CS-570 Computer Vision

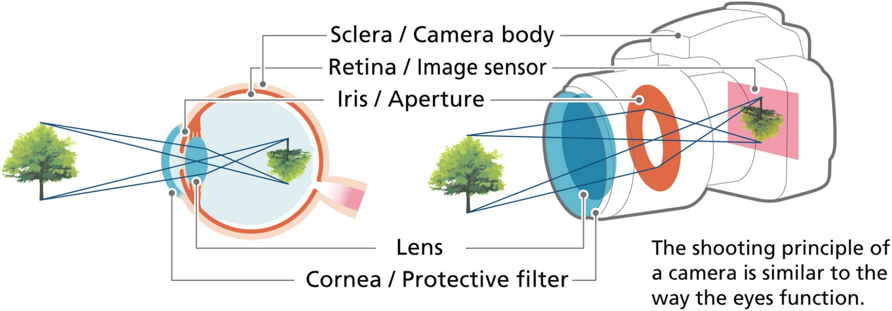
Nazar Khan

Department of Computer Science
University of the Punjab

18. Camera Geometry

Imaging Devices

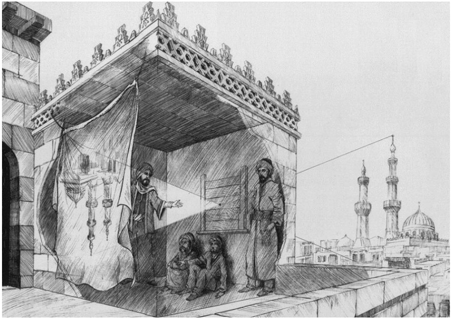
- Animal eyes and cameras share many geometric and photometric properties.



https://www.healthcare.nikon.com/en/ophthalmology-solution/valuing-eyes/

Camera Obscura

- *Camera obscura* (dark chamber) phenomena explored by ancient Chinese and Greeks.
- Extensively studied by Abu Ali al-Hassan ibn al-Haytham[1] in the 11th century.

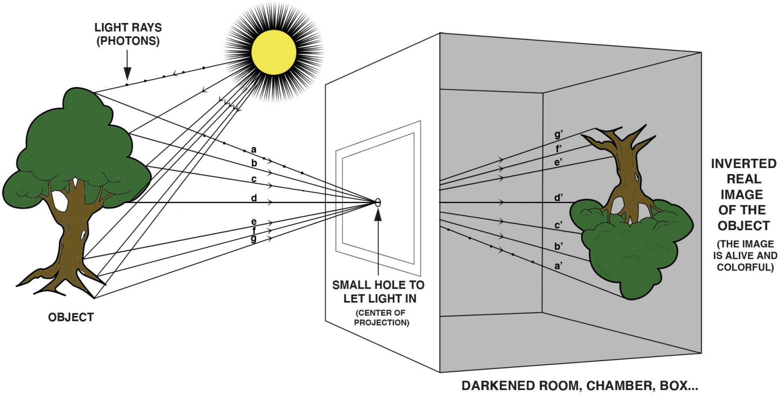


https://www.ibnalhaytham.com

[1]Father of Optics https://en.wikipedia.org/wiki/Ibn_al-Haytham

Pinhole Camera

- Pinhole used to focus light rays onto a wall or translucent plate in a dark box.



https://bonfoton.com/blogs/news/what-is-a-camera-obscura

Pinhole Camera

- Small pinhole produces sharp but dim pictures.
- Large pinhole produces brighter but blurry pictures.
- Pinholes gradually replaced by lenses to produce bright and sharp images.
- Backplane replaced by photosensitive material.
- Modern camera is essentially a camera obscura that records the amount of light striking each location of its image plane (retina).

