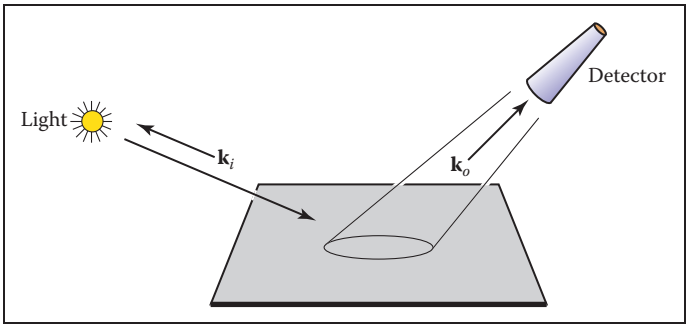


Figure 18.5. A simple measurement device for directional reflectance. The positions of light and detector are moved to each possible pair of directions. Note that both $\mathbf{k}_i$ and $\mathbf{k}_o$ point away from the surface to allow reciprocity.

illuminated to measure the light. However, for this to get an accurate reading that would not depend on the $\Delta\sigma$ of the detector, we would need the light to subtend a solid angle bigger than $\Delta\sigma$. Unfortunately, the measurement taken by our roving radiance detector in direction $\mathbf{k}_o$ will also count light that comes from points outside the new detector's cone. So this does not seem like a practical solution.

Alternatively, we can place an irradiance meter at the point on the surface being measured. This will take a reading that does not depend strongly on subtleties of the light source geometry. This suggests characterizing reflectance as a ratio:

$$\rho = \frac{L_s}{H},$$

where this fraction ρ will vary with incident and exitant directions $\mathbf{k}_i$ and $\mathbf{k}_o$, H is the irradiance for light position $\mathbf{k}_i$, and L_s is the surface radiance measured in direction $\mathbf{k}_o$. If we take such a measurement for all direction pairs, we end up with a 4D function $\rho(\mathbf{k}_i, \mathbf{k}_o)$. This function is called the *bidirectional reflectance distribution function* (BRDF). The BRDF is all we need to know to characterize the directional properties of how a surface reflects light.

Directional Hemispherical Reflectance

Given a BRDF, it is straightforward to ask, "What fraction of incident light is reflected?" However, the answer is not so easy; the fraction reflected depends on the directional distribution of incoming light. For this reason, we typically only set a fraction reflected for a fixed incident direction $\mathbf{k}_i$. This fraction is called the *directional hemispherical reflectance*. This fraction, $R(\mathbf{k}_i)$ is defined by

$$R(\mathbf{k}_i) = \frac{\text{power in all outgoing directions } \mathbf{k}_o}{\text{power in a beam from direction } \mathbf{k}_i}.$$

Note that this quantity is between zero and one for reasons of energy conservation. If we allow the incident power Φ_i to hit on a small area ΔA, then the irradiance is $\Phi_i/\Delta A$. Also, the ratio of the incoming power is just the ratio of the radiant exitance to irradiance:

$$R(\mathbf{k}_i) = \frac{E}{H}.$$

The radiance in a particular direction resulting from this power is by the definition of BRDF:

$$\begin{aligned} L(\mathbf{k}_o) &= H\rho(\mathbf{k}_i, \mathbf{k}_o) \\ &= \frac{\Phi_i}{\Delta A}. \end{aligned}$$

And from the definition of radiance, we also have

$$L(\mathbf{k}_o) = \frac{\Delta E}{\Delta\sigma_o \cos\theta_o},$$

where E is the radiant exitance of the small patch in direction $\mathbf{k}_o$. Using these two definitions for radiance we get

$$H\rho(\mathbf{k}_i, \mathbf{k}_o) = \frac{\Delta E}{\Delta\sigma_o \cos\theta_o}.$$

Rearranging terms, we get

$$\frac{\Delta E}{H} = \rho(\mathbf{k}_i, \mathbf{k}_o)\Delta\sigma_o \cos\theta_o.$$

This is just the small contribution to E/H that is reflected near the particular $\mathbf{k}_o$. To find the total $R(\mathbf{k}_i)$, we sum over all outgoing $\mathbf{k}_o$. In integral form this is

$$R(\mathbf{k}_i) = \int_{\text{all } \mathbf{k}_o} \rho(\mathbf{k}_i, \mathbf{k}_o) \cos\theta_o \, d\sigma_o.$$

Ideal Diffuse BRDF

An idealized diffuse surface is called *Lambertian*. Such surfaces are impossible in nature for thermodynamic reasons, but mathematically they do conserve energy. The Lambertian BRDF has ρ equal to a constant for all angles. This means the surface will have the same radiance for all viewing angles, and this radiance will be proportional to the irradiance.

If we compute $R(\mathbf{k}_i)$ for a a Lambertian surface with $\rho = C$ we get

$$\begin{aligned} R(\mathbf{k}_i) &= \int_{\text{all } \mathbf{k}_o} C \cos\theta_o \, d\sigma_o \\ &= \int_{\phi_o=0}^{2\pi} \int_{\theta_o=0}^{\pi/2} C \cos\theta_o \sin\theta_o \, d\theta_o \, d\phi_o \\ &= \pi C. \end{aligned}$$

Thus, for a perfectly reflecting Lambertian surface ($R = 1$), we have $\rho = 1/\pi$, and for a Lambertian surface where $R(\mathbf{k}_i) = r$, we have

$$\rho(\mathbf{k}_i, \mathbf{k}_o) = \frac{r}{\pi}.$$

This is another example where the use of a steradian for the solid angle determines the normalizing constant and thus introduces factors of π.

18.2 Transport Equation

With the definition of BRDF, we can describe the radiance of a surface in terms of the incoming radiance from all different directions. Because in computer graphics we can use idealized mathematics that might be impractical to instantiate in the lab, we can also write the BRDF in terms of radiance only. If we take a small part of the light with solid angle $\Delta\sigma_i$ with radiance L_i and "measure" the reflected radiance in direction $\mathbf{k}_o$ due to this small piece of the light, we can compute a BRDF (Figure 18.6). The irradiance due to the small piece of light is $H = L_i \cos\theta_i \Delta\sigma_i$. Thus the BRDF is

$$\rho = \frac{L_o}{L_i \cos\theta_i \Delta\sigma_i}.$$

This form can be useful in some situations. Rearranging terms, we can write down the part of the radiance that is due to light coming from direction $\mathbf{k}_i$:

$$\Delta L_o = \rho(\mathbf{k}_i, \mathbf{k}_o) L_i \cos\theta_i \Delta\sigma_i.$$

If there is light coming from many directions $L_i(\mathbf{k}_i)$, we can sum all of them. In integral form, with notation for surface and field radiance, this is

$$L_s(\mathbf{k}_o) = \int_{\text{all } \mathbf{k}_i} \rho(\mathbf{k}_i, \mathbf{k}_o) L_f(\mathbf{k}_i) \cos\theta_i d\sigma_i.$$

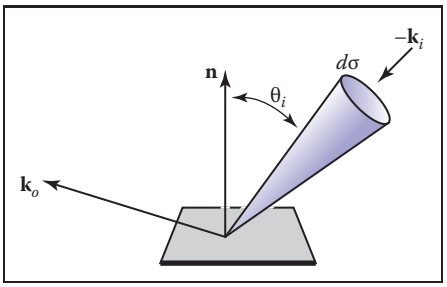


Figure 18.6. The geometry for the transport equation in its directional form.

This is often called the *rendering equation* in computer graphics (Immel, Cohen, & Greenberg, 1986).

Sometimes it is useful to write the transport equation in terms of surface radiances only (Kajiya, 1986). Note, that in a closed environment, the field radiance $L_f(\mathbf{k}_i)$ comes from some surface with surface radiance $L_s(-\mathbf{k}_i) = L_f(\mathbf{k}_i)$ (Figure 18.7). The solid angle subtended by the point $\mathbf{x}'$ in the figure is given by

$$\Delta\sigma_i = \frac{\Delta A' \cos\theta'}{\|\mathbf{x} - \mathbf{x}'\|^2},$$

where $\Delta A'$ the the area we associate with $\mathbf{x}'$. Substituting for $\Delta\sigma_i$ in terms of $\Delta A'$ suggests the following transport equation:

$$L_s(\mathbf{x}, \mathbf{k}_o) = \int_{\text{all } \mathbf{x}\text{' visible to } \mathbf{x}} \frac{\rho(\mathbf{k}_i, \mathbf{k}_o) L_s(\mathbf{x}', \mathbf{x} - \mathbf{x}') \cos\theta_i \cos\theta'}{\|\mathbf{x} - \mathbf{x}'\|^2} dA' .$$

Note that we are using a non-normalized vector $\mathbf{x} - \mathbf{x}'$ to indicate the direction from $\mathbf{x}'$ to $\mathbf{x}$. Also note that we are writing L_s as a function of position and direction.

The only problem with this new transport equation is that the domain of integration is awkward. If we introduce a visibility function, we can trade off complexity in the domain with complexity in the integrand:

$$L_s(\mathbf{x}, \mathbf{k}_o) = \int_{\text{all } \mathbf{x}'} \frac{\rho(\mathbf{k}_i, \mathbf{k}_o) L_s(\mathbf{x}', \mathbf{x} - \mathbf{x}') v(\mathbf{x}, \mathbf{x}') \cos\theta_i \cos\theta'}{\|\mathbf{x} - \mathbf{x}'\|^2} dA' ,$$

where

$$v(\mathbf{x}, \mathbf{x}') = \begin{cases} 1 & \text{if } \mathbf{x} \text{ and } \mathbf{x}\text{' are mutually visible,} \\ 0 & \text{otherwise.} \end{cases}$$

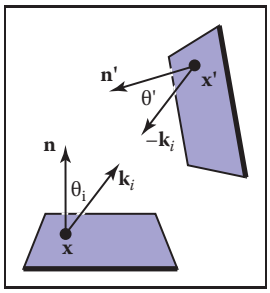


Figure 18.7. The light coming into one point comes from another point.

18.3 Photometry

For every spectral radiometric quantity there is a related *photometric quantity* that measures how much of that quantity is "useful" to a human observer. Given a spectral radiometric quantity $f_r(\lambda)$, the related photometric quantity f_p is

$$f_p = 683 \frac{\text{lm}}{\text{W}} \int_{\lambda=380\text{ nm}}^{800\text{ nm}} \bar{y}(\lambda) f_r(\lambda)\, d\lambda,$$

where $\bar{y}$ is the *luminous efficiency function* of the human visual system. This function is zero outside the limits of integration above, so the limits could be 0 and ∞ and f_p would not change. The luminous efficiency function will be

discussed in more detail in Chapter 19, but we discuss its general properties here. The leading constant is to make the definition consistent with historical absolute photometric quantities.

The luminous efficiency function is not equally sensitive to all wavelengths (Figure 18.8). For wavelengths below 380 nm (the *ultraviolet range*), the light is not visible to humans and thus has a $\bar{y}$ value of zero. From 380 nm it gradually increases until $\lambda = 555$ nm where it peaks. This is a pure green light. Then, it gradually decreases until it reaches the boundary of the infrared region at 800 nm.

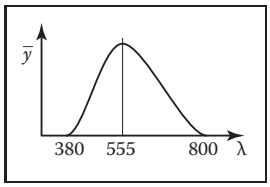


Figure 18.8. The luminous efficiency function versus wavelength (nm).

The photometric quantity that is most commonly used in graphics is *luminance*, the photometric analog of radiance:

$$Y = 683 \frac{\text{lm}}{\text{W}} \int_{\lambda=380\ \text{nm}}^{800\ \text{nm}} \bar{y}(\lambda) L(\lambda)\, d\lambda.$$

The symbol Y for luminance comes from colorimetry. Most other fields use the symbol L; we will not follow that convention because it is too confusing to use L for both luminance and spectral radiance. Luminance gives one a general idea of how "bright" something is independent of the adaptation of the viewer. Note that the black paper under noonday sun is subjectively darker than the lower luminance white paper under moonlight; reading too much into luminance is dangerous, but it is a very useful quantity for getting a quantitative feel for relative perceivable light output. The unit lm stands for *lumens*. Note that most light bulbs are rated in terms of the power they consume in watts, and the useful light they produce in lumens. More efficient bulbs produce more of their light where $\bar{y}$ is large and thus produce more lumens per watt. A "perfect" light would convert all power into 555 nm light and would produce 683 lumens per watt. The units of luminance are thus $(\text{lm}/\text{W})(\text{W}/(\text{m}^2\text{sr})) = \text{lm}/(\text{m}^2\text{sr})$. The quantity one lumen per steradian is defined to be one *candela* (cd), so luminance is usually described in units cd/m^2.

Frequently Asked Questions

• What is "intensity"?

The term *intensity* is used in a variety of contexts and its use varies with both era and discipline. In practice, it is no longer meaningful as a specific radiometric quantity, but it is useful for intuitive discussion. Most papers that use it do so in place of radiance.

• What is “radiosity”?

The term *radiosity* is used in place of radiant exitance in some fields. It is also sometimes used to describe world-space light transport algorithms.

Notes

A common radiometric quantity not described in this chapter is *radiant intensity* (I), which is the spectral power per steradian emitted from an infinitesimal point source. It should usually be avoided in graphics programs because point sources cause implementation problems. A more rigorous treatment of radiometry can be found in *Analytic Methods for Simulated Light Transport* (Arvo, 1995). The radiometric and photometric terms in this chapter are from the *Illumination Engineering Society's* standard that is increasingly used by all fields of science and engineering (American National Standard Institute, 1986). A broader discussion of radiometric and appearance standards can be found in *Principles of Digital Image Synthesis* (Glassner, 1995).

