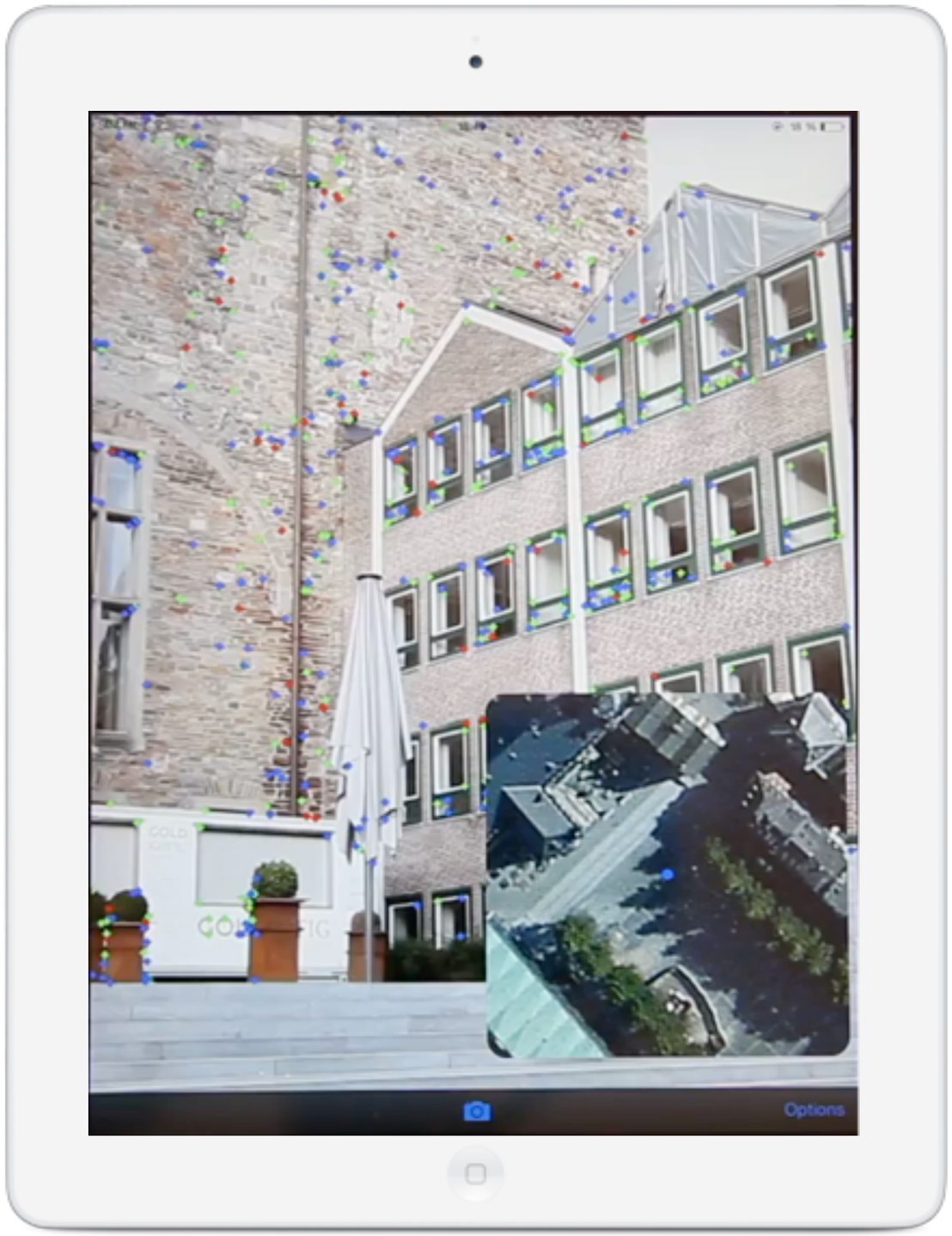
The Challenges of Large-Scale Localisation and Mapping

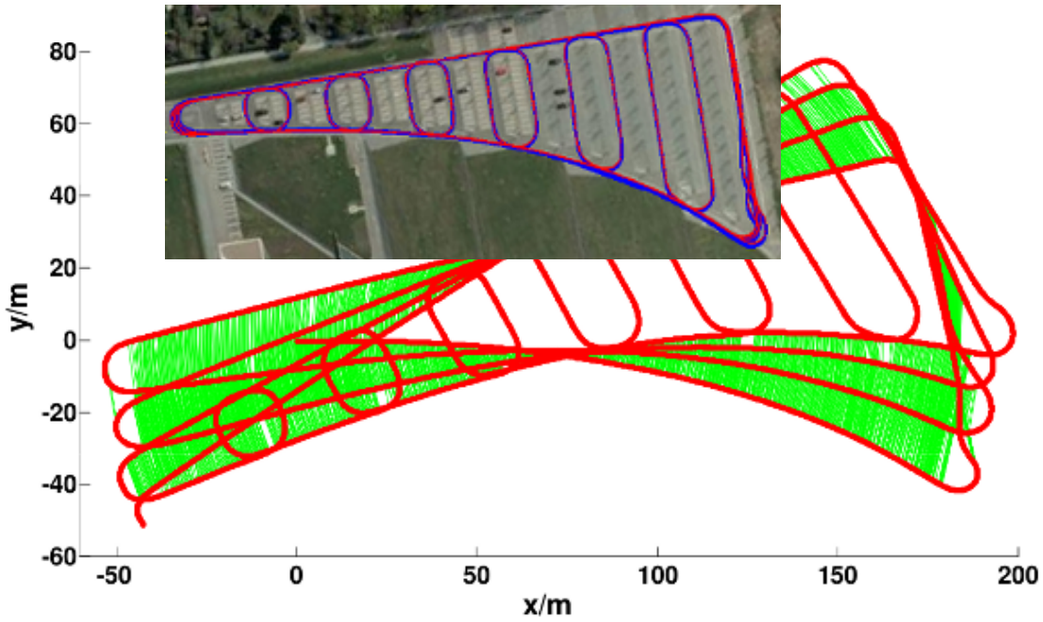
Torsten Sattler
Computer Vision and Geometry Lab
ETH Zürich

Mobile Localisation



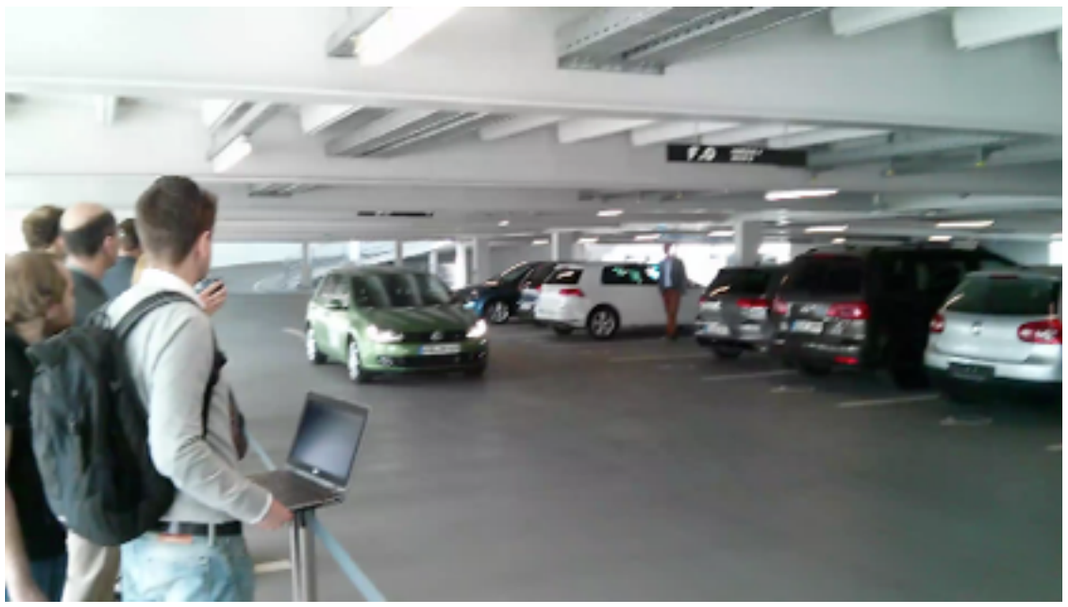
S. Middelberg, T. Sattler, O. Untzelmann, L. Kobbelt. Scalable 6-DOF Localization on Mobile Devices. ECCV 2014.

