point. If the image is formed by reflected light intensity, as in a photograph, the image records both light from primary light sources and (more usually) the light reflected off physical surfaces. We show in Chapter 3 that in certain cases we can use these kinds of images together with knowledge about physics to derive the orientation of the surfaces. If, on the other hand, the image is a computed tomogram of the human body (discussed in Section 2.3.4), the image represents tissue density of internal organs. Here orientation calculations are irrelevant, but general segmentation techniques of Chapters 4 and 5 (the agglomeration of neighboring samples of similar density into units representing organs) are appropriate.

2.2 IMAGE MODEL

Sophisticated image models of a statistical flavor are useful in image processing [Jan 1981]. Here we are concerned with more geometrical considerations.

2.2.1 Image Functions

An *image function* is a mathematical representation of an image. Generally, an image function is a vector-valued function of a small number of arguments. A special case of the image function is the *digital (discrete) image function*, where the arguments to and value of the function are all integers. Different image functions may be used to represent the same image, depending on which of its characteristics are important. For instance, a camera produces an image on black-and-white film which is usually thought of as a real-valued function (whose value could be the density of the photographic negative) of two real-valued arguments, one for each of two spatial dimensions. However, at a very small scale (the order of the film grain) the negative basically has only two densities, "opaque" and "transparent."

Most images are presented by functions of two *spatial* variables $f(\mathbf{x}) = f(x, y)$, where $f(x, y)$ is the brightness of the gray level of the image at a spatial coordinate (x, y). A multispectral image $\mathbf{f}$ is a vector-valued function with components $(f_1 \ldots f_n)$. One special multispectral image is a color image in which, for example, the components measure the brightness values of each of three wavelengths, that is,

$$\mathbf{f}(\mathbf{x}) = \left\{ f_{\text{red}}(\mathbf{x}), f_{\text{blue}}(\mathbf{x}), f_{\text{green}}(\mathbf{x}) \right\}$$

Time-varying images $\mathbf{f}(\mathbf{x}, t)$ have an added temporal argument. For special three-dimensional images, $\mathbf{x} = (x, y, z)$. Usually, both the domain and range of $\mathbf{f}$ are bounded.

An important part of the formation process is the conversion of the image representation from a continuous function to a discrete function; we need some way of describing the images as samples at discrete points. The mathematical tool we shall use is the *delta function.*

Formally, the delta function may be defined by

$$\delta(x) = \begin{cases} 0 \text{ when } x \neq 0 \\ \infty \text{ when } x = 0 \end{cases} \qquad (2.1)$$

$$\int_{-\infty}^{\infty} \delta(x)\, dx = 1$$

If some care is exercised, the delta function may be interpreted as the limit of a set of functions:

$$\delta(x) = \lim_{n \to \infty} \delta_n(x)$$

wh^ere

$$\delta_n(x) = \begin{cases} n & \text{if } |x| < \dfrac{1}{2n} \\ 0 & \text{otherwise} \end{cases} \qquad (2.2)$$

A useful property of the delta function is the *sifting property:*

$$\int_{-\infty}^{\infty} f(x)\,\delta(x-a)\, dx = f(a) \qquad (2.3)$$

A continuous image may be multipled by a two-dimensional "comb," or array of delta functions, to extract a finite number of discrete *samples* (one for each delta function). This mathematical model of the sampling process will be useful later.

2.2.2 Imaging Geometry

Monocular Imaging

Point projection is the fundamental model for the transformation wrought by our eye, by cameras, or by numerous other imaging devices. To a first-order approximation, these devices act like a pinhole camera in that the image results from projecting scene points through a single point onto an *image plane* (see Fig. 2.1). In Fig. 2.1, the image plane is behind the point of projection, and the image is reversed. However, it is more intuitive to recompose the geometry so that the point of projection corresponds to a *viewpoint* behind the image plane, and the image occurs right side up (Fig. 2.2). The mathematics is the same, but now the viewpoint is $+f$ on the z axis, with $z = 0$ plane being the image plane upon which the image is projected. (f is sometimes called the *focal length* in this context. The use of f in this section should not be confused with the use of f for image function.) As the imaged object approaches the viewpoint, its projection gets bigger (try moving your hand toward your eye). To specify how its imaged size changes, one needs only the geometry of similar triangles. In Fig. 2.2b y', the projected height of the object, is related to its real height y, its position z, and the focal length f by

$$\frac{y}{f-z} = \frac{y'}{f} \qquad (2.4)$$

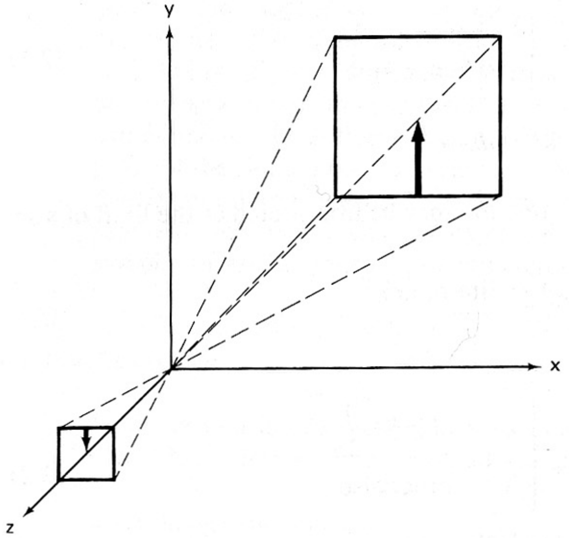


Fig. 2.1 A geometric camera model.

The case for x' is treated similarly:

$$\frac{x}{f - z} = \frac{x'}{f} \tag{2.5}$$

The projected image has $z = 0$ everywhere. However, projecting away the z component is best considered a separate transformation; the projective transform is usually thought to distort the z component just as it does the x and y. *Perspective distortion* thus maps (x, y, z) to

$$(x', y', z') = \left(\frac{fx}{f - z}, \frac{fy}{f - z}, \frac{fz}{f - z}\right) \tag{2.6}$$

The perspective transformation yields *orthographic projection* as a special case when the viewpoint is the *point at infinity* in the z direction. Then all objects are projected onto the viewing plane with no distortion of their x and y coordinates.

The perspective distortion yields a three-dimensional object that has been "pushed out of shape"; it is more shrunken the farther it is from the viewpoint. The z component is not available directly from a two-dimensional image, being identically equal to zero. In our model, however, the distorted z component has information about the distance of imaged points from the viewpoint. When this distorted object is projected orthographically onto the image plane, the result is a perspective picture. Thus, to achieve the effect of railroad tracks appearing to come together in the distance, the perspective distortion transforms the tracks so that they *do* come together (at a point at infinity)! The simple orthographic projection that projects away the z component unsurprisingly preserves this distortion. Several properties of the perspective transform are of interest and are investigated further in Appendix 1.

Binocular Imaging

Basic binocular imaging geometry is shown in Fig. 2.3a. For simplicity, we

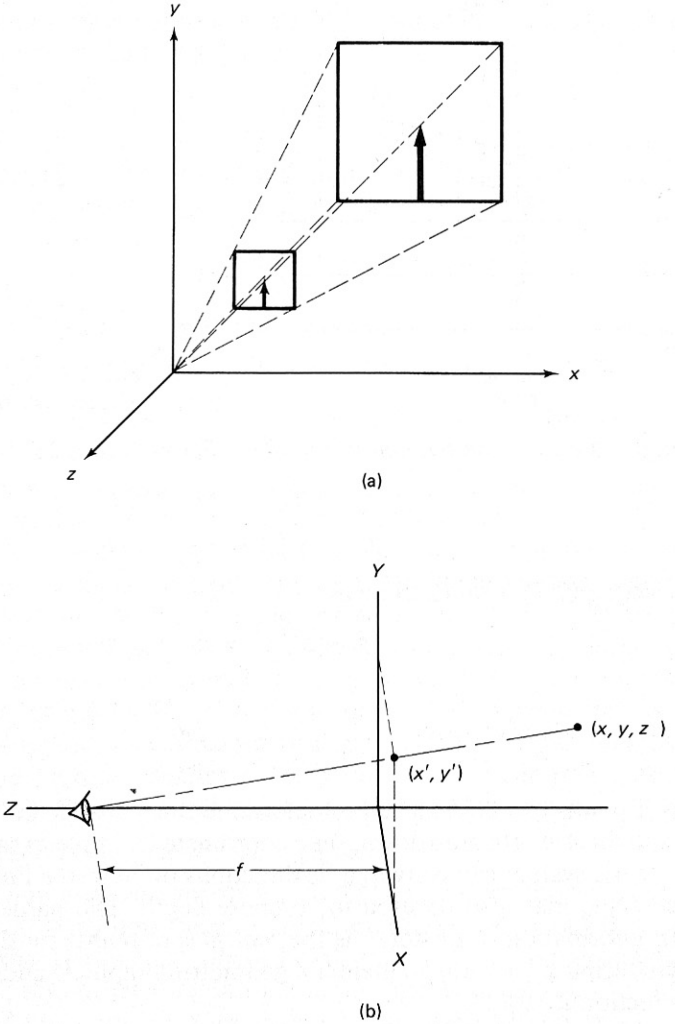


Fig. 2.2 (a) Camera model equivalent to that of Fig. 2.1; (b) definition of terms.

use a system with two viewpoints. In this model the eyes do not *converge*; they are aimed in parallel at the point at infinity in the $-z$ direction. The depth information about a point is then encoded only by its different positions (*disparity*) in the two image planes.

