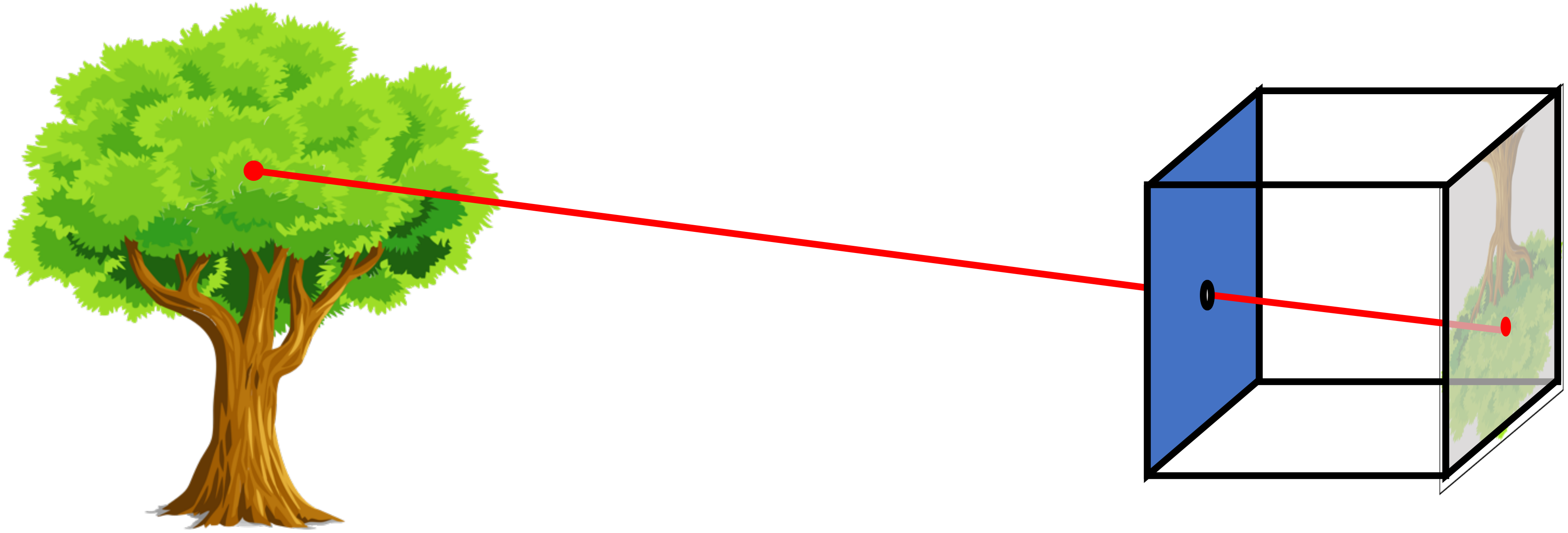
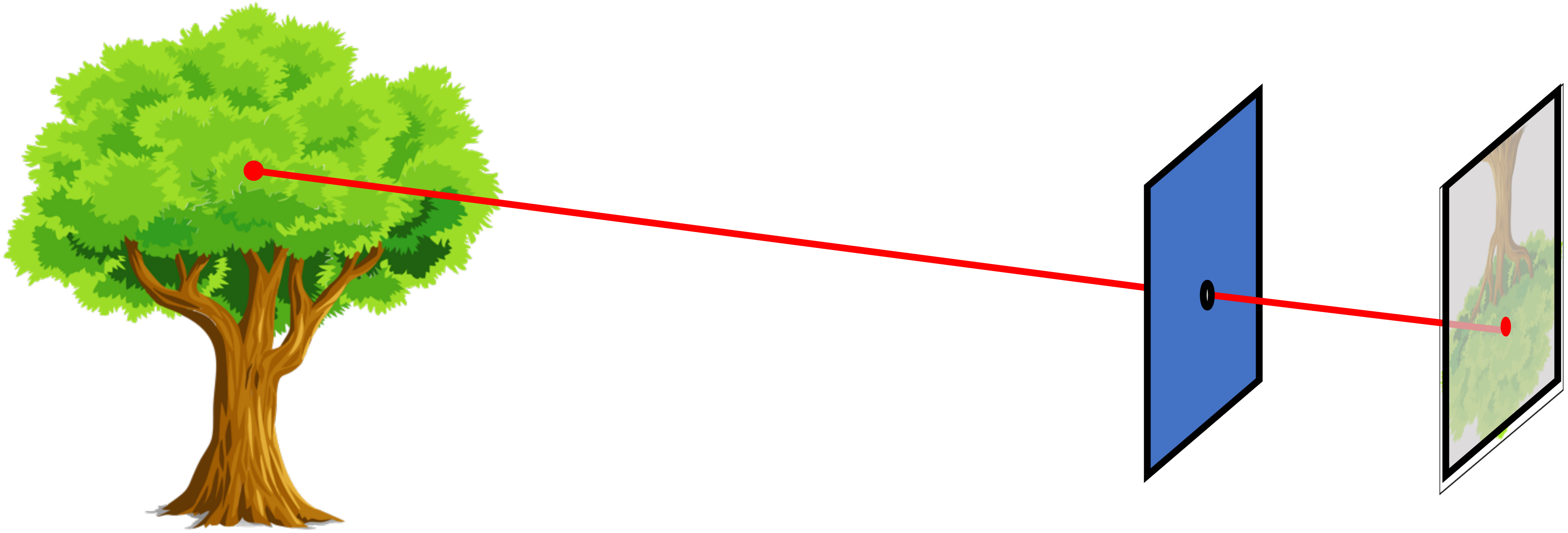
Geometry of Image Formation

