

# Verification of photo-model-based pose estimation and handling of unique clothes under illumination varieties

Khaing Win PHYU\*, Ryuki FUNAKUBO\*, Ryota HAGIWARA\*, Hongzhi TIAN\* and Mamoru MINAMI\*

\* Graduate School of Natural Science and Technology, Division of Mechanical and Systems Engineering, Okayama University

3-1-1 Tsushimanaka, Kita-ku, Okayama, 700-8530, Japan

E-mail: p6nn3hpz@s.okayama-u.ac.jp

Received: 4 September 2017; Revised: 5 February 2018; Accepted: 10 May 2018

## Abstract

Human can recognize and handle (pick and place) easily the objects with a variety of different shapes, colors, sizes, and humans' eyes are adaptable to various light environments with a certain tolerance. However, it is difficult for robots to recognize deformable objects such as cloth, string, etc., especially if an object is unique. Additionally, there have been difficulties for robots with vision sensors (cameras) to accurately detect and handle objects under various light environments. This paper proposes a cloth handling system that recognizes a unique cloth appeared in front of a robot by a photo-model-based approach. The photo-model-based approach has been adopted since the photo-model can be made at once by taking a photo of the unique cloth. In proposed clothes' pose estimation method, a photo-model projected from 3D to 2D is used, where this system does not need defining the object's size, shape, design, color and weight. It detects the cloth through model-based matching method and Genetic Algorithm (GA). The handling performance by the proposed method with dual-eyes cameras has been verified, revealing that the proposed system has leeway to recognize and handle the unique cloth in lighting varieties from 100 lx to 1300 lx. In addition, 3D recognition and handling accuracy have been confirmed to be practically effective by conducting the recognition/handling experiments under different light conditions.

**Keywords** : Photo-model-based cloth recognition, Handling, Visual servoing, Genetic Algorithm, Dual-eyes cameras, Illumination

## 1. Introduction

From the instant of birth, human beings are thought to be talented at managing their activities under such variability as climates, light environments, temperatures, etc. While human beings can conduct intended tasks in pending circumstances, an automated robot is not adept at being similarly adaptable. Therefore, the researchers have tried to improve the abilities of automated robots.

Nowadays, industrial robots have been utilized to perform a wide variety of tasks instead of human workers. These automated robots are required to handle a wide variety of deformable objects including cloths, strings, ropes, electric cables and so on. Of course, handling deformable objects is difficult than handling rigid objects. A robot control technology using visual information, called as visual servoing, has been playing an important role in the applications where deformable things are recognized and handled by a robot.

Each item of the deformable target objects has various possibilities of the poses (positions and orientations) to be recognized and handled, requiring ability with respect to both vision-based recognition and visual servoing. In (Maitin-Shepard J et al., 2010), the cloth-grasping points are detected using four cameras without using other sensors for a towel folding application by robots. The main task in (Maitin-Shepard J et al., 2010) is to detect the corners of cloth instead of recognition a whole cloth. The recognition of cloth shape based on strategic observation during handling was reported in (Yinxiao Li et al., 2014). Multi-views from trinocular stereo vision system were used in (Yinxiao Li et al., 2014). The appearance of deformable cloth is susceptible not only to how they are placed in the view of the camera but also the light condition that is one of main difficulties in visual servoing. Even though deformable cloths are handled in (Maitin-Shepard

J et al., 2010) and (Yinxiao Li et al., 2014), 3D pose estimation of deformable cloths under different light conditions has not been discussed in almost studies.

On the other hand, problems in recognizing a 3D solid object and estimating the object's pose have been thought to be a conundrum. Estimating poses of solid objects uses usually dual-eyes recognition methodology (W. E. L. Grimson, 1981)-(Z. Zhang et al., 1995). The dual-eyes image processing exploits epipolar geometry (W. E. L. Grimson, 1981) to reduce the dimension of searching space that is defined in the dual-eyes images (Z. Zhang et al., 1995). Even though epipolar geometry is effective for 3D image processing, the problem called "Corresponding Points Identification Problem" that is how to make a point in one camera image correctly correspond to a point in another camera image – to confirm whether the both two points in dual cameras' images represent a point on the 3D target object –, has yet to be solved (T. Poggio and S. Edelman, 1990), (S. Ullman and R. Basri, 1991). Then the reconstruction of an object's pose in 3D space from dual-eyes cameras has been still difficult, and estimated 3D space inevitably includes pose error.

Based on the above discussions, estimating 3D-pose of solid target with dual cameras of solid target seems not easy, then pose estimation and handling deformable cloth deems challenging. In this circumstances of robotics researches, the authors have been requested from T2K Co., Ltd (Logistics company to deal with cloths) to construct a robotic system to handle the single/unique second-hand cloths. By considering the above backdrop, a photo-model-based cloth recognition method has been devised (Funakubo R et al., 2016) - (Phyu KW et al., 2016), since the photo-model can be made at once by taking a photo of the unique cloth.

The aim of the proposed system is to be applied in a mail delivery system of cloth for T2K Co., Ltd, in which as of today employees classify a large number of second-hand cloths manually. Since the cloths are deformable objects and the cloth is single/unique, no definition of cloths can be predefined in a computer. Consequently, it is difficult to handle a wide variety of cloths that are irregular and unique in shape and size. T2K explained this condition to authors and requested to collaborate with our research group to develop a cloth handling robot. Different 12 cloths samples have been chosen by T2K so that those 12 cloths may represent enormous cloths varieties.

Despite that the authors has made efforts to find published papers concerning vision-based robot system using photo-model, we could not discover any related paper. Since the practicality of industrial robots used in factories has been indispensable, the pick and place accuracy with tolerances against illumination changing has been checked (Funakubo R et al., 2016 and Funakubo R et al., 2017). This paper is an extension of (Funakubo R et al., 2016 and Funakubo R et al., 2017), but it includes new evaluation results of illumination varieties and fitness function distribution, which can explain why the proposed photo-model-based system can be tolerable against illumination varieties. In our previous study, two different light sources, fluorescent light and light-emitting diode (LED), were used to provide with different illuminations environments. Why these illumination conditions were chosen has been discussed in (Phyu, Khaing Win, et al., 2017). According to the experimental results, the recognition performances under fluorescent light are better than light-emitting diode.

The rest of the present paper is organized as follows: Section 2 describes the system configuration, Section 3 presents the photo-model-based recognition, Section 4 describes the experimental environment, Section 5 describes the experimental contents with the results, followed by conclusion in Section 6.

## 2. System configuration

The developed vision-based robot system is shown in Fig. 1. In Fig. 1, the dual-eyes cameras that are fixed at the end-effector of a PA-10 robot perform the cloth recognition and pose estimation process based on the digital photo model. (Note that the term "photo model" will be used throughout the present paper from now on to shorten the word and to align with our previous papers.) The cloth absorption pads as shown in Fig. 1 are possible to perform the absorption of the target object (cloth). The aim of the PA-10 robot using proposed system is to pick up the cloth after recognition it and set the cloth into a desired collection box as shown in Fig. 1. Figure 2 represents the cloth handling system that consists of robot with a cloth's pose measuring dual-eyes sensor unit, which includes transport conveyor. There are three cameras (vision sensors) in this configuration. After a cloth being input on the conveyor from right-hand side of the figure, the single camera makes the photo model. The model would be used to recognize the cloth and to measure the cloth's pose at the left-hand side, where the cloth is picked up and set inside an mailing box by the robot.

An unique cloth may be recognized by barcode if attached at the cloth, but the photo-model-based whole cloth recognition is indispensable for non-erroneous identification of the unique cloth. Of course the photo-model-based recognition may be combined with barcode for enhanced reliability. After recognizing and estimating the pose of the target

cloth, handling the cloth was performed to pick and place the cloth. The proposed photo-model-based cloths recognition system is intended to save the cost of staff workers and to get better performance and higher accuracy than human workers. Moreover, this system is aimed at being applied in the real world, regardless of lighting conditions varieties. Therefore, the robustness of the proposed system against different illuminations was verified experimentally in this study.

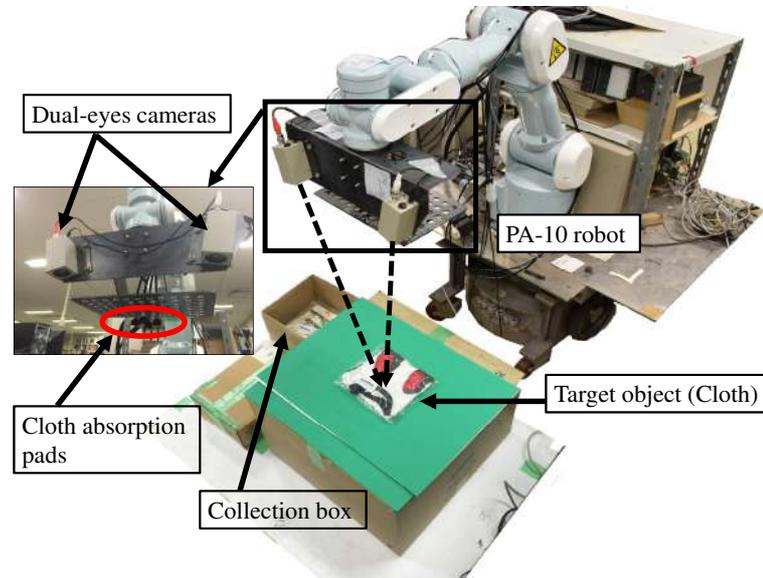


Fig. 1 A photo of a cloth handling robot system with dual-eyes cameras: PA-10 robot is equipped with two cameras (vision sensors are used as dual-eyes vision system) for recognition and vacuum cups (four absorption pads by the air compressor possible to perform the pick and place of the target object (cloth)) for handling. In the test, the robot picks up the cloth and places it into the collection box.

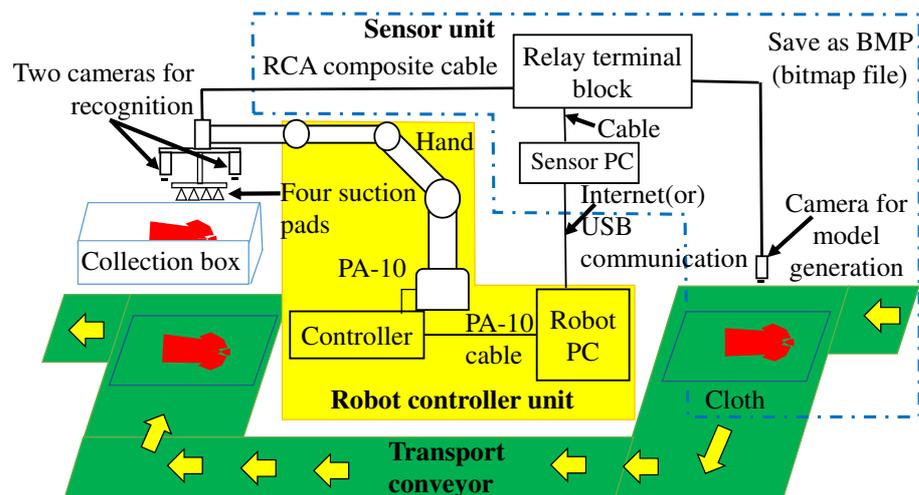


Fig. 2 System configuration of a cloth handling system: During the application process, the cloth moves along the conveyor automatically. The single camera set at the entry point of cloth in the beginning of conveyor is used to generate a photo model. The left and right cameras which are attached at the end-effector of the PA-10 robot are used to recognize and estimate the pose of the cloth that appears in the field of view of two cameras using generated model. The cloth handling application of Fig. 2 is as same as the application process of Fig. 1.

