Lecture 10.1
Introduction to Visual SLAM

Trym Vegard Haavardsholm

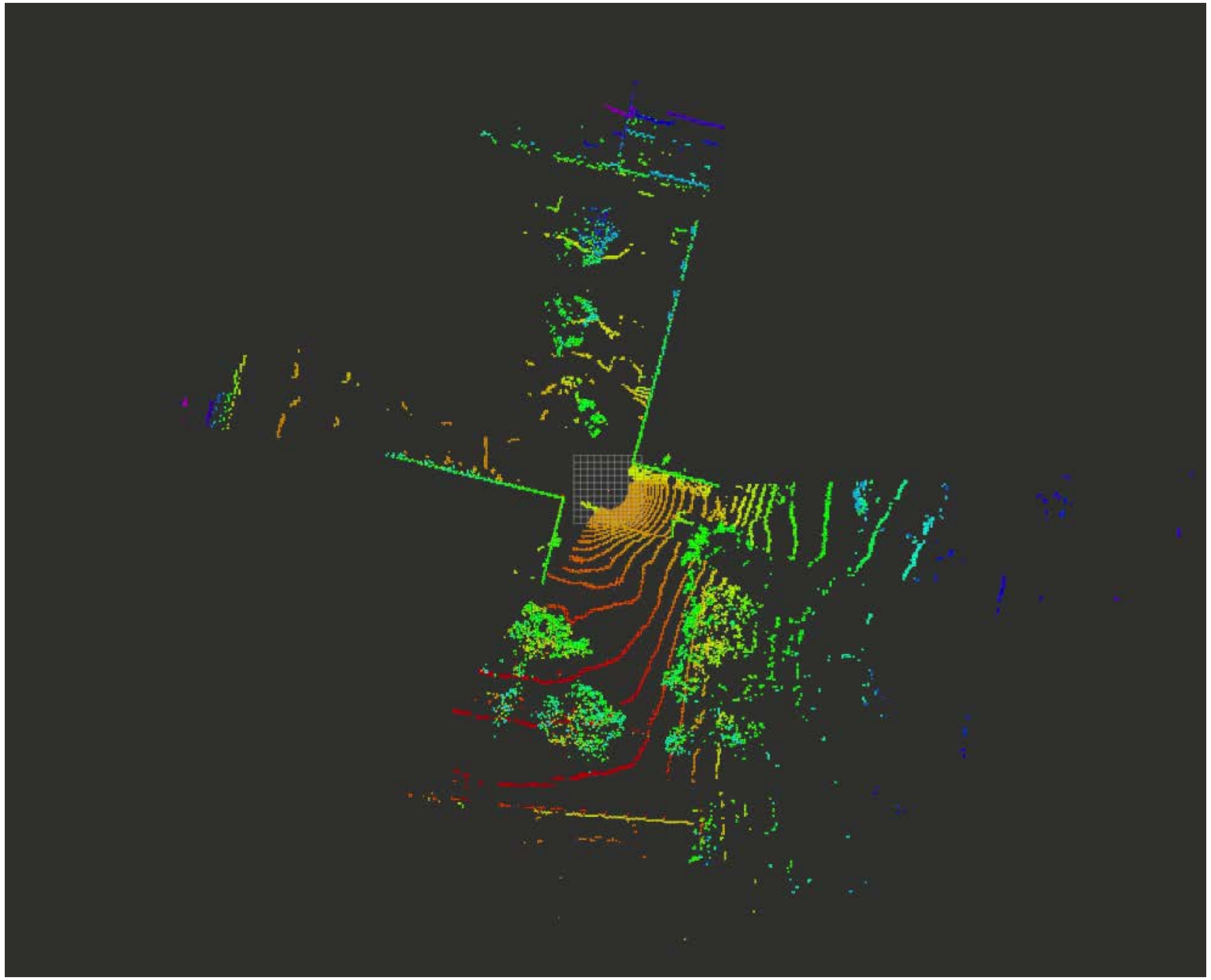
What is SLAM?

What is SLAM?

Simultaneous localization and mapping

What is SLAM?

Simultaneous localization and mapping



Simultaneous

- estimation of the state of a robot using on-board sensors
- construction of a map of the environment that the sensors are perceiving

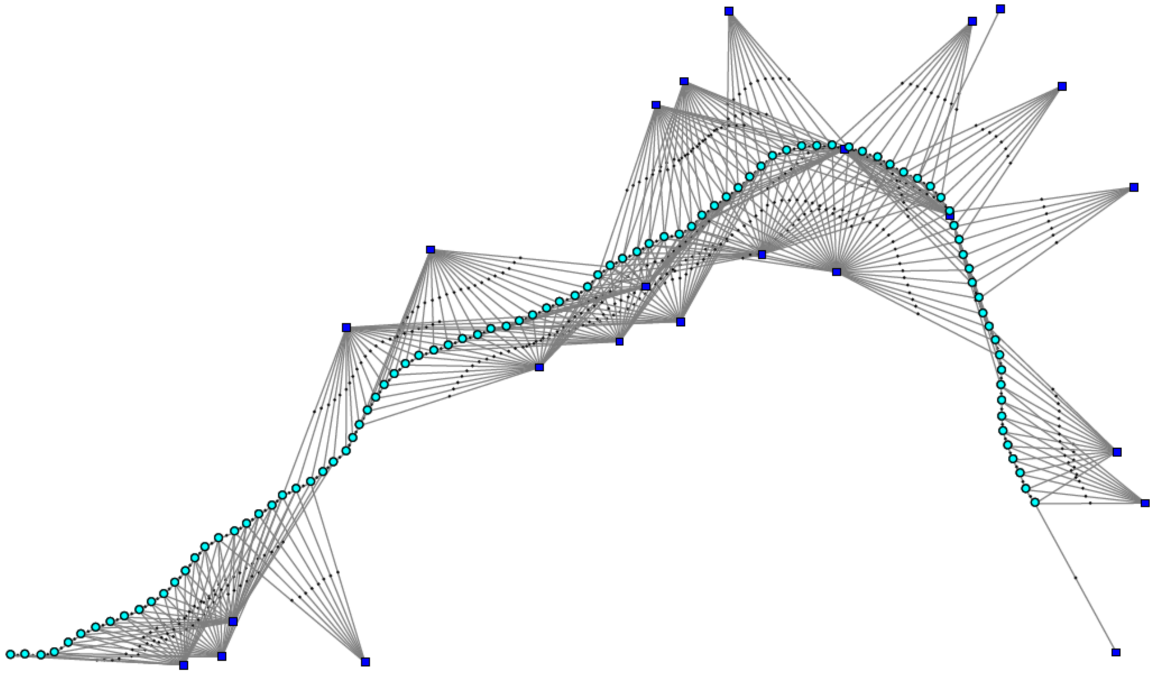
What is SLAM?

Simultaneous localization and mapping

Simultaneous

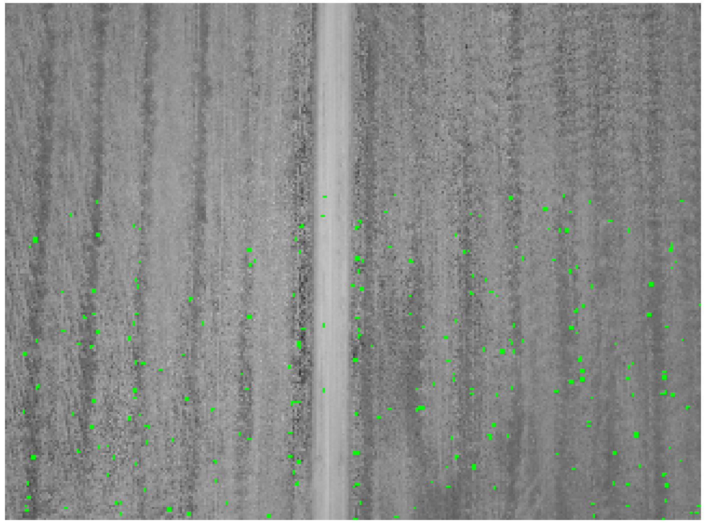
- **mapping**:
  Continuously expanding
  and optimizing a consistent map
  while exploring the environment

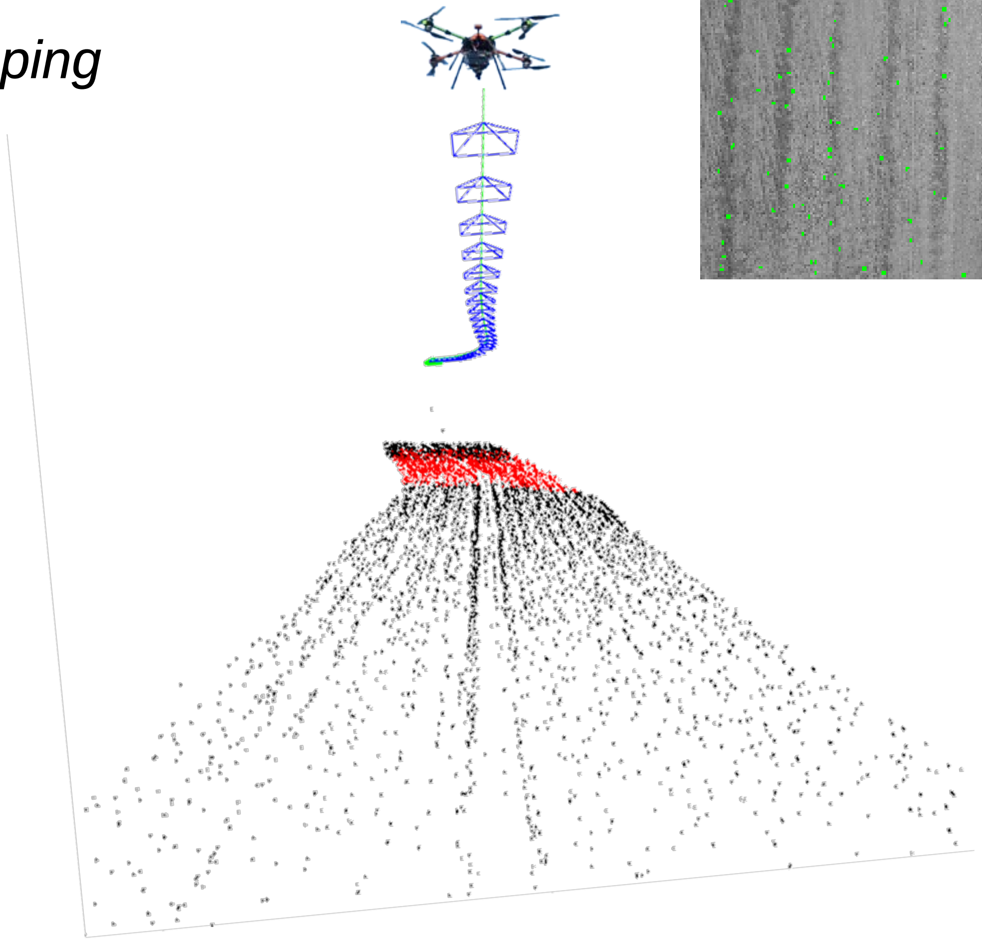
- **localization**:
  Localization within the map



What is Visual SLAM?

Visual simultaneous localization and mapping



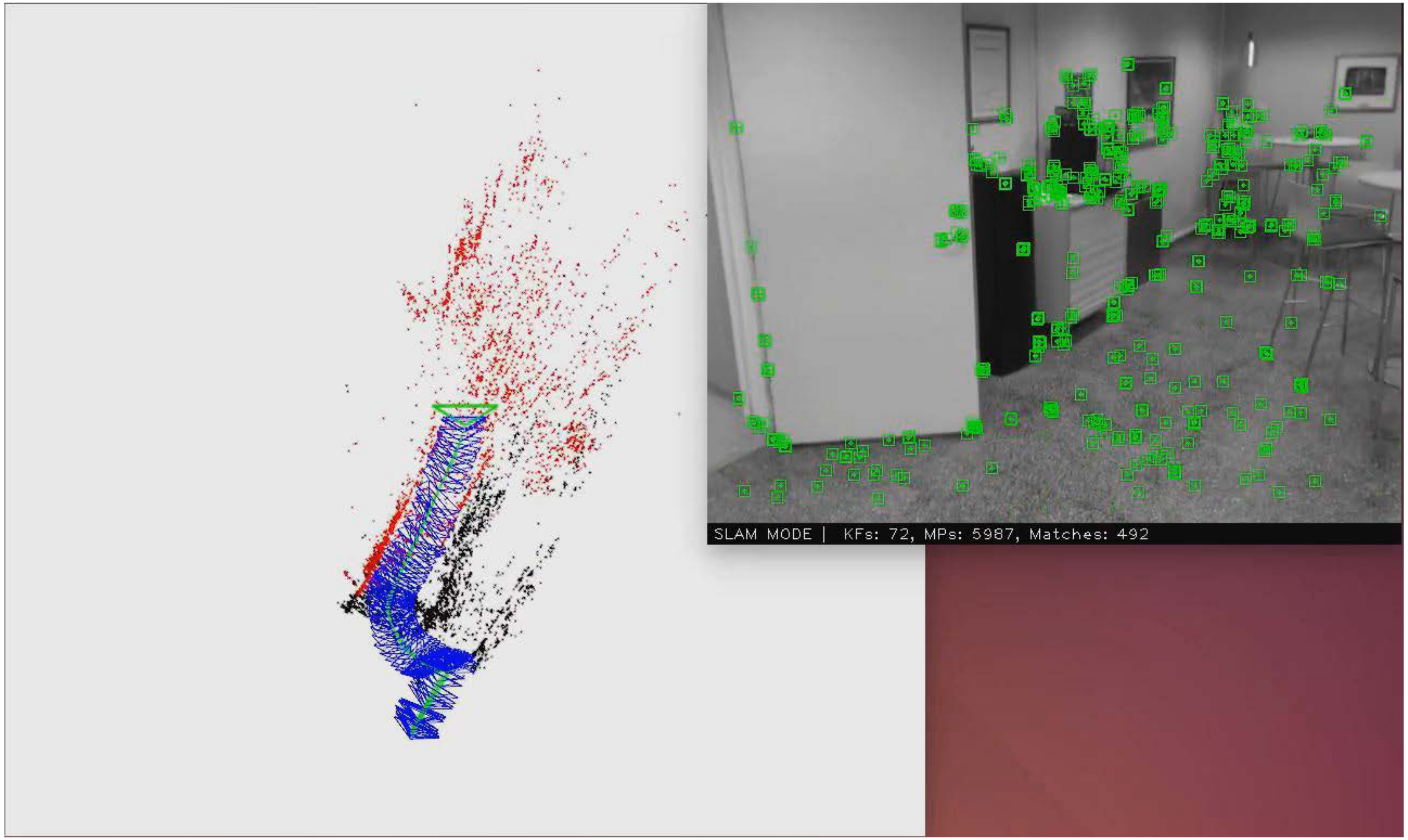


Simultaneous

- **mapping**:
  Continuously expanding
  and optimizing a consistent map
  while exploring the environment

- **localization (tracking)**:
  Localization within the map
  (tracking the map in image frames)

Visual SLAM example



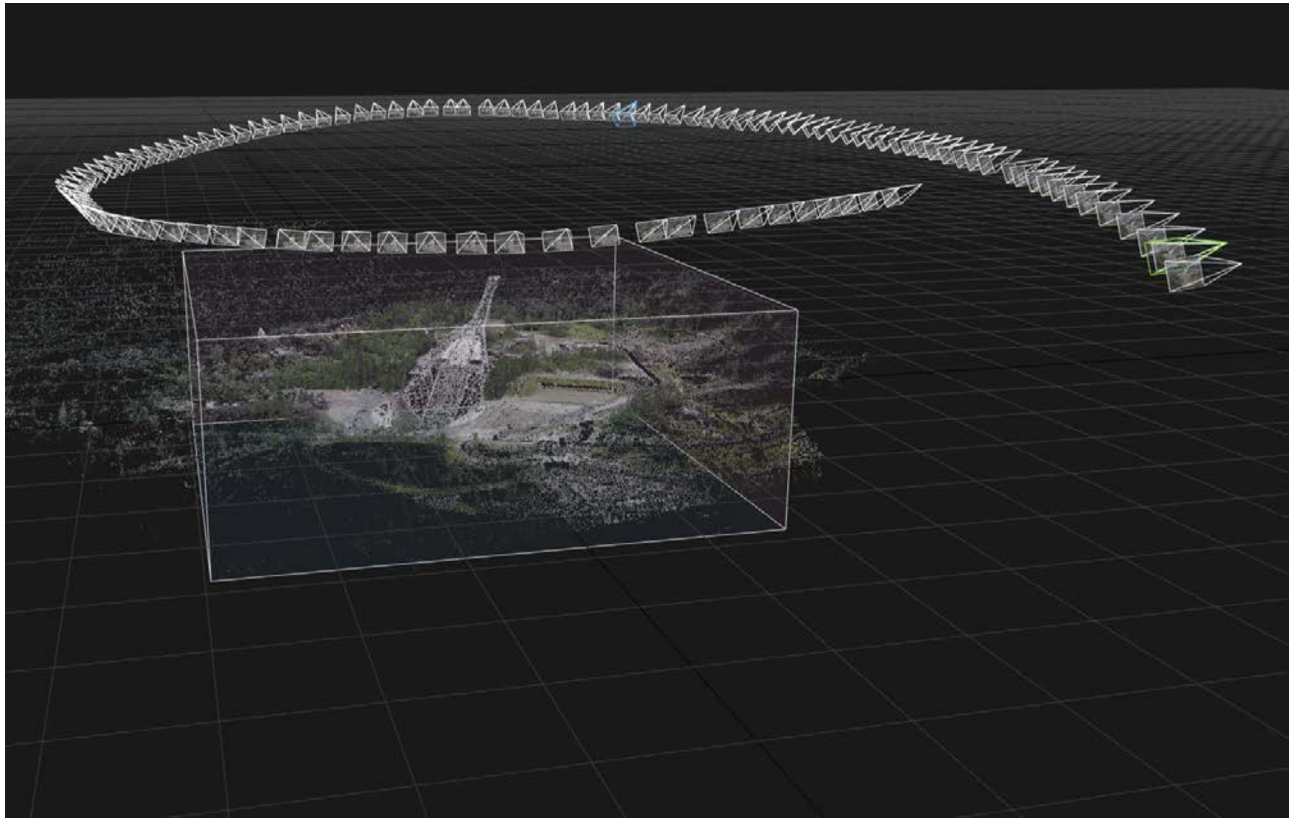
What is the map?

What is the map?

A model of the environment that lets us

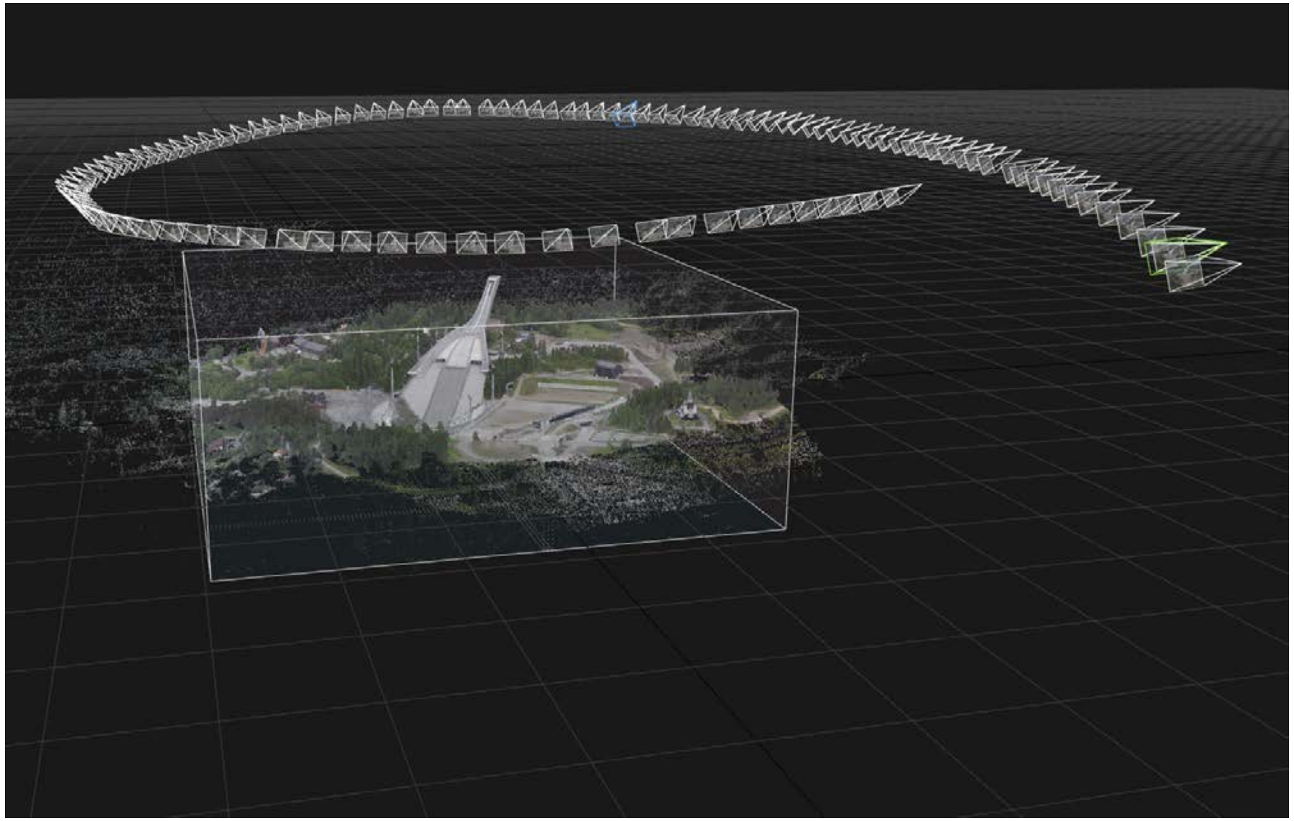
- limit the localization error
  by recognizing previously visited areas

- (support other tasks,
  such as obstacle avoidance
  and path planning)



What is the map?

