2011 IEEE International Conference on Robotics and Automation
Shanghai International Conference Center
May 9-13, 2011, Shanghai, China

Grasping with Application to an Autonomous Checkout Robot

Ellen Klingbeil, Deepak Rao, Blake Carpenter, Varun Ganapathi, Andrew Y. Ng, Oussama Khatib

Abstract— **In this paper, we present a novel grasp selection algorithm to enable a robot with a two-fingered end-effector to autonomously grasp unknown objects. Our approach requires as input only the raw depth data obtained from a single frame of a 3D sensor. Additionally, our approach uses no explicit models of the objects and does not require a training phase. We use the grasping capability to demonstrate the application of a robot as an autonomous checkout clerk. To perform this task, the robot must identify how to grasp an object, locate the barcode on the object and read the numeric code.**

We evaluate our grasping algorithm in experiments where the robot was required to autonomously grasp unknown objects. The robot achieved a success of 91.6% in grasping novel objects. We performed two sets of experiments to evaluate the checkout robot application. In the first set, the objects were placed in many orientations in front of the robot one at a time. In the second set, the objects were placed several at a time with varying amounts of clutter. The robot was able to autonomously grasp and scan the objects in 49/50 of the single-object trials and 46/50 of the cluttered trials.

I. INTRODUCTION

Grasping is a fundamental capability essential for many manipulation tasks. Enabling a robot to autonomously grasp an unknown object in uncluttered or cluttered scenes has received much attention in recent years. Due to the large variability among objects in human enviroments developing such an algorithm, assuming no prior knowledge of the object shape and given only noisy and incomplete sensor data, has proven to be very difficult.

In this paper, we present a novel grasp selection algorithm algorithm to identify the finger positions and wrist orientation for a two-fingered robot end-effector to grasp an unknown object. Our approach requires as input only the raw depth data obtained from a single frame of a 3D sensor. Additionally, our algorithm does not attempt to recognize or build a model for each object nor does it require offline training on hand-labeled data. Instead we attempt to approximately search for the geometric shape of a good region to grasp in the 3D point cloud. Our approach is motivated by the idea that identifying a good grasp pose for an object in a point cloud of a scene is approximately equivalent to finding the location of a region which is the same 3D shape as the interior of the gripper for a given orientation of the end-effector and a given gripper spread. Positioning the gripper around such a region will most likely result in a successful grasp. Reasoning about the entire local 3D region of the depth map which the robot will grasp, instead of only the finger-tip contact points, will almost always result in grasps which do not cause the gripper to be in collision with the scene, thus alleviating dependence on the motion planner to filter these grasps. An exhaustive search for these regions in a point cloud of the scene, over all the degrees of freedom defining a grasp configuration, is computationally intractable; however, we formulate an approach which approximately achieves this objective.

We use the grasping capability to present a novel application of a robot as an autonomous checkout clerk. This task requires the robot to autonomously grasp objects which can be placed in front of it at any orientation and in clutter (including objects which may be placed in contact with each other and even stacked). To perform this task, the robot must identify how to grasp an object, locate the barcode on the object and read the numeric code. In this paper, we restrict the scope of items to those which are mostly rigid (such as items found at a pharmacy) because the barcode detection is much more difficult if the object deforms when it is manipulated. To perform this task, we develop motion planning strategies to search for a barcode on the object (including a regrasping strategy for the case where the robot's initial grasp resulted in the gripper occluding the barcode). The numeric code is identified using off-the-shelf software.

We evaluate our grasping algorithm in experiments where the robot was required to autonomously grasp unknown objects. The robot achieved a success rate of 91.6%, outperforming other recent approaches to grasping novel objects. We performed two sets of experiments (using 10 different objects) to evaluate the checkout robot application. In the first set of experiments, the objects were placed one at a time at many random orientations in front of the robot. In the second set, the objects were placed several at a time with moderate to extreme clutter. The robot was able to autonomously grasp and scan the objects in 49/50 of the single-object trials and 46/50 of the cluttered scene trials.

II. RELATED WORK

Simple mobile robotic platforms are able to perform useful tasks in human environments, where those tasks do not involve manipulating the environment but primarily navigating it, such as robotic carts which deliver items to various floors of a hospital [1]. However, robots performing useful manipulation tasks in these unstructured environments is yet to be realized. We explore the capability of grasping, and utilize this capability to explore one possible application of a mobile robot as a checkout clerk.

Ellen Klingbeil is with Department of Aeronautics and Astronautics Stanford University, ellenrk7@stanford.edu. Deepak Rao, Blake Carpenter, and Varun Ganapathi are with Department of Computer Science, Stanford University, drao,blakec,varung@cs.stanford.edu. Andrew Y. Ng and Oussama Khatib are with the Faculty of Computer Science, Stanford University, ang,ok@cs.stanford.edu.

