Bits, Bytes, and Square Meals in Digital Imaging

R. W. G. Hunt
University of Derby, England

Abstract

The robust nature of digital signals makes them a very desirable vehicle for communication; but the number of bits required for pictorial images can be very high unless systems are carefully designed to fit the perceptual properties of human vision. Features where important savings can be made in the bit requirements include: adjusting the light level in scanners; using a log or power function instead of a linear function for tonal quantization; allowing for the limited reproduction gamut; using luminance and chrominance signals to achieve bandwidth compression; and allowing for the modulation transfer function of the eye. By these means, large savings in bit requirements have been achieved in both desk-top publishing and broadcast television.

Introduction

Full 24 bit color is often claimed for digital imaging systems. With 8 bits in each of the red, green, and blue channels, 16,777,216 different color signals can be produced. But what does this mean in practice? The number of colors that can be distinguished is a matter of some debate, but it has been quoted as 10 million,[1] and for a 24 bit system to reproduce these it would have to sample color space almost uniformly in visual terms. But many systems digitize linear signals, and these are notoriously non-uniform perceptually. A perceptually uniform grey scale which is widely used is that provided by the L* function of the CIELAB and CIELUV color spaces standardized by the CIE. The range of L* values that is used in imaging depends on the display medium and the viewing conditions, but the maximum can probably be taken as extending from 10 (for a black seen under ideal conditions) to 100 (for a white); if a just notice-able difference is taken to be one unit of L* then the minimum number of tonal levels required to avoid spurious contouring in gradually changing areas would be 100 - 10 = 90, which would need 7 bits. But, when linear signals (percentage reflectance) are used, the difference between L* values of 10 and 11 is the difference between 1.126 and 1.261, or 0.135, and steps of this size have to be used to cover the range of linear values from 1.126 to 100, or 98.874. The number of tonal levels therefore rises to 98.874/0.135 which is equal to 733, requiring 10 bits. Hence, using linear signals, if the same number of bits is required in each channel as is required in a grey scale, the total number of bits in the three channels rises to $10 \times 3 = 30$ in order to reproduce all the distinguishable colors.

If a system has to handle input images that can vary in average density, even higher numbers of bits are required. Acceptable photographic transparencies can vary in transmittance for their whites by a factor of about 4 to 1, and negatives by 16 to 1 or more. To cover an additional range of 16 to 1 requires a further 4 bits in each channel, making a total of $14 \times 3 = 42$ bits.

If we consider an image of spatial definition comparable with 35 mm photography, we require a pixel array of about 2300×3450, which is equal to approximately 8 million pixels. (This number is obtained by assuming a film resolution[2] of about 46 cycles, or 92 pixel-pairs, per mm). Taking now both the tonal and spatial requirements of the image, the total number of bits required becomes $14 \times 3 \times 8 = 336$ megabits, or 42 megabytes.

While numbers of this magnitude can be handled in graphic-arts printing, for other applications, such as broadcast television and desk-top publishing, they would make digital imaging unattractive, if not impossible, and means have to be found for eliminating information that is visually imperceptible. Even in graphic-arts printing, handling very high numbers of bits usually results in longer processing times which is an economic disadvantage.

Adjusting the Light Level in Scanners

One conceptually simple way of reducing the bit requirement for images of variable average density is to regulate the level of light incident on them in scanners; if this is done in inverse proportion to the average trans-mittance of the image, then it is to be expected that 4 bits could be saved in each channel, and the number of megabits comes down to $10 \times 3 \times 8 = 240$ megabits or 30 mega-bytes. An alternative, and usually more convenient, way of achieving the same result is to digitize only those levels that are included in the range from a suitably chosen image white to the area corresponding to the darkest part of the image.

Using Non-Linear Scales for Tonal Digitization

If, for digitization, signals can be used that are more visually uniform than linear signals, then the number of bits can be further reduced. A logarithmic scale is much more uniform than a linear scale, and such a scale can be represented as $50\log_{10}(100R)$, where R is the reflectance factor; this representation, which is equal to 100 minus twice the density, is chosen so that when R is 1.0 (per cent reflectance equal to 100) the signal has a value of 100. A logarithmic amplifier can be used to provide such signals. A logarithmic scale is least uniform at the light end, so that the number of levels then depends on the difference, on the

log scale, between L* values of 100 and 99, and this is the difference between 100 and 99.436, or 0.564; hence steps of this size have to be used to cover the range of values from 100 to 2.577 (the value on the log scale corresponding to L* = 10). The number of tonal levels therefore becomes (100 - 2.577)/0.564 which is equal to 173, requiring 8 bits. This is an important improvement over the 10 bits needed by the linear scale.

