Robust Teleoperation of an Anthropomorphic Robotic Hand through Gesture Classification and Continuous Mapping

Anonymous Authors

***Abstract*—Teleoperation allows for control of an anthropomorphic robot hand by leveraging the human ability to perform complex manipulation tasks. However, the mismatch between the human and robot motions affects the intuitiveness and performance of teleoperation tasks. We propose a hybrid Gesture-based mapping method, where a classifier first identifies a gesture from a predefined discrete set, and then a regression model determines the continuous gesture motion progression. The discrete nature of the classification allows for the robot to execute a predefined motion, matching the intention of the human, while the regression model allows for a continuous and thus smooth and intuitive experience. To benchmark the proposed method, we have also developed a baseline method which directly maps each human finger motion to the robot. We systematically evaluate the two methods by varying the operator to include those with varying experience levels and hand sizes, introducing signal delay of zero or one second, and testing across 20 object geometries. The proposed method outperforms the baseline through a 29% decrease in average grasping time for an expert (with and without a signal delay), an 34% increase in success rate, and 64% increase in System Usability Scale (SUS) scores for novice users.**

***Index Terms*—Robust Teleoperation, Dexterous Manipulation, Anthropomorphic Hand**

I. Introduction

ACHIEVING human-like dexterous manipulation is a longstanding challenge in robotics, often regarded as a pivotal milestone to make large impacts on flexible automation [1], agrifood [2], and human-robot interactions [3]. However, the control of an anthropomorphic robot hand is challenging due to its high degrees of freedom [4]. Teleoperation is the first step towards performing complex manipulation tasks with a dexterous robotic hand. Leveraging human expert through teleoperation can advance the use of dexterous anthropomorphic hands and also serve as a means of collecting data for developing learning based controllers [5], [6].

Mapping human motion to robot motion is essential for effective teleoperation of robotic hands. Two main approaches dominate the state of the art: continuous and discrete mapping. Continuous mapping can be joint-based [7], pose-based (Cartesian or synergy-based) [8], [9], or object(task)-based [10], [11], [12]. The second is discrete mapping, which relates specific human poses to predefined robot actions. Continuous methods generate smooth and flexible motion but are sensitive to human–robot morphology mismatch [13], [14]. Compared to simple joint-to-joint mapping, pose-based methods better capture human intent by considering finger positions [15] or low-dimensional synergies [16], though at higher control complexity. Discrete gesture mapping enables precise execution of a limited set of grasps but can produce unnatural transitions and often requires many gestures (10–30) [17], [18], [19]. More recently, neural networks have been proposed to unify mappings across robotic systems [20], but morphology mismatch remains unavoidable. Another challenge is benchmarking teleoperation performance, as differing setups often prevent fair comparisons across methods.

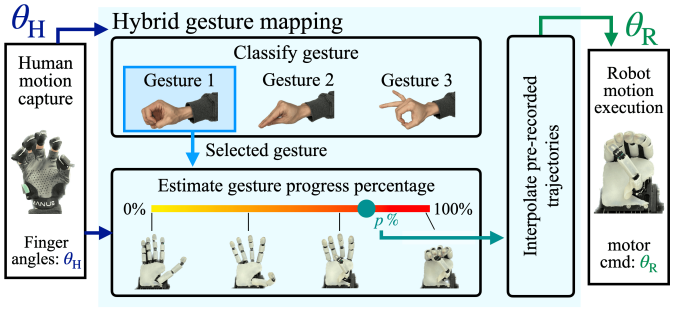

Fig. 1. Hybrid Gesture-based mapping methodology. Human motion data is first classified, and then the progress of the grasp $p\%$ is estimated. The selected gesture and estimated $p\%$ are used to return the action of the hand based upon pre-recorded trajectories of each grasp.

Robustness in teleoperation can be improved not only by achieving minimal motion mapping mismatch, but also by enhancing the system's ability to safely and reliably interact with the world (from the hardware level). For most robotic systems which operate under position control, environmental interactions can be challenging. To regulate forces for varying object geometry precisely, force estimation algorithms [21], [22] or haptic feedback [23], [24] are used. Compliant robotic hands such as the RBO Hand 3 [25] and the Pisa/IIT soft hand [26] could be a solution to demonstrate robust interaction through a simple kinematic input; however they can be limited by repeatability and controllability in the actuation.

In this work, we propose a hybrid gesture-based mapping method that differs from approaches relying only on continuous or discrete mapping, which suffer from morphology mismatch or unnatural transitions. Our method combines the smoothness of continuous mapping with the accuracy and simplicity of discrete mapping for intuitive teleoperation (see Fig. 1). During teleoperation, the operator's intent is classified from a discrete gesture set throughout the motion, not only at the final pose. A grasp percentage p is estimated to continuously actuate the robotic hand. An SVM is used for gesture classification and a regression model for p, enabling responsive mapping across different hand sizes.

