A model-driven method of estimating the state of clothes for manipulating it

Yasuyo KITA Nobuyuki KITA
Research Institute of Intelligent Systems,
National Institute of Advanced Industrial Science and Technology (AIST)
AIST Tsukuba Central 2, 1-1-1 Umezono, Tsukuba, Japan 305-8568
y.kita{n.kita}@aist.go.jp

Abstract

Aiming at manipulating clothes, a model-driven method of estimating the state of hanging clothes is proposed. We suppose a system consisting of two manipulators and a camera. The task considered in this paper is to hold a pullover at its two shoulders by two manipulators respectively, as a first step for folding it. The proposed method estimates the state of the clothes held by one manipulator in a model-driven way and indicates the position to be held next by the other manipulator. First, the possible appearances of the pullover when it is held at one point are roughly predicted. Using discriminative features of the predicted appearances, the possible states for the observed appearance are selected. Each appearance of the possible state is partially deformed so as to get close to the observed appearance. The state whose appearance succesfully approaches closest to the observed appearance is selected as the final decision. The point to be held next is determined according to the state. The results of preliminary experiments using actual images have shown the good potential of the proposed method.

1 Introduction

The handling of soft objects is attracting increasing attention in the robotics field. However, it is still challenging, though essential, to visually understand the state of largely deformed objects. Although rope handling has been studied [1], in case of dealing with clothes, complex self-occlusion makes it very difficult to understand the state. In the field of computer graphics, sophisticated models for animating cloth deformation have been developed [2]. However, research on automatic recognition of deformed clothes has just started [3].

One approach to the recognition of deformed clothes is to obtain detailed three-dimensional information by elaborate visual sensors, however, we aim to extract only the information necessary for a given task from images taken by a simple camera system, as human beings do daily. When a human recognizes deformed clothes, expectations for the object state seem to play an important role. Inferring from humans' expectation ability, the expectation need not necessarily be accurate, suggesting that even an approximate prediction could powerfully help understand processes.

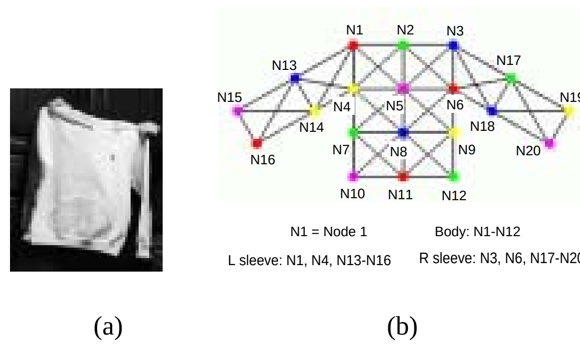


Figure 1. Pullover used for our experiments: (a) goal state held at both shoulders; (b) model consisting of 20 nodes classified into three parts, a body and two sleeves

In this paper, we propose a method which recognizes the state of deformed clothes in a model-driven way using the appearances predicted from simulation of the deformation as model appearances. As a specific example, understanding the state of a pullover held at one point is considered. The situation we assume is as follows. Our goal is to open a pullover by holding it at the two shoulders by two manipulators as shown in Fig. 1a. Currently, we assume one camera for the observation, which is calibrated relative to the manipulators. The initial state assumes that the pullover is held at a point close to its hem by one manipulator. We require to estimate the state of the pullover and to indicate the position to be held next by the second manipulator.

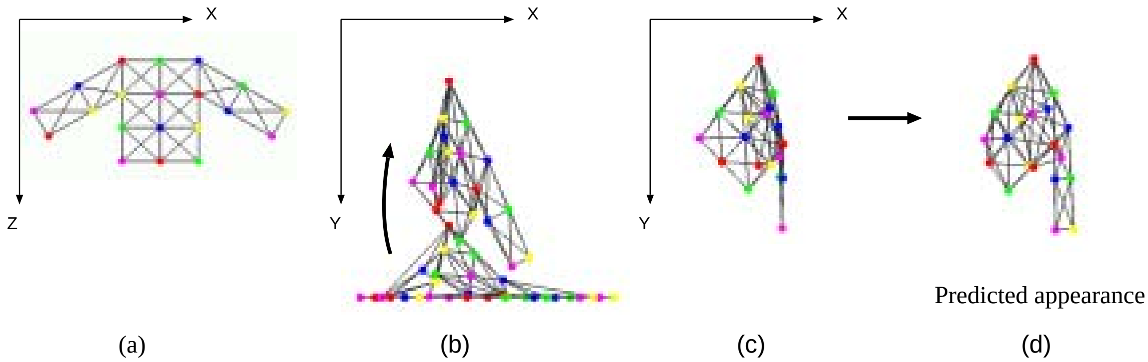


Figure 2. Simulation of pullover deformation: (a) original state (spread on a floor); (b) process of bringing up; (c) hanging at a point; (d) appearance from assumed viewing direction

