

Disclaimer: Please note that this is a pdf copy of a 3rd draft version of the original book, downloaded online. Figures and citations to bibliography are not yet all complete. This pdf file is distributed to the class to be used until students have a chance to purchase the book.
G. Gerig, August 24, 2008

Chapter 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 sixteenth century did not have lenses, but used instead a pinhole to focus light rays onto a wall or a translucent plate and demonstrate the laws of perspective discovered a century earlier by Brunelleschi. Of course pinholes have many limitations (we will come back to those in a minute) and they were replaced as early as 1550 by more and more sophisticated lenses. 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. Reprinted from [Navy, 1969], Figure 4-1.

The imaging surface of a camera is in general 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 as well. Indeed, as will be seen later in this chapter, spherical retinas are in some sense “better behaved” geometrically than planar ones: for example, a solid of revolution such as a vase will project onto a bi-laterally symmetric figure in the former case, but not in the latter one.

Imaging sensors have other characteristics as well: they may record a spatially discrete picture (as in our eyes, with their rods and cones, 35mm cameras, with their grain, and digital cameras, with their pixels), or a continuous one (in the case of old-fashioned television tubes for example). The signal at a point may itself 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 colour 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, using a pin to prick a small hole in the center of one of its sides, and then replacing the opposite side with a translucent plate. If you held that box in front of you in a dimly lit room, with the pinhole facing some light source, say a candle, you would observe 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 *image plane* of the plate, the pinhole, and some scene point.

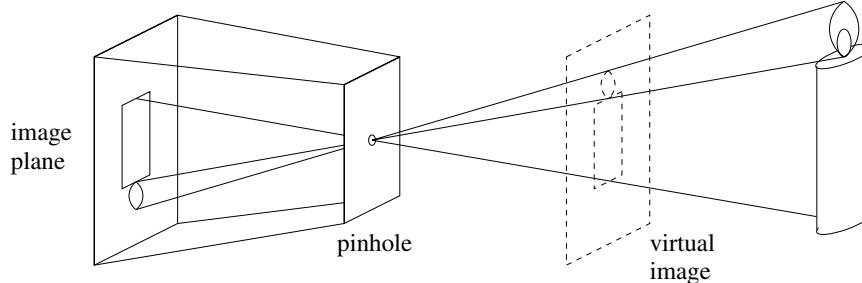


Figure 1.2. The pinhole imaging model.

In reality, the pinhole will have a finite (albeit small) size, and each point in the image plane will collect light from a cone of rays subtending a finite solid angle, so this idealized and extremely simple model of the imaging geometry will 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 fifteenth century, is mathematically convenient and, 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 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 are really half the size of B .

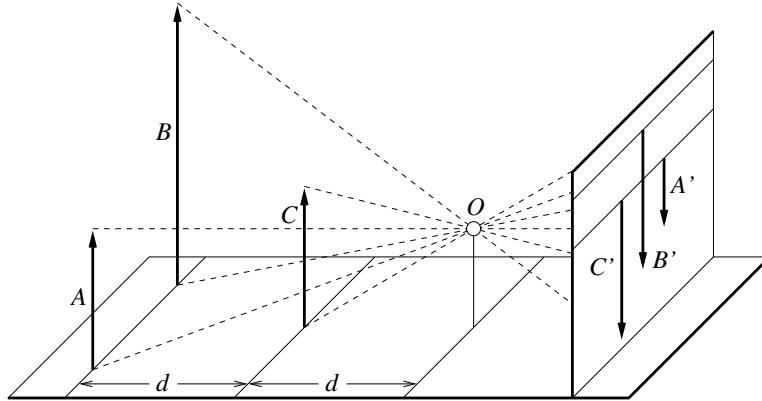


Figure 1.3. Far objects appear smaller than close ones. The distance d from the pinhole O to the plane containing C is half the distance from O to the plane containing A and B .

Figure 1.4 illustrates another well known effect: the projections of two parallel lines lying in some plane Π appear to converge on a horizon line H formed by the intersection of the image plane with the plane parallel to Π and passing through the pinhole. Note that the line L in Π that is parallel to the image plane has no image at all.

These properties are of course easy to prove in a purely geometric fashion. As usual 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 $(O, \mathbf{i}, \mathbf{j}, \mathbf{k})$ attached to a pinhole camera, whose origin O coincides with the pinhole, and vectors \mathbf{i} and \mathbf{j} form a basis for a vector plane parallel to the image plane Π' , itself located at a positive distance f' from the pinhole along the vector \mathbf{k} (Figure 1.5). The line perpendicular to Π' and passing through the pinhole is called the optical axis, and the point C' where it pierces Π' 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 and P' are colinear, we have $\overrightarrow{OP}' = \lambda \overrightarrow{OP}$ for some number

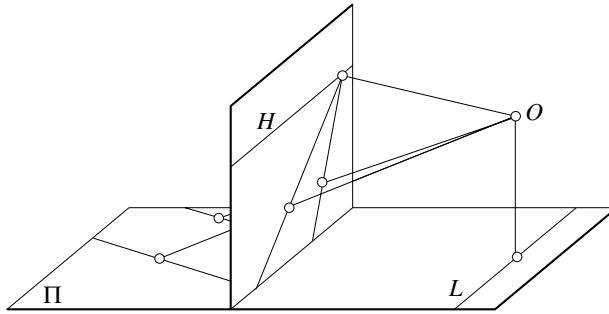


Figure 1.4. The images of parallel lines intersect at the horizon (after [Hilbert and Cohn-Vossen, 1952], Figure 127).

