2009 IEEE International Conference on Robotics and Automation
Kobe International Conference Center
Kobe, Japan, May 12-17, 2009

Clothes state recognition using 3D observed data

Yasuyo Kita*, Toshio Ueshiba*, Ee Sian Neo** and Nobuyuki Kita**
*Information Technology Institute **Intelligent Systems Institute
National Institute of Advanced Industrial Science and Technology (AIST)
{y.kita, t.ueshiba, rio.neo, n.kita}@aist.go.jp

***Abstract*— In this paper, we propose a deformable-model-driven method to recognize the state of hanging clothes using three-dimensional (3D) observed data. For the task to pick up a specific part of the clothes, it is indispensable to obtain the 3D position and posture of the part. In order to robustly obtain such information from 3D observed data of the clothes, we take a deformable-model-driven approach[4], that recognizes the clothes state by comparing the observed data with candidate shapes which are predicted in advance. To carry out this approach despite large shape variation of the clothes, we propose a two-staged method. First, small number of representative 3D shapes are calculated through physical simulations of hanging the clothes. Then, after observing clothes, each representative shape is deformed so as to fit the observed 3D data better. The consistency between the adjusted shapes and the observed data is checked to select the correct state. Experimental results using actual observations have shown the good prospect of the proposed method.**

I. Introduction

As home and welfare robots are expected to take an important role in an aged society, ability of handling daily objects is strongly required for robots. Clothes are one of such typical articles of daily use. Among techniques necessary for realizing automatic clothes handling, it is essential but still challenging to visually recognize largely deformed objects. Although handling of string-type soft objects, such as ropes and electric lines, has been studied[1][2], in case of dealing with clothes, complex self-occlusion makes it very difficult to recognize the clothes state. Here, by the term "state", we mean recognition of not only geometrical shape but also where each part of the clothes is in the shape.

Kaneko et. al[3] proposed a method which recognizes the clothes state by comparing the contour features (e. g. curvature, length-ratio) of an observed appearance with ones of model appearances under the situation that the clothes is held at two points. However, the detailed contour features are difficult to robustly extract from real observations and are very sensitive to a slight deformation of clothes. Additionally, its learning processes to obtain the model appearances from actual observations are troublesome.

In [4], a deformable-model-driven method which recognizes the state of clothes held at a point has been proposed. The method predicts possible appearances using a deformable model of the clothes and selects one which fits the observed appearance the best. Later, [5] proposed a method using these results to obtain 3D information for actual handling, such as the position and posture of the part to grasp next. Although these results were encouraging.

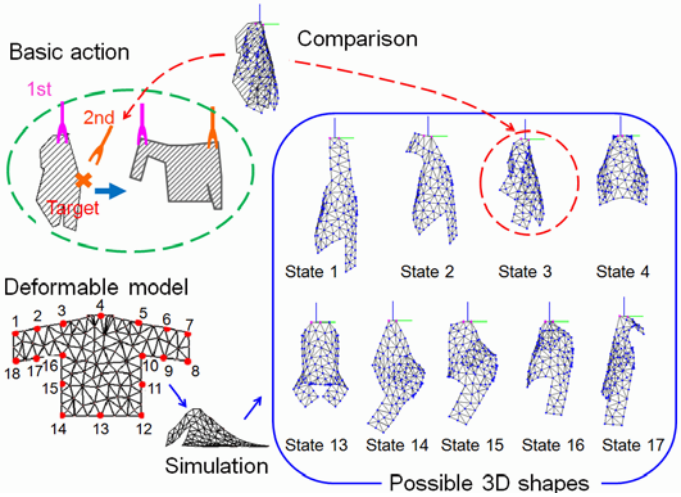

Fig. 1. Basic action and our model-driven strategy

there are some drawbacks to use the methods in practical scenes. First, some restrictions on the color and texture of the clothes are necessary to robustly extract clothes region from its background. Another point is that separate processes are required to obtain 3D information for handling after the method estimates clothes state from 2D observation.

