Picking up an Unknown Object Through Autonomous Modeling and Grasp Planning by a Mobile Manipulator

Kimitoshi Yamazaki[1], Masahiro Tomono[2], and Takashi Tsubouchi[3]

[1] The Univ. of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo, Japan
yamazaki@jsk.t.u-tokyo.ac.jp
[2] Toyo Univ., 2100 Kujirai, Kawagoe, Saitama, Japan
tomono@eng.toyo.ac.jp
[3] Univ. of Tsukuba, 1-1-1 Tennodai, Tsukuba, Ibaraki, Japan
tsubo@roboken.esys.tsukuba.ac.jp

Summary. This paper describes a novel framework for object picking and carrying task by a mobile manipulator. Conventionally, researches on mobile manipulator cope well with object manipulation task with utilizing predefined knowledge or specific tools. So these researches have an essential problem that a new target object cannot be added without relatively many preparation. On the other hand, in our framework of a robot system, because the robot can create the 3D shape model of the object and can plan a grasp pose for the object autonomously under the condition that only the position of the object is given to the robot. Experimental results show the effectiveness of our robot system to be implemented our proposed method.

1 Introduction

Issues on mobile manipulators are being actively studied. such robots can be utilized in a large variety of tasks such as carrying objects from one place to another. Conventionally, there are several researches to cope well with such challenging tasks with utilizing predefined object models [4] [5], ID tags [1] or QR codes [3]. In these issues, how to give or what kind of information of the object for grasp management is focused on. So these researches have an essential problem that a new target object cannot be added without a heavy help by manually or a special tools.

We proposed approaches which differs from conventional researches. In our approach, there are two special policies for autonomous working. That is, an image based modeling method for creating object dense model [7] was developed. Moreover, a planning method for detecting grasp pose by inputting the model [8] was developed. Utilizing these functions related to object manipulation, it is expected that the robot will be capable of picking up and carrying task to unknown objects with giving only a little information by manual.

The measure purpose of this paper is to propose a solution for the picking up and carrying task based on automatic object modeling. We aim to develop a framework of a robot system which can handle small objects which generally exist in home or office environments with following conditions:

- No additional information on the object and environment are needed.
- No such information as the object shape or how to grasp it are needed.
- 3D grasping is performed through searching the best solution from many other feasible solutions.
- The object models and planning results can be easy reused if these have already acquired at once.

That is, the robot can add its target objects without constraint of shapes or types except one constraint that the object has some texture on its surface for object modeling in our framework.

2 Assumed Task and Research Issues

2.1 Assumed Task

Object carrying task by a mobile manipulator is performed according to the following procedure:

1. The object is put on a desk and the position of the object is given to the robot manually.
2. The robot moves to the desk, and
3. picks up the object by the manipulator, and
4. conveys the object and hands it to a person.

The item 3. are a main topic in this paper.

In this task, we assume that there are no constraint on object shape and no tags or marks on the object. In the meantime, we assume that relatively

much natural texture can be found on the object surface, and the object has equivalent size that human can grasp it by one hand.

Because Jaw Gripper hand which is a compact and light weight hand has an advantage for a mobile manipulator, such hand is utilized for grasping the object. We assume that the finger tips contact with the object with some area and not with at a point.

2.2 Issues and Approach

There are two important challenges in our policy. The one is that the robot creates an object model autonomously. Utilizing multi-viewpoint images which are captured from a single camera mounted on a manipulator, the robot creates an object model as dense 3D points. The other is that the robot detects a grasp pose autonomously. Utilizing the automatic created model, grasp poses of the manipulator are planned automatically. That is, only the position information of the object is needed to the robot at manual instruction.

It is important that between modeling and grasp planning should be connected by a proper model representation. For achieving it, we apply a model representation named "oriented points". The object model is represented as 3D dense points that each point has normal information against object surface. This representation has an advantage to autonomous modeling because of its simple data structure. On the other hand, another advantage is found in grasp planning because the normal information enables to plan grasp poses effectively. In the meantime, sufficient countermeasures are needed against the shape error of the object model which is obtained from a series of images. As one of the countermeasure, the area contact in the grasp planning is performed to cancel the difference of the model shape and real shape of the object. The object modeling method is described in section 3, and the grasp planning method is described in section 4.

