Visual Slam

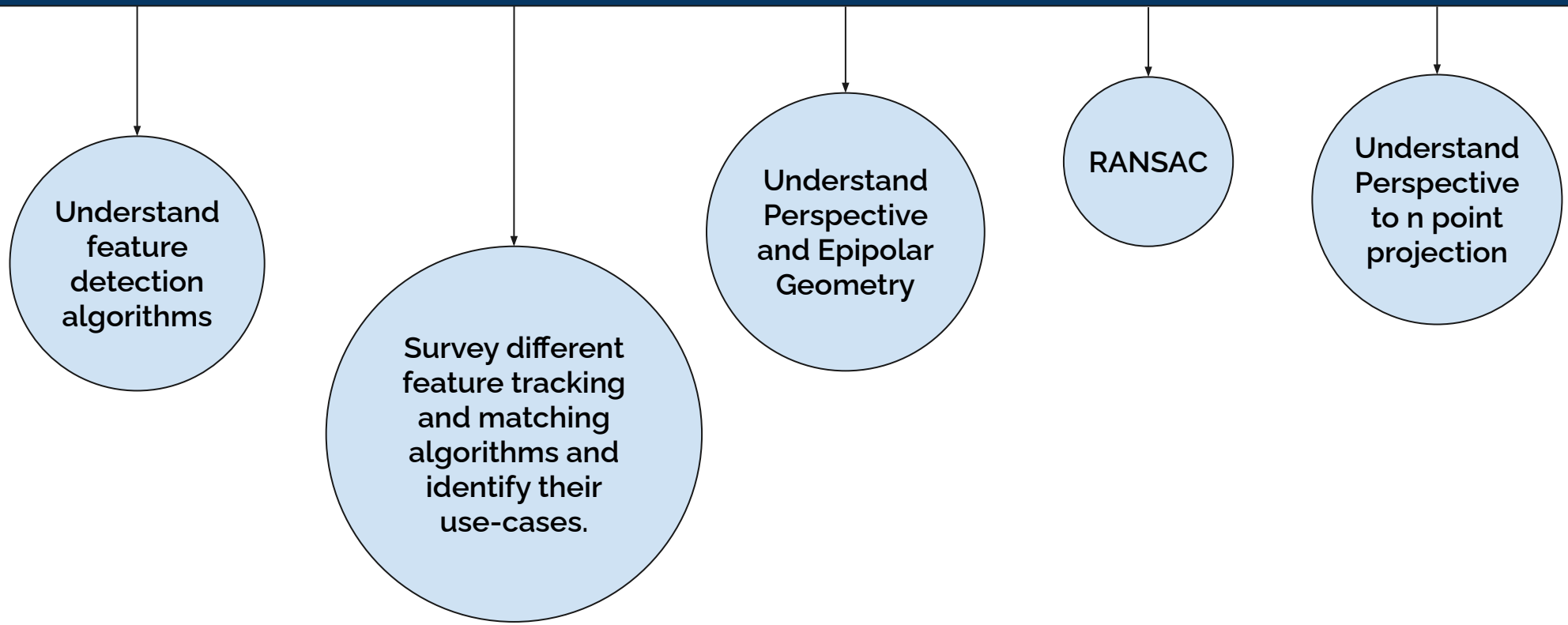
Mayur Bhise, Kyungsun Lee, Ukhyeon Shin, Aniket Gupta

Presentation Outline

- Problem Statement
- Visual Odometry
- VO vs Visual SLAM
- Stereo vs Mono VO
- Orb SLAM 3
- Orb SLAM 3 performance
- Our VO Pipeline
- Failures and Learnings
- Results
- Future Work
- Conclusion

Problem Statement

Goal: Understand Visual odometry pipeline for monocular and stereo camera and compare the performance of the frontend with Orb SLAM 3.



What is Visual Odometry?

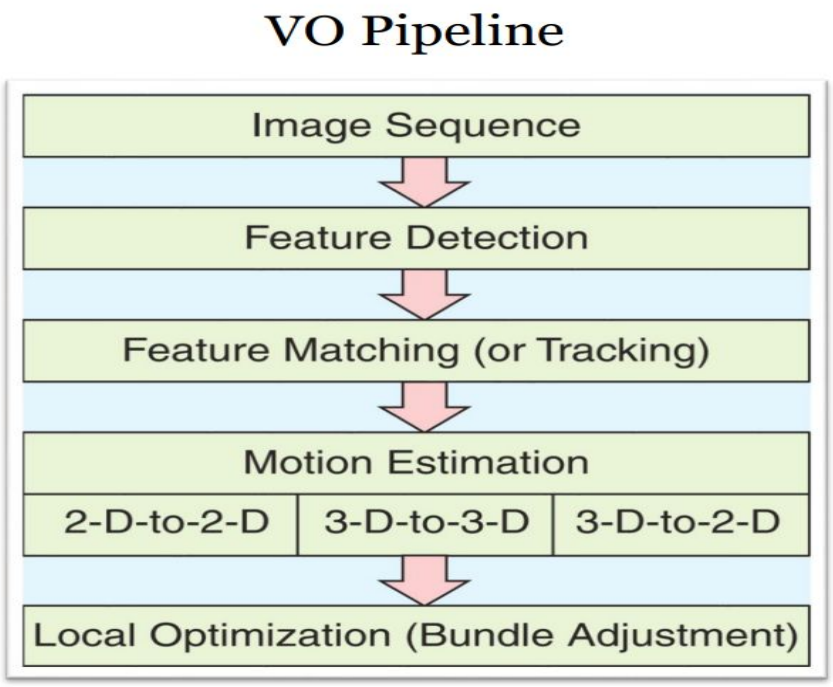
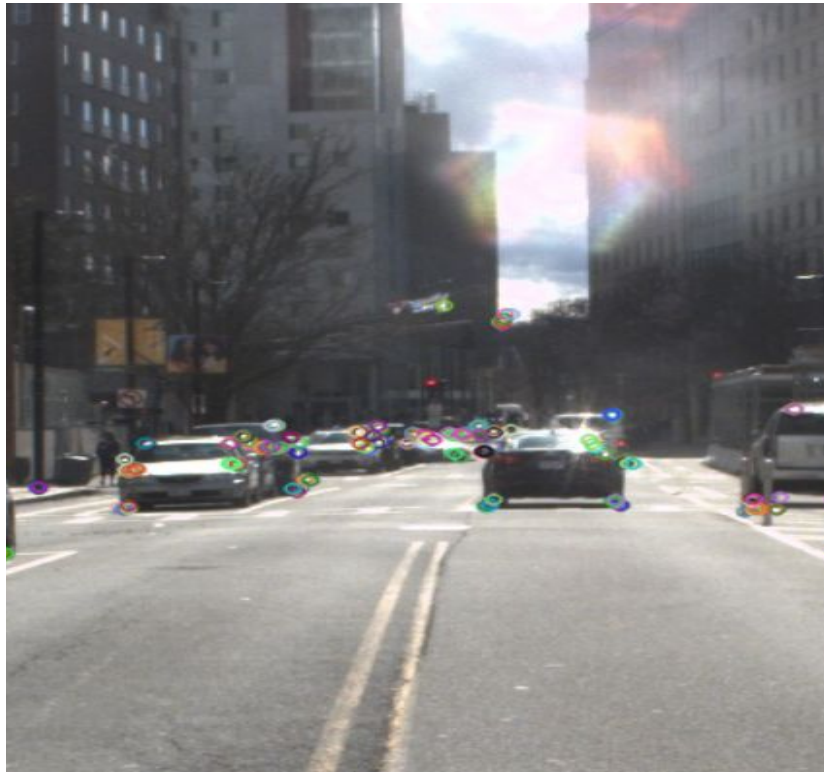
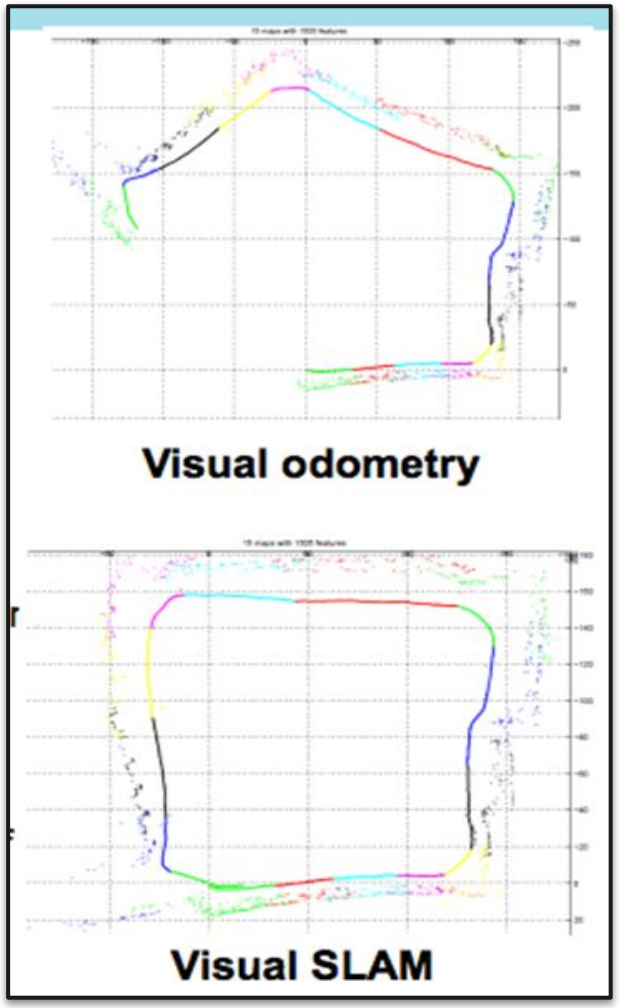