A model of the environment that lets us



- limit the localization error
  by recognizing previously visited areas

- (support other tasks,
  such as obstacle avoidance
  and path planning)

  Maybe best left as auxiliary processing?

Examples of map representations

Feature-based metric maps

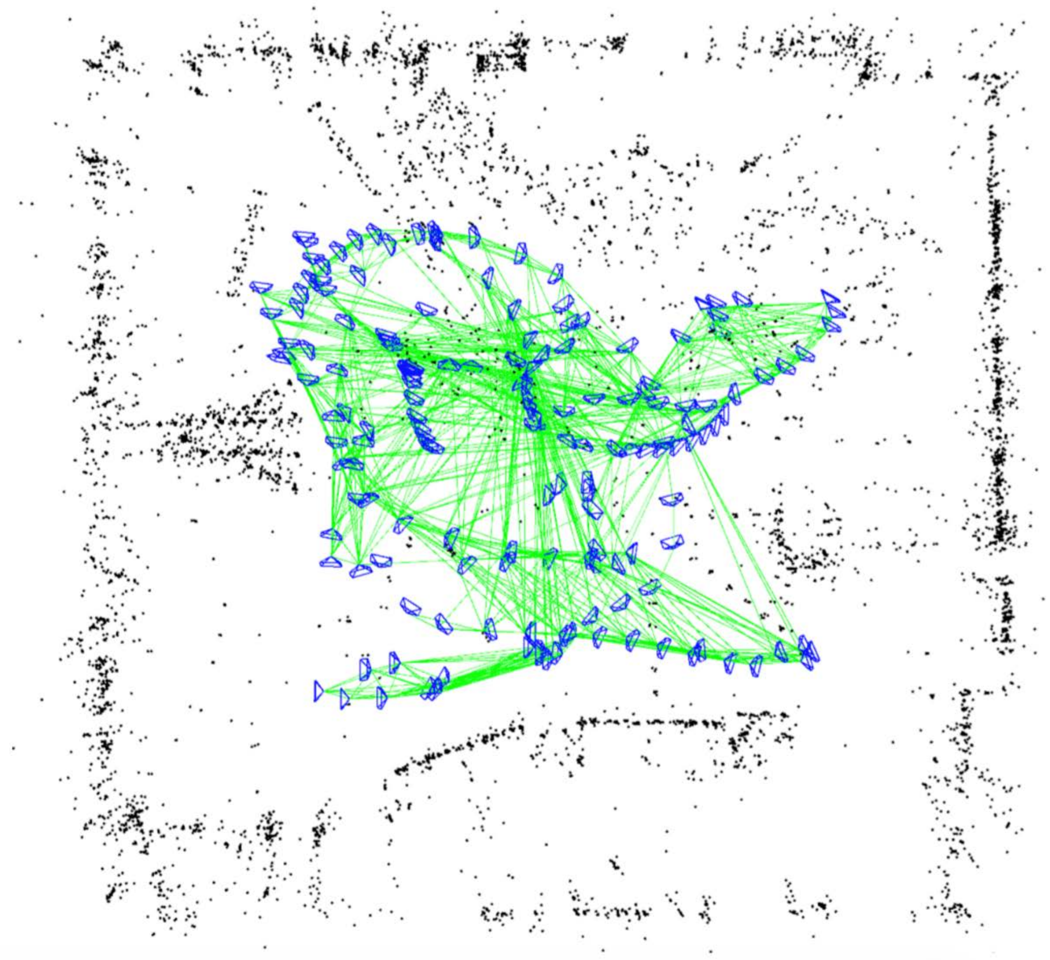


Image: Cadena, C., et al. (2016). Past, Present, and Future of Simultaneous Localization and Mapping: Toward the Robust-Perception Age. *IEEE Transactions on Robotics*, *32*(6), 1309–1332

Mur-Artal, R., Montiel, J. M. M., & Tardos, J. D. (2015). ORB-SLAM: A Versatile and Accurate Monocular SLAM System. IEEE Transactions on Robotics, 31(5), 1147–1163. https://doi.org/10.1109/TRO.2015.2463671

Examples of map representations

Dense metric maps

[DTAM:
Dense Tracking and Mapping in Real-Time]

Image: Cadena, C., et al. (2016). Past, Present, and Future of Simultaneous Localization and Mapping: Toward the Robust-Perception Age. *IEEE Transactions on Robotics*, *32*(6), 1309–1332

Newcombe, R. A., Lovegrove, S. J., & Davison, A. J. (2011). DTAM: Dense tracking and mapping in real-time. In 2011 International Conference on Computer Vision (pp. 2320–2327). IEEE

Examples of map representations

Topological maps

[FABMAP]

Image: YouTube: ORI - Oxford Robotics Institute

Cummins, M., & Newman, P. (2008). FAB-MAP: Probabilistic Localization and Mapping in the Space of Appearance. The International Journal of Robotics Research, 27(6), 647–665

Examples of map representations

Topological-metric maps

[Visual Teach & Repeat]()

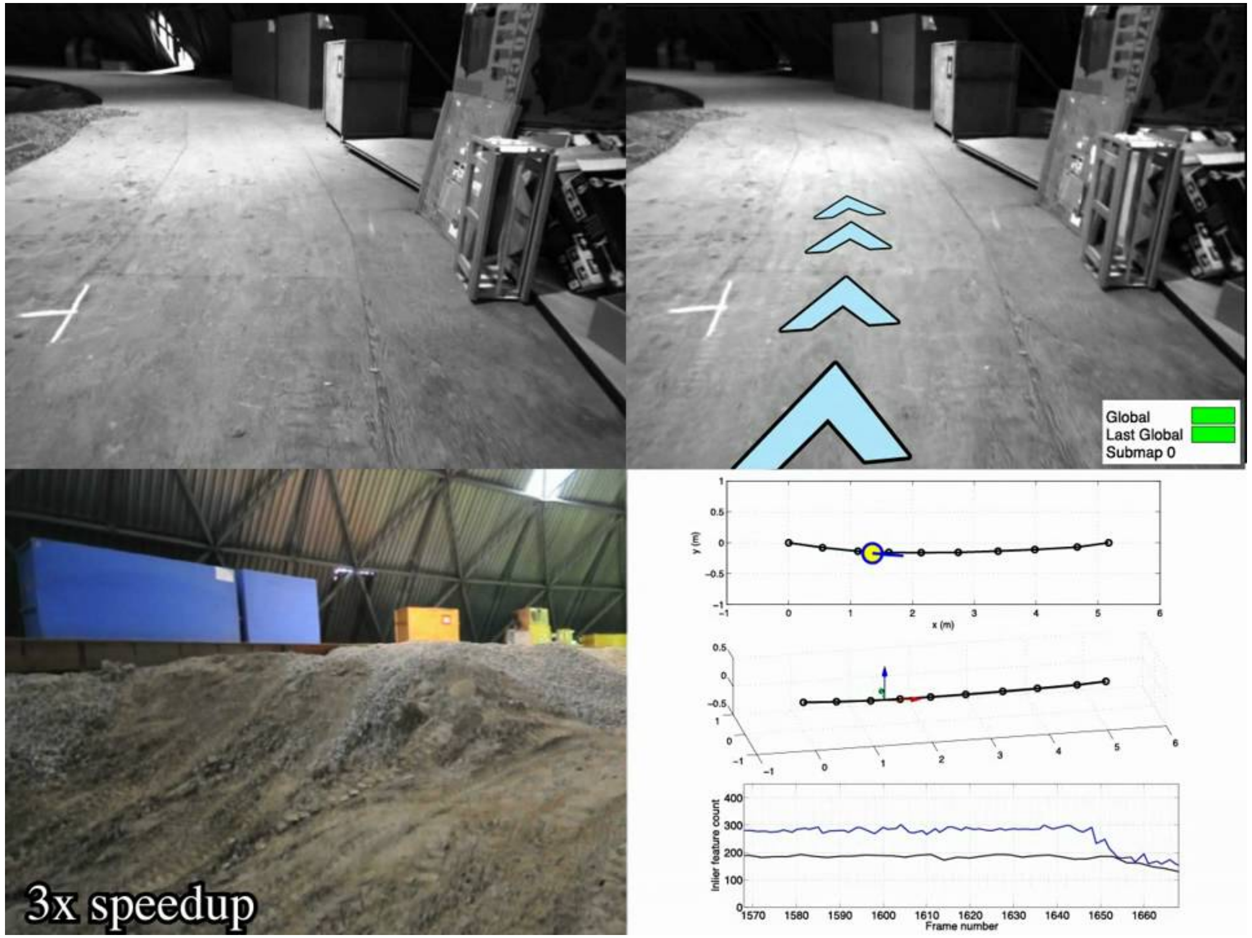
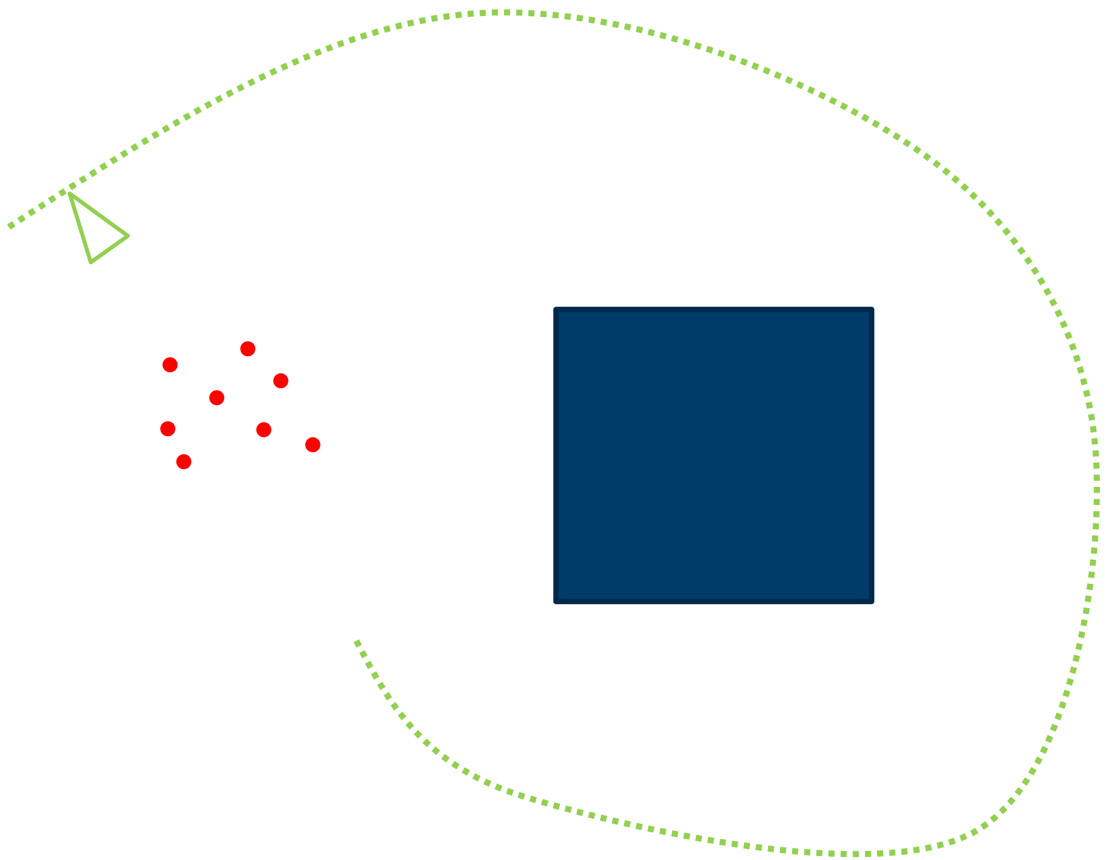


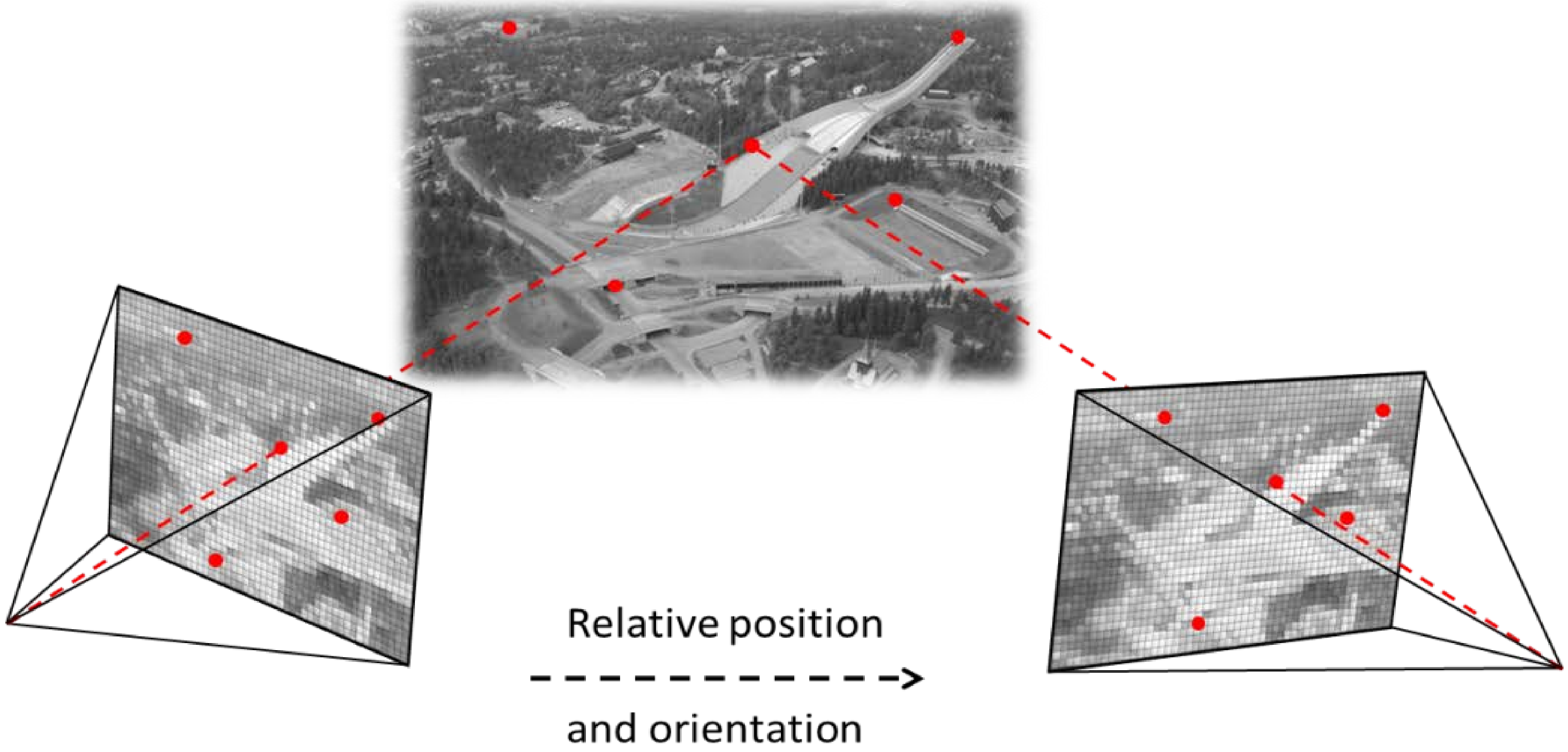
Image: YouTube: utiasASRL

Furgale P T and Barfoot T D. Visual Teach and Repeat for Long-Range Rover Autonomy.
Journal of Field Robotics, special issue on Visual mapping and navigation outdoors, 27(5): 534-560, 2010.

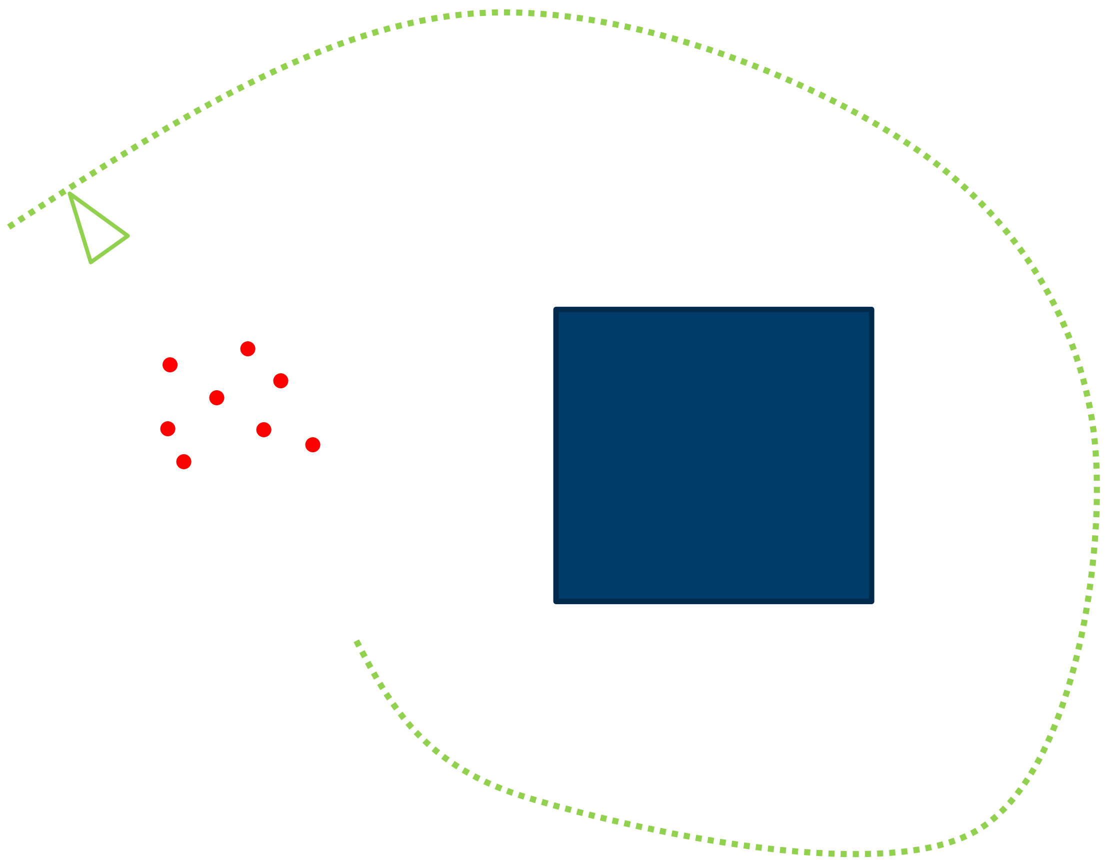
How do we build a map?



