University of Hong Kong

Data Collection in DexHand Imitation Learning

Qian Luo

05/12/2024

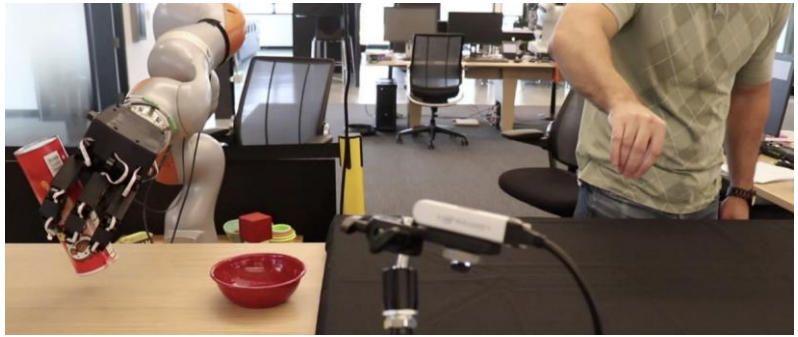
Contents

- Data Collection
  - Vision Teleporation
  - VR Teleporation
  - Motion Capture System
- Imitation Learning
  - 3D diffusion policy

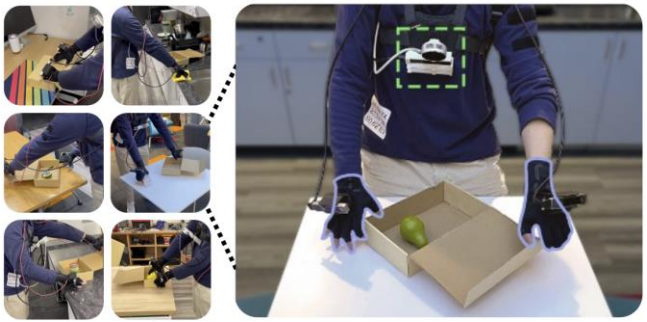
Data Collection

- Three Types
- Difference: The way captureing hand motions

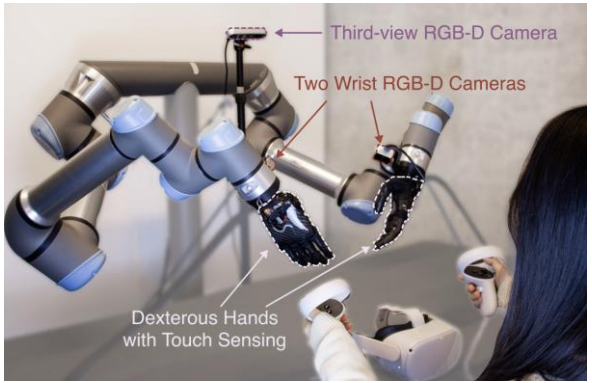
Vision Teleporation



Motion Capture System

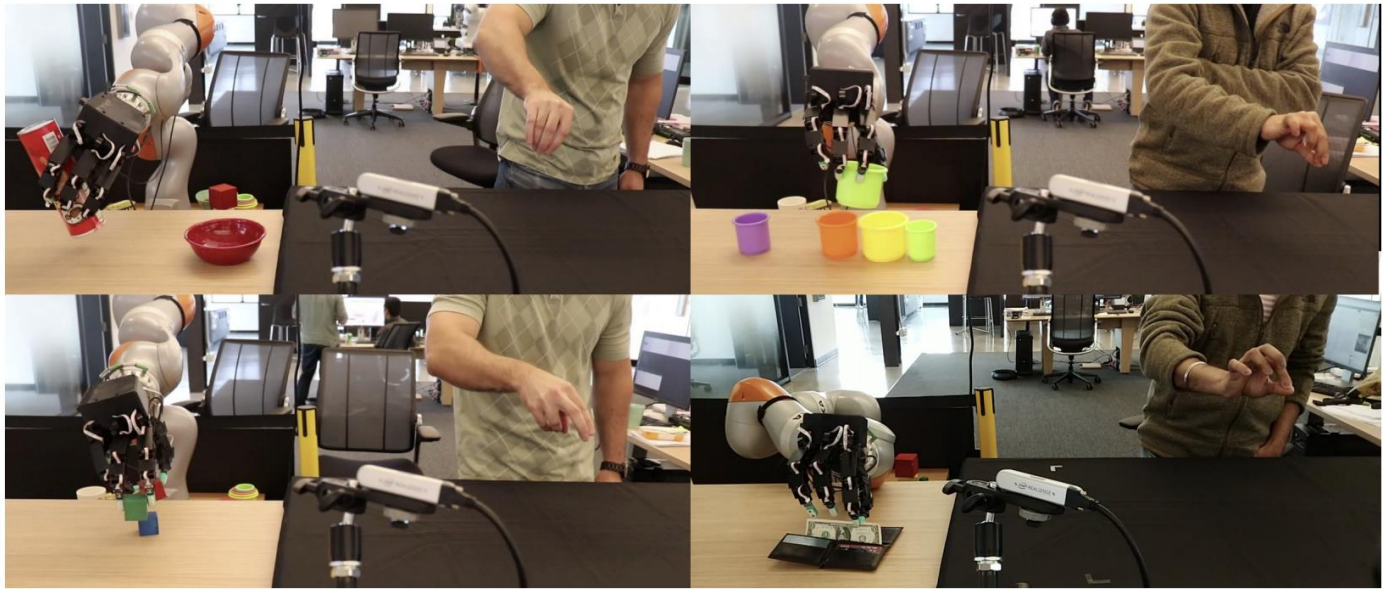


VR Teleporation



Vision Teleporation

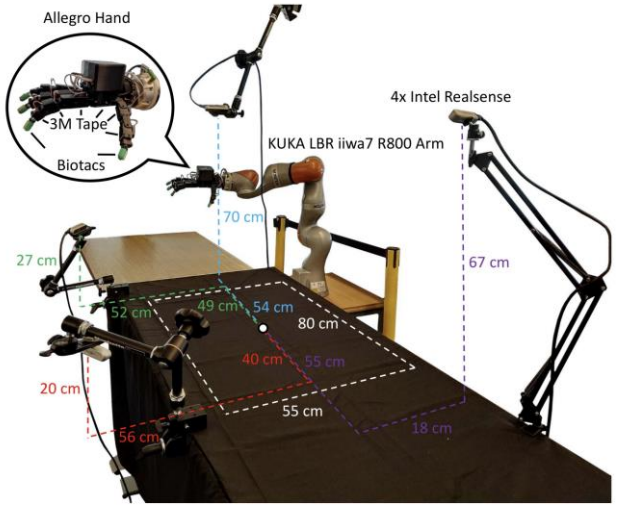
- Capturing visual information of the human hand using a camera and mapping the hand poses to a robotic hand for teleoperation.
- DexPilot(UW, 2019)