To overcome such weaknesses, we can use dense 3D information obtained by a trinocular stereo camera system[6]. If we use such 3D data, extraction of clothes region should become much easier by taking only 3D data around the holding position. Then, we get a 3D shape of the visible side of the clothes. In order to robustly recognize the position of any part from this information, similarly to [4], we take a model-driven way that compares the 3D observation with 3D possible shapes to find the most consistent states. Here, a key issue is how we prepare the 3D possible shapes which can cover wide range of shape variation of the clothes. For the purpose, we propose a two-staged method. First, small number of representative 3D shapes are calculated by simulating physical deformation of the clothes when it is held by a hand. Then, after observing clothes, each representative shape is deformed so as to absorb shape difference between the predicted shape and the observed shape. After this adjustment process, the adjusted shape which shows the best consistence with the observed data is selected as the state of the observation. From now, in Section 2, a whole flow of our model-driven method is explained. In Section3, a

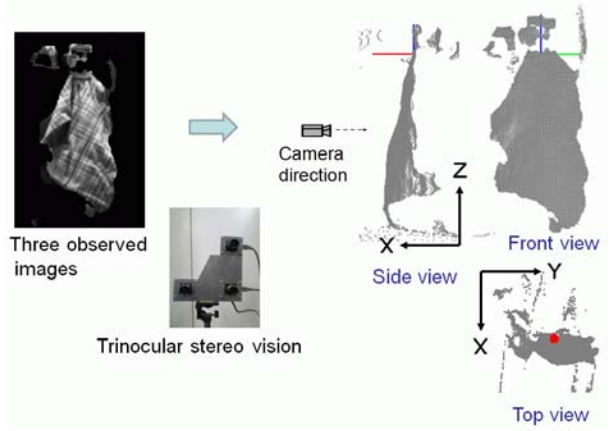

Fig. 2. Three dimensional data obtained from trinocular stereo vision system (The view direction of the camera system is $-X$)

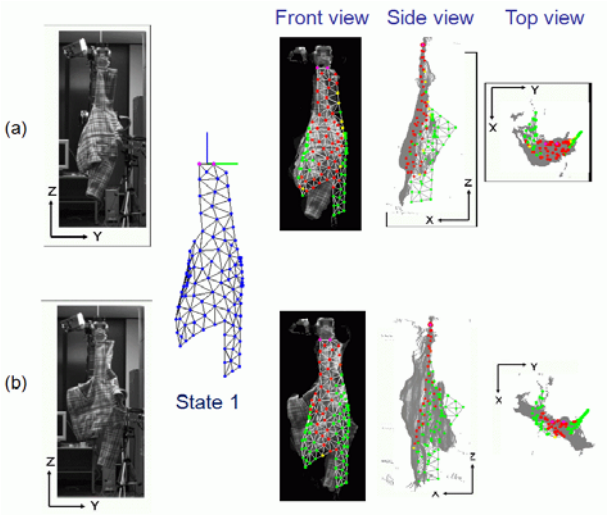

Fig. 3. Shape variation of the clothes under the same hanging condition

method of adjusting a 3D predicted shape to 3D observed data , that is the main contribution of this paper, is explained. Finally, in Section 4, experimental results using actual data are shown to discuss about the effect of the proposed method for recognizing clothes state.

II. Model-driven clothes state recognition

A. Fundamental idea

One of important and basic action for handling clothes is grasping a specific part of clothes hung in the air as shown in Fig 1. By iterating this action by two hands, clothes can be held at any goal state. For carrying out this basic action, the 3D position and posture of the target part are indispensable information. Fig. 2 shows an example of 3D observed points of clothes obtained from a real-time trinocular stereo vision system[6]. Even though we get such dense 3D information, it is almost impossible to recognize how the clothes are folded and where is each part in the observed data in a bottom-up way. Therefore, we choose a model-driven strategy based on the assumption that a simple knowledge about the target clothes, such as style of the target clothes (e.g. pullover, trousers) and its approximate sizes and softness, is known in advance.

By simulating physical deformation of the target clothes based on these information, possible 3D shapes of the clothes when it is hung at a point are obtained as shown in Fig. 1, At the current, we have done this simulation using cloth function of Maya 4.5 [7]. To make the problem simple, we assume that the front and back sides of the clothes are not separated and no thickness is given to the model. Based on this assumption, the model becomes a surface which deforms three-dimensionally. Here, we classify the clothes states according to at which position the clothes is held as shown as "State 1", "State 2" and so on in Fig. 1. We think this classification is natural since the grasping position is only one condition which explicitly determines the shape. Of course, the possible shapes obtained by the above simulation is just one of the most probable shapes since actually the clothes shape have variation depending on indefinite conditions such as subtle difference in the trail of hanging processes. It is quite important how to deal with such practically unpredictable shape variation.