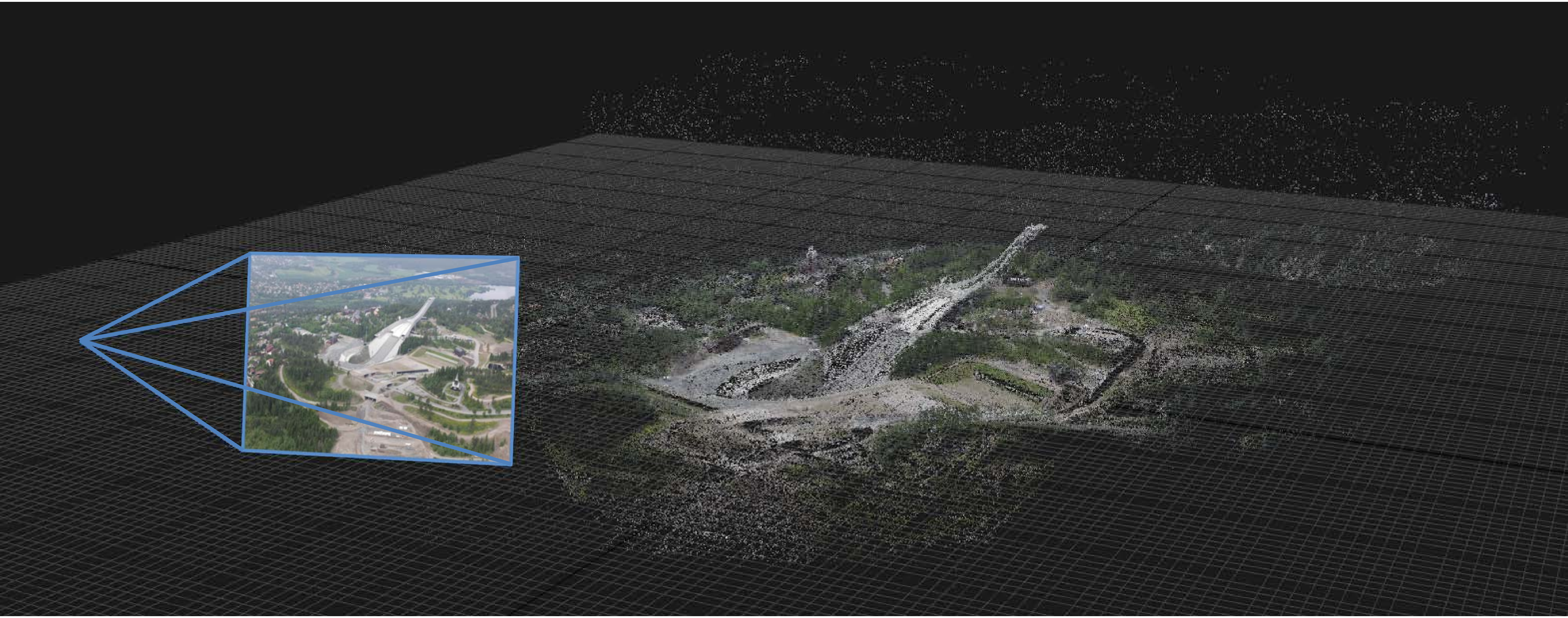
Relative pose and 3D from two views



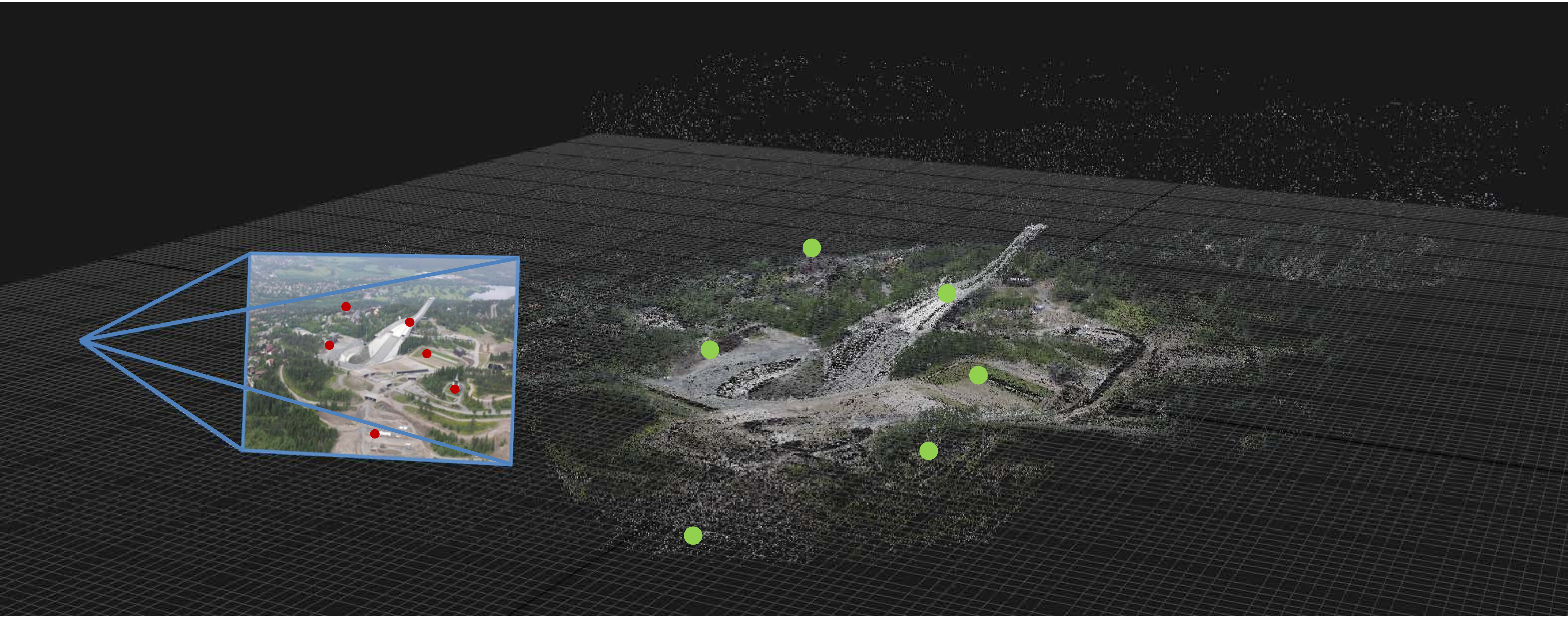
How do we track a map?



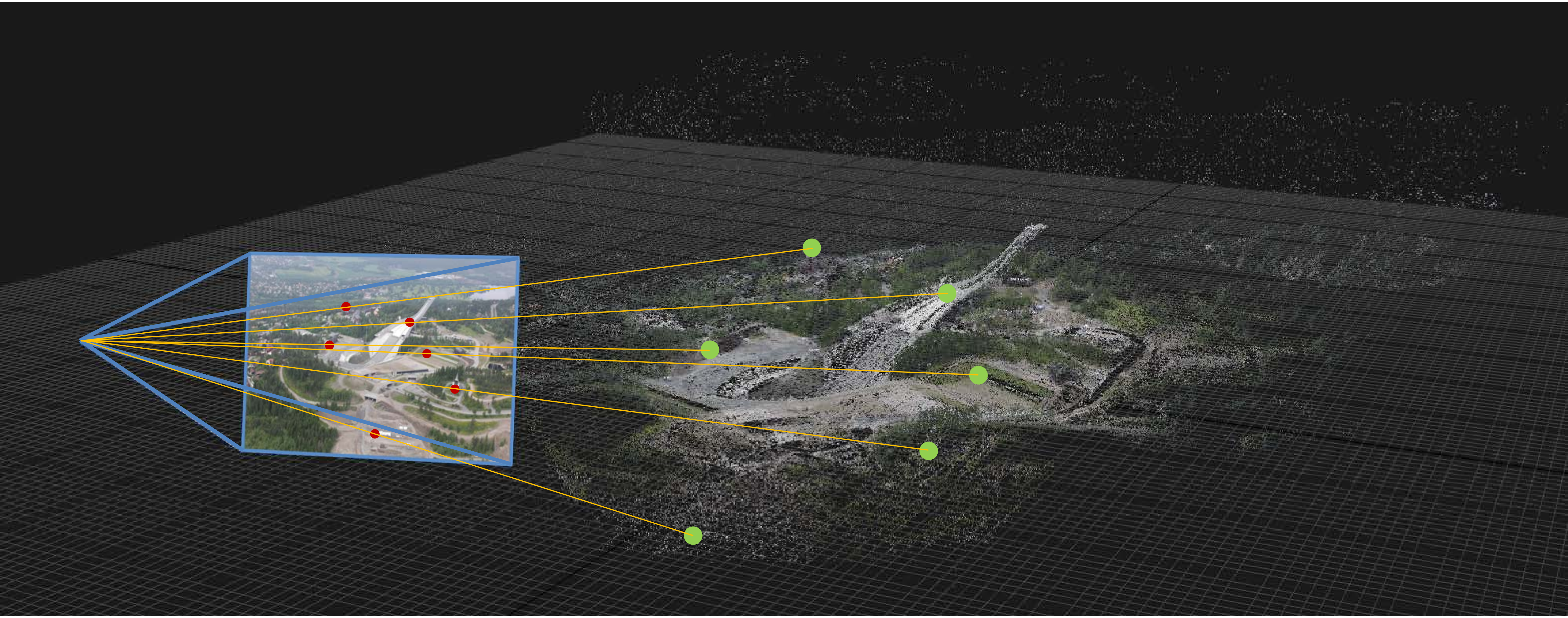
Pose from known 3D map



Pose from point correspondences



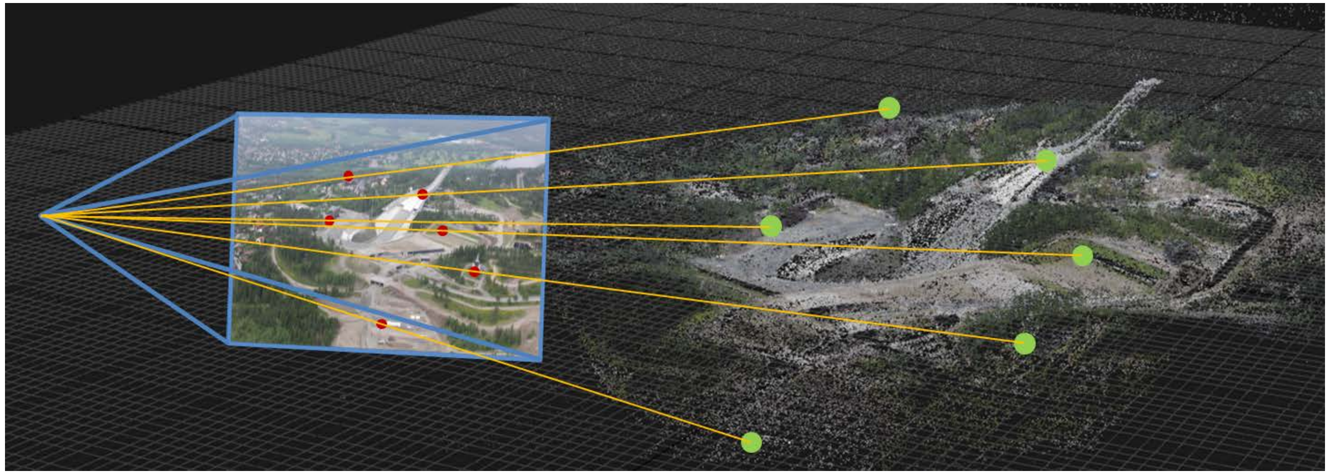
Pose from point correspondences



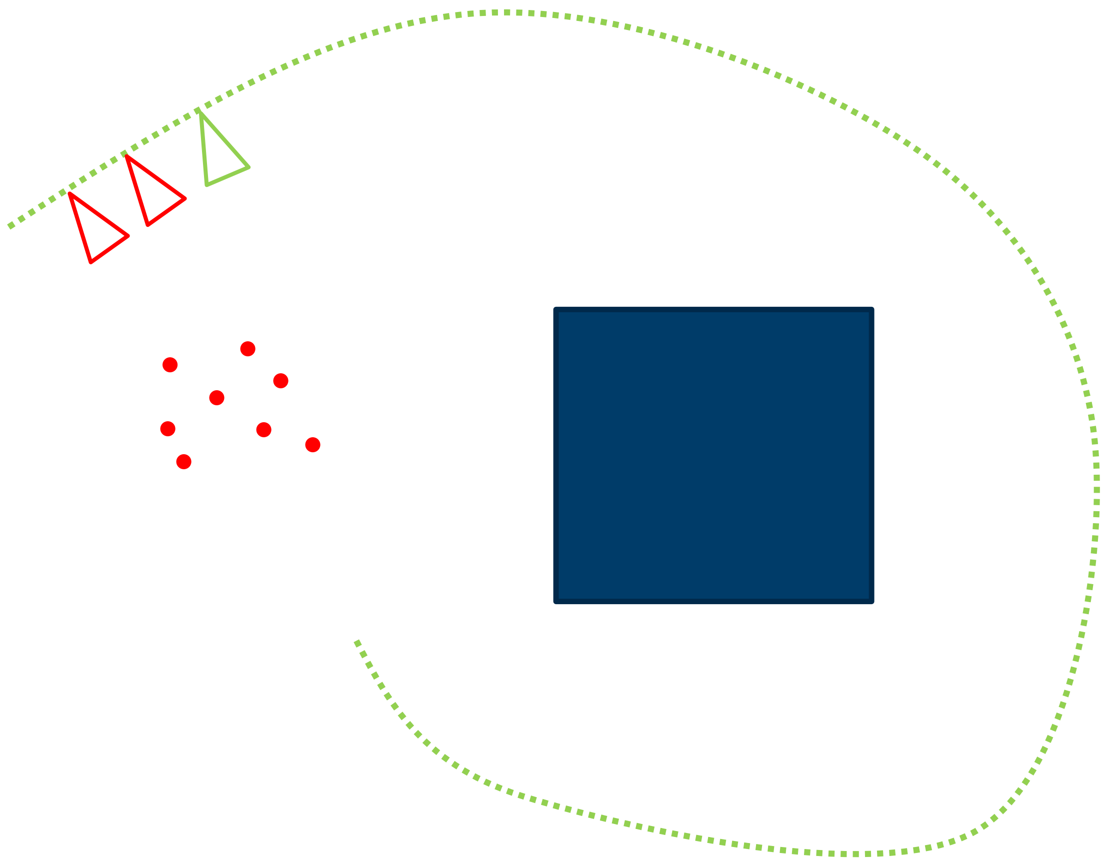
Pose from point correspondences

Minimize *geometric error*

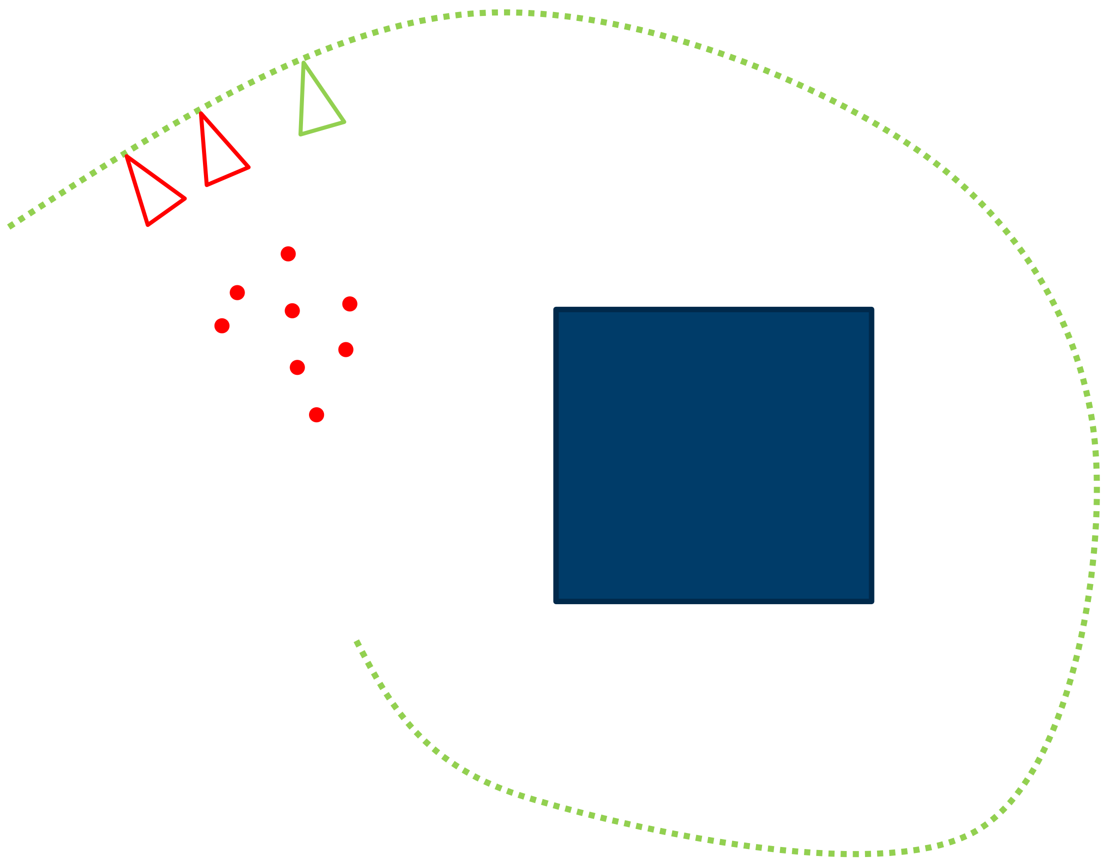
$$\mathbf{T}_{wc}^{*} = \underset{\mathbf{T}_{wc}}{\operatorname{argmin}} \sum_{i} \left\| \pi\left(\mathbf{T}_{wc}^{-1} \cdot \mathbf{x}_{i}^{w}\right) - \mathbf{u}_{i} \right\|^{2}$$



