NYU TANDON SCHOOL OF ENGINEERING

Robot Vision

SfM/vSLAM

Dr. Chen Feng
cfeng@nyu.edu

ROB-UY 3203, Spring 2024

Overview

- SfM
  - Overview
  - Globally consistent solution (Bundle adjustment)
  - Initial camera pose estimation
  - Initial object point (landmark) coordinates estimation

- vSLAM
  - A slightly different problem formulation
  - Real-time processing

References

- Forsyth & Ponce 2011
  - Chapter 8
- Szeliski 2011:
  - Chapter 11
- Corke 2022:
  - Section 14.4
- Hartley & Zisserman 2003:
  - Section 5.2, 18.1, A6

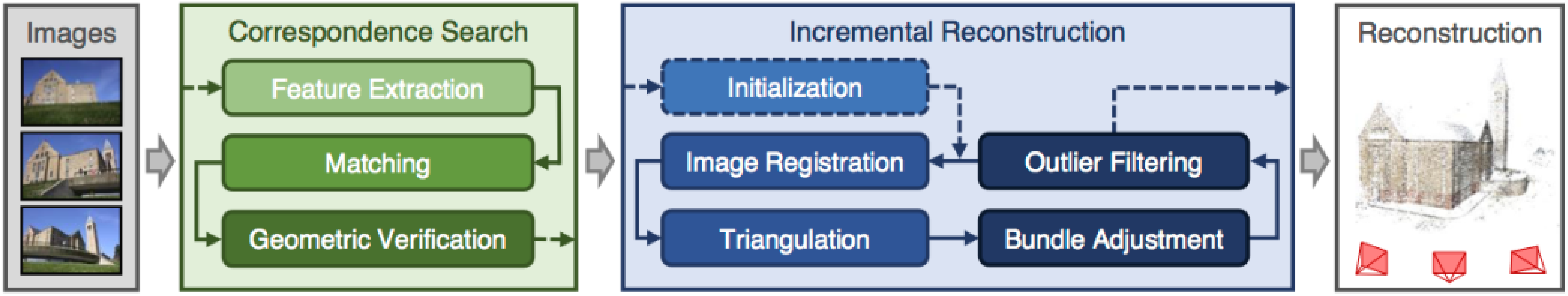
- Chen Feng, Vineet R. Kamat, and Carol C. Menassa. "Marker-Assisted Structure from Motion for 3D Environment Modeling and Object Pose Estimation." *(2016)*.

SfM: Structure from Motion

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

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

SfM Pipeline



COLMAP's incremental Structure-from-Motion pipeline.

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

SfM Pipeline – Simplified

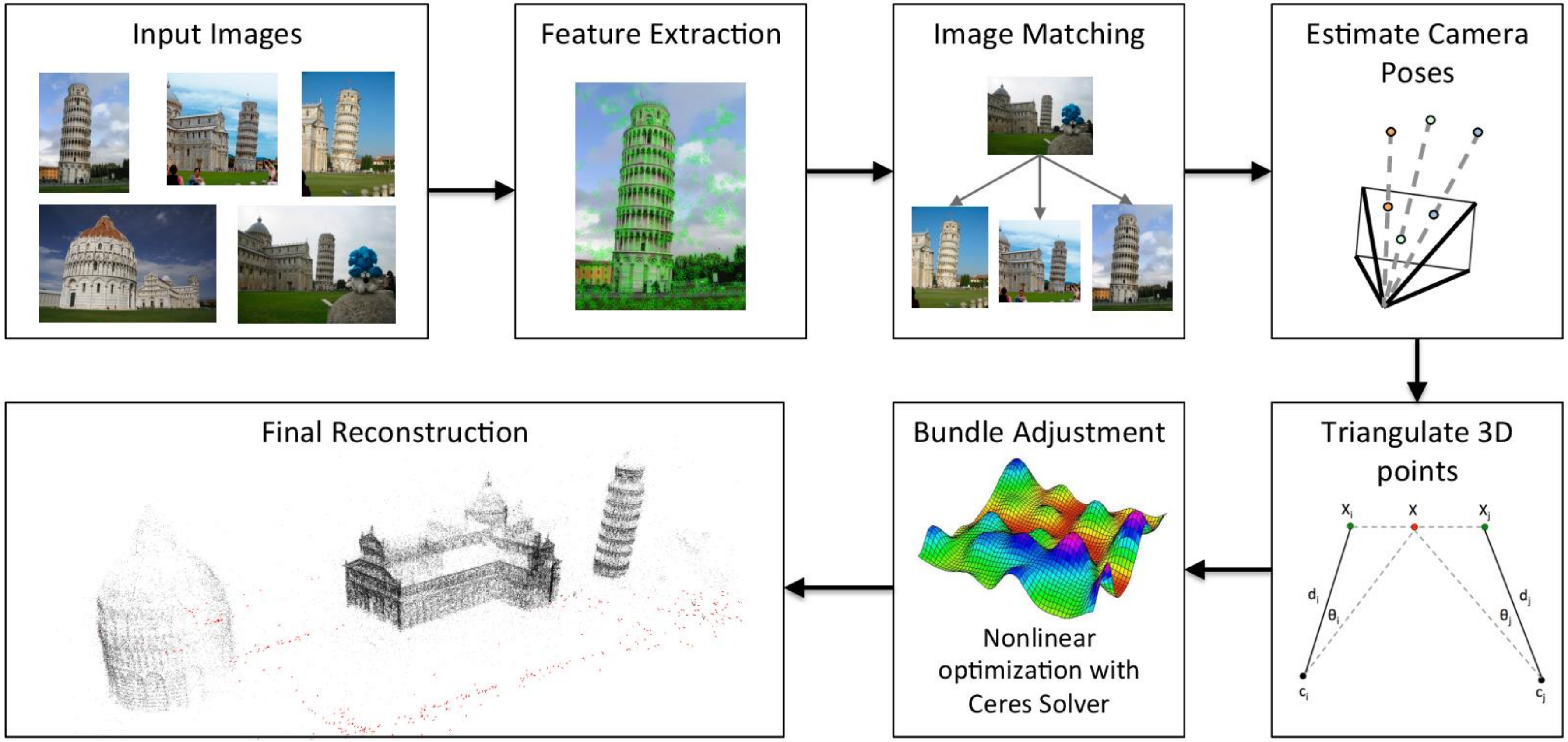


Image from: http://www.theia-sfm.org/sfm.html

SfM Pipeline – the Core

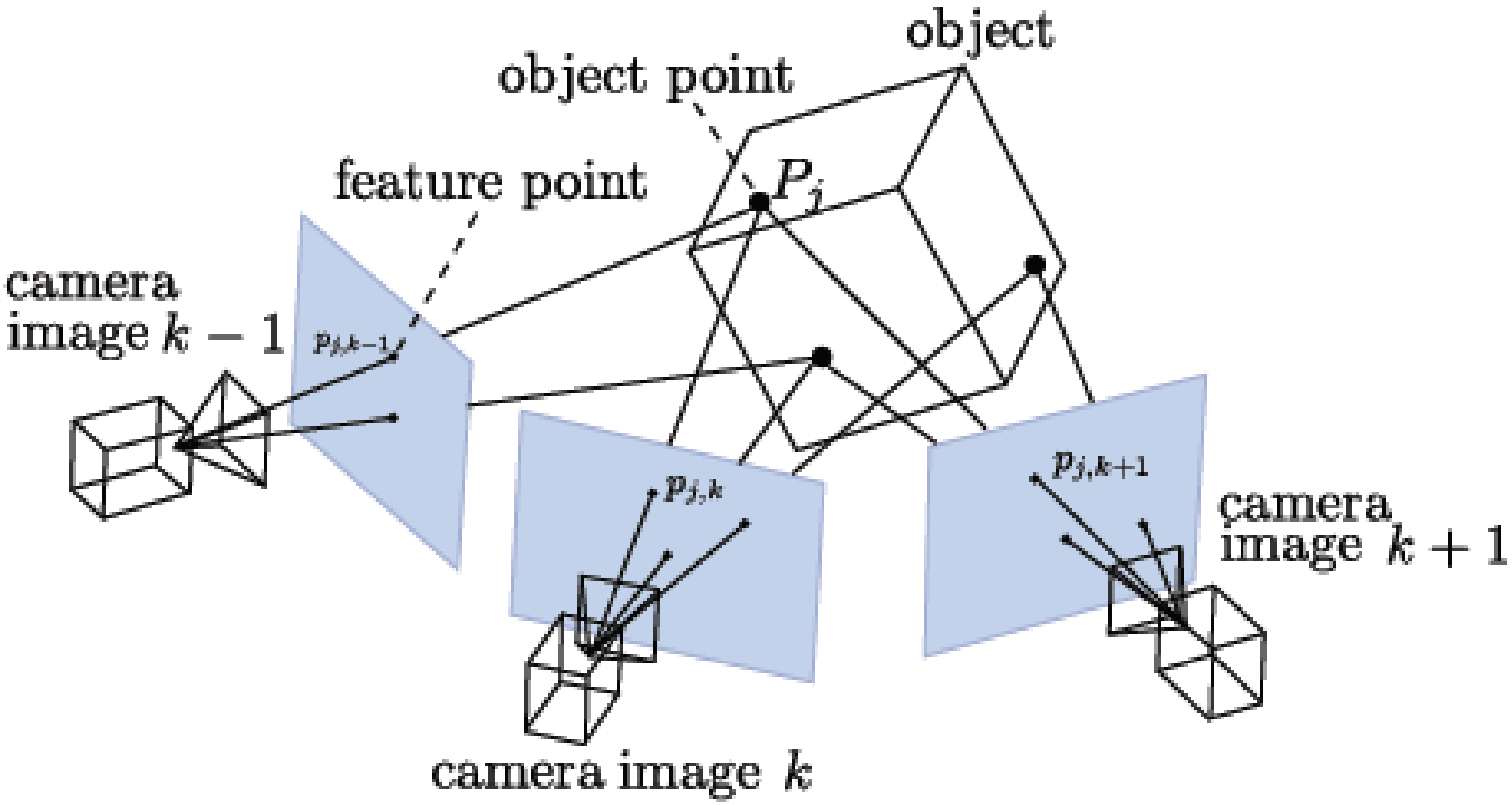


Image from:
https://openmvg.readthedocs.io/en/latest/openMVG/sfm/sfm/

SfM Problem Summary

- Input: Observed 2D image position

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

- Output:

Unknown Camera Parameters (with some guess)

