

## 2.1 ELEMENTS OF VISUAL PERCEPTION

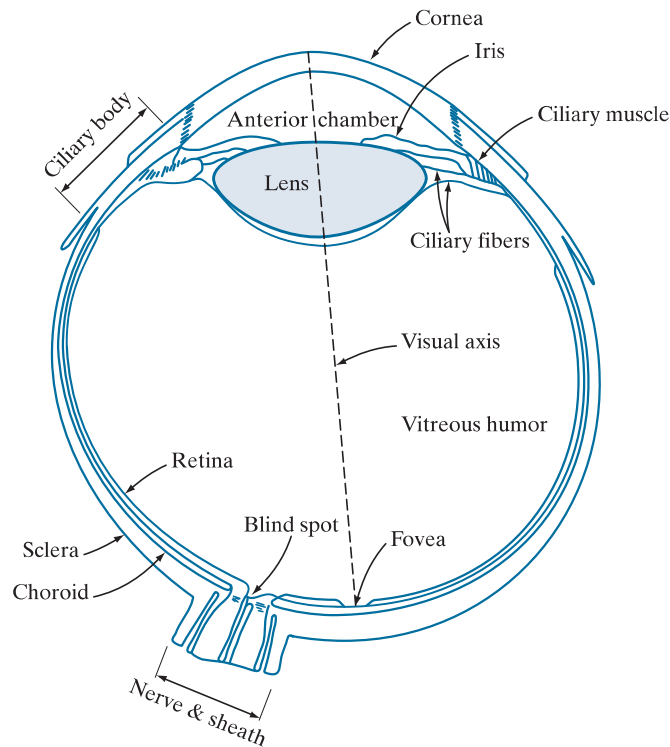
Although the field of digital image processing is built on a foundation of mathematics, human intuition and analysis often play a role in the choice of one technique versus another, and this choice often is made based on subjective, visual judgments. Thus, developing an understanding of basic characteristics of human visual perception as a first step in our journey through this book is appropriate. In particular, our interest is in the elementary mechanics of how images are formed and perceived by humans. We are interested in learning the physical limitations of human vision in terms of factors that also are used in our work with digital images. Factors such as how human and electronic imaging devices compare in terms of resolution and ability to adapt to changes in illumination are not only interesting, they are also important from a practical point of view.

### STRUCTURE OF THE HUMAN EYE

Figure 2.1 shows a simplified cross section of the human eye. The eye is nearly a sphere (with a diameter of about 20 mm) enclosed by **three membranes: the cornea and sclera outer cover; the choroid; and the retina**. The cornea is a tough, transparent tissue that covers the anterior surface of the eye. Continuous with the cornea, the sclera is an opaque membrane that encloses the remainder of the optic globe.

The choroid lies directly below the sclera. This membrane contains a network of blood vessels that serve as the major source of nutrition to the eye. Even superficial

**FIGURE 2.1**  
Simplified  
diagram of a  
cross section of  
the human eye.



injury to the choroid can lead to severe eye damage as a result of inflammation that restricts blood flow. The choroid coat is heavily pigmented, which helps reduce the amount of extraneous light entering the eye and the backscatter within the optic globe. At its anterior extreme, the choroid is divided into the *ciliary body* and the *iris*. The latter contracts or expands to control the amount of light that enters the eye. The central opening of the iris (the *pupil*) varies in diameter from approximately 2 to 8 mm. The front of the iris contains the visible pigment of the eye, whereas the back contains a black pigment.

The *lens* consists of concentric layers of fibrous cells and is suspended by fibers that attach to the ciliary body. It is composed of 60% to 70% water, about 6% fat, and more protein than any other tissue in the eye. The lens is colored by a slightly yellow pigmentation that increases with age. In extreme cases, excessive clouding of the lens, referred to as *cataracts*, can lead to poor color discrimination and loss of clear vision. The lens absorbs approximately 8% of the visible light spectrum, with higher absorption at shorter wavelengths. Both infrared and ultraviolet light are absorbed by proteins within the lens and, in excessive amounts, can damage the eye.

