# Computer Vision
## A Modern Approach

David A. Forsyth
*University of California at Berkeley*

Jean Ponce
*University of Illinois at Urbana-Champaign*

*An Alan R. Apt Book*


Prentice Hall
Upper Saddle River, New Jersey 07458

# 1

# Cameras

There are many types of imaging devices, from animal eyes to video cameras and radio telescopes. They may or may not be equipped with lenses. For example, the first models of the *camera obscura* (literally, dark chamber) invented in the 16th century did not have lenses, but instead used a *pinhole* to focus light rays onto a wall or translucent plate and demonstrate the laws of perspective discovered a century earlier by Brunelleschi. Pinholes were replaced by more and more sophisticated lenses as early as 1550, and the modern photographic or digital camera is essentially a camera obscura capable of recording the amount of light striking every small area of its backplane (Figure 1.1).

**Figure 1.1** Image formation on the backplate of a photographic camera. *Figure from US NAVY MANUAL OF BASIC OPTICS AND OPTICAL INSTRUMENTS, prepared by the Bureau of Naval Personnel, reprinted by Dover Publications, Inc., (1969).*

The imaging surface of a camera is generally a rectangle, but the shape of the human retina is much closer to a spherical surface, and panoramic cameras may be equipped with cylindrical retinas. Imaging sensors have other characteristics. They may record a spatially discrete picture (like our eyes with their rods and cones, 35 mm cameras with their grain, and digital cameras with their rectangular picture elements or pixels) or a continuous one (in the case of old-fashioned TV tubes, for example). The signal that an imaging sensor records at a point on its retina may be discrete or continuous, and it may consist of a single number (black-and-white camera), a few values (e.g., the R G B intensities for a color camera or the responses of the three types of cones for the human eye), many numbers (e.g., the responses of hyperspectral sensors), or even a continuous function of wavelength (which is essentially the case for spectrometers). Examining these characteristics is the subject of this chapter.

## 1.1 PINHOLE CAMERAS

### 1.1.1 Perspective Projection

Imagine taking a box, pricking a small hole in one of its sides with a pin, and then replacing the opposite side with a translucent plate. If you hold that box in front of you in a dimly lit room, with the pinhole facing some light source (say a candle), you see an inverted image of the candle appearing on the translucent plate (Figure 1.2). This image is formed by light rays issued from the scene facing the box. If the pinhole were really reduced to a point (which is of course physically impossible), exactly one light ray would pass through each point in the plane of the plate (or *image plane*), the pinhole, and some scene point.

In reality, the pinhole has a finite (albeit small) size, and each point in the image plane collects light from a cone of rays subtending a finite solid angle, so this idealized and extremely simple model of the imaging geometry does not strictly apply. In addition, real cameras are normally equipped with lenses, which further complicates things. Still, the *pinhole perspective* (also called *central perspective*) projection model, first proposed by Brunelleschi at the beginning of the 15th century, is mathematically convenient. Despite its simplicity, it often provides an acceptable approximation of the imaging process. Perspective projection creates inverted images, and it is sometimes convenient to consider instead a *virtual image* associated with a plane lying *in front* of the pinhole at the same distance from it as the actual image plane (Figure 1.2). This virtual image is not inverted, but is otherwise strictly equivalent to the actual one. Depending on the context, it may be more convenient to think about one or the other. Figure 1.3(a) illustrates an obvious effect of perspective projection: The apparent size of objects depends on their distance. For example, the images $B'$ and $C'$ of the posts $B$ and $C$ have the same height, but $A$ and $C$

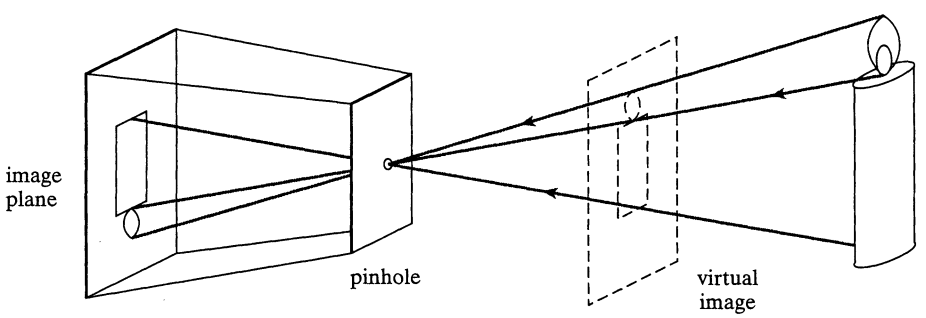


**Figure 1.2** The pinhole imaging model.

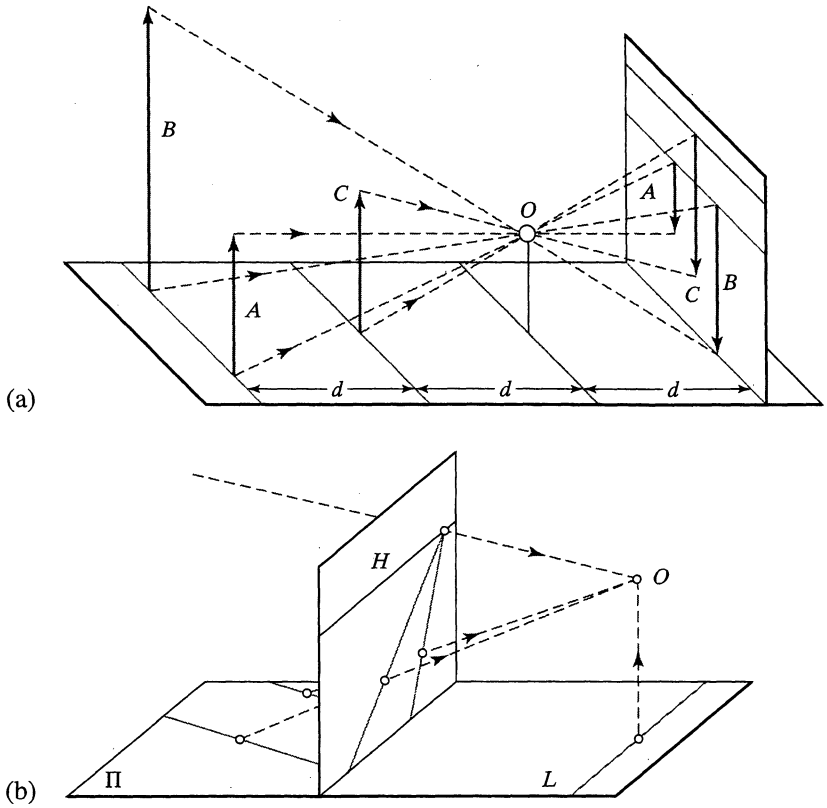


**Figure 1.3** Perspective effects: (a) far objects appear smaller than close ones: the distance $d$ from the pinhole $O$ to the plane containing $C$ is half the distance from $O$ to the plane containing $A$ and $B$; (b) the images of parallel lines intersect at the horizon (after Hilbert and Cohn-Vossen, 1952, Figure 127). Note that the image plane is *behind* the pinhole in (a) (physical retina), and *in front* of it in (b) (virtual image plane). Most of the diagrams in this chapter and the rest of this book feature the physical image plane, but a virtual one is also used when appropriate, as in (b).

are really half the size of $B$. Figure 1.3(b) illustrates another well-known effect: The projections of two parallel lines lying in some plane $\Pi$ appear to converge on a horizon line $H$ formed by the intersection of the image plane with the plane parallel to $\Pi$ and passing through the pinhole. Note that the line $L$ in $\Pi$ that is parallel to the image plane has no image at all.

