Structure from Motion

Computer Vision I

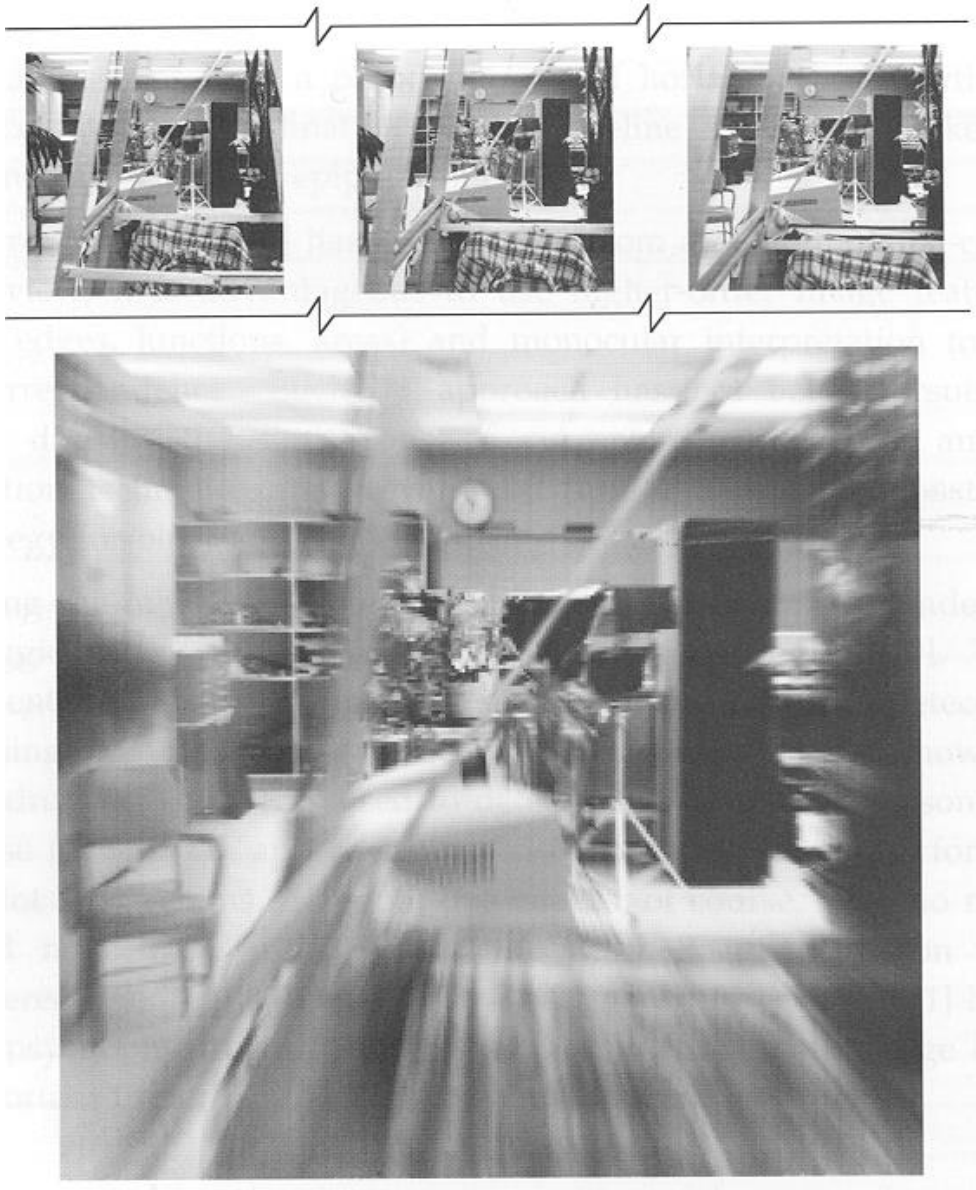
CSE 252A

Lecture 11

Announcements

- Homework 3 is due Nov 5, 11:59 PM
- Homework 4 will be assigned on Nov 5
- Reading:
  - Chapter 8: Structure from Motion
  - Optional: Multiple View Geometry in Computer Vision, 2nd edition, Hartley and Zisserman

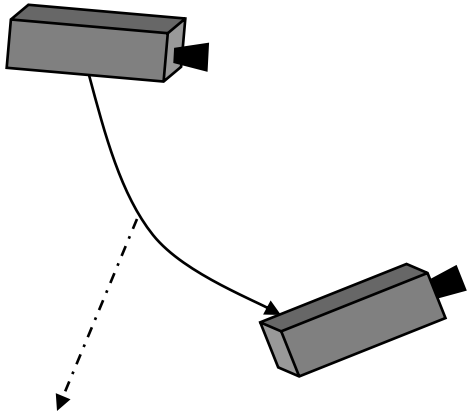
Motion



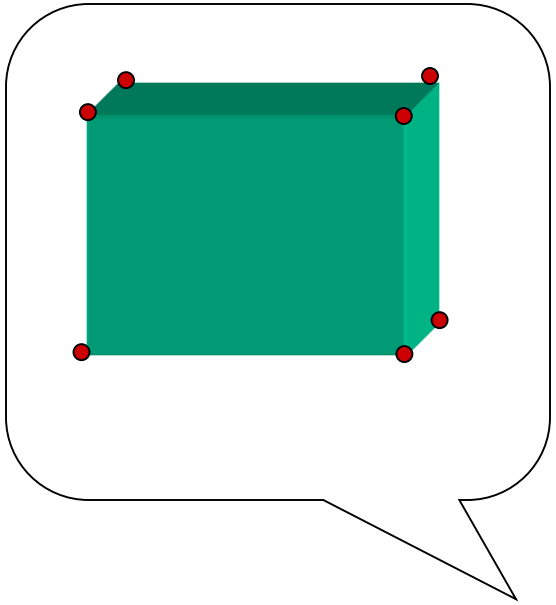
Motion

- Some problems of motion
  - Correspondence: Where have elements of the image moved between image frames?
  - Reconstruction: Given correspondences, what is 3D geometry of scene?
  - Motion segmentation: What are regions of image corresponding to different moving objects?
  - Tracking: Where have objects moved in the image? (related to correspondence and segmentation)

