Progress Report

Bardia Mojra

May 28, 2021

Robotic Vision Lab

The University of Texas at Arlington

1 Specific Research Goals

- Pose Estimation: Implement and improve.
- NBV-Grasping.
- Pose estimation survey.
- Universal pose estimation.

2 To Do

- Catch up on my reading list.
- Pose Estimation:
  - Read VEst [1].
  - Implement QuEst in Python.
  - OpenCV: Get comfortable using it.
  - Key point feature extraction.
  - Evaluate various pose estimation methods: PnP, QuEst, and else.
  - PVNet implementation: Paused. Working on a simple pose estimation for now.
  - Look into ARKit [2].
  - Look into ARCore [3].
- NBV-Grasping:
  - Update URDF and Xacro files for UR5e to include sensor, sensor mount (with offset), and the gripper.
  - Add movement constrains for tables and scene.
  - Write two IK functions for gripper and sensor, one for each. It should plug-in with MoveIt configurator.
  - Research and implement point-cloud data to training TensorFlow models.
  - UR5e in simulation: Joe might consider.
  - Learn and implement GraspIt package.
- MSI Fellowship: On pause.
- Look into methods of generating uncertainty data.

3 Reading List

- VEst [1] - On-going
- Vision-based robotic grasping from object localization, object pose estimation to grasp estimation for parallel grippers - a review [4]
- NASA papers [5]
- Leveraging feature uncertainty in the pnp problem [6]
- Berk Calli's YCB [7]
- Normalized objects [8]
- Roadmap [9]

4 Progress

The following items are listed in the order of priority:

- Pose Estimation: I continued with my OpenCV tutorial. At this point, I am trying to learn how to extract key features from objects as the primary step in object pose estimation. My goal is to develop a robust pose estimator that is primarily based on classical approach and to leverage Neural Network where it is absolutely necessary or where its benefits outweight the training time, intraceability, and inference time costs.

  Moreover, I came across Objectron [10] which is developed by a team at Google. The authors provide their precompiled library as part of MediaPipe [11] module. I tested the module and was able to detect bounding box of a cup using webcam (live feed) with about roughly 30 frames per second being processed. The performance is good as it works in a plug-and-play fashion but it is jumpy and could use some anchoring to stabilize the predicted bounding box from frame to frame. The authors contribute the fast inference time to available AR tool kits, e.g. ARKit [2] and ARCore [3], that provide sparse point clouds in 3D. They provide their data set containing 4 million densely annotated images and nearly 15 thousand video clips collected from 10 countries. I think this is an import work and I might want to

familiarize myself with the mentioned toolboxes and make use of the data set they provided.

- QuEst [12]: I still need to implement this in Python.
- VEst [1]: I am working my way through this paper. I reviewed Homography and SVD in detail. I went over EKF as well.
- NASA MSI Fellowship: Need to read more NASA papers.
- PyTorch Tutorials: Transfer learning:
- NBV Grasping Project: On Wednesday, Joe helped me build UR driver code from source on my computer which is a great news. This allows us to proceed with ROS Noetic which has Python3 natively integrated. Moreover, I am working on updating URDF file to meet our needs. Today, I am updating rvl-workstation with Ubuntu 20.04 and ROS Noetic. I also created a channel on Teams where we keep a record of our discussions and decisions. Additionally, I created a MS OneNote that allows us to seemlessly share notes and development documentations. OneNote and GitHub for this project are accessable through NBV-Grasping channel on Teams.
- UTARI:
- Implement features from PoseCNN, DOPE, and BayesOD. - On pause.

5 Immediate Plans - Summer 2021:

The following items are listed in the order of priority:

- Pose estimation:
- NBV-Grasping:
- UTARI:

6 Intermediate Goals - Fall 2021:

- Pose estimation: I must be finished with implementation, perhaps make some improvements, and should be working on a paper for ICRA or CVPR.

- Scene understanding and active learning: After pose estimation, I want to expand my research into scene understanding and active learning in the context of advanced manufacturing.
- ARIAC: Once I am up to speed, I will do the ARIAC workshops/tutorials and will talk to Jerry about possible contributions.

References

[1] A. P. Dani, N. Gans, and W. E. Dixon, "Position-based visual servo control of leader-follower formation using image-based relative pose and relative velocity estimation," in *2009 American Control Conference*, pp. 5271–5276, IEEE, 2009.

[2] "Augmented reality - apple developer." `https://developer.apple.com/augmented-reality/`. (Accessed on 05/28/2021).

[3] "Build new augmented reality experiences that seamlessly blend the digital and physical worlds." `https://developers.google.com/ar`. (Accessed on 05/28/2021).

[4] G. Du, K. Wang, S. Lian, and K. Zhao, "Vision-based robotic grasping from object localization, object pose estimation to grasp estimation for parallel grippers: a review," *Artificial Intelligence Review*, pp. 1–58, 2020.

[5] "Nasa technical reports server (ntrs)." `https://ntrs.nasa.gov/`. (Accessed on 05/07/2021).

[6] L. Ferraz Colomina, X. Binefa, and F. Moreno-Noguer, "Leveraging feature uncertainty in the pnp problem," in *Proceedings of the BMVC 2014 British Machine Vision Conference*, pp. 1–13, 2014.

[7] B. Calli, A. Singh, A. Walsman, S. Srinivasa, P. Abbeel, and A. M. Dollar, "The ycb object and model set: Towards common benchmarks for manipulation research," in *2015 international conference on advanced robotics (ICAR)*, pp. 510–517, IEEE, 2015.

[8] H. Wang, S. Sridhar, J. Huang, J. Valentin, S. Song, and L. J. Guibas, "Normalized object coordinate space for category-level 6d object pose and size estimation," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2019.

[9] "roadmap-2020.pdf." `https://cra.org/ccc/wp-content/uploads/sites/2/2020/10/roadmap-2020.pdf`. (Accessed on 04/30/2021).

[10] A. Ahmadyan, L. Zhang, J. Wei, A. Ablavatski, and M. Grundmann, "Objectron: A large scale dataset of object-centric videos in the wild with pose annotations," *arXiv preprint arXiv:2012.09988*, 2020.

[11] C. Lugaresi, J. Tang, H. Nash, C. McClanahan, E. Uboweja, M. Hays, F. Zhang, C.-L. Chang, M. G. Yong, J. Lee, *et al.*, "Mediapipe: A framework for building perception pipelines," *arXiv preprint arXiv:1906.08172*, 2019.

[12] K. Fathian, J.-P. Ramirez-Paredes, E. Doucette, J. Curtis, and N. Gans, "Quest: A quaternion-based approach for camera motion estimation from minimal feature points," *IEEE Robotics and Automation Letters*, vol. PP, pp. 1–1, 01 2018.