Proceedings of the 2016 IEEE/SICE International Symposium on
System Integration, Sapporo Convention Center, Sapporo, Japan,
December 13-15, 2016



# Recognition and Handling of Clothes with Different Pattern by Dual Hand-eyes Robotic System*

Ryuki Funakubo, Khaing Win Phyu, Hongzhi Tian and Mamoru Minami[1]

*Abstract*— **Recently, robots have been used in clothing industries for mass production with countless merits. However, there remain many challenges for robots in recognition, pose (position and orientation) detection operations, especially when the working object is deformable and every working object has unique shape and color. In this paper, pose detection of clothes through 3D recognition is proposed for the task in which the manipulator recognizes clothes, estimates relative pose and performs pick and place function. In proposed cloth recognition system, a variety of models of different clothes with unique shape and color are generated as BMP (bit map file) format extracted from the camera. Using the photograph model, recognition of cloth is performed by using images from the two cameras that are fixed at the end effector of the robot arm. 12 varieties of different clothes samples are used for this experiment. The pose of individual cloth is estimated by Genetic Algorithm (GA). 1000 times recognition and handling experiment has been executed, having shown the effectiveness of proposed Photo-Model-Based pose recognition system.**

## I. INTRODUCTION

Nowadays, most of the garment companies especially in Japan are facing with two main inconveniences as follow:

- Growing shortage of labor force because an aging population have been progressing.
- Weak point of human workers are laziness, boring and tiring due to the long working hours.

Recently, IT based technology and Robotics have began to be used in the garment (cloth) companies considering above problems. However, robotics in garment industry are capable of operations only if preconditions are met such as (1) surrounding the operating environment, the light environment is guaranteed not to change with time, (2) the handling object is able to be defined the shape of object in advance, and (3) the design of the robot hand is predefined based on the object shape. A number of research for robot in recognition deformable objects especially clothes have been done[9].

In application to cloth handling, the main tasks are to get the purchasing order of clothes from on line customer and classify these clothes and package and place in different box for storage every day. Since clothes is deformable object, no definition of clothes can be predefined in computer. Consequently it is too difficult to handle a wide variety of clothes that are in irregular shape and size. Therefore, we have developed vision-based robot system as shown in Fig.1 to solve the above problems for mass handling robot of clothes with varieties.

On the other hand, robot control technology using vi-

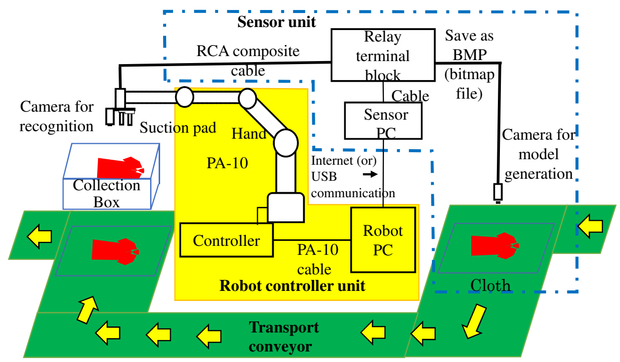


Fig. 1. System configuration

sual information, called as visual servoing, is playing an important role in industry. In our previous works, we have been developing a three-dimensional move on sensing system named 3D-MoS using two cameras as stereo vision sensor. The robot control technology using visual information has been already confirmed in guidance and control of underwater robot (3D-MoS / AUV)[6] and soil decontamination robot system (fully automated robotic system to decontaminate the radioactive contamination soil)[8]. In these previous works, model-based method is used to recognize 3D ( three dimensional) pose.

However, in cloth handling application in which the manipulator has to pick and place unique-shape-and-color clothes with random appearance in camera images, it is impossible to use models that are predefined for all clothes. Therefore, as a main contribution in this paper, we introduce a new approach to generate model for every clothes during the operation of pick and place. In proposed "Photo-model 3D-MoS", the model of object is automatically created from a photo by robot itself. Then, relative pose estimation of clothes is performed using generated model through model-based clothes recognition. PA10 with 7-DoF (Degree of Freedom) is being used for recognition and pose detection operations.