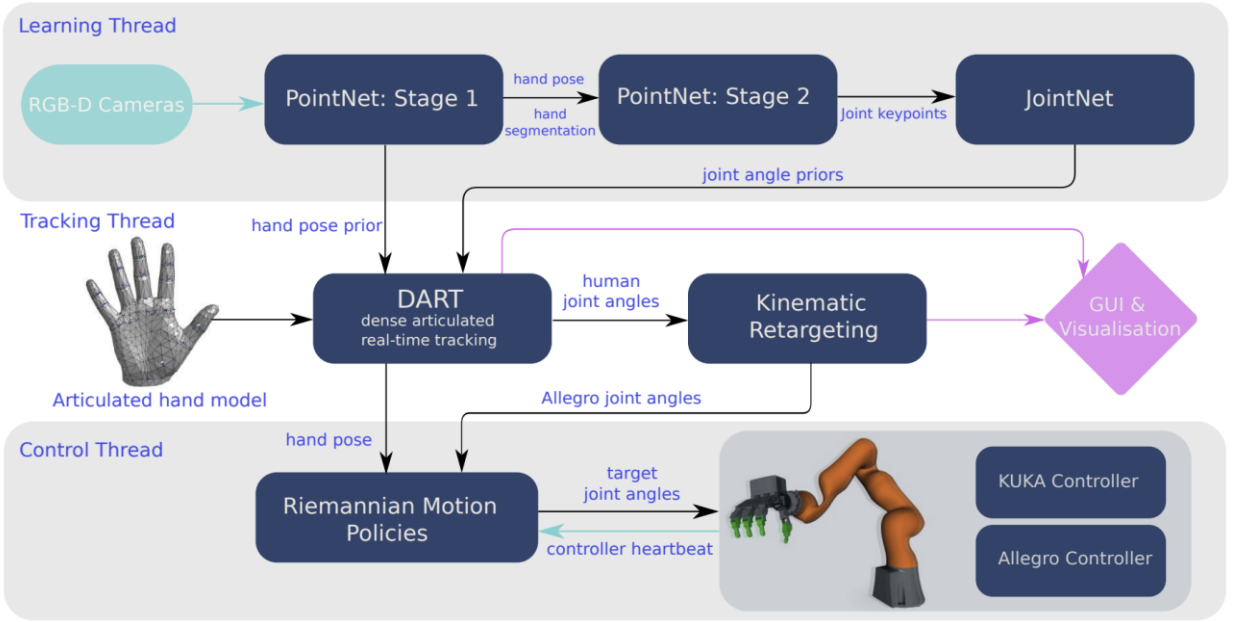


Handa et al. DexPilot: Vision Based Teleoperation of Dexterous Robotic Hand-Arm System

Vision Teleporation

- DexPilot(UW, 2019)



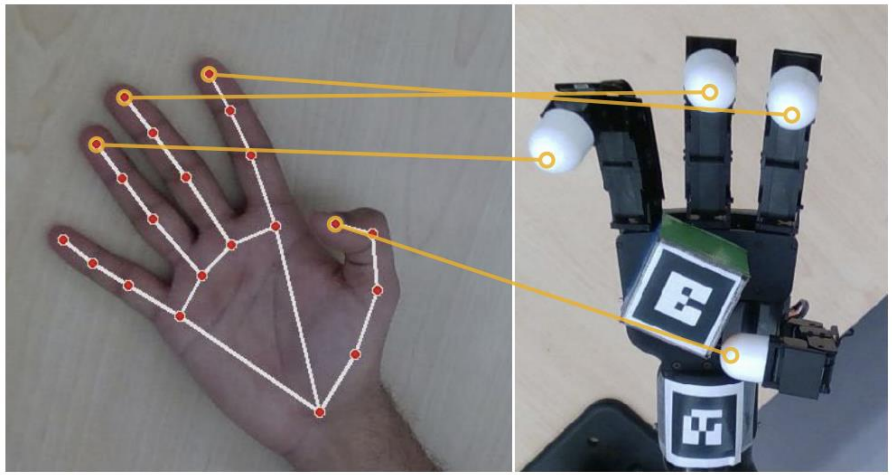


- Derive human hand poses from the 3D point cloud of the hand and re-target them to the robotic hand.

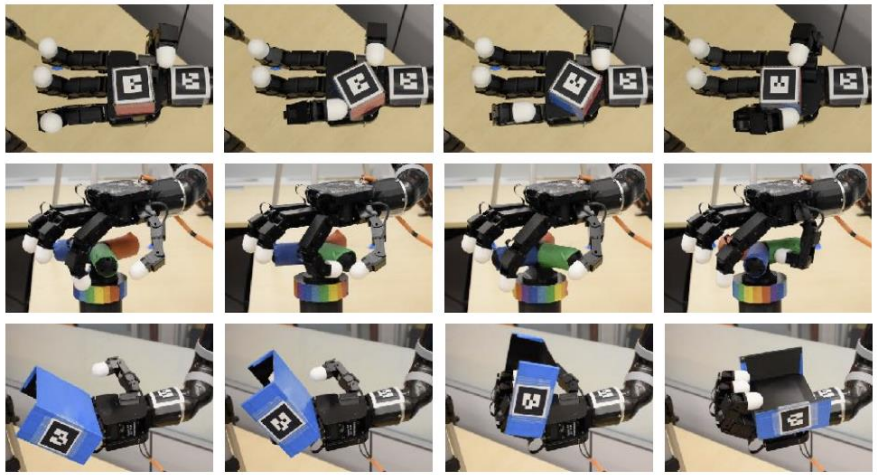
Handa et al. DexPilot: Vision Based Teleoperation of Dexterous Robotic Hand-Arm System

Vision Teleporation

- DIME(NYU, 2022)



(a) Teleoperation through a single RGB camera.