B. Key issue

Fig. 3(a) and (b) are two observations when the same pullover was hung at the same position at different times. Despite of giving the same condition, the 3D shapes are fairly different from each other. The blue mesh model in the second column shows the shape our system predicted for the corresponding state (State 1). Predicted shapes in Fig 1 are obtained by physical simulation in advance with the condition that clothes are held by a vertical planar grip. Then, after the observation, predicted shapes are placed at the holding position so that the normal direction of the clothes at the position coincide with that of the actual grip plane. Since there are two possible postures for this condition, the one closer to the observed data is selected. The third to the fifth columns of Fig. 3 show the predecited shape placed on the observed data in this way. In this example, the grip is set so that its normal has the same direction as the camera view direction. Since the both 3D observed data are rather convex toward the camera system, the predicted shape is placed just in the same way in the both cases.

The color of vertices of the model illustrates the closeness from the observed data. Concretely red vertices are ones less than 25mm from the closest observed point, green vertices are the others. As seen in the figure, the predicted shape cannot tell the correct position in many parts.

However, increasing the number of representative shapes to cover all such shape variation is not practical from the viewpoint of computational burden. Instead of that, we propose to deform the representative shapes so that they are adjusted to each observed situation. In the following section, we explain our adjustment method of the initial predicted shape to the observed data.

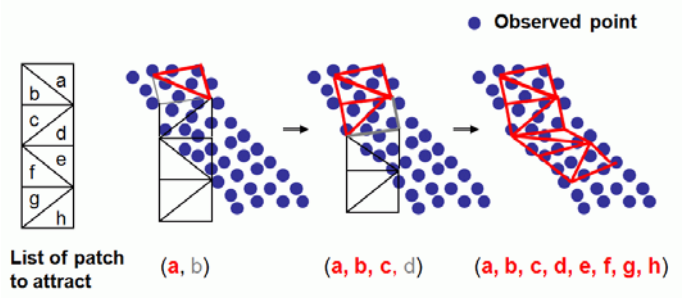

Fig. 4. Model deformation using list of patch to attract

III. Adjustment of Predicted Shape

A. Key idea

For the adjustment, we use a deformable model and deform it by exerting attractive forces towards the observed data. Some methods related to this matter have been proposed for the purpose of tracking of deformation of non-rigid objects. Yamamoto et al.[8] proposed a method of tracking deformation of paper and balloon with a deformable mesh model using high speed laser range finder. At each stage of tracking, vertices of the deformable model are attracted to the closest observed data under the constraint that all the part of the model is set close to the observed 3D data. Pilet et al. [9] has tracked cloth deformation (ex. a part of T-shirts or sail of yacht) with a deformable mesh model from time-sequential 2D images under the constraint that the object have characteristic textures.

Instead of these constraints, we utilize more natural one for the case of handling task. When we think the situation that a robot handles clothes, the position and posture of the grasped part can be actively controlled and these information is known. That means that the deformable model of the predicted shape can be placed on 3D observed data so as that the model is coincident with the data at the holding position. Therefore, at least in the vicinity of the part, we can assume that the deformable model is already close to the observed data, so that we can use Euclidean distance as a key for finding correct correspondences between model vertices and the data point. Based on this idea, we deform the deformable model gradually from the vicinity of the holding position towards the further parts as conceptually shown in Fig 4. It can be regarded as analogy of putting up a wallpaper on the wall from one point to its circumferences gradually.

B. Implementation

Our deformable clothes model is represented by triangular patches and their vertices. One reason to use triangular patches is good compatibility with Maya 4.5 [7] which we use for physical clothes simulation. The idea described in the previous subsection can be implemented by using "list of patch to attract". Only for the vertices of the patches on this list, the closest observed 3D data points are searched for and if there, the attractive force to the closest point is exerted. At the start, a deformable model of the predicted shape is placed according to the information of actual holding position and posture. Only the patch at the holding position is included into the "list of patch to attract". While exerting forces to the patches on the list, the distance of each patch of the list from the closest observed point is checked to judge if the patch already fits the observed data or not. When a patch gets on the observed data, that is, it gets sufficiently close to the observed data, its adjacent patches are newly added to the "list of patch to attract" if the added patch is visible from the view direction. Whether a patch is "visible or not" can be judged in a hidden surface removal manner. These processes are iterated until the model deformation is converged.

The deformation of the model is calculated using the following analogical forces in a similar way to general deformable model methods (ex. snakes[10]).