Mobile Localisation



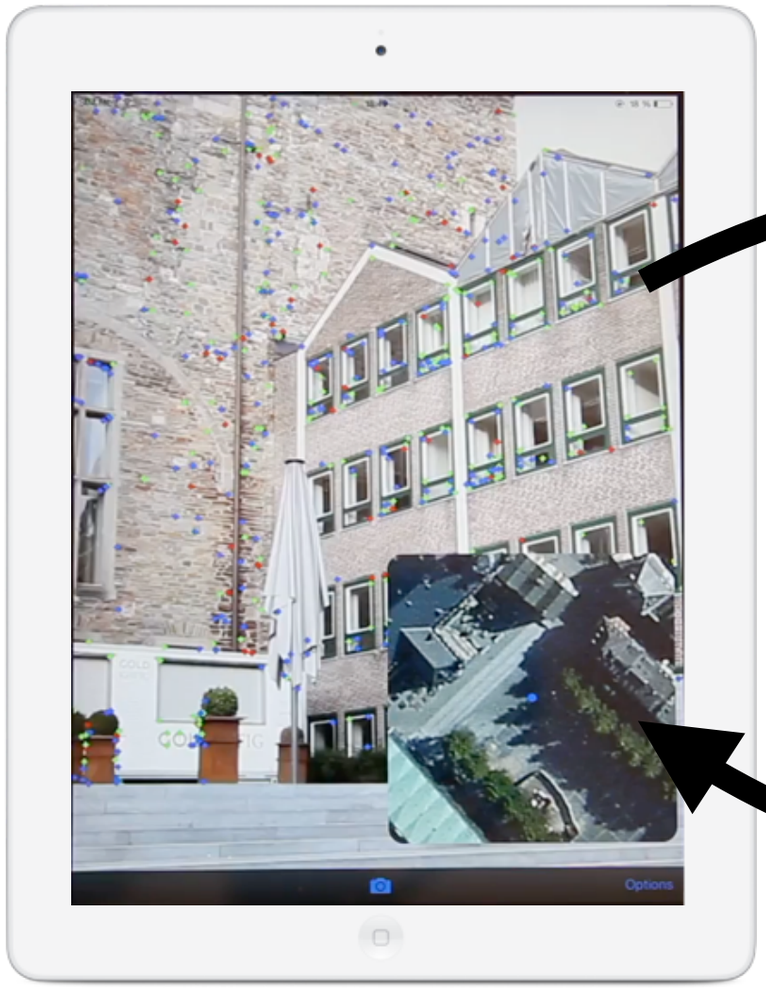
loop closures

AR / VR



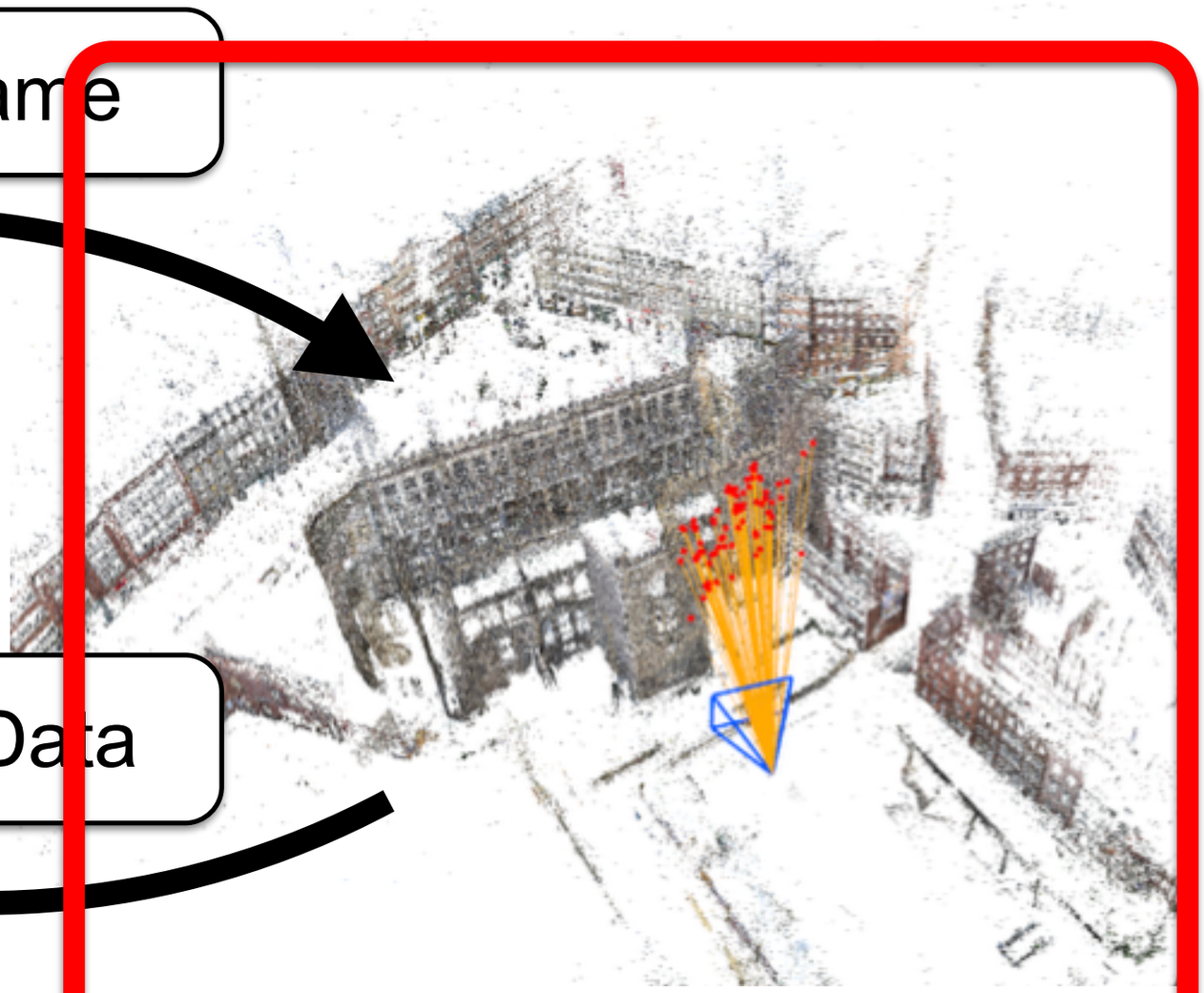
robotics / autonomous vehicles

Mobile Localisation



Keyframe

Pose Data

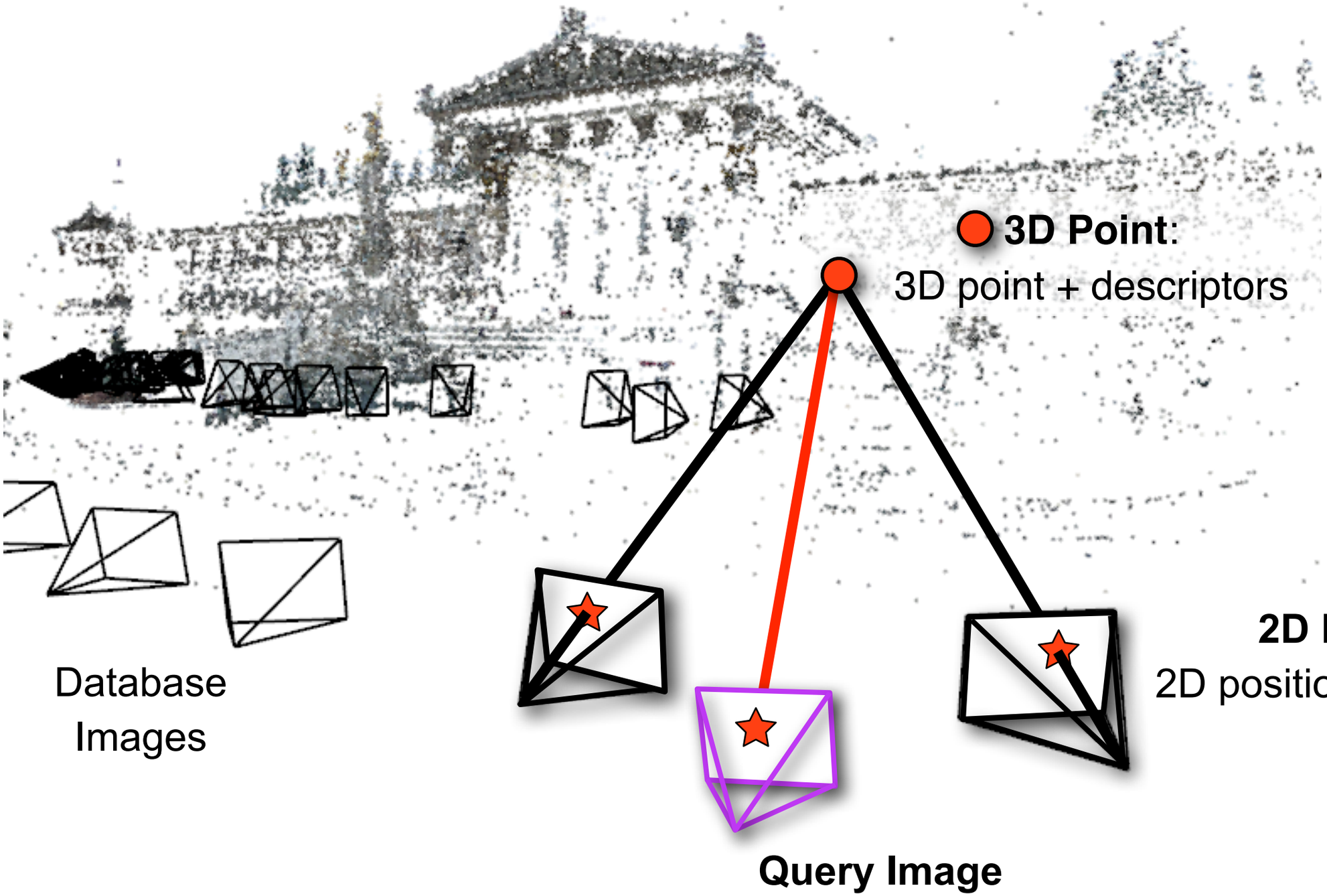


Real-time camera pose tracking

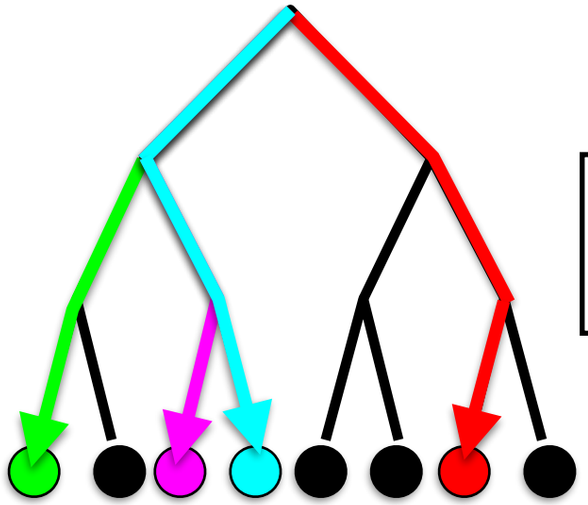
Global localisation against known map

Global Localisation

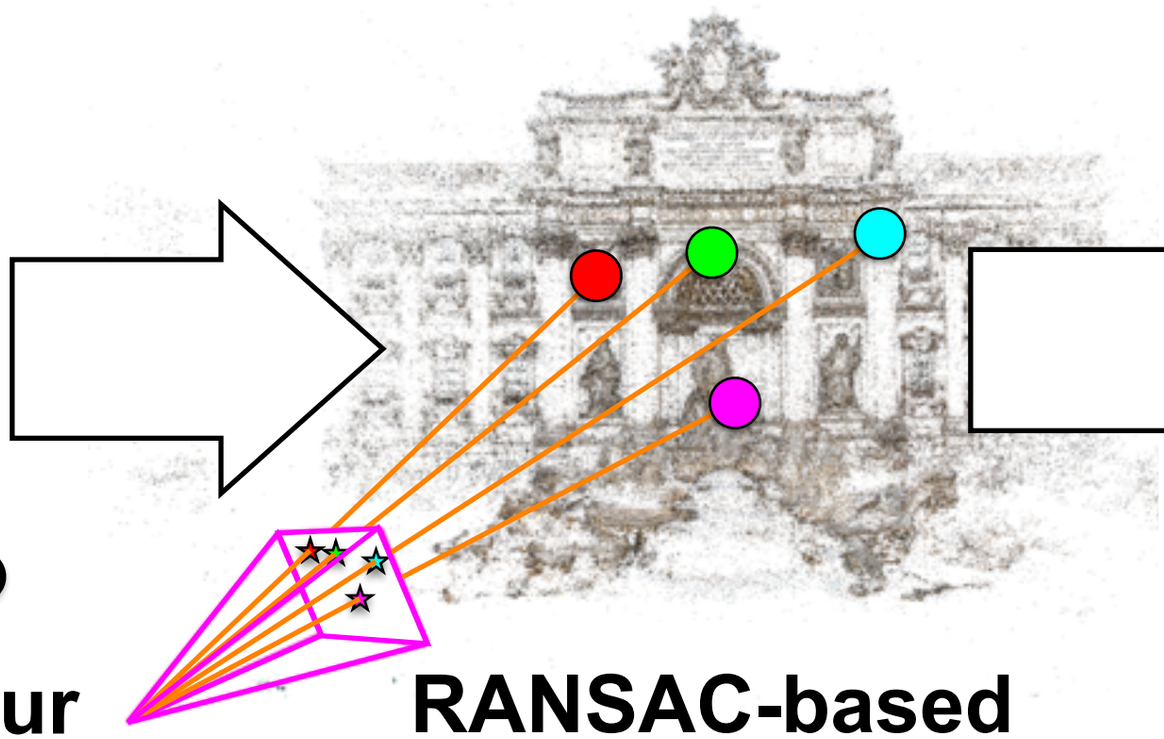
Structure-from-Motion / SLAM model



Global Localisation



Nearest Neighbour Search

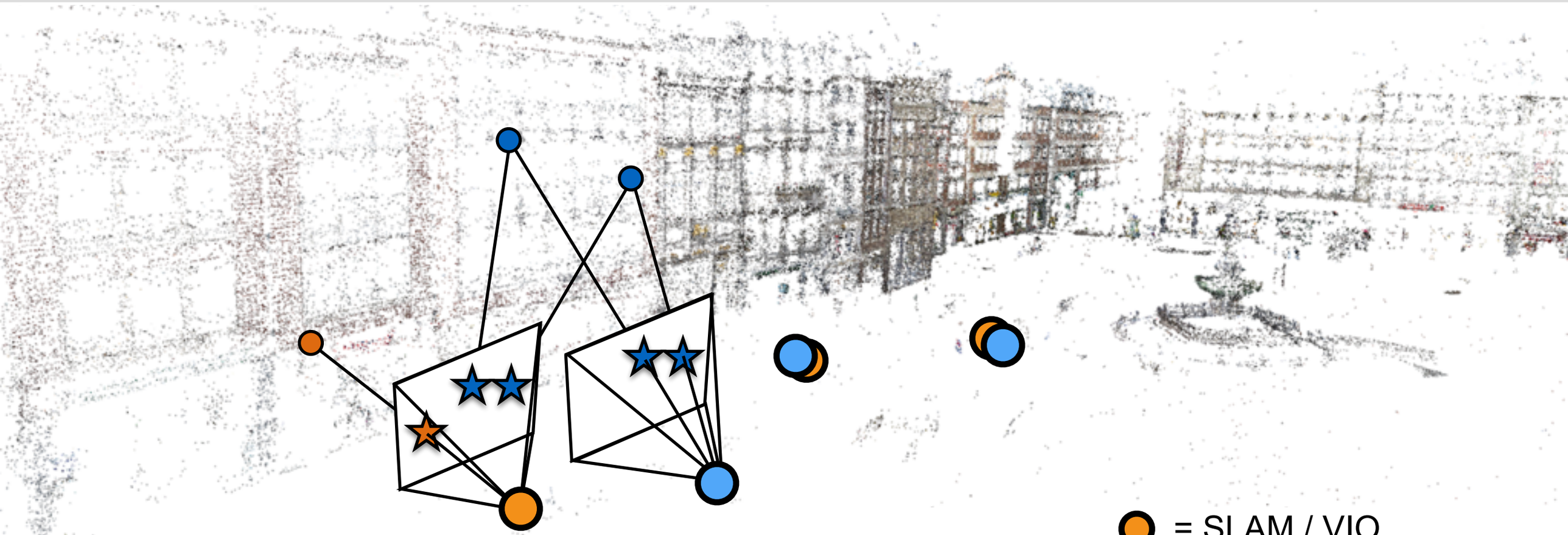


RANSAC-based Camera Pose Estimation

≥X inliers?

