Manipulate a Pile of Objects Using Pose Analysis

A-Tang LEI\textsuperscript{a}, Bing-Wei HE\textsuperscript{b,\ast}, Li-Wei ZHANG\textsuperscript{c}, Pei-Qing YAN\textsuperscript{d}

School of Mechanical Engineering and Automation, Fuzhou University, Fuzhou, PR China
\textsuperscript{a}leiatiang@163.com, \textsuperscript{b}mebwhe@fzu.edu.cn, \textsuperscript{c}lw.zhang@fzu.edu.cn, \textsuperscript{d}peiqingyf@163.com
\ast Corresponding author

Keywords: Pose analysis, Manipulate, Largest success ratio of grasping.

Abstract. We address the problem of manipulating a pile of objects on the table using pose analysis approach. The stacked objects are separated with RANSAC and the gradient segment algorithm. The CVFH feature of the point cloud of separated objects is extracted, and SVM is used for the pose estimation. Then the largest success ratio of grasping is computed by analyzing the success ratio of grasping objects from different directions. According to the current pose of objects and the largest success ratio of grasping objects, the robot hand-joints is driven to access grasping points on objects. We tested our approach using a RGB-D sensor (Xtion Pro Live) and a Nao humanoid. Results proved that our approach is with high efficiency for robustly manipulating a pile of objects on the table top.

Introduction

Mobile and autonomous manipulation has become the prerequisite for developing robotic applications to make our daily life more convenient [1]. In our work, we address the problem of manipulating a pile of objects on the table top with robot hands autonomously and reliably. It is useful in waste classification and objects sorting.

It is difficult to manipulate a pile of objects because it is limited by segmentation accuracy, precision of pose estimation and the correctness of robots’ grasping actions. Many factors lead to robots fail to clear a pile of objects: (1) Failing to segment a scene. Real scene on the table top tends to be complex because of containing varieties of objects. In addition, it happens often that objects are hidden or stacked by other objects, under-segmentation and over-segmentation may result in grasping fails. (2) Grasping position and pose. Different grasping positions and poses directly influence the success ratio of grasping. Improper grasping position or grasping pose may result grasp fails. Therefore, without considering grasp pose will result in the low success ratio of grasping or grasp fails.

Our purpose is to manipulate a pile of objects, so we have to solve some problems in our works: firstly, our system has to observe the scene in real time and segment objects precisely, reliably and stably. Secondly, pose estimation algorithm must correctly detect objects’ 6DOF (degree of freedom) pose. Finally, the robot is able to successfully grasp objects according to the position and pose of the objects.

To accomplish the task, we choose an ASUS Xtion Pro Live device as our camera and an Aldebaran NAO Humanoid Robot to grasp the object, which are shown in Fig. 1. Depth images are acquired by the camera and then a pile of objects are separated by segment algorithm. At the same time, the corresponding point cloud is used to extract CVFH features of the separated object. Thus, the pose of the object is estimated with SVM. Based on the above steps, the robot can grasp each object in the better pose and grasping direction according to the location and pose of the object obtained and processed with the RGB-D camera.
Scene Segmentation

The accuracy of segmentation will affect the success ratio of grasping. Under-segmentation may result in an attempt to grasp multiple objects, while over-segmentation may lead to a wrong parameterization of the controller, resulting in an unreliable grasping [2]. So looking for a proper algorithm is very important. In particular, when objects in the scene are hidden or stacked by other objects, finding a reliable method to separate every object is necessary.

Our algorithm combines RANSAC and the gradient segmentation algorithm. We use a RGB-D camera to get depth images of the scene. And then, gradients of the depth image are calculated, and generated depth discontinuities by setting a threshold value. On the other hand, the depth image is converted to dense point cloud, and the table plane is iteratively calculated using RANSAC algorithm. Objects are generated into RANSAC discontinuities. Finally, we overlay the depth discontinuities over the RANSAC discontinuities. The result is shown in Fig. 2.

The input (a) is an RGB image. The algorithm extracts depth discontinuities (b) and RANSAC discontinuities (c). The resulting separated objects (d) are then generated.

