

Verification of illumination tolerance for photo-model-based cloth recognition

Khaing Win Phyu¹ · Ryuki Funakubo¹ · Ryota Hagiwara¹ · Hongzhi Tian¹ · Mamoru Minami¹

Received: 16 May 2017 / Accepted: 1 August 2017
© ISAROB 2017

Abstract Globally, robots are being used in garment factories according to their advantages such as high precision, flexibility, and productivity. However, it is difficult for robots to deal with deformable objects such as cloths and strings automatically. The main tasks of such robots are to recognize a deformable object and to pick up and place it at a designated position automatically. The deformable character of cloths seems a main hindrance for automatic handling by robots, especially if the cloth is unique. In this paper, a cloth handling robotic system for unique cloths is proposed with consideration of verification of illumination tolerance. This robotic system comprises two main portions: the first portion generates a model of cloths using an image taken by a single camera, and the second portion estimates the relative pose of cloth appeared in the view of two cameras using the generated model. In our cloths' pose estimation, the 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. The illumination tolerance of the proposed system under different light conditions of different light sources was verified experimentally for evaluating the proposed system from the view point of practicality. The fluorescent light and the light-emitting diode (LED) light are used in this experiment, having confirmed that the proposed system can

recognize cloths in condition that the light environments have varieties.

Keywords Cloth recognition · Model-based matching · Genetic Algorithm · Illumination tolerance

1 Introduction

Robots equipped with hand-eye camera have been utilized widely as industrial robots, e.g., bin-picking robot with a hand-eye camera, since the robot can choose the observing point arbitrarily. As the cases that these robots use a single camera are conventional in factories, a problem has arisen that the accuracy of estimated position along with the camera sight direction tends to be deteriorated. Therefore, a recognition method [1] that combines a camera and a laser-range-finder has been researched. However, such kind of method congenitally has a difficulty named here "Object Identification Problem". The method that estimates target's pose (position and orientation) using camera and laser-range-finder assumes a preposition that the object projected into camera image and the object detected by the laser-range-finder be identical. If this assumption is not kept valid, the robot's motion may be out of order. There has been a concept called "Sensor Fusion" [2] that integrates information from multiple sensors to measure. In the sensor fusion strategy, the same kind of problem above-mentioned, i.e., Object Identification Problem, still remained unsolved, because whether the sensors detect the same object may not be justified.

In a situation to estimate 3D position and orientation from image information of multiple cameras, the similar problem above-mentioned may happen, that is whether the points on an object in 3D space correspond correctly to the

This work was presented in part at the 22nd International Symposium on Artificial Life and Robotics, Beppu, Oita, January 19–21, 2017.

✉ Mamoru Minami
minami-m@cc.okayama-u.ac.jp

Khaing Win Phyu
p6nn3hpz@s.okayama-u.ac.jp

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

points in the images of multiple cameras may not be justified. Then, the problem existing in the process to reconstruct a 3D pose from points in images of multiple cameras, i.e., “Corresponding Points Identification Problem,” may arise. It is common to search for corresponding points in multiple cameras images using epipolar geometry [3]. When the searched corresponding points are not truly identical ones, the reconstructed 3D points include inevitably position errors, meaning that the reconstructed 3D object through multiple cameras images represents an incorrect 3D shape and pose. On the other hand, concerning control method of robots by visual servoing [4], a prediction method that utilizes an object’s dynamical model and a nonlinear observer has been proposed in [5], which has been stalked by a problem that it takes some time before the recognition error reduces near zero.

To avoid above problems, a single camera system may be a possible choice. In case of single camera observation, it has been understood that the position measurement in camera depth is difficult [6] and the orientation measurement is hard to estimate correctly [7]. In [7], the position and the orientation (only heading angle) of an underwater vehicle with respect to a target object are estimated using images from a single camera. To compensate the above-mentioned demerits of pose estimation using a single camera, Luca et al. [4] have proposed a method to estimate the distance between the camera and an object by utilizing the time profile of the camera’s motion trajectory, called “Moving View Points”. As a result of above discussions concerning pose estimation by sensor fusion, multi-cameras, and single camera, there has been still some difficulties for pose estimation, even though the target object is solid and not deformable.

Based on the above discussions, handling deformable and single/unique second-hand cloth deems to be a challenging topic in robotic researches. In this circumstances of robotics researches, the authors have been requested from T2K Co., Ltd (Logistics company to deal with second-hand cloths) to construct robotic system to handle the second-hand cloths. Then, authors have started to make a robot to be able to handle single/unique and deformable cloths; a photo-model-based cloth recognition method has been devised [8–10]. A preceding work in [8] is a pioneer whose strategy is based on a combination of sensor and cameras. Then, it cannot escape from above-mentioned Object Identification Problem. In our previous work [9, 11], recognizing deformable cloths automatically and estimation of the 3D pose of the clothes using three cameras have been proposed. In the proposed cloths’ 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 proposed handling system has a merit that since it generates automatically photo-models of target

cloths, this system can recognize unique object and estimate its pose with lighting adaptiveness, enabling the system to handle unique targets in lighting condition varieties.

Since the practicality of industrial robots used in factory has been indispensable, the tolerances against illumination changing have been checked [10]. This paper is an extension of [10] that includes new evaluation of illumination varieties and fitness function distribution, which can explain why the proposed photo-model-based system can be tolerable against illumination varieties.

2 Proposed cloth handling system

Our research group has developed a robot that can automatically perform classifying and handling the garment by visual servoing. The configuration of the system is shown in Fig. 1. In this configuration, three cameras are used as vision sensors. The first camera is used for generating models. The other two cameras, which are fixed at the end-effector of the manipulator (PA10 robot), are used for recognition based on the photo-model. A robot named as PA10 with 7-DoF (Degree of Freedom) is used for recognition and pose detection operations. The aim of the robot using proposed system is to pick up the cloth after recognizing it and set the cloth into the desired place.

The proposed system is aimed to be applied to the real world with lighting condition variations that means that the working room has window. When the robots equipped with cameras are used to operate day and night in both indoor and outdoor spaces, most of them are limited to work in restricted environments in terms of illumination variation. Therefore, the robustness of visual recognition against illumination variations is important and the illumination tolerance of the vision-based system has to be confirmed.

Based on this motivation, cloths recognition experiments under different lighting conditions [100 (lx), 400 (lx), 700 (lx), 1000 (lx), and 1300 (lx)] for different light sources were conducted to verify the illumination tolerance of the

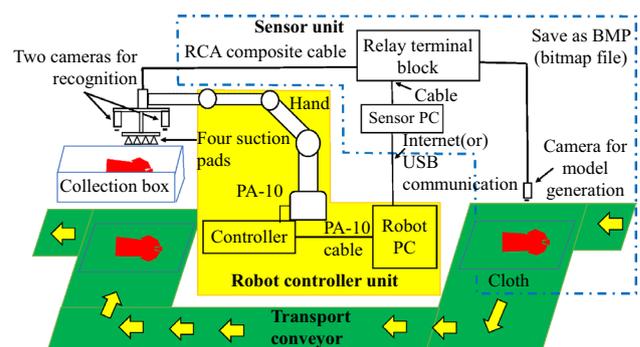


Fig. 1 System configuration of a cloth handling system

proposed system. Why these illumination conditions are chosen is discussed in Sect. 3.6. In this study, two different light sources, fluorescent and light-emitting diode (LED), are used to provide different levels of illumination for recognition experiment. The recognition performances under different light conditions are analyzed.

The remainder of the present paper is summarized as follows: Section 2 describes the proposed cloth handling system. Section 3 presents the photo-model-based cloth recognition. Section 4 describes the experimental environment. Section 5 describes the experimental results with discussion. Section 6 concludes this paper.

3 Photo-model-based cloth recognition

According to the literature review, object recognition methods are mainly classified into three categories:

- Image-based (Feature-based);
- Appearance-based (Template-based);
- Model-based.

Featured-based approach [12] is based on the error between current and desired features on the image plane, and does not implicate any evaluation of the position and orientation of the target object. The main task of this method is to select a set of visual feature points, lines, or moments of regions. This method is mostly applied to head pose estimation. In this method, the features possibly be set at points on faces like the corners of the eye or mouth that may not accurate and sometimes cannot be recognized definitely, because it cannot resist the influence of facial expression, light environment changing, and occlusion.