The innermost membrane of the eye is the *retina*, which lines the inside of the wall's entire posterior portion. When the eye is focused, light from an object is imaged on the retina. Pattern vision is afforded by discrete light receptors distributed over the surface of the retina. There are two types of receptors: *cones* and *rods*. There are between 6 and 7 million cones in each eye. They are located primarily in the central portion of the retina, called the *fovea*, and are highly sensitive to color. Humans can resolve fine details because each cone is connected to its own nerve end. Muscles rotate the eye until the image of a region of interest falls on the fovea. Cone vision is called *photopic* or *bright-light* vision.

The number of rods is much larger: Some 75 to 150 million are distributed over the retina. The larger area of distribution, and the fact that several rods are connected to a single nerve ending, reduces the amount of detail discernible by these receptors. Rods capture an overall image of the field of view. They are not involved in color vision, and are sensitive to low levels of illumination. For example, objects that appear brightly colored in daylight appear as colorless forms in moonlight because only the rods are stimulated. This phenomenon is known as *scotopic* or *dim-light* vision.

Figure 2.2 shows the density of rods and cones for a cross section of the right eye, passing through the region where the optic nerve emerges from the eye. The absence of receptors in this area causes the so-called *blind spot* (see Fig. 2.1). Except for this region, the distribution of receptors is radially symmetric about the fovea. Receptor density is measured in degrees from the visual axis. Note in Fig. 2.2 that cones are most dense in the center area of the fovea, and that rods increase in density from the center out to approximately 20° off axis. Then, their density decreases out to the periphery of the retina.

The fovea itself is a circular indentation in the retina of about 1.5 mm in diameter, so it has an area of approximately 1.77 mm<sup>2</sup>. As Fig. 2.2 shows, the density of cones in that area of the retina is on the order of 150,000 elements per mm<sup>2</sup>. Based on these figures, the number of cones in the fovea, which is the region of highest acuity

from the sensors by motion alone; they also require extensive computer processing. A 3-D digital volume consisting of stacked images is generated as the object is moved in a direction perpendicular to the sensor ring. Other modalities of imaging based on the CAT principle include magnetic resonance imaging (MRI) and positron emission tomography (PET). The illumination sources, sensors, and types of images are different, but conceptually their applications are very similar to the basic imaging approach shown in Fig. 2.14(b).

### IMAGE ACQUISITION USING SENSOR ARRAYS

Figure 2.12(c) shows individual sensing elements arranged in the form of a 2-D array. Electromagnetic and ultrasonic sensing devices frequently are arranged in this manner. This is also the predominant arrangement found in digital cameras. A typical sensor for these cameras is a CCD (charge-coupled device) array, which can be manufactured with a broad range of sensing properties and can be packaged in rugged arrays of  $4000 \times 4000$  elements or more. CCD sensors are used widely in digital cameras and other light-sensing instruments. The response of each sensor is proportional to the integral of the light energy projected onto the surface of the sensor, a property that is used in astronomical and other applications requiring low noise images. Noise reduction is achieved by letting the sensor integrate the input light signal over minutes or even hours. Because the sensor array in Fig. 2.12(c) is two-dimensional, its key advantage is that a complete image can be obtained by focusing the energy pattern onto the surface of the array. Motion obviously is not necessary, as is the case with the sensor arrangements discussed in the preceding two sections.

