Exploring the Potential of Full Body and Hand Gesture Teleoperation of Robots Inside Heterogeneous Human-Robot Teams

Seng Lee Koh, Kevin Pfeil, & Joseph J. LaViola Jr.
University of Central Florida, Orlando, FL

We present a within-subjects user study to compare robot teleoperation schemes based on traditional PC and game console input hardware, to a 3D spatial interaction interface. The goal of the study is to explore whether 3D spatial gestures can be an effective teleoperation scheme for multiple robot configurations in a heterogeneous Human-Robot Team (HRT). Our research explores the user's performance and disposition towards each teleoperation scheme so as to study their preferences regarding the efficacy of gesture-based teleoperation. Our results indicate that despite little training and lack of exposure on using 3D spatial interaction schemes to control robots, users are able to complete a complex task with the robot team with no statistically significant difference in quantitative performance. Qualitative statistics are analyzed and a discussion of user preferences is provided.

INTRODUCTION

A perennial objective of HRI interface design is to help alleviate cognitive load from human agents. As task load and complexity increase for Human-Robot Teams (HRTs) that is concomitant with the level of autonomy granted to each robot in the team, it is imperative for the underlying UI to incorporate a more natural and intuitive means of input available for human agents to interact quickly with the robotic agents. Recent HRI research thus trended along investigations or implementations extracting information from one or multiple natural modalities from human agents to guide and supervise full or semi-autonomous agents into perceiving and executing a set of tasks, either independently or in collaborative modes. Modalities can involve unimodal or fused combinations of speech, gaze, thought, emotion, and gestures involving full-body, faces, hands and fingers (Burke et al., 2013; Correa et al., 2010; Taylor et al., 2012).

Although the goal is for robots to learn from human agents in order to delegate as much cognitive load as possible to AI, there will be times when a human must intervene and take over a robot's task functions with manual control, due to safety, robot malfunction, or even task expedition (Morris et al., 2002). Another example can be found in HRT scenarios where a high level of autonomy is present in the agents in a task collaborative scheme with a human agent acting as a leader or as an "equal partner" (Shah et al., 2011). Here, we can replace the human agent in the collaborative task with a surrogate robot. Extending from this collaborative scenario, it is also possible for the human agent to switch teleoperation between robots of the same HRT, implicitly controlling the pace and execution of the collaborative task, or changing the task focus for the team. Thus, there is opportunity to explore various input designs of a multimodal teleoperation system.

This work will also gain further importance as tracking for gesture-based supervisory control of robots matures, allowing users to switch seamlessly between supervisory and manual modes without device encumberment, especially in scenarios where human agents are leading HRTs from the safety of an indoors location remotely.

The paper offers the following contributions to the human-robot interaction literature:

- A user study of three control schemes, including a 3D spatial user interface (3DUI), a gaming controller, and the mouse-keyboard combination.
- We demonstrate a system that can be physically non-obtrusive to the user in performing teleoperations of heterogeneous HRTs compared with traditional input.
- We provide lessons learned that may assist in designing future implementations for gestural control.

RELATED WORK

Using gestures to teleoperate robots is not a new idea. There has been a significant amount of literature reported about the use of vision-based sensors to capture body gestures that control robotic platforms (Uribe et al., 2011; Du et al., 2012; Pfeil et al., 2013). There is also a significant amount of literature pertaining to avateering, i.e. letting a humanoid robot imitate the pose of the human agent in control (Nguyen et al., 2012; Dragan et al., 2013; Koh et al., 2014). The findings of Pfeil et al. suggest the incorporation of descriptive metaphors into full-body gesture design enable commands to be more natural and intuitive (Pfeil et al., 2013). Our work explores this idea further, by generalizing the robot platform domain up to HRTs, and augmenting the spatial interaction experience by including hand gestures and speech. We have also included a humanoid robot as part of the HRT in the study, and applied avateering to the humanoid's arm teleoperation (Nguyen et al., 2012; Koh et al., 2014).

There has been also a significant amount of literature about using multiple modes of natural communication to interact with robots, implicitly guiding and manipulating robots at a supervisory basis. Ghidary et al. developed a prototype for interacting with robots through the use of natural language (Ghidary et al., 2001). By conjoining the spoken phrase and a hand gesture, the robot was able to visually identify an object in a room and associate it with values from

the phrase. Larochelle et al. describe a multimodal interface to command a semi-autonomous robot, receiving commands either by explicit manipulation input through a GUI, or by speaking commands such as “move forward” (Larochelle et al., 2011). In our work however, we minimize agent autonomy as we wish to investigate and explore the optimal conditions for teleoperating a diverse HRT using 3D spatial interaction techniques.

