NYU TANDON SCHOOL OF ENGINEERING

AI4CE

Robot Perception

SfM

Dr. Chen Feng

cfeng@nyu.edu

ROB-GY 6203, Fall 2022

Overview

++ Marker-based SfM

* Bundle adjustment

+ Error/uncertainty propagation

++ Feature-based SfM

* COLMAP

*: know how to code (or how to use tools)
++: know how to derive (more than just the concept)
+: know the concept

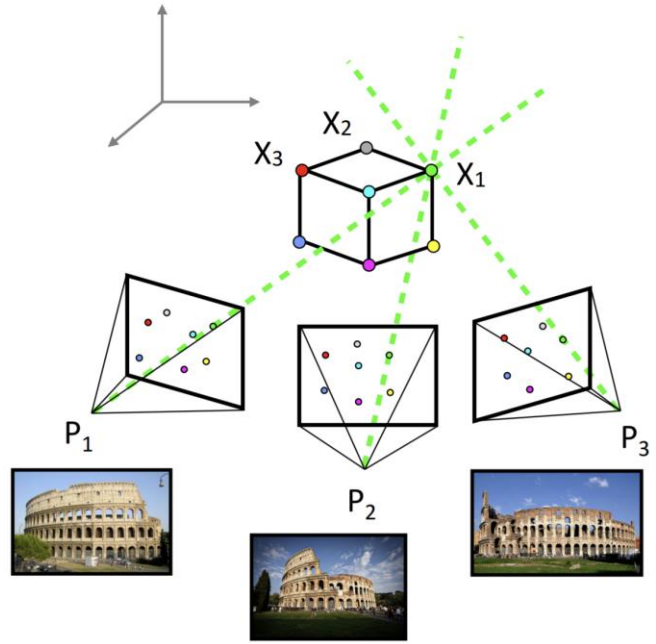
References

- Sz2022:
  - Chapter 11
- HZ2003:
  - Section 5.2, 18.1, A6.6

- https://colmap.github.io/
  - https://demuc.de/tutorials/cvpr2017/

SfM: Structure from Motion

- Joint estimation of …
  - Structure X_i
  - Cameras P_j
- … from motion, i.e.
  - images at different viewpoints

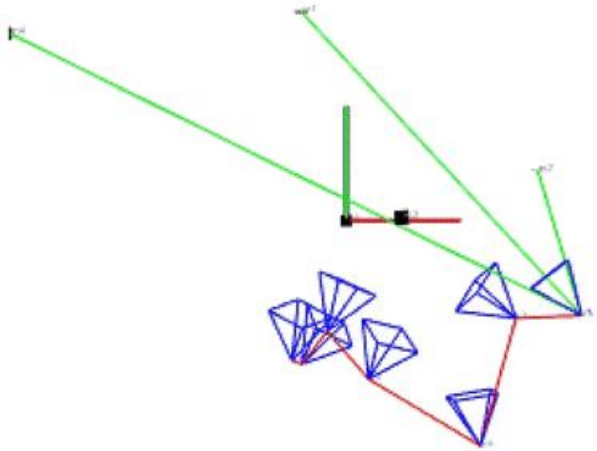


Sparse model of central Rome using 21K photos produced by COLMAP's SfM pipeline.

Image from: https://demuc.de/tutorials/cvpr2017/sparse-modeling.pdf, https://colmap.github.io/tutorial.html

Let's Start from Something Simple and Useful

- Marker-based SfM
  - Given *n* images observing *m* markers by a calibrated camera
  - Find out the *m* marker poses (the structure), and the *n* image poses (the motion)



Feng, Chen. "Camera Marker Networks for Pose Estimation and Scene Understanding in Construction Automation and Robotics." (2015).

Marker-based SfM

- Challenge: how to find the best estimation of all the marker & image poses?

Image a

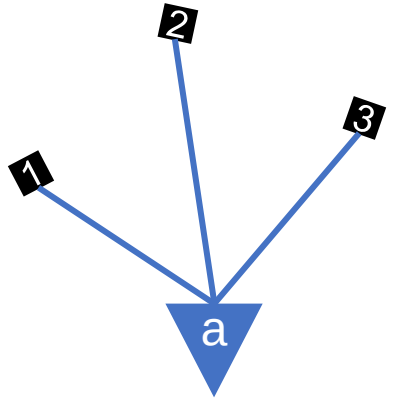


Image b

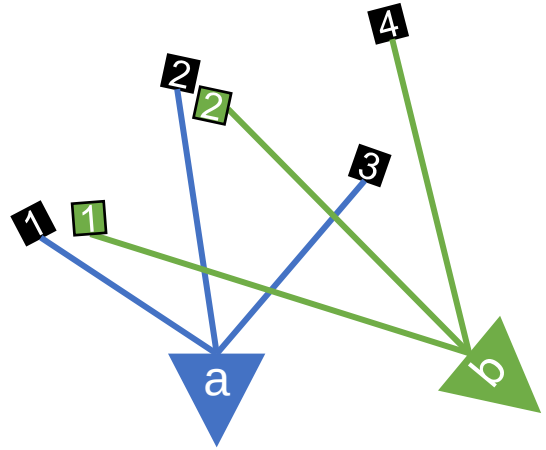
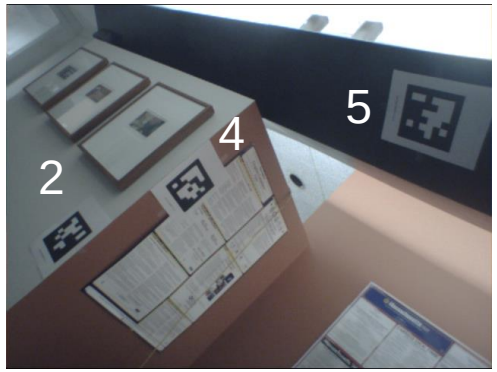
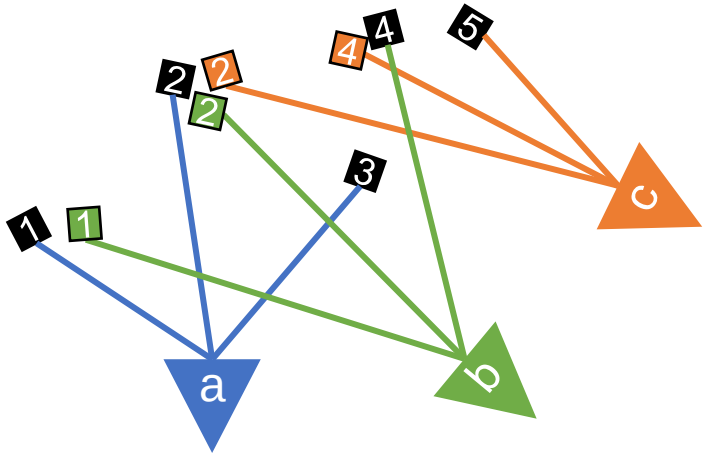


Image c

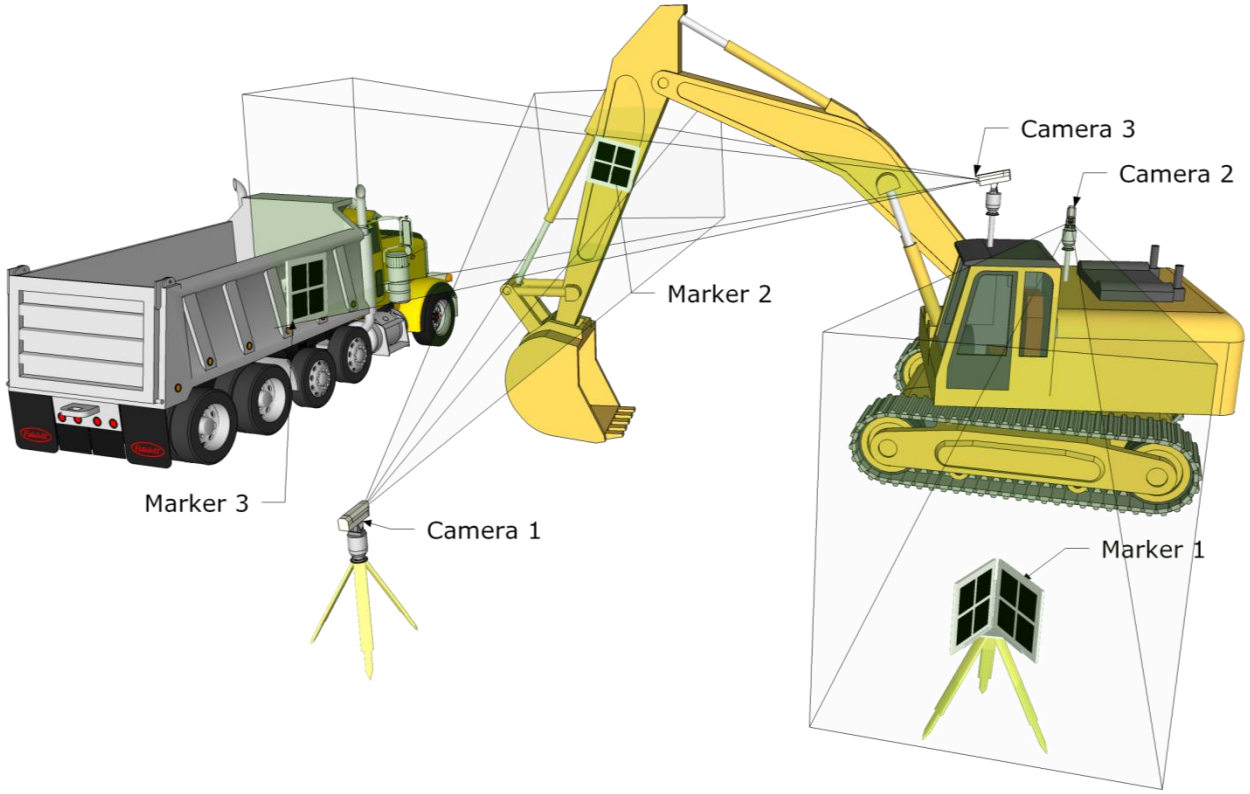




Feng, Chen. "Camera Marker Networks for Pose Estimation and Scene Understanding in Construction Automation and Robotics." (2015).

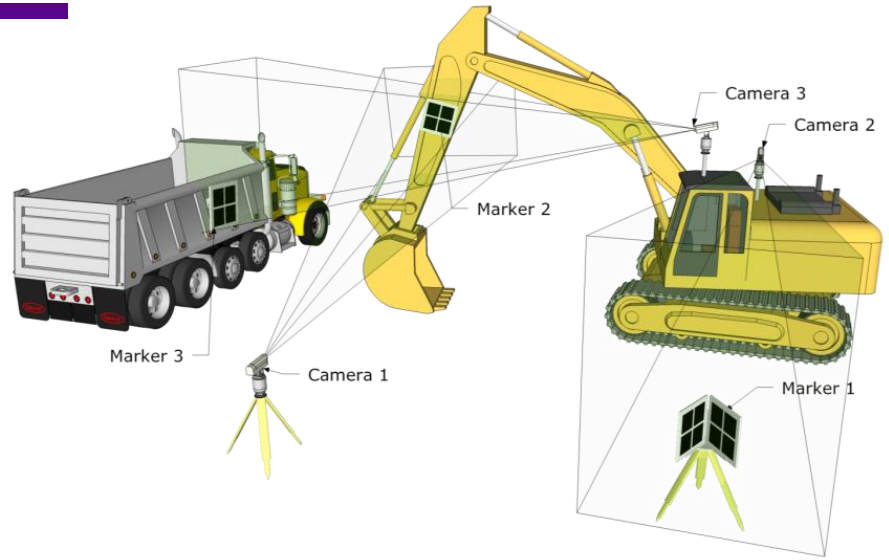
Camera Marker Networks

- Definition
  - Observation system
  - Multiple cameras or markers
  - Pose estimation of embedded object
- Multiple Cameras and Views



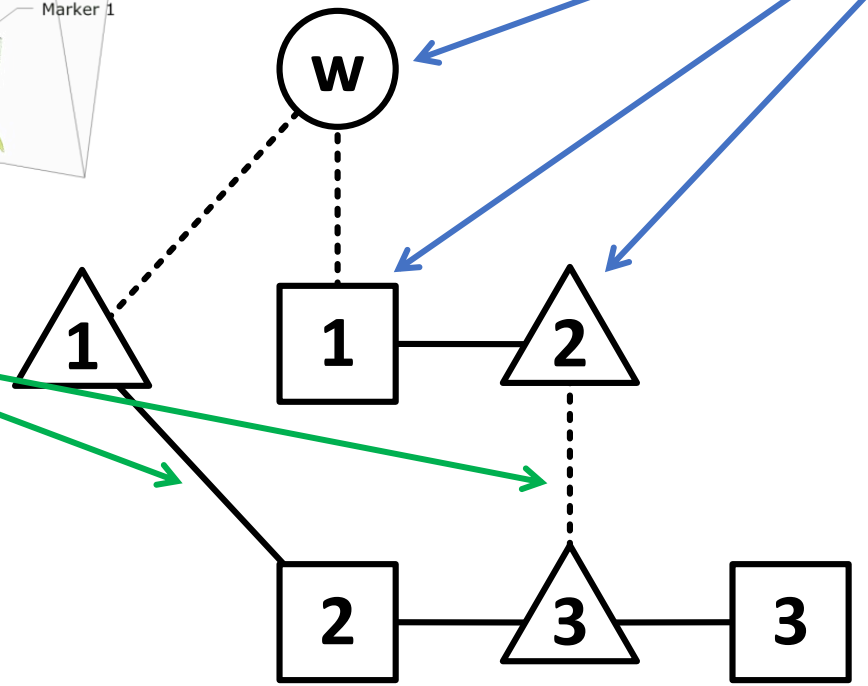
Feng, Chen, Vineet R. Kamat, and Carol C. Menassa. "Marker-Assisted Structure from Motion for 3D Environment Modeling and Object Pose Estimation." (2016).

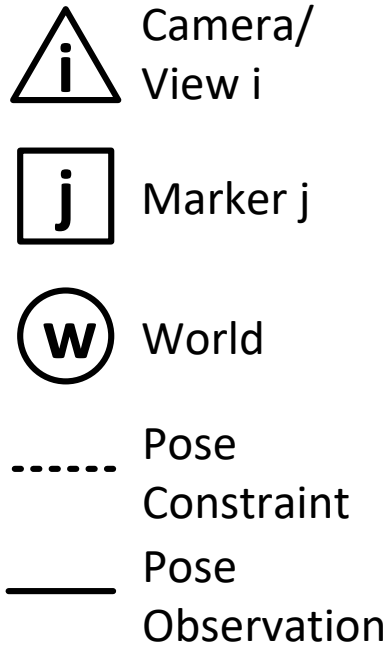
Graph Abstraction: A Unified Framework



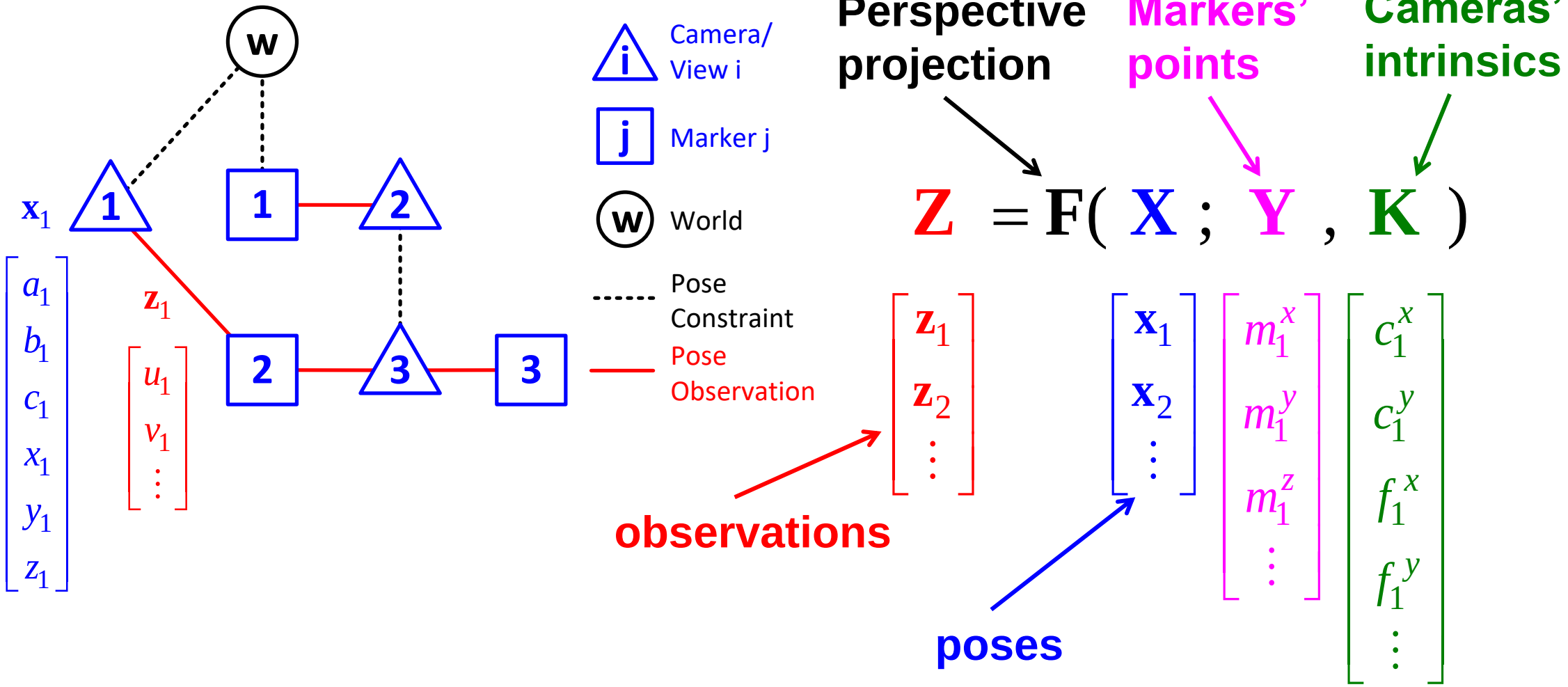
Node: pose of an object

Edge: relative relationship

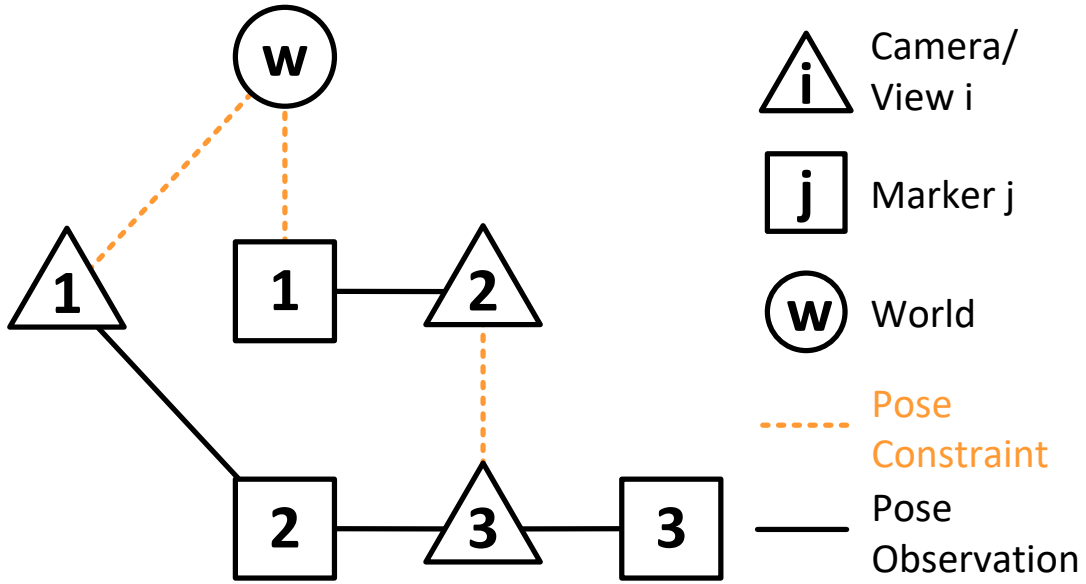




Mathematical Solution: Observation Model



Mathematical Solution: Constraint Model



$$\mathbf{G}(\mathbf{X}) = \begin{bmatrix} g_1(\mathbf{x}_{s_1}, \mathbf{x}_{e_1}) \\ \vdots \end{bmatrix} = \mathbf{0}$$

Constraint types

- Fixed-node
- Parallelism
- Perpendicularity
- Coplanarity

Mathematical Solution: Optimization

- Final solution from bundle adjustment (Triggs et al. 2000):

$$\hat{\mathbf{X}} = \underset{\mathbf{X}}{\operatorname{argmin}} \left\| \hat{\mathbf{Z}} - \mathbf{F}(\mathbf{X}; \mathbf{Y}, \mathbf{K}) \right\|^2_{\mathbf{C}_{\hat{\mathbf{Z}}}} + \|\mathbf{G}(\mathbf{X})\|^2_{\mathbf{C}_{\mathbf{K}}}$$

Actual observations

A priori covariance matrix of actual observations

Weighted constraint residuals

- Solve by Levenberg-Marquardt algorithm
  - For large problem:
    - Ceres (Agarwal and Mierle 2012)
    - g2o (Kummerle et al. 2011)
  - For small problem:
    - minimize or least_squares (scipy.optimize)
    - fminunc or lsqnonlin (matlab)

Uncertainty Propagation

- Forward (observation model $\mathbf{f}$, from given state $\mathbf{x}$):
  - For linear function
    - $f(x)=Jx$
    - $Cov[f]=JCov[x]J^T$
  - For non-linear function, approximate as
    - $f(x)\approx f(x_0)+J(x_0)(x-x_0)$
    - $Cov[f]=J(x_0)Cov[x]J(x_0)^T$
- Backward (estimate state $\mathbf{x}$, from given observation $\mathbf{f}$):
  - $f=Jx$, $P_f=C_f^{-1}$ (P_f is also known as the **information matrix, precision matrix**)
  - $x=(J^TP_fJ)^{-1}J^TP_ff$ (weighted least square estimation, minimizing **Mahalanobis distance**)
  - $Cov[x]=\{(J^TP_fJ)^{-1}J^TP_f\} * Cov[f] * \{(J^TP_fJ)^{-1}J^TP_f\}^T$ (let $Cov[f]=C_f$)
  - $= (J^TP_fJ)^{-1}J^TP_fC_f P_f^TJ(J^TP_fJ)^{-T}$
  - $= (J^TP_fJ)^{-1}J^TP_fJ(J^TP_fJ)^{-T}$
  - $= (J^TP_fJ)^{-T}$
  - $= (J^TP_fJ)^{-1}$

Uncertainty Analysis is Critical

- Reason 1: A **measure of confidence** level
  - Backward propagation of $\mathbf{C}_{\hat{\mathbf{Z}}}$, the actual observation's covariance matrix

$$\mathbf{C}_{\hat{\mathbf{X}}} = (\mathbf{J}^{\mathbf{T}}\mathbf{C}_{\hat{\mathbf{Z}}}^{-1}\mathbf{J})^{-1}$$

$$\mathbf{J} = \left.\frac{\partial \mathbf{F}}{\partial \mathbf{X}}\right|_{\hat{\mathbf{X}}}$$

Jacobian matrix of **F**
evaluated at the final solution

Estimation Uncertainty can be Visualized as Ellipsoid

https://github.com/simbaforrest/masfm

Uncertainty Analysis is Critical

- Reason 2: A **tool to evaluate stability** vs. state
  - Backward propagation at any state **X**

$$\mathbf{C}_{\mathbf{X}}(\mathbf{X}) = \left(\mathbf{J}(\mathbf{X})^{\mathbf{T}}\mathbf{C}_{\hat{\mathbf{Z}}}^{-1}\mathbf{J}(\mathbf{X})\right)^{-1}$$

