Evaluating End-Effector Modalities for Warehouse Picking: 
A Vacuum Gripper vs a 3-finger Underactuated Hand

Zakary Littlefield, Shaojun Zhu, Hristiyan Kourtev, Zacharias Psarakis
Rahul Shome, Andrew Kimmel, Andrew Dobson, 
Alberto F. De Souza and Kostas E. Bekris

Abstract—This paper studies two end-effector modalities for warehouse picking: (i) a recently developed, underactuated three-finger hand and (ii) a custom built, vacuum-based gripper. The two systems differ on how they pick objects. The first tool provides increased flexibility, while the vacuum alternative is simpler and smaller. The aim is to show how the end-effector influences the success rate and speed of robotic picking. For the study, the same planning process is followed for known poses of multiple objects with different geometries and characteristics. The resulting trajectories are executed on a real system showing that, under different conditions, different types of end-effectors can be beneficial. This motivates the development of hybrid solutions.

I. INTRODUCTION

A critical task in warehouse automation [1], which is not yet sufficiently automated, involves picking products from shelving units. The design of an appropriate end-effector, which effectively interacts with a wide variety of objects, is a critical consideration in this context. The end-effector must enable robust picking in a variety of setups, while low complexity and cost are key enablers of adoption. Versatile robotic equipment that can operate in a minimally engineered space where humans are also present is highly desirable. This could motivate the adoption of anthropomorphic hands, which aim to effectively perform human tasks by mimicking their structure. Indeed, highly articulated humanoid hands have been available for some time [27] and are improving [12]. Due to the significant effort and cost necessary to calibrate, maintain, and repair such solutions, few are actually in use.

Simpler end-effectors can provide a degree of generality yet untapped by most autonomous system [21] and are appropriate for the automation industry where, due to stringent requirements on speed, precision, and reliability, simple gripping solutions can be beneficial [23]. Underactuated and compliant solutions are especially popular, since they offer the advantage of requiring few actuators to control many degrees of freedom, and they are overall less complex and typically cheaper to produce [8], [13], [34]. The current work uses the RightHand Robotics “ReFlex” hand [24], shown in Fig. 1(a) as a representative of such solutions.

An alternative corresponds to vacuum technology. Creating constraints by grasping with a multi-finger hand requires careful alignment of opposing forces on the object, whereas vacuum requires a single area of contact. Vacuum minimizes translational and rotational movement, which makes the approach robust against wrenching forces. A drawback is the poor ability to manipulate an object. There are also objects where there is little surface where suction can be applied. Vacuum has been used in the industry for quite some time, but typically for objects that are obvious targets for suction. The potential of vacuum for general warehouse picking was exemplified by the Amazon Picking Challenge (APC) [7]. The challenge involved picking a subset of 25 different objects from the bins of an Amazon-Kiva Pod. Teams that made use of suction were proven
The Rutgers team in the APC collaborated with UniGripper Inc. to design a vacuum-based solution, shown in Fig. 1(b), for warehouse picking.

The current work describes experiments that study the grasping performance of the two end-effectors of Fig. 1 in a setup similar to the APC, as in Fig. 2. The same standard planning process is employed to find paths. This allows to evaluate the difficulty in finding successful grasps with the two end-effectors for objects with variable geometries and sizes. An extensive number of grasps are tested first in simulation and then on the real system. There are few such experimental studies, which compare vacuum and multi-finger technology. A recent comparative study aims to evaluate different designs for underactuated grippers but does not consider vacuum [26]. Another recent work describes a related solution for the APC [33], [32].

The experiments indicate that the vacuum solution is surprisingly versatile in successfully picking a variety of objects under different conditions. The volume of the 3-finger hand complicates its use in tight spaces, reduces the number of potential grasps and increases planning time. Nevertheless, there are clearly object geometries and placements where the vacuum-based solution faces challenges. This encourages the adoption of hybrid solutions, where both modalities exist on the same end-effector.

This work focuses on stable grasping, and not in-hand manipulation, such as reorienting the objects pose in hand [5], or exploiting interactions with the environment [10]. Similarly, the focus is on objects of known shape, which is reasonable for logistics applications. There is an important line of work focusing on issues related to uncertainty in the context of grasping [30], [15], [16], [9], [17]. Sensing errors are isolated from the proposed evaluation, however, to focus on the interplay between the end-effector and the planning process.

II. HARDWARE SETUP AND DESIGN

The hardware includes a dual-arm Yaskawa Motoman SDA10F (Fig. 2), a ReFlex Beta Hand and a custom built vacuum gripper. The robot has 7 DoFs on each arm and one additional DoF at its torso, which allows it to reach all 12 bins of the Kiva pod with both arms.

