PBRT: Camera Models

Sunday, December 13, 2020 1:48 PM

6 Camera models

A pinhole camera captures everything in sharp focus. The lenses of real-life cameras can't do that:

- The distribution of radiance leaving a lens system is different from the distribution entering it.
- Lenses introduce various aberrations:
    - ***Vignetting***: a darkening toward the edges of the image that occurs because less light arrives at the film/sensor's edges than it does at the center.
    - ***Pincushion or barrel distortion***: a curving of straight lines.

Virtual cameras simulate the radiometry of image formation.

Cameras produce world space rays from their position and in the direction of a sample on the film plane. Different camera models produces rays differently and the images that they produce are thereby different.

6.1 Camera model

The abstract ***base class Camera*** stores the ***transformation*** that transforms its position and direction from camera space to world space.

It also stores the ***shutter speed*** in 2 parameters: the time t_O at which it opens and the time t_C at which it closes. Rays are generated between these 2 times. By sampling during an interval as supposed to at one instant, ***motion blur*** is possible.

The final image captured by the camera is represented by an object of the Film class, which Camera stores.

And finally, the camera may be immersed in some scattering Medium. (Recall that a ray's origin is [associated with a medium].)

The film plane has a number of ***camera samples***. The ***CameraSample struct*** is used in ray generation and stores the point at which a ray intersects the lens and the point at which it intersects the film, along with the time (interpolated across the shutter's opening interval) at which the ray is sampling the scene.

6.1.1 Camera coordinate spaces

The camera's location is the origin of ***camera space***. The viewing direction is the z axis. And the up vector is the y axis. These 2 define the orientation of the camera.

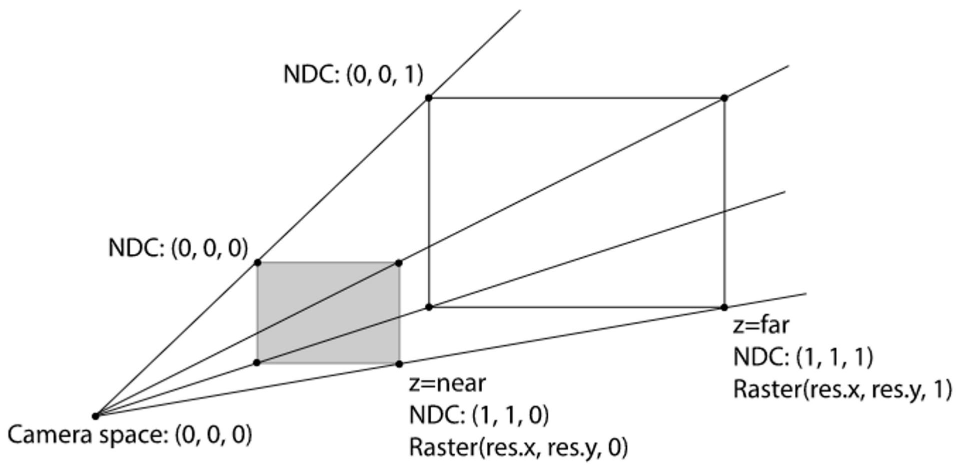
6.2 Projective camera models

OrthographicCamera and PerspectiveCamera are 2 different implementations of base abstract class ProjectiveCamera. A ***projection camera*** uses a $4x4$ ***projection matrix***.

Screen space is defined on the ***film plane***, where a projection camera projects objects that are in camera space. Screen space is a 3D space: the z coordinate, ranging from 0.0 to 1.0, places objects somewhere between the ***near clipping plane*** $(x, y, 0.0)$ and the ***far clipping plane*** $(x, y, 1.0)$. The ***film plane and the near clipping plane are the same***. Objects are projected ***onto the film/near plane***.

Normalized device coordinate (NDC) space is also a 3D space into which screen space coordinates are linearly mapped. The coordinate ranges are $[0.0,1.0]$ for $x, y, and\ z$ (although the z coordinate is exactly the same as in screen space and requires no transformation).

