Kullback Leibler Control Approach to Rubber Band Manipulation

Masaki Murase, Kimitoshi Yamazaki and Takamitsu Matsubara

Abstract— **In this paper, we study robot manipulation of rubber bands and focus on a task where a robot winds a rubber band around objects since such a task is very common in daily life. Rubber bands are highly flexible and deformable objects whose shape is changed by a force; thus, it is not easy for a robot to autonomously operate such tasks with a rubber band due to the difficulties of accurate shape modeling and shape measurement. A key insight for overcoming these difficulties is based on observing human skills; we generally pull and stretch a rubber band by applying certain force. The following are the potential benefits of stretching and pulling a rubber band: 1) reduction of the shape diversity and 2) increased friction generated on the contact points for strengthening the contact with the object. In this paper, we propose a manipulation planning method for winding rubber bands around objects that maintains the rubber band that is being stretched during the manipulation to avoid shape modeling and shape measurement difficulties. Such a method is formulated by a stochastic optimal control problem called Kullback-Leibler control. Real-world experiments were conducted with a dual-arm Baxter robot to verify the effectiveness of our proposed method. Our robot manipulated a rubber band and wound around an object with different numbers of windings and magnitudes of applied forces by changing the parameters in the cost function.**

Fig. 1: Examples of objects wound by a rubber band: keeping the book closed, holding the cover of the box, binding the multiple pens, and binding a long cable.

I. Introduction

Object manipulation is a key ingredient to integrate robots into our daily living environments. Many studies have been conducted not only with such rigid objects as cups and boxes (e.g., [1], [2]) but also with non-rigid objects such as ropes, cables, clothing. Non-rigid objects offer more complicated challenges for modeling and measurement, caused by deformation. Therefore, some studies are exploring a simplified modeling approach for deformable objects [3]–[7] or direct data-driven manipulation [8], [9] that explicitly avoids the modeling difficulty. However, no unified approach suitable for deformable objects has been established so far.

In this paper, we study the robot manipulation of rubber bands and focus on a task where a robot winds a rubber band around objects since this task is very often done with rubber bands. Even though rubber bands are used in a variety of applications, such as binding papers and cables, as shown in Fig. 1, in most cases, rubber band manipulation winds the rubber band around objects. Therefore, a manipulation planning method must be developed for such winding tasks by a robotic system.

However, like other deformable objects, since rubber bands are highly flexible and deformable objects whose shape is changed by applying a force, thus, it is not easy for a robot to autonomously operate rubber band tasks due to accurate shape modeling and shape measurement difficulties. A key insight to overcoming these difficulties is based on observing human skills; we generally pull and stretch a rubber band by applying a certain force. The potential benefits of stretching and pulling rubber bands include 1) reducing the shape diversity and 2) increasing the friction generated at the contact points for strengthening the contact with the object. In this paper, we propose a manipulation planning method for winding rubber a band around objects that maintains the rubber band while it is being stretched during the manipulation to avoid shape modeling and shape measurement difficulties. Such a constrained manipulation planning is formulated as a stochastic optimal control problem called Kullback-Leibler control [10]. By exploiting a characteristic of KL control regarding the compatibility of optimal control with free dynamics, this framework can reasonably generate a manipulation plan with constraints. Furthermore, we design a suitable cost function for the task by focusing on the topological relationship between the rubber band and the object since the desired situation is to wind a rubber band around an object. Such a topological relationship can be computed by utilizing *Gaussian linking integrals* [11].

Real-world experiments were conducted with a dual-arm Baxter robot to verify the effectiveness of our proposed

MM, TM are with Graduate School of Information Science, Nara Institute of Science and Technology, Nara, Japan.

KY is with Mechatoronics Systems Engineering, Faculty of Engineering, Shinshu University, Nagano, Japan.

Fig. 2: Experimental setup with a dual-arm baxter robot. Right hand holds the rigid object (made of wood) to be wound by the rubber band, and the left hand holds the rubber band with a tool. By controlling the right hand side, the robot tries to wind the rubber band around the objects with several conditions (e.g., magnitude of applied forces to the band or different numbers of windings).

method. Our robot manipulated a rubber band and wound it around an object with different numbers of windings and magnitudes of applied forces by changing the parameters in the cost function with our method.

