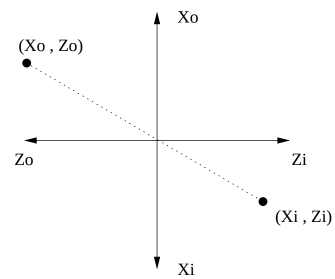
Up to now, we have taken the projection plane to be in front of the center of projection. Of course, the physical projection planes that are found in cameras (and eyes) are behind the center of the projection. For this reason, it will be convenient for us to consider two coordinate systems: (X_o, Y_o, Z_o) represents the coordinates of a point in the scene, and (X_i, Y_i, Z_i) represents a point behind the center of projection. (Subscript "o" is for object, and "i" is for image.)

Since the images on the projection plane behind the center of projectoin are upside-down and backwards, we orient the axis of X_i to be opposite to X_o and similarly for Y_i vs Y_o and Z_i versus Z_o. See figure. The two coordinate systems share the same origin.

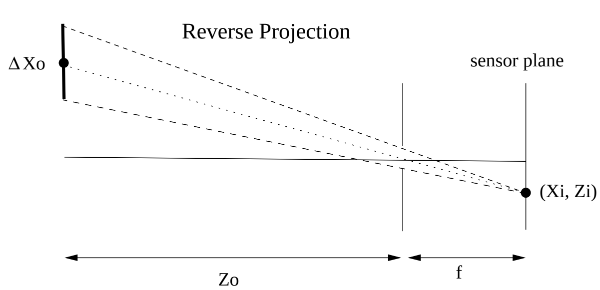


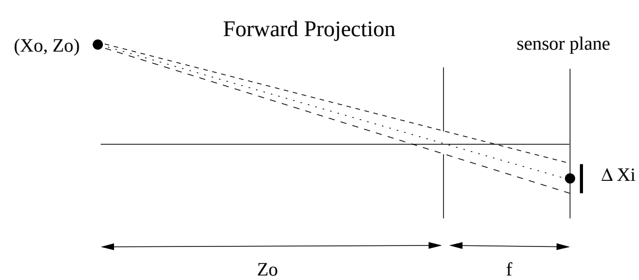
Non-pinhole cameras

The model that we have been discussing up to now assumes that we are projecting towards a single point – the center of projection. If we project to a plane behind the origin, we have a pinhole camera. The idea of a pinhole camera is that we are allowing light to pass through a tiny hole in the $Z = 0$ plane and forming an image on a plane inside a black box.

What happens if we open the pinhole so that the opening has a width A (which we refer to as the *aperture*) ? Suppose the light passes through the aperture and arrives at a plane at depth f behind the aperture. For simplicity suppose the point we are considering in the scene lies on a *surface* of constant depth Z_0 i.e. a wall that is oriented parallel to the sensor plane. The resulting image will be blurred.

We can think of the resulting blurring in two ways. See figures below. Each point on the sensor plane (X_i, Z_i) (where $Z_i = f$ in this case) will receive light from an area of points on the scene plane, namely from the point that are visible through the aperture. (This is sometimes known as *reverse projection*.)





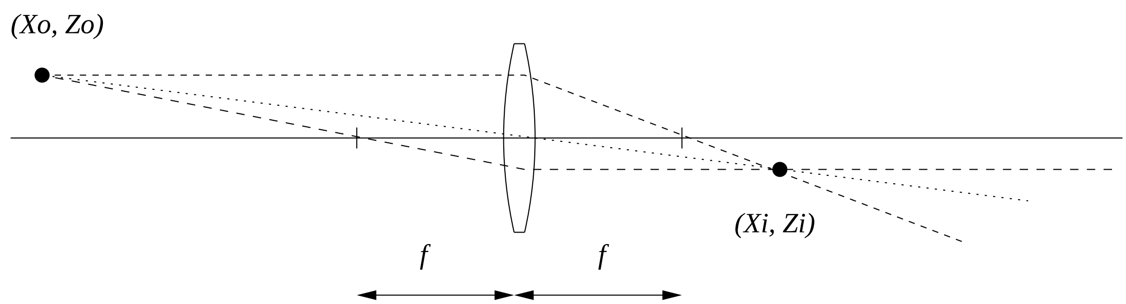
Alternatively, consider the point (X_o, Z_o) that projects to (X_i, f) through the center of the aperture, which we take to be the origin. This (X_i, f) is a single imaged point when the aperture goes to zero but for a finite aperture, there is a set of rays from (X_o, Z_o) that pass through the aperture and reach an area on the sensor plane. (This is sometimes called *forward projection*.)

