Large-scale 3D Modeling from Crowdsourced Data

Jan-Michael Frahm, UNC

Enrique Dunn, SIT

Marc Pollefeys, ETHZ, Microsoft

Jared Heinly, URCV

Johannes Schönberger, ETHZ



Sparse Modeling

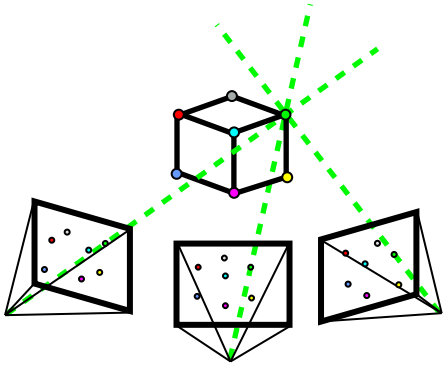
Johannes L. Schönberger

ETH zürich

Computer Vision and Geometry Lab

Outline

1. Structure-from-Motion algorithm



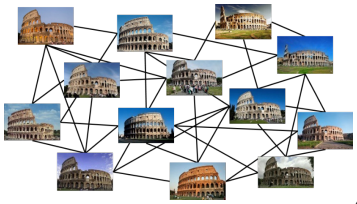
2. Challenges in crowdsourced data

Pipeline

Unstructured Images

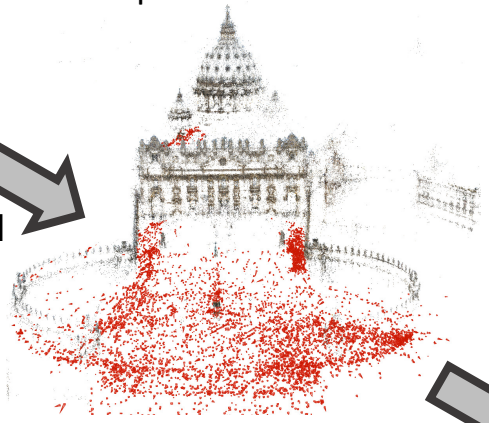
Assoc.

Scene Graph



SFM

Sparse Model



MVS

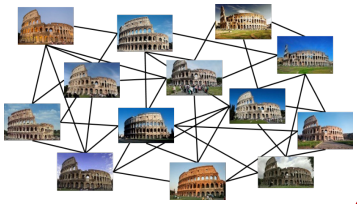
Dense Model

Pipeline

Unstructured Images

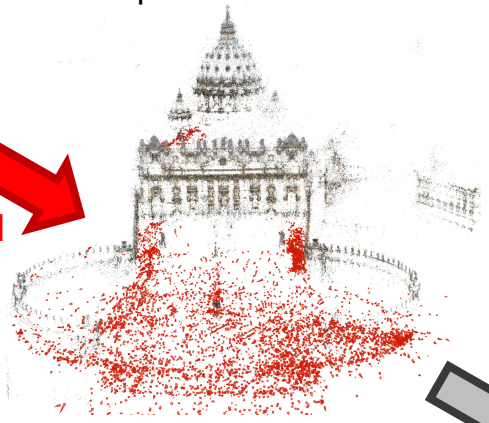
Assoc.

Scene Graph



SFM

Sparse Model

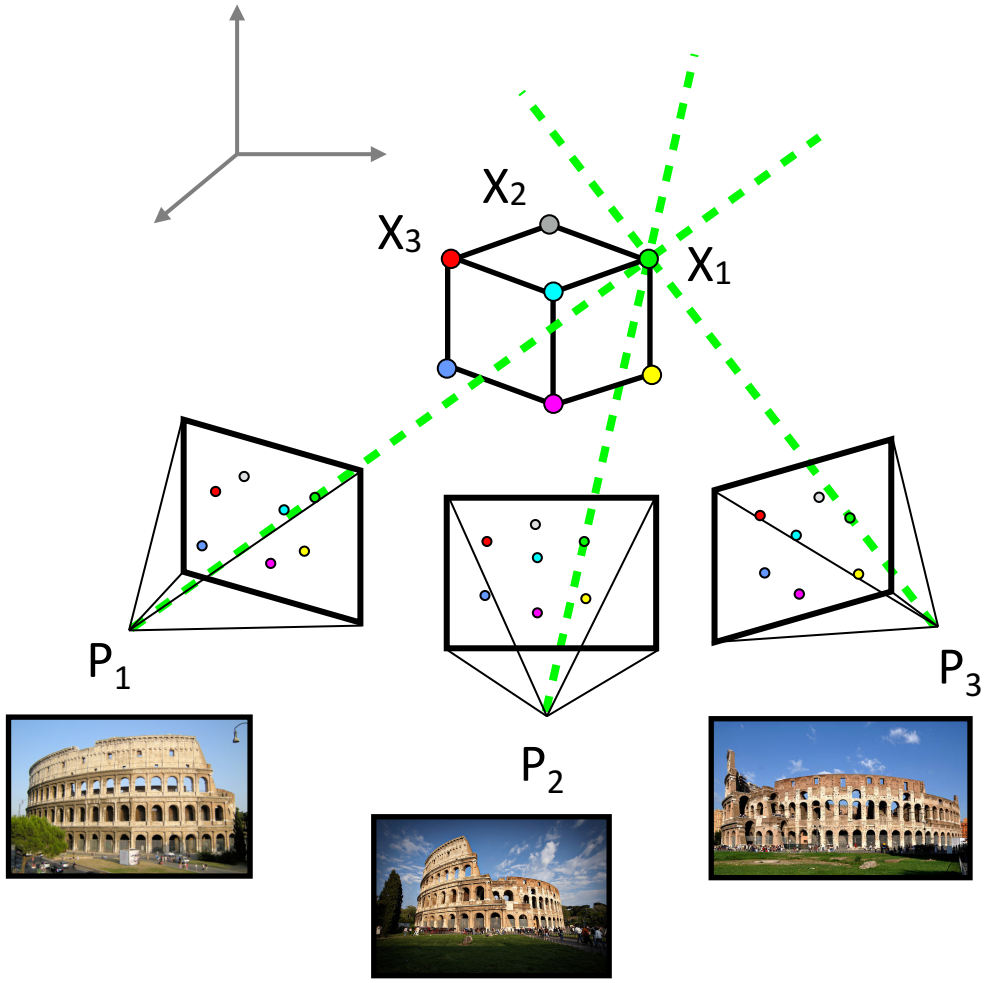


MVS

Dense Model

Structure-from-Motion

- Joint estimation of …
  - Structure $\mathbf{X}_i$
  - Cameras $\mathbf{P}_j$
- … from motion, i.e.
  - images at different viewpoints

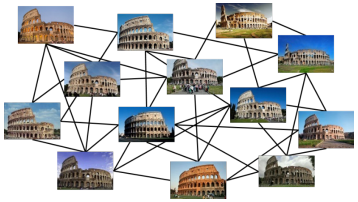


Data Association

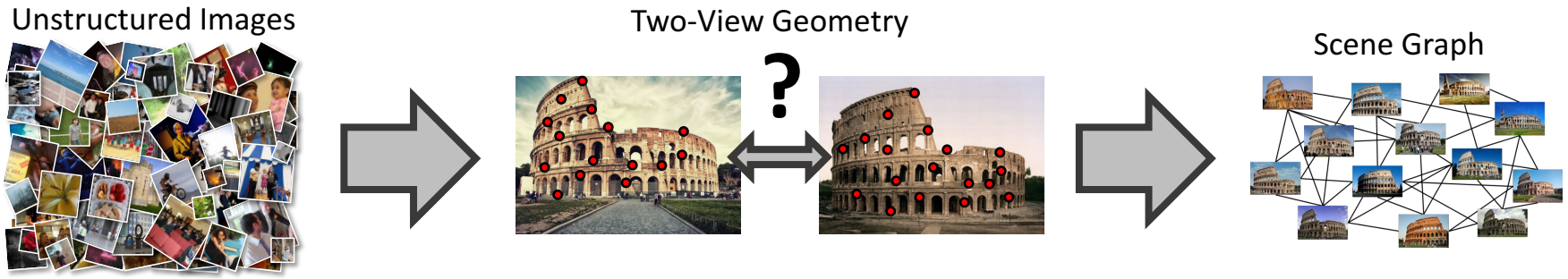
Unstructured Images

Connected Components

Scene Graph

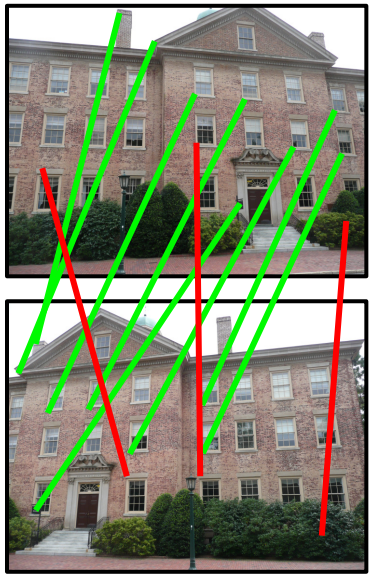


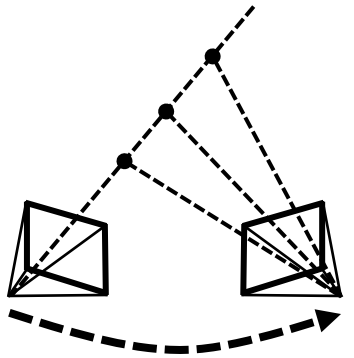
Data Association



Data Association

1. Feature Extraction → 2. Feature Matching → 3. Geometric Verification

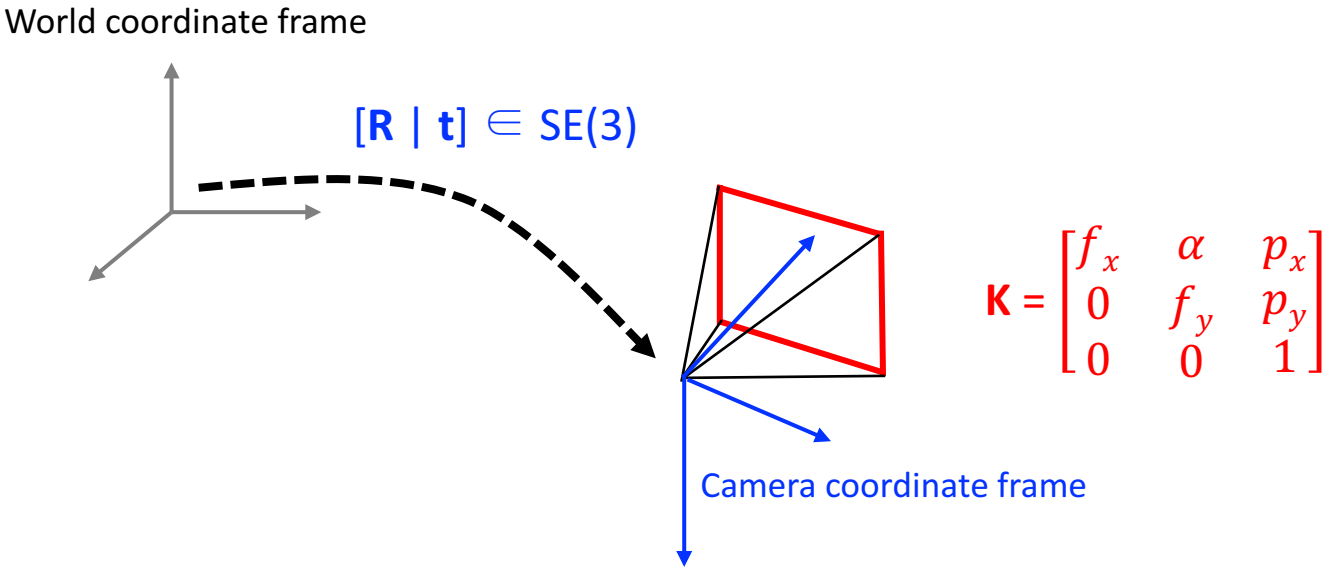




Robust Two-View Geometry Estimation

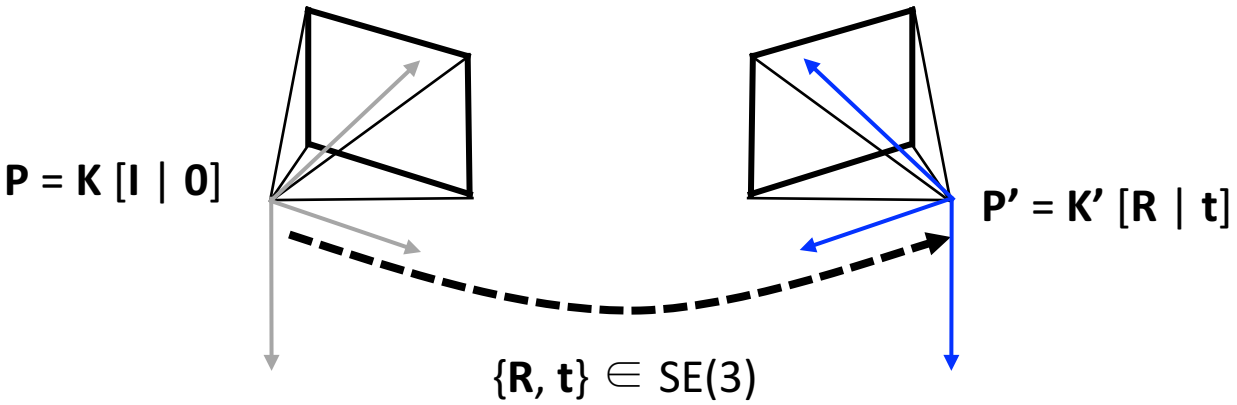
Two-View Geometry

- Absolute camera geometry: **P** = **K** [**R** | **t**]



Two-View Geometry

- Relative camera geometry: **P**, **P'**

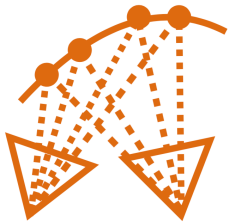


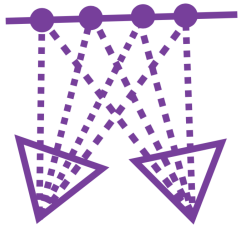
Scale of **t** unknown!

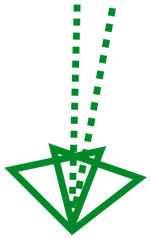
Two-View Geometry

- Model selection

General	Planar	Panoramic
• Fundamental matrix F (*uncalibrated*) • Essential matrix E (*calibrated*)	• Homography H	• Homography H
• 7 correspondences • 5 correspondences	• 4 correspondences	• 4 correspondences



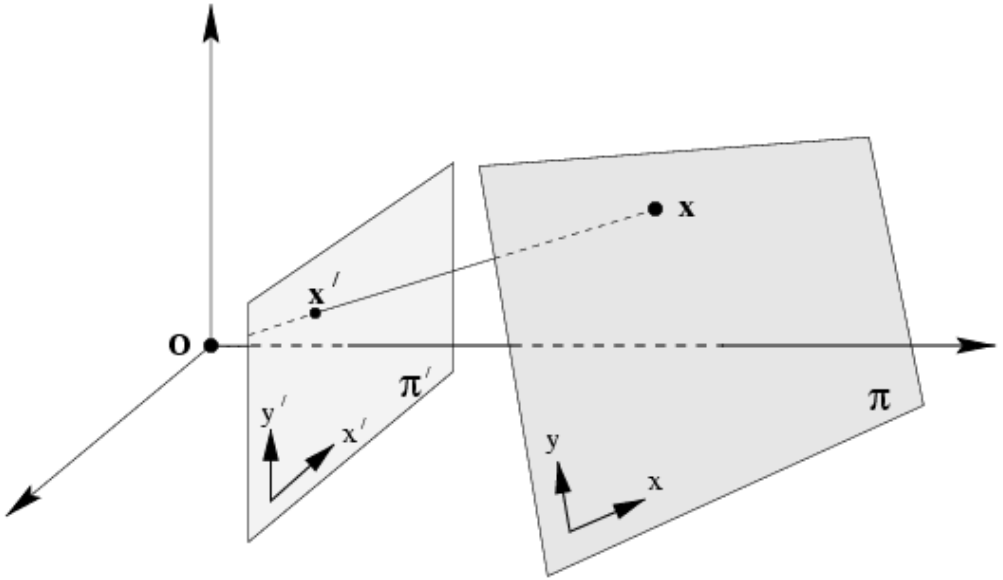




Hartley and Zisserman 2004, “Multiple View Geometry”

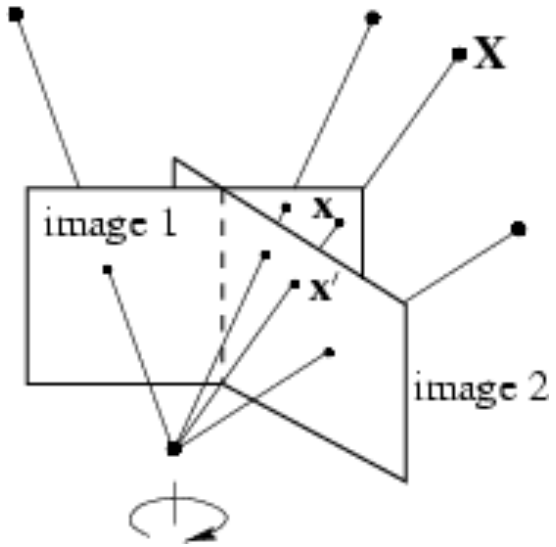
Two-View Geometry

- Homography $\mathbf{H} \in R^{3x3}$
  Linear transformation between two planes in P^2