2 State prediction

A coarse model of the pullover is built first based on three sizes, the width and length of the trunk and the length of the sleeves, as shown in Fig. 1b. We assume the front and back sides of the pullover are not separated and no thickness is given to the model. The model consists of 20 nodes which are connected to each other by springs as illustrated with the lines. Three different types of springs are used:

K_1-type springs are set between a node and its 4-neighbored node.

K_2-type springs are set between a node and its 8-neighbored but not 4-neighbored node.

K_3-type springs are set between a node and its next node but one in the 4-neighbored directions. The springs are set only inside the same part, that is the trunk, left sleeve and right sleeve. (e.g. N1-N3, N1-N7, N3-N19)

K_3-type springs are introduced to prevent folding inside the same part and have a weaker spring constant than the others.

Deformed shapes when the pullover is held at one node are automatically calculated as follows. 3D left-hand coordinates are defined so that the positive direction of the Y axis corresponds to the gravity direction. First, the model is spread on a horizontal plane, which is parallel to the X-Z plane (Fig. 2a). The gravitational forces are constantly exerted all nodes. When a selected node is picked up and vertically moved up (in the negative direction of the Y axis), the deformation of the pullover is simulated as shown in (Fig. 2b). Figure 2c shows the converged state at a fixed holding position. Holding at just a point produces uncertainty regarding the rotation around the Y axis. The actual holding we assume is not by a point but by a flat grip with a small area, so the direction of the grip specifies the rotation. By assuming the pullover is observed from the direction perpendicular to the grip plane, the viewing direction is fixed so that it coincides with the normal of the plane around the holding point. Figure 2d shows the appearance from the direction. The appearance is stored two-dimensionally to be directly compared with the region of the pullover in observed images. This appearance model still has the connection information among the nodes, but springs are no longer considered.

Figure 3 shows all the predictions, from State 1 to State 20, named after the number of the node held. For the calculation, the spring constants per unit length are manually selected so that the results show similar stiffness to the pullover and are set to 20000, 2000 and 200 for K_1-type, K_2-type and K_3-type springs respectively.

3 State Analysis

From the predicted model appearances, the states are classified into three categories:
Class A: States already held at one of the shoulders, States 1 and 3
Class B: States where it is easy to point out one of the shoulders, that is, the states having at least one non-concave and non-overlapped shoulder, States 2, 7, 9, 13, 15, 17 and 19
Class C: Others, States 4, 5, 6, 8, 10, 11, 12, 14, 16, 18 and 20

At each state, the following positions should be indicated for the purpose of approaching the goal state, that is, the state where the two shoulders are held by the two manipulators:
Class A ⇒ The position of the second shoulder
Class B ⇒ The position of one of the shoulders
Class C ⇒ The position to create Class B