3 Object Modeling

3.1 Approach to modeling

To acquire 3D model, SFM (structure from motion) method [7] is utilized. The 3D model means a reconstructed object surface which is represented by dense 3D points. Such model has simple data structure so that the model can be acquired by camera or LRF sensor autonomously without complex

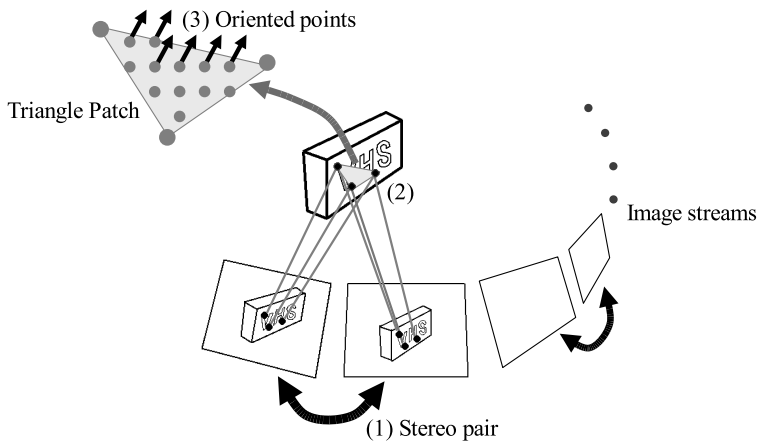


Fig. 1. Surface model reconstruction

calculation or knowledge based data transform. Moreover, it can represent the various shapes of the objects.

We take an approach to create the object model by utilizing image streams which capture the object from multiple views. Because the object model created by our method can have both 3D shape information and photometric information, it is expected that the model will be highly reusable for the purpose of object recognition by the robot.

3.2 Modeling procedure

Figure 1 shows modeling outline. The model is acquired according to following procedure: first, feature points in images are extracted and tracked from a small area which has strong intensity by using KLT-tracker [6]. From these points, object sparse model and camera poses are acquired by means of SFM. Next, the object dense shape is reconstructed from a close pair of images (Figure 1, (1)). Image pixels in each triangle patches (Figure 1, (2)) whose three vertices are adjoining and are common feature points in the pair of the images. At the same time, normal information is added to each reconstructed pixels (Figure 1, (3)) This process is applied to all the pairs of images and all the results for 3D locations of the triangular patches are integrated to obtain the 3D dense object model.

As a result, quite a number of points are reconstructed in online. Unfortunately, these are caused of consuming abundant memory because of the data

redundancy. To solve the problem, we apply to translate the model to voxel representation by means of a method described in section 5.

4 Grasp Planning

The purpose of grasp planning is to find reasonable grasp pose with making use of the reconstructed dense 3D model on the spot.

4.1 Approach to grasp planning

Grasp planning in this study has two major issues: (1) how to plan a grasp pose efficiently from the 3D model, and (2) how to find a stable grasp from the 3D model, even though the 3D model has redundant data and errors in shape. We propose grasp planning to answer the above two issues.

For detecting the best grasp pose to pick up the object, we take an approach to evaluate by utilizing two criteria. The one is that contact area between the hand and the object model. The other is that a gravity balance depending on grasp position on the object.

4.2 Planning method

In our planning, "stable grasp" is evaluated by the lowest sum total of three functions as follows [8] :

$$F = w_1 F_1(P_1, \mathbf{x}_h^o) + w_2 F_2(P_1, \mathbf{x}_h^o) + w_3 F_3(P_1, \mathbf{x}_h^o, \mathbf{x}_G, map) \tag{1}$$

where P_1 is a center point of finger plane on the hand. $\mathbf{x}_h^o$ is a hand pose in object coordinates, $\mathbf{x}_G$ is the pose of a mobile platform in world coordinates, and *map* is environmental map which is given in advance. w_i is a weight.