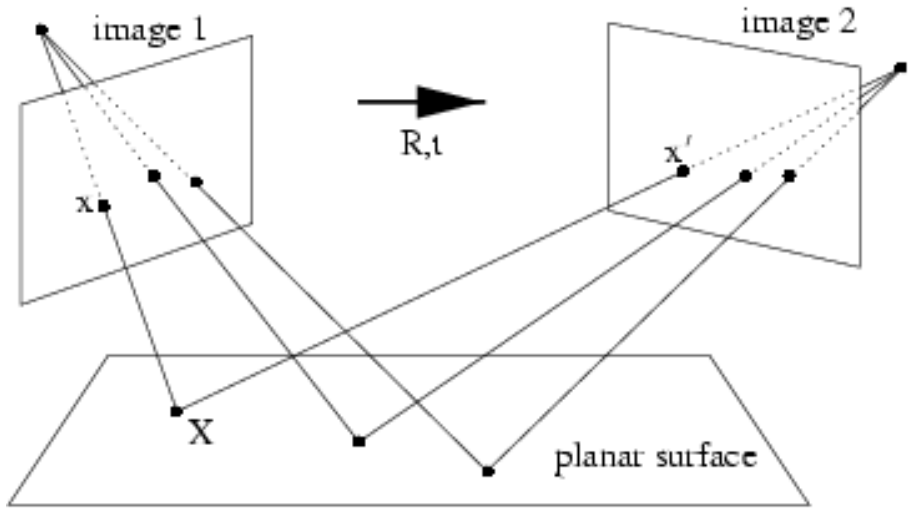


Two-View Geometry

- Homography

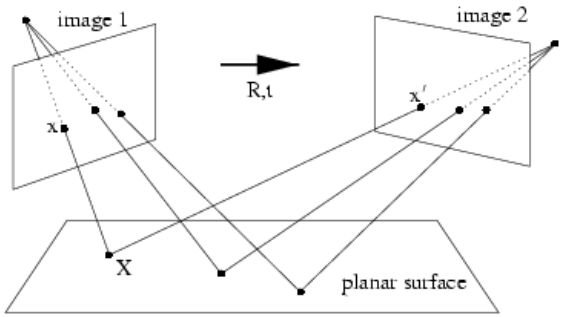


Panoramic



Planar

Two-View Geometry



- Homography
  - In the calibrated case
    - Two cameras: $P = K [I \mid 0]$ and $P' = K' [R \mid t]$
    - A plane: $\pi = (n^T, d)^T$
  - The homography is given by $x' = H x$

$$H = K' (R - t n^T/d) K^{-1}$$

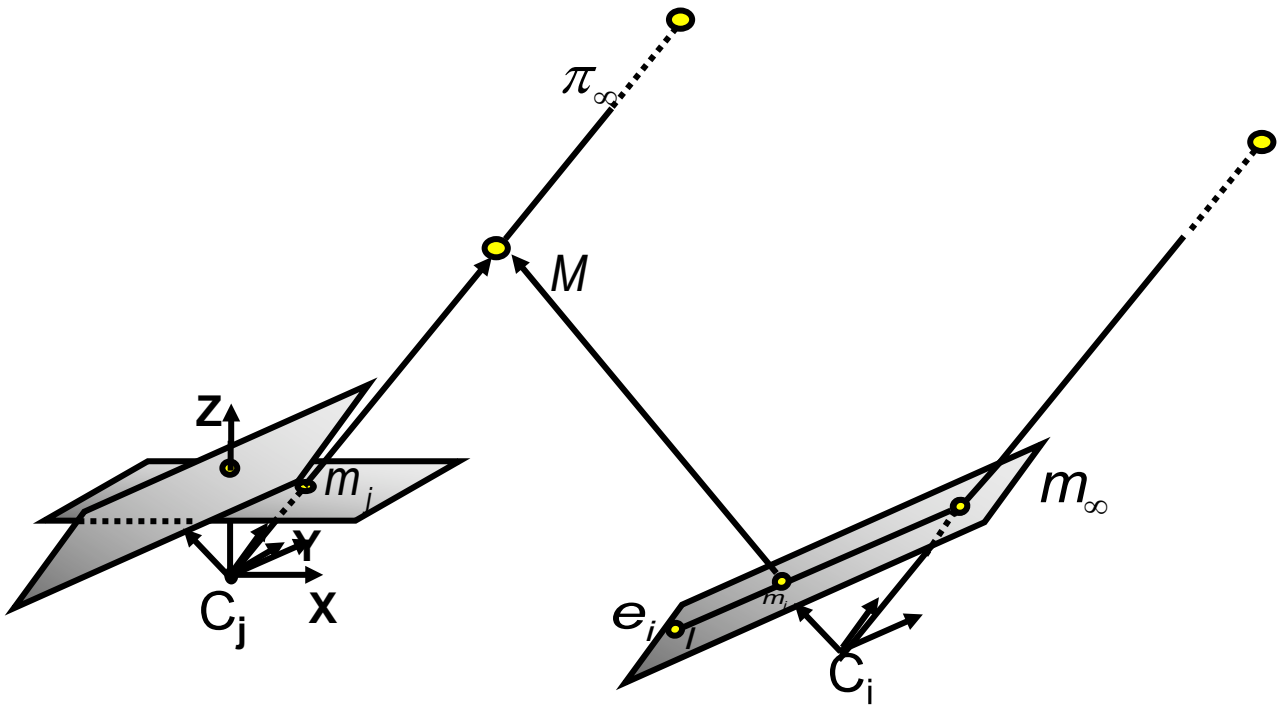
Two-View Geometry

$$m_\infty = H_\infty m_i$$

- Fundamental matrix **F**

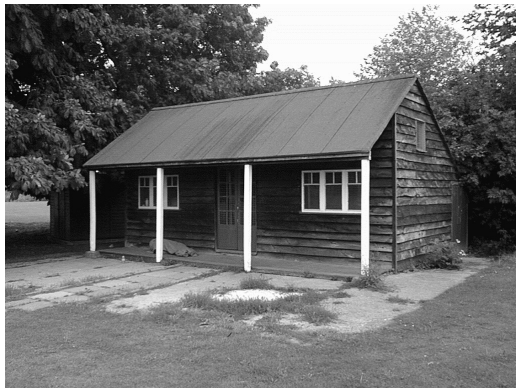
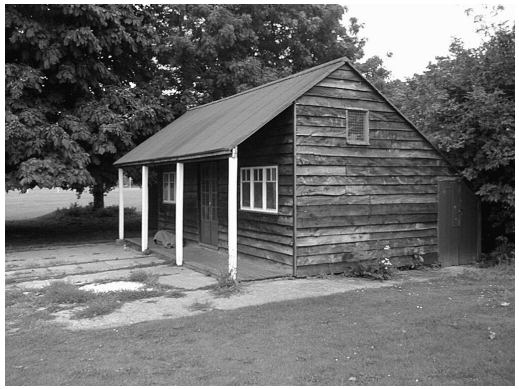
$$l = [e]_X m_\infty = \underbrace{[e]_X H_\infty}_{F} m_i$$

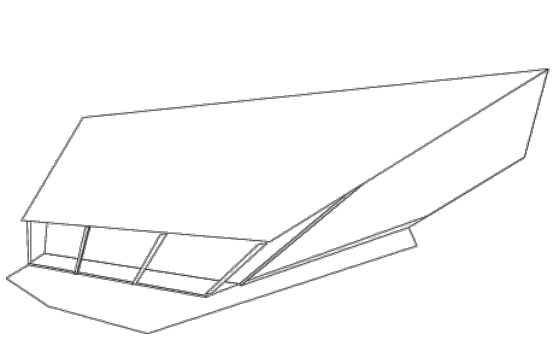
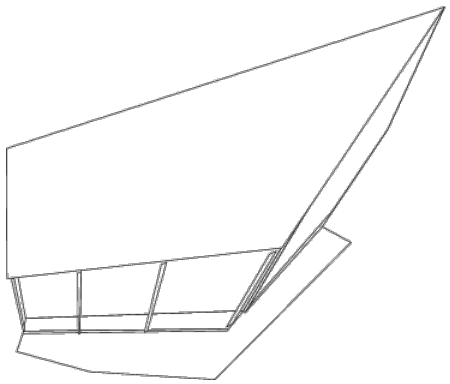
$$m_j l = \underbrace{m_j F m_i = 0}_{\text{epipolar constraint}}$$



Two-View Geometry

- Projective ambiguity

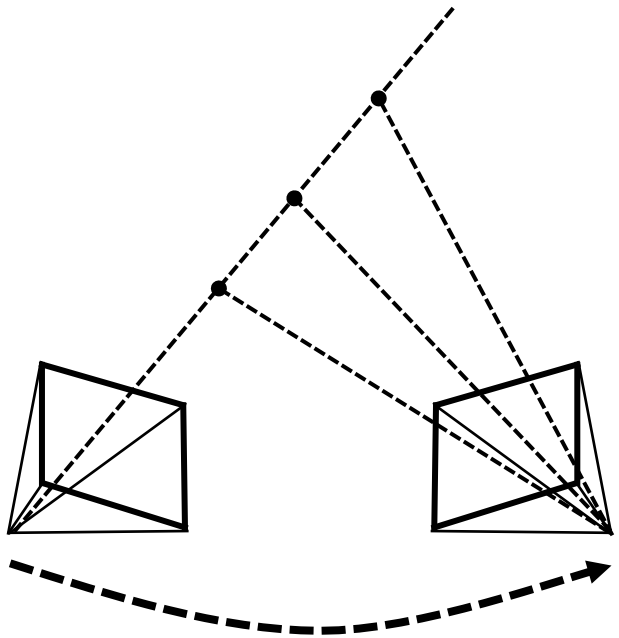



Two-View Geometry

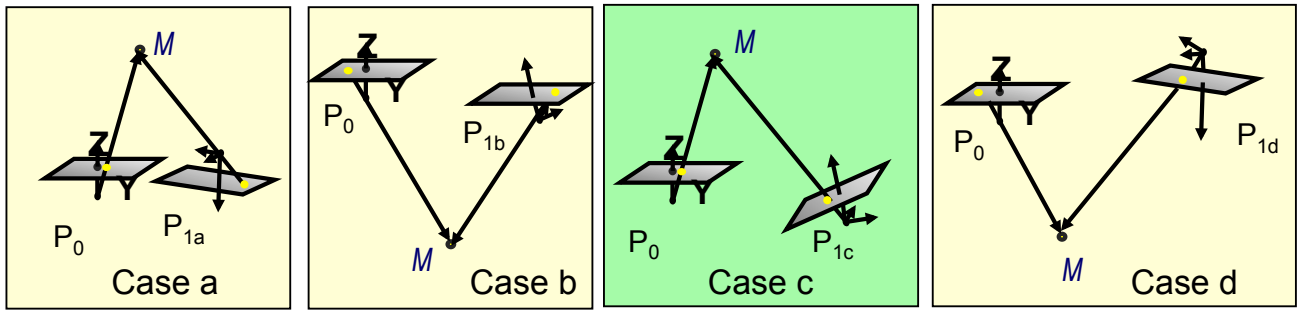
- Essential matrix **E**

$$\mathbf{E} = \mathbf{K}^T \mathbf{F} \mathbf{K}' = \mathbf{t}_x \mathbf{R}$$



Two-View Geometry

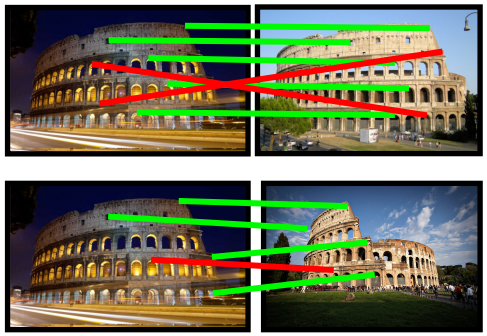
- Robust estimation using RANSAC
- Model selection (**F**, **E**, **H**)
  - Inlier maximization
  - GRIC criterion Torr 1998, “An assessment of information criteria for motion model selection"
  - QDEGSAC Frahm and Pollefeys 2006, “RANSAC for (quasi-) degenerate data (QDEGSAC)”
- Decompose **E**, **H** to {**R**, **t**}

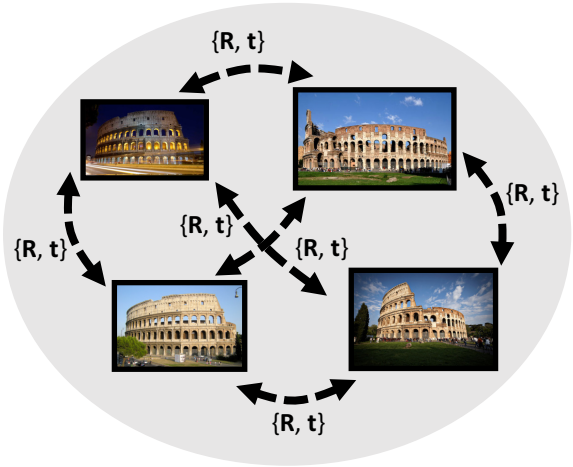


Scene Graph

Unstructured Images

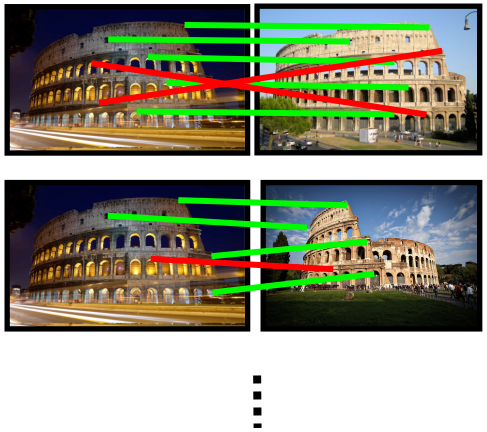
Inlier/outlier correspondences

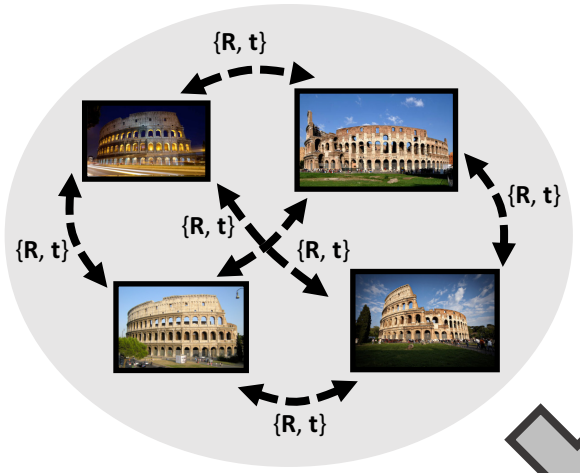


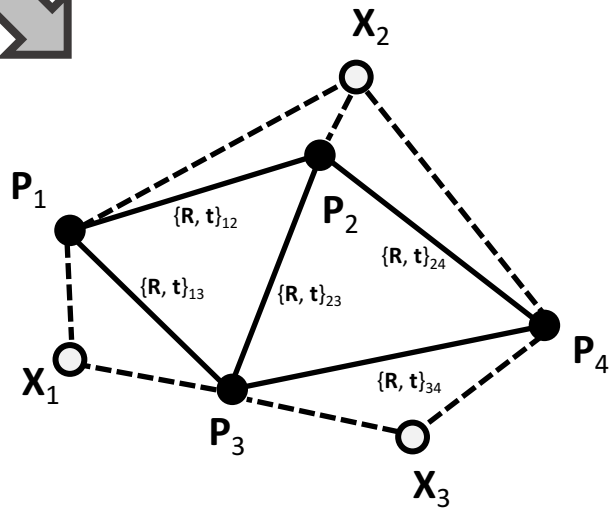


Scene Graph

Inlier/outlier correspondences

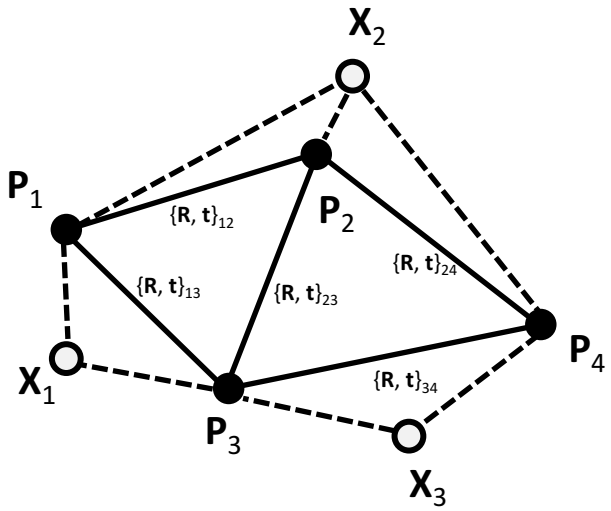


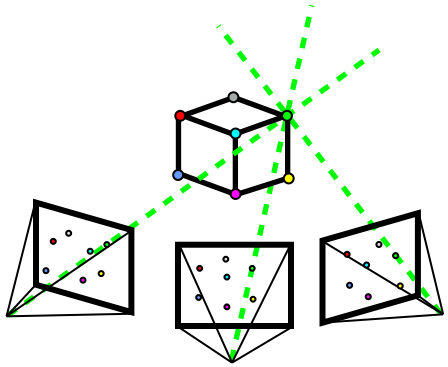




Structure-from-Motion

- From relative to absolute cameras and structure

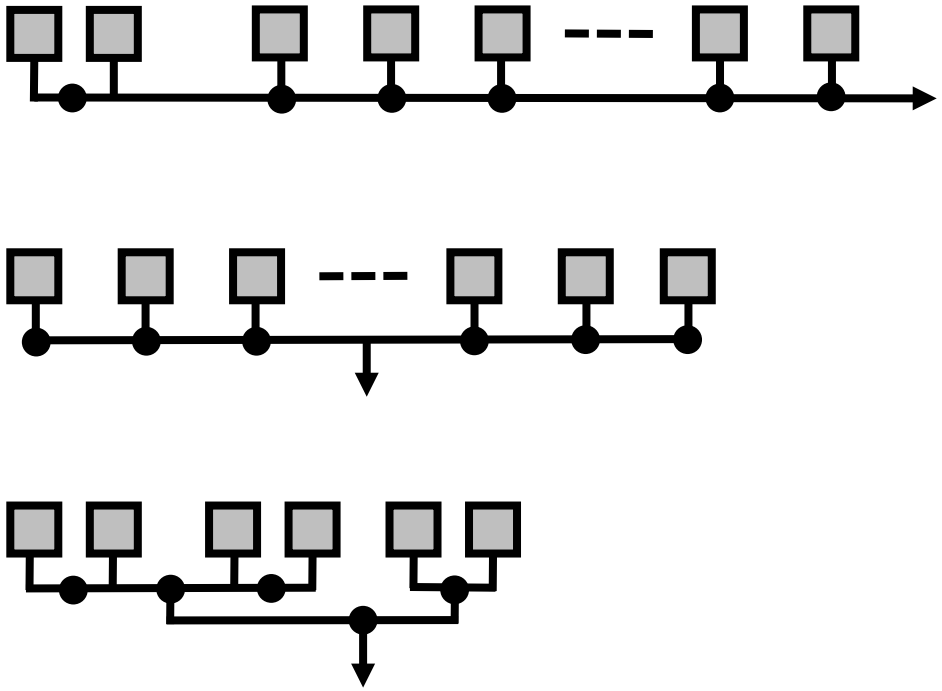




Scale of **t** unknown!
Outlier correspondences!
Outlier image pairs!