### 3. Photo-model-based recognition

This section is to explain a photo-model-based cloth recognition method. The model-based matching method (Minami M et al., 2003) has been utilized with the adoption of a set-point-model-thinking. That is all points of the solid 3D model in the 3D searching space as a group are projected onto the left and right camera image planes (2D images) without the Corresponding Points Identification Problem that has been pointed out as the difficulty existing in pose estimation by using plural cameras. Since all points on 3D model are projected into 2D camera images in our method, all projections for each point are correct. This means that forward projection of 3D object has been used, which does not raise up Corresponding Points Identification Problem. The following is description of the kinematics of stereo-vision before an explanation of the proposed system in details.

#### 3.1. Kinematics of stereo-vision

Figure 3 shows a perspective projection of the dual-eyes vision system. The coordinate systems of dual-eyes cameras and the target object (cloth) in Fig. 3 consist of world coordinate system  $\Sigma_W$ , j-th model coordinate system  $\Sigma_{Mj}$ , hand coordinate system  $\Sigma_H$ , camera coordinate systems as  $\Sigma_{CL}$  and  $\Sigma_{CR}$ , and image coordinate systems as  $\Sigma_{IL}$  and  $\Sigma_{IR}$ . In Fig. 3, the position vectors of an arbitrary i-th point of the j-th 3D model  $\Sigma_{Mj}$  based on each coordinate system are as follows:

- ${}^W\mathbf{r}_i^j$ : position of an arbitrary i-th point on j-th 3D model based on  $\Sigma_W$
- ${}^M\mathbf{r}_i^j$ : position of an arbitrary i-th point on j-th 3D model in  $\Sigma_{Mj}$ , where  ${}^M\mathbf{r}_i^j$  is constant vector
- ${}^{CR}\mathbf{r}_i^j$  and  ${}^{CL}\mathbf{r}_i^j$ : position of an arbitrary i-th point on j-th 3D model based on  $\Sigma_{CR}$  and  $\Sigma_{CL}$
- ${}^{IL}\mathbf{r}_i^j$  and  ${}^{IR}\mathbf{r}_i^j$ : projected position on  $\Sigma_{IL}$  and  $\Sigma_{IR}$  of an arbitrary i-th point on j-th 3D model

The homogeneous transformation matrix from the right camera coordinate system  $\Sigma_{CR}$  to the target object coordinate system  $\Sigma_M$  is defined as  ${}^{CR}\mathbf{T}_M(\phi_M^j, \mathbf{q})$ , where  $\phi_M^j$  is j-th model's pose and  $\mathbf{q}$  means robot's joint angle vector. Then,  ${}^{CR}\mathbf{r}_i^j$  can be calculated by using Eq. (1),

$${}^{CR}\mathbf{r}_i^j = {}^{CR}\mathbf{T}_M(\phi_M^j, \mathbf{q}) {}^M\mathbf{r}_i^j. \quad (1)$$

where  ${}^M\mathbf{r}_i^j$  is predetermined as fixed vectors since  $\Sigma_{Mj}$  is fixed on the j-th model.  ${}^{CL}\mathbf{r}_i^j$  that represents the same i-th point on j-th model based on  $\Sigma_{CL}$  is also calculated by using  ${}^{CL}\mathbf{T}_M(\phi_M^j, \mathbf{q})$ . Since  $\mathbf{q}$  can be measured by robot's joint sensors, it could be thought to have been known, then  $\mathbf{q}$  is omitted hereafter. Equation (2) represents the projective transformation matrix  $\mathbf{P}_k$ .

$$\mathbf{P}_k = \frac{1}{k_{z_i}} \begin{bmatrix} \frac{f}{\eta_x} & 0 & {}^I x_0 & 0 \\ 0 & \frac{f}{\eta_y} & {}^I y_0 & 0 \end{bmatrix}. \quad (2)$$

where,

- $k = CL, CR$ ,
- $k_{z_i}$ ; position of the i-th point in the camera sight direction in  $\Sigma_{CR}$  and  $\Sigma_{CL}$ ,
- $f$ ; focal length,
- $\eta_x$ ; [mm/pixel] in x-axis,
- $\eta_y$ ; [mm/pixel] in y-axis.

The position vector of the i-th point in the right and left camera image coordinates  ${}^{IR}\mathbf{r}_i^j$  can be described by using  $\mathbf{P}_k$  as,

$${}^{IR}\mathbf{r}_i^j = \mathbf{P}_k {}^{CR}\mathbf{r}_i^j = \mathbf{P}_k {}^{CR}\mathbf{T}_M(\phi_M^j) {}^M\mathbf{r}_i^j \quad (3)$$

Then,  ${}^{IR}\mathbf{r}_i^j$  can be described as,

$$\begin{cases} {}^{IR}\mathbf{r}_i^j(\phi_M^j) = \mathbf{f}_R(\phi_M^j, {}^M\mathbf{r}_i^j) \\ {}^{IL}\mathbf{r}_i^j(\phi_M^j) = \mathbf{f}_L(\phi_M^j, {}^M\mathbf{r}_i^j) \end{cases} \quad (4)$$

where  ${}^{IL}\mathbf{r}_i^j$  can also be described as the same manner like  ${}^{IR}\mathbf{r}_i^j$ .

#### 3.2. Cloth model generation

There are two main portions in the proposed robot handling system. The first portion is for cloth model generation and the latter is for relative pose estimation of actual cloths using generated model through model-based matching method. This subsection is to describe the first portion.

The hue value in HSV color representation is used to describe cloths' photo-model that comprises positions of plural pixel dots and the hue value of each dot. By using the hue value, it is possible to make the color recognition tolerable

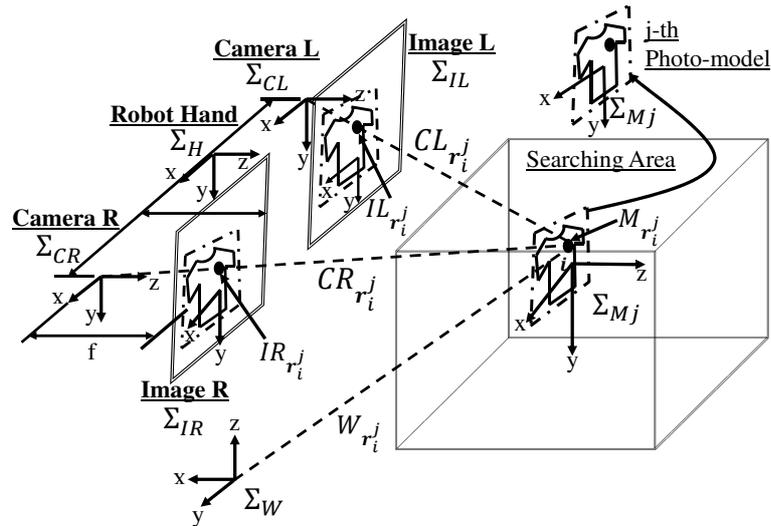


Fig. 3 Perspective projection of dual-eyes vision-system: In the searching area, a 3D solid model is represented by the picture of cloth with black point ( $j$ -th photo-model). The coordinate systems of photo-model, camera and image are represented by  $\Sigma_{Mj}$ ,  $\Sigma_{CL}$ ,  $\Sigma_{CR}$ ,  $\Sigma_{IL}$  and  $\Sigma_{IR}$  respectively. A 3D solid model that is assumed to be in the searching area is projected from 3D space to 2D left and right camera images.

against the lighting condition varieties, which is given by the character of HSV representation.

In the proposed system, three cameras are used as the vision sensors. Among them, the first camera is used for generating a model, which is depicted at the right-hand-side in Fig. 2. The model generation process is represented in Fig. 4. Firstly, a background image is captured by the first camera as shown in Fig. 4 (a) and the averaged hue value of the background image is calculated. Then, the cloth comes in on the conveyor with the green background as shown in Fig. 4 (b). In Fig. 4 (c), the hue value of each predefined pixel point in the image is compared with the averaged hue value of the background image, then the area of cloth is detected. The set made by dots on the cloth is named as  $S_{in}$ . The dots defined in enveloping strip around the  $S_{in}$  constitutes a set of  $S_{out}$ . The combined set  $S$  with  $S_{in}$  and  $S_{out}$  made of dots' position and the hue color of the dots represents the photo-model that is used for recognizing the single/unique cloth and estimating the cloth's pose. The procedure is explained next subsection.

### 3.3. 3D model-based matching

3D pose of the 3D model, including three positions and one orientation represented by angle, are defined as  $\phi_M^j = [x_M, y_M, z_M, \theta_M]^T$ . The angle  $\theta$  is an angle around the normal direction of the clothing bench, that is  $\theta$  is around  $z$ -axis of  $\Sigma_{Mj}$  as shown in Fig. 3. The upper side of the Fig. 5 shows the appearance of a generated 3D solid model in the 3D searching space, and the left and right 2D searching models (sub-figures on the left and right bottom of Fig. 5) are projected into 2D image planes.  $S(\phi_M^j)$  is made of  $S_{in}(\phi_M^j)$  (inner dotted points) and the outside strip  $S_{out}(\phi_M^j)$  enveloping  $S_{in}(\phi_M^j)$  denoted by outer dotted line. The sub-figures on the left/right bottom of Fig. 5 show the left/right 2D searching models  $S_L(\phi_M^j)$  and  $S_R(\phi_M^j)$ , where those two models are projected by using  $\phi_M^j$  that represents the pose of  $j$ -th model in the evolution process of Genetic Algorithms as one of genes.