Raster space maps NDC coordinates to the resolution of the image: $[(0.0,1.0), (0.0,1.0)] \rightarrow [(0.0, resolution.x), (0.0, resolution.y)]$.



Something important to note is that the y axis gets ***inverted*** in NDC/raster space and that the ***origin*** there lies on the ***top left corner***.

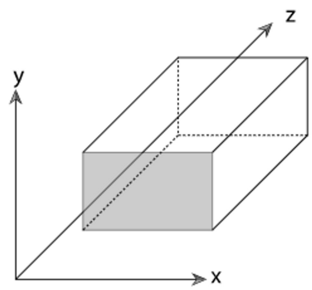
An isomorphism exists between each ***consecutive*** pair of these spaces: between camera and screen and between screen and NDC/raster:

- $(Camera \rightarrow Screen)^{-1} = (Screen \rightarrow Camera)$
- $(Screen \rightarrow Raster)^{-1} = (Raster \rightarrow Screen)$

The $(Raster \rightarrow Camera)$ transformation is the composition $(Screen \rightarrow Camera)((Raster \rightarrow Screen)(vector))$ that transforms from raster to camera space. In matrix notation: $[Screen \rightarrow Camera][Raster \rightarrow Screen](vector)$.

6.2.1 Orthographic camera

Orthographic projection transformation. This transformation takes a rectangular region of the scene and projects it onto the front face of the ***box*** that defines the region. This box is axis-aligned in camera space and is the ***view volume*** of the orthographic camera.



Objects inside the orthographic view volume (box) are projected onto the $z = near$ face of the box.

No ***foreshortening*** occurs: objects do not become smaller as they get farther away (like they do in a perspective projection).

Parallel lines and relative distances between objects are ***preserved***.

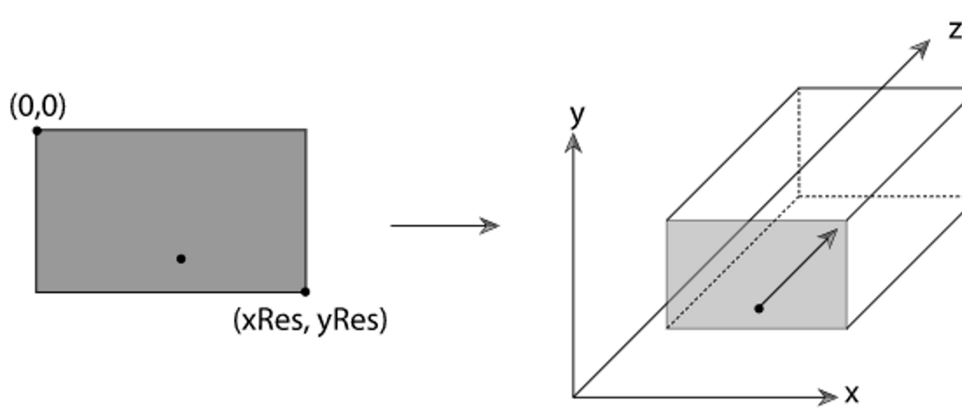
The $[Camera \rightarrow Screen]$ ***orthographic projection matrix*** is a composition of matrices that:

1. Translates camera space points only along the z axis and so that the origin of camera space lies at the z coordinate of the near plane (***x and y coordinates remain unchanged***).
2. Normalizes the z coordinate of the points contained in the orthographic view box, between the near plane $(x, y, 0)$ and the far plane $(x, y, 1)$. (Normalization is done by scaling by the reciprocal of the depth of the orthographic view box, that is, the reciprocal of the distance between the near and far planes in camera space.)

IMPORTANT: PBRT places the near plane at $z = 0$ in camera space.

An orthographic camera generates ray differentials that are parallel to the main ray (their directions are the same); what it changes is their origins, one ray differential is shifted 1 pixel in the x direction and the other is shifted 1 pixel in the y direction. The $(Raster \rightarrow Camera)$ transformation of the ProjectiveCamera is used to compute the camera space distance that corresponds to a 1-pixel shift in raster space.

