How People Use Active Telepresence Cameras in Tele-manipulation

Tsung-Chi Lin, Achyuthan Unni Krishnan and Zhi Li[1]

***Abstract*— Robot teleoperation is a reliable way to perform a variety of tasks with complex robotic systems. However, the remote control of active telepresence cameras on the robot for improved telepresence adds an additional degree of complexity while teleoperating and can thus affects the operator's performance during tele-manipulation. Our previous user study (N=16) investigates the general human performance and preference when using various wearable cameras. In this paper, we further investigate how humans adapt to the usage of telepresence cameras in terms of motion behavior. The findings from our human motion analysis inform several desired designs for robot teleoperation interfaces and assistive autonomy.**

I. Introduction

Teleoperation enables many complex robotic platforms (e.g., multi-manipulator surgical robots [1], humanoid nursing robots [2]) to perform tasks beyond the capabilities of robot autonomy, such as dexterous manipulation and high-speed navigation in dynamic, cluttered, human environments. Most of the time, the performance of many freeform tele-manipulation heavily relies upon a teleoperator's effective usage of remote telepresence cameras (see Fig. 1). As a result, efficient and intuitive interfaces are necessary not only for complex tele-action control but also for the control of active telepresence. For instance, contemporary mobile manipulators and humanoid robots (e.g., assisting humans in outer space [3], under water [4], nursing [5], manufacturing [6] and daily living tasks [7]) have several telepresence cameras, attached to head, torso, hands, base of the robot, in order to provide sufficient perception capabilities to enable robot autonomy and improve the teleoperator's situational awareness. The remote control of active telepresence cameras is difficult, because the humans develop novel motor skills to control the "alien eyes" on the robots, which are largely different from human eyes in their displacements, motion capabilities, and field of view (FOV). These elements are counter-intuitive to their intuitive, natural motions for gaze control. Learning how to use these remote telepresence cameras is critical to the development of robot teleoperation skills and therefore attracts many research efforts. For instance, prior research has extensively studied the challenges in laparoscopic camera control [8], effects of various human factors on performance [9], methods for autonomous camera assistance [10] and paradigms for surgeon training [11]. As humans and robotic systems synergize more in the future of work, robot teleoperation will become a necessary skills in many work domains. Understanding how humans learn to use the various active telepresence cameras will provide insights to the design of teleoperation interfaces and camera assistance autonomy.

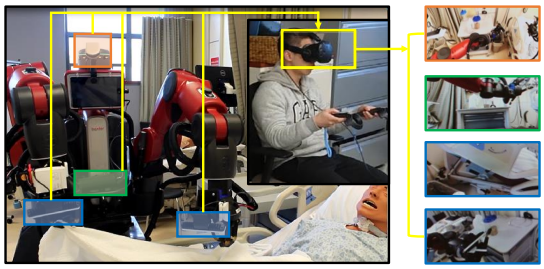

Fig. 1: Nursing robot teleoperation via a freefrom interface with feedback from telepresence cameras attached to head, torso and wrists.

This paper will leverage a novel experimental paradigm developed in our prior work to reveal human behavior and preference in the usage of various active telepresence cameras. In this experimental paradigm, participants were provided with video streams from the cameras attached to their head, torso, and hands, and a standalone camera monitoring the workspace. They then performed a comprehensive dexterous manipulation task that involves reaching, grasping, moving and stacking of objects. Our prior work has conducted the preliminary examination of the overall human performance (task completion time and errors) when using different wearable cameras, and preference of cameras in various aspects of usability (e.g., efficiency, frustration, intuitiveness, etc.) [12]. We have observed some interesting human motions of head, torso and arms across participants, which motivated us to further evaluate several hypotheses regarding the perception-action coupling in the usage of active telepresence cameras. Our insights from this observation is that: 1) a human and a robot coupled via a teleoperation interface can be considered as a cyber-human system; 2) As humans practice robot teleoperation, a new perception-action coupling will be developed for the cyber-human system, so that humans can effectively select and control the remote cameras, for the purpose of performing tele-action (e.g., reaching, moving, grasping objects) or remote perception (e.g., visual searching and tracking). Based on this insight, this work will conduct a comprehensive motion analysis, to reveal some human motor control strategies that significantly influence: 1) the telepresence camera control interface, 2) the preference of camera control motion; and 3) the integration of limited visual and haptic feedback. The novelty of our work is to extend the research on perception-action coupling from human to cyber-human system. It demonstrated that our proposed experimental paradigm provides an effective approach to understand how humans adapt to the perception and motion capabilities of teleoperated robots, and informs

[1] Robotics Engineering, Worcester Polytechnic Institute, Worcester, MA 01609, USA {tlin2,aunnikrishnan,zli11}@wpi.edu

us on how to design interface and robot autonomy to facilitate this adaptation.