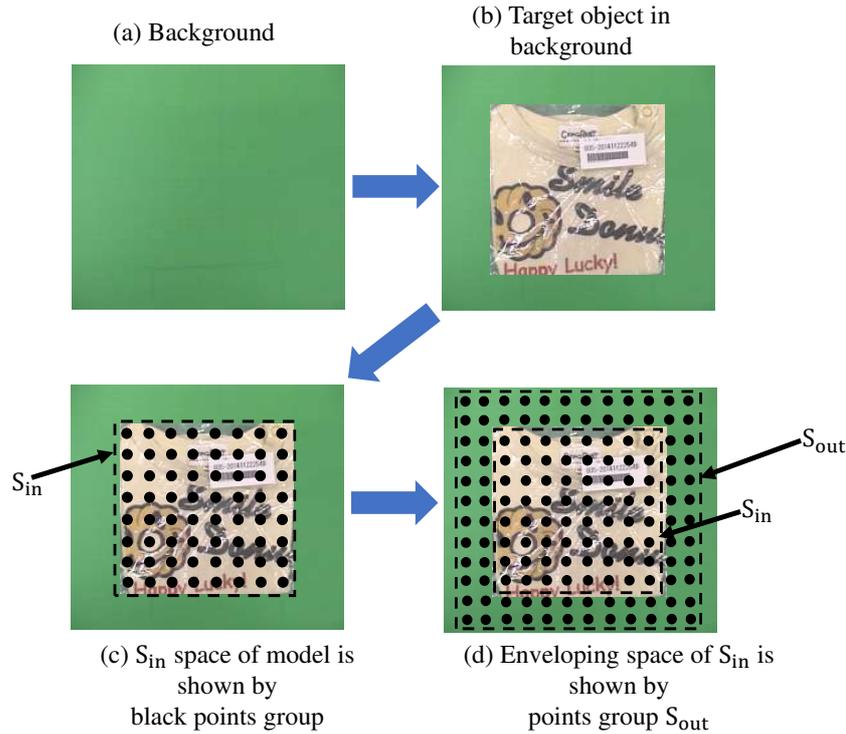


Fig. 4 Model generation process: (a) get background image and calculate the hue value, (b) get image including cloth and calculate the hue value of each point, (c) distinguish background from difference between the hue value, (d) determine the frame of model. Generated surface space of model and outside space of model are denoted as  $S_{in}$  and  $S_{out}$ .

### 3.4. Definition of the fitness function

The correlation function between the projected model and the actual cloth images input from the dual-eyes cameras attached at the end-effector is used as a fitness function (Minami M et al., 2003). In the fitness distribution, the valuable of position and orientation to give the highest peak represents the best pose of the model that coincides with the captured cloth's images from the left and right cameras as shown in Fig. 5. Then the pose  $\phi_M^j$  that gives the peak can be thought to be representing the true pose of the target cloth that is placed in the 3D searching space as shown in Fig. 5. The concept of the fitness function in this study can be said to be an extension of the work in (Minami M et al., 2003), in which different models including a rectangular shape surface-strips model was evaluated using images from a single camera. The correlation between the projected models including a pose of  $\phi_M^j$  and captured images with actual cloth that were projected on the left and right 2D searching areas is calculated by Eqs. (5) - (7).  $F(\phi_M^j)$  is calculated by averaging the fitness functions of both left camera image  $F_L(\phi_M^j)$  and right camera image  $F_R(\phi_M^j)$  as shown in Eq. (5).

$$F(\phi_M^j) = \left\{ \left( \sum_{IRr_i^j \in S_{R.in}(\phi_M^j)} p_{R.in}(IRr_i^j(\phi_M^j)) + \sum_{IRr_i^j \in S_{R.out}(\phi_M^j)} p_{R.out}(IRr_i^j(\phi_M^j)) \right) + \left( \sum_{ILr_i^j \in S_{L.in}(\phi_M^j)} p_{L.in}(ILr_i^j(\phi_M^j)) + \sum_{ILr_i^j \in S_{L.out}(\phi_M^j)} p_{L.out}(ILr_i^j(\phi_M^j)) \right) \right\} / 2$$

$$= \{F_R(\phi_M^j) + F_L(\phi_M^j)\} / 2 \quad (5)$$

The points on a 3D matching model,  $S(\phi_M^j)$  are projected to the left and right image plane. The projected points on  $S_{in}(\phi_M^j)$  and  $S_{out}(\phi_M^j)$  to the left camera image are described as  $ILr_i^j \in S_{L.in}(\phi_M^j)$  and  $ILr_i^j \in S_{L.out}(\phi_M^j)$  respectively. For detailed explanation of Eq. (5), the following definitions should be stated here.

- $S_{L.in}$ ; the inside area projected to left image plane,
- $S_{L.out}$ ; the space on a strip area surrounding  $S_{L.in}$ ,
- $H_{IL}(ILr_i^j(\phi_M^j))$ ; the hue value of the left camera image at the point  $ILr_i^j(\phi_M^j)$ ,
- $H_{ML}(ILr_i^j(\phi_M^j))$ ; the hue value of the model at the point  $ILr_i^j(\phi_M^j)$  (i-th point on the j-th model),

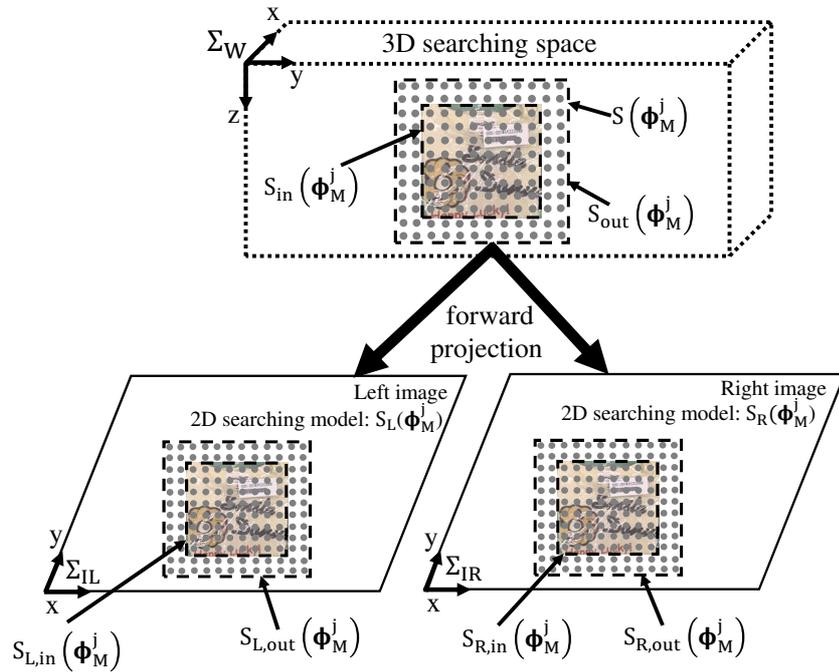


Fig. 5 A 3D solid model in the 3D searching space (sub-figure on the top of Fig. 5) and left and right 2D searching models represented as  $S_L(\Phi_M^j)$  and  $S_R(\Phi_M^j)$  (sub-figures on the left/right bottom of Fig. 5).

- $\bar{H}_B$ ; the average hue value of the background image.

The next Eqs. (6) and (7) is used for calculating  $p_{L,in}(\mathcal{I}^L \mathbf{r}_i^j(\Phi_M^j))$  and  $p_{L,out}(\mathcal{I}^L \mathbf{r}_i^j(\Phi_M^j))$  that are included in Eq. (5).

$$p_{L,in}(\mathcal{I}^L \mathbf{r}_i^j(\Phi_M^j)) = \begin{cases} 2, & \text{if } (|H_{IL}(\mathcal{I}^L \mathbf{r}_i^j(\Phi_M^j)) - H_{ML}(\mathcal{I}^L \mathbf{r}_i^j(\Phi_M^j))| \leq 30); \\ -0.005, & \text{if } (|\bar{H}_B - H_{ML}(\mathcal{I}^L \mathbf{r}_i^j(\Phi_M^j))| \leq 30); \\ 0, & \text{otherwise.} \end{cases} \quad (6)$$

$$p_{L,out}(\mathcal{I}^L \mathbf{r}_i^j(\Phi_M^j)) = \begin{cases} 0.1, & \text{if } (|\bar{H}_B - H_{ML}(\mathcal{I}^L \mathbf{r}_i^j(\Phi_M^j))| \leq 20); \\ -0.5, & \text{otherwise.} \end{cases} \quad (7)$$

Equations (6) and (7) are designed to provide a peak in the fitness value distribution,  $F(\Phi_M^j)$  when  $\Phi_M^j$  coincides with the true pose of the target cloth, which has been confirmed by (Phyu KW et al., 2016). Figure 6 (a) shows j-th model, real cloth (target object), the evaluation points of Hue value,  $\dots \mathcal{I}^L \mathbf{r}_{i-1}^j(\Phi_M^j), \mathcal{I}^L \mathbf{r}_i^j(\Phi_M^j), \mathcal{I}^L \mathbf{r}_{i+1}^j(\Phi_M^j) \dots$ , in inside area  $S_{L,in}$ , and those in outside strip  $S_{L,out}$ . Figure 6 (b) shows a situation that the overlapping area of real cloth and the model increased than the one depicted in (a). The hue value of the left camera input image at the point  $\mathcal{I}^L \mathbf{r}_i^j(\Phi_M^j)$  is represented by  $H_{IL}(\mathcal{I}^L \mathbf{r}_i^j(\Phi_M^j))$ , and the i-th point of j-th model in  $S_{L,in}$  and  $S_{L,out}$  and the hue value of the same point  $\mathcal{I}^L \mathbf{r}_i^j(\Phi_M^j)$  on the model is defined as  $H_{ML}(\mathcal{I}^L \mathbf{r}_i^j(\Phi_M^j))$ . The average hue value of background is defined as  $\bar{H}_B$ .

In Eq. (6), if the hue value of each point of captured images,  $H_{IL}(\mathcal{I}^L \mathbf{r}_i^j(\Phi_M^j))$ , which lies inside the surface model frame  $S_{L,in}$ , and the hue value of corresponding same point in a model,  $H_{ML}(\mathcal{I}^L \mathbf{r}_i^j(\Phi_M^j))$ , have similar values with a tolerance less than 30, that is  $|H_{IL}(\mathcal{I}^L \mathbf{r}_i^j(\Phi_M^j)) - H_{ML}(\mathcal{I}^L \mathbf{r}_i^j(\Phi_M^j))| \leq 30$  then this means that model's hue value and input image's hue value have close hue distance at the same checking point of  $\mathcal{I}^L \mathbf{r}_i^j(\Phi_M^j)$ . This represents photo model overlaps to the real cloth projected in left camera image in  $S_{in}$ , which are represented by dots designated by (A) in Fig. 6 (b). In this case the fitness value would be increased with the voting value of "+2." The fitness value will decrease with the value of "-0.005" for every point of cloth in the left camera image, if model's point  $\mathcal{I}^L \mathbf{r}_i^j(\Phi_M^j)$  positions on the background. This situation is described by the condition,  $|\bar{H}_B - H_{ML}(\mathcal{I}^L \mathbf{r}_i^j(\Phi_M^j))| \leq 30$  in Eq. (6), and this means model's clothes area overlaps with green background and it also represents that the model does not overlap precisely, which are represented by (B) in Fig. 6

(b). Then “-0.005” is given as a penalty to decrease  $F(\phi_M^j)$ . Otherwise, the fitness value will be “0.”