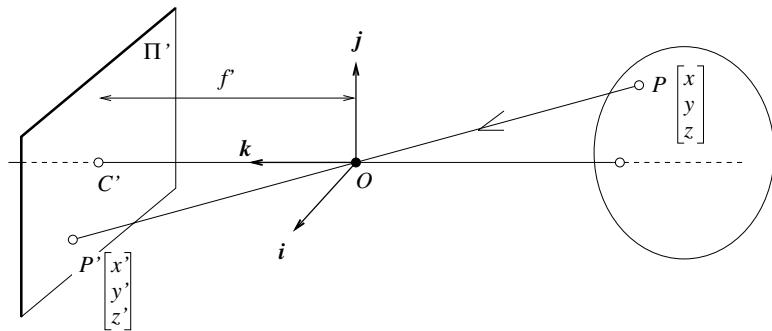


Figure 1.5. Setup for deriving the equations of perspective projection.

λ , 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

$$\begin{cases} x' = f' \frac{x}{z}, \\ y' = f' \frac{y}{z}. \end{cases} \quad (1.1.1)$$

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 even coarser approximations, called *affine projection models* (the name will be justified later), that are also useful on occasion.

Consider the *fronto-parallel* plane Π_0 defined by $z = z_0$ (Figure 1.6). For any point P in Π_0 we can rewrite the perspective projection equation (1.1.1) as

$$\begin{cases} x' = -mx \\ y' = -my \end{cases} \quad \text{where} \quad m = -\frac{f'}{z_0}. \quad (1.1.2)$$

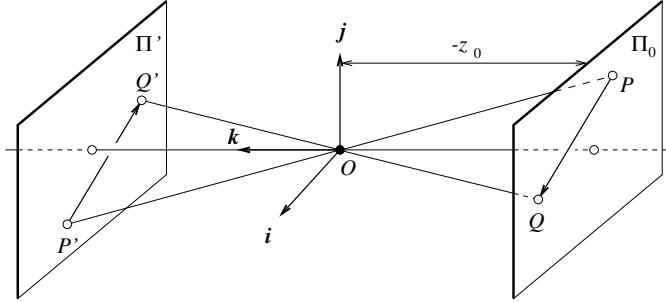


Figure 1.6. Weak perspective projection: all line segments in the plane Π_0 are projected with the same magnification.

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 Π_0 is positive. This name is justified by the following remark: consider two points P and Q in Π_0 and their images P' and Q' (Figure 1.6); 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 orthography*.

When it is a priori known that the camera will always remain 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

$$\begin{cases} x' = x, \\ y' = y, \end{cases} \quad (1.1.3)$$

with all light rays parallel to the k axis and orthogonal to the image plane Π' (Figure 1.7).

Although weak orthographic projection is an acceptable model for many imaging conditions, assuming pure orthographic projection is usually unrealistic.

1.1.3 Spherical Projection

The imaging surface, or *retina*, used in both perspective and affine projection models is a plane. One can of course imagine retinas with other simple shapes, such as

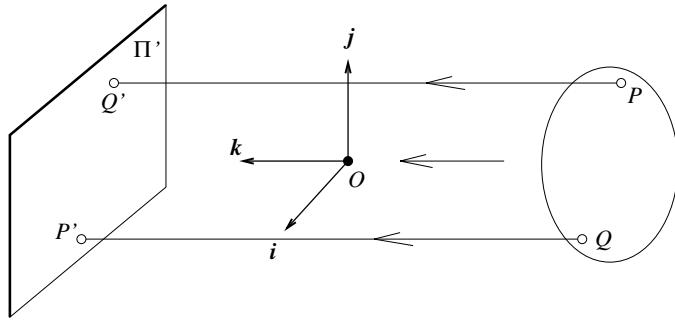


Figure 1.7. Orthographic projection.

cylinders or spheres for example. Here we consider spherical cameras where light rays passing through a pinhole form images on a spherical surface centered at the pinhole.

This model is particularly interesting because of its symmetry: consider for example a sphere observed by conventional perspective and orthographic cameras, as well as a spherical perspective camera (Figure 1.8). The outline of the sphere in the two perspective images is the intersection of the retina and a double cone tangent to the sphere with its apex located at the pinhole. Because of the symmetry of the problem, this cone is circular, and it grazes the sphere along a circle. In the planar perspective case however, the shape of the outline will depend on the orientation of the image plane (Figure 1.8, left): if this plane is perpendicular to the line joining the center of the sphere to the pinhole, the outline will be a circle, but in all other cases it will be a non-circular conic section,¹ usually an ellipse. In the spherical projection case, there is no plane orientation to account for, and the outline is always, by symmetry, a circle. Spheres also have circular outlines under orthographic projection. In this case, the tangent cone degenerates into a cylinder that intersects the image plane along a circle since its axis is always orthogonal to that plane.

In a sense, spherical perspective cameras are “better” than their planar counterparts since the pictures they produce only depend on the position of the pinhole. At the same time, they accurately capture the dependency of image size on object distance, and in that sense are also “better” than affine cameras. One could object that real cameras have planar retinas, but it is in fact easy to calibrate a planar perspective camera so it simulates a spherical one. Another drawback of spherical cameras is that they map straight lines onto circles, which at times complicates calculations. In reality these are not serious problems, and spherical camera models have their role to play in computer vision. Indeed, they have proven useful in the

¹Indeed, quadratic curves such as circles, ellipses, parabolas and hyperbolas are called conic sections because they are the various curves that you obtain when you slice a circular cone with a plane.