II. Related Work

Imagine you are an alien creature with eyes attached to the torso and hands, but not on the head. Your hands can only vaguely feel an object when touching. How would you use these novel eyes and hands to perceive and interact with your environment? Fortunately, we are confident that the human motor system is able to re-develop a "natural normal," to best utilize your new perception and action capabilities, as in motor skill training [13]–[15] and rehabilitation [16]–[18]. The human behavior and underlying human motor control strategies of the Vision-motion coupling [19], haptics-motion coupling [20], [21] and the integration of multi-sensory feedback [22]–[25] have been extensively investigated in various human motor skills. Research on human vision-motion coupling reveals that human gaze and visual attention control in daily activities can be influenced not only by the salient features [26] and surprising stimuli [27]in task environment, but also by the action and behavior goals [19] (and their associated intrinsic [28], [29] and explicit [30], [31] rewards), the benefits of collecting additional information to reduce the uncertainty in task environments [32]–[37], the memory of task-relevant objects or context cues in the environment [19], [38]–[40], and the predicted visual state in action control [41]–[52]. These findings lead effective computational models to explain, predict, and render human(-like) visual attention [53]. Moreover, frameworks such as probabilistic decision theory [19], [54]–[56], stochastic optimal control [57]–[60], and maximum likelihood estimation [25], [61]–[65] have been used to explain the vision-motion coupling of human motor control. Similarly, 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., [20], [21], [23], [61], [66], [67]. 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 [25], [61]–[65]. However, the research on perception-action coupling, from experimental human movement studies, to theoretical frameworks, to computational models, have not been extended to cyber-human systems that emerge due to recent advances in robotics. These include the humans and remote robotic systems coupled via (assisted) 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 [12]. We have observed very consistent behaviors of 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 telerobotic systems. We have also noticed humans attempt to leverage the limited available haptic feedback to compensate for the remote perception issues (e.g.,lost of depth information, limited field of view, etc), which implies the strategies for multi-sensory integration. Following the preliminary work, this paper will further analyze these observed human behaviors to reveal the perception-motion coupling in a novel context, and discuss their implications to the design of telerobotic interfaces.

III. Experimental Paradigm

Fig. 2: Experimental paradigm.

Our prior work [12] devised a novel experimental paradigm to investigate the perception-motion coordination in the usage of active telepresence. Shown in Fig. 2, human participants are asked to perform dexterous manipulation tasks (e.g., stacking light-weight plastic cups into a pyramid), with the video stream from the wearable cameras attached to their head, torso, hands, and the standalone camera in the workspace. Specifically, the *Head Camera* (C_H) was attached to the front of the VR headset, matching natural human eyesight; The *Clavicle Camera* (C_C) was attached to the chest above the sternum and between the underarms; The *Action Camera* (C_A) and *Perception Camera* (C_P) were attached to the participant's dominant and non-dominant hands respectively. A *Workspace Camera* (C_W) was set up across the workspace from the participant on a stationary tripod. The cameras in the experimental paradigm are chosen to simulate the configuration and mobility of representative active telepresence cameras commonly used in robot teleoperation, such as the pan-and-tilt or fixed cameras attached to humanoid robot head or torso/base [68], additional camera arms that track the end-effector of the manipulator arm or/and task features (often used in telerobotic surgery [10]), the eye-in-hand cameras attached to the end-effectors of manipulator arms [69], and the standalone cameras for supervising the robot workspace with a fixed viewpoint [70]. The selection of cameras in the proposed experimental paradigm also matches the cameras available to many contemporary commercial and prototype mobile humanoid nursing robots [5], [71], so that the findings from the experiments can inform the design of interfaces and assistive autonomy for nursing robots teleoperation.

A. Human Movement Study and Analysis

Our prior work conducted a user study using the proposed experimental paradigm to observe: 1) how humans perform gross manipulations (e.g., reaching, moving) and precise manipulations (e.g., grasping, stacking) using each active telepresence cameras, and 2) what cameras human prefer to use to perform different manipulation actions. During the experiment, the participants wore thick gloves to minimize their dependence on precise haptic feedback and a wireless microphone to switch the cameras using voice commands.

For each camera, a participant first practiced a cup stacking task to get familiar with the selected camera view, and then performed the cup stacking task at their comfortable pace with this camera for skill evaluation. In addition to the *selected camera trial* for each camera, the participant also performed a *mixed camera trial* in which they were allowed to switch the camera view during the cup stacking (see the detailed experiment procedure in [12]).

While the natural behavior observed in the freeform usage of active telepresence cameras are complicated to model (in terms of regular motion patterns or action sequences), our prior work did reveal some object manipulation and camera motion control behaviors demonstrated by all the participants, which implies that there may exist a perception-motion coupling strategy generally preferred by human motor control. To reveal these strategies, we conducted a comprehensive human movement analysis, followed by a participant interview, to answer two research questions: **RQ1** - Are there any consistent perception-action coupling strategies naturally preferred by human motor control in the usage of active telepresence cameras? **RQ2** - How do these perception-action coupling strategies influence robot teleoperation interfaces? To this end, we annotated the recorded video of human movements from the user study, to identify the distinguishable movement features that are largely different from how humans perform the same manipulation tasks with their eyes. These distinguishable motion features were observed in the head movements, torso movements, uni-manual and bi-manual motions for object manipulation and camera control, and physical contacts with manipulated objects and environments (with limited haptic feedback). Among all the distinguishable motion features, we identify the motions observed in all the participants, which reveals the consistent natural preference of human motor control in the usage of active telepresence. These include:

- **Head Movement**: The participant instinctively moves their head to adjust the camera view, even if the camera is not attached to their head. Some participants are aware that when the camera is not attached to their head, moving their head cannot change the camera viewpoint, while others tend to forget this. In both cases, participants are frustrated about not being able to change the view using head movement and perceive more mental workload and physical discomfort (because of uncomfortable posture of head and neck). As a result we counted the instances of head movements like turning the head around, moving the head vertically and sideways.
- **Arm Fixation**: When using the perception camera (C_P), participants always fixed their elbow joints, and mostly adjusted the camera viewpoint by turning their torsos. We measured the total time the participant held a stationary arm pose in fixed camera trial for C_P.
- **Bimanual Motion**: Whenever possible, participants used both hands for object manipulation to improve the task efficiency. These motions are observed in the usage of head (C_H), clavicle (C_C) and workspace cameras (C_w). We counted the instances when both hands were used to gather and stack cups in the fixed camera trial for head, clavicle and worksapce cameras.
- **Touch to Locate**: Even with limited haptic feedback, participants still frequently touch to identify and confirm the cup locations. For this motion, we counted the number of times a hand was used to tap the bottom of the cup in fixed camera trial for each camera.
- **Tentative Stacking**: Stacking cups on each other requires the most precise manipulation among all task components. As a result, participants always tap the cup to be stacked before carefully placing another cup on top of it. We labelled and counted instances when participants held a cup in hand to tentatively stack the cup on others.
- **Slide Cup on Table**: Whenever possible, participants slide the cup in hand on the table to move it, instead of picking it and we counted all such instances.

