Lecture Notes, Week 1: Light, Cameras, Projection

1 Light and Reflection

Visible light is really an electromagnetic radiation within a fairly narrow spectrum of wavelengths between 400nm and 700nm. Typically we do not see a pure color at a particular wavelength λ, but rather a spectrum or distribution over the range of visible wavelengths. Specifically, a ray of light may consist of many different wavelengths, which we perceive collectively as color (e.g., sunlight). Another, perhaps more intuitive way to think about light is as particles, *photons*, each of which has a specific wavelength λ. These particles travel in straight lines within a given medium (e.g., air).

The process of forming the image involves the camera, or an eye, capturing the amount of light reflected from objects in the world. The image being formed this way depends on many factors, including lighting, scene geometry, surface properties, camera optics and viewpoint. One of the key components of this process is surface reflection, which is governed by the Bidirectional Reflection Distribution Function: $BRDF(\theta_i, \phi_i, \theta_o, \phi_o)$. This function for a given incident and viewing angle tells us the fraction of light that will be reflected; (θ_i, ϕ_i) are incident angle of the light on the surface and (θ_o, ϕ_o) are viewing angle formed between the surface and the camera (both encoded in polar coordinates).

BRDF will depend on the properties of the surface. For a Lumbertian (mate) surface this function has a particularly simple form:

$$BRDF(\theta_i, \phi_i, \theta_o, \phi_o) = \frac{\rho}{\pi} \tag{1}$$

where constant ρ is called *albedo* and π is a normalization constant that ensures that reflected light (collectively) is less or equal to amount of incident light on the surface. Albedo is a measure, between 0 and 1, of how much light that hits a surface is reflected without being absorbed. The actual light being reflected is a product of not only BRDF, but also incident light intensity, and the dot product between incident light direction and the surface normal. This means that Lambertian surface appears equally bright in ALL directions and this brightness is a function of how close the incident light direction is to the normal of the surface. Overall, for Lambertian surface reflected light can be characterized by the following equation:

$$L = \frac{\rho}{\pi} I (\vec{i} \cdot \vec{n}) \tag{2}$$

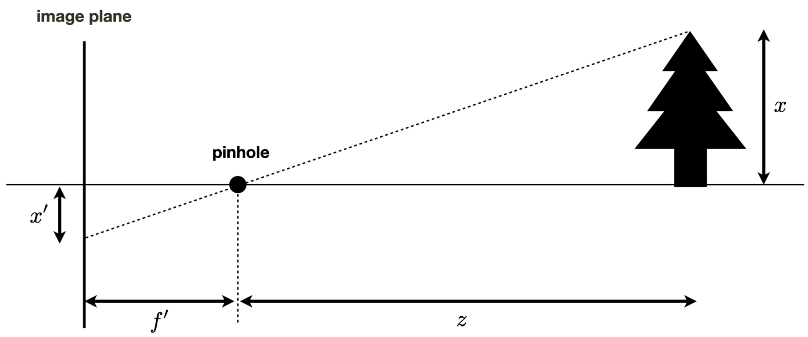
where L is the amount of light reflected, I (variable, not a function) is the amount of incident light, $\frac{\rho}{\pi}$ is the term coming from BRDF, and $\vec{i} \cdot \vec{n}$ is effectively similarity between the incident direction and the normal of the surface (coming from Lambert's cosine law).

Not all surfaces are Lumbertian in nature. For a mirror surface all of the incident light is reflected in one direction which is reciprocal of the incident direction over the normal. Unless the camera is coincident with this direction, it will not see light reflected from a mirror surface. Shiny surfaces will generally reflect light in a narrow set of directions around the the "mirror" direction with amount of light that falls off as the direction differs from the mirror direction. Most real world surfaces are a combination of Lambertian and mirror/shiny components.

2 Pinhole Camera and Perspective Projection

Pinhole camera is constructed by placing an imaging sensor in a rectangular box and making an infinitesimally small hole in the front of it. The hole is assumed to be so small that only a single photon, or light ray, will be able to pass through it. (In practice making a hole this small is impossible.) Despite simplicity of construction, pinhole camera provides an acceptable approximation to the imaging process in most cameras and our eye.

Perspective projection, which describes projection in pinhole camera, creates inverted images. The perspective projection is characterized by the following equations derived using similar triangles with respect to the illustration below, where f' is *focal length*, (x, y, z) is the point on the object being imaged and (x', y') is the corresponding point on the imaging plane. Note that coordinate system is assumed to be centered at the pinhole with $x - y$ plane being parallel to the sensor / imaging plane and z being a depth away from be pinhole.



Perspective Projection

$$
\begin{aligned}
\frac{x'}{f'} &= \frac{x}{z} \\
x' &= f'\frac{x}{z} \\
y' &= f'\frac{y}{z}
\end{aligned}
$$

Perspective projection can be characterized by the following properties:

- Each set of parallel lines in the world will meet at a different point when projected in the imaging plane. This point is called a *vanishing point.*
- Set of parallel lines on the same plane in the world lead to collinear vanishing points in the imaging plane. This line is called a *horizon line* for that plane.
- Points in the world project to points in the imaging plane.
- Lines in the world project to lines in the imaging plane.
- Planes in the world project to the whole or half planes in the imaging plane.
- Angles in the world are not preserved in the imaging plane.

