# Reinforcement Learning of Clothing Assistance with a Dual-arm Robot

Tomoya Tamei*, Takamitsu Matsubara*, Akshara Rai†, Tomohiro Shibata*
*Graduate School of Information Science,Nara Institute of Science and Technology,
8916-5 Takayama, Ikoma, Nara 630-0192, Japan
†Department of Electrical Engineering, Indian Institute of Technology Kanpur, India
Email: {tomo-tam, takam-m, tom}@is.naist.jp, akshara@iitk.ac.in

***Abstract*—This study aims at robotic clothing assistance as it is yet an open field for robotics despite it is one of the basic and important assistance activities in daily life of elderly as well as disabled people. The clothing assistance is a challenging problem since robots must interact with non-rigid clothes generally represented in a high-dimensional space, and with the assisted person whose posture can vary during the assistance. Thus, the robot is required to manage two difficulties to perform the task of the clothing assistance: 1) handling of non-rigid materials and 2) adaptation of the assisting movements to the assisted person's posture. To overcome these difficulties, we propose to use reinforcement learning with the cloth's state which is low-dimensionally represented in topology coordinates, and with the reward defined in the low-dimensional coordinates. With our developed experimental system, for T-shirt clothing assistance, including an anthropomorphic dual-arm robot and a soft mannequin, we demonstrate the robot quickly learns a suitable arm motion for putting the mannequin's head into a T-shirt.**

## I. Introduction

Robotic assistance is a growing social need due to demographic trends such as the aging population combined with a diminishing number of children. This study aims at robotic clothing assistance as it is yet an open field for robotics despite it is one of the basic and important assistance activities in daily life of elderly as well as disabled people. In this study, the scenario of the clothing assistance is as follows. First, we develop a dual-arm robot system to manipulate a cloth. Then, a human a teaches a desired trajectory of the arms by directly moving the arms by his/her hands. Since there is a possibility for an assisted person to change his/her body posture, the robot finally learns to achieve the task through reinforcement learning.

Using low-dimensional representations of state, policy and reward is necessary to achieve fast learning. Namely we need to have low-dimensional representations for the cloth as well as the arm controller. This study employs the learning scheme that Shinohara et al. have proposed [1]. Although the most popular approach to detect and grasp non-rigid materials is to use image sensors, e.g., [2], [3], it is not suitable for this study, since the non-rigid materials are generally represented in a high dimensional space, e.g., [4], which prevents the completion of reinforcement learning within reasonable time. Shinohara et al. considered that the details of the clothe, e.g., wrinkles, are not very important to achieve motor tasks, and have proposed to use the topological coordinates [5] for the state and reward representation of the cloth. They have also proposed to learn only a few parameters of the policy represented by a minimum jerk trajectory.

In this study, we developed an experimental setting with an anthropomorphic dual-arm robot and a T-shirt for a human, and designed a reward function suitable for the clothing assistance task with the topological relationship between the T-shirt and the assisted person. A result of a preliminary experiment with a soft mannequin instead of a human, we demonstrate the robot quickly learns a suitable arm motion for putting the mannequin's head into a T-shirt.

The rest of this paper is organized as follows. Section II describes the learning system we developed for learning the T-shirt clothing assistance task. Section III, the experimental setting and obtained results are shown. Finally, section IV concludes with some future directions.

## II. Learning System for the Clothing Assistance Task

In this section, we present a novel learning system for an anthropomorphic dual-arm robot to perform the clothing assistance task using reinforcement learning. The learning scheme is depicted in Fig. 1. The purpose of reinforcement learning is to optimize the policy parameters so that the expected reward becomes maximal. To initialize the policy parameters, we used a direct teaching approach under the gravity compensation control of the robot. Then, the taught trajectory was converted to the initial policy represented by via-points (see II-D in detail). To approximate the whole process as a Markov decision process (MDP), the state of relationship between the cloth and the assisted-person should be observed as much as possible. In this study, the state was represented in the topological coordinates (see II-B and II-C in detail) such that the topological relationship between the T-shirt and the assisted-person can be described by low-dimensional variables. The computer agent observes the topology coordinates and modifies the joint angle trajectories of the robot by optimizing via-points of the joint trajectories. The details of the learning system will be described below.

