Bimanual Dexterity for Complex Tasks

Kenneth Shaw[*] **Yulong Li**[*]
Jiahui Yang **Mohan Kumar Srirama** **Ray Liu** **Haoyu Xiong**
Russell Mendonca[†] **Deepak Pathak**[†]
Carnegie Mellon University

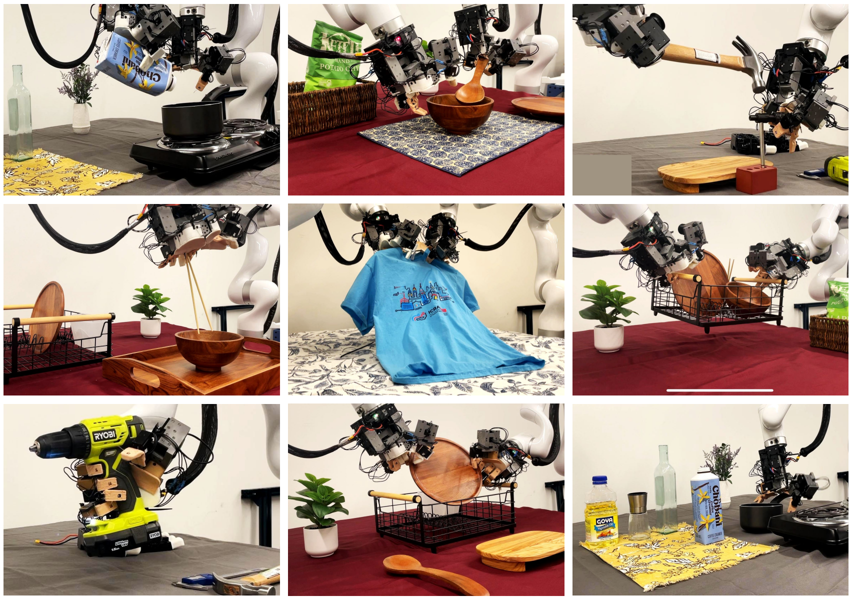

Figure 1: **Bimanual Dexterity**: Our teleoperation data collection system can perform various complex tasks of interest including pouring, scooping, hammering, chopstick picking, hanger picking, picking up basket, drilling, plate pickup and pot picking, using a bimanual dexterous system.

Abstract: To train generalist robot policies, machine learning methods often require a substantial amount of expert human teleoperation data. An ideal robot for humans collecting data is one that closely mimics them: bimanual arms and dexterous hands. However, creating such a bimanual teleoperation system with over 50 DoF is a significant challenge. To address this, we introduce BiDex, an extremely dexterous, low-cost, low-latency and portable bimanual dexterous teleoperation system which relies on motion capture gloves and teacher arms. We compare BiDex to a Vision Pro teleoperation system and a SteamVR system and find BiDex to produce better quality data for more complex tasks at a faster rate. Additionally, we show BiDex operating a mobile bimanual robot for in the wild tasks. Please refer to https://bidex-teleop.github.io for video results and instructions to recreate BiDex. The robot hands ($5k) and teleoperation system ($7k) is readily reproducible and can be used on many robot arms including two xArms ($16k).

Keywords: Bimanual, dexterous hands, behavior cloning

1 Bimanual Robot Hand and Arm System

We introduce BiDex, a system designed to enable an operator to naturally teleoperate any bimanual robot hand and arm setup. BiDex is exceptionally precise, low-cost, low-latency, portable and can

[*]Equal contribution. [†] Equal advising.

8th Conference on Robot Learning (CoRL 2024), Munich, Germany.

control any human-like pair of dexterous hands even with over 20 degrees of freedom. It achieves this accurate tracking of the human hand by utilizing a Manus VR glove based-system [1] and the human arm using a GELLO-inspired system [2]. We present the full process by which we send commands using our system for controlling bimanual hands with dextrous hands. Crucially, our solution is tailored to operate seamlessly in both tabletop and mobile environments because it does not need any external tracking devices and is very portable. In Section 3 we find that our system is highly intuitive, precise, and cost-effective compared to many commonly used approaches today such as the VR headset and SteamVR for both the tabletop and mobile manipulation settings on two different pairs of robot hands.

