Pan-Tilt-Roll Televisualization With Adjustable Baseline Stereo

Patrick Naughton[1], James Seungbum Nam[2], Joao M. C. Marques[1], Jing-Chen Peng[1], Yifan Zhu[1], Qianxi Kong[2], and Kris Hauser[1]

I. Introduction and Related Work

Telepresence robots seek to immerse a person's perception in a remote location via a robotic embodiment. Accurate depth perception is needed to perform many tasks efficiently and can be provided by virtual reality (VR) head-mounted displays (HMDs) connected to stereoscopic cameras. The contributions of this paper, which were featured on Team AVATRINA's telepresence robot in the ANA Avatar XPRIZE, demonstrated an immersive VR interface that included a 3 DoF (pan-tilt-roll) head to better match human head movement, as well as a motorized adjustable baseline stereo camera that can be matched to the operator's interpupillary distance (IPD). This paper presents the results of empirical human subjects studies relating these design choices to teleoperation task performance. We find that 1) IPD-baseline matching is most helpful at the mid-range of a robot's workspace but operators are able to cope with mismatch at near and far ranges, and 2) 2 DoF (pan-tilt) heads significantly improve hand-eye coordination over static heads but the effect of 3 DoFs over 2 DoFs is minor despite potentially improved depth estimation through parallax.

Previous studies on VR have found that mismatches between the IPD of an HMD and the user can cause discomfort and misperception of depth in the virtual world [1, 2, 3]. Stereo telepresence introduces another source of potential error: the baseline between stereo cameras may not match the operator's IPD. This paper investigates the effect of matching a robot's IPD to the operator's on hand-eye coordination on a peg-in-hole task shown in Fig. 1 at different levels of IPD-stereo baseline mismatch.

Besides stereo disparity, operators observe the remote scene from multiple viewpoints, providing depth cues from parallax and the ability to peer around occlusion. Humans use head, torso, and body movements to change viewpoint. The effect of varying DoFs of robot telepresence heads has been previously investigated in [4], which found that increasing the number of head DoFs from 0 to 3 to 6 improved the success rate and speed of completing a peg-in-hole task. However, the study had a number of limitations. The sample size was small, subjects were members of the research team, and the task was designed to make occlusion a significant factor. Moreover, the 6 DoF robot used in this study is likely prohibitively expensive for many telepresence robots and most entries in the XPRIZE exhibited 2 or 3 DoF heads. This paper conducts a larger study to evaluate 0, 2, and 3 DoF head designs. Interestingly, we find that operators do not exploit parallax (provided by the roll DoF) much to provide more accurate depth perception.

This work was partially supported by NSF Grant #2025782.

[1]P. Naughton, J. M. C. Marques, JC Peng, Y. Zhu, and K. Hauser are with the Department of Computer Science, University of Illinois at Urbana-Champaign, IL, USA. {pn10, jmc12, jcpeng2, yifan16, kkhauser}@illinois.edu

[2]J. S. Nam and Q. Kong are with the Department of Mechanical Science and Engineering, University of Illinois at Urbana-Champaign, IL, USA. {sn29, qianxik2}@illinois.edu

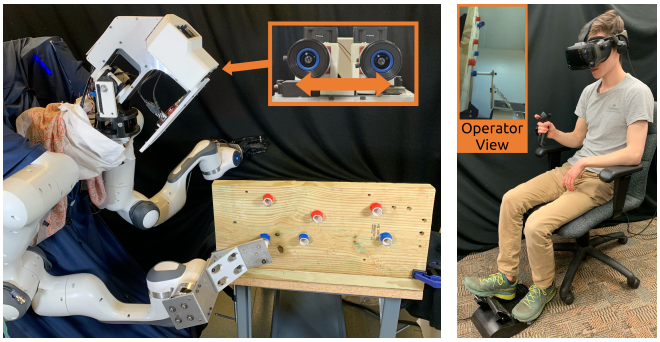

Fig. 1: (Left) A peg-in-hole experiment is used to examine how the robot's neck DoFs and stereo baseline affect operator depth perception. Two sets of holes were used to reduce learning effects between trials. (Right) Operator inserting the peg into the first hole.

II. Experimental Evaluation

Hardware. The stereo camera uses two Allied Vision Alvium 1800 U-500C 5 MP cameras with 1.67 mm focal-length wide-angle lenses providing $\sim 120^\circ \times 100^\circ$ field-of-view. The cameras are mounted to a linear rail and connected to a pair of Actuonix L12-R linear actuators that adjust the baseline with $\sim$1 mm resolution. Since the mounting is not perfectly parallel, the images are rectified before being transmitted to the operator. Stereo video was transmitted to the HMD via a WebRTC stream with latency $\sim$220 ms.

