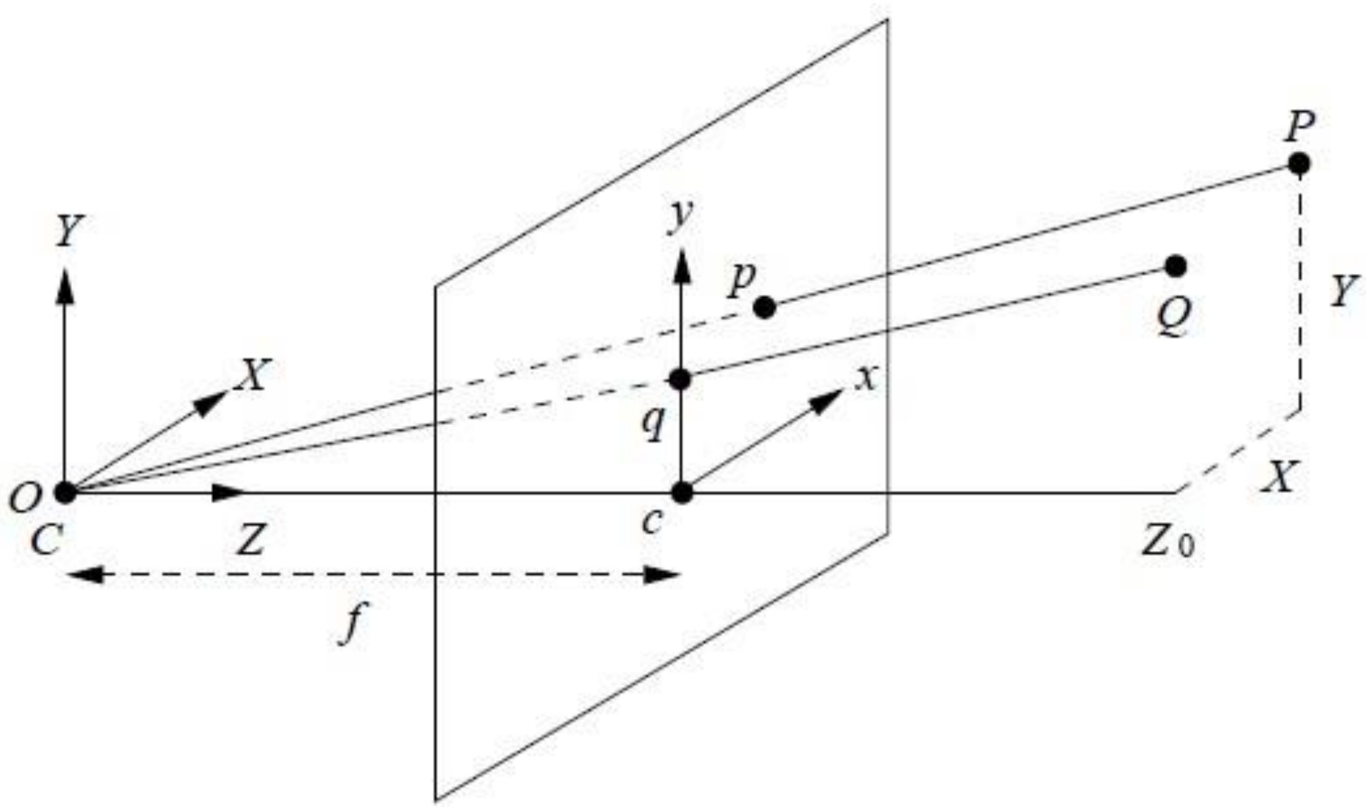
Calibration and Stereo Vision
Reminder lesson

Prof. Hichem Sahli
Abel Díaz Berenguer
VUB-ETRO
2022-2023

Scaled Orthographic Projection

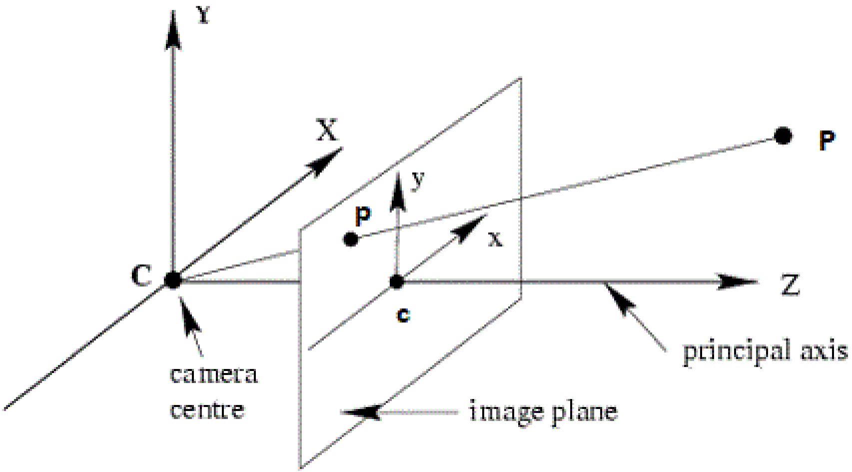


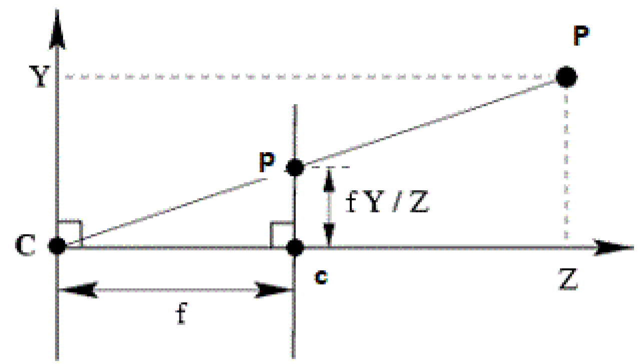
Scene depth << distance to camera.
Z is the same for all scene points, say Z_0

