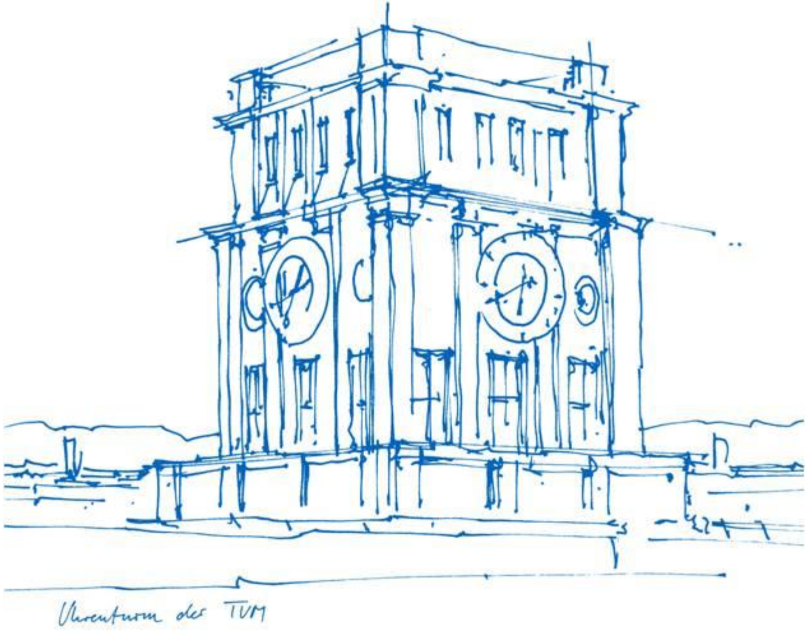
# Monocular Camera Localization in 3D LiDAR Maps

Speaker: Yunxiang Lu and Keyue Zhang

Supervisor: Simon Klenk

Vision-based Navigation

Garching b. München, 24. July 2023



# Introduction

Goal: Localize the monocular camera in a 3D LiDAR Map

- Input: Image stream + 3D point cloud map
- Output: Estimated camera pose trajectory

Method: Align the reconstructed point to 3D LiDAR Map

- Alignment by matching geometry
- Eliminate the accumulated drift

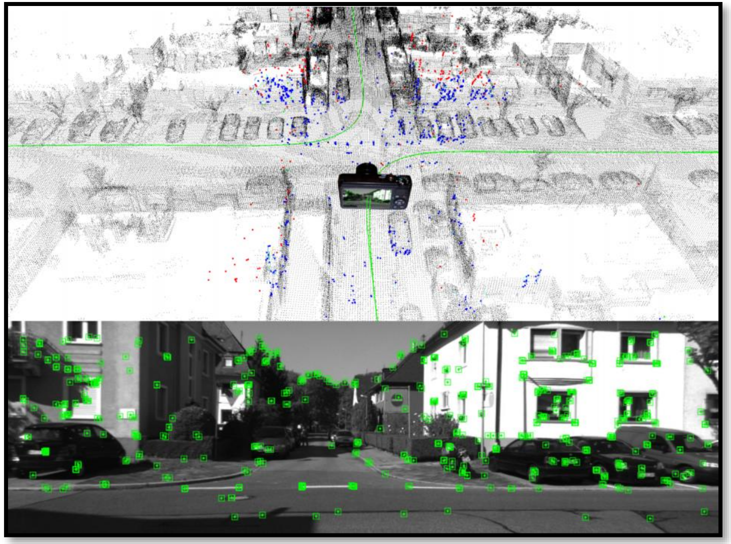
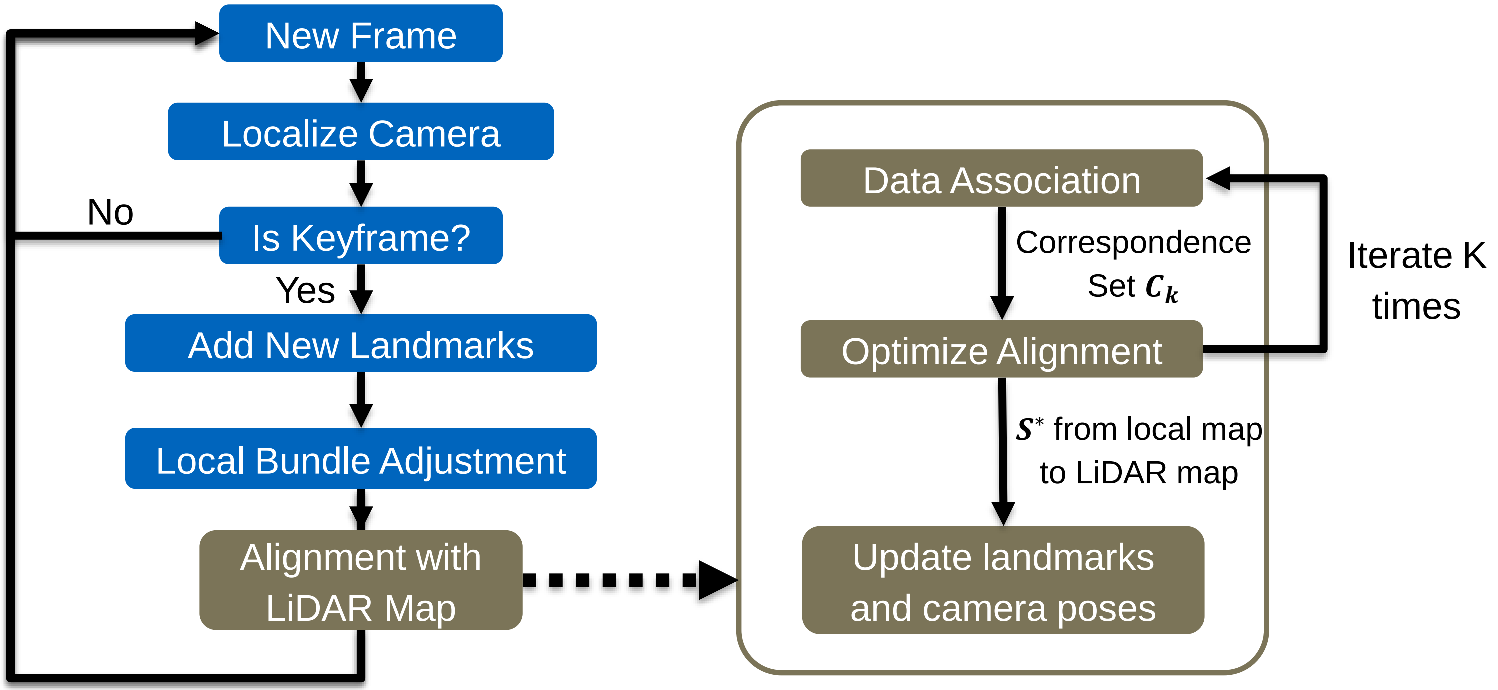


