Perception-Motion Coupling in Active Telepresence: Human Behavior and Teleoperation Interface Design

TSUNG-CHI LIN, ACHYUTHAN UNNI KRISHNAN, and ZHI LI, Worcester Polytechnic Institute, Robotics Engineering

Teleoperation enables complex robot platforms to perform tasks beyond the scope of the current state-of-the-art robot autonomy by imparting human intelligence and critical thinking to these operations. For seamless control of robot platforms, it is essential to facilitate optimal situational awareness of the workspace for the operator through active telepresence cameras. However, the control of these active telepresence cameras adds an additional degree of complexity to the task of teleoperation. In this paper we present our results from the user study that investigates: (1) how the teleoperator learns or adapts to performing the tasks via active cameras modeled after camera placements on the TRINA humanoid robot; (2) the perception-action coupling operators implement to control active telepresence cameras, and (3) the camera preferences for performing the tasks. These findings from the human motion analysis and post-study survey will help us determine desired design features for robot teleoperation interfaces and assistive autonomy.

CCS Concepts: • Human-centered computing → Interaction design theory, concepts and paradigms; *Empirical studies in interaction design*; Empirical studies in visualization;

Additional Key Words and Phrases: Perception-action coupling, active telepresence, robot teleoperation

ACM Reference format:
Tsung-Chi Lin, Achyuthan Unni Krishnan, and Zhi Li. 2023. Perception-Motion Coupling in Active Telepresence: Human Behavior and Teleoperation Interface Design. *ACM Trans. Hum.-Robot Interact.* 12, 3, Article 31 (March 2023), 24 pages.
https://doi.org/10.1145/3571599

1 INTRODUCTION

Contemporary tele-robotic systems (e.g., for nursing assistance [62], surgery [95], manufacturing [75], etc.) are usually equipped with multiple active telepresence cameras to provide the teleoperator suficient perception of the remote environment and the tasks. Deciding how to select and control them to acquire the desirable camera motion and viewpoint could be as dificult as controlling the freeform dexterous tele-manipulation, given that the remote cameras may be located at the robots head, the manipulators for task operation or camera assistance, the mobile base, or standalone in the workspace and can be moved as required (see Figure 1 [62] for example). When focusing on the tele-manipulation tasks, teleoperators often neglect effective control of the active telepresence cameras to avoid the additional cognitive workload. Although robot autonomy

Authors' address: T.-C. Lin, A. Unni Krishnan, and Z. Li, Worcester Polytechnic Institute, Robotics Engineering, Unity Hall 200A, 27 Boynton St, Worcester, Massachusetts, 01609; emails: {tlin2, aunnikrishnan, zli11}@wpi.edu.

2573-9522/2023/03-ART31
https://doi.org/10.1145/3571599

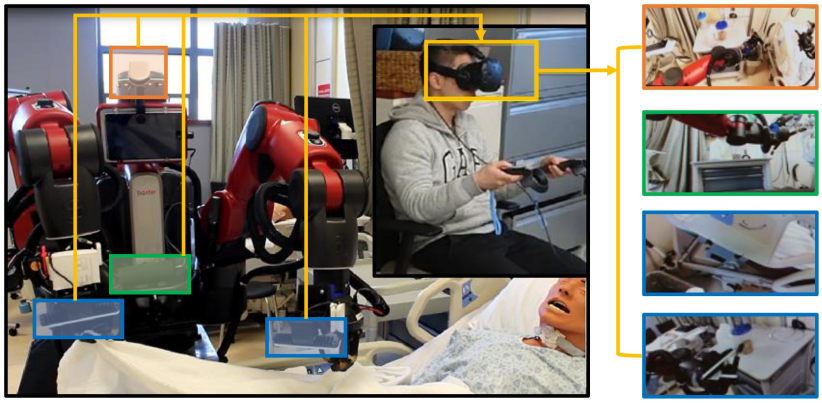

Fig. 1. Nursing robot teleoperation via a freeform interface with feedback from multiple active telepresence cameras attached to head, torso, and wrists.

for camera assistance is necessary, ill-designed camera assistance, which *do not account for the natural preference of human visual perception and visual comfort*, may confuse and frustrate the teleoperators, and reduce their performance and trust in robot autonomy.

The remote control of active telepresence cameras is dificult because the robot teleoperators need to develop novel motor skills to control the unfamiliar viewpoint of the robots, which are different from human eyes and their viewpoint in their displacements, motion capabilities, depth perception, and field of view (FOV) [24]. Controlling this foreign viewpoint of the robot is counter-intuitive to humans who are used to the location, perception capabilities and natural viewpoint control motions of human eyes. To assist the teleoperators to utilize the active telepresence cameras better, prior research efforts have developed (1) interfaces (e.g., via head/gaze tracking [5, 26]) for intuitive camera viewpoint and motion control, and (2) robot autonomy for autonomous dynamic viewpoint selection and camera motion control [79]. However, the design of interfaces and autonomy for camera assistance is mostly hand-engineered and based on empirical experience, rather than *the in-depth understanding of human natural behavior and preference of perception-action coupling, which has a strong influence on how humans prefer to coordinate the remote camera control and robot motions and actions*. They are also mostly designed for single camera robotic systems and are not capable of handling active telepresence via multiple cameras.

Overview of Research Efforts. This paper aims to transform the design philosophy for tele-robotic interfaces, based on a deep understanding of perception-action coupling of cyber-human systems. Among the many aspects of motion control, the coordination between perception and action is most critical to tele-nursing task performance. Knowledge about perception-action coupling has been leveraged in human-robot interaction to a limited extent, and has already yielded effective models and approaches for predicting human intent [10], optimizing camera motions and viewpoints [80], interactive perception [15], and sensory augmentation of human-robot interfaces for motor skill training and rehabilitation [50]. While remote robots limit human perception and motion capabilities, they also provide opportunities for the human motor system to explore. Novel perception-action coupling skills do not exist in the repertoire of human motor control, yet are critical for robot teleoperation. Through the robot teleoperation interface, the human and the robot are closely coupled as an integrated cyber-human system, and novel perception-action coordination needs to be developed to adapt this system's new perception and action capabilities. To facilitate this adaptation, both robot teleoperation interface and assistive autonomy need to be designed

based on the human behavior and preference of perception-action coupling, which has been studied extensively in human movement science [44, 93], but not at all for cyber-human systems. The research efforts in this paper aim to bridge this gap, by proposing a novel experimental paradigm that can simulate human natural behavior and preference in the usage of active telepresence. We further conduct comprehensive user studies with this experimental paradigm to (1) discover the *perception-action coupling* of a coupled human-robot system, and (2) reveal its implications to the *design of robot teleoperation interface and assistive autonomy.*

Novel Experimental Paradigm. The novel experimental paradigm we proposed was designed to study the perception-action coordination, human adaptation, and preference in the usage of active telepresence cameras. To eliminate the effort of controlling the robot, the experimental paradigm provided a simulated telepresence setting with video streams from the cameras attached to the user's head, torso, dominant and non-dominant hands as well as a standalone workspace camera while retaining the humans' ability to perform object manipulation. These video streams were used by the participants to stack lightweight plastic cups into a pyramid.

User Studies and Findings. The proposed experimental paradigm enables us to study the perception and action coupling in terms of vision-motion coupling, haptic-motion coupling, and vision-haptic coupling of sensory integration. The findings from our user study identify suitable designs for active perception camera control, the shared autonomy for camera selection, and an intuitive assisted teleoperation interface. In this paper, a novel experimental paradigm is proposed and the findings from the human motion observation from our prior works [64, 101] were extended with analysis to compare the identified motion features between the training and performing phase to reveal how a human would adapt to the control of the active telepresence. The systematic perception-action coupling, human adaptation, and preference investigation help identify the desired design of an active telepresence camera in a teleoperation interface.

The rest of the paper is organized as follows. Section 2 discusses the active perception of robot teleoperation and insights for multi-sensory integration. Section 3 describes the experimental paradigm and data analysis. Section 4 details our findings with objective and subjective data support. In Section 5, we presented the discussion of the results and future directions. Finally, Section 6 summarizes the important findings of this paper.

2 RELATED WORK

2.1 Multi-camera Telepresence for Tele-robotic Systems: Design and Limitations

The usage of multi-camera telepresence has enabled tele-robotic systems to operate in complex environments and perform tasks that require high dexterity and mobility while under the control, guidance, or supervision of remote human users. Many contemporary tele-robot systems integrate multiple cameras to increase the field of view, or to provide an additional viewpoint of robot, tasks, and environments [55]. For example, Nguyen et al. recently integrated an array of four cameras to provide a wider field of view, such that remote users could assist with wheelchair navigation [71]. Compared to panoramic cameras, the integration of multiple telepresence cameras can provide a suficiently wide camera view at lower cost and energy consumption. On the other hand, Whitney et al. proposed to display the 2D video from hand cameras of a humanoid robot along with the point cloud from its head camera. Teleoperators use both the global and local task views to efi-ciently control the robots to perform dexterous manipulation tasks such as laundry folding [105]. An interactive detail-in-context telepresence interface displays the pan-and-tilt view from a nar-row camera (in robot head) inside of a wider pannable view (attached to a pole extended from the robots back), such that the teleoperators can zoom in on details of a selected region [28, 91].

Indeed, a multi-camera telepresence system can integrate the displays from the cameras of different robots. For instance, De León et al. proposed a design of multi-camera telepresence to increase the navigation capabilities of multi-robot systems in disaster response [28]. The robot primarily responsible for the mission is provided with the external viewpoints from the cameras of the other robot teammates, in addition to the onboard camera it carries. The feed from multiple cameras on the robot can be provided simultaneously or be relayed as active camera feedback where the different viewpoints can be individually controlled. As presented by Seo et al. [91], the ability to alternate between multiple camera views simultaneously relayed to them allows the operator to get more information about the workspace and corroborate information about the workspace based on information from the various views. However, displaying multiple camera views at the same time can cause information overload overwhelming the operator and thus affecting their ability to perform the task [12]. Additionally, to fully utilize the potential of simultaneous multi-camera feeds requires the ability to spatially correlate the events between different viewpoints. However, this ability is dependent on the spatial reasoning skills of participants which are highly user-specific and as a result can result in increased cognitive workload for operators with limited spatial reasoning skills [9, 48]. With an active multi-camera telepresence network there is improved remote perception capabilities of tele-robotic systems, and improved situational awareness among the robot teleoperator or supervisors. However, tracking, managing, and controlling the feed from multiple cameras also demands additional cognitive and operational efforts. In general, related work in literature addresses this limitation by the design of: (1) *control interfaces* that use head motions and/or gaze for intuitive camera control [26, 76, 86], and (2) *robot autonomy for camera assistance* that autonomously adjust the camera viewpoint to track the object of interest [72, 78, 110], the robot end-effectors or tools [79, 80], or the features critical to task performance [19, 74, 88]. The autonomy for camera assistance can also be used to optimize the camera motions for visual comfort [20, 22, 87], or optimize camera viewpoint for information gain [51, 60], aesthetics [37, 47, 59], viewpoint familiarity [96], or other considerations. Nevertheless, these control interfaces and camera assistance are limited because: (1) they were mostly designed for single-camera systems, and (2) the strategy for camera viewpoint and motion control was proposed and evaluated case-by-case, based on empirical experience and hand-engineered criteria, instead of systematic understanding of human behavior and preference for the selection and control of active telepresence cameras.