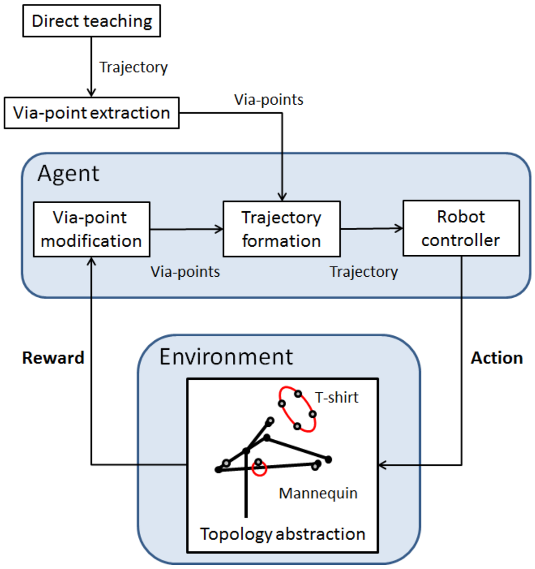


Fig. 1. Learning scheme

### A. Experimental Setting

In this study, we used a soft mannequin instead of a human, and a T-shirt as a cloth. The task was to put the mannequin's head into the T-shirt by pulling the T-shirt. Fig. 2 shows the experimental setting including a motion capturing system (MAC3D System, Motion Analysis Corp.), a dual-arm robot (WAM, Barrett), and a mannequin (flexible mannequin, Displan) with a T-shirt used in this paper.

We designed the detailed setting of the T-shirt clothing assistance task as follows. The mannequin was set in front of the robot. The initial state of the mannequin has both arms inside the sleeves of the T-shirt, while the robot holds the hem. The shirt was manually attached to the robot with the gripper which is at the end of the robotic arm (see Fig. 4). For each experiment, the mannequin was set to the same configuration by hand such that at the initial neck inclination and shoulder elevation were set to 35 degrees and 100 degrees, respectively (Fig. 3). Also, the T-shirt was set to roughly the same configuration by hand such that at the initial position both arms of the mannequin are at fixed positions inside the hem as shown in Fig. 2.

To initialize the control policy (see II-D in detail), we used a direct teaching approach in which a human held both arms of the robot and directly demonstrated how to move to put the mannequin's head into the the T-shirt. Then, in a trial-and-error manner, the robot repeated a *trial* in which a sequential movement generated by the control policy, with the data obtained during a certain number of trials, called an *episode*, used to improve the control policy. The termination condition of the trial is the end of executing a trajectory, or an abort of the robot due to a torque limitation.

