Feature-based initialization for Monocular Direct Visual Odometry

Interdisciplinary project (IDP)

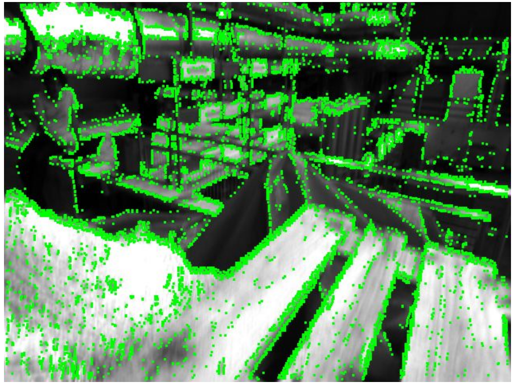
Mariia Gladkova

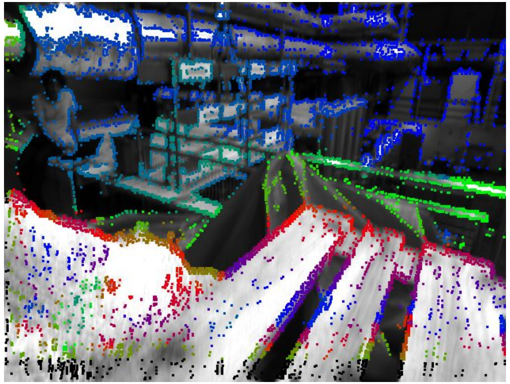
Advisor: Nikolaus Demmel
Supervisor: Prof. Daniel Cremers

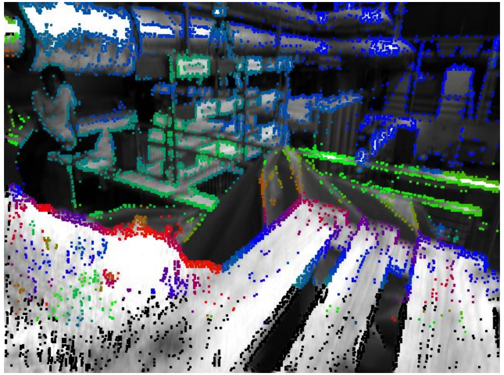
Contents

DSO Coarse Initializer

1. Select points in a grid-based manner
2. Initialize depth map to 1
3. Iteratively optimize depth map in a coarse-to-fine fashion



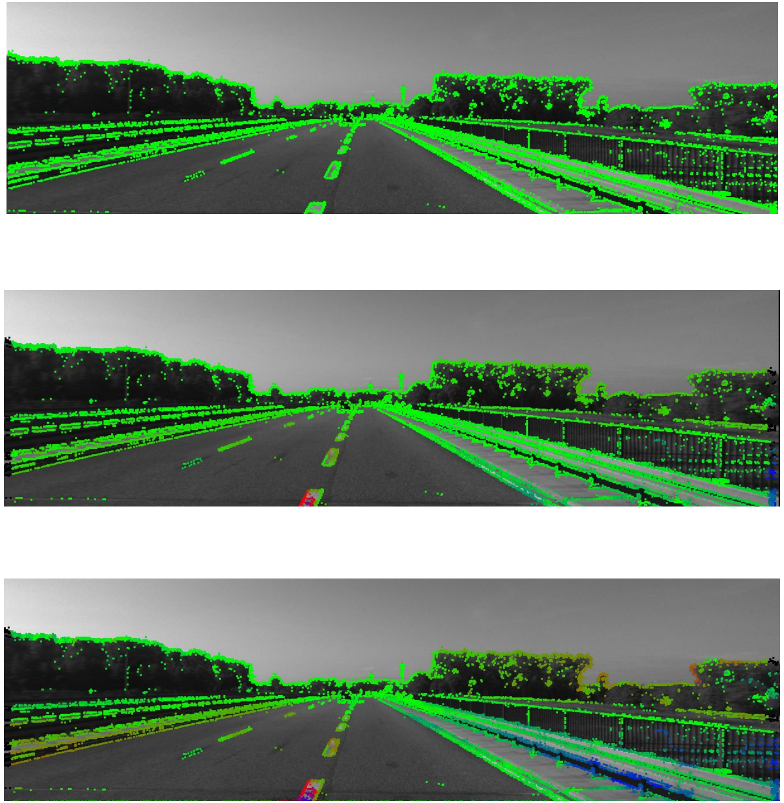




DSO Coarse Initializer

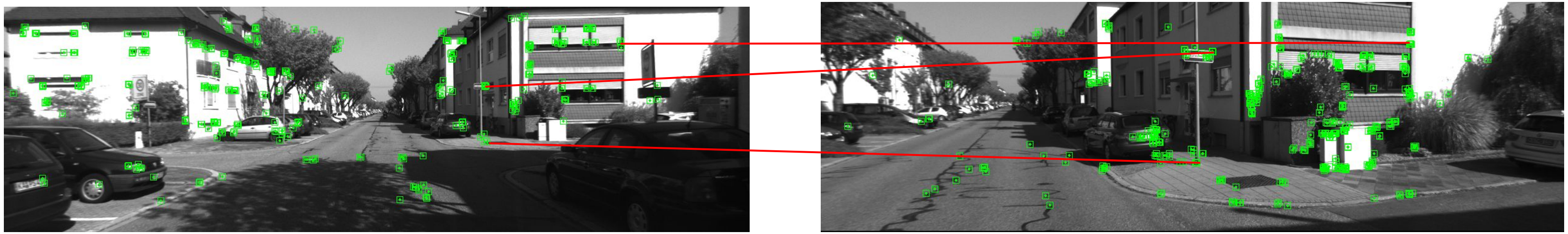
1. Select points in a grid-based manner
2. Initialize depth map to 1
3. Iteratively optimize depth map in a coarse-to-fine fashion

Needs good map initialization for convergence!