B. Participant Survey and Interview

Our prior work had conducted a post-study survey, with NASA-TLX and task-specific questions, to understand the perceived performance, mental workload, frustration, situational awareness, and the ease to perform gross and fine manipulation. Based on the preference (in ranking) for the active telepresence cameras inferred from the survey feedback, we further conducted an interview with each participant to identify the causing factors and the extent of their influence. Participant interviews are useful for connecting interface characteristics to the specific aspects of the interface usability [72], [73] but are not well-adopted in the evaluation of robot teleoperation interface and assistive autonomy design. In the interview, we presented the recorded video of the *selected camera trial* and *mixed camera trial* to the participant, and discussed about the distinguishable movement features they had demonstrated. For each distinguishable movement feature, we asked 1) if the participants was aware of the way they performed the movement, 2) why they performed the motions in this way, and 3) how they feel about the performance (e.g., efficiency, mental/physical workload, frustration, situational awareness of task-relevant and performance-critical features). In addition, we also collected detailed information about their experience in STEM education, robots, gaming, and virtual reality system. We have also asked the participants 1) how they compared the experience of using active telepresence camera with their experience in their daily manipulation activities, robot teleoperation, gaming, and virtual reality systems; and 2) Based on the experience of using these active telepresence cameras, what their idea of an efficient and intuitive camera control interface in teleoperation would be. The interview took about 1-1.5 hours for each participant and provided us an in-depth understanding of the participant's behavior.

IV. Results

A. Instinctive Head Movements

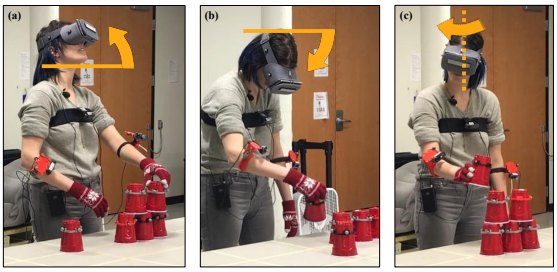

Fig. 3: Compulsive head movement: (a) raise head up; (b) hold head down; (c) turn head side way.

We noticed that people attempt to adjust the camera view using their head even for the action, perception and clavicle cameras (see the head posture in Fig. 3). The analysis of the head movement shows that using action camera causes more frequent ($p < 0.01$) head motion than the clavicle and perception cameras. Fig. 4 shows the instances of head movements with respect to the task completion time in the fixed camera trials for the clavicle, perception and action cameras. We examine the correlation between task performance (completion time) and the instances of head movements. 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. The linear regression predicts that the expected task completion time increases by approximately 9.5 (clavicle), 4.6 (perception) and 4.7 (action) seconds for each instance 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. Additionally, some participants are 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 useless for camera viewpoint control until they felt discomfort. Overall, we found that it is more difficult for the participants to realize and suppress the instinctive head movements, when the camera is more difficult to use. What the interview has revealed is also consistent with our survey results, which shows that the action camera resulted in the highest effort, frustration level and mental demand and lowest awareness of hands and cups. followed by clavicle and perception cameras.

B. Control of the Perception Hand Camera

In the usage of the perception hand camera (C_P), we found that the participants tend to fix their elbow joints when moving the perception camera around (see the arm posture in Fig. 5 a). Fig. 5 b shows the proportion of time that participants adopted a fixed arm posture with respect to the total task completion time in the fixed camera trial for the perception camera (82.9 $\pm$ 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

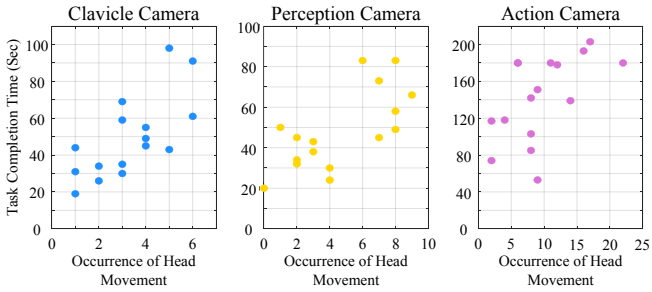

Fig. 4: Task completion time vs the instances of head movement for the clavicle, perception and action hand cameras.

camera viewpoint, which limits 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 so that they can better remember the spatial relationship of the perception hand 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.

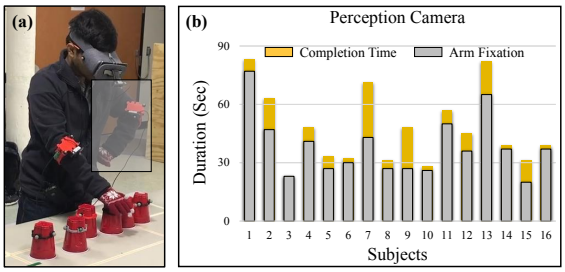

Fig. 5: (a) The fixed elbow pose for perception camera control; (b) Duration of the arm fixation w.r.t. task completion in perception camera usage.