The configuration of the system is shown in Fig. 1. In this configuration, three cameras are used as vision sensors. First camera is used for model generation. The other two are used for recognition based on the photograph model which are fixed at the end-effector of the mobile manipulator (PA10 robot). The merits of this proposed system are to save the cost of staff and to get the better performance and higher accuracy than workers in the garment company. As the final

*This work was not supported by any organization

[1]Ryuki Funakubo, Khaing Win Phyu, Hongzhi Tian and Mamoru Minami are with Graduate School of Natural Science and Technology, Okayama University Tsushimanaka 3-1-1, Okayama, Japan. pip62k65@s.okayama-u.ac.jp

978-1-5090-3328-7/16/$31.00 ©2016 IEEE

objective is to pick and place the clothes in/out individual boxes, it is necessary to estimate the pose of the clothes correctly. Therefore, we conducted 1000 times for different clothes and analyzed the recognition accuracy.

## II. PROPOSED PHOTOGRAPH MODELING-BASED CLOTHES RECOGNITION

There are two main portions in proposed system. The first portion is cloth model template generation and the latter is relative pose estimation using generated model template through model-based matching method. Here is description of kinematics of stereo-vision before explanation of proposed system in detail.

### A. Kinematics of Stereo-Vision

Fig. 2 shows the relationships between the world coordinate system of the manipulator $\Sigma_W$ and the hand coordinate system $\Sigma_H$. On the other hand, it can also be called as the coordinate system of 3D-MoS (Move on Sensing) robot. The coordinate system of dual-eyes vision system can be seen in Fig.3. An target object coordinate system is expressed in $\Sigma_M$. In the image coordinate system, the coordinate system of the left and right cameras are represented as $\Sigma_{CR}$ and $\Sigma_{CL}$. $\Sigma_{IR}$ and $\Sigma_{IL}$ are the coordinate systems of the left and right cameras' images. According to coordinate system, the j-th point of the i-th model can be represented by the following the simultaneous transformation matrix.

- ${}^{CR}T_M$:Homogeneous transformation matrix from right camera coordinate system $\Sigma_{CR}$ to the object coordinate system $\Sigma_M$
- ${}^{CR}r_i$: The object as viewed from the search point i-th on the model in $\Sigma_{CR}$
- ${}^{M}r_i$: The object as viewed from the search point i-th on the model in $\Sigma_M$

Therefore, ${}^{CR}r_i$ can be calculated by using Eq. (1)

$$ {}^{CR}r_i = {}^{CR}T_M \, {}^{M}r_i. \tag{1} $$

The homogeneous transformation matrix ${}^{W}T_{CR}$ from world coordinate system $\Sigma_W$ to the right camera coordinate system $\Sigma_{CR}$ can be obtained from Eq. (2).

$$ {}^{W}r_i = {}^{W}T_{CR} \, {}^{CR}r_i. \tag{2} $$

By using matrix $P$ according to projective transformation, the position vector of the i-th point in the right camera image ${}^{IR}r_i$ can be described as Eq. (3). Eq. (4) is described as matrix $P$.

$$ {}^{IR}r_i = P \, {}^{CR}r_i. \tag{3} $$

$$ P = \frac{1}{{}^{C}z_i} \begin{bmatrix} \frac{f}{\eta_x} & 0 & {}^{I}x_0 & 0 \\ 0 & \frac{f}{\eta_y} & {}^{I}y_0 & 0 \end{bmatrix}. \tag{4} $$

Using the same method, it is possible to obtained the position vector of the i-th point in the left camera image ${}^{IL}r_i$.

$$ {}^{CL}r_i = {}^{CL}T_M \, {}^{M}r_i. \tag{5} $$

$$ {}^{W}r_i = {}^{W}T_{CL} \, {}^{CL}r_i. \tag{6} $$

$$ {}^{IL}r_i = P \, {}^{CL}r_i. \tag{7} $$

According to Eq. (3) and Eq. (10), the relationship of Eq. (8) is connected an arbitrary point on a 3D-model with a pose ${}^{C}\phi_M$ – the pose $\Sigma_M$ based on $\Sigma_{CR}$ – with the projected point on the left camera image${}^{IL}r_i$ and right camera image ${}^{IR}r_i$ can be written as .

$$ \begin{cases} {}^{IR}r_i = f_R({}^{CR}\psi_M, \, {}^{M}r_i) \\ {}^{IL}r_i = f_L({}^{CL}\psi_M, \, {}^{M}r_i). \end{cases} \tag{8} $$