Project idea

- Address DSO initialization in a feature-based manner



1. Implement robust feature-based initializer
2. Evaluate its performance in comparison with DSO Coarse Initializer

[1] Rublee, Ethan, et al. "ORB: An efficient alternative to SIFT or SURF." *2011 International conference on computer vision*. Ieee, 2011.

Our overall pipeline

1. Relative pose estimation with homography and fundamental matrix models
    a. Extension to ORB SLAM initialization approach
2. Geometric Bundle Adjustment with outlier removal
3. Additional point extraction with depth initialization using epipolar line search
4. Photometric Bundle Adjustment with outlier removal

Relative pose estimation

1. Find feature correspondences between reference and current frames
2. Fit two models in parallel: a homography H_{cr} and a fundamental matrix F_{cr} (RANSAC scheme)
3. Select based on the score ratio and recover the transformation
4. Perform full Bundle Adjustment

$$x_c = H_{cr} x_r \Longrightarrow \text{score } S_H$$

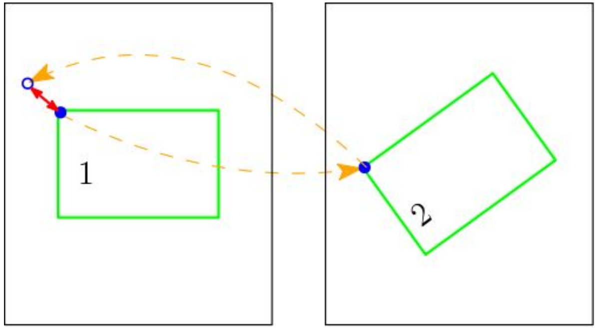
$$x_c^T F_{cr} x_r = 0 \Longrightarrow \text{score } S_F$$

$$\left.\begin{array}{l} S_H \\ S_F \end{array}\right\} \quad R_H = \frac{S_H}{S_H + S_F}$$

[1] Mur-Artal, Raul, Jose Maria Martinez Montiel, and Juan D. Tardos. "ORB-SLAM: a versatile and accurate monocular SLAM system." *IEEE transactions on robotics* 31.5 (2015): 1147-1163.

Relative pose estimation - changes to ORB SLAM initialization

- Several changes to the ORB SLAM initialization:
    - Extracting corners and tracking them with Lucas Kanade optical flow
    - Attempting to fit homography and fundamental matrix after N - 1 frames

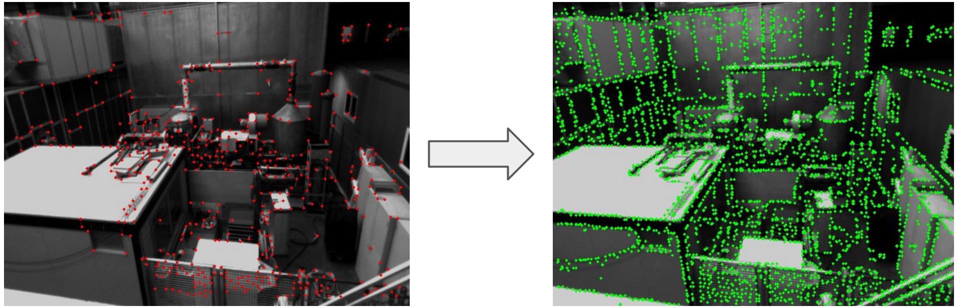


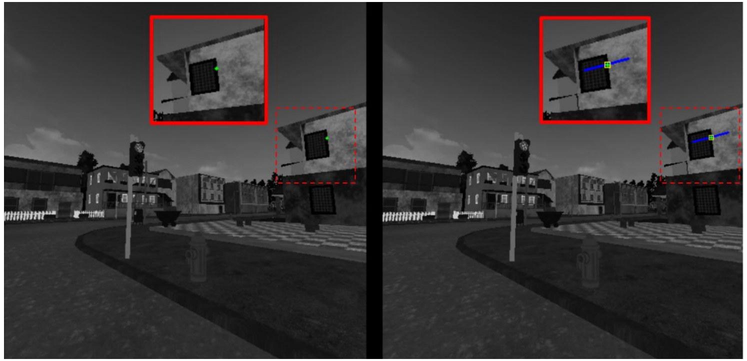
“Two-way” tracking

N - 2
frames

Point augmentation

- Grid-based point selection
- 1D search along epipolar line
  - depth prior from triangulated features
  - SSD error over 3 x 3 pixel patch
- Uncertainty propagation through N frames
  - Geometric and photometric components of depth error-variance [1]