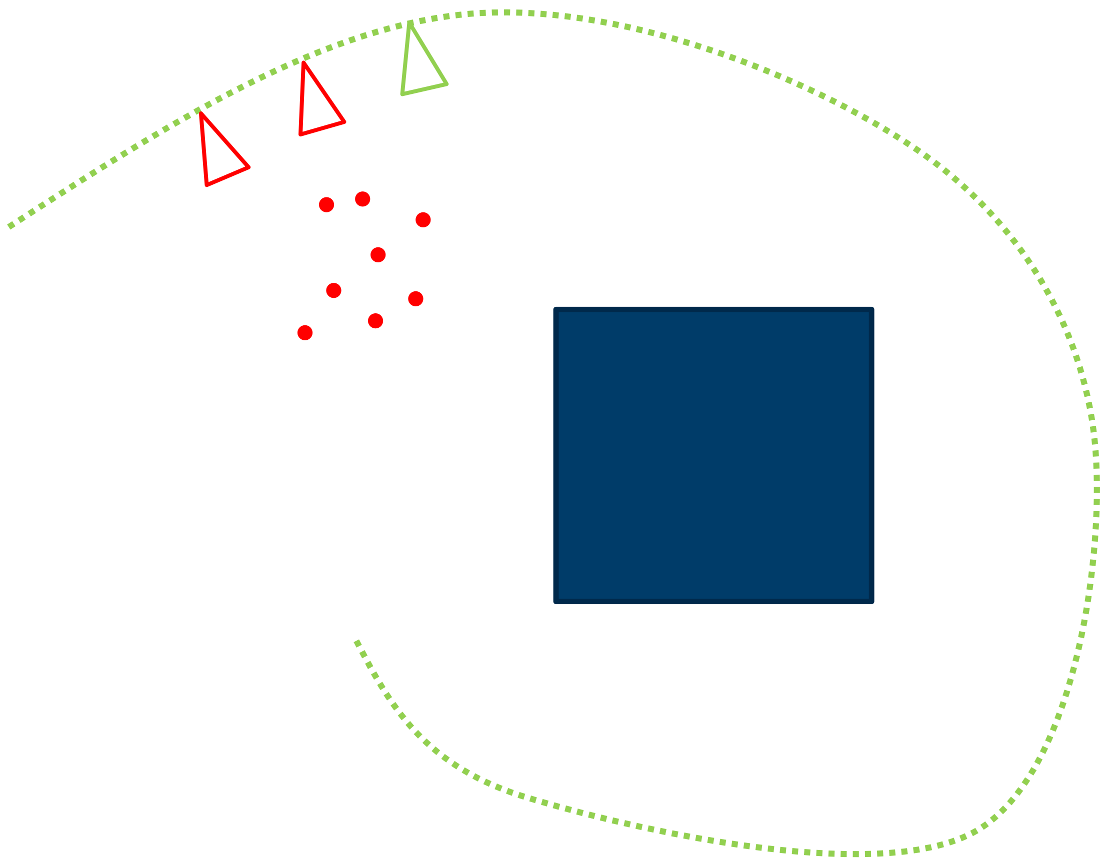
Map initialization and tracking



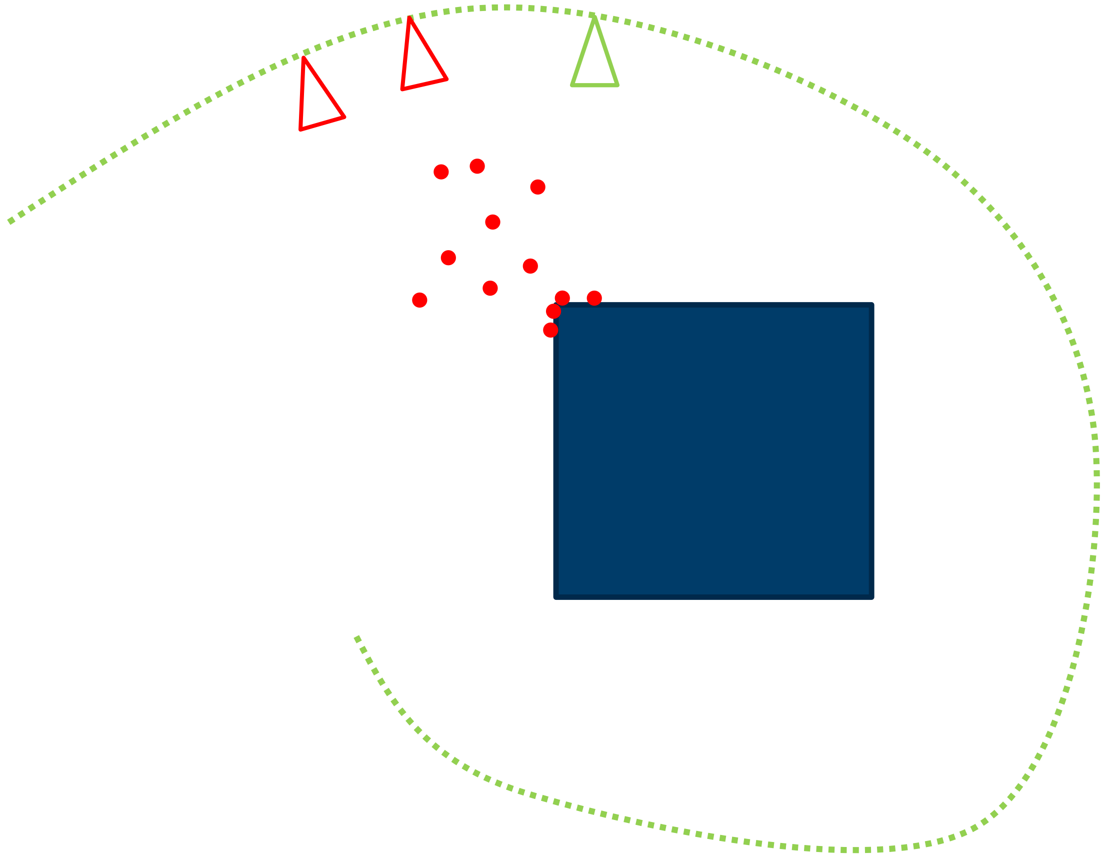
Map reinitialization and tracking



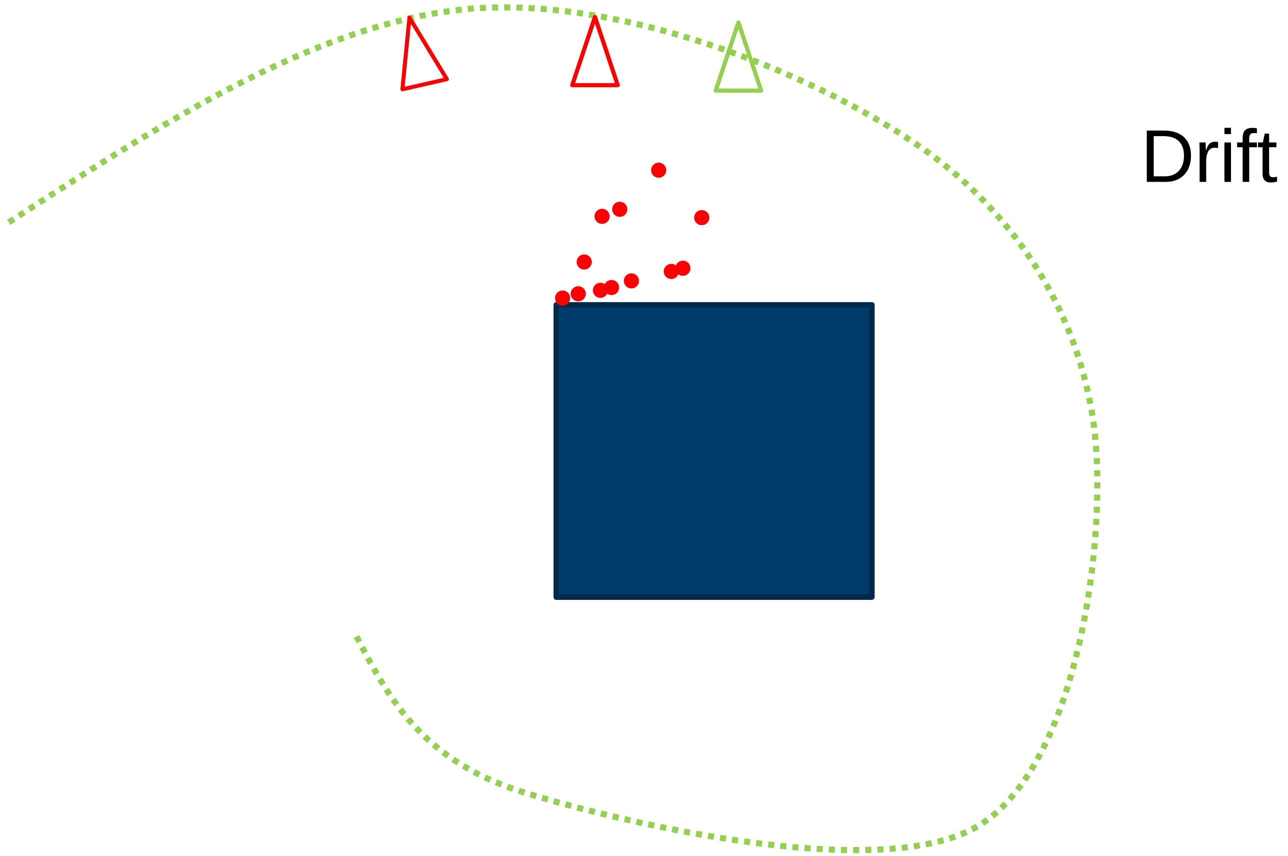
Map reinitialization and tracking



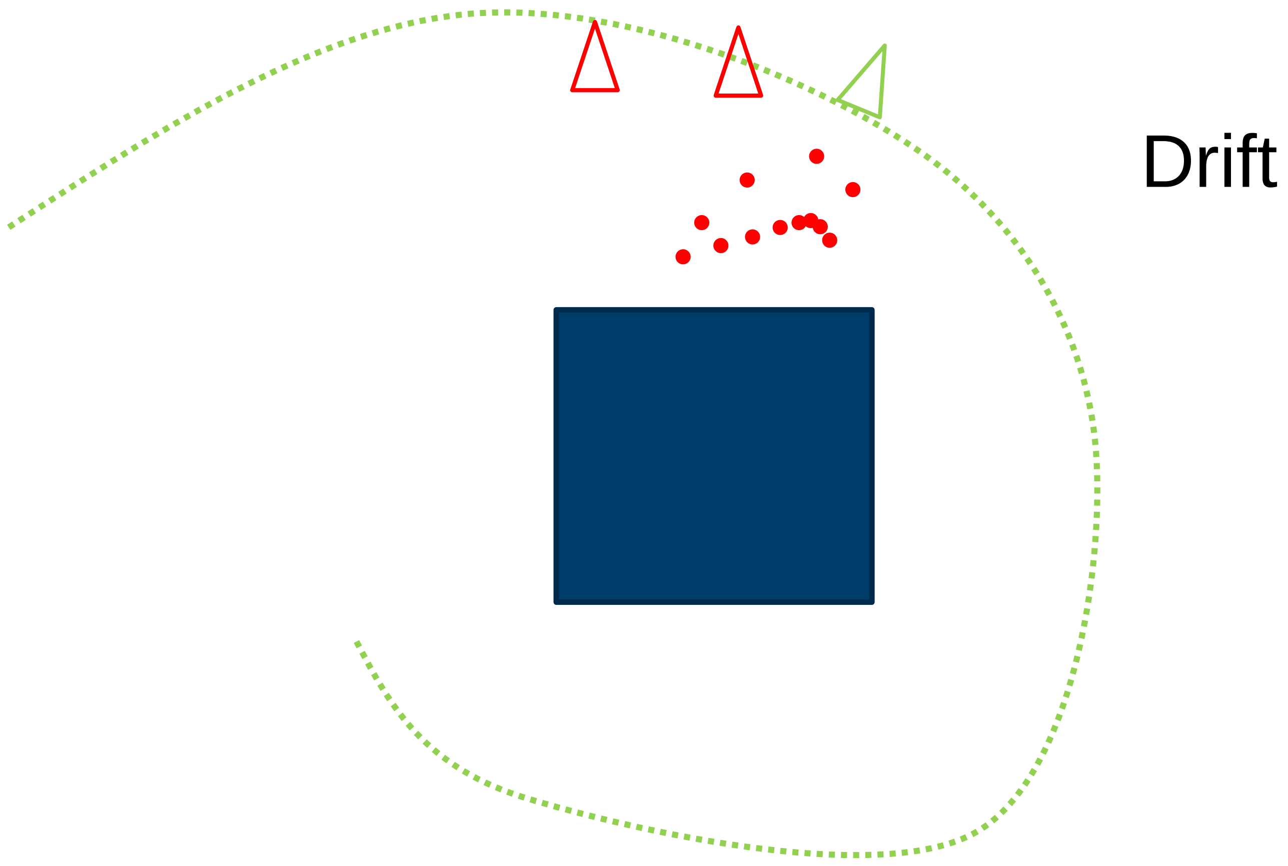
Map reinitialization and tracking



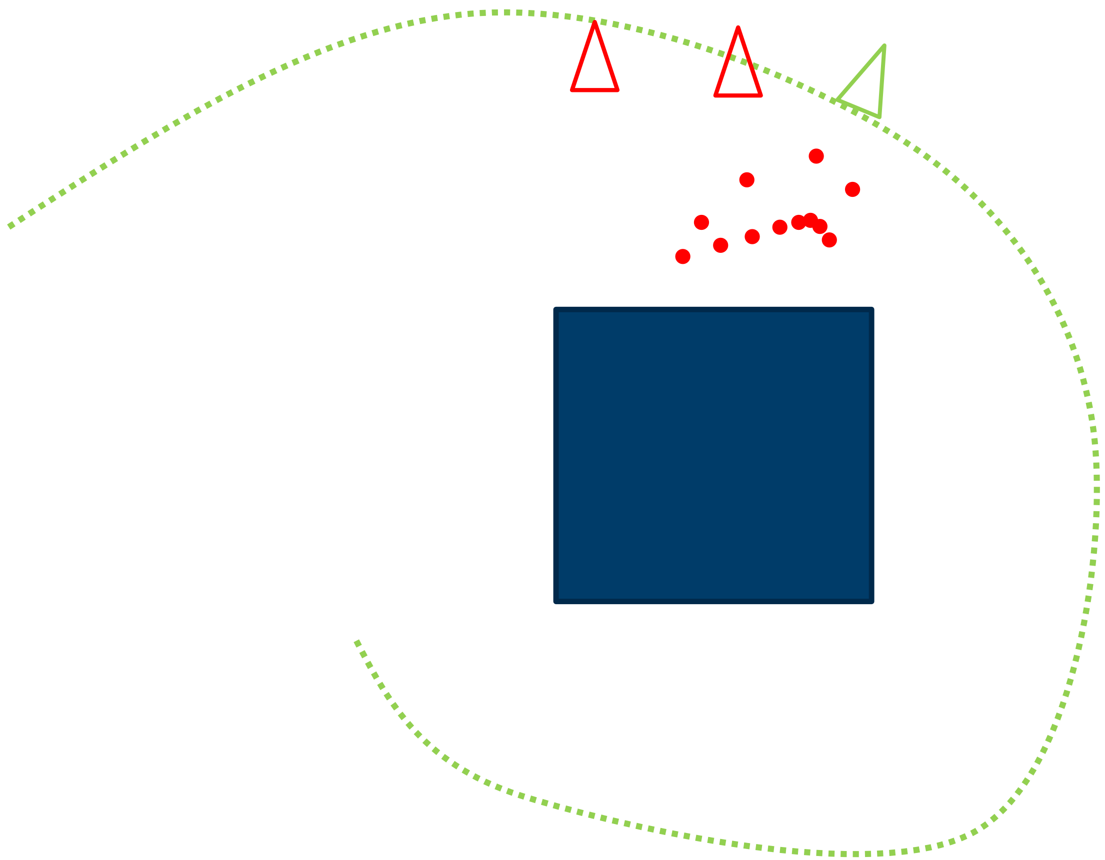
Map reinitialization and tracking



Map reinitialization and tracking



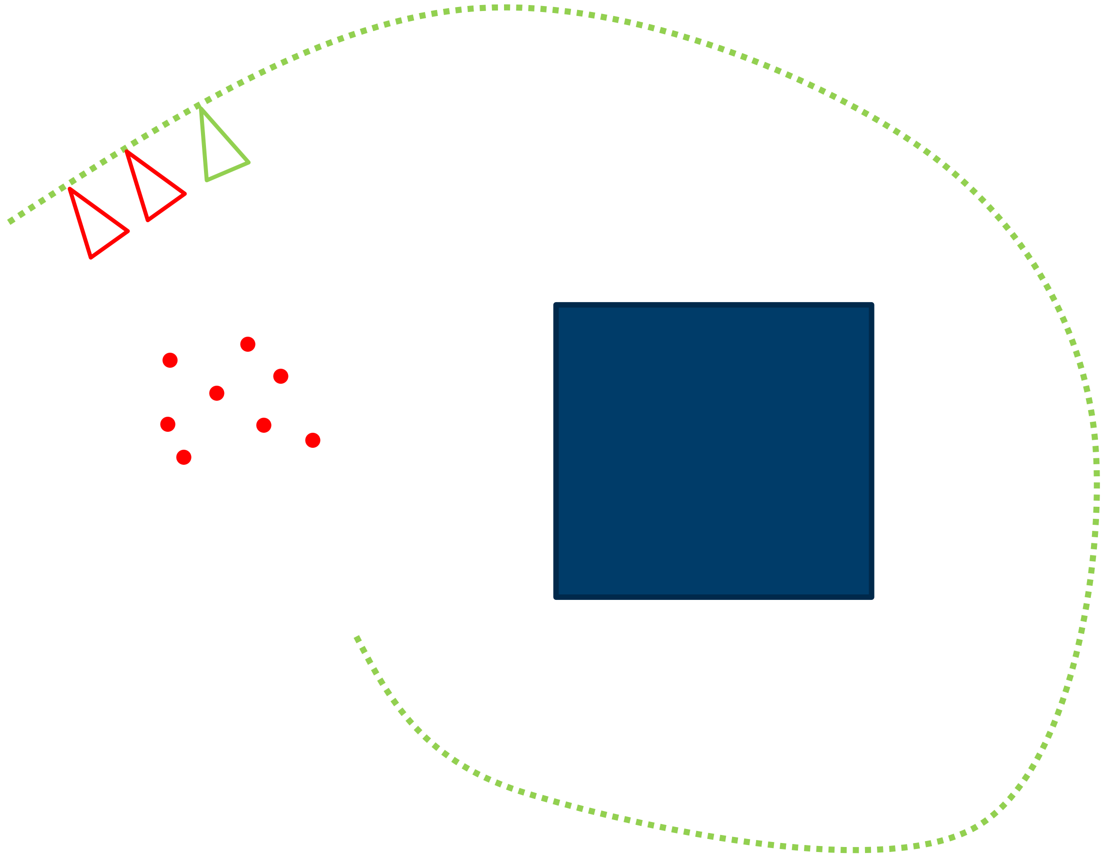
Map reinitialization and tracking



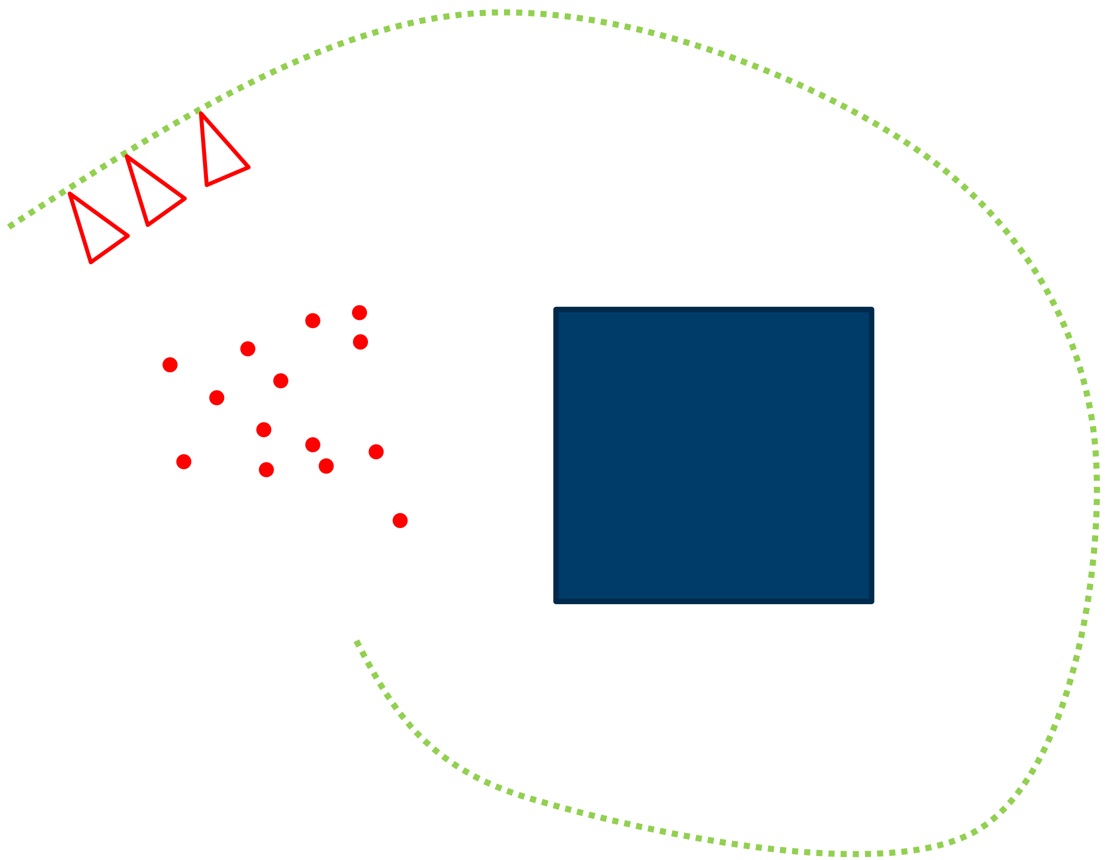
Drift

Very naïve
Visual Odometry (VO)

