CHAPTER 26

Cameras

26.1 THE PINHOLE CAMERA

A *pinhole camera* is a light-tight box with a very small hole in the front. Think about a point on the back of the box. The only light that arrives at that point must come through the hole, because the box is light-tight. If the hole is very small, then the light that arrives at the point comes from only one direction. This means that an inverted image of a scene appears at the back of the box (Figure 26.1). An appropriate sensor (CMOS sensor; CCD sensor; light sensitive film) at the back of the box will capture this image.

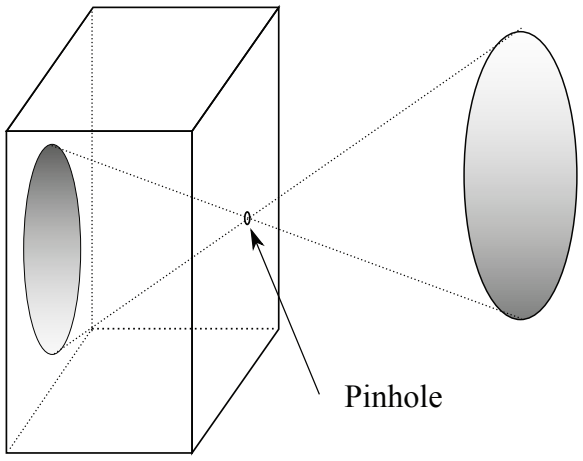


FIGURE 26.1: *In the pinhole imaging model, a light-tight box with a pinhole in it views an object. The only light that a point on the back of the box sees comes through the very small pinhole, so that an inverted image is formed on the back face of the box.*

Pinhole camera models produce an upside-down image. This is easily dealt with in practice (turn the image the right way up). An easy way to account for this is to assume the sensor is *in front* of the hole, so that the image is not upside-down. One could not build a camera like this (the sensor blocks light from the hole) but it is a convenient abstraction. There is a standard model of this camera, in a standard coordinate system (Figure 26.2). Notice that the y axis goes *down* in the image. While this is usual for image coordinate systems, there are further reasons to do this. Most people's intuition is that z *increases* as one moves into the image, and

orienting the y axis downward in the image allows me to achieve this, have x in the usual direction, and use a right-handed coordinate system. The pinhole – usually called the *focal point* – is at the origin, and the sensor is on the plane $z = f$. This plane is the *image plane*, and f is the *focal length*. We ignore any camera body and regard the image plane as infinite.

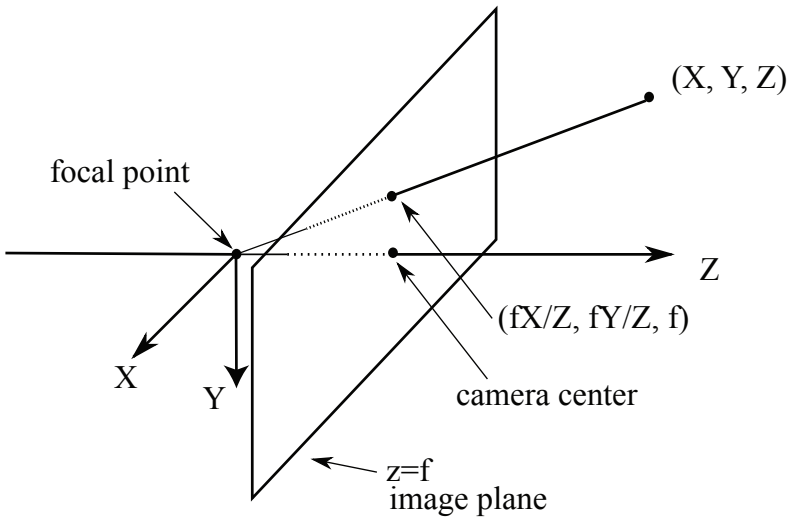


FIGURE 26.2: *The usual geometric abstraction of the pinhole model. The box doesn't affect the geometry, and is omitted. The pinhole has been moved to the back of the box, so that the image is no longer inverted. The image is formed on the plane $z = f$, by convention. Notice the y-axis goes down in the image. This allows me to use a* right handed *coordinate system and also have z* increase *as one moves into the image.*

Under this highly abstracted camera model, almost any point in 3D will map to a point in the image plane. We *image* a point in 3D by constructing a ray through the 3D point and the focal point, and intersecting that ray with the image plane. The focal point has an important, distinctive, property: It cannot be imaged, and it is the only point that cannot be imaged.

Similar triangles yields that the camera maps a point $\mathbf{X}$ in 3D to a point $\mathbf{x}$ on the image plane by:

$$\mathbf{X} = \begin{pmatrix} X \\ Y \\ Z \end{pmatrix} \rightarrow \begin{pmatrix} fX/Z \\ fY/Z \\ f \end{pmatrix} = \mathbf{x}.$$

Notice that the z-coordinate is the same for each point on the image plane, so it is quite usual to ignore it and use the model

$$\mathbf{X} = \begin{pmatrix} X \\ Y \\ Z \end{pmatrix} \rightarrow \begin{pmatrix} fX/Z \\ fY/Z \end{pmatrix} = \mathbf{x}.$$

The focal length just scales the image. In standard camera models, other scaling effects occur as well, and we write projection as if $f = 1$, yielding

$$\mathbf{X} = \begin{pmatrix} X \\ Y \\ Z \end{pmatrix} \rightarrow \begin{pmatrix} X/Z \\ Y/Z \end{pmatrix} = \mathbf{x}.$$

The projection process is known as *perspective projection*. The point where the z-axis intersects the image plane (equivalently, where the ray through the focal point perpendicular to the image plane intersects the image plane) is the *camera center* or the *principal point*.

Remember this: *Most practical cameras can be modelled as a pinhole camera. The standard model of the pinhole camera maps*

$$(X, Y, Z) \rightarrow (X/Z, Y/Z).$$

Figure 26.1 shows important terminology (focal point; image plane; camera center or principal point).

26.1.1 Perspective Effects

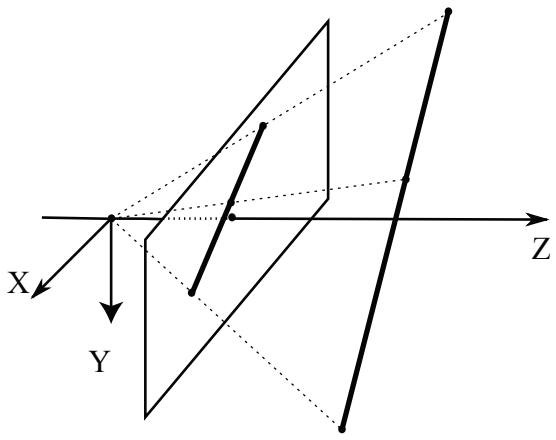


FIGURE 26.3: *Perspective projection maps almost any 3D line to a line in the image plane. Some rays from the focal point to points on the line are shown as dotted lines. The family of all such rays is a plane, and that plane must intersect the image plane in a line as long as the 3D line does not pass through the focal point.*

Perspective projection has a number of important properties, summarized as:

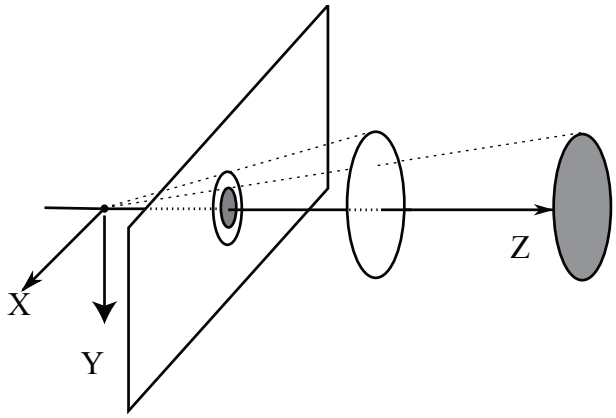


FIGURE 26.4: *Two 3D objects of the same size viewed in perspective projection. Division by Z (or, equivalently, similar triangles) means the more distant object appears smaller in the image.*