The camera is mounted to a 3 DoF robot head designed to mimic the operator's neck movement. Three Dynamixel XM430 motors are mounted with axes intersecting at a point, providing roll, pitch and yaw motion. The head tracks the operator's head orientation with respect to an operator-defined "home" orientation, which can be adjusted at any time. The operator views the remote scene with a Valve Index Headset which displays the stereo video. The operator uses one Valve Index Controller with their right hand to move the robot's right arm to complete the tasks. The operator moves the robot's target transform using a clutching mechanism as described in [5].

Study Procedure. We formulated the following hypotheses *a priori* about the system:

- **H1:** Subjects complete tasks more slowly as IPD-baseline mismatch increases.

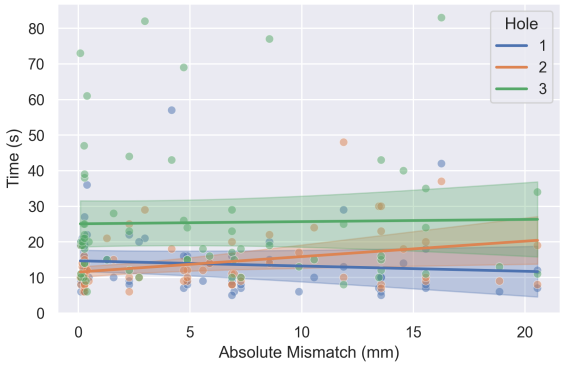

Fig. 2: Effect of IPD-baseline mismatch on peg-in-hole task completion time for each of the 3 holes. Lines show fitted regression models and shaded regions indicate 95% confidence intervals. Dots show individual measurements.

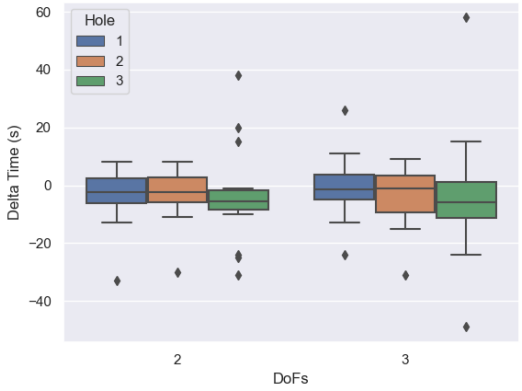

Fig. 3: Change in task completion time for each hole with respect to each subject's task completion time at 0 DoFs.

- **H2:** Subjects complete tasks more quickly as the number of head DoFs increases.

To test H1 and H2 we designed a peg-in-hole task with 3 red holes and 3 blue holes. The peg is attached to the robot's hand and has a diameter of 16.1 mm while holes have a diameter of 20.7 mm. We used a large tolerance so that novice operators could perform the task quickly without extensive training. The robot was constrained to use one arm which could translate in 3 DoFs and rotate in only 1 DoF (horizontal axis) to isolate the effects of depth perception and to avoid singularities.

We recruited 16 subjects (9 male, 7 female) from the university's student population. Subjects were of age 19–30 (mean: 24.6) and self-reported their familiarity with the avatar robot to be an average of 2.9 out of 7 on a Likert scale. Two subjects had previously been trained in how to use the robot in prior studies. Each subject measured their own IPD using a ruler and a mirror, and then fine-tuned the HMD to find the most comfortable setting, which was kept constant throughout the experiment. A researcher then trained the subject to use the head and arm, which lasted approximately 20 minutes.

First, to test the effects of IPD mismatch, the head pose was fixed and subjects completed tasks under four robot IPD settings: Matched, Average (62.72 mm [6]), Min (49.44 mm), and Max (69.88 mm). This was first done with the red holes and then repeated with the blue holes. The conditions were tested in a randomized order, which was unknown to the subjects. Then, to test the effects of robot head DoFs, the IPD was set to Matched and the subjects completed tasks under three head DoF conditions: 0, 2 (pan-tilt), and 3, first with the red holes and then with the blue holes. Each subject experienced the conditions in a randomized order and was informed of the condition since they had to consciously use their neck to use the different DoFs. Trials on the red holes were used as training trials to reduce learning effects; only times on the blue holes are analyzed.

Results and Analysis. Fig. 2 shows how subjects' task completion times changed as the absolute IPD mismatch changed for each hole. To test **H1**, we ran a generalized estimating equations (GEE) regression [7] grouped by subject with an autoregressive covariance structure [8], where distance was computed as the distance between trial indices. After applying the Bonferroni correction, we found significant correlation between IPD mismatch and completion time for Hole 2 ($\beta = 0.434$ s/mm, $\sigma_{\mathrm{M}} = 0.173$ s/mm, $p = 0.0372$) but no significant effect for Holes 1 and 3 ($p = 1.0$ for both), providing *weak evidence to support* **H1**. We note that there are more outlying data points for Holes 1 and 3 compared to Hole 2, and hypothesize that this may be because insertion into Holes 1 and 3 requires the robot's arm to be at less dexterous configurations than for Hole 2.