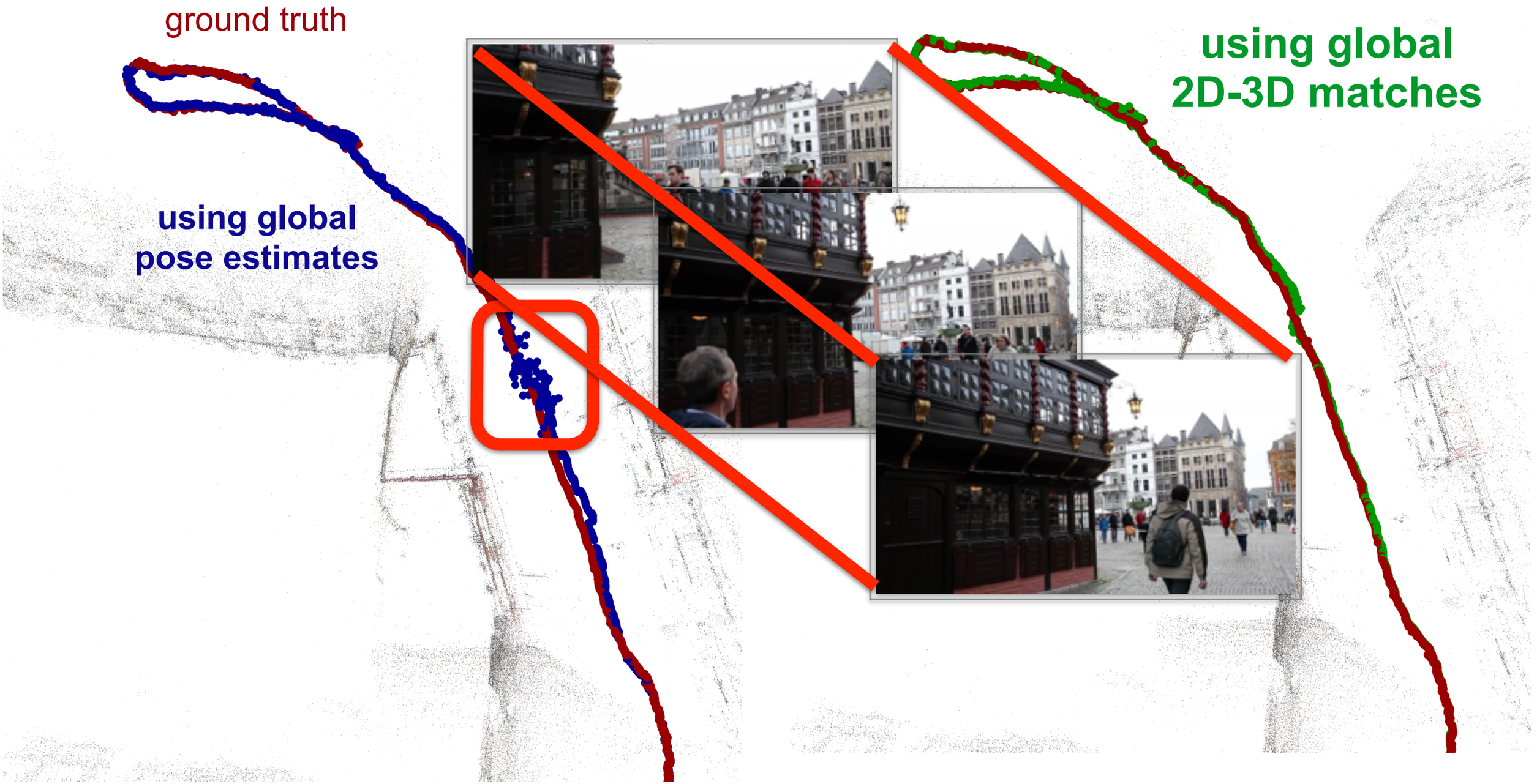
Correct Pose Estimate

Global Pose Integration



- Use inlier 2D-3D matches rather than estimate poses
  - “Control points” in Bundle Adjustment / Kalman Filter
  - Additional measurements for filter-based methods

Matches vs. Poses

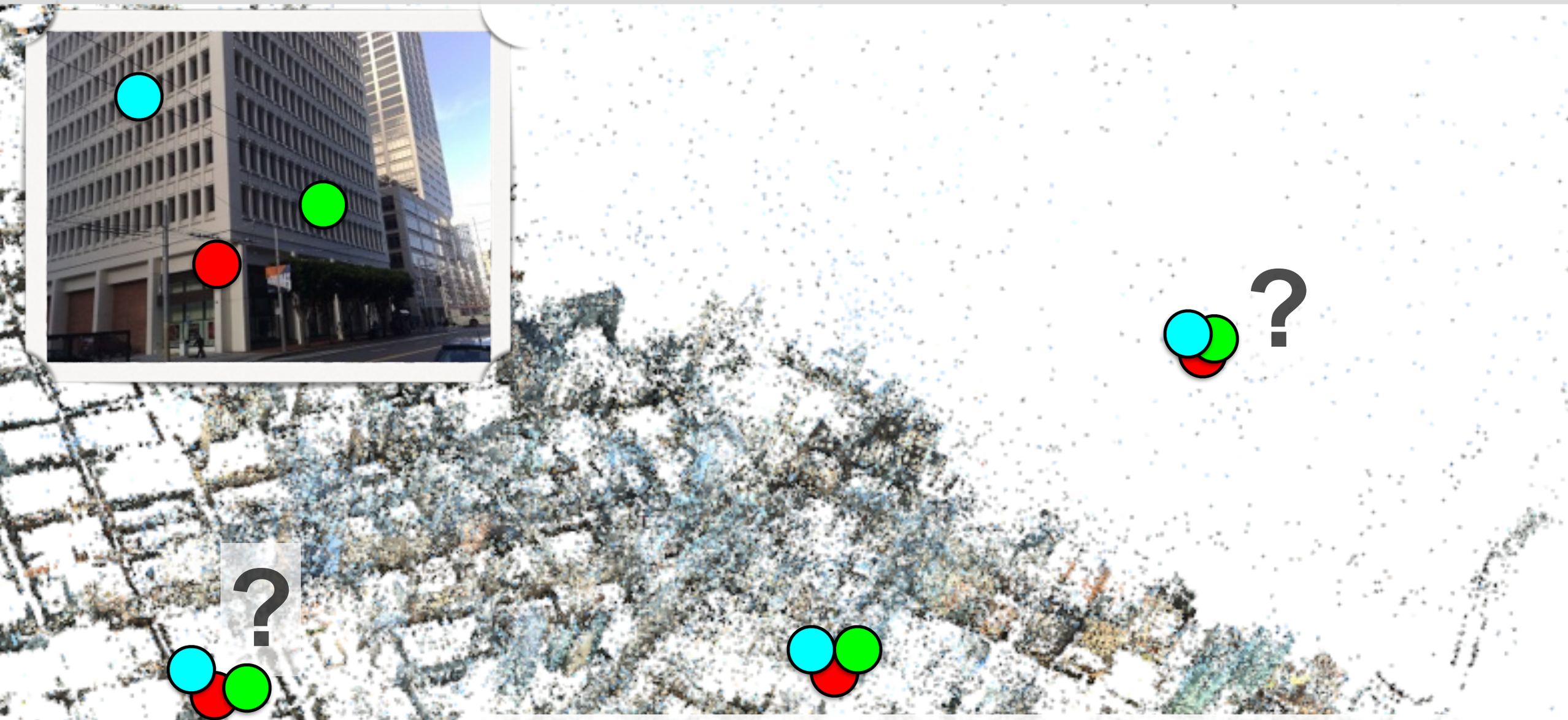


S. Middelberg, T. Sattler, O. Untzelmann, L. Kobbelt. Scalable 6-DOF Localization on Mobile Devices. ECCV 2014.

Asynchronous Localisation

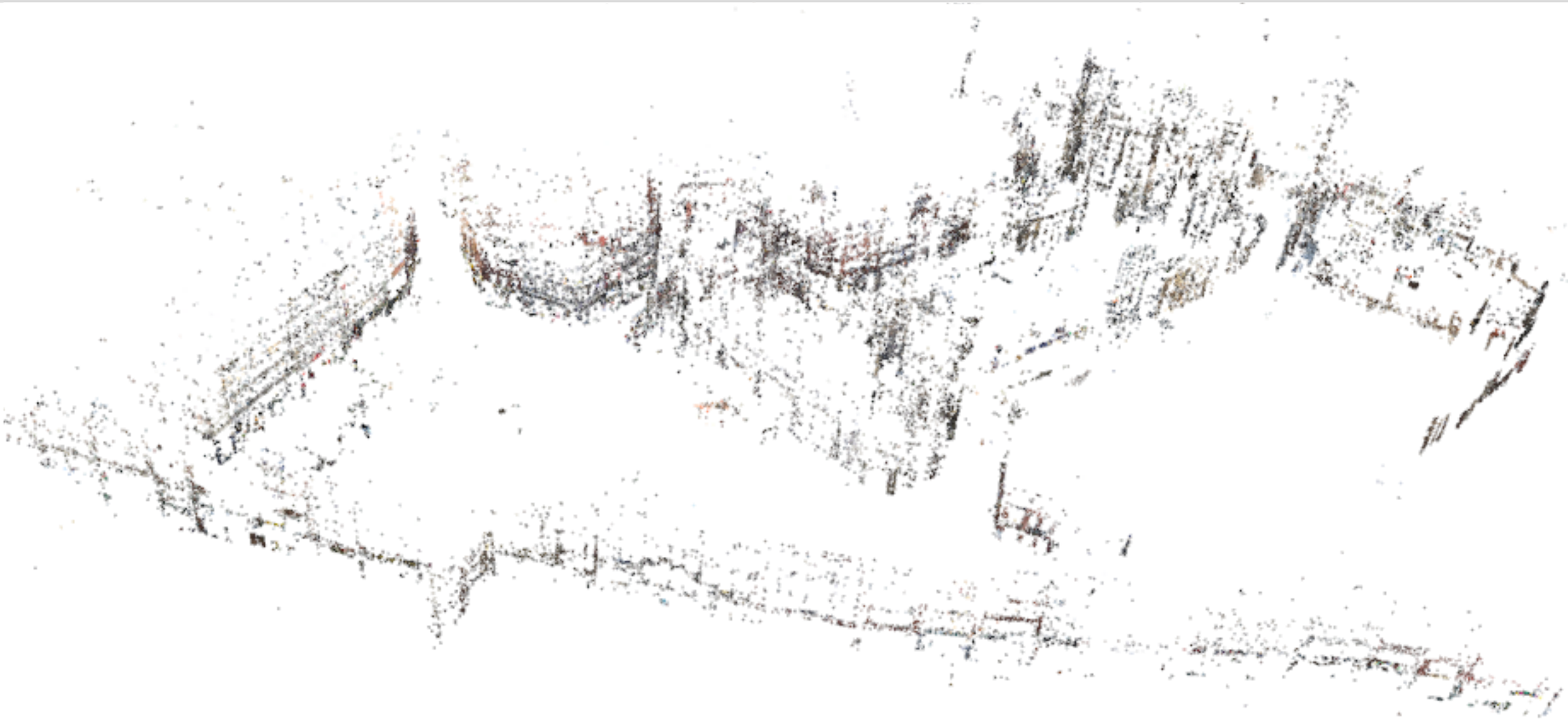
S. Lynen, T. Sattler, M. Bosse, J. Hesch, M. Pollefeys, R. Siegwart,
Get Out of My Lab: Large-scale, Real-Time Visual-Inertial Localization. RSS 2015