To evaluate the method, we compare it with a finger-wise direct joint-to-joint mapping baseline [7], [27] (Fig. 7). Both are tested on the compliant anthropomorphic ADAPT Hand [28], whose distributed compliance enables adaptive finger and whole-hand motion while reducing the need for precise haptic control or exact mapping. Using gestures from the human

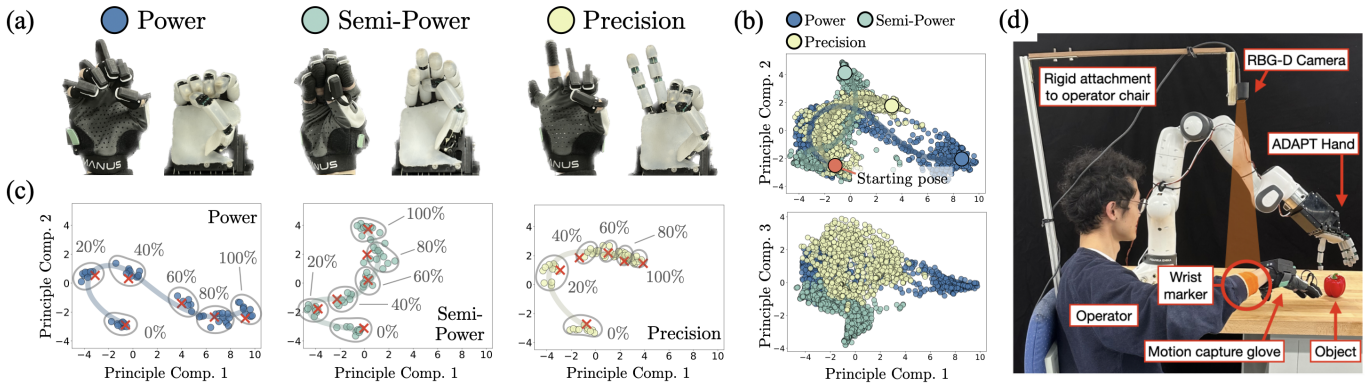

Fig. 2. (a): Three pre-defined hand grasping gestures. (b): Training data for classifier. (c): Training data for regression model. (PCA is applied only for visualization in (b) (c), not for the raw data). (d) Experimental setup for teleoperation where the user wears a glove and wrist marker to perform grasping tasks. The setup includes an ADAPT Hand mounted on a Franka Research 3 (FR3) arm. The Manus Quantum Metaglove captures finger joint pose θ_H, and the teleoperation model (running on a laptop) generates motor commands θ_R for the robot hand. A Zed2 RGB-D camera estimates wrist 3D position via HSV-based detection of an orange wrist band, fused with glove IMU data, to control the FR3 end-effector position. Details see Appendix B.

grasp taxonomy [29], we show that three gestures are sufficient for grasping objects with diverse geometries. Unlike Cartesian pose [8], [9] or object-based methods [10], [11], [12], our approach does not require fingertip sensing or kinematic models; given the ADAPT Hand's compliance, accurate modeling is difficult, making joint mapping a practical baseline. The contributions of this work can be summarized as:

- Hybrid gesture mapping which integrates discrete gesture recognition with continuous "grasp percentage" control for smooth, precise, and intuitive teleoperation.
- Learning-based & calibration-free teleoperation through SVM classification and regression to generalize across hand sizes of teleoperators without relying on kinematic models or external sensors.
- Data-efficient transferability as the mapping adapts to new robot hands with minimal additional data collection and does not require full model retraining.
- Demonstration on the ADAPT hand and evaluation against a finger-wise joint-mapping baseline, demonstrating faster, more intuitive grasping performance.

The proposed and baseline methods are evaluated across operators in teleoperation tasks with objects of varying geometry. We measure success rate and operation time for experts and novices, including an expert condition with artificial communication delay to simulate latency. The hybrid gesture-based method outperforms the baseline, reducing expert grasping time by 29% and increasing novice success rate by 34%.

II. Methods

This section presents the hybrid gesture-based mapping method (main contribution) and the direct mapping baseline, along with their integration with a 7 DoF robot arm to match the operator's wrist pose. The mapping between human joint poses θ_H and robot commands θ_R is applied to the anthropomorphic ADAPT Hand [28], which has 20 joints and 12 actuators. The hybrid method uses an SVM classifier and a linear regression model, while the baseline relies on neural networks. Details on data collection, model training, and robot arm control are described below:

A. Hybrid Gesture Mapping

In the hybrid gesture-based mapping method (Fig. 1), human hand poses $\theta_H \in \mathbb{R}^{20}$ are captured via a glove. A classifier predicts the current grasp type from a discrete set, each representing a continuous motion from open to closed. A gesture-specific regressor then estimates progression ($p\%$), from 0% (open) to 100% (grasp). The robot command $\theta_R \in \mathbb{R}^{12}$ is obtained by interpolating pre-recorded pose sequences. The classifier and regressor are trained on one expert and applied to other users.

1) Gesture selection: Inspired by human grasp taxonomy [29], [30], three gestures are selected (Fig. 2a) to cover major object geometries: power (for large objects), semi-power (for flat objects), and precision (thumb with index/middle fingers, for small objects). Using only three gestures enables grasping diverse objects with minimal types, similar to [28]. The ADAPT hand's compliance allows each gesture to adapt to object shape, improving robustness and reducing required gestures, while additional gestures can be incorporated for other robot hands.

2) Classification process: The human hand joint poses θ_H form a nonlinear trajectory in 20 dimensional(D) space from open to closed grasp. A Support Vector Machine (SVM) is used for classification due to its simplicity and robustness to noise on nonlinear boundaries. Its parameters are set to $1/(\text{number of features} \cdot \text{data variance})$ for γ and $C = 1$. Since SVM training solves a convex optimization problem [31], it guarantees a global minimum and reproducibility.

Fig. 2b shows the three gesture trajectories, visualized using PCA reducing 20D to 3D. The dataset is collected by an expert (see B1 2), with 15 repetitions per gesture (800 samples each, 2400 total) as shown in Fig. 8a (left). The data is nonlinear and noisy but forms three separable trajectories, motivating SVM boundary-based classification. Once trained, the SVM classifies not only end poses but the entire gesture trajectories.