Another alternative is to use a signal similar to the gamma-corrected luminance signal,[3] E_Y' used in broadcast television. This scale can be represented as $(100R)^{1/2.2}$ (because E_Y' is not a true luminance signal this relationship is only exactly true for neutral colors, but it is also approximately true for desaturated colors). This representtion is used so that, again, when R is 1.0 (per cent reflectance equal to 100) the signal has a value of 100. The E_Y' signal is least uniform at the dark end, so that the number of levels then depends on the difference, on the E_Y' scale, between L* values of 10 and 11 and this is the difference between 13.013 and 13.701, or 0.688; hence steps of this size have to be used to cover the range of log values from 100 to 13.013 (the value on the E_Y' scale corresponding to L* = 10). The number of tonal levels therefore becomes (100 - 3.013)/ 0.688 which is equal to 126, requiring 7 bits. However, the 128 levels provided by the 7 bits leave little margin for error, so 8 bits can be regarded as necessary for both logarithmic and E_Y' signals.

If the same number of bits is required in each channel as is required in a grey scale, the total number of bits ill the three channels for tonal resolution becomes $8 \times 3 = 24$, and the total for tonal and spatial resolution becomes $8 \times 3 \times 8$ = 192 megabits or 24 megabytes.

Allowing for the Limited Reproduction Gamut

In nature, some colors are produced by physical effects, such as interference, and when dyes or pigments are used a very great variety of colorants occurs. In imaging, the gamut of reproducible colors is always more limited. In self-luminous displays, red, green, and blue additive primaries limit the gamut to a triangle on a chromaticity diagram. In subtractive reproductions, the dyes or inks used limit the gamut both by the restrictions imposed by usually using only cyan, magenta, and yellow, and also by the unwanted absorptions which all subtractive colorants exhibit. Furthermore, when subtractive colorants are used on reflecting substrates, the unwanted absorptions are increased relative to the wanted absorption This is because, when the density is low, the light levels attenuate slowly for the multiple passes of the light between the substrate and the air-vehicle interface; but when the density is high the light levels attenuate quickly; and the densities of the unwanted absorptions are always much lower than those of the wanted absorptions.

When the image is broken up into dots, as in conventional graphic-arts printing, the effects of the unwanted absorptions are further increased. This can be seen in the following example, in which, for simplicity, a transparent image is considered (similar effects occur in reflection images, but they are more complicated to describe). Suppose a cyan colorant has a red density of 2, a green density of 0.6, and a blue density of 0.3, corresponding to percentage transmittances of 1, 25, and 50. In a continuous tone system, if the amount of colorant is reduced so that the red density is 0.297, the green and blue densities will be 0.089 and 0.045, respectively, and the corresponding percentage transmittances are 50.5, 81, and 90, respectively. In a half-tone system, 50 per cent dots would produce a red transmittance of 1/2(100 + 1) = 50.5 per cent, and the corresponding green and blue transmittances would be 1/2(100 + 25) = 62.5, and 1/2(100 + 50) = 75, respectively. Thus for the same red transmittance, compared to the con-tinuous tone system, the half-tone system results in percent-age transmittances for green of 62.5 instead of 81, and for blue of 75 instead of 90; this reduces the gamut, and makes colors containing the cyan ink darker (except where corrections can be made by reducing the amounts of magenta and yellow inks by masking). There is also a difference in hue for the intermediate amounts of cyan colorant. In a half-tone system, all the colors in a scale of cyan dots of different sizes will have the same dominant wavelength; but, in a continuous tone system, the ratio of green to blue transmittances will be greater in the mid-scale regions than at the maximum colorant end. This can be seen from the figures given above; for a red density of 0.297, the ratio of blue to green transmittance for the continuous tone system is 90 to 81, or 1.11, but for the half-tone system it is 75 to 62.5, or 1.2 to 1; whereas for a red density of 2, the blue to green transmittance ratio is 50 to 25, or 2 to 1, for both systems. Thus the continuous tone system is greener than the half-tone system in the mid-scale region.

