Topics:

Basics of projective geometry
  Points and hyperplanes in projective space
  Homography
  Estimating homography from point correspondence
The single perspective camera
  An overview of single camera calibration
  Calibration of one camera from the known scene
Scene reconstruction from multiple views
  Triangulation
  Projective reconstruction
  Matching constraints
  Bundle adjustment
Two cameras, stereopsis
  The geometry of two cameras. The fundamental matrix
  Relative motion of the camera; the essential matrix
  Estimation of a fundamental matrix from image point correspondences
  Applications of the epipolar geometry in vision
Three and more cameras
  Stereo correspondence algorithms
Scene reconstruction from multiple views

**Task:** Given matching points in $n$ images. Determine the 3D scene point.

**Basic Principle:** Back-trace the ray in 3D scene for each image. The scene point is the common intersection of all rays.

**Information needed:** To back-trace a ray in the scene space, we need to know the corresponding camera matrix $M_j$.

**Challenges:**
- In an ideal condition, $m$ back-traced rays intersect at a common point in the scene space.
- However, in real applications, due to noise and other source of errors, single-point intersection may not happen.

**How to proceed?**

**GO by Maximum likelihood estimation!**
**Triangulation**

We want to locate the 3D scene point from its projections in several cameras.

The task is simple, if we know camera projection matrices $M_j \mid j = 1, \ldots, n$

**Problem formulation:** Given image points $u_j$ and camera projection matrices $M_j \mid j = 1, \ldots, n$, solve the linear homogeneous system

$$\alpha_j u_j = M_j X \mid j = 1, \ldots, n$$

**Output:** the 3D scene point $X$

Formulate the problem into an ML optimization task (here, $[\hat{u}_j, \hat{v}_j]^T$ are measures image points)

$$\min_X \sum_{j=1}^m \left[ \left( \frac{m_{j,1}X}{m_{j,3}X} - \hat{u}_j \right)^2 + \left( \frac{m_{j,2}X}{m_{j,3}X} - \hat{v}_j \right)^2 \right]$$

**Q:** Why the error factors in measured points $[\hat{u}_j, \hat{v}_j]^T$ are not used here in the formulation ML optimization function?
**Projection reconstruction and ambiguity**

Suppose there are \( m \) scene points \( X_i \mid i = 1, \ldots, m \) and \( n \) cameras \( M_j \mid j = 1, \ldots, n \)

*Given* image points \( u_{i,j} \) and camera projection matrices \( M_j \mid j = 1, \ldots, n \), solve the linear homogeneous system

\[
\alpha_{i,j} u_{i,j} = M_j X_i \mid i = 1, \ldots, m, j = 1, \ldots, n
\]

Consider the task when both scene points \( X_i \) and camera matrices \( M_j \) are both unknown

The R.H.S. contains nonlinear terms of unknowns. Thus, it's no more a linear system problem.
Projective ambiguity

Here, we identify the natural ambiguity in the system

Let $M_j$ and $X_i$ be a solution of the system and let $T$ any non-singular $3 \times 3$ matrix. Then, assuming that $M_j' = M_j T^{-1}$ and $X_i' = T X_i$,

$$M_j' X_i' = M_j T^{-1} T X_i = M_j X_i$$

i.e., $M_j'$ and $X_i'$ are also valid solutions to the same system.

So, there exists an ambiguity in the projective reconstruction.

More, specifically, the unknown true reconstruction $\{M_j, X_i\}$ and the estimated reconstruction $\{M_j', X_i'\}$ differ by a linear transformation
Matching constraints (Initial rough estimation)

- Relations satisfied by collections of corresponding image points in $n$ views.
- It is used to solve initial and not very accurate estimates of camera matrices $M_j | j = 1, ..., n$

Remember the equation

\[
\begin{bmatrix}
S(u_1)M_1 \\
\vdots \\
S(u_n)M_n
\end{bmatrix} = W
\]

\[X = WX = 0\]