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

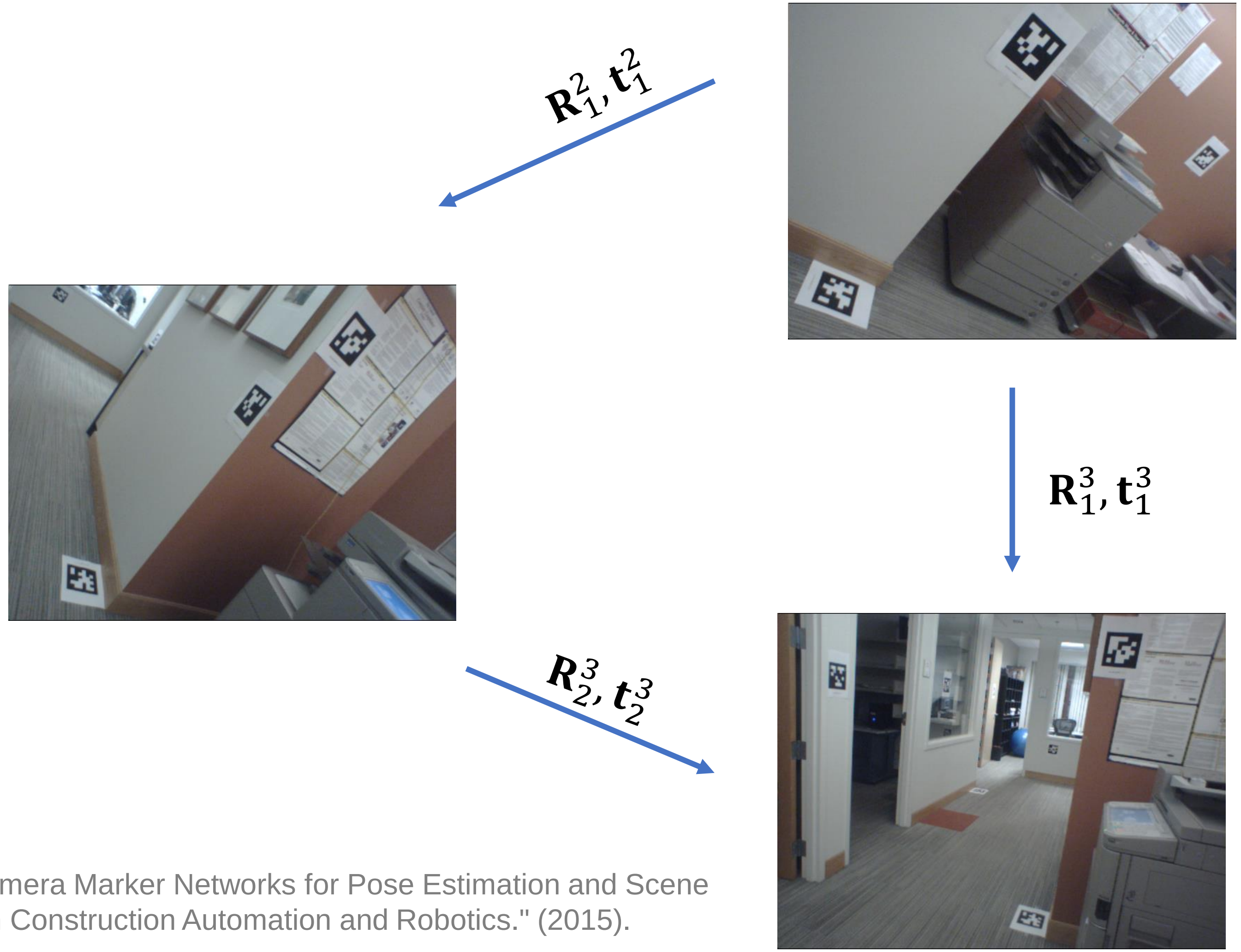
Unknown Point 3D coordinate (with some guess)

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

Slides modified from Jianxiong Xiao

Inconsistency across Multiple Relative Pose Estimations

- Recall relative pose estimation

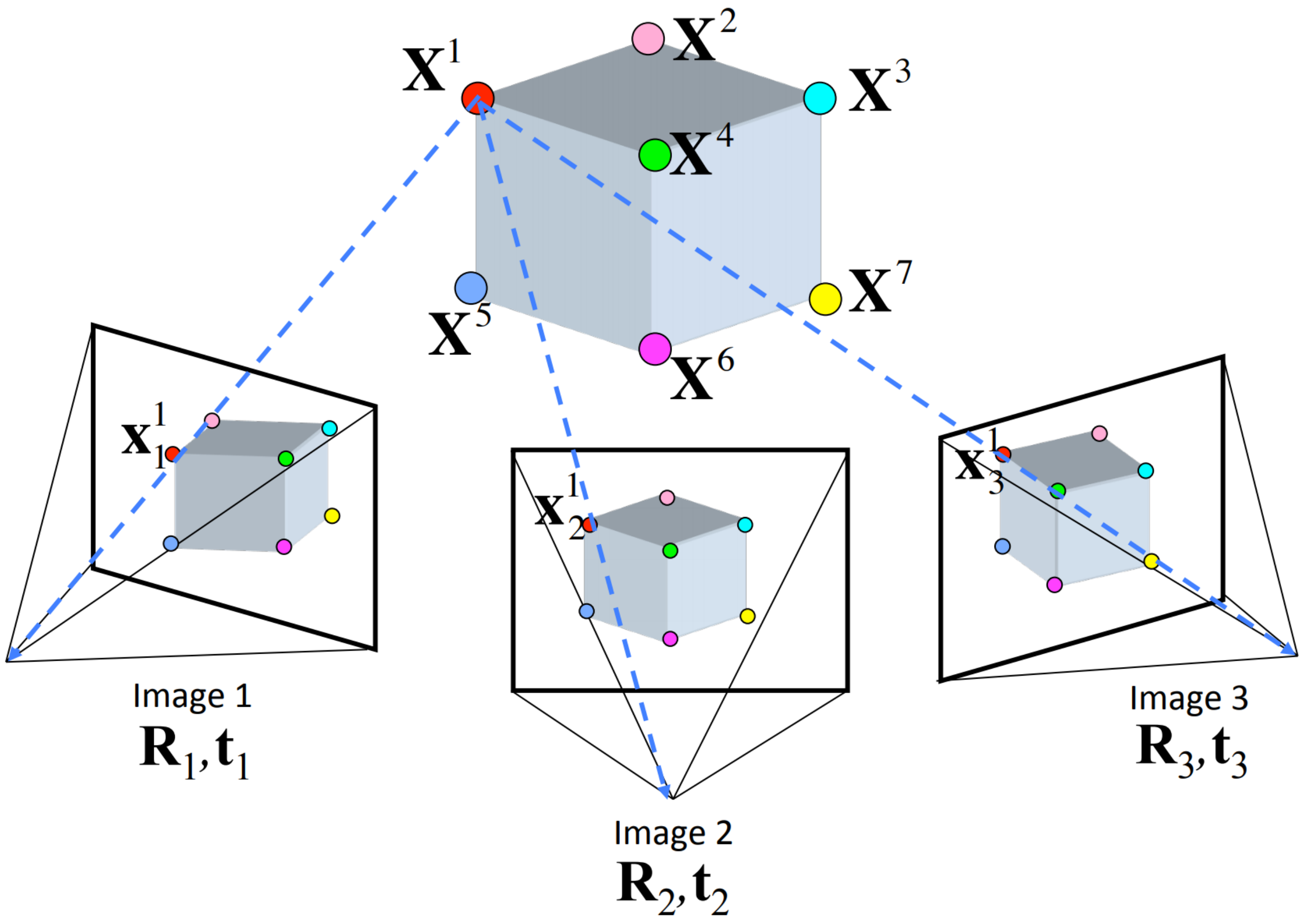

$\mathbf{R}_1^2, \mathbf{t}_1^2$
$\mathbf{R}_1^3, \mathbf{t}_1^3$
$\mathbf{R}_2^3, \mathbf{t}_2^3$

$$\begin{bmatrix} \mathbf{R}_1^3 & \mathbf{t}_1^3 \\ 0 & 1 \end{bmatrix} \overset{?}{=} \begin{bmatrix} \mathbf{R}_2^3 & \mathbf{t}_2^3 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} \mathbf{R}_1^2 & \mathbf{t}_1^2 \\ 0 & 1 \end{bmatrix}$$

How can we obtain
a **globally consistent** solution?

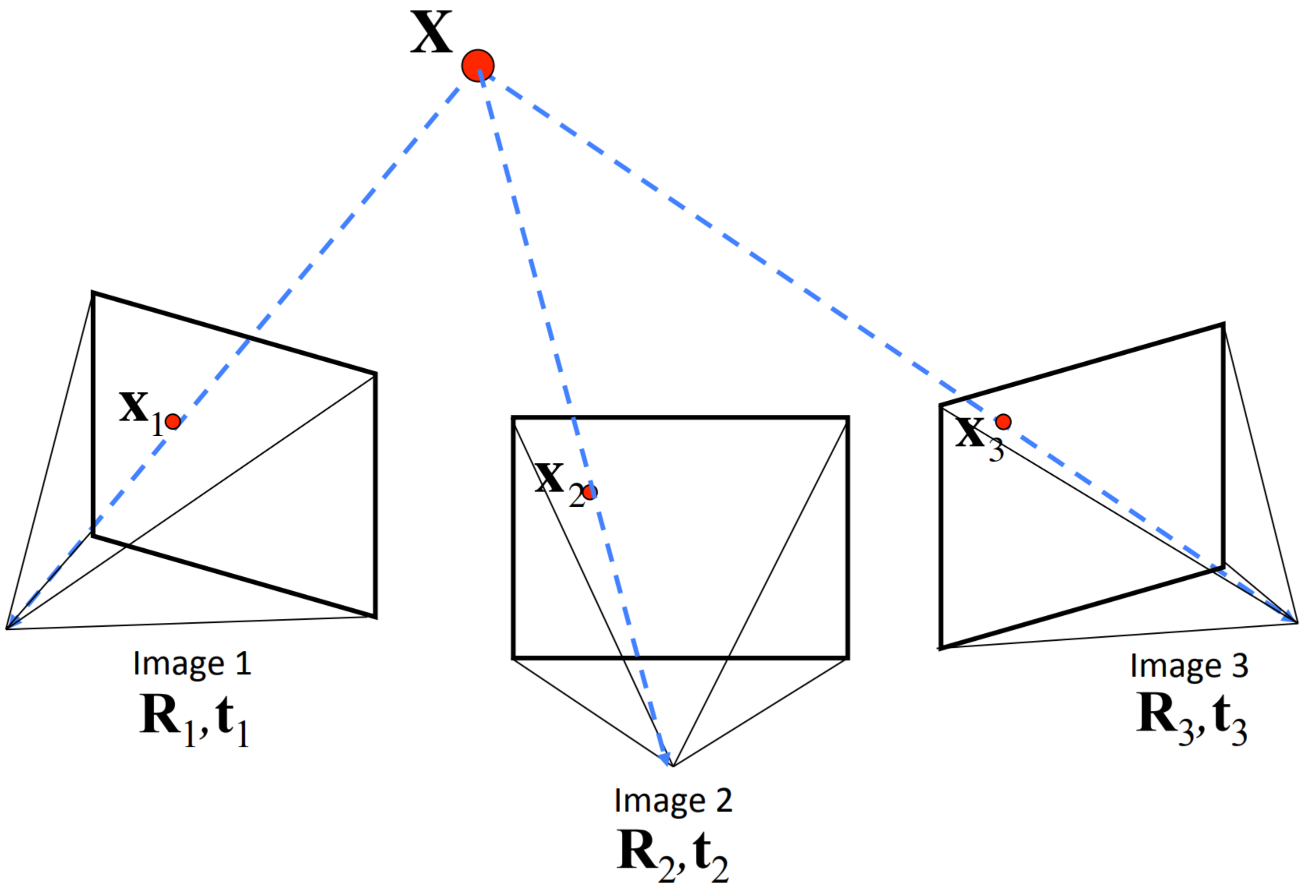
Feng, Chen. "Camera Marker Networks for Pose Estimation and Scene Understanding in Construction Automation and Robotics." (2015).