Exercises

1. For a diffuse surface with outgoing radiance L, what is the radiant exitance?
2. What is the total power exiting a diffuse surface with an area of $4\ \mathrm{m}^2$ and a radiance of L?
3. If a fluorescent light and an incandescent light both consume 20 watts of power, why is the fluorescent light usually preferred?

Erik Reinhard and Garrett Johnson

19

Color

Photons are the carriers of optical information. They propagate through media taking on properties associated with waves. At surface boundaries they interact with matter, behaving more as particles. They can also be absorbed by the retina, where the information they carry is transcoded into electrical signals that are subsequently processed by the brain. It is only there that a sensation of color is generated.

As a consequence, the study of color in all its guises touches upon several different fields: physics for the propagation of light through space, chemistry for its interaction with matter, and neuroscience and psychology for aspects relating to perception and cognition of color (Reinhard et al., 2008).

In computer graphics, we traditionally take a simplified view of how light propagates through space. Photons travel along straight paths until they hit a surface boundary and are then reflected according to a reflection function of some sort. A single photon will carry a certain amount of energy, which is represented by its wavelength. Thus, a photon will have only one wavelength. The relationship between its wavelength λ and the amount of energy it carries (ΔE) is given by

$$\lambda \, \Delta E = 1239.9,$$

where ΔE is measured in electron volts (eV).

In computer graphics, it is not very efficient to simulate single photons; instead large collections of them are simulated at the same time. If we take a very large number of photons, each carrying a possibly different amount of energy,

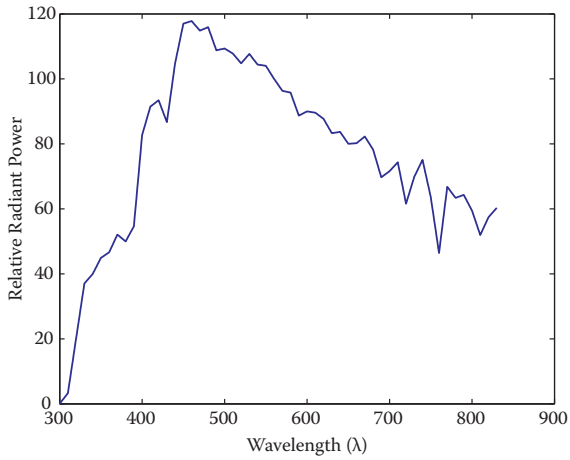


Figure 19.1. A spectrum describes how much energy is available at each wavelength λ, here measured as relative radiant power. This specific spectrum represents average daylight.

then together they represent a spectrum. A spectrum can be thought of as a graph where the number of photons is plotted against wavelength. Because two photons of the same wavelength carry twice as much energy as a single photon of that wavelength, this graph can also be seen as a plot of energy against wavelength. An example of a spectrum is shown in Figure 19.1. The range of wavelengths to which humans are sensitive is roughly between 380 and 800 nanometers (nm).

When simulating light, it would therefore be possible to trace rays that each carry a spectrum. A renderer that accomplishes this is normally called a *spectral renderer*. From preceding chapters, it should be clear that we are not normally going through the expense of building spectral renderers. Instead, we replace spectra with representations that typically use red, green, and blue components. The reason that this is possible at all has to do with human vision and will be discussed later in this chapter.

Simulating light by tracing rays takes care of the physics of light, although it should be noted that several properties of light, including, for instance, polarization, diffraction, and interference, are not modeled in this manner.

At surface boundaries, we normally model what happens with light by means of a reflectance function. These functions can be measured directly by means of *gonioreflectometers*, leading to a large amount of tabled data, which can be more compactly represented by various different functions. Nonetheless, these reflectance functions are empirical in nature, i.e., they abstract away the chemistry that happens when a photon is absorbed and re-emitted by an electron. Thus, reflectance functions are useful for modeling in computer graphics, but do not

offer an explanation as to why certain wavelengths of light are absorbed and others are reflected. We can therefore not use reflectance functions to explain why the light reflected off a banana has a spectral composition that appears to us as yellow. For that, we would have to study molecular orbital theory, a topic beyond the scope of this book.

Finally, when light reaches the retina, it is transcoded into electrical signals that are propagated to the brain. A large part of the brain is devoted to processing visual signals, part of which gives rise to the sensation of color. Thus, even if we know the spectrum of light that is reflected off a banana, we do not know yet why humans associate the term "yellow" with it. Moreover, as we will find out in the remainder of this chapter, our perception of color is vastly more complicated than it would seem at first glance. It changes with illumination, varies between observers, and varies within an observer over time.

In other words, the spectrum of light coming off a banana is perceived in the context of an environment. To predict how an observer perceives a "banana spectrum" requires knowledge of the environment that contains the banana as well as the observer's environment. In many instances, these two environments are the same. However, when we are displaying a photograph of a banana on a monitor, then these two environments will be different. As human visual perception depends on the environment the observer is in, it may perceive the banana in the photograph differently from how an observer directly looking at the banana would perceive it. This has a significant impact on how we should deal with color and illustrates the complexities associated with color.

To emphasize the crucial role that human vision plays, we only have to look at the definition of color: "Color is the aspect of visual perception by which an observer may distinguish differences between two structure-free fields of view of the same size and shape, such as may be caused by differences in the spectral composition of the radiant energy concerned in the observation" (Wyszecki & Stiles, 2000). In essence, without a human observer there is no color.

Luckily, much of what we know about color can be quantified, so that we can carry out computations to correct for the idiosyncrasies of human vision and thereby display images that will appear to observers the way the designer of those images intended. This chapter contains the theory and mathematics required to do so.

19.1 Colorimetry

Colorimetry is the science of color measurement and description. Since color is ultimately a human response, color measurement should begin with human

observation. The photodetectors in the human retina consist of rods and cones. The rods are highly sensitive and come into play in low-light conditions. Under normal lighting conditions, the cones are operational, mediating human vision. There are three cone types and together they are primarily responsible for color vision.

Although it may be possible to directly record the electrical output of cones while some visual stimulus is being presented, such a procedure would be invasive, while at the same time ignoring the sometimes substantial differences between observers. Moreover, much of the measurement of color was developed well before such direct recording techniques were available.

The alternative is to measure color by means of measuring the human response to patches of color. This leads to color matching experiments, which will be described later in this section. Carrying out these experiments have resulted in several standardized observers, which can be thought of as statistical approximations of actual human observers. First, however, we need to describe some of the assumptions underlying the possibility of color matching, which are summarized by Grassmann's laws.

19.1.1 Grassmann's Laws

Given that humans have three different cone types, the experimental laws of color matching can be summed up as the trichromatic generalization (Wyszecki & Stiles, 2000), which states that any color stimulus can be matched completely with an additive mixture of three appropriately modulated color sources. This feature of color is often used in practice, for instance by televisions and monitors which reproduce many different colors by adding a mixture of red, green, and blue light for each pixel. It is also the reason that renderers can be built using only three values to describe each color.

The trichromatic generalization allows us to make color matches between any given stimulus and an additive mixture of three other color stimuli. Hermann Grassmann was the first to describe the algebraic rules to which color matching adheres. They are known as Grassmann's laws of additive color matching (Grassmann, 1853) and are the following:

- Symmetry law. If color stimulus A matches color stimulus B, then B matches A.
- Transitive law. If A matches B and B matches C, then A matches C.
- Proportionality law. If A matches B, then αA matches αB, where α is a positive scale factor.

- Additivity law. If A matches B, C matches D, and $A + C$ matches $B + D$, then it follows that $A + D$ matches $B + C$.

The additivity law forms the basis for color matching and colorimetry as a whole.

19.1.2 Cone Responses

Each cone type is sensitive to a range of wavelengths, spanning most of the full visible range. However, sensitivity to wavelengths is not evenly distributed, but contains a peak wavelength at which sensitivity is greatest. The location of this peak wavelength is different for each cone type. The three cone types are classified as S, M, and L cones, where the letters stand for short, medium, and long, indicating where in the visible spectrum the peak sensitivity is located.

The response of a given cone is then the magnitude of the electrical signal it outputs, as a function of the spectrum of wavelengths incident upon the cone. The cone response functions for each cone type as a function of wavelength λ are then given by $L(\lambda)$, $M(\lambda)$, and $S(\lambda)$. They are plotted in Figure 19.2.

The actual response to a stimulus with a given spectral composition $\Phi(\lambda)$ is then given for each cone type by

$$L = \int_{\lambda} \Phi(\lambda)\, L(\lambda)\, d\lambda,$$
$$M = \int_{\lambda} \Phi(\lambda)\, M(\lambda)\, d\lambda,$$
$$S = \int_{\lambda} \Phi(\lambda)\, S(\lambda)\, d\lambda.$$

These three integrated responses are known as tristimulus values.

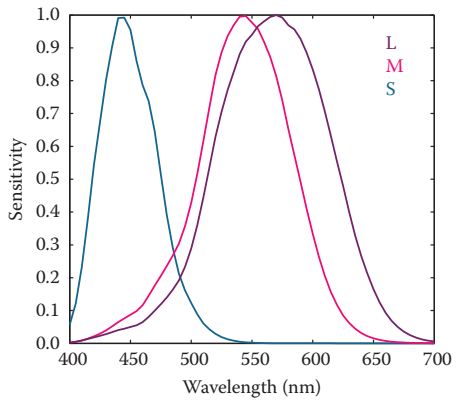


Figure 19.2. The cone response functions for L, M, and S cones.

19.1.3 Color Matching Experiments

Given that tristimulus values are created by integrating the product of two functions over the visible range, it is immediately clear that the human visual system does not act as a simple wavelength detector. Rather, our photo-receptors act as approximately linear integrators. As a result, it is possible to find two different spectral compositions, say $\Phi_1(\lambda)$ and $\Phi_2(\lambda)$, that after integration yield the same response (L, M, S). This phenomenon is known as *metamerism*, an example of which is shown in Figure 19.3.

Metamerism is the key feature of human vision that allows the construction of color reproduction devices, including the color figures in this book and anything reproduced on printers, televisions, and monitors.

Color matching experiments also rely on the principle of metamerism. Suppose we have three differently colored light sources, each with a dial to alter its intensity. We call these three light sources primaries. We should now be able to adjust the intensity of each in such a way that when mixed together additively, the resulting spectrum integrates to a tristimulus value that matches the perceived color of a fourth unknown light source. When we carry out such an experiment, we have essentially matched our primaries to an unknown color. The positions of our three dials are then a representation of the color of the fourth light source.

In such an experiment, we have used Grassmann's laws to add the three spectra of our primaries. We have also used metamerism, because the combined spectrum of our three primaries is almost certainly different from the spectrum of the

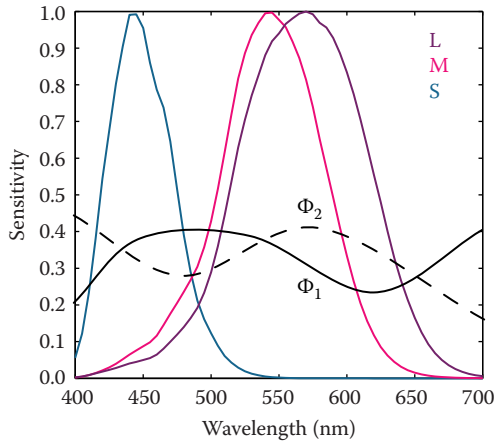


Figure 19.3. Two stimuli $\Phi_1(\lambda)$ and $\Phi_2(\lambda)$ leading to the same tristimulus values after integration.

fourth light source. However, the tristimulus values computed from these two spectra will be identical, having produced a color match.

Note that we do not actually have to know the cone response functions to carry out such an experiment. As long as we use the same observer under the same conditions, we are able to match colors and record the positions of our dials for each color. However, it is quite inconvenient to have to carry out such experiments every time we want to measure colors. For this reason, we do want to know the spectral cone response functions and average those for a set of different observers to eliminate interobserver variability.

19.1.4 Standard Observers

If we perform a color matching experiment for a large range of colors, carried out by a set of different observers, it is possible to generate an average color matching dataset. If we specifically use monochromatic light sources against which to match our primaries, we can repeat this experiment for all visible wavelengths. The resulting tristimulus values are then called *spectral tristimulus values*, and can be plotted against wavelength λ, shown in Figure 19.4.

By using a well-defined set of primary light sources, the spectral tristimulus values lead to three color matching functions. The Commission Internationale d'Eclairage (CIE) has defined three such primaries to be monochromatic light sources of 435.8, 546.1, and 700 nm, respectively. With these three monochro-

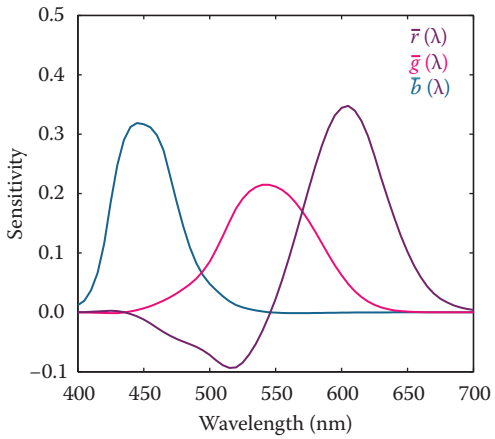


Figure 19.4. Spectral tristimulus values averaged over many observers. The primaries where monochromatic light sources with wavelengths of 435.8, 546.1, and 700 nm.

matic light sources, all other visible wavelengths can be matched by adding different amounts of each. The amount of each required to match a given wavelength λ is encoded in color matching functions, given by $\bar{r}(\lambda)$, $\bar{g}(\lambda)$, and $\bar{b}(\lambda)$ and plotted in Figure 19.4. Tristimulus values associated with these color matching functions are termed R, G, and B.