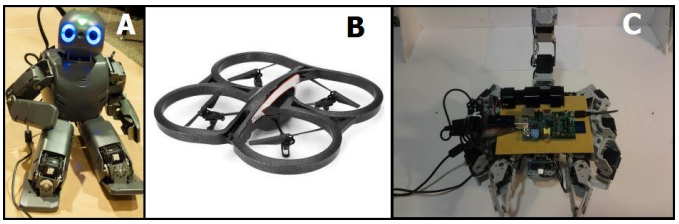

Figure 1: All robots used in the study. A: Darwin-OP humanoid. B: Parrot AR Drone 2.0. C: Scorpion constructed from Robotis Bioloid Premium kit.

HRT CONTROL SCHEMES

We developed a 3D spatial interaction prototype, with each robot having a different interaction metaphor applied to its teleoperations (Figure 1). For the user study, two other well-known teleoperation schemes were implemented based upon devices used in common PC and console gaming: a PlayStation 3 (PS3) game controller and a keyboard-mouse combination.

As a test subject would be required to switch between robots to complete a trial run, we decided to include speech as an additional input modality for all three control schemes. Speech here is used to toggle robot and, for the UAV, camera view selection. The purpose for speech in the control schemes is analogous to how gamers use speech to address their teammates in co-op game-play.

As robot autonomy is minimized for the study, neglect time for a robot becomes infinite, implying a user can interact with any number of robots (Crandall et al., 2005). However, as we require the test subject to memorize interaction command bindings for all control schemes; we limit the number of robots to three. Additionally, we want the users to interact with each robot sufficiently enough during each trial run, yet limit each run's completion time to under five minutes so as to reduce mental fatigue and learning bias as they progress through each scheme.

3D Spatial Interaction Prototype

The 3D spatial interaction system uses two cameras: one Kinect and one Leap Motion sensor. The Leap was used to track hand gestures given the inadequate sensor resolution of the Kinect. We mounted the Leap Motion sensor onto an improvised bracket to be worn upon the wrist at the top. We chose to place the Leap by the side of the user, after pilot trials.

Robot & Interaction Metaphor Selections

Humanoid. Motivated by retrospective works in Mixed Reality and full-body imitation of humanoids, we included this robot platform as part of the designated HRT (Dragone et al., 2007; Kobayashi et al., 2007; Nguyen et al., 2012; Song et al., 2012; Stanton et al., 2012). The Darwin-OP was chosen to represent the Humanoid due to its stable walking gait and teleoperable arms. From Marchal et al. and Ngyuen et al., we used the avateering metaphor for the humanoid's arm teleoperations, and the human joystick metaphor for navigation (Marchal et al., 2011; Ngyuen et al., 2012). From pilot studies, users found it uncomfortable to maintain a 'lift' pose for the humanoid as it proceeds to the drop zone after they used the humanoid to lift the brick off the platform with the avateering metaphor. Hence, we used a hand gesture detectable with the Leap Motion that let users toggle the state on whether or not the arms can be manipulated, but yet ensure the user has the option of navigating the humanoid while moving its arms. This, however, left only one arm for the user for avateering. Double exponential smoothing was used to correct the humanoid's arms gradually to the user's arm pose (LaViola, 2003).

Scorpion. The second platform was a Robotis Bioloid Premium kit configured into a scorpion model. This featured six legs, two pincers that can clasp small objects in a physical environment, and a tail that also be used for physical interaction. Due to the nature of this platform being non-anthropomorphic and non-vehicular, we found this platform essential as part of the study due to the participation of actual non-anthropomorphic robots in the. We derived the ‘Pinch’ and ‘Tail Strike’ gestures, both executed with arm and hand gestures. ‘Pinch’ allows the user to control the angle between the pincers of both scorpion claws by manipulating the distance between the thumb and index finger tracked by the Leap sensor, while ‘Tail Strike’ enables control of the scorpion tail by manipulating the angle of the elbow joint tracked by the Kinect. Similar to the humanoid, we reused the human joystick metaphor for navigation, with the lean-based gestures allowing the user to ‘Pinch’ and `Tail Strike’ while navigating the robot simultaneously.