These properties are easy to prove in a purely geometric fashion. However, it is often convenient (if not quite as elegant) to reason in terms of reference frames, coordinates, and equations. Consider, for example, a coordinate system $(O, \boldsymbol{i}, \boldsymbol{j}, \boldsymbol{k})$ attached to a pinhole camera, whose origin $O$ coincides with the pinhole, and vectors $\boldsymbol{i}$ and $\boldsymbol{j}$ form a basis for a vector plane parallel to the image plane $\Pi'$, which is located at a positive distance $f'$ from the pinhole along the vector $\boldsymbol{k}$ (Figure 1.4). The line perpendicular to $\Pi'$ and passing through the pinhole is called the *optical axis*, and the point $C'$ where it pierces $\Pi'$ is called the *image center*. This point can be used as the origin of an image plane coordinate frame, and it plays an important role in camera calibration procedures.

Let $P$ denote a scene point with coordinates $(x, y, z)$ and $P'$ denote its image with coordinates $(x', y', z')$. Since $P'$ lies in the image plane, we have $z' = f'$. Since the three points $P$, $O$,

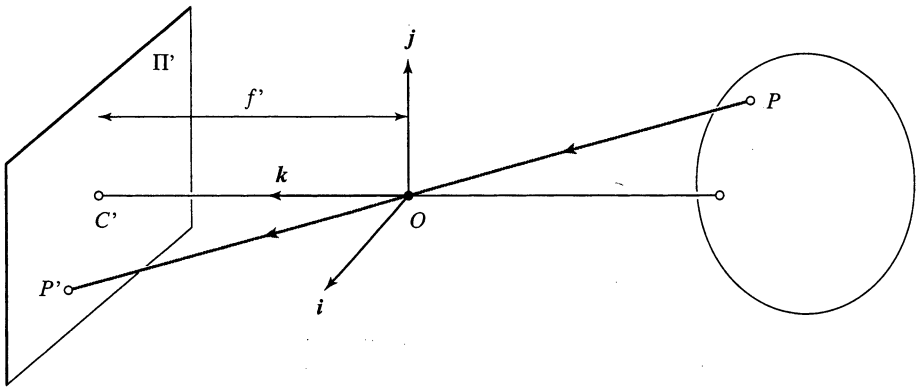


**Figure 1.4** The perspective projection equations are derived in this section from the collinearity of the point $P$, its image $P'$, and the pinhole $O$.

and $P'$ are collinear, we have $\overrightarrow{OP'} = \lambda \overrightarrow{OP}$ for some number $\lambda$, so

$$\begin{cases} x' = \lambda x \\ y' = \lambda y \\ f' = \lambda z \end{cases} \iff \lambda = \frac{x'}{x} = \frac{y'}{y} = \frac{f'}{z},$$

and therefore

$$\begin{cases} x' = f'\dfrac{x}{z}, \\ y' = f'\dfrac{y}{z}. \end{cases} \tag{1.1}$$

### 1.1.2 Affine Projection

As noted in the previous section, pinhole perspective is only an approximation of the geometry of the imaging process. This section discusses a class of coarser approximations, called *affine projection models*, that are also useful on occasion. We focus on two specific affine models—namely, *weak-perspective* and *orthographic* projections. A third one, the *paraperspective* model, is introduced in Chapter 12, where the name affine projection is also justified.

Consider the *fronto-parallel plane* $\Pi_0$ defined by $z = z_0$ (Figure 1.5). For any point $P$ in $\Pi_0$ we can rewrite the perspective projection Eq. (1.1) as

$$\begin{cases} x' = -mx \\ y' = -my \end{cases} \quad \text{where} \quad m = -\frac{f'}{z_0}. \tag{1.2}$$

Physical constraints impose that $z_0$ be negative (the plane must be in front of the pinhole), so the *magnification* $m$ associated with the plane $\Pi_0$ is positive. This name is justified by the following remark: Consider two points $P$ and $Q$ in $\Pi_0$ and their images $P'$ and $Q'$ (Figure 1.5); obviously the vectors $\overrightarrow{PQ}$ and $\overrightarrow{P'Q'}$ are parallel, and we have $|\overrightarrow{P'Q'}| = m|\overrightarrow{PQ}|$. This is the dependence of image size on object distance noted earlier.

When the scene depth is small relative to the average distance from the camera, the magnification can be taken to be constant. This projection model is called *weak perspective* or *scaled*

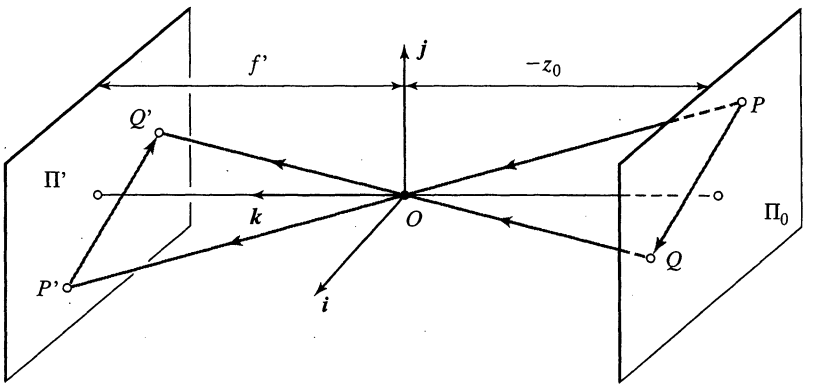


**Figure 1.5** Weak-perspective projection: All line segments in the plane $\Pi_0$ are projected with the same magnification.

