中国科学技术大学
University of Science and Technology of China

# Struction from Motion

张举勇
中国科学技术大学

# Uncontrolled (Casual) Video

# Overview

Compute 3D scene structure and camera motion from a sequence of frames.

**Topics:**

(1) Structure from Motion Problem

(2) SFM Observation Matrix

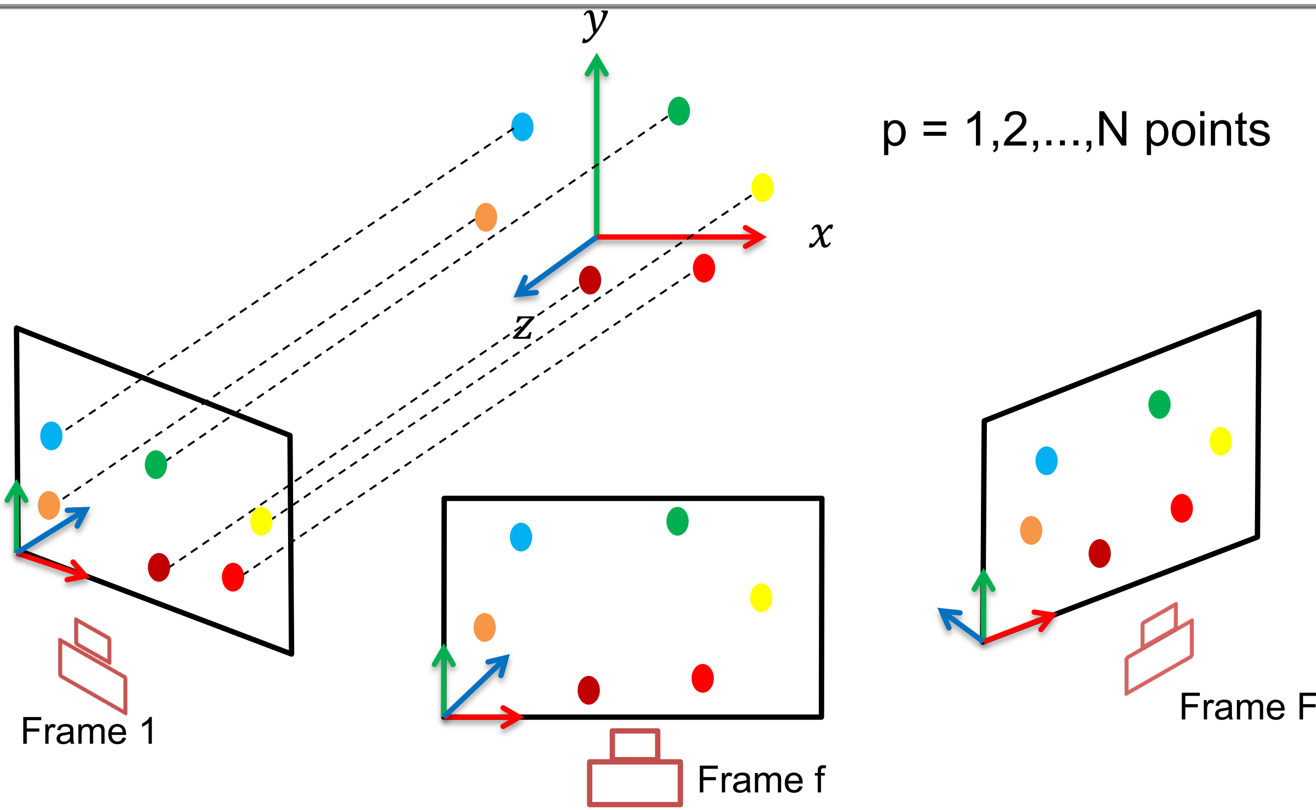
(3) Rank of Observation Matrix

(4) Tomasi-Kanade Factorization

# Feature Detection and Tracking

- Detect feature points: Corners , SIFT points , ...
- Track feature points: Template Matching, Optical Flow...

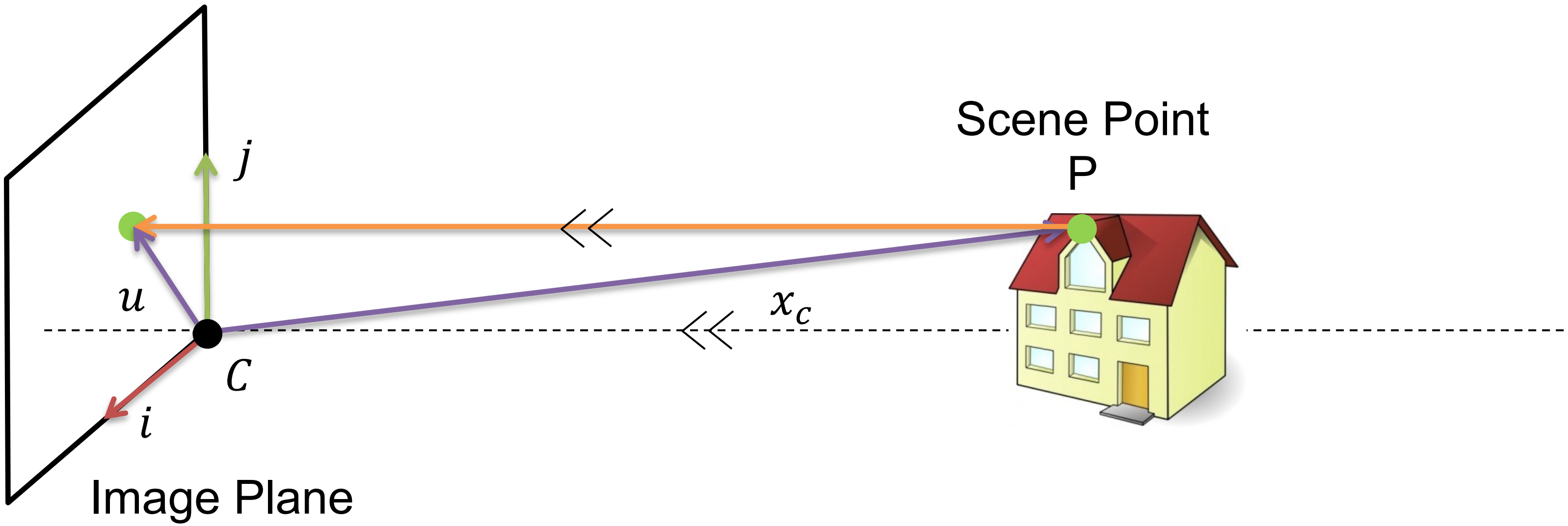
# Orthographic Structure from Motion




Given sets of corresponding image points (2D): $(u_{f,p}, v_{f,p})$

Find scene points (3D) $P_p$, assuming orthographic camera.