Estimate All Unknowns Together



Slides modified from Jianxiong Xiao

The Math Model of Multiple Photos



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

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

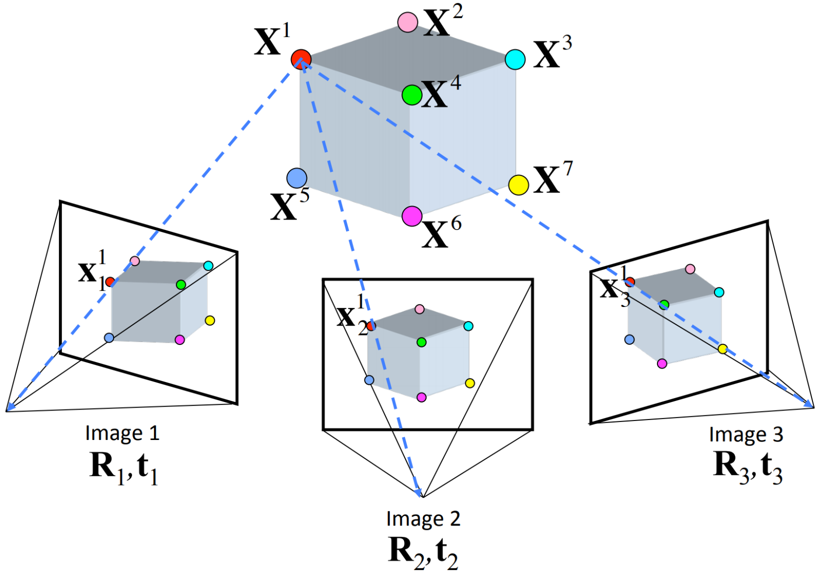
$$\mathbf{x}_3 = \mathbf{K}\left[\mathbf{R}_3 | \mathbf{t}_3\right]\mathbf{X}$$

Slides modified from Jianxiong Xiao

Write down all Equations

$$\begin{cases} \mathbf{x}_1^1 = \mathbf{K}[\mathbf{R}_1|\mathbf{t}_1]\mathbf{X}^1 \\ \mathbf{x}_1^2 = \mathbf{K}[\mathbf{R}_1|\mathbf{t}_1]\mathbf{X}^2 \\ \vdots \\ \mathbf{x}_i^j = \mathbf{K}[\mathbf{R}_i|\mathbf{t}_i]\mathbf{X}^j \\ \vdots \end{cases}$$

Unknowns: $\mathbf{R}_i$, $\mathbf{t}_i$, $\mathbf{X}^j$



	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

Bundle Adjustment

- Recall the solution for the linear equation system:

$$\mathbf{A}\mathbf{x} = \mathbf{b}$$

$$\mathbf{x} = \left(\mathbf{A}^T\mathbf{A}\right)^{-1}\mathbf{A}^T\ \mathbf{b}$$

- Now we want to solve the nonlinear equation system

$$\begin{cases} \mathbf{x}_1^1 = \mathbf{K}[\mathbf{R}_1|\mathbf{t}_1]\mathbf{X}^1 \\ \mathbf{x}_1^2 = \mathbf{K}[\mathbf{R}_1|\mathbf{t}_1]\mathbf{X}^2 \\ \vdots \\ \mathbf{x}_i^j = \mathbf{K}[\mathbf{R}_i|\mathbf{t}_i]\mathbf{X}^j \\ \vdots \end{cases}$$

$$F(\mathbf{x}) = \mathbf{b}$$

$$\mathbf{x} = \left(\mathbf{R}_1, \mathbf{R}_2, \dots, \mathbf{R}_i, \dots, \boldsymbol{t}_1, \boldsymbol{t}_2, \dots, \boldsymbol{t}_i, \dots, \mathbf{X}^1, \mathbf{X}^1, \dots \mathbf{X}^j \dots\right)^T$$

$$\mathbf{b} = \left(\mathbf{x}_1^1, \mathbf{x}_1^2, \dots, \mathbf{x}_i^j\right)^T$$

$$\hat{\mathbf{x}} = \underset{\mathbf{x}}{\operatorname{argmin}} ||F(\mathbf{x}) - \mathbf{b}||^2$$ No close-form solution!

Solution to Bundle Adjustment

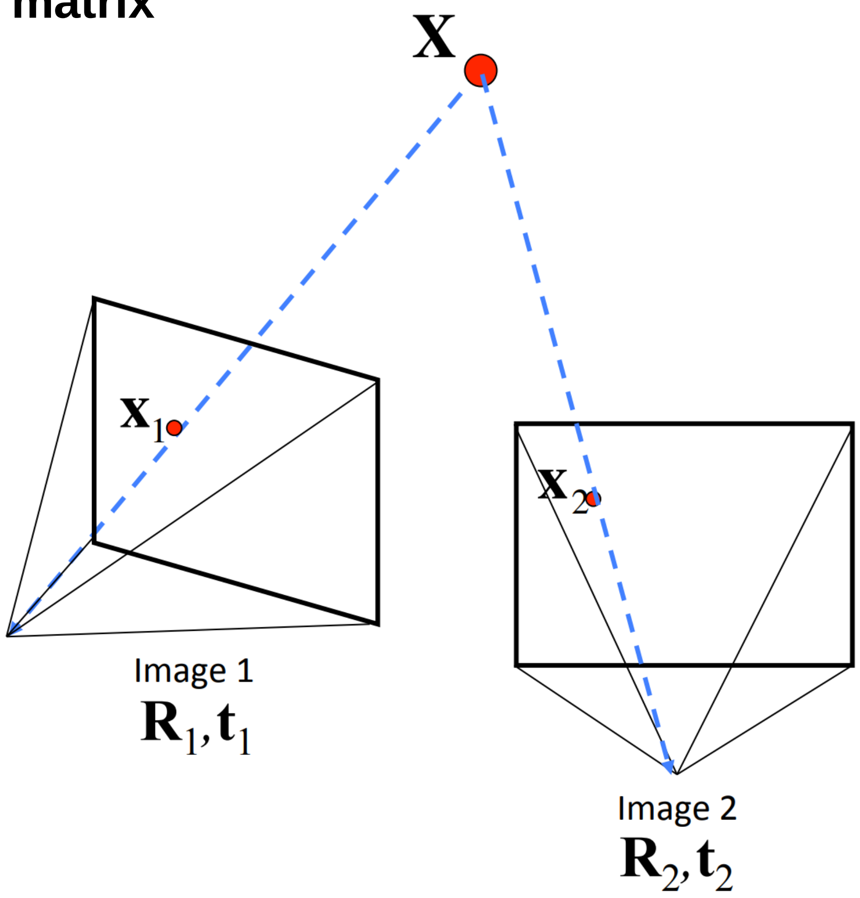
- Solve by Levenberg-Marquardt algorithm
  - Ceres (Agarwal and Mierle 2012)
  - g2o (Kummerle et al. 2011)

- Requires the initial values of $\mathbf{x}$
  - Initial values need to be close to the optimal values.
  - Bad initialization may cause the estimated values far from the optimal values.

- Camera poses can be initialized via relative camera poses.
- Landmark coordinates can be initialized via **triangulation**.

Initial Camera Pose Estimation (Relative Pose)

Estimate fundamental matrix



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

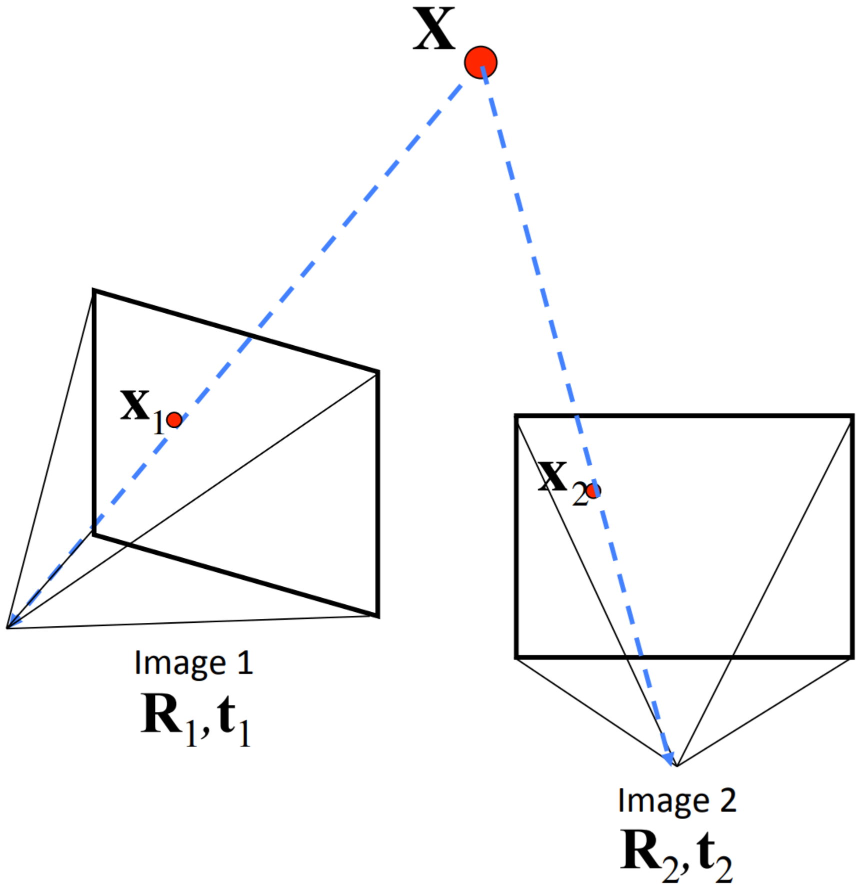
$$\mathbf{x}_1^T \mathbf{F} \mathbf{x}_2 = 0$$

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

Slides from Jianxiong Xiao

Initial Camera Pose Estimation (Relative Pose)

Fundamental matrix to essential matrix



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

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

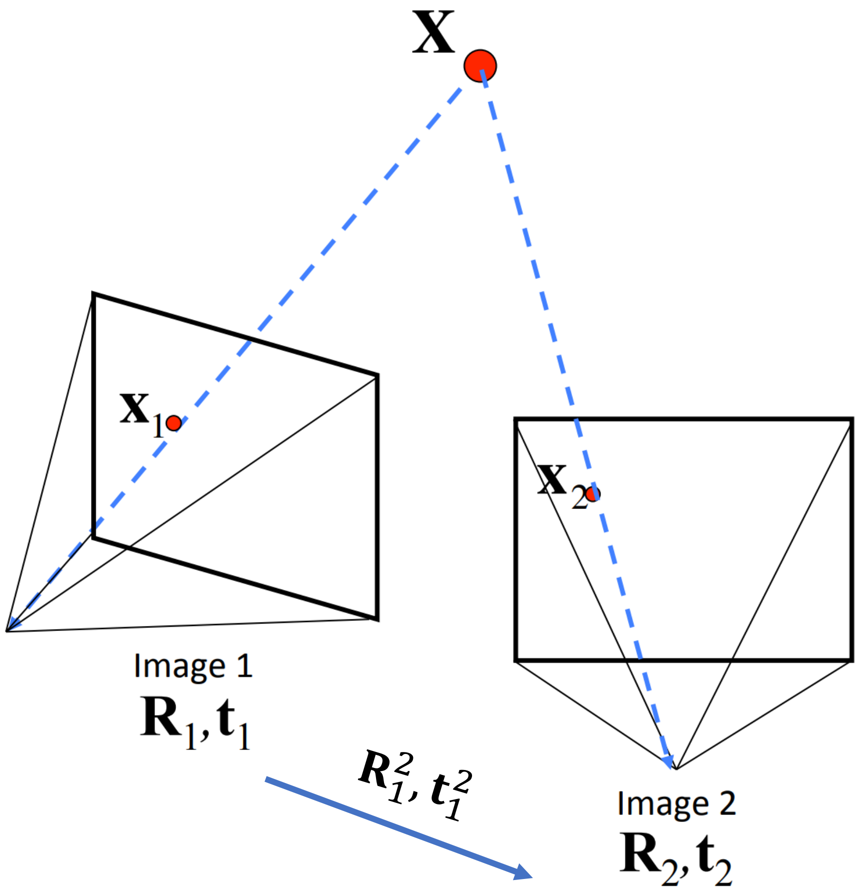
$$\mathbf{E} = \mathbf{K}_1^T \mathbf{F} \mathbf{K}_2$$

