A Hand Telerobotic System with Enhanced Dexterity towards Dexterous Manipulation Data Acquisition

Ruize Wang, Yangbin Bao, Qi Ye, Peng Cheng, Jiming Chen and Gaofeng Li*

Abstract— **Although the multi-fingered hand brings more possibilities for human-like dexterous manipulation, it also introduces serious challenges in interacting with the dynamic and uncertain physical environments. Compared with two-fingered parallel grippers, the multi-fingered hands have more joints and diverse interaction postures. This makes the operation no longer limited in the one-dimensional opening/closing degree, but can realize flexion/extension, adduction/abduction, and rotating in three-dimensional space, which are more dexterous and closer to human manipulation. However, how to precisely capture these dexterous motions and map them to a robotic hand is extremely challenging for existing technologies. In this abstract, we propose a hand telerobotic system with a highly under-actuated hand exoskeleton and an Allegro Hand towards dexterous manipulation data acquisition. Meanwhile, we design a joint mapping method with joint-space reconstruction and a Cartesian mapping method with an auxiliary frame to ensure the human hand motion can be correctly and precisely mapped to the robotic hand. To enhance system adaptability in complex tasks, we further design a hybrid mapping method combining the Joint-Cartesian space to improve the adaptability of our system. Compared with existing works, our system is able to capture the full states of humans hands and replicate them on the robotic hand, especially for the adduction/abduction motions. Three dexterous manipulation tasks with grasping, rotating, screwing and tapping actions are designed to validate the dexterity of our system.**

I. Introduction

Making robots have the humanoid dexterous manipulation ability is the most essential and central goal in robotics development [1]. Multi-fingered robotic hands are the key to achieving this goal and unlocking the robot's potential for dexterous manipulations. As the multi-fingered hands have more joints and more contact points in interactions, they can complete complex manipulation (e.g., using tools and rotation operation in hand) and provide stable and reliable grasping capabilities. While these features also bring many challenges to the planning and control of multi-fingered hands. Higher degrees of freedom (DoF) greatly make traditional model-based control impossible. Thus learning-based method based on large amount of dexterous manipulation data becomes a promising solution. However, how to acquire the dexterous manipulation data with multi-fingered hands is non-trivial.

This research is funded by the National Natural Science Foundation of China (NSFC) under Grant 62203426 and supported by NSFC under Grant 2088101.

R. Wang, Y. Bao, Q. Ye, P. Cheng, J. Chen and G. Li are with the College of Control Science and Engineering, Zhejiang University, Hangzhou, 310027, China, and with the Key Laboratory of Collaborative Sensing and Autonomous Unmanned Systems of Zhejiang Province. (e-mails: {ruize.wang, yangbin.bao, qi.ye, lunarheart, cjm, gaofeng.li}@zju.edu.cn). G. Li is the corresponding author.

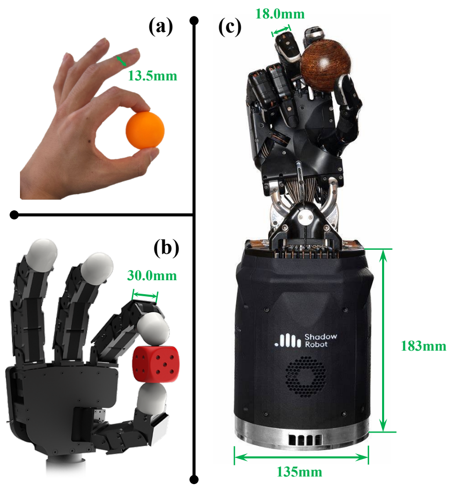

Fig. 1. Illustration of the human-to-robot gap. The comparison of (a) human hand, (b) Allegro Hand and (c) Shadow Hand. It is obvious that the Allegro Hand's fingers are much larger than human and the Shadow Hand has a huge driven box, which bring negative effects to the Arm-Hand system.

At present, there are three paradigms for acquiring dexterous manipulation data, namely: data generation based on simulation platform, data collection from human demonstration, and data collection from teleoperation demonstration. The most common method is to generate manipulation data through simulation platform [2] [3] and learn through deep reinforcement learning [4] or reinforcement learning with domain knowledge [5]. However, this method cannot get rid of the data generated by the simulation environment, which means that the sim-to-real gap makes it difficult to directly transfer the trained model to the real robotic hand.