Figure 1: Caselitz, et al. “Monocular Camera Localization in 3D LiDAR Maps.”

# Visual Odometry



# Alignment-Data Association

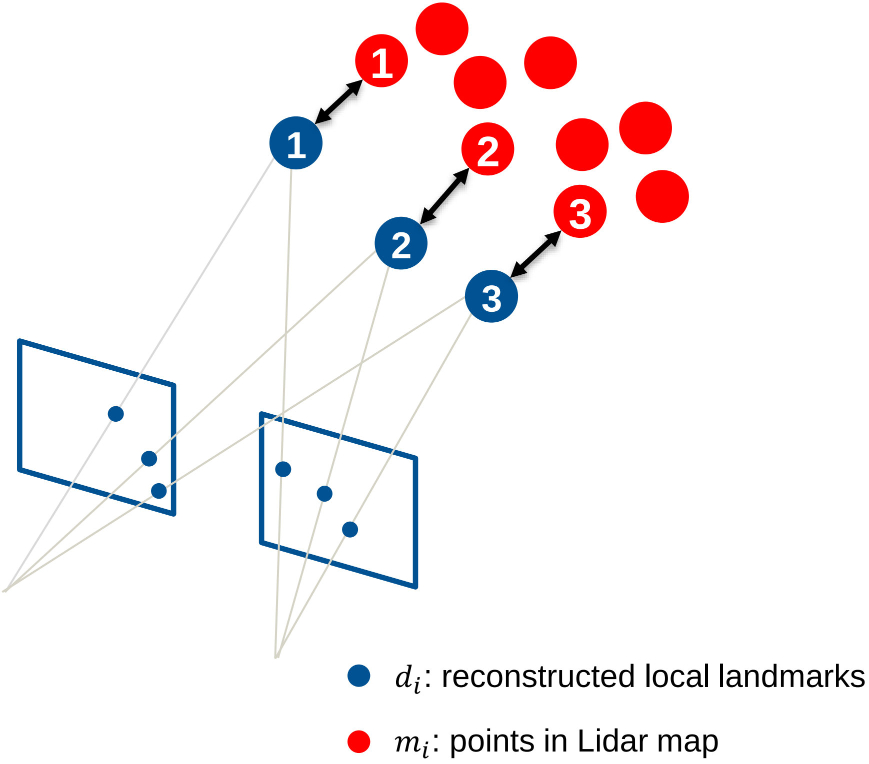
- Find correspondences between **reconstructed points** and **points in Lidar map** iteratively

$$\mathcal{C}_k = \left\{(\mathbf{d}_i, \mathbf{m}_j)\ \forall \mathbf{d}_i \in \mathcal{D} \mid \exists \operatorname*{argmin}_{\mathbf{m}_j \in \mathcal{M}} \|\mathbf{d}_i - \mathbf{m}_j\|_2 < \tau_k \right\}.$$

Local landmark

LiDAR point

Distance threshold



# Alignment-Data Association

- Find correspondences between **reconstructed points** and **points in Lidar map** iteratively

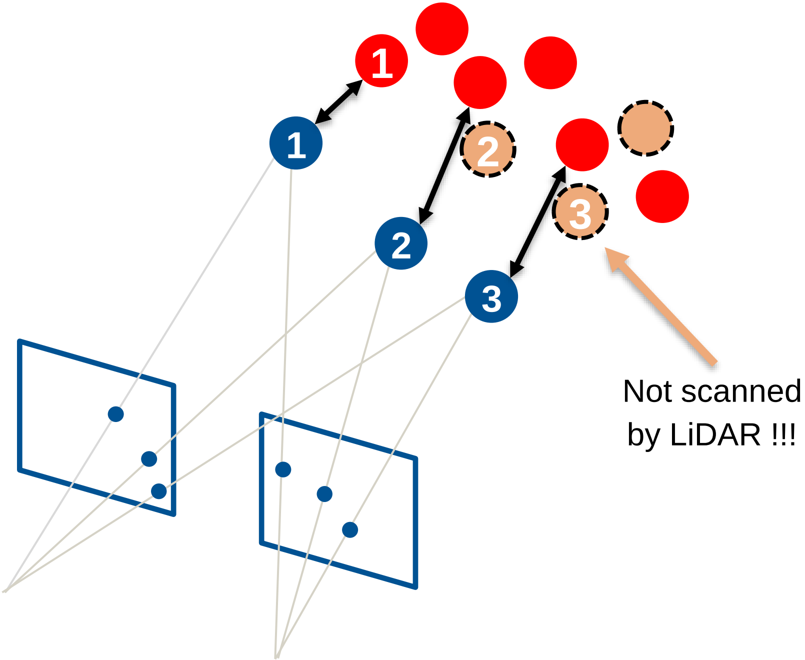
$$\mathcal{C}_k = \{(\mathbf{d}_i, \mathbf{m}_j)\ \forall \mathbf{d}_i \in \mathcal{D} \mid \exists \operatorname*{argmin}_{\mathbf{m}_j \in \mathcal{M}} \|\mathbf{d}_i - \mathbf{m}_j\|_2 < \tau_k\}.$$