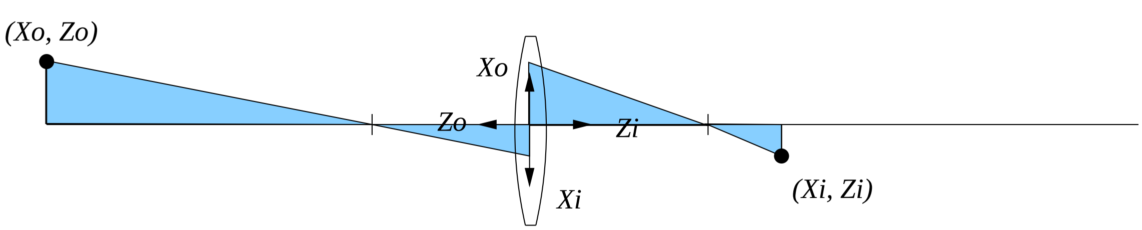
Thin lens model

Real cameras (and real eyes) indeed have apertures. These serve to allow more light to reach the sensors than a pinhole camera would. Cameras and eyes also have lenses, which redirect the light and focus it. We will consider a simple model of the optics of lenses called the *thin lens* model.

The thin lens model *assumes* that, for any point on an object (X_o, Y_o, Z_o), the light rays that diverge from that point and that pass through the lens all converge at some image point (X_i, Y_i, Z_i) behind the lens. Such points (X_o, Y_o, Z_o) and (X_i, Y_i, Z_i) are called *conjugate pairs*. Note that this is just a *model*. Real lenses behave this way only approximately, and only when certain conditions are met – you'll need to check out an Optics text if you want to understand what these conditions are.

The relationship between the coordinates of a conjugate pair can be derived as follows. Consider first the case of a point $(X_o, Y_o, Z_o) = (0, 0, \infty)$ which is the point at infinity in the direction of the optical axis – or $(0, 0, 1, 0)$ in homogeneous coordinates. The rays leaving this point and arriving at the $Z = 0$ plane are all parallel and they pass through the lens and converge at a point $(X_i, Y_i, Z_i) = (0, 0, f)$ which is also (by symmetry) on the optical axis. This constant f is called the *focal length* of the lens. This constant depends on the curvature of the two sides of the lens and on the material of the lens (e.g. glass vs. water vs ...). Note: f does not depend on the distance from the lens to the sensor plane, since obviously the lens does not know where the sensor plane is. Thus, we are using f differently from how we used it in the previous lectures! This will make more sense later.





One can relate the variables of a conjugate pair by assuming that the ray that passes through the origin (the center of the lens) does not change direction[1] and so, by similar triangles, we have

$$\frac{X_i}{Z_i} = \frac{X_o}{Z_0} .$$

Another useful relationship comes from similar triangles. There are two similar triangles in front of the lens, giving

$$\frac{X_i}{f} = \frac{X_o}{Z_0 - f}$$

and there are similar triangles behind the lens, giving

$$\frac{X_o}{f} = \frac{X_i}{Z_i - f}.$$

By rewriting each equation in terms of $\frac{X_i}{X_o}$ and then performing a few lines of algebra (do it!), one can isolate a relationship between Z_o, Z_i, f, namely the *thin lens formula*:

$$\frac{1}{Z_o} + \frac{1}{Z_i} = \frac{1}{f}$$

Notice that if $Z_o \to \infty$, then $Z_i \to f$. In particular, if an object is very far away then all the rays from that object (which will be roughly parallel when they arrive at the lens) will converge at the depth $Z_i = f$ behind the lens. We saw this earlier for the special case of a point on the optical axis, but now we see according to the model this property holds for all points at infinity.