Image from Scaramuzza and Fraundorfer, 2011



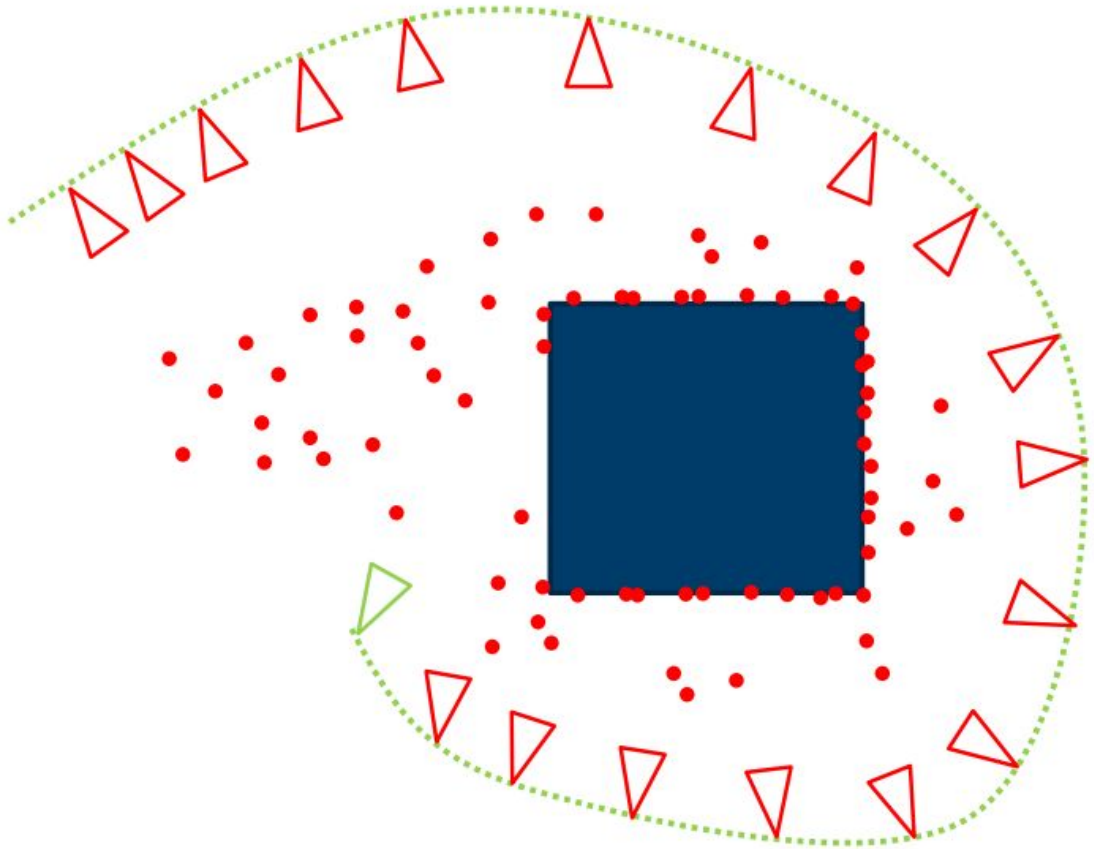
Visual Odometry vs Visual slam

1. VO only aims at the local consistency of the trajectory
2. SLAM aims to the global consistency of the trajectory and of the map
3. VO is SLAM before closing the loop!
4. VO - Consistency, Real Time
5. Visual Slam - Performance

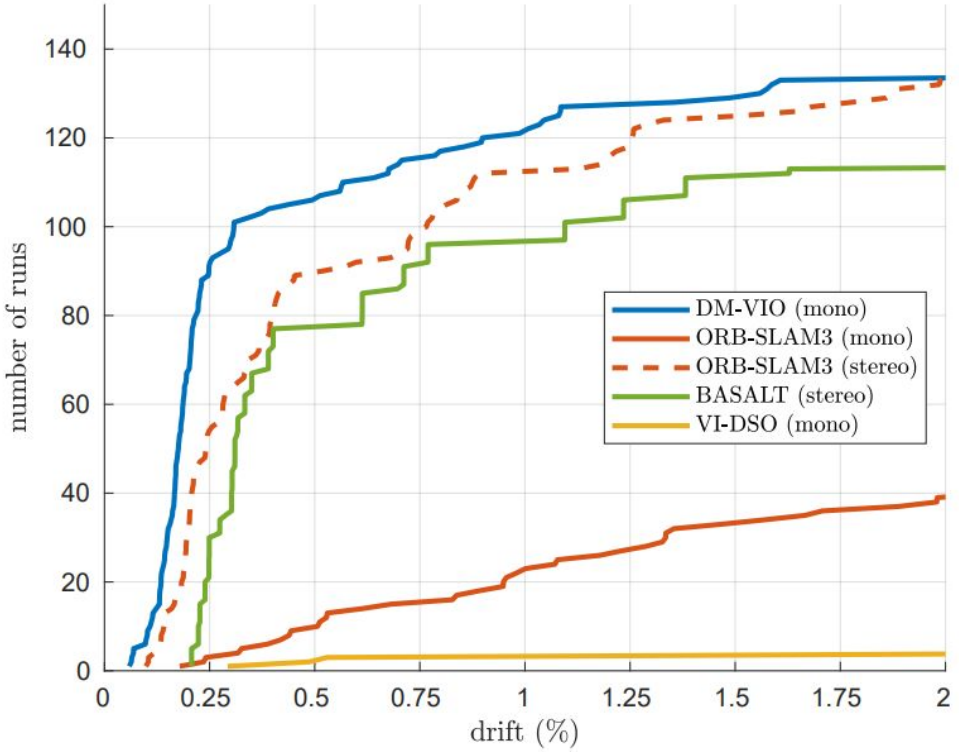


Difference between using Stereo and Mono Vision in SLAM

- Better scale prediction
- Robust in 3d regeneration of scene
- Better Localization



Delayed Marginalization VIO



https://vision.in.tum.de/research/vslam/dm-vio?redirect=1

ORB-SLAM 3 System Overview

Tracking Thread

- Keyframe-based minimization
- Robustness against tracking loss

Local Mapping Thread

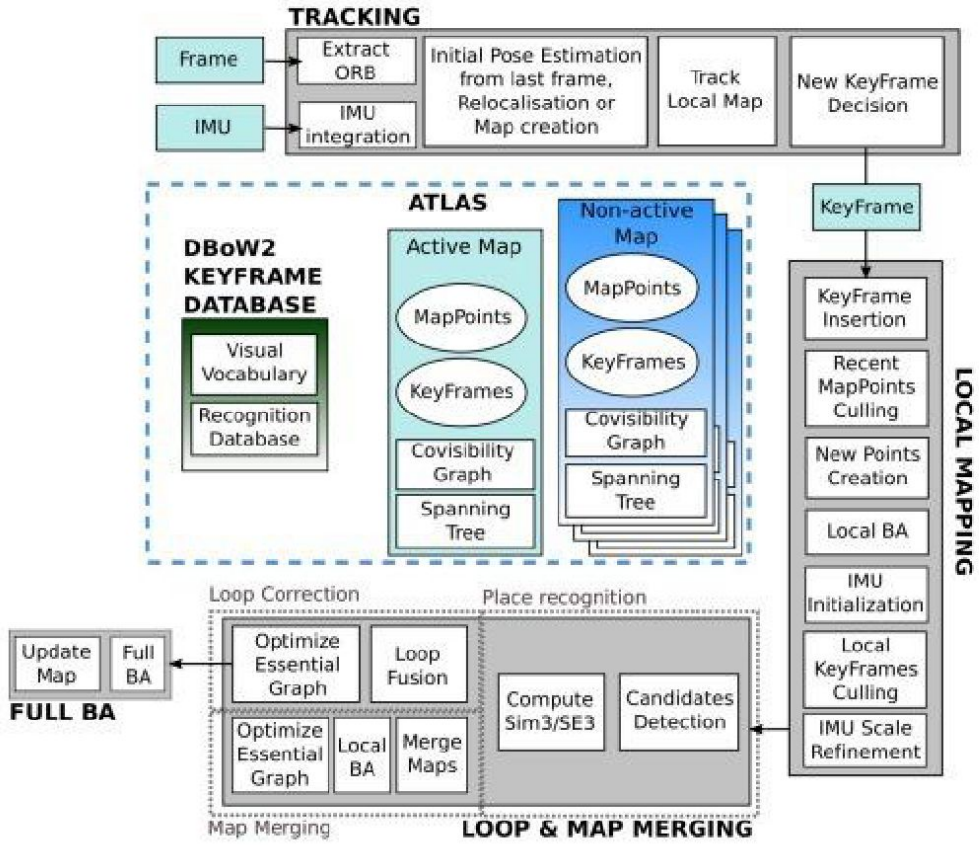
- Sliding window of keyframes
- Robustness against poor observability of inertial parameters

