Clothes handling based on recognition by strategic observation

Yasuyo Kita, Fumio Kanehiro, Toshio Ueshiba and Nobuyuki Kita
National Institute of Advanced Industrial Science and Technology(AIST),
AIST Tsukuba Central 2, 1-1-1 Umezono, Tsukuba, Ibaraki, Japan 305-8568.
{y.kita, f-kanehiro, t.ueshiba, n.kita}@aist.go.jp

***Abstract*—In this paper, we propose a method to recognize clothing shape based on strategic observation during handling. When a robot handles largely deformed objects like clothes, it is important for the robot to recognize a constantly varying shape. Large variation in shape and complex self-occlusion, however, make recognition very difficult. To address these difficulties, we have proposed a model-driven strategy using actions for informative observation and have developed some core methods based on this strategy [1][2][3]. In this paper, we show how these core methods can be used for an actual task that involves handling an item of clothing. In addition to proposing a sequence for this task, basic functions for realizing the sequence are also described. Using a robot, the experimental results demonstrated practical utility of the proposed strategy.**

I. Introduction

Owing to increasing importance of home and rehabilitation robots in an aging society, robots will be required to handle a variety of common objects including clothing. Because of its high degree of deformability, handling clothing is different from handling rigid objects. Large deformation in clothing leads to various possibility of positions and orientations of the part to be grasped, requiring considerable flexibility with respect to both visual recognition and motion control.

Compared to handling soft, string-type objects, such as ropes and electric cables [4][5], it is more difficult to recognize the state of clothing being handled because of complex self-occlusion. Here, the term "recognize the state" refers to the recognition of not only the geometric shape, but also the position of each part of the clothing within that geometric shape. This is essential information necessary for picking up a specific part (for example, the shoulder part of a sweater) , which is a basic action in many types of handling tasks such as folding clothing into a desired shape.

Although some studies have been conducted in the area of automated clothes handling [6][7][8][9][10][11][12], only a few have examined aspects related to the recognition of the complex clothing state. For example, Maitin-Shepard et al.[11] have demonstrated that a robot skillfully handles towels based on multiple-view observation. Their handling is, however, performed on the basis of corner detections rather than recognition of the state of the towels. Kaneko et al. [7] proposed a method that recognizes the clothing state by comparing the contour features (e.g., curvature, length-ratio) of observed data with models under the condition that clothes are held at two points in air. The state of clothes held at two points, however, has tremendous variations, because of the high number of two-point combinations. In addition, it is difficult to robustly extract the detailed contour from real observations; these are also highly sensitive to any slight deformation of the clothing. Cusumano-Towner et al. [12] proposed a method that estimates the clothing state in a manner similar to that proposed by Kaneko et al [7] and showed a good recognition rate using several types of clothing. However, their method requires a repeat re-grasping of the lowest point of the hanging item of clothing, which takes time and delays achieving the goal state. In addition, when the size of the clothing is large, it might be impossible for one hand to reach the lowest point, while still holding the item by the other hand.

Kita et al.[1][2] proposed a model-driven method of combining a simulation of clothing deformation with general knowledge about the target clothes, such as type (e.g., sweater, pants) and approximate size and softness, given in advance. The proposed method recognizes the state of the clothing item based on its possible 3D shape and calculates the necessary information for holding it at any specific point. In [2], they also provided additional "recognition-aid" actions such as spreading out the clothes, which can aid in understanding the state even in the case where the observed information is inadequate in estimating the state.

These studies, however, deal with isolated situations and assume that the clothes are already hung in air by holding a part on its rim. In this paper, we show how these basic techniques can be utilized to perform an actual task. As a concrete task, the robot must pick up an item of clothing that has an arbitrary shape from a desk and spread it open by holding the item by its specific parts, such as the two shoulders of a pullover or shirt or two points along the waist of trousers. This common task is useful for various purposes, such as neatly folding an item or presenting it to a person to wear. Using the above concrete task, we show a method to positively recognize a clothing state while handling the item. In Section II, we propose a processing sequence that uses informative observation. Basic actions that perform the sequence are described in Section III. In Section IV, we present some experimental results using an actual humanoid system and discuss the results.

