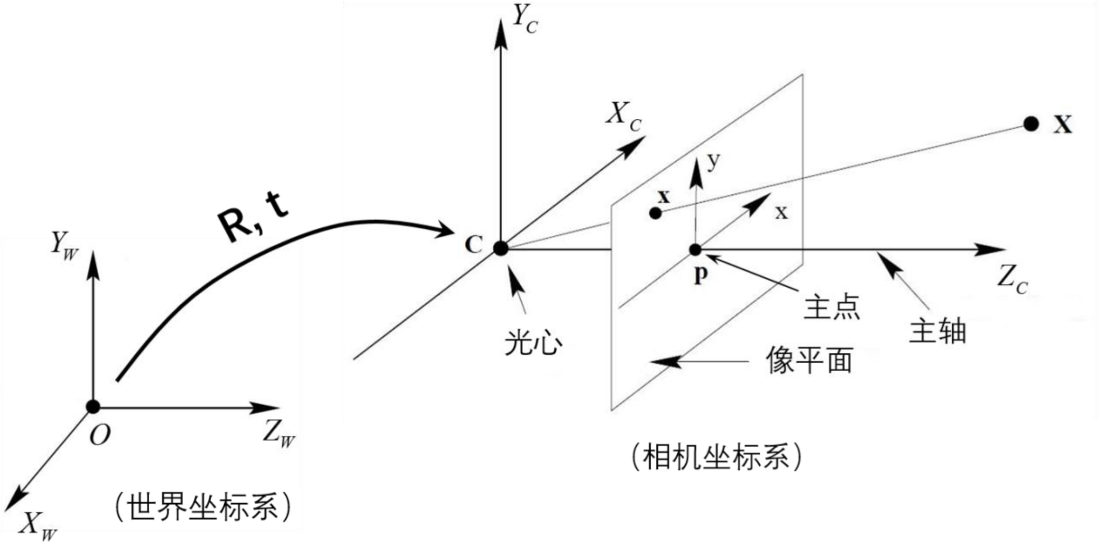
三维视觉初步——
稀疏重建与相机姿态估计

陶文兵

华中科技大学人工智能与自动化学院

2022-08-07

相机模型（Camera Model）

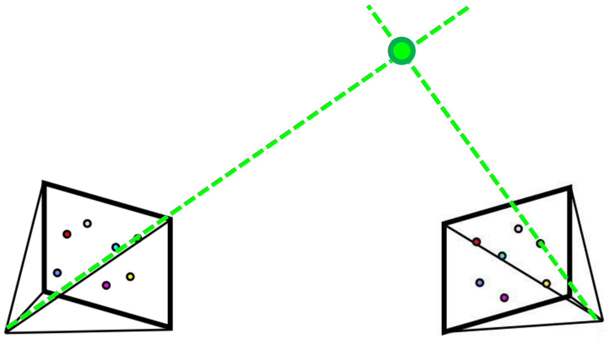


(内参数) $\mathbf{K} = \begin{bmatrix} f & 0 & u \\ 0 & f & v \\ 0 & 0 & 1 \end{bmatrix}$

(外参数) $\overset{3\times3}{\mathbf{R}} \quad \overset{3\times1}{\mathbf{t}}$

$$\begin{bmatrix} x \\ y \\ 1 \end{bmatrix} = \begin{bmatrix} f & 0 & u \\ 0 & f & v \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} R & t \\ 0^T & 1 \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix}$$

$$\mathbf{x} = \mathbf{K}\left[\mathbf{R} \middle| \mathbf{t}\right]\mathbf{X} = \boldsymbol{P}\mathbf{X}$$



