Augmented Visual Localization Using a Monocular Camera for Autonomous Mobile Robots

Ali Salimzadeh, Neel P. Bhatt, and Ehsan Hashemi

Networked Optimization, Diagnosis, and Estimation (NODE) lab

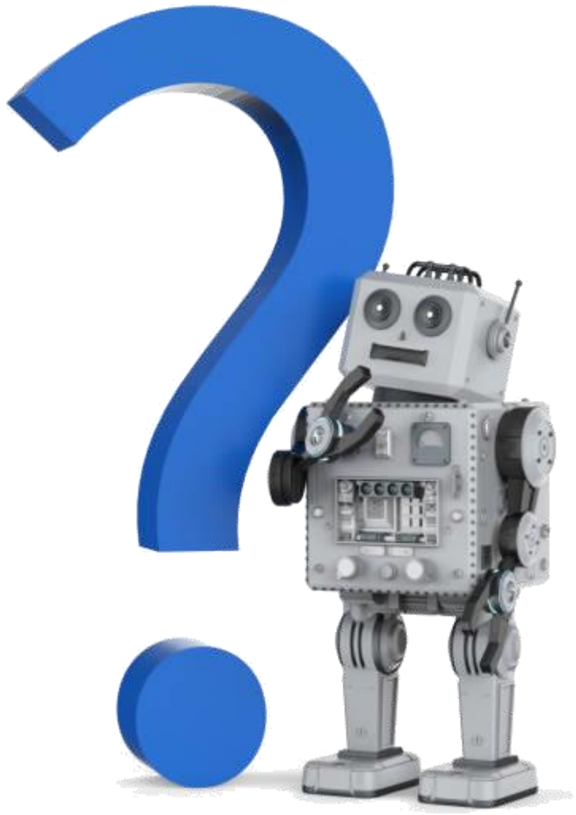
University of Alberta, Edmonton, Canada

August 2022

Automation in Navigation

Where am I?

What should I do?



Infrastructure Aided Localization

Localization with on-board sensors is prone to gradual drift.

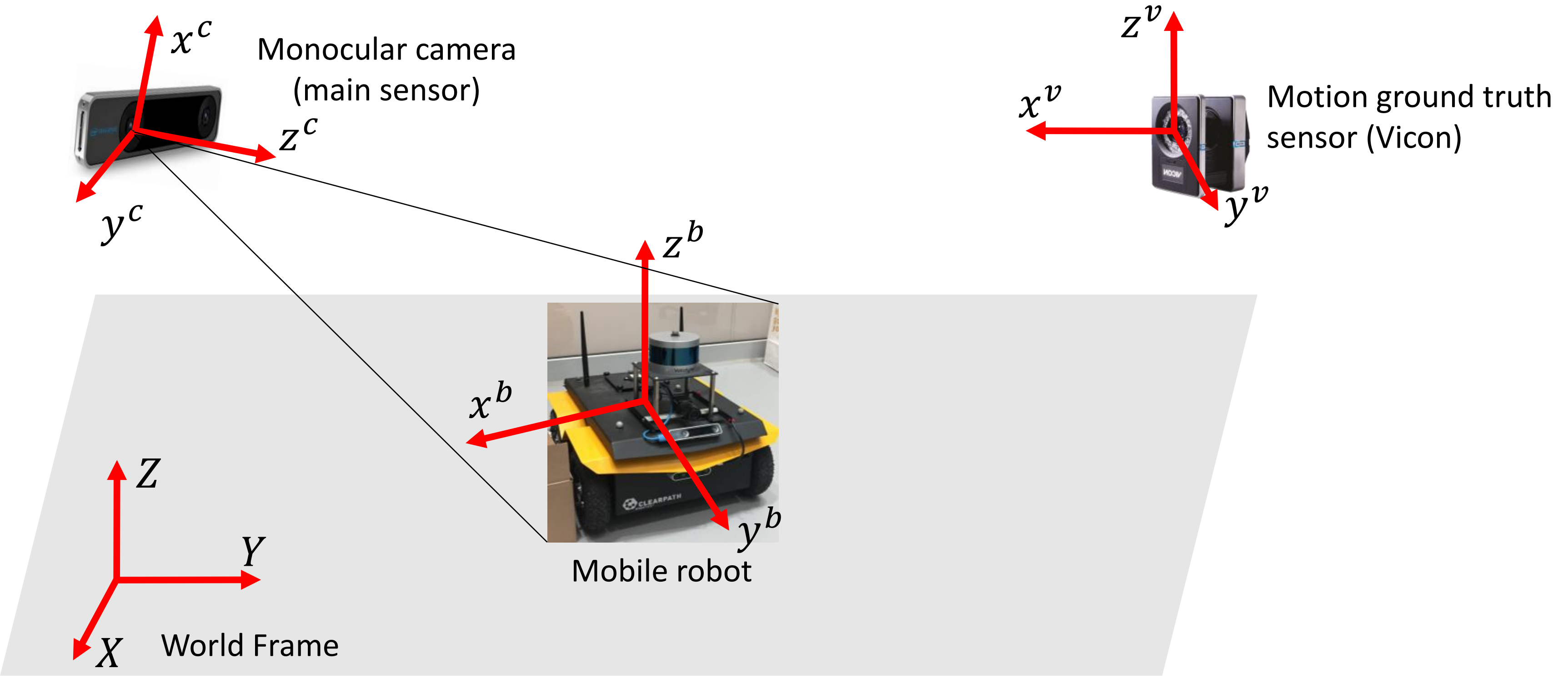
Fixed cameras can improve localization accuracy by communicating with the robot