(b) Learned policies for dexterous manipulation.

- Obtain human hand poses from a single RGB image and re-target them to the robotic hand.

Arunachalam et al. Dexterous Imitation Made Easy: A Learning-BasedFramework for Efficient Dexterous Manipulation

Vision Teleporation

- DIME(NYU, 2022)

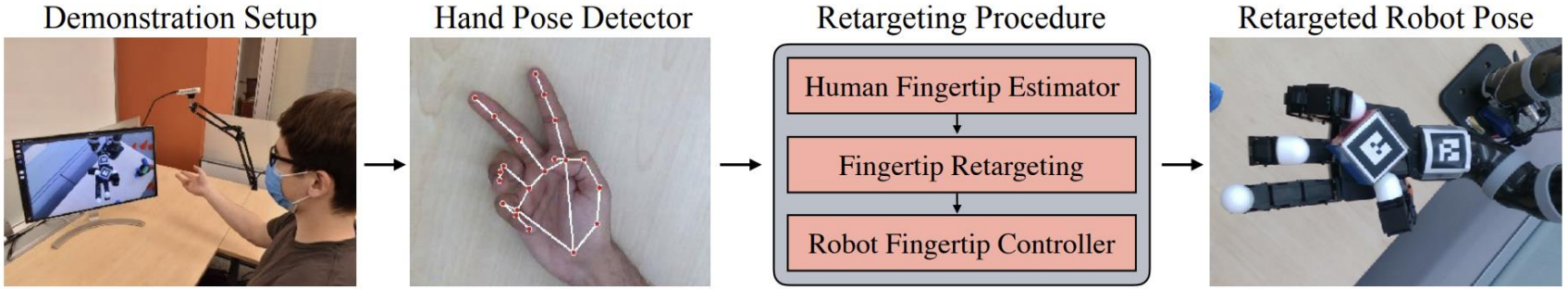
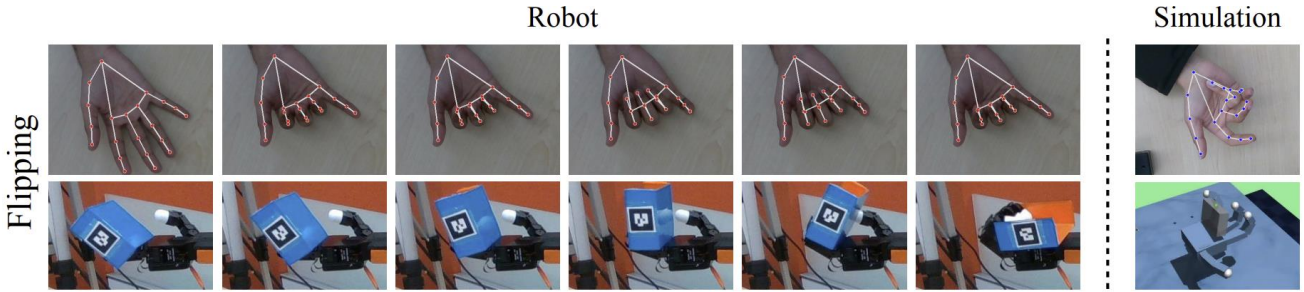


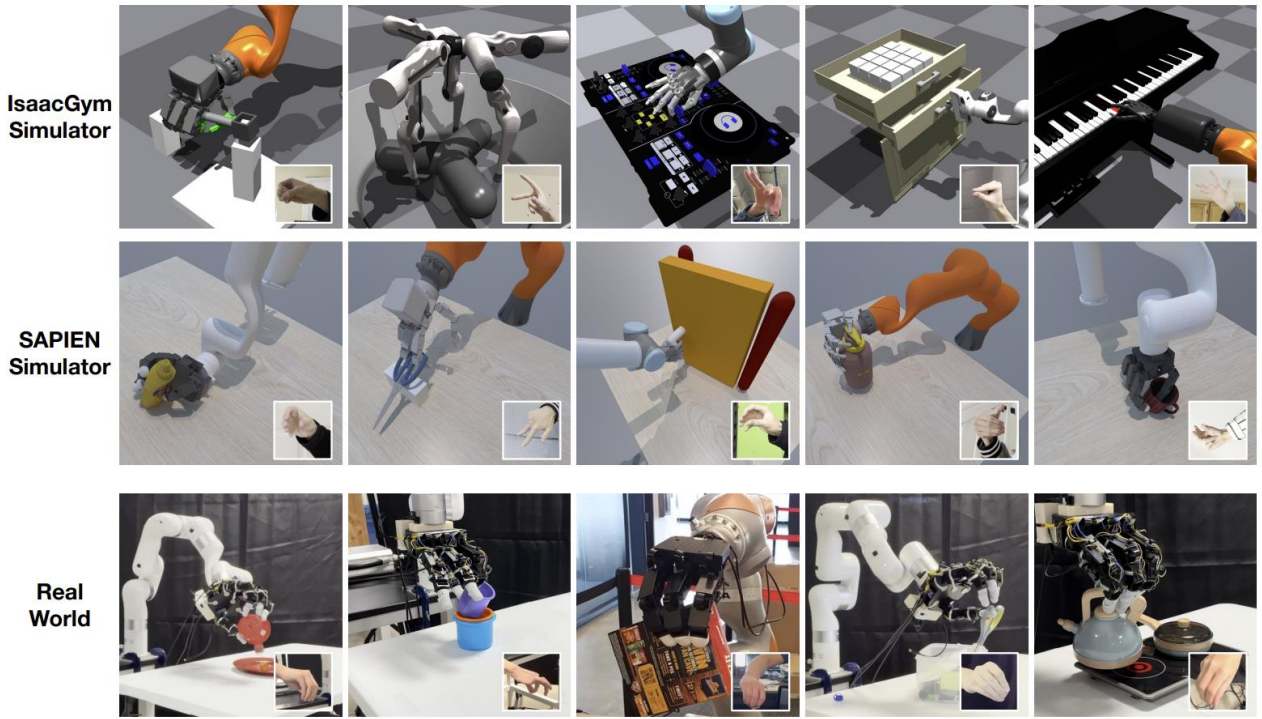
Fig. 2: Overview of the teleoperation framework in DIME. Given RGB streams of a human operator's hand, a hand pose detector followed by a retargeting procedure is used to control the fingertips of the robot's hand. Visual feedback of the teleoperated actions is then provided back to the operator for real-time teleoperation.