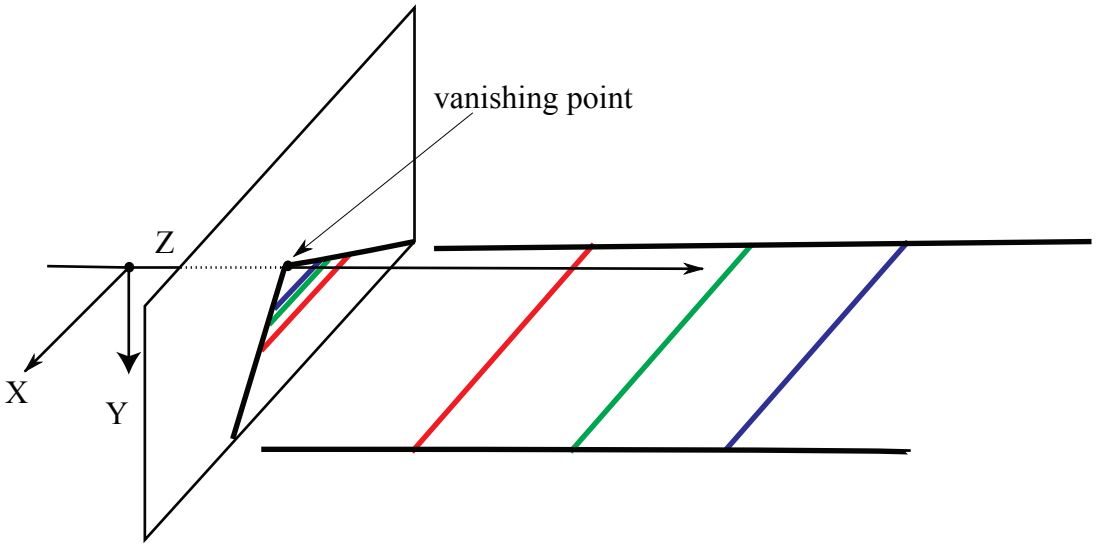


FIGURE 26.5: *Perspective projection maps a set of parallel lines to a set of lines that meet in a point. The figure shows a set of lines parallel to the z-axis, with "railway sleepers" shown. As these sleepers get further away, they get smaller in the image, meaning the projected lines must meet. The vanishing point (the point where they meet) is obtained by intersecting the ray parallel to the lines and through the focal point with the image plane.*

- lines project to lines (Figure 26.3);
- more distant objects are smaller (Figure 26.4);

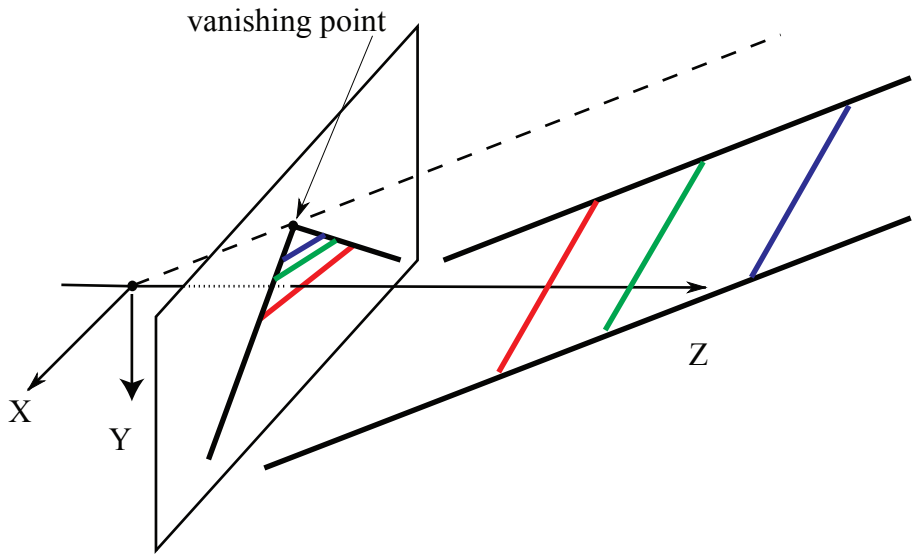


FIGURE 26.6: *Here are a different pair of parallel lines from Figure 26.5, with a different vanishing point, and "railway sleepers" still shown. Just as in Figure 26.5, as these sleepers get further away, they get smaller in the image, meaning the projected lines must meet. Just as in Figure 26.5, the vanishing point (the point where they meet) is obtained by intersecting the ray parallel to the lines and through the focal point with the image plane.*

- lines that are parallel in 3D meet in the image (Figure 26.5 and Figure 26.6);
- planes image as half-planes (Figure 26.7);
- planes have horizons (Figure 26.7);
- shapes are foreshortened (Figure **??**).

Lines project to lines: Almost every line in 3D maps to a line in the image. You can see this by noticing that the image of the 3D line is formed by intersecting rays from the focal point to each point on the 3D line with the image plane. But these rays form a plane, so we are intersecting a plane with the image plane, and so obtain a line (Figure 26.3). The exceptions are the 3D lines through the focal point – these project to points.

More distant objects are smaller: The further away an object is in 3D, the smaller the image of that object, because of the division by Z (Figure 26.4).

Lines that are parallel in 3D meet in the image: Now think about a set of infinitely long parallel railroad tracks. The sleepers supporting the tracks are all the same size. Distant sleepers are smaller than nearby sleepers, and arbitrarily distant sleepers are arbitrarily small. This means that parallel lines will meet in the image. The point at which the lines in a collection of parallel lines meet is known

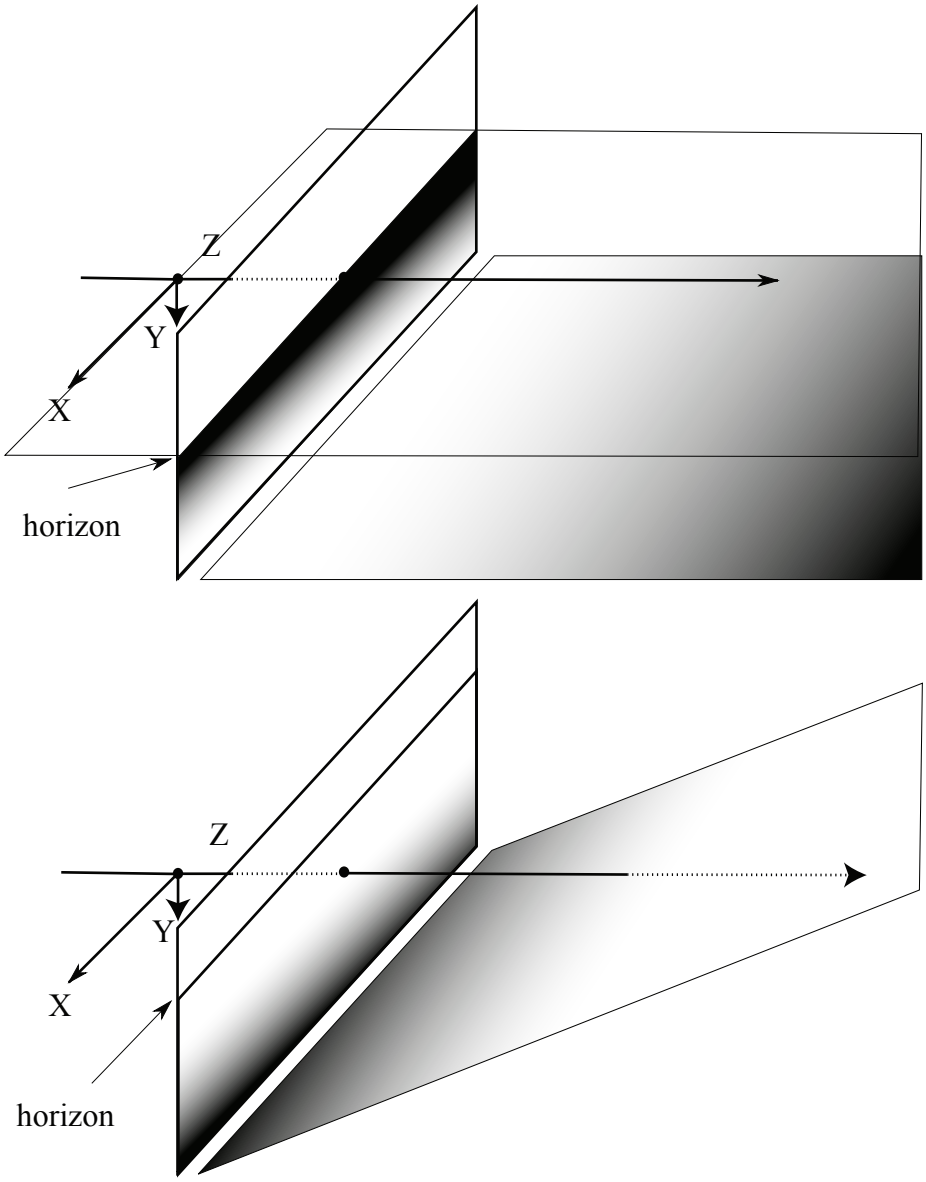


FIGURE 26.7: *Two different world planes and their horizons. The gradients on the planes indicate roughly where points on the 3D plane appear in the image plane (light points map to light, dark to dark).*

as the *vanishing point* for those lines (Figure 26.5 and Figure **??**). The vanishing point for a set of parallel lines can be obtained by intersecting the ray from the focal point and parallel to those lines with the image plane.