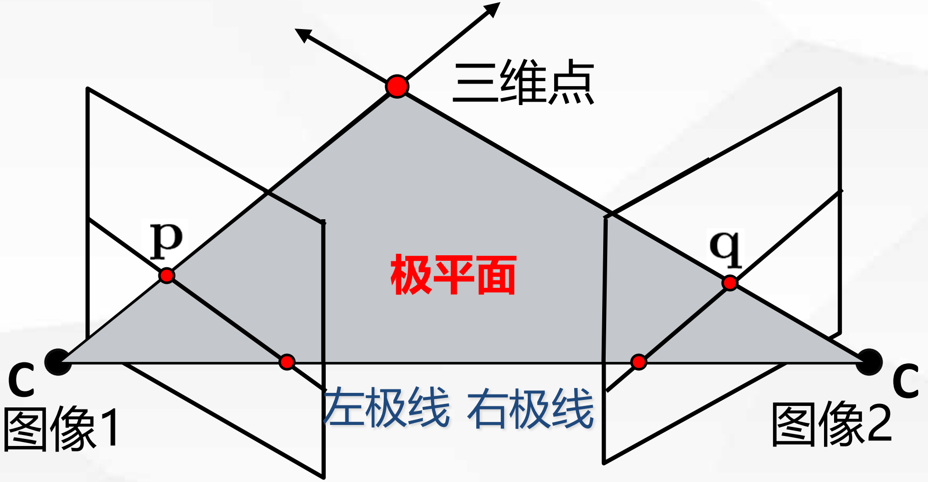
P = K [I | 0] P’ = K’ [R’ | t’]

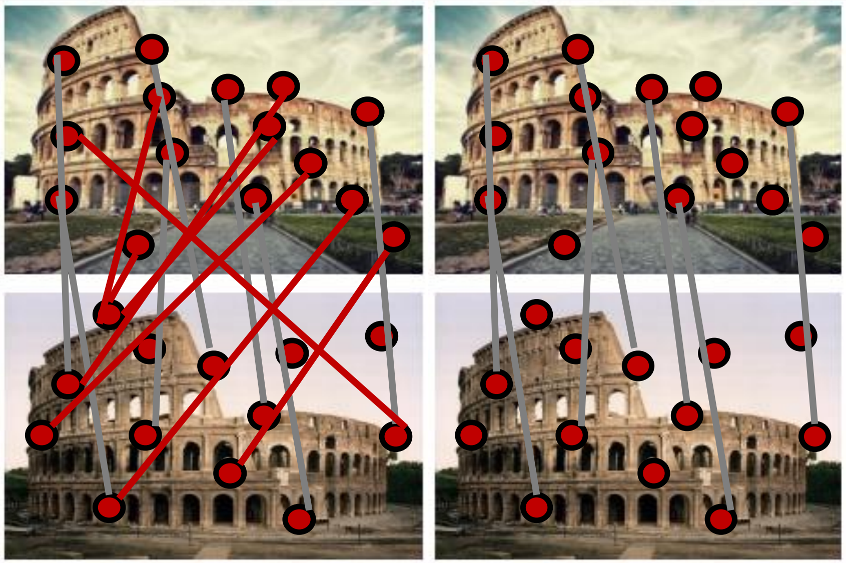
基础矩阵（Fundamental Matrix）

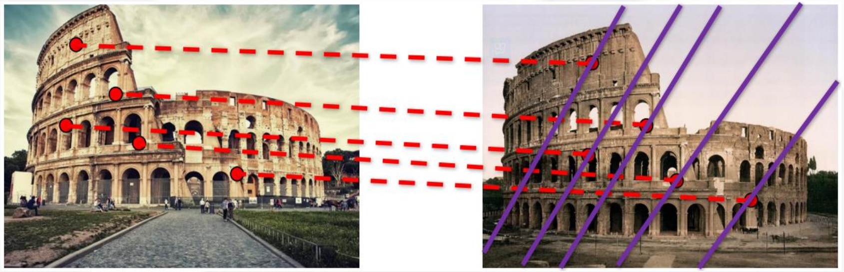
$$x'^{T} F x = 0$$

RANSAC

$$\begin{bmatrix} u_1u_1' & v_1u_1' & u_1' & u_1v_1' & v_1v_1' & v_1' & u_1 & v_1 & 1 \\ u_2u_2' & v_2u_2' & u_2' & u_2v_2' & v_2v_2' & v_2' & u_2 & v_2 & 1 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ u_nu_n' & v_nu_n' & u_n' & u_nv_n' & v_nv_n' & v_n' & u_n & v_n & 1 \end{bmatrix} \begin{bmatrix} f_{11} \\ f_{12} \\ f_{13} \\ f_{21} \\ f_{22} \\ f_{23} \\ f_{31} \\ f_{32} \\ f_{33} \end{bmatrix} = 0$$

