2024 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)
October 14-18, 2024. Abu Dhabi, UAE

Learning Bimanual Manipulation Policies for Bathing Bed-bound People

Yijun Gu and Yiannis Demiris

***Abstract*— Assistive robots hold promise in enhancing the quality of life for older adults and people with mobility impairments in daily bed bathing routines. When providing bathing assistance to bed-bound people, human caregivers often support the joints when lifting the arms and legs to properly wash and dry occluded areas. This research introduces a novel approach to robotic bed bathing manipulation, where a bimanual robot learns to lift a target limb while controlling a cleaning tool to bath the surface within safe force bounds. To ensure safe, cooperative bath manipulation, our work combines Multi-Agent Reinforcement Learning (MARL) framework with a variable impedance action space enabling adaptive interaction with the environment and carefully-designed reward functions regulating contact force on the human body. Simulation results demonstrate improved bathing area coverage compared to unimanual models and exhibit great adaptability to contact-rich interaction within a safe force boundary. We validate our approach across various human body sizes, showcasing its generalizability. We also transfer our models to a physical Baxter robot bathing a medical-grade manikin. We further incorporate a force tracking controller with the trained models to enhance adaptation to noisy real-world bathing scenarios. To the best of our knowledge, this is the first robot-assisted bed bathing application that performs autonomous bathing around the human body using bimanual robot arms.**

I. INTRODUCTION

The World Health Organization reports that 1.3 billion individuals currently experience significant disability and need assistance with daily activities, including bathing [1]. Robot caregivers provide the potential for these individuals to clean themselves, maintain personal privacy, and reduce the workload on healthcare providers. Recent research has extensively explored robotic bathing assistance using a single-arm robot [2]–[7]. Nevertheless, challenges arise when parts of the body are inaccessible, especially when the patient is lying on the bed. Most existing approaches assume humans can position themselves in a desired pose so that the robot only focuses on the exposed surface. On the other hand, human caregivers often use both arms to lift the body part in order to reveal the obstructed surfaces. Therefore, bimanual robotic bathing assistance presents significant potential to address this limitation.

In this research, we formulate the problem of bimanual robotic bed bathing manipulation: The robot employs two arms to lift the target limb of a person lying in bed while controlling a cleaning tool to bath off the dirt on the surface. This task, factorized into two sub-tasks: handling long-horizon coverage path planning and adapting safe motion in a contact-rich, dynamic environment, presents several challenges. First, designing controllers for this task requires careful consideration of complex collision avoidance and coordination behaviors. Specifically, controllers must navigate around human and inter-arm collisions, limiting the joint workspace of the robot arm, and making traditional planning difficult for this task. Second, the diversity in human shapes and resting poses causes high costs and risks during physical human-robot interaction. Moreover, contact-rich tasks involve many unforeseen changes in contact location and properties that can significantly degrade the robot's behavior. These tasks demand additional modeling and force regulation efforts. Recent studies have shown that Reinforcement Learning (RL) managed to learn robust controllers for contact-rich tasks, such as wiping [8]–[10]. Inspired by these advancements, we propose using Multi-Agent Proximal Policy Optimization (MAPPO) [11], an RL algorithm efficient for learning complex cooperative behaviors. We leverage simulation which provides a safe environment to learn the policies with a wide variety of human models. To ensure safety, we introduce a variable impedance action space to adapt the robot's stiffness and damping properties based on the interaction with the environment and carefully design a force reward to regulate contact force on the human body within a safe boundary.

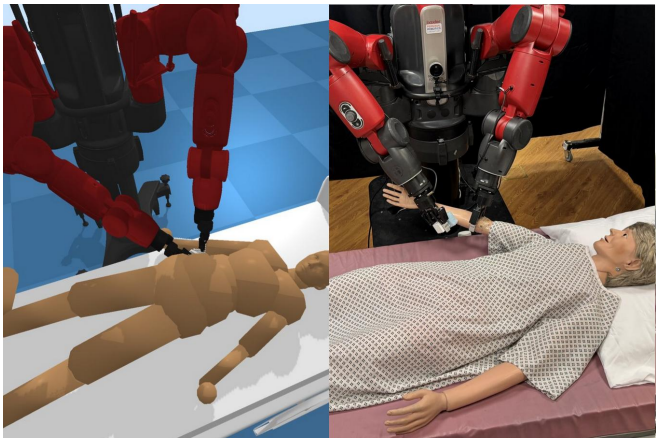

Fig. 1. A Baxter bimanual robot learns to employ dual arms for bathing bed-bound people. (Left) We train the control policies in simulation and (Right) transfer the capabilities to the real world with a physical Baxter robot and a medical-grade manikin.

In summary, we make the following contributions:

- We introduce a novel bed bathing formulation. Our

Yijun Gu and Yiannis Demiris are with the Personal Robotics Lab, Department of Electrical and Electronic Engineering, Imperial College London, London SW7 2AZ, United Kingdom. Emails: e.gu21@imperial.ac.uk; y.demiris@imperial.ac.uk. This work was supported in part by UKRI under Grant EP/V026682/1, and in part by a Royal Academy of Engineering Chair in Emerging Technologies. Videos are available on our project website: https://www.imperial.ac.uk/personal-robotics/research/chri/bimanual-bathing.

method enables a robot to employ dual arms to lift the target limb of a person lying in bed while bathing the surface around the limb.
- We propose a framework combining a Multi-Agent Reinforcement Learning method, MAPPO, with a variable impedance action space and carefully-designed reward functions. We show its efficacy for handling our bed bathing task and demonstrate the generalizability with unseen human body shapes in simulation.
- Considering the noise and sim-to-real gap, we employ a force tracking controller to regulate the actions learned from the trained policies to avoid big contact force on human body and demonstrate that bimanual bathing manipulation policies learned in simulation can be transferred to the physical dual-arm robot in real world.