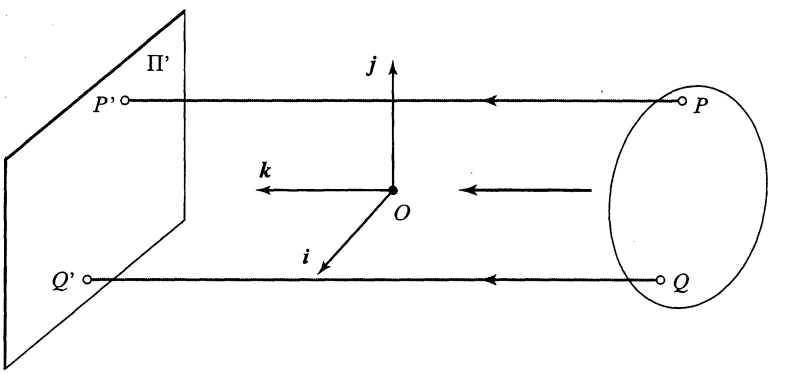


**Figure 1.6** Orthographic projection. Unlike other geometric models of the image-formation process, orthographic projection does not involve a reversal of image features. Accordingly, the magnification is taken to be negative, which is a bit unnatural, but simplifies the projection equations.

*orthography*. When it is a priori known that the camera always remains at a roughly constant distance from the scene, we can go further and normalize the image coordinates so that $m = -1$. This is *orthographic projection* defined by

$$\begin{cases} x' = x, \\ y' = y, \end{cases} \tag{1.3}$$

with all light rays parallel to the $\boldsymbol{k}$ axis and orthogonal to the image plane $\Pi'$ (Figure 1.6). Although weak-perspective projection is an acceptable model for many imaging conditions, assuming pure orthographic projection is usually unrealistic.

## 1.2 CAMERAS WITH LENSES

Most cameras are equipped with lenses. There are two main reasons for this: The first one is to gather light since a single ray of light would otherwise reach each point in the image plane under ideal pinhole projection. Real pinholes have a finite size of course, so each point in the image

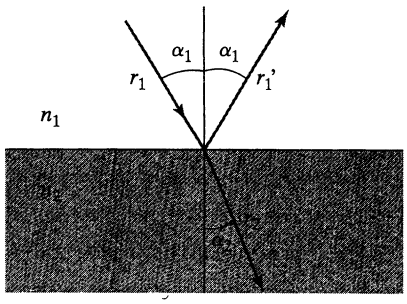


**Figure 1.7** Reflection and refraction at the interface between two homogeneous media with indexes of refraction $n_1$ and $n_2$.

plane is illuminated by a cone of light rays subtending a finite solid angle. The larger the hole, the wider the cone and the brighter the image, but a large pinhole gives blurry pictures. Shrinking the pinhole produces sharper images, but reduces the amount of light reaching the image plane, and may introduce *diffraction* effects. The second main reason for using a lens is to keep the picture in sharp focus while gathering light from a large area.

Ignoring diffraction, interferences, and other physical optics phenomena, the behavior of lenses is dictated by the laws of geometric optics (Figure 1.7): (1) light travels in straight lines (*light rays*) in homogeneous media; (2) when a ray is reflected from a surface, this ray, its reflection, and the surface normal are coplanar, and the angles between the normal and the two rays are complementary; and (3) when a ray passes from one medium to another, it is *refracted* (i.e., its direction changes). According to Snell's law, if $r_1$ is the ray incident to the interface between two transparent materials with indexes of refraction $n_1$ and $n_2$, and $r_2$ is the refracted ray, then $r_1$, $r_2$ and the normal to the interface are coplanar, and the angles $\alpha_1$ and $\alpha_2$ between the normal and the two rays are related by

$$n_1 \sin \alpha_1 = n_2 \sin \alpha_2. \tag{1.4}$$

In this chapter, we only consider the effects of refraction and ignore those of reflection. In other words, we concentrate on lenses as opposed to *catadioptric optical systems* (e.g., telescopes) that may include both reflective (mirrors) and refractive elements. Tracing light rays as they travel through a lens is simpler when the angles between these rays and the refracting surfaces of the lens are assumed to be small. The next section discusses this case.

### 1.2.1 Paraxial Geometric Optics

In this section, we consider *paraxial* (or *first-order*) geometric optics, where the angles between all light rays going through a lens and the normal to the refractive surfaces of the lens are small. In addition, we assume that the lens is rotationally symmetric about a straight line, called its *optical axis*, and that all refractive surfaces are spherical. The symmetry of this setup allows us to determine the projection geometry by considering lenses with circular boundaries lying in a plane that contains the optical axis.

Let us consider an incident light ray passing through a point $P_1$ on the optical axis and refracted at the point $P$ of the circular interface of radius $R$ separating two transparent media with indexes of refraction $n_1$ and $n_2$ (Figure 1.8). Let us also denote by $P_2$ the point where the refracted ray intersects the optical axis a second time (the roles of $P_1$ and $P_2$ are completely symmetric) and by $C$ the center of the circular interface.

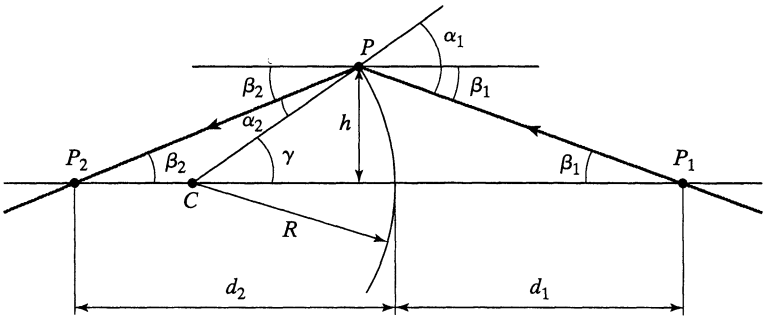


**Figure 1.8** Paraxial refraction: A light ray passing through the point $P_1$ is refracted at the point $P$ where it intersects a circular interface. The refracted ray intersects the optical axis in $P_2$. The center of the interface is at the point $C$ of the optical axis, and its radius is $R$. The angles $\alpha_1$, $\beta_1$, $\alpha_2$, and $\beta_2$ are all assumed to be small.