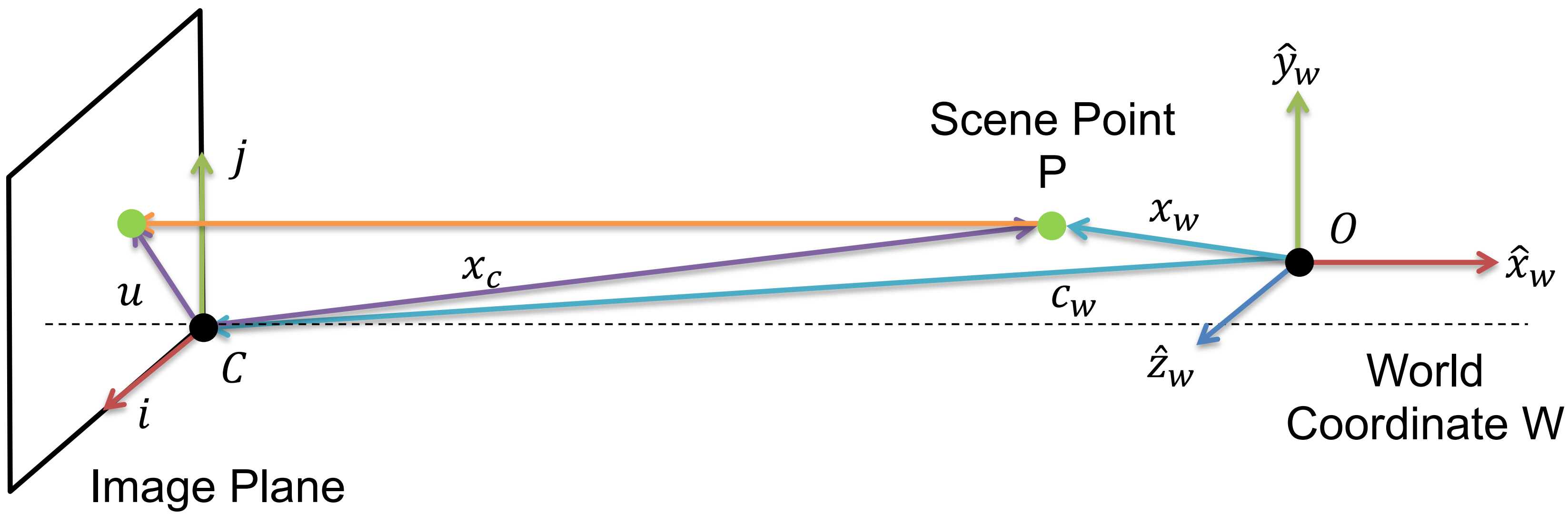
# From 3D to 2D: Orthographic Projection




$$u = i \cdot x_c = i^T x_c$$
$$v = j \cdot x_c = j^T x_c$$

Perspective cameras exhibit orthographic projection when distance of scene from camera is large compared to depth variation within scene (magnification is nearly constant).