2.2 Vision-Motion Coupling and Robot Teleoperation

If a human is subjected to a foreign viewpoint, like a viewpoint from their torso or hands, with limited haptic sensation from touch, then the human would have to adapt novel ways to use this foreign vision and haptic sensation to interact with the environment. Fortunately, we are confident that the human motor system is able to re-develop a "new normal" to best utilize the new perception and action capabilities, as seen in motor skill training [3] and rehabilitation [66, 92].

The temporal and spatial coordination of vision and movements, namely the visuomotor coordination, is essential to human motor control. The human behavior and underlying human motor control strategies of the vision-motion coupling [44] have been extensively investigated in various human motor skills. Specifically, many human factor experiments have studied the gaze pattern, visual control, or eye-hand/eye-foot/eye-head coordination in the tasks including active perception (e.g., visual search [61], target selection [27], target tracking [25, 82], scene viewing [100]), manipulation (e.g., reaching [4, 31], reaching-to-grasp [56, 98], grasping [8], interception [70, 112], bimanual coordination [94], object manipulation [57]), and locomotion (e.g., walking [13, 34], navigation [36, 41], driving [67]), tool and interface operation (e.g., laparoscopic surgery [49], video game [40]), and learning of motor skills (e.g., [14]). Such experimental studies reveal that human gaze and visual control in daily activities can be influenced not only by the salient features [97]

and surprising stimuli [52] in the task environment, but also by the action and behavior goals [44] (and their associated intrinsic [53, 68] and explicit [69, 89] rewards), the benefits of collecting additional information to reduce the uncertainty in task environments [38, 99], the memory of task-relevant objects or context cues in the environment [44], and the predicted visual state in action control [32, 33, 43]. In more recent literature, frameworks such as probabilistic decision theory [44, 107], stochastic optimal control [54, 65] have been used to explain the vision-motion coupling of human motor control, while computational models are also developed to explain, predict, and render human (-like) gaze/visual attention/active perception behavior (e.g., [16, 45]).

The natural behavior and preference of vision-motion coupling not only influence how humans perform various motor skills in daily activities (e.g., [35, 46]), but also influence how humans use robot teleoperation interfaces. In robot teleoperation, whether teleoperators can make motion control decisions depends on how well they can perceive, comprehend, and predict the remote task being operated [11], which further relies upon how well they can select and control the remote cameras in coordination with their tele-actions [111]. In the usage of a teleoperation interface, teleoperators will have less cognitive workload and better situational awareness, if the telepresence interface allows them to control the remote cameras similar to their natural gaze control, and if the robot autonomy for camera assistance can provide camera viewpoints and motions can accommodate their needs for performing tele-action and visual comfort [29, 108]. Such interface and autonomy are also important to the learning of robot teleoperation interfaces because it facilitates the development of spatial skills, including spatial visualization (perceiving objects among cluttered environments), mental rotation (rotation and visualization of an object to form different configurations), and perceptive taking (visualizing objects in different frames of reference) [23, 104].

2.3 Multi-sensory Integration

Another important human factor we need to investigate is the multi-sensory integration in the usage and learning of robot teleoperation interfaces. Similar to vision-motion coupling, the integration of visual and haptic feedback [30, 39] are also natural and essential to human motor control. The effects of haptic perception and visuo-haptic sensory integration have also been investigated in various human motor behavior and motor learning processes (e.g., [18, 84]). For example, prior research has shown that haptic perception can disambiguate visual perception of 3D shape [106] and facilitates the identification of objects [30, 58]. The integration of visual and haptic feedback also facilitates the learning of tool usage [90] and laparoscopic surgery skills [42]. In many multi-sensory tasks (e.g., grasping small objects), visual and haptic inputs are weighted based on the reliability of individual cues [39]. The framework of Maximum Likelihood Estimation (MLE) has been used to explain the weighted integration of multi-sensory cues (e.g., visual and haptic cues), in natural and synthetic environments [39, 103]. The haptic feedback provided by robot teleoperation interfaces, although limited in its accuracy, transparency, and sensitivity, can still be leveraged to compensate for lost information in the visual feedback via remote cameras.

2.4 Findings in Preliminary Work

The research on perception-action coupling, from experimental human movement studies to theoretical frameworks, to computational models, has not been extended to human-robot systems coupled via robot teleoperation interfaces. In our prior work, we have proposed a novel experimental paradigm to observe the human movements used to control the cameras attached to their head, torso, and hands, which have different configurations and mobility compared to human eyes [101]. We have observed very consistent behaviors of the human head, arm, and body movements in the usage of wearable cameras, which implies the general underlying strategies of the perception-action coupling of the integrated human and tele-robotic systems. We have also

Fig. 2. A representation of the experimental paradigm. The three images show the sequence of actions the subject uses to stack a cup while performing the experiment.

noticed humans attempt to leverage the limited available haptic feedback to compensate for the remote perception issues (e.g., loss of depth information, limited field of view, etc.), which implies the strategies for multi-sensory integration. We further analyzed these observed human behaviors to reveal the perception-motion coupling and multi-sensory integration in a novel context [64]. Following the preliminary work, this paper will extend the analysis to identify human adaptation to remote telepresence indicated by the motor skills or actions that the operators use to learn the telepresence camera control. Our observations from these experiments will be used to discuss their impact on the design of tele-robotic interfaces and assistive autonomy.

3 EXPERIMENT

In direct robot teleoperation, natural perception-action coupling in human motor control cannot be preserved due to the dissimilarity of human and robot embodiment. The added complexity of controlling the robot and vision through a motion capture system [62] might make active camera selection and control during teleoperation harder. The strong spatial skills and high mental effort required to expertly perform vision control during teleoperation might set up high barriers to entry to teleoperation. As a result, we studied human perception-action coordination in a simulated telepresence setup, where participants wearing a head-mounted display received video feeds from cameras attached to their own bodies, thereby trivializing the manipulation component of the task to encourage active camera selection and control.

3.1 Experimental Paradigm

We present a novel experimental paradigm to study the perception-action coordination in the usage of active telepresence through the stacking of cups as seen in Figure 2. The participants wore thick gloves to dampen haptic perception in the hands and hinder their sensation of friction for grasping and in-hand manipulations. This paradigm is designed to trivialize motion control for manipulation, locomotion and active telepresence, while preserving the essential perception capabilities and challenges of remote robots (e.g., 2D display, limited visual range and haptic feedback, unnatural control of camera motions). Voice commands via a wireless microphone were used to switch cameras to make this operation as straightforward as possible without interfering with the experiments. The camera switch is automated based on the command received from the participant. Participants naturally reveal effective perception-action coordination strategies as they adapt to the camera configuration and discover their preferred camera selection and control. Since the task of stacking cups is simple and straightforward, experience or skill played little role in the successful completion of the task.

The participants were instructed to perform a cup-stacking task with the camera views from various wearable and standalone cameras streamed to a VR headset. While the simultaneous display of visual feedback side by side from multiple telepresence cameras is a solution, the cognitive workload and distraction caused by this implementation can prove to be a major

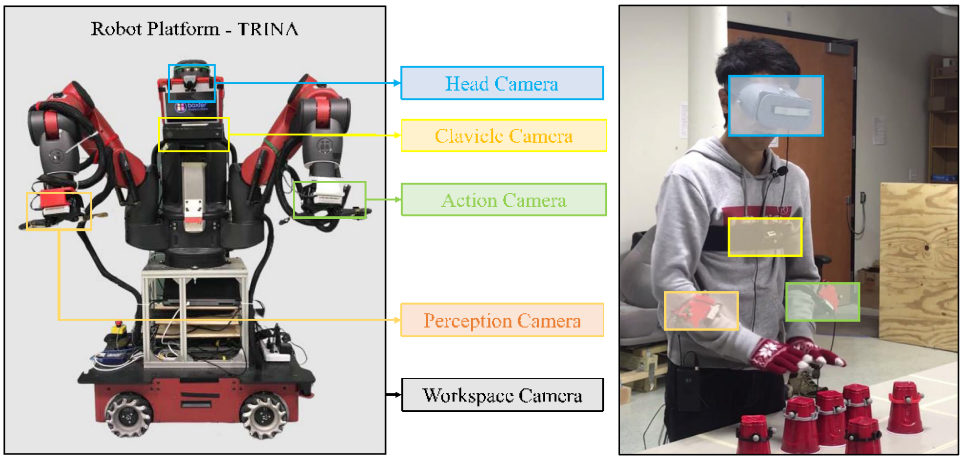

Fig. 3. The camera set-up on the operator (right) is similar to the camera set-up seen on the TRINA humanoid robot (left). The two wrist cameras correspond to perception and action hand cameras. The gloves are used to dampen haptic perception in the hands while performing the experiments.

obstacle for teleoperators [73]. Shown in Figure 3, these telepresence cameras were chosen to simulate the perception cameras equipped on a mobile humanoid nursing robot, which can perform manipulation and navigation tasks under direct teleoperation [62].

The cameras available to the participants are shown in Figure 4 and detail listed below.

- The Head Camera (C_{head}) was attached to the front of the VR headset using a strap, matching natural human eyesight.
- The Clavicle Camera ($C_{clavicle}$) was attached to the chest above the sternum and between the underarms via a strap and mimicked the limited degrees of freedom and range of motion of a robot head camera.
- The Action Camera (C_{action}) and the Perception Camera ($C_{perception}$) were mounted on the 3D printing camera mount and then attached to the dominant hand primarily responsible for manipulation and the non-dominant hand that assists manipulation using straps, respectively.
- The Workspace Camera ($C_{workspace}$) was located across the workspace from the participant on a stationary tripod.

The head, perception, and action hand cameras were the Logitech C310 HD web-camera [2] which has a maximum resolution of 1280 × 720 pixels at 30 frames per second and a diagonal field of view of 60°. The workspace camera was a AUSDOM AW615 webcamera [1] which has a maximum resolution of 1280 × 720 pixels at 30 frames per second with a field of view of 65°. The Google Daydream VR headset with an iPhone 8 mobile phone for the display was used as the Virtual Reality headset for the experiment.

3.2 Participants and Tasks

Our study recruited 16 healthy participants (8 males, 8 females, average age = 23.4 ± 3.6) including student and general populations. The experimental protocol was approved by WPI's Institutional Review Board.