Local landmark

LiDAR point

Distance threshold

**Drawback:**
the set of reconstructed points overlaps ***only partially*** with the LiDAR map



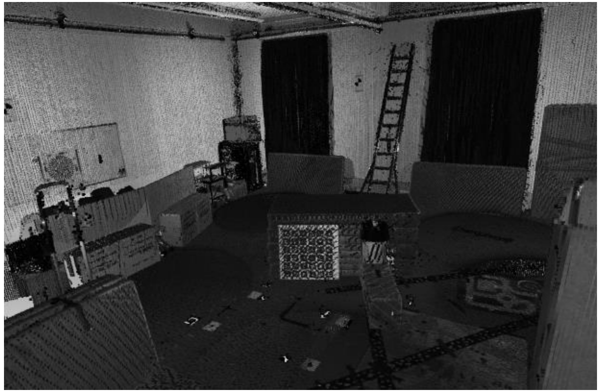

$d_i$: reconstructed local landmarks

$m_i$: points in Lidar map

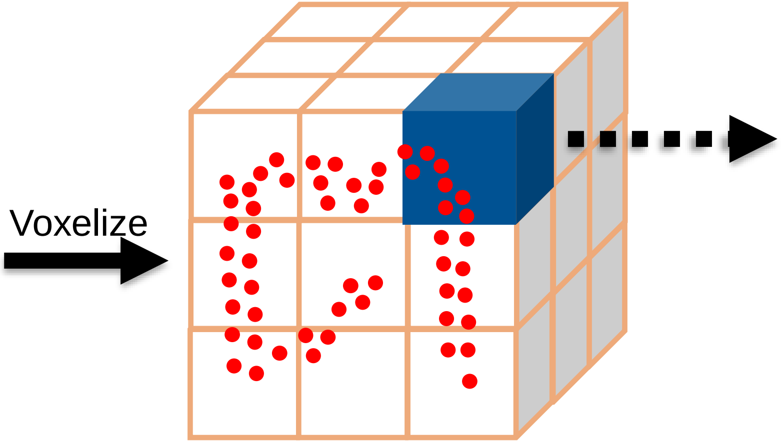
# Alignment-Local Point Distribution

Preprocessing

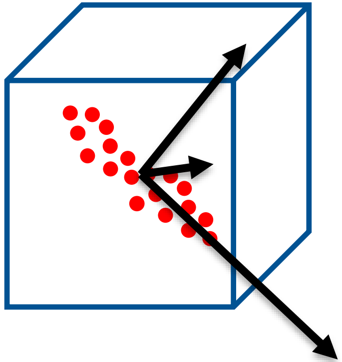
- Voxelize the point cloud map
- Use PCA to determine the local point distribution in each voxel



3D Point Cloud Map




Voxelized Map



One Voxel in the Map

# Alignment-Filter out bad correspondence

Good Conditions:

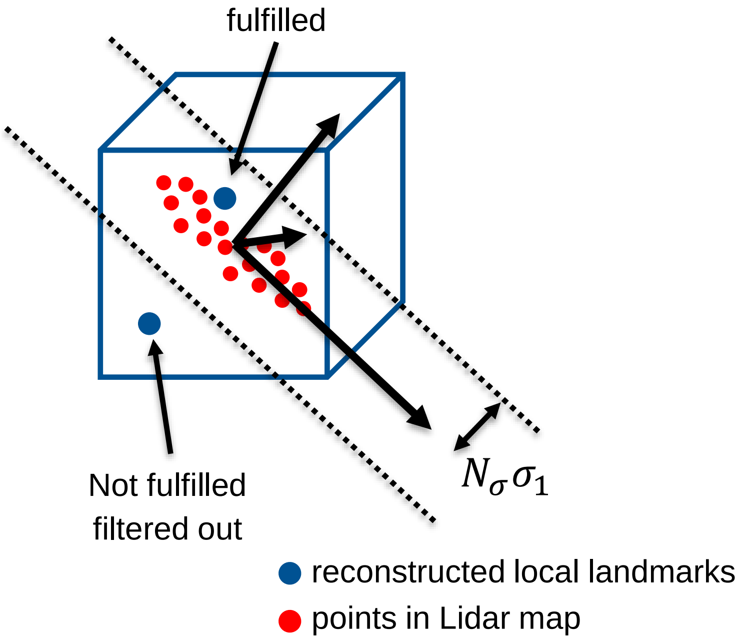
- The amount of LiDAR points in a voxel is sufficient

$$N \geq N_{min}$$

- The reconstructed local landmark lies inside a multiple standard deviation along the voxel's principle component axes

$$Td_i \leq N_{\sigma}\sigma$$

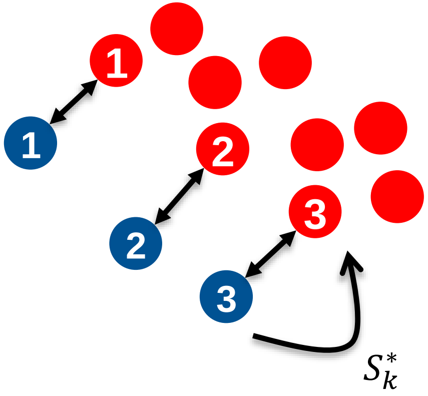
- Or any neighboring voxel fulfills above criteria



# Alignment-Optimization

Given a set of correspondences $C'_k = \{(d_1, m_1), (d_2, m_2), \dots \dots\}$

Estimate similarity transformation $S_k^*$ from local reconstruction to LiDAR map




Perform with an ICP scheme

- Update the correspondence set $C'_k$ based on current $S_k^*$ over $K$ iterations.
- Reduce the distance threshold $\tau_k$ over K iterations.

$$\tau_k = -\frac{\tau_{max} - \tau_{min}}{K} k + \tau_{max}.$$

- Error function is squared Euclidean distance between corresponding points

$$\mathbf{e}_{Data}(\mathbf{S}, \mathbf{d}_i, \mathbf{m}_j) \quad = \quad \xi(\mathbf{S}\tilde{\mathbf{d}}_i) - \mathbf{m}_j.$$

