https://doi.org/10.20965/jrm.2025.p0500

Paper:

Unconstrained Home Appliance Operation by Detecting Pointing Gestures from Multiple Camera Views

Masae Yokota*, Soichiro Majima*, Sarthak Pathak, and Kazunori Umeda****

*Precision Engineering Course, Graduate School of Science and Engineering, Chuo University
1-13-27 Kasuga, Bunkyo-ku, Tokyo 112-8551, Japan
E-mail: {yokota, majima}@sensor.mech.chuo-u.ac.jp
**Department of Precision Mechanics, Faculty of Science and Engineering, Chuo University
1-13-27 Kasuga, Bunkyo-ku, Tokyo 112-8551, Japan
E-mail: {pathak, umeda}@mech.chuo-u.ac.jp
[Received June 17, 2024; accepted October 28, 2024]

In this paper, we propose a method for manipulating home appliances using arm-pointing gestures. Conventional gesture-based methods are limited to home appliances with known locations or are device specific. In the proposed method, the locations of home appliances and users can change freely. Our method uses object- and keypoint-detection algorithms to obtain the positions of the appliance and operator in real time. Pointing gestures are used to operate the device. In addition, we propose a start gesture algorithm to make the system robust against accidental gestures. We experimentally demonstrated that using the proposed method, home appliances can be operated with high accuracy and robustness, regardless of their location or the user's location in real environments.

Keywords: intelligent room, human machine interface, control system, stereo vision

1. Introduction

Daily life is besieged by a large number of home appliances, each with its own remote control. To operate a home appliance, one must locate the remote control, pick it up, and use it. As the types and numbers of home appliances increase, the workload of the users also increases. Additionally, as home appliances become more multifunctional, their control devices become increasingly complicated. For example, dynamic home appliances such as Roomba vacuum cleaners are very difficult to operate. Increasingly, operating devices with voice recognition are becoming more widespread; however, they do not allow spatial interaction. In contrast, image recognition-based interactions allow a high degree of freedom. Consequently, research has been conducted on home appliance manipulation methods using image recognition, that is, gesture-based methods.

In this study, we constructed an equipment operation system that functions based on a clear gesture pointing to an object. These gestures are not generational and have no cultural connotation. The novelty of this study is the integration of the system as a method of spatial interaction using pointing, with the location of the manipulated object and operator completely unknown.

The originality and contributions of the proposed method are as follows:

- Localization of the operation target and operator from image information without prior measurements.
- Construction of an intuitive and easy interaction method that only requires direct pointing at the object to be operated.
- Integration to detect spatial interaction using pointing from four cameras.

2. Related Work

This section describes the research on device operation using gestures, highlighting the usefulness of directly pointing to and operating home appliances with unknown location.

Considerable research has been conducted on device operations using gestures [1–14]. Gestures that constitute natural movements in daily life are often used in daily operations. These include various categories, such as eye gestures, gestures with body movements, and hand gestures.

Although many studies on gesture recognition have used contact sensors, in many scenarios of daily life, people are assumed not to wear anything special on their bodies. In [3], a robot was remotely controlled by human hand movements using only image information. When recognizing eye gestures using only images, the camera has to be in a position to reliably capture both eyes of the user. In [4], users could click and move the cursor by capturing their own eye gestures and movements from a camera installed on the monitor in front of them. However, in daily life, it is unlikely that the user will be seated in front of the camera all the time, and situations in which the user's head is arbitrarily oriented are frequent occurrences. Second, methods that operate equipment by recognizing large body

movements such as jumping and bowing impose a large physical load on the operator [5], whereas gestures that are easily recognized from any angle in images place a low burden on the user. Thus, operating devices using gestures in daily life is appropriate. Many studies on device operation using gestures have focused on hand gestures, which require relatively little effort to perform and are easy to recognize via images, regardless of camera position.

We consider hand gestures to roughly fall into two categories: (a) methods in which an operation is recognized when a hand or arm is positioned at a predetermined location and (b) methods that associate a specific hand posture with an operation in a one-to-one relationship.

Regarding the method used in (a), systems have been proposed that tie the operation to a location in 3D coordinates [6, 7]. This method allows the user's body parts to access pre-registered 3D coordinates to execute the associated operations, requiring that each operation and its location in real space be memorized. However, large individual differences exist in how easily the operation is learned because the 3D area associated with the operation to be recognized is invisible to the user.

Regarding the method used in (b), specific hand postures are related to various operations. Several studies linked specific fingertip postures to manipulation in a one-to-one relationship [8, 9]. In [8], gesture recognition was performed using the skeletal points of detected fingers. This study demonstrated that gestures can be recognized in real time in various background environments. However, the questionnaire experiment in [8] suggested that users may not be able to perform operations intuitively because the meaning of the hand gesture itself is not related to the content of the operation. Moreover, gestures are often culture-specific and are difficult for users to understand. Therefore, selecting gestures that are easy to remember and not influenced by the user's cultural or environmental background is important. Subsequently, in [9], after reducing interference from background information and recognizing static gestures, the authors improved the generalization performance using a visual language model, thereby improving the accuracy of dynamic gesture recognition. However, conveying spatial instructions with only a one-to-one relationship between operation and gesture is difficult. Therefore, users must access devices spatially using gestures.

Therefore, a pointing gesture that is universal in all cultures and situations was chosen. Among the several studies on device manipulation using pointing gestures, a vector from the hand to the hand tip was calculated using skeletal point detection [10]. The three-dimensional (3D) vector from the skeletal point of the hand to the appliance and 3D vector from the hand to the tip of the hand were calculated, whereby the appliance which the user was pointing at was determined from the angle between them. In [11] and [12], the skeletal points of the elbow and wrist were calculated, and the angle between the spatial vector from the elbow to the home appliance and that from the elbow to the wrist was used to conduct home appliance operations. These pointing gesture-based home appliance operation methods are intuitive with no alteration of meaning depending on the situation. However, in all the previous pointing-based methods, the locations of home appliances must be known and measured beforehand. As the number of home appliances increases, the burden of measuring the precise 3D position of each appliance also increases, making the system infeasible. In addition, supporting devices that are in constant motion such as robot vacuum cleaners is difficult.