II. Related Work

A. Robot-Assisted Bed Bathing

Bed bathing is a crucial component of nursing care, requiring the active involvement of caregivers for the comfort, safety, well-being, and dignity of the individual [12]. Recent research has explored real robotic systems that provide autonomous bathing assistance to bed-bound people. King et al. introduced a compliant force control that wipes the debris off an operator-selected area of a person's arm and leg [2]. Dometios et al. integrated motion primitives with vision-based motion planning to adapt wiping motion on moving, broadly curved surfaces, such as user's back [3]. Erickson et al. proposed capacitive sensing for tracking human limb contours and planning bathing trajectories [4]. Huang et al. introduced a contour-following controller utilizing a depth camera-based soft tactile sensor, demonstrating safe cleaning of human limbs and torsos [5]. Madan et al. fused RGB-thermal multimodal perception with compliant control (RABBIT) for safe bathing manipulation [6]. Gu et al. also applied multimodal perception, utilizing visual and tactile sensing to track body contours and surfaces and employed a Transformer-based imitation learning method to learn bathing actions from human demonstrations [7]. Additionally, Liu et al. designed a unique mesoscale wearable robot that locomotes along the body surface to provide skincare [13]. Despite the effectiveness demonstrated in providing bed bathing assistance, these approaches primarily focus on a single manipulator and do not fully address bathing on occluded surfaces.

In contrast, our research proposes an approach that enables comprehensive bathing around the target limb while ensuring safe contact with the human body. We harness the capabilities of bimanual manipulators, incorporating a variable impedance action space and carefully-designed reward functions to facilitate navigation around target limbs and force interaction during the bathing process.

B. Bimanual Robotic Manipulation

Recent advancements in robotic manipulation primarily focus on single-arm robots, limiting task complexity due to dexterity constraints. Bimanual manipulation, facilitated by dual-arm systems, extends the operation workspace, enabling robots to perform intricate cooperative tasks such as food cutting [14], [15], garment manipulation [16]–[18], and wiping [19], [20]. Traditional planning methods require precise modeling and predicting of the system dynamics and careful consideration of complex collision avoidance. Also, the methods may easily cause deviations from the desired trajectory due to noise, uncertainty, and coordination errors with dual arms. RL has shown efficacy in designing effective bimanual controllers as demonstrated by Grannen et al. for bimanual scooping in food acquisition [21] and Zhu et al. with a standardized set of bimanual manipulation environments for robot learning [22]. However, RL methods face challenges in exploration and cooperation relations between the arms, promoting recent work to treat dual arms as multiple agents. Chen et al. leveraged Multi-Agent Reinforcement Learning (MARL) for dexterous hand manipulation, showing improved performance in bimanual cooperative manipulation [23].

In this work, we explore MARL in bimanual bathing, investigating how a robot utilizing dual arms can effectively bath around a target limb of a bed-bound person.

C. Force Control for In-contact Robotic Manipulation

Force control is essential for in-contact robotic manipulation where the robot frequently interacts with the environment with varying forces. Extensive research has developed robust force control strategies across different applications, such as assembly [24], [25], polishing [26], [27], surface following [8]–[10], as well as bed bathing [5], [6].

In complex contact-rich tasks, variable impedance is widely adopted as a powerful approach to address contact uncertainties during interaction. The robot learns to dynamically adjust both the impedance and arm posture to perform the task at the desired path while ensuring safe force interactions. Previous research has studied the use of RL to learn an optimal control policy by exploring the dynamics of contact-rich environments. Martín-Martín et al. proposed a variable impedance control in end-effector space (VICES), studying the effect of different action spaces for contact-rich manipulation tasks, including wiping a table [8]. Bogdanovic et al. presented a variable impedance control in joint space for surface sliding, where gains are learned with Deep Deterministic Policy Gradient (DDPG) [9].

In this work, we explore using RL to learn variable gains, enabling the robot to adjust actions to contact-rich interactions. In the real world, we further implement a force tracking controller to regulate the actions, mitigating the risk of applying big contact force on the human body during bimanual bathing manipulation.

III. Learning Bimanual Bathing Manipulation Policies for Bed-bound People

In this section, we describe the bed bathing manipulation environment in simulation and formulate the task as a MARL problem. We then present the algorithms in detail to address this problem, involving a variable impedance action

space and carefully-designed reward functions. Lastly, we introduce the training details employed in the bathing task.

A. Simulation Environment

Our bimanual bathing environment builds on Assistive Gym [28]. A bimanual Baxter humanoid robot is generated in the simulation environment equipped with two Robotiq 2F-85 adaptive grippers. The right arm of the robot holds a cleaning tool and the left arm lifts the target limb of a person lying in the bed in a randomly generated resting pose. The robot aims to move both arms simultaneously in order to clean the surface around the limb. In this work, the robot learns to bath four target limbs including the forearm, upperarm, shin, and thigh. Each target limb is represented by the positions of two supporting joints: elbow and wrist joint for forearm, shoulder and elbow joint for upperarm, knee and ankle joint for shin, and hip and knee joint for thigh.