From the observation of the predicted appearances, one

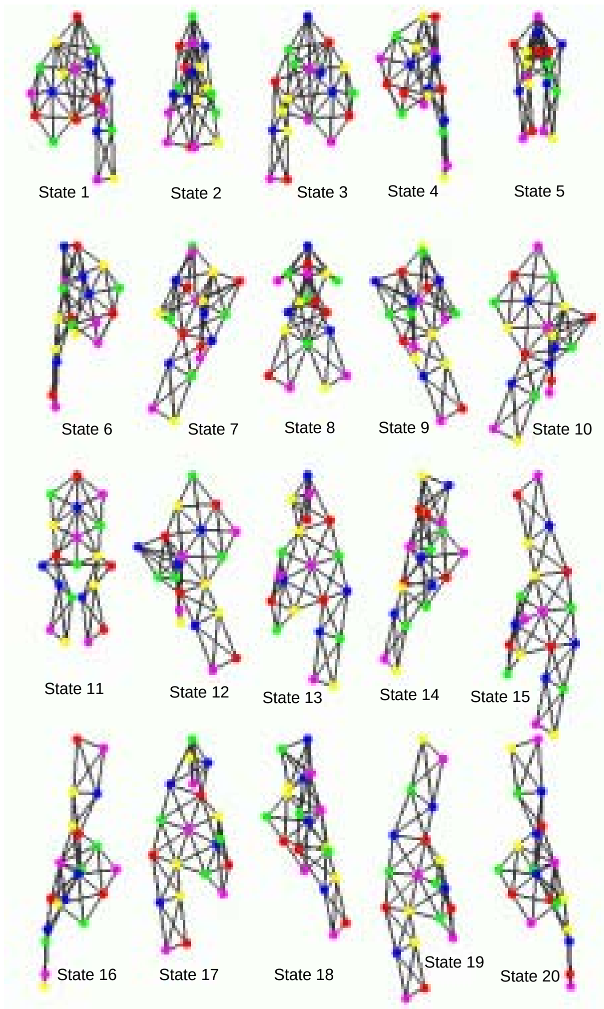


Figure 3. Predicted shapes when the pullover is held at a point

of the sleeve's tips always becomes the lowest point of the appearances and is easy to find out. Therefore, in the case of the third class, the lowest point of the observed region is indicated to guide to State 15 or 19.

After at most two iterations of the processes of this visual estimation and actual holding the point indicated, the system can get to State A, and next can hold the pullover at its two shoulders.

4 Estimation processes

Estimation processes consist of two stages: 1) selection of possible model states using simple attributes, and 2) decision of the best state based on how well its appearance corresponds with the observed region.

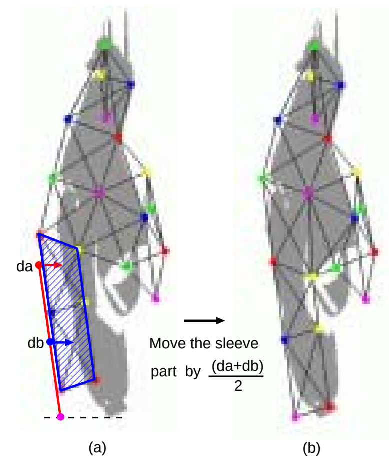


Figure 4. Modification of hanging sleeve part of model appearance

In the first stage, we use the following two features:
1. The lowest position in the appearance: $L(Lx, Ly)$
This is represented by the 2D coordinates relative to the holding position.
2. Number of hanging sleeves: N
This is the number of hanging sleeves in the vicinity of the lowest position. It should be two for States 2, 5, 8 and 11, and one for the other states.

Only model states which have the attribute values consistent with observation are selected for the second stages.

In the second stage, each model appearance is overlaid on the observed image so that its holding point coincides with that of the observed image. Because of coarse prediction and instability in appearance caused by subtle changes in boundary conditions, the model appearance may have some differences from the actual shape. Such differences can be corrected by referring the observed region, if the model appearance is sufficiently close to the region. Currently, correction of only hanging sleeve parts is implemented because of the following two reasons, here, "hanging sleeve" means the sleeve which is not held directly. First, the parts are less affected by unexpected folding because of their narrow width and tend to keep a similar shape to the prediction. The other reason is that this modification gives correct position of the shoulder which is the most important information for the current task.

The model state which shows the best overlap between its appearance and the observed region is selected as the final decision.

The flow of the actual processes is as follows.

I. Extraction of the features of the observed region

The position of holding the pullover on the image is first extracted according to the 3D position of the gripper of the manipulator. The clothes region is extracted as the largest homogeneous region under the point. The lowest point of the region is found as L_o. Concerning the second feature, N_o, the average horizontal width of the region in the upper vicinity of L_o is calculated. If the width is more than the sleeve width which is determined as one third of the length of the trunk, set $N_o = 2$, and $N_o = 1$ otherwise.

II. Selection of possible model states

Model state m_i ($m_i = 1 - 20$) is selected if all the conditions,

$$N_{m_i} = N_o$$
$$|Ly_o - Ly_{m_i}| < C_1$$
$$|L_o - L_{m_i}| < C_2$$

are satisfied. C_1 and C_2 are thresholds which should be determined according to how wide an area each model node represents. We set C_1 and C_2 to 60% and 100% of the longest distance between nodes.