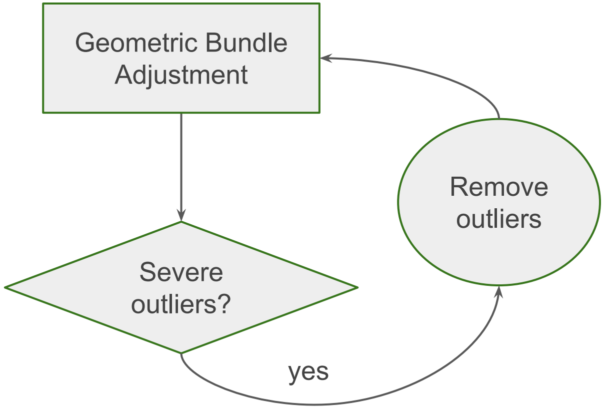


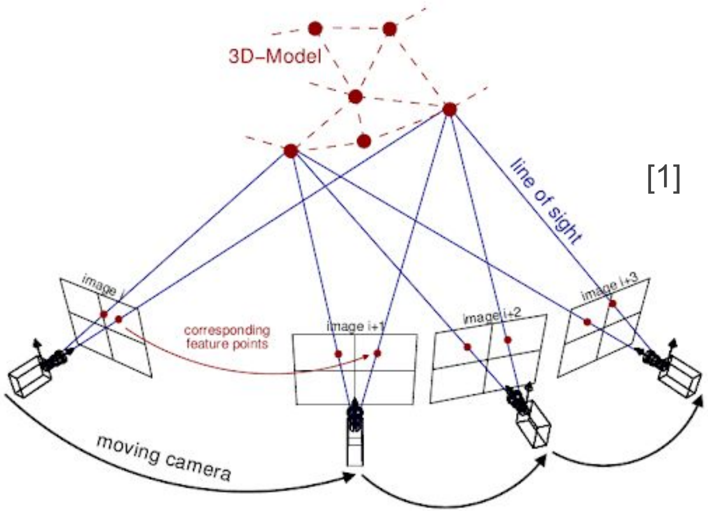
Epipolar line search

[1] Engel, Jakob, Jurgen Sturm, and Daniel Cremers. "Semi-dense visual odometry for a monocular camera." *Proceedings of the IEEE international conference on computer vision*. 2013.

Geometric Bundle Adjustment

- Geometric Bundle Adjustment
- Outlier removal based on reprojection error
    - scheme: removal - optimization - removal - optimization



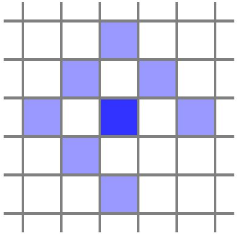


[1]

[1] http://theia-sfm.org/sfm.html

Photometric Bundle Adjustment

- Photometric Bundle Adjustment
  - 8-pixel patch
  - mean intensity normalization



[1]

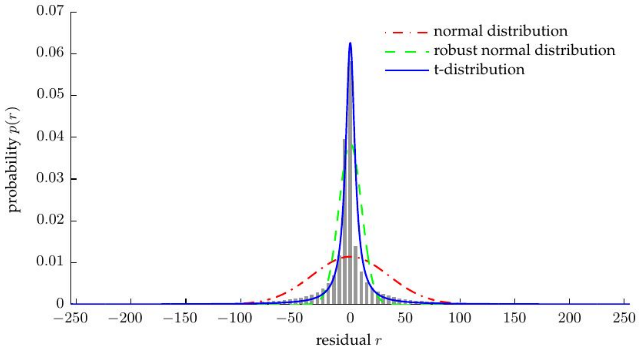
$$\min_{\{\xi_j\}_{j=1\ldots|\mathcal{C}|},\{\mathbf{x}_i\}_{i=1\ldots|\mathcal{P}|}} \sum_{i=1}^{|\mathcal{P}|} \sum_{j\in obs(\mathbf{x}_i)} \sum_{\Delta\in\mathcal{N}(\mathbf{p}_i)} ||I_j(\pi(\xi_j(\mathbf{x}_i + \pi^{-1}\Delta))) - \psi I(\pi\mathbf{x}_i + \Delta)||_\gamma$$

with $$\psi = \frac{\sum_{\Delta\in\mathcal{N}(\mathbf{p})} I_j(\pi\mathbf{T}_j(\mathbf{x} + \pi^{-1}\Delta))}{\sum_{\Delta\in\mathcal{N}(\mathbf{x})} I(\pi\mathbf{x} + \Delta)}$$

[1] Engel, Jakob, Vladlen Koltun, and Daniel Cremers. "Direct sparse odometry." *IEEE transactions on pattern analysis and machine intelligence* 40.3 (2017): 611-625.

Outlier removal

- T-distribution of residuals
  - mean, variance and degrees of freedom
  - 8 residuals per point



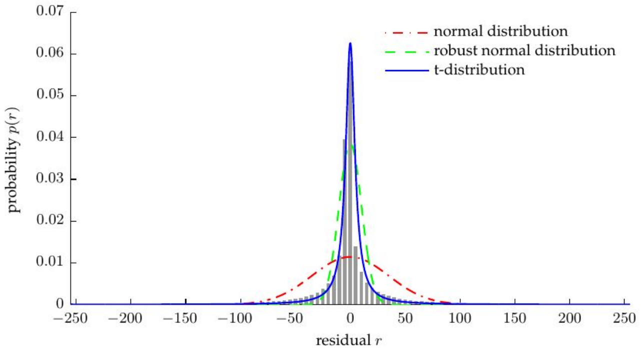
[1]

$$\sigma^2_{k+1} = \frac{1}{n}\sum_{i=1}^{n}\frac{\nu+1}{\nu+r_i^2/\sigma^2_k}r^2_i$$

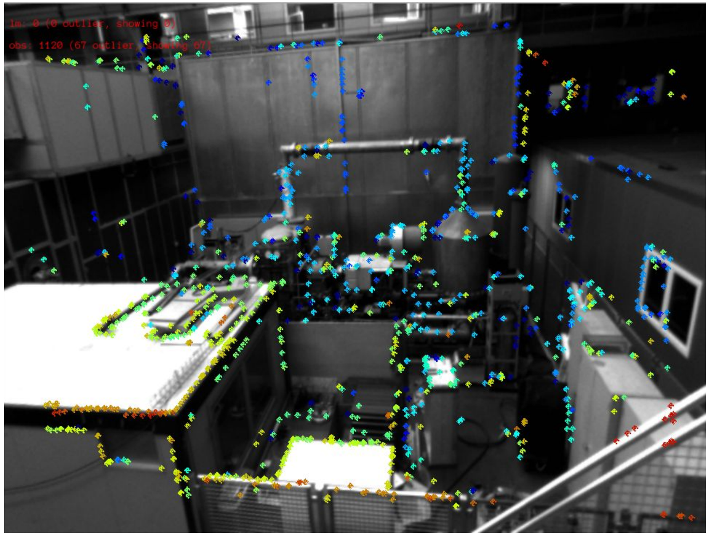
[1] Kerl, Christian. "Odometry from rgb-d cameras for autonomous quadrocopters." *Master's Thesis, Technical University* (2012).

