Hierarchical Grasp Detection for Visually Challenging Environments

D. Morrison[1,2], N. Kelly-Boxall[1,2], S. Wade-McCue[1,2], P. Corke[1,2], and J. Leitner[1,2]

I. Introduction

Robotic grasping is a widely studied topic, and many different approaches exist [1], [2]. However, many of these approaches rely on having accurate position information for the target object, which is not always feasible in real-world scenarios. This is especially true for unstructured warehouse picking, which presents many perception challenges such as occlusion, clutter and the inability of visual sensors to accurately perceive certain types of objects. While there are pose estimation strategies based on model fitting [3] and shape completion [4], these are not robust to major occlusion or a complete absence of accurate visual information.

To address these real-world issues, we have developed a grasp detection system which uses a hierarchy of three different strategies to operate successfully under varying levels of visual uncertainty. The system was used as part of our winning entry to the Amazon Robotics Challenge (ARC) [5].

II. Custom Multi-Modal End-Effector

The ARC requires robots to autonomously pick a diverse set of items – including rigid, semi-rigid, hinged, deformable and porous objects – from heavily cluttered storage systems, a task that is challenging not only with regards to perception but also in terms of manipulation. For this purpose we designed a multi-modal end-effector to complement our grasp detection system (Fig. 1). To be able to manipulate the largest possible set of items, our end-effector comprises a suction gripper and parallel-plate gripper. The two tools are selectable using a 180-degree tool change mechanism which supports 6-DOF pose control for each tool.

The dual-ended design allows a no-compromise approach to the design of each tool as they do not physically share design space, in contrast to a multi-modal end-effector that combines multiple grasping modalities at the same end-point. Our system makes effective use of redundant design in an attempt to maximise grasping performance. We achieve this by designing the two tools to be complementary, first determining which subset of items could be reliably acquired by suction and targeting the design of the parallel plate gripper at the remaining item classes (mainly those which are small, deformable or porous).

A more in-depth analysis of the design of our end-effector is provided in [6].

This research was supported by the Australian Research Council Centre of Excellence for Robotic Vision (ACRV) (project number CE140100016).
Contact: douglas.morrison@hdr.qut.edu.au
[1]Authors are with the Australian Centre for Robotic Vision (ACRV).
[2]Authors are with the Queensland University of Technology (QUT).

III. Hierarchical Grasp Detection

Our grasp detection system uses a hierarchy of three different grasp detection strategies, each of which is able to work with progressively less visual information at the expense of precision in the grasp point detection. The most appropriate grasp detection strategy is chosen to make best use of the available visual information (Fig. 2). The system relies on an external object detection and segmentation system [5].

The system addresses two main challenges of cluttered item picking. Firstly, it is able to deal with occlusion by considering only the visible portions of objects for grasping, avoiding any challenges associated with pose estimation or model fitting by relying on visual information only. Secondly, the material characteristics of some items introduce perceptual challenges. Items which are textureless, reflective, transparent or black are difficult or impossible for many modern depth cameras to accurately perceive, in which case visual information may not be accurate even if the object is not occluded.

Many objects which have regular, matte surfaces can be accurately perceived by an RGB-D camera. In this case, our system utilises its most accurate approach (Fig. 2a). This approach computes surface normals in a grid across the segmented point cloud of the object, and uses a set of heuristics to rank the grasp quality of each, similar to [7]. Grasps are ranked based on their distance to edges and boundaries, angle to vertical and height relative to nearby objects. To produce a spatially diverse set of grasps, similarly ranked grasps in close proximity to one another are removed. This approach is most suited to the suction gripper, as objects that fulfil these criteria often have suitable suction attachment