In the method proposed by [13], the mobile robot moved in the direction pointed by a person standing in front of the robot with a camera mounted on it. However, this method requires the operator to be within the field of view of the camera. Thus, a method that can perform gesture recognition regardless of the positional relationship between the operator and operation target is required.

Therefore, the purpose of this study was to build a system to control home appliances by pointing gestures without knowing the position of the home appliance or user in advance.

3. Proposed Method

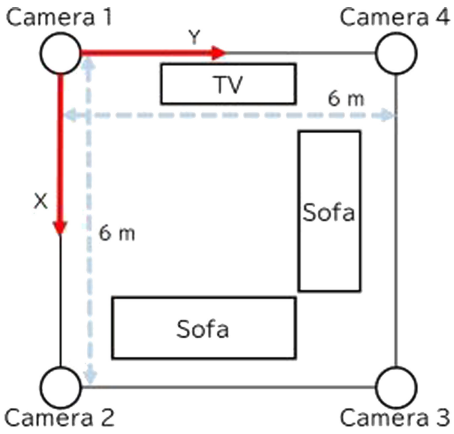
3.1. Environment

This section describes the operating environment of the system, with multiple cameras used to avoid occlusion in observing a space that resembles a living room.

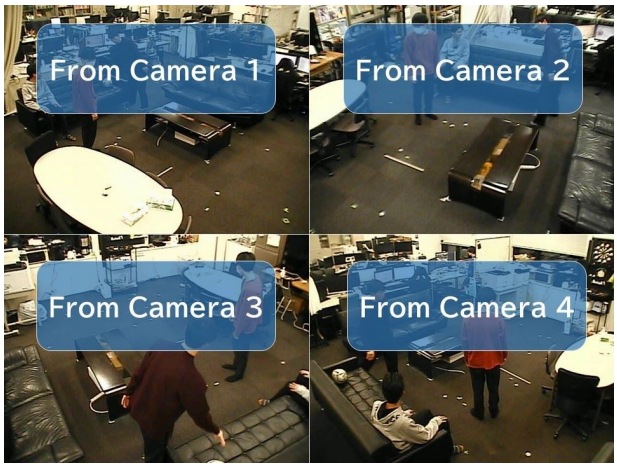
The environment in which the proposed method was applied is shown in **Fig. 1(a)**. Four charge-coupled device (CCD) cameras capable of panning, tilting, and zooming were used to acquire the images, as shown in **Fig. 2** and detailed in **Table 1**. Each camera image had a size of 640×480, generating a combined image of 1280×960, arranged as shown in **Fig. 1(b)**. The cameras were installed at the four corners of the ceiling of the target room. Using these four cameras, the 3D positions of the operator and appliances were detected. To capture images for posture recognition, a person can operate the devices from anywhere in the room using gestures. The recognition results are transmitted to each device using a smart remote control that sends infrared signals to the relevant home appliance for operation. In this study, Nature Remo 3 was used as the smart remote control.

3.2. System Overview

The proposed system allows users to initiate home appliance operations by raising their hands to operate the power source, and pointing at the home appliance with their arms, as shown in **Fig. 3(a)**. This gesture identifies the user as an "operator." A flowchart of the proposed method is shown in **Fig. 4**. YOLO [15] was used to detect appliances once every three frames and OpenPose [16] to recognize the operator's gestures once every ten frames. The 3D coordinates of the center of the appliance were calculated by triangulating the two-dimensional (2D) coordinates of the appliance obtained by YOLO using multiple cameras. OpenPose was used to extract the 2D coordinates of the operator's skeletal points, and the 3D coordinates were cal-



(a) Environmental overview



(b) Captured and combined images

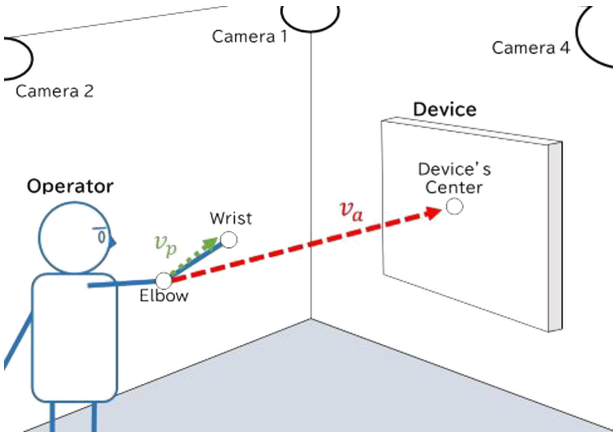
Fig. 1. Target environment of the proposed method.

Fig. 2. CCD camera used in this method: AXIS-233D.

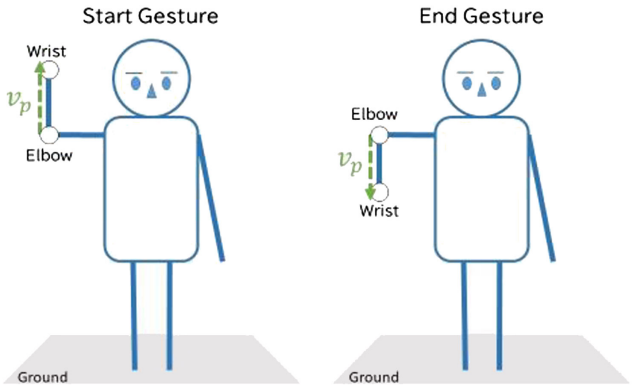
culated by triangulation. From each of the 3D coordinates obtained, a 3D vector $\vec{v}_p$ from the elbow to the wrist and a 3D vector $\vec{v}_a$ from the elbow to the home appliance's center are calculated. When the angle θ between these two spatial vectors falls below a threshold value θ_{th}, the appliance is "selected" for operation. If the selection state is held for 6 consecutive frames, the appliance is turned on/off. In order to prevent unintended home appliance operation, a gesture to initiate home appliance operation is provided as shown in **Fig. 3(b)**. This gesture is to raise the hand so that the spatial vector $\vec{v}_p$ from the elbow to the wrist is perpendicular to the ground. After the home appliance operation was complete, the system was retriggered by performing the home appliance operation start gesture.