A. The ReFlex Hand

The “ReFlex” hand (Fig. 1(a)), is the successor of the i-HY hand used in the DARPA ARM Challenge [24]. It is underactuated, 3-finger hand designed to be simple, inexpensive and durable. The fingers are interchangeable and are comprised of a proximal and a distal segment. It employs 4 Dynamixel servo motors, 3 to flex the fingers and thumb, while the fourth changes their preshape configuration. The fingers are interconnected by gears and move in sync. Communication and control is achieved over an ethernet connection via a ROS package.

The hand provides 3 pre-programmed grasps called “preshape”s: 1) Parallel grasp: fingers positioned at 0 degrees (parallel to each other and the thumb). When closed all 3 fingers interlace. It is useful for holding and operating heavy tools or other cylindrical objects. 2) Spherical/Power grasp: fingers and thumb at 120 degrees from each other. It is useful for grasping large objects or small objects manipulated in a circular pattern (e.g. screwing on a nut on a bolt). 3) Pinch grasp: fingers are spread as far as possible (180 degrees). The thumb remains stationary while the fingers are closed and meet at the fingertips. This configuration does not provide a very strong grasp but can pick small objects.

B. The UniGripper Gripper

The UniGripper tool (Fig. 3(left)) was designed by the authors and Unigripper to deal with objects inside shelves. A Kiva pod bin has about 20(width)x25(height)x40(depth) cm size. The two middle rows of bins have 20cm height. Given the depth, it is difficult to reach objects deep inside them without extending the robot’s arms. The design has a shaft of 13 cm that ends in the planar vacuum gripper in one extremity. The planar gripper is able to move up to 90 degrees according to an 1-DOF joint.

1http://www.righthandrobotics.com/main:reflex
The planar surface uses a Unigripper patented technology. A sponge glued to the planar vacuum gripper, with holes in front of each valve, allows grasping objects that do not expose planar surfaces. The on-off vacuum actuator highlighted in Fig. 1(b) was switched to a Figzelli L16-R Miniature Linear Servo actuator shown in Fig. 3(right). To control the gripper, a hardware module based on the Micro Maestro 6-channel USB Servo Controller was built. A software module implements a ROS driver for controlling the joint pose, and a ROS service to turn the vacuum on or off.

III. GRASPING PLANNING METHODOLOGY

A standard method for grasping was employed: a) generating candidate grasps, b) using an IK solver and a collision detector to verify grasp validity and c) planning collision-free trajectories.

A. Generating Grasping Poses

1) “ReFlex” Hand: GraspIt! [22], [6] was used to generate a set of predefined “virtual contact points” on the 3D model of the end-effector (Fig. 3(a)). The planner samples various hand positions and orientations to bring the contact points close to the object surface. Each grasp is then evaluated and scored according to a metric [25], [2]. Such metrics consider the objects’ 3D geometry and contact-point information, so as to detect high-quality grasps that can withstand disturbances [19], [11], [20].

Many grasps result in collisions as in the grasp of Fig. 4(b) which will collide with the resting surface unless the object is balancing on its small side, which is unlikely. To detect these cases, the object was placed on a resting surface. The hand, attached to a virtual arm link, is only allowed to approach the object from a specific direction as in Fig. 4(c). 24 axis-aligned object poses are considered, as in Fig. 4(d) to generate 30 grasps per pose. Once object pose recognition is performed, the grasps for the 2 or 3 closest axis-aligned poses are considered.

The ReFlex hand was limited to parallel (Fig. 4(b)) and pinch grasps (Fig. 4(e)). The GraspIt! eigengrasp planner used the “Hand Contacts” and “Potential Quality” functions to rank grasps and executed 40,000 “Simulated Annealing” iterations. Upon completion, GraspIt! returned the top 50 pre-grasps ranked based on “grasp energy”. Some of the resulting grasps, despite having a high score were not robust (see Fig. 4(e)) and they were visually pruned.

2) Unigripper: Each object was approximated by a cuboid. As the vacuum solution requires part of the planar gripper surface to touch the object, the process simply samples grasps on each face of the cuboid. For each pose, the object is considered to be inside a constrained setup similar to Fig. 4(c) and cannot be grasped from the bottom or the back. Grasps are sampled only for the front, top, left, and right face:

Front face: 2 positions are sampled ($p_0$: face center, and $p_1$: 1/6 of the face’s height above $p_0$) and 8 orientations for each position (multiples of 45°).

Top face: 2 positions are sampled ($p_0$: face center, and $p_1$: 1/6 of the face’s length in front of $p_0$) and 3 orientations for each position (-45, 0, and 45° rotations).