3 Weak Perspective and Orthographic Projection

Pinhole perspective is an approximation to the geometry of the imaging process. Sometimes it is useful to use even simpler (coarser) approximations. Weak Perspective and Orthographic Projection governed by equations below are such coarser approximations. Weak perspective is accurate for flat scenes or scenes where relief of an object is relatively small with respect to the distance of the object from the camera. In such cases a reasonable approximation is obtained using weak perspective projection by setting $m = \frac{f'}{z_{avg}}$, where z_{avg} is the average distance of points on the object to the camera.

Weak Perspective Projection

$$
\begin{aligned}
x' &= mx \\
y' &= my
\end{aligned}
$$

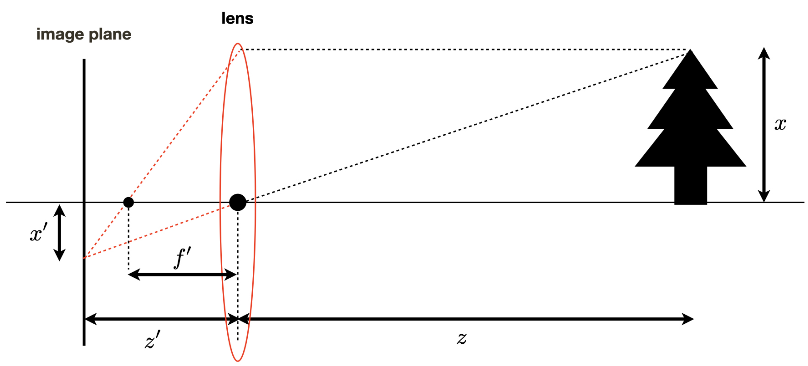
Orthographic Projection

$$
\begin{aligned}
x' &= x \\
y' &= y
\end{aligned}
$$

Note, that in orthographic projection, the image is not inverted.

4 Pinhole Camera and Perspective Projection with a Lens

Traditional pinhole camera is limited in a number of ways. The most significant ones are that images obtained using it are either dark or require a long time to capture (making it difficult to capture dynamic scenes) as is the case with small pinhole; or, alternatively, are blurry if captured with a larger pinhole. We can alleviate these issues by increasing the size of the pinhole and inserting *lens* to focus the incoming light that would pass through it. The role of a lens is to capture more light while preserving, as much as possible, the abstraction of an ideal pinhole camera. The lens will ensure that the light incident on the lens converges at a point. By making an assumption that we are using a *thin lens* – a lens is considered thin if its thickness is much less than radii of curvature of its surfaces. Under these conditions we can assume that light bends only once at the center of the lens and effects of light traveling within the lens can be ignored. For a thin lens the imaging can be characterized by the equation below:



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

or

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

(depending on convention)

where f' is again a focal length, now dictated by the properties of the lens, z' is the distance from the lens to the imaging plane and (x, y, z) is location of the point on the object in the world being imaged, with (x', y') being the corresponding projection on the imaging plane where that point will be in focus (see derivation in the slides). Note that thin lens equation dictates that for camera with a fixed lens (i.e., fixed focal length f') only objects at particular depth from the camera z will be in focus on the imaging plane z' away; objects closer or further away will, in general, be blurred. This is because the light for those objects will focus closer or further away and, as a result, on the image plane will form a *circle of confusion* – imaged rays will form a cone the base of which will dictate the amount of blur (larger base / circle of confusion = more blur). The amount of blur will be a function of focal length and aperture size.

However, consider what happens as the object moves further away, e.g., take limit of the lens equation above, as $z \rightarrow \inf$ the rays (which are becoming parallel) will focus at $z' = f'$. This means that so long as all objects are beyond certain distance from the camera they will all be imaged sharply with imaging plane at $z' = f'$. In practice, because for most cameras focal length f' is small (e.g., for iPhone 26 mm), any object which is a reasonable multiple of the focal length (e.g., 10x or 20x) away from the camera will be in focus on an imaging plane at approximately $z' = f'$.