978-1-61284-385-8/11/$26.00 ©2011 IEEE

There has been much work in recent years in the area of grasp selection using noisy, incomplete depth data (such as that obtained from a laser or stereo system). Many approaches assume that a 3D model of the object is available and focus on the planning and control to achieve grasps which meet a desired objective, such as form or force closure [2], [3], [4]. [5] decomposed the scene into shape primitives for which grasps were computed based on 3D models. [6] identified the location and pose of an object and used a pre-computed grasp based on the 3D model for each object. Accurate 3D models of objects are tedious to acquire. Additionally, matching a noisy point cloud to models in the database becomes nearly impossible when the scene is cluttered (i.e. objects in close proximity and possibly even stacked).

More recently researchers have considered approaches to robotic grasping which do not require full 3D models of the objects. Most of these approaches assume the objects lie on a horizontal surface and attempt to segment a noisy point cloud into individual clusters corresponding to each object (or assume only a single object is present). Geometric features computed from the point cluster, as well as heuristics, are used to identify candidate grasps for the object. However, identifying how to grasp a wide variety of objects given only a noisy cluster of points is a challenging task, so many works make additional simplifying assumptions. [7] presented an approach for selecting two and three-finger grasps for planar unknown objects using vision. [8] assumed that the objects could be grasped with an overhead grasp. Additionally, a human pointed out the desired object with a laser pointer. [9] considered only box-shaped objects or objects with rotational symmetry (with their rotational axis vertically aligned). [10] designed an algorithm to grasp cylindrical objects with a power grasp using visual servoing. [11] identified a bounding box around an object point cluster and used a set of heuristics to choose grasp candidates which were then ranked based on a number of geometric features. In order to obtain reliable clustering, objects were placed at least 3 cm apart.

Methods which rely on segmenting the scene into distinct objects break down when the scene becomes too cluttered, thus some researchers have considered approaches which do not require this initial clustering step. [12] developed a strategy for identifying grasp configurations for unknown objects based on locating coplanar pairs of 3D edges of similar color. The grasps are ranked by prefering vertical grasps. [13] used supervised learning with local patch-based image and depth features to identify a single "grasp point". Since the approach only identifies a single "grasp point", it is best suited for pinch-type grasps on objects with thin edges. [14] used supervised learning to identify two or three contact points for grasping, however they did not include pinch-type grasps. These supervised learning approaches require hand-labeling "good" and "bad" grasps on large amounts of data.

Methods that attempt to identify entire models suffer from low recall and the inability to generalize to different objects. The supervised learning based approaches mentioned above attempt to generalize to arbitrary object shapes by characterizing the very local regions of the fingertip contact points using a training set. However this requires a large amount of labeled training data and does not reason about the entire 3D shape of the graspable region. A key insight we make is that in order to find a stable grasp, it is not necessary to first match an entire 3D model of an object and then identify graspable locations. When the local shape of an object matches the shape of the gripper interior, then the area of contact is maximized, which leads to a stable grasp. Thus we combine the key insights from both approaches. We search the 3D scene for local shapes that match some shape the gripper can take on. This means we avoid the need for training data, but we still have high recall because we can generalize to arbitrary objects.

Many applications have been proposed to bring robots into everyday human environments. [15] presented a robot equipped with an RFID reader to automate inventory management by detecting misplaced books in a library. The recent appearance of self-checkout stands in supermarkets, which partially automate the process by having the shopper interact with a computer instead of a human cashier, provides evidence that is is profitable to consider fully automating the process of scanning objects. In this paper, we propose fully automating the process by having a robot perform the task of checking out common items found in a pharmacy. The capabilities we present could also be used in an inventory-taking application.

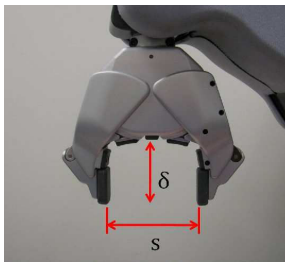
In addition to grasping, a necessary capability for an autonomous checkout robot is software that can detect and read barcodes on objects. Although RFID tags are becoming popular, barcodes are still more widely used due to their significantly lower cost. Barcode localization and recognition from a high-resolution image is a solved problem with many off-the-shelf implementations achieving near perfect accuracy [16].

III. APPROACH

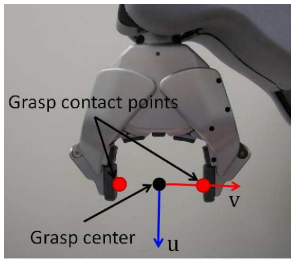
A. Grasping

Our algorithm is motivated by the observation that when a robot has a solid grasp on an object, the surface area of contact is maximized. Thus the shape of the region being grabbed should be as similar as possible to the shape of the gripper interior.

We consider an end-effector with two opposing fingers, such as the PR2 robot's parallel-plate gripper (see Figure 1). There are many possible representations to define a grasp pose for such an end-effector. One representation consists of the position, orientation, and "spread" of the gripper (the distance between the gripper pads); an equivalent representation is the position of the finger contact points and the orientation of the gripper (see Figure 1(b)). We will use both representations in our derivation, but the former will be the final representation we use to command the robot to the desired grasp pose. We introduce the additional geometric quantity of "grasp-depth" δ as the length of the portion of the object that the gripper would grab along the grasp orientation (see Figure 1(a)).