The $(Raster \rightarrow Camera)$ transformation is also used to compute camera space ***origins for rays*** (all (all of which lie on the near plane). Unlike the origins of PerspectiveCamera rays, which are all the same and lie in front $(z = 0)$ of the near plane $(z = 1)$, OrthographicCamera ray origins are all different and lie on the near plane $(z = 0)$. Points in raster space are easy to obtain: they are provided by the film. The ***direction of rays*** is always $(0,0,1)$.



6.2.2 Perspective camera

A perspective camera creates the effect of ***foreshortening***: objects that are far away are projected to be smaller than objects of the same size that are closer.

Distances, angles, and parallel lines are ***not preserved***.

The human eye perceives the world in perspective.

A perspective camera places the near plane at $z = 1$ and the camera at $z = 0$ (unlike the orthographic camera, which places it at $z = 0$ along with the camera's position).

The $[Camera \to Screen]$ ***perspective projection matrix*** is a ***canonical matrix*** scaled by a factor that depends on the FOV angle. The canonical matrix is:

$P(x, y, z, 1)$

$$= \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & \dfrac{\mathbf{z}_{far}}{(\mathbf{z}_{far} - \mathbf{z}_{near})} & -\dfrac{\mathbf{z}_{far}\,\mathbf{z}_{near}}{(\mathbf{z}_{far} - \mathbf{z}_{near})} \\ 0 & 0 & 1 & \mathbf{0} \end{bmatrix} \begin{bmatrix} x \\ y \\ z \\ 1 \end{bmatrix}$$

$$= \begin{bmatrix} 1x + 0y + 0z + 0 \\ 0x + 1y + 0z + 0 \\ 0x + 0y + z\dfrac{\mathbf{z}_{far}}{(\mathbf{z}_{far} - \mathbf{z}_{near})} - \dfrac{\mathbf{z}_{far}\,\mathbf{z}_{near}}{(\mathbf{z}_{far} - \mathbf{z}_{near})} \\ 0x + 0y + 0z + z \end{bmatrix} = \begin{bmatrix} x \\ y \\ z_{far}\dfrac{z - z_{near}}{z_{far} - z_{near}} \\ z \end{bmatrix}$$

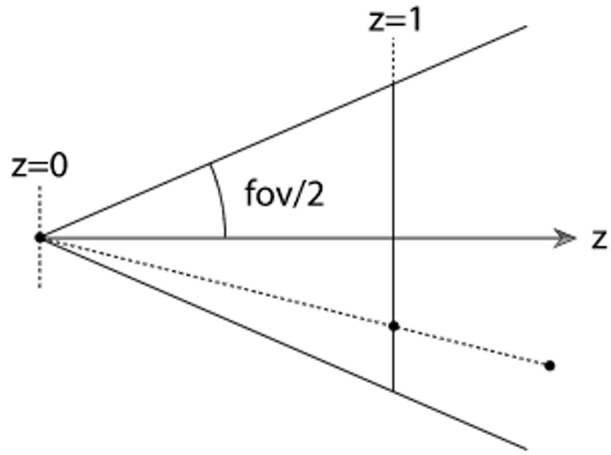
where the mapped vector is the ***homogeneous-coordinate*** vector of a camera space vector. The resulting screen space vector is thus:

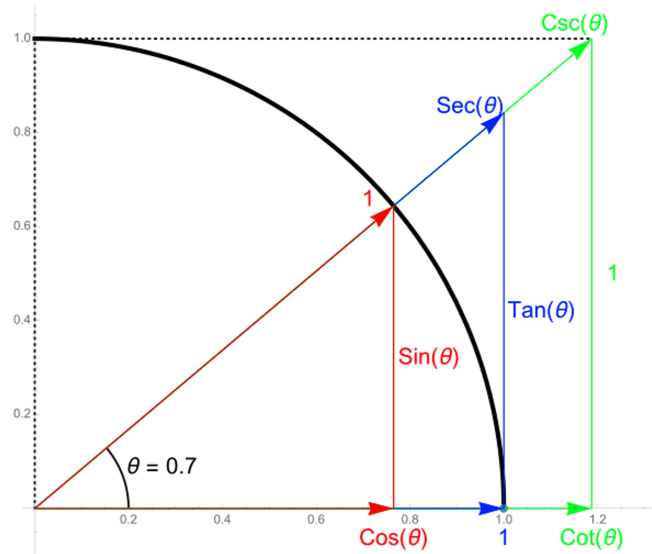
$$\begin{bmatrix} \dfrac{x}{z} \\ \dfrac{y}{z} \\ \dfrac{z_{far}\,(z - z_{near})}{z\,(z_{far} - z_{near})} \end{bmatrix}$$

The ***division*** of the x and y coordinates ***by the z*** coordinate achieves the ***foreshortening*** effect: the farther the point is in camera space, the larger its z coordinate is and the smaller the quotient will be, drawing the point closer to the origin in screen space.

As for the mapping of the z coordinate, the product of z and $\frac{(z - z_{near})}{(z_{far} - z_{near})}$ normalizes its value (i.e. maps it to the interval [0,1], where z_{near} is mapped to 0 and z_{far} is mapped to 1). I don't know where the $\frac{z_{far}}{z}$ factor comes from.

The canonical perspective matrix is then scaled by a factor that depends on the ***FOV angle***. PerspectiveCamera uses a single angle for both the vertical and horizontal FOVs.





TODO: I don't know how to explain that FOV maps camera space vector x, y coordinates to $[-1,1]$.

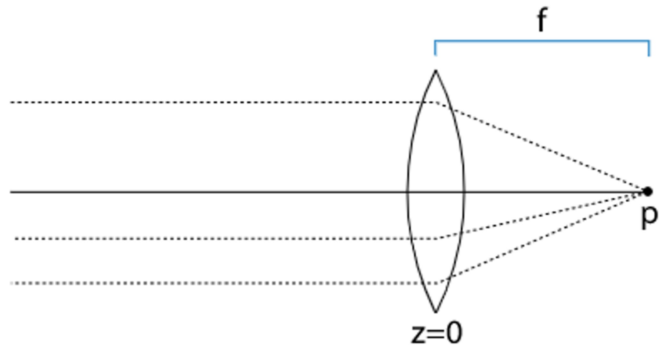
The origin of PerspectiveCamera rays and ray differentials is always the camera space origin $(0,0,0)$. The x and y differentials apply to the tip of the ray, not to its tail: the origin of the ray differentials is the origin of the main ray.

6.2.3 The thin lens model and depth of field

The ***focal plane*** is the only z plane between the near and far planes that the camera can focus. Objects in other z depths appear out of focus. The farther from the focal plane (in both directions, ***background*** and ***foreground***), the blurrier the objects are captured.

Real cameras have a ***system of lenses*** and lenses have a ***thickness*** that isn't negligible. The ***thin lens model*** is a simplification of these systems: one that only has 1 lens and its thickness doesn't have much of an effect.

Rays that enter the camera in a direction parallel to the ***optical axis*** are redirected by the lens and focused to a single point behind the lens, the ***focal point*** p. The distance f between the lens and the focal point is the ***focal length***. This is an ***immutable property of the lens*** and it can't be adjusted.

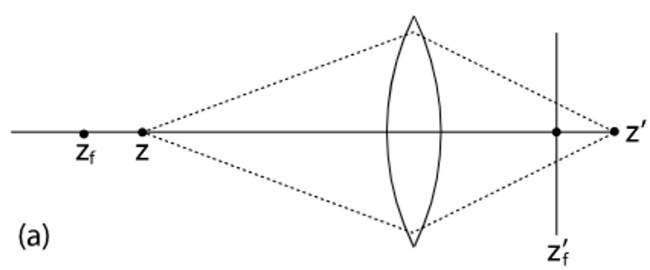


The ***location of the film*** with respect to the focal point determines the plane in the scene that will appear in focus on the image.