Arunachalam et al. Dexterous Imitation Made Easy: A Learning-BasedFramework for Efficient Dexterous Manipulation

Vision Teleporation

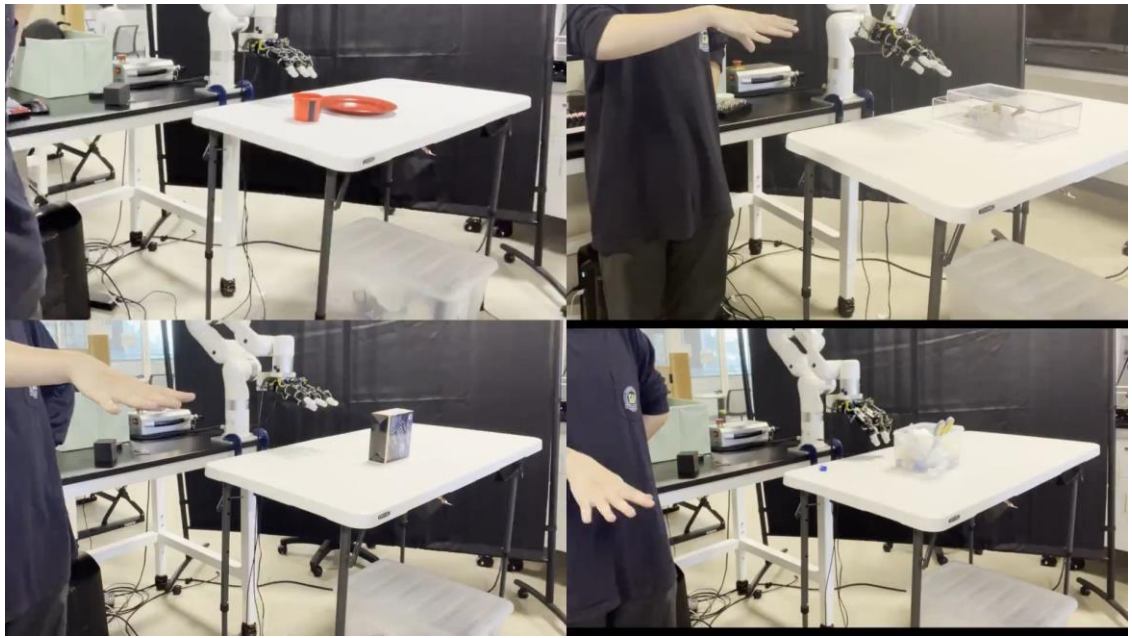
- AnyTeleop (UCSD, 2023)



Qin et al. AnyTeleop: A General Vision-Based Dexterous Robot Arm-Hand Teleoperation System

Vision Teleporation

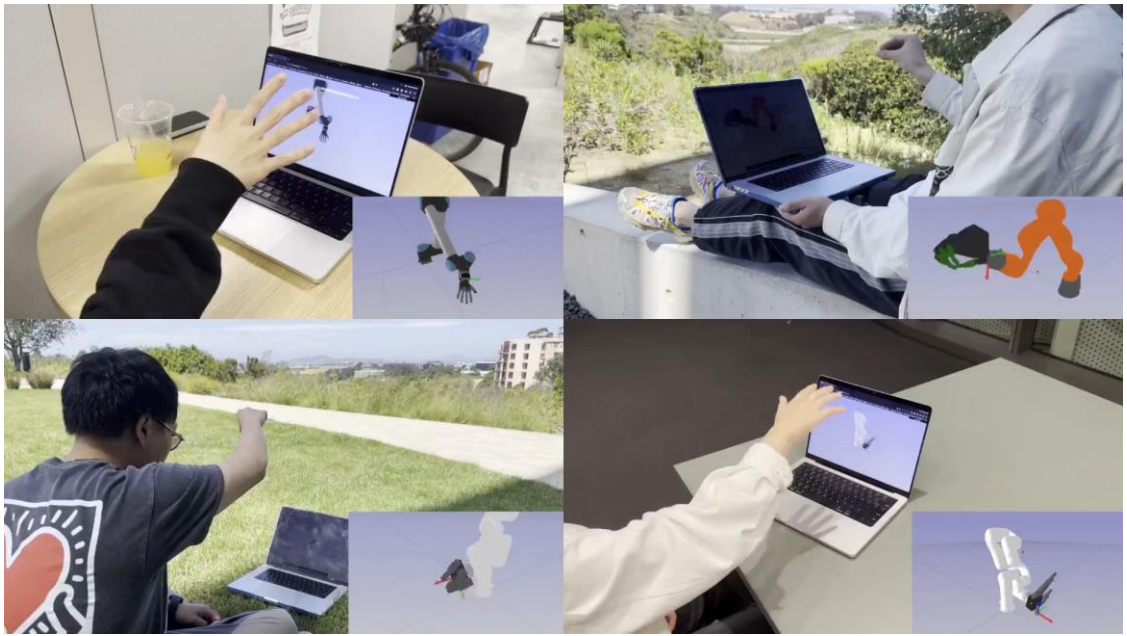
- AnyTeleop (UCSD, 2023)



Qin et al. AnyTeleop: A General Vision-Based Dexterous Robot Arm-Hand Teleoperation System

Vision Teleporation

- AnyTeleop (UCSD, 2023)