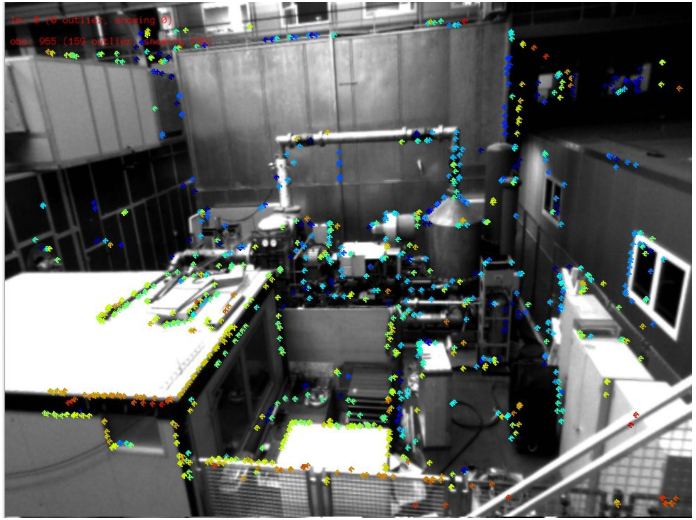
Outlier removal

- T-distribution of residuals
  - mean, variance and degrees of freedom
  - 8 residuals per point



[1]

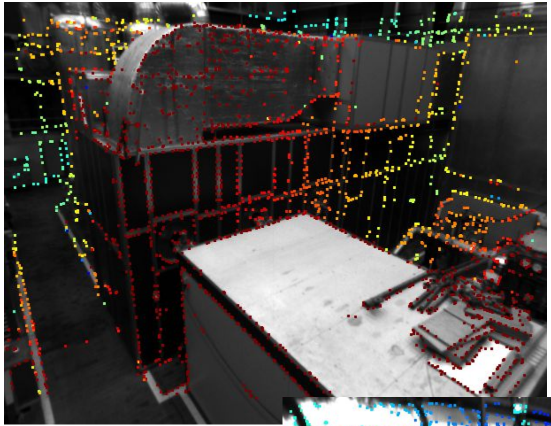


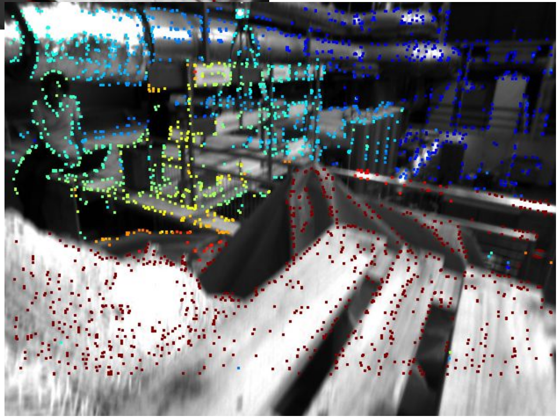


[1] Kerl, Christian. "Odometry from rgb-d cameras for autonomous quadrocopters." *Master's Thesis, Technical University* (2012).

Evaluation setup

1. KITTI dataset (sequences 00 - 10)
2. EuRoC dataset [*]





- initialization accuracy and overall robustness
  - trajectory error over 200 frames
  - different starting points for each run

[*] MH_01_easy, MH_02_easy, MH_03_medium, MH_04_difficult, MH_05_difficult, V1_01_easy, V1_02_medium, V1_03_difficult, V2_01_easy, V2_02_medium, V2_03_difficult

Evaluation setup : metrics used

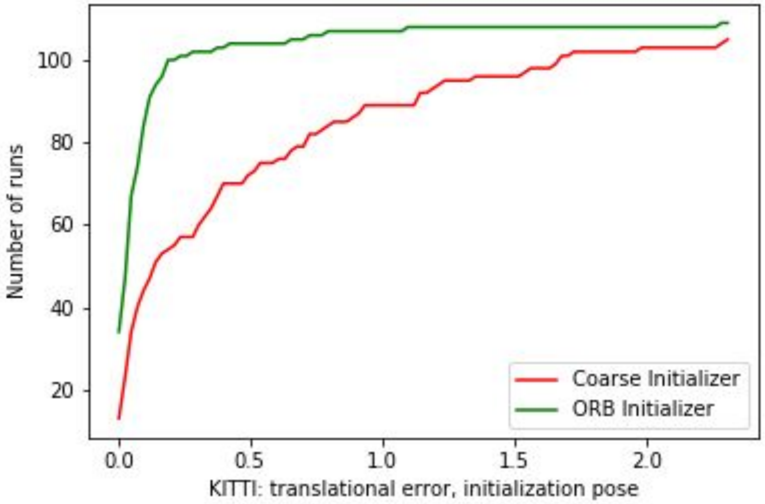
$$ATE_{1...n} = \left(\frac{1}{n} \sum_{i=1}^{n} ||transl(F_i)||_2^2 \right)^{\frac{1}{2}}$$ where $$F_i = Q_i^{-1} S P_i$$

