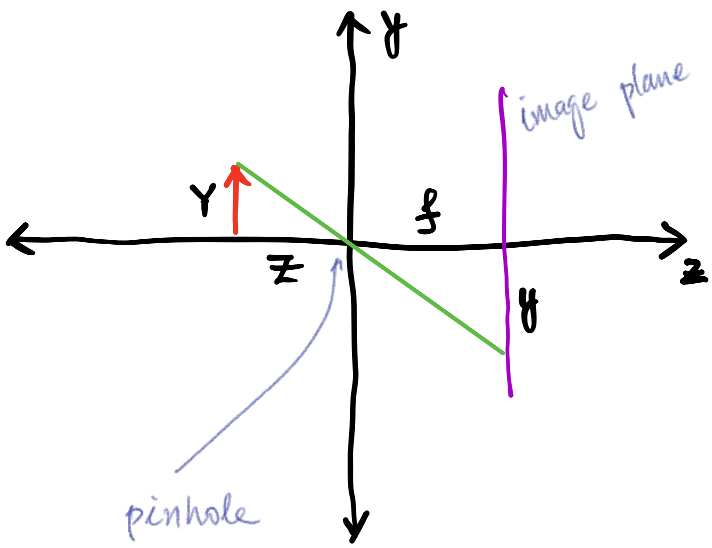
Jan 16, 2026

TOPICS

- Pinhole camera model
- Transformations
- Camera projection matrix

Pinhole Camera Model



Y = height of object

Z = depth of object

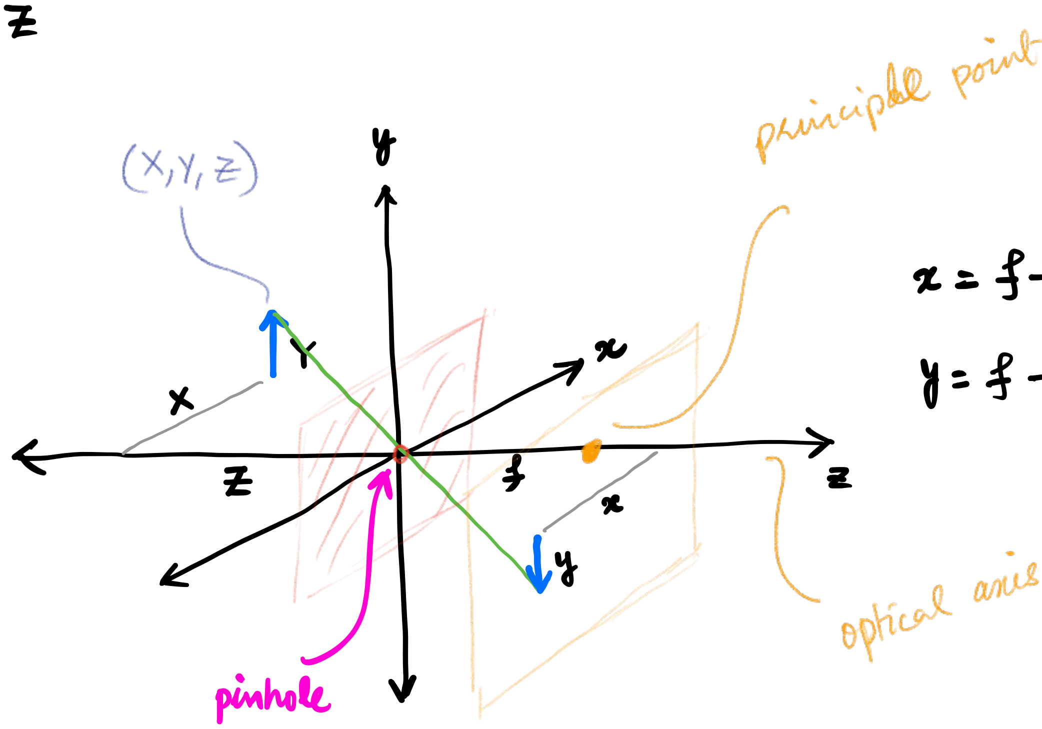
f = focal length

distance between pinhole and image plane

y = height of imaged object

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

3D



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

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

Homogeneous Coordinates

	Cartesian	Homogeneous
2D:	(x, y)	$(x, y, 1)$
	$\left(\frac{x}{w}, \frac{y}{w}\right)$	(x, y, w)
3D:	(x, y, z)	$(x, y, z, 1)$
	$\left(\frac{x}{w}, \frac{y}{w}, \frac{z}{w}\right)$	(x, y, z, w)

(i) Homogeneous coordinate (kx, ky, k) maps to Cartesian (x, y) for $k \neq 0$.

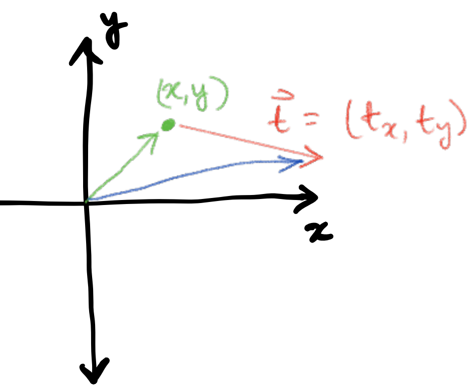
(ii) Homogeneous coordinate $(x, y, 0)$, it represents a point at infinity in the dir. (x, y).