Structure from Motion



The change in rotation and translation between the two cameras (the “motion”)



Locations of points on the object (the “structure”)

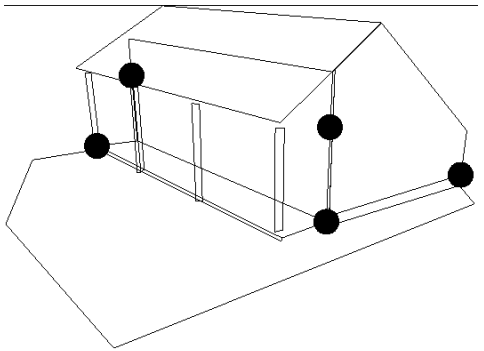
MOVING CAMERAS ARE LIKE STEREO

Structure from Motion (SFM)

- Objective
  - Given two or more images (or video frames), without knowledge of the camera poses (rotations and translations), estimate the camera poses and 3D structure of scene.
- Considerations
  - Discrete motion (wide baseline) vs. continuous (infinitesimal) motion
  - Calibrated vs. uncalibrated
  - Two views vs. multiple views
  - Orthographic (affine) vs. perspective

Discrete Motion, Calibrated

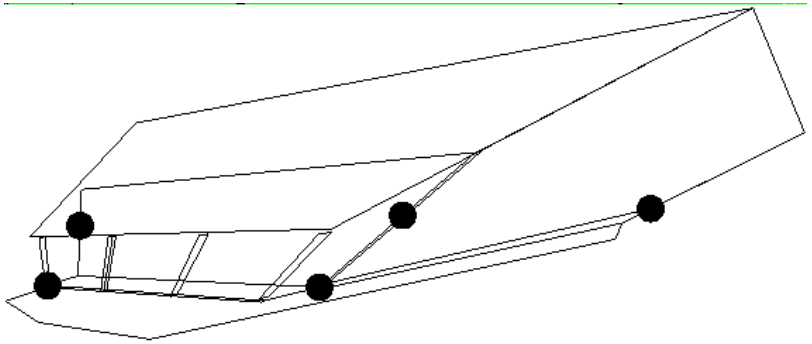
- Consider m images of n points, how many unknowns?
  - Unknowns
    - 3D Structure: $3n$
    - First normalized camera $\hat{P} = [I \mid 0]$
      - Rotations: $3(m-1)$
      - Translations (to scale): $3(m-1)-1$
    - Total: $3n + 6(m-1) - 1$
  - Measurements
    - $2nm$



- Solution when $3n + 6(m-1) - 1 \leq 2nm$

Discrete Motion, Uncalibrated

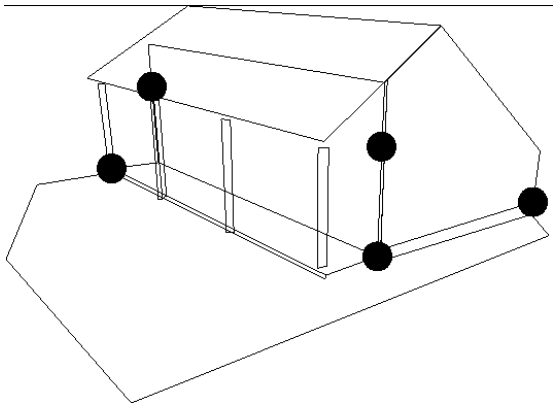
- Consider m images of n points, how many unknowns?
  - Unknowns
    - 3D Structure: $3n$
    - Cameras (to 3D projective transformation): $11m - 15$
    - Total: $3n + 11m - 15$
  - Measurements
    - $2nm$



- Solution when $3n + 11m - 15 \leq 2nm$

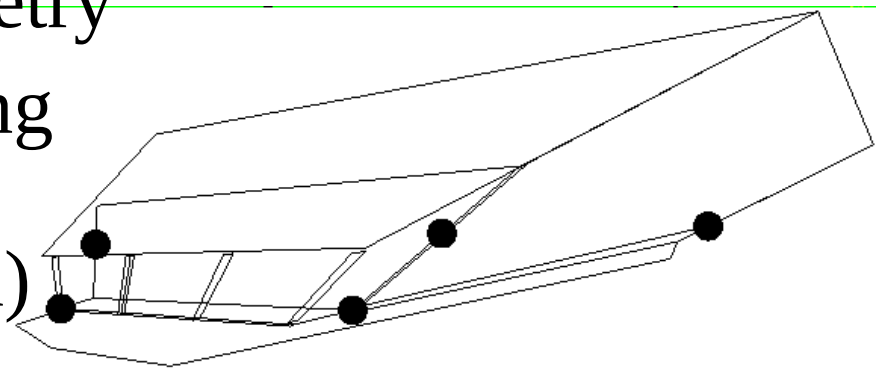
Two Views, Calibrated

- Input: Two images (or video frames)
- Detect feature points
- Determine feature correspondences
- Compute the essential matrix
- Retrieve the relative camera rotation and translation (to scale) from the essential matrix
- Optional: Perform dense stereo matching using recovered epipolar geometry
- Reconstruct corresponding 3D scene points (to scale)



