ICRA2026 Late Breaking Results Poster presented at
2026 IEEE International Conference on Robotics and Automation (ICRA 2026)
June 1-5, 2026. Vienna, Austria

Diffusion Policy for Robot-Assisted Dressing with Moving Human Arms

Haoxiang Sun and David Navarro-Alarcon

***Abstract*— Robot-assisted dressing remains challenging due to the close physical human–robot interaction and the highly deformable nature of garments. This work presents a purely vision-based approach that transfers human-mastered dressing skills to robots while accommodating dynamic human arm movements. The proposed method adopts a hierarchical structure. At the high level, a diffusion model serves as the policy to learn action distributions conditioned on point cloud observations. During execution, a diffused scalar field is constructed to infer an object-centric axial distribution of the human arm from cluttered points. Local point cloud registration across consecutive frames further captures arm motion, enabling real-time adaptation of robot actions to user dynamics. Comprehensive evaluations have been conducted in both simulation and real-world dressing scenarios using a UR10e robot with human participants of diverse genders and body types.**

I. Introduction

Dressing is an essential activity of daily living, and assistance with dressing is typically mastered by professional caregivers through extensive training. However, this task remains challenging for robots. A primary difficulty stems from the lack of generalizable knowledge of deformable objects, combined with thei infinite degrees of freedom, making deformation behavior difficult to predict accurately. Moreover, during the dressing process, interactions between garments and the human body introduce random visual occlusions and physical entanglements, which pose significant challenges for robotic manipulation.

Early studies focused on user modelling to enable personalized dressing assistance [1], [2]. Subsequently, many works explored incorporating visual observations and force feedback as inputs to reinforcement learning or generative modeling frameworks, aiming to learn garment dynamics [3] or directly infer action policies [4], [5]. However, these approaches invariably rely on a common assumption that the human arm remains static during dressing. While this assumption simplifies the problem of inferring topological information under occlusions, it also limits user comfort and natural interaction. More recent efforts have begun to address dynamic dressing scenarios [6]. Nevertheless, existing methods still lack robust solutions in terms of real-time execution and resilience to human motion variability.

In this work, we study an upper-body dressing task in which a single robot manipulates a garment sleeve onto the user's forearm and subsequently advances it along the arm toward the shoulder. Unlike prior approaches, the proposed method does not require the user to remain static during dressing. Furthermore, we do not assume the garment sleeve to be pre-positioned at the human wrist. In particular, we propose an imitation learning-based visuomotor policy that leverages point cloud observations to learn a generalizable action distribution, and incorporates a sampling and registration method to adapt to dynamic dressing scenarios with human motion.

Haoxiang Sun and David Navarro-Alarcon are with the Department of Mechanical Engineering, The Hong Kong Polytechnic University, Hung Hom, KLN, Hong Kong. hao-xiang.sun@connect.polyu.hk, dnavar@polyu.edu.hk

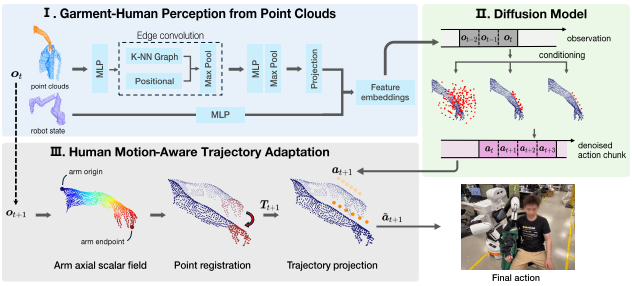

Fig. 1. Architectural diagram of the proposed robot-assisted dressing framework.

II. Methodology

As shown in Fig. 1, we propose a hierarchical framework for dynamic robot-assisted dressing. A diffusion model is trained from expert demonstrations to learn the motion policy in variations in human poses, body shapes, and garment types. At a lower level, a local object-centric point cloud sample and registration algorithm are used to adapt the trajectory to account for ongoing arm movement.

To handle the partial observability caused by occlusions between the garment and the human arm, we formulate the problem as a partially observable decision-making process (PODMP). The policy operates directly on point cloud observations and learns to generate multimodal action distributions that are robust to visual ambiguity and interaction complexity. The diffusion model generates action chunks by following the iterative denoising process:

$$a_{t-1} = \sqrt{\bar{\alpha}_{t-1}}\,\hat{a}_\theta(o_t, a_t, t) + \sqrt{1-\bar{\alpha}_{t-1}}\,\varepsilon_\theta(o_t, a_t, t) \quad (1)$$

where $\hat{a} = \frac{a_t - \sqrt{1-\bar{\alpha}_t}\,\varepsilon_\theta(o_t,a_t,t)}{\sqrt{\bar{\alpha}_t}}$. ε_θ is the noise prediction network, with denoising diffusion implicit models (DDIM) [7] used as the noise scheduler, and $\bar{\alpha}_t = \prod_{s=1}^{t} \alpha_s$ is the cumulative product of the predefined variance schedule.