Given that we are adding light, and light cannot be negative, you may have noticed an anomaly in Figure 19.4: to create a match for some wavelengths, it is necessary to subtract light. Although there is no such thing as negative light, we can use Grassmann's laws once more, and instead of subtracting light from the mixture of primaries, we can add the same amount of light to the color that is being matched.

The CIE $\bar{r}(\lambda)$, $\bar{g}(\lambda)$, and $\bar{b}(\lambda)$ color matching functions allow us to determine if a spectral distribution Φ_1 matches a second spectral distribution Φ_2 by simply comparing the resulting tristimulus values obtained by integrating with these color matching functions:

$$\begin{aligned}
\int_\lambda \Phi_1(\lambda)\, \bar{r}(\lambda) &= \int_\lambda \Phi_2(\lambda)\, \bar{r}(\lambda), \\
\int_\lambda \Phi_1(\lambda)\, \bar{g}(\lambda) &= \int_\lambda \Phi_2(\lambda)\, \bar{g}(\lambda), \\
\int_\lambda \Phi_1(\lambda)\, \bar{b}(\lambda) &= \int_\lambda \Phi_2(\lambda)\, \bar{b}(\lambda).
\end{aligned}$$

Of course, a color match is only guaranteed if all three tristimulus values match.

The importance of these color matching functions lies in the fact that we are now able to communicate and describe colors compactly by means of tristimulus values. For a given spectral function, the CIE color matching functions provide a precise way in which to calculate tristimulus values. As long as everybody uses the same color matching functions, it should always be possible to generate a match.

If the same color matching functions are not available, then it is possible to transform one set of tristimulus values into a different set of tristimulus values appropriate for a corresponding set of primaries. The CIE has defined one such a transform for two specific reasons. First, in the 1930s numerical integrations were difficult to perform, and even more so for functions that can be both positive and negative. Second, the CIE had already developed the photopic luminance response function, CIE $V(\lambda)$. It became desirable to have three integrating functions, of which $V(\lambda)$ is one and all three being positive over the visible range.

To create a set of positive color matching functions, it is necessary to define imaginary primaries. In other words, to reproduce any color in the visible spectrum, we need light sources that cannot be physically realized. The color match-

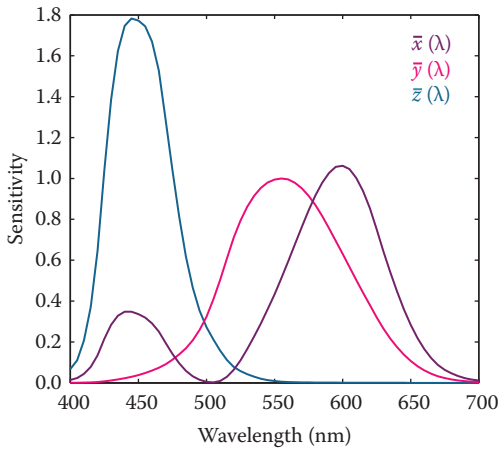


Figure 19.5. The CIE $\bar{x}(\lambda)$, $\bar{y}(\lambda)$, and $\bar{z}(\lambda)$ color matching functions.

ing functions that were settled upon by the CIE are named $\bar{x}(\lambda)$, $\bar{y}(\lambda)$, and $\bar{z}(\lambda)$ and are shown in Figure 19.5. Note that $\bar{y}(\lambda)$ is equal to the photopic luminance response function $V(\lambda)$ and that each of these functions is indeed positive. They are known as the CIE 1931 standard observer.

The corresponding tristimulus values are termed X, Y, and Z, to avoid confusion with R, G, and B tristimulus values that are normally associated with realizable primaries. The conversion from (R, G, B) tristimulus values to (X, Y, Z) tristimulus values is defined by a simple 3×3 transform:

$$\begin{bmatrix} X \\ Y \\ Z \end{bmatrix} = \frac{1}{0.17697} \begin{bmatrix} 0.4900 & 0.3100 & 0.2000 \\ 0.17697 & 0.81240 & 0.01063 \\ 0.0000 & 0.0100 & 0.9900 \end{bmatrix} \cdot \begin{bmatrix} R \\ G \\ B \end{bmatrix}.$$

To calculate tristimulus values, we typically directly integrate the standard observer color matching functions with the spectrum of interest $\Phi(\lambda)$, rather than go through the CIE $\bar{r}(\lambda)$, $\bar{g}(\lambda)$, and $\bar{b}(\lambda)$ color matching functions first, followed by the above transformation. It allows us to calculate consistent color measurements and also determine when two colors match each other.

19.1.5 Chromaticity Coordinates

Every color can be represented by a set of three tristimulus values (X, Y, Z). We could define an orthogonal coordinate system with X, Y, and Z axes and plot each color in the resulting 3D space. This is called a *color space*. The spatial extent of the volume in which colors lie is then called the color gamut.

Visualizing colors in a 3D color space is fairly difficult. Moreover, the Y-value of any color corresponds to its luminance, by virtue of the fact that $\bar{y}(\lambda)$ equals $V(\lambda)$. We could therefore project tristimulus values to a 2D space which approximates chromatic information, i.e., information which is independent of luminance. This projection is called a *chromaticity diagram* and is obtained by normalization while at the same time removing luminance information:

$$x = \frac{X}{X+Y+Z},$$
$$y = \frac{Y}{X+Y+Z},$$
$$z = \frac{Z}{X+Y+Z}.$$

Given that $x + y + z$ equals 1, the z-value is redundant, allowing us to plot the x and y chromaticities against each other in a chromaticity diagram. Although x and y by themselves are not sufficient to fully describe a color, we can use these two chromaticity coordinates and one of the three tristimulus values, traditionally Y, to recover the other two tristimulus values:

$$X = \frac{x}{y}\,Y,$$
$$Z = \frac{1-x-y}{y}\,Y.$$

By plotting all monochromatic (spectral) colors in a chromaticity diagram, we obtain a horseshoe-shaped curve. The points on this curve are called *spectrum loci*. All other colors will generate points lying inside this curve. The spectrum locus for the 1931 standard observer is shown in Figure 19.6. The purple line

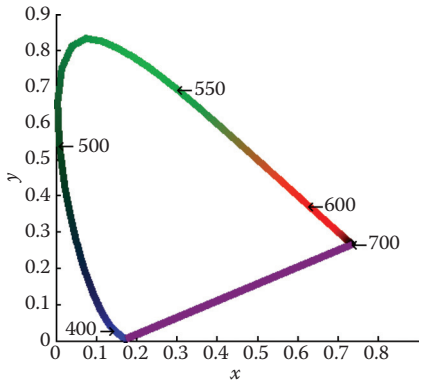


Figure 19.6. The spectrum locus for the CIE 1931 standard observer.

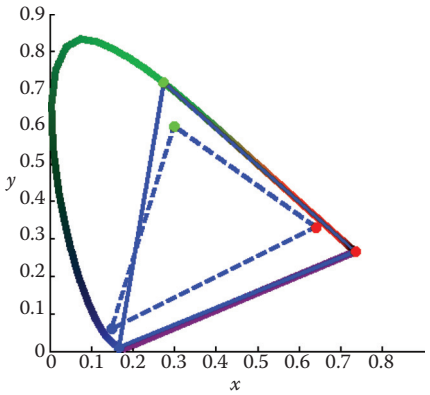


Figure 19.7. The chromaticity boundaries of the CIE RGB primaries at 435.8, 546.1, and 700 nm (solid) and a typical HDTV (dashed).

between either end of the horseshoe does not represent a monochromatic color, but rather a combination of short and long wavelength stimuli.

A (non-monochromatic) primary can be integrated over all visible wavelengths, leading to (X, Y, Z) tristimulus values, and subsequently to an (x, y) chromaticity coordinate, i.e., a point on a chromaticity diagram. Repeating this for two or more primaries yields a set of points on a chromaticity diagram that can be connected by straight lines. The volume spanned in this manner represents the range of colors that can be reproduced by the additive mixture of these primaries. Examples of three-primary systems are shown in Figure 19.7.

Chromaticity diagrams provide insight into additive color mixtures. However, they should be used with care. First, the interior of the horseshoe should not be colored, as any color reproduction system will have its own primaries and can only reproduce some parts of the chromaticity diagram. Second, as the CIE color matching functions do not represent human cone sensitivities, the distance between any two points on a chromaticity diagram is not a good indicator for how differently these colors will be perceived.

A more uniform chromaticity diagram was developed to at least in part address the second of these problems. The CIE $u'v'$ chromaticity diagram provides a perceptually more uniform spacing and is therefore generally preferred over (x, y) chromaticity diagrams. It is computed from (X, Y, Z) tristimulus values by applying a different normalization,

$$u' = \frac{4X}{X + 15Y + 3Z},$$
$$v' = \frac{9Y}{X + 15Y + 3Z}.$$

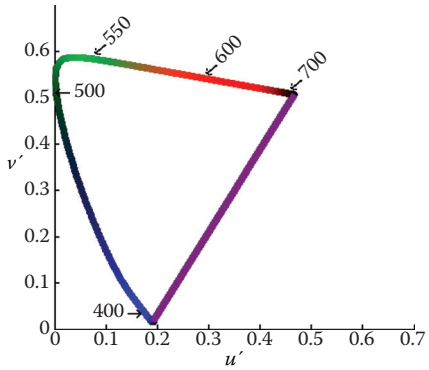


Figure 19.8. The CIE $u'v'$ chromaticity diagram.

and can alternatively be computed directly from (x, y) chromaticity coordinates:

$$u' = \frac{4x}{-2x + 12y + 3},$$
$$v' = \frac{9y}{-2x + 12y + 3}.$$

A CIE $u'v'$ chromaticity diagram is shown in Figure 19.8.

19.2 Color Spaces

As explained above, each color can be represented by three numbers, for instance defined by (X, Y, Z) tristimulus values. However, its primaries are imaginary, meaning that it is not possible to construct a device that has three light sources (all positive) that can reproduce all colors in the visible spectrum.

For the same reason, image encoding and computations on images may not be practical. There is, for instance, a large number of possible XYZ values that do not correspond to any physical color. This would lead to inefficient use of available bits for storage and to a higher requirement for bit-depth to preserve visual integrity after image processing. Although it may be possible to build a capture device that has primaries that are close to the CIE XYZ color matching functions, the cost of hardware and image processing make this an unattractive option. It is not possible to build a display that corresponds to CIE XYZ. For these reasons, it is necessary to design other color spaces: physical realizability, efficient encoding, perceptual uniformity, and intuitive color specification.

The CIE XYZ color space is still actively used, mostly for the conversion between other color spaces. It can be seen as a device-independent color space.

Other color spaces can then be defined in terms of their relationship to CIE XYZ, which is often specified by a specific transform. For instance, linear and additive trichromatic display devices can be transformed to and from CIE XYZ by means of a simple 3×3 matrix. Some nonlinear additional transform may also be specified, for instance to minimize perceptual errors when data is stored with a limited bit-depth, or to enable display directly on devices that have a nonlinear relationship between input signal and the amount of light emitted.

19.2.1 Constructing a Matrix Transform

For a display device with three primaries, say red, green, and blue, we can measure the spectral composition of the emitted light by sending the color vectors $(1, 0, 0)$, $(0, 1, 0)$, and $(0, 0, 1)$. These vectors represent the three cases namely where one of the primaries is full on, and the other two are off. From the measured spectral output, we can then compute the corresponding chromaticity coordinates (x_R, y_R), (x_G, y_G), and (x_B, y_B).

The *white point* of a display is defined as the spectrum emitted when the color vector $(1, 1, 1)$ is sent to the display. Its corresponding chromaticity coordinate is (x_W, y_W). The three primaries and the white point characterize the display and are each required to construct a transformation matrix between the display's color space and CIE XYZ.

These four chromaticity coordinates can be extended to chromaticity triplets reconstructing the z-coordinate from $z = 1-x-y$, leading to triplets (x_R, y_R, z_R), (x_G, y_G, z_G), (x_B, y_B, z_B), and (x_W, y_W, z_W). If we know the maximum luminance of the white point, we can compute its corresponding tristimulus value (X_W, Y_W, Z_W) and then solve the following set of equations for the luminance ratio scalars S_R, S_G, and S_B:

$$\begin{aligned} X_W &= x_R\, S_R + x_G\, S_G + x_B\, S_B, \\ Y_W &= y_R\, S_R + y_G\, S_G + y_B\, S_B, \\ Z_W &= z_R\, S_R + z_G\, S_G + z_B\, S_B. \end{aligned}$$

The conversion between RGB and XYZ is then given by

$$\begin{bmatrix} X \\ Y \\ Z \end{bmatrix} = \begin{bmatrix} x_R\, S_R & x_G\, S_G & x_B\, S_B \\ y_R\, S_R & y_G\, S_G & y_B\, S_B \\ z_R\, S_R & z_G\, S_G & z_B\, S_B \end{bmatrix} \begin{bmatrix} R \\ G \\ B \end{bmatrix}.$$

The luminance of any given color can be computed by evaluating the middle row of a matrix constructed in this manner:

$$Y = y_R\, S_R\, R + y_G\, S_G\, G + y_B\, S_B\, B.$$

	R	G	B	White
x	0.6400	0.3000	0.1500	0.3127
y	0.3300	0.6000	0.0600	0.3290

Table 19.1. The (x, y) chromaticity coordinates for the primaries and white point specified by ITU-R BT.709. The sRGB standard also uses these primaries and white point.

To convert between XYZ and RGB of a given device, the above matrix can simply be inverted.