In an appearance-based method [13], it is needed to perform template matching completely. The image is compared with the reference templates to determine which one closely corresponds to the image. This method may address the problem in terms of time-consuming.

Model-based approach [14] is to search the target in the image using a model, which is constructed based on known information of the target object. Our proposed method is included in this category.

3.1 Kinematics of stereo-vision

Figure 2 shows a perspective projection of dual-eye vision system. The specifications of the coordinate systems in dual-eye vision system and position vector of an arbitrary i th point on the 3D model based on each coordinate system in Fig. 2 are as follows:

- Σ_W : World coordinate system.

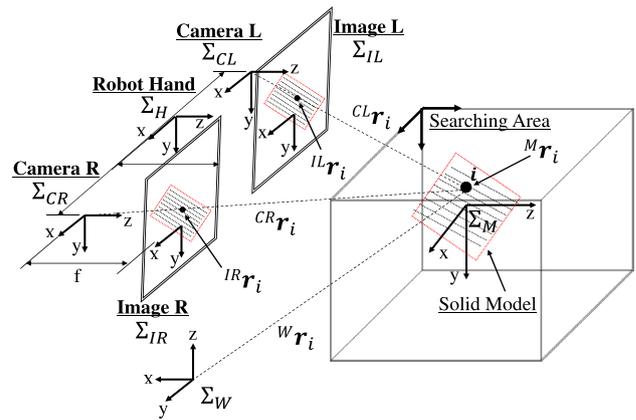


Fig. 2 Perspective projection of dual-eye vision system

- Σ_M : Coordinate system attached to a model.
- Σ_{CL} : Coordinate system of the left camera.
- Σ_{CR} : Coordinate system of the right camera.
- Σ_{IL} : Coordinate system of the left camera image.
- Σ_{IR} : Coordinate system of the right camera image.
- ${}^{CR}r_i$: Position of an arbitrary i th point on the 3D model based on Σ_{CR} .
- ${}^{CL}r_i$: Position of an arbitrary i th point on the 3D model based on Σ_{CL} .
- ${}^W r_i$: Position of an arbitrary i th point on the 3D model based on Σ_W .
- ${}^M r_i$: Position of an arbitrary i th point on the 3D model in Σ_M .
- ${}^{IL}r_i$: Position of an arbitrary i th point on the 3D model projected into the left camera image Σ_{IL} .
- ${}^{IR}r_i$: Position of an arbitrary i th point on the 3D model projected into the right camera image Σ_{IR} .

As shown in Fig. 2, a point ${}^M r_i$ in 3D space is projected into the left and right cameras images as ${}^{IL}r_i$ and ${}^{IR}r_i$, respectively. In Eq. (1), the position vector of the arbitrary i th point on the 3D model projected into the left and right cameras images Σ_{IL} and Σ_{IR} represented by ${}^{IL}r_i$ and ${}^{IR}r_i$ can be described as the functions (f_R and f_L). The function f_R and f_L include i th point on the 3D model as ${}^M r_i$, and the relative pose of the model with respect to the left and right cameras (Σ_{CL} and Σ_{CR}) as ${}^{CL}\phi_M$ and ${}^{CR}\phi_M$. This relation connects the arbitrary points being predefined and fixed on the model and projected points on the left and right images with the variables ${}^{CL}\phi_M$ and ${}^{CR}\phi_M$, which is considered to be unknown in this paper:

$$\begin{cases} {}^{IR}r_i = f_R({}^{CR}\phi_M, {}^M r_i) \\ {}^{IL}r_i = f_L({}^{CL}\phi_M, {}^M r_i). \end{cases} \quad (1)$$

Please note that we abbreviate ${}^{CL}\phi_M$ and ${}^{CR}\phi_M$ to ${}^C\phi_M$. The measurement of the pose (${}^C\phi_M$), ${}^C\phi_M = [{}^C x_M, {}^C y_M, {}^C z_M, {}^C e_{1M}, {}^C e_{2M}, {}^C e_{3M}]^T$ will be explained in the following sections.

3.2 Cloth model generation

Apart from the conventional model-based matching methods in which models are predefined in a computer system using known information of the object, a model of deformable cloths is generated directly from a photo-image of a single camera. Among three cameras that are used as the vision sensors in the proposed system, the first camera fixed in the workspace is used to generate a photograph cloth model. Figure 3 shows the process of model generation. The process of model generation technique is described as the following.

- As shown in Fig. 3a, a background image is captured by the first camera and the averaged hue value of the background image is calculated.
- The cloth is put on the background, as shown in Fig. 3b.
- The hue value of each point in the image acquired by scanning individual pixel is compared with the averaged hue value of the background image to generate a set of points representing the surface space S_{in} of the model, as shown in Fig. 3c.
- As shown in Fig. 3 (d), the outside space S_{out} of the model is generated that envelops S_{in} .

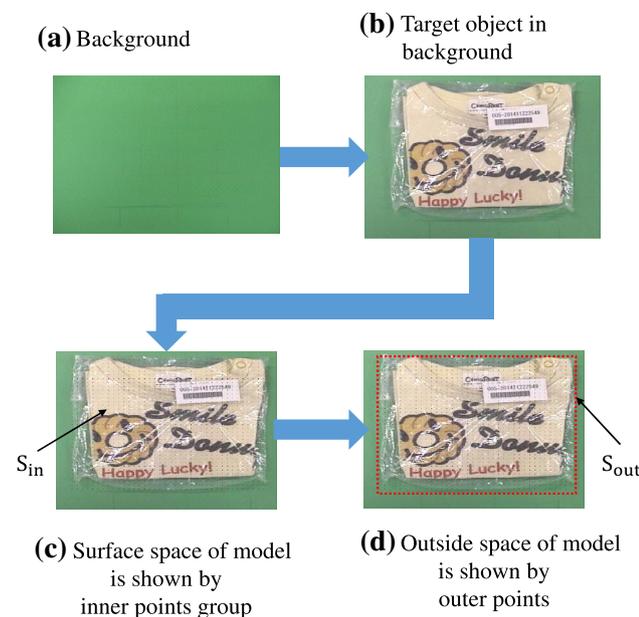


Fig. 3 Model generation process

3.3 3D model-based matching

After a model is generated by a cloth model generation process, it is used for 3D pose recognition of the target object through model-based matching. 3D pose of the 3D model ${}^C\phi_M = [{}^C x_M, {}^C y_M, {}^C z_M, {}^C e_{1M}, {}^C e_{2M}, {}^C e_{3M}]^T$ is composed of three position variables and three orientation variables in quaternion. The generated model is projected from the 3D space in the 3D searching space into the left and right 2D images planes, as shown in Fig. 4. In Fig. 4, subfigure on the top shows a 3D model composed of $S_{in}(\phi_M)$ (inner dotted points) and the outside space enveloping S_{in} denoted as $S_{out}(\phi_M)$ (outer dotted points). The subfigure on the left bottom of Fig. 4 shows the left 2D searching model, $S_L(\phi_M)$, including $S_{L,in}(\phi_M)$ and $S_{L,out}(\phi_M)$. $S_R(\phi_M)$, $S_{R,in}(\phi_M)$, and $S_{R,out}(\phi_M)$ represent as the right 2D searching model, as shown in subfigure on the right bottom of Fig. 4. Despite that the searching model assumed to set in 3D searching space is flat shape, it has a varieties of pose in the space, we named the model as 3D model.

A fitness function (explained in next section) is defined for the evaluation of the correlation between the projected models (how different models with different poses are generated in GA is explained in Sect. 3.5) and the images from the dual-eye cameras. The pose of the best model with the highest fitness value that coincides with the captured images from left and right cameras is selected as an estimated pose. Please refer to the previous works [14–16] for detailed explanation on 3D model-based matching which can be seen in the previous works.