Two Views, Uncalibrated

- Input: Two images (or video frames)
- Detect feature points
- Determine feature correspondences
- Compute the fundamental matrix
- Retrieve the relative camera 3D projective transformation from the fundamental matrix
- Optional: Perform dense stereo matching using recovered epipolar geometry
- Reconstruct corresponding 3D scene points (to 3D projective transformation)



Essential Matrix (Calibrated)

- Number of point correspondences and solutions
  - 5 point correspondences, up to 10 (real) solutions
  - 6 point correspondences, 1 solution
  - 7 point correspondences, 1 or 3 real solutions (and 2 or 0 complex ones)
  - 8 or more point correspondences, 1 solution

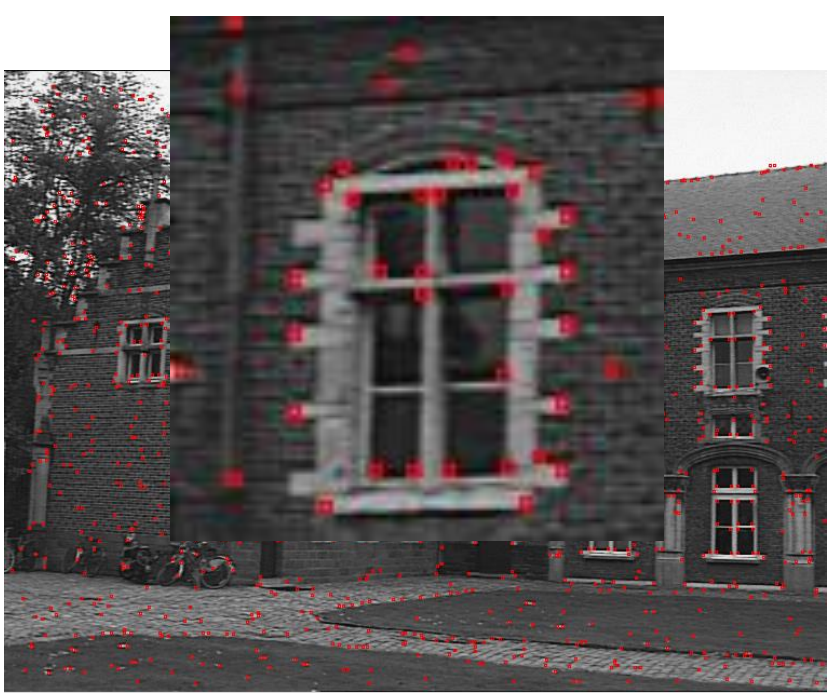
Fundamental Matrix (Uncalibrated)

- Number of point correspondences and solutions
  - 7 point correspondences, 1 or 3 real solutions (and 2 or 0 complex ones)
  - 8 or more point correspondences, 1 solution

Mathematics

- Essential matrix
  - Linear estimation (8 or more correspondences)
  - Retrieval of normalized camera projection matrices and 3D rotation and translation (to scale)
- Fundamental matrix
  - Linear estimation (8 or more correspondences)
  - Retrieval of camera projection matrices and 3D projective transformation