C. Bimanual manipulation

Whenever possible, participants preferred bimanual manipulation, to speed up the task and to increase their reaching ranges without moving the body. The usage of bimanual manipulation, in both symmetric and asymmetric forms, are observed in the head, clavicle and workspace cameras, for reaching to collect cups (Fig. 6 a), and for placing/stacking the cups in the same row (Fig. 6 b). We also found that bimanual control 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 difficult 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 efficient yet more complex to plan. The interview feedback is consistent with our survey results, which shows that when using the clavicle and workspace cameras, participants have

higher cognitive workload and worse situational awareness, which leads to less cognitive bandwidth to consider bimanual manipulations.

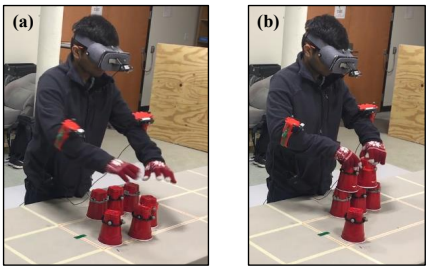

Fig. 6: Bimanual manipulation for (a) reaching, (b) stacking.

D. Dependence on Haptic Perception

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 (cups), 2) touching to feel the cup when tentatively stacking the cup, and 3) sliding the cup on the table so that they can leverage the haptic perception of table constraints to better control the moving motions. Fig. **??** shows the mean and standard deviation of touch-to-locate instances 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 least ($p < 0.01$) when using the head camera. These significant differences indicates that participant resort more to haptic feedback for the cameras more difficult 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; 2) the haptic feedback does not have to be strong and realistic if it can provide a sense of contact. This can largely reduce the mental workload and stress due to uncertainty in perception, while improving the task accuracy and efficiency.

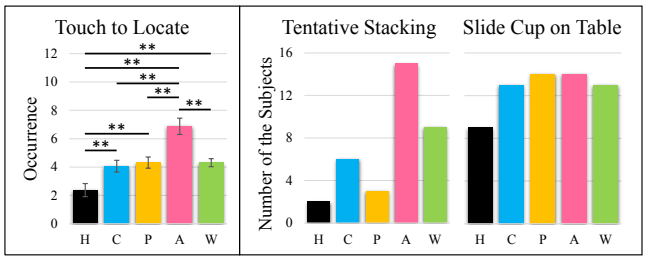

Fig. 7: Touching-to-locate, tentative-stacking and sliding cups-on-table actions observed in the usage of the head (H), clavicle (C), perception (P), action (A) and workspace (W) camera.

In addition to touch-to-locate, participants also used touch-to-feel when tentatively stacking, and sliding the cups on the table. Overall, the tentative stacking actions are more observed with the cameras identified as non-intuitive and inefficient to use. In Fig. 7 (right), the tentative stacking actions are used by 15 of 16 participants when working with the action hand camera, and by 2 of 16 participants when working with the head camera. On the other hand, sliding cups on table are observed in more than half of the participants for all the cameras. The interview feedback reveals that: 1) The gloves effectively damped most of their haptic perception; 2) the limited tactile sensing is still very helpful to the task in many cases.

V. The Implication to Assisted Teleoperation Interface Design

The findings from our human movement studies imply several design principles for robot teleoperation interfaces and assistive autonomy so that the coordinated control of remote perception and actions will comply with natural behavior and preference of human motor control. Although these principles have been practiced in the state-of-the-art robot teleoperation interface implementations to some extent, research thus far hasn't been able to reveal the human motor control strategies that make the interface design choices to be effective or provide a framework of generalizable design philosophy for the various interface hardware and robots. Our discussion here aims to bridge the gap and make an effort to extend the understanding of perception-action coupling from humans to closely coupled human-robot systems.

A. Head Control for Primary Viewpoints

Egocentric viewpoint control (using head and gaze) for remote cameras are not new to the design of tele-robotic interfaces [74]. As the human tracking technologies become more accurate, portable and affordable, head- and gaze-control have also become more adopted for the eye-in-hand cameras of manipulator and continuum robots [75], and the head camera of mobile and humanoid robots [76], [77]. While matching human eyes to robot eyes is considered to be a natural design choice, it is also not rare to see the remote cameras controlled by hands. When multiple cameras are available (as on many commercial and prototype humanoid robot platforms [5], [71]), 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 mostly rely upon to perform the task) from the head to hand camera, adapting to the hand control of camera viewpoint always leads to interruption of task performance. Lessons learned from (tele-robotic) laparoscopic surgery training also indicate that it takes much more training efforts to learn to use hand-controlled cameras [78]. The instinctive head movement we observed in the usage of clavicle, perception and action hand cameras implies that egocentric control (using head or gaze) should be adopted to control the viewpoint not only from the robot head camera, but from *any camera selected to be the primary viewpoint*. Designs that contradict this strong human

instinct will lead to high cognitive workload, frustration, physical discomfort and efforts of training.

B. Spatial Awareness of Camera Pose

Do people need to know the camera pose with respect to the robot, end-effector, manipulated object and workspace? Tele-robotic interfaces that adopted fully autonomous camera control assumes that the teleoperators do not need to remember how the camera has been moved, understand how the camera viewpoint is selected, and predict how the camera can be moved to the next desired pose. In direct teleoperation, understanding the camera pose and motions is critical to control of the robot action components (e.g., end-effectors, mobile base). Even in supervisory control, lack of the spatial awareness of camera pose will reduce the operator's situational awareness and capability to intervene if the robot autonomy is not reliable [79]. Our study reveals the strategy people used to maintain the spatial awareness of camera pose: by fixing the elbow joint and limiting the shoulder motions, the perception hand camera can only be controlled by simple translation or rotation motions with respect to the coordinate frame that the participants are very familiar with and frequently used to for motion planning (e.g., the coordinate frame attached to the torso). This observation implies two interface design principle to improve the spatial awareness of remote camera pose: 1) Similar to the constrained positioning and point-to-click interface for precise grasping orientation control [80], for a high-mobility camera (attached to a manipulator's arm), the interface should limit the degrees-of-freedom that a teleoperator can simultaneously manipulate to adjust camera viewpoint; 2) On the other hand, it is preferred to compose the motions of autonomous camera viewpoint using simple translation and rotation motion primitives, to make it easier to understand and predict the autonomous camera motions.

