



ArtiGrasp: Physically Plausible Synthesis of Bi-Manual Dexterous Grasping and Articulation

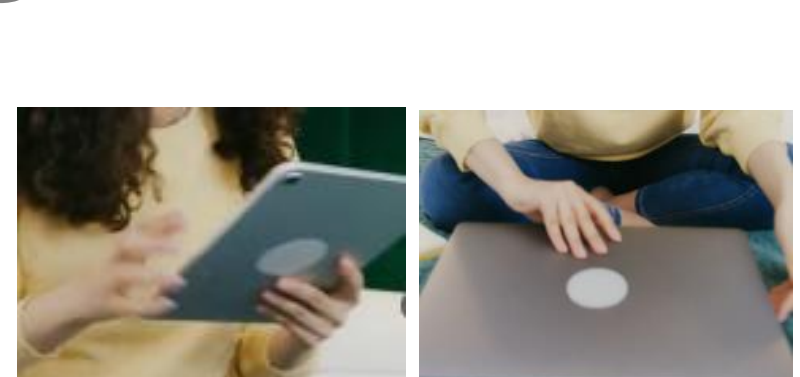
Hui Zhang*, Sammy Christen*, Zicong Fan, Luocheng Zheng, Jemin Hwangbo, Jie Song, Otmar Hilliges

* indicates equal contribution



Overview

1 Motivation



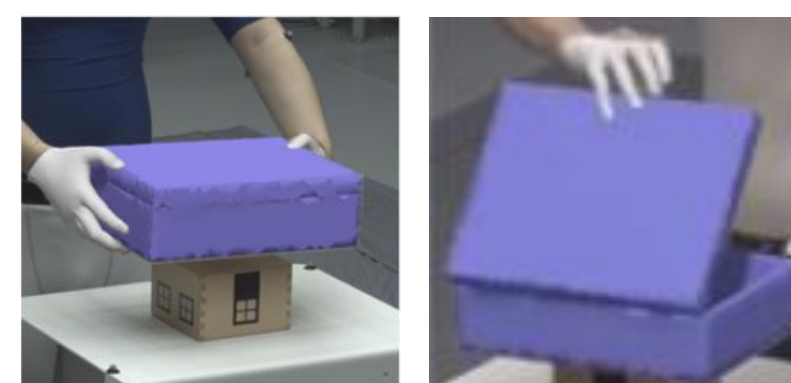
- Two hand manipulation
- Grasping & Articulation
- Expensive to capture

