Improving Grasping Performance by Segmentation of Large Planar Surfaces

Vasco Lopes
http://www.di.ubi.pt
Luís A. Alexandre
http://di.ubi.pt/~lfbaa

Departamento de Informática
Universidade da Beira Interior
Instituto de Telecomunicações
6201-001 Covilhã, Portugal

Abstract

Grasping objects is a task that humans do without major concerns. This results from learning and observing other skilled humans doing such task and with previous information, unconsciously, we know how to pick up different types of objects. However, grasping novel objects in unknown positions for a robot is a complex task which encounters many problems, such as the performance rates that are not perfect and the time consumption. In this paper we present a method that complements the state-of-the-art grasping by removing the largest planar surface of the image of the world before the grasp detector receives them. The proposed method improves the performance rate and is also capable of reducing the time consumption.

1 Introduction

Grasping novel objects is a very complex task for a robot and that's why it's an important area and with active and extensive research. In this paper we present a method to improve a current state-of-the-art algorithm that gives a robot the capability of grasping novel objects in unknown positions. Robots are getting more and more present in our daily basis, but some tasks still encounters many barriers, which is the case of grasping novel objects. The most predominant problems in the state-of-the-art methods are the incapacity of achieving perfect results in detecting grasps and the time spent on processing the algorithm for detecting such graps, because in a real-life situation, if the robot fails a grasp it can damage itself or persons that are around it, and if it spends too much time processing, the world can change and it executes movements that are not correct anymore and may collide with objects.

In a related work, Kehoe *et al*. [2] used the cloud to serve as a vast source of computation and data. The aggregation and sharing of training data proposed by this paper means that training multiple robots can occur faster than training on a single robot, which can be a way to address the problem of a robot encountering novel objects and having the cloud serving as a computation source can also decrease the run time.

Saxena *et al*. [5], presented a possible solution to the problem of grasping novel objects that the robot is perceiving for the first time thought vision. They propose a learning algorithm that doesn't require a 3D of the object, instead, the algorithm tries to identify a set of points in 2D images that corresponds to a good point at which to grasp the object, and with that point, it uses triangulation to obtain a 3D position to attempt the grasp.

Although there is work done in this area that tries to minimise the problems, there isn't a perfect solution. In this paper we propose an improvement to the method proposed in [1], which tries to calculate possible grasps by randomly selecting points from the point cloud and for each, calculating a surface normal and an axis of the major principle curvature of the object surface in the neighbourhood of that point. It generates potential hand candidates at regular orientations orthogonal to the curvature axis, and for each hand candidate it verifies if it is a possible grasp candidate and then classifies each one of the possible grasp candidates as either a viable grasp or not, using a deep neural network.

2 Proposed Method

The method proposed is an improvement to the original method described in [1]. In the original method, the grasp detector receives the information of the world directly, in our experiments using a Kinect. We propose a change to this method by introducing a segmentation node. This node receives the information of the world in form of a point cloud, calculates the largest planar surface using RANSAC, removes it from the image and sends it to the grasp detector node. The justification for removing the largest planar surface is that it usually represents a table top, the floor or a wall, and not the object to be grasped. By removing this large planar surface we are reducing the amount of data to be processed by the grasping algorithm. This can have two benefits: first, the potential grasps will not appear on the removed plane, increasing the probability that they are correctly placed on the object to be grasped. Second, since the potential grasps are more likely to be correct, the algorithm can work with a smaller number of attempts, to achieve the same grasping success rate, but using less time to do it. In figure 1, it's possible to see in a) the original image that is received by the segmentation node, in b) the plane, in red, that was calculated as being the largest one and in c) the segmented image without the plane in it.

The proposed method serves as an improvement in both time, because we are able to reduce the number of samples chosen to create grasp candidates, and performance in terms of the success in detecting viable grasps. A viable grasp is a grasp that a robot is able to perform. In figure 2 we can see two examples of where the grasp detector successfully detected viable grasps: they are indicated in the images as a blue parallel jaw gripper.

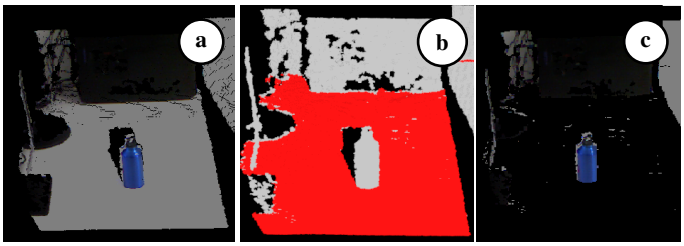

Figure 1: Three steps of the segmentation algorithm.

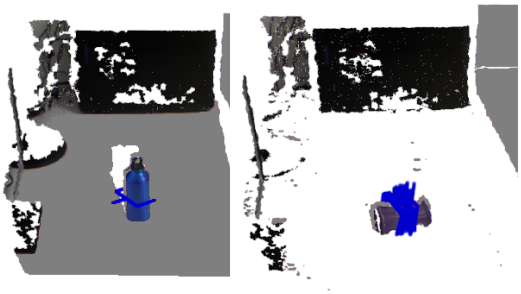

Figure 2: Two examples in which the grasp detection was successful. The left imagem shows a single grasping candidate and on the right image there are multiple succesfull candidates.

3 Experiments

We used the Robot Operating System (ROS) [3] and the Point Cloud Library (PCL) [4] for the implementation of this method. In order to compare the performance of the proposed method against the original one, we conducted three experiments, each one using the same set of 8 objects, which can be seen in figure 3. For each object we took 20 images with a Kinect in our lab, simulating a possible scenario where a robot needs to grasp an object on a table. Between each shot of an image we changed the object orientation and position so that each image differs from the others. After the images were captured, we segmented each one so we can evaluate the performance of the grasp detector with non-segmented images against segmented images. The average segmentation time was 0.21 seconds per image.