Atlas

- Multimap representation to handle unlimited submaps
- Map creation / map merging / re-localization

Loop and Map Merging Thread

- Detect common points using improved DBoW2
- Perform Loop closure (in active map) Map merge (in Atlas)

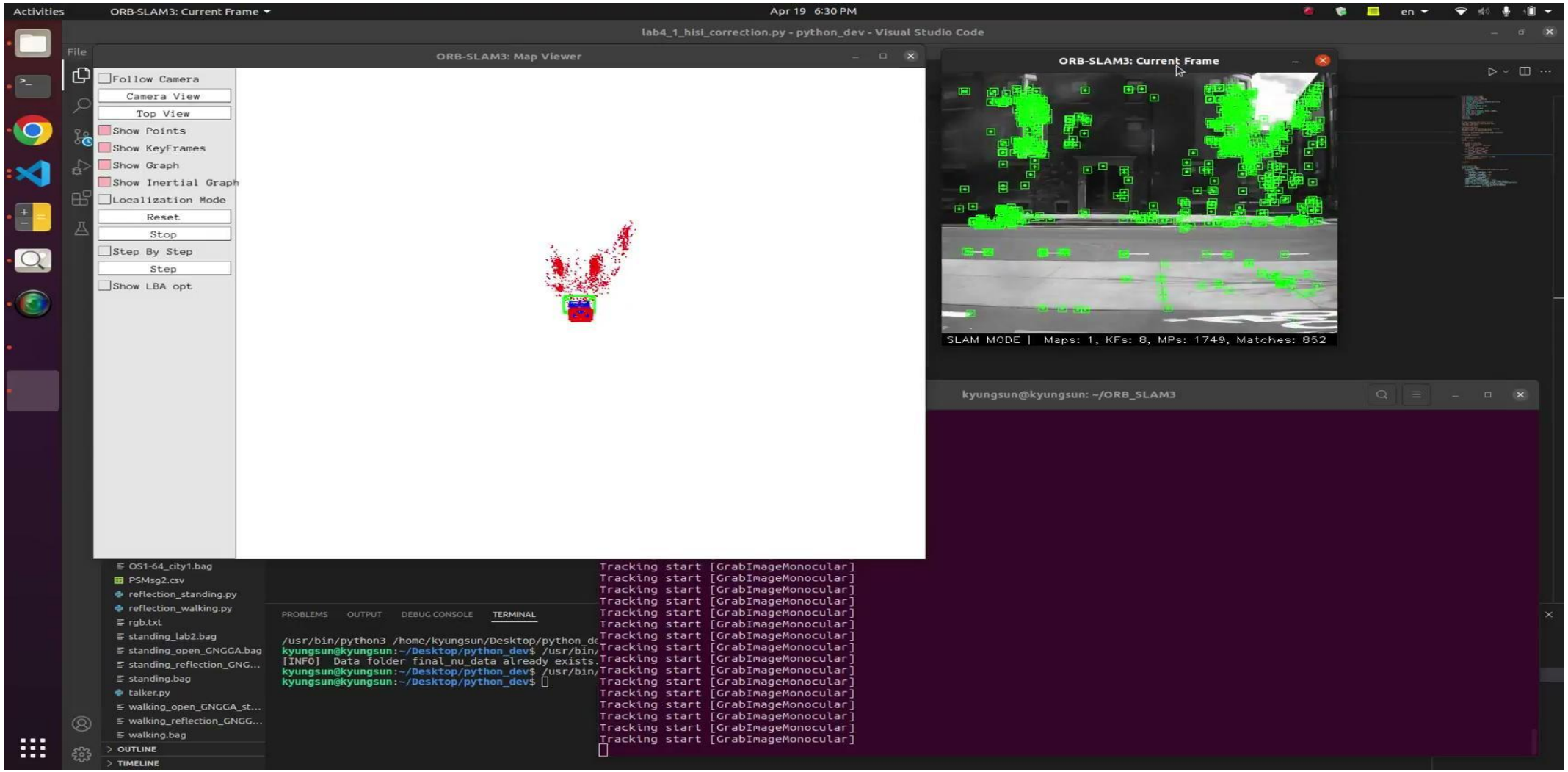


ORB SLAM3 - Monocular NUance Data

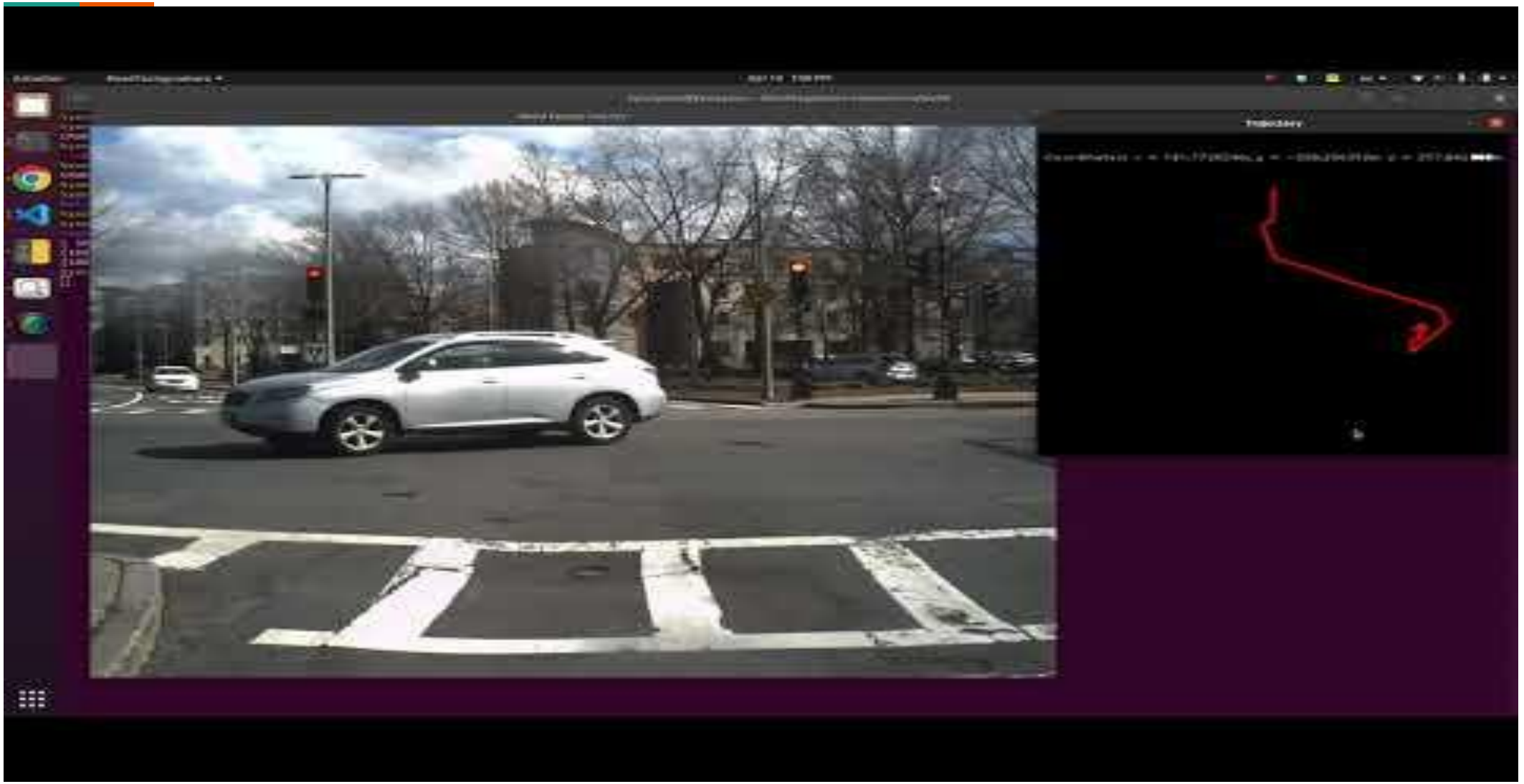
```
#--------------------------------------------
# Camera Parameters. Adjust them!
#--------------------------------------------
File.version: "1.0"

Camera.type: "PinHole"

# Camera calibration and distortion parameter
Camera1.fx: 755.378062328
Camera1.fy: 755.360037963
Camera1.cx: 245.275906054
Camera1.cy: 192.847576368

Camera1.k1: 0.0
Camera1.k2: 0.0
Camera1.p1: 0.0
Camera1.p2: 0.0
Camera1.k3: 0.0

# Camera frames per second
Camera.fps: 100

# Color order of the images (0: BGR, 1: RGB.
Camera.RGB: 1

# Camera resolution
Camera.width: 487
Camera.height: 408
```

```
#--------------------------------------------------------
# ORB Parameters
#--------------------------------------------------------

# ORB Extractor: Number of features per image
ORBextractor.nFeatures: 5000

# ORB Extractor: Scale factor between levels in the scal
ORBextractor.scaleFactor: 1.2

# ORB Extractor: Number of levels in the scale pyramid
ORBextractor.nLevels: 8

# ORB Extractor: Fast threshold
# Image is divided in a grid. At each cell FAST are extr
# Firstly we impose iniThFAST. If no corners are detecte
# You can lower these values if your images have low con
ORBextractor.iniThFAST: 10
ORBextractor.minThFAST: 3
```

