Eyeing the Camera

Robert W.G. Hunt
Color Consultant, Salisbury, England
University of Derby, Derby, England

Abstract

Imaging systems are compared to the human eye in terms of acquisition, spectral sensitivity, transmission, and display. Although the performance of imaging systems is generally of a high standard, there is still room for improvement. At the acquisition stage, the use of flash, or infra-red lighting, results in images with some unnatural features, and there is a general absence of systems that give monochrome images at very low light levels, as are provided by the rod system in the eye. The overlapping nature of the spectral sensitivities of the cones results in unwanted cone stimulations which reduce reproduction gamuts; in printing, extra colorants, such as orange, green, and violet, can be used to extend the gamut and reduce metamerism. Commercially available imaging systems incorporate spectral sensitivities that do not usually exactly match a set of color matching functions, which is a requirement for special applications where high colour accuracy is important. The transmission of image signals in broadcast television makes use of the important luminance/chrominance principle, but full benefit is not achieved because of gamma correction, and only one system makes use of the reduced resolution of the yellowness-blueness channel of the eye. Successful bit reduction in digital images is achieved by taking advantage of the reduced contrast sensitivity of the eye at high spatial frequencies and other effects, but a reduction in the consequent artefacts is desirable. In the visual system, the display is in the cortex, which has an enormous ability to interpret the retinal signals so as to recognise objects, including very efficient compensation for changes in illumination level and color; some improvements in the similar compensation provided in imaging systems are desirable, together with increases in the dynamic ranges available, especially in display devices. Automatic image-enhancement adjustments can be included when making images, but there is room for more sophisticated techniques that avoid impairing some types of scene. A low-cost image-display device for the mass market that is more convenient than the cathode-ray tube remains an important challenge.

Introduction

A pictorial imaging system is an attempt to copy the human visual system. It is therefore instructive to compare the two systems. It is convenient to do this under the headings: acquisition, spectral sensitivity, transmission, and display.

Acquisition

The retina of the human eye contains two different types of receptor, rods and cones. There are about 100 million rods and about 6 million cones in each eye. The distribution of these receptors is very far from uniform across the area of the retina, there being no rods in the central region, and no cones in the periphery, with a gradual transition from one type to the other. It is the function of the cones to provide color vision and high visual acuity in the central region, while the rods provide monochromatic vision of higher sensitivity but lower resolution in more peripheral regions. It is customary, in considering imaging systems, to ignore the rods, on the grounds that, at levels of illumination typical of daylight and good artificial light, the cones inhibit the rods. However, at levels of illumination intermediate between daylight and moonlight, the rods and cones operate together, and even at daylight levels the rods are not inactive because they provide peripheral vision, there being no cones outside a visual angle of about $\pm 40^0$.

The sensitivities of the human eye, and of cameras, depend on many factors. The receptors in the retina are situated in a geometrically random mosaic. Each cone is sensitive predominantly only to either the reddish, the greenish, or the bluish third of the spectrum, and this arrangement restricts the sensitivity. However, the spectral sensitivities of the long (ρ) and medium (γ) wavelength cones are broad (as shown in Figure 1) and this enhances their light-catching ability. The β cones have a narrower absorption band, but, as there are fewer of them (only about one twentieth as many as the γ cones), their effect on the overall sensitivity is small; the reduced resolution resulting from the small number of the β cones matches the reduced resolution of the eye to blue light when it focuses the yellow-green light to which the ρ and γ cones are most sensitive, there being no correction for chromatic aberration.