With the stereo arrangement of Fig. 2.3,

$$x' = \frac{(x - d)f}{f - z}$$

$$x'' = \frac{(x + d)f}{f - z}$$

where (x', y') and (x'', y'') are the retinal coordinates for the world point imaged

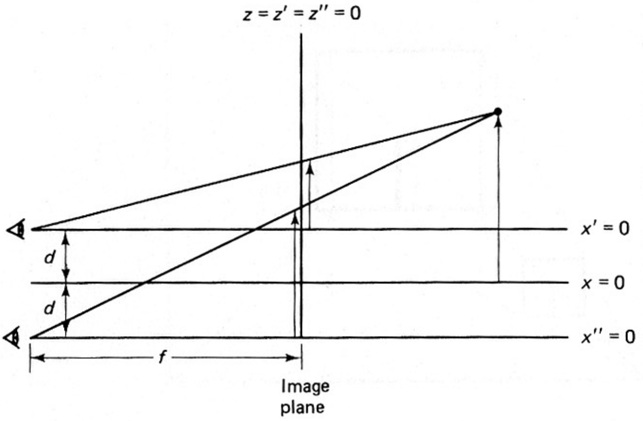


Fig. 2.3 A nonconvergent binocular imaging system.

through each eye. The *baseline* of the binocular system is $2d$. Thus

$$(f - z)x' = (x - d)f \tag{2.7}$$

$$(f - z)x'' = (x + d)f \tag{2.8}$$

Subtracting (2.7) from (2.8) gives

$$(f - z)(x'' - x') = 2df$$

or

$$z = f - \frac{2df}{x'' - x'} \tag{2.9}$$

Thus if points can be matched to determine the disparity $(x'' - x')$ and the baseline and focal length are known, the z coordinate is simple to calculate.

If the system can converge its directions of view to a finite distance, convergence angle may also be used to compute depth. The hardest part of extracting depth information from stereo is the *matching* of points for disparity calculations. "Light striping" is a way to maintain geometric simplicity and also simplify matching (Section 2.3.3).

2.2.3 Reflectance

Terminology

A basic aspect of the imaging process is the physics of the reflectance of objects, which determines how their "brightness" in an image depends on their inherent characteristics and the geometry of the imaging situation. A clear presentation of the mathematics of reflectance is given in [Horn and Sjoberg 1978; Horn 1977]. Light *energy flux* Φ is measured in watts; "brightness" is measured with respect to area and solid angle. The *radiant intensity* I of a source is the exitant flux per unit solid angle:

$$I = \frac{d\Phi}{d\omega} \qquad \text{watts/steradian} \tag{2.10}$$

Here $d\omega$ is an incremental solid angle. The solid angle of a small area dA measured perpendicular to a radius r is given by

$$d\omega = \frac{dA}{r^2} \tag{2.11}$$

in units of steradians. (The total solid angle of a sphere is 4π.)

The *irradiance* is flux incident on a surface element dA:

$$E = \frac{d\Phi}{dA} \qquad \text{watts/meter}^2 \tag{2.12}$$

and the flux exitant from the surface is defined in terms of the *radiance* L, which is the flux emitted per unit foreshortened surface area per unit solid angle:

$$L = \frac{d^2\Phi}{dA\ \cos\theta\, d\omega} \qquad \text{watts/(meter}^2\text{ steradian)} \tag{2.13}$$

where θ is the angle between the surface normal and the direction of emission.

Image irradiance f is the "brightness" of the image at a point, and is proportional to scene radiance. A "gray-level" is a quantized measurement of image irradiance. Image irradiance depends on the reflective properties of the imaged surfaces as well as on the illumination characteristics. How a surface reflects light depends on its micro-structure and physical properties. Surfaces may be *matte* (dull, flat), *specular* (mirrorlike), or have more complicated reflectivity characteristics (Section 3.5.1). The *reflectance* r of a surface is given quite generally by its Bidirectional Reflectance Distribution Function (BRDF) [Nicodemus et al. 1977]. The BRDF is the ratio of reflected radiance in the direction towards the viewer to the irradiance in the direction towards a small area of the source.

Effects of Geometry on an Imaging System

Let us now analyze a simple image-forming system shown in Fig. 2.4 with the objective of showing how the gray levels are related to the radiance of imaged objects. Following [Horn and Sjoberg 1978], assume that the imaging device is properly focused; rays originating in the infinitesimal area dA_o on the object's surface are projected into some area dA_p in the image plane and no rays from other portions of the object's surface reach this area of the image. The system is assumed to be an ideal one, obeying the laws of simple geometrical optics.

The energy flux/unit area that impinges on the sensor is defined to be E_p. To show how E_p is related to the scene radiance L, first consider the flux arriving at the lens from a small surface area dA_o. From (2.13) this is given as

$$d\Phi = dA_o \int L \cos\theta\, d\omega \tag{2.14}$$

This flux is assumed to arrive at an area dA_p in the imaging plane. Hence the irradiance is given by [using Eq. (2.12)]

$$E_p = \frac{d\Phi}{dA_p} \tag{2.15}$$

Now relate dA_o to dA_p by equating the respective solid angles as seen from the lens; that is [making use of Eq. (2.12)],

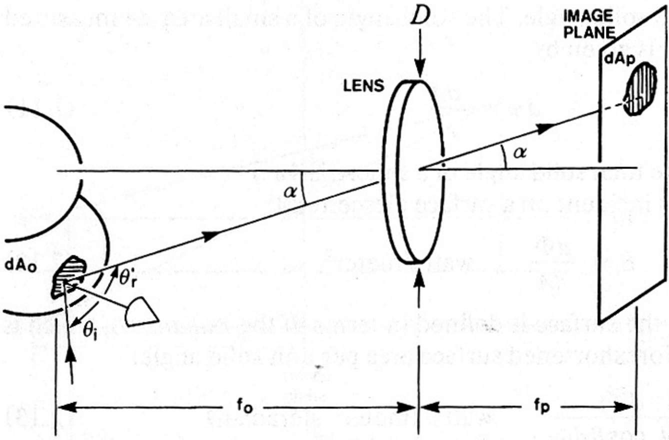


Fig. 2.4 Geometry of an image forming system.

$$dA_o \frac{\cos\theta}{f_o^2} = dA_p \frac{\cos\alpha}{f_p^2} \tag{2.16}$$

Substituting Eqs. (2.16) and (2.14) into (2.15) gives

$$E = \cos\alpha \left(\frac{f_o}{f_p} \right)^2 \int L d\omega \tag{2.17}$$

The integral is over the solid angle seen by the lens. In most instances we can assume that L is constant over this angle and hence can be removed from the integral. Finally, approximate $d\omega$ by the area of the lens foreshortened by $\cos\alpha$, that is, $(\pi/4)D^2 \cos\alpha$ divided by the distance $f_o/\cos\alpha$ squared:

$$d\omega = \frac{\pi}{4} D^2 \frac{\cos^3\alpha}{f_o^2} \tag{2.18}$$

so that finally

$$E = \frac{1}{4} \left(\frac{D}{f_p} \right)^2 \cos^4\alpha \, \pi L \tag{2.19}$$

The interesting results here are that (1) the image irradiance is proportional to the scene radiance L, and (2) the factor of proportionality includes the fourth power of the off-axis angle α. Ideally, an imaging device should be calibrated so that the variation in sensitivity as a function of α is removed.

