How to Allocate Bits To Optimize Photographic Image Quality

Joyce E. Farrell
Hewlett Packard Laboratories
Palo Alto, California

Abstract

"What is the most efficient1 way to generate grayscale – by increasing pixels/inch or by increasing bits/pixel?". Simulations, preference judgements and computational image quality metrics all converge to yield the same answer. The most efficient way to generate **high-quality photographic images** is to increase the number of graylevels/pixel. Consider, for example, two photographic images *printed at the same size*: A 3 bit/pixel grayscale image is printed at 300 DPI and takes only 0.0983 megabytes of file space. A 1 bit/pixel grayscale is printed at 1200 DPI and takes 0.5243 megabytes of file space. Although the images appear to be the same size, they do not appear to have the same image quality. Both empirical data and image quality metrics predict that the 300 DPI grayscale image will have higher perceived image quality than the 1200 binary image. Clearly, for photographic image quality it is much wiser to dedicate bits to grayscale (bits/pixel) than to DPI (pixels/inch).

Introduction

Most people use the word "resolution" to refer to the number of dots a printer can place per inch of paper (DPI). This colloquial usage often leads to confusion in technical discussions. Technically speaking, the resolution of a printer refers to the size and shape of a printed dot and NOT the number of dots per inch. Resolution is an adequate descriptor of an analog imaging system such as an optical lens (describing the spatial transformation of a point light source into a two-dimensional gaussian spot), but it is not an adequate descriptor of a digital imaging system such as a printer or display. This is because digital imaging systems place pixels, dots or spots on an addressable grid. Thus, to describe the imaging capabilities of a digital imaging system, we must specify both the device addressability" (the number of pixels/inch, dots/inch or DPI) and the device "resolution" (the size and shape of a pixel.). Most printers and displays are designed such that pixel size scales inversely with pixel addressability (hence the two terms are easily confused).

Graylevels per pixel is another imprecise term. Both the distribution of graylevels over luminance and the distribution of graylevels over space influence the perception of image quality. For example, holding the *number* of graylevels per pixel fixed, perceived image quality is influenced by the actual luminance values of those pixels (i.e. the distribution of graylevels over luminance). Again, holding the number of graylevels per pixel fixed, perceived image quality is influenced by the digital halftoning method one uses to map many graylevels into fewer graylevels. (i.e. distribution of graylevels over space). When inquiring about a printer or display, one should not only ask about how many bits/pixel and pixels/inch the device can address, but how those bits and pixels are distributed over intensity and space.

With these caveats in mind, we proceed to describe our observations of how image quality depends both on addressability (dpi) and grayscale (bits/pixel). The question we posed was: "How should we allocate our bits in print in order to optimize image quality -- by increasing dpi or by increasing bits/pixel?". The answer we found was "It depends". It depends on the content or nature of the image we want to print. There is very little improvement in the perceived image quality of ***text*** when we increase addressability beyond 600 dpi. But the perceived image quality of ***photographic images*** does continue to improve with addressabilities greater than 600 dpi. Why is this?

Text

Text is a high-frequency binary signal. When text is sampled at a relatively low frequency, the sampling artifacts (often referred as "jaggies") will be perceptible. However, if the sampling frequency is high, relative to the optical resolution of the lens of our eyes, we will not perceive the sampling artifacts. We do not need to represent or pass frequencies that are present in text but are beyond the resolution limit our our eyes. Since the lens of our eyes blur spatial frequencies greater than approximately 60 cycles per degree of visual angle [4,5] we need only sample 120 cycles/degree (the "Nyquist limit" of our optical system). At a typical viewing distance of 12 inches, a sampling rate of 120 cycles/degree corresponds to approximately 600 DPI. If

[1] I use the word "efficient" to refer to the solution that requires less disk space: If two **grayscale** images are perceptually equivalent in appearance, the image that requires less disk storage space is more "efficient". This argument does not consider the effects of image compression which can reduce disk space requirements considerably.

we sample below the Nyquist limit, with addressabilities below 600 DPI, the sampling artifacts will be visible.