Feature detection

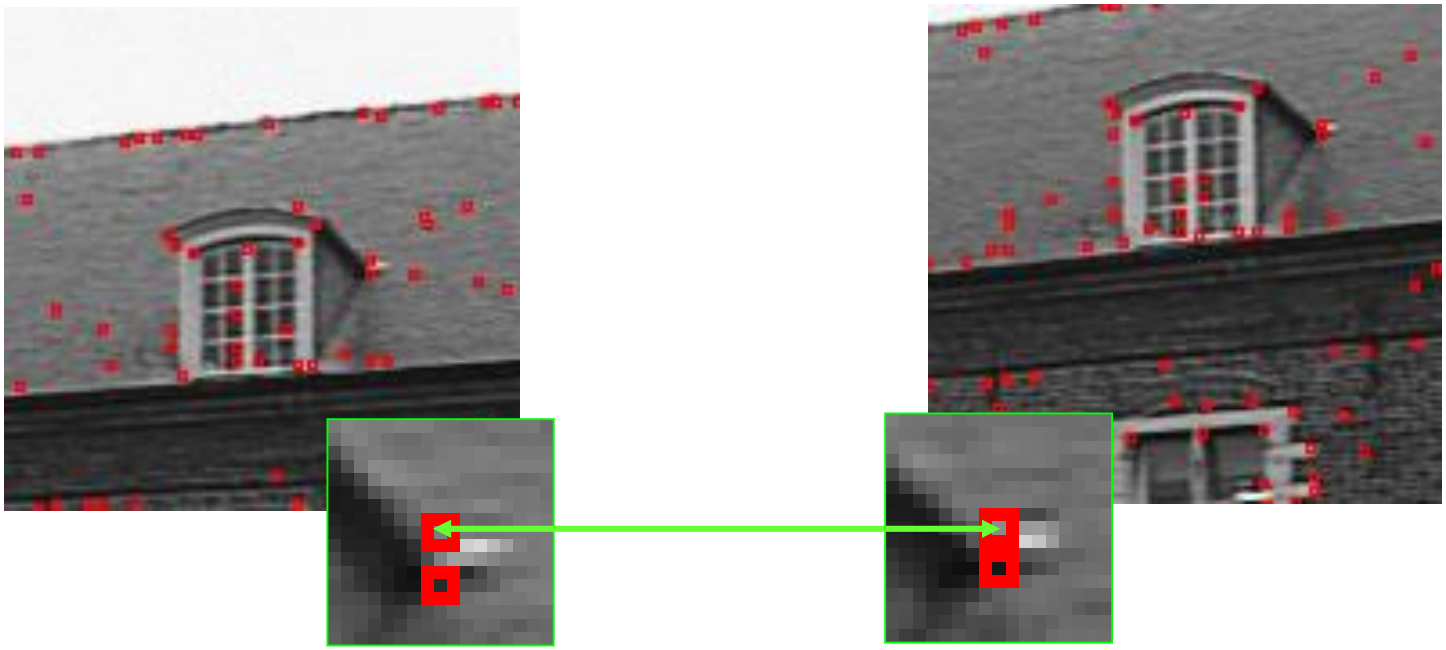


Select strongest features **(e.g. 1000/image)**

Feature matching

Evaluate normalized cross correlation (or sum of squared differences) for all features with similar coordinates

$$\text{e.g.}\ (x', y') \in \left[x - \tfrac{w}{10}, x + \tfrac{w}{10}\right] \times \left[y - \tfrac{h}{10}, y + \tfrac{h}{10}\right]$$



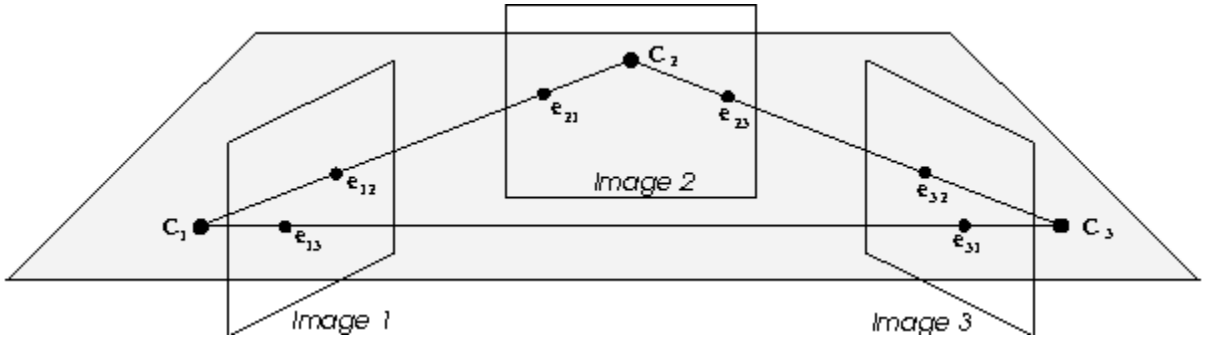
Keep mutual best matches

Still many wrong matches!

Comments

- Greedy Algorithm:
  - Given feature in one image, find best match in second image irrespective of other matches
  - Suitable for small motions, little rotation, small search window
- Otherwise
  - Must compare descriptor over rotation
  - Cannot consider all potential pairings (way too many), so
    - Manual correspondence (e.g., photogrammetry)
    - Use robust outlier rejection (e.g., RANSAC)
    - More descriptive features (line segments, SIFT, larger regions, color)
    - Use video sequence to track, but perform SFM w/ first and last image

Three Views



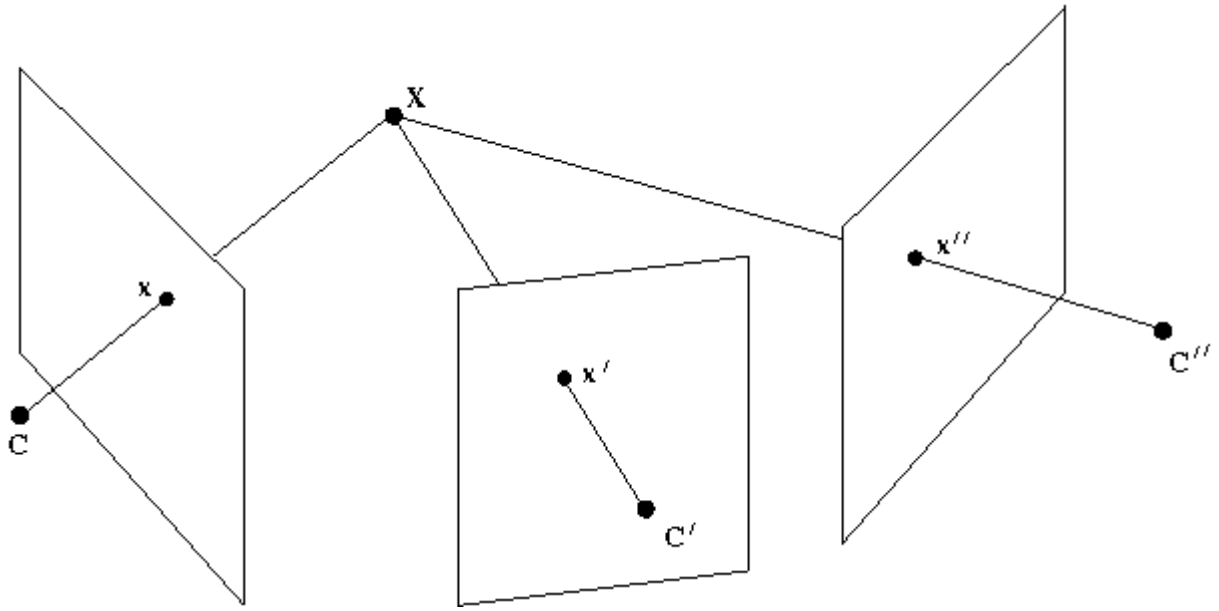
Trifocal plane

Trifocal Tensor

- 3x3x3 tensor
- 27 elements, 18 degrees of freedom
  - 33 degrees of freedom (3 camera projection matrices) minus 15 degrees of freedom (3D projective transformation)
- Uses tensor notation
  - Einstein summation
- Retrieve camera projection matrices

