2014 IEEE International Conference on Robotics & Automation (ICRA)
Hong Kong Convention and Exhibition Center
May 31 - June 7, 2014. Hong Kong, China

Fast Graspability Evaluation on Single Depth Maps for Bin Picking with General Grippers

Yukiyasu Domae[1], Haruhisa Okuda[2], Yuichi Taguchi[3], Kazuhiko Sumi[4], and Takashi Hirai[1]

***Abstract*— We present a method that estimates graspability measures on a single depth map for grasping objects randomly placed in a bin. Our method represents a gripper model by using two mask images, one describing a contact region that should be filled by a target object for stable grasping, and the other describing a collision region that should not be filled by other objects to avoid collisions during grasping. The graspability measure is computed by convolving the mask images with binarized depth maps, which are thresholded differently in each region according to the minimum height of the 3D points in the region and the length of the gripper. Our method does not assume any 3-D model of objects, thus applicable to general objects. Our representation of the gripper model using the two mask images is also applicable to general grippers, such as multi-finger and vacuum grippers. We apply our method to bin picking of piled objects using a robot arm and demonstrate fast pick-and-place operations for various industrial objects.**

I. Introduction

The task of grasping objects randomly placed in a bin, referred to as bin picking, has been studied in robotics over several decades. Bin picking is useful in industrial settings, where grasping objects from a bin reduces the required space and avoids the use of parts feeders, as well as in household robots, where grasping daily objects from cluttered scenes is an important task. However, to build practical bin-picking systems, there are still several challenges such as robust pose estimation of objects and efficient grasp planning to avoid collisions during grasping.

In this paper, we focus on the industrial settings. In particular, we take an approach of grasping an object irrespective of its pose and isolating it from other objects in the bin. If the subsequent process requires a particular pose of the object, pose estimation can be performed on the single isolated object. This system design is due to the fact that even if the pose of an object is successfully estimated in a bin, grasping can fail because of the constraints on robot motion and collision of the gripper with other objects and the bin. To minimize the cycle time, our system captures a single depth map of a bin using a 3-D depth sensor, as shown in Fig. 1. We also limit the robot motion to 4 degrees-of-freedom (DoF) to efficiently perform the object isolation task.

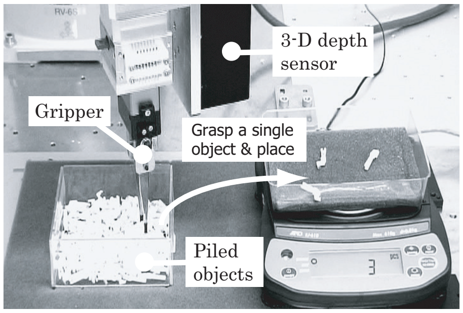

Fig. 1. How can robots quickly pick objects from cluttered scenes and place them by using a 3-D depth sensor and a general gripper? This paper presents a fast graspability evaluation method on single depth maps for the quick pick-and-place tasks.

We present a method that efficiently computes graspability measures on the single depth maps. To handle general objects and grippers, we do not assume any 3-D model of objects, and we represent a gripper by using only two mask images describing a contact region and a collision region. The contact region describes a region that should be filled by a target object for stable grasping, while the collision region denotes a region that should not be filled by other objects to avoid collision. The two mask images can be convolved with the depth map to compute pixel-wise graspability measures. To reduce the computational cost, we use a segmentation-based approach, where several segments are extracted from the depth map and the graspability measures are computed on each segment. In experiments, we demonstrate the generality of our method using several industrial objects and two different types of grippers (two-finger and vacuum grippers).

A. Contributions

Our contributions are summarized as follows.

- We present a method that evaluates graspability measures on a depth map by representing a gripper model with two mask images describing contact and collision regions.
- We present a system that enables fast pick-and-place tasks using the graspability evaluation method.
- We demonstrate the generality of our bin-picking system by using several industrial objects in experiments.

B. Related Work

There are two major approaches for grasping piled objects, depending on whether the grasping pose is determined

[1] Y. Domae and T. Hirai are with Advanced Technology R&D Center, Mitsubishi Electric Corporation, 8-1-1 Tsukaguchi-honmachi, Amagasaki, Hyogo 661-8661, JAPAN Domae.Yukiyasu@cb.MitsubishiElectric.co.jp

[2] H. Okuda is with Nagoya Works, Mitsubishi Electric Corporation, 5-1-14 Yada-minami, Higashi-ku, Nagoya, Aichi 461-8760, JAPAN

[3] Y. Taguchi is with Mitsubishi Electric Research Laboratories, 201 Broadway, Cambridge, MA 02139, USA

[4] K. Sumi is with College of Science and Engineering, Aoyama Gakuin University, 5-10-1 Fuchinobe, Chuo-ku, Sagamihara, Kanagawa 252-5258, JAPAN

978-1-4799-3684-7/14/$31.00 ©2014 IEEE

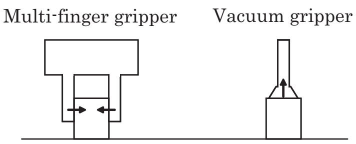

Fig. 2. Types of gripper for grasping objects by robots. Although robotic grippers have various shapes, they can be classified into two types: multi-finger and vacuum (suction). We design a fast grasping pose computation method for both gripper types.

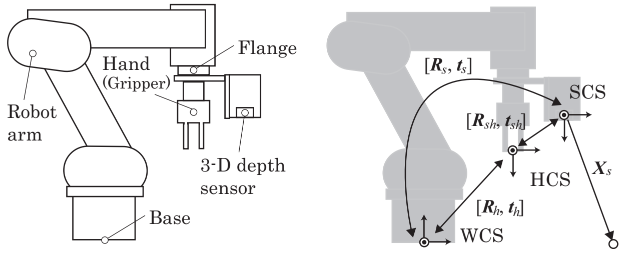

Fig. 3. System configuration and coordinate systems. The World Coordinate System (WCS) originates at the center of the robot base. The Sensor Coordinate System (SCS) is attached to the 3-D depth sensor. The Hand (gripper) Coordinate System (HCS) originates at the center of the grasping position of the grippers. The 3-D depth sensor and the hand are fixed to the flange of the robot and their coordinate systems SCS and HCS are calibrated with respect to WCS.