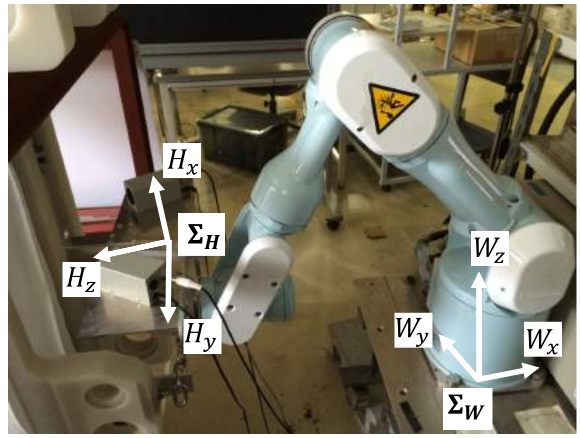


Fig. 2. Coordinate system of 3D-MoS robot

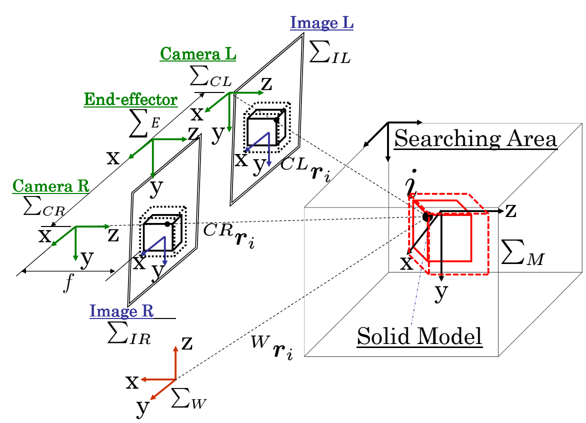


Fig. 3. Coordinate system of dual-eyes

### B. Fitness Function

A function $P({}^{IR}r_i)$ represents the matched degree of the i-th point of the model on the right image area, ${}^{IR}r_i$. Similarly, the left image area, ${}^{IL}r_i$ represented as a function $P({}^{IL}r_i)$. As shown in Eq.(9), the fitness value will be increase with the voting value of "+2" for every point of clothes in captured images that lies inside of the model frame $S_{R,in}$ and $S_{L,in}$ . The fitness value will decrease with the value of "$-0.005$" for every point of clothes in images whose hue value is similar to average of background and otherwise is "0". As shown in Eq.(10), the fitness value will be increase with the voting value of "+0.1" for every point in outside space of model frame $S_{R,out}$ and $S_{L,out}$ and otherwise is "−0.5". The correlation between the i-th point of the model (the search model) and the image having the evaluation value with such a sign is used the following Eq. (11). The whole evaluation function $F({}^{C}\phi_M)$ is obtained by the average of the fitness function of both left camera image $F_L({}^{CL}\phi_M)$ and right camera image $F_R({}^{CR}\phi_M)$.

$$p_{Rin}({}^{IR}r_i) = \begin{cases} 2 & (|H_{IR}({}^{IR}r_i) - H_{MR}({}^{IR}r_i)| \le 30) \\ -0.005 & (|\overline{H}_B - H_{IR}({}^{IR}r_i)| \le 30) \\ 0 & (otherwise) \end{cases} \quad (9)$$

$$p_{Rout}({}^{IR}r_j) = \begin{cases} 0.1 & (|\overline{H}_B - H_{IR}({}^{IR}r_j)| \le 20) \\ -0.5 & (otherwise) \end{cases} \quad (10)$$

$$\begin{aligned} F({}^{C}\phi_M) &= \Bigg\{ \Bigg( \sum_{\substack{{}^{IR}r_i \in \\ S_{R,in}({}^{CR}\phi_M)}} p({}^{IR}r_i) - \sum_{\substack{{}^{IR}r_i \in \\ S_{R,out}({}^{CR}\phi_M)}} p({}^{IR}r_i) \Bigg) \\ &\quad + \Bigg( \sum_{\substack{{}^{IL}r_i \in \\ S_{L,in}({}^{CL}\phi_M)}} p({}^{IL}r_i) - \sum_{\substack{{}^{IL}r_i \in \\ S_{L,out}({}^{CL}\phi_M)}} p({}^{IL}r_i) \Bigg) \Bigg\} / 2 \\ &= \left\{ F_R({}^{CR}\phi_M) + F_L({}^{CL}\phi_M) \right\} / 2 \end{aligned} \quad (11)$$