$$x = sX, \quad y = sY, \quad s = \frac{f}{Z_0} \text{ for all scene points.}$$

Pinhole Camera Model

3D point $\mathbf{P} = (X, Y, Z)^T$ projects to 2D image point $\mathbf{p} = (x, y)^T$.





$$\frac{X}{Z} = \frac{x}{f}, \quad \frac{Y}{Z} = \frac{y}{f}$$

$$x = f\frac{X}{Z}, \quad y = f\frac{Y}{Z}$$

Simplest form of perspective projection -> pinhole camera model => Pixel coordinates

Camera Intrinsic Parameters

- Pinhole camera model

$$x = f\frac{X}{Z}, \quad y = f\frac{Y}{Z}$$

- Camera sensor's pixels not exactly square

$$x = kf\frac{X}{Z}, \quad y = lf\frac{Y}{Z}$$

  - x, y : coordinates (pixels)
  - k, l : scale parameters (pixels/m)
  - f : focal length (m or mm)

Intrinsic Parameters

- f, k, l are not independent.
- Can rewrite as follows (in pixels):

$$f_x = kf \qquad f_y = lf$$

- So,

$$x = f_x \frac{X}{Z}, \quad y = f_y \frac{Y}{Z}$$

- Image centre or principal point c may not be at origin.
- Denote location of c in image plane as c_x, c_y.
- Then,

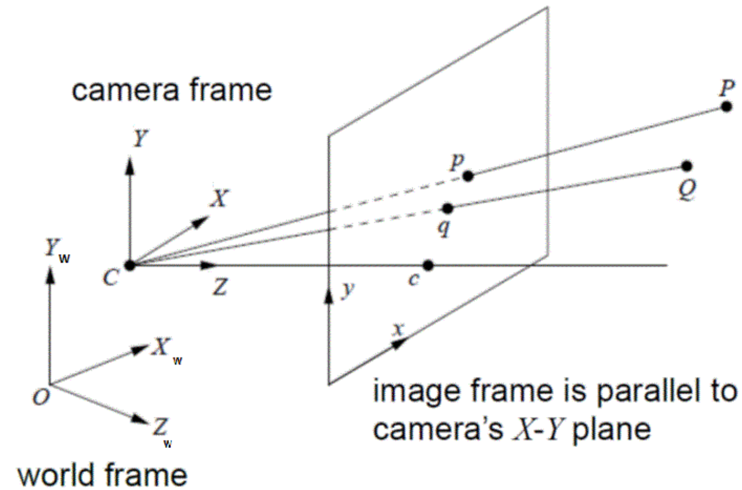
$$x = f_x \frac{X}{Z} + c_x, \quad y = f_y \frac{Y}{Z} + c_y$$

- Along optical axis, $X = Y = 0$, and $x = c_x$, $y = c_y$.

Extrinsic Parameters

- Translation matrix

$$\mathbf{T} = \begin{bmatrix} T_X \\ T_Y \\ T_Z \end{bmatrix}$$



- Rotation matrix

$$\mathbf{R} = \begin{bmatrix} r_{11} & r_{12} & r_{13} \\ r_{21} & r_{22} & r_{23} \\ r_{31} & r_{32} & r_{33} \end{bmatrix}$$

Rotation Matrix

R may be decomposed into 3 rotations about the coordinate axes:
$R = R_x R_y R_z$

If rotations are performed in the above order:

1) γ = rotation angle about z-axis
2) β = rotation angle about (new) y-axis
3) α = rotation angle about (new) x-axis

("tilt angle", "pan angle", and "nick angle" for the camera coordinate assignment shown before)

$$R_x = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos\alpha & \sin\alpha \\ 0 & -\sin\alpha & \cos\alpha \end{bmatrix}$$

$$R_y = \begin{bmatrix} \cos\beta & 0 & -\sin\beta \\ 0 & 1 & 0 \\ \sin\beta & 0 & \cos\beta \end{bmatrix}$$

$$R_z = \begin{bmatrix} \cos\gamma & \sin\gamma & 0 \\ -\sin\gamma & \cos\gamma & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

Rotation Matrix

By multiplying the 3 matrices R_x, R_y and R_z, one gets

$$R = \begin{bmatrix} \cos\beta\cos\gamma & \cos\beta\sin\gamma & -\sin\beta \\ \sin\alpha\sin\beta\cos\gamma - \cos\alpha\sin\gamma & \sin\alpha\sin\beta\sin\gamma + \cos\alpha\cos\gamma & \sin\alpha\cos\beta \\ \cos\alpha\sin\beta\cos\gamma + \sin\alpha\sin\gamma & \cos\alpha\sin\beta\sin\gamma - \sin\alpha\cos\gamma & \cos\alpha\cos\beta \end{bmatrix}$$