As internal forces to preserve the clothes-like shape, the following forces are exerted for all vertices:

1) Forces to keep distance between neighboring vertices
2) Forces to keep distance between vertices connecting via one neighboring vertex

The first and the second forces respectively correspond to the elasticity and flexural rigidity of the clothes.

As external forces,

1) Gravitational forces to all vertices,
2) Attractive forces to the closest observed points only for the vertices of the patches on the "list of patch to attract".

The movement of the 3D coordinates of the vertices, $\mathbf{x}(\mathbf{y}, \mathbf{z})$ is calculated by solving the equation of

$$\mathbf{A_t}\mathbf{x}_{t+1} + \mathbf{f}_{x,t} = -\gamma\ (\mathbf{x}_{t+1} - \mathbf{x}_t)$$

in a successive approximation way (similarly done about $\mathbf{y}$, $\mathbf{z}$) . Where, $\mathbf{x}_t$ is a vector of x coordinates of all vertices at time t; $\mathbf{A_t}$ and $\mathbf{f}_{x,t}$ are the square matrix determined by the internal forces and the vector determined by the external forces respectively; γ is a parameter to decide the step width of successive approximation.

IV. Experiments

We applied the proposed method to actual 3D data observed by the trinocular stereo vision system[6]. The main parameters for the experiments are five: two parameters of the elasticity and flexural rigidity of the deformable model, two parameters to determine the gravitational force and the attractive force to the observed data and the distance threshold to judge if the vertex is on the observed data or not. These parameters are empirically determined and fixed in all the experiments. The deformation process is stopped if any new patch is not added to "list of patch to attract" during enough number of iterations.

The same clothes was observed while changing holding points on the clothes which are corresponding to 9 states in Fig. 1. States 5 $\sim$ 12 were omitted since they are symmetrical of any of States 1 $\sim$ 3 and States 14 $\sim$ 18. State 18 was also omitted since it is fairly close to State 1. At every hanging state, the clothes was observed three times while re-holding it at the same point to obtain the data with shape variation. So totally, 27 shapes were observed. Here, the clothes were hung so as to be folded rather convexly toward the camera to avoid similar shapes which are symmetrical to each other with respect to the grip plane.

