

Classical Structure from Motion Pipelines

D.A. Forsyth

University of Illinois at Urbana-Champaign

Recall cases:

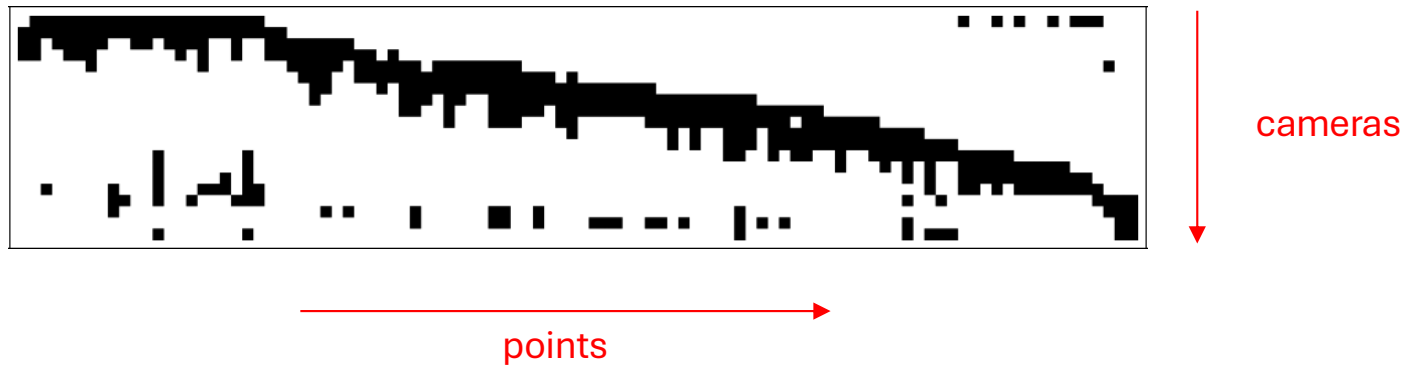
- Offline
 - work with all images
 - structure from motion
 - underlying Q:
 - how to make best use of all information
 - roughly familiar structure
 - set up huge optimization problem by constructing start point
 - efficiency is now a serious issue
- Online
 - build best estimate available up to time t
 - SLAM
 - simultaneous localization and mapping
 - build a map at the same time as determining where you are
 - new Q:
 - how do you make best use of all information up to now
 - efficiency often crucial
- Hybrids are quite usual, and important

3 or more calibrated cameras

- Basically, use recipe above BUT
 - In what order should you insert cameras?
 - matters – recipe requires shared points
 - we don't know which pictures share many points
 - some might share no points
 - Optimize everything
 - bundle adjustment

What can you see?

- Build a table of what point is visible in what camera
 - black is visible
 - notice picture is in convenient permutation!



- Issue:
 - strategy above works only if cameras share points
 - many (most) pairs don't

Representative SFM pipeline

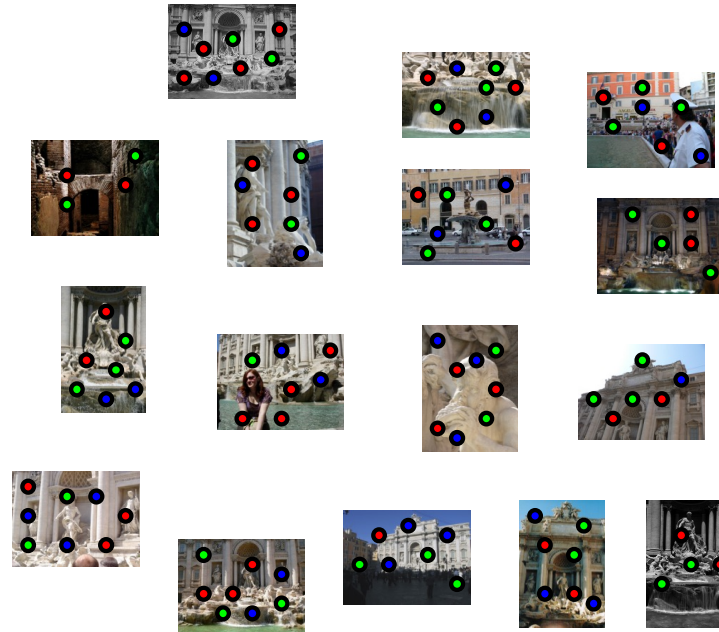


N. Snavely, S. Seitz, and R. Szeliski. [Photo tourism: Exploring photo collections in 3D](http://phototour.cs.washington.edu/). SIGGRAPH 2006
<http://phototour.cs.washington.edu/>

