The Visual Computer

Tutorial on color science

Gary W. Meyer

Program of Computer Graphics, 120 Rand Hall,
Cornell University, Ithaca, NY 14853, USA

The principles of tristimulus colorimetry are presented in tutorial fashion. The classic color matching experiments are described with an emphasis on the assumptions that are implicit in these tests and on the units of measure which should be used to record the results. The transformation to alternative sets of primaries is derived and the geometry of the resulting color spaces is illustrated. An annotated bibliography of relevant literature is also provided.

Key words: Color science – Computer graphics – Human factors

New address: Department of Computer and Information Science, College of Arts and Sciences, University of Oregon, Eugene, OR 97403-1202, USA

The principles of tristimulus colorimetry have been a fundamental topic in computer graphics ever since raster graphics introduced the use of color reproduction media such as television and film. The additive color reproduction techniques used in television are a straightforward application of the laws of color science, from filter selection for color scanning, through signal matrixing for NTSC video, to display on a color television monitor. Colorimetry is also important in working with film although its application is much more complex because the photochemical nature of the film emulsion makes its spectral sensitivity and response to light difficult to control and because the dyes in the processed film do not behave as independent primaries. Understanding the principles of color reproduction in film and television, controlling these media in order to get satisfactory results and communicating the results to other workers in the field of computer graphics all depend upon an understanding of color science.

It has also been realized that tristimulus colorimetry is important in computer graphics image synthesis. The weighted averaging process by which colors are selected in anti-aliasing problems is based on the laws of color science even if it has seldom been stated that way. Application of colorimetry has made standard color collections such as the Munsell Book of Color available for use as color palettes. Simple diffuse shading models which are expressed directly in terms of the RGB primaries of a color reproduction device have always had color science as their basis. As shading models have become more optically sophisticated the application of colorimetry has become explicit in the step which converts the synthesized spectral energy distribution into tristimulus values used to run a color reproduction device.

This article is a tutorial on the basic principles of color science. It was written because of the growing importance of this topic in computer graphics and to provide an alternative exposition of this material from the more traditional presentations. The approach stresses the fundamentals in the belief that building a good foundation is more important than treating some advanced topics superficially. An annotated bibliography is given at the end of the article where additional sources can be found.

2 The nature of color perception

When asked to describe a color or to arrange a set of random color chips, there is a natural tendency for people to organize their perceptual color space into a cylindrical coordinate system with dimensions hue, saturation, and brightness (Fig. 1). One set of definitions which has been proposed for these terms is (Burnham et al. 1963):

1. "Hue is the dimension of color that is referred to a scale of perceptions ranging from red through yellow, green, blue and (circularly) back to red."
2. "Saturation is the dimension of color that is referred to a scale of perceptions representing a color's degree of departure from an achromatic color (one lacking a distinguishable hue) of the same brightness."
3. "Brightness is the dimension of color that is referred to a scale of perceptions representing a color's similarity to some one of a series of achromatic colors ranging from very dim (dark) to very bright (dazzling)."

This simple organizational scheme will be used in this article when describing the color sensation which exists in a person's mind. It is important to note, however, that more refined versions of this nomenclature have been proposed (Hunt 1977, 1978) and are important for precise description of the color sensation.

It is fundamentally the electromagnetic energy which enters the eye and strikes the retina that produces the color sensation, but there are many variables of the stimulus which effect the hue, saturation, and brightness designation given to this sensation. The amount of enery present at each wavelength of the stimulus has the most obvious effect on the result as anyone who has seen a prism refract light will attest to. Variations in the spatial relationships between the elements in the field of view can also effect the perceived color. As examples of this, the visual angle subtended by an object effects its perceived brightness and the background against which an object is seen produces simultaneous color contrast, an effect which makes a grey square on a green background appear reddish. Temporal variations in the stimulus also effect the hue, saturation, and brightness sensation elicited, although these changes are often quite subtle and difficult to detect. Finally, an effect known as adaptation allows an observer to report the same hue, saturation, and brightness for objects in a scene illuminated by either daylight or tungsten light even though the magnitude and spectral characteristics of these types of light are quite different.

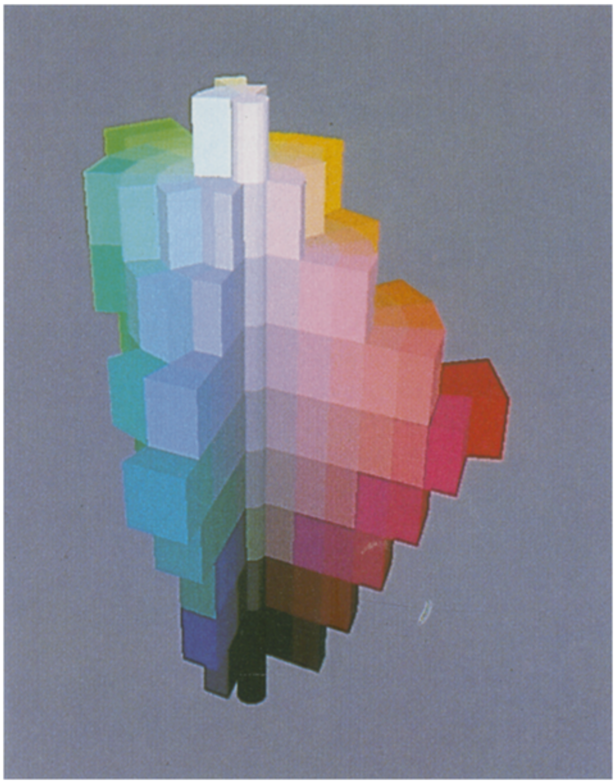

Fig. 1. Perceptual color space with dimensions hue (*angle*), saturation (*radius*), and brightness (*(height*). Cut away shows a plane of constant hue

Given the nature of the color sensation and the multitude of variables which effect it, a comprehensive human color vision model would be very complex. The color vector $\bar{c}$ with components hue h, saturation s, and brightness b would be functionally related to the wavelength λ, the spatial distribution x, y, z, and the temporal modulation t of the electromagnetic energy in the stimulus. Other less quantitative factors would also play a part such as the adaptive state of the observer a and the variation amongst observers v. All of these factors taken together plus others not listed lead to a functional description of the problem as (Burnham et al. 1963):

$$\bar{c}=(h, s, b)=f(\lambda, x, y, z, t, \ldots, a, v) \qquad (1)$$

No color vision model yet proposed incorporates this level of generality.