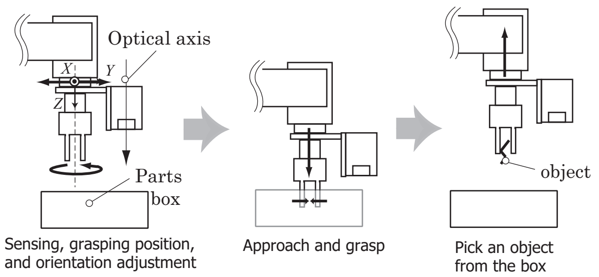

Fig. 4. Constraints of motion and layout for grasping an object from a cluttered scene. The optical axis of the 3-D depth sensor is aligned with the robot Z axis. We perform the bin-picking task using 4 DoF of the gripper pose, including X-Y-Z axis translation and rotation about the Z axis.

uniquely or arbitrarily. If the grasping pose is unique, it is necessary to know the shape information of the objects in advance to compute the grasping pose of the gripper in the reference coordinate system of the object model. If the grasping pose is arbitrary, it is sufficient to know the relationship between the depth map and the gripper, so the shape information of the objects is not necessary.

The former approach performs pose estimation of objects using known object models. Typical pose estimation algorithms first compute coarse poses of objects. Several algorithms that use a set of geometric primitives such as planes, circles, and cylinders have been proposed [1], [2], [3]. Recently, pairs of oriented points used in a voting framework [4], [5], [6] have been shown to be effective for the coarse pose estimation. The coarse poses are then refined by using an iterative closest point (ICP) algorithm [7], [8], [9]. Such pose-estimation-based approaches are able to balance factors such as the computational cost, accuracy, and success rate to a certain degree, and have begun to find applications for particular cases. Nevertheless, there are cases in which constraints on robot motion and gripper shape prevent grasping even though the pose estimation is successful.

Instead of estimating the object pose, the latter approach directly searches for the grasping pose of the gripper on a depth map or on a 3-D point cloud. As shown in Fig.2, for example, if one can find a region consisting of a set of 3D points that can be grasped by a multi-finger gripper or fitted by a vacuum gripper, then it would be possible to pick an object using the region on the basis of the depth map, without using a 3-D model of the object. Such an object grasping approach has often been adopted for household robots, which are required to grasp unknown objects [10], [11], [12], [13], [14], [15], [16]. In [10], an object region is first segmented on a depth map and then a search is made over the 3-D data for the grasping pose of a multi-finger gripper with 6 DoF. For that purpose, the position and relationship conditions of the model and data for closing the gap between the gripper fingers without gripper collision have been defined. In [14], a gripper model is defined over a depth map and brightness image and then manual training on large-volume image data is done in advance to determine what pose of the gripper model on the image leads to successful grasping. Then, the grasping pose for new data is selected by using a support vector machine (SVM) ranking algorithm [17] to classify grasping as successful or unsuccessful in a feature space that represents multiple image features. This approach is specialized for accuracy in grasping isolated objects and does not consider computational cost; thus it is not suited to the task of quickly separating piled objects with small adjustments. Furthermore, the works described above handle either multi-finger or vacuum grippers, but no method that can be used with both of those types has been proposed. We present a method that computes grasping poses fast and independently of the types of the grippers for the use in efficient bin picking.

To the best of our knowledge, Buchholz et al.'s approach [6] is the closest to ours. Their approach first estimates the poses of objects as in the former approach described above, and then computes an optimal grasping pose around the pre-defined grasping poses on the object. For the optimal grasping pose selection, they compute collisions by comparing the depth map transformed to the object coordinate system with two images generated from a CAD model of the gripper, corresponding to the upper and lower parts of the gripper. Compared to their system, we do not require any CAD model of the object since we do not compute the poses of objects. Instead, we use the mask images to compute both the grasping poses and the collisions, enabling faster operations.

II. Problem Formulation

In this section, we formulate the problem of computing the grasping pose of a gripper using a 3-D depth sensor, both of which are attached to a robot arm. Figure 3 shows our configuration. The coordinate systems shown in Fig. 3 are defined relative to the world coordinate system that originates at the center of the robot base. The sensor and the gripper are fixed to the flange of the robot, and their coordinate systems are calibrated with each other via hand-eye calibration.

We wish to compute a position and orientation with which the gripper can easily grasp an object using a depth map obtained with a 3-D depth sensor. This problem can be expressed as

$$\left[\boldsymbol{R}_h^*, \boldsymbol{t}_h^*\right] = \underset{\boldsymbol{R}_h, \boldsymbol{t}_h}{\operatorname{argmax}} f(\boldsymbol{R}_h, \boldsymbol{t}_h), \tag{1}$$

where $\boldsymbol{t}_h$ and $\boldsymbol{R}_h$ denote the position and orientation of the gripper in the world coordinate system, and $f(\boldsymbol{R}_h, \boldsymbol{t}_h)$ is a target function to be maximized for determining an optimal position and orientation. If all of the 6 DoF of the robot are considered, this problem needs to be solved in the 6 DoF space, which requires a high computational cost. Moreover, using 6 DoF robot motion would increase the time for grasping and the risk for collision with the bin. We therefore consider implementation of the minimum necessary DoF that enable bin picking by using the following robot operations:

1) In the empty space above the object, match the 3 DoF of the grasping position and orientation, which include X-Y axis translation and rotation about the Z axis, as shown in Fig. 4.
2) Move the gripper in the direction of the Z axis to the height for grasping the object.
3) Grasp the object with the gripper and move the gripper in the Z axis direction.

In this way, 2 rotational DoF are omitted and the bin-picking task can be done with 4 DoF. Furthermore, by aligning the optical axis of the 3-D depth sensor (i.e., the Z axis of the sensor coordinate system) with the robot Z axis as in Fig. 4, the orientation search can be done with the cross-section model of the gripper projected onto the depth map.

