2015 15th International Conference on Control, Automation and Systems (ICCAS 2015)
Oct. 13-16, 2015 in BEXCO, Busan, Korea

CAD-Based Pose Estimation for Random Bin-Picking of Multiple Objects Using a RGB-D Camera

Cheng-Hei Wu, Sin-Yi Jiang and Kai-Tai Song*

Institute of Electrical Control Engineering, National Chiao Tung University
Hsinchu, Taiwan, R.O.C. (ktsong@mail.nctu.edu.tw) * Corresponding author

Abstract: In this paper, we propose a CAD-based 6-DOF pose estimation design for random bin-picking of multiple different objects using a Kinect RGB-D sensor. 3D CAD models of objects are constructed via a virtual camera, which generates a point cloud database for object recognition and pose estimation. A voxel grid filter is suggested to reduce the number of 3D point cloud of objects for reducing computing time of pose estimation. A voting-scheme method was adopted for the 6-DOF pose estimation a swell as object recognition of different type objects in the bin. Furthermore, an outlier filter is designed to filter out bad matching poses and occluded ones, so that the robot arm always picks up the upper object in the bin to increase pick up success rate. A series of experiments on a Kuka 6-axis robot revels that the proposed system works satisfactorily to pick up all random objects in the bin. The average recognition rate of three different type objects is 93.9% and the pickup success rate is 89.7%.

Keywords: Industrial robot, random bin-picking, 6-DOF pose estimation.

1. INTRODUCTION

Robotic bin-picking is often needed in automated assembly lines. Vision-based bin-picking has been studied for many automated production systems. One of the main challenges in such tasks is to recognize and localize different type workpieces, which are occluded each other in the bin. Pose estimation using 2D images has been widely used in bin picking design. Choi *et. al.* used 2D keypoints feature matching to achieve robotic manipulation of 3D object recognition and tracking [1]. Ulrich *et. al.* proposed a method to apply 2D edge-based matching to eliminate object edges and pyramid level method to search for maximum similarity 2D model for estimating object 3D poses [2]. It is clear that pose estimation use 2D image in robotic automatic random bin-picking systems only work when their poses are limited to a few degrees of freedom. In recent years, 3D sensors have become more cost effective, researchers are motivated to develop robust 6-DOF pose estimation systems by using 3D data.

Model-based approaches are very often used to recognize different type objects [3]-[4]. This approach needs a 3D digital model of the target, and then a method finds the best pose estimation by matching a 3D model to a real scene image. Drost *et. al.* combines a voting scheme with hash table, which is a global model descriptor to recognize free-form objects and their 6-DOF poses [5]. Choi *et. al.* developed a family of pair features by using oriented surface points, oriented boundary points, and boundary line segments in a voting framework for practical robotic bin-picking systems [6]. Skotheim *et. al.* generated workpieces databaseby using 3D CAD model, and a Hough-like voting scheme was set up to recognize three different components [7]. In [8], a method was proposed to pick up the earliest feasible point of moving objects using a search-based algorithm. Atanasov *et. al.* developed a method to detect semantically important objects and estimate of their pose for grasping using a RGB-D camera [9]. In [10], the authors propose a method to deal with object pink up in a complex and multi-objective scene with arbitrary object positions based on a probability approach to the estimation of 6D poses. It is observed that many useful tools have been investigated for pose estimation and object picking using 3D data, however it is still challenge to recognize and localize different type workpieces in a bin more effectively. It deserves urgent attention to solve the problem of robust random bin-picking using industrial robots.

In this paper, we proposed a CAD-based 6-DOF pose estimation design for random bin-picking of multiple different objects by using a Kinect RGB-D sensor. We integrate 3D CAD data of a workpiece with a virtual camera [11] to generate CAD-model database for pose estimation system. A series of experiments of practical random bin-picking of multiple different objects reveals that our proposed system can pick up all random pose of objects until there is no objects left in a bin.

The rest of this paper is organized as follows. Section 2 describes the proposed system architecture of automatic random bin-picking of multiple different objects. Section 3 describes the proposed 6-DOF pose estimation system. Section 4 presents the design of the bin-picking system of Kuka industrial robot. In Section 5, the experimental results are presented to verify the performance of the proposed 6-DOF object pose estimation design. Section 6 summarizes the contribution of the paper.