2 New Task: Dynamic Object Grasping and Articulation

Inputs: Target + References + States



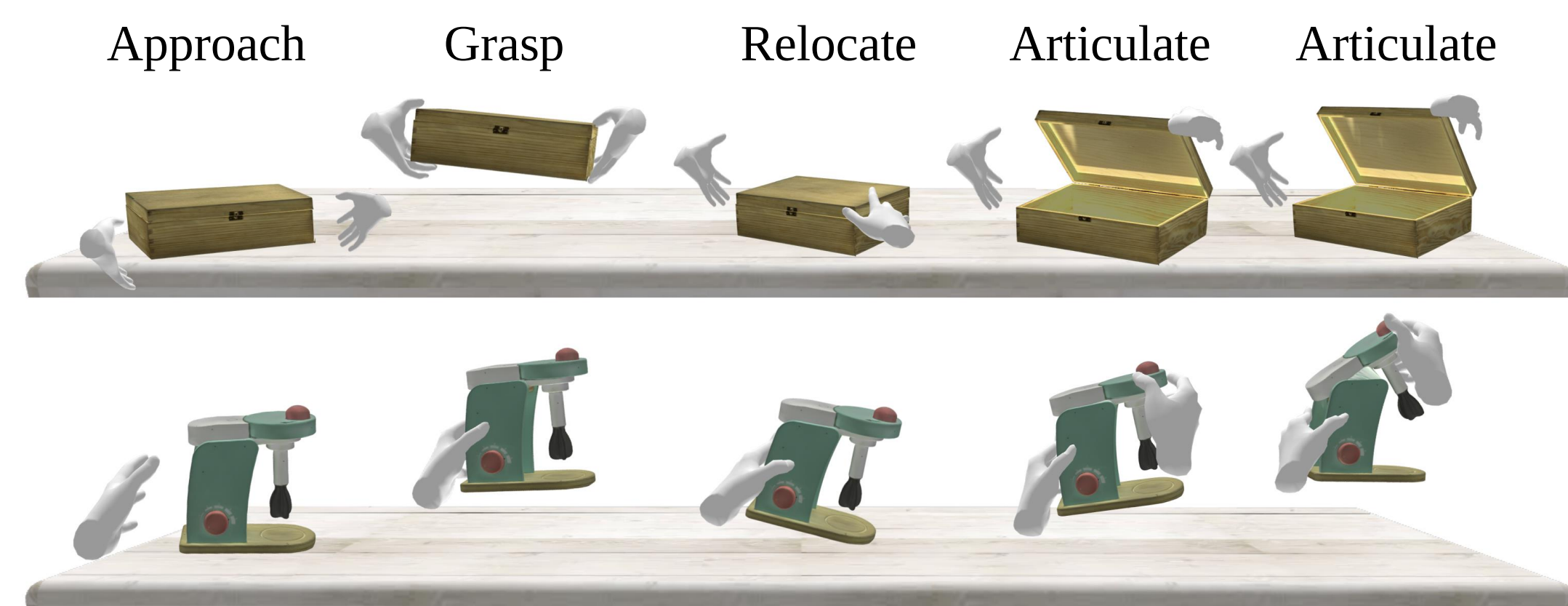
Target 7D Object Pose



A Pair of Grasp Reference Poses [2]



Output: A whole motion sequence that brings an object to target 7D pose



3 Challenges

Challenge: Expensive capture → Limited 3D motion data

Question: Fluent and natural motion with limited available data?

Challenge: Two hand manipulation + Two kinds of motions
→ Difficult control (Can easily affect each other)

