To be presented at The Fourth International Display Workshops - IDW'97
Nagoya Congress Center, Nagoya, JAPAN, November 19-21, 1997

How Many Bits Are Enough?

James Larimer[†], Jennifer Gille[††], Jeff Luszcz[††]

[†]NASA Ames Research Center, Moffett Field, CA 94035-1000
[††]Sterling Software, Ames Research Center, Moffett Field, CA 94035-1000

INTRODUCTION

Carlson and Cohen [1] suggest that "the perfect image is one that looks like a piece of the world viewed through a picture frame." They propose that the metric for the perfect image be the discriminability of the reconstructed image from the ideal image the reconstruction is meant to represent. If these two images, the ideal and the reconstruction, are noticeably different, then the reconstruction is less than perfect. If they cannot be discriminated then the reconstructed image is perfect. This definition has the advantage that it can be used to define "good enough" image quality. An image that fully satisfies a task's image quality requirements, for example text legibility, is selected to be the standard. Rendered images are then compared to the standard. Rendered images that are indiscriminable from the standard are "good enough." Test patterns and test image sets serve as standards for many tasks and are commonplace to the image communications and display industries, so this is not a new nor novel idea.

What does it take to satisfy this definition? How much information is required? The answer depends upon the reconstruction device and the observer's human visual system. Which of these two elements, the device or the observer, dominates the outcome depends upon many factors. The obvious factors are lighting and viewing distance for the observer and resolution (i.e. temporal, spatial, and chromatic) for the device. Viewed in the dark, all images look the same[2], so the human visual system dominates when the lighting is extreme. Similarly, all images look the same when viewed from a sufficient distance; here again the human visual system dominates. For any two renderings of the same image, there will always be a viewing distance and lighting condition pair at which the two renderings are indiscriminable. This distance defines the dominance boundary between the display device and the eye. This distance will also depend upon the image signal being rendered; low-information content images can be rendered 'perfectly' on low-information devices. A unique distance that does not depend upon the image signal can be defined. Over all possible images that can be rendered by the display device, the farthest distance found is the unique dominance distance. This distance depends exclusively on the attributes of the device and observer and not upon the image signal content. As rendered images move closer to the observer, the rendering engine becomes the dominant factor in determining the quality of the rendered image.

THE COST OF INFORMATION

Information in an image is expensive to gather, store, process, code, transmit, de-code, and reconstruct. Often the amount of image data communicated is more than can be rendered on a soft copy device. For example, a 300 dpi 8 bit grayscale dye sublimation printer requires 1200x3x1500 pixel values or 43.2 Mbits to render a 4x5 inch color image. If this image is rendered as soft copy on an 80 dpi screen with 8 bits of grayscale, only 3.072 Mbits of information will be rendered on the screen. This is one fourteenth of the data sent to the printer. Chances are very good that at a viewing distance of 1m or less the soft copy rendering will equal or exceed the quality of the dye sub print. How many bits are really required to achieve good image quality relative to a standard? How many bits are required to satisfy the indiscriminability test? Bits that cannot be seen are wasted, adding cost to the rendering device and possibly other system components without improving performance.

The amount of information in a static digital image is given by the Shannon formula

$$I = \log_2(\text{\# of locations X \# of levels}). \quad (1)$$

A more natural way to conceptualize image information relates the information value I to the size of the image rendered. Devices can be specified in terms of the number of controllable locations per inch or dpi. In dpi units, the information measure is

$$I = 2\log_2(\text{dpi})+\log_2(\text{sq.in.})+\log_2(\text{levels}). \quad (2)$$

Dropping the area term converts the formula to an information density, I_d, where

$$I_d = 2\log_2(\text{dpi}) + \log_2(\text{levels}). \qquad (3)$$

I_d is the sum of two components, spatial resolution and grayscale resolution. These are known to trade-off against each other [3] once a minimum spatial resolution has been achieved.

A popular printed advertisement [4] for an ink jet printer describing the image of woman in a bathing suit states, "At 300 dpi you see a lady in her bathing suit. At 720 dpi you see her bathing suit is wet. At 1440 dpi you see her bathing suit is painted on." As the resolution increases, more surface details emerge. At the lowest resolution, halftoning the image makes it look flat; at moderate resolution, sufficient surface detail is visible for surfaces to glisten. At the highest resolution, surface details suggesting depth of texture are now visible supporting the advertising claim.