In graphic-arts printing the use of a black ink is normal, in order to obtain darker blacks, and it also enables grey- component replacement to be used to reduce ink costs, facili-tate ink drying, and improve sharpness. But the use of more than three colored inks in subtractive systems is increasing for special applications in order to increase gamut and reduce metamerism;[4] typically orange, green, and violet inks are used in addition to the usual, cyan, magenta, and yellow.

So how many bits are required to quantize color space? As already mentioned, 8 bits is adequate for quantizing the L^* scale representing grays, and we must now consider how many more are required to cover the rest of the space. This can be estimated using the other two variables, a^* and b^*, of the CIELAB space.[5] If 8 bits are used for each of these signals over the range – 127 to + 128 for each of these variables, then the size of each step would be 1 ΔE unit in the CIELAB space. This might sound adequate, but it must be remembered that the CIELAB system itself has significant non-uniformities; and, if both variables changed together and there was also a change of 0.5 in L^*, then the total change would be $(1^2 + 1^2 + 0.5^2)^{1/2}$ which is equal to 1.5; a ΔE of this size could cause spurious contouring effects. However, the range of a^* and b^* values that can be reproduced is more like - 80 to + 80 in a^* and - 80 to + 110 in b^*, and by quantizing these ranges the highest value of ΔE is close to 1.0; in the presence of the random fluctuations (noise) which are always present in real systems, this ΔE level is usually found to be satisfactory.

There is no point, in a system, in processing digital signals in intermediate stages that correspond to colors that are outside the reproduction gamut; and, although values

of a^* and b^* do reach + 80 and – 80 (and + 110 in b^*), they usually only do so for colors at one level of L*. The reproduction gamut in any one hue plane is therefore approximately triangular in shape with the maximum value of a^* or b^* at one comer of the triangle and white and black at the other two corners. This means that the volume used in the color space is reduced to about a quarter, and if only these values are digitized then the total number of megabits can be reduced from $(8 + 8 + 8) \times 8 = 192$ to $(8 + 7 + 7) \times 8 = 176$, or 22 megabytes. It may be difficult in practice to restrict signals to a reproduction gamut but, if it can be done, this reduction in bit level is available.

Using Luminance and Chrominance Signals to Achieve Bandwidth Compression

It is well known that, in broadcast television, by using a luminance and two chrominance signals, considerable savings in bandwidth can be made[6]. The red-green signal can be reduced to one quarter of the bandwidth of the luminance signal, and the yellow-blue signal to about one tenth. In the NTSC system, these reductions are only made in the hori-zontal direction. In the PAL and SECAM systems, a reduction to one quarter is made in both signals it both the horizontal and vertical directions. In broadcast systems, because of gamma correction, some of the luminance is car-ried by the chrominance signals, and this limits the amount of lossless compression that can be made. In closed systems, such as desk-top publishing, if a true luminance signal could be used, one chrominance signal could be compressed to one quarter and the other to one tenth in both the vertical and horizontal directions. Compared with a system having equal bandwidth, b, in all three signals, it should therefore be possible to use only b for the luminance signal, and (l/16)b for red-green, and (l/l00)b for yellow blue. The number of megabits required could then be reduced from $(8+7 +7) \times 8 = 176$ to $8 \times 8 + 7 \times (8/16) + 7 \times (8/100) = 68.06$, or 8.51 megabytes. However, as in broadcast television, it would probably be difficult to separate luminance from chrominance completely, so a more realistic scenario is to restrict the bandwidth to a quarter for each chrominance signal; the number of megabits then becomes $8 \times 8 + 7 \times (8/4) + 7 \times (8/4) = 92$, or 11.5 megabytes.

Allowing for the Modulation Transfer Function of the Eye

As in other optical imaging systems, the modulation of the light in the eye becomes progressively smaller as the spatial frequency increases. The number of tone levels required to avoid contouring in the image therefore decreases as the spatial frequency increases.

In photography, because of scattering of the imaging light in the layers of photographic materials, the tonal resolution decreases as the spatial frequency increases, but the effects are not usually noticeable because they fit the visual modulation transfer function reasonably well.