Pinhole Camera

Real Image vs. Virtual Image

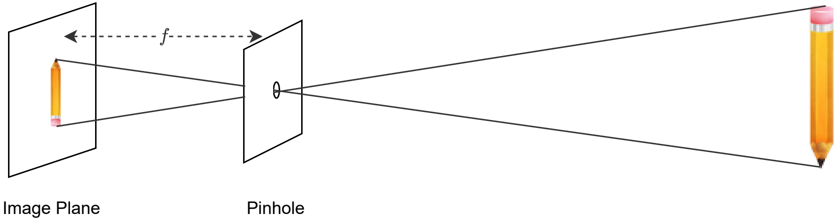


Figure: The real image is formed behind the pinhole on the image plane. The image is flipped horizontally and vertically. Author: N. Khan (2021)

Pinhole Camera

Real Image vs. Virtual Image

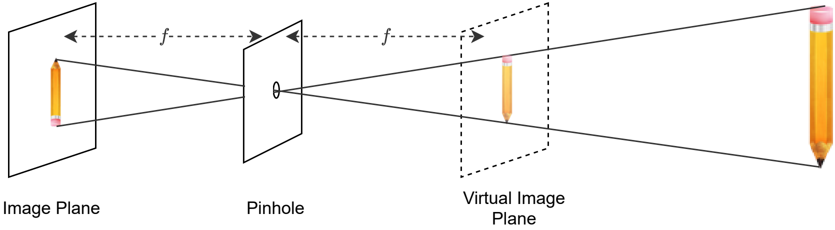


Figure: By *imagining* a virtual image plane *in front* of the pinhole, we can work with virtual images in the same orientation as the scene. The real and virtual image planes are otherwise geometrically equivalent. Author: N. Khan (2021)

Pinhole Camera

Real Image vs. Virtual Image

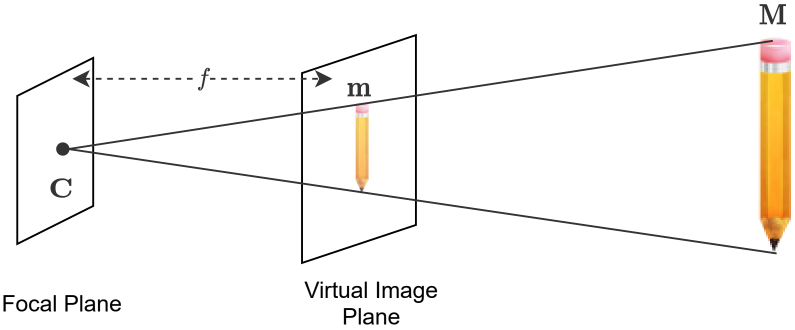


Figure: The pinhole can be modelled as the *focal point/camera center* **C**. Author: N. Khan (2021)

Pinhole Camera Model

Virtual Image Plane

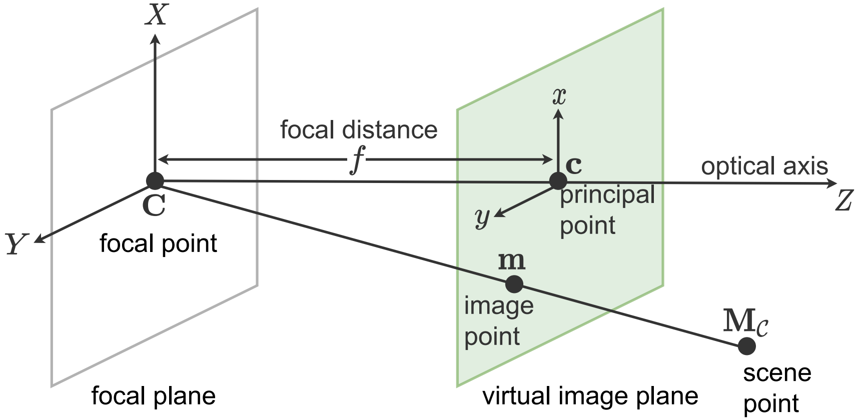


Figure: Pinhole camera model with virtual image plane. Author: N. Khan (2021)

Camera Projection Equations

- Since $\mathbf{C}, \mathbf{m} = (x, y, f)$ and $\mathbf{M}_{\mathcal{C}} = (X, Y, Z)$ are collinear

$$\overrightarrow{\mathbf{Cm}} = \lambda \overrightarrow{\mathbf{CM}_{\mathcal{C}}}$$

$$\implies \begin{cases} x = \lambda X \\ y = \lambda Y \\ f = \lambda Z \end{cases} \iff \lambda = \frac{x}{X} = \frac{y}{Y} = \frac{f}{Z}$$