As shown in Fig. 2, we represent the human using a configurable capsulized female model. As provided by Assistive Gym, we establish the human model with a fixed body size based on the published 50th percentile values. The human model is initially generated above a hospital bed with arms and legs slightly spread (20 and 10 degrees, respectively). Small variations are added to the joint positions of the human model before dropping the model from 1m above the ground onto the bed. Once the model has settled on the bed, the robot randomly generates a grasping point p_g from a region around the supporting joints: $p_g = p_{j_1} + \omega * (p_{j_2} - p_{j_1})$ where p_{j_1} and p_{j_2} are the two supporting joints, and ω is a percentage used to adjust the position of the grasping point based on the distance between p_{j_1} and p_{j_2}. The robot then moves its left end effector to p_g and grasps the target limb. We then uniformly distribute bathing markers (1cm apart) around the outer surface of the target limb for the robot to bath them off during the task. We view the markers as a point cloud M of the area to be bathed. We define a function $C_t(x, \mathcal{T})$ to classify, at time step $t \in T$, whether a given bathing marker $x \in M^t$ is bathed by the cleaning tool $\mathcal{T}$ based on whether the Euclidean distance between the marker and the tool is within a threshold distance λ = 2.5cm:

$$C_t(x, \mathcal{T}) = \begin{cases} 1, & \text{where } ||x - \mathcal{T}||_2 < \lambda,\ x \in M^t \\ 0, & \text{otherwise} \end{cases}$$

The marker is considered bathed if $C(x, \mathcal{T}) = 1$, otherwise unbathed. At each time step when the tool is in contact with the human body, we compute the function for all the markers and remove the markers that are classified as bathed.

B. Problem Formulation

In this research, we formulate our bimanual bathing manipulation task as a decentralized partially observable Markov Decision Process (DEC-POMDP). A DEC-POMDP is defined as a tuple $\mathcal{M} = (\mathcal{S}, \mathcal{A}, \mathcal{O}, \mathcal{P}, \mathcal{R}, n, \gamma)$, where $\mathcal{S}$ is the state space, $\mathcal{A}$ is the shared action space for each agent i, $o_i = \mathcal{O}(s; i)$ is the observation for agent i at global state s. $\mathcal{P}$ is the transition dynamics given by $\mathcal{P}(s'|s, A)$ given the joint action $A = (a_1, a_2, \ldots, a_n)$ for all n agents. $\mathcal{R}$ is the shared reward for all agents, and $\gamma \in [0, 1)$ is the discount factor. The goal is to learn a policy $\pi_\theta(a_i|o_i)$ that jointly optimizes the accumulated reward in an episode with T time steps: $\mathbb{E}[\sum_{t=0}^{T-1} \gamma^t \mathcal{R}(s^t, A^t)]$ where $A^t = (a_1^t, \cdots, a_n^t)$ is the joint action at time step t. We employ MAPPO to learn the task. The dual manipulators are treated as independent agents and each agent executes its own policy with individual observations and actions.

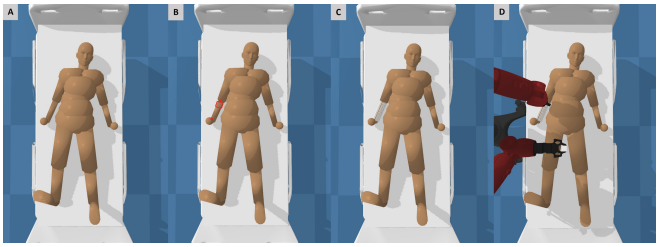

Fig. 2. Snapshots of the environment setup in simulation: (A) Settle the human model onto the bed in a randomly generated resting pose. (B) Generate a grasping point from a region around the supporting joint. (C) Uniformly distribute bathing markers (1cm apart) around the outer surface of the target limb. (D) Move the robot's left arm to grasp the target limb at the grasping point.

C. Observation Space

At every time step, we form an Agent-Specific Global State (AS) proposed by Yu et al. [11], which creates a global state for agent i by concatenating the agent-agnostic global state with agent-specific local observations for corresponding agent i. These are:

- Agent-agnostic global state:
  - the target limb position (6D) represented by the 3D position of two supporting joints
  - the tool pose consisting of tool position (3D) and tool quaternion (4D)
  - current joint angles (7D) of each arm
  - the local feature vector (9D) of the bathing marker point cloud M encoded by PointNet
- Agent-specific local observations:
  - the Cartesian linear velocity (3D) of the end effector for left arm agent
  - the relative mean position (3D) of the bathing markers to the agent for right arm agent
  - the tool force (1D) for right arm agent

D. Action Space

In order to maintain the flexibility to vary null space resolution schemes for simultaneously executing complex bimanual behaviors and avoid the jerky motion of robot arms for safety, we define an action space where the action is designed as a 7D vector representing changes in joint positions of the robot arm. The action joint space is clipped to be within [-1, 1] for the agents.

Given the output delta joint positions Δ_q, we then use a proportional-derivative (PD) controller with impedance gains to compute joint torques. The control law can be written as:

$$\tau = K_p \cdot \Delta_q - K_d \cdot \dot{q}$$

where K_p, K_d are proportional and derivative gains for joint positions. Note that in this research, we impose a fixed relationship between K_p and K_d where $K_d = \zeta\sqrt{K_p}$, ζ is a fixed damping ratio we set.

Variable gain PD control For such a contact-rich manipulation task, the policy should not only predict the joint positions as trajectory reference but also dynamically adapt the stiffness and damping properties of the controller to the interaction with the environment. We augment the action space with K_p to give the agent full control of the behavior. The control law can be written as:

$$\tau = K_p(\cdot) \cdot \Delta_q - K_d(\cdot) \cdot \dot{q}$$

The policy learns to modulate both the joint positions and gains to enable safe behaviors during contact interaction.