Each experiment had the same number of trials, 20 per object, using either original images (non-segmented) or segmented images. The grasps

were not executed by a real robot: the method [1] identifies if a grasp is viable or not. A trial is classified as successful if the algorithm detects at least one viable grasp, if otherwise, the trial is classified as a failure.

Figure 3: Set of 8 objects used to create the database.

3.1 Experiment 1

To serve as a baseline, we evaluated the performance of the original method with the original images (non-segmented). For this experiment we set the number of points to be sampled from the grasp generator, n, as 5000. The results of this experiment can be seen in table 1. The average run time was 2.63 seconds, this means that in each trial, the original method calculated the grasp candidates and classified them in 2.63 seconds, achieving a success rate of 45.63%, showing that the original method was capable of detecting a viable grasp in 73 of the 160 trials.

Table 1: Results of the experiment using non-segmented images with n=5000.

	Success	Failure	Average run time (s)	Success Rate (%)
Blue Canteen	19	1	2.65	95.0
Cardboard Box	11	9	2.69	55.0
Cardboard Cup	9	11	2.63	45.0
Clay Cup	11	9	2.66	55.0
Gel Tube	11	9	2.58	55.0
Headphones	6	14	2.64	30.0
Paper Holder	1	19	2.49	5.0
Water Bottle	5	15	2.69	25.0
Overall	73	87	2.63	45.63%

3.2 Experiment 2

In this experiment, we tested the proposed method with the same number of sampling points as in 3.1, using the segmented images of the world. This experiment had as results a success rate of 71.88% and an average run time of 3.09 seconds, as can be seen in table 2. The total average run time of this experiment is 3.3 seconds. This is the sum of the average run time (3.09 seconds) and the average segmentation time for each image (0.21 seconds), which means that in this experiment, the proposed method is capable of segmenting an image, generate grasp candidates and classifying them with success in 115 of the 160 trials in an average run of 3.3 seconds.

Table 2: Results of the experiment using segmented images with n=5000.

	Success	Failure	Average run time (s)	Success Rate (%)
Blue Canteen	20	0	3.09	100.0
Cardboard Box	12	8	3.28	60.0
Cardboard Cup	19	1	3.08	95.0
Clay Cup	17	3	3.08	85.0
Gel Tube	18	2	3.17	90.0
Headphones	12	8	3.17	60.0
Paper Holder	8	12	2.94	40.0
Water Bottle	9	11	2.90	45.0
Overall	115	45	3.09	71.88%

3.3 Experiment 3

In this experiment the proposed method was tested, with the same setup as the experiment described in 3.2 but with less sampling points. The number of points to be sampled in the image from the grasp detector in order to create grasp candidates, n, was reduced from 5000 to half, 2500. This experiment reduced the average run time of creating grasp candidates and classifying them to 1.62 seconds and achieved an success rate of 43.13% as described in table 3. This means that the total average run time is 1.83 seconds (1.62 seconds of the grasp detection run time and 0.21 seconds of segmenting the images) and that the grasp detector has able to detect viable grasps in 69 out of 160 trials.

Table 3: Results of the experiment using segmented images with n=2500.

	Success	Failure	Average run time (s)	Success Rate (%)
Blue Canteen	20	0	1.71	100.0
Cardboard Box	6	14	1.74	30.0
Cardboard Cup	10	10	1.60	50.0
Clay Cup	12	8	1.58	60.0
Gel Tube	12	8	1.64	60.0
Headphones	4	16	1.58	20.0
Paper Holder	1	19	1.51	5.0
Water Bottle	4	16	1.58	20.0
Overall	69	91	1.62	43.13%

4 Conclusions

Grasping novel objects in unknown positions is a challenging task which doesn't have a perfect success rate and normally takes several seconds to process, which can be critical in a real-life scenario.

In this paper we present a method that can be used to determine a successful grasp faster or with a higher success rate, than a current state-of-the-art approach. It is capable of reducing the average run time in about 31% and still have similar success rate, achieving a result of 43.13% in 1.83 seconds, or, if the concern is not how fast the method runs but the success rate, the proposed method is able to increase the success rate in approximately 26%, from 45.63% to 71.88%, using a similar time as the original method.

Future work will consider the possibility of ataining both benefits simultaneously, shorter execution times and higher grasping performance, by optimizing the segmentation process to further reduce the amount of data to be processed.

Acknowledgments

This work was supported by National Funding from the FCT - Fundação para a Ciência e a Tecnologia, through project UID/EEA/50008/2013.

References

[1] Marcus Gualtieri, Andreas ten Pas, Kate Saenko, and Robert Platt. High precision grasp pose detection in dense clutter. *CoRR*, abs/1603.01564, 2016. URL `http://arxiv.org/abs/1603.01564`.

[2] B. Kehoe, A. Matsukawa, S. Candido, J. Kuffner, and K. Goldberg. Cloud-based robot grasping with the google object recognition engine. In *2013 IEEE International Conference on Robotics and Automation*, pages 4263–4270, 2013.

[3] Morgan Quigley, Ken Conley, Brian Gerkey, Josh Faust, Tully B. Foote, Jeremy Leibs, Rob Wheeler, and Andrew Y. Ng. ROS: an open-source robot operating system. In *ICRA Workshop on Open Source Software*, 2009.

[4] R.B. Rusu and S. Cousins. 3D is here: Point Cloud Library (PCL). In *IEEE International Conference on Robotics and Automation (ICRA)*, Shanghai, China, May 2011.

[5] Ashutosh Saxena, Justin Driemeyer, and Andrew Y. Ng. Robotic grasping of novel objects using vision. *The International Journal of Robotics Research*, 27(2):157–173, 2008.