Image Formation, Camera Modeling and Camera Calibration

Dr. Fiseha B. Tesema

Outline

- Introduction
- Projection
  - Perspective projection (Pinhole)
  - Orthographic Projection
- Camera Modeling
- What is the effect of varying aperture size?
- Cameras and lenses
- Properties of Perspective Projection
- Propertied of Orthographic Projection
- Going to digital image space
  - Intrinsic and Extrinsic parameter
- Camera Calibration

Introduction

- The camera is one of the most essential tools in computer vision.
- It is the mechanism by which we can record the world around us and use its output - photographs - for various applications.
- Therefore, one question we must ask in computer vision is: how do we model a camera?

Source: S. Lazebnik

Projection

- A camera model is a function which maps 3-dimensional world onto a 2-dimensional plane, called the image plane.

- There are many camera models of varying complexity, and a natural dividing line which helps categorize them is whether or not they are able to capture perspective.
  - Perspective(Pinhole) Camera: Perspective, or the perspective effect is simply the property that objects far away from us appear smaller than objects up close.
  - Orthographic Camera: Cameras Which Do Not Capture The Perspective Effect

Pinhole Cameras

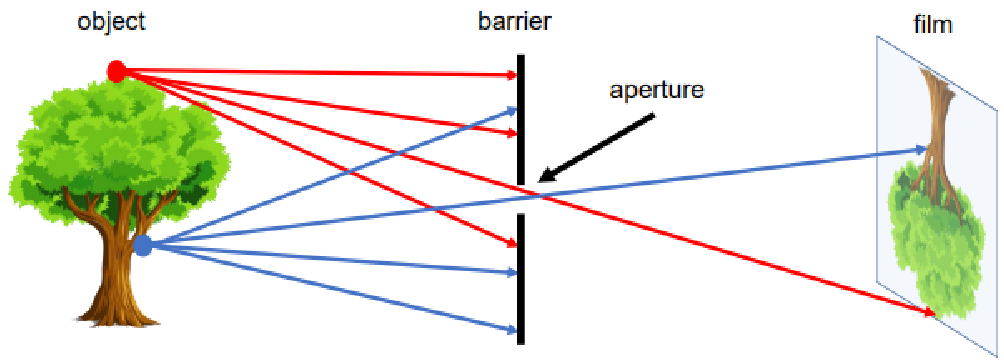


Fig: A simple working camera model: the pinhole camera model

- Camera system can be designed by placing a barrier with a small aperture between the 3D object and a photographic film or sensor.

- A pinhole camera is a simple camera without a lens but with a tiny aperture (the so-called pinhole)—effectively a light-proof box with a small hole in one side.

Pinhole Cameras

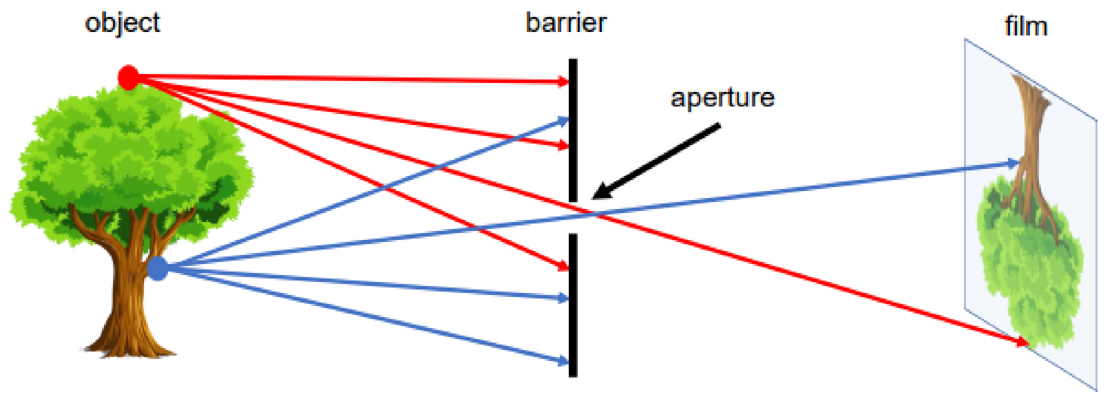
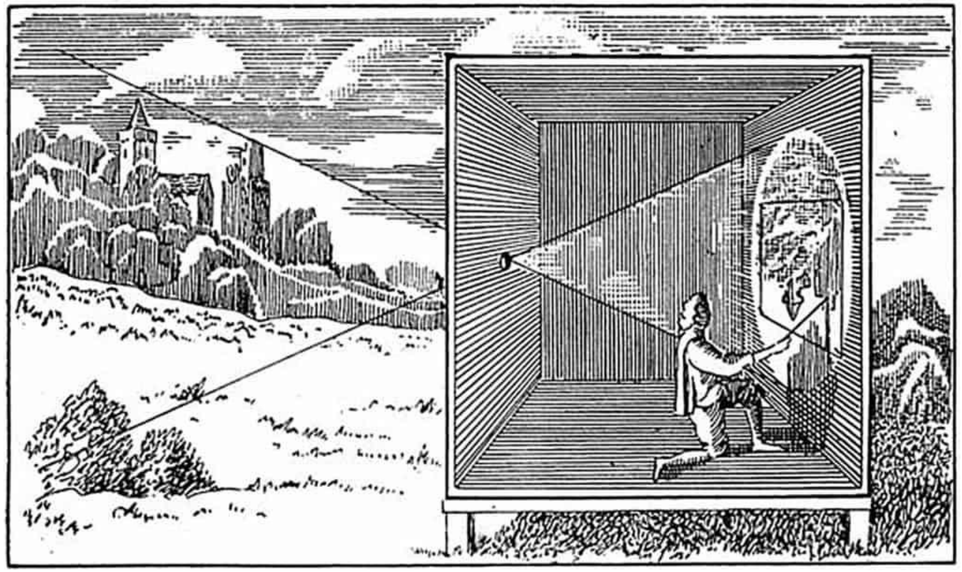


Fig: A simple working camera model: the pinhole camera model

- Each point on the 3D object emits multiple rays of light outwards.
  - Without a barrier, every point on the film will be influenced by light rays emitted from every point on the 3D object.
  - Due to the barrier, only one (or a few) of these rays of light passes through the aperture and hits the film.

Intro: History- Camera Obscura



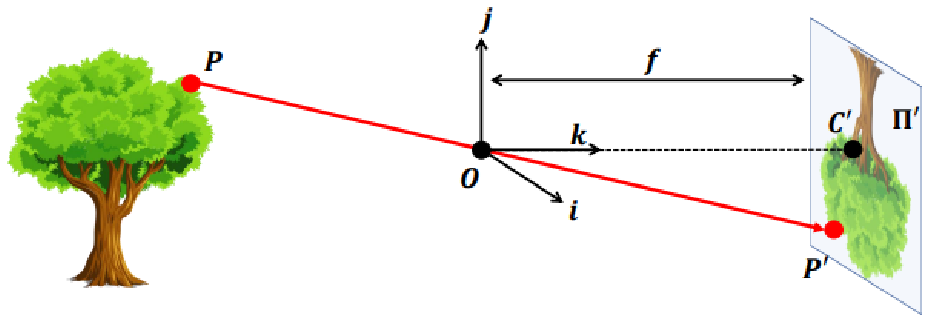
[https://magazine.artland.com/agents-of-change-camera-obscura/]

- A camera obscura (pl. camerae obscurae or camera obscuras; from Latin camera obscūra 'dark chamber') is a darkened room with a small hole or lens at one side through which an image is projected onto a wall or table opposite the hole.
- The earliest mention of the Camera Obscura dates back to Greek antiquity and the Chinese Han Dynasty (c. 468 – 391 BC).
- The Chinese philosopher Mozi was the first person to write down the principles of the Camera Obscura.

Create your own camera obscura
https://www.youtube.com/watch?v=gvzpu0Q9RTU

Camera Modeling: A formal construction of the pinhole camera model (perspective projection)

- Essential Components:
  - The film is commonly called the image or retinal plane:
    - The 2D plane where the projection of the 3D scene is captured, forming the image.
  - The aperture is referred to as the pinhole O or center of the camera.
    - The point through which all light rays from the 3D scene pass.
  - The focal length f.
    - The distance between the image plane and the pinhole O.
  - Camera Intrinsic (Will discuss in detail in the next session)
    - Parameters such as focal length, principal point (the intersection of the optical axis with the image plane), and skew (if the image axes are not perpendicular).



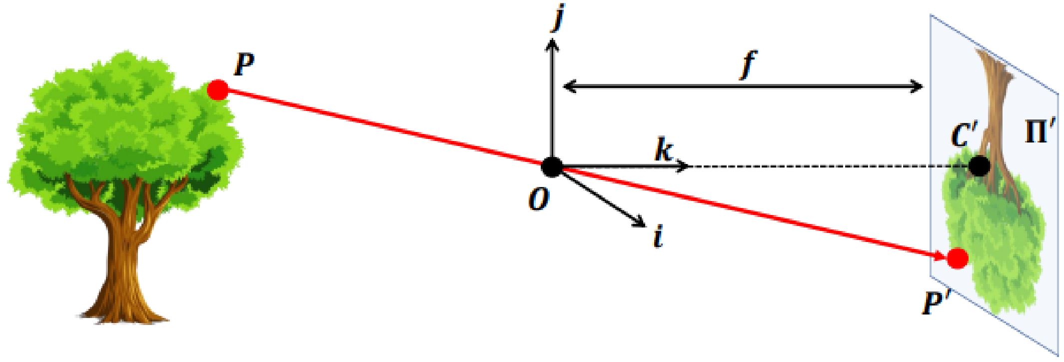
Cont. …

- Let be a point P be a point on some 3D object visible to the pinhole camera.
  - P will be mapped or projected onto the image plane π' resulting in point P'.
  - The pinhole itself can be projected onto the image plane, giving a new point C'.

$$P = \begin{bmatrix} x & y & z \end{bmatrix}^T$$

$$P' = \begin{bmatrix} x' & y' \end{bmatrix}^T$$

- Camera reference system or camera coordinate system
  - Coordinate system $[i\ j\ k]$ centered at the pinhole O such that the axis k is perpendicular to the image plane and points toward it.
  - The line defined by $C'$and O is called the optical axis of the camera system.