Another interesting observation comes when we rewrite the thin lens equation as:

$$Z_i = \frac{Z_0 f}{Z_0 - f} .$$

We now see that the transformation from (X_o, Y_o, Z_o) to (X_i, Y_i, Z_i) can be written:

$$(X_i, Y_i, Z_i) = (f\frac{X_o}{Z_0 - f}, f\frac{Y_o}{Z_0 - f}, \frac{Z_0 f}{Z_0 - f}).$$

We can represent this transformation from a scene point to its image point using homogeneous coordinates:

$$\begin{bmatrix} X_i \\ Y_i \\ Z_i \\ 1 \end{bmatrix} \equiv \begin{bmatrix} f & 0 & 0 & 0 \\ 0 & f & 0 & 0 \\ 0 & 0 & f & 0 \\ 0 & 0 & 1 & -f \end{bmatrix} \begin{bmatrix} X_o \\ Y_o \\ Z_o \\ 1 \end{bmatrix}$$

Note that this transform is its own inverse i.e. any point is the conjugate point of its conjugate point, in the sense that

$$\begin{bmatrix} f & 0 & 0 & 0 \\ 0 & f & 0 & 0 \\ 0 & 0 & f & 0 \\ 0 & 0 & 1 & -f \end{bmatrix} \begin{bmatrix} f & 0 & 0 & 0 \\ 0 & f & 0 & 0 \\ 0 & 0 & f & 0 \\ 0 & 0 & 1 & -f \end{bmatrix} \equiv \begin{bmatrix} f^2 & 0 & 0 & 0 \\ 0 & f^2 & 0 & 0 \\ 0 & 0 & f^2 & 0 \\ 0 & 0 & 0 & f^2 \end{bmatrix} \equiv \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}.$$

[1]There is a bending of the light at the lens – refraction – but it turns out that, at the origin, the bending at the front and back face cancel out

One final point: Most cameras people use these days do not have a single lens but rather they have a set or system of lenses. These lenses are all centered on the same optical axis, and the camera is designed so that the user can move them relative to each other. The main effect of this design is that it allows the user to *change the focal length* f. This is what an "optical zoom" camera does. Note that the position of the center of the lens i.e. the origin of the camera coordinate system does not necessarily correpond to the physical center of any of the lens elements. Rather, we should think of a *virtual* lens center for an equivalent thin lens of focal length f. From now on (and in Assignment 1), we will pretend there is a single lens and continue to talk about its center.

Sensor plane and blur

Suppose we put a sensor plane at distance Z_s from the lens center. The conjugate points will lie at some depth Z_s^* in the scene, according to

$$\frac{1}{f} = \frac{1}{Z_s} + \frac{1}{Z_s^*}$$

namely

$$Z_s^* = \frac{Z_s f}{Z_s - f}$$

defines the *focal plane* in the scene.

Points that do not lie at depth Z_s^* will not be "in focus", in that the rays from such points will not converge on a single point in the sensor plane. Rather they will converge on a single point that is either in front of or behind the sensor plane. Thus, rays leaving a object point will strike an *area* on the sensor plane. This is illustrated below.

If the lens aperture is a disk, then the rays from (X_o, Y_o, Z_o) will arrive at a roughly disk shaped region on the sensor. This disk is often called the *circle of confusion*. It is a roughly a circle because the lens aperture is (roughly) circularly shaped. This is easiest to imagine if you consider the scene point $(0, 0, Z_o)$. since the scene and lens would all be rotationally symmetric about the optical axis, the blur region would also be rotationally symmetric.