Similarly, in Eq. (7), if the hue value of each point in the left camera image lying in  $S_{L,out}$  has similar value to the average hue value of background  $\bar{H}_B$  with the tolerance of 20, the fitness value will be increased with the value of “+0.1.” This means  $S_{L,out}$  strip area surrounding  $S_{L,in}$  overlaps the green background, expressing the model and the clothes overlap rather correctly as (C) in Fig. 6 (b). Since this situation means that the model’s position and orientation matches to the real cloth, plus points “0.1” is given to the function  $p_{L,out}$ , which is described in Eq. (7). Otherwise, the fitness value will decrease with the penalty value of “-0.5.” This represents points on  $S_{L,out}$  overlaps with the real cloth as (D) in Fig. 6 (b). How designed fitness function explained above is effective and provides the robustness against illumination and lighting source varieties is described in section 5.2.

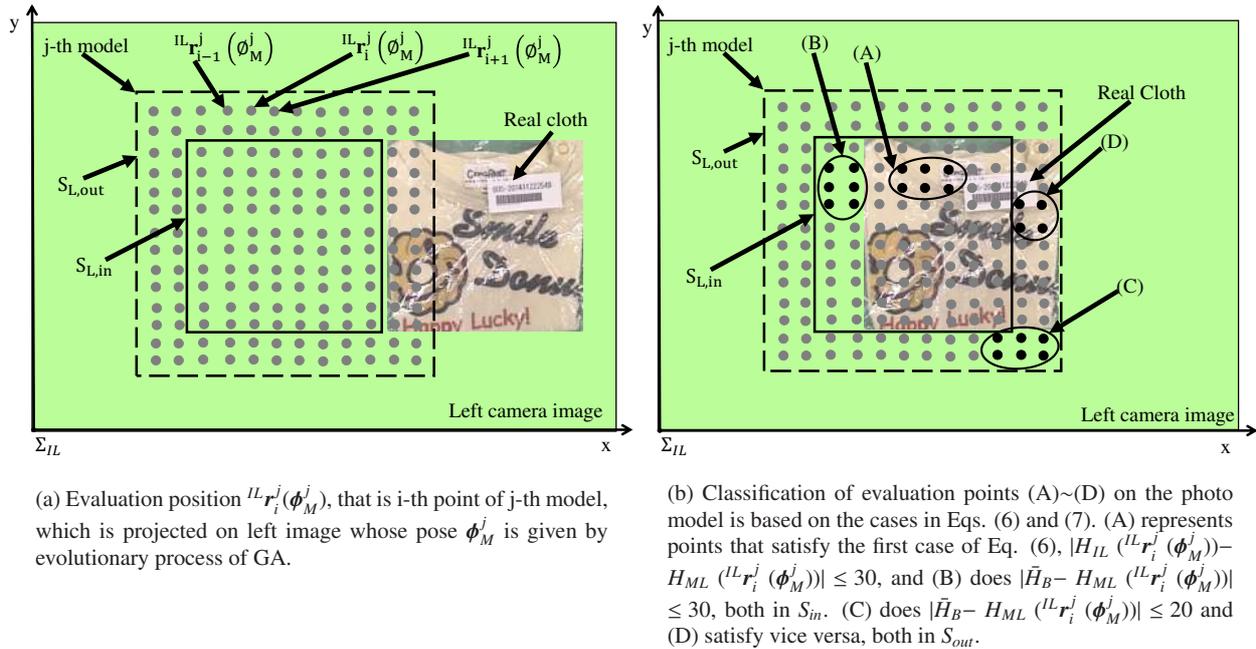


Fig. 6 Calculation of the matched degree of each point in model space ( $S_{L,in}$  and  $S_{L,out}$ )

### 3.5. Genetic Algorithm

The main problem of searching for the true pose of the object can be converted into an optimization problem because the fitness function has been designed to give the maximum value if and only if the pose of the model (GA’s gene represented by  $\phi_M^j$ ) and the target object coincide with the image in the 3D space (Phyu KW et al., 2016).

The maximum value of the fitness function can be searched by a number of optimization methods. Among them, Genetic Algorithm (GA) is applied to find the maximum value as an optimal solution because of its simplicity, effectiveness and easiness to implement.

Figure 7 shows the GA’s evolution process in which 3D models converge into the real target object (cloth). In Fig. 7, a target object is represented by cloth-shape and models are represented by a rectangular-shape with dotted lines. The models have same shape and same color information with the target object since the model is made of the photograph of the target cloth. But each model has a different pose  $\phi_M^j$  ( $j=1, 2, \dots, 60$ ) as shown in Fig. 7. The 60 individuals of GA are used in this experiment. Each individual’s chromosome consists of four variables. Each variable is coded by 12bits that can easily implement to get the optimal solution. The first three variables of a model in 3D space ( $t_x, t_y, t_z$ ) are represented as the position and the last one  $\theta$  means angle around z-axis of  $\Sigma_{Mj}$  shown in Fig. 3. And then, the genes of GA representing possible pose solution is defined as below;

$$\underbrace{01 \dots 01}_{12bits} \underbrace{00 \dots 01}_{12bits} \underbrace{11 \dots 01}_{12bits} \underbrace{01 \dots 10}_{12bits}.$$

These 60 individuals are evaluated by the fitness function value. The fitter ones are selected to regenerate the

next generations. In the final generation, the gene that gives the highest fitness value stands for the most trustful pose as shown at bottom-right part in Fig. 7. The spatial resolution in term of pixel is 0.89 [mm/pixel] for X position, and 0.92 [mm/pixel] for Y. The spatial resolutions in term of bit are 0.1 [mm/bit] for position, X, Y, Z and 0.00022 [quaternion/bit] for orientation,  $\theta$  in quaternion.

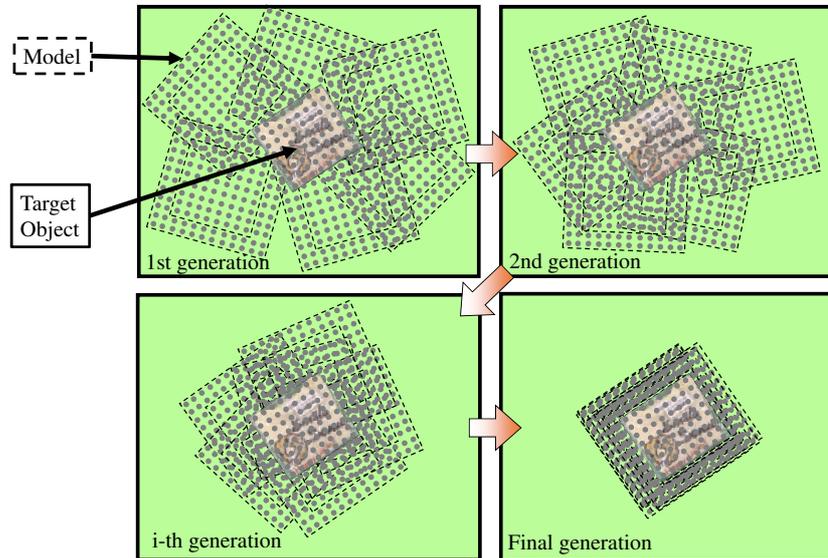


Fig. 7 GA evolution process in which 3D models with different poses converge to the real target object (cloth) through GA operation, and the pose of the model with the highest fitness function represents the estimated pose of the cloth.

#### 4. Experimental environment

Figure 8 shows the experimental layout and the coordinate systems of the cloth-handling robot, i.e. the world coordinate system ( $\Sigma_W$ ), the hand coordinate system ( $\Sigma_H$ ) and the cloth coordinate system ( $\Sigma_M$ ) that are used in the experiments respectively. The  $\Sigma_M$  that defines the center position and neutral orientation is set at  $x=0$  [mm],  $y=0$  [mm],  $z=685$  [mm], with respect to  $\Sigma_H$ . The offsets of the robot-base center and work table center in x-direction, y-direction and z-direction are described in corrected new Fig. 8. The explanation of  ${}^Wx_M$ ,  ${}^Wy_M$  and  ${}^Wz_M$  are described in the caption of Fig. 8. The structure of the robot also has been depicted in Fig. 9 with the lengths of all links being shown in the figure.

Figure 10 shows the 12 different cloths samples (No.1~ No.12) that have been chosen by collaborating company, T2K. Each cloth used in this experiment has different colors, sizes, shapes, and weights. This paper focuses on both the recognition tolerance in light varieties and handling accuracy. All 12 cloths are recognized and handled individually to confirm the influence of illumination on recognition and handling accuracy. The authors assumed that the clothing condition in the real implementation be packed in the plastic bags. The proposed method can use even reflections on plastic bag as a recognizable information in images. Even though the reflection could be changed at times depending on some happenings, and it may disturb the recognition, total accuracy could be within allowable extent for practical operations.

Since the thickness of the 12 cloths are different as shown in Table 1, the handling robot needs to detect the distance from the origin of the  $\Sigma_H$  to the surface of the vinyl package of the cloth. The authors have chosen dual-eyes pose estimation method as shown in Figs. 1 and 8 that has been used for visual servoing (CUI Y et al., 2015) ~ (Funakubo R et al., 2017), since the method has been proved to be practical and credible. Depending on the application of this proposed system, two cameras (vision sensors are used as dual-eyes vision system) for recognition and vacuum cups (four absorption pads by the air compressor possible to perform the absorption of the target cloth) for handling are attached at the end-effector of the PA-10 robot. The distance between two cameras is 323.4 [mm]. The positions of origin of  $\Sigma_M$  based on  $\Sigma_W$  are depicted as  $({}^Wx_M, {}^Wy_M, {}^Wz_M) = (-1050, 0, -180)$  [mm] in Fig. 8.

