Picking up One of the Folded and Stacked Towels by a Single Arm Robot

Satonori Demura, Kazuki Sano, Wataru Nakajima, Kotaro Nagahama*[1], Keisuke Takeshita*[2], Kimitoshi Yamazaki*[1]

***Abstract*— This paper describes methods for doing a manipulation of picking up the top one of towels folded and stacked on a table. In this task, It is necessary to guarantee the reproducibility of shape of the picked towel when placing it again. Also, it is necessary to carry out lifting motion without moving other remaining towels. To overcome these issues, we propose work procedures and also propose methods for each procedure. The topics of this paper include the grasping position detection. We introduce YOLO to correctly detect the portion where should be grasped for stable manipulation. Also, reinforcement learning is applied for a grasping motion acquisition. We introduce Q-learning to obtain a simple and successful manipulation. The effectiveness of the proposed procedure and the methods were proven through experiments using an actual robot.**

I. Introduction

Autonomous robots that support our daily lives are desirable to have an ability to properly manipulate cloth products. The ability enables to automate daily routine work such as folding clothes, ironing, wiping the table. Therefore, studies to realize such tasks using autonomous robots has been conventionally progressed.

In previous studies aimed at automating the operation of cloth products, in order to deform the cloth product from the current shape to the target shape, manipulation methods are targeted. The initial state of a cloth product is often assumed that it is placed on a table casually. That is, the initial shape is positively unfolded, and then manipulations for making a desired shape are added. On the other hand, in the case of tasks such as a pick-and-place of a neatly folded cloth product, it is required to lift it up without losing its current shape state and to place it so that its shape state can be reproduced. That is, the manipulation skill required for such task is different from skills required when folding cloth products.

The purpose of this study is to pick up only the top one of the folded and stacked cloth products. For example, consider a situation where multiple folded towels are stacked. In executing the task of picking up only one of these, the desired things are as follows:

- Towels other than the lifted towel are not moved and collapsed,
- The lifted towel can be put back in the original folded state.

That is, it is necessary to guarantee the reproducibility of shape when placing the grasped towel again. Also, it is necessary to carry out a lifting method that does not move other towels. In this paper, we propose a work procedure to achieve them and also propose methods in each procedure. Then we demonstrate the effectiveness of the proposed methods using an actual robot with single arm.

The structure of this paper is as follows: In the next section, related work is introduced. In Section III, we describe problem setting and approach. Section IV describes a method of extracting the portion to be grasped from the folded and stacked cloth products. Section V describes how to acquire appropriate grasping methods. Section VI reports the result of verifying the proposed methods using an actual robot. Section VII summarizes this study.

II. Related work

There are many studies that manipulate cloth products by autonomous robots. However, as described in Section I, almost of them targets an issue that manipulates a cloth product from unarranged shape into a desired arranged shape. Maitin-Sphepard et al. [3] achieved a folding task of several cloth products by a mobile dual-arm robot. The robot picked up a cloth product one by one, and then folded them in order. Yuba et al. [4] focused on the task of unfolding a rectangular cloth product placed on a table. ”Pinch-and-slide” motion proposed by Shibata et al. [5] was used in this study, and the behavior for unfolding was selected by means of POMDP framework. In recent years, studies for folding or unfolding of cloth products are still proceeded, and we can find results such as Li et al. [6] and Xiong et al. [7]. In the study of Li et al., the trajectory of folding a cloth product placed on a table was learned using a physics simulator, thereby achieving efficient sleeve folding.

There are also studies that attempt to achieve some objectives by manipulating cloth products in a suspended state. Osawa et al. [8] and Willimon et al. [9] succeeded in identifying the type of cloth product by a robot observing the contour and the position of the lower end point while manipulating the cloth product. Kita et al. [10] proposed a method of matching a 3D point cloud measured using a trinocular stereo camera with a deformable shape model. Doumanoglou et al. [11] succeeded in unfolding and classification of fabric products while changing a hanging state. Comparing with the results of Osawa’s work, it is progressive to be able to target parts other than the lower end point, when choosing an appropriate grasping position.