The second paradigm is to acquire data from human demonstration, in which the tasks are fulfilled by the human beings directly. Compared with the data generated by simulation, the demonstrated data in this paradigm is real data generated in the physical environment based on video [6] [7] or visual and tactile fusion [8] [9], which can greatly reduce the sim-to-real gap. But this will introduce a new human-to-robot gap. The current multi-fingered robotic hands are very different from human hands in configuration, which makes it difficult to replicate the manipulation skills learned from human demonstration on multi-fingered robotic hands directly. The comparison of multi-fingered robotic hands and human hand is shown in Fig. 1.

In order to solve those gaps in dexterous manipulation, using the telerobotic system to acquire data becomes a more efficient method. To achieve an excellent teleoperation effect

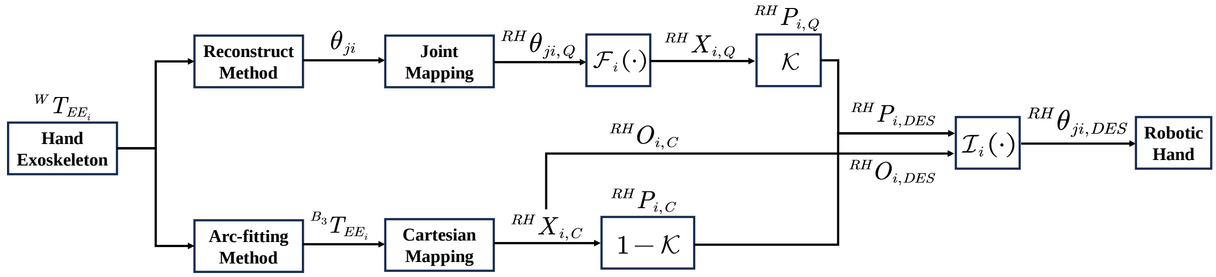

Fig. 2. Block diagram of our combined Joint-Cartesian mapping method. The two sub-problems, measurement of the motions on the human side and the combined mapping method to the robotic hand, are both addressed in our framework.

and get a high-quality dataset, two key subproblems need to be addressed: the measurement of the motions on the human side and the mapping method of movements on the robotic hand. For the first subproblem, current visual tracking [10] [11] and kinesthetic teaching [12] methods are unstable and susceptible to environmental interference. Motion capture based on tactile gloves [13] or exoskeletons [14] is more accurate and intuitive. But they could only capture flexion/extension motion, which is not enough to realize dexterous manipulation. For the second subproblem, there are joint mapping methods for power grasping [15]–[17] and Cartesian mapping methods for precise manipulation [18]–[20]. But those mapping method only focus on mapping joints or fingertips motion, which lacks adaptivity and cannot perform complex manipulation task even if full states of hands motion can be captured.

To solve the above problems, we design a telerobotic system and test it in three different dexterous manipulation tasks:

- **Full State Motion Capture of Human Hands:** A highly underactuated hand exoskeleton is used to capture human hands motion. And we design a joint mapping method with joint-space reconstruction, which can obtain the precise finger joints to the robotic hand, especially for adduction/abduction motion.
- **Precise Mapping Method and Adaptive Mapping Framework:** We design a Cartesian mapping method with an auxiliary frame to ensure the human hand motion can be precisely mapped to robotic hand. And we design a hybrid mapping framework as shown in Fig. 2 to combine the Joint-Cartesian space to map complex motions adaptively.
- **System Dexterity Verification:** We design three dexterous manipulation tasks with grasping, screwing, and tapping motions to validate the dexterity of our telerobotic system. The successful completion of the tasks demonstrates the adaptability, dexterity and accuracy of our system.

What's more, the control frequency of the telerobotic system is at the kilohertz level, which brings agile teleoperation. Thus, we can map complex human fingers motion (e.g., adduction/abduction) to the multi-fingered hand and acquire manipulation data. The acquired data can use directly in learning-base algorithm without the gap in sim-to-real and human-to-robot.

