https://doi.org/10.20965/ijat.2025.p0216

Research Paper:

Home Appliance Operation via 3D Keypoint Based Gesture Detection in Body-Relative Command Spaces

Masae Yokota*,†, Soichiro Majima*, Yushin Mochizuki*, 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
†Corresponding author, E-mail: yokota@sensor.mech.chuo-u.ac.jp
**Department of Precision Mechanics, Faculty of Science and Engineering, Chuo University, Tokyo, Japan

[Received December 7, 2024; accepted February 10, 2025]

In this paper, we propose a flexible device control method using personalized command spaces that function as buttons on a virtual remote control that follows the user. By performing two different gestures in each space, the users can control various devices in a room. This system is implemented through multiple cameras and 3D human keypoint tracking. We experimentally evaluated the influence of command spaces arrangement on gesture recognition and determined the recognition accuracies for different gestures in each command space. The system demonstrated high usability, with even inexperienced users achieving high gesture recognition accuracy.

Keywords: intelligent room, gesture recognition, stereo vision

1. Introduction

Daily life is beset by an increasing number of devices and appliances, each with its own remote control and operating method. Even in industrial environments, such as factories, workers may be in charge of multiple devices that need to be controlled and managed simultaneously. Such situations can increase the cognitive burden on users who are required to keep track of multiple remote controls and operation methods. A more intuitive and, deviceless operational method has the potential to reduce the complexity of operations. This study proposes the construction of an "intelligent room" environment that aims to simplify device operation using intuitive gestures without the use of additional devices. Our room consists of a space in which multiple cameras are fixed to enable 3D perception via triangulation. We build a virtual remote control with 3D buttons called "command spaces" that can be activated via simple hand gestures. By tracking the users' keypoints in 3D, these command spaces can follow the user enabling device-free operation from anywhere in the intelligent room. By using two different gestures and six command spaces, 12 different operations can be performed, in contrast to previous methods that allowed only one or two operations. In this paper, we describe the design and system integration involved in the construction of this Intelligent Room and command spaces.

2. Related Works

The simplest way to make a room "intelligent" and enable control of multiple devices is to use smart speakers [1] such as Alexa and Google Nest. By connecting them to the target devices, operation can be performed via voice commands. However, this approach has several disadvantages. For instance, it cannot be used in noisy environments or by multiple people in the same environment. Voice recognition also has a limited operational range, making this approach spatially restricted.

In contrast, some systems use sensors to detect visual or motion-based gestures, enabling flexible and intuitive device operations. For example, in [2], eye-gaze direction is used to control electronic devices. However, such systems require a clear line of sight to the user's eyes. Moreover, eye-gaze direction estimation is sometimes inaccurate and can lead to accidental operations. In [3], gestures are calculated using two cameras and forearm inclination correction. However, high-accuracy hand-pose recognition requires multiple camera images in which the system can be observed clearly.

In comparison, the systems presented in [4, 5] employ cameras to detect hand-waving gestures through skin color segmentation. Such gesture-based methods are superior to eye-gaze detection. However, their reliance on color information results in poor performance under varying lighting conditions, even when the HSV color space is used.

Instead of using arbitrary cues such as eye-gaze direction or skin color, it is now possible to accurately detect human keypoints using deep learning methods such as those described in [6]. The authors in [7] took advantage of this and triangulated 3D human keypoints. They implemented a pointing-based appliance control system. However, the drawback of this system is that only one operation can be performed because this method can only "select" an appli-

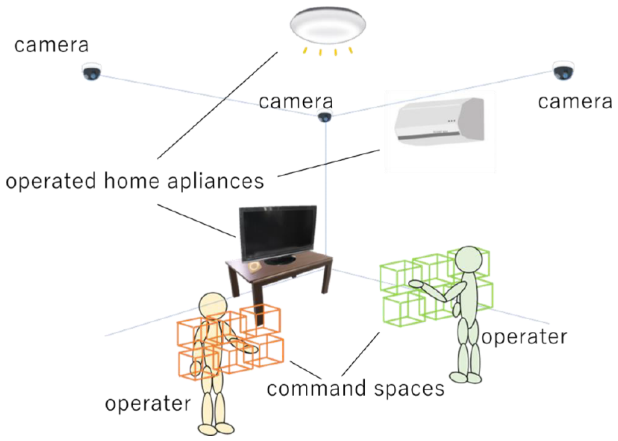


Fig. 1. Conceptual diagram of the system.

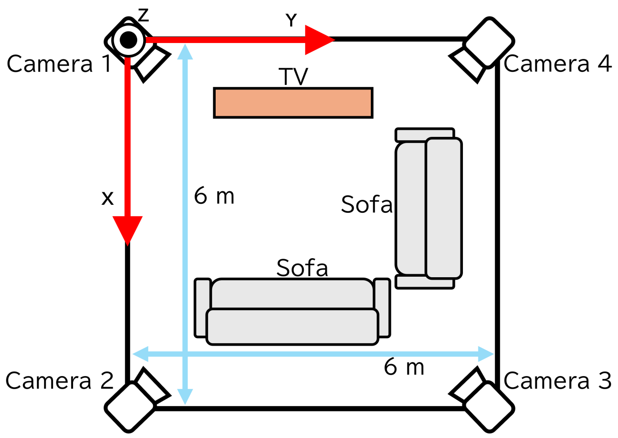


Fig. 2. Bird's eye view of the environment.