*[1] Faculty of Engineering, Shinshu University, 4-17-1 Wakasato, Nagano, Nagano, 380-8553, Japan. *[2] Toyota Motor Corp., 4-17-1 Wakasato, Nagano, Nagano, 380-8553, Japan. {kyamazaki}@shinshu-u.ac.jp

In the above-mentioned studies it was necessary to detect a grasping position to properly manipulate the cloth product. A corner portion of the fabric, a lower end portion in a suspended state, or the like is set as a candidate for the grasping position, and appropriate manipulation has been realized. This point is in common with our study. However, unlike the objective of this study, there is no consideration for ensuring the initial shape. Also, when manipulating a cloth product, the existence of other cloth products is not explicitly taken into consideration.

As a preliminary study by the authors, we achieved the task of lifting a folded cloth product [12]. In order to lift it without collapsing the folded shape, we adopted a policy to detect and grasp the thickest folded hem. Although this study is a development of the previous study, there is a difference that a plurality of cloth products are folded and stacked. In this case, a novel problem arises in the portion for detecting the portion to be grasped and the part for lifting the cloth product after grasping. This paper also presents solutions to them.

III. Problem setting and approach

A. Problem setting

There are several folded and stacked cloth products. The purpose is to pick up the top one of these. By automating this work, it can be expected to apply this function to the storage of folded washing items, hand delivery of cloth products to people, and so on. In this study, it is assumed that a single arm robot is used. We use the human support robot HSR [13] developed by Toyota Motor Corporation. HSR is a support robot aiming to be a substitute for helper dogs. It has a serial manipulator with four rotating joints. The hand is a jaw gripper. In addition, it has an omnidirectional wheelbase and a parallel shift joint for raising and lowering its manipulator and head.

The method proposed in this study is not restricted for the use on HSR. However, it is assumed that there is a three-dimensional range image sensor capable of observing cloth products viewing from above and a color camera that can be photographed close to the cloth products. In the case of HSR, a three-dimensional range image sensor embedded in the head and a color camera embedded in the wrist of the robot arm can be used. As such a sensor configuration is common for a mobile manipulator robot, the generality of the proposed method is not lost.

B. Issues

There are two major issues to lift the one of folded cloth products without collapsing the stacked state. First, it is necessary to consider folding of cloth. When grasping up a cloth product, if a proper portion is not grasped, the folding and the overlapping of the cloth easily be collapsed. Therefore, it is necessary to appropriately select the grasping portion so that the shape state of the cloth product does not collapse greatly after grasping up.

The second is to properly separate only one cloth product from others. For example, it should be avoided that the second cloth product from the top is grasped with the first one at the same time. Also, when grasping and lifting a piece of cloth products, it should also be avoided that other underlying cloth products are moved at the table. From the above, the trajectory of the hand to grasp as well as the grasping position becomes study issues.

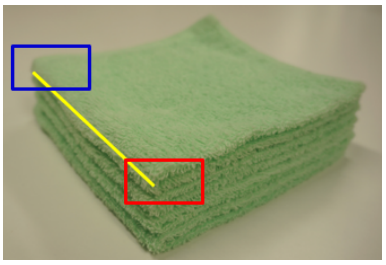

Fig. 1. Cloth placement and grasping candidate position

C. Approach

Based on the issues described in the previous subsection, we will explain the approach in grasping a towel as follows. The first one is of grasping position. Since cloth products are deformable object, the shape state can change during manipulation. Even so, it should be avoided that the shape state changes significantly. In the grasping task, for example, it is desirable to avoid situations where the originally-folded hems are collapsed, and the cloth is hung down loosely.

Fig.1 shows the candidate portions. Corners depicted by red square and blue square, and the thickest folded hem depicted by yellow line, are good grasping portions where the shape state is not collapsed so much. Furthermore, from the viewpoint of ease of recognition and ease of grasping, two places indicated by squares are appropriate as grasping position. These portions are easier to recognize than hem. In addition, such portions have an advantage that it is easier to determine the position for grasping by a single hand. The detection method of grasping points is described in the next section in detail.

For the second issue, the grasping and lifting motion, we introduce a reinforcement learning that the robot acquires proper motion automatically. The method proposed in this study should be usable even if the kind of cloth products, how to fold, the number of stacked products, are changed. In other words, it is desirable to change the grasping and lifting motions according to the above difference. However, it is troublesome to define them manually one by one. Therefore, we construct a reinforcement learning-based method to obtain an appropriate grasping posture. Details are described in Section V.