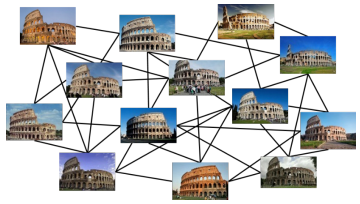
Structure-from-Motion

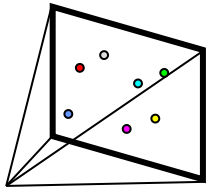
- 3 paradigms
  - Incremental
  - Global
  - Hierarchical



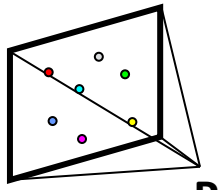
Incremental SfM

- Initialization
  1. Choose two non-panoramic views ($\|t\| \neq 0$)





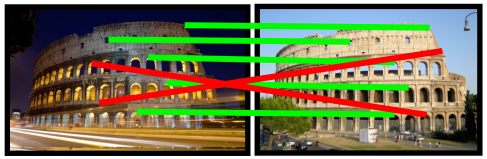
P = K [I | 0]

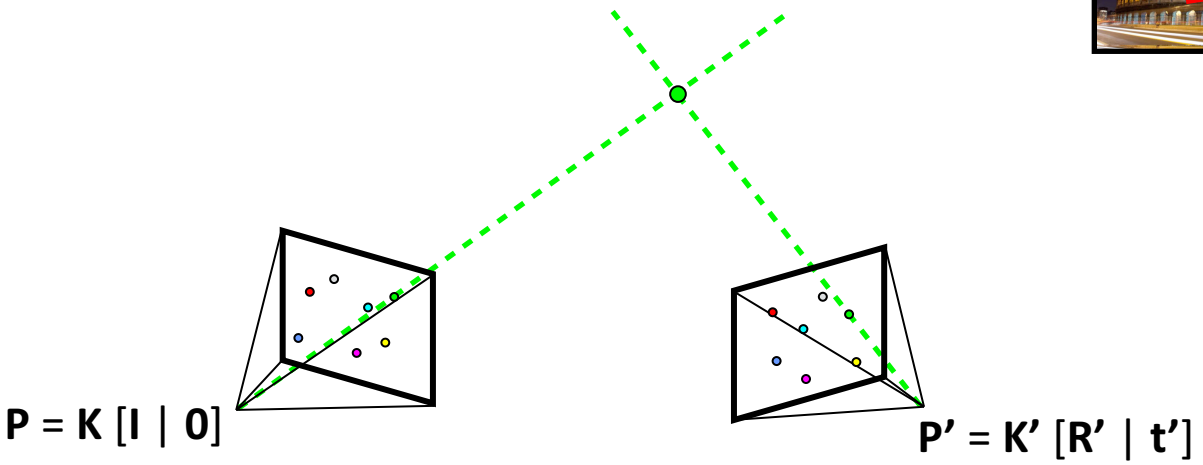


P’ = K’ [R’ | t’]

Incremental SfM

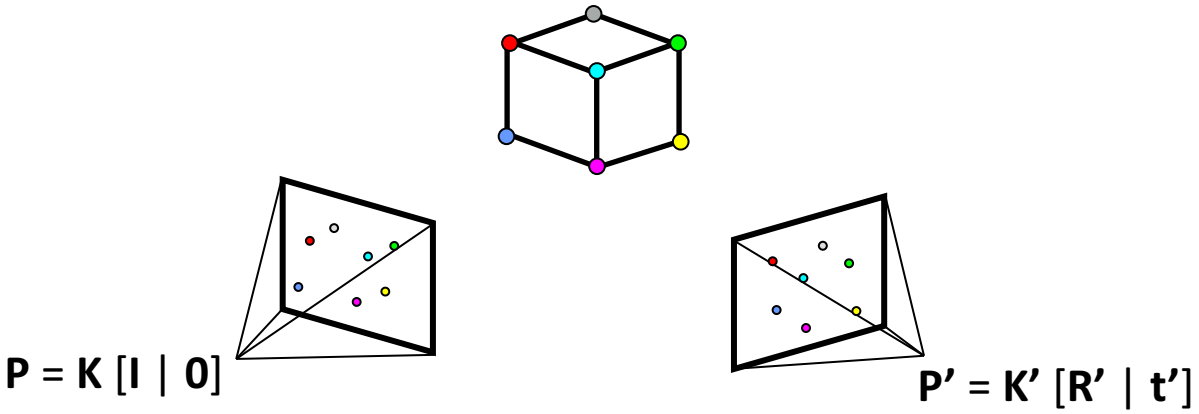
- Initialization
  1. Choose two non-panoramic views ($\|t\| \neq 0$)
  2. Triangulate inlier correspondences





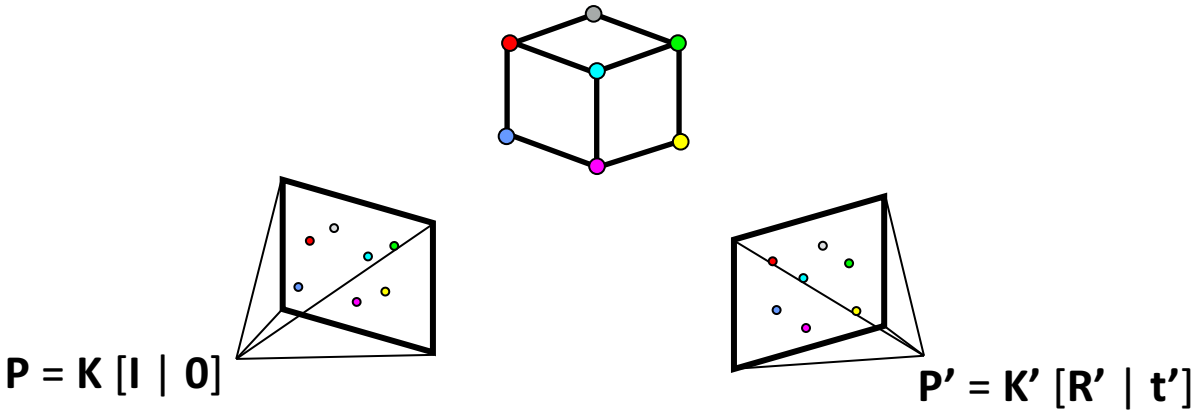
Incremental SfM

- Initialization
  1. Choose two non-panoramic views ($\|t\| = 1$)
  2. Triangulate inlier correspondences



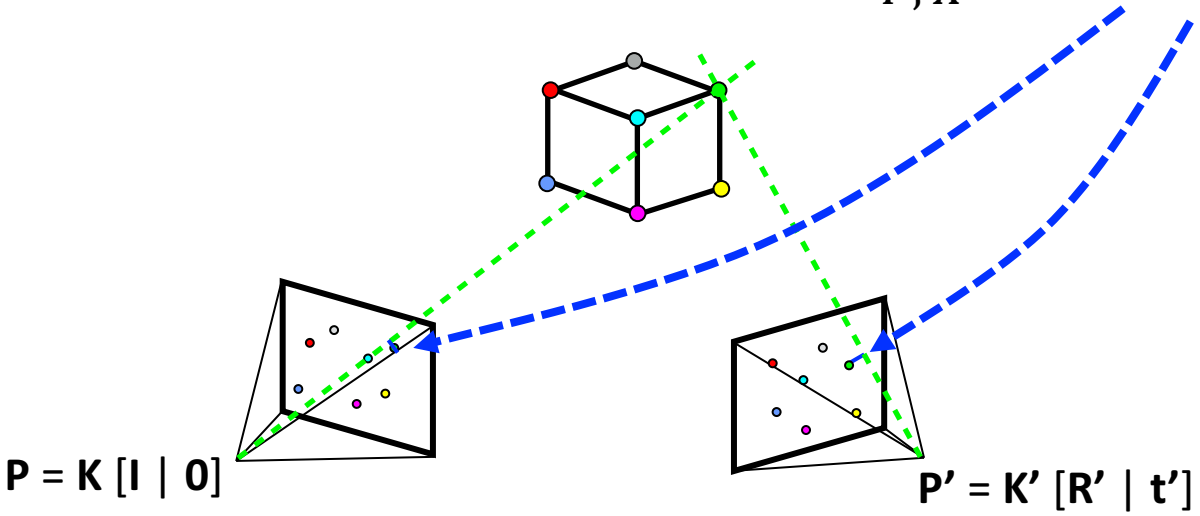
Incremental SfM

- Initialization
  1. Choose two non-panoramic views ($\|t\| = 1$)
  2. Triangulate inlier correspondences
  3. Bundle adjustment



Incremental SfM

- Bundle adjustment
  - Non-linear refinement of structure and motion
  - Minimize reprojection error: $\min_{P,X} \|x - \pi(P,X)\|$

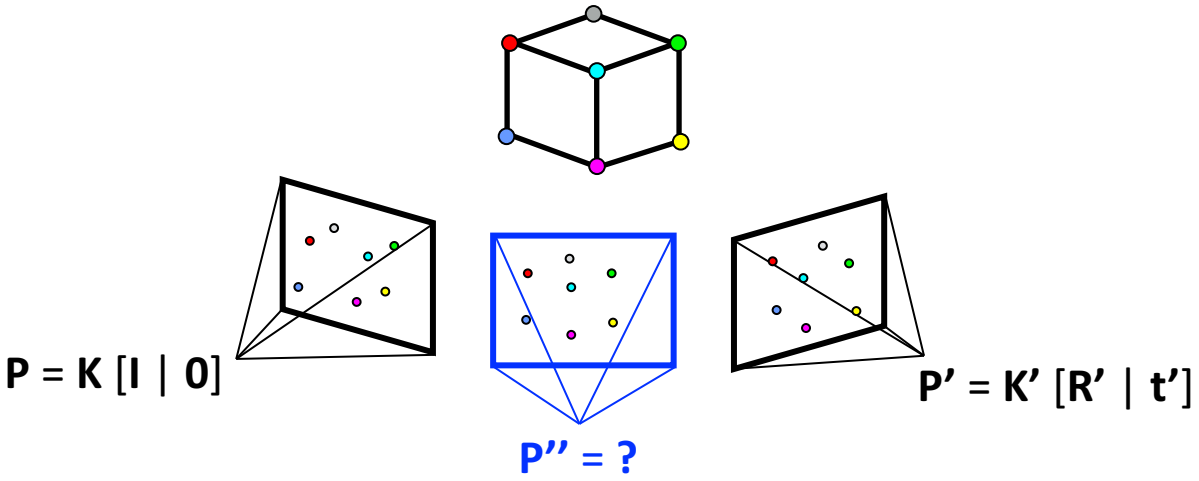


Ceres-Solver, http://ceres-solver.org/

Triggs et al., "Bundle Adjustment – A Modern Synthesis"

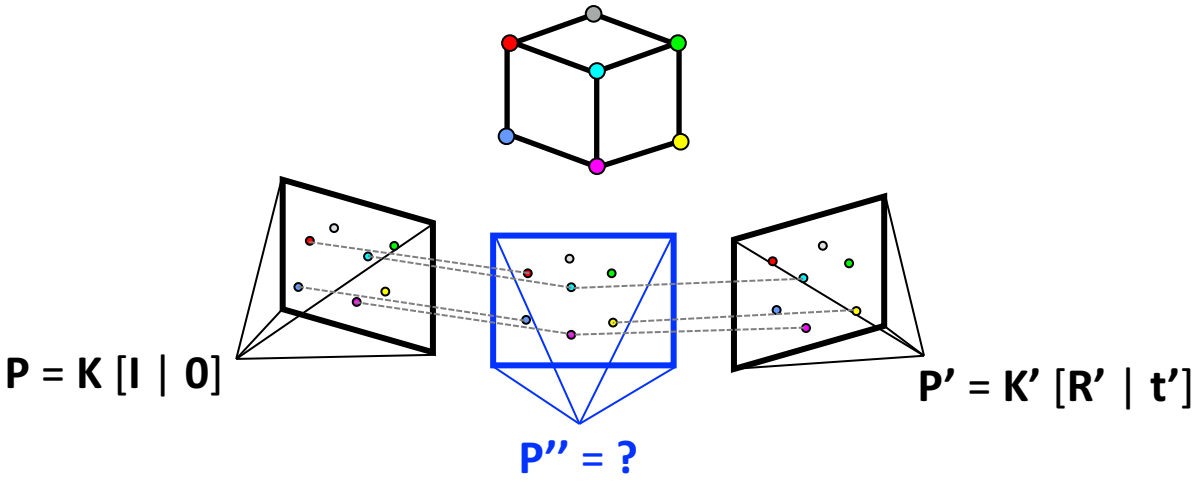
Incremental SfM

- Absolute camera registration



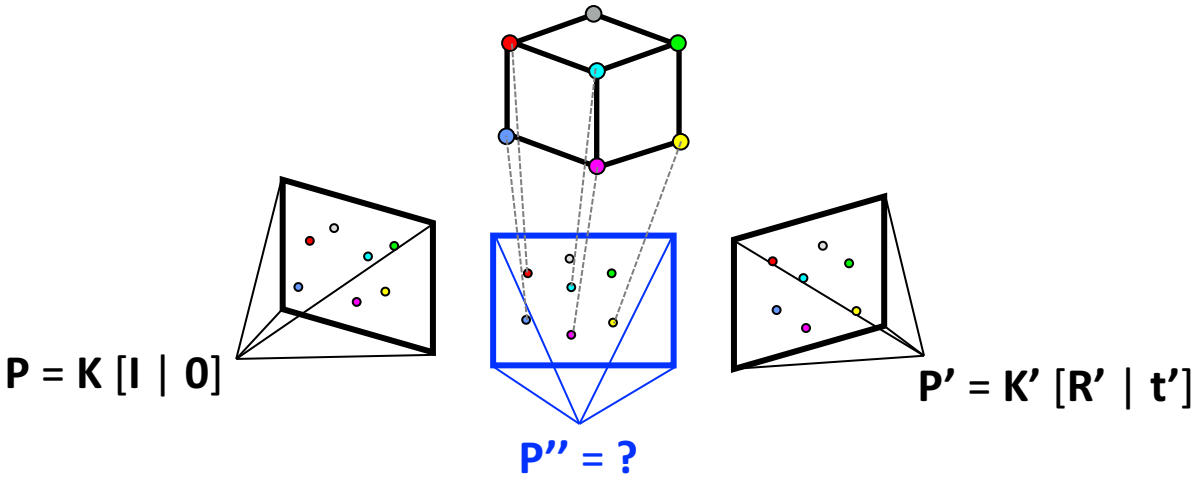
Incremental SfM

- Absolute camera registration
  1. Find 2D-3D correspondences



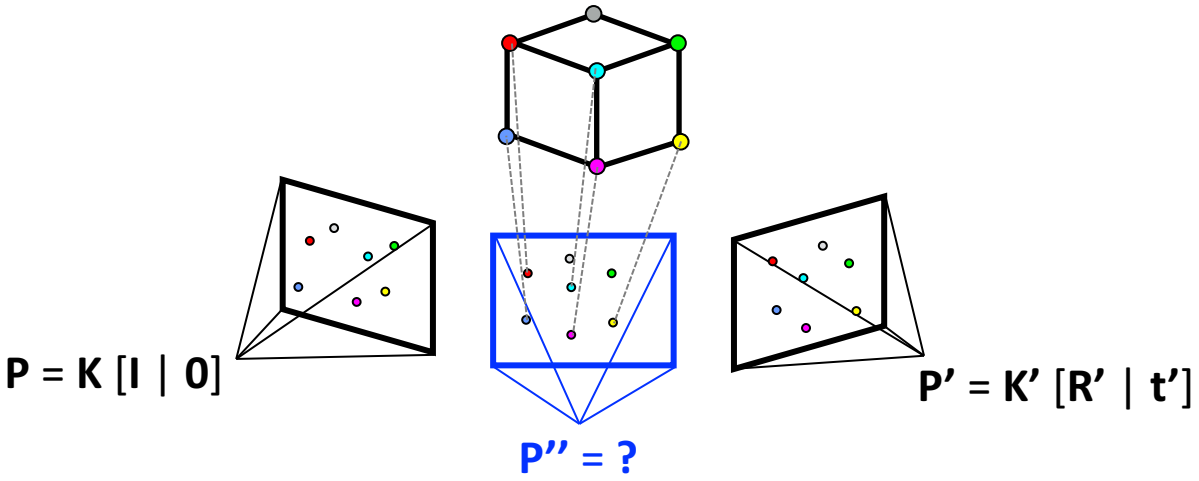
Incremental SfM

- Absolute camera registration
  1. Find 2D-3D correspondences



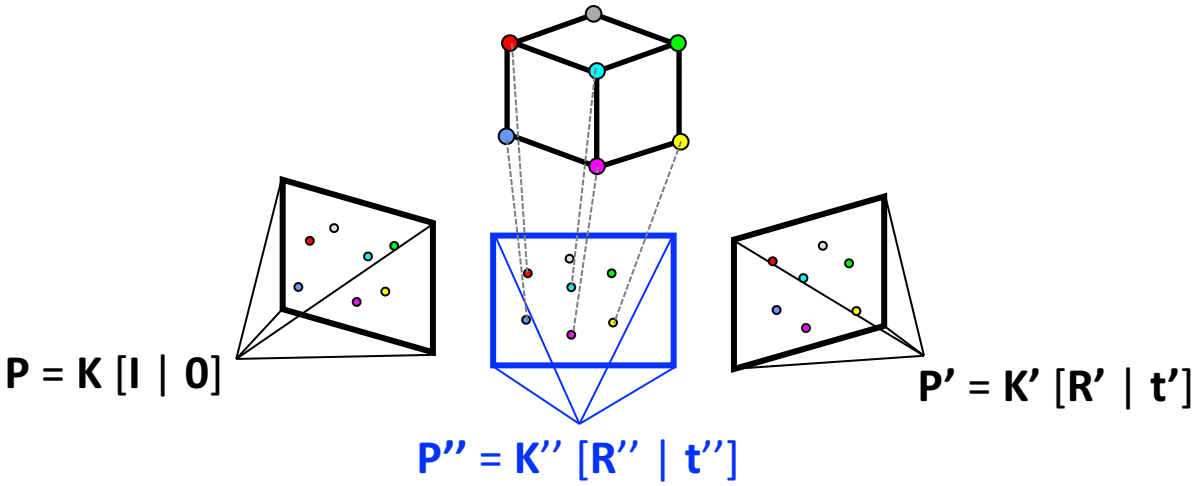
Incremental SfM

- Absolute camera registration
  1. Find 2D-3D correspondences
  2. Solve Perspective-n-Point problem

