2nd Workshop Toward Robot Avatars, IEEE International Conference on Robotics and Automation (ICRA), London, UK, 2023.

OCRA: An Optimization-based Customizable Retargeting Algorithm for Teleoperation

Mayumi Mohan[1] and Katherine J. Kuchenbecker[1]

***Abstract*— This paper presents a real-time optimization-based algorithm for mapping motion between two kinematically dissimilar serial linkages, such as a human arm and a robot arm. OCRA can be customized based on the target task to weight end-effector orientation versus the configuration of the central line of the arm, which we call the skeleton. A video-watching study (N=70) demonstrated that when this algorithm considers both the hand orientation and the arm skeleton, it creates robot arm motions that users perceive to be highly similar to those of the human operator, indicating OCRA would be suitable for telerobotics and telepresence through avatars.**

I. Introduction

Despite widespread scientific support for the health benefits of exercise, physical inactivity is on the rise worldwide [1]. While socially assistive robots have shown potential as exercise coaches [2], they are currently far from being usable in real-world scenarios, where they will need to be able to adapt to changing user needs [3]. These robots typically leverage pre-programmed gesture-based interactions in the form of demonstrations [4], [5], [6] to teach exercises to their users. However, the motions and responses of the robot can best be designed by exercise therapy experts who are familiar with the needs and therapy objectives of their patients. We believe that *an intuitive teleoperation system would facilitate the long-term personalization and effectiveness of such exercise robots*.

We are particularly interested in upper-body humanoid robotic exercise coaches due to their ability to facilitate social-physical human-robot interaction [7]. Additionally, motion-capture-based teleoperation is known to be the most effective and intuitive methodology to enable a human operator to teach gestures to a robot [8]. The so-called retargeting problem has been studied extensively by both roboticists [9] and animators [10]. Roboticists have created algorithms that rely on inverse kinematics [11], [12], optimization [13], and morphological adaptation between kinematic chains [14], [15]. However, there exists a need for a system that not only is easy to use for non-experts but also works in real time and has been carefully evaluated [9]. Thus, through this paper, we briefly explain and evaluate OCRA, our Optimization-based Customizable Retargeting Algorithm for teleoperation. We plan to release our code open-source with the final publication associated with this project. OCRA is ROS1 compatible and can easily be adapted to other robots with minimal modifications to our code.

This work is partially supported by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germany's Excellence Strategy – EXC 2120/1 – 390831618.

[1]Mayumi Mohan and Katherine J. Kuchenbecker are with the Haptic Intelligence Department of the Max Planck Institute for Intelligent Systems, Stuttgart, Germany {maymohan,kjk}@is.mpg.de

II. Algorithm

We created an optimization-based kinematic retargeting algorithm that functions in real time. OCRA takes inspiration from the method proposed by Tosun et al. [16] due to its generalizability to arbitrary serial robots. However, that method ignores end-effector orientation, matches the user pose imperfectly, and does not function in real time. Our algorithm requires only the current configurations of both the input and output kinematic chains to perform retargeting.

Assuming the human arm has been appropriately scaled and transformed so the shoulder joint coincides with the robot's first joint, OCRA minimizes the retargeting error (ϵ_R), which is a weighted sum of the squared arm skeleton error (ϵ_s) and the squared hand orientation error (ϵ_o):

$$\epsilon_R = \alpha\epsilon_s^2 + \beta\epsilon_o^2. \tag{1}$$

Each type of error (skeleton, orientation) is normalized by the maximum possible value and is thus unitless. The constant weights (α, β) are customized by the user before teleoperation and are subject to the following constraints:

$$0 \leq \alpha \leq 1, \quad 0 \leq \beta \leq 1, \quad \alpha + \beta = 1. \tag{2}$$

At each time step, the algorithm's goal is to find the robot joint values that minimize (1) for the operator's current pose. The optimization employs Newton's conjugate-gradient method [17] and uses the previous time step's solution as the next initial guess. This approach functions well in real time, achieving a mean update rate of 25.89 Hz on our system.