Here, we explain the flow from sensing to lifting of a towel. First, stacked towels are measured by the three-dimensional range image sensor mounted on the head of HSR. Plane detection process and plane removal process are applied to the point cloud obtained therefrom, and only the point cloud derived from the towels is extracted. Next, the three-dimensional position of the corner of the cloth is

Fig. 2. Emphasizing the thickest folded hem of cloth.

specified and the camera on the wrist of HSR is brought close to that position to judge whether it is a corner to be grasped or not. If it is a positive result, a target coordinate system is set there. Then HSR picks up the towel by the previously acquired grasping and lifting methods.

In the above procedure, detection of corners to be grasped might fail. Therefore, we add a manipulation to make it easy to detect the grasping position. Specifically, as shown in Fig. 2, bring a fingertip to a place where it is regarded that there is a corner, and push there to inflate a crease. As a result, the folded part of the cloth becomes larger, which makes it easy to detect a corner.

IV. Grasping position detection

A. YOLO[14]

As described in the previous section, a corner portion of the thickest folded hem is the grasping target. An image of a towel is photographed with a camera mounted on a robot manipulator and the corner portion is detected from the image. The problem here is miss-detection of the corner. Since appearance of such corners varies depending on how the cloth is folded, it is basically difficult to distinguish.

We use YOLO (You Only Look Once) as a robust and fast object detection method. YOLO is one of generic object detection algorithms using deep learning. It is faster than Faster R-CNN [15] etc., and it is possible to obtain equivalent detection accuracy. The structure is relatively simple. Moreover, since the preliminary learning result using rich image dataset such as ImageNet has been released, it is easy to use.

YOLO uses its own framework called Darknet19 [16]. Darknet19 is mainly made up of 22 convolution layers and 5 maxpool layers. It is a large difference from VGG [17] etc. because there is no fully-connected layer, so it is possible to propagate while keeping accurate position information of the feature map.

The detection method is as follows. First, an input image is divided into the cells whose number is the same at the vertical and horizontal directions. Next, a bounding box is predicted for each divided cell. After that, it calculates center coordinates, size and reliability for each bounding box, and at the same time, predicts what is included in the bounding box. Based on these results, it is possible to detect the object's class and location.

B. Learning and detection of the grasping corner

An example of a corner portion that should be detected is shown as a red rectangle in Fig. 1. The shape of a towel does not collapse if the portion can be grasped. We call this portion as grasping target corner and other corner portions as non-target corner. In order to detect corner portions and then to discriminate two types mentioned above, YOLO is used. Partial images photographing of grasping target corners and non-target corners are collected respectively, and transfer learning is performed from initial weight that is learned from ImageNet dataset.

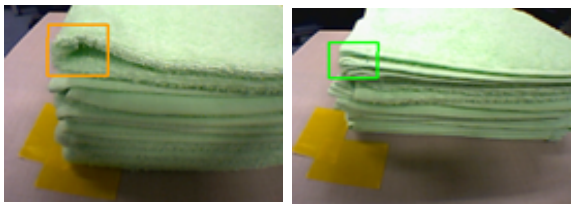

Fig. 3. Detection examples of a folded corner

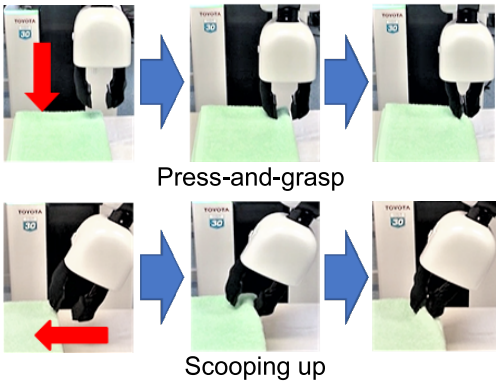

Fig. 4. Two types of grasping method for picking up a folded towel

Fig.3 shows an example of the result of detection performed on images from the camera attached to the wrist of HSR. Even if the second and subsequent towels are stacked in any orientation, grasping target corners were detected and discriminated correctly. However, there was a case that a grasping target corner did not detected correctly if the corner portion is not clearly appeared. In that case, as shown in Fig. 2, insert the manipulation to press the corner and then perform the detection process again.