III. Judgment through partial modification

Each of the model states which satisfy the above conditions is overlaid on the image so that the holding node of the appearance lies on the observed holding position. The hanging sleeve part is moved so that it overlays the observed area as described below.

1) The outer line of the hanging sleeve of the model state is moved vertically so that the lowest node has the same height as L_o, as shown by the thick gray line in Fig. 4a.

2) The closest edge point is searched horizontally from the two nodes beside the lowest one on the line. The signed distances to the edges from each node are set to da and db, which are shown with the arrows in Fig. 4a. If all the three conditions:

$$|da - db| < C3$$
$$|da| < C2$$
$$|db| < C2$$

are kept, the sleeve part is moved by $(da + db)/2$ as shown in Fig. 4b. Otherwise, the model state is rejected and the process goes back to 1) with the next possible model state. Here, C_3 determines the acceptable difference in the tangent of the sleeve lines.

3) After this partial deformation, the overlap ratio, R, which is the sum of the ratio of overlapped area to model appearance area and the ratio of overlapped area to observed area, is calculated. The state which has the highest value of R is decided as the final decision.

IV. Indication of point to be held next

When the selected state belongs to Class A or B, the edge point of the observed region that is the closest to Node 1 or 3 is indicated as the position to be held next. When a state in Class C is selected, the lowest position of the observed region is pointed out as the position to be held next. In the case, no state is selected, the method declares the reservation of judgment and indicates the lowest position of the observed region, since holding the position at least brings the state to State 15 or 19.

5 Verification processes

The appearance of the pullover after it has been held at the indicated position by the second manipulator can also be expected by the simulation of its deformation. The comparison between the expected appearance and the newly observed region assures the correctness of the current estimation of the state.

6 Experiments

We conducted preliminary experiments using images of a pullover held at a point by a human hand. In the experiments, we used 13 images taken while holding it at a point close to the nodes. Note that some nodes are fairly close to each other in the actual pullover, such as Nodes 15 and 16 or Nodes 19 and 20 at the cuffs. For such cases, the middle point between the nodes is selected. This is the reason why the number of observed images is different from that of the model nodes. The pullover model is built from the width and length of the trunk and the length of the sleeves which are given in advance. Parameters C_1 and C_2 are automatically decided according to the size. C_3 is set to allow 15 degree difference, and C_4 is set to 1.4 to allow 70% overlap as the lowest acceptable one.

Figure 5 shows an example of the results. The image coordinates of the holding position were manually given. Note that these values can be automatically extracted when the pullover is held by a manipulator whose position is actively controlled and is known. The largest bright region under this position was automatically extracted as the region of the pullover using a fixed intensity threshold. Figure 5b shows the extracted region from the original image of Fig. 5a. Figures 5c, d and e show the selected model appearances which give the first, second and third highest overlap ratio respectively before partial modification of the sleeve parts. Figure 5f shows the model state of the first one after moving its hanging sleeve part to fit the observed region. The appearances gave the highest overlap ratio, 1.64, and was selected as the final result. The point to be held next was detected as marked with the black cross.

Figure 6 shows another example. Figures 6b, c and d show the selected model appearances which give the first,