UAV. UAVs are already being used for various purposes in both military and commercial applications, with high utility and the ability to provide an eye-in-the-sky. We selected the Parrot AR Drone 2.0 to serve as our UAV platform. This particular device is a quadrotor that exhibits two on-board cameras; one facing forward, and the other facing downwards. Both cameras are used for target visualization in the user study as well as navigation. We used the Standing Proxy metaphor by Pfeil et al. for the UAV's navigational interaction, and the ‘Pinch’ gesture that manipulates the zoom factor of the UAV's video stream (Pfeil et al., 2013).

Game Controller

The game controller scheme was second-nature to users who are gamers, as command bindings on the controller were mapped to command bindings of actual console games (e.g. left vertical stick will be used for navigation). For example,

the left control-stick of the Sony PlayStation 3 (PS3) controller is used for navigation across all robot types in the team, while holding down the left-trigger button when moving the control-stick teleoperate the humanoid arms; but for the UAV, they are used for turning rather than strafing. The scorpion robot, on other hand, includes holding down the right-trigger button, besides the left-trigger, in order to teleoperate the pincers and tail motors respectively.

Keyboard & Mouse

Participants used the well-known W-A-S-D + modifier keys to teleoperate, while the mouse was reserved for manipulating the WIMP widgets such as UAV camera zooming. Analogous to the Game Controller Scheme, W-A-S-D keys are used for general navigation, while W-A-S-D + 'Shift' modifier keys teleoperate the humanoid arms; but used for UAV turning. The scorpion robot includes the 'Control' modifier key, besides 'Shift', in order to teleoperate the pincers and tail motors.

USER STUDY

We devised a user study in order to evaluate our control schemes. The following sections discuss participant demographics, task objectives, and the targeted points of data.

Participants

We recruited 14 participants from a college campus to take part in a within-subjects study. Two participants were female. The average age was 25 years; the median was 23.5 (min 19, max 40). All participants had experience using RC toys. All but one used a motion capture device. 9 of 14 participants indicated regularly playing video games using controllers. In order to gauge a participant's familiarity with the modes of input used in the study, we asked for a percentage based indication for each mode, totaling 100%. The average keyboard based gaming percentage of the participants is 61%; controller-based gaming was noted at 34%, and 3DUI gaming was at 5%.

Software and Apparatus

We used the three aforementioned robots for our study. A laptop running Ubuntu 12.04 and ROS Fuerte was used. Our input devices included the laptop's embedded keyboard, a Bluetooth PS3 Six-Axis Controller, and a combination of the Kinect and the Leap Motion sensor. We developed a QT application to provide visual feedback for the user.

Design and Procedure

We assigned the participants to use each of the control schemes in a counter-balanced design to manipulate the three robots, with specific objectives in mind. The participants are allowed to train for 5 minutes before proceeding with any control scheme. The humanoid and scorpion each needed to grab a plastic brick and bring it to a designated zone. For the humanoid, the arms were able to squeeze the brick. The scorpion was able to use the pincers in order to grab hold of handles attached to the brick. Although this task seems easy, in real scenarios where the operator is at a remote location, the robots would not be in plain view; thus we required the UAV to fly over each robot's work area. The camera would provide the user with a view of the workplace, enabling the user to complete the objective. The complete task objectives, in order, are as follows, and depicted in Figure 2:

- Use the scorpion pincers to grasp the brick
- Navigate the scorpion to the designated zone
- Fly the UAV over the scorpion to verify the object is aligned with the zone
- Release the brick from the scorpion pincers
- Switch to the humanoid robot and use the arms to grasp the brick
- Navigate the humanoid to the designated zone
- Fly the UAV over the humanoid and verify the brick is aligned with the zone
- Release the brick from the humanoid arms
- Land the UAV in the designated landing zone

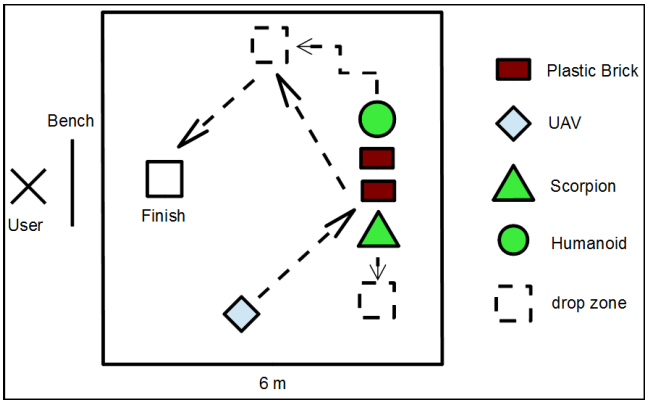

Figure 2: Diagram of the user study layout.

