Computer Vision: Algorithms and Applications

Motion and Photography

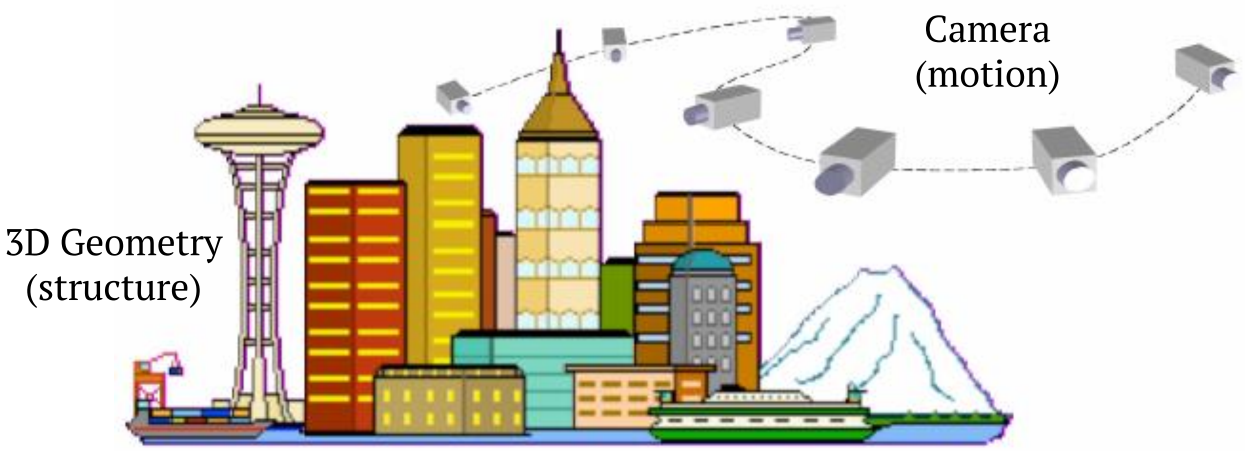
Jing Luo | Megvii Tech Talk | Mar 2018

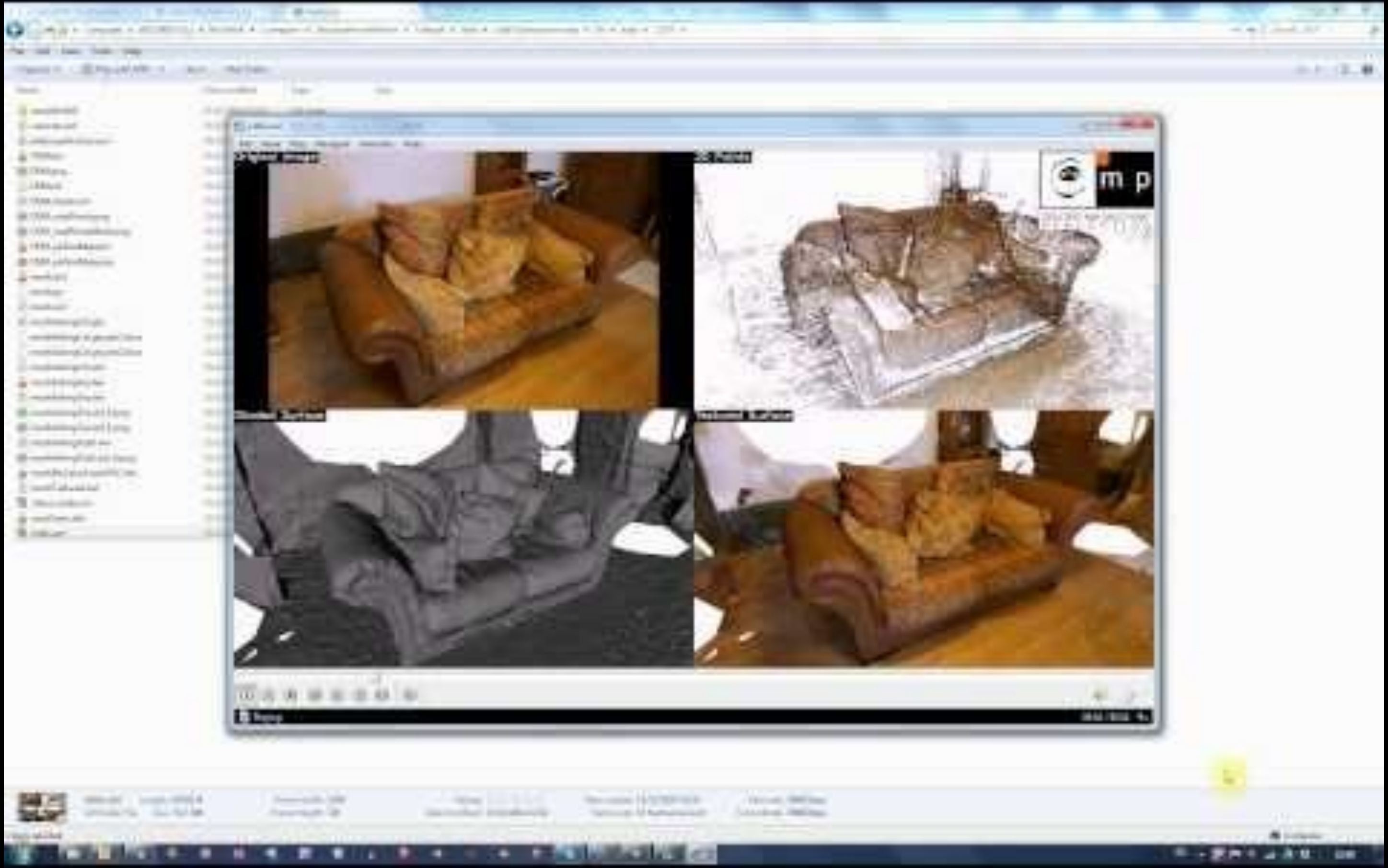
Reference: R. Szeliski. *Computer Vision: Algorithms and Applications.* 2010. 1.

1.
Structure from motion

Structure from motion

