A Method of State Recognition of Dressing Clothes Based on Dynamic State Matching

Kimitoshi Yamazaki*1, Ryosuke Oya*2, Kotaro Nagahama*2, and Masayuki Inaba*2

Abstract— **This paper describes a method of clothing recognition using dynamic state matching. One purpose of this research is to develop a function that is used to observe the progress of dressing a bottom at a toilet or a bedside. In this dressing, although we just have to do is to put my or others' legs on the bottom. some undesirable situations may occur. For instance, the feet gets hung up on the part of the bottom, and so on.**

Assuming to use a single camera, we propose a method that distinguishes whether or not a behavior of dressing a bottom is working in order. The input data is an image stream that captures a dressing sequence. Optical flow is calculated from two consecutive images, and then the distribution of the flows is used to distinguish the present clothing state by matching the optical flow with training dataset. The method outputs whether or not the present clothing is in a successful situation. The effectiveness of the proposed method was proven by means of real image data.

I. Introduction

Clothes are essential tools for our daily life. Therefore, we inevitably have many routine work involving clothing. One of the important routines we usually perform is to dress clothes. However, it is not always true for all people to dress without difficulty. That is, automation of dressing clothes will contribute to improvement of Quality of Life for the people [2].

There has been researches about dressing that is assisted by automated machine. Matsubara et al. [10] proposed dressing assist method. The method is based on topology coordinates, and the application is dressing a T-shirt to a person by a dual arm robot. Reinforcement learning is applied to make the dressing motion. As other researches, we can find several developed machines that are for daily assistance. For the case of toilet support machine, what only a care receiver has to do is to stand at the center of the machine, and then robot arms having custom-designed actuators move up and down a bottom.

These researches performed dressing work with focusing on grasping point or motion trajectory for clothing manipulation. On the other hand, they did not consider about manipulation failures that occur while dressing. That is, recognition functions that were for providing the information of dressing state were not included in.

The purpose of this research is to develop a visual function for recognition of dressing clothes. As an application, the function can be used to observe the progress of dressing a bottom at a toilet or a bedside. In this dressing, although we just have to do is to put my or other's legs on the bottom, some undesirable situations may occur. For instance, the feet gets hung up on the part of the pants, a leg is not put into a desired hole (as shown in lower left and lower right in Fig. 1, respectively). It will become more successful if the progress of the dressing action is observed by external sensors in online.

Assuming to use a single camera, we propose a method that distinguishes that a behavior of dressing clothes is working in order or not. The input of the proposed method is an image stream that captures a dressing sequence. Optical flow is calculated at two consecutive images, and then the distribution of flows is used to distinguish the present clothing state by matching with optical flow extracted from training dataset. For this reason we proposed three types of feature to describe clothing condition. The method outputs whether or not the present dressing is in a successful situation. If the dressing is in failure, the method outputs what type of failure was occur. To improve the discrimination process, transition graph is brought in.

This paper is organized as follows: In the next section, related work was introduced. Section III describes our approaches. Section IV describes three features based on optical flow. Section V and VI describe state matching method based on our state transition model. Section VII introduces experimental results, and section VIII concludes this paper.

II. Related work

Clothing manipulation have been studied in the field of intelligent robot [5] [14]. Because clothing is flexible object, we have to observe both what type of manipulation is applied, and what is happening to the clothing by the manipulation. Ono et al. [11] estimated the state of a rectangular cloth from its contour information. A group of planar states, some of which are bent its corners, was used as knowledge to be observed. There have been other researches that used silhouette information for the estimation of the state of a piece of cloth hanged down by a robot [12] [9].

On the other hand, there are researches that use a 3D database that is produced by using a physics engine. Kita et al. [6] used 3D cloth model, and estimated a cloth state by fitting the model into a 3D pointcloud Maitin-Sphepard et al. [9] proposed action selection for manipulating deformable planar objects. By using physical model, a real robot straightened a square-shaped cloth. Kita et al. [9] utilized a 3D deformable model, and obtained a correspondence between the model and an input pointcloud that was captured by a trinocular stereo camera.