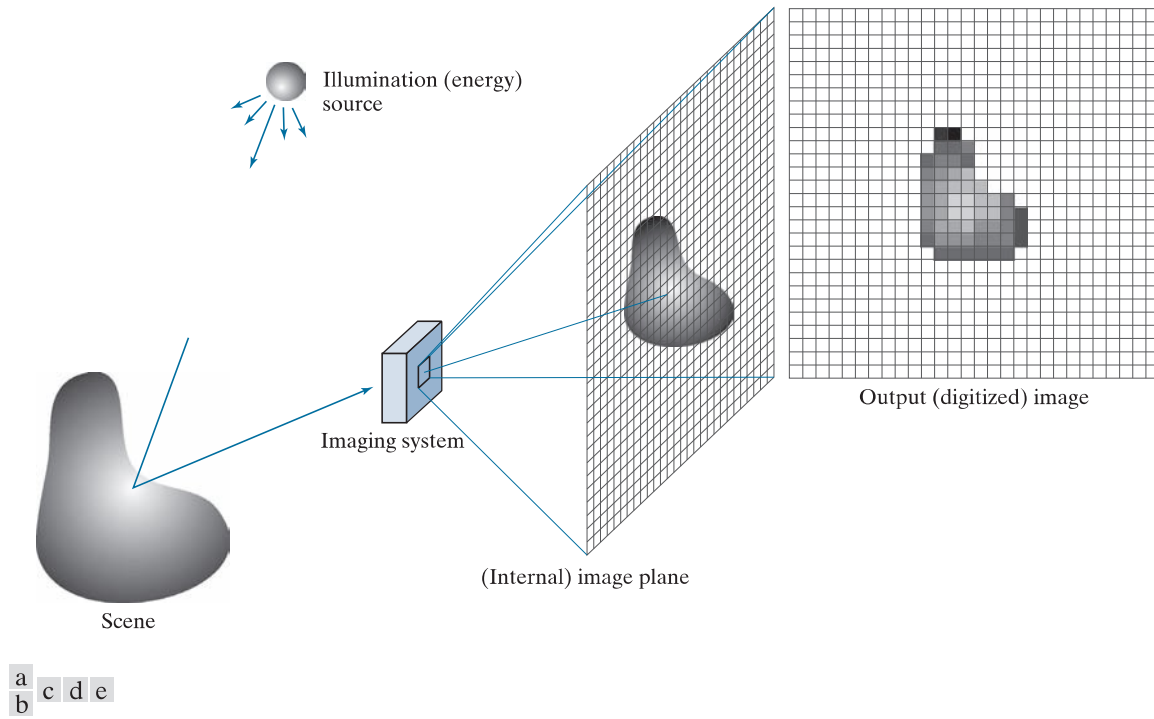
Figure 2.15 shows the principal manner in which array sensors are used. This figure shows the energy from an illumination source being reflected from a scene (as mentioned at the beginning of this section, the energy also could be transmitted through the scene). The first function performed by the imaging system in Fig. 2.15(c) is to collect the incoming energy and focus it onto an image plane. If the illumination is light, the front end of the imaging system is an optical lens that projects the viewed scene onto the focal plane of the lens, as Fig. 2.15(d) shows. The sensor array, which is coincident with the focal plane, produces outputs proportional to the integral of the light received at each sensor. Digital and analog circuitry sweep these outputs and convert them to an analog signal, which is then digitized by another section of the imaging system. The output is a digital image, as shown diagrammatically in Fig. 2.15(e). Converting images into digital form is the topic of Section 2.4.

### A SIMPLE IMAGE FORMATION MODEL

As introduced in Section 1.1, we denote images by two-dimensional functions of the form  $f(x, y)$ . The value of  $f$  at spatial coordinates  $(x, y)$  is a scalar quantity whose physical meaning is determined by the source of the image, and whose values are proportional to energy radiated by a physical source (e.g., electromagnetic waves). As a consequence,  $f(x, y)$  must be nonnegative<sup>†</sup> and finite; that is,

<sup>†</sup> Image intensities can become negative during processing, or as a result of interpretation. For example, in radar images, objects moving toward the radar often are interpreted as having negative velocities while objects moving away are interpreted as having positive velocities. Thus, a velocity image might be coded as having both positive and negative values. When storing and displaying images, we normally scale the intensities so that the smallest negative value becomes 0 (see Section 2.6 regarding intensity scaling).

In some cases, the source is imaged directly, as in obtaining images of the sun.



