The 2010 IEEE/RSJ International Conference on
Intelligent Robots and Systems
October 18-22, 2010, Taipei, Taiwan

Clothes handling using visual recognition in cooperation with actions

Yasuyo Kita, Ee Sian Neo, Toshio Ueshiba and Nobuyuki Kita

Abstract— **In this paper, we propose a method of visual recognition in cooperation with actions for automatic handling of clothing by a robot. Difficulty in visual recognition of clothing largely depends on the observed shape of the clothing. Therefore, strategy of actively making clothing into the shape easier to recognize should be effective. First, after clothing is observed by a trinocular stereo vision system, it is checked whether the observation gives enough information to recognize the clothing shape robustly or not. If not, proper "recognition-aid" actions, such as rotating and/or spreading the clothing, are automatically planned based on the visual analysis of the current shape. After executing the planned action, the clothing is observed again to recognize. The effect of the action of spreading clothes was demonstrated through experimental results using an actual humanoid.**

I. INTRODUCTION

As home and rehabilitation robots are expected to take an active role in an aging society, it becomes more important for robots to automatically handle daily necessaries. Clothing is one of such objects. Due to its large deformability, the techniques to handle clothing are fairly different from those for rigid objects. Even compared to handling soft, string-type objects, such as ropes and electric cables [1][2], additional difficulties occur to recognize the state of clothing due to its complex self-occlusion. Here, the term "recognize the state" refers to the recognition of not only the geometrical shape, but also where each part of the clothing is in the shape. This state information is necessary for executing the action of picking up a specific part (for example, the shoulder part of a sweater), which is a basic action for handling tasks such as folding clothing into a desired shape.

Although some studies have been conducted in the area of automated clothing handling [3][4][5][6][7][8], only a few have examined aspects related to the recognition of the complex clothing state. For example, Maitin-Shepard et al.[8] realized that a robot deftly handles towels based on multiple-view observation. Their handling is, however, carried out based on corner detections rather than recognition of the state of towels. Kaneko et al. [4] proposed a method which recognizes the clothing state by comparing the contour features (e.g., curvature, length-ratio) of observed data with models under the condition that the clothes is held at two points in the air. The states of clothes held at two points, however, have tremendous variations, because of huge number of combination of the two points. As a result, learning processes required for each clothes become fairly troublesome.

Y. Kita, T. Ueshiba and N. Kita are with National Institute of Advanced Industrial Science and Technology (AIST), AIST Tsukuba Central 2, 1-1-1 Umezono, Tsukuba, Ibaraki, Japan 305-8568. {y.kita, t.ueshiba, n.kita}@aist.go.jp

E. S. Neo is with Intellectual Ventures. eneo@intven.com

Kita et al.[9][10] proposed a model-driven method, which recognizes the state of clothing and calculates actions to hold its any specific part. The method predicts possible three-dimensional (3D) shapes using a deformable model of the clothing. Then, the one that best fits the observed data is selected among them. By using such a model-driven strategy, once the clothing state is recognized, the 3D position and orientation of each part can be obtained in a relatively straightforward manner. Such information is necessary for the task of grasping a specific part.

According to their preliminary experiments on state recognition, recognition failures tend to happen among certain confusing states. Especially, in the case that clothing is largely folded, it becomes very hard to distinguish the correct state from the states close to it. Here, it should be effective to actively change such ambiguous clothing shape into more understandable shape by the aid of robotic action. There is, however, few study of actions for aiding visual recognition besides changing view directions for robust processings (ex. [8]).

In this paper, we propose a strategy to make use of actions to observe the clothing in the shape which can be recognized robustly. First, when an item of clothing is observed, it is judged if the observation is enough informative for robust recognition or not. If not, "recognition-aid" actions to observe clothing in the shape easier to recognize are planned based on the current observation. After executing the actions, the clothing is observed again to recognize its state with more confidence. From now, in Section II, we briefly explain our fundamental idea for handling clothing and propose a strategy of visual recognition in cooperation with actions. In Section III, spreading-clothes action is picked up as a concrete example of "recognition-aid" actions. In Section IV, we show some experimental results using an actual humanoid system and discuss the results.

