I-014

Estimating the Positions of Target Objects Based on Object Recognition by Processing 3D Point Cloud Data

Li Qi† Jun Ohya† Atsuo Takanishi† Takashi Matsuzawa† Kenji Hashimo‡

1. Introduction

In Japan, many nuclear projects were established from the 1980s. However, in 2011, the Fukushima nuclear accident occurred due to the earthquake and tsunami [1]. After that, Japan began to promote research projects on the tele-operation of disaster-responsive robots so as to further respond to nuclear leaks, preventing incidents, and so on.

At the DARPA Robotics Challenge (DRC) [2], a very detailed plan for disaster-responsive robots was made to solve complex problems we might face in the future. There are many robots that can be used as disaster response robots through tele-operation to achieve the above operations. For examples, the HRP platform robot [3]is a representative robot of Japan. It is mainly for industrial design, and for a human-machine cooperation. Jaxon [4] is a Japanese robot that was developed by Google in December 2013. There is the Chimp robot [5] invented by Carnegie Mellon University as well. Waseda university also developed a disaster response robot called WAREC [6]. It can crawl or walk upright by changing its posture. It can achieve the grabbing operation for the target object by tele-operation. However, the tele-operation system has some problems such as operator's difficulty in feeling the distance between the robot hand and the target object; therefore, how to solve these problems is still an open issue.

In order to let WAREC grasp object by tele-operation, it is necessary to estimate the relative position of the target object with respect to the robot hand. Therefore, the main purpose of this paper is to obtain the position information of the target object by processing the 3D point cloud data captured by depth sensor. In addition, the position transformation matrix of the model object with respect to the target object in the scene is obtained, so as to confirm the rotation information of the robot wrist during the grasping process.

2. Propose Method

The proposed method is briefly explained in Fig.1. First, we collect the original 3D point cloud data by Kinect V2, after preprocessing a model point cloud data and a scene point cloud data which contains several objects, after segmentation, the scene point cloud is divided into several individual objects. Next, we extract the feature points of both model object and scene objects, matching them to recognize the target object. Finally, we estimate the position and the transformation matrix by target object's data. In this paper, by processing the 3D point cloud data obtained by Kinect v2, the target harmer in the scene is first identified. Then the 3D coordinates (x, y, z) of the target hammer in the scene are obtained to describe the position information of the hammer. Then we calculate the transformation matrix that shows how

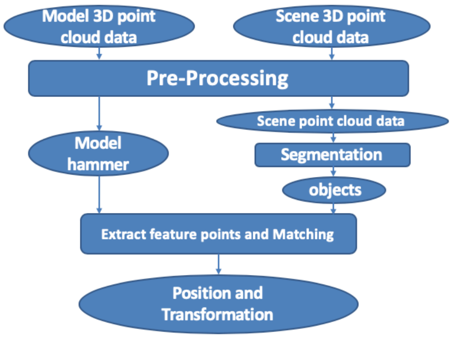


Fig.1 Procedure of Propose Method

much the transformed model hammer is registered to target hammer in the scene. By getting the position information of the target object, we can know the distance and direction of the target object with respect to the robot hand. In addition, how to determine the rotation of the wrist when grabbing an object can be derived from the transformation matrix from the target object to the model.

2.1Pre-Processing

In order to extract the objects without any noise, the pre-processing of original point cloud data is necessary.

2.1.1 Remove the Background of the 3D Point Cloud Data

In the original point cloud, the distant background and noises are generally away from the target object in the optical direction of the Kinect v2. Therefore, this paper first uses the pass-through filter that sets a threshold and removes the points on the optical axis that are farther than the threshold to initially remove the background, debris and other noise.

2.1.2 Remove the Desktop

The desktop is generally a plane in the point cloud. Using RANSAC (Random Sample Consensus), points in the 3D point cloud data that can be fitted to planes can be extracted and removed. As a result of applying RANSAC to search a plane, the inliers are points that can be fitted to the planar model (desktop). Then the inliers are removed.

2.1.3 Remove Outliers

This paper uses the statistical outlier removal filter to remove noise spots which exist sparsely in 3D space. The statistical outlier removal method is based on the calculation of the average of the distances between a point and the other points in the input data. The distance outside the threshold is defined as an outlier and can be removed from the data set.

2.2 Detecting the Target Hammer and Get the Position Information

Other objects and hammers are placed on the table together in the scene. Then the target hammer is identified by the following

† Department of Modern Mechanical Engineering, Waseda University
‡ Department of Mechanical Engineering Informatics, Meiji University

method, and the position information of the target hammer is obtained.

2.2.1 Segmentation

In the 3D point cloud data in the scene, in addition to the target hammer, there are point clouds that correspond to other objects. This paper uses Euclidean Cluster Extraction to segment the scene into each individual object.