Applications range from **warehouse or service robotics** to surveillance

Robots in one of JD.com's fully automated warehouses.

Test Setup

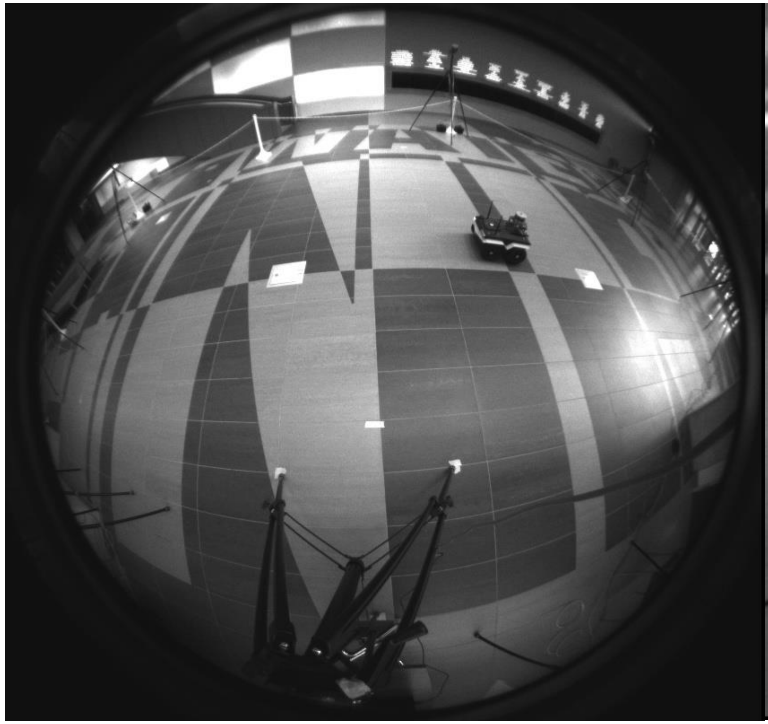
Fixed mono-camera with fish-eye lens (C) Observing the Jackal mobile robot motion capture camera system (V) used for evaluation