II. Total flow

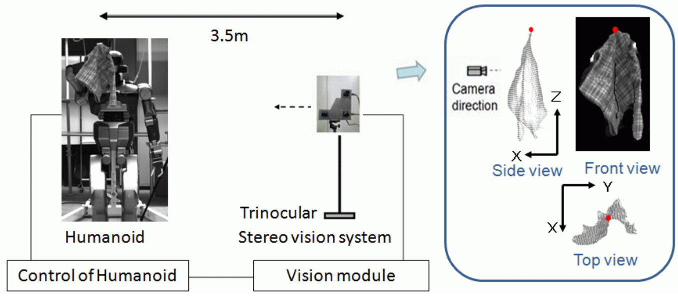
A. Our platform

Fig. 1 shows our experimental system for handling clothing, which consists of a humanoid robot, HRP2 [13], and a trinocular stereo vision system [14]. Although HRP2 has its own vision system, we currently use an external vision system, shown in Fig 1(a), to provide dense three-dimensional(3D) data of the entire clothing item. Fig 1(b) shows an example of 3D data obtained by the vision system. Texture-mapped 3D data is shown in the front view, while gray dots illustrate the 3D observed points[1] in the side and top views. The accuracy of the 3D reconstruction itself is about 1 mm, while the accuracy of the calibration of HRP2 and the stereo vision system is about $5-9$ mm. By using 3D data, the clothing region can be easily separated from the general background.

B. Flow using informative observation

In handling objects, it is essential to know the 3D information of the target part to be grasped. To facilitate re-grasping a clothing item with two hands, we create a situation in which the clothing state is easy to recognize. Fig. 2 shows the sequence of the situations. The task is to pick up an item of clothing that has an arbitrary shape from a desk and spread it open by holding it by its specific parts, such as the two shoulders of a pullover. After picking up it from the desk and holding the clothing in air using any part of its rim, variations in the shape of the clothing are drastically reduced. Actually, we have already proposed a recognition method for this constrained situation[1][2]. Fig. 3 shows the core idea of these studies. By simulating the physical deformation of the target clothing based on an approximate knowledge about the target, such as the type (e.g., sweater, pants), approximate size, and softness, possible 3D shapes of the clothing being held in air are predicted as shown in Fig. 3(a). Here, the clothing states are classified into "State 1," "State 2," ... according to the position in which the item is held. After observation, to consider the variation in the shape of clothing that can arise even when it is held at the same point, one representative shape of each state is deformed to better fit the observed 3D data. The state that shows the best fit to the observed data is selected as the correct state. After the clothing model is matched with actual observations, the position and normal direction of any part can be estimated by checking the 3D data corresponding to that part of the model. This is useful in determining the action plan for further handling of the part. Fig. 3(b) shows an example of the calculated action plan for grasping a shoulder part. The red, green and blue lines superposed on the left and right hands show the definition of their hand coordinates respectively. These lines are used for illustrating the position and orientation of the hand in this paper. For example, in Fig. 3(b), the red, green and blue lines illustrate the position and orientation of the left hand at the point when it should close its fingers to grasp the part; the hand should approach the part

[1] Although this vision system is able to capture 3D information at a rate of 30 frames per second, we currently use only static 3D data.



(a) (b)

Fig. 1. System for handling clothes: (a) configuration of robot and vision module; (b) three-dimensional data obtained from the stereo vision system (the viewing direction of the camera system is $-X$ with the red dot in the views illustrating the point at which the item is being grasped).

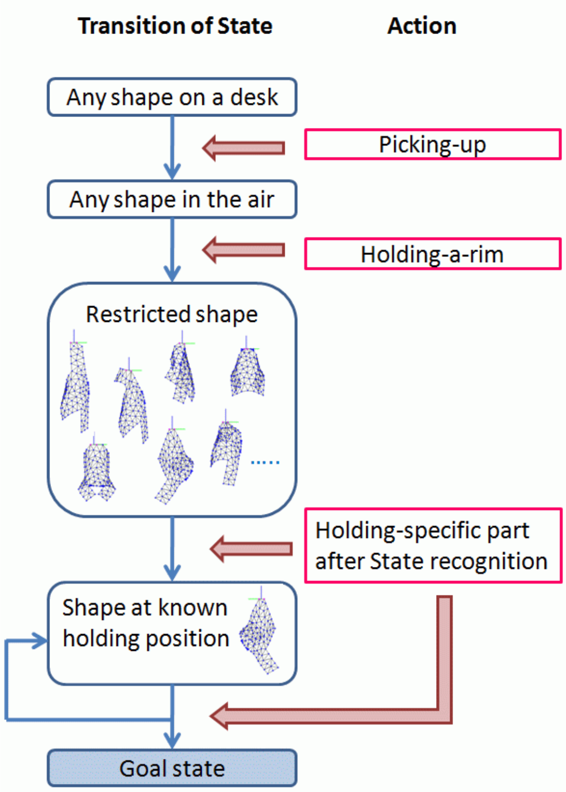


Fig. 2. Flow using informative observation

from outside along the blue line with its thumb open in the direction of the green line.

