## 5.1 Pinhole Camera

The history of the pinhole camera (or camera obscura) dates back as early as the fifth century B.C., and continues to be popular today among students, artists, and scientists. The Chinese philosopher Mo Ti is believed to be the first to notice that objects reflect light in all directions and that the light rays that pass through a small hole produce an inverted image. In its simplest form a pinhole camera is a light-tight box with a tiny hole in one end and a photo-sensitive material on the other. Remarkably, this simple device is capable of producing a photograph. However, the pinhole camera is not a particularly efficient imaging system (often requiring exposure times as long as several hours) and is more popular for its artistic value than for its practical value. Nevertheless, the pinhole camera is convenient because it affords a simple model of more complex imaging systems. That is, with a pinhole camera model, the projection of points from the three-dimensional world onto the two-dimensional sensor takes on a particularly simple form.

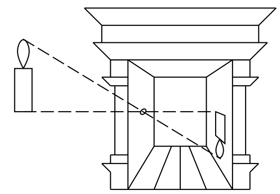

*Figure 5.1 Pinhole image formation*

Denote a point in the three-dimensional world as a column vector, $\vec{P} = (\, X \quad Y \quad Z\,)^t$ and the projection of this point onto the two dimensional image plane as $\vec{p} = (\, x \quad y\,)^t$. Note that the world and image points are expressed with respect to their own coordinate systems, and for convenience, the image coordinate system is chosen to be orthogonal to the Z-axis, i.e., the origins of the two systems are related by a one-dimensional translation along the $Z-$axis or *optical axis*. It is straight-forward to show from a similar triangles argument that the relationship between the world and image point is:

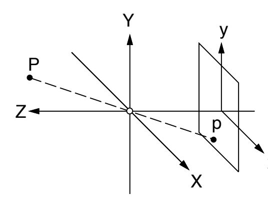


*Figure 5.2 Perspective projection*

$$x = -\frac{dX}{Z} \qquad \text{and} \qquad y = -\frac{dY}{Z}, \tag{5.1}$$

where $d$ is the displacement of the image plane along the Z-axis [6] These equations are frequently referred to as the *perspective projection* equations. Although non-linear in their nature, the perspective projection equations may be expressed in matrix form

[6] The value $d$ in Equation (5.1) is often referred to as the focal length. We do not adopt this convention primarily because it is a misnomer, under the pinhole model all points are imaged in perfect focus.

using the homogeneous equations:

$$\begin{pmatrix} x_s \\ y_s \\ s \end{pmatrix} = \begin{pmatrix} -d & 0 & 0 & 0 \\ 0 & -d & 0 & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix} \begin{pmatrix} X \\ Y \\ Z \\ 1 \end{pmatrix}, \tag{5.2}$$

where the final image coordinates are given by $( x \quad y )^t = ( \frac{x_s}{s} \quad \frac{y_s}{s} )^t$.

An approximation to the above perspective projection equations is *orthographic projection*, where light rays are assumed to travel from a point in the world parallel to the optical axis until they intersect the image plane. Unlike the pinhole camera and perspective projection equations, this model is not physically realizable and is used primarily because the projection equations take on a particularly simple linear form:

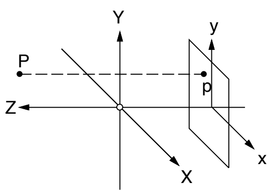

*Figure 5.3 Orthographic projection*

$$x = X \qquad \text{and} \qquad y = Y. \tag{5.3}$$

And in matrix form:

$$\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix} \begin{pmatrix} X \\ Y \\ Z \end{pmatrix} \tag{5.4}$$

Orthographic projection is a reasonable approximation to perspective projection when the difference in depth between points in the world is small relative to their distance to the image plane. In the special case when all the points lie on a single frontal-parallel surface relative to the image plane (i.e., $\frac{d}{Z}$ is constant in Equation (5.1)), the difference between perspective and orthographic is only a scale factor.

## *5.2 Lenses*

It is important to remember that both the perspective and orthographic projection equations are only approximations of more complex imaging systems. Commercial cameras are constructed with a variety of lenses that collect and focus light onto the image plane. That is, light emanates from a point in the world in all directions and, whereas a pinhole camera captures a single light ray, a lens collects a multitude of light rays and focuses the light to a small region on the image plane. Such complex imaging systems are often described with the simpler *thin-lens* model. Under the thin-lens model the projection of the central or *principal* ray obeys the rules of perspective projection, Equation (5.1): the point $\vec{P} = ( X \quad Y \quad Z )^t$ is projected onto the image plane centered about the point $( x \quad y )^t = ( \frac{-dX}{Z} \quad \frac{-dY}{Z} )^t$. If the point $\vec{P}$