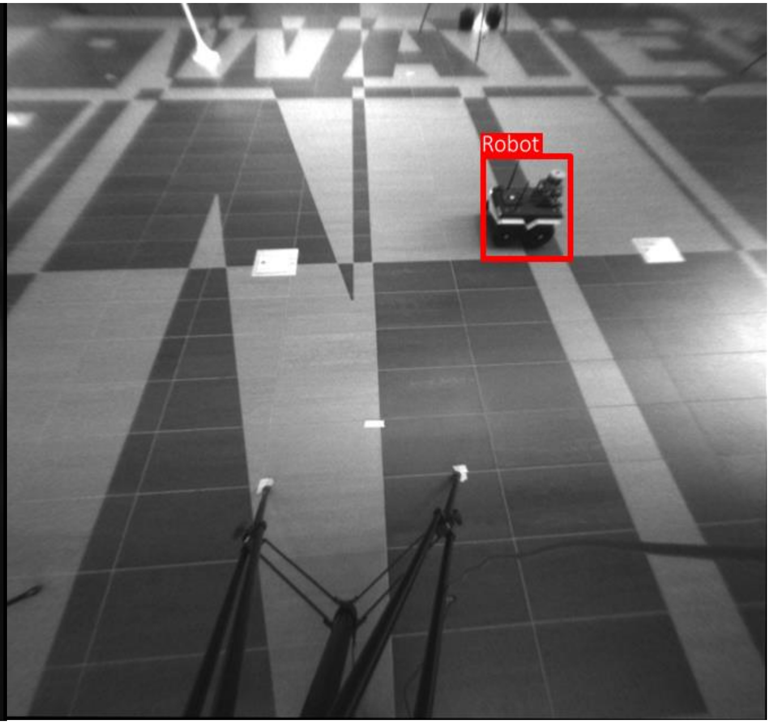
Frame Undistortion and Robot Detection

1. Image un-distortion is carried on frames with a **fish-eye camera model**[1].

2. Robot is detected in the 2D image using YOLOv.4 [2] **object detection** network.



a) Raw image



b) Un-distorted image with YOLO detection

[1]: J. Kannala and S. Brandt, “A generic camera model and calibration method for conventional, wide-angle, and fish-eye lenses,” IEEE transactions on pattern analysis and machine intelligence, vol. 28, pp. 1335–40, 09 2006
[2]: A. Bochkovskiy, C. Wang, and H. M. Liao, “Yolov4: Optimal speed and accuracy of object detection,” CoRR, vol. abs/2004.10934, 2020. [Online]. Available: https://arxiv.org/abs/2004.10934

Depth Perception

MiDas neural network [3] reconstructs a **depth map from monocular frames** for point cloud reprojection.

Advantage: no need for stereo vision or depth sensors (**cost effective**)

a) Monocular image b) Reconstructed depth map

[3]: R. Ranftl, K. Lasinger, D. Hafner, K. Schindler, and V. Koltun, “Towards robust monocular depth estimation: Mixing datasets for zeroshot cross-dataset transfer,” IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI), 2020.

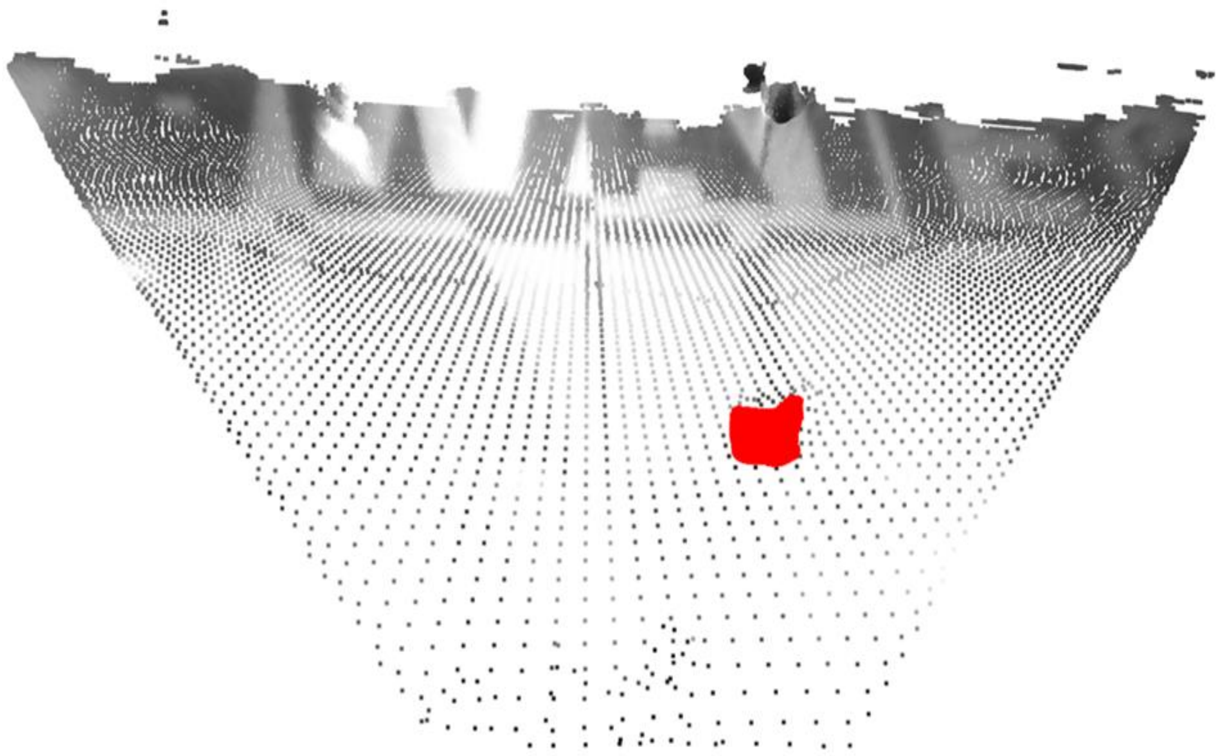
Point cloud Projection and Filtering

Using the frame and depth map, point cloud of the robot is projected with the intrinsic camera matrix.

