Multi-view 3D Reconstruction for Dummies

Jianxiong Xiao

Camera projection
- When people take a picture of a point:
  \[ x = K[R|t]X \]

Camera projection
- When people take two pictures with same camera setting:
  \[ x_1 = K[R_1|t_1]X \]
  \[ x_2 = K[R_2|t_2]X \]

SFMedu Program with Code
Download from:
http://mit.edu/jxiao/Public/software/SFMedu/
Camera projection

• When people take three pictures with same camera setting:

\[ x_1 = K[R_1|t_1]X \]
\[ x_2 = K[R_2|t_2]X \]
\[ x_3 = K[R_3|t_3]X \]

Same Camera Same Setting = Same \( K \)
Triangulation

Structure From Motion

- Structure = 3D Point Cloud of the Scene
- Motion = Camera Location and Orientation
- SFM = Get the Point Cloud from Moving Cameras
- Structure and Motion: Joint Problems to Solve

Pipeline

- Structure from Motion (SFM)
- Multi-view Stereo (MVS)
Two-view Reconstruction

Keypoints Detection
Descriptor for each point

Point Match for correspondences

Same for the other images

Point Match for correspondences
Fundamental Matrix

Given a correspondence \( x_1 \leftrightarrow x_2 \)

\[ x_1^T F x_2 = 0 \]

Estimating Fundamental Matrix

- Given a correspondence \( x_1 \leftrightarrow x_2 \)
- The basic incidence relation is

\[
\begin{bmatrix}
f_{11} \\
f_{12} \\
f_{13} \\
f_{21} \\
f_{22} \\
f_{23} \\
f_{31} \\
f_{33}
\end{bmatrix} = 0
\]

Need 8 points

RANSAC to Estimate Fundamental Matrix

- For many times
  - Pick 8 points
  - Compute a solution for \( F \) using these 8 points
  - Count number of inliers
- Pick the one with the largest number of inliers

Direct Linear Transformation (DLT)
Result 9.19. For a given essential matrix 
\[ E = U \text{diag}(1, 1, 0) V^T, \]
and the first camera matrix \( P_1 = [I | 0] \), there are four possible choices for the second camera matrix \( P_2 \): 

\[
\begin{align*}
P_2 &= [UWV^T + u_3] \\
P_2 &= [UWV^T - u_3] \\
P_2 &= [UW^T V^T + u_3] \\
P_2 &= [UW^T V^T - u_3]
\end{align*}
\]

\[ W = \begin{bmatrix} 0 & -1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix} \]

Fig. 9.12. The four possible solutions for calibrated reconstruction from \( E \). Between the left and right sides there is a baseline reversal. Between the top and bottom rows camera \( B \) rotates 180° about the baseline. Note, only in (a) is the reconstructed point in front of both cameras.
In front of the camera?

- Camera Extrinsic $[R|t]$
  \[
  \begin{bmatrix}
  X_{\text{world}} \\
  Y_{\text{world}} \\
  Z_{\text{world}}
  \end{bmatrix} = \begin{bmatrix}
  X_{\text{cam}} \\
  Y_{\text{cam}} \\
  Z_{\text{cam}}
  \end{bmatrix} + t = R \begin{bmatrix}
  X_{\text{cam}} \\
  Y_{\text{cam}} \\
  Z_{\text{cam}}
  \end{bmatrix} - R^t t
  \]

- Camera Center
  \[
  \begin{bmatrix}
  X_{\text{world}} \\
  Y_{\text{world}} \\
  Z_{\text{world}}
  \end{bmatrix} = \begin{bmatrix}
  X_{\text{cam}} \\
  Y_{\text{cam}} \\
  Z_{\text{cam}}
  \end{bmatrix} = R \begin{bmatrix}
  0 \\
  0 \\
  0
  \end{bmatrix} - R^t t = -R^t t
  \]

- View Direction
  \[
  \begin{bmatrix}
  0 \\
  0 \\
  1
  \end{bmatrix} \rightarrow \begin{bmatrix}
  R^t \\
  -R^t t
  \end{bmatrix} - (C) = (R(3,:)^t - R^t t) - (-R^t t) = R(3,:)^t
  \]

In front of the camera?

- A point $X$
- Direction from camera center to point $X - C$
- Angle Between Two Vectors
- Angle Between $X - C$ and View Direction
- Just need to test
  \[
  (X - C) \cdot R(3,:)^t > 0?
  \]

Pick the Solution

With maximal number of points in front of both cameras.

Two-view Reconstruction

* Fig. 6.12. The four possible solutions for a cinedward reconstruction from $\Sigma$. Between the left and right images there is a baseline normal. Because the top and bottom row camera $B$ rotates $180^\circ$ about the baseline. Note, only one of the reconstructed points is from each camera.
Multi-view Stereo (MVS)
Structure from Motion (SFM)
Pipeline

There can be only one $[R_t|t]$
Merge Two Point Cloud

• From the 1st and 2nd images, we have 
  $[R_1|t_1]$ and $[R_2|t_2]$

• From the 2nd and 3rd images, we have 
  $[R_2|t_2]$ and $[R_3|t_3]$

• Exercise: How to transform the coordinate system of the second point cloud to align with the first point cloud so that there is only one $[R_2|t_2]$?

Oops

Bundle Adjustment

See From a Different Angle
Rethinking the SFM problem

- **Input:** Observed 2D image position
  \[
  \begin{bmatrix}
  \tilde{x}_1^1 & \tilde{x}_2^1 \\
  \tilde{x}_1^2 & \tilde{x}_2^2 \\
  \tilde{x}_1^3 & \tilde{x}_2^3 \\
  \end{bmatrix}
  \]

- **Output:**
  - Unknown Camera Parameters (with some guess)
    \[
    \begin{bmatrix}
    R_1 | t_1 \\
    R_2 | t_2 \\
    R_3 | t_3 \\
    \end{bmatrix}
    \]
  - Unknown Point 3D coordinate (with some guess)
    \[
    X^1, X^2, X^3, \ldots
    \]

**Camera projection**

- **Image 1**
  \[
  x_1^1 = K[R_1 | t_1] X^1
  \]

- **Image 2**
  \[
  x_1^2 = K[R_2 | t_2] X^1
  \]

- **Image 3**
  \[
  x_1^3 = K[R_3 | t_3] X^1
  \]

Same Camera Same Setting = Same \( K \)

---

**Bundle Adjustment**

A valid solution \( [R_1 | t_1] [R_2 | t_2] [R_3 | t_3] \) and \( X^1, X^2, X^3, \ldots \) must let

- **Re-projection**
  \[
  \begin{bmatrix}
  x_1^1 = K[R_1 | t_1] X^1 \\
  x_1^2 = K[R_2 | t_2] X^1 \\
  x_1^3 = K[R_3 | t_3] X^1
  \end{bmatrix}
  \]

- **Observation**
  \[
  \begin{bmatrix}
  \tilde{x}_1^1 & \tilde{x}_2^1 \\
  \tilde{x}_1^2 & \tilde{x}_2^2 \\
  \tilde{x}_1^3 & \tilde{x}_2^3 \\
  \end{bmatrix}
  \]
Bundle Adjustment

A valid solution \([R_1|t_1][R_2|t_2][R_3|t_3]\) and \(X^1, X^2, X^3, \ldots\)

must let the Re-projection close to the Observation, i.e. to minimize the reprojection error

\[
\min \sum \sum \left( \tilde{x}_i - K [R_i|t_i] X^j \right)^2
\]

Solving This Optimization Problem

- Theory:
  The Levenberg–Marquardt algorithm

- Practice:
  The Ceres-Solver from Google