The remainder of this paper is structured as follows. Section II presents a short summary of KL control as a preliminary to subsequent sections. Section III describes our proposed method for the manipulation planning of rubber bands. Section IV presents the experimental validation of our method, and Section V concludes with discussion and potential future work.

II. Kullback-Leibler control

Assuming a finite state space $\mathcal{X} = \{1, ..., N\}$, let $q_t(x) \geq 0$ $(0 \leq t < T)$ be the cost functions of the non-terminal states, and $g_T(x)$ be the terminal cost imposed at $t = T$, where T is called the horizon. Let $p(x'|x)$ be the free dynamics describing the system's behavior in the absence of controls. The control policy can impose any dynamics as $u_t(x'|x)$. However, it pays a price with the KL divergence between u and p. Therefore, the instantaneous cost is defined as follows:

$$\ell_t(x, u_t(\cdot|x)) = q_t(x) + \mathrm{KL}(u_t(\cdot|x)||p(\cdot|x)). \tag{1}$$

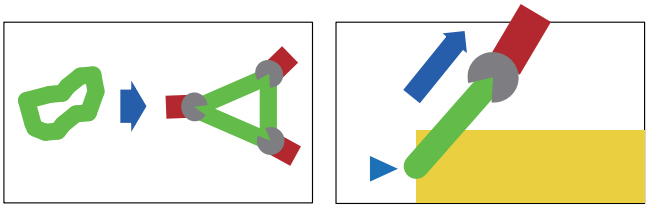

Fig. 3: Illustration of key insights to make rubber band manipulation tractable. By pulling the band, its shape gets to follow the contact point polygons, and the contact with the object gets strong and rigid due to frictions.

The objective is to construct an optimal policy $\pi_t^*(x) = u_t^*(\cdot|x)$ that minimizes the expected total cost as follows:

$$\pi^* \leftarrow \arg\min_{\pi} v_0^{\pi}(x) \tag{2}$$

where $v_t^{\pi}(x) = \mathrm{E}\left[g_T(x_T) + \sum_{i=t}^{T-1} \ell_i(x_i, \pi_i(x_i))\right]$, and $\pi^* = [\pi_0^*, \pi_2^*, \cdots, \pi_{T-1}^*]$.

We can obtain the optimal policy as follows by defining the *desirability function* $z_t(x) = \exp(-v_t(x))$, where $v_t = v_t^{\pi^*}$, and the linear operator $\mathcal{G}[z_t](x) = \sum_{x'} p(x'|x) z_t(x')$:

$$u_t^*(x'|x) = \frac{p(x'|x) z_{t+1}(x')}{\mathcal{G}[z_{t+1}](x)}. \tag{3}$$

The recursive formula of the desirability function, called the linear Bellman equation, is presented as follows:

$$z_t(x) = \exp(-q_t(x))\, \mathcal{G}[z_{t+1}](x). \tag{4}$$

Thus, we can efficiently solve the problem starting from the terminal cost backward in time by dynamic programming. Note that finding such an analytic form of optimal policy in stochastic optimal control is impossible in general.

The optimal policy of KL control only takes state transitions allowed by free dynamics. We exploit this characteristic to manipulate a rubber band to wind it around objects. In particular, when we design free dynamics, if the rubber band's shape does not follow the polygon composed of the contact points with the hand and the object due to insufficient stretching force, or if it loses the contact, such a state transition is set to 0 in probability as a constraint. With this setting, since the optimal policies hold the constraints and minimize the cost function, suitable behaviors for rubber band manipulation are created by avoiding such undesirable situations as ill-formed shape and loss of contact due to insufficient stretching and pulling force (Fig. 3).

III. Proposed method

In this section, we present our proposed manipulation planning method for winding rubber bands around the objects. Our method follows a formulation with KL control. All the ingredients are explained in details as follows.