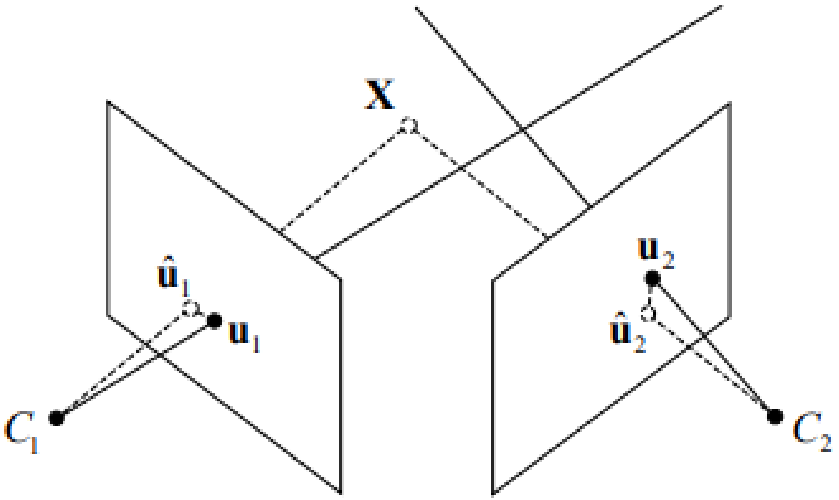
- Given a set of flow fields or displacement vectors from a moving camera time, determine:
  - the sequence of camera poses
  - the 3D structure of the scene2D points





Triangulation

- Determining a point's 3D position from a set of corresponding image locations and known camera positions.