# Alignment-Optimization

ICP scheme

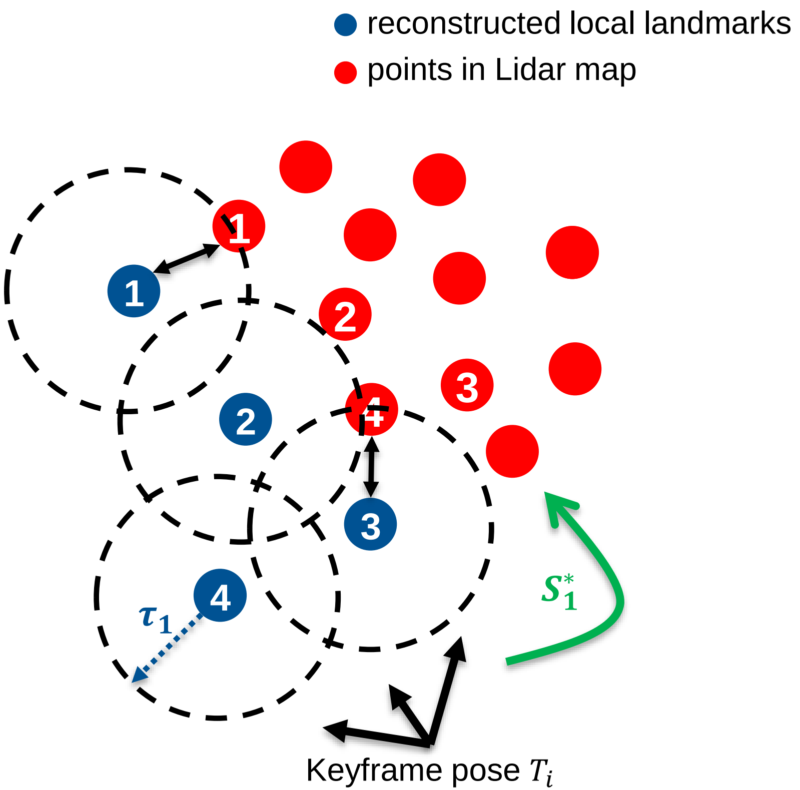
k=1

Data Association:

$C'_k = \{(d_1, m_1), (d_3, m_4)\}$

Optimization:

Estimate $S_1^*$



# Alignment-Optimization

● reconstructed local landmarks
● points in Lidar map

ICP scheme

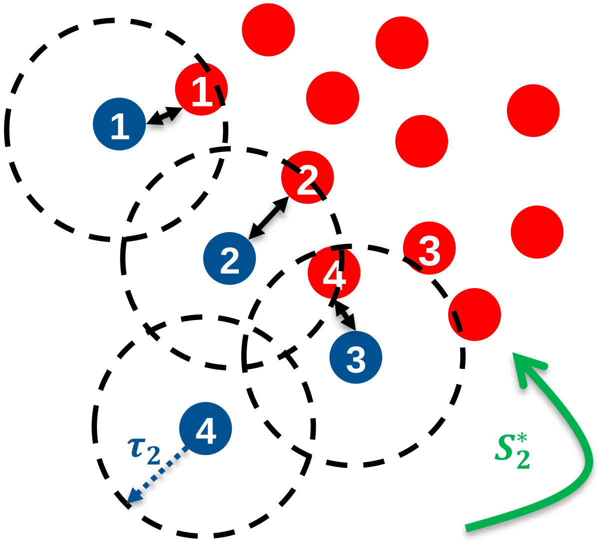
k=2

Data Association:

$C'_k = \{(d_1, m_1), (d_2, m_2), (d_3, m_4)\}$

Optimization:

Estimate $\boldsymbol{S}_2^*$



# Alignment-Optimization

ICP scheme

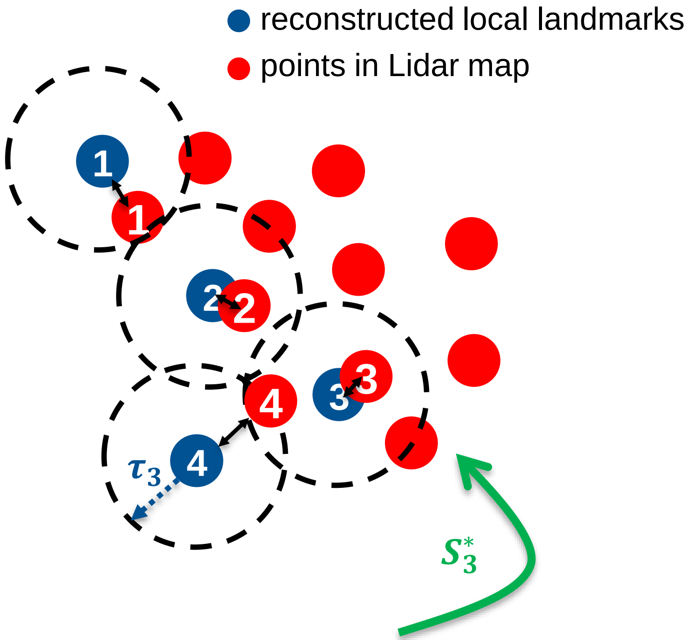
k=3

Data Association:

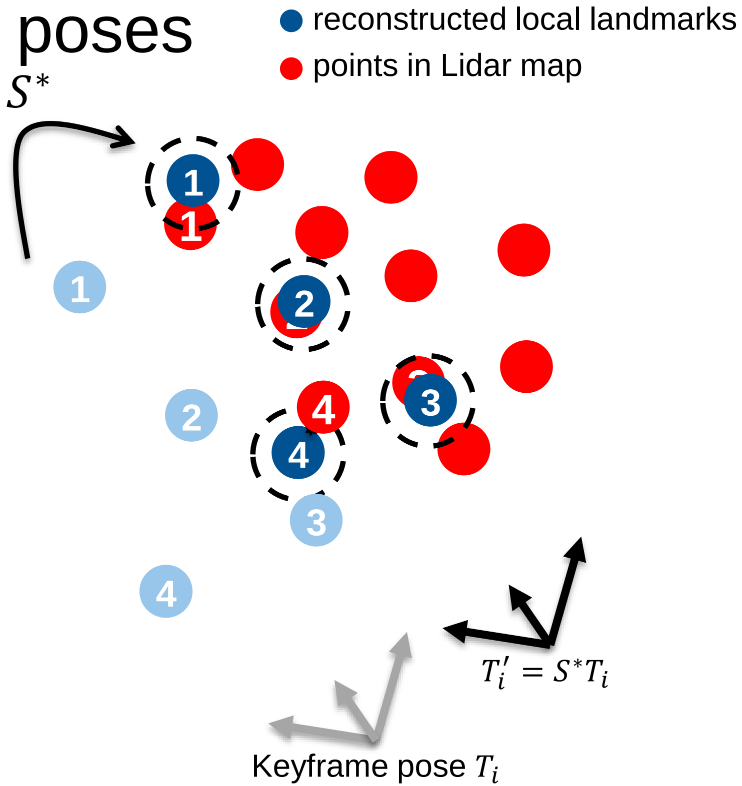
$C'_k = \{(d_1, m_1), (d_2, m_2), (d_3, m_3), (d_4, m_4)\}$

Optimization:

Estimate $S_3^*$



# Alignment-Update landmarks and poses




After K iterations

$C'_k = \{(d_1, m_1), (d_2, m_2), (d_3, m_3), (d_4, m_4)\}$

From origin landmarks to optimized landmarks:

$$S^* = \prod_{k=0}^{K-1} S^*_{K-k}$$

Transform all point positions $d_i$ and keyframe poses $T_i$

$D' = \{d_i' = S^* d_i \;, \forall d_i \in D\}$

$T' = \{T_i' = S^* T_i \;, \forall T_i \in T\}$

# Result-Stereo Camera for Euroc V1_01_easy

Stereo without alignment
1 min 23 s (~30Hz)

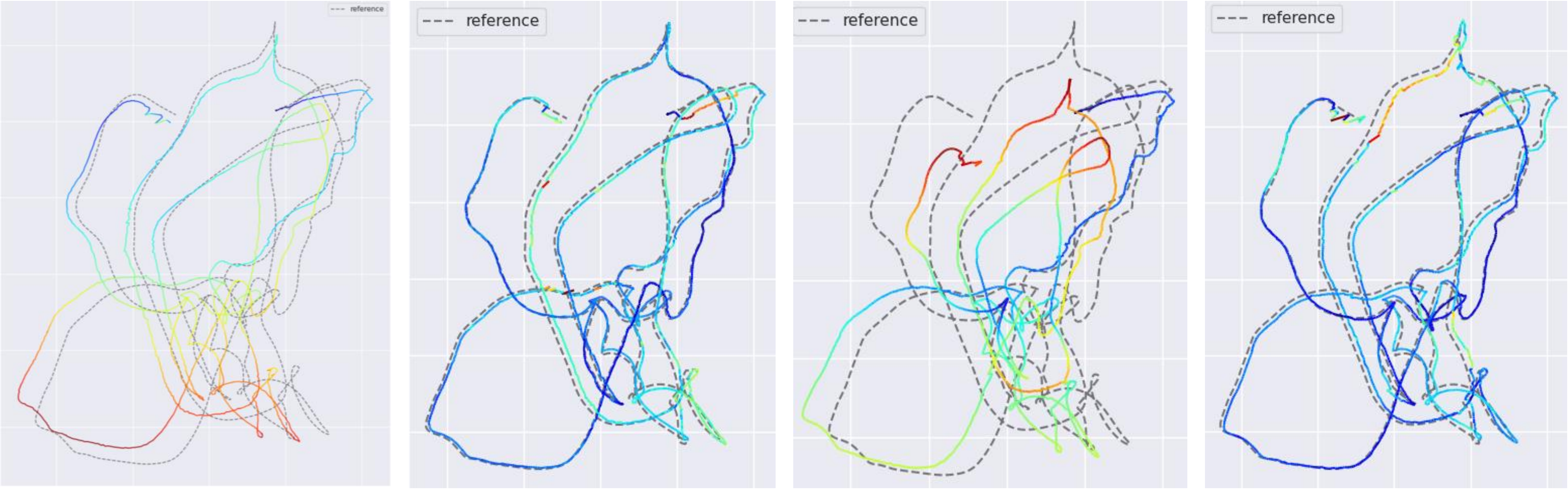
Stereo with alignment
3 min 53 s (~12Hz)

# Result-Monocular Camera for Euroc V1_01_easy

Monocular without alignment
1 min 09 s (~40Hz)

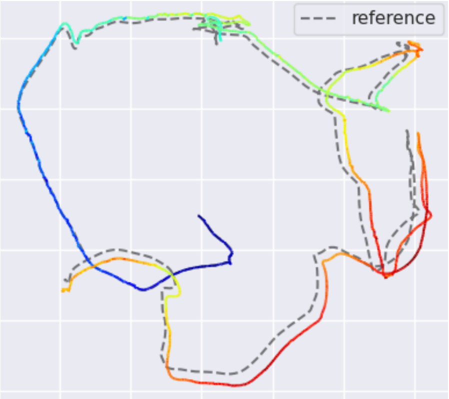
Monocular with alignment
2 min 57 s (~15Hz)