# From 3D to 2D: Orthographic Projection

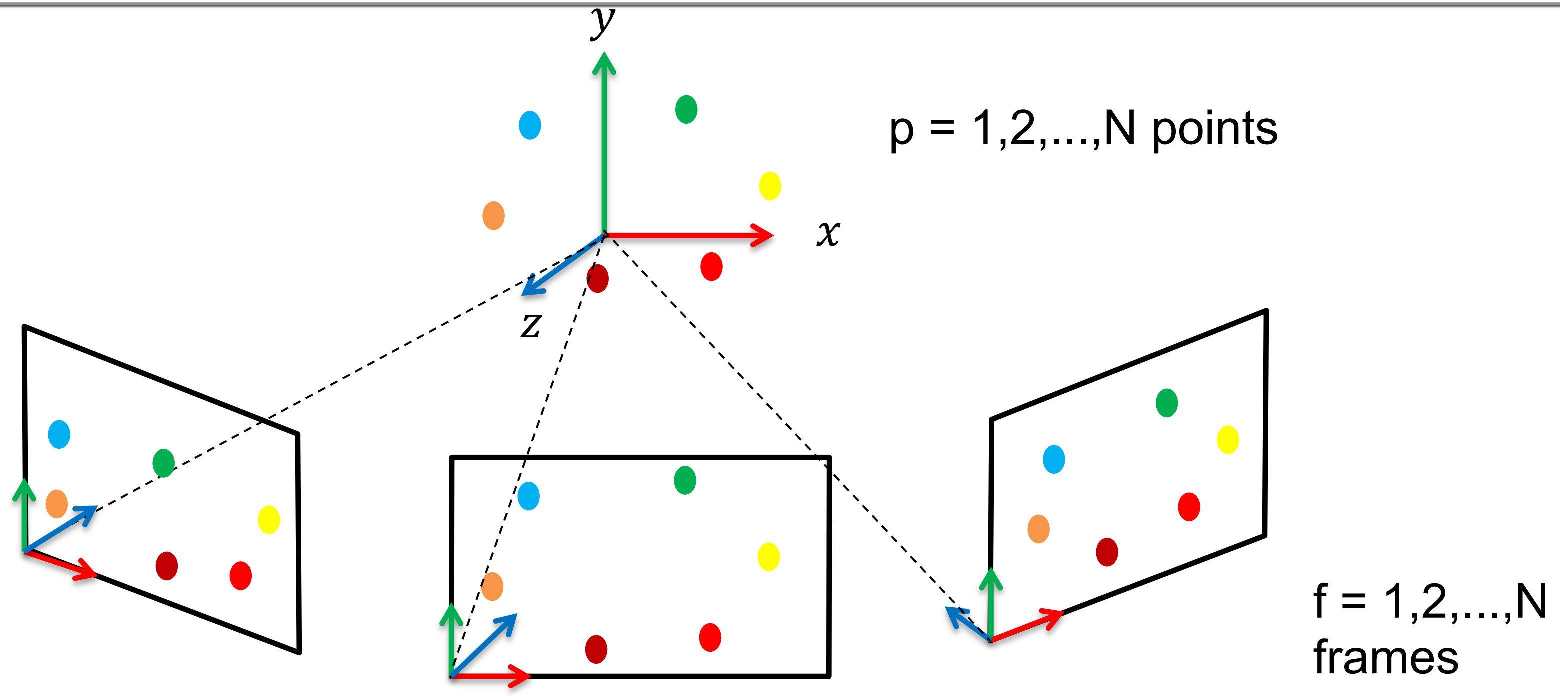



$$u = i^T x_c = i^T(x_w - c_w) = i^T(P - C)$$

$$v = j^T x_c = j^T(x_w - c_w) = j^T(P - C)$$

$$\boxed{\begin{aligned} u &= i^T(P - C) \\ v &= j^T(P - C) \end{aligned}}$$

# Orthographic SFM




Given corresponding image points (2D) $(u_{f,p}, v_{f,p})$

Find scene points $\{P_p\}$.

Camera Positions $\{C_f\}$, camera orientations $\{(i_f, j_f)\}$ are unknown.

# Orthographic SFM

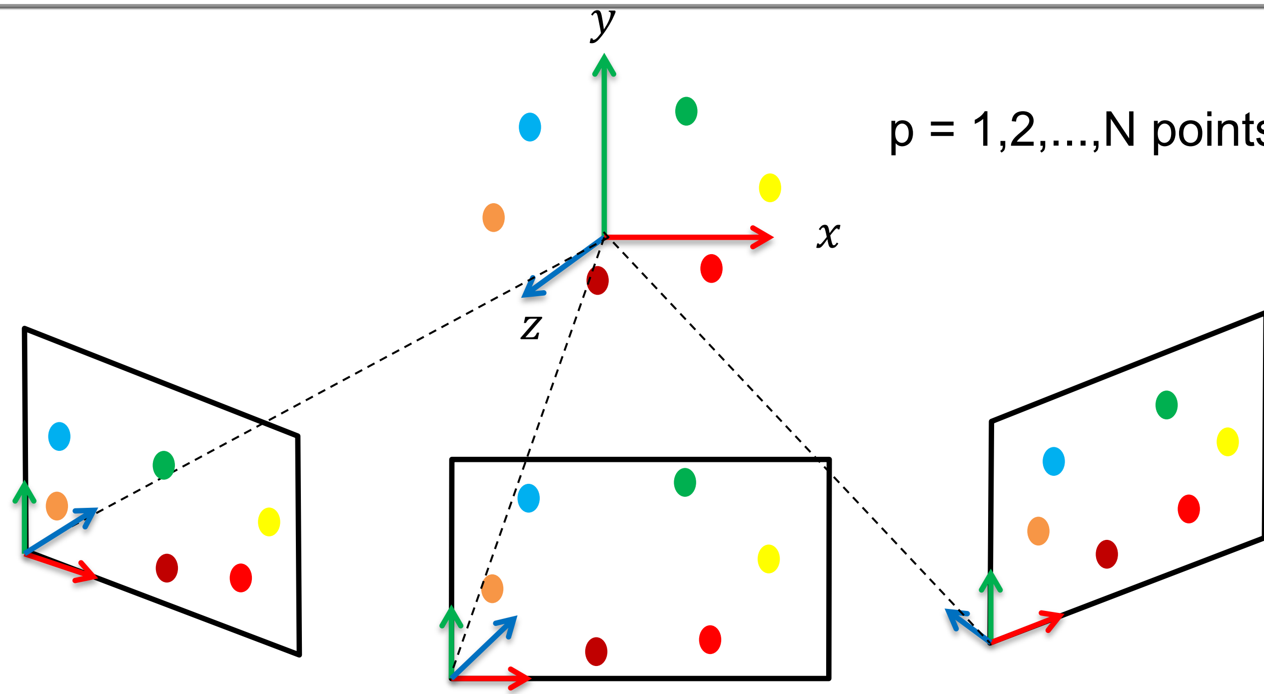


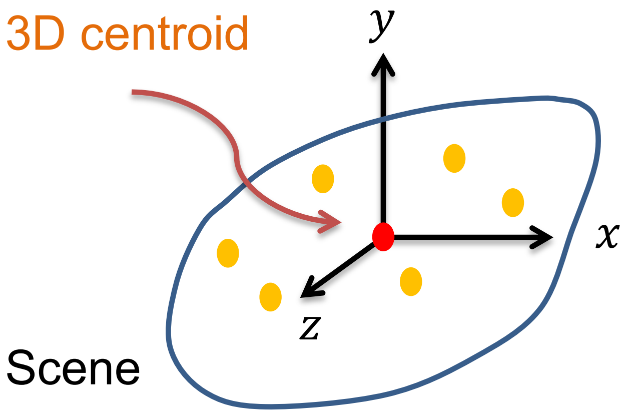

Image of point P in camera frame f :

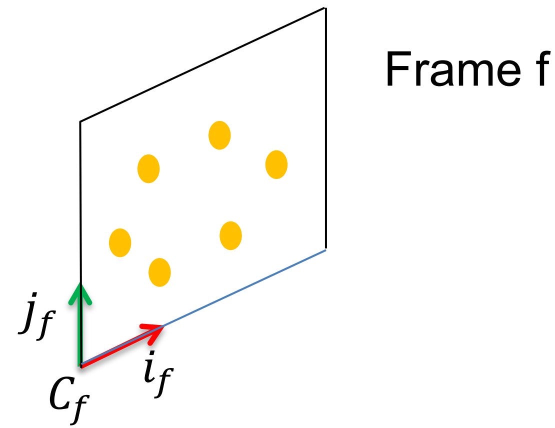
$$\underbrace{u_{f,p}}_{\text{Known}} = \underbrace{i_f^T\,(P_p - C_f)}_{\text{Unknown}}$$

$$\underbrace{v_{f,p}}_{\text{Known}} = \underbrace{j_f^T\,(P_p - C_f)}_{\text{Unknown}}$$

We can remove C from equations to simply SFM problem.