Planes image as half-planes: The **top** of Figure 26.7 shows the plane

$y = -1$ being viewed by a perspective camera. The intersection of the plane through the focal point parallel to the 3D plane (in this case, $y = 0$) and the image plane, forms an image line called the *horizon*. This line cuts the image plane into two parts. This applies to any plane in 3D, as the **bottom** of the Figure shows. Construct the ray through the focal point and a point $\mathbf{x}$ in the image plane. For $\mathbf{x}$ on one side of the horizon, this ray will intersect the 3D plane in the half space $z > 0$ (and so in front of the camera, shown here). If $\mathbf{x}$ is on the other side of the horizon, the intersection will be in the half space $z < 0$ (and so behind the camera, where it cannot be seen).

Planes have horizons: Figure 26.7 shows the horizons of two different planes being viewed by a perspective camera. The horizon is a line in the image formed by the set of vanishing points for all pairs of parallel lines on the plane. Alternatively and equivalently, the horizon is the line of image points where the ray through the focal point and the image point intersects the world plane infinitely far away. Alternatively and equivalently, the horizon is the intersection between the image plane and the plane parallel to the world plane through the focal point.

On images of planes, closer to the horizon is further away: This is an occasionally useful little fact that is regularly rediscovered. Write $\mathbf{x}$ and $\mathbf{x}'$ for two image points that are the image of two points $\mathbf{X}$ and $\mathbf{X}'$ lying on a world plane *that isn't parallel to the image plane*. Write $d_h(\mathbf{x})$ for the perpendicular distance from $\mathbf{x}$ to the horizon of that plane, etc. Write $d_f(\mathbf{X})$ for the distance from the focal point to $\mathbf{X}$. It is an exercise to show that

$$d_h(\mathbf{x}') > d_h(\mathbf{x}) \equiv d_f(\mathbf{X}') > d_f(\mathbf{X})$$

and

$$d_h(\mathbf{x}') = d_h(\mathbf{x}) \equiv d_f(\mathbf{X}') = d_f(\mathbf{X}).$$

Shapes are foreshortened: Figure **??** shows rectangles (the boxes between the "sleepers" in the figure) being imaged as parallelipipeds. It is an exercise to show that, in a perspective camera, a circle can be imaged as an ellipse; as a parabola; or as one branch of a hyperbola. The general term for these and other changes of shape is *foreshortening*.

26.1.2 Scaled Orthographic Projection

Under some circumstances, perspective projection can be simplified. Assume the camera views a set of points which are close to one another compared with the distance to the camera. Write $\mathbf{X}_i = (X_i, Y_i, Z_i)$ for the i'th point, and assume that $Z_i = Z(1+\epsilon_i)$, where ϵ_i is quite small. In this case, the distance to the set of points is much larger than the *relief* of the points, which is the distance from nearest to furthest point. The i'th point projects to $(fX_i/Z_i, fY_i/Z_i)$, which is approximately $(f(X_i/Z)(1-\epsilon_i), f(Y_i/Z)(1-\epsilon_i))$. Ignoring ϵ_i because it is small, we have the projection model

$$(X, Y, Z) \to (f/Z)(X, Y) = s(X, Y).$$

This model is usually known as *scaled orthograpic projection*. A geometric view of this model is that points in 3D "slide" down rays perpendicular to the image plane to form their image (Figure 26.8). It is an exercise to show that parallel lines do

not have vanishing points and that planes do not have horizons in scaled orthographic cameras. However, scaled orthographic cameras *do* foreshorten shapes, as Figure 26.8 and the exercises show.

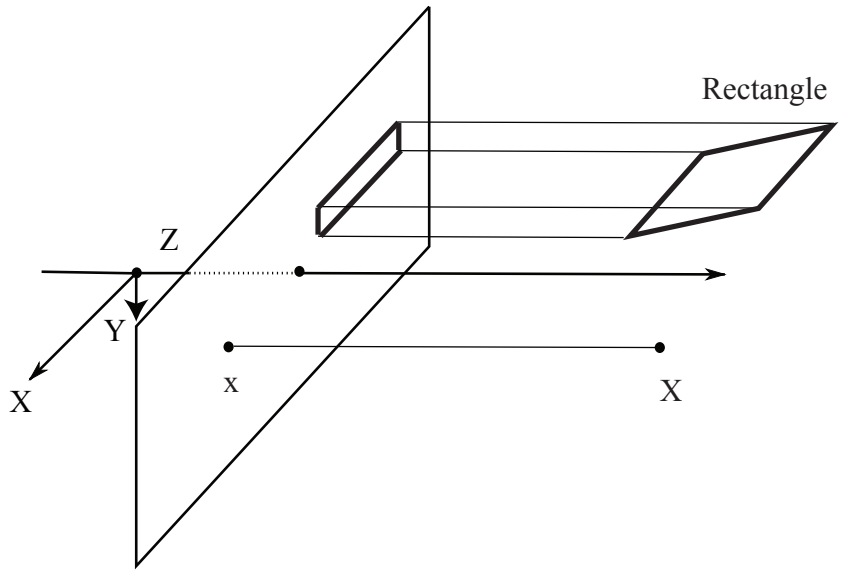


FIGURE 26.8: *Scaled orthographic projection with $s = 1$ is particularly easy to draw (and if s is not 1, one can then rescale the image as required). Here a point $\mathbf{X}$ in 3D projects to a point $\mathbf{x}$ in the image plane by "sliding" down a line perpendicular to the image plane – which is a plane of constant Z, so the line is parallel to the Z axis. In the top right corner, a rectangle is imaged as a foreshortened rectangle. The shortened edges of the rectangle are shortened because they are not parallel to the image plane (exercises).*

The model applies quite often. One important example is pictures of people. Very often, all body parts are roughly the same distance from the camera — think of a side view of a pedestrian seen from a motor car. Scaled orthographic projection applies in such cases. It is not always an appropriate model. For example, when a person is holding up a hand to block the camera's view, perspective effects can be significant (Figure 26.9).

26.1.3 Orthographic Projection

Occasionally, it is useful to rescale the camera (or assume that $f/Z = 1$), yielding $(X, Y, Z) \rightarrow (X, Y)$. This is known as *orthographic projection*.

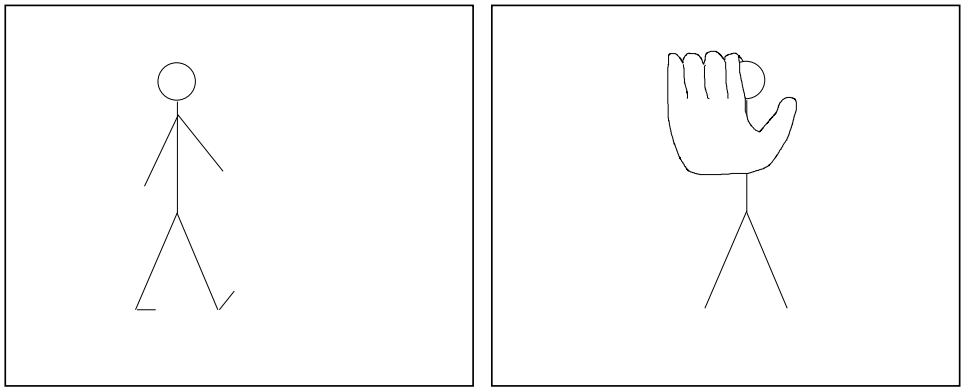


FIGURE 26.9: *The pedestrian on the* **left** *is viewed from some way away, so the distance to the pedestrian is much larger than the change in depth over the pedestrian. In this case, which is quite common for views of people, scaled orthography will apply. The celebrity on the* **right** *is holding a hand up to prevent the camera viewing their face; the hand is quite close to the camera, and the body is an armslength away. In this case, perspective effects are strong. The hand looks big because it is close, and the head looks small because it is far.*

Remember this: *Under perspective projection:*

- *points project to points;*
- *lines project to lines;*
- *more distant objects are smaller;*
- *lines that are parallel in 3D meet in the image;*
- *planes image as half-planes.*
- *planes have horizons;*
- *and shapes are foreshortened.*

Scaled orthographic projection applies when the distance to the points is much greater than their relief. Scaled orthographic projection maps

$$(X, Y, Z) \rightarrow s(X, Y)$$

where s is some scale. Shapes are foreshortened under scaled orthographic projection.
Orthographic projection maps

$$(X, Y, Z) \rightarrow (X, Y)$$

Shapes are foreshortened under orthographic projection.

