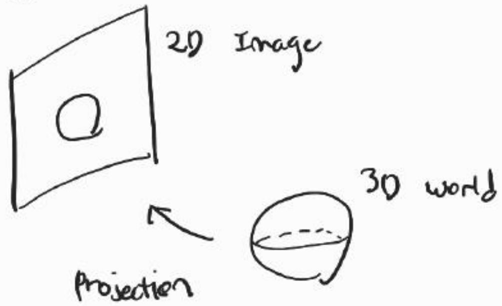
• Camera



In previous lectures, learn projective transformations of

$P^2 \to P^2$ or $P^3 \to P^3$

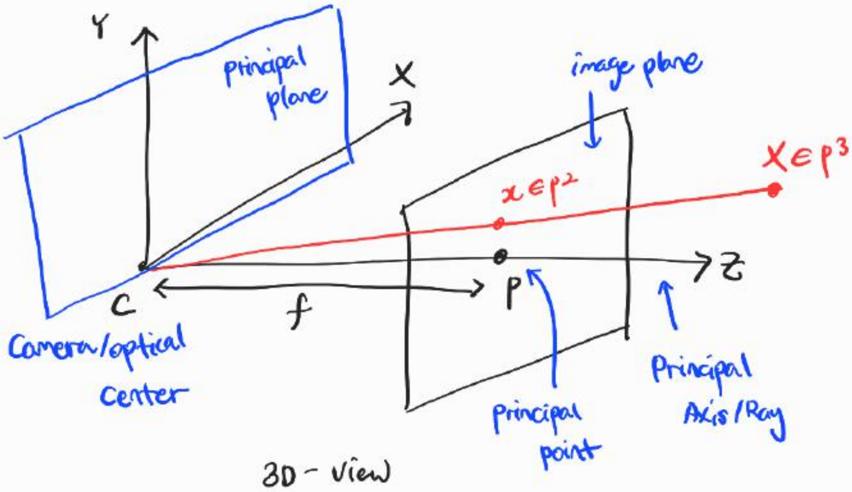
Camera : $P^3 \to P^2$

• Camera models

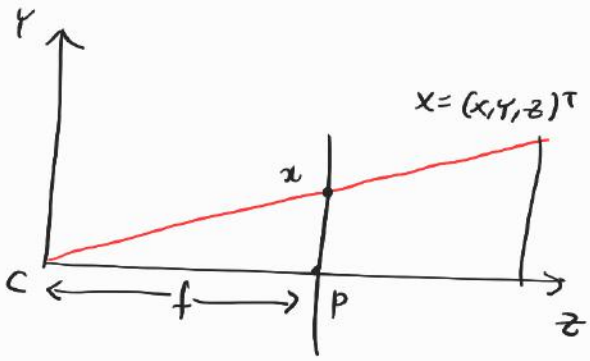
Camera model w/ central projection
⇒ all rays converge at single center

[finite center - Projective camera
[center at "infinity" (π_∞) - affine camera

• Projective camera : Basic pinhole model



3D - view



2D (YZ plane) - view

Calculate x coordinates using similar triangle b/w ΔCpx and ΔCZX :

$$\frac{y}{f} = \frac{Y}{Z} \rightarrow y = \frac{fY}{Z} \quad , \text{ similarly } \quad x = \frac{fX}{Z}$$

∴ mapping from $(X, Y, Z)^T \rightarrow \left(\frac{fX}{Z}, \frac{fY}{Z}, f\right)^T$

or ignoring final coordinate, $(X, Y, Z)^T \rightarrow \left(\frac{fX}{Z}, \frac{fY}{Z}\right)^T \quad (\mathbb{R}^3 \to \mathbb{R}^2)$