Challenges at Large-Scale



- Memory consumption
- Distinguish correct vs. wrong localisations

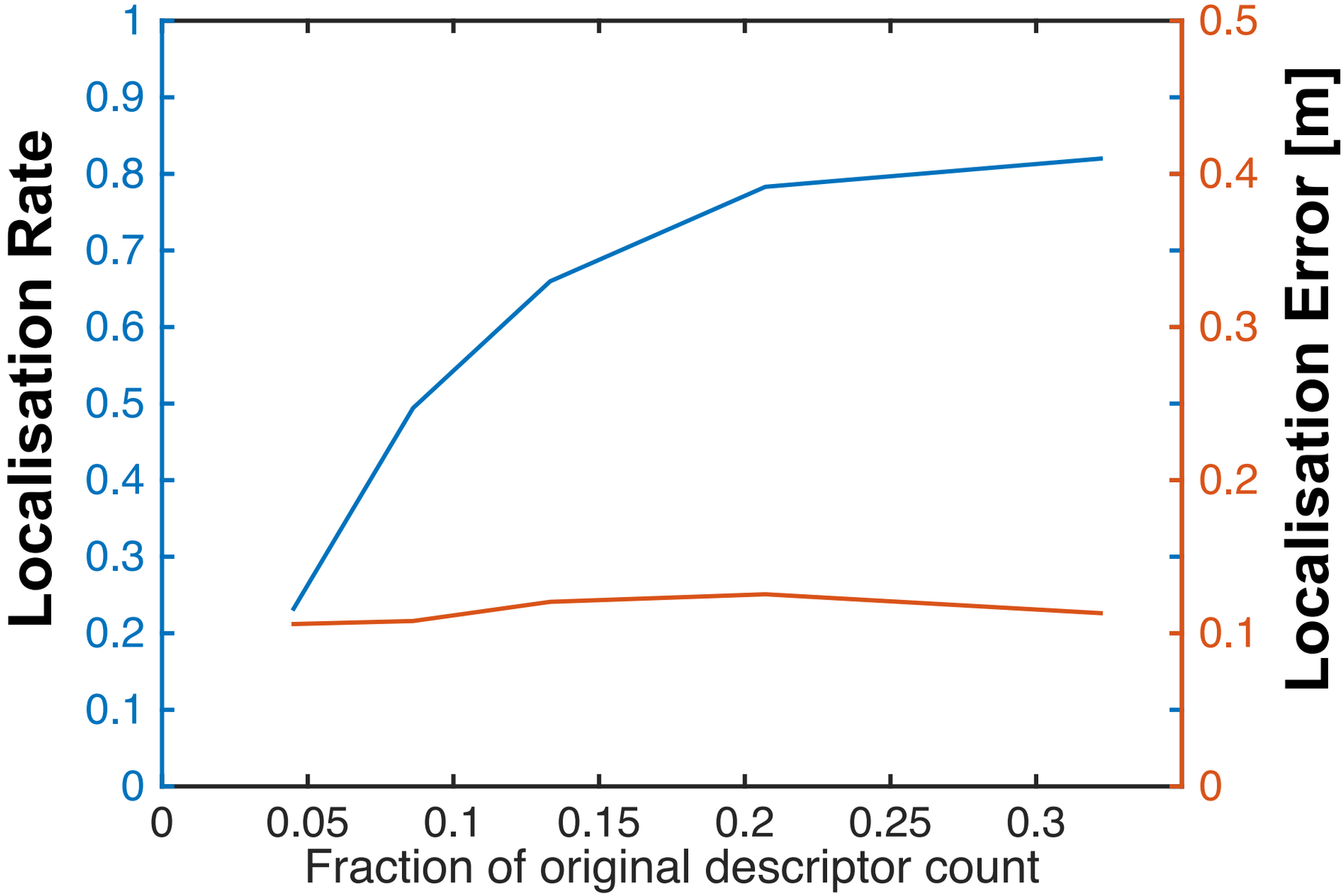
Compact Maps



- Redundancy: Not every point required for localisation
- Select subset of points s.t. every database image observes ≥N points

[Li et al., ECCV'10] [Cao & Snavely, CVPR'14]

Compact Maps



S. Lynen, T. Sattler, M. Bosse, J. Hesch, M. Pollefeys, R. Siegwart,
Get Out of My Lab: Large-scale, Real-Time Visual-Inertial Localization. RSS 2015

Compact Maps

Video credit: Simon Lynen

S. Lynen, T. Sattler, M. Bosse, J. Hesch, M. Pollefeys, R. Siegwart,
Get Out of My Lab: Large-scale, Real-Time Visual-Inertial Localization. RSS 2015

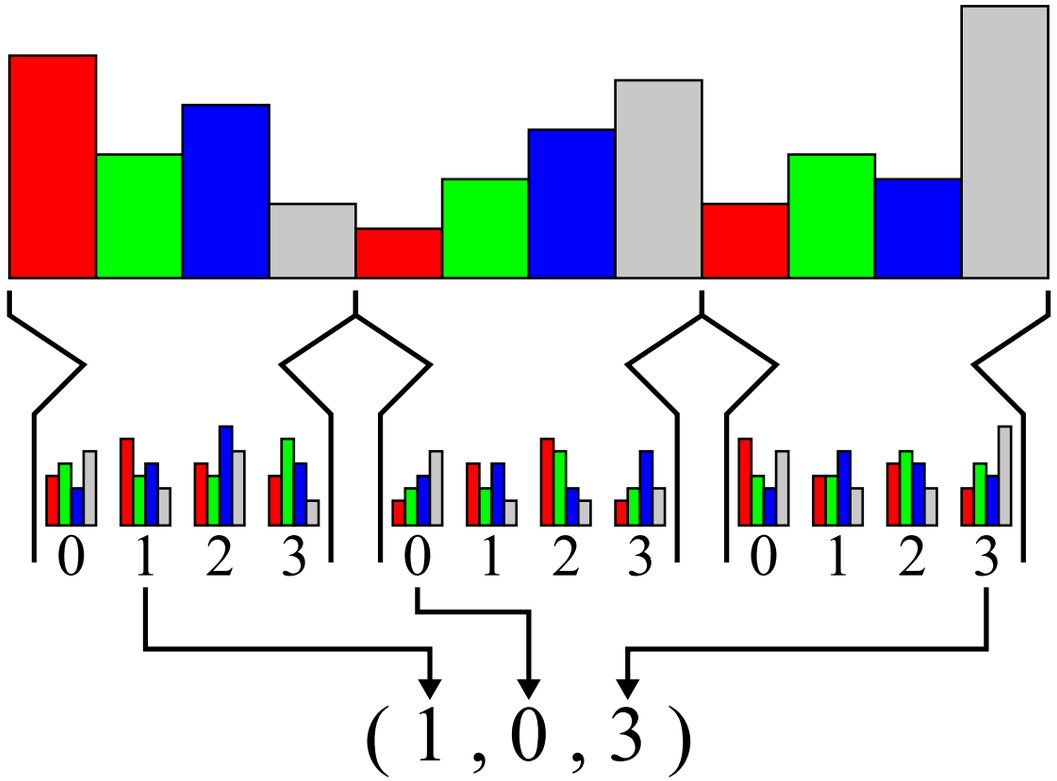
Compression by Quantisation

3D position 12 bytes	N descriptors $40 \cdot N$ bytes	N image IDs $4 \cdot N$ bytes	Add. data

3D point reconstructed from N database images

- Need certain number of points to enable localisation
- Further compression by descriptor quantisation

Product Quantisation



Original descriptor

Vocabulary per component

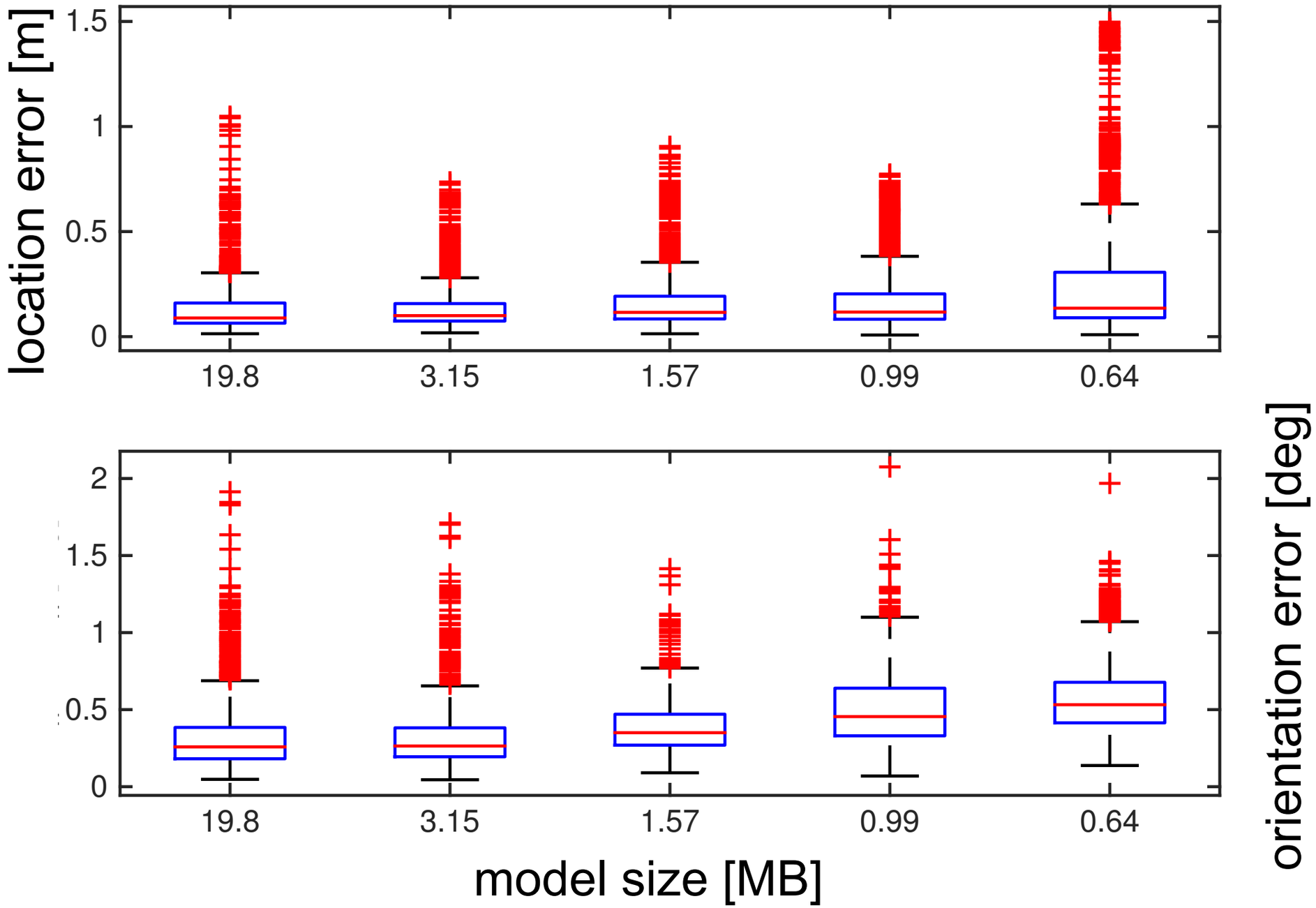
Quantised descriptor

Example: 8 components with 256 words each

- Quantised descriptor requires 8 bytes
- $256^8 = 2^{64}$ product words
- Note: Many words not “meaningful”

