Progress Report

Bardia Mojra

June 6, 2022

Robotic Vision Lab

The University of Texas at Arlington

1 Specific Research Goals

- VPQEKF (May 30th): Work on the paper.
- DLO Manipulation Dataset (ICRA - Sept. 1st)

2 To Do

- QEKF Paper - 30% extension (June 30th):
  - Read and summarize recent QEKF papers.
- QEKF/QuEst+VEst Implementation (May 30th):
  - Feature point extraction: implement semantic segmentation
  - Address scale factor (depth-scale) issues: DL solutions?
  - Address "hand off" issue when objects enter or leave field of view
  - Experiments - On-going
  - Noise issue: noise cannot be modeled - revisit
  - Adding plots - On-going
  - Add semantic segmentation for detecting moving object pixels and rejecting matched features in those regions
- DLO Manipulation: ICRA - Sept. 1st
  - Find other ICRA dataset papers and summarize the structure. — On-going.
  - Dynamic Dataset Collection System with Reinforcement Learning:
    * Design dynamic DLO data collection system.
    * Build work cell.
    * Collect data and create a dataset.
    * Define evaluation metrics.
    * Create a high frequency RGBD dataset with UV-frames and open-loop input control actions as the ground truth.
  - Real-Time Preception
    * Deep learning methods for keypoint pose estimation in real-time.

        * Use UV dye dataset
        * Use PVNet-based approach for known-objects
    - Learning DLO Dynamics and System Identification
        * List feasible approached for learing DLO dynamics
        * Model dynamics and deformity in a latent space
    - Real-Time Control
        * Time model inference, using auto-encoders generate the lowest dimensional representation for each object.
        * Use another GAN model for object deformity for each object.
        * Evaluate encoded representation for accuracy.
        * Used another GAN to explore other abstraced representations from individual encoded representation. In theory, we can create a low dimensionsal representation for multiple similar objects, given all individual low-dimensional representations. This is inspired by "fundamental principles first" approach which has universal applicability.

3 Progress

The following items are listed in the order of priority:

- XEst (RAL - April 30st, 2022): I met with Dr. Gans and we went over the XEst results. I made the modules for pose and velocity estimations more distinct. Although the QuEst and VEst estimation methods are independent methods, I named pose and velocity modules QuEst+ and VEst+ respectively. QuEst+ and VEst+ modules can contain any number of pose and velocity estimation modules which are the used by a corresponding QEKF filter. Right now, the QEKF filter instances only correspond to selected pose estimation methods but additional instances could be spawned for different velocity estimation methods on demand. This feature allows for benchmarking any methods and not just limited to pose and velocity estimations. Below are the results for QuEst+, VEst+, and QEKF modules. The entire system is called *XEst* with *X* referring modularity of the QEKF estimation system. Note that there are two types of error, GT-X and Z-X represent the groundtruth vs. the state error and the observation vs. the state, respectively. It is clear to me that there are some issues with computation so I am plotting the data logs to debug them. I need to focus on this and just wrap it up. I will spend this week on reading RAL EKF papers and make come up with a list of contributions. I am confident I can wrap this up quickly. Moreover, we could follow up on this paper with a double-Quaternion formulation of the same algorithms. It would be very simple and straight forwards paper, perhaps a journal paper.

	EightPt	Nister	Kukelova	QuEst	VEst
T err mean	0.0492	0.1345	0.1495	0.0613	0.0382
T err std	0.0484	0.1163	0.1391	0.0656	0.0411
T err med	0.0145	0.0492	0.0535	0.0118	0.0099
T err Q1	0.0108	0.0405	0.0326	0.0073	0.0064
T err Q3	0.0950	0.2605	0.3001	0.1346	0.0733
Q err mean	0.0636	0.0060	0.0139	0.0032	0.0058
Q err std	0.0912	0.0069	0.0181	0.0031	0.0078
Q err med	0.0029	0.0020	0.0021	0.0013	0.0015
Q err Q1	0.0022	0.0009	0.0008	0.0007	0.0009
Q err Q3	0.1153	0.0104	0.0260	0.0059	0.0091

Figure 1: QuEst+ module output error statistics on KITTI dataset, running with EightPoint, Nister, Kukelova, QuEst and VEst algorithms for pose estimation.

