Designing Robotic Camera Systems to Enable Synchronous Remote Collaboration

Pragathi Praveena
Department of Computer Sciences,
University of Wisconsin–Madison,
Madison, WI, USA
pragathi@cs.wisc.edu

Michael Gleicher
Department of Computer Sciences,
University of Wisconsin–Madison,
Madison, WI, USA
gleicher@cs.wisc.edu

Bilge Mutlu
Department of Computer Sciences,
University of Wisconsin–Madison,
Madison, WI, USA
bilge@cs.wisc.edu

ABSTRACT

Collaborative robots have the potential to be *intelligent, embodied agents* that can contribute to remote human collaboration. We explore this paradigm through the design of robot-mounted camera systems for remote assistance. In this extended abstract, we discuss our iterative design process to develop interaction techniques that leverage shared control-based methods to distribute camera control between the agentic robot and human collaborators.

CCS CONCEPTS

• **Human-centered computing → Computer supported cooperative work**; • **Computer systems organization → Robotics**.

KEYWORDS

telepresence, remote collaboration, shared control, mixed-initiative

ACM Reference Format:
Pragathi Praveena, Michael Gleicher, and Bilge Mutlu. 2023. Designing Robotic Camera Systems to Enable Synchronous Remote Collaboration. In *Companion of the 2023 ACM/IEEE International Conference on Human-Robot Interaction (HRI '23 Companion), March 13–16, 2023, Stockholm, Sweden.* ACM, New York, NY, USA, 3 pages. https://doi.org/10.1145/3568294.3579974

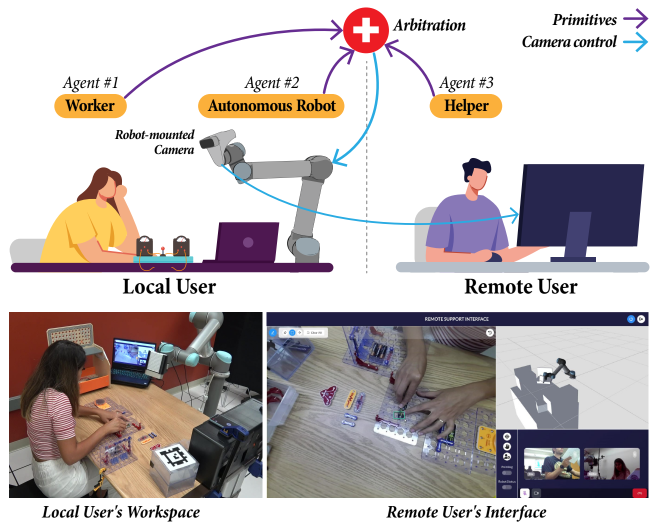

Figure 1: We design and evaluate robotic camera systems that help two people collaborate remotely on physical tasks. In our past work, called *Periscope*, a local worker completes an assembly task with guidance from a remote helper who can view the workspace through a robot-mounted camera.

1 INTRODUCTION

Technologies that support remote collaboration have advanced significantly in recent decades, and they were indispensable during the COVID-19 pandemic: tools such as Zoom for video conferencing and TeamViewer for remote desktop access supported collaborative work, especially in the information economy. However, remote support for jobs that involve physical work, such as nursing or factory work, was rare. This highlights opportunities for the use of robots that can leverage their physical embodiment to extend people's abilities in remote spaces. Advancements in robotics have enabled collaborative robots that are safe, easy to use, and capable, opening up a promising space for designing systems that support remote human collaboration. In this work, we focus on the design of robotic camera systems. We contextualize our work in scenarios where a *local user* completing manual tasks is co-located with a robotic arm and collaborates with a *remote user* who views the local workspace through a camera mounted on the robot.

We draw from the extensive body of work on telepresence robots [18], but our work leverages a robot form that is kinematically more capable than the prototypical tablet-on-wheels robot [9] or tabletop robot [12, 22] in existing work. This improved capability allows the remote user to have diverse and detailed views of the workspace but also increases the complexity of camera control [3]. We posit that users may focus better on the collaborative task at hand if they can offload part of the camera control to an autonomous robot [12, 21]. Thus, our research approach draws from a robot control paradigm, called *shared control* [13], to distribute the control of the robot-mounted camera between the local user, the remote user, and the autonomous robot depending on task needs. Thus, the key research questions that guide our work are:

RQ.1 *How can we design interaction techniques for camera control that effectively combine human and artificial intelligence to support remote collaboration?*
RQ.2 *How does a collaborative robot influence human-human interaction in a remote collaboration setting?*

This work was supported by an NSF award 1830242 and a NASA University Leadership Initiative (ULI) grant awarded to the UW-Madison and The Boeing Company (Cooperative Agreement #80NSSC19M0124). We thank Bengisu Cagiltay, Carter Sifferman, Emmanuel Senft, & Mike Hagenow for their input. Fig. 1 used art by jcomp on Freepik.