3 Color science

Producing satisfactory color reproductions does not require the comprehensive color vision model outlined in the previous section. A model adequate for color reproduction takes as input a subset of the stimuli from the environment and produces as output enough information about the color sensation produced in the observer's mind to permit accurate color reproduction. Due to the nature of the color reproduction process, the only variable considered as input to the model is the light reaching the observer from any point in the environment. Other characteristics of the stimuli, which in the general case affect color perception, are considered to remain constant. As output, the model does not have to provide a hue, saturation, and brightness specification of the color sensation produced in the observer's mind, but need only tell whether this color sensation matches the sensation produced by another stimulus. A match implies that the observer will report an identical hue, saturation and brightness specification for both stimuli, but the model need be no more specific than to predict the match.

The assumption that only the light varies while everything else remains constant warrants further discussion. Modern color reproduction can be considered to discretize a scene into a dot structure – albeit at a finer resolution than that used by pointillist painters or needlepoint artists. The significance of this is that the color reproduction problem now centers on reproducing each dot of the image independent of the rest of the scene. This means, for example, that the analysis technique need not consider the visual angle subtended by the stimulus because the size of the dot is fixed. Although there is a time varying component to television imagery, the rate of change is above the critical flicker frequency, and the analysis technique (at least for static scenes) can also consider the intensity of the dot to be constant when television or photographs are the final reproduction medium. The analysis technique can also assume that the ambient lighting under which the entire image and hence the dot are viewed is constant. One factor that the analysis technique cannot ignore is the area surrounding a dot being analyzed. If two stimuli are removed from identical backgrounds and a match is predicted when they are evaluated on the background used by the analysis technique, the match must continue to hold when the stimuli are returned to their original backgrounds even though the hue, saturation, and brightness of the stimuli are different on each background. Assuming that matches continue to hold when evaluated on different backgrounds, the only parameter left for the analysis technique to consider is the light coming from each dot in the image.

These assumptions lead to the following experimental approach for deriving the color vision model to be used in color reproduction. Two patches of light are placed side by side on a homogeneous background. The patch on the left displays light from a dot in the original image. The patch on the right displays light which is to be used to create this dot in the reproduction. An observer is allowed to adjust some as yet unspecified controls for the light on the right until the two lights are matched. The background surrounding the patches in the matching environment can be different than the background surrounding the dots in the original and the reproduction, but it is assumed that the match holds in both settings even if the hue, saturation, and brightness change. All other conditions in the environment where the matching is performed are similar to the conditions in both the original and the reproduction environment where the dots are viewed. Ultimately the observer is replaced by a model which can analyze the light from both patches and can determine whether the observer would accept the match. If they do not match, the model can predict the adjustments to the right hand light necessary to create a match.

3.1 The nature of light

Having limited the variables of the color producing stimuli to just the light reaching the observer's eye, let us pause to consider the nature of this stimuli in more detail. Electromagnetic energy is released in bundles or quanta which travel in waves of different lengths and heights but at the same velocity, about 186,000 miles per second in air. The fundamental relationship between the speed of electromagnetic energy c, its wavelength λ, and its frequency υ is:

$$c = \upsilon \lambda \tag{2}$$

The entire electromagnetic spectrum includes gamma rays, x-rays, ultraviolet rays, infrared rays, and radio waves. The electromagnetic energy incident at the observer's eye is composed of quanta traveling in waves of many different lengths. In this paper our attention is restricted to the range of electromagnetic energy to which the eyes are sensitive. This is the relatively small band between approximately 380 and 770 nanometers called the visible range. The term "light" as used in this paper refers only to this range of the electromagnetic spectrum.

An "absolute spectral energy distribution" $E_{a1}(\lambda)$ indicates how the energy of the stimulus is distributed across the component wavelengths. The composite energy, or radiant flux, for the entire stimulus is found from the relation:

$$P_{a1} = \int E_{a1}(\lambda)\, d\lambda \tag{3}$$

where λ is wavelength (nm), $E_{a1}(\lambda)$ is the radiant flux per unit area per unit solid angle per unit wavelength interval (watt $m^{-2}\, sr^{-1}\, nm^{-1}$), and P_{a1} is the radiant flux per unit area per unit solid angle (watt $m^{-2}\, sr^{-1}$). If the intensity of the light source is modified by applying a constant K_{a2} to the original absolute spectral energy distribution $E_{a1}(\lambda)$ a new absolute spectral energy distribution $E_{a2}(\lambda)$ is created which has radiant flux:

$$P_{a2} = \int E_{a2}(\lambda)\, d\lambda = K_{a2} \int E_{a1}(\lambda)\, d\lambda = K_{a2}\, P_{a1} \tag{4}$$

In this paper this is referred to as modulating the intensity of a light source. The effect is to scale the absolute spectral energy distribution. The preceeding paragraph suggests that a unitless relative spectral energy distribution $E(\lambda)$ be established for the light source which has the property:

$$\int E(\lambda)\, d\lambda = 1 \tag{5}$$

and a constant K with units watt $m^{-2}\, sr^{-1}$ be used as a scaling factor to create absolute spectral energy distributions $KE(\lambda)$. The constant K can be thought of as a rheostat controlling the current to a light fixture. For convenience, the units for K will be specified in watts alone in this article.

Another concept which needs to be defined is the energy distribution of a spectral or monochromatic light source. This is a light source with spectral energy distribution 1 nm in width centered at a particular wavelength. The relative spectral energy distribution for a spectral light source 1 nm in width has height 1.0 from Eq. 5.

3.2 The trichromatic generalization

Recall the color matching experiment outlined above where an observer attempts to match a patch of light in the left half of the field of view by adjusting another patch of light in the right half of the field of view. To this point in the paper, the controls for the adjustable right hand patch have not been specified. Clearly, if the light in the right hand patch is the composite of the light from 390 spectral light sources (each with spectral energy distribution 1 nm in width) which includes all of the visible wavelengths of electromagnetic energy between 380 and 770 nanometers, it should be possible to approximate any spectral energy distribution presented in the left patch. This could be accomplished by using an optical measuring instrument to adjust the individual spectral light sources until an approximation to the continous spectral energy distribution is created in the right patch. This experiment configuration is shown in Fig. 2. However, a somewhat surprising result emerges when the visual system of a human observer is used to determine the adjustments of the spectral light sources necessary to obtain a match. Although the observer does create some spectral energy distributions on the right which are very close approximations to the spectral energy distributions on the left, in the majority of the cases, when the observer feels that the hue, saturation, and brightness of the patches match, the spectral energy distributions for the two patches are completely different. In fact, most test patches can be matched by a limitless number of different spectral energy distributions. Spectral energy distributions that differ in their composition, but which are still considered a match by a human observer, are called "metamers."

