Commodity Telepresence with Team AVATRINA's Nursebot in the ANA Avatar XPRIZE Finals

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

I. Introduction and Related Work

Telepresence shows promise for extending human skills beyond geographic boundaries, enabling long-range skill transfer through robot avatars in applications like telemedicine [1, 2], search-and-rescue operations [3, 4], and remote environment exploration [5, 6, 7]. Early efforts to forward this technology, like the DARPA Robotics Challenge [8], were primarily focused on the robot's ability to accomplish locomotion and manipulation tasks in adversarial conditions while guided by trained operators. Aiming to democratize the use of telepresence beyond specialized applications, the $10M ANA Avatar XPRIZE evaluated an avatar system's intuitiveness, immersiveness, the avatar's social capabilities, their robustness and manipulation capacity. It consisted of 10 distinct tasks, ranging from face-to-face communication and weight sensing, to surface texture rendering and remote tool use. External judges were asked to complete the tasks using the robot after a 45-minute training session on its usage.

Our entry, AVATRINA, placed 4th in the XPRIZE finals, being among the only 4 teams, out of 17 finalists, that completed all 10 tasks at the competition. This paper presents our system, highlights our unique system features, and shares insights and lessons learned from the competition.

II. TRINA Avatar System for XPRIZE Finals

The design philosophy espoused by our team was to use relatively inexpensive, accessible operator interfaces to democratize access to telepresence robots, which may cost $100,000 or more. We pursued this approach for the XPRIZE Semifinals as described in a prior paper [9]. We summarize the major design changes to our final system (Figure 1) in Table I, and we highlight major features of our system below.

Our avatar system is comprised of the TRINA robot and a lightweight operator station connected by a communications network. The TRINA robot can operate without a tether with 2 hours of battery life, which was deemed more than sufficient for completing the finals requirements. The robot was also redesigned to have a width of 0.7 m, which is slim enough to fit comfortably through standard doorways.

We employ 7 DoF Franka Emika Panda robots as TRINA's arms, since their kinematic redundancy allows us to maintain a human-like appearance while tracking motion targets. In addition, their high-frequency torque control interface allows

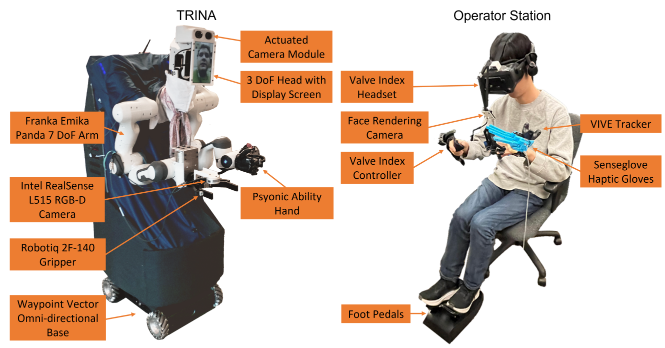

Fig. 1: An overview of the TRINA robot (left) and operator station (right). Note the absence of heavy exoskeletons.

for highly compliant contact behavior. The arm-to-torso mount design was selected to maximize the robot's dexterous workspace. This was achieved by comparing different mount options according to how effectively the arms could track pre-recorded end-effector trajectories corresponding to typical manipulation tasks.

We chose an asymmetrical design for TRINA's grippers – with an anthropomorphic left hand (Psyonic Ability Hand) and a parallel jaw gripper (Robotiq-140) on the right hand – because different gripper designs are better suited for different tasks: the Psyonic's power grasp can hold arbitrarily shaped objects securely, while the parallel jaw's ease of operation and robustness makes it suitable for operation in low visibility spaces and manipulating prismatic objects. The right gripper is equipped with a LiDAR camera to scan surfaces and determine their texture, like Shin and Choi [10]. This information is presented to the operator via augmented reality (AR) visual and haptic rendering, shown in Figure 2.

Additionally, since previous research in virtual reality [11, 12, 13] has shown that a mismatch between the interpupillary distance (IPD) of the VR headset and the user can result in reduced depth perception, we provide the first real-time adjustable stereo baseline camera system integrated with a teleoperated mobile manipulator. Our custom stereo camera system consists of two high-resolution and field-of-view cameras (Alvium 1800 U-500C with 1.67 mm focal-length wide-angle lenses) mounted on linear actuators that allow us to adjust TRINA's stereo baseline to match the operator's IPD. The camera is mounted on a 3D-printed 3 DoF neck assembly which mimics the operator's head rotation, enabling human-like head motion. Video from the cameras is transmitted via WebRTC to the headset (Valve Index), minimizing latency.