The pinhole camera



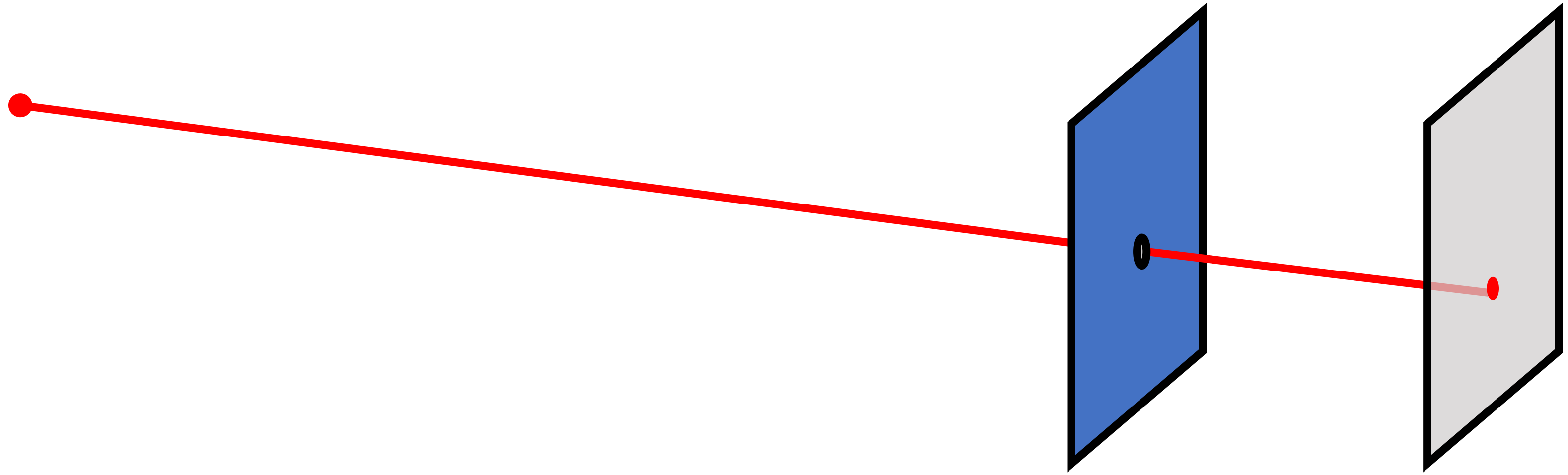
- Let’s abstract out the details

The pinhole camera



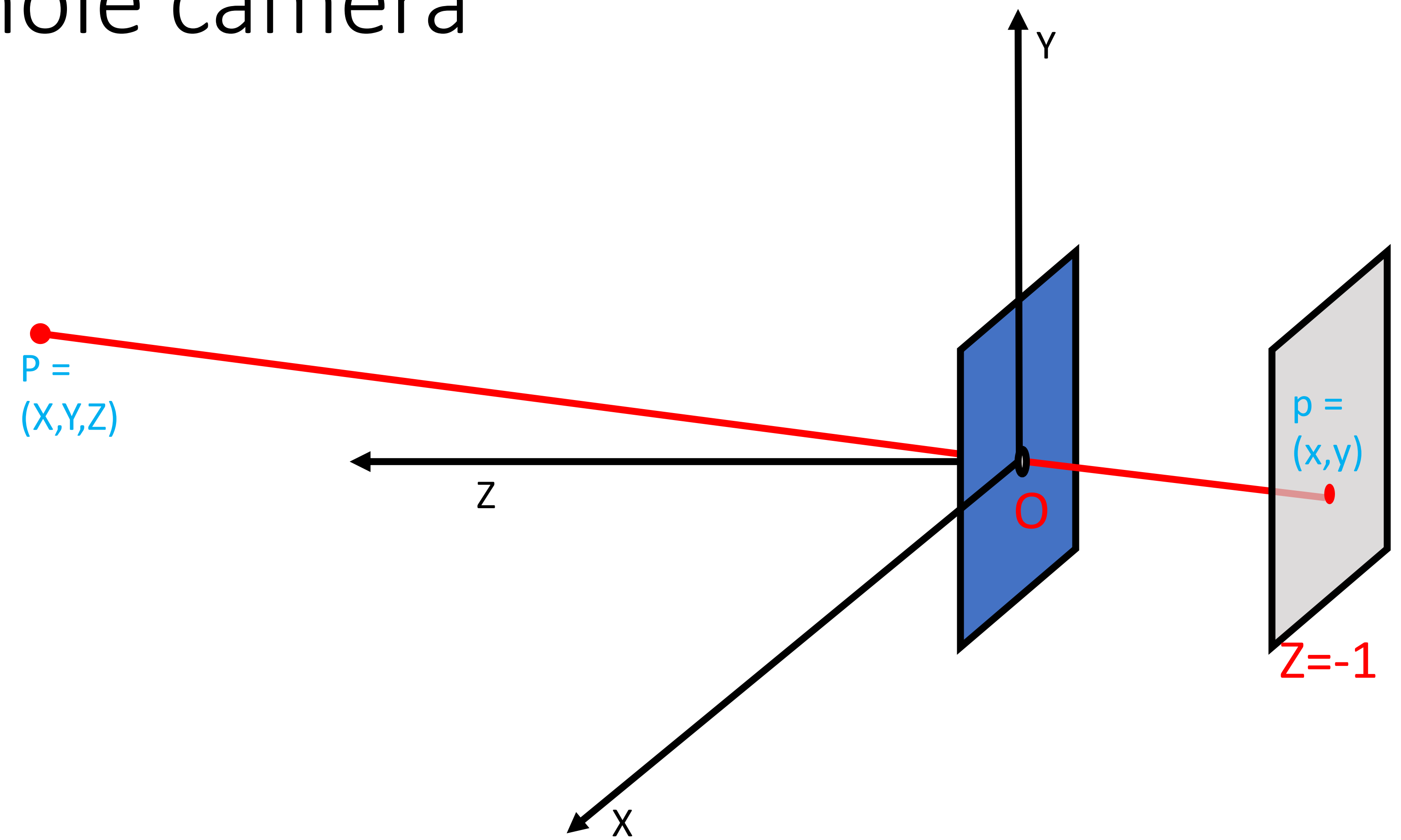
- We don’t care about the other walls of the box, so let’s remove those

The pinhole camera



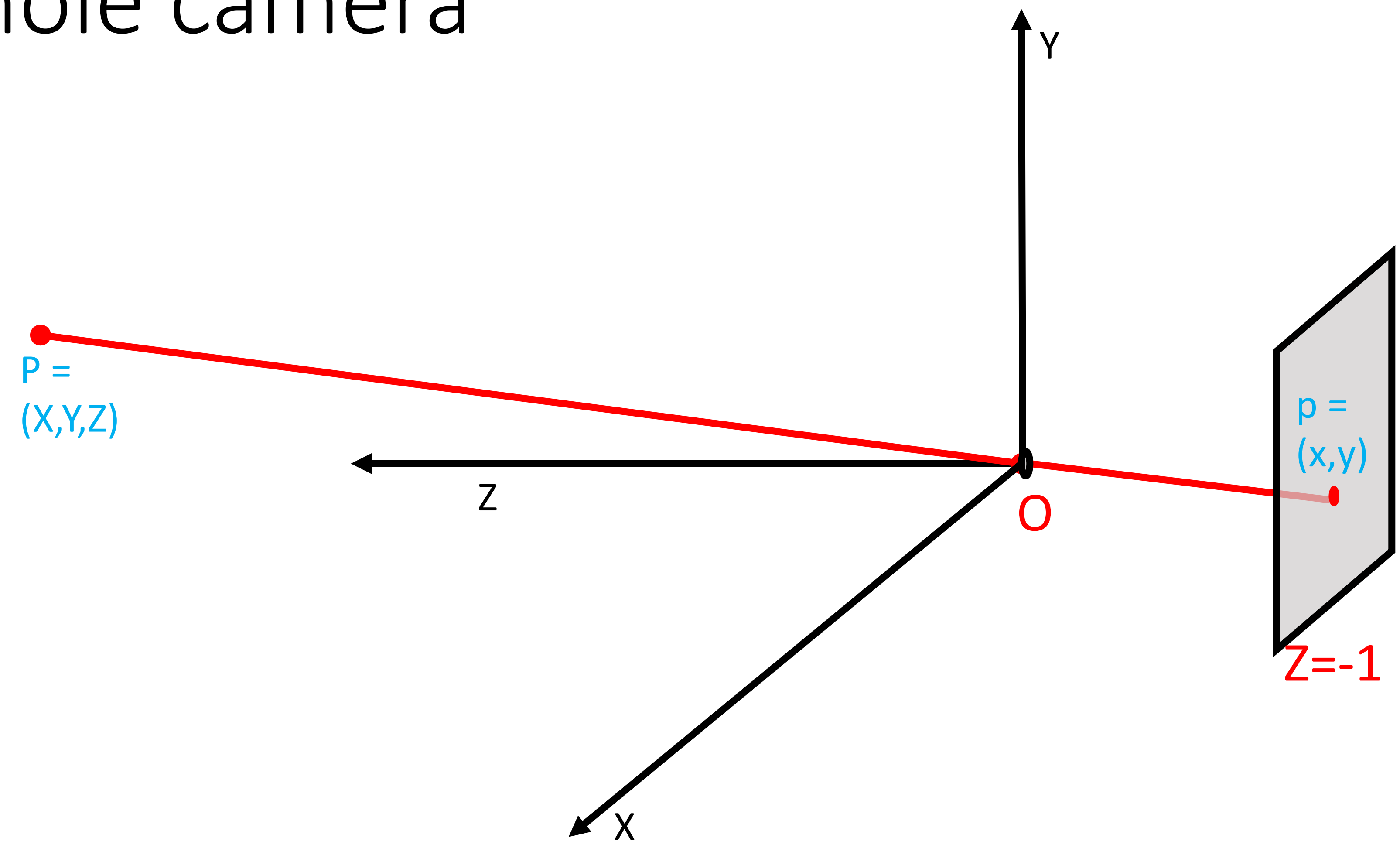
- Let’s look at a individual points in the world and not worry about what they are.

The pinhole camera



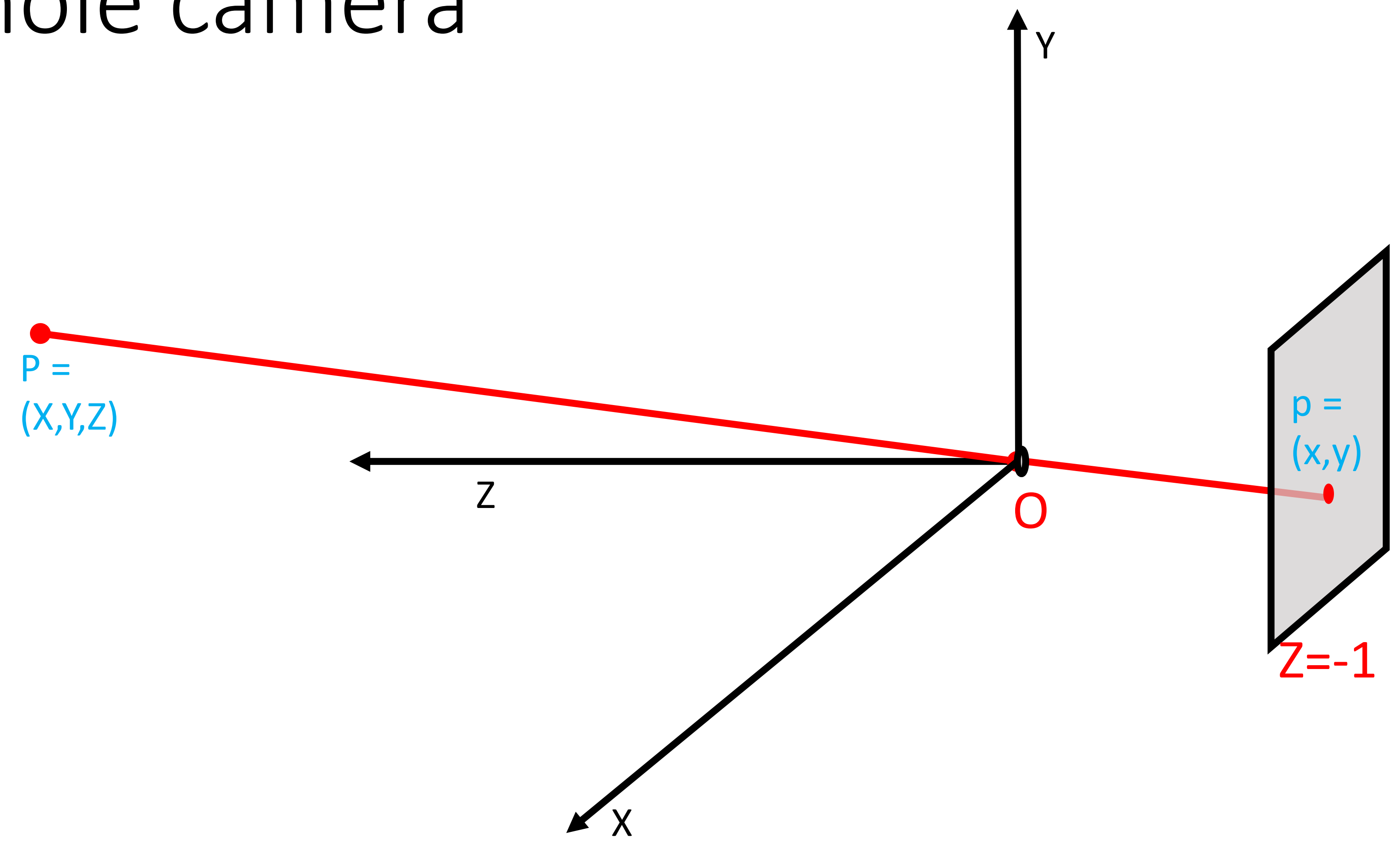
- Let’s place the origin at the pinhole, with Z axis pointing away from the screen (called *camera plane*)