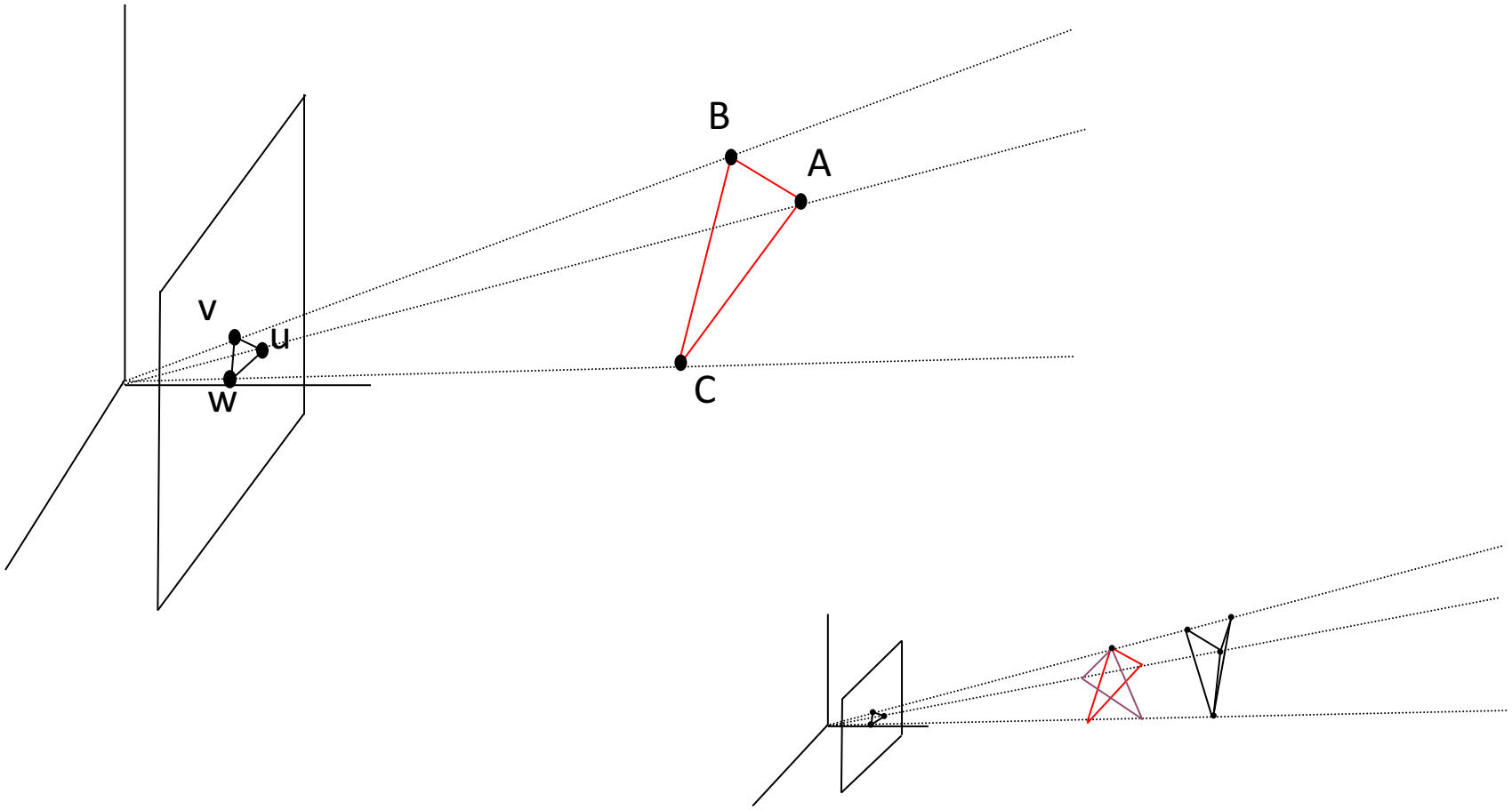


Incremental SfM

- Absolute camera registration
  1. Find 2D-3D correspondences
  2. Solve Perspective-n-Point problem

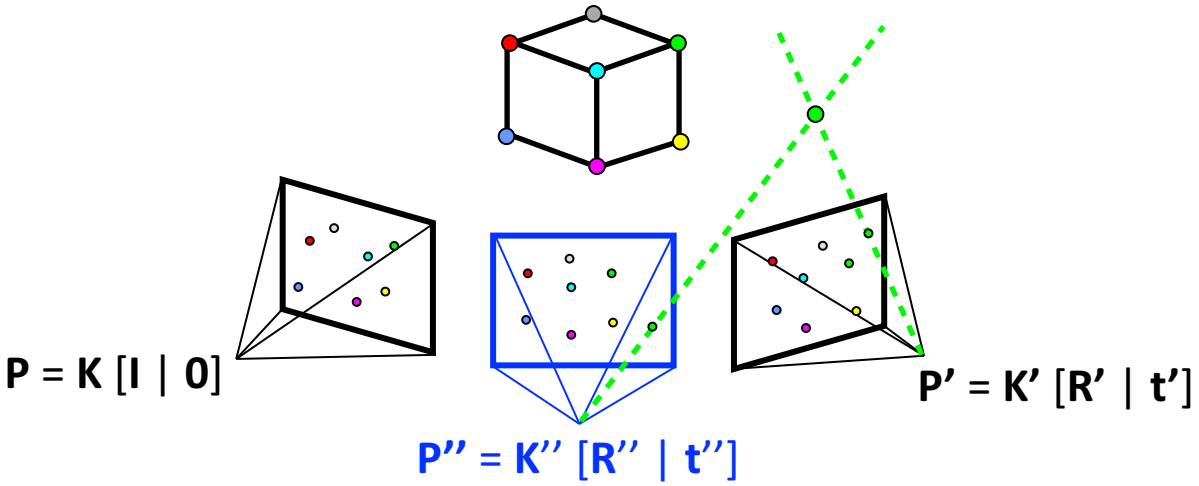


Perspective-3-Point Problem



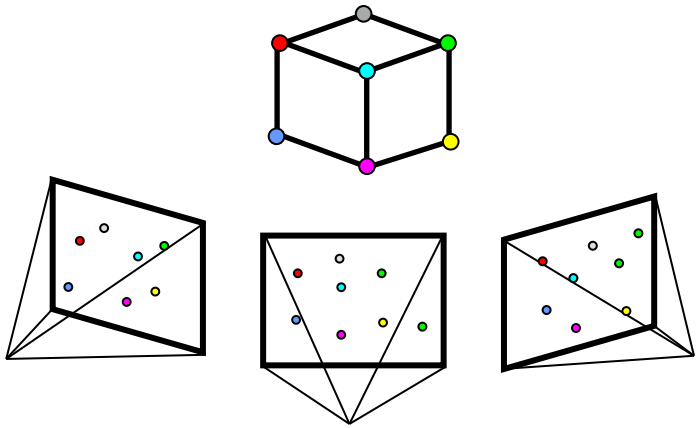
Incremental SfM

- Absolute camera registration
  1. Find 2D-3D correspondences
  2. Solve Perspective-n-Point problem
  3. Triangulate new points



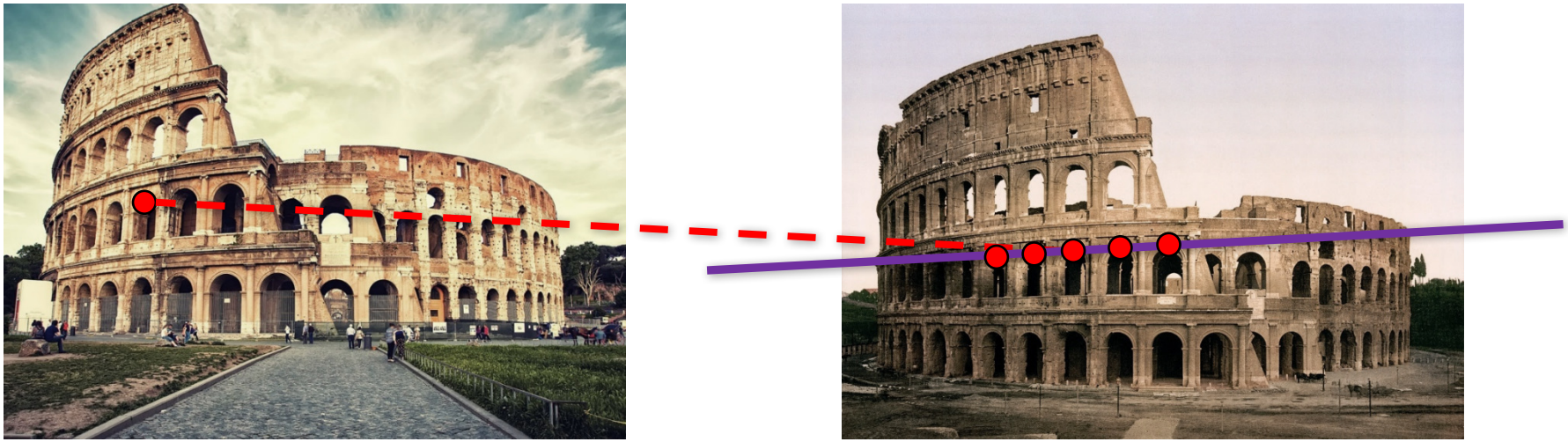
Incremental SfM

- Bundle adjustment $\min_{P,X} \|x - \pi(P,X)\|$



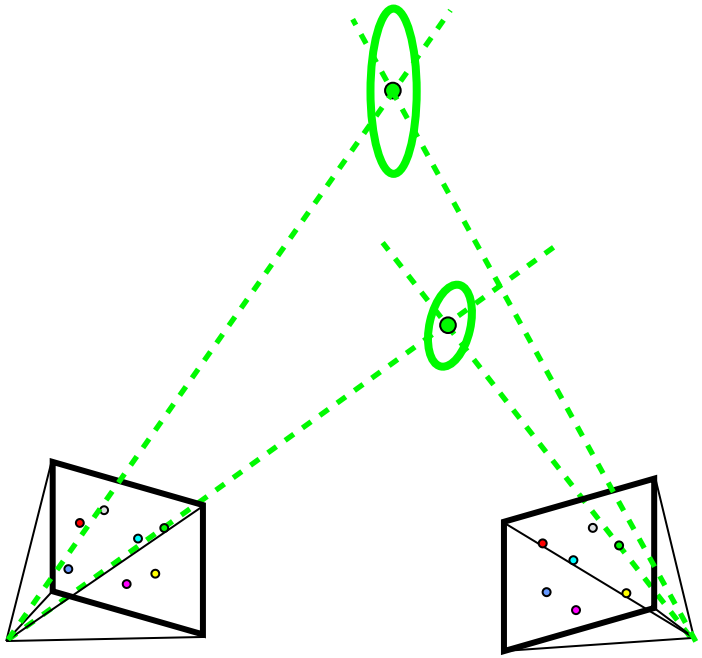
Incremental SfM

- Outlier filtering
  - Remove points with large reprojection error

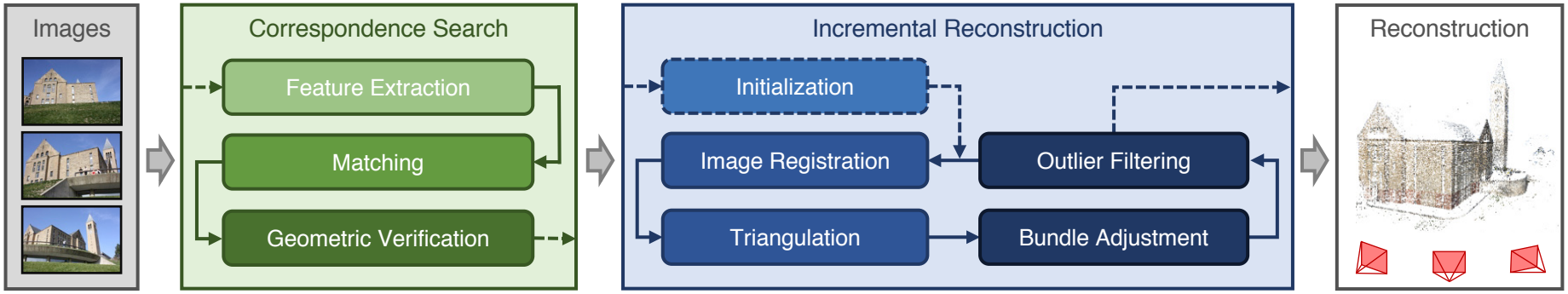


Incremental SfM

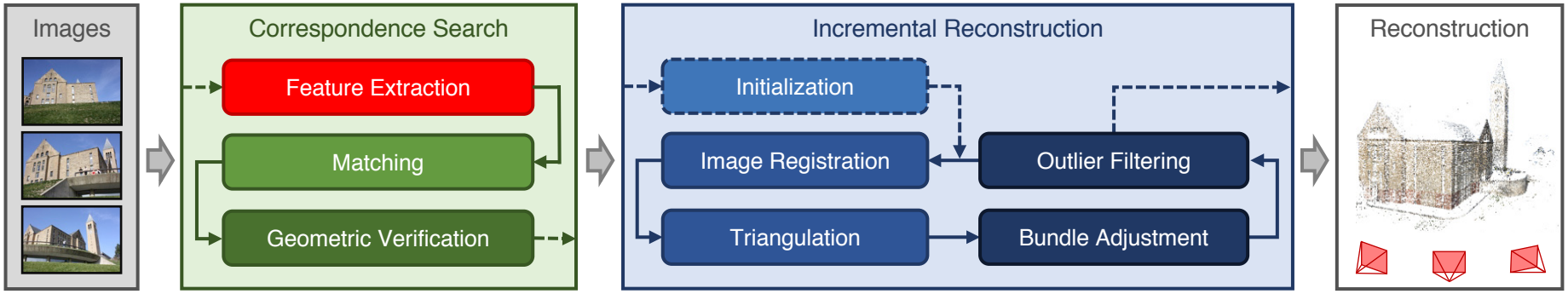
- Outlier filtering
  - Remove points with large reprojection error
  - Remove points at “infinity”