C. Preference of Bimanual Operation

Our study reveals that when the camera is intuitive to use, people have more cognitive bandwidth for planning complex motions. Human motor control naturally prefers motion symmetry (in synchronous or anti-phase motions), due to the inter-hemispheric coupling effects [81]. In the usage of active telepresence cameras, the preference of symmetric bimanual motions still exists, and could be leveraged to improve task efficiency and accuracy. Comparing the bimanual operation when using the head and clavicle cameras, we found that it is preferred to ensure the perceptive action control will not interfere with the manipulation action control. In many contemporary tele-robotic interfaces, robot manipulation and mobility actions unavoidably affect the active telepresence camera viewpoints, because the remote cameras are attached to the robot base or the end-effector. The findings from our study suggest that whenever possible, we should select the camera for the primary viewpoint to be the one that has the least control inference with robot actions.

D. The Need for Visuo-Haptic Sensory Integration

Our study reveals that people actively resort to every possible haptic feedback, to compensate for the lost depth information of the visual feedback via active telepresence. The desire for haptic feedback is more prominent when performing precise manipulation. Indeed, human motor control has the instinct to pursue visuo-haptic sensory integration when they perform tasks with their own bodies [82] as well as via tele-robotic interfaces. Unfortunately, the state-of-the-art haptic feedback sending and rendering technologies cannot enable the teleoperation interface to provide the most realistic haptic perception. Will that be a problem? It actually depends on how much haptic feedback we need to compensate for the limitation of active telepresence visual feedback. 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; 2) for general purpose manipulation tasks, adding a little bit 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.

E. General Design Philosophy for Tele-robotic Interfaces

Inspired by findings from our study, we propose a philosophy for tele-robotic interface and assistive autonomy design: The ultimate goal for interface design is to facilitate human 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 [83]; 2) we may also **replace** haptic display with augmented reality visual display [84]; 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 [68].

VI. Conclusion and Future Work

This paper analyzed human motion behaviors from the user study performed primarily with visual feedback from various wearable cameras in a simulated telepresence setting. The results identify the participants preferred designs for the teleoperation interfaces and robot autonomy including: (1) head movement should be mapped to control the primary viewpoint of the active telepresence cameras; (2) interface should limit the degrees-of-freedom when adjusting the camera viewpoint and simplify the camera motions to a supervisory scenario to improve the spatial and situational awareness of camera pose; (3) haptic feedback is needed to compensate for the limited visual feedback and the complexity of the haptic sensation should be integrated with visual feedback. The user study was recorded along with the participant's view from the selected camera. Our future work will analyze the gaze allocation to investigate the vision for perception

and action. We will also investigate the learning effect to see if there are any consistent human behaviors that can be applied to normal usage.

REFERENCES

[1] B. S. Peters, P. R. Armijo, C. Krause, S. A. Choudhury, and D. Oleynikov, "Review of emerging surgical robotic technology," *Surgical endoscopy*, vol. 32, no. 4, pp. 1636–1655, 2018.

[2] E. Ackerman, "Toyota gets back into humanoid robots with new t-hr3," *IEEE Spectrum*, 2017.

[3] S. J. Jorgensen, M. W. Lanighan, S. S. Bertrand, A. Watson, J. S. Altemus, R. S. Askew, L. Bridgwater, B. Domingue, C. Kendrick, J. Lee *et al.*, "Deploying the nasa valkyrie humanoid for ied response: An initial approach and evaluation summary," in *2019 IEEE-RAS 19th International Conference on Humanoid Robots (Humanoids)*. IEEE, 2019, pp. 1–8.

[4] M. de la Cruz, G. Casañ, P. Sanz, and R. Marín, "Preliminary work on a virtual reality interface for the guidance of underwater robots," *Robotics*, vol. 9, no. 4, p. 81, 2020.

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

[6] J. I. Lipton, A. J. Fay, and D. Rus, "Baxter's homunculus: Virtual reality spaces for teleoperation in manufacturing," *IEEE Robotics and Automation Letters*, vol. 3, no. 1, pp. 179–186, 2017.

[7] D. Park, H. Kim, Y. Hoshi, Z. Erickson, A. Kapusta, and C. C. Kemp, "A multimodal execution monitor with anomaly classification for robot-assisted feeding," in *2017 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2017, pp. 5406–5413.

[8] M. S. Amin, A. Aydin, N. Abbud, B. Van Cleynenbreugel, D. Veneziano, B. Somani, A. S. Gözen, J. P. Redorta, M. S. Khan, P. Dasgupta *et al.*, "Evaluation of a remote-controlled laparoscopic camera holder for basic laparoscopic skills acquisition: a randomized controlled trial," *Surgical Endoscopy*, pp. 1–9, 2020.

[9] G. Zhang, J. P. Hansen, and K. Minakata, "Hand-and gaze-control of telepresence robots," in *Proceedings of the 11th ACM Symposium on Eye Tracking Research & Applications*, 2019, pp. 1–8.