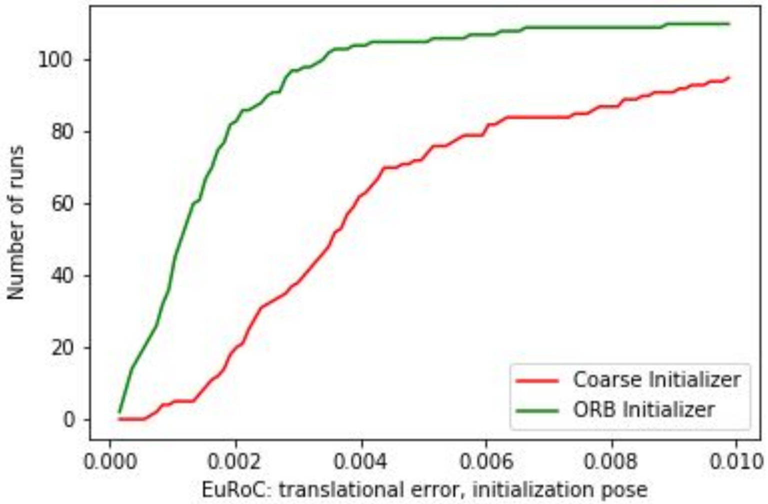
$$RPE_{1...n} = \left(\frac{1}{n-1} \sum_{i=1}^{n-1} ||transl(E_{i,i+1})||_2^2 \right)^{\frac{1}{2}}$$ where $$E_{i,j} = (Q_i^{-1} Q_j)^{-1} (P_i^{-1} P_j)$$

[1] Sturm, Jürgen, et al. "A benchmark for the evaluation of RGB-D SLAM systems." *2012 IEEE/RSJ International Conference on Intelligent Robots and Systems*. IEEE, 2012.
[2] Choice of the metrics: ATE and number of frames used for initialization are based on the evaluation done by Xingwei Qu in his master thesis *Initialization Methods for Visual and Visual-inertial SLAM*

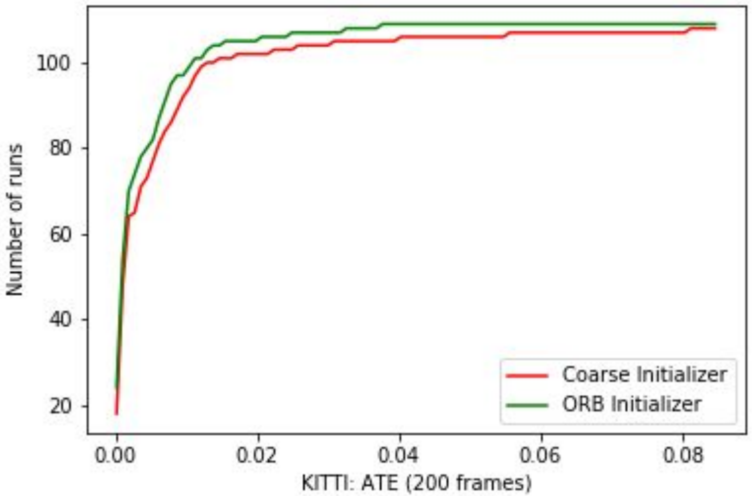
Evaluation (initialization)

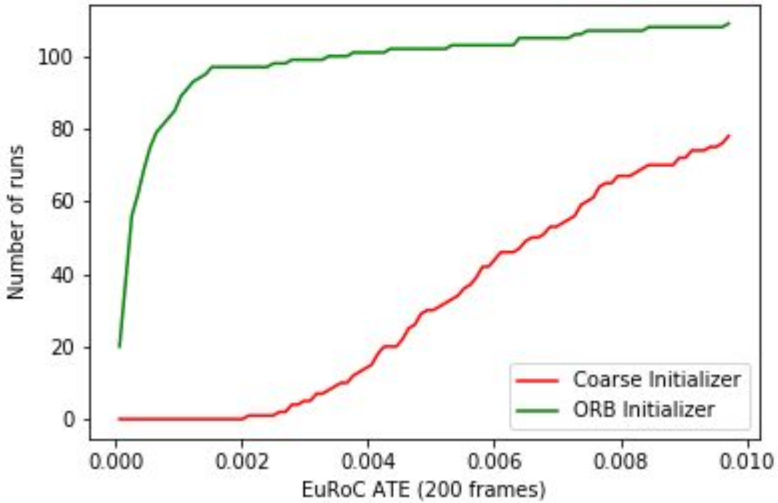
Average number frames	KITTI	EuRoC
Coarse Initializer	7.3	7.82
ORB Initializer	4.78	5.28