We designed the task to be simple to understand and perform. The stacking task involved three distinct actions: (1) world exploration to observe the environment without interaction, (2) gross manipulation to reach for and carry objects, and (3) fine manipulation of objects with hands. These

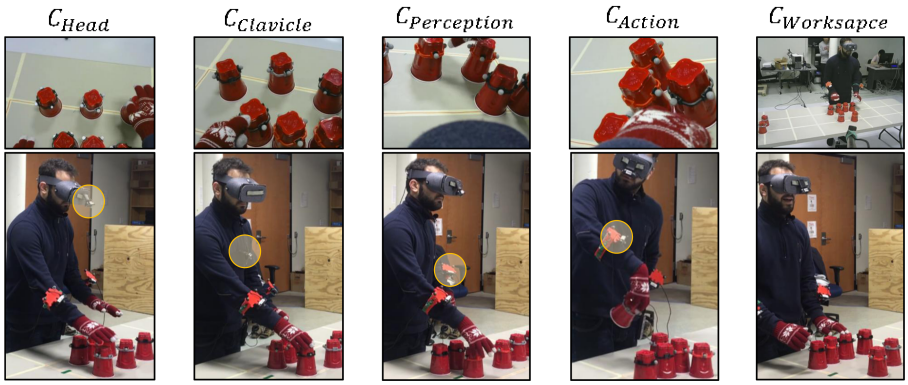

Fig. 4. The demonstration of the video streams from head, clavicle, perception, action, and workspace cameras.

actions, and combinations thereof, span a wide variety of tasks a tele-manipulation system may need to perform. The cups were easy to grasp and manipulate, yet their low-friction surface and light weight made manipulation errors easy to observe.

3.3 Experimental Procedure

Before the experiment, the experimenter equipped the participant with wearable cameras, a VR headset, a microphone and gloves, and introduced the task of stacking lightweight plastic cups into a pyramid. Also, participants were allowed to make small adjustments to the camera field of view to their preference. The available camera adjustments include:

- C_{head}: The angle between the front of the VR headset and the camera lens.
- C_{action} and $C_{perception}$: The location on the forearm (between the elbow and the wrist), the rotation of the mounting bracket around the forearm, and the angle between the mounting plate and the camera lens.
- $C_{clavicle}$: The angle between the sternum mounting strap and the camera lens.
- $C_{workspace}$: The location of the camera tripod relative to the participant and workspace, the angle between the tripod mount and the camera lens, and the focal length of the camera image. The $C_{workspace}$ image was flipped horizontally based on user feedback during a pilot study.

Participants were first asked to stack six cups using the feedback from the telepresence cameras (single camera trials = 2 trials × 5 cameras). For each camera, a participant had a three-minute practice section to get familiar with the selected camera view. The first completed trial was extracted to represent the trial before practice (training phase). This second trial (performing phase) is used to evaluate the operator's skill and workload using the selected camera. The order of camera selection was randomized for each participant to minimize task-learning effects. Camera adjustment was permitted before, during, and after the practice trial, but the wearable camera locations and angles with the mounting point remained static during the performance trial. The participants were asked to prioritize the speed of completing the task (without compromising on comfort) and avoid errors, like knocking over cups, and misaligning while stacking, when performing the task.

For the final trial, participants were instructed to stack ten cups and were able to use and switch the camera view at will (multi-camera trial). This trial did not include the head camera (C_{head}) because in practice, VR telepresence systems may be uncomfortable to use for long periods of time like traditional healthcare worker schedules [24]; we used the C_{head} condition to represent an

ideal camera control baseline against which the other cameras can be compared. The participants were allowed to select the starting camera view of their own preference and were instructed to perform the final trial at a comfortable pace. Before the final trial, participants practiced using voice commands to switch cameras.

3.4 Data Processing

The methodology used while annotating the user study videos for identifying strategies developed with regard to perception-action coupling and human adaptation towards camera control while teleoperating will be expanded upon in the following sections. The annotation of user study videos involved two observers and one supervisor. The supervisor frequently held group discussions to address any conflicts in observations and converge on a conclusion.

3.4.1 General Task Performance.

- Task Completion Times: The time taken to perform the experimental tasks during both the single and multi-camera trials were recorded. The task completion times help us get an objective evaluation of how a particular camera feed aids in performing a task eficiently and intuitively.
- Number of Errors: The number of errors that occurs during the practice and performance phases of the single and multi-camera trials was recorded. These errors include misalignment of cups and knocking down of cups while stacking. Misalignment of cups implies cups were placed in the wrong location while stacking due to lost information from the camera's video feed. These errors help us objectively evaluate how a camera feed enables the correct performance of the tasks with suficient visual feedback provided.
- Camera Selection: During the multi-camera trials the number of camera switches between the various camera views was counted. These results can help identify the preferences of the participant for completing the task and helps objectively validate the responses provided by the participants' responses to the post-study survey.

3.4.2 Human Behavior Analysis.

- Instinctive Head Movement: Participants tended to try and control their camera/vision using their head motion even when the camera is not connected to the head. The head motion was counted as a non-trivial rotation when it is along the transverse and longitudinal axes. The instances of these head motions were compared with task completion times to identify how the user instinctively desires to move their head or go for their natural mode of perception with the complexity of the camera view indicated by task completion time. These motions were counted for the training and performing phase of single camera trials and for all the camera feeds except the head and workspace camera.
- Body Coordination: In the performing phase of the $C_{clavicle}$ camera experiment, the participants moved their upper body or walked sideways to improve their field of vision. The instances of torso motion and walking motion were counted.
- Bimanual Manipulation: Bimanual Manipulation is the eficient way of performing tasks and thus the number of participants performing bimanual manipulation during the performing phase in single camera trials for all the camera feeds and in the multi-camera trial was counted. These motions were counted for $C_{workspace}$, $C_{clavicle}$, and C_{head} when both hands were used to gather and stack cups.
- Fixed Elbow: During the performing phase of the $C_{perception}$, the time during which the perception camera was stationary while performing the experiment was recorded from the

user study videos. This action usually involved the user holding a stationary pose for their elbow on which the perception camera was mounted with respect to their body.

- Saccade Ahead: We observed that some participants looked ahead at the location of the future cup placement before grasping it. This motion was counted for the training and performing phase in single camera trials and for all the camera feeds except the action hand and workspace camera.
- Touch to Locate: Even with limited haptic feedback, participants attempt to identify the cup location and the position of their hands using their ability to touch surfaces. We counted the number of times a hand was used to tap the bottom of the cup to identify the subject's reliance on haptic feedback. This motion was counted for the training and performing phase in single camera trials for all the camera feeds and in the multi-camera trial.
- Tentative Stacking: Participants also tended to stack tentatively by tapping the bottom of the cup they are trying to stack against the surface where they intend to stack to precisely align their cup while stacking. This motion was counted for the training and performing phase in single camera trials for all the camera feeds and in the multi-camera trial.
- Slide Cup on Table: We also counted the number of times the participants slid the cup across the table's surface rather than picking it up. This motion was counted for the training and performing phase in single camera trials for all the camera feeds and in the multi-camera trial.
- Touch for Alignment: While in the stacking phase, we noticed that participants try to use one hand to hold the bottom cup and another hand to make the alignment. This motion was counted for the training and performing phase in single camera trials for all the camera feeds and in the multi-camera trial.

3.4.3 Subjective Survey. The preference of participants for different camera views was verified by the time they spent using different camera views while performing the multi-camera stacking trial. A subjective camera preference survey was performed to record the participant's perceived preferences for different cameras while performing the various components involved in the stacking operation like a choice of the camera in exploring, reaching, grasping, and for the overall performance of the task. They were also asked to provide specific feedback about certain camera viewpoints and configurations and their thoughts on improving the system. Additionally, the participants also participated in a post-study interview at a later stage where the experimental video was replayed to them and questions pertaining to their reasoning for performing the action mentioned in Section 3.4.2. The survey and interview address our results and conclusions highlighted in Section 4.3 and 5.

4 RESULTS

4.1 Perception-Action Coupling

As mentioned in the previous section, we analyzed the human behavior from the performing phase in the single camera trial and combined the multi-camera trial to reveal the vision, haptic, and motion coordination while performing the cup stacking task for each camera usage.

4.1.1 Vision-Motion Coupling. We noticed that people attempt to adjust the camera view using their head not only for the head camera but even for the action, perception, and clavicle cameras (see the head posture in Figure 5). The ANOVA analysis of the Instinctive Head Movement from the performing phase in the single camera trial shows that using action camera causes significantly more frequent futile head motion than the clavicle ($F(1,15) = 24.4$, $p < 0.01$) and perception ($F(1,15) = 22.8$, $p < 0.01$) cameras. We further examined the correlation between task performance

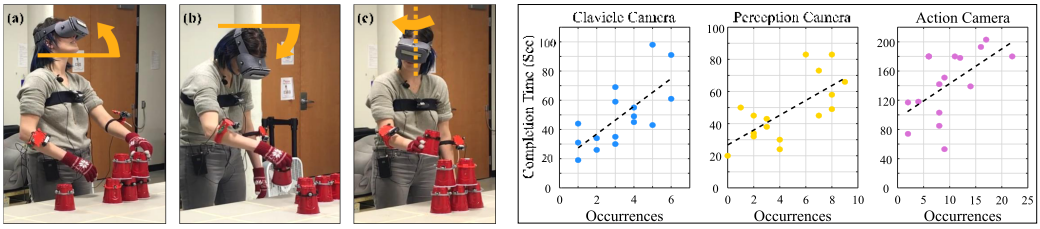

Fig. 5. Compulsive head movement: (a) raise head up; (b) hold head down; (c) turn head side way. Task completion time versus the occurrences of head movement for the clavicle, perception and action hand camera.

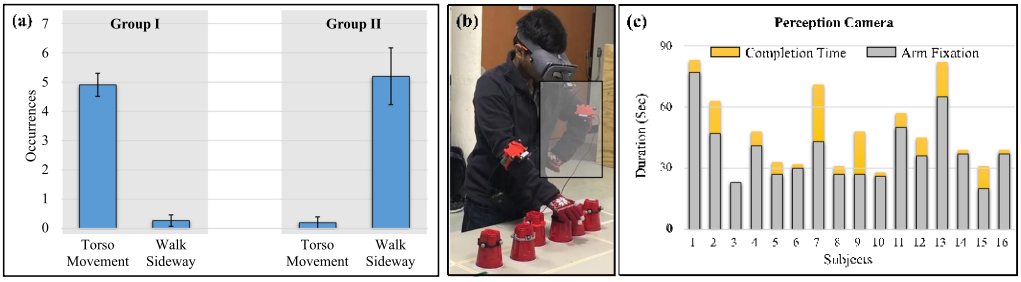

Fig. 6. (a) Two groups of the body coordination while using clavicle camera; (b) The fixed elbow pose for perception camera control; (c) Duration of the arm fixation w.r.t. task completion in perception camera usage.