Quantitative Metrics

We measured the interaction time from the point where the first command was assigned to a robot, until the point where the UAV touched the ground. Upon the occurrence of a mistake that disrupted task completion, such as the UAV crashing, a ground robot dropping the brick, or a ground robot moving off the designated area, the time was paused and an error was logged. After correction by the study proctor, the timing resumed. We logged the total completion time and the number of errors in order to determine if there was a clear advantage in terms of efficiency as well as accuracy.

Qualitative Metrics

After using a control scheme, the participant was asked to fill out a questionnaire. This survey contained questions that measured the user's disposition to that individual technique. These questions asked the user to rate particular factors on a 7-point Likert scale. After all techniques were completed, the participant was asked to complete one last

survey to rank the techniques on various metrics, with no ties allowed. Finally, we captured any comments the users may have had, on both surveys.

RESULTS AND DISCUSSION

Using a repeated measures ANOVA test, no significant difference was found for completion time ($F_{2,13} = 1.103$, $p = 0.347$). The Keyboard scheme did exhibit the best average completion time, as it is a very common input device; as per the demographics collected from the pre-survey, the keyboard was used the most, while the 3DUI is rarely used by the test subjects (Figure 3).

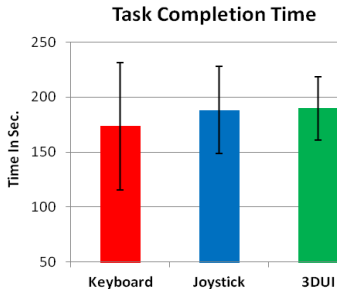

Figure 3: Mean Task Completion Time Across Teleoperation Schemes

Similarly, the quantity of errors seemed to be comparable across control schemes. The keyboard exhibited a lower average number of errors, while the controller and 3DUI tied. However, the median was 0 for 3DUI but 0.5 for the controller; more participants encountered at least one error when using the controller, compared to the 3DUI.

Although participants were given time to become accustomed to the robots and controls for each technique, their prior experience and comfort with the keyboard control assisted with the completion of the trial. However, the data offers evidence for 3D spatial interaction being an appropriate alternative to the current control schemes of HRTs. As there was little difference in completion times between the techniques, further iterations for 3D gesture design may prove to bolster performance, by allowing an even more natural interface.

Qualitative Analysis

Rating Results. We performed non-parametric Friedman tests on the rating results. We found two results showing statistical difference; we then performed Wilcoxon signed rank tests on these results to find difference between the 3DUI and the other control schemes. We used Holm's Sequential Bonferroni adjustment to control Type I errors (Holm, 1979).

In terms of how *comfortable* it was to complete the entire task, statistical significance was found ($\chi^2 = 9.814$, $p < 0.007$). For this metric, the 3DUI control scheme's performance was statistically different than the keyboard. This is attributed to the unfavorable perception of avateering for the humanoid robot. Statistical difference was found for the control schemes regarding the humanoid, ($\chi^2 = 11.023$, $p < 0.004$). The perceived level of comfort for the other two robots did not see any statistical difference between input modalities.

Ranking Results. We performed non-parametric Friedman tests on the ranking results. None of the ranking metrics showed any statistical difference, except for the rankings for Ease of Use. Having found significance, we performed a Wilcoxon signed rank test between the control schemes. Statistical difference was found between the 3DUI and the keyboard ($\chi^2 = 7.000$, $p < 0.030$). This is an indication that details how lack of familiarity harmed performance.

Similar to the quantitative results, we believe the users did not find the 3DUI favorable due to their significant experience, familiarity, and comfort with the keyboard and game controller. However, user disposition should increase with further training, and we would expect performance results to improve as well.

Discussion

There were many factors that contributed to the participants' displeasure with the 3DUI control scheme. It seems that the participants did not appreciate the human joystick gesture when navigating and turning the humanoid and scorpion robots. During setup for the 3DUI scheme, we offered users the option of calibrating their own leaning gestures for navigating the ground-robots, and although they confirmed comfortable poses for this gesture, many users still rated the control scheme as uncomfortable. Due to inexperience with motion control, we believe these users under-calibrated; if they had performed a less demanding gesture, they could have had a more favorable experience. Though users perceive the avateering metaphor to control the humanoid limbs positively, using hand gestures to create context for the humanoid's arms was not well received. For the scorpion robot, many users regard the 'Pinch' and 'Tail Strike' gestures fun and easy to recall, but uncomfortable compared to the traditional control schemes. Users found navigating the UAV based on the Standing Proxy metaphor was natural and slightly more comfortable.