When a grasping position is computed in the depth map, the position is represented in the sensor coordinate system as $\boldsymbol{X}_s = \left[X_s, Y_s, Z_s\right]$. Using the transformation between the sensor and hand coordinate systems $\left[\boldsymbol{R}_{sh}, \boldsymbol{t}_{sh}\right]$, the position in the world coordinate system is obtained as

$$\boldsymbol{t}_h = \boldsymbol{R}_{sh}\boldsymbol{X}_s + \boldsymbol{t}_{sh} + \boldsymbol{d}. \tag{2}$$

Depending on the gripper shape, it may be necessary to lower the grasping position relative to the measured object position. We therefore add the term $\boldsymbol{d}$. From the operation constraints of Fig. 4, $\boldsymbol{d} = \left[0, 0, d\right]^T$, but the value d varies with the design of the gripper. In the case of a vacuum gripper that applies suction to the surface of the object, for example, the object surface is at the computed position, so $d = 0$. The only unknown term in Eq. (2) is $\boldsymbol{X}_s$.

Next, we represent the desired grasping orientation $\boldsymbol{R}_h$ as

$$\boldsymbol{R}_h = R_z(C), \tag{3}$$

where $R_z(C)$ denotes the rotation around the Z axis of the world coordinate system with an angle C. Representing the gripper orientation at the time the object is imaged by the 3-D depth sensor as (A^v, B^v, C^v), the constraints on operation and orientation mean that $A^v = 0$ and $B^v = 0$, so the only unknown is the angle of rotation about the Z axis, which can be represented as follows:

$$C = \theta + \theta_{\text{offset}}. \tag{4}$$

Here, θ represents the angle of in-plane rotation that is computed for an optimal grasping in the depth map. The θ_{offset} term is the offset for adjusting the angle to the actual gripper angle; it is uniquely determined when the gripper and camera are calibrated. In Eq. (4), θ is the only unknown variable. Thus, the problem of computing the grasping position and orientation for bin picking is expressed as follows:

$$\left[X_s^*, Y_s^*, \theta^*\right] = \underset{X_s, Y_s, \theta}{\operatorname{argmax}} f(X_s, Y_s, \theta), \tag{5}$$

$$\boldsymbol{t}_h^* = \boldsymbol{R}_{sh}\boldsymbol{X}_s^* + \boldsymbol{t}_{sh} + \boldsymbol{d}, \tag{6}$$

$$\boldsymbol{R}_h^* = R_z(\theta^* + \theta_{offset}). \tag{7}$$

Because the Z_s at the measurement position can be calculated from (X_s, Y_s) given a 3-D depth sensor, it is not included as an independent variable in Eq. (5). According to Eq. (5), if the position (X_s^*, Y_s^*) on the depth map obtained from the 3-D depth sensor and the in-plane rotation angle θ^* can be computed, then the grasping position and orientation $\left[\boldsymbol{R}_h^*, \boldsymbol{t}_h^*\right]$ can be obtained from Eq. (6) and Eq. (7). The problem is therefore solved by designing the target function $f(X_s, Y_s, \theta)$.

III. Fast Graspability Evaluation

In this section, we propose a method for efficiently solving the 3 DoF problem that is defined in the previous section. We first define a state of graspability that can be represented on a depth map and describe a gripper model for representing that graspable state for a general gripper. We then present a method for evaluating graspability from the gripper model and the depth map. Our method considers a depth map a 2D gray-scale image and uses 2-D image processing for efficiently computing the graspability.

A. Graspable State

To design the target function of Eq. (5), we first define the ideal grasping pose. The evaluation indicators for the gripper include form closure and force closure [18], but those are understood as an equilibrium force in an existing stable gripping state or a geometrically confined state. The problem of concern here requires consideration of a "graspable state," which is a state that exists prior to the transition to a stable state in which an object is being grasped. In this paper, we take the graspable state to be "a gripper position and orientation for which an operation such as opening and closing or application of suction enables the grasping of an object."

The graspable state is defined differently for different types of grippers. Here, the gripper types include the multi-finger type and the vacuum type as shown in Fig. 2. For example, five-finger grippers and jamming grippers [19] are all in the class of multi-finger grippers, in the sense that they grasp the target object[1]. Grippers that use suction or electromagnetic force to directly attach to an object are classified as vacuum grippers. According to this classification, the graspable state can be considered in the following way. For a multi-finger gripper, a graspable state is "a state in which there is nothing to collide with the fingers of the gripper and there is contact with the target object in the gap between the fingers." For a vacuum gripper, a graspable state is "a state in which the vacuum pad is in full contact with the surface of the target object." The cross section illustrations in Fig. 5 show the graspable states for various grippers.

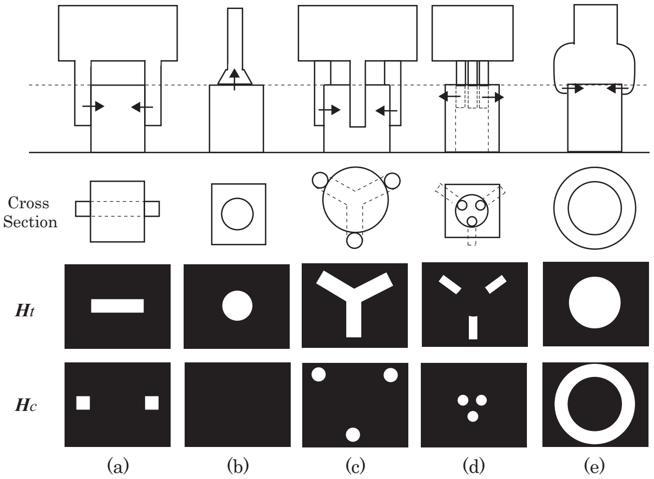

Fig. 5. Models for various grippers. (a) Two-finger gripper. (b) Single-pad vacuum gripper. (c) Three-finger gripper that closes the fingers to grasp. (d) Three-finger gripper that opens the fingers to grasp. (e) Jamming gripper [19] (classified as a multi-finger gripper). The masks corresponding to the contact region $\boldsymbol{H}_t$ and the collision region $\boldsymbol{H}_c$ are modeled separately for each gripper. The masks $\boldsymbol{H}_t$ and $\boldsymbol{H}_c$ are represented by binary values, with the white regions denoting one and the black regions denoting zero.