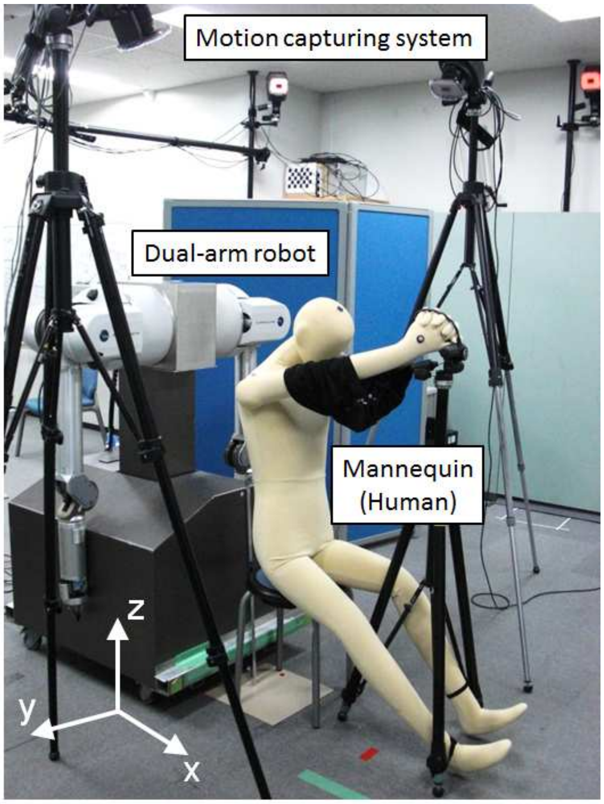


Fig. 2. Experimental setting

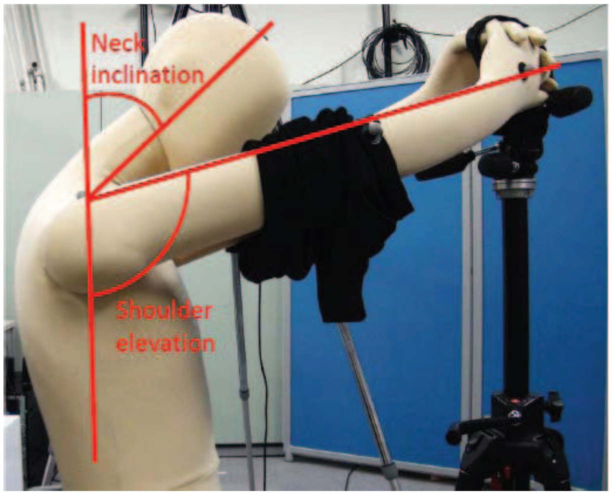


Fig. 3. Posture definition of the mannequin

### B. Topological Coordinates for learning motor skills of the robot

A study of Shinohara et al. [1] shares similar research aspects with our study. They considered that the relationship between the configuration of the robot and the nonrigid material is very important to achieve motor tasks, and have proposed to use the topological coordinates [5] for the state and reward representation of the cloth. For designing a

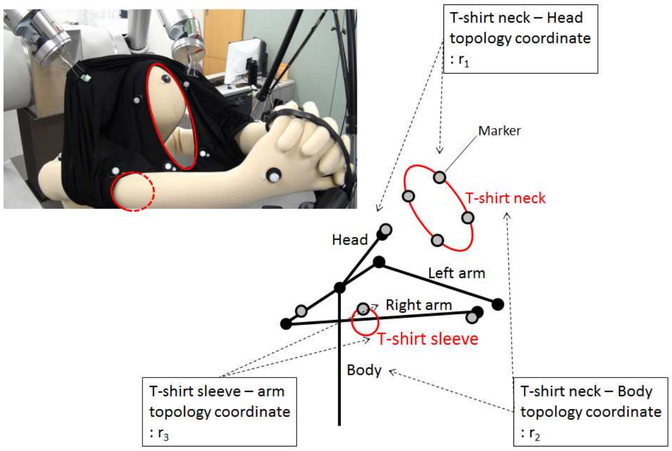


Fig. 4. Illustration of the setting of topology coordinates for the T-shirt clothing assistance task by an dual-arm robot. The topology coordinates are defined between the T-shirt neck and mannequin's head, T-shirt neck and mannequin's body, and between the T-shirt sleeve and mannequin's right arm.

reward function suitable for the clothing assistance task, we utilize the topological relationship between the configuration of the mannequin and a clothe with topology coordinates. We used topological relationships between the T-shirt neck and mannequin's head, between the T-shirt neck and mannequin's body, and between the T-shirt sleeve and mannequin's right arm as an approximation to the state variables, as depicted in Fig. 4. The T-shirt neck and sleeves, mannequin body and arm were detected by a motion capturing system.

