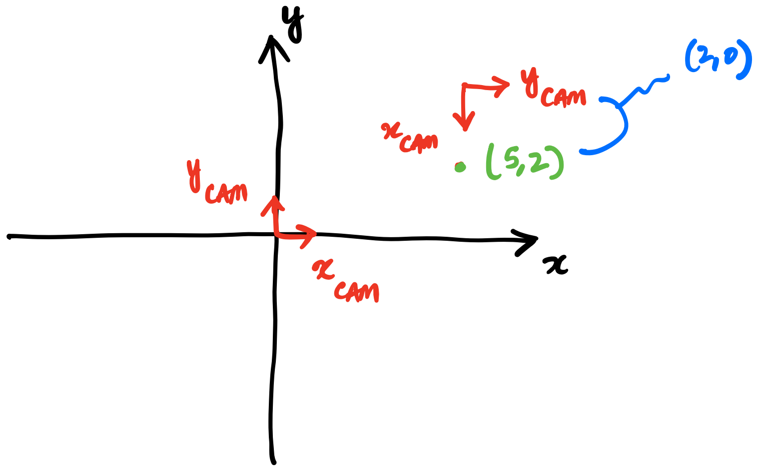
Jan 21, 2026



Pinhole Model

$$\vec{x} = \underbrace{K}_{\text{intrinsic matrix}} \underbrace{[R|\vec{t}]}_{\text{extrinsic matrix}} \vec{X}$$

intrinsic matrix.
- it maps points in the camera coordinate into image coordinate.

Matrix $[R|\vec{t}]$ maps points in the world coordinates into camera coordinates.

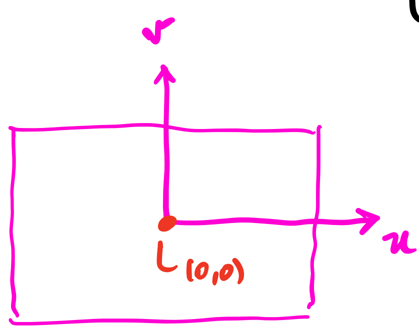
$$\begin{bmatrix} x \\ y \\ w \end{bmatrix} = \begin{bmatrix} f & & \\ & f & \\ & & 1 \end{bmatrix} \begin{bmatrix} r_{11} & r_{12} & r_{13} & t_x \\ r_{21} & r_{22} & r_{23} & t_y \\ r_{31} & r_{32} & r_{33} & t_z \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix}$$

image coordinates (homogeneous)

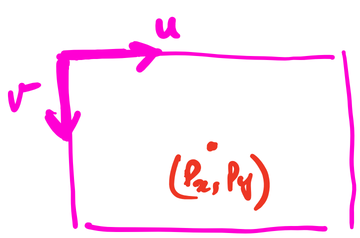
world coordinates (homogeneous)

Cartesian

$$u = \frac{x}{w}, \quad v = \frac{y}{w}$$

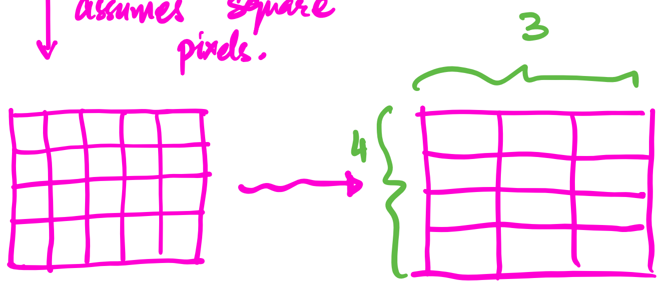


Shift the principle point.



$$\begin{bmatrix} x \\ y \\ w \end{bmatrix} = \begin{bmatrix} f & & p_x \\ & f & p_y \\ & & 1 \end{bmatrix} \begin{bmatrix} r_{11} & r_{12} & r_{13} & t_x \\ r_{21} & r_{22} & r_{23} & t_y \\ r_{31} & r_{32} & r_{33} & t_z \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix}$$

$\downarrow$ assumes square pixels.



# of pixels per unit length is different in u, v direction.

m_x = # of pixels in unit length in u (x)

m_y = # of pixels in unit length in v (y)

$$\begin{bmatrix} x \\ y \\ w \end{bmatrix} = \begin{bmatrix} \alpha_x & & x_0 \\ & \alpha_y & y_0 \\ & & 1 \end{bmatrix} \begin{bmatrix} r_{11} & r_{12} & r_{13} & t_x \\ r_{21} & r_{22} & r_{23} & t_y \\ r_{31} & r_{32} & r_{33} & t_z \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix}$$

where $\alpha_x = f m_x$, $\alpha_y = f m_y$ and $x_0 = m_x p_x$ and $y_0 = m_y p_y$

what if image plane is not perfectly aligned with xy plane.

$$\begin{bmatrix} x \\ y \\ w \end{bmatrix} = \begin{bmatrix} \alpha_x & s & x_0 \\ & \alpha_y & y_0 \\ & & 1 \end{bmatrix} \begin{bmatrix} r_{11} & r_{12} & r_{13} & t_x \\ r_{21} & r_{22} & r_{23} & t_y \\ r_{31} & r_{32} & r_{33} & t_z \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix}$$

$$= K\,[R|\vec{t}\,] \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix}$$

$K \in \mathbb{R}^{3\times 3}$

$[R|\vec{t}\,] \in \mathbb{R}^{3\times 4}$

$R \in \mathbb{R}^{3\times 3}$

$\vec{t} \in \mathbb{R}^{3\times 1}$

$$= P \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix}$$

3x4 projection matrix.