One method for reducing the visibility of sampling artifacts generated when we sample text at frequencies below our Nyquist limit is to pre-blur the text before sampling. This method, often referred to as “anti-aliasing”, uses grayscale information to represent the blurred images [6]. Although the sampling artifacts may be less visible, blurred text will never be as pleasant to read as text sampled with higher frequencies [7]. Therefore, it always better to sample at the Nyquist limit of our optical system (600 dpi at a 12 inch viewing distance). Sampling with frequencies higher than the Nyquist limit does not generate improvement in the perceived image quality of text, however, because these higher frequencies have already been removed by the optics of our visual system.

Images

Why, then, does the perceived image quality of ***photographic images*** appear to increase when we sample beyond the Nyquist limit of our visual system? Increasing device addressability beyond 600 dpi effectively increases the number of perceived graylevels: Perceived graylevel is determined by the number of pixels that are blurred together as a result of the optical point spread function of the lens of our eye. The number of perceived graylevels one can create by blurring pixels together is determined both by the area over which the eye blurs and the number of pixels per area [8].

There are many ways to increase the number of graylevels/area in a printed image. One can vary the density of dots that fall within a fixed area (defined by the optical point spread function of our eyes) or one can vary the dye density of the dots that fall on a particular spot within the fixed area. And, of course, one can vary both dots/area and density/dot. This brings us back to our original question: “What is the most efficient way to generate grayscale – by increasing pixels/inch or by increasing bits/pixel?”. To answer this question, we developed methods for simulating the output of devices with different addressability and grayscale capabilities. We developed methods for quantifying how people perceive the image quality of such devices. And we developed metrics to predict these subjective judgements.

Printer Addressability (pixels/inch) and Grayscale (bits/pixel)

Printer Simulations

To investigate how grayscale and addressability affect image quality in regions of unexplored printer design space, we created a 1200 dpi device with 8 bits of addressable grayscale [9]. We placed a relatively high-resolution 24 bit color CRT at the end of a long tunnel. The tunnel was lined with black felt cloth to eliminate depth information about the actual location of the CRT. Inside the tunnel, we placed two camera lenses between the CRT at one end and a small hole at the other end. The camera lenses enabled us to both minify and focus a virtual image of the CRT display at a distance of 12 inches from the viewing hole. The focused virtual image had an effective visual addressability of 1200 dpi.

We used the 1200 dpi display to present simulations of lower-resolution images (200, 300, 400 and 600 dpi) with varying grayscale capability [10]. We first created a high resolution (2Kx2K) 8 bit grayscale image. We decimated the image (low-pass followed by sampling) to create lower resolution images. After correcting for printer and display non-linearities, we mapped the 8 bit grayscale map to fewer bits using the Floyd and Steinberg [11] error diffusion algorithm. To create the simulated images, we interpolated (upsampled) the lower resolution images to 2Kx2K using a modified gaussian model of a printed dot. We applied gamma correction before displaying the images.

To test the validity of our printer model and our ability to render simulations of printed output on our 1200 dpi 24 bit color display, we conducted a control experiment in which we compared subjective judgements of displayed simulations of 200 dpi grayscale images with subjective judgements of printed 200 dpi grayscale images. Printed images were generated using the same image processing pipeline described above, with the exception that the interpolation stage was unnecessary. Subjects were shown the printed and displayed images at two different times. In the printer condition, subjects were asked to rank order the different grayscale images (2, 4, ... 256 levels at 200 dpi) from worst to best image quality. In the display condition, subjects were shown pairwise combinations of the different grayscale images and asked to indicate which of the two images they preferred. Image quality ratings were obtained by summing the number of time subjects preferred (or ranked) one image over the other.