2. 3D CAD-MODEL DATABASE

The system architecture of the proposed bin-picking system is shown in Fig. 1. It consists of three main parts: (1) CAD-model database system, (2) 6-DOF pose estimation module, and (3) target selection system. In the offline phase, a CAD-model database is generated by using 3D CAD data of different type objects. For online phase, the Kinect sensor captures RGB-D images and generates point clouds of the objects. A voting-scheme is applied such that the pose estimation

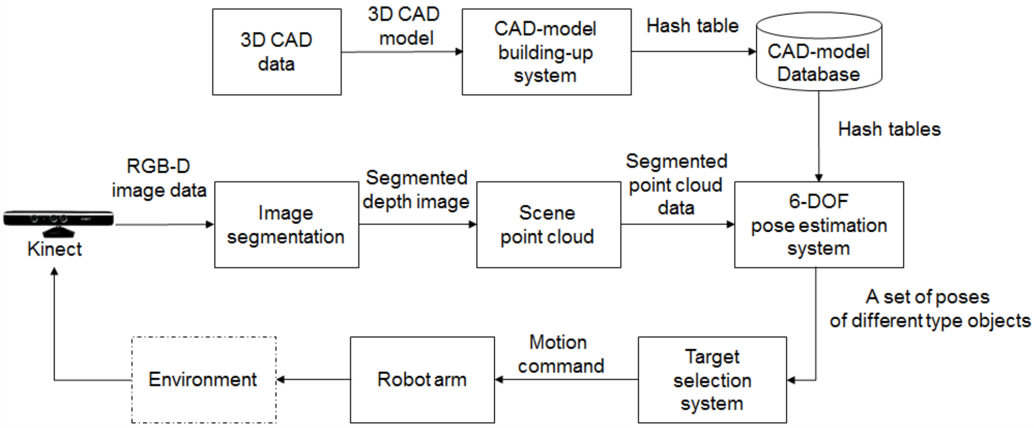


Fig. 1. System architecture of robotic random bin-picking.

system estimates the type and 6-DOF pose of different objects in the bin. The target selection system selects the best-to-pick-up target from all recognized objects, and determines the most suitable pose of the robot gripper to pick up the target.

We take advantage that 3D CAD data can be generated rapidly and easily to construct the CAD-model database for the 6-DOF pose estimation. For acquiring useful data from a 3D CAD data to generate CAD-model database, a simulation system has been built to integrate 3D CAD data of workpieces to a virtual camera [11]. In this process, we first use the virtual camera system to generate depth images of the workpiece, and then converte these depth images into point cloud data. In this system, we adopt the voxel grid filter [7] to reduce the number of the point cloud of 3D CAD data for reducing computing time of pose estimation. The next step is to estimate surface normal for each point cloud of 3D CAD data. For building a global model description of 3D CAD data, we compute all point pair features descriptor from the point cloud data of 3D CAD data, and then group them in a hash table [6]. The hash table is a 3D CAD model, and it is stored in the CAD-model database for pose estimation system.

2.1 Virtual Depth Camera

For acquiring useful data from 3D CAD model to generate CAD-model database, we employ a virtual depth camera to generate a depth image from a 3D CAD data of workpiece. The pose of the virtual depth camera is set on a virtual sphere, and a 3D CAD model is set at center of the virtual sphere. We define the parameter of camera pose to allow the virtual depth camera to capture a depth image from desired orientation of the 3D CAD data. The procedure is shown in Fig. 2.

2.2 Voxel Grid Filter

Since the amount of 3D CAD data is very large, we apply a voxel grid filter to decrease the amount of points in the point cloud. The voxel grid filter works to create a 3D voxel grid with a specified unit size over the point cloud and by replacing all the points that fall inside each voxel with the centroid of these points. Figure 3 shows

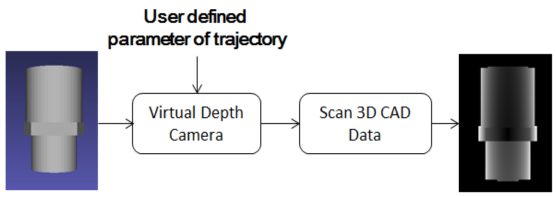


Fig. 2. Procedure of using the virtual depth camera. A depth image of 3D CAD data is captured by using virtual depth camera with user defined trajectory.

an example that the number of the point cloud of 3D CAD data is considerably reduced by using voxel grid filter, but shape of the point cloud is kept. One of the advantages of the voting-scheme based matching algorithm is that it tends to work well even for sparse point clouds. The voxel grid filter allows system to increase the pose estimation speed, without significant decrease the recognition rate of pose estimation.