Filtering:

- Ground is removed by prior knowledge about the environment
- Points are filtered based on their neighborhood density to reject outliers

Centroid of the filtered point cloud is the location measurement



Filtered point cloud showing detection

Uncertain State Estimation Model

Discrete-time uncertain state estimation model has been designed based on a constant acceleration motion model

$$x_{k+1} = Ax_k + Bx_k + \varrho_k$$
$$y_k = Cx_k + \nu_k$$

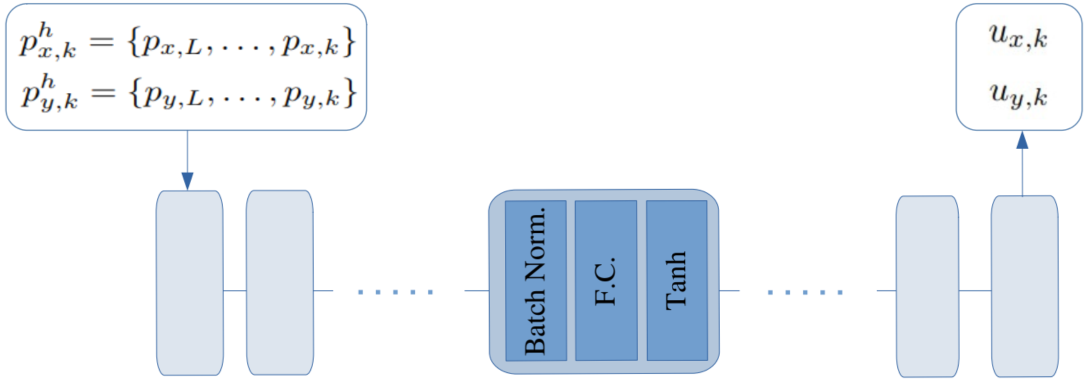
$$A = \begin{bmatrix} 1 & 0 & T_s & 0 \\ 0 & 1 & 0 & T_s \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}, B = \begin{bmatrix} \frac{T_s^2}{2} & 0 \\ 0 & \frac{T_s^2}{2} \\ T_s & 0 \\ 0 & T_s \end{bmatrix}, C = I_{4\times4}, T_s = 100ms$$

ϱ_k and ν_k are process and measurement noises accordingly and are assumed to be independent of each other.

Input Estimation

In order to deal with double derivation noise for input calculation, a deep neural network has been designed to estimate the input to the motion model (linear acceleration)

- Consisted of 15 of **fully-connected** layers.
- Each neuron has a **Tanh** activation function.
- **Batch normalization** has been used to speed up training process and increase the network accuracy.



- Input to this network is a **moving horizon** of location measurement in lateral and longitudinal directions
- Output is the estimated **acceleration** in each of the directions mentioned above

Uncertainty Aware Kalman Filter

To estimate robot states (position and velocity) a Kalman filter is designed which benefits from **adaptive covariance tuning**.

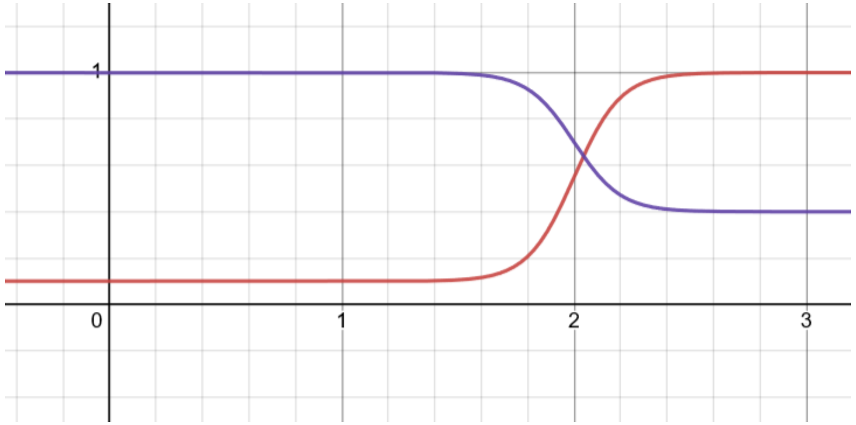
Idea:

- Visual information degrade with depth
- State estimation relies more on the process in greater depth instances rather than the measurement

$$\bar{Q}_k = Q_d \left[\frac{1-\gamma_Q}{2} \tanh\left(s_Q \times \tilde{d}\right) + \frac{1+\gamma_Q}{2} \right]$$

$$\bar{R}_k = R_d \left[\frac{1-\gamma_R}{2} \tanh\left(s_R \times \tilde{d}\right) + \frac{1+\gamma_R}{2} \right]$$

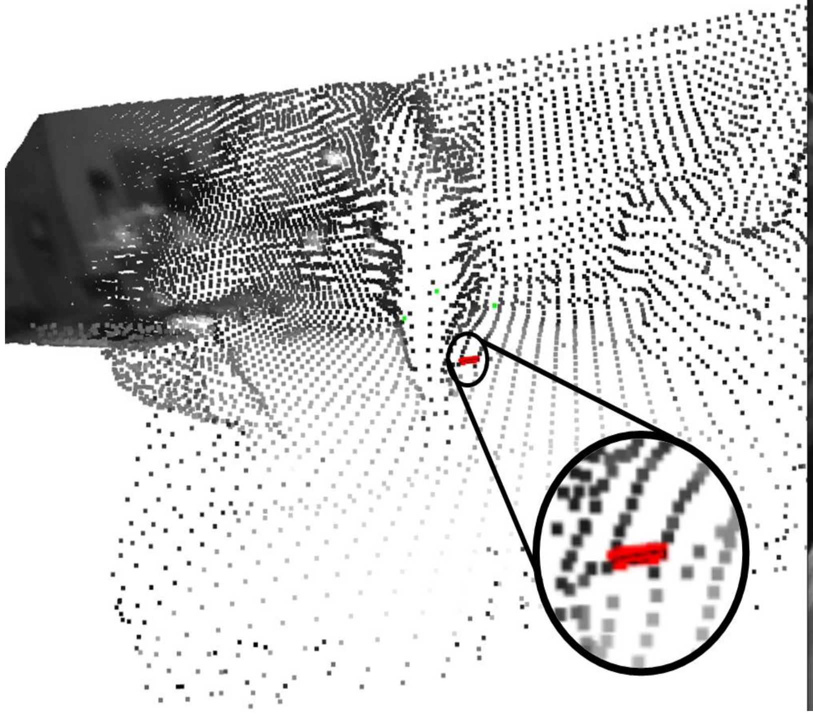
$$\tilde{d} = d_k - \bar{d}$$

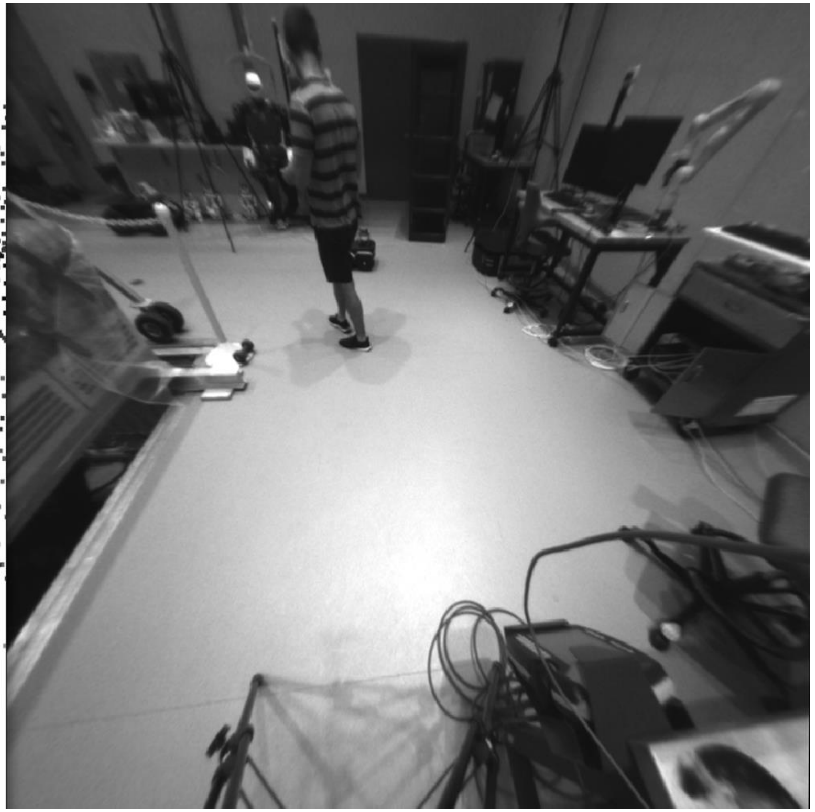


Covariance gain switching example

Detection with Occlusion

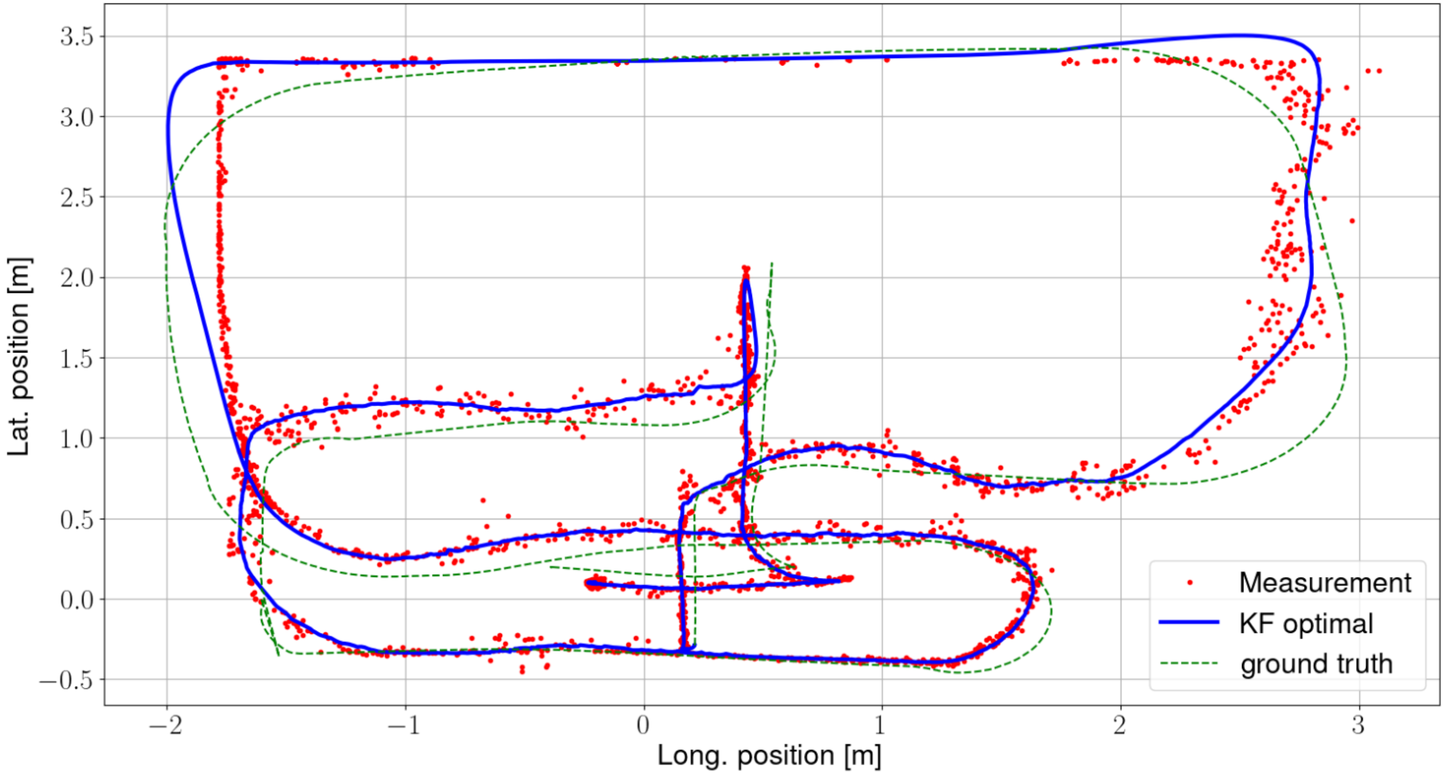
Showcasing the performance of the detection module with **significant occlusion** from human presence in the scene