To obtain the topological relationship between the T-shirt and mannequin, we first divide each link of the arm and the sleeve into a number of small segments. Then those small segments can be used to obtain topology coordinates which compactly represent the topological relationships between the arm and the sleeve. The writhe $w$ counts how much the two curves are twisting around each other, which is an approximation of the Gauss Linking Integral (GLI) [6] as

$$GLI(\gamma_1, \gamma_2) = \frac{1}{4\pi} \int_{\gamma_1} \int_{\gamma_2} \frac{d\gamma_1 \times d\gamma_2 \cdot (\gamma_1 - \gamma_2)}{||\gamma_1 - \gamma_2||^3}, \quad (1)$$

where, $\gamma_1$ and $\gamma_2$ are assumed to be the curved lines of the robot and sleeve. The center $\mathbf{c}$, composed of two scalars, explains the center of location and the density $d$ presents the difference of the quantity of twisting of the robot and the sleeve.

The topology coordinates are defined mathematically as serial chains of segments. It is assumed that two chains $S_1$ and $S_2$ which are composed $n_1$ and $n_2$ line segments, connected by revolute, universal or gimbal joints. Thus, the total writhe is computed as summation of the writhes by each pair of segments:

$$w = GLI(S_1, S_2) = \sum_{i=1}^{n_1} \sum_{j=1}^{n_2} T_{i,j}, \quad (2)$$

where $T_{i,j}$ is the writhe between segment $i$ on $S_1$ and $j$ on $S_2$. An analytical solution for computing $T_{i,j}$ is presented in [5]. A $n_1 \times n_2$ matrix $\mathbf{T}$ which has $(i, j)$-th element of $T_{i,j}$ is called writhe matrix. The center is defined as center of twisted chains in topology space by the following equation:

$$\begin{aligned} \mathbf{c} &= (x_g, y_g) \\ &= \left( \frac{\sum_{i=1}^{n_1} \sum_{j=1}^{n_2} i \cdot T_{i,j}}{w} - \frac{n_2}{2}, \frac{\sum_{i=1}^{n_1} \sum_{j=1}^{n_2} i \cdot T_{i,j}}{w} - \frac{n_1}{2} \right). \end{aligned} \quad (3)$$

The definition of the density is led by the orientation of the principal axis of writhe matrix. The density is computed as the angle between the principal axis and the diagonal line.

### *C. Reward Function for clothing assistance with topology coordinates*

Motor skills required for achieving this task are 1) to put the mannequin's head into the T-shirt neck, and 2) to put the mannequin's shoulder into the T-shirt sleeve. The achievement of each skill can be represented as the state by using the topology coordinates. The topological relationship between the T-shirt neck and the mannequin's head was defined using the center of topology coordinates as $s_1 = [c_2^{head}]$. The topological relationship between the T-shirt neck and the mannequin's body as $s_2 = [w^{body}]$ using writhe. The topological relationship between the T-shirt sleeve and the mannequin's right arm as $s_3 = [c_2^{arm}]$ using center. With the above definitions of the state, we defined the reward function as:

$$r_i = -||s_i^{\text{target}} - s_i||^2 \qquad (i = 1, 2, 3), \quad (4)$$

where $s_i^{\text{target}}$ is the target state of neck-head, neck-body and the sleeve-arm in terms of the topology coordinates. $s_i$ is the topology coordinates obtained at the end of a trial. The total reward was calculated as:

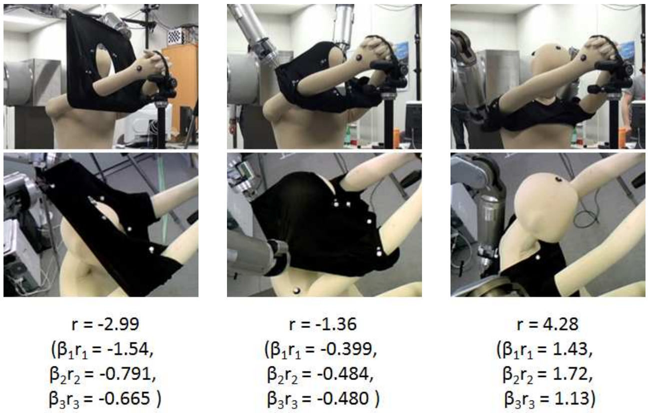
$$r(\mathbf{s}) = \sum_{i=1}^{3} \frac{r_i - \mu_i}{\sigma_i}, \quad (5)$$

where $\mu_i$ and $\sigma_i$ are average and variance value of case examples of failure and success.

The T-shirt neck and sleeves, mannequin body and arm were detected by a motion capturing system. A total of 8 markers for the motion capturing were attached to the T-shirt and mannequin as shown in Fig. 2 and Fig. 4 T-shirt neck and sleeve were divided into 80 and 30 segments respectively for computing the coordinates (see II-B). In this study, the target states were also determined by the direct teaching approach as [12.3] for the neck-head, [0.725] for the neck-body and [-46.4] for the arm-sleeve topological relationships. Fig. 5 shows the reward values obtained in example failure and success cases.

### *D. Control Policy*

The action of the control policy was defined as updated joint angles $q^d_{0:\Delta t:T} = \{q^d_0, q^d_{\Delta t}, \cdots, q^d_T\}$ , where $\Delta t$ is a fixed small-time duration. The parameter for the $n$-th joint $\boldsymbol{\theta}^n$ in the policy corresponds to a set of via-points (intermediate points) of the trajectory as $\boldsymbol{\theta}^n = [q_1^{\text{via}}, \cdots, q_I^{\text{via}}]^T$. The control policy is defined as a deterministic function $L^n(\boldsymbol{\theta}^n)$ with



(a) Failure case 1 (b) Failure case 2 (c) Success case

Fig. 5. Example reward values with different topology coordinates

a minimum jerk criterion [7]. For example, the generated trajectory by the control policy with two sequential via-points $\hat{\boldsymbol{\theta}}^n = [q_i^{\text{via}}, q_{i+1}^{\text{via}}]^T$ which are corresponding to the point at time $\tau_i$ and $\tau_{i+1}$ are written as follows:

$$\begin{aligned} q^d_{\tau_i:\Delta t:\tau_{i+1}} &= L^n(\hat{\boldsymbol{\theta}}^n) \\ &\leftarrow \operatorname*{argmin}_{q_{T_i:\Delta t:T_{i+1}}} \int_{T_i}^{T_{i+1}} \left(\frac{d\ddot{q}(t)}{dt}\right)^2 dt \\ \text{s.t.} \quad & q^d_{\tau_i} = q_i^{\text{via}},\ q^d_{\tau_{i+1}} = q_{i+1}^{\text{via}} \end{aligned} \tag{6}$$

and the solution of Eq. (6) can be analytically obtained as a 5-th order time polynomial function [7]. The control policy and policy parameters are independently set for all joints.

### E. Policy Improvement

We employed a finite difference policy gradient method [7], [8] for reinforcement learning for policy improvement. The finite difference policy gradient is a policy-gradient method which can be easily implemented. The advantages of the policy gradient method is that the policy representation can be chosen so that it is meaningful for the task and can incorporate domain knowledge. This often requires fewer parameters in the learning process than in value-function based methods. In addition, the policy gradient method is a model-free approach. In the policy gradient method, the gradient information of the expected reward $\eta(\boldsymbol{\theta})$ must be estimated with respect to the policy parameter $\boldsymbol{\theta}$ as $\frac{\partial\eta(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}$. For the gradient estimation, the changed policies are made from the current policy parameter $\theta$ by small perturbations $\pm\Delta\theta$ and then executed to obtain expected returns $r(\theta \pm \Delta\theta)$. This allows to estimate the policy gradient of the $i$-th element of the policy parameter as

$$\frac{\partial\eta(\theta_i)}{\partial\theta_i} \approx \frac{r(\theta_i + \Delta\theta) - r(\theta_i - \Delta\theta)}{2\Delta\theta}. \tag{7}$$