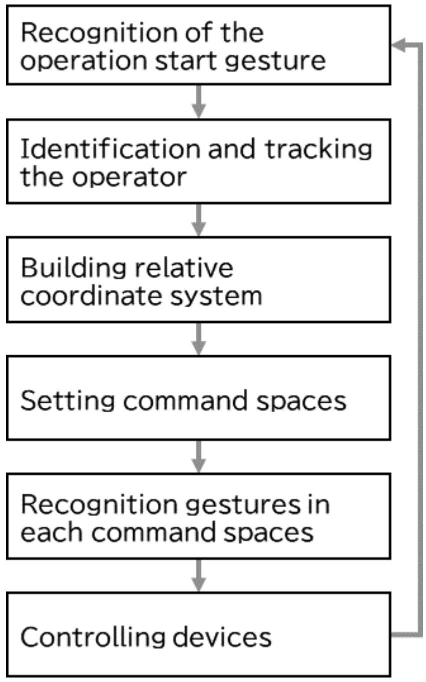


Fig. 3. Outline of the proposed system.

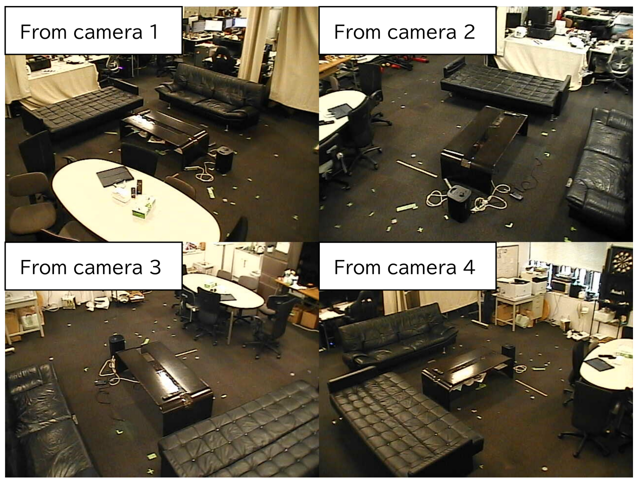


Fig. 4. Captured images.

ance. Even simple appliances, such as a TV, have multiple functions and require a more flexible approach.

In this study, we implement a new method inspired by the work of [8, 9], where "command spaces," which function as buttons on a virtual remote control, are used to enable multiple commands for device operation. By defining a person-based coordinate system and tracking users' 3D keypoints, our proposed method permits the command spaces to follow the user as required. This also allows simultaneous operation by multiple users, as each user can have their own control method.

3. Methods

3.1. Overview of the Entire Proposed Method

In our system, the operation is performed using two simple gestures, in dedicated command spaces consisting of six cubes positioned around the operator based on a relative coordinate system, as shown in **Fig. 1**. The operator can perform 12 different operations by executing either of the two gestures in each command space. The environment for the proposed method is illustrated in **Fig. 2**. Cameras are installed in the four corners of the ceiling in a space with multiple home appliances and furniture to prevent the deterioration of recognition accuracy due to occlusion. When the epipolar line error is minimal between the first and subsequent camera detections of a skeletal, those cameras are considered a valid pair for triangulation.

The system outline is depicted in **Fig. 3**. First, the operator is identified and tracked after a "start gesture" is recognized. A relative coordinate system is constructed using the operator's skeletal points. Six command spaces are set based on this relative coordinate system, fixed to the user's body. Finally, a command linked to a particular gesture and command space is executed. In this study, it is assumed that the operator is standing still while performing the gesture. Operator 3D keypoint detection and gesture recognition are performed using images captured from cameras installed on the ceiling of a space, as shown in **Fig. 4**.

Fig. 5. Detection of skeletal points.

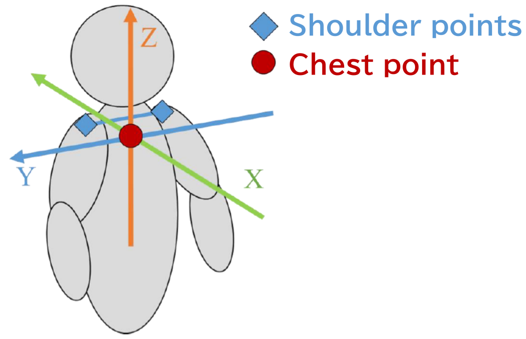


Fig. 6. Body-relative coordinates.

Table 1. 3D coordinates of the shoulders and chest.