V. Grasping method acquisition

A. Basic policy of grasping

We examined two kinds of grasping methods assuming a jaw gripper. The one is to press a place slightly away from the corner with a fingertip from the top, and then to grasp. Hereafter, this is called a pressing method. As shown in the upper panels of Fig. 4, the pressing method is a method that takes advantage of deformable characteristics of cloth. That is, there is a merit that grasping becomes easy because it is possible to float the portion to be grasped. Meanwhile, there is a disadvantage that the shape of floated corner is diverse depending on the state of force and the position to be pressed.

The other way to grasp is to insert fingertip between the topmost towel and the second towel and then to scoop up

the top one. Hereafter, this method is called a scooping method. As shown in the lower panels of Fig. 4, the scooping method is a relatively orthodox method, which is a way close to the action of human. Unlike the pressing method, the manipulation becomes simple as it is not necessary to consider the force adjustment. However, when inserting a fingertip between towels, if the position of the fingertip shifts, there are possibility that the shape of the towel to be grasped might collapse or the robot grasps two towels at the same time.

In our advance proof experiments using an actual robot, the pressing method was more appropriate. Therefore, in the next section we describe the automatic acquisition of the grasping motion on the basis of the pressing method.

B. Acquisition of grasping and lifting methods

The grasping policy described in the previous subsection was decided based on observations made by human developers. Therefore, the effectiveness is qualitatively ensured. On the other hand, the joint angle sequence actually given to the robot arm should be adjusted according to the type of cloth and how to place towels. An example is the difference of number of stacked towels. If the number of the cloth is large, the robot can grasp the topmost towel and can vertically lift it up. On the other hand, when the number of cloths is small, the friction between the table and the towels is small, so that the remained towels might slip when lifting the topmost one. Therefore, it is necessary to move the hand so as to prevent it. On the other hand, as another viewpoint, manipulation trajectory should be as simple as possible. As the number of steps increases, it takes time to execute the work accordingly.

From the above, it is desirable to appropriately select the manipulation trajectory according to the type of cloth product and the number stacked. However, it is burdensome to manually set it. Therefore, we take an approach to acquire robot motion using reinforcement learning. There are two kinds of actions: grasping of cloth by pressing method and lifting after grasping.

C. Acquisition of grasping method using Q-learning

In Q-Learning, a value called Q value is given to each of all possible actions a at the state s. Every time doing an action, Q value is updated using the update formula, and the behavior with the maximum Q value is finally obtained. The update formula is as follows:

$$Q(s_t, a_t) \rightarrow \\ Q(s_t, a_t) + \alpha\{R + \gamma \max Q(s_{t+1}, a_{t+1}) - Q(s_t, a_t)\}, \tag{1}$$

where α is a learning rate, R is a reward, and γ is a discount rate. $Q(s_t, a_t)$ is a Q value at time t, $\max Q(s_{t+1}, a_{t+1})$ is a Q value at time $t+1$. It is the maximum value of the Q value. In this way, since the Q value at the time t is updated using the Q value at the next time $t+1$, the influence of the previous action immediately appears.

The selection of action is done using ϵ-greedy method. This is a method to act randomly with a certain probability ϵ and to take action based on a high Q value with the probability of $1-\epsilon$. In general, $\epsilon = 0.1$, so we set it as well. This will act randomly with a probability of 10% and take actions with high Q value with a probability of 90%.

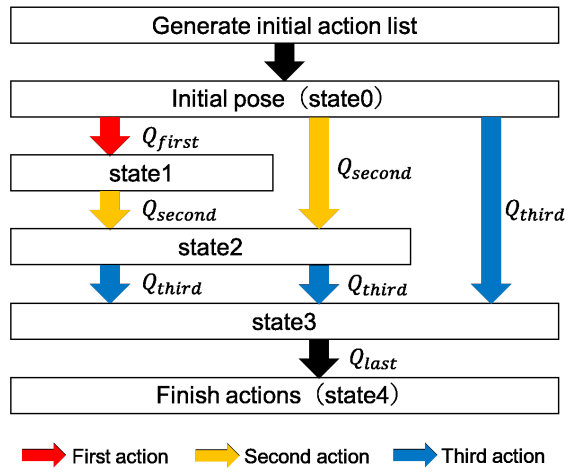

Fig. 5. A sequence for multi-step actions generation

Fig. 5 is the flow of manipulation learning. In a single trial, during the period from the initial state to the final state of the manipulation, an appropriate action is searched based on ϵ-greedy method. An action consists of 1 to 3 steps, and this number is randomly selected. It is most preferable when grasping or lifting is completed in one step. However, in some cases, a manipulation might not succeed unless it is a multi-step. That is, since how many steps are necessary is unknown, the number of actions is also considered as a searching target.