By applying the above process repeatedly for all elements of the parameter, we can approximately estimate the gradient as $\frac{\partial\eta(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}$, and thus could improve the control policy in a gradient ascent manner as $\boldsymbol{\theta} \leftarrow \boldsymbol{\theta} + \alpha\frac{\partial\eta(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}$, where $\alpha$ is a small positive scholar.

## III. Experiments

In this section, we describe the experimental settings and the obtained results. Section III-A shows the initialization for setting the control policy and desired values of the topology coordinates in the reward function. In this study, both are provided by using a direct teaching approach. Section III-B demonstrates the obtained results by applied the reinforcement learning in a real environment.

### A. Direct Teaching for Setting Policy and Reward

In this study, we decided to improve the policy parameters for only the first joint of the both arms, through empirical investigation. In Fig. 6, the curve depicted with the solid line shows the trajectory of the robot's first joints, demonstrated by a human (initial trajectory) when the mannequin's neck inclination was 45 degrees and shoulder elevation was 105 degrees. Circles on the curve are via-points extracted with the method proposed by Wada et al. [9], and were used to initialize the control policy representing the demonstrated trajectory.

The demonstrated trajectory was also used to determine the targets $\mathbf{s}^{\text{target}}$ in the reward function (Eq. (4)).

### B. Learning Results

Fig. 6 also shows acquired trajectories of right arm (upper panel) and left arm (lower panel) through 4 episodes. Fig. 7 shows an example end-effector's trajectory of the right arm, from a lateral view, that was calculated from joint trajectories using inverse kinematics (see Fig. 2 for coordinates definition). Fig. 8 shows the mean of the reward obtained over 3 experiments. Example sequential snapshots in an initial and a learned movement are depicted in Fig. 9. Though the T-shirt was pulled over on the mannequin's head in the initial episode (episode zero), the mannequin's head successfully came out of the T-shirt neck in the episode 3. This result shows that, through the learning process, the robot was able to acquire the movements for successful clothing assistance even when the posture of the mannequin was slightly changed from the initial movement.

## IV. Conclusion

In this paper we have presented a novel learning system for an anthropomorphic dual-arm robot to perform the clothing assistance task. The keys of our system are to apply a reinforcement learning method for coping with the posture variation of the assisted person, and to define a low-dimensional state representation utilizing the topological relationship between the assisted person and the non-rigid material. With our developed experimental system, for T-shirt clothing assistance, including an anthropomorphic dual-arm robot and a soft mannequin, we demonstrate the robot quickly learns to modify its arm motion to put the mannequin's head into a T-shirt.