HRI '23 Companion, March 13–16, 2023, Stockholm, Sweden

ACM ISBN 978-1-4503-9970-8/23/03.
https://doi.org/10.1145/3568294.3579974

In past work [17], we designed and implemented the *Periscope* system (see Figure 1, §2), which serves as an exploration of RQ.1 and a platform for further investigation into both research questions.

2 THE PERISCOPE SYSTEM

Periscope is a robotic system for remote assistance that is used by a local worker to complete assembly tasks with guidance from a remote helper who observes the workspace through the robot-mounted camera. A screen interface (see Figure 1) for each collaborator shows the camera view and a simulated 3D view of the robot and its surroundings. The interface is also equipped with screen annotation tools to support referential communication, allows video conferencing, and enables control of the robot. The camera view provides a dynamic *shared visual context* [6], which is task-relevant visual information that the worker and the helper have in common and is critical for them to maintain task awareness [5] and develop a mutual understanding for effective communication [4].

In *Periscope*, both collaborators can adjust the shared view for their task needs (e.g., asking questions or providing guidance). The robot can also adjust the view autonomously to assume part of the workload of camera control, such as tracking the worker's hand and maintaining it in the camera view. A key contribution of our system is a formulation for *shared camera control* that consists of a set of *primitives* and *modes* that enables camera control to be shared between the three agents so that both the worker and the helper can shape the view with assistance from the autonomous robot. A core aspect of shared control systems is the design of *arbitration*, which is the division of control among agents when completing a task. Arbitration in shared control should allow all agents to contribute and change the type of contribution they make over time [13]. In this work, we designed primitives and modes (based on our prior experience designing robotic systems [15, 16]) to uniquely arbitrate the control of the camera between the three agents.

Primitives are basic elements of our camera control formulation that give an agent the ability to modify the shared view. For example, *Set target* is a primitive that sets the viewing direction of the camera. Primitives can be triggered by the helper and the worker through user input (such as a mouse click or hand gesture) and by the robot through sensory input (such as depth data from a camera). Primitives serve as building blocks that can be organized in various combinations. We posit that, depending on the varying needs over the course of a task, each agent may require a different amount of control of the camera view [11, 14]. Thus, we organize the primitives into three modes that users can toggle between, each of which is primarily driven by one of the three agents but also allows the other agents to exert some influence. The inputs (or primitives) from the three agents are arbitrated in real-time to generate robot motion and acquire the co-constructed view. The arbitration is done at a *discrete* level using the three modes and at a *continuous* level using an optimization-based method similar to prior work [19, 20].

3 CURRENT AND FUTURE WORK

Our past work provides insight into RQ.1 through the design of primitives and modes (for arbitration). Our current work includes the analysis of data collected during our usability testing of *Periscope* with 12 dyads to understand human interactions in robot-mediated collaboration (RQ.2). We will continue to investigate our research questions through an iterative design process where we improve the *arbitration* mechanisms and the *specification* of primitives within *Periscope* and assess how they impact collaboration.

Evaluation (Ongoing)—We are conducting a thematic analysis [2] of video recordings from our usability testing of *Periscope* to qualitatively understand robot-supported, dyadic collaboration (RQ.2). Our early findings include: (1) helpers utilizing the robot's kinematic capabilities to frequently move the camera and obtain diverse and detailed views; and (2) workers using the robot's physical embodiment for situation awareness, such as recognizing what the helper is looking at in the workspace or if the robot is tracking their hand. Additionally, the robot may serve different roles during the collaboration process as a tool [8] (e.g., a camera holder), a surrogate [10] (e.g., representing the helper's gaze), or a collaborator [21] (e.g., tracking an object of interest autonomously during periods of helper inattention). We plan to use insights from the analysis of our multi-user, multi-modal data to derive hypotheses and design decisions for our future systems and establish quantitative metrics to determine the quality of collaboration supported by our systems.

Arbitration (Planned)—Our usability testing of *Periscope* revealed issues with the representation of discrete arbitration as conditional statements. The logic of UI flow is not explicit in this representation, making it challenging to validate and debug system behavior. As a result, there were some unforeseen errors, such as becoming caught in an infinite loop (a piece of code that does not terminate and repeats indefinitely). Thus, we plan to improve the representation in future work using *deterministic finite-state Mealy automata*.