If an image is represented in an RGB color space for which the primaries and white point are unknown, then the next best thing is to assume that the image was encoded in a standard RGB color space. A reasonable choice is then to assume that the image was specified according to ITU-R BT.709, which is the specification used for encoding and broadcasting of HDTV. Its primaries and white point are specified in Table 19.1. Note that the same primaries and white point are used to define the well-known sRGB color space. The transformation between this RGB color space and CIE XYZ is and vice versa given by

$$\begin{bmatrix} X \\ Y \\ Z \end{bmatrix} = \begin{bmatrix} 0.4124 & 0.3576 & 0.1805 \\ 0.2126 & 0.7152 & 0.0722 \\ 0.0193 & 0.1192 & 0.9505 \end{bmatrix} \begin{bmatrix} R \\ G \\ B \end{bmatrix};$$

$$\begin{bmatrix} R \\ G \\ B \end{bmatrix} = \begin{bmatrix} 3.2405 & -1.5371 & -0.4985 \\ -0.9693 & 1.8706 & 0.0416 \\ 0.0556 & -0.2040 & 1.0572 \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \end{bmatrix}.$$

By substituting the maximum RGB values of the device, we can compute the white point. For ITU-R BT.709, the maximum values are $(R_W, G_W, B_W) = (100, 100, 100)$, leading to a white point of $(X_W, Y_W, Z_W) = (95.05, 100.00, 108.90)$.

In addition to a linear transformation, the sRGB color space is characterized by a subsequent nonlinear transform. The nonlinear encoding is given by

$$R_{\text{sRGB}} = \begin{cases} 1.055\, R^{1/2.4} - 0.055 & R > 0.0031308, \\ 12.92\, R & R \leq 0.0031308; \end{cases}$$

$$G_{\text{sRGB}} = \begin{cases} 1.055\, G^{1/2.4} - 0.055 & G > 0.0031308, \\ 12.92\, G & G \leq 0.0031308; \end{cases}$$

$$B_{\text{sRGB}} = \begin{cases} 1.055\, B^{1/2.4} - 0.055 & B > 0.0031308, \\ 12.92\, B & B \leq 0.0031308. \end{cases}$$

This nonlinear encoding helps minimize perceptual errors due to quantization errors in digital applications.

19.2.2 Device-Dependent RGB Spaces

As each device typically has its own set of primaries and white point, we call the associated RGB color spaces device-dependent. It should be noted that even if all these devices operate in an RGB space, they may have very different primaries and white points. If we therefore have an image specified in some RGB space, it may appear very different to us, depending upon which device we display it.

This is clearly an undesirable situation, resulting from a lack of color management. However, if the image is specified in a known RGB color space, it can first be converted to XYZ, which is device independent, and then subsequently it can be converted to the RGB space of the device on which it will be displayed.

There are several other RGB color spaces that are well defined. They each consist of a linear matrix transform followed by a nonlinear transform, akin to the aforementioned sRGB color space. The nonlinear transform can be parameterized as follows:

$$R_{\text{nonlinear}} = \begin{cases} (1+f)\,R^{\gamma} - f & t < R \leq 1, \\ s\,R & 0 \leq R \leq t; \end{cases}$$

$$G_{\text{nonlinear}} = \begin{cases} (1+f)\,G^{\gamma} - f & t < G \leq 1, \\ s\,G & 0 \leq G \leq t; \end{cases}$$

$$B_{\text{nonlinear}} = \begin{cases} (1+f)\,B^{\gamma} - f & t < B \leq 1, \\ s\,B & 0 \leq B \leq t. \end{cases}$$

The parameters s, f, t and γ, together with primaries and white point, specify a class of RGB color spaces that are used in various industries. Several common transformations are listed in Table 19.2.

19.2.3 LMS Cone Space

The aforementioned cone signals can be expressed in terms of the CIE XYZ color space. The matrix transform to compute LMS signals from XYZ and vice versa are given by

$$\begin{bmatrix} L \\ M \\ S \end{bmatrix} = \begin{bmatrix} 0.38971 & 0.68898 & -0.07868 \\ -0.22981 & 1.18340 & 0.04641 \\ 0.00000 & 0.00000 & 1.00000 \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \end{bmatrix};$$

$$\begin{bmatrix} X \\ Y \\ Z \end{bmatrix} = \begin{bmatrix} 1.91019 & -1.11214 & 0.20195 \\ 0.37095 & 0.62905 & 0.00000 \\ 0.00000 & 0.00000 & 1.00000 \end{bmatrix} \begin{bmatrix} L \\ M \\ S \end{bmatrix}.$$

Color space	XYZ to RGB matrix	RGB to XYZ matrix	Nonlinear transform
sRGB	$\begin{bmatrix} 3.2405 & -1.5371 & -0.4985 \\ -0.9693 & 1.8760 & 0.0416 \\ 0.0556 & -0.2040 & 1.0572 \end{bmatrix}$	$\begin{bmatrix} 0.4124 & 0.3576 & 0.1805 \\ 0.2126 & 0.7152 & 0.0722 \\ 0.0193 & 0.1192 & 0.9505 \end{bmatrix}$	$\gamma = 1/2.4 \approx 0.42$ $f = 0.055$ $s = 12.92$ $t = 0.0031308$
Adobe RGB (1998)	$\begin{bmatrix} 2.0414 & -0.5649 & -0.3447 \\ -0.9693 & 1.8760 & 0.0416 \\ 0.0134 & -0.1184 & 1.0154 \end{bmatrix}$	$\begin{bmatrix} 0.5767 & 0.1856 & 0.1882 \\ 0.2974 & 0.6273 & 0.0753 \\ 0.0270 & 0.0707 & 0.9911 \end{bmatrix}$	$\gamma = \frac{1}{2\frac{51}{256}} \approx \frac{1}{2.2}$ $f =$ N.A. $s =$ N.A. $t =$ N.A.
HDTV (HD-CIF)	$\begin{bmatrix} 3.2405 & -1.5371 & -0.4985 \\ -0.9693 & 1.8760 & 0.0416 \\ 0.0556 & -0.2040 & 1.0572 \end{bmatrix}$	$\begin{bmatrix} 0.4124 & 0.3576 & 0.1805 \\ 0.2126 & 0.7152 & 0.0722 \\ 0.0193 & 0.1192 & 0.9505 \end{bmatrix}$	$\gamma = 0.45$ $f = 0.099$ $s = 4.5$ $t = 0.018$
NTSC (1953)/ ITU-R BT.601-4	$\begin{bmatrix} 1.9100 & -0.5325 & -0.2882 \\ -0.9847 & 1.9992 & -0.0283 \\ 0.0583 & -0.1184 & 0.8976 \end{bmatrix}$	$\begin{bmatrix} 0.6069 & 0.1735 & 0.2003 \\ 0.2989 & 0.5866 & 0.1145 \\ 0.0000 & 0.0661 & 1.1162 \end{bmatrix}$	$\gamma = 0.45$ $f = 0.099$ $s = 4.5$ $t = 0.018$
PAL/SECAM	$\begin{bmatrix} 3.0629 & -1.3932 & -0.4758 \\ -0.9693 & 1.8760 & 0.0416 \\ 0.0679 & -0.2289 & 1.0694 \end{bmatrix}$	$\begin{bmatrix} 0.4306 & 0.3415 & 0.1783 \\ 0.2220 & 0.7066 & 0.0713 \\ 0.0202 & 0.1296 & 0.9391 \end{bmatrix}$	$\gamma = 0.45$ $f = 0.099$ $s = 4.5$ $t = 0.018$
SMPTE-C	$\begin{bmatrix} 3.5054 & -1.7395 & -0.5440 \\ -1.0691 & 1.9778 & 0.0352 \\ 0.0563 & -0.1970 & 1.0502 \end{bmatrix}$	$\begin{bmatrix} 0.3936 & 0.3652 & 0.1916 \\ 0.2124 & 0.7010 & 0.0865 \\ 0.0187 & 0.1119 & 0.9582 \end{bmatrix}$	$\gamma = 0.45$ $f = 0.099$ $s = 4.5$ $t = 0.018$
SMPTE-240M	$\begin{bmatrix} 2.042 & -0.565 & -0.345 \\ -0.894 & 1.815 & 0.032 \\ 0.064 & -0.129 & 0.912 \end{bmatrix}$	$\begin{bmatrix} 0.567 & 0.190 & 0.193 \\ 0.279 & 0.643 & 0.077 \\ 0.000 & 0.073 & 1.016 \end{bmatrix}$	$\gamma = 0.45$ $f = 0.1115$ $s = 4.0$ $t = 0.0228$
Wide Gamut	$\begin{bmatrix} 1.4625 & -0.1845 & -0.2734 \\ -0.5228 & 1.4479 & 0.0681 \\ 0.0346 & -0.0958 & 1.2875 \end{bmatrix}$	$\begin{bmatrix} 0.7164 & 0.1010 & 0.1468 \\ 0.2587 & 0.7247 & 0.0166 \\ 0.0000 & 0.0512 & 0.7740 \end{bmatrix}$	$\gamma =$ N.A. $f =$ N.A. $s =$ N.A. $t =$ N.A.

Table 19.2. Transformations for standard RGB color spaces (after (Pascale, 2003)).

This transform is known as the Hunt-Pointer-Estevez transform (Hunt, 2004) and is used in chromatic adaptation transforms as well as in color appearance modeling.

19.2.4 CIE 1976 $L^*a^*b^*$

Color opponent spaces are characterized by a channel representing an achromatic channel (luminance), as well as two channels encoding color opponency. These are frequently red-green and yellow-blue channels. These color opponent chan-

nels thus encode two chromaticities along one axis, which can have both positive and negative values. For instance, a red-green channel encodes red for positive values and green for negative values. The value zero encodes a special case: neutral which is neither red or green. The yellow-blue channel works in much the same way.

As at least two colors are encoded on each of the two chromatic axes, it is not possible to encode a mixture of red and green. Neither is it possible to encode yellow and blue simultaneously. While this may seem a disadvantage, it is known that the human visual system computes similar attributes early in the visual pathway. As a result, humans are not able to perceive colors that are simultaneously red and green, or yellow and blue. We do not see anything resembling reddish-green, or yellowish-blue. We are, however, able to perceive mixtures of colors such as yellowish-red (orange) or greenish-blue, as these are encoded across the chromatic channels.

The most relevant color opponent system for computer graphics is the CIE 1976 $L^*a^*b^*$ color model. It is a perceptually more or less uniform color space, useful, among other things, for the computation of color differences. It is also known as CIELAB.

The input to CIELAB are the stimulus (X, Y, Z) tristimulus values as well as the tristimulus values of a diffuse white reflecting surface that is lit by a known illuminant, (X_n, Y_n, Z_n). CIELAB therefore goes beyond being an ordinary color space, as it takes into account a patch of color in the context of a known illumination. It can thus be seen as a rudimentary color appearance space.

The three channels defined in CIELAB are L^*, a^*, and b^*. The L^* channel encodes the lightness of the color, i.e., the perceived reflectance of a patch with tristimulus value (X, Y, Z). The a^* and b^* are chromatic opponent channels. The transform between XYZ and CIELAB is given by

$$\begin{bmatrix} L^* \\ a^* \\ b^* \end{bmatrix} = \begin{bmatrix} 0 & 116 & 0 & -16 \\ 500 & -500 & 0 & 0 \\ 0 & 200 & -200 & 0 \end{bmatrix} \begin{bmatrix} f\left(X/X_n\right) \\ f\left(Y/Y_n\right) \\ f\left(Z/Z_n\right) \\ 1 \end{bmatrix}.$$

The function f is defined as

$$f(r) = \begin{cases} \sqrt[3]{r} & \text{for } r > 0.008856, \\ 7.787\, r + \dfrac{16}{116} & \text{for } r \leq 0.008856. \end{cases}$$

As can be seen from this formulation, the chromatic channels do depend on the luminance Y. Although this is perceptually accurate, it means that we cannot plot the values of a^* and b^* in a chromaticity diagram. The lightness L^* is normalized

between 0 and 100 for black and white. Although the a^* and b^* channels are not explicitly constrained, they are typically in the range $[-128, 128]$.

As CIELAB is approximately perceptually linear, it is possible to take two colors, convert them to CIELAB, and then estimate the perceived color difference by computing the Euclidean distance between them. This leads to the following color difference formula:

$$\Delta E^*_{\text{ab}} = \left[(\Delta L^*)^2 + (\Delta a^*)^2 + (\Delta b^*)^2 \right]^{1/2}.$$

The letter E stands for difference in sensation (in German, Empfindung) (Judd, 1932).

Finally, the inverse transform between CIELAB and XYZ is given by

$$X = X_n \begin{cases} \left(\frac{L^*}{116} + \frac{a^*}{500} + \frac{16}{116} \right)^3 & \text{if } L^* > 7.9996, \\ \frac{1}{7.787} \left(\frac{L^*}{116} + \frac{a^*}{500} \right) & \text{if } L^* \leq 7.9996, \end{cases}$$

$$Y = Y_n \begin{cases} \left(\frac{L^*}{116} + \frac{16}{116} \right)^3 & \text{if } L^* > 7.9996, \\ \frac{1}{7.787} \frac{L^*}{116} & \text{if } L^* \leq 7.9996, \end{cases}$$

$$Z = Z_n \begin{cases} \left(\frac{L^*}{116} - \frac{b^*}{200} + \frac{16}{116} \right)^3 & \text{if } L^* > 7.9996, \\ \frac{1}{7.787} \left(\frac{L^*}{116} - \frac{b^*}{200} \right) & \text{if } L^* \leq 7.9996. \end{cases}$$

19.3 Chromatic Adaptation

The CIELAB color space just described takes as input both a tristimulus value of the stimulus and the tristimulus value of light reflected off a white diffuse patch. As such, it forms the beginnings of a system in which the viewing environment is taken into account.