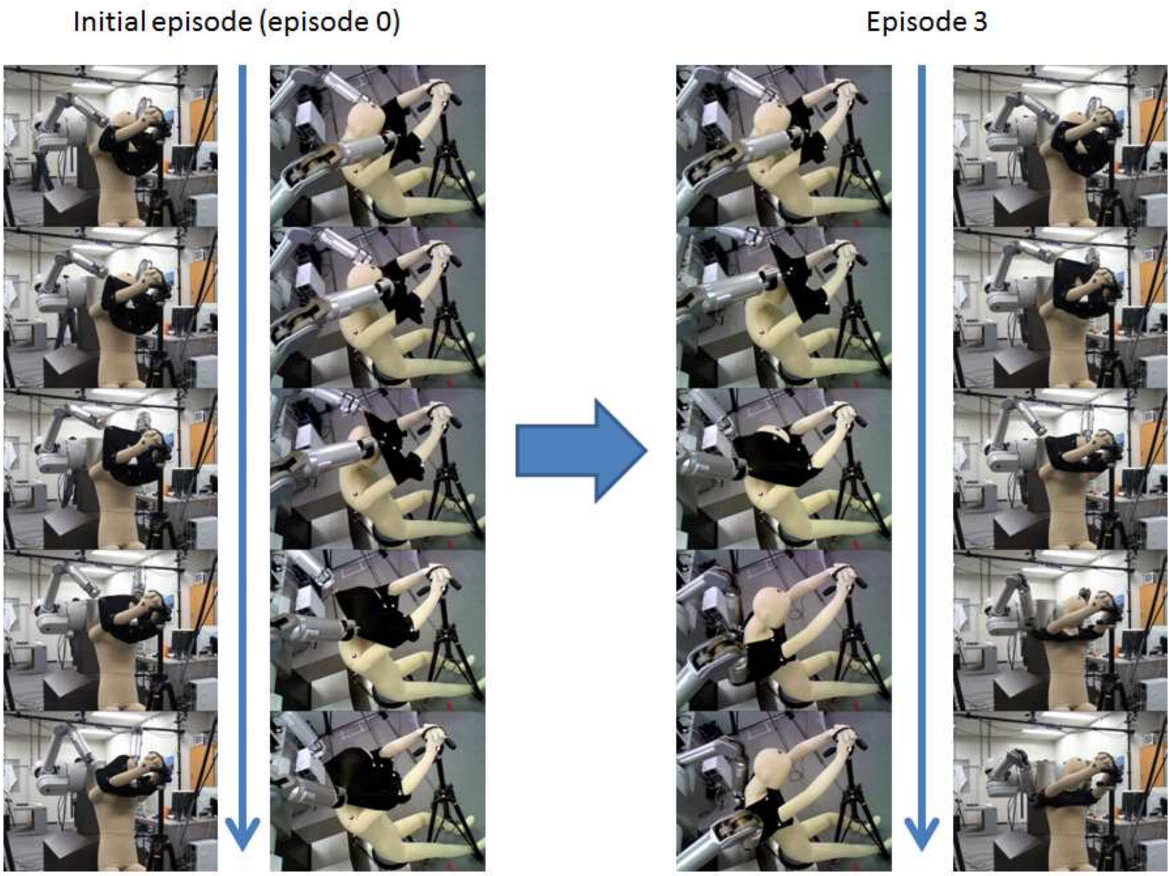


Fig. 9. Sequential snapshots of a successful learning process for the clothing assistance task.

Validation of the proposed method with more experiments with more participants is our near future work. Application of the proposed method to elderly or disordered persons will also be our future work.

## ACKNOWLEDGEMENT

This work was partly supported by Grant-in-Aid for Scientific Research from Japan Society for the Promotion of Science, No. 23240028.

## REFERENCES

[1] D. Shinohara, T. Matsubara, and M. Kidode, "A learning framework for motor skills with non-rigid materials based on reinforcement learning," in *Proc. of the 29th Ann. Conf. of The Robotics Society of Japan*, 2011, 3P3-4.

[2] J. Maitin-Shepard, M. Cusumano-Towner, J. Lei, and P. Abbeel, "Cloth grasp point detection based on multiple-view geometric cues with application to robotic towel folding," in *Proc. of the Int. Conf. on Robotics and Automation*, 2010, pp. 2308–2315.

[3] K. Yamazaki and M. Inaba, "A cloth detection method based on image wrinkle feature for daily assistive robots," in *Proc. in IAPR Conf. on Machine Vision Applications*, 2009, pp. 366–369.

[4] Y. Kita, T. Ueshiba, E. S. Neo, and N. Kita, "Clothes state recognition using 3D observed data," in *Proc. of the Int. Conf. on Robotics and Automation*, 2009, pp. 1220–1225.

[5] S. Edomond and T. Komura, "Character motion synthesis by topology coordinates," in *EUROGRAPHICS2009*, vol. 28, no. 2, 2009, pp. 299 – 308.