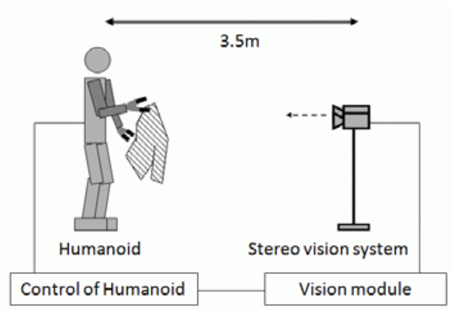
II. FUNDAMENTAL IDEA

A. Basis of our system for handling clothing

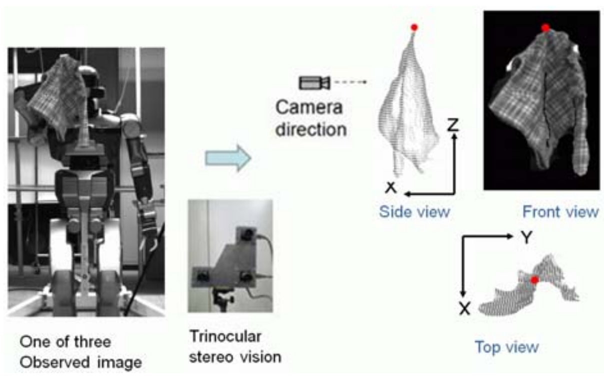
Fig. 1(a) shows the configuration of our experimental system for handling clothing, which consist of a humanoid robot, HRP2 [11], and a trinocular stereo vision system [12]. Although HRP2 has its own vision system, currently, we do not use it because it is unable to provide 3D data of the entire clothing item [1]. This stereo vision system is capable of recording dense 3D information with high accuracy as shown in Fig. 1(b). Here, the XYZ world coordinates are

[1] Aiming at effective usage of humanoid's eyes, we are studying a camera with wide field of view, in parallel.

978-1-4244-6676-4/10/$25.00 ©2010 IEEE



(a)



(b)

Fig. 1. System for handling clothing: (a) configuration of humanoid and vision module; (b) three-dimensional data obtained from the stereo vision system (the view direction of the camera system is $-X$, with the red dot in the views illustrating the holding position).

set as follows: the direction of Z-axis is the opposite to the gravity vector, and the direction of X-axis is the opposite direction to the view direction; the Y-axis completes the right-handed coordinate system. Texture-mapped 3D data is shown in the front view, while, in the side and top views, gray dots illustrates the 3D observed points. 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. The 3D data corresponding to the target clothing can be extracted as a connected 3D data region which includes the 3D position of the tip of the holding hand (the right hand in the case of Fig. 1(b)). 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 $\sim$ 9 mm.

One basic action for handling clothing is grasping a specific part of the clothing, that is held by one hand, by the other hand as shown in Fig 2. By iterating this action with two hands by turns, clothing can be held at a variety of desired states. For carrying out this basic action, the 3D position and orientation of the target part are essential information. We take a model-driven approach [9] based on the assumption that simple knowledge about the target clothing, such as type (e.g., sweater, pants etc.) and approximate size and softness, is known in advance. By simulating physical deformation of the target clothing based on this information, possible 3D shapes of the hanging clothing are obtained as shown in Fig. 2, Here, we classify the clothing states according to the position at which the clothing is held as shown as "State 1," "State 2," and so on. In Fig. 2, some