Let z be the distance from the lens to a point in the scene (called ***focal distance***) and z' be the distance behind the lens at which that point is in focus.

If the film is placed exactly at the lens' focal length, then a point that is infinitely far away will be in focus. That is, the focal plane will be $z = \infty$.

If the film is placed at a point z'_f between the lens and the lens' focal point, then a point at a focal distance $z_f < \infty$ will be in focus. That is, the focal plane will be z for some $z < \infty$.



The ***Gaussian lens equation*** relates z and z':

$$\frac{1}{z'} - \frac{1}{z} = \frac{1}{f}$$

where f is the lens' focal length. Note that as $z \to \infty$, $z' \to f$ (a point at infinity is in focus at the lens' focal length).

Given a depth z, we can make it the focal distance by solving for z' and placing the film at that distance from the lens:

$$z' = \frac{fz}{f+z}$$

A point in focus will be mapped to a point on the film. A point ***out of focus*** will be ***mapped to a disk*** on the film. The boundary of this disk is called the ***circle of confusion***. The size of the circle of confusion is ***determined by*** the diameter of the aperture, the focal distance, and the distance between the point in the scene and the lens.

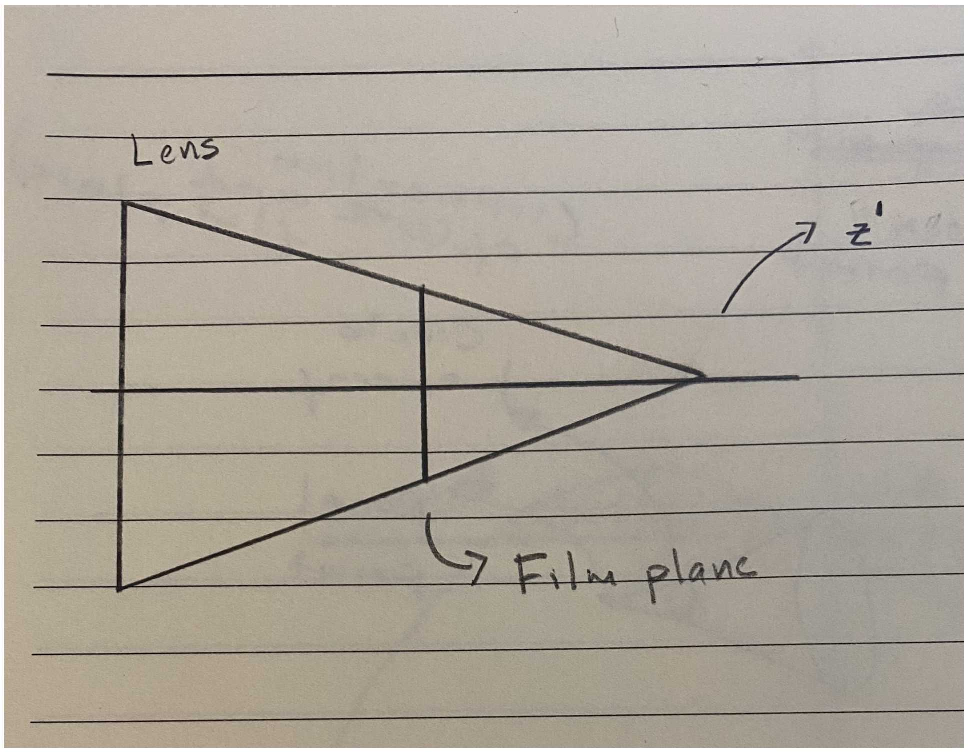
The circle of confusion is the intersection of the cone and the film plane. It's diameter is computed as follows:

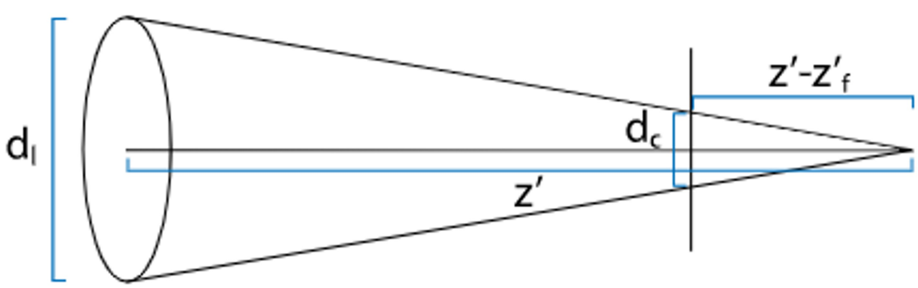
$$d_c = \left|\frac{d_l(z' - z'_f)}{z'}\right|$$

where d_l is the diameter of the lens, z' is the distance behind the lens at which depth z in the scene is in focus, and z'_f is the film.

Which comes from the following ***similar triangles relation***:

$$\frac{d_l}{z\prime} = \frac{d_c}{|z' - z\prime_f|}$$





Since z' is not known directly, its value must be computed using the Gaussian lens equation:

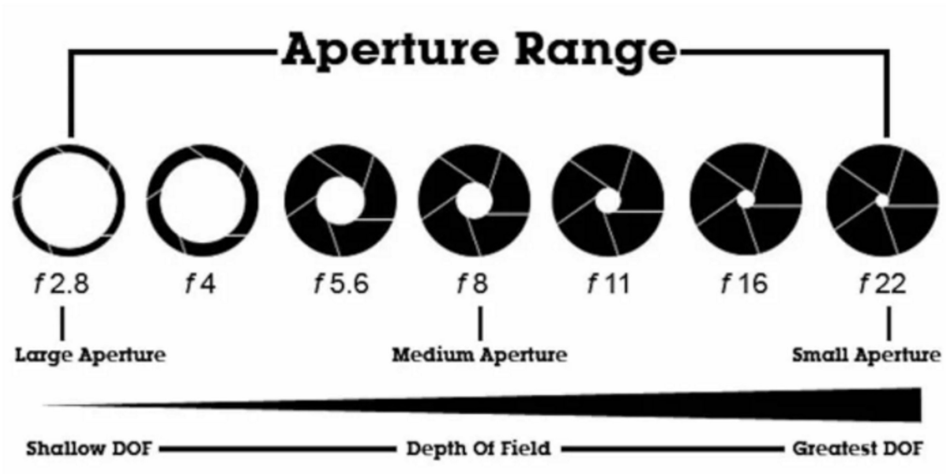
$$z' = \frac{fz}{f + z}$$

Substituting:

$$d_c = \left|\frac{d_l\, f\, (z - z_f)}{z\, (f + z_f)}\right|$$

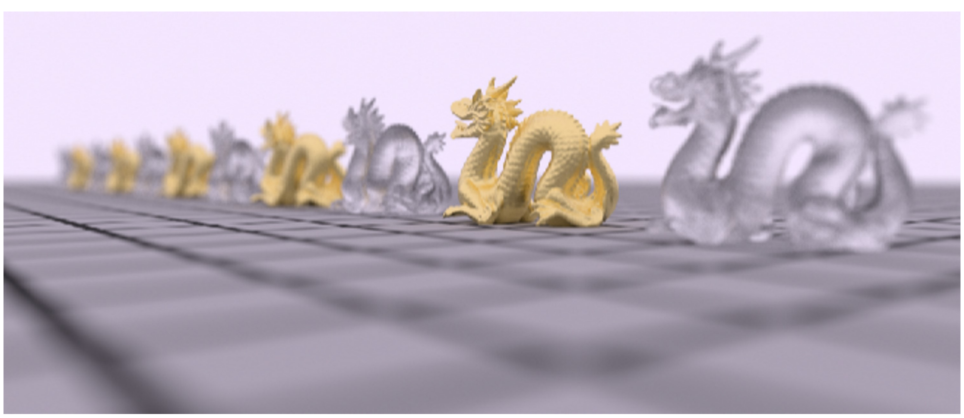
where f is the lens' focal length. Note that the diameter of the circle of confusion for a given depth z is 0 when z' equals the distance from the lens to the film.