2.3 Hash Table

To build a global model description of 3D CAD data, we compute all point pair features descriptor from the point cloud of 3D CAD data and group them in a hash table [6]. The hash table of an object consists of feature pairs of 3D CAD data, and these feature pairs are saved in a CAD-model database for 6-DOF pose estimation. These CAD mode are represented by a set of point-pair features with similar feature descriptors which are grouped together in the same slot. The feature descriptors $F_m(m_r, m_i)$ of the CAD model will be matched by those from scene feature descriptors

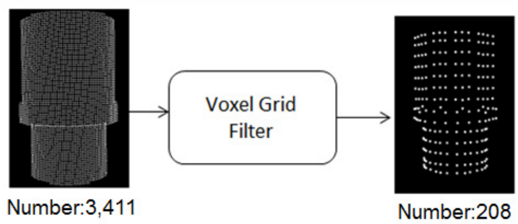


Fig. 3. One example of voxel grid filter. The number of point cloud of the 3D CAD data decrease from 3411 points to 208 points by using the voxel grid filter.

$F_s(s_r, s_i)$ using the hash table.

3. POSE ESTIMATION DESIGN

The 6-DOF pose estimation system includes a workpiece recognizing and localizing. Input to the system is a scene depth image, and output is a set of 6-DOF poses of different type objects. The pose estimation module consists of six steps: i) converting scene depth image into scene point cloud, ii) using voxel grid filter to decrease the number of scene point cloud, iii) estimating surface normal for scene point cloud, vi) estimating 6-DOF pose of scene point cloud by using voting-scheme matching algorithm and CAD-model database built in offline phase, v) removing isolated 6-DOF poses with low scores by using pose clustering algorithm, and vi) increasing accuracy of 6-DOF poses by using ICP algorithm to get refined 6-DOF poses.

3.1 Voting Scheme Matching

The voting scheme matching algorithm of [6] is adopted in this design. The algorithm consists of an offline phase and an online phase. In the offline phase, a global model descriptor is built by using the 3D CAD data. The global model descriptor is a hash table that allows feature descriptor matching to quickly search model points pair with a given scene points pair feature descriptor. In the online phase, all the points in the scene are checked and select one by one as the reference point s_r. First, each reference point s_r is the paired against every other point s_i in the scene, and its feature descriptor $\mathrm{F}(s_r, s_i)$ is calculated. Second, the feature descriptor $\mathrm{F}(s_r, s_i)$ used as a key to the hash table to search for point pairs (m_r, m_i) that have a similar distance and normal orientation as (s_r, s_i). Third, for every point pairs (m_r, m_i), we use local coordinates to aligned point pairs (m_r, m_i) with a scene point pairs (s_r, s_i). The transformation from the local model coordinates to the scene coordinates is defined as below:

$$s_i = T_{s\to g}^{-1} R_x(\alpha) T_{m\to g} m_i \qquad (1)$$

We first use transformation $T_{s\to g}$ and $T_{m\to g}$ to align the points s_r and m_r, and align their surface normal n_r^s and n_r^m to be parallel to the x axis. The final step employs $R_x(\alpha)$ to rotate the oriented model point m_i along the x axis with an angle α to align with oriented scene point s_i. Each (m_r, α) is termed as local coordinates of the model with respect reference point s_r. Finally, we construct a two-dimensional accumulator table, one axis with the model point number, $r = 1, 2, \ldots, \mathrm{N}$, and the other axis with a discretized set of angles, $\alpha = 1, 2, \ldots, \mathrm{M}$. A vote is cast for the local coordinates (m_r, α). After all points s_i are processed by using voting-scheme matching, all peaks in the accumulator table that gained a high number votes are the pose estimation results.

