Lecture 02: Image Formation Part I

1 Image Formation

How are objects in the world captured in an image? In this chapter the image formation process will be introduced.

1.1 Optics

1.1.1 Preliminary Example

Place a film in front of an object and illuminate the object. The light is then reflected on the film. This setup is reported in Figure 2. You will obtain an image which is not **reasonable**. In fact, the rays don't converge in the same point, causing an unsharp and blurred version of the real image.

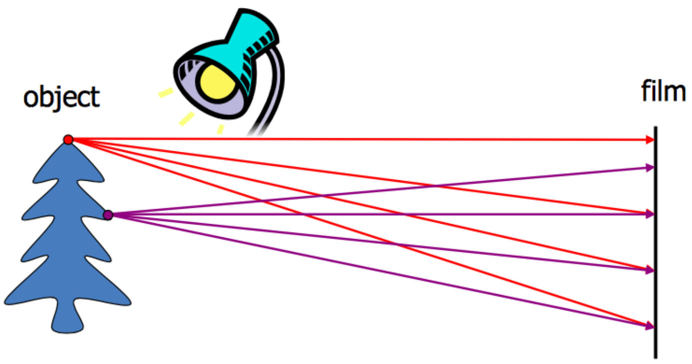

Figure 1: Basics of Image Formation .

Definition 1. In optics, a **circle of confusion (blur circle)** is an optical spot caused by a cone of light rays from a lens, not coming to a perfect focus when imaging a point source.

1.1.2 The Pinhole Camera

Recalling the preleminary example above and adding a barrier layer with a pinhole between the object and the film, one gets a *camera obscura*. This is depicted in Figure 1. It holds:

- The opening is called **aperture**.
- The pinhole reduces blurring.
- The **ideal pinhole** allows only one ray of light reaching each point on the film. A bigger aperture translates into a blurrier image.

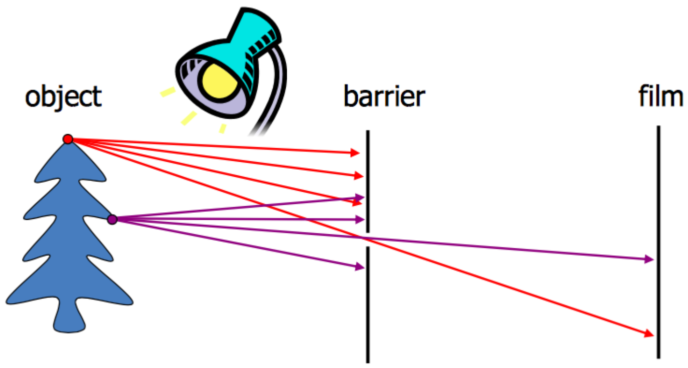

Figure 2: Pinhole Camera .

Now that we know tha a large aperture translates into blurry images, why not choosing it as small as possible? Diffraction effects cause **interference** in waves as we approach their wavelength and less light gets through.

Remark. Camera obscura is a latin term, which means *dark room*. The basic principle originates from Mozi (470-390 BC) and Aristotle (388-322 BC). Artists used it to paint landscapes (drawing aid for artists by Leonardo Da Vinci, 1452-1519).

1.1.3 Converging Lens

If one substitutes the pinhole barrier with a converging lens in the previous setup, one gets the setup depicted in Figure 3. This setup is extremely useful, because:

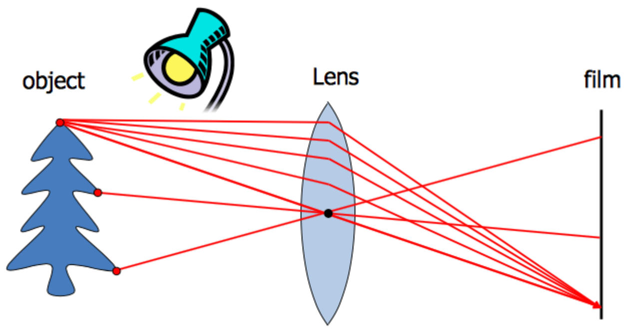

Figure 3: Setup with converging lens .