The pinhole camera



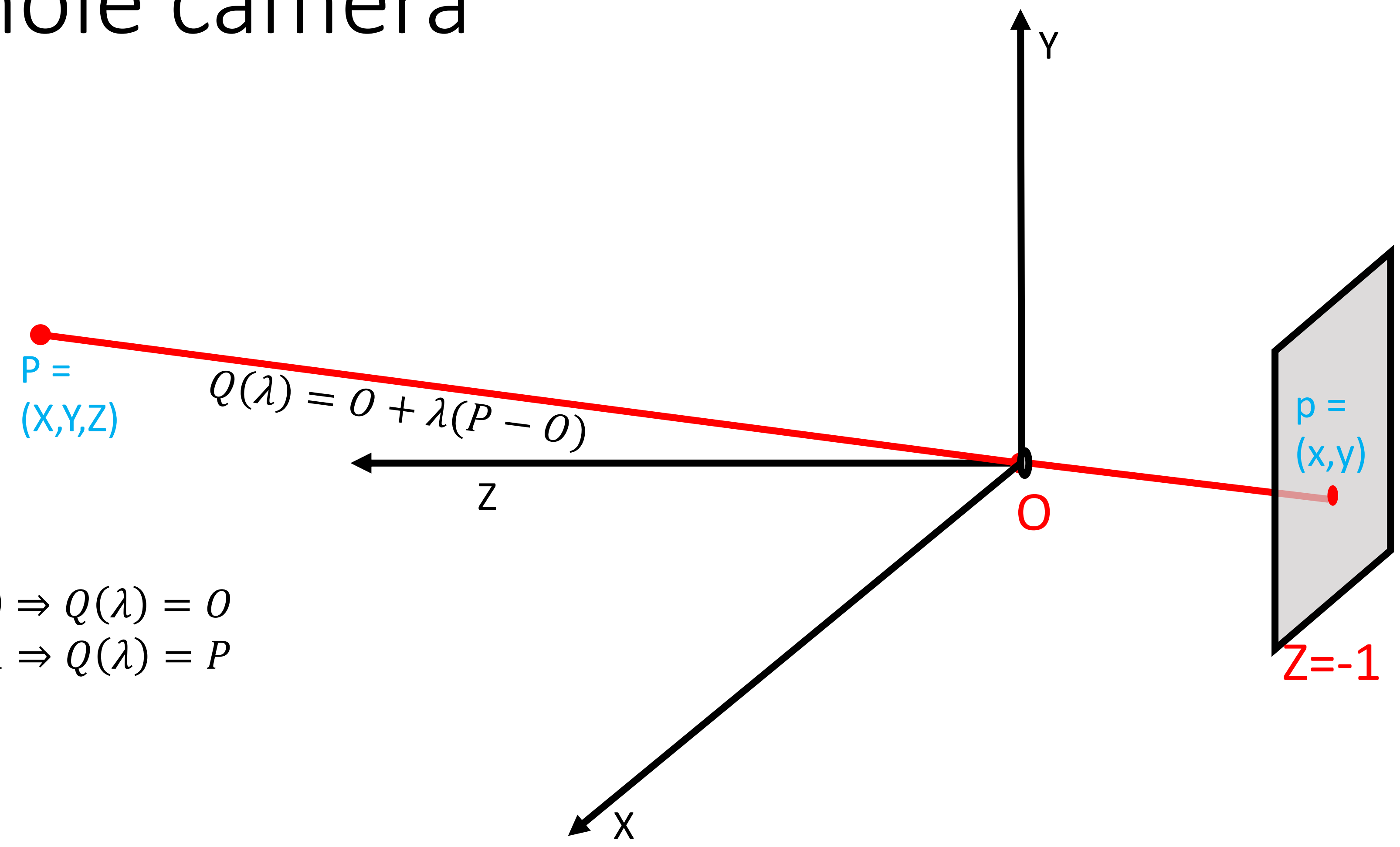
- Let's remove the wall with the pinhole: all we care about is that all light rays of interest *must pass through the pinhole, i.e., the origin*

The pinhole camera



- Question: Where will we see the “image” of point P on the camera plane?

The pinhole camera



$\lambda = 0 \Rightarrow Q(\lambda) = O$
$\lambda = 1 \Rightarrow Q(\lambda) = P$

$$
\begin{aligned}
&Q(\lambda) \\
&= \big(0 + \lambda(X - 0), 0 + \lambda(Y - 0), 0 + \lambda(Z - 0)\big) \\
&= (\lambda X, \lambda Y, \lambda Z)
\end{aligned}
$$

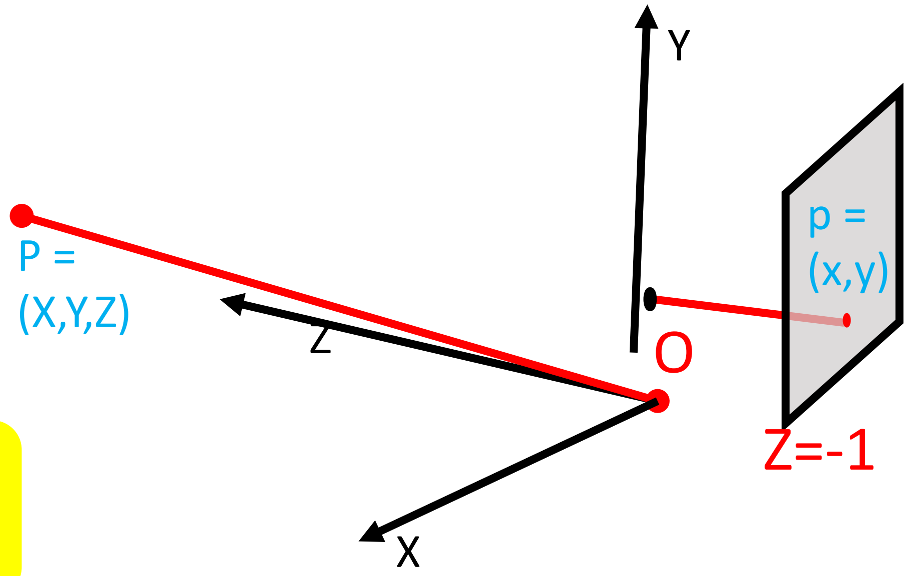
The pinhole camera

- Pinhole camera collapses *ray OP to point p*
- Any point on ray OP = $O + \lambda(P - O) = (\lambda X, \lambda Y, \lambda Z)$
- For this point to lie on Z=-1 plane:

$$\lambda^* Z = -1$$
$$\Rightarrow \lambda^* = \frac{-1}{Z}$$

- Coordinates of point p:

$$(\lambda^* X, \lambda^* Y, \lambda^* Z) = (\frac{-X}{Z}, \frac{-Y}{Z}, -1)$$



The projection equation

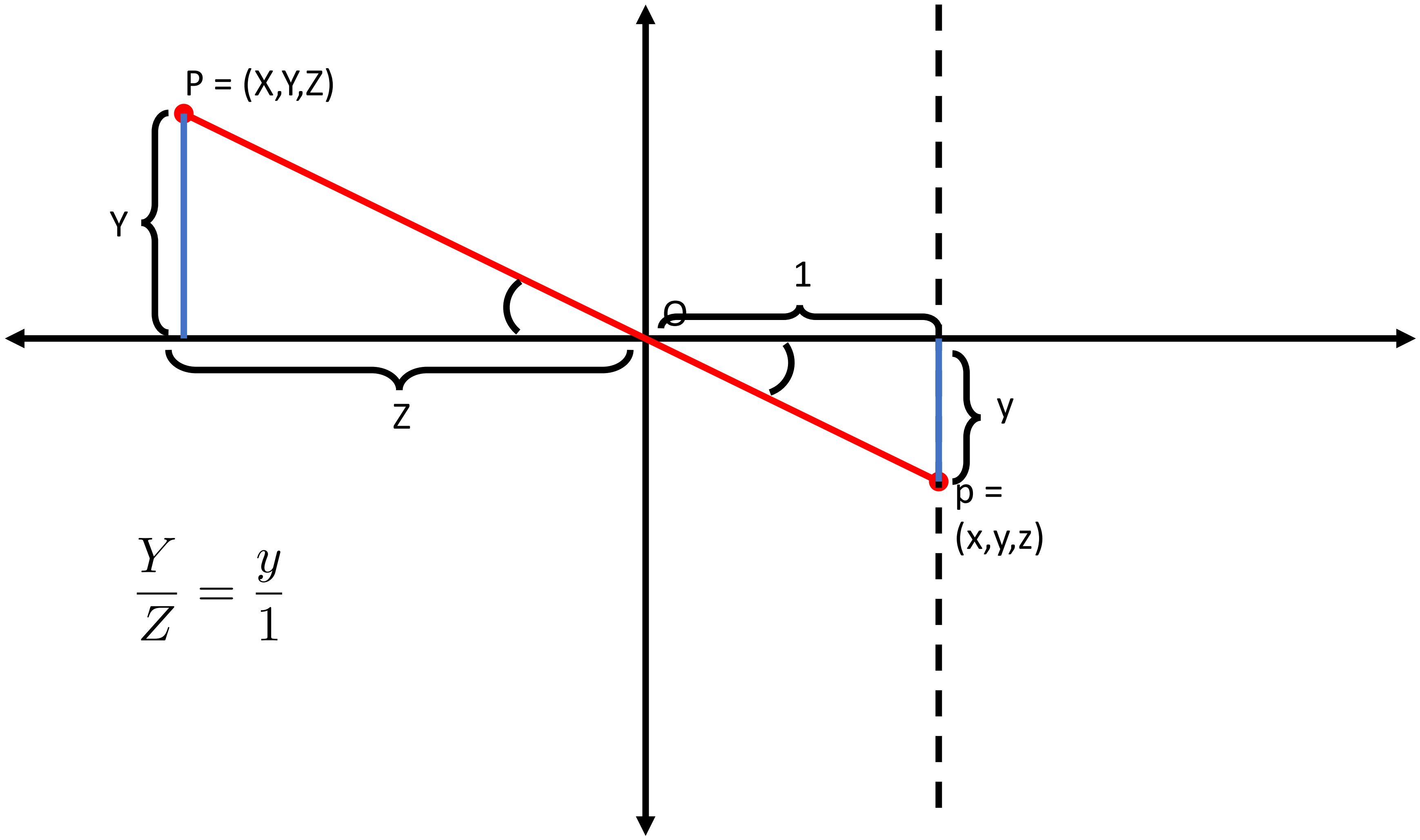
- A point P = (X, Y, Z) in 3D projects to a point p = (x,y) in the image

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

- But pinhole camera's image is inverted, invert it back!

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

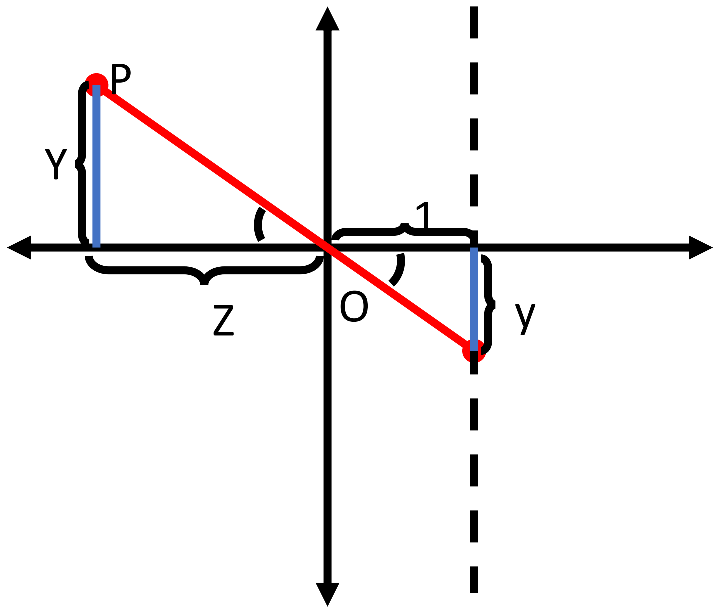
Another derivation

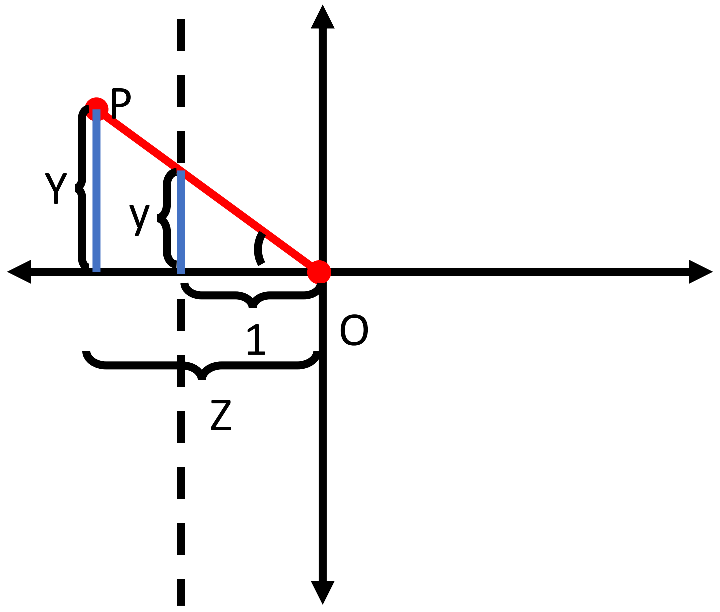


$$\frac{Y}{Z} = \frac{y}{1}$$

A virtual image plane

- A pinhole camera produces an inverted image
- Imagine a ”virtual image plane” in the front of the camera





The projection equation

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

Consequence 1: Farther away objects are smaller

Image of foot: $(\frac{X}{Z}, \frac{Y}{Z})$

Image of head: $(\frac{X}{Z}, \frac{Y+h}{Z})$

$$\frac{Y+h}{Z} - \frac{Y}{Z} = \frac{h}{Z}$$

Consequence 2: Parallel lines converge at a point

- Point on a line passing through point A with direction D:
$$Q(\lambda) = A + \lambda D$$
- Parallel lines have the same direction but pass through different points
$$Q(\lambda) = A + \lambda D$$
$$R(\lambda) = B + \lambda D$$
-

Consequence 2: Parallel lines converge at a point

- Parallel lines have the same direction but pass through different points

$$Q(\lambda) = A + \lambda D$$
$$R(\lambda) = B + \lambda D$$

- $A = (A_X, A_Y, A_Z)$
- $B = (B_X, B_Y, B_Z)$
- $D = (D_X, D_Y, D_Z)$

Consequence 2: Parallel lines converge at a point

- $Q(\lambda) = (A_X + \lambda D_X, A_Y + \lambda D_Y, A_Z + \lambda D_Z)$
- $R(\lambda) = (B_X + \lambda D_X, B_Y + \lambda D_Y, B_Z + \lambda D_Z)$
- $q(\lambda) = \left(\frac{A_X+\lambda D_X}{A_Z+\lambda D_Z}, \frac{A_Y+\lambda D_Y}{A_Z+\lambda D_Z}\right)$
- $r(\lambda) = \left(\frac{B_X+\lambda D_X}{B_Z+\lambda D_Z}, \frac{B_Y+\lambda D_Y}{B_Z+\lambda D_Z}\right)$
- Need to look at these points as Z goes to infinity
- Same as $\lambda \to \infty$

Consequence 2: Parallel lines converge at a point

- $q(\lambda) = \left(\frac{A_X+\lambda D_X}{A_Z+\lambda D_Z}, \frac{A_Y+\lambda D_Y}{A_Z+\lambda D_Z}\right)$
- $r(\lambda) = \left(\frac{B_X+\lambda D_X}{B_Z+\lambda D_Z}, \frac{B_Y+\lambda D_Y}{B_Z+\lambda D_Z}\right)$

$$\lim_{\lambda\to\infty} \frac{A_X + \lambda D_X}{A_Z + \lambda D_Z} = \lim_{\lambda\to\infty} \frac{\frac{A_X}{\lambda} + D_X}{\frac{A_Z}{\lambda} + D_Z} = \frac{D_X}{D_Z}$$

$$\lim_{\lambda\to\infty} q(\lambda) = (\frac{D_X}{D_Z}, \frac{D_Y}{D_Z}) \qquad \lim_{\lambda\to\infty} r(\lambda) = (\frac{D_X}{D_Z}, \frac{D_Y}{D_Z})$$

Consequence 2: Parallel lines converge at a point

- Parallel lines have the same direction but pass through different points

$$Q(\lambda) = A + \lambda D$$
$$R(\lambda) = B + \lambda D$$

- Parallel lines converge at the same point $(\frac{D_X}{D_Z}, \frac{D_Y}{D_Z})$
- This point of convergence is called the *vanishing point*
- What happens if $D_Z = 0$?