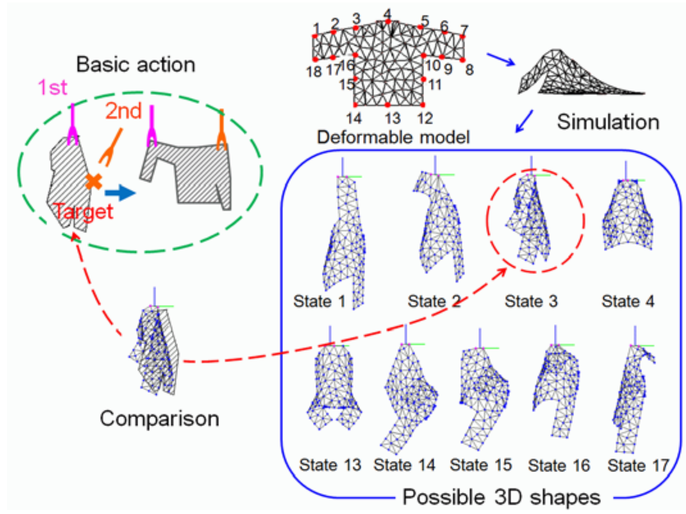
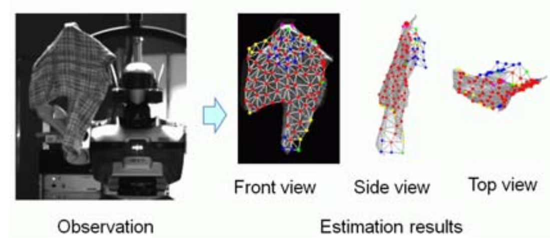
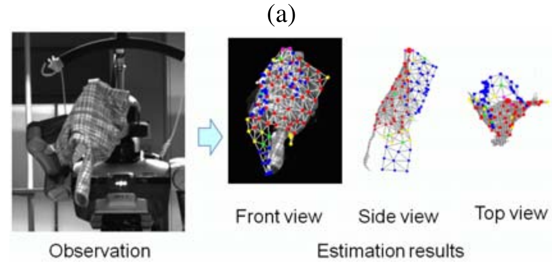


Fig. 2. Model-driven strategy to recognize clothing state.



(a)



(b)

Fig. 3. Example of "State" recognition results of sweater held at its side

examples of these states for the case of a sweater are shown. After observing the clothing, each representative shape is deformed to better fit the observed 3D data. The state which shows the best fit to the observed data is selected as the correct state. Once the clothing state is recognized in this way, the necessary information for target-grasping actions is straightforwardly obtained also in a model-driven way[10].

B. Strategy of visual recognition in cooperation with actions

As described in Section I, recognition failures tend to happen among certain ambiguous states. For example, in the case of Fig. 2, State 2, 3, 15 and 16 are sometimes confused with each other, while State 1, 4, and 14 tend to be robustly recognized owing to the characteristic vertical length of their observed 3D data. Actually, from preliminary experiments[9], success rate in recognizing State 16 was around 60%, while that for State 1 was almost 100%. Fig. 3(a) and Fig. 3(b) show one success and one failure examples for State 16 respectively. In the figures, the recognition

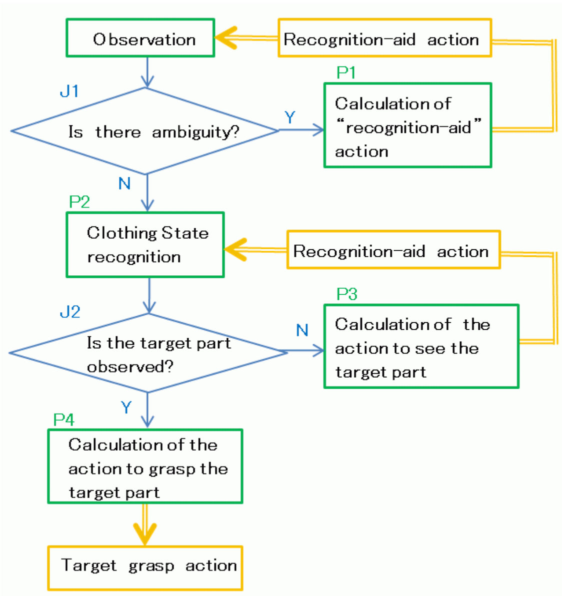


Fig. 4. Process flow of visual recognition in cooperation with actions

results are shown by clothing models superposed on the 3D data. In Fig. 3(b), the state symmetrical to the correct one was wrongly selected. As seen from the comparison of the two cases, large folding of the clothing makes the recognition difficult because the folding decreases observable area of the clothing. To actively change this ambiguous situations, the following two "recognition-aid" actions can be considered:

∗ Rotating the holding hand to show the self-occluded part of the clothing

∗ Spreading the clothing toward the vision system

The rotating action might cause new occlusion of the part which is observed at first. In such a case, integration of multiple 3D observed data are required. This task is not so simple when the clothing is deformed by the rotation action. In this paper, we leave this rotating action as a future subject and concentrate on spreading action.