# Centering Trick



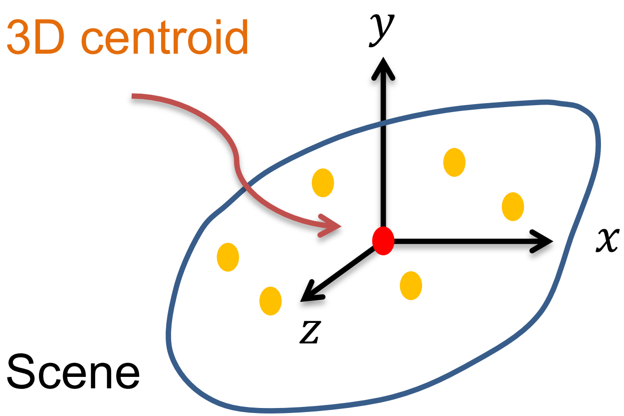



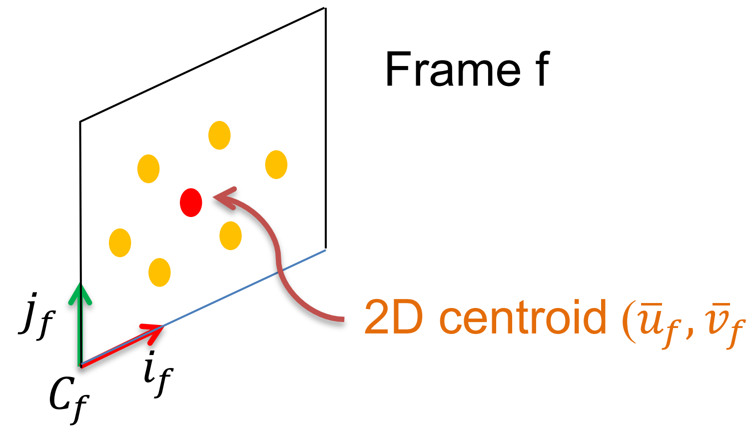

Assume origin of world at centroid of scene points:

$$\frac{1}{N}\sum_{p=1}^{N} P_p = \bar{P} = 0$$

We will compute scene points w.r.t their centroid!

# Centering Trick



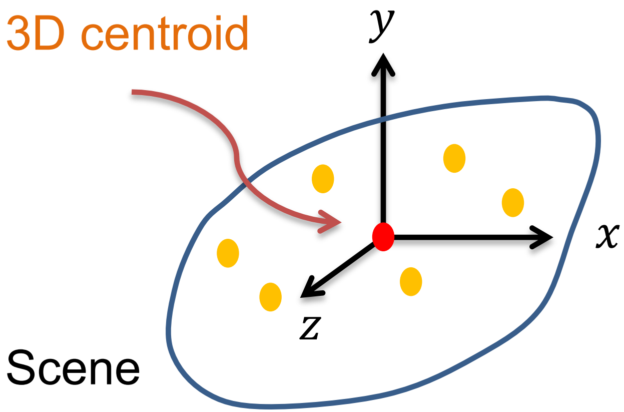

Centroid $(\bar{u}_f, \bar{v}_f)$ of the image points in frame $f$ :

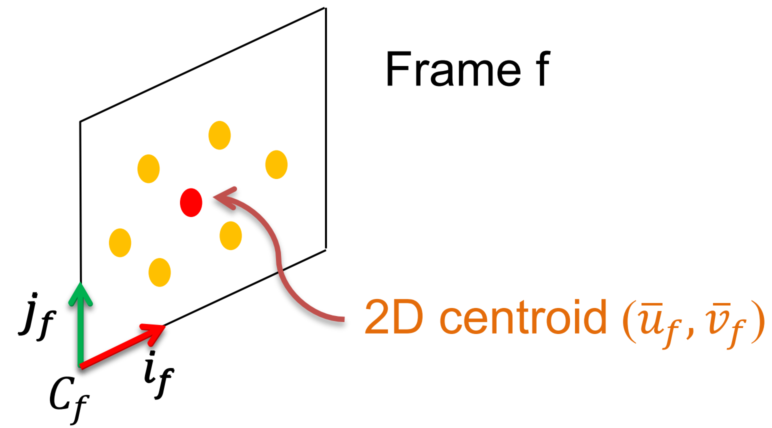
$$\bar{u}_f = \frac{1}{N}\sum_{p=1}^{N} u_{f,p} = \frac{1}{N}\sum_{p=1}^{N} i_f^T\,(P_p - C_f) \qquad \bar{v}_f = \frac{1}{N}\sum_{p=1}^{N} v_{f,p} = \frac{1}{N}\sum_{p=1}^{N} j_f^T\,(P_p - C_f)$$

$$\bar{u}_f = \cancel{\frac{1}{N} i_f^T \sum_{p=1}^{N} P_p} - \frac{1}{N}\sum_{p=1}^{N} i_f^T\,C_f \qquad \bar{v}_f = \cancel{\frac{1}{N} j_f^T \sum_{p=1}^{N} P_p} - \frac{1}{N}\sum_{p=1}^{N} j_f^T\,C_f$$

$$\boxed{\bar{u}_f = -i_f^T\,C_f} \qquad \boxed{\bar{v}_f = -j_f^T\,C_f}$$

# Centering Trick




Shift camera origin to the centroid $(\bar{u}_f, \bar{v}_f)$ .

Image points w.r.t. $(\bar{u}_f, \bar{v}_f)$ :

$$\tilde{u}_{f,p} = u_{f,p} - \bar{u}_f = i_f^T \left(P_p - C_f\right) + i_f^T C_f \qquad \tilde{v}_{f,p} = v_{f,p} - \bar{v}_f = j_f^T \left(P_p - C_f\right) + j_f^T C_f$$

$$\boxed{\tilde{u}_{f,p} = i_f^T P_p} \qquad \boxed{\tilde{v}_{f,p} = j_f^T P_p}$$

Camera locations $C_f$ now removed from equations.