E-matrix are estimated by F-matrix with camera intrinsic parameters

Slides from Jianxiong Xiao

Initial Camera Pose Estimation (Relative Pose)

Relative camera pose estimation from essential matrix (recall previous classes)



$$\mathrm{E} = [\mathrm{t}]_\times \mathrm{R}$$

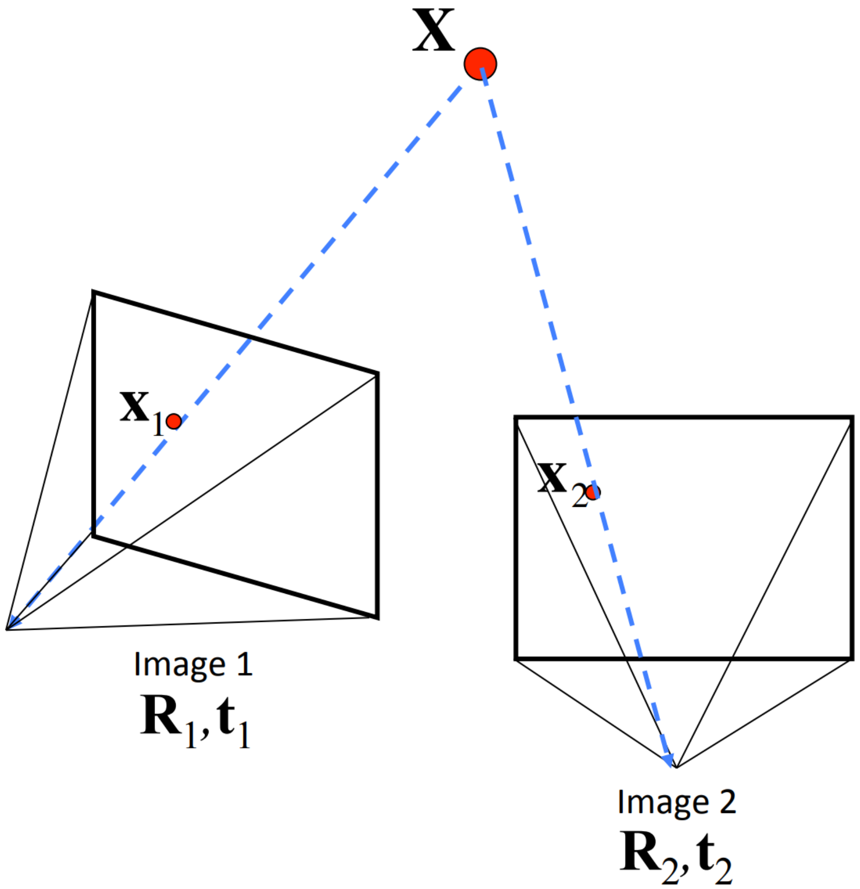
Assume $\mathbf{R}_1 = \mathbf{I}, \mathbf{t}_1 = \mathbf{0}$

$$\mathbf{R}_2 = \mathbf{R}_1^2, \mathbf{t}_2 = \mathbf{t}_1^2$$

Triangulation

- Solve X=?

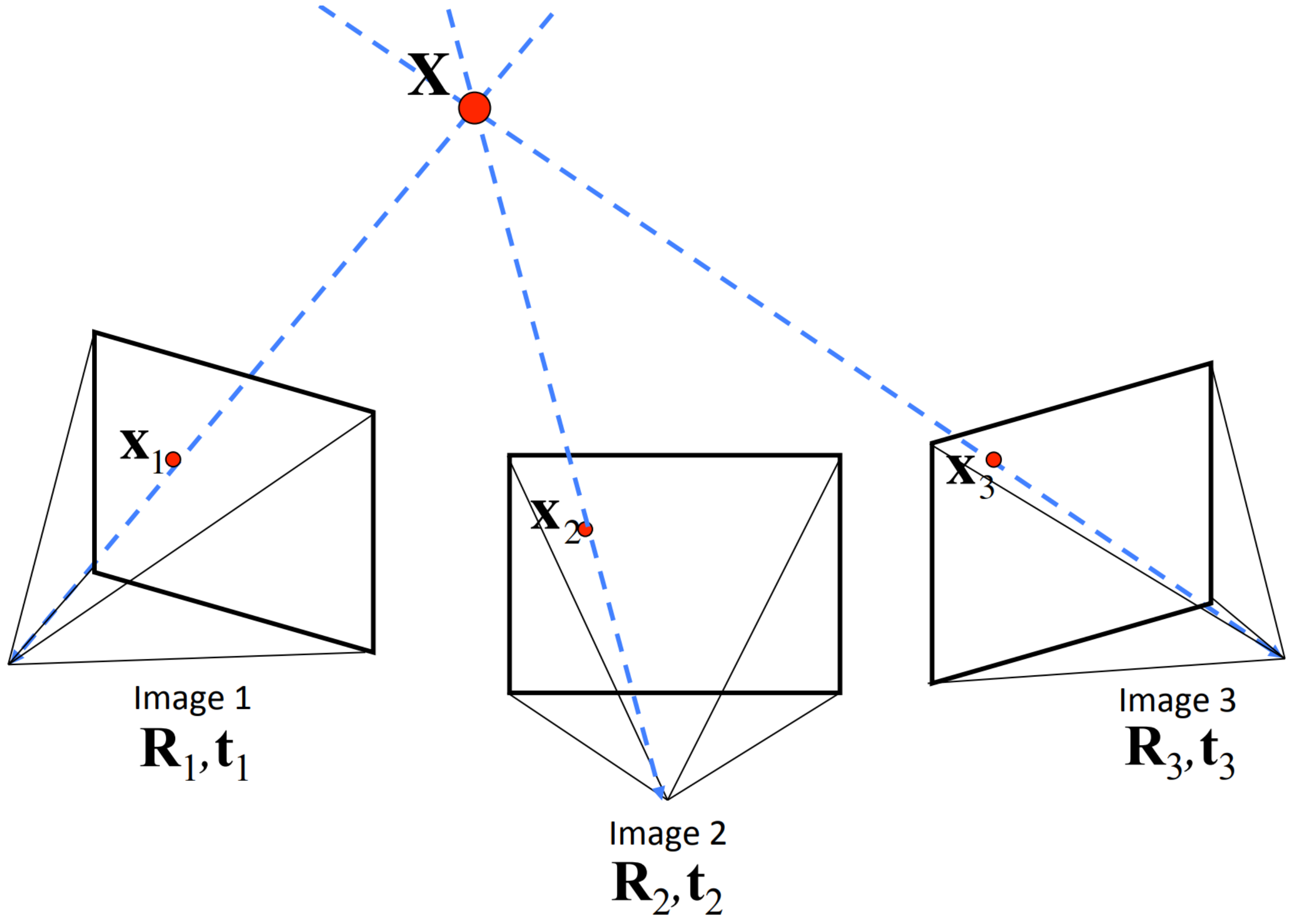
$$\begin{cases} \mathbf{x_1} = \mathbf{K}[\mathbf{R}_1|\mathbf{t}_1]\mathbf{X} \\ \mathbf{x_2} = \mathbf{K}[\mathbf{R}_2|\mathbf{t}_2]\mathbf{X} \end{cases}$$



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

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

Triangulation from Multiple Views

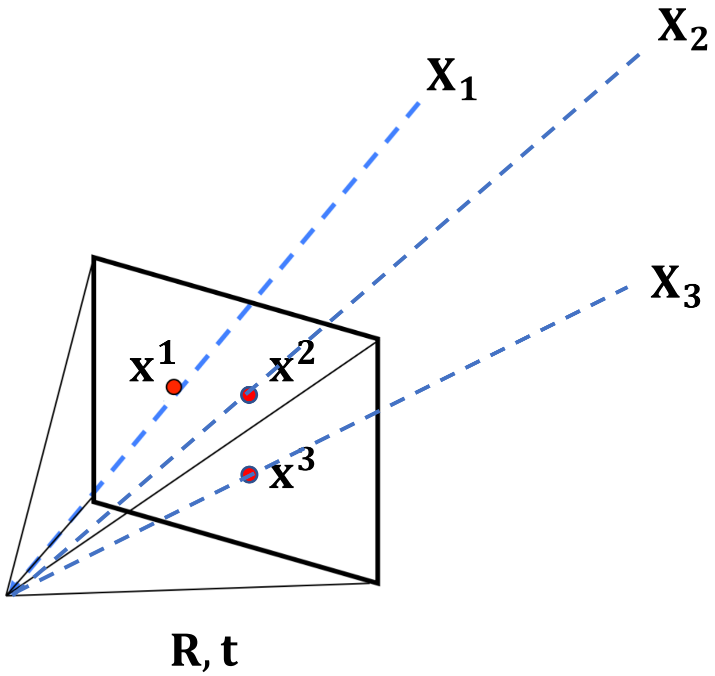

$\mathbf{X}$
$\mathbf{x}_1$
$\mathbf{x}_2$
$\mathbf{x}_3$
Image 1
$\mathbf{R}_1, \mathbf{t}_1$
Image 2
$\mathbf{R}_2, \mathbf{t}_2$
Image 3
$\mathbf{R}_3, \mathbf{t}_3$

Slides from Jianxiong Xiao

Initial Camera Pose Estimation (Perspective-n-Point)

- Solve **R**, **t** =?

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



Slides from Jianxiong Xiao

Camera Pose and Landmark Coordinates Initialization

1. Initialize two camera poses via essential matrix;
2. Triangulate landmark coordinates using initialized camera poses;
3. Initialize new camera poses via solving Perspective-n-Point (PnP);
4. Repeat 2-3 until all camera poses and landmark coordinates are initialized.

Slides from Jianxiong Xiao

Bundle Adjustment: Take Home Message

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

$$\text{Re-projection}\left\{\begin{array}{lll} \mathbf{x}_1^1=\mathbf{K}[\mathbf{R}_1|\mathbf{t}_1]\mathbf{X}^1 & \mathbf{x}_1^2=\mathbf{K}[\mathbf{R}_1|\mathbf{t}_1]\mathbf{X}^2 & \\ \mathbf{x}_2^1=\mathbf{K}[\mathbf{R}_2|\mathbf{t}_2]\mathbf{X}^1 & \mathbf{x}_2^2=\mathbf{K}[\mathbf{R}_2|\mathbf{t}_2]\mathbf{X}^2 & \mathbf{x}_2^3=\mathbf{K}[\mathbf{R}_2|\mathbf{t}_2]\mathbf{X}^3 \\ \mathbf{x}_3^1=\mathbf{K}[\mathbf{R}_3|\mathbf{t}_3]\mathbf{X}^1 & & \mathbf{x}_3^3=\mathbf{K}[\mathbf{R}_3|\mathbf{t}_3]\mathbf{X}^3 \end{array}\right.$$

$$=$$

$$\text{Observation}\left\{\begin{array}{lll} \tilde{\mathbf{x}}_1^1 & \tilde{\mathbf{x}}_1^2 & \\ \tilde{\mathbf{x}}_2^1 & \tilde{\mathbf{x}}_2^2 & \tilde{\mathbf{x}}_2^3 \\ \tilde{\mathbf{x}}_3^1 & & \tilde{\mathbf{x}}_3^3 \end{array}\right.$$

Slides from Jianxiong Xiao

Bundle Adjustment == Least Squares

A valid solution $\left[\mathbf{R}_1 | \mathbf{t}_1\right], \left[\mathbf{R}_2 | \mathbf{t}_2\right], \left[\mathbf{R}_3 | \mathbf{t}_3\right]$ 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

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

Slides from Jianxiong Xiao

vSLAM References

- 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.

