7

Image Processing Basics

7.0 INTRODUCTION

In the previous six chapters, we studied the fundamentals of 2-D digital signal processing. In the next four chapters, we will study digital image processing, an important application of 2-D digital signal processing theories.

Digital image processing has many practical applications. One of the earliest applications was processing images from the Ranger 7 mission at the Jet Propulsion Laboratory in the early 1960s. The imaging system mounted on the spacecraft had a number of constraints imposed on it, such as size and weight, and images received had such degradations as blurring, geometric distortions, and background noise. These images were successfully processed by digital computers, and since then images from space missions have been routinely processed by digital computers. The striking pictures of the moon and the planet Mars we see in magazines have all been processed by digital computers.

The image processing application which probably has had the greatest impact on our lives is in the field of medicine. Computed tomography, which has its basis in the projection-slice theorem discussed in Section 1.4.3, is used routinely in many clinical situations, for example, detecting and identifying brain tumors. Other medical applications of digital image processing include enhancement of x-ray images and identification of blood vessel boundaries from angiograms.

Another application, much closer to home for the average person, is the improvement of television images.* The image that we view on a television monitor has flickering, limited resolution, a ghost image, background noise, and motion

*It may be argued that improving the quality of television programs is a much more urgent problem. We point out, though, that improving image quality probably will not make mediocre programs any worse.

crawling due to line interlace. Digital televisions are not far from realization, and digital image processing will have a major impact on improving the image quality of existing television systems and on developing such new television systems as high definition television.

One major problem of such video communications as video conferencing and video telephone has been the enormous bandwidth required. A straightforward coding of broadcast quality video requires on the order of 100 million bits per second. By sacrificing quality and using digital image coding schemes, systems that transmit intelligible images at bit rates lower than 100 thousand bits per second have become commercially available.

Robots are expected to play an increasingly important role in industries and homes. They will perform jobs that are very tedious or dangerous and jobs that require speed and accuracy beyond human ability. As robots become more sophisticated, computer vision will play an increasingly important role. Robots will be asked not only to detect and identify industrial parts, but also to "understand" what they "see" and take appropriate actions. Digital image processing will have a major impact on computer vision.

In addition to these well-established application areas of digital image processing, there are a number of less obvious ones. Law enforcement agents often take pictures in uncooperating environments, and the resulting images are often degraded. For example, snapshots of moving cars' license plates are often blurred; reducing blurring is essential in identifying the car. Another potential application is the study of whale migration. When people study the migratory behavior of lions, tigers, and other land animals, they capture animals and tag them on a convenient tail or ear. When the animals are recaptured at another location, the tags serve as evidence of migratory behavior. Whales, however, are quite difficult to capture and tag. Fortunately, whales like to show their tails, which have features that can be used to distinguish them. To identify a whale, a snapshot of its tail, taken from shipboard, is compared with a reference collection of photographs of thousands of different whales' tails. Successive sightings and identifications of an individual whale allow its migration to be tracked. Comparing photographs, though, is extremely tedious, and digital image processing may prove useful in automating the task.

The potential applications of digital image processing are limitless. In addition to the applications discussed above, they include home electronics, astronomy, biology, physics, agriculture, geography, defense, anthropology, and many other fields. Vision and hearing are the two most important means by which humans perceive the outside world, so it is not surprising that digital image processing has potential applications not only in science and engineering, but also in any human endeavor.

Digital image processing can be classified broadly into four areas, depending on the nature of the task. These are image enhancement, restoration, coding, and understanding. In image enhancement, images either are processed for human viewers, as in television, or are preprocessed to aid machine performance, as in object identification by machine. Image enhancement is discussed in Chapter 8. In image restoration, an image has been degraded in some manner such as blurring,

and the objective is to reduce or eliminate the effect of degradation. Image restoration is closely related to image enhancement. When an image is degraded, reducing the image degradation often results in enhancement. There are, however, some important differences between restoration and enhancement. In image restoration, an ideal image has been degraded and the objective is to make the processed image resemble the original as much as possible. In image enhancement, the objective is to make the processed image look better in some sense than the unprocessed image. To illustrate this difference, note that an original, undegraded image cannot be further restored, but can be enhanced by increasing sharpness. Image restoration is discussed in Chapter 9. In image coding, one objective is to represent an image with as few bits as possible, preserving a certain level of image quality and intelligibility acceptable for a given application such as video conferencing. Image coding is related to image enhancement and restoration. If we can enhance the visual appearance of the reconstructed image, or if we can reduce degradation from such sources as quantization noise from an image coding algorithm, then we can reduce the number of bits required to represent an image at a given level of image quality and intelligibility. Image coding is discussed in Chapter 10.

In image understanding, the objective is to symbolically represent the contents of an image. Applications of image understanding include computer vision, robotics, and target identification. Image understanding differs from the other three areas in one major respect. In image enhancement, restoration, and coding, both the input and the output are images, and signal processing has been the backbone of many successful systems in these areas. In image understanding, the input is an image, but the output is typically some symbolic representation of the contents of the input image. Successful development of systems in this area involves both signal processing and artificial intelligence concepts. In a typical image understanding system, signal processing is used for such lower-level processing tasks as reduction of degradation and extraction of edges or other image features, and artificial intelligence is used for such higher-level processing tasks as symbol manipulation and knowledge base management. We treat some of the lower-level processing techniques useful in image understanding as part of our general discussion of image enhancement, restoration, and coding. A more complete treatment of image understanding is beyond the scope of this book.

The theoretical results we studied in the first six chapters are generally based on a set of assumptions. In practice, these assumptions rarely are satisfied exactly. Some results may not be useful in image processing; others may have to be modified. We need to know the basics of image processing if we are to understand the theories' applicability and limitations and to modify them when necessary to adapt to real-world problems. Moreover, the first six chapters have focused on general theories that apply not only to image processing, but also to other 2-D signal processing problems such as geophysical data processing. Some important theoretical results specific to image processing have not yet been discussed. Some basic knowledge of image processing is needed to understand these theories. In this chapter, we present the basics of image processing. These basics will lay a foundation for later chapters' discussion of image enhancement, restoration, and coding. In Section 7.1, we discuss basics of the images we process. In Sections 7.2 and 7.3, we discuss the basics of the human visual system. In Section 7.4, we discuss the basics of a typical image processing environment.

7.1 LIGHT

7.1.1 Light as an Electromagnetic Wave

Everything that we view is seen with light. There are two types of light sources. One type, called a *primary light source*, emits its own light. Examples of primary light sources include the sun, lamps, and candles. The other type, called a *secondary light source*, only reflects or diffuses the light emitted by another source. Examples of secondary light sources include the moon, clouds, and apples.

Light is part of a vast, continuous spectrum of electromagnetic radiation. An electromagnetic wave carries energy, and the energy distribution of the wave passing through a spatial plane can be represented by $c(x, y, t, \lambda)$, where x and y are two spatial variables, t is the time variable, and λ is the wavelength. The function $c(x, y, t, \lambda)$ is called *radiant flux* per (area $\times$ wavelength) or *irradiance* per wavelength. The wavelength λ is related to the frequency f by

$$\lambda = c/f \tag{7.1}$$

where c is the speed* of an electromagnetic wave, approximately 3×10^8 m/sec in vacuum and air. Although the function $c(x, y, t, \lambda)$ can be expressed in terms of the frequency f, it is more convenient to use the wavelength λ. The unit associated with $c(x, y, t, \lambda)$ is energy per (area $\times$ time $\times$ wavelength) and is joules/(m^3 sec) in the MKS (meter, kg, second) system. If we integrate $c(x, y, t, \lambda)$ with respect to λ, we obtain irradiance that has the unit of joules/(m^2 sec) or watts/m^2. Radiation from the sun that passes through a spatial plane perpendicular to the rays has 1350 watts/m^2 of irradiance in the absence of atmospheric absorption. If we integrate $c(x, y, t, \lambda)$ with respect to all four variables x, y, t and λ, we obtain the total energy (in joules) of the electromagnetic wave that passes through the spatial plane.

Light is distinguished from other electromagnetic waves—for instance, radio transmission waves—by the fact that the eye is sensitive to it. Suppose we consider a fixed spatial point (x', y') and a fixed time t'. The function $c(x, y, t, \lambda)$ can be viewed as a function of λ only. We can express it as $c(x', y', t', \lambda)$, or $c(\lambda)$ for convenience. An example of $c(\lambda)$ for the radiation from the sun is shown in Figure 7.1. The eye is sensitive to electromagnetic waves over an extremely narrow range of λ, that is, approximately from 350 nm to 750 nm. (1 nm = 10^{-9} meter). Figure 7.2 shows different types of electromagnetic waves as a function of the wavelength λ. Electromagnetic radiation with large λ, from a few centimeters to several thousand meters, can be generated by electrical circuits. Such radiation is used

*The variable c is used both as the speed and as the energy distribution function of an electromagnetic wave. Which is meant will be apparent from the context.

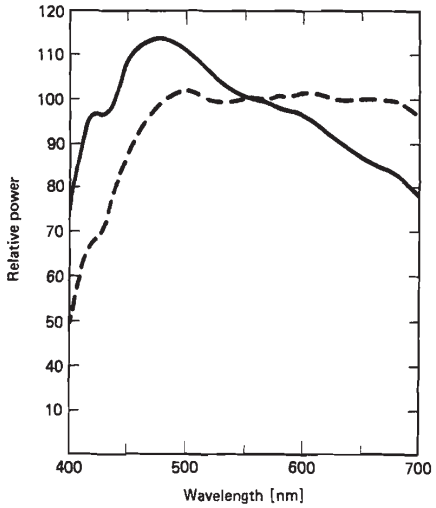


Figure 7.1 Spectral contents of the sun's radiation, above the earth's atmosphere (solid line) and on the ground at noon in Washington (dotted line). After [Hardy].

for radio transmission and radar. Radiation with λ just above the visible range is called *infrared*; with λ just below the visible range, it is called *ultraviolet*. Both infrared and ultraviolet radiations are emitted by typical light sources such as the sun. Radiation with λ far below the visible range includes X rays, γ rays, and cosmic rays; for cosmic rays, λ is less than 10^{-5} nm or 10^{-14} m.

7.1.2 Brightness, Hue, and Saturation

Human perception of light with $c(\lambda)$ is generally described in terms of brightness, hue, and saturation. *Brightness* refers to how bright the light is. *Hue* refers to the color, such as red, orange, or purple. *Saturation*, sometimes called *chroma*, refers to how vivid or dull the color is. Brightness, hue, and saturation are perceptual terms, and they depend on a number of factors, including the detailed shape of $c(\lambda)$, the past history of the observer's exposure to visual stimuli, and the specific environment in which the light is viewed. Nevertheless, it is possible to relate them very approximately to specific features of $c(\lambda)$.

To relate the human perception of brightness to $c(\lambda)$, it is useful to define photometric quantities. The quantities associated with $c(\lambda)$, such as radiant flux, irradiance, and watts/m^2, are called *radiometric units*. These physical quantities can be defined independent of a specific observer. The contributions that $c(\lambda_1)$ and $c(\lambda_2)$ make to human perception of brightness are in general quite different for $\lambda_1 \neq \lambda_2$ even though $c(\lambda_1)$ may be the same as $c(\lambda_2)$. For example, an electromagnetic wave with $c(\lambda)$ is invisible to a human observer as long as $c(\lambda)$ is zero in the visible range of λ, no matter how large $c(\lambda)$ may be outside the visible range. Even within the visible range, the brightness depends on λ. For this reason, a simple integral of $c(\lambda)$ over the variable λ does not relate well to the

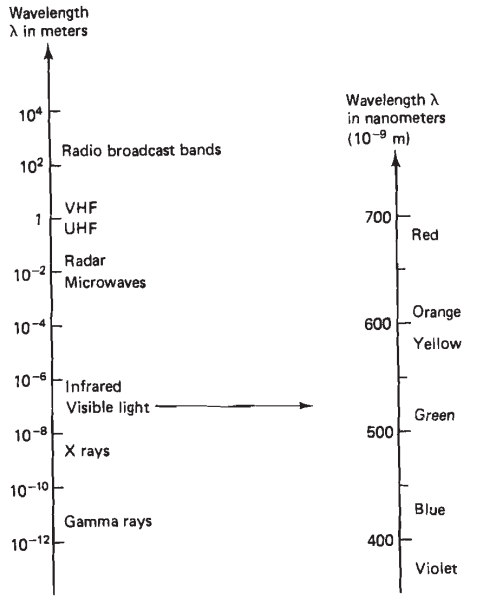


Figure 7.2 Different types of electromagnetic waves as a function of the wavelength λ.

perception of brightness. The quantities which take the human observer's characteristics into account, thus relating to brightness better than the integral of $c(\lambda)$, are called *photometric quantities*.