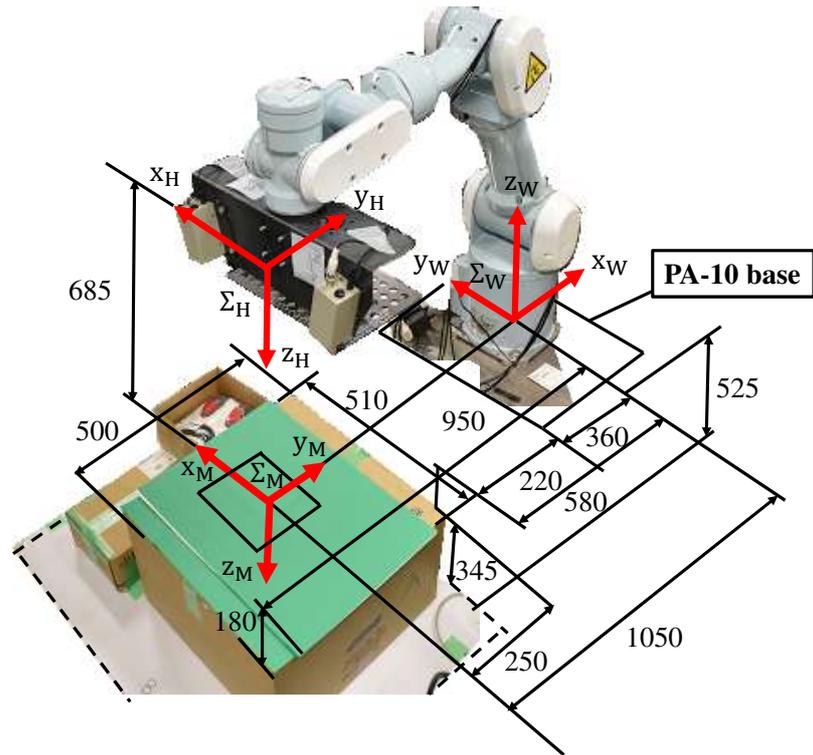


Fig. 8 Coordinate systems of robot and end-effector: (hand coordinate system ( $\Sigma_H$ ), world coordinate system ( $\Sigma_W$ ) and target object coordinate system ( $\Sigma_M$ )). Note that  ${}^W x_M, {}^W y_M, {}^W z_M$  are offsets of robot-base center and work table center in x-, y-, z-directions. (unit is [mm] in Fig. 8.)

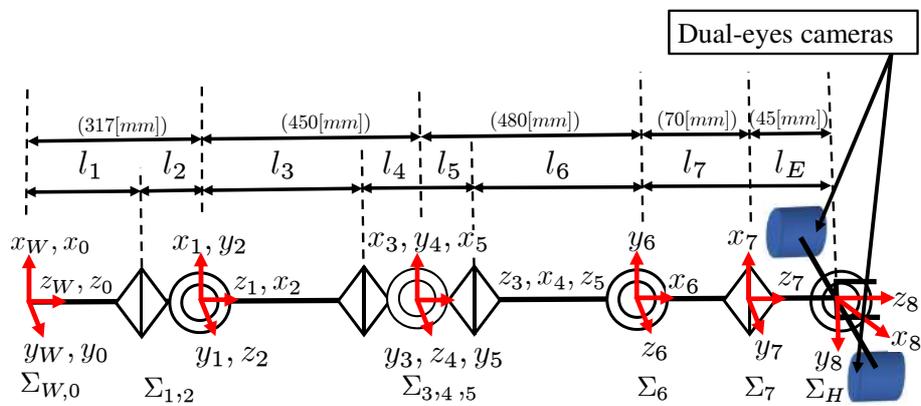


Fig. 9 Sketch map of the manipulator. (Note that the distances,  $l_2, l_4, l_5$  are zero [mm], but the joints are illustrated in figure to be seen clearly.)



Fig. 10 Target objects (No.1 ~ No.12) cloths: each has different colors, sizes, shapes, patterns and weights (unit is [mm] and g in Table 1).

Table 1 Size, thickness and weight of all cloths (No.1 ~ No.12).

No.	Size [mm]×[mm]	Thickness [mm]	Weight [g]
1	140×170	10	38
2	145×145	10	37
3	200×200	5	53
4	190×200	5	58
5	150×150	15	69
6	200×150	13	46
7	130×130	14	31
8	190×200	5	45
9	130×205	15	50
10	205×125	10	55
11	200×260	30	177
12	220×260	40	199

## 5. Experiments

### 5.1. Accuracy with fixed illumination

In this section, each pose ( $x, y, z$  in  $\Sigma_H$  in Fig. 8 and  $\theta$  around  $z$ -axis) of all cloths in Table 1 is estimated 1000 times repeatedly under fixed illumination of 700 lx. The average pose estimation errors and their extent of  $\pm 3\sigma$  of all cloths (No.1 ~ No.12) are shown in Fig. 10. Errors and the extent of  $\pm 3\sigma$  concerning  $x$  and  $y$  positions of all 12 cloths are less than 10 [mm], and those of  $\theta$  is less than 10 [degree] with the probability of 99.7%. However, the errors of  $z$  direction is less than 30 [mm], which is 3 times bigger than the cases of  $x$  and  $y$ . The numerical data of Fig. 11 are listed in Table 2. The frequency distributions concerning  $x, y, \theta$  of cloth No.6 are shown in Fig. 12. We have chosen No.6 to show actual data repeated 1000 times, where the reason of the choice of No.6 is that the cloth represents comparatively large  $3\sigma$  value in  $x, y$  and  $z$  direction. From the graph of (c) in Fig. 12, the recognition system tends to miscalculate the  $z$  value to be nearer than the one in fact (the minus value indicates nearer position).

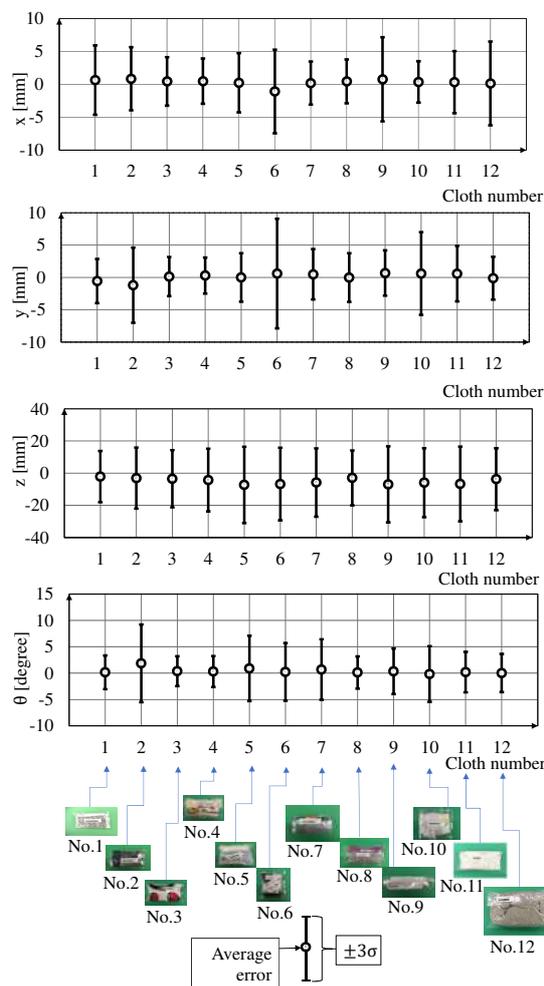


Fig. 11 Average error and  $\pm 3\sigma$  concerning three positions ( $x, y, z$ ) and one orientation ( $\theta$ ) of different cloths (No.1 ~ No.12) that have been confirmed by the experimental result of 1000 times recognition under fixed illumination 700 lx, where  $\sigma$  is standard deviation.

Table 2 Average error and standard deviation ( $3\sigma$ ) measured by 1000 times recognition experiments of different cloths (No.1 ~ No.12) under fixed illumination 700 lx.

Cloth No.	Average error : upper row Standard Deviation ( $3\sigma$ ) : lower row			
	$\bar{x}$ [mm]	$\bar{y}$ [mm]	$\bar{z}$ [mm]	$\theta$ [degree]
No.1	0.652	-0.565	-2.13	0.274
	5.28	3.42	15.9	3.84
No.2	0.858	-1.20	-3.06	1.86
	4.80	5.82	18.9	7.38
No.3	0.449	0.136	-3.46	0.379
	3.69	3.03	17.8	2.83
No.4	0.0494	0.294	-4.26	0.31
	3.45	2.79	19.5	2.94
No.5	0.244	0.00557	-7.31	0.908
	4.50	3.75	23.7	6.21
No.6	-1.08	0.601	-6.75	0.235
	6.36	8.49	22.5	5.49
No.7	0.197	0.488	-5.80	0.691
	3.27	3.90	21.2	5.76
No.8	0.453	-0.0191	-2.94	0.128
	3.33	3.75	17.0	3.06
No.9	0.764	0.683	-6.95	0.360
	6.39	3.51	23.6	4.35
No.10	0.373	0.598	-5.91	-0.159
	3.15	6.39	21.4	5.28
No.11	0.346	0.586	-6.70	0.205
	4.71	4.29	23.2	3.84
No.12	0.137	-0.120	-3.74	0.0310
	6.36	3.30	19.2	3.63

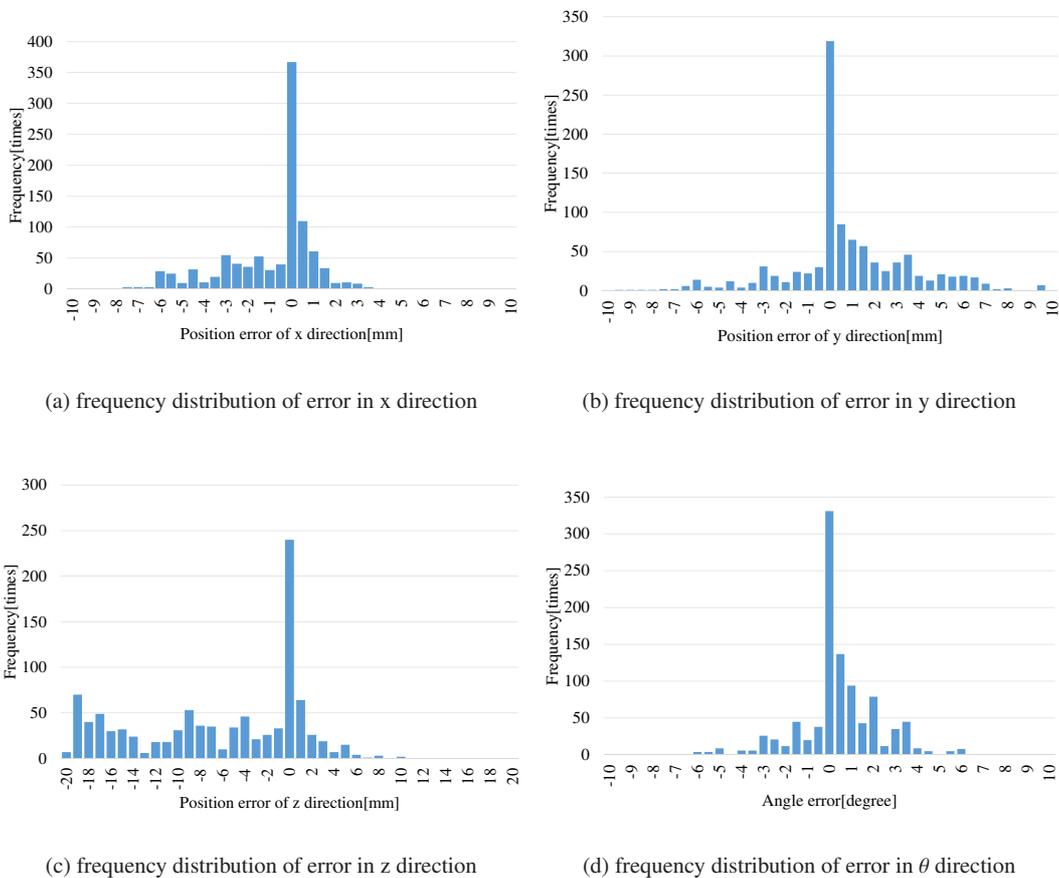


Fig. 12 Frequency distribution concerning x, y, z and  $\theta$  of cloth No.6 that has been resulted by conducting 1000 times recognition experiment.