Skeletal point	3D coordinate
Left shoulder	(x_1, y_1, z_1)
Right shoulder	(x_2, y_2, z_2)
Chest	(x_c, y_c, z_c)

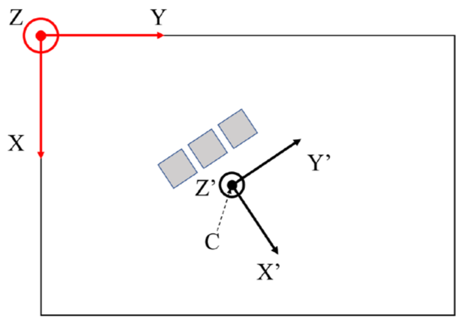


Fig. 7. Relationship between the two coordinate systems.

3.2. Identification of the Operator

In this study, skeletal point information is obtained from four camera images using OpenPose [6], as shown in **Fig. 5**.

The operator is identified when their elbow is raised higher than their chest. First, using the images acquired from the four cameras, candidate 2D coordinates of skeletal points at the elbow and center of the chest of each person are obtained. Next, using the 2D coordinates of skeletal point candidates that satisfy the epipolar constraint, the 3D coordinates of the elbow (x_e, y_e, z_e) and chest (x_c, y_c, z_c) are calculated using the principles of stereo vision [10]. When $z_e > z_c$ is maintained for several frames, the system recognizes the gesture as the start of operation, and identifies the person as the operator. This ensures that operation is performed intentionally by the person who executed the start gesture, and not accidentally.

3.3. Construction of Body Relative Coordinate System

3.3.1. Overview

The body-relative coordinate system is a coordinate system constructed relative to the operator's body. This coordinate system is defined with the operator's body as its origin. There are two advantages to establishing this coordinate system:

- Operable at any time, regardless of the operator's position.
- Easy to understand the location of the command space using the operator's body as a reference point, even though it is invisible.

3.3.2. Construction of the Body Relative Coordinate System

The body-relative coordinate system is defined using the 3D coordinates of three skeletal points obtainable from OpenPose: the chest, and the left and right shoulders. As illustrated in **Fig. 6**, the chest serves as the origin of the coordinate system. First, the y-axis is set as a straight line extending from the left shoulder to the right shoulder, translating it through the origin. The z-axis is then aligned with the direction of the world coordinate system, and the x-axis is defined as the cross product of the y- and z-axes. To ensure the y-axis remains horizontal, the difference in z-coordinate values between the left and right shoulders is disregarded.

The 3D coordinates of the left shoulder, right shoulder, and chest are listed in **Table 1**. To facilitate understanding of the command space location, the y-axis of the relative coordinate system is adjusted to be parallel to the xy-plane of the world coordinate system. Therefore, the values of $z = 1$ and z_2 should be equal to z_c. The relationship between the world coordinate system (x, y, z) and relative coordinate system (x', y', z') is depicted in **Fig. 7**, and defined by Eqs. (1) and (2).

$$\begin{pmatrix} x' \\ y' \\ z' \end{pmatrix} = \mathbf{R} \begin{pmatrix} x \\ y \\ z \end{pmatrix} + \begin{pmatrix} x_c \\ y_c \\ z_c \end{pmatrix} \tag{1}$$

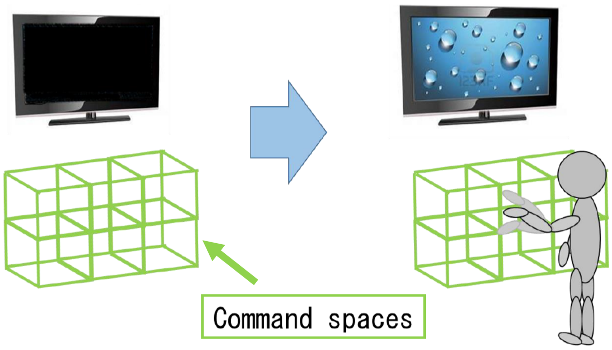


Fig. 8. Conceptual diagram of command spaces.

$$\boldsymbol{R} = \begin{pmatrix} \frac{y_2 - y_1}{(x_2 - x_1)^2 + (y_2 - y_1)^2} & \frac{x_1 - x_2}{(x_2 - x_1)^2 + (y_2 - y_1)^2} & 0 \\ \frac{x_2 - x_1}{(x_2 - x_1)^2 + (y_2 - y_1)^2} & \frac{y_2 - y_1}{(x_2 - x_1)^2 + (y_2 - y_1)^2} & 0 \\ 0 & 0 & 1 \end{pmatrix} \tag{2}$$

3.4. Gesture Recognition in Command Spaces

3.4.1. Overview of Gesture Recognition Method

This system employs command spaces, as displayed in **Fig. 8**, in which commands for equipment operation are linked to specific real-world spaces. When an operator performs a gesture within a command space, the equipment operation associated with that command space and gesture combination is executed. The system recognizes two gestures types, distinguished by elbow and wrist heights. Thus, a total of 12 possible commands can be performed, opposed to only one in [7].