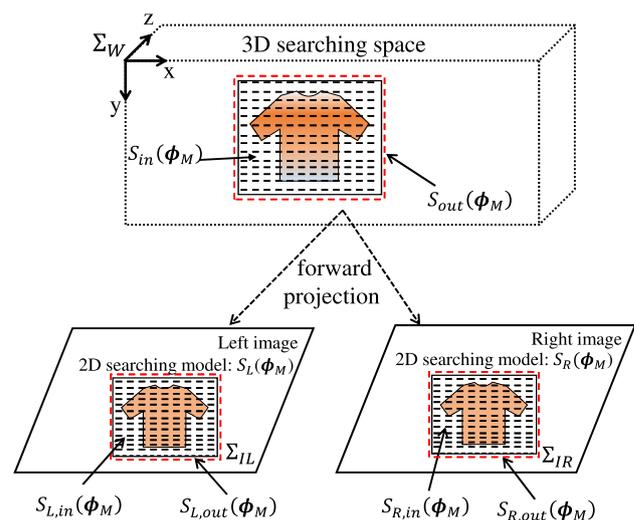


Fig. 4 Definition of a solid model in the 3D searching area (subfigure on the top of Fig. 4) and left/right 2D searching models (subfigures on the left and right bottom of Fig. 4)

3.4 Definition of the fitness function

The fitness function is constructed to estimate how much the matching degree between the projected model defined by its pose (ϕ_M) and the captured images on the left and right 2D searching areas. If the projected 2D searching models ($S_L(\phi_M)$ and $S_R(\phi_M)$) absolutely coincide with the captured target object (cloth) in the left and right images, the fitness function that represents correlation of that model with the pose of (ϕ_M) has been designed to have a maximum value. Therefore, the fitness value distribution for all models will represent a mountain shape with a peak that represents the pose of the real 3D target object. The concept of fitness function in this study can be said to be an extension of the work in [17] in which different models including a rectangular shape surface-strips model were evaluated using images from a single camera. The evaluation function $F(\phi_M)$ is designed as follows:

$$F(\phi_M) = \left\{ \left(\sum_{\substack{IRr_i \in \\ S_{R,in}(CR \phi_M)}} p(IRr_i) + \sum_{\substack{IRr_i \in \\ S_{R,out}(CR \phi_M)}} p(IRr_i) \right) + \left(\sum_{\substack{ILr_i \in \\ S_{L,in}(CL \phi_M)}} p(ILr_i) + \sum_{\substack{ILr_i \in \\ S_{L,out}(CL \phi_M)}} p(ILr_i) \right) \right\} / 2$$

$$= \{F_R(CR \phi_M) + F_L(CL \phi_M)\} / 2 \tag{2}$$

Since evaluation functions for left and right cameras are the same and total fitness function is average of them, let us explain the contents of Eq. (2) here only in the case of left camera. The evaluation of every points in the input image that lie inside the surface model frame and outside area of the model frame are represented as $ILr_i \in S_{L,in}(\phi_M)$ and $ILr_i \in S_{L,out}(\phi_M)$ respectively. $p_{L,in}(ILr_i)$ and $p_{L,out}(ILr_i)$ are calculated using the following equations:

$$p_{L,in}(ILr_i) = \begin{cases} 2, & \text{if } (|H_{IL}(ILr_i) - H_{ML}(ILr_i)| \leq 30); \\ -0.005, & \text{if } (|\bar{H}_B - H_{IL}(ILr_i)| \leq 30); \\ 0, & \text{otherwise.} \end{cases} \tag{3}$$

$$p_{L,out}(ILr_i) = \begin{cases} 0.1, & \text{if } (|\bar{H}_B - H_{IL}(ILr_i)| \leq 20); \\ -0.5, & \text{otherwise.} \end{cases} \tag{4}$$

where $S_{L,in}$ is the space of coordinates on the surface area of the model, $S_{L,out}$ is the space of coordinates on the outside area of the model, $H_{IL}(ILr_i)$ is the hue value of the left camera image at the point ILr_i (ith point in $S_{L,in}$), $H_{ML}(ILr_i)$ is the hue value of the point ILr_i (ith point in $S_{L,in}$) on the model, \bar{H}_B : the average hue value of the background image. Please note that hue value is used in the functions of $p_{L,in}(ILr_i)$ and $p_{L,out}(ILr_i)$, because hue value is less sensitive to the illumination variation. Eqs. (3) and (4) are designed to provide a peak in fitness value distribution by reducing noises. The evaluation values are tuned experimentally.

Here, detailed explanation on Eqs. (3) and (4) is presented. In Eq. (3), if the hue value of each point of captured images, which lies inside the surface model frame $S_{L,in}$, is similar to the hue value of each point in a model, i.e., the difference be less than 30, the fitness value will increase with the voting value of “+2.” The fitness value will decrease with the value of “-0.005” for every point of cloths in the left camera image that are similar to the average hue value of the background. Similarly, in Eq. (4), if the hue value of each point in the left camera image, which are in $S_{L,out}$, is the same to the hue value of background with the tolerance of 20, the fitness value will increase with the value of “0.1.” Otherwise, the fitness value will be decreased with the value of “-0.5.” The detailed explanation of fitness function is discussed in [14, 17].

3.5 Genetic algorithm (GA)

The main problem of searching for the pose of the target cloths 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 in GAs gene coincides with the target in the image.

The maximum value of the fitness function can be searched in numerous ways. Among them, the full search method is the simplest and easiest way. It is envisaged that the peak in fitness distribution will be searched by scanning all possible points. However, this method addresses a drawback in terms of computing time. GA can recognize the target in a short time, without the need to scan all the possible points. Therefore, GA process is applied to find the maximum value in the fitness distribution as an optimization solution.

In Fig. 5, target object is represented by solid line as cloth and each model is represented by dotted lines. Many models, as shown in Fig.5, have the same shape and color with the target object, but each model has various pose. The randomly generated GA individuals are come from the different poses of the different models. The 60 individuals of GA are used in this experiment. Each individual chromosome consists of six variables. Each variable is coded by 12 bits. The first three variables of a model in 3D space (t_x, t_y, t_z)

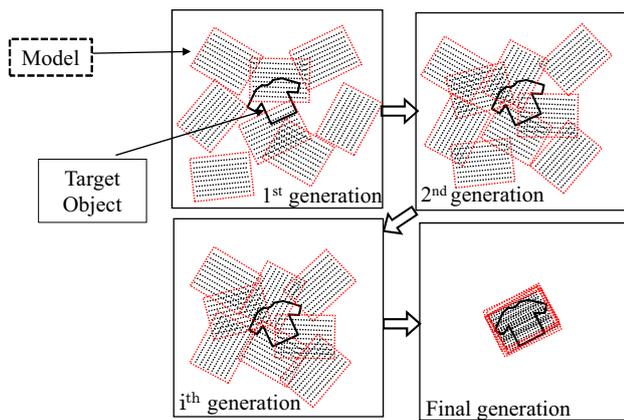


Fig. 5 GA evolution process, representing that 3D models converge into the real target object (cloth), through evaluations on fitness function

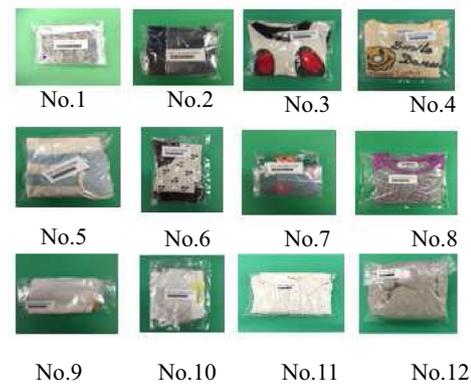


Fig. 7 Target objects (No.1–No.12) cloths: each has different colors, sizes, shapes, and weights

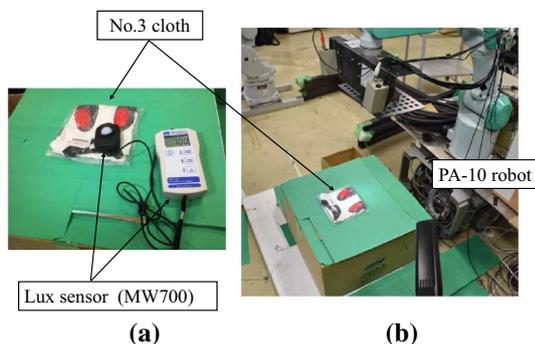


Fig. 6 Photo of experiment for recognition of cloth No.3 under 700 (lx) **a** (lx) is measured by a Lux sensor (MW700); **b** recognition under 700 (lx) using PA10 robot

represent the position and the last three variables ($\epsilon_1, \epsilon_2, \epsilon_3$) to do the orientation. And then, the definition of GA individual is shown as follows:

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

The above 60 individuals are evaluated by the fitness function value, as shown in Eq. (2). The fitter one is selected to regenerate the next generations. In the final generation, the best model with the highest fitness value represents the most trustful pose, as shown in Fig. 5.