E. Reward Functions for Efficient Exploration

We carefully design a set of reward functions: $R_d(S)$, $R_b(S)$, $R_t(S)$, $R_v(s)$, $R_f(s)$ for the bimanual bathing manipulation task:

- $R_d(S)$: Reward for distance between the tool and the target positions. We divide the reward into two phases. In the first phase, the agent aims to move the tool and the target limb close to each other. We set

  $$R_d(S) = -\omega_d \cdot ||\mathcal{T} - L||_2$$

  which penalizes for the large distance between the tool $\mathcal{T}$ and the limb L. Once the tool is close enough to the limb within a distance of 10 cm, we compute the reward based on the average distance between the tool and bathing markers (N is the number of markers) as

  $$R_d(S) = -\omega_d \cdot (1/N \cdot \sum_{x \in M} ||\mathcal{T} - x||_2)$$

  , encouraging the tool nearer to the bathing area.
- $R_b(S)$: Reward for wiping off newly bathed bathing markers. According to official CNA instructions, caregivers are expected to perform smooth and continuous wiping motions. To encourage such actions, we design $R_b(S)$ as an adaptive weighted reward. At every time step t, if the robot successfully wipes off new bathing markers, we measure the similarity between the previous marker point cloud M^{t-1} and the current marker point cloud M^t using the Hausdorff distance:

  $$\begin{aligned} \mathrm{H}(M^{t-1}, M^t) = &\frac{1}{2} \max_{x \in M^{t-1}} |x - \mathrm{NN}(x, M^t)| \\ &+ \frac{1}{2} \max_{x' \in M^t} |x' - \mathrm{NN}(x', M^{t-1})| \end{aligned}$$

  where $\mathrm{NN}(x, M) = \mathrm{argmin}_{x' \in M} \|x - x'\|$ is the nearest neighbor function.
  If $\mathrm{H}(M^{t-1}, M^t) > 0.1$, which means the robot baths an area far from its last action, we compute $R_b(S)$ by adding a discount factor γ to the number of newly bathed markers N_b. $R_b(S)$ can be expressed as:

  $$R_b(S) = \omega_b \cdot \begin{cases} N_b, & \text{where } \mathrm{H}(M^{t-1}, M^t) < 0.1 \\ \gamma N_b, & \text{otherwise} \end{cases}$$

- $R_t(S)$: Reward for the trajectory error between the desired joint positions given by the policy at time step t, ${q_d}^t$, and the actual positions achieved at the next time step q^{t+1}. Following [9], we set

  $$R_t(S) = -\omega_t \cdot ||{q_d}^t - q^{t+1}||_2$$

  to encourage the policy to generate actions that can be effectively tracked.
- $R_v(S)$: Reward for the velocity to prevent the end-effectors from moving too fast to the person.
- $R_f(S)$: Reward for the contact force. We split $R_f(S)$ to two sub-rewards:
  - $R_{f,f}(S)$: prevent the right arm from applying force f_f away from the target bathing area. If $f_f > 0$, we penalize the force by calculating

    $$R_{f,f}(S) = -\omega_{f,f} \cdot f_f$$

  - $R_{f,c}(S)$: ensure that the contact force f_c exerted on the target bathing area remains within a desired safe force range of $[f_{c,min}, f_{c,max}]$. If f_c falls outside the range, we penalize the force based on the proximity of the contact force to the range as

    $$\begin{aligned} R_{f,c}(S) = &-\omega_{f,c} \cdot \\ &min(abs(f_c - f_{c,min}), abs(f_c - f_{c,max})) \end{aligned}$$

    Here we set this safe force range as $[5N, 10N]$ for all the tasks, as commonly adopted in relevant literature [4], [7].

F. Training Details

We train policies for bathing four target limbs with diverse shapes, weights, and biomechanical relationships (joint mechanics, biomechanical constraints, etc.), including the forearm, upperarm, shin, and thigh. Note that this research focus on learning to bath the right parts of the human from the right side of the bed, given the limitations in robot joint space, however, the procedures for the other side should be the same. At each time step, the robot receives observations from the system's state, executes an action output from the policy, and receives a reward calculated using appropriately designed reward functions. We utilize MAPPO with a variable impedance action space to train policies. Four policies are individually trained, each for one target limb. The actuators of both robot arms were fine-tuned: for left arm, the strengths are adjusted based on varying weights and biomechanical relationships of target limbs to provide optimal joint support; for right arm, strengths are universally capped at 10N to enhance safety during contact interactions. Additionally, we add random variations in human pose to better transfer policies to the real world. Each policy uses 36 concurrent simulation actors for a total of 30,000 simulation rollouts (trials), each consisting of 1000 time steps (20 seconds of simulation time at 10 time steps per second). We perform 50 SGD updates to the policy with a learning rate of 5e-5. All the policies are trained with the same hyperparameters on two desktops with Nvidia GeForce GTX 3060 Ti GPUs. Training a policy took ∼5 hours for each task.

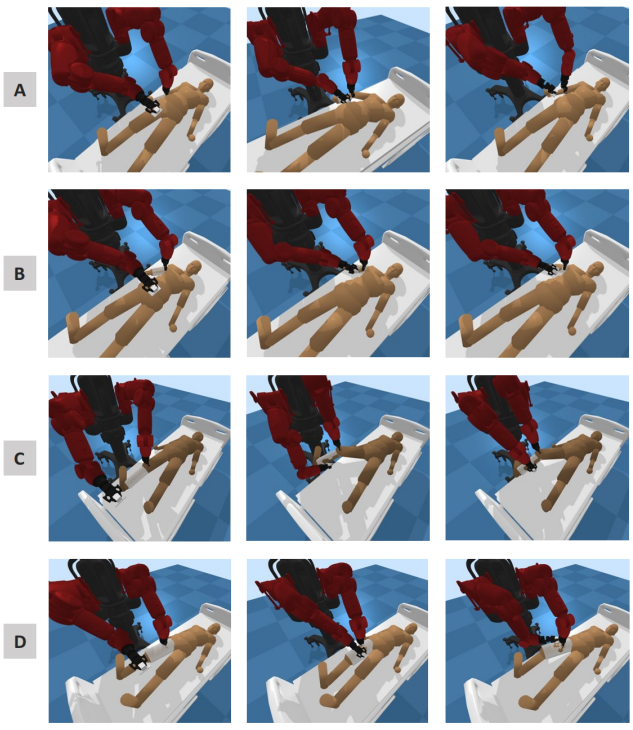

Fig. 3. Snapshots of the learned MAPPO models providing bathing assistance for four different body parts (A) Forearm (B) Upperarm (C) Shin (D) Thigh. The example trials are provided in the supplementary videos.