3) Regression process: Given correct gesture classification, a linear regression model estimates the gesture progress percentage. A first-degree polynomial linear regression model is chosen for its simplicity and trade-off between complexity

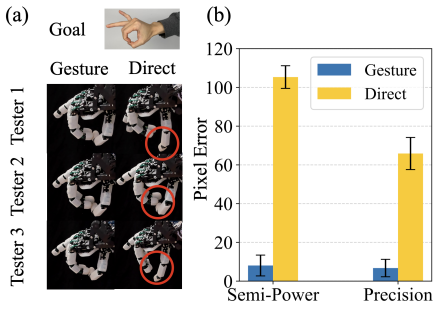

Fig. 3. (a) Visual demonstration of the variability and error that emerges when users are required to perform a precision grasping gesture ("pinch grasp") using Gesture-based and Direct mapping, images of 3 out of 5 testers are shown. (b) Quantitative assessment of the error, as measured in the pixel error between contact points of the fingers for two grasps and the two teleoperation methods.

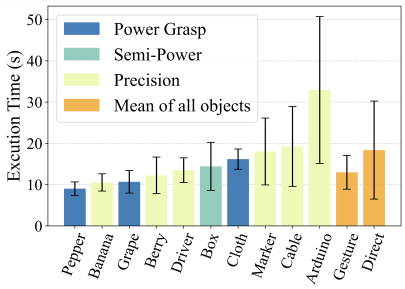

Fig. 4. Average time to complete the grasping task for each object with both the gesture based and baseline method also with and without delay (last two orange bars show the average values across all objects using Gesture-based and Direct mapping methods). The classification of the grasp type (in Gesture-based method) - the most suitable type for each corresponding object - is also shown

and representation; its training solves a convex optimization problem [32], guaranteeing a global minimum and reproducibility. Training data (Fig. 2c) is collected by an expert (see B1 2) performing each gesture at discrete progress levels $(0\%, 20\%, 40\%, 60\%, 80\%, 100\%)$, where 0% is the initial motion and 100% the fully closed grasp. This process (Fig. 8a, middle) is repeated 10 times, yielding 60 samples per gesture. The data for the three gestures is shown in Fig. 2c, and intermediate progress values are inferred via regression.

4) Robot Gesture Playback: Since the classifier and regressor only output the detected gesture and progress p, the corresponding robot motor command θ_R must be generated. A "signal mixer" control box (Fig. 8c) is used to manually adjust servo positions and record θ_R. For each gesture, robot motor commands are recorded at six waypoints (0%, 20%, ..., 100%) matching the regression labels, as shown in Fig. 8c. Intermediate commands are then obtained via linear interpolation between these waypoints. Although manual recording is required for each gesture, it is a one-time process. When switching to a different robot hand, only this step must be repeated to collect new motor trajectories, while the classifier and regressor remain unchanged, improving adaptability across robot morphologies.

Additional method details about the baseline (direct mapping) and the developed pipeline of our method for new robot hands are provided; please refer to A.

III. Results

This section presents the experimental results. The teleoperation performance of the proposed method is first evaluated in isolation (without grasping or environmental interaction), focusing on classification accuracy and mapping accuracy. It is then assessed in grasping tasks performed by both expert and novice operators.

A. Motion Mapping

Firstly, classification accuracy is evaluated with five operators of different hand sizes (Table I). Each participant performs the three gestures repeatedly, with classification recorded while moving from rest (0%) to full grasp (100%). Near 0%, confusion occurs due to similar starting poses, but accuracy reaches 100% near 100% completion. This confirms correct classification without overfitting, as training and test data share a similar distribution across different hand sizes. Adding more gestures may reduce accuracy, and misclassifications can degrade regression-based mapping. To address this, misclassified samples should be collected and added to the training set, enabling iterative retraining until robust performance is achieved.

Secondly, mapping accuracy is compared between Gesture-based and Direct mapping methods. Fig. 3 shows thumb–index pixel error across five testers. The Gesture-based method consistently achieves lower error (Fig. 3b), reaching only 7.6% and 10.2% of Direct Mapping error for semi-power and precision grasps, respectively. Direct mapping also shows high variability in final robot poses across operators (Fig. 3a), whereas Gesture-based mapping is more consistent, since it follows a shared trajectory and is only influenced by the regression-based progress estimate given correct classification.

B. Grasping Assessment

In addition to the isolated evaluation, both teleoperation methods are tested in full grasping tasks with objects of varying geometries. The robot hand starts about 10 cm above the object and is teleoperated to grasp and lift it. The operator may attempt multiple grasps within a 60 s window; otherwise the trial is marked as a failure. Execution time is used as the performance metric.

1) Expert Teleoperation Tests: In experiments with the expert operator, the robot is teleoperated to grasp all items in the *core* object set using both mapping methods. Each object is grasped five times consecutively and manually returned to its original position. To evaluate robustness, the same experiment is repeated with an artificial 1 s delay to simulate communication latency in deployment.

a) Average time of grasping tests: Fig. 4 shows the average grasping time across methods with and without artificial delay. Bar colors indicate the hand gesture used in the Gesture-based mapping. The average time reflects task difficulty, with the Red pepper being the easiest and the Arduino the most difficult. The last two orange bars show that the Gesture-based method outperforms Direct mapping, reducing average grasping time by 29% across all objects.