The basic photometric quantity is luminance, adopted in 1948 by the C.I.E. (Commission Internationale de l'Eclairage), an international body concerned with standards for light and color. Consider a light with $c(\lambda)$ that is zero everywhere except at $\lambda = \lambda_r$, where λ_r denotes a fixed reference wavelength. A light that consists of only one spectral component (one wavelength) is called a *monochromatic light*. Suppose we ask a human observer to compare the brightness from a monochromatic light with $c(\lambda_r)$ with that from another monochromatic light with $c'(\lambda_t)$ where λ_t is a test wavelength. Suppose further that the observer says that $c(\lambda_r)$ matches $c'(\lambda_t)$ in brightness. The equal brightness points $c(\lambda_r)$ and $c'(\lambda_t)$ can be obtained by such experiments as showing two patches of light with a fixed $c(\lambda_r)$ and a variable $c'(\lambda_t)$ and asking the observer to decrease or increase the amplitude of $c'(\lambda_t)$ until they match in brightness. The ratio $c(\lambda_r)/c'(\lambda_t)$ where $c(\lambda_r)$ and $c'(\lambda_t)$ match in brightness, is called the *relative luminous efficiency* of a monochromatic light with λ_t relative to λ_r, and is approximately independent of the amplitude of $c(\lambda_r)$ under normal viewing conditions. The wavelength λ_r used is 555 nm (yellow-green light), at which a typical observer has maximum brightness sensitivity. For this choice of λ_r, the relative luminous efficiency $c(\lambda_r)/c'(\lambda_t)$ is always less than or equal to 1, since $c(\lambda_r)$ is not greater than $c'(\lambda_t)$; that is, it takes less energy at λ_r to produce the same brightness. The relative luminous efficiency as a function of

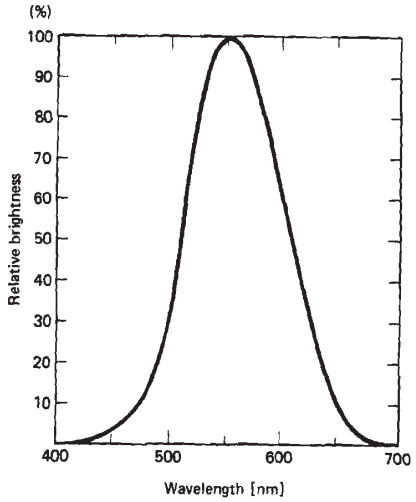


Figure 7.3 C.I.E. relative luminous efficiency function. After [Hardy].

λ is called the *relative luminous efficiency function* and is denoted by $v(\lambda)$. Two monochromatic lights with $c_1(\lambda_1)$ and $c_2(\lambda_2)$ appear equally bright to an observer when*

$$c_1(\lambda_1)v(\lambda_1) = c_2(\lambda_2)v(\lambda_2). \tag{7.2}$$

The relative luminous efficiency function $v(\lambda)$ depends on the observer. Even with a single observer, a slightly different $v(\lambda)$ is obtained when measured at different times. To eliminate this variation, the *C.I.E. standard observer* was defined in 1929, based on experimental results obtained from a number of different observers. The resulting function $v(\lambda)$ is called the *C.I.E. relative luminous efficiency function* and is shown in Figure 7.3. The C.I.E. function is roughly bell-shaped with a maximum of 1 at $\lambda = 555$ nm.

The basic unit of luminance is the lumen (lm). The luminance per area l of a light with $c(\lambda)$ can be defined by

$$l = k \int_{\lambda=0}^{\infty} c(\lambda)v(\lambda)\, d\lambda. \tag{7.3}$$

In (7.3), the quantity l is in units of lumens/m^2, k is 685 lumens/watt, $c(\lambda)$ is in units of watts/m^3, $v(\lambda)$ is unitless, and λ has a unit of m. A monochromatic light with an irradiance of 1 watt/m^2 produces 685 lumens/m^2 when $v(\lambda) = 1$. This

*Our discussions in this section are brief and some reasonable assumptions are made. For example, (7.2) is based on the transmittivity law, which states that if A and B are equally bright, and B and C are equally bright, then A and C are equally bright. This transmittivity law has been approximately verified experimentally.

occurs when $\lambda = 555$ nm. At other wavelengths, $v(\lambda) < 1$, so the irradiance of the monochromatic light must be greater than 1 watt/m^2 to generate a luminance per area of 685 lumens/m^2. Many other photometric units such as footcandle (lumen/ft^2) and phot (lumen/cm^2) can be defined in terms of lumen.

It is important to note that the luminance or luminance per area does not measure human perception of brightness. For example, a light with 2 lumens/m^2 does not appear twice as bright to a human observer as a light with 1 lumen/m^2. It is also possible to create an environment where a light with a smaller luminance per area looks brighter than a light with a larger luminance per area. However, luminance per area is related more directly than an integral of $c(\lambda)$ to human perception of brightness. Furthermore, in typical viewing conditions (light neither too weak nor excessive), a light with larger luminance per area is perceived to be brighter than a light with smaller luminance per area.

Hue is defined as that attribute of color which allows us to distinguish red from blue. In some cases, the hue of a color can be related to simple features of $c(\lambda)$. Light with approximately constant $c(\lambda)$ in the visible range appears white or colorless. Under normal viewing conditions, a monochromatic light appears colored and its color depends on λ. When an observer is shown a succession of monochromatic lights side by side, the color changes smoothly from one hue to another. Light can be split into a succession of monochromatic lights by a prism, as shown in Figure 7.4. This was first done in 1666 by Newton. Newton divided the color spectrum in the visible range into seven broad categories: red, orange, yellow, green, blue, indigo, and violet, in the order of longer to shorter λ. These are known as the seven colors of the rainbow. Newton originally began with only red, yellow, green, blue and violet. He later added orange and indigo to bring the number to seven (in keeping with the tradition of dividing the week into seven days, the musical notes into seven, and so on).

When a light is not monochromatic but its $c(\lambda)$ is narrow band in the sense that most of its energy is concentrated in $\lambda' - \Delta\lambda < \lambda < \lambda' + \Delta\lambda$ for small $\Delta\lambda$, the perceived hue roughly corresponds to monochromatic light with $\lambda = \lambda'$. The color will appear less pure, however, than a monochromatic light of a similar hue. When $c(\lambda)$ is some arbitrary function, it is difficult to relate the hue to some simple features of $c(\lambda)$. By proper choice of $c(\lambda)$, it is possible to produce hues that do not correspond to any monochromatic light. By mixing red and blue lights, for example, it is possible to produce purple light.

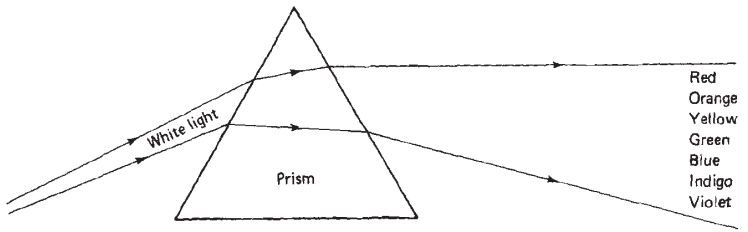


Figure 7.4 White light split into a succession of monochromatic lights by a prism.

Saturation refers to the purity or vividness of the color. A monochromatic light has very pure spectral contents, and looks very vivid and pure. It is said to be highly saturated. As the spectral content of $c(\lambda)$ widens, the color is perceived as less vivid and pure, and the color is said to be less saturated. Color saturation is related very approximately to the effective width of $c(\lambda)$.

7.1.3 Additive and Subtractive Color Systems

When two lights with $c_1(\lambda)$ and $c_2(\lambda)$ are combined, the resulting light has $c(\lambda)$ given by

$$c(\lambda) = c_1(\lambda) + c_2(\lambda). \tag{7.4}$$

Since the lights add in (7.4), this is called an *additive color system*. By adding light sources with different wavelengths, many different colors can be generated. For example, the lighted screen of a color television tube is covered with small, glowing phosphor dots arranged in groups of three. Each of these groups contains one red, one green, and one blue dot. These three colors are used because by proper combination they can produce a wider range of colors than any other combination of three colors; they are the primary colors of the additive color system. Colors of monochromatic lights change gradually, and it is difficult to pinpoint the specific wavelength corresponding to red (R), green (G), and blue (B). The C.I.E. has chosen $\lambda = 700$ nm for red, $\lambda = 546.1$ nm for green, and $\lambda = 435.8$ nm for blue.

The three primary colors of the additive color system are shown in Figure 7.5. In the additive color system, a mixture of equal amounts of blue and green produces cyan. A mixture of equal amounts of red and blue produces magenta, and a mixture of equal amounts of red and green produces yellow. The three colors yellow (Y), cyan (C), and magenta (M) are called the secondary colors of the additive color system. When roughly equal amounts of all three colors R, G, and B are combined, the result is white. When roughly equal amounts of R, G, and B components are used in a color TV monitor, therefore, the result is a black-and-white image. By combining different amounts of the R, G, and B components, other colors can be obtained. A mixture of a red light and a weak green light with no blue light, for example, produces a brown light.

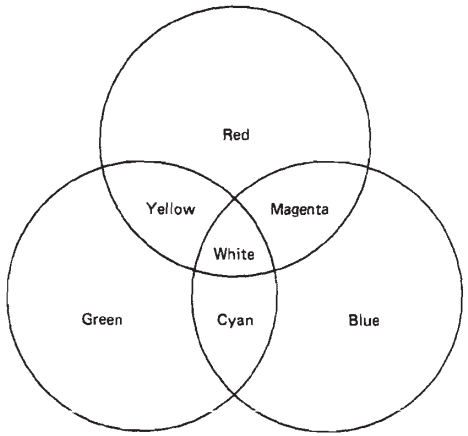


Figure 7.5 Primary colors of the additive color system.

Nature often generates color by filtering out, or subtracting, some wavelengths and reflecting others. This process of wavelength subtraction is accomplished by molecules called *pigments*, which absorb particular parts of the spectrum. For example, when sunlight, which consists of many different wavelengths, hits a red apple, the billions of pigment molecules on the surface of the apple absorb all the wavelengths except those corresponding to red. As a result, the reflected light has a $c(\lambda)$ which is perceived as red. The pigments subtract out certain wavelengths, and a mixture of two different types of pigments will result in a reflected light whose wavelengths are further reduced. This is called a *subtractive color system*. When two inks of different colors are combined to produce another color on paper, the subtractive color system applies.

The three primary colors of a subtractive color system are yellow (Y), cyan (C), and magenta (M), which are the secondary colors of the additive color system. The three colors are shown in Figure 7.6. By mixing the proper amounts of these colors (pigments), a wide range of colors can be generated. A mixture of yellow and cyan produces green. A mixture of yellow and magenta produces red. A mixture of cyan and magenta produces blue. Thus, the three colors, red, green, and blue, the primary colors of the additive color system, are the secondary colors of the subtractive color system. When all three primary colors Y, C, and M are combined, the result is black; the pigments absorb all the visible wavelengths.

It is important to note that the subtractive color system is fundamentally different from the additive color system. In the additive color system, as we add colors (lights) with different wavelengths, the resulting light consists of more wavelengths. We begin with black, corresponding to no light. As we then go from

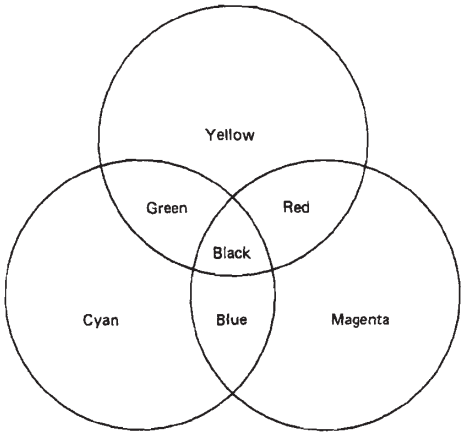


Figure 7.6 Primary colors of the subtractive color system.

the primary colors (RGB) to the secondary colors (YCM) and then to white, we increase the spread of the wavelengths in the resulting light. In a subtractive color system, we begin with white, corresponding to no pigments. As we go from the primary colors (YCM) to the secondary colors (RGB) and then to black, we decrease the spread of the wavelengths in the resulting reflected light.

In an additive color system, we can think of red, green, and blue light as the result of passing white light through three different bandpass filters. Mixing two colors can be viewed as passing white light through a filter which is a parallel combination of the two corresponding bandpass filters. In a subtractive color system, we can think of yellow, cyan, and magenta as the result of passing white light through three different bandstop filters. Mixing two colors can be viewed as passing white light through a cascade of the two corresponding bandstop filters.

7.1.4 Representation of Monochrome and Color Images

For a black-and-white image, a light with $c(\lambda)$ can be represented by one number I given by

$$I = k \int_{\lambda=0}^{\infty} c(\lambda) s_{BW}(\lambda)\, d\lambda \tag{7.5}$$

where $s_{BW}(\lambda)$ is the spectral characteristic of the sensor used and k is some scaling constant. Since brightness perception is the primary concern with a black-and-white image, $s_{BW}(\lambda)$ is typically chosen to resemble the relative luminous efficiency function discussed in Section 7.1.2. The value I is often referred to as the *luminance*, *intensity*, or *gray level* of a black-and-white image. Since I in (7.5) represents power per unit area, it is always nonnegative and finite; that is,

$$0 \le I \le I_{MAX} \tag{7.6}$$

where I_{MAX} is the maximum I possible. In image processing, I is typically scaled such that it lies in some convenient arbitrary range, for example, $0 \le I \le 1$ or $0 \le I \le 255$. In these cases 0 corresponds to the darkest possible level and 1 or 255 corresponds to the brightest possible level. Because of this scaling, the specific radiometric or photometric units associated with I become unimportant. A black-and-white image has, in a sense, only one color. Thus, it is sometimes called a monochrome image.

A color image can be viewed as three monochrome images. For a color image, a light with $c(\lambda)$ is represented by three numbers which are called *tristimulus values*. One three-number set that is frequently used in practice is R, G, and B, representing the intensity of the red, green, and blue components. The tristimulus values R, G, and B are obtained by

$$R = k \int_{\lambda=0}^{\infty} c(\lambda) s_R(\lambda)\, d\lambda \tag{7.7a}$$

$$G = k \int_{\lambda=0}^{\infty} c(\lambda) s_G(\lambda)\, d\lambda \tag{7.7b}$$

and

$$B = k \int_{\lambda=0}^{\infty} c(\lambda) s_B(\lambda)\, d\lambda \tag{7.7c}$$

where $s_R(\lambda)$, $s_G(\lambda)$, and $s_B(\lambda)$ are spectral characteristics of the red, green, and blue sensors (filters) respectively. Like the gray level I in a monochrome image, R, G, and B are nonnegative and finite. One possible set of $s_R(\lambda)$, $s_G(\lambda)$, and $s_B(\lambda)$ is shown in Figure 7.7. Examples of $f_R(x, y)$, $f_G(x, y)$ and $f_B(x, y)$, which represent the red, green and blue components of a color image, are shown in Figures 7.8(a), (b), and (c), respectively (see color insert). The color image that results when the three components are combined by a color television monitor is shown in Figure 7.8(d).