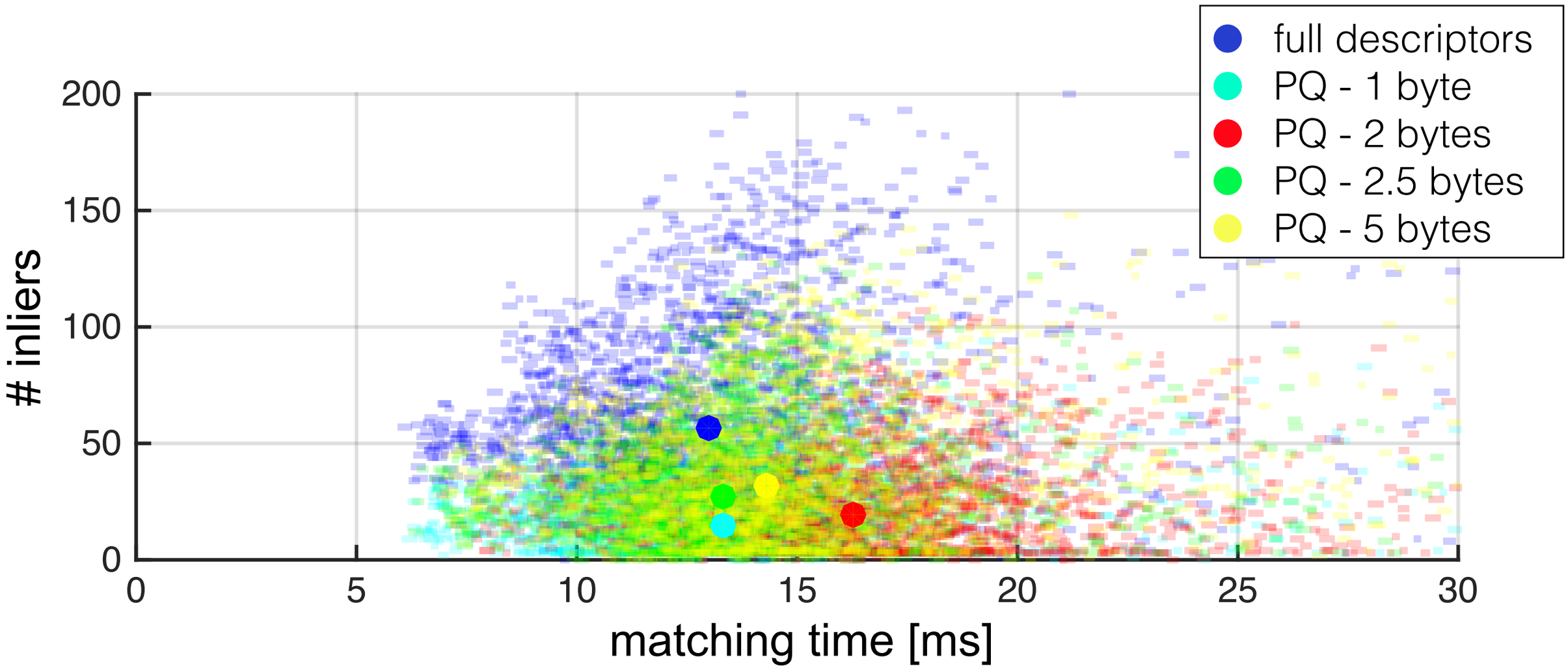
[Jégou et al., PAMI’11]

Product Quantisation



S. Lynen, T. Sattler, M. Bosse, J. Hesch, M. Pollefeys, R. Siegwart,
Get Out of My Lab: Large-scale, Real-Time Visual-Inertial Localization. RSS 2015

The Price of Quantisation

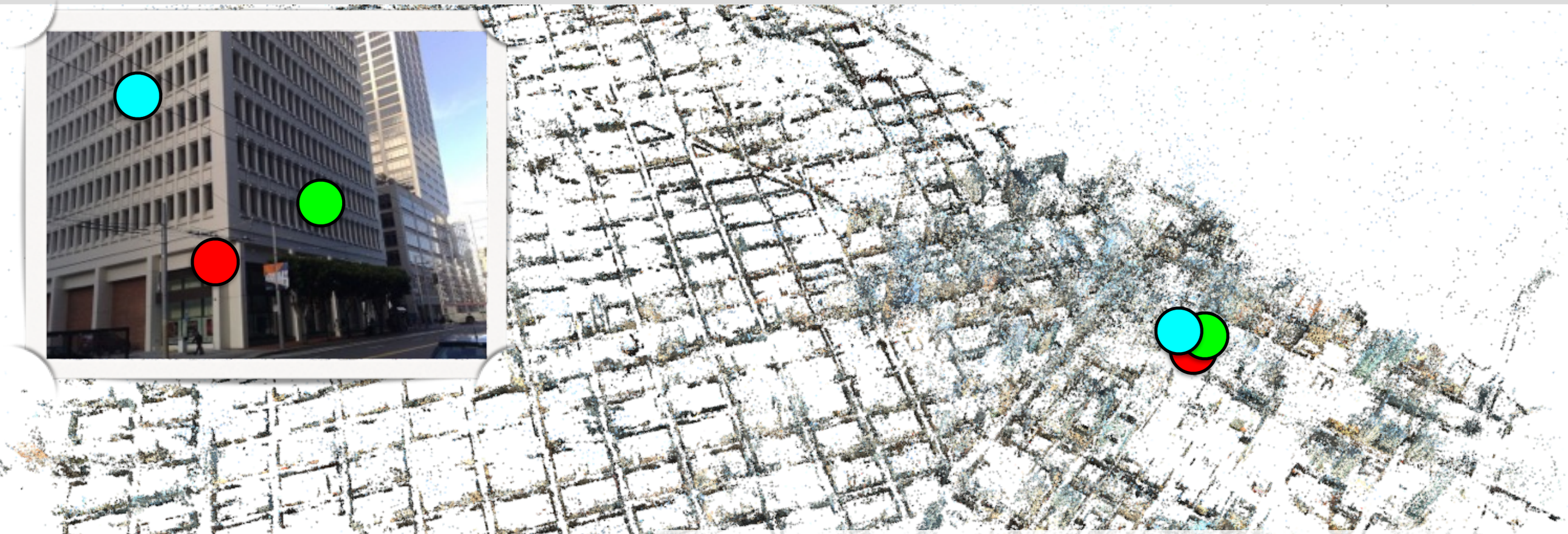


S. Lynen, T. Sattler, M. Bosse, J. Hesch, M. Pollefeys, R. Siegwart,
Get Out of My Lab: Large-scale, Real-Time Visual-Inertial Localization. RSS 2015

Lessons Learned

- Significant compression by descriptor quantisation
  - Storing descriptors not a bottleneck anymore
- Price of compression:
  - Fewer inliers
  - Reduced localisation rate
- Trade-off feasible for robust SLAM / VIO
  - Also applicable for large-scale, one-shot localisation?

Working on a Large Scale

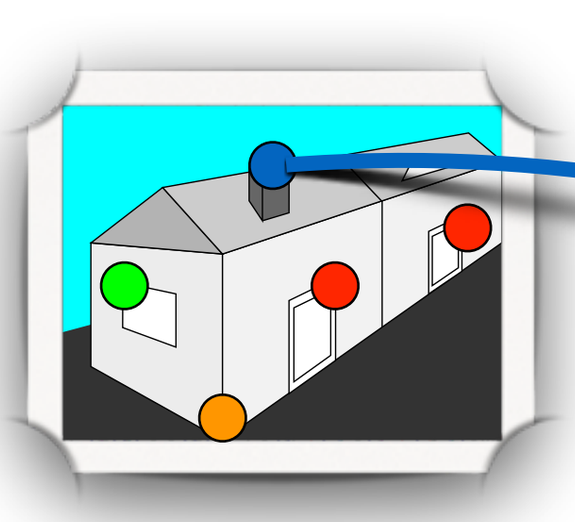


San Francisco (SF-0 model)

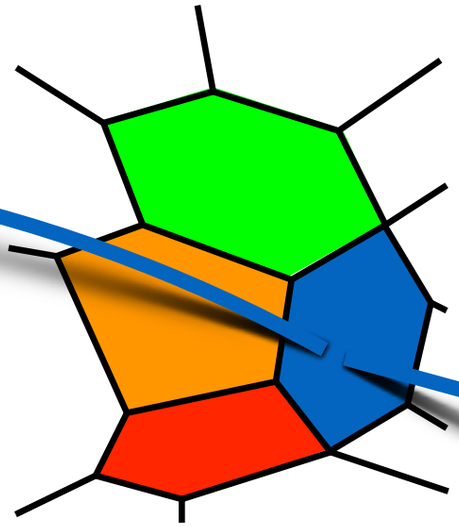
- 30M 3D points
- 149M SIFT descriptors (~17.7GB)
- 611k database images
- Task: Landmark recognition