# Observation Matrix W

$$\tilde{u}_{f,p} = i_f^T \, P_p$$

$$\tilde{v}_{f,p} = j_f^T \, P_p$$

$$\begin{bmatrix} \tilde{u}_{f,p} \\ \tilde{u}_{f,p} \end{bmatrix} = \begin{bmatrix} i_f^T \\ j_f^T \end{bmatrix} P_p$$

$$
\begin{array}{l}
\\
\text{Image 1}\\
\text{Image 2}\\
\\
\text{Image F}\\
\text{Image 1}\\
\text{Image 2}\\
\\
\text{Image F}
\end{array}
\begin{array}{c}
\begin{array}{cccc} \text{Point 1} & \text{Point 2} & & \text{Point N} \end{array}\\
\left[\begin{array}{cccc}
\tilde{u}_{1,1} & \tilde{u}_{1,2} & \cdots & \tilde{u}_{1,N}\\
\tilde{u}_{2,1} & \tilde{u}_{2,2} & \cdots & \tilde{u}_{2,N}\\
\vdots & \vdots & \vdots & \vdots\\
\tilde{u}_{F,1} & \tilde{u}_{F,2} & \cdots & \tilde{u}_{F,N}\\
\hline
\tilde{v}_{1,1} & \tilde{v}_{1,2} & \cdots & \tilde{v}_{1,N}\\
\tilde{v}_{2,1} & \tilde{v}_{2,2} & \cdots & \tilde{v}_{2,N}\\
\vdots & \vdots & \vdots & \vdots\\
\tilde{v}_{F,1} & \tilde{v}_{F,2} & \cdots & \tilde{v}_{F,N}
\end{array}\right]
\end{array}
=
\left[\begin{array}{c}
i_1^T\\
i_2^T\\
\vdots\\
i_N^T\\
\hline
j_1^T\\
j_2^T\\
\vdots\\
j_N^T
\end{array}\right]
\begin{array}{c}
\begin{array}{cccc} \text{Point 1} & \text{Point 2} & & \text{Point N} \end{array}\\
\begin{bmatrix} P_1 & P_2 & \cdots & P_N \end{bmatrix}
\end{array}
$$

$W_{2F\times N}$

Centroid-Subtracted Feature Points (Known)

$M_{2F\times 3}$

Camera Motion (Unknown)

$S_{3\times N}$

Scene Struction (Unknown)

# Observation Matrix W

$$
\begin{array}{cc}
 & \begin{array}{cccc}\text{Point 1} & \text{Point 2} & & \text{Point N}\end{array} \\
\begin{array}{c}\text{Image 1}\\ \text{Image 2}\\ \\ \text{Image F}\\ \text{Image 1}\\ \text{Image 2}\\ \\ \text{Image F}\end{array} &
\left[\begin{array}{cccc}
\tilde{u}_{1,1} & \tilde{u}_{1,2} & \cdots & \tilde{u}_{1,N} \\
\tilde{u}_{2,1} & \tilde{u}_{2,2} & \cdots & \tilde{u}_{2,N} \\
\vdots & \vdots & \vdots & \vdots \\
\tilde{u}_{F,1} & \tilde{u}_{F,2} & \cdots & \tilde{u}_{F,N} \\
\hdashline
\tilde{v}_{1,1} & \tilde{v}_{1,2} & \cdots & \tilde{v}_{1,N} \\
\tilde{v}_{2,1} & \tilde{v}_{2,2} & \cdots & \tilde{v}_{2,N} \\
\vdots & \vdots & \vdots & \vdots \\
\tilde{v}_{F,1} & \tilde{v}_{F,2} & \cdots & \tilde{v}_{F,N}
\end{array}\right]
\end{array}
=
\left[\begin{array}{c}
i_1^T \\ i_2^T \\ \vdots \\ i_N^T \\ \hdashline j_1^T \\ j_2^T \\ \vdots \\ j_N^T
\end{array}\right]
\begin{array}{c}
\begin{array}{cccc}\text{Point 1} & \text{Point 2} & & \text{Point N}\end{array} \\
\begin{bmatrix} P_1 & P_2 & \cdots & P_N \end{bmatrix}
\end{array}
$$

$W_{2F\times N}$

Centroid-Subtracted
Feature Points (Known)

$M_{2F\times 3}$

Camera Motion
(Unknown)

$S_{3\times N}$

Scene Struction
(Unknown)

Can we find M and s from W ?

# Linear Independence of Vectors

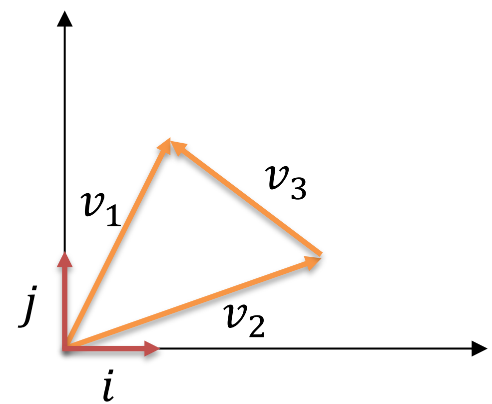
A set of vectors $\{v_1, v_2, \dots v_n\}$ is said to be linearly independent if no vector can be represented as a weighted linear sum of the others.

$\{i, j\}$ is linearly independent.

$\{i, j, v_1\}$ is linearly dependent.

$\{i, j, v_3\}$ is linearly dependent.

$\{v_1, v_2, v_3\}$ is linearly dependent.