### C. Genetic Algolithm(GA)

Recognition problem of the object can be converted to a searching problem of maximum value $F({}^{C}\phi_M)$. There are various ways in finding the maximum value of the fitness function. The simplest and easiest way is the full search method. It is intended to find the maximum value by scanning all possible pixels. However, it has inefficient drawback in term of computing time. Even though there are powerful optimization methods, GA with long history is selected in this work. By applying the GA evaluation process as an optimization solution, the maximum value search processing can be completed efficiently in a short period of time. In this experiment, GA has 60 individuals representing different poses of model. Each individual chromosome has six variables. Each variable are coded by 12 bits. The former 36 bits represent for the position coordinate of the 3D-model and the last 36bits represent for the orientation of the 3D-model. The characteristics of GA individual is defined as

$$\underbrace{\overbrace{01\cdots01}^{t_x}}_{12bit}\underbrace{\overbrace{00\cdots01}^{t_y}}_{12bit}\underbrace{\overbrace{11\cdots01}^{t_z}}_{12bit}\underbrace{\overbrace{01\cdots01}^{\varepsilon_1}}_{12bit}\underbrace{\overbrace{01\cdots11}^{\varepsilon_2}}_{12bit}\underbrace{\overbrace{01\cdots10}^{\varepsilon_3}}_{12bit}.$$

These 60 chromosomes are evaluated by fitness value. Fitter ones are selected to regenerate next generations. Finally the best chromosome that has the most trustful pose is achieved as shown in Fig.4.

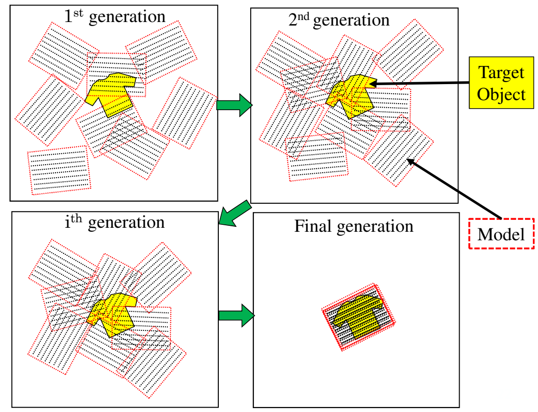


Fig. 4. GA evolution process

### D. Photo Model Recognition

In our system, three cameras are used as vision sensors. Among these three cameras, the first camera that is fixed in the workspace for capturing photos of clothes is used for cloth model generation. Captured photos are saved as BMP (bit map file) format. Note that generated model is not for just matching between saved ones and current images. Instead, generated model is used to estimate relative pose with respect to end effector using images from two cameras attached together with end effector. In proposed model generation technique, firstly, the background photo is taken as shown in Fig. 5 (a) and average hue value is calculated. Secondly, the cloth (target object) is put on the background and hue value of each point in the image is calculated as shown in Fig. 5 (b). Thirdly, the individual pixel of captured image is compared by scanning with the average of the hue value of the background image to define model frame based on error. Then, inner surface space of model $S_{in}$ is generated by sampling hue value of each point inside defined frame. Finally, the outside space $S_{out}$ of model is generated as shown in Fig. 5 (c).

### E. Model-Based Recognition

After generating a model from a bitmap image, the model is used for recognition the cloth (target object). Here, an overview of the recognition method with respect to the camera image is given as a description. 3D pose of the 3D model ${}^{C}\phi_M = [{}^{CR}x_M, {}^{CR}y_M, {}^{CR}z_M, {}^{CR}\varepsilon_{1M}, {}^{CR}\varepsilon_{2M}, {}^{CR}\varepsilon_{3M}]^T$ is determined using model-based matching method. Generated models with different poses are projected from 3D-model in searching area onto the left and right 2D images plane as shown in Fig.6. By comparing the projected models with images from two cameras attached at the end effector, relative pose is estimated by using fitness function $F({}^{C}\phi_M)$ to evaluate. It means the pose of the best model, that is