2.2.2 Extracting the Descriptor of Point Features

After the segmentation, we have point cloud data corresponding to the model hammer, the segmented target hammer, and other objects. Next, we match the model hammer and the segmented point cloud separately to determine which point cloud in the scene corresponds to the hammer.

For each individual point in the point cloud, we can describe the characteristics of this point by the relationship between this point and its neighbors, such as the angle between the normal and the distance. The characteristics of all points in the point cloud can be described by FPFH (Fast Point Feature Histograms). In FPFH, relationship between two points can be expressed by relationship between the normal vectors to the two points.

2.2.3 Matching Descriptors

For the descriptor, the four elements are $\alpha, \emptyset, \theta, d$, which spans a 4-dimantional system. The proposed method uses KNN (k-nearest neighbor) algorithm to search for the matching relationship of descriptors in two 3D point cloud data.

2.2.4 Location of the Scene Hammer

The point cloud in the split point cloud which corresponds to the hammer is determined. Then the position information on the point cloud is that of the target hammer in the scene. The position information of the hammer is represented by the coordinates of its centroid. The average of the coordinate values of all the points in the 3D point cloud data, is the position information of the hammer in the scene.

2.3 Transformation Matrix

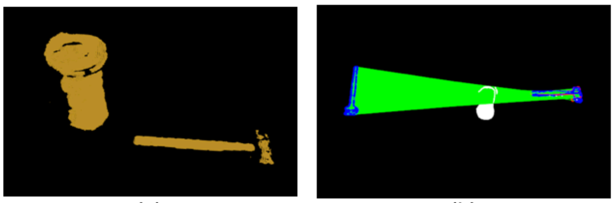
To obtain how the model hammer is transformed to the target hammer in the scene, awe need to confirm the transformation matrix of the relation between them. The transformation matrix is able to be calculated by applying SVD (singular value decomposition) theory.

3. Experiments and Evaluation

In order to evaluate the accuracy of this proposed method, an experiment was settled which used Kinect V2 to take the point cloud data, and a hammer and a bottle were placed on the table.

Fig.2 shows the result of this experiment. (a) is the point cloud data after preprocessing, which does not contain noises. (b) shows how the hammer be recognized and transformed by the rotation and translation matrix. The Blue points indicate the Points of the model hammer and the target hammer. The red points indicate Transformed model hammer. And the Green lines indicate the Matching between the model hammer and target hammer. (c) shows the coordinate and distance of hammer, and the transformation matrix. Red circle part is the rotation matrix and the blue circle part is the translation matrix.

The propose method in this paper still sometimes mis-recognizes and not-recognizes objects. The results of all experiments show the average relative error of estimating the position is



(a) (b)

X [m]	Y [m]	Z [m]
-0.0187	0.0828	0.8649

True d [cm]	Measured D [cm]	Absolut E [cm]	Relative e [%]
86.5	86.9	0.4	0.462

R			t
0.0533279	-0.832034	0.552156	-0.448306
0.222731	0.548924	0.805651	-0.66423
-0.97342	0.0800185	0.214593	0.669404

(c)

Fig.2 Results of Experiment

0.835%, which is within the required accuracy, and it can correctly obtain the rotation matrix.

4. Conclusion and Future Work

This paper has proposed an effective method for estimating the positions of target objects, which is processing the 3D point cloud data and then recognize the target object, finally confirm the position information which are useful for supporting grasping.

In the future, it is necessary to improve the method for extracting feature points and the method of matching feature points in order to increase the accuracy of recognition. In addition, using deep-learning to recognize objects is also a possibility.

Reference

[1] “Fukushima accident,” ENCYCLOPAEDIA BRITANNICA, https://www.britannica.com/event/Fukushima-accident (2018).

[2] “DARPA Robotics Challenge Features Disaster-response Tasks,” U.S. DEPARTMENT OF DEFENSE, https://www.defense.gov/News/Article/Article/604236/ (2018).

[3] Kenji Kaneko, Fumio Kanehiro, Shuuji Kajita et al., “Humanoid Robot HRP-2,” in IEEE International Conference on Robotics & Automation New Orleans, LA, (2004).

[4] Kunio Kojima, Tatuhi Karasawa, Toyotaka Kozuki, et al., “Development of life-sized high-power humanoid robot JAXON for real-world use,” in IEEE-RAS 15th International Conference on Humanoid Robots (2015).

[5] “Meet Chimp, A Disaster Response Robot with Four-Limb Drive,” POPULAR SCIENCE, https://www.popsci.com/meet-robots-designed-disaster-response (2018).

[6] “Development of Disaster Response Robot for Extreme Environment,” Takanishi Lab., WASEDA University, http://www.takanishi.mech.waseda.ac.jp/top/research/rescue/index.htm (2018).