Electronic cameras usually use arrays of charge-coupled devices (CCD) or complementary metal-oxide semi-conductors (CMOS). For high quality television, cameras incorporate highly efficient dichroic prism blocks to image the red, green, and blue light on to three separate detector arrays; in this case very little light need be wasted. However, in camcorders, and digital still cameras, the high cost and bulk of the prism assembly precludes its use, and a single array is normally used with a pattern of filters over its pixels. This results in both reduced sensitivity, on account of the light absorbed in the filters, and also reduced resolution on account of the fact that not every pixel provides red, green, and blue signals. Some filtered arrays, instead of using red, green, and blue filters, each with their loss of about two-thirds of the light, incorporate some cyan, magenta, yellow, or colorless filters; this increases the amount of light incident on the cells, but signal-to-noise ratios deteriorate when signals are subtracted from one another to recover the red, green, and blue information.

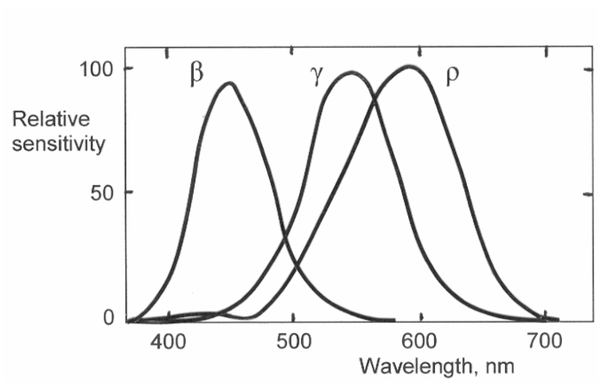

Figure 1. Curves typical of those believed to represent the spectral sensitivities of the three types of cone of the human eye.

In the earliest form of color photography, arrays of red, green, and blue filters were used over black-and-white emulsions. However, the consequent loss of light and resolution of this arrangement has led to them being superseded by multi-layer subtractive systems. In these systems, red-sensitive, green-sensitive, and blue-sensitive layers are coated on top of one another, an arrangement in which little or no light need be lost. CMOS sensors in which red, green, and blue light is absorbed at different depths of the silicon layer have been described,[1] and these too have the potential of wasting little or no light. However, multi-layer acquisition devices may result in loss of resolution because of diffusion by the upper layers, and loss of speed because of absorption in the upper layers. In films sensitive enough for use in cameras, because fast emulsions have a natural blue sensitivity, it is necessary to coat the blue-sensitive layer at the top, whereas it is desirable to have the green-sensitive layer at the top because this contributes most to the sharpness of the image. In print films, where speed is less important, this is done; in camera films the top blue-sensitive layer is made to have as low a turbidity as possible to minimise the reduction in resolution of the lower layers.

The sensitivity achieved in film cameras is now such that, by using the fastest films and large aperture lenses, usable images can be obtained down to illuminances of a few lux, and electronic cameras can reach a similar level; this high sensitivity is obtained at the expense of reductions in resolution and color rendering, and increases in noise. The use of flash lighting with still cameras (both photographic and electronic) avoids these impairments, but results in unnatural lighting geometry and some artefacts such as red-eye; to avoid the unpleasantness of a continuous bright light, some camcorders provide infra-red illumination when there is little or no light, but the images then obtained do not correspond to those provided by vision either in spectral rendering or in geometry of illumination. To obtain images that match the quality of vision, at low light levels, without the use of flash or other supplementary lighting, remains a challenge.

Cameras usually only give color pictures, thus representing the operation of the visual cone system. But experimental color films have been made with a fast black-and-white emulsion layer, coated above the color-sensitive layers, to give black-and-white images in low levels of illumination, the layer being inhibited at high levels as is the case with the rods and the cones.[2] Electronic cameras can provide a facility for averaging the red, green, and blue signals from the sensor at low light levels to give similar black-and-white pictures; by averaging the signals over groups of several cells of each color, further increases of signal-to noise ratio can be achieved to give black-and-white pictures at even lower levels of illumination, and, although the resolution is reduced, this is also the case for the eye.

The relatively high entry cost of electronic cameras, as compared to single-use film cameras, represents a limitation for their application to the mass market, particularly in the less developed parts of the world.