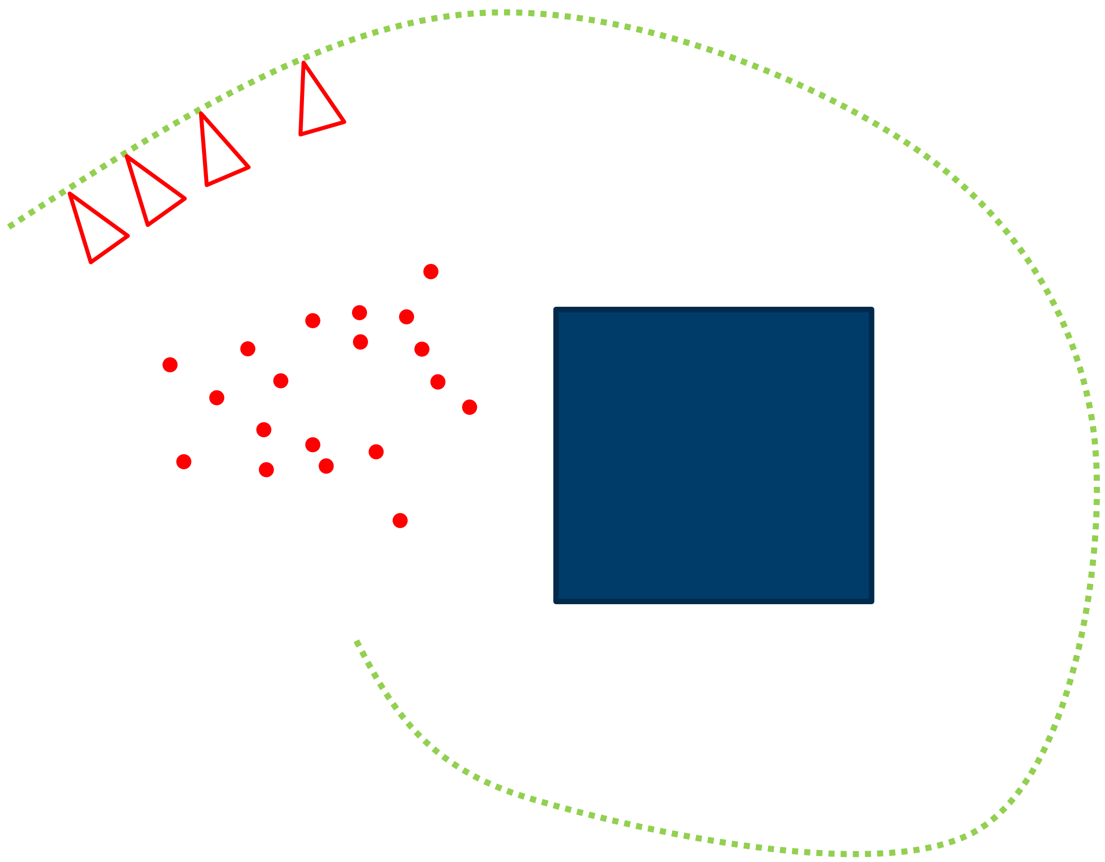
Multi-view mapping



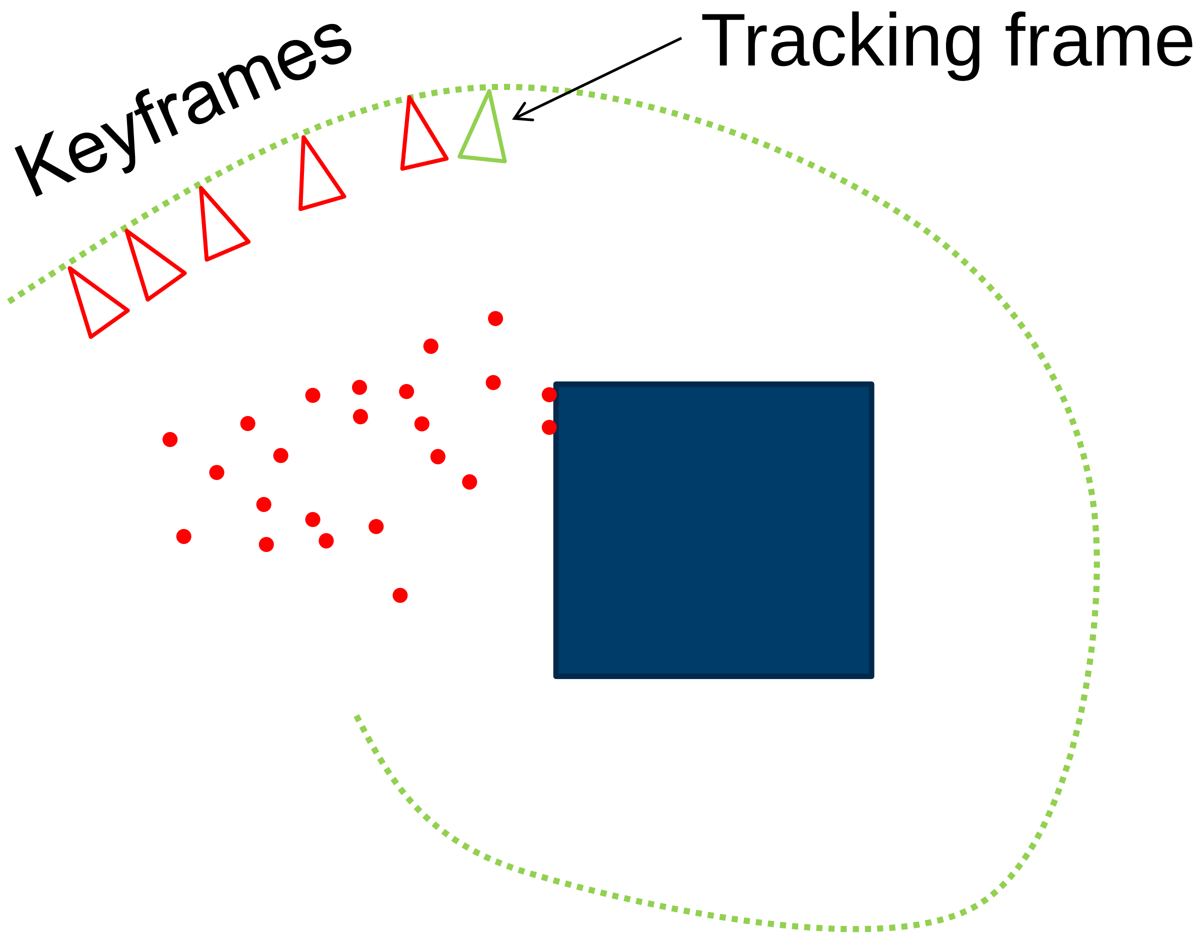
Multi-view mapping



Multi-view mapping



Multi-view mapping

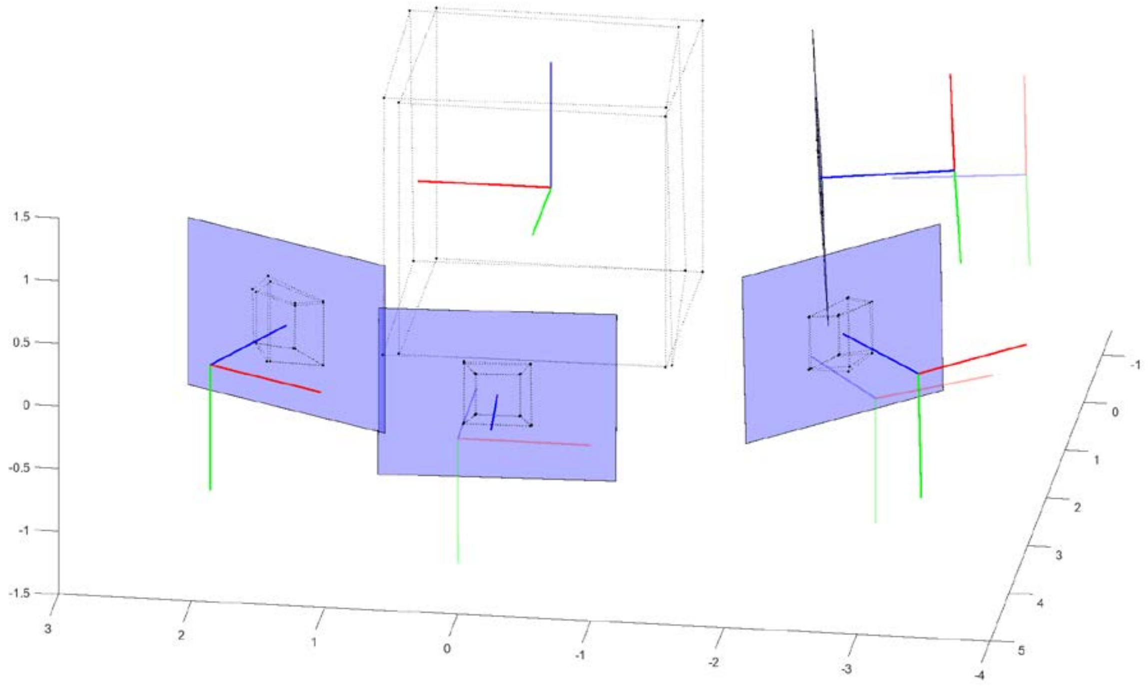


Maximum a Posteriori Inference

Interested in the unknown state variables X,
given the measurements Z.

The most often used estimator for X
is the MAP estimate:

$$X^{\text{MAP}} = \underset{X}{\operatorname{argmax}}\; p(X \mid Z)$$



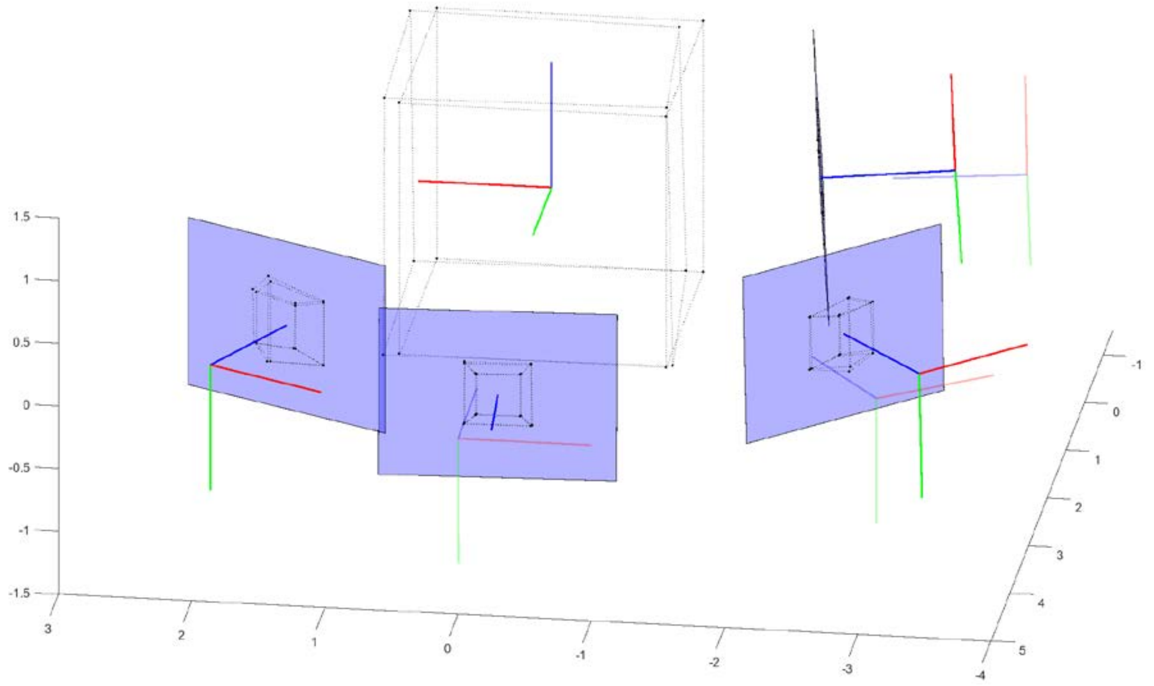
Maximum a Posteriori Inference

Interested in the unknown state variables X, given the measurements Z.

The most often used estimator for X is the MAP estimate:

$$
\begin{aligned}
X^{\text{MAP}} &= \underset{X}{\operatorname{argmax}}\, p(X \mid Z) \\
&= \underset{X}{\operatorname{argmax}} \frac{p(Z \mid X)\, p(X)}{p(Z)} \\
&= \underset{X}{\operatorname{argmax}}\, l(X;Z)\, p(X)
\end{aligned}
$$

$$l(X;Z) \propto p(Z \mid X)$$



Maximum a Posteriori Inference

Measurement model:

$$\mathbf{z}_i = h_i(X_i) + \eta, \qquad \eta \sim N(\mathbf{0}, \mathbf{\Sigma}_i)$$

Measurement prediction function:

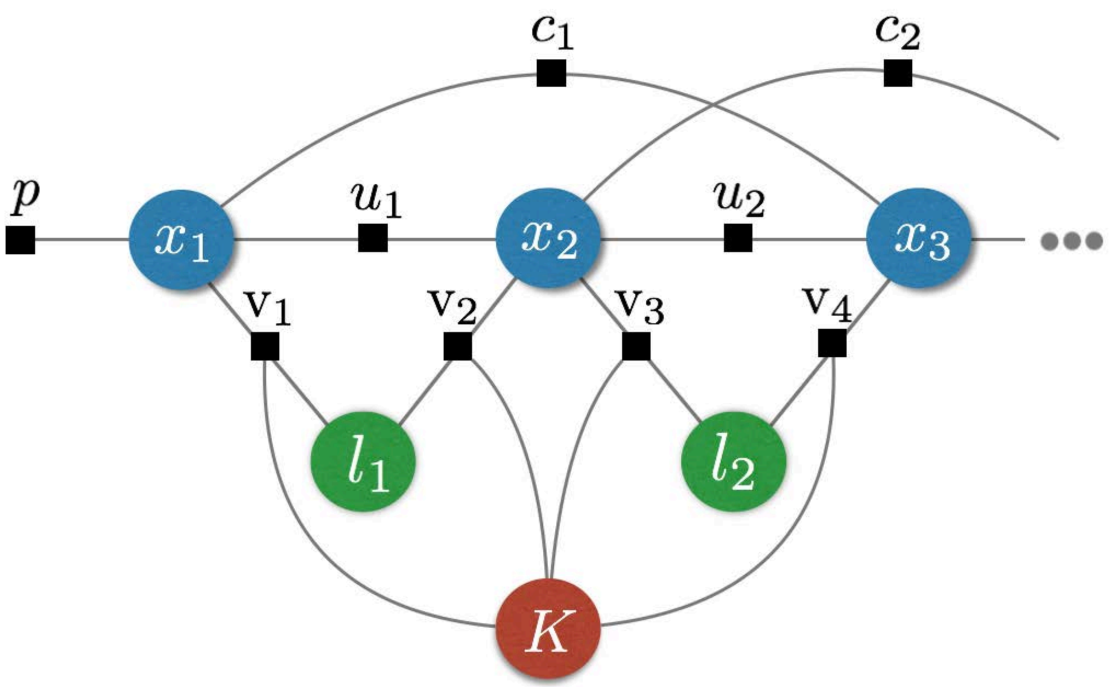
$$\hat{\mathbf{z}}_i = h_i(X_i)$$

Measurement likelihood:

$$p(\mathbf{z}_i \mid X_i) \propto l(X_i; \mathbf{z}_i) = \exp\left(-\frac{1}{2} \left\| h_i(X_i) - \mathbf{z}_i \right\|_{\mathbf{\Sigma}_i}^2 \right)$$

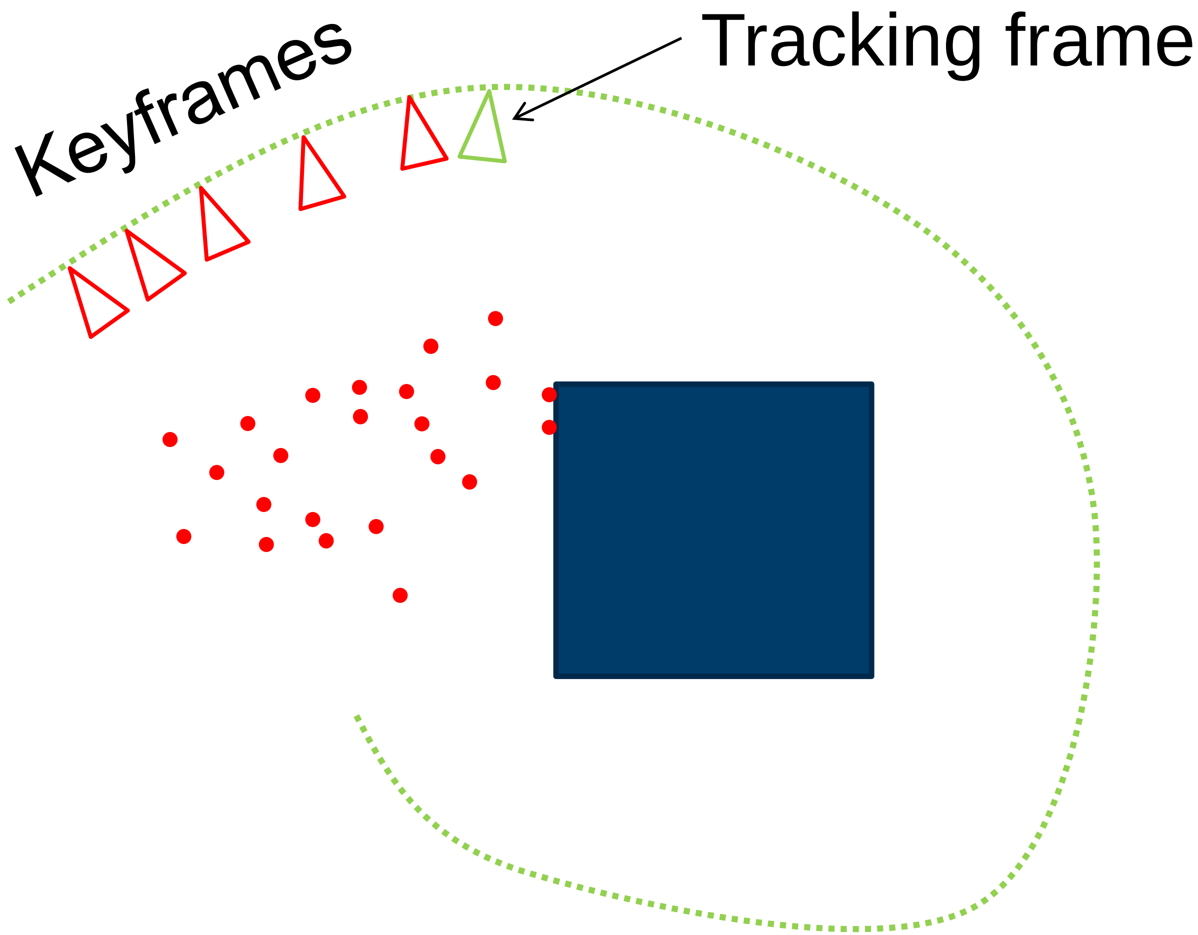
MAP estimate:

$$X^{\text{MAP}} = \underset{X}{\operatorname{argmin}} \sum_i \left\| h_i(X_i) - \mathbf{z}_i \right\|_{\mathbf{\Sigma}_i}^2$$

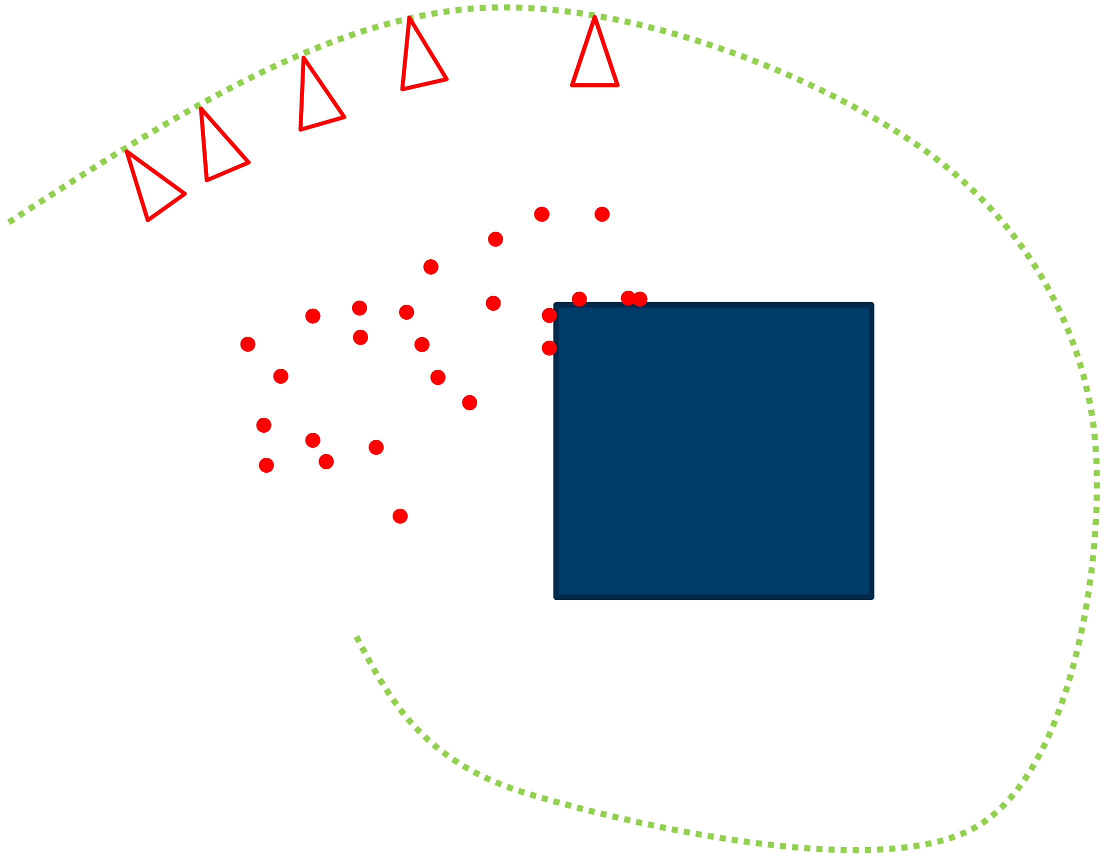


Cadena, C., et al. (2016). Past, Present, and Future of Simultaneous Localization and Mapping: Toward the Robust-Perception Age. *IEEE Transactions on Robotics*, *32*(6), 1309–1332

