3D Reconstruction from Depth Data using CNNs

3D-Rekonstruktion aus Tiefendaten unter Verwendung von CNN-Architekturen

IDP Proposal

An ongoing research interest in the area of 3D reconstruction is the minimization of viewpoints necessary to reconstruct the 3D geometry of an object to a certain degree of accuracy. Specifically, 3D reconstruction methods using convolutional neural networks (CNN) are of particular interest. A recent publishing of a large-scale 3D model dataset, called ModelNet, has made possible the rendering of arbitrary viewpoints to evaluate these methods. The intent is to apply this insight to aid robot manipulation, to allow a robot arm to view an object from a few viewpoints and then grasp it at an optimal angle.

An interdisciplinary project is being proposed to join research in the Computer Science and Electrical Engineering communities. From the Computer Science side, the project will utilize computer vision for the task of 3D geometry reconstruction. Computer graphics will be used to render models at arbitrary viewpoints and to voxelize models for use in training. Finally, machine learning, namely a deep learning technique, will be heavily used to learn 3D geometries of a set of objects. From the Electrical Engineering side, specifically robotics, the project focuses on the topics of robot manipulation and robot vision. Within robot manipulation, the student will need to understand inverse kinematics, motion planning techniques, methods for state estimation and sensor fusion and collision avoidance techniques.

The lecture “Dynamic Human-Robot Interaction” offered by the chair of Dynamic Human-Robot Interaction for Automation Systems is a suitable lecture for a computer science student interested in becoming better informed about the robotics theory in this project. The lecture contents include inverse kinematics, motion control with regression models, probabilistic state estimation methods and reactive collision avoidance approaches. Beyond this, depth data is widely used in robot interaction applications, so best practices of utilizing depth data will very likely be presented throughout the course.

The lecture also includes a section on cognitive human-robot interaction, which is very much in line with what this research is attempting to explore. On the practical side, the course includes tutorials on Robotic Operating System (ROS) as well as MATLAB's Simulink, either of which may be necessary to implement the desired end result.

Project Milestones

1. Conduct a preliminary survey of suitable object categories out of the ModelNet dataset for this application.
2. Rendering of depth maps of each model at various viewpoints and generation accompanying voxelizations.
3. Creation of a CNN architecture and training thereof with one viewpoint of each model.
4. Evaluation of using multiple viewpoints at one once as well as finer vs. coarser output resolutions. Of additional interest is evaluating the generalizing power of the model to reconstruct objects of unseen categories.
5. Application of these trained models to depth data acquired from real objects using a robotic arm and depth sensor.
6. Computation of feature points based on the output of the trained model for optimal object grasping.

A final report will be written detailing the project through the various milestones and a final presentation will be given to share the insight gained over the course of this interdisciplinary project. This is expected to connect the student's knowledge in computer vision and research in machine learning with real-world problems in robotics.