For formula manipulations, one tries to avoid the trigonometric functions and takes

$$R = \begin{bmatrix} r_{11} & r_{12} & r_{13} \\ r_{21} & r_{22} & r_{23} \\ r_{31} & r_{32} & r_{33} \end{bmatrix}$$

Note that the coefficients of R are constrained:
A rotation matrix is <u>orthonormal</u>:

$R\,R^T = I$ (unit matrix)

Rotation Matrix

- Rotation matrix is orthonormal:

$$\mathbf{R}^\top \mathbf{R} = \mathbf{I}$$

- So, $\mathbf{R}^{-1} = \mathbf{R}^\top$.
- In particular, let

$$\mathbf{R} = \begin{bmatrix} \mathbf{R}_1^\top \\ \mathbf{R}_2^\top \\ \mathbf{R}_3^\top \end{bmatrix}$$

Then,

$$\mathbf{R}_i^\top \mathbf{R}_j = \begin{cases} 1 & \text{if } i = j \\ 0 & \text{otherwise} \end{cases}$$

Homogeneous coordinates

Is Perspective Projection a linear transformation? $x = f\frac{X}{Z}, \quad y = f\frac{Y}{Z}$

- no—division by z is nonlinear

Trick: add one more coordinate:

$$(x, y) \Rightarrow \begin{bmatrix} x \\ y \\ 1 \end{bmatrix}$$

homogeneous image coordinates

$$(x, y, z) \Rightarrow \begin{bmatrix} x \\ y \\ z \\ 1 \end{bmatrix}$$

homogeneous scene coordinates

Converting *from* homogeneous coordinates

$$\begin{bmatrix} x \\ y \\ w \end{bmatrix} \Rightarrow (x/w, y/w)$$

$$\begin{bmatrix} x \\ y \\ z \\ w \end{bmatrix} \Rightarrow (x/w, y/w, z/w)$$

Camera parameters

Summary - A camera is described by several parameters

- Translation T of the optical center from the origin of world coords
- Rotation R of the image plane
- focal length f, principle point (Cx, Cy), pixel size (k, l)

Projection equation - correspondence between image (u,v) and scene (X, Y,Z)

$$x = \begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = \begin{bmatrix} * & * & * & * \\ * & * & * & * \\ * & * & * & * \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix} = M\mathbf{X}$$

- The projection matrix models the cumulative effect of all parameters
  - Useful to decompose into a series of operations

identity matrix

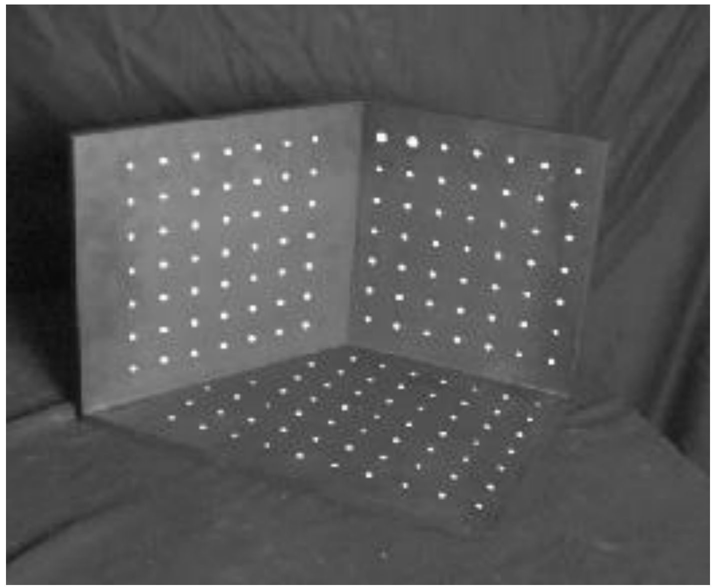
$$M = \underbrace{\begin{bmatrix} k_x & s & c_x \\ 0 & k_y & c_y \\ 0 & 0 & 1 \end{bmatrix}}_{\text{intrinsics}} \underbrace{\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix}}_{\text{projection}} \underbrace{\begin{bmatrix} \mathbf{R}_{3x3} & \mathbf{0}_{3x1} \\ \mathbf{0}_{1x3} & 1 \end{bmatrix}}_{\text{rotation}} \underbrace{\begin{bmatrix} \mathbf{I}_{3x3} & \mathbf{T}_{3x1} \\ \mathbf{0}_{1x3} & 1 \end{bmatrix}}_{\text{translation}}$$

K

$$\rho \tilde{\mathbf{x}} = \mathbf{K}(\mathbf{R}\mathbf{X}_W + \mathbf{T})$$

Estimating the projection matrix

- Place a known object in the scene
  - identify correspondence between image (ui,vi) and scene (Xi, Yi, Zi)



  - compute mapping from scene to image (previous slide)

$$\begin{bmatrix} u_i \\ v_i \\ 1 \end{bmatrix} \cong \begin{bmatrix} m_{00} & m_{01} & m_{02} & m_{03} \\ m_{10} & m_{11} & m_{12} & m_{13} \\ m_{20} & m_{21} & m_{22} & m_{23} \end{bmatrix} \begin{bmatrix} X_i \\ Y_i \\ Z_i \\ 1 \end{bmatrix}$$