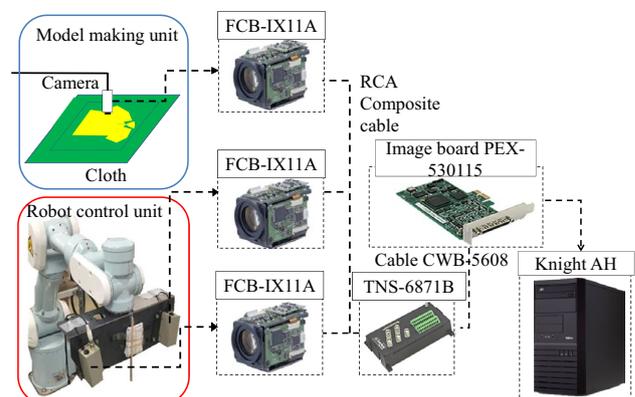


Fig. 8 Sensor unit configuration

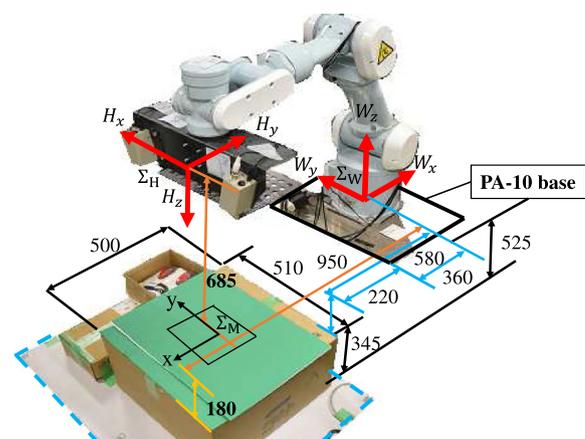
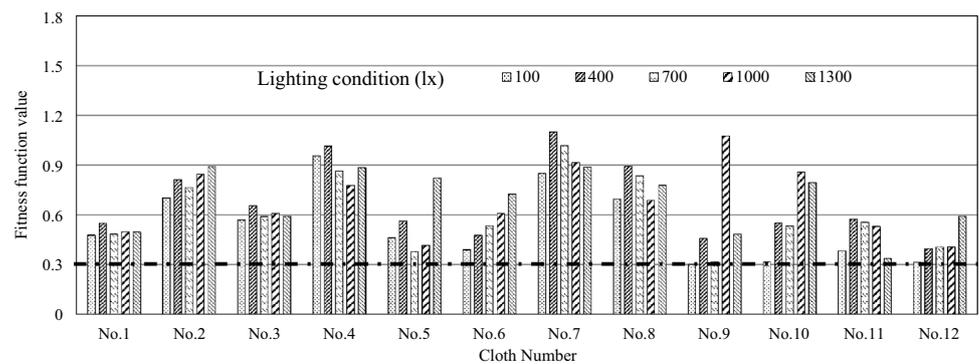


Fig. 9 Coordinate systems of robot and end-effector (unit is (mm) in Fig. 9)

Fig. 10 Fitness function value of each cloth (No.1– No.12) under fluorescent light (see Table 1 for the entire numerical results of Fig. 10)



3.6 Verification of cloth recognizable illumination range

The different types of lighting sources applied in this experiment are as follows:

- Fluorescent;
- Light-emitting diode (LED).

Three common light conditions such as 100 (lx), 700 (lx), and 1300 (lx) are provided using both lighting types. According to recommendation from Japanese Industrial Standard (JIS), 75 (lx)–150 (lx) is commonly used in the warehouse and 750 (lx)–1500 (lx) is recommended value of fine visual work of inspection. In addition, 700 (lx) is used for ordinary illumination. Therefore, 100 (lx), 700 (lx), and 1300 (lx) are selected in the experiment to verify the proposed system's robustness in all possible working areas. Figure 6 shows the photo of experiment for recognition of cloth No.3 under 700 (lx) lighting condition, where all tested cloths are listed in Fig. 7. Illuminance is measured by a Lux sensor (MW700), as shown in Fig. 6a, and recognition under 700 (lx) using PA-10 robot is shown in Fig. 6b, respectively.

The initial aim of the proposed system is to be applied in a mail-sending procedures that have been conducted by the T2K Co., Ltd., in which employees classify a large number of second-hand cloths manually every day. Since the

cloths are deformable objects and unique, since it is second hand, no definition of cloths can be predefined in a computer beforehand. Consequently, it is difficult to handle a wide variety of cloths that are irregular in shape and size. T2K explained us this condition and requested to collaborate with our research group to solve the above-mentioned problem. Therefore, we conducted a joint research with T2K Co., Ltd. Different 12 cloth samples have been chosen by the cooperation of T2K Co., Ltd., as shown in Fig. 7. T2K Co., Ltd., though the 12 samples represent general varieties of cloths, so they have chosen 12 cloth samples to conduct in the experiment.

3.7 Sensor unit

The configuration of the sensor unit is shown in Fig. 8. In this configuration, three cameras (FCIX11A) are used as vision sensors. The first camera is used for generating models which are placed on model making unit, as shown in Fig. 8. The other two cameras, which are fixed at the end-effector of the manipulator (PA10 robot), are used for recognition based on the photograph model. In Fig. 8, the robot control unit is used for recognition and pose detection operations. The sensor unit named as NS001 is operated in response to an instruction from the robot PC by serial

Fig. 11 Fitness function value of each cloth (No.1– No.12) under LED light (see Table 2 for the entire numerical results of Fig. 11)

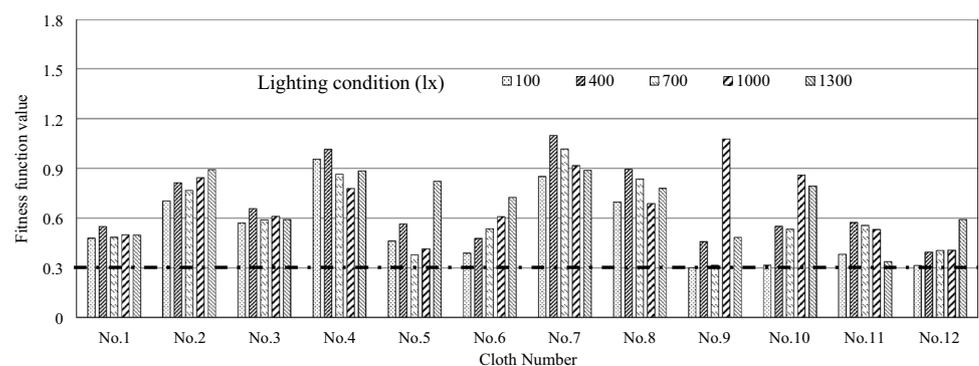


Table 1 Fitness function value under different light conditions [100 (lx), 400 (lx), 700 (lx), 1000 (lx), and 1300 (lx)] using fluorescent light

Illuminance (lx)	Fitness function value of each cloth (No.1–No.12)											
	No.1	No.2	No.3	No.4	No.5	No.6	No.7	No.8	No.9	No.10	No.11	No.12
100	1.005	0.475	1.293	1.512	0.738	0.372	0.477	0.973	0.965	0.756	1.550	1.019
400	0.675	0.722	1.203	1.296	0.757	0.515	0.739	0.683	0.887	0.715	1.148	0.915
700	0.685	0.668	1.14	1.27	0.864	0.475	0.784	0.859	0.965	0.756	1.55	1.02
1000	0.625	0.705	1.147	1.260	0.911	0.662	0.861	0.748	0.535	0.631	0.837	0.994
1300	0.511	0.713	1.10	1.25	0.789	0.749	0.778	0.696	0.459	0.750	0.766	0.937

Table 2 Fitness function value under different light conditions (100 (lx), 400 (lx), 700 (lx), 1000 (lx) and 1300 (lx)) using LED light