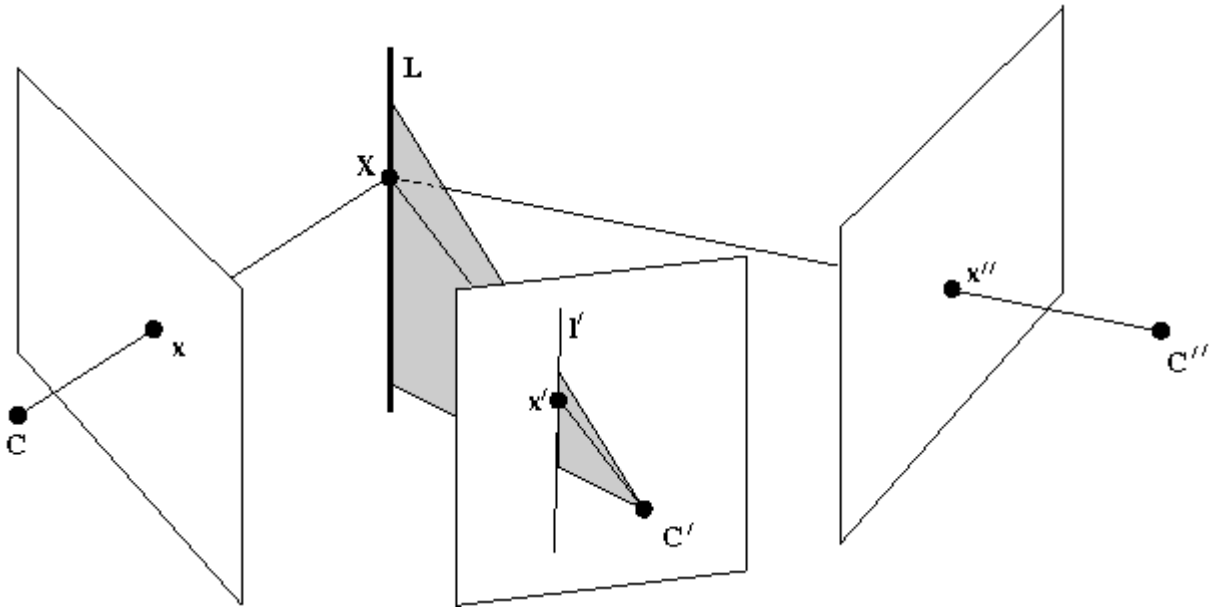
3 Points

- Point-Point-Point



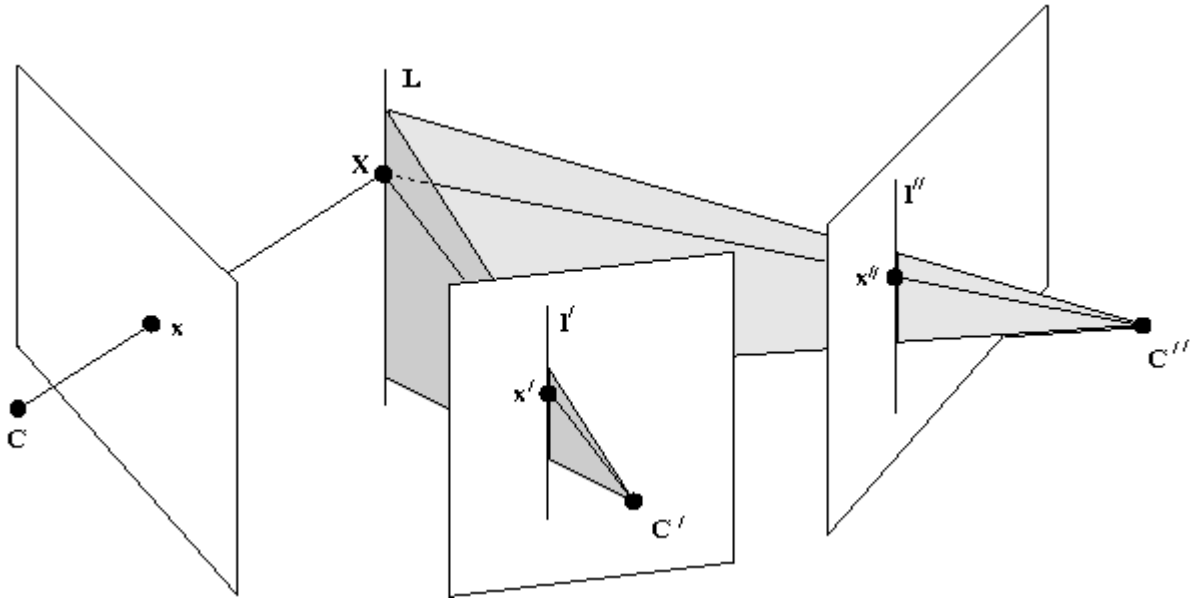
2 Points, 1 Line

- Point-Line-Point
  - Note: image line must pass through corresponding image point