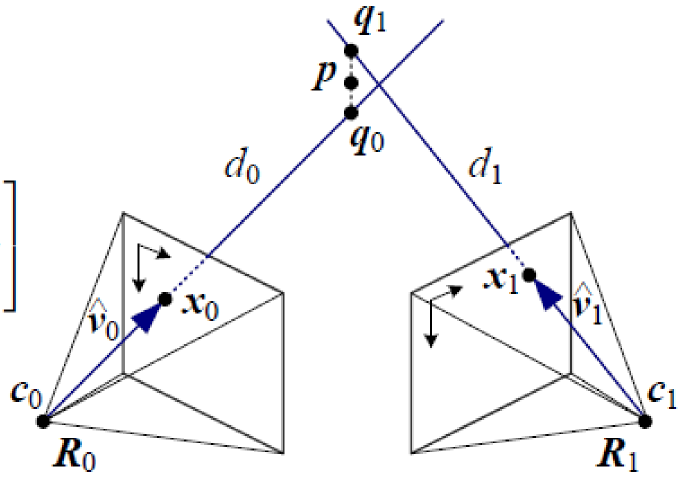


Triangulation

- Determining a point's 3D position from a set of corresponding image locations and known camera positions.

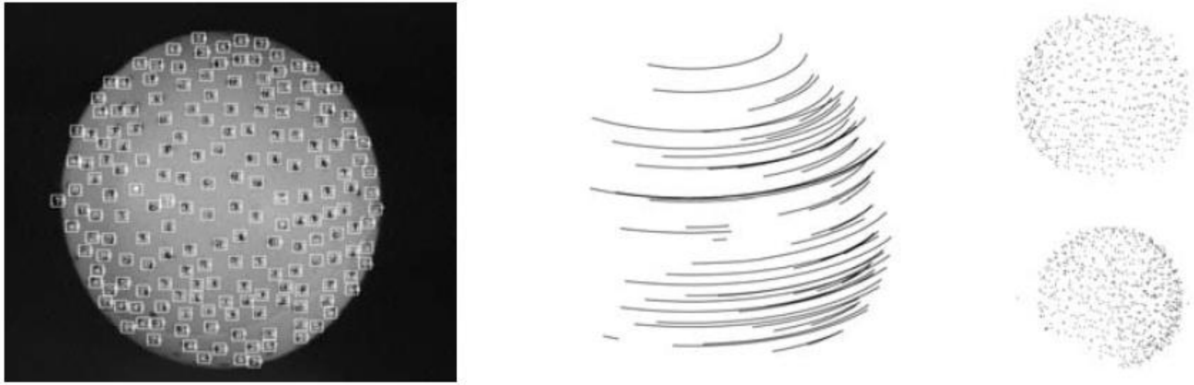
$$\|c_j + d_j \hat{v}_j - p\|^2$$

$$p = \left[\sum_j (I - \hat{v}_j \hat{v}_j^T)\right]^{-1} \left[\sum_j (I - \hat{v}_j \hat{v}_j^T) c_j\right]$$



Factorization

- When processing video sequences, we often get extended feature tracks which it is possible to recover the structure and motion using a process called factorization.

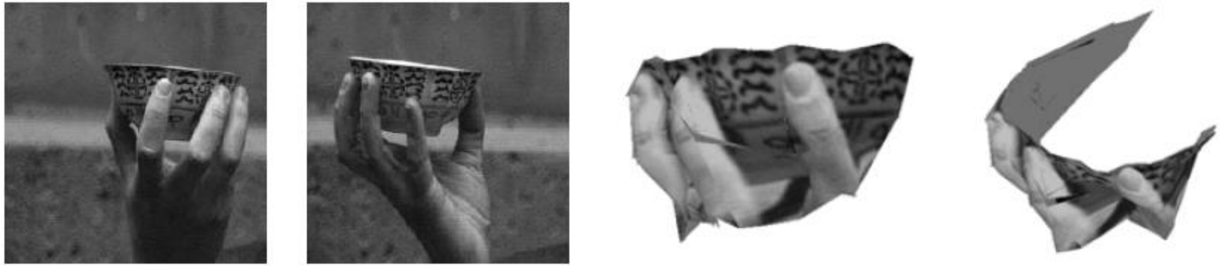


(a) (b) (c)

Figure 7.5 3D reconstruction of a rotating ping pong ball using factorization (Tomasi and Kanade 1992) © 1992 Springer: (a) sample image with tracked features overlaid; (b) sub-sampled feature motion stream; (c) two views of the reconstructed 3D model.