(task completion time) and the instances of head movements (see Figure 5). A significant linear regression was found for clavicle ($F(1,13) = 12.8$, $p < 0.01$, with an R^2 of 0.5), perception ($F(1,13) = 14.2$, $p < 0.01$, with an R^2 of 0.52) and action ($F(1,13) = 5.9$, $p < 0.05$, with an R^2 of 0.32) cameras. Linear regression of this data predicts that the expected task completion time increases by approximately 9.5 (clavicle), 4.6 (perception), and 4.7 (action) seconds for each occurrence of head movements. Our interview reveals that not being able to control the camera viewpoint using their head movements caused a lot of frustration for every participant. Some participants were able to remind themselves that head movements are not effective for camera viewpoint control and try to suppress this instinct, while others only realized the head movements are ineffective for camera viewpoint control until they felt discomforts like motion sickness or physical fatigue due to activity. Overall, we found that it is more dificult for the participants to realize and suppress the instinctive head movements when the camera is considered more dificult to use.

Based on the usage of the clavicle camera ($C_{clavicle}$), we found that the participants can be separated into two groups by their Body Coordination. The result from the performing phase (Figure 6(a)) shows that one group of participants tend to explore the environment through torso motions to control the camera view instead of walking around while the other group walked around in the workspace for the same.

As shown in Figure 6(c), the proportion of task time that the participants employed a Fixed Elbow Posture (refer Figure 6(b)) while using the perception camera for the fixed camera trial was 82.9 ± 12.7 percent. We also found that the majority of the participants (11 of 16) tend to fix their shoulder joints and move their torso to control the perception camera viewpoint, thus limiting the perception hand camera motions with respect to the base frame of the torso. Our interview reveals that: most participants intentionally limit the elbow and shoulder motions of the perception camera arm to better remember the spatial relationship of the perception hand camera

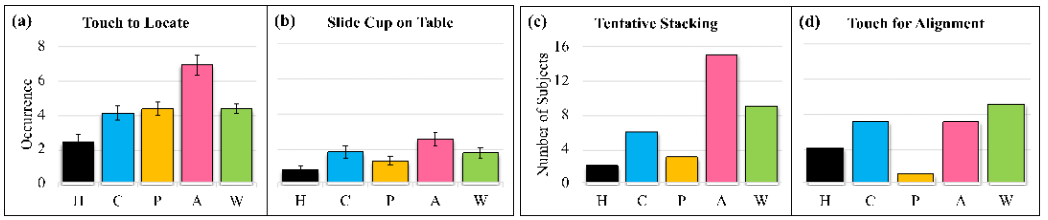

Fig. 7. (a) Touching-to-locate, (b) sliding cups-on-table, (c) tentative-stacking, and (d) touching-for-alignment actions observed in the usage of the head (H), clavicle (C), perception (P), action (A), and workspace (W) camera.

with respect to their body. This lets them coordinate the camera motions with the motions of their manipulating hand, object, and workspace. Some participants indicated that they unconsciously choose the elbow angle so that the perception camera is not too far away from their body, making it easy and comfortable to move and look around the workspace. Overall, the situational awareness of the perception camera pose with respect to their body is critical to the planning of coordinated perception and manipulation actions.

Whenever possible, participants preferred Bimanual Manipulation, to speed up the task and to increase their reach without moving the body. The usage of bimanual manipulation, in both symmetric and asymmetric forms, is observed when using the head, clavicle and workspace cameras, for reaching to collect cups, and for placing/stacking the cups in the same row. We also found that bimanual control/manipulation as discussed in Section 3.4 is more frequent with the head camera (13/16 participants) than the clavicle (3/16 participants) and workspace cameras (4/16 participants). Our interview shows that bimanual manipulations are more dificult when using the clavicle camera because reaching both hands forward to objects caused the torso to lean forward which reduces the viewpoint control of the clavicle camera. Compared to unimanual manipulation, bimanual manipulation is more eficient yet more complex to plan.

4.1.2 Haptic-Motion Coupling. Our experimental paradigm limited the haptic perception of the participants so that they had to rely mostly on the visual feedback from RGB cameras to perform the tasks. However, participants still learned to utilize the limited haptic feedback received through the thick gloves they wore to compensate for reduced visual feedback. Across all the participants and camera viewpoints, we observed the participants (1) Touching to Locate the cups to build the *contact* sensation, (2) Sliding Cup on the Table so that they can leverage the haptic perception of table *constraints* to better control the moving motions, (3) Stacking Tentatively to get the better placement location, and (4) Touching for Alignment using the bottom of the cup.

As mentioned in Section 3.4, haptic-motion coupling actions like touch to locate, sliding the cup on the table, tentative stacking, and touching for alignment were counted. Figure 7(a) shows the mean and standard deviation of touch-to-locate occurrences across participants for different cameras. The ANOVA analysis shows that using an action camera causes significantly more frequent ($p < 0.01$) touch-to-locate actions than all other cameras. Also, touch-to-locate actions occurred the least ($p < 0.01$) when using the head camera. These significant differences indicate that participants resort more to haptic feedback for the cameras more dificult to use (as indicated in our survey feedback). Both the observed human behavior and the interview feedback indicate that (1) touching-to-locate an object is the most necessary haptic perception to complement the loss of depth information and limited field of view while using active telepresence; and (2) the haptic feedback does not have to be strong and realistic if it can provide a sense of contact. We hypothesize that this can largely reduce the mental workload and stress due to uncertainty in perception while improving task accuracy and eficiency.

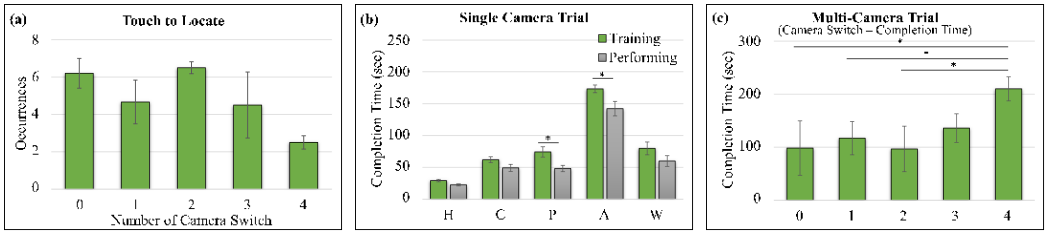

Fig. 8. (a) The correlation between camera switches and touch to locate the action in multi-camera trial. (b) The comparison of the completion time between training and performing phase in single camera trial. (c) The comparison of the completion time across the number of camera switches in multi-camera trial.

In addition to touch-to-locate, participants also used touch-for-alignment when tentatively stacking, aligning, and sliding the cups on the table. Overall, haptic compensations were required for the cameras identified as non-intuitive and ineficient to use. In Figure 7(b), sliding cups on the table are observed the most in action hand camera usage. On the other hand, the tentative stack-ing actions are used by 15 of 16 participants when working with the action hand camera, and by two of 16 participants when working with the head camera (see Figure 7(c)). While in Figure 7(d), touch for alignment is observed in more than half of the participants for the workspace cameras followed by action hand and clavicle cameras. The interview feedback reveals that: (1) the gloves effectively damped most of their haptic perception; and (2) the limited tactile sensing is still very helpful to the task in many cases.

4.1.3 Vision-Haptic Coupling. In the single camera trial, the participant used the video feedback from a single camera to perform the cup stacking task. This helps us compare the performance and human behavior across cameras. Figure 8(a) shows the concept of vision-haptic coupling, where information gathering while touching to locate is offset by the increased number of camera switches and vice-versa. Unlike the single camera trial that receives vision feedback from one camera viewpoint, the multi-camera trial allows participants to switch the viewpoint across cameras. The need for haptic compensation like loss of depth information can be compensated by switching the camera view to a different view like the perception hand camera. However, our interview feedback suggests that the cognitive workload increases when having to involve more camera switches thus limiting the bandwidth to perform other actions.

4.2 Human Adaptation

We performed the analysis of human behavior in the single camera trial to investigate perception-action coupling while using the active telepresence cameras. In this section, we further compare the performance and the human behavior between the training and performing phases in single camera trial to disclose the impact of the practice and motor learning process. We also combine the information from multi-camera trial to better understand how the skill sets learned from single camera trial transfer to multi-camera trial.

4.2.1 General Performance. We compared the performance in terms of *task completion time* and *number of errors* between the training and performing phase in single camera trial. An error can occur when the cup drops due to: (1) misalignment during stacking, and (2) collision while moving hands around. Figure 8(b) shows the comparison of the task completion times between the training and performing phases in single camera trial. The ANOVA analysis shows that the completion time had significantly reduced after practice while using the perception ($F(1,30) = 7.3$, $p < 0.05$) and action ($F(1,30) = 5.6$, $p < 0.05$) hand camera. These significant differences indicate

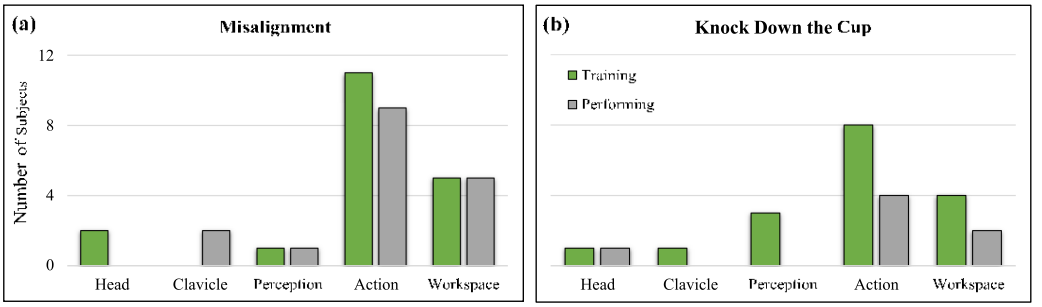

Fig. 9. The errors occurred in single camera trial in the type of (a) misalignment; (b) colliding with the cup.

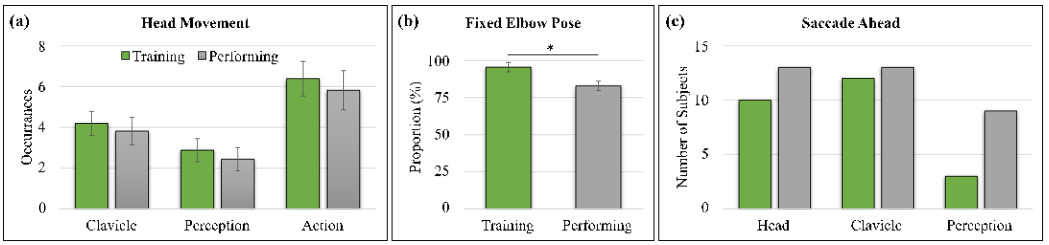

Fig. 10. The comparison of the human behavior between training and performing phase for: (a) head movement, (b) arm fixation, and (c) looking ahead while using active telepresence camera.

that the comprehensive practice section is necessary for dificult cameras. In addition, Figure 8(c) shows the correlation between task completion time and the number of camera switches in multi-camera trial. Based on the ANOVA analysis, the time needed to complete the task is significantly longer when participants had the most number of camera switches than switches twice ($F(1,4)$ = 11.3, $p < 0.05$), once ($F(1,3)$ = 12.5, $p < 0.05$) and none ($F(1,5)$ = 8.1, $p < 0.05$).