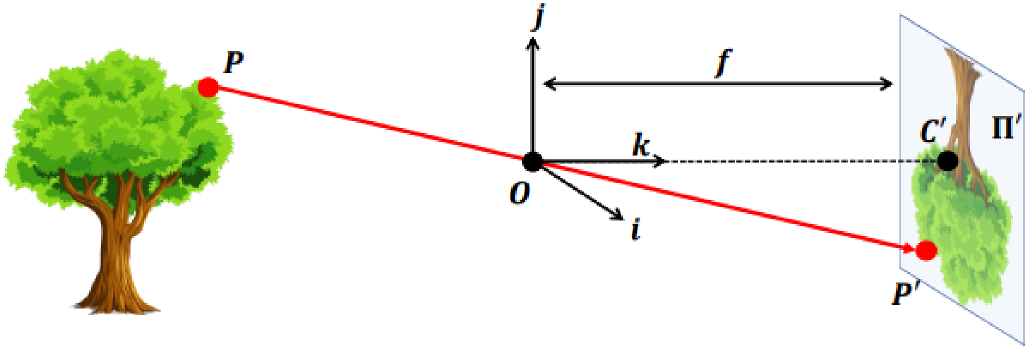


Cont. …

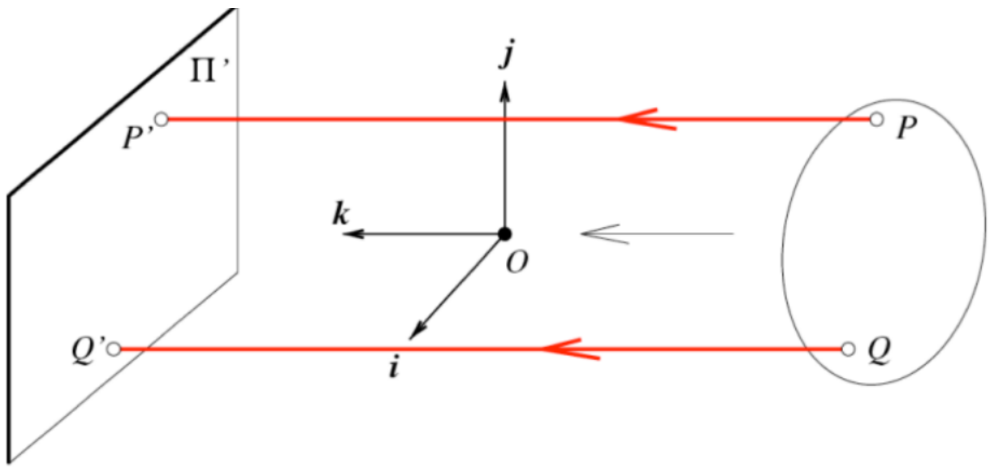
- Triangle $P'C'O$ is similar to the triangle formed by P, O and $(0, 0, Z)$. Therefore, using the law of similar triangles we find that:

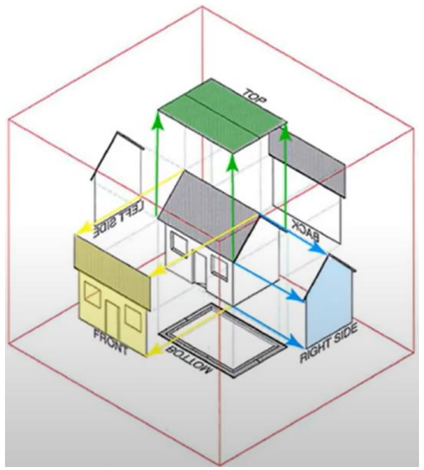
$$\frac{f}{z} = \frac{P'}{P}$$

$$P' = \begin{bmatrix} x' & y' \end{bmatrix}^T = \begin{bmatrix} f\frac{x}{z} & f\frac{y}{z} \end{bmatrix}^T \qquad (1)$$



Orthographic Projection



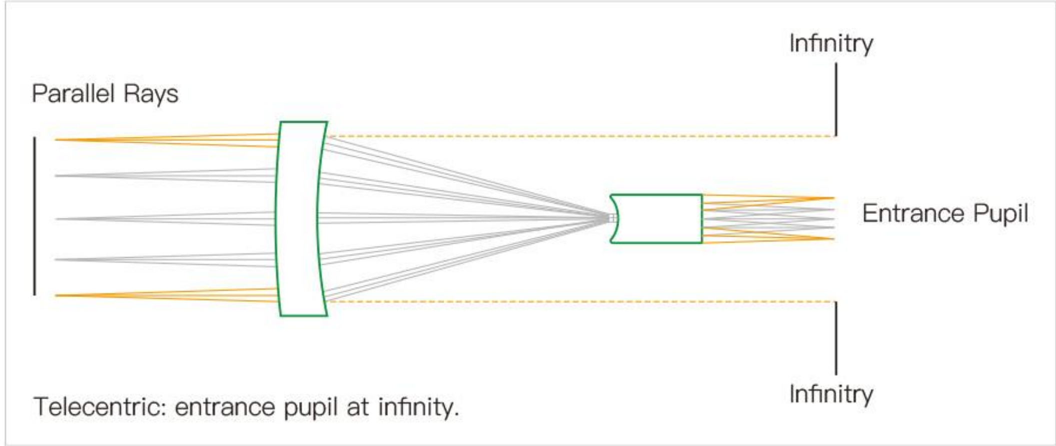


- The projection rays are now perpendicular to the retinal plane. As a result, this model ignores depth altogether.

$$x' = x$$
$$y' = y$$

Orthographic Projection

- Can be achieved using specialized hardware (e.g., telecentric lenses) or simulated in software.
- Orthographic cameras are particularly useful in technical and scientific applications where preserving accurate scale and proportions is essential.

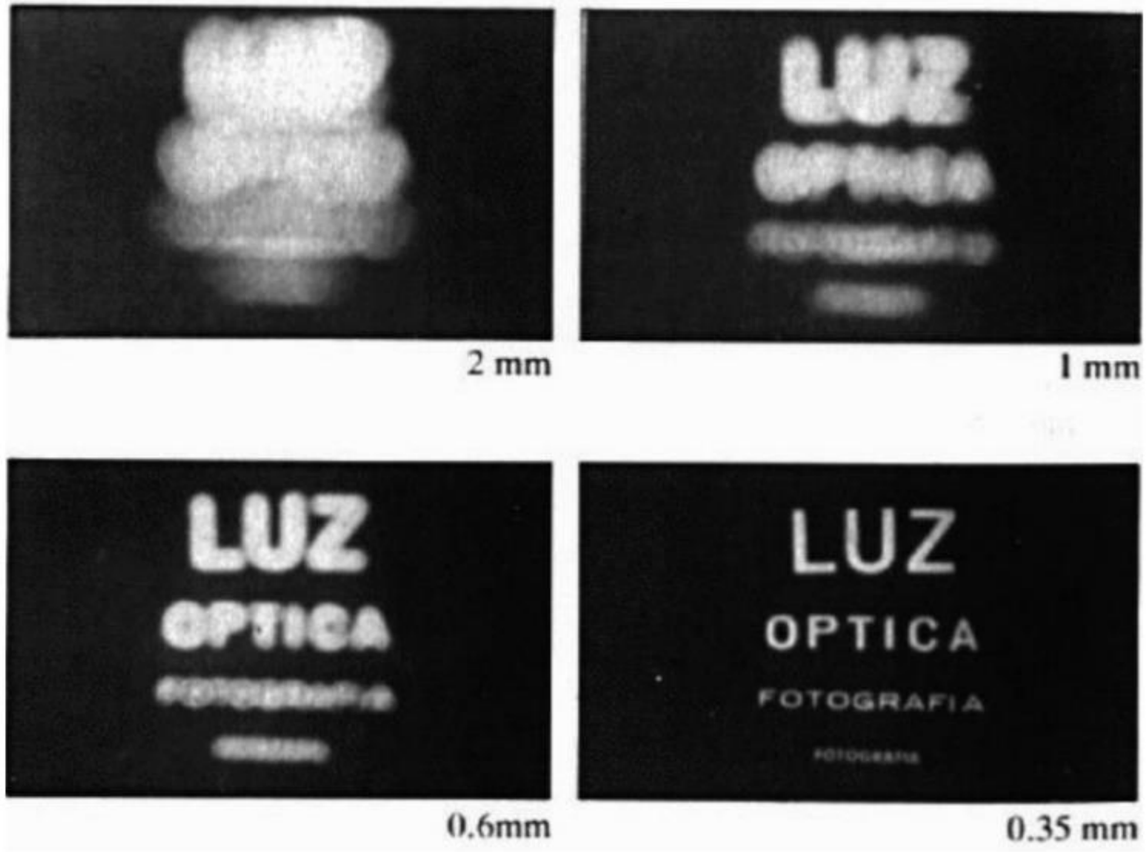


What is the effect of varying aperture size?

- As the aperture size increases, the number of light rays that passes through the barrier consequently increases.
    - Then each point on the film may be affected by light rays from multiple points in 3D space, blurring the image.

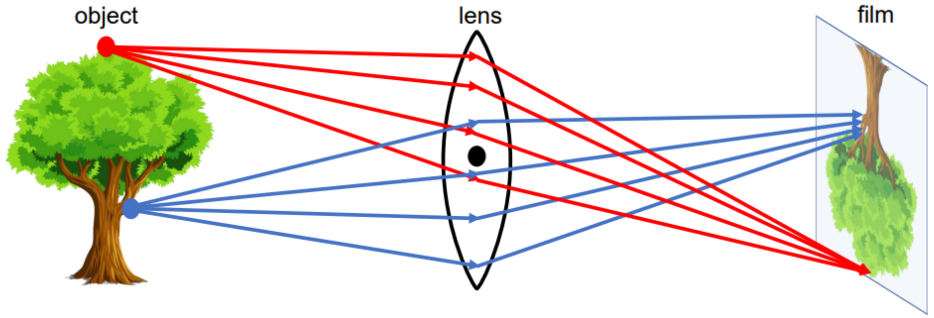
- A smaller aperture size causes less light rays to pass through, resulting in crisper but darker images.

- How can we solve this?



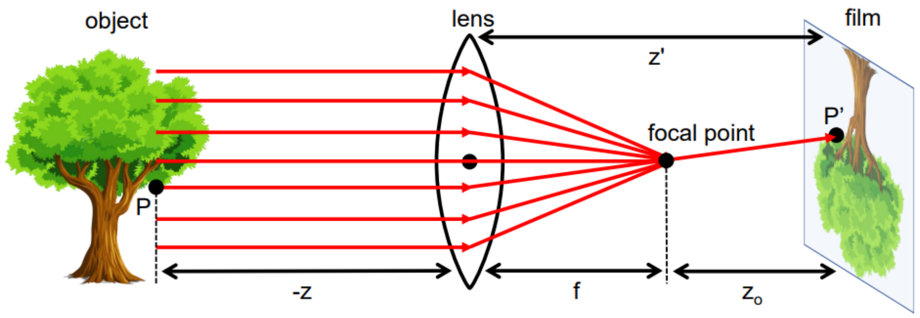
Cameras and lenses

- Lenses as a Solution:
  - Lenses mitigate this conflict by **focusing light** rather than simply blocking it.
  - A lens refracts (bends) light rays such that all rays emitted from a single point P in the 3D world converge to a single point P′ on the image plane.
  - This creates a sharp and bright image.



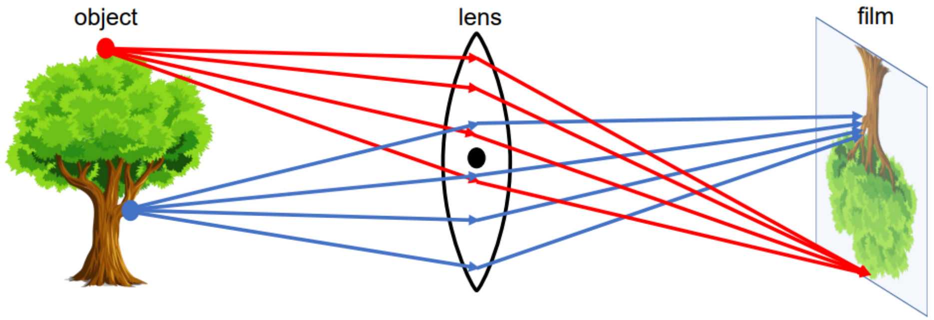
Cameras and lenses

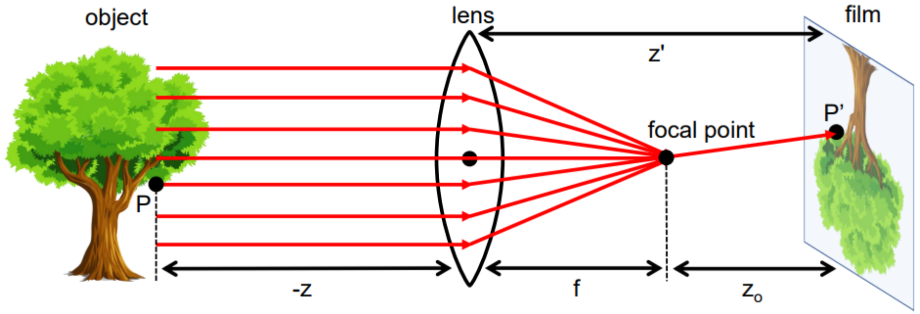
- Focal Point:
  - A lens focuses all light rays traveling parallel to the optical axis (the central axis of the lens) to a single point called the focal point.
- Focal Length:
  - The distance between the focal point and the optical center of the lens is called the focal length (f).