Although the amount of information rendered by the printers in this example is dramatically changing, the signal information content is not necessarily changing at the same rate. Often dithering can be accomplished by spatially up-sampling the image signal without requiring additional image information at the interpolated locations. Because the amount of information required to render an image as soft versus hard copy can be different, image signals often contain more information than will be rendered on the soft copy device. When the signal is rendered as soft copy, the image signal must be downsampled, and this can produce visible artifacts if not done properly. Dithering trades spatial resolution for grayscale resolution. The cost to dither is that more display locations must be controlled in the rendered image. Another cost of dithering is its impact on image quality. The dither itself can become visible, adding a high spatial frequency noise component called 'fixed pattern noise' to the rendered image, degrading image quality by producing a visible pattern that can mask details in the image in addition to adding an objectionable texture to the rendered image.

IMAGE RENDERING COSTS

There are four elements that affect the cost of rendering information as soft or hard copy. These are: (1) the cost of the number of locations in the rendered image, (2) the cost of controlling the locations, (3) the cost of communicating of the image data, and (4) the cost of processing the image data. These factors are not independent. The number of pixels rendered is the number of locations. For soft copy rendering devices, increasing the number of pixels can result in more gates, a smaller fill factor, and less luminous efficiency. Controlling more pixels requires more row and column drivers and higher bandwidth connectivity from the signal source to the rendering device. Driver complexity can be reduced by temporal and spatial dithering of the image signal or by subsampling the image data. By simply downsampling the grayscale, a 2- to 8-fold savings can be generated for many displays. This can result in simpler and cheaper drivers, lower bandwidth connectivity, and reduced EMI, but at a cost of more image processing. It is often simpler to understand the technology trade-off costs in terms of the complexity of manufacturing, the cost of components and architectures, and the cost of added processing, than it is to understand the trade-off cost in terms of image quality. Thus it is important to know the limitations that the human visual system imposes on this trade-off. As previously noted, such limitations are less severe with increased viewing distance and lower brightness.

HUMAN VISUAL SYSTEM

Spatial Sampling Osterberg [5] in 1935 reported measurements of cone and rod photoreceptor densities of the human retina. The retina is the sensory mechanism that is sensitive to light and transduces it into a neural signal. Osterberg counted linear densities of approximately 120 cones per degree visual angle in the fovea, the small 1° retinal region of highest spatial acuity. More recently, Curcio and her collaborators [6-9] have measured the cone mosaic linear sampling densities and find individual variations from as low as 90 to as high as 190 cones per degree visual angle in the foveal region. Campbell and Gubisch [10] have characterized the point-spread function of the eye's optics. The blur measurements and cone density data taken together correspond well with the limiting acuity of the average eye of approximately 1 arc minute. They are also consistent with the variation of best corrected acuity in the population.

To be presented at The Fourth International Display Workshops - IDW'97
Nagoya Congress Center, Nagoya, JAPAN, November 19-21, 1997

Measurements of the spatial contrast sensitivity of the eye, however, show that there are additional attenuation factors that impact the eye's ability to resolve details [11,12]; these factors are believed to be due to neural processing.

The variation in cone sampling densities in the population implies that the dominance boundary defined earlier will not be the same for all viewers. Some individuals will require greater or lesser distances to satisfy an indiscriminability criterion for any given comparison. The 120 samples per degree as a detector density norm is consistent with population acuity norms and therefore is a good figure-of-merit for evaluating display resolution requirements for the average observer.