# Result for V1_01_easy



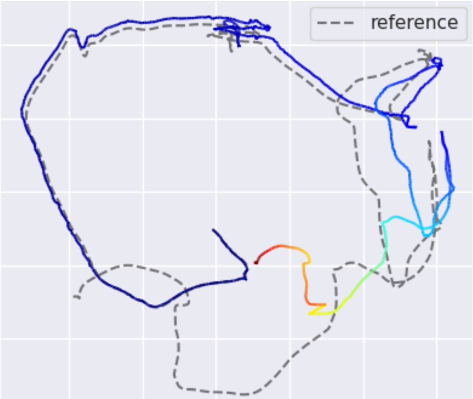

Stereo | Stereo + LiDAR map | Monocular | Monocular + LiDAR map

# Result for V2_01_easy

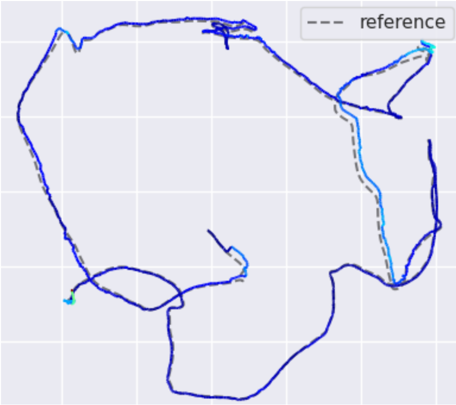



Stereo


Stereo + LiDAR map




Monocular




Monocular + LiDAR map

# Other good results

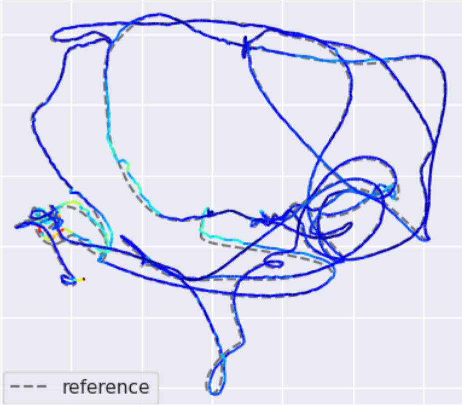
## V1_02_medium


Monocular


Monocular + LiDAR map

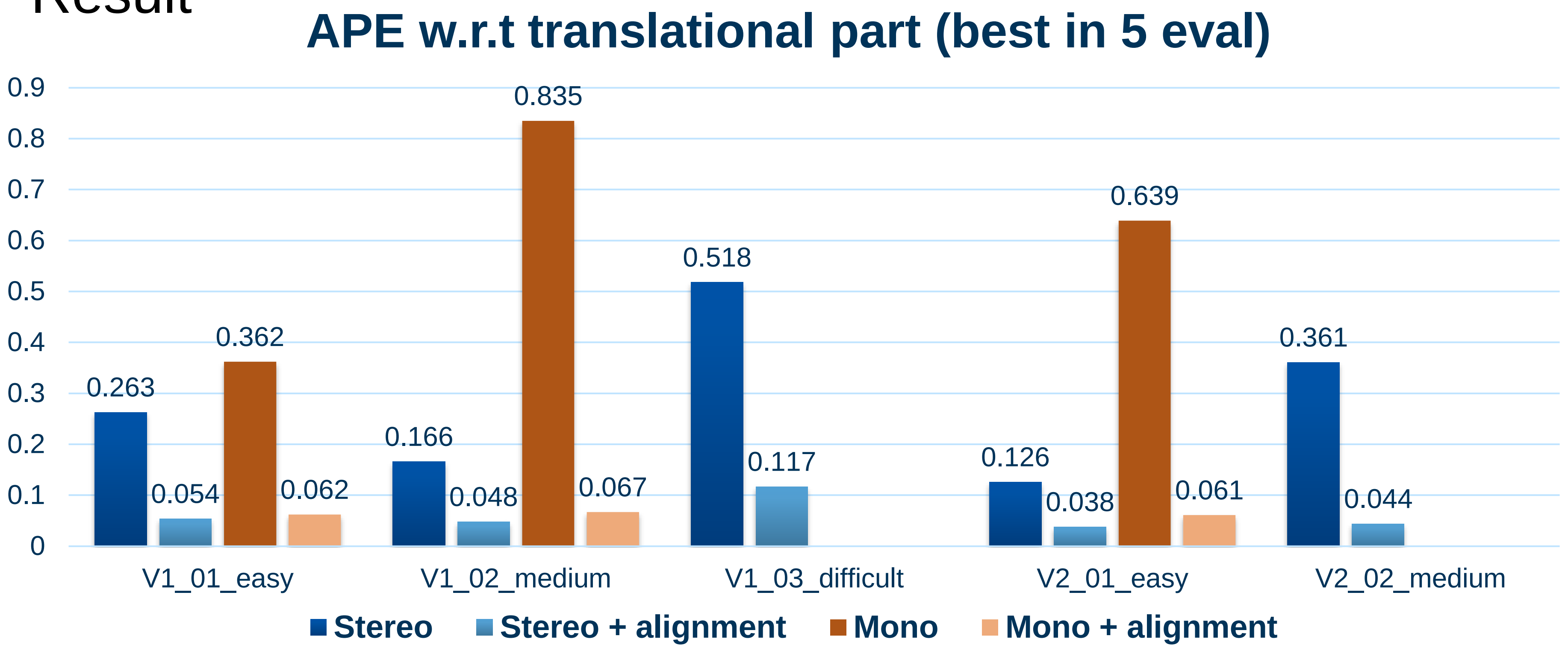
## V2_02_medium


Stereo



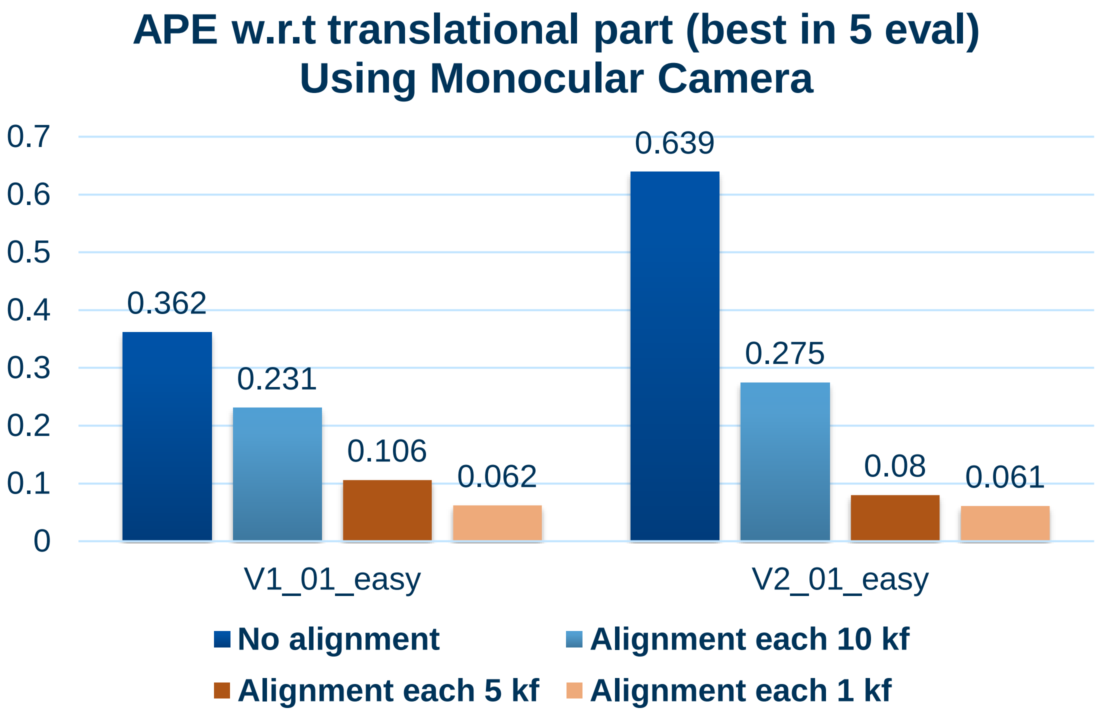

Stereo + LiDAR map

# Result



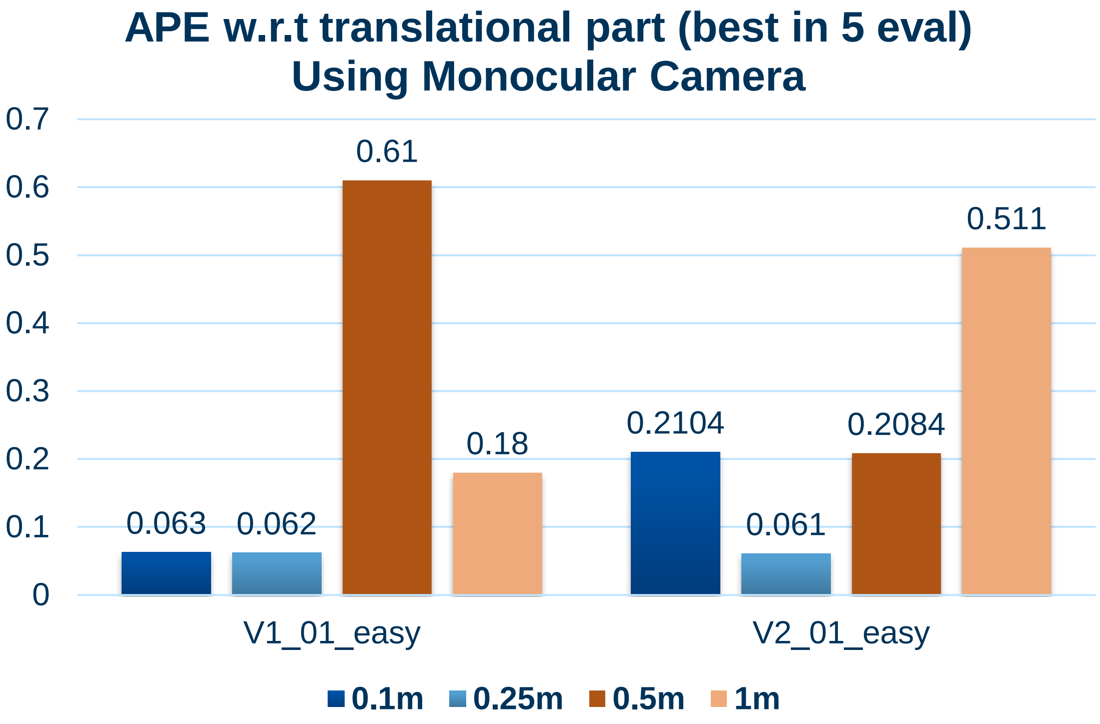
APE w.r.t translational part (best in 5 eval)