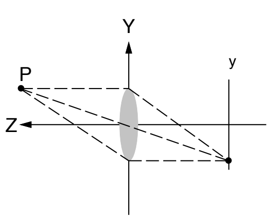

*Figure 5.4 Thin lens*

is in perfect focus, then the remaining light rays captured by the lens also strike the image plane at the point $\vec{p}$. A point is imaged in perfect focus if its distance from the lens satisfies the following *thin-lens equation*:

$$\frac{1}{Z} + \frac{1}{d} = \frac{1}{f}, \tag{5.5}$$

where $d$ is the distance between the lens and image plane along the optical axis, and $f$ is the focal length of the lens. The focal length is defined to be the distance from the lens to the image plane such that the image of an object that is infinitely far away is imaged in perfect focus. Points at a depth of $Z_o \neq Z$ are imaged onto a small region on the image plane, often modeled as a blurred circle with radius $r$:

$$r = \frac{R}{\frac{1}{f} - \frac{1}{Z_o}} \left| \left( \frac{1}{f} - \frac{1}{Z_o} \right) - \frac{1}{d} \right|, \tag{5.6}$$

where $R$ is the radius of the lens. Note that when the depth of a point satisfies Equation (5.5), the blur radius is zero. Note also that as the lens radius $R$ approaches 0 (i.e., a pinhole camera), the blur radius also approaches zero for all points independent of its depth (referred to as an infinite depth of field).

Alternatively, the projection of each light ray can be described in the following more compact matrix notation:

$$\begin{pmatrix} l_2 \\ \alpha_2 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ -\frac{1}{R}\left(\frac{n_2 - n_1}{n_2}\right) & \frac{n_1}{n_2} \end{pmatrix} \begin{pmatrix} l_1 \\ \alpha_1 \end{pmatrix}, \tag{5.7}$$

where $R$ is the radius of the lens, $n_1$ and $n_2$ are the index of refraction for air and the lens material, respectively. $l_1$ and $l_2$ are the height at which a light ray enters and exits the lens (the thin lens idealization ensures that $l_1 = l_2$). $\alpha_1$ is the angle between the entering light ray and the optical axis, and $\alpha_2$ is the angle between the exiting light ray and the optical axis. This formulation is particularly convenient because a variety of lenses can be described in matrix form so that a complex lens train can then be modeled as a simple product of matrices.

Image formation, independent of the particular model, is a three-dimensional to two-dimensional transformation. Inherent to such a transformation is a loss of information, in this case depth information. Specifically, all points of the form $\vec{P}_c = (\, cX \quad cY \quad cZ \,)^t$, for any $c \in \mathcal{R}$, are projected to the same point $(\, x \quad y \,)^t$ - the projection is not one-to-one and thus not invertible. In addition to this geometric argument for the non-invertibility of image formation, a similarly straight-forward linear algebraic argument holds.

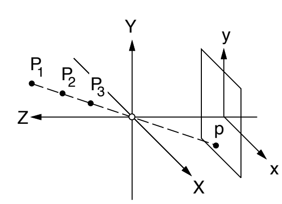

*Figure 5.5*
*Non-invertible projection*

In particular, we have seen that the image formation equations may be written in matrix form as, $\vec{p} = M_{n \times m}\vec{P}$, where $n < m$ (e.g., Equation (5.2)). Since the projection is from a higher dimensional space to a lower dimensional space, the matrix $M$ is not invertible and thus the projection is not invertible.

### *5.3 CCD*

To this point we have described the geometry of image formation, how light travels through an imaging system. To complete the image formation process we need to discuss how the light that strikes the image plane is recorded and converted into a digital image. The core technology used by most digital cameras is the charge-coupled device (CCD), first introduced in 1969. A basic CCD consists of a series of closely spaced metal-oxide-semiconductor capacitors (MOS), each one corresponding to a single image pixel. In its most basic form a CCD is a charge storage and transport device: charge is stored on the MOS capacitors and then transported across these capacitors for readout and subsequent transformation to a digital image. More specifically, when a positive voltage, $V$, is applied to the surface of a P-type MOS capacitor, positive charge migrates toward ground. The region depleted of positive charge is called the depletion region. When photons (i.e., light) enter the depletion region, the electrons released are stored in this region. The value of the stored charge is proportional to the intensity of the light striking the capacitor. A digital image is subsequently formed by transferring the stored charge from one depletion region to the next. The stored charge is transferred across a series of MOS capacitors (e.g., a row or column of the CCD array) by sequentially applying voltage to each MOS capacitor. As charge passes through the last capacitor in the series, an amplifier converts the charge into a voltage. An analog-to-digital converter then translates this voltage into a number (i.e., the intensity of an image pixel).

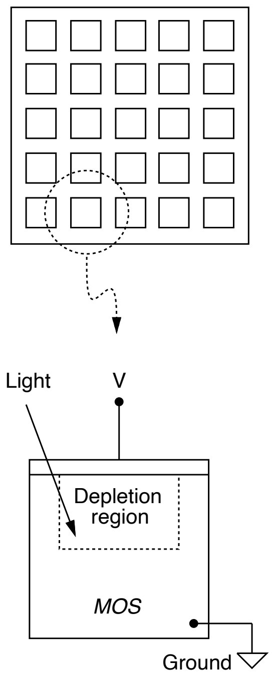


*Figure 5.6 MOS capacitor*