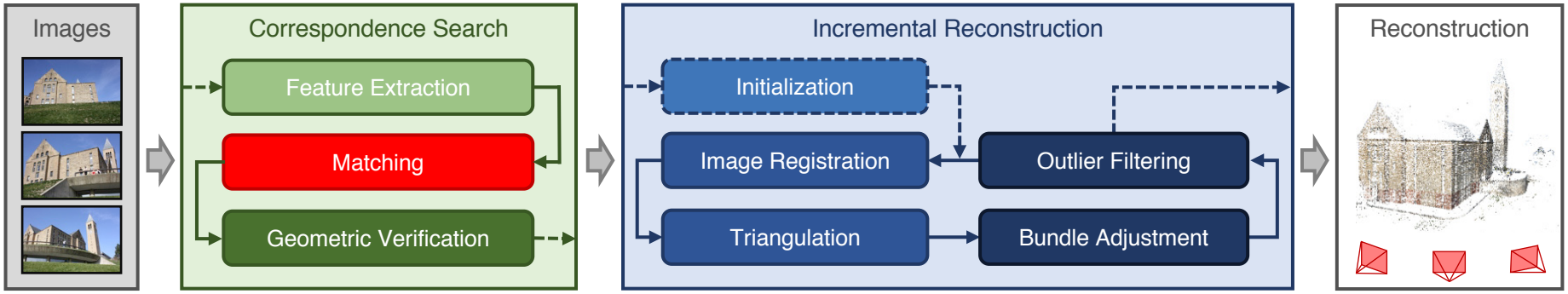
Incremental SfM

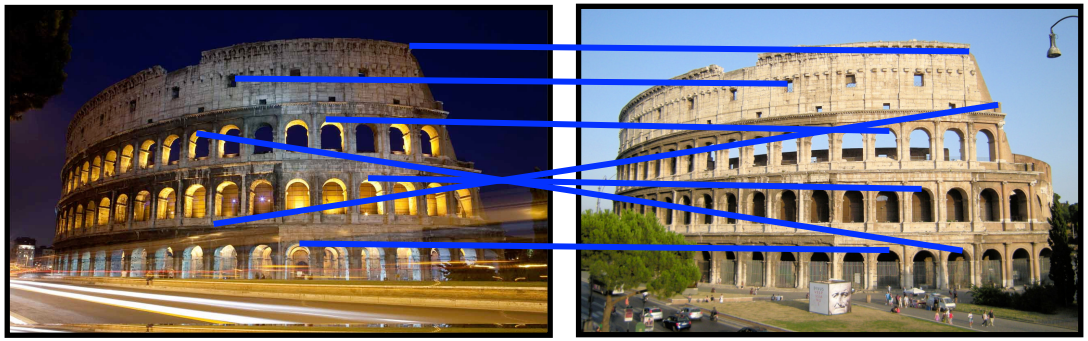


Incremental SfM

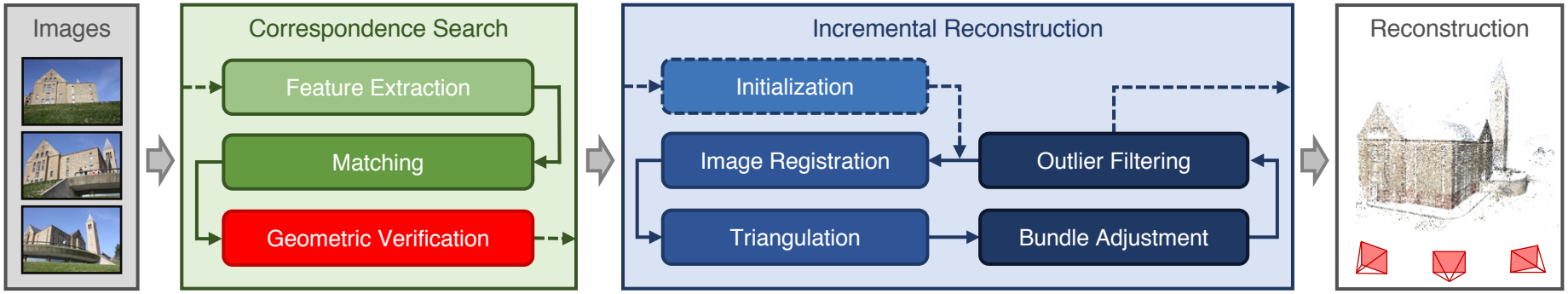


Incremental SfM

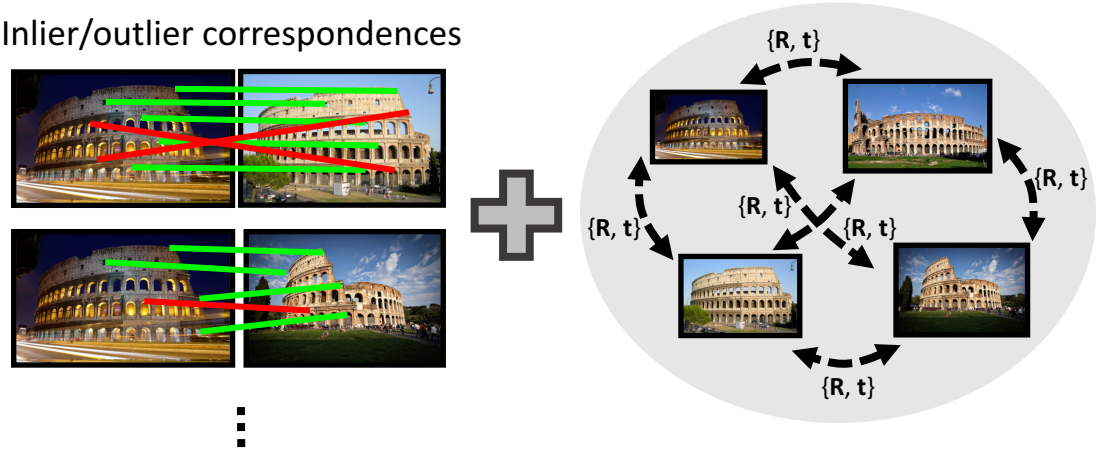




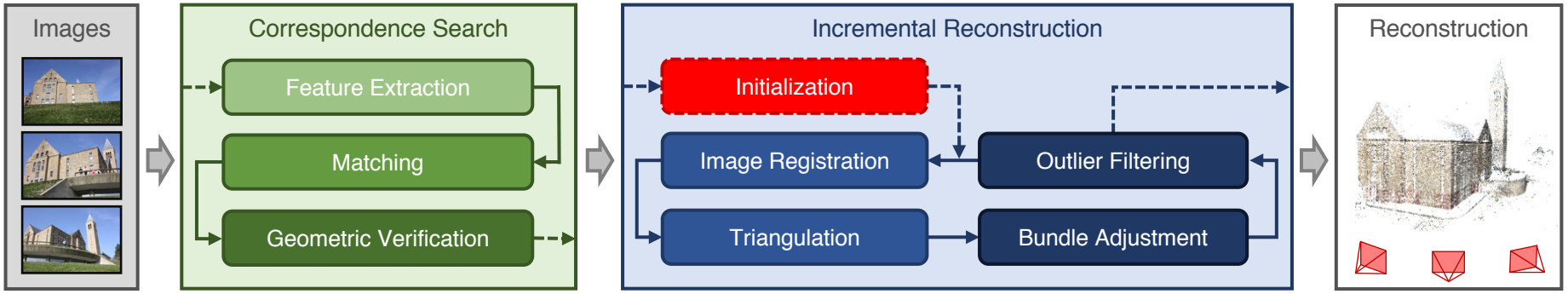
Incremental SfM

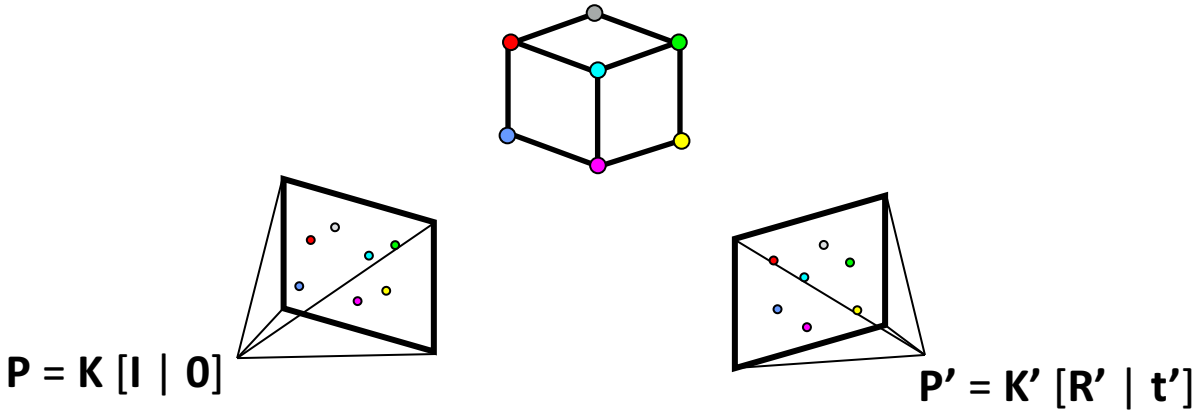


Inlier/outlier correspondences

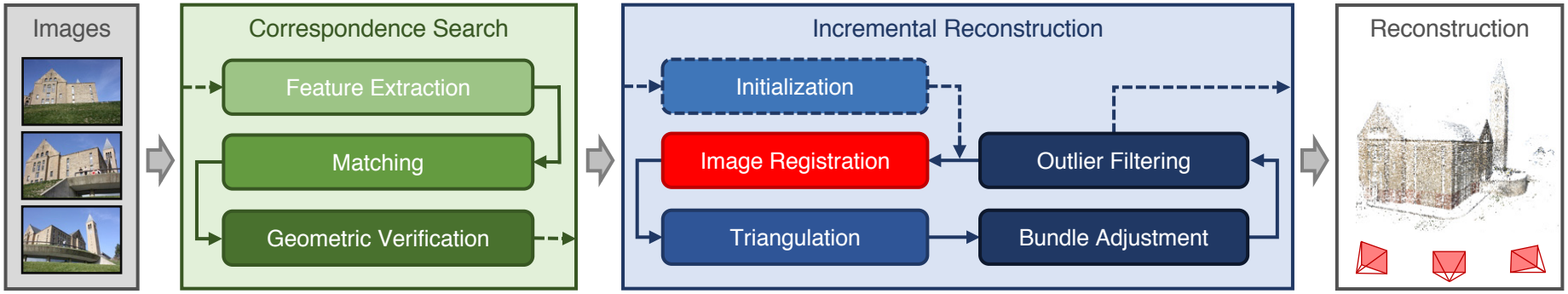


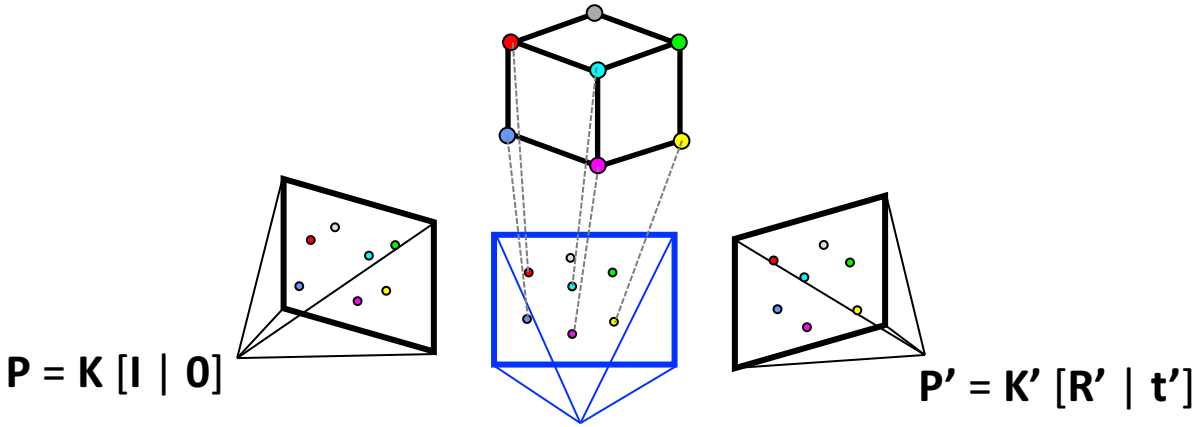
Incremental SfM



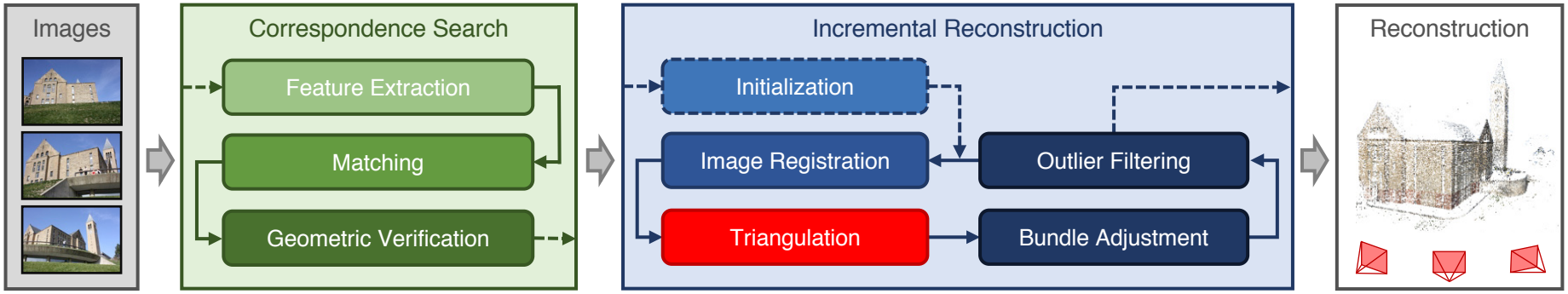


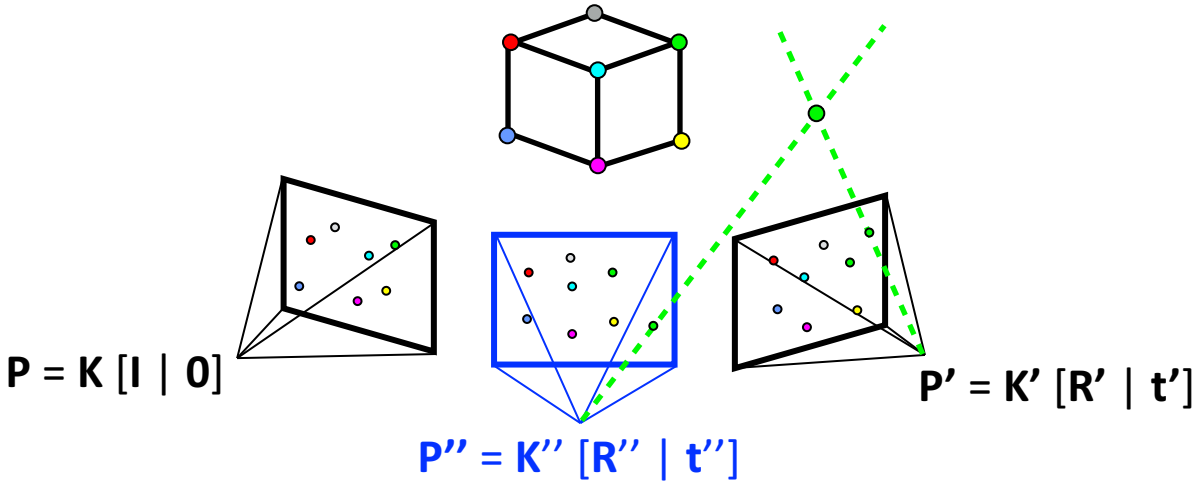
Incremental SfM



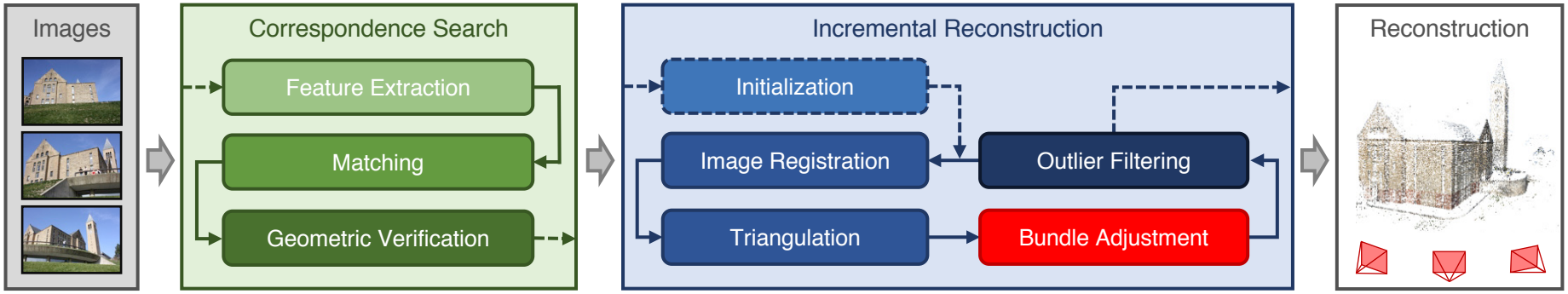


Incremental SfM



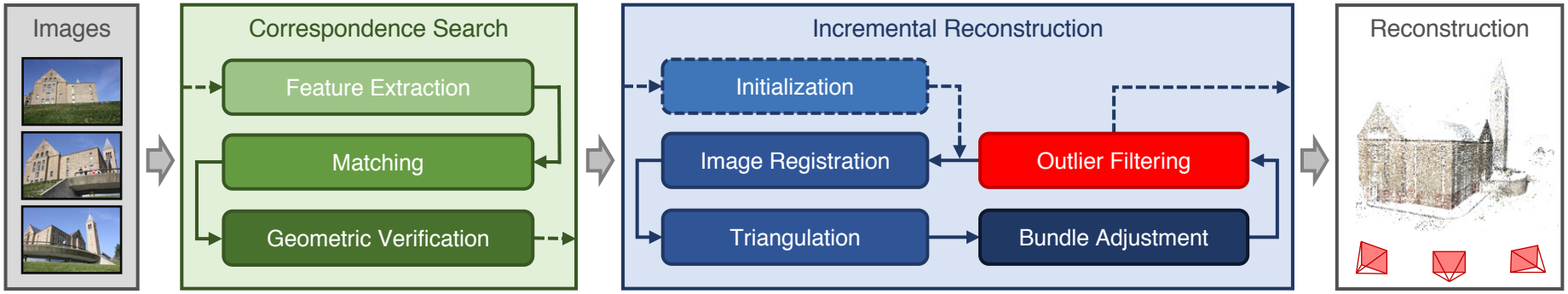


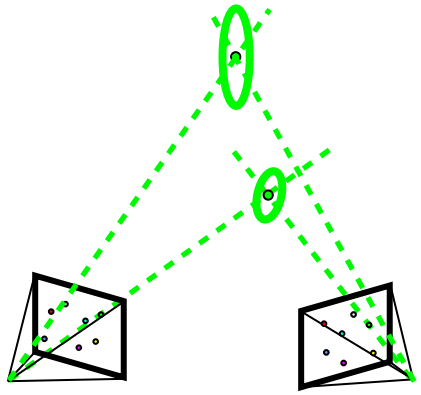
Incremental SfM



$$\min_{P,X} \|x - \pi(P,X)\|$$

Incremental SfM





Demo: Incremental SfM

Global SfM

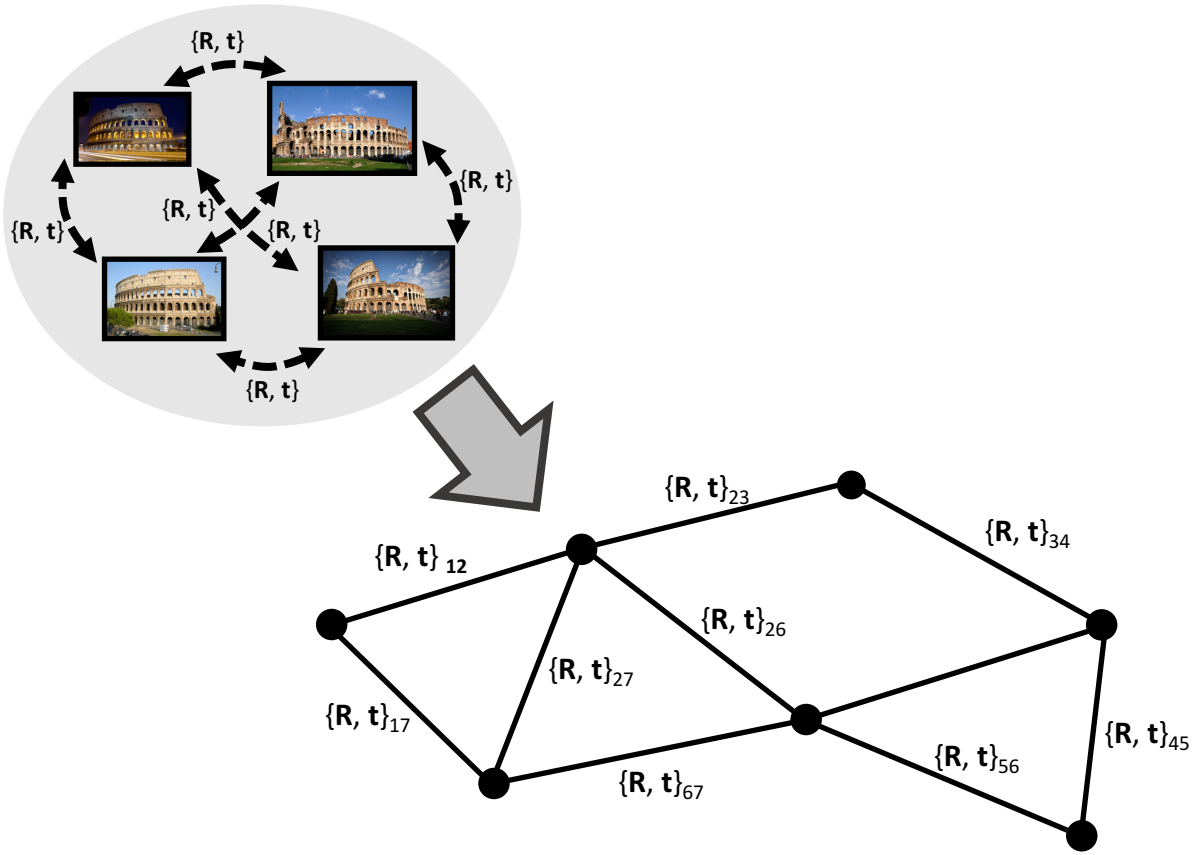


Image
Two-View Geometry
{R, t}
{R, t}
{R, t}
{R, t}
{R, t}
{R, t}
{R, t}$_{23}$
{R, t}$_{34}$
{R, t}$_{12}$
{R, t}$_{26}$
{R, t}$_{27}$
{R, t}$_{17}$
{R, t}$_{45}$
{R, t}$_{56}$
{R, t}$_{67}$

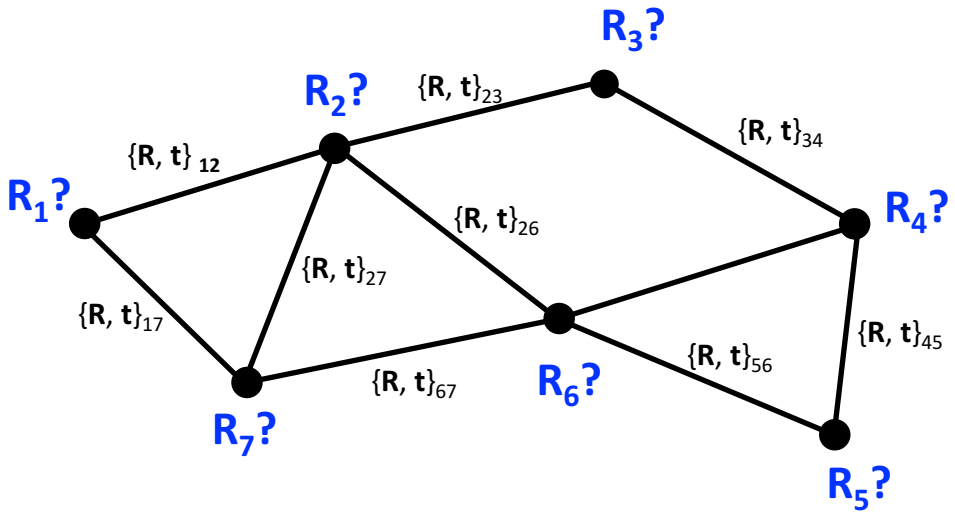
Global SfM

Image

Two-View Geometry

1. Estimate global rotations: $\min_{R} \|R_{ij} - R_j R_i^T\|$

Chatterje and Govindu 2013, "Efficient and Robust Large-Scale Rotation Averaging"



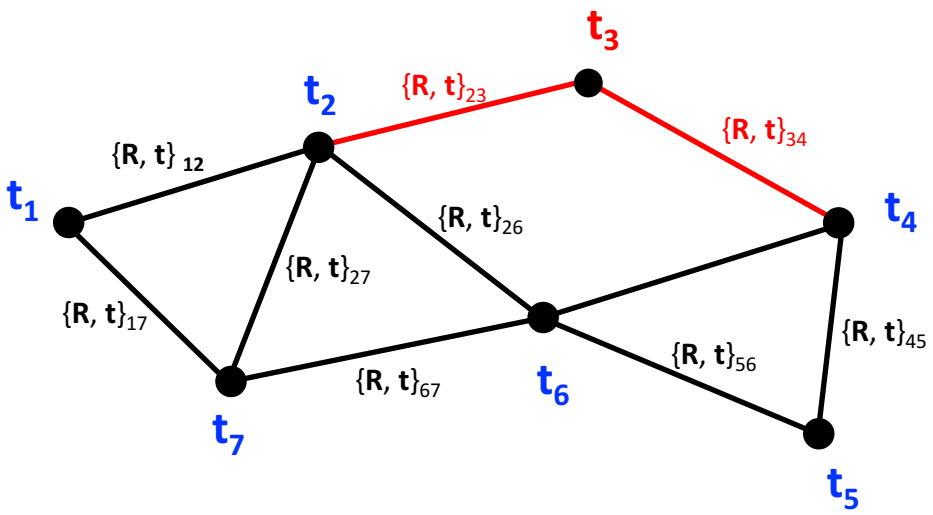
Global SfM

Image

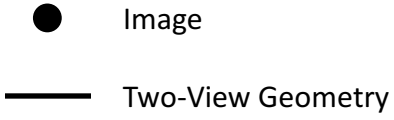
Two-View Geometry