    - The **theoretically best/smallest** pose estimation **uncertainty** that one can expect at **X**
    - The system stability at **X**

Optimize Configuration to Reduce Uncertainty

- Improve camera marker network designs by:

$$\hat{\mathbf{X}} = \underset{\mathbf{X}}{\operatorname{argmin}} \operatorname{Cost}(\mathbf{C}_{\mathbf{X}}(\mathbf{X}))$$

  - Trust region reflective (Coleman and Li 1996), etc.
  - MATLAB's *fmincon*
- **Intractable** for large network if using Monte Carlo simulation (Luhmann 2009)

Feature-based SfM Pipeline – Overview

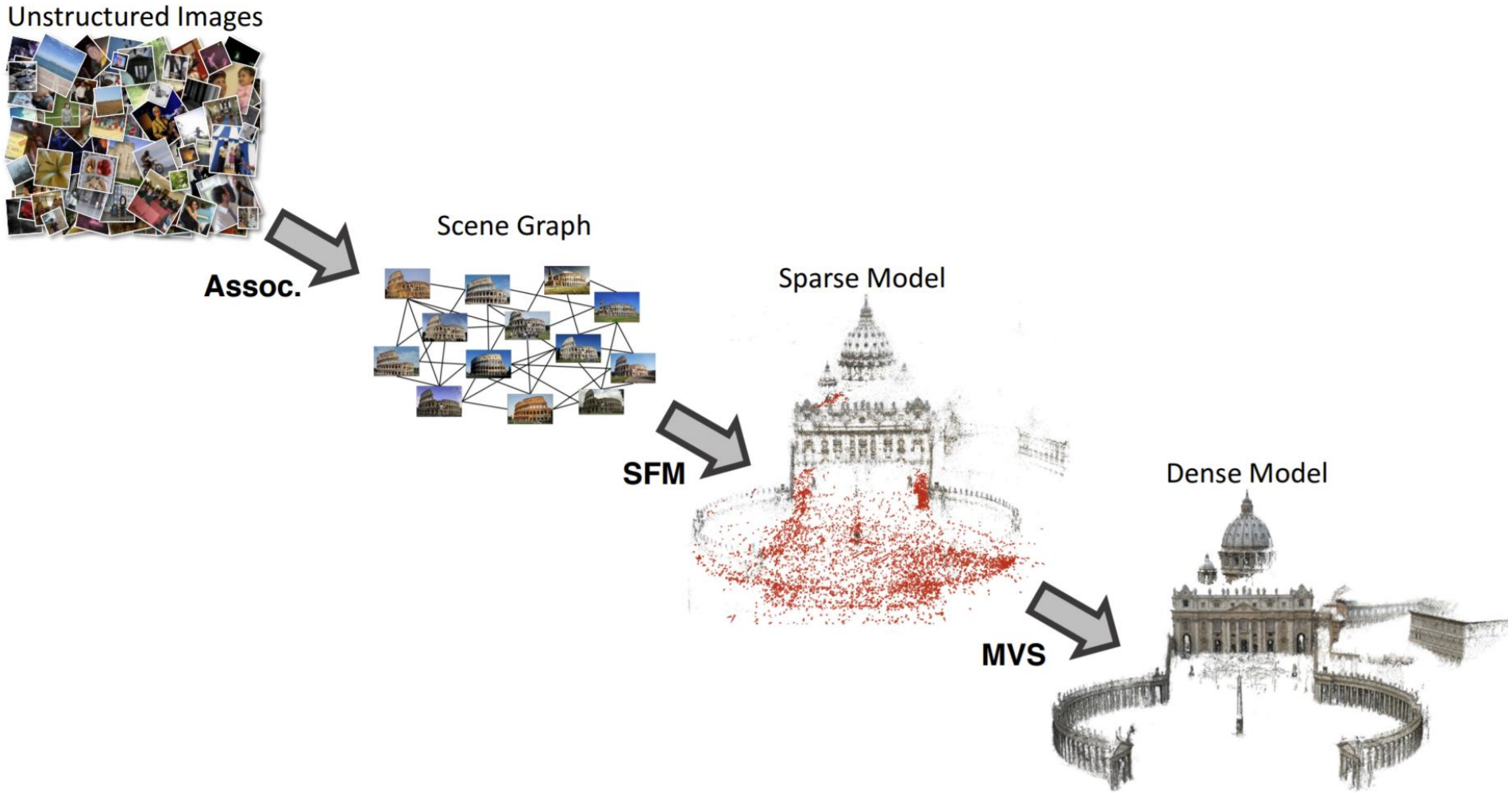
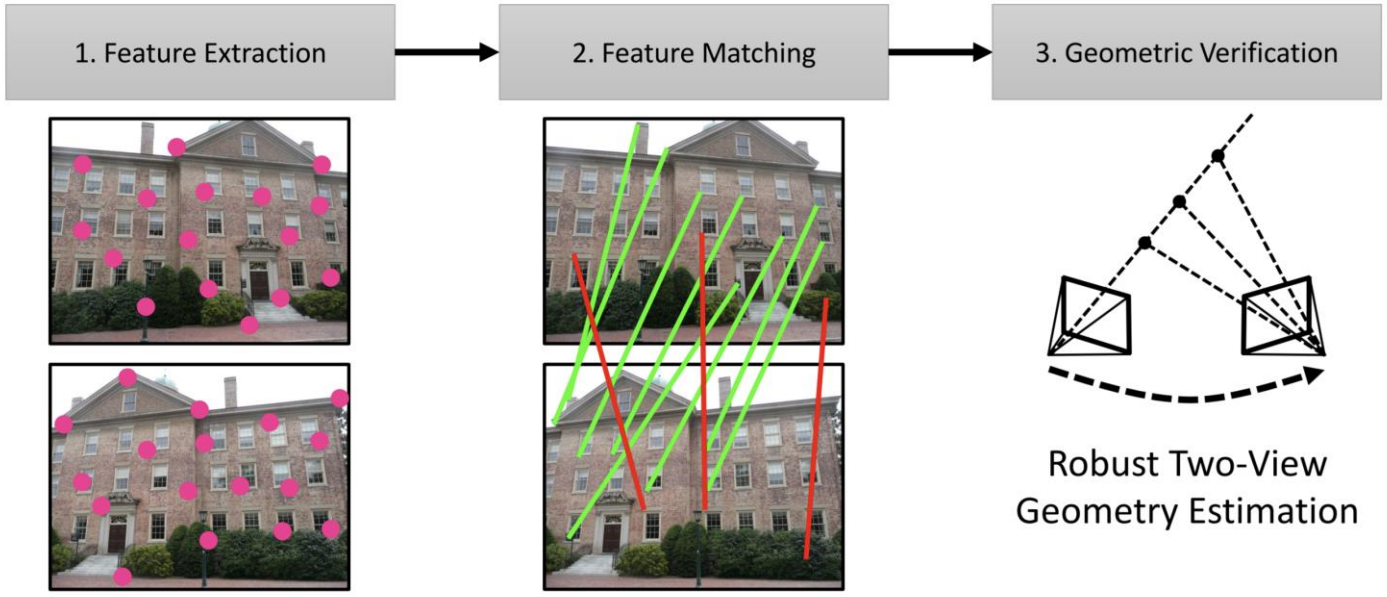


Image from: https://demuc.de/tutorials/cvpr2017/sparse-modeling.pdf

Data Association

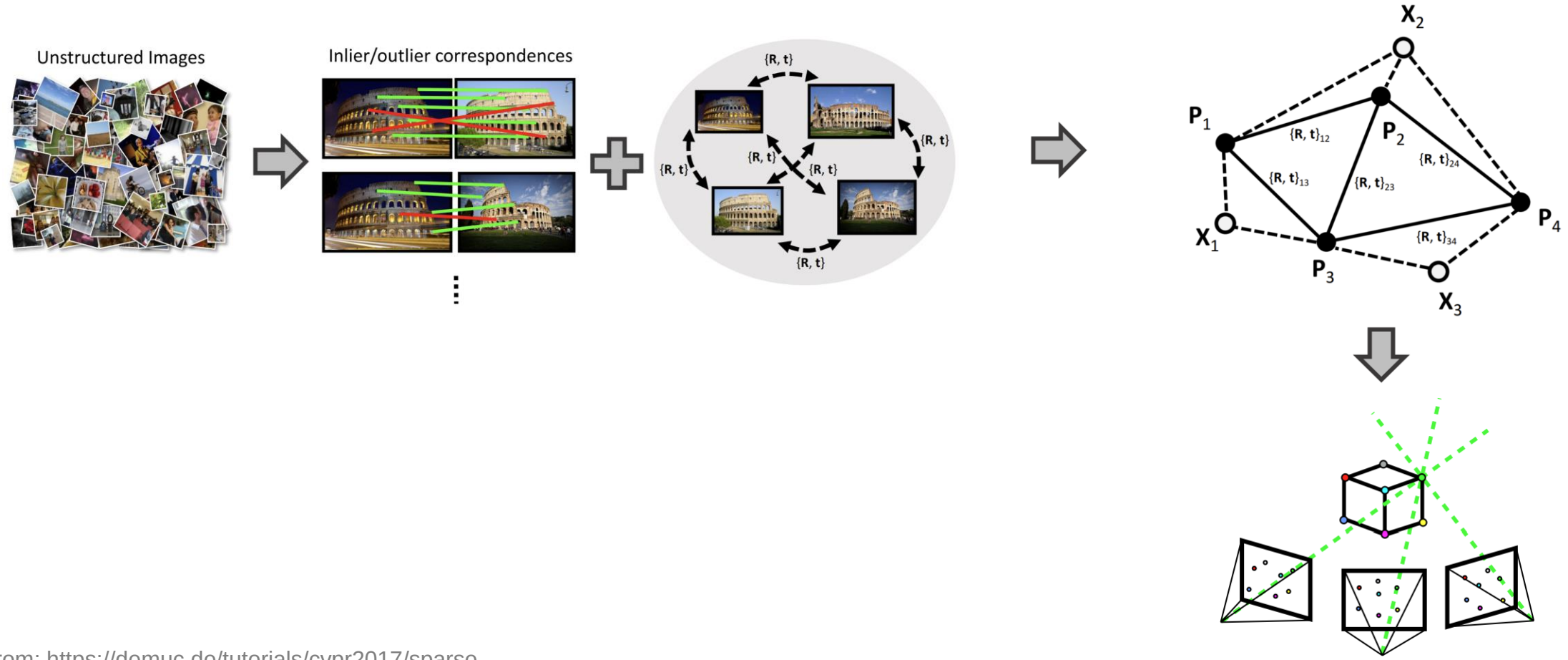


General	Planar	Panoramic
• Fundamental matrix F (*uncalibrated*) • Essential matrix E (*calibrated*)	• Homography H	• Homography H
• 7 correspondences • 5 correspondences	• 4 correspondences	• 4 correspondences

Images from: https://demuc.de/tutorials/cvpr2017/sparse-modeling.pdf

Data Association

- Data association creates a graph of cameras/views and landmark points
  - Similar to the camera marker graph



Images from: https://demuc.de/tutorials/cvpr2017/sparse-modeling.pdf

The Math Core of SfM

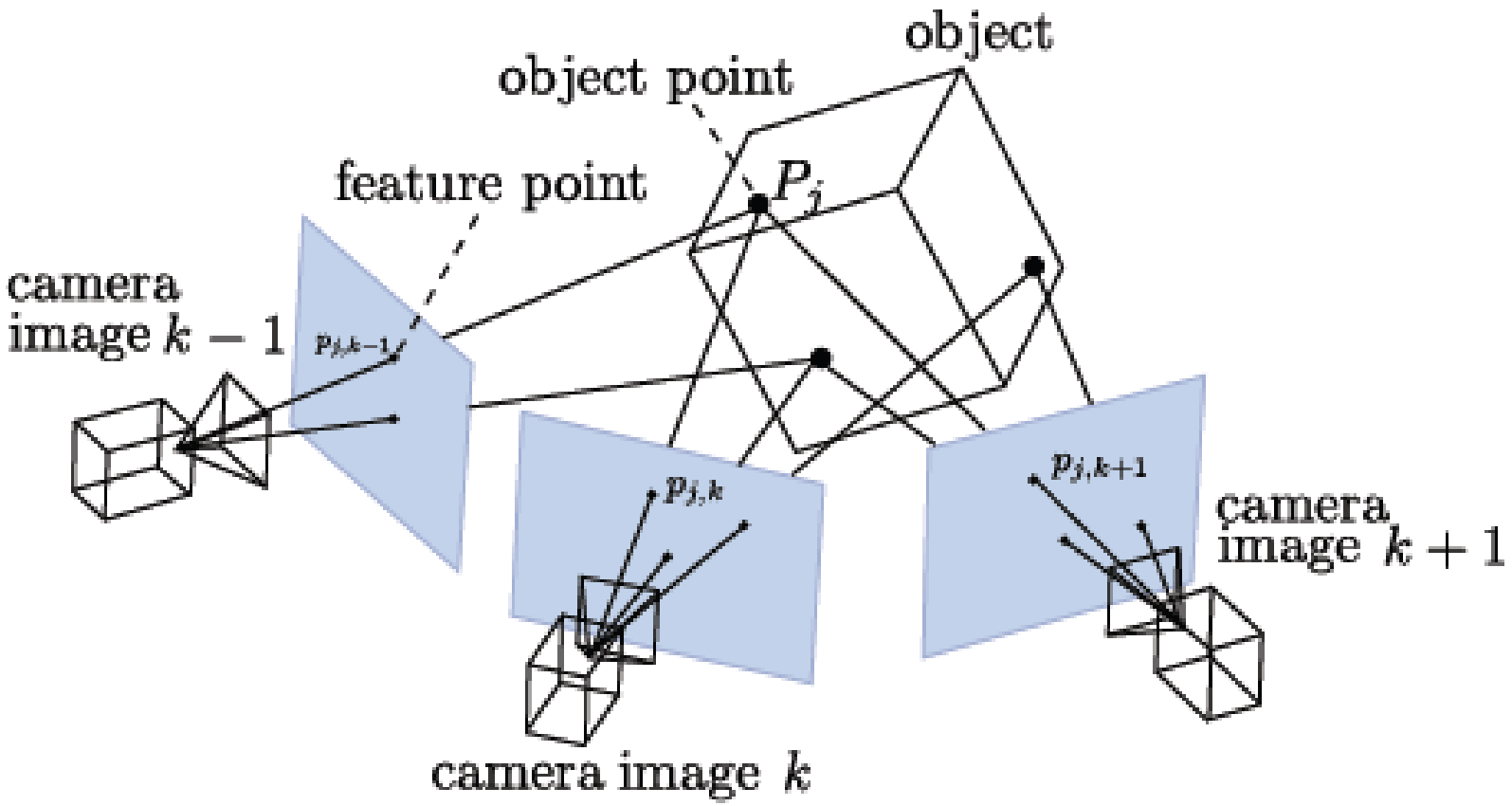
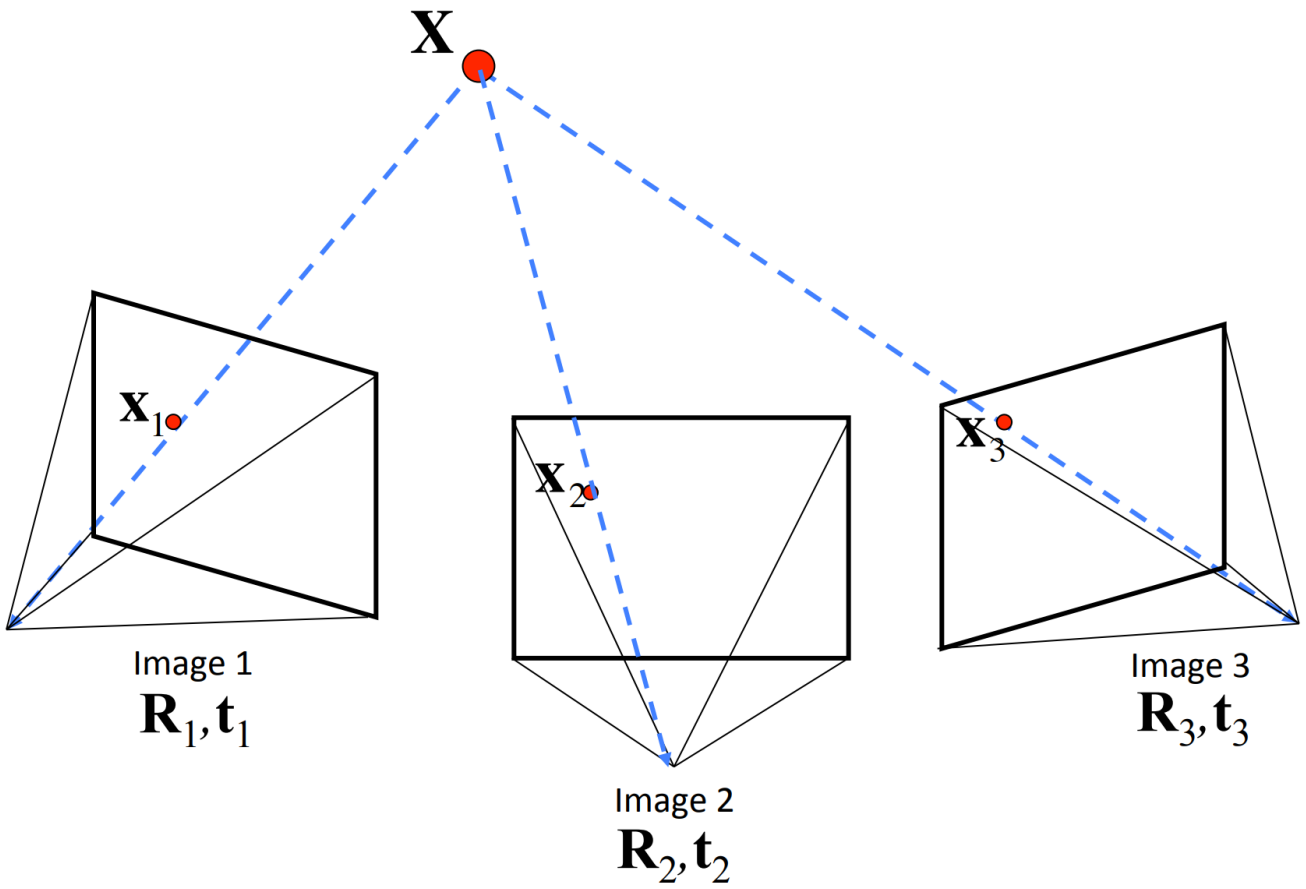


Image from:
https://openmvg.readthedocs.io/en/latest/openMVG/sfm/sfm/

The Math Model of Multiple Photos

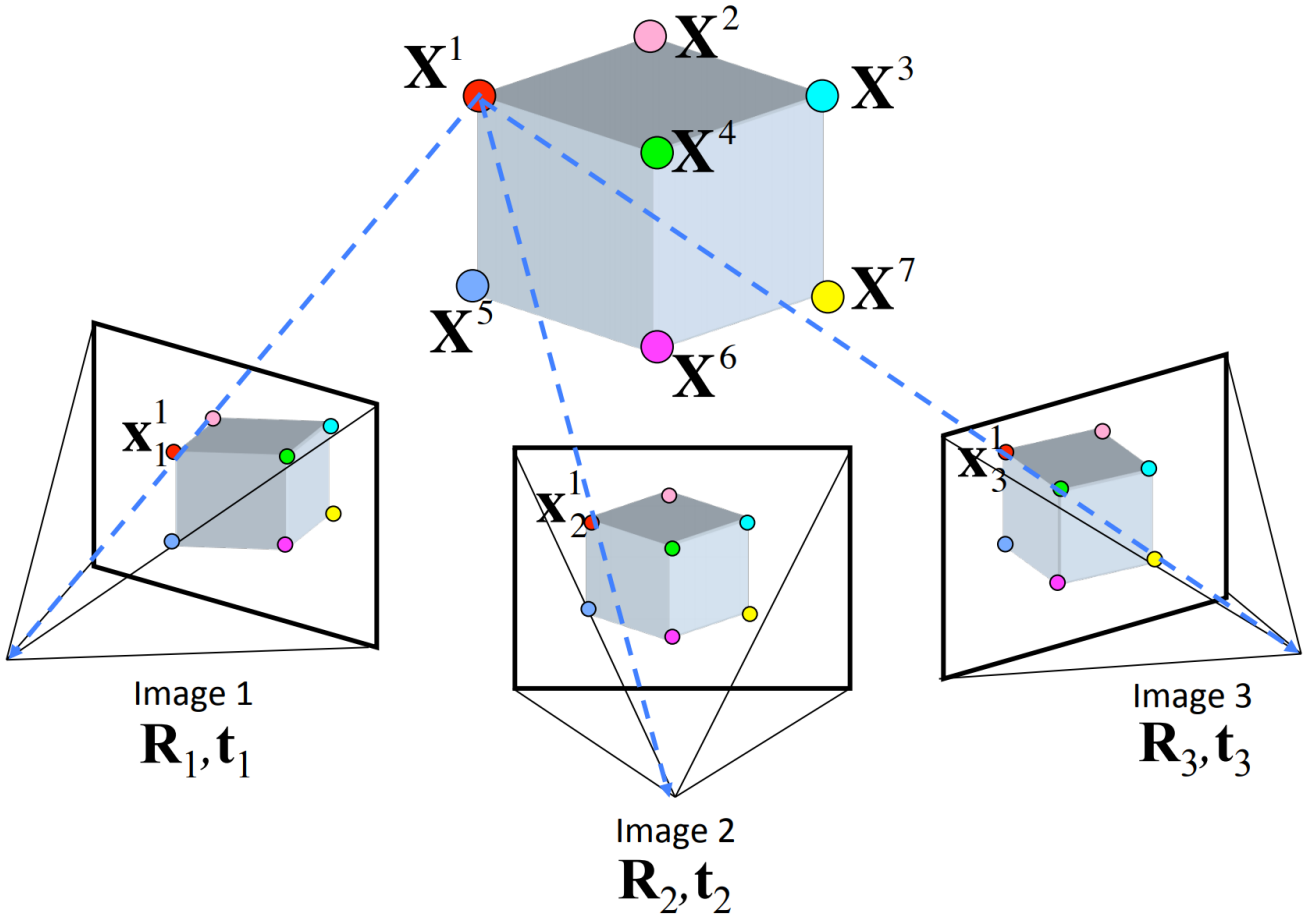


$$\mathbf{x}_1 = \mathbf{K}\left[\mathbf{R}_1 \middle| \mathbf{t}_1\right]\mathbf{X}$$