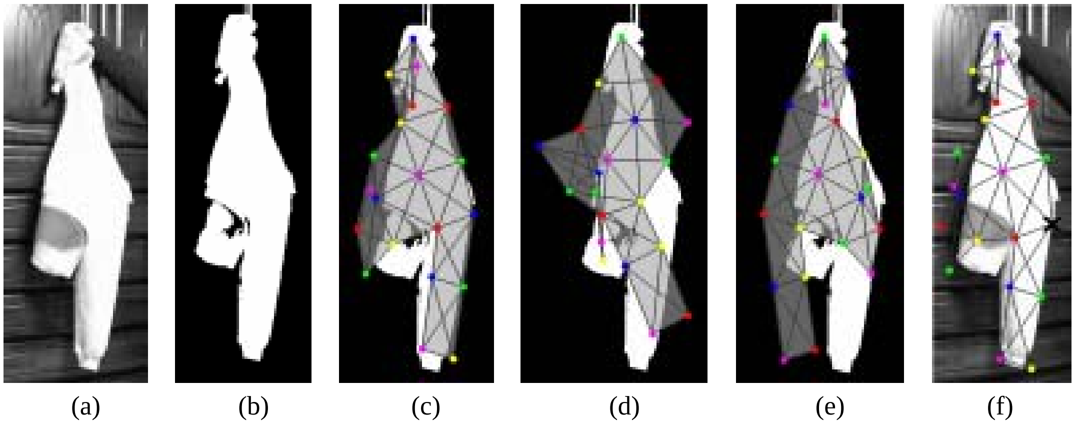


Figure 5. Experimental result 1: (a) original image; (b) observed clothes region; (c)(d)(e) selected candidates which give the first, second and third highest overlap ratio before partial modification (State 13, State 12, State 17); (f) indication of the position to be held next (the black cross).

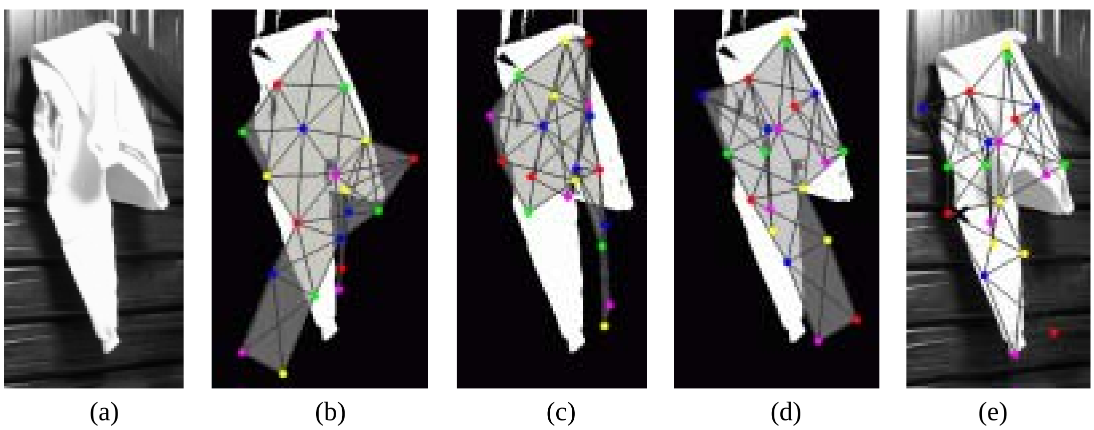


Figure 6. Experimental result 2: (a) original image; (b)(c)(d) selected candidates which give the first, second and third highest overlap ratio before partial modification (State 10, State 4, State 9); (f) indication of the position to be held next (the black cross).

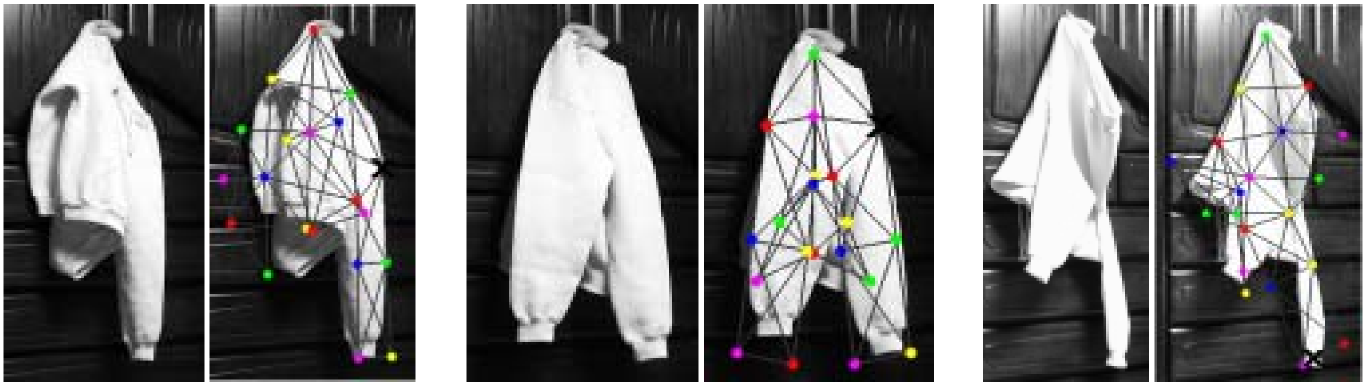
second and third highest overlap ratio respectively for the observed image (Fig. 6a) before partial modification of the sleeve parts. In this case, the third candidate gave the highest overlap ratio after the sleeve parts modification.