Consequence 2: Parallel lines converge at a point

What about planes?

$$N_X X + N_Y Y + N_Z Z = d$$

$$\Rightarrow N_X \frac{X}{Z} + N_Y \frac{Y}{Z} + N_Z = \frac{d}{Z}$$

$$\Rightarrow N_X x + N_Y y + N_Z = \frac{d}{Z}$$

Take the limit as Z approaches infinity

$$\boxed{N_X x + N_Y y + N_Z = 0}$$

Vanishing line of a plane

What about planes?

$$N_X X + N_Y Y + N_Z Z = d$$

Normal: (N_X, N_Y, N_Z)

What do parallel planes look like?

$$N_X X + N_Y Y + N_Z Z = d \qquad\qquad N_X X + N_Y Y + N_Z Z = c$$

$$\boxed{N_X x + N_Y y + N_Z = 0} \qquad\qquad \boxed{N_X x + N_Y y + N_Z = 0}$$

Vanishing lines

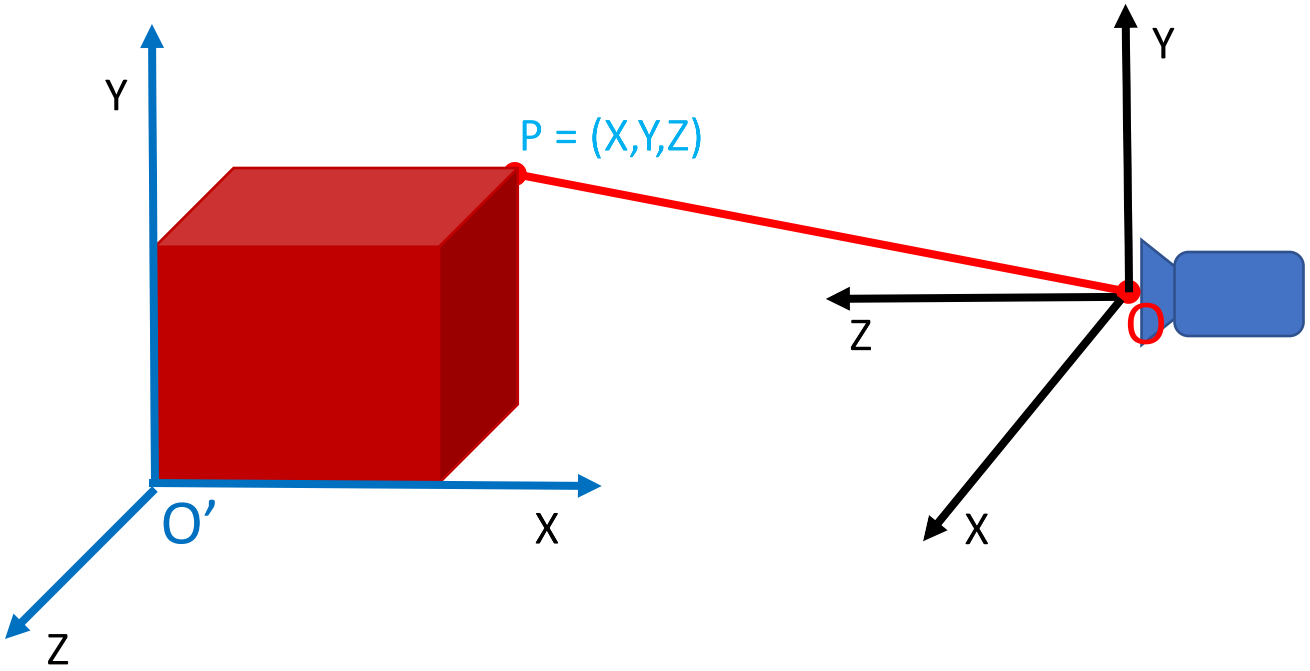
Parallel planes converge!

Vanishing line

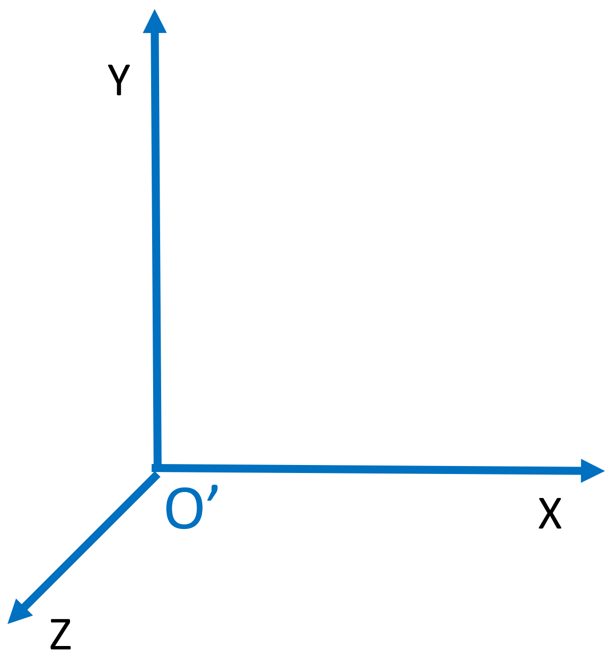
$$N_X X + N_Y Y + N_Z Z = d$$

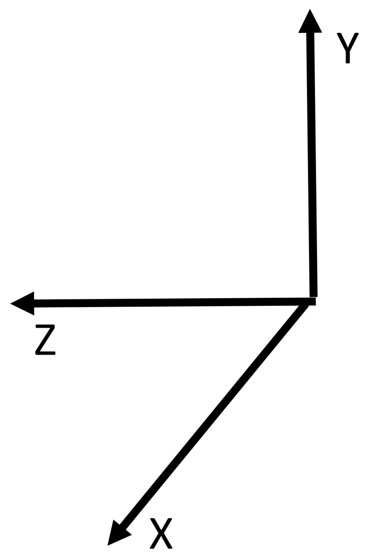
- What happens if $N_X = N_Y = 0$?
- Equation of the plane: $Z = c$
- Vanishing line?