SLAM as Estimation

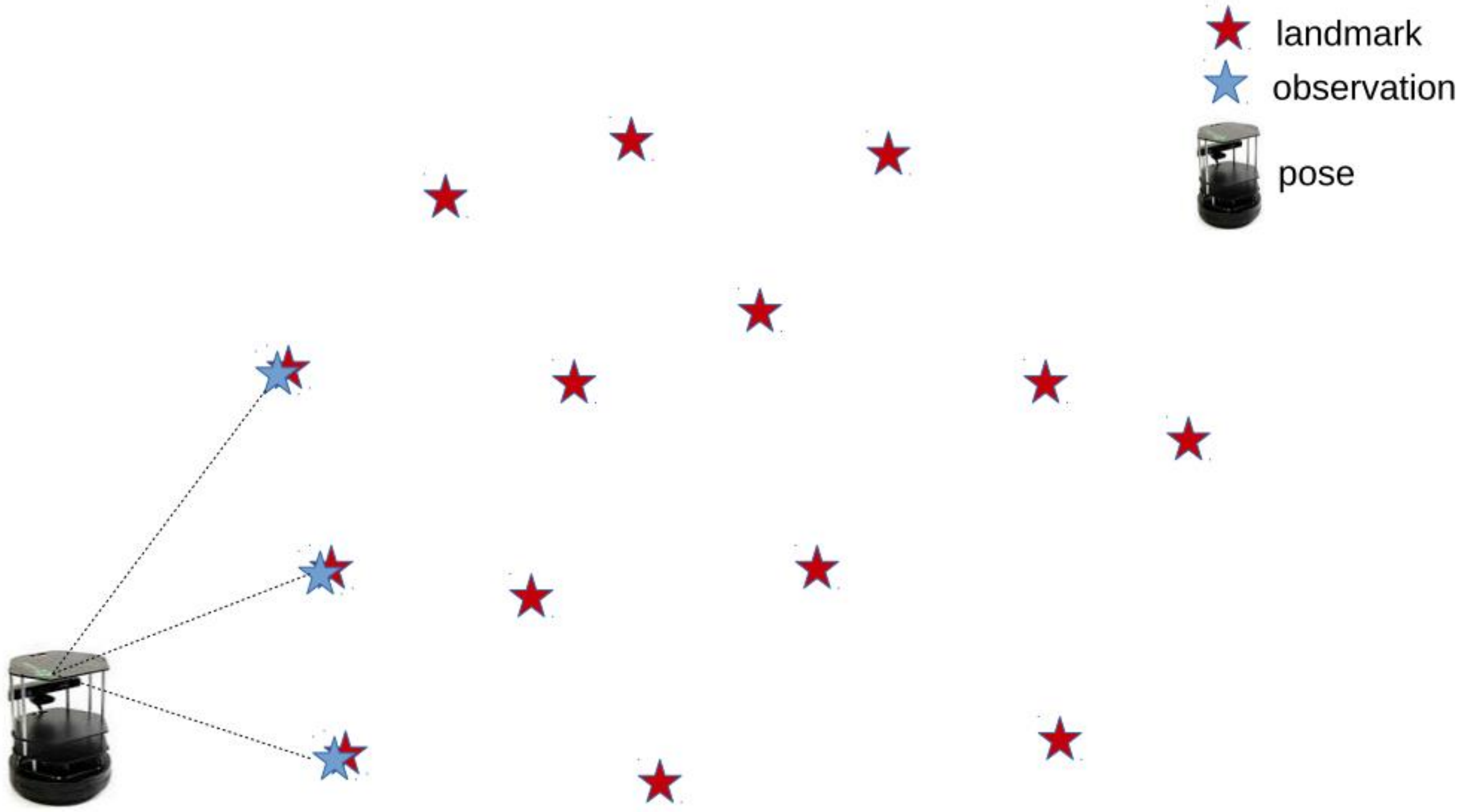


Image from Giorgio Grisetti 2016

SLAM as Estimation

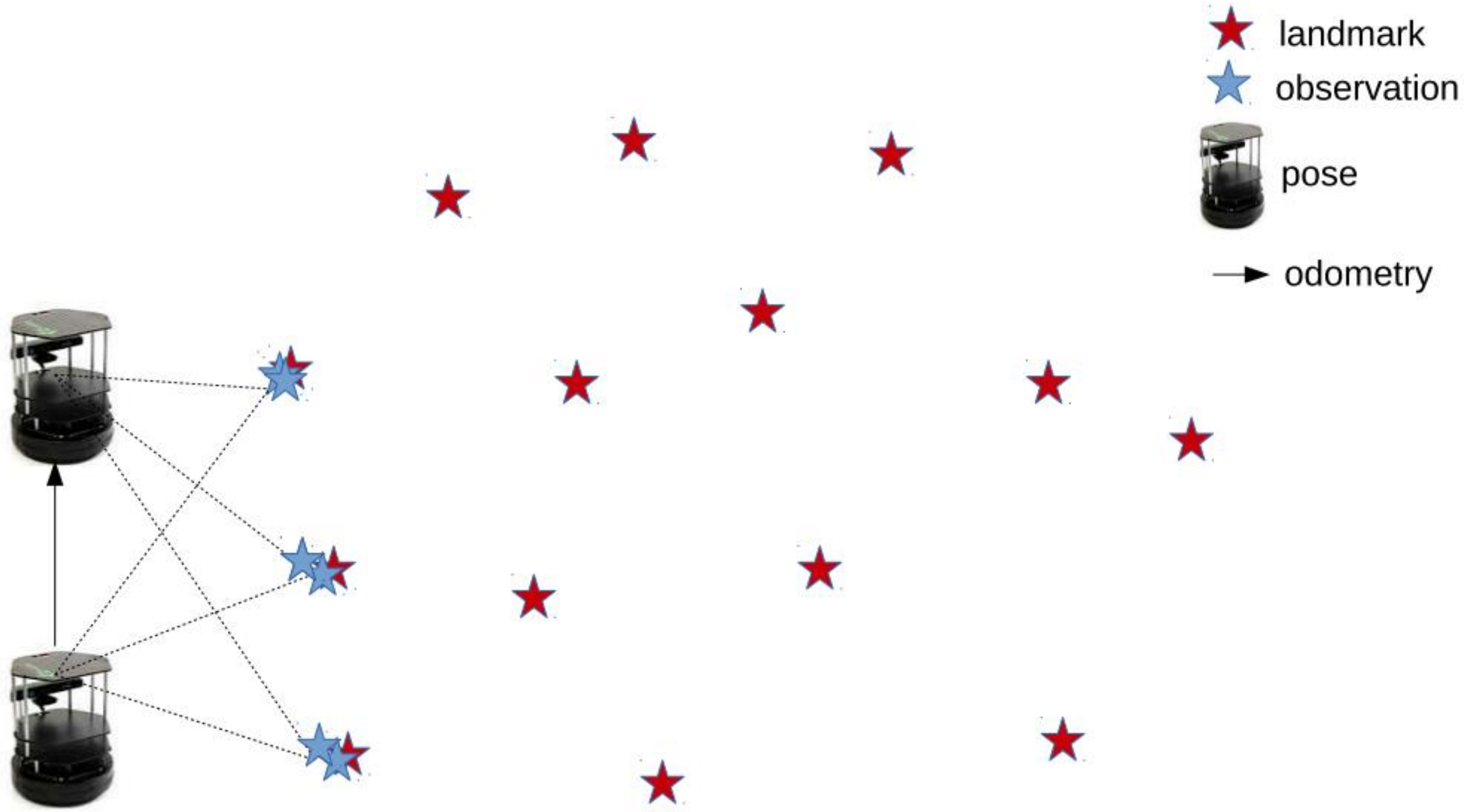


Image from Giorgio Grisetti 2016

SLAM as Estimation

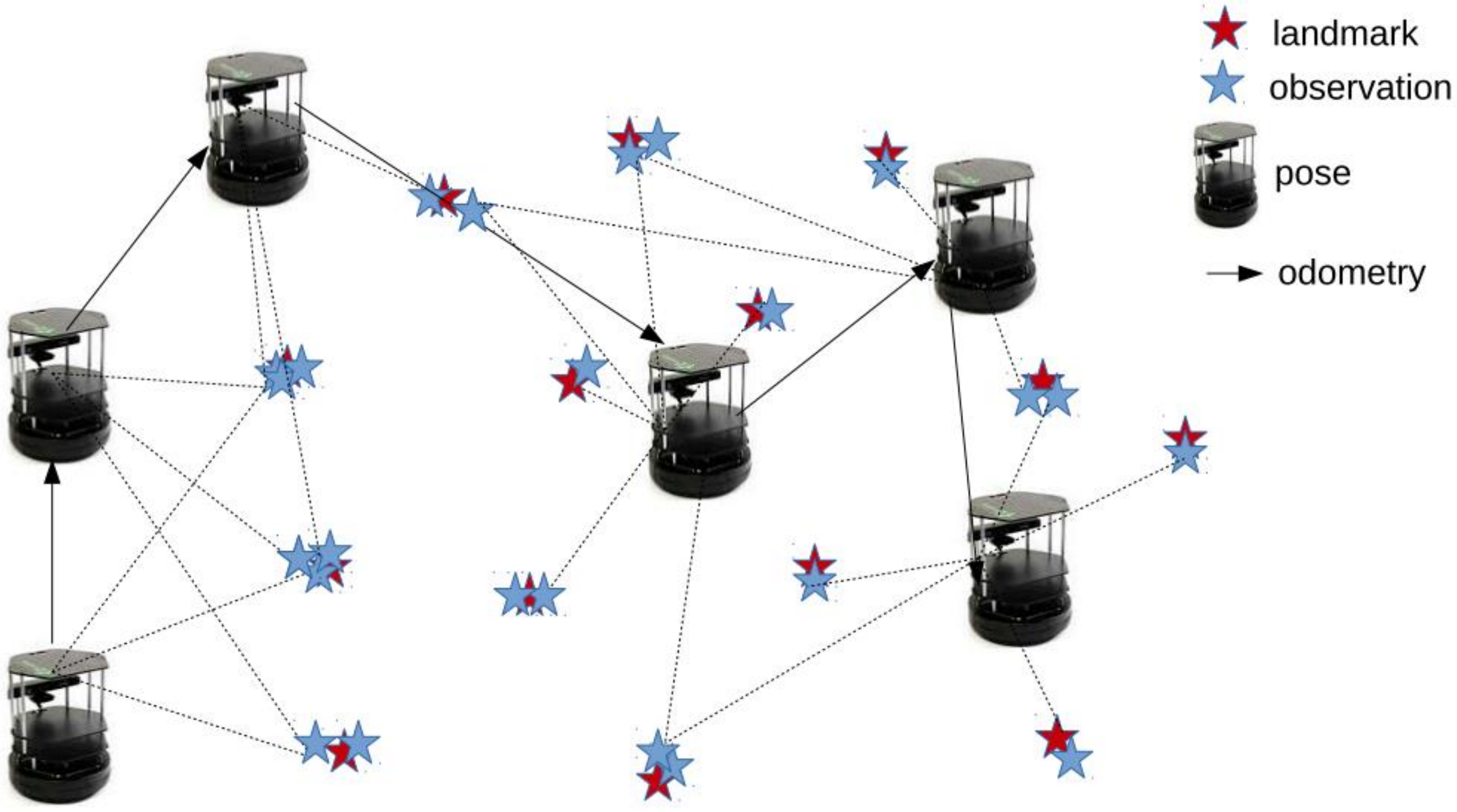


Image from Giorgio Grisetti 2016

SLAM as Estimation

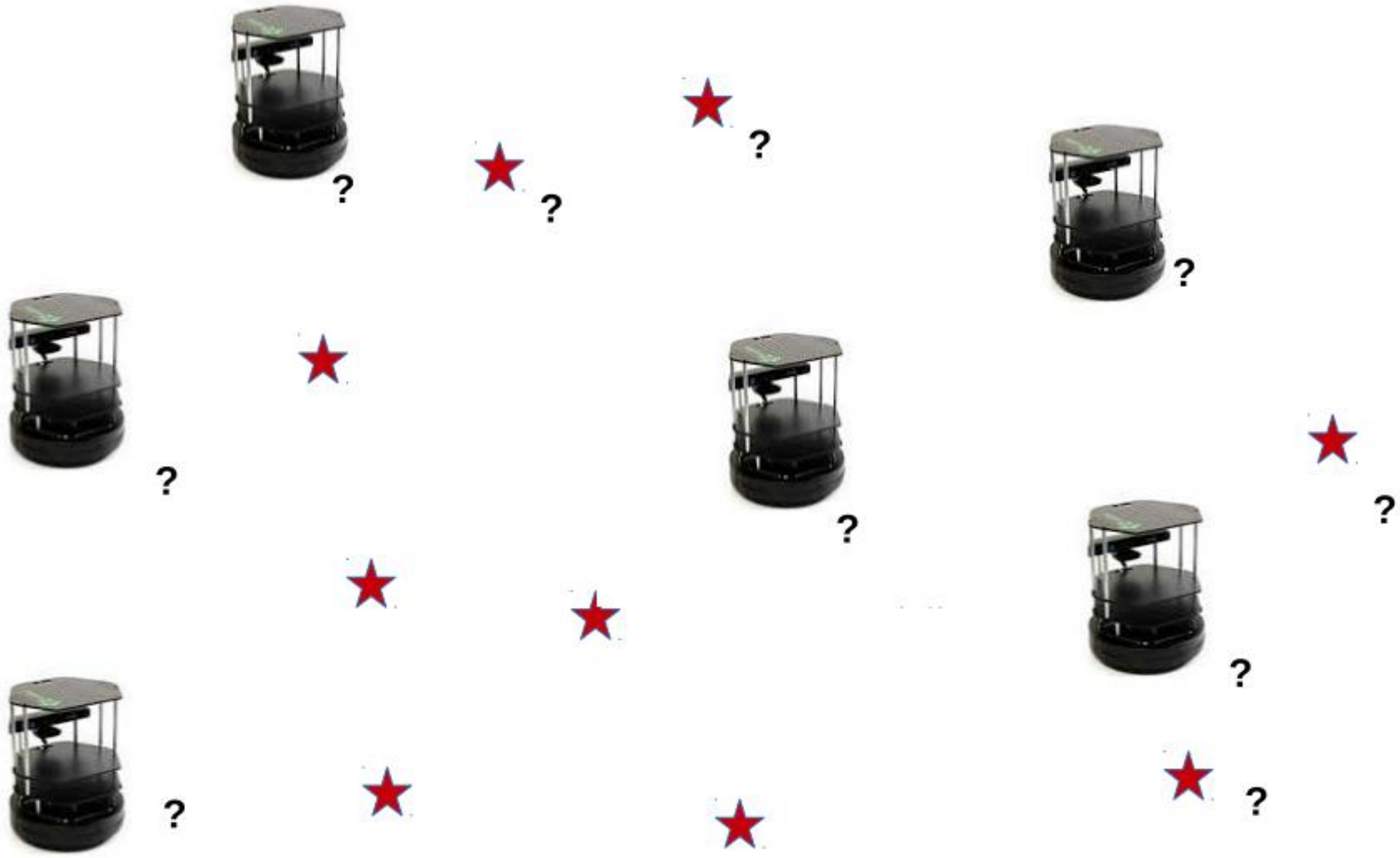
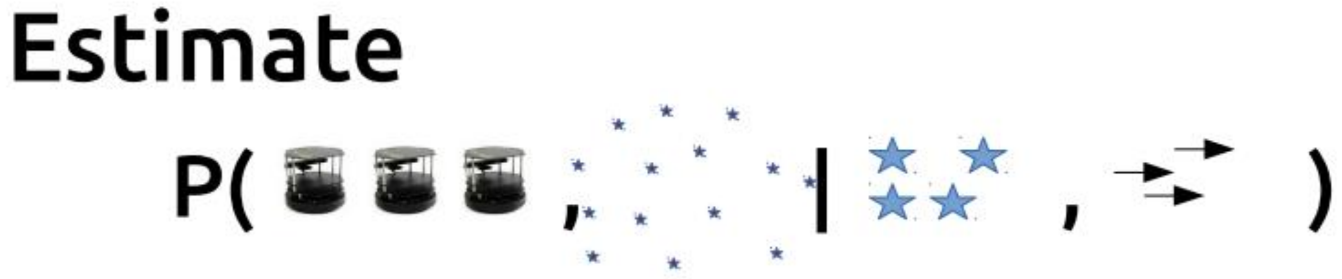


Image from Giorgio Grisetti 2016

SLAM from a Probabilistic Perspective