Figure 9(a) shows that trials using the action hand camera had more participants who misaligned cups compared to the other camera views and displayed limited improvement after practice (11/16 to 9/16 participants). This implies the non-intuitive camera usage in terms of loss of depth information which may lead to failure of the task despite the practice session. In Figure 9(b), the action hand camera still caused most participants to knock down the cup while moving their hands around. However, the practice helped prevent the collision with the cup when using: perception (3/16 to 0/16 participants), action (8/16 to 4/16 participants), and workspace (4/16 to 2/16 participants) cameras. This implies the narrow field of view and complex camera control can be adapted to by practice.

4.2.2 Motor Learning. We identified several actions that we constantly observed from both the training and performing phases in the single camera trial. As shown in Figure 10(a), we found no significant differences for the Instinctive Head Movement in the clavicle, perception, and action hand cameras between the training and performing phase. This solidifies that it is dificult to suppress the head movement though participants are able to realize that the camera cannot be controlled by the head during the training phase. Figure 10(b) shows the duration (the proportion with respect to task completion time) of fixing the elbow in a certain posture while using the perception hand camera (including the training and performing phase). We found that the duration of the Fixed Elbow posture significantly reduces ($F(1,30)$ = 13.5, $p < 0.01$) after practice. This implies that the training session helped improve the participant's understanding of the spatial

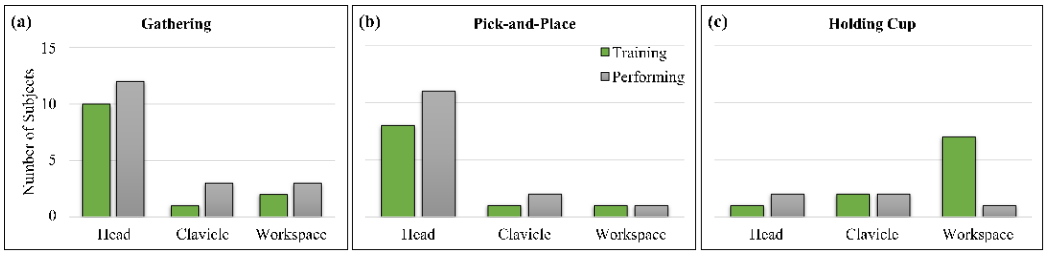

Fig. 11. Bimanual manipulation for: (a) gathering, (b) pick-and-place, and (c) holding cup.

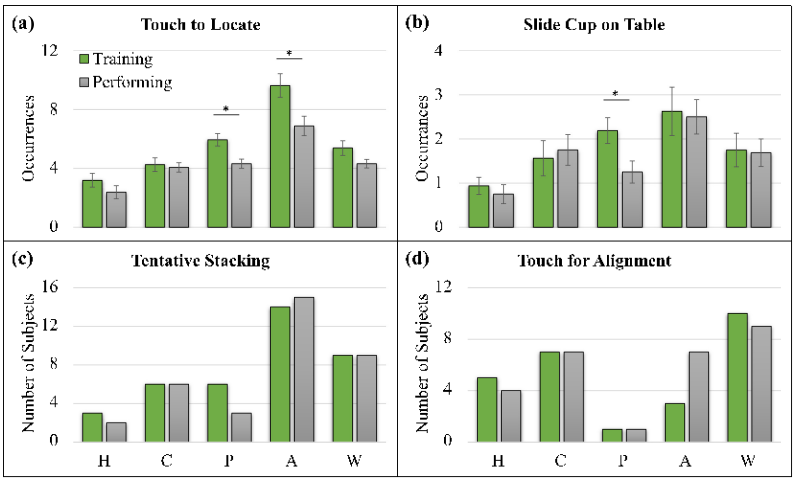

Fig. 12. Comparison of the haptic compensation between training and performing phase including: (a) touch-to-locate, (b) slide cup on the table, (c) tentative stacking, and (d) touch-for-alignment.

relationship of the perception of hand camera with respect to their body. We noticed that some participants made a Saccade Ahead of the cup, just before grasping it, to a location on the future placement. In Figure 10(c), there is a noticeable increase in the participants (from 3/16 to 9/16) who looked ahead when performing pick-and-place motion in the perception hand camera trial after the practice session. This observation implies the practice section can improve the cognitive bandwidth when controlling a non-intuitive camera.

We divided the Bimanual Manipulation into gathering (maneuvering multiple cups in the workspace), picking/stacking (picking up and stacking actions of the cup in the workspace), and holding the cup (holding and carrying a cup). Figures 11(a) and 11(b) show the increase in the number of participants who performed the bimanual gathering and picking/stacking after practice in the head camera trial which was identified as the most intuitive camera view to control. However, lesser participants used both hands to hold the cup after practice while using the workspace camera (see Figure 11(c)). Our interview feedback indicates that they try to eliminate the mirror effect that occurs while the workspace camera by holding a cup with both hands.

We also investigated the differences in Haptic Compensations between the training phase and performing phase to better understand how humans adapt to the different active telepresence cameras usage. Figures 12(a) and 12(b) show the mean and standard deviation of touch-to-locate and slide-cup-on-table occurrences across participants for different cameras in the training and performing phase. The ANOVA analysis indicates that using the perception ($F(1,30) = 9.5$,

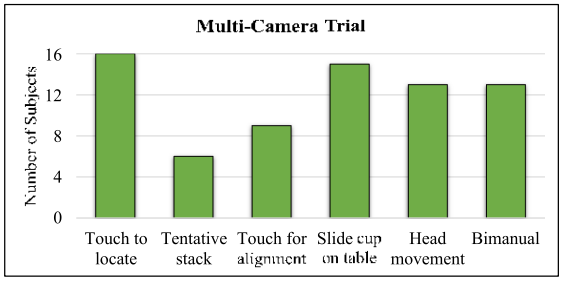

Fig. 13. The number of subjects performed the haptic compensation, head movement, and bimanual manipulation in the multi-camera trial.

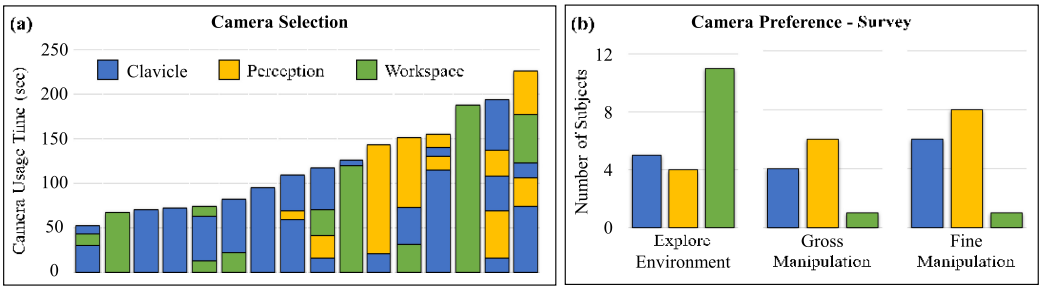

Fig. 14. The subjective assessment about the camera selection and preference from multi-camera trial.

$p < 0.01$) and action ($F(1,30) = 7.1$, $p < 0.05$) hand cameras significantly reduces the touch-to-locate actions and perception hand camera significantly reduces ($F(1,30) = 5.9$, $p < 0.05$) the slide-cup-on-table action after practice. These differences imply that haptic feedback helped improve the usage of the more dificult, limited field-of-view cameras by virtue of being close to and focused on the object of manipulation. Figures 12(c) and 12(d) show the number of participants who performed the tentative stacking and touch-to-alignment actions for different cameras in the training and performing phase. We found a slight decrease in the participants who performed the tentative stacking while using the perception hand camera and an increase in the participants who performed the touch-for-alignment in the action hand camera after practice.

We further analyzed the identified actions including head movement, bimanual operation, and haptic compensations in multi-camera trials to investigate the process of human adaptation in the usage of active telepresence cameras. As shown in Figure 13, more than half of the participants use the touch-to-locate (16/16 participants), slide-cup-on-table (15/16 participants), bimanual manipulation (13/16 participants) and touch-for-alignment (9/16 participants) actions while using the active telepresence cameras. Furthermore, we observed 13 out of 16 participants still tried to control the camera viewpoint using their heads.

4.3 Camera Selection and Preference

We conducted the analysis of the camera preference as indicated by camera selection while performing the multi-camera trial and post-study survey. Figure 14(a) shows the correlation between the duration of camera usage and the number of camera switches in the order of total task completion time. We found that fewer camera switches and participants who had a higher proportion of clavicle camera usage led to better performance (in terms of task completion time). These observations aligned with the recent study of multi-view interface design where it was observed

that autonomous switching might ease the control effort [77]. The camera preference survey indicates that the workspace camera is preferred while exploring the environment and the perception hand camera for gross and fine manipulation followed by the clavicle camera (see Figure 14(b)). It is to be noted that the action hand camera was never selected during the multi-camera experiment. From a human action perspective, these results aligned with the findings from the recent design of the adaptive viewpoint in telemanipulation where the performance was improved with the shared-autonomous camera control [80].

5 DISCUSSION

In this paper, we investigate how humans coordinate perception-action coupling during active telepresence through a novel experimental paradigm that emphasizes limited haptic feedback. The results from participant task performance, human motion analysis, and user feedback reveal the integration of vision, motion and haptic feedback, human natural motor learning, and preferences. In this section, we will further discuss the implication of suitable camera control and selection as well as the preferable robot teleoperation interface design.

5.1 Desirable Characteristics of Viewpoint Control and Selection

Tele-nursing robots need different viewpoints from strategically placed telepresence cameras to provide a comprehensive view of the environment and the task workspace. A natural approach to control and select the cameras is necessary to reduce the cognitive workload and increase the transparency in robot teleoperation. The findings from our human motion analysis identify several components for camera control and selection so that the perception-action coupling complies with the natural human motor control.

As human tracking technologies become more accurate, portable and affordable, head- and gaze-control are getting increasingly adopted for the control of eye-in-hand cameras of manipulator and continuum robots [81], and the head camera of mobile and humanoid robots [21, 85]. While matching human eyes to robot eyes is considered to be a natural design choice, it is also not rare to see remote cameras controlled by robot hands. When multiple cameras are available (as on many commercial and prototype humanoid robot platforms [6, 62]), head and hand control are usually only used for the head and eye-in-hand cameras, respectively. When a teleoperator switches their primary viewpoint (i.e., the camera view they primarily rely upon to perform the task) from the head-to-hand camera, adapting to control of camera viewpoint via hands always causes interruption of task performance. Lessons learned from (tele-robotic) laparoscopic surgery training also indicate that it takes much more training effort to learn to use hand-controlled cameras [102]. The intuitive nature of the head motion observed in the clavicle and hand camera trials, highlights the need for egocentric control (usage of head to control gaze) to control any camera viewpoint selected as the primary viewpoint. This head-controlled dynamic viewpoint aligned with the recent mobile manipulator implementation [83].