Feature detection

Detect SIFT features

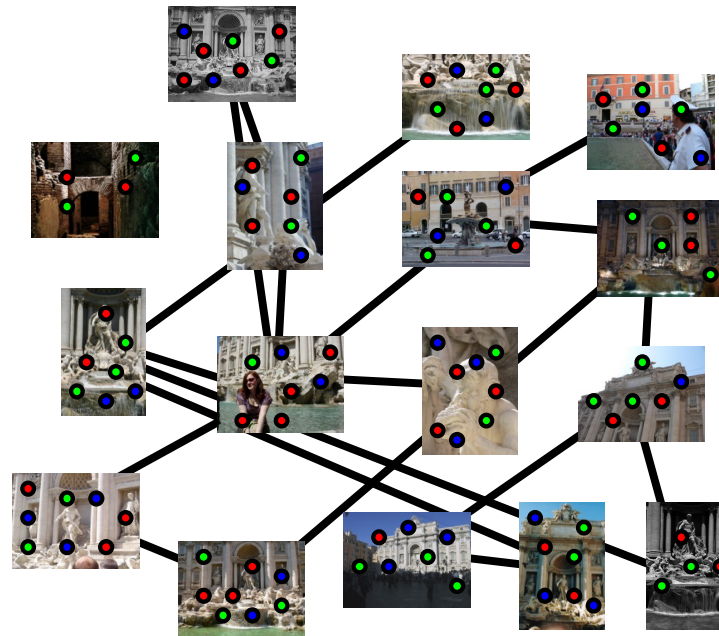
Other popular feature types: [SURF](#), [ORB](#), [BRISK](#), ...



Source: N. Snavely

Feature matching

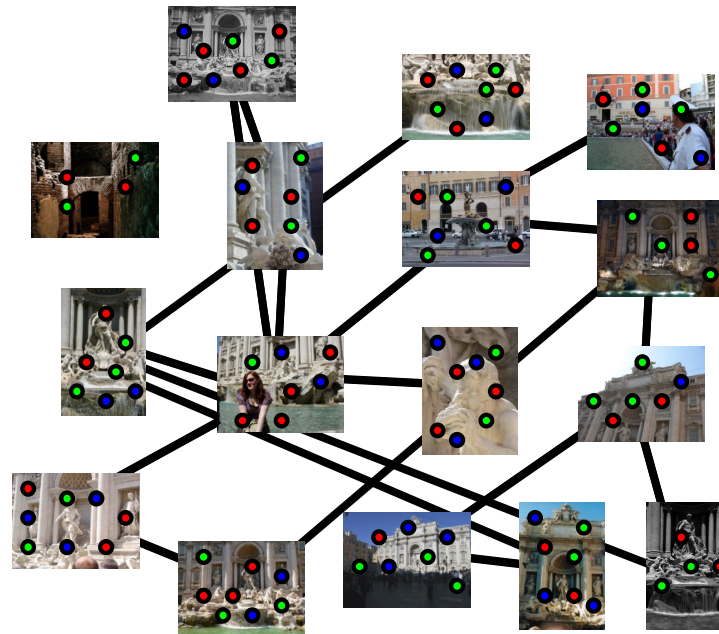
Match features between each pair of images