For proceeding Q learning, set the movable range and step size of each joint angle in advance. Also, set the height of fingers and step width. Then, as shown in Fig. 5, *First action, Second action*, and *Third action* are prepared assuming that the action has a maximum of three steps. This enables to create an action list that stores the joint angle and the height of the hand in each action.

The ways of setting and update of Q value are as follows. In the initial state, Q values are set to 0. When one action is completed, the Q value set for the action is updated based on the update formula, shown in Eq.(1). For each of state 0 to state 4, it is represented as follows.

$$\begin{aligned} &state0 \rightarrow state1 \\ &\quad Q_{first} \rightarrow Q_{first} + \alpha(0 + \gamma \max Q_{second} - Q_{first}) \\ &state1 \rightarrow state2 \\ &\quad Q_{second} \rightarrow Q_{second} + \alpha(0 + \gamma \max Q_{third} - Q_{second}) \\ &state2 \rightarrow state3 \\ &\quad Q_{third} \rightarrow Q_{third} + \alpha(0 + \gamma \max Q_{last} - Q_{third}) \\ &state3 \rightarrow state4 \\ &\quad Q_{last} \rightarrow Q_{last} + \alpha(R - Q_{last}) \end{aligned} \tag{2}$$

The rule of giving reward R at the update formula is as follows. If the work has failed after manipulation, the reward is set to 0. If it was successful, the following three patterns are selected:

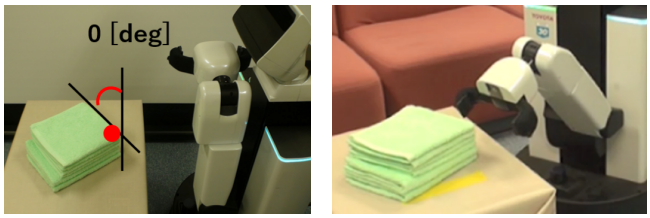

Fig. 6. An experimental situation. Left: positional relationship of initial robot position and cloth position, right: the observation of corner portion using in-hand camera.

- Set 5.0 if the number of states was 1,
- Set 3.0 if the number of states was 2,
- Set 1.0 if the number of states was 3.

Based on the above, it becomes to select a successful manipulation with small number of actions.

VI. Experiments

A. The experimental method

As shown on the left of Fig.6, the six rectangular towels were folded in quarters and stacked on a table. The orientation of the towels was randomly placed within the range of ± 40 deg. Here the angle of the thickest folded hem was set to 0 degree that is transverse direction with respect to the front of the robot. The head of the robot was tilted in advance so that the towels can be measured by a three-dimensional range image sensor embedded on the head.

Corner detection by YOLO was prepared as follows. The towels folded and stacked were photographed by the wrist camera of HSR. Then, 130 images of the corner portion to be grasped and 60 images of other corner portions were collected, respectively. After modifying the final layer of the YOLO network so as to detect these two classes, transfer learning was performed using the weight previously learned with ImageNet dataset.

The overall flow of a manipulation experiment was as follows. First, towels placed on a table were measured by a 3D range image sensor on the robot head, and the inclination around the vertical axis of the towels was calculated. Next, the position of the corner portion (the red circle portion on the left side of Fig. 6) was extracted from 3D point cloud, and transformed to the position from the coordinate system fixed to the wheelbase of the robot. Then, the robot moved to a place where the robot and the towels had a predetermined positional relationship. After that, the target posture of the robot hand was determined, and the manipulation plan was made. Using the result, robot moved its hand, and brought a camera close to the corner (Fig. 6, right). When the grasping target corner could be detected from the photographed image, the grasping manipulation explained in Section III was performed. If it could not be detected, an additional action to emphasize folded portion was performed. Then, the detection process was performed again, and when the detection succeeded, the robot grasped and lifted the topmost towel.

TABLE I
Success rates of two types of grasping method

Method	Success rate
Push-and-grasp (with emphasis)	86.7%
Push-and-grasp (without emphasis)	73.3%
Scooping up	60.0%

In grasping manipulation, both of the two methods described in V-A were tried. The motions of the joints for grasping and lifting were given manually.