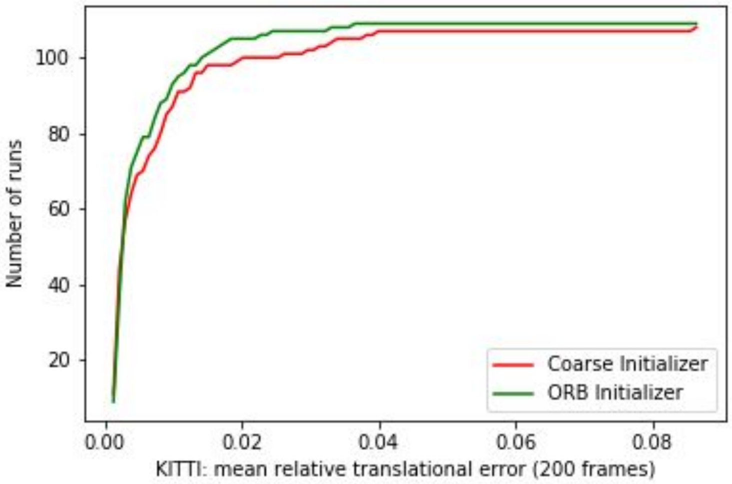


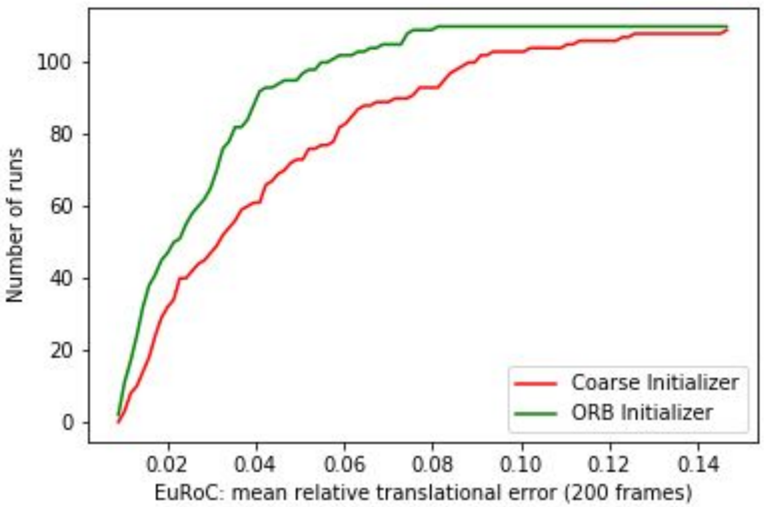
Evaluation (ATE)





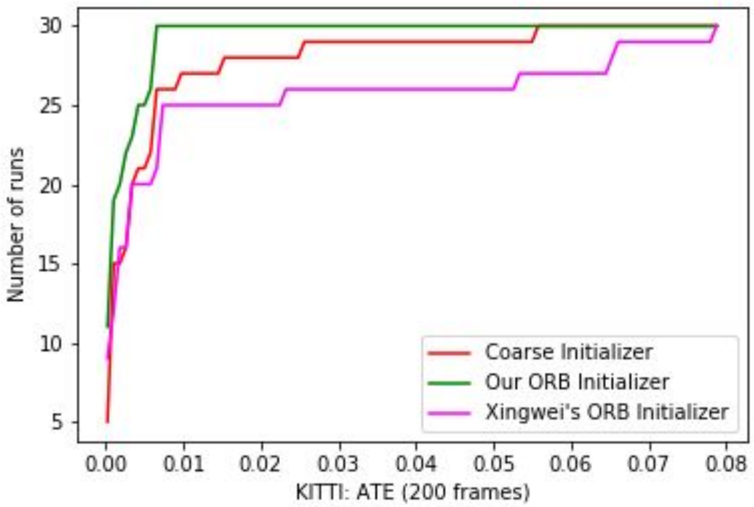
Evaluation (RPE)





Evaluation with work by Xingwei Qu

- KITTI sequences (00, 01, 02)



	Coarse Initializer	Xingwei’s ORB Initializer	Our ORB Initializer
Average number frames	7.0	12.367	4.8

Conclusion and future work

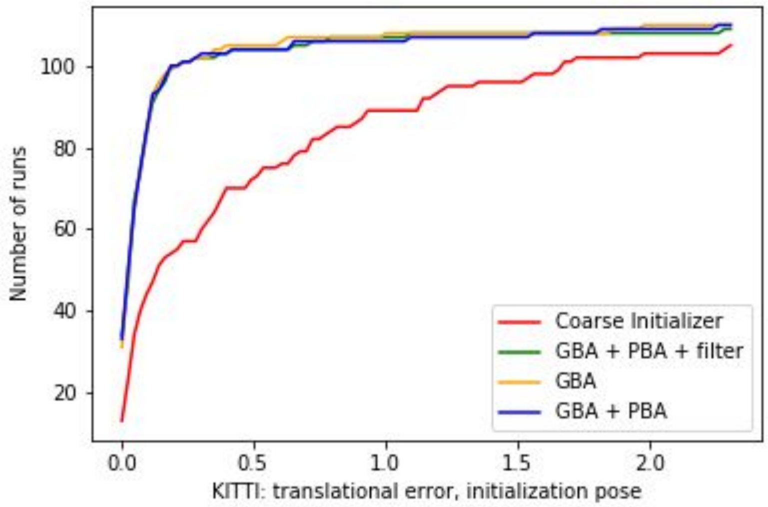
- Improvement of runtime
- Further tuning of the outlier removal parameters to boost PBA performance
- Optimization of affine light transform parameters
- Learned features (e.g. SuperPoint)

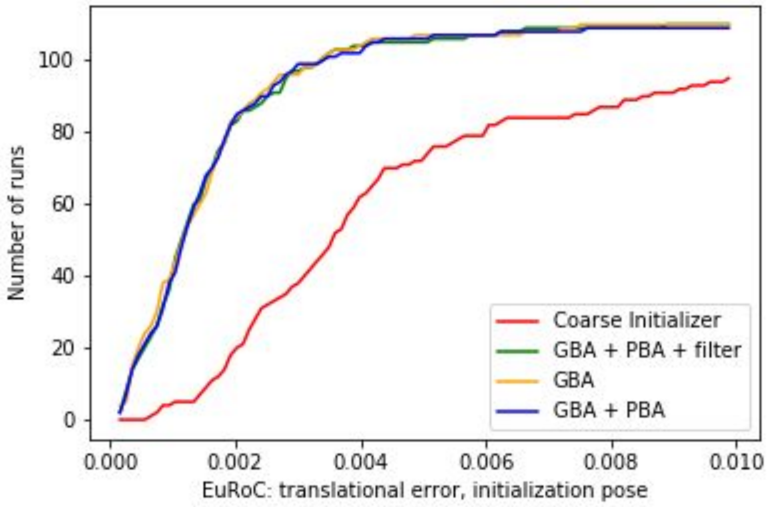
Thank you for your attention!