Figure 2. Segmentation algorithm.

Although depth discontinuities can separate the stacked objects, some boundaries of objects are fuzzy. On the contrary, RANSAC discontinuities can clearly present the boundary of multiple objects. Our method combining RANSAC and gradient segment algorithm has the following advantages: 1) fully separating all objects even if some objects are hidden or stacked by others, 2) solve the problem of under-segmentation caused by similar color between objects.
Object Recognition and Pose Estimation

The CVFH [3] has its roots in the VFH descriptor, it based on the idea that objects have a certain structure that allows to split them in a certain number N of disjoint smooth regions. The CVFH can not only maintain the strong recognition results, but also add in viewpoint variance to distinguish different poses of the object.

In this paper, our robot grasps objects with different poses, grasping points and grasping directions, so we need to know the categories and poses of the objects in advance, and use the pose of objects to analysis the grasping point and grasping direction. Therefore, we extract CVFH features of objects’ point cloud, which are used by a linear SVM for pose estimation.

Preparing for object classification, we collect some daily objects (such as napkin, glue, chewing gum box and cup) in different pose. Fig. 3 shows the CVFH feature of a pack of napkin.

![Figure 3. The CVFH feature of a pack of napkin.](image)

Grasping

In order to improve success ratio of grasping, we define (1) the pose of objects and (2) a variety of grasping directions. Two ways to improve success ratio of grasping are analyzed as follows.

Object Pose, Grasping Direction and Grasping Point

Different poses of objects are generated from different placed way. For different poses of objects, the success ratio of grasping will be low if the robot grasps it in only one way. The CVFH feature and SVM are used to pose estimation and generate varieties of grasping directions according to the pose of objects. The robot grasps the objects based on the pose and grasping direction. In this way, the success ratio of grasping is higher.

![Figure 4. Different poses and grasping directions of a pack of napkin.](image)
There are some poses and grasping directions of a pack of napkin in Fig. 4. An object has more than one pose (a, b, c), and a pose has more than one grasping direction (a and d, b and e, c and f). In addition, we defined the grasping points (the black dot) which make robot grasp objects more accurately, as shown in the Fig. 4.

**The Success Ratio of Grasping**

In this section, the success ratio of grasping is discussed as it changes with different grasping directions. For example, the success ratio of grasping in Fig. 4(a) is higher than Fig. 4(d). The success ratio of grasping directly influences whether the robots can success grasp objects or not. We analyze the success ratio of grasping of different grasping directions in different objects’ poses, compare the success ratio of grasping in all grasping directions, and find the grasping direction with the highest success ratio.

The scene includes more than one object and exists blocking between objects. In order not to affect other objects when the robot is grasping objects, we redefine the success ratio of grasping:

\[
\Pi_i = \begin{cases} 
\Pi, & \text{there are no objects in grasping direction} \\
0, & \text{there are objects in grasping direction}
\end{cases}
\]  

(1)

where \(\Pi_i\) is the success ratio of grasping when considering that there is no blocking between objects.

We have considered all factors include the pose of objects, the grasping directions, the grasping points and the success ratio of grasping in our grasping process, which is shown as follows.

1. Analyzing the success ratio of grasping of every object after segmentation, and finding the pose and the grasping direction with the largest success ratio.
2. Getting the grasping point according to the pose and the grasping direction.
3. Driving the robot hand to the intermediate grasping point which is at a distance of 10 mm from the grasping point in the opposite grasping direction.
4. Driving the robot hand to the grasping point in the grasping direction.
5. Grasping the object and placing it according to the category of the object.

The process is continuing until there are no objects on the table top.

**Experimental Evaluation**

To evaluate our algorithm, we implemented experiments using an ASUS Xtion Pro Live camera and a Nao Humanoid. The Xtion Pro Live camera was used to capture depth images. Point Cloud Library (PCL) [4] was used to extract CVFH feature of point cloud of the separated objects. The computed values and 6DOF poses were then passed to the robot for controlling joints. The robot grasped the object by choosing the best grasping direction. We detailed our experiment in section A and B as follows.