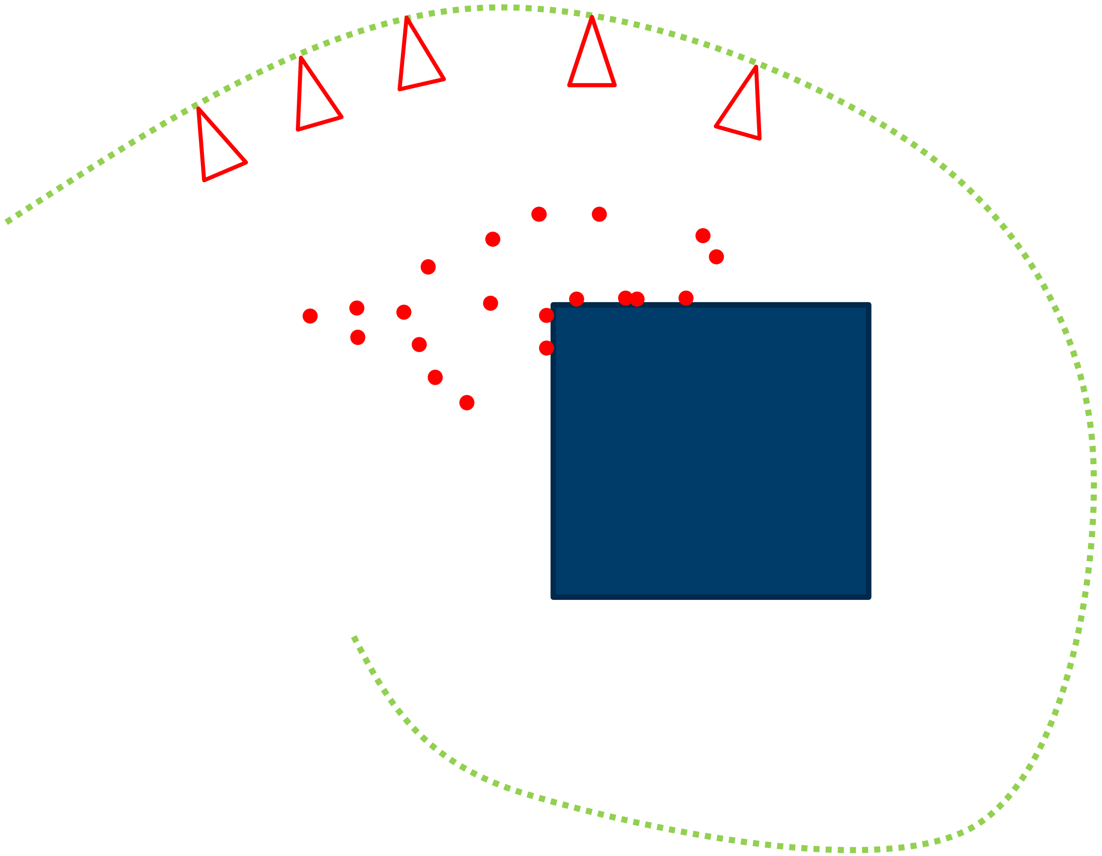
Multi-view mapping



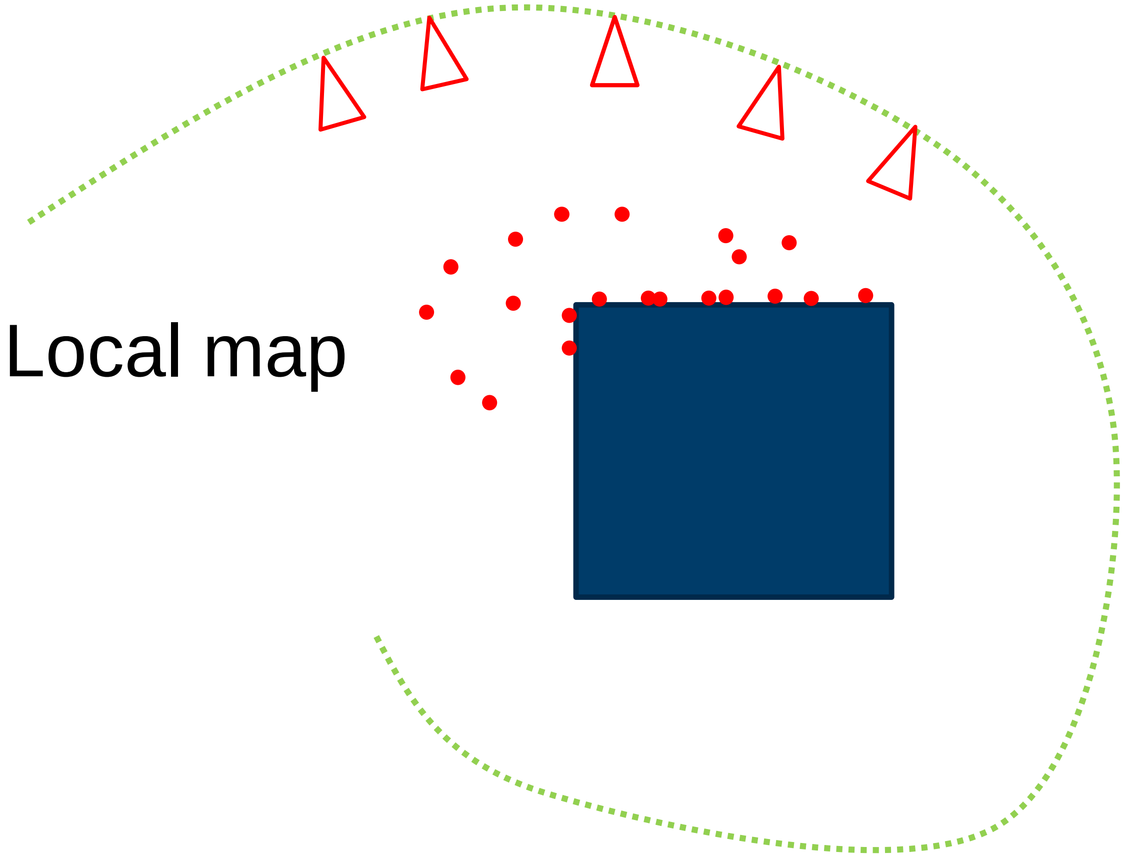
Sliding window mapping



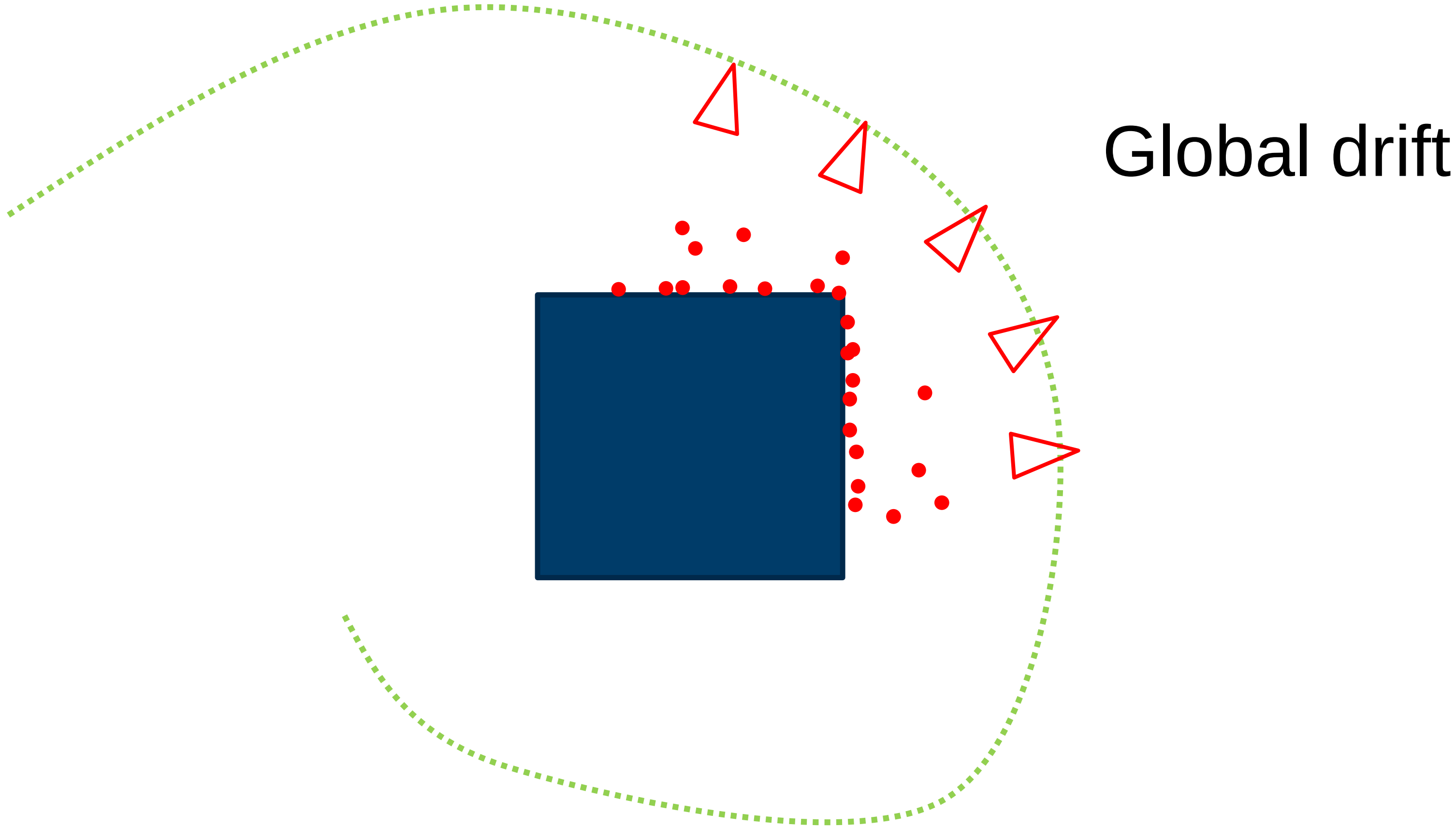
Sliding window mapping



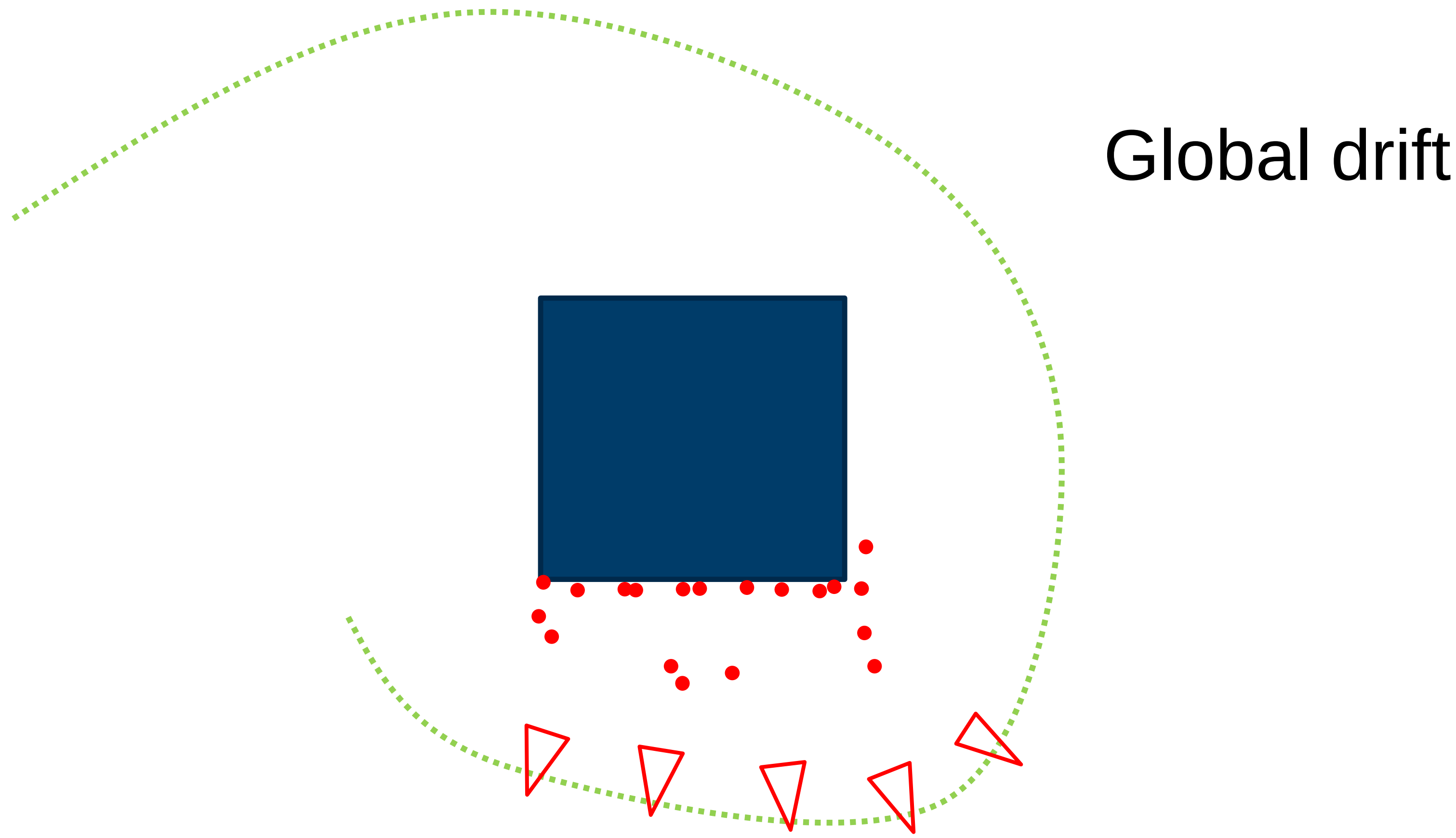
Sliding window mapping



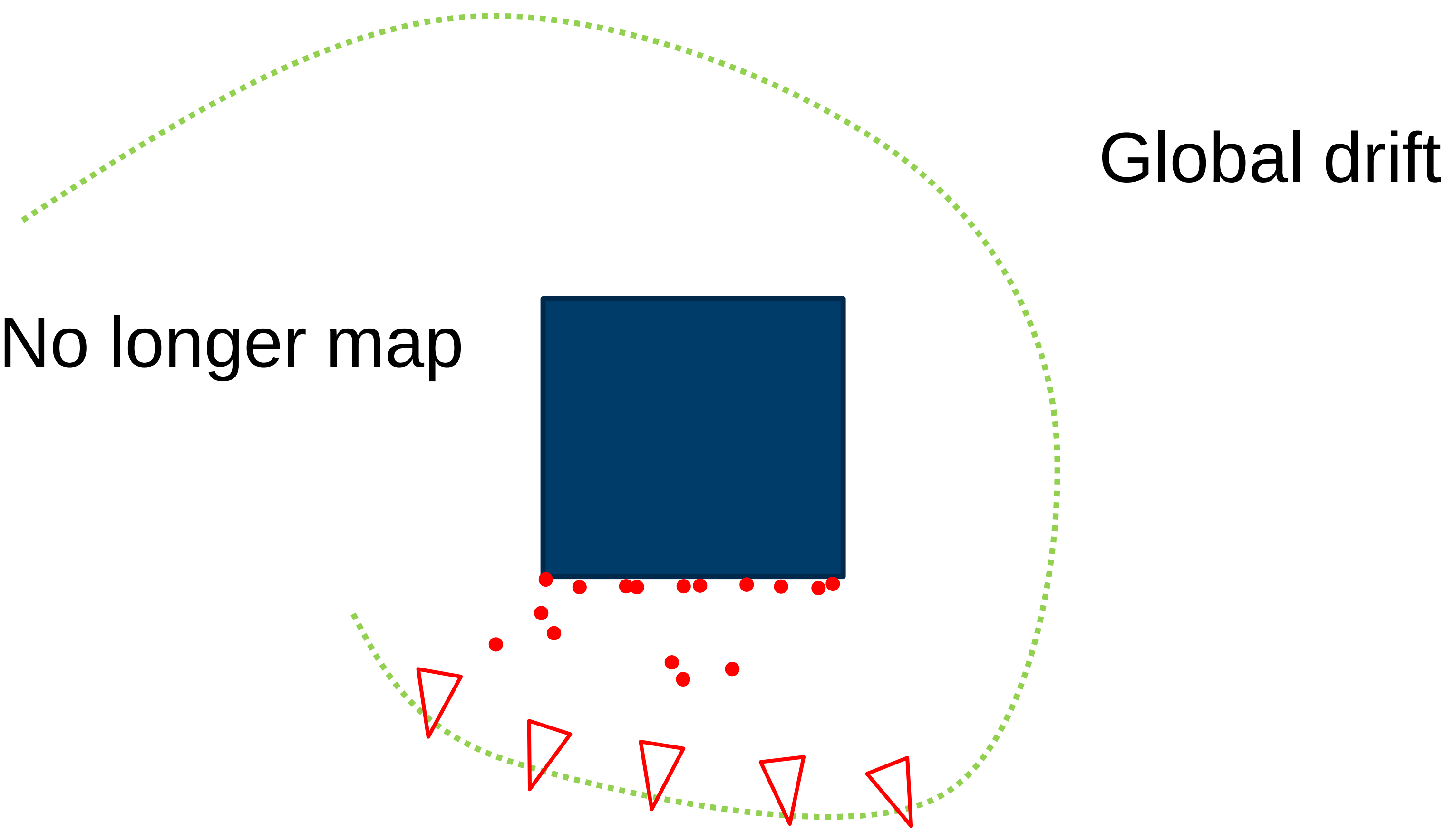
Sliding window mapping



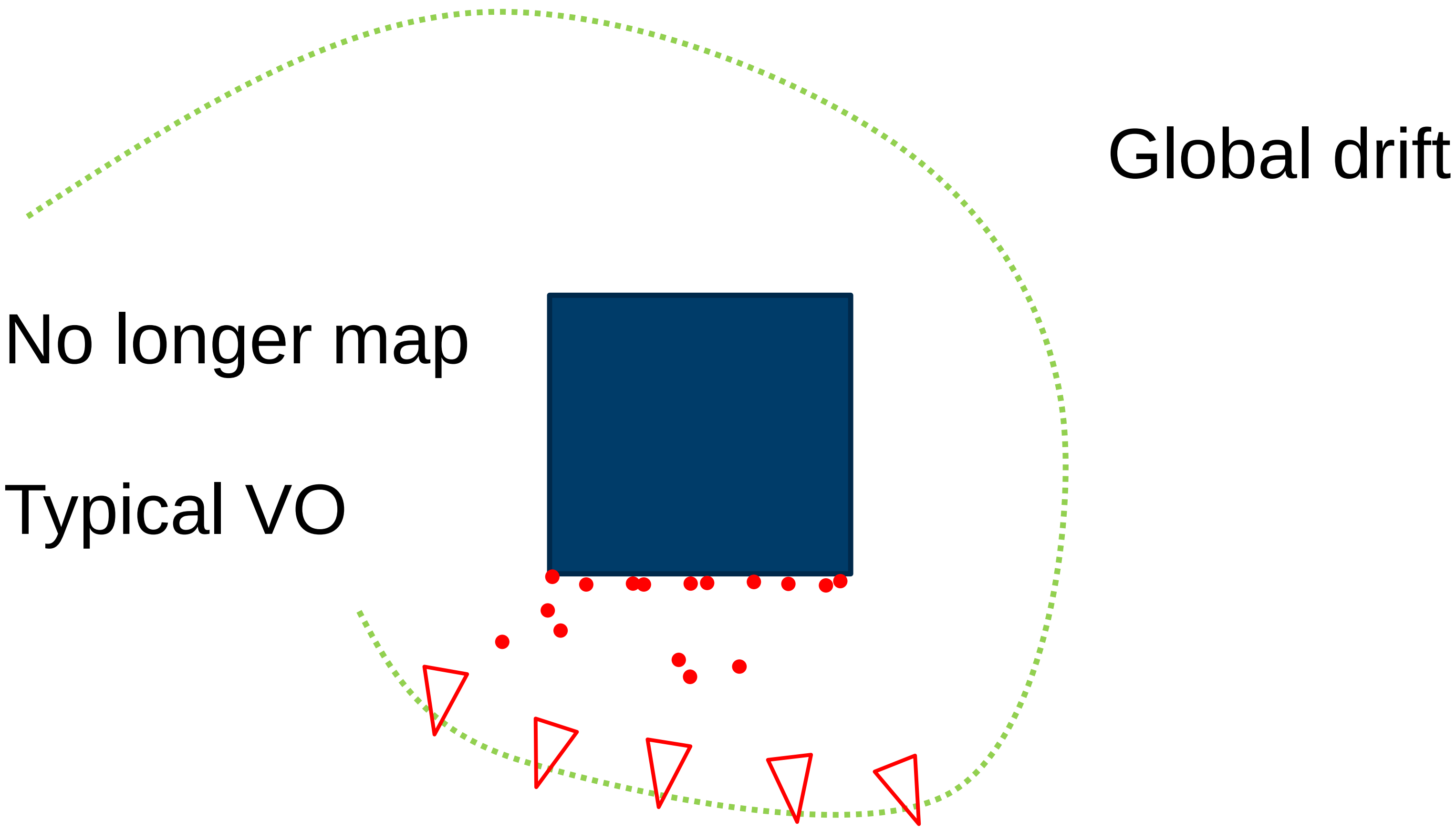
Sliding window mapping



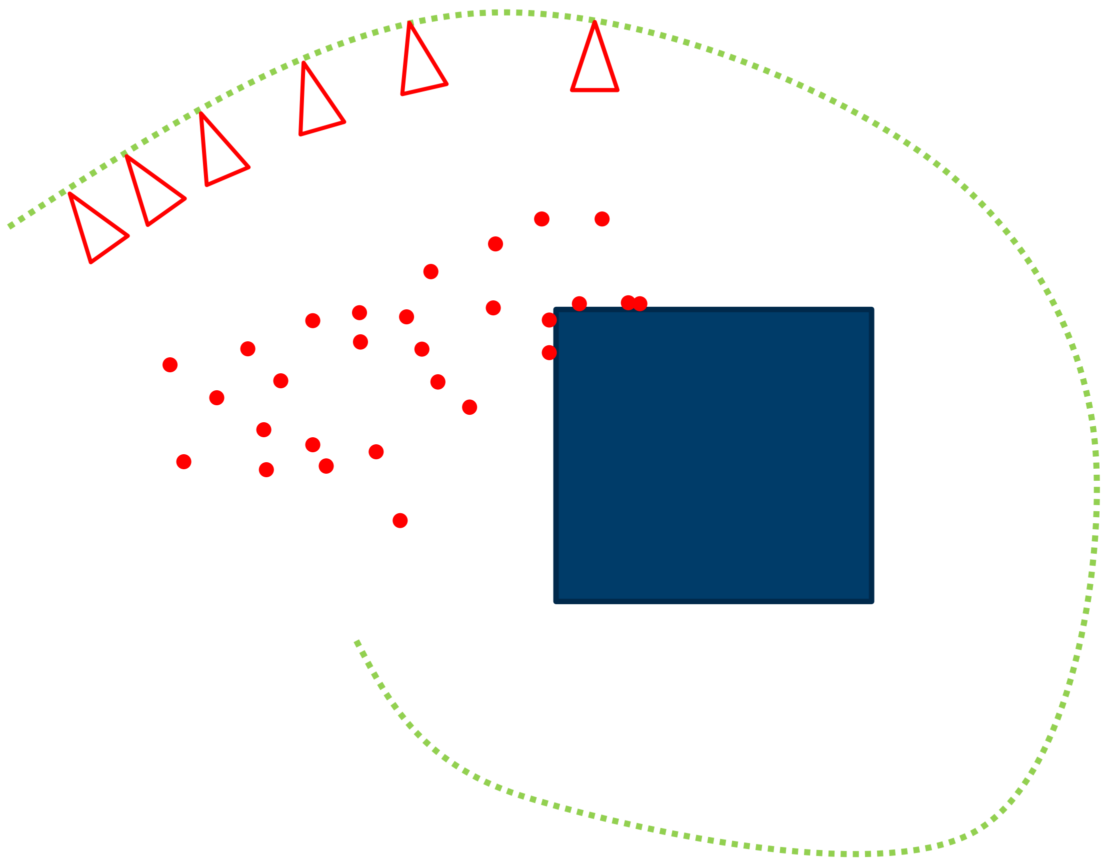
Sliding window mapping



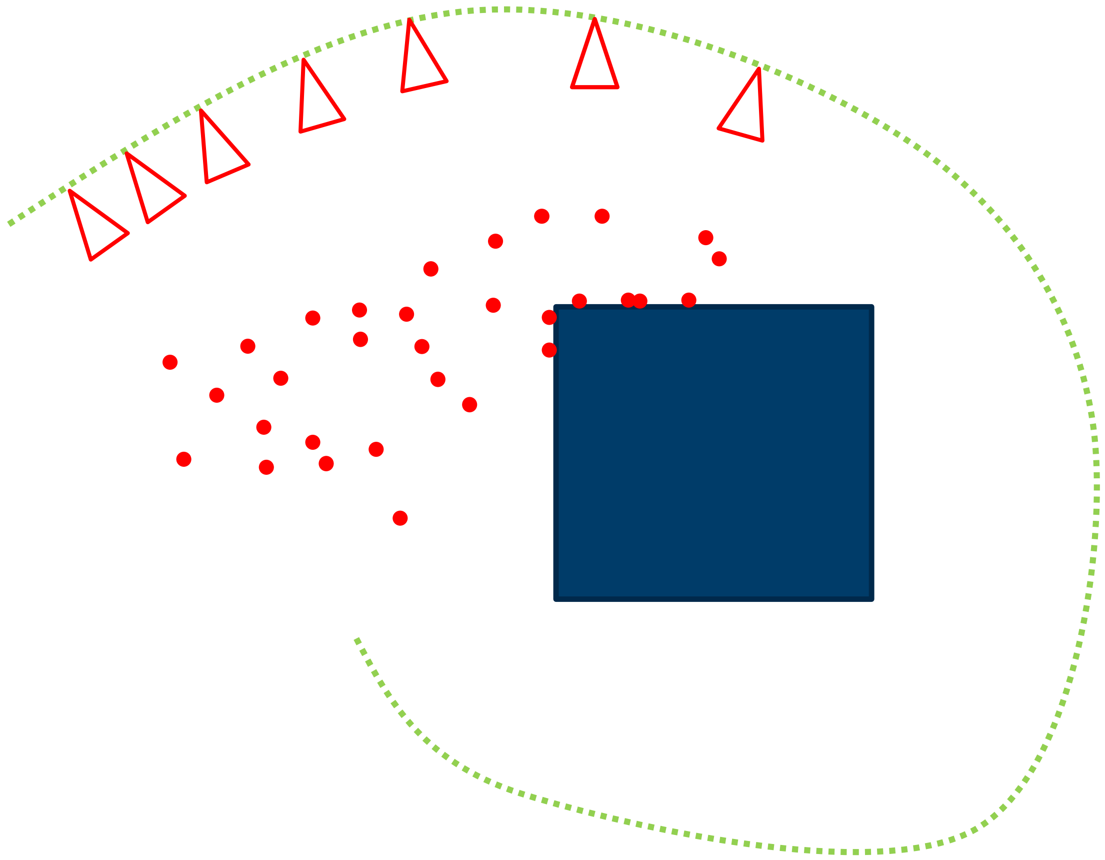
Sliding window mapping



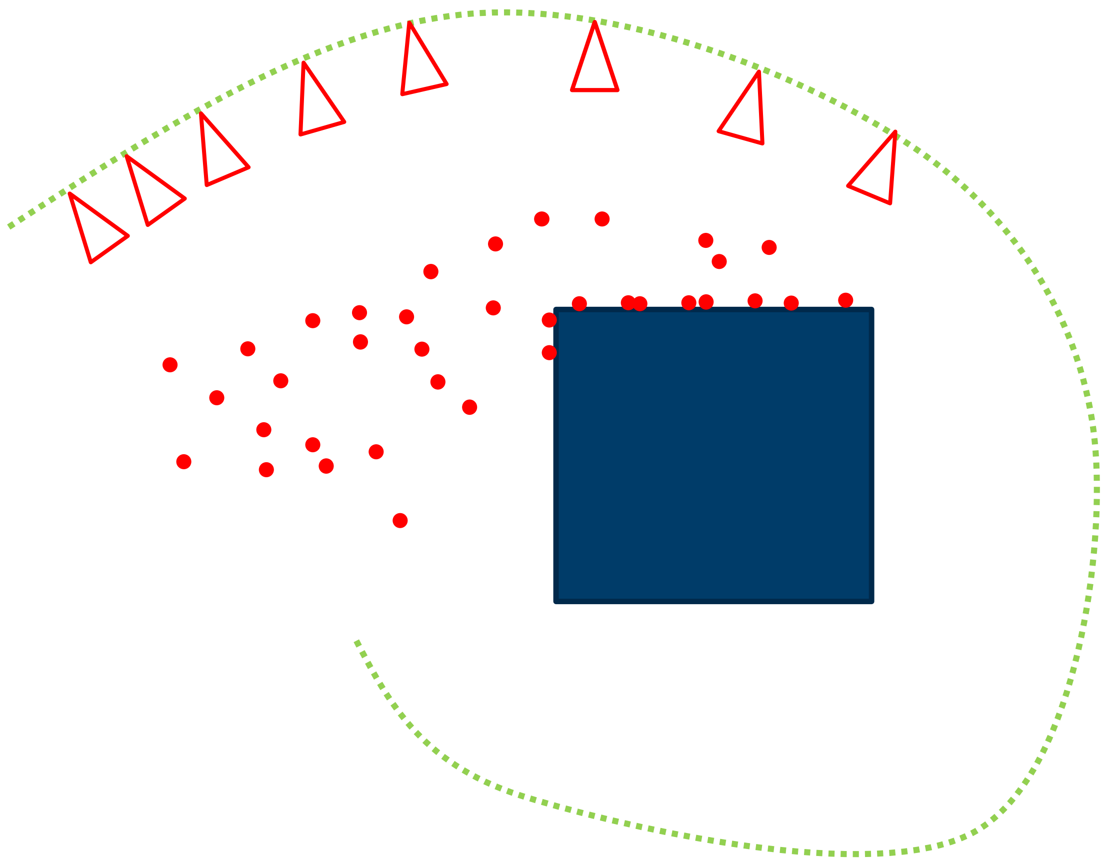
Monocular Visual SLAM



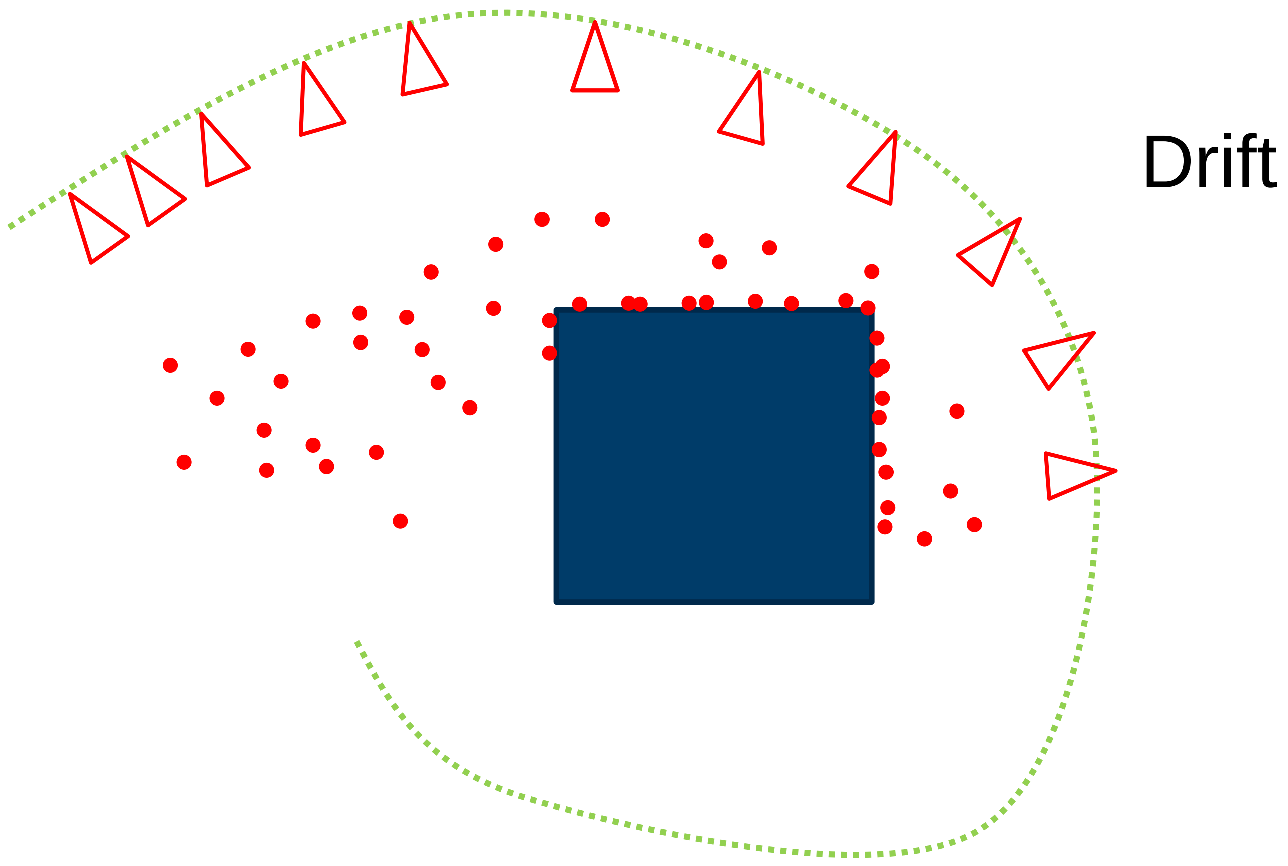
Monocular Visual SLAM



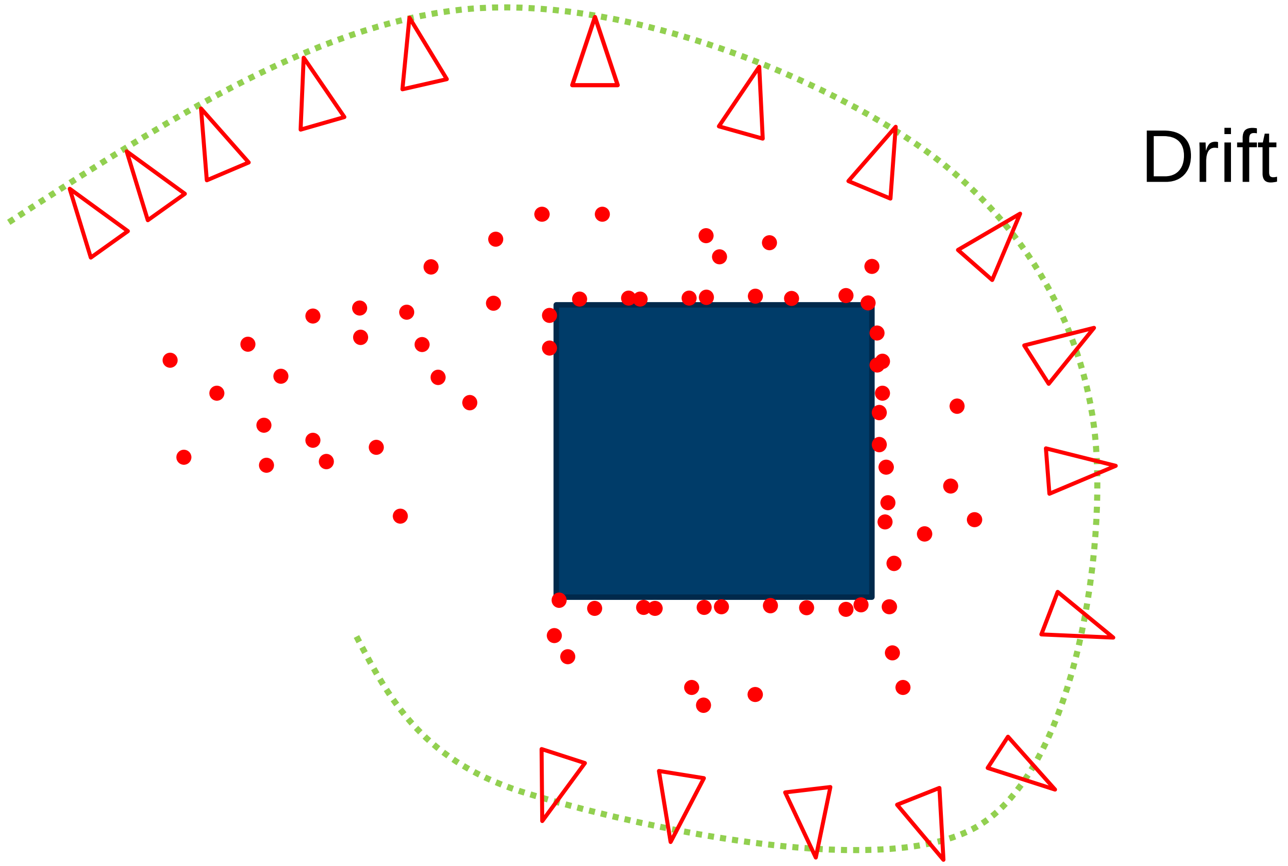
Monocular Visual SLAM



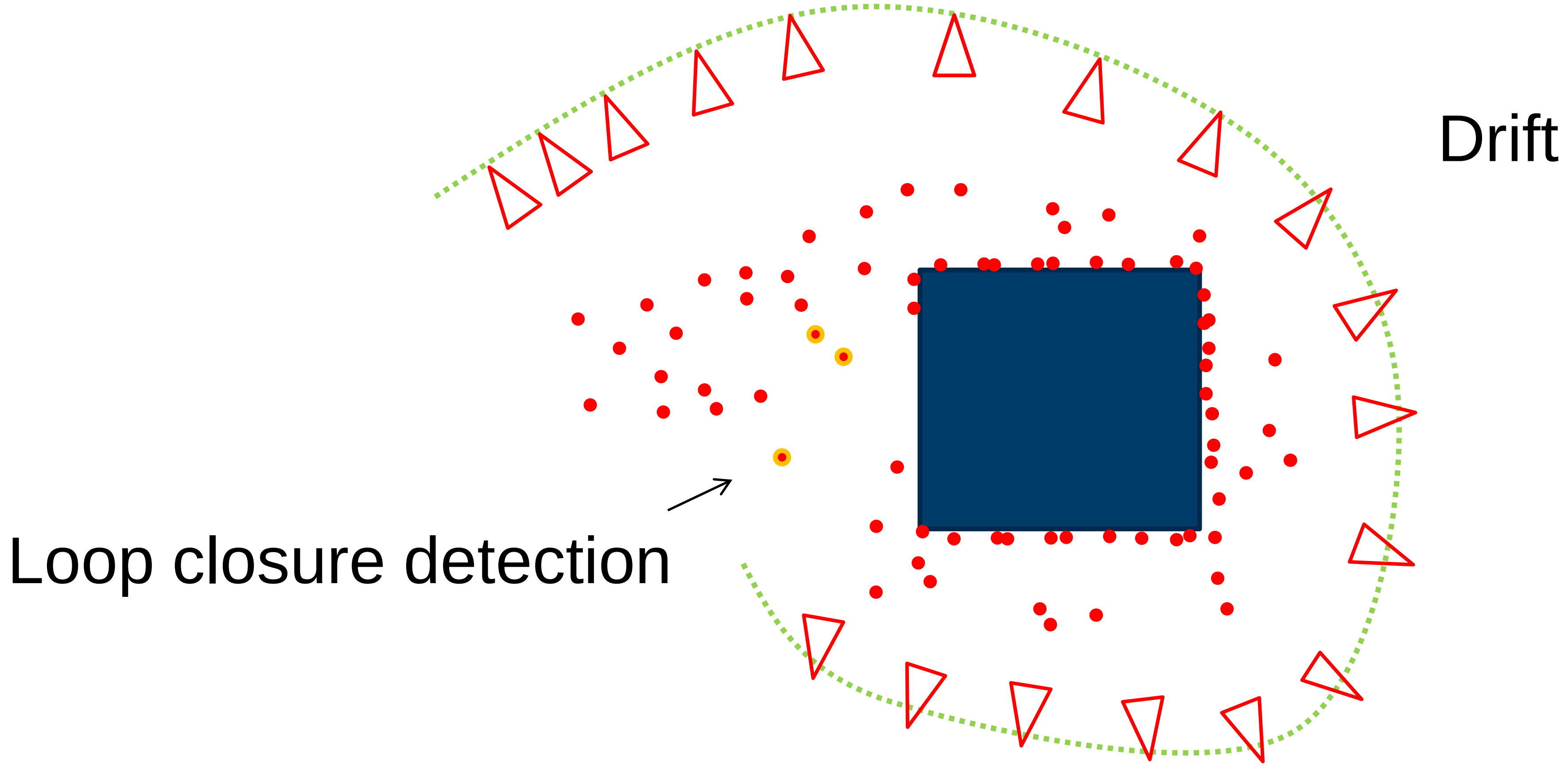
Monocular Visual SLAM



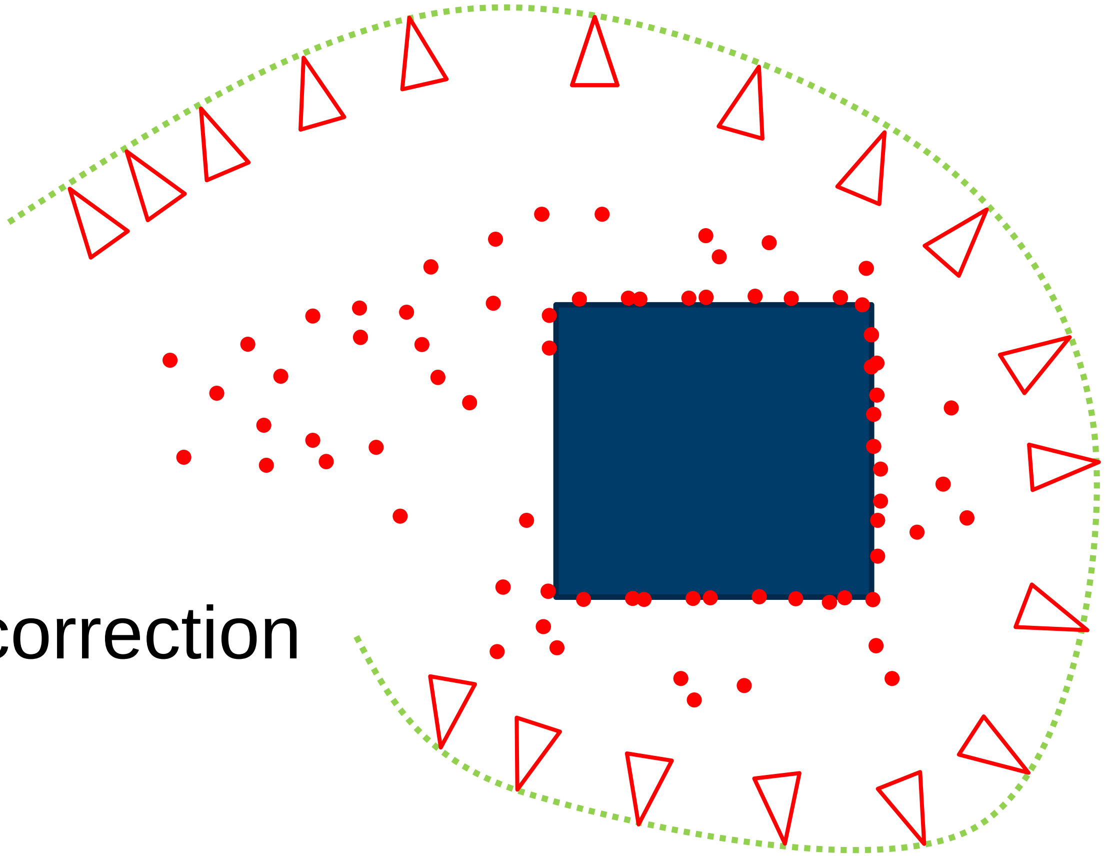
Monocular Visual SLAM



Monocular Visual SLAM

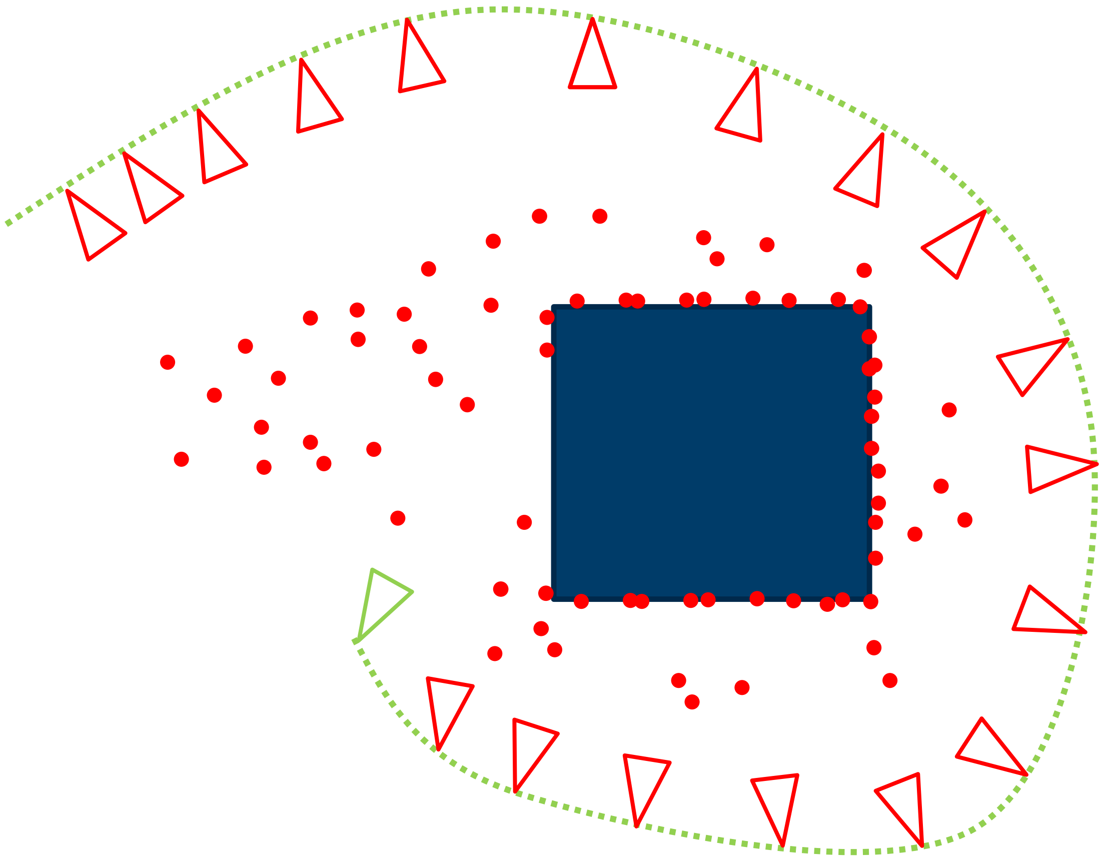


Monocular Visual SLAM

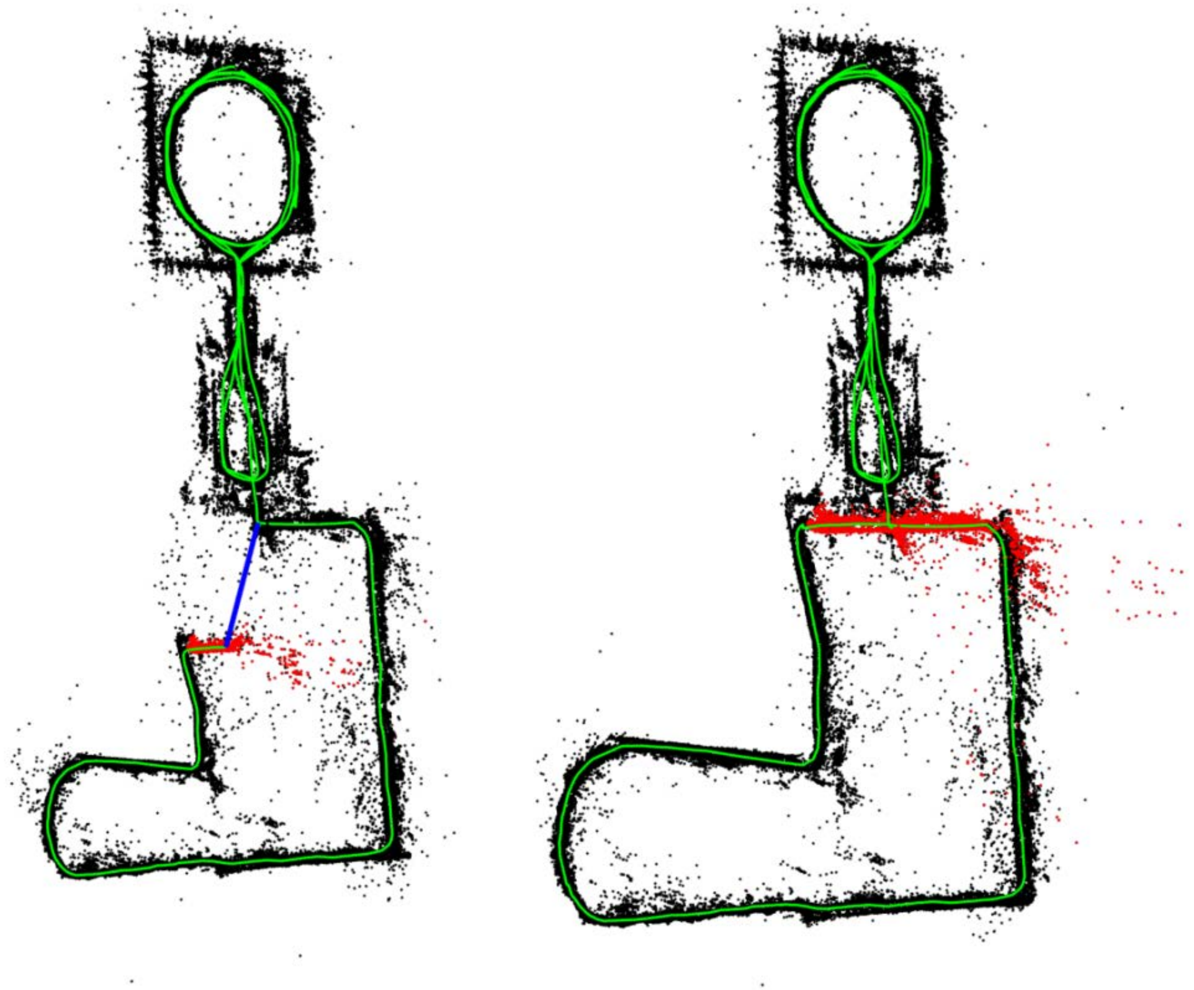


Loop closure correction

Monocular Visual SLAM

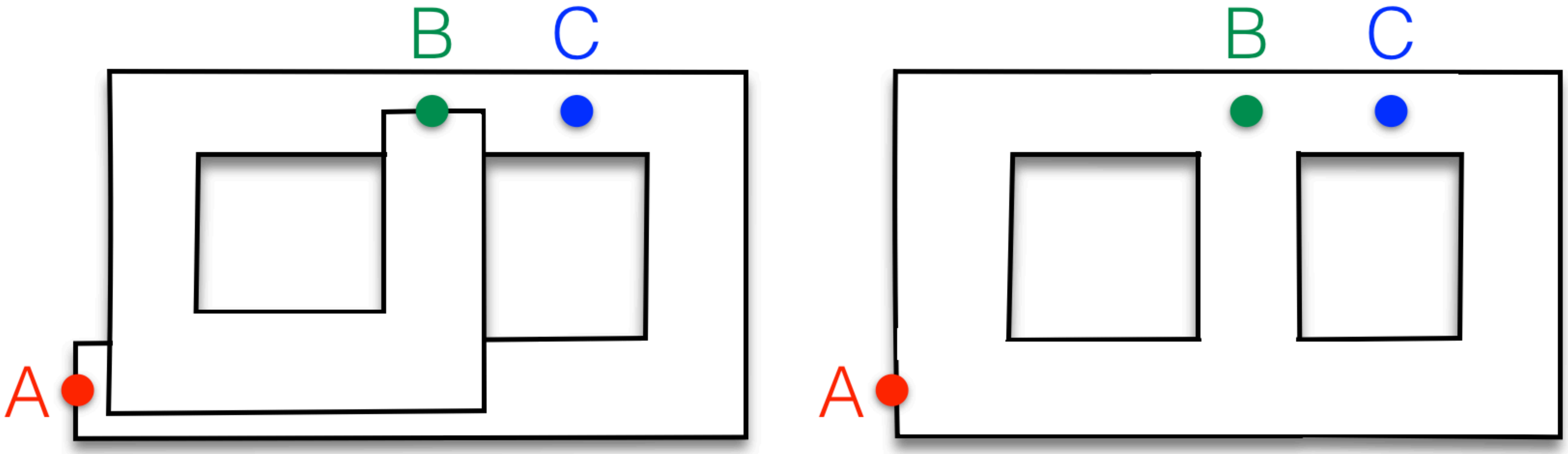