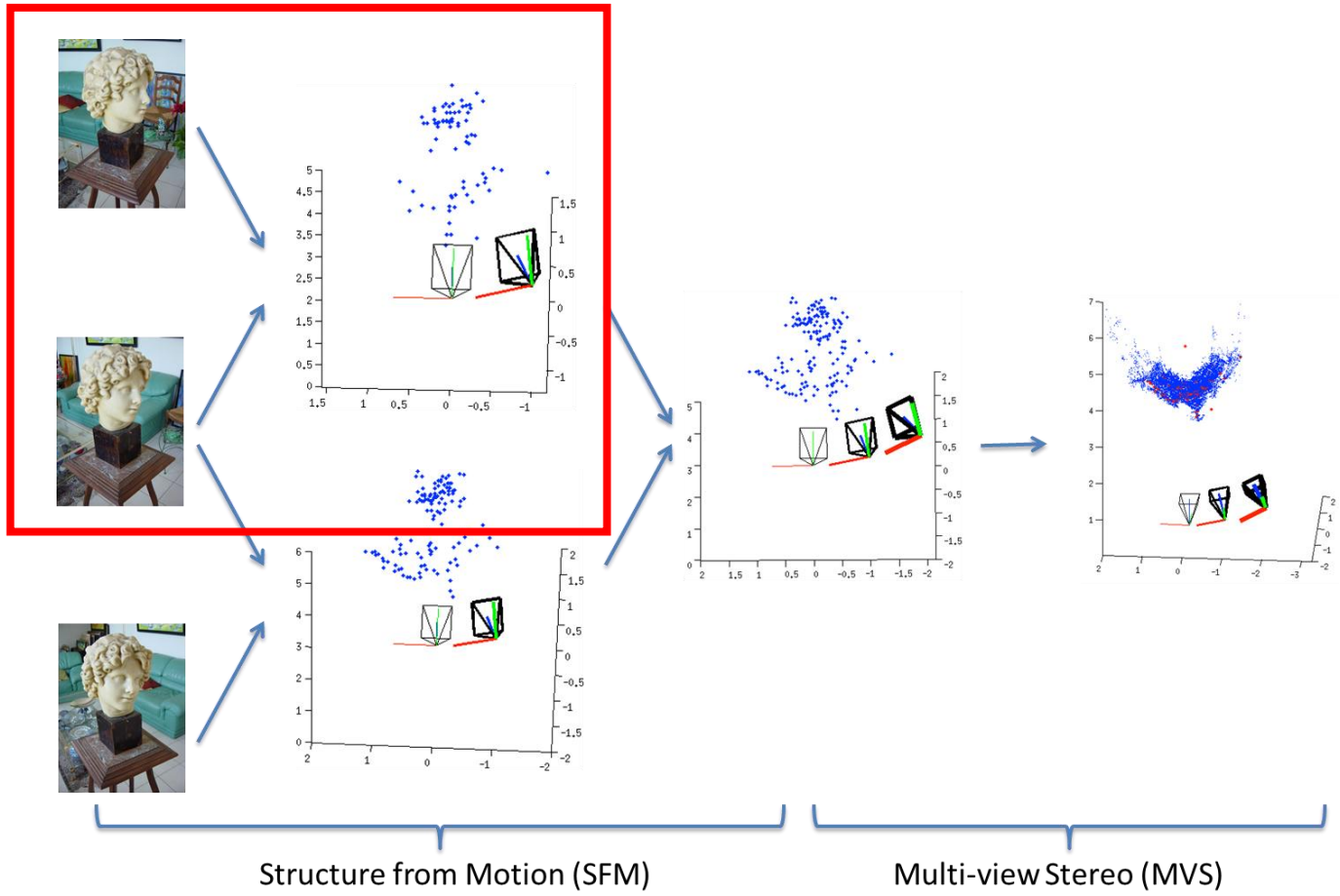
$$\mathbf{x}_2 = \mathbf{K}\left[\mathbf{R}_2 \middle| \mathbf{t}_2\right]\mathbf{X}$$

$$\mathbf{x}_3 = \mathbf{K}\left[\mathbf{R}_3 \middle| \mathbf{t}_3\right]\mathbf{X}$$

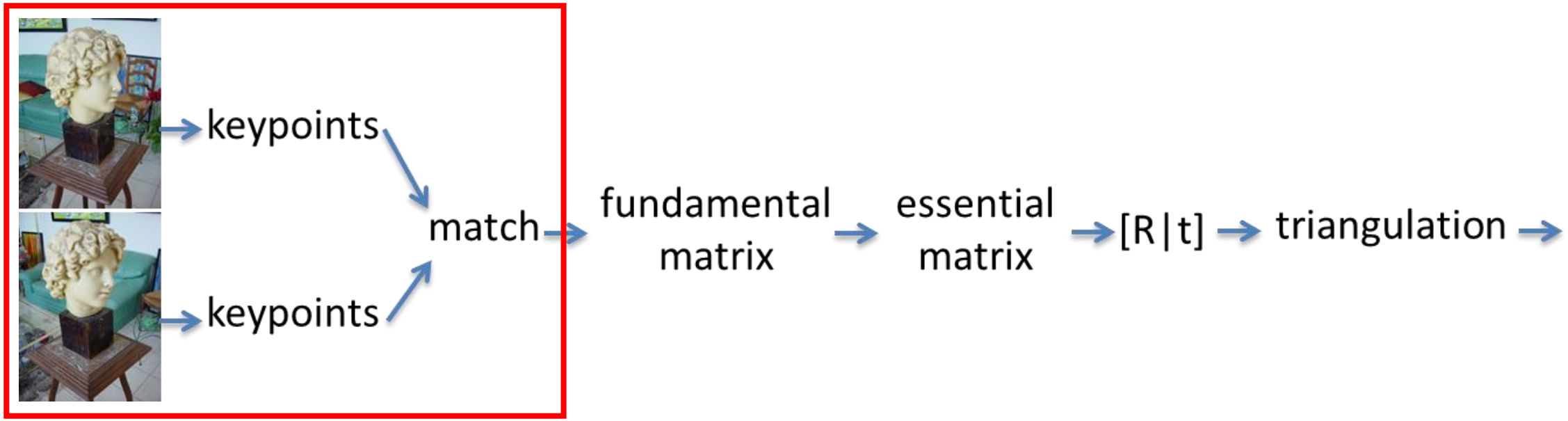
Multiple Photos of Multiple Points



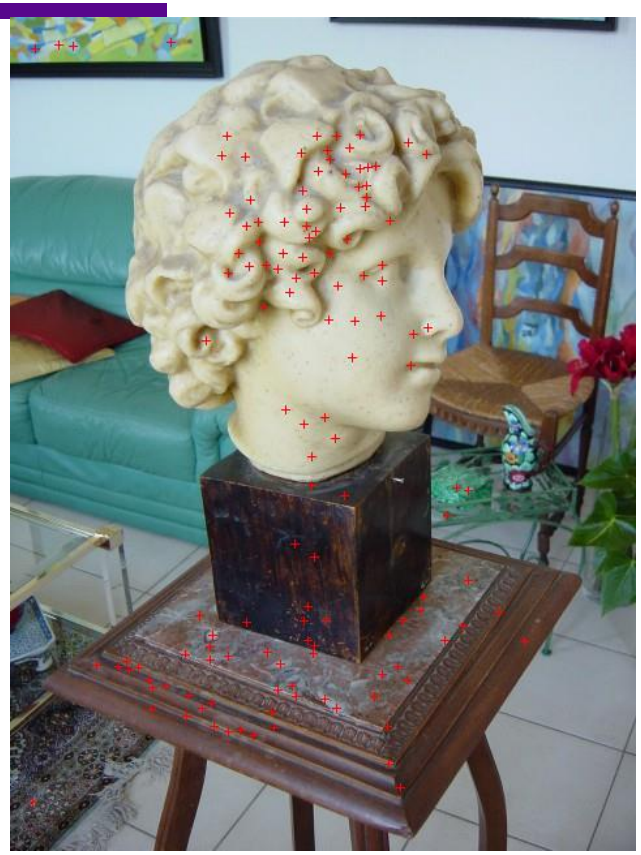
Let's Look At a Simple Example

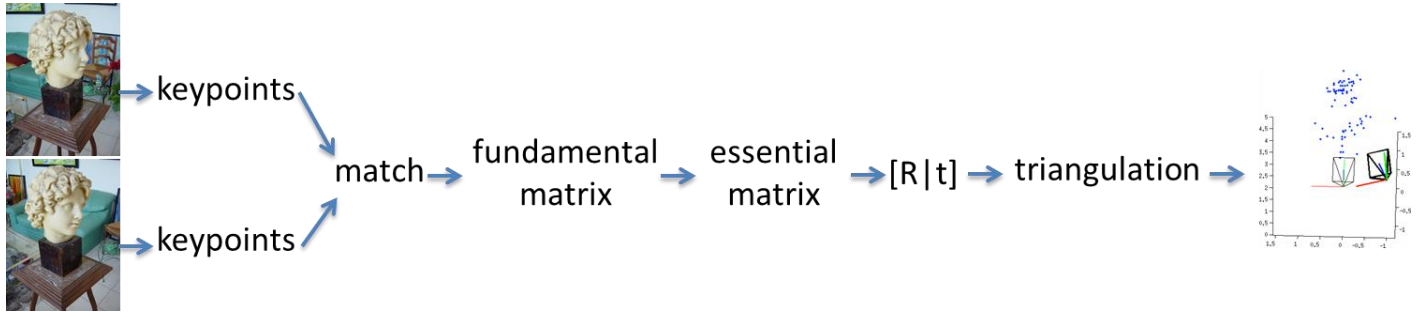


Two-view Reconstruction

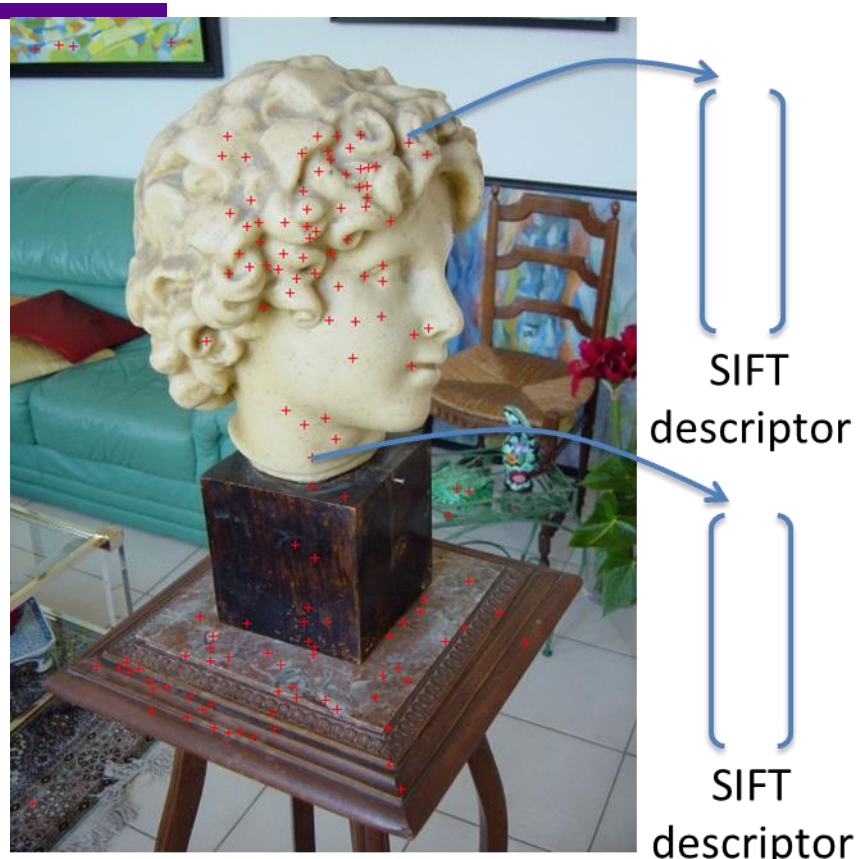


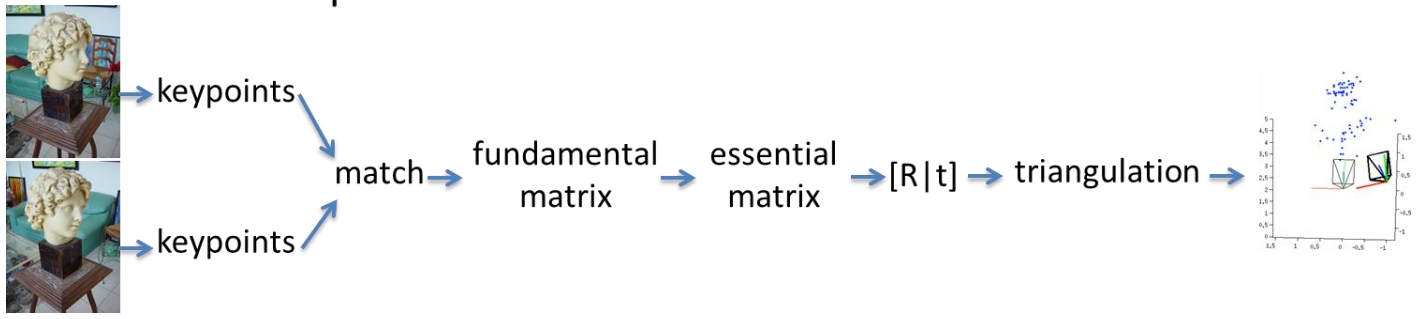
Keypoint Detection



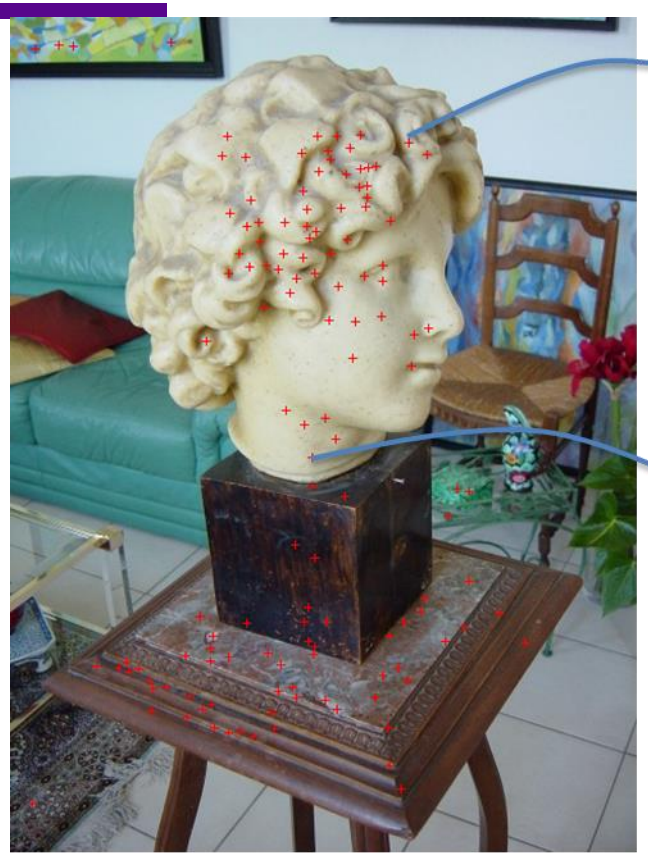


Keypoint Description





Repeat for the other image

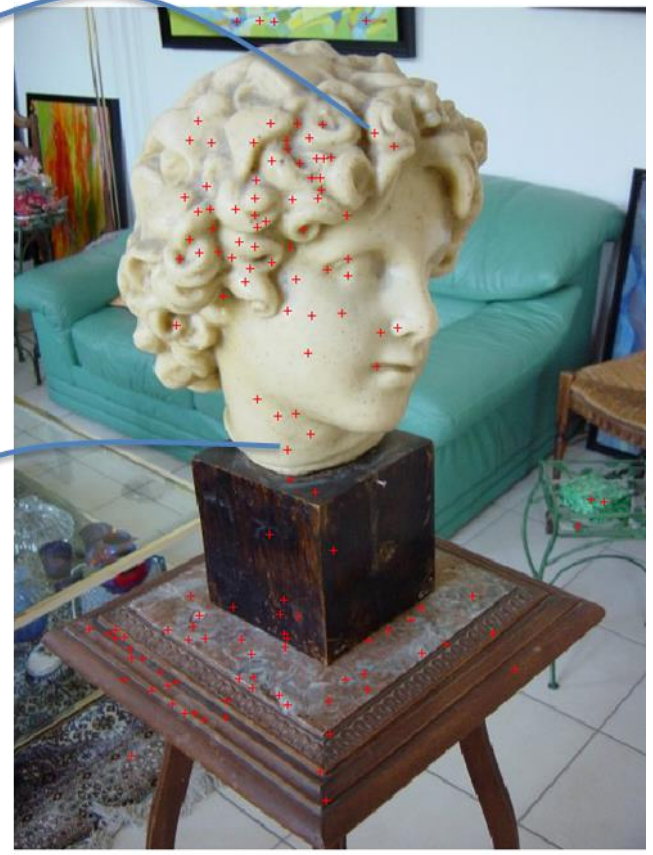


SIFT descriptor

SIFT descriptor

SIFT descriptor

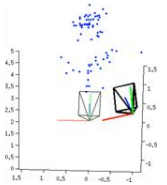
SIFT descriptor



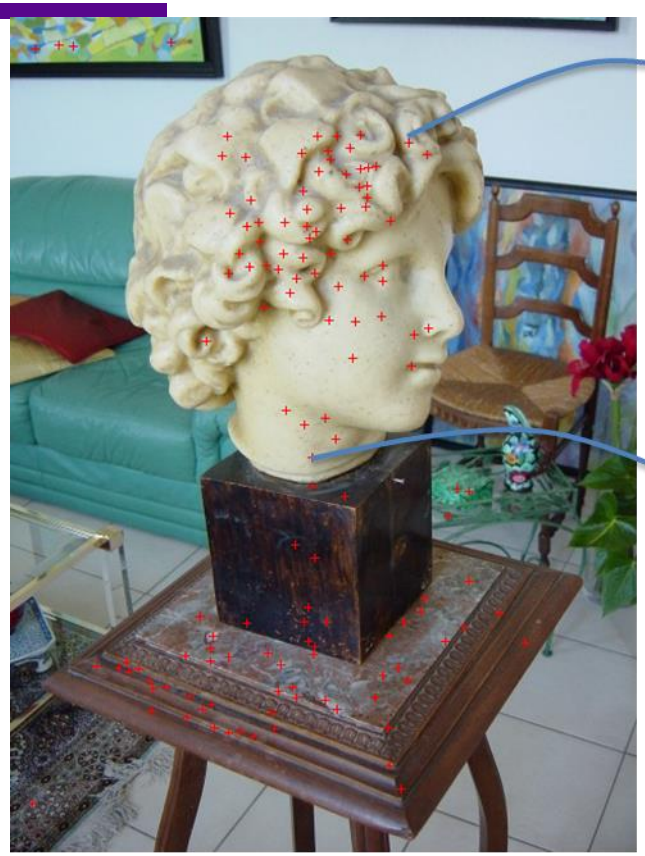
keypoints

keypoints

match → fundamental matrix → essential matrix → [R|t] → triangulation →



Keypoint Matching

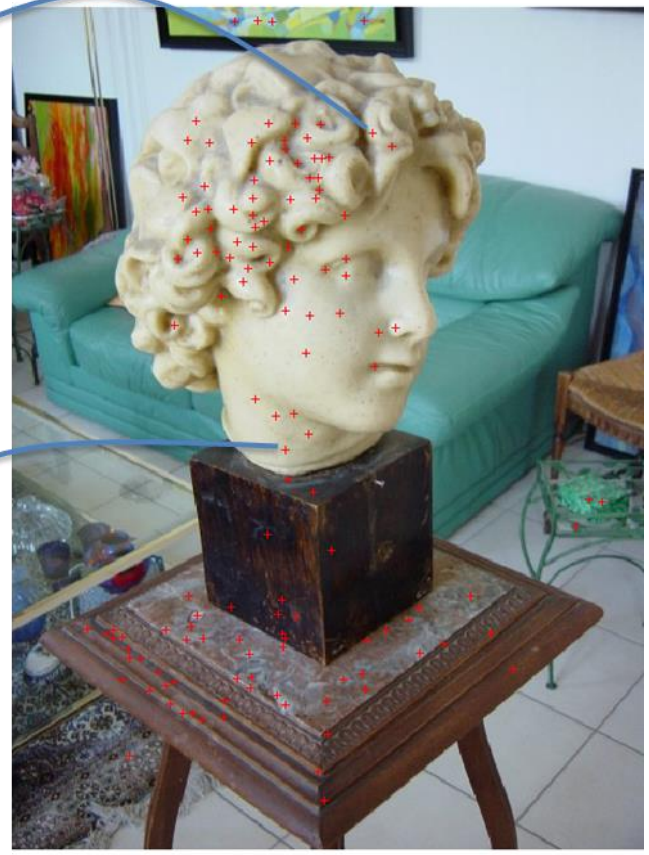


SIFT descriptor

SIFT descriptor

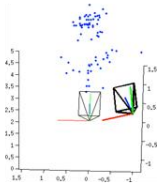
SIFT descriptor

SIFT descriptor

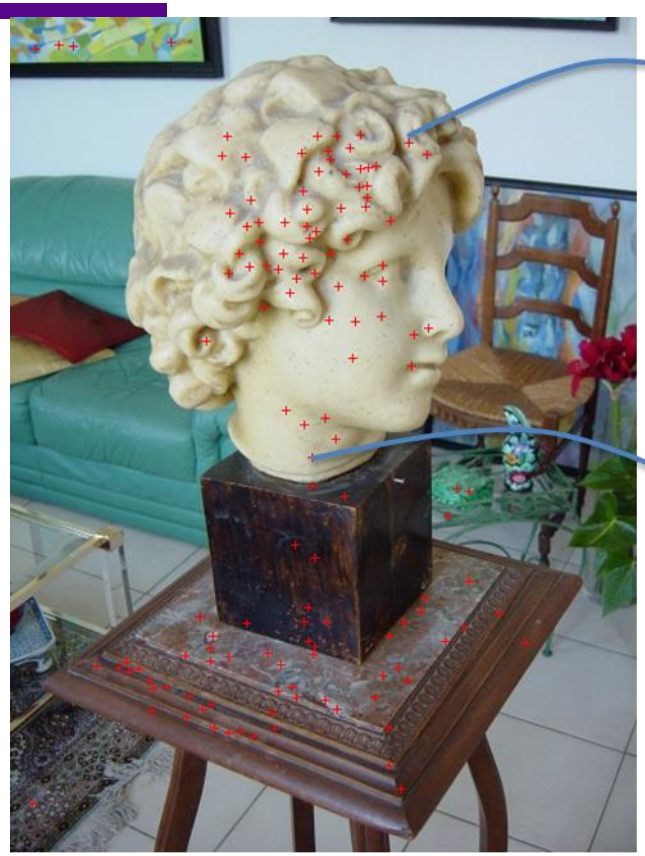


keypoints → match → fundamental matrix → essential matrix → [R|t] → triangulation →