2.2.4 Spatial Properties

The Fourier Transform

An image is a spatially varying function. One way to analyze spatial variations is the decomposition of an image function into a set of orthogonal functions, one such set being the Fourier (sinusoidal) functions. The Fourier transform may be used to transform the intensity image into the domain of *spatial frequency.* For notational convenience and intuition, we shall generally use as an example the continuous one-dimensional Fourier transform. The results can readily be extended to the discrete case and also to higher dimensions [Rosenfeld and Kak 1976]. In two dimensions we shall denote transform domain coordinates by (u, v). The one-dimensional Fourier transform, denoted $\mathfrak{F}$, is defined by

$$\mathfrak{F}\,[f(x)] = F(u)$$

where

$$F(u) = \int_{-\infty}^{+\infty} f(x) \exp(-j2\pi ux)\,dx \tag{2.20}$$

where $j = \sqrt{(-1)}$. Intuitively, Fourier analysis expresses a function as a sum of sine waves of different frequency and phase. The Fourier transform has an *inverse* $^{-1}[F(u)] = f(x)$. This inverse is given by

$$f(x) = \int_{-\infty}^{\infty} F(u) \exp(j2\pi ux)\,du \tag{2.21}$$

The transform has many useful properties, some of which are summarized in Table 2.1. Common one-dimensional Fourier transform pairs are shown in Table 2.2.

The transform F(u) is simply another representation of the image function. Its meaning can be understood by interpreting Eq. (2.21) for a specific value of x, say x_0:

$$f(x_0) = \int F(u) \exp(j2\pi ux_0)\,du \tag{2.22}$$

This equation states that a particular point in the image can be represented by a weighted sum of complex exponentials (sinusoidal patterns) at different spatial frequencies u. $F(u)$ is thus a *weighting function* for the different frequencies. Low-spatial frequencies account for the "slowly" varying gray levels in an image, such as the variation of intensity over a continuous surface. High-frequency components are associated with "quickly varying" information, such as edges. Figure 2.5 shows the Fourier transform of an image of rectangles, together with the effects of removing low- and high-frequency components.

The Fourier transform is defined above to be a continuous transform. Although it may be performed instantly by optics, a discrete version of it, the "fast Fourier transform," is almost universally used in image processing and computer vision. This is because of the relative versatility of manipulating the transform in the digital domain as compared to the optical domain. Image-processing texts, e.g., [Pratt 1978; Gonzalez and Wintz 1977] discuss the FFT in some detail; we content ourselves with an algorithm for it (Appendix 1).

The Convolution Theorem

Convolution is a very important image-processing operation, and is a basic operation of linear systems theory. The convolution of two functions f and g is a function h of a displacement y defined as

$$h(y) = f*g = \int_{-\infty}^{\infty} f(x) g(y-x)\,dx \tag{2.23}$$

Table 2.1

PROPERTIES OF THE FOURIER TRANSFORM

	Spatial Domain	*Frequency Domain*
	$f(x)$ $g(x)$	$F(u) = \mathfrak{F}[f(x)]$ $G(u) = \mathfrak{F}[g(x)]$
(1)	Linearity $c_1 f(x) + c_2 g(x)$ c_1, c_2 scalars	$c_1 F(u) + c_2 G(u)$
(2)	Scaling $f(ax)$	$\frac{1}{\lvert a \rvert} F\left(\frac{u}{a}\right)$
(3)	Shifting $f(x - x_0)$	$e^{-2\pi j x_0} F(u)$
(4)	Symmetry $F(x)$	$f(-u)$
(5)	Conjugation $f^*(x)$	$F^*(-u)$
(6)	Convolution $h(x) = f * g = \int_{-\infty}^{\infty} f(x') g(x - x')\, dx'$	$F(u) G(u)$
(7)	Differentiation $\frac{d^n f(x)}{dx^n}$	$(2\pi j u)^n F(u)$

Parseval's theorem:

$$\int_{-\infty}^{\infty} |f(x)|^2 dx = \int_{-\infty}^{\infty} |F(\xi)|^2 d\xi$$

$$\int_{-\infty}^{\infty} f(x) g^*(x)\, dx = \int_{-\infty}^{\infty} F(\xi) G^*(\xi)\, d\xi$$

$f(x)$	$F(\xi)$
Real(R)	Real part even (RE) Imaginary part odd (IO)
Imaginary (I)	RO,IE
RE,IO	R
RE,IE	I
RE	RE
RO	IO
IE	IE
IO	RO
Complex even (CE)	CE
CO	CO

Table 2.2

FOURIER TRANSFORM PAIRS

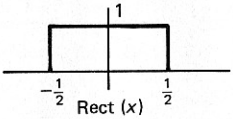
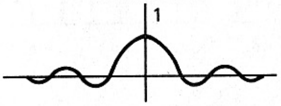
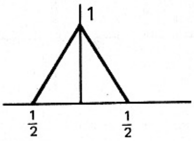
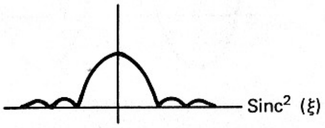
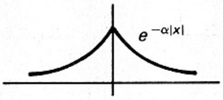
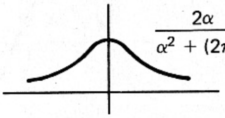
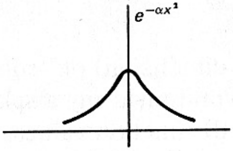
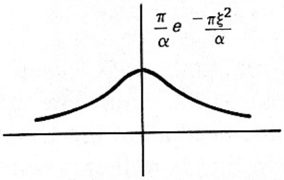
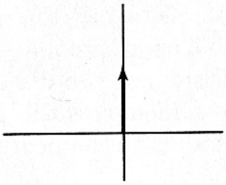
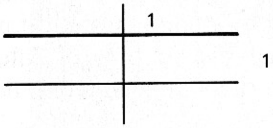
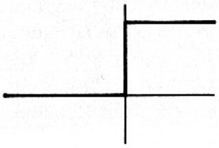
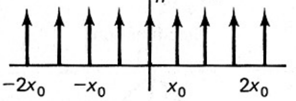
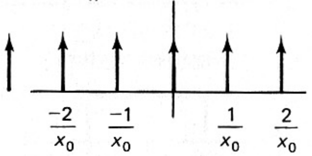
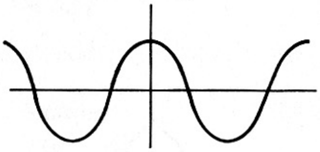
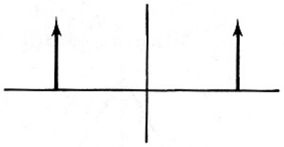
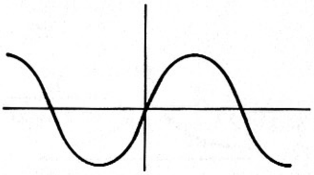
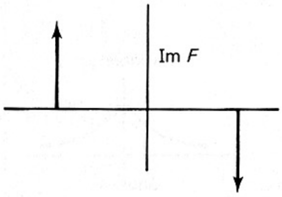
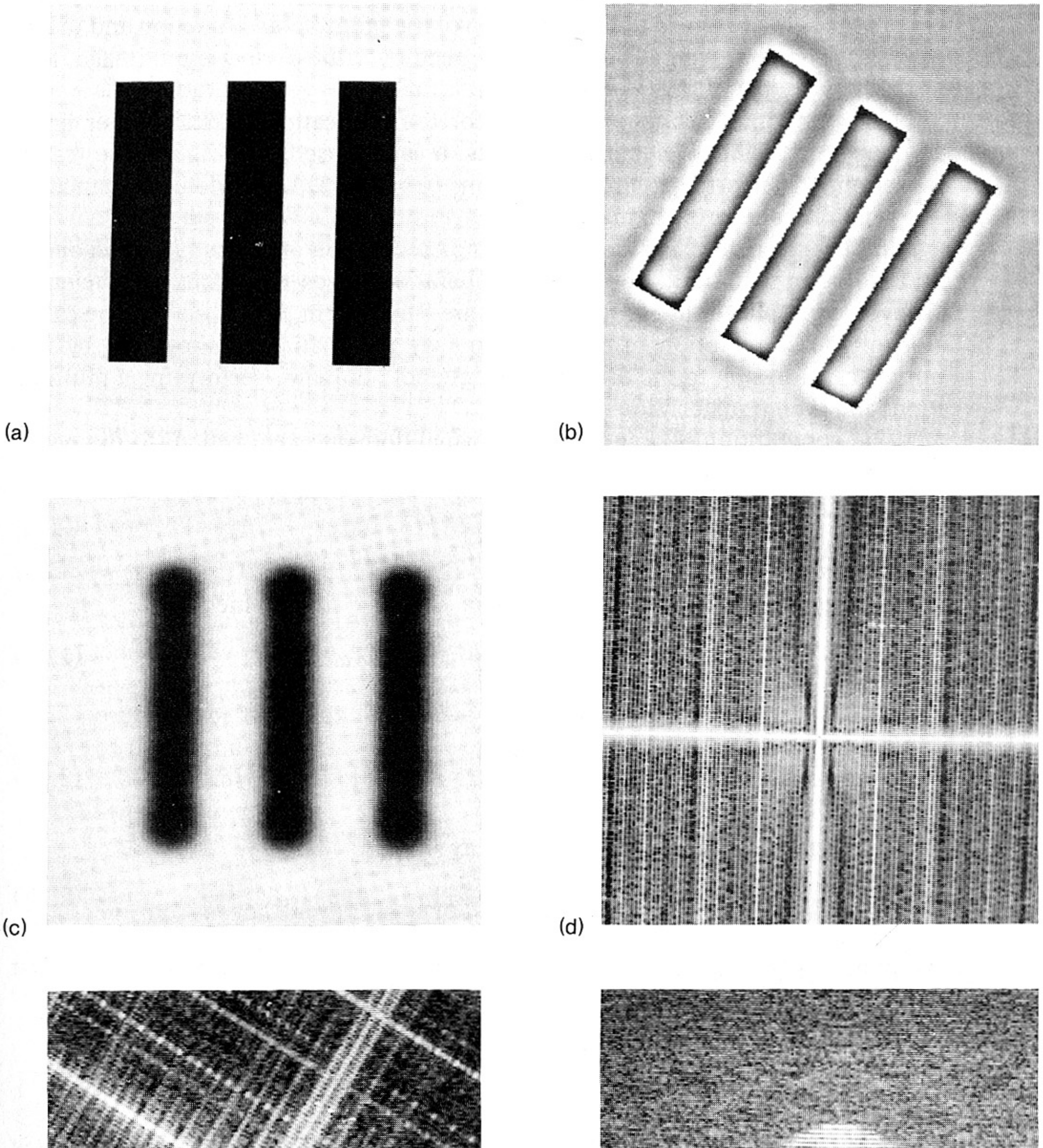
$f(x)$	$F(\xi)$
Rectangle function  	Sinc function  $\text{Sinc}\,(\xi) = \frac{\sin \pi\xi}{\pi\xi}$
Triangle function  	
Exponential  	 $\frac{2\alpha}{\alpha^2 + (2\pi\xi)^2}$
Gaussian  	 $\frac{\pi}{\alpha} e^{-\frac{\pi\xi^2}{\alpha}}$
Unit impulse $\delta(x)$ 	
Unit step 	$\frac{1}{2}\,\delta\,(\xi) + \frac{1}{2\pi j\xi}$