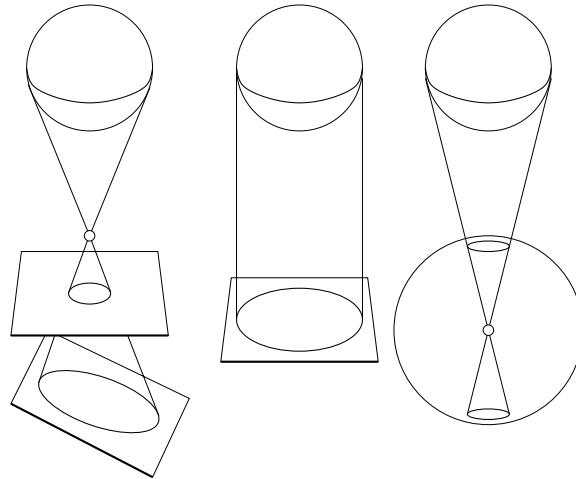


Figure 1.8. Different pictures of a sphere. From left to right: planar perspective, orthographic and spherical perspective projections.

study of shape (e.g., solids of revolution project onto bi-lateral symmetries under spherical perspective projection) and the analysis of motion (e.g., to recover the shape of smooth objects from sequences of pictures).

Let us close this section by noting that, although the eye has a (roughly) spherical retina, it does not obey the projection model described above since its “pinhole” (the pupil) is not located at the center of the corresponding sphere center.

1.2 Cameras with Lenses

Most real 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 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 (Figure 1.9). Shrinking the pinhole produces sharper images but reduces the amount of light reaching the image plane, and may introduce diffraction effects. Keeping the picture in sharp focus while gathering light from a large area is the second main reason for using a lens.

Ignoring diffraction, interferences and other physical optics phenomena, the behavior of lenses is dictated by the laws of geometric optics (Figure 1.10): light travels in straight lines (*light rays*) in homogeneous media. 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. When a ray passes

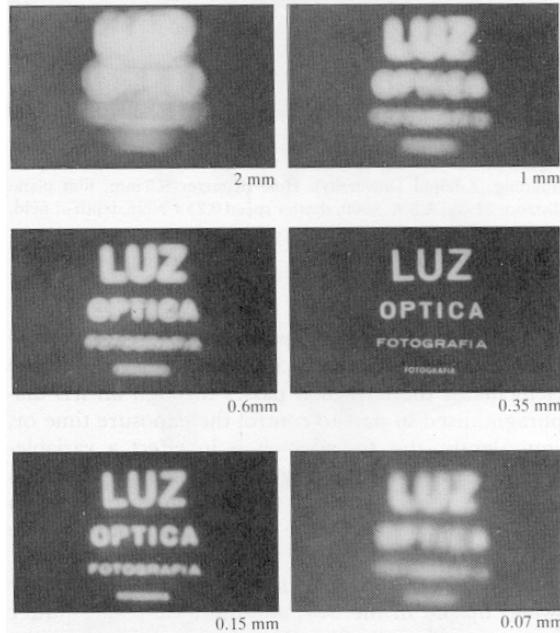


Figure 1.9. Images of some text obtained with shrinking pinholes: large pinholes give bright but fuzzy images but pinholes that are too small also give blurry images because of diffraction effects. Reprinted from [Hecht, 1987], Figure 5.108.

from one medium to another, it is *refracted*, i.e., its direction changes: according to Descartes' law, if r_1 is the ray incident to the interface between two transparent materials with indices 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 α_1 and α_2 between the normal and the two rays are related by

$$n_1 \sin \alpha_1 = n_2 \sin \alpha_2. \quad (1.2.1)$$

In this chapter, we will only consider the effects of refraction and ignore those of reflection. In other words, we will 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, and the next section discusses this case.

1.2.1 First-Order Geometric Optics

We consider in this section first-order (or *paraxial*) 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

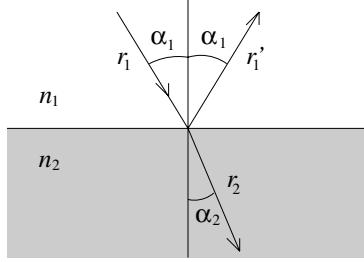


Figure 1.10. Reflection and refraction at the interface between two homogeneous media with indices of refraction n_1 and n_2 .

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 indices of refraction n_1 and n_2 (Figure 1.11). 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 symmetrical), and by C the center of the circular interface.

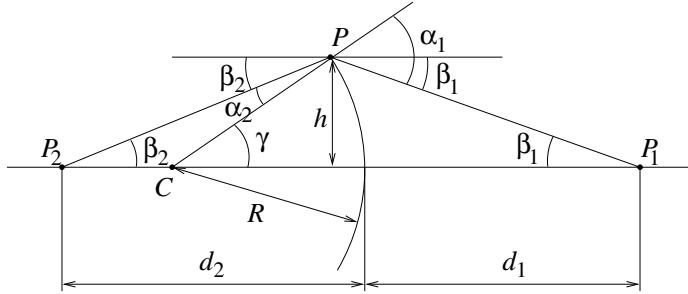


Figure 1.11. Paraxial refraction.