# Rank of a Matrix

- Column Rank: The number of linearly independent columns of the matrix.
- Row Rank: The number of linearly independent rows of the matrix.

$$\begin{array}{cc} m & \begin{bmatrix} & & \\ & A & \\ & & \end{bmatrix} \\ & n \end{array} = \begin{bmatrix} c_1 & c_2 & \cdots & c_n \end{bmatrix} = \begin{bmatrix} r_1^T \\ r_2^T \\ \vdots \\ r_n^T \end{bmatrix}$$

$ColumnRank(\mathrm{A}) \le \mathrm{n}$ $\qquad$ $ColumnRank(\mathrm{A}) \le \mathrm{m}$

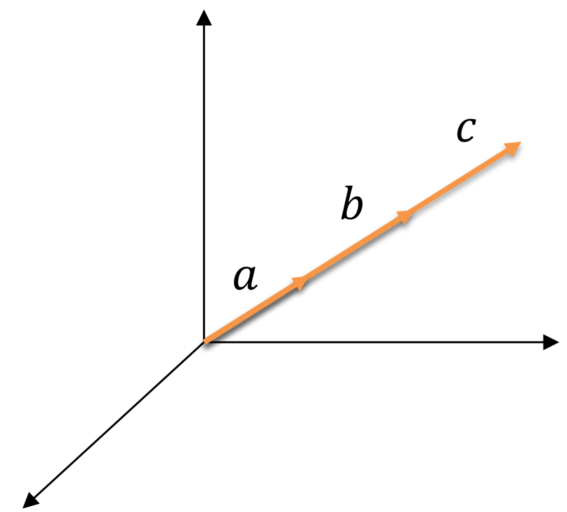
$$\boxed{\begin{gathered} ColumnRank(\mathrm{A}) = RowRank(\mathrm{A}) = Rank(\mathrm{A}) \\ Rank(\mathrm{A}) \le \min(\mathrm{m,n}) \end{gathered}}$$

# Geometric Meaning of Matrix Rank

Rank is the dimensionality of the space spanned by column or row vectors of the matrix.

$$A = \begin{bmatrix} a_1 & b_1 & c_1 \\ a_2 & b_2 & c_2 \\ a_3 & b_3 & c_3 \end{bmatrix} = [\boldsymbol{a} \quad \boldsymbol{b} \quad \boldsymbol{c}]$$

$Rank(\mathrm{A}) = 1$

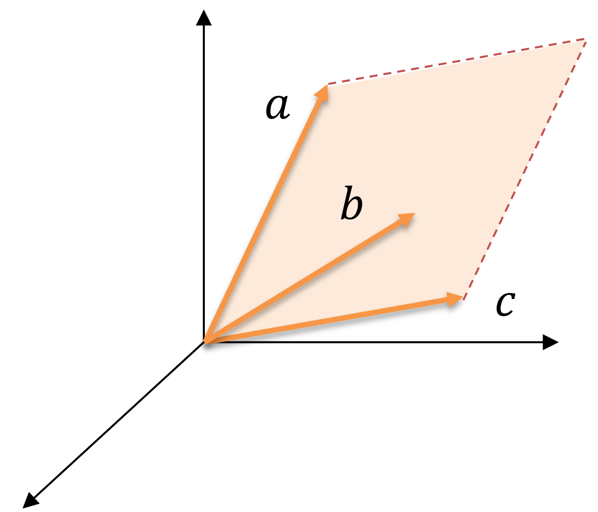


# Geometric Meaning of Matrix Rank

Rank is the dimensionality of the space spanned by column or row vectors of the matrix.

$$A = \begin{bmatrix} a_1 & b_1 & c_1 \\ a_2 & b_2 & c_2 \\ a_3 & b_3 & c_3 \end{bmatrix} = [\boldsymbol{a} \quad \boldsymbol{b} \quad \boldsymbol{c}]$$

$Rank(\mathrm{A}) = 2$ $\Leftrightarrow$

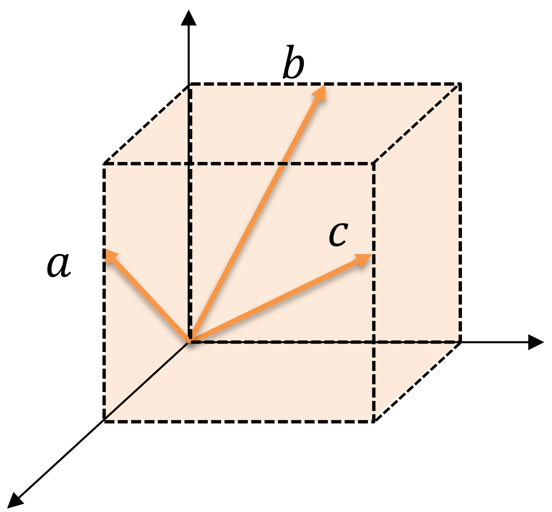


# Geometric Meaning of Matrix Rank

Rank is the dimensionality of the space spanned by column or row vectors of the matrix.

$$A = \begin{bmatrix} a_1 & b_1 & c_1 \\ a_2 & b_2 & c_2 \\ a_3 & b_3 & c_3 \end{bmatrix} = [\boldsymbol{a} \quad \boldsymbol{b} \quad \boldsymbol{c}]$$

$Rank$(A) =3 ⟺



# Important Properties of Matrix Rank

- $Rank\,(A^T) = Rank(A)$

- $Rank(A_{m\times n}\, B_{n\times p}) = \min\,(Rank(A_{m\times n}), Rank(B_{n\times p}))$
  $\le \min(\mathrm{m, n, p})$