Visual SLAM vs visual odometry



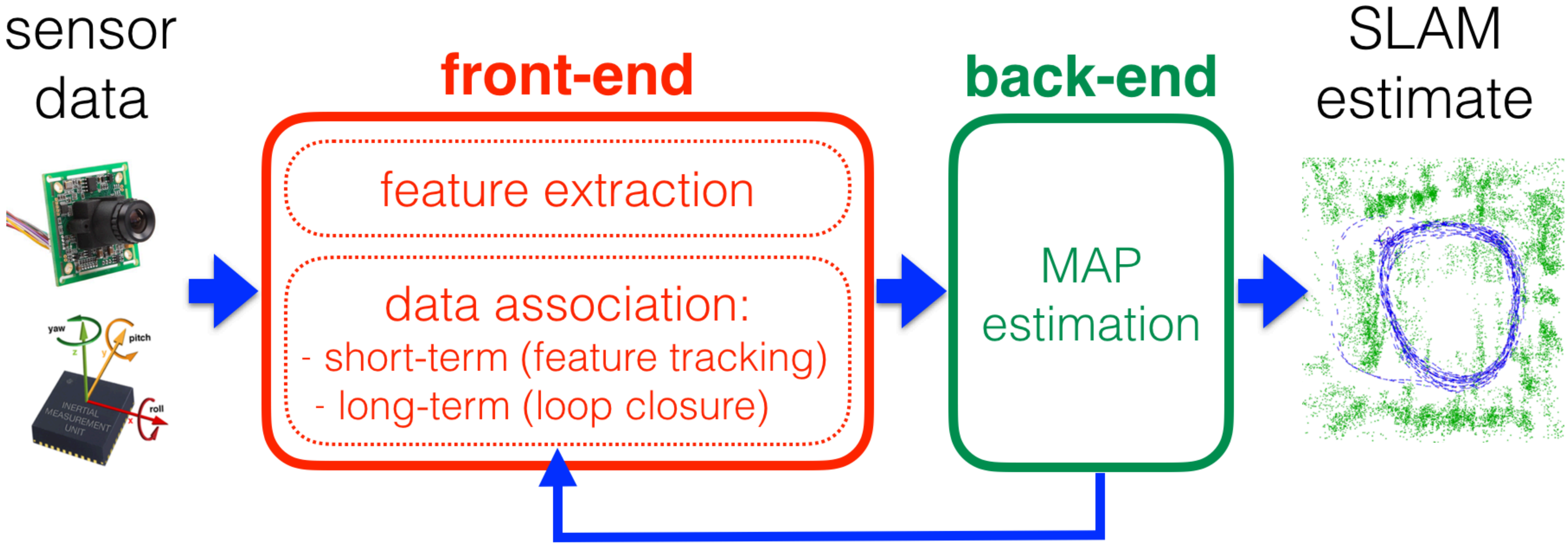
Mur-Artal, R., Montiel, J. M. M., & Tardos, J. D. (2015). ORB-SLAM: A Versatile and Accurate Monocular SLAM System. *IEEE Transactions on Robotics*, *31*(5), 1147–1163

Visual SLAM vs visual odometry



Cadena, C., et al. (2016). Past, Present, and Future of Simultaneous Localization and Mapping:
Toward the Robust-Perception Age. *IEEE Transactions on Robotics*, *32*(6), 1309–1332

Components of SLAM



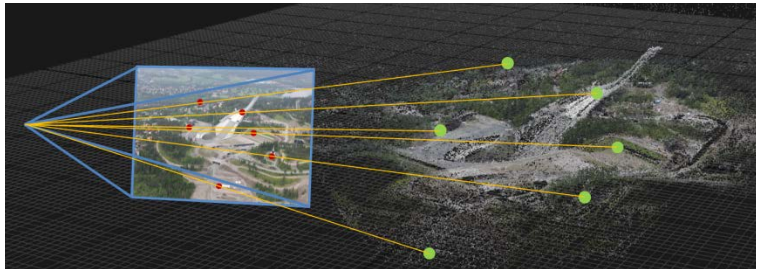
Cadena, C., et al. (2016). Past, Present, and Future of Simultaneous Localization and Mapping: Toward the Robust-Perception Age. *IEEE Transactions on Robotics*, *32*(6), 1309–1332

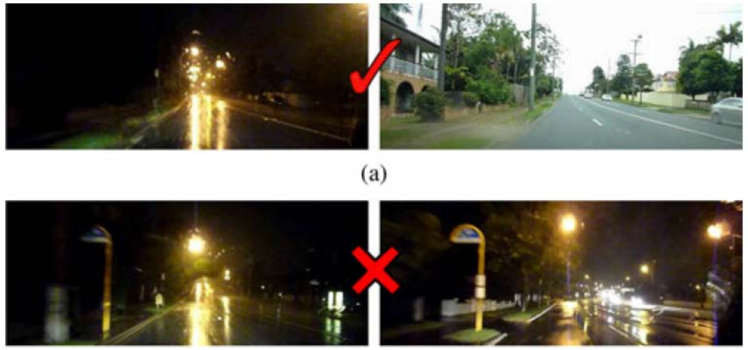
Components of VSLAM

- Short-term tracking
  - Pose estimation given the map
  - Keyframe proposals

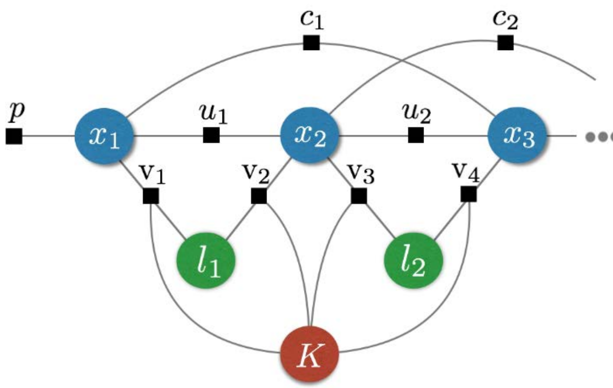
- Long-term tracking
  - Visual place recognition
  - Loop closure detection over keyframes

- Mapping
  - Building and optimizing the map over keyframes
  - Data fusion



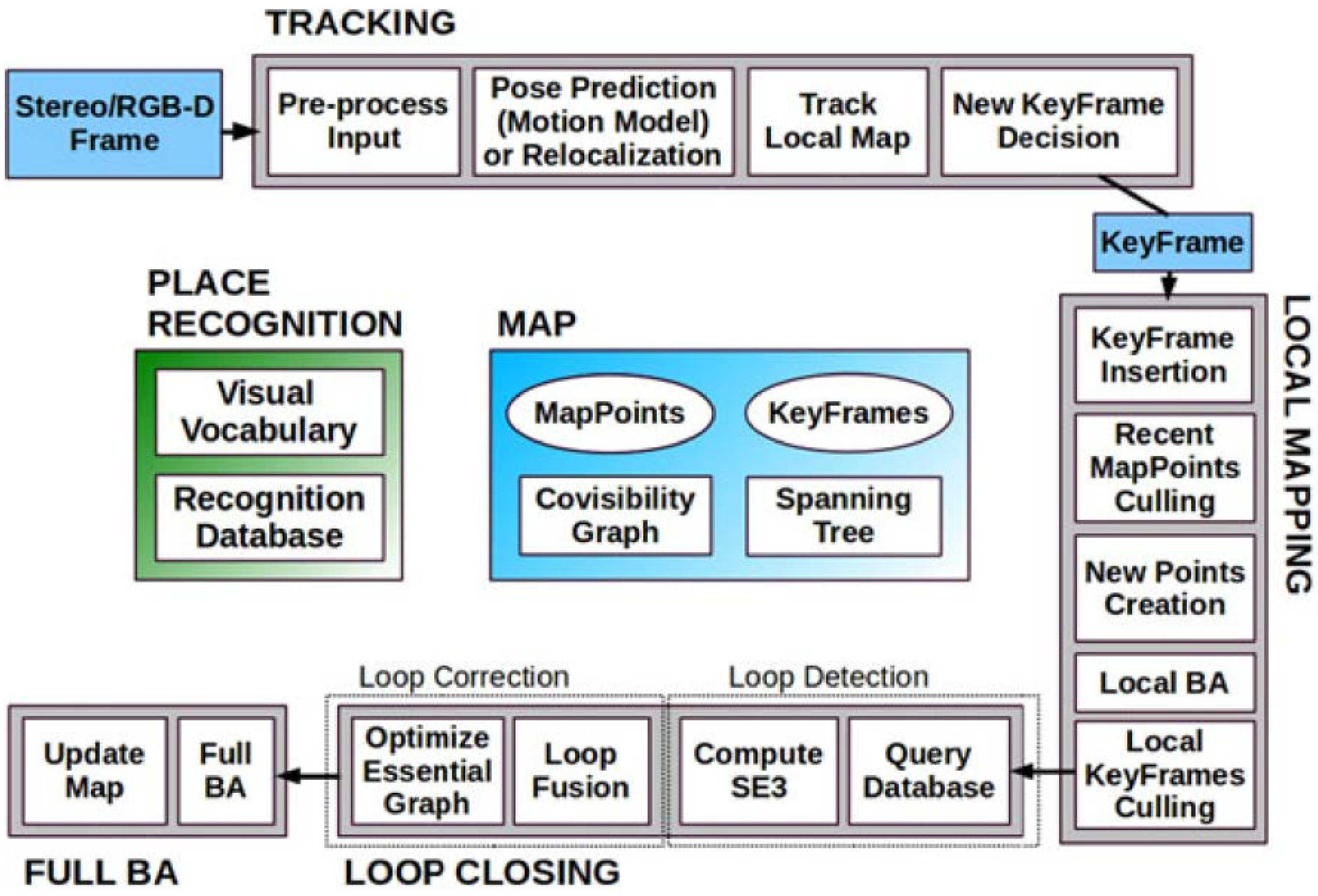