3.4.2. Construction of Command Spaces

The command space is set up on the body-relative coordinate system constructed relative to the operator. If the gesture to initiate operation is performed with the right hand, the command space is configured for right-hand operation; if the gesture is performed with the left hand, the command space is configured for left-hand operation. This allows for a natural posture when performing gestures in the command space on the side opposite the operating hand.

3.4.3. Gesture Recognition

This study uses gestures types, as shown in **Fig. 9**, to enable two distinct operations per command space. Both gestures involve holding the hand within the command space; however, the height difference between the wrist and elbow is significant for gesture 1 and minimal for gesture 2.

The 3D coordinates of the skeletal points of the wrist and elbow are calculated as displayed in **Table 2**. When the 3D coordinates of the wrist are within a command space, a gesture is recognized as being performed. Gesture 1 is recognized when the condition defined by the first equation of Eq. (3) is maintained across several consecutive frames. Similarly, gesture 2 is recognized when the condition defined by the second equation of Eq. (3) is sustained across several consecutive frames. When the expression is the third equation of Eq. (3), no gesture is recognized.

If a gesture is not recognized, but the same gesture is recognized in both the preceding and following frames, that gesture is inferred for the intervening frame.

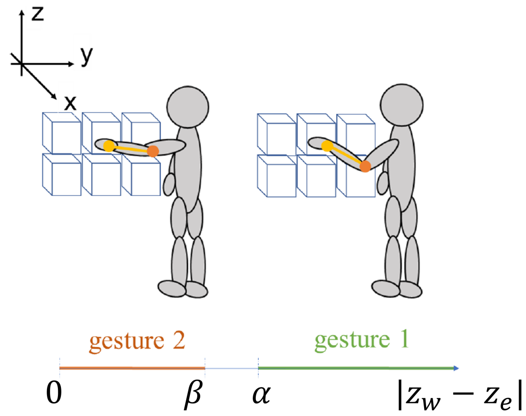


Fig. 9. Definition of two types of gestures.

Table 2. 3D coordinates of the elbow and wrist.

Skeletal point	3D coordinate
Elbow	(x_e, y_e, z_e)
Wrist	(x_w, y_w, z_w)

$$\begin{cases} |z_e - z_w| > \alpha & \text{Gesture 1} \\ |z_e - z_w| < \beta & \text{Gesture 2} \\ \beta \le |z_e - z_w| \le \alpha & \text{Unrecognized} \end{cases} \tag{3}$$

4. Evaluation

4.1. Overview of Evaluation

Experiments were conducted to evaluate the recognition accuracy of the proposed system. Two types of experiments were performed. In each experiment, both subjects familiar with the proposed method and those unfamiliar with it participated, resulting in four categories of experimental results. In each experiment, subjects were presented with a feedback screen generated using OpenGL, as illustrated in **Fig. 10**. A color-coded command space, dynamically-adjusted to the operator's position, was continuously displayed. The result of each operation was indicated by a change in the color of a large rectangle in front of the screen. The operator's wrist position was represented by a blue circle, allowing for continuous objective monitoring of their wrist position. To make the positional relationship easier to understand for the operator to understand, the sofa positions were represented by a combination of black rectangles. These positions were fixed.

To evaluate the influence of command space placement on recognition accuracy and ease of use in Experiment 1,

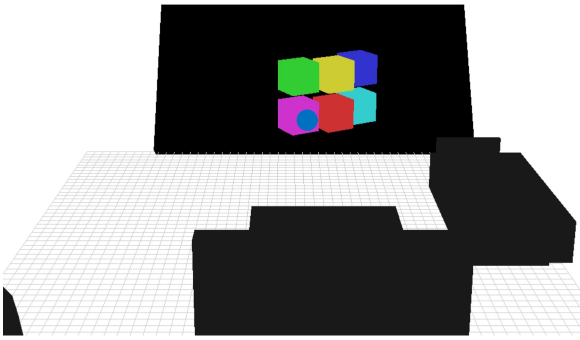


Fig. 10. Feedback screen.

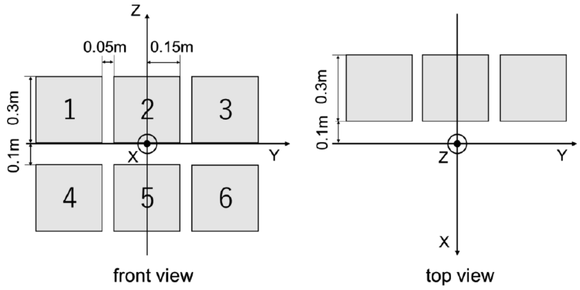
a comparative study was conducted using the three placements shown in **Fig. 11**. In **Fig. 11(a)**, command space is positioned directly in front of and centered on the operator. In **Figs. 11(b)** and **(c)**, the command space is placed in front of the operator but offset slightly toward their dominant hand. Specifically, the command space is shifted to the right for right-handed operators, and to the left for left-handed operators. **Figs. 11(b)** and **(c)** both show the right-handed case. While the command space is planar in **Figs. 11(a)** and **(b)**, it is curved to approximate the trajectory of the operator's arm in **Fig. 11(c)**.