One approach to processing a color image is to process three monochrome images, R, G, and B, separately and then combine the results. This approach is simple and is often used in practice. Since brightness, hue, and saturation each depends on all three monochrome images, independent processing of R, G, and B may affect hue and saturation, even though the processing objective may be only modifying the brightness.

The three tristimulus values R, G, and B can be transformed into a number of other sets of tristimulus values. One particular set, known as *luminance-chrominance*, is quite useful in practice. When R, G, and B are the values used on a

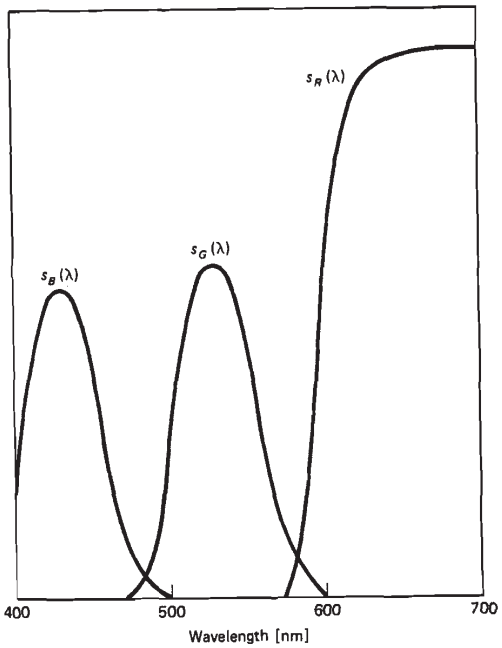


Figure 7.7 Example of spectral characteristics of red, green, and blue color sensors.

(a)

(b)

(c)

Figure 7.9 Y, I, and Q components of the color image in Figure 7.8(d). (a) Y component; (b) I component; (c) Q component.

TV monitor (in the NTSC* color system), the corresponding luminance-chrominance values Y, I, and Q are related to R, G, and B by

$$\begin{bmatrix} Y \\ I \\ Q \end{bmatrix} = \begin{bmatrix} 0.299 & 0.587 & 0.114 \\ 0.596 & -0.274 & -0.322 \\ 0.211 & -0.523 & 0.312 \end{bmatrix} \begin{bmatrix} R \\ G \\ B \end{bmatrix} \tag{7.8a}$$

*NTSC stands for National Television Systems Committee.

and

$$\begin{bmatrix} R \\ G \\ B \end{bmatrix} = \begin{bmatrix} 1.000 & 0.956 & 0.621 \\ 1.000 & -0.273 & -0.647 \\ 1.000 & -1.104 & 1.701 \end{bmatrix} \begin{bmatrix} Y \\ I \\ Q \end{bmatrix}. \tag{7.8b}$$

The Y component is called the luminance component, since it roughly reflects the luminance l in (7.3). It is primarily responsible for the perception of the brightness of a color image, and can be used as a black-and-white image. The I and Q components are called chrominance components, and they are primarily responsible for the perception of the hue and saturation of a color image. The $f_Y(x, y)$, $f_I(x, y)$, and $f_Q(x, y)$ components, corresponding to the color image in Figure 7.8, are shown as three monochrome images in Figures 7.9(a), (b), and (c), respectively. Since $f_I(x, y)$ and $f_Q(x, y)$ can be negative, a bias has been added to them for display. The mid-gray intensity in Figures 7.9(b) and (c) represents the zero amplitude of $f_I(x, y)$ and $f_Q(x, y)$. One advantage of the YIQ tristimulus set relative to the RGB set is that we can process the Y component only. The processed image will tend to differ from the unprocessed image in its appearance of brightness. Another advantage is that most high-frequency components of a color image are primarily in the Y component. Therefore, significant spatial lowpass filtering of I and Q components does not significantly affect the color image. This feature can be exploited in the coding of a digital color image or in the analog transmission of a color television signal.

When the objective of image processing goes beyond accurately reproducing the "original" scene as seen by human viewers, we are not limited to the range of wavelengths visible to humans. Detecting objects that generate heat, for example, is much easier with an image obtained using a sensor that responds to infrared light than with a regular color image. Infrared images can be obtained in a manner similar to (7.7) by simply changing the spectral characteristics of the sensor used.

7.2 THE HUMAN VISUAL SYSTEM

7.2.1 The Eye

The human visual system is one of the most complex in existence. Our visual system allows us to organize and understand the many complex elements of our environment. For nearly all animals, vision is just an instrument of survival. For humans, vision is not only an aid to survival, but an instrument of thought and a means to a richer life.

The visual system consists of an eye that transforms light to neural signals, and the related parts of the brain that process the neural signals and extract necessary information. The eye, the beginning of the visual system, is approximately spherical with a diameter of around 2 cm. From a functional point of view, the eye is a device that gathers light and focuses it on its rear surface.

A horizontal cross section of an eye is shown in Figure 7.10. At the very front of the eye, facing the outside world, is the cornea, a tough, transparent membrane. The main function of the cornea is to refract (bend) light. Because

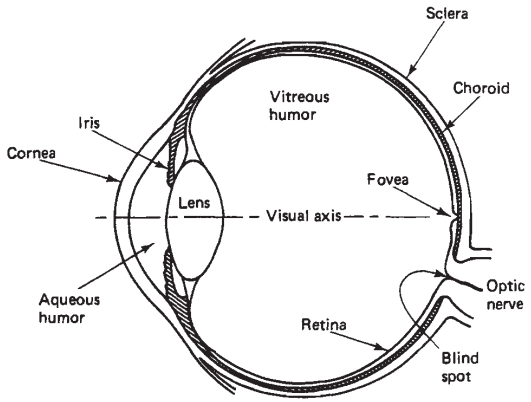


Figure 7.10 Horizontal cross section of a right human eye, seen from above.

of its rounded shape, it acts like the convex lens of a camera. It accounts for nearly two-thirds of the total amount of light bending needed for proper focusing.

Behind the cornea is the aqueous humor, which is a clear, freely flowing liquid. Through the cornea and the aqueous humor, we can see the iris. By changing the size of the pupil, a small round hole in the center of the iris, the iris controls the amount of light entering the eye. Pupil diameter ranges between 1.5 mm ~ 8 mm, with smaller diameter corresponding to exposure to brighter light. The color of the iris determines the color of the eye. When we say that a person has blue eyes, we mean blue irises. Iris color, which has caught the attention of so many lovers and poets, is not functionally significant to the eye.

Behind the iris is the lens. The lens consists of many transparent fibers encased in a transparent elastic membrane about the size and shape of a small bean. The lens grows throughout a person's lifetime. Thus, the lens of an eighty-year-old man is more than fifty percent larger than that of a twenty-year old. As with an onion, cells in the oldest layer remain in the center, and cells in newer layers grow further from the center. The lens has a bi-convex shape and a refractive index of 1.4, which is higher than any other part of the eye through which light passes. However, the lens is surrounded by media that have refractive indices close to its own. For this reason, much less light-bending takes place at the lens than at the cornea. The cornea has a refractive index of 1.38, but faces the air, which has a refractive index of 1. The main function of the lens is to accurately focus the incoming light on a screen at the back of the eye called the retina. For a system with a fixed lens and a fixed distance between the lens and the screen, it is possible to focus objects at only one particular distance. If faraway objects are in sharp focus, for example, close objects will be focused behind the screen. To be able to focus close objects at one time and distant objects at some other time, a camera changes the distance between the fixed lens and the screen. This is what the eyes of many fish do. In the case of the human eye, the shape of the lens, rather than the distance between the lens and screen, is changed. This process of changing shape to meet the needs of both near and far vision is called *accommodation*. This adjustability of the shape is the most important feature of the lens. Accommodation takes place almost instantly and is controlled by the ciliary body, a group of muscles surrounding the lens.

Behind the lens is the vitreous humor, which is a transparent jelly-like substance. It is optically matched so that light which has been sharply focused by the lens keeps the same course. The vitreous humor fills the entire space between the lens and the retina and occupies about two-thirds of the eye's volume. One of its functions is to support the shape of the eye.

Behind the vitreous humor is the retina, which covers about 65% of the inside of the eyeball. This is the screen on which the entering light is focused and light-receptive cells convert light to neural signals. All of the other eye parts we have discussed so far serve the function of placing a sharp image on this receptor surface. The fact that an image is formed on the retina, so the eye is simply an image catching device, was not known until the early seventeenth century. Even though the ancient Greeks knew the structure of an eye accurately and performed delicate surgery on it, they theorized that light-like rays emanate from the eye, touch an object, and make it visible. After all, things appear "out there." In 1625, Scheiner demonstrated that light enters the eye and vision stems from the light that enters the eye. By exposing the retina of an animal and looking through it from behind, he was able to see miniature reproductions of the objects in front of the eyeball.

There are two types of light-receptive cells in the retina. They are called cones and rods because of their shape. The cones, which number about 7 million, are less sensitive to light than rods, and are primarily for day (photopic) vision. They are also responsible for seeing color. The three types of cones are most sensitive to red, green, and blue light, respectively. This is the qualitative physiological basis for representing a color image with red, green, and blue monochrome images. The rods, which number about 120 million, are more sensitive to light than cones, and are primarily for night (scotopic) vision. Since the cones responsible for color vision do not respond to dim light, we do not see color in very dim light.

Rods and cones are distributed throughout the retina. However, their distribution is highly uneven. The distribution of the rods and cones in the retina is shown in Figure 7.11. Directly behind the middle point of the pupil is a small

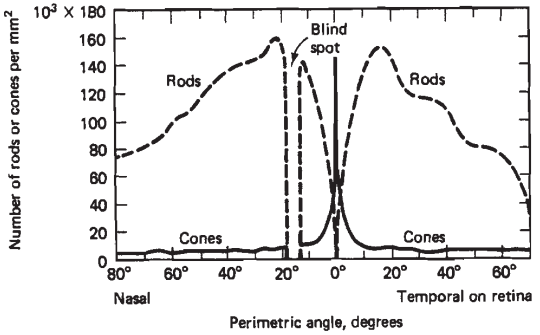


Figure 7.11 Distribution of rods (dotted line) and cones (solid line) on the retina. After [Pirenne].

depressed dimple on the retina called the fovea. There are no rods in this small region, and most of the cones are concentrated here. Therefore, this is the region for the most accurate vision in bright light. When we look straight ahead at an object, the object is focused on the fovea. Since the fovea is very small, we constantly move our attention from one region to another when studying a larger region in detail. The rods, which function best in night vision, are concentrated away from the fovea. Since there are no rods in the fovea, an object focused in the fovea is not visible in dim light. To see objects at night, therefore, we look at them slightly sideways.

There are many thin layers in the retina. Even though cones and rods are light-receptive cells, so that it would be reasonable for them to be located closer to the vitreous humor, they are located farther away from it. Therefore, light has to pass through other layers of the retina, such as nerve fibers, to reach the cones and rods. This is shown in Figure 7.12. It is not clear why nature chose to do it this way, but the arrangement works. In the fovea, at least, the nerves are pushed aside so that the cones are directly exposed to light. Due to this particular arrangement, the optic nerve fibers have to pass through the light-receptive cell layers

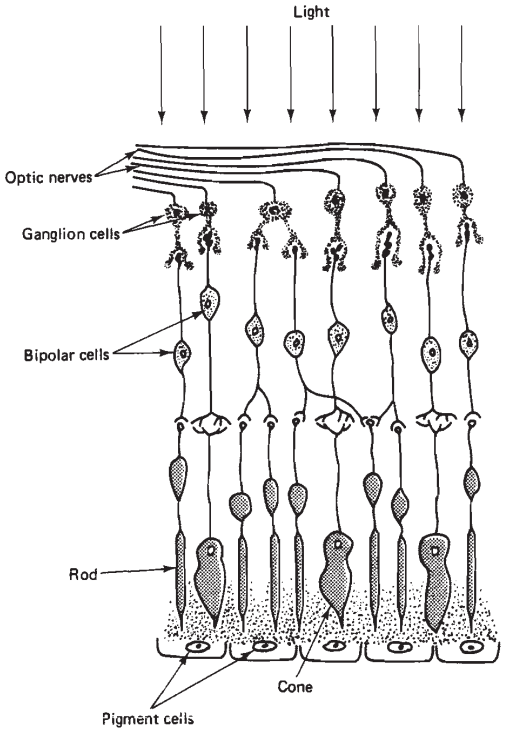


Figure 7.12 Layers in the retina. Note that light has to travel through several layers before it reaches light-sensitive cells.

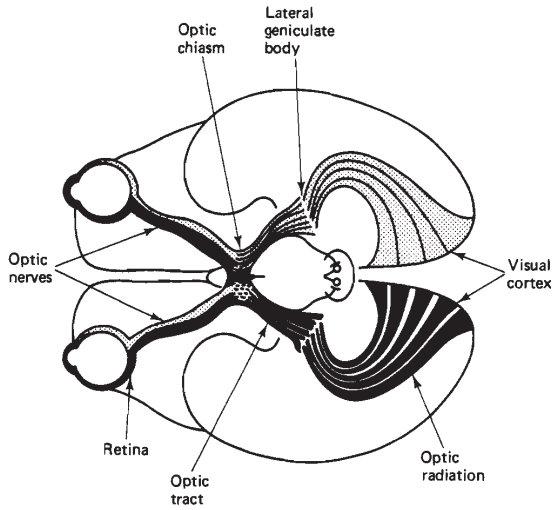


Figure 7.13 Path that neural signals travel from the retina to the visual cortex.

on the way to the brain. Instead of crossing the light-receptive cell layers throughout the retina, they bundle up at one small region the size of a pinhead in the retina, known as the blind spot. Since there are no light receptive cells in this region, we cannot see light focused on the blind spot.