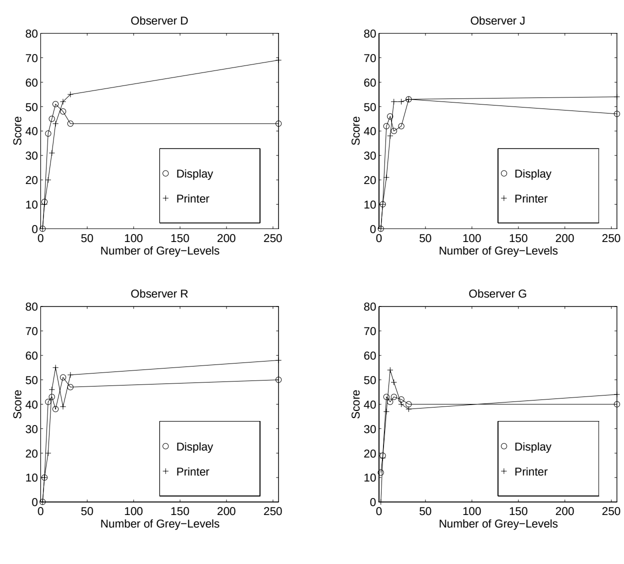

Figure 1. Image quality score (estimated by the percentage of trials in which subjects preferred (display condition) or ranked (printer condition) one image over the number plotted as a function of number of graylevels. Data are shown for four subjects.

Figure 1 compares image quality ratings for printed and displayed simulations for four subjects. The relationship between image quality scores and number of graylevels is similar for both the printed and displayed simulations. We concluded that the displayed simulations were a reasonable approximation to the appearance of printed images (See Anthony and Farrell [3] for details about the printer simulation, empirical data and analysis.)

Subjective Evaluation

The method of pairwise comparison generates reliable and informative data about perceived image quality above threshold. In this method, subjects are presented with two stimuli at any given time and asked to indicate which of the two stimuli “looks better” to them. This method requires no apriori decisions (assumptions) about the factors determining their judgements. Rather, it enables us to test assumptions about how many different factors (such as addressability and graylevels, for example) affect supra-threshold judgements about image quality.

We presented all pairwise combinations of images that differed in both addressability (300, 600 and 1200 DPI) and number of graylevels (2, 4, 8 and 12) and asked subjects to indicate which of any two presented combinations they preferred [10]. Since there were 12 different stimuli, this required a minimum of 132 comparisons. Each of the four people who volunteered to be subjects in our experiment viewed each comparison 10 times over the course of a week. Thus for each subject, we collected data over 1320 trials (see [12] for an adaptive pairwise comparison method that reduces the number of trials while maximizing the information content of the confusion matrix).

To test the hypothesis that addressability and graylevels directly tradeoff we analyzed the stimulus comparison matrix to determine if the stimuli could be ordered along one dimension (or preference vector) and if different combinations of grayscale and addressability resulted in equivalence along this dimension. We used several different statistical methods [12, 14] to determine that the dimensionality (or rank) of the stimulus matrix was 1. This result is significant because it demonstrates that grayscale and addressability tradeoff, such that one can obtain equivalent preference judgements by different combinations of grayscale and addressability. When two images have the same DPI, subjects prefer the image with the higher number of graylevels. Conversely, when two images have the same number of graylevels, they prefer the image that has the higher DPI. Figure 2 shows that one can offset a decrease in DPI with an increase in number of graylevels to keep perceived image quality constant. Similarly, one can decrease the number of graylevels if one increases DPI and keep perceived image quality constant. In other words, grayscale and addressability map into a single dimension of perceived image quality. We turn now to consider a metric that predicts image quality as a function of grayscale depth and addressability.

