Primitive Shape Recognition Based on Local Point Cloud for Object Grasp

Qirong Tang(✉), Lou Zhong, Zheng Zhou, Wenfeng Zhu, and Zhugang Chu

Laboratory of Robotics and Multibody System, School of Mechanical Engineering, Tongji University, Shanghai 201804, People's Republic of China

Abstract. Object recognition and grasping are important means of interaction between robot and environment, and also two of the main tasks of robot. Due to the rich information provided by depth sensors, it has paved the way for the object recognition. The geometric information is more conducive to the primitive recognition of the object, and the primitive shape information is used as the input information for the robot to grasp. This study proposes a primitive shape recognition method using local point cloud. First, 900 sets of point cloud data including three primitive shapes was created. Then the PointNet network using the point cloud data to recognize the primitive shape of the objects was trained. Experiments in simulation and physical world shows our recognition method can effectively recognize the primitive shape of the object.

Keywords: Primitive shape · Object recognition · Point cloud · PointNet

1 Introduction

Robot is widely used in all aspects of production and life because it can replace human heavy labor, realize production automation and keep human safety. Manipulator grasping is one of the important means for robot to interact with the outside world. It is usually divided into three aspects: perception, planning and control. Model-based control methods have always dominated the field of manipulator grasping, such as model predictive control [1,2] and force control. With the emergence of machine learning and neural networks, data-driven methods [3–5] provide new ways for robot grasping.

The modeling process of model-based methods is complex and the generalization performance of grasping unknown objects does not work well. Meanwhile, the data-driven methods are more robust, but they require enormous data for train. To avoid the disadvantages, many researches are based on the grasping of the primitive shapes.

Grasping based on the primitive shape is an approach from another point of view. In this method, the objects are not accurately modeled. The object is sampled into the primitive shape with prior knowledge, and the grasping posture

Y. Tan and Y. Shi (Eds.): ICSI 2021, LNCS 12690, pp. 85–91, 2021.
https://doi.org/10.1007/978-3-030-78811-7_8

is selected from a small number of grasp candidates. There is no need to do a lot of searches, and alleviate problems that require enormous data. At present, there are two kinds of shape-based grasping methods: selecting the grasping posture according to the predefined grasping postures [6,7] or sorting the grasping quality according to the known grasping modes [8].

The first of grasping with primitive is to complete the target shape recognition. What's more, the development of the depth sensors have paved the way for 3D shape recognition due to the additional 3D information. The deep learning provides an efficient and accurate method for primitive shape recognition. These methods are mainly by extracting RGB-D image [9–11] or point cloud [12–14] features.

This study mainly investigates the 3D shape recognition of the objects' point cloud using PointNet. This study is aiming at

- point cloud dataset of primitive shape creation,
- a primitive shape recognition method using local point cloud.

This paper is organized in the following manner: Sect. 2 introduces the method of establishing primitive shape point cloud dataset and the PointNet network. The effectiveness of the proposed methods is verified through a set of experiments in simulation and physical world and the results are shown in Sect. 3. Section 4 concludes the work.

2 Method of Primitive Recognition

2.1 Primitive Shape Dataset Creation

First, a small dataset is prepared with respect to PointNet structure and the dataset contains point clouds solid objects from three categories: sphere, cylinder, cuboid. The depth image was created using a Kinect V2 in V-REP. These objects are uniform in color and texture information. The size of the objects are different for each. For example, the height-diameter ratio of a cylinder is different. In the V-REP simulation platform, the object and the Kinect sensor are put at different angles and the primitive shape model is rotated around the axis. Therefore, for each shape, depth images were taken from different angles and distances by Kinect sensor. Each primitive shape includes 300 samples depth images.

In order to remove the background depth information, the background subtraction method is used. As shown in Fig. 1, background depth images are subtracted from original images to retain the object depth information. If the final information is less than 0, it is retained. The retained information adds the background information to restoring the original depth of the object. So, the ground is black expect for the objects. Then the depth information is converted to point clouds based on the intrinsics, as in Eq. (1).

$$ZP_{uv} = \begin{bmatrix} u \\ v \\ z \end{bmatrix} = \begin{bmatrix} f_x & 0 & c_x \\ 0 & f_y & c_y \\ 0 & 0 & 1 \end{bmatrix} P = KP, \tag{1}$$

(a) Original depth (b) Background depth (c) Final depth

Fig. 1. Depth image subtraction

where P_{uv} represents the position (u, v) in the pixel coordinate system, z is the depth in the position (u, v) in the pixel coordinate, f_x, f_y is the focal length and c_x, c_y is the offset, K is the intrinsics, P is the camera coordinate system.

The 1024 point cloud represents the object after sampling under the point clouds. Then, the point cloud is normalized, the center of point clouds is translated to the origin of coordinates, as in Eq. (2), and the size of point cloud is scaled to the unit sphere, as in Eq. (3),

$$P = P - \bar{P}, \tag{2}$$

$$P = P/max\{P\}, \tag{3}$$

where the $\bar{P}$ is the center of point clouds and $max\{P\}$ is the max distance of the points to origin.