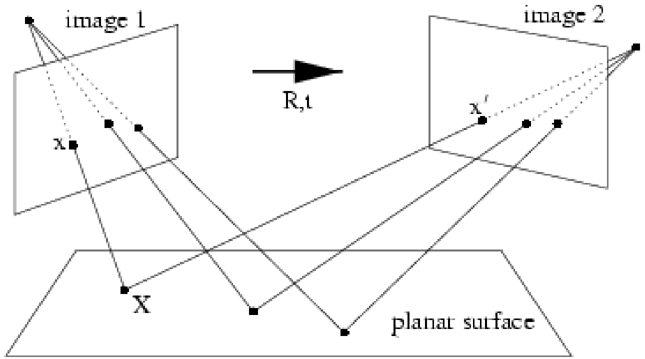






单应矩阵（Homography Matrix）

$$\begin{bmatrix} x' \\ y' \\ z' \end{bmatrix} = \begin{bmatrix} h_{11} & h_{12} & h_{13} \\ h_{21} & h_{22} & h_{23} \\ h_{31} & h_{32} & h_{33} \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \boldsymbol{H} \begin{bmatrix} x \\ y \\ z \end{bmatrix}$$

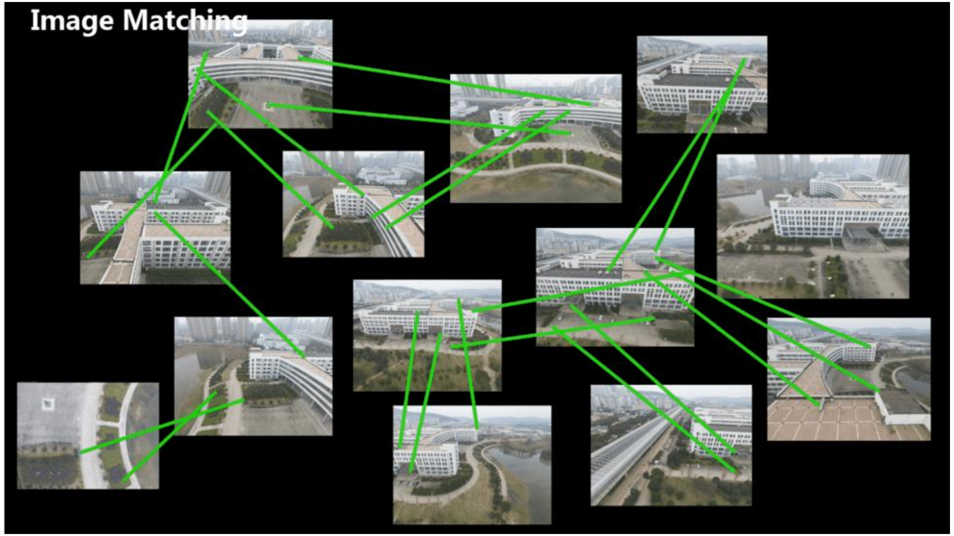


$$H = K'(R - t\frac{n^T}{d})K^{-1}$$

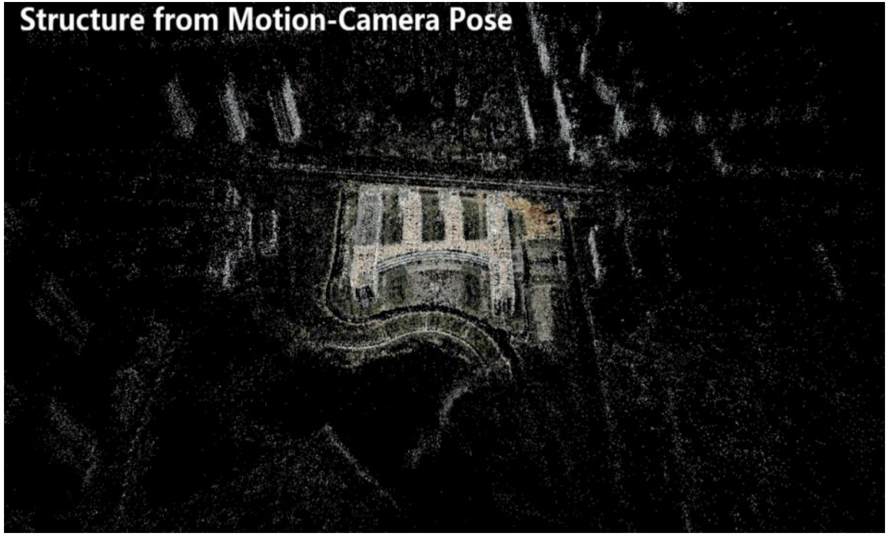
If n 3D points lie on the same planar surface, then the n matchs corresponding these 3D points satisfiy one homography transformation.