Table 1. Properties of the camera.

Angle of view	55.8° × 43.3°
Resolution	640 × 480
Minimum illuminance	0.5 lx, 30 IRE
Shutter speed	from 1/30000 to 1/2



(a) Pointing gestures for device operation



(b) Start and end gestures for device operation

Fig. 3. Overview of the gestures used in this method.

The proposed method integrates YOLO and OpenPose to develop a pointing-based spatial interaction system and sets a variable threshold depending on the distance to the manipulation target when judging arm pointing operations. In addition, epipolar constraints were used to obtain the center points of skeletal points and home appliances through triangulation, thereby reducing the negative effects of false positives on recognition accuracy.

3.3. Detection via YOLO

Appliances are detected by YOLOv4 [17]. When an object is detected, a bounding box appears as shown in **Fig. 5**. The objects detected are a sofa, a TV, and a humidifier. In this paper, two types of home appliances were considered: TV and humidifiers. We annotated 122 photographs taken

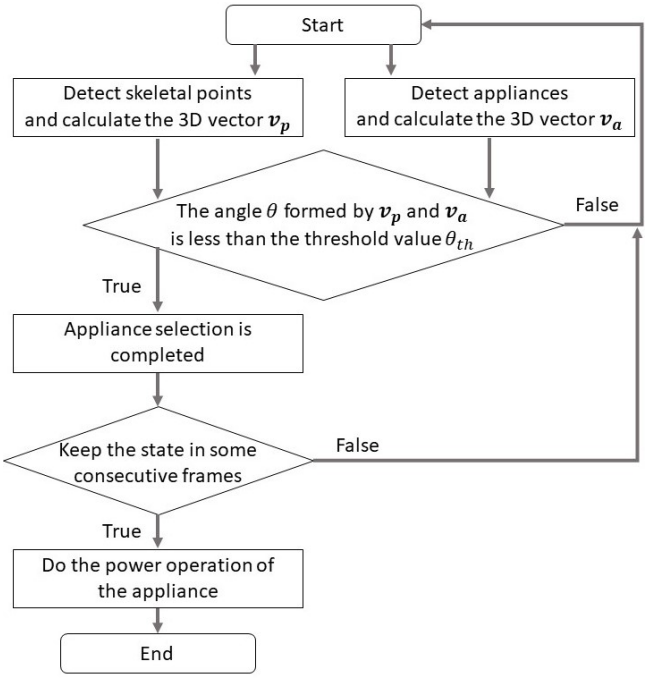


Fig. 4. Flowchart of the proposed system.

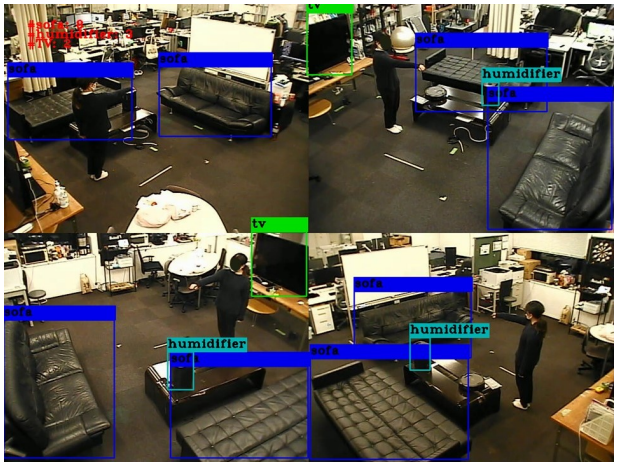


Fig. 5. Object detection results: bounding boxes.

in the experimental environment with different positions of the home appliances to create a dataset that could detect the TV and humidifier with high accuracy. The 122-pair dataset after annotation was expanded to 2928 pairs by adding noise, adjusting brightness, rotating, and other processing. Of the 2928-pair dataset, 2050 pairs were used for training and the remaining 878 for testing. Test results showed that the average compliance rates for the sofa, TV, and humidifier classes were 100%, 100%, and 99.1%, respectively.

3.4. Detection of Skeletal Points

Skeletal points were detected using OpenPose [16], and the 2D pose of a person was estimated from an image. OpenPose is a CNN-based algorithm that performs person pose estimation by cascading heatmaps and part-affinity fields. Since a confidence value was calculated for each skeletal point obtained using OpenPose, skeletal points with a confidence level of 0.60 or higher were used for gesture recognition. The detected skeletal points were shown in **Fig. 6**. Our method performs gesture recognition by observing the skeletal points of the right elbow and right wrist.

Fig. 6. Skeletal points detection results.

3.5. Recognition of Gestures

As shown in **Fig. 5**, bounding boxes of the same appliance are visible in each camera. The center point of the bounding box is considered to be the center point of the home appliance. The 3D coordinates of the center point of the home appliance are calculated from the 2D coordinates of the home appliance center point obtained from each camera by using triangulation [18]. Similarly, the 2D coordinates of the elbow and wrist skeletal points obtained from each camera were used to calculate their 3D coordinates via triangulation. Let (x_a, y_a, z_a) be the 3D coordinates of the center point of the home appliance, and (x_e, y_e, z_e) and (x_w, y_w, z_w) be the 3D coordinates of the right elbow and right wrist, respectively.