State space: is designed with the following assumptions. 1) rubber band is in contact with both the left hand tool and the right hand object. Moreover, sufficient forces are applied to the rubber band so that it is stretched and pulled, thus, we can approximately capture the shape of the band by a

Fig. 4: Snapshots of direct teaching and visual manual confirmation to collect only compatible data. It satisfies that sufficient tension is assumed to be applied to the band so that we can approximately capture the shape of the band by a contact point polygon, and the contact may not be taken out thanking for the friction and tension.

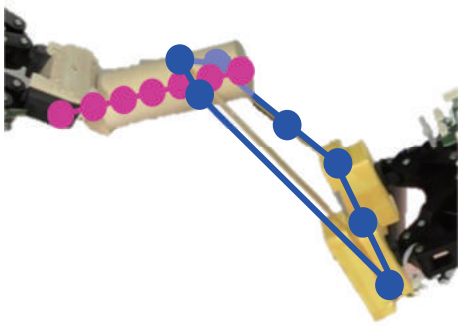

Fig. 5: Approximate models of rubber band and object, to be used for GLI computation, each of which is divided into seven line segments.

contact point polygon, and all the contacts are maintained thanking for the sufficient friction forces. Based on them, we set the state space by the robot's right arm's joint positions and velocities (7DOFs, 14 dim.).

Free dynamics: is designed so that each state transition does not meet the following situations: 1) the shape of the rubber band does not follow the polygon composed of the contact points with the hand and the object due to insufficient forces to stretch and pull, 2) it gets out of contact. In other words, state transitions that meet such situations are all set by 0 in probability. Thus, undesirable situations never happen when the robot follows this dynamics. In our experiment (see in Section IV.), direct teaching and visual and manual confirmation are used to properly collect only compatible data. Some of the snapshots of such direct teaching are shown in Fig. 4.

Since the robot's state is continuous variable, the standard KL control framework as shown in the previous section cannot be directly applied. A naive approach is to discretize the continuous state space into discrete states, however, since our target system has 14 dim, it is difficult to do due to the curse of dimensionality. Here we take an alternative approach: we used the latent KL control framework with Hidden Markov Models (HMMs) [12]. HMMs with discrete latent variables and continuous observations are used to *embed* the KL control with continuous state space into latent space of HMMs. In [12], it is shown that the curse can be alleviated. For the sack of space, the details of the method are omitted. See the [12] for more details.

Cost function: is designed to capture the topological relationship between the rubber band and the object related to how much the rubber band is being wound around the object. Such a topological relationship can be captured by the Gaussian Linking Integrals (GLI) as defined as follows:

$$w(\gamma_1, \gamma_2) = \frac{1}{4\pi} \int_{\gamma_1} \int_{\gamma_2} \frac{d\gamma_1 \times d\gamma_2 \cdot (\gamma_1 - \gamma_2)}{\|\gamma_1 - \gamma_2\|^3} \tag{5}$$

where γ_1 and γ_2 are the curves representing the shapes of the rubber band and the object, respectively.

Such a cost function has been first utilized in computer graphics [11] as a part of topology coordinates, and applied for robot control in simulation [13] and real robots [14]–[16] subsequently. Note that it does not require to measure the detailed shape of rubber band, if the shape follows the contact points polygon and all the contact positions in 3D space and object shape are known.

To compute GLI in Eq. (5) for the object and rubber band, we need to approximate the curves by multiple line segments. As shown in Fig. 5, we made such approximations by seven line segments. More precisely, for the object shape, the length between the right hand tip and the end of the object was divided evenly in seven line segments. For the rubber band shape, three vertices were allocated on the left hand, and other three were allocated on the objects along the rubber band, then the remaining one was allocated on the middle along the rubber band in the robot body side. The detailed scheme for computation is given in Appendix.

IV. Experiment

The experimental settings and results are described, respectively.