	VEst
T err mean	0.5672
T err std	0.4587
T err med	0.8913
T err Q1	0.0083
T err Q3	0.9521
Q err mean	0.0058
Q err std	0.0078
Q err med	0.0015
Q err Q1	0.0009
Q err Q3	0.0091

Figure 2: VEst+ module output error statistics on KITTI dataset, running only with the VEst algorithm for velocity estimation.

	EightPt	Nister	Kukelova	QuEst	VEst
GT-X T err mean	0.0594	0.1444	0.1947	0.0785	0.0458
GT-X T err std	0.0616	0.1296	0.1609	0.0860	0.0436
GT-X T err med	0.0195	0.0492	0.1726	0.0134	0.0197
GT-X T err Q1	0.0104	0.0385	0.0225	0.0072	0.0093
GT-X T err Q3	0.1111	0.2719	0.3666	0.1771	0.0887
GT-X Q err mean	0.3334	0.3334	0.3334	0.3334	0.3334
GT-X Q err std	0.0007	0.0007	0.0007	0.0007	0.0007
GT-X Q err med	0.3335	0.3335	0.3335	0.3335	0.3335
GT-X Q err Q1	0.3326	0.3326	0.3326	0.3326	0.3326
GT-X Q err Q3	0.3341	0.3341	0.3341	0.3341	0.3341
GT-X V err mean	0.5634	0.5798	0.5846	0.5517	0.5586
GT-X V err std	0.4501	0.4137	0.3786	0.4437	0.4506
GT-X V err med	0.8535	0.8706	0.8348	0.8523	0.8852
GT-X V err Q1	0.0204	0.1048	0.1982	0.0128	0.0100
GT-X V err Q3	0.9541	0.9150	0.8974	0.9421	0.9303
Z-X T L1 mean	0.4606	1.0581	1.3635	0.5910	0.4025
Z-X T L1 std	0.3965	0.8347	0.5471	0.4974	0.3322
Z-X T L1 med	0.3204	0.8213	1.3703	0.6247	0.2590
Z-X T L1 Q1	0.0711	0.2621	0.9690	0.0333	0.1857
Z-X T L1 Q3	0.8714	1.9459	1.7082	1.0772	0.6076
Z-X Q L1 mean	1.1693	1.0330	1.0719	1.0158	1.0222
Z-X Q L1 std	0.2295	0.0357	0.0918	0.0137	0.0297
Z-X Q L1 med	1.0152	1.0108	1.0108	1.0070	1.0046
Z-X Q L1 Q1	1.0107	1.0056	1.0054	1.0051	1.0034
Z-X Q L1 Q3	1.3101	1.0591	1.1331	1.0268	1.0357
Z-X V L1 mean	0.0488	0.0632	0.0590	0.0520	0.0516
Z-X V L1 std	0.0417	0.0555	0.0575	0.0424	0.0421
Z-X V L1 med	0.0243	0.0480	0.0305	0.0243	0.0289
Z-X V L1 Q1	0.0211	0.0238	0.0229	0.0214	0.0229
Z-X V L1 Q3	0.0697	0.0821	0.0798	0.0809	0.0734
Z-X T L2 mean	0.3338	0.7324	1.3257	0.3978	0.2525
Z-X T L2 std	0.3969	0.7806	0.9242	0.4305	0.3772
Z-X T L2 med	0.0844	0.3607	1.1590	0.1548	0.0566
Z-X T L2 Q1	0.0022	0.0280	0.8175	0.0005	0.0358
Z-X T L2 Q3	0.6853	1.4451	1.6334	0.8748	0.3684
Z-X Q L2 mean	1.0000	1.0000	1.0000	1.0000	1.0000
Z-X Q L2 std	0.0000	0.0000	0.0000	0.0000	0.0000
Z-X Q L2 med	1.0000	1.0000	1.0000	1.0000	1.0000
Z-X Q L2 Q1	1.0000	1.0000	1.0000	1.0000	1.0000
Z-X Q L2 Q3	1.0000	1.0000	1.0000	1.0000	1.0000
Z-X V L2 mean	0.0033	0.0044	0.0039	0.0034	0.0034
Z-X V L2 std	0.0054	0.0068	0.0065	0.0053	0.0053
Z-X V L2 med	0.0005	0.0014	0.0005	0.0005	0.0005
Z-X V L2 Q1	0.0003	0.0005	0.0004	0.0003	0.0004
Z-X V L2 Q3	0.0046	0.0057	0.0050	0.0050	0.0048