- Therefore, camera projection equations are

$$x = f\frac{X}{Z}$$
$$y = f\frac{Y}{Z}$$

World Coordinates to Camera Coordinates

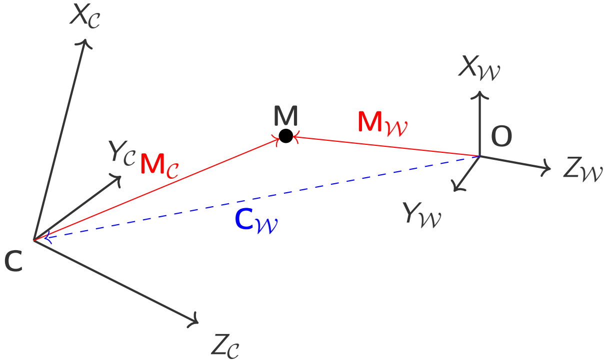


Figure: Any 3D location M has different representations in different coordinate systems. The camera center C itself is a 3D location represented in a world coordinate system. Author: N. Khan (2018)

World to Camera Coordinates

- Change of coordinates from world to camera frame in nonhomogenous coordinates can be obtained as

$$\mathbf{M}_{\mathcal{C}} = R\mathbf{M}_{\mathcal{W}} + \mathbf{t}$$

where the 3×3 matrix R represents a $3D$ rotation and $\mathbf{t}$ is a $3D$ translation vector.
- In homogenous coordinates, the same rigid transformation can be performed as

$$\mathbf{M}_{\mathcal{C}} = T\mathbf{M}_{\mathcal{W}}$$

where

$$T = \begin{bmatrix} R & \mathbf{t} \\ \mathbf{0}^T & 1 \end{bmatrix}$$

is a 4×4 matrix.

3D Rotations

- In $3D$, any arbitrary rotation is represented by a 3×3 matrix

$$\begin{bmatrix} r_{11} & r_{12} & r_{13} \\ r_{21} & r_{22} & r_{23} \\ r_{31} & r_{32} & r_{33} \end{bmatrix}$$

- It can be decomposed into a sequence[2] of rotations around the X-, Y- and Z-axes by angles θ_x, θ_y and θ_z respectively.

$$\underbrace{\begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos\theta_x & -\sin\theta_x \\ 0 & \sin\theta_x & \cos\theta_x \end{bmatrix}}_{X\text{-axis}} \quad \underbrace{\begin{bmatrix} \cos\theta_y & 0 & \sin\theta_y \\ 0 & 1 & 0 \\ -\sin\theta_y & 0 & \cos\theta_y \end{bmatrix}}_{Y\text{-axis}} \quad \underbrace{\begin{bmatrix} \cos\theta_z & -\sin\theta_z & 0 \\ \sin\theta_z & \cos\theta_z & 0 \\ 0 & 0 & 1 \end{bmatrix}}_{Z\text{-axis}}$$

- Therefore, a $3D$ rotation is parameterized by 3 rotation angles only.

[2] Remember that order matters!

Extrinsic Parameters

- The transformation T from world to camera coordinates has 6 parameters
  - 3 for rotation: $\theta_x, \theta_y, \theta_z$
  - 3 for translation: t_x, t_y, t_z
- They represent the *extrinsic parameters* of the camera.

Projection

- After transforming world coordinates $\mathbf{M}_{\mathcal{W}}$ to camera coordinates $\mathbf{M}_{\mathcal{C}}$, the *image coordinates* can be obtained via the projection equations

$$x = \frac{fX_{\mathcal{C}}}{Z_{\mathcal{C}}}$$
$$y = \frac{fY_{\mathcal{C}}}{Z_{\mathcal{C}}}$$

- In homogenous coordinates

$$\begin{bmatrix} f & 0 & 0 & 0 \\ 0 & f & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} X_{\mathcal{C}} \\ Y_{\mathcal{C}} \\ Z_{\mathcal{C}} \\ W_{\mathcal{C}} \end{bmatrix}$$

- Note that if units for focal length f were inches, then the projections x and y are still in inches.