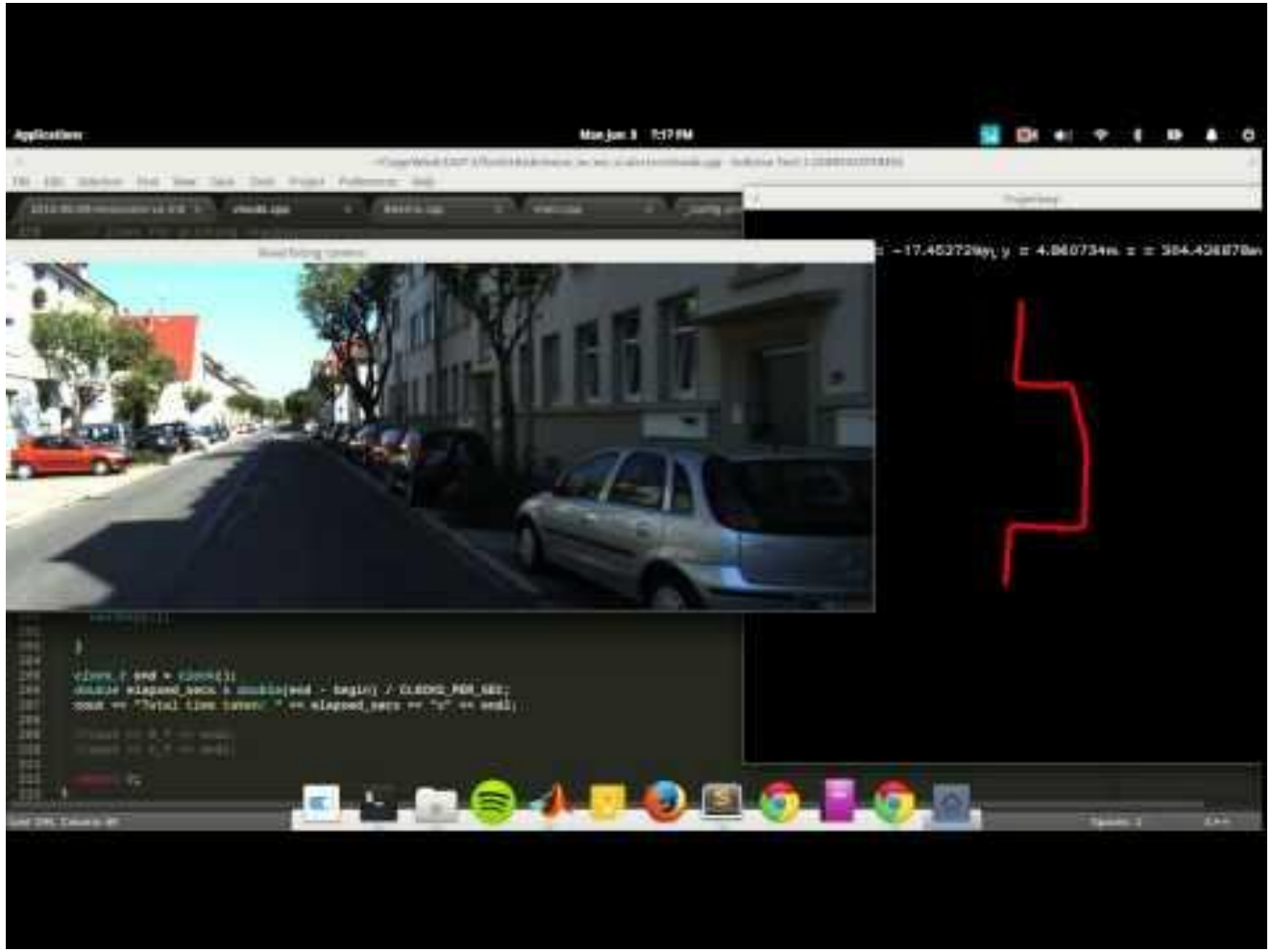
ORB SLAM3 - Monocular NUance Data



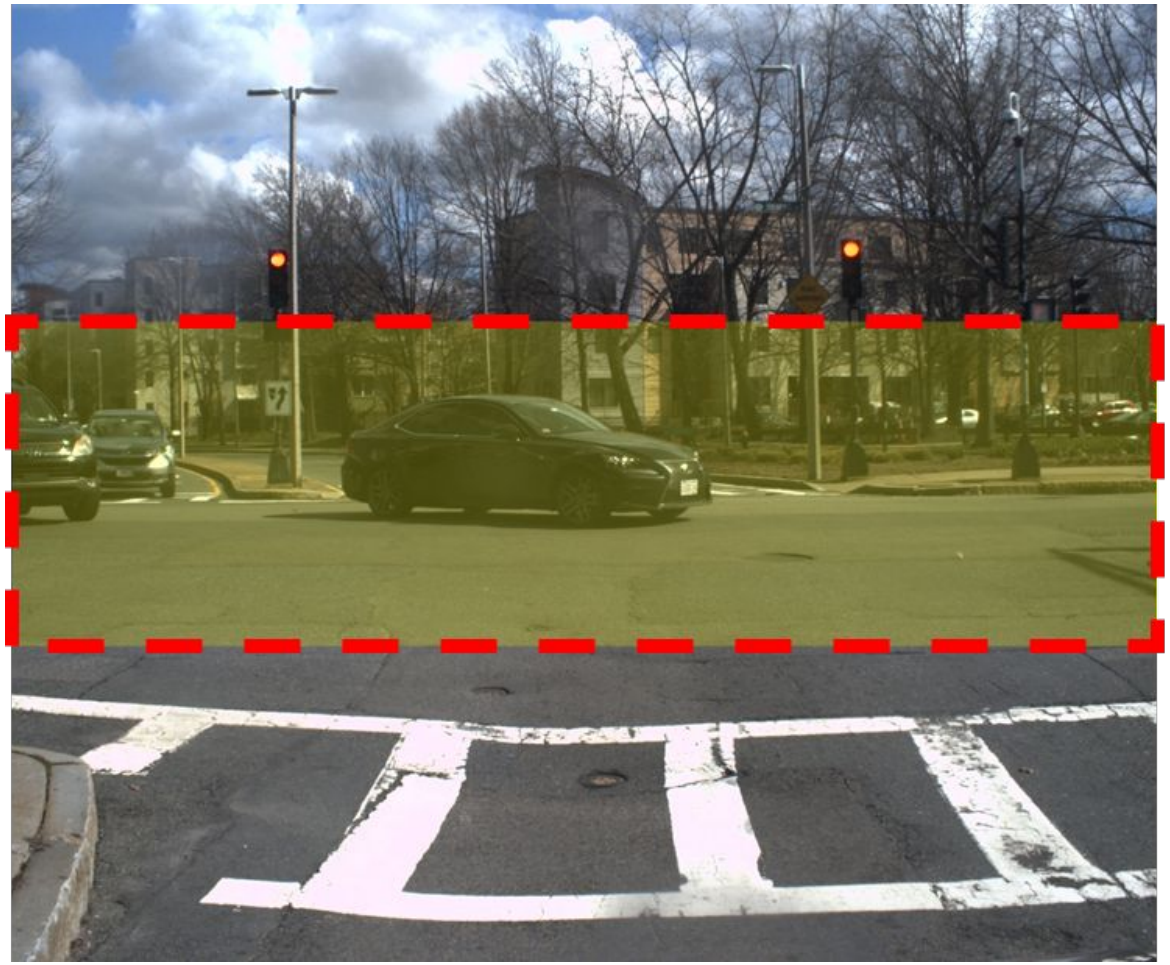
Failure Case: Moving objects



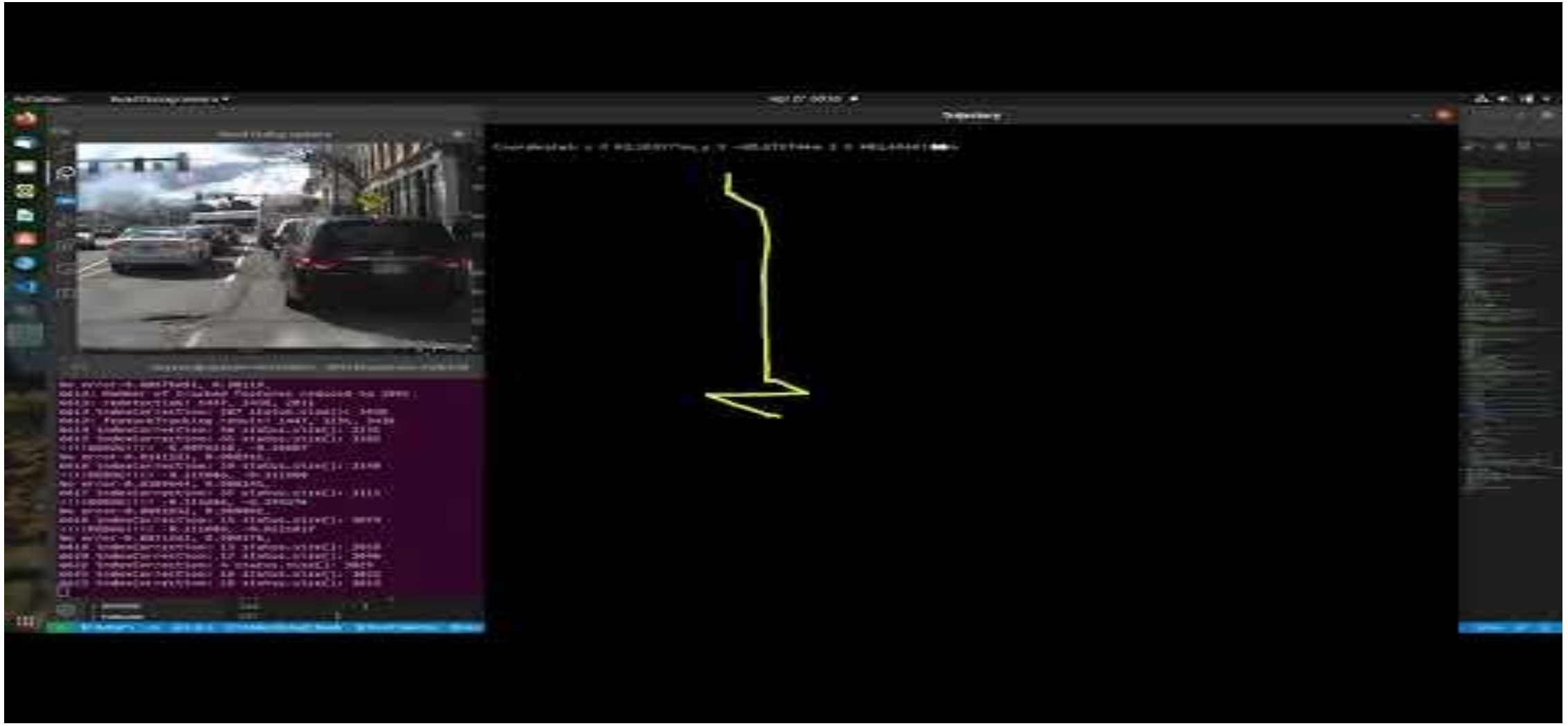
Needs Heuristics : Moving object feature