Direct linear calibration

$$\begin{bmatrix} u_i \\ v_i \\ 1 \end{bmatrix} \cong \begin{bmatrix} m_{00} & m_{01} & m_{02} & m_{03} \\ m_{10} & m_{11} & m_{12} & m_{13} \\ m_{20} & m_{21} & m_{22} & m_{23} \end{bmatrix} \begin{bmatrix} X_i \\ Y_i \\ Z_i \\ 1 \end{bmatrix}$$

$$\begin{aligned} u_i &= \frac{m_{00}X_i + m_{01}Y_i + m_{02}Z_i + m_{03}}{m_{20}X_i + m_{21}Y_i + m_{22}Z_i + m_{23}} \\ v_i &= \frac{m_{10}X_i + m_{11}Y_i + m_{12}Z_i + m_{13}}{m_{20}X_i + m_{21}Y_i + m_{22}Z_i + m_{23}} \end{aligned}$$

$$u_i(m_{20}X_i + m_{21}Y_i + m_{22}Z_i + m_{23}) = m_{00}X_i + m_{01}Y_i + m_{02}Z_i + m_{03}$$
$$v_i(m_{20}X_i + m_{21}Y_i + m_{22}Z_i + m_{23}) = m_{10}X_i + m_{11}Y_i + m_{12}Z_i + m_{13}$$

$$\begin{bmatrix} X_i & Y_i & Z_i & 1 & 0 & 0 & 0 & 0 & -u_iX_i & -u_iY_i & -u_iZ_i & -u_i \\ 0 & 0 & 0 & 0 & X_i & Y_i & Z_i & 1 & -v_iX_i & -v_iY_i & -v_iZ_i & -v_i \end{bmatrix} \begin{bmatrix} m_{00} \\ m_{01} \\ m_{02} \\ m_{03} \\ m_{10} \\ m_{11} \\ m_{12} \\ m_{13} \\ m_{20} \\ m_{21} \\ m_{22} \\ m_{23} \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

Direct linear calibration

$$\begin{bmatrix} X_1 & Y_1 & Z_1 & 1 & 0 & 0 & 0 & 0 & -u_1X_1 & -u_1Y_1 & -u_1Z_1 & -u_1 \\ 0 & 0 & 0 & 0 & X_1 & Y_1 & Z_1 & 1 & -v_1X_1 & -v_1Y_1 & -v_1Z_1 & -v_1 \\ & & & & & & \vdots & & & & & \\ X_n & Y_n & Z_n & 1 & 0 & 0 & 0 & 0 & -u_nX_n & -u_nY_n & -u_nZ_n & -u_n \\ 0 & 0 & 0 & 0 & X_n & Y_n & Z_n & 1 & -v_nX_n & -v_nY_n & -v_nZ_n & -v_n \end{bmatrix} \begin{bmatrix} m_{00} \\ m_{01} \\ m_{02} \\ m_{03} \\ m_{10} \\ m_{11} \\ m_{12} \\ m_{13} \\ m_{20} \\ m_{21} \\ m_{22} \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \\ 0 \end{bmatrix}$$

Can solve for m_{ij} by linear or non-linear methods.

- Estimate *K*, *R*, and *T* from *M*

Camera Calibration

SVD

The singular value decomposition of a matrix M is the factorization of M into the product of three matrices

$$M = UDV^T$$

where the columns of U and V are orthonormal and the matrix D is diagonal with positive real entries.

Since U and V have orthogonal columns:

$$U^T U = I$$

and

$$V^T V = I$$

SVD

$$\begin{bmatrix} X_1 & Y_1 & Z_1 & 1 & 0 & 0 & 0 & 0 & -u_1X_1 & -u_1Y_1 & -u_1Z_1 & -u_1 \\ 0 & 0 & 0 & 0 & X_1 & Y_1 & Z_1 & 1 & -v_1X_1 & -v_1Y_1 & -v_1Z_1 & -v_1 \\ & & & & & & \vdots & & & & & \\ X_n & Y_n & Z_n & 1 & 0 & 0 & 0 & 0 & -u_nX_n & -u_nY_n & -u_nZ_n & -u_n \\ 0 & 0 & 0 & 0 & X_n & Y_n & Z_n & 1 & -v_nX_n & -v_nY_n & -v_nZ_n & -v_n \end{bmatrix} \begin{bmatrix} m_{00} \\ m_{01} \\ m_{02} \\ m_{03} \\ m_{10} \\ m_{11} \\ m_{12} \\ m_{13} \\ m_{20} \\ m_{21} \\ m_{22} \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \\ 0 \end{bmatrix}$$

$$M = UDV^T$$

Solution: the smallest element (null) of the diagonal of D (last row (NumPy) or column (Matlab) of V)

Intrinsic & Extrinsic parameters from M

- Projection Equation

$$\rho\,\tilde{\mathbf{x}} = \mathbf{K}(\mathbf{R}\mathbf{X}_W + \mathbf{T})$$

Using the notation $\mathbf{r}_i = (r_{i1}\; r_{i2}\; r_{i3})$.