To enable responsive adaptation during dressing, we construct an object-centric axial representation of the human arm from partial observations by solving a diffused scalar field:

$$u_{\tau-1} = (M - \tau C)^{-1} M u_0 \quad (2)$$

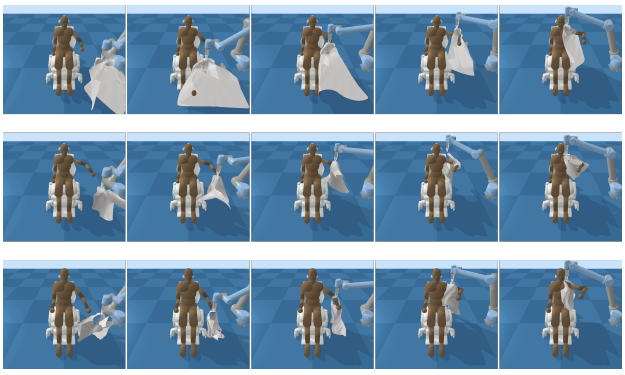

Fig. 2. The proposed method enables the UR10e robot to successfully perform dressing across 3 garment types, 4 body types, and 10 movement scenarios.

where u denotes a scalar field defined along the arm axis, C is the Laplacian matrix, M is the mass matrix and τ is the diffusion time parameter. Based on this representation, we estimate local geometric transformations across consecutive frames to align the planned motion with the current arm configuration. This allows the system to react to non-static, non-Markovian arm movements without requiring full reconstruction of the arm. The adapted actions are then projected and executed by the manipulator.

III. Results

We evaluate the proposed method and its baselines in simulation. The dressing scenarios include three types of garments with different sleeve configurations: long-sleeve, short-sleeve, and sleeveless. We consider four human body models covering variations in gender and body shape. To evaluate performance under varying human motions, we consider ten arm motion patterns, including eight planar displacements with different directions and two rotational motions (clockwise and counterclockwise). We further assess the method under three motion speed profiles: static, constant-speed ($v = 1.0$), and fast motion ($v = 2.0$).

We compare our full method against several baselines and ablations:

- *Ours w/o TA*: ablating the trajectory adaptation module.
- *PointNet-Diffusion (DP3)*: replacing the proposed encoder to a PointNet encoder.
- *Image based Diffusion Policy (DP-image)*: using image-based visual inputs.
- *Diff-MPC*: a diffusion model with feature encoding (f.e.) of end-effector actions to capture garment dynamics, combined with an MPC controller that outputs optimal actions.
- *BC-LSTM*: recurrent imitation learning policy.
- *Implicit Behavioral Cloning (IBC)*: behavioral cloning policy.

Performance is evaluated using two metrics: sleeve insertion success, which measures whether the sleeve opening is successfully inserted onto the arm, and dressing ratio, defined as the ratio of the dressed arm length to the total arm length. Results are reported across all combinations of arm motions, motion speeds, garment types, and human body models. For each configuration, 70 trials are conducted with randomized initial conditions, and average performance is reported.

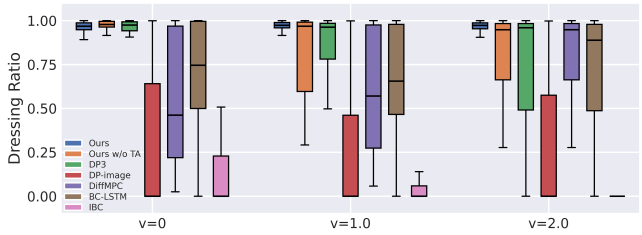

Fig. 3. Average dressing ratio in simulation, evaluated over 70 randomized trials per baseline.

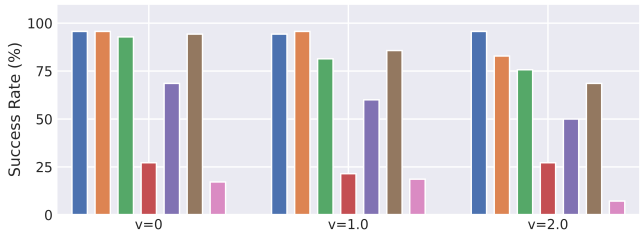

Fig. 4. Average success rate for fully insert the openings of garments in simulation, evaluated over 70 randomized trials per baseline.

IV. Conclusion

We presented a method for robot-assisted dressing with moving human arms. The proposed approach achieves high performance and demonstrates generalization across different body types and arm motion patterns, including cases beyond those covered in expert demonstrations. By combining learned motion generation with online adaptation, the system is able to maintain alignment with the evolving human arm configuration during dressing. Future work will investigate the integration of force-based modalities to further improve physical interaction and user comfort.

References

[1] Y. Gao, H. J. Chang, and Y. Demiris, "User modelling for personalised dressing assistance by humanoid robots," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst*, 2015, pp. 1840–1845.

[2] F. Zhang, A. Cully, and Y. Demiris, "Probabilistic Real-Time User Posture Tracking for Personalized Robot-Assisted Dressing," *IEEE Trans. Robot.*, vol. 35, no. 4, pp. 873–888.

[3] S. Kotsovolis and Y. Demiris, "Garment diffusion models for robot-assisted dressing," *IEEE Robot. Automat. Lett.*, vol. 10, no. 2, pp. 1217–1224, 2025.

[4] Y. Wang, Z. Sun, Z. Erickson, and D. Held, "One Policy to Dress Them All: Learning to Dress People with Diverse Poses and Garments," in *Proc. Robot.: Sci. Syst*, 2023.

[5] Z. Sun, Y. Wang, D. Held, and Z. Erickson, "Force-constrained visual policy: Safe robot-assisted dressing via multi-modal sensing," *IEEE Robot. Automat. Lett.*, vol. 9, no. 5, pp. 4178–4185, 2024.

[6] A. Y. Hao, Y. Wang, N. S. Ravie, B. Hegde, D. Held, and Z. Erickson, "Force-modulated visual policy for robot-assisted dressing with arm motions," in *Proc. Conf. Robot Learn*, 2025, pp. 5483–5505.

[7] J. Song, C. Meng, and S. Ermon, "Denoising diffusion implicit models," *arXiv:2010.02502*, 2020.