Fig. 4 shows our general flow using "recognition-aid" actions. First the ambiguity of observation is checked (Judgement 1 (J1)). If the shape of the clothing is judged as a difficult case to recognize, then necessary actions for robust recognition are calculated by analyzing the first observation (Process 1 (P1)). After executing the resultant recognition-aid actions, the clothing is observed again. If the observed shape does not include ambiguity, it is analyzed in the same way proposed in [9] (P2). In the following section, as a typical example of recognition-aid actions, we propose a method to check the folding condition of the observed clothing and to derive the necessary spreading actions based on the observation.

III. Action of Spreading Clothing

A. Ambiguity check

First, the vertical length of the 3D observed region is checked to see whether it has confusing candidates or not. If so, the degree of the folding of the clothes is measured. In this paper, we consider the case that the clothes of interest

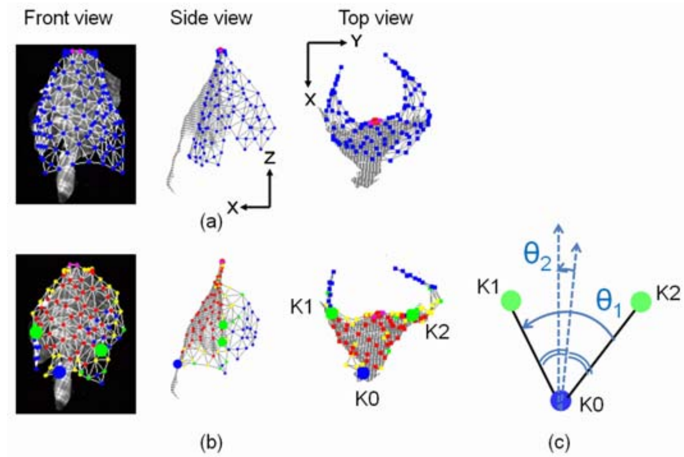


Fig. 5. Process to check the folding conditions

is not so soft to have many drapes. In such cases, the degree of folding can be measured as follows:

I) Deformable surface fitting
In order to estimate the convexity/concavity of the clothing surface below the holding position, a deformable surface is fitted to the observed data. As a temporal implementation for this fitting process, we use the deformable clothing model of the initial shape of State 4[2]. Figs. 5(a) and 5(b) show examples of the initial and fitted shape respectively for the data of Fig. 3(b). The red vertices illustrate ones which get on the observed 3D data.

II) Determination of three characteristic points, $K0, K1$ and $K2$ among the fitted vertices
Let $K1$ and $K2$ the vertices which had the smallest and largest Y coordinates at the initial shape before fitting. Then the vertex which produces the larger angle between $\overrightarrow{K0'K1'}$ and $\overrightarrow{K0'K2'}$ in the fitted shape, $K0$, is detected. Here, P' represents the projected point of a point P on the X-Y plane. That is, $P' = (P_x, P_y, 0)$ when $P = (P_x, P_y, P_z)$.

III) Calculation of characteristic angles, θ_1, θ_2
θ_1: angle of $\overrightarrow{K0'K1'}$ from $\overrightarrow{K0'K2'}$
θ_2: angle of the bisector of $\overrightarrow{K0'K1'}$ and $\overrightarrow{K0'K2'}$ from the view direction

The sign and absolute magnitude of θ_1 represent convexity/concavity feature and its degree respectively. We use a threshold of the absolute value of θ_1, T_{θ_1}, to discriminate between ambiguous/unambiguous cases. The value of θ_2 represents the degree of the deviation of the current view direction from the best view direction.