Arm Skeleton Error: This error term is a curve similarity metric that measures how different the current configuration of the source chain (human arm) is from that of the target chain (robot arm) based on Fréchet's distance [18], [19]. We define the arm skeleton error as the sum of the shortest distances from each joint of the source kinematic chain to the target chain and vice versa. Let s_i be the shortest positive distance from the ith joint of the source chain to the closest segment of the target chain, and let t_j be the shortest distance from the target chain's joint j to the closest segment of the source chain. The source chain has m joints, and the target chain has n joints. To represent the maximum possible value, the normalization factor, ℓ, is the sum of the cumulative sums of the segment lengths of the source and target chains. The normalized skeleton error can then be defined as follows:

$$\epsilon_s = \frac{\sum_{i=1}^{m} s_i + \sum_{j=1}^{n} t_j}{\ell}. \tag{3}$$

End-effector Orientation Error: The orientation of the hand plays a vital role in both task-based teleoperation and social interaction. Thus, we add a second metric that quantifies errors in the orientation of the end-effector frame. Since robot and motion-capture manufacturers choose end-effector frames somewhat arbitrarily, we first define a custom frame that is positioned and oriented in a suitable way on each end-effector. For our application, it is in the middle of the human's palm and at the center of the robot's palm, which coincides with the base of the parallel jaw gripper. Let q_ς and q_τ be the quaternions representing the orientations of the source and target end-effector frames, respectively. We can then obtain the quaternion $Q_d = q_\varsigma / q_\tau$ that represents the orientation difference between the two frames. From Q_d we calculate the positive relative angle between these frames, θ_d, using the axis-angle representation. This angle can have a maximum value of π radians. Thus, normalized end-effector orientation error is defined as $\epsilon_o = \theta_d/\pi$.

III. Implementation and Study Design

We tested OCRA on a Rethink Robotics Baxter Research Robot [20] using human motion-capture data collected via an Xsens Link [21]. Human perception of the motion of others stems from our immense experience in performing activities ourselves [22]. Thus, an intuitive teleoperation system should create human-like motions. We conducted a two-factor within-subjects study to discover how users perceived robot motions created by OCRA (with different weights) compared to those of the human operator.

70 online participants (age: M=35.4, SD=12.3) were presented with videos of two-armed human motion positioned next to and synchronized with a corresponding Baxter robot motion. They saw four exercise-related motions of differing complexity, each rendered with four values for the arm-skeleton weight: $\alpha = [0.001, 0.33, 0.67, 1.00]$. Fig. 1 shows sample images from the study, which was designed for an online format following relevant guidelines [23]. Participants rated how similar each robot motion was to the human motion on a visual analog scale from 0 (worst) to 100 (best) with five standard labels ranging from "very different" to "very similar". Additionally, we posed final questions to identify the participant's reasoning for their ratings. We hypothesized that *retargeting that combines both types of errors will outperform retargeting that considers only skeleton or only orientation error* (H1) and that *the relative performance of different weight sets will be consistent across motions* (H2). These hypotheses were tested via an aligned-rank-transform two-way ANOVA [24].

IV. Results

Selected results from the user study can be seen in Fig. 2. Participants rated conditions having a substantial combination of skeleton and orientation error to be significantly more similar to the human motion than the tested conditions with purely orientation or purely skeleton error. User comments also demonstrate that participants cared most about the orientation of the robot's hand and the shape of

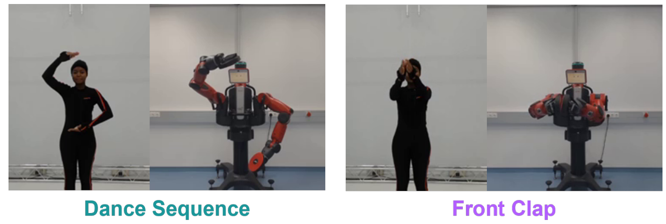

Fig. 1. Sample screenshots of two movement videos shown in the study for the $\alpha = 0.67$ condition. The two-armed robot tries to match the human operator's hand orientations while prioritizing matching their arm skeletons.