- The lens focuses light onto the film.
- The rays passing through the **Optical Center** are not deviated.
- All rays parallel to the **Optical Axis** converge at the **Focal Point**

Using similar triangles and Figure 4 one can write

$$\begin{aligned}\frac{B}{A} &= \frac{e}{z} \quad \text{and} \\ \frac{B}{A} &= \frac{e-f}{f} = \frac{e}{f} - 1.\end{aligned} \tag{1.1}$$

Putting these two relations together, one gets the so called **thin lens equation**.

$$\frac{e}{f} - 1 = \frac{e}{z}. \tag{1.2}$$

This result can be used to determine focus properties: any object point satisfying Equation 1.2 is in **focus**. This is used for retrieving the distance to an object (depth from focus). What would happen if $z \gg f$? One would get

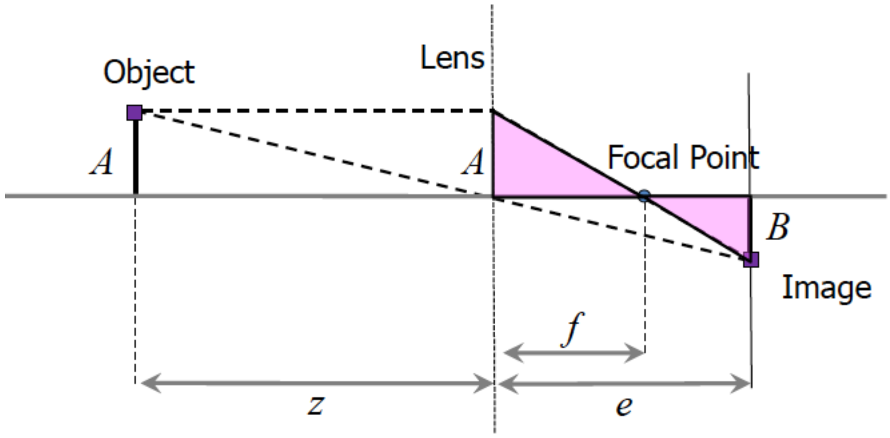

Figure 4: Thin lens .

$$\frac{1}{f} \approx \frac{1}{e} \Rightarrow f \approx e. \tag{1.3}$$

This is why we can think of a lens of focal length f as being equivalent to a pinhole distance f from the focal plane. We need to adjust the image plane such that objects at infinity are in focus, namely

$$\frac{h'}{h} = \frac{f}{z} \Rightarrow h' = \frac{f}{z}h, \tag{1.4}$$

where h and h' are defined in Figure 5. The dependence of the apparent size of the object on its depth is known as **perspective** (e.g. far away objects appear smaller).

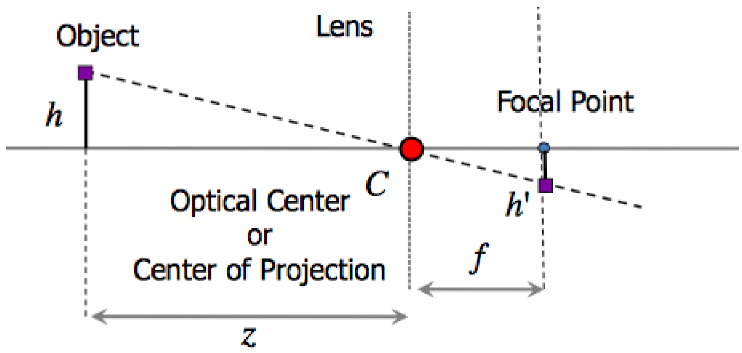

Figure 5: Thin lens for the case $z \gg f$.

But what does it mean to be **in focus**? It holds

- There is a specific distance from the lens at which world points are in focus in the image.
- Other points project onto a **blur circle** in the image, with radius

$$R = \frac{L\delta}{2e}. \tag{1.5}$$

From this it follows that a minimal pinhole gives a minimal R and that R should remain smaller than the image resolution.