2D

Transformations

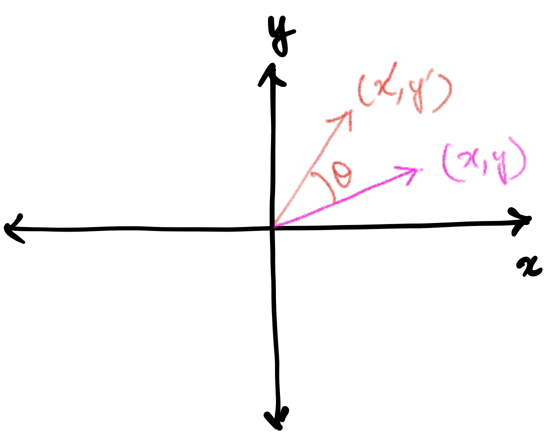
2D: Translation

Rigid body transformations



$$x' = x + t_x$$
$$y' = y + t_y$$

Rotation

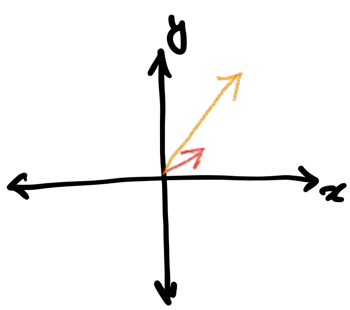


$$\begin{bmatrix} x' \\ y' \end{bmatrix} = \begin{bmatrix} c\theta & -s\theta \\ s\theta & c\theta \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix}$$

$$c\theta = \cos(\theta)$$
$$s\theta = \sin(\theta)$$

Scaling

$$(x', y') = (s_x x, s_y y)$$



Shearing

$$(x', y') = (x + q_x y, \; q_y x + y)$$

Goal: Represent these as matrix-vector products.

e.g. scaling.

$$\begin{bmatrix} x' \\ y' \end{bmatrix} = \begin{bmatrix} s_x & 0 \\ 0 & s_y \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix}$$

* For translation

$$\begin{bmatrix} x' \\ y' \\ 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 & t_x \\ 0 & 1 & t_y \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ 1 \end{bmatrix}$$

homogeneous coordinates.