*1 Faculty of Engineering, Shinshu University, 4-17-1 Wakasato, Nagano, Nagano, 380-8553, Japan. PREST, JST. *2 Graduate School of Systems and Information Engineering, The University of Tokyo, 1-1-1 Hongo, Bunkyo-ku, Tokyo, 305-8573, Japan. yamazaki@shinshu-u.ac.jp

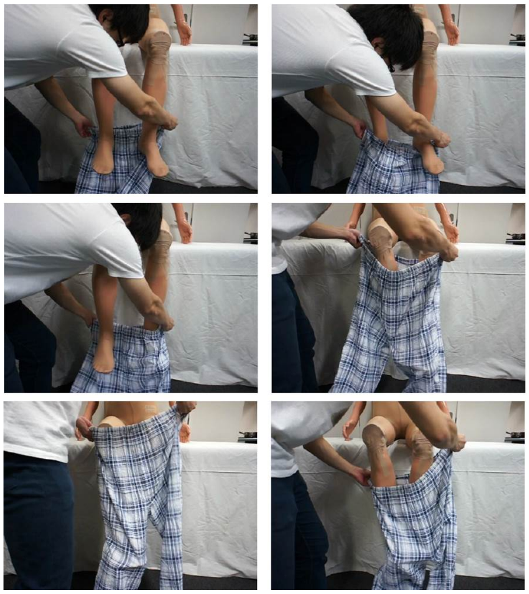

Fig. 1. Failure cases

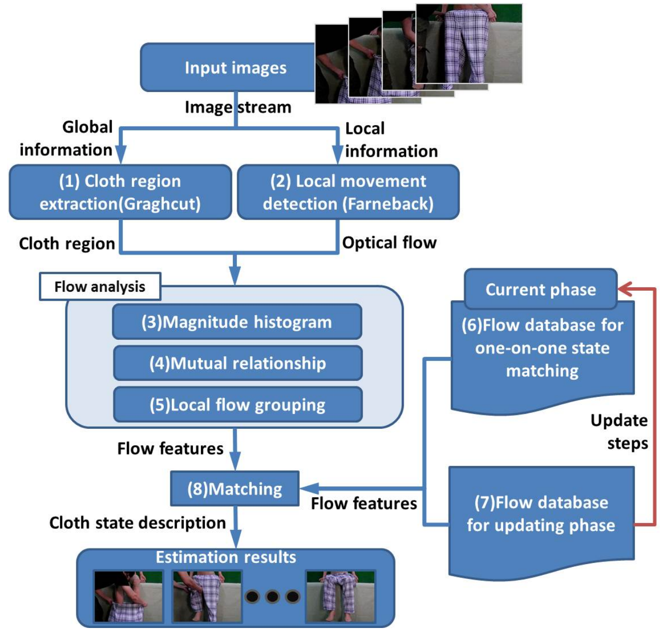

Fig. 2. A framework for the estimation of dressing clothes

These previous methods assumed that a piece of cloth exists with static condition when recognition was performed. However, methods that are for recognizing the state of clothing state while manipulating it should cope with dynamic condition for more effective estimation.

III. Approach

Figure 2 shows the configuration of the proposed method. Major components are as follows:

Preprocessing (Fig. 2(1), (2)): First, each image included in an input image stream is divided into cloth region and background region. After that, optical flow is calculated at the cloth region. Therefore the cloth states can be extracted as a set of tiny movements at local regions. The details are described in section IV.B.

Feature extraction (Fig. 2(3), (4), (5)): Because of cloth's flexibility, the transition of flows is not uniform. For representing the variety of the cloth movement, following three kinds of features are calculated; (i) the magnitude of the flows (section IV.C), (ii) the mutual relationship of the flows (section IV.D), (iii) the local movement of the flows (section IV.E).