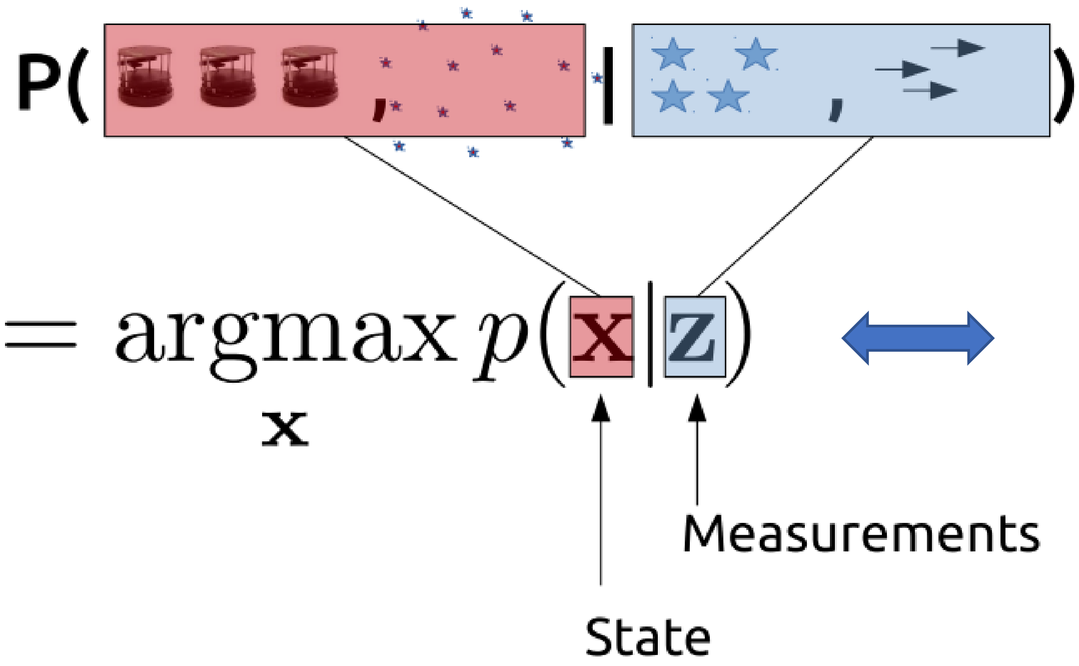


$$\mathbf{x}^* = \operatorname*{argmax}_{\mathbf{x}}\, p(\mathbf{x}|\mathbf{z})$$

Slide from Giorgio Grisetti 2016

SLAM as Maximum Likelihood Estimation (MLE)

Estimate



$$\mathbf{x}^* = \underset{\mathbf{x}}{\operatorname{argmax}}\, p(\boxed{\mathbf{x}}|\boxed{\mathbf{z}}) \quad \longleftrightarrow \quad \hat{\mathbf{x}} = \underset{\mathbf{x}}{\operatorname{argmin}} ||F(\mathbf{x}) - \boldsymbol{z}||^2$$