1. Estimate global rotations: $\min_R \|R_{ij} - R_j R_i^T\|$

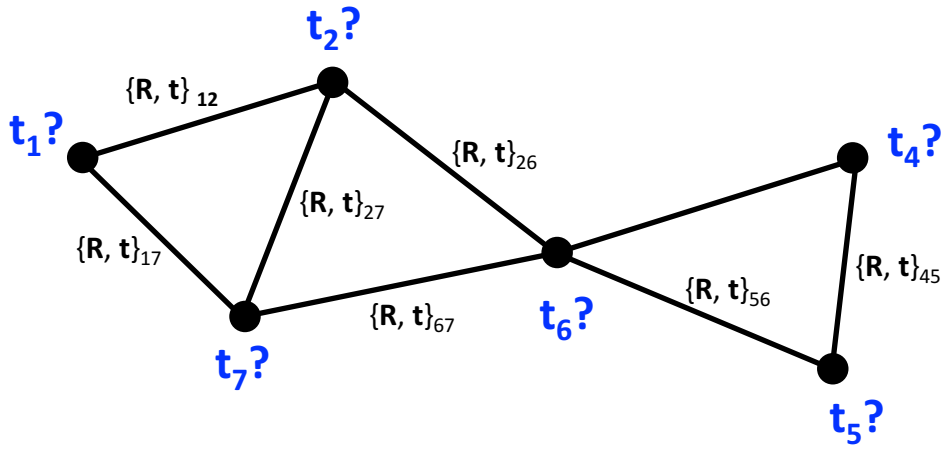
   Filter relative rotations: $\|R_{ij} - R_j R_i^T\| > \epsilon$



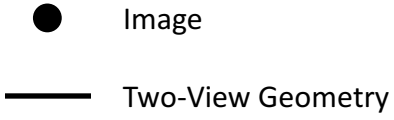
Global SfM



1. Estimate and filter global rotations
2. Estimate global translations: $\min_t \left\| t_{ij} - \frac{t_i - t_j}{\|t_i - t_j\|} \right\|$

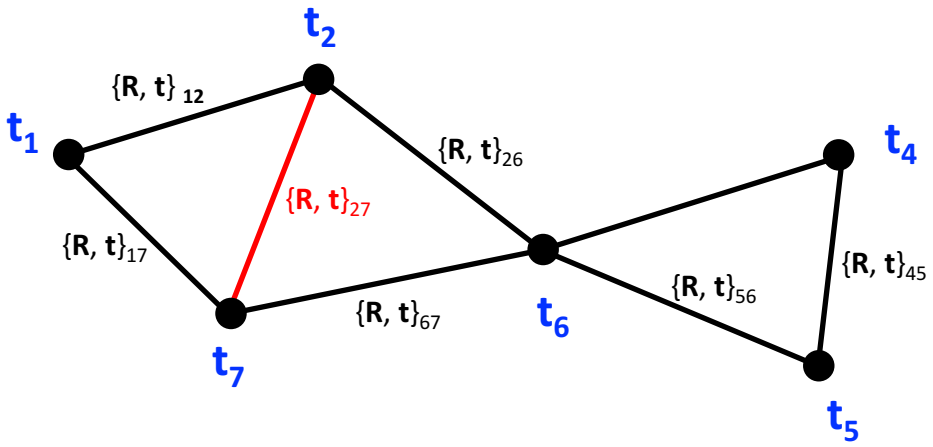


Global SfM



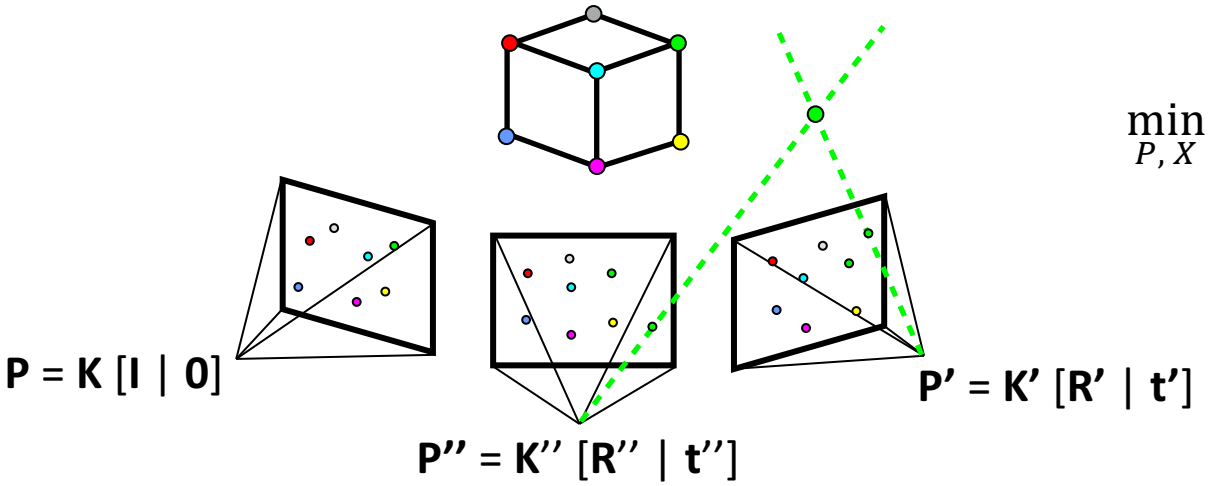
1. Estimate and filter global rotations
2. Estimate global translations: $\min_R \left\| t_{ij} - \frac{t_i - tj}{\|t_i - tj\|} \right\|$

Filter relative translations: $\left\| t_{ij} - \frac{t_i - t_j}{\|t_i - t_j\|} \right\| > \epsilon$



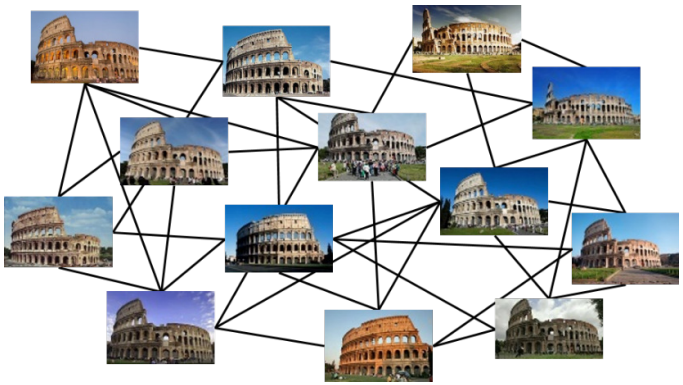
Global SfM

1. Estimate and filter global rotations
2. Estimate and filter global translations
3. Triangulate and refine with bundle adjustment



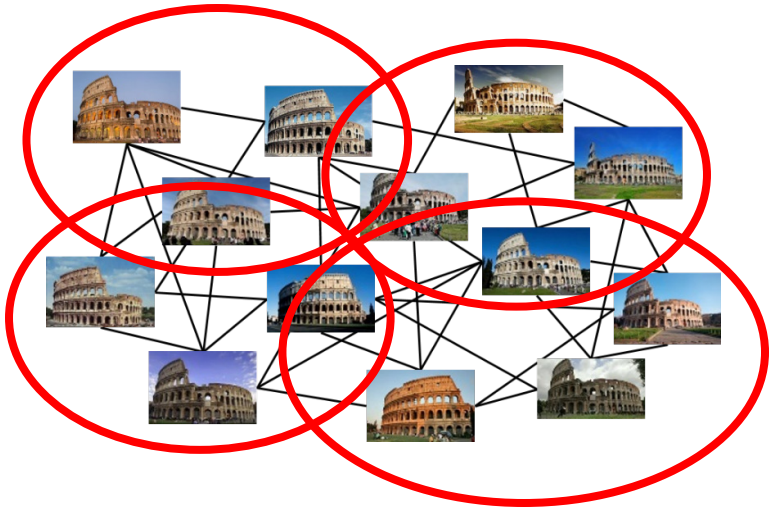
$$\min_{P,X} \|x - \pi(P,X)\|$$

Hierarchical SfM



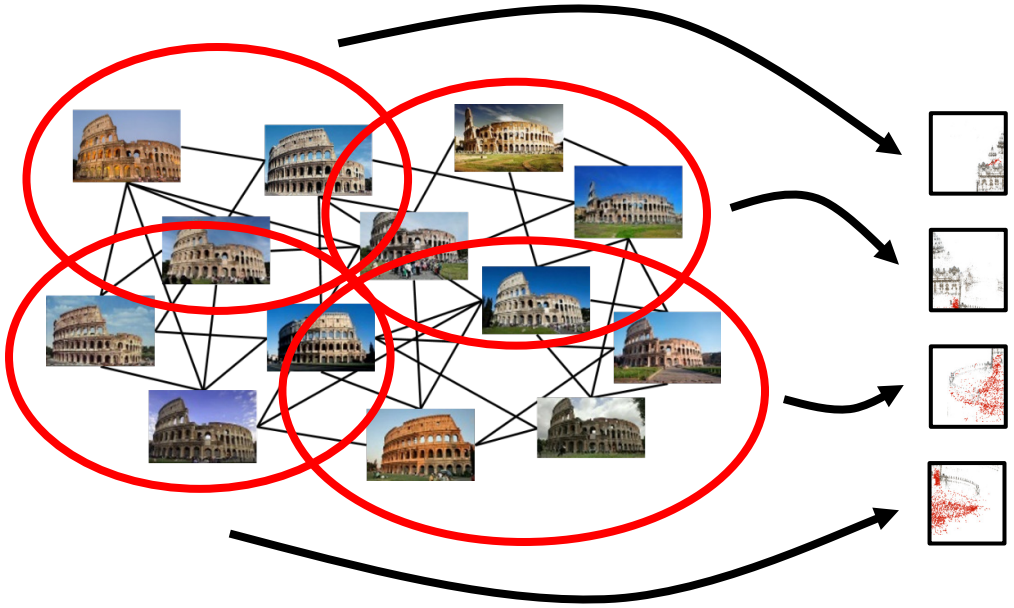
Hierarchical SfM

1. Hierarchical clustering of scene graph



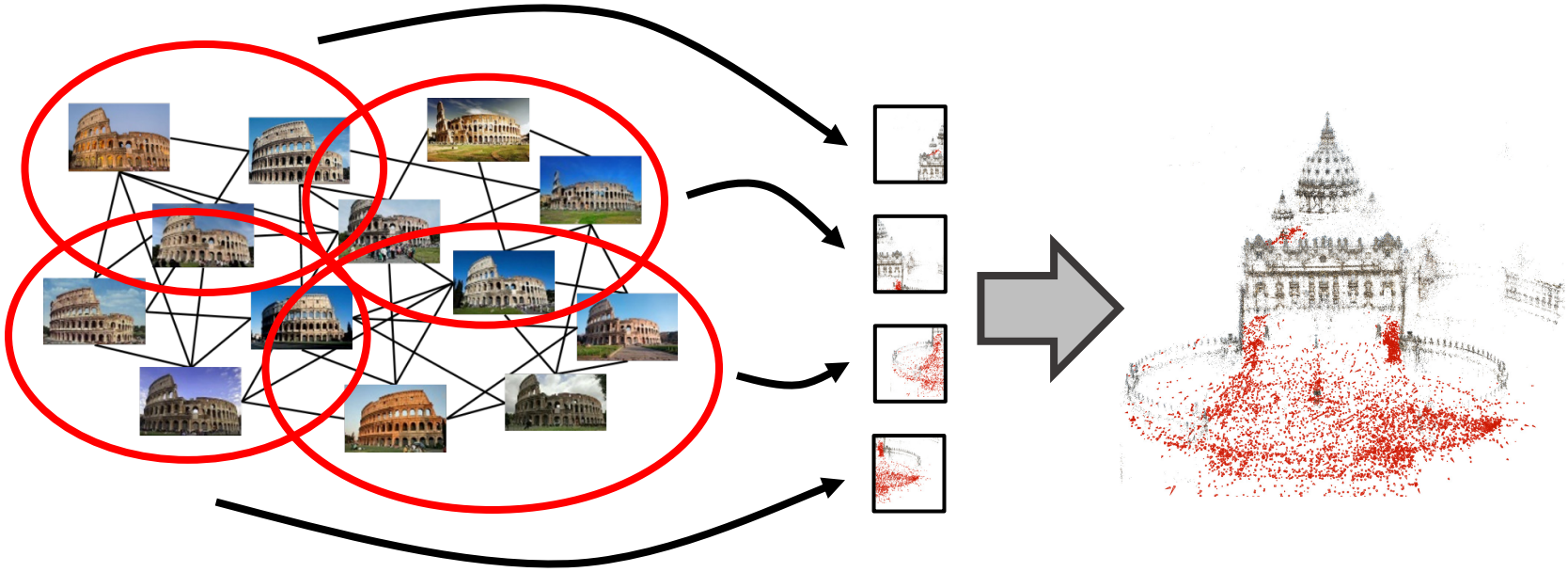
Hierarchical SfM

1. Hierarchical clustering of scene graph
2. Reconstruct clusters independently (using incremental or global SfM)



Hierarchical SfM

1. Hierarchical clustering of scene graph
2. Reconstruct clusters independently
3. Merge clusters using similarity transformations