Factorization

- Assumption
  - Pixel correspondence
    - via tracking
  - Projection model
    - classic methods are orthographic
    - newer methods use perspective
  - The positions of “P” points in “F” frames (F>=3), which are not all coplanar, and have been tracked.



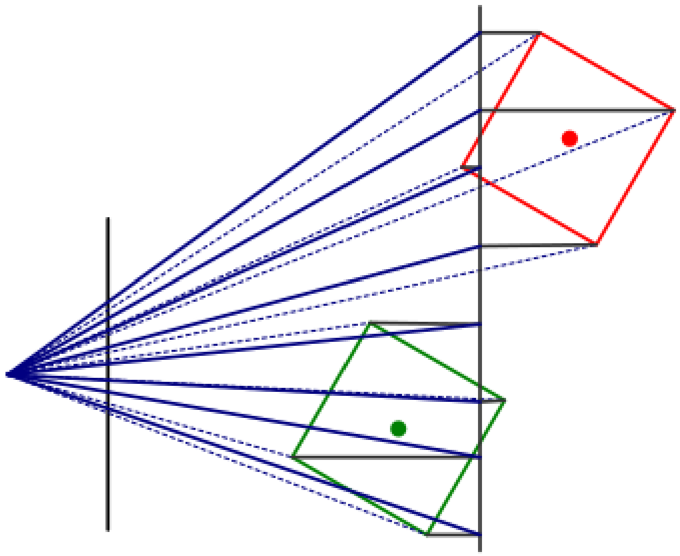
(a) (b) (c) (d)

3D to 2D projections

- Orthography

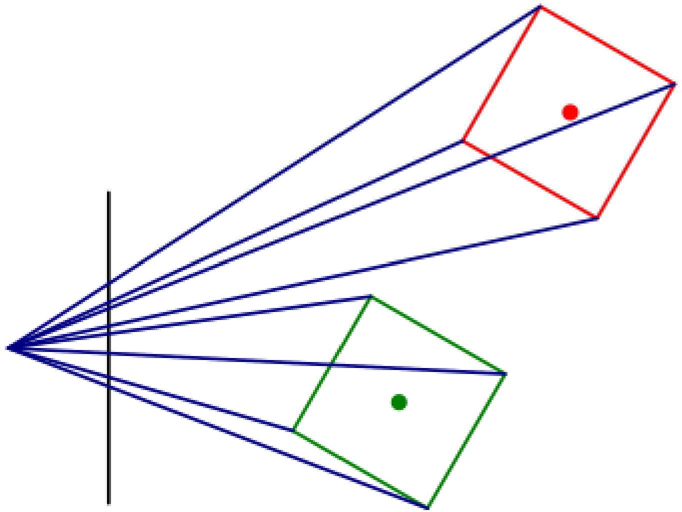
$$\boldsymbol{x} = [\boldsymbol{I}_{2\times 2}|\boldsymbol{0}]\,\boldsymbol{p}$$

$$\tilde{\boldsymbol{x}} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \tilde{\boldsymbol{p}}$$



3D to 2D projections

- 3D perspective



$$\bar{\boldsymbol{x}} = \mathcal{P}_z(\boldsymbol{p}) = \begin{bmatrix} x/z \\ y/z \\ 1 \end{bmatrix}$$

$$\tilde{\boldsymbol{x}} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \tilde{\boldsymbol{p}}$$

Factorization

- Feature points

$$\left\{(u_{fp}, v_{fp}) \mid f = 1,\ldots,F,\ p = 1,\ldots,P\right\}$$

$$W = \begin{bmatrix} u_{11} \ldots u_{1p} \\ \vdots \\ u_{F1} \ldots u_{FP} \\ v_{11} \ldots v_{1P} \\ \vdots \\ v_{F1} \ldots v_{FP} \end{bmatrix} \qquad W = \begin{bmatrix} U \\ - \\ V \end{bmatrix}$$