1.1.4 Projective Geometry

How can you perceive a 3D scene by viewing its 2D representation (image)? The use of perspective prjection is known to be used during the first century BC for frescoes. During Renaissance (1480-1515), artists developed methods to determine perspective projection. In the transition from 3D to 2D, there are elements which are preserved and other which are lost. It holds:

- Straight lines are still straight.
- Lengths and angles are lost.
- Parallel lines in the world intersect in the image at a **vanishing point**.
- Parallel planes in the world intersect in the image at a **vanishing line**.

Example 1. A good example of misleading 3D perception is the Ames room. An example of this is depicted in Figure 6.

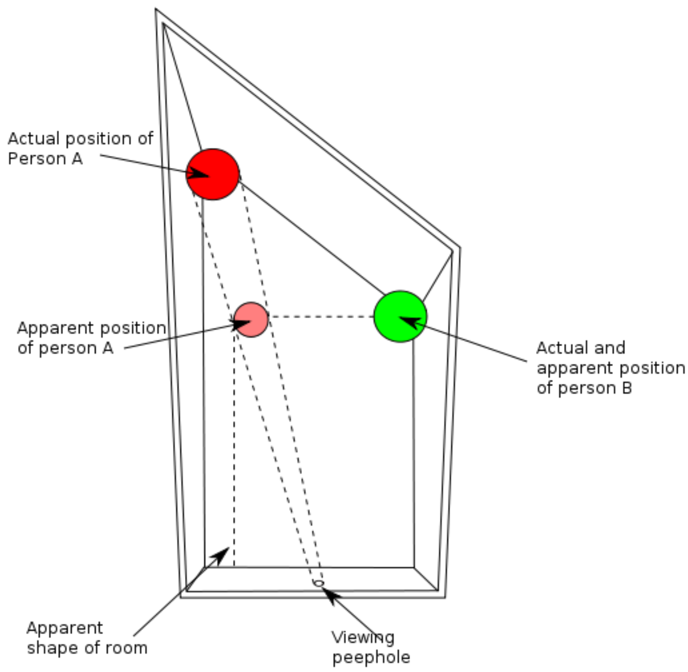

Figure 6: Example of Ames Room

Exercise 1. Prove that world's parallel lines intersect at a vanishing point in the camera image.

Solution.

Proof. Let's define two parallel 3D lines:

$$\begin{aligned} \begin{pmatrix} X \\ Y \\ Z \end{pmatrix} &= \begin{pmatrix} X_1 \\ Y_1 \\ Z_1 \end{pmatrix} + s \cdot \begin{pmatrix} l \\ m \\ n \end{pmatrix}, \\ \begin{pmatrix} X \\ Y \\ Z \end{pmatrix} &= \begin{pmatrix} X_2 \\ Y_2 \\ Z_2 \end{pmatrix} + s \cdot \begin{pmatrix} l \\ m \\ n \end{pmatrix}. \end{aligned} \tag{1.6}$$

The perspective projection equations in calibrated coordinates are

$$\begin{aligned} x &= \frac{X}{Z} \\ y &= \frac{Y}{Z}. \end{aligned} \tag{1.7}$$

Plugging the line equations into the perspective projection equation and letting $s \to \infty$ results in

$$\begin{aligned} \lim_{s\to\infty} \frac{X_i + sl}{Z_i + sn} &= \frac{l}{n} = x_{VP} \\ \lim_{s\to\infty} \frac{Y_i + sm}{Z_i + sn} &= \frac{m}{n} = y_{VP}. \end{aligned} \tag{1.8}$$

This result depends only on the direction vector of the line. □

Other Parameters

Definition 2. The **depth of field (DOF)** is the distance between the nearest and the farthest objects in a 3D scene that appear acceptably sharp in a 2D image.

Remark. Although a lens can precisely focus at only one distance at a time, the decrease in sharpness is gradual on each side of the focused distance, so that within the DOF, the unsharpness is imperceptible under normal viewing conditions.
The *aperture* influences the depth of field: a smaller aperture increases the range in which the object appears approximately in focus, but reduces the amount of light into the camera.

Definition 3. The **field of view (FOV)** is an angular measure of the portion of 3D space seen by the camera.

