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.

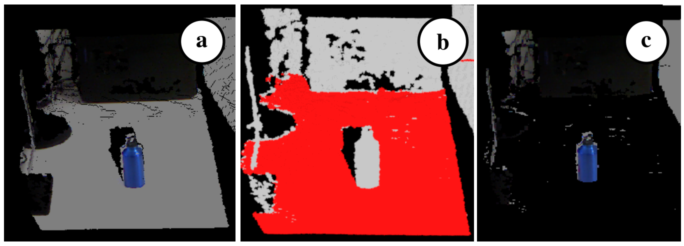

Figure 1: Three steps of the segmentation algorithm.

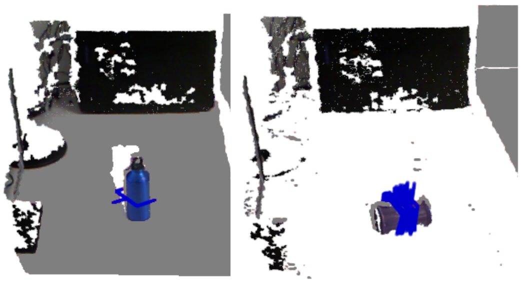

Figure 2: Two examples in which the grasp detection was successful. The left imagem shows a single grasping candidate and on the right image there are multiple succesfull candidates.

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