Database (Fig. 2 (6), (7)): Image streams that capture several patterns of dressing clothes are used to generate database. A consecutive set of optical flow is calculated from each image stream, and it is added in the database with two-stage classification. In the first stage, a set of flows is extracted per frame pair, and they are registered for estimating the state of the dressing progress. In the second stage, a set of the optical flow is divided into a couple of phases according to the change of manipulation. For instance, a changing point is set to just after pulling up bottoms along shins. The details are described in section V.

State matching (Figure 2 (8)): Current state is estimated by searching corresponding optical flow in the database. If a failure is found, the type of the failure is also identified.

IV. State description of dressing clothes

A. Preprocessing 1: Domain segmentation

Before state description, image region of a target clothing is extracted by means of dynamic graphcut [7]. It is minimizes a cost function as follows:

$$E(X) = \sum_{v \in V} g_v(X_v) + \sum_{(u,v) \in N} h_{uv}(X_u, X_v), \tag{1}$$

where V is a group of image pixels, and $N \subset V$ is a group of pixels whose positions are neighbor to a target pixel. X_v is a label assigned to each pixel. The first term $g_v(\cdot)$ is called 'data term', which represents likelihood of a label assigned to a pixel. The second term $h_{uv}(\cdot)$ is called 'smoothing term', which represents the cost of disconnection between a present pixel and neighbor pixels.

In general, graphcut needs seed points before optimization process. We manually define them as two small regions, one is from clothing region and another is from background region.

B. Preprocessing 2: Optical flow calculation

Optical flow is calculated from the consecutive clothes region in an image stream. A number of methods have been proposed to calculate optical flow [8]. As clothes is not always having distinctive texture, methods that do not need that kind of texture are suited to our case. For this reason, we apply a method proposed by Farneback [3]. It corresponds

Fig. 3. An example of optical flow when only lower left part was manipulated. Left: original image, center: detected flows depicted by color segments, right: the distribution of flow magnitude.

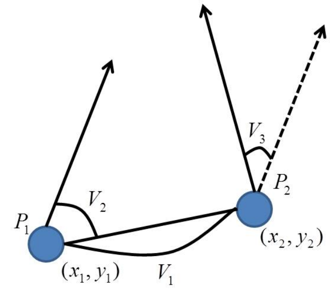

Fig. 4. Mutual relationship between two flows

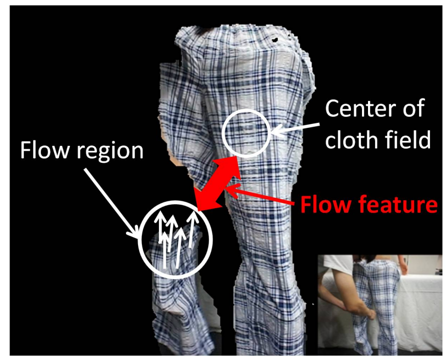

Fig. 5. Relationship between overall flows and local flows

two local image window $\mathbf{W}$ by using following error vector ε :

$$\varepsilon = \int\int_{\mathbf{W}} [J(\mathbf{x} + \mathbf{d}) - I(\mathbf{x})]^2 w(\mathbf{x}) d\mathbf{x}, \tag{2}$$

where $\mathbf{x}$ is a pixel coordinates, and $\mathbf{d}$ is displacement that should be estimated. $J(\cdot)$ and $I(\cdot)$ denotes two consecutive images.

Because the method outputs high-density flows from even less texture region, it has high expressive power about detailed shape change of the clothes.

C. The magnitude of the flows F_m

While a piece of cloth is manipulated, the amount of its motion can vary from place to place because it is a soft object. One of methods to capture the motion characteristics globally is to use the magnitude of the flows.

Our first feature description is calculated by the following procedure. Unnecessary flows whose magnitude is too small or too large are removed by means of threshold processing. Remaining flows are normalized based on the average magnitude of them. The reason for the normalization is to use the flows for matching process without influence by the difference of dressing speed.

The center of Figure 3 shows color-coded flows according to their directions. The right figure shows the magnitude of the flows (moving speed) by means of brightness difference. F_m is generated by calculating density histogram from the gray scale image.

