Multimodal Sensor Fusion in Differentiable Filters Using A Learned Proposal Distribution - Milestone Report

Rabeh, Ali
Technical University of Munich
Munich, Germany
ali.rabeh@tum.de

Gerstewitz, Tim
Technical University of Munich
Munich, Germany
tim.gerstewitz@tum.de

***Abstract*—Traditional Bayesian filters rely on analytical models, thus making them less effective if the underlying transition or measurement models are unknown. To address this limitation, differentiable filters have been introduced as a data-driven extension to Bayesian filters. The aim of this project is to utilize differentiable filters for state estimation of shape poses in pushing scenarios, using the MIT Push dataset as a reference.**

I. Used tools

A. Differentiable Particle Filter

The dynamics of pushing in a real-world setting are inherently discontinuous due to a possible loss of contact. Additionally, results from the literature ([1], [2]) indicate that differentiable particle filters show better performance in pushing scenarios than the family of differentiable Kalman filters. As a result, we choose a differentiable particle filter as the filter of choice within this project and have so far implemented a simplistic but working version from scratch in PyTorch.

B. MIT Push Dataset

For evaluating the filter, we use the MIT Push dataset [4], which consists of over a million individual pushing sequences of eleven different-shaped objects pushed across four different surface materials recorded with a sampling rate of 250Hz. Each pushing sequence contains recordings of the pushed object's pose (x, y and orientation) as well as the end effector pose of the robotic pusher and its forces and torques used during the push. Additionally, the authors of [1] have published a script for rendering synthetic gray scale images of the scene. So far, we have used all signals except the synthetic images in our project.

II. First Experiment

A. Experimental Setup

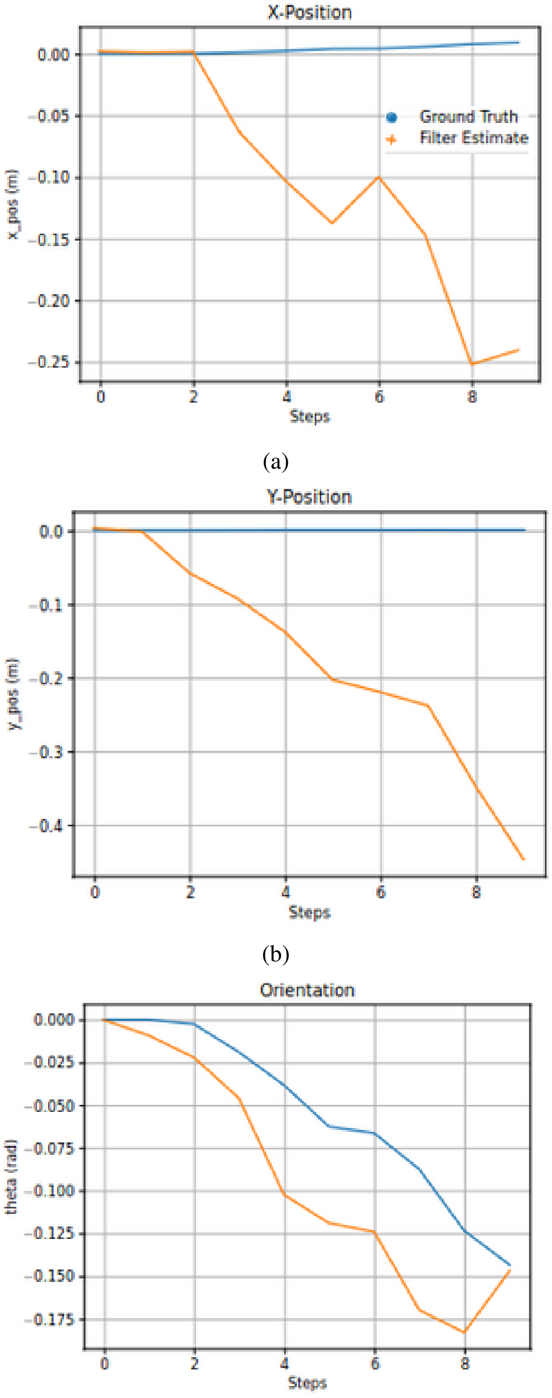
For our first experiment, we used a subset of the MIT Push dataset: we included only pushing sequences in which a small metallic rectangle was pushed across plywood with a constant speed of 100mm/s. This left us with 20 individual sequences for training and 12 sequences for validation and testing each. All signals were downsampled to 50Hz. The filter we used features a forward model consisting of two hidden layers and ReLU type activation functions, with the pose at the previous timestep and the current forces and torques chosen as inputs. Our current observation model receives the state prediction by this forward model as an input, together with the current end-effector pose and again forces and torques. The model here consists of three hidden layers with ReLU activation functions and a sigmoid as the final layer to map the resulting outputs into the $[0, 1]$ range. We trained our filter end-to-end with no pretraining and no resampling over 400 epochs, using an L2-Loss. The training objective was set as minimizing the estimation error over a single timestep into the future. During both training and testing, 500 particles were used.

B. Results and Discussion

We show the results of testing sequence 10 in Fig. 1 over the ten initial timesteps. While the filter estimate stays close to the ground truth over a single timestep (which we expect after training), both the x and y position estimates diverge rapidly over the subsequent steps, as can be seen in Fig. 1a and Fig. 1b. This is plausible since we only trained for minimizing single-step estimation errors. We therefore expect the filter estimates to become more consistent and accurate when switching to minimizing multiple-step errors during training. Additional performance boosts might come from introducing resampling, which we have so far neglected, and normalizing the states while propagating them through the filter. Still, our filter seems to track the orientation quite well already, as can be seen in Fig. 1c.

III. Next Steps

In the following milestone, we intend to explore several additional aspects to improve the capabilities of our model for pose state estimation. Firstly, we intend to incorporate resampling techniques and to pretrain the motion model to enhance the accuracy of predictions over multiple steps. Secondly, we aim to extend the applicability of our model by incorporating image data into the training data. Moreover, we plan to investigate the potential of using differentiable filters to estimate physical parameters of the system, such as the coefficient of friction. Furthermore, we will conduct a comparison between differentiable filters and recurrent neural networks (RNNs). We will assess the advantages and disadvantages of each approach based on the outcomes.



(a)

(b)

(c)

Fig. 1: Estimated pose versus ground truth pose of testing sequence 10 over ten initial timesteps.

References

[1] A. Kloss, G. Martius, and J. Bohg, “How to Train Your Differentiable Filter”, in Autonomous Robots, Vol. 45, 2021, pp. 562—578.

[2] M. A. Lee, B. Yi, R. Martin-Martin. S. Salvarese, and J. Bohg, “Multimodal Sensor Fusion with Differentiable Filters”, in 2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), 2020, pp. 10444–10451.

[3] L. Roestel, L. Sievers, J. Pitz, and B. Baeuml, “Learning a State Estimator for Tactile In-Hand Manipulation”, in 2022 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), 2022, pp. 4749–4756.

[4] K. T. Yu, M. Bauza, N. Fazeli, and A. Rodriguez, “More than a Million Ways to Be Pushed: A High-Fidelity Experimental Data Set of Planar Pushing”, in 2016 IEEE/RSJ Internation Conference on Intelligent Robots and Systems (IROS), 2016, pp.30–37.