Sparse Reconstruction and Pose Estimation

Image Feature Matching



Camera Pose Estimation



Build tracks from matches

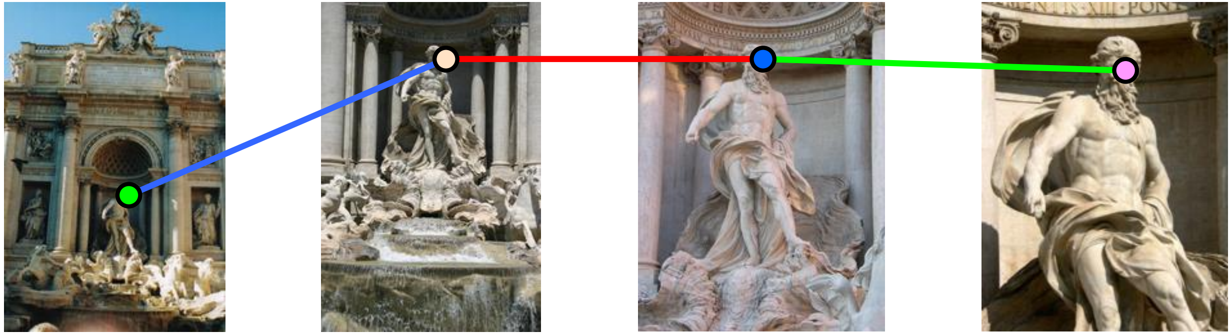


Image 1 Image 2 Image 3 Image 4

- Link up matches between pairs of images into tracks between multiple images
- Each track corresponds to a 3D point

Pose estimation—— Incremental Structure from Motion(SFM)

- Initialization
  - 1. Choose two views

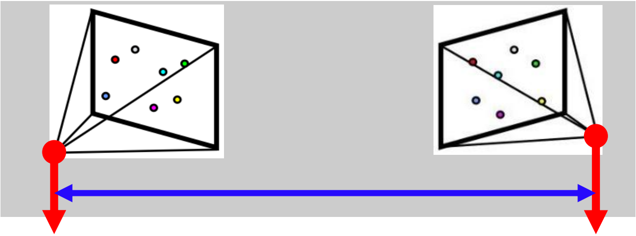
■ *They have the most number of correspondences*
■ *They have wide baselines*

lots of correspondences

narrow baseline

wide baseline

few correspondences

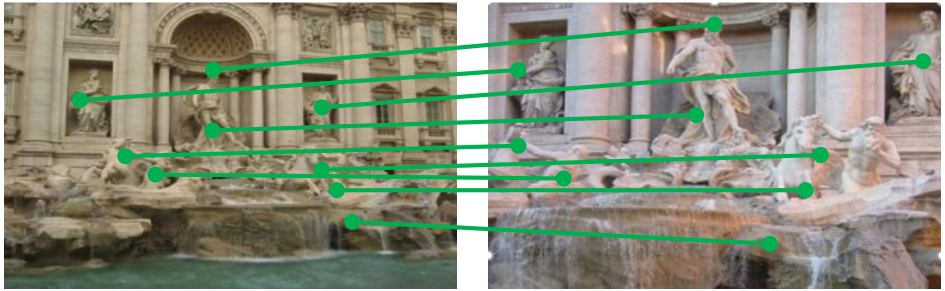


wide baseline

lots of correspondences

Incremental Structure from Motion (SFM)

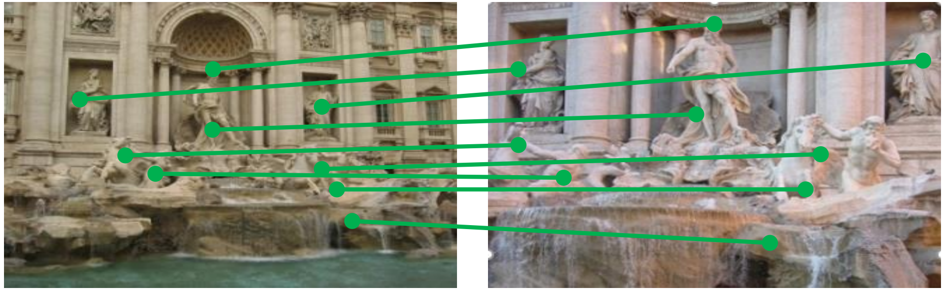
- Initialization
  - 1. Choose two views
  - 2. Estimate relative pose using two-view geometry