Y. Li, N. Snavely, D. Huttenlocher, P. Fua, Worldwide Pose Estimation using 3D Point Clouds. ECCV 2012

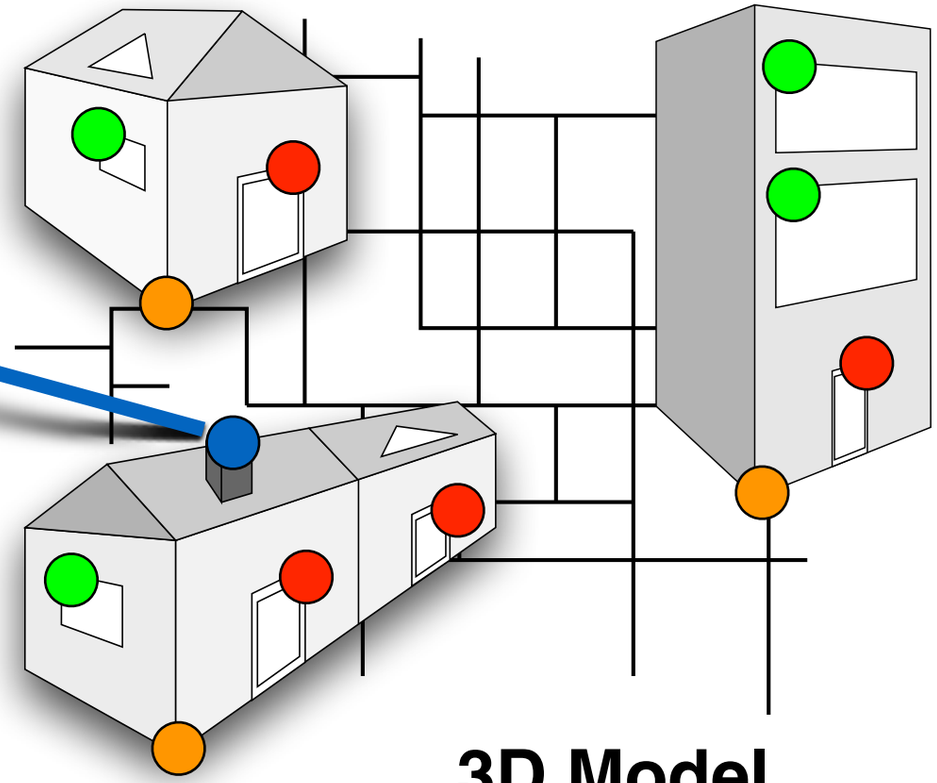
Matching with Hyperpoints



Query Image

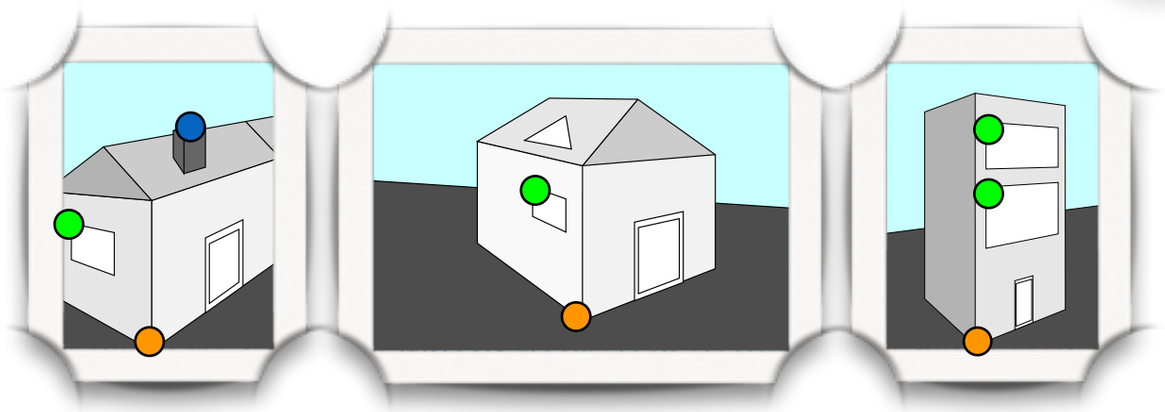


16M words



3D Model

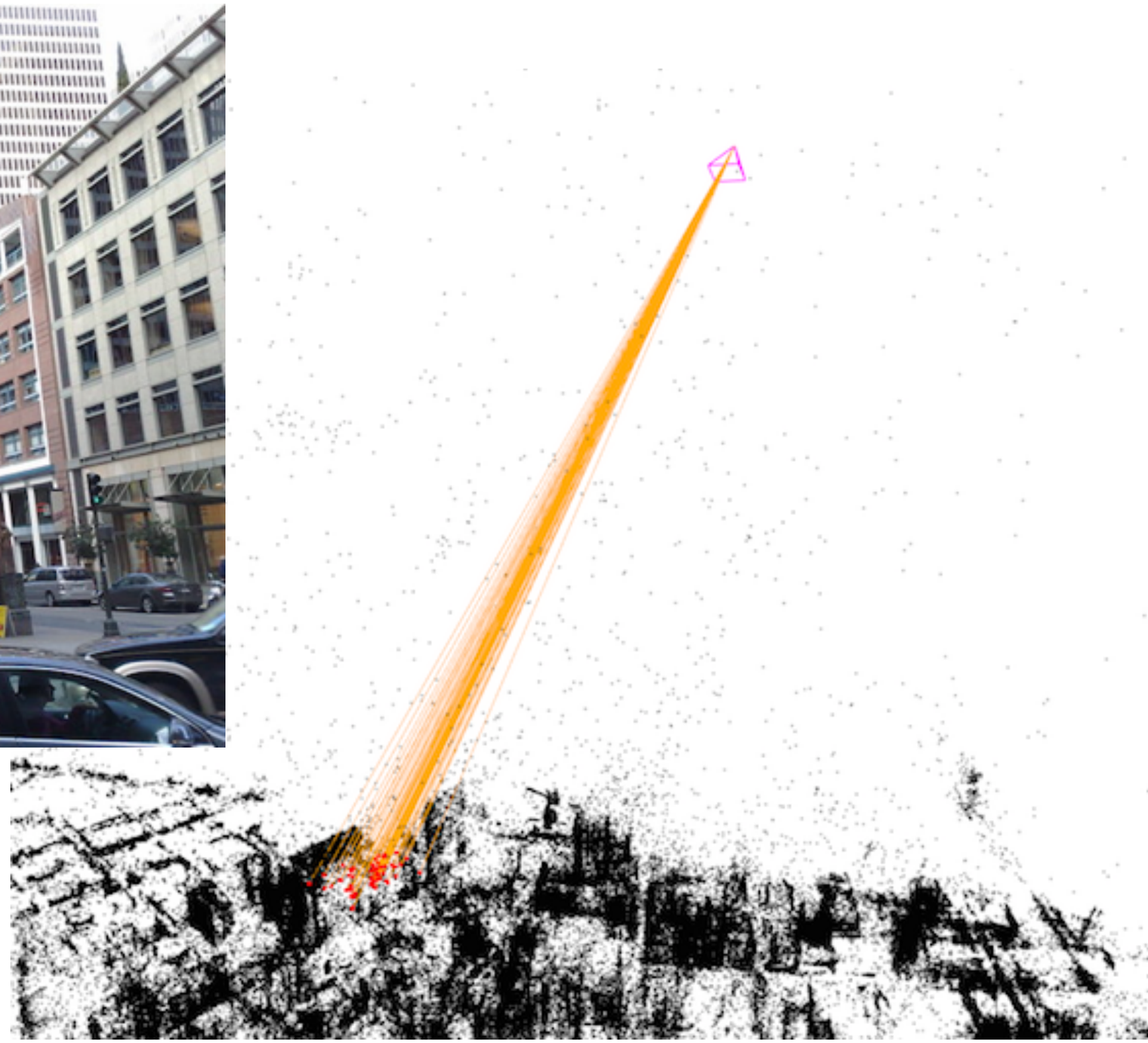
Database Images



T. Sattler, M. Havlena, F. Radenovic, K. Schindler, M. Pollefeys,
Hyperpoints and Fine Vocabularies for Large-Scale Location Recognition. ICCV 2015

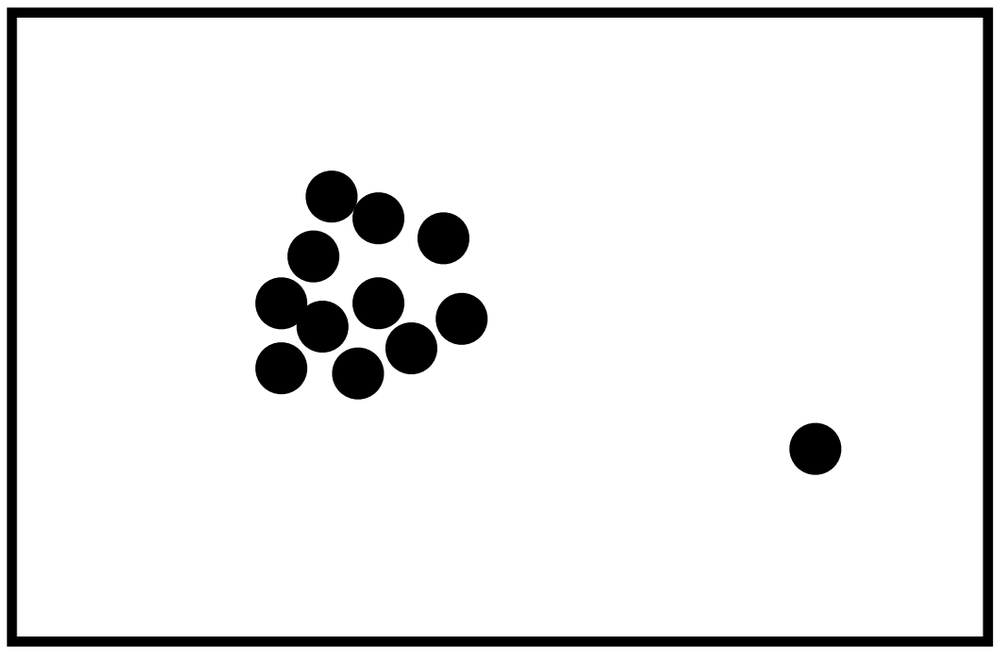
Large-Scale Quantised Localisation

30 inliers

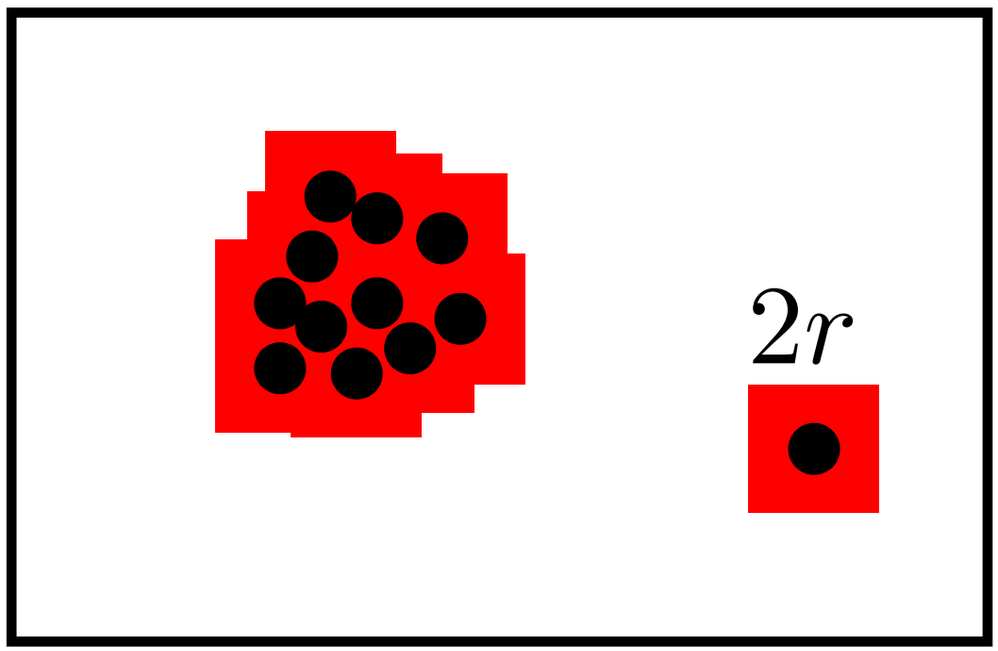


T. Sattler, M. Havlena, F. Radenovic, K. Schindler, M. Pollefeys,
Hyperpoints and Fine Vocabularies for Large-Scale Location Recognition. ICCV 2015