Let $\alpha_1$ and $\alpha_2$, respectively, denote the angles between the two rays and the chord joining $C$ to $P$. If $\beta_1$ (resp. $\beta_2$) is the angle between the optical axis and the line joining $P_1$ (resp. $P_2$) to $P$, the angle between the optical axis and the line joining $C$ to $P$ is, as shown by Figure 1.8, $\gamma = \alpha_1 - \beta_1 = \alpha_2 + \beta_2$. Now let $h$ denote the distance between $P$ and the optical axis and $R$ the radius of the circular interface. If we assume all angles are small and thus, to first order, equal to their sines and tangents, we have

$$\alpha_1 = \gamma + \beta_1 \approx h\left(\frac{1}{R} + \frac{1}{d_1}\right) \quad \text{and} \quad \alpha_2 = \gamma - \beta_2 \approx h\left(\frac{1}{R} - \frac{1}{d_2}\right).$$

Writing Snell's law for small angles yields the *paraxial refraction equation*:

$$n_1\alpha_1 \approx n_2\alpha_2 \Longleftrightarrow \frac{n_1}{d_1} + \frac{n_2}{d_2} = \frac{n_2 - n_1}{R}. \tag{1.5}$$

Note that the relationship between $d_1$ and $d_2$ depends on $R$, $n_1$, and $n_2$, but not on $\beta_1$ or $\beta_2$. This is the main simplification introduced by the paraxial assumption. It is easy to see that Eq. (1.5) remains valid when some (or all) of the values of $d_1$, $d_2$, and $R$ become negative, corresponding to the points $P_1$, $P_2$, or $C$ switching sides.

Of course, real lenses are bounded by at least two refractive surfaces. The corresponding ray paths can be constructed iteratively using the paraxial refraction equation. The next section illustrates this idea in the case of thin lenses.

### 1.2.2 Thin Lenses

Let us now consider a lens with two spherical surfaces of radius $R$ and index of refraction $n$. We assume that this lens is surrounded by vacuum (or, to an excellent approximation, by air), with an index of refraction equal to 1, and that it is *thin* (i.e., that a ray entering the lens and refracted at its right boundary is immediately refracted again at the left boundary).

Consider a point $P$ located at (negative) depth $z$ off the optical axis and denote by $(PO)$ the ray passing through this point and the center $O$ of the lens (Figure 1.9). As shown in the exercises, it follows from Snell's law and Eq. (1.5) that the ray $(PO)$ is not refracted and that all other rays passing through $P$ are focused by the thin lens on the point $P'$ with depth $z'$ along

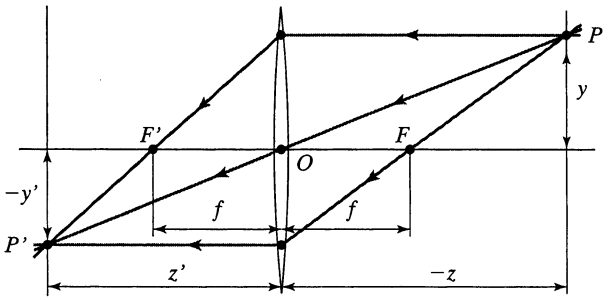


**Figure 1.9** A thin lens. Rays passing through the point $O$ are not refracted. Rays parallel to the optical axis are focused on the focal point $F'$.

$(PO)$ such that

$$\frac{1}{z'} - \frac{1}{z} = \frac{1}{f}, \tag{1.6}$$

where $f = \frac{R}{2(n-1)}$ is the *focal length* of the lens.

Note that the equations relating the positions of $P$ and $P'$ are exactly the same as under pinhole perspective projection if we take $z' = f'$, since $P$ and $P'$ lie on a ray passing through the center of the lens, but that points located at a distance $-z$ from $O$ are only in sharp focus when the image plane is located at a distance $z'$ from $O$ on the other side of the lens that satisfies Eq. (1.6) (i.e., the *thin lens equation*). Letting $z \to -\infty$ shows that $f$ is the distance between the center of the lens and the plane where objects such as stars, which are effectively located at $z = -\infty$, focus. The two points $F$ and $F'$ located at distance $f$ from the lens center on the optical axis are called the *focal points* of the lens.

In practice, objects within some range of distances (called *depth of field* or *depth of focus*) are in acceptable focus. As shown in the exercises, the depth of field increases with the *f number* of the lens (i.e., the ratio between the focal length of the lens and its diameter). The *field of view* of a camera is the portion of scene space that actually projects onto the retina of the camera. It is not defined by the focal length alone, but also depends on the effective area of the retina (e.g., the area of film that can be exposed in a photographic camera, or the area of the CCD sensor in a digital camera; Figure 1.10).

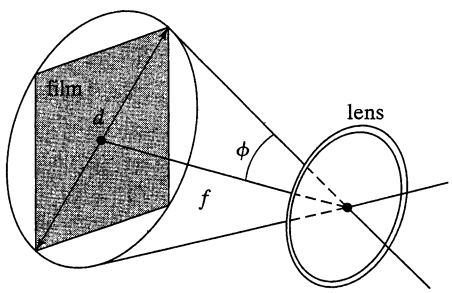


**Figure 1.10** The field of view of a camera is $2\phi$, where $\phi \stackrel{\text{def}}{=} \arctan \frac{d}{2f}$, $d$ is the diameter of the sensor (film or CCD chip) and $f$ is the focal length of the camera. When $f$ is (much) shorter than $d$, we have a wide-angle lens with rays that can be off the optical axis by more than 45°. Telephoto lenses have a small field of view and produce pictures closer to affine ones.

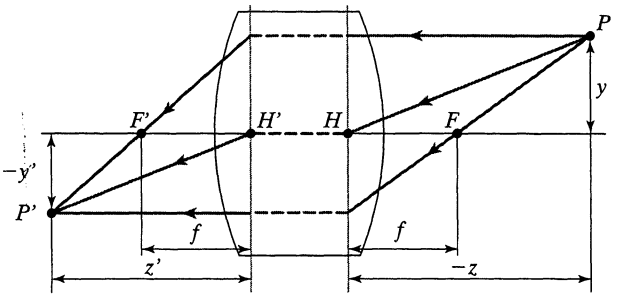


**Figure 1.11** A simple thick lens with two spherical surfaces.

### 1.2.3 Real Lenses

A more realistic model of simple optical systems is the *thick lens*. The equations describing its behavior are easily derived from the paraxial refraction equation, and they are the same as the pinhole perspective and thin lens projection equations except for an offset (Figure 1.11): If $H$ and $H'$ denote the *principal points* of the lens, then Eq. (1.6) holds when $-z$ (resp. $z'$) is the distance between $P$ (resp. $P'$) and the plane perpendicular to the optical axis and passing through $H$ (resp. $H'$). In this case, the only undeflected ray is along the optical axis.

Simple lenses suffer from a number of *aberrations*. To understand why, let us remember first that the paraxial refraction Eq. (1.5) is only an approximation—valid when the angle $\alpha$ between each ray along the optical path and the optical axis of the length is small and $\sin\alpha \approx \alpha$. For larger angles, a third-order Taylor expansion of the sine function yields the following refinement of the paraxial equation:

$$\frac{n_1}{d_1} + \frac{n_2}{d_2} = \frac{n_2 - n_1}{R} + h^2 \left[ \frac{n_1}{2d_1} \left( \frac{1}{R} + \frac{1}{d_1} \right)^2 + \frac{n_2}{2d_2} \left( \frac{1}{R} - \frac{1}{d_2} \right)^2 \right].$$

Here, $h$ denotes, as in Figure 1.8, the distance between the optical axis and the point where the incident ray intersects the interface. In particular, rays striking the interface farther from the optical axis are focused closer to the interface.

