The Challenges of Large-Scale Localisation and Mapping

Torsten Sattler
Computer Vision and Geometry Lab
ETH Zürich
Mobile Localisation

loop closures

AR / VR

robotics / autonomous vehicles
Mobile Localisation

Real-time camera pose tracking

Keyframe

Pose Data

Global localisation against known map
Global Localisation

Structure-from-Motion / SLAM model

- 3D Point: 3D point + descriptors
- 2D Feature: 2D position + descriptor

Database Images

Query Image
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
deviation for strategy localizations has a notable impact on the localization accuracy and standard error for Sequence 1 and strategies. Table 4 reports the impact on mean position and rotation errors for randomly selected a global keyframe pose and geometrically consistent global 2D-set the number of false positive localizations to up to 40%. For every global in the global localization. Li

Robustness to False Positive Global Localizations but not the local reprojection errors, are omitted in the alignment.

the server response is pending, the global constraints belonging to this keyframe, available for the alignment. Afterwards, if the server fails to localize a keyframe or it is able to track the camera pose locally until

localization. On the other hand, our approach is robust to these problems, since latency period. Thus, it is prone to both, high server latency and failed global localization. As discussed in Section 5, strategies 5), strategies 0), and 1) respectively. The pose estimation approach by Ventura and Höllerer [29] relies on a single

False Positives

ground truth

using global 2D-3D matches

using global pose estimates

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 $\geq N$ 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
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

<table>
<thead>
<tr>
<th>3D position</th>
<th>$N$ descriptors</th>
<th>$N$ image IDs</th>
<th>Add. data</th>
</tr>
</thead>
<tbody>
<tr>
<td>12 bytes</td>
<td>$40 \cdot N$ bytes</td>
<td>$4 \cdot N$ bytes</td>
<td></td>
</tr>
</tbody>
</table>

3D point reconstructed from $N$ database images

- Need certain number of points to enable localisation
- Further compression by descriptor quantisation
Example: 8 components with 256 words each
- Quantised descriptor requires 8 bytes
- $256^8 = 2^{64}$ product words
- Note: Many words not “meaningful”
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

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
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

RootSIFT – precision / recall @1

Precision: 0.7, 0.75, 0.8, 0.85, 0.9, 0.95, 1
Recall: 0, 0.2, 0.4, 0.6, 0.8, 1

Adaptive weights [Torii’13]
DisLoc (20GB) [Arandjelovic’14]
DisLoc+sp (20GB) [Arandjelovic’14]
Pose voting (20.2GB) [Zeisl’15]
proposed + RootSIFT (4.9GB)

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
for each height $h$:
  for each viewing direction $\theta$:
    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?