2.2 Primitive Shape Recognition Using PointNet

The proposed approach hypothesizes that common objects can be divided into three categories: cylinder, sphere and cuboid. This part of recognition objects in terms of shape features using local point clouds extracted from depth image. Since the point clouds has the permutation invariance and rigid transformation robustness, it is necessary to pay attention to these two properties when performing point clouds feature recognition.

The state-of-the-art of deep neural networks are specifically designed to handle the irregularity of point clouds. This approach was proposed by PointNet [12]. The PointNet provides a unified architecture for object classification. Since the point cloud has the permutation invariance and rigid transformation robustness, it is necessary to pay attention to these two properties when performing point cloud feature recognition. The PointNet solves the problems by rotating transformation and constructing a symmetric function. The point clouds is the input of the PointNet, The number of outputs, N, corresponds to the numbers of class.

As shown in Fig. 2, the network structure rotates the input point cloud, and then uses the multi-layer perceptron to arise the three-dimensional point cloud to 1024-dimensional. The maximum pooling layer solves the problem of

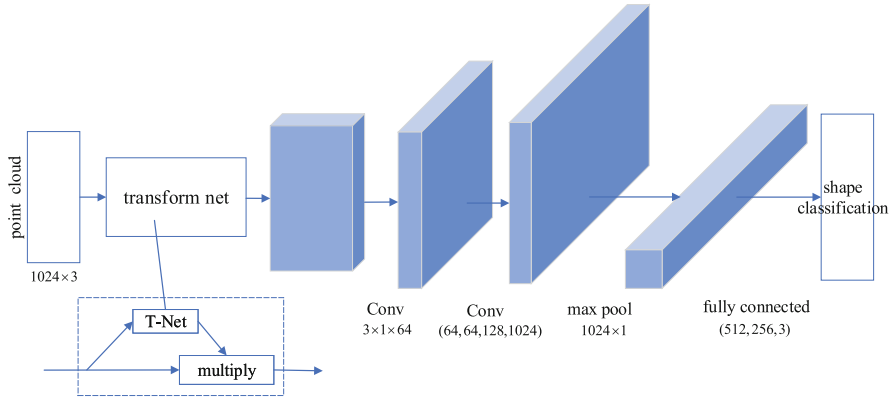

Fig. 2. Modified PointNet model

permutation invariance and extracts the most important points in the point cloud. Finally, shape classification is performed through multiple fully connected layers.

In this study, the PointNet was used to recognize the primitive shape features of the object based on local point clouds. To simplify the model, only the incoming point clouds are spatially transformed, and the rest are consistent with the PointNet.

3 Evaluation

3.1 Model Train

The input data of the PointNet is the three-dimensional point clouds, and the output is the label of the category of the object. If the output label is consistent with the shape label of the object, the shape recognition is considered correct. Those 900 local point cloud data sets created by the V-REP simulation platform are divided into training sets and test sets, which are used for network training and testing, respectively. In the training process, random point cloud interference is used to increase the diversity of training data. After training, objects recognition performance experiment are carried out in simulation and in the actual environment.

3.2 Results of Simulation Experiment

In simulation, the 3D model of the object was downloaded from the YCB model library for verifying the accuracy of the recognition. The objects are similar to the features of sphere, cylinder, and cuboid, such as cups, tennis balls, oranges, etc. Compared with standard primitive shape, the surface texture of the objects is more complicated. Meanwhile, each primitive shape contains three objects.

The method of obtaining the target point cloud is consistent with the method of the primitive 3D model point cloud. The test point cloud data contains 10 sets of point cloud data for each objects and includes 9 types of objects, totally 90 sets of point cloud data (sphere: 30, cuboid: 30 and cylinder: 30) are considered. The point cloud data only uses the trained network model for shape recognition. The PointNet extracts the features of the point cloud to recognize the primitive shape of objects. In the simulation, the primitive shape recognition results are shown in Table 1.

Table 1. The accuracy of shape recognition in simulation (%)

Objects	Can1	Can2	Cup	Cookie box	Candy box	Block	Golf	Tennis	Orange
Recognition accuracy rate of each object	100	100	80	90	100	90	100	100	100
Recognition accuracy rate of each shape	93.3			93.3			100		
Average	96								

3.3 Results of Physical Experiment

Preprocessing of Actual Depth Image. In this part, Kinect is used to obtain the 3D point cloud of the objects in the actual world to verify the model. Considering the noise of depth images in the actual environment, it is necessary to preprocess the image of the depth sensor to better recover the point cloud and reduce the interference of environmental noise.

(a) Depth image before filtering (b) Depth image after filtering

Fig. 3. Depth image median filter

It consists of two parts: target region extraction of depth image and filtering the scene containing objects using median filter. As shown in Fig. 3, it is a bottle filtered before and after image. In Fig. 3(a), the black area is the noise of the depth image. And the Fig. 3(b) is a better depth image after median filtering.

Results in Real World. The other processes are the same which the process in simulation. 15 sets of point cloud data for each type of object were acquired by Kinect. Thus, there are a total of 135 sets of data. The primitive shape recognition results in physical world are shown in Table 2.