B. Determination of spreading-clothes actions

Using Fig. 6, we explain how to determine the position and pose of the hands for executing spreading-clothes actions. Here, we consider the case that the clothing is currently held

[2] Any planar deformable surface should work for this purpose. Just for labor-saving, we used clothing model in the shape at State 4, because of its symmetrical characteristics.

by the right hand of the robot. The hand coordinates of the left hand of our humanoid are shown in Fig. 6(a). The origin, O_h, is set at the center of the rotation axis of the wrist. The Y_h-axis is set along the rotation axis of the wrist, and the X_h-axis is set perpendicular to the flat front surface of the hand. The Z_h-axis completes the right-handed coordinate system. We carry out a spreading-clothes action by moving the hand from "Hand state1" to "Hand state2". Hence, our sub-goal becomes to determine "Hand state1" and "Hand state2".

1) Convex case ($\theta_1 \geq 0$)
In the case of convex towards the vision system as shown in Fig.6(b), the half-side of the clothing closer to the other hand is pushed toward the vision system. A plan for realizing this actions can be automatically determined as follows:

Hand State1:

$$\overrightarrow{Z_h} = \frac{\overrightarrow{K0'K2'}}{\left|\overrightarrow{K0'K2'}\right|}, \qquad \overrightarrow{X_h} = (0,0,1), \qquad \overrightarrow{Y_h} = \overrightarrow{Z_h} \times \overrightarrow{X_h}$$

The position of P_{touch}:

$$\left(\frac{K0_x + K2_x}{2}, \frac{K0_y + K2_y}{2}, z_0\right)$$

$$z_0 = H_{holding} - L_0$$

Here, $H_{holding}$ and L_0 is the height of the position of holding the clothing and a fixed length respectively.

Hand State2:

$$\overrightarrow{Z_h} = (0,1,0), \qquad \overrightarrow{X_h} = (0,0,1), \qquad \overrightarrow{Y_h} = \overrightarrow{Z_h} \times \overrightarrow{X_h}$$

The position of P_{touch}:

$$\left(K0_x, K0_y + \frac{\left|\overrightarrow{K0'K2'}\right|}{2}, z_0\right)$$

2) Concave case ($\theta_1 < 0$)

In the case of concave towards the vision system as shown in Fig.6(c), the furthest point from the vision system, K0, is pushed along the bisector direction between $\overrightarrow{K0'K1'}$ and $\overrightarrow{K0'K2'}$:

Hand State1:

$$\overrightarrow{Z_h} = (\beta, -\alpha, 0), \qquad \overrightarrow{X_h} = (0,0,1), \qquad \overrightarrow{Y_h} = \overrightarrow{Z_h} \times \overrightarrow{X_h}$$

The position of P_{touch}:

$$(K0_x, K0_y, z_0)$$

Here, (α, β, γ) is the direction cosine of the bisector.

Hand State2:

$$\overrightarrow{Z_h} = (\beta, -\alpha, 0), \qquad \overrightarrow{X_h} = (0,0,1), \qquad \overrightarrow{Y_h} = \overrightarrow{Z_h} \times \overrightarrow{X_h}$$

The position of P_{touch}:

$$(K0_x + L_1\alpha, K0_y + L_1\beta, z_0)$$

Here, L_1 is the distance of $K0'$ from the line of $K1'K2'$.

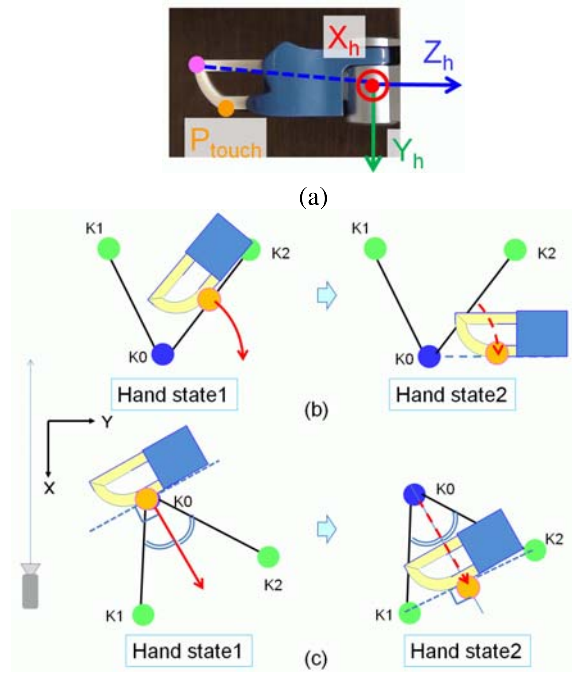


Fig. 6. Actions for spreading clothes toward the view direction: (a) HRP2 hand coordinates; (b) convex case; (c) concave case.