IV. Evaluation

In this section, we first summarize and analyze the results of the trained policies in simulation. Fig. 3 portrays a sequence of snapshots from bathing trials for each target limb. We conduct ablation studies with baselines to investigate the effectiveness of our method: MAPPO with a variable impedance action space and carefully-designed reward functions. We evaluate our policies with varying-sized human models to show their generalizability. We further compare and demonstrate our method using a bimanual Baxter humanoid robot in a real-world bed bathing task.

A. Policy Evaluation in Simulation

We evaluate and compare our method with the following baselines:

- *ours*: We train the policy using MAPPO with variable gains and a force regulation reward function $R_{f,c}(S)$ to control a bimanual robot.
- *w.o. Bimanual*: To demonstrate the effectiveness of using bimanual robot, we compare against the policy trained in an environment where only a single arm is utilized to hold the cleaning tool and move to bath.
- *w.o. MAPPO*: We treat both manipulators as one agent and train the policy using PPO [29] to show the effectiveness of MARL. The observation space for this policy is a fusion of agent-agnostic global state and agent-specific local observations.
- *w.o. VI*: To investigate the efficacy of variable impedance, we use fixed gains for the controller.

TABLE I

Average Coverage Percentage of the Policies over 100 Trials

Method	Forearm	Upperarm	Shin	Thigh
w.o.Bimanual	45.8%	26.5%	28.7%	15.1%
w.o.MAPPO	69.8%	53.7%	53.4%	27.0%
w.o.VI	79.3%	65.8%	65.9%	38.9%
w.o.FR	80.9%	66.5%	**74.5%**	**47.1%**
ours	**82.6%**	**76.2%**	68.6%	41.7%

TABLE II

Average Coverage Percentage of the Policies across Varied Body Sizes over 100 Trials

Method	Forearm	Upperarm	Shin	Thigh
w.o.MAPPO	60.9%	45.2%	42.6%	20.5%
w.o.VI	75.8%	61.3%	57.6%	27.7%
w.o.FR	79.4%	66.8%	**69.8%**	**42.9%**
ours	**81.3%**	**72.0%**	65.9%	40.8%

- *w.o. FR*: We train the policy without $R_{f,c}(S)$ to evaluate its effectiveness on regulating force during the task.

We compute and analyze the performance of the models from three metrics:

1) Ability of Bathing: To evaluate the ability of bathing, we calculate a coverage percentage of each trial (1000 time steps) with the ratio of the number of bathed markers N_B to the total number of bathing markers N: $P = \frac{N_B}{N}$. We compute the average coverage percentage over 100 randomized simulation trials.

From the results presented in Table I, for all the tasks, multi-agent bimanual policies demonstrate a significant 40% higher capability compared to unimanual policies, and a 10% improvement over single-agent policies in learning cooperative manipulation of both robot arms. Those policies effectively control the left arm to raise the target limb to positions where the right arm bath the obstructed areas. In contrast, policies with a single arm only bath the top surface of the target limb and the right arm controlled by *w.o. MAPPO* frequently collides with the human body as well as the left arm. Among the multi-agent bimanual policies, *ours* achieves the highest coverage percentage for bathing upper limbs, while *w.o. FR* shows the highest for bathing the lower limbs. One potential reason is that, when bathing the lower limbs, which are larger in size and involve narrower biomechanical joint limits, the robot must navigate within a smaller workspace to avoid collisions between its arms and the lower limbs. This constraint poses a challenge for *ours* to achieve greater coverage while ensuring that contact forces remain within the desired safe range. In contrast, *w.o. FR* does not need to consider such constraints, potentially contributing to its higher coverage percentage performance in bathing the lower limbs.

2) Ability of Safe Contact Manipulation: We evaluate the effectiveness of our variable impedance action space and the force regulation reward function by monitoring the contact forces applied to the human body. As shown in

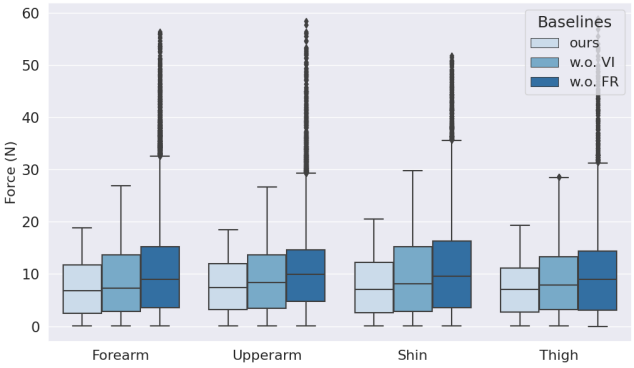

Fig. 4. Evaluation boxplot for the distribution of contact force across three baselines: *ours*, *w.o. VI*, and *w.o. FR* over 100 randomized trials. Notably, without force regulation, peak forces can reach up to 60N.

Fig. 5. The recorded contact forces applied on the human body during a representative trial (shin) across three baselines: *ours*, *w.o. VI*, and *w.o. FR*.

Fig. 4, *ours* consistently manages to maintain contact forces within the desired safe range most, aligning closely with the least dispersion in force data observed across all tasks. These observations indicate the robustness of our approach in adapting to interactions with the environment. Conversely, in the absence of either variable impedance action space or force regulation reward function, the data demonstrates a significantly wider range of contact forces, with the maximum force even reaching 60N, which can easily put people at risk in the real world. These results underscore the critical role of our variable impedance action space and force regulation reward in ensuring safe human-robot interaction.