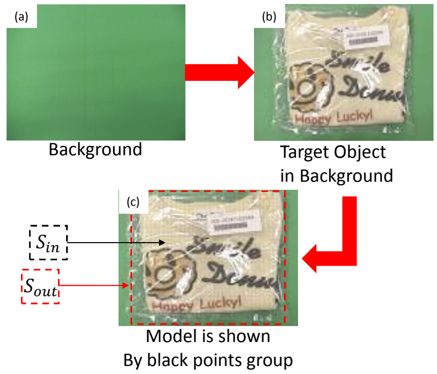


Fig. 5. Model generation technique

fully matched with captured images from left and right cameras, is selected as estimated pose. The top of Fig. 6 is represented as searching area of a 3D-model named S to search for a cloth(target object). $S_{in}$ is depicted by the space of coordinates on the surface of the model and $S_{out}$ was enveloped the outside space of $S_{in}$. The left and right 2D searching model are named as $S_L$ and $S_R$. In order to evaluate, the evaluation and change in hue of the surrounding of the object as shown in the interior region is represented as $S_{R,in}$, $S_{L,in}$ and the outside space enveloping $S_{R,in}$, $S_{L,in}$ is defined as $S_{R,out}$, $S_{L,out}$.

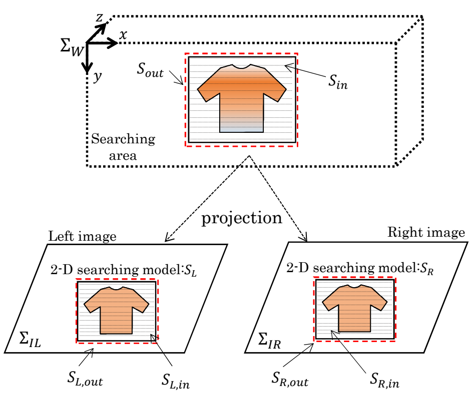


Fig. 6. Searching model

## III. EXPERIMENTAL ENVIRONMENT

There are two units in experimental environment. One is for generation cloth model including one camera as shown in Fig. 7. Another one is end effector equiped with two cameras installed in manipulator's end effector as shown in Fig. 8 and Fig. 9. In Fig. 7, the distance from the camera lens to the model creating plane is 400 mm and the plane color is green. The size of clothes models can be up to 250mm × 200mm. Each coordinate system of the robot and the cloth used in this experiments are shown in Fig. 8 and Fig. 9. The cloth coordinate system is represented as $\Sigma_M$ and $\Sigma_H$ is defined as the hand coordinate system of the robot – end effector –. $\Sigma_M$ can be viewed from (x=0, y=0, z=685mm). It is centered on the recognition range of the position as a reference of the 510mm × 390mm. After defining about the position, the recognition range of the angles are from 53° to −53°. The size of the collection box is a 220mm × 220mm. However, in this experiment, we mainly emphasized to the recognition experiment and handling experiment is our follow-up work.

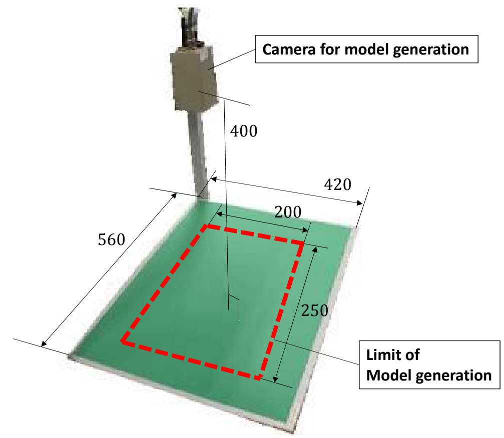


Fig. 7. Coordinate system of target object (unit is mm in Figure 7)

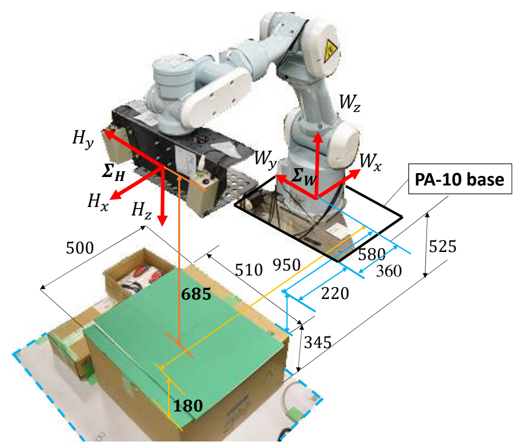


Fig. 8. Environment of model generation (unit is mm in Figure 8)