The same phenomenon occurs for a lens and it is the source of two types of *spherical aberrations* (Figure 1.12[a]): Consider a point $P$ on the optical axis and its paraxial image $P'$. The distance between $P'$ and the intersection of the optical axis with a ray issued from $P$ and refracted by the lens is called the *longitudinal spherical aberration* of that ray. Note that if an image plane $\Pi'$ were erected in $P$, the ray would intersect this plane at some distance from the axis, called the *transverse spherical aberration* of that ray. Together, all rays passing through $P$ and refracted by the lens form a circle of confusion centered in $P$ as they intersect $\Pi'$. The size of that circle changes when we move $\Pi'$ along the optical axis. The circle with minimum diameter is called the *circle of least confusion*, and it is not (in general) located in $P'$.

Besides spherical aberration, there are four other types of *primary aberrations* caused by the differences between first- and third-order optics—namely, *coma*, *astigmatism*, *field curvature*, and *distortion*. A precise definition of these aberrations is beyond the scope of this book. Suffice to say that, like spherical aberration, they degrade the image by blurring the picture of every object point. Distortion plays a different role and changes the shape of the image as a whole (Figure 1.12[b]). This effect is due to the fact that different areas of a lens have slightly different focal lengths. The aberrations mentioned so far are monochromatic (i.e., they are independent of the response of the lens to various wavelengths). However, the index of refraction of a transparent

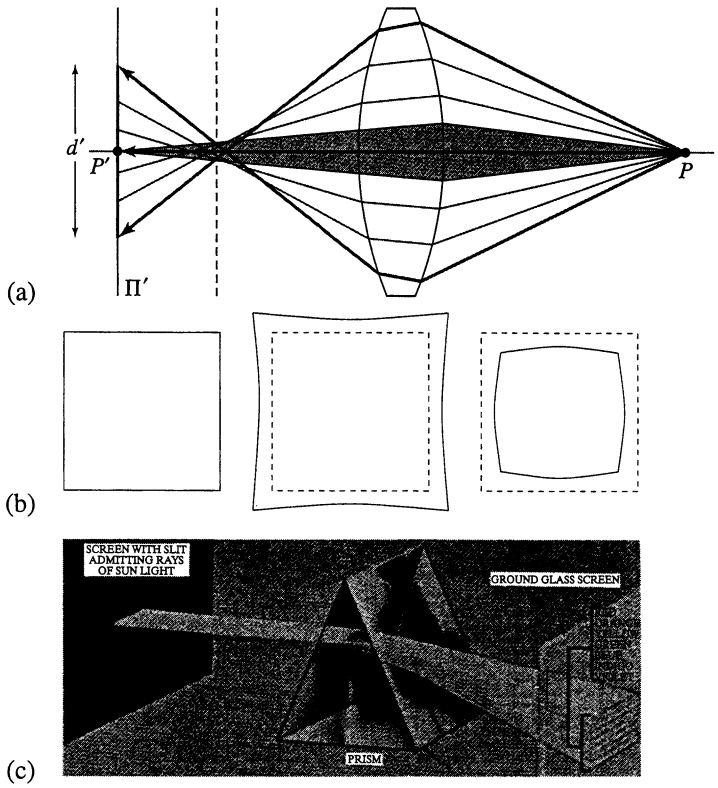


**Figure 1.12** Aberrations. (a) Spherical aberration: The grey region is the paraxial zone where the rays issued from $P$ intersect at its paraxial image $P'$. If an image plane $\Pi'$ is erected in $P'$, the image of $P'$ in that plane forms a circle of confusion of diameter $d'$. The focus plane yielding the circle of least confusion is indicated by a dashed line. (b) Distortion: From left to right, the nominal image of a fronto-parallel square, pincushion distortion, and barrel distortion. (c) Chromatic aberration: The index of refraction of a transparent medium depends on the wavelength (or color) of the incident light rays. Here, a prism decomposes white light into a palette of colors. *Figure from US NAVY MANUAL OF BASIC OPTICS AND OPTICAL INSTRUMENTS, prepared by the Bureau of Naval Personnel, reprinted by Dover Publications, Inc., (1969).*

medium depends on wavelength (Figure 1.12[c]), and it follows from the thin lens Eq. (1.6) that the focal length depends on wavelength as well. This causes the phenomenon of *chromatic aberration*: Refracted rays corresponding to different wavelengths intersect the optical axis at different points (*longitudinal chromatic aberration*) and form different circles of confusion in the same image plane (*transverse chromatic aberration*).

Aberrations can be minimized by aligning several simple lenses with well-chosen shapes and refraction indexes, separated by appropriate stops. These *compound lenses* can still be modeled by the thick lens equations. They suffer from one more defect relevant to machine vision: Light beams emanating from object points located off-axis are partially blocked by the various apertures (including the individual lens components) positioned inside the lens to limit aberrations (Figure 1.13). This phenomenon, called *vignetting*, causes the brightness to drop in the image periphery. Vignetting may pose problems to automated image analysis programs, but it is not as important in photography thanks to the human eye's remarkable insensitivity to smooth

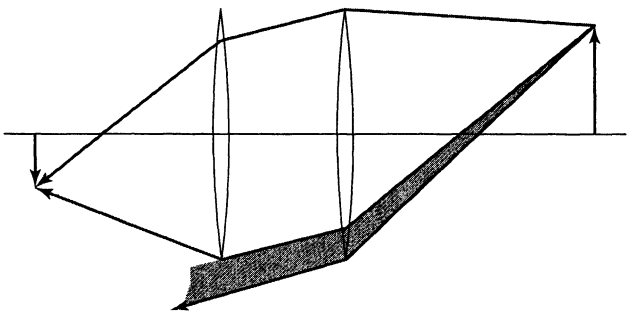

**Figure 1.13** Vignetting effect in a two-lens system. The shaded part of the beam never reaches the second lens. Additional apertures and stops in a lens further contribute to vignetting.

brightness gradients. Speaking of which, it is time to look at this extraordinary organ in a bit more detail.

## 1.3 THE HUMAN EYE

Here we give a (brief) overview of the anatomical structure of the eye. It is largely based on the presentation in Wandell (1995), and the interested reader is invited to read this excellent book for more details. Figure 1.14 (left) is a sketch of the section of an eyeball through its vertical plane of symmetry, showing the main elements of the eye: the *iris* and the *pupil*, which control the amount of light penetrating the eyeball; the *cornea* and the crystalline *lens*, which together refract the light to create the retinal image; and finally the *retina*, where the image is formed.