- $Rank(A\,A^T) = Rank(A^T A) = Rank(A) = Rank(A^T)$

- $A_{m\times m}$ is invertible iff $Rank(A_{m\times m}) = \mathrm{m}$

# ...Back to Observation Matrix W

$$
\begin{array}{rc}
 & \begin{array}{cccc}\text{Point 1} & \text{Point 2} & & \text{Point N}\end{array} \\
\begin{array}{r}\text{Image 1}\\ \text{Image 2}\\ \\ \text{Image F}\\ \text{Image 1}\\ \text{Image 2}\\ \\ \text{Image F}\end{array} &
\left[\begin{array}{cccc}
\tilde{u}_{1,1} & \tilde{u}_{1,2} & \cdots & \tilde{u}_{1,N} \\
\tilde{u}_{2,1} & \tilde{u}_{2,2} & \cdots & \tilde{u}_{2,N} \\
\vdots & \vdots & \vdots & \vdots \\
\tilde{u}_{F,1} & \tilde{u}_{F,2} & \cdots & \tilde{u}_{F,N} \\
\hdashline
\tilde{v}_{1,1} & \tilde{v}_{1,2} & \cdots & \tilde{v}_{1,N} \\
\tilde{v}_{2,1} & \tilde{v}_{2,2} & \cdots & \tilde{v}_{2,N} \\
\vdots & \vdots & \vdots & \vdots \\
\tilde{v}_{F,1} & \tilde{v}_{F,2} & \cdots & \tilde{v}_{F,N}
\end{array}\right]
\end{array}
=
\left[\begin{array}{c}
i_1^T \\ i_2^T \\ \vdots \\ i_N^T \\ \hdashline j_1^T \\ j_2^T \\ \vdots \\ j_N^T
\end{array}\right]
\begin{array}{c}
\begin{array}{cccc}\text{Point 1} & \text{Point 2} & & \text{Point N}\end{array} \\
\begin{bmatrix} P_1 & P_2 & \cdots & P_N \end{bmatrix}
\end{array}
$$

$W_{2F\times N}$

Centroid-Subtracted Feature Points (Known)

$M_{2F\times 3}$

Camera Motion (Unknown)

$S_{3\times N}$

Scene Struction (Unknown)

# Rank of Observation Matrix

$$\underset{2F\times N}{W} = \underset{2F\times 3}{M} \times \underset{3\times N}{S}$$

We know:

$$Rank(MS) \leq Rank(M) \qquad Rank(MS) \leq Rank(S)$$

$$\Rightarrow \quad Rank(MS) \leq min(3,2F) \qquad Rank(MS) \leq min(3,N)$$

$$\Rightarrow \quad Rank(W) = Rank(MS) \leq min(3,N,2F)$$

Rank throem : $\boxed{Rank(W) \leq 3}$

We can "factorize" W into M and S!

# Singular Value Decomposition (SVD)

For any matrix A there exists a factorization:

$$A_{M\times N} = U_{M\times M} \cdot \Sigma_{M\times N} \cdot V_{N\times N}^{T}$$

Where $U_{M\times M}$ and $V_{N\times N}^{T}$ are orthonormal and $\Sigma_{M\times N}$ is diagonal.

Mathlab : [U,S,V] = svd(A)

$$\Sigma_{M\times N} = \begin{bmatrix} \sigma_1 & 0 & 0 & 0 & \cdots & 0 \\ 0 & \sigma_2 & 0 & 0 & \cdots & 0 \\ 0 & 0 & \sigma_3 & 0 & \cdots & 0 \\ 0 & 0 & 0 & \sigma_4 & \cdots & 0 \\ 0 & 0 & 0 & 0 & \ddots & 0 \\ 0 & 0 & 0 & 0 & \cdots & \sigma_N \\ 0 & 0 & 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \cdots & \vdots \end{bmatrix}$$

$\sigma_1, \ldots, \sigma_N$ : Singular Values

If *Rank*(A) = r then A has r non-zero singular values.

# Enforcing Rank Constraint

Using SVD:

$$W = U\Sigma V^T$$

$$= \underbrace{\begin{bmatrix} & & \\ & U & \\ & & \end{bmatrix}}_{2F\times 2F} \underbrace{\begin{bmatrix} \sigma_1 & 0 & 0 & 0 & \cdots & 0 \\ 0 & \sigma_2 & 0 & 0 & \cdots & 0 \\ 0 & 0 & \sigma_3 & 0 & \cdots & 0 \\ 0 & 0 & 0 & \sigma_4 & \cdots & 0 \\ 0 & 0 & 0 & 0 & \ddots & 0 \\ 0 & 0 & 0 & 0 & \cdots & \sigma_N \\ 0 & 0 & 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \cdots & \vdots \end{bmatrix}}_{2F\times N} \underbrace{\begin{bmatrix} & & \\ & V^T & \\ & & \end{bmatrix}}_{N\times N}$$

Where: $\sigma_1 \geq \sigma_2 \geq \ldots \geq \sigma_N$ are the singular values of $\Sigma$.

# Enforcing Rank Constraint

Using SVD:

$$W = U\Sigma V^T$$

$$= \underbrace{\begin{bmatrix} & & \\ & U & \\ & & \end{bmatrix}}_{2F\times 2F} \underbrace{\begin{bmatrix} \sigma_1 & 0 & 0 & 0 & \cdots & 0 \\ 0 & \sigma_2 & 0 & 0 & \cdots & 0 \\ 0 & 0 & \sigma_3 & 0 & \cdots & 0 \\ 0 & 0 & 0 & 0 & \cdots & 0 \\ 0 & 0 & 0 & 0 & \ddots & 0 \\ 0 & 0 & 0 & 0 & \cdots & 0 \\ 0 & 0 & 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \cdots & \vdots \end{bmatrix}}_{2F\times N} \underbrace{\begin{bmatrix} & & \\ & V^T & \\ & & \end{bmatrix}}_{N\times N}$$

Since $Rank(W) \leq 3$, $Rank(\Sigma) \leq 3$.

All expect first 3 diagonal elements of $\Sigma$ must be 0.

# Enforcing Rank Constraint

Using SVD:

$$W = U\Sigma V^T$$

$$= \underbrace{\begin{bmatrix} \underbrace{U_1}_{3} & \underbrace{U_2}_{2F-3} \end{bmatrix}}_{2F\times 2F} \underbrace{\begin{bmatrix} \sigma_1 & 0 & 0 & 0 & \cdots & 0 \\ 0 & \sigma_2 & 0 & 0 & \cdots & 0 \\ 0 & 0 & \sigma_3 & 0 & \cdots & 0 \\ 0 & 0 & 0 & 0 & \cdots & 0 \\ 0 & 0 & 0 & 0 & \ddots & 0 \\ 0 & 0 & 0 & 0 & \cdots & 0 \\ 0 & 0 & 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \cdots & \vdots \end{bmatrix}}_{2F\times N} \underbrace{\begin{bmatrix} V_1^T \\ V_2^T \end{bmatrix}}_{N\times N} \begin{matrix} 3 \\ N-3 \end{matrix}$$

Since $Rank(W) \leq 3$, $Rank(\Sigma) \leq 3$ .