We also further visualize the contact forces applied to the human body during a representative trial. From Fig. 5, *ours* displays its ability to maintain contact force within the safe force range. Compared to *w.o. FR*, there are longer no-contact periods in *ours* and *w.o. VI*, because the policies with force regulation reward take time to adjust the arm joint positions to avoid the robot from causing high contact force beyond the safe force range. We can also clearly see that *w.o. VI* generates forces with higher oscillations, while *ours* consistently keeps in a safer contact force range.

Fig. 6. Snapshots for bimanual bed bathing manipulation task execution in real world. Our experiment includes a bimanual Baxter Humanoid robot, a medical manikin, a hospital bed, and a cleaning tool. In real world, we only evaluate for forearm and upperarm due to the limitation of Baxter's payload. Example trials can be found in the supplementary video.

3) Ability of Generalization to Unseen Human Bodies: We further explore whether the learned policies have the ability to generalize to human models with varying body sizes unseen. Provided by Assistive Gym [28], [30], we generate an SMPL-X body mesh by employing 10 uniformly sampled body shape parameters $\beta \in U(0, 4)$ and modified the parameters of the capsulized human model to align with the mesh. The modified human model is within a range between 160cm and 185cm in height, exhibiting different body shapes. We evaluate the models with randomly generated human body sizes over 100 rollouts. Different body sizes have different sizes of bathing areas. For a fair comparison, we change the time steps of each trial based on the total number of bathing markers.

Similar to the results observed with trained human models, our findings indicate that policies with variable gains outperform those employing fixed gains or single-agent RL. The significant performance drops occur with single-agent policies. Since PPO relies on full observations during policy training, the policies are more susceptible to overfitting by the environment and harder to generalize than multi-agent methods. Furthermore, our analysis reveals that while the performance of policies without variable gains remains consistent for upper limb tasks, a noticeable drop of 8.3% and 11.2% is observed for lower limb tasks. This observation shows the robust adaptability and generalizability of variable impedance action space in navigating complex unseen contact-rich environments. These findings highlight the importance of incorporating variable impedance mechanisms with MARL to enhance the performance of our system.

B. Whole Pipeline Evaluation in Real World

To enable the robot to successfully perform bathing tasks in the real world, we first revise a few settings in simulation. In the early study, we discovered discrepancies in joint limitations between the manikin and the human model employed in simulation from [28]. For example, the manikin's elbow is constrained to only extension rotation while fixed for pronation and supination motions; also the flexion of the shoulder is limited up to 20 degrees. While this observation encourages consideration of developing strategies adaptable to varying human physical capabilities, especially among individuals with disabilities, it is not the primary focus of our current research but presents a potential avenue for future exploration [31]. Consequently, in this study, we directly

TABLE III

AVERAGE COVERAGE PERCENTAGE OF BASELINES ON THE REAL ROBOT OVER 10 TRIALS

Target	ours	ours-realworld	ours-realworld+force
Forearm	66.8%	57.7%	55.3%
Upperarm	61.5%	56.0%	54.7%

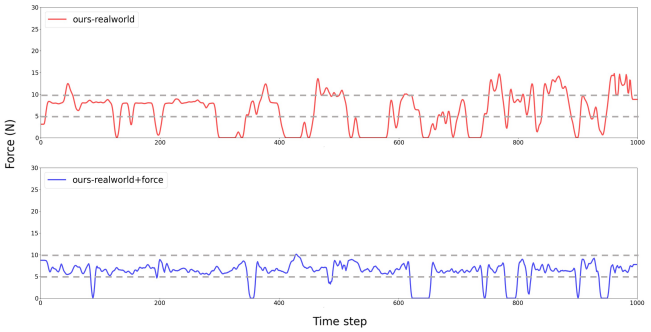

Fig. 7. The recorded measured tool force during a representative trial (forearm) across two baselines: *ours-realworld*, *ours-realworld+force* in real world.

align the joint limitations of the simulated human model with those of the manikin and retrain the model for real-world evaluations. We evaluate the average coverage percentage of the new model. We find that the percentage performance of the policies decreases due to the joint limitations.

We conduct the real-world experiments in a scenario where a manikin is positioned supine in a hospital bed while a Baxter humanoid robot holds a 3D printed cleaning holder attached to a bathing sponge similar to in the simulation (See Fig. 6). Prior to the experiment, we match the relative height between the robot and the hospital bed to the prescribed height. An Intel Realsense L515 camera is mounted on the head of the robot. AprilTags [32] are placed at the center of the bed to localize the robot with respect to the bed and on supporting limb joints to localize the initial grasping point. Due to occlusions caused by the robot arms, precise tracking of bathing markers is challenging in the real world. Therefore, we assume that the supporting joints maintain a consistent linear alignment with the grasping point by the robot's right arm. We position the markers based on the grasping point and predict whether the markers are bathed by their proximity to the tool position. Considering real-world constraints including robot payload and gripper opening limitations, we specifically investigate bathing the forearm and upperarm. Taking advantage of MAPPO's power on multiple agents, alternative manipulators could be explored in future work, such as UR10, which offers a higher payload capacity.

Force tracking controller Instead of using a force sensor, in this paper, we estimate the tool force transforming from the external force applied to the end-effector of the Baxter robot. We calculate this external end-effector force from the measured joint torques using gravity compensation and spring compensation torque, followed by a deadband filter for noise reduction. In the early study, we noticed that the noise in tool force affects the policy's capability to adapt actions in real-world environment. To address this sim-to-real gap and better maintain the contact forces within the safe force range, we introduce a proportional force tracking controller to regulate the actions output from the policy:

$$\tau_f = J^T \cdot (K_{f,p} \cdot (f_d - f_t)),$$
$$\text{where } f_d = \begin{cases} f_{\min}, & \text{if } f_t < f_{\min} \\ f_{\max}, & \text{elif } f_t > f_{\max} \end{cases}$$

where τ_f denotes the desired force regulation joint commands, $K_{f,p}$ is the proportional gain for force, $(f_{\min}, f_{\max})$ is the desired contact force range same as in simulation. J^T is the pseudoinverse of the robot's Jacobian matrix.