Spectral Sensitivity

As has already been mentioned, the spectral sensitivities of the three cone types are broad; they also overlap one another considerably. Although the broadness is good for collecting light, it results, together with the overlap, in less wavelength discrimination than would be provided by sensitivities that were narrower and more separated; however, evidence has been found for sharpened sensitivities[3,4] and this suggests that some differencing of the cone signals may occur at an early stage along the visual pathway.

The overlapping of the cone sensitivities has a profound effect on trichromatic imaging. Although red light can be found that stimulates the ρ cones alone, and blue light can be found that stimulates the β cones almost alone, there is no green light that stimulates the γ cones without resulting in substantial stimulation of the ρ cones and some stimulation of the β cones. Even if lasers are used (which tend to be inconvenient and costly), these unwanted stimulations cannot be avoided; with practicable primaries, such as the phosphors used in cathode-ray tubes, substantial unwanted stimulations occur in all three channels This means that displays using the addition of any red, green, and blue primaries result in images having a restricted gamut of reproducible colors. The regions of wavelengths controlled by cyan, magenta, and yellow colorants, as are used in the subtractive imaging systems, also result in large unwanted stimulations, and hence a restricted gamut. These restricted gamuts are widely used in imaging systems that are commercially very successful, and this is because the gamut of colors in most real scenes is usually only slightly, if at all, larger. In printing, however, extra colorants, such as orange, green, and violet, in addition to the usual cyan, magenta, yellow, and black, are sometimes used, and this extends the gamut usefully and also offers the possibility of reducing metamerism between an original object and its reproduction, a result which is desirable in mail-order catalogues.

The unwanted stimulations result in the theoretically correct spectral sensitivities for cameras having negative, as well as positive, parts to their functions. The negative parts can be simulated in electronic imaging systems by matrixing circuits, and this can result in correct color reproduction for colors that lie within the reproduction gamut; however, this is only achieved if the spectral sensitivities of the camera correspond to a set of color matching functions, and this is not usually the case. In films, matrixing is not possible, and the spectral sensitivity of the red layer is usually shifted to longer wavelengths, that of the blue layer to shorter wavelengths, and that of the green layer made narrower; this results in improved color reproduction for most colors, but some errors occur, an example being the tendency for some blue flowers to be reproduced pink. In one color film, a blue-green sensitive layer is incorporated in addition to the usual red-, green-, and blue-sensitive layers; this extra layer sends a developer-inhibitor to the red-sensitive layer in an attempt to simulate the negative part of the spectral sensitivity of that layer. But the effect is to subtract contrast from the red image, whereas what is required is to subtract exposure, so that other errors occur. Another approach is to lessen the shift of the red-layer sensitivity to longer wavelengths, and to restore the lost saturation of colors with greater inter-image effects. In systems where the film is always going to be scanned, it is possible to make the spectral sensitivities of the layers the same as those of an all-positive set of color matching functions and then to introduce matrixing at the electronic stage. This requires the following sequence of steps: converting the signals from the scanner to log signals; removing the effects of the unwanted absorptions of the image dyes; removing the inter-image effects; removing the effects of the tone reproduction of the film; anti-logging the signals; and finally applying the appropriate matrix.

Transmission

After absorption of the light in the cones and rods, analogue voltage signals are generated and passed through a complicated network of neurons. The exact functions of this solid-state micro-electronic network are not fully understood, but there is strong evidence that at some stage the ρ, γ, and β signals are converted into an achromatic signal indicating the amount of light, and two color-difference signals, one indicating whether the color is reddish or greenish, and the other whether it is yellowish or bluish. When the signals reach the ganglion cells in the retina they are changed from analogue voltages to pulses of constant voltage, of about 70 milli-volts each; signal modulation is then conveyed in terms of the frequency of transmission of these pulses, the frequencies ranging from a few per second to a maximum of about 400 per second. The signals are thus binary, but not digital. Another important feature of the visual transmission system is that it is restricted to about 1 million nerve fibres connecting each retina to the brain. This number of fibres has to service 6 million cones and 100 million rods, and hence a regime for compressing the amount of information sent has to be in place. In the centre of the retina, there are about as many fibres as receptors thus giving excellent resolution; but in the periphery as many as 100 or more rods feed into the same nerve fibre, which, although giving poor resolution, results in an important gain in sensitivity. Movement of the eyes in their sockets, and head movements, are then used to bring into the centre part of the visual field those parts that are needed to be scrutinised at high resolution.