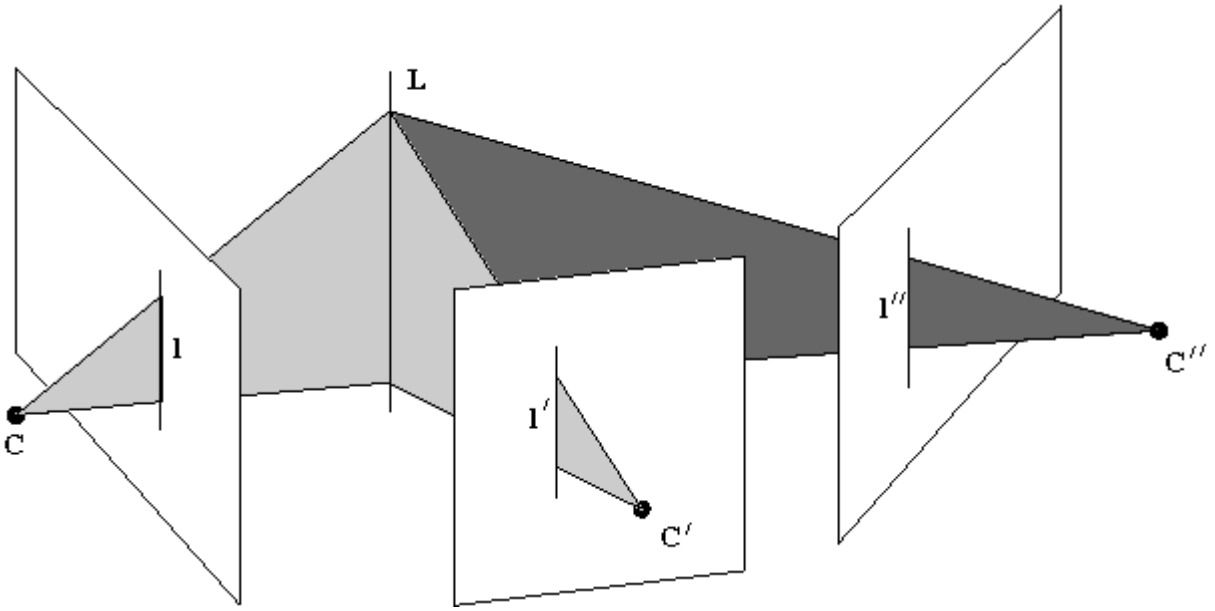
1 Point, 2 Lines

- Point-Line-Line
  - Note: image lines do not need to correspond, but must pass through corresponding image points



3 Lines

- Line-Line-Line



N-View Geometry

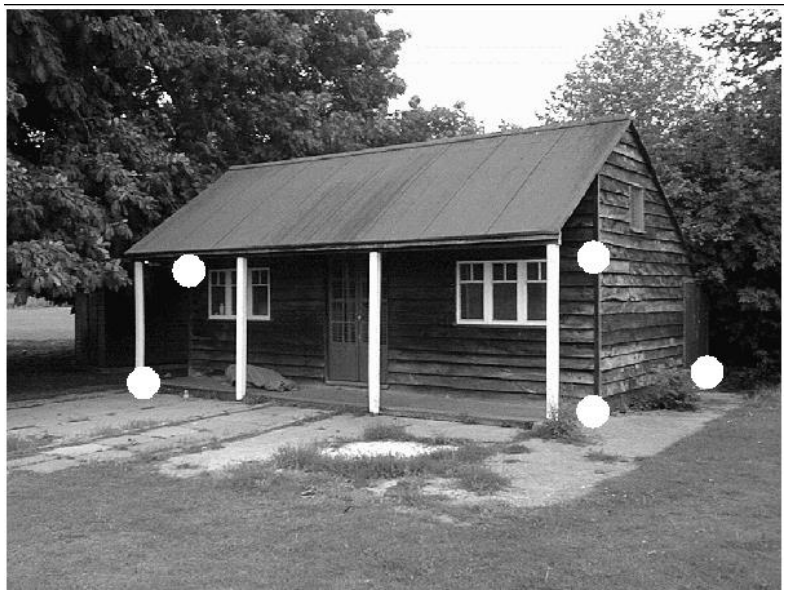
- Reconstruction
  - Bundle adjustment
    - Simultaneous adjustment of parameters for all cameras and all 3D scene points
    - Minimize reprojection error in all images

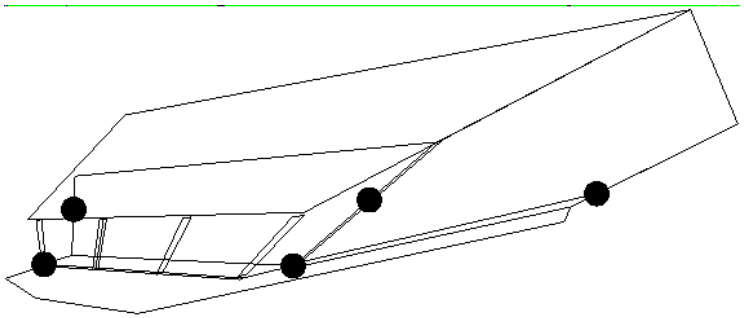
      $$\min_{\hat{\mathrm{P}}^i, \hat{\mathbf{X}}_j} \sum_{ij} d(\hat{\mathrm{P}}^i \hat{\mathbf{X}}_j, \mathbf{x}_j^i)^2$$