Using Figure 7, one can compute the **FOV** as

$$\tan\left(\frac{\theta}{2}\right) = \frac{W}{2f}, \tag{1.9}$$

which implies

$$f = \frac{W}{2}\left[\tan\left(\frac{\theta}{2}\right)\right]^{-1}. \tag{1.10}$$

It follows that a larger focal length results in a smaller field of view. The FOV is affected by the focal length f:

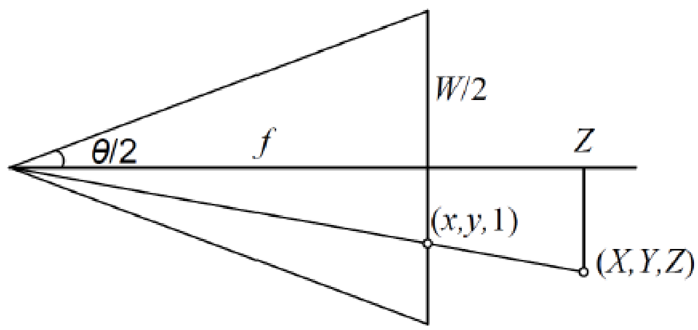

Figure 7: Field of view computation.

- As f gets smaller, the image becomes more *wide angle*, i.e. more world points project onto the finite image plane.
- As f gets larger, the image becomes more *narrow angle*, i.e. a smaller part of the world projects onto the finite image plane.

1.1.5 Digital Camera

Instead of using a film we use a sensor array and we convert detected information in numbers (e.g. $[0, 255]$ for 8 bytes), as depicted in Figure 8. Each pixel is characterized by

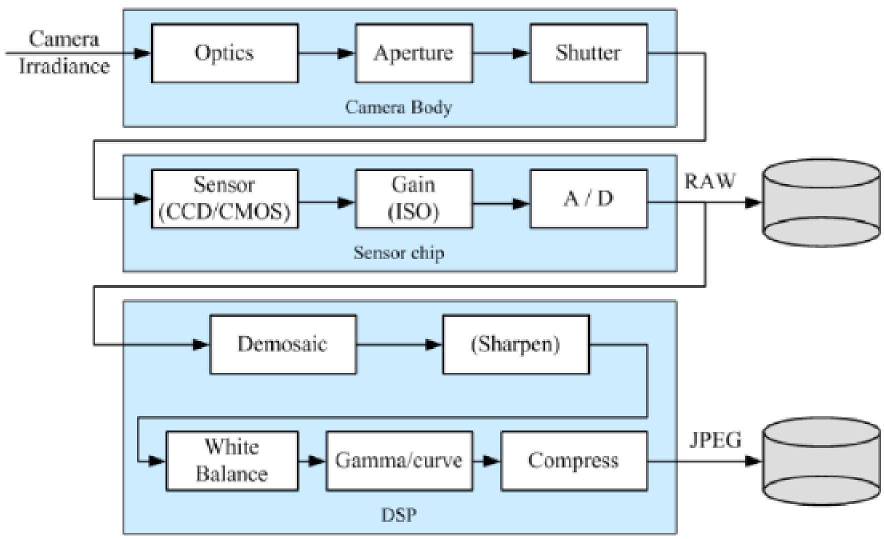

Figure 8: Working principle for digital cameras.

its row and column number and its intensity.

Color Sensing

In order to sense colors, the following procedure is used:

- Bayer pattern (1976) places green filters over half of the sensors and red and blue filters over the remaining ones. This is because the luminance signal is mostly determined by green values and the human visual system is much more sensitive to high frequency detail in luminance than in chrominance.

  Remark. This method has the disadvantage that the number of detected pixels is cut by 4.

- The three-chip color camera splits in three color filters the light and each chip measures the light intensity for one color. Here, resolution is preserved.

- Estimate missing components from neighboring values: **demosaicing**.