Fig. 3 shows the subjects' changes in task completion time from the 0 DoF condition for the 2 and 3 DoF conditions. We ran a Shapiro-Wilk test [9] on the data for each hole at each of the three conditions, indicating significant deviation from normality. To test **H2**, we ran a Friedman test for each hole with Bonferroni correction and found no significant effects ($p = 1.0$, $p = 1.0$, $p = 0.140$ for Holes 1, 2, and 3 respectively). Post hoc pairwise one-sided Wilcoxon-signed-rank comparisons provided weak evidence that completion time for Hole 3 is smaller when using 2 DoFs compared to 0 DoFs ($M = 3.94$ s, $SD = 16.54$ s, $p = 0.0719$) and when using 3 DoFs compared to 0 DoFs ($M = 4.00$ s, $SD = 21.35$ s, $p = 0.0877$), but not when using 3 DoFs compared to 2 DoFs ($p = 0.470$). This corresponds to a 14.7% and a 10.6% average reduction in task completion time when switching from a 0 to 2 DoF neck and from 0 to 3 DoF neck respectively. *These results weakly support* **H2** and corroborate previous research [4] showing that increasing head DoFs improves task performance on tasks that require depth perception and occlusion resolution, but with diminishing returns.

III. Conclusion

Experiments suggest that matching stereo baseline to the operator's IPD improves telemanipulation proficiency in certain regions of a robot's workspace. It may also improve comfort for long-term use as suggested by prior VR studies. Increasing head DoFs improves manipulation proficiency as well, but pan-tilt may be sufficient for many tasks. Fixed-baseline cameras near the average human IPD may be satisfactory for some applications but may impair performance for operators with large or small IPD.

REFERENCES

[1] A. Takagi, S. Yamazaki, Y. Saito, and N. Taniguchi, "Development of a stereo video see-through HMD for AR systems," in *Proceedings IEEE and ACM International Symposium on Augmented Reality (ISAR 2000)*, Munich, Germany: IEEE, 2000, pp. 68–77. [Online]. Available: http://ieeexplore.ieee.org/document/880925/.

[2] P. B. Hibbard, L. C. van Dam, and P. Scarfe, "The Implications of Interpupillary Distance Variability for Virtual Reality," in *2020 International Conference on 3D Immersion (IC3D)*, Dec. 2020.

[3] R. S. Renner *et al.*, "The Influence of the Stereo Base on Blind and Sighted Reaches in a Virtual Environment," en, *ACM Transactions on Applied Perception*, vol. 12, no. 2, Apr. 2015. [Online]. Available: https://dl.acm.org/doi/10.1145/2724716 (visited on 03/01/2023).

[4] M. Schwarz and S. Behnke, "Low-Latency Immersive 6D Televisualization with Spherical Rendering," in *2020 IEEE-RAS 20th International Conference on Humanoid Robots (Humanoids)*, Munich, Germany: IEEE, Jul. 2021, pp. 320–325. [Online]. Available: https://ieeexplore.ieee.org/document/9555797/ (visited on 03/23/2023).

[5] J. M. C. Marques, P. Naughton, Y. Zhu, N. Malhotra, and K. Hauser, "Commodity Telepresence with the AvaTRINA Nursebot in the ANA Avatar XPRIZE Semifinals," en, in *RSS 2022 Workshop on "Towards Robot Avatars: Perspectives on the ANA Avatar XPRIZE Competition"*, IEEE, 2022.

[6] C. C. Gordon *et al.*, "2012 anthropometric survey of u.s. army personnel: Methods and summary statistics," 2014.

[7] K.-Y. Liang and S. L. Zeger, "Longitudinal data analysis using generalized linear models," en, *Biometrika*, vol. 73, no. 1, pp. 13–22, 1986. [Online]. Available: https://academic.oup.com/biomet/article-lookup/doi/10.1093/biomet/73.1.13 (visited on 03/30/2023).

[8] B. Rosner and A. Munoz, "Autoregressive modelling for the analysis of longitudinal data with unequally spaced examinations," en, *Statistics in Medicine*, vol. 7, no. 1-2, pp. 59–71, Jan. 1988. [Online]. Available: https://onlinelibrary.wiley.com/doi/10.1002/sim.4780070110 (visited on 03/30/2023).

[9] S. S. Shapiro and M. B. Wilk, "An Analysis of Variance Test for Normality (Complete Samples)," *Biometrika*, vol. 52, no. 3/4, pp. 591–611, 1965. [Online]. Available: https://www.jstor.org/stable/2333709 (visited on 03/28/2023).