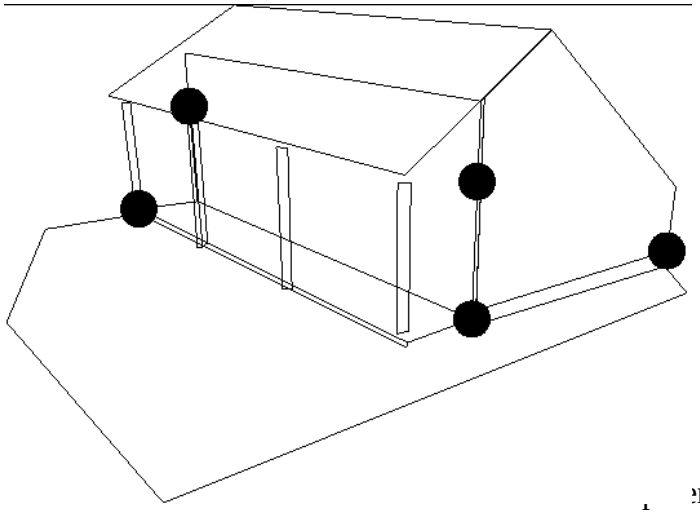
    - Reconstruction of cameras and 3D scene points to similarity (calibrated) or projective (uncalibrated) ambiguity
  - Factorization (see text)