Once this recognition process is achieved, the holding position is tracked all subsequent re-grasping processes, to increasingly limit the variations in the possible shape.

III. Basic actions

In the next section, we explain how each action is realized. We call hand holding the item of clothing the "holding-hand", while the other hand, which is used for the next grasping action, is the "grasping-hand".

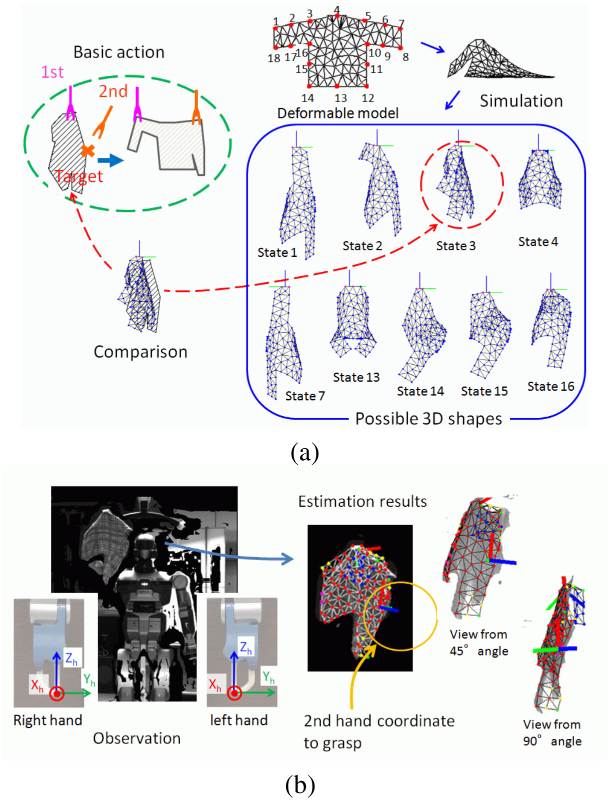


Fig. 3. Model-driven strategy of clothes recognition for automatic handling: (a) basic action and model-driven recognition; (b) calculation of action to perform the basic action; to grasp a specific part, the hand should approach the part from outside along the blue line with its thumb open in the direction of the green line.

A. Picking-up clothing from a desk

The action of picking up clothing from a desk consists of the following three parts:

1. Extract 3D clothing data

Assuming that the robot knows that the clothing item is on the table placed at a known position, the 3D data of the item can be separated from the background by extracting the connected observed region around the expected position. Fig. 4(b) shows an example of the result for the observation of Fig. 4(a), in which the expected position is illustrated by an orange point.

2. Calculate the height of the desk

To avoid that the robot hand hits the table, the height of the desk surface should be known. We estimate this from the lowest height of observed clothing region. The yellow line in the Fig. 4(b) shows the estimated desk height.

3. Select best grasping point

A part of observed clothing surface that is as vertical as possible and is higher than a few centimeters above the desk surface, is selected as the grasping position. The pose of the hand is determined so that the normal of the grip surface coincides with the normal of the part. In Fig. 4(b), the pink dot and pink line in the top and side views show the position and normal orientation of the detected part. Based on this information, the position and orientation of the grasping-hand