When light hits cones and rods, a complex electrochemical reaction takes place, and light is converted to neural impulses, which are transmitted to the brain through the optic nerve fibers. There are about 130 million light-receptive cells (cones and rods), but only about 1 million nerve fibers. This means that one nerve fiber, on the average, serves more than 100 light-receptive cells. The nerve fibers are not shared equally. Some cones in the fovea are served by one nerve fiber each, increasing the visual acuity in this region. The rods, however, always share nerve fibers. This is one reason why visual acuity at night is not as good as it is during the day even though there are many more rods than cones.

After the optic nerve bundles leave the two eyes, the two bundles meet at an intersection called the optic chiasm. This is shown in Figure 7.13. Each of the two bundles is divided into two branches. Two branches, one from each of the two bundles, join together to form a new bundle. The remaining two branches form another bundle. This crossing of the nerve fibers from two eyes is partly responsible for our stereoscopic vision, which mixes the images from each eye to allow the visual field to be perceived as a 3-D space. These two new bundles go to the left and right lateral geniculate bodies, respectively. The original fibers end here and new fibers continue to the visual cortex, where the neural signals are processed and vision takes place. The visual cortex is a small part of the cortex, a mass of gray matter forming two hemispheres in the back of the brain. Little is known about how the visual neural signals are processed in the visual cortex.

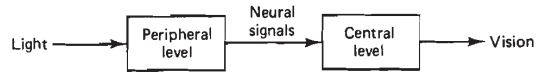


Figure 7.14 Human visual system as a cascade of two systems. The first system represents the peripheral level of the visual system and converts light to neural signals. The second system represents the central level and processes neural signals to extract necessary information.

7.2.2 Model for Peripheral Level of Visual System

The human visual system discussed in Section 7.2.1 can be viewed as a cascade of two systems, as shown in Figure 7.14. The first system, which represents the peripheral level of the visual system, converts light to a neural signal. The second system, which represents the central level of the visual system, processes the neural signal to extract information.

Unlike central level processing, about which little is known, peripheral level processing is fairly well understood, and many attempts have been made to model it. One very simple model [Stockham] for a monochrome image that is consistent with some well-known visual phenomena is shown in Figure 7.15. In this model, the monochrome image intensity $I(x, y)$ is modified by a nonlinearity, such as a logarithmic operation, that compresses the high level intensities but expands the low level intensities. The result is then filtered by a linear shift-invariant (LSI) system with spatial frequency response $H(\Omega_x, \Omega_y)$. The nonlinearity is motivated by the results of some psychophysical experiments that will be discussed in the next section. The LSI system $H(\Omega_x, \Omega_y)$, which is bandpass in character, is motivated by the finite size of the pupil, the resolution limit imposed by a finite number of light-receptive cells, and the lateral inhibition process. The finite size of the pupil and the resolution limit due to a finite number of light receptive cells are responsible for the lowpass part of the bandpass nature of $H(\Omega_x, \Omega_y)$. The lateral inhibition process stems from the observation that one neural fiber responds to many cones and rods. The response of the neural fiber is some combination of the signals from the cones and rods. While some cones and rods contribute positively, others contribute negatively (inhibition). This lateral inhibition process is the rationale for the highpass part of the bandpass character of $H(\Omega_x, \Omega_y)$. Even though the model in Figure 7.15 is very simple and applies only to the peripheral level of the human visual system, it is consistent with some of the visual phenomena that are discussed in the next section.

One way to exploit a model such as the one in Figure 7.15 is to process an image in a domain closer to where vision takes place. This can be useful in some applications. In image coding, for example, the information that is in the image but is discarded by the visual system does not need to be coded. By processing an image in a domain closer to where vision takes place, more emphasis can be

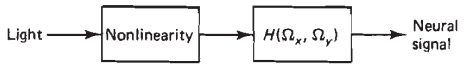


Figure 7.15 Simple model of peripheral level of human visual system.

placed on what is important to the visual system. This is one reason why some image processing operations are performed in the log intensity domain rather than the intensity domain.

In addition to the model in Figure 7.15, more sophisticated models for monochrome images and models for color images have also been proposed in the literature [Budrikis; Mannos and Sakrison; Hall and Hall; Faugeras; Granrath].

7.3 VISUAL PHENOMENA

7.3.1 Intensity Sensitivity

One way to quantify our ability to resolve two visual stimuli which are the same except for their intensities or luminances is by measuring the just-noticeable difference (j.n.d.). The j.n.d. can be defined and measured in a variety of ways. One way is through a psychophysical experiment called *intensity discrimination*. Suppose we present the visual stimulus in Figure 7.16 to an observer. The inside region is a monochrome image of uniform intensity I_{in}, which is randomly chosen to be either I or $I + \Delta I$. The outside region is a monochrome image of intensity I_{out}, which is chosen to be $I + \Delta I$ when $I_{in} = I$ and I when $I_{in} = I + \Delta I$. The observer is asked to make a forced choice as to which of the two intensities I_{in} and I_{out} is brighter. When ΔI is very large, the observer will give a correct answer most of the time, correct in the sense that the region with $I + \Delta I$ is chosen. When ΔI is very small, the observer will give a correct answer about half of the time. As we move away from a very large ΔI to a very small ΔI, the percentage of the observer's correct responses decreases continuously, and we can define ΔI at which the observer gives correct responses 75% of the time as the j.n.d. at the intensity I.

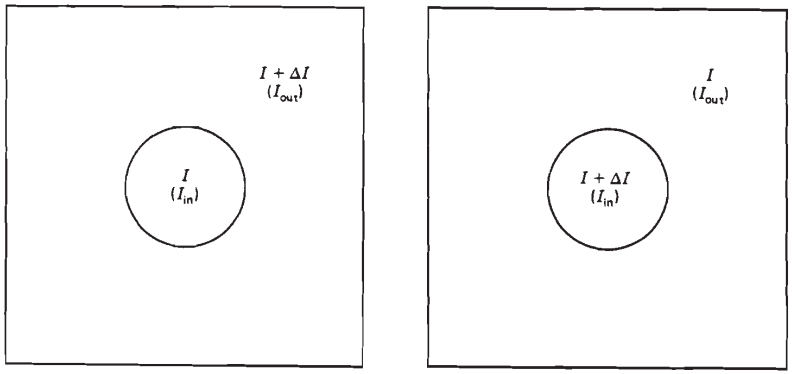


Figure 7.16 Two stimuli used in an intensity discrimination experiment. Each trial consists of showing one of the two stimuli to an observer and asking the observer to make a forced choice of which between I_{in} and I_{out} appears brighter. The stimulus used in a trial is chosen randomly from the two stimuli. Results of this experiment can be used to measure the just noticeable difference ΔI as a function of I.

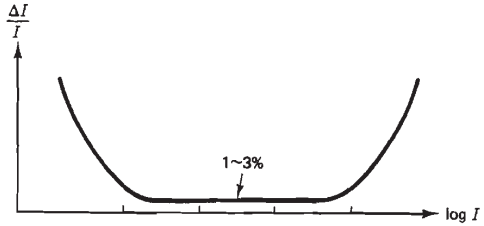


Figure 7.17 Plot of $\Delta I/I$ as a function of I. Incremental intensity ΔI is the just noticeable difference. Over a wide range of I, $\Delta I/I$ is approximately constant. This relationship is called Weber's law.

The result of plotting $\Delta I/I$ as a function of I where ΔI is the j.n.d. is shown in Figure 7.17. From the figure, for a wide range of I,

$$\frac{\Delta I}{I} \approx \text{constant}. \tag{7.9}$$

This relationship is called *Weber's law*. Weber's law states that the j.n.d. ΔI is proportional to I. As we increase I, we need a larger ΔI to make $I + \Delta I$ noticeably different from I. This is one way in which the visual system remains sensitive to a wide dynamic range of the visual stimulus intensity. Weber's law holds approximately not only for vision, but also for all other human sense modalities: hearing, smell, taste, and touch.

As we let ΔI approach 0, (7.9) can be written as

$$\frac{dI}{I} = d(\log I) \approx \text{constant}. \tag{7.10}$$

From (7.10), the j.n.d. is constant in the $\log I$ domain for a wide range of I. This is consistent with the notion that a nonlinear operation such as the log is applied to the image intensity in the simple model in Figure 7.15. The intensity discrimination experiment involves a very simple task on the part of the observer, and

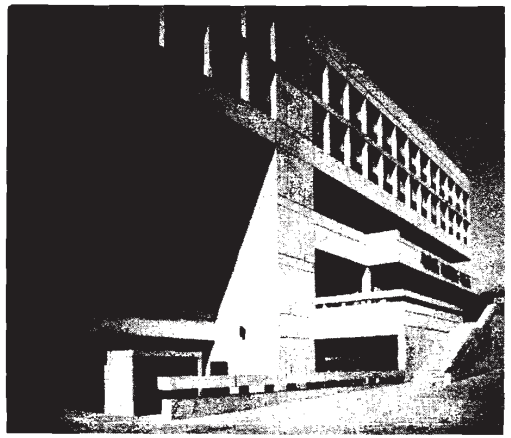
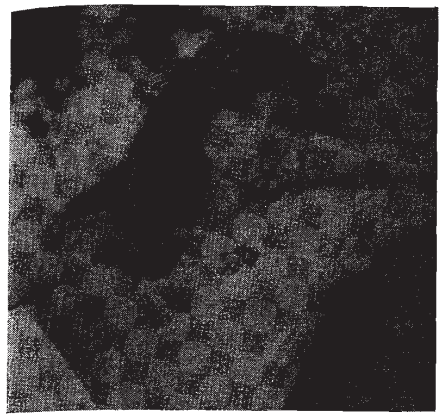
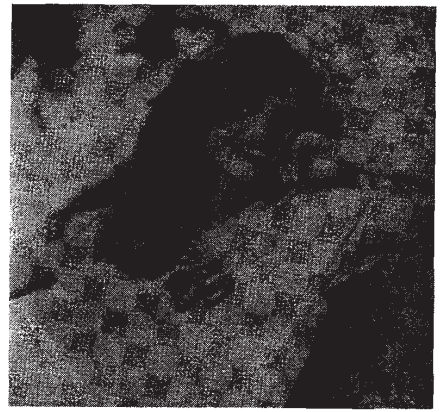


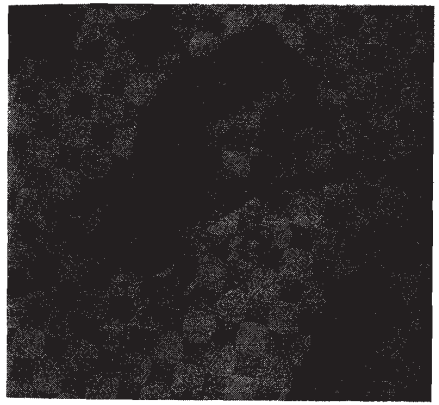
Figure 7.18 Image of 512 × 512 pixels degraded by zero-mean white noise with a uniform probability density function. Same level of noise is more visible in a dark region relative to a bright region. Same level of noise is more visible in a uniform background region than in a region with edges.



(a)



(b)

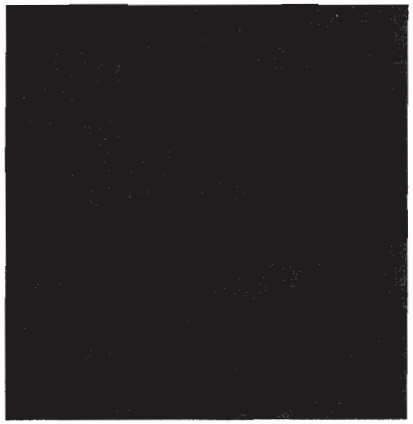


(c)

(d)

Figure 7.8 Red, green, and blue components of a color image. (a) Red component; (b) green component; (c) blue component; (d) color image of 512 × 512 pixels.

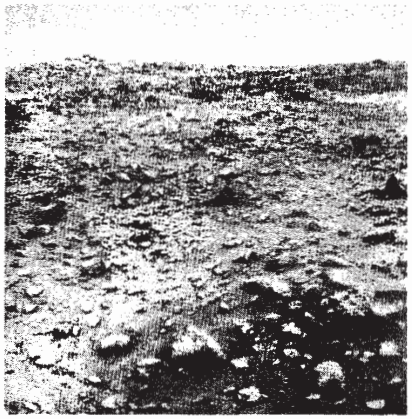
Figure 8.7 Example of gray scale modification. (a) Original color image of 512 × 512 pixels; (b) processed color image.



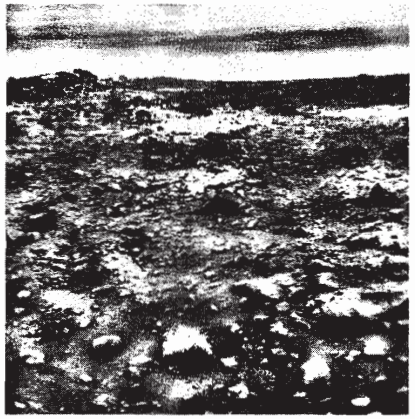
(a)

(b)

Figure 8.49 Creation of a color image from a black-and-white image by mapping the result of lowpass filtering the black-and-white image to the blue component, the result of bandpass filtering to the green component, and the result of highpass filtering to the red component.