The environment in which we observe objects and images has a large influence on how we perceive those objects. The range of viewing environments that we encounter in daily life is very large, from sunlight to starlight and from candlelight to fluorescent light. The lighting conditions not only constitute a very large range in the amount of light that is present, but also vary greatly in the color of the emitted light.

The human visual system accommodates these changes in the environment through a process called adaptation. Three different types of adaptation can be distinguished, namely light adaptation, dark adaptation, and chromatic adaptation. Light adaptation refers to the changes that occur when we move from a very dark to a very light environment. When this happens, at first we are dazzled by the light, but soon we adapt to the new situation and begin to distinguish objects in our environment. Dark adaptation refers to the opposite—when we go from a light environment to a dark environment. At first, we see very little, but after a given amount of time, details will start to emerge. The time needed to adapt to the dark is generally much longer than for light adaptation.

Chromatic adaptation refers to our ability to adapt, and largely ignore, variations in the color of the illumination. Chromatic adaptation is, in essence, the biological equivalent of the white balancing operation that is available on most modern cameras. The human visual system effectively normalizes the viewing conditions to present a visual experience that is fairly consistent. Thus, we exhibit a certain amount of color constancy: object reflectances appear relatively constant despite variations in illumination.

Although we are able to largely ignore changes in viewing environment, we are not able to do so completely. For instance, colors appear much more colorful on a sunny day than they do on a cloudy day. Although the appearances have changed, we do not assume that object reflectances themselves have actually changed their physical properties. We thus understand that the lighting conditions have influenced the overall color appearance.

Nonetheless, color constancy does apply to chromatic content. Chromatic adaptation allows white objects to appear white for a large number of lighting conditions, as shown in Figure 19.9.

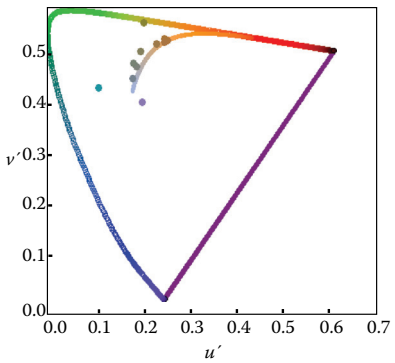


Figure 19.9. A series of light sources plotted in the CIE $u'v'$ chromaticity diagram. A white piece of paper illuminated by any of these light sources maintains a white color appearance.

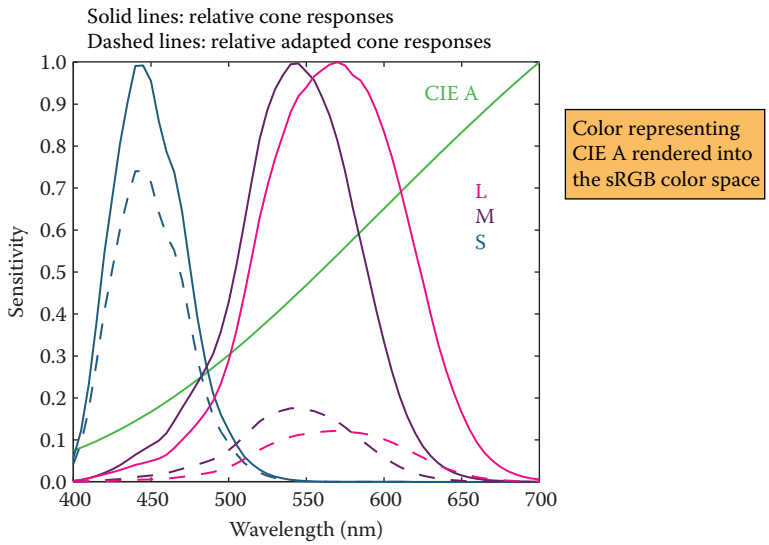


Figure 19.10. An example of von Kries–style independent photoreceptor gain control. The relative cone responses (solid line) and the relative adapted cone responses to CIE illuminant A (dashed) are shown. The separate patch of color represents CIE illuminant A rendered into the sRGB color space.

Computational models of chromatic adaptation tend to focus on the gain control mechanism in the cones. One of the simplest models assumes that each cone adapts independently to the energy that it absorbs. This means that different cone types adapt differently dependent on the spectrum of the light being absorbed. Such adaptation can then be modeled as an adaptive and independent rescaling of the cone signals:

$$L_a = \alpha\, L,$$
$$M_a = \beta\, M,$$
$$S_a = \gamma\, S,$$

where (L_a, M_a, S_a) are the chromatically adapted cone signals, and α, β, and γ are the independent gain controls which are determined by the viewing environment. This type of independent adaptation is also known as von-Kries adaptation. An example is shown in Figure 19.10.

The adapting illumination can be measured off a white surface in the scene. In the ideal case, this would be a Lambertian surface. In a digital image, the adapting illumination can also be approximated as the maximum tristimulus values of the scene. The light measured or computed in this manner is the adapting white, given by (L_w, M_w, S_w). Von Kries adaptation is then simply a scaling by the reciprocal

of the adapting white, carried out in cone response space:

$$\begin{bmatrix} L_a \\ M_a \\ S_a \end{bmatrix} = \begin{bmatrix} \frac{1}{L_w} & 0 & 0 \\ 0 & \frac{1}{M_w} & 0 \\ 0 & 0 & \frac{1}{S_w} \end{bmatrix} \begin{bmatrix} L \\ M \\ S \end{bmatrix}.$$

In many cases, we are interested in what stimulus should be generated under one illumination to match a given color under a different illumination. For example, if we have a colored patch illuminated by daylight, we may ask ourselves what tristimulus values should be generated to create a matching color patch that will be illuminated by incandescent light.

We are thus interested in computing corresponding colors, which can be achieved by cascading two chromatic adaptation calculations. In essence, the previously mentioned von Kries transform divides out the adapting illuminant—in our example, the daylight illumination. If we subsequently multiply in the incandescent illuminant, we have computed a corresponding color. If the two illuminants are given by $(L_{w,1}, M_{w,1}, S_{w,1})$ and $(L_{w,2}, M_{w,2}, S_{w,2})$, the corresponding color (L_c, M_c, S_c) is given by

$$\begin{bmatrix} L_c \\ M_c \\ S_c \end{bmatrix} = \begin{bmatrix} L_{w,2} & 0 & 0 \\ 0 & M_{w,2} & 0 \\ 0 & 0 & S_{w,2} \end{bmatrix} \begin{bmatrix} \frac{1}{L_{w,1}} & 0 & 0 \\ 0 & \frac{1}{M_{w,1}} & 0 \\ 0 & 0 & \frac{1}{S_{w,1}} \end{bmatrix} \begin{bmatrix} L \\ M \\ S \end{bmatrix}.$$

There are several more complicated and, therefore, more accurate chromatic adaptation transform in existence (Reinhard et al., 2008). However, the simple von Kries model remains remarkably effective in modeling chromatic adaptation and can thus be used to achieve white balancing in digital images.

The importance of chromatic adaptation in the context of rendering, is that we have moved one step closer to taking into account the viewing environment of the observer, without having to correct for it by adjusting the scene and rerendering our imagery. Instead, we can model and render our scenes, and then, as an image postprocess, correct for the illumination of the viewing environment. To ensure that white balancing does not introduce artifacts, however, it is important to ensure that the image is rendered to a floating-point format. If rendered to traditional 8-bit image formats, the chromatic adaptation transform may amplify quantization errors.

19.4 Color Appearance

While colorimetry allows us to accurately specify and communicate color in a device-independent manner, and chromatic adaptation allows us to predict color matches across changes in illumination, these tools are still insufficient to describe what colors actually look like.

To predict the actual perception of an object, we need to know more information about the environment and take that information into account. The human visual system is constantly adapting to its environment, which means that the perception of color will be strongly influenced by such changes. Color appearance models take into account measurements of the stimulus itself, as well as the viewing environment. This means that the resulting description of color is independent of viewing condition.

The importance of color appearance modeling can be seen in the following example. Consider an image being displayed on an LCD screen. When making a print of the same image and viewing it in a different context, more often than not the image will look markedly different. Color appearance models can be used to predict the changes required to generate an accurate cross-media color reproduction (Fairchild, 2005).

Although color appearance modeling offers important tools for color reproduction, actual implementations tend to be relatively complicated and cumbersome in practical use. It can be anticipated that this situation may change over time. However, until then, we leave their description to more specialized textbooks (Fairchild, 2005).

Notes

Of all the books on color theory, Reinhard et al.'s work (Reinhard et al., 2008) is most directly geared toward engineering disciplines, including computer graphics, computer vision, and image processing. Other general introductions to color theory are given by Berns (Berns, 2000) and Stone (Stone, 2003). Wyszecki and Stiles have produced a comprehensive volume of data and formulae, forming an indispensable reference work (Wyszecki & Stiles, 2000). For color reproduction, we recommend Hunt's book (Hunt, 2004). Color appearance models are comprehensively described in Fairchild's book (Fairchild, 2005). For color issues related to video and HDTV Poynton's book is essential (Poynton, 2003).

William B. Thompson

20

Visual Perception

The ultimate purpose of computer graphics is to produce images for viewing by people. Thus, the success of a computer graphics system depends on how well it conveys relevant information to a human observer. The intrinsic complexity of the physical world and the limitations of display devices make it impossible to present a viewer with the identical patterns of light that would occur when looking at a natural environment. When the goal of a computer graphics system is physical realism, the best we can hope for is that the system be *perceptually effective*: displayed images should "look" as intended. For applications such as technical illustration, it is often desirable to visually highlight relevant information and perceptual effectiveness becomes an explicit requirement.

Artists and illustrators have developed empirically a broad range of tools and techniques for effectively conveying visual information. One approach to improving the perceptual effectiveness of computer graphics is to utilize these methods in our automated systems. A second approach builds directly on knowledge of the human vision system by using perceptual effectiveness as an optimization criterion in the design of computer graphics systems. These two approaches are not completely distinct. Indeed, one of the first systematic examinations of visual perception is found in the notebooks of Leonardo da Vinci.

The remainder of this chapter provides a partial overview of what is known about visual perception in people. The emphasis is on aspects of human vision that are most relevant to computer graphics. The human visual system is extremely complex in both its operation and its architecture. A chapter such as this

can at best provide a summary of key points, and it is important to avoid over generalizing from what is presented here. More in-depth treatments of visual perception can be found in Wandell (1995) and Palmer (1999); Gregory (1997) and Yantis (2000) provide additional useful information. A good computer vision reference such as Forsyth and Ponce (2002) is also helpful. It is important to note that despite over 150 years of intensive research, our knowledge of many aspects of vision is still very limited and imperfect.

20.1 Vision Science

Light:

- travels far
- travels fast
- travels in straight lines
- interacts with stuff
- bounces off things
- is produced in nature
- has lots of energy

—Steven Shafer

Figure 20.1. The nature of light makes vision a powerful sense.

Vision is generally agreed to be the most powerful of the senses in humans. Vision produces more useful information about the world than does hearing, touch, smell, or taste. This is a direct consequence of the physics of light (Figure 20.1). Illumination is pervasive, especially during the day but also at night due to moonlight, starlight, and artificial sources. Surfaces reflect a substantial portion of incident illumination and do so in ways that are idiosyncratic to particular materials and that are dependent on the shape of the surface. The fact that light (mostly) travels in straight lines through the air allows vision to acquire information from distant locations.

The study of vision has a long and rich history. Much of what we know about the eye traces back to the work of philosophers and physicists in the 1600s. Starting in the mid-1800s, there was an explosion of work by perceptual psychologists exploring the phenomenology of vision and proposing models of how vision might work. The mid-1900s saw the start of modern neuroscience, which investigates both the fine-scale workings of individual neurons and the large-scale architectural organization of the brain and nervous system. A substantial portion of neuroscience research has focused on vision. More recently, computer science has contributed to the understanding of visual perception by providing tools for precisely describing hypothesized models of visual computations and by allowing empirical examination of computer vision programs. The term *vision science* was coined to refer to the multidisciplinary study of visual perception involving perceptual psychology, neuroscience, and computational analysis.

Vision science views the purpose of vision as producing information about objects, locations, and events in the world from imaged patterns of light reaching the viewer. Psychologists use the term *distal stimulus* to refer to the physical world under observation and *proximal stimulus* to refer to the retinal image.[1] Us-

[1] In computer vision, the term *scene* is often used to refer to the external world, while the term *image* is used to refer to the projection of the scene onto a sensing plane.

ing this terminology, the function of vision is to generate a description of aspects of the distal stimulus given the proximal stimulus. Visual perception is said to be *veridical* when the description that is produced accurately reflects the real world. In practice, it makes little sense to think of these descriptions of objects, locations, and events in isolation. Rather, vision is better understood in the context of the motor and cognitive functions that it serves.

20.2 Visual Sensitivity

Vision systems create descriptions of the visual environment based on properties of the incident illumination. As a result, it is important to understand what properties of incident illumination the human vision system can actually detect. One critical observation about the human vision system is that it is primarily sensitive to *patterns* of light rather than being sensitive to the absolute magnitude of light energy. The eye does not operate as a photometer. Instead, it detects spatial, temporal, and spectral patterns in the light imaged on the retina and information about these patterns of light form the basis for all of visual perception.

There is a clear ecological utility to the vision system's sensitivity to variations in illumination over space and time. Being able to accurately sense changes in the environment is crucial to our survival.[2] A system which measures changes in light energy rather than the magnitude of the energy itself also makes engineering sense, since it makes it easier to detect patterns of light over large ranges in light intensity. It is a good thing for computer graphics that vision operates in this manner. Display devices are physically limited in their ability to project light with the power and dynamic range typical of natural scenes. Graphical displays would not be effective if they needed to produce the identical patterns of light as the corresponding physical world. Fortunately, all that is required is that displays be able to produce similar patterns of spatial and temporal change to the real world.

