3D Computer Vision
Syllabus

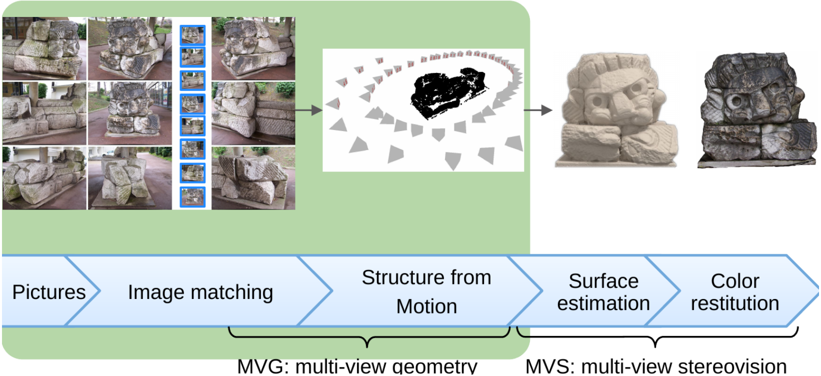
Pascal Monasse
pascal.monasse@enpc.fr

IMAGINE/LIGM, École nationale des ponts et chaussées

ÉCOLE NATIONALE DES PONTS ET CHAUSSÉES

IP PARIS

Structure from Motion

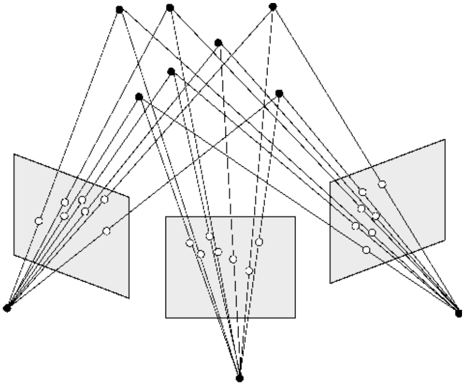


We will cover the first part of this Stereo pipeline, MVG, and first part of MVS but only with two views.

- Output of MVG: point cloud and camera positions and orientations
- Output of MVS: a mesh surface with colors

Bundle Adjustment

- Observed points in images are central projections of unknown 3D points.
- Each image provides a bundle of half-lines containing the 3D points.
- We must orient the bundles so that the lines intersect.



Source: Manolis Lourakis

One equation to rule them all

$$\arg \min_{\{R_i\},\{T_i\},\{\mathrm{X}_j\}} \sum_{i,j} \epsilon_{ij} \; d\left(\mathrm{x}_{ij}, \Pi_i(R_i \mathrm{X}_j + T_i)\right)^2 .$$

- x_{ij}: point number j (among n) in view i (among m), pixel coordinates.
- $\epsilon_{ij} \in \{0, 1\}$: visibility of point j in view i.
- d: Euclidean distance in image.
- Π_i: 2D projection operator in view i from 3D.

Unknowns:

- X_j: 3D point in a fixed coordinate frame.
- R_i: rotation matrix of coordinate frame wrt view i.
- T_i: position of origin O of coordinate frame wrt view i.

Session 1

$$\arg \min_{\{R_i\},\{T_i\},\{\mathrm{x}_j\}} \sum_{i,j} \epsilon_{ij} \; d\left(\mathrm{x}_{ij}, \Pi_i(R_i \mathrm{X}_j + T_i)\right)^2 .$$

- Expression of projection model Π_i (i.e., camera model)
- Case $m = 2$, $T_L = T_R = 0$ or all X_j on a single plane:

$$\forall i \in \{L, R\}, \forall j, \; d\left(\mathrm{x}_{ij}, \Pi_i(R_i \mathrm{X}_j + 0)\right) = 0.$$

  Then $\mathrm{x}_{Rj} = H_{\Pi_R, R_R}(\mathrm{x}_{Lj})$, and the *homography* H can be determined from four pairs $\{\mathrm{x}_{Lj}, \mathrm{x}_{Rj}\}$. Application: panorama.
- Case $m = 1$, $n \geq 6$ pairs $(\mathrm{x}_j, \mathrm{X}_j)$, recover all info about the camera (camera calibration from 3D rig):

$$\arg \min_{\Pi, R, T} \sum_{j} d\left(\mathrm{x}_{ij}, \Pi(R \mathrm{X}_j + T)\right)^2 .$$

- Case $m \geq 3$, single camera Π, 4 (or more) pairs $(\mathrm{x}_j, \mathrm{X}_j)$, all X_j on a common plane (camera calibration from 2D rig):

$$\arg \min_{\Pi, \{R_i\}, \{T_i\}} \sum_{ij} d\left(\mathrm{x}_{ij}, \Pi(R_i \mathrm{X}_j + T_i)\right)^2 .$$

Session 2

$$\arg \min_{\{R_i\},\{T_i\},\{\mathrm{X}_j\}} \sum_{i,j} \epsilon_{ij} \; d\left(\mathrm{x}_{ij}, \Pi_i(R_i \mathrm{X}_j + T_i)\right)^2 .$$

Case $m = 2$ views.