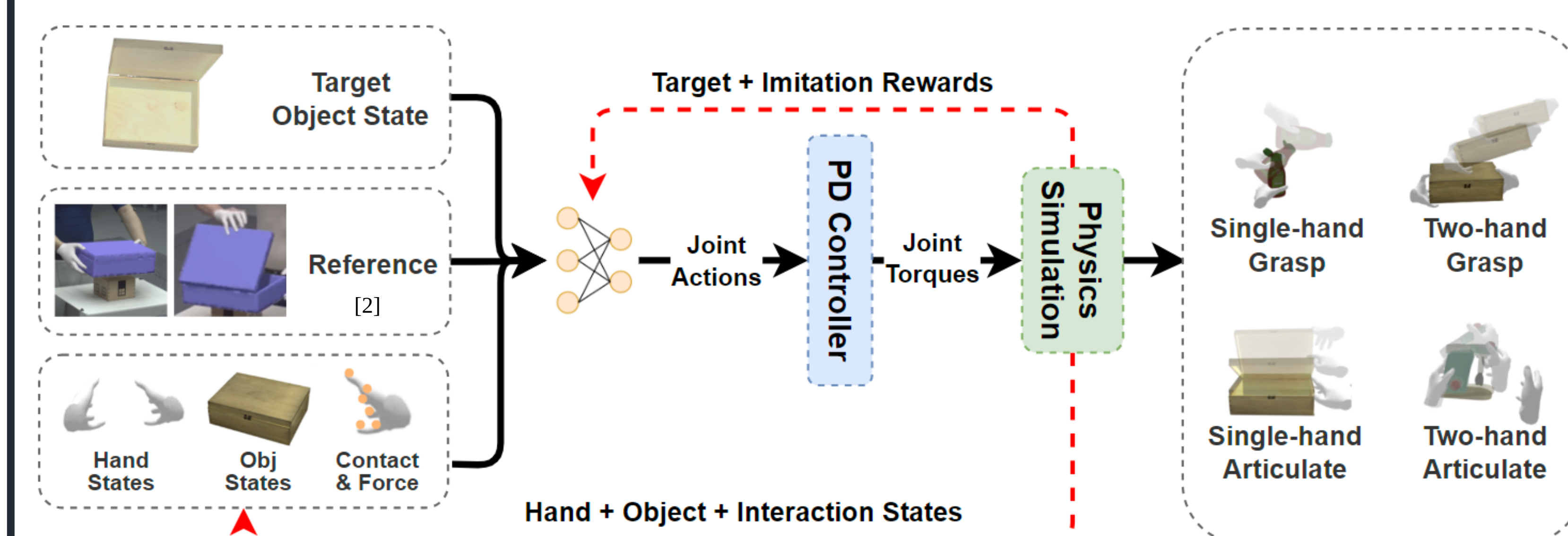
Inaccurate wrist motion

Inaccurate finger motion

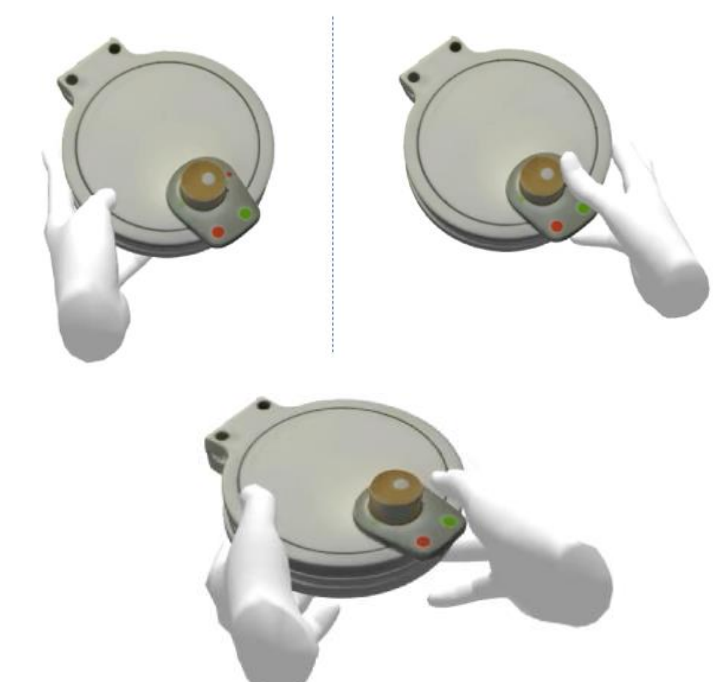
Question: Precise wrist & finger control
+ collaboration without affecting each other?

Method

1 Framework: Reinforcement Learning + Physics Simulation



2 Learning Curriculum

Setting: Stationary Object + Individual Hands
Goal: Simple Interaction + Precise Finger MotionSetting: Non-stationary Object + Two Hand Collaboration
Goal: Collaboration+ Precise Wrist Motion

3 Reward Function

$$r = r_{imitation} + r_{task}$$

Imitate
reference poseComplete
the task

$$r_{imitation} = r_{pose} + r_{contact} + r_{reg}$$

Punish Finger
Pose & Position
ErrorsEncourage
Target Contact
& ForcesPunish Object
Linear & Angular
Velocities

$$r_{task} = r_{move} + r_{angle}$$

Punish Object
MovementPunish Object
Joint Angle Error

$$= -w_{tq} ||\bar{\omega} - \omega|| - w_{tx} ||\bar{\mathbf{p}}^0 - \mathbf{p}||^2$$

 $\bar{\omega}, \omega$: target and current object joint angles $\bar{\mathbf{p}}^0, \mathbf{p}$: object initial and current positions

$$r_{pose} = - \sum_{i=1}^L w_{px}^i ||\bar{\mathbf{x}}^i - \mathbf{x}^i||^2 - w_{pq} ||\bar{\mathbf{q}} - \mathbf{q}||$$