Let α_1 and α_2 respectively denote the angles between the two rays and the chord joining C to P . If β_1 (resp. β_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 the diagram above, $\gamma = \alpha_1 - \beta_1 = \alpha_2 + \beta_2$. Now, let us denote by h the distance between P and the optical axis, and by 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 (1.2.1) for small angles yields the *paraxial refraction equation*:

$$n_1 \alpha_1 \approx n_2 \alpha_2 \iff \frac{n_1}{d_1} + \frac{n_2}{d_2} = \frac{n_2 - n_1}{R}. \quad (1.2.2)$$

Note that the relationship between d_1 and d_2 depends on R , n_1 and n_2 but not on β_1 or β_2 . This is the main simplification introduced by the paraxial assumption. It is easy to see that (1.2.2) remains valid when some (or all) of the values of d_1 , d_2 and R becomes 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: Geometry

Let us now consider a lens with two spherical surfaces of radius R and index of refraction n . We will 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 and distance y from the optical axis, and let r_0 denote a ray passing through P and intersecting the optical axis in P_0 at (negative) depth z_0 and the lens in Q at a distance h from the optical axis (Figure 1.12).

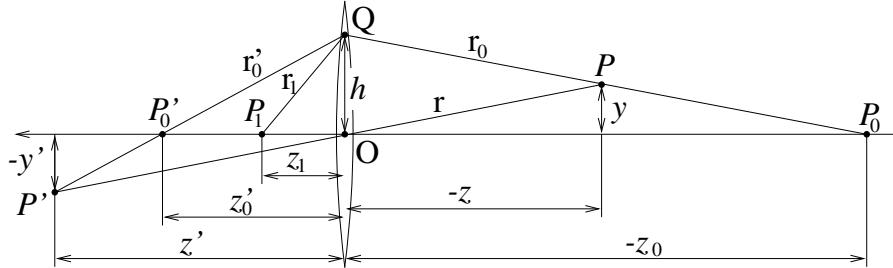


Figure 1.12. Image formation in the case of a thin lens.

Before constructing the image of P , let us first determine the image P'_0 of P_0 on the optical axis: after refraction at the right circular boundary of the lens, r_0 is transformed into a new ray r_1 intersecting the optical axis at the point P_1 whose depth z_1 verifies, according to (1.2.2),

$$\frac{1}{-z_0} + \frac{n}{z_1} = \frac{n-1}{R}.$$

The ray r_1 is immediately refracted at the left boundary of the lens, yielding a new ray r'_0 that intersects the optical axis in P'_0 . The paraxial refraction equation

can be rewritten in this case as

$$\frac{n}{-z_1} + \frac{1}{z'_0} = \frac{1-n}{-R},$$

and adding these two equation yields:

$$\frac{1}{z'_0} - \frac{1}{z_0} = \frac{1}{f}, \quad \text{where } f = \frac{R}{2(n-1)}. \quad (1.2.3)$$

Let r denote the ray passing through P and the center O of the lens, and let P' denote the intersection of r and r_0 , located at depth z' and at a distance $-y'$ of the optical axis (Figure 1.12). We have the following relations among the sides of similar triangles:

$$\left\{ \begin{array}{l} \frac{y}{h} = \frac{z - z_0}{-z_0} = (1 - \frac{z}{z_0}), \\ \frac{-y'}{h} = \frac{z' - z'_0}{z'_0} = -(1 - \frac{z'}{z'_0}), \\ \frac{y'}{z'} = \frac{y}{z}. \end{array} \right.$$

Eliminating h , y and y' between these equations and using (1.2.3) yields

$$\frac{1}{z'} - \frac{1}{z} = \frac{1}{f}. \quad (1.2.4)$$

In particular, all rays passing through the point P are focused by the thin lens on the point P' . 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 will only be in sharp focus when the image plane is located at a distance z' from O on the other side of the lens that satisfies (1.2.4).

The distance f is called the *focal length* of the lens, and (1.2.4) is called the *thin lens equation*. Letting $z \rightarrow -\infty$ shows that f is the distance between the center of the lens and the plane where points located at $z = -\infty$ (the sun, stars..) will 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 (Figure 1.13).

The equations relating the positions of P and P' are exactly the same as under pinhole perspective projection since these two points lie on a ray passing through the center of the lens, but points located at a distance $-z$ from O will only be in sharp focus when the image plane is located at the distance z' from O on the other side of the lens that satisfies (1.2.4). In practice, objects within some range of distances (called *depth of field* or *depth of focus*) will be in acceptable focus. 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.

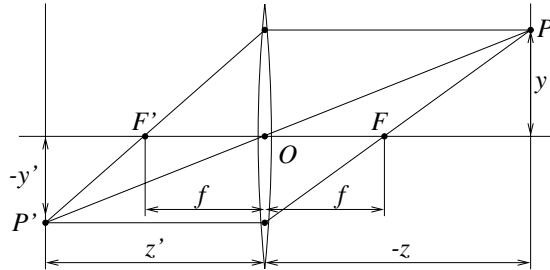


Figure 1.13. The focal points of a thin lens.

Note that the *field of view* of a camera, i.e., the portion of scene space that actually projects onto the retina of the camera, 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 1.14).

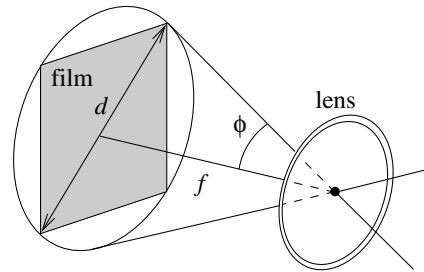


Figure 1.14. The field of view of a camera. It can be defined as 2ϕ , where $\phi \stackrel{\text{def}}{=} \arctan \frac{d}{2f}$, d is the diameter of the sensor (film or CCD chip) and f is the focal length of the camera.