Appendix: Pipeline overview

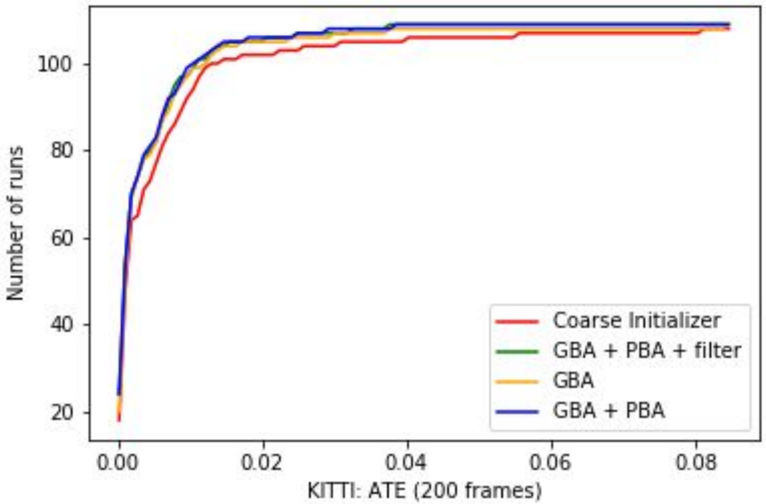
```
if |frame.frame_id - ref_frame_id| ≥ N then
    success ← map.computeRelativeTransform(frame, ref_frame);
    if success then
        do
            map.performGeometricBA();
            severeOutliers ← map.removeSevereOutliers();
        while severeOutliers > 0;
        map.populateStructure3D();
        do
            map.performGeometricBA();
            severeOutliers ← map.removeSevereOutliers();
        while severeOutliers > 0;
        if map.performPhotometricBA() then
            map.fitTDistribution();
            map.performPhotometricBA();
        return SUCCESS;
return FAILURE;
```

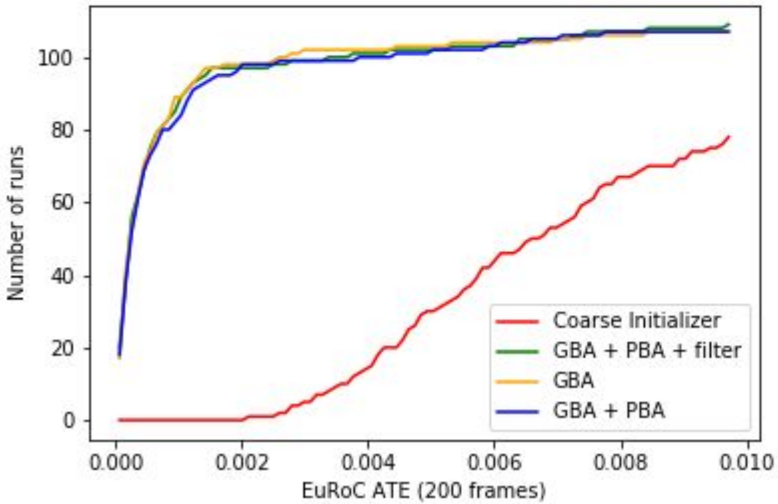
Appendix: evaluation of ORB Initializer versions (initial pose)



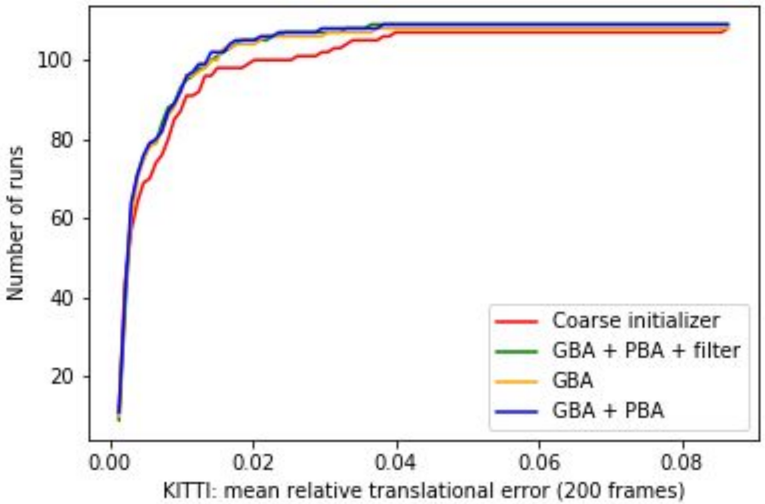


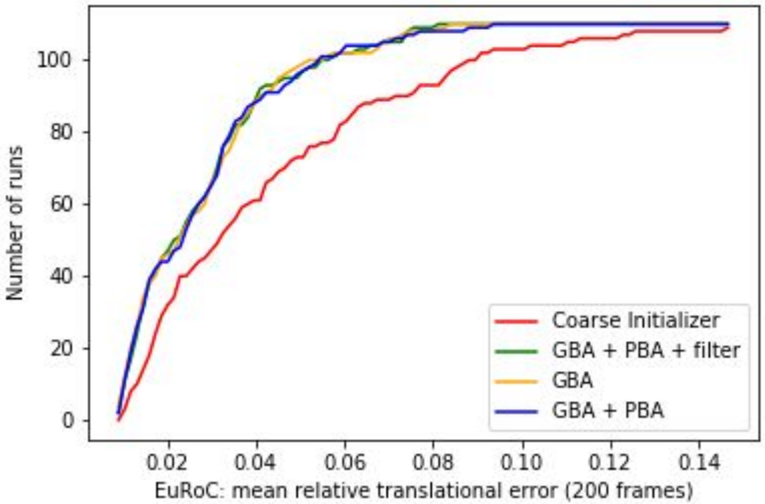
Appendix: evaluation of ORB Initializer versions (ATE)