[10] J. Sandoval, M. A. Laribi, and S. Zeghloul, "Autonomous robot-assistant camera holder for minimally invasive surgery," in *IFToMM International Symposium on Robotics and Mechatronics*. Springer, 2019, pp. 465–472.

[11] I. Rivas-Blanco, C. J. Perez-del Pulgar, C. López-Casado, E. Bauzano, and V. F. Muñoz, "Transferring know-how for an autonomous camera robotic assistant," *Electronics*, vol. 8, no. 2, p. 224, 2019.

[12] A. Valiton and Z. Li, "Perception-action coupling in usage of telepresence cameras," in *2020 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2020, pp. 3846–3852.

[13] B. I. Bertenthal, J. L. Rose, and D. L. Bai, "Perception–action coupling in the development of visual control of posture." *Journal of experimental psychology: human perception and performance*, vol. 23, no. 6, p. 1631, 1997.

[14] C. Craig, "Understanding perception and action in sport: how can virtual reality technology help?" *Sports Technology*, vol. 6, no. 4, pp. 161–169, 2013.

[15] T. Abe, N. Raison, N. Shinohara, M. S. Khan, K. Ahmed, and P. Dasgupta, "The effect of visual-spatial ability on the learning of robot-assisted surgical skills," *Journal of surgical education*, vol. 75, no. 2, pp. 458–464, 2018.

[16] C. K. Williams and H. Carnahan, "Motor learning perspectives on haptic training for the upper extremities," *IEEE transactions on haptics*, vol. 7, no. 2, pp. 240–250, 2014.

[17] A. Shafti, P. Orlov, and A. A. Faisal, "Gaze-based, context-aware robotic system for assisted reaching and grasping," in *2019 International Conference on Robotics and Automation (ICRA)*. IEEE, 2019, pp. 863–869.

[18] R. Lokesh and R. Ranganathan, "Haptic assistance that restricts the use of redundant solutions is detrimental to motor learning," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 2020.

[19] M. M. Hayhoe, "Vision and action," *Annual review of vision science*, vol. 3, pp. 389–413, 2017.

[20] E. G. Roelofsen, J. Bosga, D. A. Rosenbaum, M. W. Nijhuis-van der Sanden, W. Hullegie, R. van Cingel, and R. G. Meulenbroek, "Haptic feedback helps bipedal coordination," *Experimental brain research*, vol. 234, no. 10, pp. 2869–2881, 2016.

[21] S. Monaco, G. Króliczak, D. J. Quinlan, P. Fattori, C. Galletti, M. A. Goodale, and J. C. Culham, "Contribution of visual and proprioceptive information to the precision of reaching movements," *Experimental brain research*, vol. 202, no. 1, pp. 15–32, 2010.

[22] S. Serwe, K. P. Körding, and J. Trommershäuser, "Visual-haptic cue integration with spatial and temporal disparity during pointing movements," *Experimental brain research*, vol. 210, no. 1, pp. 67–80, 2011.

[23] R. Sigrist, G. Rauter, R. Riener, and P. Wolf, "Augmented visual, auditory, haptic, and multimodal feedback in motor learning: a review," *Psychonomic bulletin & review*, vol. 20, no. 1, pp. 21–53, 2013.

[24] G. Desmarais, M. Meade, T. Wells, and M. Nadeau, "Visuo-haptic integration in object identification using novel objects," *Attention, Perception, & Psychophysics*, vol. 79, no. 8, pp. 2478–2498, 2017.

[25] T. L. Gibo, W. Mugge, and D. A. Abbink, "Trust in haptic assistance: weighting visual and haptic cues based on error history," *Experimental Brain Research*, vol. 235, no. 8, pp. 2533–2546, 2017.

[26] B. W. Tatler, M. M. Hayhoe, M. F. Land, and D. H. Ballard, "Eye guidance in natural vision: Reinterpreting salience," *Journal of vision*, vol. 11, no. 5, pp. 5–5, 2011.

[27] L. Itti and P. Baldi, "Bayesian surprise attracts human attention," *Vision research*, vol. 49, no. 10, pp. 1295–1306, 2009.

[28] J. Jovancevic-Misic and M. Hayhoe, "Adaptive gaze control in natural environments," *Journal of Neuroscience*, vol. 29, no. 19, pp. 6234–6238, 2009.

[29] J. S. Matthis and B. R. Fajen, "Humans exploit the biomechanics of bipedal gait during visually guided walking over complex terrain," *Proceedings of the Royal Society B: Biological Sciences*, vol. 280, no. 1762, p. 20130700, 2013.

[30] V. Navalpakkam, C. Koch, A. Rangel, and P. Perona, "Optimal reward harvesting in complex perceptual environments," *Proceedings of the National Academy of Sciences*, vol. 107, no. 11, pp. 5232–5237, 2010.

[31] A. C. Schütz, J. Trommershäuser, and K. R. Gegenfurtner, "Dynamic integration of information about salience and value for saccadic eye movements," *Proceedings of the National Academy of Sciences*, vol. 109, no. 19, pp. 7547–7552, 2012.

[32] J. Najemnik and W. S. Geisler, "Optimal eye movement strategies in visual search," *Nature*, vol. 434, no. 7031, pp. 387–391, 2005.

[33] ——, "Eye movement statistics in humans are consistent with an optimal search strategy," *Journal of Vision*, vol. 8, no. 3, pp. 4–4, 2008.

[34] B. T. Sullivan, L. Johnson, C. A. Rothkopf, D. Ballard, and M. Hayhoe, "The role of uncertainty and reward on eye movements in a virtual driving task," *Journal of vision*, vol. 12, no. 13, pp. 19–19, 2012.