D. The mutual relationship of the flows F_r

The second feature is designed to describe position and direction relationship between a pair of flows. The intention is to describe the complexity of flows, that is, a group of flows is in well-aligned or disturbed condition.

Let (x_i, y_i) be the starting position of the each flow. Using optical flow calculated between a pair of images, mutual relationship between local parts is described. This description is inspired from SPRH [13] proposed by Wahl et al. In the feature generation, two flows are selected, and following three dimensional vector is calculated.

1) The distance between two flows (Fig. 4, V_1)

$$V_1 = \sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2} \tag{3}$$

2) The angle between one flow direction and the position vector between two flows. (Fig. 4, V_2)
3) The angle between one flow and another (Fig. 4, V_3)

These calculations are performed against many of flow pairs that are randomly sampled, and then the results are voted into the three dimensional parameter space whose axes corresponds to above item 1) to 3) F_r is generated by calculating a frequency histogram from the voted space. That is, the space is regarded as divided into small voxels, and number of points in each voxel are counted. Each number is an element of a bin of the histogram. In our experiments, the number of element of the frequency histogram was set to $5^3 = 125$. It means that each of three axes was divided into five sections,

E. The local movement of the flows F_l

Wrinkles and stretches can be found on cloth because of its flexibility. That is, when a part of a target cloth is manipulated, the cloth can often transmute around the manipulated part. For representing such partial movement, possible flows in local area are extracted and used.

In preparation for the feature description, the center position of the cloth region are calculated from image segmentation result. On the other hand, only flows that constitute a region where the flows densely exist are selected. RANSAC [4] is used for this process. F_l is generated by calculating the relationship between the center position and the local flows as shown in Fig. 5. The feature consists of two criteria: (i) the direction from the local flows to the center position, and (ii) the ratio of the area of local flows and cloth region.

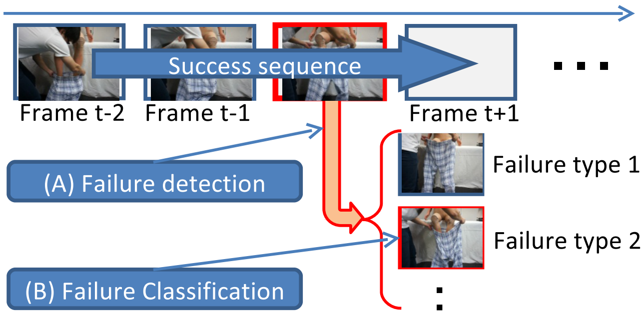

Fig. 6. Online classification of clothing sequence

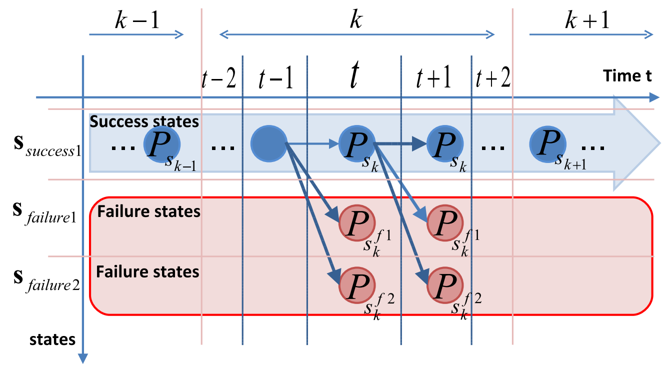

Fig. 7. State transition model

V. State transition model and matching

There are three kinds of estimation targets that describe clothing state.

1) Dressing clothes is in success or in failure (Fig. 6, (A)),
2) if in success, which state it is in the present,
3) if in failure, what kind of failure it is (Fig. 6, (B)).

For making state classification, we prepare a database that is consist of feature vectors calculated from a set of training data. Several image streams about successful dressing clothes are captured. On the other hand, several series of image streams with the failure of dressing action are also captured. Optical flow is calculated from each of the image streams, and three features F_m, F_r and F_l are calculated. The database has a list of the features to which information of dressing state is appended.