Measurements

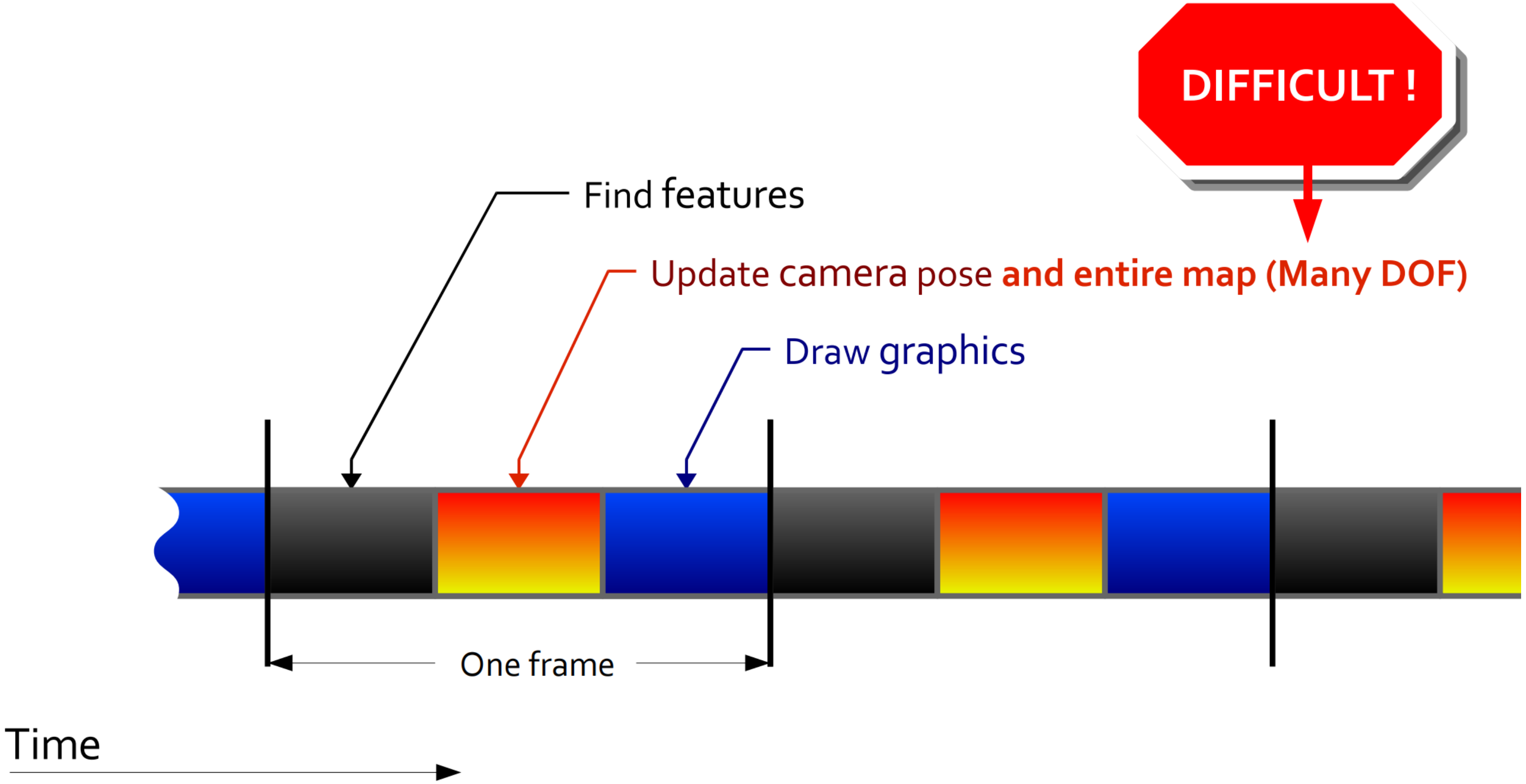
State

x*: state most consistent with observations

Slide from Giorgio Grisetti 2016

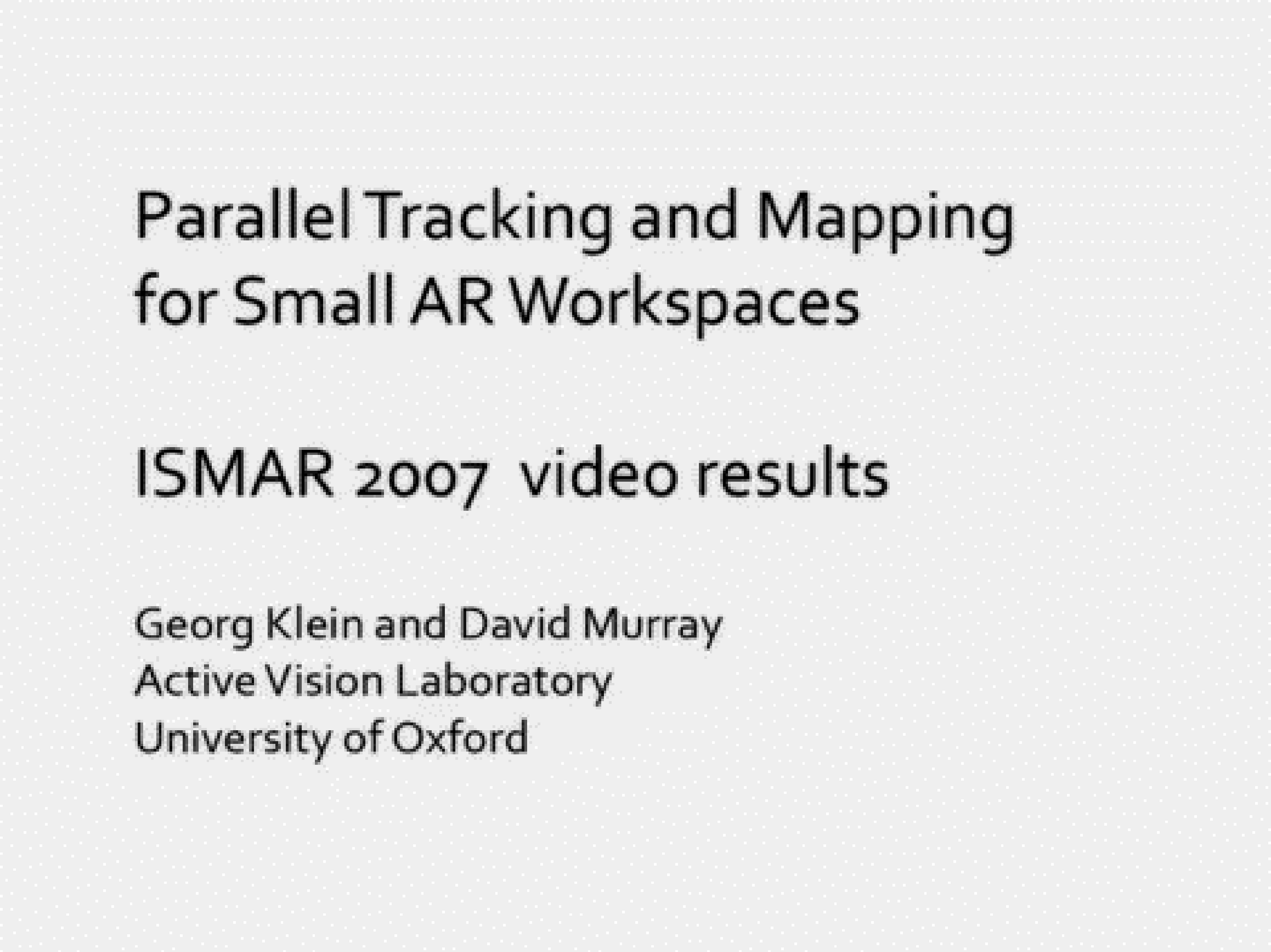
Frame-by-frame SLAM in Practice

- Updating entire map every frame is expensive
- Mandates “sparse map of high-quality features”



Slide from Klein & Murray 2007

PTAM



Parallel Tracking and Mapping
for Small AR Workspaces

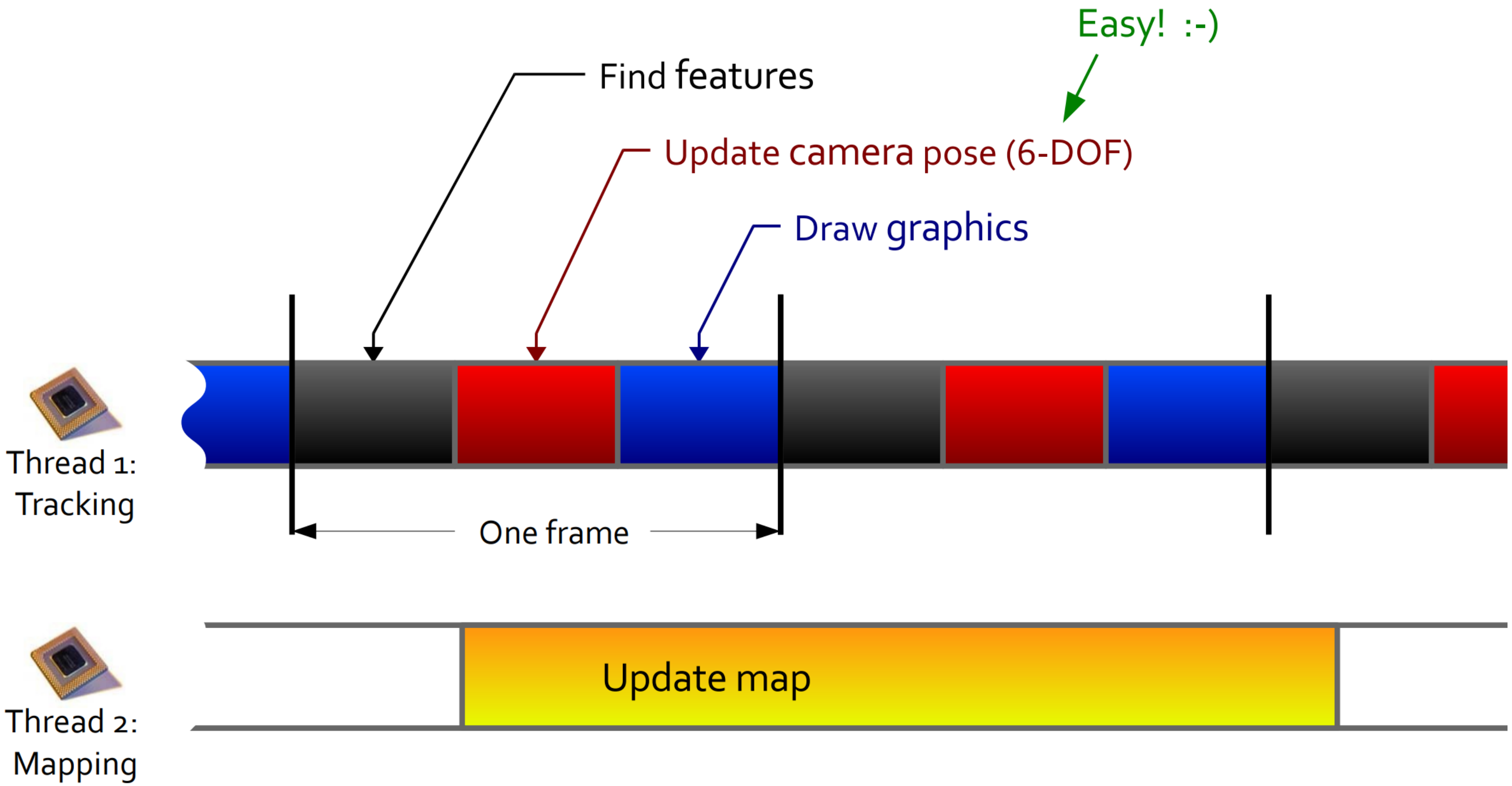
ISMAR 2007 video results

Georg Klein and David Murray
Active Vision Laboratory
University of Oxford

Klein, Georg, and David Murray. "Parallel tracking and mapping for small AR workspaces." *Mixed and Augmented Reality, 2007. ISMAR 2007. 6th IEEE and ACM International Symposium on*. IEEE, 2007.

