Robotics, Vision, & Mechatronics for Manufacturing

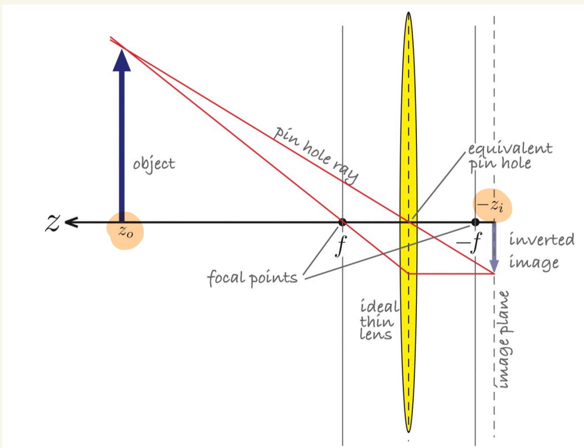
Image Formation

Xu Chen

2021 Spring

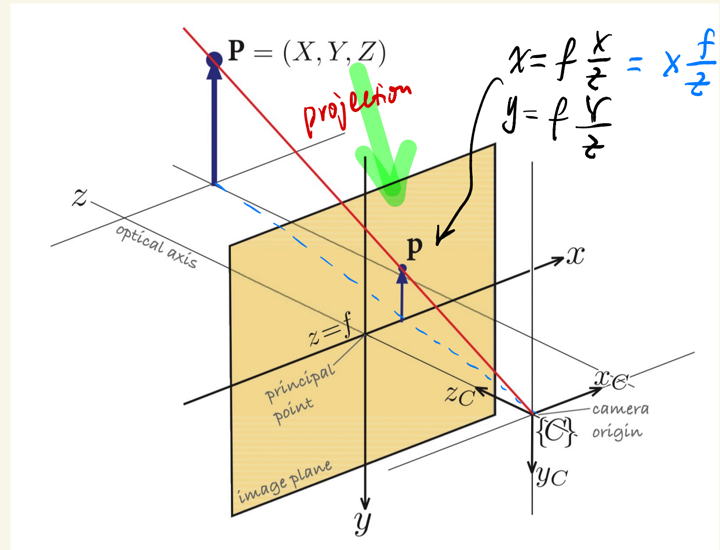
Perspective Camera

* fundamental geometry of image formation of a thin lens



$$\frac{1}{z_0} + \frac{1}{z_i} = \frac{1}{f}$$

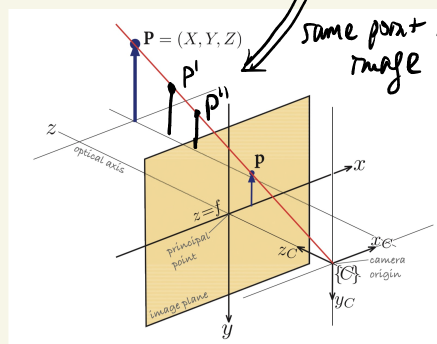
* the central perspective image model



Perspective Projection from the world to the image

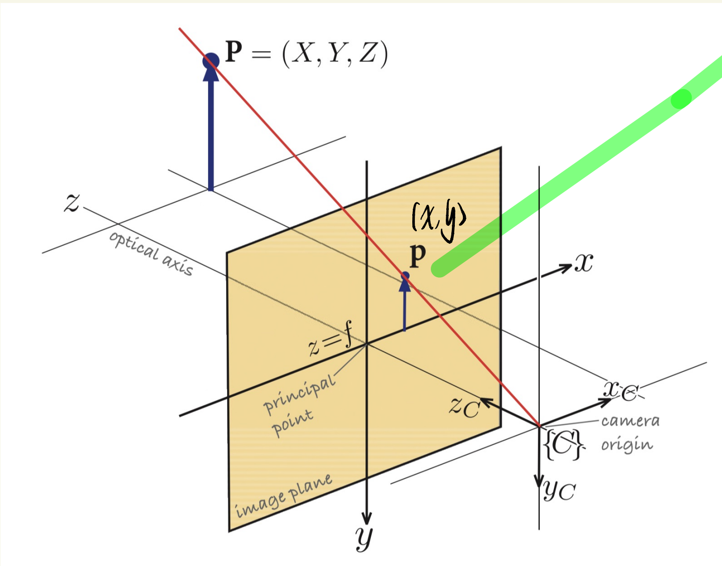
Properties:

* 3D → 2D
* parallel lines → intersecting lines
* circles → circles or ellipses
* Mapping is not one-on-one
  no unique inverse exists



same point in image

Modeling a Perspective Camera



* write $\underbrace{x = \frac{\tilde{x}}{\tilde{z}} \qquad y = \frac{\tilde{y}}{\tilde{z}}}_{\text{retinal/nonhomogeneous image-plane coordinates}}$

* define: $\left.\begin{aligned} \tilde{x} &= xZ = fX \\ \tilde{y} &= yZ = fY \end{aligned}\right\}$ homogeneous image-plane coordinates

$\tilde{z} = Z$

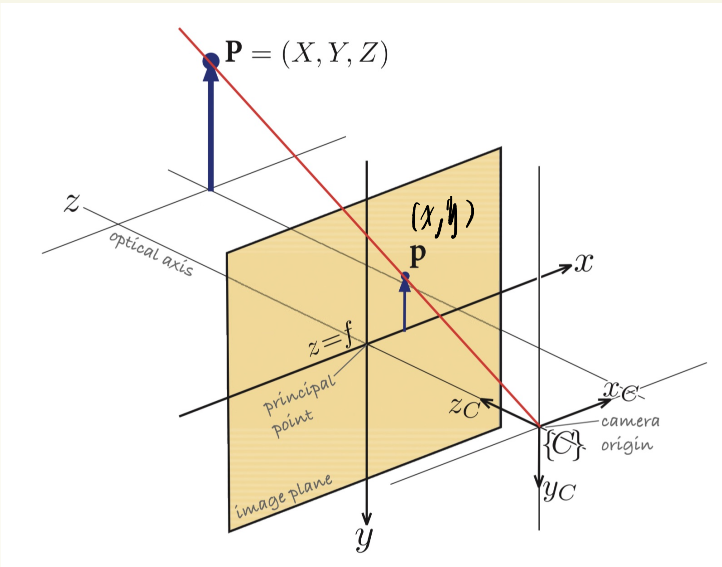
$$\Leftrightarrow \tilde{p}_i \triangleq \begin{bmatrix} \tilde{x} \\ \tilde{y} \\ \tilde{z} \end{bmatrix} = \begin{bmatrix} f & 0 & 0 \\ 0 & f & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \end{bmatrix}$$

* write coordinate in camera frame in homogeneous form too. ${}^{c}\tilde{p} = \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix}$

$$\Rightarrow \tilde{p}_i = \begin{bmatrix} \tilde{x} \\ \tilde{y} \\ \tilde{z} \end{bmatrix} = \underbrace{\begin{bmatrix} f & 0 & 0 & 0 \\ 0 & f & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix}}_{\text{the camera matrix: } C_i} {}^{c}\tilde{p}$$

$$= \begin{bmatrix} f & 0 & 0 \\ 0 & f & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} {}^{c}\tilde{p}$$

Modeling a Perspective Camera



Alternative approach:

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

"homogeneous coordinate" in image plane

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

"normalizing":

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

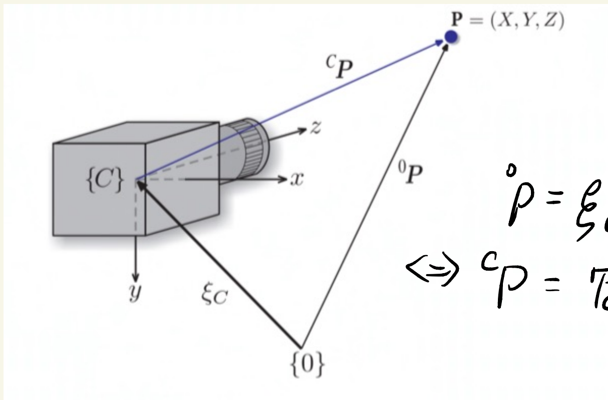
$$= \begin{bmatrix} f & 0 & 0 & 0 \\ 0 & f & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix}$$

homogeneous form in 3D.