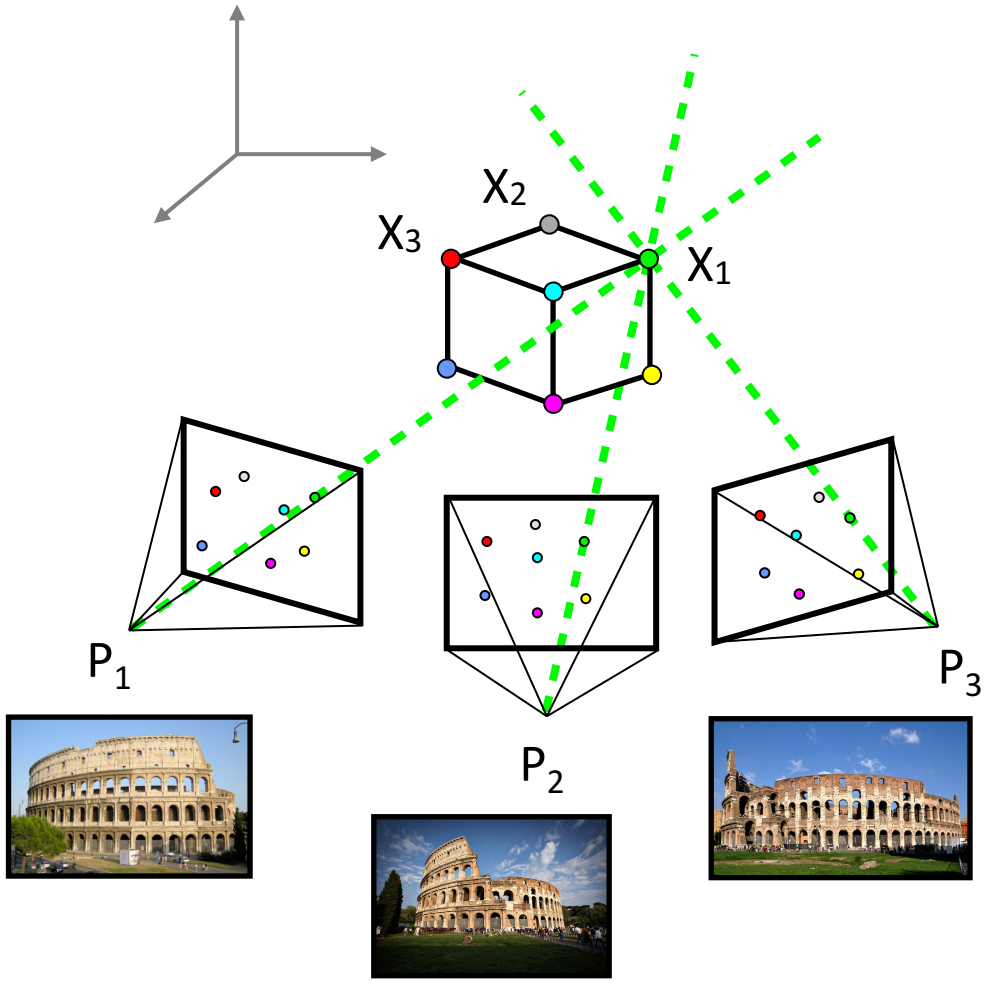
Structure-from-Motion

- Comparison

Method	Efficiency	Robustness	Accuracy
Incremental	-	++	+
Global	+	+	+
Hierarchical	++	-	-

Structure-from-Motion

- Joint estimation of …
  - Structure $\mathbf{X}_i$
  - Cameras $\mathbf{P}_j$
- … from motion, i.e.
  - images at different viewpoints

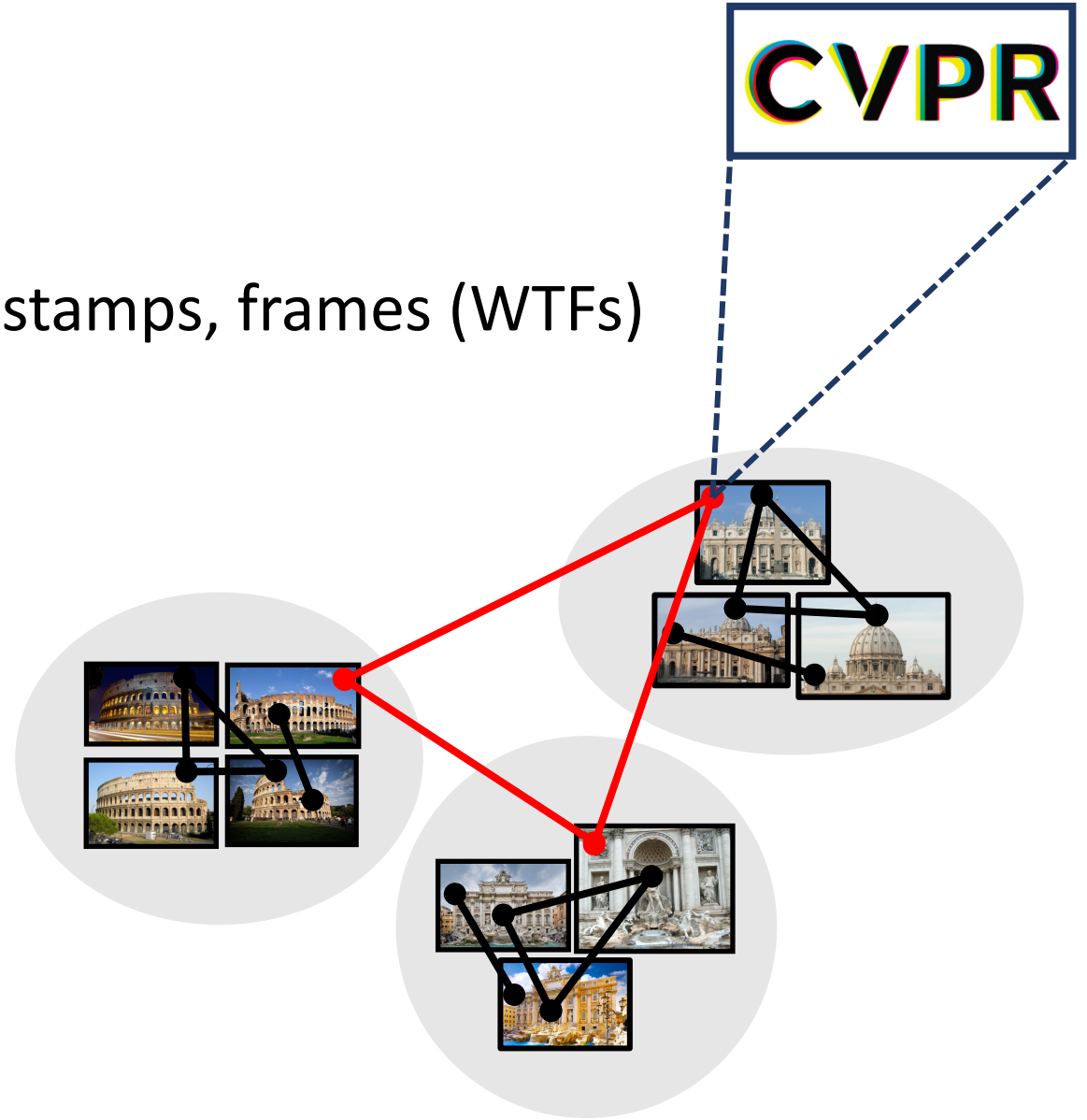


Challenges

Crowdsourced Data

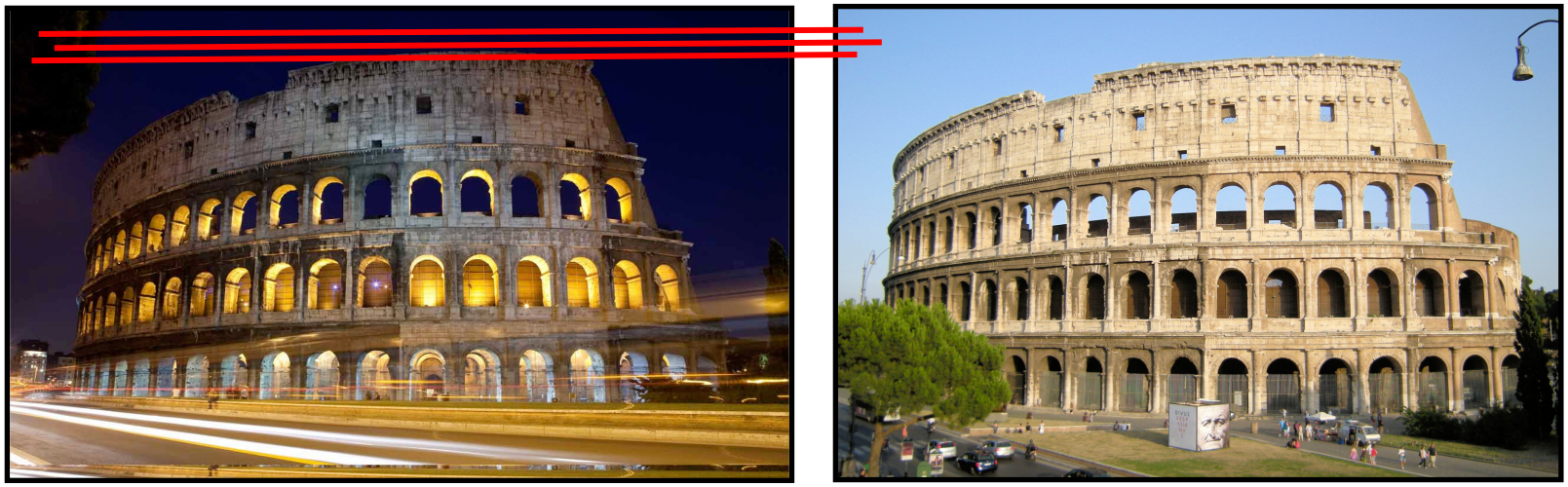
WTFs

- Watermarks, timestamps, frames (WTFs)



WTFs

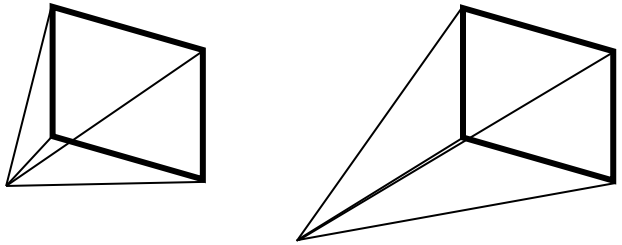
- Detect pure translation at image border



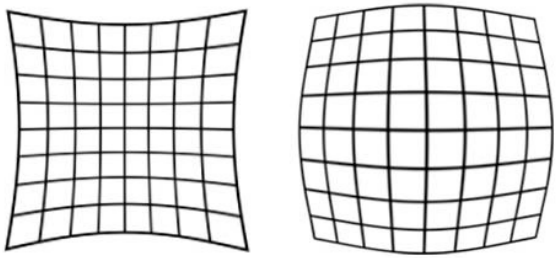
Weyand et al. 2015, “Fixing WTFs: Detecting Image Matches caused by WTFs in Internet Photos”
Heinly et al. 2015, “Reconstructing the World in Six Days”

Calibration

- Focal length unknown
  - EXIF inaccurate/missing



- Image distortion



Calibration

- Self-calibration
  - Two-view geometry estimation, e.g.
    - Stewenius et al. 2008, “A minimal solution for relative pose with unknown focal length”
    - Kukelova et al. 2015, “Efficient Solution to the Epipolar Geometry for Radially Distorted Cameras”
  - Absolute camera estimation
    - P4Pf, P4Pfr, etc.
      Bujnak, Kukelova, Pajdla et al.
    - P3.5P
      Wu 2015, “Pose Estimation with Unknown Focal Length”
  - Refinement in bundle adjustment

$$\min_{P,\,X,\,C} \|x - \pi(P, X, C)\|$$

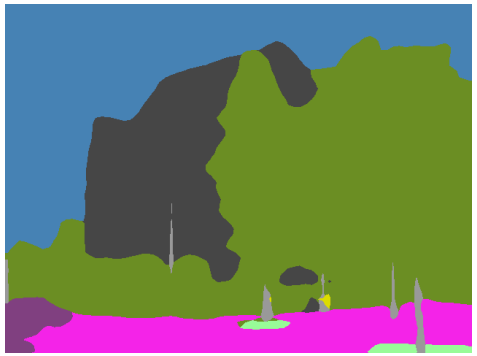
Scale ambiguity

- Inherent scale ambiguity of SfM

Scale ambiguity

- Use GPS EXIF tags for geo-registration


- Use semantics to infer scale
  (e.g., prior on size of cars, trees, people, etc.)

Dynamic objects

- Standard SfM formulation only for static objects

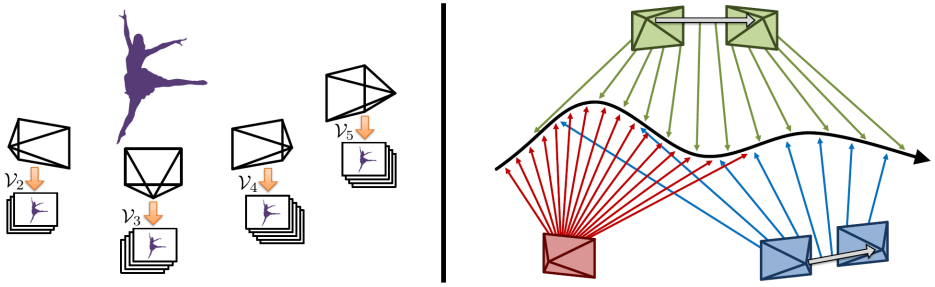
Dynamic objects

- From crowdsourced videos

  Zheng et al. 2015, “Sparse Dynamic 3D Reconstruction from Unsynchronized Videos”
  Ji et al. 2014, “3D Reconstruction of Dynamic Textures in Crowdsourced Data”

- From crowdsourced images

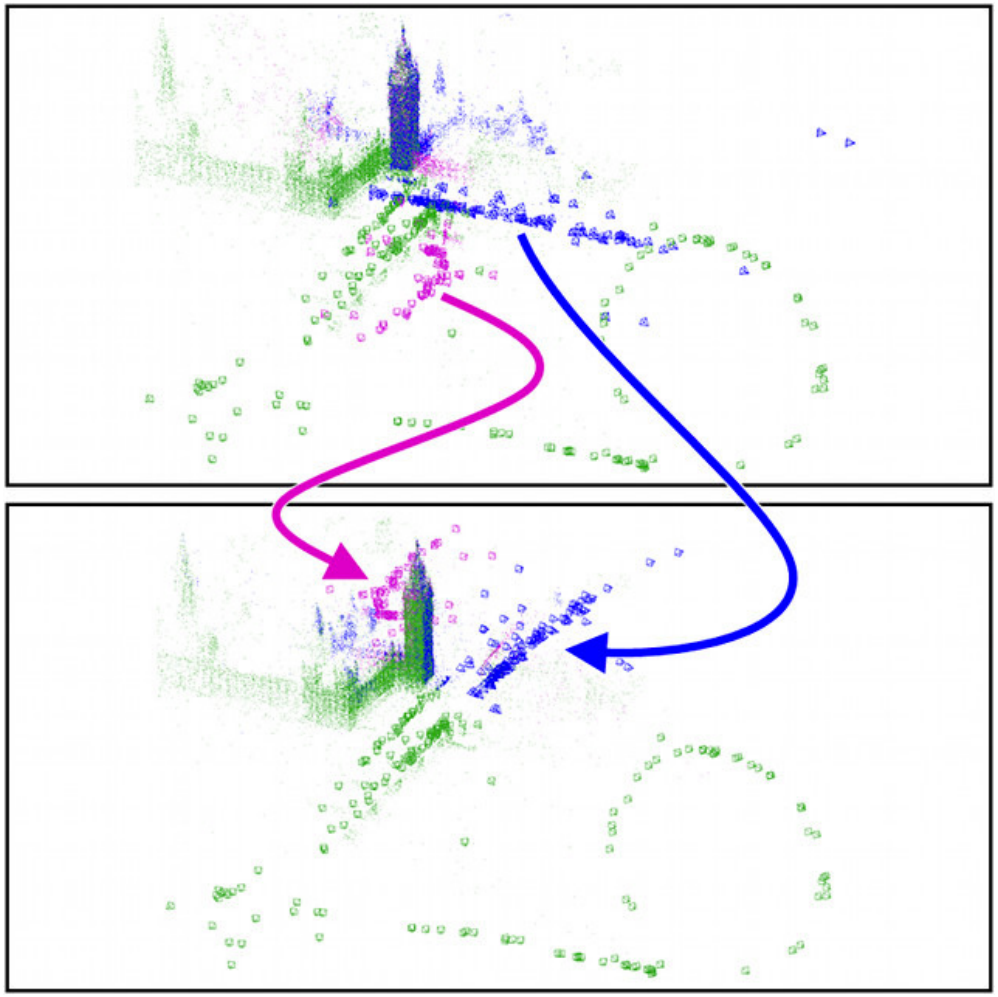
  Zheng et al. 2014, “Joint Object Class Sequencing and Trajectory Triangulation (JOST)”



