



## II. PROPOSED PHOTOGRAPH MODELING-BASED CLOTH RECOGNITION

There are two main portions in the 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 a description of the kinematics of stereo-vision before the explanation of proposed system in details.

### A. Kinematics of Stereo-Vision

The relationship with the world coordinate system of the manipulator  $\Sigma_W$ , the hand coordinate system  $\Sigma_H$  and the target object coordinate system  $\Sigma_M$  are shown in Fig. 2. Perspective projection of dual-eyes vision system is shown in Fig. 3,  $\Sigma_{CR}$  and  $\Sigma_{CL}$  represent the coordinate system of the left and right cameras and the left and right cameras' images are represented as  $\Sigma_{IR}$  and  $\Sigma_{IL}$ . According to the coordinate system of dual-eyes in Fig. 3, the  $i$ -th point of the solid model can be represented as the simultaneous transformation matrix  ${}^{CR}T_M$ . Firstly, a homogeneous transformation matrix from right camera coordinate system  $\Sigma_{CR}$  to the target object coordinate system  $\Sigma_M$  is defined as  ${}^{CR}T_M$ . Secondly, the arbitrary  $i$ -th point on the target object defined on the model in  $\Sigma_{CR}$  is named as  ${}^{CR}r_i$  and similarly, the object as viewed from the search point  $i$ -th on the model in  $\Sigma_M$  is named as  ${}^M r_i$ . Then,  ${}^{CR}r_i$  can be calculated by using (1),

$${}^{CR}r_i = {}^{CR}T_M {}^M r_i. \quad (1)$$

where  ${}^M r_i$  is predetermined fixed vector.

By using the homogeneous transformation matrix  ${}^W T_{CR}$  from the  $i$ -th point world coordinate system in  $\Sigma_W$  to the right camera coordinate system  $\Sigma_{CR}$  and the left camera coordinate system  $\Sigma_{CL}$  are achieved as (2) and (3),

$${}^W r_i = {}^W T_{CR} {}^{CR}r_i. \quad (2)$$

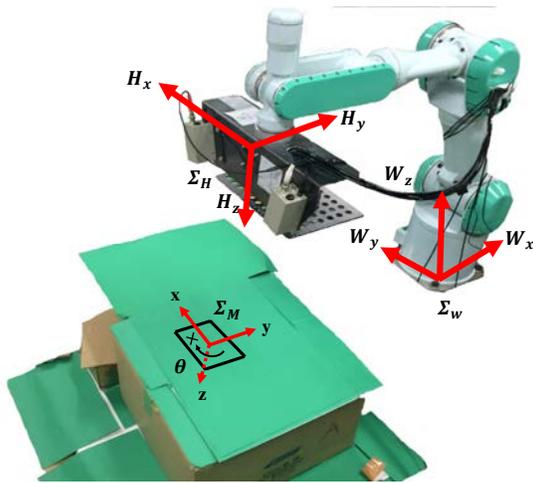


Fig. 2. Coordinate systems of 3D-MoS robot

$${}^W r_i = {}^W T_{CL} {}^{CL}r_i. \quad (3)$$

Equation (4) represents the projective transformation matrix  $P$ .

$$P = \frac{1}{c_{z_i}} \begin{bmatrix} \frac{f}{\eta_x} & 0 & l_{x_0} & 0 \\ 0 & \frac{f}{\eta_y} & l_{y_0} & 0 \end{bmatrix}. \quad (4)$$

The position vector of the  $i$ -th point in the right and left camera image coordinates  ${}^{IR}r_i$  and  ${}^{IL}r_i$  can be described by using  $P$  as shown in (4) of the camera as (5) and (6).

$${}^{IR}r_i = P {}^{CR}r_i. \quad (5)$$

$${}^{IL}r_i = P {}^{CL}r_i. \quad (6)$$

By using the same procedure,  ${}^{CL}r_i$  is possible to define as the kinematic relation from  $\Sigma_{CL}$  to  $\Sigma_M$  and  ${}^{CL}T_M$  whether it is described by the following (7).

$${}^{CL}r_i = {}^{CL}T_M {}^M r_i. \quad (7)$$

Then (5) and (6) are connected by an arbitrary point on a 3D-model  ${}^M r_i$  in  $\Sigma_M$  with the pose of  $\Sigma_M$  that is based on  $\Sigma_{CR}$  and  $\Sigma_{CL}$  ( ${}^C \phi_M$ ) to the projected point on the left and right camera's images  ${}^{IL}r_i$  and  ${}^{IR}r_i$  which can be rewritten as (8).