In Experiment 2, we verified the gesture recognition accuracy using arrangement 3, which yielded the best results in Experiment 1. The two gesture types defined in Section 3.4.3 were used, and the recognition accuracy was calculated for each command space.

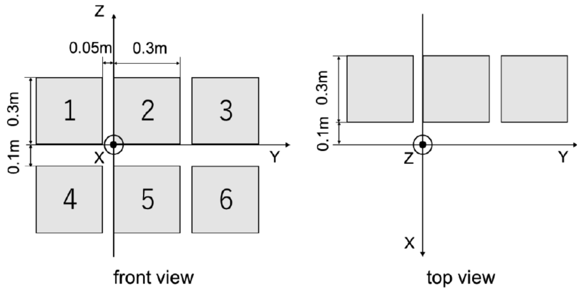
In both Experiments 1 and 2, the operation sequence was consistent. A clockwise operation was performed using gesture 1 in each command space. After all operations were completed with gesture 1 across all command spaces, the process was repeated using gesture 2, following the same clockwise order. The experimental results were evaluated based on the following two components:

- Probability that the wrist position is recognized as being within the correct command space: position recognition rate.
- Probability that the two gestures are correctly recognized: gesture recognition rate.

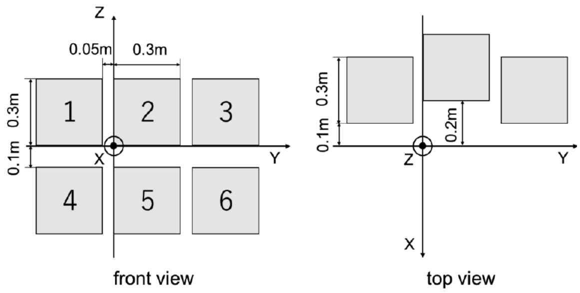
Verbal instructions were provided to the subjects for each operation. If the correct operation was not performed within 5 s, the gesture recognition was considered a failure. Each experiment began with the subjects performing a gesture to start the operation, after which the experiment proceeded. The feedback screen was continuously displayed, allowing subjects to refer to it when necessary. Subjects mainly stood in front of the TV set, but their precise position was not specified. Two or three people, including the subject, were present during the experiments, while the non-subjects were seated.



(a) Arrangement 1: alignment with the front and middle of the body



(b) Arrangement 2: displaced toward the dominant hand



(c) Arrangement 3: arc-shaped and displaced in the direction of the dominant hand

Fig. 11. Three command space arrangements.

4.2. Experiment 1: Comparative Experiment of Gesture Recognition Accuracy with Different Command Space Arrangements

Experiment 1 was conducted to compare the effects of different command space placements on position recognition accuracy. Three command space arrangements, as depicted in **Fig. 11**, were evaluated. No distinction was made between gestures, and operators simply extended their arms and positioned their wrists within the target command space. The success rate was then calculated.

4.2.1. Experiment 1 with an Experienced Subject

A subject, who was familiar with the operation performed the start gesture with their right hand and held then positioned their hand 10 times within each command space. **Table 3** presents the success rates for the three command space arrangements configured in three different ways, as illustrated in **Fig. 11**.

The average position recognition rate was 73% for arrangement 1 , 97% for arrangement 2, and 99% for arrangement 3.

Table 3. Position recognition rate for Experiment 1 with experienced subjects [%].

Command	1	2	3	4	5	6	Average
Arrangement 1	30	100	90	20	100	100	73
Arrangement 2	90	100	100	100	100	90	97
Arrangement 3	100	98	100	95	100	100	99

Table 4. Position recognition rate for Experiment 1 with inexperienced subjects [%].

Command	1	2	3	4	5	6	Average
Arrangement 1	16	84	96	8	72	92	61
Arrangement 2	88	88	96	60	88	96	86
Arrangement 3	68	92	92	84	96	88	87

4.2.2. Experiment 1 with Inexperienced Subjects

This experiment was conducted with five subjects unfamiliar with command space operations. Each operator, remaining in a fixed standing position, placed their wrists within the targeted command space. The success rates are listed in **Table 4**.

The average position recognition rate was 61% for arrangement 1, 86% for arrangement 2, and 87% for arrangement 3.

4.3. Experiment 2: Evaluation of Gesture Recognition Accuracy

Experiment 2 was conducted to evaluate the recognition rates for the two gesture types within each command space. Each gesture was performed 30 times per command space, and the recognition rate was evaluated. This experiment utilized the command space of the arrangement displayed in **Fig. 11(c)**.

4.3.1. Experiment 2 with an Experienced Subject

This experiment was conducted by an operator person experienced with the system. The operator performed gestures within each command space at the same position and the recognition accuracy was determined, as described in Section 4.1.

