triangles per second. Under these assumptions, the maximum triangle rendering rate required to saturate the human visual system is:

$$60 \text{ Hz} \cdot 2 \text{ eyes} \cdot 14.83\text{M pixels} \cdot 6 \text{ DC} \cdot \frac{1\Delta}{\text{pixel}} = 10.68\text{B} \frac{\Delta}{\text{sec}}$$

just over ten billion triangles per second. For most traditional display devices, the saturation number is under half a billion triangles per second. The numbers presented here are neither information theoretic minimum nor maximum calculations; they are conservative "average case" estimates; changes in any of the assumptions can have a large effect on the results.

4 CONCLUSIONS

A model of the perception limits of the human visual system was given, resulting in a maximum estimate of approximately 15 million variable-resolution pixels per eye. Assuming 60 Hz stereo display with a depth complexity of 6, it was estimated a rendering rate of approximately ten billion triangles per second is sufficient to saturate the human visual system. Display and human visual system limitations were also computed for 17 representative display configurations, illustrating the current state of the art.

REFERENCES

1   Cook, Robert, L. Carpenter, and E. Catmull. The Reyes Image Rendering Architecture. Proceedings of SIGGRAPH '87 (Anaheim, CA, July 27-31, 1987). In *Computer Graphics* 21, 4 (July 1987), 95-102.

2   De Valois, Russell, and K. De Valois. *Spatial Vision*, Oxford University Press, 1988.

3   Grigsby, Scott, and B. Tsou. Visual Processing and Partial-Overlap Head-Mounted Displays. In J*ournal of the Society for Information Display* 2, 2 (1994), 69-74
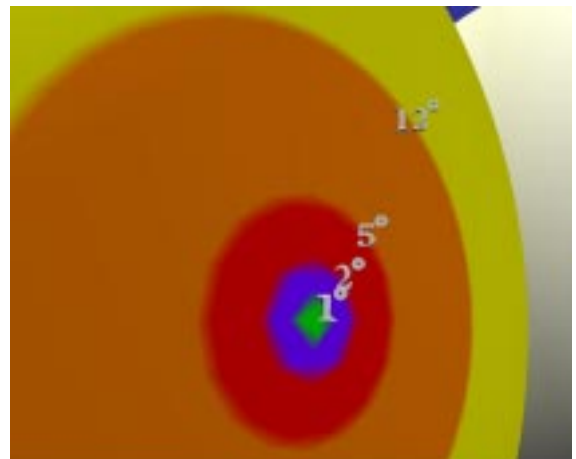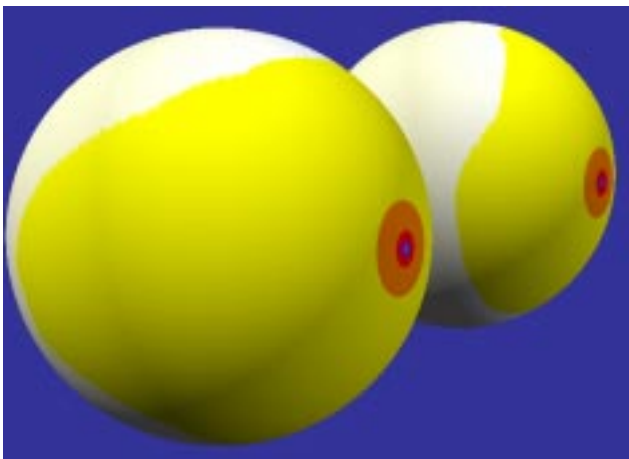
Figure 4a and 4b. Human eye variable density regions.