In direct teleoperation, understanding the camera pose and motions is critical to control the robot action components (e.g., end-effectors, mobile base). Even in supervisory control, lack of spatial awareness due to sub-optimal camera pose will reduce the operator's situational awareness and capability to intervene if the robot autonomy is not reliable [17]. The elbow joint fixation we observed from the single-camera trial highlights the strategy that humans adopted to maintain the spatial awareness of the camera pose with respect to their bodies. A preferable method for camera control thus should limit the degree of freedom to be controlled by simple translation or rotation. In the case of supervisory control, the trajectory of the autonomous camera system should be easy to understand and predictable for the operator. Learning preferences for camera viewpoints for specific tasks increases situational awareness crucial for supervisory control of remote robots.

In the fixed camera usage, our study reveals that intuitive camera control is preferred for better performance (faster completion time and fewer errors) and lower cognitive workload. When people have more cognitive bandwidth, they are able to perform complex motions. This is supported by the fact that most participants perform bimanual operations and look-ahead motions to place the cups when using the head camera. On the other hand, our camera preference survey indicates the correlation between the purpose of the action and the preferred camera for this action in a multi-camera setup. However, the camera choice in the multi-camera trial shows a large variance with no consistency across participants. These outcomes imply that customization of autonomous camera selection with respect to user groups, or even personalization, is necessary.

As part of our future work, we will further develop an intuitive method to control multi-camera active telepresence. A user study will also be devised to investigate if the perception and action hand camera could be controlled using the head, hand, or a mixture of head and hand motions as well as to understand the human behavior, preference, and rationals in the usage of a multi-camera active telepresence system. The entire experiment was performed in a motion capture enclosure with motion capture markers located on the VR headset, wrist camera mounts, and cups. However, the motion capture data did not yield any significant results due to the lack of quality. We will further utilize the VR trackers to get meaningful data to investigate human behavior objectively. We will also explore the use of autonomous camera control and selection to reduce operator workload and improve task performance in supervisory control.

5.2 Design Philosophy for Multi-sensory Integration

The experiment paradigm enables participants to manipulate the object with their own hands, which is more capable of moving and sensing through proprioception. In object manipulation, the benefit of proprioception is limited because visual information is still required to locate the target and a freely moving arm will not help in locating the object. The feedback from participants also indicated that they need to place their hand in the view to better understand the relationship between the arm and the target object implying the limited usage of proprioception during object manipulation. However, if proprioception combines with the human's memory of the workspace, it will indeed ease the effort in object manipulation because the direction towards the target can be identified.

Our human motion analysis indicates that people tend to use haptic feedback to compensate for the loss of depth information and narrow limited vision of the visual feedback via active telepresence. The desire for haptic feedback ranges from precise or gross manipulation to general environment exploration. Indeed, human motor control has the instinct to pursue visuo-haptic sensory integration when they perform tasks with their own bodies as well as via teleoperation interfaces. Unfortunately, state-of-the-art haptic feedback rendering technologies cannot enable the teleoperation interface to provide the most realistic haptic perception. The tradeoff between the complexity and suitable level of the haptic feedback to compensate for the limitation of active telepresence visual feedback needs to be studied. Our study reveals that: (1) human motor control can achieve very effective visuo-haptic sensory integration with active telepresence visual feedback and limited haptic feedback; and (2) for general-purpose manipulation tasks, adding a little bit of haptic feedback to indicate the contacts with the remote physical environment will be much more simple and effective than fabricating complicate strategies for the optimization of camera control and selection.

Inspired by findings from our study, we propose a philosophy for visuo-haptic sensory integration to re-establish the perception-motion coupling with the perception and action capabilities of

the remote robotic system. From a high-level perspective, there are three strategies to achieve this goal. Take several designs in literature and our prior work for example: (1) we may restore the lost haptic perception by adding vibrotactile feedback to indicate contacts with the remote environment [109]; (2) we may also replace haptic display with augmented reality visual display [7]; and (3) we may delegate the task components that heavily rely upon haptic feedback to reliable robot autonomy, to eliminate the need for remote perception-action coupling [63]. Our future work will implement the proposed philosophy and conduct a user study to compare the eficacy of each haptic compensation approach and user acceptance as well as preference for the use of robot teleoperation.

5.3 Limitations

Advanced Gaze Analysis. There was no gaze detection used in this paper. Our future work regarding perception studies will involve the utilization of a gaze tracker to accurately track human gaze motion and collect more reliable data. This will help us accurately determine where the operator is looking at different parts of the task improving our ability to draw information regarding camera view usage.

Influence of Human Sensation. As mentioned in Section 4, the participants felt that the usage of multiple gloves effectively reduced the haptic feedback while performing the task. However, a systematic manner to dampen the haptic sensation was not implemented. Studying the impacts of varying levels of haptic dampening and their impact of camera interface control will be an interesting avenue of future research.

Integration with Teleoperation Systems. Ideal teleoperation needs to consider both remote perception and robot control. As the first step, the proposed experiment paradigm provided the simulated telepresence setting and retained the human's ability to manipulate the object which relaxed the control effort and focused on the investigation of active telepresence design in remote perception. A further investigation of teleoperating the robot with the preferred active telepresence design will be conducted along with the suitable robot control interface.

6 CONCLUSION

This paper analyzed human motion behaviors and camera selection preferences for the user study conducted primarily with visual feedback from various wearable cameras in a simulated telepresence setting. The results primarily identify the impact different camera feeds have on stacking via telepresence. These results help us identify the preferred design philosophy for visuo-haptic sensory feedback for teleoperation interfaces as well as the preferred mode of viewpoint control and selection for active telepresence. The main findings of this article and the suggested designs are:

(1) Intuitive Control of Multi-Camera Active Telepresence — The instinctive head motion we constantly observed to not only control the head camera but the cameras attached to their torso and hands indicates the head should control for any camera selected for telepresence.
(2) Active Telepresence Assistance for Supervisory Control — The participants intended to maintain the arm posture for better spatial awareness of camera pose implies that the motions of the shared autonomy camera should follow the simple translation or rotation to make it easier to understand and predict by the users.
(3) The Need for Visuo-Haptic Sensory Integration — People tend to resort to using every possible haptic sensation to compensate for the limitation of the visual feedback reiterating the importance of integrating vision and haptic feedback in robot teleoperation interfaces.

REFERENCES

[1] [n.d.]. AW615 Webcam. https://ausdom.com/product/full-hd-1080p-wide-angle-view-webcam-with-anti-distortion-ausdom-aw61.

[2] [n.d.]. Logitech C310 HD Webcam, 720p Video with Noise Reducing Mic. https://www.logitech.com/en-us/products/webcams/c310-hd-webcam.960-000585.html.

[3] Takashige Abe, Nicholas Raison, Nobuo Shinohara, M. Shamim Khan, Kamran Ahmed, and Prokar Dasgupta. 2018. The effect of visual-spatial ability on the learning of robot-assisted surgical skills. *Journal of Surgical Education* 75, 2 (2018), 458–464.

[4] Naotoshi Abekawa and Hiroaki Gomi. 2015. Online gain update for manual following response accompanied by gaze shift during arm reaching. *Journal of Neurophysiology* 113, 4 (2015), 1206–1216.

[5] Evan Ackerman. 2015. Oculus Rift-Based System Brings True Immersion to Telepresence Robots. https://spectrum.ieee.org/automaton/robotics/robotics-hardware/upenn-dora-platform.

[6] E. Ackerman. 2018. Moxi prototype from diligent robotics starts helping out in hospitals. *IEEE Spectrum*. https://spectrum.ieee.org/automaton/robotics/industrial-robots/moxi-prototype-from-diligent-robotics-starts-helping-out-in-hospitals.

[7] Jacopo Aleotti, Giorgio Micconi, Stefano Caselli, Giacomo Benassi, Nicola Zambelli, Manuele Bettelli, and Andrea Zappettini. 2017. Detection of nuclear sources by UAV teleoperation using a visuo-haptic augmented reality interface. *Sensors* 17, 10 (2017), 2234.

[8] Sai Krishna Allani, Brendan John, Javier Ruiz, Saurabh Dixit, Jackson Carter, Cindy Grimm, and Ravi Balasubramanian. 2016. Evaluating human gaze patterns during grasping tasks: Robot versus human hand. In *Proceedings of the ACM Symposium on Applied Perception*. 45–52.

[9] Luis Almeida, Paulo Menezes, and Jorge Dias. 2020. Interface transparency issues in teleoperation. *Applied Sciences* 10, 18 (2020), 6232.

[10] Reuben M. Aronson and Henny Admoni. 2020. Eye gaze for assistive manipulation. In *Companion of the 2020 ACM/IEEE International Conference on Human-Robot Interaction*. 552–554.

[11] Ruzena Bajcsy, Yiannis Aloimonos, and John K. Tsotsos. 2018. Revisiting active perception. *Autonomous Robots* 42, 2 (2018), 177–196.

[12] Michael Barnes, Linda R. Elliott, Julia Wright, Angelique Scharine, and Jessie Chen. 2019. *Human-Robot Interaction Design Research: From Teleoperations to Human-Agent Teaming*. Technical Report. CCDC Army Research Laboratory Aberdeen Proving Ground United States.

[13] Sean Barton, Scott Steinmetz, Gabe Diaz, Jonathan Matthis, and Brett Fajen. 2017. The visual control of walking over complex terrain with flat versus raised obstacles. *Journal of Vision* 17, 10 (2017), 707–707.

[14] Bennett I. Bertenthal, James L. Rose, and Dina L. Bai. 1997. Perception–action coupling in the development of visual control of posture. *Journal of Experimental Psychology: Human Perception and Performance* 23, 6 (1997), 1631.

[15] Jeannette Bohg, Karol Hausman, Bharath Sankaran, Oliver Brock, Danica Kragic, Stefan Schaal, and Gaurav S. Sukhatme. 2017. Interactive perception: Leveraging action in perception and perception in action. *IEEE Transactions on Robotics* 33, 6 (2017), 1273–1291.

[16] Ali Borji and Laurent Itti. 2012. State-of-the-art in visual attention modeling. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 35, 1 (2012), 185–207.

[17] Mark Boyer, Mary L. Cummings, Lee B. Spence, and Erin T. Solovey. 2015. Investigating mental workload changes in a long duration supervisory control task. *Interacting with Computers* 27, 5 (2015), 512–520.

[18] Chiara Bozzacchi, Robert Volcic, and Fulvio Domini. 2016. Grasping in absence of feedback: Systematic biases endure extensive training. *Experimental Brain Research* 234, 1 (2016), 255–265.

[19] Berk Calli, Wouter Caarls, Martijn Wisse, and P. Jonker. 2018. Viewpoint optimization for aiding grasp synthesis algorithms using reinforcement learning. *Advanced Robotics* 32, 20 (2018), 1077–1089.

[20] Kieran Carnegie and Taehyun Rhee. 2015. Reducing visual discomfort with HMDs using dynamic depth of field. *IEEE Computer Graphics and Applications* 35, 5 (2015), 34–41.