20.2.1 Brightness and Contrast

In bright light, the human visual system is capable of distinguishing gratings consisting of high-contrast parallel light and dark bars as fine as 50–60 cycles/degree. (In this case, a "cycle" consists of an adjacent pair of light and dark bars.)

[2] It is sometime said that the primary goals of vision are to support eating, avoiding being eaten, reproduction, and avoidance of catastrophe while moving. Thinking about vision as a goal-directed activity is often useful, but needs to be done so at a more detailed level.

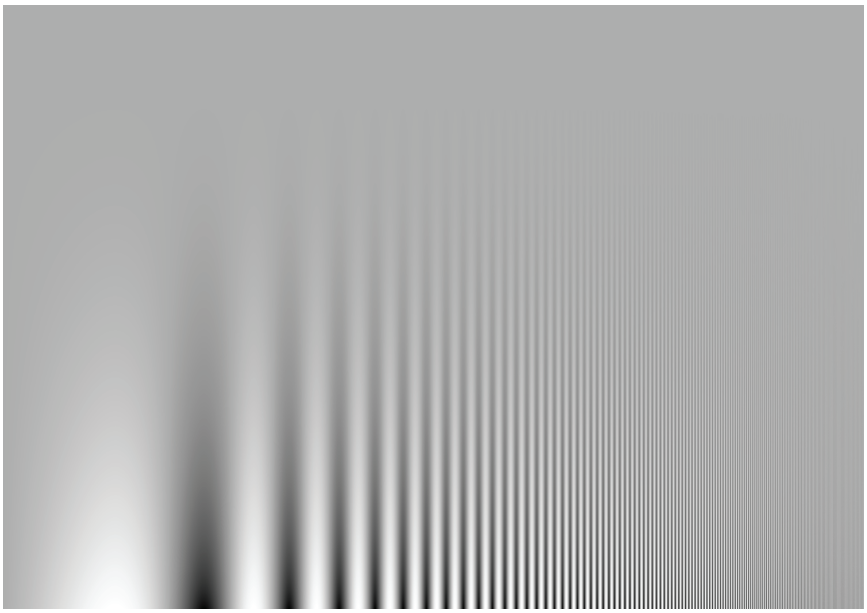


Figure 20.2. The contrast between stripes increases in a constant manner from top to bottom, yet the threshold of visibility varies with frequency.

For comparison, the best currently available LCD computer monitor, at a normal viewing distance, can display patterns as fine as about 20 cycles/degree. The minimum contrast difference at an edge detectable by the human visual system in bright light is about 1% of the average luminance across the edge. In most 8-bit displays, differences of a single gray level are often noticeable over at least a portion of the range of intensities due to the nature of the mapping from gray levels to actual display luminance.

Characterizing the ability of the visual system to detect fine scale patterns (*visual acuity*) and to detect changes in brightness is considerably more complicated than for cameras and similar image acquisition devices. As shown in Figure 20.2, there is an interaction between contrast and acuity in human vision. In the figure, the scale of the pattern decreases from left to right while the contrast increases from top to bottom. If you view the figure at a normal viewing distance, it will be clear that the lowest contrast at which a pattern is visible is a function of the spatial frequency of the pattern.

There is a linear relationship between the intensity of light L reaching the eye from a particular surface point in the world, the intensity of light I illuminating that surface point, and the reflectivity R of the surface at the point being observed:

$$L = \alpha I \cdot R, \tag{20.1}$$

Figure 20.3. *Lightness constancy.* Cast a shadow over one of the patterns with your hand and notice that the apparent brightness of the two center squares remains nearly the same.

where α is dependent on the relationship between the surface geometry, the pattern of incident illumination, and the viewing direction. While the eye is only able to directly measure L, human vision is much better at estimating R than L. To see this, view Figure 20.3 in bright direct light. Use your hand to shadow one of the patterns, leaving the other directly illuminated. While the light reflected off of the two patterns will be significantly different, the apparent brightness of the two center squares will seem nearly the same. The term *lightness* is often used to describe the apparent brightness of a surface, as distinct from its actual luminance. In many situations, lightness is invariant to large changes in illumination, a phenomenon referred to as *lightness constancy*.

The mechanisms by which the human visual system achieves lightness constancy are not well understood. As shown in Figure 20.2, the vision system is relatively insensitive to slowly varying patterns of light, which may serve to discount the effects of slowly varying illumination. Apparent brightness is affected by the brightness of surrounding regions (Figure 20.4). This can aid lightness constancy when regions are illuminated dissimilarly. While this *simultaneous contrast* effect is often described as a modification of the perceived lightness of

(a) (b)

Figure 20.4. (a) Simultaneous contrast: the apparent brightness of the center bar is affected by the brightness of the surrounding area; (b) The same bar without a variable surround.

(a) (b)

Figure 20.5. The Munker-White illusion shows the complexity of simultaneous contrast. In Figure 20.4, the central region looked lighter when the surrounding area was darker. In (a), the gray strips on the left look *lighter* than the gray strips on the right, even though they are nearly surrounded by regions of white; (b) shows the gray strips without the black lines.

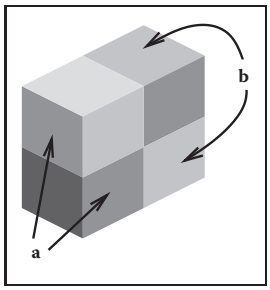
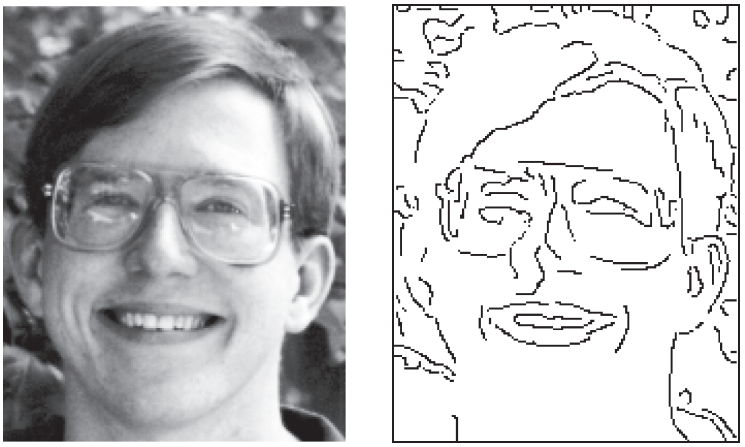


Figure 20.6. The perception of lightness is affected by the perception of 3D structure. The two surfaces marked (a) have the same brightness, as do the two surfaces marked (b) (after Adelson (1999)).

one region based on contrasting brightness in the surrounding region, it is actually much more complicated than that (Figures 20.5 and 20.6). For more on lightness perception, see (Gilchrist et al., 1999) and (Adelson, 1999).

While the visual system largely ignores slowly varying intensity patterns, it is extremely sensitive to *edges* consisting of lines of discontinuity in brightness. Edges in imaged light intensity often correspond to surface boundaries or other important features in the environment (Figure 20.7). The vision system can also detect localized differences in motion, stereo disparity, texture, and several other



(a) (b)

Figure 20.7. (a) Original gray scale image, (b) image *edges*, which are lines of high spatial variability in some direction.

Figure 20.8. The visual system sometimes sees "edges" even when there are no sharp discontinuities in brightness, as is the case at the right side of the central pattern in this image.

image properties. The vision system has very little ability, however, to detect spatial discontinuities in color when not accompanied by differences in one of these other properties.

Perception of edges seems to interact with perception of form. While edges give the visual system the information it needs to recognize shapes, slowly varying brightness can appear as a sharp edge if the resulting edge creates a more complete form (Figure 20.8). Figure 20.9 shows a *subjective contour*, an extreme form of this effect in which a closed contour is seen even though no such contour exists in the actual image. Finally, the vision system's sensitivity to edges also appears to be part of the mechanism involved in lightness perception. Note that the region enclosed by the subjective contour in Figure 20.9 appears a bit brighter than the surrounding area of the page. Figure 20.10 shows a different interaction between edges and lightness. In this case, a particular brightness profile at the edge has a dramatic effect on the apparent brightness of the surfaces to either side of the edge.

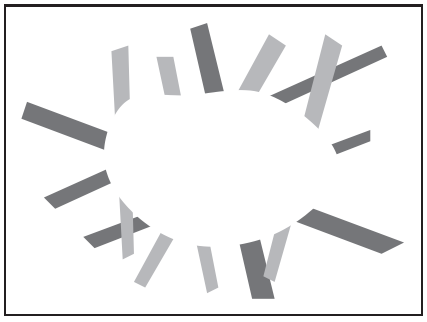


Figure 20.9. Sometimes, the visual system will "see" *subjective contours* without any associated change in brightness.

Figure 20.10. Perceived lightness depends more on local contrast at edges than on brightness across surfaces. Try covering the vertical edge in the middle of the figure with a pencil. This figure is an instance of the *Craik-O'Brien-Cornsweet illusion*.

As indicated above, people can detect differences in the brightness between two adjacent regions if the difference is at least 1% of the average brightness. This is an example of *Weber's law*, which states that there is a constant ratio between the *just noticeable differences* (jnd) in a stimulus and the magnitude of the stimulus:

$$\frac{\Delta I}{I} = k_1, \tag{20.2}$$

where I is the magnitude of the stimulus, ΔI is the magnitude of the just noticeable difference, and k_1 is a constant particular to the stimulus. Weber's law was postulated in 1846 and still remains a useful characterization of many perceptual effects. *Fechner's law*, proposed in 1860, generalized Weber's law in a way that allowed for the description of the strength of any sensory experience, not just jnd's:

$$S = k_2 \log(I), \tag{20.3}$$

where S is the perceptual strength of the sensory experience, I is the physical magnitude of the corresponding stimulus, and k_2 is a scaling constant specific to the stimulus. Current practice is to model the association between perceived and actual strength of a stimulus using a power function (*Stevens's law*):

$$S = k_3 I^b, \tag{20.4}$$

where S and I are as before, k_3 is another scaling constant, and b is an exponent specific to the stimulus. For a large number of perceptual quantities involving vision, $b < 1$. The CIE $L^*a^*b^*$ color space, described elsewhere, uses a modified Stevens's law representation to characterize perceptual differences between brightness values. Note that in the first two characterizations of the perceptual strength of a stimulus and in Stevens's Law when $b < 1$, changes in the stimulus when it has a small average magnitude create larger perceptual effects than do the same physical change in the stimulus when it has a larger magnitude.

The "laws" described above are not physical constraints on how perception operates. Rather, they are generalizations about how the perceptual system responds to particular physical stimuli. In the field of perceptual psychology, the quantitative study of the relationships between physical stimuli and their perceptual effects is called *psychophysics*. While psychophysical laws are empirically derived observations rather than mechanistic accounts, the fact that so many perceptual effects are well modeled by simple power functions is striking and may provide insights into the mechanisms involved.

20.2.2 Color

In 1666, Isaac Newton used prisms to show that apparently white sunlight could be decomposed into a *spectrum* of colors and that these colors could be recombined to produce light that appeared white. We now know that light energy is made up of a collection of photons, each with a particular wavelength. The *spectral distribution* of light is a measure of the average energy of the light at each wavelength. For natural illumination, the spectral distribution of light reflected off of surfaces varies significantly depending on the surface material. Characterizations of this spectral distribution can therefore provide visual information for the nature of surfaces in the environment.

Most people have a pervasive sense of color when they view the world. Color perception depends on the frequency distribution of light, with the visible spectrum for humans ranging from a wavelength of about 370 nm to a wavelength of about 730 nm (see Figure 20.11). The manner in which the visual systems derives a sense of color from this spectral distribution was first systematically examined in 1801 and remained extremely controversial for 150 years. The problem is that the visual system responds to patterns of spectral distribution very differently than patterns of luminance distribution.

"The history of the investigation of colour vision is remarkable for its acrimony."
—Richard Gregory (1997)

Even accounting for phenomena such as lightness constancy, distinctly different spatial distributions almost always look distinctly different. More importantly given that the purpose of the visual system is to produce descriptions of the distal stimulus given the proximal stimulus, perceived patterns of lightness correspond at least approximately to patterns of brightness over surfaces in the environment.

Figure 20.11. The visible spectrum. Wavelengths are in nanometers.

The same is not true of color perception. Many quite different spectral distributions of light can produce a sense of any specific color. Correspondingly, the sense that a surface is a specific color provides little direct information about the spectral distribution of light coming from the surface. For example, a spectral distribution consisting of a combination of light at wavelengths of 700 nm and 540 nm, with appropriately chosen relative strengths, will look indistinguishable from light at the single wavelength of 580 nm. (Perceptually indistinguishable colors with different spectral compositions are referred to as *metamers*.) If we see the color "yellow," we have no way of knowing if it was generated by one or the other of these distributions or an infinite family of other spectral distributions. For this reason, in the context of vision the term *color* refers to a purely perceptual quality, not a physical property.

There are two classes of photoreceptors in the human retina. *Cones* are involved in color perception, while *rods* are sensitive to light energy across the visible range and do not provide information about color. There are three types of cones, each with a different spectral sensitivity (Figure 20.12). *S-cones* respond to short wavelengths in the blue range of the visible spectrum. *M-cones* respond to wavelengths in the middle (greenish) region of the visible spectrum. *L-cones* respond to somewhat longer wavelengths covering the green and red portions of the visible spectrum.