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

### B. Cloth Model Generation

In this system, three cameras are used as the vision sensors. A background image is captured by the first camera as shown in Fig. 4 (a) and the average hue value of the background image is calculated. The next step is to put a cloth in the background as shown in Fig. 4 (b). Then, the hue value of each point in the image acquired by scanning individual pixel is compared with the average hue value of the background image which is generated the surface space

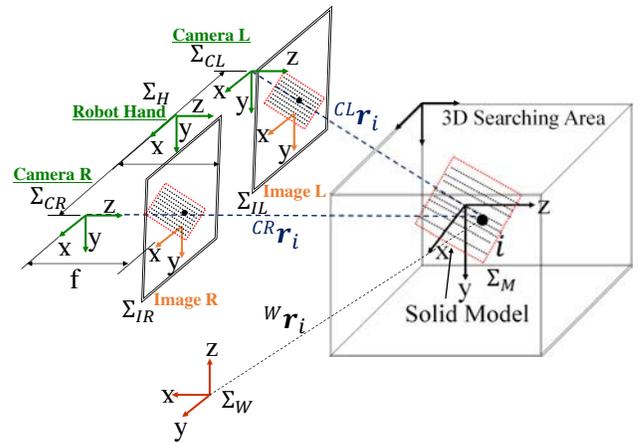


Fig. 3. 3D to 2D projection with coordinate systems

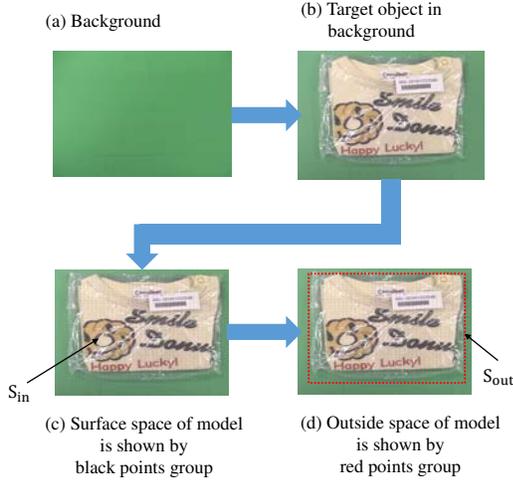


Fig. 4. Process of model generation

$S_{in}$  of the model as shown in Fig. 4 (c). Finally, the outside space  $S_{out}$  of the model is generated by enveloping  $S_{in}$  as shown in Fig. 4 (d).

### C. Model-based Matching Method

In the top of Fig. 5, the space on the surface of the model is specified as  $S_{in}(\phi_M)$ . The outside space enveloping  $S_{in}(\phi_M)$  is denoted as  $S_{out}(\phi_M)$ . In the left bottom of Fig. 5, there is a left 2D searching model named  $S_L(\phi_M)$ , including  $S_{L,in}(\phi_M)$  and  $S_{L,out}(\phi_M)$ . In the right bottom of Fig. 5, there is a right 2D searching model named  $S_R(\phi_M)$ , including  $S_{R,in}(\phi_M)$  and  $S_{R,out}(\phi_M)$ . These models in left and right cameras' images are projected from the model in 3D space. The models are randomly generated in the 3D searching area with each relative pose. In Fig. 5, the 3D solid model is projected from 3D searching area to 2D (left and right) image planes. The projected models on the left and right cameras' images planes are located in the corresponding pose of the 3D solid model. The matching degree between the projected model and the captured image are determined by using the fitness function. If the captured image coincides with the projected image, the pose of the real target object will be obtained precisely as shown in Fig. 5.