Camera lenses: properties

- Rays parallel to the optical axis converge at the focal point.
- Rays passing through the optical center of the lens are not deviated and continue in a straight line.
- All rays emitted from a point P in the 3D world are refracted by the lens and converge to a single point P′ on the image plane (if P is in focus).





Cameras and lenses: Problem with lenses

- **Focus**:
  - A lens can only perfectly focus light from points at a specific distance (the **focal plane**). For example, if the lens is focused on point P, then P will appear sharp in the image.

  - Points at other distances (e.g., Q, which is closer or farther than P) will not converge to a single point on the image plane. Instead, they form a **blur circle** (or **circle of confusion**), resulting in a blurred image.

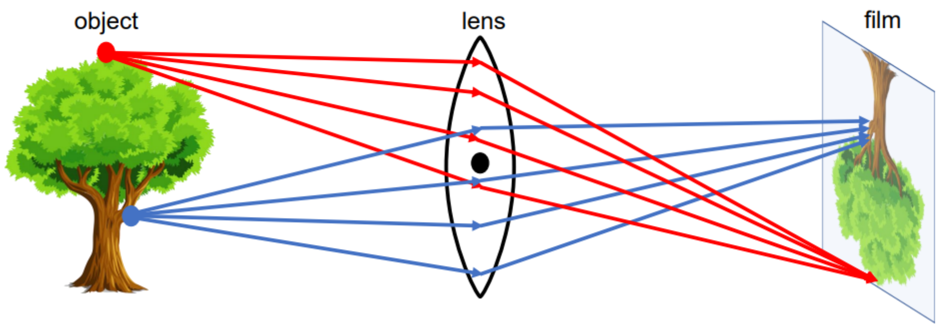


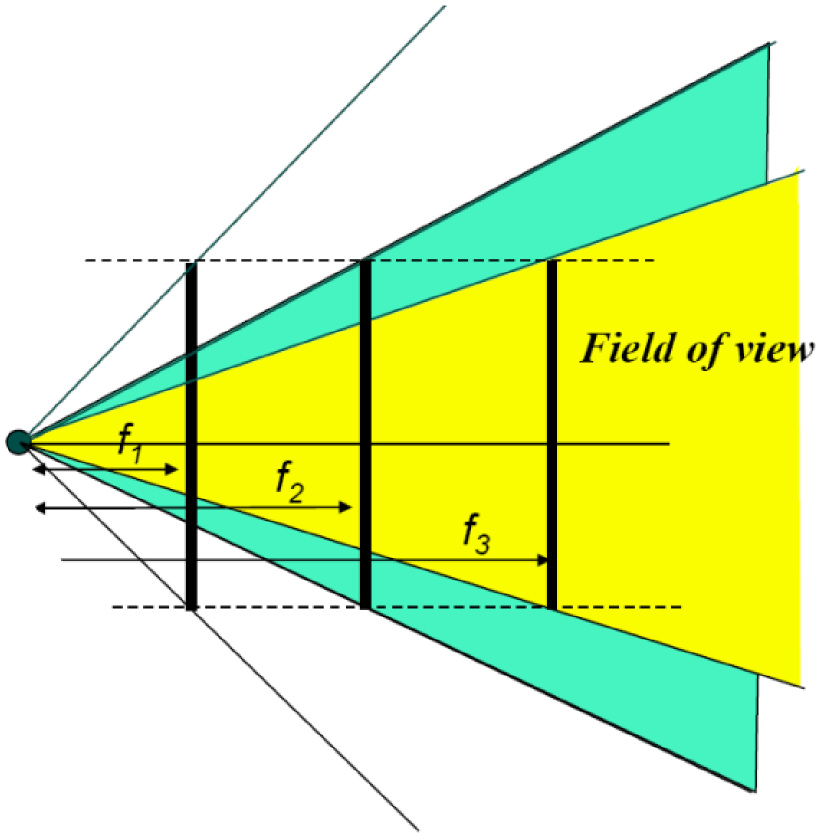
Figure: a point at a different distance away from the lens results in rays not converging perfectly on the film. E.g. The tree branch(blue point).

Cameras and lenses: Problem with lenses

- **Depth of Field**:
  - The range of distances over which objects appear acceptably sharp is called the **depth of field**.
  - Depth of field is influenced by factors such as aperture size, focal length, and distance to the subject.
  - A shallow depth of field means only a narrow range of distances is in focus,
    - Isolation; controlled by **wide apertures, long lenses, and close distances**.
  - A deep depth of field means a wider range of distances is in focus.
    - Sharpness throughout; controlled by **narrow apertures, wide lenses, and far distances**.

Cameras and lenses

- The focal length determines the lens's field of view and magnification:
  - A shorter focal length provides a wider field of view.
  - A longer focal length provides a narrower field of view (zoom).



Cameras and lenses

- Effect of focal length

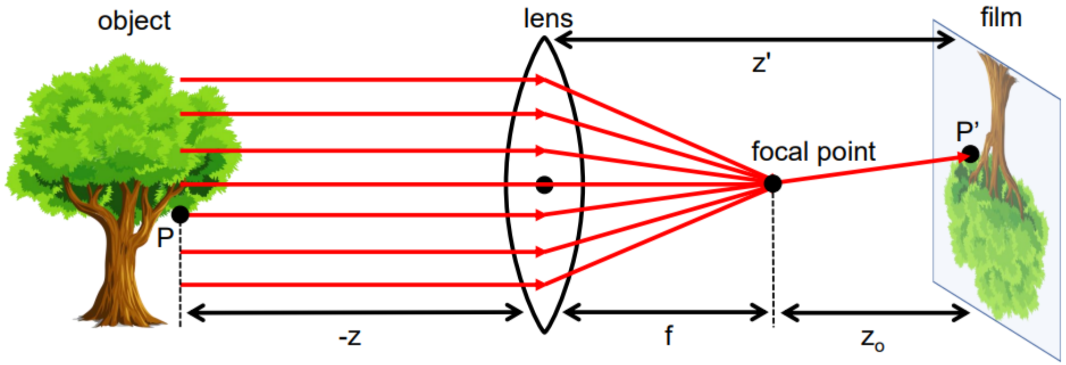
Field of view: portion of 3D space seen by the camera

Camera lenses: properties

- We thus can arrive at a similar construction to the pinhole model that relates a point P in 3D space with its corresponding point P' in the image plane

$$P' = \begin{bmatrix} x' \\ y' \end{bmatrix} = \begin{bmatrix} z' \frac{x}{z} \\ z' \frac{y}{z} \end{bmatrix} \qquad (2)$$

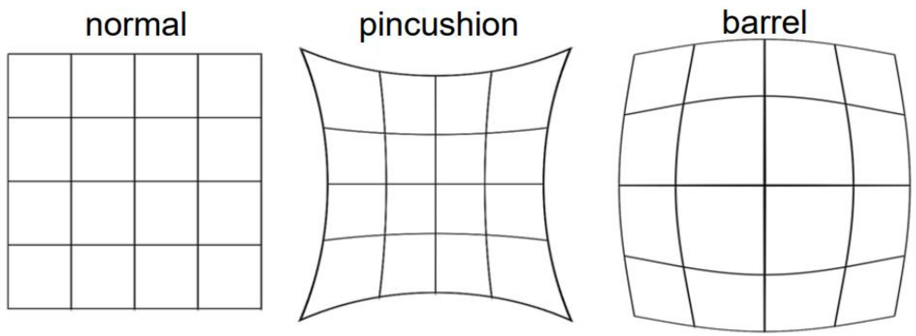
$$z' = f + z_0$$



- This derivation takes advantage of the paraxial or "thin lens" assumption, it is called the paraxial refraction model.

Please notice that in the pinhole model $z' = f$, while in this lens-based model $z' = f + z_0$

Paraxial refraction model



https://seanmichaelpritchard.com/2022/10/24/lens-distortion-an-essential-photography-guide/

https://www.lifeafterphotoshop.com/lens-aberrations-and-what-you-can-do-about-them/

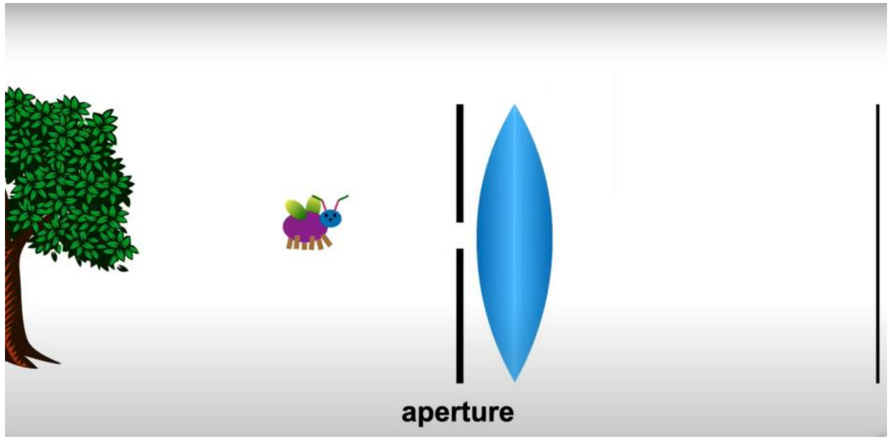
- The derivation takes advantage of the paraxial or "thin lens" assumption, it is called the paraxial refraction model.
- The paraxial refraction model approximates using the thin lens assumption, a number of aberrations can occur.
  - The most common one is referred to as radial distortion, which causes the image magnification to decrease or increase as a function of the distance to the optical axis.
    - Pincushion distortion when the magnification increases and
    - Barrel distortion when the magnification decrease

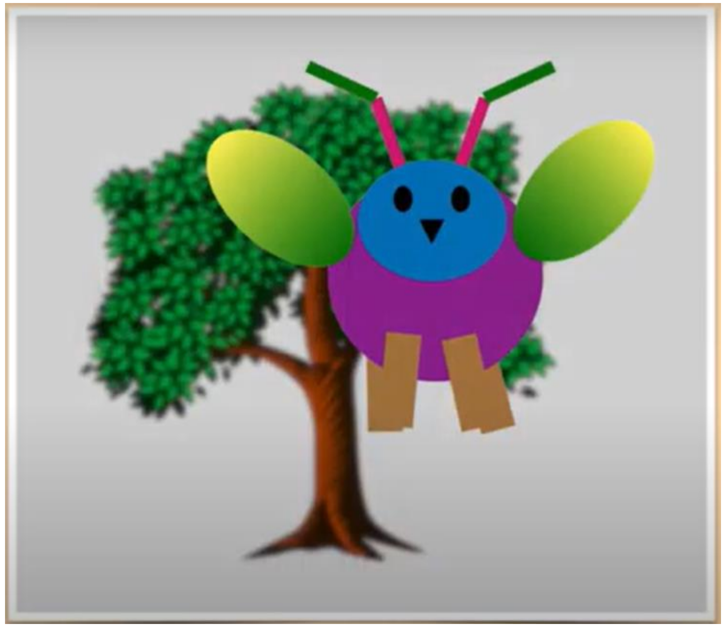
Perspective projection properties

- Many-to-One Mapping
  - Many-to-one mapping means that multiple points in 3D space can map to the same point in the 2D projection. This happens because perspective projection "flattens" the 3D world onto a 2D plane, losing depth information.

