Multimodal Sensor Fusion in Differentiable Filters Using A Learned Proposal Distribution

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

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

I. Objective

Recently, differentiable filters have been proposed as a data-driven expansion to classical Bayesian filters [1]. Since such classical filters rely on analytical models, this is especially useful when the transition or measurement models of the filtered process are not known exactly. The approach has demonstrated impressive performance: on various state estimation tasks, it has been shown to outperform unstructured LSTMs [2].

For differentiable filters (DF), [2] has recently investigated different methods for multimodal sensor fusion based on a real-world pushing dataset [4]. In [3] on the other hand, a method to improve the efficiency of a DF for state estimation using a learned proposal distribution was presented and succesfully validated in an in-hand manipulation experiment employing only tactile sensors.

The goal of this work is to combine the modified DF in [3] with the pushing scenario in [2]. Specifically, we want to investigate the modified DFs' performance using a learned proposal distribution incorporating multimodal information on the MIT Push dataset [4] compared to the standard DF used in [2].

II. Related Work

A. How to Train Your Differentiable Filter

[1] provides practical guidance for applying DFs and compare them to manually-tuned filtering algorithms. They implemented 4 type of DFs with different filtering algorithms. In addition, they implemented two novel DFs based on the unscented Kalman filter.

B. Multimodal Sensor Fusion in DFs

In [2], fusion of state estimates from DFs using tactile, proprioceptive and visual information is investigated on a real-world pushing dataset [4]. Specifically, the authors look at three architectures to fuse the estimates of unimodal DFs into a multimodal one and show that this outperforms any of the unimodal estimates. However, they do not consider the case of a single, multimodal DF.

C. Learned Proposal Distributions for pose estimation in 3D with DFs

[3] recently proposed a modified DF algorithm which replaces the usual motion model for prediction in DFs with a learned proposal distribution incorporating the most recent measurement. They show that this helps mitigate particle degeneracy. They also modify the DF algorithm to work on the SE(3) Lie-Group for pose estimation in 3D. In an in-hand manipulation experiment, it is shown that a DF incorporating these changes outperforms an unmodified DF. However, the authors do not investigate fusing multimodal sensor data in the learned proposal distribution.

III. Technical Outline

In Robotics state estimation is crucial in some scenarios such as grasping and motion planning. For adapting the Deep Differentiable Proposal Particle Filter (D2P2F) from [3] to a pushing scenario, we therefore intend to do the following:

We plan to modify the proposal distribution (represented through a feed-forward neural network) such that it can incorporate the tactile, proprioceptive and visual information the MIT Push dataset [4] provides. Since pushing is mainly a 2D task, we also have to modify the algorithm to work on SE(2) instead of SE(3). These two changes might offer a more principled alternative to the fusion methods presented in [2]. Our results on the MIT dataset will lastly be compared to that of a recurrent neural network in the last part of the project to investigate the advantages and disadvantages between them.

A. Milestones

Hence, we envision the following milestones for our project:

- Implement a differentiable filter for estimating the state of a pushed object.
- Adapt the algorithm presented in [3] from 3D state estimation to the 2D pushing scenario on SE(2).
- Fuse tactile, proprioceptive and visual measurements in the learned proposal distribution in a novel way such as in a single network.
- Compare the modified algorithm's results to a recurrent neural network and show advantages and disadvantages based on the results.

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.