### D. Definition of Fitness Function

The matching degree of each point in model space ( $S_{L,in}(\phi_M)$  and  $S_{L,out}(\phi_M)$ ) and that of captured image can be calculated from designed fitness values as shown in (9) and (10). Similarly, a function  $pR_{in}({}^{IR}r_i)$  and  $pR_{out}({}^{IR}r_j)$  are calculated for the right camera image. Equation (11) represents the evaluation function  $F({}^C\phi_M)$ .  $F({}^C\phi_M)$  is achieved by the average of the fitness function of both left camera image  $F_L({}^{CL}\phi_M)$  and right camera image  $F_R({}^{CR}\phi_M)$ .

$$pL_{in}({}^{IL}r_i) = \begin{cases} 2, & \text{if } (|H_{IL}({}^{IL}r_i) - H_{ML}({}^{IL}r_i)| \leq 30); \\ -0.005, & \text{if } (|\bar{H}_B - H_{IL}({}^{IL}r_i)| \leq 30); \\ 0, & \text{otherwise.} \end{cases} \quad (9)$$

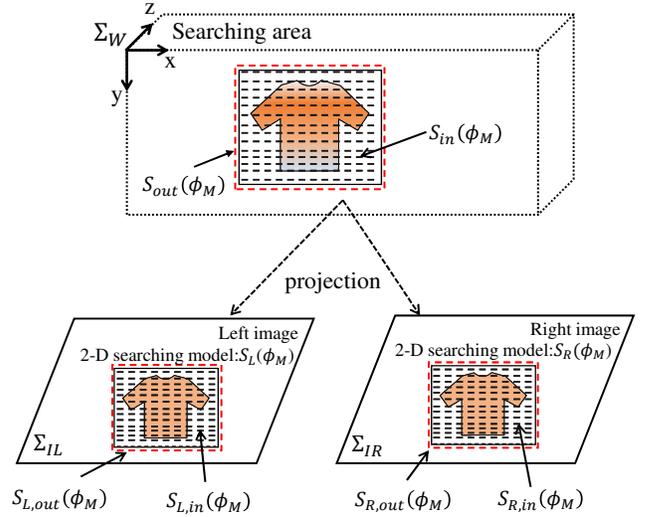


Fig. 5. Model-based matching

$$pL_{out}({}^{IL}r_j) = \begin{cases} 0.1, & \text{if } (|\bar{H}_B - H_{IL}({}^{IL}r_j)| \leq 20); \\ -0.5, & \text{otherwise.} \end{cases} \quad (10)$$

where

- $H_{IL}({}^{IL}r_i)$ : the  $i$ -th point in  $S_{L,in}$  and the hue value of the left camera image at the point  ${}^{IL}r_i$ ,
- $H_{ML}({}^{IL}r_i)$ : the  $i$ -th point in  $S_{L,in}$  and the hue value of the left camera image at the point  ${}^{IL}r_i$  on the model,
- $H_{IL}({}^{IL}r_j)$ : the  $j$ -th point in  $S_{L,out}$  and the hue value of the left camera image at the point  ${}^{IL}r_j$ ,
- $\bar{H}_B$ : the average hue value of the background image

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

### E. GA (Genetic Algorithm)

In proposed 3D model-based recognition method, the problem of recognition the target object and detecting its pose are converted into an optimization problem. GA is used as an optimization method for cloth recognition in this experiment. The reason why we choose the GA is based on its simplicity, repeatable ability and especially effectiveness in the recognition performance. Fig. 6 shows the GA evolution