When the focal length is (much) shorter than the effective diameter of the retina, 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. In addition, specially designed *telecentric* lenses offer a very good approximation of orthographic projection (Figure 1.15). These are useful in many situations, including part inspection.

1.2.3 Thin Lenses: Radiometry

Let us now consider an image patch $\delta A'$ centered in P' where a lens concentrates the light radiating from a scene patch δA centered in P (Figure 1.16). If $\delta\omega$ denotes



Figure 1.15. A telecentric lens: the NAVITAR TC-5028 model (<http://navitar.com/oem/tc5028.htm>).

the solid angle subtended by δA (or $\delta A'$) from the center O of the lens, we have

$$\left\{ \begin{array}{l} \delta\omega = \frac{\delta A' \cos \alpha}{|\overrightarrow{OP'}|^2} = \frac{\delta A' \cos \alpha}{(z'/\cos \alpha)^2}, \\ \delta\omega = \frac{\delta A \cos \beta}{|\overrightarrow{OP}|^2} = \frac{\delta A \cos \beta}{(z/\cos \alpha)^2}, \end{array} \right.$$

and it follows that

$$\frac{\delta A}{\delta A'} = \frac{\cos \alpha}{\cos \beta} \left(\frac{z}{z'}\right)^2.$$

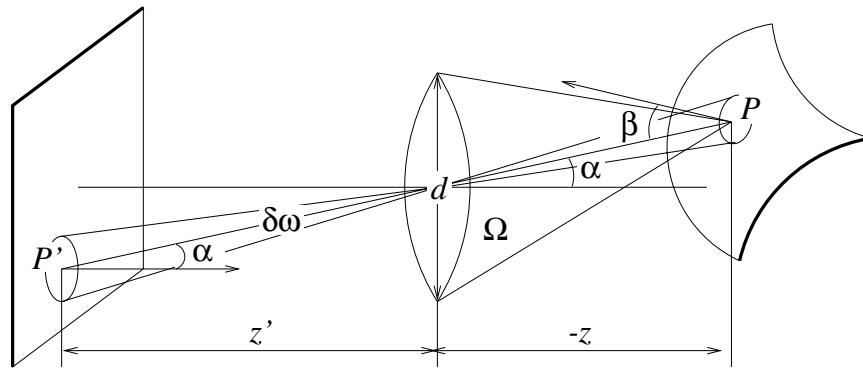


Figure 1.16. Object radiance and image irradiance.

Now, the area of a lens with diameter d is $\frac{\pi}{4}d^2$, and if Ω denotes the solid angle

sustended by the lens from P , we have

$$\Omega = \frac{\pi}{4} \frac{d^2 \cos \alpha}{(z/\cos \alpha)^2} = \frac{\pi}{4} \left(\frac{d}{z}\right)^2 \cos^3 \alpha.$$

Let L be the object radiance, the power δP emitted from the patch δA and falling on the lens is

$$\delta P = L \Omega \delta A \cos \beta = \frac{\pi}{4} \left(\frac{d}{z}\right)^2 L \delta A \cos^3 \alpha \cos \beta.$$

This power is concentrated by the lens on the patch $\delta A'$ of the image plane. If E denotes the image irradiance, we have

$$E = \frac{\delta P}{\delta A'} = \frac{\pi}{4} \left(\frac{d}{z}\right)^2 L \frac{\delta A}{\delta A'} \cos^3 \alpha \cos \beta,$$

and substituting the value of $\delta A/\delta A'$ in this equation finally yields

$$E = \left[\frac{\pi}{4} \left(\frac{d}{z'}\right)^2 \cos^4 \alpha \right] L. \quad (1.2.5)$$

This relationship is important for several reasons: first, it shows that the image irradiance is proportional to the object radiance. In other words, what we measure (E) is proportional to what we are interested in (L)! Second, the irradiance is proportional to the area of the lens and inversely proportional to the distance between its center and the image plane. The quantity $a = d/f$ is the *relative aperture* and is the inverse of the f number defined earlier. Equation (1.2.5) shows that E is proportional to a^2 when the lens is focused at infinity. Finally, the irradiance is proportional to $\cos^4 \alpha$ and falls off as the light rays deviate from the optical axis. For small values of α , this effect is hardly noticeable.

1.2.4 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 1.17): if H and H' denote the *principal points* of the lens, then (1.2.4) 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 equation (1.2.2) is only an approximation, valid when the angle α between each ray along the optical path and the optical axis of the length is small and $\sin \alpha \approx \alpha$. This corresponds to a first-order Taylor

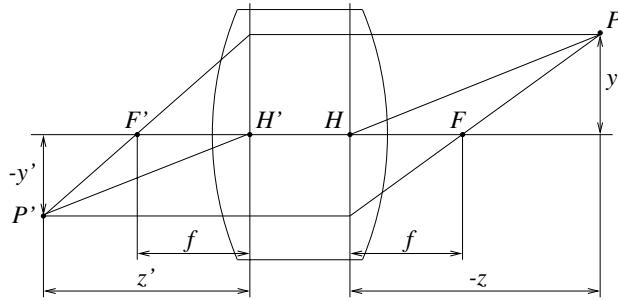


Figure 1.17. A simple thick lens with two spherical surfaces.

expansion of the sine function. For larger angles, additional terms yield a better approximation:

$$\sin \alpha = \alpha - \frac{\alpha^3}{3!} + \frac{\alpha^5}{5!} - \frac{\alpha^7}{7!} + \dots$$

In particular, it is easy to show that a third-order Taylor expansion yields the following refinement of the paraxial equation:

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

where h denotes, as in Figure 1.11, 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 1.18(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 Π' was 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 Π' . The size of that circle will change if we move Π' 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*. Like spherical aberration, coma, astigmatism and field curvature 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 1.18(b)). This effect is due to the fact that different areas of a lens have slightly

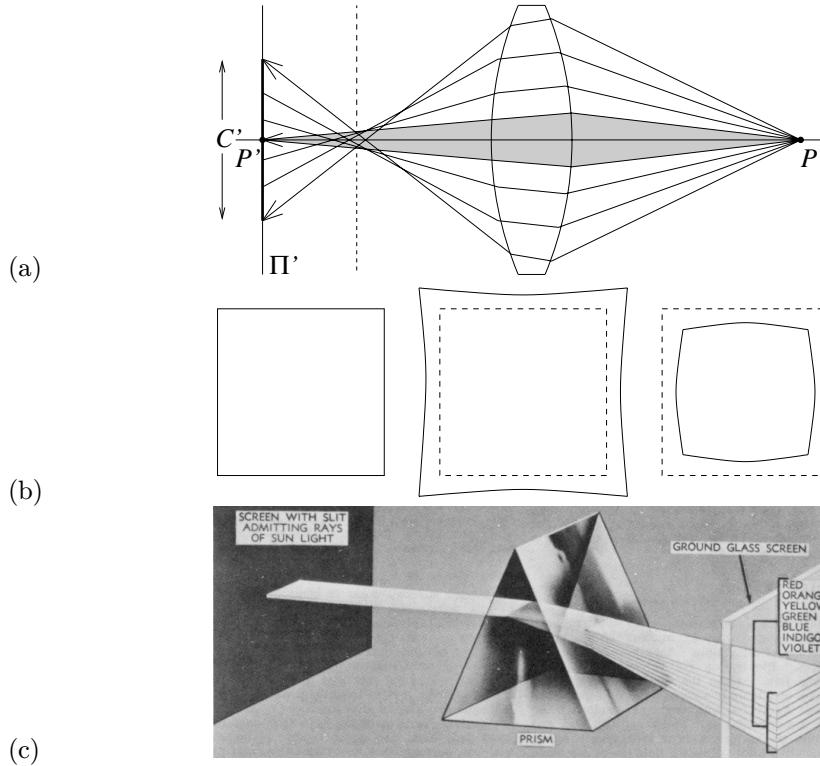


Figure 1.18. Aberrations. (a) Spherical aberration: the grey region is the paraxial zone where the rays issued from P intersect at its paraxial image P' . If an image plane Π' is erected in P' , the image of P' in that plane will form a circle of confusion of diameter C' . The focus plane yielding the 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 colour) of the incident light rays. Here, a prism decomposes white light into a palette of colours. Reprinted from [Navy, 1969], Figure 3-17.

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 medium depends on wavelength (Figure 1.18(c)), and it follows from the thin lens equation (1.2.4) that the focal length depends of wavelength as well. This causes the phenomenon of *chromatic aberration*: refracted rays corresponding to different wavelengths will 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 indices, separated by appropriate stops. These compound lenses can still be modelled by the thick lens equations. Figure 1.19 shows the general configuration of a few well-known photographic lenses.

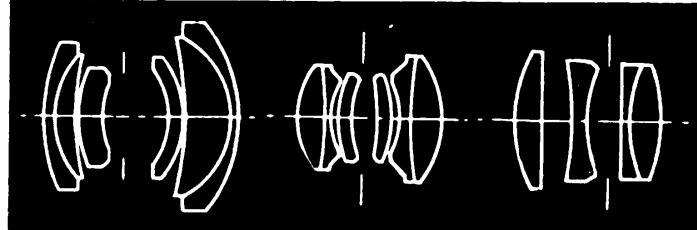


Figure 1.19. Photographic lenses. Reprinted from [Montel, 1972], p. 54.

These complex lenses 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 themselves) positioned inside the lens to limit aberrations. This phenomenon, called *vignetting*, causes the irradiance to drop in the image periphery. The vignetting fall-off can be much more noticeable than the $\cos^4 \alpha$ effect noted in the last section in typical situations. Note that vignetting is not as relevant in photography since the human eye is remarkably insensitive to smooth brightness gradients.

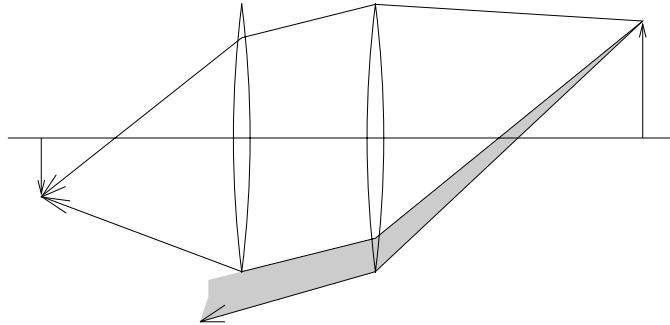


Figure 1.20. Vignetting effect in a two-lens system. The shaded part of the beam never reaches the second lens. Additional apertures and stops in a lens further contribute to vignetting.

Human Vision: The Structure of the Eye

Here we give a (very 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 1.21 (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*, that control the amount of light penetrating the eyeball; the *cornea* and the crystalline *lens*, that together refract the light to create the retinal image; and finally the *retina*, where the image is formed.

