3D Computer Vision
Syllabus

Pascal Monasse
pascal.monasse@enpc.fr

IMAGINE, École des Ponts ParisTech

This work is licensed under the Creative Commons Attribution 4.0 International License. To view a copy of this license, visit http://creativecommons.org/licenses/by/4.0/.
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

\[\arg \min_{\{R_i\}, \{T_i\}, \{x_j\}} \sum_{i,j} \epsilon_{ij} \ d(x_{ij}, \Pi_i(R_i x_j + T_i))^2.\]

- \(x_{ij}\): 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.
- \(\Pi_i\): 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 \(i\).
arg\ \min_{\{R_i\}, \{T_i\}, \{x_j\}} \sum_{i,j} \epsilon_{ij} \ d (x_{ij}, \Pi_i (R_i x_j + T_i))^2 .

Expression of projection model $\Pi_i$ (i.e., camera model)

Case $m = 2$, $T_L = T_R = 0$ or all $x_j$ on a single plane:

$$\forall i \in \{L, R\}, \forall j, d (x_{ij}, \Pi_i (R_i x_j + 0)) = 0.$$ 

Then $x_{Rj} = H_{\Pi, R, R} (x_{Lj})$, and the homography $H$ can be determined from four pairs $\{x_{Lj}, x_{Rj}\}$. Application: panorama.

Case $m = 1$, $n \geq 6$ pairs $(x_j, X_j)$, recover all info about the camera (camera calibration from 3D rig):

$$\arg\ \min_{\Pi, R, T} \sum_j d (x_{ij}, \Pi (RX_j + T))^2 .$$

Case $m \geq 3$, single camera $\Pi$, 4 (or more) pairs $(x_j, X_j)$, all $x_j$ on a common plane (camera calibration from 2D rig):

$$\arg\ \min_{\Pi, \{R_i\}, \{T_i\}} \sum_{ii} d (x_{ij}, \Pi (R_i x_j + T_i))^2 .$$
Session 2

\[
\arg \min_{\{R_i\}, \{T_i\}, \{X_j\}} \sum_{i,j} \epsilon_{ij} d(x_{ij}, \Pi_i(R_iX_j + T_i))^2.
\]

Case \(m = 2\) views.

▷ From \(n \geq 5\) pairs \(\{x_{Lj}, x_{Rj}\}\) and known \(\Pi_L, \Pi_R\), deduce a bilinear constraint \(\forall j, E_{RR}, T_R(x_{Lj}, x_{Rj}) = 0\) (essential matrix).
Recover \((R_R, T_R)\) from \(E\).

\[
\arg \min_{\{R_i\}, \{T_i\}, \{X_j\}} \sum_{i \in \{L,R\}, j} \epsilon_{ij} d(x_{ij}, \Pi_i(R_iX_j + T_i))^2.
\]

▷ From \(n \geq 7\) pairs \(\{x_{Lj}, x_{Rj}\}\) and unknown \(\Pi_L, \Pi_R\), deduce a bilinear constraint \(\forall j, F(x_{Lj}, x_{Rj}) = 0\) (fundamental matrix).

▷ From \(n\) pairs \(\{x_{Lj}, x_{Rj}\}\), maximize the number of matching pairs up to tolerance \(\tau \geq 0\) (RANSAC algorithm)

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

s.t. \(\exists \Pi_L, \Pi_R, \{R_i\}, \{T_i\}, \{X_j\}, d(x_{ij}, \Pi_i(R_iX_j + T_i)) \leq \tau\).
Sessions 3&4

\[
\arg \min_{\{R_i\}, \{T_i\}, \{x_j\}} \sum_{i,j} \epsilon_{ij} d (x_{ij}, \Pi_i(R_i x_j + T_i))^2.
\]

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.
arg \min_{\{R_i\},\{T_i\},\{X_j\}} \sum_{i,j} \epsilon_{ij} d (x_{ij}, \Pi_i(R_iX_j + T_i))^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):

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

⇒ incremental multi-view pipeline.
- Special case $T_i = B_i v$ with $v$ a fixed vector and known $B_i$:

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

⇒ Light field imagery, epipolar plane imagery
\[
\arg \min_{\{R_i\}, \{T_i\}, \{x_j\}} \sum_{i,j} \epsilon_{ij} d(x_{ij}, \Pi_i(R_i x_j + T_i))^2.
\]

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