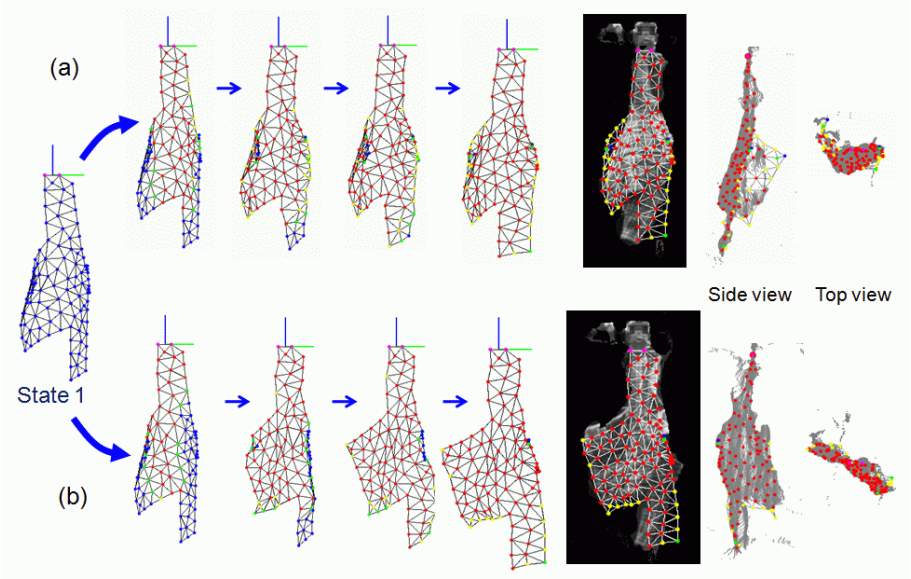

Fig. 5. Absorption of shape variation 1

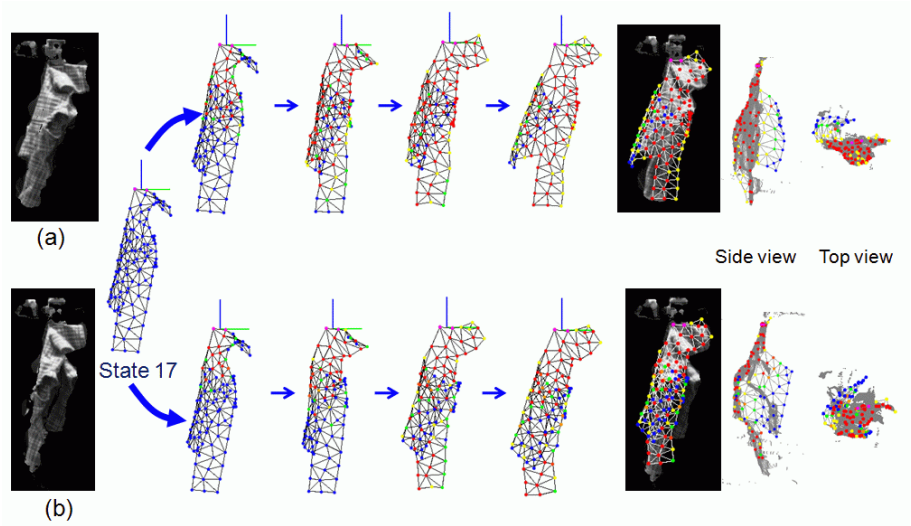

Fig. 6. Absorption of shape variation 2

A. *Absorption of shape variation*

Fig. 5 shows examples of deformation processes and the final deformed shape for the cases of Fig. 3. At each model, blue vertices show vertices which are not on the "list of patch to attract", green ones are vertices on the list, red to yellow ones are vertices on the list which are enough close to the observed data. More red represents closer to the data. In the both cases, the resultant shapes are well consistent with the actual 3D state, In other words, the 3D position and posture of each part (ex. shoulder, armpit and so on) of the clothes model are fairly close to the actual ones.

For all 27 data, similar results were obtained. Fig. 6 shows examples of the case corresponding to State 17, where clothes are largely self-occluded and the amount of observed 3D data is relatively small. Even in the cases, the proposed method realized good adjustment and gave good estimation of the 3D information of each part of the clothes. It is one of strong points of a model-driven approach that the 3D information of even unobserved part can be inferred.

B. *Effect for state recognition*

In the previous subsection, we showed how the proposed adjustment method can cover the shape variation in the case that we give the correct holding position. In this subsection, we consider the situation that we do not know at where the clothes is held and use the method under the model-driven recognition strategy illustrated in Fig. 1. Here, it is important to understand what happens when a wrong predicted shape