Because several different settings of the 390 spectral lights can match a test patch, it appears that the human visual system does not have individual receptors for each wavelength of light. A logical next question is: How few spectral lights are neces-

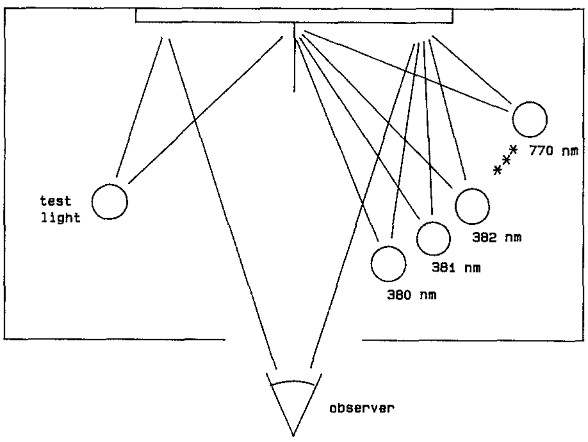

Fig. 2. Using spectral light sources to approximate a continuous spectral energy distribution

sary to adequately stimulate whatever receptor system exists and produce matches for a majority of the test patches? If only a single spectral matching light is used, it is found that the test patch on the left can only be completely matched when it consists of monochromatic light of the same wavelength as the matching light. However, if the observer can discount the difference in hue and saturation between the two patches, the brightness of the two patches can be matched. This limited form of color matching is known as photometry, and it leads to gray scale image reproduction instead of full color reproduction. The use of two spectral lights allows the observer to match more of the test patches than could be matched with a single spectral light, but no matter which two spectral lights are chosen, the majority of the test patches cannot be matched. The introduction of a third spectral matching light leads to a dramatic increase in the number of test patches that can be matched. When the three lights are chosen so that one comes from the extremely low wavelengths of the visible spectrum, one comes from the middle wavelengths, and the third comes from the high wavelengths, almost all of the test patches can be matched. The only restriction on the selection of the three wavelengths is that it must not be possible to match one of the spectral matching lights by the use of the other two alone. This would mean that any possible contribution of this third light could be accomplished by the other two lights alone. This vast reduction in the number of spectral lights necessary to create a match is what makes color reproduction practical. The study of color matching using three lights is known as colorimetry.

The goal is now to replace the human observer in the color matching experiment with an analysis technique which takes the spectral energy distribution of the left patch as input and which predicts the settings necessary to create a complete match when three spectral lights are used on the right side. As noted earlier, the continuous spectral energy distribution in the left hand test patch can be approximated by the light from 390 spectral light sources were the energy of each light source has been set by an optical measuring instrument to equal the energy present at that wavelength in the continuous distribution. The light in the test patch can then be matched one wavelength at a time by the human observer using three spectral lights (or one spectral light in the case of brightness matching). The continuous spectral energy distribution is then matched by the composite light from the trios of spectral lights used to match each wavelength. If the same spectral lights are used for each trio, then the total contribution from each light source can be summed and all of the individual trios replaced by one set of three spectral lights. Further generality can be obtained by performing the matching experiment only once with a set of "equal energy" spectral light sources. The results for any spectral energy distribution can then be found by appropriately scaling the results obtained with the "equal energy" spectral light sources. This is the approach that is taken in the following section.

3.3 Colorimetry

The use of three spectral light sources to match the test light in the matching experiment leads to the study of colorimetry. The experimental setup employed is shown in Fig. 3. The test stimuli to be matched is presented in the left patch and the three spectral light beams used to match it are shown coincident with one another on the right patch. The three spectral lights selected to do the matching in this experiment are in the low, medium, and high wavelength ranges of the visible spectrum. Their wavelengths are 444 nm, 526 nm, and 645 nm, respectively.

Each of the spectral lights has its own rheostat to control the wattage of the light: L_1 for the low wavelength 444 nm light, M_1 for the medium wavelength 525 nm light, and H_1 for the high wavelength 645 nm light. The relative spectral energy

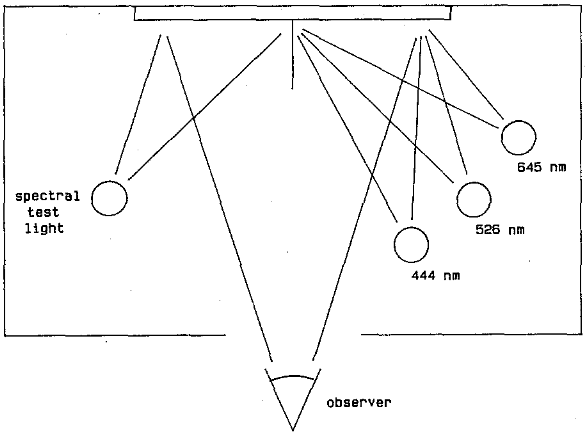

Fig. 3. Experimental setup used to perform color matching experiment

distributions (which are spikes as shown in Fig. 4) are given by $E_{L1}(\lambda)$, $E_{M1}(\lambda)$, and $E_{H1}(\lambda)$. The absolute spectral energy distributions for the adjustable light sources created by the use of the rheostats are given by $L_1\ E_{L1}(\lambda)$, $M_1\ E_{M1}(\lambda)$, and $H_1\ E_{H1}(\lambda)$. In the sequence of experiments that are performed, the visible spectrum is analyzed one wavelength at a time, in anticipation that continuous spectral energy distributions can then be approximated by breaking them into their component wavelengths.

Having described the apparatus and the experimental procedure, the data gathering begins. Suppose that the test patch consists of 100 watts of light from a 476 nm spectral light source. The observer is allowed to adjust the rheostats of the matching lights to try and obtain a match. Surprisingly, there is no setting of the three lights which makes the patches look identical to the observer. Figuring that this is just one anomalous wavelength, another spectral light at 500 nm is tried. Again the observer is unable to obtain a match. In fact, it is quickly discovered that with the exception of the three wavelengths that correspond to the wavelengths of the matching lights themselves, the observer is unable to match any spectral lights.

Backtracking, the observer is shown a previously matched test light which has an absolute spectral energy distribution with a 100 watt spike at 476 nm and a 100 watt spike at 580 nm. Gradually the amount of 580 nm light is decreased to 25 watts. The amount of the three primaries necessary to match are recorded in Table 1. If the data is extrapolated to the point where the amount of 580 nm light has gone to zero and only 100 watts of 476 nm light remains, the predicted amounts of low, medium, and high wavelength light required to match is $L_1 = 55.4$ watts, $M_1 = 16$ watts, and $H_1 = -22$ watts. It is clearly impossible to have a negative amount of light. An alternative is to move the high wavelength light to the test patch side of the matching experiment and to count its contribution as a negative amount. When the experimental setup is rearranged to accomplish this and the observer adjusts the 444 nm and 526 nm lights in the matching patch and the 645 nm light in the test patch, the amount required of each primary coincides with the prediction. This approach of adding one of the matching lights to the test patch and counting it as a negative amount is used in the remainder of the experiment whenever a match can not be obtained with the usual experimental setup.