In total, for 9 of the 13 images, the correct state was selected and the correct position to be held next was indicated. Figure 7 shows other examples of Class A (State 1), Class B (State 2) and Class C (State 12). As noted in Section 3, in the case of Class C, the lowest point is selected.

For two of the remaining images, the system declared the reservation of judgment since all states were rejected. Figure 8a shows one example, Although the correct state, State 5, which is superimposed on the image, was selected with the highest overlap ratio, it was finally rejected since the ratio was less C_4. In the other case, the correct state was also selected and showed the highest overlap ratio before partial modification processes, but was rejected in the process because the observed sleeves had unexpected foldings. In both cases, the lowest position is indicated as the point to be held next as shown in Fig. 8a.

Only two of the 13 images were wrongly estimated. Figure 8b shows one of the results. State 1 was selected instead of the correct state, State 4. The main reason for this mistake was that the hanging sleeve region of the appearance was too narrowly expected.

We also conducted experiments using 16 additional images which were taken while holding the pullover at the middle points between two connected nodes. For 12 images, the correct state was selected and the correct position to be held next was indicated. Three cases gave wrong es-

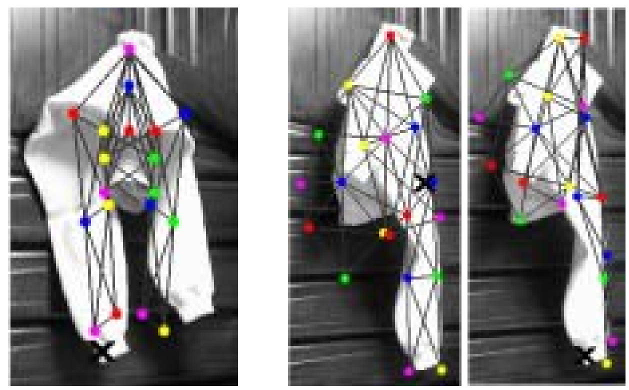


(a) (b) (c) (d) (e) (f)

Figure 7. Experimental result 3: (a) (c) (e) original images; (b) (d) (f) indications of the position to be held next (the black cross).

timations. For the remaining one image, all states were rejected due to the failure of extraction of the pullover region from the observed image. Figure 9a, b show examples of success and failure respectively. In the left image of Fig. 9a, the model appearance of a node close to the actual holding point deviates from the actual region because of the displacement in the holding position between the actual and the one used for the simulation. The right image of Fig. 9a shows its final result. By correcting the hanging sleeve part, the location of the shoulder was correctly indicated as shown by the black cross. In the failure case of Fig. 9b, State 1, superimposed on the left image, was selected instead of the correct state, State 12, superimposed on the right image. The main reason for this mistake was folding of the hanging sleeve part which was not expected by the current simulation. As shown in this example, the displacement in the holding position between the actual and the one used for the simulation was not the major cause in all the three cases of failure.

Trials of the verification processes are shown in Fig. 10. Figure 10b shows the image after holding the pullover at the position indicated in Fig. 5f. The process to pick up the point is simulated as shown in Fig. 10a and the final result is superimposed in Fig. 10b. The predicted shape corresponds well to the observed clothes region. On the other hand, Fig. 10d is the observed image after holding the pullover at the position wrongly indicated, which is illustrated by the cross in Fig. 10b. As shown, the large unoverlapped area clearly indicates that the prediction is false.