Table 2.2 (cont.)

Comb function $\sum_{n=-\infty}^{\infty} \delta(x - nx_0)$  	$\frac{1}{x_0} \sum_{n=-\infty}^{\infty} \delta(\xi - \frac{n}{x_0})$  $\frac{-2}{x_0}$ $\frac{-1}{x_0}$ $\frac{1}{x_0}$ $\frac{2}{x_0}$
$\cos 2\pi\omega_0 x$ 	$\frac{1}{2}[\,\delta(\xi - \omega_0) + \delta(\xi + \omega_0)\,]$ 
$\sin 2\pi\omega_0 x$ 	$\frac{1}{2}\,j\,[-\delta(\xi - \omega_0) + \delta(\xi + \omega_0)\,]$ 

Intuitively, one function is "swept past" (in one dimension) or "rubbed over" (in two dimensions) the other. The value of the convolution at any displacement is the integral of the product of the (relatively displaced) function values. One common phenomenon that is well expressed by a convolution is the formation of an image by an optical system. The system (say a camera) has a "point-spread function," which is the image of a single point. (In linear systems theory, this is the "impulse response," or response to a delta-function input.) The ideal point-spread function is, of course, a point. A typical point-spread function is a two-dimensional Gaussian spatial distribution of intensities, but may include such phenomena as diffraction rings. In any event, if the camera is modeled as a linear system (ignor-

Fig. 2.5 (on facing page) (a) An image, $f(x, y)$. (b) A rotated version of (a), filtered to enhance high spatial frequencies. (c) Similar to (b), but filtered to enhance low spatial frequencies. (d), (e), and (f) show the logarithm of the power spectrum of (a), (b), and (c). The power spectrum is the log square modulus of the Fourier transform $F(u, v)$. Considered in polar coordinates (ρ, θ), points of small ρ correspond to low spatial frequencies ("slowly-varying" intensities), large ρ to high spatial frequencies contributed by "fast" variations such as step edges. The power at (ρ, θ) is determined by the amount of intensity variation at the frequency ρ occurring at the angle θ.


(a)
(b)
(c)
(d)
(e)
(f)

ing the added complexity that the point-spread function usually varies over the field of view), the image is the convolution of the point-spread function and the input signal. The point-spread function is rubbed over the perfect input image, thus blurring it.

Convolution is also a good model for the application of many other linear operators, such as line-detecting templates. It can be used in another guise (called correlation) to perform matching operations (Chapter 3) which detect instances of subimages or features in an image.

In the spatial domain, the obvious implementation of the convolution operation involves a shift–multiply–integrate operation which is hard to do efficiently. However, multiplication and convolution are "transform pairs," so that the calculation of the convolution in one domain (say the spatial) is simplified by first Fourier transforming to the other (the frequency) domain, performing a multiplication, and then transforming back.

The convolution of f and g in the spatial domain is equivalent to the pointwise product of F and G in the frequency domain,

$$\mathfrak{F}(f*g) = FG \tag{2.24}$$

We shall show this in a manner similar to [Duda and Hart 1973]. First we prove the *shift theorem.* If the Fourier transform of $f(x)$ is $F(u)$, defined as

$$F(u) = \int_x f(x) \exp[-j2\pi(ux)]dx \tag{2.25}$$

then

$$\mathfrak{F}[f(x-a)] = \int_x f(x-a) \exp[-j2\pi(ux)]dx \tag{2.26}$$

changing variables so that $x' = x - a$ and $dx = dx'$

$$= \int_{x'} f(x') \exp\{-j2\pi[u(x'+a)]\}dx' \tag{2.27}$$

Now $\exp[-j2\pi u(x'+a)] = \exp(-j2\pi ua)\exp(-j2\pi ux')$, where the first term is a constant. This means that

$$\mathfrak{F}[f(x-a)] = \exp(-j2\pi ua)F(u) \qquad \text{(shift theorem)}$$

Now we are ready to show that $\mathfrak{F}[f(x)*g(x)] = F(u)G(u)$.

$$\mathfrak{F}(f*g) = \int_y \{\int_x f(x)g(y-x)\} \exp(-j2\pi uy)\, dx\, dy \tag{2.28}$$

$$= \int_x f(x)\{\int_y g(y-x) \exp(-j2\pi uy)\, dy\}\, dx \tag{2.29}$$

Recognizing that the terms in braces represent $\mathfrak{F}[g(y-x)]$ and applying the shift theorem, we obtain

$$\mathfrak{F}(f*g) = \int_x f(x)\exp(-j2\pi ux)G(u)\, dx \tag{2.30}$$

$$= F(u)G(u) \tag{2.31}$$

2.2.5 Color

Not all images are monochromatic; in fact, applications using multispectral images are becoming increasingly common (Section 2.3.2). Further, human beings intuitively feel that color is an important part of their visual experience, and is useful or even necessary for powerful visual processing in the real world. Color vision provides a host of research issues, both for psychology and computer vision. We briefly discuss two aspects of color vision: color spaces and color perception. Several models of the human visual system not only include color but have proven useful in applications [Granrath 1981].

Color Spaces

Color spaces are a way of organizing the colors perceived by human beings. It happens that weighted combinations of stimuli at three principal wavelengths are sufficient to define almost all the colors we perceive. These wavelengths form a natural basis or coordinate system from which the color measurement process can be described. Color perception is not related in a simple way to color measurement, however.

Color is a perceptual phenomenon related to human response to different wavelengths in the visible *electromagnetic spectrum* [400 (blue) to 700 nanometers (red); a nanometer (nm) is 10^{-9} meter]. The sensation of color arises from the sensitivities of three types of neurochemical sensors in the retina to the visible spectrum. The relative response of these sensors is shown in Fig. 2.6. Note that each sensor responds to a range of wavelengths. The illumination source has its own spectral composition $f(\lambda)$ which is modified by the reflecting surface. Let $r(\lambda)$ be this reflectance function. Then the measurement R produced by the "red" sensor is given by

$$R = \int f(\lambda) r(\lambda) h_R(\lambda)\ d\lambda \tag{2.32}$$

So the sensor output is actually the integral of three different wavelength-dependent components: the source f, the surface reflectance r, and the sensor h_R.