Perspective projection

- $\tilde{p}_i = C_i \, {}^C\tilde{P} = \begin{bmatrix} f & 0 & 0 & 0 \\ 0 & f & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} {}^C\tilde{P}$ is a linear operation given f

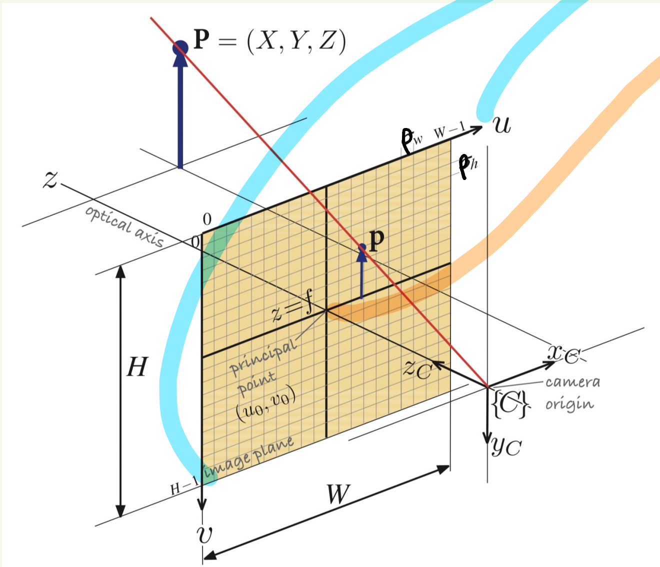
- To obtain ${}^C P$, we can use the usual pose composition:



$${}^0P = \xi_C \, {}^C P = T_C \, {}^C P$$

$$\Leftrightarrow \; {}^C P = T_C^{-1} \, {}^0 P$$

Pixels & Discrete Image Plane



pixel axis: index u & v are nonnegative

principal point: geometric center

(u_0, v_0)

⇒ coordinates in pixel domain:

$$u = \frac{x}{\underbrace{\rho_w}_{\text{pixel width}}} + u_0 \qquad v = \frac{y}{\underbrace{\rho_h}_{\text{pixel height}}} + v_0$$

$$\Rightarrow \begin{bmatrix} u \\ v \end{bmatrix} = \begin{bmatrix} \frac{1}{\rho_w} & 0 & u_0 \\ 0 & \frac{1}{\rho_h} & v_0 \end{bmatrix} \begin{bmatrix} x \\ y \\ 1 \end{bmatrix}$$

homogeneous form

$$\Rightarrow \begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = \begin{bmatrix} \frac{1}{\rho_w} & 0 & u_0 \\ 0 & \frac{1}{\rho_h} & v_0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ 1 \end{bmatrix}$$

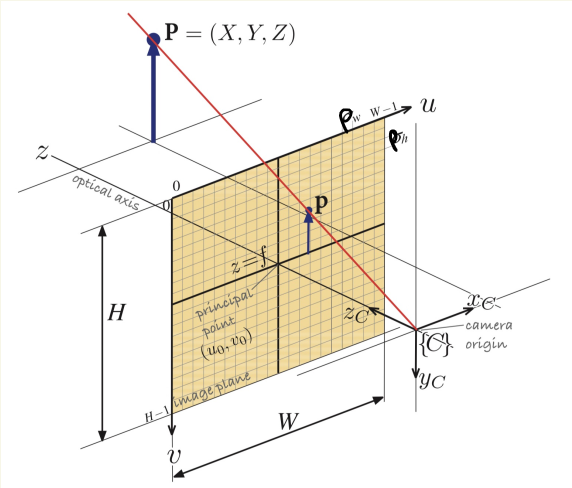
$$\begin{bmatrix} \tilde{x} \\ \tilde{y} \\ \tilde{z} \end{bmatrix} = \begin{bmatrix} f & 0 & 0 & 0 \\ 0 & f & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \overbrace{\begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix}}^{{}^C\tilde{P}}$$