A. Setting

We designed and conducted the experiment to verify the effectiveness of our proposed method. The experimental system is shown in Fig. 2. The goal of manipulation planning is to wind the rubber band around the object with different numbers of windings and with different magnitude of applied forces on the rubber band. To this end, we designed the cost function as follows:

$$q_t(x) = \|H_{ref} - H_t\|^2 \ (0 \le t < T), \tag{6}$$

$$g_T(x) = \|w_{ref} - w_T\|^2 \tag{7}$$

where H_{ref} is the target and H_t is the actual hand distance at $t = t$. Since the contact points are not taken out (in our assumption), the magnitude of applied forces to the rubber band could be controlled by the distance of both hands in

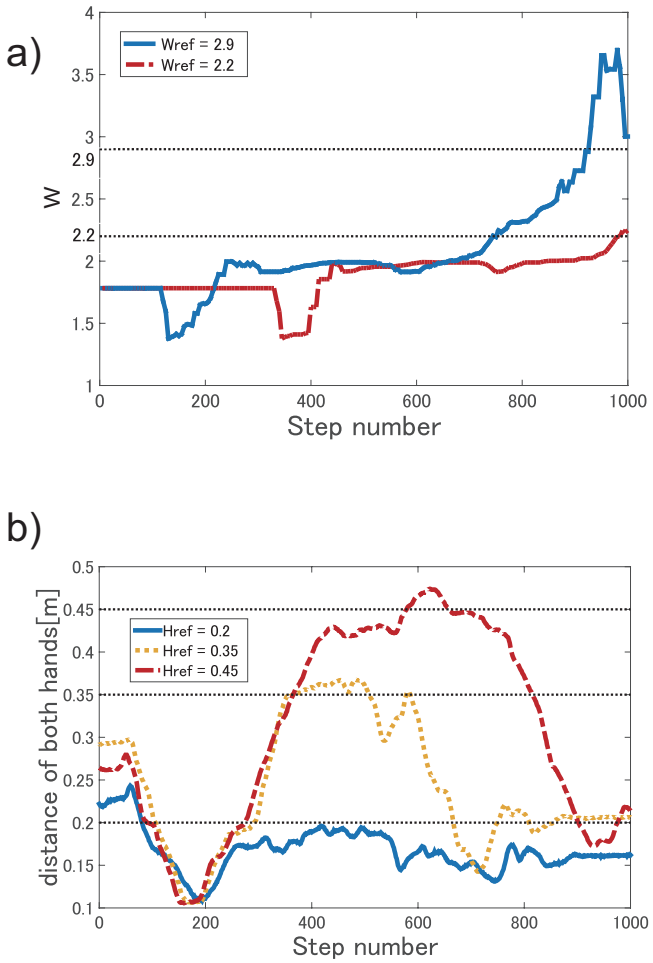

Fig. 6: Trajectories of a) GLI and b) Hand distance during the motion execution.

Euclidean space. w_{ref} is the target and w_T is the actual GLI between the rubber band and the object at $t = T$.

To compute GLI, the information of all the contact positions in 3D space and object shape were used.

For constructing the free dynamics, we collected compatible data from human demonstrations. In total, ten sets of trial data (each is 10 s with 1 kHz) were collected by direct teaching of double windings of the rubber band around the object and all the trials were certainly confirmed that the shape of the band could be approximately captured by the contact point polygon, and the contact might not be taken out thanking for the friction. Regarding the HMMs, we set 2000 latent states, and Gaussian observation model was allocated for each latent state. The horizon of the KL control was set by $T = 1000$. The standard EM algorithm was used to optimize all the parameters of HMMs including latent state transition model that was used as the free dynamics for KL control.

We set two different values in w_{ref} as $\{2.2, 2.9\}$ which were given from the human demonstration data as corresponding to the values of the single and double windings, and three different values in H_{ref} as $\{0.2, 0.35, 0.45\}$, respectively. Then, we investigate the effectiveness of our proposed method whether it could generate the rubber band manipulation for winding the rubber band around the object with different number of windings and different tensions without measuring the detailed shape of rubber band.