Surprisingly, only weighted combinations of three delta-function approximations to the different $f(\lambda) h(\lambda)$, that is, $\delta(\lambda_R)$, $\delta(\lambda_G)$, and $\delta(\lambda_B)$, are necessary to

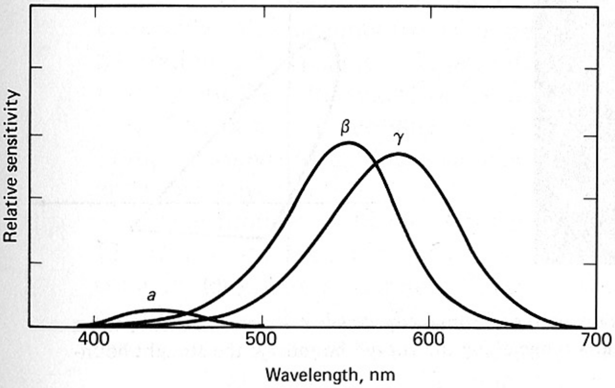


Fig. 2.6 Spectral response of human color sensors.

produce the sensation of nearly all the colors. This result is displayed on a *chromaticity diagram.* Such a diagram is obtained by first normalizing the three sensor measurements:

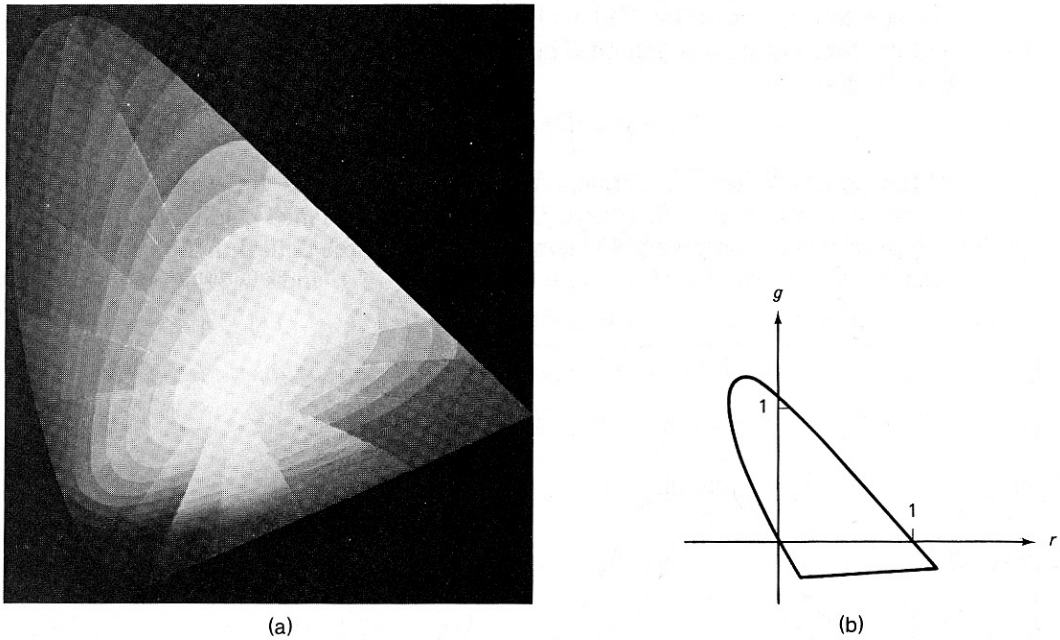
$$
\begin{aligned}
r &= \frac{R}{R+G+B} \\
g &= \frac{G}{R+G+B} \qquad (2.33) \\
b &= \frac{B}{R+G+B}
\end{aligned}
$$

and then plotting perceived color as a function of any two (usually red and green). Chromaticity explicitly ignores intensity or brightness; it is a section through the three-dimensional color space (Fig. 2.7). The choice of $(\lambda_R, \lambda_G, \lambda_B) = (410, 530, 650)\,nm$ maximizes the realizable colors, but some colors still cannot be realized since they would require negative values for some of r, g, and b.

Another more intuitive way of visualizing the possible colors from the *RGB* space is to view these measurements as Euclidean coordinates. Here any color can be visualized as a point in the unit cube. Other coordinate systems are useful for different applications; computer graphics has proved a strong stimulus for investigation of different color space bases.

Color Perception

Color perception is complex, but the essential step is a transformation of three input intensity measurements into another basis. The coordinates of the new



(a) (b)

Fig. 2.7 (a) An artist's conception of the chromaticity diagram—*see color insert*; (b) a more useful depiction. Spectral colors range along the curved boundary; the straight boundary is the line of purples.

basis are more directly related to human color judgments.

Although the *RGB* basis is good for the acquisition or display of color information, it is not a particularly good basis to explain the perception of colors. Human vision systems can make good judgments about the relative surface reflectance $r(\lambda)$ despite different illuminating wavelengths; this reflectance seems to be what we mean by surface color.

Another important feature of the color basis is revealed by an ability to perceive in "black and white," effectively deriving intensity information from the color measurements. From an evolutionary point of view, we might expect that color perception in animals would be compatible with preexisting noncolor perceptual mechanisms.

These two needs—the need to make good color judgments and the need to retain and use intensity information—imply that we use a transformed, non-*RGB* basis for color space. Of the different bases in use for color vision, all are variations on this theme: Intensity forms one dimension and color is a two-dimensional subspace. The differences arise in how the color subspace is described. We categorize such bases into two groups.

1. *Intensity/Saturation/Hue (IHS).* In this basis, we compute intensity as

$$\text{intensity:} = R + G + B \tag{2.34}$$

The saturation measures the lack of whiteness in the color. Colors such as "fire engine" red and "grass" green are saturated; pastels (e.g., pinks and pale blues) are desaturated. Saturation can be computed from *RGB* coordinates by the formula [Tenenbaum and Weyl 1975]

$$\text{saturation:} = 1 - \frac{3 \min (R,\ G,\ B)}{\text{intensity}} \tag{2.35}$$

Hue is roughly proportional to the average wavelength of the color. It can be defined using *RGB* by the following program fragment:

$$\text{hue:} = \cos^{-1}\left\{\frac{\{\tfrac{1}{2}[(R - G) + (R - B)]\}}{\sqrt{(R - G)^2 + (R - B)(G - B)}^{1/2}}\right\} \tag{2.36}$$

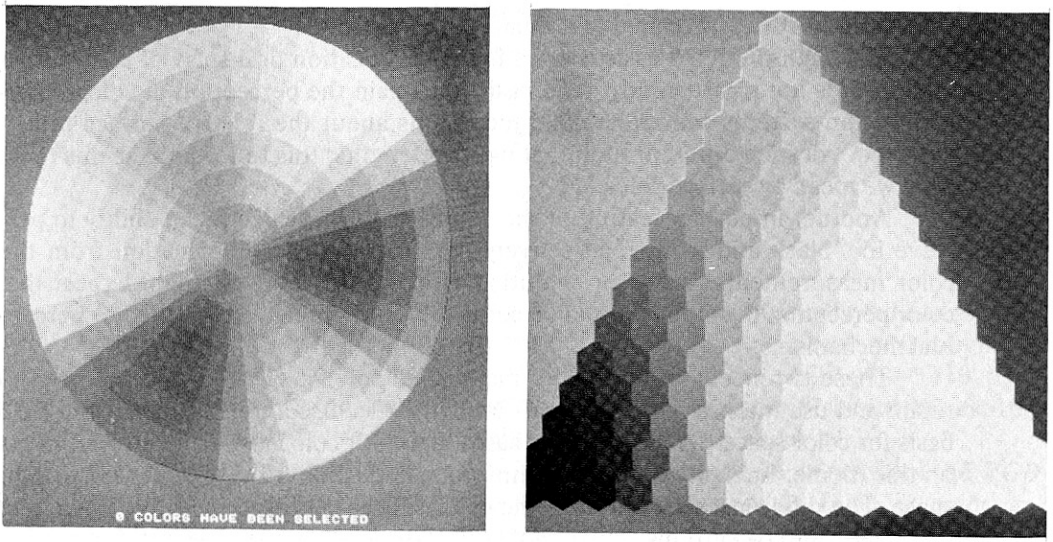
$$\text{If } B > G \text{ then hue:} = 2pi - \text{hue}$$

The IHS basis transforms the *RGB* basis in the following way. Thinking of the color cube, the diagonal from the origin to (1, 1, 1) becomes the intensity axis. Saturation is the distance of a point from that axis and hue is the angle with regard to the point about that axis from some reference (Fig. 2.8).