 L : number of hand links $\bar{\mathbf{x}}, \mathbf{x}$: target and current link positions $\bar{\mathbf{q}}, \mathbf{q}$: target and current finger joint angles

$$r_{contact} = w_{cc} \frac{\bar{\mathbf{c}}^T \mathbf{f}_{>0}}{\bar{\mathbf{c}}^T \mathbf{c}} + w_{cf} \min(\bar{\mathbf{c}}^T \mathbf{f}, \lambda m_o)$$

 $\bar{\mathbf{c}}$: target contact vector of each hand link $\mathbf{f}$: current force vector of each hand link m_o : object mass

$$r_{reg} = -w_{rh} ||\dot{\mathbf{T}}||^2 - w_{ro} ||\dot{\Omega}||^2$$

 $\dot{\mathbf{T}}$: hand linear and angular velocities $\dot{\Omega}$: hand linear and angular velocities

Comparison to SOTA

	Grasping			Articulation		
Model	Suc. G ↑	PE ↓	AE ↓	Suc. A ↑	AAE ↓	SD ↓
PD+IK	0.13	1.20	1.50	0.28	0.80	0.39
D-Grasp [1]	0.72	0.12	0.62	0.22	0.93	0.49
Ours	0.71	0.13	0.69	0.55	0.57	0.01

Models	Suc. T ↑	PE ↓	AE ↓	AAE ↓
D-Grasp [1]	0.11	0.05	0.15	0.66
Ours	0.50	0.03	0.10	0.41