Left/Right faces: 3 positions ($p_0$: face center, $p_1$: 1/6 of face’s height above $p_0$, and $p_2$: 1/6 of face’s length in front of $p_1$) and 3 orientations (-45, 0, and 45°).

This results in 40 grasps per object pose. Grasps that collide with the shelf or grasps that do not provide enough overlap between the gripper and the object face are filtered out. The overlap must be either more than half of the total area of the sponge or the area of the object face, whichever smaller. Eventually the resulting databases stored on average 30 grasps for each pose.
B. Inverse Kinematics and Collision Checking

Given an object pose, the planner identifies the three databases with the most similar axis-aligned poses. These databases are combined into a single one, and any repeated grasps are removed. The planner iterates over the resulting grasps and queries an external package [3] for a valid IK solution. Collision checking is performed. If the grasping state is collision free, the planner computes local “reaching” and “retracting” plans. These plans allow the end effector to both reach and retract from a pre-specified offset from the object.

C. Planning and Control for Grasping

Once a valid grasp has been found, the planner must figure out arm trajectories. The components of this process are shown in Fig. 6.

- **Transit Plan**: Takes the arm from an initial state to a “connection” state. This plan is retrieved from a precomputed PRM* roadmap [14], which takes into account the arm and the static geometry. To find the shortest collision-free path given new objects in the world, an iterative lazy evaluation process is followed.

- **Transit Connection Plan**: Takes the arm from the “connection” to the “retracted” state. This is computed online using a goal-biased RRT [18]. The start state is the “retracted” state and the candidate goals are the nearest roadmap nodes to the “retracted” state. The plan is executed in reverse.

- **Reaching Plan**: Takes the arm from the “retracted” to the “grasped” state. To accommodate for the time it takes for the Unigripper to establish suction with an object, the reaching portion of the plan was slowed down. A slight “pushing” action, to enable better contact between the suction sponge and the object surface.

- **Retracting Plan**: Takes the arm from the “grasped” to the “retracted” state. It results in the arm to move a short distance away from the initial object pose.

- **Transfer Connection Plan**: Takes the arm from the “retracted” state to a new “connection” state. This plan is computed online using a goal-biased RRT, similar to the transit connection, but with the object in hand.

- **Transfer Plan**: Takes the arm from the new “connection” to the “target” state. This plan can be retrieved from a precomputed roadmap using lazy evaluation. The resulting path is accepted only if it is collision-free given that the arm is holding the object.

IV. Evaluation of End Effectors

Every experiment starts from a collision free state where the arms are outside the shelf. The objects are placed in a specified pose inside the shelf and must be dropped at a target location outside the shelf. The parameters that vary are the following:

- 2 end-effectors: Unigripper and ReFlex hand.
- 12 objects that vary in shape, weight, and material.
- 2 bins of the APC shelf of different size.
- 3 poses in each bin starting with the pose shown in Fig. 7. The second pose is rotated 45° about +Z (pointing up) and the third is rotated 90° about +X (pointing into the shelf) relative to the initial pose.

A. Planning Evaluation

All the grasps for the corresponding object/end-effector database are attempted. For collision-free grasps, the motion planner is invoked and the ratio of successful planning attempts is recorded. The duration of the generated motions is used to measure the quality of the solution. For every object/pose/bin/arm combination, the two grasps that yield the best solution quality are tested in a real-world execution.

Table I: The first column identifies the ratio of precomputed grasps reachable in terms of IK and collision free. About 4.5% of the Unigripper and 3.33% of the ReFlex grasps end up being valid. Then, the motion planner solves the majority of the corresponding problems (63% to 64%). The resulting path quality is equivalent for the two end-effectors. Grasping evaluation is significantly faster than motion planning. Planning takes 3 seconds on average when a plan is found and about 12 seconds upon a failure.

Table II: It shows that the picking problem is harder for bigger objects (cheezit, feline and mark twain), which are difficult to maneuver out of the bin. The object pose also affects whether grasping succeeds (e.g., mark twain, kong duck, elmers glue, and outlet plugs do not allow the Reflex fingers to wrap around the object in every pose). For objects like feline and elmers glue, a narrow face may be not allow a UniGripper grasp.
TABLE I
PLANNING STATISTICS

<table>
<thead>
<tr>
<th>End Effector</th>
<th>Grasp success ratio</th>
<th>Planning success ratio</th>
<th>Solution quality (sec)</th>
<th>Comp. time upon success (sec)</th>
<th>Comp. time upon grasping failure (sec)</th>
<th>Comp. time upon planning failure (sec)</th>
<th>Average failure time (sec)</th>
</tr>
</thead>
<tbody>
<tr>
<td>UniGripper</td>
<td>0.0449</td>
<td>0.6406</td>
<td>47.1128</td>
<td>3.3817</td>
<td>0.1014</td>
<td>12.5685</td>
<td>0.3098</td>
</tr>
<tr>
<td>Reflex</td>
<td>0.0333</td>
<td>0.6335</td>
<td>46.1100</td>
<td>3.3590</td>
<td>0.0993</td>
<td>12.4044</td>
<td>0.2552</td>
</tr>
</tbody>
</table>

TABLE II
THE PLANNING SUCCESS RATIO FOR EVERY OBJECT.