$$\mathbf{A} = \begin{pmatrix} r_{11} & r_{12} & r_{13} & t_x \\ r_{21} & r_{22} & r_{23} & t_y \\ r_{31} & r_{32} & r_{33} & t_z \\ 0 & 0 & 0 & 1 \end{pmatrix} = \begin{pmatrix} \mathbf{R} & \mathbf{t} \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} \mathbf{r}_1 & t_x \\ \mathbf{r}_2 & t_y \\ \mathbf{r}_3 & t_z \\ \mathbf{0} & 1 \end{pmatrix}$$

$$\begin{pmatrix} su \\ sv \\ s \end{pmatrix} = \mathbf{KA} \begin{pmatrix} X \\ Y \\ Z \\ 1 \end{pmatrix}_w \quad \Rightarrow \quad \begin{pmatrix} su \\ sv \\ s \end{pmatrix} = \mathbf{M} \begin{pmatrix} X \\ Y \\ Z \\ 1 \end{pmatrix}_w$$

Intrinsic & Extrinsic parameters from M

- Projection Equation

$$\begin{pmatrix} su \\ sv \\ s \end{pmatrix} = \begin{pmatrix} m_{11} & m_{12} & m_{13} & m_{14} \\ m_{21} & m_{22} & m_{23} & m_{24} \\ m_{31} & m_{32} & m_{33} & m_{34} \end{pmatrix} \begin{pmatrix} X \\ Y \\ Z \\ 1 \end{pmatrix}_w$$

Using the notation $\mathbf{m}_i = (m_{i1}\ m_{i2}\ m_{i3})$.

$$\mathrm{M} = \begin{pmatrix} m_{11} & m_{12} & m_{13} & m_{14} \\ m_{21} & m_{22} & m_{23} & m_{24} \\ m_{31} & m_{32} & m_{33} & m_{34} \end{pmatrix} = \begin{pmatrix} \mathbf{m}_1 & m_{14} \\ \mathbf{m}_2 & m_{24} \\ \mathbf{m}_3 & m_{34} \end{pmatrix}$$

Using the notation $\mathbf{r}_i$ defined for the matrix

$$M = KA = \begin{pmatrix} f_x \boldsymbol{r}_1 + c_x \boldsymbol{r}_3 & f_x t_x + c_x t_z \\ f_y \boldsymbol{r}_2 + c_y \boldsymbol{r}_3 & f_y t_y + c_y t_z \\ \boldsymbol{r}_3 & t_z \end{pmatrix}$$

Intrinsic & Extrinsic parameters from M

- Expressing M in function of KA, and taking into account the orthogonality of the matrix R, the camera parameters can be estimated as:

$$
\begin{cases}
\boldsymbol{r}_3 = \boldsymbol{m}_3 \\
c_x = \boldsymbol{m}_1 \cdot \boldsymbol{m}_3 \\
c_y = \boldsymbol{m}_2 \cdot \boldsymbol{m}_3 \\
f_x = ||\boldsymbol{m}_1 \wedge \boldsymbol{m}_3|| \\
f_y = ||\boldsymbol{m}_2 \wedge \boldsymbol{m}_3|| \\
\boldsymbol{r}_1 = 1/f_x(\boldsymbol{m}_1 - c_x\boldsymbol{m}_3) \\
\boldsymbol{r}_2 = 1/f_y(\boldsymbol{m}_2 - c_y\boldsymbol{m}_3) \\
t_x = 1/f_x(m_{14} - c_x m_{34}) \\
t_y = 1/f_y(m_{24} - c_y m_{34}) \\
t_z = m_{34}
\end{cases}
$$

With $\boldsymbol{m}_i=(\boldsymbol{m}_{i1}, \boldsymbol{m}_{i2}, \boldsymbol{m}_{i3})$