The results are summarized in **Table 5**. Gesture recognition took approximately 1–4 s. On average, the position recognition rates were 98% for gesture 1 and 99% for gesture 2. The average recognition rates for gestures 1 and 2 were 94% and 96%, respectively.

4.3.2. Experiment 2 with Inexperienced Subjects

This experiment involved five participants who were inexperienced with this method of operation. Each participant performed two types of gestures, five times each, within every command space at the same position. The recognition rate was then described in Section 4.1.

The results are summarized in **Table 6**. On average, the position recognition rates were 86% and 87% for gestures 1 and 2. The average gesture recognition rates for gestures 1 and 2 were 73% and 60%, respectively.

Table 5. Recognition rates for Experiment 1 [%].

	Command	1	2	3	4	5	6	Average
Gesture 1	Gesture recognition	100	100	90	87	93	93	94
	Position recognition	100	100	100	90	100	100	98
Gesture 2	Gesture recognition	90	97	97	93	100	100	96
	Position recognition	100	97	100	100	100	100	99
Average	Gesture recognition	95	99	94	90	97	97	95
	Position recognition	100	99	100	95	100	100	99

Table 6. Recognition rates for Experiment 2 with inexperienced subjects [%].

	Command	1	2	3	4	5	6	Average
Gesture 1	Gesture recognition	80	88	92	44	64	68	73
	Position recognition	80	88	96	72	92	88	86
Gesture 2	Gesture recognition	36	60	52	56	80	76	60
	Position recognition	74	88	96	76	96	92	87
Average	Gesture recognition	58	74	72	50	72	72	66
	Position recognition	77	88	96	74	94	90	87

5. Discussion

This section discusses the results of the experiments. Based on the outcomes of Experiment 1, we analyzed the impact of command space layout on ease of operation. Then, we used the outcomes of Experiment 2 to examine the recognition accuracy of each gesture within each command space.

5.1. Effects of Command Space Placement

In Experiment 1, we calculated the recognition rate of subjects' wrist skeletal points within the command space for three different command space arrangements, and the results are presented in **Tables 3** and **4**. The average position recognition rate for placement 1 is 73% for experienced operators and 61% for inexperienced operators. The average position recognition rate for placement 2 is 97% for experienced operators and 86% for inexperienced operators. The average position recognition rate for placement 3 is 99% for experienced operators and 87% for inexperienced operators.

The location recognition rates for the experienced operator show that arrangement 1 yields the lowest values, whereas arrangements 2 and 3 are nearly identical. This operator is right-handed. The position recognition rate for placement 1 is notably low for commands 1 and 4, which

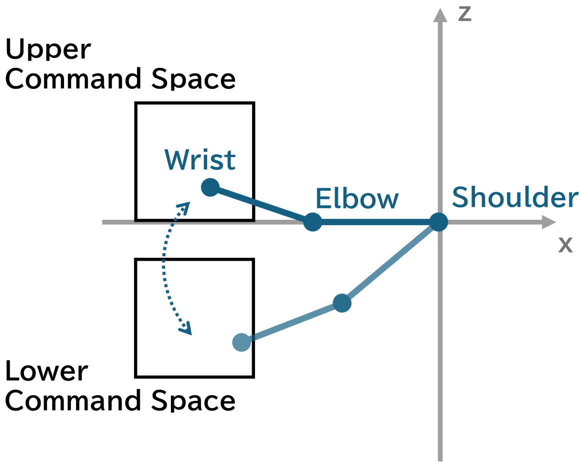
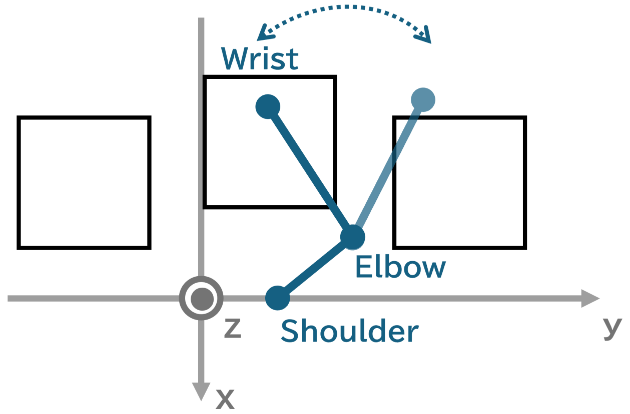
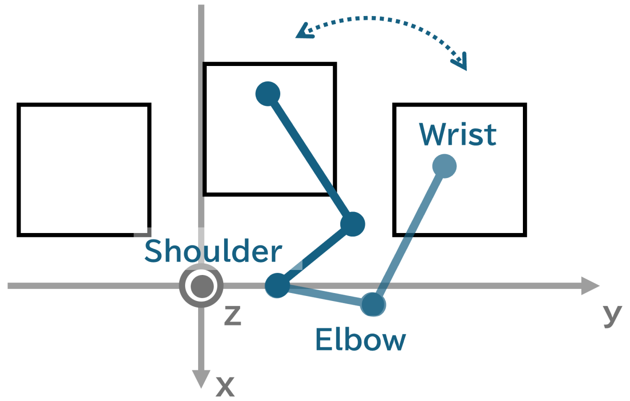


Fig. 12. Side view: manipulation of command space by vertical movement of the wrist.