The main roles of the database are as follows:

1) Phase estimation: One dressing sequence is divided into several phases. The database is used to recognize the switching of the phases.
2) One-on-one matching: A pair of consecutive images produce a set of optical flow. The result is used for frame to frame matching that is to recognize present dressing condition.

Based on them, we implement two types of feature set as shown in Fig. 2 (6), (7).

A. Evaluation formula

The estimation of dressing clothes is achieved by searching a feature sequence that is the most similar to a feature set calculated from current image pairs. Let $\mathbf{F}$ be a feature set, that is also written as $\mathbf{F} = \{\mathbf{F}_m, \mathbf{F}_r, \mathbf{F}_l\}$. Each of $\mathbf{F}_*$ is a time-series feature set; for instance, $\mathbf{F}_m = \{\mathbf{f}^1_m, \mathbf{f}^2_m, \cdots, \mathbf{f}^T_m\}$, where T denotes a serial number of clock time, and each of $\mathbf{f}_m$ consists of a set of flows calculated from a pair of images.

Because we have three different feature descriptions, similarity calculation should include integration procedure for three features. It is achieved by the following equation:

$$P_{\mathbf{s}_k} = \prod_{i=\{m,r,l\}} dist(\mathbf{F}_i, \mathbf{F}^{\mathbf{s}_k}_i), \tag{4}$$

where $P_{\mathbf{s}_k}$ represents the similarity with phase $\mathbf{s}_k$ in the database. A matching result is a result that shows the smallest value in the equation (4). $\mathbf{s}_k$ denotes kth phase of dressing clothes. $\mathbf{s}_k$ consists of a group of state s^t_k, which we try to represent by using three features $\mathbf{f}^t_m$, $\mathbf{f}^t_r$ and $\mathbf{f}^t_l$. That is, $\mathbf{s}_k = \{s^{t-a}_k, \cdots, s^t_k, \cdots, s^{t+b}_k\}$ (where a and b are positive number). If any one of s^*_k included in $\mathbf{s}_k$ is well matched with current optical flow, the current phase is estimated as k.

The similarity of $\mathbf{F}_m$ is calculated by means of Bhattacharyya distance [1]. Pairs of features whose distance is under the pre-defined threshold are candidates.

$$dist(\mathbf{F}_m, \mathbf{F}^{\mathbf{s}_k}_m) = \sqrt{1 - \sum_{n=1}^{N_m} \sqrt{H_m(n) H^{\mathbf{s}_k}_m(n)}}, \tag{5}$$

where $H_m(n)$ denotes an element of density histogram $\mathbf{H}_m$ that is derived from flow magnitude. n is the serial number of the histogram bins. $\mathbf{H}^{\mathbf{s}_k}_m$ denotes a histogram derived from a state s^t_k in the database.

The similarity of $\mathbf{F}_r$ is also calculated as a distance between two histograms.

$$dist(\mathbf{F}_r, \mathbf{F}^{\mathbf{s}_k}_r) = \sum_{n=1}^{N_r} \sqrt{(H_r(n) - H^{\mathbf{s}_k}_r(n))^2}, \tag{6}$$

where $\mathbf{H}_r$ and $\mathbf{H}^{\mathbf{s}_k}_r$ are relation histograms for current optical flow and that in the database, respectively.

The similarity of $\mathbf{F}_l$ is calculated by using the distance between two feature vectors:

$$dist(\mathbf{F}_l, \mathbf{F}^{\mathbf{s}_k}_l) = \sum_{n=0}^{N_l} \sqrt{(H_l(n) - H^{\mathbf{s}_k}_l(n))^2}. \tag{7}$$

VI. Improvement of effectiveness and robustness of state matching

One of the approaches of searching a state that is similar to current state is to match the current state with all of states in the database. However, such approach is time consuming and easy to produce mismatching. For this reason, we apply state transition model so that the matching process becomes more effective and robust.