(a) "Grasp-depth" and "spread".



(b) Grasp pose parameters.

Fig. 1. Image of the PR2 parallel-plate gripper, labeling relevant parameters. (In figure (b), u, v describe the gripper orientation.)

We can find regions which approximately "fit" the interior of the gripper by searching for approximately concave regions of width s and height δ. Searching for such a region will result in grasps with contact points such as those shown in Figure 2. Although both grasps are stable, in order to be robust to slippage it is more desirable to place the fingers in the positions denoted by the "o's" than the positions denoted by the "x's". To identify the grasp pose which places the fingers closer to the center of the object, more global properties must be considered than just the local region at the contact points.

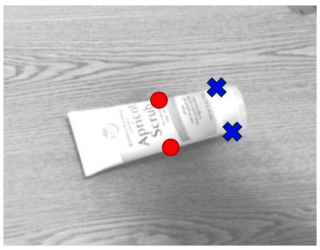
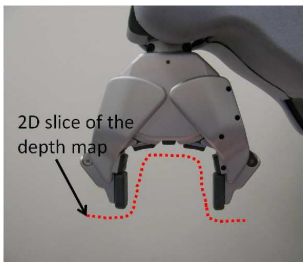


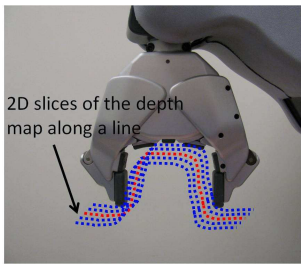
Fig. 2. Possible finger placements on an object. To be more robust to slippage, the "o's" are more desirable than the "x's".

We develop our algorithm by first solving a more restricted problem of finding a grasp in a 2D slice of the depth map. We will refer to this 2D grasp as a "planar-grasp". An example of a good region to grasp in a 2D slice of the depth map is shown in Figure 3(a). We observe that a good grasp region in 3D for our end-effector will consist of similar planar-grasps which lie along a line (see Figure 3(b)). Given a solution to the problem of finding planar-grasps in a 2D slice of the depth map, we can search across all possible planes (in principle) to identify the best planar-grasps. We then sample pairs of similar planar-grasps and use these as a model to evaluate the 3D grasp shape along the line defined by the two planar grasps. This approach is computationally faster than directly attempting to search for the full 3D shape of a good grasp region.

1) Identifying Planar-Grasps: The input to our algorithm consists of a depth map and registered image provided by the active stereo sensor on the PR2 robot ([17]). We interpolate and smooth the point cloud using the registered stereo depth map and intensity image to provide a dense map of the scene (i.e. a depth value is provided for each pixel in the image).



(a)



(b)

Fig. 3. These images show an example of a single 2D slice (a) and a line of 2D slices (b) in the depth map which are a good region to grasp.

Let $Z = [z^1 ....z^M]$ where z^k is the depth measurement for the k^{th} pixel in the image. Let n^k be the normal vector of the depth surface at the k^{th} pixel. We will identify planar cross-sections in the point cloud which approximately match the 2D cross-section of the gripper interior by first introducing a geometric quantity we will refer to as a "planar-grasp". We use a planar-grasp to approximately describe the shape of the planar cross-section of the depth map in the region between two given pixels in the depth map (see Figure 4). We define the planar-grasp between pixels i and j as a vector quantity

$$g^{i,j} = [z^i, z^j, n^i, n^j, \delta^{i,j}, s^{i,j}]$$

where z^i, z^j, n^i, n^j are as defined above, $\delta^{i,j}$ is the "grasp-depth" (i.e. the length of the portion of the object that the gripper would grab along the grasp orientation), and $s^{i,j}$ is the spread of the grasp (the distance between the gripper fingers) (see Figure 4). We will formulate an objective function to locate planar regions in the point cloud which approximately fit the 2D cross-section of the gripper interior.

Lets define a feature vector f for a planar-grasp $g^{i,j}$ which captures the following desirable characteristics:

1) The surface normals at the contact points and the vector connecting the two contact points are aligned.
2) The "grasp-depth" is large enough to be robust to slip and provide a stable grasp.
3) The grasp "spread" is large enough to provide a stable grasp.

To simplify notation, we will refer to $f(g^{i,j})$ as $f^{i,j}$.

$$f^{i,j} = \left[min(-n^i \cdot c^{i,j}, n^j \cdot c^{i,j}),\ min\left(\frac{\delta^{i,j}}{\delta_{des}}, \frac{\delta_{des}}{\delta^{i,j}}\right),\ min\left(\frac{s^{i,j}}{s_{des}}, \frac{s_{des}}{s^{i,j}}\right)\right]$$