$F_1(\cdot)$ represents the function of contact area between the hand and the object. The evaluation value become smaller if the hand pose has more contact area. $F_2(\cdot)$ represents the function of a gravity balance of the object. The evaluation value become small if a moment of the object is small. $F_3(\cdot)$ represents the function of the grasping pose. The evaluation value becomes small if manipulability related to grasp by the mobile manipulator is large [9]. The policy of grasp planning is to find P_1, $\mathbf{x}_h^o$ and $\mathbf{x}_G$ which minimize the function of F.

Oriented points touched to P_1 are selected in order, grasp poses are evaluated utilizing eq.(1). In this planning, it is important to reduce vain contact

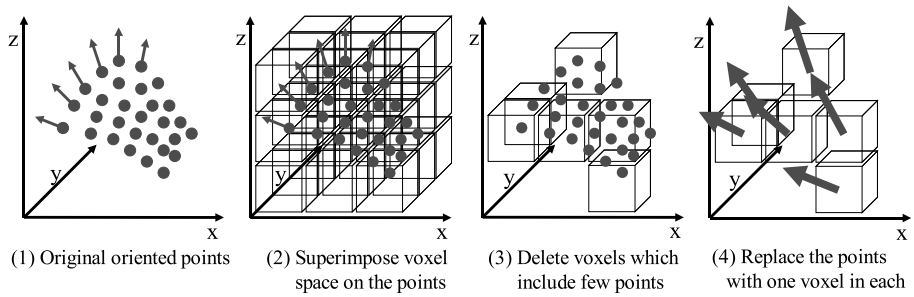


Fig. 2. Surface model reconstruction

between finger and the object model and to select oriented points which can have good evaluation. The former can be achieved to restrict the direction of the contact by utilizing normal information of each point because the normal indicates the shape of object surface. The latter is described in next Section.

5 Model Representation for Efficient Implementation

To improve the performance of our framework, we take an approach to translate an object model into voxelized representation. A voxel in this paper means a small cubic bin, so an object model is represented by a group of voxels.

5.1 Voxel Representation

Oriented points described in section 3 has redundant data for grasp planning. By transforming from these points to voxelized model, redundant data can be reduced. Moreover, this process has effects of reducing the shape error of the object model. The size of voxel is set with 2mm or 5mm based on an allowable shape error in the grasp planning.

A new voxelized model is acquired as "thin" model according to the following procedure. At first, 3D-space is squared up fine cubic voxels (Figure 2, (2)). Voxels are deleted except special voxels which include several oriented points in it (Figure 2, (3)). Next, orientation component is averaged in each voxel (Figure 2, (4)). Finally, oriented points are acquired as each position is the center of each voxel and the orientation is an average value.

As it is necessary to yield the moment of inertia of the object for calculation of $F_2(\cdot)$, the model must be volumetric. For this purpose, once a voxel space including all the part of the model is defined. Then, the voxels of outside of the

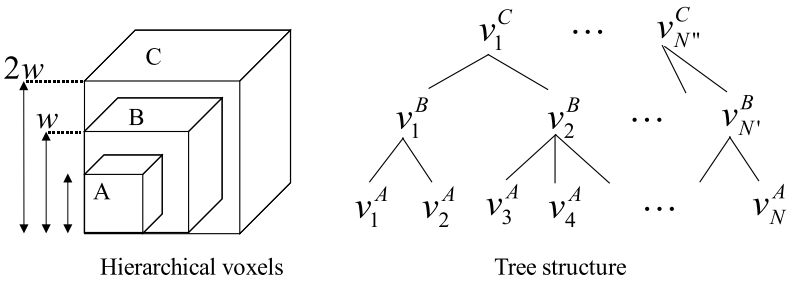


Fig. 3. Hierarchy model representation

object are pruned away. The reminder of the voxels is the volumetric model of the object. This process consumes few time. So this is one of the advantage of voxelized model.

5.2 Hierarchical Data Representation

The method as mentioned in 5.1 can reduce the number of grasp pose searching. However, the searching has potential to be still capable of improving. We notice that there are somewhat points which obviously need not to be checked by eq.(1). From this reason, we propose hierarchical data representation which can exclude points before judging the quality of grasp pose. Using the new formed model, the searching can be performed at some parts of object model where will have rich contact area with fingers.