Perspective projection properties

- Scaling/Foreshortening
  - The distance to an object is inversely proportional to its image size.

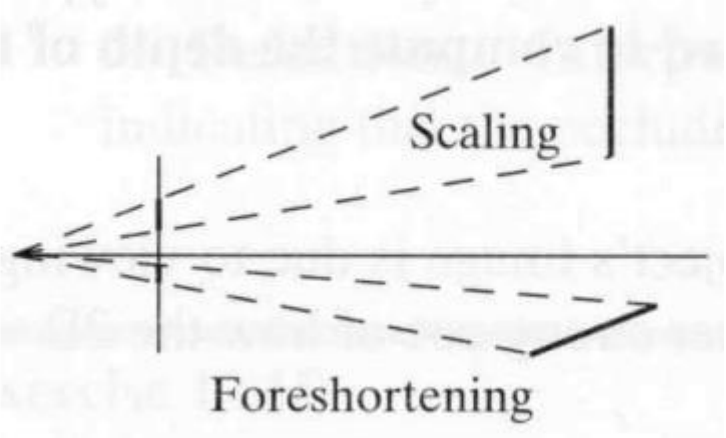




- Object further away from the camera appears smaller
- This creates a sense of depth in the 2D image.

Perspective projection properties

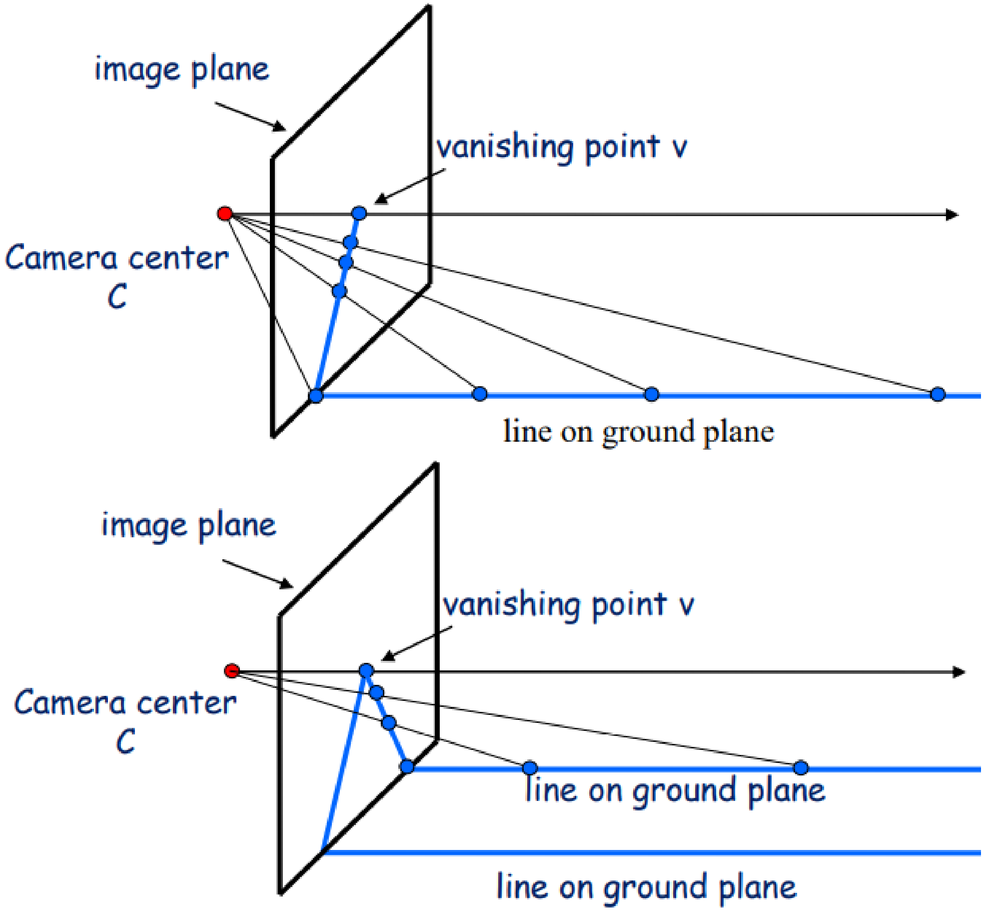
- Scaling/Foreshortening
  - When a line (or surface) is parallel to the image plane, the effect of perspective projection is scaling.
  - When a line (or surface) is not parallel to the image plane, we use the term foreshortening to describe the projective distortion (i.e., the dimension parallel to the optical axis is compressed relative to the frontal dimension).



Perspective projection properties

- Lines, distances, angles
  - Lines in 3D project to lines in 2D.
  - Distances and angles are not preserved.
  - Parallel lines do not in general project to parallel lines (unless they are parallel to the image plane).

Perspective projection properties



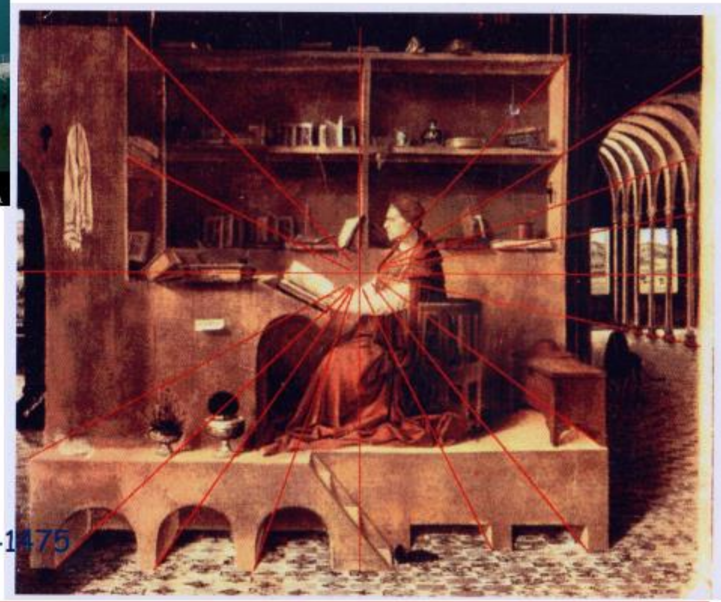
- Parallel lines in the scene intersect in the image
- Converge in image on horizon line

- An image may have more than one vanishing point

Perspective projection properties

- Perspective effect

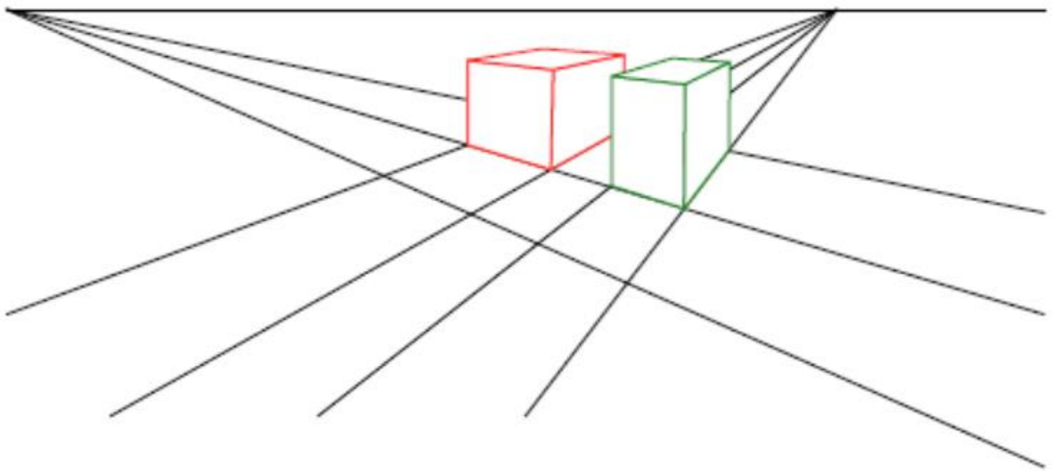
Image credit: S. Seitz



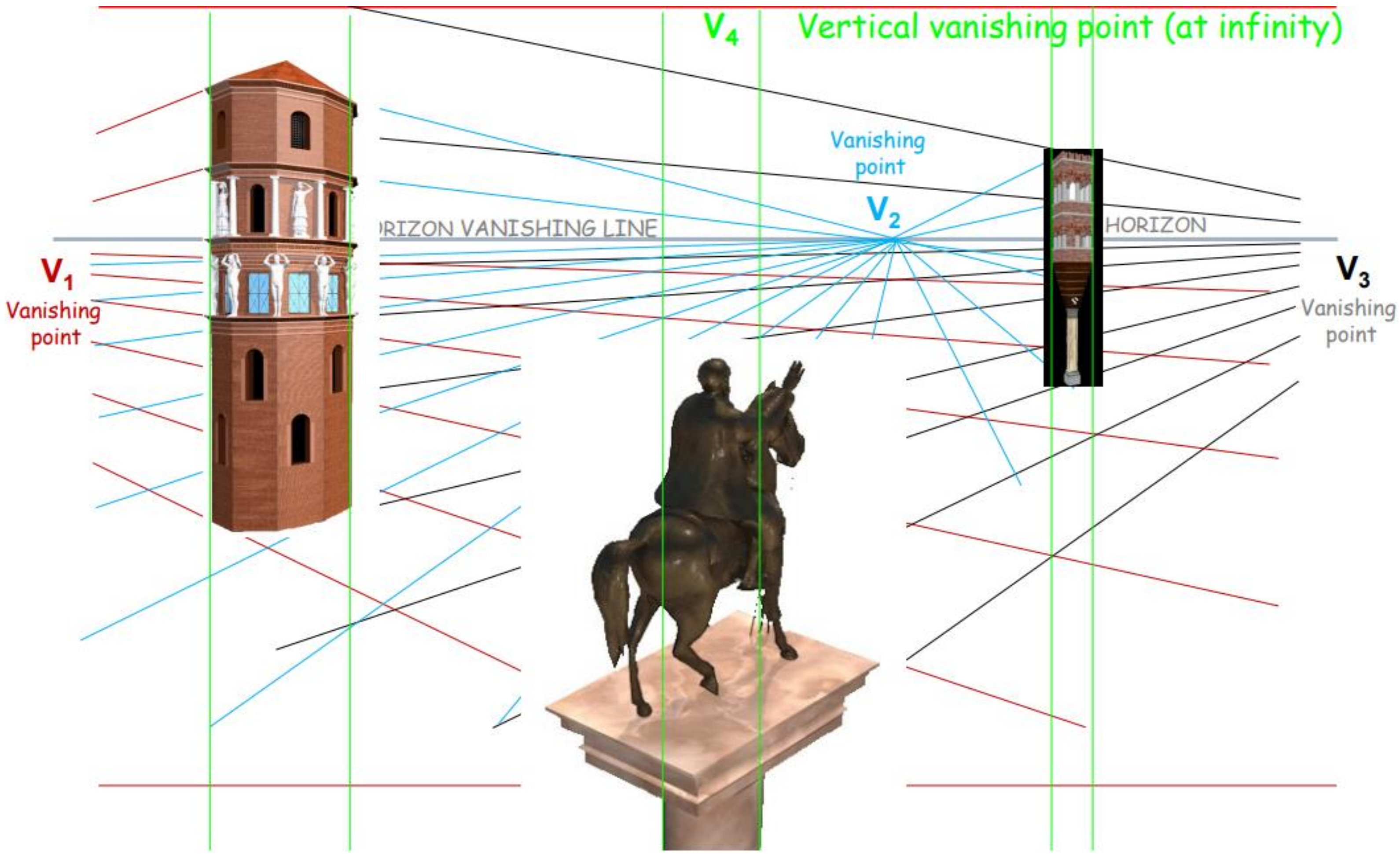
San Girolamo nello studio
Antonello da Messina, 1474-1475
London, National Gallery