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
0.263
0.054
0.362
0.062
0.166
0.048
0.835
0.067
0.518
0.117
0.126
0.038
0.639
0.061
0.361
0.044
V1_01_easy
V1_02_medium
V1_03_difficult
V2_01_easy
V2_02_medium
Stereo
Stereo + alignment
Mono
Mono + alignment

# Ablation Study-Frequency of alignment




- More alignments leads to better performance
- Also leads to more running time!

# Ablation Study-Voxel Size



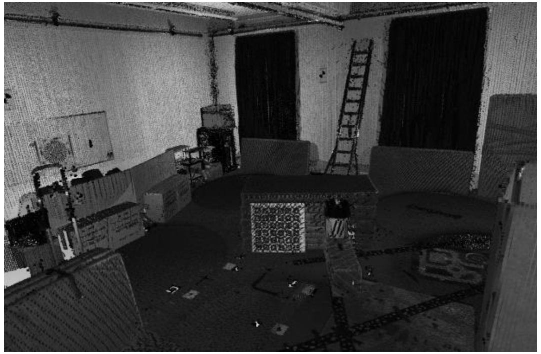

- Voxel size is highly **scene-specific**
- **Trade-off** between voxel size and running time

# KITTI Attempt

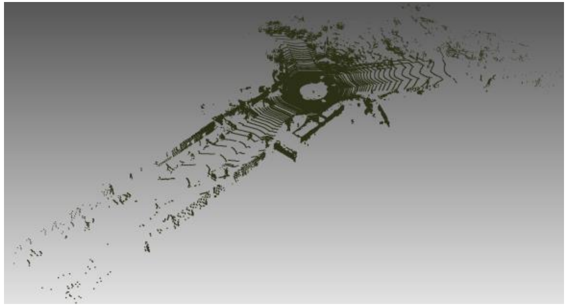
**Euroc:** LiDAR map of the whole scene (~100 Mb/scene)