* $\vec{x} = P\vec{X}$

(i) $P = \left[\begin{array}{c|c} M & p_4 \end{array}\right]$ $\quad M \in \mathbb{R}^{3\times3}$, $p_4 \in \mathbb{R}^{3\times1}$

(ii) $P = \begin{bmatrix} — & p_1^T & — \\ — & p_2^T & — \\ — & p_3^T & — \end{bmatrix}$ $\quad p_1^T \in \mathbb{R}^{1\times4}$, $p_2^T \in \mathbb{R}^{1\times4}$, $p_3^T \in \mathbb{R}^{1\times4}$

$$P = \begin{bmatrix} p_{11} & p_{12} & p_{13} & p_{14} \\ p_{21} & p_{22} & p_{23} & p_{24} \\ p_{31} & p_{32} & p_{33} & p_{34} \end{bmatrix}$$

M

Given P, we can find the location of the camera as follows:

$$P\vec{c} = \emptyset$$

* How do we estimate P?

$$w\begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = P\begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix}$$

DLT algorithm for estimating $P \in \mathbb{R}^{3\times4}$. P maps world (X, Y, Z) to image (u, v)

(i) We need a set of known mappings. $(X_i, Y_i, Z_i) \longrightarrow (u_i, v_i)$

(ii) Move 3D and 2D points such that their respective centroids is $(0,0,0)$ and $(0,0)$

(iii) Scale (shrink/grow) points such that average distance from origin for (u,v) points is $\sqrt{2}$ and average distance for (X,Y,Z) points is $\sqrt{3}$.

(iv) $\vec{x}_i = (u_i, v_i, 1)$ and $\vec{X}_i = (X_i, Y_i, Z_i, 1)$

(v) $$\underbrace{\vec{x}_i}_{\in \mathbb{R}^{3\times1}} = \underline{\underline{\lambda}} \underbrace{P\vec{X}_i}_{\in \mathbb{R}^{3\times1}}$$

(vi) $\vec{x}_i$ and $\lambda P\vec{X}_i$ are colinear. $\therefore$ $\vec{x}_i \underline{\underline{\times}} P\vec{X}_i = 0$

cross-product symbol

(vii) $P\vec{X}_i = \begin{bmatrix} \underline{\vec{P}_1^T \vec{X}_i} \\ \vec{P}_2^T \vec{X}_i \\ \vec{P}_3^T \vec{X}_i \end{bmatrix}$ *

$\vec{P}_1^T \vec{X}_i = P_{11}X_i + P_{12}Y_i + P_{13}Z_i + P_{14}$

(vii) $\vec{a} \times \vec{b} = \begin{bmatrix} a_2b_3 - a_3b_2 \\ a_3b_1 - a_1b_3 \\ a_1b_2 - a_2b_1 \end{bmatrix}$

(viii) $\vec{x}_i = \begin{bmatrix} u_i \\ v_i \\ 1 \end{bmatrix}$ *

(ix) $\vec{x}_i \times P\vec{X}_i = 0$

row 1: $\boxed{v_i(\vec{P}_3^T \vec{X}_i) - 1(\vec{P}_2^T \vec{X}_i) = 0}$ — $\times u_i$

row 2: $1(\vec{P}_1^T \vec{X}_i) - u_i(\vec{P}_3^T \vec{X}_i) = 0$ — $\times v_i$

row 3: $u_i(\vec{P}_2^T \vec{X}_i) - v_i(\vec{P}_1^T \vec{X}_i) = 0$

* we actually got only two equations. Discard 3rd row.

(x) we are trying to estimate P: $\underline{\underline{\vec{P}_1^T}}$, $\underline{\underline{\vec{P}_2^T}}$ and $\underline{\underline{\vec{P}_3^T}}$.

(xi) Re-write row 1 and row 2 to create a system of linear equations.

$$\begin{bmatrix} 0 & 0 & 0 & 0 & -X_i & -Y_i & -Z_i & -1 & v_iX_i & v_iY_i & v_iZ_i & v_i \\ X_i & Y_i & Z_i & 1 & 0 & 0 & 0 & 0 & -u_iX_i & -u_iY_i & -u_iZ_i & -u_i \end{bmatrix} \begin{bmatrix} | \\ P_1 \\ | \\ \hline | \\ P_2 \\ | \\ \hline | \\ P_3 \\ | \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

✓ ✓

?

$\boxed{A\vec{p} = \vec{\emptyset}}$

P has 12 elements. Therefore, P needs 12 equations.

Each mapping $(\vec{x}_i, \vec{X}_i)$ provides 2 equations.

$\vec{x}_i$ → image, $\vec{X}_i$ → world

∴ Need at least 6 mappings.

* In reality P can only be estimated upto a scale. So P has 11 unknowns. We need 5.5 point mappings.

* In practices we need more mappings and the points should not be colinear / coplanar.

* So say you have N mappings. You have setup the equation

$$A \vec{p} = \vec{0}$$

if $\vec{p}$ is a solution, than $2\vec{p}$, $100\vec{p}$, ... are also solutions.

Solve this system with constraint $\|\vec{p}\| = 1$

(i) Decompose $A = U \Sigma V^T$ (singular value decomposition)

(ii) Extract the last column of V. This is the solution.

This corresponds to the smallest eigenvalue.

$\boxed{P \checkmark}$

* We know that $P = K[R|t]$.

How to compute K, R, t from P.

(i) $P = [KR \mid Kt]$
$= [M \mid P_4]$

(ii) How to compute KR from M

We will use RQ decomposition. It decomposes M as a product of an upper-triangular matrix and an orthogonal matrix.

not unique
may have
-ve diagonals.

(iii) Estimate $\vec{\underline{t}}$ and then camera center is $\tilde{C} = -R^{-1}\vec{t}$