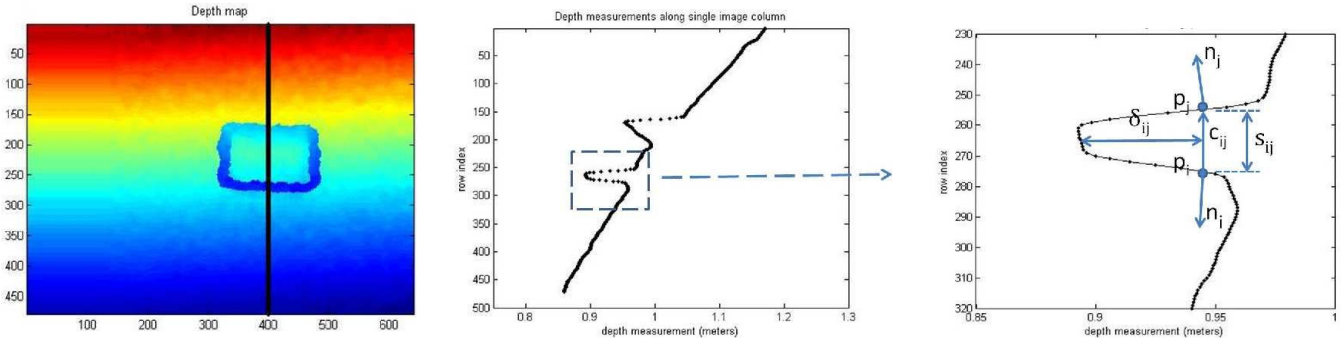
where

$$c^{i,j} = \frac{p^j - p^i}{\|p^j - p^i\|};\ p^i, p^j \text{ are the 3D points at pixels } i, j$$

$$\delta_{des} = \text{desired grasp-depth}$$

$$s_{des} = \text{desired grasp spread}$$

See Figure 4 for a description of p^i, p^j, and $c^{i,j}$.



(a) Depth map denoting search column. (b) Depth measurements for the column. (c) An example "planar-grasp" in the depth map.

Fig. 4. Figures defining a "planar-grasp" in the depth map of an open rectangular container.

We choose an objective function J_g which is linear in the feature vector. For planar-grasp $g^{i,j}$,

$$J_g^{i,j} = \theta^T f^{i,j}$$

where $\theta \in R^3$ is a vector of weights.

We want to find the planar-grasp for which this function is maximized over all possible planar-grasps in the depth map. For an image with M pixels, there are $M!/(2!(M-2)!)$ distinct pairs of points (and thus possible planar-grasps). For a 640x480 image, this is a very large number. We therefore prune possibilities that are clearly suboptimal, such as those which are infeasible for the geometry of our gripper (i.e. those with grasp-depth or spread which are too large defined by δ_{max} and s_{max}).

$$\begin{aligned} \text{maximize}_{i,j} \quad & J_g^{i,j} \\ \text{subject to} \quad & i \in 1...M \\ & j \in 1...M \\ & \delta^{i,j} \leq \delta_{max} \\ & s^{i,j} \leq s_{max} \end{aligned}$$

where

$$s_{max} = \text{maximum possible gripper spread}$$
$$\delta_{max} = \text{distance between the gripper tip and palm}$$

We search for the top n planar-grasps (i.e. those which come closest to maximizing our objective) by considering feasible planar-grasps along each column of the image. By rotating the image in small enough increments and repeating, all possible planar-grasps could be considered. However, despite ruling out infeasible planar-grasps by the spread and depth constraints, the number of possible grasps in the depth map is still very large. Thus we consider a representative subset of all the possible planar-grasps by rotating the image in larger increments.

Figure 5 displays the detected planar-grasp points for a rectangular-shaped storage container for search angles of 0°, 45°, 90°, and 135°. Notice how the 0° search (which is equivalent to searching down a column of the depth map) produces many planar-grasps on the front and back edges of the container, while the 90° search (which is equivalent to searching across a row of the depth map) produces many planar-grasps on the side edges of the container. Searches at angles of 45° and 135° produce planar-grasps along all the object edges. From these results, we see that a fairly coarse search generalizes well enough to identify planar-grasps at many orientations in the depth map. [1]

2) Identifying Robust Grasps: We have a weighted distribution of planar-grasps for each orientation of the image in which we searched. From Figure 5 we see that the highest scoring planar-grasps are distributed in regions on the object which visually appear to be good grasp regions. However, a single planar grasp alone is not sufficient to define a good grasp region because our gripper is 3-dimensional, but a series of similar planar-grasps along a line will fully define a good grasp region for our gripper.

To find regions in the depth map which approximately match the 3D shape of the gripper interior, we consider pairs of similar planar-grasps which are separated by a distance slightly larger than the width of the gripper fingertip pads (providing a margin of error to allow for inaccuracies in the control and perception). By "similar planar-grasps", we mean those which have similar grasp-depth and spread. For simplicity of notation, we will denote a pair of planar-grasps as g^μ and g^η, where $\mu = (i,j)$ and $\eta = (k,l)$. Let us define some additional geometric quantities which will aid in computing a grasp pose for the end-effector from the planar-grasp pair geometry. We define the planar-grasp center for g^μ as $p^\mu = (p^i + p^j)/2$ where p^i and p^j are the 3D points for the pixels i and j. Given the planar-grasp pair g^μ and g^η, we define the vector oriented along the planar-grasp centers as $v^{\mu,\eta} = p^\eta - p^\mu$. The contact points associated with the pair of planar-grasps approximately forms a plane and we denote the normal vector to this plane as $u^{\mu,\eta}$. Figure 1 shows how $u^{\mu,\eta}$ and $v^{\mu,\eta}$ define the grasp orientation.