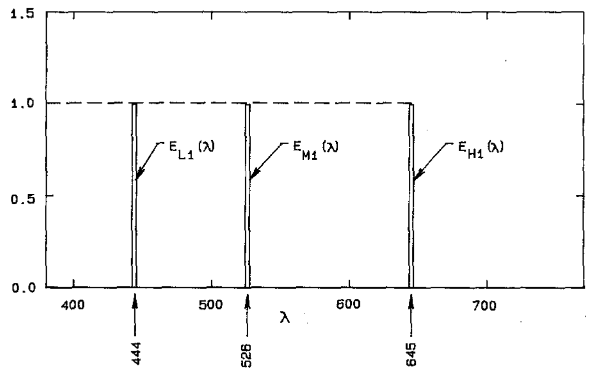

Fig. 4. Superimposed spectral energy distributions for the three spectral matching lights

Table 1. Amounts L_1, M_1, and H_1 of 444 nm, 526 nm, and 645 nm primaries necessary to match test light

Test light		Matching lights		
Watts at 476 nm	Watts at 580 nm	L_1 (watts)	M_1 (watts)	H_1 (watts)
100	100	53.4	84	210
100	75	53.9	67	152
100	50	54.4	50	94
100	25	54.9	33	36
100	0	55.4	16	−22

The data in Table 1 shows that a linear change in the test stimulus resulted in a linear change in the matching stimuli. This is a crucial result. It means that when a test stimulus is scaled, the matching stimuli are scaled by the same amount. The ratio of the rheostat settings to the power of the test light remains constant and is what will be recorded as *l*, *m*, and *h* for each wavelength. Since the ratios are constant for all possible power settings of the test light, it suffices to examine each test light only once when it is emitting 100 watts of radiant flux.

By performing the matching experiment at each wavelength of the visible spectrum and then interpolating the results the three curves in Fig. 5 are obtained. These curves are referred to as the matching functions $l_1(\lambda)$, $m_1(\lambda)$, and $h_1(\lambda)$. Note that at the 476 nm wavelength the curves take on the values $l_1(476) = 0.554$, $m_1(476) = 0.16$, and $h_1(476) = -0.22$ as previously determined. It is

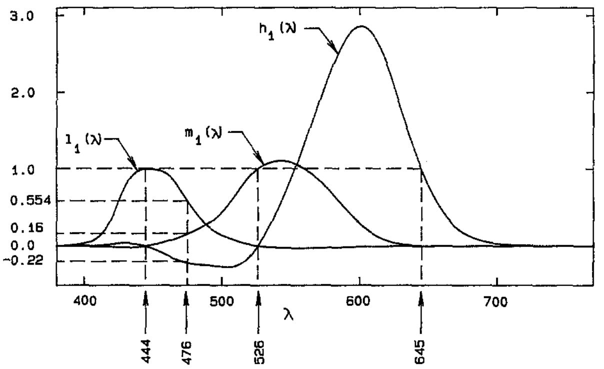

Fig. 5. Results of color matching experiment using 444 nm, 526 nm, and 645 nm lights

also interesting to examine the results at 444 nm, 526 nm, and 645 nm, the wavelengths of the matching stimuli. As expected, two of the matching stimuli are zero while the one which has the same wavelength as the test stimulus operates at a power setting identical to the setting for the test stimulus.

The amount of the 444 nm, 526 nm, and 645 nm lights necessary to match a general absolute spectral energy distribution $E_a(\lambda)$ is:

$$\begin{aligned} L_1 &= \int E_a(\lambda)\ l_1(\lambda)\, d\lambda \\ M_1 &= \int E_a(\lambda)\, m_1(\lambda)\, d\lambda \\ H_1 &= \int E_a(\lambda)\ h_1(\lambda)\, d\lambda \end{aligned} \tag{6}$$

The numbers L_1, M_1, and H_1 are referred to as the tristimulus values for the absolute spectral energy distribution $E_a(\lambda)$. Negative tristimulus values are possible in the case of the general stimulus $E_a(\lambda)$ just as they are for the spectral lights examined earlier. This is due to the negative portions of the matching functions in Fig. 5.

3.3.1 Color plotting space

The tristimulus nature of the color matching experiment suggests plotting, in an appropriate coordinata system, the values obtained using Eq. 6. To accomplish this, an orthogonal coordinate system, with axes L_1, M_1, and H_1 is established. An absolute spectral energy distribution quantified as a triplet by using Eq. 6, plots as a point in this space.

When the 476 nm spectral light source was matched in the preceeding section, the proportion of each of the three primaries in the matching mixture remained constant and was independent of the wattage of the light source. Geometrically, this means that the locus of L_1, M_1, H_1 triplets necessary to match various amounts of the 476 nm test light is a ray. This ray, with orientation determined by the ratio of the primaries, starts at the origin and extends indefinitely into the plotting space.

Now consider the absolute spectral energy distribution composed of spectral spikes at 476 nm and 580 nm. Taken as a whole, this spectral energy distribution can be matched experimentally as recorded in Table 1 or analytically by use of Eq. 6. Alternatively each of the two wavelengths can be considered separately with the amount necessary to match the entire distribution taken as the sum of the amounts necessary to match each component wavelength. This latter approach can be thought of as vector addition. The 476 nm portion of the distribution yields a vector in the direction of the ray described in the preceeding paragraph, the 580 nm portion produces a vector in some other direction, and the resultant tristimulus value for the entire distribution is found by adding these two vectors together.

Using the vector addition model it is evident that all tristimulus values which result from absolute spectral energy distributions composed of these two wavelengths must lie in the plane which contains the two vectors. Furthermore, if only positive amounts of these two spectral lights are possible the domain of possible tristimulus values is restricted to the region of the plane interior to the angle formed by the two vectors.

The addition of a third spectral spike to the 476 nm and 580 nm spikes already present in the absolute spectral energy distribution, introduces a third vector into the vector addition equation. In the general case, the tristimulus values produced by this new vector alone might lie on the plane containing the original two vectors and interior to the angle which the original two vectors form. The new vector would therefore fail to increase the region of possible tristimulus values. It is more likely, however, that the new vector is not coplanar with the first two. In this case, the region of potential tristimulus values becomes the interior of a three-sided pyramid. The sides of the pyramid are the planer angles formed by taking combinations of the three vectors two at a time. The point of the pyramid is at the origin and its sides extend indefinitely out into the plotting space.