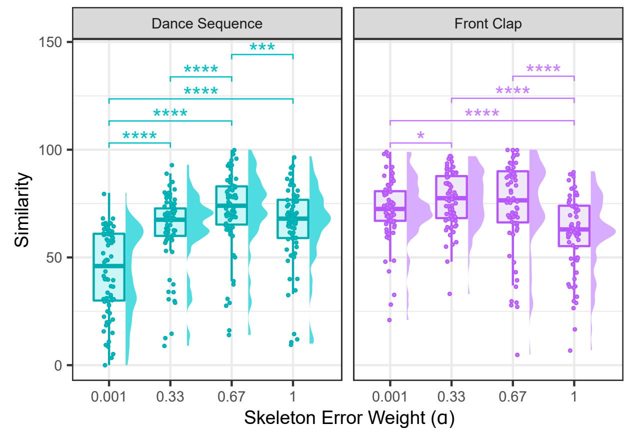

Fig. 2. Participant ratings for two of the movements shown in the user study, each with four weights for skeleton error. Significant pairwise differences between medians (the central bars in the boxes) are indicated by lines with stars: $\star$ $p < 0.05$, $\star\star\star$ $p < 0.001$, and $\star\star\star\star$ $p < 0.0001$.

its arm skeleton, rather than the position of the hand or the smoothness of the motion. These observations support our first hypothesis. Thus, skeleton error and orientation error can indeed be considered suitable ingredients for creating human-like robot motion from motion-capture data.

Interestingly, our second hypothesis (H2) was not fully substantiated by our results (see Fig. 2). There were variations in participant ratings across the two presented motions, which we attribute to their different complexity levels. Users distinguished between $\alpha = 0.33$ and $\alpha = 0.67$ when the motions were more elaborate, as seen in the dance sequence.

V. Conclusions

This workshop paper briefly presented OCRA, a customizable algorithm to retarget human arm motion onto a robot arm. A perceptual user experiment with videos showed that our algorithm produces human-like motions for moderate weight values. This algorithm is generalizable to all rigid robots with revolute joints; it may need to be adapted to work well with prismatic joints. Our next step is to evaluate this teleoperation system via a real-time usability experiment.

Acknowledgments

The authors thank Chitvan Killawala, Ivan Oreshnikov, Gokhan Serhat, Behnam Khojasteh, and Towela Tembo for supporting this research project. We also thank the International Max Planck Research School for Intelligent Systems (IMPRS-IS) for supporting Mayumi Mohan.

REFERENCES

[1] C. Ozemek, C. J. Lavie, and Ø. Rognmo, "Global physical activity levels – need for intervention," *Progress in Cardiovascular Diseases*, vol. 62, no. 2, pp. 102–107, 2019.

[2] M. J. Matarić and B. Scassellati, "Socially assistive robotics," in *Springer Handbook of Robotics*. Springer, 2016, pp. 1973–1994.

[3] A. M. Okamura, M. J. Matarić, and H. I. Christensen, "Medical and health-care robotics," *IEEE Robotics & Automation Magazine*, vol. 17, no. 3, pp. 26–37, 2010.

[4] J. C. Pulido, C. Suarez-Mejias, J. C. Gonzalez, A. D. Ruiz, P. F. Ferri, M. E. M. Sahuquillo, C. E. R. De Vargas, P. Infante-Cossio, C. L. P. Calderon, and F. Fernandez, "A socially assistive robotic platform for upper-limb rehabilitation: a longitudinal study with pediatric patients," *IEEE Robotics & Automation Magazine*, vol. 26, no. 2, pp. 24–39, 2019.

[5] Y. Matsusaka, H. Fujii, T. Okano, and I. Hara, "Health exercise demonstration robot TAIZO and effects of using voice command in robot-human collaborative demonstration," in *Proceedings of the IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)*, 2009, pp. 472–477.

[6] L. Lewis, T. Metzler, and L. Cook, "Evaluating human-robot interaction using a robot exercise instructor at a senior living community," in *Proceedings of the International Conference on Intelligent Robotics and Applications*. Springer, 2016, pp. 15–25.

[7] N. T. Fitter, M. Mohan, K. J. Kuchenbecker, and M. J. Johnson, "Exercising with Baxter: Preliminary support for assistive social-physical human-robot interaction," *Journal of NeuroEngineering and Rehabilitation*, vol. 17, no. 19, Feb. 2020.