(a)



(b)

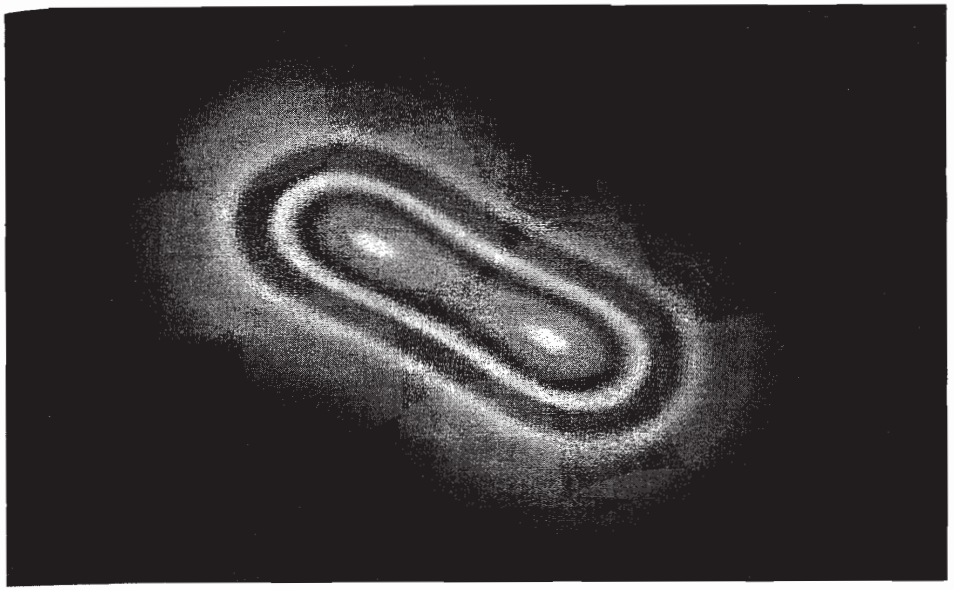
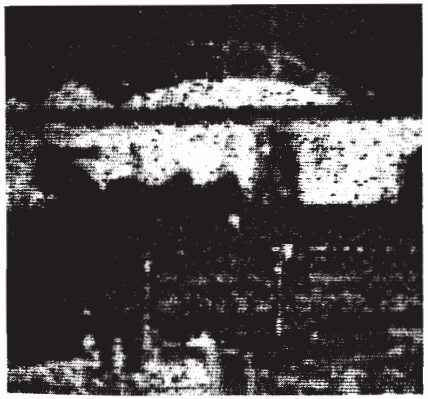
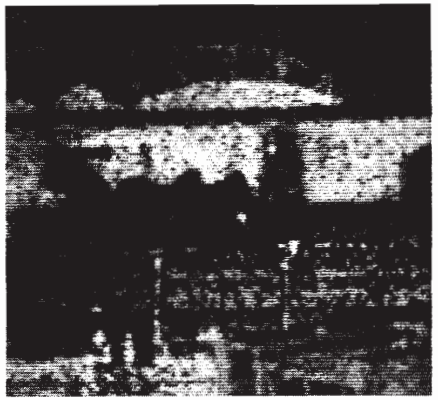


Figure 8.50 (b)



(a)



(b)

Figure 8.51 Display of range information with color. After [Sullivan et al.]. (a) Intensity image from an infrared radar imaging system with range information discarded; (b) image in (a) with range information displayed with color.

(a)

(b)

Figure 10.58 Example of color image coding using the scene adaptive coder. Courtesy of Wen-Hsiung Chen. (a) Original image of 512 × 512 pixels; (b) reconstructed image at 0.4 bit/pixel.

(a)

(b)

Figure 10.59 Another example of color image coding using the scene adaptive coder. Courtesy of Wen-Hsiung Chen. (a) Original image of 512 × 512 pixels; (b) reconstructed color image of 0.4 bit/pixel.

complicated central level processing probably is not needed. Consequently, the result of the intensity discrimination experiment can be related to peripheral level processing in the visual system.

The result of the intensity discrimination experiment states that the j.n.d. ΔI increases as I increases. This partly explains why a uniform level of random noise is more visible in a darker region than in a brighter region. This is illustrated in Figure 7.18. The image in Figure 7.18 is the result of adding zero-mean white noise with a uniform probability density to an original undegraded image. The grainy appearance, which is due to noise, is more pronounced in the darker uniform background region than in the brighter uniform background region. Since the j.n.d. ΔI is larger for a larger I, a higher level of noise in the brighter region is required for it to be equally visible as a lower level of noise in the darker region. The implication of this observation for image processing is that reducing noise in the darker region is more important than reducing noise in the brighter region.

7.3.2 Adaptation

In the intensity discrimination experiment discussed above, the intensities shown at any given time are I and $I + \Delta I$. If the observer is assumed to spend some time before making a decision, the result is obtained when the observer is adapted to the intensity level I. When the intensity level the observer is adapted to is different from I, the observer's intensity resolution ability decreases. Suppose we run the same intensity discrimination experiment discussed in Section 7.3.1, but with I and $I + \Delta I$ surrounded by a much larger region with intensity I_0, as shown in Figure 7.19. The result of plotting $\Delta I/I$ as a function of I and I_0 is shown in

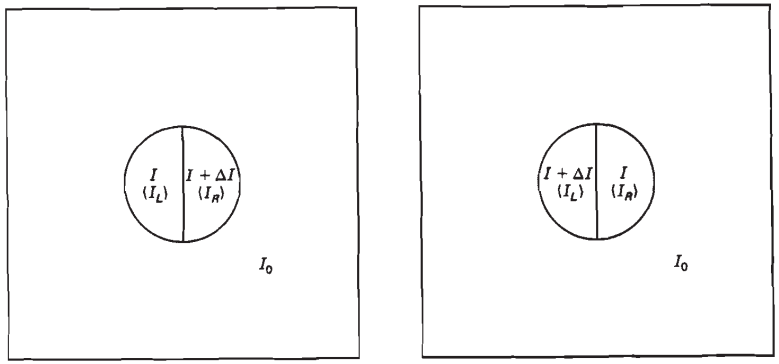


Figure 7.19 Two stimuli used in studying the effect of adaptation on the intensity sensitivity. Each trial consists of showing one of the two stimuli to an observer and asking the observer to make a forced choice of which between I_R or I_L appears brighter. The stimulus used in a trial is chosen randomly from the two stimuli. Results of this experiment can be used to measure the just noticeable difference ΔI as a function of I and I_0.

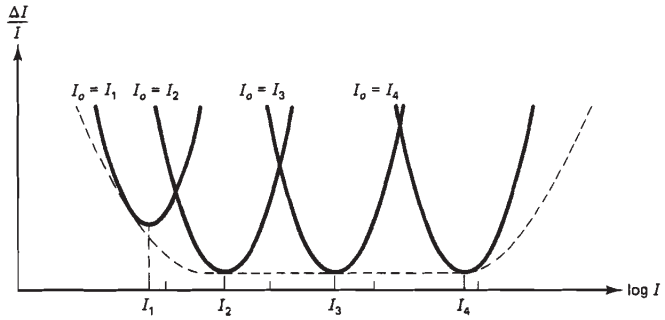


Figure 7.20 Plot of $\Delta I/I$ as a function of I and I_0. When I_0 equals I, $\Delta I/I$ is the same as that in Figure 7.17 (dotted line in this figure). When I_0 is different from I, $\Delta I/I$ increases relative to the case $I_0 = I$. This means that the observer's sensitivity to intensity decreases.

Figure 7.20. When I_0 is equal to I, the result is the same as that in Figure 7.17. When I_0 is different from I, however, the j.n.d. ΔI increases relative to the case $I_0 = I$, indicating that the observer's sensitivity to intensity decreases. The result shows that sensitivity to intensity is highest near the level that the observer is adapted to. This is another way in which the visual system responds to a wide range of intensities at different times.

7.3.3 Mach Band Effect and Spatial Frequency Response

Consider an image whose intensity is constant along the vertical dimension but increases in a staircase manner along the horizontal dimension, as shown in Figure 7.21(a). The intensities along the horizontal direction are sketched in Figure 7.21(b). Even though the intensity within each rectangular region is constant, each region looks brighter towards the left and darker towards the right. This is known as the *Mach band effect*. This phenomenon is consistent with the presence of spatial filtering in the peripheral-level model of the visual system in Figure 7.15. When a filter is applied to a signal with sharp discontinuities, an overshoot and undershoot occur. This is partly responsible for uneven brightness perception within the region of uniform intensity. This suggests that precise preservation of the edge shape is not necessary in image processing.

The presence in the visual system of a spatial bandpass filter can be seen by looking at the image in Figure 7.22. The image $I(x, y)$ in Figure 7.22 is given by

$$I(x, y) = I_0(y) \cos(\omega(x)x) + \text{constant} \tag{7.11}$$

where the constant is chosen such that $I(x, y)$ is positive for all (x, y). As we move in the horizontal direction from left to right, the spatial frequency $\omega(x)$ increases. As we move in the vertical direction from top to bottom, the amplitude $I_0(y)$ increases. If the spatial frequency response were constant across the frequency range, sensitivity to intensity would be constant along the horizontal direction.

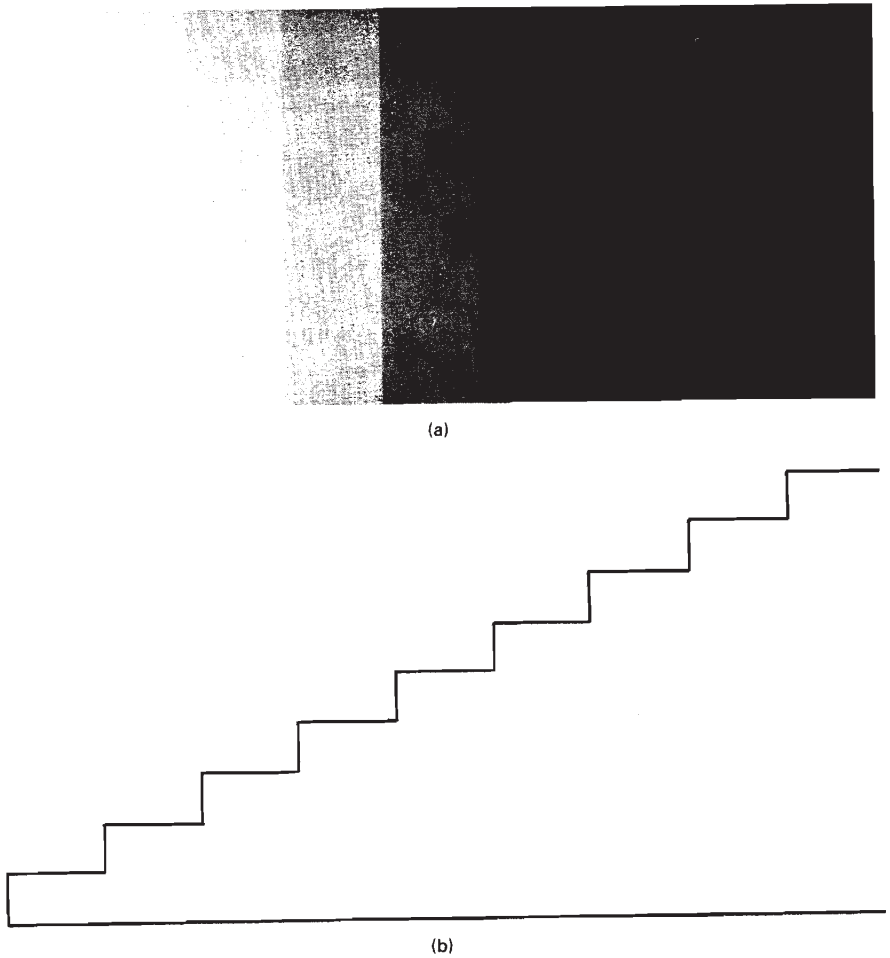


Figure 7.21 Illustration of Mach band effect.

rection. In Figure 7.22, we are more sensitive to the contrast in mid-frequency regions than in low- and high-frequency regions, indicating the bandpass character of the visual system. A spatial filter frequency response $H(\Omega_x, \Omega_y)$, which is more accurately measured by assuming the model in Figure 7.15 is correct, is shown in Figure 7.23. The horizontal axis is the spatial frequency/angle of vision. The perceived spatial frequency of an image changes as a function of the distance between the eye and the image. As the distance increases, the perceived spatial frequency increases. To take this effect into account, the spatial frequency/angle

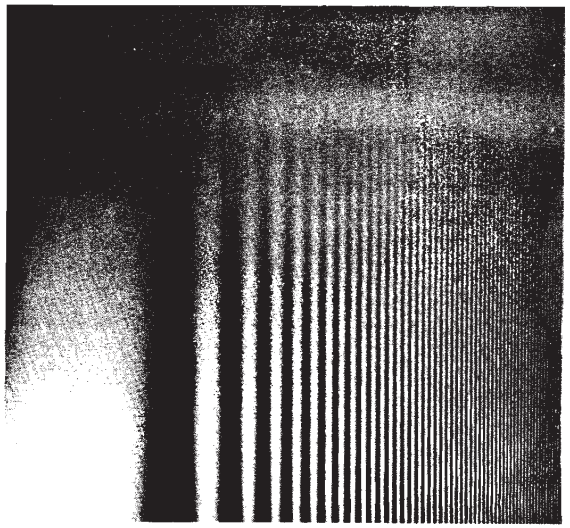


Figure 7.22 Modulated sinewave grating that illustrates the bandpass character of the peripheral level of the human visual system.

of vision (spatial frequency relative to the spatial domain in the retina) is often used in determining $H(\Omega_x, \Omega_y)$. The frequency response $H(\Omega_x, \Omega_y)$ is maximum at the spatial frequency in the range of approximately 5 ~ 10 cycles/degree and decreases as the spatial frequency increases or decreases from 5 ~ 10 cycles/degree.