The hierarchical representation is similar to octree. Octree is often used for judging collision in the field of computer graphics. The transformation procedure is as follows: at first, initial voxels which construct original voxelized model are set hierarchical A. Next, other voxel space which is constructed w times larger voxels than hierarchical A is superimposed on the voxels of hierarchical A. A new model is represented by the larger voxels which are set hierarchical B. In this processing, only voxels of hierarchical B are adopted if these voxels include over predefined threshold about the number of voxels of hierarchical A which have similar orientation. The same hierarchy construction is performed from hierarchical B to hierarchical C, too. As a result, one voxel of hierarchical C includes several voxels of hierarchical A as shown in Figure 3. Because these voxels of hierarchcal A exist in neighbor and has similar orientation, the area can be expected that it supplies rich contact area with finger.

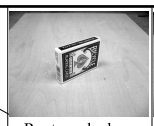
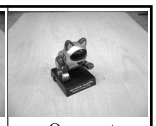
Types of objects / Results	 Rectangular box	 Cup	 Toy	 Stapler	 Ornament
Voxels	807	986	966	934	1531
Grasp pose Candidates	64	37	49	48	58
Processing time (A) [sec]	0.96	1.01	0.94	0.86	1.36
Processing time (B) [sec]	9.3	6.3	6.7	7.2	16.3

Fig. 4. Trials to several shapes of objects

In the grasp pose searching, voxels of hierarchical C are selected in order and evaluate by eq.(1) about inner voxels which belong to hierarchical A. This can achieve efficient searching with selecting only voxels which are guaranteed good evaluation result about contact area.

6 Experiments

6.1 Trials to Several Shapes of Objects

Object modeling and grasp planning are tried to several shapes and types of objects. Figure 4 shows the results. There are five objects which have relatively rich texture on its surface. The processing time of these object modeling took about 100 sec through 134th images in offline.

Notice that the planning results are not related to object shape. In these results, dozens of grasp poses can be found from each created models about 1 second (Pentium M, 2.0 GHz) as shown in processing times (A). Processing times (B) as shown in Figure 4 indicate the time of no utilizing hierarchy data representation describe in section 5.1. The more performance of the planning is achieved by utilizing our proposed method.

On the other hand, some problems are cleared up through this experiments. For instance, because an area where has no texture cannot reconstructed by our modeling method, grasp poses which touch to inner of the cup are not selected in grasp planning.

6.2 Implementation to Assumed Task

Figure 5 shows a mobile manipulator system to be used and an experimental environment. A small PET bottle is set on a desk as a target object. The goal