We further use AR and heads-up-display (HUD) technologies to augment the operator's situational awareness during

This work is partially supported by NSF Grant #2025782.

[1]J. M. C. Marques, P. Naughton, J.C. Peng, Y. Zhu, J. S. Nam, and K. Hauser are with the Grainger College of Engineering, University of Illinois at Urbana-Champaign, IL, USA. {jmc12,pn10, jcpeng2, yifan16, sn29, kkhauser}@illinois.edu

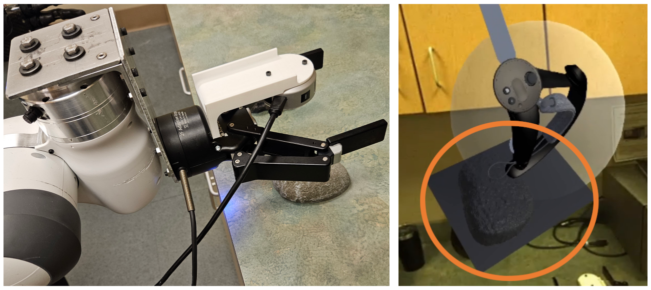

Fig. 2: A rock being scanned (left) and reconstructed as an AR mesh displayed in the headset (right). The operator can also “feel” the virtual object as the controller moves across it, as we map the reconstructed texture to vibrotactile and audio feedback.

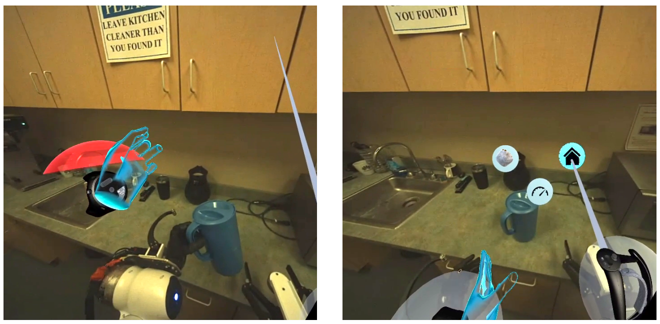

Fig. 3: (Left) View from the operator’s perspective, showing force-feedback spheres while lifting a heavy object. (Right) Heads-up display menu enables semi-autonomous functionalities: arm “homing” (right), texture-sensing mode (left), and base speed adjustment (center) icons.

TABLE I: Major TRINA changes from XPRIZE Semifinals to Finals.

Feature	Semifinals	Finals
Power (run time)	Power tether	1534 Wh battery (≈ 2 h)
Communications	Ethernet tether	WiFi
Neck Assembly	2 DoF	3 DoF
Eye Cameras	ZED-Mini	IPD-Adjustable Stereo
VR System	Meta Quest 2	Valve Index + PC
Left Gripper	Parallel Jaw	6 DoF Ability Hand
Right Gripper	4 DoF Claw	Parallel Jaw
Arms	6 DoF UR5-e	7 DoF Pandas
Force Haptics	Vibrotactile	Vibrotactile + AR force-spheres
Texture Sensing	Pressure Sensors	LiDAR Camera
Texture Rendering	Vibrotactile	Vibrotactile, auditory, AR Mesh
Arm Control	Oculus Controllers	Valve Index Controller, Senseglove
Assistive Modes	High Precision, Homing	Homing, Base Speed, Texture
Video Streaming	RTMP	WebRTC
Arm Clutching	Oculus Triggers	Foot Pedals

navigation: a togglable birds-eye view camera system lets the operator see obstacles to the sides and behind the robot, while ultrasonic proximity sensors warn the operator of any imminent collision risks through directional auditory cues.

We also enhance manipulation through AR and semi-autonomous functions. We employ force hemispheres to indicate the direction and magnitude of forces felt by the arms, with the force magnitude also conveyed by controller vibration intensity. Semi-autonomous functions are demonstrated in Figure 3 alongside these AR indicators.

To enable any operator to use the robot’s full workspace, which is typically larger than a human’s, TRINA’s arms are controlled by relative positioning. This is achieved with a clutching mechanism controlled by foot pedals. A SenseGlove DK1 is used on the operator’s left hand to command the anthropomorphic gripper and provide per-finger vibrotactile feedback based on gripper-mounted pressure sensors.