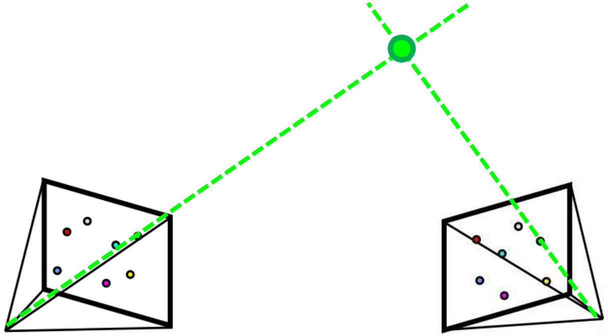
- *Camera intrinsics known*
  Essential matrix, ***E*** (5 points)
- *Camera intrinsics unknown*
  Fundamental matrix, ***F*** (7 points)

Incremental Structure from Motion

- Initialization
  - 1. Choose two views
  - 2. Estimate relative pose using two-view geometry
  - 3. Triangulate inlier correspondences

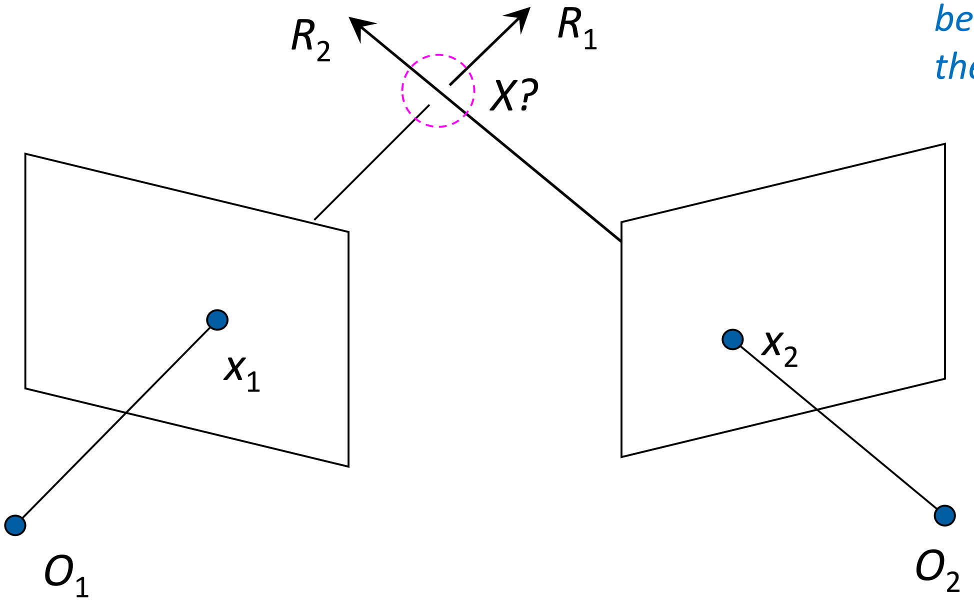


- *Given projections of a 3D point in two or more images (with known camera matrices), find the coordinates of the point*



Incremental Structure from Motion

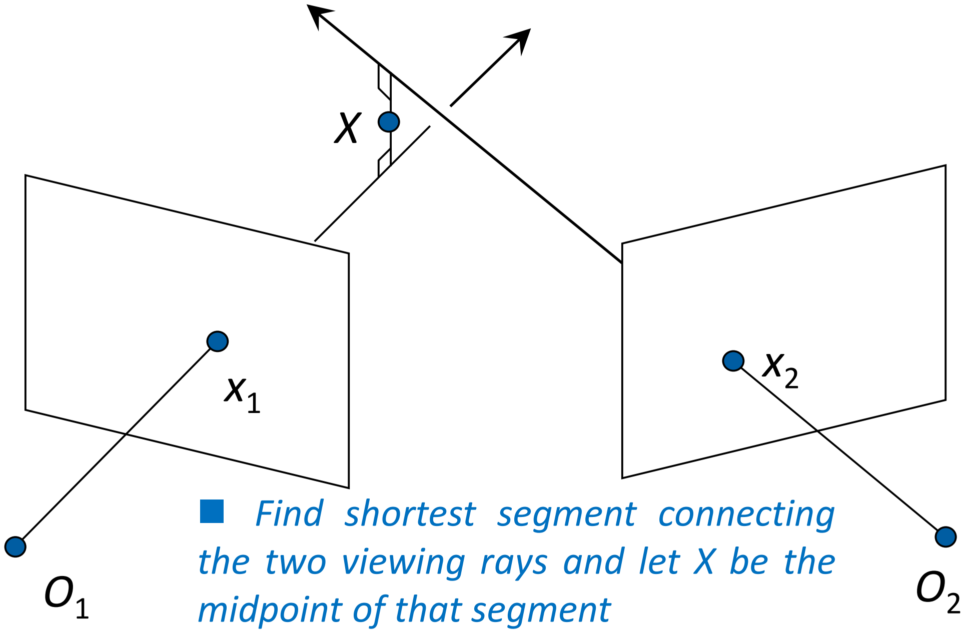
- Initialization
  - 1. Choose two views
  - 2. Estimate relative pose using two-view geometry
  - 3. Triangulate inlier correspondences