While it is common to describe the three types of cones as *red*, *green*, and *blue*, this is neither correct terminology nor does it accurately reflect the cone sensitivities shown in Figure 20.12. The *L-cones* and *M-cones* are broadly tuned, meaning that they respond to a wide range of frequencies. There is also substantial overlap between the sensitivity curves of the three cone types. Taken together, these two properties mean that it is not possible to reconstruct an approximation

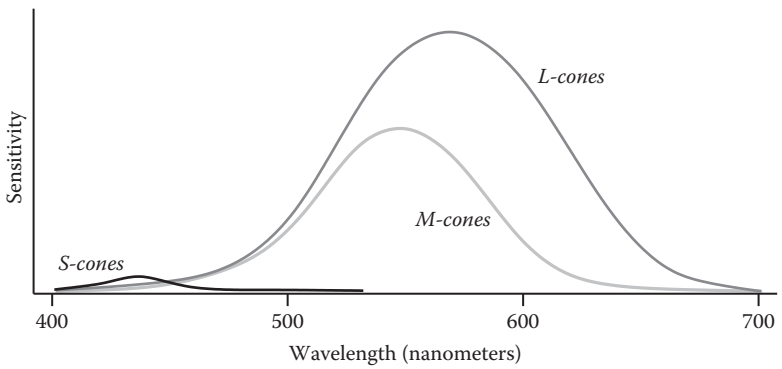


Figure 20.12. Spectral sensitivity of the *short*, *medium*, and *long* cones in the human retina.

to the original spectral distribution given the responses of the three cone types. This is in contrast to spatial sampling in the retina (and in digital cameras), where the receptors are narrowly tuned in their spatial sensitivity in order to be able to detect fine detail in local contrast.

The fact that there are are only three types of color sensitive photoreceptors in the human retina greatly simplifies the task of displaying colors on computer monitors and in other graphical displays. Computer monitors display colors as a weighted combination of three fixed-color distributions. Most often, the three colors are a distinct red, a distinct green, and a distinct blue. As a result, in computer graphics, color is often represented by a *red-green-blue* (RGB) triple, representing the intensities of red, green, and blue primaries needed to display a particular color. Three *basis colors* are sufficient to display most perceptible colors, since appropriately weighted combinations of three appropriately chosen colors can produce metamers for these perceptible colors.

There are at least two significant problems with the RGB color representation. The first is that different monitors have different spectral distributions for their red, green, and blue primaries. As a result, perceptually correct color rendition involves remapping RGB values for each monitor. This is, of course, only possible if the original RGB values satisfy some well-defined standard, which is often not the case. (See Chapter 19 for more information on this issue.) The second problem is that RGB values do not define a particular color in a way that corresponds to subjective perception. When we see the color "yellow," we do not have the sense that it is made up of equal parts of red and green light. Rather, it looks like a single color, with additional properties involving brightness and the "amount" of color. Representing color as the output of the S-cones, M-cones, and L-cones is no help either, since we have no more phenomenological sense of color as characterized by these properties than we do as characterized by RGB display properties.

There are two different approaches to characterizing color in a way that more closely reflects human perception. The various CIE color spaces aim to to be "perceptually uniform" so that the magnitude of the difference in the represented values of two colors is proportional to the perceived difference in color (Wyszecki & Stiles, 2000). This turns out to be a difficult goal to accomplish, and there have been several modifications to the CIE model over the years. Furthermore, while one of the dimensions of the CIE color spaces corresponds to perceived brightness, the other two dimensions that specify chromaticity have no intuitive meaning.

The second approach to characterizing color in a more natural manner starts with the observation that there are three distinct and independent properties that

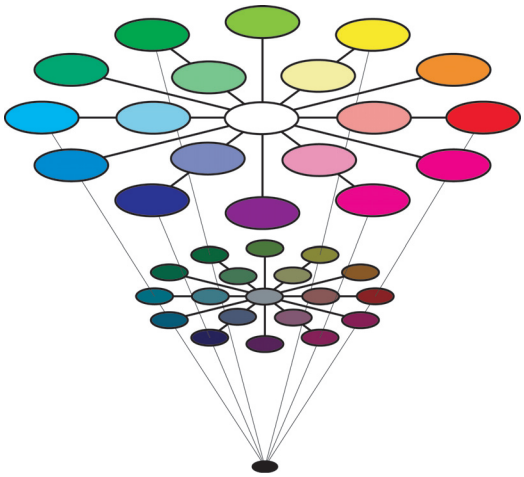


Figure 20.13. HSV color space. Hue varies around the circle, saturation varies with radius, and value varies with height.

dominate the subjective sense of color. *Lightness*, the apparent brightness of a surface, has already been discussed. *Saturation* refers to the purity or vividness of a color. Colors can range from totally unsaturated gray to partially saturated pastels to fully saturated "pure" colors. The third property, *hue*, corresponds most closely to the informal sense of the word "color" and is characterized in a manner similar to colors in the visible spectrum, ranging from dark violet to dark red. Figure 20.13 shows a plot of the hue-saturation-lightness (HSV) color space. Since the relationship between brightness and lightness is both complex and not well understood, HSV color spaces almost always use brightness instead of attempting to estimate lightness. Unlike wavelengths in the spectrum, however, hue is usually represented in a manner that reflects the fact that the extremes of the visible spectrum are actually similar in appearance (Figure 20.14). Simple transformations exist between RGB and HSV representations of a particular color value. As a result, while the HSV color space is motivated by perceptual considerations, it contains no more information than does an RGB representation.

Figure 20.14. Which color is closer to red: green or violet?

The hue-saturation-lightness approach to describing color is based on the spectral distribution at a single point and so only approximates the perceptual response to spectral distributions of light distributed over space. Color perception is subject to similar constancy and simultaneous contrast effects as is lightness/brightness, neither of which are captured in the RGB representation and as a result are not captured in the HSV representation. For an example of color constancy, look at a piece of white paper indoors under incandescent light and outdoors under direct sunlight. The paper will look "white" in both cases, even though incandescent light has a distinctly yellow hue and so the light reflected off of the paper will also have a yellow hue, while sunlight has a much more uniform color spectrum.

Another aspect of color perception not captured by either the CIE color spaces or HSV encoding is the fact that we see a small number of distinct colors when looking at a continuous spectrum of visible light (Figure 20.11) or in a naturally occurring rainbow. For most people, the visible spectrum appears to be divided into four to six distinct colors: red, yellow, green, and blue, plus perhaps light blue and purple. Considering non-spectral colors as well, there are only 11 basic color terms commonly used in English: *red*, *green*, *blue*, *yellow*, *black*, *white*, *gray*, *orange*, *purple*, *brown*, and *pink*. The partitioning of the intrinsically continuous space of spectral distributions into a relatively small set of perceptual categories associated with well-defined linguistic terms seems to be a basic property of perception, not just a cultural artifact (Berlin & Kay, 1969). The exact nature of the process, however, is not well understood.

20.2.3 Dynamic Range

Natural illumination varies in intensity over 6 orders of magnitude (Figure 20.15). The human vision system is able to operate over this full range of brightness levels. However, at any one point in time, the visual system is only able to detect variations in light intensity over a much smaller range. As the average brightness to which the visual system is exposed changes over time, the range of discriminable brightnesses changes in a corresponding manner. This effect is most obvious if we move rapidly from a brightly lit outdoor area to a very dark room. At first, we are able to see little. After a while, however, details in the room start to become apparent. The *dark adaptation* that occurs involves a number of physiological changes in the eye. It takes several minutes for significant dark adaptation to occur and 40 minutes or so for complete dark adaptation. If we then move back into the bright light, not only is vision difficult but it can actually be painful. *Light adaptation* is required before it is again possible to see clearly. Light adaptation

direct sunlight	10^5
indoor lighting	10^2
moonlight	10^{-1}
starlight	10^{-3}

Figure 20.15. Approximate luminance level of a white surface under different types of illumination in candelas per meter squared (cd/m^2). (Wandell, 1995).

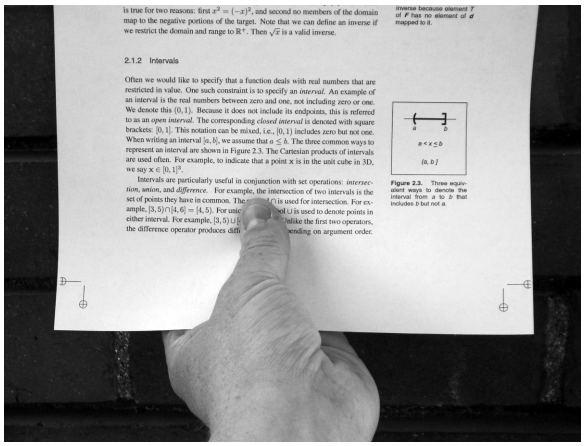
occurs much more quickly than dark adaptation, typically requiring less than a minute.

The two classes of photoreceptors in the human retina are sensitive to different ranges of brightness. The cones provide visual information over most of what we consider normal lighting conditions, ranging from bright sunlight to dim indoor lighting. The rods are only effective at very low light levels. *Photopic* vision involves bright light in which only the cones are effective. *Scotopic* vision involves dark light in which only the rods are effective. There is a range of intensities within which both cones and rods are sensitive to changes in light, which is referred to as *mesopic* conditions (see Chapter 21).

20.2.4 Field-of-View and Acuity

Each eye in the human visual system has a field-of-view of approximately 160° horizontal by 135° vertical. With binocular viewing, there is only partial overlap between the fields-of-view of the two eyes. This results in a wider overall field-of-view (approximately 200° horizontal by 135° vertical), with the region of overlap being approximately 120° horizontal by 135° vertical.

With normal or corrected-to-normal vision, we usually have the subjective experience of being able to see relatively fine detail wherever we look. This is an illusion, however. Only a small portion of the visual field of each eye is actually sensitive to fine detail. To see this, hold a piece of paper covered with normal-sized text at arm's length, as shown in Figure 20.16. Cover one eye with the


is true for two reasons: first $x^2 = (-x)^2$, and second no members of the domain map to the negative portions of the target. Note that we can define an inverse if we restrict the domain and range to $\mathbb{R}^+$. Then $\sqrt{x}$ is a valid inverse.

inverse because element 7 of F has no element of d mapped to it.

2.1.2 Intervals

Often we would like to specify that a function deals with real numbers that are restricted in value. One such constraint is to specify an *interval*. An example of an interval is the real numbers between zero and one, not including zero or one. We denote this $(0, 1)$. Because it does not include its endpoints, this is referred to as an *open interval*. The corresponding *closed interval* is denoted with square brackets: $[0, 1]$. This notation can be mixed, i.e., $[0, 1)$ includes zero but not one. When writing an interval $[a, b]$, we assume that $a \leq b$. The three common ways to represent an interval are shown in Figure 2.3. The Cartesian products of intervals are used often. For example, to indicate that a point x is in the unit cube in 3D, we say $\mathbf{x} \in [0, 1]^3$.

Intervals are particularly useful in conjunction with set operations: *intersection*, *union*, and *difference*. For example, the intersection of two intervals is the set of points they have in common. The ∩ is used for intersection. For example, $[3, 5) \cap [4, 6] = [4, 5)$. For unio ∪ is used to denote points in either interval. For example, $[3, 5) \cup [$ Unlike the first two operators, the difference operator produces diff pending on argument order.

$a < x \leq b$

$(a, b]$

Figure 2.3. Three equivalent ways to denote the interval from a to b that includes b but not a.

Figure 20.16. If you hold a page of text at arm's length and stare at your thumb, only the text near your thumb will be readable. *Photo by Peter Shirley.*

hand not holding the paper. While staring at your thumb and not moving your eye, note that the text immediately above your thumb is readable while the text to either side is not. High acuity vision is limited to a visual angle slightly larger than your thumb held at arm's length. We do not normally notice this because the eyes usually move frequently, allowing different regions of the visual field to be viewed at high resolution. The visual system then integrates this information over time to produce the subjective experience of the whole visual field being seen at high resolution.

There is not enough bandwidth in the human visual cortex to process the information that would result if there was a dense sampling of image intensity over the whole of the retina. The combination of variable density photoreceptor packing in the retina and a mechanism for rapid eye movements to point at areas of interest provides a way to simultaneously optimize acuity and field-of-view. Other animals have evolved different ways of balancing acuity and field-of-view that are not dependent on rapid eye movements. Some have only high acuity vision, but limited to a narrow field-of-view. Others have wide field-of-view vision, but limited ability to see detail.

The eye motions which focus areas of interest in the environment on the fovea are called *saccades*. Saccades occur very quickly. The time from a triggering stimulus to the completion of the eye movement is 150–200 ms. Most of this time is spent in the vision system planning the saccade. The actual motion takes 20 ms or so on average. The eyes are moving very quickly during a saccade, with the maximum rotational velocity often exceeding 500°/second. Between saccades, the eyes point toward an area of interest (*fixate*), taking 300 ms or so to acquire fine detail visual information. The mechanism by which multiple fixations are integrated to form an overall subjective sense of fine detail over a wide field of view is not well understood.

Figure 20.17 shows the variable packing density of cones and rods in the human retina. The cones, which are responsible for vision under normal lighting, are packed most closely at the *fovea* of the retina (Figure 20.17). When the eye is fixated at a particular point in the environment, the image of that point falls on the fovea. The higher packing density of cones at the fovea results in a higher sampling frequency of the imaged light (see Chapter 9) and hence greater detail in the sampled pattern. Foveal vision encompasses about 1.7°, which is the same visual angle as the width of your thumb held at arm's length.