II. Proposed Method

A. Exoskeleton Design

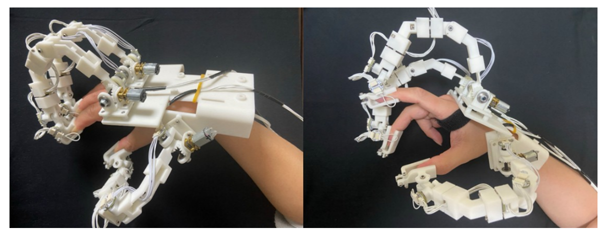

Fig. 3. The highly under-actuated hand exoskeleton used for tracking the position and orientation of human hand.

Based on the HEXOTRAC [21], we design a hand exoskeleton with the abilities for full-space motion capture of human fingertips and reconstruction of finger configuration. The general assembly of the hand exoskeleton is shown in Fig. 3. The back of the human hand is tied to the hand exoskeleton base, while the thumb, index finger, and middle fingertips are respectively fixed to the second cross link at the end of the three serial kinematic chains.

Each finger has a total of 6 DoF, including 3 DoF in the pitch direction, 2 degrees of freedom in the roll direction, and 1 DoF in the yaw direction, which allows for effective haptic interactions within the functional workspace of the hand without constraints to the fingers. This high-freedom design fully satisfies the finger movement of wearers of different sizes in the workspace.

B. Human Finger's Joint-space Reconstruction

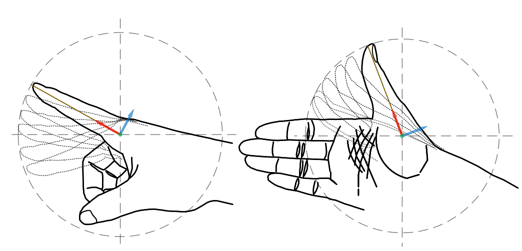

Fig. 4. The index and the thumb fingers moved as a circular arc around the base frame.

Although HEXOTRAC has provided the capability of accurate finger motion capture, two questions still remain:

First, the transformation $^{W}T_{B_i}$ from world frame to each finger's base frame is uncertain for different wearers.

Besides, the lengths of each phalanxes are uncertain for different people.

Therefore, it is important to apply our finger reconstruction method [22]. As shown in Fig. 4, a finger can be modeled as a serial link manipulator by simple assumptions: each fingers' last three parallel joints are placed on a certain plane, the ratios of each phalanx are considered to be fixed for all people in medical science [23], and the fingertip is placed on the sphere.

So we can summarize the method for obtaining the precise finger joints in three steps:

First, let $P_i = \{x_B^i, y_B^i, z_B^i\}$ represent the position of $\{B_i\}$ relative to $\{W_i\}$, let $P_i^{ee} = \{x, y, z\}$ represent the position of the fingertip. we can calibrate the translation from $\{W\}$ to $\{B_i\}$ using an arc fitting algorithm:

$$(x - x_B^i)^2 + (y - y_B^i)^2 + (z - z_B^i)^2 = L_{Total}^2. \quad (1)$$

With the arc fitting algorithm, we can estimate the position of $\{B_i\}$ in the $\{W\}$:

$$L_{Total} = \sqrt{(x_B^i)^2 + (y_B^i)^2 + (z_B^i)^2 - \hat{d}}, \quad (2)$$

Although length of fingers varies from person to person, the ratios of each phalanxes are considered to be fixed for all human being. With knowing the total lengths L_{Total}, we each easily obtain the lengths of each phalanx.

Finally, the joint angles can be calculated with inverse kinematics ${}^{B_i}T_{EE_i}$ which is the transformation from base to fingertip needs to be captured to get the accurate joint configurations. According to the kinematics tree, ${}^{B_i}T_{EE_i}$ can be calculated as follow:

$${}^{B_i}T_{EE_i} = ({}^{W}T_{B_i})^{-1} \cdot {}^{W}T_{EE_i}, \quad (3)$$

${}^{W}T_{EE_i}$ is the transformation from world to fingertip which can be captured by hand exoskeleton. And the accurate finger joints can be calculated with the precise lengths of phalanx estimated above.