Direct Reconstruction
Projective to Euclidean

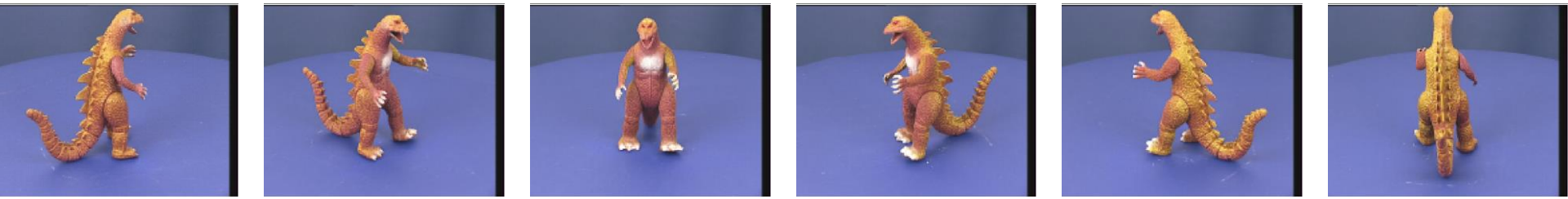
5 (or more) points correspondences

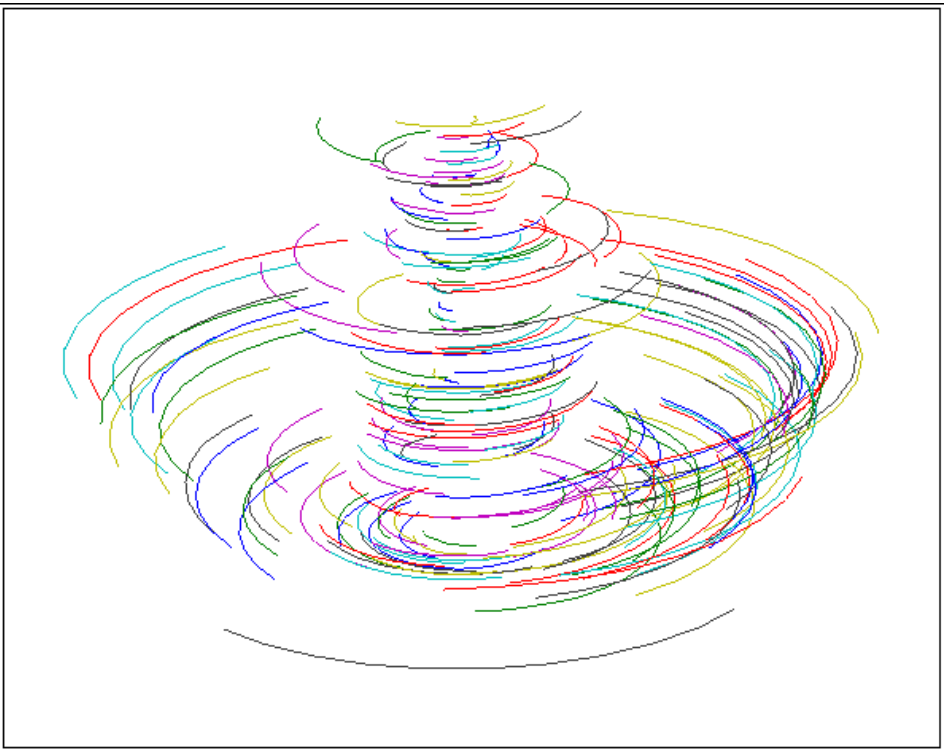


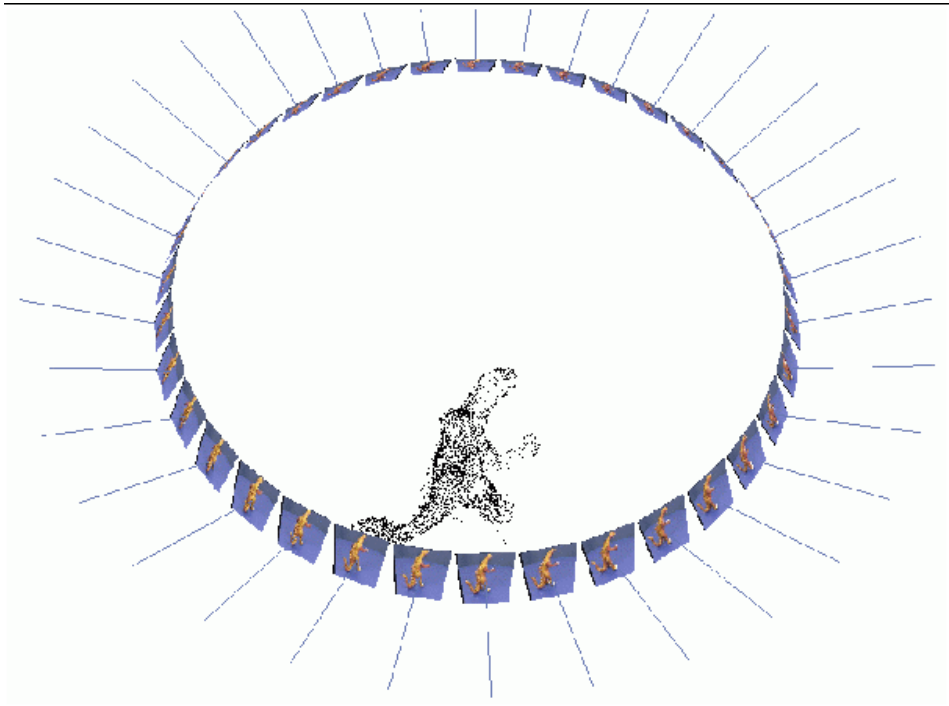




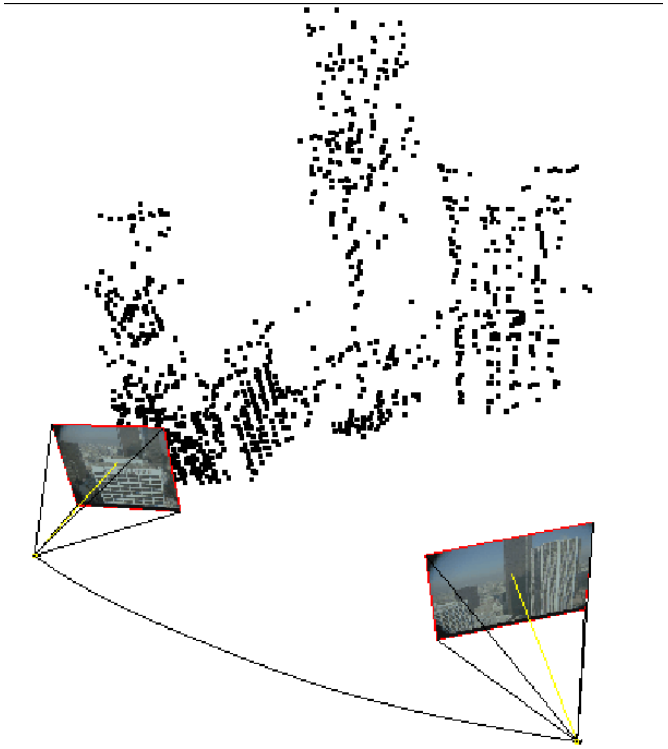
N-View Geometry





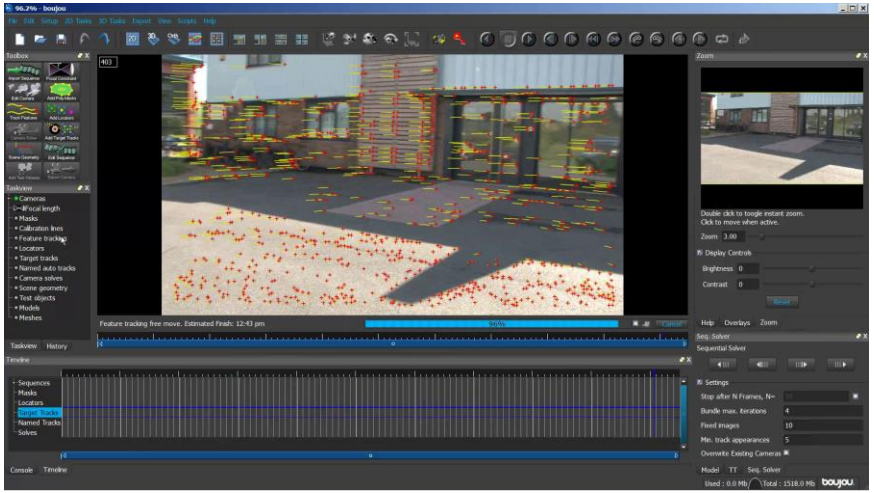


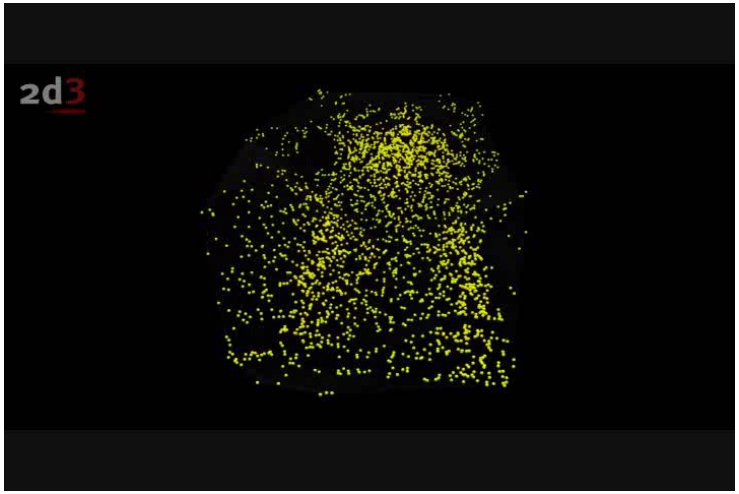
N-View Geometry



N-View Geometry

- Example results





Next Lecture

- Mid-level vision
  - Grouping and model fitting
- Reading:
  - Chapter 10: Grouping and Model Fitting