While a version of Figure 20.17 appears in most introductory texts on human visual perception, it provides only a partial explanation for the neurophysiological limitations on visual acuity. The output of individual rods and cones is pooled in various ways by neural interconnects in the eye, before the information is shipped

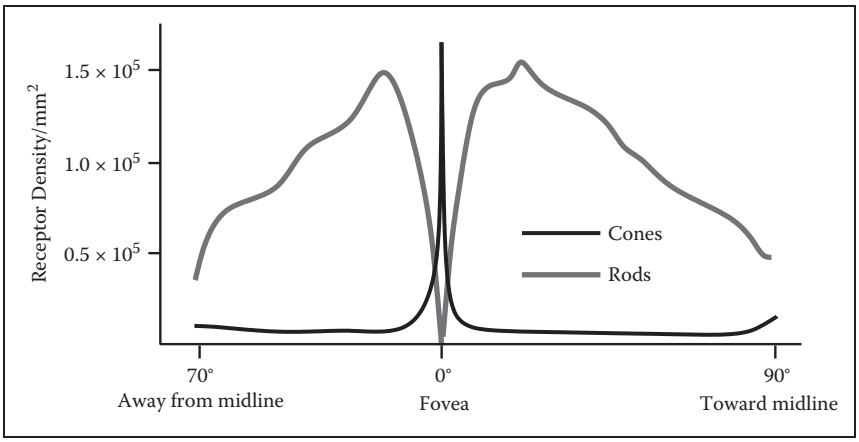


Figure 20.17. Density of rods and cone in the human retina (after Osterberg (1935)).

along the optic nerve to the visual cortex.[3] This pooling filters the signal provided by the pattern of incident illumination in ways that have important impacts on the patterns of light that are detectable. In particular, the farther away from the fovea, the larger the area over which brightness is averaged. As a consequence, spatial acuity drops sharply away from the fovea. Most figures showing rod and cone packing density indicate the location of the retinal *blind spot*, where the nerve bundle carrying optical information from the eye to the brain passes through the retina, and there is no sensitivity to light. By and large, the only practical impact of the blind spot on real-world perception is its use as an illusion in introductory perception texts, since normal eye movements otherwise compensate for the temporary loss of information.

As shown in Figure 20.17, the packing density of rods drops to zero at the center of the fovea. Away from the fovea, the rod density first increases and then decreases. One result of this is that there is no foveal vision when illumination is very low. The lack of rods in the fovea can be demonstrated by observing a night sky on a moonless night, well away from any city lights. Some stars will be so dim that they will be visible if you look at a point in the sky slightly to the side of the star, but they will disappear if you look directly at them. This occurs because when you look directly at these features, the image of the features falls only on the cones in the retina, which are not sufficiently light sensitive to detect the feature. Looking slightly to the side causes the image to fall on the more light-sensitive cones. Scotopic vision is also limited in acuity, in part because

[3] All of the cells in the optic nerve and almost all cells in the visual cortex have an associated retinal *receptive field*. Patterns of light hitting the retina outside of a cell's receptive field have no effect on the firing rate of that cell.

of the lower density of rods over much of the retina and in part because greater pooling of signals from the rods occurs in the retina in order to increase the light sensitivity of the visual information passed back to the brain.

20.2.5 Motion

When reading about visual perception and looking at static figures on a printed page, it is easy to forget that motion is pervasive in our visual experience. The patterns of light that fall on the retina are constantly changing due to eye and body motion and the movement of objects in the world. This section covers our ability to detect visual motion. Section 20.3.4 describes how visual motion can be used to determine geometric information about the environment. Section 20.4.3 deals with the use of motion to guide our movement through the environment.

The detectability of motion in a particular pattern of light falling on the retina is a complex function of speed, direction, pattern size, and contrast. The issue is further complicated because simultaneous contrast effects occur for motion perception in a manner similar to that observed in brightness perception. In the extreme case of a single small pattern moving against a contrasting, homogenous background, perceivable motion requires a rate of motion corresponding to 0.2°–0.3°/second of visual angle. Motion of the same pattern moving against a textured pattern is detectable at about a tenth this speed.

With this sensitivity to retinal motion, combined with the frequency and velocity of saccadic eye movements, it is surprising that the world usually appears stable and stationary when we view it. The vision system accomplishes this in three ways. Contrast sensitivity is reduced during saccades, reducing the visual effects generated by these rapid changes in eye position. Between saccades, a variety of sophisticated and complex mechanisms adjust eye position to compensate for head and body motion and the motion of objects of interest in the world. Finally, the visual system exploits information about the position of the eyes to assemble a mosaic of small patches of high-resolution imagery from multiple fixations into a single, stable whole.

The motion of straight lines and edges is ambiguous if no endpoints or corners are visible, a phenomenon referred to as the *aperture problem* (Figure 20.18). The aperture problem arises because the component of motion parallel to the line or edge does not produce any visual changes. The geometry of the real world is sufficiently complex that this rarely causes difficulties in practice, except for intentional illusions such as barber poles. The simplified geometry and texturing found in some computer graphics renderings, however, has the potential to introduce inaccuracies in perceived motion.

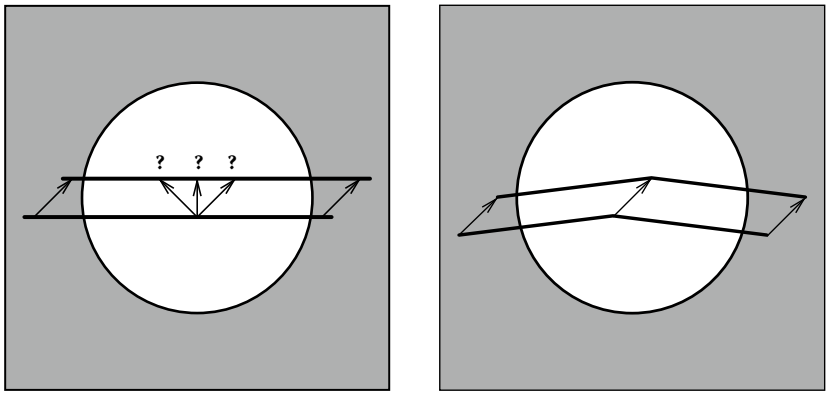


Figure 20.18. The aperture problem: (a) If a straight line or edge moves in such a way that its endpoints are hidden, the visual information is not sufficient to determine the actual motion of the line. (b) 2D motion of a line is unambiguous if there are any corners or other distinctive markings on the line.

Real-time computer graphics, film, and video would not be possible without an important perceptual phenomena: discontinuous motion, in which a series of static images are visible for discrete intervals in time and then move by discrete intervals in space, can be nearly indistinguishable from continuous motion. The effect is called *apparent motion* to highlight that the appearance of continuous motion is an illusion.

Figure 20.19 illustrates the difference between continuous motion, which is typical of the real world, and apparent motion, which is generated by almost all dynamic image display devices. The motion plotted in Figure 20.19 (b) consists of an average motion comparable to that shown in Figure 20.19 (a), modulated by a high space-time frequency that accounts for the alternation between a stationary pattern and one that moves discontinuously to a new location. Apparent perception of continuous motion occurs because the visual system is insensitive to the high-frequency component of the motion.

A compelling sense of apparent motion occurs when the rate at which individual images appear is above about 10 Hz, as long as the positional changes between successive images is not too great. This rate is not fast enough, however, to produce a satisfying sense of continuous motion for most image display devices. Almost all such devices introduce brightness variation as one image is switched to the next. In well-lit conditions, the human visual system is sensitive to this varying brightness for rates of variations up to about 80 Hz. In lower light, detectability is present up to about 40 Hz. When the rate of alternating brightness is sufficiently high, *flicker fusion* occurs and the variation is no longer visible.

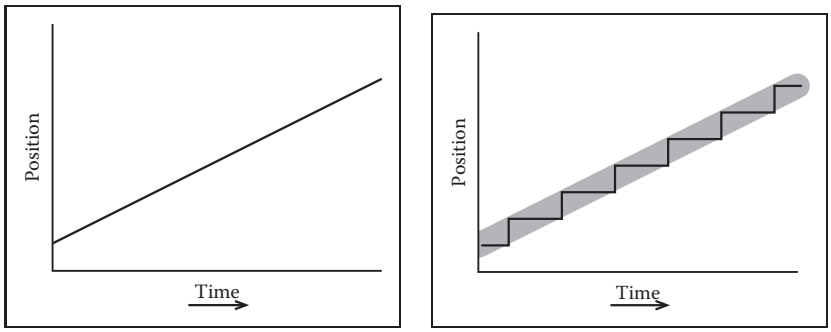


Figure 20.19. (a) Continuous motion. (b) Discontinuous motion with the same average velocity. Under some circumstances, the perception of these two motion patterns may be similar.

To produce a compelling sense of visual motion, an image display must therefore satisfy two separate constraints:

- images must be updated at a rate $\geq$ 10 Hz;
- any flicker introduced in the process of updating images must occur at a rate $\geq$ 60–80 Hz.

One solution is to require that the image update rate be greater than or equal to 60–80 Hz. In many situations, however, this is simply not possible. For computer graphics displays, the frame computation time is often substantially greater than 12–15 msec. Transmission bandwidth and limitations of older monitor technologies limit normal broadcast television to 25–30 images per second. (Some HDTV formats operate at 60 images/sec.) Movies update images at 24 frames/second due to exposure time requirements and the mechanical difficulties of physically moving film any faster than that.

Different display technologies solve this problem in different ways. Computer displays refresh the displayed image at $\sim$70–80 Hz, regardless of how often the contents of the image change. The term *frame rate* is ambiguous for such displays, since two values are required to characterize this display: *refresh rate*, which indicates the rate at which the image is redisplayed and *frame update rate*, which indicates the rate at which new images are generated for display. Standard non-HDTV broadcast television uses a refresh rate of 60 Hz (NTSC, used in North America and some other locations) or 50 Hz (PAL, used in most of the rest of the world). The frame update rate is half the refresh rate. Instead of displaying each new image twice, the display is *interlaced* by dividing alternating horizontal image lines into even and odd *fields* and alternating the display of these even and

odd fields. Flicker is avoided in movies by using a mechanical shutter to blink each frame of the film three times before moving to the next frame, producing a refresh rate of 72 Hz while maintaining the frame update rate of 24 Hz.

The use of apparent motion to simulate continuous motion occasionally produces undesirable artifacts. Best known of these is the *wagon wheel illusion* in which the spokes of a rotating wheel appear to revolve in the opposite direction from what would be expected given the translational motion of the wheel. The wagon wheel illusion is an example of temporal aliasing. Spokes, or other spatially periodic patterns on a rotating disk, produce a temporally periodic signal for viewing locations that are fixed with respect to the center of the wheel or disk. Fixed frame update rates have the effect of sampling this temporally periodic signal in time. If the temporal frequency of the sampled pattern is too high, undersampling results in an aliased, lower temporal frequency appearing when the image is displayed. Under some circumstances, this distortion of temporal frequency causes a spatial distortion in which the wheel appears to move backwards. Wagon wheel illusions are more likely to occur with movies than with video, since the temporal sampling rate is lower.

Problems can also occur when apparent motion imagery is converted from one medium to another. This is of particular concern when 24 Hz movies are transferred to video. Not only does a non-interlaced format need to be translated to an interlaced format, but there is no straightforward way to move from 24 frames per second to 50 or 60 fields per second. Some high-end display devices have the ability to partially compensate for the artifacts introduced when film is converted to video.

20.3 Spatial Vision

One of the critical operations performed by the visual system is the estimation of geometric properties of the visible environment, since these are central to determining information about objects, locations, and events. Vision has sometimes been described as *inverse optics*, to emphasize that one function of the visual system is to invert the image formation process in order to determine the geometry, materials, and lighting in the world that produced a particular pattern on light on the retina. The central problem for a vision system is that properties of the visible environment are confounded in the patterns of light imaged on the retina. Brightness is a function of both illumination and reflectance, and can depend on environmental properties across large regions of space due to the complexities of light transport. Image locations of a projected environmental location at best can

be used to constrain the position of that location to a half-line. As a consequence, it is rarely possible to uniquely determine the nature of the world that produced a particular imaged pattern of light.

Determining *surface layout*—the location and orientation of visible surfaces in the environment—is thought to be a key step in human vision. Most discussions of how the vision system extracts information about surface layout from the patterns of light it receives divide the problem into a set of *visual cues*, with each cue describing a particular visual pattern which can be used to infer properties of surface layout along with the needed rules of inference. Since surface layout can rarely be determined accurately and unambiguously from vision alone, the process of inferring surface layout usually requires additional, nonvisual information. This can come from other senses or assumptions about what is likely to occur in the real world.

Visual cues are typically categorized into four categories. *Ocularmotor cues* involve information about the position and focus of the eyes. *Disparity cues* involve information extracted from viewing the same surface point with two eyes, beyond that available just from the positioning of the eyes. *Motion cues* provide information about the world that arises from either the movement of the observer or the movement of objects. *Pictorial cues* result from the process of projecting 3D surface shapes onto a 2D pattern of light that falls on the retina. This section deals with the visual cues relevant to the extraction of geometric information about individual points on surfaces. More general extraction of location and shape information is covered in Section 20.4.

20.3.1 Frames of Reference and Measurement Scales

Descriptions of the location and orientation of points on a visible surface must be done within the context of a particular frame of references that specifies the origin, orientation, and scaling of the coordinate system used in representing the geometric information. The human vision system uses multiple frames of reference, partially because of the different sorts of information available from different visual cues and partly because of the different purposes to which the information is put (Klatzky, 1998). *Egocentric* representations are defined with respect to the viewer's body. They can be subdivided into coordinate systems fixed to the eyes, head, or body. *Allocentric* representations, also called *exocentric* representations, are defined with respect to something external to the viewer. Allocentric frames of reference can be local to some configuration of objects in the environment or can be globally defined in terms of distinctive locations, gravity, or geographic properties.