Progress Report

Bardia Mojra

October 23, 2020

Robotic Vision Lab

University of Texas at Arlington

1   Progress

Following items are listed in order of priority:

- Pose Estimation: This week I dissected [1] and wrote a literature review. The notes are on my Task Manager OneNote. Dope uses synthesized data for training their model and they works on so-called reality gap. Reality gap occurs when a model is trained on simulated data and tested in real environment and is characterized by models failure to perform as well as expected based on simulation dataset. Dope uses Domain Randomization (DR) as a feasible solution to reality gap problem. Authors used Falling Things synthetic dataset and well as their own and combined DR images with photorealistic images to yield better training results than using only DR dataset. In essence, Dope is deep neural network that infer the 2D coordinates of projected 3D bounding boxes and then derives the 3D coordinates using EPnP. In their approach, first they estimate a belief map of 2D keypoints of all objects of interest in the image coordinate system. In the second stage they compute (EPnP) 6D pose estimate based on the most likelihood estimate (MLE) from the previous step. Moreover for training, developers generated their ground truth training data in simulation using Unreal Engine as oppose to PoseCNN's training data, YCB data set, which is a set of high resolution scanned items limited to few objects and training images. DOPE's data set is much larger and allows for much deeper neural net training. It is also embedded or fused with eight intercepting lines from camera's focal point to the eight corners of each object's bounding box and another line to the center of the box for each object.

- OCRTOC: Jerry helped me setup a Docker container for DOPE but we need to convert the code from python v2 to v3. I also need to install Unreal Engine on my laptop and start modifying the dataset with Nvidia's NDDS plugin to generate our own embeddings. I think with additional embeddings we can achieve better results or at least broaden the application range. Moreover, I am almost out of space on my laptop so I am going to upgrade it with a 2TB M2 drive and perhaps an additional 2TB SSD for Windows and internal backup. Unreal Engine alone needs about 150GB, Matlab is another 50GB. I see why robotics is going more towards deep learning and that maybe because it is too difficult for most to derive equations of motion a

simple LTI system.

- Reading list: [2] and [3].

- MoreFusion [4]: Still need to write a literature review on this.

- Project Alpe with Nolan: Nolan and I met today and discussed the direction of this side project. We don't want to spend more than 2-3 hours a week on this and we want to make it as productive as possible. Our objective is to record one-hour bi-weekly podcast on computer vision, AI and robotics where we discuss papers we read, what we learned recently and some of our projects. We will each create a professional website and incrementally post about our academic progress.

- TensorFlow: Now that I have some clue what I am doing in regards to pose estimation, I am shifting more time to implementation. I am working on chapter 2 on [5] which begins with a computer vision example using TensorFlow and Keras. I will try to finish it by end of the weekend. Everything makes more sense now, even when I review previous notes and reports I find new connections and familiar patterns.

- Quaternions: I think I understand this much better now. I don't think Hamilton intended to describe a 4th spatial dimension as described by some sources. I believe 6D and 7D notations are merely to simplify complex calculation. It is much easer to describe motion of an object with an additional variable term (a new mathematical dimension) than to calculate it for every instance using differentials of position with respect to time. It is an abstractions technique similar to the 3 angular "dimensions".

- UR5e:

- Fellowship:

2   Plans

Following items are listed in order of priority:

- (On pause) Continue with ROS Industrial tutorials and documentation.

- (On pause) Resume Robotic Perception course as soon as possible.

- (On pause) Read Digital Image Processing by Gonzalez and Woods.

References

[1] J. Tremblay, T. To, B. Sundaralingam, Y. Xiang, D. Fox, and S. Birchfield, "Deep object pose estimation for semantic robotic grasping of household objects," *arXiv preprint arXiv:1809.10790*, 2018.

[2] J. Lampinen and A. Vehtari, "Bayesian approach for neural network-sreview and case studies," *Neural networks*, vol. 14, no. 3, pp. 257–274, 2001.

[3] R. Li, Z. Liu, and J. Tan, "A survey on 3d hand pose estimation: Cameras, methods, and datasets," *Pattern Recognition*, vol. 93, pp. 251–272, 2019.

[4] K. Wada, E. Sucar, S. James, D. Lenton, and A. J. Davison, "Morefusion: Multi-object reasoning for 6d pose estimation from volumetric fusion," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 14540–14549, 2020.

[5] B. Planche and E. Andres, "Hands-on computer vision with tensorflow 2," 2019.