New Task

- Higher success rates (articulation & new task)
- Smaller errors

Suc. T: Task success rate

Suc. G: Grasp success rate

Suc. A: Articulation success rate

PE: Position error

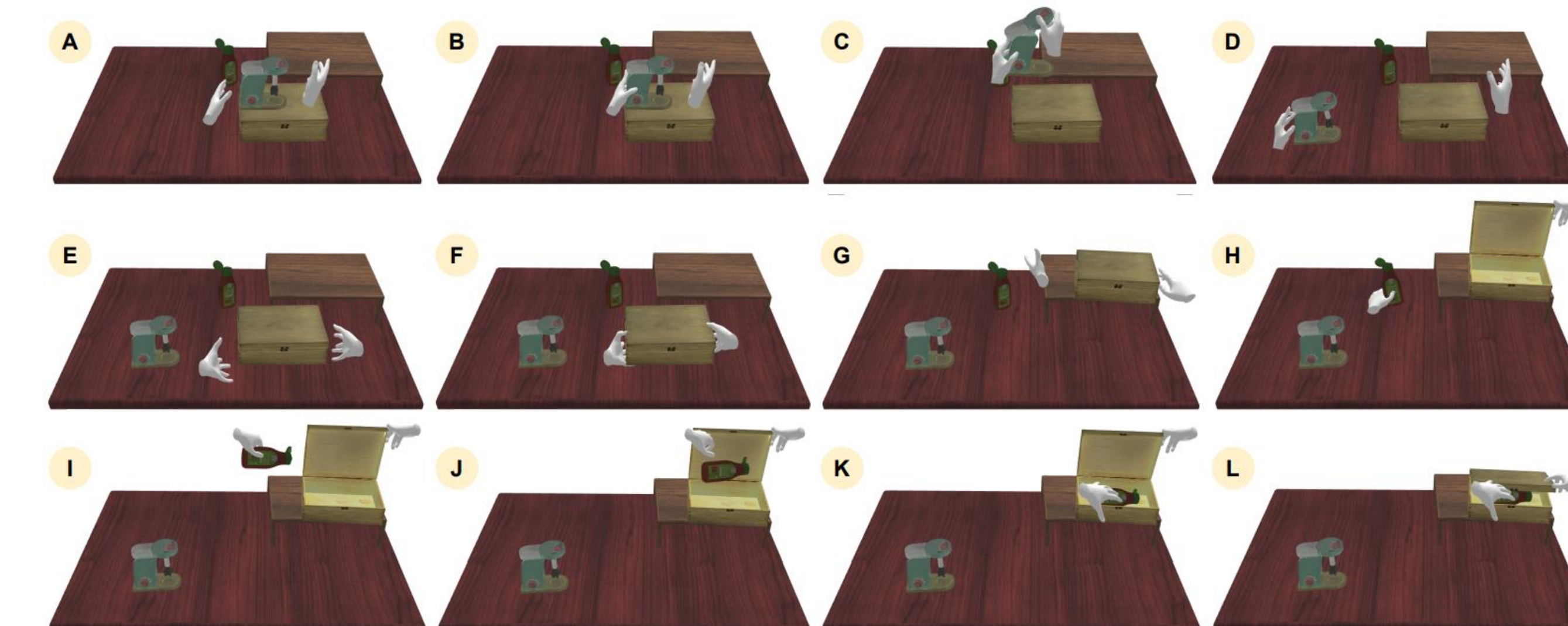
AE: Angle (orientation) error

AAE: Joint angle error

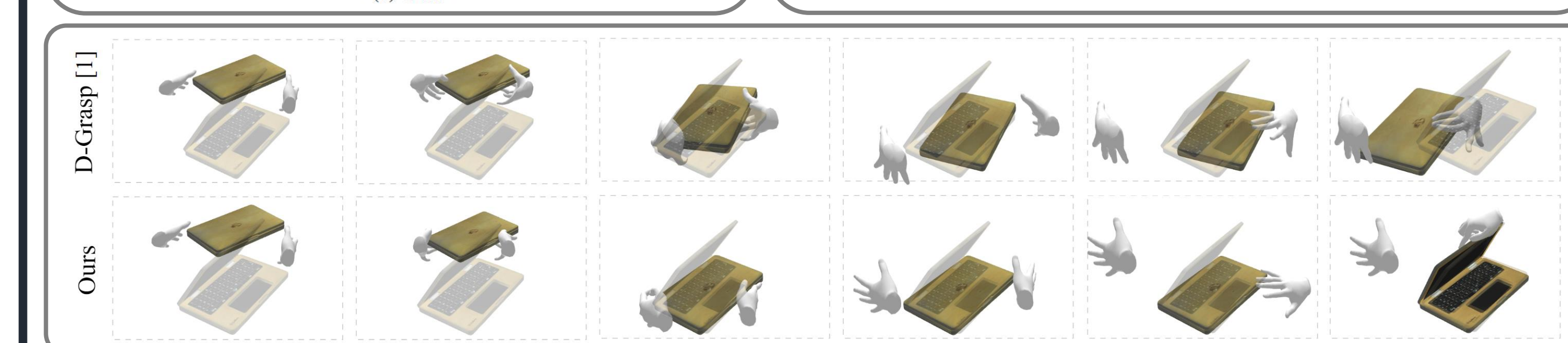
SD: Simulated distance

Qualitative Results

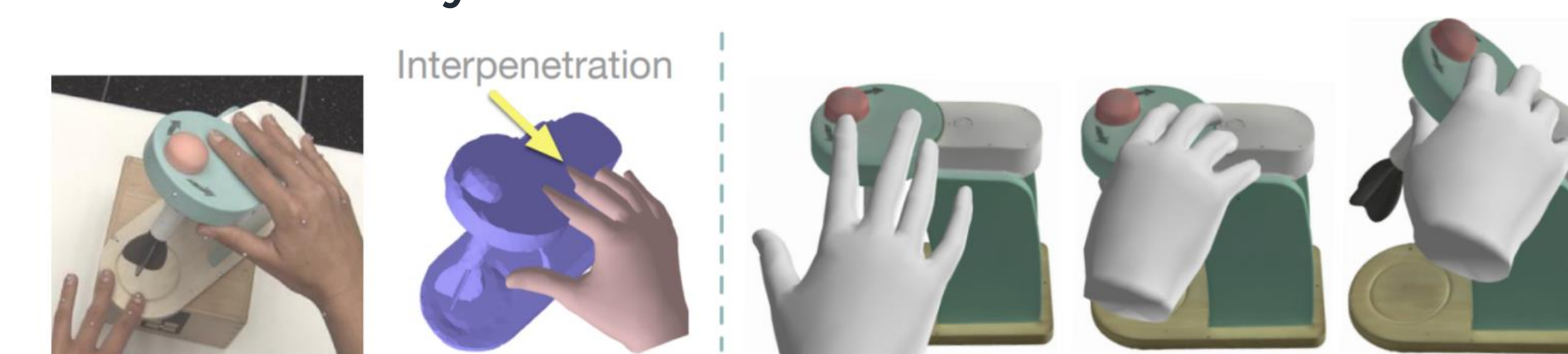
1 The Long Sequence Generated with Our Method



2 Qualitative Comparison with Baselines