Qin et al. AnyTeleop: A General Vision-Based Dexterous Robot Arm-Hand Teleoperation System

Vision Teleporation

- AnyTeleop (UCSD, 2023)

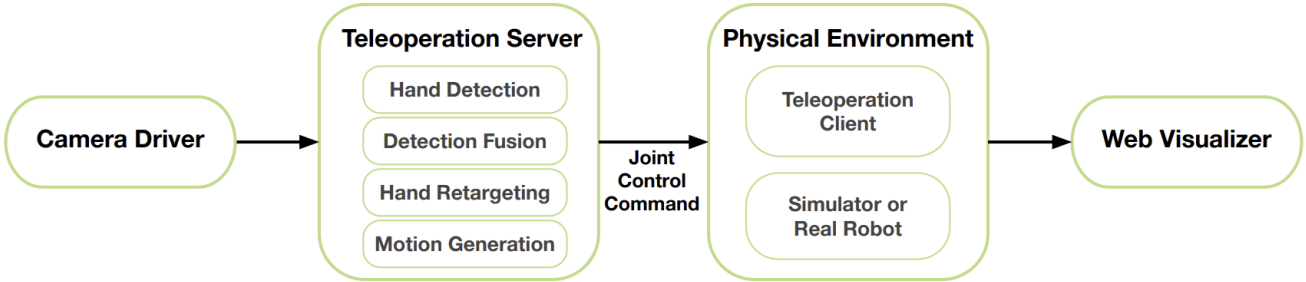


Fig. 3: **System Architecture.** AnyTeleop is composed of four components: (i) camera driver, which captures the human hand pose in RGB or RGB-D format; (ii) teleportation server, the core component in our system, which performs hand pose detection and converts detection results to robot control commands; (iii) teleoperated robot, which is either a real robot or a simulated robot in a virtual environment; (iv) web visualizer, which enables remote visualization across the internet.

- Human hand pose detection + re-targeting + motion generation.

Qin et al. AnyTeleop: A General Vision-Based Dexterous Robot Arm-Hand Teleoperation System

Vision Teleporation

Pros：

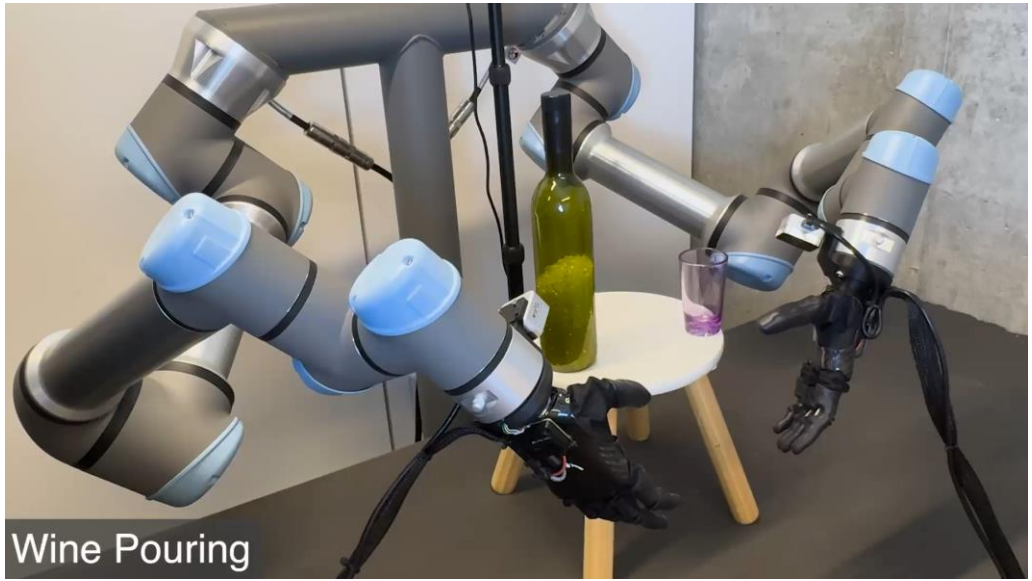
- Portable
- Real-time,low latency
- Realistic(real robot data, easy for policy training)

Cons：

- Relatively low accuracy (the occlusion problem is challenging to resolve).

VR Teleporation

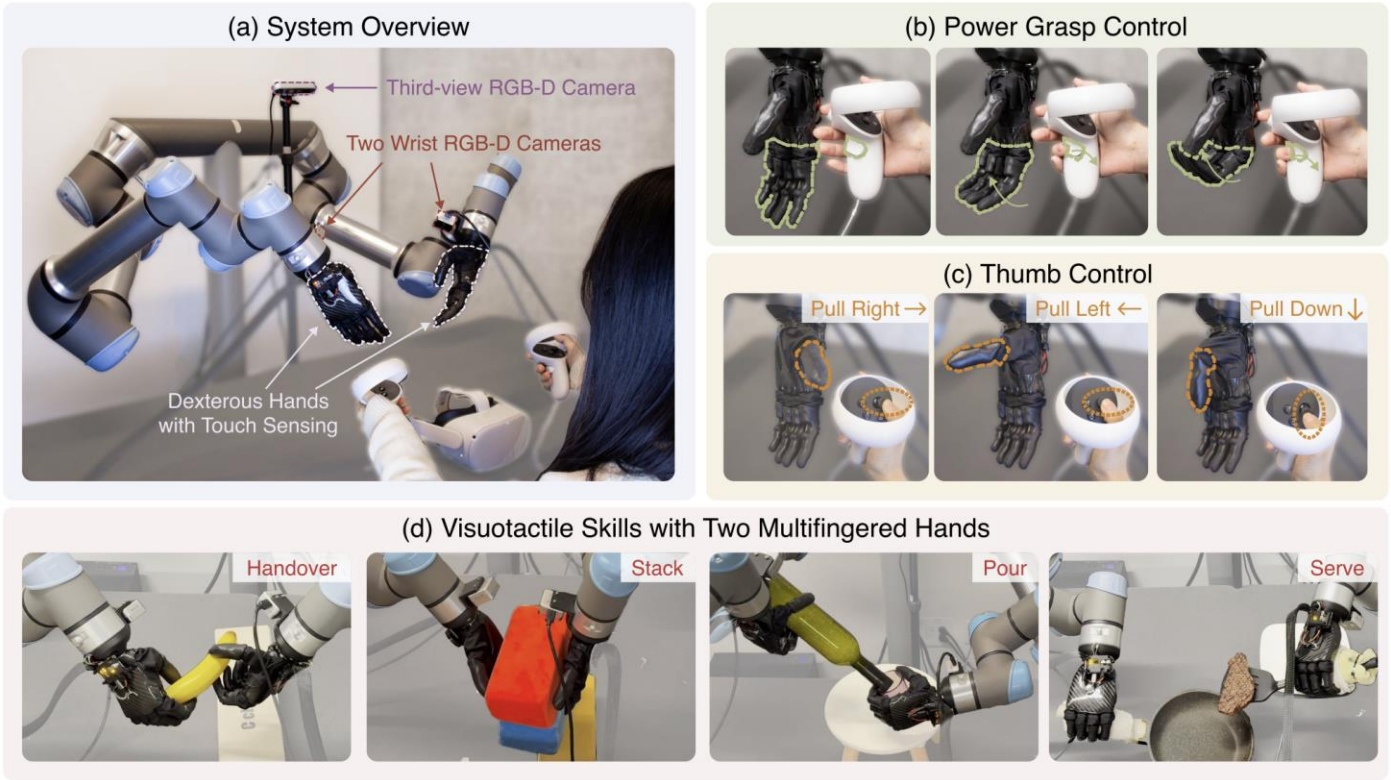
- Human teleoperate the robotic hand's movements through a (VR) headset.
- HATO(UCB, 2024)