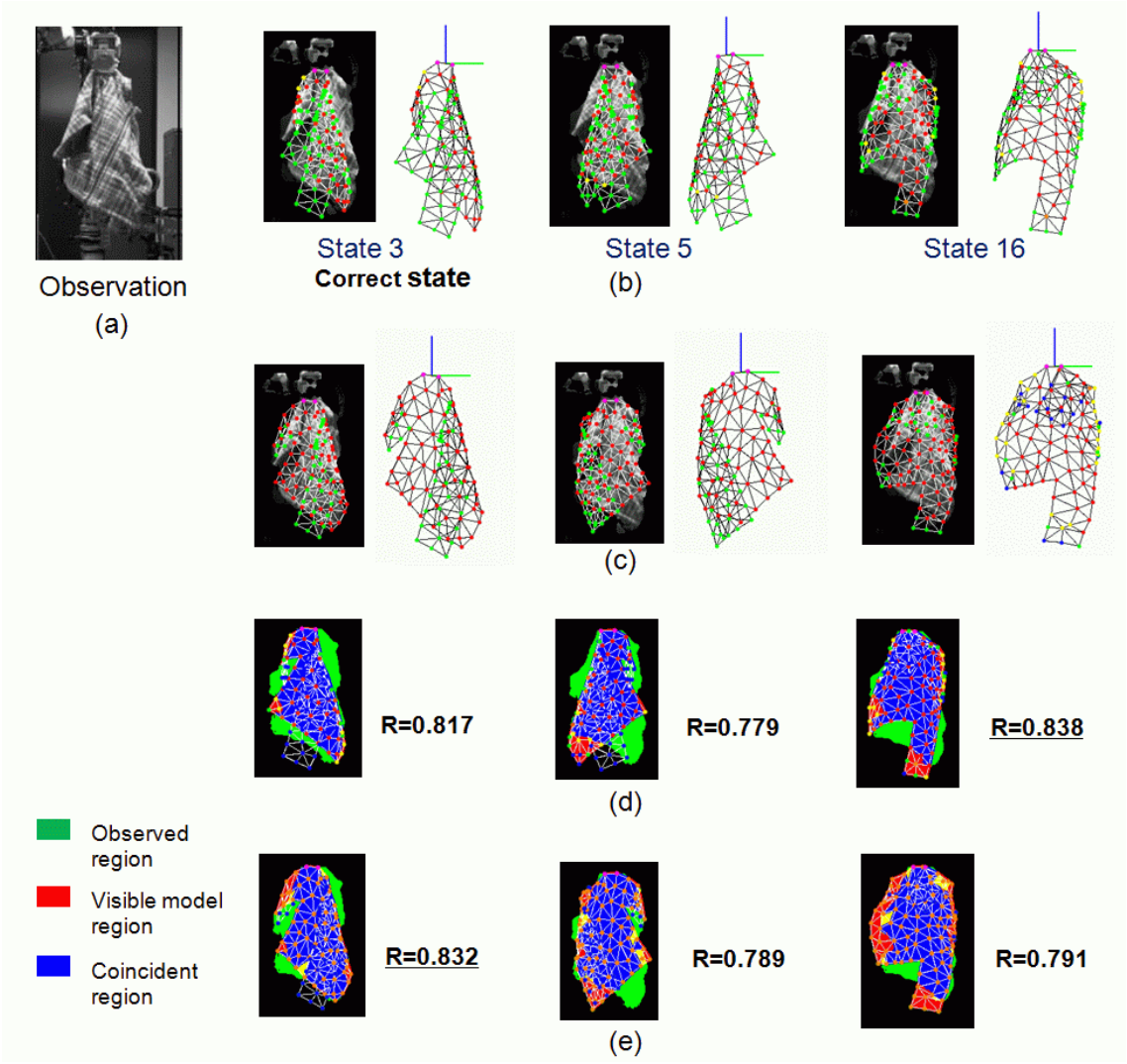

Fig. 7. Effect for state recognition: (a) observed image; (b) predicted shape obtained from physical simulation; (c) modified shape after the proposed adjustment; (d) overlap ratio after two-dimensional adjustment[5]; (e) overlap ratio after the proposed adjustment.

is given as the initial shape for the adjustment process.

Fig. 7 shows a case where the shapes hung at some different holding points have 3D observed data close to each others. The clothes observed in Fig. 7(a) is actually held at the left shoulder (State 3). However, the shape hung at the other shoulder (State 5), and the shape hung at one of the armpits (State 16) give similar 3D observations. Predicted shapes for the cases obtained through the simulation of physical deformation are shown in Fig. 7(b). Actually, if we count the vertices of the predicted shapes which are enough close to the observed 3D points (red vertices in the Fig. 7(b)), a wrong state, State 16, has the largest number. Fig. 7(c) shows adjusted shapes after applying the proposed method to these predicted shapes. All predicted shapes are modified so as to better fit to the observed 3D data. However, in the correct case, after the adjustment, the adjusted model gets overlapped on the observed data with few excess and deficiency. On the other hand, in the two wrong cases, some parts of observed data do not have corresponding model parts, and vice versa, some model parts do not have corresponding observed data.

To evaluate this point quantitatively, the following criteria is calculated: the overlap ratio, R,

$$R = \left(\frac{\textbf{coincident area}}{\textbf{observed area}} + \frac{\textbf{coincident area}}{\textbf{visible model surface area}} \right) / 2$$

Here "area"s are calculated on the 2D image plane of one of stereo cameras after projecting the 3D observed data and the 3D adjusted shape on the plane respectively.

Fig. 7(e) shows the results after the adjustment of the proposed method. R values are 0.832, 0.789 and 0.791 respectively for State 3, State 5, and State 16. The consistency is well presented by the values.

For the comparison with the method using 2D observed data[5], the results of overlap ratios after two-dimensional adjustment were shown in Fig. 7(d). Concretely, each model appearance is horizontally shrunk or extended so as to have the same width as the observed region on the image plane.