## 5.2. Robustness confirmation against illumination and lighting source varieties

If the lighting source changes (e.g. light bulb, fluorescent lamp, mercury-arc lamp, sunlight, etc), it will be necessary to check the robustness of the system against different light sources. We have already confirmed the verification of illumination variations under two different lighting sources. Three different illumination conditions (100 lx, 700 lx, and 1300 lx) using fluorescent light and light-emitting diode (LED) separately were given as experimental environments in (Funakubo R et al., 2017). In (Phyu, Khaing Win, et al., 2017), five different illumination conditions (100 lx, 400 lx, 700 lx, 1000 lx and 1300 lx) using fluorescent light and light-emitting diode (LED) were separately simulated for experiments. Fluorescent light was used for the experiments of the present paper.

Figures 13 and 14 show the fitness distributions for position and orientation of cloth No.3 under the five different illuminations (100 lx, 400 lx, 700 lx, 1000 lx and 1300 lx) and the two different lighting sources (Fig. 13; fluorescent light and Fig. 14; LED). In the experiments, cloth No.3 was chosen because of its distinct characteristics such as colorful patterns, small size and light weight. According to experimental results in Figs. 13 and 14, it can be seen that the height of the peak of the fitness distribution changes with illumination variations. The difference between the height of fitness distribution in position under 100 lx and 1300 lx can be clearly seen in Figs. 13 (a) and 14 (a). Even though the height of the peak changes with illumination strength, it can be seen from Figs. 13 and 14 that there exist peaks at the position of the cloth, and the positions represented by the peaks are maintained in all cases. It means that the proposed system is robust against illumination and lighting source varieties. The reason why the system is robust is that the searching problem is converted into the optimization problem in our proposed system. The conversion enables the pose estimation system to be robust against lighting condition varieties since the optimization procedures do not care about height of the peaks, but the existence of the highest peak at the pose to be estimated. These Figs. 13 and 14 are introduced from our previous paper (Phyu, Khaing Win, et al., 2017).

The first conditions in Eqs. (6) and (7) contributed to make the peak higher, and second conditions of penalties in Eqs. (6) and (7) helped lower peaks that were generated by image noises deleted. In the Figs. 13 (a) and 14 (a), the fitness function values with minus sign were all replaced by zero, then all the fitness distributions in Figs. 13 (a) and 14 (a) look like external form with single peak. The values set in Eqs. (6) and (7), that is, 2, -0.005, 0.1 and -0.5 are experimentally set by adjusting the valuables that had been done before pose estimation and handling experiments.

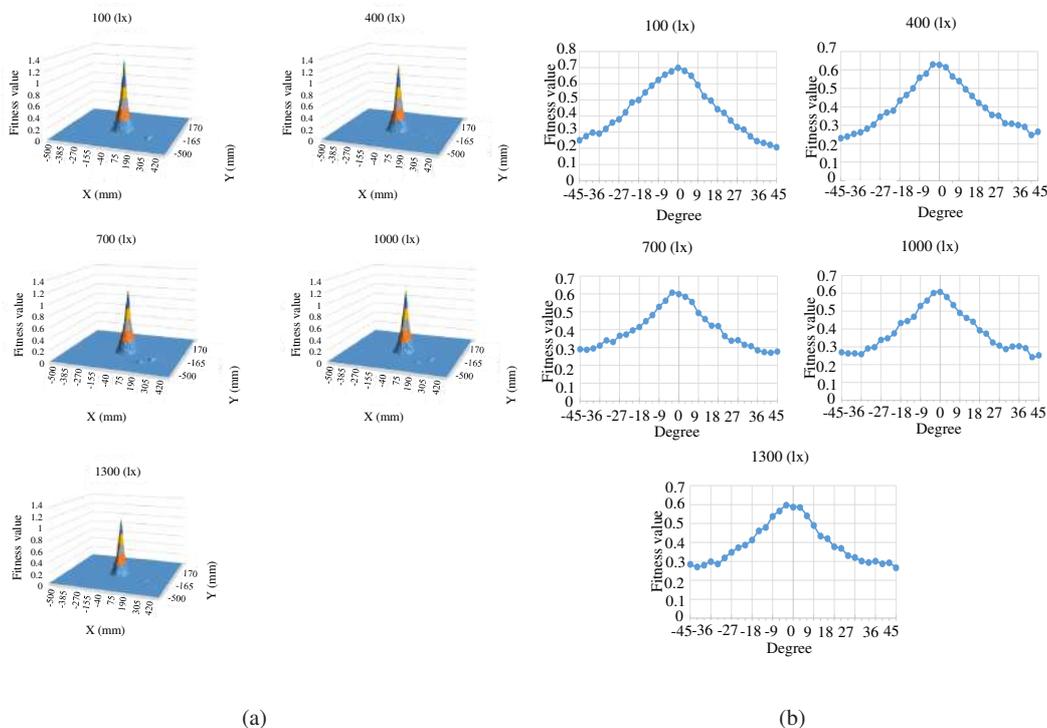


Fig. 13 Fitness function distribution of cloth No.3 in (a) x-y plane and (b) orientation (fluorescent light)

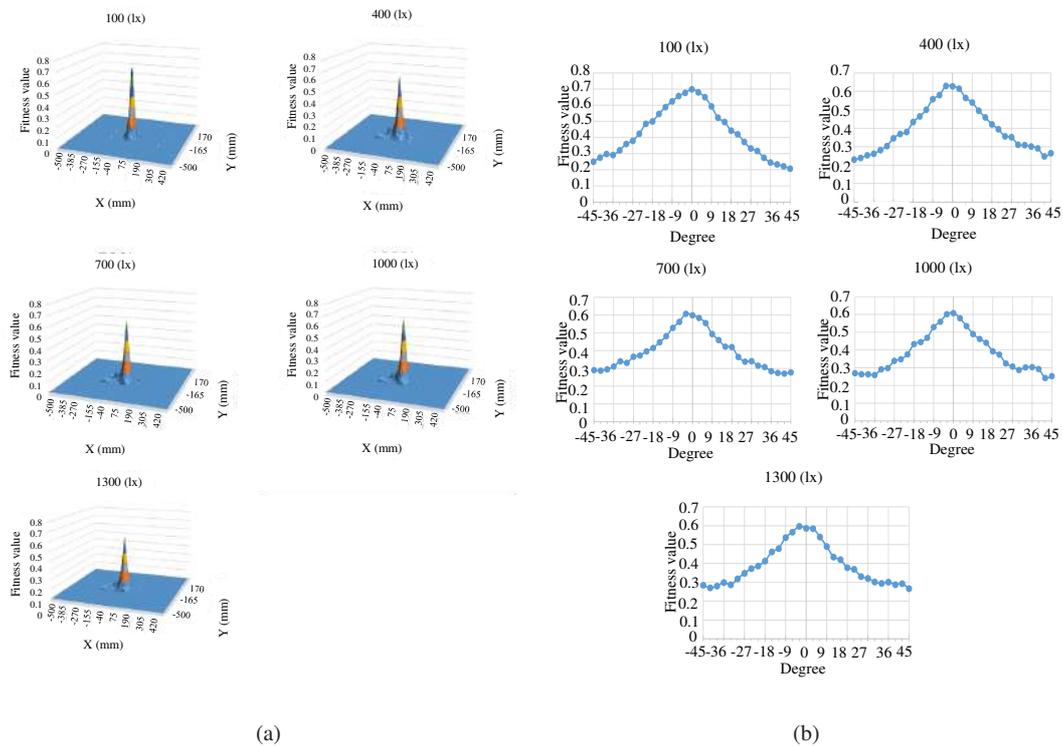


Fig. 14 Fitness function distribution of cloth No.3 in (a) x-y plane and (b) orientation (LED light)

### 5.3. Accuracy with illumination varieties

Figures 15 to 19 show the average error of each cloth No.1~No.12 under five different illuminations (100 lx, 400 lx, 700 lx, 1000 lx and 1300 lx). The numerical values of average error of pose estimation results are listed in Table 3 and the standard deviation are in Table 4. From Figs. 15 to 19, the 100 times recognition experiment results of average error under five different illuminations for all cloths (No.1~No.12) have almost same tendency with Fig. 12. With the variation of cloths and also with the varieties of light conditions (100 lx~1300 lx), it has been confirmed that the  $\pm 3\sigma$  of the x, y position are less than 10 [mm], z position being 30 [mm],  $\theta$  being 10 [degree].

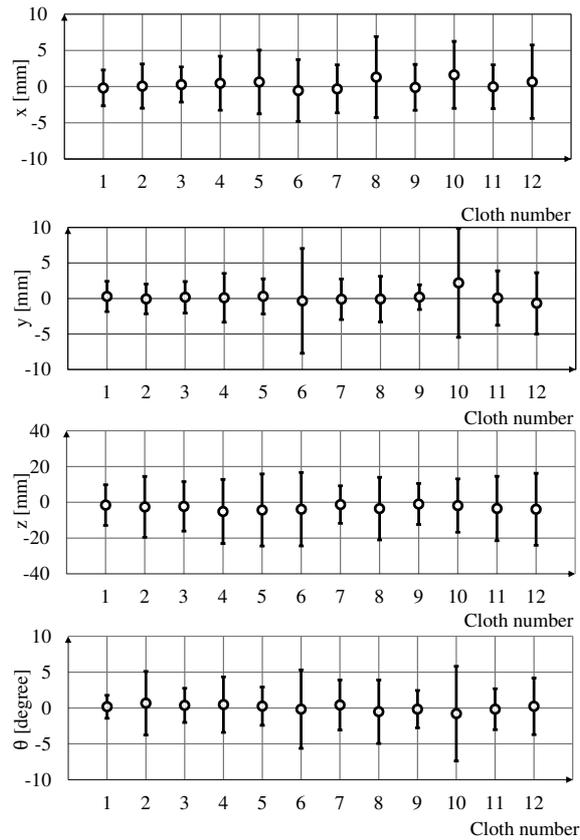


Fig. 15 Average error and average error  $\pm 3\sigma$  of 100 times recognition experimental result of 12 unique cloths under 100 lx.