C. Cartesian Mapping Method

Direct Cartesian Mapping typically involves processing the human fingertip poses ${}^{B_3}T_{EE_i}$ to apply scaling, optimization, or ad hoc transformations based on specific design criteria. A new Cartesian method based on an auxiliary frame is proposed. The method can be summarized into two main steps: Unify the rotation of world frame and calculate the transformation T' to bridge the gap from human fingers to robotic fingers,

$${}^{RW}T_{REE_i}(\Theta) = T_i^{'} \cdot {}^{B_3} T_{EE_i}(\Theta) \cdot ({}^{REE_i}T_{EE_i})^{-1}, \quad (4)$$

$$T_1^{'} = ({}^{W'}T_{RW})^{-1} \cdot {}^{W'} T_{B_3}, \quad (5)$$

where ${}^{W'}T_{RW}$ and ${}^{W'}T_{B_3}$ can be captured by the hand exoskeleton, with the compensation values obtained for different fingers $T_i' \in \mathbb{R}^{4\times4}$, then for any finger at configuration Θ, ${}^{W'}T_{REE_i}(\Theta)$ can be calculated as follow:

$${}^{W'}T_{REE_i}(\Theta) = T_i^{'} \cdot {}^{W'} T_{EE_i}(\Theta) \cdot {}^{REE_i} T_{EE_i}, \quad (6)$$

where ${}^{REE_i}T_{EE_i}$ represents the gap between robotic fingers and human fingers.

D. Hybrid Joint-Cartesian Mapping Method

In the real dexterous manipulation task, it is important to have the ability to switch powerful gasping mode and precise operating mode adaptively.

Here we design a direct joint mapping method [22], which can fully map the range of human finger joint configurations to a robotic hand linearly. The advantage of this algorithm is its ability to achieve complete control of the robotic hand across the whole joint space, which is considered effective in power grasping.

Based on the proposed Joint and Cartesian Mapping methods, we use a hybrid mapping method to combine the Joint-Cartesian space. ${}^{RH}P_{i,DES} \in \mathbb{R}^{1\times3}$ and ${}^{RH}R_{i,DES} \in \mathbb{R}^{3\times3}$ are the position part and rotation part of the desire poses of the robotic hand, respectively. Then, we have:

$${}^{RH}P_{i,DES} = {}^{RH}P_{i,Q} \cdot \mathcal{K} + {}^{RH}P_{i,C} \cdot (\mathcal{I} - \mathcal{K}), \quad (7)$$

$${}^{RH}R_{i,DES} = {}^{RH}R_{i,Q}, \quad (8)$$

where $\mathcal{I} \in \mathbb{R}^{3\times3}$ is the identity matrix and $\mathcal{K} \in \mathbb{R}^{3\times3}$ is a smooth, sigmoidal gain governing the transition between joint and Cartesian mappings [24].

$$k_I = \begin{cases} 1 & \text{if} \quad \delta_\mathrm{I} < \mathrm{r_{in}} \\ \frac{1}{2}\left(1 - cos\left(\frac{\delta_I \pi}{r_{out} - r_{in}}\right)\right) & \text{if} \quad \mathrm{r_{in}} \le \delta_\mathrm{I} \le \mathrm{r_{out}}, \\ 0 & \text{if} \quad \delta_\mathrm{I} > \mathrm{r_{out}} \end{cases} \quad (9)$$

where r_{in} and r_{out} ($r_{in} < r_{out}$) represent the radii of two spheres centered in the human hand thumb fingertip, $\delta_I = \|{}^{B_3}P_{EE_1} - {}^{B_3}P_{EE_2}\|$

represents the distance between human thumb fingertip and human index fingertip.

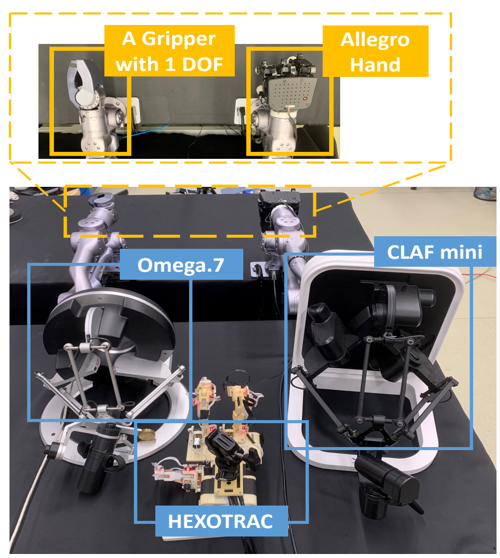

Fig. 5. Experimental platform setups. A dual-arm teleoperation system with an anthropomorphic robotic hand is established to perform the fine manipulations.