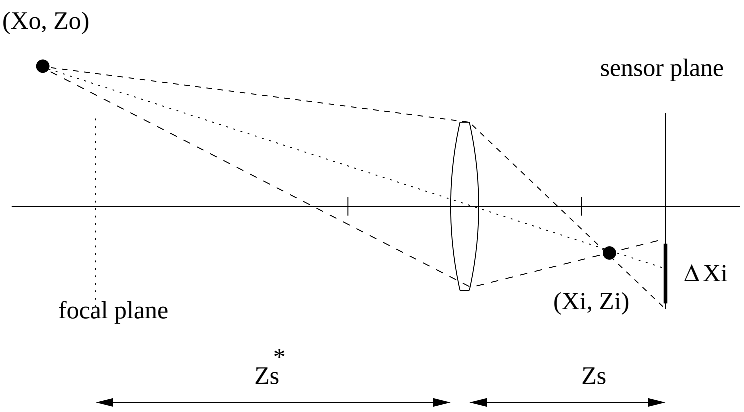
Let's next derive an expression for the blur width ΔX_i. We now consider a point at depth Z_o such that $Z_o \neq Z_s^*$. Let ΔX_i be the diameter of the *blur disk*. (In the figure, the blur disk is a dark line on the sensor plane. For a 2D sensor plane, it is a disk.) Let A be the diameter of the lens aperture. Then, by similar triangles,

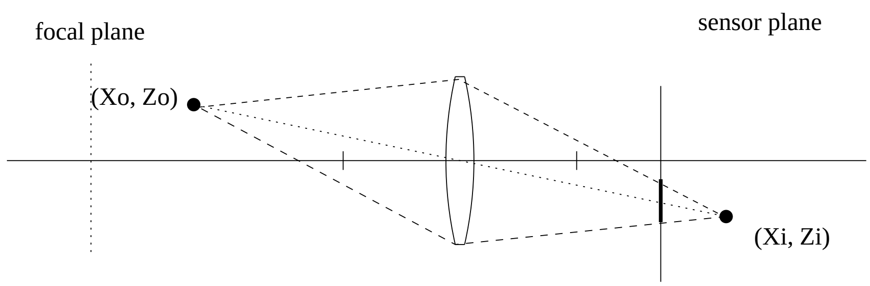
$$\frac{A}{Z_i} = \frac{\Delta X_i}{\mid Z_s - Z_i \mid}$$

and so

$$\Delta X_i = A \mid \frac{Z_s}{Z_i} - 1 \mid \; = A \mid Z_s \left(\frac{1}{f} - \frac{1}{Z_o} \right) - 1 \mid$$

For a given photograph, all the terms on the right hand side are constant except for Z_o, and so we see that blur width is a constant plus another constant times $\frac{1}{Z_o}$. (Verify that if $Z_o = Z_s^*$, then the blur width is 0.) This is a very simple relationship. For example, recall from lecture 1 that if there is a plane in the scene then inverse depth $\frac{1}{Z}$ varies linearly across the image coordinates (x, y). Hence the blur width would vary linearly across the image as well. We will return to this in Part 3 of the course.





f-number (f-stop)

As we will see in an upcoming lecture, the amount of light that arrives at a point on the sensor plane depends on the number of rays that arrive at that point, and this in turn depends on the angle subtended by the lens. The angle subtended by the lens is approximately $\frac{A}{f}$ radians, where A is the width of the aperture and f is the distance from the aperture to the sensor. (More precisely, the angle is approximately $\frac{1}{2} atan(\frac{A}{2f})$ radians. The inverse of this ratio is called the *f-number* (or *f-stop*):

$$N \equiv \frac{f}{A}$$

While is very common in photography to refer to the focal length of the lens explicitly (in units of mm), it is much less common to refer to aperture explicitly. Instead, one refers to the aperture as a particular "$f/\#$" where # denotes a particular f-number (i.e. N). For example, "$f/4$" refers to an aperture (in mm units) that corresponds to a particular focal length f and f-number of 4. This is confusing for novice photographers, and indeed even experienced photographers sometimes write "f/4" when talking about the f-number 4, when in fact "f/4" refers to an aperture.