Illuminance (lx)	Fitness function value of each cloth (No.1–No.12)											
	No.1	No.2	No.3	No.4	No.5	No.6	No.7	No.8	No.9	No.10	No.11	No.12
100	0.478	0.702	0.570	0.955	0.460	0.389	0.851	0.696	0.301	0.317	0.382	0.315
400	0.548	0.812	0.655	1.014	0.563	0.475	1.097	0.893	0.456	0.549	0.572	0.3956
700	0.484	0.766	0.588	0.864	0.378	0.535	1.02	0.835	0.317	0.533	0.556	0.406
1000	0.498	0.843	0.609	0.777	0.414	0.606	0.915	0.687	1.076	0.857	0.529	0.408
1300	0.497	0.891	0.591	0.884	0.823	0.724	0.887	0.779	0.482	0.794	0.338	0.590

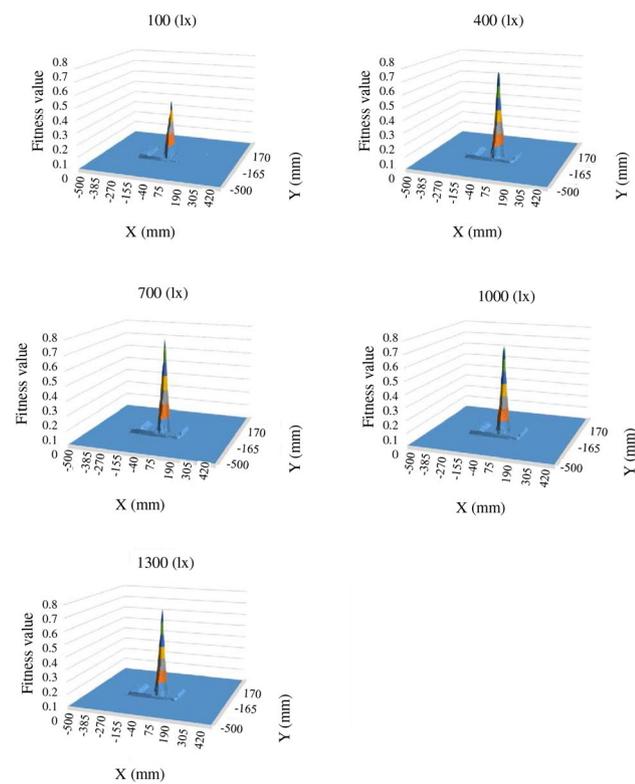


Fig. 12 Fitness function distribution of cloth No.2 in x–y plane (fluorescent light)

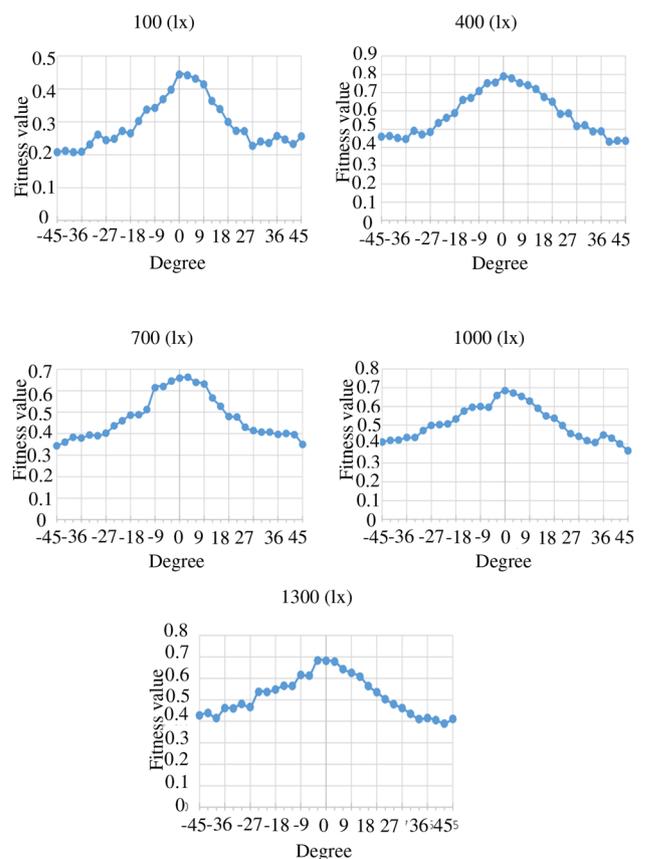


Fig. 13 Fitness function distribution of angle for cloth No.2 (fluorescent light)

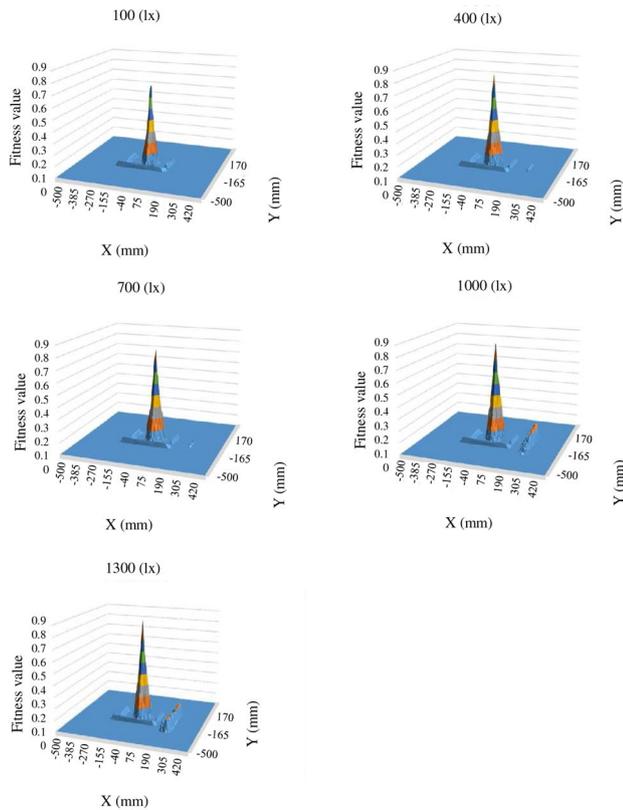


Fig. 14 Fitness function distribution of cloth No.2 in x-y plane (LED light)

communication. Configuration of the sensor unit NS001 is as follows:

- Image board PEX-530115;
- Relay terminal block TNS-6871B;
- Two cameras FCB-IX11A;
- Camera mounting base.

4 Experimental environment

Figure 9 shows the experimental layout and the coordinate systems of the system, i.e., the world coordinate system (Σ_W), the hand coordinate system (Σ_H), and the cloth coordinate system (Σ_M), that are used in the experiments, respectively. The cloth coordinate system (Σ_M) is set at $x = 0$, $y = 0$, and $z = 685$ (mm) with respect to the hand coordinate system (Σ_H). Figure 7 shows the 12 different cloths samples (No.1–No.12), each has different colors, sizes, shapes, and weights, used in this experiment. Each item of cloth is recognized individually to confirm the influence of illumination on recognition performance. This experiment was mainly focused on the recognition experiment and handling experiment is our follow-up work.

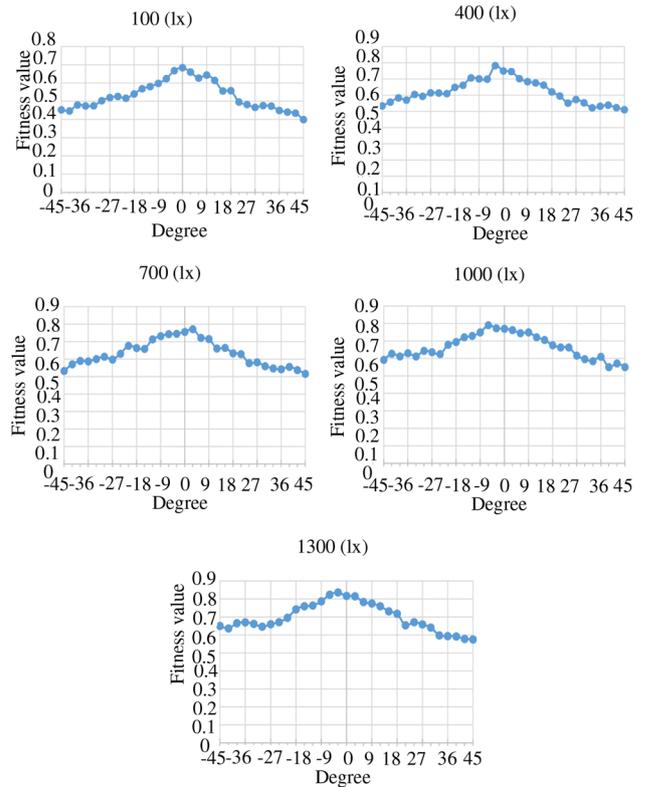


Fig. 15 Fitness function distribution of angle for cloth No.2 (LED light)

5 Results and discussion

Figures 10 and 11 show the fitness function value of 12 cloths separately recognition under five common light conditions [100 (lx), 400 (lx), 700 (lx), 1000 (lx), and 1300 (lx)] using fluorescent light and LED light. In each figure, the vertical axis represents the highest fitness function value that is found by GA and the horizontal axis shows the cloth number (No.1–No.12).