The transition model is designed based on dressing phases. For instance, if a present phase is in a situation that both legs are run into a bottom that is pulled up by both hands, next

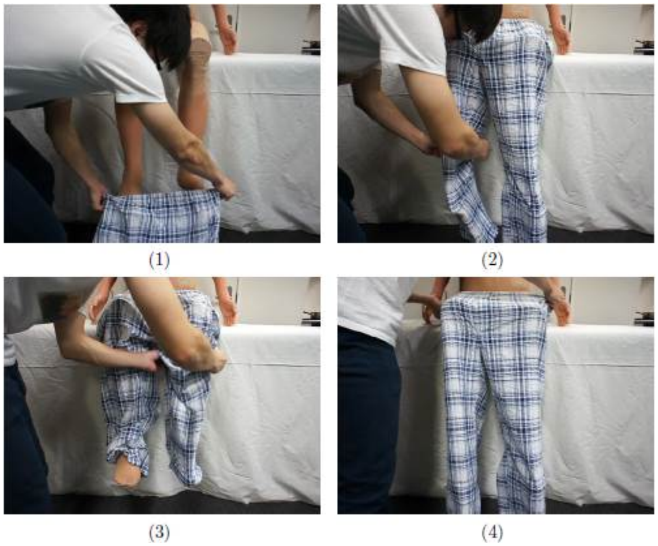

Fig. 8. 4 phases of dressing a bottom

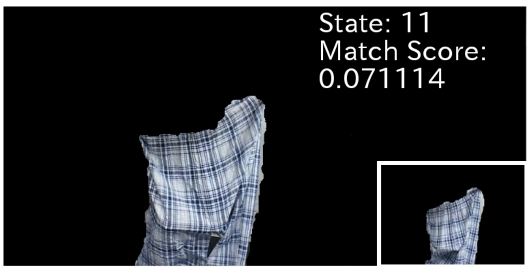

Fig. 9. Current state estimation

phase can be limited to a manipulation for running one foot into the bottom. If such transition is not detected, it is viewed a present manipulation as a failure.

A transition from one phase to another is decided according to the following equation;

$$\mathbf{s}_{k'} = \max_{\mathbf{s} \in \mathbf{S}} \{ w_{\mathbf{s}_k \rightarrow \mathbf{s}_{k'}} \cdot P_{\mathbf{s}_{k'}} \}, \tag{8}$$

where $\mathbf{S} = \{\mathbf{s}_1, \mathbf{s}_2, ..., \mathbf{s}_K\}$, and $w_{\mathbf{s}_k \rightarrow \mathbf{s}_{k'}}$ denotes a weight coefficient that is for a transition from phase $\mathbf{s}_k$ to $\mathbf{s}_{k'} \in \mathbf{S}$. If there is no way to $\mathbf{s}_{k'}$, $w_{\mathbf{s}_k \rightarrow \mathbf{s}_{k'}}$ is set to 0. $P_{\mathbf{s}_{k'}}$ is the score of the matching with $\mathbf{s}_{k'}$, as shown in equation (4).

Fig. 7 shows a conceptual diagram of the state transition model. Normally, a transition occurs only on phases in success. but if in failure, a phase transits to failure states such as s_{f1} and s_{f2}.

VII. Proof experiments

A. Settings

A mannequin was set on a horizontal board whose height was 700 [mm]. A camera placed at 750 [mm] height and 1500 [mm] away from the mannequin was used to capture image streams. Each image stream is VGA size (640 $\times$ 480 [pixel]) with 30 [fps]. Under the setting, a person clothed in a bottom to the mannequin.

