Opening Cabinets and Drawers in the Real World using a Commodity Mobile Manipulator

Arjun Gupta⋆ Michelle Zhang⋆ Rishik Sathua Saurabh Gupta
University of Illinois at Urbana-Champaign
https://arjung128.github.io/opening-cabinets-and-drawers

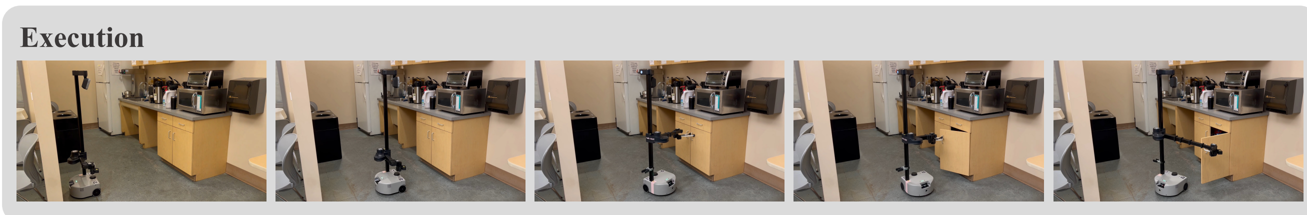

Figure 1. This paper presents a system for opening cabinets and drawers in novel real world environments using a commodity mobile manipulator. Here we visualize an example execution of our system interacting with a novel object in an unseen environment. We include the following frames: before navigation, after navigation, pre-grasp pose, during manipulation, and at the end of manipulation. We evaluate our system across 13 unseen environments from 10 distinct buildings for a total of 31 unique articulated objects in the real world.

Abstract

Pulling open cabinets and drawers presents many difficult technical challenges in perception (inferring articulation parameters for objects from onboard sensors), planning (producing motion plans that conform to tight task constraints), and control (making and maintaining contact while applying forces on the environment). In this work, we build an end-to-end system that enables a commodity mobile manipulator (Stretch RE2) to pull open cabinets and drawers in diverse previously unseen real world environments. We conduct 4 days of real world testing of this system spanning 31 different objects from across 13 different real world environments. Our system achieves a success rate of 61% on opening novel cabinets and drawers in unseen environments zero-shot. An analysis of the failure modes suggests that errors in perception are the most significant challenge for our system. We will open source code and models for others to replicate and build upon our system.

1. Introduction

This paper develops and evaluates a system for pulling open cabinets and drawers in diverse previously unseen real world environments (Figure 1). Opening articulated objects like cabinets and drawers presents hard technical challenges spanning perception, planning, and last centimeter control. These include accurate perception of object handles that are typically small and shiny, whole-body planning to drive the end-effector along the task constraint (*i.e.* trajectory dictated by the articulating handle), and dealing with execution errors in a task with low tolerances. All of these pieces have been studied at length in isolation [1, 5, 6]. Yet, how these modules interact with one another and what matters for successfully completing the task are not well understood. End-to-end learning via imitation or reinforcement circumvents these issues but is itself difficult because of the sample efficiency of learning and the unavailability of large-scale datasets for learning [2]. We take a modular approach and bring to bear state-of-the-art modules for perception and planning with a specific focus on studying how the different modules play with one another.

Specifically, for perception we extend a Mask RCNN model [3] to also output articulation parameters. For planning, we extend SeqIK, the recently proposed trajectory optimization framework [2] to produce whole body motion plans. Contrary to our expectation, just putting these two modules together did not lead to a successful system because of last centimeter errors in execution. Even slight inaccuracies in navigation and extrinsic camera calibration cause the end-effector to just be slightly off from the handle preventing handle grasping. To tackle this problem, we close the loop with proprioceptive feedback: predictions from visual sensors gets the end-effector in the vicinity of