Parallel Tracking and Mapping

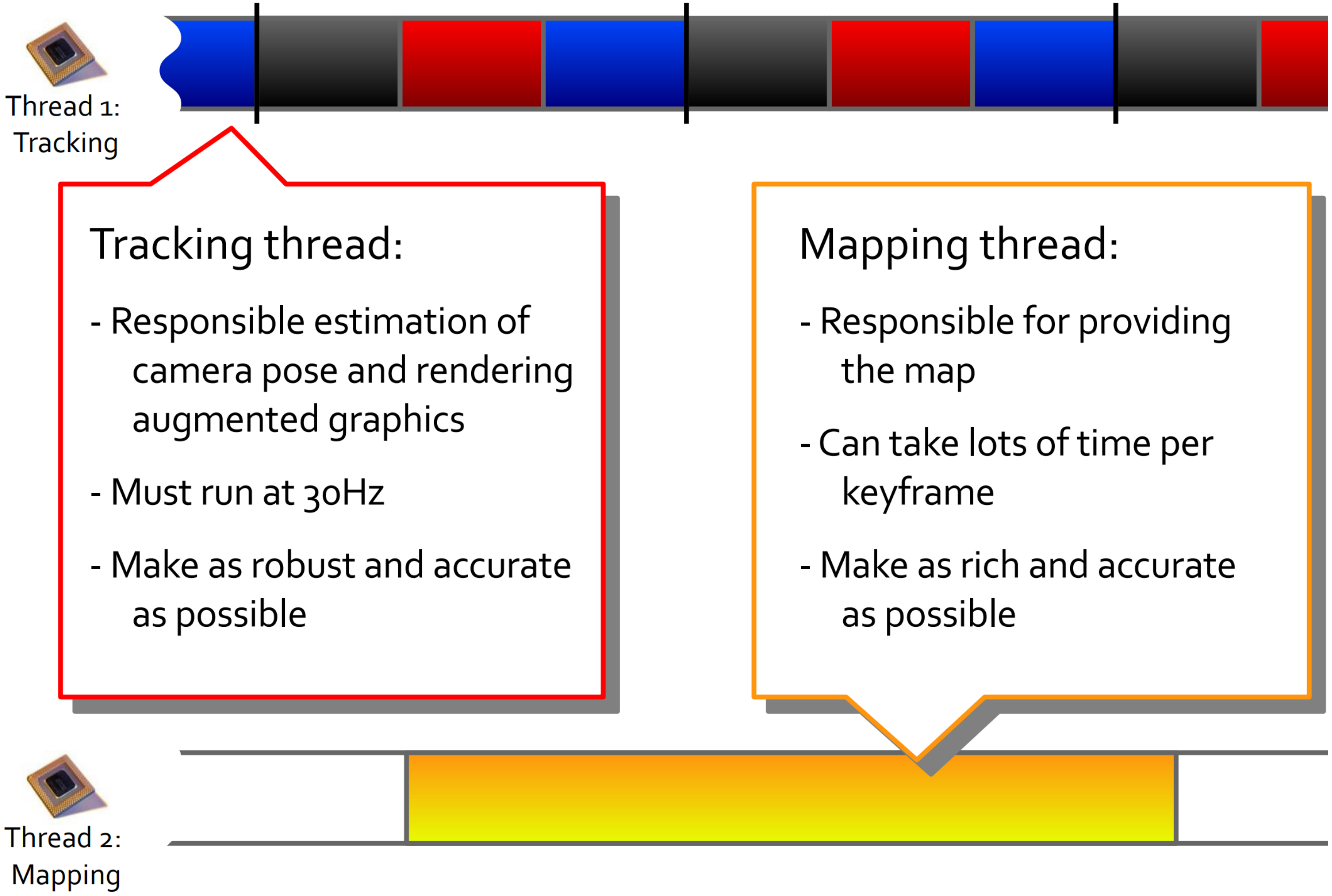
- Use dense map (of low-quality features)
- Don't update the map every frame: **Keyframes**
- Split the tracking and mapping into two threads



Klein, Georg, and David Murray. "Parallel tracking and mapping for small AR workspaces." *Mixed and Augmented Reality, 2007. ISMAR 2007. 6th IEEE and ACM International Symposium on*. IEEE, 2007.

Slide from Klein & Murray 2007

Multi-threads is Common in Modern vSLAM



Tracking thread:

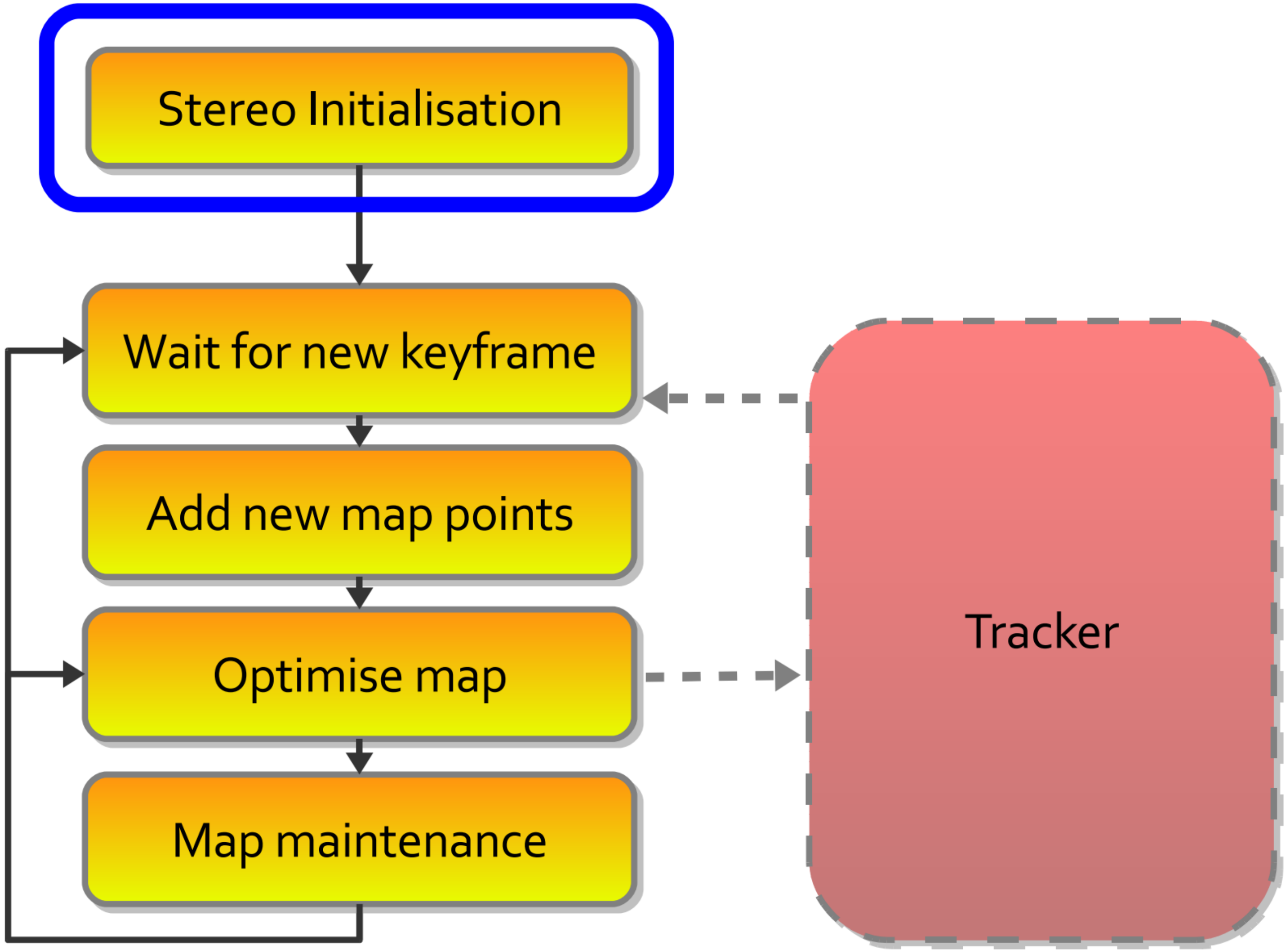
- Responsible estimation of camera pose and rendering augmented graphics
- Must run at 30Hz
- Make as robust and accurate as possible

Mapping thread:

- Responsible for providing the map
- Can take lots of time per keyframe
- Make as rich and accurate as possible

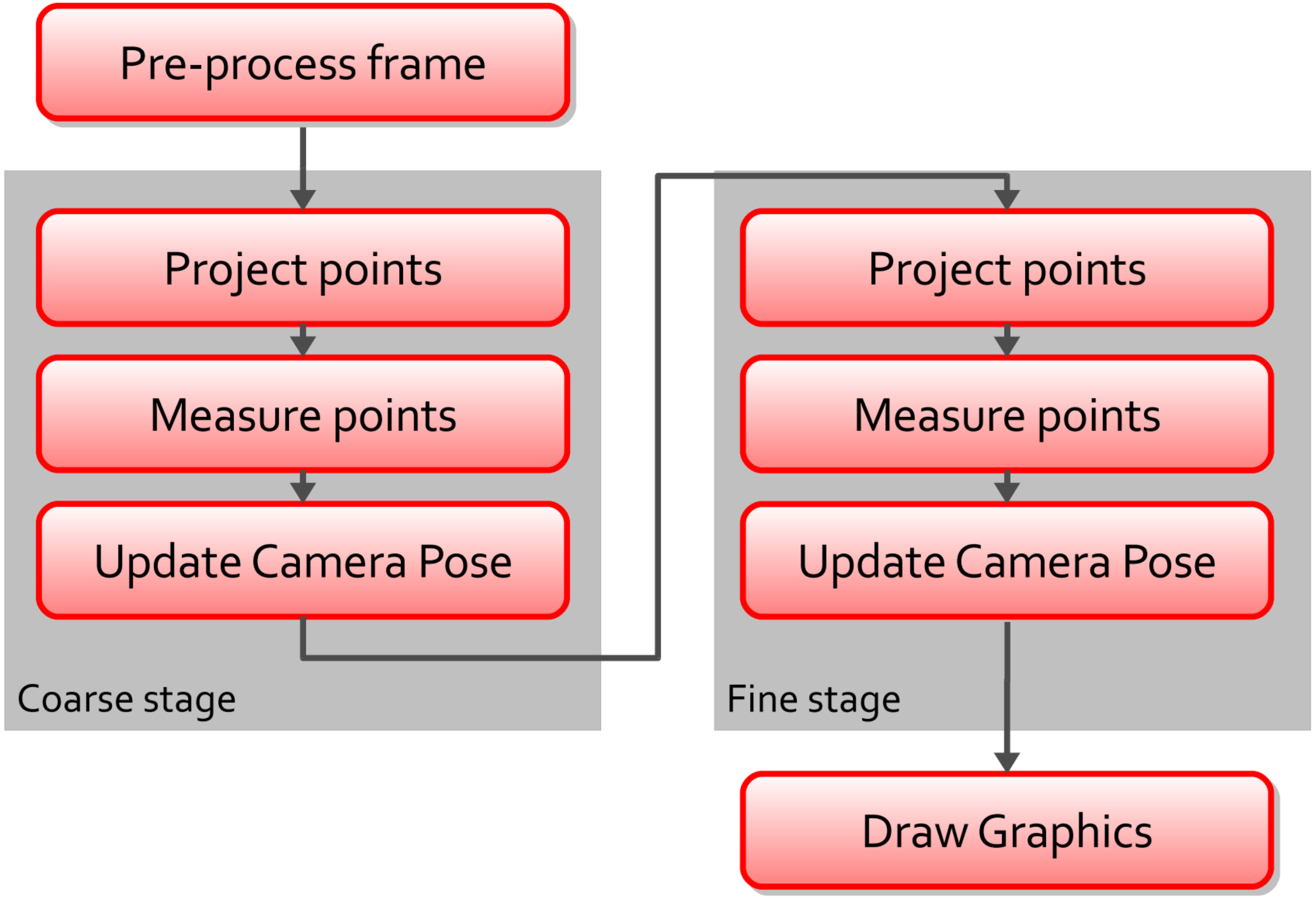
Slide from Klein & Murray 2007

Mapping Thread



Slide from Klein & Murray 2007

Tracking Thread



Slide from Klein & Murray 2007

Next Week

- Quiz on March 02
  - 80min
  - Online (Google Form)
  - Close-book, but 1-page study sheet allowed
  - True-or-False or Multiple-Choice questions
  - Covering everything we’ve discussed so far.

References for the Week after Quiz1

- Corke 2017:
  - Section 14.5
- Forsyth & Ponce 2011
  - Section 14.3
- Szeliski 2022:
  - Section 13.2.1

- Umeyama, S. (1991). Least-squares estimation of transformation parameters between two point patterns. IEEE Transactions on Pattern Analysis & Machine Intelligence, (4), 376-380.
- Low, K.L., 2004. Linear least-squares optimization for point-to-plane icp surface registration. *University of North Carolina Chapel Hill*, *4*(10).
- Park, J., Zhou, Q. Y., & Koltun, V., (2017). Colored point cloud registration revisited. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (pp. 143-152).