Finally, we deploy an adapted live talking head animation machine learning pipeline to render a live headset-free video of the operator on the robot [9], similar to [14], but not requiring on-site training or extensive headset modifications.

III. XPRIZE Finals Performance

During the finals, our team was able to complete all 10 tasks, being one of only 4 teams to do so. During the first competition run, the judge operated TRINA to complete all 10 tasks in a time of 24:47. The second competition run was executed at a faster pace, but a hardware emergency stop was triggered on the final task, ending the run.

Schwarz et al. [15] provides a detailed per-task breakdown of each successful team’s performance, and notes that our robot had the slowest locomotion system of these four teams. This was due to the Waypoint Vector base’s built-in obstacle detection system generating spurious detections on the dark course, limiting its maximum speed. Execution was also slowed by a network disconnection during Task 6 that required an operator station restart. The operator noted that the semi-autonomous texture sensing mode was helpful in reaching and identifying the rocks effectively.

During run 2 the operator started the last task (object identification through texture) at the 12:17 mark, much faster than the previous day. However, while reaching to feel the object’s texture, TRINA’s arm collided with the environment with high force, causing it to engage mechanical brakes. Since the arm could not be remotely recovered, this prevented the operator from completing the task. We note that other successful teams also faced emergency stops [15, 16], but recovered from them through system robustness or redundancy.

IV. Conclusion & Lessons Learned

The relatively successful performance of AVATRINA indicates that immersive and effective teleoperation can be achieved through minimal use of haptic load force feedback (LFF), which uses external devices to apply load forces to the operator’s arm. AVATRINA was one of the four teams to complete all tasks and had no LFF devices, while Pollen Robotics (second place) had a simple wearable LFF device and the other two used costly exoskeletons. This suggests that rich user interfaces with augmented reality and operator assistance are viable low-cost alternatives to LFF.

We also highlight our team’s use of semi-autonomous functions to assist with both routine and specialized tasks. Our modular software interface [9] allowed us to quickly develop a task-specific texture scanning mode for the finals, which greatly simplified the task of reaching and feeling the rocks, but note that integration of semi-autonomous tasks in shared autonomy remains an active area of research [17, 18].

References

[1] S. Mehrdad, F. Liu, M. T. Pham, A. Lelevé, and S. F. Atashzar, "Review of Advanced Medical Telerobots," *Applied Sciences*, vol. 11, no. 1, 2021. [Online]. Available: https://www.mdpi.com/2076-3417/11/1/209

[2] L. Qian, A. Deguet, and P. Kazanzides, "ARssist: augmented reality on a head-mounted display for the first assistant in robotic surgery," *Healthcare Technology Letters*, vol. 5, no. 5, pp. 194–200, 2018. [Online]. Available: https://ietresearch.onlinelibrary.wiley.com/doi/abs/10.1049/htl.2018.5065

[3] T. Klamt, M. Schwarz, C. Lenz, L. Baccelliere, D. Buongiorno, T. Cichon, A. DiGuardo, D. Droeschel, M. Gabardi, M. Kamedula, N. Kashiri, A. Laurenzi, D. Leonardis, L. Muratore, D. Pavlichenko, A. S. Periyasamy, D. Rodriguez, M. Solazzi, A. Frisoli, M. Gustmann, J. Roßmann, U. Süss, N. G. Tsagarakis, and S. Behnke, "Remote mobile manipulation with the centauro robot: Full-body telepresence and autonomous operator assistance," *Journal of Field Robotics*, vol. 37, no. 5, pp. 889–919, 2020. [Online]. Available: https://onlinelibrary.wiley.com/doi/abs/10.1002/rob.21895

[4] T. Klamt, D. Rodriguez, M. Schwarz, C. Lenz, D. Pavlichenko, D. Droeschel, and S. Behnke, "Supervised Autonomous Locomotion and Manipulation for Disaster Response with a Centaur-Like Robot," in *2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2018, pp. 1–8.