IV. Experiments

A. Relation between θ_1 and clothing state ambiguity

Using eight observations corresponding to State 16, we examined the relation between θ_1 and ambiguity in clothing state recognition. To measure the degree of ambiguity in clothing state recognition, we use the following value M:

$$M = \frac{R_{correct\ State}}{\max_{i \neq correct\ State}\{R_i\}}$$

Here, R_i is the overlap ratio which evaluates the consistency between observation and the predicted shape of State i and gets larger when the consistency increases [9]. Then, the value of M over 1.0 means that the correct state has the largest R_i among the possible States i. Larger M means more stable selection of the correct State. The graph of Fig. 7 shows the relations between θ_1 and M for the eight cases. The tendency that smaller θ_1 produces smaller M clearly appears. From this observation, the threshold for using recognition-aid actions, T_{θ_1}, should be around 90 degrees.

B. Experiments using actual system

We have applied the proposed method under the situation that HRP2 holds a sweater by its right hand. In the experiments, the sweater was intentionally folded into both convex/concave shapes under the states of State 16, State 16' which is symmetrical state of State 16 and State 3. The parameter L_0 is set 15 cm in consideration of the size of the sweater.

In Fig. 8 and Fig. 9, we show convex and concave examples using the case of State 16. (A video showing these two handling processes is also provided as Supplementary Material.) Figs. 8(a) and 9(a) show original observations.

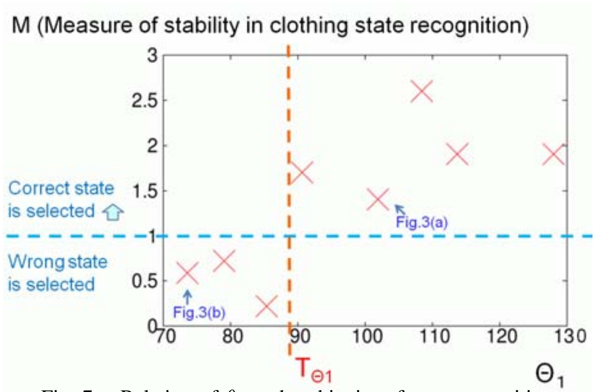


Fig. 7. Relation of θ_1 and ambiguity of state recognition

Table 1 Experimental results using humanoid

State V:Convex C:Concave	θ_1	Action Plan	Execution	Recog. Before	Recog. After
State 16(V)	83.4	○	○	×	○
State 16(C)	-81.7	○	○	○	○
State 16'(V)	74.4	○	○ –	○	○
State 16'(C)	-106.4	○	–	○	–
State 3(V)	84.8	○	×	○	–
State 3(C)	-91.4	○	○ –	○	○

In Figs. 8(b) and 9(b), detected $K0, K1$ and $K2$ from the observed data were shown by green and blue dots. The values (in degrees) of (θ_1,θ_2) were (83.4, -2.5) and (-81.7, 17.6) respectively. As a result, the hand coordinates for spreading-clothes action were determined as shown by lines in Figs. 8(b) and 9(b), respectively. Here, colored solid and dashed lines correspond to the lines in Fig. 6(a). Figs. 8 (c) and 9 (c) shows the observation after the spreading-clothes actions were executed by the left hand of HRP2. By applying our model-based recognition method[9] to these observations, the clothing states were successfully recognized as shown by the clothing models superposed on the 3D data in Figs. 8(d) and 9(d). After the left hand had been taken away from the sweater, the clothing was observed again as shown in Figs. 8(e) and 9(e). This newly observed 3D data was recognized by using the previous recognized shape as its initial 3D shape, which is superposed on the data in Figs. 8(e) and 9(e). By deforming this initial shape to fit the newly observed data, the current shape was obtained as shown by the superposed clothing model in Figs. 8(f) and 9(f). Based on this recognition result, action plans for the task of grasping one shoulder were obtained as shown by lines in Figs. 8(f) and 9(f) [10]. Figs. 8(g) and 9(g) show the handling processes while executing these grasping actions. In the case of Fig. 8, the target shoulder was successfully held up. Unfortunately, the holding-up action failed in the case of Fig. 9 because the calculated position for holding was about 1 cm outside from the actual clothing hem. Fig.9(g) shows this failure.