7.3.4 Spatial Masking

When random noise of a uniform level is added to an image, it is much more visible in a uniform background region than in a region with high contrast. This effect

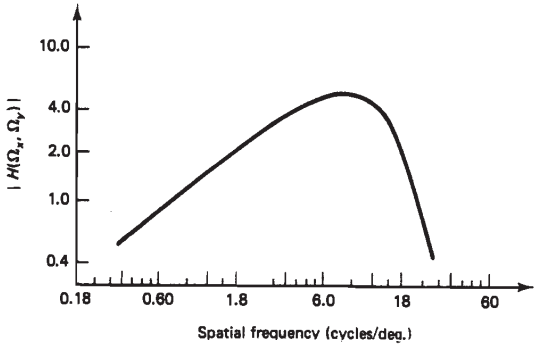


Figure 7.23 Frequency response $H(\Omega_x, \Omega_y)$ in the model in Figure 7.15. After [Davidson].

is much more pronounced than the effect of the brightness level on noise visibility discussed in Section 7.3.1. Consider the image in Figure 7.18, which illustrates the effect of overall brightness level on noise visibility. In the figure, the noise is much less visible in the edge regions than in the uniform background regions. In addition, the noise in the dark edge regions is less visible than the noise in the bright uniform background regions. One way to explain this is to consider the local signal-to-noise ratio (SNR). If we consider the local SNR to be the ratio of the signal variance to the noise variance in a local region, at the same level of noise, the SNR is higher in the high-contrast region than in the uniform background. Another related view is spatial masking. In a high-contrast region, signal level is high and tends to mask the noise more.

The spatial masking effect can be exploited in image processing. For example, attempting to reduce background noise by spatial filtering typically involves some level of image blurring. In high-contrast regions, where the effect of blurring due to spatial filtering is more likely to be pronounced, the noise is not as visible, so little spatial filtering may be needed.

7.3.5 Other Visual Phenomena

It is well known that a sharper image generally looks even more pleasant to a human viewer than an original image. This is often exploited in improving the appearance of an image for a human viewer. It is also a common experience that an unnatural aspect catches a viewer's attention. A positive aspect of this phenomenon is that it can be exploited in such applications as production of television commercials. A negative aspect is that it sometimes makes it more difficult to develop a successful algorithm using computer processing techniques. For example, some image processing algorithms are capable of reducing a large amount of background noise. In the process, however, they introduce noise that has an artificial tone to it. Even when the amount of the artificial noise introduced is much less than the amount by which the background noise is reduced, the artificial noise may catch a viewer's attention more, and a viewer may prefer the unprocessed image over the processed image.

The visual phenomena discussed in the previous sections can be explained simply, at least at a qualitative level; however, many other visual phenomena cannot be explained simply, in part due to our lack of knowledge. For example, a visual phenomenon that involves a fair amount of central level processing cannot be explained in a satisfactory manner. Figure 7.24 shows a sketch consisting of just a small number of lines. How we can associate this image with Einstein is not clear. The example does demonstrate, however, that simple outlines representing the gross features of an object are important for its identification. This can be exploited in such applications as object identification in computer vision and the development of a very low bit-rate video telephone system for the deaf.

The visual phenomena discussed above relate to the perception of light that shines continuously. When light shines intermittently, our perception depends a great deal on its frequency. Consider a light that flashes on for a brief duration N times per second. When N is small, the light flashes are perceived to be separate.

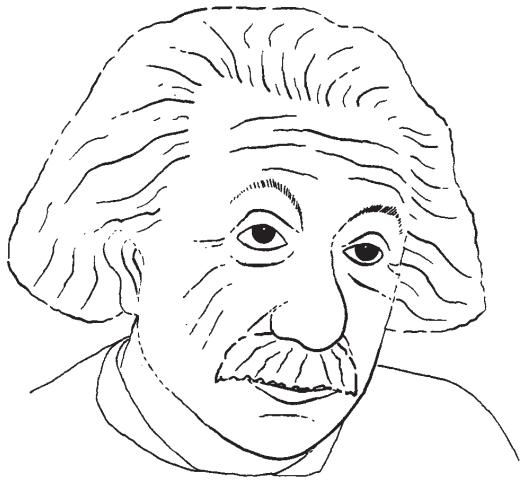


Figure 7.24 Sketch of Einstein's face with a small number of lines.

As we increase N, an unsteady flicker that is quite unpleasant to the human viewer occurs. As we increase N further, the flicker becomes less noticeable, and eventually the observer can no longer detect that the light intensity is changing as a function of time. The frequency at which the observer begins perceiving light flashes as continuous light is called the *critical flicker frequency* or *fusion frequency*. The fusion frequency increases as the size and overall intensity of the flickering object increase. The fusion frequency can be as low as a few cycles/sec for a very dim, small light and may exceed 100 cycles/sec for a very bright, larger light. When a flicker is perceived, visual acuity is at its worst.

Intermittent light is common in everyday vision. Fluorescent lights do not shine continuously, as they appear to, but flicker at a sufficiently high rate (over 100 times/sec) that fusion is reached in typical viewing conditions. Avoiding the perception of flicker is an important consideration in deciding the rate at which a CRT (cathode ray tube) display monitor is refreshed. As is discussed further in Section 7.4, CRT display monitors are illuminated only for a short period of time. For an image to be displayed continuously without the perception of flicker, the monitor has to be refreshed at a sufficiently high rate. Typically, a CRT display monitor is refreshed 60 times per second. With 2:1 interlace, which is discussed further in Section 7.4, this corresponds to 30 frames/sec. The current NTSC (National Television Systems Committee)* television system employs 30 frames/sec with 2:1 interlace. For motion pictures, 24 frames per second are shown, with one frame shown twice. The effective flicker rate is therefore 48 frames/sec. In addition, the typical viewing condition in a cinema is very dark, decreasing the fusion frequency to below 40 cycles/sec. For this reason, flickering is not visible in a motion picture, even though the screen is dark for approximately half of the time.

*The NTSC is a group formed in the United States in 1940 to formulate standards for monochrome television broadcast service. It was reconvened in 1950 to study color television systems and recommended a standard system for the United States. The NTSC system is used in North America and Japan. Both the SECAM (Sequential Couleur a Memoire) System used in France, the USSR and Eastern Europe, and the PAL (Phase Alternating Line) system used in South America, Africa, and Western Europe except France employ 25 frames/sec with 2:1 interlace.

Even though each frame of a motion picture or television broadcast is actually still, and only a finite number of frames are shown in a second, the objects in the scene appear to be moving in a continuous manner. This effect, known as *motion rendition*, is closely related to the phi phenomenon [Kinchla et al]. Consider two pulsating light sources separated by approximately 1 degree of an observer's viewing angle. When the lights shine for one msec each with a separation of 10 msec, the light is perceived to move continuously from one source to the other. When the time difference between the two lights is on the order of 1 msec, they appear to flash simultaneously. When the time difference is more than 1 second, they are perceived as two separate flashes. This is known as the *phi phenomenon*.

In general, frame rates that are sufficiently high to avoid flicker are adequate for motion rendition. The fact that an object appears to move in a continuous manner in a motion picture or television broadcast does not necessarily imply that the sampling rate along the temporal dimension is above the Nyquist rate. For objects with sufficiently rapid motion, sampling the temporal dimension 24 times/sec or 30 times/sec is far lower than the Nyquist rate, and temporal aliasing occurs. Temporal aliasing does not always cause motion discontinuity. In a movie, we sometimes see a wheel that moves continuously, but backwards. In this case, motion rendition is present, but significant temporal aliasing has occurred. Our current knowledge of the flicker effect, motion rendition, the temporal aliasing effect, and their interrelations is far from complete. A comprehensive understanding of this topic would be useful in a number of applications, such as bit rate reduction by frame elimination in a sequence of image frames.

7.4 IMAGE PROCESSING SYSTEMS

7.4.1 Overview of an Image Processing System

A typical image processing system that involves digital signal processing is shown in Figure 7.25. The input image source $I(x, y)$ is generally an object or a natural scene, but it may be an image produced by another system, such as a filter, a cathode ray tube (CRT) display monitor, or a video cassette recorder (VCR). The digitizer converts the input source to an electrical signal whose amplitude represents the image intensity and digitizes the electrical signal using an analog-to-digital (A/D) converter.

The sequence $f(n_1, n_2)$ that results from the digitizer is then processed by a digital image processing algorithm. The algorithm may be implemented on a general purpose computer, a microprocessor, or a special purpose hardware. The specific algorithm used depends on the objective, which may involve image en-

Figure 7.25 Typical overall image processing system.

hancement, restoration, coding, understanding, or any combination of them. The result of processing is then displayed, generally for human viewing, but sometimes as an input to another system. The display typically used is a CRT monitor, but may be a photograph or VCR tape. If the result is some symbolic representation, as in image understanding, the display used can also be just a printer.

7.4.2 The Digitizer

The digitizer converts the input image source to an electrical signal and then samples the electrical signal, using an A/D converter. The specific functions performed by a digitizer depend on the input image source. When the input is already in the form of an electrical signal, as in the case of a VCR tape, the digitizer is interfaced to the input source and is used in sampling the electrical signal following the format used in converting the input source to an electrical signal.

When the input source is in the form of an image, an electronic camera converts the image to an electrical signal, and the result is digitized by using an A/D converter. In some camera systems, parallel paths allow light intensities at many spatial points to be measured simultaneously. In typical systems, however, there is only one path, and the light intensity at only one spatial point can be measured at a given time. In this case, the entire image is covered by scanning. In most scanning systems, a small aperture searches the image following a certain pattern called a *raster*. The light intensity integrated over a small aperture is measured, is converted to an electrical signal, and is considered to be the image intensity at that spatial point. This process can be viewed as convolving the input image intensity $I(x, y)$ with the aperture and then sampling the result of the convolution. The effect of the aperture is, therefore, lowpass filtering $I(x, y)$. This limits the spatial resolution of $I(x, y)$ and can be used for antialiasing necessary in an A/D converter. For a still image, the image is generally scanned once, but may be scanned more times for noise reduction through frame averaging. For a moving scene, the image is scanned at periodic time intervals.

When the input is a film or a photograph, a common device used to convert the image intensity to an electrical signal is a flying spot scanner. In this arrangement, a small spot of light scans the input source, and the light that is reflected by the photograph or transmitted through the film is collected by wide-area photodetectors. The source of the small light spot is a CRT screen. The CRT is discussed in greater detail in Section 7.4.3.

When the input image source is an object or a natural scene, the most common device for converting light intensity to an electrical signal has been the *vidicon* and its relatives such as a Saticon and Newvicon. The vidicon and its relatives were employed until the early 1980s in practically all TV applications, including broad-

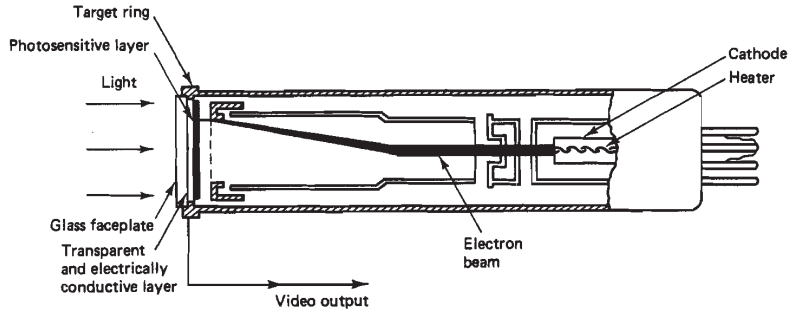


Figure 7.26 Construction of a Vidicon camera.

casting, small portable video cameras, and surveillance cameras. The construction of a vidicon camera tube is shown in Figure 7.26. At one end (the left end in the figure) of the tube inside the glass envelope is an image plate. The plate has two layers. Facing the light from the input source is a thin layer of tin oxide coating that is transparent to light but is electrically conductive. Facing the electron gun is the second layer. This second layer has a coating of photosensitive material, antimony trisulfide for a basic vidicon. The light from the input source passes through an optical lens that is the focusing mechanism, through an optically flat glass plate, and through the first layer of the image plate. The light is then focused on the second layer. The photosensitive image plate (the second layer) is scanned by an electron gun, and the resulting electrical current is the camera signal that is digitized by an A/D converter. The scanning pattern used is from right to left, bottom to top. Since the input source is inverted by the lens, this scanning pattern is equivalent to scanning left to right, top to bottom, in the input image plane.

The photosensitive layer is a semiconductor, which acts as an insulator when no light is present. As light hits this layer, electrons move to the electrically conductive tin oxide layer, creating positive charges on the image-plate surface facing the electron gun. The number of electrons that move, or alternatively the number of positive charges facing the electron gun, represents the image intensity at that spatial point. As the low-energy electron beam from the electron gun scans the image plate, it drops enough electrons to neutralize the positive charges. This discharge current is collected at a metal target ring that is electrically attached to the tin oxide layer. The current at the metal target ring is the camera signal. The electrons originate at the cathode, which is at the other end of the vidicon camera tube. Electrons converge to a narrow beam by means of the electrostatic lens and magnetic focusing.

The spectral response of a basic vidicon for a black and white image is very similar to the C.I.E. relative luminous efficiency function discussed in Section 7.1.2. For color images, a color camera optically separates the incoming light into red, green, and blue components. Each component is the input to a vidicon camera tube. A color camera, therefore, houses three separate tubes.

The camera signal, which represents the intensity of the input image source, is sampled by an A/D converter to obtain digital images. Common digital image sizes are 128 × 128, 256 × 256, 512 × 512, and 1024 × 1024 pixels. As we reduce the number of pixels, the spatial resolution, which is also referred to as *definition*, is decreased and the details in the image begin to disappear. Examples of images of different sizes can be found in Figure 1.11. The amplitude of each pixel is typically quantized to 256 levels (represented by 8 bits). Often, each level is denoted by an integer, with 0 corresponding to the darkest level and 255 corresponding to the brightest. As we decrease the number of amplitude quantization levels, the signal-dependent quantization noise begins to appear first as random noise and then as false contours. Examples of images quantized at different numbers of quantization levels can be found in Figure 1.10. For a color image, each of the red, green, and blue components is typically quantized to 8 bits/pixel, a total of 24 bits/pixel.