1.1 Multi-fingered Hand Tracking

A hand tracking system must provide low-latency skeleton information for the human hand, which has over 20 degrees of freedom. Many current vision-based tracking systems, like those using FrankMocap, struggle with inaccuracies due to occlusions and lighting changes. [4, 5] In contrast, recent motion capture gloves with EMF sensors offer greater accuracy without significant cost and are unaffected by occlusions, delivering detailed joint-structure data while allowing comfortable wear. We chose the Manus Glove for its reliable tracking and minimal overheating or calibration issues compared to alternatives like the Rokoko gloves. However, mapping commands from a human hand to a robot hand remains challenging due to their different structures. Previous research has addressed this by ensuring consistent pinch grasping between the two. Effective mapping has also been demonstrated by optimizing fingertip and penultimate joint positions using an SDLS IK solver. [6] We apply a similar inverse kinematics approach with the Manus gloves, achieving precise pinch grasps and proper thumb placement.

1.2 Arm Tracking

Our system must track the human wrist pose accurately to control two robot arms. Traditionally, there are many approaches from both the motion capture community and robotics community alike. However, many of these approaches rely on calibrated external tracking devices which either are costly have high latency or are not portable. Instead, we leverage key insights from Zhao et al. [7], Wu et al. [2] which both use lighter teacher arms attached to the human arm to control a robot arm and hand system. Specifically we follow the GELLO system from Wu et al. [2] to teleoperate a full size robot arms. A key question is how to mount this arm-tracking system on a human wrist and hand. If the robot hand is mounted on the arm in a human-like way, then the glove needs to be mounted in a human-like way on the GELLO to match. However, this orientation means that the human arm will be parallel with the GELLO and constantly collide with it. In BiDex, we mount the robot hands underneath the arms in the same orientation as if it were a gripper. When mirroring this in the GELLO, the human arm and GELLO output are perpendicular to each other which is more comfortable.

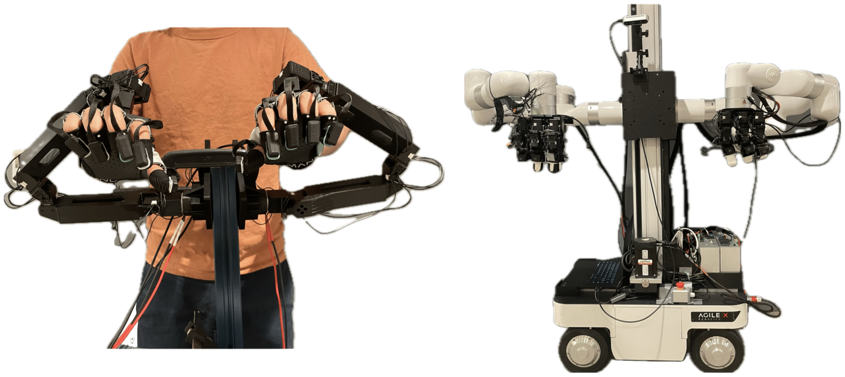

Figure 2: **Mobile bimanual teleoperation system** Left: An operator strapped into BiDex. Right: Our bimanual robot setup including two xArm robot arms, two LEAP Hands [3] and three cameras on an AgileX base.

1.3 Robot Configurations

Tabletop Manipulation For our tabletop setup, the robotic arms are positioned to face each other, while the GELLO teaching arms mirror this configuration. Compared to a side-by-side configuration, this setup has three main benefits: 1) the human operators avoid collisions with the GELLO arms; 2) the setup allows better visibility of the workspace, which will be otherwise occluded by a side-by-side robot arm configuration; 3) and finally, the robot arms have a wider shared workspace.

Mobile Manipulation BiDex does not require external tracking systems and is very lightweight so it can easily be used in the mobile setting. The teacher arms are instead mounted onto a compact mobile cart. Our cameras are all egocentric, one is mounted on the torso and two others are mounted on each of the wrists. For the mobile robot, we mount two robot arms onto an articulated torso that can move up towards high objects and down towards the ground similar to PR2 [8]. The robotic assembly, including arms and torso, is mounted on an AgileX Ranger Mini enabling movement in any SE(2) direction required by the task [9, 10]. A secondary operator manages the mobile base with a joystick, manages the task resets, and handles the data.

2 Experiment Setup

2.1 Baseline Teleoperation Approaches

Vision based VR Headset In recent years, the accessibility of low-cost VR headsets using multi-camera hand tracking has made them popular for teleoperation such as in [11, 12] As a baseline we use the Apple Vision Pro which returns both finger data similar to MANO parameters [13] and wrist coordinate frame data. The finger data is used in the same way as with BiDex through inverse kinematics-based retargeting and commanded onto the robot hands. The wrist data is reoriented, passed through inverse kinematics and the final joint configuration is commanded to the arms.