Since the diameter of the circle of confusion is related to the lens' diameter by a similar triangles relation, a change in the lens' diameter causes a proportional change in the circle of confusion. The diameter of the lens can't change, but the ***aperture*** simulates just such a change. This is what ***aperture*** is:

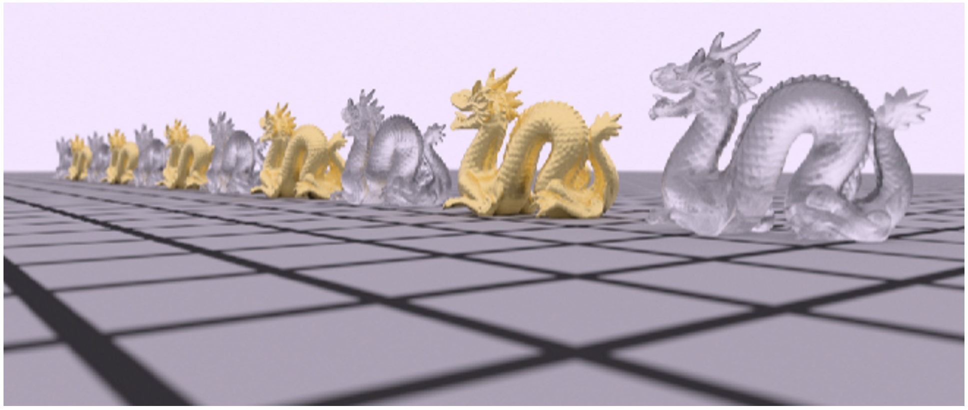


As the size of the aperture increases, blurriness increases the farther a point is from the focal plane. Which makes sense because the larger the aperture is, the larger the circle of confusion is. This is also the reason why every point is in focus in a ***pinhole camera*** no matter what its depth is: the aperture is infinitesimally small.

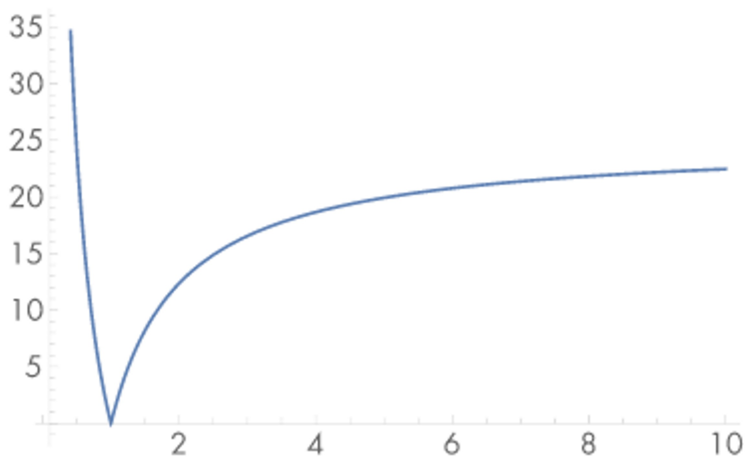
Large aperture: focal plane is at the 2nd dragon



Small aperture: focal plane is at the 2nd dragon



This graph of d_c as a function of depth z for $z_f = 1\,m$, $f = 50\,mm$ and $25\,mm$ aperture shows how quickly the circle of confusion grows for objects in the foreground (depths $z < z_f$) the closer they are to the camera. Growth is much slower for objects in background the farther away they are.



The range of distances from the lens at which objects appear in focus (which includes the focal distance, of course) is called the lens' ***depth of field***. Outside this range, objects appear out of focus.