The vidicon and its relatives are called *photo-conductive sensors* or *tube sensors*, and were employed until the early 1980s in practically all TV applications. Since the mid-1980s, however, there has been a rapid growth in *solid-state sensors*. In a typical solid-state sensor, a 2-D array of sensor elements is integrated on a chip. One sensor element is located spatially at each pixel location and senses the light intensity at the pixel, the value of which is then read by a scanning mechanism.

The charge coupled device (CCD) is an example of a solid-state sensor element. When a CCD array is exposed to light, charge packets proportional to the light intensity develop. The stored charge packets are shifted to the storage CCD array which is not exposed to light. The light intensity values are then read from the storage array. Depending on how the imaging and storage CCD arrays are configured, different methods [Flory] have been developed to read the light intensity values from the storage array.

Solid-state sensors have many advantages over photo-conductive sensors. They are inherently more stable and compact. A solid-state sensor also has a well-defined structure and the location of every pixel is known accurately in both space and time. As a result, color extraction requires simpler signal processing and better color uniformity can be obtained. Solid-state sensors also have the potential for much higher sensitivity of light detection. In a typical photo-conductive sensor, each pixel is examined one at a time by a single light sensor. The time for light detection is measured in microseconds in a typical TV application and the sensitivity of light detection is low. In a typical solid-state sensor, an array of sensor elements, one for each pixel, is used. Each sensor element, therefore, needs to be checked once per picture. The light energy can be integrated over the time of one frame rather than one pixel, increasing the potential for light sensitivity by orders of magnitude. A solid-state sensor also has a lower lag factor than a photo-conductive sensor. The *lag* is the residual output of a sensor after the light intensity is changed or removed.

Solid-state sensors have some disadvantages in comparison with photo-conductive sensors. One is the spatial resolution. A higher spatial resolution requires

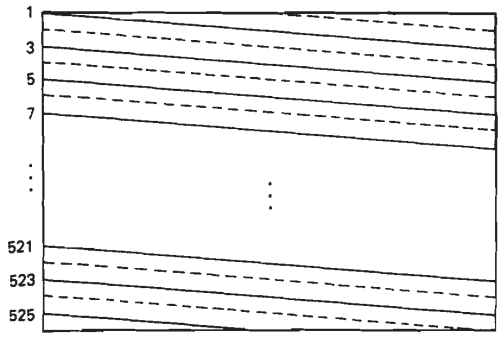


Figure 7.27 Odd (solid line) and even (dotted line) fields in a 2:1 interlace.

a larger number of pixels, meaning a larger number of sensor elements that need to be integrated on a chip. Another is the relatively low signal-to-noise ratio. Despite some of these disadvantages, the solid-state technology is advancing rapidly, and it is expected that solid-state sensors will replace photo-conductive sensors in almost all TV applications in the near future. More details on solid-state sensors can be found in [Baynon and Lamb; Kosonocky; Flory; Declerck].

In the NTSC television broadcasting, 30 frames are transmitted every second. Each frame consists of 525 horizontal scan lines, which are divided into two fields, the odd field and the even field. Each of the two fields is scanned from left to right and from top to bottom, and covers 262½ horizontal lines. The odd field consists of odd-numbered lines, and the even field, of even-numbered lines. The horizontal lines in the odd and even fields are interlaced to form a frame. This is shown in Figure 7.27 and is called a 2:1 interlace. The 2:1 interlace is used so that the vertical resolution will be 525 lines per frame at the rate of 30 frames/sec, but the flickering frequency will be 60 cycles/sec, reducing the perception of flicker at the receiver monitor. Without the interlace, a frame would have to be displayed twice to achieve a flicker frequency of 60 cycles/sec, and this requires frame storage or more transmission bandwidth. The spatial resolution of a television frame displayed as a still image is similar to that of a 512 × 512-pixel digital image. In the television broadcast, the signal remains in the same analog form both at the transmitter and receiver. Digital processing of these signals involves sampling using an A/D converter.

The output of the digitizer is a sequence of numbers. Although the output is represented by a 2-D sequence $f(n_1, n_2)$ in Figure 7.25, the output may be three sequences, $f_R(n_1, n_2)$, $f_G(n_1, n_2)$, and $f_B(n_1, n_2)$, corresponding to the red, green, and blue components for a color image. The output may also be a 3-D sequence $f(n_1, n_2, n_t)$, which is a function of two spatial variables and a time variable for a sequence of frames. These signals are then processed by digital image processing algorithms, which may be implemented on a general purpose computer, a microprocessor, or a special purpose hardware. Digital processing of these signals is the topic of Chapters 8, 9, and 10.

7.4.3 Display

The most common display device in an image processing environment is a CRT. A CRT consists of an electron gun and a phosphor screen, as shown in Figure 7.28. The electron gun produces a beam of electrons that are focused in a narrow region on the phosphor screen through the electrostatic lens, which involves the use of an electrostatic and a magnetic field. The electron beam excites the phosphor to generate light. For monochrome tubes, one beam is used. For color tubes, three separate beams are used to excite the phosphor for each of the three colors. The screen is typically scanned from left to right and from top to bottom to cover the whole screen.

The phosphor chemicals are generally light metals, such as zinc, in the form of sulfide and sulfate. The phosphor material is processed to obtain very fine particles, which are applied to the inside of the glass plate. For monochrome tubes, the phosphor coating is a uniform layer. For color tubes, the phosphor is deposited in dots or vertical layers for each color.

When the high-velocity beams excite the phosphor, electrons in the atoms of the phosphor move to a high energy level. As the electron beams move to a different spot on the screen in raster scanning, the electrons move back to a lower energy level and emit light. The radiation of light from the screen is called *luminescence*. When the light is extinguished, the screen fluoresces. The time it takes for light emitted from the screen to decay to 1% of the maximum value is called the *screen persistence*. For medium to short persistence, generally used for television monitors, the decay time is about 5 msec. Long persistence phosphor exists, but can cause merging of two frames, resulting in substantial blurring for a sequence of images with motion.

Since their screen persistence is short, most CRTs used for display are brightened for short periods of time as the electron beams scan the screen. To display a still image or a sequence of images, the screen must be refreshed continuously. Display monitors often have semiconductor random access memory (RAM) for

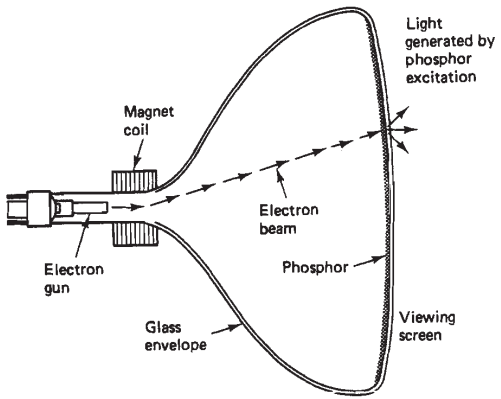


Figure 7.28 Construction of a cathode ray tube.

refresh. The refresh memory, which can typically hold more than one image frame of a reasonable size, is used to refresh the screen.

The display monitor used in an image processing environment is typically a high quality industrial type. A typical monitor has an aspect ratio (width to height) of 1* and is typically calibrated to display a square frame. Standard display sizes are 512 × 512 and 1024 × 1024, but other sizes, such as 640 × 480 and 1024 × 1280, are available. The refresh rate has traditionally been 30 frames/sec with a 2:1 interlace, the same as for television broadcasting. Individual scan lines are sharply defined on high quality monitors. A sharp horizontal line, therefore, may appear only in the odd field or only in the even field, so the flicker rate for this scan line would become 30 cycles/sec. At this rate, the flicker may be visible. To avoid this problem, displays which refresh the screen at 50 frames/sec or 60 frames/sec without interlace are available. This provides a display that is essentially flicker-free and is more pleasant to the human viewer.

In many displays, the refresh memory is used not only for refreshing the screen but also for processing. The displayed image can be zoomed, scrolled, or processed with a simple filter in the refresh memory in real time. As hardware advances continue, the display unit will gain even more sophisticated capabilities.

An image displayed on a CRT screen is considered soft since it is in a sense a temporary image. One approach to making a hard copy is to photograph the CRT screen on film. Special camera systems that have their own small CRT screens are available for such purposes. There are many additional hard copy devices [Schreiber]. To obtain multiple copies, for example, methods such as letterpress, lithography (offset), and gravure exist.

REFERENCES

For books on digital image processing, see [Rosenfeld (1969); Lipkin and Rosenfeld; Andrews (1970); Huang (1975); Andrews and Hunt; Pratt (1978); Hall (1979); Castleman; Pratt (1979); Moik; Huang (1981); Rosenfeld and Kak; Green; Braddick and Sleigh; Ekstrom; Rosenfeld (1984); Yaroslavsky; Niblack; Jensen; Huang (1986); Young and Fu; Gonzalez and Wintz]. For collections of selected papers on digital image processing, see [Andrews (1978); Chellapa and Sawchuck]. For special journal issues on digital image processing, see [Andrews and Enloe; Andrews (1974); Netravali and Habibi; Hunt (1981); Hwang (1983); Netravali and Prasada (1985)].

For readings on light and human vision, see [Graham, et al.; Mueller, et al.; Le Grande; Gregory; Cornsweet; Kupchella]. For a computational view of visual perception, see [Barrow and Tenenbaum]. For readings on color and color vision, see [Wyszecki and Stiles; Yule; Boynton]. On models of human visual system, see [Stockham; Budrikis; Mannos and Sakrison; Hall and Hall; Faugeras; Gran-

*Currently, television receivers have an aspect ratio of 4:3, which is the same ratio used in conventional motion pictures. In high resolution or high definition television systems, the aspect ratio will increase to approximately 5:3 or 16:9.

rath]. On mathematical models of images, see [Jain]. For readings on image processing systems, including cameras and displays, see [Hunt and Breedlove; Weimer; Reader and Hubble; Grob; Flory; Schreiber; Declerck].

H. C. Andrews, *Computer Techniques in Image Processing.* New York: Academic Press, 1970.

H. C. Andrews, ed., *Special Issue on Digital Image Processing, Computer*, Vol. 7, May 1974.

H. C. Andrews, ed., *Tutorial and Selected Papers in Digital Image Processing.* New York: IEEE Press, 1978.

H. C. Andrews and L. H. Enloe, eds., *Special Issue on Digital Picture Processing, Proc. IEEE*, Vol. 60, July 1972.

H. C. Andrews and B. R. Hunt, *Digital Image Restoration.* Englewood Cliffs, NJ: Prentice-Hall, 1977.

H. C. Barrow and J. M. Tenenbaum, Computational vision, *Proc. IEEE*, Vol. 69, May 1981, pp. 572–595.

J. D. Baynon and D. R. Lamb, *Charge Coupled Devices and Their Applications.* London: McGraw-Hill, 1980.

R. M. Boynton, *Human Color Vision.* New York: Holt, Rinehart and Winston, 1979.

O. J. Braddick and A. C. Sleigh, eds., *Physical and Biological Processing of Images.* Berlin: Springer-Verlag, 1983.

Z. L. Budrikis, Visual fidelity criterion and modeling, *Proc. IEEE*, Vol. 60, 1972, pp. 771–779.

K. R. Castleman, *Digital Image Processing.* Englewood Cliffs, NJ: Prentice-Hall, 1979.

R. Chellapa and A. A. Sawchuck, eds., *Digital Image Processing and Analysis: Vol. 1*, Digital Image Processing. New York: IEEE Press, 1985.

T. N. Cornsweet, *Visual Perception.* New York: Academic Press, 1970.

M. L. Davidson, Perturbation approach to spatial brightness interaction in human vision, *J. Opt. Soc. of Am.*, Vol. 58, Sept. 1968, pp. 1300–1308.

G. J. Declerck, ed., *Solid State Imagers and Their Applications, SPIE*, Vol. 591, 1986.

M. P. Ekstrom, ed., *Digital Image Processing Techniques.* Orlando, FL: Academic Press, 1984.

O. D. Faugeras, Digital color image processing within the framework of a human visual model, *IEEE Trans. Acoust. Speech, and Sig. Proc.*, Vol. ASSP-27, August 1979, pp. 380–393.

R. E. Flory, Image Acquisition Technology, *Proc. IEEE*, Vol. 73, April 1985, pp. 613–637.

R. C. Gonzalez and P. Wintz, *Digital Image Processing*, 2nd ed. Reading, MA: Addison-Wesley, 1987.

C. H. Graham et al., eds., *Vision and Visual Perception.* New York: Wiley, 1965.

D. J. Granrath, The role of human visual models in image processing, *Proc. IEEE*, Vol. 69, May 1981, pp. 552–561.

W. B. Green, *Digital Image Processing, A Systems Approach.* New York: Van Nostrand and Reinhold Comp., 1983.

R. L. Gregory, Visual Illusions, *Scientific American*, November 1968, pp. 66–76.

B. Grob, *Basic Television and Video Systems.* 5th ed. New York: McGraw-Hill, 1984.

E. L. Hall, *Computer Image Processing and Recognition.* New York: Academic Press, 1979.

C. F. Hall and E. L. Hall, A nonlinear model for the spatial characteristics of the human visual system, *IEEE Trans. Syst., Man. and Cybern.*, Vol. SMC-7, March 1977, pp. 161–170.

A. C. Hardy, *Handbook of Colorimetry.* Cambridge, MA: MIT Press, 1936.

T. S. Huang, ed., *Picture Processing and Digital Filtering.* Berlin: Springer-Verlag, 1975.

T. S. Huang, ed., *Image Sequence Analysis.* Berlin: Springer-Verlag, 1981.

T. S. Huang, ed., *Advances in Computer Vision and Image Processing, Vol. 2.* New York: JAI Press, 1986.

B. R. Hunt, ed., *Special Issue on Image Processing, Proc. IEEE*, May 1981.

B. R. Hunt and J. R. Breedlove, Scan and display considerations in processing images by digital computer, *IEEE Trans. Computers*, Vol. C-24, Aug. 1975, pp. 848–853.