Lin et al. Learning Visuotactile Skills with Two Multifingered Hands

VR Teleporation

- HATO (UCB, 2024)



Lin et al. Learning Visuotactile Skills with Two Multifingered Hands

VR Teleporation

Pros：

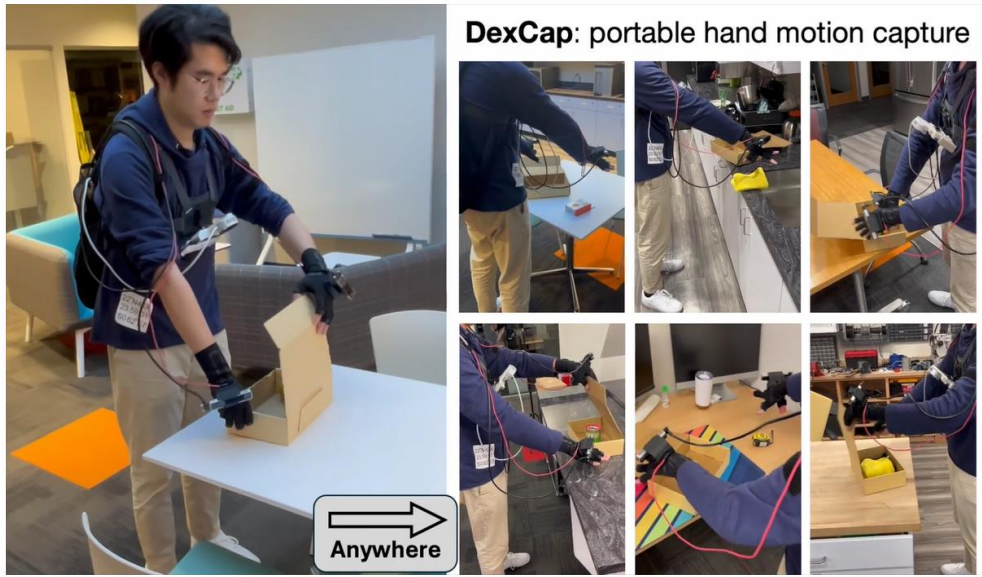
- Immersive teleporation
- Real-time,low latency
- Realistic(real robot data, easy for policy training)

Cons：

- Relatively low accuracy (the occlusion problem is challenging to resolve).

Motion Capture System

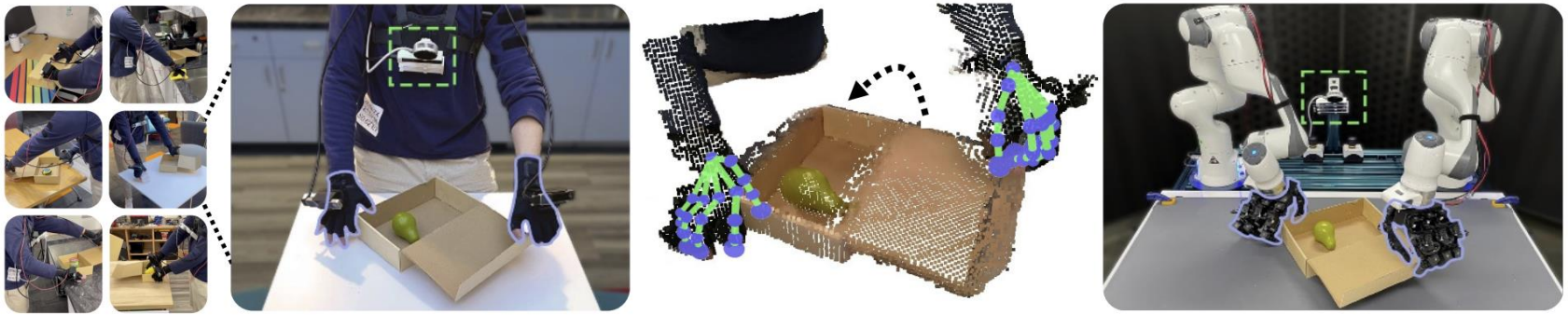
- Human hand directly performs the task, capturing hand poses during the process.
- DexCap (Stanford, 2024)



Wang et al. DexCap: Scalable and Portable Mocap Data Collection System for Dexterous Manipulation

Motion Capture System

- DexCap (Stanford, 2024)



(a) DexCap: Portable motion capture system (b) Mocap data and 3D scene (c) DexIL: Dexterous imitation learning