Perspective projection properties

- Vanishing points:
  - Each set of parallel lines (=direction) meets at a different point
    - The vanishing point for this direction
  - Sets of parallel lines on the same plane lead to collinear vanishing points.
    - The line is called the horizon for that plane

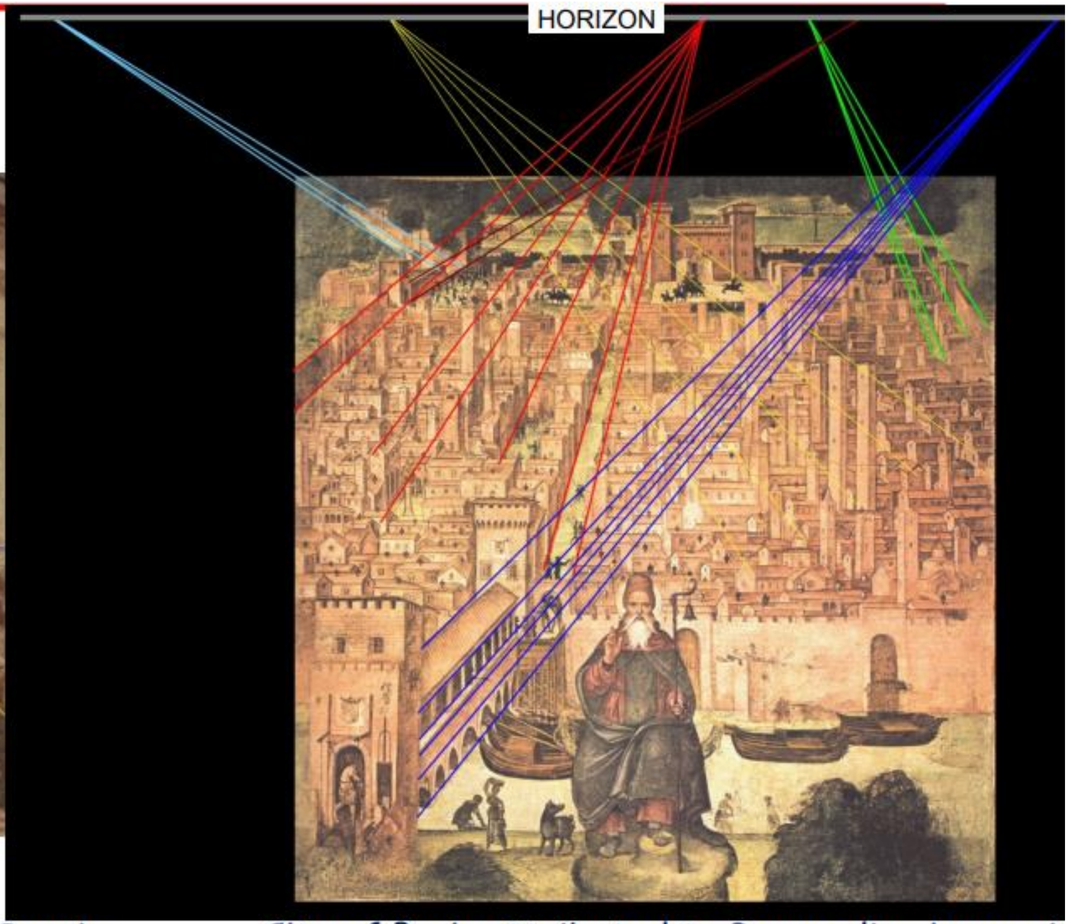


Perspective projection properties



Perspective projection properties

The School of Athens by Raffaello Sanzio which dates from 1508-1511



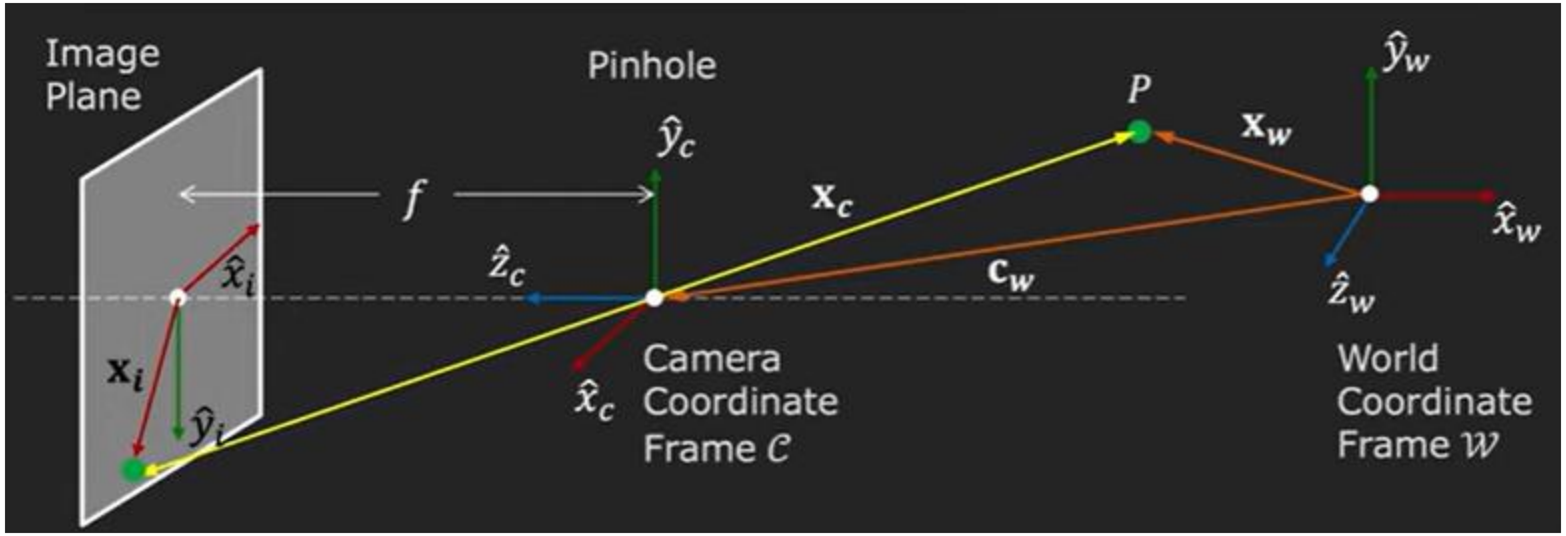
City of Pavia attributed to Bernardino Lanzani which dates from 1522

Properties of orthographic projection

- Parallel lines project to parallel lines.
- Size does not change with distance from the camera.

Going to digital image space

3D to 2D Mapping



Perspective Projection

3D to 3D Coordinate Transformation

$$p_i = \begin{bmatrix} x_i \\ y_i \end{bmatrix} \longleftarrow P_c = \begin{bmatrix} x_c \\ y_c \\ z_c \end{bmatrix} \longleftarrow P_w = \begin{bmatrix} x_w \\ y_w \\ z_w \end{bmatrix}$$

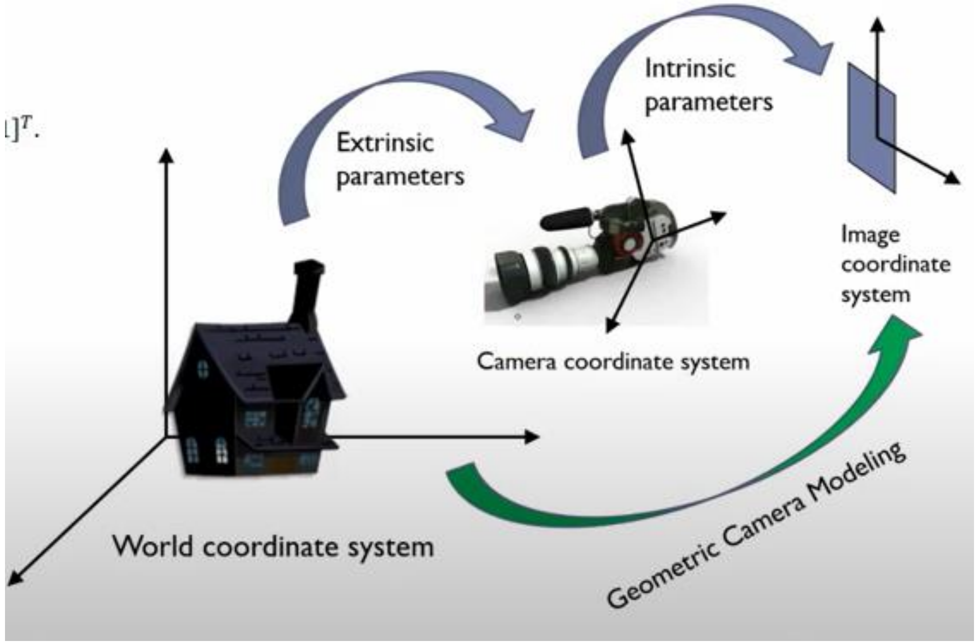
Image Coordinates

Camera Coordinates

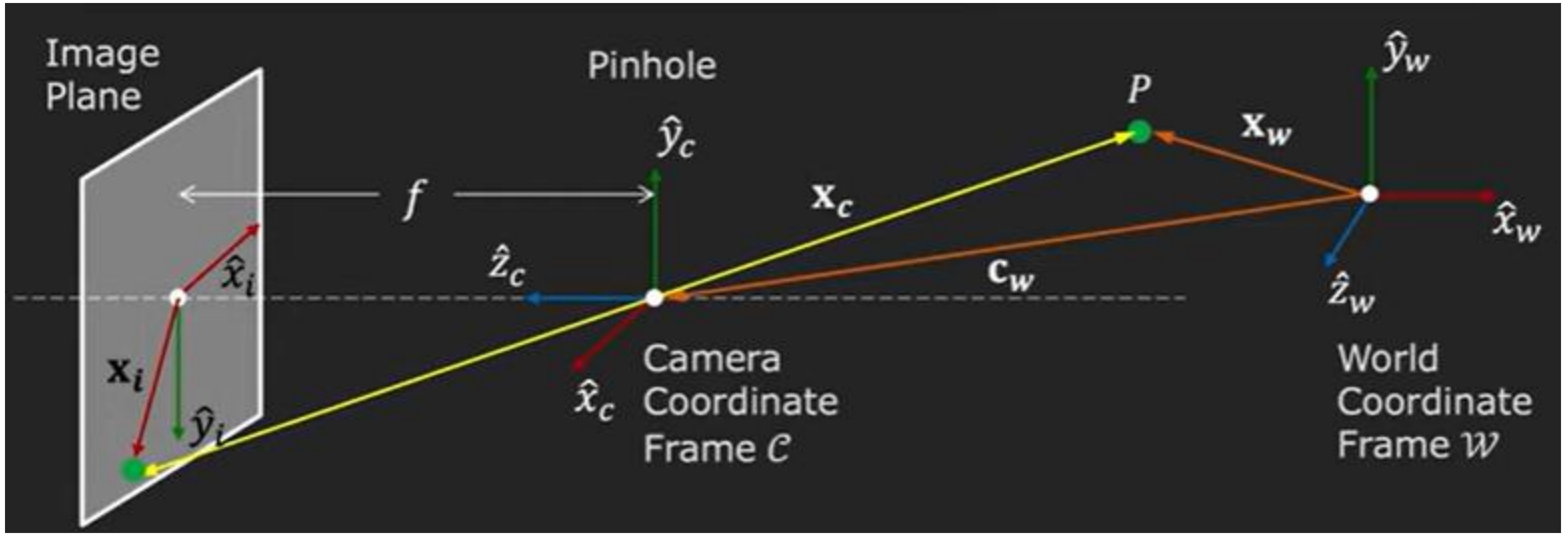
World Coordinates

Geometric Camera Modeling

1. Intrinsic parameters: Define the camera's internal geometry
2. Extrinsic parameters: Define the camera's position and orientation in the world.