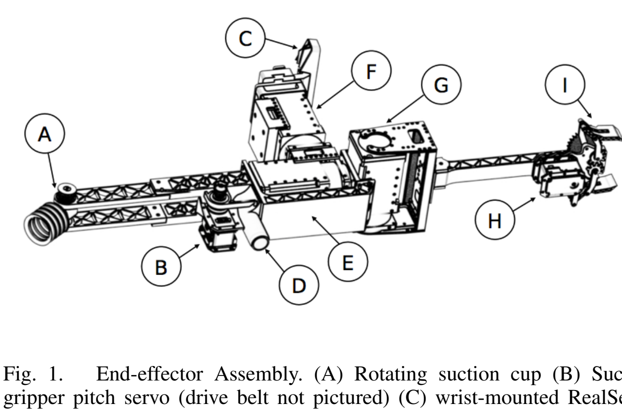

Fig. 1. End-effector Assembly. (A) Rotating suction cup (B) Suction gripper pitch servo (drive belt not pictured) (C) wrist-mounted RealSense camera (D) suction hose attachment (E) Roll motor (F) Yaw (tool-change) motor (G) Gripper pitch motor (H) Gripper servo (I) Parallel plate gripper [5]

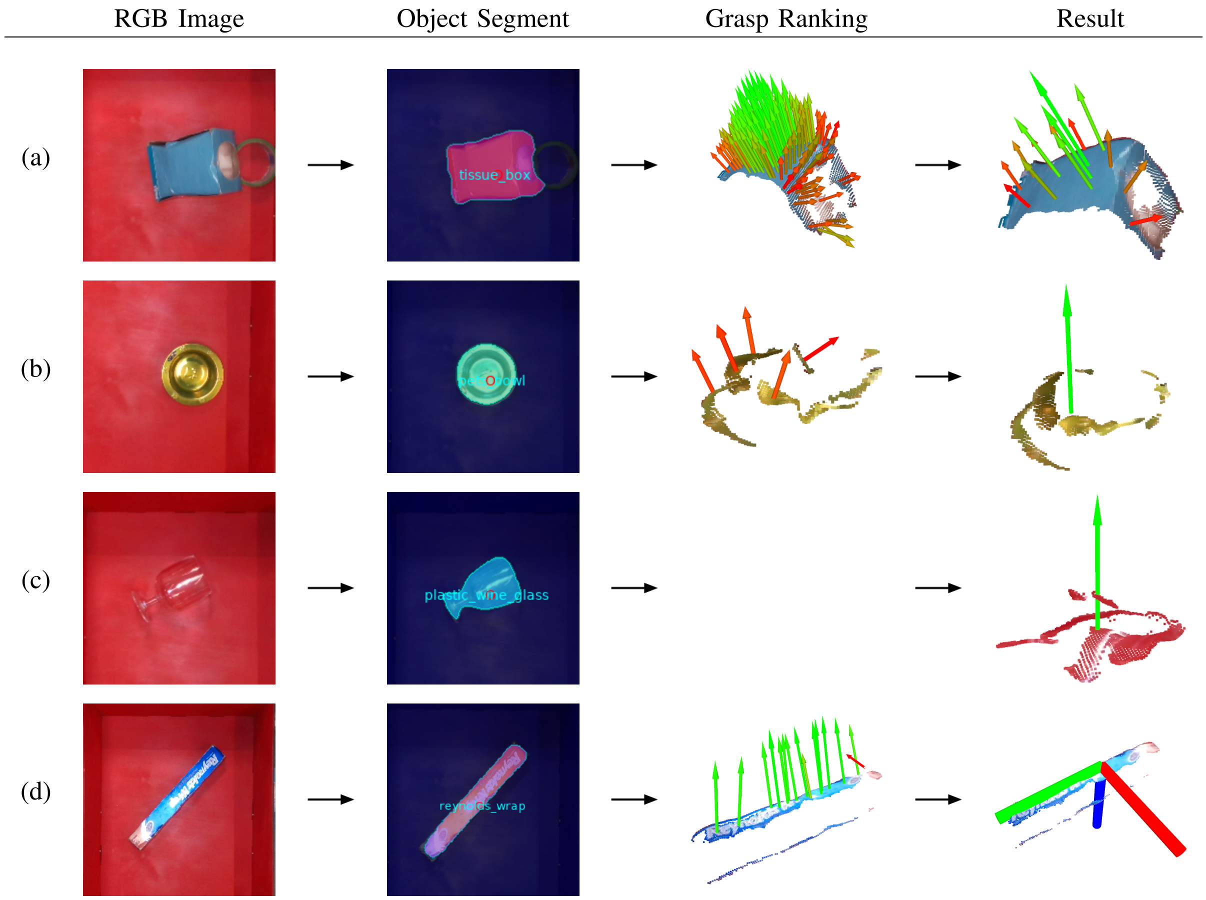

Fig. 2. Grasp detection under varying visual quality. Grasps are represented as vectors opposite to the direction of approach, with grasp rankings shown by colour (red to green = low to high) and length (longer is better). (a) Grasp ranking on the tissue box are chosen heuristically, (b) The incomplete point cloud of the metallic bowl does not produce any quality grasps, so the point cloud centroid is used, and (c) the plastic wine glass gives no valid depth data so the RGB segment is used to estimate its position in 3D space. (d) estimation of the cling wrap pose based on the RGB segment's principal component.

points.

Objects which are reflective or partially transparent often result in scattered point clouds. However, generally, many of the scattered points can be removed through a filtering process such as statistical outlier removal, resulting in an incomplete point cloud that can still be used to estimate the object's position (Fig. 2b). The above grasp ranking approach can be used on such a point cloud, but results in only lowly-rated grasp candidates. In this case, our system falls back to using the centroid of the filtered point cloud with a vertical orientation as a grasp point, which is applicable to both suction and gripping grasps.

Transparent and black objects rarely produce any valid depth information. In this case, where the above methods can't be applied, our system uses the known camera parameters to estimate the position of the object in 3D space using the centre-of-mass of the RGB segment (Fig. 2c). To counteract any uncertainty in the depth estimation, other sensors such as weight or pressure are used to detect contact.