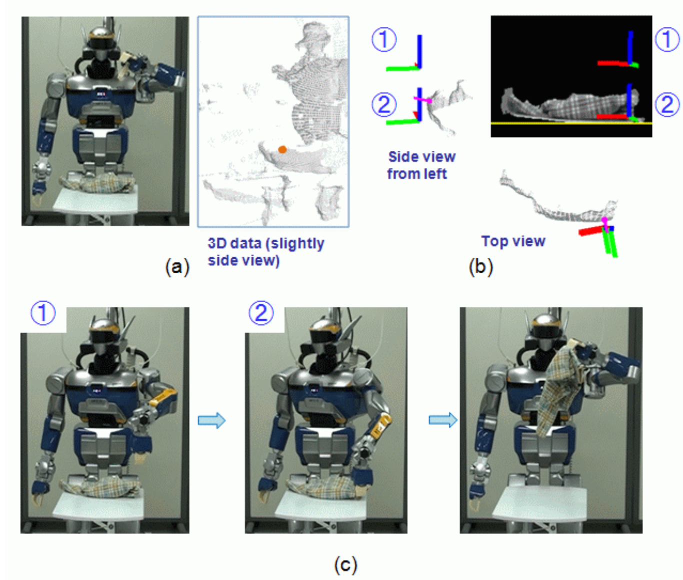


Fig. 4. Picking-up actions: (a) observation of the initial state; (b) calculated action plan; (c) actions performed on the basis of the plan.

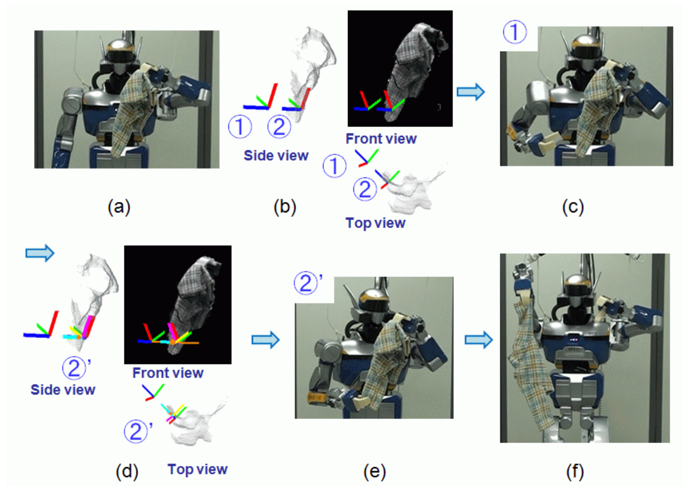


Fig. 5. Holding-a-rim actions: (a) observation at the initial state; (b) calculated action plan; (c) observation at the ready position; (d) recalculated action plan; (e),(f) actions performed on the basis of the new plan.

(left hand) at the ready and grasp positions are determined as shown by the red, green, and blue lines.

As shown, in our system, grasping actions are determined by the position and orientation of the grasping-hand at the ready and grasp positions. The grasp position is the position for closing the grip to grasp the clothing and the ready position is 10 to 20 cm away from the grasp position to set its approaching direction. These hand configurations are realized by a prioritized inverse kinematics solver[15] that calculates joint angles necessary to realize the given position and orientation of the hand while also considering joint limits and self-collision avoidance while maintaining the robot's center of gravity. Fig. 4(c) shows actions performed on the basis of the action plan of Fig. 4(b), where the final position is fixed in advance.

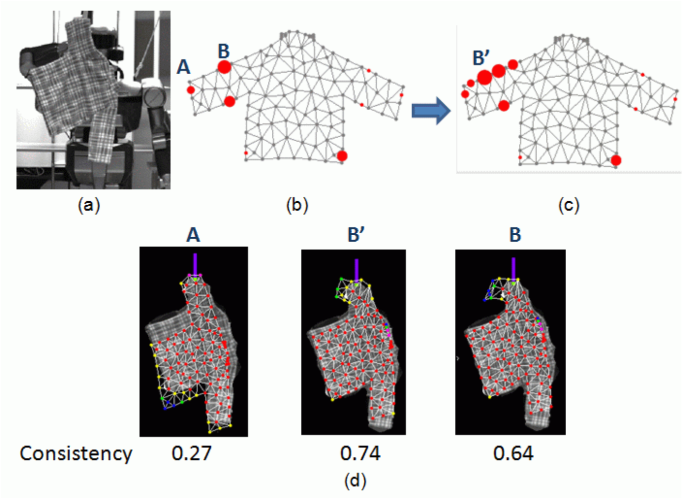


Fig. 6. Refinement of holding position

B. Holding a part of the rim

Actions of holding a part of the rim consist of the following three parts:

1. Extract 3D clothing data

When clothing is held by hand, its surface should be observed just under the tip of the holding-hand. Considering this, the 3D data of clothing can be separated stably from the background. Fig. 5(b) shows an example of the result of the observation in Fig. 5(a).