Factorization

- Mean normalize feature points

$$a_f = \frac{1}{P}\sum_{p=1}^{P} u_p \qquad b_f = \frac{1}{P}\sum_{p=1}^{P} v_p$$

$$\widetilde{u}_{fP} = u_{fP} - a_{fP}$$

$$\widetilde{v}_{fP} = v_{fP} - b_{fP}$$

Factorization

- Orthographic projection

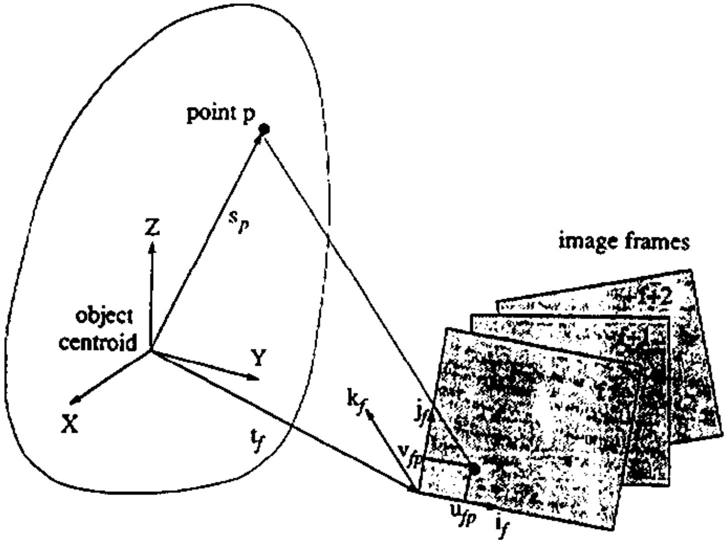
$$s_p = (X_p, Y_P, Z_P)$$

$$k_f = i_f \times j_f$$

i, j, k are unit vectors along X, Y, Z

$$u_{fP} = i_f^T (s_P - t_f)$$

$$v_{fP} = j_f^T (s_P - t_f)$$



Factorization

- Subtract mean

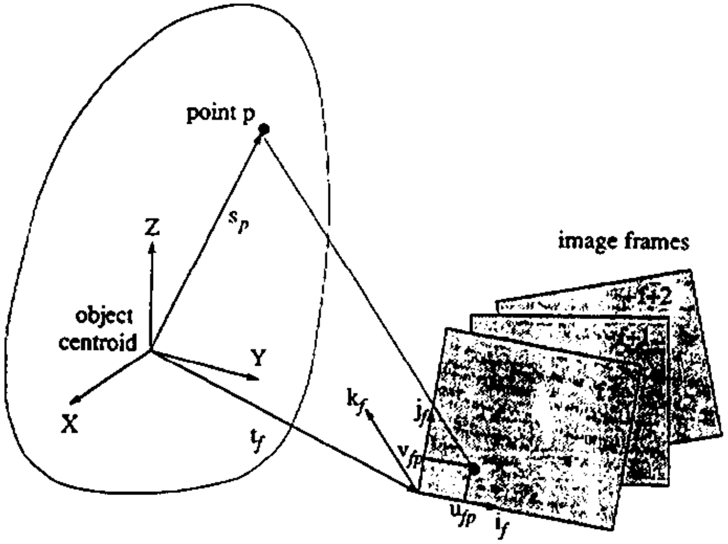
$$\tilde{u}_{fp} = u_{fP} - a_f$$

$$u_{fP} = i_f^T (s_P - t_f) \qquad a_f = \frac{1}{P}\sum_{p=1}^{P} u_p$$

$$= i_f^T (s_p - t_f) - \frac{1}{P}\sum_{q=1}^{P} i_f^T (s_q - t_f)$$

$$= i_f^T \left[s_P - \frac{1}{P}\sum_{q=1}^{P} s_q \right]$$

$$= i_f^T s_P$$



If Origin of world is at the centroid of object points, second term is zero.