26.2 LENSES

One practical version of a pinhole camera is a *camera obscura* – the box is built as a room, and you can stand in the room and see the view on the back wall (some examples are at `https://www.atlasobscura.com/lists/camera-obscura-places`; the internet yields amusing disputes about the correct plural form of the term). You can also build a simple pinhole camera with a matchbox, some tape, a pin, and some light sensitive film do the trick. Getting good images takes trouble, however.

A large hole in front of the camera will cause the image at the back to be brighter, but blurrier, because each point on the sensor will average light over all directions that happen to go through the hole. If the hole is smaller, the image will get sharper, but darker. A *lens* is a piece of refracting material shaped so as to focus light.

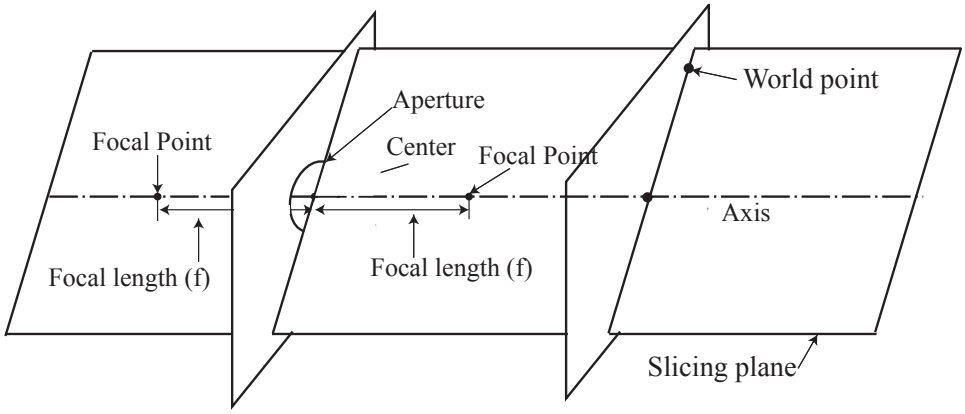


FIGURE 26.10: *The inset shows a 3D drawing of the abstract thin lens – a plane with a circular aperture and refractive material in the aperture. This is more usually drawn in 2D (as in Figure 26.11). The 2D drawing shows a section of the 3D geometry cut by a plane through the axis of the lens and the world point of interest.*

26.2.1 The Thin Lens

The simplest abstract model of a lens is a *thin lens*. A thin lens can be modelled as a plane with a circular hole in it. The center of this hole is the center of the lens. The line perpendicular to the plane and passing through the focal point is the axis of the lens. Light leaves some point in the world and arrives at the lens. The geometry of the lens determines where this light ends up. Slice the 3D geometry with a plane through the axis and the point of interest, as in Figure 26.10). The geometry of the lens is most easily understood using a drawing on this plane (Figure 26.11). Associated with any thin lens are (a) its *focal length* (traditionally written f: more detail below) and (b) its *aperture* (the circular hole). The focal length defines two *focal points* on the axis of the lens, one on either side of the lens and each one focal length away from the center of the lens.

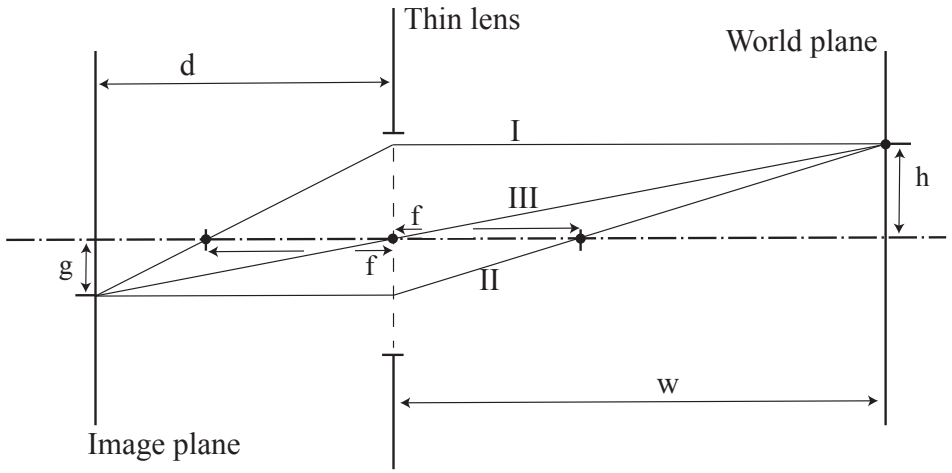


FIGURE 26.11: *A slice of the 3D drawing of Figure 26.10 by a plane through the axis and the world point: the dot-dashed line is the axis of the lens, and the dashed line shows the aperture. The world point defines a plane normal to the axis of the lens (and so parallel to the plane of the lens) shown here as a line. This world plane is w away from the center of the lens. The point that the world point images to can be constructed using the properties of the lens. Light path I shows a ray of light entering the lens perpendicular to the plane of the lens and emerging to pass through the focal point on the other side. Light path II shows a ray of light passing through the focal point on one side then striking the lens and emerging parallel to the axis of the lens on the other side. These two rays intersect in a point. This point – the image of the world point, or the image point – in turn defines a plane, normal to the axis of the lens and passing through the image point, which is d away from the center of the lens. Any point on a given world plane will image to some point on the corresponding image plane. Light path III is a ray leaving the world point and passing through the center of the lens without changing direction. It is an exercise to show that this ray will pass through the image point.*

Thin lenses have three properties that apply to rays of light that strike the lens aperture.

- Any ray of light entering the lens perpendicular to the plane of the lens will emerge to pass through the focal point on the other side.
- Any ray of light passing through the focal point on one side then striking the lens will emerge parallel to the axis of the lens on the other side.
- Any ray of light striking the center of the lens passes through lens without changing direction.

All rays of light that strike the plane of the lens outside the aperture are absorbed. Now place a plane perpendicular to the axis at w from the lens center. Call this the world plane. From these properties, it is straightforward to establish that three

specific rays of light leaving a point on this plane and entering the lens are focused to arrive at a corresponding point on a second plane on the other side (the image plane). Figure 26.11 shows the construction. It is an exercise to show that the image plane is d from the lens center, where

$$\frac{1}{w} + \frac{1}{d} = \frac{1}{f}.$$

It is a property of the lens that all rays leaving the world point and entering the lens will arrive at the image point. The image point of a lens will be brighter than the image point from a pinhole camera, because it collects power from more of the directions leaving the world point. The larger the aperture of the lens, the more power it collects from a world point (roughly, because it collects more rays).

For many cameras, the lens is at a fixed distance from the imaging device. Any point that lies on the world plane will be imaged as a point on the image plane, but points off the world plane will image as small circles on the image plane. These circles are sometimes known as *circles of confusion*. As Figure 26.12 shows, the radius of the circles of confusion grows as the point being imaged moves away from the world plane. If the circle of confusion is small enough (say, radius less than r_c), the point is in focus. There is a range of depths, known as the *depth of field*, where points are in focus because the radius of the circle of confusion is small enough.

Changing the aperture of the lens changes the depth of field, as Figure 35.1 shows. A larger aperture will collect more light, but have smaller depth of field. A smaller aperture will collect less light but have larger depth of field. As the aperture becomes arbitrarily small, the camera is more like a pinhole camera.