Projection to pixels

- To convert to pixels, multiply by pixels-per-inch.
- In imaging sensors with rectangular pixels, pixels-per-inch will be different for x and y directions.
- Let p_x be the pixels-per-inch in x-direction and p_y be the pixels-per-inch in y-direction.
- Projection equations *in pixels* become

$$x = p_x \frac{fX_{\mathcal{C}}}{Z_{\mathcal{C}}}$$
$$y = p_y \frac{fY_{\mathcal{C}}}{Z_{\mathcal{C}}}$$

- Denoting $f_x = p_x f$ and $f_y = p_y f$

$$x = f_x \frac{X_{\mathcal{C}}}{Z_{\mathcal{C}}}$$
$$y = f_y \frac{Y_{\mathcal{C}}}{Z_{\mathcal{C}}}$$

Changing the origin

- To change origin from principal point **c** to some corner of the image, add the coordinates (u_0, v_0) of **c** with respect to new origin.
- Projection equations in pixels *with respect to new origin* become

$$x = p_x \frac{fX_\mathcal{C}}{Z_\mathcal{C}} + u_0$$
$$y = p_y \frac{fY_\mathcal{C}}{Z_\mathcal{C}} + v_0$$

Skewed Pixels

- Sometimes, sensor pixels can be slightly skewed due to manufacturing error.
- This means that x and y directions are not orthogonal. Instead, they have an angle θ (close to 90°) between them.
- Projection equations become

$$x = f_x \frac{X_\mathcal{C}}{Z_\mathcal{C}} - f_x \cot\theta \frac{Y_\mathcal{C}}{Z_\mathcal{C}} + u_0$$
$$y = \frac{f_y}{\sin\theta} \frac{Y_\mathcal{C}}{Z_\mathcal{C}} + v_0$$

- The 5 parameters $f_x, f_y, \theta, u_0, v_0$ are known as the *intrinsic parameters* of the camera.

Camera Matrix

- A 3D point in homogeneous world coordinates $(X_w, Y_w, Z_w, 1)^T$ is mapped to a 2D image point with homogeneous pixel coordinates $(u, v, w)^T$ as

$$\begin{pmatrix} u \\ v \\ w \end{pmatrix} = \underbrace{\begin{pmatrix} f_x & -f_x \cot\theta & u_0 \\ 0 & f_y / \sin\theta & v_0 \\ 0 & 0 & 1 \end{pmatrix}}_{\text{intrinsic}} \underbrace{\begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix}}_{\text{projection}} \underbrace{\begin{pmatrix} r_{11} & r_{12} & r_{13} & t_1 \\ r_{21} & r_{22} & r_{23} & t_2 \\ r_{31} & r_{32} & r_{33} & t_3 \\ 0 & 0 & 0 & 1 \end{pmatrix}}_{\text{extrinsic}} \begin{pmatrix} X_w \\ Y_w \\ Z_w \\ 1 \end{pmatrix}$$

$$= \underbrace{\begin{pmatrix} p_{11} & p_{12} & p_{13} & p_{14} \\ p_{21} & p_{22} & p_{23} & p_{24} \\ p_{31} & p_{32} & p_{33} & p_{34} \end{pmatrix}}_{\text{full projection matrix}} \begin{pmatrix} X_w \\ Y_w \\ Z_w \\ 1 \end{pmatrix}$$

- 12 parameters in total but with 1 free scaling parameter. So 11 degrees of freedom: 6 extrinsic plus 5 intrinsic.

Summary

- We have seen how a point in world coordinates is converted into its corresponding pixel coordinates via a single matrix multiplication in homogenous coordinates.

$$\mathbf{m} = P\mathbf{M}$$

- The whole process can be decomposed into into a sequence of 3 matrix multiplications
  1. Intrinsic
  2. Projection
  3. Extrinsic
- We have seen how parallel lines in a plane intersect under perspective projection.
- Next lecture: anatomy of the camera matrix P.