B. Experimental results of manipulation

Table I shows the experimental result. 15 trials were conducted and success rate was 86.7% by the pressing method after emphasizing the thickest folded hem. Meanwhile 73.3% by only the pressing method, and 60.0% by the scooping method. From the above, it was found that the combination of the emphasizing manipulation with the pressing method are effective. As failure cases seen in this experiment, there were mis-recognition of corner portion, and collapsing the shape of the towel after grasping.

C. Grasping method acquisition

In the above experiment, the motion of grasping and lifting was manually given. This subsection introduces the verification result of the method described in Section V. Some variations of the number and the kind of cloth products were examined in this experiment. The manipulation to be acquired were two kinds; the manipulation until a cloth product is grasped and the manipulation to lift the cloth after grasping. Based on the experimental result shown in the previous subsection, basic trajectory was set along the pressing method. Then, appropriate manipulation was acquired by means of reinforcement learning within a range that it does not deviate from the basic trajectory.

Judgment of the success/failure of the manipulations was done visually. Specifically, human operator described as failures if there were the case that the towel could not be grasped, the case where the shape of the cloth collapsed, and the case where remained towels were moved. For the parameters of Eq. (1), $\alpha = 0.1$ and $\gamma = 0.9$ were set.

Fig. 7 shows a manipulation acquired when six towels were stacked. A one-step grasping action to lower the hand straight down and a one-step lifting action to raise the hand straight up were acquired, respectively. The number of learnings until the Q value converged was about 60 and 70, respectively. On the other hand, Fig. 8 is the result of carrying out the experiment with two stacked towels. The grasping manipulation became one step as the same as the abovementioned case. On the other hand, in the lifting manipulation, it needed three steps; raised the hand straight, twisted the wrist, and raised the hand again. In this case, one step manipulation did not work because the remained towel was moved on the table during lifting. The numbers of learning in this experiment were about 60 and

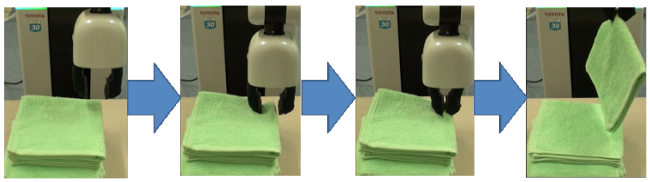

Fig. 7. An example of picking up a folded towel placed on other five towels

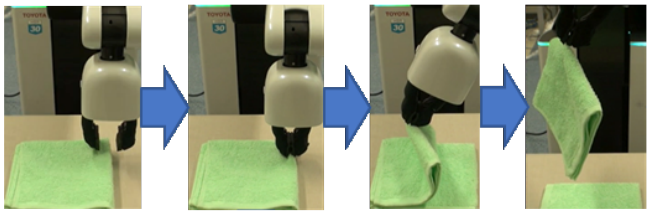

Fig. 8. An example of picking up a folded towel placed on another towel

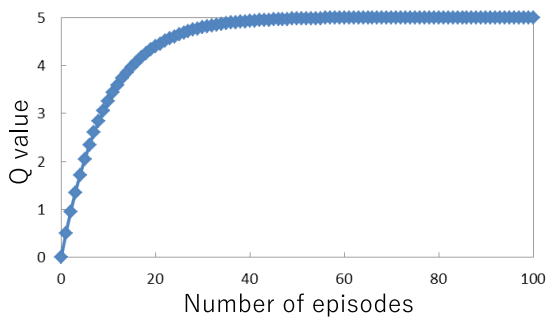

Fig. 9. Number of episodes vs. Q value. This graph shows a result of reinforcement learning in the case that six folded towels were stacked.

250, respectively. The same experiment was also conducted for the case where two bath towels were stacked. This result was the same as the case of two towels. The numbers of learning were about 70 and 300 respectively.

Fig. 9 shows the transition of the Q value in the lifting experiment in the case of six towels. Since it was possible to lift by the one step action, it converged to the maximum value 5.0 of the reward. On the other hand, when using two towels, the Q value converged to 1.0. At this time, the shape of the learning curve was almost the same shape as Fig. 9. In both cases, stable learning was performed.