3D to 2D Mapping



Perspective Projection

3D to 3D Coordinate Transformation

$$p_i = \begin{bmatrix} x_i \\ y_i \end{bmatrix} \longleftarrow P_c = \begin{bmatrix} x_c \\ y_c \\ z_c \end{bmatrix} \longleftarrow P_w = \begin{bmatrix} x_w \\ y_w \\ z_w \end{bmatrix}$$

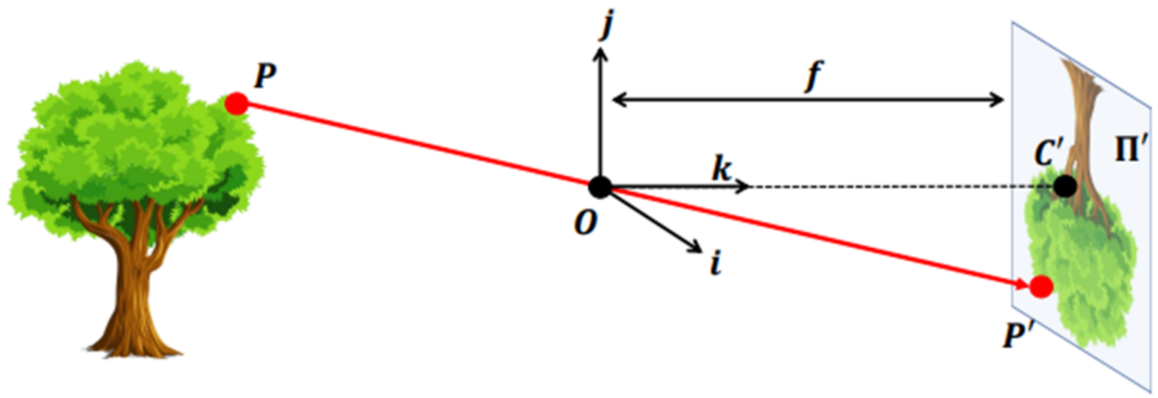
Image Coordinates

Camera Coordinates

World Coordinates

Going to digital image space

- A point P in 3D space mapped (or projected) into a 2D point P' in the image plane π′
- This $R^3 \rightarrow R^2$ mapping is referred to as a projective transformation.



Going to digital image space

- This projection of 3D points into the image plane does not directly correspond to what we see in actual digital images for the following reasons.
  1) Points in the digital images are in a different reference system than those in the image plane.
  2) Digital images are divided into discrete pixels, whereas points in the image plane are continuous.

Image Plane

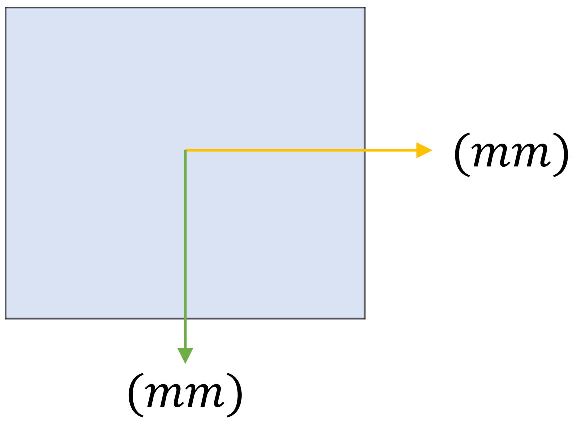
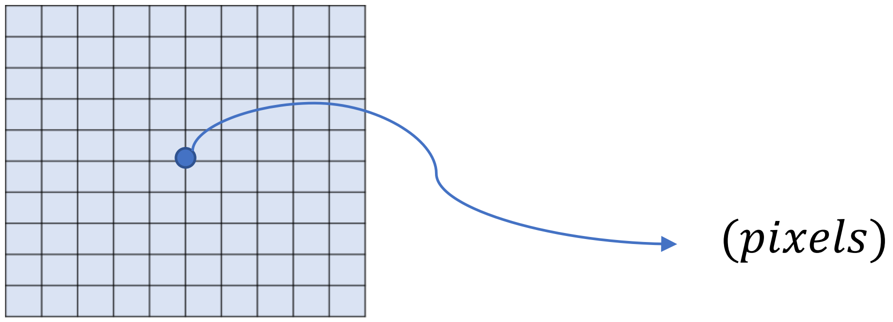
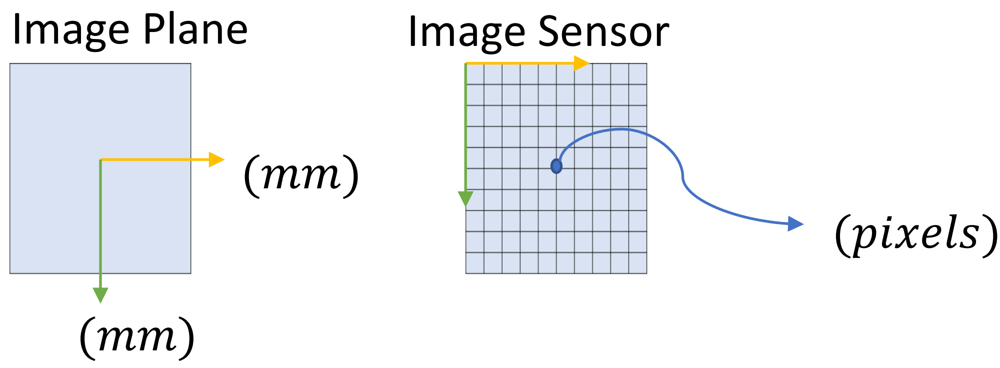


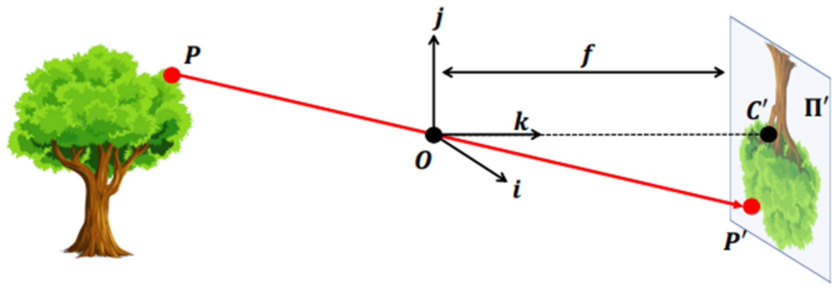
Image Sensor



The Camera Matrix Model and Homogeneous Coordinates

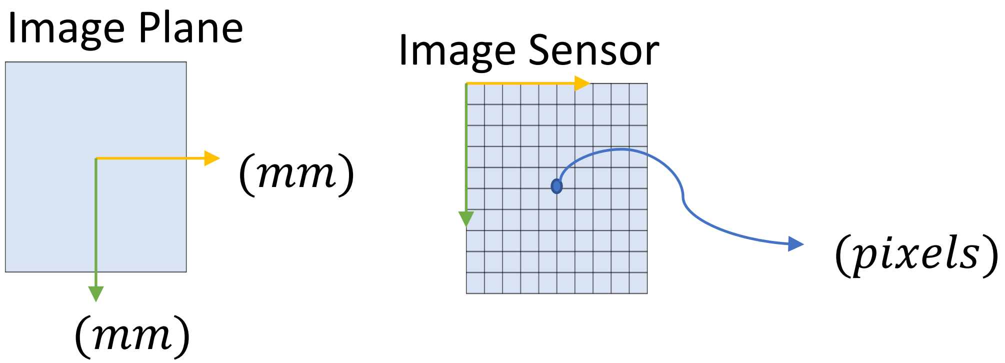
- Introduction to the Camera Matrix Model
  - Describes a set of important parameters that affect how a world point P is mapped to image coordinates P'.
- The first parameter:
  - The principal point, C_x *and* C_y, describe how image plane and digital image coordinates can differ by a translation.
    - Image plane coordinates have their origin C_0 at the image center where the k axis intersects the image plane.
    - Digital image coordinates typically have their origin at the upper-left corner of the image. Thus, 2D points in the image plane and 2D points in the image are offset by a translation vector $[C_x, C_y]^T$.





$$P' = \begin{bmatrix} x' \\ y' \end{bmatrix} = \begin{bmatrix} f\frac{x}{z} + c_x \\ f\frac{y}{z} + c_y \end{bmatrix} \quad (3)$$

The Camera Matrix Model and Homogeneous Coordinates



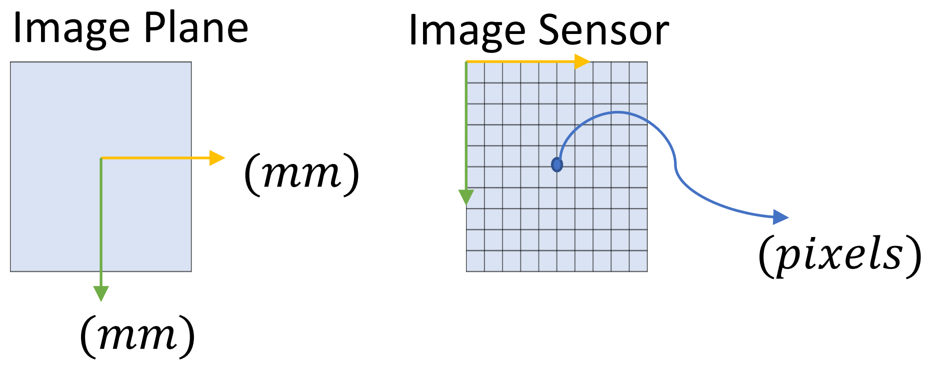
- Example: If the image center C' in digital image coordinates is at $(C_x, C_y) = (320{,}240)$ (for a (640×480) image), then:
  - A point $(x, y) = (10{,}20)$ in image plane coordinates corresponds to:
    - $x' = 10 + 320$
    - $y' = 20 + 240$

    in digital image coordinates.

$$P' = \begin{bmatrix} x' \\ y' \end{bmatrix} = \begin{bmatrix} f\frac{x}{z} + c_x \\ f\frac{y}{z} + c_y \end{bmatrix} \quad (3)$$

The Camera Matrix Model and Homogeneous Coordinates

- Introduction to the Camera Matrix Model
  - The second parameter,
    - Points in digital images are expressed in pixels, while points in image plane are represented in physical measurements (e.g. centimeters).
    - To accommodate this change of units, we must introduce two new parameters k and l.
    - These parameters, whose units would be something like $\frac{pixels}{mm}$, correspond to the change of units in the two axes of the image plane.
    - Note that k and l may be different because the aspect ratio of a pixel is not guaranteed to be one. If $k = l$, we often say that the camera has square pixels.



$$P' = \begin{bmatrix} x' \\ y' \end{bmatrix} = \begin{bmatrix} fk\frac{x}{z} + c_x \\ fl\frac{y}{z} + c_y \end{bmatrix} = \begin{bmatrix} \alpha\frac{x}{z} + c_x \\ \beta\frac{y}{z} + c_y \end{bmatrix} \tag{4}$$

The Camera Matrix Model and Homogeneous Coordinates

$$P' = \begin{bmatrix} x' \\ y' \end{bmatrix} = \begin{bmatrix} fk\frac{x}{z} + c_x \\ fl\frac{y}{z} + c_y \end{bmatrix} = \begin{bmatrix} \alpha\frac{x}{z} + c_x \\ \beta\frac{y}{z} + c_y \end{bmatrix} \quad (4)$$

- **Example:** Suppose:
  - $k = 200\frac{pixels}{mm}$ (horizontal scaling factor).
  - $l = 200\frac{pixels}{mm}$ (vertical scaling factor).
  - Principal point $(C_x, C_y) = (320{,}240)$ (for a (640×480) image.
  - A point in the image plane has coordinates $(C_x, C_y) = (0.5\, mm, 0.3mm)$
  - The corresponding digital image coordinates are:
    - $x' = k.x + C_x = 200 \times 0.5 + 300 = 100 + 320 = 420$
    - $y' = y' = k.y + C_y = 200 \times 0.3 + 240 = 60 + 240 = \mathit{300}$
- So, the point (0.5mm,0.3mm) in the image plane corresponds to the pixel (420,300) in the digital image.