A finite-state machine (FSM) representation offers a scalable computational model to document and control execution flow within our system. A Mealy machine's output depends on the present state as well as the present input (from the three agents), which lends itself well to our shared control approach. A deterministic FSM allows the interface to behave in predictable and legible ways to the user. Modeling arbitration using such automata allows precise and comprehensive specification of complex dependencies that may exist between various inputs (from multiple users) and states of the system. However, there is a risk that the FSM may become too complicated to function according to the expectations of the user [1]. To mitigate this, we plan to use methods from automata learning to simplify the FSM based on task context. This approach may additionally help us to identify alternate, non-modal ways of organizing primitives for arbitration.

Specification (Planned)—In *Periscope*, the helper used through-the-lens camera control [7], where the camera view is specified through controls in the image plane, such as clicking and dragging left/right/up/down to orbit the camera around a 3D point in the workspace. The worker could physically move the co-located robot or use hand gestures to direct the camera. In future work, we want to improve how users can specify the desired view. For example, we envision enhancing the capabilities of the simulated 3D view for object-centric input such as specifying the camera to view the *right side* of an object or stay *normal* to a plane. Additionally, we plan to conduct co-design sessions with experts familiar with film production and cinematography to generate better view specifications and understand opportunities for other applications of robotic cameras, such as remote filming (e.g., filming how-to videos of manual tasks).

REFERENCES

[1] F. Belli. 2001. Finite State Testing and Analysis of Graphical User Interfaces. In *Proceedings 12th International Symposium on Software Reliability Engineering*. 34–43. https://doi.org/10.1109/ISSRE.2001.989456

[2] Virginia Braun and Victoria Clarke. 2006. Using Thematic Analysis in Psychology. *Qualitative Research in Psychology* 3, 2 (2006), 77–101. https://doi.org/10.1191/1478088706qp063oa

[3] Marc Christie, Patrick Olivier, and Jean-Marie Normand. 2008. Camera Control in Computer Graphics. *Computer Graphics Forum* 27, 8 (2008), 2197–2218. https://doi.org/10.1111/j.1467-8659.2008.01181.x

[4] Herbert H. Clark and Catherine R. Marshall. 1981. Definite Knowledge and Mutual Knowledge. In *Elements of Discourse Understanding*, Aravind K. Joshi, Bonnie L. Webber, and Ivan A. Sag (Eds.). Cambridge, UK: Cambridge University Press, 10–63.

[5] Mica R. Endsley. 1995. Measurement of Situation Awareness in Dynamic Systems. *Human Factors* 37, 1 (1995), 65–84. https://doi.org/10.1518/001872095779049499

[6] Susan R. Fussell, Robert E. Kraut, and Jane Siegel. 2000. Coordination of Communication: Effects of Shared Visual Context on Collaborative Work. In *Proceedings of the 2000 ACM Conference on Computer Supported Cooperative Work* (Philadelphia, Pennsylvania, USA) *(CSCW '00)*. Association for Computing Machinery, New York, NY, USA, 21–30. https://doi.org/10.1145/358916.358947

[7] Michael Gleicher and Andrew Witkin. 1992. Through-the-Lens Camera Control. In *Proceedings of the 19th Annual Conference on Computer Graphics and Interactive Techniques (SIGGRAPH '92)*. Association for Computing Machinery, New York, NY, USA, 331–340. https://doi.org/10.1145/133994.134088

[8] Pavel Gurevich, Joel Lanir, Benjamin Cohen, and Ran Stone. 2012. TeleAdvisor: A Versatile Augmented Reality Tool for Remote Assistance. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Austin, Texas, USA) *(CHI '12)*. Association for Computing Machinery, New York, NY, USA, 619–622. https://doi.org/10.1145/2207676.2207763

[9] Steven Johnson, Irene Rae, Bilge Mutlu, and Leila Takayama. 2015. Can You See Me Now? How Field of View Affects Collaboration in Robotic Telepresence. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems* (Seoul, Republic of Korea) *(CHI '15)*. Association for Computing Machinery, New York, NY, USA, 2397–2406. https://doi.org/10.1145/2702123.2702526

[10] Hideaki Kuzuoka, Gen Ishimoda, Yushi Nishimura, Ryutaro Suzuki, and Kimio Kondo. 1995. Can the GestureCam be a Surrogate?. In *Proceedings of the Fourth European Conference on Computer-Supported Cooperative Work ECSCW'95*, Hans Marmolin, Yngve Sundblad, and Kjeld Schmidt (Eds.). Springer, Dordrecht, 181–196. https://doi.org/10.1007/978-94-011-0349-7_12