The pointing vector $\vec{v}_p$ from the right elbow to the right wrist is obtained by Eq. (1). The vector $\vec{v}_a$ from the right elbow to the center of the appliance is obtained by Eq. (2):

$$\vec{v}_p = \begin{bmatrix} x_e - x_w \\ y_e - y_w \\ z_e - z_w \end{bmatrix}, \tag{1}$$

$$\vec{v}_a = \begin{bmatrix} x_a - x_e \\ y_a - y_e \\ z_a - z_e \end{bmatrix}. \tag{2}$$

The gesture to initiate home appliance operation is shown in **Fig. 3(b)**. The appliance operation start gesture is recognized when the angle between the space vector $\vec{v}_p$ and the space vector parallel to the ground is in the range from 1.25 rad to 1.58 rad and the right wrist is above the right elbow. On the other hand, the operation end gesture is recognized when the angle between the space vector $\vec{v}_p$

and the space vector parallel to the ground is in the range from 1.25 rad to 1.58 rad and the right wrist is below the right elbow.

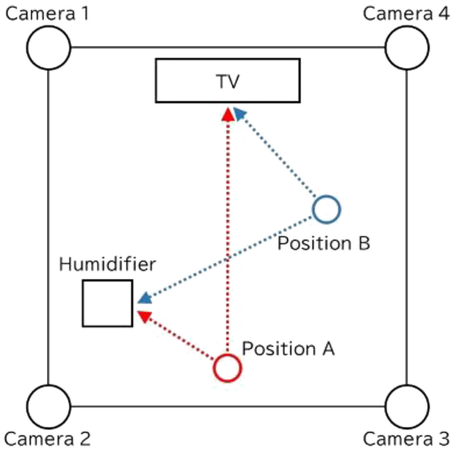
After the home appliance operation start gesture is recognized, the operator can turn on the home appliance by pointing towards its center. The angle θ between these two spatial vectors is used to determine appliance operation and is obtained by Eq. (3). The appliance selection state is activated when the angle θ between the vectors is less than or equal to a threshold value θ_{th}:

$$\theta = \arccos\left(\frac{\vec{v}_p \cdot \vec{v}_a}{|\vec{v}_p||\vec{v}_a|}\right). \tag{3}$$

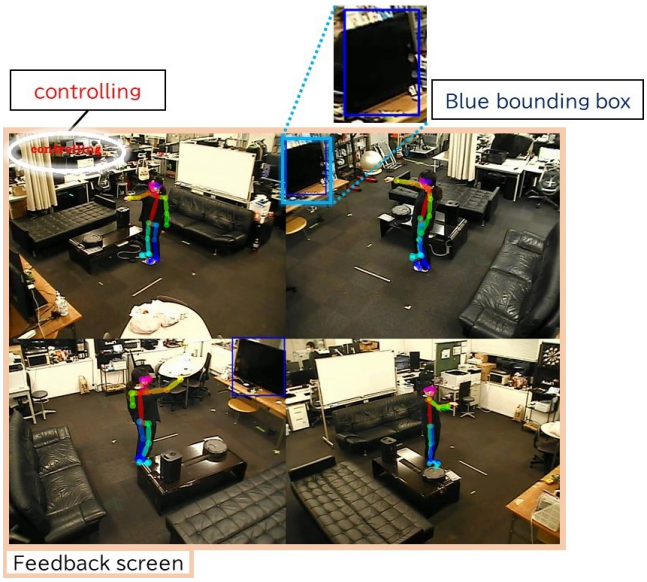
Assuming that the distance to the object being pointed at has an effect when pointing at the center of the object, we used the threshold θ_{th} to determine whether the home appliance operation selection varied depending on the 3D distance $|\vec{v}_a|$ between the center of the home appliance and right elbow of the operator. A video of pointing at the center of the home appliance was acquired from positions near and far from the center of the home appliance. The relationship between the distance from the elbow to the center of the appliance was determined along with the angle for one operator instructed to keep pointing at the center of the appliance. As a result, we found that there is a negative correlation between the 3D distance $|\vec{v}_a|$ and angle θ. Thus, we empirically obtained Eqs. (4) and (5), to determine whether the operator is pointing at the center of the TV or humidifier, respectively:

$$(TV)\theta_{th} = 0.50 - 0.10|\vec{v}_a|, \tag{4}$$

$$(Humidifier)\theta_{th} = 0.46 - 0.05|\vec{v}_a|. \tag{5}$$



(a) Experimental environment



(b) Feedback screen shown to participants

Fig. 7. Experimental situation.

4. Experimental Evaluation

4.1. Experimental Conditions

To verify the recognition rate, operational accuracy, and usability of the proposed method, three experiments were conducted in an actual environment, as shown in **Fig. 7(a)**. After a brief explanation, ten subjects were asked to participate in the experiment, and the screen shown in **Fig. 7(b)** was presented as feedback. When a gesture to start home appliance operation was recognized, the word "controlling" was displayed in red letters in the upper left corner, and when a gesture to end home appliance operation was recognized, these letters disappeared. A bounding box with blue lines around the target home appliance was displayed when a pointing operation was recognized. The operations chosen in this study involved switching the devices on or off.

Experiment 1 evaluated the accuracy of gesture recognition using skeletal points in a real environment. The subject performed the gesture of the appliance control start/end, as shown in **Fig. 3(b)**. The recognition accuracy (success rates) of the operations being recognized by the system was recorded. In Experiment 2, the subjects actually operated the TV and humidifier using pointing gestures, and the recognition accuracy (success rate) of the operations was recorded. In Experiment 3, the usability of the system was evaluated based on the subjects' responses to a questionnaire without conducting a review, as we were told by our university's Ethical Review Office that we need not apply for a review. In Experiment 4, recognition accuracy (success rate) was evaluated for random positions of appliances and people. In Experiment 5, we used a smart speaker to control a home appliance from a random position for comparison with gesture-based methods.