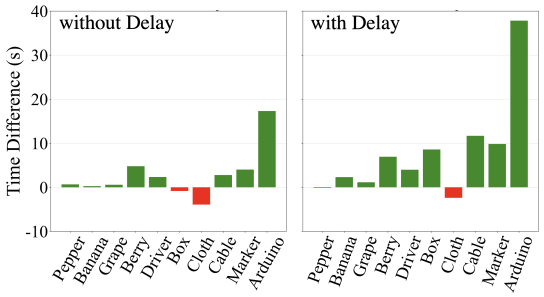

Fig. 5. Execution time difference between two teleopration methods, time difference: $t_{\text{direct}} - t_{\text{gesture}}$. In the experiment of right subfigure, a artificial delay is added to investigate the potential impact of communication or data processing delays on teleoperation performance in real applications.

b) Execution time difference between two teleoperation mapping: The average time difference between Gesture-based and Direct mapping is shown in Fig. 5. Green and red bars indicate whether the Gesture-based method reduces execution time. Without artificial delay (Fig.5 left), our method generally performs faster. For easy objects (all dimensions >5 cm), especially with Power Grasp, the difference is negligible (<3 s), while for the most difficult object (Arduino chip), it saves 17.3 s due to reduced alignment difficulty in Direct mapping. This stems from large mismatch errors in Direct mapping. With artificial delay (Fig.5 right), the advantage increases as manual compensation becomes harder, particularly for precision grasps (USB cable, Marker pen, Arduino chip), where the time savings are 4.2, 2.4, and 2.2 times higher than without delay. For objects like Red pepper and Grapes, where precise positioning is less critical, improvements are small or negative. The baseline performs better for soft objects such as clothing, since they tolerate varied finger configurations and do not require precise grasp poses. In these cases, the advantage of predefined gesture control is reduced, while Direct mapping allows more intuitive and adaptive grip adjustment.

2) Beginner User Teleoperation Tests: To evaluate usability for beginners, ten users are selected. Each receives a brief introduction and two practice trials. The order of Direct and Gesture-based methods is randomized to avoid learning bias. Each participant performs five grasp attempts on a red pepper and a marker pen (highly different geometries), while success rate and execution time are recorded. The system starts from a fixed initial state, and timing stops once the object is lifted. No artificial delay is used.

Results are shown in Fig. 6. For the red pepper (power grasp), both methods achieve similar times (7 s). For the marker pen (precision grasp), beginners often fail, with failures counted as 60 s. Overall, Gesture-based teleoperation reduces average time by 2.1× compared to Direct mapping and improves success rate by 34% in precision tasks.

Fig. 6b shows teleoperation time across five trials for ten beginner users in the precision grasp task. With Direct mapping, 4–5 users fail in the first two trials. Although performance improves over time as users learn to compensate for mismatch errors, it is inconsistent and non-monotonic, reaching an average of 35.7 s by the fifth trial, indicating that adaptation is difficult within a short period. In contrast,

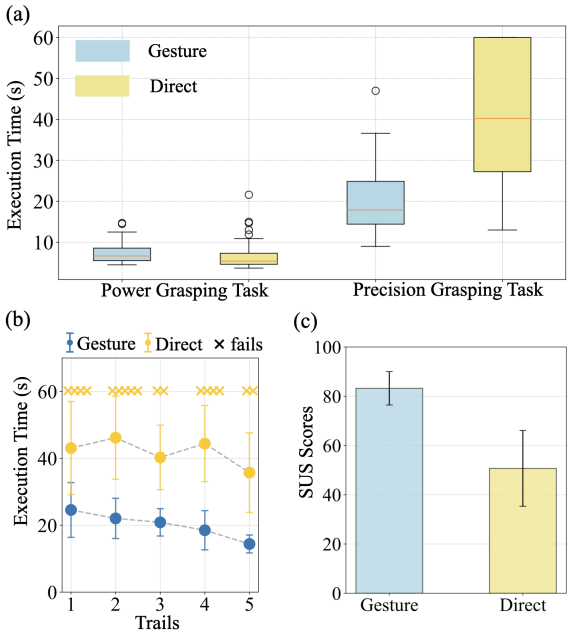

Fig. 6. Execution time and SUS scores of 10 beginner operators in grasping tasks, (a): average execution time of power (for large object) and precision grasping (for small object) using two teleoperation methods. (b): Average of individual execution time (small object) across all operators of 5 repetitions of the precision grasping task. (c): System Usability Scale (SUS) scores of two teleoperation methods

our method yields lower and monotonically decreasing times, with no failures across trials, attributed to reduced mismatch in gesture-based replay and improved intuitiveness.

System Usability Scale (SUS) scores in Fig. 6c show 83.2 for Gesture-based and 50.7 for Direct Mapping. According to [33], 50.7 is in the "OK" range (avg. 50.9), while 83.2 falls in the "Excellent" range (avg. 85.5), highlighting the superior usability of the proposed method.

Additional results on the generalizability of both methods across a wide range of objects are provided in Appendix C.

IV. Discussion & Conclusion

We present a teleoperation framework for an anthropomorphic robotic hand in grasping tasks. To improve robustness across operators and reduce mapping mismatch, we use a gesture classifier with a regression model to estimate progression and generate smooth robot trajectories. Compared to direct mapping, the method reduces execution time by 29% (experts) and 47.6% (novices), and improves novice SUS scores by 64%, with larger gains for precise small-object tasks and under communication delay. When adapting to new robot hands, only pre-recorded trajectories are updated, while models remain unchanged. Despite restricting the action space, results on 20 diverse objects show minimal impact on success.

Future work will extend the gesture set for more general manipulation and enable large-scale demonstrations data collection for imitation learning. The framework also supports offline human-to-robot mapping, making it promising for learning from video data.

ACKNOWLEDGEMENTS

This research project was supported by the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No. 945363.

REFERENCES

[1] B. Burger et al., "A mobile robotic chemist," *Nature*, vol. 583, no. 7815, pp. 237–241, 2020.

[2] C. Armanini et al., "Soft robotics for farm to fork: Applications in agriculture & farming," *Bioinspiration & Biomimetics*, 2024.

[3] M. BARAKAZI, "The use of robotics in the kitchens of the future: The example of'moley robotics'," *Journal of Tourism & Gastronomy Studies*, vol. 10, no. 2, pp. 895–905, 2022.

