Bootstrapping Object Pose Detection

Tushar Agrawal
SCS, RI, MRSD
tagrawa1@andrew.cmu.edu

Cole Gulino
SCS, RI, MRSD
cgulino@andrew.cmu.edu

Erik Sjoberg
SCS, RI, MRSD
esjoberg@andrew.cmu.edu

Keywords (optional): [computer vision, bootstrap, convnets, object pose estimation]

Introduction

Object pose detection has many applications in the realm of robotics including industrial applications for robots that pick and place objects. Many state of the art algorithms are designed to use computer vision and other perception techniques (iterative closest point, etc.) in order to estimate the pose of an object on the table. Often dense 3D models of the objects and classic computer vision techniques are utilized in realms where the object's shape is readily known to improve perception accuracy.

Problem Statement and Challenges

The basic problem statement is that we want to determine if we can use data generated from current classic methods to bootstrap a neural network to detect object pose reliably.

We posit that we may need to still use ICP in order to fine tune our solution from the neural network.

The main problem with current techniques, specifically ICP, is that it requires a good initialization of pose in order to be effective. This can be problematic and slow if you are required to try many initialization techniques. We believe that by bootstrapping a deep network to learn an estimate of the pose directly from images, we can use that as a good initialization scheme for the ICP algorithm if direct estimate of pose from the neural network is unreliable.

Datasets

The idea of our approach is that we want to use datasets that contain RGB-D information so that we can use the data for multiple techniques including using depth data for iterative closest point.

Here we have seen a few datasets that will fit our needs including Berkeley's Bigbird dataset [4] and Rutgers' dataset [5] designed for pick and place robotics.

Valid Metrics

Comparing with other methods that we have observed, we will use median pose error (in 3D) and percentage of objects who have a median pose error of greater than 15 degrees as our metrics. We will compare these to other state of the art and our own methods including (NN as an initialization to ICP, full NN, and ICP).

Literature Survey

Several approaches have been used to predict pose from RGB-D data. Schwarz et al[1] classify objects and detect poses using transfer learning on features CNNs which are pre-trained for image categorization and provide a rich, semantically meaningful feature set.

Zeng et al[2] consider the problem from a warehouse automation perspective. They use a CNN to segment and label multiple views of a scene and then fit pre-scanned 3D object models to the resulting segmentation to get the 6D object pose. In addition, they propose a self-supervised method to generate a large labeled dataset without tedious manual segmentation.

Yu et al[3] proposed a method utilizing images to generate poses. They use a novel object detection system to segment objects from background before recognition. Then a Max-pooling CNN is used to recognize objects and estimate their poses.

Plan

For the first milestone on November 7:

- Gather data from various datasets
- Create pipeline for generating our own data
- Setup neural network architecture to begin training

For the end of project on December 5:

- Train Neural Network fully to learn pose of objects
- Produce output and metrics using full NN, NN as initialization, and classic methods ICP
- Compare these methods in their strengths and weaknesses

References

[1] Schwarz, M. & Schulz, H. & Behnke, S. (2015) RGB-D Object Recognition and Pose Estimation based on Pre-trained Convolutional Neural Network Features. *ICRA Seattle May 2015*.

[2] Zeng, A. & Yu, K. & Song, S. & Suo, D. & Walker, E. & Rodriguez, A. & Xiao, J. (2016) Multi-view Self-supervised Deep Learning for 6D Pose Estimation in the Amazon Picking Challenge, *arXiv:1609.09475*.

[3] Yu, J. & Weng, K. & Liang, G. & Xie, G. (2013) A Vision-based Robotic Grasping System Using Deep Learning for 3DObject Recognition and Pose Estimation, *International Conference on Robotics and Biomimetics (ROBIO) Shenzhen, China 2013*.

[4] Singh, A. & Sha, J. & Narayan, K. S. & Achim, T. & Abbeel, P. (2014) "Bigbird: A large-scale 3d database of object instances" *ICRA 2014*

[5] Rennie, C. & Shome, R. & Bekris, K. E. & De Souza, A. F. (2016) "A dataset for improved rgbd-based object detection and pose estimation for warehouse pick-and-place." *Robotics and Automation Letters 2016*