2022 ACM/IEEE International Conference on Human-Robot Interaction (HRI 2022), Virtual, March 2022

A Haptic Multimodal Interface with Abstract Controls for Semi-Autonomous Manipulation

Brett Stoddard, Miranda Cravetz, Timothy Player, and Heather Knight
Collaborative Robotics and Intelligent Systems (CoRIS) Institute
Oregon State University
Corvallis, OR

***Abstract*—Even as autonomous capabilities improve, many robot manipulation tasks require human(s)-in-the-loop to resolve high-level problems in uncertain environments or ambiguous situations. Prior work in highly autonomous applications tends to use interfaces with few human interface modalities, potentially missing out on the benefits that multimodal interfaces have demonstrated in lower-level operation. In this work, we demonstrate a system with a multimodal interface for a controlling a robot at a high level of autonomy. This example highlights how multiple modalities could enable redundant and robust interactions, increased situational awareness, and compact representations of complex commands, such as how to grasp an object.**

***Index Terms*—Multimodal User Interfaces, Manipulation, Teleoperation**

I. Introduction

Recent advances in robotic capabilities have allowed robots to improve human quality of life by completing real-world tasks with high levels of autonomy. Robots today are picking fruit [1], finding lost hikers [2], and collecting marine samples [3] without requiring complete remote control. In these semi-autonomous systems, human operators are still required to to ensure that each task is successful: specifying if apples are ripe, identifying regions to search, or selecting marine sample to be collected. High-level human input can improve or is necessary in uncertain environments or ambiguous tasks.

Alongside advances in robot autonomy, researchers have identified engaging and effective interfaces for controlling robotic systems. Specifically, multimodal interfaces leveraging multiple modes or channels of input and output to improve performance and reduce operator workload [4]. Typically, these rich multimodal interfaces are applied to systems with low-levels of autonomy, such as direct teleoperation, with a minority of works extending these techniques to higher levels of autonomy.

Combining these topics, we asked "**can systems with both high levels-of-autonomy and multimodal human interfaces lead to compounded gains?**" In this demonstration, we present a complete semi-autonomous system with a multi-modal peripheral, the *Ultrahaptic*, to a robotic pick-and-place task with a high level of robot autonomy. The Ultrahaptic renders tactile sensations on the operator's hand in midair as they decide which object to grasp and how to grasp it (Fig. 1). Multimodal interaction allows the operator to intuitively relay detailed information about the desired grasp while visual and haptic feedback is rendered so that if either channel is disrupted then operation of the robot is not. We hope that this system, which implements a thoughtful user experience for a common mixed-autonomy task, provides inspiration and comparison for further interest in what other synergies are achievable with controls at a high level of abstraction using a multimodal interface.

This work was supported in part by ONR award N0014-21-1-2052 and ONR/NAVSEA contract N00024-10-D-6318/DO#N0002420F8705 (Task 2: Fundamental Research in Autonomous Subsea Robotic Manipulation.

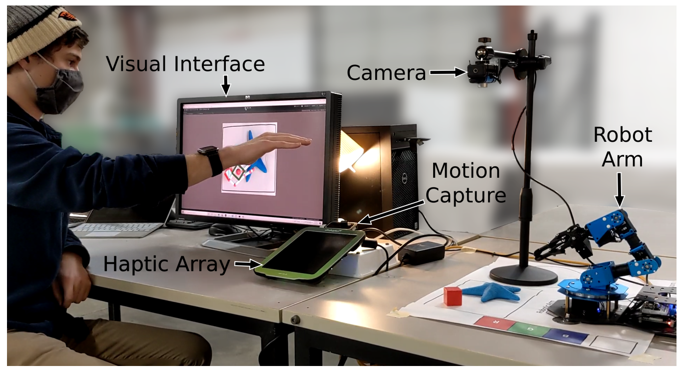

Fig. 1. Our system with visual and haptic and user I/O devices for a semi-autonomous pick and place task. With the visual and haptic interface, the user intuitively selects an object for the robot to autonomously grasp.

II. Related Work

Based on a preliminary review of robotic teleoperation methods, we found a negative correlation between a robot's level of control abstraction and the number of modes of feedback in the operator interface. The lower left half of the graph in Fig. 2 includes all interfaces we found in our preliminary literature search (a systematic review is left to future work). This figure shows an apparent gap in the literature for rich multimodal interfaces with a high level of control abstraction. In this section, we discuss how works at the edge of this frontier, and that use the Ultrahaptic peripheral, compare to our demonstration.

While multimodal input devices have been adopted to improve system performance and safety in diverse commercial applications (such as self-driving cars [5]), in research multi-modal input systems in robotics are most frequently applied to *low-level* tasks requiring direct control of robot systems.

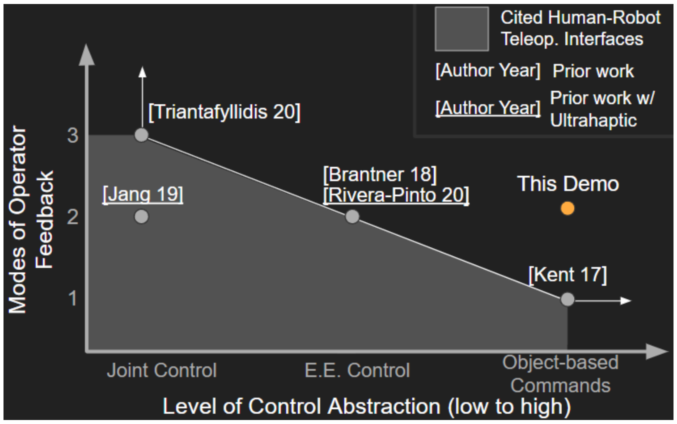

Fig. 2. Chart showing interface feedback modality vs. control abstraction for manipulating robots showing our relation to prior work. The level of abstraction ranges from joint control to objective tasks (i.e., "grab that object"). Works at the Pareto front, and using an Ultrahaptic are cited.