The human eyeball, despite its globular shape, is functionally similar to a camera with a field of view covering a 160° (width) $\times 135^\circ$ (height) area, and like any other optical system, it suffers 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 1.21(right) shows one of them, Helmholtz's schematic eye. There are only three refractive surfaces, with an infinitely thin cornea and a homogeneous lens. The constants given in Figure 1.21 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.

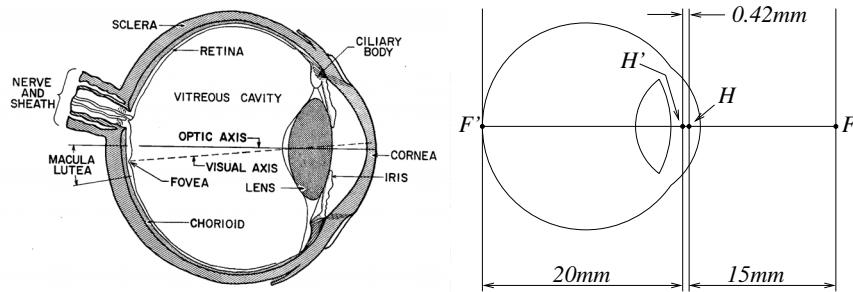


Figure 1.21. The eye. Left: the main components of the eye. Reprinted from [Thompson et al., 1966], Figure 11-2. Right: Helmholtz's schematic eye as modified by Laurance (after [Driscoll and Vaughan, 1978], Section 2, Figure 1). The distance between the pole of the cornea and the anterior principal plane is 1.96mm, and the radii of the cornea, anterior and posterior surfaces of the lens are respectively 8mm, 10mm and 6mm.

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 coloured and opaque surface of the iris. The pupil is an opening at the center of the iris whose diameter varies from about 1 to 8mm in response to illumination changes, controlling the amount of energy that reaches the retina and limiting the amount of image blurring due to imperfections of the eye's optics. The refracting power (reciprocal of the focal length) of the eye is in large part an effect of refraction at 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 ($1\text{diopter}=1\text{m}^{-1}$), corresponding to a range of focal lengths between 15 and 17mm. The retina itself is a thin, layered membrane populated by two types of photoreceptors, the *rods* and the *cones*, that respond to light in the 330-730nm wavelength (violet to red) range. As already mentioned in Chapter 4, there are three types of cones with different spectral sensitivities, and these play a key role

in the perception of colour.

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 1.21). The highest concentration of cones occurs in the *fovea*, a depression in the middle of the macula lutea where it peaks at $1.6 \times 10^5/\text{mm}^2$, with the centers of two neighboring cones separated by only half a minute of visual angle (Figure 1.22). Conversely, there are no rods in 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.

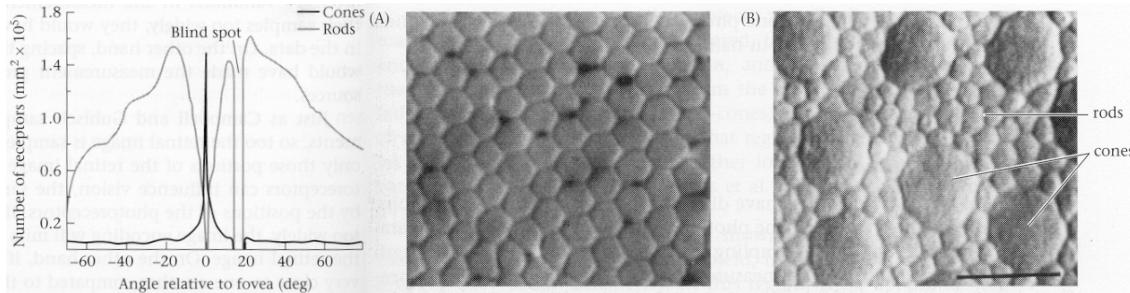


Figure 1.22. Rods and cones. Left: the distribution of rods and cones across the retina. Right: (A) cones in the fovea; (B) rods and cones in the periphery. Note that the size of the cones increases with eccentricity and that the presence of the rods disrupts the regular arrangement of the cones. Reprinted from [Wandell, 1995], Figures 3.1 and 3.4.

The rods are extremely sensitive photoreceptors, 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 levels, but the signal output by each cone in the fovea is encoded by several neurons, yielding a very 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, etc.. 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 will come back to various aspects of this endeavor later in this book.

1.3 Sensing

What differentiates a camera (in the modern sense of the world) from the portable camera obscura of the seventeenth 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 sun light, 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 in order 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) shown in Figure 1.23.

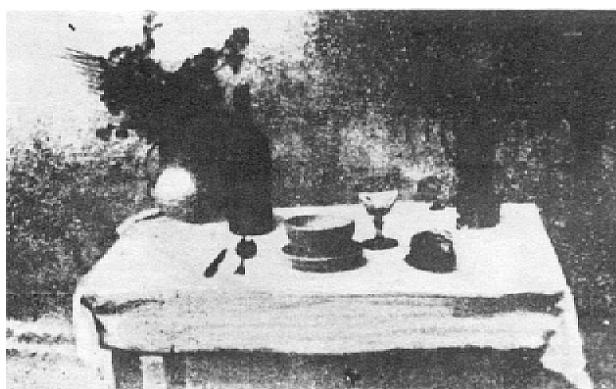


Figure 1.23. The first photograph on record, “la table servie”, obtained by Nicéphore Niepce in 1822. Reprinted from [Montel, 1972], p. 9.

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, and “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, that required the pictures to be developed on the spot but produced excellent negatives; the invention of the gelatin process by Maddox in 1870, that 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 colour photography in 1908.