In homogenous coordinates, above central projection becomes

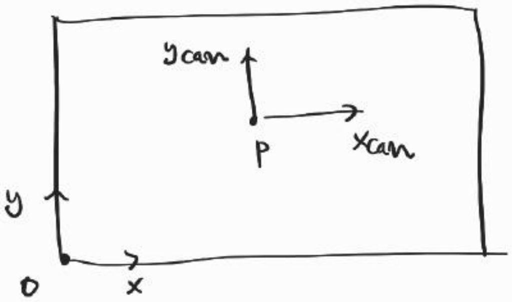
$$\begin{pmatrix} X \\ Y \\ Z \\ 1 \end{pmatrix} \rightarrow \begin{pmatrix} fX \\ fY \\ Z \end{pmatrix} = \begin{bmatrix} f & & & 0 \\ & f & & 0 \\ & & 1 & 0 \end{bmatrix} \begin{pmatrix} X \\ Y \\ Z \\ 1 \end{pmatrix}$$

or $x = PX$ where $P = \mathrm{diag}(f, f, 1)\,[I \mid 0]$

P : 3x4 homogenous projection matrix

- Principal point offset

Usually, origin of image plane $\neq$ principal point



$\therefore$ $p = (p_x, p_y)^T$ r.t. image frame

$$\therefore \begin{pmatrix} X \\ Y \\ Z \\ 1 \end{pmatrix} \rightarrow \begin{pmatrix} fX + Zp_x \\ fY + Zp_y \\ Z \end{pmatrix} = \begin{bmatrix} f & & p_x & 0 \\ & f & p_y & 0 \\ & & 1 & 0 \end{bmatrix} \begin{pmatrix} X \\ Y \\ Z \\ 1 \end{pmatrix}$$

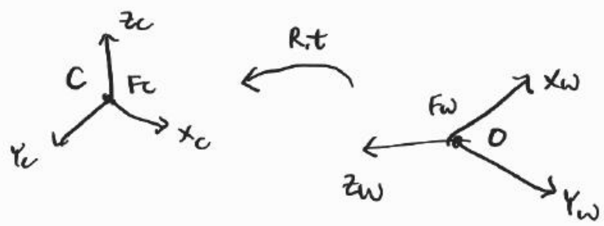
- Camera Calibration Matrix

$\updownarrow$ same expression

$$K = \begin{bmatrix} f & & p_x \\ & f & p_y \\ & & 1 \end{bmatrix} \quad \text{where} \quad x = K[I \mid 0] X_{cam}$$

- Camera Rotation and Translation

Camera coordinate frame $\neq$ world coordinate frame



two frames related by rigid transformation (R,t)

$$X_{cam} = \begin{bmatrix} R & -RC \\ 0 & 1 \end{bmatrix} \begin{pmatrix} X \\ Y \\ Z \\ 1 \end{pmatrix} = \underbrace{\begin{bmatrix} R & -RC \\ 0 & 1 \end{bmatrix}}_{\text{4x4 isometric transformation}} X_{world}$$

where $t = -RC$ and C is camera center in world coordinates

- World $\rightarrow$ camera $\rightarrow$ Image

$$x = K[I \mid 0] X_{cam}$$

$$= K[I \mid 0] \begin{bmatrix} R & -RC \\ 0 & 1 \end{bmatrix} X_{world}$$

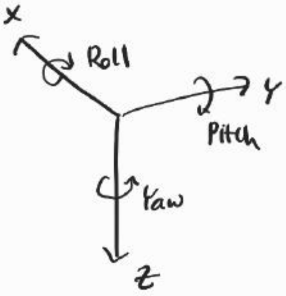
$$= \underset{3\times3}{K} \underset{3\times4}{[R \mid -RC]} \underset{4\times1}{X_{world}}$$

$3\times1 \leftarrow$

- Euler Angles to Rotation Matrix

  Represent rotations in 3 dof

  i.e. 3X3 matrix representions are overparameterized!