4.2. Experiment 1

In addition to **Fig. 3(b)**, **Table 2** summarizes the start/end gestures and combinations of operations performed.

These two types of gestures were performed ten times each at positions A and B, as shown in **Fig. 7(a)**. The gestures were performed for 5 s each, and each gesture

Table 2. Combination of the performed operation and the gesture for switching.

Operation	Gesture
Start	Raise right forearm perpendicular to the floor
End	Lower right forearm perpendicularly to the floor

Table 3. The recognition accuracy for the switching operations [%].

	Position A	Position B	Average
Start	89	73	81
End	87	70	79

Table 4. Recognition accuracy of each operation [%].

Appliance	Status	Position A	Position B	Average
TV	Standing	57	75	66
	Sitting	34	61	48
Humidifier	Standing	95	72	84
	Sitting	92	88	90

was considered successful if the operation was recognized within 5 s. The results of the gesture recognition accuracy in this experiment are shown in **Table 3**.

The recognition accuracy of the gesture for initiating the home appliance operation was 89% at position A and 73% at position B. The recognition accuracy of the gesture to terminate the home appliance operation was 87% at position A and 70% at position B. Therefore, the average recognition accuracy of the gesture to start the home appliance operation was 81% and to end the home appliance operation was 79%. The results showed that real-world gesture recognition using skeletal points can be performed with high accuracy.

4.3. Experiment 2

In Experiment 2, the subjects performed the device operation by gesturing towards the center of the appliances, thereby evaluating the recognition accuracy by which the operations are recognized. The appliances were operated by pointing 10 times in the standing and sitting positions from positions A and B, as shown in **Fig. 7(a)**. The objects to be operated were a TV and a humidifier. **Table 4** shows the experimental results.

When the user operated the TV while standing, the recognition accuracy was 57% at position A and 75% at position B. The average recognition accuracy was 66%. When the user operated the TV while seated, the recognition accuracy was 34% at position A and 61% at position B, with an average recognition accuracy of 48%. The recognition accuracy of the humidifier operation in the standing state was 95% at position A and 72% at position B, with an average recognition accuracy of 84%. In the seated position, the recognition accuracy of the humidifier operation was 92% at position A and 88% at position B, with an average recognition accuracy of 90%. The average recognition accuracy exceeded 50% for all three patterns except for the seated TV operation.

Table 5. The SUS score of this system.

A	B	C	D	E	F	G	H	I	J	Average
52.5	85	77.5	75	75	90	65	82.5	75	50	72.75

The results in **Table 4** show that the average recognition accuracies for the TV and humidifier operations were 57% and 87%, respectively.

4.4. Experiment 3

To evaluate the usability of the system of the proposed method based on the questionnaire, system usability scale (SUS) [19], to evaluate the user's subjective and overall satisfaction with the system. The questionnaire used a 5-point Likert scale [20], with 1 being "completely disagree" and 5 being "completely agree." Subjects were asked to select one of the five levels they fell into for each question. The contents of the questionnaire are as follows:

Q1. I would like to use this system often.

Q2. I thought the system was unnecessarily complicated.

Q3. I thought the system was easy to use.

Q4. I think we need the support of a technician to use this system.

Q5. I thought the various functions were well integrated.

Q6. I thought this system had many inconsistencies.

Q7. I think most people will be able to use it right away.

Q8. I found the system very difficult to use.

Q9. I felt very confident in using this system.

Q10. I had to learn a lot of things to master this system.

The following is the method used to calculate the score for each subject in the SUS:

- Subtract 1 from the response score of the odd-numbered question.
- Subtract the response score from 5 for even-numbered questions.
- Add up all the converted scores and multiply by 2.5.

The calculated SUS scores for each subject and their average scores are shown in **Table 5**. The SUS score of this system was calculated to be approximately 72.8.

4.5. Experiment 4

In this experiment, the operation recognition rate was evaluated for five subjects in a situation in which the positions of the home appliance and standing operator were random. In each trial, the subjects performed the operation

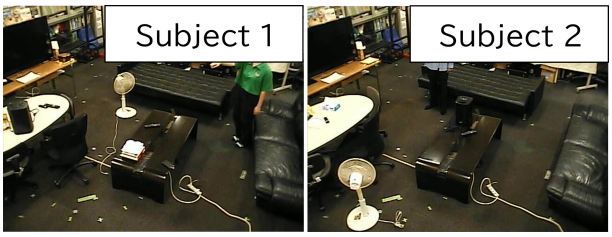


Fig. 8. Image from camera 2: changes in humidifier and fan positions for each subject.

Table 6. Recognition accuracy of each operation [%].

Appliance	Average
TV	53
Humidifier	93
Fan	93

Table 7. The recognition accuracy of each situation via a smart speaker [%].

Situation	Average
One-person situation	95
Multi-person situation	35

by pointing at the appliance from a random position. The participants operated each appliance once. Three types of home appliances were operated: TVs, humidifiers, and fans. As shown in **Fig. 8**, the positions of the humidifier and fan were changed by the subject for each run, as well as the subject's standing position:

$$(Fan)\theta_{th} = 1.05 - 0.11|\vec{v_a}|, \tag{6}$$

The operation of the electric fan was judged empirically using Eq. (6).

The experimental results are presented in **Table 6**. The average recognition accuracy exceeded 50% for all the appliances. The results in **Table 6** show that the average recognition accuracy for the TV operation was 53%, and for the humidifier and fan, 93% each.

4.6. Experiment 5

An experiment was conducted using a smart speaker with one subject. The results were compared with our gesture-based method in terms of accuracy, thereby identifying future issues with the proposed method. In the same experimental environment, a smart speaker using Google Voice Search was placed at the center of the space shown in **Fig. 1(a)** and instructed to operate the home appliance 20 times from a random position. The experiments were conducted in two different environments: one in which only one operator was present and the other in which multiple people conversed with each other.