1.2 Perspective Camera Model

In the following, we will refer to Figure 9 for the description of the perspective camera model. With $\boldsymbol{C}$ we represent the optical center (the center of the lens), with $\boldsymbol{O}$ the principal point, with $\boldsymbol{W}$ the center of the world frame and with $\boldsymbol{Z_c}$ the optical axis. For convenience, the image plane is usually represented in front of $\boldsymbol{C}$, such that the image preserves the same orientation (i.e. is not flipped). The general procedure reads

1. Find pixel coordinates (u, v) of point ${}_W\boldsymbol{P}$.
2. Convert world point ${}_W\boldsymbol{P}$ to camera point ${}_C\boldsymbol{P}$.
3. Find pixel coordinates (u, v) of point ${}_C\boldsymbol{P}$ in the camera frame.
4. Convert ${}_C\boldsymbol{P}$ to image-plane coordinates (x, y).
5. Convert ${}_C\boldsymbol{P}$ to discretized pixel coordinates (u, v).

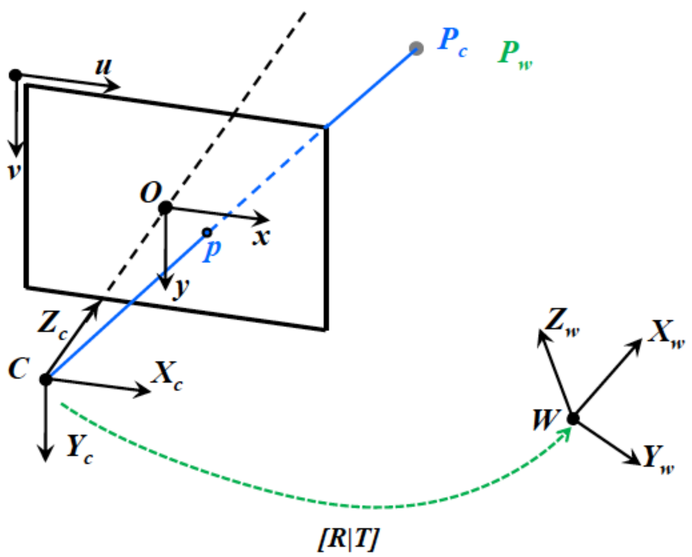

Figure 9: Frames.

1.2.1 Perspective Projection

We use the notation from euclidean coordinates to homogeneous ones:

$$\underbrace{\begin{pmatrix} u \\ v \\ w \end{pmatrix}}_{Homog.} \rightarrow \underbrace{\begin{pmatrix} u/w \\ v/w \\ 1 \end{pmatrix}}_{Eucl.} \tag{1.11}$$

Using similar triangles and Figure 10, one gets

$$\begin{aligned} \frac{x}{f} &= \frac{X_c}{Z_c} \Rightarrow x = \frac{fX_c}{Z_c} \\ \frac{y}{f} &= \frac{Y_c}{Z_c} \Rightarrow y = \frac{fY_c}{Z_c} \end{aligned} \tag{1.12}$$

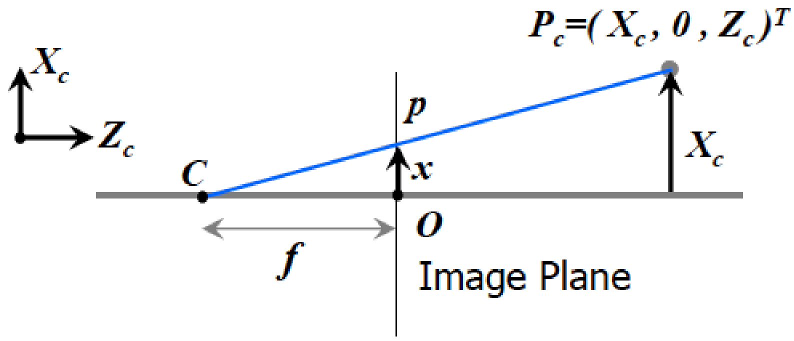

Figure 10: Figure for perspective projection.