In broadcast television, the red, green, and blue signals from the camera are similarly converted into an achromatic (luminance) signal and two color-difference (chrominance) signals. This conversion has two advantages. First, the luminance signal can be used to provide monochrome pictures on black-and-white receivers; this *compatibility* was important when color television was first introduced, but is now a minor factor. Second, the resolution required in the chrominance signals can be reduced substantially, thus saving bandwidth. The reason for this is as follows.[5] In the visual achromatic channel, a luminance discontinuity in the field can be detected if it corresponds on the retina to about one cone diameter, because all cones feed into that channel. But to detect a discontinuity in redness-greenness the discontinuity must correspond to a distance on the retina sufficient to ensure that it contains at least one ρ and one γ cone; because of the geometrically random pattern of the cones, this corresponds to a distance equal to about four cone diameters. The red-green resolution of the eye along a line is therefore only a quarter of that of the luminance resolution, and there is a similar reduction in a direction at right-angles to the line. On an area basis the red-green resolution is therefore only one sixteenth of that of the luminance signal.

In the PAL system of broadcast television the bandwidth devoted to the red-green signal is therefore only a quarter of that devoted to the luminance signal, and the signals are averaged over four picture lines so that the vertical resolution is also reduced to a quarter, resulting in a reduction to one sixteenth on an area basis. The PAL system also reduces the resolution of the yellowness-blueness signal similarly. However, because of the small proportion of the β cones, the visual resolution of yellowness-blueness discontinuities is much less, and in the NTSC system of broadcast television the yellowness-blueness chrominance signal is reduced to about one-tenth of that of the luminance signal; but in this system there is no reduction in chrominance resolution in the vertical direction. However, because of the effects of gamma correction, the luminance signal in these systems does not carry all of the luminance information, some being carried by the chrominance signals; so in high definition television (HDTV) the chrominance resolution is only reduced by a factor of two in each direction, resulting in a reduction by a factor of four on an area basis.

The standard adopted for color facsimile transmissions (color fax) uses signals equivalent to the L^*, a^*, and b^* variables of the CIELAB color space, and, although the consequent use of a true luminance signal would justify the full reduction to one sixteenth, the reduction used is still only to a half along a line.

In the eye, contrast sensitivity reduces as the spatial frequency increases. This also occurs in photographic materials. In digital imaging this means that fewer bits are required to avoid contouring as the spatial frequency increases; this provides the basis for an important part of the bit reduction provided by the JPEG and MPEG algorithms. These algorithms are so successful that in HDTV it is only necessary to provide 0.3 bits per pixel in the signal.[6] In JPEG 2000, instead of dealing with the image in sub-blocks of 8 by 8 pixels, the whole image is passed through a series of high pass and low pass filters horizontally and vertically, and this avoids a tendency for the final image to be 'blocky' in appearance. A further elaboration is to identify regions of interest (ROI), and to assign smaller numbers of bits to other areas, a technique similar to that whereby the eye fixates areas of interest so that they are imaged in the centre of the retina with its high resolution, the rest of the field being assigned to regions of lower resolution. Photographers can achieve similar results by using a limited depth of field to render sharp one part of a scene, such as a person's face or a close-up of a flower, while the rest of the scene is out of focus. In some digital systems the number of colors is limited to about 200 in a palette, and this can cause errors in color rendering. Artefacts caused by digitisation can be noticeable, and their reduction is a desirable goal.