Introduce again homogeneous notation:

$$u = \frac{\tilde{u}}{\tilde{w}}, \quad v = \frac{\tilde{v}}{\tilde{w}} \qquad \tilde{w} = Z$$

$$\xrightarrow{\text{normalizing}} \begin{bmatrix} \tilde{u} \\ \tilde{v} \\ \tilde{w} \end{bmatrix} \triangleq Z \begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = \underbrace{\begin{bmatrix} \frac{1}{\rho_w} & 0 & u_0 \\ 0 & \frac{1}{\rho_h} & v_0 \\ 0 & 0 & 1 \end{bmatrix}}_{K} \overbrace{Z \begin{bmatrix} x \\ y \\ 1 \end{bmatrix}} = K \begin{bmatrix} f & 0 & 0 & 0 \\ 0 & f & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} {}^C\tilde{P}$$

Camera Matrix Revisited (the full camera matrix from global coordinate to pixel domain)



We have:

– pixel domain:

$$\tilde{p} = \begin{bmatrix} \tilde{u} \\ \tilde{v} \\ \tilde{w} \end{bmatrix} = \underbrace{\begin{bmatrix} \frac{1}{\rho_w} & 0 & u_0 \\ 0 & \frac{1}{\rho_h} & v_0 \\ 0 & 0 & 1 \end{bmatrix}}_{K} \begin{bmatrix} f & 0 & 0 & 0 \\ 0 & f & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \underbrace{\begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix}}_{{}^{c}\tilde{P}}$$

– object in world frame:

$$\tilde{P} = {}^{0}T_{c}\, {}^{c}\tilde{P} \Rightarrow {}^{c}\tilde{P} = ({}^{0}T_{c})^{-1}\tilde{P}$$

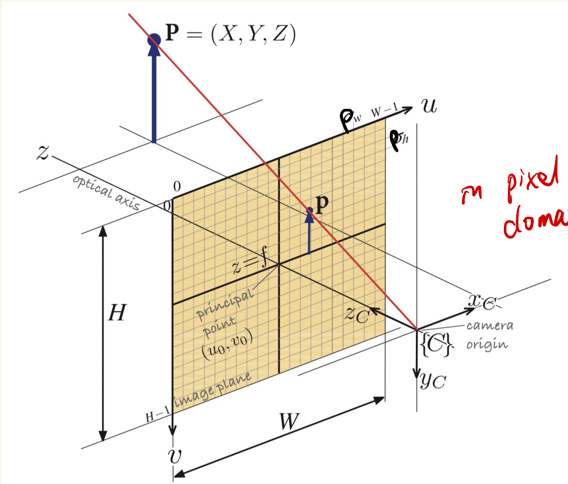
camera pose

$$\Rightarrow \tilde{p} = \underbrace{\underbrace{\begin{bmatrix} \frac{f}{\rho_w} & 0 & u_0 \\ 0 & \frac{f}{\rho_h} & v_0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix}}_{\text{intrinsic}} \underbrace{({}^{0}T_{c})^{-1}}_{\text{extrinsic}}}_{C} \tilde{P} = C\tilde{P}$$

in pixel domain

in world frame

Camera Matrix Revisited (the full camera matrix from global coordinate to pixel domain)



in pixel domain

$$\tilde{p} = \underbrace{\underbrace{\begin{bmatrix} \frac{f}{\rho_w} & 0 & u_0 \\ 0 & \frac{f}{\rho_h} & v_0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix}}_{\text{intrinsic : } 3\times 4} \underbrace{\left({}^{0}T_C\right)^{-1}}_{\text{extrinsic : } 4\times 4}}_{C : 3\times 4} \tilde{P} = C\tilde{P}$$

in world frame

Intrinsic parameters : $f, \rho_w, \rho_h, u_0, v_0$
(5 parameters)