B. Gripper Models

The graspable states described above can be evaluated by the relationship between the measured depth map of the object and the position and orientation of the gripper model. We design a gripper model that has a high degree of generality for various types and shapes of grippers to implement this evaluation. As shown in Fig. 5, we use two mask images modeling a contact region $\boldsymbol{H}_t$ and a collision region $\boldsymbol{H}_c$ for each gripper. The illustrations from left to right in Fig. 5 are for a two-finger gripper, a single-pad vacuum gripper, a three-finger gripper that closes the fingers to grasp, a three-finger gripper that opens the fingers to grasp, and a jamming gripper (classified as a multi-finger gripper). The masks $\boldsymbol{H}_t$ and $\boldsymbol{H}_c$ are represented by binary values, with the white regions denoting one and the black regions denoting zero. For the vacuum gripper, there is no collision region. For the jamming gripper, the contact region depends on the shape of the target object surface because a flexible material is used. Note that the collision regions can also include other parts of the system, such as the 3-D depth sensor and other parts of the hand, if there is a possibility for those parts to collide with the object and its surroundings. Our gripper model can be designed using only physical dimension parameters.

In addition, for determining the grasping orientation, we use N gripper models corresponding to different in-plane orientations of the gripper. We denote as $\boldsymbol{H}_t^i$ and $\boldsymbol{H}_c^i$ ($i = 1, \ldots, N$) the contact and collision regions corresponding to the ith orientation, generated by rotating $\boldsymbol{H}_t$ and $\boldsymbol{H}_c$ with the angle $\frac{\pi i}{N}$.

C. Evaluating Graspability

Here we design the target function for evaluating graspability from the relationship between the gripper model and the depth map. We define a contact region $\boldsymbol{W}_t$ and a collision region $\boldsymbol{W}_c$ for the scene in the depth map, similar to the contact region $\boldsymbol{H}_t^i$ and the collision region $\boldsymbol{H}_c^i$ for the gripper. As shown in Fig. 6, the height of the target object in the depth map, h, and the depth to which the gripper advances when grasping, d, are used to define the contact region $\boldsymbol{W}_t$ and collision region $\boldsymbol{W}_c$ for some object in the scene. Denoting the value of $\boldsymbol{W}_t$ at position (x, y) as $W_t(x, y)$, the contact region of the object is expressed as

$$W_t(x,y) = \begin{cases} 1 \text{ if } W(x,y) \geq h \\ 0 \text{ otherwise} \end{cases}, \tag{8}$$

where $W(x, y)$ represents the value of the depth map at position (x, y). The object collision region is expressed as

$$W_c(x,y) = \begin{cases} 1 \text{ if } W(x,y) \geq h - d \\ 0 \text{ otherwise} \end{cases}. \tag{9}$$

If the height of the target object surface is not uniform, the minimum value of the height distribution of the target object surface is used for the threshold h to eliminate the possibility of collision of the multi-finger gripper with the surrounding collision regions.

A position that has a large intersection of object and gripper contact regions and no intersection of their collision regions can be considered as a position where the graspability is high. The intersection of the object and gripper contact regions can be computed as

$$\boldsymbol{T}^i = \boldsymbol{H}_t^i \otimes \boldsymbol{W}_t, \tag{10}$$

where $\boldsymbol{T}^i$ is a binary value that represents whether or not there is contact between the gripper contact model with the ith in-plane rotation angle, $\boldsymbol{H}_t^i$, and the object contact model $\boldsymbol{W}_t$. The value is one if there is contact and zero otherwise. The operator $\otimes$ represents convolution. The intersection of the collision regions can be similarly computed as

$$\boldsymbol{C}^i = \boldsymbol{H}_c^i \otimes \boldsymbol{W}_c, \tag{11}$$

where $\boldsymbol{C}^i$ is a binary value that represents whether or not there is collision of the gripper collision model with the ith in-plane rotation angle, $\boldsymbol{H}_c^i$, with the object collision model

[1] Although jamming grippers are pushed down onto an object from above, they grasp by shaping a flexible gripper to bind the object rather than by applying suction to the object.

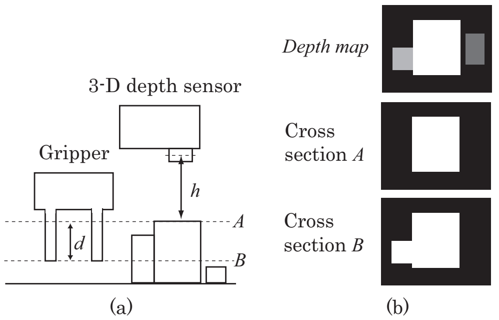

Fig. 6. Scene modeled by contacts and collisions between a gripper and objects. (a) Scene and system setup. (b) A depth map, contact region $\boldsymbol{W}_t$, and collision region $\boldsymbol{W}_c$ for the scene. The height of the target object in the depth map, h, and the depth to which the gripper advances when grasping, d, are used to define the regions.

$\boldsymbol{W}_c$ in the scene. The region of contact without collision of the gripper and object can thus be expressed as $(\boldsymbol{T^i} \cap \bar{\boldsymbol{C}}^{\boldsymbol{i}})$. We define $\boldsymbol{G}^i$ so as to obtain the peak of that region as

$$\boldsymbol{G}^i = (\boldsymbol{T^i} \cap \bar{\boldsymbol{C}}^{\boldsymbol{i}}) \otimes \boldsymbol{g}, \tag{12}$$

where $\boldsymbol{g}$ denotes a Gaussian. $\boldsymbol{G}^i$ is the graspability map for the gripper model that has the ith in-plane rotation angle and the target object. The search for the maximum value in that map is equivalent to the search for the grasping position for that gripper model. The process for computing the graspability map using Eqs. (10)–(12) is illustrated in Fig. 7. Denoting the value of $\boldsymbol{G}^i$ at position (x, y) as $G^i(x, y)$, we can design the target function as follows for each gripper model.