keypoints →

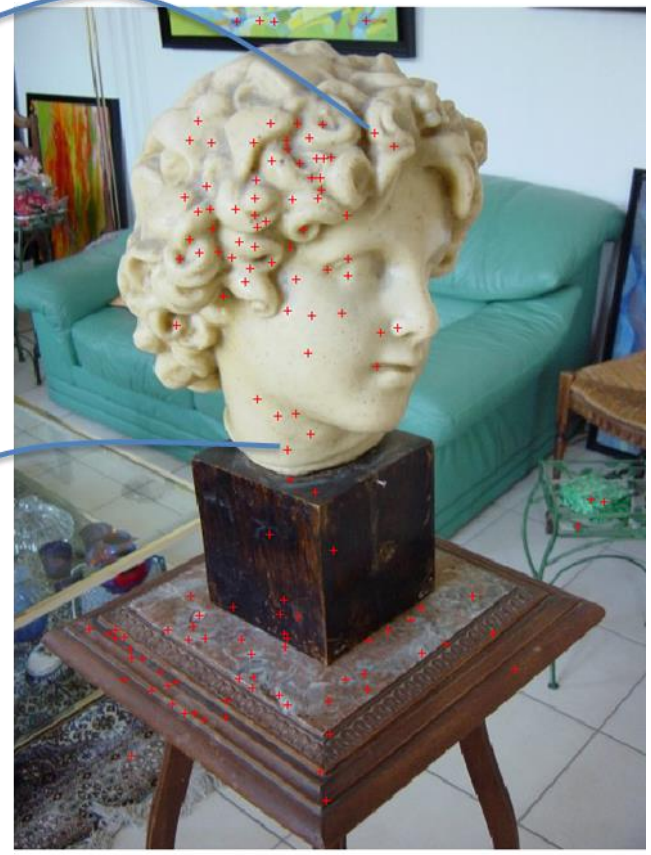


Find Keypoint Correspondences



SIFT descriptor

SIFT descriptor



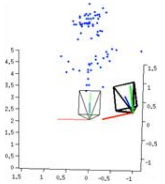
SIFT descriptor

SIFT descriptor

keypoints

keypoints

match → fundamental matrix → essential matrix → [R|t] → triangulation →

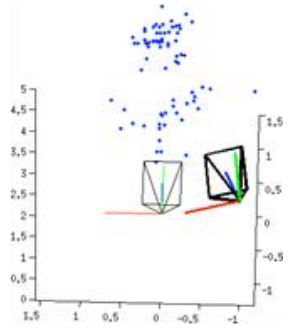


Two-view Reconstruction

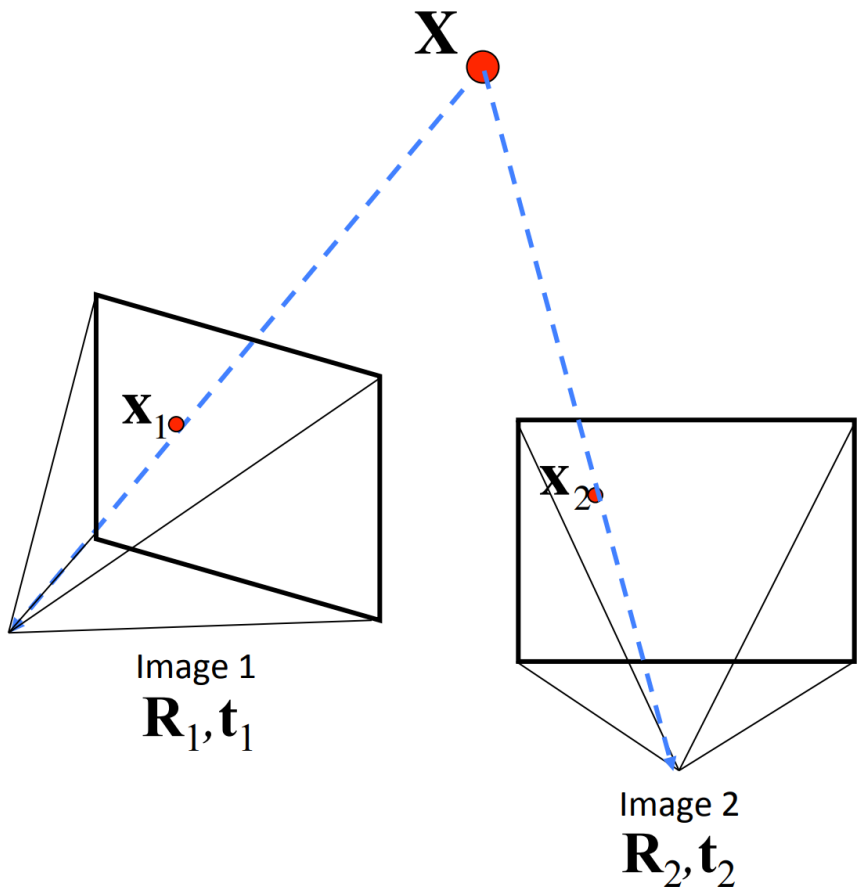

keypoints

keypoints

match → fundamental matrix → essential matrix → [R|t] → triangulation



Relative Pose ←→ Fundamental Matrix



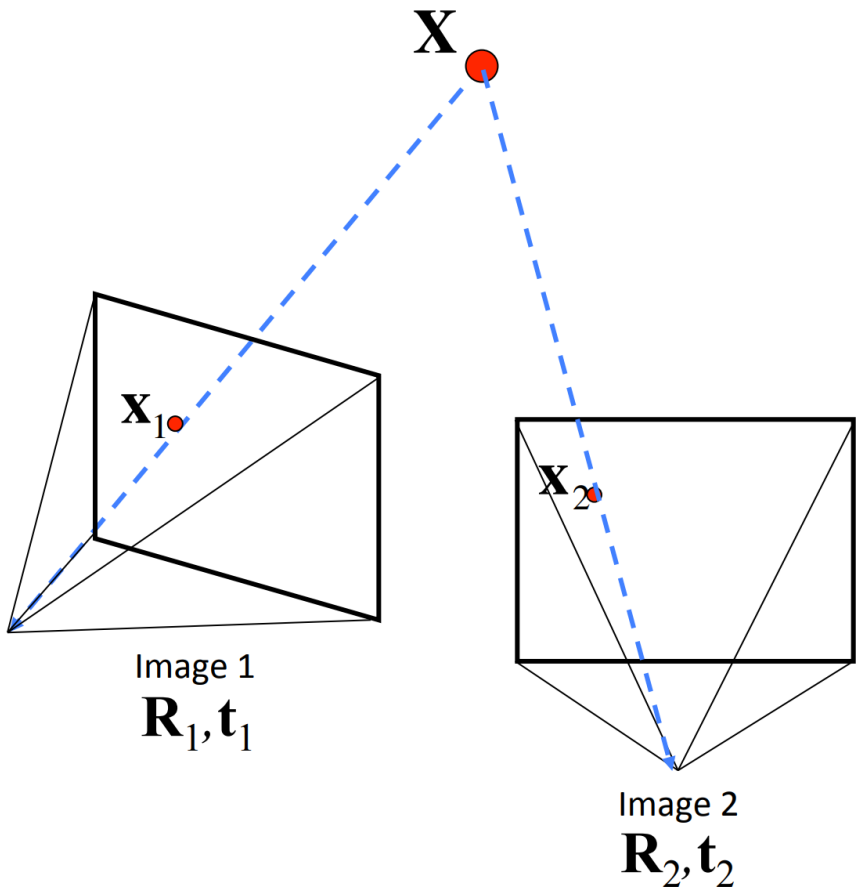
$$\mathbf{x}_1 \leftrightarrow \mathbf{x}_2$$

$$\mathbf{x}_1^T \mathbf{F} \mathbf{x}_2 = 0$$

F-matrix are estimated by finding multiple pairs of such corresponding keypoints ($\mathbf{x}_1$, $\mathbf{x}_2$)

Slides from Jianxiong Xiao

Fundamental Matrix ←→ Essential Matrix

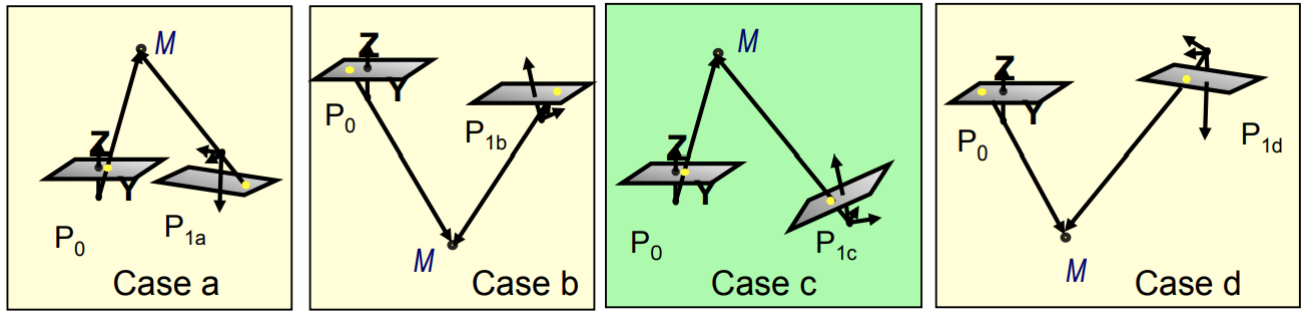


$$\mathbf{x}_1 \leftrightarrow \mathbf{x}_2$$

$$\mathbf{x}_1^T \mathbf{F} \mathbf{x}_2 = 0$$

$$\mathbf{E} = \mathbf{K}_1^T \mathbf{F} \mathbf{K}_2$$

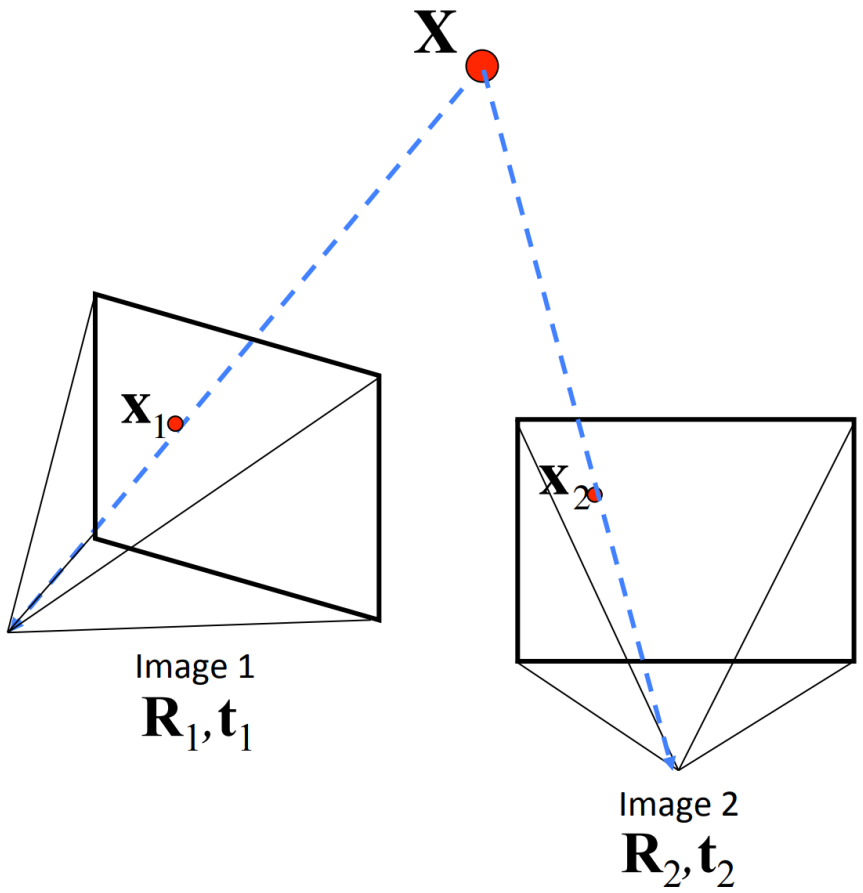
E-matrix are estimated by F-matrix with camera intrinsic parameters, and are used to estimate relative pose via decomposition $[t]_x R$



Slides from Jianxiong Xiao

Triangulation

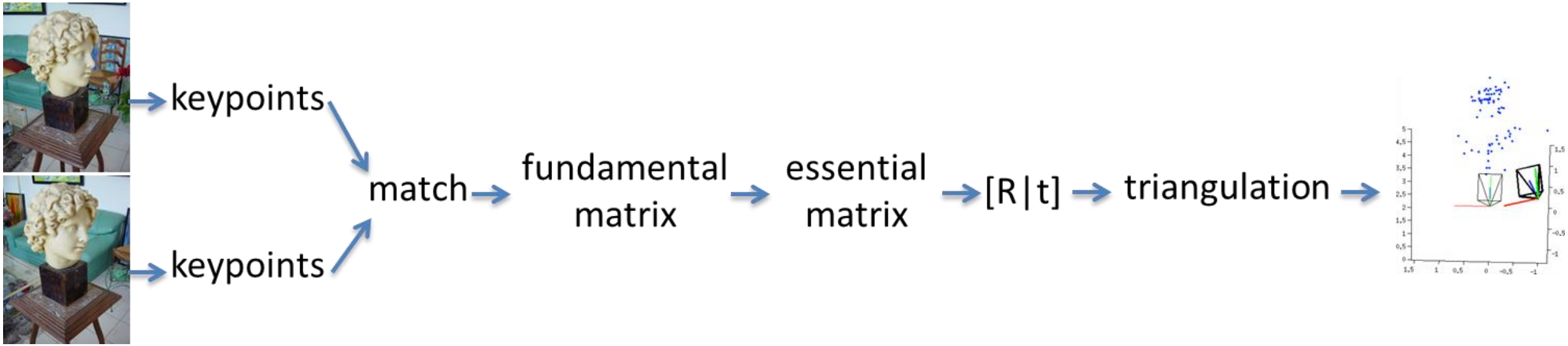
- Solve X=?



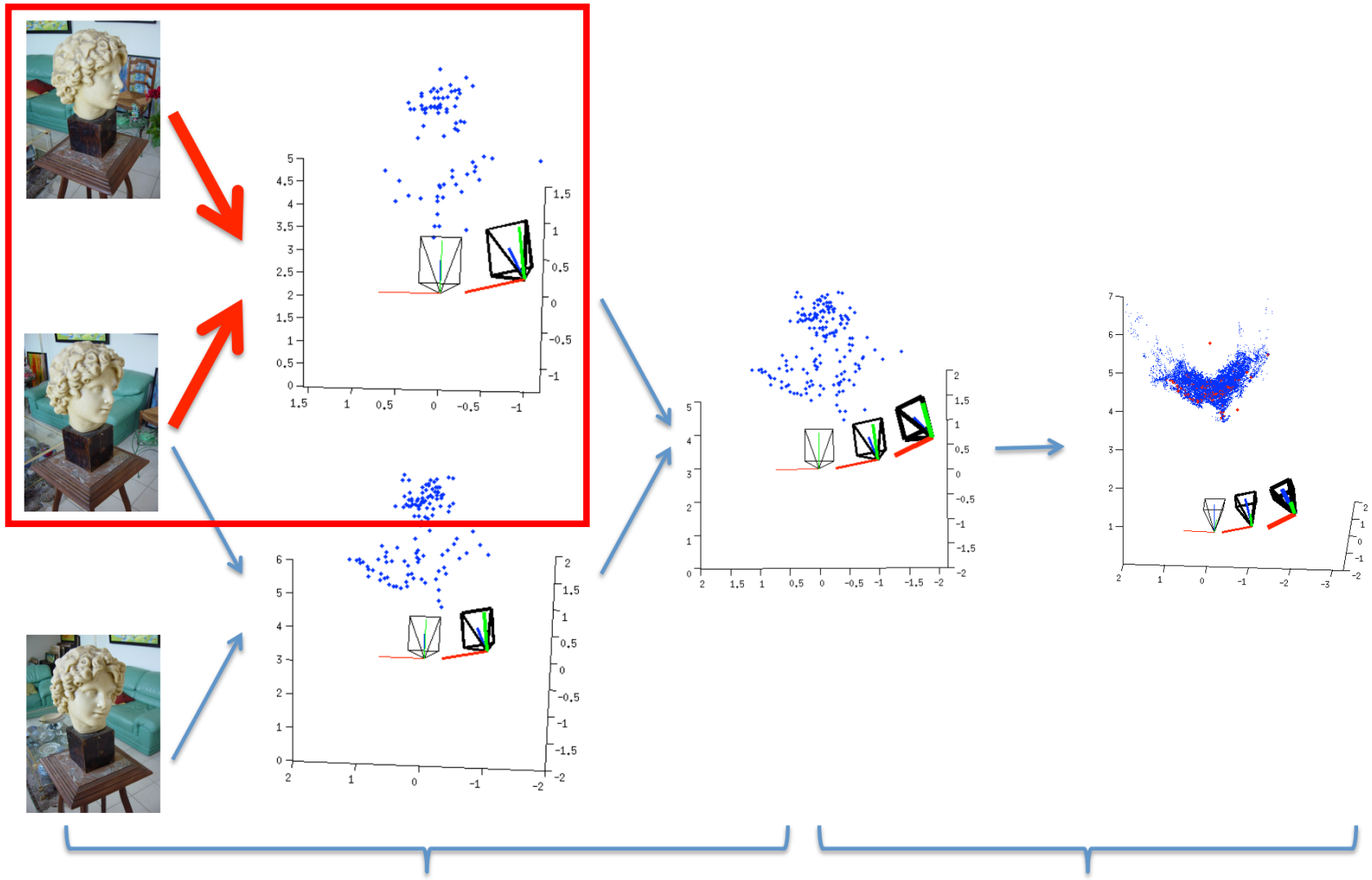
$$\mathbf{x}_1 = \mathbf{K}\left[\mathbf{R}_1 \middle| \mathbf{t}_1\right]\mathbf{X}$$

$$\mathbf{x}_2 = \mathbf{K}\left[\mathbf{R}_2 \middle| \mathbf{t}_2\right]\mathbf{X}$$

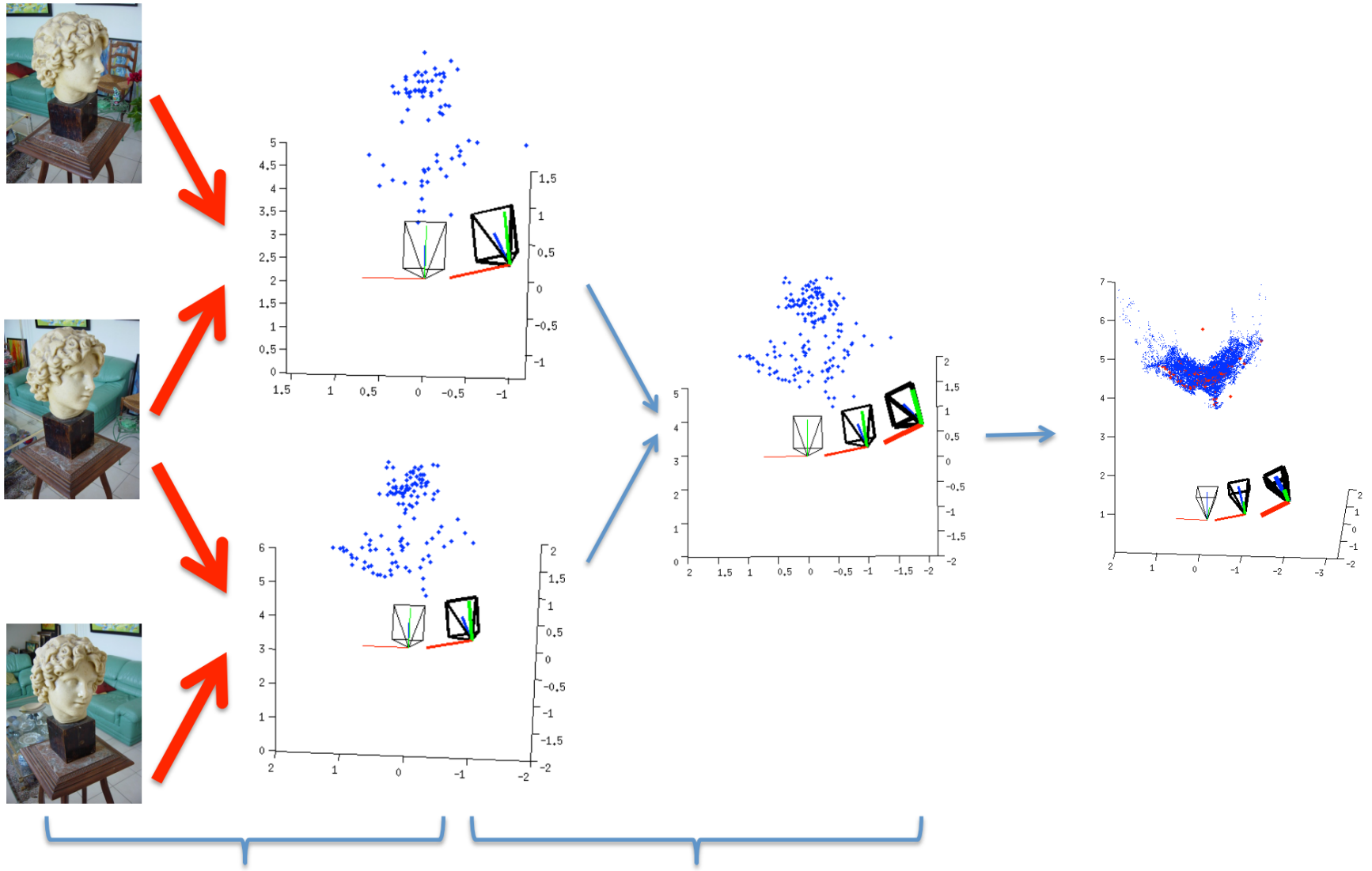
Recap: Two-view Reconstruction



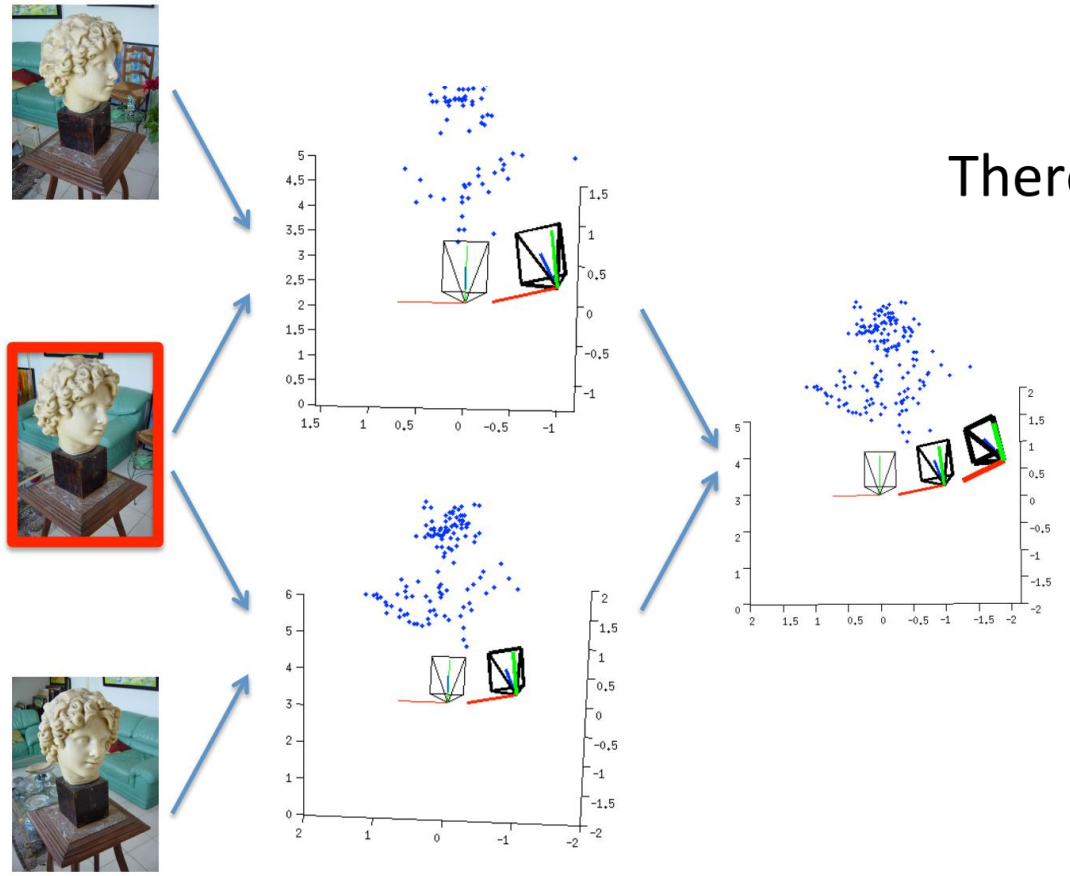
Multi-view Reconstruction



Multi-view Reconstruction

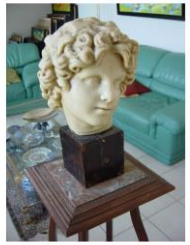


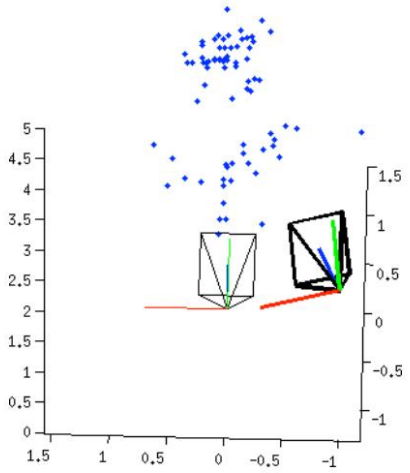
How to Merge Two Point Clouds?