Source: N. Snavely

Fundamental matrix

Use RANSAC to estimate fundamental matrix between each pair



Source: N. Snavely

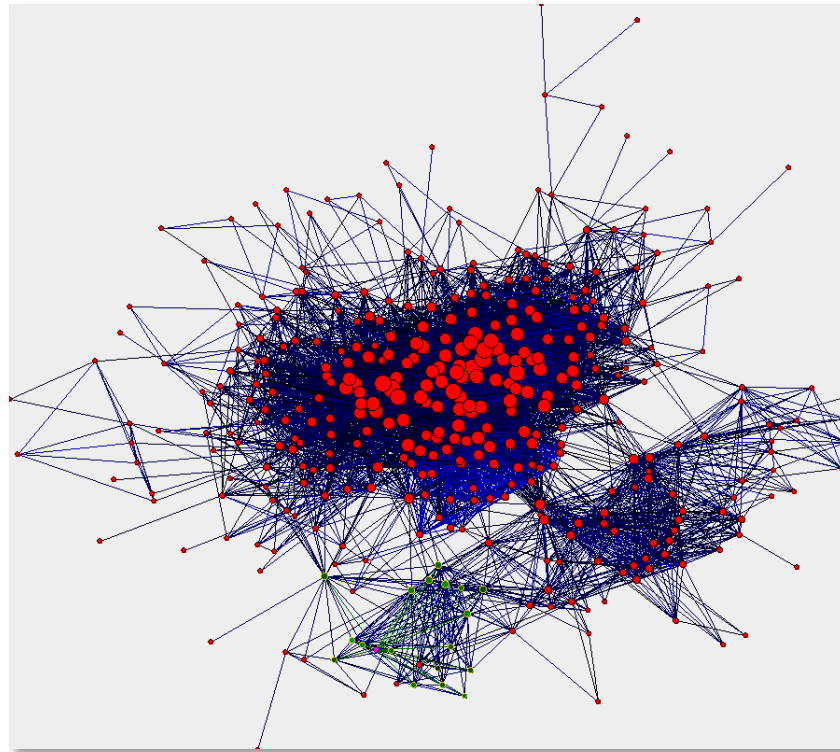
Fundamental matrix

Use RANSAC to estimate fundamental matrix between each pair



[Image source](#)

Image connectivity graph

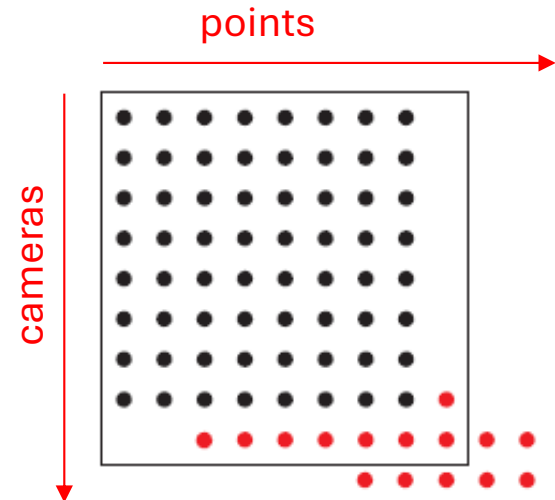


(graph layout produced using the Graphviz toolkit: <http://www.graphviz.org/>)

Source: N. Snavely

Incremental structure from motion

- Initialize motion from two images using fundamental matrix
- Initialize structure by triangulation
- For each additional view:
 - Determine projection matrix of new camera using all the known 3D points that are visible in its image – **calibration**
 - Refine and extend structure: compute new 3D points, re-optimize existing points that are also seen by this camera – **triangulation**
- Refine structure and motion: bundle adjustment



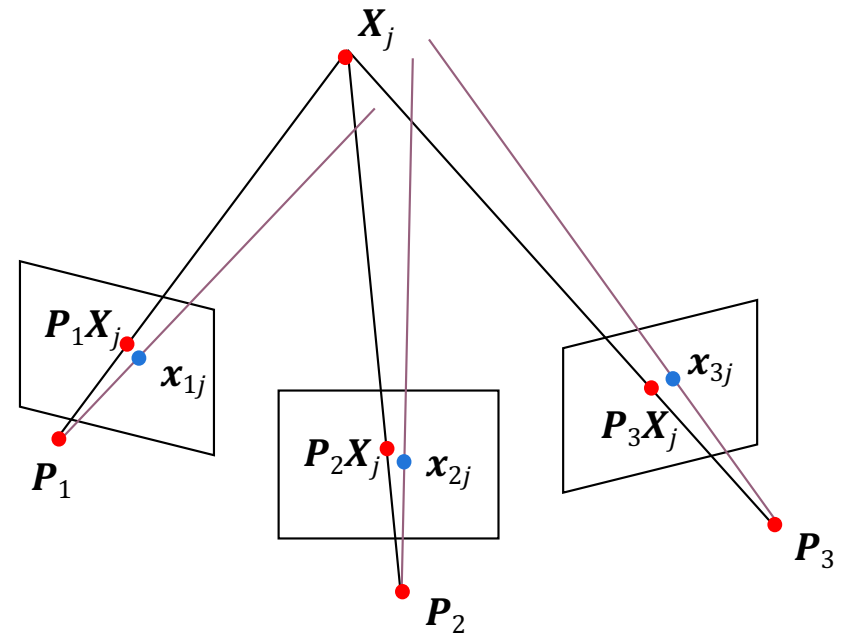
Bundle adjustment

- Non-linear method for refining structure and motion
- Minimize reprojection error (with lots of bells and whistles):

$$\sum_{i=1}^m \sum_{j=1}^n w_{ij} d \left(\mathbf{x}_{ij} - \text{proj}(\mathbf{P}_i \mathbf{X}_j) \right)^2$$

This might be a robust loss

visibility flag: is point j visible in view i ?



SFM software

- [Bundler](#)
- [OpenSfM](#)
- [OpenMVG](#)
- [VisualSFM](#)
- [COLMAP](#)
- See also [Wikipedia's list of toolboxes](#)