Large-Scale Quantised Localisation



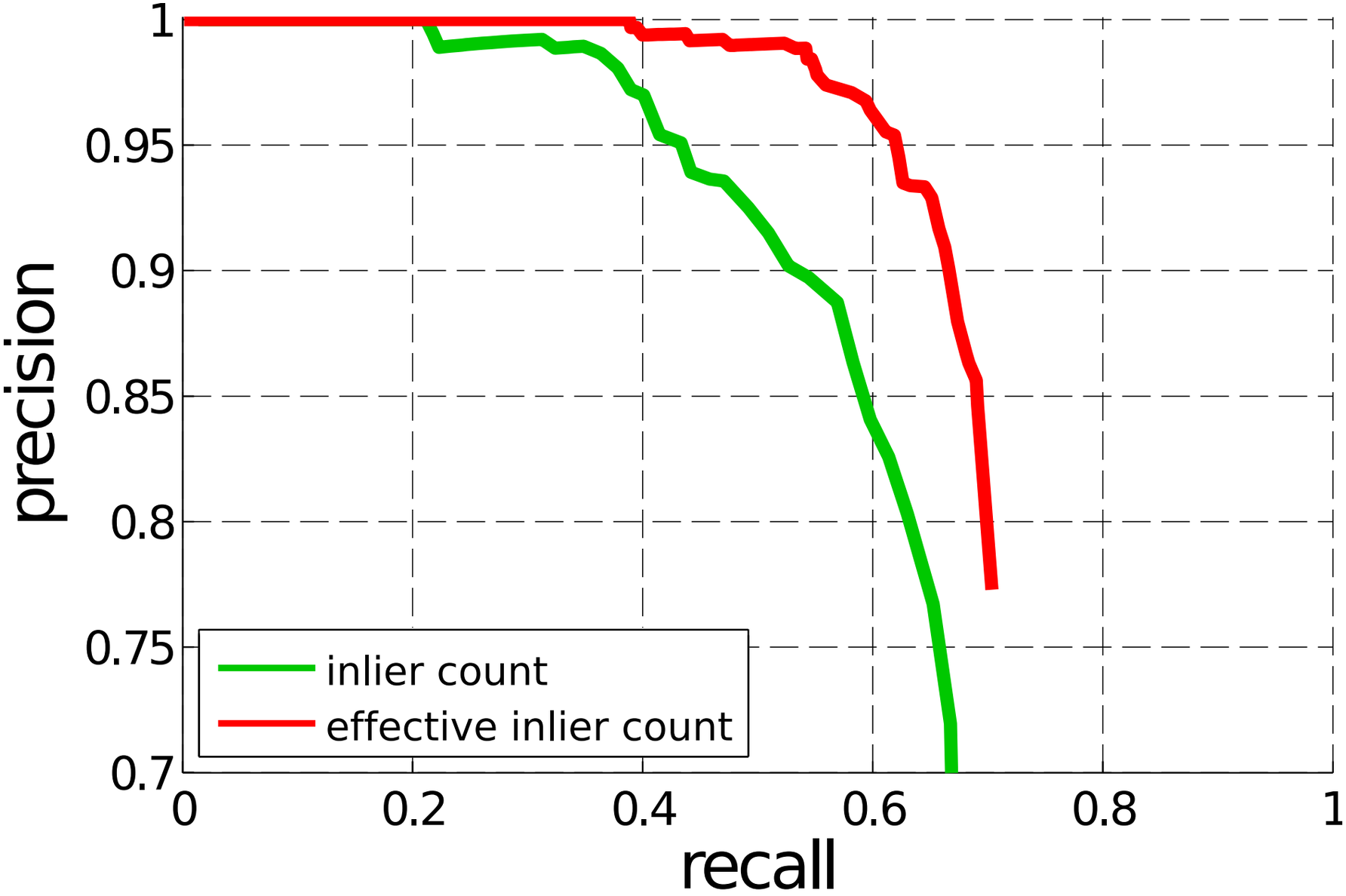
N inliers



$\text{area}(\blacksquare)/(4r^2 \cdot N) \cdot N$ inliers

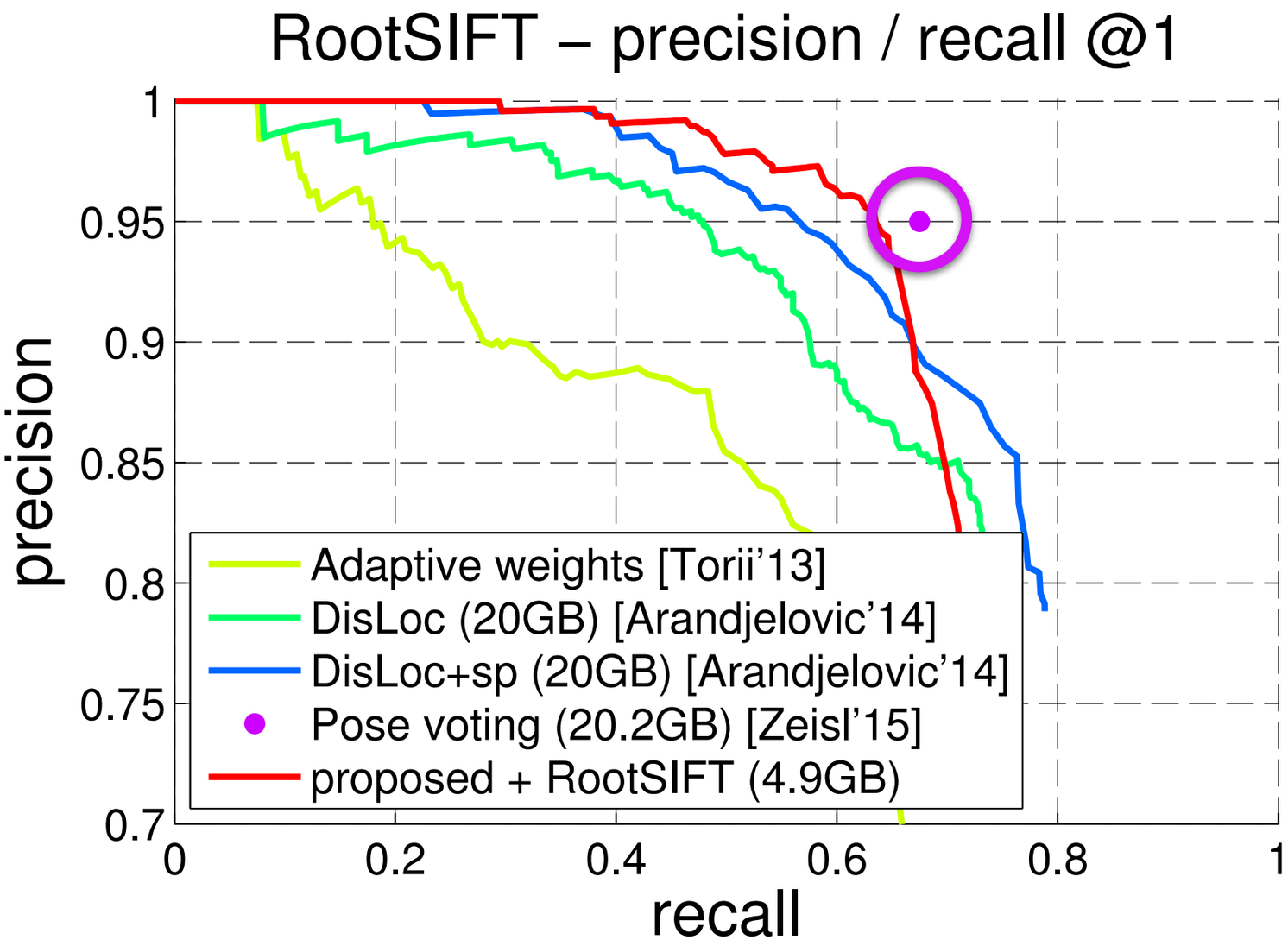
A. Irschara, C. Zach, J.-M. Frahm, H. Bischof,
From Structure-from-Motion Point Clouds to Fast Location Recognition. CVPR 2009

Large-Scale Quantised Localisation



T. Sattler, M. Havlena, F. Radenovic, K. Schindler, M. Pollefeys,
Hyperpoints and Fine Vocabularies for Large-Scale Location Recognition. ICCV 2015

Large-Scale Quantised Localisation

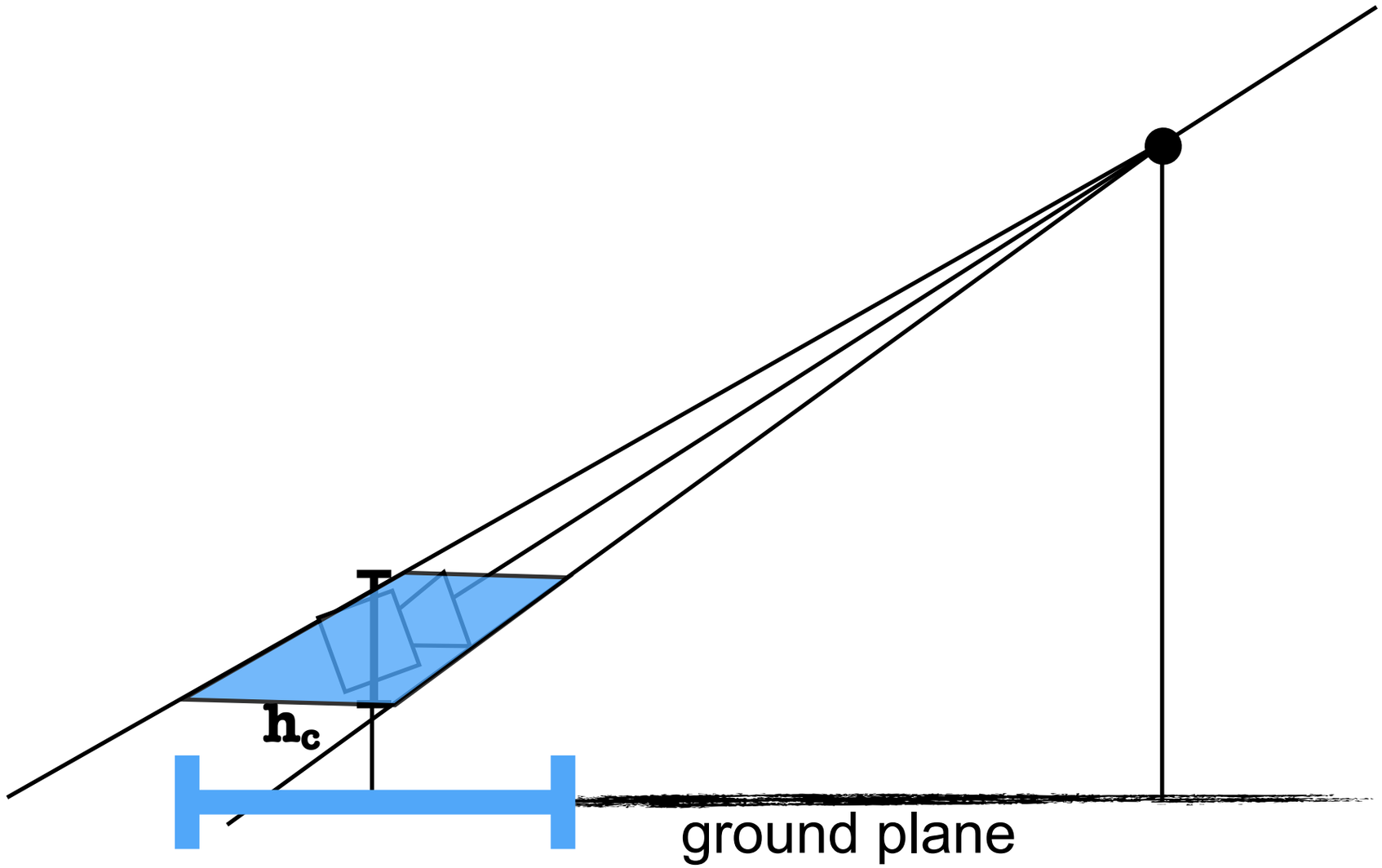


k-means vocabulary, *specifically learned* for dataset

T. Sattler, M. Havlena, F. Radenovic, K. Schindler, M. Pollefeys,
Hyperpoints and Fine Vocabularies for Large-Scale Location Recognition. ICCV 2015