**FIGURE 2.15** An example of digital image acquisition. (a) Illumination (energy) source. (b) A scene. (c) Imaging system. (d) Projection of the scene onto the image plane. (e) Digitized image.

$$0 \leq f(x, y) < \infty \quad (2-3)$$

Function  $f(x, y)$  is characterized by two components: (1) the amount of source illumination incident on the scene being viewed, and (2) the amount of illumination reflected by the objects in the scene. Appropriately, these are called the *illumination* and *reflectance* components, and are denoted by  $i(x, y)$  and  $r(x, y)$ , respectively. The two functions combine as a product to form  $f(x, y)$ :

$$f(x, y) = i(x, y)r(x, y) \quad (2-4)$$

where

$$0 \leq i(x, y) < \infty \quad (2-5)$$

and

$$0 \leq r(x, y) \leq 1 \quad (2-6)$$

Thus, reflectance is bounded by 0 (total absorption) and 1 (total reflectance). The nature of  $i(x, y)$  is determined by the illumination source, and  $r(x, y)$  is determined by the characteristics of the imaged objects. These expressions are applicable also to images formed via transmission of the illumination through a medium, such as a

chest X-ray. In this case, we would deal with a *transmissivity* instead of a *reflectivity* function, but the limits would be the same as in Eq. (2-6), and the image function formed would be modeled as the product in Eq. (2-4).

**EXAMPLE 2.1: Some typical values of illumination and reflectance.**

The following numerical quantities illustrate some typical values of illumination and reflectance for visible light. On a clear day, the sun may produce in excess of  $90,000 \text{ lm/m}^2$  of illumination on the surface of the earth. This value decreases to less than  $10,000 \text{ lm/m}^2$  on a cloudy day. On a clear evening, a full moon yields about  $0.1 \text{ lm/m}^2$  of illumination. The typical illumination level in a commercial office is about  $1,000 \text{ lm/m}^2$ . Similarly, the following are typical values of  $r(x, y)$ : 0.01 for black velvet, 0.65 for stainless steel, 0.80 for flat-white wall paint, 0.90 for silver-plated metal, and 0.93 for snow.

Let the intensity (gray level) of a monochrome image at any coordinates  $(x, y)$  be denoted by

$$\ell = f(x, y) \quad (2-7)$$

From Eqs. (2-4) through (2-6) it is evident that  $\ell$  lies in the range

$$L_{\min} \leq \ell \leq L_{\max} \quad (2-8)$$

In theory, the requirement on  $L_{\min}$  is that it be nonnegative, and on  $L_{\max}$  that it be finite. In practice,  $L_{\min} = i_{\min} r_{\min}$  and  $L_{\max} = i_{\max} r_{\max}$ . From Example 2.1, using average office illumination and reflectance values as guidelines, we may expect  $L_{\min} \approx 10$  and  $L_{\max} \approx 1000$  to be typical indoor values in the absence of additional illumination. The units of these quantities are  $\text{lum/m}^2$ . However, actual units seldom are of interest, except in cases where photometric measurements are being performed.

The interval  $[L_{\min}, L_{\max}]$  is called the *intensity* (or *gray*) *scale*. Common practice is to shift this interval numerically to the interval  $[0, 1]$ , or  $[0, C]$ , where  $\ell = 0$  is considered black and  $\ell = 1$  (or  $C$ ) is considered white on the scale. All intermediate values are shades of gray varying from black to white.

## 2.4 IMAGE SAMPLING AND QUANTIZATION

As discussed in the previous section, there are numerous ways to acquire images, but our objective in all is the same: to generate digital images from sensed data. The output of most sensors is a continuous voltage waveform whose amplitude and spatial behavior are related to the physical phenomenon being sensed. To create a digital image, we need to convert the continuous sensed data into a digital format. This requires two processes: *sampling* and *quantization*.

### BASIC CONCEPTS IN SAMPLING AND QUANTIZATION

Figure 2.16(a) shows a continuous image  $f$  that we want to convert to digital form. **An image may be continuous with respect to the  $x$ - and  $y$ -coordinates, and also in**

The discussion of sampling in this section is of an intuitive nature. We will discuss this topic in depth in Chapter 4.