[4] C. Piazza, G. Grioli, M. G. Catalano, and A. Bicchi, "A century of robotic hands," *Annual Review of Control, Robotics, and Autonomous Systems*, vol. 2, pp. 1–32, 2019.

[5] W. Si, N. Wang, and C. Yang, "A review on manipulation skill acquisition through teleoperation-based learning from demonstration," *Cognitive Computation and Systems*, vol. 3, no. 1, pp. 1–16, 2021.

[6] C. Pan, K. Junge, and J. Hughes, "Vision-language-action model and diffusion policy switching enables dexterous control of an anthropomorphic hand," *arXiv preprint arXiv:2410.14022*, 2024.

[7] M. V. Liarokapis, P. K. Artemiadis, and K. J. Kyriakopoulos, "Mapping human to robot motion with functional anthropomorphism for teleoperation and telemanipulation with robot arm hand systems," in *2013 IEEE/RSJ International Conference on Intelligent Robots and Systems*, IEEE, 2013, pp. 2075–2075.

[8] R. Li, H. Wang, and Z. Liu, "Survey on mapping human hand motion to robotic hands for teleoperation," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 32, no. 5, pp. 2647–2665, 2021.

[9] M. V. Liarokapis, P. Artemiadis, C. P. Bechlioulis, and K. J. Kyriakopoulos, "Directions, methods and metrics for mapping human to robot motion with functional anthropomorphism: A review," *School of Mechanical Engineering, National Technical University of Athens, Tech. Rep*, 2013.

[10] F. Ficuciello, A. Villani, T. Lisini Baldi, and D. Prattichizzo, "A human gesture mapping method to control a multi-functional hand for robot-assisted laparoscopic surgery: The musha case," *Frontiers in Robotics and AI*, vol. 8, p. 741 807, 2021.

[11] G. Salvietti, M. Malvezzi, G. Gioioso, and D. Prattichizzo, "On the use of homogeneous transformations to map human hand movements onto robotic hands," in *2014 IEEE International Conference on Robotics and Automation (ICRA)*, IEEE, 2014, pp. 5352–5357.

[12] G. Gioioso, G. Salvietti, M. Malvezzi, and D. Prattichizzo, "Mapping synergies from human to robotic hands with dissimilar kinematics: An approach in the object domain," *IEEE Transactions on Robotics*, vol. 29, no. 4, pp. 825–837, 2013.

[13] S. Li, R. Rameshwar, A. M. Votta, and C. D. Onal, "Intuitive control of a robotic arm and hand system with pneumatic haptic feedback," *IEEE Robotics and Automation Letters*, vol. 4, no. 4, pp. 4424–4430, 2019.

[14] S. Ekvall and D. Kragic, "Interactive grasp learning based on human demonstration," in *IEEE International Conference on Robotics and Automation, 2004. Proceedings. ICRA'04. 2004*, IEEE, vol. 4, 2004, pp. 3519–3524.

[15] Y. Inoue, F. Kato, and S. Tachi, "Master-slave robot hand control method based on congruence of vectors for telexistence hand manipulation," in *2019 IEEE International Symposium on Measurement and Control in Robotics (ISMCR)*, IEEE, 2019, A1–1.

[16] C. Meeker, T. Rasmussen, and M. Ciocarlie, "Intuitive hand teleoperation by novice operators using a continuous teleoperation subspace," in *2018 IEEE International Conference on Robotics and Automation (ICRA)*, IEEE, 2018, pp. 5821–5827.

[17] M. T. Wolf, C. Assad, M. T. Vernacchia, J. Fromm, and H. L. Jethani, "Gesture-based robot control with variable autonomy from the jpl biosleeve," in *2013 IEEE International Conference on Robotics and Automation*, IEEE, 2013, pp. 1160–1165.

[18] M. A. Simao, O. Gibaru, and P. Neto, "Online recognition of incomplete gesture data to interface collaborative robots," *IEEE Transactions on Industrial Electronics*, vol. 66, no. 12, pp. 9372–9382, 2019.

[19] K. Higashi, K. Koyama, R. Ozawa, K. Nagata, W. Wan, and K. Harada, "Functionally divided manipulation synergy for controlling multi-fingered hands," in *2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, IEEE, 2020, pp. 9190–9197.

[20] Y. Qin et al., "Anyteleop: A general vision-based dexterous robot arm-hand teleoperation system," *Robotics: Science and Systems*, 2023.

[21] C. Zeng et al., "Learning compliant grasping and manipulation by teleoperation with adaptive force control," in *2021 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, IEEE, 2021, pp. 717–724.

[22] C. Zeng, S. Li, Z. Chen, C. Yang, F. Sun, and J. Zhang, "Multifingered robot hand compliant manipulation based on vision-based demonstration and adaptive force control," *IEEE Transactions on Neural Networks and Learning Systems*, 2022.

[23] C. Li, A. Fahmy, and J. Sienz, "Development of a neural network-based control system for the dlr-hit ii robot hand using leap motion," *IEEE Access*, vol. 7, pp. 136 914–136 923, 2019.

[24] J. Bimbo, C. Pacchierotti, M. Aggravi, N. Tsagarakis, and D. Prattichizzo, "Teleoperation in cluttered environments using wearable haptic feedback," in *2017 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, IEEE, 2017, pp. 3401–3408.

[25] S. Puhlmann, J. Harris, and O. Brock, "Rbo hand 3: A platform for soft dexterous manipulation," *IEEE Transactions on Robotics*, vol. 38, no. 6, pp. 3434–3449, 2022.

[26] C. Della Santina, C. Piazza, G. Grioli, M. G. Catalano, and A. Bicchi, "Toward dexterous manipulation with augmented adaptive synergies: The pisa/iit softhand 2," *IEEE Transactions on Robotics*, vol. 34, no. 5, pp. 1141–1156, 2018.