We will represent a grasp pose as the position of the gripper $p \in R^3$, the quaternion defining the gripper orientation $q \in R^4$, and the gripper spread. The planar-grasp pair g^μ and g^η define a grasp pose which we will denote as $G^{\mu,\eta}$.

$$G^{\mu,\eta} = [p^{\mu,\eta},\ q^{\mu,\eta},\ s^{\mu,\eta}]$$

[1] This search can be parallelized over four CPU cores to provide increased computational speed.

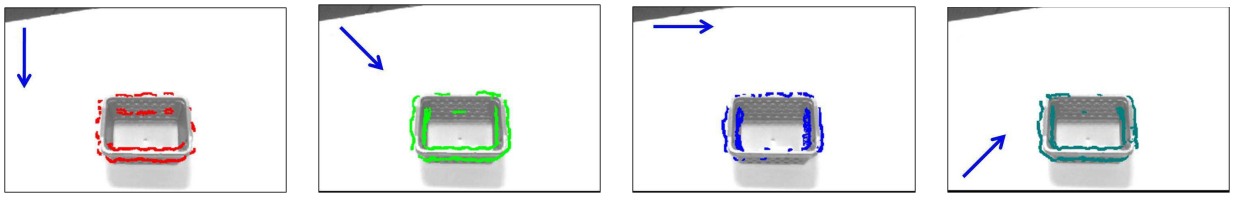
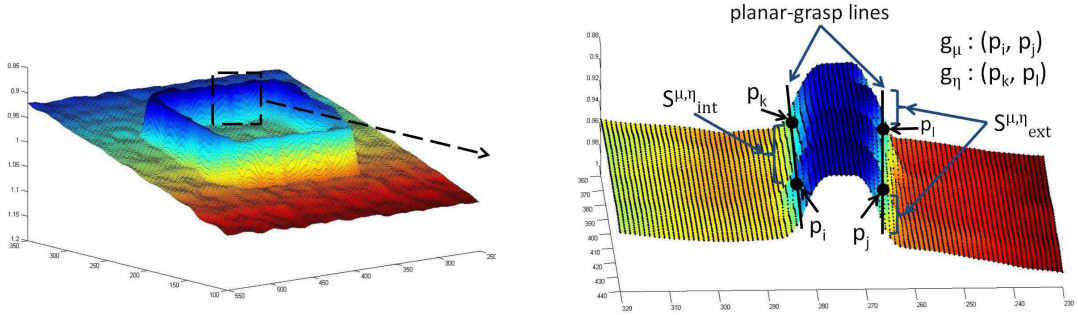


Fig. 5. Finger contact locations for the "best" planar-grasps found by searching the depth map at angles of 0°, 45°, 90°, and 135° (search direction shown by blue arrows). Note how different search angles find good planar-grasps on different portions of the container.



(a) Depth map denoting region defined by a pair planar-grasp pair. (b) Geometry of a planar-grasp pair.

Fig. 6. Figures defining the "planar-grasp" pair geometry in the depth map of an open rectangular container.

where

$$p^{\mu,\eta} = (p^\mu + p^\eta)/2$$
$$q^{\mu,\eta} = h(v^{\mu,\eta}, u^{\mu,\eta})$$
$$s^{\mu,\eta} = (s^\mu + s^\eta)/2$$

where h is a function that converts two vectors (or equivalently a rotation matrix) representing the orientation of the gripper to a quaternion.

We now define an objective function J_G to assign a score to a planar-grasp pair (which defines a grasp pose for the end-effector according to the equations given above). The region between two planar-grasps will define a good grasp if they are similar to each other and similar to the planar-grasps which lie along the line between them (i.e. we would like all the neighboring planar-grasps to have similar depth measurements for the two points and similar grasp-depth and spread). However, we also want the objective function to capture more global properties to choose grasps closer to the center of an object (Figure 2). Thus our objective function also considers the planar-grasps that lie along the line extending past the edge of the grasp region defined by the planar-grasp pair.

Let $S^{\mu,\eta}_{int}$ be the set of planar-grasps which lie along the line between the planar-grasp pair g^μ and g^η. Let $S^{\mu,\eta}_{ext}$ be the set of planar-grasps which lie along the line extending a given distance past the edge of the grasp region defined by the planar-grasp pair g^μ and g^η (see Figure 6). Given two planar-grasps of similar grasp-depth and spread g^μ and g^η,

$$J^{\mu,\eta}_G = \sum_{g \in S^{\mu,\eta}_{int}} J_g * discount_{int} + \sum_{g \in S^{\mu,\eta}_{ext}} J_g * discount_{ext}$$

where J_g is as defined in section III-A.1 and

$$discount_{int} = \begin{cases} 0 & \text{if } g \text{ does not exist,} \\ -\infty & \text{if } \delta > \delta_{max} \text{ or } s > s_{max} \\ \frac{1}{1+\Delta} & \text{otherwise} \end{cases}$$

$$discount_{ext} = \begin{cases} 0 & \text{if } g \text{ does not exist,} \\ \frac{1}{1+\Delta} & \text{otherwise} \end{cases}$$

where Δ is the distance between the desired location for the contact points of the planar-grasp and the actual location.

