Solving the "Last Centimeter" Problem with Autonomous Grasping Reflexes

Andrew SaLoutos, Menglong Guo, Elijah Stanger-Jones, Hongmin Kim, Sangbae Kim
Department of Mechanical Engineering, MIT
Email: [saloutos, guomeng, elijahsj, davidhmk, sangbae]@mit.edu

***Abstract*—Many modern manipulation systems rely on closing feedback loops around vision data, which reduces system bandwidth and performance speed. By developing autonomous grasping reflexes that rely on high-bandwidth force, contact, and proximity data, the overall system speed and robustness can be increased while also reducing the reliance on vision data. We are developing a new system built around a low-inertia, high-speed arm with nimble fingers that combines a high-level trajectory planner operating at less than 1Hz with low-level autonomous reflex controllers operating upwards of 200Hz. In preliminary experiments, the system is able to clear clutter consisting of 5 household objects from a cupboard in approximately 35 seconds.**

Fig. 1. **Manipulation system setup.** Left: Manipulation platform with 6-DoF arm, 6-DoF gripper, multimodal fingertip sensing, and Intel Realsense D435i camera. Right: Objects are picked from the upper shelf and placed in a dishes bin, on the lower shelf, or dropped into a trashcan.

I. Introduction

Achieving human-like versatility in robotic manipulation will depend on developing hands that are as nimble and reactive as human hands. Much work has been done on developing taxonomies and design requirements for hands [1, 4], but state-of-the-art manipulation systems have not yet been able to replicate the human hand's functionality.

Instead, many modern systems use relatively limited hardware and deploy learning algorithms that depend on large amounts of vision data to carefully plan grasps [3, 5, 8]. These algorithms can plan the entire manipulation process, from arm motion down to fingertip contacts, but the high latency introduced by the vision systems results in grasping controllers that are unable to properly react to high-bandwidth object interaction information. Even if contact and force data are used in the planning algorithms, the bandwidth of the vision system limits the execution speed and usually requires that the manipulation is quasi-static.

We propose a different approach to building a manipulation system, starting from autonomous low-level behaviors, which we call reflexes, instead of relying on a single integrated algorithm. As reflexes are constructed to reason about contact interactions, finger motions, and potentially arm motions, the scope of a higher-level planner is reduced to reasoning only about the manipulation task. The reflexes are autonomous from the planner, and are based on high-bandwidth information, such as force, contact, and proximity data, that is directly sensed. We use these reflexes to close the grasping feedback loop locally in the hand, without needing vision data or adding unnecessary planning complexity. In this abstract, we present initial results from our proposed system architecture for a clutter-clearing task.

II. Manipulation System

Our manipulation system is shown in the left of Fig. 1. It consists of a low-inertia, high-speed arm with six degrees of freedom and a planar six degree-of-freedom gripper with two three-link fingers. Each fingertip has a multimodal force/contact sensor [2] combined with four time-of-flight sensors (two facing inwards, one forwards, and one outwards). There is an additional time-of-flight sensor in the palm. The reflexes make use of the arm and finger kinematics, the contact kinematics (when in contact), and the force and proximity data. All of the sensors are sampled at 200Hz, with minimal time for processing overhead, and the motors are controlled at 500Hz.

To generate high-level manipulation plans, a Realsense D435i camera is used to capture a frame of the shelf environment. The user (emulating a planning algorithm) selects a region for grasping, and the average $[x,y,z]$ coordinates of the selected pixels are sent to the arm controller as the next desired grasping location. As the arm moves to the desired grasp location, the reflexes attempt to grasp the object.

III. Autonomous Grasping Reflexes

A flowchart detailing the reflexive grasping controller is shown in Fig. 2. Once a grasping target has been received, the arm moves towards the location. During this stage, the fingers use the proximity data to avoid collisions with external objects and detect occlusions in the forward direction. The inwards-facing proximity sensors are used to detect the desired object and allow the fingers to follow the object surface as the wrist advances and to trigger the grasp. Once a grasp attempt is triggered, the fingers close into contact with the object and the grasp is evaluated. If the grasp is successful, with good fingertip contact angles and forces and no extraneous fingertip motion, the arm trajectory is stopped and the object is moved to the desired location. During this motion, the shear force data from the fingertips is used to calculate pinch forces that will prevent the object from slipping out of the gripper. If the grasp is unsuccessful, the finger kinematics and contact angles are used to choose whether to attempt a power or antipodal pinch grasp or to drag the object out from clutter before trying to grasp again. If executing the grasp takes too long or too many attempts are required, the grasp fails and the arm returns to a home position to wait for another target location.

IV. Results and Discussion

Despite noisy vision data and inconsistent user inputs, the reflex controllers can locate and grasp the desired object. In our first results videos here and here, we send the initial object location and then purposely perturb the object. The fingers react to the displaced object, even if it is moved in real-time, and are able to complete the desired grasp.