3.2 Pose Clustering

The voting scheme matching result is a set of 6-DOF poses, each with an associated number of votes. A pose clustering procedure is applied to identify clusters of 6-DOF poses that are close together in translation and rotation for filtering out error poses to increase the accuracy of the pose estimation result [6]. All 6-DOF poses are clustered such that all poses in one cluster are similar as each other. The score of one cluster is the sum of the scores of all the contained poses which scores are gained in the voting scheme matching. Finally, one cluster with the maximum score is obtained. The resulting pose is calculated by averaging the poses contained in the cluster. Pose clustering increases accuracy of the 6-DOF pose estimates obtained by using voting scheme matching.

3.3 ICP Refinement

After we obtain a set of coarse 6-DOF poses by using pose clustering, the coarse poses can be further refined by using an iterative closet point (ICP) algorithm [12]. The ICP algorithm is employed to minimize the distance between two clouds of points. First, we render the point cloud of 3D CAD data using the current coarse 6-DOF pose. Second, we compute the closet point in the scene point cloud for each point in the 3D CAD data. Finally, we update the pose estimate result to obtain refined 6-DOF pose. After ICP refinement, we obtain the registration error which is given by the average distance between the 3D CAD data and scene points. If the registration error is small, the refined pose is accuracy.

4. ROBOTIC PICK UP MOTION PLANNING

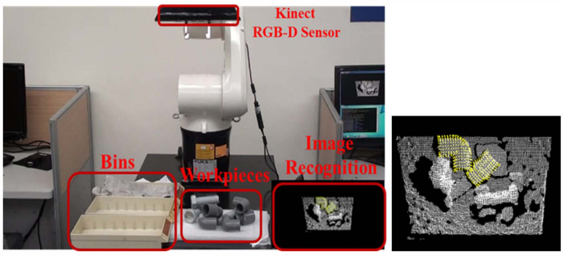
After obtaining a set of 6-DOF poses of different type objects from 6-DOF pose estimation system , the robot needs to determine which recognized object is the best-to-pick-up target, and what pose of robot gripper is the most suitable to pick up the random pose target. A target selection system is designed to resolve this problem. First, the scene point cloud and a set of point clouds of recognized CAD models are used to filter out bad matching poses and occluded ones by using outlier filter and then the best-to-pick-up target is selected by finding the highest value of z axis of each central point of retained point clouds of recognized CAD models. The grasp pose of robot gripper is decided by figuring out z axis rotation of the target relative to robot gripper.

4.1Grasp Pose Determination

For selecting the best-to-pick-up target, we compute 3D centroid points of all recognized workpieces, and find the maximum value of Z-axis of them. The maximum value of Z axis of the target is the highest workpiece in the bin, so the robot gripper to grasp the target is not be obstructed by the other workpieces. Different type workpieces have different shapes. For robot gripper to grasp a workpiece by using correct pose. The yaw rotation angle of target relative to robot gripper is determined to generate the grasp pose of robot gripper.

5. EXPERIMENTAL RESULTS

The setup of our random bin-picking system shown in



(a) (b)

Fig. 4. System overview. (a) Setup of the random bin-picking system. A Kinect RGB-D sensor attached at the end-effector of a Kuka 6-axis industrial robot arm. (b) A 6-DOF object pose estimation result.

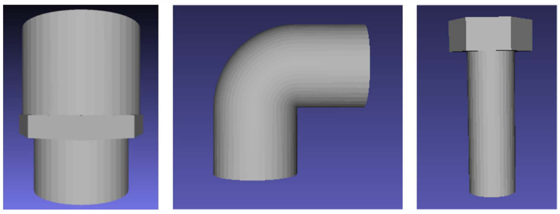


Fig. 5. 3D CAD data of three different type workpieces. From left to right: Object 1, Object 2, Object 3.

Fig. 4. In the experiments, three objects were used, as shown in Fig. 5. Our program runs on a PC with an Intel Core i5-2400 CPU (3.1 Ghz) and 4 GB RAM. The operating system of PC is Win 7. Our system uses a Kinect RGB-D sensor which attached on the end-effector of a Kuka 6-axis industrial robot to recognize a set of 6-DOF poses of different types of objects randomly placed in a bin. The Kinect RGB-D sensor has been calibrated with respect of the robot arm, so that the robot can pick up an object using the estimated 6-DOF pose.