The Camera Matrix Model and Homogeneous Coordinates

- One way to solve this problem is to change the Coordinate System.
- To convert a Euclidean vector $(v_1,\dots.\ v_n)$ to homogeneous coordinates, we simply append a 1 in a new dimension to get $(v_1,\dots.\ v_n, 1)$

For example:

- $P' = (\ x'\ ,\ y'\ )\ becomes\ \ (\ x'\ ,\ y'\ ,1)$
- $P\ = (x, y, z)\ becomes\ (x, y, z, 1)$

This augmented space is referred to as the homogeneous coordinate system.

- The equality between a vector and its homogeneous coordinates only occurs when the final coordinate equals one.
- When converting back from arbitrary homogeneous coordinates $(v_1,\dots.\ v_n, w)$ we get Euclidean coordinates $(\frac{v_1}{w}, \dots., \frac{v_1}{w})$.

The Camera Matrix Model and Homogeneous Coordinates

1) Rewrite the projection Equations (4)

$$x' = \alpha\frac{x}{z} + c_x \implies x'z = \alpha x + c_x z,$$

$$y' = \beta\frac{y}{z} + c_y \implies y'z = \beta y + c_y z.$$

Now, let:

$$x'_h = \alpha x + c_x z, \quad y'_h = \beta y + c_y z, \quad w = z.$$

This Gives:

$$P'_h = \begin{bmatrix} x'_h \\ y'_h \\ w \end{bmatrix} = \begin{bmatrix} \alpha x + c_x z \\ \beta y + c_y z \\ z \end{bmatrix}.$$

The Camera Matrix Model and Homogeneous Coordinates

2. Express as a Matrix-Vector Product

We want to write P_h' as $MP_h = (x, y, z, 1)$. To do this, we construct M such that:

$$\begin{bmatrix} x_h' \\ y_h' \\ w \end{bmatrix} = \begin{bmatrix} \text{Row 1} \\ \text{Row 2} \\ \text{Row 3} \end{bmatrix} \begin{bmatrix} x \\ y \\ z \\ 1 \end{bmatrix}.$$

The Camera Matrix Model and Homogeneous Coordinates

Using homogeneous coordinates, we can formulate

$$P'_h = \begin{bmatrix} \alpha x + c_x z \\ \beta y + c_y z \\ z \end{bmatrix} = \begin{bmatrix} \alpha & 0 & c_x & 0 \\ 0 & \beta & c_y & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \\ 1 \end{bmatrix} = \begin{bmatrix} \alpha & 0 & c_x & 0 \\ 0 & \beta & c_y & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} P_h \qquad (5)$$

Drop the h index, so any point $P \; or \; P'$ can be assumed to be in homogeneous coordinates

$$P' = \begin{bmatrix} x' \\ y' \\ z \end{bmatrix} = \begin{bmatrix} \alpha & 0 & c_x & 0 \\ 0 & \beta & c_y & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \\ 1 \end{bmatrix} = \begin{bmatrix} \alpha & 0 & c_x & 0 \\ 0 & \beta & c_y & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} P = MP \qquad (6)$$

We can decompose this transformation a bit further into

$$P' = MP = \begin{bmatrix} \alpha & 0 & c_x \\ 0 & \beta & c_y \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} I & 0 \end{bmatrix} P = K \begin{bmatrix} I & 0 \end{bmatrix} P \qquad (7)$$

The matrix K is often referred to as the camera matrix.

The Camera Matrix Model and Homogeneous Coordinates

- The Complete Camera Matrix Model:
  - The camera matrix K contains some of the critical parameters that describes a camera's characteristics and its model, including the α, β, c_x, c_y
  - Two parameters are currently missing this formulation: skewness and distortion.
  - We often say that an image is skewed when the camera coordinate system is skewed, meaning that the angle between the two axes is slightly larger or smaller than 90 degrees.
  - Most cameras have zero-skew, but some degree of skewness may occur because of sensor manufacturing errors.

$$K = \begin{bmatrix} \alpha & -\alpha \cot\theta & c_x \\ 0 & \frac{\beta}{\sin\theta} & c_y \\ 0 & 0 & 1 \end{bmatrix} \tag{8}$$

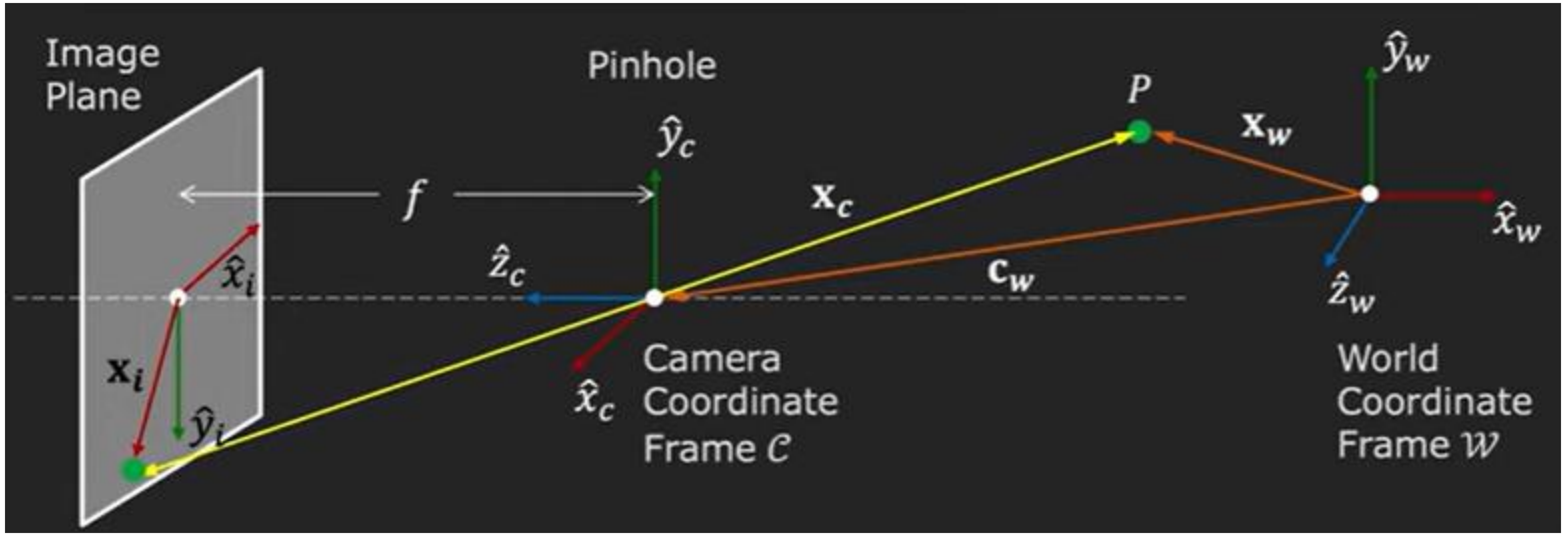
The Camera Matrix Model and Homogeneous Coordinates

- The Complete Camera Matrix Model:
  - Most methods that we introduce in this class ignore distortion effects, therefore our class camera matrix K has 5 degrees of freedom:
    - 2 for focal length, 2 for offset, and 1 for skewness.

$$K = \begin{bmatrix} \alpha & -\alpha \cot\theta & c_x \\ 0 & \frac{\beta}{\sin\theta} & c_y \\ 0 & 0 & 1 \end{bmatrix} \tag{8}$$

These parameters are collectively known as the intrinsic parameters.

The Camera Matrix Model and Homogeneous Coordinates (Extrinsic Parameters)



Perspective Projection

3D to 3D Coordinate Transformation

$$p_i = \begin{bmatrix} x_i \\ y_i \end{bmatrix} \longleftarrow P_c = \begin{bmatrix} x_c \\ y_c \\ z_c \end{bmatrix} \longleftarrow P_w = \begin{bmatrix} x_w \\ y_w \\ z_w \end{bmatrix}$$

Image Coordinates

Camera Coordinates

World Coordinates

Cont. …

- Extrinsic Parameters:
  - we need to include an additional transformation that relates points from the world reference system to the camera reference system.
  - This transformation is captured by a rotation matrix R and translation vector T.
    - The camera's position in the world is described by a **translation vector** $T = [tx, ty, tz]$.
      - This tells you where the camera is located in the world coordinate system.
    - The camera's orientation is described by a **rotation matrix R** , which is a 3x3 matrix that encodes the camera's orientation (e.g., roll, pitch, and yaw).
      - This tells you how the camera is rotated relative to the world coordinate system.

Cont. …

- Extrinsic Parameters:
  - given a point in a world reference system p_W,

$$P = \begin{bmatrix} R & T \\ 0 & 1 \end{bmatrix} P_w \tag{9}$$

These parameters R and T are known as the extrinsic parameters because they are external to and do not depend on the camera.

Substituting Eq. 9 in equation (7) and simplifying gives

$$P' = MP = \begin{bmatrix} \alpha & 0 & c_x \\ 0 & \beta & c_y \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} I & 0 \end{bmatrix} P = K \begin{bmatrix} I & 0 \end{bmatrix} P \tag{7}$$

$$P' = K \begin{bmatrix} R & T \end{bmatrix} P_w = MP_w \tag{10}$$

The Camera Matrix Model and Homogeneous Coordinates

$$P' = K \begin{bmatrix} R & T \end{bmatrix} P_w = M P_w \tag{10}$$

- This completes the mapping from a 3D point P in an arbitrary world reference system to the image plane.
- To reiterate, we see that the full projection matrix M consists of the two types of parameters introduced above: Intrinsic and extrinsic parameters.
  - Intrinsic:
    - All parameters contained in the camera matrix K are the intrinsic parameters, which change as the type of camera changes.
    - These define the internal properties of the camera, such as focal length, principal point, and pixel aspect ratio.
  - Extrinsic
    - The extrinsic parameters include the rotation and translation, which do not depend on the camera's build.
    - Overall, we find that the 3 × 4 projection matrix M has 11 degrees of freedom:
      - 5 from the intrinsic camera matrix, 3 from extrinsic rotation, and 3 from extrinsic translation.

Where does Camera Model Leads?

- It also leads into camera calibration.
- We need it to understand stereo and 3D reconstruction.