In Fig. 10, the fitness values of recognition of cloth No.1 are 1.0 for 100 (lx), 0.675 for 400 (lx), 0.685 for 700 (lx), 0.625 for 1000 (lx), and 0.511 for 1300 (lx), respectively. The fitness function value of No.3 cloth is nearly around 1.0–1.3 for all illuminances. No.4 and No.11 cloths for illuminance of 100 (lx) have larger fitness function value than the other cloths numbers under different illuminances. There is no significant difference among fitness values for cloth No.5 under three illuminance conditions with fitness value of 0.738 [100 (lx)], 0.757 [400 (lx)], 0.864 [700 (lx)], and 0.789 [1300 (lx)]. The highest fitness value for cloth No.5 under 1000 (lx) is about 0.911. Similarly, cloth No.10 has nearly equal fitness function value for all illuminances with the value of around 0.75. No.7 cloth has approximately equal value of 0.7 at 700 (lx) and 1300 (lx). Then, the two fitness values of No.8 cloth are nearly equal to 1.0 at 100

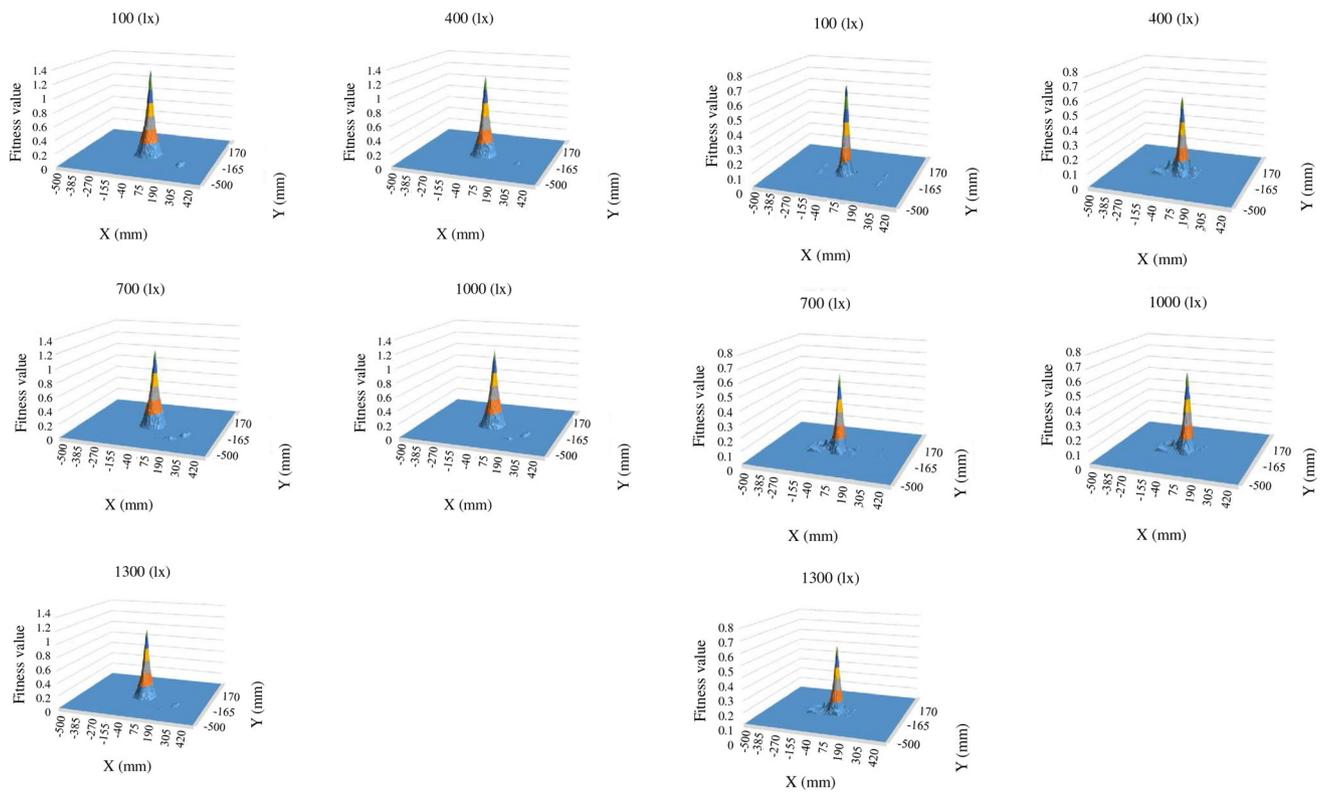


Fig. 16 Fitness function distribution of cloth No.3 in x–y plane (fluorescent light)

Fig. 18 Fitness function distribution of cloth No.3 in x–y plane (LED light)

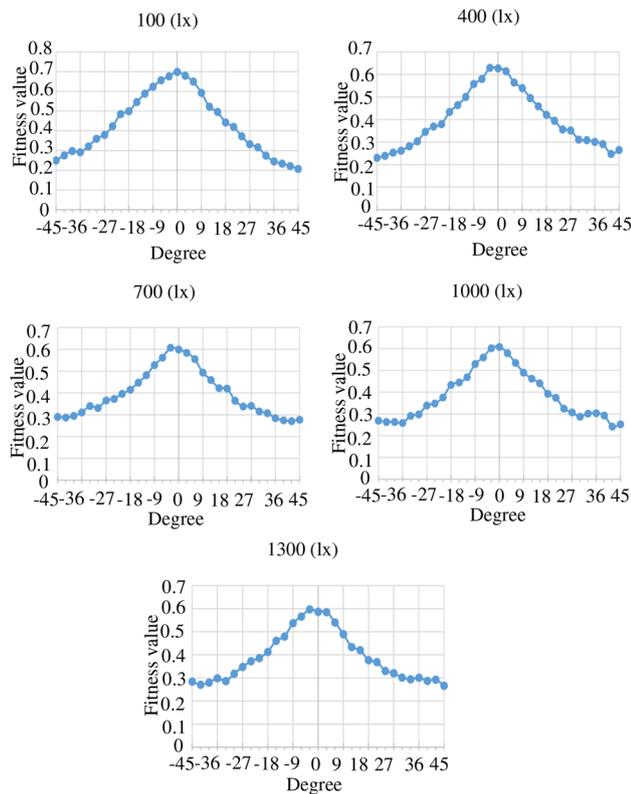


Fig. 17 Fitness function distribution of angle for cloth No.3 (fluorescent light)

(lx) and 700 (lx). For cloth No.9, the one fitness value close to the minimum value (0.3) with fitness values of 0.459 [1300(lx)]. However, the fitness value is about 1.0 at 100 (lx). The fitness value of No.12 cloth remains nearly 1.0 for five light conditions in case of fluorescent light environment. In Fig. 11, the fitness function values for all different cloths numbers are not prominently changed under all light conditions. However, the fitness value for No.9, No.10, and No.12 cloths under 100 (lx) is 0.301, 0.317, and 0.315, respectively. The highest fitness value appears at cloth No.7 with 1.097 [400 (lx)].