2. Select one point along the rim

For stable grasping, the area to be gripped should be comparable to the grip size. The area of the clothing that is the closest to the grasping-hand (the right hand in the case of Fig. 5) is selected. In Fig. 5(b), the red, green and blue lines illustrate the action plan that was calculated on the basis of the 3D position and normal orientation of the selected area. Fig. 5(c) shows the grasping-hand in the ready state.

3. Recalculate 3D information of the grasping point

In preparing to perform a grasp, the holding-hand moves slightly to maintain the robot's center of gravity, which is necessary when the grasping-hand makes large movements. Therefore, to gain a firm hold on the rim of the clothing, we make one more observation at the ready position, which is the observation of Fig. 5(c). The pink, yellow, and light-blue lines in Fig. 5(d), illustrate the recalculated position and pose of the hand in the grasp position using this new observed data. Figs. 5(e) and (f) show the actions that are actually performed on the basis of the recalculated action plan. The final position is determined on the basis of the distance between the holding position of the holding-hand and the grasping position of the grasping-hand.

C. Model-driven state recognition

At this stage, the holding position is restricted to any point along the rim of the clothes. Based on this situation, the method explained in Section II [1][2] is applied to recognize the clothing state. In this section, we propose an improvement in the method for increasing the accuracy and robustness of the recognition process.

Fig. 6(b) shows the results of the method[1][2] for the observation shown in Fig. 6(a). The size of the red points along the rim illustrates the degree of consistency between the observed data and the 3D shape that was predicted for the clothing when held in that position. Please refer to Fig. 3(a) for understanding the relation between 3D shapes and holding positions on the rim of clothing; the 3D shapes corresponding to the symmetrical points along the rim become also symmetrical, such as State 1 and State 7 corresponding to point 1 and point 7. Because the 3D shape of the hanging item of clothing should gradually change as holding position is moved along the rim, the consistency with the observation should change gradually, too. Considering this, we take a coarse-to-fine approach. That is, after this coarse process, points along the holding position that shows the highest consistency are used to check the consistency in shape when the clothing is held at those points. Fig. 6(c) shows the refined consistency after this process around the point B. By this refinement process, we can obtain a more accurate holding position and model fitting as shown by B' in Fig 6(d).

IV. Experiments

As preliminary experiments for the total task, we conducted experiments on sequences of handling clothing using our platform, namely "picking-up", "holding-a-rim", and "state recognition". We have performed the following 11 sequences using a pullover; the holding positions after holding-a-rim action became the end of a sleeve(4), a corner of the bottom of the body(2), the neck (2), the side of the body(1), the shoulder(1), and the elbow(1). Figs. 7 and 8 show two examples of the "side of the body" (No.1 in Table 1) and the "end of the sleeve"(No.7). A video showing an example of "elbow" holding (No. 10 in Table 1) is also provided as Supplementary Material. In Figs. 7 and 8, (a)−(d) show the picking-up process, (e)−(h) show the holding-a-rim process, and (i)−(j) show the recognition results of the observation of (h).

A. Picking-up action

In the experiments, the pullover was arbitrarily thrown on the table. Owing to the mechanism of the hand of our humanoid, the hand cannot pick up clothing items when the item lies flat on the table. Except such mechanically impossible cases, the picking-up actions were reliably performed for the other cases. Since we have experienced that the holding position of an item determined based on only its height often led a failure in the holding, it is essential to consider the normal direction of the holding position for system robustness.

B. Holding-a-rim action

The grasping positions that were calculated for the holding-a-rim action are listed in Table 1. The third column shows