B. Result

We applied our method and generated policies were executed by the robot. Fig. 6 shows actual trajectories of GLI between the rubber band and the object, and hand distance. As you can see, the target GLI values are almost achieved at end for all the cases. Simultaneously, the hand distances are also tracked to the target values well. When $w_{ref} = 2.2$ was used, the resulted number of windings was one, while $w_{ref} = 2.9$ resulted in two, thus, wound twice sequentially. Some of the robot snapshots are shown in Fig. 7. For all the cases, the contacts were almost kept while it sometimes was slipped. From all the results, the effectiveness of our proposed method is verified.

C. Demonstration

As a demonstration of usefulness of our method, we replaced the rigid object used in the previous experiments to a paper tube. The result was illustrated in Fig. 8. Since the paper tube is very soft and fragile, it is important to control the magnitude of force applied to the rubber band when it is binded by the rubber band. With $H_{ref} = 0.2$, it was successful; the paper was properly wound by the rubber band without getting it jammed. On the other hand, with $H_{ref} = 0.45$, during the task execution, the paper got jammed, and the rubber band got slipped and the task was failed.

V. Conclusion

In this paper, we explored rubber band manipulation by a dual arm robot by focusing on the task of winding the object. We proposed a manipulation planning method for a rubber band based on KL control. The method was applied to a real robot system and its effectiveness was investigated through experiments. By changing the parameters in the cost function, single and double winding behaviors of the rubber band to the object with different hand distances were automatically generated by our proposed method with human-teaching and sensor feedback.

The approach of exploiting the nature of KL control for robot control was explored in dynamic imitation learning for humanoid robots [17] previously, where the dynamic balancing of the robot is considered as a constraint for mimicking human motions. However, such an approach has never been explored for deformable object manipulation as in this paper.

Our future work includes developing a self compatible-data collection system with several sensors rather than human demonstrations and manual confirmation. Another direction is generalizing the method to be applicable for several different deformable objects. Current framework requires rough object shape model for computing the GLI. This limitation