26.2.2 Lens Systems

Very few cameras have just one thin lens. In modern cameras, achieving an image that is both bright and focused is the job of the lens system. There may be one or several lenses that light passes through before reaching the sensor at the back of the camera. Thin lens concepts (focal length; aperture; circle of confusion; depth of field) apply to a lens system, though the details of the definitions require some adjusting. It is usually possible to adjust the aperture of the lens system, and it is often possible to change the focal length. Lenses tend to be heavy relative to cameras. The shape and position of the lenses, together with the refractive index of the materials they are built of, determine the path that light follows through the lens system. Generally, the lens system is designed to collect as much light as possible at the input and produce a focused image on the image plane. Remarkably, most lens systems result in an imaging geometry that can be modelled with a pinhole camera model, and lens system effects are ignored in all but quite specialized applications of computer vision.

Lens systems are designed and modelled using geometric optics, but lens designs always involve compromises. The result is that cameras with lenses differ from pinhole cameras in some ways that are worth knowing about, although they are not always important. Manufacturing difficulties and cost considerations mean that lenses can display various inconvenient properties. Lenses may cause *geometric distortions* of the image. The most noticeable effect of these distortions is that

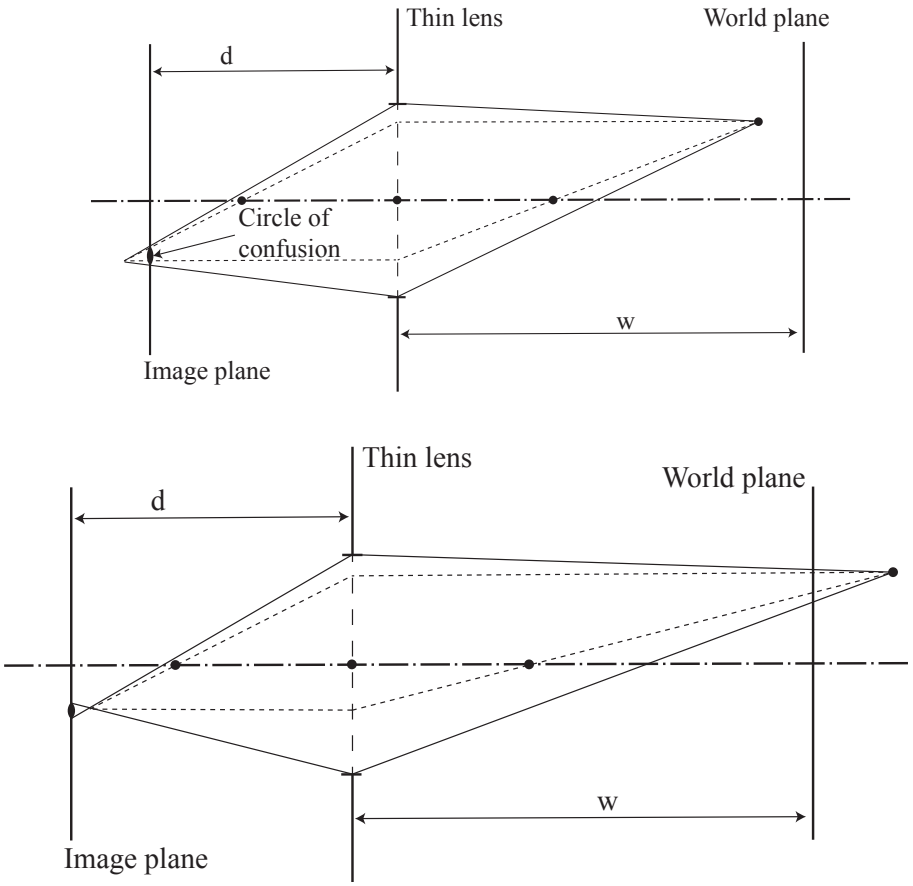


FIGURE 26.12: *In the* **top** *figure, the world point has moved towards the lens from the original world plane of Figure 26.11, and in the* **bottom** *figure it has moved away from the lens. In each case, the image point has moved away from the image plane. The dashed lines show the rays used to construct the location of the image point. The full lines indicate rays that just enter the aperture of the lens. The family of rays leaving the world point focuses on the image point, and so creates a small circle on the image plane. The radius of this circle will be increased by increasing the aperture.*

straight lines in the world may project to curves in the image. Most common is *barrel distortion*, where a square is imaged as a bulging barrel; *pincushion distortion*, where the square bulges in rather than out, can occur (Figure 26.14). Geometric distortions can be calibrated away (Section **??**). Lenses can display *chromatic aberrations*, where colors are less crisp at boundaries and objects have "halos" of color. Chromatic aberration occurs because light of different wavelengths takes slightly different paths through a refracting object. Various lens coatings can correct chro-

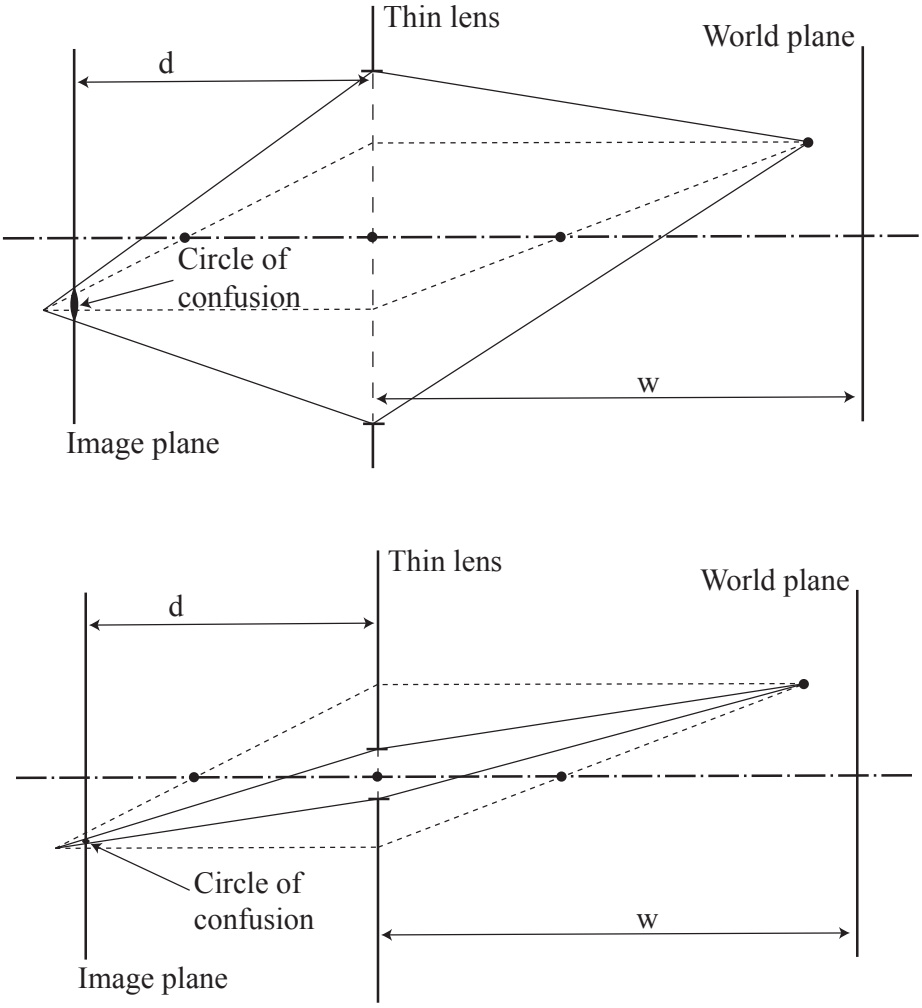


FIGURE 26.13: *This figure compares a thin lens with a large aperture (***top***) to one with a small aperture (***bottom***). The dashed lines show the rays used to construct the location of the image point. The full lines indicate rays that just enter the aperture of the lens. The family of rays leaving the world point focuses on the image point, and so creates a small circle on the image plane. Notice that for the same location of the world point, the circle of confusion is smaller for the lens with smaller aperture. This means that the depth of field of the lens will be larger, because the image point will be in focus (i.e. have a circle of confusion smaller than the critical radius) for a larger range of depths to the world.*

matic aberration, but the resulting lens system will be more expensive (I am not aware of methods that can calibrate chromatic aberrations for a given lens). In most lens systems, the periphery of the image tends to be brighter than it would

be in a pure pinhole camera. For more complex lens systems, an effect in the lens known as *vignetting* can darken the periphery somewhat.