Changing coordinate systems

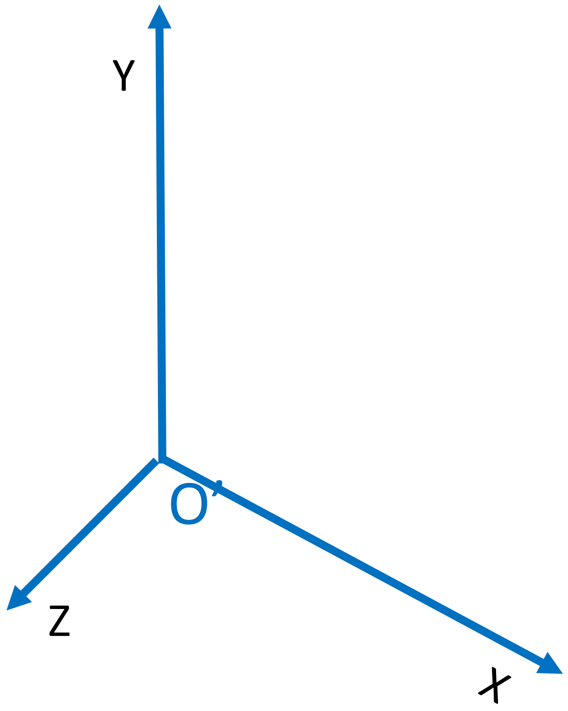


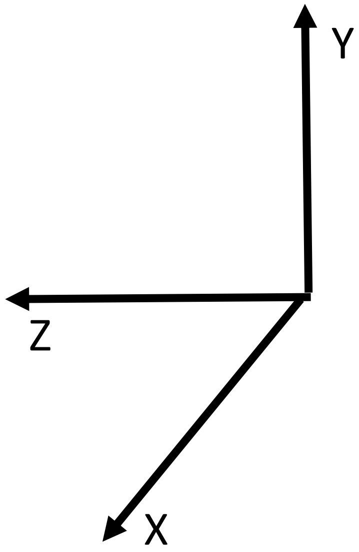
Changing coordinate systems



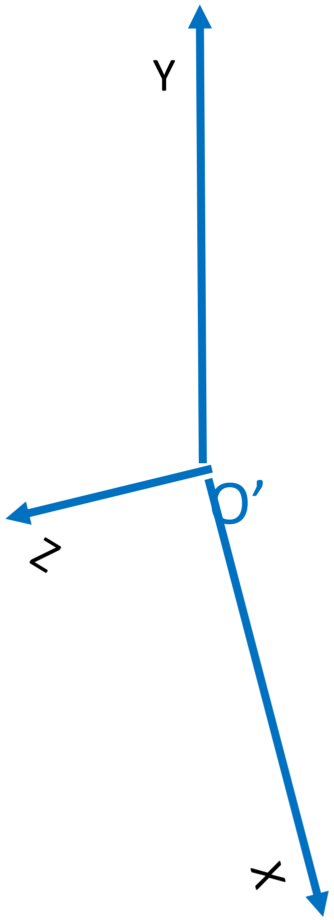


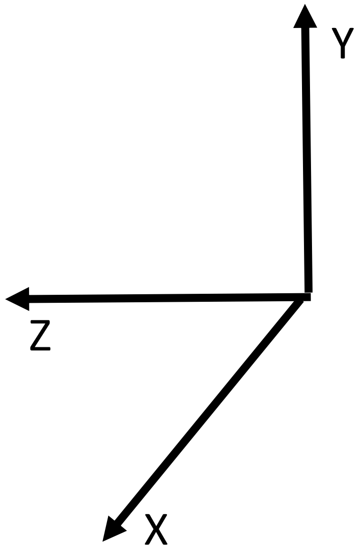
Changing coordinate systems





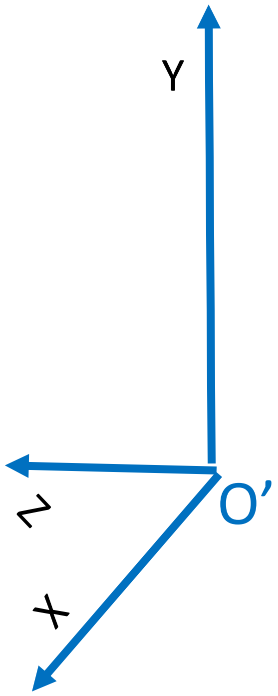
Changing coordinate systems

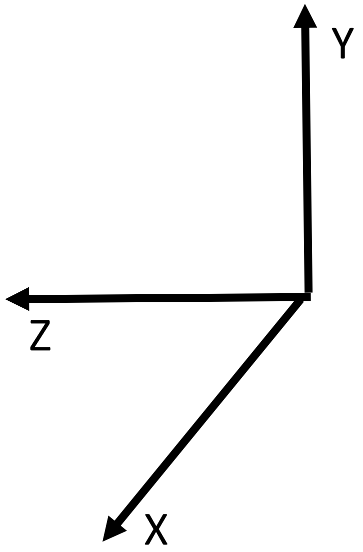




Changing coordinate systems

Changing coordinate systems





Rotations and translations

- How do you represent a rotation?
- A point in 3D: (X,Y,Z)
- Rotations can be represented as a matrix multiplication

$$\mathbf{v}' = R\mathbf{v}$$

- What are the properties of rotation matrices?

Properties of rotation matrices

- Rotation does not change the length of vectors

$$\mathbf{v}' = R\mathbf{v}$$

$$\|\mathbf{v}'\|^2 = \mathbf{v}'^T\mathbf{v}'$$

$$= \mathbf{v}^T R^T R\mathbf{v}$$

$$\|\mathbf{v}\|^2 = \mathbf{v}^T\mathbf{v}$$

$$\Rightarrow R^T R = I$$

Properties of rotation matrices

$$\Rightarrow R^T R = I$$

$$\Rightarrow det(R)^2 = 1$$

$$\Rightarrow det(R) = \pm 1$$

$$det(R) = 1$$

Rotation

$$det(R) = -1$$

Reflection

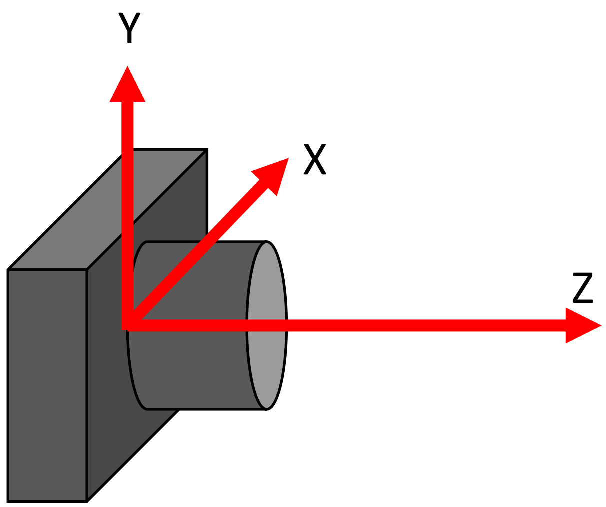
Translations

$$\mathbf{x}' = \mathbf{x} + \mathbf{t}$$

- Can this be written as a matrix multiplication?

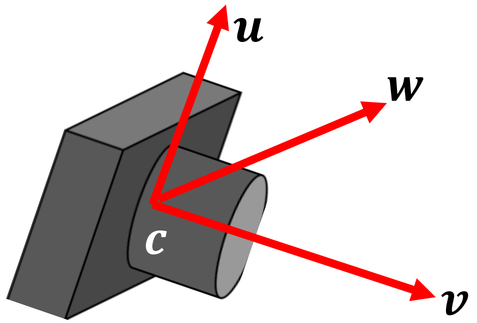
Till now

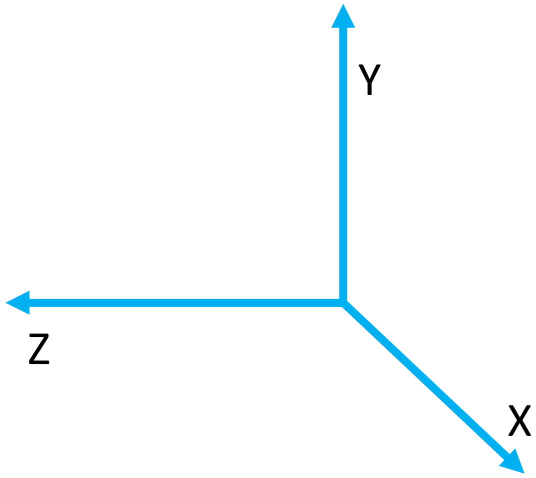
- We assumed coordinate system:
  - With origin at pinhole
  - With Z axis along viewing direction
  - Let us call this “camera coordinate system”
  - What about the world coordinate system?



Aligning coordinate systems

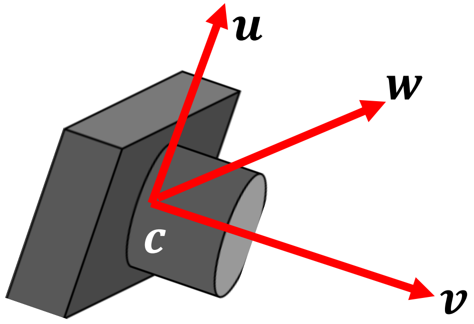
- Suppose according to the world coordinate system
  - Camera is at $\boldsymbol{c}$
  - Pointing in direction $\boldsymbol{v}$
  - The camera's vertical direction is $\boldsymbol{u}$
  - And its horizontal direction is $\boldsymbol{w}$

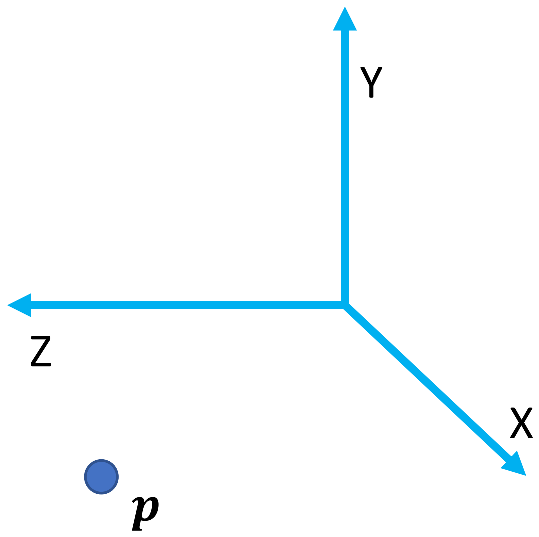




Aligning coordinate systems

- Consider a point with coordinates $\boldsymbol{p}$ in the world coordinates
- To express it in camera coordinates, we need to rotate and translate
  - $\boldsymbol{q} = R\boldsymbol{p} + \boldsymbol{t}$
  - Where R is a rotation matrix
- What should R and t be?