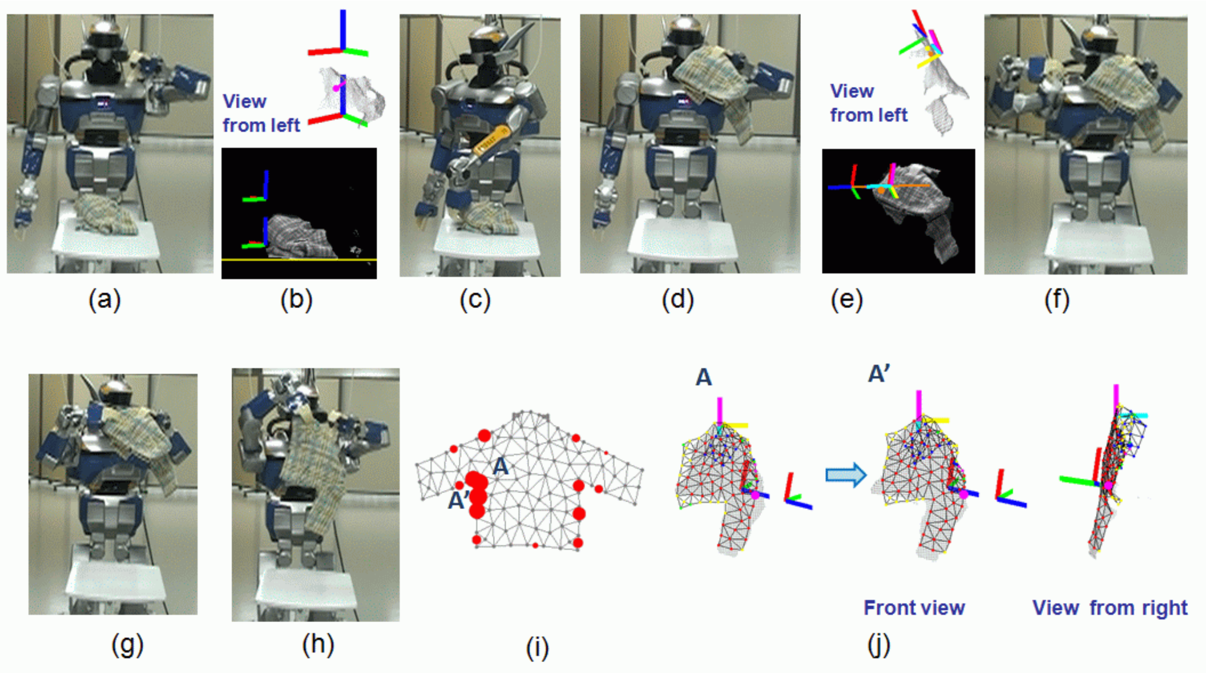


Fig. 7. Experiment 1 results

Table 1 Grasping position for holding-a-rim action

Data No.	Distance from surface (mm)	1st grasping position (mm)	Recalculated grasping position (mm)
1	8.6(-8.3,1.1,2.1)	(474.6,32.3,1288.4)	(466.0,-13.8,1287.6)
2	0.4(-0.2,0.3,0.2)	(502.2,-29.7,1214.4)	(515.2,-49.4,1213.1)
3	3.8(-2.7, 1.9, 1.9)	(506.4, 25.3,1227.6)	(490.1,-6.5,1221.0)
4	29.8(-29.6,3.1,1.7)	(452.0, 9.6,1142.8)	(444.8,-17.9,1139.5)
5	No data	(497.3, 46.2,1251.8)	(522.7, 74.5,1261.4)
6	35.3(35.3, 0.5, -1.2)	(485.4,-14.2,1256.3)	(498.6,9.4,1251.1)
7	11.8(-11.4, 1.5, 2.7)	(506.6,-12.0,1196.2)	(505.3,-44.2,1204.8)
8	166.5(-165.5, 4.4,18.2)	(440.1,-11.4,1302.3)	(606.3,-21.3,1286.6)
9	4.6(4.3, -0.1, 1.7)	(418.3, -4.0,1003.0)	(404.1,-18.0,1003.7)
10	2.7(-0.9, 0.4, 2.6)	(470.8, 45.9,1255.6)	(462.9,1.0,1255.6)
11	8.7(8.4, 2.1, 0.4)	(478.8, 7.3,1308.7)	(467.6,-33.9,1307.8)

the grasping position that was obtained on the basis of the first observation, such as Figs. 7(d) and 8(d). The second column shows the distance of these points to the clothing's surface after moving the grasping-hand to the ready position (Ex. Figs. 7(f) and 8(f)). These distances are calculated by searching the closest clothing surface along the line connecting the point and the optical center of the basic camera. If the holding-hand(left hand) never moves and the clothing is as it was in the first observation, then the distance should be zero, since these points are determined on the clothing surface in the first observation. Actually, in all cases, the holding-hand moved slightly and the clothing also moved. As shown in the table, in many cases, this movement is less than 1cm. However, in some cases, the distance becomes more than a few centimeters. In the case of No. 5, "No data" means no surface was observed along the view direction in the new observation. which means the clothes moved in a direction perpendicular to the view direction. In the case of No. 8, after the movement, a different part of the clothing occluded the clothing's surface, which was observed in the first observation. This causes a 16.65cm difference from the first observed point. Using a new observation, the holding positions are recalculated as shown in the forth column, which contributes to the holding-a-rim action as shown in Figs. 7(g) and (h) and 8(g) and (h).