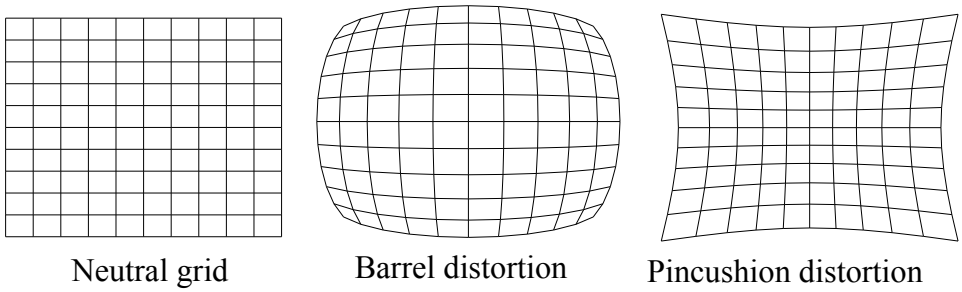


FIGURE 26.14: *On the* **left** *a neutral grid observed in a non-distorting lens (and viewed frontally to prevent any perspective distortion).* **Center** *shows the same grid, viewed in a lens that produces barrel distortion.* **Right**, *the same grid, now viewed in a lens that produces pincushion distortion.*

Remember this:

- *Lens systems ensure the image is both bright and focused. For our purposes, the concepts associated with a thin lens are sufficient to describe a lens system.*
- *Lens systems have a focal length which determines which world plane will be focused on the imaging device. Points off this world plane will be imaged as circles of confusion. A sufficiently small circle of confusion means the point is in focus. The depth of field of the lens is the volume in 3D of points that are in focus.*
- *Lens systems have an aperture. A larger aperture admits more light and results in a brighter image, but results in a smaller depth of field. A smaller aperture admits less light and results in a darker image, but results in a larger depth of field.*
- *A camera with a lens system can usually be modelled as a pinhole camera.*
- *Lenses can cause geometric distortions; these can be calibrated away.*
- *Lenses can cause color distortions, which are hard to remove.*

26.3 THE IMAGE SENSOR

Very bright sunlight is very much brighter than faint starlight (the factor is about 10^9). A complex interaction between mechanical systems in the iris of the eye, chemical systems in the photoreceptors, and neural systems in the retina, optic nerve and brain means that people can see in either illumination condition. However, for a single scene the *dynamic range* (ratio of brightest intensity to darkest intensity) is rather smaller than 10^9, but still big enough to present serious engineering challenges.

26.3.1 Film

Not that long ago, cameras mostly sensed light through changes to chemicals in a film at the back of the camera. This process is intrinsically non-linear. If a small amount of light arrives at a location, a small fraction of chemical will change. This means that if slightly more light arrives, there is still a lot of unchanged chemical to react. If a large amount of light arrives, a large fraction of the chemical will change, so that if more light arrives, there is little chemical left to react and the increase in response is smaller. However, it will take a great deal of light to cause all of the chemical to react. When an increase in the amount of light causes no increase in the response of the sensor, the sensor is said to be *saturated*. The non-linear property of film is convenient – film can measure light over a very wide dynamic range without saturating.

26.3.2 Electronic Sensors

The intensity of light travelling through a point in space in some direction is represented with a unit known as *radiance*. The intensity of light arriving at a point on a surface averaged over some range of directions is known as *irradiance*. Sensors average the irradiance over the area of a pixel to obtain incoming power P. This power is summed for some time period Δt – the *shutter interval* – to obtain the amount of energy the pixel receives. In turn, the energy determines the pixel intensity value reported by the imaging system. A property called *reciprocity* means that the response is a function of $P\Delta t$ alone. In particular, we will get the same outcome if we collect power P for time Δt or collect power P/k for time $k\Delta t$.

The sensors in modern cameras are solid-state electronic devices and so are linear. These sensors function over a significantly smaller dynamic range than those in the eye. The range depends on the camera, but sensors typically have a range between about 1000:1 (for simple consumer cameras) to 20, 000:1 (very high end sensors). Most cameras encode intensity with 8 or sometimes 10 bits. This range tends to be narrower than the range over which the sensor will function, and the sensor response is adjusted in software to compress the range of observed intensities.

26.3.3 The Camera Response Function

The *camera response function* or *CRF* determines what value is reported at each location. The response that the camera produces is obtained by applying the camera response function to $P\Delta t$. The fact that there is a camera response function is easily observed (Figure 26.15). Now write $\mathbf{X}$ for a point in space that projects to $\mathbf{x}$ in

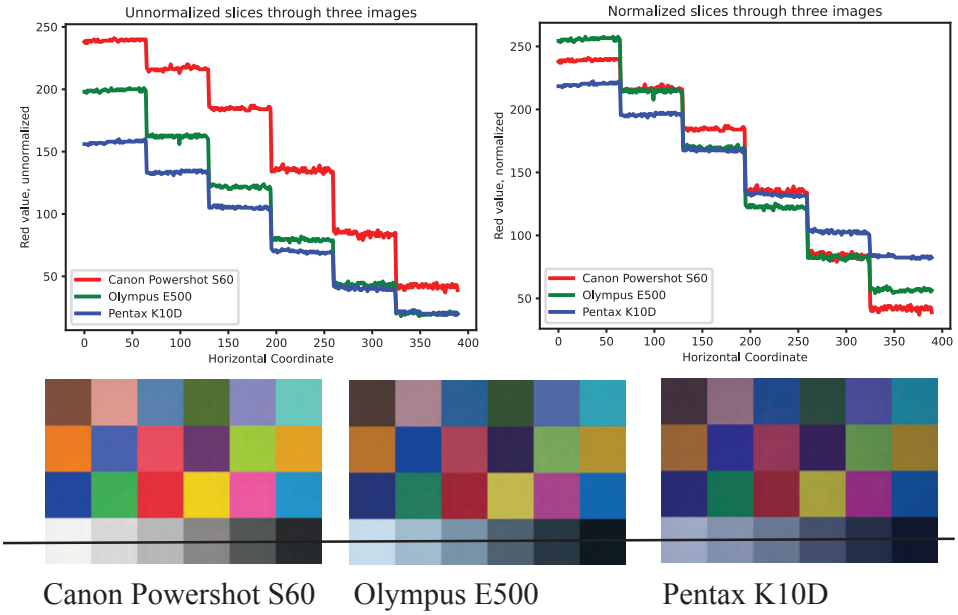


FIGURE 26.15: *It is easy to see the effects of camera response functions by looking at images. The* **bottom row** *shows images of the same scene, taken with different cameras. The* **top** *shows plots of the red intensity along the line.* **Top left** *is the red intensity – the cameras have given different outputs for the same scene. This might be the result of different camera gains or changes in the illumination, but the* **top right** *shows these cross sections scaled and translated to align with the Canon cross-section. They do not lie on one another, establishing that whatever happened isn't just a scaling of the intensity values, so is not caused by camera gain or by change in illumination. This figure was constructed from the dataset published at* `https://github.com/zyfccc/Representing-camera-response-function`, *described in the paper "Representing Camera Response Function by a Single Latent Variable and Fully Connected Neural Network", by Y. Zhao, S. Ferguson, H. Zhou and K. Rafferty in Signal, Image and Video Processing, 2022.*

the image, $P_{\mathbf{X}\to\mathbf{x}}$ for the power that leaves the patch and is collected at the pixel, $C(\cdot)$ for the camera response function, E for the energy collected at the sensor, and $I_{camera}(\mathbf{x})$ for the intensity reported by the camera at $\mathbf{x}$. Then our model is:

$$I_{camera}(\mathbf{x}) = C(E) = C(P_{\mathbf{X}\to\mathbf{x}}\Delta t).$$

Camera response functions must fit a larger range of light intensities into the relatively small dynamic range encoded by the digitized pixel values. Typical camera response functions do so by "stretching" the camera response to dark pixels and "squashing" the response at bright pixels (Figure 26.16). A rather rough argument about albedo (written ρ, details in Section 36.3) justifies the stretching/squashing strategy. One would like to be able to resolve some fixed change in albedo $\Delta\rho$ in