The viewing distance to the typical office soft or hard copy display device is approximately 0.5m. At this distance, one inch subtends three degrees of visual angle. At 120 cone photoreceptors per degree, it would require 360 dpi to match one pixel to each cone photoreceptor in the foveal region. A 600 dpi laser printer is at a higher spatial resolution than the eye at standard viewing distances even for a high-acuity individual. One might believe that at 600 dpi the laser printer would render 'perfect' natural images. Black and white images that do not contain grayscale values between the min and max can be 'perfectly' rendered on these devices. The grayscale of a laser printer, however, is only one bit. Grayscale values between the max and min reflectances can only be created by dithering. The dither trade-off lowers effective spatial resolution to less than 150 dpi, and because of the inherently high contrast of the pixels, the fixed pattern noise produced by dithering becomes quite visible at this distance. Remember that the spatial Fourier power spectrum of a point contains frequency energy over a broad range of spatial frequencies and therefore can be detected by mechanisms of vision that are tuned to any of these frequencies. Soft copy devices are not limited to one bit of grayscale, and can produce good-looking images at lower spatial resolutions than laser printers because of the added degrees of freedom in grayscale at each pixel location.

Intensity Sampling The grayscale resolution of the eye is limited by two factors. Below some minimum signal level the visual system cannot resolve signal from noise. First, the eye, like any detector system, is limited by noise both internal and external. A very elegant experiment and theoretical analysis conducted over 50 years ago by Hecht, et al. [13] revealed that human vision is remarkably immune to noise in the visual environment and is limited only by the quantum nature of light itself. For example, we would not be able to see laser speckle, which can be quite random due to mode switching in the laser, if the eye were less noise immune. The noise limits of vision are primarily due to internal or neural noise.

Secondly, the nervous system has a compressive saturating non-linearity: above some contrast no additional internal signal is generated even though the input signal is increasing. The eye rapidly adapts, however, so it can be very difficult to measure this saturating limit, and thus this limitation of the human visual system is not a significant factor in grayscale resolution. There is an additional reason that the saturating nonlinearity is not a factor. The most frequently encountered viewing conditions and lighting for standard direct-view soft-copy displays all lie within a quite limited range.

Office automation displays have a peak brightness in the range of 15 to 300 cd m^{-2}. A typical display will have a peak brightness of 75 cd m^{-2}. Ambient lighting in most offices is in a range of 5 to 75 cd m^{-2} measured with a perfectly reflecting white screen, or approximately 100 to 1000 Lux. Under these viewing conditions, the peak contrast of office devices is not sufficient to generate saturating signals in the visual pathways and there is little visually-significant adaptation.

The signal-to-noise ratio of the visual system has been measured empirically in a variety of experimental settings. The classical measurements were made by Stiles [14]. Stiles used a chromatic adaptation methodology to isolate the cone pathways and then attempted to measure the signal-to-noise ratio in each cone system. The measurement is made by viewing a steady field of fixed intensity, called the adapting field, and then measuring the intensity increment of a brief flash required to just see the increment. The minimal increment that is visible is the minimal signal required to evoke a sensory experience; it is a measure of the signal-to-noise ratio of the system. It is now understood that Stiles measured the signal-to-noise ratio at various stages within the cascade of neural processing taking place within the visual system [15,16]. Nonetheless, his measurements are useful measures of the limiting resolution of the eye. He found that for two of the cone systems and their associated primary pathways the minimal ratio was approximately 2% and for the third cone system 8.7%.

Above a minimal steady-field level, the ratio is independent of brightness. That is to say, the incremental or decremental signal required to see a change is a fixed proportion of the adapting level [17] once the dc level is above the threshold level for the field. This is a property of sensory systems and is called Weber's Law [18]; the ratio is often referred to as the Weber-Fechner fraction. We measured the Weber-Fechner fraction for soft copy displays [19] and found that the red and green primaries of a typical rendering device have signal-to-noise ratios just under 2% and for the blue primary it is around 4%.

The discrepancy between our finding and that of Stiles is due in part to the non-isolation condition we used to measure these ratios. Unlike Stiles, we made no attempt to isolate the cone pathways, so our signal could have been sensed by any or all of the cone mechanisms and the subsequent mechanisms that process their signals. Since more than one mechanism can detect the increment, there are multiple chances to detect it. Our results are consistent with simple probability summation over these pathways [20]. Since the spectral tuning of the cone systems is highly overlapping, the cones are broadband detectors in the spectral domain, and the standard rendering-device RGB primaries produce highly correlated signals within the cone photoreceptors.