## IV. EXPERIMENTAL CONTENT

### A. 1000 Times Recognition Experiment

In order to prove the high efficiency of our recognition system, we made use of 12 unique clothes(No.1, No.2, ... , No,12)samples as shown in Fig. 10 to conduct a verification experiment. In this experiment, we have conducted for 1000 times recognition and analyzed in term of average error and standard deviation. Illuminance is 700Lx during this experiment.

Generally, on the condition that data is similar to normal distribution, the probability that the data is in the range of the average±3$\sigma$ is 99.7% , meaning the error in x and y direction is less than 10mm and the one of $\theta$ is less than 10° with

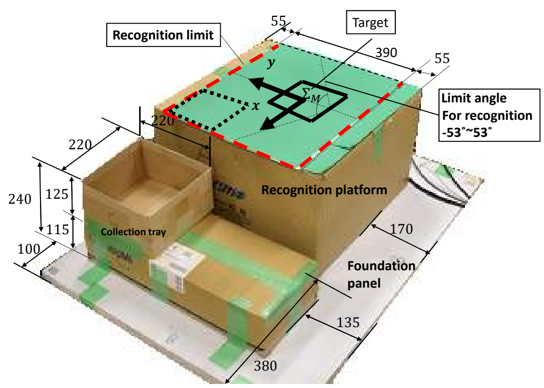


Fig. 9. Coordinate system of robot and end-effector (unit is mm in Figure 9)

probability of 99.7% . Though the error and the standard deviation in z direction is big, if x, y and $\theta$ can be detected accurately, clothes can be handling, with a force sensor being fixed at the robot hand.In this experiment, purpose is that the error average$\pm 3\sigma$ is within a 10mm for position(x-y) and the orientation for angle $\theta$ is 10°.


Fig. 10. Target objects (No.1 ～ No.12)

Fig. 11 shows the error average and standard deviation all clothes. The error average $\pm$ $3\sigma$ ( standard deviation ) is within a 10 mm for position (x-y) and the orientation for angle $\theta$ is 10 °. As Fig. 11 , we achieved our purpose.

### B. Handling Experiment

When used in practice in the factory, handling of clothes must be able to reproduce the accuracy with which human beings are done by hand. It confirmed the accuracy of GA recognition in the previous chapter. We have conducted verification experiment of handling precision in this chapter. Specifically, the experiment was carried out with No. 2 clothes. In the experiment the robot adsorbed clothes with the