[35] M. H. Tong, O. Zohar, and M. M. Hayhoe, "Control of gaze while walking: task structure, reward, and uncertainty," *Journal of vision*, vol. 17, no. 1, pp. 28–28, 2017.

[36] P. Verghese, "Active search for multiple targets is inefficient," *Vision Research*, vol. 74, pp. 61–71, 2012.

[37] S. Ghahghaei and P. Verghese, "Efficient saccade planning requires time and clear choices," *Vision research*, vol. 113, pp. 125–136, 2015.

[38] J. R. Brockmole, M. S. Castelhano, and J. M. Henderson, "Contextual cueing in naturalistic scenes: Global and local contexts." *Journal of Experimental Psychology: Learning, Memory, and Cognition*, vol. 32, no. 4, p. 699, 2006.

[39] A. Hollingworth, "Two forms of scene memory guide visual search: Memory for scene context and memory for the binding of target object to scene location," *Visual Cognition*, vol. 17, no. 1-2, pp. 273–291, 2009.

[40] M. L.-H. Võ and J. M. Henderson, "The time course of initial scene processing for eye movement guidance in natural scene search," *Journal of Vision*, vol. 10, no. 3, pp. 14–14, 2010.

[41] M. F. Land and S. Furneaux, "The knowledge base of the oculomotor system," *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences*, vol. 352, no. 1358, pp. 1231–1239, 1997.

[42] M. F. Land and P. McLeod, "From eye movements to actions: how batsmen hit the ball," *Nature neuroscience*, vol. 3, no. 12, pp. 1340–1345, 2000.

[43] P. Han, D. R. Saunders, R. L. Woods, and G. Luo, "Trajectory

prediction of saccadic eye movements using a compressed exponential model," *Journal of vision*, vol. 13, no. 8, pp. 27–27, 2013.

[44] G. Diaz, J. Cooper, C. Rothkopf, and M. Hayhoe, "Saccades to future ball location reveal memory-based prediction in a virtual-reality interception task," *Journal of vision*, vol. 13, no. 1, pp. 20–20, 2013.

[45] K. P. Körding and D. M. Wolpert, "Bayesian integration in sensorimotor learning," *Nature*, vol. 427, no. 6971, pp. 244–247, 2004.

[46] H. Tassinari, T. E. Hudson, and M. S. Landy, "Combining priors and noisy visual cues in a rapid pointing task," *Journal of Neuroscience*, vol. 26, no. 40, pp. 10 154–10 163, 2006.

[47] G. Diaz, J. Cooper, and M. Hayhoe, "Memory and prediction in natural gaze control," *Philosophical Transactions of the Royal Society B: Biological Sciences*, vol. 368, no. 1628, p. 20130064, 2013.

[48] J. Fukushima, T. Akao, S. Kurkin, C. R. Kaneko, and K. Fukushima, "The vestibular-related frontal cortex and its role in smooth-pursuit eye movements and vestibular-pursuit interactions," *Journal of Vestibular Research*, vol. 16, no. 1, 2, pp. 1–22, 2006.

[49] V. P. Ferrera and A. Barborica, "Internally generated error signals in monkey frontal eye field during an inferred motion task," *Journal of Neuroscience*, vol. 30, no. 35, pp. 11 612–11 623, 2010.

[50] T. Nyffeler, S. Rivaud-Pechoux, N. Wattiez, and B. Gaymard, "Involvement of the supplementary eye field in oculomotor predictive behavior," *Journal of Cognitive Neuroscience*, vol. 20, no. 9, pp. 1583–1594, 2008.

[51] N. Shichinohe, T. Akao, S. Kurkin, J. Fukushima, C. R. Kaneko, and K. Fukushima, "Memory and decision making in the frontal cortex during visual motion processing for smooth pursuit eye movements," *Neuron*, vol. 62, no. 5, pp. 717–732, 2009.

[52] J. A. Assad and J. H. Maunsell, "Neuronal correlates of inferred motion in primate posterior parietal cortex," *Nature*, vol. 373, no. 6514, pp. 518–521, 1995.

[53] A. Borji and L. Itti, "State-of-the-art in visual attention modeling," *IEEE transactions on pattern analysis and machine intelligence*, vol. 35, no. 1, pp. 185–207, 2012.

[54] L. T. Maloney and H. Zhang, "Decision-theoretic models of visual perception and action," *Vision research*, vol. 50, no. 23, pp. 2362–2374, 2010.

[55] D. W. Franklin and D. M. Wolpert, "Computational mechanisms of sensorimotor control," *Neuron*, vol. 72, no. 3, pp. 425–442, 2011.

[56] D. M. Wolpert and M. S. Landy, "Motor control is decision-making," *Current opinion in neurobiology*, vol. 22, no. 6, pp. 996–1003, 2012.

[57] D. Liu and E. Todorov, "Evidence for the flexible sensorimotor strategies predicted by optimal feedback control," *Journal of Neuroscience*, vol. 27, no. 35, pp. 9354–9368, 2007.

[58] E. Todorov and M. I. Jordan, "Optimal feedback control as a theory of motor coordination," *Nature neuroscience*, vol. 5, no. 11, pp. 1226–1235, 2002.

[59] D. C. Knill, A. Bondada, and M. Chhabra, "Flexible, task-dependent use of sensory feedback to control hand movements," *Journal of Neuroscience*, vol. 31, no. 4, pp. 1219–1237, 2011.

[60] M. L. Latash, J. P. Scholz, and G. Schöner, "Motor control strategies revealed in the structure of motor variability," *Exercise and sport sciences reviews*, vol. 30, no. 1, pp. 26–31, 2002.

