Progress Report

Bardia Mojra

February 12, 2021

Robotic Vision Lab

The University of Texas at Arlington

1 To Do

- Implement pose estimation: Train OD on YCB data.
- Implement pose estimation: Train PE on YCB data.
- Implement DOPE with added dropout before each layer to estimate variational Bayesian inference.
- Data Association and Localization of Classified Objects in Visual SLAM [1]: Finished reading and annotated a virtual copy.
- Pose Estimation Uncertainty paper: I wrote down some ideas and section drafts for the paper.
- Implement PoseCNN, DOPE, and BayesOD.
- Pose Estimation Servery: Will begin writing as soon as this weekend.
- Vision-based robotic grasping from object localization, object pose estimation to grasp estimation for parallel grippers - a review, [2]:

2 Progress

The following items are listed in the order of priority:

- Instead of writing a review on [3], I thought it would be better if I focus on my primary task and write down my ideas for pose estimation. I need to stay focused.
- Object Pose Estimation with Uncertainty: I finished setting up basic object detection, next I will train a model on YCB dataset. I went over PoseCNN and BayesOD code once again and it slightly more sense to me. At some point I want to add an object motion classification feature where the OD subsystem makes predictions about whether each object is stationary, moving passively, or moving actively. This object level classification feature will help with more advanced pose estimations by providing informative priors.

  I believe pose estimation becomes much easer task if there are known dimensions within the frame. Given known objects, or even some known objects in the image frame, one should be able to make accurate inference about most unknown linear dimensions in the image frame. To reduce computational load, I will train a network to detect keypoints for each segmented object and only perform pose estimation inference for those keypoints.

Keypoint Pose Estimation: I believe overall object pose estimation could increase considerably with this proposed method, especially in occlusion cases. Once keypoints are detected, another network estimates xyz pose for each point with uncertainty which I will explain in sections bellow.

Keypoint Pose Uncertainty: Once keypoints are segmented in each frame, a certainty region is then defined about that point in 3D. This region is divided into regions either as a hypercube or hypersphere. We do the same for all keypoints of all objects present in the scene and connect the keypoints for each object. This method eliminates the need to make a similar angular classification regions.

For each frame, detected objects with high confidence are considered as acceptable predictions by the system and draws a 3D box containing all keypoints for each object. A network would be trained to predict a 3D bounding box given keypoint corners depth, dimensions of the known object, spatial relationship of high confidence objects (need to develop more) in form of embeddings. An IoU loss function [4] could be implemented to optimize 3D bounding boxes.

This method mitigates uncertainty in pose estimation by only considering new evidence if previous process made high confidence predictions and that new evidence should be *consistent* with all previous evidence or this could increase uncertainty of the priors which is an inherent problem in all *chaotic systems*. Robust control tackles this issue by assigning more weight to new evidence and slowly forgetting older evidence. Similar approaches in computer science have become more promenant in machine learning with the introduction transformers (long-short memory, I am not sure).

Similar to weights of a neuron, certainty about a prior should decrease over time unless new supporting evidence observed or it is classified as an anchor with lasting weight and high certainty. Anchors would be used to track keypoint from frame to frame. This brings me to embodiment and expansion into active learning where the agent interacts with the environment, seeks to reduce uncertainty and approaches similar to Next Best View (cite chris's paper)

- YCB Dataset [5]: Right now I am trying to figure out how to train a model using this dataset.

- Normalized Objects [6]:
- Implement features from PoseCNN, DOPE, and BayesOD.

3 Plans

The following items are listed in the order of priority:

- Pose Estimation in Simulation [7]: Use Nvidia Isaac SDK for in-simulation pose estimation training.
- Look into domain randomization and adaptation techniques.
- Project Alpe with Nolan: On pause for right now.
- UR5e: Finish ROS Industrial tutorials.

4 2021 Goals and Target Journals/Conferences

- Submit a paper on pose estimation with uncertainty to ICIRS.
- Get comfortable with TensorFlow and related Python modules.
- Keep writing.

References

[1] A. Iqbal and N. R. Gans, "Data association and localization of classified objects in visual slam," *Journal of Intelligent & Robotic Systems*, vol. 100, pp. 113–130, 2020.

[2] 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.

[3] L. Pinto, D. Gandhi, Y. Han, Y.-L. Park, and A. Gupta, "The curious robot: Learning visual representations via physical interactions," in *European Conference on Computer Vision*, pp. 3–18, Springer, 2016.

[4] J. Yu, Y. Jiang, Z. Wang, Z. Cao, and T. Huang, "Unitbox: An advanced object detection network," in *Proceedings of the 24th ACM international conference on Multimedia*, pp. 516–520, 2016.

[5] 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.

[6] 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.

[7] Nvidia, "Nvidia isaac sdk — nvidia developer." `https://developer.nvidia.com/Isaac-sdk`, 2021. (Accessed on 02/05/2021).