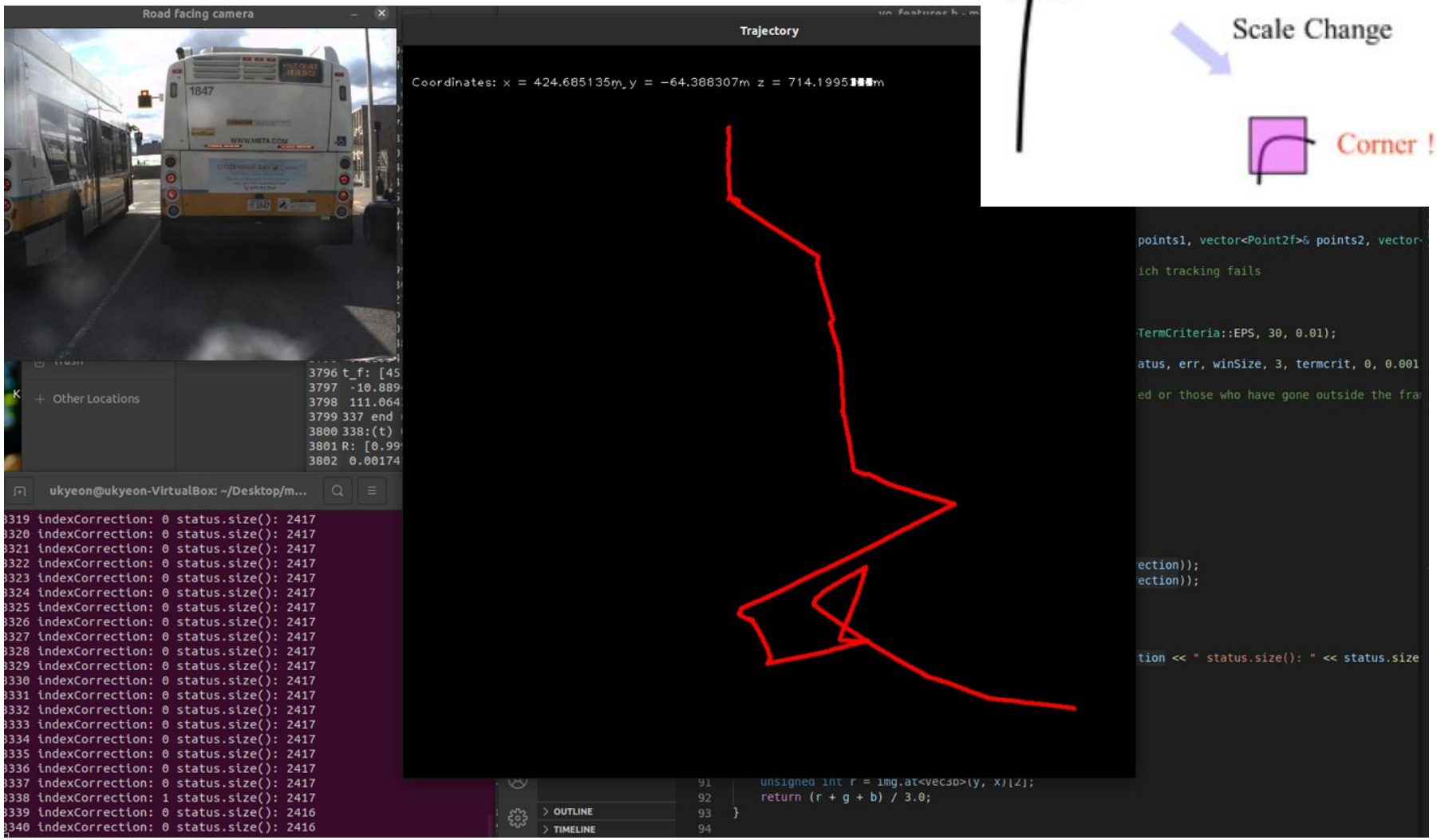
Harris Corner + Partial Gaussian Filtering (Monocular NUance Data)

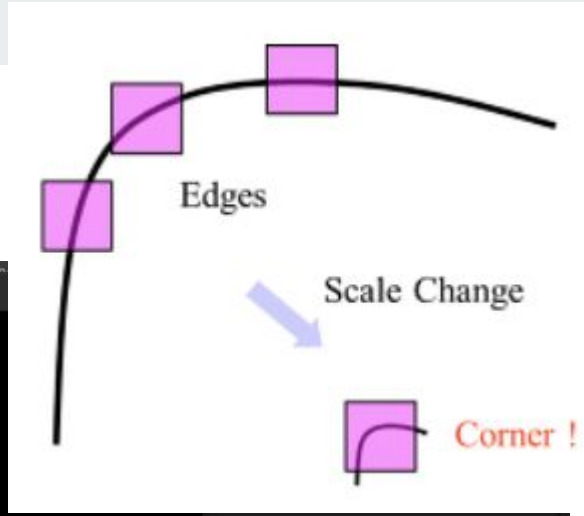


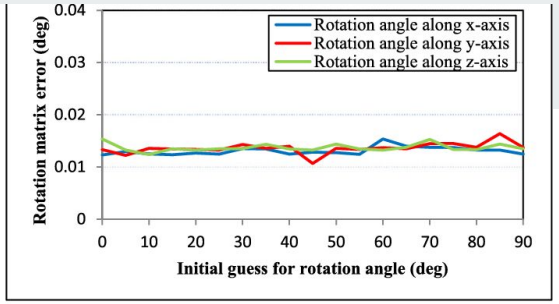
Results after partial Gaussian Filtering