[6] W. Pohl, "The self-linking number of a closed space curve," *Journal of Mathematics and Mechanics*, vol. 17, pp. 875–985, 1968.

[7] H. Miyamoto, S. Schaal, F. Gandolfo, H. Gomi, Y. Koike, R. Osu, E. Nakano, Y. Wada, and M. Kawato, "A kendama learning robot based on bi-directional theory," *Neural Networks*, vol. 9, no. 8, pp. 1281–1302, 1996.

[8] J. Peters and S. Schaal, "Reinforcement learning of motor skills with policy gradients," *Neural Networks*, vol. 21, no. 4, pp. 682–697, 2008.

[9] Y. Wada and M. Kawato, "A theory for cursive handwriting based on the minimization principle," *Biological Cybernetics*, vol. 73, no. 1, pp. 3–13, 1995.

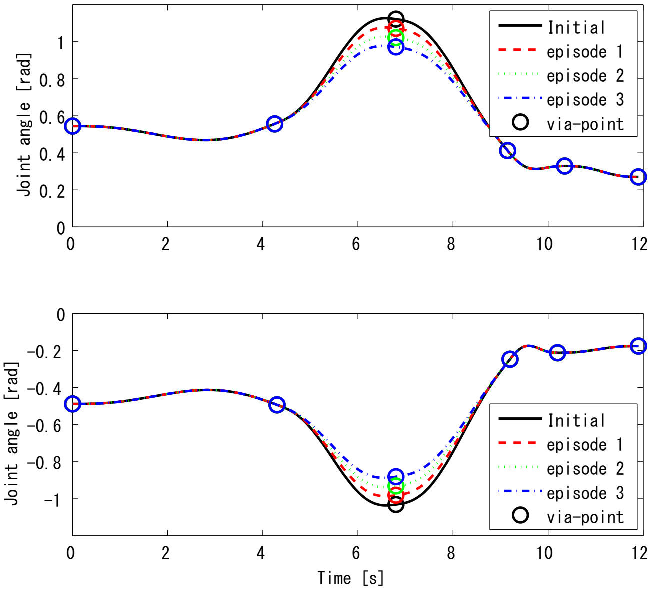


Fig. 6. Initial and acquired trajectories of right arm (upper panel) and left arm (lower panel)

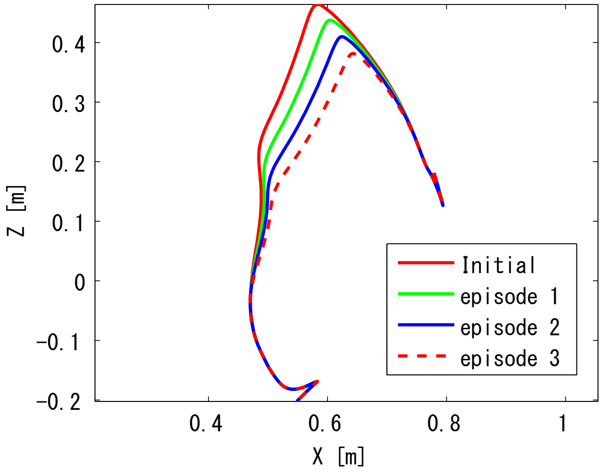


Fig. 7. Initial and acquired trajectories of robot's right end-effector

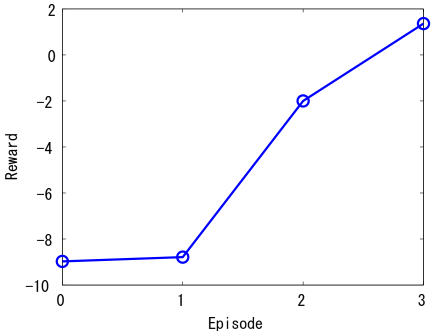


Fig. 8. Mean of the reward obtained over 3 experiments