Target function for multi-finger grippers:

$$f(X_s, Y_s, \theta) = \begin{cases} G^i(x, y) \text{ if } C^i(x, y) = 0 \\ 0 \text{ otherwise} \end{cases}. \tag{13}$$

Here, $C^i(x, y)$ is the value of $\boldsymbol{C}^i$ at position (x, y). If there is no collision between the gripper and the object or its surroundings, there is graspability.

Target function for vacuum grippers:

$$f(X_s, Y_s, \theta) = \begin{cases} G^i(x, y) \text{ if } T^i(x, y) \geq S_t^i \\ 0 \text{ otherwise} \end{cases}. \tag{14}$$

Here, S_t^i denotes the surface area of the gripper contact region (the sum of the white regions in $\boldsymbol{H}_t^i$) and $T^i(x, y)$ is the value of $\boldsymbol{T}^i$ at position (x, y). The value S_t^i is determined by the gripper shape and is used to decide whether or not the gripper's pad can be sufficiently attached to the object surface. Note that we do not have to compute $\boldsymbol{C}^i$ in Eq. (11) for the vacuum grippers since their collision region $\boldsymbol{H}_c^i$ is zero.

From the peak position (x^{i*}, y^{i*}) and the orientation index i^* computed in the graspability map with Eq. (13) or Eq. (14), we obtain the values (X_s^*, Y_s^*, θ^*) as

$$X_s^* = \frac{Z_s}{f} x^{i*}, \quad Y_s^* = \frac{Z_s}{f} y^{i*}, \quad \theta^* = \frac{\pi i^*}{N}, \tag{15}$$

where f is the focal length of the 3-D depth sensor and Z_s is the depth value at the pixel location (x^{i*}, y^{i*}). Then we obtain $\boldsymbol{R}_h^*$ and $\boldsymbol{t}_h^*$ using Eqs. (6) and (7).

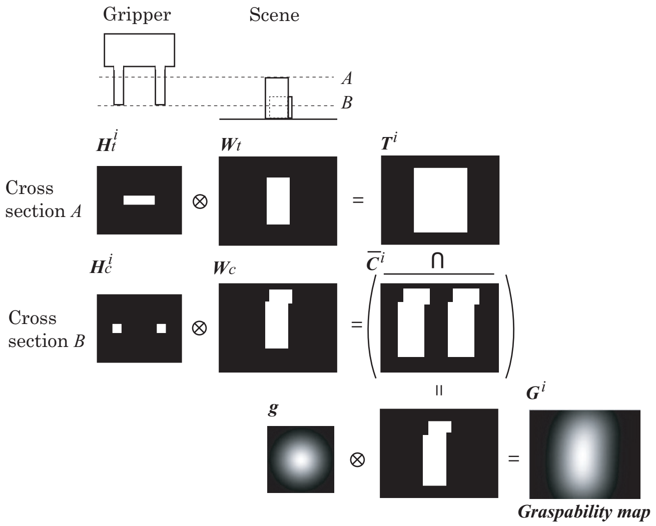

Fig. 7. Graspability evaluated by gripper and object models. $\boldsymbol{T}^i$ is a binary value that represents whether or not there is contact between the contact model of the gripper that has the ith in-plane rotation angle, $\boldsymbol{H}_t^i$, and the object contact model $\boldsymbol{W}_t$. The value is one if there is contact and zero otherwise. $\boldsymbol{C}^i$ is a binary value that represents whether or not there is collision of the gripper collision model that has the ith in-plane rotation angle, $\boldsymbol{H}_c^i$, with the object collision model $\boldsymbol{W}_c$ in the scene. The region of contact without collision of the gripper and object can thus be expressed as $(\boldsymbol{T^i} \cap \bar{\boldsymbol{C}}^{\boldsymbol{i}})$. The graspability map, $\boldsymbol{G}^i$, is computed to obtain the peak of that region.

IV. Implementation

In the previous sections, we reduced the problem of estimating the grasping pose to the problem of finding the peaks in a graspability map computed using 2-D image processing on a depth map. Obtaining such a pixel-wise graspability map would involve the computation on the order of $w \times h \times x_t \times y_t \times N$, where w and h are the width and height of the depth map, and x_t and y_t are the width and height of the gripper contact and collision models. To further reduce the processing time, rather than computing the pixel-wise graspability map, we use a segmentation-based approach where we first extract object candidate regions using segmentation and then limit the search scope to the rectangle regions, each of which bounds a candidate region. Also, because Eq. (10) and Eq. (11) are binary images, the convolution operation is accelerated by representing each pixel with one bit and taking the logical AND for each bit.

Extracting Object Candidate Regions: To implement the segmentation-based approach, we segment a depth map into regions bounded by edges, and fit a plane or a curved surface for each region to extract object candidates. We sort the candidates according to their average heights and keep the K highest candidates, considering that the object located at the highest position in the pile should be the easiest to grasp. The graspability is computed for each of the K candidate regions in the rectangle bounding the region. If we find multiple peak positions in the candidate regions, we select a peak closest to the center of one of the regions so that the grasping position can be close to the center of gravity of the object.