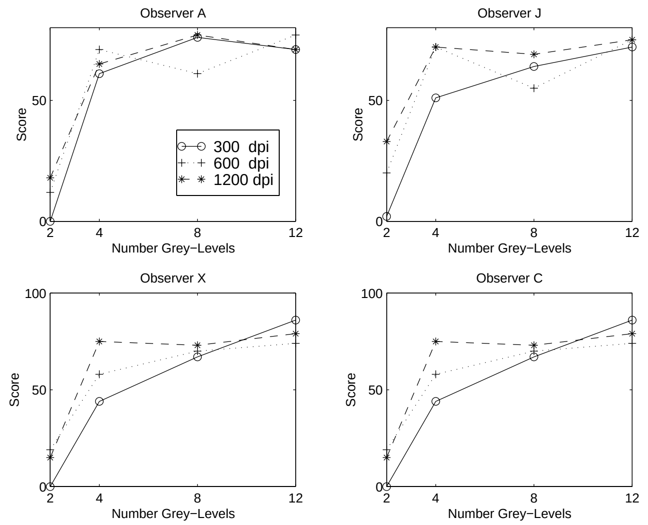

Figure 2. Image quality score plotted as a function of the number of graylevels with addressability (dpi) as a parameter. Data for four subjects are shown.

Metrics

Just as we use printer modeling and simulation tools to present images to subjects that would be generated by a device, in the absence of the device, we can use computational image quality metrics to predict how people would judge images, in the absence of people. We have at our disposal a collection of different metrics designed to predict the visibility of film grain, toner particles, halftone texture, printer banding and JPEG compression artifacts [15]. Since we hypothesized that subjects’ judgements in the grayscale/addressability study were determined in large part by the visibility of the halftone texture, we were interested in comparing the predictions of a metric developed to predict halftone texture visibility to the empirical data we collected in our experiments [3, 10].

The halftone texture visibility metric we used was developed by Zhang and Wandell [16,17] as a spatial extension to CIELAB. This metric, referred to as S-CIELAB, imposes three pre-processing stages before the computation of the CIELAB color difference metric, ΔE. First, the input image is converted from a device- dependent space into a device-independent representation consisting of one luminance and two chrominance color components. Second, each component image is passed through a spatial filter that represents the spatial sensitivity of the human visual system for that color component. Third, the filtered images are transformed into the CIE-XYZ format so that standard CIELAB color difference metrics can be computed.

One of the advantages of the S-CIELAB metric is that it is backwardly compatible with CIELAB in the sense that for large uniform targets the S-CIELAB predictions are the same as the CIELAB predictions. For textured regions, however, the two formulae make very different predictions. Another advantage of S-CIELAB is that the units of the metric already have special meaning in the engineering community. Color scientists and engineers are accustomed to reporting perceived color differences in units of ΔE.

We used the metric to make predictions about the visibility of halftone texture as a function of device addressability and number of graylevels. To compare these predictions with our empirical data, we used the same halftoning method (error diffusion) and the same original image. The results, shown in Figure 3, are predictions based on a standard test pattern that we developed later. The test pattern is an exponential grayscale ramp that spans 3 degrees of visual angle [18]. We prefer the standard test pattern because it samples a wider range of grayscale values. (The predictions are comparable to the predictions based on the original image.)

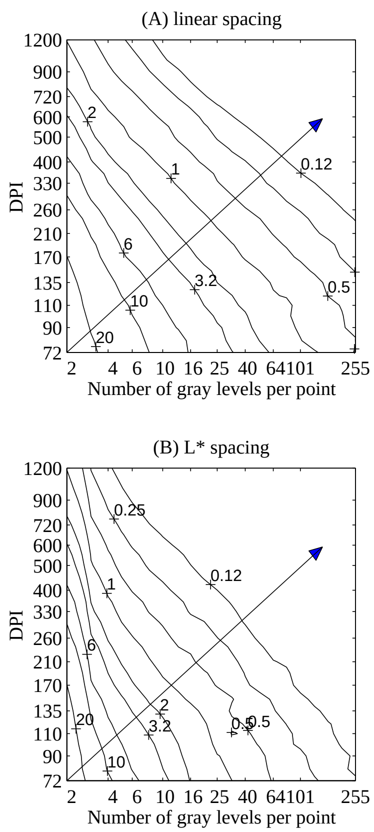

Figure 3. Iso-quality contour plots for grayscale ramp image: Each curve represents the combinations of grayscale and addressability that generate the same S-CIELAB ΔE value. The direction of the arrow indicates increasing image quality. (A) Grayscale ramp was halftoned using linear spacing of luminance. (B) Grayscale ramp was halftoned with levels equally spaced in L.*

The predictions shown in Figure 3 are based on the assumption that a standard visual observer (modeled by the S-CIELAB metric) viewed the test pattern from a distance of 12 inches. We computed the S-CIELAB differences between the continuous grayscale ramp and different possible halftoned ramps. The halftoned ramps differed in addressability and grayscale.