The experimental results are shown in **Table 7**, which shows that operation instructions were recognized with 95% accuracy in a one-person environment and 35% accuracy in a multiple-person situation.

5. Discussion

The experimental results are discussed in this section based on Experiments 1, 2, and 4 in relation to the accuracy of home appliance operations based on gesture recognition. Subsequently, the usability of the system is discussed based on the results of Experiment 3. Finally, the challenges and prospects of the proposed gesture-based method are compared with experimental results using smart speakers, based on the results of Experiment 5.

5.1. Evaluation of Recognition Gestures

The results of Experiment 1 show that the recognition accuracy of gestures for switching the start/end of home appliance operations is high. This indicates that gestures are effective in preventing misoperation. However, the recognition rate at position B is lower than that at position A. This may be because the operator's head was located close to the camera at position B, and skeletal point detection did not work well because the right arm was hidden by the head.

The results of Experiment 2 confirm that the accuracy of recognizing home appliance operations by pointing the arm is greater than 50%, except in certain situations. **Table 4** shows that the recognition accuracy of the TV power operation is lower than that of the humidifier. This is attributed to the fact that the TV is larger than the humidifier, making it difficult to target the center of the TV. In fact, in the TV manipulation experiment, the subjects sometimes changed the direction in which they pointed their arms.

Moreover, the operation recognition accuracy was lower in the sitting state than in the standing state, regardless of position, only when operating the TV. This may be due to changes in eye level. As the humidifier was in a low position, it could be viewed from above while sitting or standing. Therefore, the ease of pointing did not change. However, the TV was placed at a high position. Therefore, when the subjects were standing, they could point to the center of the TV from above, whereas when seated, they pointed to the center of the TV from below, making it inaccurate. This experiment suggests that a change in posture reduces the accuracy of the operation. However, for operations on moving objects such as robots, gesture-based methods have been shown to be more intuitive than button-based methods when posture changes are considered [21]. This suggests that in the future, gesture-based operations may be useful in the operation of moving home appliances such as robot vacuum cleaners.

Experiment 4 evaluated the operational recognition accuracy when the appliances were placed at random positions. The accuracy of the TV operation was 53%, and that of the humidifier and fan operation was 93% each. The reason for the lower accuracy of the TV operation compared with the operation of other home appliances may be the same as in Experiment 2. This may be due to the fact that the area of the home appliance was visually larger than that of other home appliances, making it difficult to point to the center of the home appliance. In addition, the humidifier

and the fan were sometimes placed at high position by the subjects during the experiment and were at the same height as the TV. However, the humidifier and fan were successfully operated even in such cases, suggesting that the height of the home appliance did not affect the accuracy of the operation. Therefore, we found that two factors affect the accuracy of the operation: a change in viewpoint necessitating a change in posture and a change in the visual area of the appliance. In some cases, misidentification occurred when fans and humidifiers were located close to each other. In most cases, one of the home appliance operations was recognized more often than the other. However, when two home appliances are in close proximity, operations are required for correct identification when misrecognition occurs.

In the future, a new gesture recognition method that considers the operator's posture will be proposed so that changes in the line of sight do not affect the accuracy of operation recognition. Camera pairs should be selected to avoid occlusion not to affect the accuracy of gesture recognition. In addition to the above improvements, a function is required for determining and correcting misrecognition. Misrecognition can be identified by determining the user's posture and the situation in which the user is attempting to perform the gesture. If a discrepancy exists between the posture or scene and the gesture being recognized, an effective workaround is to not perform the operation associated with the gesture. Furthermore, the system provides feedback to the user on the discrepancy determined by the system for the user to correct the operation. This reduces the number of erroneous operations and allows the user to employ the system with confidence. Currently, the feedback screen shown in **Fig. 7(b)** is presented to the user; however, it should be more user-friendly.

In addition, only the simple operation of turning the power on and off was performed in this experiment; however, in practice, it would be desirable to operate various functions. Thus, we are considering the following two developments:

- Function selection and adjustment by pointing up or down.
- Shortcuts and detailed function operations using the command space.

Operations such as volume control can be performed by changing the pointing in the up/down direction. For TVs that display functions on the screen, pointing can be used to select the function to be executed by moving the pointer up, down, left, or right from the initial position.

Home appliances perform various functions. In remote-control operations, in many cases, a function is used by searching for and pressing the button for each function in the remote control, or by performing multiple operations. Therefore, shortcuts to frequently used functions and complex hierarchical operations would increase the practicality of gesture-based operations. The command space [7] that we have been studying for some time can be used to shortcut complex and detailed operations that cannot be handled by pointing alone. To improve the learning rate of operation techniques as operations become more complex, we are conducting research on mixed reality goggles to visualize the command space.

In addition, for future practical applications, image information from various home appliances should be added to the dataset, which would be useful for extending the method by referring to [22].

5.2. Evaluation of Usability

In [23], Sauro explained that the average SUS score is 68.1 points for data revealing the relationship between the SUS score and percentiles based on more than 5,000 SUS score measurements. The SUS score for this method was found to be 72.8 in Experiment 3. This indicates that the system exhibits excellent usability.

Currently, the only operation that can be performed using this system is to switch the device on or off. The burden on the user is expected to increase as the types of operations and the complexity of the operations increase. To make an appliance control system multifunctional while maintaining high usability requires a system that is easy to understand from the beginning and is capable of recognizing robustly and quickly to various user gestures.

The usability evaluation was conducted to ensure that the minimum usability level was met. In the future, we would like to devise methods to ensure that even if operations become more complex, they do not fall below a certain threshold value. The SUS of commercially available remote controls varies depending on the panel design. As no study exists with exactly the same problem, we plan to create a remote control that can adapt to the functionality of the device in future studies.