In summary, according to the graphical results shown in Figs. 10 and 11, it can be generally said that the overall fitness function value of recognition for all cloths under fluorescent light is higher than that of LED light condition. In Fig. 10, the brighter the light condition, the less the fitness function value become except for No.2 and No.6 cloths. On the other hand, in Fig. 11, the brighter the light condition, the higher the fitness function value except for No.11 cloth. According to the experimental result, it is confirmed that all cloths can be recognized with the minimum fitness value of 0.3 under different light conditions. The results of the fitness values are summarized in Tables 1 and 2 numerically.

The cloth recognition experiment has been conducted under the five illuminations 100 (lx), 400 (lx), 700 (lx),

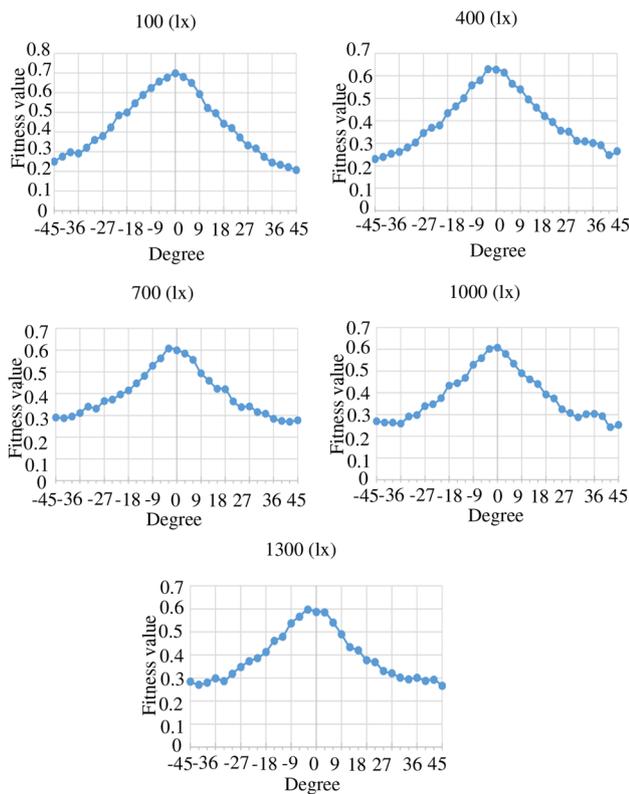


Fig. 19 Fitness function distribution of angle for cloth No.3 (LED light)

1000 (lx), and 1300 (lx) for the verification of illumination tolerance. The number of differences between each illumination range is 300 (lx) steadily. The difference between each illumination is small and detail. The detailed analysis of the experimental result has been confirmed to apply in the different lighting environments. These results intend to apply the actual working environment.

Figures 12, 14, 16, 18, 20, and 22 show the fitness distribution in x - y plane of cloths No.2, No.3, and No.11 with five different illuminations [100 (lx), 400 (lx), 700 (lx), 1000 (lx), and 1300 (lx)] under fluorescent light condition and LED light condition. Overall, the fitness distribution in x - y plane results of cloths No.2, No.3, and No.11 can also be seen as graphical results in Figs. 10 and 11 and numerical value in Tables 1 and 2, respectively. In addition, Figs. 13, 15, 17, 19, 21, and 23 show the fitness distribution for orientation of cloths No.2, No.3, and No.11 with five different illuminations [100 (lx), 400 (lx), 700 (lx), 1000 (lx), and 1300 (lx)] under two different light sources (fluorescent and LED). The proposed system has been analyzed in orientation of three cloths because of their criteria. These three different cloths have different characteristic as follows.

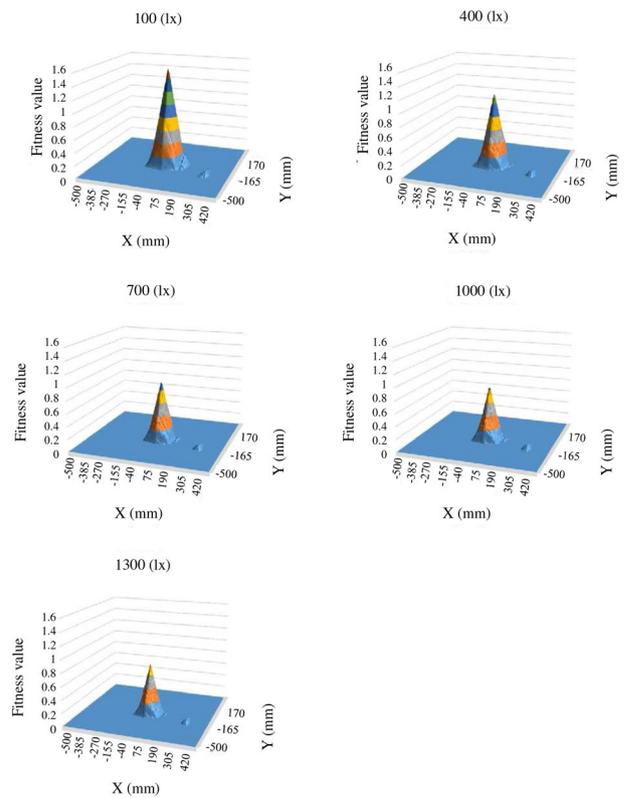


Fig. 20 Fitness function distribution of cloth No.11 in x - y plane (fluorescent light)

- Cloth No.2 has a black color, small size, and with some reflection.
- Cloth No.3 has colorful and small size.
- Cloth No.11 has a white color without colorful patterns and large size.

According to experimental results, it can be analyzed that how the different light illuminations effect on the fitness distribution of different cloths whose color, texture, and pattern are different. For example, there is a significant difference in the height of fitness value of cloth No.11 between the case of 100 (lx) and 1300 (lx), as shown in Figs. 20 and 22. However, the positions recognized by the proposed system do not change, even though the heights of the fitness distributions are affected by illumination. These experimental results show the robustness of the proposed system against illumination conditions. However, there are some errors in recognition of orientation effected by illumination. For example, the highest fitness distribution of cloth No.11 in 1300 (lx) of fluorescent light, and that in 100 (lx), 400 (lx), 1000 (lx), and 1300 (lx) of LED light do not appear at 0 (degree), as shown in Figs. 21 and 23. However, it can be

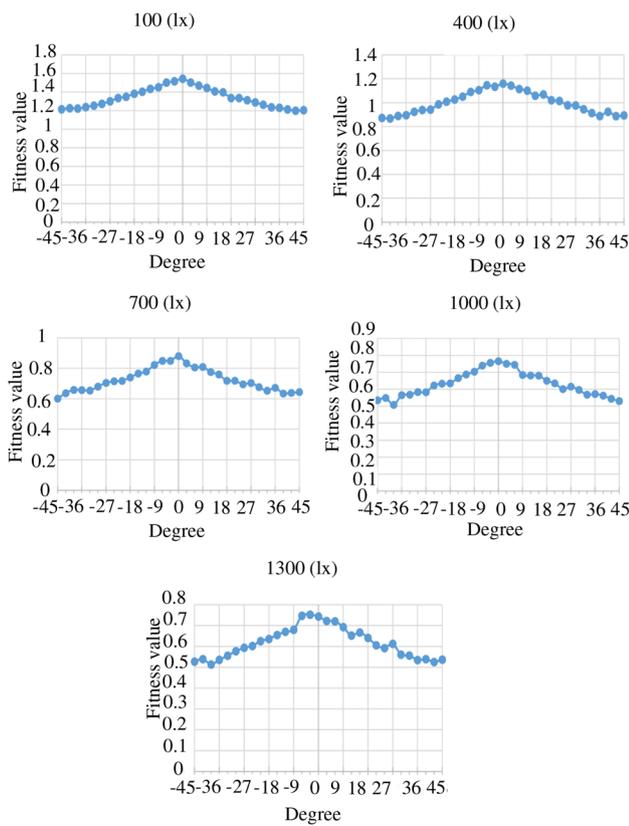


Fig. 21 Fitness function distribution of angle for cloth No.11 (fluorescent light)

confirmed experimentally that the system can recognize the different cloth under different light conditions with fitness value above 0.3 that is the threshold of the fitness value to guarantee the recognition of cloths.

The reason why the proposed cloth recognition system seems to come from that the criterion of recognition is set as the optimization result, allowing that the system to give a tolerance derived from the optimization does not matter how much the peaks of fitness functions are [16].