[8] E. Hedlund, M. Johnson, and M. Gombolay, "The effects of a robot's performance on human teachers for learning from demonstration tasks," in *Proceedings of the ACM/IEEE International Conference on Human-Robot Interaction (HRI)*. ACM, mar 2021, pp. 207–215.

[9] K. Darvish, L. Penco, J. Ramos, R. Cisneros, J. Pratt, E. Yoshida, S. Ivaldi, and D. Pucci, "Teleoperation of humanoid robots: A survey," *IEEE Transactions on Robotics*, pp. 1–22, 2023.

[10] J.-S. Monzani, P. Baerlocher, R. Boulic, and D. Thalmann, "Using an intermediate skeleton and inverse kinematics for motion retargeting," in *Proceedings of the Eurographics Computer Graphics Forum*, vol. 19. Wiley Online Library, 2000, pp. 11–19. [Online]. Available: https://doi.org/10.1111/1467-8659.00393

[11] Z. Ju, C. Yang, Z. Li, L. Cheng, and H. Ma, "Teleoperation of humanoid Baxter robot using haptic feedback," in *Proceedings of the IEEE International Conference on Multisensor Fusion and Information Integration for Intelligent Systems (MFI)*, 2014, pp. 1–6.

[12] C. Lenz and S. Behnke, "Bimanual telemanipulation with force and haptic feedback through an anthropomorphic avatar system," *Robotics and Autonomous Systems*, vol. 161, p. 104338, 2023.

[13] B. Dariush, M. Gienger, A. Arumbakkam, Y. Zhu, B. Jian, K. Fujimura, and C. Goerick, "Online transfer of human motion to humanoids," *International Journal of Humanoid Robotics*, vol. 6, no. 02, pp. 265–289, 2009.

[14] P. Sousa, J. L. Oliveira, L. P. Reis, and F. Gouyon, "Humanized robot dancing: humanoid motion retargeting based in a metrical representation of human dance styles," in *Progress in Artificial Intelligence: Proceedings of the 15th Portuguese Conference on Artificial Intelligence, EPIA*. Lisbon, Portugal: Springer, Oct. 2011, pp. 392–406.

[15] G. Lentini, A. Settimi, D. Caporale, M. Garabini, G. Grioli, L. Pallottino, M. G. Catalano, and A. Bicchi, "Alter-Ego: A Mobile Robot With a Functionally Anthropomorphic Upper Body Designed for Physical Interaction," *IEEE Robotics & Automation Magazine*, vol. 26, no. 4, pp. 94–107, 2019.

[16] T. Tosun, R. Mead, and R. Stengel, "A general method for kinematic retargeting: Adapting poses between humans and robots," in *Proceedings of the ASME International Mechanical Engineering Congress and Exposition (IMECE)*, 2014. [Online]. Available: https://doi.org/10.1115/IMECE2014-37700

[17] S. G. Nash, "Newton-type minimization via the Lanczos method," *Journal on Numerical Analysis*, vol. 21, no. 4, pp. 770–788, 1984.

[18] M. Fréchet, "Sur la distance de deux lois de probabilité," *Comptes Rendus Hebdomadaires des Seances de L Academie des Sciences*, no. 6, pp. 689–692, 1957.

[19] H. Alt and M. Godau, "Computing the Fréchet distance between two polygonal curves," *International Journal of Computational Geometry & Applications*, vol. 5, no. 01n02, pp. 75–91, 1995.

[20] "Rethink Robotics Baxter robot," https://robots.ieee.org/robots/baxter/, accessed: 2023-03-27.

[21] Movella Inc., "Xsens MVN Link," https://www.movella.com/products/motion-capture/xsens-mvn-link, accessed: 2023-03-27.

[22] R. Blake and M. Shiffrar, "Perception of human motion," *Annual Review of Psychology*, vol. 58, pp. 47–73, 2007.

[23] International Telecommunication Union, "Methodology for the subjective assessment of the quality of television pictures," https://www.itu.int/rec/R-REC-BT.500-14-201910-I/en, accessed: March 29, 2023.

[24] J. O. Wobbrock, L. Findlater, D. Gergle, and J. J. Higgins, "The aligned rank transform for nonparametric factorial analyses using only ANOVA procedures," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 2011, pp. 143–146.