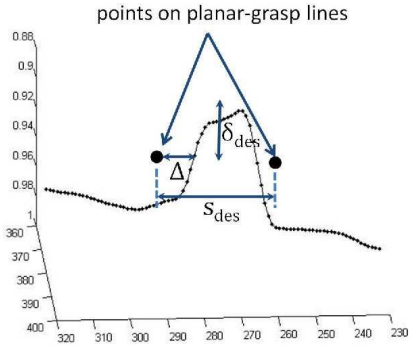


Fig. 7. A planar-grasp in the set $S^{\mu,\eta}_{int}$ is shown along with the geometry for the planar-grasp model defined by (g^μ,g^η).

The discount factor function accounts for the fact that the planar-grasps contained in the set $S^{\mu,\eta}$ will not exactly lie along the lines defined by the planar-grasp pair (g^μ,g^η) (see Figure 7). Planar-grasps which lie closer to the lines defined by the planar-grasp pair better fit the grasp model defined by the pair and are thus weighted more highly.

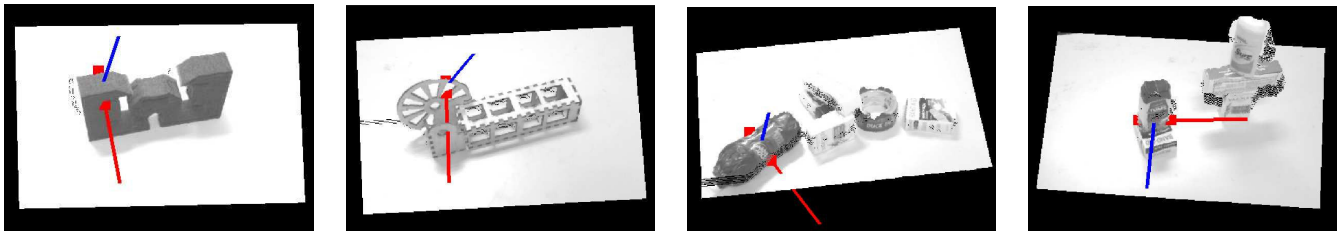


Fig. 8. 3D visualizations of the best detected grasp in several scenes. The colored axes represent the orientation of the grasp (blue is along the gripper wrist and red is normal to the gripper pads (as shown in Figure 1).

Finally, we want to find grasp $G^{\mu,\eta}$ defined by the planar-grasp pair (g^{μ},g^{η}), which maximizes J_G over all possible pairs of planar-grasps.

$$G^{\star} = \text{argmax}\, J_G^{\mu,\eta}$$

for all combinations of planar-grasp pairs (g^{μ}, g^{η}).

However, due to computational constraints, we search a subset of this large space by randomly sampling pairs of planar-grasps from the weighted distribution and repeating for a set number of iterations. Our approach is motivated by RANSAC ("RANdom SAmple Consensus") [18]. In the standard RANSAC formulation, the minimum number of data points required to fit a desired model are chosen and then the fit all of the data to this proposed model is evaluated (with outlier removal). In our algorithm discussed above, two planar-grasps define a full grasp pose. We randomly sample a pair of planar-grasps and evaluate an objective function to determine how well the nearby data fits the proposed grasp. We keep the top n distinct grasps resulting from this procedure for a given scene. Figure 8 displays examples of the grasps found for various objects.

B. Barcode Localization and Recognition

We use a standard barcode reader software [16] to recognize barcodes on objects after grasping them. A simple integration of grasping and a barcode reader is not sufficient due to high variance in the shape of objects and the location of barcodes on them. Thus we use a motion planning strategy to provide the barcode reader with images of a given object from different orientations. The intuition is to increase the odds of finding the barcode by feeding images of every surface of the object to the barcode reader.

Once the robot has picked up the object, it moves the object close to the high resolution camera (Figure 13(c)). To locate the barcode, the robot rotates its wrist in increments of 30° in view of the high-resolution camera (Figure 13(d)-(f)). The wrist rotations are followed by a maneuver to expose the surfaces that were previously hidden (Figure 13(g)). An image of the object at every rotation is fed to the barcode reader. Barcode readers require barcodes to be in a particular orientation. Thus, each image fed to the barcode reader is additionally rotated in the image plane in increments of 45° and then scanned for finding barcodes. [2]

We propose regrasping to handle the case of the barcode being occluded by the gripper. If the robot fails to find the barcode, it replaces the object back on the table in a different orientation, and the entire pipeline runs in a loop until the barcode is found. See Figure 9 for examples of detected barcodes on various objects.

