Depth-Based Object Detection and Pose Estimation

Object pose estimation is vital for many robotic tasks such as grasping or determining object relations. Most state-of-the-art approaches focus on the extraction of RGB-features for pose estimation and employ depth data only as refinement step. This methodology is wasteful because it does not exploit the depth information in the primary pose estimation phase. Consequently, information regarding the shape of objects is not used for the task, which is highly important in certain domains that deal with texture-less objects.

The aim of this project is to build on state-of-the-art deep networks for object detection and develop a new multi-task, end-to-end architecture to estimate the pose of rigid and texture-less objects. This will require familiarization and modification of recent deep learning architectures that apply pose regression to object detection networks. The modifications should be flexible such that different object detection methods can be integrated and the performance compared. Lastly, different inputs and target layouts should be tested and compared.

Tasks

- Become familiar with the state of the art of deep architectures for object detection and pose estimation.
- Adapt architectures to enable multi-task, end-to-end learning for pose estimation that use depth data only.
- Conduct experiments using different architectures, backbones, input and target data.
- Train the model on existing synthetic data and compare it to a model trained on real data.

Workload split

- Research and theory: 30%
- Programming and implementation: 50%
- Writing: 20%

Contact
Stefan Thalhammer, thalhammer@acin.tuwien.ac.at