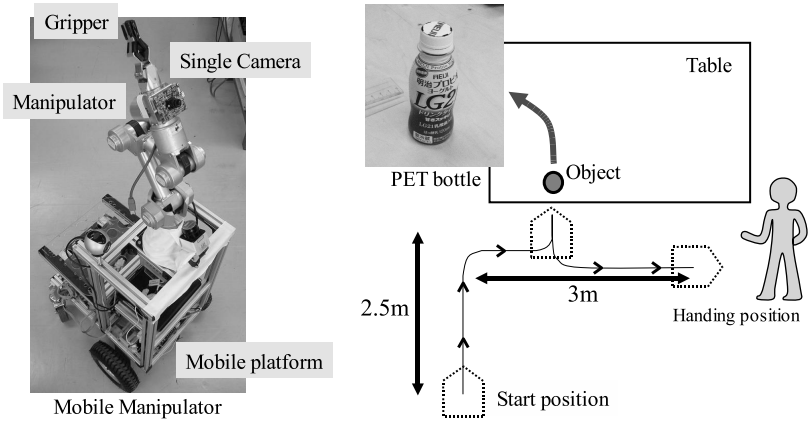


Fig. 5. Settings

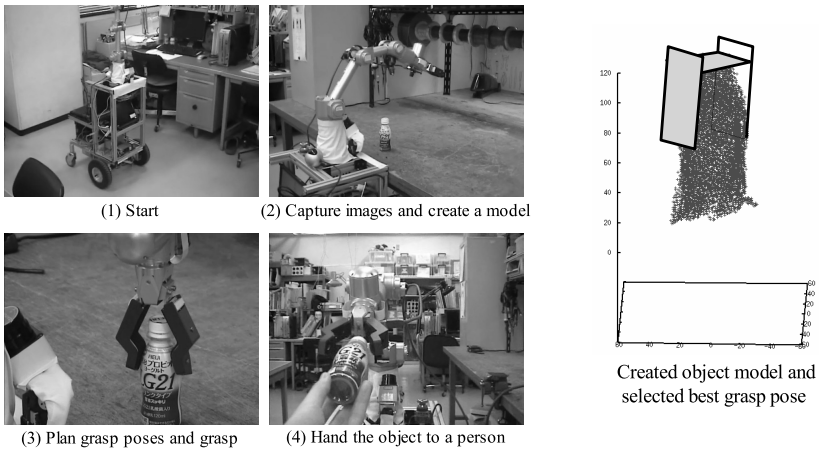


Fig. 6. Experimental result

of this experiment is to hand the object to a person who stands another place from the object position.

Environment map which includes the position of the object was given in advance. Moreover, the initial position of the robot and the position of the person were given in advance, too. In this condition, the robot planned its motion trajectory automatically by using artificial potential method [2].

Figure 6 shows experimental result. The robot (1)moved to the front of the desk, (2) created object model over capturing images, (3) grasped the object by applying the result of grasp planning, (4)moved to the front of a person, and

handed it. All processing was performed by two laptop computers (Pentium M, 2.0GHz, Celeron, 1.0GHz) on the robot. As a result, image capturing and object modeling was needed about 360 sec and grasp planning was needed 0.65 sec in this experiment. In this experiment, final grasp pose is detected from only the sum of $F_1(\cdot)$ and $F_2(\cdot)$, and the inverse kinematics is calculated instead of $F_3(\cdot)$.

7 Conclusion

In this paper, a mobile manipulator system which can pick up an object in real environment autonomously was introduced. In this system, because automatic object modeling and grasp planning are applied with having proper data relation, object picking and carrying task can be achieved under the condition that only the position of the object is given to the robot.

References

1. N. Y. Chong and K. Tanie: "Object Directive Manipulation Through RFID", Proc. of Int. Conf. on Control, Automation, and Systems, pp.22–25, 2003.
2. C. Connolly, J. Burns and R. Weiss: "Path Planning using Laplace's Equation", Proc. of the IEEE Intl. Conf. on Robotics and Automation, pp.2102–2106, 1990.
3. R. Katsuki, J. Ohta, T. Mizuta, T. Kito, T. Arai, T. Ueyama and T. Nishiyama: "Design of Artificial Marks to Determine 3D Pose By Monocular Vision", Proc. 2003 IEEE Int. Conf. Robotics and Automation, pp.995–1000, 2003.
4. J. Miura, Y. Shirai and N. Shimada: "Development of a Personal Service Robot with User-Friendly Interfaces", 4th Int. Conf. on Field and Service Robotics, pp.293– 298, 2003.
5. K. Nagatani and S. Yuta: "Autonomous Mobile Robot Navigation Including Door Opening Behabior-System Integration of Mobile Manipulator to Adapt Real Environment-", Intl. Conf. on FSR, pp.208–215, 1997.
6. J. Shi and C. Tomasi: "Good Features to Track", IEEE Computer Society Conf. on Computer Vision and Pattern Recognition, pp.593–600, 1994.
7. K.Yamazaki, M.Tomono, T.Tsubouchi and S.Yuta: "3D Object Modeling by a Camera mounted on a Mobile Robot", Proc. of the 2004 IEEE Int. Conf. Robotics and Automation, 2004.
8. K.Yamazaki, M.Tomono, T.Tsubouchi and S.Yuta: "A Grasp Planning for Picking up an Unknown Object for a Mobile Manipulator", Proc. of the 2006 IEEE Int. Conf. Robotics and Automation, 2006.
9. K.Yamazaki, M.Tomono, T.Tsubouchi and S.Yuta: "Motion Planning for a Mobile Manipulator Based on Joint Motions for Error Recovery", Proc. of the 2006 IEEE Int. Conf. Intelligent Robots and Systems, pp. 7–12, 2006.