Figure 3: QEKF module output error statistics on KITTI dataset, running with various pose estimation algorithms.

- XEst - Semantic segmentation (RAL - April 30st, 2022): I found [1] for semantic segmentation of moving cars and pedestrians. The matched feature points that are labelled as moving object will be removed from pose and velocity estimation methods as their inclusion results in increased error.

- DLO Dataset: There are number of recent papers that have made significant contribution to real-time manipulation of DLOs, [2], [3], [4], and [5]. Based on my understanding at the moment, the are two main approaches for real-time manipulation of DLO, quasi-static and non-quasi-static approaches. Quasi-static modeling of a DLO defines our initial approach where we aimed to model the object in form of a set of dynamic segments with locally linear dynamic behavior. Non-quasi-static approaches define the system as inherently non-linear and aim to model the system in a non-linear and often latent space. The two are two completely different regimes of mathematical analysis and the manipulation task of DLOs the latter is the superior approach. MuJuCo is the only state of the art contact-based physics engine that was designed from ground up for this specific purpose. In addition, it has built-in API with Unity rendering engine which also features API for digital twin, business intelligence and ROS. Please see attachments for more detailed analysis. Moreover, for dynamic dataset collection, we need a high frequency auto-annotation system to capture accurate dynamics, essentially with every other frame as a UV ground truth frame. I can design and build a simple embedded switcher circuit for automating that. But this has to be after the XEst paper, or after I finish the initial draft.

- DLO Control: No update.

- DLO Perception: No update. – might not be needed.

- Grasping Project (DLO-03): I am making this a part of the DLO project.

- PyTorch Tutorials: Transfer learning.

4 Intermediate Goals - Fall 2021:

- QEKF: Finish paper.

- UR5e: Do the tutorials.

References

[1] I. Ballester, A. Fontan, J. Civera, K. H. Strobl, and R. Triebel, "Dot: dynamic object tracking for visual slam," in *2021 IEEE International Conference on Robotics and Automation (ICRA)*, pp. 11705–11711, IEEE, 2021.

[2] W. Zhang, K. Schmeckpeper, P. Chaudhari, and K. Daniilidis, "Deformable linear object prediction using locally linear latent dynamics," in *2021 IEEE International Conference on Robotics and Automation (ICRA)*, pp. 13503–13509, IEEE, 2021.

[3] H. Zhang, J. Ichnowski, D. Seita, J. Wang, H. Huang, and K. Goldberg, "Robots of the lost arc: Self-supervised learning to dynamically manipulate fixed-endpoint cables," *arXiv preprint arXiv:2011.04840*, 2020.

[4] Y. Narang, B. Sundaralingam, M. Macklin, A. Mousavian, and D. Fox, "Sim-to-real for robotic tactile sensing via physics-based simulation and learned latent projections," in *2021 IEEE International Conference on Robotics and Automation (ICRA)*, pp. 6444–6451, IEEE, 2021.

[5] R. Laezza, R. Gieselmann, F. T. Pokorny, and Y. Karayiannidis, "Reform: A robot learning sandbox for deformable linear object manipulation," in *2021 IEEE International Conference on Robotics and Automation (ICRA)*, pp. 4717–4723, IEEE, 2021.