Intrinsic & Extrinsic parameters from M

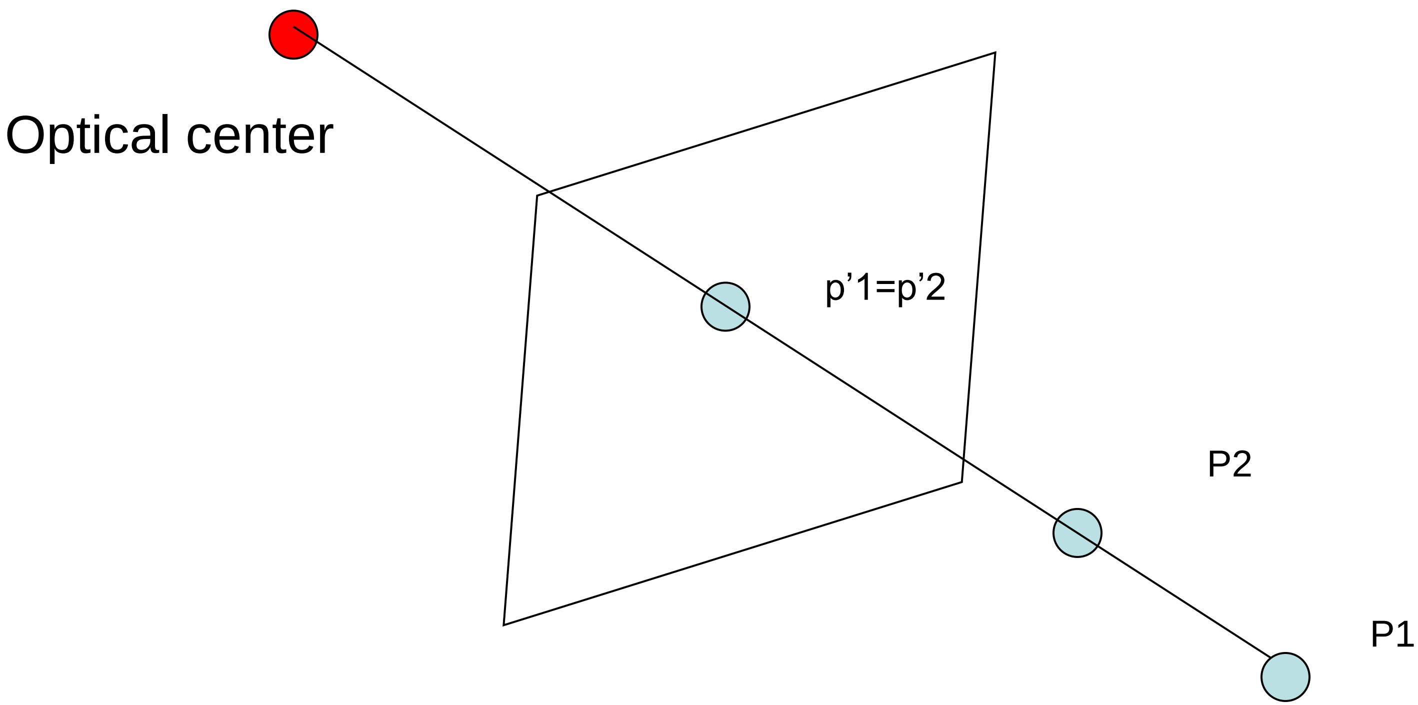
- Expressing M in function of KA, and taking into account the orthogonality of the matrix R, the camera parameters can be estimated as:

$$
\begin{cases}
\boldsymbol{r}_3 = \boldsymbol{m}_3 \\
c_x = \boldsymbol{m}_1 \cdot \boldsymbol{m}_3 \\
c_y = \boldsymbol{m}_2 \cdot \boldsymbol{m}_3 \\
f_x = ||\boldsymbol{m}_1 \wedge \boldsymbol{m}_3|| \\
f_y = ||\boldsymbol{m}_2 \wedge \boldsymbol{m}_3|| \\
\boldsymbol{r}_1 = 1/f_x(\boldsymbol{m}_1 - c_x\boldsymbol{m}_3) \\
\boldsymbol{r}_2 = 1/f_y(\boldsymbol{m}_2 - c_y\boldsymbol{m}_3) \\
t_x = 1/f_x(m_{14} - c_x m_{34}) \\
t_y = 1/f_y(m_{24} - c_y m_{34}) \\
t_z = m_{34}
\end{cases}
$$

With $\boldsymbol{m}_i$=$(\boldsymbol{m}_{i1}, \boldsymbol{m}_{i2}, \boldsymbol{m}_{i3})$

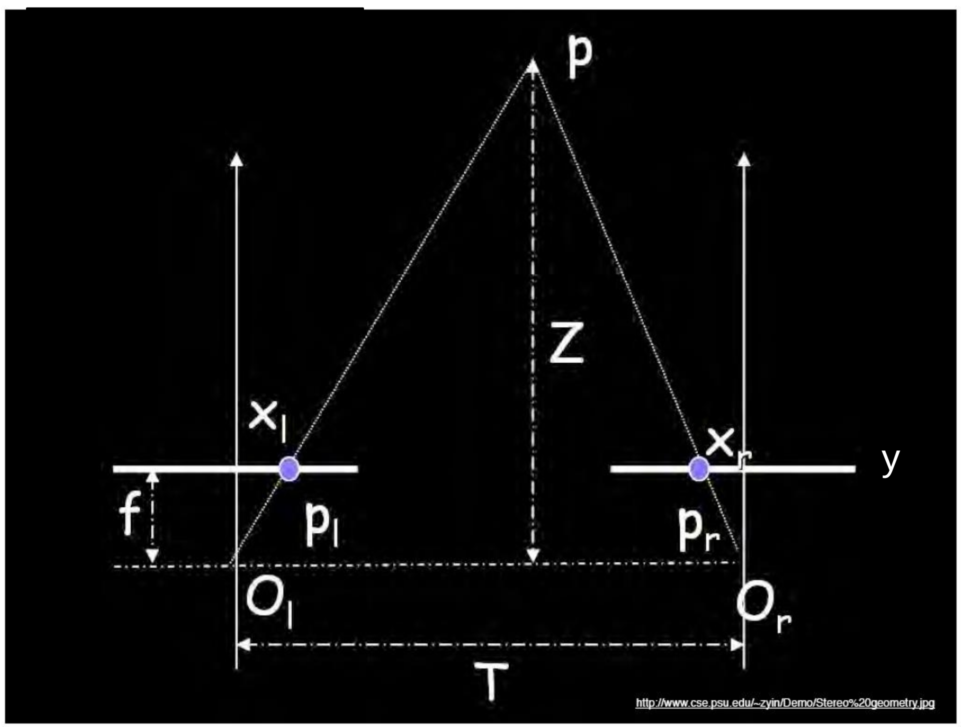
Why multiple views?