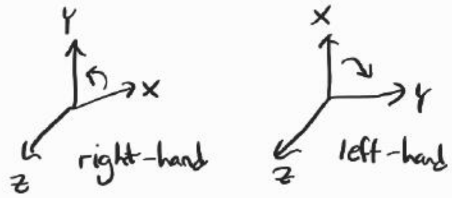


- properties of Rotation matrix

  1) square matrix (2d - 2x2 w/ 1dof, 3d - 3x3 w/ 3dof)

  2) orthonormal matrix

  (i) $\det(R)$ $\begin{cases} 1 & \text{if right-hand coordinate frame} \\ -1 & \text{if left-hand} \quad \text{—} \end{cases}$



  ∵ Rx change the area/volume spanned by basis vectors

  (ii) $R^T = R^{-1}$

  (iii) $r_i \times r_j = r_k$ where $R = [r_i, r_j, r_k]$

  (iv) $r_i^T r_j = 0$

  (v) $\|r_i\| = \|r_j\| = \|r_k\| = 1$

- The basic pinhole model

  Previously, $x = K[R|-RC]X_{world}$ as world $(P^3) \rightarrow$ image (P^2) mapping

  Then, $x = KR[I|-C]X$ where $P = \underbrace{K}_{\substack{\text{Intrinsic} \\ 3\times3}}\underbrace{R[I|-C]}_{\substack{\text{Extrinsic} \\ 3\times4}}$ $\rightsquigarrow$ has 9dof

  (K(3), R(3), C or t(3))

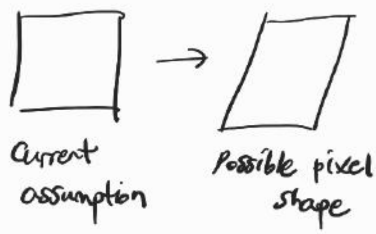
  or similarly,

  $P = K[R|t]$ where $t = -RC$

- Non-square and skewed pixels

  Pixels might be non-square and skewed in real cameras

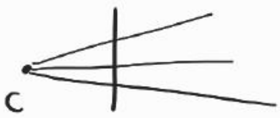
  non-square: $f_x \neq f_y$ skewed: s (x-direction)



  $\therefore K = \begin{bmatrix} f_x & s & p_x \\ 0 & f_y & p_y \\ 0 & 0 & 1 \end{bmatrix}$ w/ 5 dof

  ∴ P has 11 dof $(K(5) + R(3) + t/c(3))$

- Finite Projective cameras



$P = KR[I|-C] = M[I|M^{-1}p_4]$ where $P = [p_1, p_2, p_3, p_4]$ and

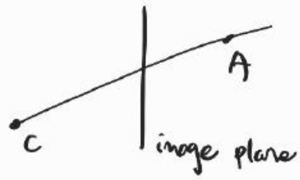
$M = [p_1, p_2, p_3]$ non-singular matrix

- Camera center

  $PC = 0$: projection of center is undefined point $(0,0,0)^T$

  $\because$ Suppose line $X(\lambda) = \lambda A + (1-\lambda)C$

  All points on the line project to PA



  $\Rightarrow x = PX(\lambda) = \lambda PA + (1-\lambda)PC = PA$

  $\because PC = 0$

- Column vectors of P

  for $P = [p_1, p_2, p_3, p_4]$
  - p_1, p_2, p_3 are vanishing points of X, Y, Z
  - p_4 is image/projection of world origin

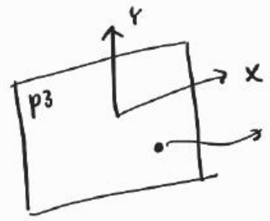
  e.g. $p_3 = P\begin{pmatrix}0\\0\\1\\0\end{pmatrix}$: Projection of z-axis $(0,0,1)^T$ direction of vanishing point

  $p_4 = P\begin{pmatrix}0\\0\\0\\1\end{pmatrix}$: Projection of world origin $(0,0,0,1)^T$

- Row vectors of P

  For $P = \begin{bmatrix}p^{1T}\\p^{2T}\\p^{3T}\end{bmatrix}$, each row represents a particular world plane