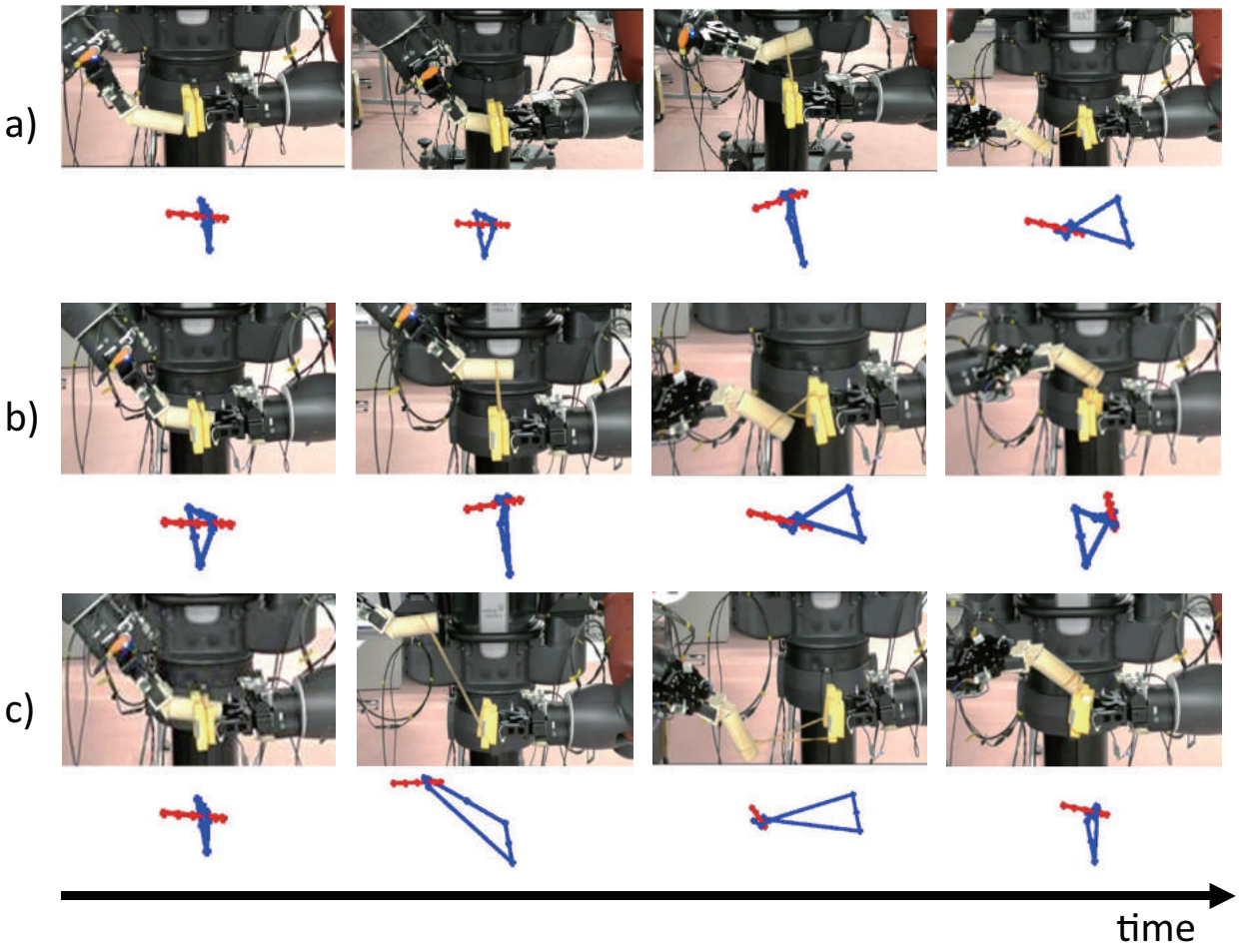

Fig. 7: Snapshots of robots during task executions. The red line indicates the object shape, and the blue line indicates the rubber band shape. a) short hand distance, single winding (w_{ref} = 2.2, H_{ref} = 0.2 m), b) short hand distance, double winding (w_{ref} = 2.9, H_{ref} = 0.2 m), c) long hand distance, double winding (w_{ref} = 2.9, H_{ref} = 0.45 m).

will be removed by using a RGB-D sensor and 3D object shape estimation.

Appendix: Gaussian Linking Integrals

This appendix briefly summarizes the computations of the GLI for two curves γ_1 and γ_2 proposed by [11]. First it is required to divide each curve into a number of small segments. Then those small segments are used to calculate the GLI of the two curves.

For two segments $\mathbf{r}_{ab}$ and $\mathbf{r}_{cd}$ (one from each curve), where $\mathbf{a} \in \mathbb{R}^3$ and $\mathbf{b} \in \mathbb{R}^3$ are starting and ending points of vector $\mathbf{r}_{ab}$, following vectors are calculated as

$$\begin{aligned} \mathbf{n}_a &= \frac{\mathbf{r}_{ac} \times \mathbf{r}_{ad}}{\|\mathbf{r}_{ac} \times \mathbf{r}_{ad}\|}, \quad \mathbf{n}_b = \frac{\mathbf{r}_{ad} \times \mathbf{r}_{bd}}{\|\mathbf{r}_{ad} \times \mathbf{r}_{bd}\|}, \\ \mathbf{n}_c &= \frac{\mathbf{r}_{bd} \times \mathbf{r}_{bc}}{\|\mathbf{r}_{bd} \times \mathbf{r}_{bc}\|}, \quad \mathbf{n}_d = \frac{\mathbf{r}_{bc} \times \mathbf{r}_{ac}}{\|\mathbf{r}_{bc} \times \mathbf{r}_{ac}\|}. \end{aligned} \tag{8}$$

With the above vectors, the *writhe* of these two segments can be analytically calculated by

$$\begin{aligned} T_{i,j} = \; & \arcsin\left(\mathbf{n}_a^T \mathbf{n}_b\right) + \arcsin\left(\mathbf{n}_b^T \mathbf{n}_c\right) + \arcsin\left(\mathbf{n}_c^T \mathbf{n}_d\right) \\ & + \arcsin\left(\mathbf{n}_d^T \mathbf{n}_a\right). \end{aligned} \tag{9}$$

Then, the GLI w, the writhe of the two curves, is obtained by

$$w(\gamma_1, \gamma_2) = \sum_{i=1}^{n_1} \sum_{j=1}^{n_2} T_{i,j} \tag{10}$$

where n_1 and n_2 are the number of segments for curves γ_1 and γ_2, respectively, and matrix $\mathbf{T}$ in which the (i, j) element is $T_{i,j}$ is called *writhe matrix*.