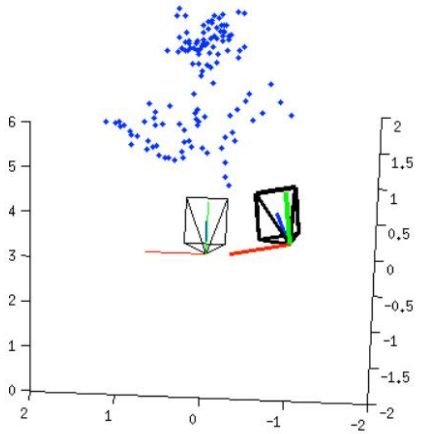


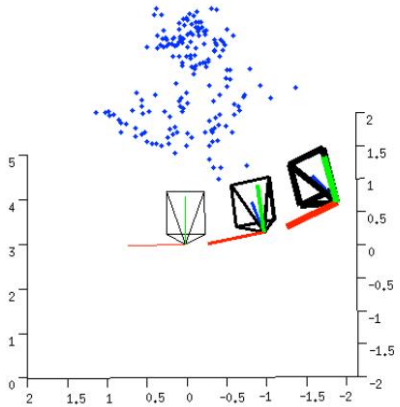
There can be only one $\left[\mathbf{R}_2 \middle| \mathbf{t}_2\right]$

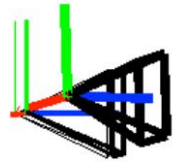
Oops

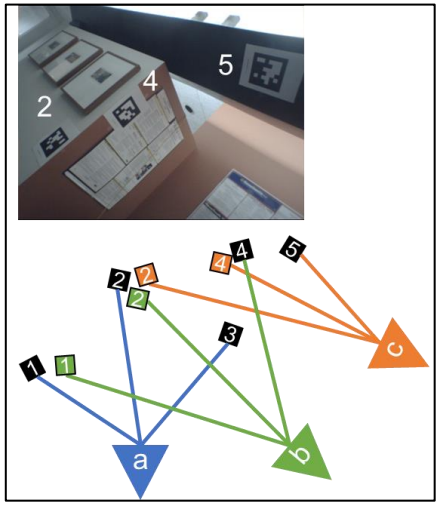


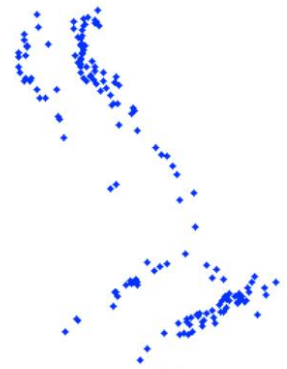






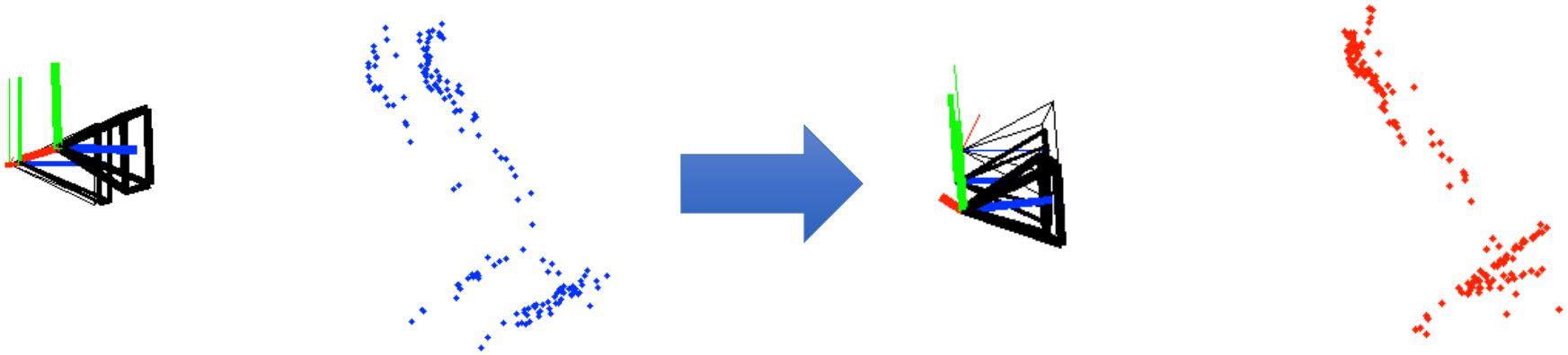






See From a Different Angle

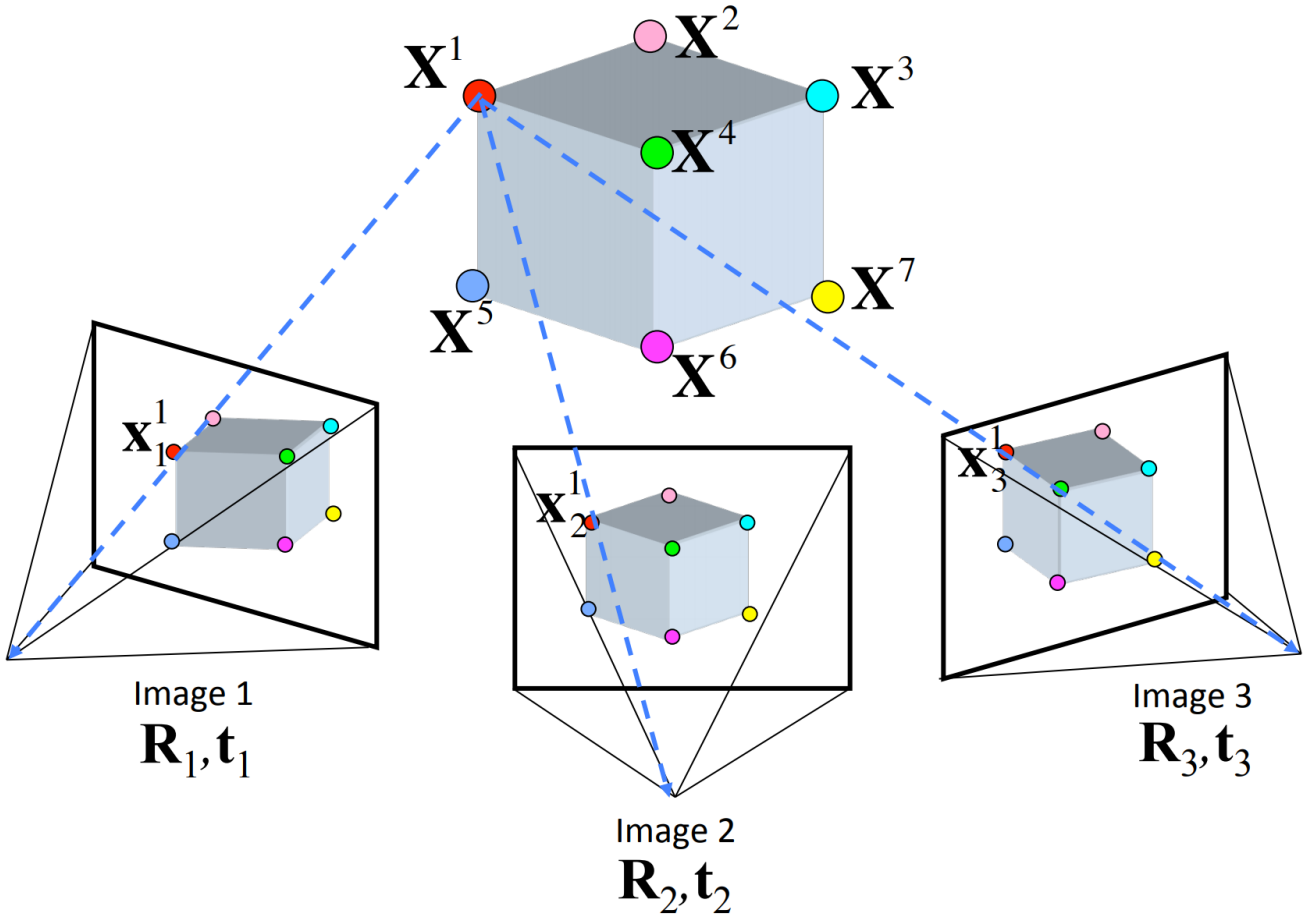
Bundle Adjustment Come to the Rescue



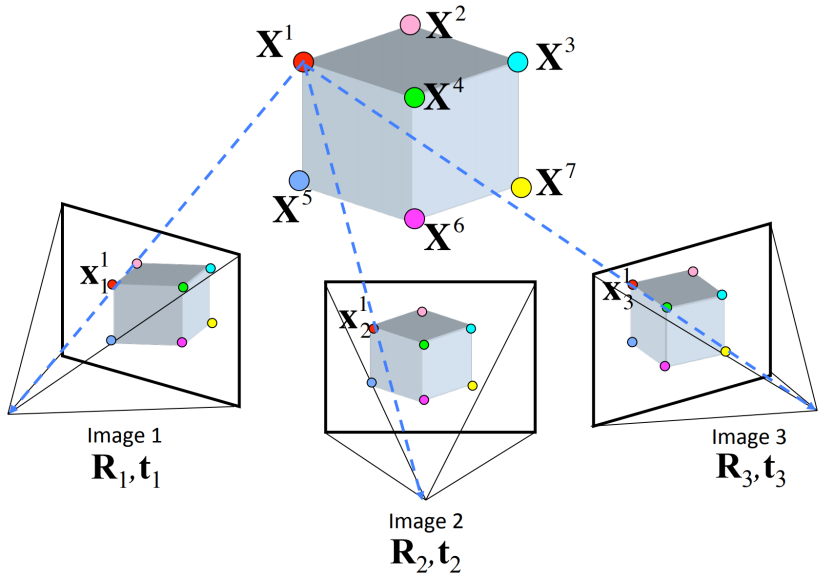
If assume all [R,t] are correct, error are all shown in point cloud.

Jointly optimize all [R,t] and points

Recall: Multiple Photos of Multiple Points



Writing in Math



	Point 1	Point 2	Point 3
Image 1	$\mathbf{x}_1^1 = \mathbf{K}\left[\mathbf{R}_1 \middle\| \mathbf{t}_1\right]\mathbf{X}^1$	$\mathbf{x}_1^2 = \mathbf{K}\left[\mathbf{R}_1 \middle\| \mathbf{t}_1\right]\mathbf{X}^2$	
Image 2	$\mathbf{x}_2^1 = \mathbf{K}\left[\mathbf{R}_2 \middle\| \mathbf{t}_2\right]\mathbf{X}^1$	$\mathbf{x}_2^2 = \mathbf{K}\left[\mathbf{R}_2 \middle\| \mathbf{t}_2\right]\mathbf{X}^2$	$\mathbf{x}_2^3 = \mathbf{K}\left[\mathbf{R}_2 \middle\| \mathbf{t}_2\right]\mathbf{X}^3$
Image 3	$\mathbf{x}_3^1 = \mathbf{K}\left[\mathbf{R}_3 \middle\| \mathbf{t}_3\right]\mathbf{X}^1$		$\mathbf{x}_3^3 = \mathbf{K}\left[\mathbf{R}_3 \middle\| \mathbf{t}_3\right]\mathbf{X}^3$

Slides modified from Jianxiong Xiao

Formulating the SfM Problem Mathematically

- Input: Observed 2D image position

$$\begin{matrix} \tilde{\mathbf{x}}_1^1 & \tilde{\mathbf{x}}_1^2 & \\ \tilde{\mathbf{x}}_2^1 & \tilde{\mathbf{x}}_2^2 & \tilde{\mathbf{x}}_2^3 \\ \tilde{\mathbf{x}}_3^1 & & \tilde{\mathbf{x}}_3^3 \end{matrix}$$

- Output:

Unknown Camera Parameters (with some guess)

$$[\mathbf{R}_1|\mathbf{t}_1], [\mathbf{R}_2|\mathbf{t}_2], [\mathbf{R}_3|\mathbf{t}_3]$$

Unknown Point 3D coordinate (with some guess)

$$\mathbf{X}^1, \mathbf{X}^2, \mathbf{X}^3, \ldots$$

Bundle Adjustment

A valid solution $[\mathbf{R}_1|\mathbf{t}_1],[\mathbf{R}_2|\mathbf{t}_2],[\mathbf{R}_3|\mathbf{t}_3]$ and $\mathbf{X}^1,\mathbf{X}^2,\mathbf{X}^3,\cdots$ must let

Re-projection

$$\begin{cases} \mathbf{x}_1^1 = \mathbf{K}[\mathbf{R}_1|\mathbf{t}_1]\mathbf{X}^1 & \mathbf{x}_1^2 = \mathbf{K}[\mathbf{R}_1|\mathbf{t}_1]\mathbf{X}^2 & \\ \mathbf{x}_2^1 = \mathbf{K}[\mathbf{R}_2|\mathbf{t}_2]\mathbf{X}^1 & \mathbf{x}_2^2 = \mathbf{K}[\mathbf{R}_2|\mathbf{t}_2]\mathbf{X}^2 & \mathbf{x}_2^3 = \mathbf{K}[\mathbf{R}_2|\mathbf{t}_2]\mathbf{X}^3 \\ \mathbf{x}_3^1 = \mathbf{K}[\mathbf{R}_3|\mathbf{t}_3]\mathbf{X}^1 & & \mathbf{x}_3^3 = \mathbf{K}[\mathbf{R}_3|\mathbf{t}_3]\mathbf{X}^3 \end{cases}$$

=

Observation

$$\begin{cases} \tilde{\mathbf{x}}_1^1 & \tilde{\mathbf{x}}_1^2 & \\ \tilde{\mathbf{x}}_2^1 & \tilde{\mathbf{x}}_2^2 & \tilde{\mathbf{x}}_2^3 \\ \tilde{\mathbf{x}}_3^1 & & \tilde{\mathbf{x}}_3^3 \end{cases}$$

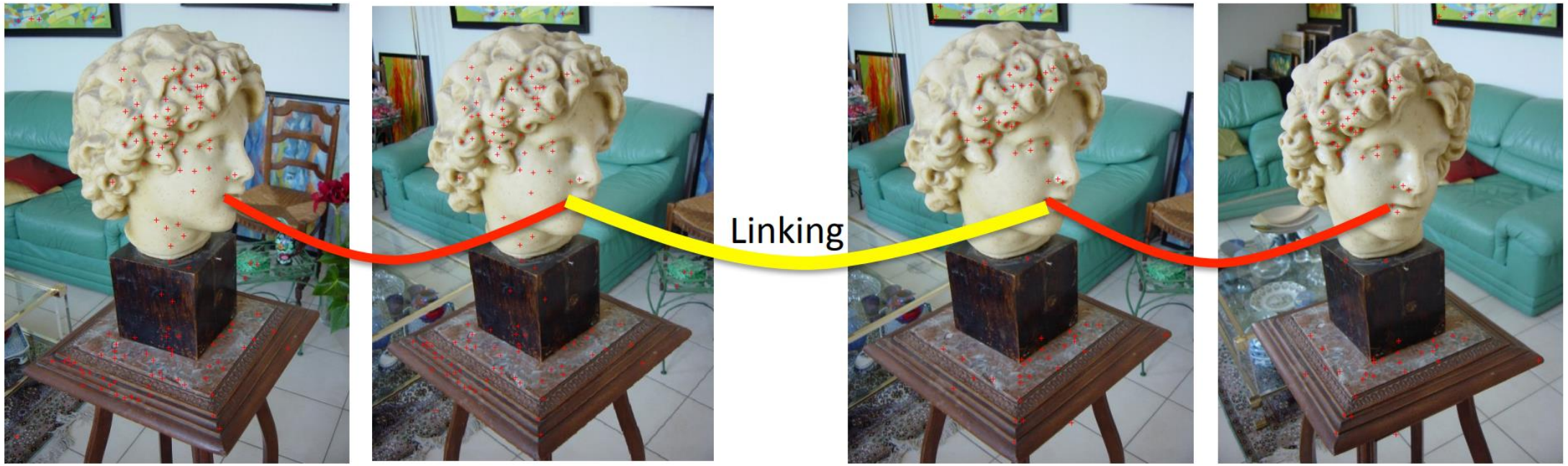
Slides from Jianxiong Xiao

Bundle Adjustment is a Least Squares Problem

A valid solution $[\mathbf{R}_1|\mathbf{t}_1],[\mathbf{R}_2|\mathbf{t}_2],[\mathbf{R}_3|\mathbf{t}_3]$ and $\mathbf{X}^1,\mathbf{X}^2,\mathbf{X}^3,\cdots$ must let the Re-projection close to the Observation, i.e. to minimize the reprojection error

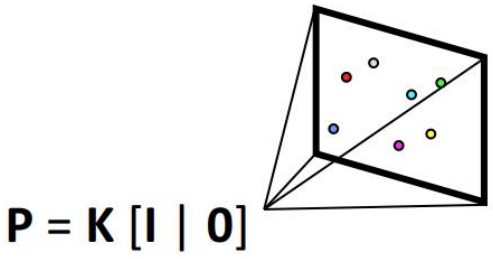
$$\min \sum_i \sum_j \left(\tilde{\mathbf{x}}_i^j - \mathbf{K}[\mathbf{R}_i|\mathbf{t}_i]\mathbf{X}^j \right)^2$$

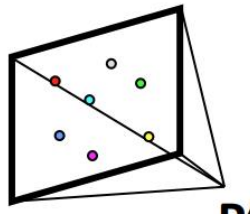
Keypoint Linking: Build the Objective Function



What We Just Learnt: Incremental SfM

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

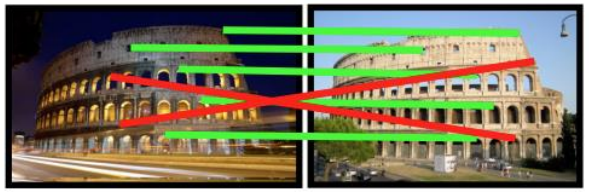


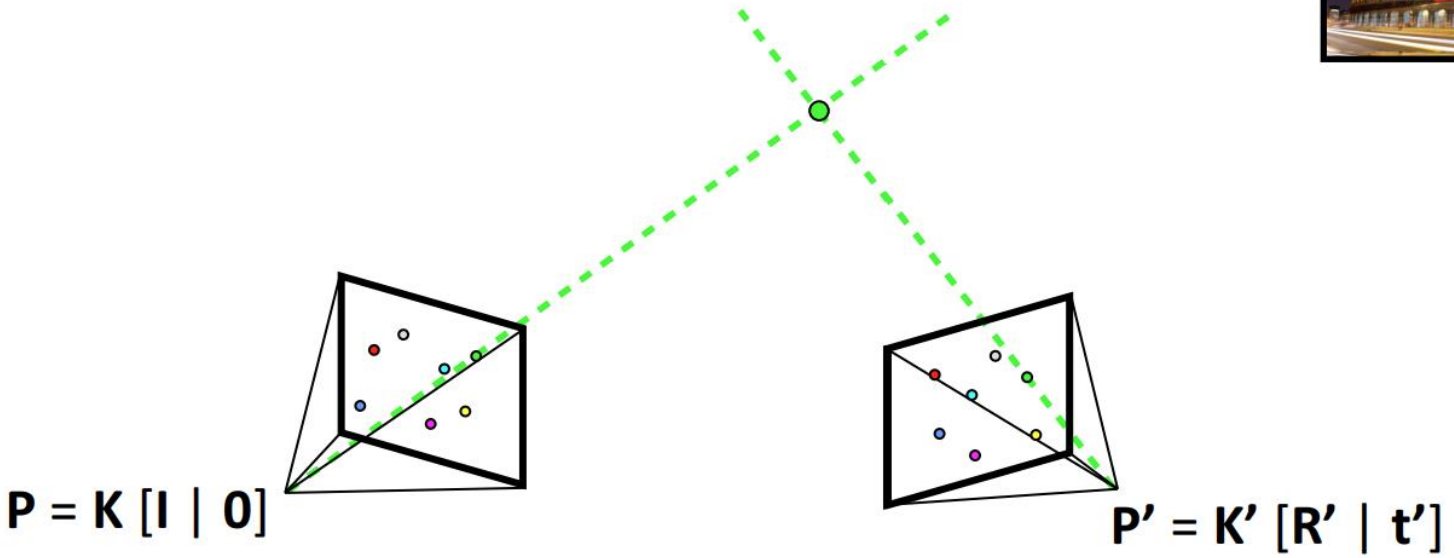


Slides from Image from: https://demuc.de/tutorials/cvpr2017/sparse-modeling.pdf

Incremental SfM

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

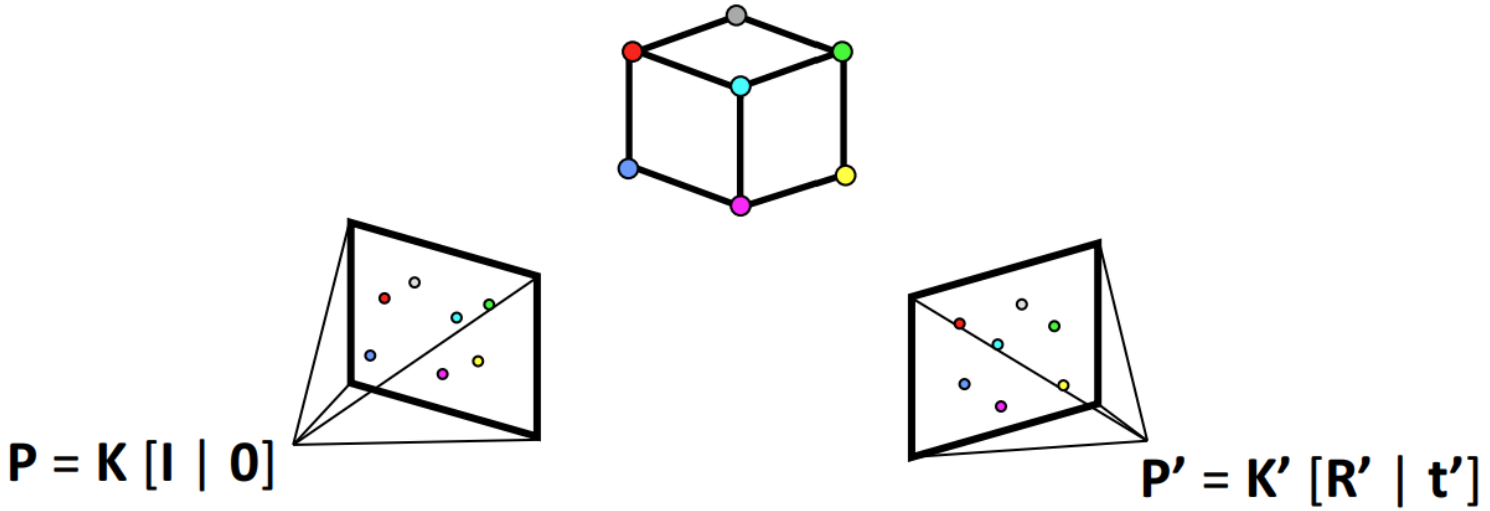




Slides from Image from: https://demuc.de/tutorials/cvpr2017/sparse-modeling.pdf

Incremental SfM

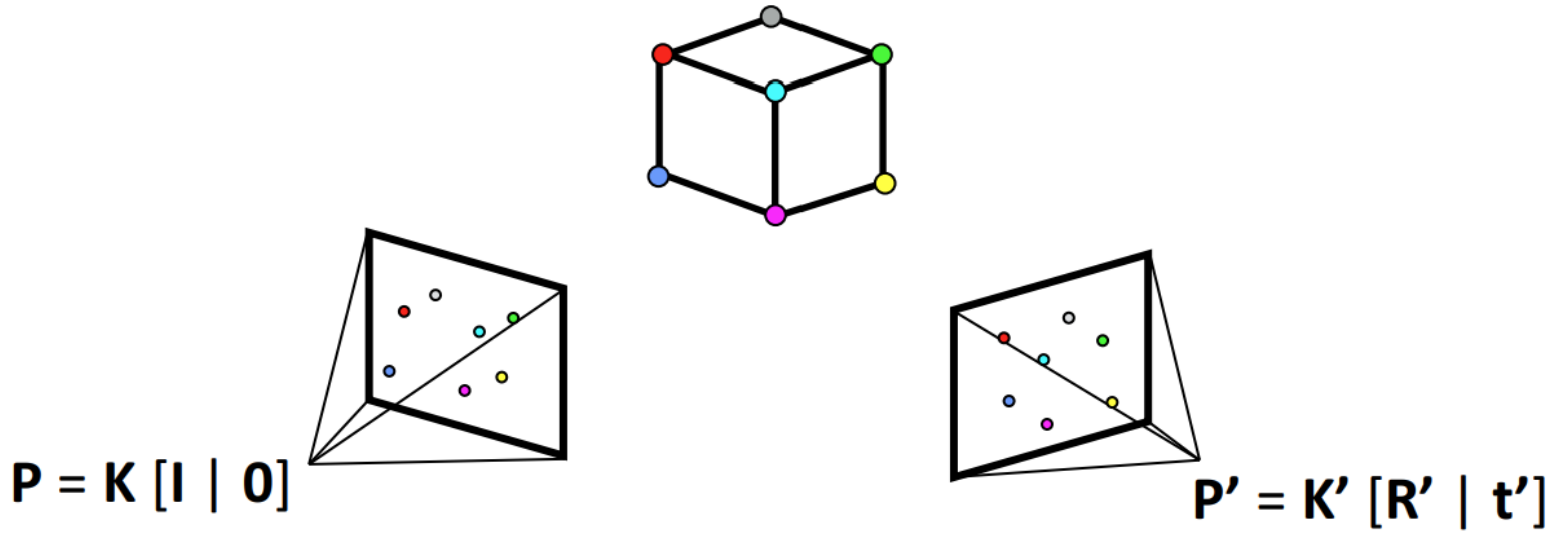
- Initialization
  1. Choose two non-panoramic views ($\|t\| = 1$)
  2. Triangulate inlier correspondences