  (i) Principal plane $= p^{3T}$



  Every point X in principal point project to line of infinity

  $\Rightarrow PX = (x, y, 0)^T$

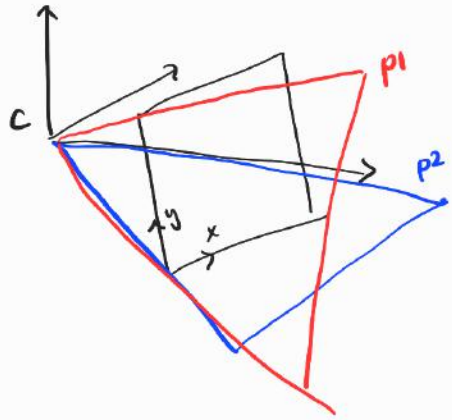
  $$\begin{bmatrix}p_1^T\\p_2^T\\p_3^T\end{bmatrix} X = \begin{bmatrix}x\\y\\0\end{bmatrix}$$

  $\therefore p_3^T X = 0$
  incidence relation b/w plane p_3 & point X

  (ii) Axis plane $= p_1^T, p_2^T$

  p^1 : plane defined by C and $x = 0$

  p^2 : ——— C and $y = 0$



$\because$ Set of points X on plane p^2 should satisfy

$p^{2T}X = 0$

Also, $PX = (x, 0, w)^T$ as $y=0$ for $p2$ plane

$\Leftrightarrow$ second row of P ($= p2$) should be 0

$PC = 0 \Rightarrow p^{2T}C = 0$

$\therefore C$ also lies on p^2 plane

- The principal point

  $\Rightarrow$ intersection of principle axis and image plane

  = normal of principle plane

  = direction vector of $P3$

  $= (P_{31}, P_{32}, P_{33})^T \leftarrow$ can also treat as ideal point

  $x_0 = PP_3$: point P_3 projects to image at principal point x_0

  ↳ or similarly, $x_0 = Mm^3$ where $P = [M \mid p_4]$ and m^{3T} is third row of M

- The principle axis vector

  In theory, $x = PX$ where any 3D points X projects to 2D image plane

  ↓

  In reality, only half of points lie front of camera can project

  $V = \det(M)\, m^3$ is direction of principle axis to the front of camera

  ↳ original principle vector

  ↳ direction of principle axis : signed area equivalent

- Action of a projective camera on points

  - Forward projection

    maps point in space $X \in \mathbb{P}^3$ to image point $x \in \mathbb{P}^2$: $x = PX$

    for vanishing point $D = (d^T, 0)^T$,

    $x = PD = [M \mid p_4]D = Md$ $\quad\because$ Last element of $D = 0$

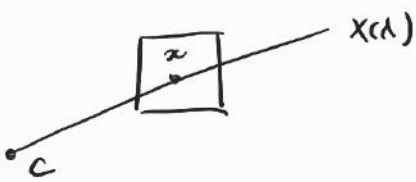
    $\therefore$ only affected by M

  - Back-projection

    maps image point to ray : $X(\lambda) = P^+x + \lambda C$

    Pseudo-inverse ↵

    $\because P$ is non-invertible



$X(\lambda)$ is join of 2 points $\left[\begin{array}{l}\text{Camera center } C \quad s.t.\ PC=0 \\ \text{Point } P^{+}x = P^{T}(PP^{T})^{-1}\end{array}\right.$

for finite camera, line can be expressed as :

$$X(\mu) = \mu \underbrace{\begin{pmatrix} M^{-1}x \\ 0 \end{pmatrix}}_{\mu \cdot \text{ ideal point}} + \underbrace{\begin{pmatrix} M^{-1}p_4 \\ 1 \end{pmatrix}}_{\text{Camera center}} = \begin{pmatrix} M^{-1}(\mu x - p_4) \\ 1 \end{pmatrix}$$

= dir of C to x

$\because P\begin{pmatrix} M^{-1}x \\ 0 \end{pmatrix} = [M|p_4]\begin{pmatrix} M^{-1}x \\ 0 \end{pmatrix} = x$

↳ Any point on direction $M^{-1}x$ projects to x on image plane

$\longrightarrow PC = [M|p_4]\begin{pmatrix} M^{-1}p_4 \\ 1 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}$