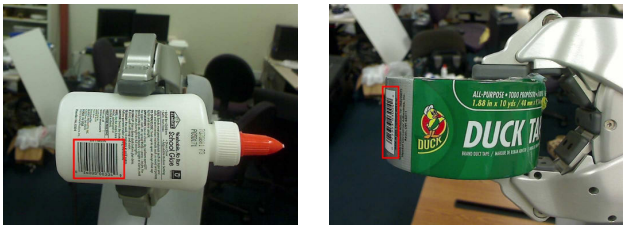


Fig. 9. Examples of detected barcodes.

IV. EXPERIMENTS

A. Hardware / Software

We demonstrate our perception and planning algorithms on the PR2 mobile robot platform. The input to our grasp detection algorithm is provided by the projected texture stereo system on the robot head [17]. To provide high-resolution, color images for the barcode reader, we use a 2 MP Logitech Webcam Pro 9000 mounted on the robot head.

B. Results

We first evaluate our grasping algorithm on a set of 6 unknown objects (see Figure 10) and compare the results to recent works on grasping novel objects. We performed 20 trials for each object by placing the individual objects on the table at various orientations. As shown in Table I, we achieved success rates equal to or better than the previous approaches for all objects.

We then evaluate the autonomous checkout robot application in two sets of experiments. For these experiments, we randomly selected 10 common items found in a pharmacy (see Figure 11). We restricted the set of objects to those which the PR2 gripper is mechanically capable of grasping at any possible orientation of the object. In the first set of experiments, each object was placed at 5 unique orientations

[2]The barcode search is executed in a parallel pipelined fashion in which planar rotation and barcode finding are run simultaneously on different CPU cores.

Fig. 10. Objects used for grasping experiments.

Fig. 11. Objects used for checkout experiments.

Fig. 12. Example of cluttered and very cluttered scenes.

in front of the robot one at a time. In the second set, the objects were placed several at a time with varying amount of clutter (see Figure 12). A trial is considered a success if the robot grasps the object, lifts it from the table, successfully locates and reads the barcode, and then drops the object into a paper bag. The robot was able to autonomously perform the required sequence in 49/50 of the single-object trials (see Table II) and 46/50 of the cluttered scene trials (see Table III). Regrasping was necessary 10 times due to the robot occluding the barcode with its gripper in the initial grasp. The barcode reader had a 100% accuracy on all the objects. The failures encountered were due to the object slipping from the gripper and falling outside the camera field of view or 2 objects being grasped at once.

TABLE I

PERFORMANCE OF OUR METHOD WITH PREVIOUS METHODS FOR THE TASK OF GRASPING NOVEL OBJECTS

Objects	Method in [19]	Method in [14]	Our Method
Mug	80%	90%	90%
Helmet	90%	75%	95%
Robot Arm	70%	80%	85%
Foam	70%	85%	95%
Cup	70%	85%	90%
Bowl	75%	-	95%
Mean/Std	75.8 $\pm$ 8.0%	83.0 $\pm$ 5.7%	91.6 $\pm$ 4.1%

See Figure 13 for snapshots of the robot experiments. Videos of these experiments are available at:

http://www.stanford.edu/∼ellenrk7/CheckoutBot

V. CONCLUSION

In this paper, we proposed a novel algorithm for grasping uknown objects given raw depth data obtained from a single frame of a 3D sensor. The reliability and robustness of our algorithm allowed us to create a system that enabled the robot to autonomously checkout items in a pharmacy store

TABLE II

SUCCESS RATE FOR SINGLE OBJECT CHECKOUT EXPERIMENTS.

Object	Grasping	Barcode Id.	Overall
Band-Aid	5/5	5/5	5/5
Hydrocortisone Cream	5/5	5/5	5/5
Glue	5/5	5/5	5/5
Cotton Rounds	4/5	4/4	4/5
Deodorant	5/5	5/5	5/5
Soap Bar	5/5	5/5	5/5
Duct Tape	5/5	5/5	5/5
Apricot Scrub	5/5	5/5	5/5
Soda Bottle	5/5	5/5	5/5
Tums	5/5	5/5	5/5

TABLE III

SUCCESS RATE FOR MULTIPLE OBJECTS CHECKOUT EXPERIMENTS.

Scene No.	Scenario	Overall Success Rate
1	Cluttered	4/4
2	Cluttered	4/4
3	Cluttered	4/4
4	Cluttered	4/5
5	Cluttered	4/4
6	Very Cluttered	6/6
7	Very Cluttered	5/5
8	Very Cluttered	4/5
9	Very Cluttered	5/6
10	Very Cluttered	6/7

setting with near perfect accuracy. Our grasping experiments also showed the effectiveness of our algorithm in comparison with other competitive methods.

We point out limitations of our approach to motivate potential areas of future work. Our approach will fail to identify a grasp for objects which are too large to fit in the gripper or objects which are too small and/or thin (e.g. a piece of paper or pencil lying on a table) to provide reliable depth measurements. For cluttered scenes, it is possible that the selected grasp will result in the robot simultaneously picking up two contacting objects (however, this only occurred once in our experiments). [3] Since our algorithm only uses a single view, occluded regions inherently introduce uncertainty in the grasp selection. This uncertainty could be reduced by adding additional views. Although we considered a two-fingered gripper, the general idea behind our approach is

[3]One could consider adding a feature to prefer a grasp region to be uniform in color (however, many of the objects in our experiments had very colorful, non-uniform patterns).