6 Conclusion

In this paper, cloth model generation method and pose estimation method using model-based cloth recognition are presented. Using a photo-model of each cloth, the target objects (cloths) with different colors, shapes, sizes, weights, and patterns can be recognized. The merits of the proposed system are as follows.

- Model-based approach is used to search the target in the image using a model, which is constructed based on known information of the target object.

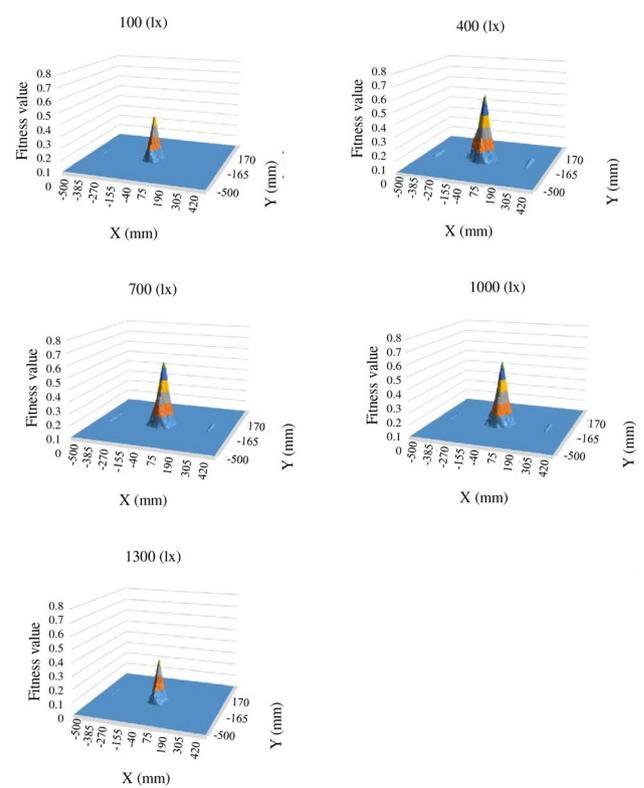


Fig. 22 Fitness function distribution of cloth No.11 in x-y plane (LED light)

- The proposed method can resist the influence of light environment changing.
- The best matching between the target object and model can get quickly and accurately without time-consuming.
- Photo-model-based method can recognize the deformable different cloths automatically.
- 3D-pose measurement is possible.
- Photo-model-based 3D-pose recognition is not limited to clothing; any object can be recognized by the proposed system.

The experiments were conducted to confirm the performance of GA recognition under fluorescent light and LED light using 12 cloths with different colorful patterns and multiple sizes. Illumination tolerance for cloth recognition was analyzed in terms of fitness function value. According to the experimental results, this system can recognize the cloths under different light environments.

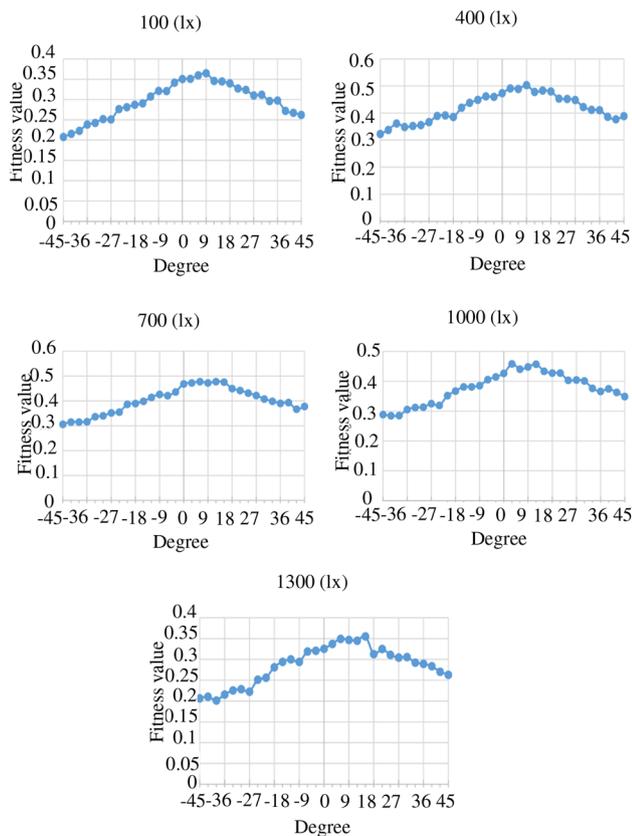


Fig. 23 Fitness function distribution of angle for cloth No.11 (LED light)

References

- Awai M, Shimizu T, Yamashita A, Kaneko T (2011) Person following and autonomous returning by mobile robot equipped with monocular camera and laser range finder. In: Proceedings of the 2011 JSME Conference on Robotics and Mechatronics, pp 2A1-H10 (in Japanese)
- Ishikawa M (1998) Sensor fusion system-mechanism for integration of sensory information. *J Robot Soc Jpn* 6(3):251–255 (in Japanese)
- Matsuyama T, Kuno Y, Imiya A (1998) Computer vision: technical review and future view. *N Technol Communication*, pp 85–95. (in Japanese)
- De Luca A, Oriolo G, Giordano PR (2007) On-line estimation of feature depth for image-based visual servoing schemes. In: Proceedings of 2007 IEEE International Conference on Robotics and Automation (ICRA2007), pp 2823–2828
- Hashimoto K, Kimura H (1995) Visual servoing-nonlinear observer approach. *J Robot Soc Jpn* 13(7):986–993
- Ono K, Ogawa T, Maeda Y, Nakatani S, Nagayasu G, Shimizu R, Ouchi N (2013) Recognition and bin-picking of coil springs by stereo vision. *Trans Jpn Soc Mech Eng Ser C* 79(804):2769–2779 (in Japanese)
- Son-Cheol Yu (2001) Tamaki Ura. Teruo Fujii, Hayato Kondo, Navigation of autonomous underwater vehicles based on artificial underwater landmarks. In: OCEANS, MTS/IEEE Conference and Exhibition, vol 1, pp 409–416
- Yinxiao L, Chih-Fan C, Peter KA (2014) Recognition of deformable object category and pose, Robotics and Automation (ICRA). In: IEEE International Conference, pp 5558–5564, Hong Kong, May 31–June 7 2014
- Funakubo R, Phyu KW, Tian H, Minami M (2016) Recognition and handling of clothes with different pattern by dual hand-eyes robotic system. In: IEEE/SICE International Symposium, pp 742–747
- Funakubo R, Phyu KW, Hagiwara R, Tian H, Minami M (2017) Verification of illumination tolerance for clothes recognition. In: The Twenty-Second International Symposium on artificial life and robotics 2017 (AROB 22nd 2017), B-Con Plaza, Beppu, Japan, January 19–21
- Phyu KW, Cui Y, Tian H, Hagiwara R, Funakubo R, Yanou A, Minami M (2016) Accuracy on photo-model-based clothes recognition. In: SICE Annual Conference, Tsukuba, Japan, September 20–23 2016
- Gee A, Cipolla R (1994) Determining the gaze of faces in images. *Image Vis Comput* 12(10):639–647
- Huang C-Y, Camps OI, Kanungo T (1997) Object recognition using appearance-based parts and relations. In: Computer vision and pattern recognition, IEEE Computer Society Conference on, pp 877–883, 1997
- Yu F, Minami M, Song W, Zhu J, Yanou A (2012) On-line head pose estimation with binocular hand-eye robot based on evolutionary model-based matching. *J Comput Inform Technol* 2(1):43–54
- Myint M, Yonemori K, Yanou A, Lwin KN, Minami M, Ishiyama S (2016) Visual Servoing for underwater vehicle using dual-eyes evolutionary real-time pose tracking. *J Robot Mechatron* 28(4):543–558
- Lwin KN, Yonemori K, Myint M, Naoki M, Minami M, Yanou A, Matsuno T (2016) Performance analyses and optimization of real-time multi-step GA for visual-servoing based underwater vehicle. In: Techno-Ocean, pp 519–526, 2016
- Minami M, Agbanhan J, Asakura T (2003) Evolutionary scene recognition and simultaneous position/orientation detection, in *Soft Computing in Measurement and Information Acquisition*. Springer, Berlin, pp 178–207