5.1 Accuracy Test for Pose Estimation

To verify the accuracy of the random bin-picking system, we used a single object placed on a plane in this experiment. The Kinect sensor captured depth image and converted it into scene point cloud. The image segmentation method was implemented to obtain point cloud of the object from scene point cloud. Its 6-DOF pose was then estimated by using the CAD model. To obtain translational error, we subtracted the center point of object segment and the recognized CAD model. To obtain rotational error, we manually selected two reference points of the object segment and computed its vector. We then chose the same point at the same place from the recognized CAD model and computed its vector. The rotational error is obtained by computing the angle between these two vectors.

In this experiment, we placed random pose of three objects on a plane 50 times, and computed their average absolute translational error and rotational error. The average absolute errors of pose estimation are shown in Table I. The pose estimation results give average absolute errors of less than 1.5mm for all translations and less than 4.6° for rotations.

5.2 Random Bin-Picking Experiment

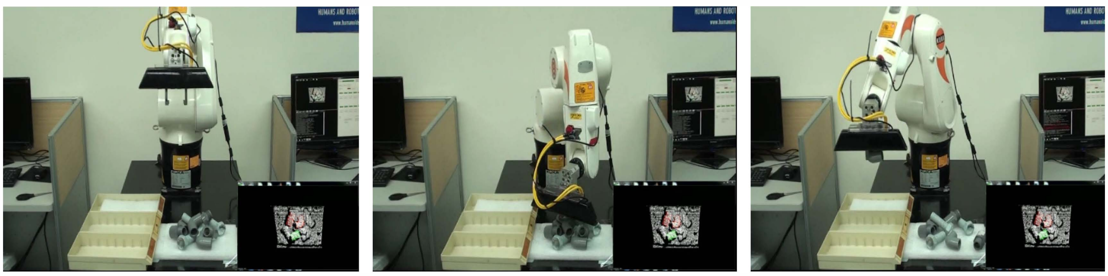
In this experiment, we used 3 different objects to verify the proposed method. We measured the recognition success rate and robot pickup success rate by placing three different types of multiple number objects which are randomly placed in the bin as shown in Fig. 5. Figure 6 shows the experiment of random bin-picking of multiple different workpieces. After repeating 23 times of the experiment, the recognition rate of different type objects are obtained and depicted in Table II. When the robot gripper picks up one object, and places it in the correct bin, we count it as a success. The experimental result show that the recognition rate of objects 1 is 100%, the recognition rate of objects 2 is 100%, the recognition rate of objects 3 is 89.7%. The average cycle time of random bin-picking of the experiment is 28.37 s. The average object recognition time is 7.67 s. The major part of cycle time of the random bin-picking experiment is robot arm motion of object grasping and placing. The pickup success rate

Table I Average absolute error of pose estimation.

X [mm]	Y [mm]	Z [mm]	Rotation [°]
1.06	1.48	0.72	4.58

Table II Recognition rate of different type objects.

Object type	Recognition total times	Recognition fail times	Recognition rate
Object 1	58	0	100%
Object 2	50	0	100%
Object 3	76	14	81.53%
Average recognition rate			93.9%



(a) (b) (c)

Fig. 6. Experiment of random bin-picking of multiple different objects. (a) Locate the object (b) Grasp the object (c) Put the object in the right box.

Table III Pickup success rate.

Total trial	Success	Failure	Success rate
184	165	19	89.7%

and average recognition rate are shown in Table III. The average recognition rate of three different type objects is 93.9% and the pickup success rate is 89.7%. Some objects are recognized but not grasped due to the thickness of the gripper itself, which might touch other objects during grasping the target. A video clip of this experiment can be found in [13].