(a) detect a grasp (b) execute grasping (c) move object close to the camera (d) barcode search

(e) barcode search (f) barcode search (g) barcode detected and read (h) drop object into the paper bag

Fig. 13. Snapshots of our robot performing the checkout experiment.

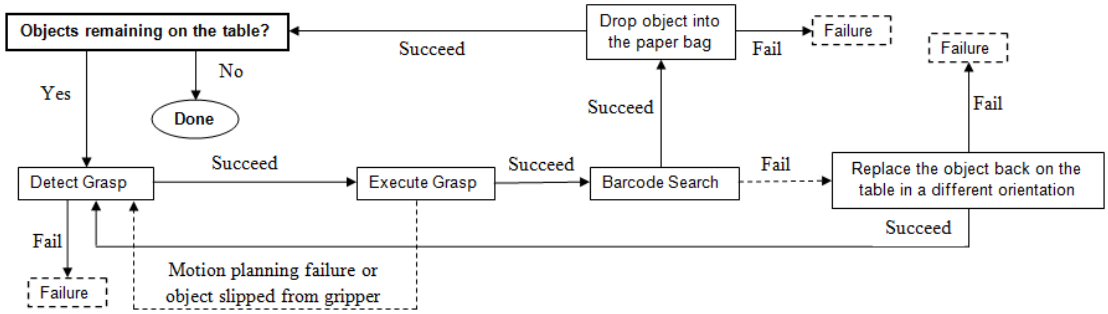


Fig. 14. The state machine model of the checkout appliation pipeline: dashed lines indicate failure recovery cases.

extend-able to other end-effectors. For example, there are several common pre-shapes the human hand takes on to perform common tasks, so the algorithm could perform a search (with appropriate choice of the objective function) for each pre-shape. Despite these limitations, the robustness of our algorithm to variations in object shape and the freedom from offline training on hand-labeled datasets allows our approach to be readily used in many novel applications.

REFERENCES

[1] T. Sakai, H. Nakajima, D. Nishimura, H. Uematsu, and K. Yukihiko, "Autonomous mobile robot system for delivery in hospital," in *MEW Technical Report*, 2005.

[2] A. Bicchi and V. Kumar, "Robotic grasping and contact: a review," in *ICRA*, 2000.

[3] V. Nguyen, "Constructing stable grasps," *IJRR*, 1989.

[4] K. Lakshminarayana, "Mechanics of form closure," in *ASME*, 1978.

[5] A. T. Miller, S. Knoop, H. I. Christensen, and P. K. Allen, "Automatic grasp planning using shape primitives," in *ICRA*, 2003.

[6] S. Srinivasa, D. Fergusun, M. V. Weghe, R. Diankov, D. Berenson, C. Helfrich, and H. Strasdat, "The robotic busboy: Steps towards developing a mobile robotic home assistant," in *10th International Conference on Intelligent Autonomous Systems*, 2008.

[7] A. Morales, P. J. Sanchez, A. P. del Pobil, and A. H. Fagg, "Vision-based three-finger grasp synthesis contrained by hand geometry," in *Robotics and Autonomous Systems*, 2006.

[8] A. Jain and C. Kemp, "El-e: An assistive mobile manipulator that autonomously fetches objects from flat surfaces," *Autonomous Robots*, 2010.

[9] M. Richtsfeld and M. Vincze, "Detection of grasping points of unknown objects based on 2 1/2d point clouds," 2008.

[10] A. Edsinger and C. Kemp, "Manipulation in human environments," in *IEEE-RAS International Conference on Humanoid Robotics*, 2006.

[11] K. Hsiao, S. Chitta, M. Ciocarlie, and E. G. Jones, "Contact-reactive grasping of objects with partial shape information," in *IROS*, 2010.

[12] M. Popovic, D. Kraft, L. Bodenhagen, E. Baseski, N. Pugeault, D. Kragic, T. Asfour, and N. Kruger, "A strategy for grasping unknown objects based on co-planarity and colour information," *Robotics Autonomous Systems*, 2010.

[13] A. Saxena, J. Driemeyer, and A. Y. Ng, "Robotic grasping of novel objects using vision," *IJRR*, 2008.

[14] Q. V. Le, D. Kamm, A. Kara, and A. Y. Ng, "Learning to grasp objects with multiple contact points," in *ICRA*, 2010.

[15] I. Ehrenberg, C. Floerkemeier, and S. Sarma, "Inventory management with an rfid-equipped mobile robot," in *"International Conference on Automation Science and Engineering"*, 2007.

[16] "Softek barcode reader," http://www.softeksoftware.co.uk/linux.html, 2010.

[17] K. Konolige, "Projected texture stereo," in *ICRA*, 2010.

[18] M. A. Fischler and R. C. Bolles, "Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography," in *Communications of the ACM*, 1981.

[19] A. Saxena, L. Wong, and A. Y. Ng, "Learning grasp strategies with partial shape information," in *AAAI*, 2008.