6

PERCEPTION OF BRIGHTNESS AND COLOR

In previous chapters we addressed the problems of separating occluding objects, generating suitable descriptions for them, and recognizing them, *assuming* that the boundaries of the objects were known. We now turn our attention to early processing of visual information and the important problems associated with obtaining such boundaries. These processes include *figure-ground* separation or *scene segmentation*—that is, identifying the regions of an image that correspond to physical objects of interest.

For controlled scenes of polyhedral objects, scene segmentation is simple, at least in principle. The surfaces of the objects can be arranged to have continuous brightness values that are distinctly different from the backgrounds. However, in general, the images of our daily experience are much more complex. An object surface is not always characterized by a simple uniformity of, say, intensity or color. Rather, the patterns of intensity and color over one object surface seem to be different from those of others. For example, a grass field is not uniformly green, but the pattern of elongated blades in it is different from the patterns caused by waves in a nearby lake. We call such patterns the *texture* of a surface.

Human vision uses many sources of information to aid in the segmentation process. An important source is our ability to extract relative three-dimensional positions of objects using two eyes or even from a single view using the variations in the surface brightness and texture, and other cues. Many of these operations are poorly understood, and only weak mechanisms for machines to perform the same operations have been developed. Segmentation using different sources of information is discussed in Chapters 7, 8, and 9. In this chapter we first consider our perception of brightness and color, which in itself is a complex process.

6.1 BRIGHTNESS AND COLOR

Measurement of the brightness or the color of an isolated single pixel is relatively straightforward. However, our perception of the brightness (or color) of a part of an image is strongly influenced by the brightness (color) values of the neighboring parts. For example, see Fig. 6-1, which consists of a small square in each of the three larger squares. The leftmost small square *appears* to be much brighter than the rightmost small square, the center one having an intermediate value. In fact, the three small squares have the same intensity. (A similar effect is also observed for color perception.)

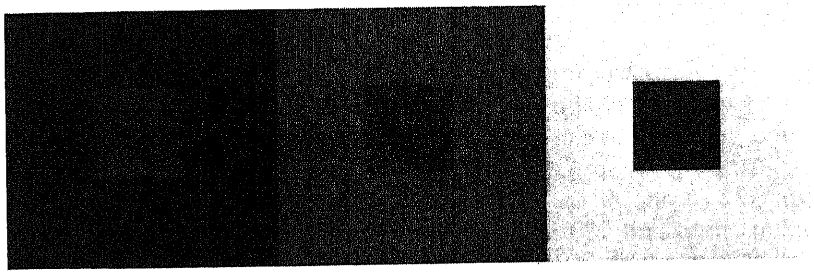


Figure 6-1: An example illustrating simultaneous contrast

Another important effect is the relative constancy of the perceived *lightness* and color of a surface under different illumination conditions. A grey piece of paper taken from inside a poorly lit room to the bright outside sunshine still appears grey, even though the reflected light from it is several orders of magnitude larger than from a white piece of paper indoors. Only a small part of the change is explained by the change in the size of the iris (the lens opening) of our eyes. Similarly, a surface of a certain color retains its perceived color under most conditions even when the color of illumination changes. These effects, known as brightness- and color-constancy phenomena, are clearly desirable properties for a perceiving machine to have. (The brightness and color constancies are only approximate, and they do break down under extreme conditions such as very dim light.)

6.1.1 The Human Eye

The human visual system may be simply thought of as consisting of two eyes that image the environment and pass the information to the cortex in the brain. Besides forming the image, some very important processing also takes place in the eye itself, before the information reaches the cortex. All descriptions presented here are highly simplified; much remains unknown about the human visual system.

A simple model of the human eye consists of a lenslike device that forms an image on a *retina.* The retinal surface is nearer to being spherical than planar, as for ideal camera systems of Chapter 3. The opening of the lens is adjusted by an iris. However, adaptability to wide variations in light intensity, say between night and day viewing, is accomplished by the variations in the sensitivity of the sensing elements, called *receptors.*

The human retina contains two types of receptors, called rods and cones. The cones are responsible for color vision, but the rods have a higher sensitivity to light. The spatial distribution of these receptors is not uniform over the retina. Most of the cones are concentrated in a central region known as the fovea. Thus, we have poor color vision in the dark and at the periphery of the field of view.