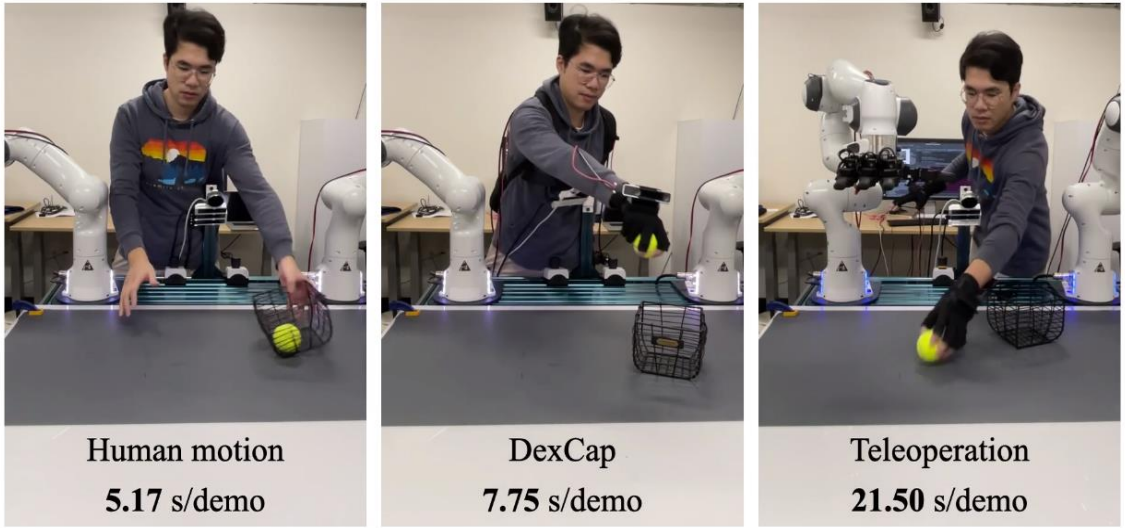
Fig. 1: **DexCap** facilitates the in-the-wild collection of high-quality human hand motion capture data and 3D observations. Leveraging this data, **DexIL** adapts it to the robot embodiment and trains control policy to perform the same task.

- Unlike teleoperation, DexCap directly uses data collected from the human hand for training and executes tasks on the robotic hand.

Wang et al. DexCap: Scalable and Portable Mocap Data Collection System for Dexterous Manipulation

Motion Capture System

- DexCap (Stanford, 2024)

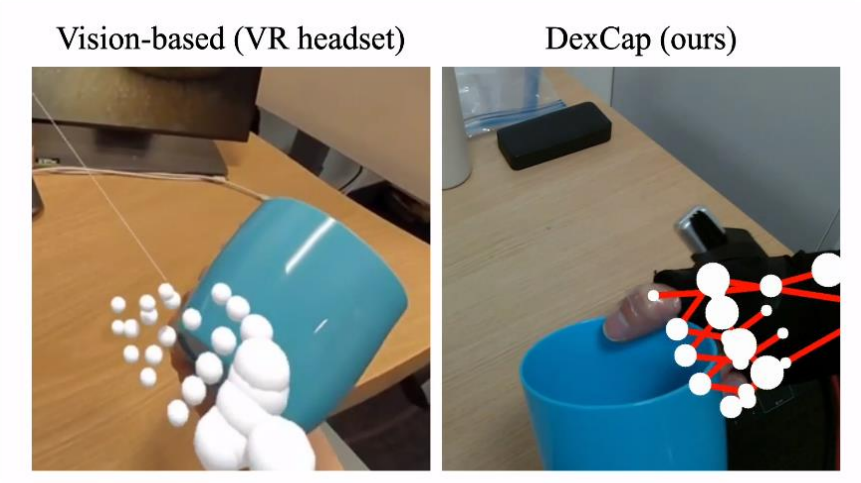


- Compared to teleoperation, motion capture is more efficient.

Wang et al. DexCap: Scalable and Portable Mocap Data Collection System for Dexterous Manipulation

Motion Capture System

- DexCap (Stanford, 2024)

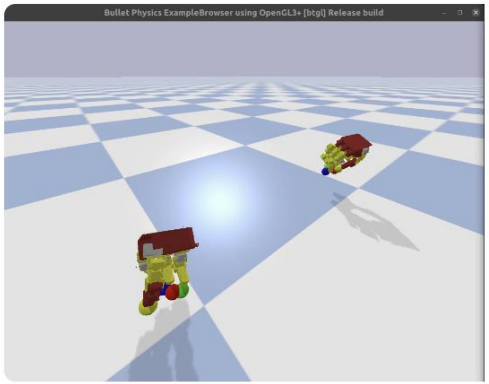


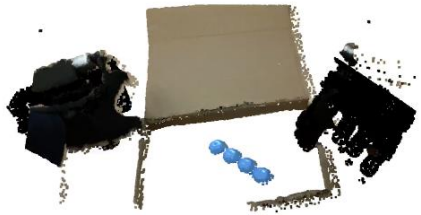
- Compared to vision-based capture, (IMU-based) motion capture offers higher accuracy.

Wang et al. DexCap: Scalable and Portable Mocap Data Collection System for Dexterous Manipulation

Motion Capture System

- DexCap (Stanford, 2024)





Visual gap: To further bridge the visual gap between human hand and robot hand, we use forward kinematics to genrate a point cloud mesh of the robot hand and add it to the pointcloud observation as is shown in this video.

- A visual gap exists. It is necessary to align the human hand and the robotic hand visually.