**The Success Ratio of Grasping a Single Object**

To analyze the success ratio of a single object, we choose 18 daily objects in our experiments. Tab. 1 shows the success ratios of grasping in different pose and grasping direction of napkin, glue, chewing gum box and cup. In the table, “down” represents robot grasping objects from top to bottom, “left” represents robot grasping objects leftward, and we only define “left” grasping direction because the success ratios of grasping are equal when robot grasps objects leftward or rightward. The robot grasps objects hundreds of times in each pose and acquires the success ratio
for robot grasping plan. From the results, we know that the success of grasping differs with different poses and different grasping directions. Fig. 5 shows the comparison between grasping without considering success ratio and grasping based on our algorithm. It shows that the success ratio of grasping based on our algorithm is higher than grasping without considering success ratio.

| Object pose | Grasping direction | Down | Left | Down | Left | Down | Left | Down | Left | Down | Left |
|-------------|--------------------|------|------|------|------|------|------|------|------|------|------|
| Success ratio of grasping | 0.98 | 0.86 | 0.96 | 0.84 | 0.92 | 0.94 | 0.68 | 0.86 | 0.91 | 0.20 | 0.90 | 0.15 |

Number 1-4 are napkin, glue, chewing gum box and cup.

Fig. 5 Comparing the success ratio of grasping between grasping without considering success ratio and grasping based on our algorithm

**Manipulating in the Scene**

Grasping multiple objects in clutter is a complex task. Our experiment scene is shown in Fig. 1, and the steps of grasping are: (1) segmentation in the scene; (2) pose estimation of separated object; (3) ensuring the grasping direction and grasping point; (4) grasping the object. These steps are continuing until all the objects are removed.

Fig. 6 and Fig. 7 are steps that the robot grasps objects in a scene. From the results of experiment, we know that Nao Humanoid can successfully grasp the objects with the highest success ratio sequence, like cup, chewing gum box, cup, and napkin.

Using our segmentation algorithm, we obtained separated objects (different kinds of color represent different objects) before robot began to grasp objects.
In all our experiments the robot was able to remove all objects on the table top. In Fig. 6, the robot begins by grasping a cup. Next, the chewing gum box, napkin and two pack of napkin are grasped. The process is completed successfully.

Figure 6. A sequence of steps of grasping objects in a pile of objects.

Figure 7. The Xtion’s view of the scene during the experiment in Fig. 6.

Summary
The problem of manipulating a pile of objects on the table is represented. Our segment algorithm, which combines RANSAC and the gradient segment algorithm, can successfully separate objects from a pile of objects. Poses, grasping directions and grasping points of objects are analyzed to search the highest success ratio of grasping, and then the robot grasps objects using the highest success ratio of grasping. The experimental results show that our method can greatly improve the ability of the robot to successfully grasp objects in clutter. In our future work, our algorithm will be improved to enhance the ability of the robot in grasping some more objects with complex shapes. A data library, storing poses, grasping directions, grasping points and the corresponding success ratios of grasping of daily objects, will also be built to improve the efficiency of manipulating.

Acknowledgement
This work was supported by the National Natural Science Foundation of China (Project No. 61473090 and 51175087), Fujian Natural Science Funds for Distinguished Young Scholar (Project No. 2013J06013).
References

[1] Umar Asif, Mohammed Bennamoun, and Ferdous Sohel, Clearing a pile of unknown objects using interactive perception. Robotics and Automation (ICRA), 2013 IEEE International Conference on, 2013, pp. 154–161.

[2] Lim, Young Won, and Sang Uk Lee, On the color image segmentation algorithm based on the thresholding and the fuzzy c-means techniques. Pattern recognition, vol. 23, no. 9, pp. 935-952, 1990.

[3] Aldoma A., Vincze M., Blodow N., et al, CAD-model recognition and 6DOF pose estimation using 3D cues. Computer Vision Workshops IEEE International Conference on, 2011, pp. 585-592.

[4] Point cloud library on http://www.pointclouds.org/.