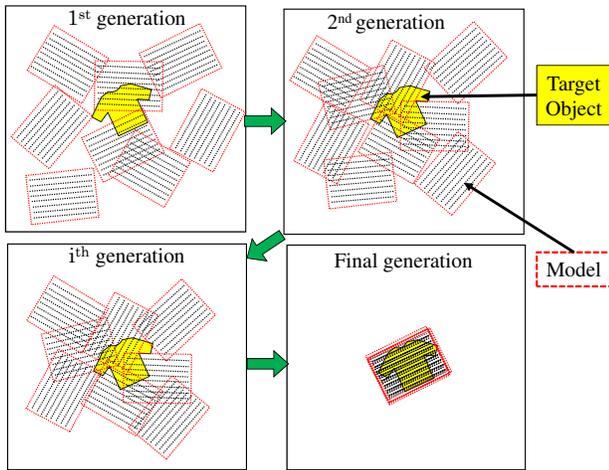


Fig. 6. GA evolution process

process for recognition. In this figure, the cloth picture is the target object and the rectangular shape dotted line is the model. There are many models located in the searching area with the same shape and color but the models have different poses. Firstly, the GA individuals have randomly generated in the searching area as the 1<sup>st</sup> generation. Secondly, the fitness value of each individual is calculated and sorted the calculated results based on the rank of the fitness value. And then the best individuals are selected from the current population and the weak individuals are removed. The next generation is reproduced by making crossover and mutation from the selected individuals. Finally, the true pose of the target object is obtained with the highest fitness value through the GA evolution process. Each individual chromosome has six variables ( $12\text{bits} \times 6 = 72 \text{ bits}$ ). The first three variables (36bits) represent for the position of a model in 3D space as ( $t_x, t_y, t_z$ ) and the last three variables (36bits) represent for the orientation of the 3D model as ( $\epsilon_1, \epsilon_2, \epsilon_3$ ) in quaternion. And then, the characteristics of GA individuals as shown in below;

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

### III. EXPERIMENTAL ENVIRONMENT

There are two units in an experimental environment. One is for cloth model generation including one camera as shown in Fig. 7. Another one is end-effector equipped with two cameras installed in manipulator's end-effector as shown in Fig. 8. In Fig. 7, the distance from the camera lens to the model creating plane is 400 mm and the color of the background plane is green. The size of cloth models can be up to 250mm  $\times$  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  $\Sigma_H$  defined as the hand coordinate system of

the robot end-effector.  $\Sigma_M$  can be viewed from ( $x=0, y=0, z=580\text{mm}$ ). It is centered on the recognition range of the position as a reference of the 510mm  $\times$  390mm. The size of the collection box is a 220mm  $\times$  220mm as shown in Fig. 9. 12 different cloths (No.1, No.2, ... , No.12) have unrepeatable color, size, shape and weight samples which are used in this experiment as shown in Fig. 10. Each item of cloth must be recognized individually to confirm the recognition accuracy of the proposed system.

### IV. EXPERIMENTAL RESULTS AND DISCUSSION

Fig. 11 shows the different experiments of cloth No.3. Fig. 11 (a), (b) and (c) show environment without disturbance, recognition of partly hidden cloth and identification of two pieces of cloths such as No.3 and No.4 cloths respectively. To evaluate the recognition accuracy, the full search method is used to compare with the estimated pose by proposed system. In full search method, the pose of the target cloth is searched by scanning all possible pixels in the entire searching area. Fig. 13 shows the recognition comparison of fitness distribution in x-y plane between the full search and the GA search process. Fig. 12 shows the recognition accuracy of all cloths (No.1~No.12).

#### A. Recognition without disturbance

Among 12 different cloths, we will discuss in detail about the experiments of cloth No.3 in this section according to the three different features of the cloth No.3. There are small size, colorful pattern and light weight. As shown in Fig. 11 (a), cloth No.3 is placed in the vicinity of ( $x,y$ )=(0,0). Fig. 13 (a) shows the fitness distribution in the x-y plane of cloth No.3 as 2D view and Fig. 13 (b) shows as 3D view. Fig. 13 (b) shows clearly that the maximum fitness function value of a peak in the fitness function distribution can be searched by the evaluation process of GA model after 100 times evaluation. The fitness value reaches the maximum at

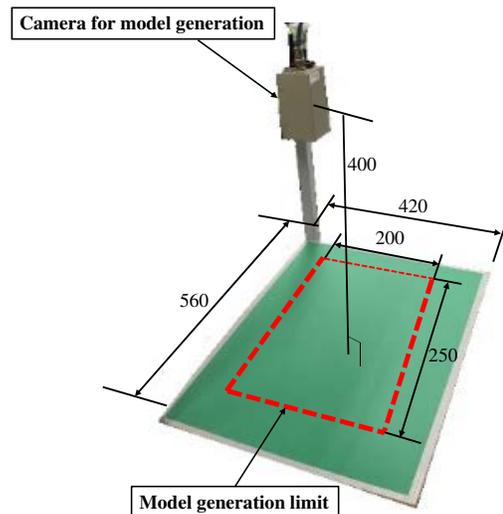


Fig. 7. Environment of model generation (unit is (mm) in Figure 7)

