Edge Grasp Classification in Deep Learning

Dian Wang

Northeastern University, Boston, MA  
wang.dian@husky.neu.edu

Abstract

This paper considers the problem of the influence of table edge or shelf edge to grasp pose detection in point clouds. A model is trained to take point cloud and grasp configurations as input to classify each grasp as either edge grasp that is pointing to an edge of table or shelf, or a non-edge grasp that is pointing to an actual object. This paper benchmark two representation of potential grasp, and two convolutional neural networks is built corresponding to the two representations. Experiments were performed to compare those two methods. The analysis shows…

Introduction

Grasping is one of the most intuitive and the most widely applied technology in robotics. Recently, researchers have proposed grasp detection methods that estimate potential grasps independently of object identity. Those methods take RGBD image or point cloud from depth sensor as input and generate output of various grasp configurations.

Those methods are promising and reliable. (ten Pas, Gualtieri et al. 2017) achieved a 93% end-to-end grasp success rate for novel objects presented in dense clutter with their grasp pose detection (GPD) method. However, one problem of such method is that in practical grasping applications which aim to grasp object on table or shelf, some grasp pointing to the edge of table or shelf may be regarded as good grasp.

In this paper, a series of methods are proposed that make GPD avoid that problem. First, 2 grasp representation methods are used to turn raw data from point cloud as well as grasp configuration into image that can be used for neural network input. Second, 2 convolutional neural networks are built respectively for those 2 representations. They take the representation image as input and output the probability for that grasp being edge grasp or non-edge grasp. Third, experiments are made to evaluate the performance of the 2 models.

Background

Grasping involves observing the environment as well as generating potential grasps from observed information. Some grasp detection method like the ROS grasp pipeline (Chitta, Jones et al. 2012) build a CAD model of the target object, then calculate a trajectory that grasp the localized object. However, this approach assume that an accurate CAD model could be built for every grasping object. In contrast, GPD (ten Pas, Gualtieri et al. 2017) detect potential grasps by analyzing the local geometry and/or graspable surface.

|  |
| --- |
| **Algorithm 1** Grasp Pose Detection |
| **Input:** a viewpoint cloud, ; a region of interest, ; a hand, ; a positive integer,  **Output:** a set of 6-DOF grasp candidates,  1:  2:  3:  4:  5:  6: |

Their algorithm, as is shown in Algorithm 1, works as follow. Step 1 preprocess the original point cloud. Step 2 characterize the region of interest (ROI), which is the region of the target location. Step 3 samples a series of grasp candidates from ROI. Step 4 encodes each grasp candidate as a multi-channel image. Step 5 assign each candidate a score. Those scores indicate how likely that candidate is a grasp. Step 6 select a number of grasps based on the score from step 5.

Related Work

Choosing Grasp Heuristically

By manually confine grasp configuration based on the known knowledge of the environment, edge grasps could be avoided. For example, a “top grasp” (Gualtieri, Kuczynski et al. 2017), which means the gripper approaches the object from the top down, is always a non-edge grasp since the edge is always perpendicular to the gravity vector. However, this requires enough space above the grasp position. If there is only “side grasps”, the problem still exists.

Multi-view Representation of Point Cloud

One approach to apply machine learning methods to point cloud is using multiple views. Multi-view CNN (Su, Maji et al. 2015) takes 2D images from 12 different views from a 3D shape model as input, outputs object class predictions.

Based on multi-view CNN, GPD used 15 channels image representation for each grasp. For each of the x, y, and z axis, they have an averaged height map of the occupied points, an averaged height map of the unobserved region, and 3 averaged surface normal images.

Occupancy Grid Representation of Point Cloud

3D convolutional neural networks are also used when applying machine learning methods to point cloud data. VoxNet (Maturana and Scherer 2015) provided a basic 3D CNN architecture that can be applied to create fast and accurate object class detectors for point cloud data. Their method takes point cloud as input, generate occupancy grid based on that point cloud, and uses a 2-layer 3D CNN for classification.

Project Description

Overview of Edge Grasp Classification Algorithm

Given a point cloud and a transformation matrix representing a grasp, the edge grasp classification problem is to compute the probability of that grasp being an edge grasp or a non-edge grasp.

|  |
| --- |
| **Algorithm 2** Edge Grasp Classification |
| **Input:** a point cloud, ; a grasp transformation matrix, *T*  **Output:** a probability of that grasp being edge grasp,  1:  2: |

My algorithm follows the steps shown in Algorithm 2. Step 1 encodes the grasp candidate into a multi-channel image from the original point cloud and the transformation matrix of that grasp. Step 2 generate the probability of that grasp being an edge grasp.

Grasp Description

Gr

This Is an Example Subsection Heading

The text below a second-level heading begins without indentation. Use of the subsection heading style is required.

This is the second paragraph. It in formatted with the Text-indent style. Use of the Text-Indent style is required.

This Is a Subsubsection Heading

This is example text. It is 10 point Times New Roman. This is example text. It is 10 point Times New Roman.

This is example indented text. It is 10 point Times New Roman. This is example indented text. It is 10 point Times New Roman. This is example indented text. It is 10 point Times New Roman. This is example indented text. It is 10 point Times New Roman..

This is an example of an extract or quotation. Note the indent on both sides. Quotation marks are not neces­sary if you offset the text in a block like this, and properly identify and cite the quotation in the text.

This Is an Example of a Figure Caption. Use of the figure caption style is required.

This Is a Section Heading

Use of the section heading style is required.

Margins

Papers **must** be printed in two-column format on 8-1⁄2 x 11 inch paper. The margins if you use 8-1/2 x 11 inch paper must be **exactly** as follows:

• Top margin: 3⁄4 of an inch

• Left margin: 3⁄4 inch

• Right margin: 3⁄4 inch

• Bottom margin: 1-1⁄4 inches

The actual image size for your paper (minus margins) is 7

x 9 inches (17.78 x 22.86 cm). Papers that deviate from these measurements will not be published.

Style Sheet Use and Alterations

You are required to use the style sheets embedded in this document without alteration.

Fonts

All fonts must be embedded for your paper to be published. No exceptions.

References

Chitta, S., E. G. Jones, M. Ciocarlie and K. Hsiao (2012). "Mobile Manipulation in Unstructured Environments: Perception, Planning, and Execution." IEEE Robotics & Automation Magazine **19**(2): 58-71.

Gualtieri, M., J. Kuczynski, A. M. Shultz, A. Ten Pas, R. Platt and H. Yanco (2017). Open world assistive grasping using laser selection. Robotics and Automation (ICRA), 2017 IEEE International Conference on, IEEE.

Maturana, D. and S. Scherer (2015). VoxNet: A 3D Convolutional Neural Network for real-time object recognition. 2015 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS).

Su, H., S. Maji, E. Kalogerakis and E. Learned-Miller (2015). Multi-view convolutional neural networks for 3d shape recognition. Proceedings of the IEEE international conference on computer vision.

ten Pas, A., M. Gualtieri, K. Saenko and R. Platt (2017). "Grasp Pose Detection in Point Clouds." The International Journal of Robotics Research **36**(13-14): 1455-1473.