Acknowledgment

We gratefully acknowledge the support from the New Energy and Industrial Technology Development Organization (NEDO) for this research.

References

[1] A. T. Miller, S. Knoop, H. I. Christensen, and P. K. Allen, "Automatic grasp planning using shape primitives," in *IEEE International Conference on Robotics and Automation*, vol. 2, 2003, pp. 1824–1829.

[2] D. Gunji, Y. Mizoguchi, S. Teshigawara, A. Ming, A. Namiki, M. Ishikawa, and M. Shimojo, "Grasping force control of multi-fingered robot hand based on slip detection using tactile sensor," in *IEEE International Conference on Robotics and Automation*, 2008, pp. 2605–2610.

[3] M. Cani-Gascuel and M. Desbrun, "Animation of deformable models using implicit surfaces," *IEEE Transactions on Visualization and Computer Graphics*, vol. 3, no. 1, pp. 39–50, 1997.

[4] P. Volino, M. Courchesne, and N. Magnenat Thalmann, "Versatile and efficient techniques for simulating cloth and other deformable objects," in *22th Annual Conference on Computer Graphics and Interactive Techniques*, 1995, pp. 137–144.

[5] J. Barbič, M. da Silva, and J. Popović, "Deformable object animation using reduced optimal control," *ACM Trans. Graph.*, vol. 28, no. 3, pp. 53:1–53:9, 2009.

[6] B. Lloyd, G. Szekely, and M. Harders, "Identification of spring parameters for deformable object simulation," *IEEE Transactions on Visualization and Computer Graphics*, vol. 13, no. 5, pp. 1081–1094, 2007.

[7] G. Bianchi, B. Solenthaler, G. Székely, and M. Harders, "Simultaneous topology and stiffness identification for mass-spring models based on fem reference deformations," in *7th Medical Image Computing and Computer-Assisted Intervention*, 2004, pp. 293–301.

[8] A. X. Lee, H. Lu, A. Gupta, S. Levine, and P. Abbeel, "Learning force-based manipulation of deformable objects from multiple demonstrations," in *IEEE International Conference on Robotics and Automation (ICRA)*, 2015, pp. 177–184.

[9] 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 *IEEE International Conference on Robotics and Automation*, 2010, pp. 2308–2315.

[10] E. Todorov, "Efficient computation of optimal control," *PNAS*, vol. 106, no. 28, pp. 11 478–11 483, 2009.

[11] E. S. L. Ho and T. Komura, "Character motion synthesis by topology coordinates," *Computer Graphics Forum*, vol. 28, no. 2, pp. 299–308, 2009.

[12] T. Matsubara, V. Gómez, and H. J. Kappen, "Latent Kullback Leibler Control for Continuous-State Systems using Probabilistic Graphical Models," in *the Thirtieth Conference on Uncertainty in Artificial Intelligence*, 2014, pp. 583–592.

[13] E. S. L. Ho, T. Komura, S. Ramamoorthy, and S. Vijayakumar, "Controlling humanoid robots in topology coordinates," in *IEEE/RSJ International Conference on Intelligent Robots and Systems*, 2010, pp. 178–182.

[14] T. Matsubara, D. Shinohara, and M. Kidode, "Reinforcement learning of a motor skill for wearing a t-shirt using topology coordinates," *Advanced Robotics*, vol. 27, no. 7, pp. 513–524, 2013.

[15] T. Tamei, T. Matsubara, A. Rai, and T. Shibata, "Reinforcement learning of clothing assistance with a dual-arm robot," in *11th IEEE-RAS International Conference on Humanoid Robots (Humanoids)*, 2011, pp. 733–738.