Factorization

- Measurement matrix

$$\tilde{u}_{fP} = i_f^T s_P$$

$$\tilde{v}_{fP} = j_f^T s_P$$

$$\tilde{W} = \begin{bmatrix} \tilde{U} \\ - \\ \tilde{V} \end{bmatrix} = \begin{bmatrix} \tilde{u}_{11} \ldots \tilde{u}_{1p} \\ \vdots \\ \tilde{u}_{F1} \ldots \tilde{u}_{FP} \\ \tilde{v}_{11} \ldots \tilde{v}_{1P} \\ \vdots \\ \tilde{v}_{F1} \ldots \tilde{v}_{FP} \end{bmatrix} = \underbrace{\begin{bmatrix} i_1^T \\ \vdots \\ i_F^T \\ j_1^T \\ \vdots \\ j_F^T \end{bmatrix}}_{2F \times 3} \underbrace{\begin{bmatrix} s_1 & \ldots & s_P \end{bmatrix}}_{3 \times P} = RS$$

Factorization

- Rank theorem

$$\widetilde{W} = \begin{bmatrix} i_1^T \\ \vdots \\ i_F^T \\ j_1^T \\ \vdots \\ j_F^T \end{bmatrix} \begin{bmatrix} s_1 & \dots & s_P \end{bmatrix} = RS$$

2F × 3

3 × P

Centered measurement matrix

Without noise, the centered measurement matrix is at most of rank three.
Proof.

Motion
(camera rotation)

Structure
(3D scene points)
Rank of ***S*** is 3, because points in 3D space are not Co-planar.

Factorization

- Importance of rank theorem
  - Shows that video data is highly redundant
  - Precisely quantifies the redundancy
  - Suggests an algorithm for solving SFM
- SVD

$$\widetilde{W} = O_1 \Sigma O_2$$

- Theorem:
  - Any m by n matrix A, for which $m \geq n$,can be written as

$$\underset{m\times n}{A} = \underset{m\times n}{O_1}\,\underset{n\times n}{\Sigma}\,\underset{n\times n}{O_2}$$

Σ is diagonal

O_1, O_2 are orthogonal

$O_1^T O_1 = O_2^T O_2 = I$

Factorization

- Importance of rank theorem
  - Shows that video data is highly redundant
  - Precisely quantifies the redundancy
  - Suggests an algorithm for solving SFM
- SVD

$$\tilde{W} = O_1 \Sigma O_2$$

- Theorem:
  - Any m by n matrix A, for which $m \geq n$, can be written as

$$\underset{mxn}{A} = \underset{mxn}{O_1} \underset{nxn}{\Sigma} \underset{nxn}{O_2}$$