In our second results video here, we show an example trial of clearing objects off of a shelf. The objects have different sizes, shapes, and stiffness, and they are all successfully picked from the shelf using just the approximate initial location sent by the user. Our system takes an average of 5-6 seconds per object with nominal trajectory speeds, while a lab member performing the same shelf-clearing task at a comfortable speed took about 2 seconds per object.

The current version of the gripper uses off-the-shelf finger actuators with limited current control bandwidth, which reduces their reaction speeds and thus the average time per grasp. This can be improved with custom actuators and motor drivers that are designed for high-bandwidth current control.

It is also important to note that our low-level reflexes could be easily integrated in place of the grasping controllers in most state-of-the-art learning pipelines, such as those in [5, 6, 7]. We would expect this to increase the execution speed and robustness of the high-level manipulation plans.

V. Conclusion

We have presented promising early results for our proposed autonomous reflexive grasping system. The reflexes improve closed-loop manipulation speed and robustness while also reducing planner complexity. Future work includes expanding the reflexes, improving the reliability of the gripper hardware, further testing and characterization of the overall system behavior, and possible integration with state-of-the-art manipulation planners.

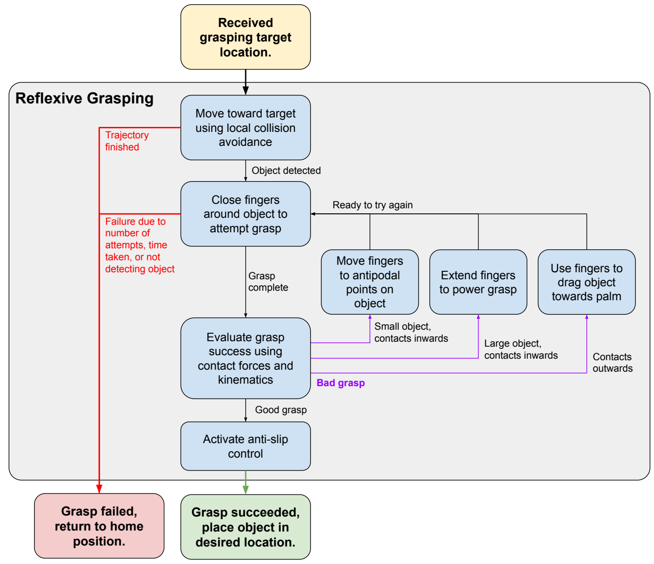

Fig. 2. **Controller for reflexive grasping.** As the arm moves the gripper to the target location, the autonomous reflexes attempt to grasp the object.

References

[1] Ian M Bullock, Raymond R Ma, and Aaron M Dollar. A hand-centric classification of human and robot dexterous manipulation. *IEEE transactions on Haptics*, 6(2):129–144, 2012.

[2] Lindsay Epstein, Andrew SaLoutos, Donghyun Kim, and Sangbae Kim. Bi-modal hemispherical sensors for dynamic locomotion and manipulation. In *2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 7381–7381. IEEE, 2020.

[3] Nima Fazeli, Miquel Oller, Jiajun Wu, Zheng Wu, Joshua B Tenenbaum, and Alberto Rodriguez. See, feel, act: Hierarchical learning for complex manipulation skills with multisensory fusion. *Science Robotics*, 4(26): eaav3123, 2019.

[4] Thomas Feix, Javier Romero, Heinz-Bodo Schmiedmayer, Aaron M Dollar, and Danica Kragic. The grasp taxonomy of human grasp types. *IEEE Transactions on human-machine systems*, 46(1):66–77, 2015.

[5] Wei Gao and Russ Tedrake. kpam 2.0: Feedback control for category-level robotic manipulation. *IEEE Robotics and Automation Letters*, 6(2):2962–2969, 2021.

[6] Douglas Morrison, Peter Corke, and Jürgen Leitner. Closing the loop for robotic grasping: A real-time, generative grasp synthesis approach. *arXiv preprint arXiv:1804.05172*, 2018.

[7] Joseph Redmon and Anelia Angelova. Real-time grasp detection using convolutional neural networks. In *2015 IEEE international conference on robotics and automation (ICRA)*, pages 1316–1322. IEEE, 2015.

[8] Andy Zeng, Shuran Song, Kuan-Ting Yu, Elliott Donlon, Francois R Hogan, Maria Bauza, Daolin Ma, Orion Taylor, Melody Liu, Eudald Romo, et al. Robotic pick-and-place of novel objects in clutter with multi-affordance grasping and cross-domain image matching. In *2018 IEEE international conference on robotics and automation (ICRA)*, pages 3750–3757. IEEE, 2018.