Aligning coordinate systems

- Let's first figure out R

$$R\boldsymbol{v} = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}$$

$$R\boldsymbol{u} = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}$$

$$R\boldsymbol{w} = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}$$

Aligning coordinate systems

- Let's first figure out R

$$\Longrightarrow R[\boldsymbol{w} \quad \boldsymbol{u} \quad \boldsymbol{v}] = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} = \boldsymbol{I}$$

$$\Longrightarrow R^T R[\boldsymbol{w} \quad \boldsymbol{u} \quad \boldsymbol{v}] = R^T$$

$$\Rightarrow R^T = [\boldsymbol{w} \quad \boldsymbol{u} \quad \boldsymbol{v}]$$

$$\Rightarrow R = [\boldsymbol{w} \quad \boldsymbol{u} \quad \boldsymbol{v}]^T$$

Aligning coordinate systems

- Let us next figure out $\boldsymbol{t}$

$$R\boldsymbol{c} + \boldsymbol{t} = \boldsymbol{0}$$
$$\Rightarrow \boldsymbol{t} = -R\boldsymbol{c}$$

- Thus:

$$\boxed{\begin{aligned} R &= [\boldsymbol{w} \quad \boldsymbol{u} \quad \boldsymbol{v}]^{\mathrm{T}} \\ \boldsymbol{t} &= -R\boldsymbol{c} \end{aligned}}$$

Putting everything together

- Change coordinate system so that center of the coordinate system is at pinhole and Z axis is along viewing direction

$$\mathbf{x}'_w = R\mathbf{x}_w + \mathbf{t}$$

- Perspective projection

$$\mathbf{x}'_w \equiv (X, Y, Z)$$
$$\mathbf{x}'_{img} \equiv (x, y)$$

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

Putting everything together

- Change coordinate system so that center of the coordinate system is at pinhole and Z axis is along viewing direction

$$\mathbf{x}'_w = R\mathbf{x}_w + \mathbf{t}$$

- Perspective projection

$$\mathbf{x}'_w \equiv (X, Y, Z)$$
$$\mathbf{x}'_{img} \equiv (x, y)$$

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

The projection equation

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

- Is this equation linear?
- Can this equation be represented by a matrix multiplication?

Is projection linear?

$$X' = aX + b$$
$$Y' = aY + b$$
$$Z' = aZ + b$$

$$x' = \frac{aX + b}{aZ + b}$$

$$y' = \frac{aY + b}{aZ + b}$$

Can projection be represented as a matrix multiplication?

Matrix multiplication

$$\begin{bmatrix} a & b & c \\ p & q & r \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \end{bmatrix} = \begin{bmatrix} aX + bY + cZ \\ pX + qY + rZ \end{bmatrix}$$

Perspective projection

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

Homogenous coordinates

- Cartesian coordinates : Each point represented by 2 coordinates (x,y)
- Homogenous coordinates : Each “point” represented by *3* coordinates (x,y,z), BUT:
  - $(\lambda x, \lambda y, \lambda z) \equiv (x, y, z)$

- Mapping Cartesian to Homogenous :

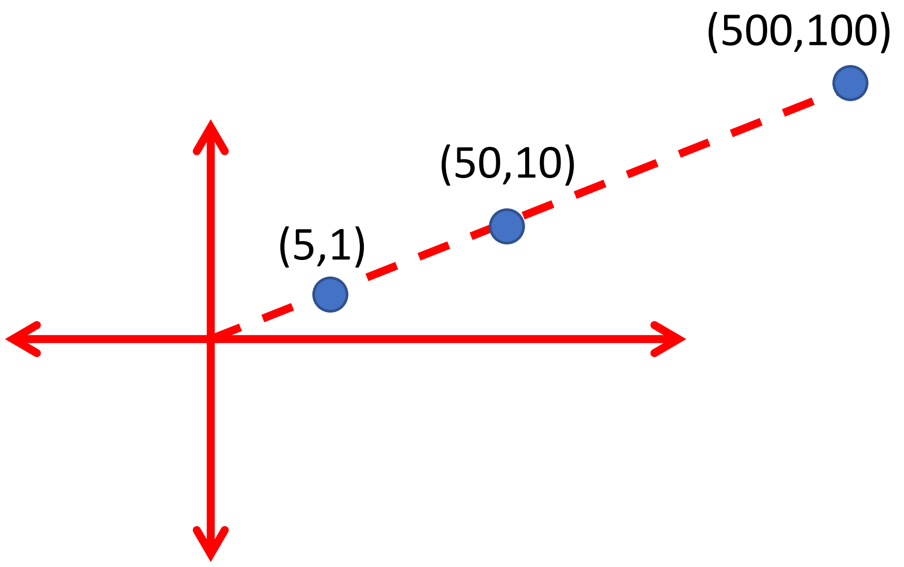
$$(x, y) \rightarrow (x, y, 1)$$

- Mapping Homogenous to Cartesian :

$$(x, y, z) \rightarrow (\frac{x}{z}, \frac{y}{z})$$

Points at infinity

- 2D Points in homogenous coordinates: (x, y, z)
- What if $z = 0$?
- Corresponding cartesian coordinates: $\left(\frac{x}{z}, \frac{y}{z}\right) = ?$
- Suppose x=5, y= 1
  - $(5,1,1) \rightarrow \left(\frac{5}{1}, \frac{1}{1}\right) = (5,1)$
  - $(5,1,0.1) \rightarrow \left(\frac{5}{0.1}, \frac{1}{0.1}\right) = (50, 10)$
  - $(5,1,0.01) \rightarrow \left(\frac{5}{0.01}, \frac{1}{0.01}\right) = (500, 100)$
  - $(5,1,0) \rightarrow$ point "at infinity " along the same direction



Homogenous coordinates

- In standard Cartesian coordinates
  - 2D points : (x,y)
  - 3D points : (x,y,z)

- In homogenous coordinates
  - 2D points : (x,y,1)
  - 3D points : (x,y,z,1)

Why homogenous coordinates?

$$\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix} = \begin{bmatrix} X \\ Y \\ Z \end{bmatrix} \equiv \begin{bmatrix} \frac{X}{Z} \\ \frac{Y}{Z} \\ 1 \end{bmatrix}$$

Homogenous coordinates of world point

Homogenous coordinates of image point

Why homogenous coordinates?

$$\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix} = \begin{bmatrix} X \\ Y \\ Z \end{bmatrix} \equiv \begin{bmatrix} \frac{X}{Z} \\ \frac{Y}{Z} \\ 1 \end{bmatrix}$$

$$P\vec{\mathbf{x}}_w = \vec{\mathbf{x}}_{img}$$

- Perspective projection is matrix multiplication in homogenous coordinates!

Why homogenous coordinates?

$$\begin{bmatrix} 1 & 0 & 0 & t_x \\ 0 & 1 & 0 & t_y \\ 0 & 0 & 1 & t_z \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix}$$

- Translation is matrix multiplication in homogenous coordinates!

Homogenous coordinates

$$\begin{bmatrix} a & b & c & t_x \\ d & e & f & t_y \\ g & h & i & t_z \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix} = \begin{bmatrix} aX + bY + cZ + t_x \\ dX + eY + fZ + t_y \\ gX + hY + iZ + t_z \\ 1 \end{bmatrix}$$

$$\begin{bmatrix} M & \mathbf{t} \\ \mathbf{0}^T & 1 \end{bmatrix} \begin{bmatrix} \mathbf{x}_w \\ 1 \end{bmatrix} = \begin{bmatrix} M\mathbf{x}_w + \mathbf{t} \\ 1 \end{bmatrix}$$

Homogenous coordinates

$$\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix} = \begin{bmatrix} X \\ Y \\ Z \end{bmatrix} \equiv \begin{bmatrix} \frac{X}{Z} \\ \frac{Y}{Z} \\ 1 \end{bmatrix}$$

$$\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \Rightarrow \begin{bmatrix} I & \mathbf{0} \end{bmatrix}$$

Perspective projection in homogenous coordinates

$$\vec{\mathbf{x}}_{img} = \begin{bmatrix} I & \mathbf{0} \end{bmatrix} \begin{bmatrix} R & \mathbf{t} \\ \mathbf{0}^T & 1 \end{bmatrix} \vec{\mathbf{x}}_w$$

$$\vec{\mathbf{x}}_{img} = \begin{bmatrix} R & \mathbf{t} \end{bmatrix} \vec{\mathbf{x}}_w$$

More about matrix transformations

$\begin{bmatrix} I & \mathbf{0} \end{bmatrix}$ 3 x 4 : Perspective projection

$\begin{bmatrix} I & \mathbf{t} \\ \mathbf{0}^T & 1 \end{bmatrix}$ 4 x 4 : Translation

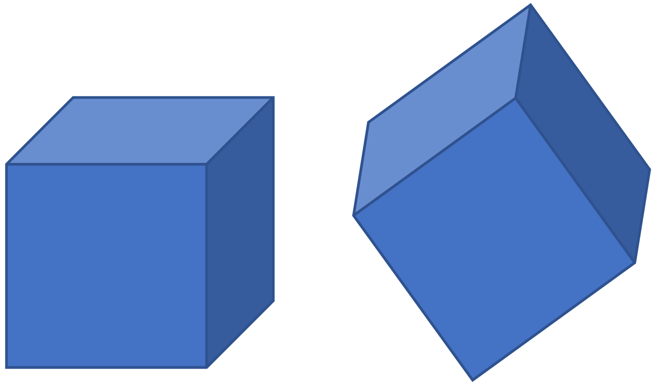
$\begin{bmatrix} M & \mathbf{t} \\ \mathbf{0}^T & 1 \end{bmatrix}$ 4 x 4 : Affine transformation (linear transformation + translation)