TABLE I

Estimate state and calculate scores

	Input Pattern	Result	Avg. score
Success cases	No.1 to No.4	No.9	0.5242
	No.1 to No.4	No.1 to No.4	0.3620
	No.1 to No.4	No.10	0.3926
	No.1 to No.4	No.1 to No.4	0.5690
	No.1 to No.4	No.1 to No.4	0.3204
	No.1 to No.4	No.1 to No.4	0.3712
Failure cases	No.5	No.5	0.3110
	No.6	No.6	0.4762
	No.7	No.11	0.4307
	No.7	No.7	0.3901
	No.8	No.11	0.3455
	No.8	No.8	0.2826
	No.8	No.8	0.2937
	No.9	No.9	0.4027
	No.9	No.8	0.2993
	No.10	No.7	0.3210
	No.10	No.11	0.3641
	No.11	No.11	0.3508
	No.11	No.11	0.4013

TABLE II

Compare to trial without one feature

	All	F_r, F_l	F_m, F_l	F_m, F_r
Success rate [%]	58–73	21–36	5–20	21–42
False positive [%]	0–8	15–30	15–38	0
False negative [%]	33–66	50–83	33–83	84–100
Failure state estimation [%]	69–77	15–30	15–30	30–45

Before state matching experiments, we captured 6 different image streams that were of dressing clothes in success. For representing state transition about the success of dressing clothes, each of image stream was divided into four phases. They were named as No.1 to No.4 (Fig. 8 shows starting frames of the 4 phases). On the other hand, 7 different image streams that included failure state were also captured. They were named as No.5 to No.11. Feature descriptions based on optical flow calculated from these 13 image streams were used to constitute a database.

The purpose of this experiments was to estimate whether or not the dressing clothes was completed. That is, the result of state matching will transition from phase No.1 to phase No.4 in order if the dressing action succeeds. If this is not the case, the matching result was classified into failure, and the type of failure is identified from phase No.5 to No.11 in database.

B. Success / failure classification and failure state estimation

Fig. 9 shows an example of state matching. Current state is consistently matched with the flows in database (shown in (6) in Fig. 2). A large image on the left side is an image corresponding to the current optical flow, and a small image on the lower right side is an image corresponding to matched features in database. These images show a failure that both legs fell into one hole while hitching up a bottom.

Table I shows experimental results. In this experiment, 19 series of image sterams were used as test data. The second column describes the success / failure pattern of input image

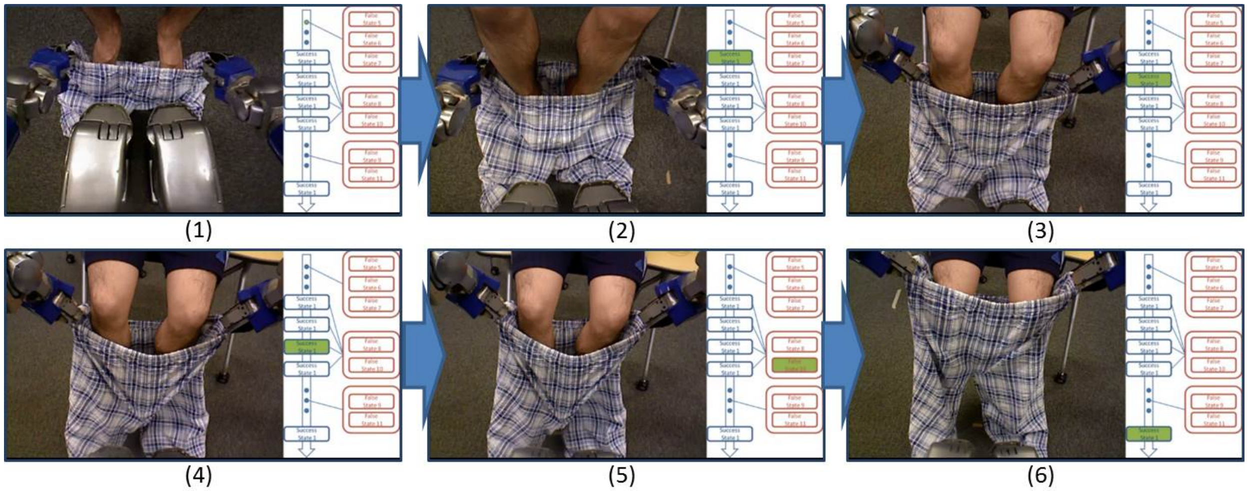

Fig. 10. Online experiment

streams, and the third column describes that of matching results. Avg. score shows average similarity value of the matching. The matching was failed with 2 of the 6 trials about success case. On the other hand, all of failure cases were correctly estimated as failure. However, failure type estimation was failed with 4 of the 13 trials. As a result, the final success rate was 73 %.