[16] P. Vinayavekhin, S. Kudoh, J. Takamatsu, Y. Sato, and K. Ikeuchi, "Representation and mapping of dexterous manipulation through task primitives," in *IEEE International Conference on Robotics and Automation*, 2013, pp. 3722–3729.

[17] Y. Ariki, T. Matsubara, and S. H. Hyon, "Latent Kullback-Leibler control for dynamic imitation learning of whole-body behaviors in humanoid robots," in *IEEE-RAS 16th International Conference on Humanoid Robots (Humanoids)*, 2016, pp. 946–951.

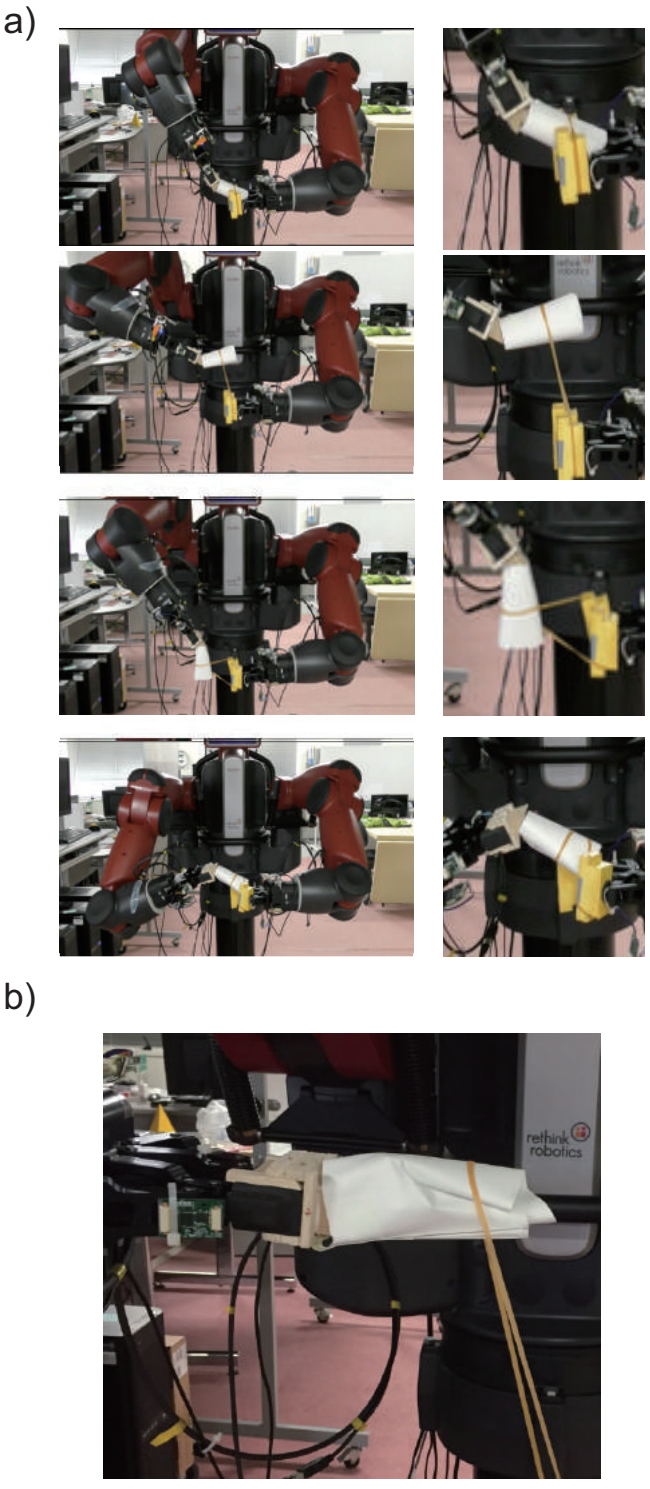

Fig. 8: Demonstration for paper tube. a) successful case, b) failure case. The paper tube gets jammed due to too strong applied force and the rubber band gets slipped during the task execution.