C. State recognition

After the holding-a-rim action, the holding position was estimated by the state recognition method. In 8 of 11 cases, the states were determined accurately. In the remaining three cases, recognition failed mainly because of the observations were inadequate to judge the state owing to inappropriate view direction and/or tight fold in the clothes. Figs. 8(i) and (j) show examples of such failures. The state symmetrical to the correct one was wrongly detected. To avoid such failures, we can combine some "recognition-aid" actions, such as the horizontal rotation of the holding-hand or spreading out the clothing [3].

In the case of Figs. 7(i) and (j), the state was correctly detected and the holding position was refined from A to A' by the process described in Section III. This refinement decreases the distance from the ideal grasping position from 2.9 cm to 1.3 cm.

V. Conclusion

In this paper, we propose a method to recognize clothing shape based on strategic observation during handling. By using the action of holding a part of rim, the large shape variations of clothing are drastically reduced so that we can find the correct state from a relatively small number of possible shapes. In addition, taking advantage of consecutive shape changes created

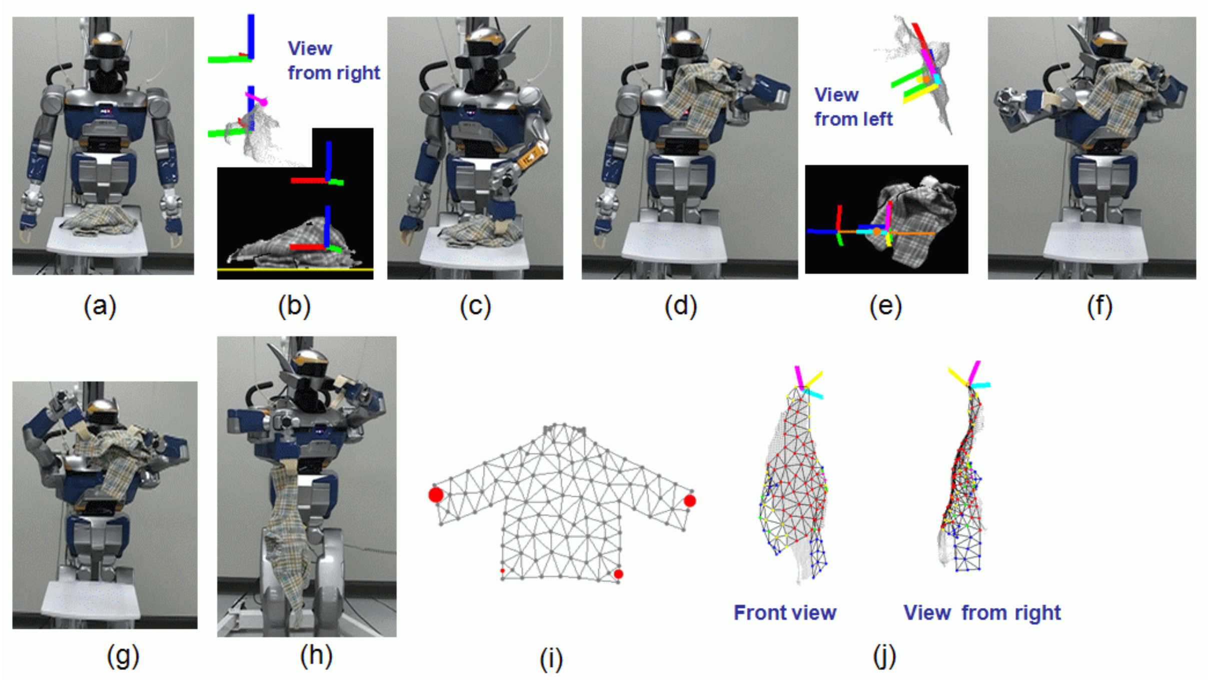


Fig. 8. Experiment 2 results

by changing the holding position along the rim, a coarse-to-fine process can be used to determine the holding position with a high resolution and the clothing state accurately. The experimental results have shown that a reasonable number of representative shapes can drive post-processing that can decide the holding position precisely. Using the example of holding-a-rim action, we also show the effectiveness of re-observation after a large motion is made by the robot.