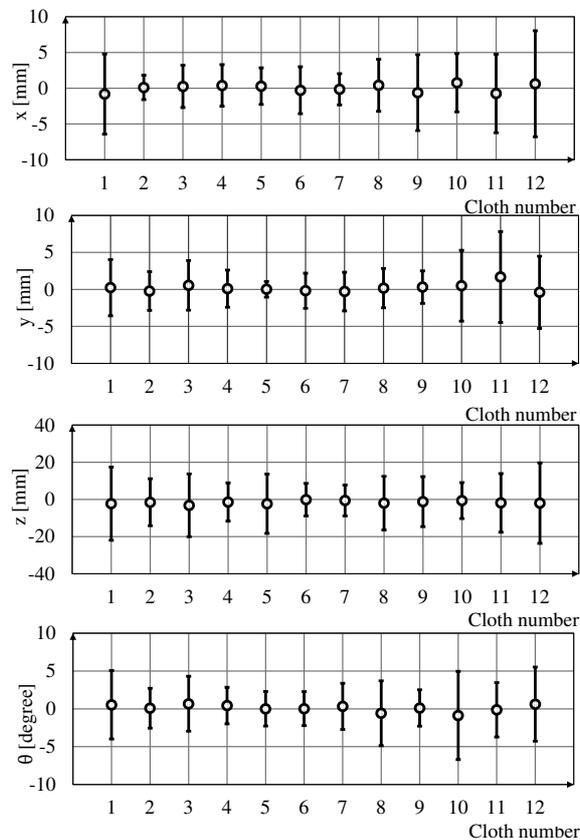


Fig. 16 Average error and average error  $\pm 3\sigma$  of 100 times recognition experimental result of 12 unique cloths under 400 lx.

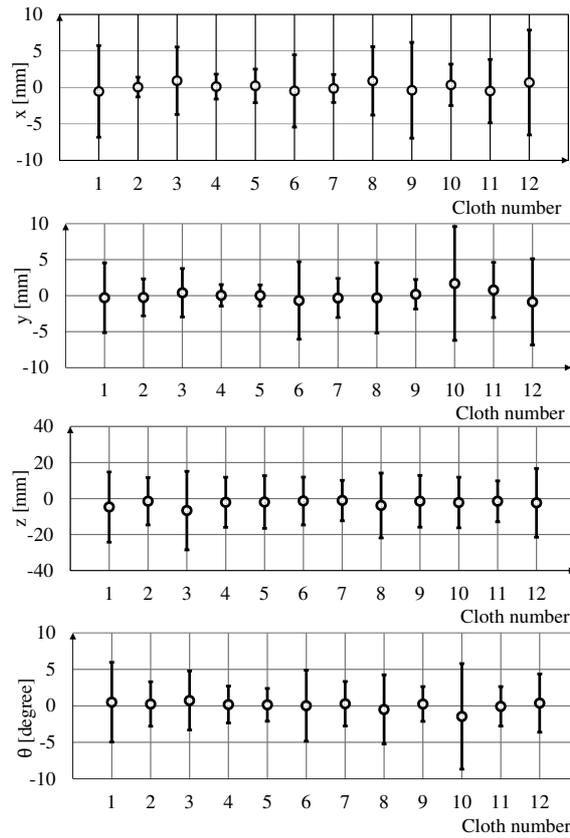


Fig. 17 Average error and average error  $\pm 3\sigma$  of 100 times recognition experimental result of 12 unique cloths under 700 lx.

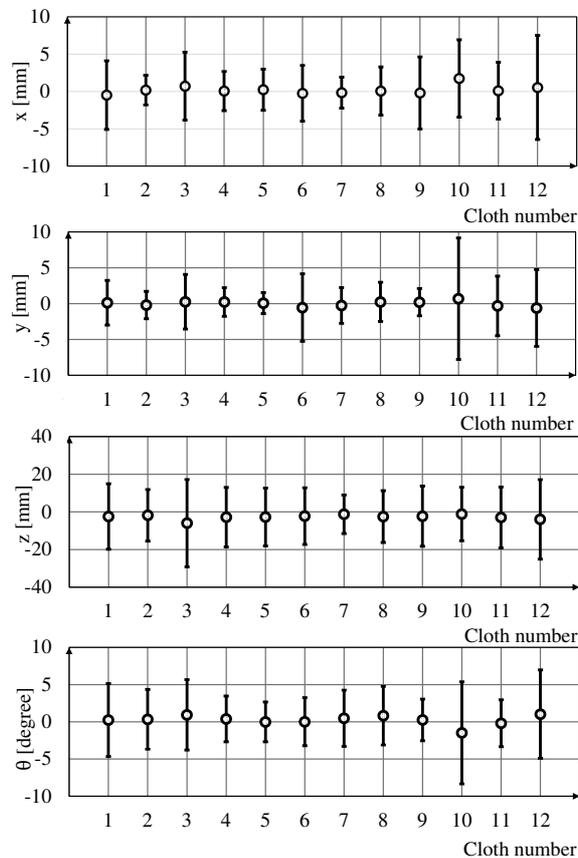


Fig. 18 Average error and average error  $\pm 3\sigma$  of 100 times recognition experimental result of 12 unique cloths under 1000 lx.

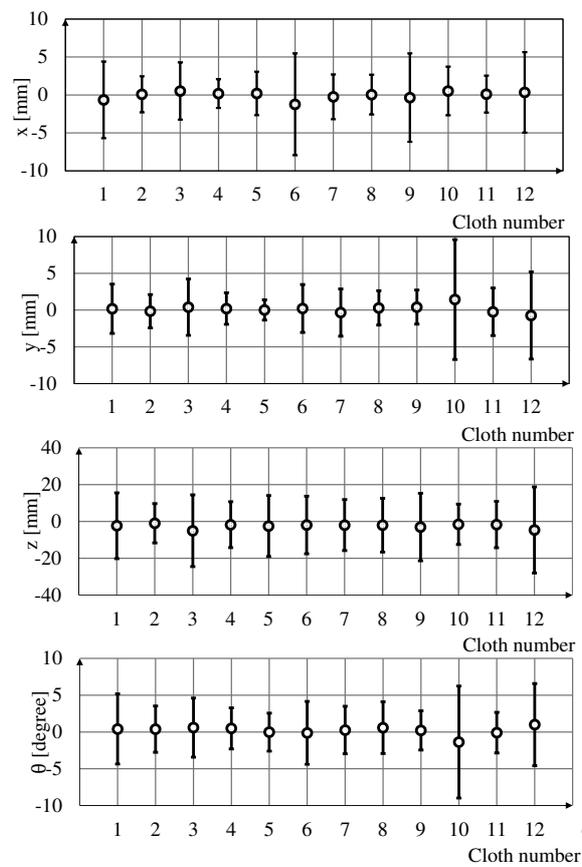


Fig. 19 Average error and average error  $\pm 3\sigma$  of 100 times recognition experimental result of 12 unique cloths under 1300 lx.

Table 3 Average error measured by 100 times recognition experiments of each cloth No.1~No.12 under different light conditions (100 lx, 400 lx, 700 lx, 1000 lx and 1300 lx).

Cloth No.	Illuminance [lx]	Average error			
		$\bar{x}$ [mm]	$\bar{y}$ [mm]	$\bar{z}$ [mm]	$\theta$ [degree]
No.1	100	-0.179	0.282	-1.586	0.169
	400	-0.813	0.234	-2.273	0.533
	700	-0.550	-0.291	-4.727	0.499
	1000	-0.487	0.119	-2.459	0.233
	1300	-0.645	0.181	-2.310	0.408
No.2	100	0.074	-0.067	-2.57	0.667
	400	0.109	-0.227	-1.52	0.084
	700	0.043	-0.239	-1.49	0.251
	1000	0.182	-0.204	-1.84	0.335
	1300	0.0840	-0.156	-0.968	0.381
No.3	100	0.287	0.170	-2.32	-0.364
	400	0.250	0.543	-3.17	0.676
	700	0.913	0.407	-6.68	0.7290
	1000	0.715	0.254	-5.98	0.933
	1300	0.521	0.399	-5.04	0.600
No.4	100	0.462	0.095	-5.124	-0.462
	400	0.392	0.099	-1.416	0.435
	700	0.113	0.042	-2.061	0.186
	1000	0.057	0.222	-2.810	0.390
	1300	0.185	0.208	-1.749	0.486
No.5	100	0.641	0.296	-4.312	0.2496
	400	0.291	0.015	-2.314	0.008
	700	0.209	0.027	-1.944	0.139
	1000	0.234	0.062	-2.730	-0.017
	1300	0.198	0.019	-2.466	-0.021
No.6	100	-0.545	-0.341	-3.880	-0.164
	400	-0.278	-0.181	-0.100	0.027
	700	-0.484	-0.676	-1.356	0.021
	1000	-0.248	-0.561	-2.246	0.011
	1300	-1.22	0.215	-1.909	-0.126
No.7	100	-0.308	-0.115	-1.260	0.413
	400	-0.145	-0.291	-0.571	0.334
	700	-0.143	-0.321	-1.112	0.288
	1000	-0.168	-0.268	-1.257	0.472
	1300	-0.253	-0.336	-1.973	0.263
No.8	100	1.319	-0.094	-3.551	-0.523
	400	0.416	0.173	-1.984	0.582
	700	0.893	-0.2976	-3.864	-0.492
	1000	0.052	0.234	-2.495	0.821
	1300	0.047	0.300	-2.018	0.5871
No.9	100	-0.103	0.186	-0.947	-0.166
	400	-0.614	0.304	-1.185	0.116
	700	-0.393	0.195	-1.507	0.255
	1000	-0.197	0.195	-2.300	0.255
	1300	-0.349	0.410	-3.043	0.213
No.10	100	1.622	2.194	-1.850	-0.795
	400	0.773	0.497	-0.613	-0.878
	700	0.349	1.691	-2.214	-1.449
	1000	1.739	0.679	-1.153	-1.480
	1300	0.525	1.437	-1.551	-1.376
No.11	100	-0.016	0.059	-3.49	-0.168
	400	-0.740	1.67	-1.86	-0.122
	700	-0.506	0.790	-1.55	-0.073
	1000	0.103	-0.325	-2.95	0.197
	1300	0.109	-0.235	-1.68	-0.092
No.12	100	0.674	-0.687	-3.910	0.226
	400	0.625	-0.394	1.981	0.622
	700	0.678	-0.861	-2.381	0.372
	1000	0.537	-0.620	-3.981	1.028
	1300	0.340	-0.733	-4.625	1.000

Table 4 Standard deviation ( $\sigma$ ) measured by 100 times recognition experiments of each cloth No.1~No.12 under different light conditions (100 lx, 400 lx, 700 lx, 1000 lx and 1300 lx).