■ *We want to intersect the two visual rays corresponding to x1 and x2, but because of noise and numerical errors, they don't meet exactly*

Incremental Structure from Motion

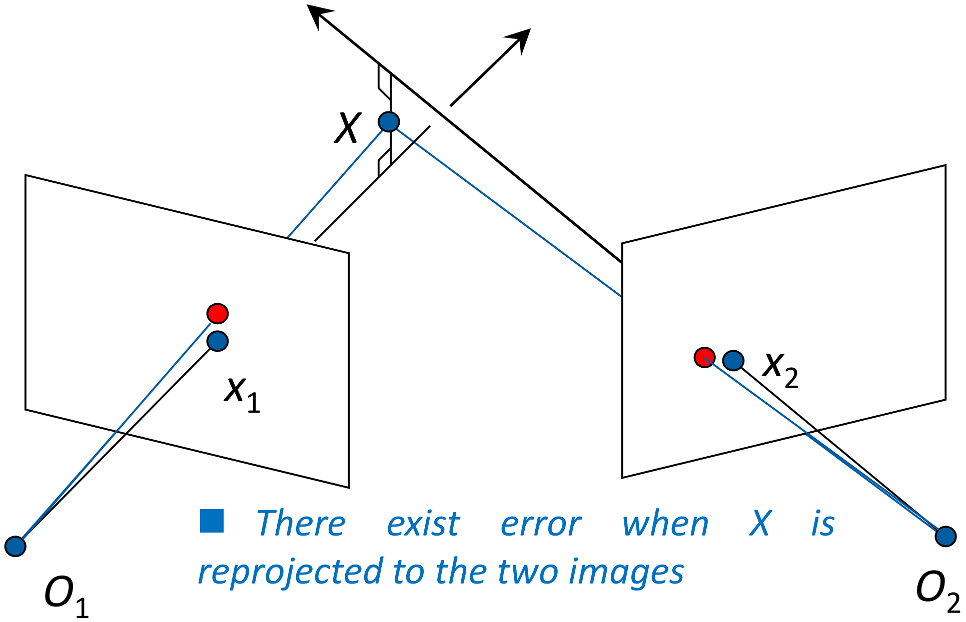
- Initialization
  - 1. Choose two views
  - 2. Estimate relative pose using two-view geometry
  - 3. Triangulate inlier correspondences
  - 4. Bundle adjustment



■ *Find shortest segment connecting the two viewing rays and let X be the midpoint of that segment*