We found in addition that below field levels of approximately 0.34 cd m^{-2} (0.1 fL), the Weber-Fechner fractions were increasing. This is the brightness operating level at which sensitivity is dominated by absolute threshold of the cone mechanisms. At this level the field is not generating any appreciable signal, so only the increment matters. For a display device this means that grayscale differences below this level are very difficult to see, since all input signals below this intensity level are generating essentially the same internal event. Added grayscale steps below this value are wasted.

ENOUGH INFORMATION

At IDRC'94, we [3] measured the trade-off in grayscale and spatial resolution for dithered images empirically with real human observers and in theory using a computational model of human vision. A schematic representation of Fig. 3 from that report is shown in Fig. 1. Fig. 1 plots the locus of points in a space defined by equation (3), grayscale resolution in $\log_2$levels versus spatial resolution in $2\log_2$dpi, that are indiscriminable from a high-resolution rendered image. In the experiment we conducted [3], we measured the discriminability of a very high resolution rendered image to downsampled and dithered renderings of the same image signal. The rendering device we used in the empirical part of our experiment had a peak luminance of 25 fL and a measured contrast of 50:1. Additional details of the experiment can be found in [3]. The trade-off locus fell along a curve that could be described as two line segments meeting at a common point or elbow. This locus is represented as two solid line segments in Fig. 1.

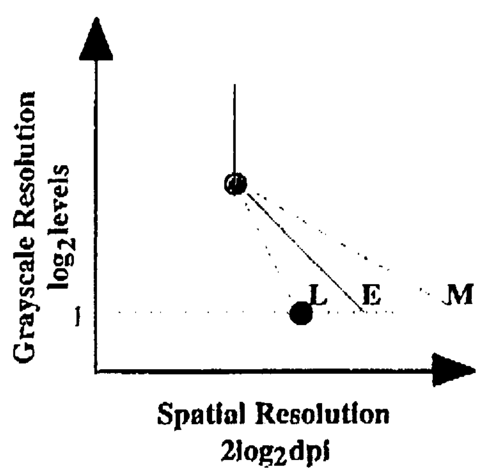

Figure 1. Schematic representation of trade-off locus shown in Fig. 3 of [3].

All the points on this locus are downsampled and dithered renderings of a zone plate [21] and they are indiscriminable from a very high resolution rendering of this image signal. The high resolution image can be thought of as a point representing some large number of gray levels (possibly continuous) and some very large number of dpi (possibly continuous as well). In Fig. 1 the high resolution point is in the upper right-hand corner. Our results indicated that below the elbow the line segment had a slope close to -1. One bit of grayscale was equivalent to one bit of spatial resolution in our empirical data [3]. This locus is labeled as 'E' for 'Equivalent' in the figure. The elbow is indicated as an open circle in Fig. 1 and was empirically found to be approximately at 140 dpi and 8 levels of grayscale (i.e. the point <14.4,3> in this information space) [see 3, Fig. 3].

Note that all of the points to the right and above this locus of indiscriminable rendered images require more information to render them. If the slope of the lower segment is -1 (as it was approximately found to be in the results of the empirical and computational experiment [3]), then the information content of the point at the elbow and all points below it on the indiscriminability locus are the lowest possible bits required to be indiscriminable from a continuous rendering of the image. In this case we call the elbow the 'Superior Information Nodal Point' or Superior Nodal Point for point of least information content. The point where this line intersects the horizontal line defined by bits = 1 (the smallest possible number of quantized grayscale levels) is called the 'Lesser Nodal Point.' For the case where the slope is -1, these two Nodal Points correspond to the same amount of information rendered.

The information content of the Superior Nodal Point is the least information required to be indiscriminable from a continuous image when the data follow a locus with slope greater than -1. This possible locus is shown as the dashed line segment labeled 'M' for 'More' in Fig. 1. Here the Lesser Nodal Point requires more information to render. Finally, if the locus of points below the elbow has a slope of less than -1, then the point of least information content is below the elbow and to the left of the line of slope -1. In this case the point of least information would be the Lesser Nodal Point. A line segment representing this condition is shown in Fig. 1 as a dashed line labeled 'L' for 'Less' information. The Lesser Nodal Point for this case is indicated at the solid black circle at the intersection of this line segment and 1 bit.