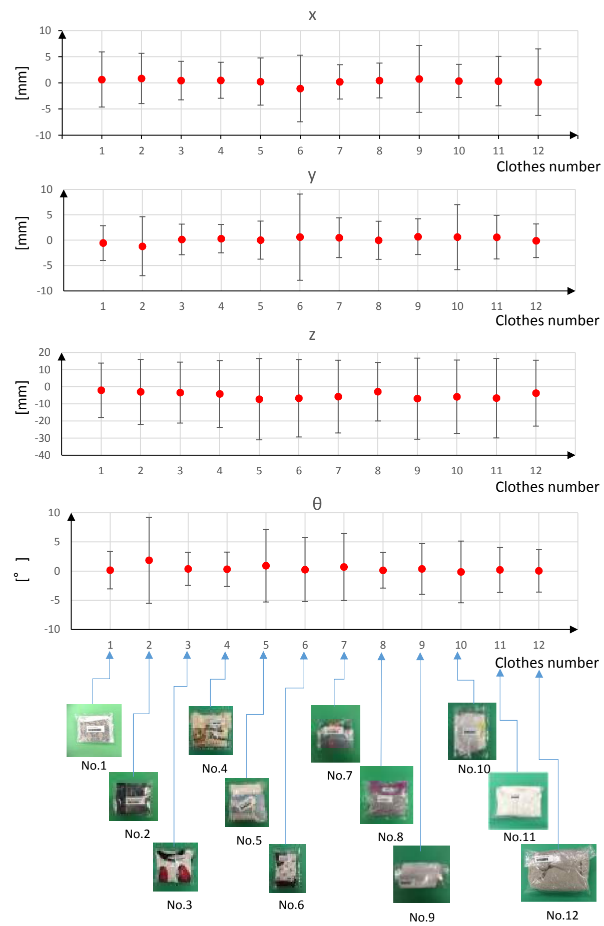


Fig. 11. 1000 times recognition experimental result

suction pad which is on the tip of the robot hand and handled it to a storage box. Then we analyze the obtained error data of the whole process. Fig. 12 shows flowchart of handling experiment. Shown below the sequence of experiments.

- First recognition
- Move to the pose (position and orientation) of the recognition
- Second recognition
- Handle based on the recognition result
- Robot's hand uplift to get the same pose relationship between clothes and hand as the second recognition
- Third recognition
- Place the clothes on the first recognition plane randomly

We have taken experiment in 100, 400, 700, 1000 and 1300Lx lighting condition. In each illuminance the experiment was carried out 100 times. And the third time recognition result was obtained in every experiment. In this experiment, purpose is that the error average$\pm 3\sigma$ is within a 15mm for position(x-y) and the orientation for angle $\theta$ is 15°.

Fig. 13 shows the error average and standard deviation illuminance of 100Lx ～ 1300Lx. The error average $\pm$ $3\sigma$ ( standard deviation ) is within a 15 mm for position (x-y) and the orientation for angle $\theta$ is 15 °. As Fig. 13 , we achieved

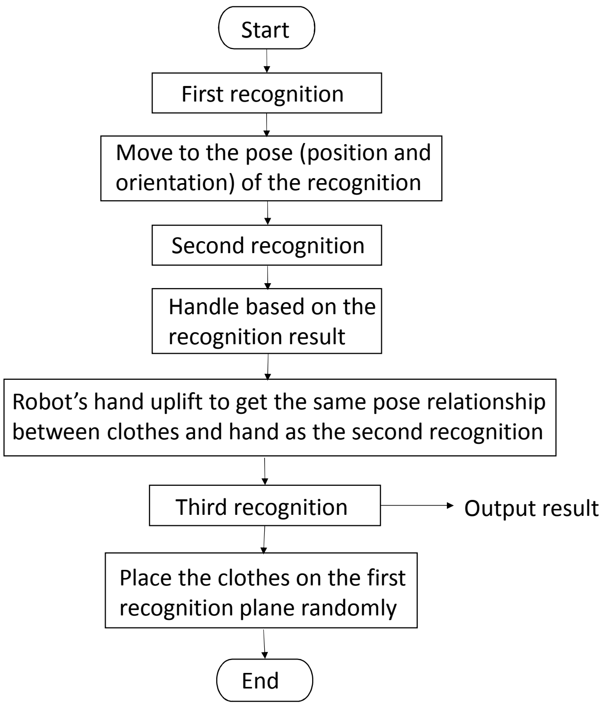


Fig. 12. Flowchart of handling experiment

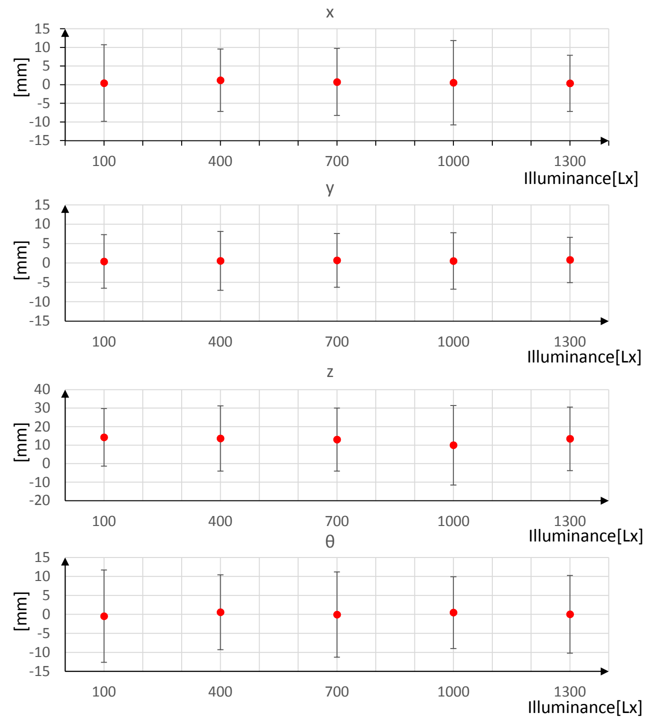


Fig. 13. Handling experimental result

our purpose.