[21] C. Carreto, D. Gêgo, and L. Figueiredo. 2018. An eye-gaze tracking system for teleoperation of a mobile robot. *Journal of Information Systems Engineering & Management* 3, 2 (2018), 16.

[22] Harvey Cash and Tony J. Prescott. 2019. Improving the visual comfort of virtual reality telepresence for robotics. In *International Conference on Social Robotics*. Springer, 697–706.

[23] Jessie Y. C. Chen and Michael J. Barnes. 2014. Human–agent teaming for multirobot control: A review of human factors issues. *IEEE Transactions on Human-Machine Systems* 44, 1 (2014), 13–29.

[24] Jessie Y. C. Chen, Ellen C. Haas, and Michael J. Barnes. 2007. Human performance issues and user interface design for teleoperated robots. *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)* 37, 6 (2007), 1231–1245.

[25] Frederic R. Danion and J. Randall Flanagan. 2018. Different gaze strategies during eye versus hand tracking of a moving target. *Scientific Reports* 8, 1 (2018), 1–9.
[26] Tareq Dardona, Shahab Eslamian, Luke A. Reisner, and Abhilash Pandya. 2019. Remote presence: Development and usability evaluation of a head-mounted display for camera control on the da Vinci Surgical System. *Robotics* 8, 2 (2019), 31.
[27] Dibyendu Kumar Das, Mouli Laha, Somajyoti Majumder, and Dipnarayan Ray. 2018. Stable and consistent object tracking: An active vision approach. In *Advanced Computational and Communication Paradigms*. Springer, 299–308.
[28] Jorge de León, Mario Garzón, David Garzón, Eduardo Narváez, Jaime del Cerro, and Antonio Barrientos. 2016. From video games multiple cameras to multi-robot teleoperation in disaster scenarios. In *2016 International Conference on Autonomous Robot Systems and Competitions (ICARSC)*. IEEE, 323–328.
[29] Brian P. DeJong, J. Edward Colgate, and Michael A. Peshkin. 2004. Improving teleoperation: Reducing mental rotations and translations. In *IEEE International Conference on Robotics and Automation, 2004. Proceedings. ICRA'04. 2004*, Vol. 4. IEEE, 3708–3714.
[30] Geneviève Desmarais, Melissa Meade, Taylor Wells, and Mélanie Nadeau. 2017. Visuo-haptic integration in object identification using novel objects. *Attention, Perception, & Psychophysics* 79, 8 (2017), 2478–2498.
[31] Jonathan S. Diamond, Daniel M. Wolpert, and J. Randall Flanagan. 2017. Rapid target foraging with reach or gaze: The hand looks further ahead than the eye. *PLoS Computational Biology* 13, 7 (2017), e1005504.
[32] Gabriel Diaz, Joseph Cooper, and Mary Hayhoe. 2013. Memory and prediction in natural gaze control. *Philosophical Transactions of the Royal Society B: Biological Sciences* 368, 1628 (2013), 20130064.
[33] Gabriel Diaz, Joseph Cooper, Constantin Rothkopf, and Mary Hayhoe. 2013. Saccades to future ball location reveal memory-based prediction in a virtual-reality interception task. *Journal of Vision* 13, 1 (2013), 20–20.
[34] Javier Dominguez-Zamora, Shaila Gunn, and Daniel Marigold. 2017. Does uncertainty about the terrain explain gaze behavior during visually guided walking? *Journal of Vision* 17, 10 (2017), 709–709.
[35] Digby Elliott, Werner F. Helsen, and Romeo Chua. 2001. A century later: Woodworth's (1899) two-component model of goal-directed aiming. *Psychological Bulletin* 127, 3 (2001), 342.
[36] Francesca C. Fortenbaugh, John C. Hicks, Lei Hao, and Kathleen A. Turano. 2006. High-speed navigators: Using more than what meets the eye. *Journal of Vision* 6, 5 (2006), 3–3.
[37] Christoph Gebhardt, Stefan Stevšić, and Otmar Hilliges. 2018. Optimizing for aesthetically pleasing quadrotor camera motion. *ACM Transactions on Graphics (TOG)* 37, 4 (2018), 1–11.
[38] Saiedeh Ghahghaei and Preeti Verghese. 2015. Eficient saccade planning requires time and clear choices. *Vision Research* 113 (2015), 125–136.
[39] Tricia L. Gibo, Winfred Mugge, and David A. Abbink. 2017. Trust in haptic assistance: Weighting visual and haptic cues based on error history. *Experimental Brain Research* 235, 8 (2017), 2533–2546.
[40] Jerry L. Grifith, Patricia Voloschin, Gerald D. Gibb, and James R. Bailey. 1983. Differences in eye-hand motor coordination of video-game users and non-users. *Perceptual and Motor Skills* 57, 1 (1983), 155–158.
[41] Sahar N. Hamid, Brian Stankiewicz, and Mary Hayhoe. 2010. Gaze patterns in navigation: Encoding information in large-scale environments. *Journal of Vision* 10, 12 (2010), 28–28.
[42] Felix G. Hamza-Lup, Crenguta M. Bogdan, Dorin M. Popovici, and Ovidiu D. Costea. 2019. A survey of visuo-haptic simulation in surgical training. *arXiv preprint arXiv:1903.03272* (2019).
[43] Peng Han, Daniel R. Saunders, Russell L. Woods, and Gang Luo. 2013. Trajectory prediction of saccadic eye movements using a compressed exponential model. *Journal of Vision* 13, 8 (2013), 27–27.
[44] Mary M. Hayhoe. 2017. Vision and action. *Annual Review of Vision Science* 3 (2017), 389–413.
[45] Mary M. Hayhoe and Jonathan Samir Matthis. 2018. Control of gaze in natural environments: Effects of rewards and costs, uncertainty and memory in target selection. *Interface Focus* 8, 4 (2018), 20180009.
[46] Werner F. Helsen, Digby Elliott, Janet L. Starkes, and Kathryn L. Ricker. 2000. Coupling of eye, finger, elbow, and shoulder movements during manual aiming. *Journal of Motor Behavior* 32, 3 (2000), 241–248.
[47] Chong Huang, Fei Gao, Jie Pan, Zhenyu Yang, Weihao Qiu, Peng Chen, Xin Yang, Shaojie Shen, and Kwang-Ting Cheng. 2018. Act: An autonomous drone cinematography system for action scenes. In *2018 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 7039–7046.
[48] Thomas Huk. 2006. Who benefits from learning with 3D models? The case of spatial ability. *Journal of Computer Assisted Learning* 22, 6 (2006), 392–404.
[49] J. A. Ibbotson, C. L. MacKenzie, C. G. L. Cao, and A. J. Lomax. 1999. Gaze patterns in laparoscopic surgery. *Studies in Health Technology and Informatics* (1999), 154–160.
[50] Danut C. Irimia, Woosang Cho, Rupert Ortner, Brendan Z. Allison, Bogdan E. Ignat, Guenter Edlinger, and Christoph Guger. 2017. Brain-computer interfaces with multi-sensory feedback for stroke rehabilitation: A case study. *Artificial Organs* 41, 11 (2017), E178–E184.

[51] Masato Ito and Kosuke Sekiyama. 2015. Optimal viewpoint selection for cooperative visual assistance in multi-robot systems. In *2015 IEEE/SICE International Symposium on System Integration (SII)*. IEEE, 605–610.
[52] Laurent Itti and Pierre Baldi. 2009. Bayesian surprise attracts human attention. *Vision Research* 49, 10 (2009), 1295–1306.
[53] Jelena Jovancevic-Misic and Mary Hayhoe. 2009. Adaptive gaze control in natural environments. *Journal of Neuroscience* 29, 19 (2009), 6234–6238.
[54] David C. Knill, Amulya Bondada, and Manu Chhabra. 2011. Flexible, task-dependent use of sensory feedback to control hand movements. *Journal of Neuroscience* 31, 4 (2011), 1219–1237.
[55] Annica Kristoffersson, Silvia Coradeschi, and Amy Loutfi. 2013. A review of mobile robotic telepresence. *Advances in Human-Computer Interaction* 2013 (2013).
[56] Jessica R. Kuntz, Jenni M. Karl, Jon B. Doan, Melody Grohs, and Ian Q. Whishaw. 2020. Two types of memory-based (pantomime) reaches distinguished by gaze anchoring in reach-to-grasp tasks. *Behavioural Brain Research* 381 (2020), 112438.
[57] Johannes Kurz, Mathias Hegele, Mathias Reiser, and Jörn Munzert. 2017. Impact of task dificulty on gaze behavior in a sequential object manipulation task. *Experimental Brain Research* 235, 11 (2017), 3479–3486.
[58] Simon Lacey and K. Sathian. 2020. Visuo-haptic object perception. *Multisensory Perception* (2020), 157–178.
[59] Kai Lan and Kosuke Sekiyama. 2019. Autonomous robot photographer system based on aesthetic composition evaluation using yohaku. In *2019 IEEE International Conference on Systems, Man and Cybernetics (SMC)*. IEEE, 101–106.
[60] George Leifman, Elizabeth Shtrom, and Ayellet Tal. 2016. Surface regions of interest for viewpoint selection. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 38, 12 (2016), 2544–2556.
[61] Chia-Ling Li, M. Pilar Aivar, Dmitry M. Kit, Matthew H. Tong, and Mary M. Hayhoe. 2016. Memory and visual search in naturalistic 2D and 3D environments. *Journal of Vision* 16, 8 (2016), 9–9.
[62] Zhi Li, Peter Moran, Qingyuan Dong, Ryan J. Shaw, and Kris Hauser. 2017. Development of a tele-nursing mobile manipulator for remote care-giving in quarantine areas. In *2017 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 3581–3586.
[63] Tsung-Chi Lin, Achyuthan Unni Krishnan, and Zhi Li. 2020. Shared autonomous interface for reducing physical effort in robot teleoperation via human motion mapping. In *2020 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 9157–9163.
[64] Tsung-Chi Lin, Achyuthan Unni Krishnan, and Zhi Li. 2021. How people use active telepresence cameras in telemanipulation. *To Appear in the 2021 IEEE International Conference on Robotics and Automation (ICRA)*.
[65] Dan Liu and Emanuel Todorov. 2007. Evidence for the flexible sensorimotor strategies predicted by optimal feedback control. *Journal of Neuroscience* 27, 35 (2007), 9354–9368.
[66] Rakshith Lokesh and Rajiv Ranganathan. 2020. Haptic assistance that restricts the use of redundant solutions is detrimental to motor learning. *IEEE Transactions on Neural Systems and Rehabilitation Engineering* (2020).
[67] Kristen L. Macuga, Andrew C. Beall, Roy S. Smith, and Jack M. Loomis. 2019. Visual control of steering in curve driving. *Journal of Vision* 19, 5 (2019), 1–1.
[68] Jonathan Samir Matthis and Brett R. Fajen. 2013. Humans exploit the biomechanics of bipedal gait during visually guided walking over complex terrain. *Proceedings of the Royal Society B: Biological Sciences* 280, 1762 (2013), 20130700.
[69] Vidhya Navalpakkam, Christof Koch, Antonio Rangel, and Pietro Perona. 2010. Optimal reward harvesting in complex perceptual environments. *Proceedings of the National Academy of Sciences* 107, 11 (2010), 5232–5237.
[70] José A. Navia, Matt Dicks, John van der Kamp, and Luis M. Ruiz. 2017. Gaze control during interceptive actions with different spatiotemporal demands. *Journal of Experimental Psychology: Human Perception and Performance* 43, 4 (2017), 783.
[71] Tuan Nghia Nguyen, Hung T. Nguyen, et al. 2016. Real-time video streaming with multi-camera for a telepresence wheelchair. In *2016 14th International Conference on Control, Automation, Robotics and Vision (ICARCV)*. IEEE, 1–5.
[72] Davide Nicolis, Marco Palumbo, Andrea Maria Zanchettin, and Paolo Rocco. 2018. Occlusion-free visual servoing for the shared autonomy teleoperation of dual-arm robots. *IEEE Robotics and Automation Letters* 3, 2 (2018), 796–803.
[73] Curtis W. Nielsen, Michael A. Goodrich, and Robert W. Ricks. 2007. Ecological interfaces for improving mobile robot teleoperation. *IEEE Transactions on Robotics* 23, 5 (2007), 927–941.
[74] Timothy Patten, Michael Zillich, Robert Fitch, Markus Vincze, and Salah Sukkarieh. 2015. Viewpoint evaluation for online 3-D active object classification. *IEEE Robotics and Automation Letters* 1, 1 (2015), 73–81.
[75] Lorenzo Peppoloni, Filippo Brizzi, Emanuele Ruffaldi, and Carlo Alberto Avizzano. 2015. Augmented reality-aided tele-presence system for robot manipulation in industrial manufacturing. In *Proceedings of the 21st ACM Symposium on Virtual Reality Software and Technology*. 237–240.
[76] Thammathip Piumsomboon, Arindam Day, Barrett Ens, Youngho Lee, Gun Lee, and Mark Billinghurst. 2017. Exploring enhancements for remote mixed reality collaboration. In *SIGGRAPH Asia 2017 Mobile Graphics & Interactive Applications*. 1–5.