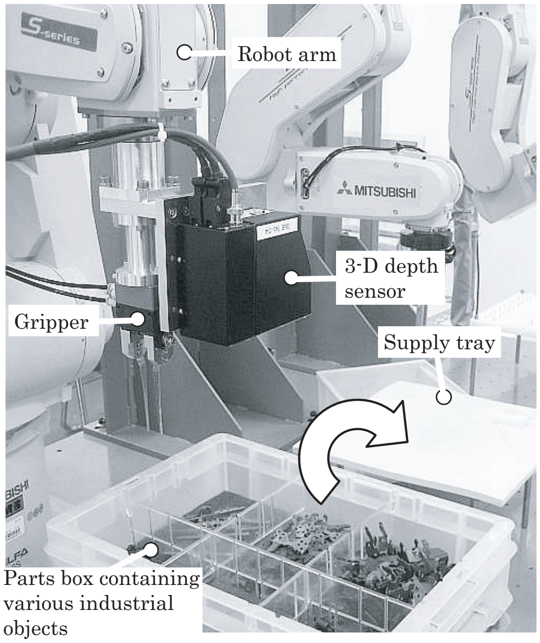

Fig. 8. Experimental setup. A parts box divided into compartments was placed in the environment and piled industrial parts were placed in the compartments.

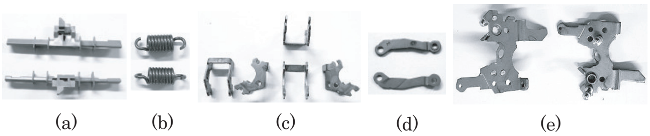

Fig. 9. Five types of objects (industrial parts) used in experiments, having different shapes and materials.

Parameters: The processing parameters defined so far are summarized below.

1) h: the surface height of the target segment; the minimum value of the measured height is used to prevent collisions when there is dispersion in the surface height values.
2) d: the depth to which the gripper is to proceed from the measured position; this is zero for a vacuum gripper and depends on the gripper design for a multi-finger gripper.
3) N: the number of gripper models with different orientations; a larger N means higher accuracy; this relates to processing and accuracy.
4) K: the number of object candidates extracted; this relates to processing and grasping success rate.

In addition to these parameters, the parameter for conversion of the edge strengths to binary values during the segmentation process must be set. That parameter can be automatically set from the design of the 3-D depth sensor being used and the desired edge height separation.

V. Experiments

We performed extensive evaluation of our method using the setup shown in Fig. 8. A parts box divided into compartments was placed in the environment, and five types of objects of different shapes and materials (Fig. 9) were placed in the compartments. The objects were picked up by two types of air-driven grippers (2-finger and vacuum shown in Fig. 10) attached to a robot arm, and then placed on a supply tray. The picking process was repeated 100 times for each object and for each gripper. The 3-D depth sensor consisted of a camera and laser, and used structured light to measure a depth map with 640×480 pixels and 8 bit resolution. The time for measuring a depth map was about 1 second. We adjusted the sensor to have sub-millimeter depth accuracy at a working distance of 300 to 400 mm. The robot arm was a Mitsubishi Electric RV-6SL. For the computation, a standard PC (2 GHz Core2 Duo with 2 G RAM) was connected directly to the robot controller and 3-D depth sensor via Ethernet. The processing parameters were set as $N = 8$ and $K = 9$. The other parameters were calculated from the actual dimensions of the experimental environment. The grasping position and orientation were recalculated for each grasping attempt. If the height of the grasping position estimation result was bigger than the height of the floor of the depth maps, the result was classified as a failure. Whether or not the pick-and-place operation was successful was decided automatically according to the total weight of objects in the part supply tray. Please refer to the supplementary video, which demonstrates our system performing the pick-and-place tasks for several different objects.

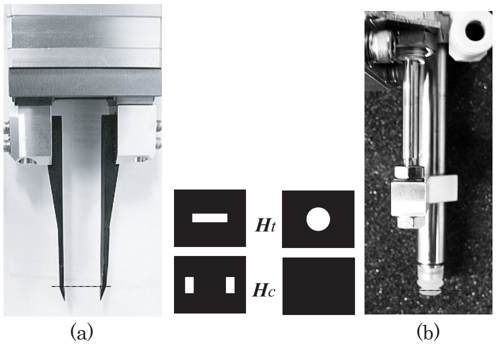

Fig. 10. Two types of grippers and their models used in experiments: (a) 2-finger gripper and (b) vacuum gripper. Both grippers were air-driven.

A. Evaluation Results

Table I summarizes the experimental results for the pick-and-place operations of the five different objects using the two different grippers. Here the *graspable rate* indicates a percentage where a specific object is physically possible to be grasped by a specific gripper. Specifically, we computed it as the number of stable poses of the object that is graspable by the gripper over the number of all stable poses of the object. For a multi-finger gripper, an object is basically 100% graspable if there is a height difference, but for a vacuum gripper, grasping is not possible unless there is a flat part that can cover the entire surface area of the suction pad. Therefore, the object (b), which is a coil spring, cannot be grasped by a vacuum gripper. Also, the object (c) has many