SteamVR Tracking SteamVR, commonly used in the video gaming community has also seen recent interest in the robotics community from industry [14] and academia alike. [15, 16] It uses active powered laser lighthouses that must be carefully placed around the perimeter of the workspace. Wearable pucks with IMUs and laser receivers are worn on the body of the operator. In our experiment the operator wears one tracker on each wrist and one tracker on their belly. The wrist location is calculated with respect to the belly pose, mapped to the robot arm and the joint angles are calculated using inverse kinematics. The hand tracking gloves are the same as BiDex.

2.2 Choice of Dexterous End-Effectors

Leap Hand LEAP Hand, introduced by Shaw et al. [3] is a low-cost, easy-to-assemble robot hand with 16 DOF and 4 fingers. LEAP Hand introduces a novel joint configuration that optimizes for dexterity as well as human-like grasping. We use this hand for many experiments in the paper as it is a readily available dexterous hand available for comparison studies.

DLA Hand We would like a hand that is smaller and more compliant than LEAP Hand. DLA Hand is crafted to mimic the suppleness and strength of the human hand with fingers that have a 3D-printed flexible outer skin paired with a sturdy inner framework resembling bones. These fingers do not break but instead bend and flex upon impact. We also introduce an active articulated palm which integrates two motorized joints, one spanning the fingers and another for the thumb, enabling natural tight grasping.

3 Results

3.1 Bimanual Dexterous Teleoperation Results

BiDex provides more stable arm tracking. Because the teacher arm system is wired, BiDex is very reliable and does not fail. There is almost no jitter and very little latency which makes it ideal for arm tracking. The GELLO is very light and does not get in the way of the user more than the

	Completion Rate			Time Taken		
	Handover	Cup Stacking	Bottle Pouring	Handover	Cup Stacking	Bottle Pouring
Vision Pro VR	60	40	70	21.6	38.8	35.5
SteamVR	80	85	60	17.5	16.5	15.5
BiDex	95	75	85	6.5	15.5	14.9

Table 1: **Tabletop Teleoperation**: We compare BiDex on the handover, cup stacking, and bottle pouring tasks to two baseline methods, SteamVR and Vision Pro. BiDex enables more reliable and faster data collection, especially for harder tasks like bottle pouring.

	Completion Rate			Time Taken		
	Chair Pushing	Box Carry	Clear Trash	Chair Pushing	Box Carry	Clear Trash
Vision Pro VR	75	75	50	15.0	33.7	79.8
BiDex	95	95	75	16.4	29.7	74.6

Table 2: **Mobile Teleoperation**: Completion rate and time taken averaged across 20 trials using a mobile bimanual system with LEAP Hand [3], for different tasks. BiDex is versatile and compact enough to be adopted to successfully collect data for mobile tasks.

weight of the gloves. The kinematic feedback of arm resistance is very light but it helps the operator feel and work around arm singularities naturally. As seen in Table 1 and Table 2, BiDex achieves a higher completion rate while requiring less time for teleoperators. The Vision Pro often has jittery arm tracking which makes it difficult to teleoperate more difficult tasks. Low-pass filtering can help somewhat, but the latency added is undesirable. Occasionally the system will stop working completely which is jarring for the user.

BiDex provides more accurate hand tracking. With BiDex, the fingertip tracking is very accurate with the Manus glove. When mapping to different robots, the inverse kinematics only has to be tuned slightly with each operator if their hand size is significantly different. The abduction-adduction of the MCP side joint is also very accurate in any condition. These advantages are especially apparently in DLA Hand where these accuracies are very important in doing more complex tasks.

With the Vision Pro, the hand size often changes slightly with different lighting conditions which makes it difficult to retarget to the robot hands. The abduction-adduction estimation for the fingers also changes with occlusions which makes it difficult to do more complex tasks. The latency is noticeable but this is not an issue when teleoperating for quasi-static tasks.

3.2 Training Dexterous Visuomotor Policies with BiDex

To ensure that the data that is collected by our system is high quality and useful for machine learning we train single task closed loop behavior cloning policies. Specifically, we train an action chunking transformer from [7] with a horizon length of 16 at 30hz using pretrained weights from [17] on around 50 demonstrations. The state space is the current joint angles of the robot hand and the images from the camera. The action space in the case of LEAP Hand is 16 dimensions for each hand and 6 dimensions for each arm for a total of 44 dimensions. During rollouts, the behavior of the policies are very smooth, exhibiting the high quality of the teleop data. In tasks such as the YCB Pringles can [18] handover, we even see good generalization of the policy to different initial locations of the can.