Adding a fourth spectral spike to the three already present in the absolute spectral energy distribution will only increase the region of potential tristimulus values if the vector which represents the contribution of the new spike does not lie within the region defined by the other three spikes alone. The new region of possible tristimulus values is the union of all three-sided pyramids formed by taking combinations of the four vectors three at a time. In general, introducing the n'th new spike into the absolute spectral energy distribution increases the domain of possible tristimulus values by creating $(n-1)(n-2)/2$ new three-sided pyramids to be unioned with the domain defined by the first $n-1$ spikes.

When the absolute spectral energy distribution has a spike for each wavelength in the visible spectrum, it approximates a continuous distribution. To begin with, assume that each spike has the same wattage. The triplet L_1, M_1, H_1 necessary to match each of these spikes can be determined by simply reading the values off of Fig. 5 at the appropriate wavelength and scaling each value by a constant amount. This equal energy spectrum locus plots as the curve in Fig. 6. One can also consider this curve to be the locus of endpoints for a set of vectors representing the contribution of each spectral spike in the equal energy distribution. Furthermore, by appropriately scaling each vector (but leaving the direction of each unchanged), one can create the set of vectors representing the contribution of each spectral spike in an absolute energy distribution where the energy at each wavelength is arbitrary. The surface upon which all such vectors must lie is the union of all rays which emanate from the origin and pass through the equal energy spectrum locus. This surface is basically cone shaped with an opening on one side (see Fig. 6).

The triplet necessary to match a continuous absolute spectral energy distribution is found by vector addition, where the vector for each component wavelength lies on the cone-shaped surface of Fig. 6. It was previously shown that the region of possible tristimulus values for any collection of vectors is found by unioning the three-sided pyramids formed by taking all possible combinations of the vectors three at a time. This yields the interior of a cone with one flat side when applied to Fig. 6. For this reason, the surface in Fig. 6 and the plane which makes it an enclosed volume are referred to as the "cone of realizable color" because together they bound the region of the plotting space within which all possible tristimulus values must lie.

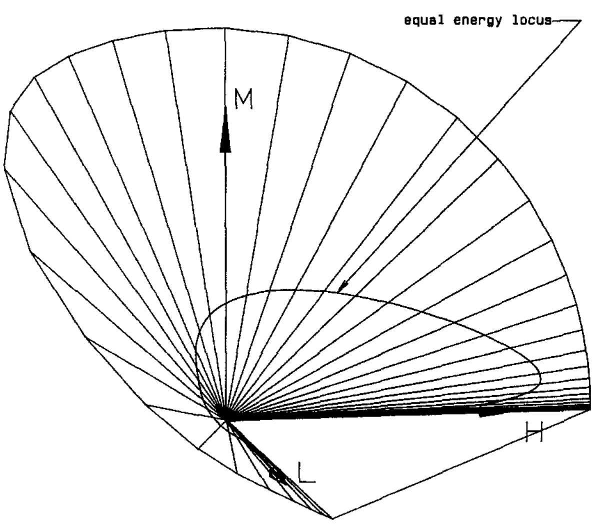

Fig. 6. Cone of realizable color

3.3.2 Transformation of primaries

Using Eq. 6 it is possible to predict the amount of 444 nm, 526 nm, and 645 nm light necessary to match an arbitrary absolute spectral energy distribution. What if three new light sources are available which emit radiation at not just one wavelength, but at all wavelengths? Can such lights even be used to match a test stimuli, and if they can, will the matching experiments have to be repeated?

Suppose the relative spectral energy distributions for the new light sources can be described by $E_{L2}(\lambda)$, $E_{M2}(\lambda)$, and $E_{H2}(\lambda)$. Rheostats with settings L_2, M_2, H_2 are provided to control the lights and they produce absolute spectral energy distributions $L_2\,E_{L2}(\lambda)$, $M_2\,E_{M2}(\lambda)$, and $H_2\,E_{H2}(\lambda)$.

It is easy to determine the amounts L_1, M_1, and H_1 of the original primaries necessary to match the absolute spectral energy distribution $L_2\,E_{L2}(\lambda)$ of one of the new primaries. Eq. 6 yields the relationships:

$$\begin{aligned} L_1 &= \int L_2\,E_{L2}(\lambda)\;\; l_1(\lambda)\,d\lambda \\ &= L_2 \int E_{L2}(\lambda)\;\; l_1(\lambda)\,d\lambda = L_2\,\alpha \\ M_1 &= \int L_2\,E_{L2}(\lambda)\,m_1(\lambda)\,d\lambda \\ &= L_2 \int E_{L2}(\lambda)\,m_1(\lambda)\,d\lambda = L_2\,\beta \\ H_1 &= \int L_2\,E_{L2}(\lambda)\;\; h_1(\lambda)\,d\lambda \\ &= L_2 \int E_{L2}(\lambda)\;\; h_1(\lambda)\,d\lambda = L_2\,\gamma \end{aligned} \tag{7}$$

where the symbols α, β, and γ are introduced to represent the values of the constant integrals. A similar analysis can be used to determine the amount of the original primaries necessary to match the other two new primaries. This leads to the relationship:

$$\begin{bmatrix} L_1 \\ M_1 \\ H_1 \end{bmatrix} = \begin{bmatrix} \alpha & \delta & \theta \\ \beta & \varepsilon & \rho \\ \gamma & \zeta & \sigma \end{bmatrix} \begin{bmatrix} L_2 \\ M_2 \\ H_2 \end{bmatrix} = \begin{bmatrix} \\ T \\ \\ \end{bmatrix} \begin{bmatrix} L_2 \\ M_2 \\ H_2 \end{bmatrix} \tag{8}$$

The new primaries can now be used to match the arbitrary absolute spectral energy distribution $E_a(\lambda)$ that was matched using the original primaries. One can solve for L_2, M_2, and H_2 by inverting the matrix in Eq. 8:

$$\begin{bmatrix} L_2 \\ M_2 \\ H_2 \end{bmatrix} = \begin{bmatrix} \\ T^{-1} \\ \\ \end{bmatrix} \begin{bmatrix} L_1 \\ M_1 \\ H_1 \end{bmatrix} \tag{9}$$

A new plotting space for the tristimulus values is a by-product of this transformation. The orthogonal axes labelled L_2, M_2, H_2 in this space are the three rays which represented the new primaries in the original plotting space.