Although we use only one item of clothing in the experiments of this paper, we believe that the proposed strategy is useful in common to other clothing. Verification to assure this generality is one of our future subjects. Through our experiments, we realize that the folding of clothes produces considerable difficulty in state recognition. To reliably perform all processes of the task of "opening the clothes", we plan to combine "recognition-aid" actions[3] into this framework.

ACKNOWLEDGMENT

This work was supported by a Grant-in-Aid for Scientific Research, KAKENHI (22240019). We are thankful to Dr. K. Yokoi and Dr. T. Nagami for their support in this research.

REFERENCES

[1] Y. Kita, T. Ueshiba, E. S. Neo and N. Kita: "Clothes state recognition using 3D observed data", In *Proc. of International Conference on Robotics and Automation*, pp. 1220–1225, 2009.

[2] Y. Kita, T. Ueshiba, E. S. Neo and N. Kita: "A method for handling a specific part of clothes by dual arms", In *Proc. of International Conference on Robots and Systems*, pp. 480–485, 2009.

[3] Y. Kita, T. Ueshiba, E. S. Neo and N. Kita: "Clothes handling using visual recognition in cooperation with actions", In *Proc. of International Conference on Robots and Systems*, (To appear), 2010.

[4] M. Inaba and H. Inoue: "Hand eye coordination in rope handling", *JRSJ (Journal of Robotics Society of Japan)*, Vol. 3, No. 6, pp. 32–41, 1985.

[5] H. Nakagaki, K. Kitagaki, T. Ogasawara and H. Tsukune: "Study of Deformation and Insertion Tasks of a Flexible Wire", In *Proc. of IEEE International Conference on Robotics and Automation*, Vol.3, pp. 2397–2402, 1997.

[6] E. Ono, H. Okabe, H. Ichijo and N. Aisaka: "Robot Hand with Sensor for Cloth Handling", In *Proc. 1990, Japan.U.S.A. Symp. on Flexible Automation*, pp. 1363–1366, 1990.

[7] M. Kaneko and M. Kakikura: "Planning strategy for putting away laundry –Isolating and unfolding task –", In *Proc. of the 4th IEEE International Symposium on Assembly and Task Planning*, pp. 429–434, 2001.

[8] Y. Kita, F. Saito and N. Kita: "A deformable model driven visual method for handling clothes", In *Proc. of International Conference on Robotics and Automation*, pp.3889–3895, 2004.

[9] K.Yamazaki and M.Inaba: "A Cloth Detection Method Based on Wrinkle Features for Daily Assistive Robots", In *Proc. the IAPR Conf. on Machine Vision Applications*, pp. 366–369, 2009.

[10] M. Shibata, T. Ota, and S. Hirai: "Wiping Motion for Deformable Object Handling", In *Proc. of IEEE Int. Conf. on Robotics and Automation*, pp. 134–139, 2009.

[11] 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 IEEE Int'l Conf. on Robotics and Automation (ICRA '10)*, 2010.

[12] Marco Cusumano-Towner, Arjun Singh, Stephen Miller, James O'Brien, Pieter Abbeel: "Bringing Clothing into Desired Configurations with Limited Perception", In *Proc. of IEEE Int'l Conf. on Robotics and Automation (ICRA '11)*, 2010.

[13] Kenji Kaneko and Fumio Kanehiro and Shuuji Kajita and Masaru Hirata and Kazuhiko Akachi and Takakatsu Isozumi: "Humanoid Robot HRP-2", In *Proc. of IEEE Int'l Conf. on Robotics and Automation (ICRA '04)*, pp.1083–1090, 2004.

[14] T. Ueshiba: "An Efficient Implementation Technique of Bidirectional Matching for Real-time Trinocular Stereo Vision", In *Proc. of 18th International Conference on Pattern Recognition*, pp.1076–1079, 2006.

[15] Oussama Kanoun and Florent Lamiraux and Fumio Kanehiro and Eiichi Yoshida and Jean-Paul Laumond and Pierre-Brice Wieber: "Prioritizing Linear Equality and Inequality Systems: Application to Local Motion Planning for Redundant Robot", In *Proc. of IEEE Int'l Conf. on Robotics and Automation (ICRA '99)*, pp.2939–2944, 2009.