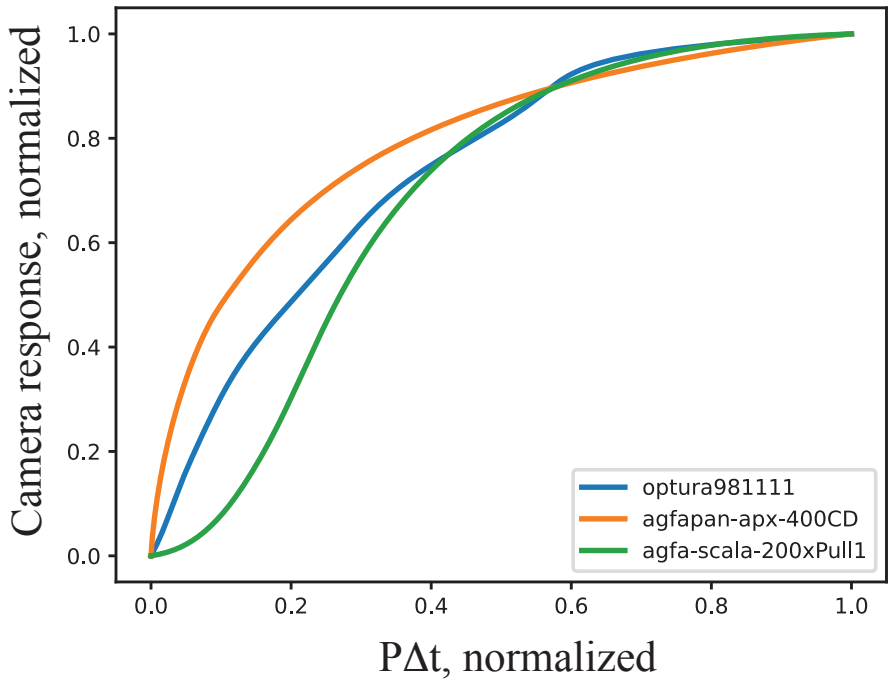


FIGURE 26.16: *Camera response functions for three different cameras, plotted from the comprehensive dataset available at* `https://cave.cs.columbia.edu/repository/DoRF`. *The horizontal axis is the "input" – the $P\Delta t$ observed by the camera, scaled to $0-1$. The vertical axis is the "output" – the response of the camera, again scaled to $0-1$. Small inputs are "stretched" and large inputs are "squashed".*

the image. When the input is small (i.e. dark), a small change will result in a big change of output. In turn, dark regions of the scene (where the shading is s, and quite small) produce dark image regions where one can resolve different scene brightnesses that are quite close. So one can resolve $s\rho$ and $s(\rho+\Delta\rho)$, which will be rather close because s is small. When the input is large (i.e. bright), a large change in input is needed to produce a small change in output. In bright regions of the scene (where the shading is S and quite large), quite a large change in brightness is required to produce a change in image intensity that can be resolved. In this case, $S\rho$ and $S(\rho+\Delta\rho)$ will be quite far apart, but we can still resolve them.

In principle, pixel values are hard to interpret when the camera response function is unknown. In practice, it is quite usual to assume that cameras are linear and proceed without comment. This practice is widespread because for many applications it is successful. For some applications, it is helpful to know the camera response function. For some cameras, one can simply look up the camera response function. Alternatively, camera response functions can be calibrated from images

using one of a variety of methods (Section 27.4.3). Camera response functions are always monotonic, and tend to have a quite simple form, so that estimation is quite reliable.

Many cameras are willing to report a RAW image, which is (typically) a direct report of what the sensor observed. If one has access to the camera, then RAW images are often very useful. For example, one does not need to know the CRF. Most "found" images – those observed on the internet, say, or the ones your relatives send you – are not RAW images.

26.3.4 High Dynamic Range Imaging

One reason to know a camera response function is *high dynamic range imaging* or *HDR imaging*. It is quite easy to find scenes where the dynamic range is so big that images in a reasonable camera lose information. Either the brightest points are saturated or the darkest points are very close to zero, or both. Color and relative intensity information is then lost. Now imagine we obtain two images of the scene, one with a neutral density filtthe camera and another However, if we have multiple images of the scene, obtained with different values of Δt, then we can recover information that would otherwise be lost. Using a small Δt will allow very bright locations to be measured accurately (though mid range locations will be dark, and dark locations will be lost). Similarly, using a large Δt will allow very dark locations to be measured accurately (though mid range locations will be bright, and bright locations will be lost). If the CRF is known, then for each location at each Δt_i we can compute the value of $E\Delta t_i$ and so recover E for each location with higher precision than any single image allows. Many modern cameras are willing to perform this computation with minimal provocation.

EXERCISES

26.1. Passing from left to right handed coordinate systems is straightforward.

(a) Show that the coordinate system obtained by reversing the direction of the z-axis in Figure 26.1, but otherwise leaving the figure unchanged, is a right handed coordinate system.

(b) Show that in this coordinate system, the camera images a 3D point $\mathbf{X}$ to a point $\mathbf{x}$ on the image plane by:

$$\mathbf{X} = \begin{pmatrix} X \\ Y \\ Z \end{pmatrix} \rightarrow \begin{pmatrix} fX/Z \\ fY/Z \\ -f \end{pmatrix} = \mathbf{x}.$$

26.2. Lines map to lines: A pinhole camera with focal point at the origin and image plane at $z = f$ views a line in 3D given by $\mathbf{u} + t\mathbf{w}$, where t is the parameter along the line. Write $\mathbf{w} = [w_1, w_2, w_3]^T$, etc. and use a left-handed coordinate system as in Figure **??**.

(a) Show that this line projects to the parametric curve in the image given by

$$\left(f\frac{u_1 + tw_1}{u_3 + tw_3}, f\frac{u_2 + tw_2}{u_3 + tw_3} \right).$$

(b) Now write

$$(x(t), y(t)) = \left(f\frac{u_1 + tw_1}{u_3 + tw_3}, f\frac{u_2 + tw_2}{u_3 + tw_3} \right).$$

Show that

$$[u_2w_3 - u_3w_2]\, x(t) - [u_1w_3 - u_3w_1]\, y(t) + [u_1w_2 - u_2w_1] = 0$$

and use this to argue that the curve is a line *except* when either $\mathbf{u} = (0,0,0)^T$ or $\mathbf{v} = (0,0,0)^T$.

(c) What happens when $\mathbf{u} = (0,0,0)^T$? and when $\mathbf{w} = (0,0,0)^T$?

(d) The signs and permutations of the subscripts in the previous subexercise should suggest a determinant is involved; it is. Show that

$$\mathbf{c} = \begin{pmatrix} u_2w_3 - u_3w_2 \\ -u_1w_3 + u_3w_1 \\ u_1w_2 - u_2w_1 \end{pmatrix}$$

has the properties: $\mathbf{c}^T\mathbf{u} = 0$; and $\mathbf{c}^T\mathbf{w} = 0$.

(e) Write

$$\mathcal{M} = \begin{bmatrix} u_1 & w_1 & c_1 \\ u_2 & w_2 & c_2 \\ u_3 & w_3 & c_3 \end{bmatrix}$$

and show that the determinant of $\mathcal{M}$ is 0.

26.3. Vanishing points: A pinhole camera with focal point at the origin and image plane at $z = f$ views two parallel lines $\mathbf{u} + t\mathbf{w}$ and $\mathbf{v} + t\mathbf{w}$. Write $\mathbf{w} = [w_1, w_2, w_3]^T$, etc. and use a left-handed coordinate system as in Figure **??**

(a) Show that the vanishing point of these lines, on the image plane, is given by $(f\frac{w_1}{w_3}, f\frac{w_2}{w_3})^T$.

26.4. Planes have horizons: A pinhole camera with focal point at the origin and image plane at $z = f$ views two parallel lines $\mathbf{u} + t\mathbf{w}$ and $\mathbf{v} + t\mathbf{w}$. Write $\mathbf{w} = [w_1, w_2, w_3]^T$, etc. and use a left-handed coordinate system as in Figure **??** Now we parametrize the direction of the lines by writing $\mathbf{w}(r, s) = r\mathbf{a} + s\mathbf{b}$, for (r, s) parameters. A pair of parallel lines in the direction $\mathbf{w}(r, s)$ is given by $\mathbf{u} + t\mathbf{w}(r, s)$ and $\mathbf{v} + t\mathbf{w}(r, s)$.

(a) Show that all the lines parametrized in this way lie on a single plane, shared by all lines.