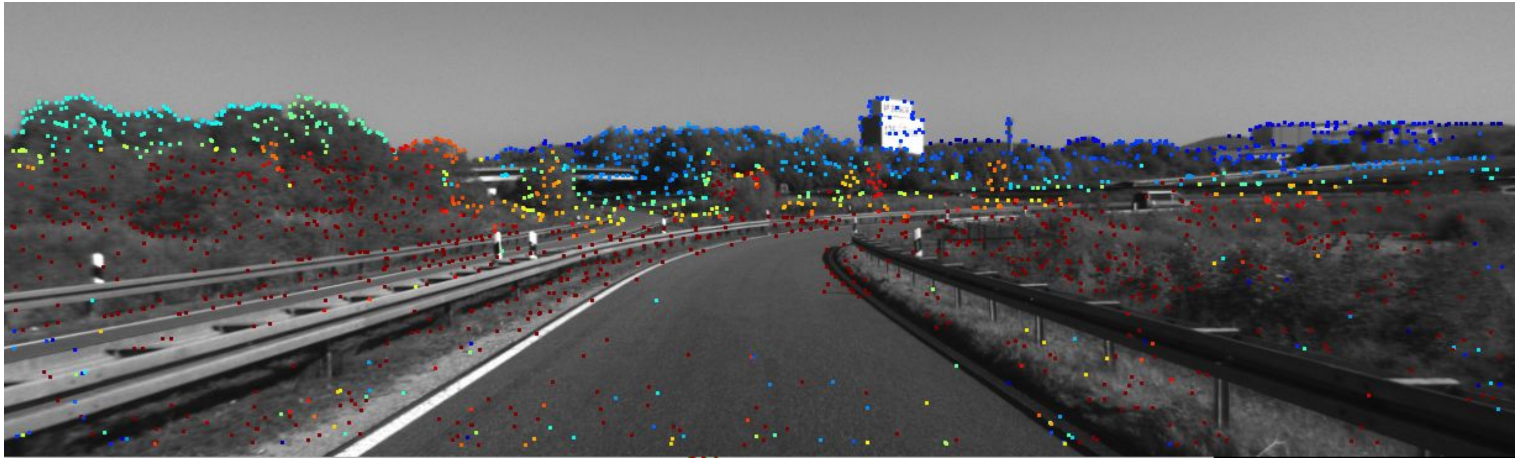


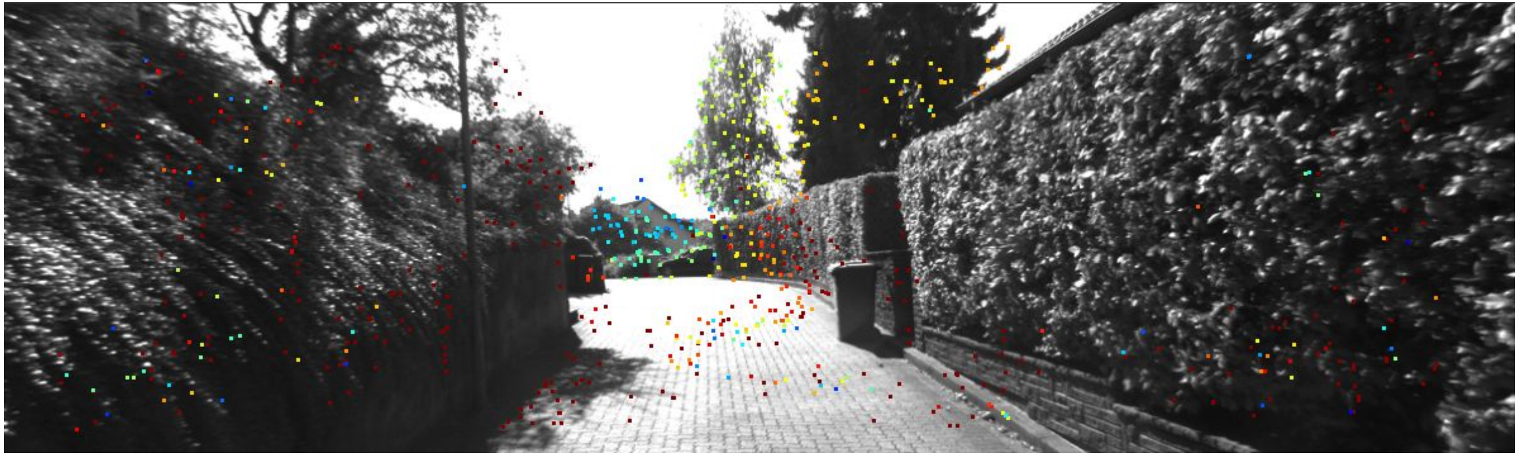
Appendix: evaluation of ORB Initializer versions (RPE)





Failed cases: wrong depth estimation





Failed cases: incorrect model chosen