5.3. Challenges and Future Prospects Compared to Speech Recognition Methods

The results of Experiment 5 show that the accuracy of the smart speaker in operating home appliances was 95% in the one-operator situation and 35% in the multiple-operator situation. The highest average accuracy was 93% in Experiment 4, in which only one operator was in the room and operated the home appliance from a random position. Based on these results, the accuracy of controlling home appliances by voice using a smart speaker and using the proposed method are comparable, and both are useful. However, the accuracy of the proposed method changed significantly, depending on the operator's posture and the position of the home appliance. In the future, we aim to devise a method that can solve these problems and make the method as versatile as a smart speaker. We also aim to make the proposed method usable in multiperson situations.

6. Conclusion and Future Work

In this study, we propose a system for controlling home appliances using multiple camera views that do not require

prior information about the locations of the appliances or users. This system allows the operator to turn home appliances on and off using arm-pointing gestures. The system can also switch between the start and end of home appliance operation to prevent accidental triggering of the home appliance. The proposed method was experimentally evaluated and found to have high recognition accuracy and usability in experiments. However, the operation accuracy was found to be easily affected by changes in the operator's posture and the size of the home appliance compared to existing voice recognition-based operation methods using smart speakers.

In the future, this system should be improved to be applicable to multiperson situations. In this study, the system was constructed such that only the operator was present in the environment. However, adding a function that would distinguish the operator in a multiperson situation would make it more robust. Moreover, robots that coexist in space (such as robotic vacuum cleaners) will be expected to move to the destination pointed to by the arm and perform tasks.

Tools that allow spatial interaction improve the communication between people and intelligent devices. For example, when a person who has difficulty moving requires assistance, a simple gesture can be used to call a care robot and provide it with spatial instructions.

References:

[1] D. Coquin, E. Benoit, H. Sawada, and B. Ionescu, "Gestures recognition based on the fusion of hand positioning and arm gestures," J. Robot. Mechatron., Vol.18, No.6, pp. 751-759, 2006. https://doi.org/10.20965/jrm.2006.p0751

[2] J. Güttler, D. Bassily, C. Georgoulas, T. Linner, and T. Bock, "Unobtrusive tremor detection while gesture controlling a robotic arm," J. Robot. Mechatron., Vol.27, No.1, pp. 103-104, 2015. https://doi.org/10.20965/jrm.2015.p0103

[3] K. Hoshino, T. Kasahara, M. Tomida, and T. Tanimoto, "Gesture-world environment technology for mobile manipulation – Remote control system of a robot with hand pose estimation –," J. Robot. Mechatron., Vol.24, No.1, pp. 180-190, 2012. https://doi.org/10.20965/jrm.2012.p0180

[4] A. Dongre, R. Pinto, A. Patkar, and M. Lopes, "Computer cursor control using eye and face gestures," 2020 11th Int. Conf. on Computing, Communication and Networking Technologies (ICCCNT), 2020. https://doi.org/10.1109/ICCCNT49239.2020.9225311

[5] F. Deboeverie, S. Roegiers, G. Allebosch, P. Veelaert, and W. Philips, "Human gesture classification by brute-force machine learning for exergaming in physiotherapy," 2016 IEEE Conf. on Computational Intelligence and Games (CIG), 2016. https://doi.org/10.1109/CIG.2016.7860414

[6] M. Niitsuma, H. Hashimoto, and H. Hashimoto, "Spatial memory as an aid system for human activity in intelligent space," IEEE Trans. on Industrial Electronics, Vol.54, No.2, pp. 1122-1131, 2007. https://doi.org/10.1109/TIE.2007.892730

[7] S. Yan, Y. Ji, and K. Umeda, "A system for operating home appliances with hand positioning in a user-definable command space," 2020 IEEE/SICE Int. Symp. on System Integration (SII), pp. 366-370, 2020. https://doi.org/10.1109/SII46433.2020.9025978

[8] Y. Muranaka, M. Al-Sada, and T. Nakajima, "A home appliance control system with hand gesture based on pose estimation," 2020 IEEE 9th Global Conf. on Consumer Electronics (GCCE), pp. 752-755, 2020. https://doi.org/10.1109/GCCE50665.2020.9291877

[9] S. Wan, L. Yang, K. Ding, and D. Qiu, "Dynamic gesture recognition based on three-stream coordinate attention network and knowledge distillation," IEEE Access, Vol.11, pp. 50547-50559, 2023. https://doi.org/10.1109/ACCESS.2023.3278100

[10] A. I. D. Viaje et al., "Selection of appliance using skeletal tracking of hand to hand-tip for a gesture controlled home automation," 2020 Int. Conf. on Electronics and Sustainable Communication Systems (ICESC), pp. 575-580, 2020. https://doi.org/10.1109/ICESC48915.2020.9155860

[11] A. Fernández, L. Bergesio, A. M. Bernardos, J. A. Besada, and J. R. Casar, "A Kinect-based system to enable interaction by pointing in smart spaces," 2015 IEEE Sensors Applications Symp. (SAS), 2015. https://doi.org/10.1109/SAS.2015.7133613

[12] M. A. Iqbal, S. K. Asrafuzzaman, M. M. Arifin, and S. K. A. Hossain, "Smart home appliance control system for physically disabled people using Kinect and X10," 2016 5th Int. Conf. on Informatics, Electronics and Vision (ICIEV), pp. 891-896, 2016. https://doi.org/10.1109/ICIEV.2016.7760129

[13] T. Ikeda, N. Noda, S. Ueki, and H. Yamada, "Gesture interface and transfer method for AMR by using recognition of pointing direction and object recognition," J. Robot. Mechatron., Vol.35, No.2, pp. 288-297, 2023. https://doi.org/10.20965/jrm.2023.p0288

[14] Y. Tamura, M. Sugi, T. Arai, and J. Ota, "Target identification through human pointing gesture based on human-adaptive approach," J. Robot. Mechatron., Vol.20, No.4, pp. 515-525, 2008. https://doi.org/10.20965/jrm.2008.p0515