(b) Now show that the vanishing point for the (r, s)'th pair of lines is

$$\begin{pmatrix} f\frac{ra_1+sb_1}{ra_3+sb_3} \\ f\frac{ra_2+sb_2}{ra_3+sb_3} \end{pmatrix}.$$

(c) The collection of vanishing points given by changing the values of r and s is a parametric curve in the image, even though there are two parameters. Show this by substituting $r' = kr$ and $s' = ks$ for some $k \neq 0$ and seeing what happens to the collection.

(d) Show that this parametric curve is a straight line in the image. Do this by using the results of the previous subexercise to show that the curve parameterized by (r, s) is the same as the curve parametrized by $(1, s/r)$ (and ignoring the case where $r = 0$ for the moment). Now use the results of the previous exercise to construct $\mathbf{c}$ such that $\mathbf{c}^T\mathbf{a} = \mathbf{c}^T\mathbf{b} = 0$. Now write

$$\begin{pmatrix} x(1, s/r) \\ y(1, s/r) \end{pmatrix} = \begin{pmatrix} f\frac{a_1+(s/r)b_1}{a_3+(s/r)b_3} \\ f\frac{a_2+(s/r)b_2}{a_3+(s/r)b_3} \end{pmatrix}$$

and use the results of the previous exercise to show that $c_1x(1, s/r) + c_2y(1, s/r) + c_3 = 0$.

(e) Use the results of the previous subexercise to argue that observing the horizon of a plane in a perspective camera yields the coefficients of the normal of that plane with respect to the camera's coordinate system.

26.5. On images of planes, closer to the horizon is further away: Write $\mathbf{x}$ and $\mathbf{x}'$ for two image points that are the image of two points $\mathbf{X}$ and $\mathbf{X}'$ lying on a world plane *that isn't parallel to the image plane.* Write $d_h(\mathbf{x})$ for the perpendicular distance from $\mathbf{x}$ to the horizon of that plane, etc. Write $d_f(\mathbf{X})$ for the distance from the focal point to $\mathbf{X}$. **Advice:** This problem is very easily worked with a drawing (start with Figure 26.7) but something of a nuisance if you write out the algebra and thrash.

(a) Show that

$$d_h(\mathbf{x}') > d_h(\mathbf{x}) \equiv d_f(\mathbf{X}') > d_f(\mathbf{X}).$$

(b) Show that

$$d_h(\mathbf{x}') = d_h(\mathbf{x}) \equiv d_f(\mathbf{X}') = d_f(\mathbf{X}).$$

(c) Now if

$$d_h(\mathbf{x}') < d_h(\mathbf{x})$$

what happens? (**Hint:** swap the names of the points).

(d) What happens if the world plane is parallel to the image plane?

26.6. (a) Do parallel lines have a vanishing point in a scaled orthographic camera? Why?

(b) Does a plane have a horizon in a scaled orthographic camera? Why?

26.7. Foreshortening: A pinhole camera with focal point at the origin and image plane at $z = f$ views a circle in the world (and lying on a world plane, as circles do). **Advice:** This problem is very easily worked with a drawing, harder if you write out the algebra and thrash, and very hard if you don't know (or look up) anything about conic sections.

(a) Construct the family of lines joining the points on the circle with the focal point. Show that this is a right circular cone whose axis is the line joining the center of the circle to the focal point.

(b) Show that the image of this circle will be a circle if the world plane is parallel to the image plane.

(c) Show that the image of this circle will be a parabola if the axis of the cone is parallel to the image plane.

(d) Show that the image of this circle could be an ellipse or one branch of a hyperbola.

26.8. Foreshortening: A pinhole camera with focal point at the origin and image plane at $z = 10$ views a rectangle in the world. The vertices of the rectangle are

$$\mathbf{v}_1 = \begin{bmatrix} 0 \\ 1 \\ 15 \end{bmatrix}, \mathbf{v}_2 = \begin{bmatrix} 1 \\ 1 \\ 15 \end{bmatrix}, \mathbf{v}_3 = \begin{bmatrix} 1 \\ 2 \\ 20 \end{bmatrix} \text{ and } \mathbf{v}_4 = \begin{bmatrix} 0 \\ 2 \\ 20 \end{bmatrix}$$

What is the image of the rectangle?

26.9. Foreshortening in scaled orthographic cameras: A scaled orthographic camera with image plane at $z = 0$ views a circle in the world (and lying on a world plane, as circles do). **Advice:** This problem is easier if you write out the algebra and thrash, as long as you use a sensible coordinate system. A drawing might help get that coordinate system.

(a) Construct the family of lines perpendicular to the image plane and passing through points on the circle. Show that this is a cylinder, but it's cross section may not be a circle.

(b) Show that the cross-section of this cylinder is an ellipse.

(c) What difference does the choice of scale make?

(d) Does the choice of Z-value where the image plane is located make any difference at all (i.e. what happens when the image plane is $z = c$, for c some constant different from zero)?

26.10. Foreshortening in scaled orthographic cameras: An orthographic camera camera with image plane at $z = 10$ views a rectangle in the world. The vertices of the rectangle are

$$\mathbf{v}_1 = \begin{bmatrix} 0 \\ 1 \\ 15 \end{bmatrix}, \mathbf{v}_2 = \begin{bmatrix} 1 \\ 1 \\ 15 \end{bmatrix}, \mathbf{v}_3 = \begin{bmatrix} 1 \\ 2 \\ 20 \end{bmatrix} \text{ and } \mathbf{v}_4 = \begin{bmatrix} 0 \\ 2 \\ 20 \end{bmatrix}$$

(a) What is the image of the rectangle?

(b) What shows that the rectangle has been foreshortened?

26.11. Foreshortening in orthographic cameras: An orthographic camera camera with image plane at $z = 10$ views a line segment in the world. The line segment is given by $\mathbf{u} + t\mathbf{w}$, where $0 \leq t \leq L$, the length of the line segment is L, and

$$\mathbf{w} = \begin{bmatrix} \cos\theta\cos\phi \\ \sin\theta\cos\phi \\ \sin\phi \end{bmatrix}$$

(which is a unit vector).

(a) What is the length of the projected line segment? **Advice:** This is much easier if you remember that the length of a line segment does not change if you translate it.

(b) A line segment of known length L is viewed in a scaled orthographic camera. Show that it is possible to recover most of the configuration of the 3D line segment from the image, but there is a two-fold ambiguity and the distance to the line segment is unknown.

26.12. Refer to Figure 26.11 for thin lens notation. In this figure, d, w, f, h and g are distances, and so have positive sign.

(a) Use ray I and similar triangles to show that

$$\frac{h}{f} = \frac{g+h}{d}$$

(b) Use ray II and similar triangles to show that

$$\frac{g}{f} = \frac{g+h}{w}$$

(c) Use ray III and similar triangles to show that

$$\frac{g}{d} = \frac{h}{w}$$

(d) Now use ray I and ray II to show that

$$\frac{1}{d} + \frac{1}{w} = \frac{1}{f}$$

(e) Three rays do not, in general, intersect at one point. Show that rays I, II and III leaving the point (w, h) intersect at the point $(-d, -g)$. Do this by showing that the intersection point for ray I and ray III is

$$\left(-\frac{fw}{w-f}, -\frac{hf}{w-f}\right)$$

and the intersection point for ray II and ray III is the same point.

26.13. All points on the projective plane with homogeneous coordinates $(U, V, 0)$ lie "at infinity" (divide by zero). As we have seen, these points form a projective line.

(a) Show this line is represented by the vector of coefficients $(0, 0, C)$.

(b) A homography $\mathcal{M} = \left[\mathbf{m}_1^T; \mathbf{m}_2^T; \mathbf{m}_3^T\right]$ is applied to the projective plane. Show that the line whose coefficients are $\mathbf{v}_3$ maps to the line at infinity.

(c) Now write the homography as $\mathcal{M} = [\mathbf{m}'_1, \mathbf{m}'_2, \mathbf{m}'_3]$ (so $\mathbf{m}'$ are columns). Show that the homography maps the points at infinity to a line given in parametric form as $s\mathbf{m}'_1 + t\mathbf{m}'_2$. Now write $\mathbf{n}$ for a non-zero vector such that $\mathbf{n}^T\mathbf{m}'_1 = \mathbf{n}^T\mathbf{m}'_2 = 0$. Show that $\mathbf{n}$ is the vector of coefficients for the image of the line at infinity in the homography. Is $\mathbf{n}$ unique?