We combined the force tracking controller with the actions output from the policy. Here is the final equation:

$$\tau = K_p(\cdot) \cdot \Delta_q - K_d(\cdot) \cdot \dot{q} + J^T \cdot (K_{f,p} \cdot (f_d - f_t))$$

We evaluate the following baselines in real world:

- *ours-realworld*: We transfer *ours* to control a bimanual Baxter humanoid robot with a medical manikin in the real world.
- *ours-realworld+force*: We further employ a proportional force tracking controller to ensure safe contact force for actions learned from *ours*.

For each target limb, we evaluate the corresponding policies over ten trials. Table III summarizes the average coverage percentage in both (sim-to-real) simulation and real-world trials. Both *ours-realworld* and *ours-realworld+force* present a relatively similar coverage percentage with *ours*, but still decrease for about 10%. One reason for this sim-to-real gap is that due to the joint limitation of the manikin, the manikin itself applies forces to restrict the robot from lifting the supporting joint to certain positions. Also, in this task, we predict human joint positions and bathing markers by a fixed transformation due to partial occlusions caused by the robot arms. This assumption causes the robot to intermittently deviate from the correct bathing areas. In the future, this could be solved by incorporating multiple cameras to continually capture the target limb from diverse views [33]–[35]. Fig. 7 visualizes the measured tool forces during a representative trial. *ours-realworld+force* proves that an additional force tracking controller can improve the robot's ability to maintain contact force within the safe force range. Visual demonstrations of the results can be found in the supplementary video. Overall, the average coverage percentage and the contact forces in the real world are consistent with the simulation results, indicating promising prospects for transferring simulation-trained policies to real-world robots for bimanual bed bathing manipulation.

V. CONCLUSION

This research introduces a novel formulation for robot-assisted bed bathing where a dual-arm robot is employed to lift and bath the target limb of a bed-bound person. We employ MAPPO with variable impedance action space and

carefully-designed reward functions for such a collaborative contact-rich manipulation task. We compare the approaches for four different target limbs. Our empirical findings demonstrate that these simulation-trained models enable the robot to lift and safely bath around the limb surface and exhibit robust generalization capabilities across unseen body sizes. Furthermore, we provide evidence of the successful transfer of our policies to a physical dual-arm robot, indicating the potential viability of our approach for practical applications in real-world bathing care.

REFERENCES

[1] "Disability, world health organization," https://www.who.int/news-room/fact-sheets/detail/disability-and-health, 2023.

[2] C.-H. King, T. L. Chen, A. Jain, and C. C. Kemp, "Towards an assistive robot that autonomously performs bed baths for patient hygiene," in *2010 IEEE/RSJ International Conference on Intelligent Robots and Systems*. IEEE, 2010, pp. 319–324.

[3] A. C. Dometios, Y. Zhou, X. S. Papageorgiou, C. S. Tzafestas, and T. Asfour, "Vision-based online adaptation of motion primitives to dynamic surfaces: application to an interactive robotic wiping task," *IEEE Robotics and Automation Letters*, vol. 3, no. 3, pp. 1410–1417, 2018.

[4] Z. Erickson, H. M. Clever, V. Gangaram, G. Turk, C. K. Liu, and C. C. Kemp, "Multidimensional capacitive sensing for robot-assisted dressing and bathing," in *2019 IEEE 16th International Conference on Rehabilitation Robotics (ICORR)*. IEEE, 2019, pp. 224–231.

[5] I. Huang, D. Chow, and R. Bajcsy, "Soft tactile contour following for robot-assisted wiping and bathing," in *2022 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2022, pp. 7797–7802.

[6] R. Madan, S. Valdez, D. Kim, S. Fang, L. Zhong, D. T. Virtue, and T. Bhattacharjee, "Rabbit: A robot-assisted bed bathing system with multimodal perception and integrated compliance," in *Proceedings of the 2024 ACM/IEEE International Conference on Human-Robot Interaction*, 2024, pp. 472–481.

[7] Y. Gu and Y. Demiris, "Vttb: A visuo-tactile learning approach for robot-assisted bed bathing," *IEEE Robotics and Automation Letters*, 2024.

[8] R. Martín-Martín, M. A. Lee, R. Gardner, S. Savarese, J. Bohg, and A. Garg, "Variable impedance control in end-effector space: An action space for reinforcement learning in contact-rich tasks," in *2019 IEEE/RSJ international conference on intelligent robots and systems (IROS)*. IEEE, 2019, pp. 1010–1017.

[9] M. Bogdanovic, M. Khadiv, and L. Righetti, "Learning variable impedance control for contact sensitive tasks," *IEEE Robotics and Automation Letters*, vol. 5, no. 4, pp. 6129–6136, 2020.

[10] A. Allshire, R. Martín-Martín, C. Lin, S. Manuel, S. Savarese, and A. Garg, "Laser: Learning a latent action space for efficient reinforcement learning," in *2021 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2021, pp. 6650–6656.

[11] C. Yu, A. Velu, E. Vinitsky, J. Gao, Y. Wang, A. Bayen, and Y. Wu, "The surprising effectiveness of ppo in cooperative multi-agent games," *Advances in Neural Information Processing Systems*, vol. 35, pp. 24 611–24 624, 2022.

[12] A. G. Perry, P. A. Potter, and W. Ostendorf, *Clinical nursing skills and techniques*. Elsevier Health Sciences, 2013.