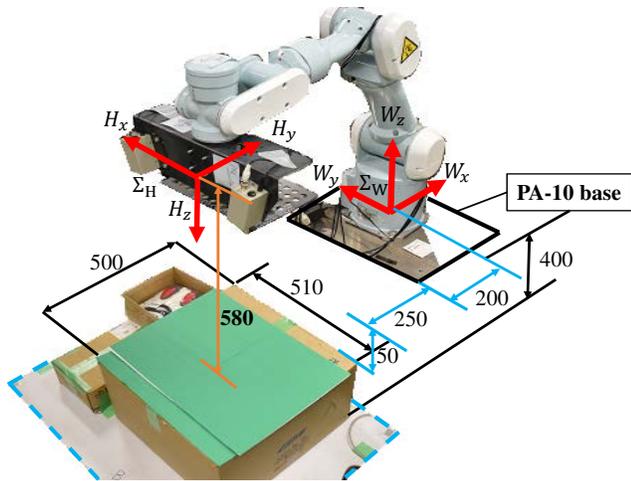


Fig. 8. Coordinate system of robot and end-effector (unit is (mm) in Figure 8)

the peak of a mountain at 0.85 and the position is (9,11) (mm) in Y-plane. There was only one peak in the fitness distribution which means that the target object (cloth No.3) and the model are matching well.

### B. Recognition of partly hidden cloth

In this experiment, about 45% of cloth is hidden by the blue paper as shown in Fig. 11 (b) and that hidden cloth No.3 is put near (x,y)=(0,0) for recognition performance. After 100 times evaluation, the evolution process of GA is shown in Fig. 13 (c) and (d). The maximum fitness value 0.32 can be seen at (x,y)=(-6,0) in the fitness distribution as the peak of the mountain shape. Fig. 13 (c) and (d) confirmed that GA converged correctly with the pose of the target object and recognized the partly hidden No.3 cloth.

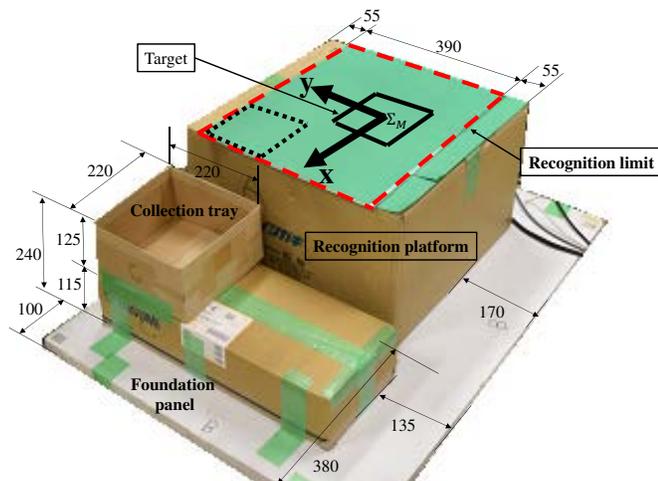


Fig. 9. Coordinate system of target object (unit is (mm) in Figure 9)

### C. Identification of two cloths

In Fig. 11 (c), cloth No.3 put in the range of  $x < 0$  and cloth No.4 put in the range of  $x > 0$ . The evolution process of GA at 100 times evaluation is indicated in Fig. 13 (e) and (f). The peak of cloth No.3 appear at (x,y)=(-111,21) and the peak of cloth No.4 can be seen at (x,y)=(89,24) with respective fitness value of 0.788 and 0.66. GA converged individually in this experiment. The maximum value of the highest mountain of the peak represents for the cloth No.3 and also it has been confirmed that the cloth No.3 has been recognized among two different cloths.

### D. 1000 times recognition experiment

Fig. 12 shows the histogram of the pose estimation error for 1000 times recognition experiment for all cloths. Fig. 12 (a) and (b) show the mean error  $\pm 3\sigma$  (standard deviation) which are within  $\pm 10$  (mm) for position X and Y. The orientation for angle  $\theta$  is within  $\pm 10^\circ$  as shown in Fig. 12 (c). From these histogram results, all the cloths (No.1~No.12) can be recognized without any problems.

## CONCLUSION

In this paper, we have proposed a photo-model-based matching method to detect the target cloth and estimate the pose of the target object. GA can converge correctly between the generated model and the undefined target cloth. Moreover, in this experiment, 100 times generation is applied for the evolutionary process. The recognition accuracy is analyzed in terms of the histogram of pose estimation error. Recognizing the deformable different cloths automatically, estimation of the 3D pose of the target object and recognizing the cloth under changing and unknown environment by our proposed photo-model-based cloth recognition system have been confirmed experimentally to be applied in the garment factories.