III. EXPERIMENTS

A. Experimental Setups

We design a telerobotic system for experimental validation. As illustrated in Fig. 5, Omega.7 is used to control the left manipulator equipped with a gripper, while the right manipulator is operated via CLAF mini. Additionally, the Allegro Hand is teleoperated using our exoskeleton, which offers six DoFs motion capture for each finger. To demonstrate the effectiveness in dexterous manipulation, we design three fine teleoperation tasks.

1) The first task, shown in Fig. 6, involves **screwing a screwdriver into a threaded hole**. This task requires the accurately dexterous manipulation with continuous adduction/abduction motion, ensuring the screwdriver remains stable throughout the process.

2) The second task, shown in Fig. 7, involves **unscrewing a bottle cap** in three steps: grabbing the bottle, exchanging the bottle, and unscrewing the bottle cap. The grab action requires motion capture and mapping in flexion/extension directions, while unscrewing action requires motion capture and mapping in adduction/abduction directions. However, this task requires not only the full state motion capture of human hands, but also adaptive adjustment of mapping methods for different steps. The first step involves power grasping using joint mapping to securely grip the bottle, while the third step requires precious manipulation through the proposed Cartesian mapping to complete the task. The hybrid mapping method is employed to ensure a smooth and adaptive transition between two mapping methods.

3) The final task involves **holding a panel and tapping specific points** with the robotic thumb. We selected three points along a straight line and an additional point positioned randomly away from the line to demonstrate the ability of human hands motion capture in large workspace and the precision of the finger's adduction/abduction motion mapping. The process is illustrated in Fig. 8.

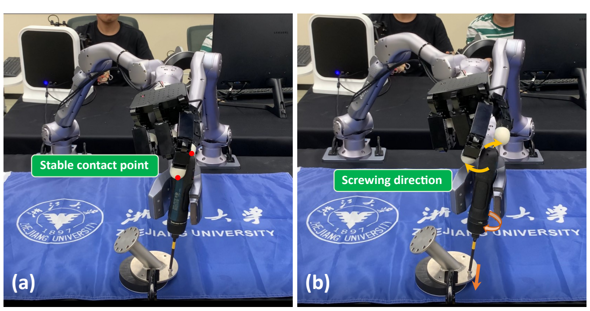

Fig. 6. The task of screwing a screwdriver. This task requires the Allegro Hand (a) contact with screwdriver stably and (b) screw the screwdriver accurately.

B. Performance in Three Dexterous Manipulation Tasks

We intend to validate the accuracy and dexterity of our telerobotic system through a user study in the bottle cap unscrewing task. Seven volunteers (aged 20–28, seven men, without the foundation of robot knowledge) are invited to finish the task. As shown in Table.I, the hybrid mapping outperforms the joint mapping both in success rate and average (AVG) time, as the joint mapping is more conducive to grasping and difficult to accommodate the dexterous manipulation. Although the success rate of Cartesian mapping is equal to hybrid mapping, its average time is longer as the volunteers need more time to get used to grab the bottle. Meanwhile, two basically trained operators using our system to perform the remaining tasks as they are too difficult for volunteers. Five consecutive test trials are conducted by each operator and the average time to finish the tasks are $242s$ and $12.5s$ in screwing screwdriver task and tapping the panel task, respectively. The results validate the adaptability, dexterity and stability of our telerobotic system.

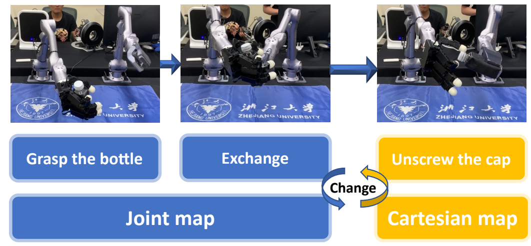

Fig. 7. The bottle cap unscrewing task and experimental steps. The yellow blocks show the third step of task with Cartesian map, while the blue blocks show other steps of task with Joint map.