- From $n \geq 5$ pairs $\{\mathrm{x}_{Lj}, \mathrm{x}_{Rj}\}$ and known Π_L, Π_R, deduce a bilinear constraint $\forall j$, $E_{R_R, T_R}(\mathrm{x}_{Lj}, \mathrm{x}_{Rj}) = 0$ (essential matrix). Recover (R_R, T_R) from E.

$$\arg \min_{\{R_i\},\{T_i\},\{\mathrm{X}_j\}} \sum_{i \in \{L,R\},j} \epsilon_{ij} \; d\left(\mathrm{x}_{ij}, \Pi_i(R_i \mathrm{X}_j + T_i)\right)^2 .$$

- From $n \geq 7$ pairs $\{\mathrm{x}_{Lj}, \mathrm{x}_{Rj}\}$ and *unknown* Π_L, Π_R, deduce a bilinear constraint $\forall j$, $F(\mathrm{x}_{Lj}, \mathrm{x}_{Rj}) = 0$ (fundamental matrix).
- From n pairs $\{\mathrm{x}_{Lj}, \mathrm{x}_{Rj}\}$, maximize the number of matching pairs up to tolerance $\tau \geq 0$ (RANSAC algorithm)

$$\max \sum_{i \in \{L,R\},j} \epsilon_{ij}$$

$$\text{s.t. } \exists \Pi_L, \Pi_R, \{R_i\}, \{T_i\}, \{X_j\},\; d\left(\mathrm{x}_{ij}, \Pi_i(R_i \mathrm{X}_j + T_i)\right) \leq \tau.$$

Sessions 3&4

$$\arg\min_{\{R_i\},\{T_i\},\{X_j\}} \sum_{i,j} \epsilon_{ij}\; d\left(x_{ij}, \Pi_i(R_i X_j + T_i)\right)^2 .$$

Case $m = 2$ views, $R_L = R_R = I,\ T_L = 0,\ T_R = Be_1$ (rectified pair).

- Go from generic poses to rectified pair with fundamental matrix F.
- For all x_{Lj}, $j \in \{\text{pixels of left image}\}$, find corresponding pixel x_{Rj}, hence X_j by triangulation.
- Two categories of methods: local/global.

Session 5

$$\arg \min_{\{R_i\},\{T_i\},\{X_j\}} \sum_{i,j} \epsilon_{ij}\, d\left(x_{ij}, \Pi_i(R_i X_j + T_i)\right)^2 .$$

Case $m \geq 3$, multi-view stereovision.

- For $m = 3$, trilinear constraints $\mathcal{T}(x_{1j}, x_{2j}, x_{3j}) = 0$. Recovery of tensor $\mathcal{T}$.
- Recover R_{m+1} and T_{m+1} from known pairs (x_{m+1j}, X_j) (resection, PnP Perspective from n Points):

$$\arg \min_{R_{m+1}, T_{m+1}} \sum_{j} \epsilon_{m+1j}\, d\left(x_{ij}, \Pi_{m+1}(R_{m+1} X_j + T_{m+1})\right)^2 .$$

$\Rightarrow$ incremental multi-view pipeline.

- Special case $T_i = B_i v$ with v a fixed vector and known B_i:

$$\arg \min_{\{R_i\},\{X_j\}} \sum_{i,j} \epsilon_{ij}\, d\left(x_{ij}, \Pi_i(R_i X_j + B_i v)\right)^2 .$$

$\Rightarrow$ Light field imagery, epipolar plane imagery

Session 6

$$\arg \min_{\{R_i\},\{T_i\},\{\mathrm{X}_j\}} \sum_{i,j} \epsilon_{ij} \; d\left(\mathrm{x}_{ij}, \Pi_i(R_i \mathrm{X}_j + T_i)\right)^2 .$$

Case $m = 2$ views, find interest points $\{\mathrm{p}_{Lk}\}$ and $\{\mathrm{p}_{Rk'}\}$. Deduce pairs $j = (k, k')$ of matching points $(\mathrm{x}_{Lj}, \mathrm{x}_{Rj})$ with $\mathrm{x}_{Lj} = \mathrm{p}_{Lk}$ and $\mathrm{x}_{Rj} = \mathrm{p}_{Rk'}$.