$\therefore M^{-1}p_4$ is inhomogeneous camera center

- Depth of points

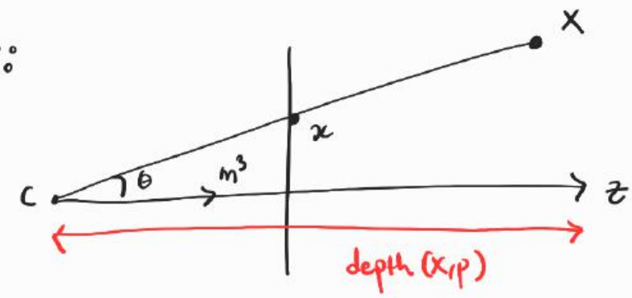
Let $P = [M|p_4]$ projects 3D point $X = (X, Y, Z, T)^T$ to $x = w(x, y, 1)^T$

$P(X, Y, Z, T)^T = w(x, y, 1)^T$

Then, the depth of X is :

$$\text{depth}(X, P) = \frac{\text{sign}(\det M)\, w}{T\|m^3\|}$$

$\because$



3D point : $X = (X, Y, Z, 1)^T = (\tilde{X}^T, 1)^T$

Center : $C = (\tilde{C}, 1)^T$

Image point : $x = w(x, y, 1)^T = PX$

$w = p^{3T}X$ $\quad \because$ Last element of $x = w$

$= p^{3T}(X - C)$ $\quad \because PC = 0$

$= m^{3T}(\tilde{X} - \tilde{C})$ $\quad \because$ homogeneous → non-homogeneous

$= \|m^{3T}\|\,\|\tilde{X} - \tilde{C}\| \cos\theta$ $\quad \because$ Dot product definition

$= \text{sign}(\det M)\, w$ $\quad \because$ considers the sign of 3D point X

$$\therefore \text{depth}(X, P) = \|\tilde{X} - \tilde{C}\| \cos\theta = \frac{\text{sign}(\det M)\, w}{\|m^{3T}\|}$$

- Decomposition of camera matrix

Given P, decompose into
- K : f_x, f_y, s, p_x, p_y
- R : roll, pitch, yaw
- $\tilde{C}$: $\tilde{c}_x, \tilde{c}_y, \tilde{c}_z$

- Finding camera center C

C = null-space of P = intersecting point of axis plane (p^{1T}, p^{2T}) & principle plane (p^{3T})

$\therefore$ Take $SVD(P) = U\Sigma V^T$,

then C = orthogonal vector v w/ least singular value

- Finding K, R

$P = [M \mid -M\tilde{C}] = K[R \mid -R\tilde{C}]$, then take $RQ(M) = KR$
$\downarrow$
RQ decomposition = (upper triangular) x (orthogonal)

- Euclidean vs. projective spaces

Camera model
- World / camera coordinate systems : Euclidean space
- camera itself : projective mapping $P^3 \to P^2$

$\therefore$ Projective camera can be viewed as

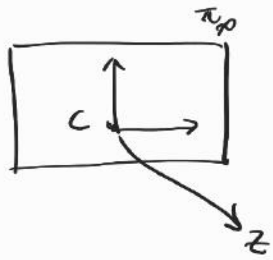
$$P = [3\times3 \text{ homography}] \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} [4\times4 \text{ homography}]$$

$\downarrow$ $P^2 \to P^2$ $\downarrow$ $P^3 \to P^2$ (image) $\downarrow$ $P^3 \to P^3$

- Cameras at infinity : Affine Camera

Affine Camera is in form $P_A = \begin{bmatrix} m_{11} & m_{12} & m_{13} & t_1 \\ m_{21} & m_{22} & m_{23} & t_2 \\ 0 & 0 & 0 & 1 \end{bmatrix}$

↳ Camera center lies on plane of infinity



$\therefore$ $p^{3T} = (0,0,0,1)^T$ infers principal plane is at π_∞

$C = (d, 0)^T$ is ideal point, thus $PC = 0$ is $M_{2\times3}d = 0$

$[M \mid p_4]\begin{bmatrix} d \\ 0 \end{bmatrix}$

* left 3x3 block of P_A is singular $\because$ last row $=0$ (linearly dependent)