6. CONCLUSION AND FUTURE WORK

In this paper, we propose a CAD-based 6-DOF pose estimation for random bin-picking of multiple different objects based on a Kinect RGB-D sensor. We built the 3D CAD model of objects via a virtual camera, which generated a point cloud database for object pose estimation. A voting-scheme was adopted for 6-DOF pose estimation system as well as for recognizing a set of 6-DOF poses of different type objects from the bin. A series of experiments of practical random bin-picking of multiple different objects reveals that the proposed system can pick up all random pose of objects in the bin. The average recognition rate of three different type objects is 93.9% and the pickup success rate is 89.7%. In the future, method to accelerate the processing time will be studied. A more accurate depth sensor will be investigate for bin-picking of small objects.

ACKNOWLEDGEMENTS

This work was supported by the Ministry of Science and Technology of Taiwan R.O.C under grant NSC 102-2218-E-009-012-MY3.

REFERENCES

[1] C. Choi and H. I. Christensen, "Real-time 3D Model-based Tracking Using Edge and Keypoint Features for Robotic Manipulation," in *Proc. of IEEE In International Conference on Robotics and Automation (ICRA),* Anchorage, Alaska, USA, 2010, pp. 4048-4055.

[2] M. Ulrich, C. Wiedemann and C. Steger, "CAD-Based Recognition of 3D Objects in Monocular Images," in *Proc. of IEEE International Conference on Robotics and Automation (ICRA)*, Kobe, Japan, 2009, pp.1191-1198.

[3] S. Akizuki and M. Hashimoto, "High-speed and Reliable Object Recognition Using Distinctive 3-D Vector-Pairs in a Range Image." *in Proc. of IEEE International Symposium on Optomechatronic Technologies (ISOT)*, Chukyo, Japan, 2012.

[4] C. Choi and H. I. Christensen, "3D Pose Estimation of Daily Objects Using an RGB-D Camera," *in Proc. of IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS),* Vilamoura, Algarve, Portugal, 2012, pp.3342-3349.

[5] B. Drost, M. Ulrich, N. Navab and S. Ilic, "Model Globally, Match Locally: Efficient and Robust 3D Object Recognition," in *Proc. of IEEE International Conference on Computer Vision and Pattern Recognition(CVPR)*, Munich, Germany, 2010, pp.998-1005.

[6] C. Choi, Y. Taguchi, O. Tuzel, M. Liu and S. Ramalingam, "Voting- Based Pose Estimation for Robotic Assembly Using a 3D Sensor," in *Proc. of IEEE International Conference on Robotics and Automation (ICRA)*, Minnesota, USA, 2012, pp.1724-1731.

[7] Ø. Skotheim, M. Lind, P. Ystgaard and S. Fjerdingen, "A Flexible 3D Object Localization System for Industrial Part Handling," in *Proc. of IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, Vilamoura, Algarve, Portugal, 2012, pp.3326-3333.

[8] A. Menon, B. Cohen and M. Likhachev, "Motio n Planning for Smooth Pickup of Moving Objec ts," in *Proc. of the IEEE International Conferen ce on Robotics and Automation (ICRA)*, Hong K ong, 2014, pp 453-460.

[9] N. Atanasov, B. Sankaran, J. L. Ny, G. J. Pappas, K. Daniilidis, "Nonmyopic View Planning for Active Object Classification and Pose Estimation", *IEEE Transactions on Robotics (TRO)*, Vol. 30, No. 5, pp. 1078-1090, October 2014.

[10] R. Eidenberger and J. Scharinger, "Active perception and scene modeling by planning with probabilistic 6D object poses," in *Proc. IEEE/RSJ IEEE/RSJ International Conference on Intelligent Robots and Systems*, Taipei, Taiwan, 2010, pp. 1036-1043.

[11] S. Y. Jiang, N. Y. C. Chang, C. C. Wu, C. H. Wu and K. T. Song, "Error Analysis and Experiments of 3D Reconstruction Using a RGB-D Sensor," in *Proc. of IEEE International Conference on Automation Science and Engineering (CASE)*, Taipei, Taiwan, 2014, pp. 1020-1025.

[12] P. J. Besl and N. D. McKay, "A Method for Registration of 3D Shapes," *IEEE Trans on Pattern Analysis and Machine Intelligence,* vol. 14, no. 2, 1992, pp. 239-259.

[13] http://isci.cn.nctu.edu.tw/video/ICCAS2015/RB P.wmv.