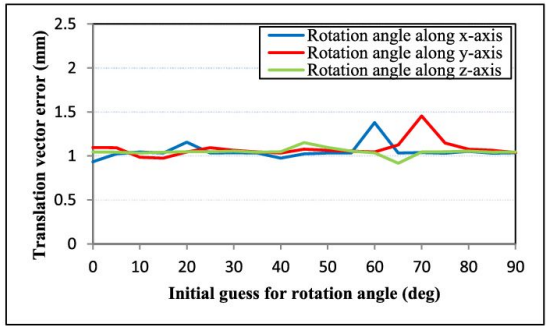
Issues - Harris Corners



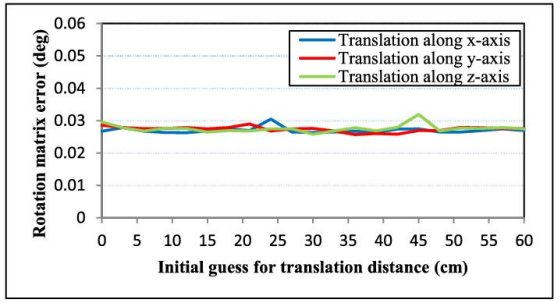




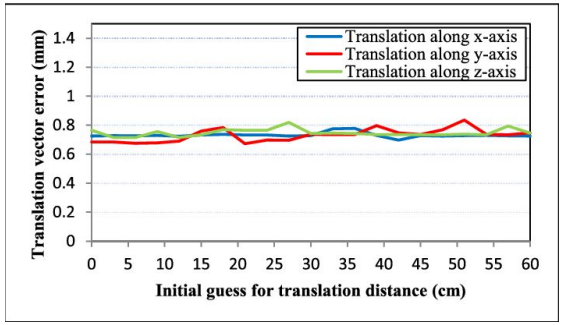
(a)



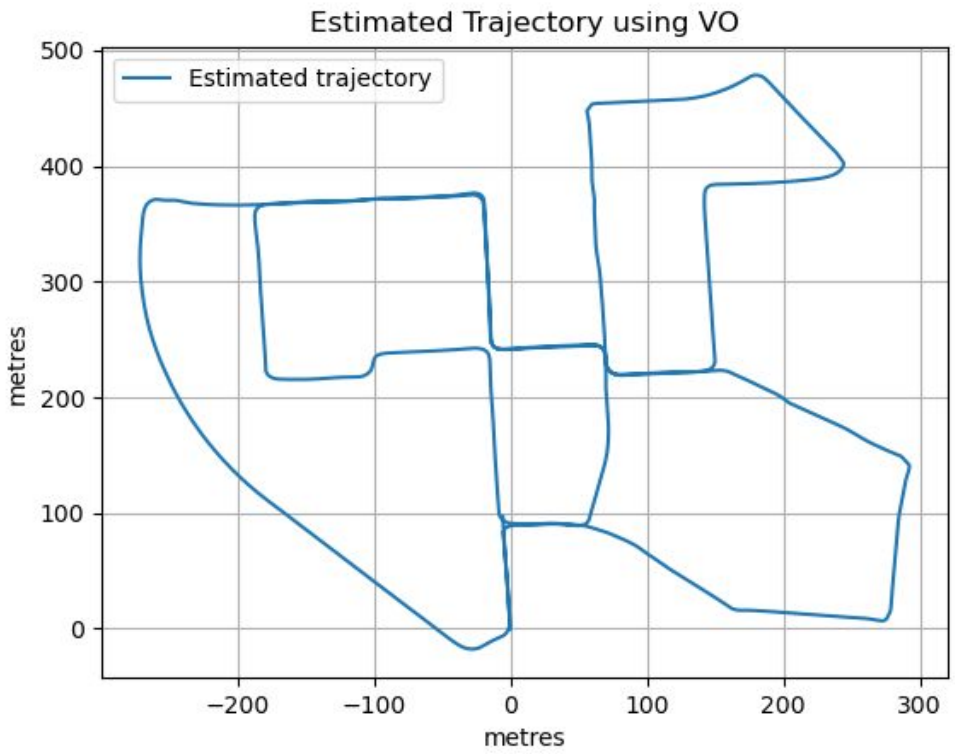
(b)



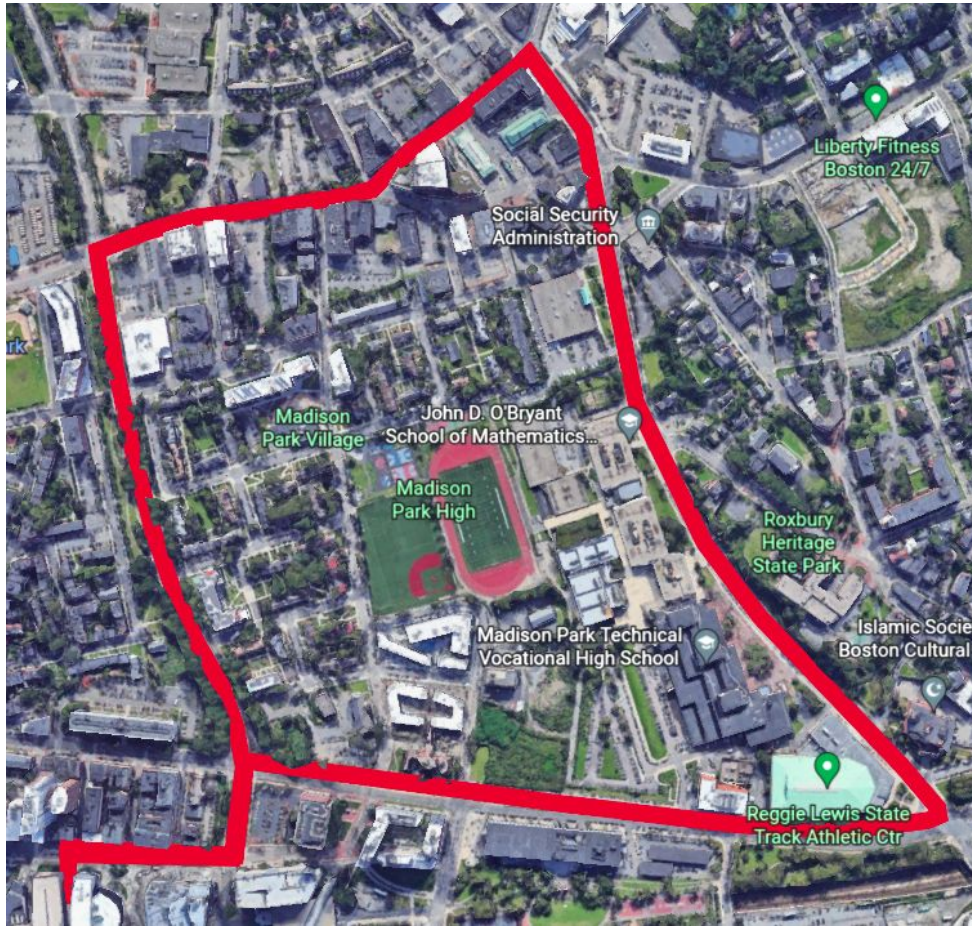
(c)