- Decomposition of affine camera

$$P_A = [3\times3 \text{ affine}] \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} [4\times4 \text{ affine}]$$

affine e.g. $X_\infty \mapsto X_\infty$

orthographic projection

$P^3 \to P^2$

OR,

$$P_A = \begin{bmatrix} K_{2\times2} & \tilde{x}_0 \\ \hat{0}^T & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} R & t \\ 0^T & 1 \end{bmatrix}$$

calibration matrix

Rigid transformation

$$= \begin{bmatrix} \alpha_x & s & p_x \\ 0 & \alpha_y & p_y \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} r^{1T} & t_1 \\ r^{2T} & t_2 \\ 0 & 1 \end{bmatrix}$$

where $(p_x, p_y)^T = (0,0)$ usually b/c there is no projective distortion

$r^{3T} = 0,\ t_3 = 1$

P_A has 8 dof : $3(\alpha_x, \alpha_y, s) + 3(\text{rotation}) + 2(t_1, t_2)$

3 because original rotation aren't linearly independent i.e. $r_1 \times r_2 = r_3$

Also, $P_A = \begin{bmatrix} m_{11} & m_{12} & m_{13} & t_1 \\ m_{21} & m_{22} & m_{23} & t_2 \\ 0 & 0 & 0 & 1 \end{bmatrix}$ has rank $=2$ ($*$ rank$(P_p) = 3$)

- Affine properties of Camera at infinity

(i) Plane at infinity (P^3) maps into points at infinity (P^2) at the image

$\because P_A(X, Y, Z, 0)^T = (X, Y, 0)^T$

(ii) Parallel world lines (P^3) maps into parallel image lines (P^2)

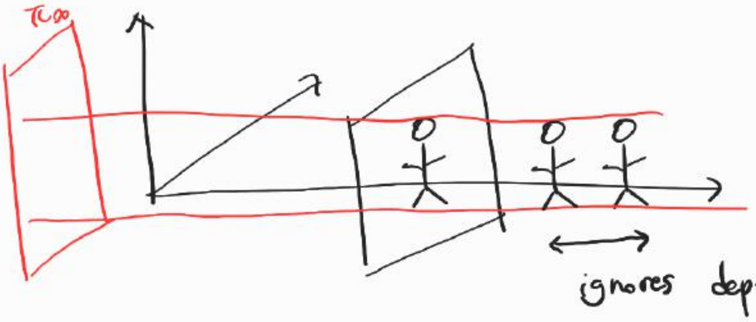
$\because$ Plane of infinity $\rightarrow$ point of infinity (from (i))

$\quad$ || $\qquad\qquad\qquad\qquad$ ||

intersection of parallel world lines $\quad$ intersection of parallel image lines

- Hierarchy of affine cameras

1) orthographic projection



× change in scale → $K = I$

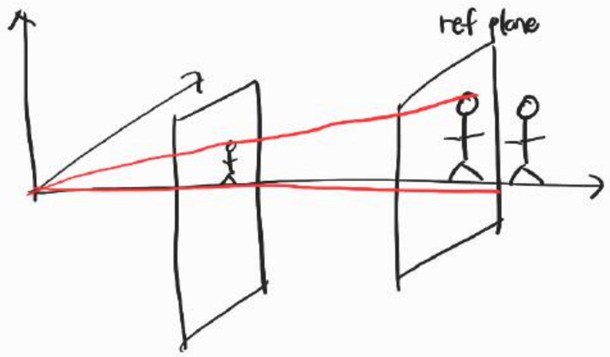
Parallel rays → camera center at π_∞

$$P = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} R & t \\ 0^T & 1 \end{bmatrix} = \begin{bmatrix} r_1^T & t_1 \\ r_2^T & t_2 \\ 0^T & 1 \end{bmatrix}$$