Figure 3 shows iso-quality contours for the test pattern halftoned with different number of graylevels and different addressability. Each curve in the figures plots the combinations of grayscale and addressability that generate the same S-CIELAB ΔE values. *Figure 3a* shows measurements using linear halftone level spacing and *Figure 3b* shows measurements using L*-spacing of halftone levels. Together the curves support the conclusions of our empirical investigations of grayscale/addressability tradeoffs. When equated for the number of halftone levels, L*-spacing of the levels is predicted to have better image quality than linear-spacing. Halftone errors do not decrease linearly with the increase of DPI or number of graylevels. Rather, as the halftone levels increase beyond 16, or dpi increases beyond 800 dpi, the halftone quality improves very little [17].

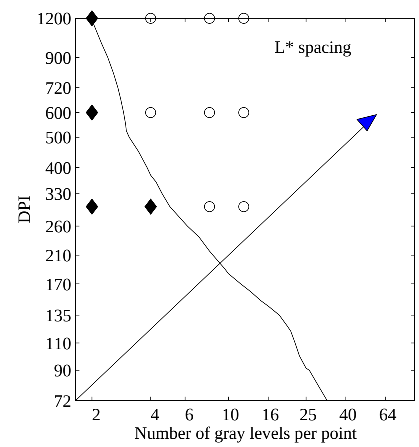

Figure 4. Comparision of stimulus conditions that generate S-CIELAB ΔE values of 1.0 (see Figure 3) to stimulus conditions that generated "threshold" and "suprathreshold" judgements. Threshold judgements (denoted with "o") refer to stimulus conditions that were indiscriminable from the image rendered at 1200 dpi with 256 graylevels. Suprathreshold judgements (denoted with "+") refer to stimulus conditions that were always discriminable from the 1200 dpi, 256 graylevel image. Again, the arrow indicates the direction of increasing image quality.

S-CIELAB, like CIELAB, predicts perceptual **thresholds** for detecting the difference between two retinal images and, in this case, the difference between a contone and a halftone grayscale image. Each curve in Figure 3 denotes a particular threshold value. Figure 4 compares the predictions of the S-CIELAB metric to the empirical data we collected in our experiments. The empirical data are plotted as a function of the pixel addressability conditions (bits/pixel and pixels/inch) and categorized into one of two categories: threshold and suprathreshold judgements. Threshold judgements refer to conditions in which images were perceptually equivalent to the 1200 dpi, 8 bit (256 levels) grayscale image (1200 dpi with 4, 8 and 12 levels, 600 dpi with 4, 8 and 12 levels and 300 dpi with 8 and 12 levels). Suprathreshold judgements refer to conditions in which subjects could always tell the difference between the 1200 dpi, 8 bit grayscale image and the halftoned image

(300, 600 and 1200 dpi with 2 levels and 300 dpi with 4 levels). The data are plotted in this way to illustrate the following observation: S-CIELAB values of 1.0 separate the threshold and suprathreshold stimulus conditions. When the S-CIELAB difference metric was greater than 1.0, subjects could always perceive the difference between a halftone and contone image. In other words, halftone texture was visible in images with S-CIELAB values greater than 1.0. When the S-CIELAB difference metric was less than 1.0, subjects could not perceive the difference between a halftone and contone image.

S-CIELAB makes predictions that are also consistent with other experiments we have conducted on the visibility of halftone texture in color images [17, 18]. For example, the metric predicts that the increase in image quality with increasing grayscale depth is greater for black, magenta and cyan, in that order. S-CIELAB predicts that there is no improvement in image quality with increasing the number of levels for the yellow inks. These predictions are consistent with our own observations and support the design decisions we made for the HP Photosmart Printer.