This basis is essentially that used by artists [Munsell 1939], who term saturation *chroma.* Also, this basis has been used in graphics [Smith 1978; Joblove and Greenberg 1978].

One problem with the IHS basis, particularly as defined by (2.34) through (2.36), is that it contains essential singularities where it is impossible to define the color in a consistent manner [Kender 1976]. For example, hue has an essential singularity for all values of $(R,\ G,\ B)$, where $R = G = B$. This means that special care must be taken in algorithms that use hue.

2. *Opponent processes.* The opponent process basis uses Cartesian rather than



(a) (b)

Fig. 2.8 An IHS Color Space. (a) Cross section at one intensity; (b) cross section at one hue—*see color inserts.*

cylindrical coordinates for the color subspace, and was first proposed by Hering [Teevan and Birney 1961]. The simplest form of basis is a linear transformation from R, G, B coordinates. The new coordinates are termed "$R - G$", "$Bl - Y$", and "$W - Bk$":

$$\begin{bmatrix} R - G \\ Bl - Y \\ W - Bk \end{bmatrix} = \begin{bmatrix} 1 & -2 & 1 \\ -1 & -1 & 2 \\ 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} R \\ G \\ B \end{bmatrix}$$

The advocates of this representation, such as [Hurvich and Jameson 1957], theorize that this basis has neurological correlates and is in fact the way human beings represent ("name") colors. For example, in this basis it makes sense to talk about a "reddish blue" but not a "reddish green." Practical opponent process models usually have more complex weights in the transform matrix to account for psychophysical data. Some startling experiments [Land 1977] show our ability to make correct color judgments even when the illumination consists of only two principal wavelengths. The opponent process, at the level at which we have developed it, does not demonstrate how such judgments are made, but does show how stimulus at only two wavelengths will project into the color subspace. Readers interested in the details of the theory should consult the references.

Commercial television transmission needs an intensity, or "$W - Bk$" component for black-and-white television sets while still spanning the color space. The National Television Systems Committee (NTSC) uses a "YIQ" basis extracted from RGB via

$$\begin{bmatrix} I \\ Q \\ Y \end{bmatrix} = \begin{bmatrix} 0.60 & -0.28 & -0.32 \\ 0.21 & -0.52 & 0.31 \\ 0.30 & 0.59 & 0.11 \end{bmatrix} \begin{bmatrix} R \\ G \\ B \end{bmatrix}$$

This basis is a weighted form of

$$(I,\ Q,\ Y) = (\text{“}R\text{–cyan,” “magenta–green,” “}W\text{–}Bk\text{”})$$

2.2.6 Digital Images

The *digital images* with which computer vision deals are represented by m-vector discrete-valued image functions $\boldsymbol{f}(\mathbf{x})$, usually of one, two, three, or four dimensions.

Usually $m = 1$, and both the domain and range of $f(\mathbf{x})$ are discrete. The domain of f is finite, usually a rectangle, and the range of f is positive and bounded: $0 \leqslant f(\mathbf{x}) \leqslant M$ for some integer M. For all practical purposes, the image is a continuous function which is represented by measurements or *samples* at regularly spaced intervals. At the time the image is sampled, the intensity is usually *quantized* into a number of different *gray levels.* For a discrete image, $f(\mathbf{x})$ is an integer gray level, and $\mathbf{x} = (\mathrm{x}, \mathrm{y})$ is a pair of *integer* coordinates representing a sample point in a two-dimensional image plane. Sampling involves two important choices: (1) the *sampling interval,* which determines in a basic way whether all the information in the image is represented, and (2) the *tesselation* or spatial pattern of sample points, which affects important notions of connectivity and distance. In our presentation, we first show qualitatively the effects of sampling and gray-level quantization. Second, we discuss the simplest kinds of tesselations of the plane. Finally, and most important, we describe the sampling theorem, which specifies how close the image samples must be to represent the image unambiguously.

The choice of integers to represent the gray levels and coordinates is dictated by limitations in sensing. Also, of course, there are hardware limitations in representing images arising from their sheer size. Table 2.3 shows the storage required for an image in 8-bit bytes as a function of m, the number of bits per sample, and N, the linear dimension of a square image.

For reasons of economy (and others discussed in Chapter 3) we often use images of considerably less spatial resolution than that required to preserve fidelity to the human viewer. Figure 2.9 provides a qualitative idea of image degradation with decreasing spatial resolution.

As shown in Table 2.3, another way to save space besides using less spatial resolution is to use fewer bits per gray level sample. Figure 2.10 shows an image represented with different numbers of bits per sample. One striking effect is the "contouring" introduced with small numbers of gray levels. This is, in general, a problem for computer vision algorithms, which cannot easily discount the false contours. The choice of spatial and gray-level resolution for any particular computer vision task is an important one which depends on many factors. It is typical in

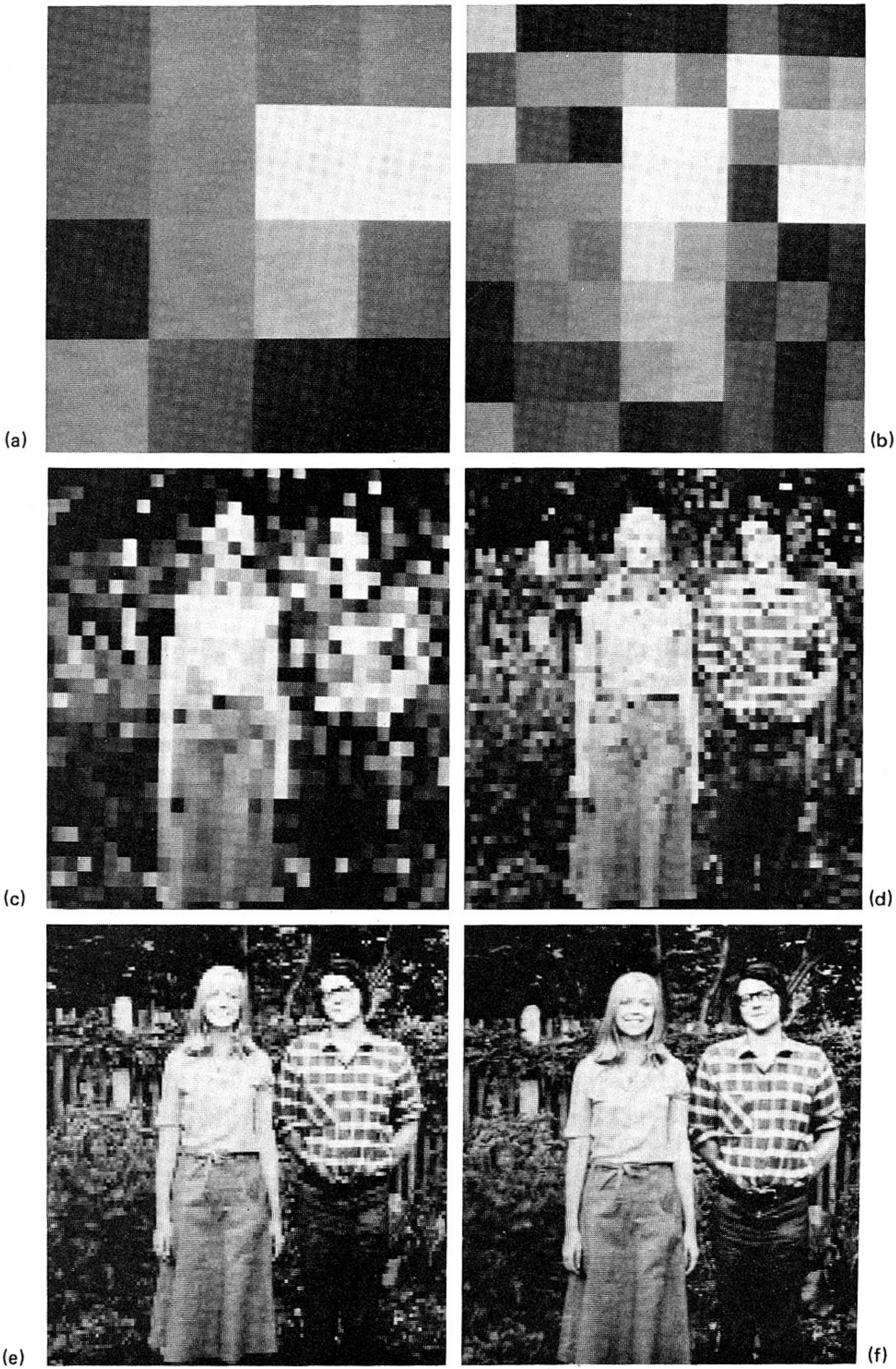


Fig. 2.9 Using different numbers of samples. (a) $N = 16$; (b) $N = 32$; (c) $N = 64$; (d) $N = 128$; (e) $N = 256$; (f) $N = 512$.