[11] Joel Lanir, Ran Stone, Benjamin Cohen, and Pavel Gurevich. 2013. Ownership and Control of Point of View in Remote Assistance. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Paris, France) *(CHI '13)*. Association for Computing Machinery, New York, NY, USA, 2243–2252. https://doi.org/10.1145/2470654.2481309

[12] Jiannan Li, Maurício Sousa, Chu Li, Jessie Liu, Yan Chen, Ravin Balakrishnan, and Tovi Grossman. 2022. ASTEROIDS: Exploring Swarms of Mini-Telepresence Robots for Physical Skill Demonstration. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems* (New Orleans, LA, USA) *(CHI '22)*. Association for Computing Machinery, New York, NY, USA, Article 111, 14 pages. https://doi.org/10.1145/3491102.3501927

[13] Dylan P. Losey, Craig G. McDonald, Edoardo Battaglia, and Marcia K. O'Malley. 2018. A Review of Intent Detection, Arbitration, and Communication Aspects of Shared Control for Physical Human–Robot Interaction. *Applied Mechanics Reviews* 70, 1 (02 2018). https://doi.org/10.1115/1.4039145

[14] Helena M. Mentis, Yuanyuan Feng, Azin Semsar, and Todd A. Ponsky. 2020. Remotely Shaping the View in Surgical Telementoring. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) *(CHI '20)*. Association for Computing Machinery, New York, NY, USA, 1–14. https://doi.org/10.1145/3313831.3376622

[15] Pragathi Praveena, Luis Molina, Yeping Wang, Emmanuel Senft, Bilge Mutlu, and Michael Gleicher. 2022. Understanding Control Frames in Multi-Camera Robot Telemanipulation. In *Proceedings of the 2022 ACM/IEEE International Conference on Human-Robot Interaction* (Sapporo, Hokkaido, Japan) *(HRI '22)*. IEEE Press, 432–440. https://doi.org/10.5555/3523760.3523818

[16] Pragathi Praveena, Daniel Rakita, Bilge Mutlu, and Michael Gleicher. 2020. Supporting Perception of Weight through Motion-Induced Sensory Conflicts in Robot Teleoperation. In *Proceedings of the 2020 ACM/IEEE International Conference on Human-Robot Interaction* (Cambridge, United Kingdom) *(HRI '20)*. Association for Computing Machinery, New York, NY, USA, 509–517. https://doi.org/10.1145/3319502.3374841

[17] Pragathi Praveena, Yeping Wang, Emmanuel Senft, Michael Gleicher, and Bilge Mutlu. 2023. Periscope: A Robotic Camera System to Support Remote Physical Collaboration. In *(Under Review)*.

[18] Irene Rae, Gina Venolia, John C. Tang, and David Molnar. 2015. A Framework for Understanding and Designing Telepresence. In *Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing* (Vancouver, BC, Canada) *(CSCW '15)*. Association for Computing Machinery, New York, NY, USA, 1552–1566. https://doi.org/10.1145/2675133.2675141

[19] Daniel Rakita, Bilge Mutlu, and Michael Gleicher. 2019. Remote Telemanipulation with Adapting Viewpoints in Visually Complex Environments. In *Proceedings of Robotics: Science and Systems XV*. Freiburg im Breisgau, Germany. https://doi.org/10.15607/RSS.2019.XV.068

[20] Emmanuel Senft, Michael Hagenow, Pragathi Praveena, Robert Radwin, Michael Zinn, Michael Gleicher, and Bilge Mutlu. 2022. A Method For Automated Drone Viewpoints to Support Remote Robot Manipulation. In *2022 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. 7704–7711. https://doi.org/10.1109/IROS47612.2022.9982063

[21] Janis Stolzenwald and Walterio W. Mayol-Cuevas. 2019. Reach Out and Help: Assisted Remote Collaboration through a Handheld Robot. *CoRR* abs/1910.02015 (2019). arXiv:1910.02015 http://arxiv.org/abs/1910.02015

[22] Ana M Villanueva, Ziyi Liu, Zhengzhe Zhu, Xin Du, Joey Huang, Kylie A Peppler, and Karthik Ramani. 2021. RobotAR: An Augmented Reality Compatible Teleconsulting Robotics Toolkit for Augmented Makerspace Experiences. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems* (Yokohama, Japan) *(CHI '21)*. Association for Computing Machinery, New York, NY, USA, Article 477, 13 pages. https://doi.org/10.1145/3411764.3445726