5 dof

$r^1 \perp r^2$, $\|r^1\| = \|r^2\|$

2) Scaled orthographic projection



Orthographic + perspective (Scaling)

$$P = \begin{bmatrix} k & & \\ & k & \\ & & 1 \end{bmatrix} \begin{bmatrix} r^{1T} & t_1 \\ r^{2T} & t_2 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} r^{1T} & t_1 \\ r^{2T} & t_2 \\ 0 & 1/k \end{bmatrix}$$

6 dof

3) Weak perspective projection

$$P = \begin{bmatrix} \alpha_x & & \\ & \alpha_y & \\ & & 1 \end{bmatrix} \begin{bmatrix} r^{1T} & t_1 \\ r^{2T} & t_2 \\ 0 & 1 \end{bmatrix}$$

7 dof

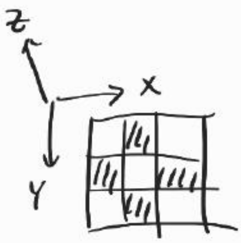
$\alpha_x \neq \alpha_y$, $r^1 \perp r^2$, but × required $\|r_1\| = \|r_2\|$

- Calibration of Projective camera

Finding 11 dof of $P = K[R\ t]$ → true calibration / camera resectioning

Commonly by 2D calibration pattern e.g. checkerboard

refered to Zhang's method



For simplicity, let $Z = 0$

↳ checkerboard on XY plane

$x = PX \rightarrow s\begin{bmatrix} x \\ y \\ 1 \end{bmatrix} = K[r_1\ r_2\ r_3\ t]\begin{bmatrix} X \\ Y \\ 0 \\ 1 \end{bmatrix}$ $\because z = 0$

$= K[r_1\ r_2\ t]\begin{bmatrix} X \\ Y \\ 1 \end{bmatrix}$ $\because$ As $z=0$, r_3 can be ignored

mapping from $\begin{bmatrix} X \\ Y \\ 1 \end{bmatrix}$ to $\begin{bmatrix} x \\ y \\ 1 \end{bmatrix}$ is $P^2 \mapsto P^2$

$\therefore$ If $s\underline{[h_1\ h_2\ h_3]} = K[r_1\ r_2\ t]$

↳ homography H as 3D-2D $(X,Y)-(x,y)$ corresponds to $P^2 \to P^2$ mapping

* known parameters : Image point (x,y) and world point (X,Y)

Unknown parameters : Homography H, scale s

↓ Find s, H using rotation properties

From $s[h_1\ h_2\ h_3] = K[r_1\ r_2\ t]$, $sK^{-1}h_1 = r_1$ and $sK^{-1}h_2 = r_2$

Using two properties
$$\left[\begin{array}{ll} r_1^T r_2 = 0 & \Rightarrow h_1^T K^{-T} K^{-1} h_2 = 0 \\ \|r_1\| = \|r_2\| & \Rightarrow h_1^T \underline{K^{-T}K^{-1}} h_1 = h_2^T K^{-T} K^{-1} h_2 \\ (r_1^T r_1 = r_2^T r_2) & \end{array}\right.$$

↳ Image of absolute conic

Let $K^{-T}K^{-1} = B = \begin{bmatrix} - B_1 - \\ - B_2 - \\ - B_3 - \end{bmatrix}$ → since B is symmetric (A^TA form) and positive definite
$B = [B_{11}, B_{12}, B_{13}, B_{22}, B_{23}, B_{31}]^T$ (i.e. 6 dof)

Then, $h_1^T B h_2 = (h_1^T h_2) B = 0$

$h_1^T B h_1 - h_2^T B h_2 = (h_1^T h_1 - h_2^T h_2) B = 0$] stack into linear constraint $ab = 0$

homography terms (2x6) → a ; unknown K (6x1) → b

$\therefore$ Each 2D-3D correspondences gives h_1, h_2 $(=a)$

checkerboard (world) Image

Requires at least 3 different views

$\therefore a \in R^{2(\#\text{ of views}) \times 6}$

↳ ≥ 6 to solve 6 unknowns of b

Requires 4 point correspondences to calculate $H(h_1, h_2)$ using $\begin{pmatrix} x \\ y \\ 1 \end{pmatrix} = H\begin{pmatrix} X \\ Y \\ 1 \end{pmatrix}$

$\therefore$ for $n \geq 3$ views, $Ab = 0$ where $A \in \mathbb{R}^{2n \times 6}$

Solve least-square problem by (i) Take $SVD(A) = U \Sigma V^T$

(ii) solution b is null-space of A

$\Leftrightarrow$ vector v w/ least singular value

In practice, use more than 3 views.