Table 2.3

NUMBER OF 8-BIT BYTES OF STORAGE FOR VARIOUS VALUES OF N AND M

N	32	64	128	256	512
m					
1	128	512	2,048	8,192	32,768
2	256	1,024	4,096	16,384	65,536
3	512	2,048	8,192	32,768	131,072
4	512	2,048	8,192	32,768	131,072
5	1,024	4,096	16,384	65,536	262,144
6	1,024	4,096	16,384	65,536	262,144
7	1,024	4,096	16,384	65,536	262,144
8	1,024	4,096	16,384	65,536	262,144

computer vision to have to balance the desire for increased resolution (both gray scale and spatial) against its cost. Better data can often make algorithms easier to write, but a small amount of data can make processing more efficient. Of course, the image domain, choice of algorithms, and image characteristics all heavily influence the choice of resolutions.

Tesselations and Distance Metrics

Although the spatial samples for $f(\mathbf{x})$ can be represented as points, it is more satisfying to the intuition and a closer approximation to the acquisition process to think of these samples as finite-sized cells of constant gray-level partitioning the image. These cells are termed *pixels*, an acronym for *picture elements.* The pattern into which the plane is divided is called its *tesselation.* The most common regular tesselations of the plane are shown in Fig. 2.11.

Although rectangular tesselations are almost universally used in computer vision, they have a structural problem known as the "connectivity paradox." Given a pixel in a rectangular tesselation, how should we define the pixels to which it is connected? Two common ways are *four-connectivity* and *eight-connectivity*, shown in Fig. 2.12.

However, each of these schemes has complications. Consider Fig. 2.12c, consisting of a black object with a hole on a white background. If we use four-connectedness, the figure consists of four disconnected pieces, yet the hole is separated from the "outside" background. Alternatively, if we use eight-connectedness, the figure is one connected piece, yet the hole is now connected to the outside. This paradox poses complications for many geometric algorithms. Triangular and hexagonal tesselations do not suffer from connectivity difficulties (if we use three-connectedness for triangles); however, *distance* can be more difficult to compute on these arrays than for rectangular arrays.

The distance between two pixels in an image is an important measure that is fundamental to many algorithms. In general, a distance d is a *metric.* That is,

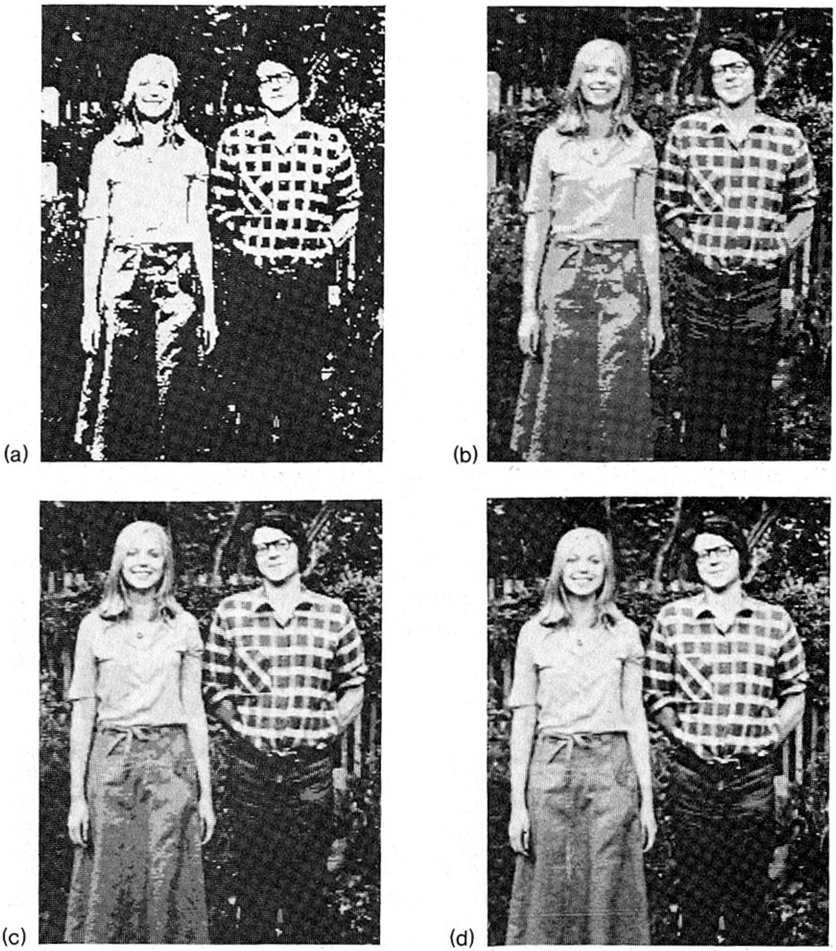


Fig. 2.10 Using different numbers of bits per sample. (a) $m = 1$; (b) $m = 2$; (c) $m = 4$; (d) $m = 8$.

(1) $d(\mathbf{x}, \mathbf{y}) = 0$ iff $\mathbf{x} = \mathbf{y}$

(2) $d(\mathbf{x}, \mathbf{y}) = d(\mathbf{y}, \mathbf{x})$

(3) $d(\mathbf{x}, \mathbf{y}) + d(\mathbf{y}, \mathbf{z}) \geqslant d(\mathbf{x}, \mathbf{z})$

For square arrays with unit spacing between pixels, we can use any of the following common distance metrics (Fig. 2.13) for two pixels $\boldsymbol{x} = (x_1, y_1)$ and $\boldsymbol{y} = (x_2, y_2)$.

Euclidean:

$$d_e(\mathbf{x}, \mathbf{y}) = \sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2} \tag{2.37}$$

City block:

$$d_{cb}(\mathbf{x}, \mathbf{y}) = |x_1 - x_2| + |y_1 - y_2| \tag{2.38}$$

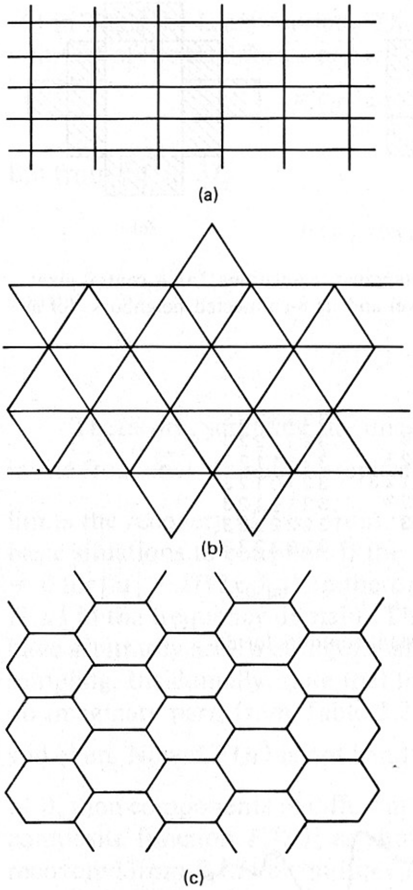


Fig. 2.11 Different tesselations of the image plane. (a) Rectangular; (b) triangular; (c) hexagonal.

Chessboard:

$$d_{ch}(\mathbf{x}, \mathbf{y}) = \max\left\{|x_1 - x_2|, |y_1 - y_2|\right\} \tag{2.39}$$

Other definitions are possible, and all such measures extend to multiple dimensions. The tesselation of higher-dimensional space into pixels usually is confined to (n-dimensional) cubical pixels.