Wang et al. DexCap: Scalable and Portable Mocap Data Collection System for Dexterous Manipulation

Motion Capture System

Pros：

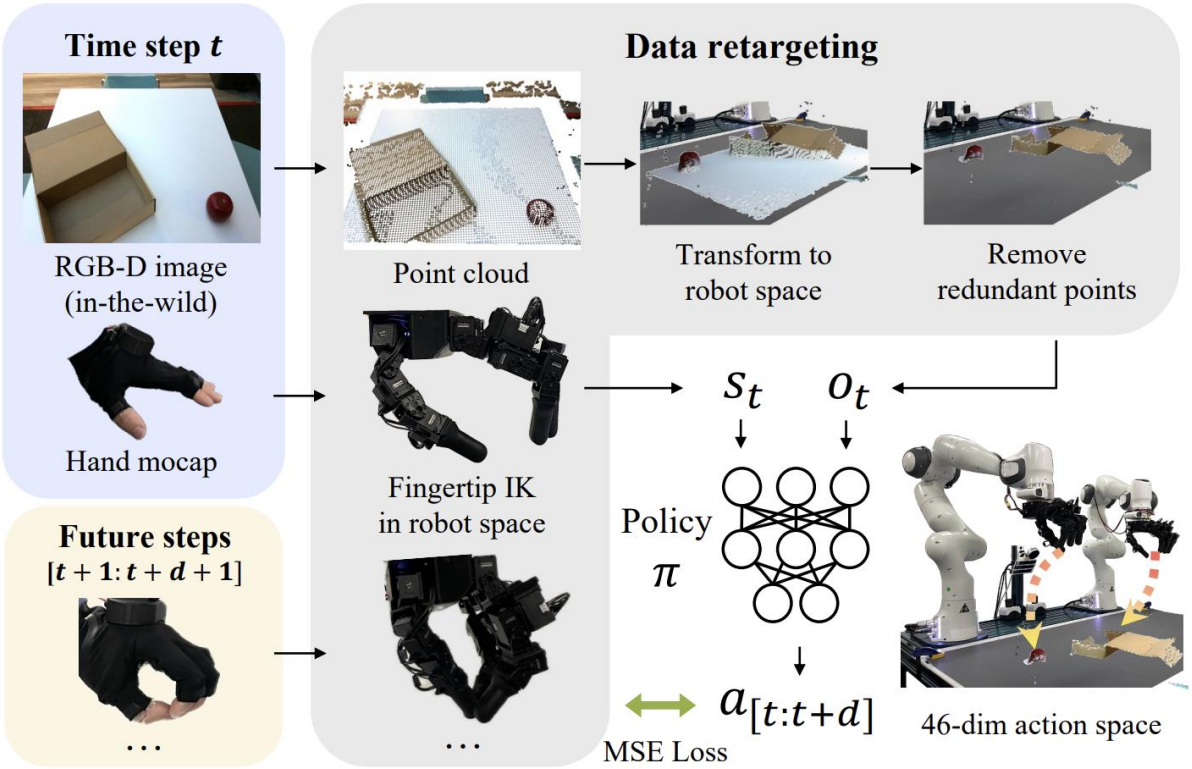
- Highly portable
- No latency
- High accuracy (IMU data)

Cons：

- It is necessary to compensate for the error between the human hand and the robotic hand.

Imitation Learning

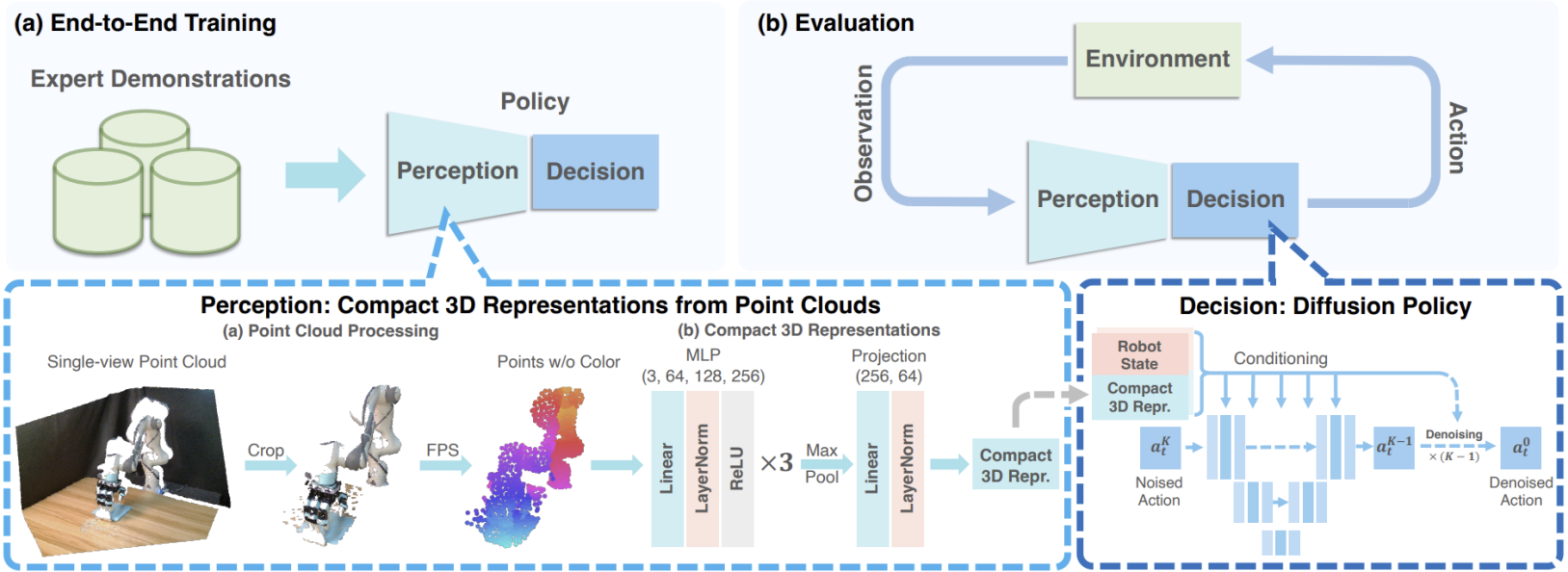
- Pipeline



Wang et al. DexCap: Scalable and Portable Mocap Data Collection System for Dexterous Manipulation

Imitation Learning

- 3D diffusion policy (THU, 2024)



- Perception: 3D point cloud Decision-making: Diffusion policy

Ze et al. 3D Diffusion Policy: Generalizable Visuomotor Policy Learning via Simple 3D Representations

Thank you for listening!