Despite its globular shape, the human eyeball is functionally similar to a camera with a field of view covering a 160° (width) × 135° (height) area. Like any other optical system, it suffers

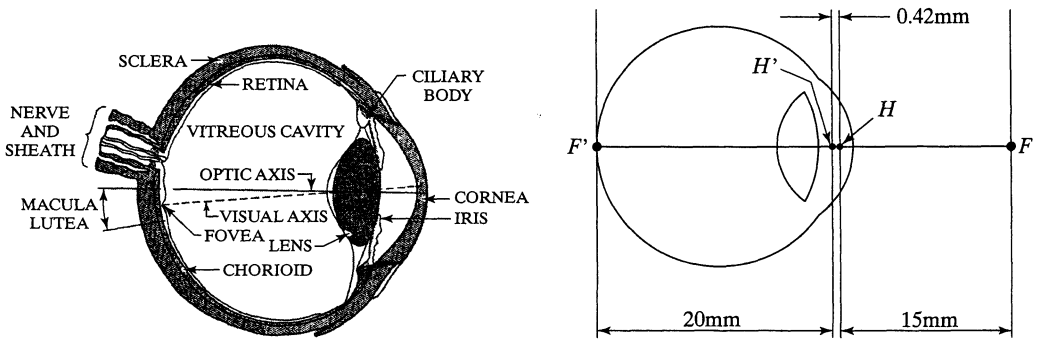


**Figure 1.14** Left: the main components of the human eye. *Reproduced with permission, the American Society for Photogrammetry and Remote Sensing. A.L. Nowicki, "Stereoscopy." MANUAL OF PHOTOGRAMMETRY, edited by M.M. Thompson, R.C. Eller, W.A. Radlinski, and J.L. Speert, third edition, pp. 515–536. Bethesda: American Society of Photogrammetry, (1966).* Right: Helmoltz's schematic eye as modified by Laurance (after Driscoll and Vaughan, 1978). The distance between the pole of the cornea and the anterior principal plane is 1.96 mm, and the radii of the cornea, anterior, and posterior surfaces of the lens are respectively 8 mm, 10 mm, and 6 mm.

from various types of geometric and chromatic aberrations. Several models of the eye obeying the laws of first-order geometric optics have been proposed, and Figure 1.14 (right) shows one of them, *Helmoltz's schematic eye*. There are only three refractive surfaces, with an infinitely thin cornea and a homogeneous lens. The constants given in Figure 1.14 are for the eye focusing at infinity (*unaccommodated eye*). This model is of course only an approximation of the real optical characteristics of the eye.

Let us have a second look at the components of the eye one layer at a time: the cornea is a transparent, highly curved, refractive window through which light enters the eye before being partially blocked by the colored and opaque surface of the iris. The pupil is an opening at the center of the iris whose diameter varies from about 1 to 8 mm in response to illumination changes, dilating in low light to increase the amount of energy that reaches the retina and contracting in normal lighting conditions to limit the amount of image blurring due to spherical aberration in the eye. The refracting power (reciprocal of the focal length) of the eye is, in large part, an effect of refraction at the the air–cornea interface, and it is fine tuned by deformations of the crystalline lens that accommodates to bring objects into sharp focus. In healthy adults, it varies between 60 (unaccommodated case) and 68 diopters (1 diopter $= 1\ \mathrm{m}^{-1}$), corresponding to a range of focal lengths between 15 and 17 mm. The retina itself is a thin, layered membrane populated by two types of photoreceptors—*rods* and *cones*—that respond to light in the 330 to 730 nm wavelength range (violet to red). As mentioned in Chapter 6, there are three types of cones with different spectral sensitivities, and these play a key role in the perception of color. There are about 100 million rods and 5 million cones in a human eye. Their spatial distribution varies across the retina: The *macula lutea* is a region in the center of the retina where the concentration of cones is particularly high and images are sharply focused whenever the eye fixes its attention on an object (Figure 1.14). The highest concentration of cones occurs in the *fovea*, a depression in the middle of the macula lutea where it peaks at $1.6 \times 10^5/\mathrm{mm}^2$, with the centers of two neighboring cones separated by only half a minute of visual angle (Figure 1.15). Conversely, there are no rods in the center of the fovea, but the rod density increases toward the periphery of the visual field. There is also a *blind spot* on the retina, where the ganglion cell axons exit the retina and form the optic nerve.

The rods are extremely sensitive photoreceptors; they are capable of responding to a single photon, but they yield relatively poor spatial detail despite their high number because many rods converge to the same neuron within the retina. In contrast, cones become active at higher light

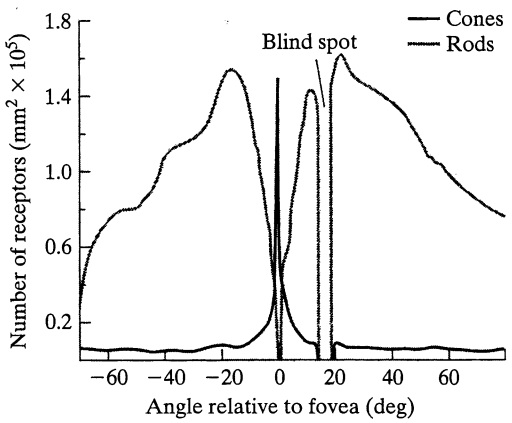


**Figure 1.15** The distribution of rods and cones across the retina. *Reprinted from FOUNDATIONS OF VISION, by B. Wandell, Sinauer Associates, Inc., (1995). © 1995 Sinauer Associates, Inc.*

levels, but the signal output by each cone in the fovea is encoded by several neurons, yielding a high resolution in that area. More generally, the area of the retina influencing a neuron's response is traditionally called its *receptive field*, although this term now also characterizes the actual electrical response of neurons to light patterns.

Of course, much more could (and should) be said about the human eye—for example how our two eyes verge and fixate on targets, cooperate in stereo vision, and so on. Besides, vision only starts with this camera of our mind, which leads to the fascinating (and still largely unsolved) problem of deciphering the role of the various portions of our brain in human vision. We come back to various aspects of this endeavor later in this book.

## 1.4 SENSING

What differentiates a camera (in the modern sense of the world) from the portable camera obscura of the 17th century is its ability to record the pictures that form on its backplane. Although it had been known since at least the Middle Ages that certain silver salts rapidly darken under the action of sunlight, it was only in 1816 that Niepce obtained the first true photographs by exposing paper treated with silver chloride to the light rays striking the image plane of a camera obscura, then fixing the picture with nitric acid. These first images were negatives, and Niepce soon switched to other photosensitive chemicals to obtain positive pictures. The earliest photographs have been lost, and the first one to have been preserved is *la table servie* (the set table) reproduced in Figure 1.16.