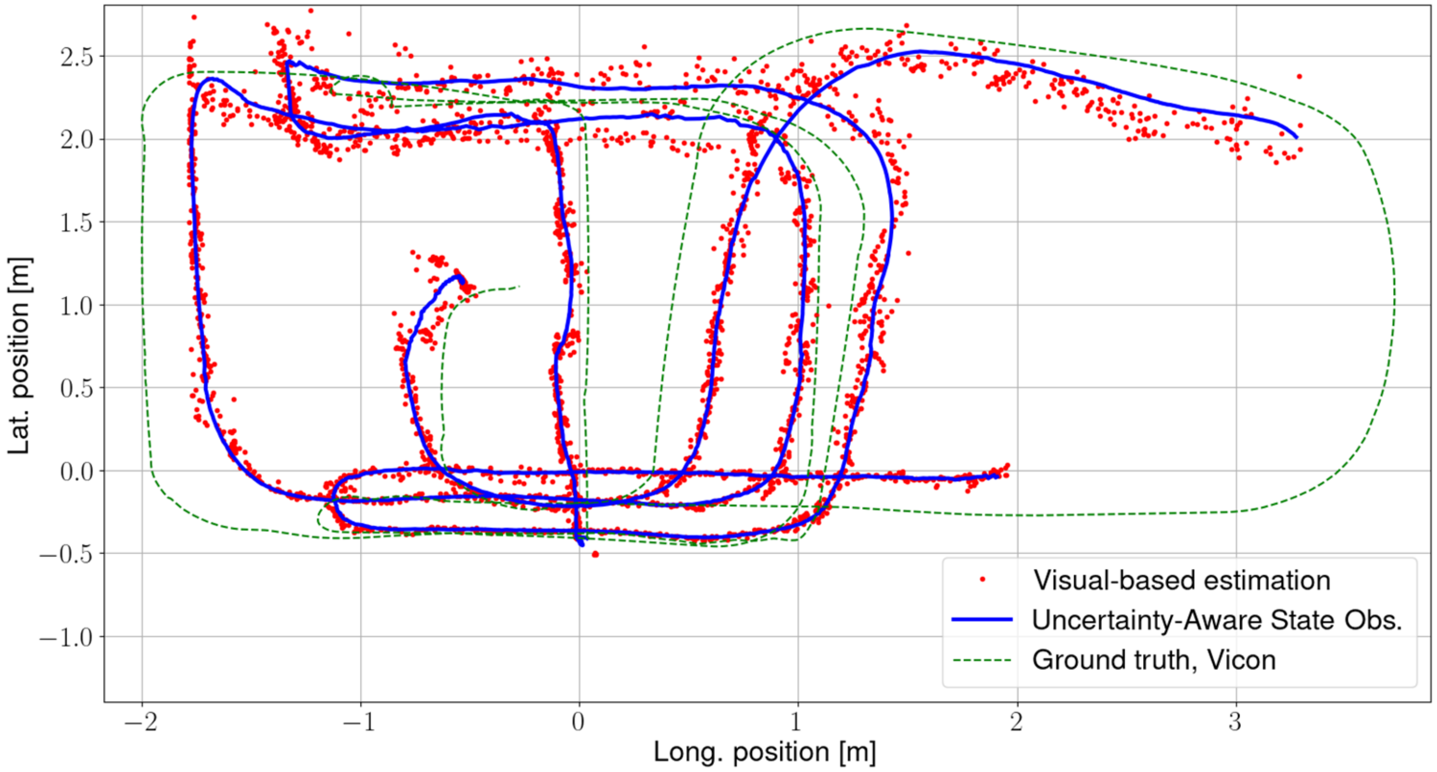
Results

Precise state estimation even when **trajectory is complex**.