## V. CONCLUSIONS

In this paper, we introduced new cloth model generation method and pose estimation method using model-based clothes recognition. The merits of the photo-model-based clothes recognition system are shown as follow:

- Photo-model-based allows the deformable different clothes recognized automatically.
- 3D-pose measurement is possible.
- Photo-model-based 3D-pose recognition is not limited to clothing, and any object can recognized by using proposed system.

By using a photo model of each clothes, the target object with different shape and pattern can be recognized. We conducted 1000 times of experiment using 12 different clothes with different colorful pattern and multiple size. Recognition accuracy is analyzed in term of fitness distribution, histogram of pose estimation error. According to the experimental results, it can be confirmed that pose estimation of clothes for mass production can be implemented successfully using proposed photo-model-based clothes recognition system.

## APPENDIX

Appendixes should appear before the acknowledgment.

## ACKNOWLEDGMENT

## REFERENCES

[1] W. Song, M. Minami, Y. Mae, S. Aoyagi, On-line Evolutionary Head Pose Measurement by Feedforward Stereo Model Matching,*International Conference on Robotics and Automation (ICRA)*.

[2] J. Stavnitzky, D. Capson, Mutiple Camera Model-Based 3-D Visual Servoing , *IEEE Trans. on Robotics and Automation*, vol. 16, no. 6, December 2000.

[3] C. Dune, E. Marchand, C. leroux, One Click Focus with Eye-inhand/Eye-to hand Cooperation" , *IEEE Int. Conf. on Robotics and Automation (ICRA)*, pp.2471-2476, 2007.

[4] Wei Song, Mamoru Minami, Fujia Yu, Yanan Zhang and Akira Yanou : " 3-D Hand and Eye-Vergence Approaching Visual Servoing with Lyapunouv-Stable Pose Tracking" , *IEEE International Conference on Robotics and Automation Shanghai International Conference Center*, May 9-13, 2011, Shanghai China.

[5] S. Nakamura, K. Okumura, A. Yanou and M. Minami : "Visual Servoing of Patient Robot's Face and Eye-Looking Direction to Moving Human," *SICE Annual Conference* pp.1314-1319 2011.

[6] Myo Myint, Kenta Yonemori, Akira Yanou, Shintaro Ishiyama and Mamoru Minami, "Robustness of Visual-Servo against Air Bubble Disturbance of Underwater Vehicle Using Three-dimensional Marker and Dual-eye Cameras", *International Conference OCEANS 15 MTS/IEEE, Washington DC, USA*, October 19-22, 2015.

[7] Koichi Maeda, Mamoru Minami, Akira Yanou, Hiroaki Matsumoto, Fujia Yu, Sen Hou: "Frequency Response Experiments of 3-D Full Tracking Visual Servoing with Eye-Vergence Hand-Eye Robot System *"SICE Annual Conference* pp.101-107, 2012.

[8] Y. Cui, X. Li, M. Minami, A. Yanou, M. Yamashita and S. ISHIYAMA, "Robot for Decontamination with 3D Move on Sensing", *Nuclear Safety and Simulation*, Vol. 6, No. 2, pp. 142-154 , June 2015.

[9] Yinxiao Li, Chih-Fan Chen, and Peter K.Allen, "Recognition of Deformable Object Category and Pose", *Robotics and Automation (ICRA), 2014 IEEE International Conference*, pp.5558-5564, May 31-June 7, 2014, Hong Kong.

[10] Kimitoshi Yamazaki, Masayuki Inaba, " A Cloth Detection Method Based on Image Wrinkle Feature for Daily Assistive Robots", *MVA 2009 IAPR Conference on Machine Vision Applications*, May 20-22, 2009, Yokohama, Japan.

[11] S.Miller, M.Fritz, T.Darrell and P.Abbeel. "Parametrized Shape Models for Clothing", *IEEE International Conference on Robotics and Automation (ICRA)*.