Niepce invented photography, but Daguerre would be the one to popularize it. After the two became associates in 1826, Daguerre went on to develop his own photographic process using mercury fumes to amplify and reveal the latent image formed on an iodized plating of silver on copper. *Daguerréotypes* were an instant success when Arago presented Daguerre's process at the French Academy of Sciences in 1839, three years after Niepce's death. Other milestones in the long history of photography include the introduction of the wet-plate negative/positive process by Legray and Archer in 1850, which required the pictures to be developed on the spot but produced excellent negatives; the invention of the gelatin process by Maddox in 1870, which eliminated the need for immediate development; the introduction in 1889 of the photographic film (that has replaced glass plates in most modern applications) by Eastman; and the invention by the Lumière brothers of cinema in 1895 and color photography in 1908.

**Figure 1.16** The first photograph on record, *la table servie*, obtained by Nicéphore Niepce in 1822. *Collection Harlinge–Viollet.*

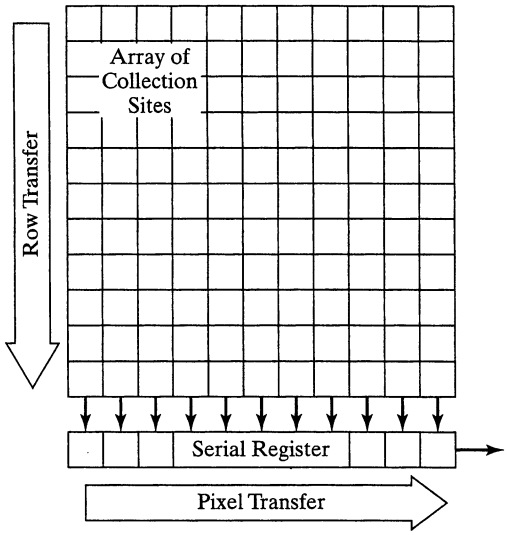


**Figure 1.17** A CCD Device.

The invention of television in the 1920s by people like Baird, Farnsworth, and Zworykin was of course a major impetus for the development of electronic sensors. The *vidicon* is a common type of TV vacuum tube. It is a glass envelope with an electron gun at one end and a faceplate at the other. The back of the faceplate is coated with a thin layer of photoconductor material laid over a transparent film of positively charged metal. This double coating forms the *target*. The tube is surrounded by focusing and deflecting coils that are used to repeatedly scan the target with the electron beam generated by the gun. This beam deposits a layer of electrons on the target to balance its positive charge. When a small area of the faceplate is struck by light, electrons flow through, locally depleting the charge of the target. As the electron beam scans this area, it replaces the lost electrons, creating a current proportional to the incident light intensity. The current variations are then transformed into a video signal by the vidicon circuitry.

### 1.4.1 CCD Cameras

Let us now turn to *charge-coupled-device (CCD)* cameras that were proposed in 1970 and have replaced vidicon cameras in most modern applications, from consumer camcorders to special-purpose cameras geared toward microscopy or astronomy applications. A CCD sensor uses a rectangular grid of electron-collection sites laid over a thin silicon wafer to record a measure of the amount of light energy reaching each of them (Figure 1.17). Each site is formed by growing a layer of silicon dioxide on the wafer and then depositing a conductive gate structure over the dioxide. When photons strike the silicon, electron-hole pairs are generated (*photo-conversion*), and the electron are captured by the *potential well* formed by applying a positive electrical potential to the corresponding gate. The electrons generated at each site are collected over a fixed period of time $T$.

At this point, the charges stored at the individual sites are moved using *charge coupling*: Charge packets are transfered from site to site by manipulating the gate potentials, preserving the separation of the packets. The image is read out of the CCD one row at a time, each row being transfered in parallel to a serial output register with one element in each column. Between two row reads, the register transfers its charges one at a time to an output amplifier that generates a signal proportional to the charge it receives. This process continues until the entire image has

been read out. It can be repeated 30 times per second (TV rate) for video applications or at a much slower pace, leaving ample time (seconds, minutes, even hours) for electron collection in low-light-level applications such as astronomy. It should be noted that the digital output of most CCD cameras is transformed internally into an analog video signal before being passed to a *frame grabber* that constructs the final digital image.

Consumer-grade color CCD cameras essentially use the same chips as black-and-white cameras, except that successive rows or columns of sensors are made sensitive to red, green or blue light often using a filter coating that blocks the complementary light. Other filter patterns are possible, including mosaics of $2 \times 2$ blocks formed by two green, one red, and one blue receptors (*Bayer patterns*). The spatial resolution of single-CCD cameras is of course limited, and higher-quality cameras use a beam splitter to ship the image to three different CCDs via color filters. The individual color channels are then either digitized separately (*RGB* output) or combined into a composite color video signal (*NTSC* output in the United States, *SECAM* or *PAL* in Europe and Japan) or into a *component video* format separating color and brightness information.

### 1.4.2 Sensor Models

For simplicity, we restrict our attention in this section to black-and-white CCD cameras: Color cameras can be treated in a similar fashion by considering each color channel separately and taking the effect of the associated filter response explicitly into account.

The number $I$ of electrons recorded at the cell located at row $r$ and column $c$ of a CCD array can be modeled as

$$I(r, c) = T \int_{\lambda} \int_{p \in S(r, c)} E(p, \lambda) R(p) q(\lambda) dp \, d\lambda,$$

where $T$ is the electron-collection time and the integral is calculated over the spatial domain $S(r, c)$ of the cell and the range of wavelengths to which the CCD has a nonzero response. In this integral, $E$ is is the power per unit area and unit wavelength (i.e., the *irradiance*, see chapter 4 for a formal definition) arriving at the point $p$, $R$ is the spatial response of the site, and $q$ is the *quantum efficiency* of the device (i.e., the number of electrons generated per unit of incident light energy). In general, both $E$ and $q$ depend on the light wavelength $\lambda$, and $E$ and $R$ depend on the point location $p$ within $S(r, c)$.

The output amplifier of the CCD transforms the charge collected at each site into a measurable voltage. In most cameras, this voltage is then transformed into a low-pass-filtered[1] video signal by the camera electronics with a magnitude proportional to $I$. The analog image can be once again transformed into a digital one using a frame grabber that spatially samples the video signal and quantizes the brightness value at each image point or *pixel* (from *picture element*).

There are several physical phenomena that alter the ideal camera model presented earlier: *Blooming* occurs when the light source illuminating a collection site is so bright that the charge stored at that site overflows into adjacent ones. It can be avoided by controlling the illumination, but other factors such as fabrication defects, thermal and quantum effects, and quantization noise are inherent to the imaging process. As shown next, these factors are appropriately captured by simple statistical models.