(a) Overhead view: movement of a person with the habit of moving only the forearm



(b) Overhead view: movement of a person with the habit of moving only the entire arm

Fig. 13. Differences in arm movement.

are located at the opposite end of the command space from the dominant hand. For the inexperienced operator, the position recognition rate for arrangement 1 is also the lowest, mirroring the result for the experienced operator, with minimal difference between arrangements 2 and 3. Again, all subjects are right-handed. The position recognition rates for commands 1 and 4, located opposite the dominant hand, are also particularly low for arrangement 1. However, the position recognition rates for placements 2 and 3 exceed 80%, regardless of the operator's experience level. Therefore, shifting the command space toward the dominant-hand operational accuracy.

Body movement characteristics also affected the results. Observations of the inexperienced operator revealed that when moving the wrist from an upper to a lower command space, the arm often followed an arc on the zx-plane, as presented in **Fig. 12**. This frequently resulted in the wrist being positioned at the edge of the lower command space, potentially contributing to the lower position recognition rate. **Fig. 12** shows a side view of the wrist at the edge of the command space during vertical movement.

Comparison of the results for arrangements 2 and 3 revealed further insights. The differences likely stem from the operator movement habits. The optimal command space arrangement may vary depending on whether the operator moves their arm using the elbow or shoulder as the fulcrum. As shown in **Fig. 13**, when only the elbow tip is moved with the elbow as the fulcrum, the wrist tends to be positioned deeper than intended within the command space. In contrast, if the entire arm is moved with the shoulder as the fulcrum, the wrist can be positioned correctly within the command space. Therefore, it is advisable to determine the command space layout based on the physical characteristics of the operator's movements during operation.

5.2. Evaluation of Recognition Accuracy of Each Gesture in Each Command Space

In Experiment 2, using arrangement 3 from **Fig. 11**, we confirmed the accurate recognition of both. **Table 5** presents the results for a subject with operating experience, while **Table 6** shows the results for inexperienced subjects.

Table 5 shows that the position recognition rates for both gestures 1 and 2 exceed 90%, indicating that the system accurately identified the operator's intended command space. The gesture recognition rates for both gestures also surpassed 90%, demonstrating near-perfect accuracy for participants with operating experience.

As presented in **Table 6**, the average position recognition rates for both gestures exceed 80%, even among inexperienced operators. However, some individual command spaces, specifically commands 1 and 4, exhibit position recognition rates below 80%. This may be because all operators are right-handed, making it relatively difficult for them to position their wrists accurately in command spaces located on the opposite side of their dominant hand. The average gesture recognition rates exceed 60% for both gestures. Interestingly, the gesture recognition rate of gesture 1 is lower than that of gesture 2 for commands 4–6 in the lower row. Conversely, for gesture 2, the gesture recognition rate is lower for commands 1–3 in the upper row than for gesture 1. This discrepancy may be due to the inherent difficulty in distinguishing gestures based on the difference in height between the elbow and wrist when the command spaces are located above and below each other. Adjusting the position of the elbow and wrist to match the height of the target command space likely felt unnatural and posed a challenge for inexperienced users. Therefore, future work should focus on developing methods that allow even inexperienced users to quickly become proficient with the operations.

In addition, during observation of the skeletal point detection process, we found that the accuracy of skeletal point recognition varied depending on the color of the subject's clothing. Although OpenPose was employed in this method, a more accurate skeletal point detection method suitable for real-time processing is required. For intelligent room to be truly usable in daily life, real-time operation and minimization of user burden due to unintended movements are crucial. OpenPose processing time, for skeletal detection increases with the number of people in the image. Moreover, in crowded scenes, skeletal points of nearby individuals are misidentified, and the detection results are sometimes mixed up. Because the image is captured from an overhead perspective to avoid occlusion, posture is often poor for people distant from the camera. However, increasing the resolution to solve this problem, further lengthens processing time. Therefore, future work should focus on identifying and implementing a tool that is more suitable for the proposed method than OpenPose, considering both real-time performance and detection accuracy.

6. Conclusion

In this paper, we presented a system that allows only the operator to control home appliances by introducing an operation start gesture and an operator-specific command space. We conducted experiments to evaluate the accuracy of our proposed gesture-based method for the remote control of devices. Our experiments yielded the following findings:

- Placing the command space considering the operator's dominant hand is effective for improving recognition accuracy.
- Therefore, it is advisable to dynamically adjust the command space layout according to the operator's arm movement tendencies.
- The compatibility between command space arrangement and gesture characteristics should be carefully considered.