Overall, we find that the under-appreciation of the 3DUI control scheme occurred for two main reasons: (1) Participants simply did not have enough experience using this form of input; it was very unfamiliar and therefore did not allow for higher levels of interaction, and (2) The human-joystick metaphor may be intuitive but highly uncomfortable for navigating ground-robots, especially when under-calibrated. However, a positive note is of the 3DUI control scheme task completion time, which was statistically insignificant and comparable to the traditional controls. It remains to be seen how task performance would be affected had the users been given an extended amount of time to familiarize with the 3DUI scheme; we expect, however, for the timing and error data to improve.

FUTURE WORK

We aim at redesigning the gestures to be more comfortable, in an effort to find interaction techniques that would be viewed positively by the participants, while

decreasing the amount of time needed to complete our task. We anticipate performing future user studies to measure performance of these new gestural commands.

We plan on extending this work to study methodology to reduce cognitive load for HRT operators. Future research will include more robotic platforms including alternative ground systems, as well as underwater and surface vehicles.

Additional modes of interaction should also be considered as alternative methods of HRT teleoperation. Touch or sketch-based interfaces could provide alternative modes than traditional forms (Correa et al., 2010). It would be interesting to compare task performance and qualitative metrics when using or in combination with this modality.

Further, we plan on studying the effects on user cognitive load, when including higher levels of autonomy.

CONCLUSIONS

It is evident through our research that user studies in the field of HRI are very necessary. By selecting the UAV interaction technique that was highly regarded in a previous study, we were able to bolster the participants' perception towards our system. However, a formal study with further iterations is needed to explore optimal metaphors and natural modalities that are suited for ground-robot teleoperation. We believe that a user study for humanoid and non-anthropomorphic robot control would enable an accurate redesign of our described control scheme, which would then allow users to perform our task more naturally and comfortably.

Regardless, we have shown that task completion time between traditional modes of input and 3DUI are comparable. We envision HRT supervisors with a well-designed gestural control scheme having the ability to switch between and command multiple robots seamlessly for teleoperation, especially in the presence of a mixed-unit team. By incorporating speech and natural gestures, operators should be able to perform duties with reduced cognitive load.

REFERENCES

Burke, D., Schurr, N., Ayers, J., Rousseau, J., Fertitta, J., Carlin, A., & Dumond, D. (2013, May). Multimodal interaction for human-robot teams. In *Unmanned Systems Technology XV* (Vol. 8741, p. 87410E). International Society for Optics and Photonics.

Correa, A., Walter, M. R., Fletcher, L., Glass, J., Teller, S., & Davis, R. (2010, March). Multimodal interaction with an autonomous forklift. In *Proceedings of the 5th ACM/IEEE international conference on Human-robot interaction* (pp. 243-250). IEEE Press.

Crandall, J. W., Goodrich, M. A., Olsen, D. R., & Nielsen, C. W. (2005). Validating human-robot interaction schemes in multitasking environments. *IEEE Transactions on Systems, Man, and Cybernetics-Part A: Systems and Humans*, 35(4), 438-449.

Dragan, A., & Srinivasa, S. (2013, May). A policy-blending formalism for shared control. In *International Journal of Robotics Research*. 32(7), 790-80

Dragone, M., Holz, T., & O'Hare, G. M. (2007, August). Using mixed reality agents as social interfaces for robots. In *Robot and Human interactive Communication, RO-MAN 2007*. The 16th IEEE International Symposium on Robot and Human interactive Communication (pp. 1161-1166). IEEE.

Du, G., Zhang, P., Mai, J., & Li, Z. (2012). Markerless kinect-based hand tracking for robot teleoperation. *International Journal of Advanced Robotic Systems*, 9(2), 36.

Ghidary, S. S., Nakata, Y., Saito, H., Hattori, M., & Takamori, T. (2001). Multi-modal human robot interaction for map generation. In *Intelligent Robots and Systems*, 2001. Proceedings. 2001 IEEE/RSJ International Conference on (Vol. 4, pp. 2246-2251). IEEE.

Holm, S. (1979). A simple sequentially rejective multiple test procedure. *Scandinavian journal of statistics*, (pp. 65-70).