<table>
<thead>
<tr>
<th></th>
<th>spark plug</th>
<th>cheezit</th>
<th>crayola</th>
<th>rolodex</th>
<th>elmers glue</th>
<th>feline</th>
<th>sticky notes</th>
<th>kong duck</th>
<th>kyjen</th>
<th>mark twain</th>
<th>outlet plugs</th>
<th>duck toy</th>
<th>Average</th>
</tr>
</thead>
<tbody>
<tr>
<td>UniGripper</td>
<td>0.9487</td>
<td>0.1573</td>
<td>0.6327</td>
<td>0.6623</td>
<td>0.9565</td>
<td>0.5185</td>
<td>0.8214</td>
<td>0.8095</td>
<td>1.0000</td>
<td>0.4615</td>
<td>0.6667</td>
<td>0.7813</td>
<td>0.6371</td>
</tr>
<tr>
<td>Reflex</td>
<td>0.9667</td>
<td>0.2143</td>
<td>0.6552</td>
<td>0.8889</td>
<td>0.7879</td>
<td>0.5882</td>
<td>0.9412</td>
<td>0.6667</td>
<td>0.8750</td>
<td>0.3000</td>
<td>0.5789</td>
<td>0.9375</td>
<td>0.7215</td>
</tr>
</tbody>
</table>

Fig. 7. Successful grasps out of two attempts for each object/pose/bin combination. The objects are placed in pose 1 in the photos. If an object was successfully retrieved from the shelf, the corresponding end effector image is present for that pose entry. (left: UniGripper, right: ReFlex)

Fig. 8. A visual breakdown of the performance of the two end-effectors in real-world experiments. The number of successful grasps and retrievals from the shelf is plotted, along with the number of failures with respect to grasping the object and transferring it.

B. Physical Evaluation

The two best quality solutions were replayed for each pose provided by simulation. A grasp may fail because the object was moved before grasping, or because the item’s shape prevented a successful grasp. Even if the item was grasped, it may be dropped during its transfer. In Figure 8, the success count for each end effector and item is presented, as well as the reason for failure.

Figure 7 demonstrates that the use of a specific end effector may be more advantageous. For instance, the “mark twain” item, was successfully grasped by the UniGripper. The ReFlex hand, however, is unable to grab a book while lying on a flat surface. “Rolodex” was almost impossible for the UniGripper to grasp, while the ReFlex was successful in most attempts. Even for cases that solutions were provided from planning for every pose of an item (i.e., “kyjen”), UniGripper was unable to perform a successful grasp.

Overall, 122 trajectories were tested for the UniGripper in real-world experiments. Out of those, 63 succeeded. The majority of failures arise from three objects (“rolodex”, “kyjen” and “duck toy”) for which it is difficult to establish suction. For the ReFlex hand, 85 trajectories were computed, which is a smaller number. Out of those, however, 62 trajectories were successful. These failures are more equally distributed over objects.
and more unpredictable. Out of the 78 considered poses (2 bins x 3 poses x 12 objects), it was possible to find a successful grasp with either end effector 57 times. Thus, small-profile end effectors with both modalities of caged grasps and vacuum may be the most successful.

V. DISCUSSION

The UniGripper is suited to objects with smaller profiles and flat surfaces, such as books and boxes. The ReFlex hand is more suited to non-uniform objects that can be caged. This indicates that incorporating multiple grasping modalities may help warehouse picking. Due to the sheer variety of objects and the numerous poses these objects can achieve, using a single type of simple end-effector may not be sufficient.

The picking success rate is influenced by the planning process, which introduces some bottlenecks. Alternatives avoid IK solvers by utilizing the manipulator’s Jacobian to guide towards a grasp [29], [31]. A recent method employs dimensionality reduction to avoid pre-computed grasps [28]. Learning can also be employed, where techniques synthesize grasps by sampling candidates, ranking them based on a metric using experience generated through demonstrations, labeled data, or heuristics [4]. Considering sensing uncertainty is a possible extension, as well as including additional simple but flexible grippers, such as one that combines both modalities. It would also be interesting to consider more complex grasping strategies for the two end-effectors, such as exploiting interactions with the environment.

REFERENCES