Quantum physics effects introduce an inherent uncertainty in the photoconversion process at each site (*shot noise*). More precisely, the number of electrons generated by this process can be modeled by a random integer variable $N_I(r, c)$ obeying a Poisson distribution with mean $\beta(r, c) I(r, c)$, where $\beta(r, c)$ is a number between 0 and 1 that reflects the variation of the spatial response and quantum efficiency across the image and also accounts for bad pixels. Electrons

[1] That is, roughly speaking, spatially or temporally averaged; more on this later.

freed from the silicon by thermal energy add to the charge of each collection site. Their contribution is called *dark current* and it can be modeled by a random integer variable $N_{DC}(r, c)$ whose mean $\mu_{DC}(r, c)$ increases with temperature. The effect of dark current can be controlled by cooling down the camera. Additional electrons are introduced by the CCD electronics (*bias*), and their number can also be modeled by a Poisson-distributed random variable $N_B(r, c)$ with mean $\mu_B(r, c)$. The output amplifier adds read-out noise that can be modeled by a real-valued random variable $R$ obeying a Gaussian distribution with mean $\mu_R$ and standard deviation $\sigma_R$.

There are other sources of uncertainty (e.g., charge transfer efficiency), but they can often be neglected. Finally, the discretization of the analog voltage by the frame grabber introduces both geometric effects (*line jitter*), which can be corrected via calibration, and a quantization noise, which can be modeled as a zero-mean random variable $Q(r, c)$ with a uniform distribution in the $[-\frac{1}{2}\delta, \frac{1}{2}\delta]$ interval and a variance of $\frac{1}{12}\delta^2$, where $\delta$ is the quantization step. This yields the following composite model for the digital signal $D(r, c)$:

$$D(r, c) = \gamma(N_I(r, c) + N_{DC}(r, c) + N_B(r, c) + R(r, c)) + Q(r, c).$$

In this equation, $\gamma$ is the combined gain of the amplifier and camera circuitry. The statistical properties of this model can be estimated via radiometric camera calibration: For example, dark current can be estimated by taking a number of sample pictures in a dark environment ($I = 0$).

## 1.5 NOTES

The classical textbook by Hecht (1987) is an excellent introduction to geometric optics. It includes a detailed discussion of paraxial optics as well as the various aberrations briefly mentioned in this chapter (see also Driscoll and Vaughan, 1978). Vignetting is discussed in Horn (1986) and Russ (1995). Wandell (1995) gives an excellent treatment of image formation in

**TABLE 1.1** Reference card: Camera models.

| | | |
|---|---|---|
| Perspective projection | $\begin{cases} x' = f'\frac{x}{z} \\ y' = f'\frac{y}{z} \end{cases}$ | $x, y$: world coordinates ($z < 0$)<br>$x', y'$: image coordinates<br>$f'$: pinhole-to-retina distance |
| Weak-perspective projection | $\begin{cases} x' = -mx \\ y' = -my \\ m = -\frac{f'}{z_0} \end{cases}$ | $x, y$: world coordinates<br>$x', y'$: image coordinates<br>$f'$: pinhole-to-retina distance<br>$z_0$: reference-point depth ($< 0$)<br>$m$: magnification ($> 0$) |
| Orthographic projection | $\begin{cases} x' = x \\ y' = y \end{cases}$ | $x, y$: world coordinates<br>$x', y'$: image coordinates |
| Snell's law | $n_1 \sin\alpha_1 = n_2 \sin\alpha_2$ | $n_1, n_2$: refraction indexes<br>$\alpha_1, \alpha_2$: normal-to-ray angles |
| Paraxial refraction | $\frac{n_1}{d_1} + \frac{n_2}{d_2} = \frac{n_2 - n_1}{R}$ | $n_1, n_2$: refraction indexes<br>$d_1, d_2$: point-to-interface distances<br>$R$: interface radius |
| Thin lens equation | $\frac{1}{z'} - \frac{1}{z} = \frac{1}{f}$ | $z$: object-point depth ($< 0$)<br>$z'$: image-point depth ($> 0$)<br>$f$: focal length |

the human visual system. The Helmoltz schematic model of the eye is detailed in Driscoll and Vaughan (1978).

CCD devices were introduced in Boyle and Smith (1970) and Amelio *et al.* (1970). Scientific applications of CCD cameras to microscopy and astronomy are discussed in Aiken *et al.* (1989), Janesick *et al.* (1987), Snyder *et al.* (1993), and Tyson (1990). The statistical sensor model presented in this chapter is based on Snyder *et al.* (1993), with an additional term for the quantization noise taken from Healey and Kondepudy (1994). These two articles contain interesting applications of sensor modeling to image restoration in astronomy and radiometric camera calibration in machine vision.

Given the fundamental importance of the notions introduced in this chapter, the main equations derived in its course have been collected in Table 1.1 for reference.

## PROBLEMS

**1.1.** Derive the perspective equation projections for a virtual image located at a distance $f'$ *in front* of the pinhole.

**1.2.** Prove geometrically that the projections of two parallel lines lying in some plane $\Pi$ appear to converge on a horizon line $H$ formed by the intersection of the image plane with the plane parallel to $\Pi$ and passing through the pinhole.

**1.3.** Prove the same result algebraically using the perspective projection Eq. (1.1). You can assume for simplicity that the plane $\Pi$ is orthogonal to the image plane.

**1.4.** Use Snell's law to show that rays passing through the optical center of a thin lens are not refracted, and derive the thin lens equation.

Hint: consider a ray $r_0$ passing through the point $P$ and construct the rays $r_1$ and $r_2$ obtained respectively by the refraction of $r_0$ by the right boundary of the lens and the refraction of $r_1$ by its left boundary.

**1.5.** Consider a camera equipped with a thin lens, with its image plane at position $z'$ and the plane of scene points in focus at position $z$. Now suppose that the image plane is moved to $\hat{z}'$. Show that the diameter of the corresponding blur circle is

$$d\frac{|z' - \hat{z}'|}{z'},$$

where $d$ is the lens diameter. Use this result to show that the depth of field (i.e., the distance between the near and far planes that will keep the diameter of the blur circles below some threshold $\varepsilon$) is given by

$$D = 2\varepsilon f z(z + f)\frac{d}{f^2 d^2 - \varepsilon^2 z^2},$$

and conclude that, for a *fixed* focal length, the depth of field increases as the lens diameter decreases, and thus the f number increases.

Hint: Solve for the depth $\hat{z}$ of a point whose image is focused on the image plane at position $\hat{z}'$, considering both the case where $\hat{z}'$ is larger than $z'$ and the case where it is smaller.

**1.6.** Give a geometric construction of the image $P'$ of a point $P$ given the two focal points $F$ and $F'$ of a thin lens.

**1.7.** Derive the thick lens equations in the case where both spherical boundaries of the lens have the same radius.