$\because$ Noise exists, Ab cannot be exactly 0

- Calculate K

1) Convert $B \in \mathbb{R}^{6 \times 1}$ into $B \in \mathbb{R}^{3 \times 3}$

2) Use cholesky decomposition : $cholesky(A) = LL^T$

since $B = K^{-T}K^{-1} = (K^TK)^{-1}$ $\therefore$ $cholesky(B^{-1}) = K^TK$

$\Rightarrow f_x, f_y, K_s, P_x, P_y$

- Calculate s, R, t

1) since $r_1 = sK^{-1}h_1$, $r_2 = sK^{-1}h_2$, find s using $s = \frac{1}{\|K^{-1}h_1\|} = \frac{1}{\|K^{-1}h_2\|}$ $\because \|r_1\| = \|r_2\| = 1$

2) Calculate r_1, r_2 using above property

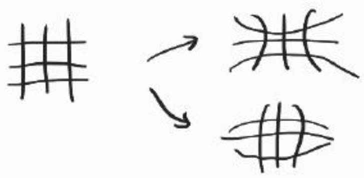
3) Calculate r_3 using the property $r_3 = r_1 \times r_2$

4) Calculate t : $t = sK^{-1}h_3$

$*$ s is fixed for all different views / R, t are different

- Lens distortion

1) Radial distortion (common)



Negative radial distortion / pincushion

Positive radial distortion / barrel

$$x_r \text{ (distorted)} = \begin{bmatrix} x_r \\ y_r \end{bmatrix} = (1 + K_1 r^2 + K_2 r^4 + K_5 r^6) \begin{bmatrix} x \\ y \end{bmatrix}$$

where $r^2 = x^2 + y^2$

k_1, k_2, k_5 are unknown parameters

$\Rightarrow x_{distort} = f(x_{undistort})$

2) Tangential distortion (not common)

Caused by poor manufacturing - camera sensors and lens aren't parallel

$$dx = \begin{bmatrix} 2k_3xy + k_4(r^2 + 2x^2) \\ k_3(r^2 + 2y^2) + 2k_4xy \end{bmatrix}$$

where $r^2 = x^2 + y^2$

k_3, k_4 are unknown parameters

$$\therefore x_d = x_r + dx = (1 + K_1 r^2 + K_2 r^4 + K_5 r^6) \begin{bmatrix} x \\ y \end{bmatrix} + \begin{bmatrix} 2k_3xy + k_4(r^2 + 2x^2) \\ k_3(r^2 + 2y^2) + 2k_4xy \end{bmatrix}$$

- Maximum Likelihood Estimation

  1) Estimate intrinsic/extrinsic Parameters w/o distortions

  2) Initialize all lens distortion parameters ($k_1 - k_5$) to 0

     ∵ These values are usually very small → 0 is good starting point!

  3) Minimize total reprojection error (geometric error)

$$\underset{K,R_i,t_i,X}{\operatorname{argmin}} \sum_{i=1}^{n}\sum_{j=1}^{m} \|x_{ij} - \pi(K, R_i, t_i, k, X_j)\|^2 \rightarrow$$

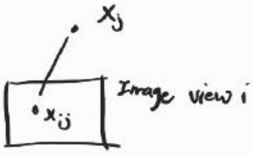


Image point

Projection of world point to the image

⇒ Continuous unconstrained optimization solved by e.g. Levenberg - Marquardt method

* More accurate than algebric error e.g. Solve $Ab = 0$ previously

- Lens distortion correction