A new set of matching functions $l_2(\lambda)$, $m_2(\lambda)$, and $h_2(\lambda)$ can be ontained from

$$\begin{bmatrix} l_2(\lambda) \\ m_2(\lambda) \\ h_2(\lambda) \end{bmatrix} = \begin{bmatrix} \\ T^{-1} \\ \\ \end{bmatrix} \begin{bmatrix} l_1(\lambda) \\ m_1(\lambda) \\ h_1(\lambda) \end{bmatrix} \tag{10}$$

This means that given an arbitrary light source, the amount L_2, M_2, H_2 necessary to match the distribution can be determined directly from the absolute spectral energy distribution $E_a(\lambda)$. Only one matching experiment need ever be performed, the results being transformed for use with whatever set of primaries is convenient. Furthermore, these new primaries need not be spectral light sources.

The transformation matrix T^{-1} was derived using three light sources for which the relative spectral energy distributions were known. One could also select the elements of the matrix T^{-1} directly (as long as a singular matrix doesn't result), without being concerned whether they correspond to actual light sources. In this way, tristimulus values L_2, M_2, H_2 and matching functions $l_2(\lambda)$, $m_2(\lambda)$, $h_2(\lambda)$ which meet certain constraints can be generated. If at some point in the process it is necessary to reproduce the color, the tristimulus values can be expressed in terms of primaries for which actual light sources exist.

3.4 CIE XYZ System

There can be other objectives when selecting the elements of the transformation matrix in Eq. 9 than the original objective of expressing the tristimulus values in terms of a new set of real primaries. For example, since the observer matches the hue, saturation, *and* brightness of the two patches in the matching experiment, the results of a brightness matching experiment should be implicit in the color matching functions of Fig. 5. Thus it makes sense for one of the color matching functions in the new coordinate system to correspond to the brightness matching function, more commonly known as the luminous efficiency function. Secondly, the negative tristimulus values which occurred during the matching experiment are a mathematical nuisance. Eliminating the negative lobes from the matching functions in Fig. 5 would solve this problem but will make it impossible for the new set of primaries to correspond to real light sources. Finally, the most commonly occurring spectral energy distributions are relatively flat across the visible spectrum without extreme variations. It therefore makes sense for a spectral energy distribution which is constant across the spectrum (an equal energy spectral energy distribution) to yield tristimulus values which are all of the same order of magnitude. This can be accomplished by making the area under each color matching function identical so that the tristimulus values for the equal energy spectral energy distribution will all be the same.

In 1931 the CIE standardized a transformation which meets the above constraints. Due to technological limitations present at that time, the transformation was applied to color matching functions for which the primaries were measured by brightness matching (Guild 1931; Wright 1928) instead of by energy measurement (Stiles and Burch 1955) as was done in the above derivation. The "standard observer" matching functions $\bar{x}(\lambda)$, $\bar{y}(\lambda)$, and $\bar{z}(\lambda)$ which resulted are shown in Fig. 7. Note that the $\bar{y}(\lambda)$ matching function is identical to the luminous efficiency function, that the curves have no negative portions, and that the area under each curve is identical. The tristimulus values X, Y, and Z are found from the relations:

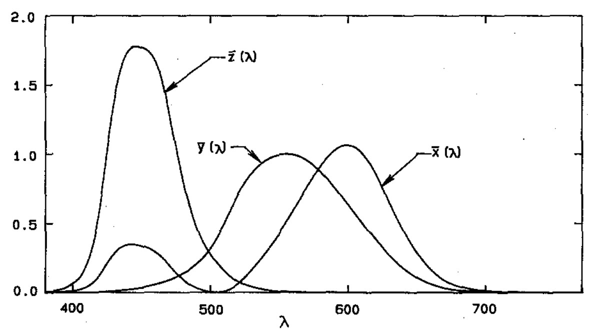

Fig. 7. CIE "standard observer" matching functions

$$
\begin{aligned}
X &= C \int E(\lambda)\, \bar{x}(\lambda)\, d\lambda \\
Y &= C \int E(\lambda)\, \bar{y}(\lambda)\, d\lambda \qquad (11) \\
Z &= C \int E(\lambda)\, \bar{z}(\lambda)\, d\lambda
\end{aligned}
$$

where $E(\lambda)$ is a spectral energy distribution and C is a constant used to normalize the results. The position of the cone of realizable color in the positive octant of the resulting coordinate system is shown in Fig. 8.

Up to this point only real light sources have been considered and results from equations similar to Eq. 11 have been expressed in terms of watts (c.f. Eq. 6). Now that "fictitous" primaries are being used the units are no longer fixed. If $E(\lambda)$ is an absolute spectral energy distribution, then C can be selected to equal K_m (680 lumen/watt). Using this factor, Y gives the luminous flux of the stimulus. If $E(\lambda)$ is a relative spectral energy distribution then it is only the ratio between the tristimulus values which is important and their absolute magnitude can be ignored. In this case C is usually chosen such that Y has the value 100.0.

3.4.1 Chromaticity coordinates

In the above discussion of the color matching experiments it has been seen that adjusting the rheostat of a light source produces a locus of tristimulus values which forms a ray emanating from the origin out into the plotting space. The direction of this ray is clearly an important colorimetric property of the light source. Chromaticity coordinates are numbers which describe the direction of such a ray. Although chromaticity coordinates are applicable to any tristimulus system, the following development uses the CIE XYZ system.

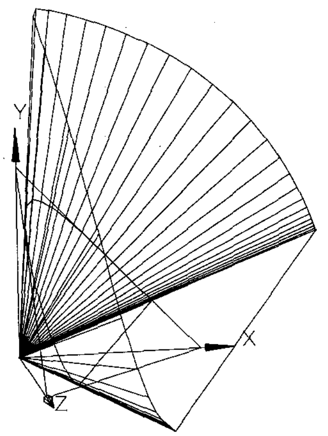

Fig. 8. Cone of realizable color and unit plane in CIE XYZ space

Because the locus of tristimulus values produced by a realizable light source must lie within the cone of realizable color, the ray produced by the locus must intersect the unit plane $X+Y+Z=1$. This point of interscetion can therefore be used to describe the direction of the ray. Figure 8 shows the unit plane and the cone of realizable color. The horseshoe shaped line of intersection between the plane and the cone bounds all possible points at which a ray representing a realizable light source might intersect the unit plane. The linear portion of this locus, which results from the flat side of the cone, is referred to as the "purple line." These points of intersection with the unit plane are referenced by taking an orthographic projection of the unit plane onto the XY plane as shown in Fig. 9. This is called a chromaticity diagram and chromaticity coordinates x and y are used to locate points on this two dimensional figure.