[61] A. M. Wing, M. Doumas, and A. E. Welchman, "Combining multisensory temporal information for movement synchronisation," *Experimental brain research*, vol. 200, no. 3-4, pp. 277–282, 2010.

[62] R. Volcic and I. Camponogara, "How do vision and haptics combine in multisensory grasping?" *Journal of Vision*, vol. 18, no. 10, pp. 64–64, 2018.

[63] R. J. van Beers, C. M. van Mierlo, J. B. Smeets, and E. Brenner, "Reweighting visual cues by touch," *Journal of vision*, vol. 11, no. 10, pp. 20–20, 2011.

[64] K. N. de Winkel, J. Weesie, P. J. Werkhoven, and E. L. Groen, "Integration of visual and inertial cues in perceived heading of self-motion," *Journal of vision*, vol. 10, no. 12, pp. 1–1, 2010.

[65] J. S. Butler, S. T. Smith, J. L. Campos, and H. H. Bülthoff, "Bayesian integration of visual and vestibular signals for heading," *Journal of vision*, vol. 10, no. 11, pp. 23–23, 2010.

[66] C. Bozzacchi, R. Volcic, and F. Domini, "Grasping in absence of feedback: systematic biases endure extensive training," *Experimental brain research*, vol. 234, no. 1, pp. 255–265, 2016.

[67] A. Sengül, G. Rognini, M. van Elk, J. E. Aspell, H. Bleuler, and O. Blanke, "Force feedback facilitates multisensory integration during robotic tool use," *Experimental brain research*, vol. 227, no. 4, pp. 497–507, 2013.

[68] T.-C. Lin, A. U. Krishnan, and Z. Li, "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, 2020, pp. 9157–9163.

[69] J. Shaw and K. Cheng, "Object identification and 3-d position calculation using eye-in-hand single camera for robot gripper," in *2016 IEEE International Conference on Industrial Technology (ICIT)*. IEEE, 2016, pp. 1622–1625.

[70] V. Lippiello, B. Siciliano, and L. Villani, "Eye-in-hand/eye-to-hand multi-camera visual servoing," in *Proceedings of the 44th IEEE Conference on Decision and Control*. IEEE, 2005, pp. 5354–5359.

[71] Z. Li, P. Moran, Q. Dong, R. J. Shaw, and K. Hauser, "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, 2017, pp. 3581–3586.

[72] T. Shibata and K. Tanie, "Influence of a priori knowledge in subjective interpretation and evaluation by short-term interaction with mental commit robot," in *Proceedings. 2000 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS 2000)(Cat. No. 00CH37113)*, vol. 1. IEEE, 2000, pp. 169–174.

[73] M. K. Lee, K. P. Tang, J. Forlizzi, and S. Kiesler, "Understanding users! perception of privacy in human-robot interaction," in *2011 6th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*. IEEE, 2011, pp. 181–182.

[74] J. P. Hansen, A. Alapetite, M. Thomsen, Z. Wang, K. Minakata, and G. Zhang, "Head and gaze control of a telepresence robot with an hmd," in *Proceedings of the 2018 ACM Symposium on Eye Tracking Research & Applications*, 2018, pp. 1–3.

[75] R. Reilink, G. de Bruin, M. Franken, M. A. Mariani, S. Misra, and S. Stramigioli, "Endoscopic camera control by head movements for thoracic surgery," in *2010 3rd IEEE RAS & EMBS International Conference on Biomedical Robotics and Biomechatronics*. IEEE, 2010, pp. 510–515.

[76] C. Carreto, D. Gêgo, and L. Figueiredo, "An eye-gaze tracking system for teleoperation of a mobile robot," *Journal of Information Systems Engineering & Management*, vol. 3, no. 2, p. 16, 2018.

[77] A. Roncone, U. Pattacini, G. Metta, and L. Natale, "A cartesian 6-dof gaze controller for humanoid robots." in *Robotics: science and systems*, vol. 2016, 2016.

[78] S. J. Vine, R. S. Masters, J. S. McGrath, E. Bright, and M. R. Wilson, "Cheating experience: Guiding novices to adopt the gaze strategies of experts expedites the learning of technical laparoscopic skills," *Surgery*, vol. 152, no. 1, pp. 32–40, 2012.

[79] M. Boyer, M. L. Cummings, L. B. Spence, and E. T. Solovey, "Investigating mental workload changes in a long duration supervisory control task," *Interacting with Computers*, vol. 27, no. 5, pp. 512–520, 2015.

[80] D. Kent, C. Saldanha, and S. Chernova, "A comparison of remote robot teleoperation interfaces for general object manipulation," in *Proceedings of the 2017 ACM/IEEE International Conference on Human-Robot Interaction*, 2017, pp. 371–379.

[81] P. Treffner and M. Turvey, "Symmetry, broken symmetry, and handedness in bimanual coordination dynamics," *Experimental Brain Research*, vol. 107, no. 3, pp. 463–478, 1996.

[82] S. Ladwig, C. Sutter, and J. Müsseler, "Intra-and intermodal integration of discrepant visual and proprioceptive action effects," *Experimental brain research*, vol. 231, no. 4, pp. 457–468, 2013.

[83] L. Xiong, C. B. Chng, C. K. Chui, P. Yu, and Y. Li, "Shared control of a medical robot with haptic guidance," *International journal of computer assisted radiology and surgery*, vol. 12, no. 1, pp. 137–147, 2017.

[84] J. Aleotti, G. Micconi, S. Caselli, G. Benassi, N. Zambelli, M. Bettelli, and A. Zappettini, "Detection of nuclear sources by uav teleoperation using a visuo-haptic augmented reality interface," *Sensors*, vol. 17, no. 10, p. 2234, 2017.