[13] F. Liu, V. Patil, Z. Erickson, and Z. Temel, "Characterization of a meso-scale wearable robot for bathing assistance," in *2022 IEEE International Conference on Robotics and Biomimetics (ROBIO)*. IEEE, 2022, pp. 2146–2152.

[14] M. C. Gemici and A. Saxena, "Learning haptic representation for manipulating deformable food objects," in *2014 IEEE/RSJ International Conference on Intelligent Robots and Systems*. IEEE, 2014, pp. 638–645.

[15] K. Zhang, M. Sharma, M. Veloso, and O. Kroemer, "Leveraging multimodal haptic sensory data for robust cutting," in *2019 IEEE-RAS 19th International Conference on Humanoid Robots (Humanoids)*. IEEE, 2019, pp. 409–416.

[16] I. Garcia-Camacho, M. Lippi, M. C. Welle, H. Yin, R. Antonova, A. Varava, J. Borras, C. Torras, A. Marino, G. Alenya *et al.*, "Benchmarking bimanual cloth manipulation," *IEEE Robotics and Automation Letters*, vol. 5, no. 2, pp. 1111–1118, 2020.

[17] Y. Avigal, L. Berscheid, T. Asfour, T. Kröger, and K. Goldberg, "Speedfolding: Learning efficient bimanual folding of garments," in *2022 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2022, pp. 1–8.

[18] S. Kotsovolis and Y. Demiris, "Bi-manual manipulation of multi-component garments towards robot-assisted dressing," in *2023 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2023, pp. 9865–9871.

[19] A. M. Kabir, K. N. Kaipa, J. Marvel, and S. K. Gupta, "Automated planning for robotic cleaning using multiple setups and oscillatory tool motions," *IEEE Transactions on Automation Science and Engineering*, vol. 14, no. 3, pp. 1364–1377, 2017.

[20] V. Girbés-Juan, V. Schettino, L. Gracia, J. E. Solanes, Y. Demiris, and J. Tornero, "Combining haptics and inertial motion capture to enhance remote control of a dual-arm robot," *Journal on Multimodal User Interfaces*, pp. 1–20, 2022.

[21] J. Grannen, Y. Wu, S. Belkhale, and D. Sadigh, "Learning bimanual scooping policies for food acquisition," in *Conference on Robot Learning*. PMLR, 2023, pp. 1510–1519.

[22] Y. Zhu, J. Wong, A. Mandlekar, R. Martín-Martín, A. Joshi, S. Nasiriany, and Y. Zhu, "robosuite: A modular simulation framework and benchmark for robot learning," *arXiv preprint arXiv:2009.12293*, 2020.

[23] Y. Chen, T. Wu, S. Wang, X. Feng, J. Jiang, Z. Lu, S. McAleer, H. Dong, S.-C. Zhu, and Y. Yang, "Towards human-level bimanual dexterous manipulation with reinforcement learning," *Advances in Neural Information Processing Systems*, vol. 35, pp. 5150–5163, 2022.

[24] C. C. Beltran-Hernandez, D. Petit, I. G. Ramirez-Alpizar, and K. Harada, "Variable compliance control for robotic peg-in-hole assembly: A deep-reinforcement-learning approach," *Applied Sciences*, vol. 10, no. 19, p. 6923, 2020.

[25] M. A. Lee, Y. Zhu, P. Zachares, M. Tan, K. Srinivasan, S. Savarese, L. Fei-Fei, A. Garg, and J. Bohg, "Making sense of vision and touch: Learning multimodal representations for contact-rich tasks," *IEEE Transactions on Robotics*, vol. 36, no. 3, pp. 582–596, 2020.

[26] F. Tian, C. Lv, Z. Li, and G. Liu, "Modeling and control of robotic automatic polishing for curved surfaces," *CIRP Journal of Manufacturing Science and Technology*, vol. 14, pp. 55–64, 2016.

[27] F. Chen, H. Zhao, D. Li, L. Chen, C. Tan, and H. Ding, "Contact force control and vibration suppression in robotic polishing with a smart end effector," *Robotics and Computer-Integrated Manufacturing*, vol. 57, pp. 391–403, 2019.

[28] Z. Erickson, V. Gangaram, A. Kapusta, C. K. Liu, and C. C. Kemp, "Assistive gym: A physics simulation framework for assistive robotics," in *2020 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2020, pp. 10 169–10 176.

[29] J. Schulman, F. Wolski, P. Dhariwal, A. Radford, and O. Klimov, "Proximal policy optimization algorithms," *arXiv preprint arXiv:1707.06347*, 2017.

[30] K. Puthuveetil, C. C. Kemp, and Z. Erickson, "Bodies uncovered: Learning to manipulate real blankets around people via physics simulations," *IEEE Robotics and Automation Letters*, vol. 7, no. 2, pp. 1984–1991, 2022.

[31] A. Clegg, Z. Erickson, P. Grady, G. Turk, C. C. Kemp, and C. K. Liu, "Learning to collaborate from simulation for robot-assisted dressing," *IEEE Robotics and Automation Letters*, vol. 5, no. 2, pp. 2746–2753, 2020.

[32] J. Wang and E. Olson, "Apriltag 2: Efficient and robust fiducial detection," in *2016 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2016, pp. 4193–4198.

[33] M. Wu, Y. Wang, Q. Hu, and J. Yu, "Multi-view neural human rendering," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2020, pp. 1682–1691.

[34] M. Hofmann and D. M. Gavrila, "Multi-view 3d human pose estimation in complex environment," *International journal of computer vision*, vol. 96, pp. 103–124, 2012.

[35] S. E. Ovur and Y. Demiris, "Naturalistic robot-to-human bimanual handover in complex environments through multi-sensor fusion," *IEEE Transactions on Automation Science and Engineering*, 2023.