lect 17 Structure from Motion

Different from previous stereo problem, we know 3-D info.

we want to know camera's motion (movement's, param).

If we have m views, → m, projection matrix P_i } unknown
n points. X_i

$$\lambda_{ij} x_{ij} = P_j X_i \quad j = 1 \cdots n, \quad i = 1 \cdots m$$

(P_j: 3×4, X_i: 4×1)

actually, not uniquely solvable.

say. $\frac{P_j}{k} \cdot k X_i$

$= P_i \cdot X_i$

$= P_j Q^{-1} Q X_i$

We don't know which type of transformation (in 3D, not 2D)

Projection 15-DoF 4×4 $Q = \begin{bmatrix} A & t \\ v^T & v \end{bmatrix}$ Preserve intersection and tangency

Affine 12-DoF 4×4 $Q = \begin{bmatrix} A & t \\ 0^T & 1 \end{bmatrix}$ Preserve parallel, volumn ratio

Similarity 7 DoF $Q = \begin{bmatrix} sR & t \\ 0 & 1 \end{bmatrix}$ Preserve angle, length ratio

Euclidean 6 DoF $Q = \begin{bmatrix} R & t \\ 0^T & 1 \end{bmatrix}$ Preserve, angle, length

So, depends on which kind of Q we want, we may have projection up to {projective, affine, similarity, euclidean} Ambiguity.

Start with weak perspective camera.

① Orthographic Projection

$$\underbrace{\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}}_{P} \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix} = \begin{bmatrix} X \\ Y \\ 1 \end{bmatrix}$$

② Affine Camera

Image/Film may not be orthogonal to ray/projection direction

$$P = \underbrace{\begin{bmatrix} 3\times 3 \text{ affine} \end{bmatrix}}_{\text{Affine Trans for image}} \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \underbrace{\begin{bmatrix} 4\times 4 \text{ affine} \end{bmatrix}}_{\text{Affine Trans for 3D-point}} = \underbrace{\begin{bmatrix} a_{11} & a_{12} & a_{13} & b_1 \\ a_{21} & a_{22} & a_{23} & b_2 \\ 0 & 0 & 0 & 1 \end{bmatrix}}_{\text{8-DoF}} = \begin{bmatrix} A & b \\ 0 & 1 \end{bmatrix}$$

camera center ↓

$$P \cdot \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix} = \underset{2\times 3}{A} \cdot \begin{bmatrix} X \\ Y \\ Z \end{bmatrix} + \begin{bmatrix} b_1 \\ b_2 \end{bmatrix}$$

When this problem is sovable?

If we have m cameras, n points (2-D) → known $2mn$

$8m + 3n$ → unknown

↑ Camera ↑ 3D-point

Affine 12-DoF ↓

P has 12-DoF. So, if $2mn \geq 8m + 3n - 12$

Step:

1. Centering (Remove "b")

$$\hat{x}_{ij} = x_{ij} - \frac{1}{n}\sum x_{ij} = A_i X_j + b_i - \frac{1}{n}\sum_{k}^{n}(A_i X_k + b_i)$$

$$= A_i \hat{X}_j$$

$$\uparrow$$

$$X_j - \frac{1}{n}\sum_{k=1}^{n} X_k$$

2. Create $2m \times n$ data matrix

$$D = \begin{bmatrix} \hat{x}_{11} & & \hat{x}_{1n} \\ \hat{x}_{21} & \cdots & \vdots \\ \vdots & & \vdots \\ \hat{x}_{m1} & & \hat{x}_{mn} \end{bmatrix} \quad 2m \qquad \hat{x}_{ij} = \begin{pmatrix} \hat{x}_{ij} \\ \hat{y}_{ij} \end{pmatrix}$$

n

$$= \begin{bmatrix} A_1 \\ \vdots \\ A_m \end{bmatrix} [X_1 \cdots X_n] \quad \text{and } D \text{ is rank-3}$$

$2m \times 3$ $\qquad$ $3 \times n$

$\downarrow$

SVD, → choose top 3 - u, λ, v.

↳ $D = U \Sigma V$ $\qquad$ assume $\begin{bmatrix} A_1 \\ \vdots \\ A_m \end{bmatrix} = U \cdot \Sigma^{1/2}$. $[X_1 \cdots X_n] = W^{1/2} V^T$

$2m\times3 \quad 3\times3 \quad 3\times n$

But, it's true that this assumption may be wrong.

Say switch. columns, ...

3. More constrain

We want $\tilde{A}_i \tilde{A}_i^T = I$ (we want orthographic projection)

So, we find $A_i C_i (A_i C_i)^T = A_i (C_i C_i^T) A_i^T = I$, where $L = C_i C_i^T$

we find L, then use

Cholesky decomposition to find C

For projective Structure of Motion.

We can't use affine transformation's method, althrough the formula is still something like

$$\lambda_{ij} x_{ij} = P_i X_j$$

But, it is of homogenious representation, $\| P_i X_i - \lambda_{ij} x \|$ is not distance.

1, caliberate two camera use fundamental matrix F.

Say using vanishing point, special designed 3D known structure.

2.