The Joint Photographic Experts Group (JPEG) of the International Standards Organization has devised a very effective way of using this visual feature in digital imaging to provide further large economies in bit level. In the JPEG algorithm the image is dealt with in square sub-areas consisting of blocks of 8×8 pixels (the square meals). In each block the signals are transmitted through electronic filters that divide them into spatial frequencies. The filters range from d.c. to 4 cycles, with intermediate frequencies of 1, 1 ½, 2, 2 ½, 3, and 3 ½ both the vertical and horizontal directions, in all possible combinations, making a total of 64 filters altogether (since cosine functions are used, this process is known as the Discrete Cosine Transform DCT). Such a system is similar to a two-dimensional Fourier transform of the signal, and, in the absence of any compression, should make reconstruction of the original signal possible. But, by imposing increasingly restricted numbers of tonal levels pemmitted as the spatial frequency increases, it is possible to reduce the total number of bits required without serious impaimment of the final image.

Important further reductions in bit level are obtained in the JPEG algorithm by using two other techniques. First, because, in most pictorial images, neighbouring pixels tend to have similar values, only the differences in successive pixel values are encoded for the d.c. components; this differential d.c. coding is a powerful means of reducing the bit level, and is also used in other compression algorithms, such as that used in Photo CD. Second, the statistical nature of the values in the other filters is used to encode the information more efficiently, and this further reduces the bit level.

According to the application, the JPEG algorithm can be used to provide compression ratios of between 12 and 100. Compression that is perceptually lossless requires the lower end of this range. This reduces the number of megabits required from $8 \times 8 + 7 \times (8/4) + 7 \times (8/4) = 92$ to 92/12, or 7.67 megabytes. As there are 8 million pixels, this represents an average of about 1 bit per pixel, a remarkable degree of compression.

In half-tone systems, the tonal modulation has to be produced by using dots of different sizes, and in digital sys-tems these are usually produced by using different numbers of micro-dots in arrays of about 12×12 pixels. Such arrays can produce 144 different dot sizes (and hence 145 tone levels including the white). If the dots are used in a transparency, or on metal, when there is no diffusion of the light, then the tonal levels are spaced linearly and the number of levels that can be produced is far too few to avoid contouring. But, when the dots are printed on to paper, the light passes through the dot structure twice, once on the way into the paper, and a second time, after diffusion in the paper, on the way out; this has the effect of increasing the effective density of the mid tones, a phenomenon known as *dot gain*[7], and this makes the relationship between per cent dot size and L^* more nearly linear. The 145 levels then provide an average spacing of about 0.55 in L^* over the range of 20 to 100 in L^* which is typically the most that is available on paper. In some desk-top publishing systems, such as those using ink-jet technology, the number of micro-dots per inch may be about 300; higher numbers of micro-dots per inch may be desirable for text, but 300 dots per inch can give good pictorial results in continuous tone systems, and also in desk-top publishing if special techniques are used as will be described below. In 300 dot-per-inch systems, if a 12×12 array is used for dot formation, then the half-tone resolution becomes only 25 lines per inch, which is very poor. Several methods are used to

overcome this problem including: using fewer than 144 dot-sizes to obtain better spatial resolution at the expense of some risk of contouring; using *dispersed dots,* in which the micro-dots are dispersed over the pixel array instead of being clustered at the centre to form a traditional type of half-tone dot; using *stochastic screening* in which the micro-dots are dispersed over the array randomly (instead of regularly as in dispersed dots); and *error diffusion* in which the error from an area being either inked or not inked is carried forward to subsequent areas on an ongoing basis. Some desk-top publishing systems also take advantage of the modulation transfer function of the eye, but by using simpler algorithms than that used in JPEG. In the Iris system, any number of microdots of ink up to 32 can be printed in each pixel, giving 32 tonal levels. This is not a sufficient number of levels to avoid contouring so a 4×4 pixel dithering pattern (in which the average reflectance can have any one of 16 different values) is superimposed which, together with the 32 levels of inking, gives 512 tonal levels. The system operates at 300 dots per inch, but the dithering pattern reduces this to 75 dots per inch. Coarse detail is therefore reproduced at 75 dots per inch with 512 tonal levels, giving more than enough tonal levels at low spatial frequencies. In fine detail the dithering pattern is ineffective, so that the resolution increases to the full 300 dots per inch, but with only 32 tonal levels; however, at these high frequencies this number of tonal levels is more than enough because of the low modulation transfer factor of the eye at these high spatial frequencies. Other combinations of inking levels and dithering are possible, and a particularly interesting one is to use 8 inking levels with a 4×4 dithering pattern; at 300 dots per inch, this provides 128 levels at 75 dots per inch for coarse detail, and 8 levels at 300 dots per inch for fine detail.[8]

