Deep-Learning Solution for a Robotic Bin-Picking Task

Degree programme : BSc in Micro- and Medical Technology | Specialisation : Robotics
Thesis advisor : Prof. Dr. Gabriel Gruener
Expert : Matthias Höchemer (CSEM, SA)

A novel approach has been developed to estimate the pose of parts stored in bulk. It promises high accuracy and real-time speed, while only requiring affordable RGB cameras. The system relies on machine learning to predict the poses of the objects and uses a clustering algorithm to combine the estimates from multiple views.

Introduction

Grasping a part stored in bulk with a robotic arm is called bin-picking. It is of great value for industrial automation facilities, as it allows parts to be stored in any standard box and can make the use of customized feeding systems obsolete. An industrial partner has commissioned the project. Together with the research center CSEM SA and the HuCE Laboratory for Robotics at Bern University of Applied Sciences, they have developed working bin-picking solutions that require expensive camera technology.

The goal of this project is to develop a machine-learning solution to find the pose of at least one robot-pickable part stored in bulk with many other parts of the same type by using images from multiple affordable RGB cameras.

Methods

Bin-picking is an active field of development. Many methods have been published in recent years, most of them need depth images as input. Here, a novel end-to-end pipeline is presented that avoids using any depth information and directly estimates the pose from RGB images. The pose and visibility for each part in a scene is estimated independently for four cameras using a convolutional neural network. The estimates are transformed to a global reference frame to group together pose estimates of the same part from different views, using a clustering algorithm. Finally, a pose estimate and a confidence score is calculated for each group.

The orientation is computed by averaging the estimates from each view. It was demonstrated that depth estimation based on known part size cannot be accurate enough for this application and is therefore not used. By having several cameras around the bin, the parts can be precisely localized in 3D using only the accurate position estimates in the corresponding image planes. The scene is replicated in the game engine Unity to synthesize images that can be used to train the machine learning algorithm. The algorithm is trained only on synthetic data, because gathering data from real bin-picking scenes is difficult and time consuming. To achieve a successful transfer from synthetic to real data, domain randomization is used. This includes randomizing the colors and textures of objects, the lights, the size and shape of the box as well as adding random distractor objects.

Florian Dominic Burri
f6.burri@gmail.com

Results

The pipeline has been implemented in Python. It is expected to be real-time capable, processing multiple scenes per second.

Outlook

The next step is to grasp the parts with a robot, given the pose estimates.

Bin-picking scene.

Synthetic images.