K. Hwang, ed., *Special Issue on Computer Architectures for Image Processing, Computer*, Vol. 16, January 1983.

A. K. Jain, Advances in mathematical models for image processing, *Proc. IEEE*, Vol. 69, March 1981, pp. 502–528.

J. R. Jensen, *Introductory Digital Image Processing. A Remote Sensing Perspective.* Englewood Cliffs, NJ: Prentice-Hall, 1986.

D. H. Kelly, Theory of flicker and transient responses, I: Uniform Fields, *J. Opt. Soc. Am.*, Vol. 61, April 1971, pp. 537–546.

R. A. Kinchla et al., A theory of visual movement perception, *Psych. Rev.*, Vol. 76, 1968, pp. 537–558.

W. F. Kosonocky, Visible and infra-red solid-state sensors, *1983 IEDM Tech. Dig.*, Int. Elec. Dev. Meeting, Dec. 1983, pp. 1–7.

C. E. Kupchella, *Sights and Sounds.* Indianapolis, IN: The Bobbs-Merrill Company, 1976.

Y. LeGrand, *Light, Color, and Vision*, translated from *Lumiere et Couleurs.* Chapman and Hall, 1968.

B. S. Lipkin and A. Rosenfeld, eds., *Picture Processing and Psychopictorics.* New York: Academic Press, 1970.

J. L. Mannos and D. J. Sakrison, The effects of a visual fidelity criterion on the encoding of images, *IEEE Trans. Inf. Theory*, Vol. IT-20, July 1974, pp. 525–536.

J. K. Moik, *Digital Processing of Remotely Sensed Images.* NASA, 1980.

C. G. Mueller, M. Rudolph, and the editors of LIFE, *Light and Vision.* New York: Time, 1966.

W. Niblack, *An Introduction to Digital Image Processing.* Englewood Cliffs, NJ: Prentice-Hall, 1986.

A. N. Netravali and A. Habibi, *Special Issue on Picture Communication Systems, IEEE Trans. Commun.*, Vol. COM-29, Dec. 1981.

A. N. Netravali and B. Prasada, *Special Issue on Visual Communications Systems, Proc. IEEE*, Vol. 73, April 1985.

M. H. Pirenne, *Vision and the Eye*, 2nd ed. London: Associated Book Publishers, 1967.

W. K. Pratt, *Digital Image Processing.* New York: Wiley, 1978.

W. K. Pratt, ed., *Image Transmission Techniques.* New York: Academic Press, 1979.

C. Reader and L. Hubble, Trends in image display systems, *Proc. IEEE*, Vol. 69, May 1981, pp. 606–614.

A. Rosenfeld, *Picture Processing by Computer.* New York: Academic Press, 1969.

A. Rosenfeld, ed., *Multiresolution Image Processing and Analysis.* Berlin: Springer-Verlag, 1984.

A. Rosenfeld and A. C. Kak, *Digital Picture Processing*, 2nd ed. New York: Academic Press, 1982.

W. F. Schreiber, *Fundamentals of Electronic Imaging Systems.* Berlin: Springer-Verlag, 1986.

T. G. Stockham, Jr., Image processing in the context of a visual model, *Proc. IEEE*, Vol. 60, July 1972, pp. 828–842.

P. K. Weimer, A historical review of the development of television pickup devices (1930–1976), *IEEE Trans. Electron. Devices*, Vol. ED-23, July 1976, pp. 739–752.

G. W. Wyszecki and W. S. Stiles, *Color Science.* New York: Wiley, 1967.

L. P. Yaroslavsky, *Digital Picture Processing: An Introduction.* Berlin: Springer-Verlag, 1985.

T. Y. Young and K. S. Fu, eds., *Handbook of Pattern Recognition and Image Processing.* Orlando, FL: Academic Press, 1986.

J. A. C. Yule, *Principles of Color Reproduction.* New York: Wiley, 1967.

PROBLEMS

7.1. Let λ denote the wavelength of an electromagnetic wave propagating in a vacuum and let f denote its frequency. For each of the following wavelengths, determine the corresponding frequency f.

(a) $\lambda = 100$ m (radio broadcast)
(b) $\lambda = 1$ m (VHF, UHF)
(c) $\lambda = 10^{-1}$ m (radar)
(d) $\lambda = 5 \cdot 10^{-7}$ m (visible light)
(e) $\lambda = 10^{-9}$ m (X rays)
(f) $\lambda = 10^{-11}$ m (gamma rays)

7.2. Let $c(\lambda)$ denote the energy distribution of a light as a function of its wavelength λ. Let A and B denote two lights with energy distributions of $c_A(\lambda)$ and $c_B(\lambda)$, respectively, as shown below.

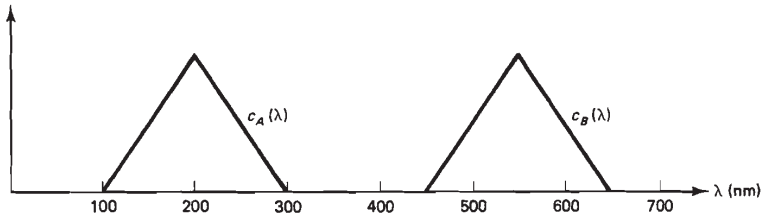


Figure P7.2

Under identical normal viewing conditions, which light will appear brighter to a human viewer?

7.3. Consider a monochromatic light with wavelength of 555 nm which has irradiance of 50 watts/m^2. We wish to have another monochromatic light with wavelength of 450 nm to appear to the C.I.E. standard observer to be as bright, under identical normal viewing conditions, as the monochromatic light with wavelength of 555 nm. Determine the irradiance of the monochromatic light with wavelength of 450 nm.

7.4. Let $c(\lambda)$ denote the energy density of a light as a function of the wavelength λ. Let A and B denote two lights with energy density of $c_A(\lambda)$ and $c_B(\lambda)$, respectively, as shown below.

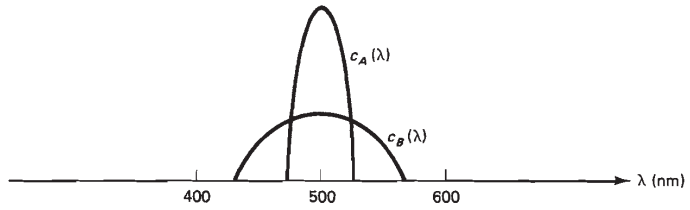


Figure P7.4

Under identical normal viewing conditions, which light will appear to be more saturated (more pure or vivid) to a human viewer?

7.5. Let the energy densities of two lights be denoted by $c_1(\lambda)$ and $c_2(\lambda)$, as shown in the following figure.

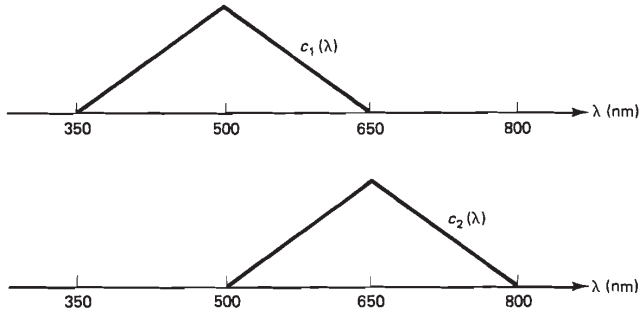


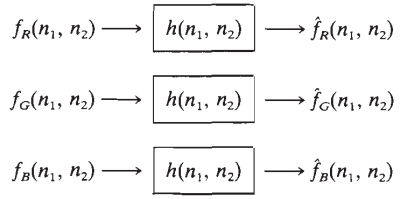
Figure P7.5

(a) Suppose we mix the two lights. Let $c(\lambda)$ denote the energy density of the combined light. Sketch $c(\lambda)$.

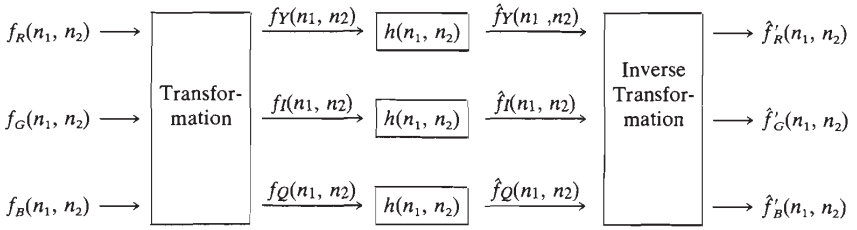
(b) Suppose that $c_1(\lambda)$ denotes the energy density of light reflected by a particular ink illuminated by white light. Suppose further that $c_2(\lambda)$ denotes the energy density of light reflected by a second ink illuminated by the same white light. We now mix the two inks in equal amounts and illuminate the mixture with the same white light. Sketch $c(\lambda)$, the light reflected by the mixture of the two inks.

7.6. A color image can be viewed as three monochrome images, which we will denote by $f_R(n_1, n_2)$, $f_G(n_1, n_2)$, and $f_B(n_1, n_2)$, representing the red, green, and blue components, respectively. Alternatively, we can transform the $f_R(n_1, n_2)$, $f_G(n_1, n_2)$, and $f_B(n_1, n_2)$

components to the luminance component $f_Y(n_1, n_2)$ and the two chrominance components $f_I(n_1, n_2)$ and $f_Q(n_1, n_2)$. Consider the system shown in the following figure.



In the above system, each of the three components $f_R(n_1, n_2)$, $f_G(n_1, n_2)$, and $f_B(n_1, n_2)$ is filtered by a linear shift-invariant system with impulse response $h(n_1, n_2)$. Now consider the system shown in the following figure.



In the figure, $h(n_1, n_2)$ is the same as in the previous system. How are $\hat{f}'_R(n_1, n_2)$, $\hat{f}'_G(n_1, n_2)$, and $\hat{f}'_B(n_1, n_2)$, related to $\hat{f}_R(n_1, n_2)$, $\hat{f}_G(n_1, n_2)$, and $\hat{f}_B(n_1, n_2)$?

7.7. Consider a light represented in terms of its three components, red (R), green (G), and blue (B). Suppose the intensities R, G, and B are in the following ranges:

$$0 \le R \le R_{MAX}$$

$$0 \le G \le G_{MAX}$$

$$0 \le B \le B_{MAX}$$

Note that R, G, and B are nonnegative. Another way of representing the same light is in terms of the luminance component Y and the two chrominance components I and Q.

(a) The possible ranges of Y, I, and Q are

$$Y_{MIN} \le Y \le Y_{MAX}$$

$$I_{MIN} \le I \le I_{MAX}$$

$$Q_{MIN} \le Q \le Q_{MAX}$$

Determine Y_{MIN}, Y_{MAX}, I_{MIN}, I_{MAX}, Q_{MIN}, and Q_{MAX}.

(b) Does any set of Y, I, and Q in the ranges obtained in (a) correspond to a valid light, valid in the sense that it corresponds to a set of nonnegative R, G, and B?

7.8 Figure 7.22 illustrates that the peripheral level of the human visual system is approximately bandpass in character. Design an experiment that can be used in more accurately determining the frequency response of the bandpass filter. State the assumptions that you make in designing your experiment.

7.9. Weber's law, which states that the just-noticeable intensity difference ΔI at a given intensity I is proportional to I, appears to hold over a wide range of I not only for human vision, but for all other sense modalities, including hearing, smell, taste, and touch. Discuss why evolution may have favored such a mechanism rather than a system whereby the just-noticeable intensity difference ΔI is constant independent of I.

7.10. Supppose Weber's law holds strictly. The just-noticeable intensity difference ΔI when $I = 100$ has been measured to be 1. How many just-noticeable differences are there between $I = 10$ and $I = 1000$?

7.11. A physical unit of spatial frequency is cycles/cm. A unit of spatial frequency more directly related to our perception of spatial frequency is cycles/viewing angle. Suppose we have a sinusoidal intensity grating at a horizontal frequency of 10 cycles/cm. In answering the following questions, assume that the image is focused on the retina, the sinusoidal grating is directly displayed before an eye, and the eye looks at the image straight on.

(a) When the image is viewed at a distance of 50 cm away from the eye, what is the horizontal spatial frequency in units of cycles/degree in viewing angle?

(b) Answer (a) when the image is viewed at a distance of 100 cm from the eye.

(c) From the results of (a) and (b), and the spatial frequency response of a human eye, discuss why the details of a distant scene are less visible than those of nearby objects.

7.12. The lens of a human eye is convex. As a result, the image formed on the retina is upside down, as shown in the following figure. Nevertheless, the object is perceived to be right side up. Discuss possible explanations for this phenomenon.

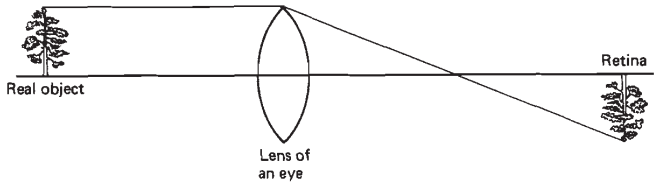


Figure P7.12

7.13. In the region on the retina known as the blind spot, there are no light-sensitive cells, so light focused on this region is not visible. The blind spot often goes unnoticed in our everyday experience. Design an experiment that can be used to perceive the blind spot.

7.14. In a typical image digitization system, a small aperture $A(x, y)$ searches an image following a certain pattern called a raster. Let $A(x, y)$ be given by

$$A(x, y) = \begin{cases} 1, & \sqrt{x^2 + y^2} < r \\ 0, & \text{otherwise} \end{cases}$$

The light intensity integrated over the small aperture is measured and is considered the image intensity at the spatial point.

(a) Discuss how this process can be viewed as convolving the input intensity $I(x, y)$ with the aperture $A(x, y)$ and then sampling the result of convolution.