- To hold the equality, $W$ must be a rank-deficient matrix
- Each row of $S(u_j)$ is a line and each leads to a plane in the scene space with the transformation $M_1 = a$ row of $W$
- Thus the determinant from any four rows of $W$ is zero, i.e., the four planes have a common intersection
**Bundle adjustment (optimum solution)**

Nonlinear optimization function (here, $[\hat{u}_j, \hat{v}_j]^T$ are measures image points)

$$
\min_{\mathbf{X}} \sum_{i=1}^{m} \sum_{j=1}^{n} \left[ \left( \frac{\mathbf{m}_{j,1} \mathbf{X}_i}{\mathbf{m}_{j,3} \mathbf{X}_i} - \hat{u}_{i,j} \right)^2 + \left( \frac{\mathbf{m}_{j,2} \mathbf{X}_i}{\mathbf{m}_{j,3} \mathbf{X}_i} - \hat{v}_{i,j} \right)^2 \right]
$$
Two cameras, stereopsis

**Objective:** Create 3D machine vision using images from two cameras – similar to the principle of human vision

**Major steps:**
- Camera calibration
- Establishing point correspondence between two pairs of points from the left and the right images
- Reconstruction of 3D coordinates of the points in 3D scene space

We will start with understanding **Epipolar geometry** and **Fundamental matrix**

**MATH:** Points and lines in $\mathcal{P}^2$
Let $u$ and $v$ be two points on a projection plane $\mathcal{P}^2$; a line $l$ passing through the two points are expressed as $l = u \times v$. Also, it may be shown that

$$l = S(u)v$$

Any point $w$ lying on the line satisfies $l^T w = 0$
Epipolar geometry and Fundamental matrix

- Optical centers
- Baseline
- Epipoles
- Epipolar plane
- Epipolar line

Epipolar constraints:
\[ l'^T u' = 0 \]
\[ l^T u = 0 \]

Fundamental matrix \((F)\): The transformation matrix relating matching points in two images.

Find the relation between fundamental matrix and camera geometry
\[ l' = e' \times u' = e' \times M'X = e' \times M'M^+u \]

\[ l' = S(e')M'M^+u = Fu, \quad \text{where,} \quad F = S(e')M'M^+ \]

Using the epipolar constraint, \( l'^T u' = 0 \Rightarrow u'^T l' = 0 \Rightarrow u'^T Fu = 0 \)

Also, \( u'^T F^T u' = 0 \)
A closer look at the Fundamental matrix

Consider Case I

\[ M = [I | 0] \]

Following the projective ambiguity, we can always find a \( T \) s.t. the first camera matrix satisfies the above form.

Now, the center \( C \) is projected at the origin, i.e.,

\[ MC = 0 \Rightarrow C = [0,0,0,1]^T \]

Assume, \( M' = [\tilde{M}' | d] \)

Then following, \( M'C = e' \), \( d \) must be equal to \( e' \)

Now, \( M^+ = M^T(MM^T)^{-1} = [I | 0]^T \)

Thus, \( F = S(e')M'M^+ = S(e')M'[I | 0] = S(e')M' = S(M'C)\tilde{M}' \)
A closer look at the Fundamental matrix

Case II
Case I ignores the affine transform between image Euclidean space (ideal image space) and image affine space (acquired image space).

Case II solves the fundamental matrix under more realistic environment

\[ M = K[I|0] \mid K: \text{intrinsic callib. matrix} \]

The second camera matrix may be expressed in the form

\[ M' = K'[R| - Rt] \quad \text{Explain!} \]

As in Case I,

\[ MC = 0 \Rightarrow C = [0,0,0,1]^T \]

Now, \[ M^+ = \begin{bmatrix} K^{-1} \\ 0^T \end{bmatrix} \]

Thus,

\[ F = S(e')M'M^+ = S(M'C)K'[R - Rt] \begin{bmatrix} K^{-1} \\ 0^T \end{bmatrix} = S(-K'Rt)K'RK^{-1} \]

Using \[ S(Hu) = H^{-1T}S(u)H^{-1}, F = K'^{-T}RS(t)K^{-1} \]