Display

In the visual system it is the function of the cortex to interpret the signals it receives from the retina, and much emphasis is placed on the recognition of objects. Ambiguous diagrams demonstrate that the same signals from the retina can be interpreted by the cortex in different ways. The cortex can also extract different parts of an image, as is the case in recognising faces; this is illustrated by the technique of building composite face images from constituent parts to help witnesses provide security services with usable depictions of suspects.

By means of adaptation, the visual system adjusts its sensitivity, both overall, and differentially in its chromatic channels, to compensate for changes in both the level and the color of the illumination. Cameras with photo-electric exposure-controls provide automatic compensation for illumination level; film cameras without such controls depend on the exposure latitude, and subsequent printing control, of negative films, or on manual adjustments for slide films. Compensation in cameras for the color of the illumination is provided by electronic white-balance algorithms in electronic cameras; in film cameras it is provided by exposure latitude and printing control for negative films, and by the use of different films or colored filters for slide film. The visual system is extremely reliable in making these compensations, so that object recognition is largely unaffected by changes in illumination level and color such as occur in everyday life. The compensations made by cameras and printers work reasonably well most of the time, but some improvements are desirable.

The dynamic range of the visual system is wide enough to accommodate both sunlit and shadow luminances simultaneously, a range of about 1000 to 1. Acquisition of this type of range in imaging devices is possible, for example in negative film and in detector arrays; the dynamic range in CMOS arrays can be extended by capturing images at a series of gradually increasing exposure times.[7] But imaging display devices have dynamic ranges that are much less than 1000 to 1, their effective ranges usually being at best about 100 to 1, and for printing on newsprint only about 20 to 1. Some improvement in the rendering of brightly and dimly lit areas in the same image can be achieved by appropriately increasing (burning) or decreasing (dodging) the density at which they are rendered, but the result is not the same as that which would be offered by a display having a wide dynamic response.

When the image information is available as electronic signals in an intermediate stage, it is possible to incorporate various improvements, such as adjustments of tone reproduction, either overall, or locally (burning and dodging). Artefacts such as red-eye can also be eliminated. These image-enhancing adjustments are possible not only in images derived from electronic cameras, but also in those derived from film, by scanning the film.

In imaging, the most widely used display is the reflection print. If, instead of using optical printing, the film is scanned, the image-enhancing adjustments described above can be incorporated, as is done in digital photo-finishing.

The second most widely used display is the directly viewed self-luminous image, and, as this is produced by electronic signals, the image-enhancements can again be included. Self-luminous displays can be produced either by a cathode-ray tube (CRT), a liquid crystal display (LCD), a plasma display (PD), light-emitting diodes (LEDs), or organic light-emitting diodes (OLEDs) which have the potential for having a multi-layer structure so that every pixel can produce red, green, and blue light instead of only one of the three. The CRT is the most widely used technology, the quality being very high, and the price low as a result of very high volume manufacture. However, the device is bulky and heavy, and not very well suited to HDTV; strenuous efforts have been made to find a replacement technology suitable for the mass market, preferably thin enough to provide 'picture on the wall' presentation, but so far without much success. This remains as an important challenge to the imaging business.

A third display technology which is widely used is the projected image. If, instead of using direct optical projection, electronic signals are used to form the image, then, again, the image-enhancing adjustments can be incorporated. Electronic projection technologies include triple projection cathode-ray tubes, monochrome transmitting or reflecting liquid crystal arrays in beam-splitting projectors, and digital mirror devices. Considerable increases in the luminance of the displayed image can be achieved by using back-projection, in which the image is viewed by transmission on a translucent screen.

One avenue that is worth exploring is to try to develop a luminance-chrominance display,[8] so that the increase in the number of pixels for color could be reduced from three times (red, green, and blue) to only one and three-sixteenths (luminance and (red + green + blue)/16), thus hopefully reducing the complexity and cost of the device substantially.