Extrinsic parameters : homogeneous transformation matrix
(6 parameters)

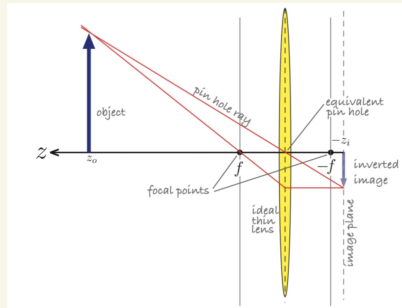
11 unknowns

on the other hand : $C : 3\times 4 \Rightarrow$ 12 parameters

one unconstrained parameter in $C_{3\times 4}$
usually the overall scaling can be arbitrarily chosen when identifying $C_{3\times 4}$

Lens Distortion

— No lenses are perfect.



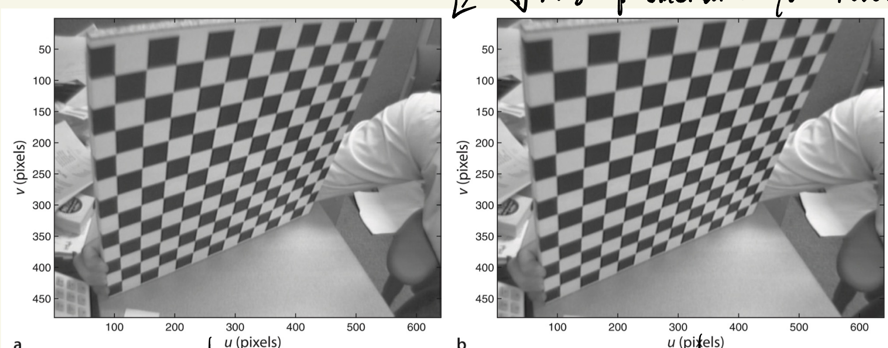
Distortion modeling:

$$\begin{cases} u^d = u + \delta_u \\ v^d = v + \delta_v \end{cases}$$

$$\begin{bmatrix} \delta_u \\ \delta_v \end{bmatrix} = \underbrace{\begin{bmatrix} u(k_1 r^2 + k_2 r^4 + k_3 r^6 + \dots) \\ v(k_1 r^2 + k_2 r^4 + k_3 r^6 + \dots) \end{bmatrix}}_{\text{radial}} + \underbrace{\begin{bmatrix} 2p_1 uv + p_2(r^2 + 2u^2) \\ p_1(r^2 + 2v^2) + 2p_2 uv \end{bmatrix}}_{\text{tangential}}$$

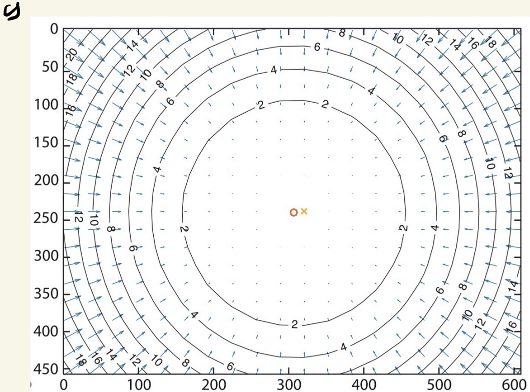
— common distorsions: geometric distortion, color fringing, variations in focus across the scene

↓ most problematic for robotics



distorted

undistorted



radial distortion

Camera Calibration :

$$C = \begin{bmatrix} \frac{f}{\rho_w} & 0 & u_0 \\ 0 & \frac{f}{\rho_h} & v_0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \left({}^{0}T_c\right)^{-1}$$