[5] N. Y. Lii, D. Leidner, P. Birkenkampf, B. Pleintinger, R. Bayer, and T. Krueger, "Toward scalable intuitive telecommand of robots for space deployment with METERON SUPVIS Justin," in *The 14th Symposium on Advanced Space Technologies for Robotics and Automation (ASTRA)*. European Space Agency, 6 2017. [Online]. Available: https://elib.dlr.de/113125/

[6] J. O. Burns, D. A. Kring, J. B. Hopkins, S. Norris, T. J. W. Lazio, and J. Kasper, "A lunar L2-Farside exploration and science mission concept with the Orion Multi-Purpose Crew Vehicle and a teleoperated lander/rover," *Advances in Space Research*, vol. 52, no. 2, pp. 306–320, 2013. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0273117712006990

[7] G. G. Muscolo, S. Marcheschi, M. Fontana, and M. Bergamasco, "Dynamics Modeling of Human–Machine Control Interface for Underwater Teleoperation," *Robotica*, vol. 39, no. 4, p. 618–632, 2021.

[8] E. Krotkov, D. Hackett, L. Jackel, M. Perschbacher, J. Pippine, J. Strauss, G. Pratt, and C. Orlowski, "The DARPA Robotics Challenge Finals: Results and Perspectives," *Journal of Field Robotics*, vol. 34, no. 2, pp. 229–240, 2017. [Online]. Available: https://onlinelibrary.wiley.com/doi/abs/10.1002/rob.21683

[9] 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," in *RSS 2022 Workshop on "Towards Robot Avatars: Perspectives on the ANA Avatar XPRIZE Competition*, 2022.

[10] S. Shin and S. Choi, "Geometry-based haptic texture modeling and rendering using photometric stereo," in *2018 IEEE Haptics Symposium (HAPTICS)*, 2018, pp. 262–269.

[11] R. S. Renner, E. Steindecker, M. MüLler, B. M. Velichkovsky, R. Stelzer, S. Pannasch, and J. R. Helmert, "The Influence of the Stereo Base on Blind and Sighted Reaches in a Virtual Environment," *ACM Trans. Appl. Percept.*, vol. 12, no. 2, 3 2015. [Online]. Available: https://doi.org/10.1145/2724716

[12] G. Bruder, A. Pusch, and F. Steinicke, "Analyzing Effects of Geometric Rendering Parameters on Size and Distance Estimation in On-Axis Stereographics," in *Proceedings of the ACM Symposium on Applied Perception*, ser. SAP '12. New York, NY, USA: Association for Computing Machinery, 2012, p. 111–118. [Online]. Available: https://doi.org/10.1145/2338676.2338699

[13] G. Gonçalves, H. Coelho, P. Monteiro, M. Melo, and M. Bessa, "Systematic Review of Comparative Studies of the Impact of Realism in Immersive Virtual Experiences," *ACM Comput. Surv.*, vol. 55, no. 6, 12 2022. [Online]. Available: https://doi.org/10.1145/3533377

[14] A. Rochow, M. Schwarz, M. Schreiber, and S. Behnke, "VR Facial Animation for Immersive Telepresence Avatars," in *2022 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, Kyoto, Japan, 2022, pp. 2167–2174.

[15] M. Schwarz, C. Lenz, R. Memmesheimer, B. Pätzold, A. Rochow, M. Schreiber, and S. Behnke, "Robust Immersive Telepresence and Mobile Telemanipulation: NimbRo wins ANA Avatar XPRIZE Finals," *arXiv preprint arXiv:2303.03297*, 3 2023. [Online]. Available: http://arxiv.org/abs/2303.03297

[16] R. Luo, C. Wang, C. Keil, D. Nguyen, H. Mayne, S. Alt, E. Schwarm, E. Mendoza, T. Padır, and J. P. Whitney, "Team Northeastern's Approach to ANA XPRIZE Avatar Final Testing: A Holistic Approach to Telepresence and Lessons Learned," *arXiv preprint arXiv:2303.04932*, 3 2023. [Online]. Available: http://arxiv.org/abs/2303.04932

[17] F. Vanderhaegen, "Heuristic-based method for conflict discovery of shared control between humans and autonomous systems - A driving automation case study," *Robotics and Autonomous Systems*, vol. 146, p. 103867, 2021. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0921889021001524

[18] D. A. Abbink, T. Carlson, M. Mulder, J. C. F. de Winter, F. Aminravan, T. L. Gibo, and E. R. Boer, "A Topology of Shared Control Systems—Finding Common Ground in Diversity," *IEEE Transactions on Human-Machine Systems*, vol. 48, no. 5, pp. 509–525, 2018.