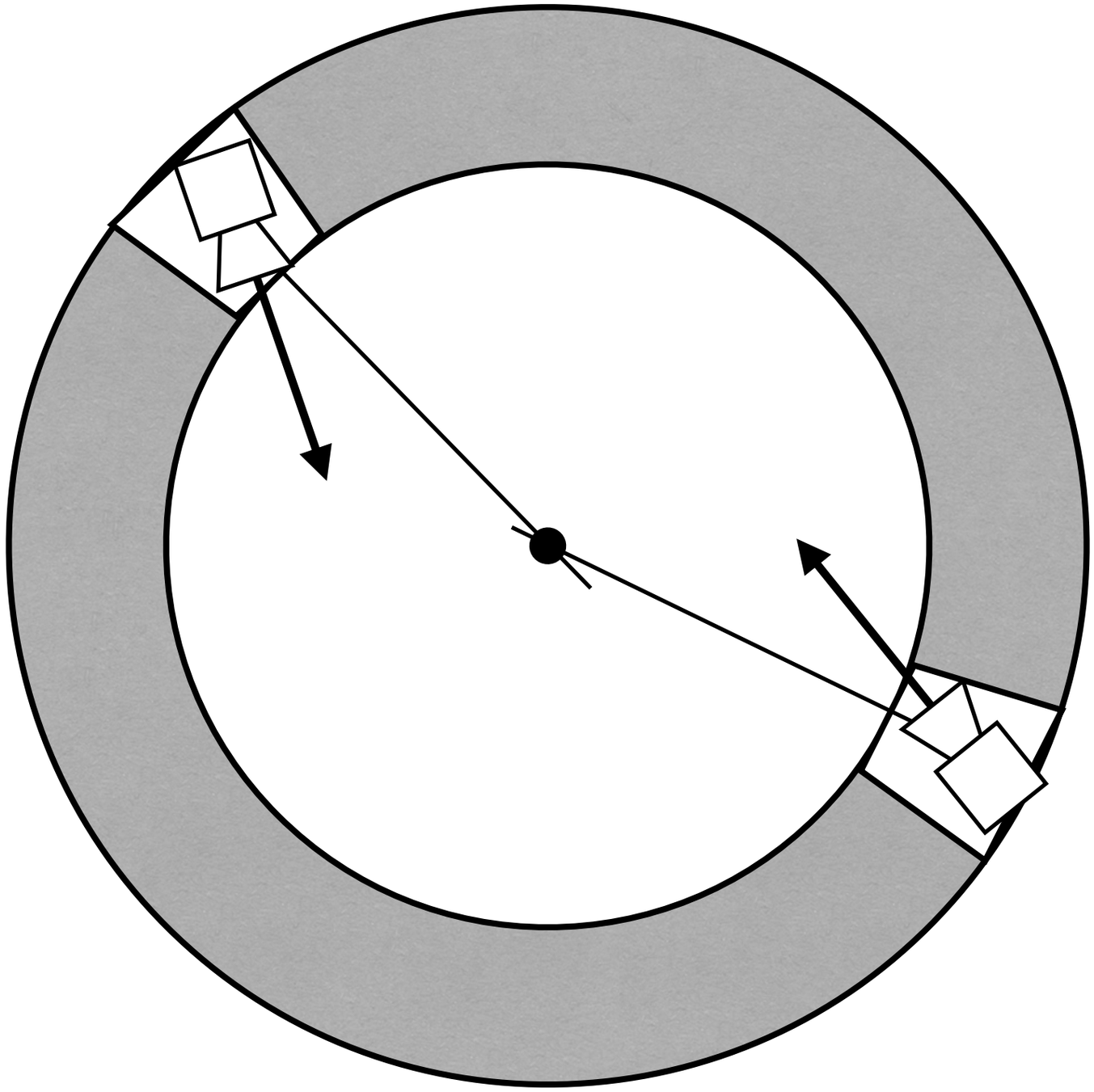
Local Registration Problem

Known: Intrinsics, gravity direction



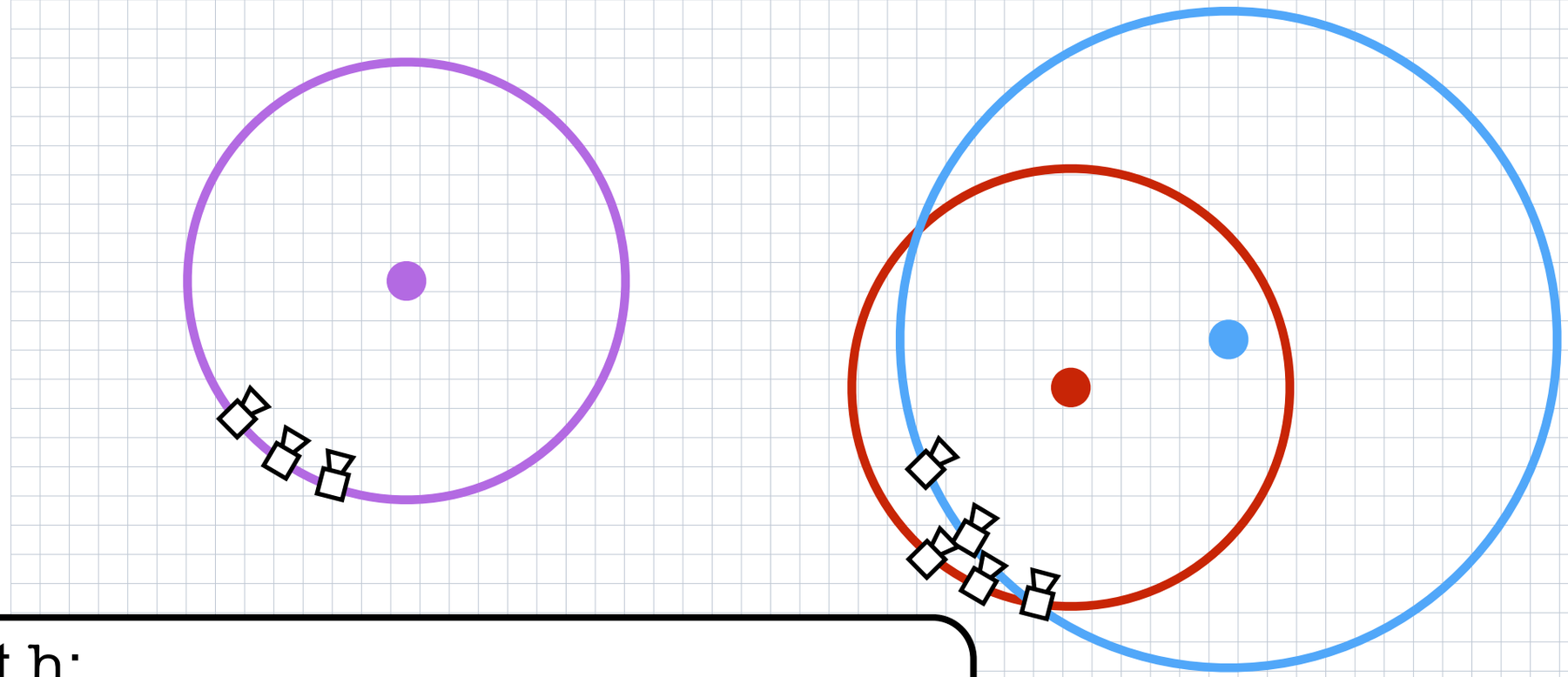
L. Svärm, O. Enqvist, M. Oskarsson, F. Kahl, Accurate Localization and Pose Estimation for Large 3D Models, CVPR'14

Global Pose Estimation



B. Zeisl, T. Sattler, M. Pollefeys, Camera Pose Voting for Large-Scale Image-Based Localization, ICCV 2015

Camera Pose Voting



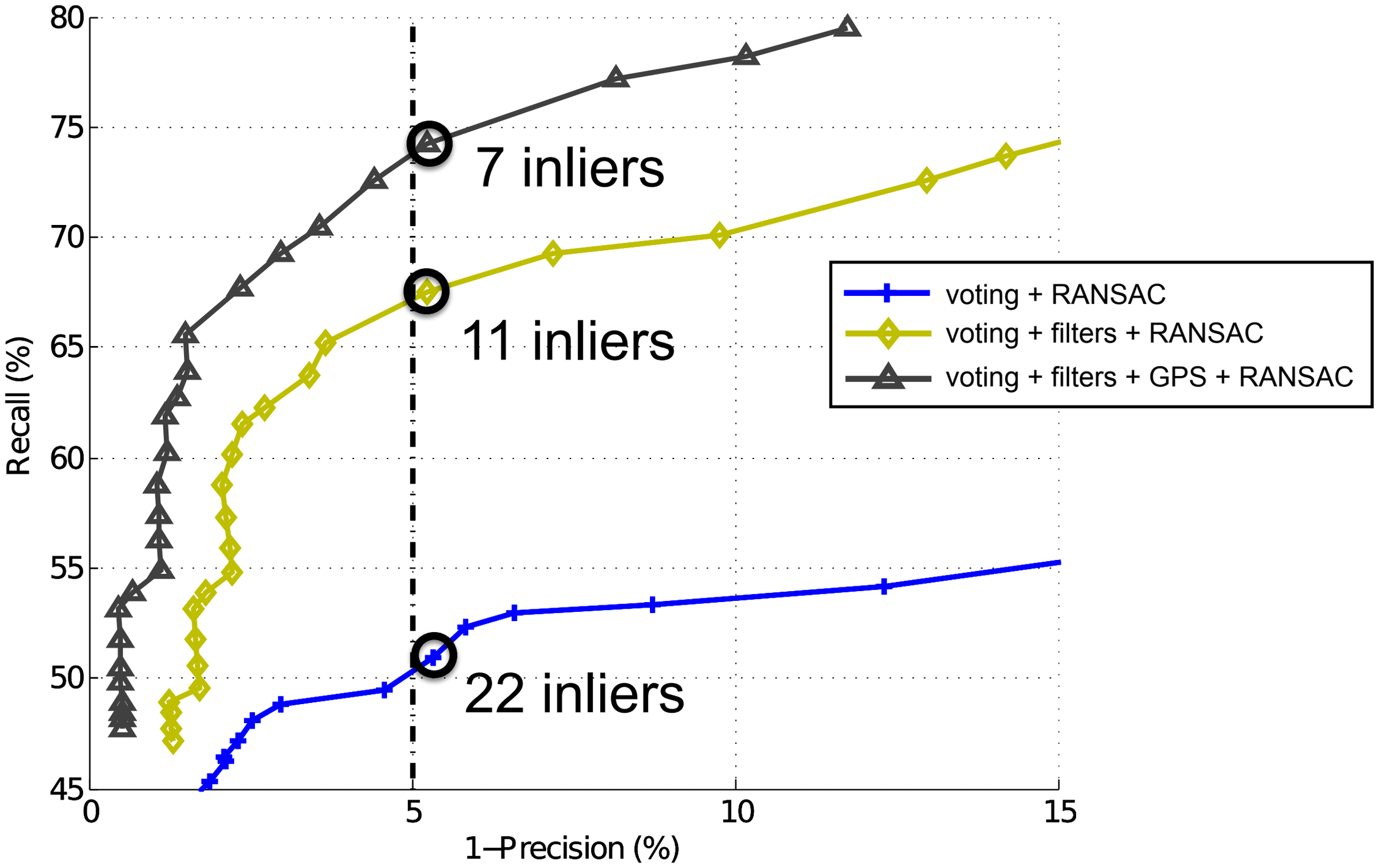
for each height `h`:
  for each viewing direction θ:
    for each 2D-3D match:
      vote for position
    find best position, update best pose
Refine pose using RANSAC

B. Zeisl, T. Sattler, M. Pollefeys, Camera Pose Voting for Large-Scale Image-Based Localization, ICCV 2015

Matching with Full Descriptors

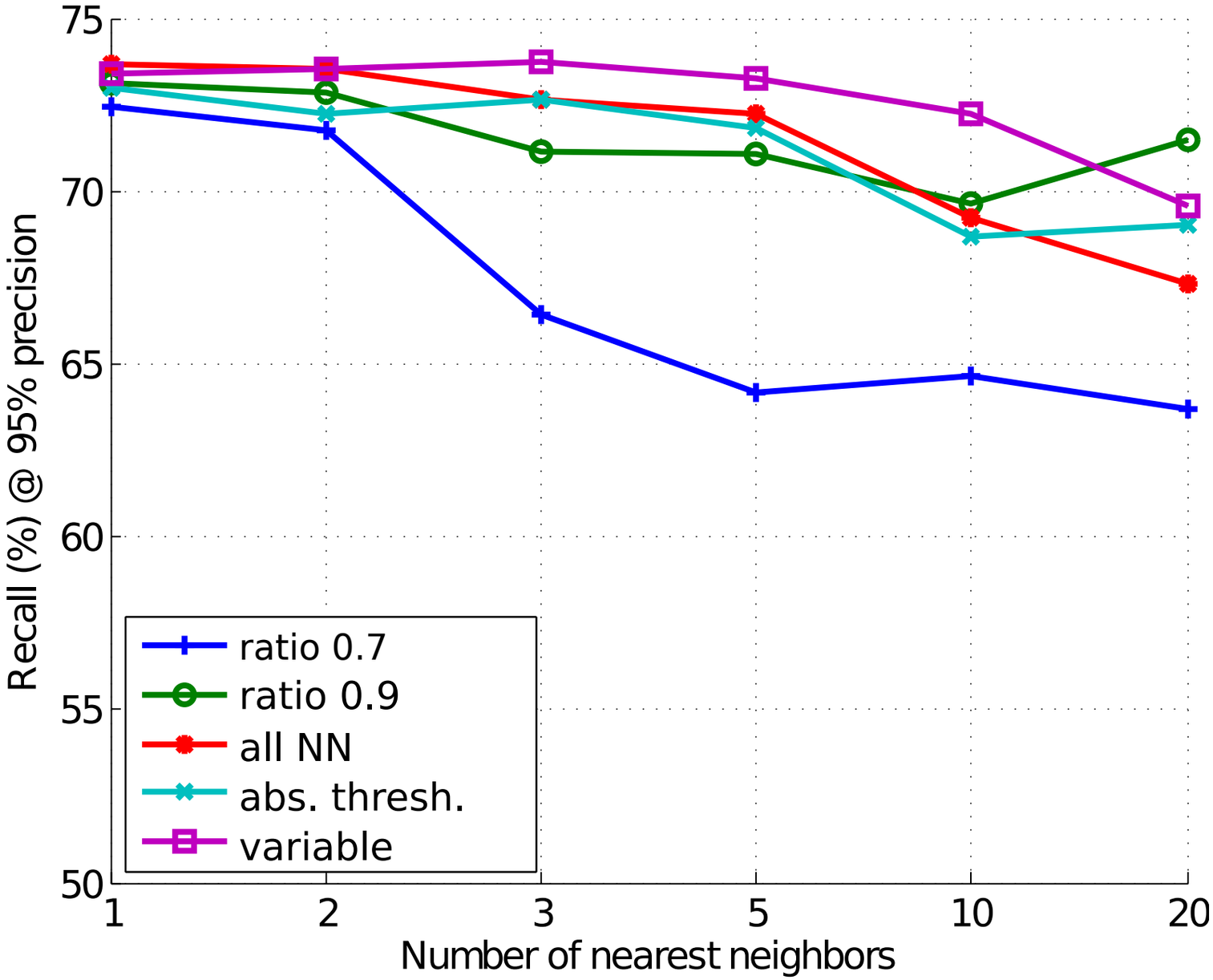
B. Zeisl, T. Sattler, M. Pollefeys, Camera Pose Voting for Large-Scale Image-Based Localization, ICCV 2015

Matching with Full Descriptors



B. Zeisl, T. Sattler, M. Pollefeys, Camera Pose Voting for Large-Scale Image-Based Localization, ICCV 2015

Matching with Full Descriptors



B. Zeisl, T. Sattler, M. Pollefeys, Camera Pose Voting for Large-Scale Image-Based Localization, ICCV 2015

Lessons Learned

- Individual feature descriptors become less discriminative at large scale
  - Exploiting feature geometry helps…
  - but does not resolve the problem

- Matching with quantised descriptors works surprisingly well
  - … but better understanding required

- #inliers not good for distinguishing between correct and wrong camera poses

The Future of (Real-Time) SLAM?

- Compact map representations?

- Better understanding on when to trust camera pose estimates
  - Semantic understanding of scenes?

- Handling more challenging scenes
  - Natural scenes (forests, …)
  - Long-term localisation and mapping
  - Nighttime localisation against daytime maps?