High Definition Television

When compared to normal television, in typical high-definition television, the number of lines is about twice as many, and the number of resolvable pixel-pairs along each line is also about twice (in order to match the larger number of lines) and this is increased by a further factor of 4/3 to allow for the increase of the aspect ratio from 4/3 to 16/9. These considerations increase the band-width required by about $2 \times 2 \times 4/3 = 5.33$; if the normal system is regarded as having a bandwidth of 5 MHz, then the bandwidth required for a high-definition system becomes about 27 MHz. If this is digitized using 8 bits to represent the amplitude of the signal, then the bit rate required becomes $27 \times 8 = 216$ megabits per second. High definition systems usually use separate bandwidth for the chrominance signals, and these are normally compressed, in both the horizontal and vertical directions, to half of that used for the luminance signal; the total amount of bandwidth is thus increased by $1 + (1/2)^2 + (1/2)^2 = 1.5$, making a total of $216 \times 1.5 = 324$ megabits per second. In motion pictures, as are used in television, except immediately after a scene break, the information in each frame can usually be related to that in the previous frame by means of motion vectors. Use is made of this in the MPEG (Motion Picture Experts Group) algorithm used in high definition television, and this, together with a JPEG type of algorithm, can result in a compression factor of 18, enabling the bit rate to be reduced to 324/18 = 18 megabits per second. In a 30 pictures per-second system this is equivalent to 18/30 = 0.6 megabits per picture; a high definition television picture usually has about 2 million pixels, so that this represents 0.6/2 = 0.3 bits per pixel. The still picture considered earlier required about 1 bit per pixel; the lower value of 0.3 for the television picture can be taken to illustrate the benefit of being able to use motion vectors, but other factors will also have contributed to the difference in this comparison.

Conclusions

The advantages of digital imaging can be obtained without the penalties associated with requiring very high levels of bits in the images, by using various means of compression. Those that are useful include: adjusting light levels in scanners; using non-linear scales for tonal digitization; allowing for the limited reproduction gamut; using luminance chrominance signals to achieve band saving; and allowing for the modulation transfer function of the eye. By engineering color[9] in these ways, the number of bits per pixel required in a high resolution image can be reduced from about 42 to about 1 for still images, and to about 0.3 for motion pictures.

Acknowledgments

I am very grateful to Tony Johnson and Lindsay MacDonald for kindly helping me in the preparation of this paper.

References

1. D. B. Judd and G.Wyszecki, *Color in Business, Science, and Industry.* 3rd edition, Page 388, Wiley, New York (1975).
2. R. W. G. Hunt, *The Reproduction of Colour.* 5th edition. Page 421, Fountain Press, Kingston-upon Thames, England (1995).
3. M. Anderson, R. Motta, S. Chandrasekar, and M. Stokes, Proposal for a standard default color space for the internet - sRGB, *Proc. IS&T/SID Fourth Color Imaging Conf.,* 238-246 (1996); (see page 198, this publication).
4. L. W. MacDonald, J. M. Deane, and D. N. Rughani. Ex--tending the colour gamut of printed images. *J. Phot. Sci.*, **42**, 97-99 (1994).
5. L. W. MacDonald and J. M. Deane, A comparative study of the errors caused by quantizing colour images in the CIELAB and PhotoYCC colour spaces, *J. Phot. Sci.,* **41**, 106-107 (1993).
6. R. W. G. Hunt, Why is black-and-white so important in color? *Proc. IS&T/SID Fourth Color Imaging Conf.,* 54-57 (1996); (see page 189, this publication).
7. R. W. G. Hunt, Spatial and tonal resolution in desk-top publishing, *J. Phot. Sci.,* **42**, 85 (1994).
8. R. W. G. Hunt, *The Reproduction of Colour,* 5th edition, Page 697, Fountain Press, Kingston-upon Thames, England (1995).
9. J. C. King, The color engineer, *Proc. IS&T/SID First Color Imaging Conf.,* 99-100 (1993).

☆ This paper was previously published in *IS&T's 5th Color Imaging Conference Proc.,* p. 1 (1997).