Table 2. The accuracy of shape recognition in physical world (%)

Objects	Bottle1	Drinkbottle	Bottle2	Box1	Box2	Rubik'scube	Golf	Tennis	Orange
Recognition accuracy rate of each object	60	100	86.7	93.3	100	93.3	86.7	100	100
Recognition accuracy rate of each shape	82.2			95.6			95.6		
Average	91.1								

Comparing Table 1 and Table 2, the accuracy of the primitive three dimensional shapes recognition of the target in the simulation is higher than that in the actual environment. In the physical world, the recognition accuracy is lower due to the noise of depth image and the point cloud of the object. Pleasantly, primitive shape recognition has both acceptable performance both in simulation and physical world. And the recognition accuracy of ball objects are better than the other two shapes.

4 Conclusions

A method of recognizing the object primitive shape using local point cloud is proposed by this study. First of all, a point cloud dataset with three primitive shapes is established via converting depth images into point cloud. Then, PointNet is used to train the data set. The recognition accuracy rate for objects of different sizes is up to 96% in simulation. In the actual environment, the accuracy rate of object shape recognition is about 91.1%.

In the future research, the authors would like to apply the primitive shape recognition for grasp process. The shape information of the object is used as the input information of the robot controller, which may contribute to a better grasping performance of the robot.

Acknowledgements. This work is supported by the projects of National Natural Science Foundation of China (No. 61873192; No. 61603277; No. 61733001), the Quick Support Project (No. 61403110321), and Innovative Project (No. 20-163-00-TS-009-125-01). Meanwhile, this work is also partially supported by the Fundamental Research Funds for the Central Universities and the Youth 1000 program project. It is also partially sponsored by International Joint Project Between Shanghai of China and Baden-Württemberg of Germany (No. 19510711100) within Shanghai Science and Technology

Innovation Plan, as well as the projects supported by China Academy of Space Technology and Launch Vehicle Technology. All these supports are highly appreciated.

References

1. Hogan, F., Grau, E., Rodriguez, A.: Reactive planar manipulation with convex hybrid MPC. In: Proceedings of IEEE International Conference on Robotics and Automation, pp. 247–253, Brisbane, Australia (2018)
2. Righetti, L., et al.: An autonomous manipulation system based on force control and optimization. Auton. Rob. **36**(1–2), 11–30 (2014)
3. Liang, H., et al.: PointNetGPD: detecting grasp configurations from point sets. In: Proceedings of IEEE International Conference on Robotics and Automation, pp. 3629–3635, Montreal, Canada (2019)
4. Mahle, J., et al.: Dex-Net 2.0: deep learning to plan robust grasps with synthetic point clouds and analytic grasp metrics. In: Proceedings of Robotics: Science and Systems, pp. 1–12, Massachusetts, USA (2017)
5. Mahler, J., et al.: Learning ambidextrous robot grasping policies. Sci. Rob. **4**(26), 1–12 (2019)
6. Tsai, J., Lin, P.: A low-computation object grasping method by using primitive shapes and in-hand proximity sensing. In: IEEE International Conference on Advanced Intelligent Mechatronics, pp. 497–502, Munich, Germany (2017)
7. Beltran-Hernandez, C., Petit, D., Ramirez-Alpizar, I., Harada, K.: Learning to grasp with primitive shaped object policies. In: IEEE/SICE International Symposium on System Integration, pp. 468–473, Paris, France (2019)
8. Lin, Y., Tang, C., Chu, F., Vela, P.: Using synthetic data and deep networks to recognize primitive shapes for object grasping. In: IEEE International Conference on Robotics and Automation, pp. 10494–10501, Paris, France (2020)
9. Eitel, A., Springenberg, J., Spinello, L., Riedmiller, M., Burgard, W.: Multimodal deep learning for robust RGB-D object recognition. In: IEEE/RSJ International Conference on Intelligent Robots and Systems, pp. 681–687, Hamburg, Germany (2015)
10. Carlucci, F., Russo, P., Caputo. B.: $(DE)^2$CO: deep depth colorization. IEEE Rob. Autom. Lett. **3**(3), 2386–2393 (2018)
11. Loghmani, M., Planamente, M., Caputo, B., Vincze, M.: Recurrent convolutional fusion for RGB-D object recognition. IEEE Rob. Autom. Lett. **4**(3), 2878–2885 (2019)
12. Charles, R., Su, H., Kaichun, M., Guibas, L.: PointNet: deep learning on point sets for 3D classification and segmentation. In: 34th IEEE Conference on Computer Vision and Pattern Recognition, pp. 77–85, HI, USA(2017)
13. Charles, R., Su, H., Kaichun, M., Guibas, L.: PointNet++. deep hierarchical feature learning on point sets in a metric space. In: 31st Annual Conference on Neural Information Processing System, pp. 5100–5109, Long Beach, CA, USA (2018)
14. Wu, W., Qi, Z., Fu, L.: PointConv: deep convolutional networks on 3D point clouds. In: 35th IEEE Conference on Computer Vision and Pattern Recognition, pp. 9613–9622, Long Beach, CA, USA (2019)