Conclusions

The systems of imaging available at present provide a remarkable level of performance, availability, and ease of use. Ideally, they should achieve WYSYWYG, what you see is what you get. By comparing the different parts of imaging systems with the eye, it is possible to identify areas where improvement is desirable. At the acquisition stage, image-making down to levels of illumination in the region of a few lux is possible, but, because of impairments to picture quality, widespread use is made in still cameras of flash, and in camcorders of infra-red lighting; these devices result in departures from visual appearance, such as unnatural geometry of illumination, the occurrence of red-eye, and unusual spectral rendering. Moreover, the eye provides a service down to very much lower levels by switching from the cone system to the monochrome rod system with its greater sensitivity; imaging systems do not usually provide monochrome images at low light levels. The production, in imaging systems, of color pictures at low photopic light levels, and monochrome pictures at scotopic levels, without the use of supplementary lighting, is therefore a challenge. The overlapping nature of the cone spectral sensitivities results in limited reproduction gamuts, and in printing more use could be made of extra colorants to extend the gamut and reduce metamerism. As far as the spectral sensitivity of cameras is concerned, commercially available systems provide spectral sensi-tivities that do not usually correspond exactly to a set of color matching functions; the resulting errors of color rendering are not high, and are often unnoticeable in pictorial images, but, for special applications, where accuracy of color rendering is important, improvements in this area are desirable in both electronic cameras and in film. The transmission of images in broadcast television makes use of the luminance/chrominance principle although full advantage is not achieved because of the effects of gamma correction; only the NTSC system makes use of the reduced resolution in the yellowness-blueness channel. In digital imaging systems the JPEG and MPEG algorithms make excellent use of the reduced contrast sensitivity of the eye at high spatial frequencies, but a reduction in the consequent artefacts is desirable. The display in the visual system takes place in the cortex, where much interpretation of the image takes place, including very efficient compensation for changes in illumination level and color; some improvements in the similar compensation provided in imaging systems are desirable, together with increases in the dynamic ranges available. At electronic stages of imaging image-enhancing adjustments are possible, but there is room for progress in developing sophisticated automatic methods of improving images without impairing the quality of some types of scene. A more convenient replacement for the cathode-ray tube, that provides picture-on-the-wall presentation, remains a challenge in imaging; at present no display device makes use of the luminance/chrominance principle.

Acknowledgements

I am grateful to Dr Michael R. Pointer and to Dr Paul M. Hubel for helpful comments in the preparation of this paper.

References

1. *RPS Journal*, The Foveon X3 R-G-B chip. Page 141 (April 2002).
2. D. M. Zwick, Private communication. (1960).
3. D. H. Foster, Changes in field spectral sensitivities of red-, green-, and blue-sensitive colour mechanisms obtained on small background fields. *Vis. Res.*, **21**, 1433-1455 (1981).
4. A. B. Poirson and B. A. Wandell, Task dependent color discrimination. *J. Opt. Soc. Am.*, **7,** 776-782 (1990).
5. R.W.G. Hunt, The strange journey from retina to brain. *J. Royal Television Soc.,* **11,** 220-229 (1967).
6. R.W.G. Hunt, Bits, bytes, and square meals in digital imaging. *IS&T/SID 5th Color Imaging Conference, Scottsdale, Arizona,* page 1 (1997).
7. B. A. Wandell, A. E. Gamal and B. Girod, Common principles of image acquisition systems and biological vision. *Proc. IEEE,* **90**, No. 1 January, 5-17 (2002).
8. P. Hubel and R.W.G. Hunt, Configurations for color displays by the use of lenticular optics. US Patent Application (2001).

Biography

Robert Hunt worked for 36 years at the Kodak Research Laboratories in Harrow, England, taking early retirement as Assistant Director of Research in 1982. Since then he has been working as an independent color consultant. He has had two books published: The Reproduction of Colour, now in its fifth edition, and Measuring Colour, in its third edition. He has attended all the previous Color Imaging Conferences at Scottsdale and is a regular contributor of keynote papers.