Slides from Image from: https://demuc.de/tutorials/cvpr2017/sparse-modeling.pdf

Incremental SfM

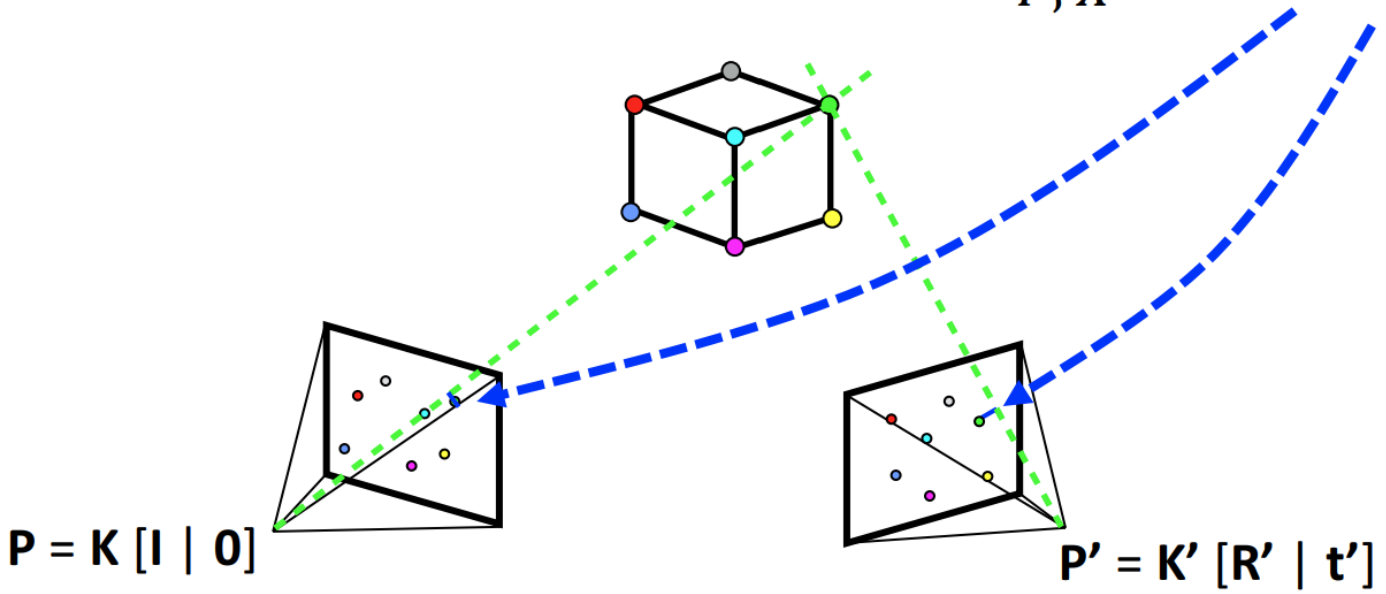
- Initialization
  1. Choose two non-panoramic views ($||t|| = 1$)
  2. Triangulate inlier correspondences
  3. Bundle adjustment



Slides from Image from: https://demuc.de/tutorials/cvpr2017/sparse-modeling.pdf

Incremental SfM

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

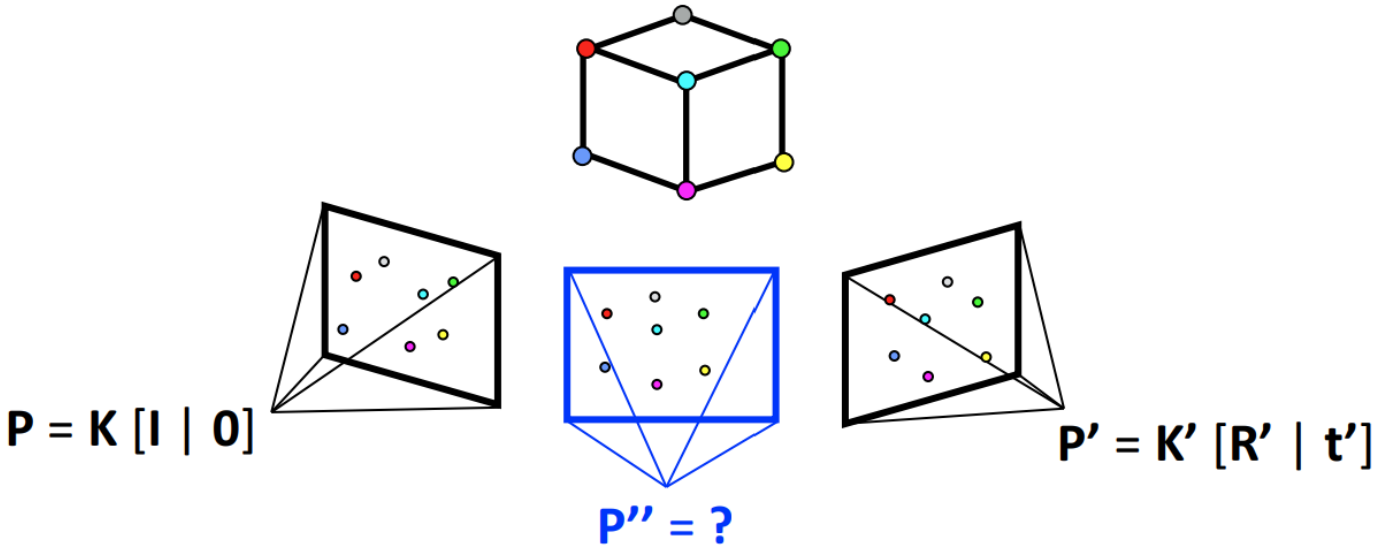


Ceres-Solver, http://ceres-solver.org/

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

Incremental SfM

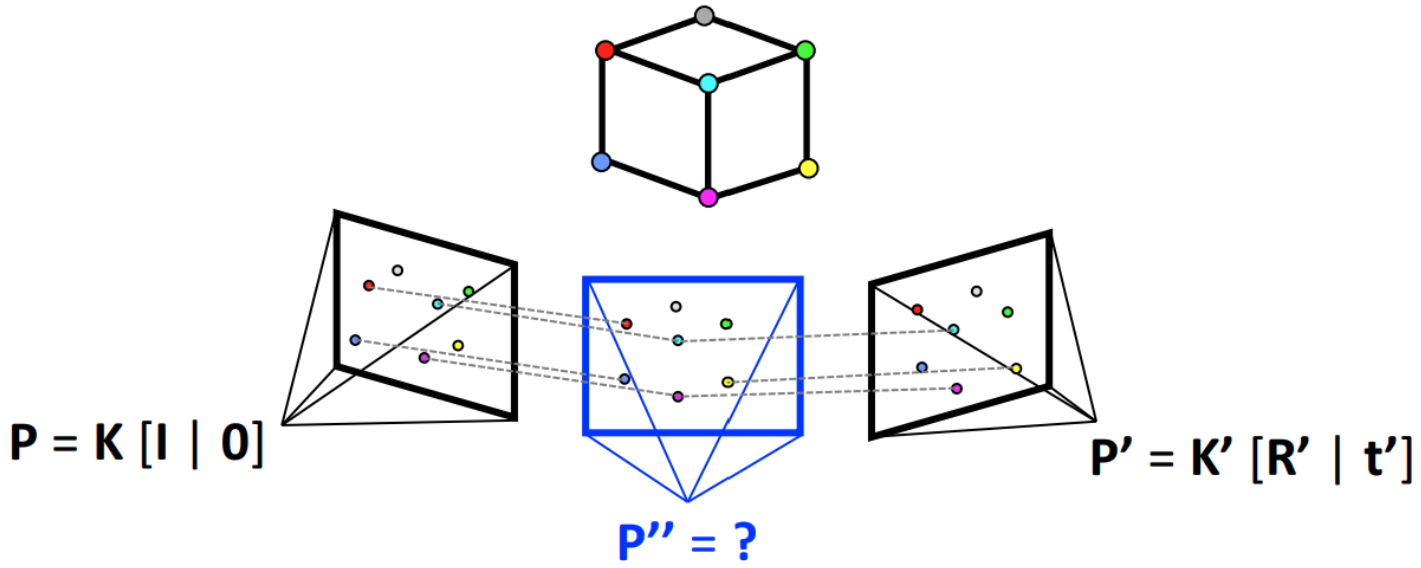
- Absolute camera registration



Slides from Image from: https://demuc.de/tutorials/cvpr2017/sparse-modeling.pdf

Incremental SfM

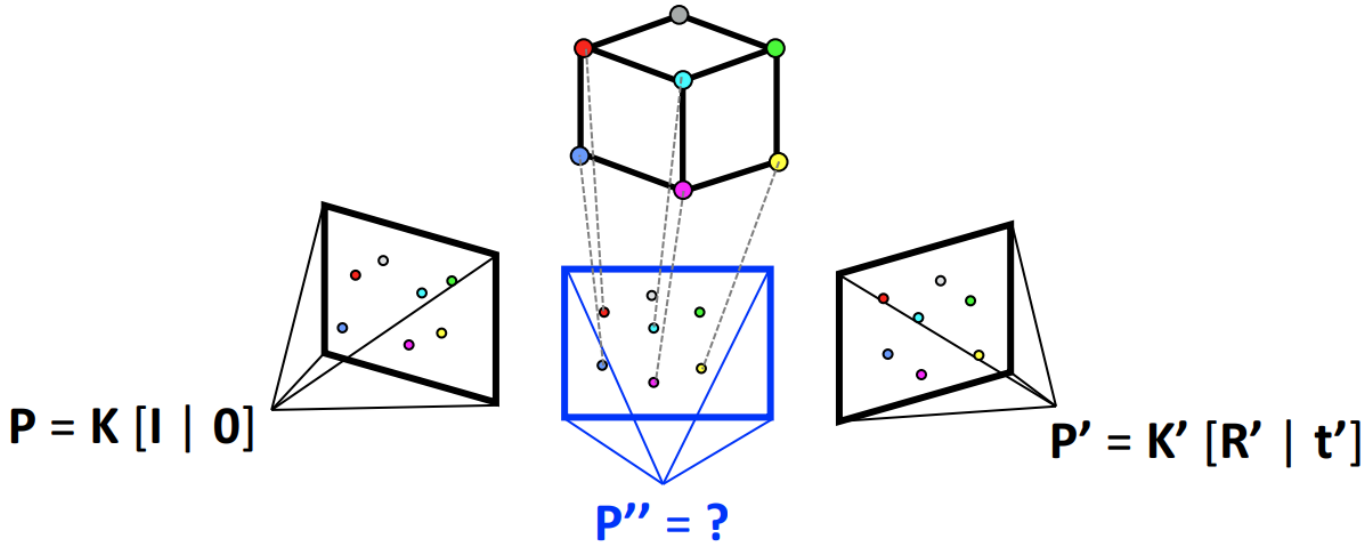
- Absolute camera registration
  1. Find 2D-3D correspondences



Slides from Image from: https://demuc.de/tutorials/cvpr2017/sparse-modeling.pdf

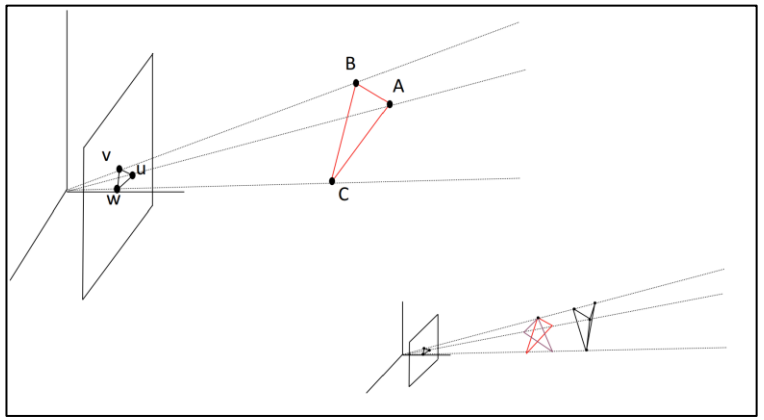
Incremental SfM

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

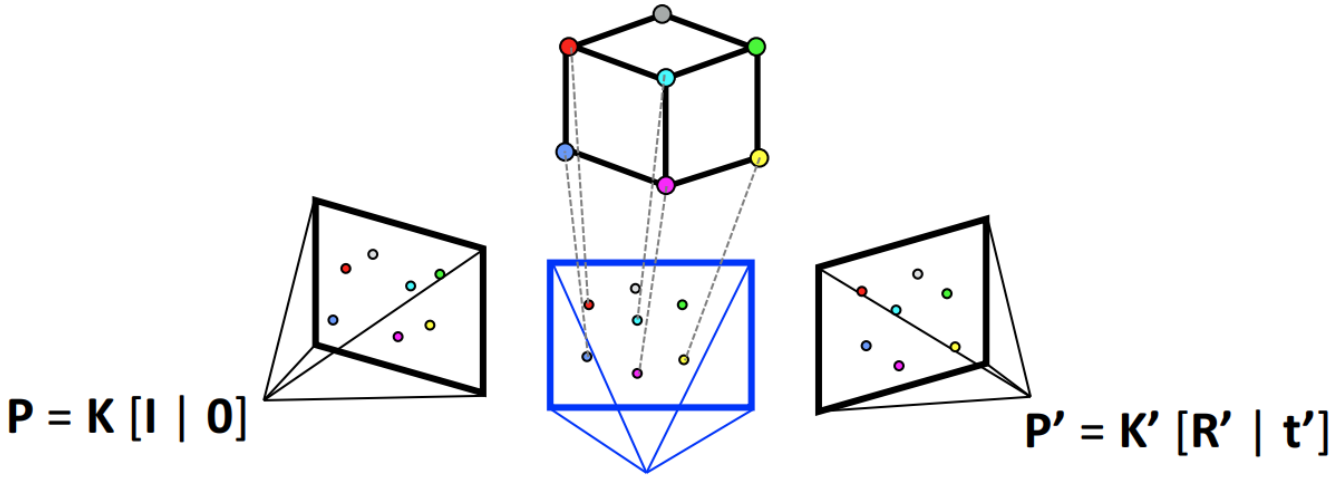


Slides from Image from: https://demuc.de/tutorials/cvpr2017/sparse-modeling.pdf

Incremental SfM



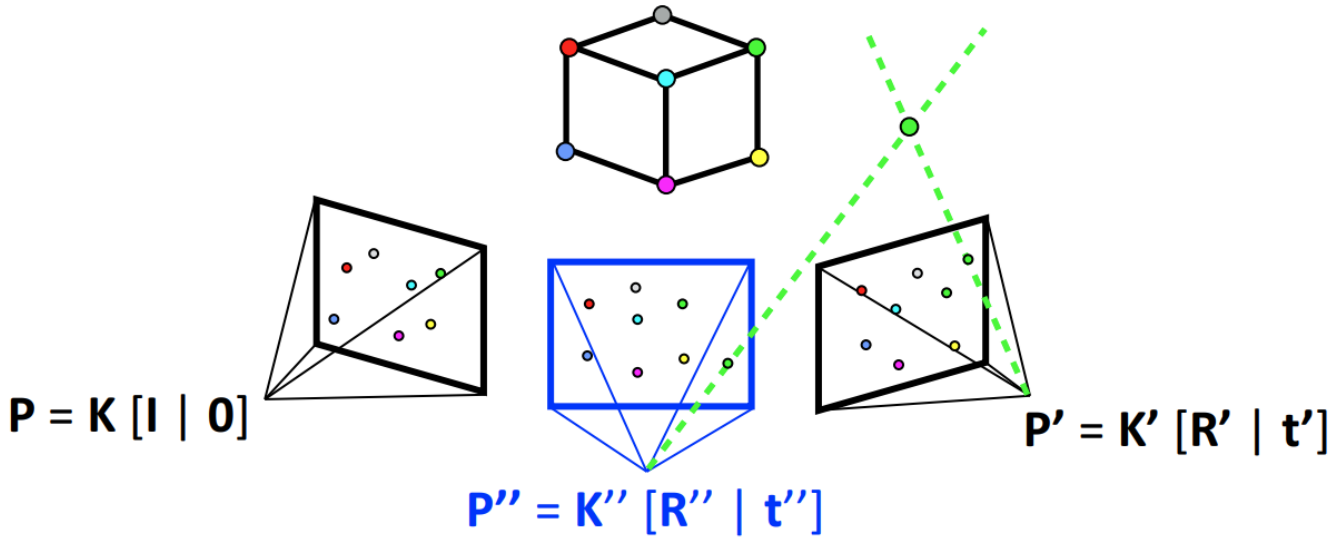
- Absolute camera registration
  1. Find 2D-3D correspondences
  2. Solve Perspective-n-Point problem



Slides from Image from: https://demuc.de/tutorials/cvpr2017/sparse-modeling.pdf

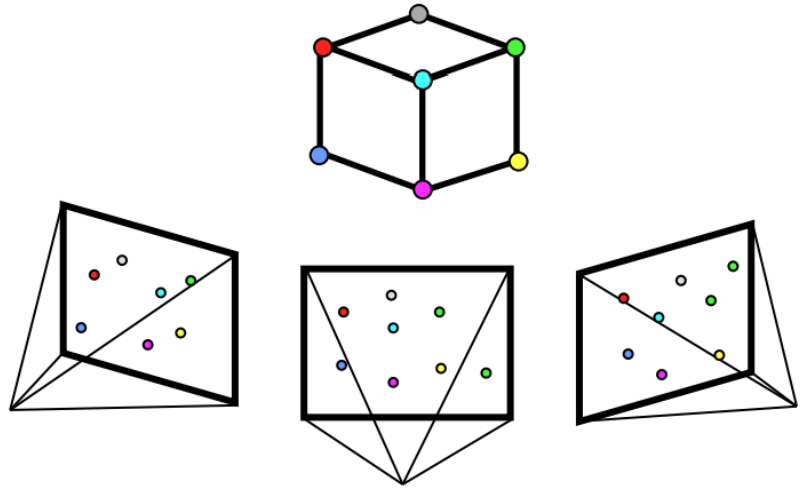
Incremental SfM

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



Incremental SfM

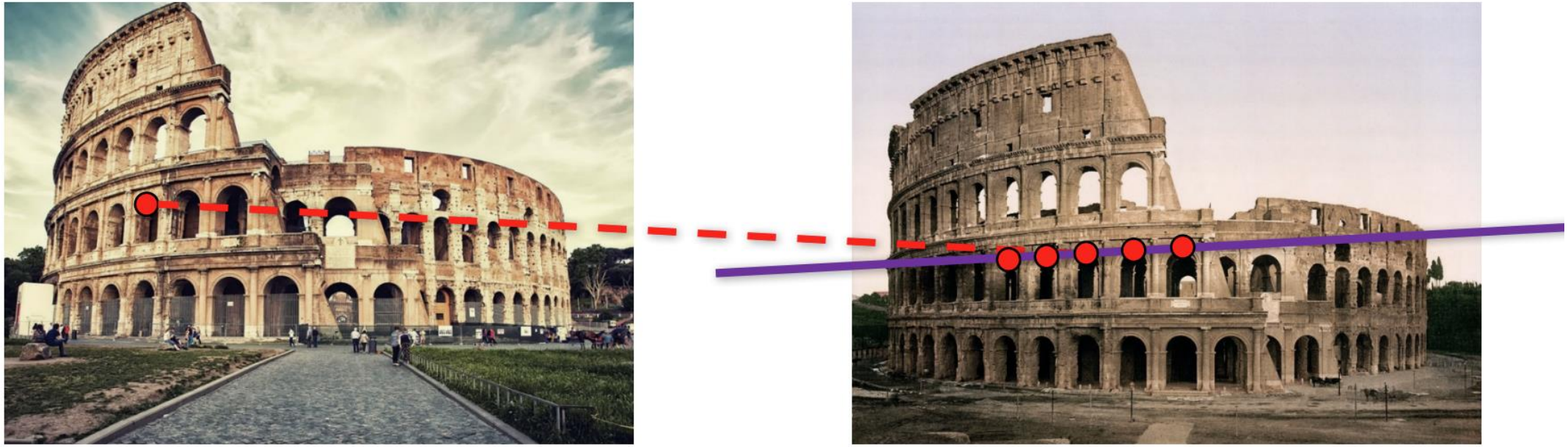
- Bundle adjustment $\min_{P,X} \|x - \pi(P,X)\|$