Concatenate Transformations

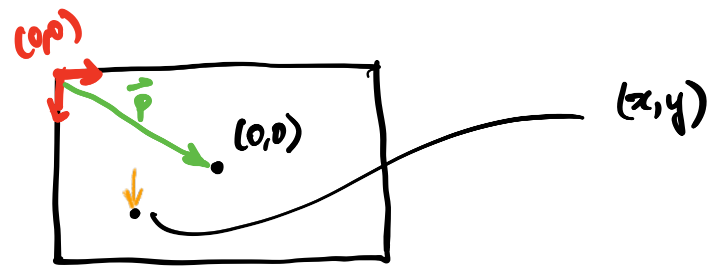
$$\begin{bmatrix} x' \\ y' \\ 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 & t_x \\ 0 & 1 & t_y \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} c\theta & -s\theta & 0 \\ s\theta & c\theta & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ 1 \end{bmatrix}$$

$$= \left[\begin{array}{cc|c} c\theta & -s\theta & t_x \\ s\theta & c\theta & t_y \\ \hline 0 & 0 & 1 \end{array}\right] \begin{bmatrix} x \\ y \\ 1 \end{bmatrix}$$

In 3D:

$$\begin{bmatrix} w x' \\ w y' \\ w z' \\ w \end{bmatrix} = \underbrace{\left[\begin{array}{c|c} R & \begin{matrix} t_x \\ t_y \\ t_z \end{matrix} \\ \hline \emptyset & 1 \end{array}\right]}_{\in \mathbb{R}^{4 \times 4}} \begin{bmatrix} x \\ y \\ z \\ 1 \end{bmatrix}, \quad R \in \mathbb{R}^{3 \times 3}$$

Image Coordinates.



①

$$\begin{bmatrix} x \\ y \\ w \end{bmatrix} = \begin{bmatrix} f & 0 & 0 \\ 0 & f & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \end{bmatrix}$$

homogeneous image coordinates

world point *

$$x = fX$$
$$y = fY$$
$$w = Z$$

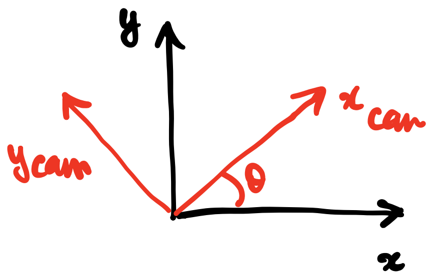
cartesian

$$\left(\frac{fX}{w}, \frac{fY}{w} \right)$$

For a principal point (p_x, p_y)

$$\begin{bmatrix} f & 0 & p_x \\ 0 & f & p_y \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \end{bmatrix}$$

Thus far we have assumed that pinhole sits at (0,0,0). How do we deal with camera translation & rotation?



$$x_{cam} = R_\theta x$$

$$y_{cam} = R_\theta y$$

Recall $R_\theta = \begin{bmatrix} c\theta & -s\theta \\ s\theta & c\theta \end{bmatrix}$

$$P_{camera} = R(-\theta) P_{world}$$

$$R(-\theta) = R(-\theta)^T = R \in \mathbb{R}^{3\times 3}$$

camera rotation matrix

Camera center: $\tilde{C}$

rotation center (pinhole)

$$\underset{\mathbb{R}^{3\times3}}{P_{camera}} = \overset{\mathbb{R}^{3\times1}}{R} (P_{world} - \underset{\mathbb{R}^{3\times1}}{\underline{\underline{\tilde{C}}}})$$

world point

$$= R P_{world} - R\tilde{C}$$

Putting everything together.

$$\begin{bmatrix} wx \\ wy \\ w \end{bmatrix} = \underbrace{\begin{bmatrix} f & 0 & P_x \\ 0 & f & P_y \\ 0 & 0 & 1 \end{bmatrix}}_{\in \mathbb{R}^{3\times 3}} \underbrace{\left[\, R \;\middle|\; -R\tilde{C} \,\right]}_{\in \mathbb{R}^{3\times 4}} \underbrace{\begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix}}_{\in \mathbb{R}^{4\times 1}}$$

image   camera   world

intrinsic   extrinsic