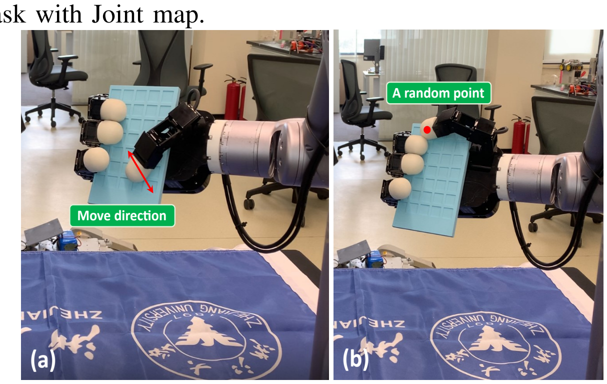

Fig. 8. The task of holding a panel and tapping. This task requires the Allegro Hand (a) tap three point in a line continuously and (b) tap a random point far from the line.

TABLE I: The result of volunteers test.

	Success rate[1]	AVG time(s)
Joint mapping	2/7	125.4
Proposed Cartesian mapping	**6/7**	118
Proposed hybrid mapping	**6/7**	**106**

[1] The trial is denoted as successful if the cap is unscrewed.

IV. Conclusion

In this paper, we designed a telerobotic system that can accurately capture the full states of human hands and adaptively map the motion to the robotic hand. Through our system, human operation data can be better collected and mapped to robotic hands, bridging the gap of sim-to-real and human-to-robot. The experimental results demonstrate that our system exhibits excellent performance in dexterous manipulation. In the future, we will focus more on force and tactile feedback for the operator which leads to a more immersive feeling and abundant dataset for dexterous manipulation.

References

[1] Aude Billard and Danica Kragic. Trends and challenges in robot manipulation. *Science*, 364(6446):eaat8414, 2019.

[2] Haoming Li, Qi Ye, Yuchi Huo, Qingtao Liu, Shijian Jiang, Tao Zhou, Xiang Li, Yang Zhou, and Jiming Chen. Tpgp: Temporal-parametric optimization with deep grasp prior for dexterous motion planning. In *2024 IEEE International Conference on Robotics and Automation (ICRA)*, pages 18106–18112, 2024.

[3] Yufei W · Zhou X · Feng C · Johnson Tsun-Hsuan W · Yian W · Katerina Fragkiadaki · Zackory Erickson · David Held · Chuang G. Robogen: Towards unleashing infinite data for automated robot learning via generative simulation. In *International Conference on Machine Learning*, pages 51936–51983, 2024.

[4] Lingfeng Tao, Jiucai Zhang, and Xiaoli Zhang. Stable in-hand manipulation with finger-specific multi-agent shadow critic consensus and information sharing. *IEEE Robotics and Automation Letters*, 10(3):2407–2413, 2025.

[5] Gagan Khandate, Cameron Paul Mehlman, Xingsheng Wei, and Matei Ciocarlie. Dexterous in-hand manipulation by guiding exploration with simple sub-skill controllers. In *2024 IEEE International Conference on Robotics and Automation (ICRA)*, pages 6551–6557, 2024.

[6] Kenneth Shaw, Shikhar Bahl, and Deepak Pathak. Videodex: Learning dexterity from internet videos. In *Proceedings of The 6th Conference on Robot Learning*, pages 654–665, 2023.

[7] Shikhar Bahl, Russell Mendonca, Lili Chen, Unnat Jain, and Deepak Pathak. Affordances from human videos as a versatile representation for robotics. In *2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 01–13, 2023.

[8] Qingtao Liu, Yu Cui, Zhengnan Sun, Gaofeng Li, Jiming Chen, and Qi Ye. VTDexmanip: A dataset and benchmark for visual-tactile pretraining and dexterous manipulation with reinforcement learning. In *The Thirteenth International Conference on Learning Representations*, 2025.

[9] Qingtao Liu, Qi Ye, Zhengnan Sun, Yu Cui, Gaofeng Li, and Jiming Chen. Masked visual-tactile pre-training for robot manipulation. In *2024 IEEE International Conference on Robotics and Automation (ICRA)*, pages 13859–13875, 2024.