The Sampling Theorem

Consider the one-dimensional "image" shown in Fig. 2.14. To digitize this image one must sample the image function. These samples will usually be separated at regular intervals as shown. How far apart should these samples be to allow reconstruction (to a given accuracy) of the underlying continuous image from its samples? This question is answered by the Shannon sampling theorem. An excellent rigorous presentation of the sampling theorem may be found in [Rosenfeld and Kak 1976]. Here we shall present a shorter graphical interpretation using the results of Table 2.2. For simplicity we consider the image to be periodic in order to avoid small edge effects introduced by the finite image domain. A more rigorous

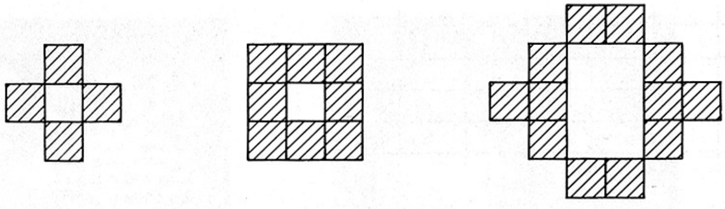


Fig. 2.12 Connectivity paradox for rectangular tesselations. (a) A central pixel and its 4-connected neighbors; (b) a pixel and its 8-connected neighbors; (c) a figure with ambiguous connectivity.

```
    2 3 2            3          3 3 3 3 3 3 3
  3 2 2 2 3        3 2 3        3 2 2 2 2 2 3
2 2 1 1 1 2 2    3 2 1 2 3      3 2 1 1 1 2 3
3 2 1 0 1 2 3  3 2 1 0 1 2 3    3 2 1 0 1 2 3
2 2 1 1 1 2 2    3 2 1 2 3      3 2 1 1 1 2 3
  3 2 2 2 3        3 2 3        3 2 2 2 2 2 3
    2 3 2            3          3 3 3 3 3 3 3

    (a)             (b)              (c)
```

Fig. 2.13 Equidistant contours for different metrics.

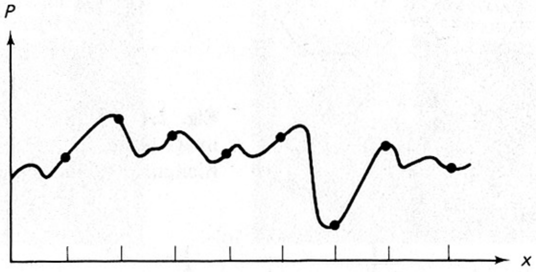


Fig. 2.14 One-dimensional image and its samples.

treatment, which considers these effects, is given in [Andrews and Hunt 1977].

Suppose that the image is sampled with a "comb" function of spacing x_0 (see Table 2.2). Then the sampled image can be modeled by

$$f_s(x) = f(x) \sum_n \delta(x - nx_0) \tag{2.40}$$

where the image function modulates the comb function. Equivalently, this can be written as

$$f_s(x) = \sum_n f(nx_0)\,\delta(x - nx_0) \tag{2.41}$$

The right-hand side of Eq. (2.40) is the product of two functions, so that property

(6) in Table 2.1 is appropriate. The Fourier transform of $f_s(x)$ is equal to the convolution of the transforms of each of the two functions. Using this result yields

$$F_s(u) = F(u) * \frac{1}{x_0}\sum_n \delta\left(u - \frac{n}{x_0}\right) \tag{2.42}$$

But from Eq. (2.3),

$$F(u) * \delta\left(u - \frac{n}{x_0}\right) = F\left(u - \frac{n}{x_0}\right) \tag{2.43}$$

so that

$$F_s(u) = \frac{1}{x_0}\sum_n F\left(u - \frac{n}{x_0}\right) \tag{2.44}$$

Therefore, sampling the image function $f(x)$ at intervals of x_0 is equivalent in the frequency domain to replicating the transform of f at intervals of $\frac{1}{x_0}$. This limits the recovery of $f(x)$ from its sampled representation, $f_s(x)$. There are two basic situations to consider. If the transform of $f(x)$ is *bandlimited* such that $F(u) = 0$ for $|u| > 1/(2x_0)$, then there is no overlap between successive replications of $F(u)$ in the frequency domain. This is shown for the case of Fig. 2.15a, where we have arbitrarily used a triangular-shaped image transform to illustrate the effects of sampling. Incidentally, note that for this transform $F(u) = F(-u)$ and that it has no imaginary part; from Table 2.2, the one-dimensional image must also be real and even. Now if $F(u)$ is not bandlimited, i.e., there are $u > \frac{1}{2x_0}$ for which $F(u) \neq 0$, then components of different replications of $F(u)$ will interact to produce the composite function $F_s(u)$, as shown in Fig. 2.15b. In the first case $f(x)$ can be recovered from $F_s(u)$ by multiplying $F_s(u)$ by a suitable $G(u)$:

$$G(u) = \begin{cases} 1 & |u| < \dfrac{1}{2x_0} \\ 0 & \text{otherwise} \end{cases} \tag{2.45}$$

Then

$$f(x) = \mathcal{F}^{-1}[F_s(u)G(u)] \tag{2.46}$$

However, in the second case, $F_s(u)G(u)$ is very different from the original $F(u)$. This is shown in Fig. 2.15c. Sampling a $F(u)$ that is not bandlimited allows information at high spatial frequencies to interfere with that at low frequencies, a phenomenon known as *aliasing.*

Thus the sampling theorem has this very important result: As long as the image contains no spatial frequencies greater than one-half the sampling frequency, the underlying continuous image is unambiguously represented by its samples. However, lest one be tempted to insist on images that have been so sampled, note that it may be useful to sample at lower frequencies than would be required for total reconstruction. Such sampling is usually preceded by some form of blurring of

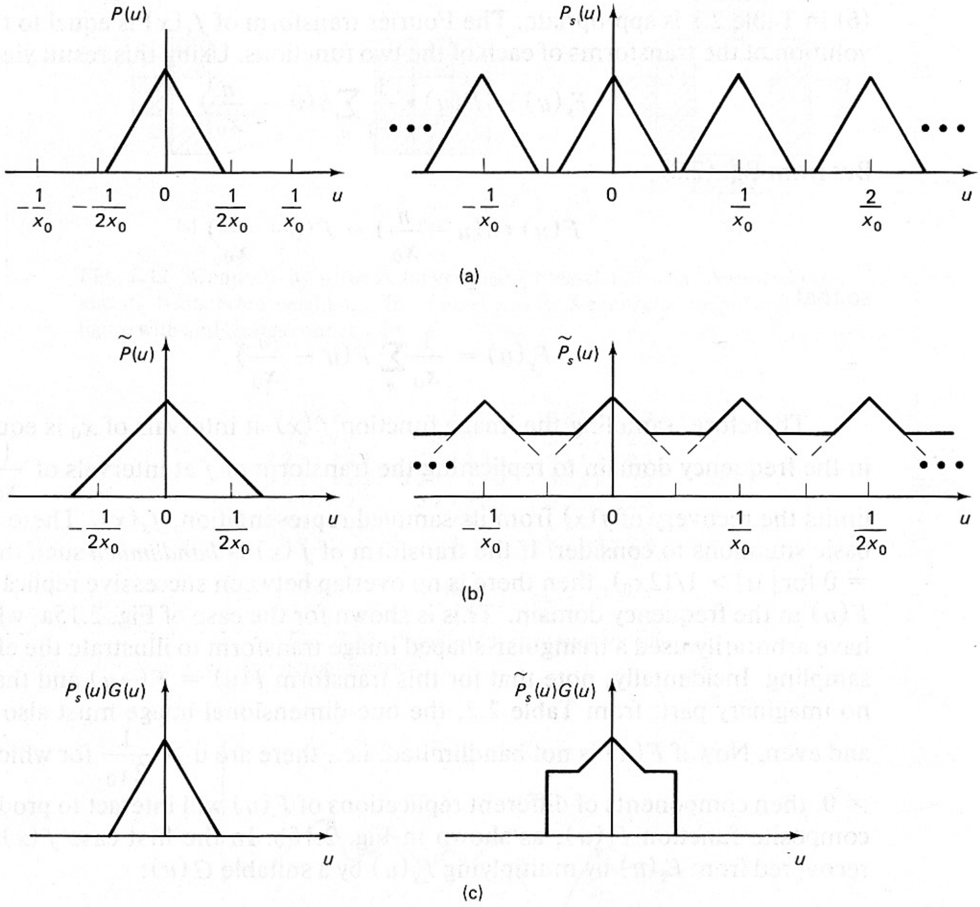


Fig. 2.15 (a) $F(u)$ bandlimited so that $F(u) = 0$ for $|u| > 1/2x_0$. (b) $F(u)$ not bandlimited as in (a). (c) reconstructed transform.

the image, or can be incorporated with such blurring (by integrating the image intensity over a finite area for each sample). Image blurring can bury irrelevant details, reduce certain forms of noise, and also reduce the effects of aliasing.

2.3 IMAGING DEVICES FOR COMPUTER VISION

There is a vast array of methods for obtaining a digital image in a computer. In this section we have in mind only "traditional" images produced by various forms of radiation impinging on a sensor after having been affected by physical objects.

Many sensors are best modeled as an *analog* device whose response must be *digitized* for computer representation. The types of imaging devices possible are limited only by the technical ingenuity of their developers; attempting a definitive