Lowry, S. et al. (2016). Visual Place Recognition: A Survey. IEEE Transactions on Robotics, 32(1), 1–19.



Cadena, C., et al. (2016). Past, Present, and Future of Simultaneous Localization and Mapping: Toward the Robust-Perception Age. *IEEE Transactions on Robotics*, *32*(6), 1309–1332

Example: ORB-SLAM 2



R. Mur-Artal and J. D. Tardos, "ORB-SLAM2: An Open-Source SLAM System for Monocular, Stereo, and RGB-D Cameras," *IEEE Trans. Robot.*, pp. 1–8, 2017.

Supplementary material

- “Parallel Tracking and Mapping for Small AR Workspaces”,
  Klein and Murray,
  In Proc. International Symposium on Mixed and Augmented Reality (ISMAR'07, Nara), 2007
  https://www.robots.ox.ac.uk/~vgg/rg/papers/klein_murray__2007__ptam.pdf

- “Past, Present, and Future of Simultaneous Localization And Mapping: Towards the Robust-Perception Age”,
  Cadena et al., IEEE Transactions on Robotics 32 (6) pp 1309-1332, 2016
  https://arxiv.org/abs/1606.05830

- “Visual Place Recognition: A Survey“,
  Lowry, S. et al., IEEE Transactions on Robotics, 32 (1), pp 1–19, 2016
  https://ieeexplore.ieee.org/document/7339473