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

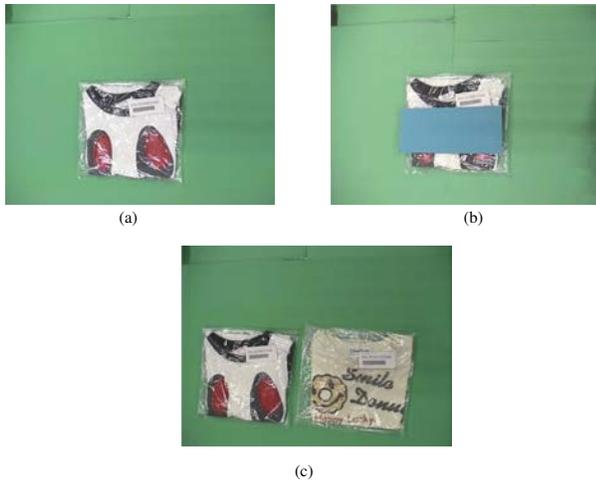


Fig. 11. Experiments of No.3 cloth (a) recognition without disturbance (b) recognition of partly hidden cloth (c) identification of two cloths

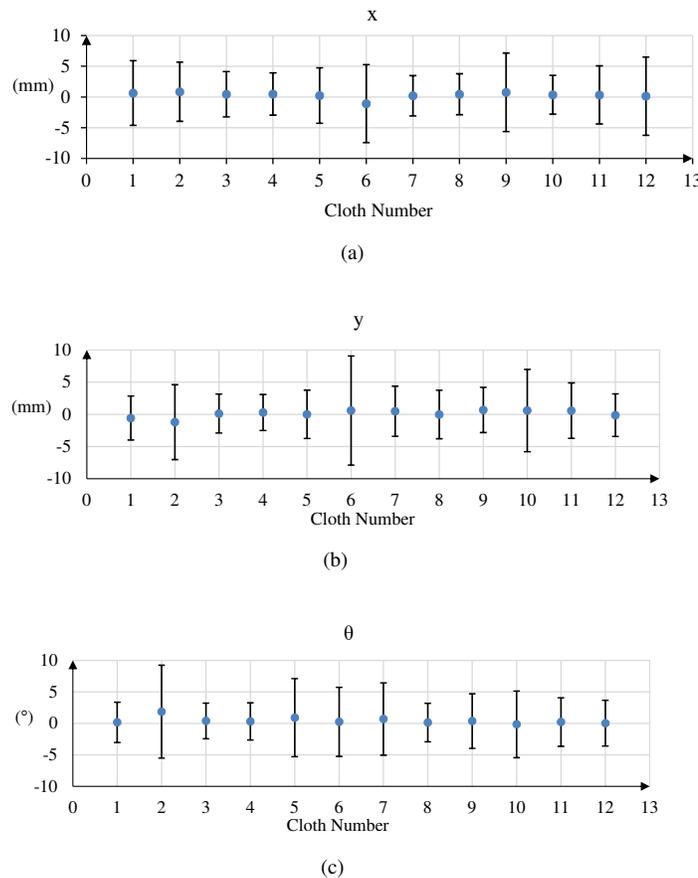


Fig. 12. Histogram of error in (a) position x(mm) (b) position y(mm) (c) angle (°) (No.1~No.12 Clothes)

## REFERENCES

[1] 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.

[2] Pengyu Liu, Mamoru Minami, Sen Hou, Koichi Maeda, Akira Yanou:"Eye-Vergence Visual Servoing on Moving Target,"*System Control Information Society Research Presentation Lecturer,Okayama University*, 2013.5.15-17

[3] 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)*

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

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

[6] 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.

[7] 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.

[8] Man, Kim-Fung, Kit-Sang Tang, and Sam Kwong. "Genetic algorithms: concepts and applications [in engineering design]." *IEEE transactions on Industrial Electronics* 43.5 (1996): 519-534.

[9] 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.

[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)*.

[12] 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.