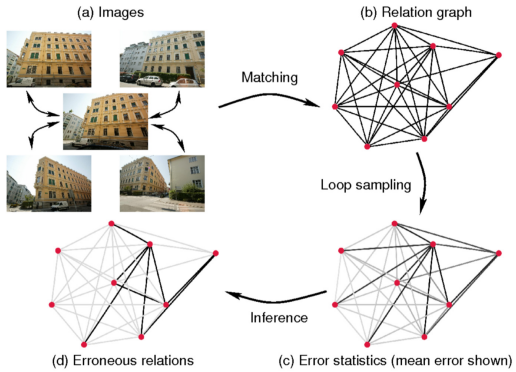
Repetitive Structures

- Symmetries in man-made structures

Repetitive Structures



Repetitive Structures

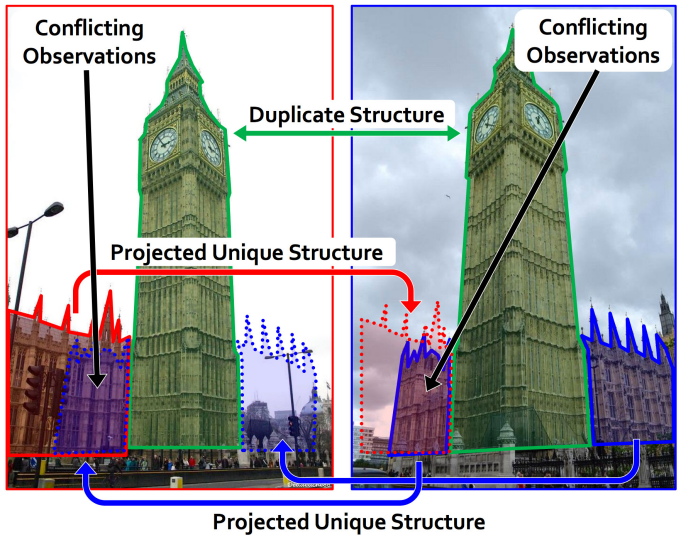


- Pre-processing
  - Remove inconsistent scene graph edges

    Zach et al. 2010, “Disambiguating visual relations using loop constraints”
    Wilson and Snavely 2013, “Network Principles for SfM: Disambiguating Repeated Structures with Local Context”

- Post-processing
  - Identify and correct duplicate structures

    Heinly et al. 2014, “Correcting for Duplicate Scene Structure in Sparse 3D Reconstruction”



Illumination Change

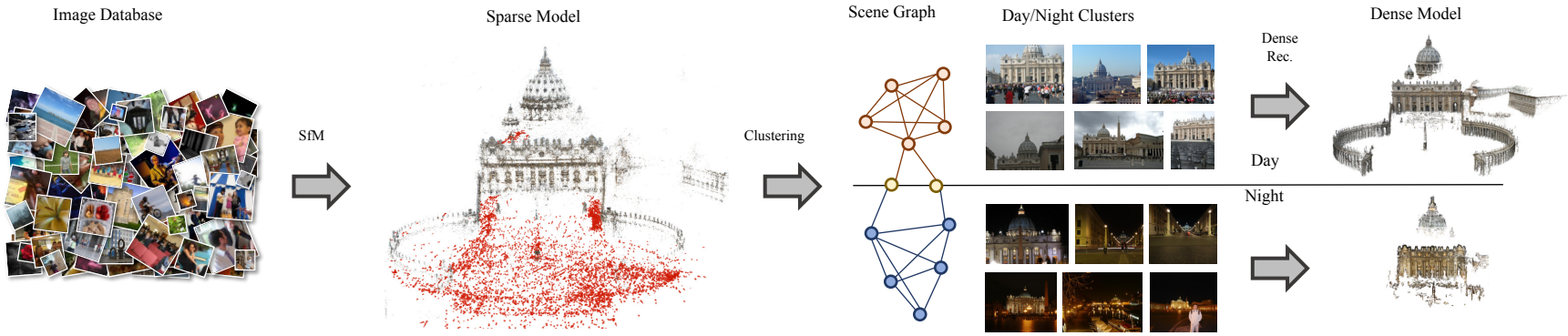
- Day-night matching difficult / not possible

Night

Day

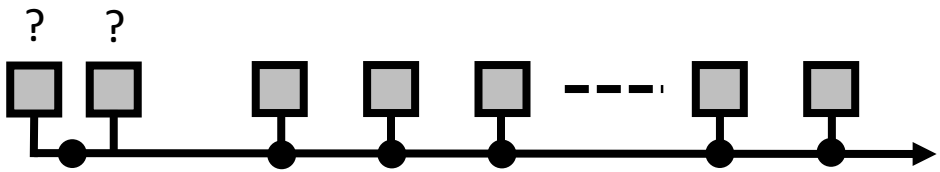
Illumination Change

- Leverage transition images during dusk / dawn



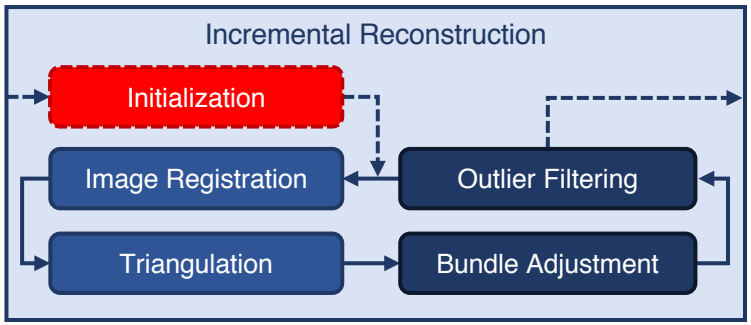
Zhou et al. 2016, “Evaluating Local Features for Day-Night Matching”
Verdie et al. 2015, “TILDE: A Temporally Invariant Learned Detector”
Radenovic et al. 2016, “From Dusk Till Dawn: Modeling in the Dark”

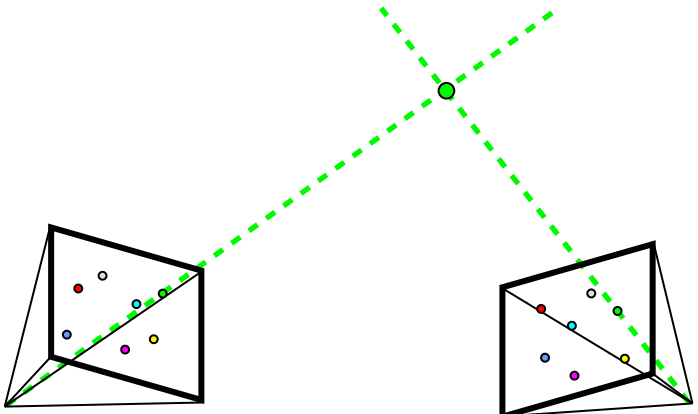
Initial View Selection



Trade-off:

- Triangulation angle
- Num. correspondences

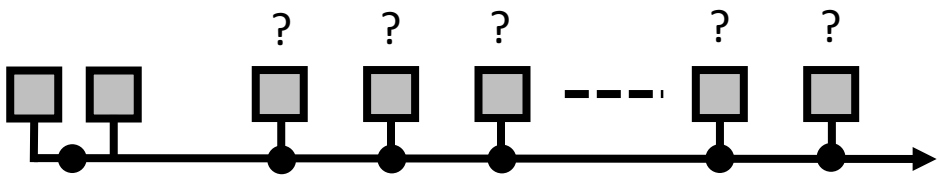


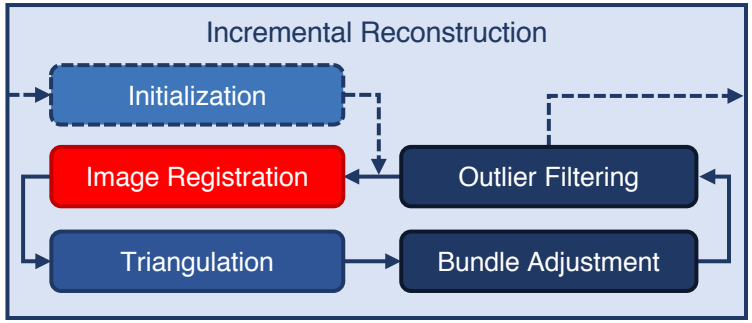


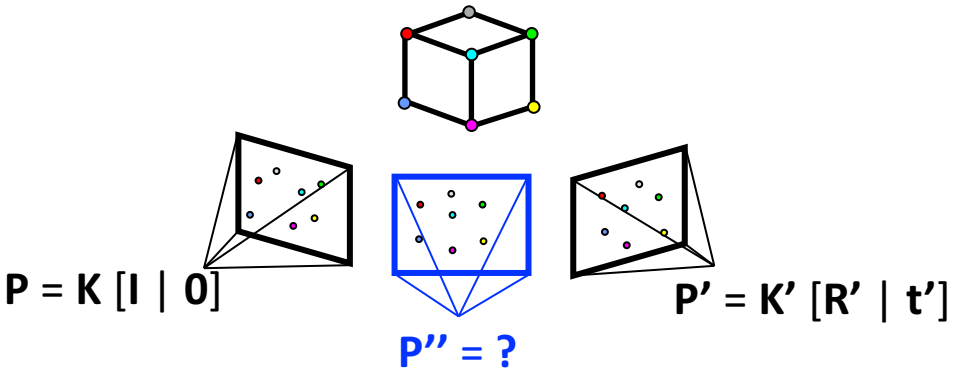
$P = K [I \mid 0]$

$P' = K' [R' \mid t']$

Next-Best View Selection

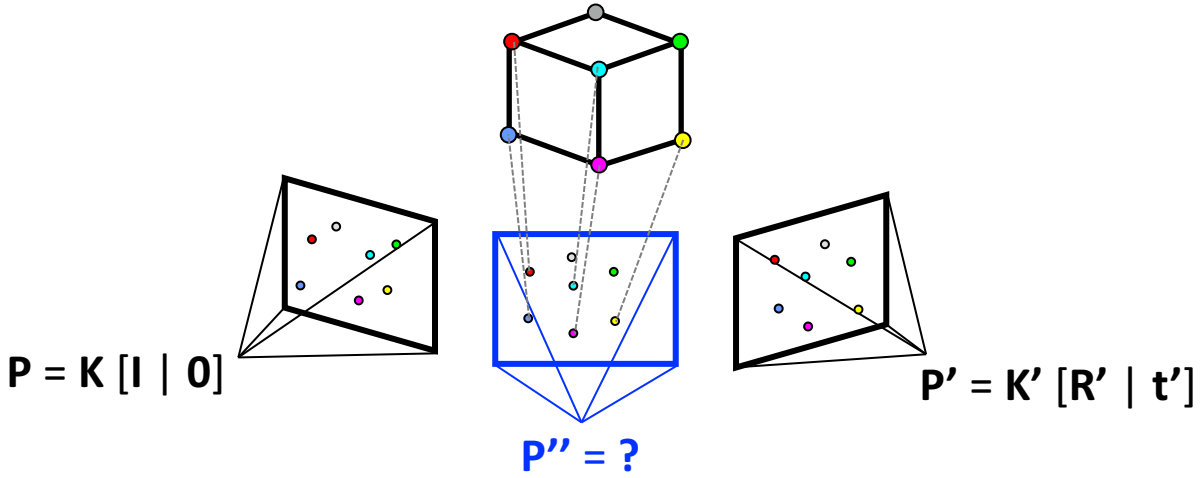


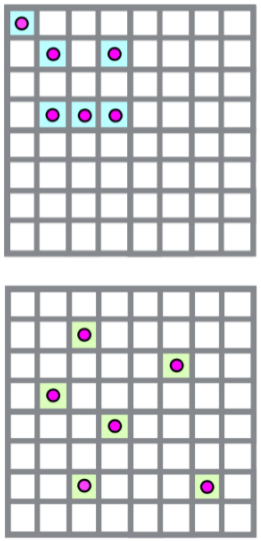




Next-Best View Selection

- Maximize number of 2D-3D correspondences

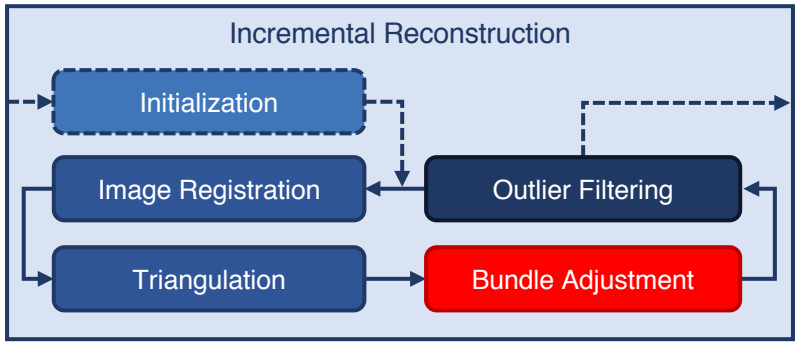


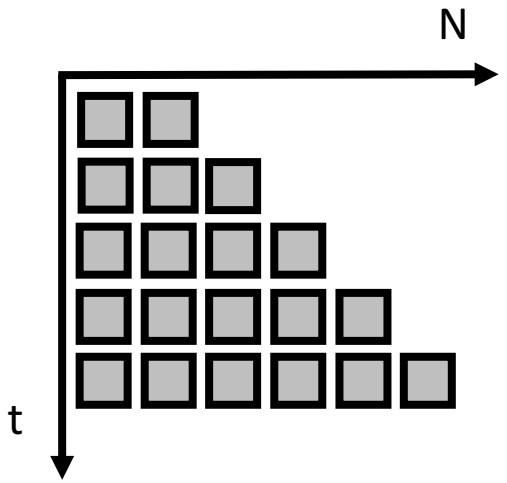


Snavely 2008, "Scene reconstruction and visualization from internet photo collections"
Schönberger and Frahm 2016, "Structure-from-Motion Revisited"

Scalability

- $O(N^3)$ complexity in bundle adjustment



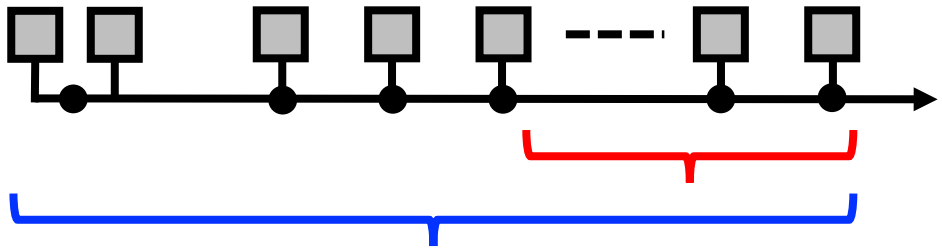


Scalability

- Efficient bundle adjustment
  1. *Local* after each camera registration
  2. *Global* after model grows by constant factor
- Inexact step algorithms (PCG)

  ➢ O(N) complexity

Agarwal et al. 2010, "Bundle-Adjustment in the Large"
Wu 2013, "Towards Linear-Time Incremental Structure-from-Motion"

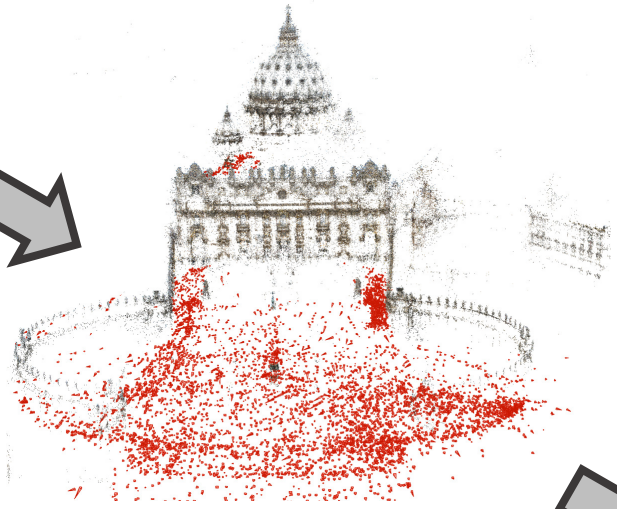


Open Source Pipeline

Unstructured Images

SfM

Sparse Model



MVS

Dense Model

COLMAP - 3D reconstruction pipeline:
https://github.com/colmap/colmap

Questions?

Structure-from-Motion Revisited

Johannes L. Schönberger, Jan-Michael Frahm

CVPR 2016

Code available at:
https://github.com/colmap/colmap