The invention of television in the 1920s by Baird, Farnsworth, Zworykin and a few others was of course a major impetus for the development of electronic sensors.

The *vidicon* is a common type of television 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.

1.3.1 CCD cameras

Let us now turn to *charge-coupled-device* (or *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 1.24). Each site is formed by growing a layer of silicon dioxide on the wafer then depositing a conductive gate structure over the dioxide. When a photon strikes the silicon, an electron-hole pair is generated, and the electron is 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.

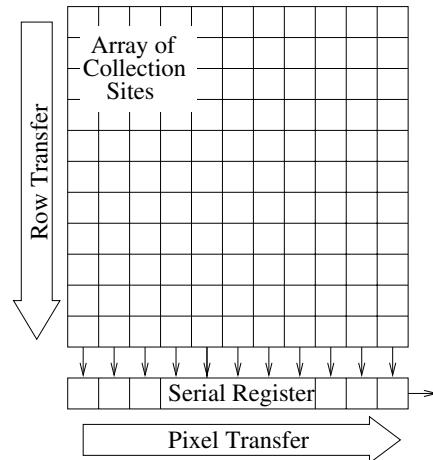


Figure 1.24. A CCD Device.

At this point, the charges stored at the individual sites are moved using *charge*

coupling: charge packets are transferred 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 transferred 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 (television 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 will construct the final digital image.

Consumer-grade colour CCD cameras essentially use the same chips as black-and-white cameras, except for the fact 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 4×4 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 colour filters. The individual colour channels are then either digitized separately (RGB output), or combined into a composite colour video signal (NTSC output in the United States) or into a component video format separating colour and brightness information.

1.3.2 Sensor Models

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

The number I of electrons recorded at a collection site can be modelled as

$$I = \int_t \int_\lambda \int_y \int_x E(x, y, \lambda, t) S(x, y) q(\lambda) dx dy d\lambda dt,$$

where E is the irradiance, S 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 λ , and E and S depend on the location of the site point considered. E also depends on time, but it is generally assumed to be constant over the integration interval.

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 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 below, these factors are appropriately captured by simple statistical models.

Fabrication defects cause the value of the spatial response and quantum efficiency to undergo small variations across the image. Lumping together the two effects, the distribution of electrons collected at each site can be modeled by KI , where K is a random variable with a mean of 1 and a spatial variance of σ_K^2 over the sites. Electrons freed from the silicon by thermal energy may add to the charge of each collection site (their contribution is called *dark current*). The number N_t of these electrons is proportional to the integration time and increases with temperature. It may also undergo small variations from site to site. The effect of dark current can be controlled by cooling down the camera. Quantum physics effects introduce an inherent uncertainty in the number of electrons stored at each site (*shot noise* N_s). Both dark current and shot noise can be modelled by integer random variables with Poisson distributions [Snyder *et al.*, 1993]. The output amplifier adds a read noise N_a that dominates shot noise for low intensities and can be modelled by a real-valued random variable with a Gaussian distribution. Finally, the discretization of the analog voltage by the frame grabber introduces both geometric effects (*line jitter*), that can be corrected via calibration, and a quantization noise that can be modeled as a zero-mean random variable N_q with a uniform distribution in the $[-\frac{1}{2}q, \frac{1}{2}q]$ interval and a variance of $\frac{1}{12}q^2$.

There are other sources of uncertainty (e.g., charge transfer efficiency) but they can often be neglected. This yields the following composite model for the digital signal D :

$$D = A(KI + N_t + N_s + N_a) + N_q,$$

where A 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 ($I = 0$), etc.

1.4 Notes

There are of course many books on geometric optics. [Hecht, 1987] is an excellent introduction, and 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], as well as [Navy, 1969] that provides a very accessible and colourful alternative. Vignetting is discussed in [Horn, 1986; Russ, 1995].

As noted earlier, [Wandell, 1995] gives an excellent treatment of image formation in the human visual system. The Helmholtz schematic model of the eye is detailed

in [Driscoll and Vaughan, 1978].

As noted earlier, spherical projection models have been used in shape and motion analysis. See [Cipolla and Blake, 1992; Nalwa, 1987].

CCD devices were introduced in [Boyle and Smith, 1970; Amelio *et al.*, 1970]. Scientific applications of CCD cameras to microscopy and astronomy are discussed in [Aikens *et al.*, 1989; Janesick *et al.*, 1987; Snyder *et al.*, 1993; Tyson, 1990]. The sensor model presented in this chapter is based on [Healey and Kondepudy, 1994], that also presents a method for radiometric camera calibration. A different model and its application to astronomical image restoration are discussed in [Snyder *et al.*, 1993].

1.5 Assignments

Exercises

1. Derive the perspective equation projections for a virtual image located at a distance f' *in front* of the pinhole.
2. Prove geometrically that the projections of two parallel lines lying in some plane Π appear to converge on a horizon line H formed by the intersection of the image plane with the plane parallel to Π and passing through the pinhole.
3. Prove the same result algebraically, using the perspective projection equation (1.1.1). You can assume for simplicity that the plane Π is orthogonal to the image plane.
4. What is the image of a circle under perspective projection? What about orthographic projection?
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 . What is the size of the blur circle obtained by imaging a point located at position $z + \delta z$ on the optical axis?
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.
7. Derive the thick lens equations in the case where both spherical boundaries of the lens have the same radius.
8. Derive the relationship between the scene radiance and image irradiance for a pinhole camera with a pinhole of diameter d .
9. Derive the relationship between the scene radiance and image irradiance for a spherical camera with a pinhole of diameter d .