[27] S. Li et al., "Vision-based teleoperation of shadow dexterous hand using end-to-end deep neural network," in *2019 International Conference on Robotics and Automation (ICRA)*, IEEE, 2019, pp. 416–422.

[28] K. Junge and J. Hughes, "Spatially distributed biomimetic compliance enables robust anthropomorphic robotic manipulation," *Communications Engineering*, vol. 4, no. 1, p. 76, 2025.

[29] T. Feix, J. Romero, H.-B. Schmiedmayer, A. M. Dollar, and D. Kragic, "The grasp taxonomy of human grasp types," *IEEE Transactions on human-machine systems*, vol. 46, no. 1, pp. 66–77, 2015.

[30] M. Santello, M. Flanders, and J. F. Soechting, "Postural hand synergies for tool use," *Journal of neuroscience*, vol. 18, no. 23, pp. 10 105–10 115, 1998.

[31] J. Cervantes, F. Garcia-Lamont, L. Rodríguez-Mazahua, and A. Lopez, "A comprehensive survey on support vector machine classification: Applications, challenges and trends," *Neurocomputing*, vol. 408, pp. 189–215, 2020.

[32] S. Boyd and L. Vandenberghe, *Convex optimization*. Cambridge university press, 2004.

[33] J. R. Lewis, "The system usability scale: Past, present, and future," *International Journal of Human–Computer Interaction*, vol. 34, no. 7, pp. 577–590, 2018.

[34] A. Kapandji, "Clinical test of apposition and counter-apposition of the thumb," *Annales de chirurgie de la main: organe officiel des societes de chirurgie de la main*, vol. 5, no. 1, pp. 67–73, 1986.

[35] R. Meattini, R. Suarez, G. Palli, and C. Melchiorri, "Human to robot hand motion mapping methods: Review and classification," *IEEE Transactions on Robotics*, vol. 39, no. 2, pp. 842–861, 2022.

[36] B. R. Galarza, P. Ayala, S. Manzano, and M. V. Garcia, "Virtual reality teleoperation system for mobile robot manipulation," *Robotics*, vol. 12, no. 6, p. 163, 2023.

[37] A. Amaya, D. D. Arachchige, J. Grey, and I. S. Godage, "Evaluation of human-robot teleoperation interfaces for soft robotic manipulators," in *2021 30th IEEE International Conference on Robot & Human Interactive Communication (RO-MAN)*, IEEE, 2021, pp. 412–417.

APPENDIX

A. Additional Methods Details

1) Direct Mapping: Baseline: As the baseline approach, a joint-to-joint Direct mapping (similar to methods in [7], [27]) between the human and robot is established through finger-wise regression models described in Fig. 7. Whilst a straight forward method is to directly map finger joint angle from the motion capture glove to the robot hand, this method leads to very poor performance due to structural differences (especially in the thumb), variations in finger segment lengths, and the fact that the first two robot finger joints are driven together by single tendon/motor, making it difficult to achieve a one-to-one joint mapping.

a) Models for Direct Mapping: Therefore, as a baseline we propose instead a data-driven approach where a neural network is used to build this Direct mapping. For example, to control the motion of the index finger, one regression model (neural network with architecture of 32-64-32 for latent layers) maps the three joint angle measurements to the two actuator inputs of the ADAPT Hand index finger. A total of six regression networks are used, corresponding to the five fingers and the spread (abduction/adduction) motion of the fingers. The data collection for each finger is illustrated in Fig. 8b - a single measurement with the human finger joint angle θ_H and the corresponding motor command θ_R which matches the robot finger pose.

b) Data Collection & Model Training: During the data collection for each finger, a human operator randomly adjusts the joint angles of each finger. The robotic hand, controlled by the signal mixer box, mirrors the motion of the human fingers. Human hand joint angles θ_H of thumb, four fingers and spread (abduction/adduction) motion have degrees of freedom of 3, 4, and 4. They are mapped into θ_R of the robot hand with 2, 3, and 1 degree of freedom correspondingly. 140, 70, and 30 data pairs are collected for the thumb, four fingers, and the spread motion respectively. Adam is used as the optimizer for training of all networks, the learning rate is 0.008 and the training is stopped when MSE loss changes are smaller than 1e-6 for 20 consecutive epochs. For reproducibility, the random seed of the optimizer is set as 42.

2) Developed Pipelines for New Robot Hands: Fig. 8 compares the two development pipelines, namely the proposed Gesture-based method and the Direct mapping baseline, when adapting the existing teleportation framework to new robot hands. As shown in Fig. 8a, for adaptation to a new robot hand, Gesture-based method can reuse the well-trained classifier and regressor. Only the pre-recorded trajectories need to be updated, which involves recording 6 data pairs for each hand gesture, totaling 18 data pairs. This process takes approximately 15 minutes. Notably, this approach can accommodate various types of robotic grippers/hands with different structures and morphologies, such as those with three or four fingers.

However, for Direct mapping method (see Fig. 8b), transitioning to a different robot hand requires repeating the entire data collection (250 data pairs) and model training process This involves collecting 250 data pairs and training the model again to minimise mismatches with the robot and human hand, which takes around 45 minutes, three times longer than the Gesture-based method.

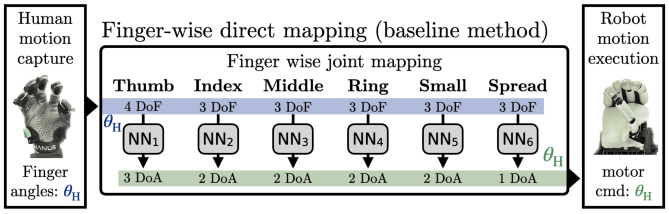

Fig. 7. Direct Mapping method - Baseline teleoperation method where each finger is mapped via neural networks to the degrees of freedom of the robots finger.