The experimental results of all six cases are summed up in Table 1. In all cases, "recognition-aid" actions were well calculated as shown in the column of "Action plan". The forth column, "Execution", show the results of actual execution of these action plans. Here, – and ○– mean that we failed or partially failed in generating arm-joint angles which realize the calculated hand position and pose, respectively. Execution results besides the case of – are shown in Fig. 10. Only in the case of State 3(V), the calculated action failed in spreading the sweater by pushing its side so as to fold it more as shown in the second row of Fig. 10. The reason of this is that the side part was so deeply folded. Such deep folding can be detected while fitting a deformable surface in the "ambiguity check" process. Then, this problem should be solved by combining "rotating actions". In the four cases where the recognition-aid actions had been successfully executed, the success rate of the recognition was improved from 3/4 to 4/4 after the "recognition-aid" actions as shown in the fifth and the sixth columns in Table 1.

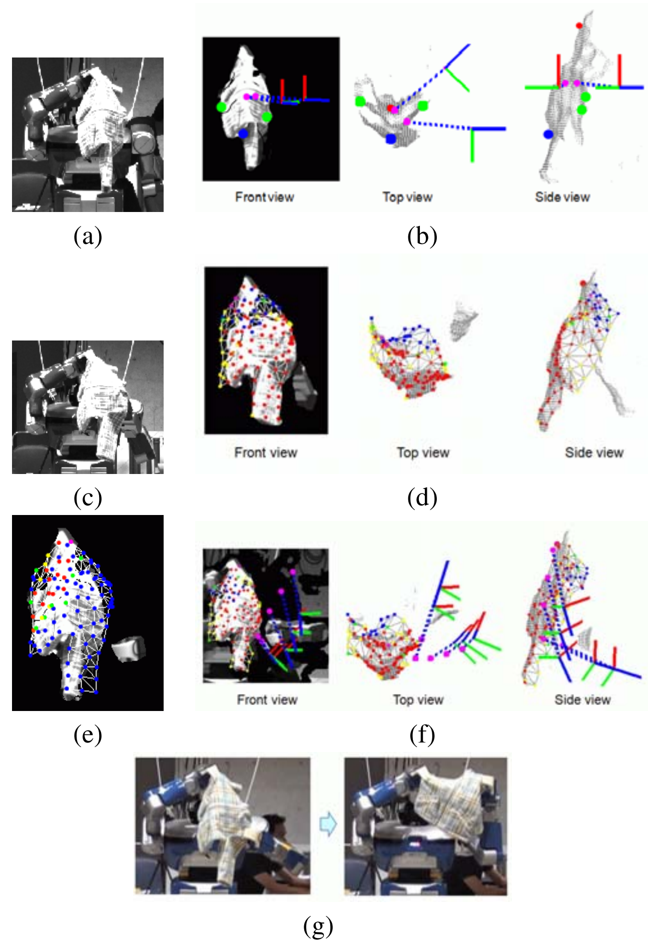


Fig. 8. Experimental example of convex case: (a) original observation; (b) calculated spreading action; (c) observation after the action; (d) recognition results from the observation, (e) observation after removing the left hand; (f) calculated action for picking up a shoulder; (g) picking up action.

V. Conclusions

We proposed a strategy of visual recognition in cooperation with actions and presented an implementation of spreading-clothes actions as one of the recognition-aid actions. Through the experiments using actual humanoid, the followings were shown:

1) Proper plans for carrying out spreading-clothes actions can be automatically calculated from the first visual information.