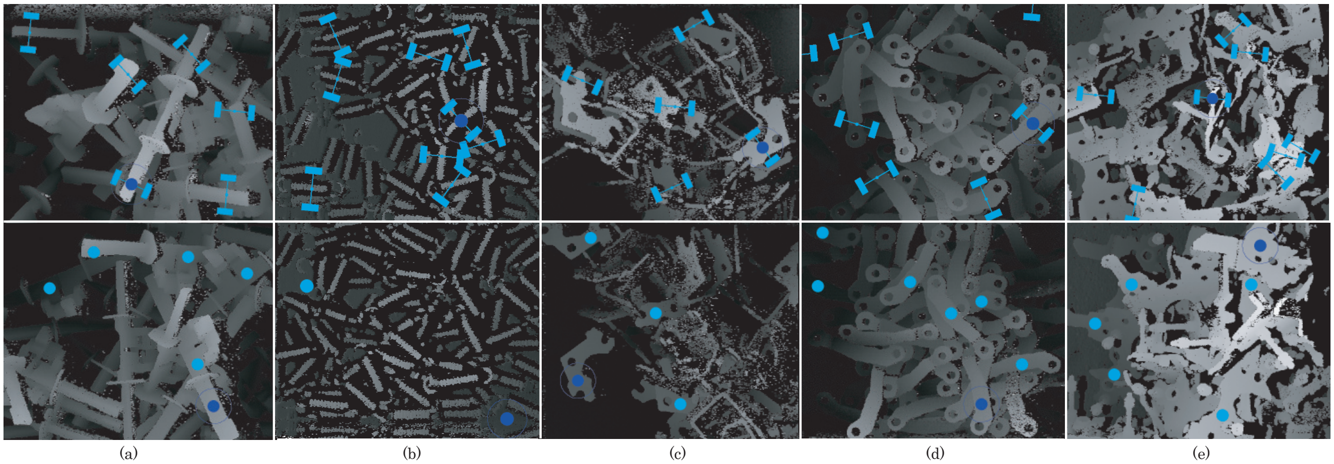

Fig. 11. Grasping position and orientation estimation results for five types of industrial parts by using the 2-finger gripper model (top) and the vacuum gripper model (bottom). (a)–(e) correspond to the objects (a)–(e) shown in Fig. 9. Sky-blue colored areas denote the gripper poses that our method estimated as graspable, while the blue colored circle denotes the best pose with a maximum value of the graspability. The robot can grasp the objects using the position and orientation. Please see the supplementary video, demonstrating bin picking from these scenes.

TABLE I

SUCCESS RATE FOR PICK-AND-PLACE OPERATIONS OF PILED OBJECTS.

Object type	Gripper type	Graspable rate	Grasping rate (multiple)	Successful rate (multiple)
(a)	2-finger	100	95 (1)	95 (96)
	vacuum	100	89 (0)	89 (89)
(b)	2-finger	100	87 (8)	87 (95)
	vacuum	0	0 (0)	—
(c)	2-finger	100	78 (12)	78 (90)
	vacuum	40	31 (6)	75 (92)
(d)	2-finger	100	79 (6)	79 (85)
	vacuum	100	79 (2)	79 (81)
(e)	2-finger	100	67 (5)	67 (72)
	vacuum	100	92 (4)	92 (96)

poses that do not present a flat surface, so the graspable rate is low. The *grasping rate* is the ratio of the number of successful pick-and-place operations to the total number of attempts in the experiment for one object. The *grasping rate (multiple)* is the ratio of cases where multiple objects were picked up in an entangled state in a single attempt. The *success rate* is defined as $100 \times \frac{GraspingRate}{GraspableRate}$ to take into account cases in which grasping is physically impossible. The *success rate (multiple)* includes cases where multiple objects were picked up at the same time.

The success rate averaged over the five different objects were 81.2% (87.6%) for the 2-finger gripper and 83.75% (89.5%) for the vacuum gripper; the overall average success rate was 82.47% (88.55%) (the values in parentheses are the success rates that include multiple object picking). Even at the current success rates, adequate system operation is possible when the method is used together with an error recovery mechanism in which grasping failure is detected and another candidate is picked. Nevertheless, a higher success rate would allow a simpler system configuration and increase the system throughput.

Figure 11 shows results of the grasping position and orientation computations for the five objects using the two gripper models. Sky-blue colored areas denote the grasping poses that our method estimated as graspable, and the blue colored circle denotes the best pose with a maximum value of the graspability. As described above, the object (b) cannot be grasped by the vacuum gripper; the method computed flat regions in the background as graspable regions, leading to a failure.

The average computation time required for the proposed method was 0.31 seconds (0.04 seconds for extracting object candidate regions and 0.27 seconds for the graspability evaluation) for the multi-finger gripper and 0.17 seconds (0.04 seconds for extracting object candidate regions and 0.13 seconds for the graspability evaluation) for the vacuum gripper. The processing time is shorter for the vacuum gripper because $\boldsymbol{H}_c = \mathbf{0}$ in the collision model and the computation for Eq. (11) is not performed. The cycle time of our system was 4.5 seconds (1.0 second for the depth measurement, 0.3 seconds for the proposed method, 2.5 seconds for the robot motion, and the remaining time for the wait for anti-vibration). Note that the computation was fast enough and could be performed during the robot motion. To further improve the cycle time, we could place the 3-D depth sensor separately from the robot arm; then the robot would move continuously and the cycle time would be determined by only the time for the robot motion.

B. Discussion

The experiments confirmed that our method enables grasping for both multi-finger and vacuum grippers if the scene includes objects that are graspable. However, there are still grasping failures, which we discuss here.

The frequency of failures varied with the object shape, but there were two types of causes: 1) the object was grasped, but the object fell down because of the weight of the object itself or the weight of surrounding objects; and 2) the impact of opening and closing during grasping created a moment that resulted in grasping failure. Grasping failures due to object weight were frequent when the object (e) was grasped with the multi-finger gripper, where the success rate was the lowest at 67%. Such failures could be reduced by maintaining the amount of gripping force or suction force needed to lift the object regardless of what part of the object is grasped, or by using time-consuming object pose estimation to grasp objects that are not occluded with other objects. The second type of failure occurred most often when objects that are thin (target objects no more than 1 mm thick), light-weight, and long and narrow, such as the object (d), were grasped at the end of the gripper. Such failures could be dealt with by reducing the gripping speed or putting priority on grasping at the object center of gravity. Those various methods could be selected to increase the success rate for either type of failures according to the system configuration or the target cycle time.

As described in Table I, we also observed cases where multiple objects were picked up in a single attempt. For the multi-finger hand, we had the case where multiple objects lie within the gap of the opened gripper, which occurred for the object (a). This problem could be addressed by using a gripper that has an adjustable span and setting the span so that multiple grasping does not occur. However, the main cause of multiple grasping was objects entangled with each other; in particular, objects of complex shape, such as the object (c), can be easily entangled (which might be hidden under the objects) and are difficult to separate. Even if we used sophisticated pose estimation algorithms to estimate poses of all the objects in a scene and locate hidden entanglement, and placed grasping priority on objects that are not entangled, it would be ultimately necessary to disentangle objects. Therefore, an efficient approach would be to pick the entangled objects and then either separate them or discard them after picking.