Incremental Structure from Motion

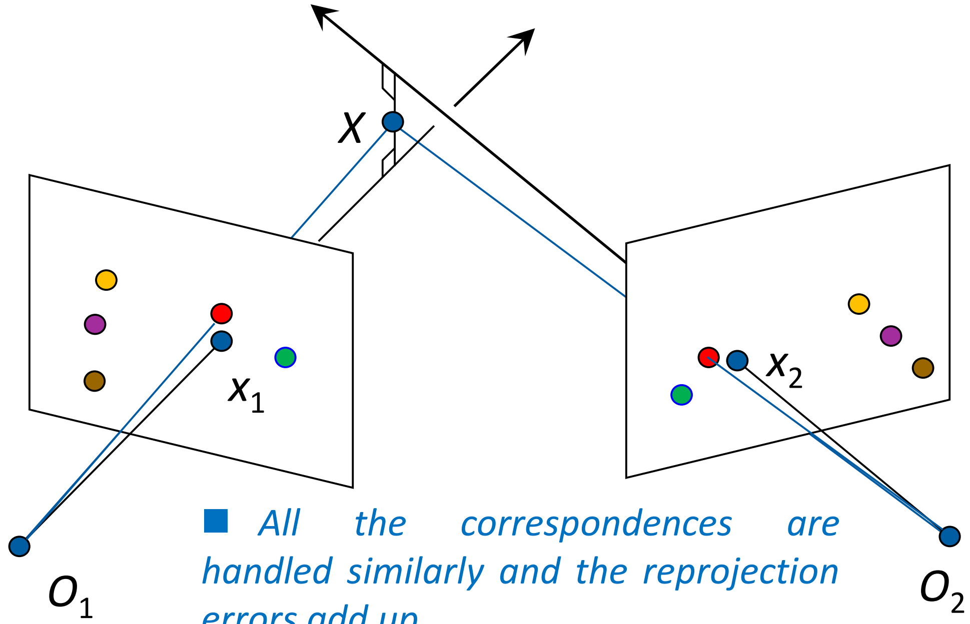
- Initialization
  - 1. Choose two views
  - 2. Estimate relative pose using two-view geometry
  - 3. Triangulate inlier correspondences
  - 4. Bundle adjustment



■ *There exist error when X is reprojected to the two images*

Incremental Structure from Motion

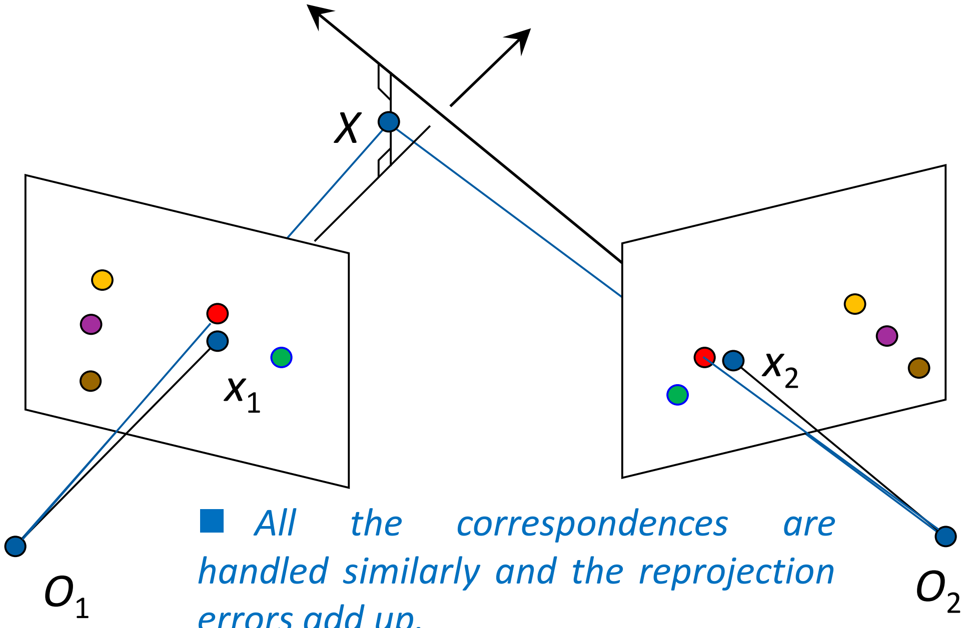
- Initialization
  - 1. Choose two views
  - 2. Estimate relative pose using two-view geometry
  - 3. Triangulate inlier correspondences
  - 4. Bundle adjustment



■ *All the correspondences are handled similarly and the reprojection errors add up.*

Incremental Structure from Motion

- Initialization
  - 1. Choose two views
  - 2. Estimate relative pose using two-view geometry
  - 3. Triangulate inlier correspondences
  - 4. Bundle adjustment



- *All the correspondences are handled similarly and the reprojection errors add up.*

- *refine 3D points*
- *refine camera parameters*
- *Minimize reprojection error:*

$$E(\mathbf{P},\mathbf{X}) = \sum_{i=1}^{m}\sum_{j=1}^{n} w_{ij} D\left(\mathbf{x}_{ij}, \mathbf{P}_i \mathbf{X}_j\right)^2$$

w_{ij} indicator variable for visibility of point $\mathbf{X}_j$ in camera $\mathbf{P}_i$

- *Non-linear least squares problem*
- *Minimizing by Levenberg-Marquardt (LM), is called **Bundle Adjustment (BA)***