B. Experimental Setup and Protocols

The robotic setup is shown in Fig. 2(d).

a) ADAPT Hand: The ADAPT Hand [28] is used throughout the experiment The hand has 12 degrees of freedom and can be actuated to complete the Kapandji test [34] and achieve all 33 grasping taxonomies of human grasp types defined in [29], which demonstrates the hand's dexterous capabilities. The robot hand motion is manually programmed by varying each motor position independently via a "Signal Mixer" (see Fig. 8c) comprised of 12 linear potentiometers.

The compliance is the key feature of this hand. When combined with an impedance controlled Franka Research 3 robot arm, the robot hand is compliant in the skin (material compliance), finger (series elastic actuation), and wrist (impedance control). This compliance can help improve the robustness teleoperation using the proposed discrete Gestures-based method. Similar to how humans are able to the same grasp motion (e.g.: power grasp) for a variety of object geometries, the compliance allows for the same gesture to adapt to multiple object shapes and enhance the robustness of teleoperation.

b) Wrist teleoperation: To complete the teleoperation setup, a system to control the wrist position and orientation is introduced, using the integrated IMU in the motion capture glove, combined with an RBG-D camera tracking the operator's wrist position. Note that impedance control is applied in wrist teleoperation to mitigate collision risks, although it can introduce a control delay during operation. This wrist teleoperation system is used in all grasping tests for both teleoperation mapping methods. In practice, the operator has control of the relative motion of the robot when holding a "dead man's switch", rather than the absolute pose. Only when pressing the switch the robot is active, while the relative pose is zeroed on release.

1) Experimental Subjects: In the evaluation experiments, the teleoperation capabilities of the proposed method will be first assessed with regard to motion mapping, followed by performing grasping tests.

Since the difference in hand size between human and robot hands can adversely affect the performance of teleoperation mapping [35], we emphasize the robustness of our proposed

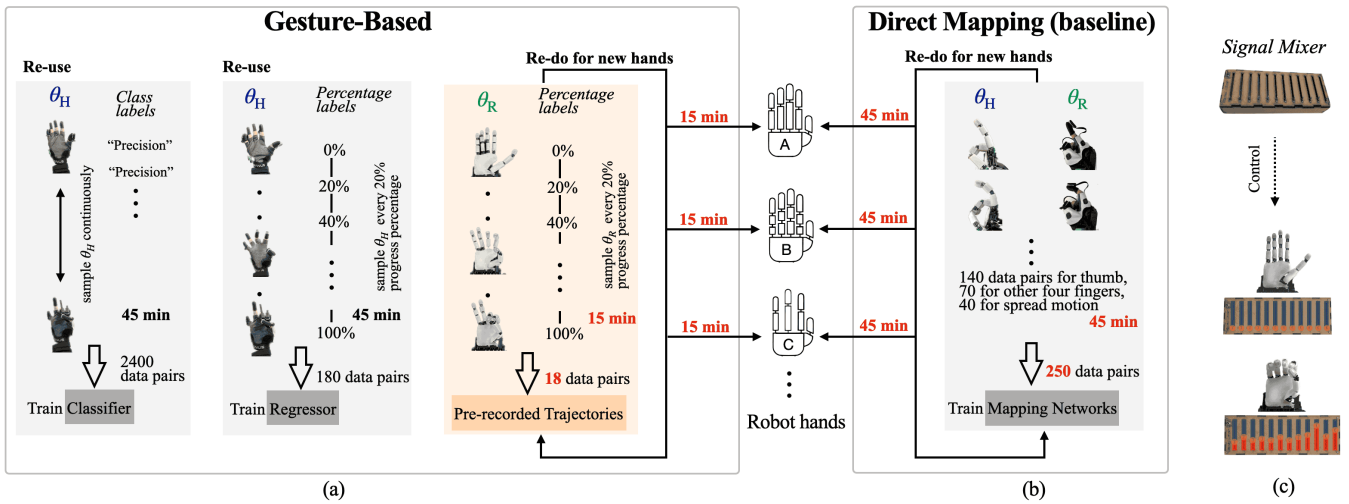

Fig. 8. : Development pipelines of teleoperation for (a) Gesture-based method and (b) Direct mapping method. (c) Control of robot hand through Signal Mixer during data collection

mappings. For the motion mapping assessment, both classification accuracy and the mapping accuracy (defined as the spatial distance difference across fingertips between the human and robot hand) of the teleoperation are assessed with five operators of varying hand sizes. In the test about mapping accuracy of the teleoperation, five operators (see table II) perform precision and semi-power gestures for both the mapping methods, and an image is taken with the same view point. Power grasping is not evaluated because it is not designed for precision grasping tasks involving small objects. From this the pixel distance between the thumb and the index fingertip alongside qualitative comparisons of robot gestures photographs are used to evaluate the mapping accuracy. In summary, two groups of operators are invited to classification accuracy and the mapping accuracy tests. The dimensions of their hands are shown in Table I, II. Fig. 9 illustrates the comparison of hand sizes between the expert (from whom training data was collected), and testers involved in the classification experiment.

In the grasping tasks, the usability of teleoperation methods are assessed by expert operators and beginner / novice operators. The *Expert* is an individual with strong technical knowledge of the underlying system with 10+ hours of experience in data collection, teleoperation methods development, and trial tests on the setup. Additionally, ten *Beginner* operators (first time users of the teleoperation system) were selected for the evaluation. The age of the expert and all beginner operators range from 25 - 33. Among the beginner operators, two are female, while the others and the expert are male. Similar to work in [36], [37], the subjective experience of users with the system is evaluated by inviting them to complete the System Usability Scale (SUS) questionnaire after the testing. The necessary Human Research Ethics Committee (HREC) approvals were sought for the experiments.