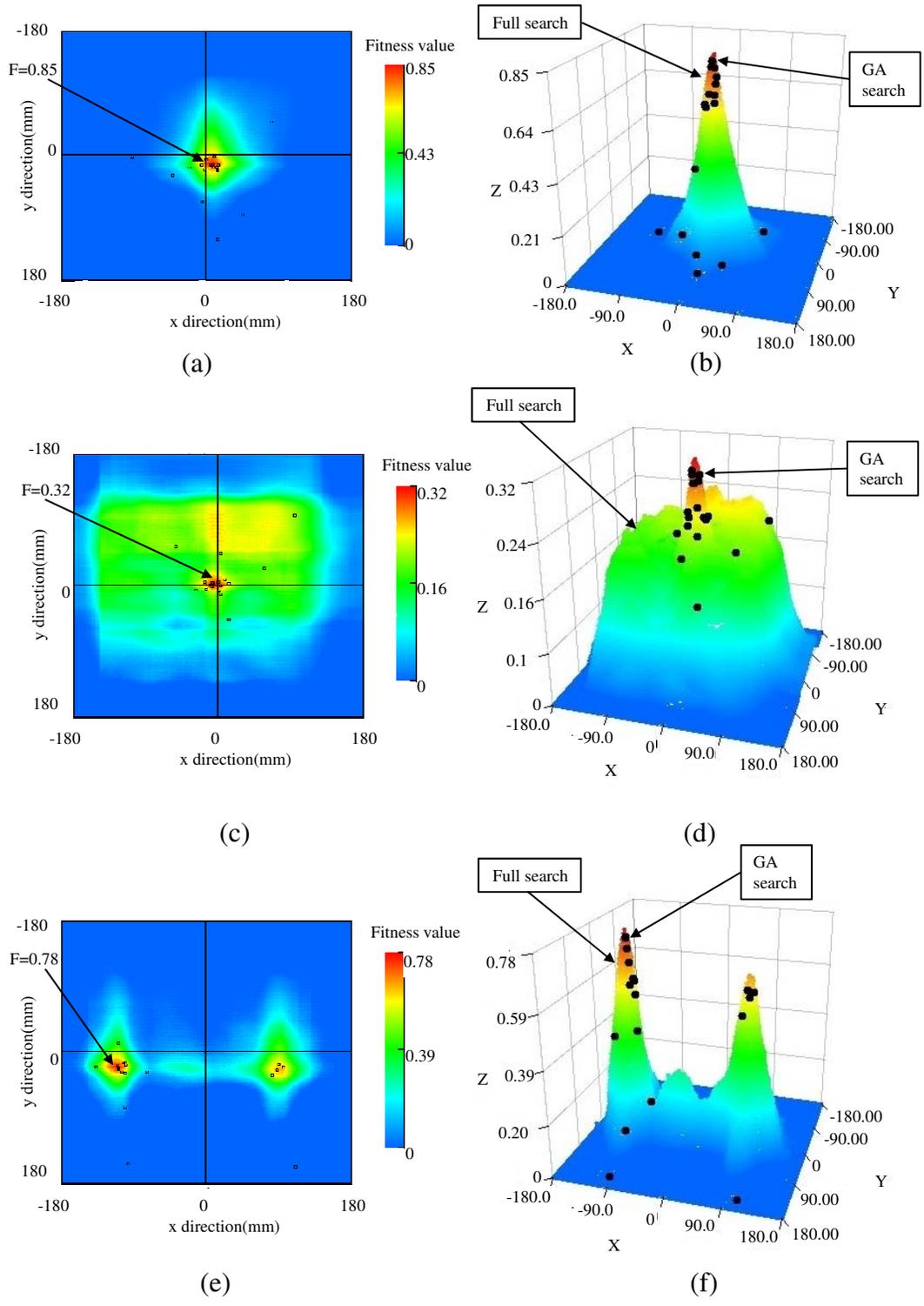


Fig. 13. Fitness function distribution of No.3 cloth in x-y plane (a) recognition without disturbance as 2D view (b) recognition without disturbance as 3D view (c) recognition of partly hidden cloth as 2D view (d) recognition of partly hidden cloth as 3D view (e) identification of two cloths as 2D view (f) identification of two cloths as 3D view (each black point in Fig.13 represents the result of GA which indicates the pose of each model and mountain means pose searched by full search method)