The Camera Model

George Tang

September 2018

1 Introduction

Computer vision seeks to automate the process of human vision through computational, mathematical, and physics-based techniques. Its has numerous, interdisciplinary applications, ranging from manufacturing to medicine.

Computer vision is an *inverse problem*: given insufficient visual information, construct a model to determine some unknowns, which could be a higher quality image or locating a region of interest.

2 3-D World

We live in a 3D world where every location is specified by three coordinates (X,Y,Z). Analogously, computer vision models place a camera and the surrounding objects in a 3D right-handed coordinate system. The images the camera captures is a 2D projection of the 3D space, and every pixel is specified by two coordinates (X,Y). The camera and objects are often translated and rotated. This is the subject of the next lecture.

3 Image Formation

The images the camera captures depends on lighting, surface properties, scene geometry, and lens optics. One should always keep these factors in mind when before applying computer vision. We will return to many of these concepts when we discuss computer graphics.

3.1 Lighting

Light sources in a scene are responsible for visibility and illumination. Emitted light bounces off surfaces and some eventually reach the lens of the camera.
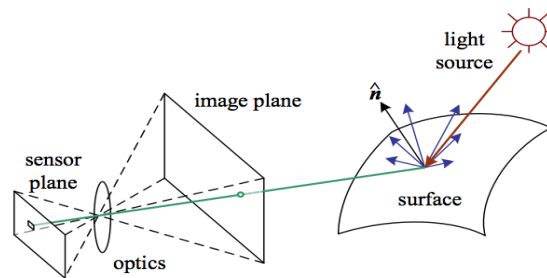
How images form

Light scattering is modeled by a Bidirectional Reflectance Distribution Function (BDRF). A BDRF takes an input direction and returns the proportion of light reflected in an output direction at point $\mathbf{p}$. Because of the physics of light, swapping the input and output directions yields the same result.

$$BDRF = f(\mathbf{p}, \mathbf{w_i}, \mathbf{w_o})$$

To determine the amount of light reflected in a given a direction given an input direction, we can multiply the BDRF with the amount of incoming light in that direction times the *foreshorting term*, which accounts for a decrease in reflectance as the angle between θ_i the incoming light and surface normal is increased. This is because the surface area exposed to the light is larger at oblique angles, and less light intensity is concentrated at a point.

$$L(\mathbf{p}, \mathbf{w_o}) = L(\mathbf{p}, \mathbf{w_i})f(\mathbf{p}, \mathbf{w_i}, \mathbf{w_o})cos^+(\theta_i)$$
$$cos^+(\theta_i) = max(0, cos(\theta_i))$$

Given many input directions, we can write:

$$L(\mathbf{p}, \mathbf{w_o}) = \sum_i L(\mathbf{p}, \mathbf{w_i})f(\mathbf{p}, \mathbf{w_i}, \mathbf{w_o})cos^+(\theta_i)$$

And taking into account incoming light from the entire surroundings, we can write:

$$L(\mathbf{p}, \mathbf{w_o}) = \int L(\mathbf{p}, \mathbf{w_i})f(\mathbf{p}, \mathbf{w_i}, \mathbf{w_o})cos^+(\theta_i)d\mathbf{w_i}$$

BDRF for a surface are usually obtained by physical modeling. Many BDRFs can be split into *diffuse* and *specular* components, allowing us to compute it easily.

3.1.1 Diffuse

The diffuse component accounts for shading. It scatters light uniformly in all directions, so the diffuse component of the BDRF is constant at $\mathbf{p}$.

$$L_d(\mathbf{p}, \mathbf{w_o}) = \int L(\mathbf{p}, \mathbf{w_i})f_d(\mathbf{p})cos^+(\theta_i)d\mathbf{w_i}$$

3.1.2 Specular

The specular component accounts for glossy reflections. Instead of scattering light, it reflects a single beam. The output direction thus has the same angle as the incoming light by Law of Reflection. Note the models define f_s as dependent on the angle, so we do not have to multiply it with the foreshorting term. For instance, the Phong BDRF defines f_s to be a power of the cosine angle.

$$\mathbf{s_i} = \mathbf{v}_\| - \mathbf{v}_\perp$$
$$L_s(\mathbf{p}, \mathbf{w_o}) = \int L(\mathbf{p}, \mathbf{w_i})f_s(\mathbf{p})d\mathbf{w_i}$$
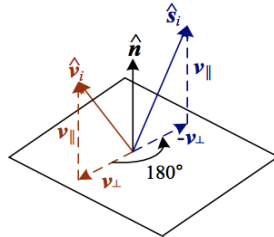

Law of Reflection

3.2 Lens Optics

After light reaches the camera, it must past through the lens.