To facilitate anti-podal gripping of objects, our system requires an estimate of the object's pose. As our system is designed to work in the presence of severe occlusions and incomplete visual information, it does not perform any model fitting. Instead, the grasp pose is aligned to the principal component of the object's RGB segment (Fig.2(d)).

In long term testing, our robotic vision and grasping system is able to pick a wide range of perceptually challenging objects in dense clutter with a 72% success rate using this hierarchical grasp detection system [5].

References

[1] J. Bohg, A. Morales, T. Asfour, and D. Kragic, "Data-driven grasp synthesis - a survey," *IEEE Transactions on Robotics*, vol. 30, no. 2, pp. 289–309, 2014.

[2] A. Sahbani, S. El-Khoury, and P. Bidaud, "An overview of 3d object grasp synthesis algorithms," *Robotics and Autonomous Systems*, vol. 60, no. 3, pp. 326–336, 2012.

[3] M. Ciocarlie, K. Hsiao, E. G. Jones, S. Chitta, R. B. Rusu, and I. A. Şucan, "Towards reliable grasping and manipulation in household environments," in *Experimental Robotics*. Springer, 2014, pp. 241–252.

[4] J. Varley, C. DeChant, A. Richardson, A. Nair, J. Ruales, and P. Allen, "Shape completion enabled robotic grasping," *arXiv preprint arXiv:1609.08546*, 2016.

[5] D. Morrison, A. Tow, M. McTaggart *et al.*, "Cartman: The low-cost cartesian manipulator that won the amazon robotics challenge," *arXiv preprint arXiv:1709.06283*, 2017.

[6] S. Wade-McCue, N. Kelly-Boxall *et al.*, "Design of a multi-modal end-effector and grasping system: How integrated design helped win the amazon robotics challenge," *arXiv preprint arXiv:1710.01439*, 2017.

[7] C. Lehnert, A. English, C. Mccool *et al.*, "Autonomous Sweet Pepper Harvesting for Protected Cropping Systems," *IEEE Robotics and Automation Letters*, vol. 2, no. 2, pp. 872–879, 2017.