Kobayashi, K., Nishiwaki, K., Uchiyama, S., Yamamoto, H., & Kagami, S. (2007, November). Viewing and reviewing how humanoids sensed, planned and behaved with mixed reality technology. In *Humanoid Robots 2007*, 7th IEEE-RAS International Conference on Humanoid Robots (pp. 130-135). IEEE.

Kobayashi, K., Nishiwaki, K., Uchiyama, S., Yamamoto, H., Kagami, S., & Kanade, T. (2007, November). Overlay what humanoid robot perceives and thinks to the real-world by mixed reality system. In *Proceedings of the 2007 6th IEEE and ACM International Symposium on Mixed and Augmented Reality* (pp. 1-2). IEEE Computer Society.

Koh, S. L., Pfeil, K. & LaViola, J. (2014). Enhancing the robot avateering metaphor discreetly with an assistive agent and its effect on perception. In *Robot and Human interactive Communication, RO-MAN 2014*. The 23rd IEEE International Symposium on Robot and Human interactive Communication (pp. 1095-1102). IEEE.

Larochelle, B., Kruijff, G. J. M., Smets, N., Mioch, T., & Groenewegen, P. (2011). Establishing human situation awareness using a multi-modal operator control unit in an urban search & rescue human-robot team (pp. 229-234). IEEE.

LaViola, J. J. (2003, May). Double exponential smoothing: an alternative to Kalman filter-based predictive tracking. In *Proceedings of the workshop on Virtual environments* 2003 (pp. 199-206). ACM.

Lee, J. H. (2012, December). Full-body imitation of human motions with kinect and heterogeneous kinematic structure of humanoid robot. In *System Integration (SII)*, 2012 IEEE/SICE International Symposium on (pp. 93-98). IEEE.

Marchal, M., Pettré, J., & Lécuyer, A. (2011, March). Joyman: A human-scale joystick for navigating in virtual worlds. In *3D User Interfaces* (3DUI), 2011 IEEE Symposium on (pp. 19-26). IEEE.

Morris, A. C., Smart, C. K., & Thayer, S. M. (2002). Adaptive Multi-Robot, Multi-Operator Work Systems. In *Multi-Robot Systems: From Swarms to Intelligent Automata* (pp. 203-211). Springer, Dordrecht.

Nguyen, V., & J.H. Lee (2012). Full-body imitation of human motions with kinect and heterogeneous kinematic structure of humanoid robot. In *System Integration (SII)*, 2012 International Symposium on System Integration, (pp. 93–98), IEEE/SICE.

Pfeil, K., Koh, S. L., & LaViola, J. (2013, March). Exploring 3d gesture metaphors for interaction with unmanned aerial vehicles. In *Proceedings of the 2013 international conference on Intelligent user interfaces* (pp. 257-266). ACM.

Shah, J., Wiken, J., Williams, B., & Breazeal, C. (2011, March). Improved human-robot team performance using chaski, a human-inspired plan execution system. In *Proceedings of the 6th international conference on Human-robot interaction* (pp. 29-36). ACM.

Sian, N., Sakaguchi, T., Yokoi, K., Kawai, Y., & Maruyama, K. (2006, August). Operating humanoid robots in human environments. In *Proc. RSS Workshop: Manipulation for Human Environments*.

Song, W., Guo, X., Jiang, F., Yang, S., Jiang, G., & Shi, Y. (2012, August). Teleoperation humanoid robot control system based on kinect sensor. In *Intelligent Human-Machine Systems and Cybernetics (IHMSC)*, 2012 4th International Conference on (Vol. 2, pp. 264-267). IEEE.

Stanton, C., Bogdanovych, A., & Ratanasena, E. (2012, December). Teleoperation of a humanoid robot using full-body motion capture, example movements, and machine learning. In *Proc. Australasian Conference on Robotics and Automation*.

Taylor, G., Frederiksen, R., Crossman, J., Quist, M., & Theisen, P. (2012, May). A multi-modal intelligent user interface for supervisory control of unmanned platforms. In *Collaboration Technologies and Systems (CTS)*, 2012 International Conference on (pp. 117-124). IEEE.

Uribe, A., Alves, S., Rosário, J. M., Ferasoli Filho, H., & Pérez-Gutiérrez, B. (2011, October). Mobile robotic teleoperation using gesture-based human interfaces. In *Robotics Symposium, 2011 IEEE IX Latin American and IEEE Colombian Conference on Automatic Control and Industry Applications (LARC)* (pp. 1-6). IEEE.