Table II shows results with several combinations of the features. It was tried with 4 times for each combination. Values in the cells show the range of success rate. In each case of using only two types of features, the success rate was much lower than the combination of three features. This fact tells us that three types of feature descriptions proposed in this paper have different representation power that is needed for estimation of dressing behavior.

Fig.10 shows an experimental result that is an application to dressing by an autonomous robot. Left side of each image is captured from a camera mounted on the robot, and left side shows dressing phases ordered by temporeal sequence. If one blue rectangle is filled by green color, it means that present phases is in success. Otherwise, it is in failure. In this experiment, the dressing procedure was temporarily classified into failure because the bottom got stuch with left toe (Fig. 10, (5)). However, the dressing motion was retried based on the estimation result. As a result, dressing behavior was continued until the bottom was pulled up over knees.

VIII. Conclusions

This paper describes a method of estimation of dressing clothes. The input of the method is an image stream that captures the progress of the dressing action. It recognizes whether or not the present dressing state is in a successful situation, and if in failure, what type it is. Through experiments about dressing a bottom at the bedside, we confirmed the effectiveness of the method.

For future work, application to the automation of dressing clothes by robots should be considered.

References

[1] D. Comaniciu, V. Ramesh, P. Meer: "Real-Time Tracking of Non-Rigid Objects using Mean Shift," in Proc. of IEEE Conf. on Computer Vision and Pattern Recognition, Vol. 2, 142-149, 2000.

[2] B. Dudgeon, J. Hoffman, M. Ciol, A. Shumway-Cook, K. Yorkston, L. Chan: "Managing Activity Difficulties at Home: A Survey of Medicare Beneficiaries," Archives of Physical Medicine and Rehabilitation, Vol. 89, No. 7, pp. 1256 - 1261, 2008.

[3] G. Farneback: "Two-Frame Motion Estimation Based on Polynomial Expansion," in Proc. of the 13th Scandinavian Conference on Image Analysis, LNCS 2749, pp. 363 – 370, 2003.

[4] M. A. Fischler and Robert C. Bolles: "Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography," Communications of the ACM, Vol. 24, No. 6, pp. 381 – 395, 1981.

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

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

[7] P. Kohli and P.H.S. Torr: "Efficiently solving dynamic Markov random fields using graph cuts," in Proc of IEEE Int'l Conf. on Computer Vision, Vol. 2, pp. 922 – 929, 2005.

[8] D. Lucas and T. Kanade: "An Iterative Image Registration Technique with an Application to Stereo Vision," in Proc. of Int'l. Joint Conf. on Artificial Intelligence, pp.674 - 679, 1981.

[9] J. Maitin-Sphepard, M. Cusumano-Towner, and P. Abbeel J. Lei: "Cloth Grasp Point Detection based on Multiple-View Geometric Cues with Application to Robotic Towel Folding," in Proc of Int'l. Conf. on Robotics and Automation, pp. 2308 – 2315, 2010.

[10] T. Matsubara, D. Shinohara and Masatsugu Kidode: "Reinforcement Learning of a Motor Skill for Wearing a T-shirt using Topology Coordinates," Advanced Robotics, Vol. 27, Issue 7, pp.513 - 524, 2013.

[11] 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, pp. 1363 - 1366, 1990.

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

[13] E. Wahl and U. Hillenbrand and G. Hirzinger: "Surflet-Pair-Relation Histograms: A Statistical 3D-Shape Representation for Rapid Classification," in Proc. of 3-D Digital Imaging and Modeling, pp. 474 - 482, 2003.

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