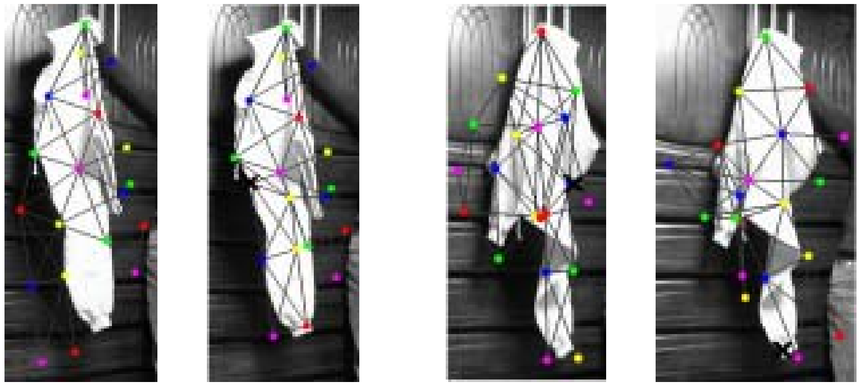
(a) (b)

Figure 8. Failure examples: (a) image data at State 8 (with predicted appearance of State 8 superimposed); (b) image data at State 4 (with predicted appearance of State 7 (left) and State 4 (right) superimposed).

7 Conclusion

We proposed a model-driven method of estimating the clothes state and indicating the position to be held next according to a given task. In the preliminary experiments using 29 actual images, the method estimated the correct states for 21 images, declared the reservation of judgment for three images, and gave wrong estimations for five images. As a result, for 24 images, the point to be held next was preferably indicated. We consider the results to be very good for this challenging subject and to show the good potential of the proposed strategy.

We also suggested verification processes by extension of the same strategy. At present, this part has not yet been sufficiently examined. We plan to build a hand-eye system for handling soft clothes by combining this method with actual manipulators. The verification processes will surely play an important role for practically carrying out a given task by the system.



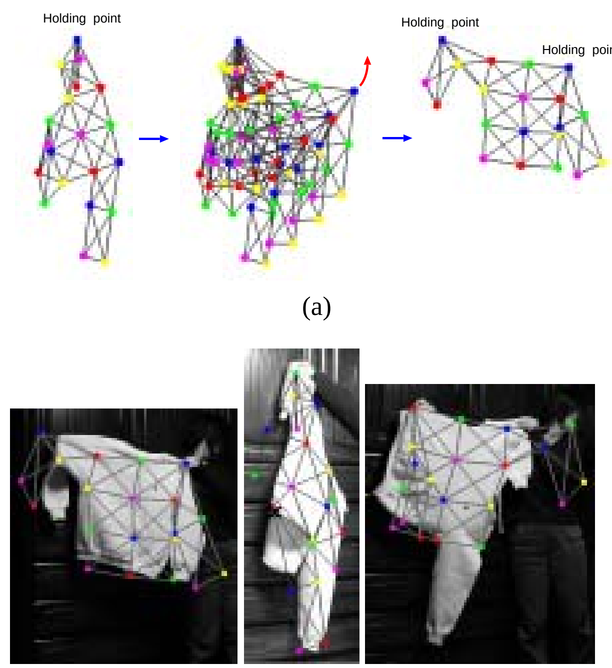
(a) (b)

Figure 9. Experimental results when the pullover is held at a middle point between adjacent nodes: (a) success example(left: initial overlap before partial modification, right: result); (b) failure example(with predicted appearance of the resultant state (left) and the correct state (right) superimposed).

The clothes model chosen for simulating the deformation in this paper is fairly coarse, partly because we aimed to derive a solution which requires less prior knowledge of the clothes in question. Although in the experiments of this paper we assumed the three sizes and stiffnesses of the clothes are approximately known, it may also be possible to automatically acquire such information by handling and observing the clothes.

Even if we use an elaborate clothes model, perfect expectation of the deformed clothes shape is almost impossible, because it is affected by slight differences in the boundary conditions, which are difficult to know. We believe that a key point to overcome this difficulty is to make good use of the information obtained in the observed image for modifying the approximate prediction. In the proposed method, the processes of III in Section 4 correspond to this.

One of our future works is to examine the generality of this strategy. For this purpose, we are now conducting experiments using different types of clothes with different stiffness. Other aspects we will consider include improvement of the simulation and modification processes, which should reinforce the robustness of the estimation. Although we use a single view in this paper, we are also starting to use stereo images by extending this strategy.



(b) (c) (d)

Figure 10. Verification examples: (a) simulation processes for verification (b) verification after the indication in Fig. 5f; (c) example of wrong indication; (d) verification after the indication in Fig. 9c.

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] D. H. House and D. E. Breen: “*Cloth modeling and animation*,” A. K. Peters, Ltd., 2000.

[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.