Exercise: Consider a scenario with a lens who's focal length is 50 mm = 0.05 m, where an object being imaged is 5 m away. Where should we place and imaging plane to ensure the object is in focus?

$$
\begin{aligned}
\frac{1}{z'} &= \frac{1}{f'} + \frac{1}{z} \\
\frac{1}{z'} &= \frac{1}{0.05} + \frac{1}{-5} \\
\frac{1}{z'} &= 19.8 \\
z' &= \frac{1}{19.8} = 0.0505\cdots \approx 0.05 \text{ m}
\end{aligned}
$$

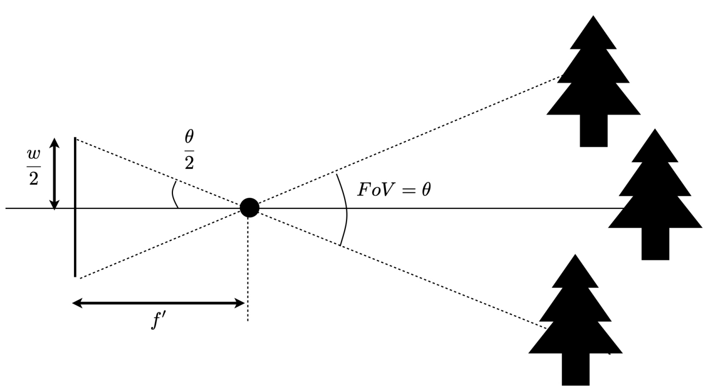
Why is the optimal placement of the imaging plane so close to the focal length? Because relative to focal length of the lens, the object is very far away, i.e., could be considered almost at infinity.

Exercise: Consider a scenario with a lens who's focal length is $5m$ (not realistic, but humor me), where an object being imaged is $5m$ away. Where should we place and imaging plane to ensure the object is in focus?

$$
\begin{aligned}
\frac{1}{z'} &= \frac{1}{f'} + \frac{1}{z} \\
\frac{1}{z'} &= \frac{1}{5} + \frac{1}{-5} \\
\frac{1}{z'} &= 0.0 \\
z' &\approx \infty\, m
\end{aligned}
$$

5 Field of View and Depth of Field

An infinite size sensor can image any object in front of the camera. However, in practice, our sensors are small and hence can only image effectively a cone encompassing part of the world. *Field of view* (FoV) is defined as the extent of this cone in terms of the angle being captured. FoV can be easily computed from the focal length and the size of the sensor (e.g., $w \times w$). The illustration of this below.



$$
\tan\left(\frac{\theta}{2}\right) = \frac{w}{2f'}
$$

$$
\theta = 2 \text{ atan}\left(\frac{w}{2f'}\right)
$$

Exercise: What is the field of view of a full frame (35mm) camera with a 50mm lens? 100mm lens?

$$FoV_{50mm} = 2\ \text{atan}\left(\frac{0.035}{2 * 0.050}\right) = 2\ \text{atan}(0.35) = 38.6\,\text{deg}$$
$$FoV_{100mm} = 2\ \text{atan}\left(\frac{0.035}{2 * 0.10}\right) = 2\ \text{atan}(0.175) = 19.9\,\text{deg}$$

The *depth of field* is the distance between the nearest and furthest elements in a scene that appear to be sharp in the image. In other words, how much of the scene in depth will be sharp. Note that this notion does not really makes sense for pinhole camera without a lens, as everything is always "sharp". In the pinhole camera with a lens the depth of field will be a function of the focal length and the aperture. Small aperture will have large *depth of field*, meaning that nearly everything in field of view would be in focus. This is easy to see if one takes aperture to the limit, which would result in a pinhole camera with light rays only traveling through the center of the lens (where they do not bend). With large aperture the *depth of field* will be small, meaning that objects even slightly away from an object in focus will appear significantly blurred.

6 Issues with Lenses

Spherical Aberration. Because lenses are not perfect, often times refraction will change as you get off center. Rays that strike closer to the edge of lens will generally focus closer. This means that no optimal position for the location of the imaging plane could be chosen. One typically is forced to place an image plane in the location that minimizes the blur (i.e., circle of confusion). This causes artifacts is called *aberration*. Aberrations can be minimizes by aligning several simpler lenses. These are called compound lenses that can be be modeled by *thick lens* equation. Analysis of this is beyond the scope of this class.

Vignetting. Vignetting is an artifact that arises in camera systems with multiple aligned lenses. Part of the beam of light refracted by the front lens may not hit the back lens. In this case portion of the light at the edge of the image is lost and does not reach a sensor. As a result edges appear darker.

Chromatic Aberration. In general, index of refraction depends on wavelength, λ, of light. Meaning that different wavelengths will be refracted differently by the lens and will focus at different depth. As a result no single placement of the imaging plane will result in all colors being in focus at the same time. Similar to spherical aberration, one typically is forced to place an image plane in the location that minimizes the blur (i.e., circle of confusion) among the spectra of color.

7 Human Eye

Human eye acts much like a pinhole camera. The projection in the human eye is similarly inverted and brain is responsible for upright interpretation. The *pupil* takes the role of the pinhole and *retina* serves as a curved imaging plane / sensor. We are able to change the shape of the pupil which, in turn, changes the focal length and ensures that image is focused at the retina. The retina contains light receptors called *rods* and *cones*. Rods are highly sensitive and can responding to a single photon; however, they yield relatively poor spatial detail and not involved in color vision. Cones are fewer and largely reside in foveal region. They are less sensitive (i.e., operate in high light) and yield high resolution color vision.