[77] 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*. 432–440.

[78] Sina Radmard, A. Jung Moon, and Elizabeth A. Croft. 2019. Impacts of visual occlusion and its resolution in robot-mediated social collaborations. *International Journal of Social Robotics* 11, 1 (2019), 105–121.

[79] Daniel Rakita, Bilge Mutlu, and Michael Gleicher. 2018. An autonomous dynamic camera method for effective remote teleoperation. In *Proceedings of the 2018 ACM/IEEE International Conference on Human-Robot Interaction*. 325–333.

[80] Daniel Rakita, Bilge Mutlu, and Michael Gleicher. 2019. Remote telemanipulation with adapting viewpoints in visually complex environments. *Robotics: Science and Systems XV* (2019).

[81] Rob Reilink, Gart de Bruin, Michel Franken, Massimo A. Mariani, Sarthak Misra, and Stefano Stramigioli. 2010. Endoscopic camera control by head movements for thoracic surgery. In *2010 3rd IEEE RAS & EMBS International Conference on Biomedical Robotics and Biomechatronics*. IEEE, 510–515.

[82] Evgeny Rezunenko, Kasper van der El, Daan M. Pool, Marinus M. van Paassen, and Max Mulder. 2018. Relating human gaze and manual control behavior in preview tracking tasks with spatial occlusion. In *2018 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*. IEEE, 3440–3445.

[83] Pollen Robotics. 2021. *A New Mobile Base Offers Seamless and Self-detecting Navigation to the Robot*. https://www.pollen-robotics.com/.

[84] Eefje G. J. Roelofsen, Jurjen Bosga, David A. Rosenbaum, Maria W. G. Nijhuis-van der Sanden, Wim Hullegie, Robert van Cingel, and Ruud G. J. Meulenbroek. 2016. Haptic feedback helps bipedal coordination. *Experimental Brain Research* 234, 10 (2016), 2869–2881.

[85] Alessandro Roncone, Ugo Pattacini, Giorgio Metta, and Lorenzo Natale. 2016. A cartesian 6-DoF gaze controller for humanoid robots. In *Robotics: Science and Systems*, Vol. 2016.

[86] Nina Rudigkeit and Marion Gebhard. 2019. AMiCUS-a head motion-based interface for control of an assistive robot. *Sensors* 19, 12 (2019), 2836.

[87] Juan Sandoval, Med Amine Laribi, and Saïd Zeghloul. 2019. Autonomous robot-assistant camera holder for minimally invasive surgery. In *IFToMM International Symposium on Robotics and Mechatronics*. Springer, 465–472.

[88] Akanksha Saran, Branka Lakic, Srinjoy Majumdar, Juergen Hess, and Scott Niekum. 2017. Viewpoint selection for visual failure detection. In *2017 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 5437–5444.

[89] Alexander C. Schütz, Julia Trommershäuser, and Karl R. Gegenfurtner. 2012. Dynamic integration of information about salience and value for saccadic eye movements. *Proceedings of the National Academy of Sciences* 109, 19 (2012), 7547–7552.

[90] Ali Sengül, Giulio Rognini, Michiel van Elk, Jane Elizabeth Aspell, Hannes Bleuler, and Olaf Blanke. 2013. Force feedback facilitates multisensory integration during robotic tool use. *Experimental Brain Research* 227, 4 (2013), 497–507.

[91] Stela H. Seo, Daniel J. Rea, Joel Wiebe, and James E. Young. 2017. Monocle: Interactive detail-in-context using two pan-and-tilt cameras to improve teleoperation effectiveness. In *2017 26th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)*. IEEE, 962–967.

[92] Ali Shafti, Pavel Orlov, and A. Aldo Faisal. 2019. Gaze-based, context-aware robotic system for assisted reaching and grasping. In *2019 International Conference on Robotics and Automation (ICRA)*. IEEE, 863–869.

[93] Roland Sigrist, Georg Rauter, Robert Riener, and Peter Wolf. 2013. Augmented visual, auditory, haptic, and multimodal feedback in motor learning: A review. *Psychonomic Bulletin & Review* 20, 1 (2013), 21–53.

[94] Stephan P. Swinnen, Yong Li, Nicole Wenderoth, Natalia Dounskaia, Winston Byblow, Cathy Stinear, and Johan Wagemans. 2004. Perception-action coupling during bimanual coordination: The role of visual perception in the coalition of constraints that govern bimanual action. *Journal of Motor Behavior* 36, 4 (2004), 394–398.

[95] Arpád Takács, Dénes Ákos Nagy, Imre Rudas, and Tamás Haidegger. 2016. Origins of surgical robotics: From space to the operating room. *Acta Polytechnica Hungarica* 13, 1 (2016), 13–30.

[96] Yubo Tao, Qirui Wang, Wei Chen, Yingcai Wu, and Hai Lin. 2016. Similarity voting based viewpoint selection for volumes. In *Computer Graphics Forum*, Vol. 35. Wiley Online Library, 391–400.

[97] Benjamin W. Tatler, Mary M. Hayhoe, Michael F. Land, and Dana H. Ballard. 2011. Eye guidance in natural vision: Reinterpreting salience. *Journal of Vision* 11, 5 (2011), 5–5.

[98] Matsya R. Thulasiram, Ryan W. Langridge, Hana H. Abbas, and Jonathan J. Marotta. 2020. Eye–hand coordination in reaching and grasping vertically moving targets. *Experimental Brain Research* 238 (2020), 1433–1440.

[99] Matthew H. Tong, Oran Zohar, and Mary M. Hayhoe. 2017. Control of gaze while walking: Task structure, reward, and uncertainty. *Journal of Vision* 17, 1 (2017), 28–28.

[100] Hans A. Trukenbrod, Simon Barthelmé, Felix A. Wichmann, and Ralf Engbert. 2019. Spatial statistics for gaze patterns in scene viewing: Effects of repeated viewing. *Journal of Vision* 19, 6 (2019), 5–5.

[101] Alexandra Valiton and Zhi Li. 2020. Perception-action coupling in usage of telepresence cameras. In *2020 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 3846–3852.

[102] Samuel J. Vine, Rich S. W. Masters, John S. McGrath, Elizabeth Bright, and Mark R. Wilson. 2012. Cheating experience: Guiding novices to adopt the gaze strategies of experts expedites the learning of technical laparoscopic skills. *Surgery* 152, 1 (2012), 32–40.

[103] Robert Volcic and Ivan Camponogara. 2018. How do vision and haptics combine in multisensory grasping? *Journal of Vision* 18, 10 (2018), 64–64.

[104] Chunhui Wang, Yu Tian, Shanguang Chen, Zhiqiang Tian, Ting Jiang, and Feng Du. 2014. Predicting performance in manually controlled rendezvous and docking through spatial abilities. *Advances in Space Research* 53, 2 (2014), 362–369.

[105] David Whitney, Eric Rosen, Daniel Ullman, Elizabeth Phillips, and Stefanie Tellex. 2018. ROS reality: A virtual reality framework using consumer-grade hardware for ROS-enabled robots. In *2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 1–9.

[106] Maarten W. A. Wijntjes, Robert Volcic, Sylvia C. Pont, Jan J. Koenderink, and Astrid M. L. Kappers. 2009. Haptic perception disambiguates visual perception of 3D shape. *Experimental Brain Research* 193, 4 (2009), 639–644.

[107] Daniel M. Wolpert and Michael S. Landy. 2012. Motor control is decision-making. *Current Opinion in Neurobiology* 22, 6 (2012), 996–1003.

[108] Julia L. Wright, Jessie Y. C. Chen, and Michael J. Barnes. 2018. Human–automation interaction for multiple robot control: The effect of varying automation assistance and individual differences on operator performance. *Ergonomics* 61, 8 (2018), 1033–1045.

[109] Linfei Xiong, Chin Boon Chng, Chee Kong Chui, Peiwu Yu, and Yao Li. 2017. Shared control of a medical robot with haptic guidance. *International Journal of Computer Assisted Radiology and Surgery* 12, 1 (2017), 137–147.

[110] Guang Yang, Shuoyu Wang, Junyou Yang, and Bo Shen. 2018. Viewpoint selection strategy for a life support robot. In *2018 IEEE International Conference on Intelligence and Safety for Robotics (ISR)*. IEEE, 82–87.

[111] Alfred B. Yu and Jeffrey M. Zacks. 2017. Transformations and representations supporting spatial perspective taking. *Spatial Cognition & Computation* 17, 4 (2017), 304–337. https://doi.org/10.1080/13875868.2017.1322596

[112] Huaiyong Zhao, Dominik Straub, and Constantin A. Rothkopf. 2019. The visual control of interceptive steering: How do people steer a car to intercept a moving target? *Journal of Vision* 19, 14 (2019), 11–11.

Received 9 April 2021; revised 24 July 2022; accepted 3 November 2022