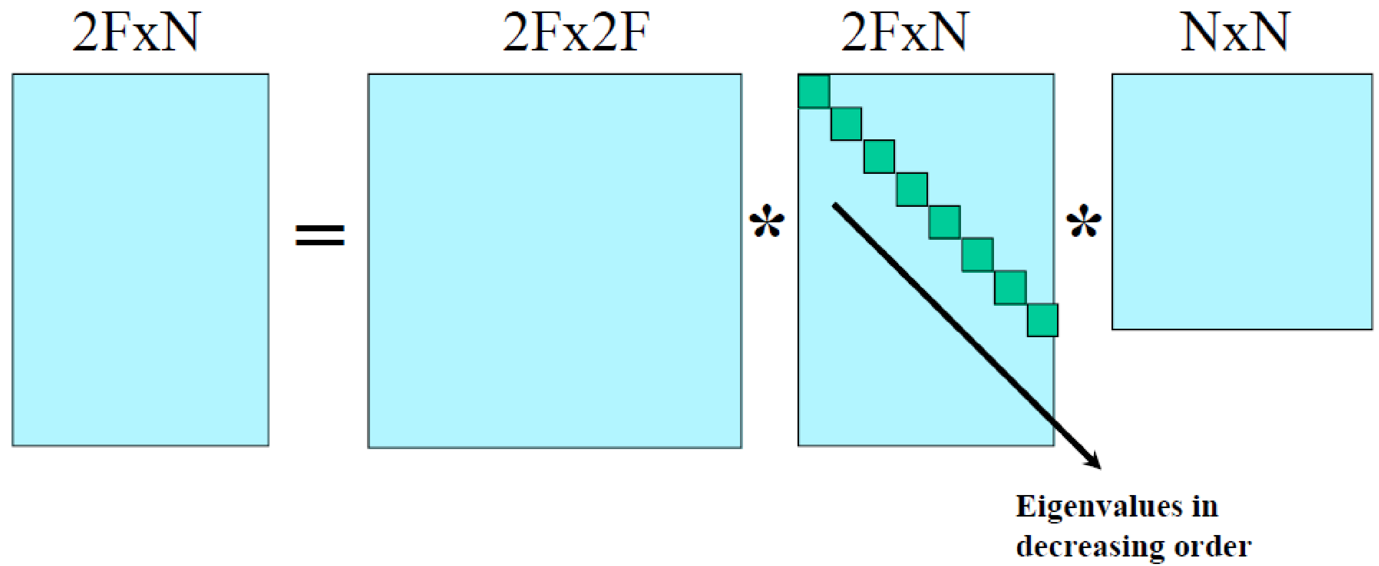
Σ is diagonal

O_1, O_2 are orthogonal

$O_1^T O_1 = O_2^T O_2 = I$

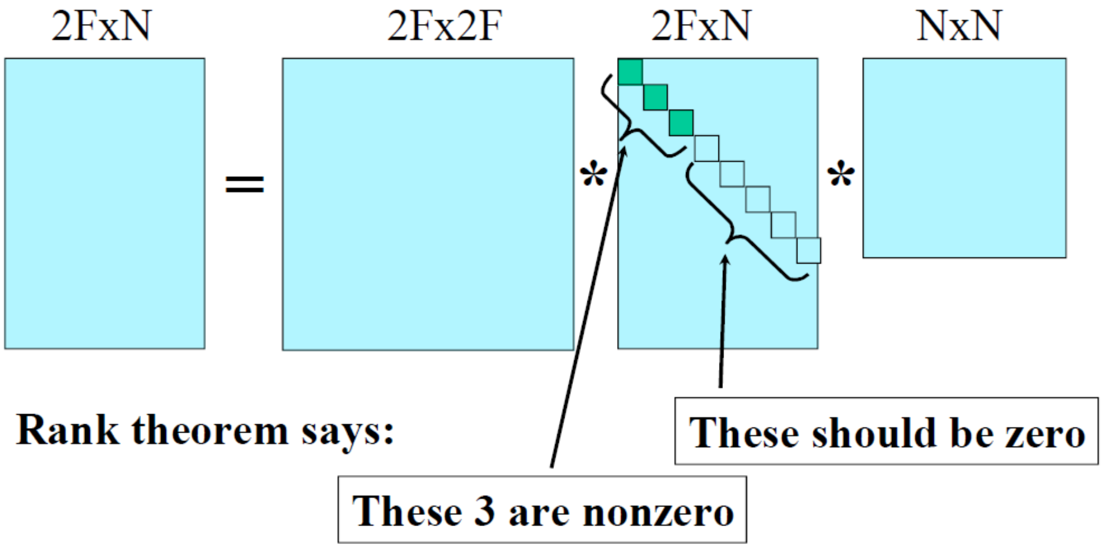
Factorization

- SVD



Factorization

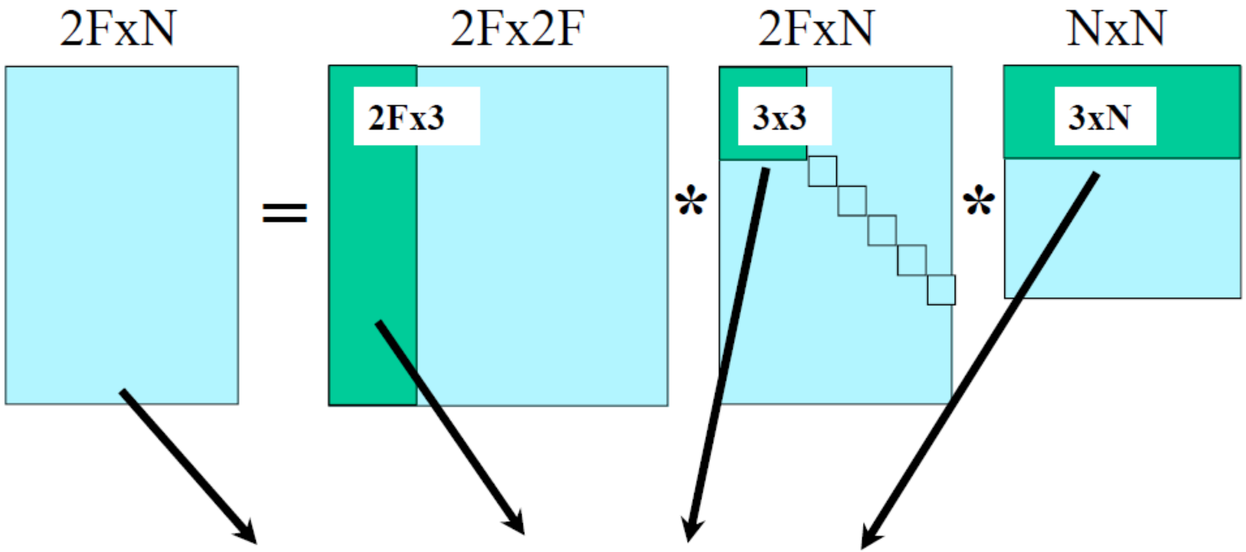
- SVD



In practice, due to noise, there may be more than 3 nonzero eigenvalues, but rank theorem tells us to ignore all but the largest three.

Factorization

- SVD



$$W = U' \; D' \; V'^{T}$$

Factorization

- Factorization approach

Observed image points

$$W \overset{SVD}{=} U' \; D' \; V'^{T}$$

$$\underset{2F\times N}{W} = \underset{2F\times 3}{\underline{U' \; D'^{1/2}}} \; \underset{3\times N}{\underline{D'^{1/2} \; V'^{T}}}$$

$$W = M \; S$$

Camera motion (M) **Scene structure** (S)

Factorization

- Summary

Assumptions

- orthographic camera

- N non-coplanar points tracking in F>=3 frames

Form the centered measurement matrix $W=[\tilde{X} ; \tilde{Y}]$

- **where $\tilde{x}_{ij} = x_{ij} - mx_j$**
- **where $\tilde{y}_{ij} = y_{ij} - my_j$**
- **mx_j and my_j are mean of points in frame i**
- **j ranges over set of points**

Rank theorem: The centered measurement matrix has a rank of at most 3

Factorization

1) **Form the centered measurement matrix W from N points tracked over F frames.**
2) **Compute SVD of $W = U D V^T$**
   - **U is 2Fx2F**
   - **D is 2FxN**
   - **V^T is NxN**
3) **Take largest 3 eigenvalues, and form**
   - **D' = 3x3 diagonal matrix of largest eigenvalues**
   - **U' = 2Fx3 matrix of corresponding column vectors from U**
   - **V'^T = 3xN matrix of corresponding row vectors from V^T**
4) **Define**

   $$M = U' D'^{1/2} \quad \text{and} \quad S = D'^{1/2} V'^T$$

5) **Solve for Q that makes appropriate rows of M orthogonal**
6) **Final solution is**

   $$M^* = M Q \quad \text{and} \quad S^* = Q^{-1} S$$