There is a simple analytical formulation which accomplishes the operations of intersection and projection described above. The coordinates of the point of intersection between the unit plane and a ray through the color with tristimulus values X, Y, and Z are found from the relations:

$$x=\frac{X}{X+Y+Z} \quad y=\frac{Y}{X+Y+Z} \quad z=\frac{Z}{X+Y+Z} \qquad (12)$$

These are referred to as the chromaticity coordinates of the color. The projection of the unit plane onto the XY plane can be accomplished by considering only the x and y members of the trio since $z=1-x-y$.

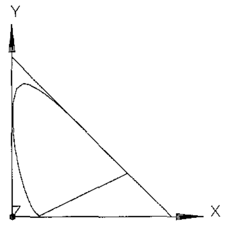

Fig. 9

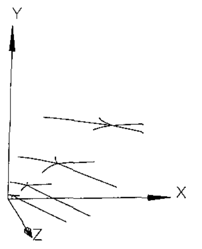

Fig. 10

Fig. 9. Orthographic projection of the unit plane onto the XY plane, commonly known as a chromaticity diagram

Fig. 10. Tristimulus values for hue, saturation, and brightness color samples in CIE XYZ space

The chromaticity coordinates x and y together with one of the original tristimulus values, form a new way of specifying a color. Although any one of the three original tristimulus values could be used to pick off a point along the ray described by the chromaticity coordinates, typically the Y value is selected because it contains the luminance information. A chromaticity and luminance color specification (x, y, Y) can be transformed into tristimulus values X, Y, and Z through the relations:

$$X = x\frac{Y}{y} \qquad Y = Y \qquad Z = (1-x-y)\frac{Y}{y} \tag{13}$$

3.4.2 Hue, saturation, and brightness in CIE XYZ space

A set of color swatches from a paint store can be organized using the hue, saturation, and brightness vocabulary developed in an earlier section of the paper. This collection can subsequently be resolved to tristimulus values by using Eq. 11, and can finally be plotted in CIE XYZ space. Each resulting "spider" in Fig. 10 has constant brightness, and each spider arm has constant hue and brightness. Saturation varies from zero at the center of each spider to a maximum at the end of each arm.

There are certain correspondences that can be established between the hue, saturation, and brightness of these color swatches and the position of their tristimulus values in the plotting space:

1. As predicted by the photometry experiments, each constant brightness spider has a constant Y value and therefore a constant luminance. However the spiders, which differ from one another by the same amount of brightness, are not equally spaced in CIE XYZ space. This means that although luminance can predict a brightness match it is not a direct measure of brightness magnitude.
2. The achromatic axis of the hue, saturation, and brightness system runs through the centers of the spiders and hence through the center of the cone of realizable color. The chromaticity coordinates for these neutral colors are close to the center of the chromaticity diagram where $x = 1/3$ and $y = 1/3$.
3. There is a correspondence between change in wavelength and change in hue. Comparing Figs. 8 and 10 one sees that constant hue spider arms intersect the cone of realizable color at distinct wavelengths. Sections of the chromaticity diagram can therefore be labelled by hue.
4. Saturation increases from a minimum at the center of the chromaticity diagram to a maximum at the spectrum locus.

It must be emphasized that the above correspondence between CIE XYZ tristimulus values and the hue, saturation, and brightness notation system is not fixed. If, for example, the size or the background of the swatches is altered, the hue, saturation, and brightness specifications will change but the CIE XYZ tristimulus values will remain constant. The use of hue, saturation, and brightness to refer to locations in CIE XYZ space is therefore not recommended. However, since this practice is so common in the literature, it will be used occasionally in the remainder of this article.

3.4.3 Practical color space transformations

In a preceeding section it was shown that colors expressed in terms of one set of primaries can be expressed in terms of a new set of primaries by means of a linear transformation. The situation was summarized in Eq. 9. The CIE XYZ system was introduced as a standard set of primaries in which to express colors. In this section the practical problem of taking the available information about a new set of primaries and constructing a linear transformation from these primaries to CIE XYZ space is addressed. The notation R, G, B is frequently used to refer to the amounts of the new primaries since the widest gamut of matchable colors occurs when the low, medium, and high wavelength primaries have chromaticities which place them in the blue, green, and red regions respectively of the chromaticity diagram. In the following derivation it is assumed that the units of the RGB primaries have been selected so that $R=1$, $G=1$, $B=1$, yields a chromaticity of approximately $x=1/3$, $y=1/3$ and hence a "white" color sensation when viewed by the observer. The derivation is equally valid when the RGB primaries are measured in watts or lumens, although the chromaticity of the color produced by $R=1$ watt, $G=1$ watt, and $B=1$ watt, for example, may no longer lie in the white section of the chromaticity diagram.

The elements α, β, γ, ..., σ of the matrix in Eq. 8 were determined by knowing how much of one set of primaries was necessary to match unit amounts of the other set of primaries. Although this information is not generally available, assume that for unit amounts of the RGB primaries the amount of the CIE XYZ primaries necessary to create the match is known. For example when $R=1$, $G=0$, and $B=0$ then $X=X_r$, $Y=Y_r$, and $Z=Z_r$. The entire expression is:

$$\begin{bmatrix} X \\ Y \\ Z \end{bmatrix} = \begin{bmatrix} X_r & X_g & X_b \\ Y_r & Y_g & Y_b \\ Z_r & Z_g & Z_b \end{bmatrix} \begin{bmatrix} R \\ G \\ B \end{bmatrix} \tag{14}$$

If the chromaticity coordinates (x_r, y_r), (x_g, y_g), and (x_b, y_b) of the RGB primaries in the CIE XYZ system are known, then for the R primary:

$$\begin{aligned} x_r &= \frac{X_r}{X_r+Y_r+Z_r} \\ y_r &= \frac{Y_r}{X_r+Y_r+Z_r} \\ (1-x_r-y_r) &= \frac{Z_r}{X_r+Y_r+Z_r} \end{aligned} \tag{15}$$

Using similar expressions for the other two primaries and introducing the constants $C_r=X_r+Y_r+Z_r$, $C_g=X_g+Y_g+Z_g$, $C_b=X_b+Y_b+Z_b$ the transformation of Eq. 14 can be expressed as:

$$\begin{bmatrix} X \\ Y \\ Z \end{bmatrix} = \begin{bmatrix} x_r C_r & x_g C_g & x_b C_b \\ y_r C_r & y_r C_g & y_b C_b \\ (1-x_r-y_r) C_r & (1-x_g-y_g) C_g & (1-x_b-y_b) C_b \end{bmatrix} \begin{bmatrix} R \\ G \\ B \end{bmatrix} \tag{16}$$

where additional information is necessary to determine C_r, C_g, and C_b.