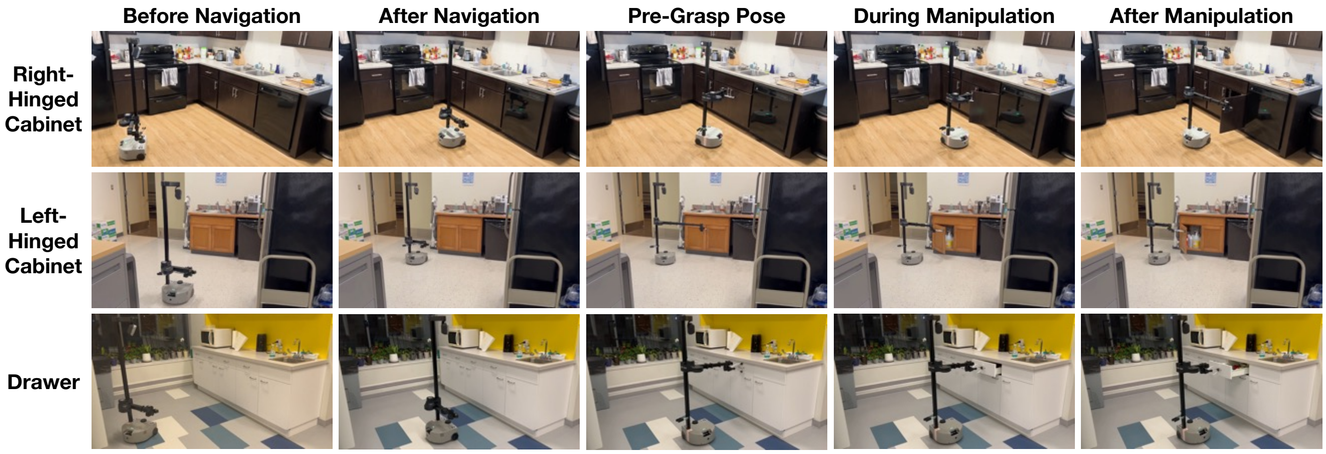

Figure 2. Example roll outs of our full system in various unseen environments. For each environment, we show the following frames: before navigation, after navigation, pre-grasp pose, during manipulation, and at the end of manipulation.

the handle and the actual grasping is done by leveraging contact sensors in the gripper and the arm.

Two other unique aspects of our study are a) the use of a *commodity mobile manipulator* and b) extensive testing in previously unseen diverse real world environments. Many previous papers have demonstrated specialized systems for similar problems [4]. Constructing specialized hardware for a given task can simplify the task at hand, at the cost of generality to other tasks. Therefore, we test our proposed system using the Stretch RE2, a *general purpose* commodity mobile manipulator, *without any hardware modifications*. Furthermore, this testing is conducted across 31 different articulated objects in 10 different buildings. Testing sites include offices, classrooms, apartments, office kitchenettes, and lounges. Our system achieves a 61% success rate in a zero-shot manner across this challenging testbed.

This broad study has allowed us to answer numerous questions about deploying such a system in the real world: a) what are the current bottlenecks in deploying a system for articulating objects in novel environments, b) how accurate should motion plans be for articulating objects, and c) what aspects of the current pipeline could benefit from machine learning? We find that perception is a major bottleneck for such a system, where inability to detect objects and handles accounts for 59% of the failures of the system. This calls for broad datasets with labels for cabinet and drawer articulation parameters in *diverse* settings. Our study also reveals that control is surprisingly robust to misestimations in the articulation parameters. Once the end-effector has acquired a firm grasp of the handle, the system is able to open the cabinet a non-trivial amount even with a radius error of up to $10cm$. Finally, errors from earlier parts of the pipeline (e.g. calibration and navigation) compound to lead to failures in grasping of the handle. This motivates the need for a learned visual last-centimeter grasping policy to correct for compounding errors.

In summary, this paper presents the design and evaluation of a mobile manipulation system to open cabinets and drawers using a commodity mobile manipulator. Large scale testing across 13 test sites in 10 buildings and 31 different cabinets and drawers reveals guidance for practitioners aiming to build similar systems.

2. Experiments

We work with the Stretch RE2 robot for our experiments. We present our full end-to-end system test results, in which our system is evaluated across 10 buildings and a total of 31 novel articulated objects. This test set of objects does not overlap with the set used for development. We assume the robot previously navigated toward the target object, so for each test it is positioned approximately $1.5m$ from the object with the camera oriented to have the target in view. We allow for some variance across tests in the exact positioning and orientation of the robot base due to environmental constraints and potential variance in the ending pose of any previous navigation. In particular, the base orientation is randomly chosen to be facing forward, oriented slightly to the left, or oriented slightly to the right. We represent each trajectory by ten end-effector waypoints, for which our whole-body motion planner attempts to find joint angles. We define a successful opening of an object if our system is able to execute at least 7 out of 10 waypoints. For cabinets, this corresponds to opening the cabinet over 60-degrees.

End-to-end System Test. Overall, our system achieves a 61% success rate across 31 unseen cabinets and drawers in unseen real world environments. For example deployments of our full pipeline in the testing environments, please refer to Figure 2 and the project webpage.

References

[1] Dmitry Berenson, Siddhartha Srinivasa, and James Kuffner. Task space regions: A framework for pose-constrained manipulation planning. *IJRR*, 30(12):1435–1460, 2011. 1

[2] Arjun Gupta, Max Shepherd, and Saurabh Gupta. Predicting motion plans for articulating everyday objects. In *International Conference on Robotics and Automation (ICRA)*. IEEE, 2023. 1

[3] Kaiming He, Georgia Gkioxari, Piotr Dollár, and Ross Girshick. Mask r-cnn. In *ICCV*, pages 2961–2969, 2017. 1

[4] Advait Jain and Charles C Kemp. Pulling open doors and drawers: Coordinating an omni-directional base and a compliant arm with equilibrium point control. In *2010 IEEE International Conference on Robotics and Automation*, pages 1807–1814. IEEE, 2010. 2

[5] Yiannis Karayiannidis, Christian Smith, Francisco Eli Vina Barrientos, Petter Ögren, and Danica Kragic. An adaptive control approach for opening doors and drawers under uncertainties. *IEEE Transactions on Robotics*, 32(1):161–175, 2016. 1

[6] Xiaohao Sun, Hanxiao Jiang, Manolis Savva, and Angel Xuan Chang. Opdmulti: Openable part detection for multiple objects, 2023. 1