1.2.2 Pixel Coordinates

From local image-plane coordinates (x, y) to the pixel coordinates (u, v), with scale factors $k_{u,v}$ (inverse of the effective pixel size along the u (v) direction, measured in pixel m^{-1}) and the pixel coordinates of the camera optical center $\boldsymbol{O} = (u_0, v_0)$:

$$\begin{aligned} u &= u_0 + k_u x \Rightarrow u = u_0 + \frac{k_u f X_c}{Z_c} \\ v &= v_0 + k_v y \Rightarrow v = v_0 + \frac{k_v f Y_c}{Z_c}, \end{aligned} \tag{1.13}$$

which, expressed in matrix form read

$$\begin{aligned} \begin{pmatrix} \lambda u \\ \lambda v \\ \lambda \end{pmatrix} &= \begin{pmatrix} \alpha_u & 0 & u_0 \\ 0 & \alpha_v & v_0 \\ 0 & 0 & 1 \end{pmatrix} \cdot \begin{pmatrix} X_c \\ Y_c \\ Z_c \end{pmatrix} \\ &= K \begin{pmatrix} X_c \\ Y_c \\ Z_c \end{pmatrix}, \end{aligned} \tag{1.14}$$

with the focal lengths $\alpha_{u,v} = k_{u,v} f$ and the **calibration matrix** (intrinsic parameter matrix) K. Using homogeneous coordinates, one can write

$$\lambda \cdot \begin{pmatrix} u \\ v \\ 1 \end{pmatrix} = \underbrace{K}_{3\times 3,\ intrinsics} \cdot \underbrace{\begin{pmatrix} R & | & \boldsymbol{t} \end{pmatrix}}_{3\times 4,\ extrinsics} \cdot \begin{pmatrix} X_w \\ Y_w \\ Z_w \\ 1 \end{pmatrix}. \tag{1.15}$$

Exercise 2.

a) Determine the intrinsic parameter matrix K for a digital camera with image size 640 $\times$ 480 pixels and horizontal field of view equal to 90°. Assume the principal point in the center of the image and squared pixels.

b) What is the vertical field of view?

Solution.

a) From the formula for the field of view, we recover that

$$\begin{aligned} f &= \frac{W}{2}\left[\tan\left(\frac{\theta}{2}\right)\right]^{1} \\ &= \frac{640}{2\tan\left(\frac{90}{2}\right)} \\ &= 320 \text{ pixels.} \end{aligned} \tag{1.16}$$

This means that

$$K = \begin{pmatrix} 329 & 0 & 320 \\ 0 & 320 & 240 \\ 0 & 0 & 1 \end{pmatrix}. \tag{1.17}$$

b) It holds

$$\begin{aligned} \theta_{\mathrm{V}} &= 2\mathrm{arctan}\left(\frac{H}{2f}\right) \\ &= 2\mathrm{arctan}\left(\frac{480}{2\cdot 320}\right) \\ &= 73.74°. \end{aligned} \tag{1.18}$$

1.2.3 Lens Distortion

Radial Distortion

The amount of distortion of the coordinates of the observed image is a nonlinear function of their radial distance. This is a transformation from ideal to distorted coordinates ((u, v) to $(u_\mathrm{d}, v_\mathrm{d})$. For most lenses, one writes a simple quadratic model

$$\begin{pmatrix} u_\mathrm{d} \\ v_\mathrm{d} \end{pmatrix} = (1 + k_1 r^2) \cdot \begin{pmatrix} u - u_0 \\ v - v_0 \end{pmatrix} + \begin{pmatrix} u_0 \\ v_0 \end{pmatrix}, \tag{1.19}$$

with

$$r^2 = (u - u_0)^2 + (v - v_0)^2. \tag{1.20}$$

Depending on the amount of distortion, one can introduce higher order terms:

$$\begin{pmatrix} u_d \\ v_d \end{pmatrix} = (1 + k_1 r^2 + k_2 r^4 + k_3 r^6) \cdot \begin{pmatrix} u - u_0 \\ v - v_0 \end{pmatrix} + \begin{pmatrix} u_0 \\ v_0 \end{pmatrix} \tag{1.21}$$

1.3 Understanding Check

Are you able to:

- *Explain what is a Blur Circle?*
- *Derive the thin lens equation and perform the pinhole approximation?*
- *Define vanishing points and lines?*
- *Prove that parallel lines intersect at vanishing points?*
- *Explain how to build an Ames room?*
- *Derive a relation between the field of view and the focal length?*
- *Explain the perspective projection equation, including lens distortion and world to camera projection?*