More about matrix transformations

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

$$M^T M = I$$

Euclidean



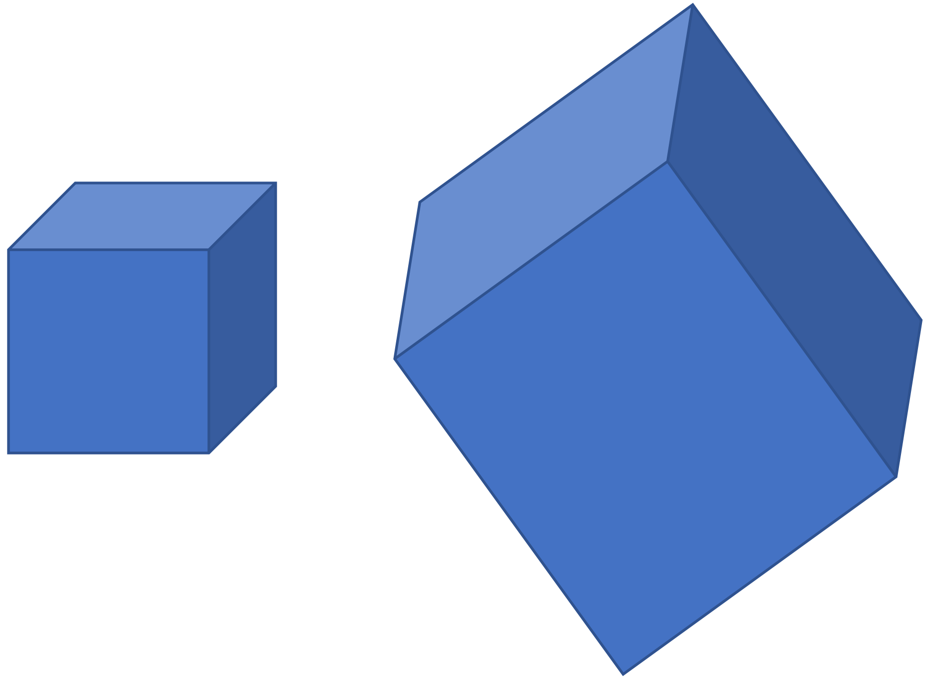
More about matrix transformations

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

$$M = sR$$

$$R^T R = I$$

Similarity transformation

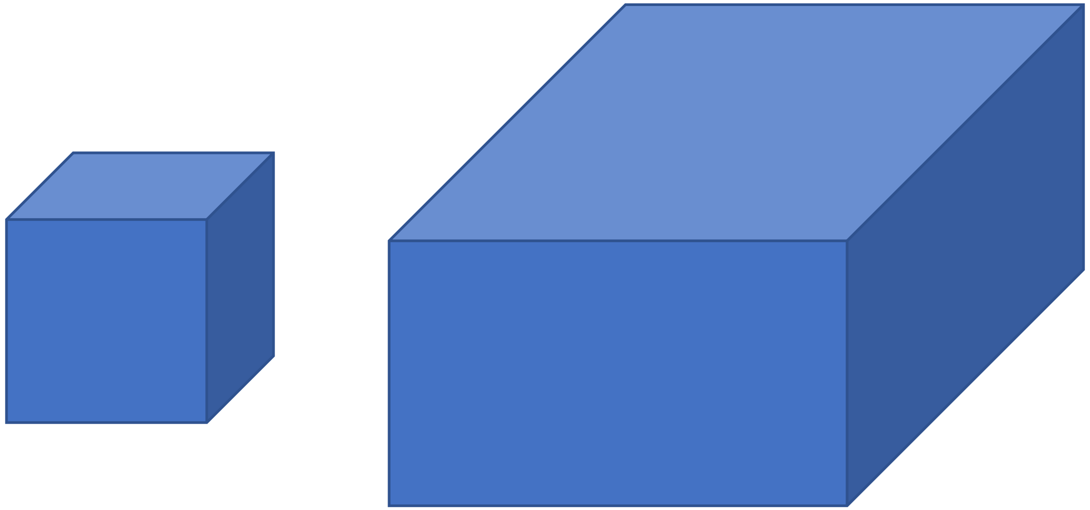


More about matrix transformations

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

$$M = \begin{bmatrix} s_x & 0 & 0 \\ 0 & s_y & 0 \\ 0 & 0 & s_z \end{bmatrix}$$

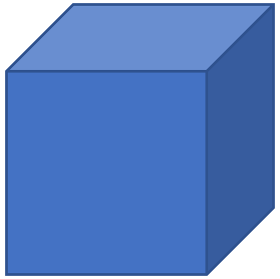
Anisotropic scaling and translation

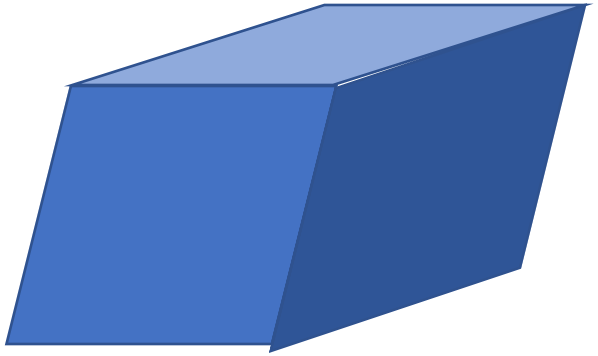


More about matrix transformations

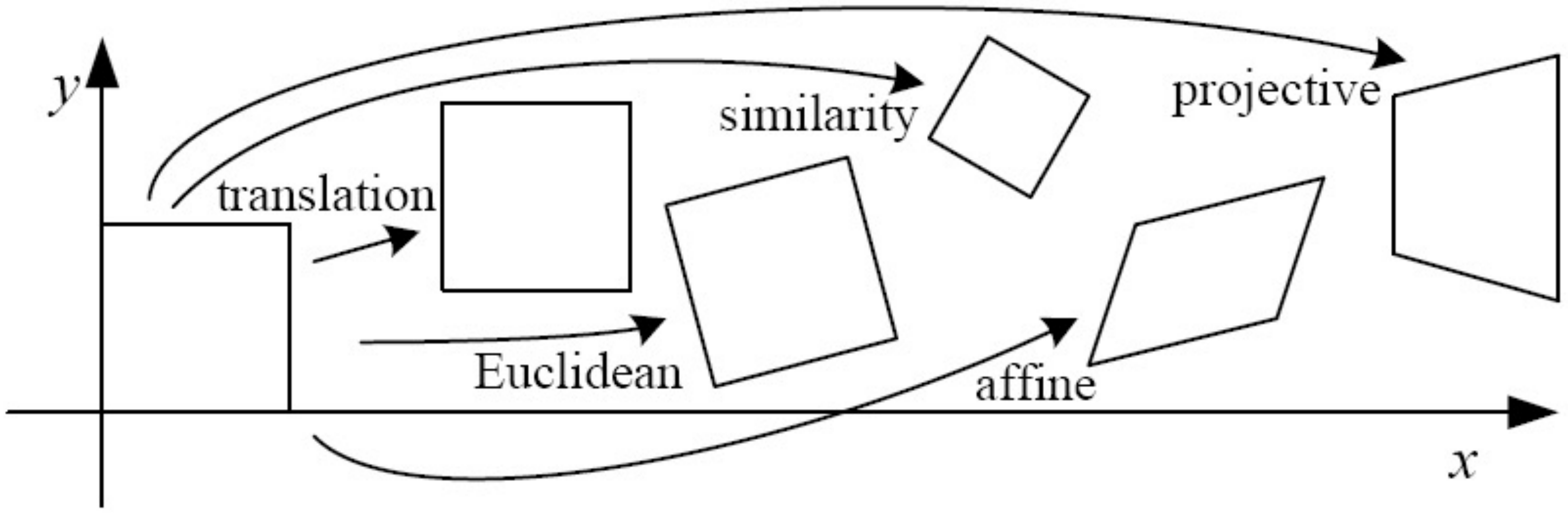
$$\begin{bmatrix} M & \mathbf{t} \\ \mathbf{0}^T & 1 \end{bmatrix}$$

General affine transformation





Matrix transformations in 2D



Perspective transformation till now

$$\vec{\boldsymbol{x}}_{img} \equiv [\,R \quad \boldsymbol{t}\,]\,\vec{\boldsymbol{x}}_w$$

- What about image coordinate system?
  - Origin not in center but on top left
  - Units are different (dependent on resolution)
  - Assume a general homogenous transformation K
  - If only translation and scale, then $K \equiv \begin{bmatrix} s_x & 0 & u \\ 0 & s_y & v \\ 0 & 0 & 1 \end{bmatrix}$
  - Final equation:

$$\boxed{\vec{\boldsymbol{x}}_{img} \equiv K[\,R \quad \boldsymbol{t}\,]\,\vec{\boldsymbol{x}}_w}$$