- Structure and depth are inherently ambiguous from single views.
- => Many-to-one: any points along same ray map to same point in image



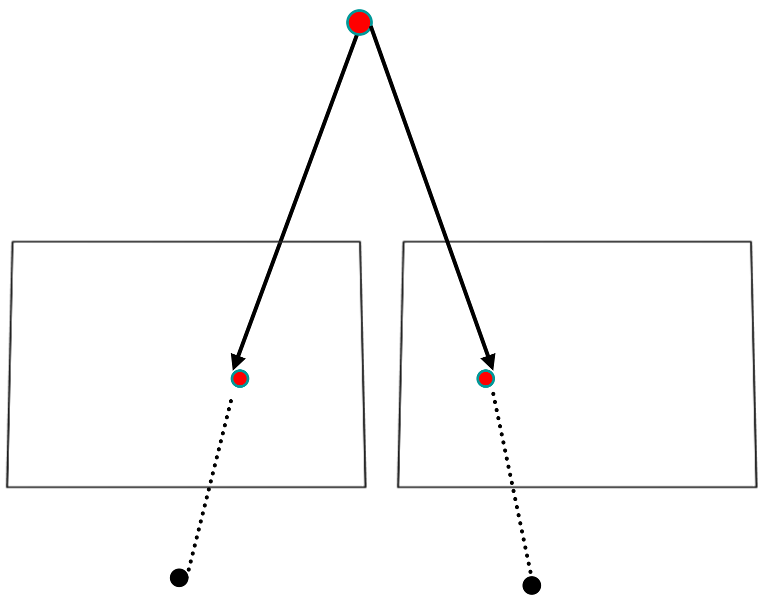
Geometry for a simple stereo system

- First, assuming parallel optical axes, known camera parameters (i.e., calibrated cameras):

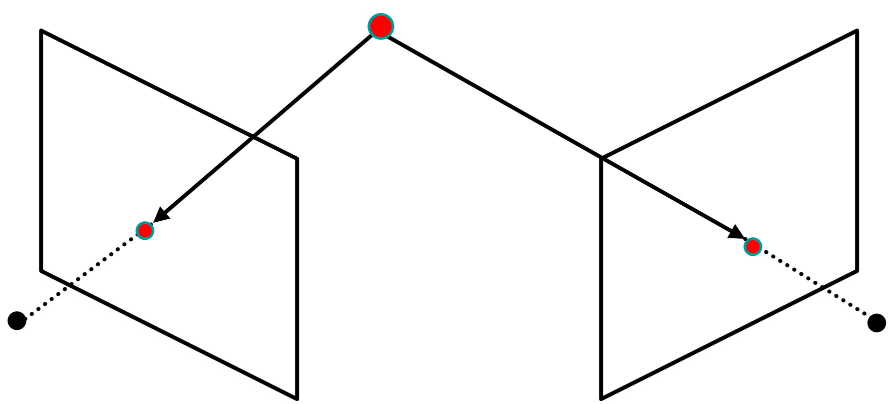


General case, with calibrated cameras

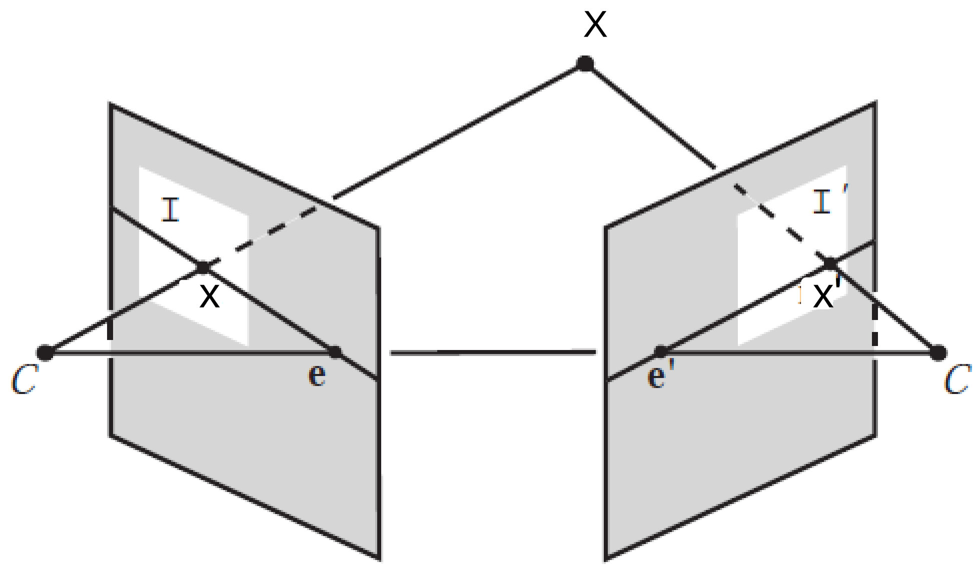
- The two cameras will not have parallel optical axes.



VERSUS



Triangulation



Having two points x (on left image) and x’ (right image) projections of a scene point X -> How to estimate X?