VI. Conclusions

We presented a bin-picking system using a 3-D depth sensor enabling fast pick-and-place tasks of piled objects. We reduced the problem of computing the poses of objects using a depth map to the problem of determining graspable regions directly on the depth map given a gripper model. We then proposed a method that efficiently evaluates the graspability using two mask images corresponding to the gripper model. Applied to an actual robot system, the proposed method achieved an average success rate of 82.47% and a processing time of 0.31 seconds or less. The proposed method has already been implemented as an algorithm in a 3-D depth sensor product for industrial robots. In future work, we wish to extend our method to other types of applications, such as those using humanoid and service robots.

Acknowledgments

We are grateful to Makito Seki, Akio Noda, and Kenichi Tanaka of Mitsubishi Electric and Professor Shunichi Kaneko of Hokkaido University for advice on robot vision research. We are also indebted to Yasuo Kitaaki for assistance in the implementation of our method. This research was funded by Next-generation Intelligent Robot Technology Development Project commissioned by the New Energy and Industrial Technology Development Organization (NEDO).

References

[1] K. Harada, K. Nagata, T. Tsuji, N. Yamanobe, A. Nakamura, and Y. Kawai, "Probabilistic approach for object bin picking approximated by cylinders," in *Proc. IEEE Int'l Conf. Robotics and Automation (ICRA)*, May 2013, pp. 3742–3747.

[2] M. Nieuwenhuisen, D. Droeschel, D. Holz, J. Stückler, A. Berner, J. Li, R. Klein, and S. Behnke, "Mobile bin picking with an anthropomorphic service robot," in *Proc. IEEE Int'l Conf. Robotics and Automation (ICRA)*, May 2013, pp. 2327–2334.

[3] M. Berger, G. Bachler, and S. Scherer, "Vision guided bin picking and mounting in a flexible assembly cell," in *Proc. Int'l Conf. Industrial Engineering Applications of Artificial Intelligence and Expert systems (IEA/AIE)*, vol. 1821, June 2000, pp. 109–117.

[4] B. Drost, M. Ulrich, N. Navab, and S. Ilic, "Model globally, match locally: Efficient and robust 3D object recognition," in *Proc. IEEE Conf. Computer Vision and Pattern Recognition (CVPR)*, June 2010, pp. 998–1005.

[5] C. Choi, Y. Taguchi, O. Tuzel, M.-Y. Liu, and S. Ramalingam, "Voting-based pose estimation for robotic assembly using a 3D sensor," in *Proc. IEEE Int'l Conf. Robotics and Automation (ICRA)*, May 2012, pp. 1724–1731.

[6] D. Buchholz, M. Futterlieb, S. Winkelbach, and F. M. Wahl, "Efficient bin-picking and grasp planning based on depth data," in *Proc. IEEE Int'l Conf. Robotics and Automation (ICRA)*, May 2013, pp. 3245–3250.

[7] P. J. Besl and N. D. McKay, "A method for registration of 3-D shapes," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 14, no. 2, pp. 239–256, Feb. 1992.

[8] D. Chetverikov, D. Svirko, D. Stepanov, and P. Krsek, "The trimmed iterative closest point algorithm," in *Proc. Int'l Conf. Pattern Recognition (ICPR)*, vol. 3, Aug. 2002, pp. 545–548.

[9] S. Rusinkiewicz and M. Levoy, "Efficient variants of the ICP algorithm," in *Proc. Int'l Conf. 3-D Digital Imaging and Modeling (3DIM)*, May 2001, pp. 145–152.

[10] D. Rao, Q. V. Le, T. Phoka, M. Quigley, A. Sudsang, and A. Y. Ng, "Grasping novel objects with depth segmentation," in *Proc. IEEE/RSJ Int'l Conf. Intelligent Robots and Systems (IROS)*, Oct. 2010, pp. 2578–2585.

[11] J. Bohg and D. Kragic, "Grasping familiar objects using shape context," in *Proc. Int'l Conf. Advanced Robotics (ICAR)*, June 2009, pp. 1–6.

[12] Y. Kimitoshi, M. Tomono, and T. Tsubouchi, "Autonomous 3D shape modeling and grasp planning for handling unknown objects," in *Robot Manipulators Trends and Development*, A. Jimenez and B. M. A. Hadithi, Eds. InTech, Mar. 2010, ch. 22.

[13] E. Klingbeil, D. Rao, B. Carpenter, V. Ganapathi, A. Y. Ng, and O. Khatib, "Grasping with application to an autonomous checkout robot," in *Proc. IEEE Int'l Conf. Robotics and Automation (ICRA)*, May 2011, pp. 2837–2844.

[14] Y. Jiang, S. Moseson, and A. Saxena, "Efficient grasping from RGBD images: Learning using a new rectangle representation," in *Proc. IEEE Int'l Conf. Robotics and Automation (ICRA)*, May 2011, pp. 3304–3311.

[15] A. Saxena, J. Driemeyer, and A. Ng, "Robotic grasping of novel objects using vision," *International Journal of Robotics Research (IJPR)*, vol. 27, no. 2, pp. 157–173, 2008.

[16] M. Popović, G. Kootstra, J. A. Jørgensen, D. Kragic, and N. Krüger, "Grasping unknown objects using an early cognitive vision system for general scene understanding," in *Proc. IEEE/RSJ Int'l Conf. Intelligent Robots and Systems (IROS)*, Sept. 2011, pp. 987–994.

[17] T. Joachims, "Optimizing search engines using clickthrough data," in *Proc. ACM SIGKDD Int'l Conf. Knowledge Discovery and Data Mining*, 2002, pp. 133–142.

[18] E. Rimon and J. Burdick, "On force and form closure for multiple finger grasps," in *Proc. IEEE Int'l Conf. Robotics and Automation (ICRA)*, vol. 2, Apr. 1996, pp. 1795–1800.

[19] J. R. Amend, E. Brown, N. Rodenberg, H. M. Jaeger, and H. Lipson, "A positive pressure universal gripper based on the jamming of granular material," *IEEE Trans. Robotics*, vol. 28, no. 2, pp. 341–350, Apr. 2012.