	Can Handover	Cup Stacking	Bottle Pouring
Leap Hand	7/10	14/20	16/20

Table 3: **Imitation learning**: We train ACT from [7] using data collected by BiDex and find that our system can perform well even in this 44 dimension action space. This demonstrates that our robot data is high quality for training robot policies.

Acknowledgments

We thank Ankur Handa, Arthur Allshire, Toru Lin and Nancy Pollard for discussions about the paper. We also thank Maarten Witteveen and Sarah Shaban at Manus Meta for their assistance on their gloves. We thank Diya Dinesh and Lukas Kebuladze for helping with teleoperation. This work is supported in part by DARPA Machine Commonsense grant, AFOSR FA9550-23-1-0747, ONR N00014-22-1-2096, AIST CMU grant and Google Research Award.

References

[1] Manus. https://www.manus-meta.com/. [VR haptic gloves].

[2] P. Wu, Y. Shentu, Z. Yi, X. Lin, and P. Abbeel. Gello: A general, low-cost, and intuitive teleoperation framework for robot manipulators. *arXiv preprint arXiv:2309.13037*, 2023.

[3] K. Shaw, A. Agarwal, and D. Pathak. Leap hand: Low-cost, efficient, and anthropomorphic hand for robot learning. *arXiv preprint arXiv:2309.06440*, 2023.

[4] A. Handa, K. Van 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. IEEE, 2020.

[5] A. Sivakumar, K. Shaw, and D. Pathak. Robotic telekinesis: Learning a robotic hand imitator by watching humans on youtube. *arXiv preprint arXiv:2202.10448*, 2022.

[6] S. R. Buss and J.-S. Kim. Selectively damped least squares for inverse kinematics. *Journal of Graphics tools*, 10(3):37–49, 2005.

[7] T. Z. Zhao, V. Kumar, S. Levine, and C. Finn. Learning fine-grained bimanual manipulation with low-cost hardware. *arXiv preprint arXiv:2304.13705*, 2023.

[8] J. Bohren, R. B. Rusu, E. G. Jones, E. Marder-Eppstein, C. Pantofaru, M. Wise, L. Mösenlechner, W. Meeussen, and S. Holzer. Towards autonomous robotic butlers: Lessons learned with the pr2. In *2011 IEEE International Conference on Robotics and Automation*, pages 5568–5575. IEEE, 2011.

[9] H. Xiong, R. Mendonca, K. Shaw, and D. Pathak. Adaptive mobile manipulation for articulated objects in the open world. *arXiv preprint arXiv:2401.14403*, 2024.

[10] AgileX Robotics. Ranger mini. https://global.agilex.ai/products/ranger-mini, 2023. Omnidirectional mobile robot platform.

[11] R. Ding, Y. Qin, J. Zhu, C. Jia, S. Yang, R. Yang, X. Qi, and X. Wang. Bunny-visionpro: Bimanual dexterous teleoperation with real-time retargeting using vision pro. 2024.

[12] Y. Park and P. Agrawal. Using apple vision pro to train and control robots, 2024. URL https://github.com/Improbable-AI/VisionProTeleop.

[13] G. Pavlakos, D. Shan, I. Radosavovic, A. Kanazawa, D. Fouhey, and J. Malik. Reconstructing hands in 3D with transformers. In *CVPR*, 2024.

[14] Shadow Robot Company. https://www.shadowrobot.com/. [Robotic hand].

[15] P. Mannam, K. Shaw, D. Bauer, J. Oh, D. Pathak, and N. Pollard. Designing anthropomorphic soft hands through interaction. In *2023 IEEE-RAS 22nd International Conference on Humanoid Robots (Humanoids)*, pages 1–8, 2023. doi:10.1109/Humanoids57100.2023.10375195.

[16] A. Agarwal, S. Uppal, K. Shaw, and D. Pathak. Dexterous functional grasping. In *Conference on Robot Learning*, pages 3453–3467. PMLR, 2023.

[17] S. Dasari, M. K. Srirama, U. Jain, and A. Gupta. An unbiased look at datasets for visuo-motor pre-training. In *Conference on Robot Learning*, pages 1183–1198. PMLR, 2023.

[18] B. Calli, A. Singh, J. Bruce, A. Walsman, K. Konolige, S. Srinivasa, P. Abbeel, and A. M. Dollar. Yale-cmu-berkeley dataset for robotic manipulation research. *The International Journal of Robotics Research*, 36(3):261–268, 2017.