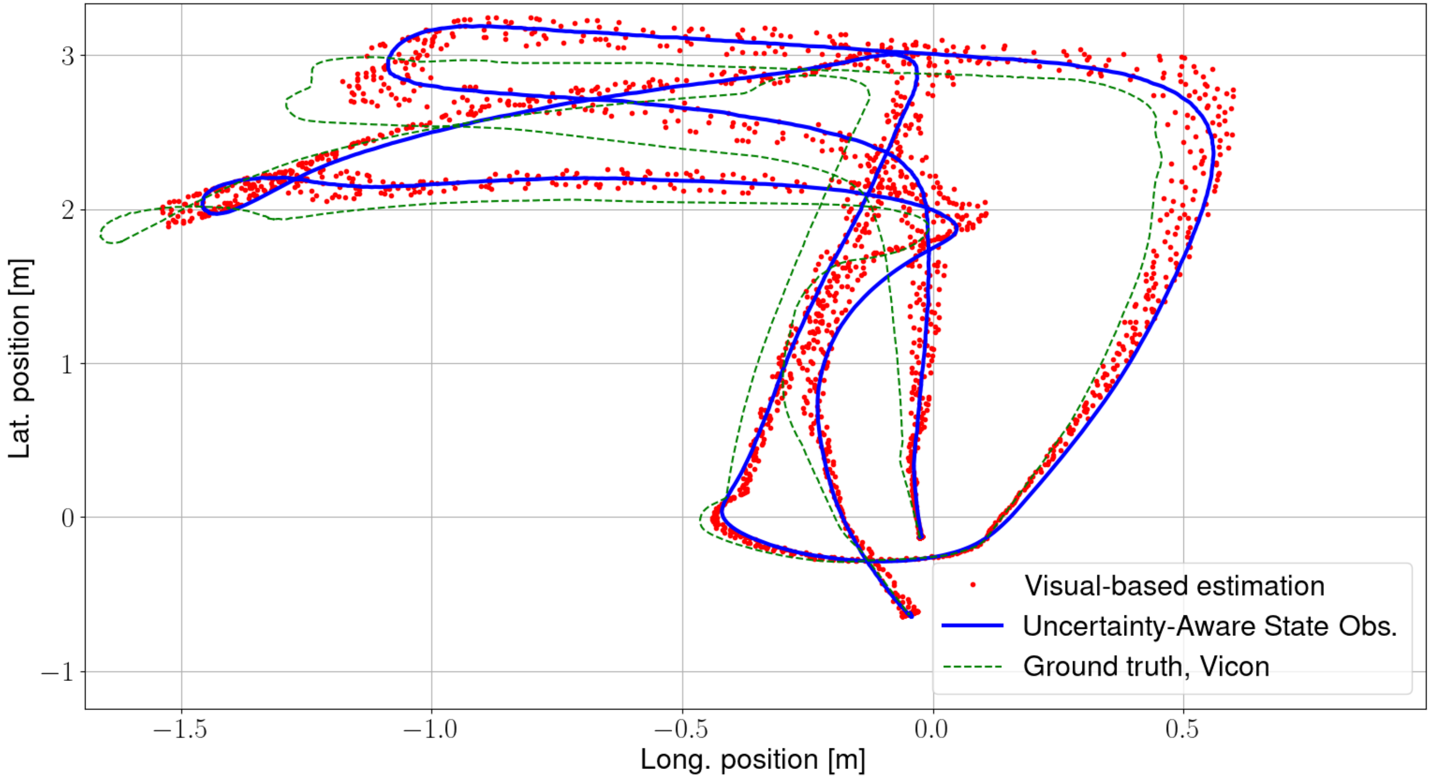
Results

Robot moves **out of the frame**, but proposed method is able to reinitialize state estimation.



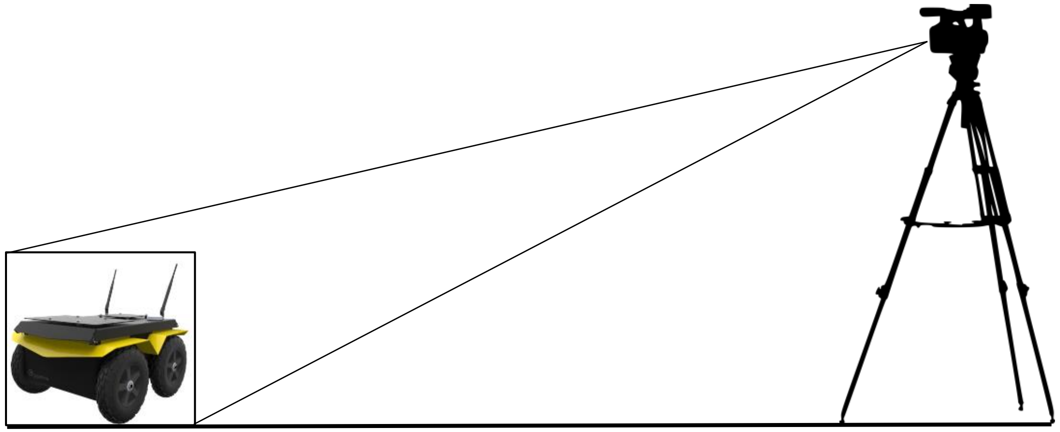
Results

Symbiotic movement of the robot and humans resulting in **sporadic occlusions**.



Conclusion and Future Work

Infrastructure aided mobile robot localization using **fisheye monocular vision** as the sole source of information



Future work:

- Improving computational efficiency for sampling times < 20ms
- Augmenting the motion model to include robot lateral dynamics using sideslip angles for harsh scenarios
- Inclusion of robot heading angle for a more detailed localization

Funding Agencies and Acknowledgements

NSERC CRSNG

The authors would like to acknowledge the technical support of the University of Waterloo's RoboHub

NODE Lab

e-mail: salimzad@ualberta.ca, npbhatt@uwaterloo.ca, ehashemi@ualberta.ca