Lecture 28:
Structure from Motion and Visual SLAM
• Structure from Motion
• Correspondence
• Optimization
• Visual SLAM
• 4D Reconstruction
• Dense Surface Reconstruction
Photo Tourism


http://phototour.cs.washington.edu/
3D Models from Community Databases

E.g., Google image search on “Dubrovnik”
3D Models from Community Databases

5K images, 3.5M points, >10M factors

Movie by Aggarwal et al.
Building Rome in a Day

Sameer Agarwal, Noah Snavely, Ian Simon, Steven M. Seitz and Richard Szeliski

International Conference on Computer Vision, 2009, Kyoto, Japan.

http://grail.cs.washington.edu/rome/
Outline

• Structure from Motion
  • Correspondence
• Optimization
• Visual SLAM
• 4D Reconstruction
• Dense Surface Reconstruction
A Correspondence Problem
Feature detection

• Detect features using SIFT [Lowe, IJCV 2004]
Feature matching

Refine matching using RANSAC [Fischler & Bolles 1987] to estimate fundamental matrices between pairs
Feature Matching!
Bad News: Good correspondences are hard to find

- Good news: Geometry constrains possible correspondences.
  - 4 DOF between $x$ and $x'$; only 3 DOF in $X$.
  - Constraint is manifest in the **Fundamental matrix**
    
    $x'^T F x = 0$.

- $F$ can be calculated either from camera matrices or a set of good correspondences.
Fundamental Matrix
Fundamental Matrix $F$

- $3 \times 3 = 9$ DOF
- However, scale, rank 2!
- $\Rightarrow 7$ DOF
Epipoles inside the image: zoom-like setup.

Epipoles are where the other camera is!
Epipoles in near-stereo config.

Epipoles are where the other camera is!
<table>
<thead>
<tr>
<th>Trinocular Camera rig</th>
</tr>
</thead>
<tbody>
<tr>
<td><img src="https://www.skydio.com/" alt="Skydio 2 Image" /></td>
</tr>
</tbody>
</table>

## Navigation Camera System

<table>
<thead>
<tr>
<th>Configuration</th>
<th>6x cameras in trinocular configuration top and bottom</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sensor Type</td>
<td>Sony 1/3&quot; 4K color CMOS</td>
</tr>
<tr>
<td>Lens Aperture</td>
<td>f/1.8</td>
</tr>
<tr>
<td>Field-of-view</td>
<td>200°</td>
</tr>
<tr>
<td>Environment Coverage</td>
<td>True 360°</td>
</tr>
</tbody>
</table>
Trifocal Geometry

\[ [x']_\times (\sum_i x^iT_i)[x''']_\times = 0_{3 \times 3} \]
Outline

- Structure from Motion
- Correspondence
- Optimization
- Visual SLAM
- 4D Reconstruction
- Dense Surface Reconstruction
An Optimization Problem

• Find the **most likely** structure and motion $\Theta$
Optimization

= Non-linear Least-Squares!

\[ \sum_{i=1}^{m} \sum_{k=1}^{K_i} \| u_{ik} - h(m_i, x_{j_{ik}}) \|^2 \]

\[ m_i \quad \text{Image } i \quad \text{Image } i' \quad m_i' \]
Sparse nonlinear least squares

• Simple 1-Dimensional Example
• $p = 2$ cameras and 4 points: \{c_1 \ c_2 \ l_1 \ l_2 \ l_3 \ l_4\}
• $f(u_{ik}, p) = \text{difference in x position} = l_{j(ik)} - c_i$
Model with Factor Graphs

- Connectivity = sparsity!
- Factor is function of small set.
Sparse Jacobian and Hessian

\[ A' \cdot A = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ -1 & 0 & 1 & 0 & 0 & 0 \\ -1 & 0 & 0 & 1 & 0 & 0 \\ -1 & 0 & 0 & 0 & 1 & 0 \\ 0 & -1 & 1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 1 & 0 & 0 \\ 0 & -1 & 0 & 0 & 1 & 0 \\ 0 & -1 & 0 & 0 & 0 & 1 \end{bmatrix} \]

\[ b = \begin{bmatrix} 5 \\ -5 \\ 5 \\ 10 \\ -15 \\ -5 \\ 0 \\ 5 \end{bmatrix} \]

\[ (A' \cdot A) \backslash A' \cdot b = \begin{bmatrix} 5.0000 \\ 15.0000 \\ 0.0000 \\ 10.0000 \\ 15.0000 \\ 20.0000 \end{bmatrix} \]
Structure from Motion (Chicago, movie by Yong Dian Jian)

180 cameras, 88723 points
458642 projections
active camera: 4
• Structure from Motion
• Correspondence
• Optimization
• Visual SLAM
• 4D Reconstruction
• Dense Surface Reconstruction
Visual SLAM: SfM for Robots
Visual SLAM Factor Graph

\[ P(X, M) = k^* \prod_{i=1}^{M} P(x_i | x_{i-1}, u_i) \times \prod_{k=1}^{K} P(z_k | x_{i_k}, l_{i_k}) \]

- Trajectory of Robot
- Landmark Measurements
- “Landmarks”
Visual SLAM Factor Graph
End result: Solution + Uncertainty
Example: Underwater SLAM

9831 camera poses, 185261 landmarks, and 350988 factors
Outline

• Structure from Motion
• Correspondence
• Optimization
• Visual SLAM

• 4D Reconstruction
• Dense Surface Reconstruction
Spatiotemporal Reconstruction

Historical Image Collection

Supported by NSF CAREER, Microsoft
Recent revival: NSF NRI award on 4D crops for precision agriculture...

Grant Schindler
4D Reconstruction of Lower Manhattan

4D Structure over Time
4D crop monitoring (Jing Dong)
Results: video (by Jing Dong)

4D reconstruction results (by PMVS) and its cross section
• Structure from Motion
• Correspondence
• Optimization
• Visual SLAM
• 4D Reconstruction
  • Dense Surface Reconstruction
Multi-view Stereo

Multi-View Stereo for Community Photo Collections
Michael Goesele, Noah Snavely, Brian Curless, Hugues Hoppe, and Steven M. Seitz
ICCV 2007
Multi-view Stereo

• Poisson Surface Reconstruction

Compared with Laser-Scanner
Neural Radiance Fields (NeRF)

- Original NeRF paper:
  - [https://www.matthewtancik.com/nerf](https://www.matthewtancik.com/nerf)
- See two blog posts:
  - [https://dellaert.github.io/NeRF/](https://dellaert.github.io/NeRF/)
  - [https://dellaert.github.io/NeRF21/](https://dellaert.github.io/NeRF21/)