Many unknowns/variations :

u_0, v_0 : in general not at the center

f : accurate to 4% of nominal value

intrinsic parameters also change if remounted / lens is

Homogeneous Transformation Approach

photosite dimension ρ_w & ρ_h are usually accurate

$$\begin{cases} \tilde{p} = \begin{bmatrix} \tilde{u} \\ \tilde{v} \\ \tilde{w} \end{bmatrix} = C \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix} = \begin{pmatrix} C_{11} & C_{12} & C_{13} & C_{14} \\ C_{21} & C_{22} & C_{23} & C_{24} \\ C_{31} & C_{32} & C_{33} & C_{34} \end{pmatrix} \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix} \\ u = \frac{\tilde{u}}{\tilde{w}} \quad v = \frac{\tilde{v}}{\tilde{w}} \end{cases}$$

can measure

can measure

$$\Rightarrow \begin{cases} \tilde{u} = C_{11}X + C_{12}Y + C_{13}Z + C_{14} \\ \tilde{v} = C_{21}X + C_{22}Y + C_{23}Z + C_{24} \\ \tilde{w} = C_{31}X + C_{32}Y + C_{33}Z + C_{34} \end{cases}$$

$$\Rightarrow \begin{cases} u(C_{31}X + C_{32}Y + C_{33}Z + C_{34}) - (C_{11}X + C_{12}Y + C_{13}Z + C_{14}) = 0 \\ v(C_{31}X + C_{32}Y + C_{33}Z + C_{34}) - (C_{21}X + C_{22}Y + C_{23}Z + C_{24}) = 0 \end{cases}$$

$$\Rightarrow \begin{bmatrix} X & Y & Z & 1 & 0 & 0 & 0 & 0 & -uX & -uY & -uZ \\ 0 & 0 & 0 & 0 & X & Y & Z & 1 & -vX & -vY & -vZ \end{bmatrix} \begin{bmatrix} C_{11} \\ C_{12} \\ \vdots \\ C_{33} \end{bmatrix} = \begin{bmatrix} u \\ v \end{bmatrix}$$

a linear equation

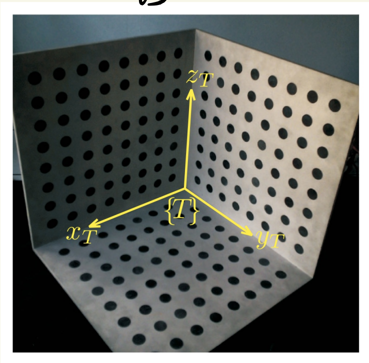
can take a lot of X, Y, Z, u, v to estimate C

Camera Calibration: Solution concept.

$$\begin{bmatrix} X & Y & Z & 1 & 0 & 0 & 0 & 0 & -uX & -uY & -uZ \\ 0 & 0 & 0 & 0 & X & Y & Z & 1 & -vX & -vY & -vZ \end{bmatrix} \begin{bmatrix} c_{11} \\ c_{12} \\ \vdots \\ c_{33} \end{bmatrix} = \begin{bmatrix} u \\ v \end{bmatrix}$$

take a lot of measurements

$\Rightarrow$



$$\begin{pmatrix} X_1 & Y_1 & Z_1 & 1 & 0 & 0 & 0 & 0 & -u_1X_1 & -u_1Y_1 & -u_1Z_1 \\ 0 & 0 & 0 & 0 & X_1 & Y_1 & Z_1 & 1 & -v_1X_1 & -v_1Y_1 & -v_1Z_1 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ X_N & Y_N & Z_N & 1 & 0 & 0 & 0 & 0 & -u_NX_N & -u_NY_N & -u_NZ_N \\ 0 & 0 & 0 & 0 & X_N & Y_N & Z_N & 1 & -v_NX_N & -v_NY_N & -v_NZ_N \end{pmatrix} \begin{bmatrix} c_{11} \\ c_{12} \\ \vdots \\ c_{33} \end{bmatrix} = \begin{bmatrix} u_1 \\ v_1 \\ \vdots \\ u_N \\ v_N \end{bmatrix}$$