Under normal viewing, the fovea focuses on different parts of the scene in an apparently pseudo-random way, but concentrating on "interesting" parts. The points of focus seem to be the busy areas of the scene, but it is likely that the choices are based on more complex processing. The spatial resolution of the human system is extremely high (1/2 second of arc at the fovea). Each eye is estimated to have 6 million cones and 120 million rods. The light sensitivity of the rods is also amazingly high; it is estimated that a single photon of light activates a rod, and a few photons suffice for reliable detection of the presence of light. Details of the human visual sensory system may be found in [1, 2].

6.1.2 Local Measurements

Local measurement of the brightness of a small area of an image is made by a sensor whose response is proportional to the average light intensity in a small neighborhood. For the human system, these local measurements may be considered to be the responses of the individual rods and cones. The response, for the human system and also for many electronic sensors, is roughly proportional to the logarithm of the image intensity. The perceived brightness, however, is dependent on the intensities of the surrounding pixels and possibly even on the global

context.

Perception of local color is more complex. Humans are able to distinguish between two homogeneous surfaces of different colors by three independent attributes: intensity, hue and saturation. *Intensity* is the attribute also used for noncolor or achromatic images (commonly called black-and-white or grey-level images). *Hue* is the common English language sense of color; for example, red, green, and blue are three different hues. *Saturation* is a measure of the amount of white mixed with a pure hue. Our color sensation can be viewed as requiring three independent parameters for its characterization.

For a monochromatic source of light, our perception of hue is directly dependent on the wavelength. However, in our normal viewing environments, objects of a certain hue, say red, do not reflect just the light of a single wavelength, but rather a broad spectrum that has a distribution different from that of the white light. Different hues are also obtained by mixing colors; for example, green hue is obtained by mixing blue and yellow, even though the mixed light may contain no green-wavelength component at all. It is known, from empirical data, that all colors distinguished by humans can be synthesized by a mixture of three *primary* colors. The commonly used primary colors are red, green, and blue, but the choice of primaries is not unique. Note that the sufficiency of the three primaries corresponds to the characterization of color by three independent components. The primary colors need not be monochromatic but may have a wide-spectrum distribution of light intensity. Broadcast television relies on standard primary color spectra in the three light sources of the receivers. The three primary distributions have peaks at the nominal red, green, and blue wavelengths and overlap.

The receptors in the human retina, or the color TV cameras, do not directly measure the perceived attributes of intensity, hue, and saturation. Instead, the measured attributes correspond to the intensity of the three primary color components in the viewed surface. For a color camera, the three measurements are from three sensors that respond to light according to the response curves for the standard primaries. Let the three intensity measurements be R, G, and B, corresponding to the nominal red, green and blue primaries. Similar measurements could be obtained from a noncolor camera by using three different color filters in front of the lens. For human vision, the cone receptors are believed to be of three types with three different spectral responses, as shown in Fig. 6-2. (This figure is not intended to give highly accurate psychophysical data, but only an illustrative estimate.)

There are many theories as to how the attributes of intensity, hue, a saturation are inferred from the brightness values for the three primaries—that is, the R, G, and B measurements. A simple method is

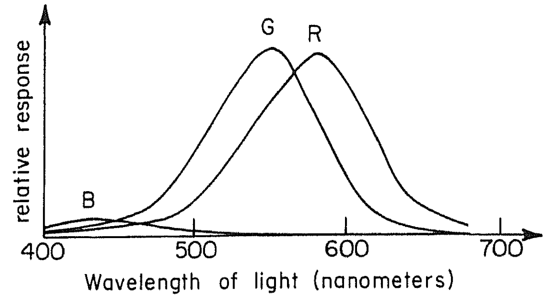


Figure 6-2: Relative responses of three color sensors in the human retina (illustrative data only)

presented here. The intensity I is given simply by a weighted sum of the three components:

$$I = c_1R + c_2G + c_3B \tag{6-1}$$

The constants c_1, c_2, and c_3 depend on the choice of the primary colors and can be determined by measurements made on a white surface. (See [3] for details.) For a camera calibrated according to United States TV broadcast standards (known as N.T.S.C., the National Television Systems Committee Standards), the three constants in order are 0.299, 0.587, and 0.114.