From perspective projection we have:

$$x = \mathrm{M}\mathbf{X} \qquad x' = \mathrm{M}'\mathbf{X}$$

General Triangulation

- Use the equations $x = \mathrm{M}\mathbf{X}$ and $x' = \mathrm{M}'\mathbf{X}$ to solve $\mathbf{X}$
  - For the first camera:

$$\mathrm{M} = \begin{bmatrix} m_{11} & m_{12} & m_{13} & m_{14} \\ m_{21} & m_{22} & m_{23} & m_{24} \\ m_{31} & m_{32} & m_{33} & m_{34} \end{bmatrix} = \begin{bmatrix} \boldsymbol{m}^{1\top} \\ \boldsymbol{m}^{2\top} \\ \boldsymbol{m}^{3\top} \end{bmatrix}$$

where $\boldsymbol{m}^{i\top}$ are the rows of M

  - Eliminate the unknown scale $sx = \mathrm{MX}$ by forming a cross product $x \times \mathrm{MX} = 0$

$$x(\boldsymbol{m}^{3\top}\mathbf{X}) - (\boldsymbol{m}^{1\top}\mathbf{X}) = 0$$
$$y(\boldsymbol{m}^{3\top}\mathbf{X}) - (\boldsymbol{m}^{2\top}\mathbf{X}) = 0$$
$$x(\boldsymbol{m}^{2\top}\mathbf{X}) - y(\boldsymbol{m}^{1\top}\mathbf{X}) = 0$$

  - Rearrange as (first to equations only)

$$\begin{bmatrix} x\boldsymbol{m}^{3\top} - \boldsymbol{m}^{1\top} \\ y\boldsymbol{m}^{3\top} - \boldsymbol{m}^{2\top} \end{bmatrix} \mathbf{X} = 0$$

Advantage: extends to more than two views

General Triangulation

- Use the equations $x = \mathrm{M}\mathbf{X}$ and $x' = \mathrm{M}'\mathbf{X}$ to solve $\mathbf{X}$
  - Similarly for the second camera:

$$\begin{bmatrix} x'\boldsymbol{m}'^{3\top} - \boldsymbol{m}'^{1\top} \\ y'\boldsymbol{m}'^{3\top} - \boldsymbol{m}'^{2\top} \end{bmatrix} \mathbf{X} = 0$$

  - Collecting together gives

$$\mathrm{A}\mathbf{X} = 0$$

where A is a 4×4 matrix

$$\mathrm{A} = \begin{bmatrix} x\boldsymbol{m}^{3\top} - \boldsymbol{m}^{1\top} \\ y\boldsymbol{m}^{3\top} - \boldsymbol{m}^{2\top} \\ x'\boldsymbol{m}'^{3\top} - \boldsymbol{m}'^{1\top} \\ y'\boldsymbol{m}'^{3\top} - \boldsymbol{m}'^{2\top} \end{bmatrix}$$

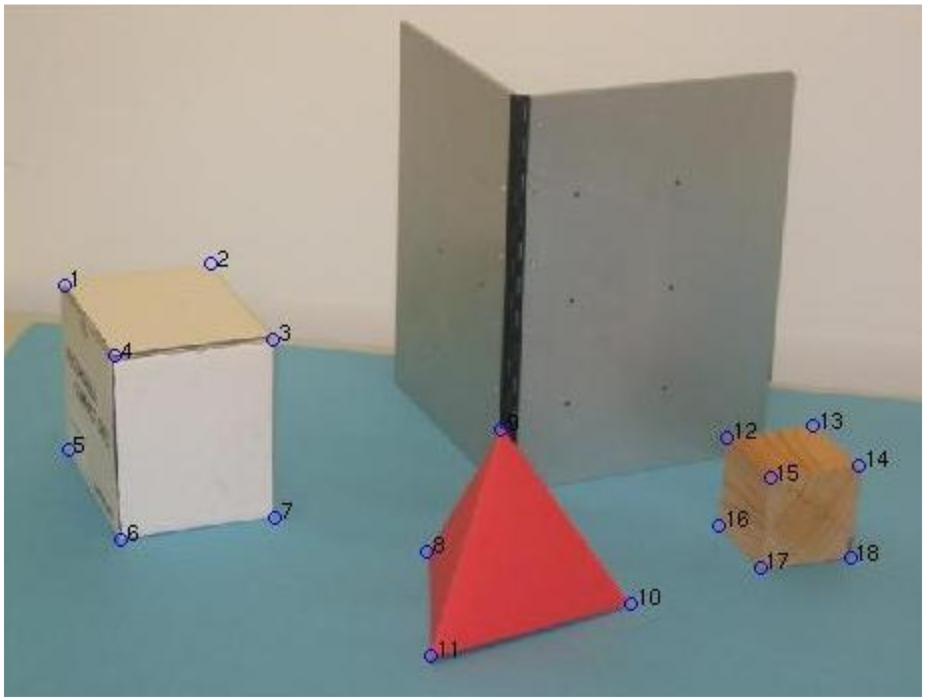
  - From which we can solve using SVD

$$\mathrm{A} = UDV^T$$

Solution: the smallest element (null) of the diagonal of D (last row (NumPy) or column (Matlab) of V)

General Triangulation

- **How do we check our reconstruction?**


1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18


11
12
5
6
9
10
3
4
7
8
Z
1
2
Y
X
O