2) Objects Test Set: Fig. 10 shows the 20 objects used in grasping tests alongside their rough dimensions. These objects are selected to encompass a diverse range of sizes, shapes, and forms (similar to [28]), facilitating testing across

TABLE I
HAND DIMENSIONS OF EXPERT AND 5 TEST USERS FOR THE CLASSIFICATION EXPERIMENT.

	Expert (Training)	Tester 1	Tester 2	Tester 3	Tester 4	Tester 5
Length (cm)	18.3	22.6	20.4	18.3	21.3	17.7
Width (cm)	10.5	12.3	11.0	9.6	12.8	10.9

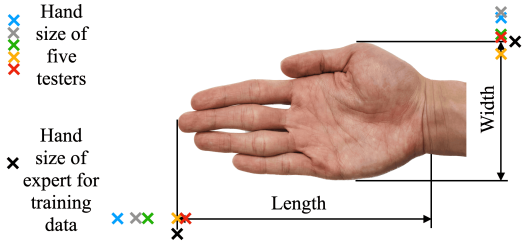

Fig. 9. Hand size of expert (for training data collection) and testers for classification experiment.

different grasp types, including precision, power, and semi-power grasps. Objects of varying geometries were selected and split into two sets: the *core* set used for a detailed teleoperation assessment, and an *extended* set used to measured the broader generalizability of the teleoperation method to different objects.

C. Additional Results

1) Object generalization: To measure the generalizability of the two methods to a wide variety of objects, the expert operator teleoperates to grasp all the 20 objects. For each object, two teleoperation methods are tested to measure the execution time. For the proposed Gesture-based mapping method, each grasping gesture is tested separately for all objects.

The grasping results or grasping time are depicted in Fig. 11. The depth of the color bar indicates the operation time. Objects can be categorized into challenging groups (labeled based on their size characteristics) and a ease-of-grasp group.

TABLE II
HAND DIMENSIONS OF EXPERT AND 5 TEST USERS FOR THE MAPPING ACCURACY EXPERIMENT

	Expert (Training)	Tester 1	Tester 2	Tester 3	Tester 4	Tester 5
Length (cm)	18.3	22.3	20.8	19.4	22.7	17.2
Width (cm)	10.5	11.2	11.1	11.0	12.1	10.4

In the challenging group, certain objects with large surfaces (horizontal cross-sectional area, i.e., Scale, Case) can only be grasped effectively using power grasping, as the other two hand gestures do not spread the fingers wide enough to cover the object surface. For the object such as a standing Board, which is thin and has vertical side surface requiring sufficient force and contact area to provide friction for lifting, only semi-precision works due to the pinch power between the finger pulp of four fingers and the thumb. For objects with small surfaces and thin height (i.e., Battery, Cable), both power grasping and precision grasping are viable, but precision grasping typically takes less time to complete. This suggests that our three grasping gestures have their respective strengths in dealing with different types of objects.

In comparison, although the Direct mapping methods work for most of objects in the challenging group, the teleoperation time is significantly longer. For objects in ease-of-grasp group, the results of all methods are similar.

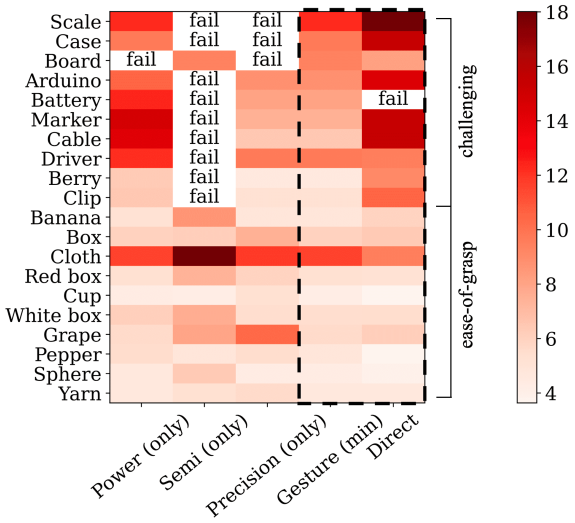

Fig. 11. Execution time of expert operator in one-shot grasping tasks, using single gesture (power, semi-power, precision) of Gesture-based method and Direct mapping methods. "Gesture (min)" refers to the shortest execution time among three grasping gestures, the corresponding hand gesture is the most suitable one for grasping the object. Grasping test that requires more than 20 seconds is considered as a failure.

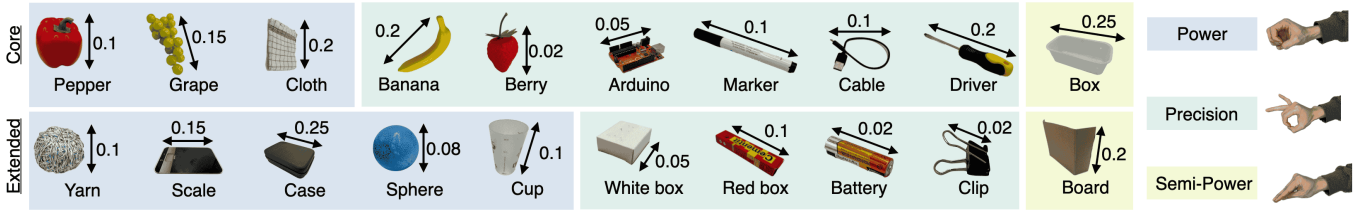

Fig. 10. Images and dimensions (in meters) of the different objects used in the grasping experiments.