To compute hue and saturation components, it is convenient to introduce two additional attributes T_1 and T_2 defined as follows:

$$T_1 = \frac{R}{R + G + B} \tag{6-2}$$

$$T_2 = \frac{G}{R + G + B} \tag{6-3}$$

The $T_1 - T_2$ plane contains all of the chromaticity (non-brightness related) information. Note that the transformation is undefined when $R = G = B = 0$; a detailed discussion of this singularity may be found in [4]. The three primary colors R, G, and B are at the three points in the $T_1 - T_2$ plane as shown in Fig. 6-3. The triangle connecting these points is known as the *color triangle*; all observed colors must fall within this triangle. White is represented by the point $W(1/3,\ 1/3)$. For a given color represented by the point P in the color triangle, its distance Y from the point W, as shown in Fig. 6-3, determines its saturation and the angle θ of the line joining P to W with the T_1 axis determines the hue.

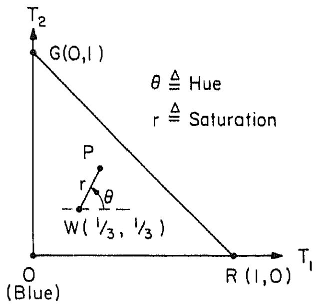


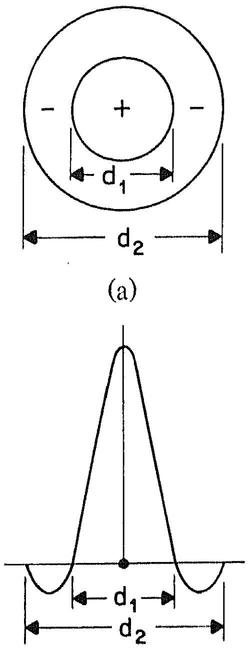
Figure 6-3: A color triangle

6.2 LATERAL INHIBITION AND LIGHTNESS COMPUTATION

In the human system, each receptor output does not go to the cortex in the brain. It is estimated that the outputs from about 120 million receptors in each eye are transmitted along only a million nerve fibers to the brain. This reduction in the member of output signals is a result of very important, and only partially understood, processing at the retina. The retina contains layers of cells in addition to the receptor cells and is capable of extensive computation. Only very simple models of processing at the retinal level are given below.

The most common model of interaction between receptors at the retina is that of *lateral inhibition*. Outputs from a small neighborhood of receptors, known as a receptive field, are combined by forming a weighted sum of the individual outputs. The receptors in a central excitatory neighborhood contribute positively and those in a larger surrounding inhibitory neighborhood have negative weights. Figure 6-4(a) shows a schematic receptive field and Fig. 6-4(b) shows the weights of the receptors along a diameter. The lateral inhibition operation may also be viewed as performing a 2-D bandpass filtering of the input signal. (The filter is the spatial Fourier transform of the lateral inhibition weighting function.) Such filtering explains the observed phenomenon of Mach bands, shown in Fig. 6-5. In this figure, near the center of the strip, the intensity changes linearly from dark to bright, but the reader should see a thin dark strip on the left side and a thin light strip on the right side of the area of change.

Lateral inhibition is also commonly used to explain the



(a)

(b)

Figure 6-4: Lateral-inhibition operator: (a) receptive field, (b) weighted contributions of the receptors

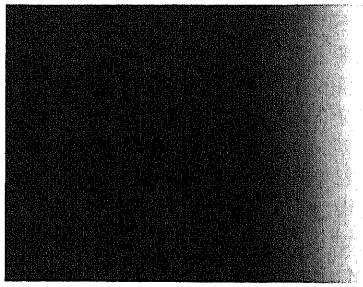


Figure 6-5: An example of Mach bands effect

phenomenon of simultaneous contrast—that is, the dependence of perceived brightness on the surround, as in Fig. 6-1 (for example, see [1]). Modeling the receptor response as logarithmic, the lateral inhibition operation can be viewed as computing the ratio of central brightness with peripheral brightness. However, this only explains perceived brightness near the boundaries and not the perceived uniform brightness of each square in Fig. 6-1. Similar arguments also apply to

lateral inhibition explaining the phenomenon of brightness constancy, where perceived brightness is again computed relative to those of other surrounding surfaces. We will later see, in Chapter 7, that an operation like lateral-inhibition is also useful for edge detection.