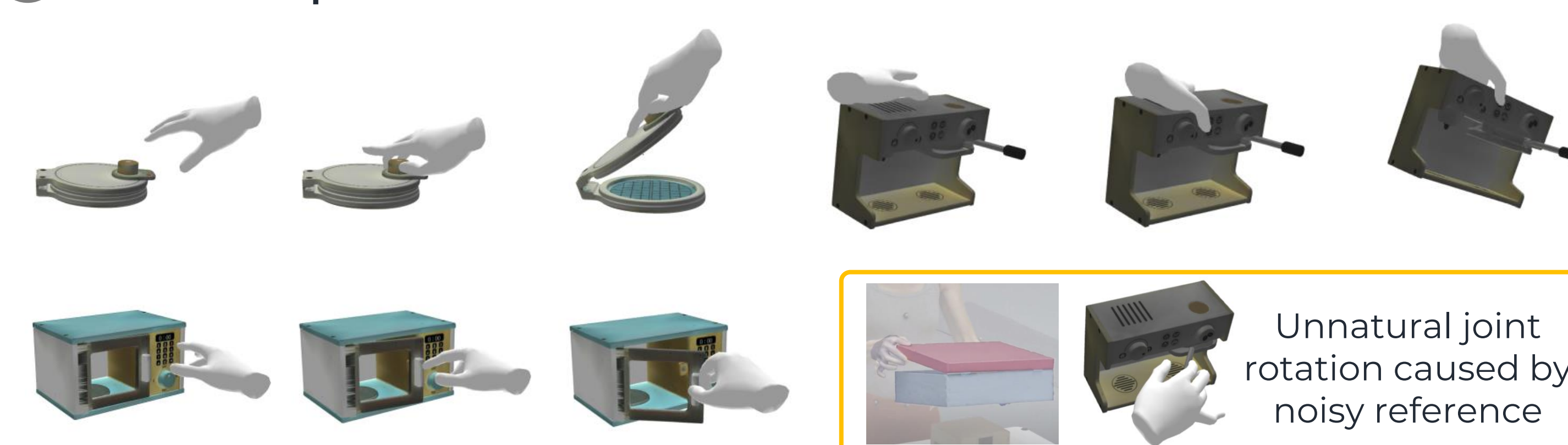
3 Synthesis with Noisy References



(a) Input image (b) Noisy recon. (c) Our motion generation

Generated Motion with a Noisy Reference Reconstructed from a Single RGB Image

4 More Examples & Limitations

Correspondence huizhang@ethz.ch

References

[1] Christen et. al, "D-Grasp: Physically plausible dynamic grasp synthesis for hand-object interactions", CVPR 2022

[2] Fan et. al, "ARCTIC: A dataset for dexterous bimanual hand-object manipulation", CVPR 2023