Conclusion

This paper posed the question: What is the best way to allocate bits to optimize photographic image quality? To answer this question, we developed methods for simulating the output of devices with different addressability and grayscale capabilities. We developed methods for quantifying how people perceive the image quality of such devices. And we developed metrics to predict these subjective judgements. All three methods converged upon the same answer: The most *efficient* way to generate high-quality photographic images is to increase the number of bits/pixel or graylevels/dot.

References

1. J. E. Farrell and B. A. Wandell, Scanner linearity, *Journal of Electronic Imaging*, **2**, 225-230, (1993).
2. P. L. Vora, J. E. Farrell, J. D. Tietz, and D. H. Brainard, Linear models for digital cameras, *Proceedings of the IST 50th Annual Conference*, (1997).
3. W. R. Anthony and J. E. Farrell, CRT-Display simulation of printed output, *SID 95 Digest*, 209-212, (1995).
4. R. W. Campbell and R. W. Gubisch, Optical quality of the human eye. *Journal of Physiology*, **186**, 558-578, (1966).
5. R. W. Campbell, The human eye as an optical filter. *Proceedings of the IEEE*, **56**, 1009-1014, (1968).
6. J. Kajiya and M. Ullner, Filtering high quality text for display on raster scan displays, *Computer Graphics*, **15**, 7-15, (1981).
7. J. E. Farrell, Grayscale/Resolution tradeoffs in image quality. *Raster Imaging and Digital Typography II*, (ed. R. Morris and J. Andre'), Cambridge University Press, 63-80 (1991).
8. J. E. Farrell, J. Is GPA a good measure of IQ?, *Proceedings of the SPIE*, February, (1997).
9. L. D. Silverstein, J. H. Krantz, F. E. Gomer, Y. Yeh, and R. W. Monty, Effects of spatial sampling and luminance quantization on the image quality of color matrix displays. *Journal of the Optical Society of America*, **7**, 1955-1968 (1990).
10. J. E. Farrell, Grayscale and resolution tradeoffs for photographic image quality, *Proceedings of the SPIE 3016*, February (1997).
11. R. Floyd and L. Steinberg, An adaptive algorithm for spatial gray scale, *SID 75 Digest*, 36-37, (1975).
12. D. A. Silverstein and J. E. Farrell, Quantifying perceptual quality, *Proceedings of the 51st Annual Meeting of the IST*, May (1998).
13 D. A. Silverstein and J. E. Farrell, The relationship between image fidelity and image quality, *Proceedings ICIP-96*, **1**, 881-884, (1996).
14. A. J. Ahumada and C. H. Null, Image quality: A multidimensional problem, *SID 92 Digest*, 851-854, (1992).
15. J. E. Farrell, X. M. Zhang, C. J. Van den Branden Lambrecht, and D. A. Silverstein, Image quality metrics based on single and multi-channel models of visual processing, *Proceedings of the IEEE COMPCON* , (1997).
16. X. M. Zhang and B. A. Wandell, A spatial extension to CIELAB for digital color image reproduction, *SID 96 Digest*, 731-734, (1996).
17. X. M. Zhang, J. E. Farrell, and B. A. Wandell, Applications of a spatial extension to CIELAB, *Proceedings of the IS&T/SPIE Symposium on Electronic Imaging*, (1997).
18. X. M. Zhang, D. A. Silverstein, J. E. Farrell, and B. A. Wandell, Color image quality metric S-CIELAB and its application on halftone texture visibility. *In Proceedings of the IEEE COMPCON* Spring (1997).

Biography

Joyce Farrell received her B.A. in Experimental Psychology from the University of California at San Diego in 1976 and her Ph.D. in Visual Psychophysics from Stanford University in 1981. Prior to joining Hewlett Packard in 1985, Dr. Farrell conducted research at Copenhagen University in Denmark, New York University, the NASA Ames Research Center and the Xerox Palo Alto Research Center.

farrell@hpl.hp.com