We found in the results of both our computational and empirical experiments on the grayscale resolution trade-off that the Lesser Nodal Point was at 300 dpi. At first this would seem wrong since natural images rendered on 300+ dpi laser printers certainly do not look like continuous images. But there is an important difference. Our experiments were done with images that had a 100% spatial fill factor. Laser printers do not have this property. The sharpening of the dots in the laser printing process tends to put more information into lower spatial frequency bands, making the dots more visible by providing an input to the spatial mechanisms of vision that are tuned to lower spatial frequencies. The peak tuning spatial frequency of human vision is around 2 cycles per degree[11]. The dots used in our experiment had 100% fill; this acts like a bandlimiting filter reducing the alaising energy present in smaller dots like those used in laser printing.

Combining the findings in [3] and [19] we can calculate the number of bits to look 'good enough' relative to a very high resolution standard zone plate. In the case of image rendered at 25 fL peak brightness with 50:1 contrast or less a information density of 14.4+3 or 17.4 bits is enough. To optimize the image quality, the steps should be geometrically spaced [19].

CONCLUSION

The number of bits required to render a perfect image depends upon the human visual system and the rendering engine. There is a dominance boundary at which one of these two systems determine image quality. Finding this boundary and measuring the Nodal Point defines the minimal information required to render images that are indiscriminable from an ideal image. These dominance boundaries are near the dpi rates currently being produced by LCD manufacturers and used in laptop computers and other information appliance devices.

To be presented at The Fourth International Display Workshops - IDW'97
Nagoya Congress Center, Nagoya, JAPAN, November 19-21, 1997

REFERENCES

[1] Carlson, C. & Cohen, R. (1978) **Visibility of displayed information. Image descriptors for displays.** Office of Naval Research Report, CR213-120-4F.

[2] There are two senses to viewing in the dark. A self-luminous image can be viewed in a dark room, but not in the dark, since it sets the minimum level by virtue of being self-luminous. Viewing an emissive display in a dark room maximizes image contrast by eliminating the veiling glare of ambient light. If in addition one views such images by looking at them through attenuating filters, then they can be viewed in the common sense notion of dark, i.e. with little light.

[3] Gille, J., Martin, R., Larimer, J. (1994) **Proc. IDRC '94**, 381-385.

[4] Epson Color Ink Jet Printer Advertisement

[5] Osterberg, G. (1935) **Acta Ophthal.**, Kbh., Suppl., 6:1-102.

[6] Curico, C., Allen, K.,Sloan,K., Lerea, C., Hurley, J., Klock, I., Milam, A. (1991) **J. Comp. Neurol.**, 312:610-624.

[7] Curcio, C., Sloan, K. (1992) **Vis. Neurosci.**,9:169-180.

[8] Curico, C., Sloan, K., Kalina, R., Hendrickson, A. (1990) **J. Comp. Neurol.**, 292:497-523.

[9] Curcio, C., Sloan, K., Packer, O, Hendrickson, A., Kalina, R. (1987) **Science**, 236:579-582.

[10] Campbell, F., Gubisch, R. (1966) **J. Physiol.**, 186, 558-578.

[11] van Nes, F., Bouman, M. (1967) **JOSA**, 57, 401-406.

[12] Campbell, F., Robson, J. (1968) **J. Physiol.**, 197, 551-566.

[13] Hecht, S., Schlaer, S., Pirenne, M. (1942) **J Gen. Physiol.**, 25, 819-840.

[14] Stiles, W.S. (1959) Color vision: The approach through increment-threshold sensitivity. **Proc. Nat. Acad. Sci.**, 45, 100-114.

[15] Pugh, E.N., Mollon, J. (1979) **Vis. Res.**, 19, 293-312.

[16] Pugh, E.,N., Larimer, J. (1980) **Vis. Res.**, 20, 779-788.

[17] Whittle, P.

[18] Fechner, G.T. (1860/1966) **Elements of psychophysics**. D.H. Howes & E.G. Boring (eds.), H.E. Adler (trans.) New York:Holt, Rinehart, and Winston.

[19] Larimer, J., Gille, J., Luszcz, J. (1997) **Proc. IDRC'97**, 262-265.

[20] We did not directly test the probably summation model in our measurement, but the data are consistent with that interpretation.

[21] A zone plate is a radially symmetric spatial sinewave chirp or fuzzy bulls eye pattern.