Slides from Image from: https://demuc.de/tutorials/cvpr2017/sparse-modeling.pdf

Incremental SfM: Important Details

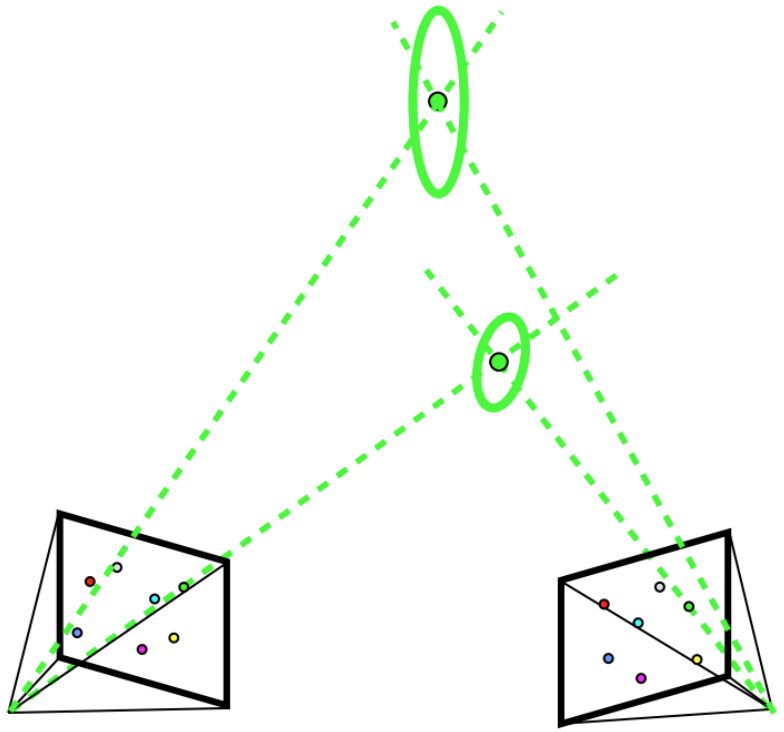
- Outlier filtering
  - Remove points with large reprojection error



Slides from Image from: https://demuc.de/tutorials/cvpr2017/sparse-modeling.pdf

Incremental SfM: Important Details

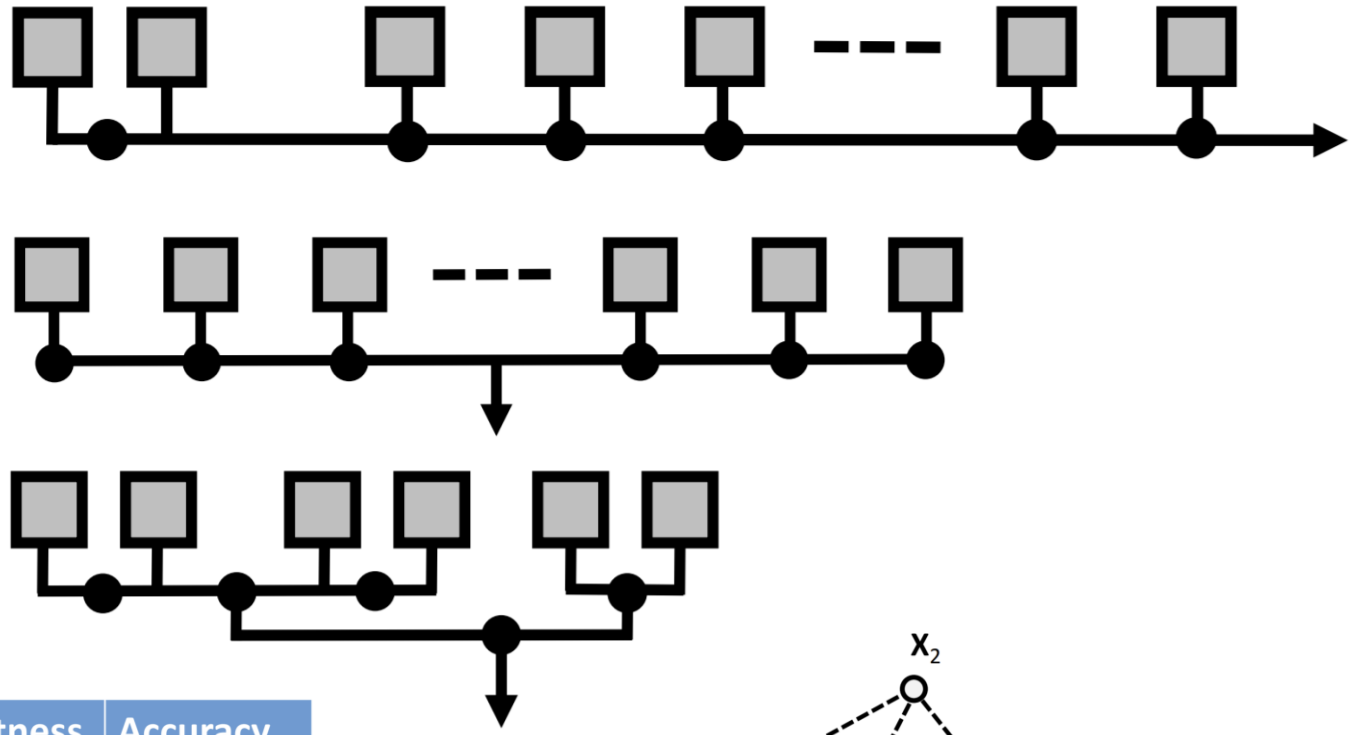
- Outlier filtering
  - Remove points with large reprojection error
  - Remove points at “infinity”



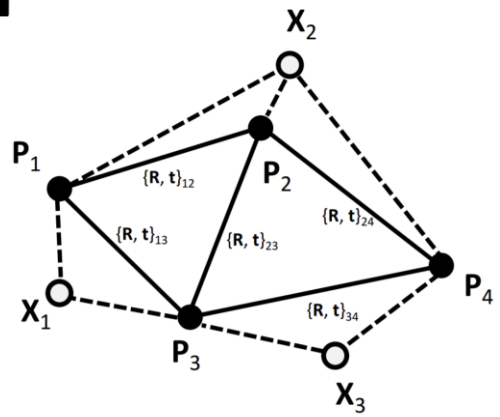
Slides from Image from: https://demuc.de/tutorials/cvpr2017/sparse-modeling.pdf

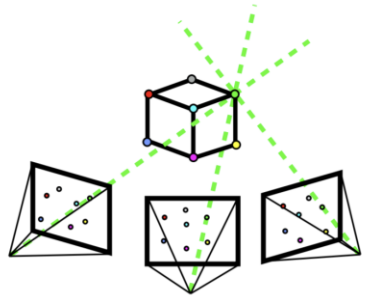
SfM Pipeline – Sparse 3D Reconstruction

- 3 paradigms
  - Incremental
  - Global
  - Hierarchical



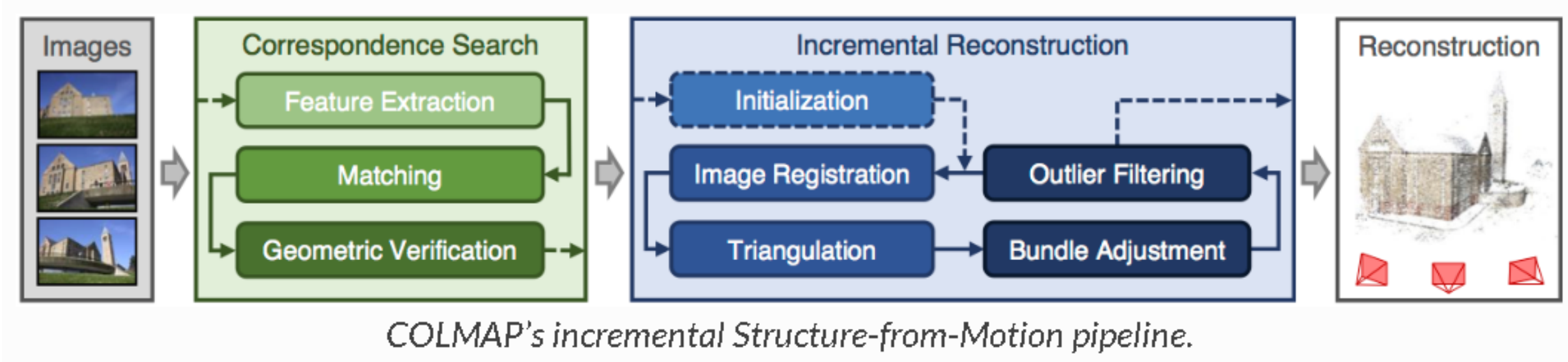
Method	Efficiency	Robustness	Accuracy
Incremental	-	++	+
Global	+	+	+
Hierarchical	++	-	-





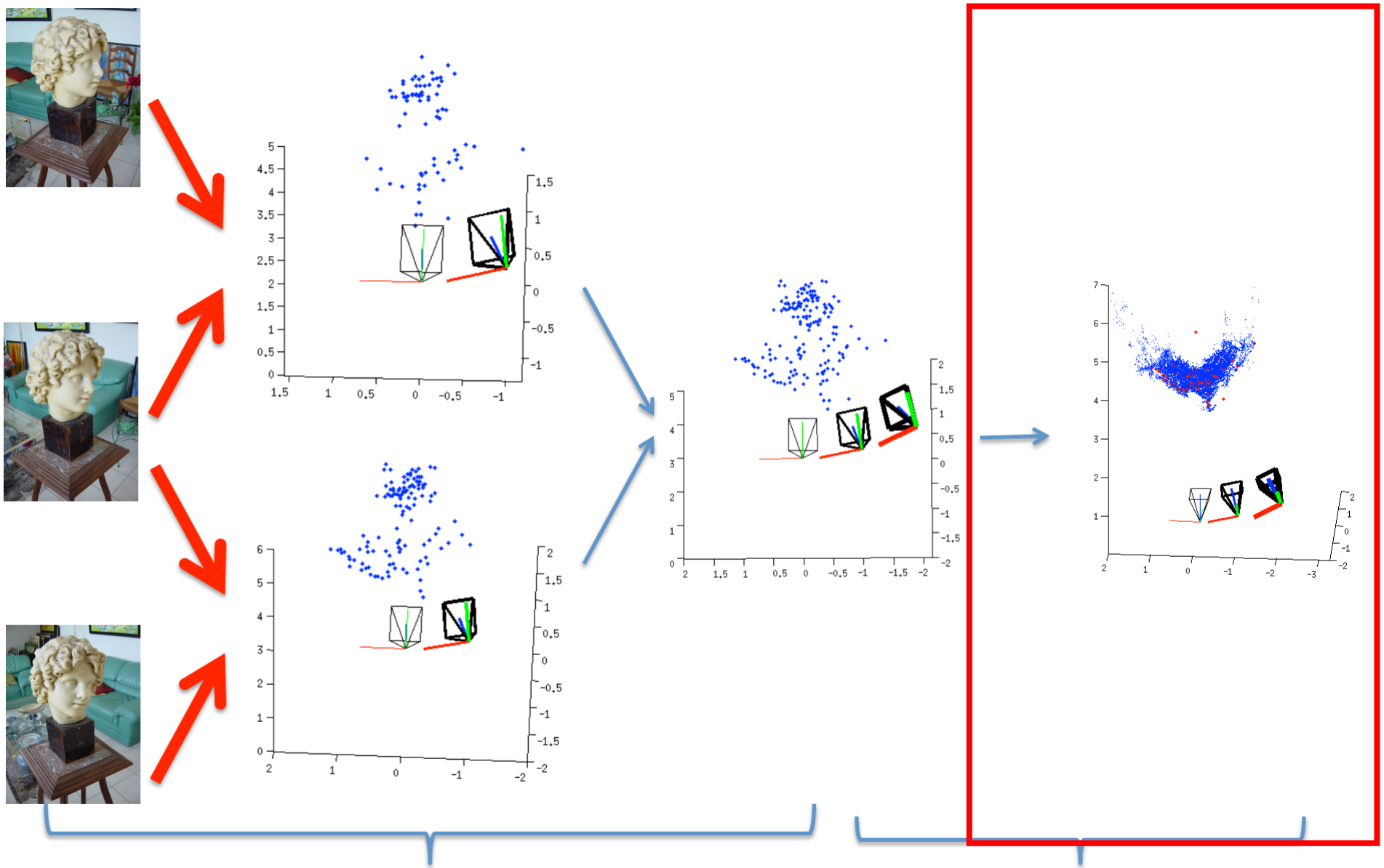
Images from: https://demuc.de/tutorials/cvpr2017/sparse-modeling.pdf

SfM Pipeline

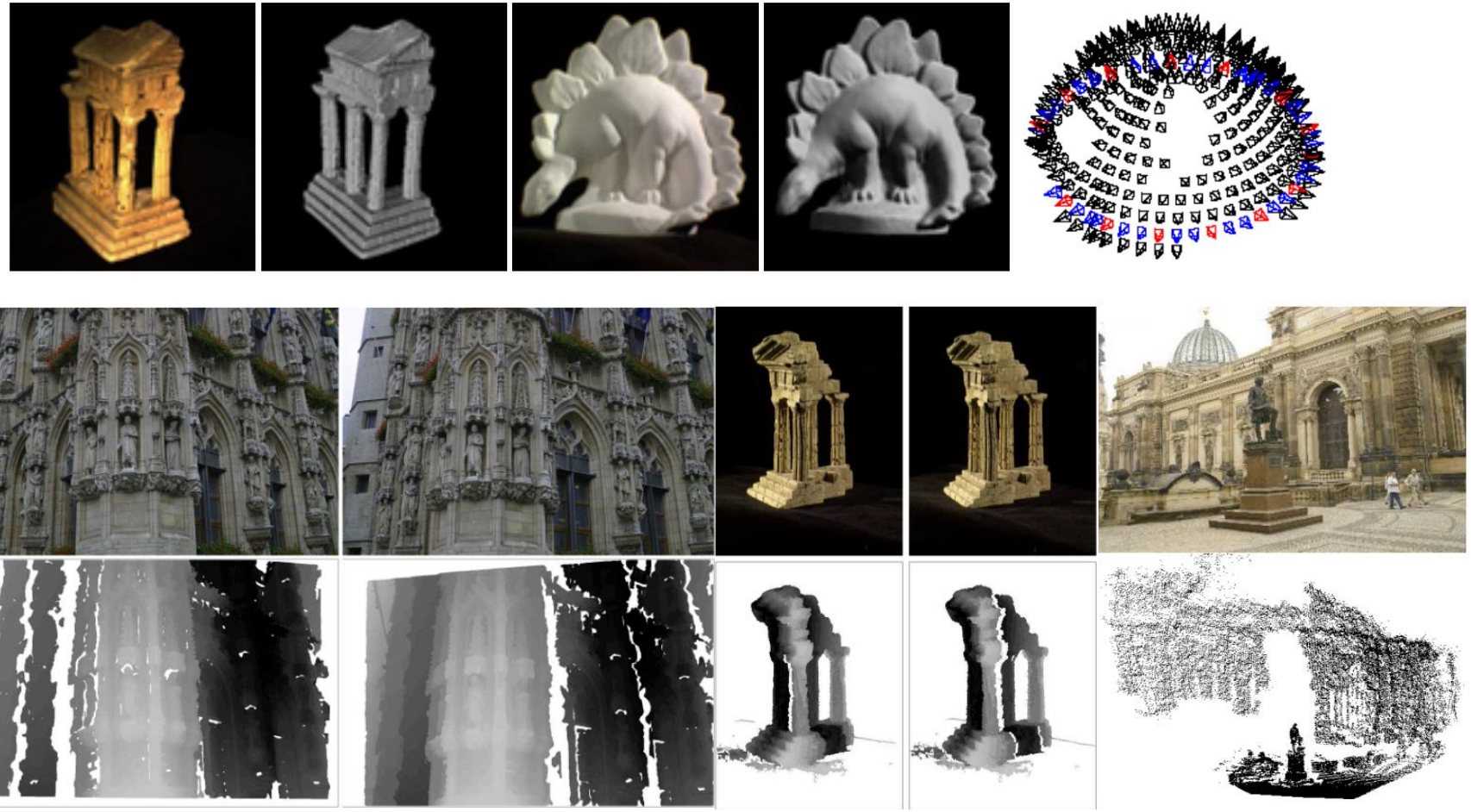


COLMAP's incremental Structure-from-Motion pipeline.

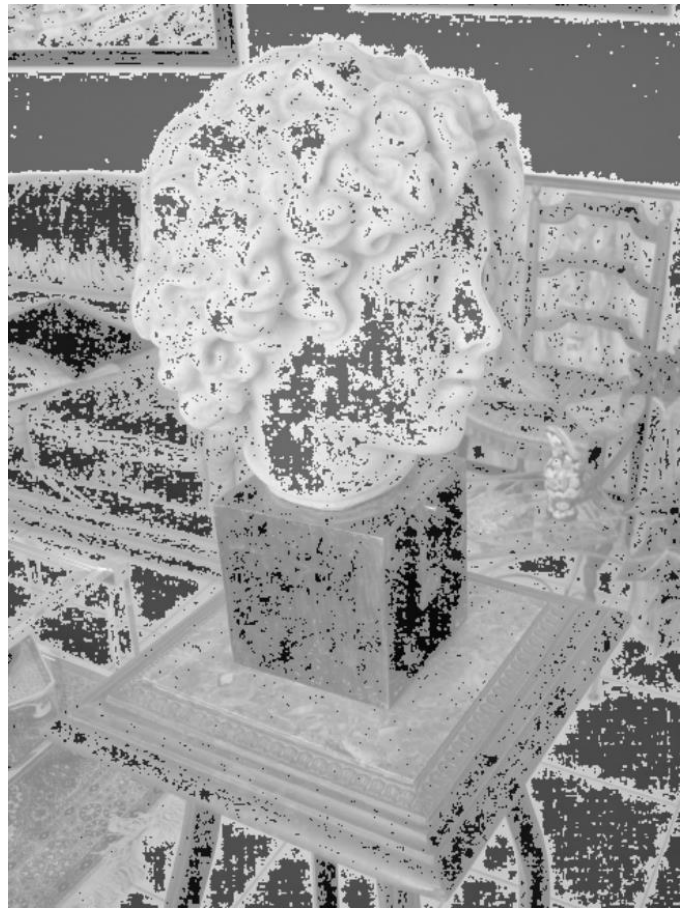
Image from: https://colmap.github.io/tutorial.html


Taught
Next

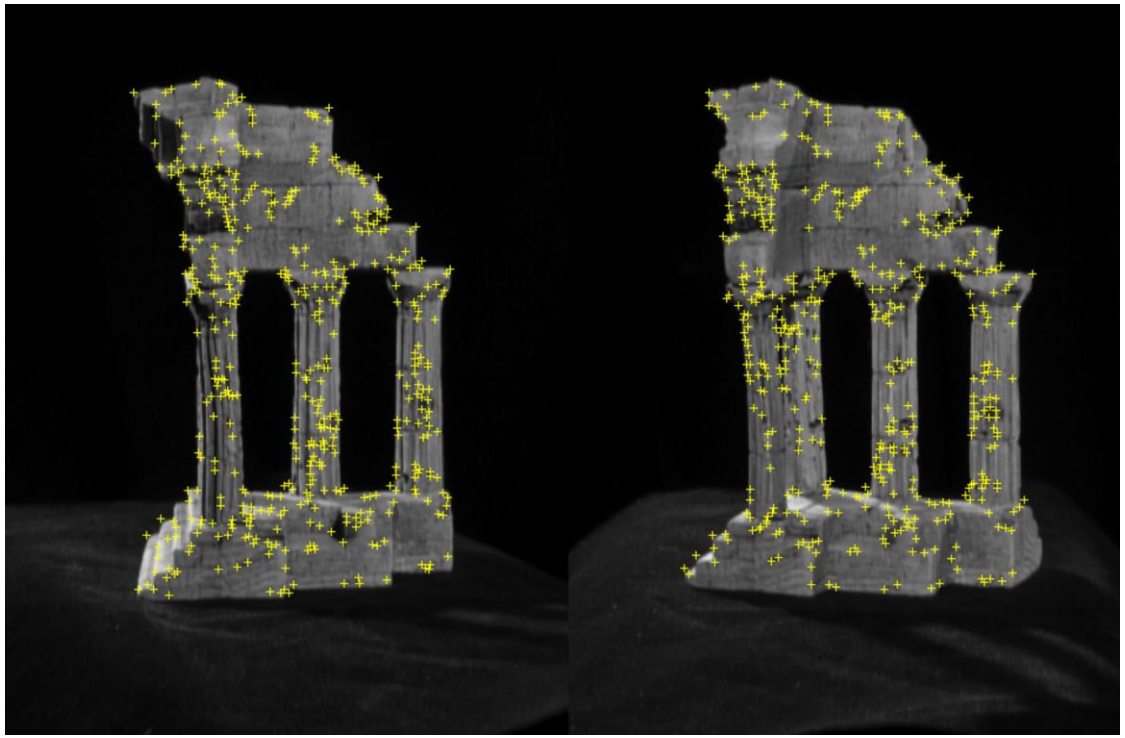
Multi-View Stereo: Matching Propagation

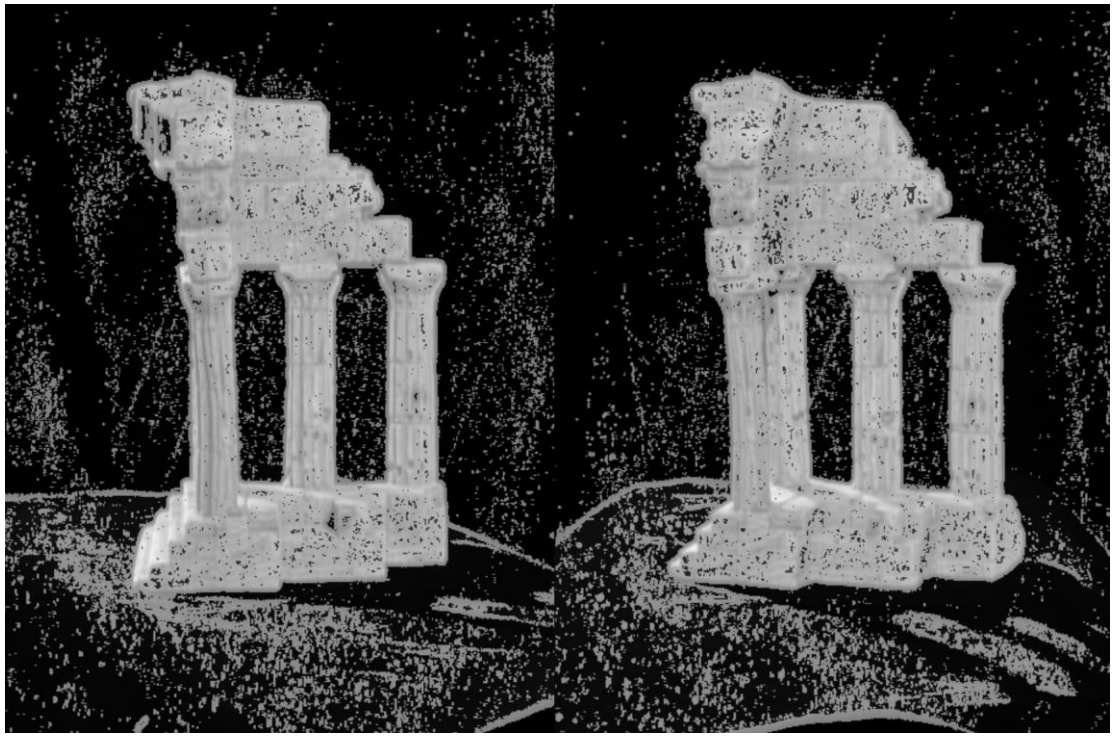


Start With the Seeding Pairs from SfM

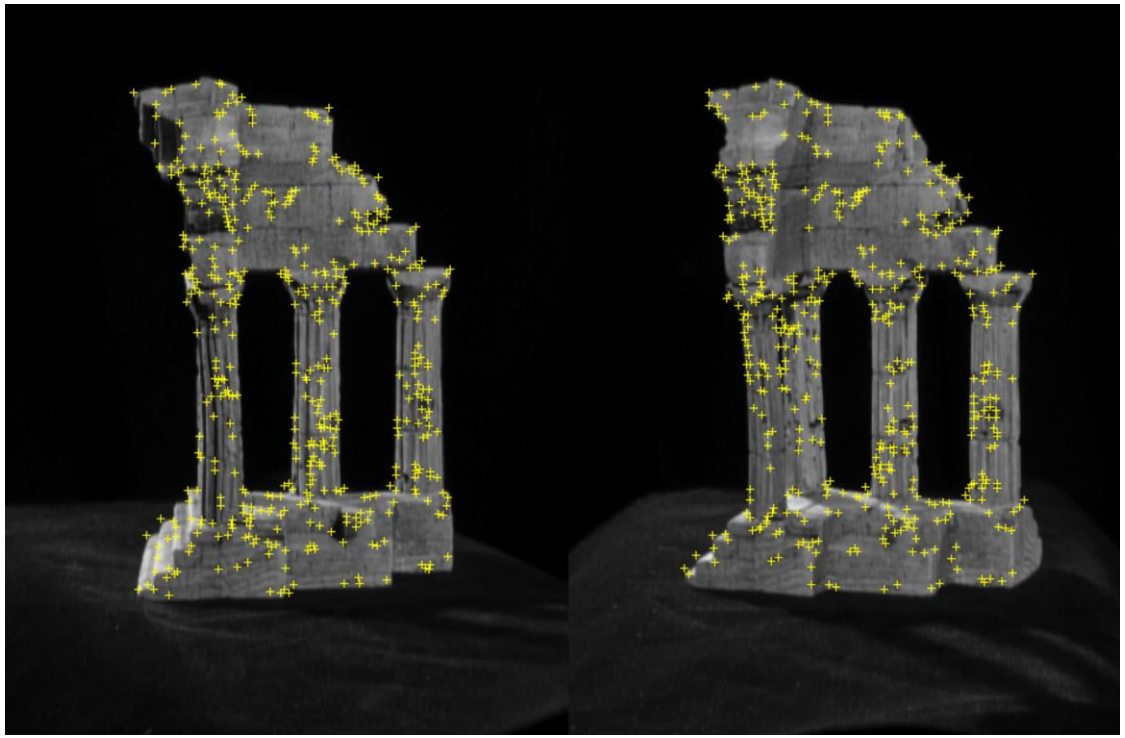


Select Only Matchable Area (with strong enough gradient)