**KITTI:** Each frame has a LiDAR map (.bin file) (~10 Gb/scene)

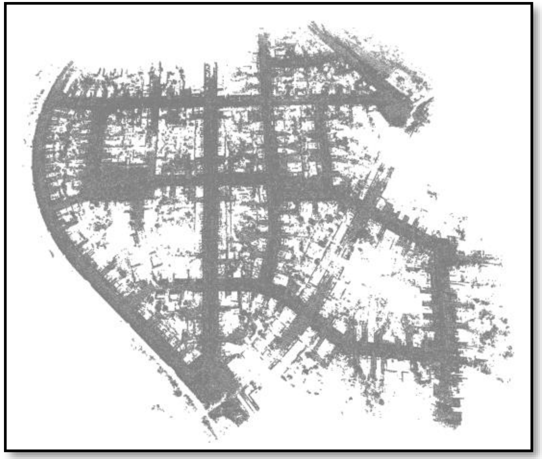
Using downsample to preprocess a map for whole scene



data.ply for one Euroc Scene

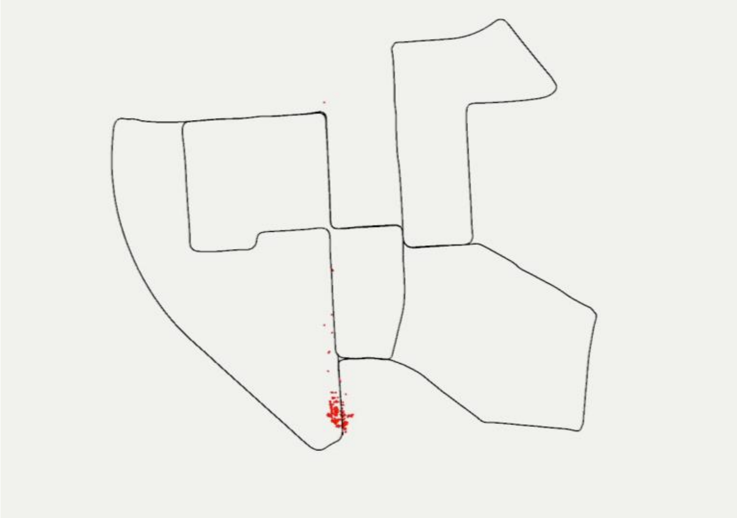


000000.bin in KITTI Sequence (~130k points)

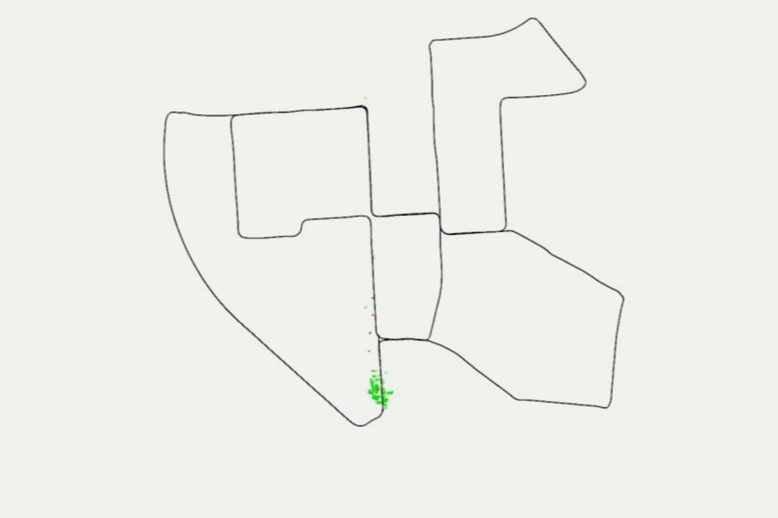


Our LiDAR Map for whole scene

# KITTI Attempt-Result



Stereo



Stereo + whole LiDAR map

# Summary

## Contribution

- Alignment with LiDAR map can **eliminate the accumulated drift**
- Matching based on geometry is **robust to light changes**
- The alignment performance is still **restricted by the based VO performance**

## Improvement potential

- Powerful **CPU**
- Accurate **LiDAR Map**
- Powerful **based VO**

# Thanks for your attention!

# Additional Materials

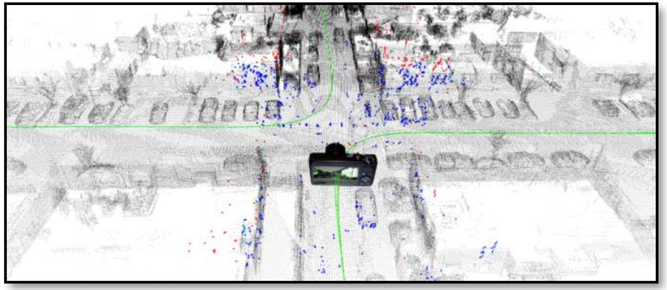
# KITTI-Attempt

Paper's method

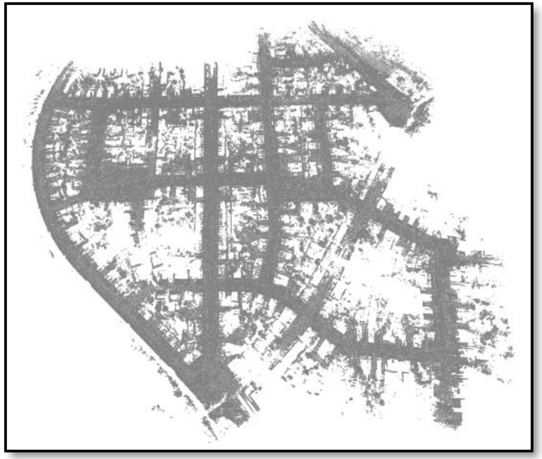
- Use a LIDAR-based SLAM system to get GT trajectory (for loop closure)
- Build a map at resolution of 20 cm

Our method

- Preprocess a LiDAR map for whole scene
  - Put point clouds for all frames together
  - Downsample
  - Build a map at resolution of 50cm



Paper's LiDAR Map for whole scene



Our LiDAR Map for whole scene