3.2.1 Pinhole Camera Model

Usually it is sufficient to model the camera aperture, the location that allows the light through, as a pinhole. This means that all light rays that hit the camera converge at one focal point, so the image is simply a scaled 2D projection of the 3D space. This projection is known as a perspective transformation, and it along with other transformations will be covered in the next lecture. Notice in diagram 2, the pinhole is behind the image place. This is because by the location of the image plane behind the pinhole is just the front image plane flipped upside-down.
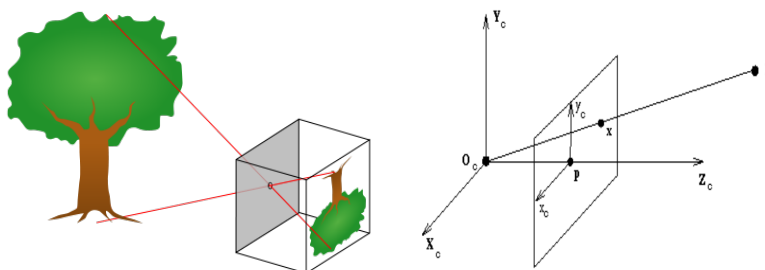
Pinhole Camera Model; Perspective Projection

However, if we want to account focus, exposure, etc, we must account for the lens.

3.2.2 Thin Lens Focus

Assume we have a thin convex lens. Let us define the distance the horizontal distance (z-axis in the diagram) of the object from the lens as S_1 and the distance from the lens to the plane to where the image formed is focused S_2. We also define the distance from the lens to the focus of the lens as f. In a pinhole camera, although there is no lens, we define $f = S_1$.

Because our lens is thin, by the thin lens approximation, rays passing through the center of the lens are not refracted. Thus, we can derive the relationship:

$$\frac{1}{f} = \frac{1}{S_1} + \frac{1}{S_2}$$

If the focal plane is moved away from the ideal position, the a fuzzy *circle of confusion* results. The acceptable range of object locations given a image plane setting is known as the *depth of field*. Pinhole cameras have an infinite depth of field.

As the object is moved towards ∞, note that $f = S_1$, which is exactly the property of the pinhole camera. This is because the light rays become straighten out and become parallel, and the lens results in a pinhole camera.
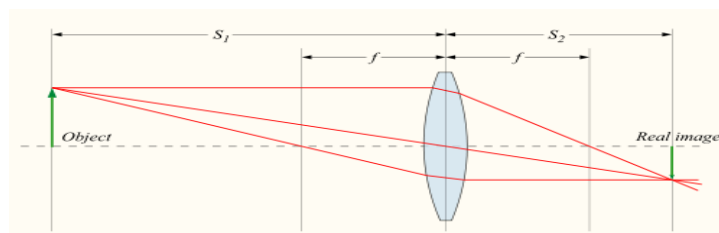
Simple Lens Law Schematic

4 The Digital Camera

Finally the light has reached the part of the camera that will capture the image in digital memory! Photons are picked up by the *active sensing area*, integrated into an representation for a exposure duration (camera parameter) by the *shutter*, and finally passed through a set of processes to yield the final result.
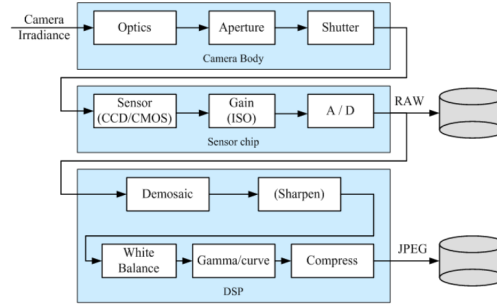
The Digital Camera

4.1 Color

Images are composed of pixels of RGB values that specify a unique color. The digital camera must integrate different wavelengths of incoming light into discrete RGB values. In place of each pixel is a R,G, or B sensor. For instance, the Bayer arrangement:
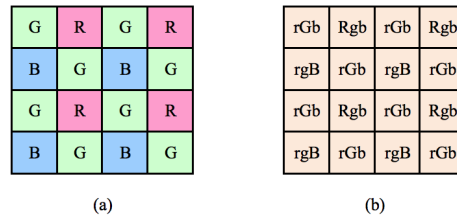
(a)Bayer arrangement, (b) interpolated colors

4.2 Compression

The last step is to compress an image to save memory. There are many compression algorithms out there that are far beyond the scope of this lecture.

We take the time to introduce the *peak signal-to-noise ration* (PSNR) measures the quality of an image, which is a standard metric used during compression and many other cases:

$$PSNR = 10log(\frac{I_{max}^2}{MSE})$$

where MSE is the mean squared error of the image and I_{max} is the maximum possible signal value (255 for 8-bit images).

$$MSE = \frac{1}{n}\sum_{i}[I(i) - I(\hat{i})]^2$$

where n is the number of pixels, $I(i)$ is the pixel value of the uncompressed image, and $I(\hat{i})$ is the value of its compressed counterpart. PSNR is used in assessing image quality in signal processing, computer vision, computer graphics, and computational photography.