Submatrices $U_2$ and $V_2^T$ do not contribute to W.

# Enforcing Rank Constraint

Using SVD:

$$W = U \Sigma V^T$$

$$
= \underbrace{\begin{bmatrix} \underset{3}{U_1} & \underset{2F-3}{U_2} \end{bmatrix}}_{2F\times 2F}
\underbrace{\begin{bmatrix}
\sigma_1 & 0 & 0 & 0 & \cdots & 0 \\
0 & \sigma_2 & 0 & 0 & \cdots & 0 \\
0 & 0 & \sigma_3 & 0 & \cdots & 0 \\
0 & 0 & 0 & 0 & \cdots & 0 \\
0 & 0 & 0 & 0 & \ddots & 0 \\
0 & 0 & 0 & 0 & \cdots & 0 \\
0 & 0 & 0 & 0 & \cdots & 0 \\
\vdots & \vdots & \vdots & \vdots & \cdots & \vdots
\end{bmatrix}}_{2F\times N}
\underbrace{\begin{array}{c} \begin{bmatrix} V_1^T \\ V_2^T \end{bmatrix} \end{array}
\begin{array}{l} 3 \\ N-3 \end{array}}_{N\times N}
$$

$$\boxed{W = U_1 \; \Sigma_1 \; V_1^T}$$

$$(2F\times 3)(3\times 3)(3\times P)$$

# Factorization (Finding M , S)

$$W = U_1 (\Sigma_1)^{\frac{1}{2}} (\Sigma_1)^{\frac{1}{2}} V_1^T$$

$(2F \times 3)$ $(3 \times N)$

$= M$ ? $= S$ ?

Not so fast. Decomposition not unique!

For any 3X3 non-singular matrix Q :

$$W = U_1 (\Sigma_1)^{\frac{1}{2}} Q \; Q^{-1} (\Sigma_1)^{\frac{1}{2}} V_1^T$$ is also valid.

$(2F \times 3)$ $(3 \times N)$

$= M$ $= S$ for some Q

How to find the matrix Q ?

# Orthonormality of M

The Motion Matrix M:

$$M = \begin{bmatrix} i_1^T \\ \vdots \\ i_F^T \\ j_1^T \\ \vdots \\ j_F^T \end{bmatrix} = U_1(\Sigma_1)^{1/2} Q = \begin{bmatrix} \hat{i}_1^T \\ \vdots \\ \hat{i}_F^T \\ \hat{j}_1^T \\ \vdots \\ \hat{j}_F^T \end{bmatrix} Q = \begin{bmatrix} \hat{i}_1^T Q \\ \vdots \\ \hat{i}_F^T Q \\ \hat{j}_1^T Q \\ \vdots \\ \hat{j}_F^T Q \end{bmatrix}$$

Orthonormality Constraints:

$$i_f \cdot i_f = i_f^T i_f = 1$$
$$j_f \cdot j_f = j_f^T j_f = 1$$
$$i_f \cdot j_f = i_f^T j_f = 0$$

$$\Rightarrow$$

$$\hat{i}_f^T Q Q^T \hat{i}_f = 1$$
$$\hat{j}_f^T Q Q^T \hat{j}_f = 1$$
$$\hat{i}_f^T Q Q^T \hat{j}_f = 0$$

# Orthonormality of M

- We have computed $(\hat{i}_f^T, \hat{j}_f^T)$ for f = 1,…,F.

$$\left.\begin{aligned} \hat{i}_f^T Q Q^T \hat{i}_f &= 1 \\ \hat{j}_f^T Q Q^T \hat{j}_f &= 1 \\ \hat{i}_f^T Q Q^T \hat{j}_f &= 0 \end{aligned}\right\}$$

Q is unknown.

- Q is 3x3 matrix, 9 variables, 3F quadratic equations.
- Q can be solved with 3 or more images (F≥ 3) using Newton's method.

Final Solution: $\boxed{M = U_1 (\Sigma_1)^{\frac{1}{2}} Q}$ $\boxed{S = Q^{-1} (\Sigma_1)^{\frac{1}{2}} V_1^T}$

Camera Motion Scene struction

# Summary: Orthographic SFM

1. Detect and track feature points.
2. Create the centroid subtracted matrix w of corresponding feature points.
3. Compute SVD of W and enforce rank constraint.

$$W = U\,\Sigma\,V^T = \underset{(2F\times 3)}{U_1}\ \underset{(3\times 3)}{\Sigma_1}\ \underset{(3\times P)}{V_1^T}$$

4. Set $M = U_1\,(\Sigma_1)^{\frac{1}{2}}\,Q$ and $S = Q^{-1}\,(\Sigma_1)^{\frac{1}{2}}\,V_1^T$.
5. Find Q by enforcing the orthonormality constraint.

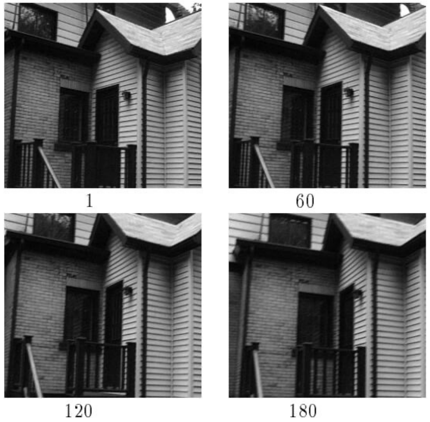
# Result



Input image sequence

Estimated 3D points

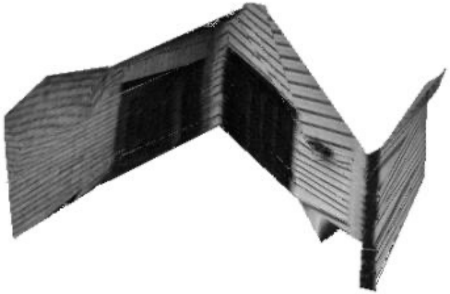
# Result




Input image sequence

Tracked features

3D reconstruction



3D reconstruction

# Structure from Motion Result