We believe that this work represents a significant step toward realizing intelligent room. Our proposed method, operatable with just two simple gestures, requires no complex operational knowledge. Moreover, it can be used even when the location of the device is unknown, because the operation only needs to be linked to the command space. However, because the command space is currently invisible, beginners require some time to become accustomed to its operation. Therefore, we are conducting further research on visualizing command spaces to accelerate the learning process [11].

Future work includes the following:

- Implementing more robust skeletal point detection methods and suitable for real-time processing.
- Exploring the use of MR goggles to dynamically adjust the position and size of the command space, thereby enhancing user accessibility and ease of interaction.

References:

[1] R. Faizrakhmanov, A. Platunov, and M. R. Bahrami, "Smart home user interface: Overview," 2023 Int. Conf. on Industrial Engineering, Applications and Manufacturing (ICIEAM), pp. 595-600, 2023. https://doi.org/10.1109/ICIEAM57311.2023.10139050

[2] H. Nakayama, N. Yabuki, H. Inoue, Y. Sumi, and T. Tsukutani, "A control system for electrical appliances using eye-gaze input," 2012 Int. Symp. on Intelligent Signal Processing and Communications Systems, pp. 410-413, 2012. https://doi.org/10.1109/ISPACS.2012.6473521

[3] S. Sugimura and K. Hoshino, "Wearable hand pose estimation for remote control of a robot on the moon," J. Robot. Mechatron., Vol.29, No.5, pp. 829-837, 2017. https://doi.org/10.20965/jrm.2017.p0829

[4] T. Nagayasu, H. Asano, K. Terabayashi, and K. Umeda, "Improvement of an intelligent room that detects hand waving motion for operation of home appliances," SICE Annual Conf. 2011, pp. 821-826, 2011.

[5] K. Irie, M. Takahashi, K. Terabayashi, H. Ogishima, and K. Umeda, "Skin color registration using recognition ofwaving hands," J. Robot. Mechatron., Vol.22, No.3, pp. 262-272, 2010. https://doi.org/10.20965/jrm.2010.p0262

[6] Z. Cao, T. Simon, S.-E. Wei, and Y. Sheikh, "Realtime multi-person 2d pose estimation using part affinity fields," 2017 IEEE Conf. on Computer Vision and Pattern Recognition, pp. 7291-7299, 2017. https://doi.org/10.1109/CVPR.2017.143

[7] M. Yokota, S. Majima, S. Pathak, and K. Umeda, "Intuitive arm-pointing based home-appliance control from multiple camera views," 2023 32nd IEEE Int. Conf. on Robot and Human Interactive Communication (RO-MAN), pp. 179-184, 2023. https://doi.org/10.1109/RO-MAN57019.2023.10309557

[8] T. Kano, T. Kawamura, H. Asano, T. Nagayasu, and K. Umeda, "Hand waving in command spaces: A framework for operating home appliances," Advanced Robotics, Vol.32, No.18, pp. 999-1006, 2018. https://doi.org/10.1080/01691864.2018.1515661

[9] 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

[10] R. Hartley, "Multiple view geometry in computer vision," Cambridge University Press, 2003.

[11] Y. Mochizuki, M. Yokota, S. Pathak, and K. Umeda, "Visualisable and adjustable command spaces for gesture-based home appliance operation system via hololens2," The 2025 IEEE/SICE Int. Symp. on System Integrations (SII), pp. 1405-1410, 2025. https://doi.org/10.1109/SII59315.2025.10870954

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. 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:
Hitachi Atago Bldg., 15-12 Nishi Shimbashi 2-chome, Minato-ku, Tokyo 105-8410, Japan

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

Name:
Yushin Mochizuki

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:
2024 Received B.Eng. in Precision Mechanics from Chuo University

Membership in Academic Societies:
- The Japan Society for Precision Engineering (JSPE)
- Institute of Electrical and Electronics Engineers (IEEE)

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 (IITM)
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:
- S. Pathak, A. Moro, H. Fujii, A. Yamashita, and H. Asama, "Spherical Video Stabilization by Estimating Rotation from Dense Optical Flow Fields," J. Robot. Mechatron., Vol.29, No.3, pp. 566-579, 2017.
- D. Kim, S. Pathak, A. Moro, A. Yamashita, and H. Asama, "Self-Supervised Optical Flow Derotation Network for Rotation Estimation of a Spherical Camera," Advanced Robotics, Vol.35, No.2, pp. 118-128, 2021.
- S. Pathak, A. Moro, A. Yamashita, and H. Asama, "A Decoupled Virtual Camera Using Spherical Optical Flow," Proc. of the 2016 IEEE Int. Conf. on Image Processing (ICIP2016), 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. in Precision Machinery Engineering from The University of Tokyo
1994- Lecturer, Chuo University
2003-2004 Visiting Worker, National Research Council of Canada

Main Works:
- M. Shinozaki, M. Kusanagi, K. Umeda, G. Godin, and M. Rioux, "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.
- T. Kuroki, K. Terabayashi, and K. Umeda, "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 (IROS2010), pp. 5972-5977, 2010.
- S. Yan, S. Pathak, and K. Umeda, "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)