In a practical situation the information frequently available in addition to the chromaticity coordinates may be the luminance Y_r, Y_g, and Y_b of unit amounts of the RGB primaries. Then from Eq. 16 the constants Y_r, Y_g, and Y_b can be determined as:

$$C_r = \frac{Y_r}{y_r} \quad C_g = \frac{Y_g}{y_g} \quad C_b = \frac{Y_b}{y_b} \tag{17}$$

and the transformation is fully specified. Alternatively, the additional piece of information may be the tristimulus values X_w, Y_w, and Z_w for the "white" color sensation produced when $R=1$, $G=1$, and $B=1$. Then Eq. 16 can be re-expressed as:

$$\begin{bmatrix} X_w \\ Y_w \\ Z_w \end{bmatrix} = \begin{bmatrix} x_r & x_g & x_b \\ y_r & y_g & y_b \\ (1-x_r-y_r) & (1-x_g-y_g) & (1-x_b-y_b) \end{bmatrix} \begin{bmatrix} C_r \\ C_g \\ C_b \end{bmatrix} \tag{18}$$

from which the constants C_r, C_g, and C_b can be determined. When the "white" color is specified by chromaticity coordinates x_w, y_w and luminance Y_w instead of tristimulus values X_w, Y_w, and Z_w, the result of solving Eq. 18 is:

$$\begin{aligned} C_r &= \frac{Y_w}{y_w} \left\{ \frac{x_w(y_g-y_b) - y_w(x_g-x_b) + x_g y_b - x_b y_g}{x_r(y_g-y_b) - x_g(y_b-y_r) + x_b(y_r-y_g)} \right\} \\ C_g &= \frac{Y_w}{y_w} \left\{ \frac{x_w(y_b-y_r) - y_w(x_b-x_r) + x_r y_b - x_b y_r}{x_r(y_g-y_b) - x_g(y_b-y_r) + x_b(y_r-y_g)} \right\} \\ C_b &= \frac{Y_w}{y_w} \left\{ \frac{x_w(y_r-y_g) - y_w(x_r-x_g) + x_r y_g - x_g y_r}{x_r(y_g-y_b) - x_g(y_b-y_r) + x_b(y_r-y_g)} \right\} \end{aligned} \tag{19}$$

and the transformation in Eq. 16 is again fully specified. The inverse transformations from CIE XYZ space to RGB space can be found by inverting the matrix in Eq. 16.

3.5 Color science bibliography

This tutorial on color science is a distillation of material from many sources. In order to set the major concepts into a framework which ties them tightly together, many of the topics covered in these references have been omitted from this tutorial. Among things not discussed are the 10 degree of field standard observer, the scotopic luminous efficiency curve, the CIE standard sources, and uniform chromaticity diagrams. This section lists a number of books and articles used in the preparation of this tutorial where the omitted material may be found.

Other tutorial presentations of color science can be found in Billmeyer and Saltzman (1966) and Wentworth (1955). Cornsweet (1970) contains an excellent discussion of "wavelength mixture space." A tightly reasoned summary of tristimulus colorimetry is given in outline form in Burnham et al. (1963).

Wyszecki and Stiles (1982) is the fundamental reference book in color science. All of the data from CIE (1971) is included along with information from other related areas of research. Judd and Wyszecki (1975) is also a good source along with OSA (1953). A fascinating collection of some of the earliest attempts to quantify color is found in MacAdam (1970).

Because color science was so important to the development of color television, several books and articles surveying the subject were written as this new technology emerged. Wentworth (1955) contains an excellent discussion of colorimetry as applied to color television. Other references appearing at that time include Fink (1957) and Wintringham (1951). More recent publications include Neal (1973) and Pearson (1975).

The above list is by no means comprehensive. It only represents those references which this author has found particularly useful.

4 Summary

An understanding of the principles of color science is important for solving color reproduction and image synthesis problems in computer graphics. Color perception in all its generality is a complex phenomena which is not completely understood. However, when the objective is simply to predict a color match and the environmental circumstances under which the match is evaluated are tightly controlled, the theory becomes much simpler. A trichromatic model of human color vision which is linear over the practical range of operation is found to be adequate. This leads to a scheme which quantifies a color using tristimulus values expressed in terms of three independent primaries. Standards such as the CIE XYZ color specification system promote the use of a specific set of primaries. Image synthesis techniques used in computer graphics must apply the laws of color science to convert spectral energy distributions to tristimulus values suitable for driving a color reproduction device.

Acknowledgement. This work was performed under partial funding by National Science Foundation grant DCR-8203979, "Interactive Computer Graphics Input and Display Techniques."

References

Billmeyer FW, Saltzman M (1966) Principles of color technology. Interscience, New York

Burnham RW, Hanes RM, Bartleson CJ (1963) Color: A guide to basic facts and concepts. John Wiley and Sons, New York

CIE (1971) Colorimetry, Publication No. 15, Commission Internationale de l'Eclairage, Paris

Cornsweet TN (1970) Visual perception. Academic Press, New York

Fink DG (ed) (1957) Television engineering handbook. McGraw-Hill, New York

Foley JD, Van Dam A (1982) Fundamentals of interactive computer graphics. Addison-Wesley, Reading

Guild J (1931) The colorimetric properties of the spectrum. Phil Trans Roy Soc London A230:149–187

Hunt RWG (1977) The specification of colour appearance. I. concepts and terms. Color Res Appl 2:55–68

Hunt RWG (1978) Colour terminology. Color Res Appl 3:79–87

Judd DB, Wyszecki G (1975) Color in business, science and industry. John Wiley and Sons, New York

MacAdam DL (ed) (1970) Sources of color science. The MIT Press, Cambridge

Neal CB (1973) Television colorimetry for receiver engineers. IEEE Transactions on Broadcast and Television Receivers 19:149–162

Newman WM, Sproull RF (1979) Principles of interactive computer graphics. McGraw-Hill, New York

OSA Committee on Colorimetry (1953) The science of color. Thomas Y. Crowell Co., New York

Pearson DE (1975) Transmission and display of pictoral information. John Wiley and Sons, New York

Stiles WS, Burch JM (1955) Interim report to the Commission Internationale de l'Eclairage, Zurich, 1955, on the National Physical Laboratory's investigation of colour-matching. Optica Acta 2:168–181

Wentworth JW (1955) Color television engineering. McGraw-Hill, New York

Wintringham WT (1951) Color television and colorimetry. Proc IRE 39:1135–1172

Wright WD (1928) A re-determination of the trichromatic coefficients of the spectral colors. Trans Opt Soc London 30:141

Wyszecki G, Stiles WS (1982) Color science (2th ed.). John Wiley and Sons, New York