A better explanation for brightness constancy and simultaneous contrast is given by the *retinex* theory suggested by Land and McCann [5, 6]. The light intensity at the receptor, say $p'(x)$, is a product of the incident light, $s'(x)$, and surface reflectivity, $r'(x)$. Given an image and hence $p'(x)$, we seek to compute $r'(x)$ without a priori knowledge of $s'(x)$. Taking logarithms of p', s', and r', we get

$$p(x) = s(x) + r(x) \tag{6-4}$$

where $p(x) = \log p'(x)$, and so on.

Taking derivatives, we get

$$d(x) = D(p(x)) = D(s(x)) + D(r(x)) \tag{6-5}$$

Now, if we assume that the incident illumination $s(x)$ changes smoothly across the image whereas the reflectivity $r(x)$ changes abruptly at the object boundaries, then $D(s(x))$ is finite and $D(r(x))$ consists of infinite impulses at the object boundaries (see Fig. 6-6.) $D(r(x))$ can be easily separated from $D(s(x))$. In the discrete case, the derivative is approximated by a differential, and thresholding $d(x)$ gives $D(r(x))$. The reflectivity (or lightness) $r(x)$ can now be recovered by integrating the thresholded differential, except for a constant term. Note that in Fig. 6-6, lightness $l(x)$ has negative values. A constant of integration must be added to get physically realizable values.

Land and McCann's theory was formulated for continuous domain processing along a line. For 2-D processing, they suggest use of randomly selected lines. Horn has generalized the theory for 2-D directly by using a Laplacian operation instead of the derivative and approximating the Laplacian for discrete images [7]. Marr has given a neurological model for this computation in the human retina [8].

Computation of the lightness of a surface, independent of the illumination, explains brightness constancy. Simultaneous contrast is explained by perception of the differences in the logarithms of the lightness. Color constancy is explained by assuming that the lightness for the three components is first computed independently and that the ratios of the lightness values, rather than the intensity values, are used for chromatic perception. However, this model applies only to surfaces with uniform lightness (sometimes called Mondrian surfaces). It also fails to explain some of the other observed effects of human perception,

Vague.

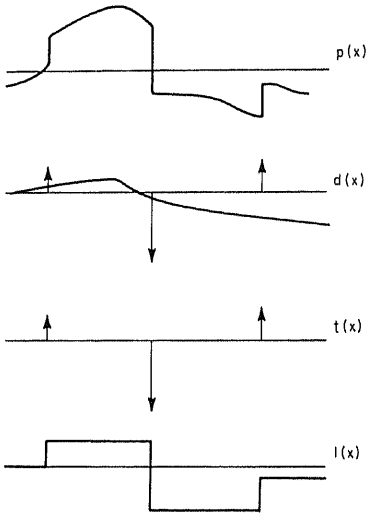


Figure 6-6: Steps in computing lightness

such as Mach bands.

Owing to the limitations of the human visual models, their use in machine processing is not widespread. Instead, only the single pixel brightness and color are frequently used for further processing.

6.3 SUMMARY

The perception of apparently simple properties, such as lightness and color, is seen to be rather complex and only partially understood. The major effects are due to the interaction with the neighboring cells. Two models, those of lateral inhibition and retinex, have been described.

REFERENCES

[1] T. N. Cornsweet, *Visual Perception*, Academic Press, New York, 1970.
[2] I. Rock, *Introduction to Perception*, Macmillan, New York, 1975.
[3] W. K. Pratt, *Digital Image Processing*, John Wiley & Sons, New York, 1978.
[4] J. Kender, "Saturation, Hue, and Normalized Color: Calculation, Digitization Effects and Use," Carnegie-Mellon University,

Department of Computer Science, Pittsburgh, November 1976.
[5] E. Land, "Experiments in Color Vision," *Scientific American*, May 1959, pp. 84-99.
[6] E. Land and J. J. McCann, "Lightness Theory," *Journal of the Optical Society*, Vol. 61, No. 1, January 1971, pp. 1-11.
[7] B. K. P. Horn, "Determining Lightness from an Image," *Computer Graphics and Image Processing*, Vol. 3, 1974, pp. 277-299.
[8] D. Marr, "The Computation of Lightness by the Primate Retina," *Vision Research*, Vol. 14, 1974, pp. 1377-1388.