[10] Ankur Handa, Karl Van Wyk, Wei Yang, Jacky Liang, Yu-Wei Chao, Qian Wan, Stan Birchfield, Nathan Ratliff, and Dieter Fox. Dexpilot: Vision-based teleoperation of dexterous robotic hand-arm system. In *2020 IEEE International Conference on Robotics and Automation (ICRA)*, pages 9164–9170, 2020.

[11] Sridhar Pandian Arunachalam, Sneha Silwal, Ben Evans, and Lerrel Pinto. Dexterous imitation made easy: A learning-based framework for efficient dexterous manipulation. In *2023 IEEE International Conference on Robotics and Automation (ICRA)*, pages 5954–5961, 2023.

[12] Katharina Mülling, Jens Kober, Oliver Kroemer, and Jan Peters. Learning to select and generalize striking movements in robot table tennis. *The International Journal of Robotics Research*, 32(3):263–279, 2013.

[13] Heni Ben Amor, Oliver Kroemer, Ulrich Hillenbrand, Gerhard Neumann, and Jan Peters. Generalization of human grasping for multi-fingered robot hands. In *2012 IEEE/RSJ International Conference on Intelligent Robots and Systems*, pages 2043–2050, 2012.

[14] Shu-Wei Pu, Yu-Cheng Pei, and Jen-Yuan Chang. Decoupling finger joint motion in an exoskeletal hand: A design for robot-assisted rehabilitation. *IEEE Transactions on Industrial Electronics*, 67(1):686–697, 2020.

[15] I. Cerulo, F. Ficuciello, V. Lippiello, and B. Siciliano. Teleoperation of the SCHUNK S5FH under-actuated anthropomorphic hand using human hand motion tracking. *Robotics and Autonomous Systems*, 89:75–84, 2017.

[16] H. Liu, X. Xie, M. Millar, M. Edmonds, F. Gao, Y. Zhu, V. J. Santos, B. Rothrock, and S. C. Zhu. A glove based system for studying hand-object manipulation via joint pose and force sesing. In *2017 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), Canada*, pages 1033–1040, 2017.

[17] H. Liu, Z. Zhang, X. Xie, Y. Zhu, Y. Liu, Y. Wang, and S. C. Zhu. High-fidelity grasping in virtual reality using a glove-based system. In *2019 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), Canada*, pages 5180–5186, 2019.

[18] L. Cui, U. Cupcic, and J.S. Dai. An optimization approach to teleoperation of the thumb of a humanoid robot hand: Kinematic mapping and calibration. *Journal of Mechanical Design*, 136(9), 2014.

[19] U. Scarcia, R. Meattini, and C. Melchiorri. Mapping human hand fingertips motion to an anthropomorphic robotic hand. In *2017 IEEE International Conference on Robotics and Biomimetics (ROBIO)*, pages 774–779, 2017.

[20] A. Handa, K. V. Wyk, W. Yang, J. Liang, Y. W. Chao, Q. Wan, S. Birchfield, N. Ratliff, and D. Fox. Dexpilot: Vision-based teleoperation of dexterous robotic hand-arm system. In *2020 IEEE International Conference on Robotics and Automation (ICRA)*, pages 9164–9170, 2020.

[21] I. Sarakoglou, A. Brygo, D. Mazzanti, N. G. Hernandez, D. G. Caldwell, and N. G. Tsagarakis. , HEXOTRAC: A highly under-actuated hand exoskeleton for finger tracking and force feedback. In *2016 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS),Daejeon*, pages 1033–1040, 2016.

[22] Y. Su, G. Li, Y. Deng, I. Sarakoglou, N. G. Tsagarakis, and J. Chen. The joint-space reconstruction of human fingers by using a highly under-actuated exoskeleton. In *2024 IEEE International Conference on Robotics and Automation (ICRA)*, 2024.

[23] J. Lenarčič and T. Bajd and M.M. Stanišić. *RobotMechanisms*. Springer, 2013.

[24] R. Meattini, D. Chiaravalli, G. Palli, and C. Melchiorri. Simulative evaluation of a joint-cartesian hybrid motion mapping for robot hands based on spatial in-hand information. *Fronties in Robotics and AI*, 2022.