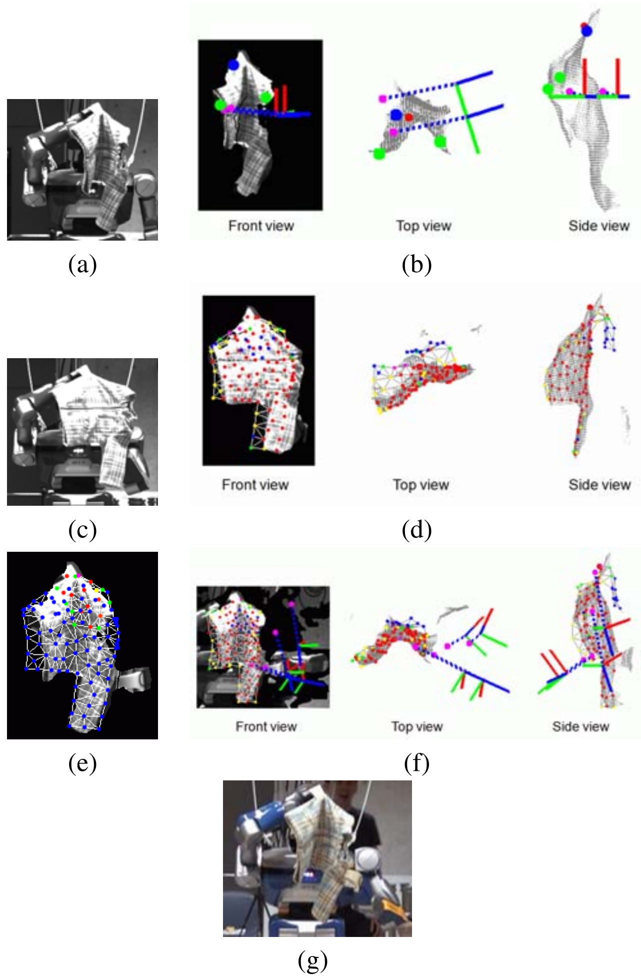


Fig. 9. Experimental example of concave case: (a) original observation; (b) calculated spreading-action; (c) observation after the action; (d) recognition results from the observation; (e) observation after removing the left hand; (f) calculated action for picking up a shoulder; (g) picking up action (failed).

2) The resultant actions can produce a new observation which is easy to recognize.

Although the number of experiments is not big, the high success rate of the experimental results encourages us to carry forward the proposed strategy. The computational time for calculation of proper recognition-aid actions and recognition after the action is about 5 and 10-20 sec respectively (Intel Xeon 3.0GHz dual core). Most of the time is used for the process of deforming a clothing model. Although acceleration of the computational time is desired for real application, it is allowable since this recognition process is done only at first time in a sequence of handling an item of clothing.

One of our future subjects is to combine several recognition-aid actions for realizing more smooth handling.

Acknowledgment

This work was supported by a Grant-in-Aid for Scientific Research, KAKENHI (19300066). We are thankful to Dr. K. Yokoi and Dr. K. Sakaue for their support in this research.

References

[1] 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.
[2] 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.
[3] 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.
[4] M. Kaneko and M. Kakikura: “Plannning 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.
[5] 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.
[6] 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.
[7] 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.
[8] 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.
[9] Y. Kita, T. Ueshiba, E. S. Neo and N. Kita: “Clothes state recoginition using 3D observed data”, In *Proc. of International Conference on Robotics and Automation*, pp. 1220–1225, 2009.
[10] 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.
[11] E. S. Neo, K. Yokoi, S. Kajita and K. Tanie:: “Whole-Body Motion Generation Integrating Operator’s Intention and Robot’s Autonomy in Controlling Humanoid Robots”, *IEEE Trans. on Robotics*, Vol. 23, No. 4, pp. 763–775, 2007.
[12] 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.

Fig. 10. Results of executing spreading actions in the case of State 16’(C), State 3(V) and State 3(C)