[15] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You only look once: Unified, real-time object detection," 2016 IEEE Conf. on Computer Vision and Pattern Recognition (CVPR), pp. 779-788, 2016. https://doi.org/10.1109/CVPR.2016.91

[16] Z. Cao, G. Hidalgo, T. Simon, S.-E. Wei, and Y. Sheikh, "OpenPose: Realtime multi-person 2D pose estimation using part affinity fields," IEEE Trans. on Pattern Analysis and Machine Intelligence, Vol.43, No.1, pp. 172-186, 2021. https://doi.org/10.1109/TPAMI.2019.2929257

[17] A. Bochkovskiy, C.-Y. Wang, and H.-Y. M. Liao, "YOLOv4: Optimal speed and accuracy of object detection," arXiv:2004.10934, 2004. https://doi.org/10.48550/arXiv.2004.10934

[18] R. Hartley and A. Zisserman, "Multiple View Geometry in Computer Vision," 2nd Edition, Cambridge University Press, 2004. https://doi.org/10.1017/CBO9780511811685

[19] J. Brooke, "SUS: A 'quick and dirty' usability scale," P. W. Jordan, B. Thomas, I. L. McClelland, and B. Weerdmeester (Eds.), "Usability Evaluation in Industry," pp. 189-194, CRC Press, 1996.

[20] R. Likert, "A technique for the measurement of attitudes," Archives of Psychology, No.140, 1932.

[21] Q. Yao, T. Terakawa, M. Komori, H. Fujita, and I. Yasuda, "Effect of viewpoint change on robot hand operation by gesture- and button-based methods," J. Robot. Mechatron., Vol.34, No.6, pp. 1411-1423, 2022. https://doi.org/10.20965/jrm.2022.p1411

[22] Y. Ishida and H. Tamukoh, "Semi-automatic dataset generation for object detection and recognition and its evaluation on domestic service robots," J. Robot. Mechatron., Vol.32, No.1, pp. 245-253, 2020. https://doi.org/10.20965/jrm.2020.p0245

[23] J. Sauro, "A Practical Guide to the System Usability Scale: Background, Benchmarks & Best Practices," Measuring Usability LLC, 2011.

Name:
Masae Yokota

Affiliation:
Precision Engineering Course, Graduate School of Science and Engineering, Chuo University

Address:
1-13-27 Kasuga, Bunkyo-ku, Tokyo 112-8551, Japan

Brief Biographical History:
2023 Received B.Eng. degree in Precision Mechanics from Chuo University

Membership in Academic Societies:
- The Japan Society of Mechanical Engineers (JSME)
- Institute of Electrical and Electronics Engineers (IEEE)

Name:
Soichiro Majima

Affiliation:
Hitachi Global Life Solutions, Inc.

Address:
1-13-27 Kasuga, Bunkyo-ku, Tokyo 112-8551, Japan

Brief Biographical History:
2023 Received M.Eng. degree in Precision Engineering from Chuo University
2023- Joined Hitachi Global Life Solutions, Inc.

Name:
Sarthak Pathak

ORCID:
0000-0002-5271-1782

Affiliation:
Assistant Professor, Department of Precision Mechanics, Chuo University

Address:
1-13-27 Kasuga, Bunkyo-ku, Tokyo 112-8551, Japan

Brief Biographical History:
2014 Received Bachelor of Technology and Master of Technology degrees from Department of Engineering Design, Indian Institute of Technology Madras
2017 Received Ph.D. degree from Department of Precision Engineering, The University of Tokyo
2017- Postdoctoral Researcher, Department of Precision Engineering, The University of Tokyo
2018- JSPS Postdoctoral Research Fellow, Department of Precision Engineering, The University of Tokyo
2020- Project Assistant Professor, Department of Precision Engineering, The University of Tokyo
2021- Assistant Professor, Department of Precision Mechanics, Chuo University

Main Works:
• "Spherical video stabilization by estimating rotation from dense optical flow fields," J. Robot. Mechatron., Vol.29, No.3, pp. 566-579, 2017.
• "Self-supervised optical flow derotation network for rotation estimation of a spherical camera," Advanced Robotics, Vol.35, No.2, pp. 118-128, 2021.
• "A decoupled virtual camera using spherical optical flow," 2016 IEEE Int. Conf. on Image Processing (ICIP), pp. 4488-4492, 2016.

Membership in Academic Societies:
• Institute of Electrical and Electronic Engineers (IEEE)
• The Japan Society for Precision Engineering (JSPE)
• The Robotics Society of Japan (RSJ)

Name:
Kazunori Umeda

ORCID:
0000-0002-4458-4648

Affiliation:
Professor, Department of Precision Mechanics, Chuo University

Address:
1-13-27 Kasuga, Bunkyo-ku, Tokyo 112-8551, Japan

Brief Biographical History:
1994 Received Ph.D. degree in Precision Machinery Engineering from The University of Tokyo
1994- Lecturer, Chuo University
2003-2004 Visiting Worker, National Research Council of Canada

Main Works:
• "Correction of color information of a 3D model using a range intensity image," Computer Vision and Image Understanding, Vol.113, No.11, pp. 1170-1179, 2009.
• "Construction of a compact range image sensor using multi-slit laser projector and obstacle detection of a humanoid with the sensor," 2010 IEEE/RSJ Int. Conf. on Intelligent Robots and Systems, pp. 5972-5977, 2010.
• "PointpartNet: 3D point-cloud registration via deep part-based feature extraction," Advanced Robotics, Vol.36, No.15, pp. 724-734, 2022.

Membership in Academic Societies:
• The Robotics Society of Japan (RSJ)
• The Japan Society for Precision Engineering (JSPE)
• The Japan Society of Mechanical Engineers (JSME)
• The Society of Instrument and Control Engineers (SICE)
• The Institute of Electronics, Information and Communication Engineers (IEICE)
• Institute of Electrical and Electronics Engineers (IEEE)