VII. Conclusions

In this paper, we focus on a task to pick up one of the topmost towels which are folded and stacked on a table. We proposed an appropriate manipulation procedure and also proposed methods to realize the procedure. To guarantee the reproducibility of the shape of cloth products when placing a grasped cloth again, we found that grasping portion selection and additional behavior for emphasizing the corner are important. We confirmed that our methods enable to appropriately pick up a folded towel with 86.7% success rate. On the other hand, we examined to obtain effective manipulation behavior. It depends on the number and types of the stacked towels, so we took an approach to reinforcement learning. We confirmed that tens or several hundred times trials are needed for the manipulation acquisitions. This effectiveness was proven through experiments using an actual robot.

For future work we examine the proposed methods to various types of cloth products, e.g. folded shirts and trousers. In addition, more effective manipulation should be developed. We revealed that two types of grasping method are feasible, but the verification was done by dividing motion sequence into a few discretized postures. As this approach lacks the smoothness of motion, more detailed motion planning is required.

References

[1] K. Hamajima and M. Kakikura: "Planning Strategy for Unfolding Task of Clothes – Isolation of clothes from washed mass –," in Proc. of Int'l. Conf. on Robots and Systems, pp. 1237 – 1242, 2000.

[2] S. Cuén-Rochín, J. Andrade-Cetto and C. Torras: "Action Selection for Robotic Manipulation of Deformable Planar Objects," in Proc. of Frontier Science Conference Series for Young Researchers: Experimental Cognitive Robotics, pp. 1 – 6, 2008.

[3] J. Maitin-Sphepard et al.: "Cloth Grasp Point Detection based on Multiple-View Geometric Cues with Application to Robotic Towel Folding," Int'l. Conf. on Robotics and Automation, pp.2308 – 2315, 2010.

[4] H. Yuba, S. Arnold and K. Yamazaki: "Unfolding of a Rectangular Cloth Based on Action Selection Depending on Recognition Uncertainty," in Proc. of IEEE/SICE Int'l Symp. on System Integration, pp. 623 - 628, 2015.

[5] M. Shibata, T. Ota, Y. Endo and S. Hirai: "Handling of hemmed fabrics by a single-armed robot," in Proc. of IEEE International Conference on Automation Science and Engineering pp. 882 - 887, 2008.

[6] Y. Li, C-F Chen, and P. K. Allen: "Recognition of deformable object category and pose," In Proc. of IEEE Int ' l Conf. on Robotics and Automation, pp. 5558 5564, 2014.

[7] C. Xiong, N. Shukla, W. Xiong, and S. Zhu: "Robot Learning with a Spatial, Temporal, and Causal And-Or Graph," in Proc. of IEEE ICRA, pp. 2144 2151, 2016.

[8] F. Osawa, H. Seki, and Y. Kamiya: "Unfolding of Massive Laundry and Classification Types by Dual Manipulator," Journal of Advanced Computational Intelligence and Intelligent Informatics, Vol.11 No.5, pp. 457 – 463 , 2007.

[9] B. Willimon, S. Birchfield, I. Walker: "Classification of clothing using interactive perception," in Proc. of IEEE Int'l Conf. on Robotics and Automation, pp. 1862 - 1868, 2011.

[10] Y. Kita, F. Saito and N. Kita: "A deformable model driven method for handling clothes," in Proc. of Int.l Conf. on Pattern Recognition, Vol.4, pp. 3889 – 3895, 2004.

[11] A. Doumanoglou, A. Kargakos, T. Kim, S. Malassiotis: "Autonomous Active Recognition and Unfolding of Clothes using Random Decision Forests and Probabilistic Planning," in Proc. of Int'l Conf. on Robotics and Automation, pp. 987 - 993, 2014.

[12] Y. Moriya, D. Tanaka, K. Yamazaki, K. Takeshita: "A Method of Picking up a Folded Fabric Product by a Single-Armed Robot," ROBOMECH Journal 5:1. https://doi.org/10.1186/s40648-017-0098-y, 2018.

[13] Toyota Personal Assist Robot: http://www.toyota-global.com/innovation/partner_robot/family_2.html

[14] Redmon, Joseph and Farhadi, Ali: "YOLO9000: Better, Faster, Stronger," arXiv:1612.08242, 2016.

[15] S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: Towards Real-time Object Detection with Region Proposal Networks," in Proc. of NIPS, 2015.

[16] Darknet: Open Source Neural Networks in C: https://pjreddie.com/darknet/

[17] Karen Simonyan, Andrew Zisserman: "Very Deep Convolutional Networks for Large-Scale Image Recognition," arXiv:1409.1556, 2014.