Cloth No.	Illuminance [lx]	Standard deviation ( $\sigma$ )			
		x [mm]	y [mm]	z [mm]	$\theta$ [degree]
No.1	100	0.829	0.713	3.796	0.535
	400	1.869	1.270	6.577	1.510
	700	2.093	1.619	6.508	1.823
	1000	1.534	1.041	5.805	1.635
	1300	1.685	1.117	5.964	1.591
No.2	100	1.02	0.699	5.68	1.48
	400	0.570	0.874	4.22	0.877
	700	0.455	0.858	4.39	1.02
	1000	0.664	0.640	4.58	1.34
	1300	0.793	0.755	3.57	1.05
No.3	100	0.812	0.739	4.62	0.795
	400	0.988	1.12	5.64	1.21
	700	1.54	1.12	7.28	1.35
	1000	1.52	1.27	7.73	1.58
	1300	1.26	1.28	6.50	1.34
No.4	100	1.124	1.148	5.988	1.289
	400	0.974	0.837	3.440	0.808
	700	0.571	0.500	4.646	0.844
	1000	0.876	0.668	5.279	1.026
	1300	0.634	0.716	4.154	0.932
No.5	100	1.468	0.824	6.729	0.889
	400	0.855	0.355	5.325	0.762
	700	0.697	0.489	4.884	0.752
	1000	0.918	0.490	5.128	0.892
	1300	0.957	0.460	5.515	0.863
No.6	100	1.426	2.462	6.847	1.821
	400	1.096	0.794	2.922	0.748
	700	1.653	1.797	4.439	1.626
	1000	1.249	1.572	5.023	1.078
	1300	2.237	1.087	5.217	1.425
No.7	100	1.105	0.954	3.495	1.166
	400	0.730	0.875	2.767	1.019
	700	0.639	0.907	3.747	1.015
	1000	0.692	0.835	3.421	1.261
	1300	0.984	1.074	4.625	1.079
No.8	100	1.865	1.071	5.846	1.476
	400	1.213	0.888	4.816	1.429
	700	1.565	1.630	6.011	1.582
	1000	1.078	0.911	4.587	1.314
	1300	0.872	0.782	4.863	1.174
No.9	100	1.060	0.578	3.825	0.869
	400	1.773	0.734	4.493	0.803
	700	2.192	0.686	4.807	0.795
	1000	1.612	0.634	5.336	0.933
	1300	1.944	0.772	6.114	0.889
No.10	100	1.542	2.557	4.978	2.202
	400	1.360	1.596	3.220	1.946
	700	0.950	2.637	4.686	2.411
	1000	1.731	2.828	4.735	2.286
	1300	1.067	2.720	3.655	2.536
No.11	100	1.01	1.28	6.01	0.952
	400	1.83	2.05	5.26	1.20
	700	1.45	1.28	3.79	0.905
	1000	1.27	1.39	5.39	1.05
	1300	0.813	1.08	4.20	0.924
No.12	100	1.693	1.442	6.711	1.314
	400	2.472	1.629	7.233	1.635
	700	2.398	1.999	6.368	1.331
	1000	2.325	1.787	7.038	1.979
	1300	1.768	1.98	7.801	1.864

#### 5.4. Handling experiment

Regarding the handling of clothing by the PA-10 robot, the four absorption pads that are inhaled by air pump are used for picking up the cloth and putting into the collection box automatically. Figure 21 shows the four absorption pads used in the experiment. Even though the required accuracy for grabbing is not high, but it relates to handling accuracy finally, which has been confirmed and been shown in Fig. 22. This has shown that the proposed system can be useful in practical view point.



Fig. 21 Four absorption pads under robot hand

In the mail order system of the company T2K, human workers classify and handle a large number of cloths manually every day. The robot to help human workers should be capable of automatically handling to classify cloths. Results of 100 times handling experiment at different light conditions (100 lx, 400 lx, 700 lx, 1000 lx and 1300 lx) were summarized in Table 5 numerically and shown in Fig. 22. Table 5 lists the numerical data of average error and standard deviation  $\pm 3\sigma$  when handling No.2 cloth. As being depicted that maximum  $3\sigma$  of z-axis is about 30 [mm], the hand of the robot needs some adaptive mechanism to pick up the cloths, like spring-rubber siphon absorbing mechanism to adjust possible z-axis hand position errors. Since the horizontal position errors in x, y-axis are less than 10 [mm] and orientation error is roughly less than 15 [degree], the proposed handling robot can insert the cloths into a box with a size being 20 [mm] larger the biggest size of the cloths' varieties. Then the experimental results as shown in Fig. 22 and Table 5 have confirmed experimentally that the proposed system is able to handle the 12 different cloths (No.1 ~ No.12) under different light conditions without the need for human assistance.

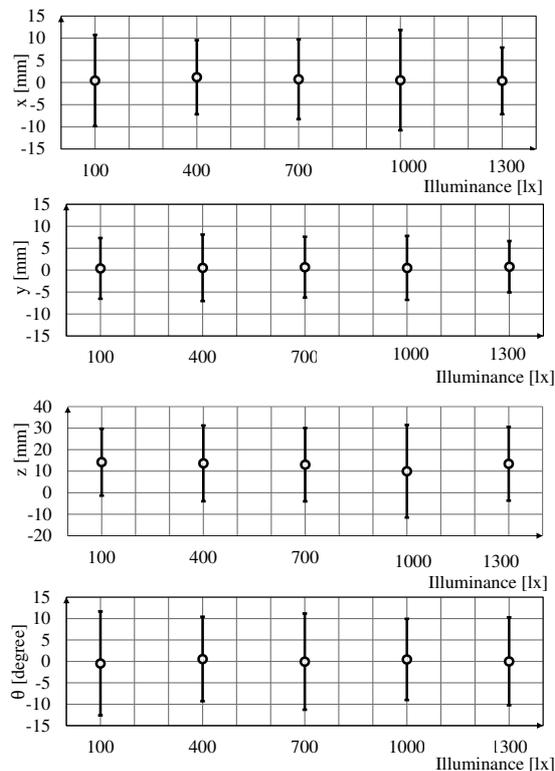


Fig. 22 Average error and average error  $\pm 3\sigma$  of 100 times handling experiments of No.2 cloth under different light conditions (100 lx, 400 lx, 700 lx, 1000 lx and 1300 lx).

Table 5 Average error and standard deviation ( $3\sigma$ ) measured by 100 times handling experiments of cloth No.2 under different light conditions.

Illuminance [lx]	Average error : upper row			
	Standard Deviation ( $3\sigma$ ): lower row			
	$\bar{x}$ [mm]	$\bar{y}$ [mm]	$\bar{z}$ [mm]	$\bar{\theta}$ [degree]
100	0.436	0.392	14.2	-0.466
	10.2	6.93	15.5	12.1
400	1.20	0.527	13.6	0.559
	8.37	7.59	17.6	9.87
700	0.732	0.666	13.0	-0.0440
	9.00	6.93	17.0	11.2
1000	0.525	0.502	9.96	0.464
	11.3	7.29	21.4	9.48
1300	0.366	0.772	13.4	0.0171
	7.53	5.85	17.1	10.2

## 6. Conclusion

Verification of the unique cloth recognition and handling performance using the photo-model-based cloth recognition under different illuminations under 100 lx to 1300 lx is presented. In addition, the handling performance by PA-10 robot has been verified. The experimental results indicated that errors and the extent of  $\pm 3\sigma$  concerning x and y positions of all 12 cloths are less than  $\pm 10$  [mm], and those of  $\theta$  around z-axis is less than 10 [degree] with the probability of 99.7% in recognition experiment with illumination varieties. According to the experimental result, the proposed system has been confirmed to be able to recognize and handle the 12 unique cloths under different light conditions without the need for human assistance.

## Acknowledgment

The authors would like to express their thanks to T2K Co., Ltd for their help and support.

## References

- Castillo CD and Jacobs DW., Using stereo matching with general epipolar geometry for 2d face recognition across pose, *IEEE Transactions on Pattern Analysis and Machine Intelligence* (2009), 31(12), pp.2298-2304.
- CUI Y, LI X, MINAMI M, YANOU A, YAMASHITA M and ISHIYAMA S., Robot for decontamination with 3D move on sensing, *Nuclear Safety and Simulation*, 6(2) (2015), pp.142-154.
- Funakubo R, Phyu KW, Tian H and Minami M., Recognition and handling of clothes with different pattern by dual hand-eyes robotic system, *IEEE/SICE International Symposium* (2016), pp. 742-747.
- Funakubo R, Phyu KW, Hagiwara R, Tian H and Minami M., Verification of illumination tolerance for clothes recognition, *The Twenty-Second International Symposium on Artificial Life and Robotics 2017 (AROB 22nd 2017)*, B-Con Plaza (January 19-21 2017), Beppu, Japan.
- Maitin-Shepard J, Cusumano-Towne M, Lei J and Abbeel., Cloth grasp point detection based on multiple-view geometric cues with application to robotic towel folding, *In Robotics and Automation (ICRA), IEEE International Conference* (May 2010), pp. 2308-2315.
- Minami M, Agbanhan J and Asakura T., Evolutionary scene recognition and simultaneous position/orientation detection, *in Soft Computing in Measurement and Information Acquisition*, Springer Berlin Heidelberg (2003), pp. 178-207.
- Myint Myo, Kenta Yonemori, Akira Yanou, Shintaro Ishiyama and Mamoru Minami., Robustness of visual-servo against air bubble disturbance of underwater vehicle system using three-dimensional marker and dual-eye cameras, *In OCEANS'15 MTS/IEEE Washington* (2015), pp. 1-8.
- Phyu KW, Cui Y, Tian H, Hagiwara R, Funakubo R, Yanou A and Minami M., Accuracy on Photo-Model-Based Clothes Recognition, *SICE Annual Conference* (September 20-23 2016), Tsukuba, Japan.
- Paton M, Pomerleau F, Barfoot TD., *In the dead of winter: Challenging vision-based path following in extreme conditions*, *In Field and Service Robotics Springer International Publishing* (2016), pp. 563-576.

- Phyu, Khaing Win, Ryuki Funakubo, Ryota Hagiwara, Hongzhi Tian, and Mamoru Minami, Verification of illumination tolerance for photo-model-based cloth recognition, *Artificial Life and Robotics* (2017), DOI 10.1007/s10015-017-0391-0.
- S. Ullman and R. Basri., Recognition by linear combinations of models, *IEEE Trans. PAMI*, Vol. 13, No. 10 (1991), pp. 992-1106.
- T. Poggio and S. Edelman., A network that learns to recognize 3d objects, *Nature* (1990), Vol. 343, pp. 263-266.
- W. E. L. Grimson., A computer implementation of a theory of human stereo vision, *Philosophical Transactions of the Royal Society of London, B.*, Vol. 292, No. 1058 (1981), pp. 217-253.
- Yinxiao Li, Chih-Fan Chen and Peter K Allen., Recognition of deformable object category and pose, *Robotics and Automation (ICRA)*, IEEE International Conference, Hong Kong (May 31-June 7 2014), pp.5558-5564.
- Z. Zhang, R. Deriche, O. Faugeras, and Q.-T. Luong., A robust technique for matching two uncalibrated images through the recovery of the unknown epipolar geometry, *Artificial Intelligence Journal* (1995), Vol. 78, pp. 87-119.