Results

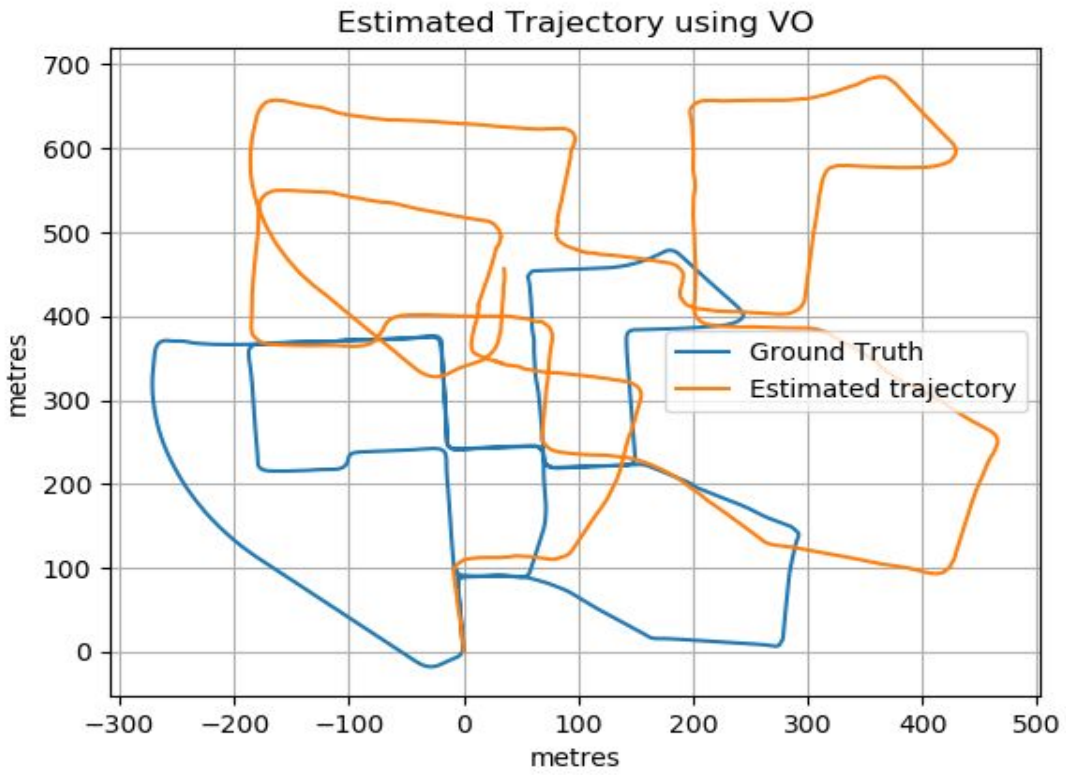


Kitti sequence ground truth trajectory

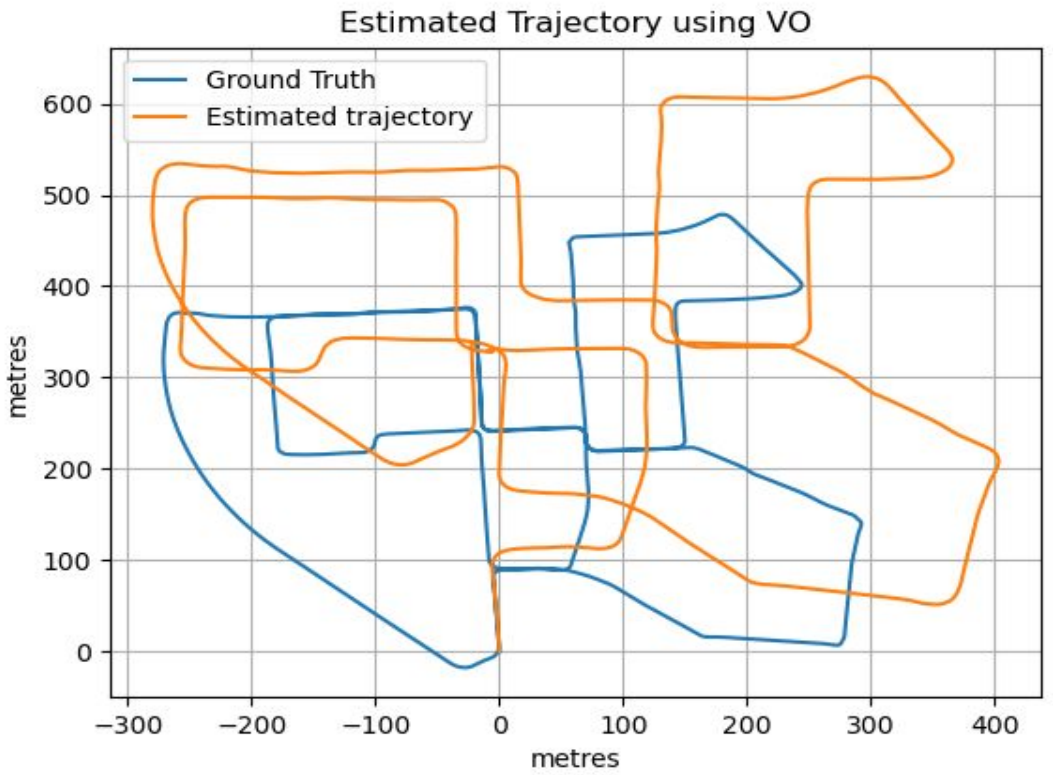


Nuance data ground truth trajectory

Results

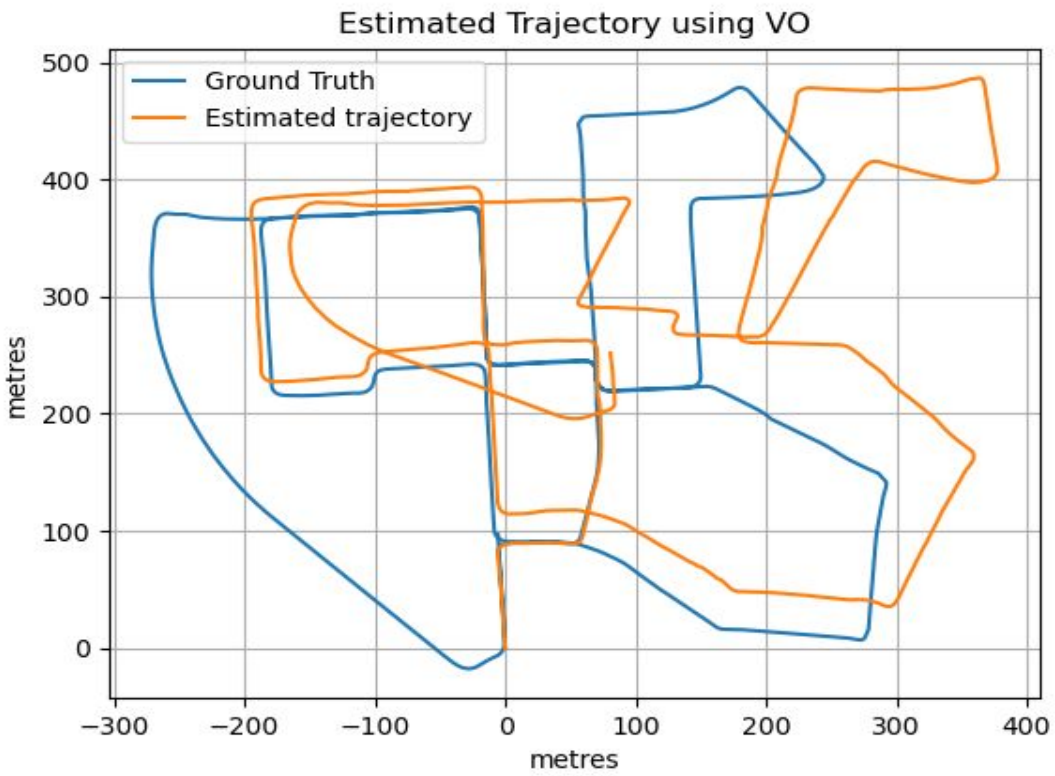


Monocular VO with ORB descriptor
MAE: 109.739
RMSE: 123.751

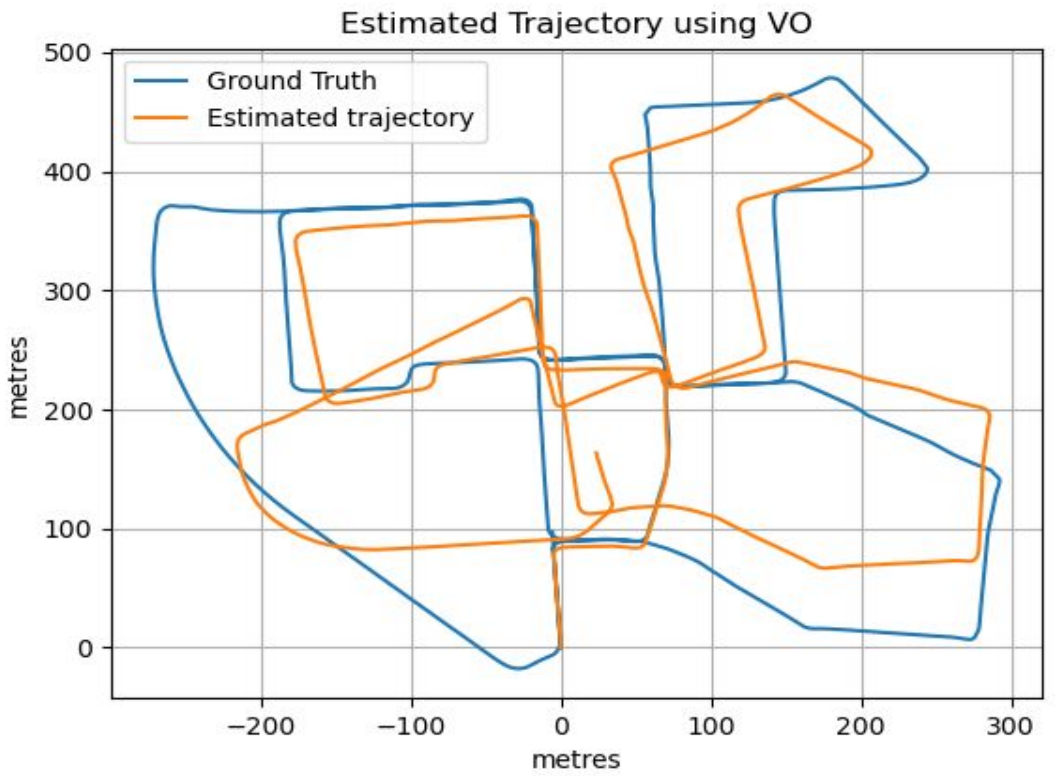


Monocular VO with SIFT descriptor
MAE: 96.819
RMSE: 101.265

Results

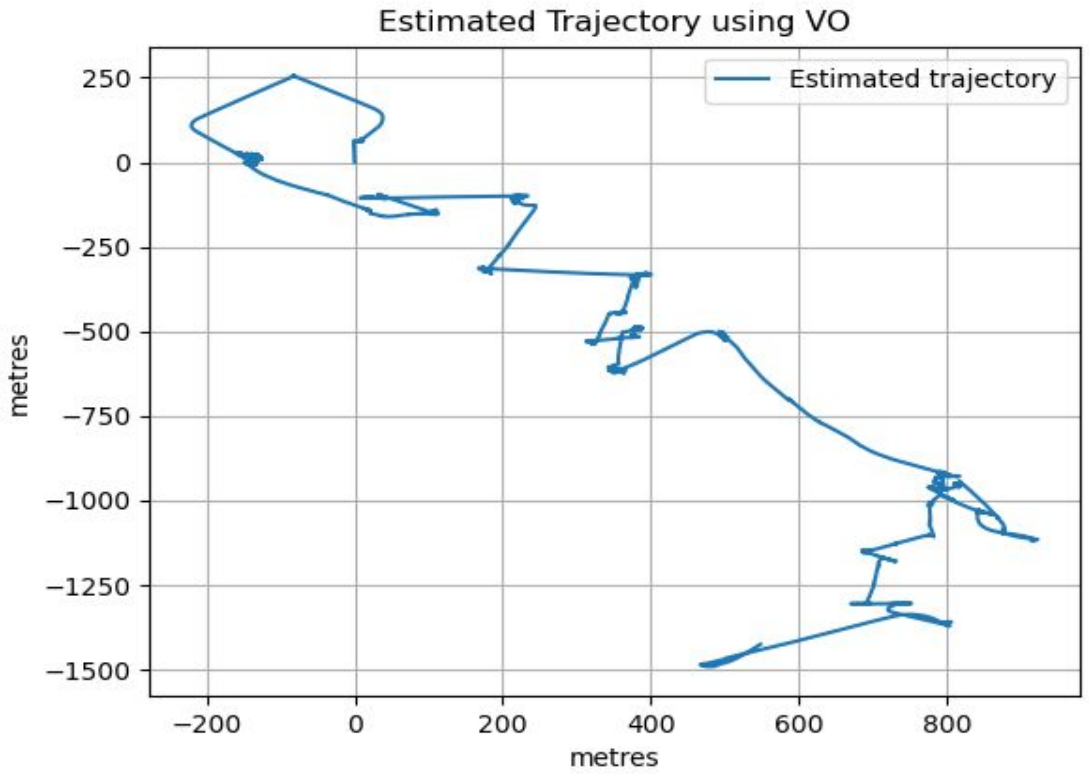


Stereo VO with ORB descriptor
MAE: 65.739
RMSE: 87.751

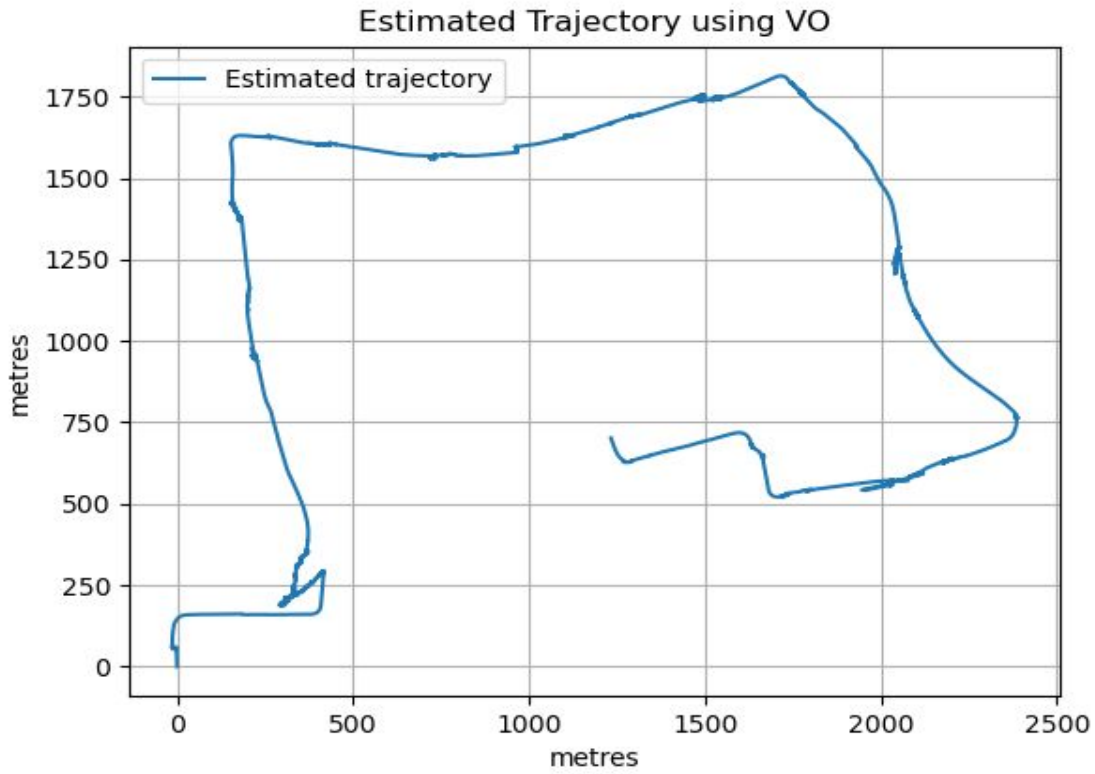


Stereo VO with SIFT descriptor
MAE: 45.739
RMSE: 57.751

Results



Mono VO with ORB descriptor

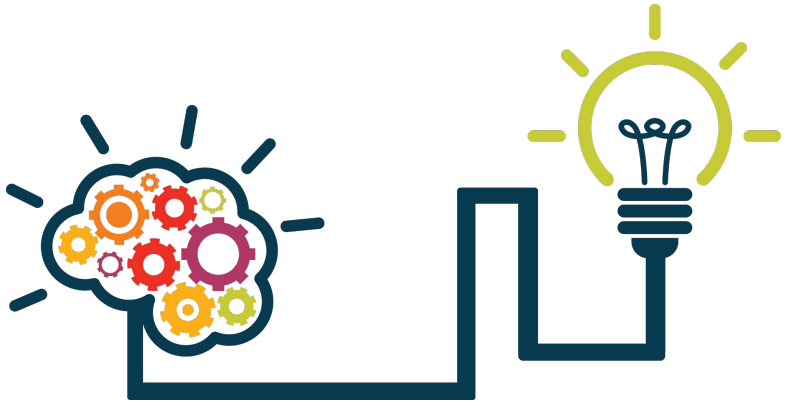


Mono VO with SIFT descriptor

Impact of Backend on VSLAM

Conclusion

- Mono VO suffers from scale ambiguity which can be somewhat corrected with Stereo cameras.
- Stereo VO suffers when the features are too far away. The depth estimate for these points is not correct and thus triangulation fails.
- VO cannot be directly used in real-world systems without any corrective measures or backend optimization.
- VSLAM can be a very good system if tuned properly and with a proper frontend and backend. Though in featureless areas, such as off-road scenarios, the applications are limited.
- Orb SLAM 3 while being a really good system, breaks often on sharp turns which is still a open problem.



Future Work

- Implement Bundle adjustment and keyframe optimization.
- Implement Bag of Words for Loop Closure.
- Test the performance of Optical flow based feature trackers.
- Utilize GPU based implementations for feature matching and bundle adjustment.

Thank You