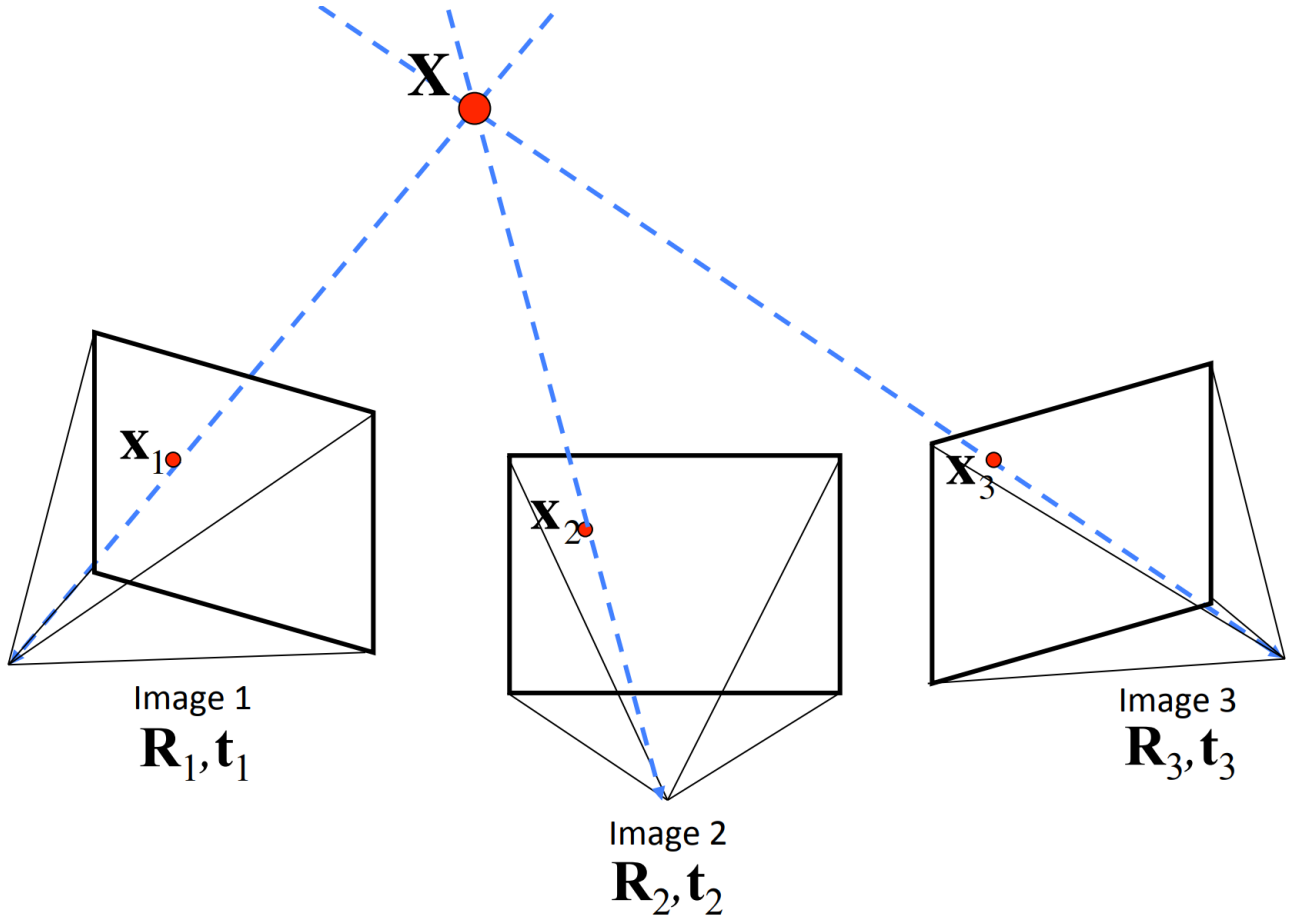




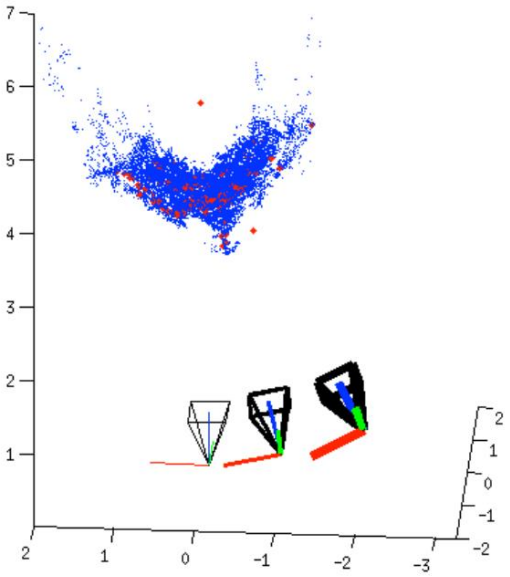
Select Only Matchable Area (with strong enough gradient)

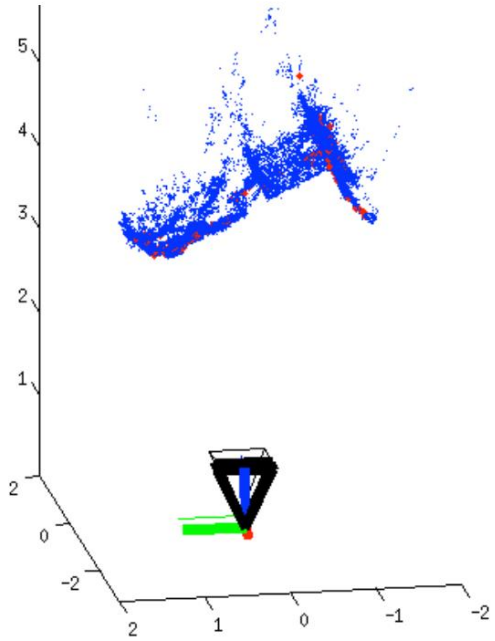


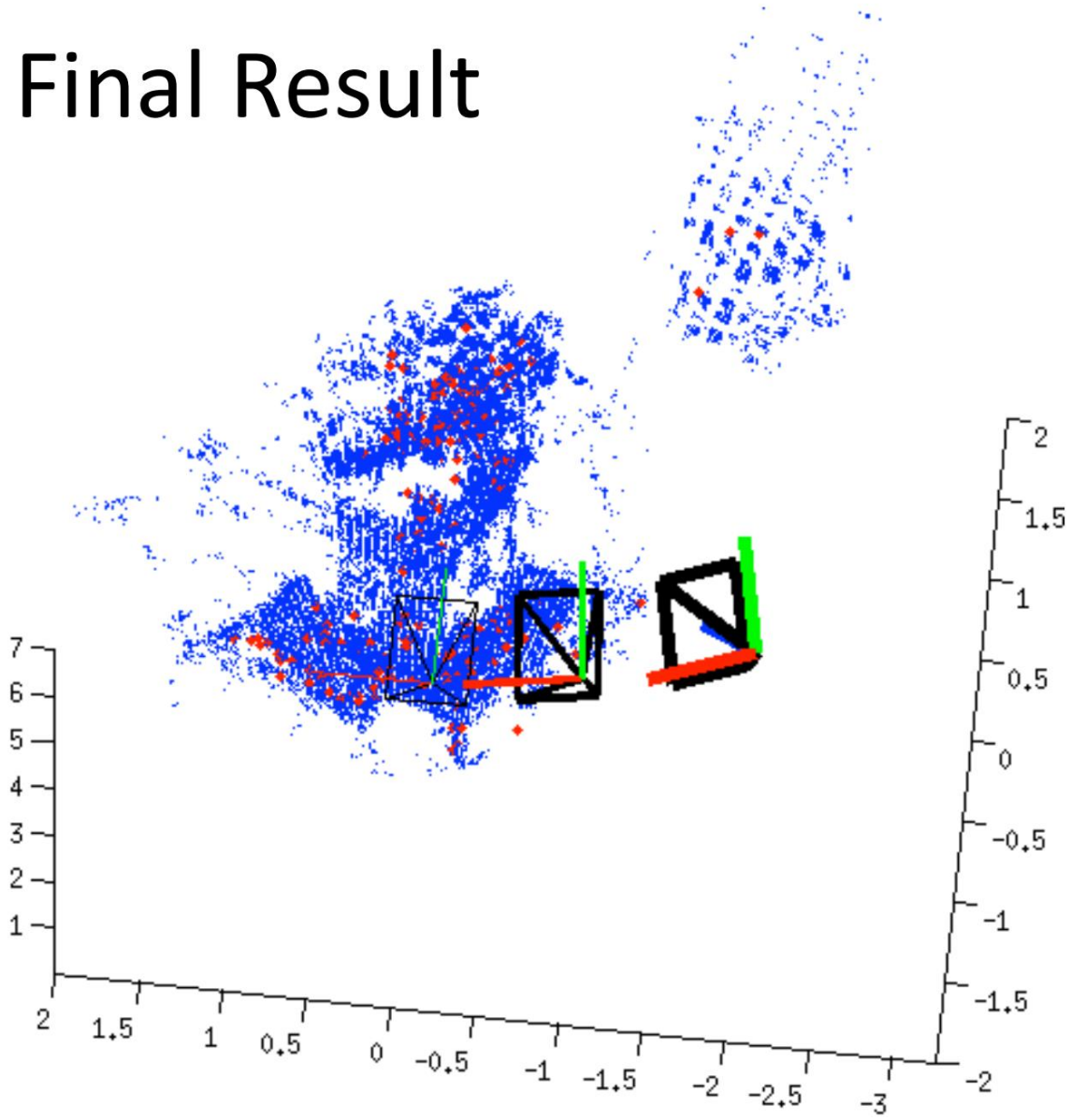
Triangulation from Multiple Views



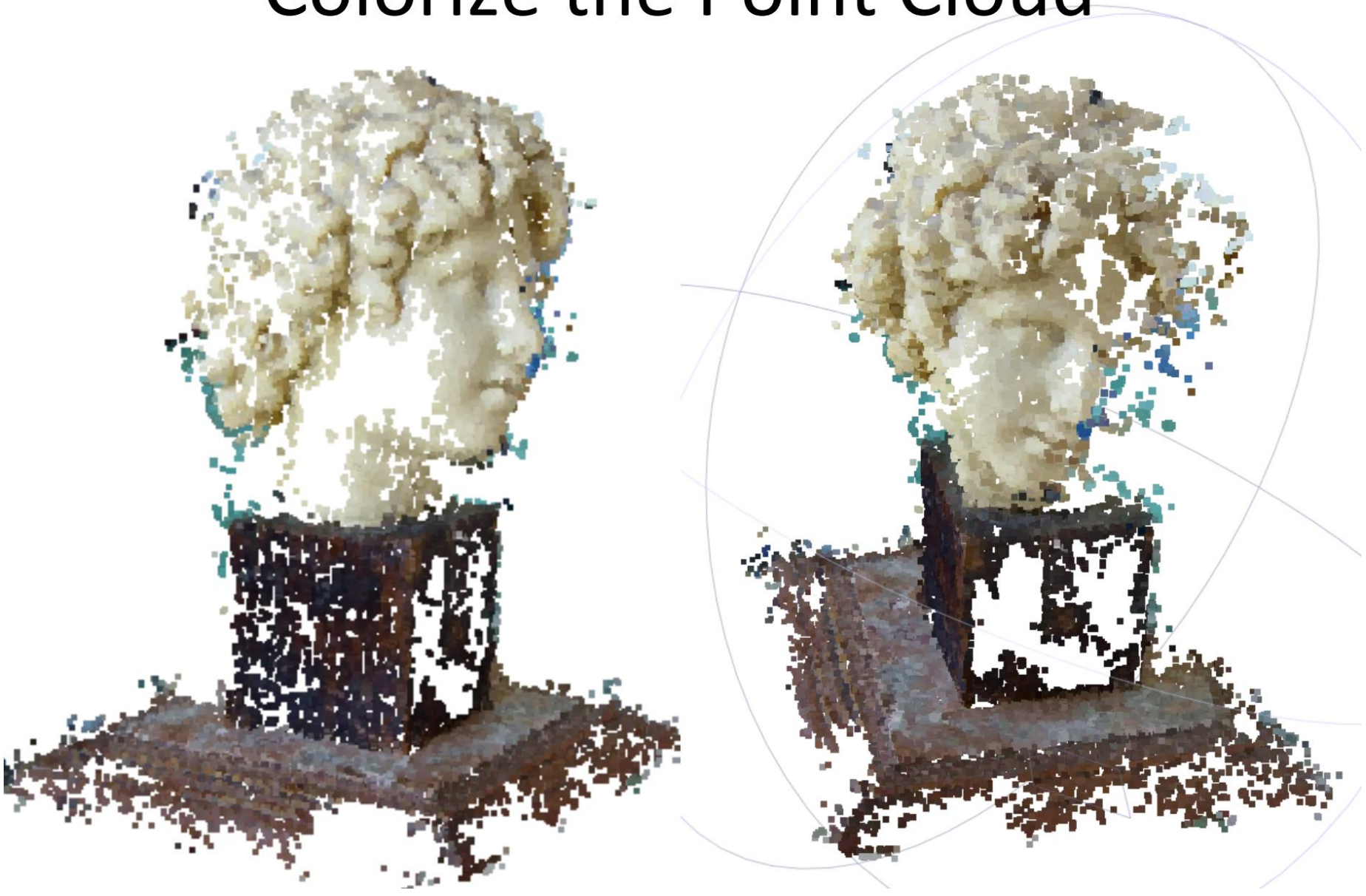
Final Result





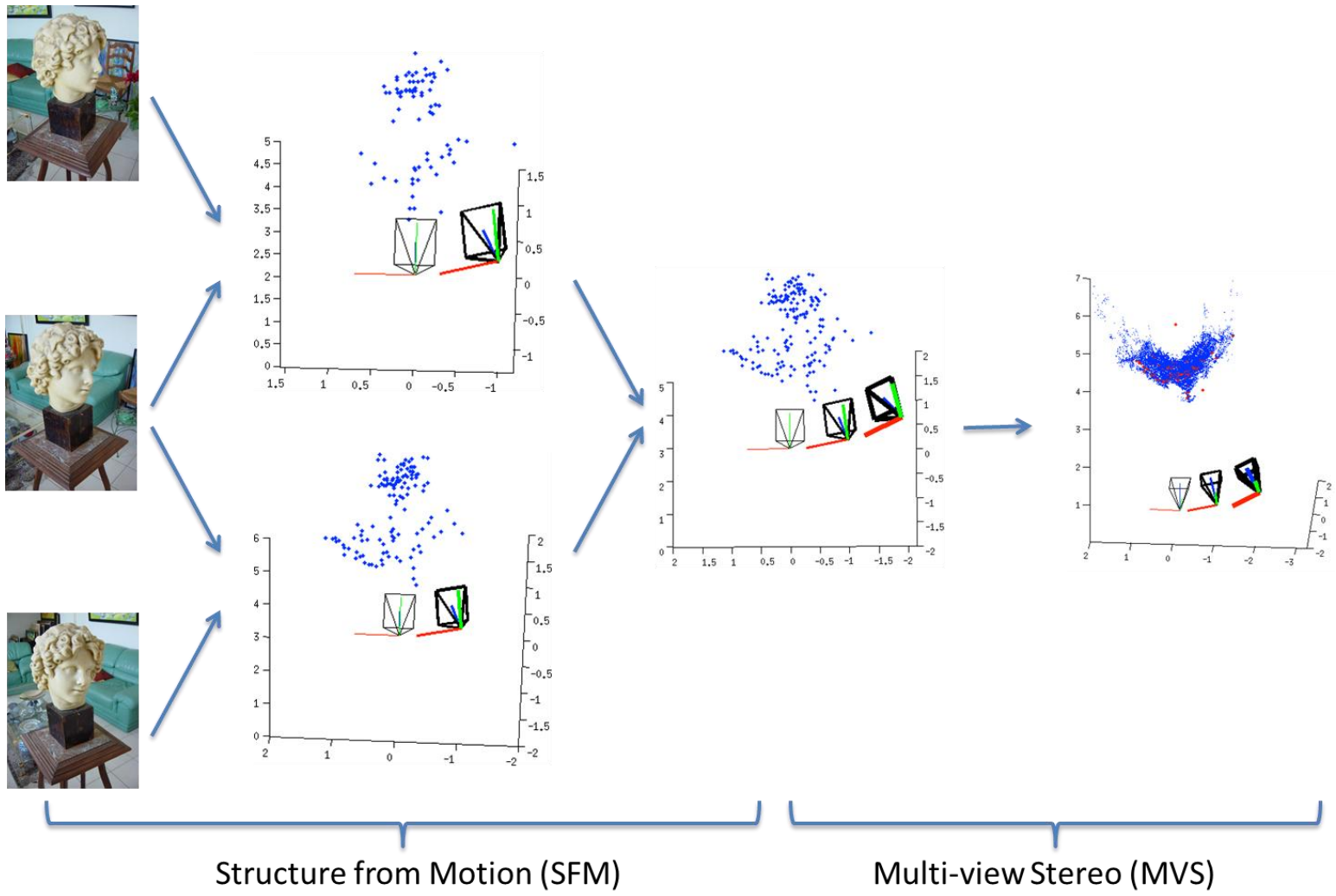


Colorize the Point Cloud



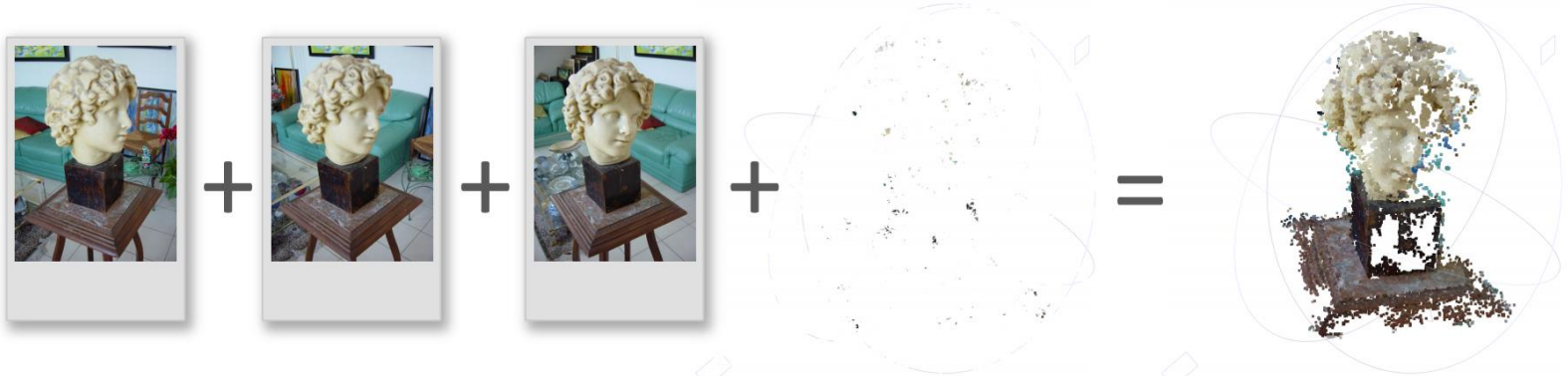
Slides from Jianxiong Xiao

Recap: SfM + MVS



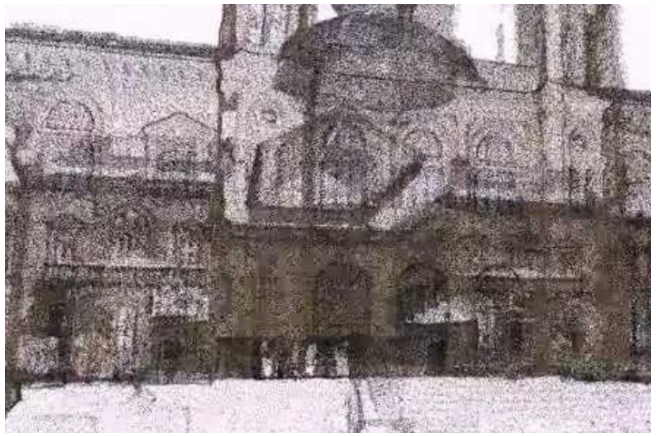
Major Steps of Image-based 3D Modeling

- Images → Sparse Points
  - SfM
- Images → Dense Points
  - Multiple View Stereo



Major Steps of Image-based 3D Modeling

- Images → Sparse Points
  - SfM
- Images → Dense Points
  - Multiple View Stereo
- Points → Meshes
  - Mesh Generation
- Meshes → Textured Meshes
  - Texture Mapping



https://www.youtube.com/watch?v=dXeXtj0PPVI

Next Week

++ Graph-based SLAM

+ SLAM Formalism (state, observation, error/observation model)

* 1D SLAM example

+ A probabilistic perspective of SLAM

+ Parallel Tracking And Mapping (PTAM)

+ Deep-learning-based SLAM and NeRF

*: know how to code (or how to use tools)
++: know how to derive (more than just the concept)
+: know the concept

References for Next week

- Visual SLAM Tutorial, CVPR'14
- Dellaert, Frank. "Visual SLAM Tutorial: Bundle Adjustment." (2014).
- Grisetti, Giorgio, et al. "A tutorial on graph-based SLAM." IEEE Intelligent Transportation Systems Magazine 2.4 (2010): 31-43.
- Klein, G. and Murray, D., 2007, November. Parallel tracking and mapping for small AR workspaces. In 2007 6th IEEE and ACM international symposium on mixed and augmented reality (pp. 225-234). IEEE.
- Mildenhall, B., Srinivasan, P.P., Tancik, M., Barron, J.T., Ramamoorthi, R. and Ng, R., 2021. Nerf: Representing scenes as neural radiance fields for view synthesis. Communications of the ACM, 65(1), pp.99-106.