Camera Calibration

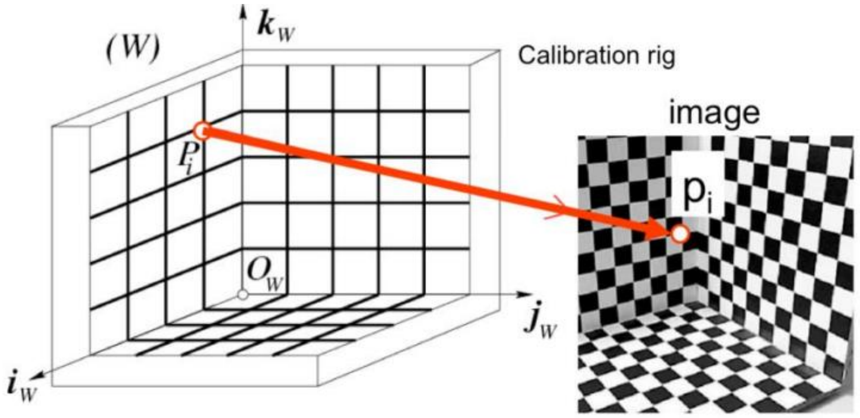
Camera Calibration

- If given an arbitrary camera, we may or may not have access to these parameters. We do, however, have access to the images the camera takes. Therefore, can we find a way to deduce them from images?
- This problem of estimating the extrinsic and intrinsic camera parameters is known as camera calibration.

Camera Calibration

- We do this by solving for the intrinsic camera matrix K and the extrinsic parameters R, T from Equation 10. We can describe this problem in the context of a calibration rig, such as the one show in Figure.
- The rig usually consists of a simple pattern (i.e. checkerboard) with known dimensions.
- Furthermore, the rig defines our world reference frame with origin O_w and axes $i_w, j_w k_w$.
  - From the rig's known pattern, we have known points in the world reference frame $P_1, P_2, \ldots, P_n$.
  - Finding these points in the image we take from the camera gives corresponding points in the image $P_1, P_2, \ldots, P_n$.

$$P' = K \begin{bmatrix} R & T \end{bmatrix} P_w = M P_w \tag{10}$$



Camera Calibration

- We set up a linear system of equations from n correspondences such that for each correspondence P_i, p_i and camera matrix M whose rows are m_1, m_2, m_3 :

$$p_i = \begin{bmatrix} u_i \\ v_i \end{bmatrix} = MP_i = \begin{bmatrix} \frac{m_1 P_i}{m_3 P_i} \\ \frac{m_2 P_i}{m_3 P_i} \end{bmatrix} \tag{11}$$

As we see from the above equation, each correspondence gives us two equations and, consequently, two constraints for solving the unknown parameters contained in m.

$$u_i(m_3 P_i) - m_1 P_i = 0$$
$$v_i(m_3 P_i) - m_2 P_i = 0$$

Camera Calibration

- From before, we know that the camera matrix has 11 unknown parameters. This means that we need at least 6 correspondences to solve this. However, in the real world, we often use more, as our measurements are often noisy.

$$
\begin{aligned}
u_1(m_3P_1) - m_1P_1 &= 0 \\
v_1(m_3P_1) - m_2P_1 &= 0 \\
&\vdots \\
u_n(m_3P_n) - m_1P_n &= 0 \\
v_n(m_3P_n) - m_2P_n &= 0
\end{aligned}
$$

This can be formatted as a matrix-vector product shown below:

$$
\begin{bmatrix} P_1^T & 0^T & -u_1P_1^T \\ 0^T & P_1^T & -v_1P_1^T \\ & \vdots & \\ P_n^T & 0^T & -u_nP_n^T \\ 0^T & P_n^T & -v_nP_n^T \end{bmatrix} \begin{bmatrix} m_1^T \\ m_2^T \\ m_3^T \end{bmatrix} = \mathbf{P}m = 0 \qquad (12)
$$

Camera Calibration

- Therefore, to constrain our solution, we complete the following minimization:

$$\begin{aligned} \underset{m}{\text{minimize}} \quad & \|\mathbf{P}m\|^2 \\ \text{subject to} \quad & \|m\|^2 = 1 \end{aligned} \tag{13}$$

- To solve this minimization problem, we simply use singular value decomposition.

Camera Calibration

Here, r_1^T, r_2^T, and r_3^T are the three rows of R. Dividing by the scaling parameter gives

$$M = \frac{1}{\rho}\begin{bmatrix} \alpha r_1^T - \alpha\cot\theta r_2^T + c_x r_3^T & \alpha t_x - \alpha\cot\theta t_y + c_x t_z \\ \frac{\beta}{\sin\theta} r_2^T + c_y r_3^T & \frac{\beta}{\sin\theta} t_y + c_y t_z \\ r_3^T & t_z \end{bmatrix} = \begin{bmatrix} A & b \end{bmatrix} = \begin{bmatrix} a_1^T \\ a_2^T \\ a_3^T \end{bmatrix} \begin{bmatrix} b_1 \\ b_2 \\ b_3 \end{bmatrix}$$

Solving for the intrinsics gives

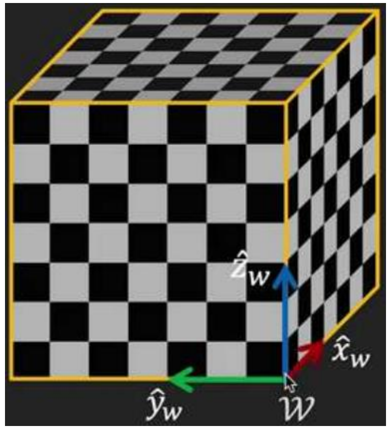
$$\begin{aligned} \rho &= \pm\frac{1}{\|a_3\|} \\ c_x &= \rho^2(a_1 \cdot a_3) \\ c_y &= \rho^2(a_2 \cdot a_3) \\ \theta &= \cos^{-1}\left(-\frac{(a_1 \times a_3)\cdot(a_2 \times a_3)}{\|a_1 \times a_3\| \cdot \|a_2 \times a_3\|}\right) \\ \alpha &= \rho^2\|a_1 \times a_3\|\sin\theta \\ \beta &= \rho^2\|a_2 \times a_3\|\sin\theta \end{aligned} \tag{15}$$

The extrinsics are

$$\begin{aligned} r_1 &= \frac{a_2 \times a_3}{\|a_2 \times a_3\|} \\ r_2 &= r_3 \times r_1 \\ r_3 &= \rho a_3 \\ T &= \rho K^{-1} b \end{aligned} \tag{16}$$

Camera Calibration Procedure

Step 1: Capture an image of an object with known geometry.

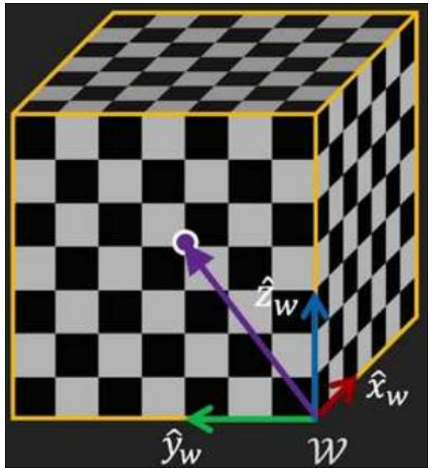


Object of known Geometry

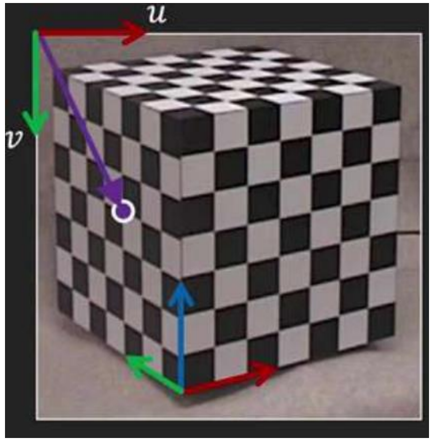
(1) a 3D object of known geometry.
(2) it is located in a known position in space.
(3) it is generating image features which can be located accurately.

Camera Calibration Procedure

Step 2: Identify correspondences between 3D scene points and image points



Object of known Geometry



Capture Image

- Identify correspondences :
  - Corner detection Algorithm:
  - Feature Detection and Matching
  - Manually

$$X_x = \begin{bmatrix} x_w \\ y_w \\ Z_w \end{bmatrix} = \begin{bmatrix} 0 \\ 3 \\ 4 \end{bmatrix}$$

(inches)

$$\text{U} = \begin{bmatrix} u \\ v \end{bmatrix} = \begin{bmatrix} 56 \\ 115 \end{bmatrix}$$

(pixels)

Camera Calibration Procedure

Step 3: Use the formula to calculate both intrinsic and extrinsic parameter

Here, r_1^T, r_2^T, and r_3^T are the three rows of R. Dividing by the scaling parameter gives

$$M = \frac{1}{\rho}\begin{bmatrix} \alpha r_1^T - \alpha \cot\theta r_2^T + c_x r_3^T & \alpha t_x - \alpha \cot\theta t_y + c_x t_z \\ \frac{\beta}{\sin\theta} r_2^T + c_y r_3^T & \frac{\beta}{\sin\theta} t_y + c_y t_z \\ r_3^T & t_z \end{bmatrix} = \begin{bmatrix} A & b \end{bmatrix} = \begin{bmatrix} a_1^T \\ a_2^T \\ a_3^T \end{bmatrix} \begin{bmatrix} b_1 \\ b_2 \\ b_3 \end{bmatrix}$$

Solving for the intrinsics gives

$$\begin{aligned} \rho &= \pm\frac{1}{\|a_3\|} \\ c_x &= \rho^2(a_1 \cdot a_3) \\ c_y &= \rho^2(a_2 \cdot a_3) \\ \theta &= \cos^{-1}\left(-\frac{(a_1 \times a_3)\cdot(a_2 \times a_3)}{\|a_1 \times a_3\| \cdot \|a_2 \times a_3\|}\right) \\ \alpha &= \rho^2\|a_1 \times a_3\| \sin\theta \\ \beta &= \rho^2\|a_2 \times a_3\| \sin\theta \end{aligned} \tag{15}$$

The extrinsics are

$$\begin{aligned} r_1 &= \frac{a_2 \times a_3}{\|a_2 \times a_3\|} \\ r_2 &= r_3 \times r_1 \\ r_3 &= \rho a_3 \\ T &= \rho K^{-1} b \end{aligned} \tag{16}$$

Why Camera Calibration

- You can use these parameters to
  - Correct for lens distortion,
  - Measure the size of an object in world units, or
  - Determine the location of the camera in the scene
  - 3D reconstruction

Recap

- Camera modeling
- Camera Calibration

Reference

- Szeliski 2.1 and especially 2.1.4 cover the geometry of image formation
- Additional reading:
  - Hartley and Zisserman, “Multiple View Geometry in Computer Vision,” Cambridge University Press 2004.
    - Chapter 6 of this book has a very thorough treatment of camera models.
  - Richter-Gebert, “Perspectives on Projective Geometry,” Springer 2011.
    - A great math textbook on everything to do with projective geometry.
  - Gortler, “Foundations of 3D Computer Graphics,” MIT Press 2012.
    - Chapter 10 of this book has a nice discussion of pinhole cameras from a graphics point of view.
  - Zhang, “A flexible new technique for camera calibration,” PAMI 2000.
    - The paper that introduced camera calibration from multiple views of a planar target.
  - Yu and McMillan, “General Linear Cameras,” ECCV 2004.
    - This paper presents a very general model and classification of linear cameras.

Next
Stereo Vision