Table 1 State estimation results

State No.	1	2	3	4	13	14	15	16	17	total
Candidate Number	2	8	5	1	7	6	12	10	10	-
1st	3	1	3	3	3	3	0	3	3	22/27
up to 3rd	3	3	3	3	3	3	1	3	3	25/27

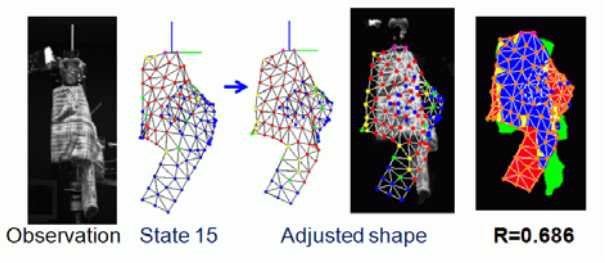

Fig. 8. Failure example in state estimation

The adjustment did not work well in this case. As a result, a wrong state, State 16, has the highest R.

Using this evaluation value, we conducted experiments to recognize clothes state. The results for all 27 cases are summed up in Table 1. At each observation, before applying the adjustment procedures, states which have vertical length consistent to the observed data are selected as possible states. The number of remained possible states after this length check are different depending on observations and shown in the second row in Table 1. Then, for every possible state, the overlapped ratio R was calculated after each predicted shape was deformed by the proposed method. The third and forth rows show the resultant order of the correct shape when sorting the candidates in ascending order of R. For 22 data, the correct state had the highest score, and for three of the remaining data, the correct states came up to the third position. You can see detailed numbers for each state in Table 1. It is clear that failures happened intensively in the case of two states, especially State 15. In most of the failure cases, the values of R of the correct state were low. Fig 8 shows the adjusted model shape of one of the failure cases. The adjusted shape still far from the actual state around the hanging sleeve. Actually, a slight deviation between the model and observed data around the hanging sleeve are also shown in the case of Figs. 5, 6. Accumulation of errors can be one reason for this phenomena since the part is the furthest from the holding position. However, we infer that the biggest reason is the accuracy of the initial shape predicted by physical simulation around the part. The foldings occurred around shoulders generally look more complex than the simulation we used. The improvement of this simulation process should lead better results.

V. Conclusions

We proposed a deformable-model-driven method for clothes state recognition, which consists of two stages: shape prediction by simulating physical deformation of the clothes and adjustment of the predicted shape to 3D observed data. The main contribution of this paper is proposal of an adjustment method for the second stage. Owing to successful absorption of shape variation of clothes, state of clothes were well recognized in the experiments using actual data observed by our trinocular stereo vision system. It is one of future subjects to find the best number of representative shapes and a way to select good representative shapes.

The computational time is 5-20 sec (Intel Xeon 3.0GHz dual core) for each adjustment process. It is better to accelerate the time for real application, but is allowable since this recognition process is done only at the first time. Once we understand the clothes state, we can just track the clothes deformation based on the known state.

After the adjustment of the second stage, most parts of the adjusted model show almost correct 3D position of the corresponding part of the actual clothes. That means, once we get the results, the 3D position and posture of a specific target part can be known if the part appears on the observed data. Even in the case that the target part is unobserved, the position of the part on the model can give the information of where and from which direction the clothes should be observed next to obtain its actual 3D information. To affirm these characteristics, we plan to conduct experiments of actual handling by a humanoid.

Acknowledgment

This work was supported by Grant-in-Aid for Scientific Research,KAKENHI(19300066). We are thankful to Dr. K. Yokoi and Dr. T. Nagami for their support to this research and are also grateful to Mr. T. Matsukawa for data acquisition and so on.

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] 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.

[4] Y. Kita and N. Kita: "A model-driven method of estimating the state of clothes for manipulating it", In *Proc. of 6th Workshop on Applications of Computer Vision*, pp.63–69, 2002.

[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] 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.

[7] D. A. D. GOULD: *"Complete Maya Programming"*, Morgan Kaufmann Pub, 2004.

[8] M. Yamamoto, P. Boulanger, J. -A. Beraldin and M. Rioux: "Direct estimation of range flow on deformable shape from a video rate range camera", *IEEE Trans. on Pattern Analysis and Machine Intelligence*, Vol. 15, No.1, pp. 82–89, 1993.

[9] J. Pilet, V. Lepetit and P. Fua: "Fast Non-rigid surface detection, registration andrealistic augmentation", *International Journal of Computer Vision*, Vol 76, No. 2, pp. 109–122, 2008.

[10] M. Kass, A. Witkin, and D. Terzopoulos: "Snakes: active contour models", *International Journal of Computer Vision*, Vol. 1, No. 4, pp. 321–331, 1988.