Likewise, work which highlights the benefits of higher levels of autonomy is often limited to a single modality, such as in a study by Kent et. al. [6].

Multimodal interfaces have been utilized to improve aspects of teleoperation such as situational awareness and performance. For example, the Ocean One robot utilized a hierarchical controller wherein an operator controlled the robot's end effector positions with the robot autonomously compensating to optimize its 14 degrees of freedom [3]. This was effectively tested during multiple voyages to recover delicate artifacts and marks an exciting new frontier in underwater robotic manipulation and telerobotics [3]. Another study found that the inclusion of audio and haptic cues improved performance on a low-level telemanipulation task [7]. Additionally, multimodal interfaces improved factors of human-swarm interaction for an observation task [8]. We anticipate that these benefits extend to manipulating robots with higher levels of autonomy.

Prior work with the Ultrahaptic showed it improving performance. These leveraged two different levels of autonomy to control a robotic hand: controlling joint states [9], and end effector's (EE) position [10]. In this work, we show how the Ultrahaptic could be used effectively leveraged at a higher level of autonomy.

III. Interface Features

Our proposed system consists of three elements (Fig. 1): a perception system that detects graspable objects in the robot's workspace, a multimodal human interface, and a motion planner that uses the human input to plan and execute actions.

The **perception system** consists of a computer vision program tuned to detect colored cubes and starfish (shown in Fig. 1). Frame by frame, a color threshold finds candidate "blobs" which are classified using Hu moment matching [11]. Detected object data is periodically sent to the interface.

The **human interface** consists of visual and haptic elements. Visually, positions of objects identified by the perception system are overlaid on a live feed of the camera. With Ultrahaptic's motion capture camera, the operator can move a digital twin of their hand in the scene. When this digital hand gets near an object, a tickling cutaneous shape is rendered based on the object's class (square for cube, circle for starfish) by the Ultrahaptic's ultrasonic array: an additional haptic affordance. With a gripping gesture, the operator initiates a grasp with parameters of the grasp, such as grasp location and angle, sent via HTTP to the motion planner.

The **motion planner** executes grasps with open-loop control according to what object was selected by the interface and the angle at which to grasp. Using inverse kinematics, the gripper first moves to the location of an object and then the gripper moves downward and the jaws close at the specified angle.

IV. Conclusion

Our demonstration shows a single implementation of this multimodal input in a highly autonomous system, but we believe this form of human-in-the-loop interaction can be extended to a variety of application domains where autonomous systems are increasingly viable but human input is still advantageous. Specifically, we believe that multimodal interaction can benefit tree pruning, which requires user preference on where to cut, or underwater biological sampling collection, where the diversity of flora and fauna creates ambiguous decisions about what to collect. Rich multimodal human interfaces do not need to be reserved for tasks where detailed, low-level input is needed. By incorporating these interfaces into semi-autonomous systems, we can improve both the human experience and, ideally, the safety and efficacy of robots.

References

[1] F. A. e. a. Cheein, "Human-robot interaction in precision agriculture: Sharing the workspace with service units," in *Proc. IEEE Int. Conf. Ind. Technol.* IEEE, 2015, pp. 289–295.

[2] D. C. Schedl, I. Kurmi, and O. Bimber, "An autonomous drone for search and rescue in forests using airborne optical sectioning," *Science Robotics*, vol. 6, no. 55, 2021.

[3] G. Brantner and O. Khatib, "Controlling ocean one: Human–robot collaboration for deep-sea manipulation," *Journal of Field Robotics*, vol. 38, no. 1, pp. 28–51, 2021.

[4] J. Y. Chen, E. C. Haas, and M. J. Barnes, "Human performance issues and user interface design for teleoperated robots," *IEEE Trans. Syst., Man, Cybern. Syst.*, vol. 37, no. 6, pp. 1231–1245, 2007.

[5] L.-M. Lundqvist and L. Eriksson, "Age, cognitive load, and multimodal effects on driver response to directional warning," *Applied ergonomics*, vol. 76, pp. 147–154, 2019.

[6] D. Kent, C. Saldanha, and S. Chernova, "Leveraging depth data in remote robot teleoperation interfaces for general object manipulation," *The Int. Journal of Robotics Research*, vol. 39, no. 1, pp. 39–53, 2020.

[7] E. Triantafyllidis, C. Mcgreavy, J. Gu, and Z. Li, "Study of multimodal interfaces and the improvements on teleoperation," *IEEE Access*, vol. 8, pp. 78 213–78 227, 2020.

[8] E. Haas, M. Fields, S. Hill, and C. Stachowiak, "Extreme scalability: Designing interfaces and algorithms for soldier-robotic swarm interaction," ARMY RESEARCH LAB ABERDEEN PROVING GROUND MD, Tech. Rep., 2009.

[9] I. Jang, J. Carrasco, A. Weightman, and B. Lennox, "Intuitive barehand teleoperation of a robotic manipulator using virtual reality and leap motion," in *Annual Conference Towards Autonomous Robotic Systems*. Springer, 2019, pp. 283–294.

[10] A. Rivera-Pinto and J. Kildal, "Visuo-tactile mixed reality for offline cobot programming," in *HRI Companion*, 2020, pp. 403–405.

[11] M.-K. Hu, "Visual pattern recognition by moment invariants," *IRE transactions on information theory*, vol. 8, no. 2, pp. 179–187, 1962.