Advanced Robotics

ISSN: 0169-1864 (Print) 1568-5535 (Online) Journal homepage: http://www.tandfonline.com/loi/tadr20

Hand waving in command spaces: a framework for operating home appliances

Takumi Kano, Takuya Kawamura, Hidetsugu Asano, Takeshi Nagayasu & Kazunori Umeda

ADVANCED ROBOTICS
https://doi.org/10.1080/01691864.2018.1515661

FULL PAPER

Hand waving in command spaces: a framework for operating home appliances

Takumi Kano[a], Takuya Kawamura[b], Hidetsugu Asano[c], Takeshi Nagayasu[a] and Kazunori Umeda[b]

[a]Course of Precision Engineering, Chuo University, Tokyo, Japan; [b]Department of Precision Mechanics, Chuo University, Tokyo, Japan; [c]Research & Development Division, Pioneer Corporation, Saitama, Japan

ABSTRACT
In this paper, hand waving in a 'command space' is proposed as a new framework for operating home appliances. A command space is associated with the operation of a home appliance; an operation is retrieved by waving a hand in the command space. A home appliance operation system with the function of multiplexing command space modules and integrating multiple commands is constructed. With the proposed framework and system, home appliance operation is possible only by hand waving, an intuitive gesture for a human. Experiments were conducted that verified that complex operations, such as scheduling the recording of a TV program, can be performed with the system. The usability of the constructed system is evaluated using Brooke's System Usability Scale (SUS). The average SUS score was 66.8, which indicates that the subjects had relatively positive impression on the system in spite of long operation time.

ARTICLE HISTORY
Received 30 March 2018
Revised 14 July 2018
Accepted 7 August 2018

KEYWORDS
Gesture recognition; human interface; image processing; intelligent room; command space

1. Introduction

In recent years, there has been increasing demand for human–machine interfaces that can operate devices more easily because the machines that exist in human living environments are becoming more complicated and diversified. Of these interfaces, gesture interface, i.e. the operation of appliances by a gesture, is superior in that it does not require operators to have a device such as a remote control or to touch a specific device; thus, they can intuitively operate appliances easily anytime and anywhere. Various methods have been proposed to realize gesture interfaces [1–11]; some of them utilize images. There are several approaches for them, such as detecting skin color regions and recognizing gestures from the shape or motion of the regions [1,2], using an optical flow [3], extracting a person by background subtraction, and estimating his/her motion [4], etc. Peshkova et al. present a survey of gesture interfaces using images for controlling unmanned aerial vehicles [5]. Some studies utilize range images, such as extracting a whole person from range images and recognizing his/her gesture [6], and some focus only on the position of the hand region [7]. Plouffe and Cretu realized complex hand gesture recognition by extracting 3D hand shape using a Kinect camera [8]. In the methods using these images, the operator does not need to have a specific device for operation and can operate the appliance at any time. However, it is necessary to face a camera within a specific range of the camera when performing the gesture; thus, restrictions on places still exist. One method uses a position sensor and an acceleration sensor to recognize a gesture by attaching a glove-type device and obtaining the spatial position of the hand [9]. Fukui et al. proposed a wrist contour measuring device to recognize hand shape [10]. Baraldi et al. proposed to use a wearable camera mounted on a head for hand gesture recognition [11]. With these methods, the operator can operate the machine anywhere and at any time, but it is necessary to attach a device. To overcome these issues, a system called an Intelligent Room [12,13] was proposed that continuously detects a gesture to be carried out by using multiple cameras in the room to detect the hand gestures of the operator and to operate home appliances intuitively anytime, anywhere. However, the system requires an operator to perform multiple gestures in a specific order to perform an operation, which is not a simple interface.

In this paper, we propose a new framework for operating home appliances—hand waving in 'command spaces'—with which an operator can operate various home appliances with only one gesture, i.e. hand waving, by using position information where the gesture is performed.

The rest of the paper is organized as follows. In section 2, we give an overview of the Intelligent Room

CONTACT Kazunori Umeda umeda@ mech.chuo-u.ac.jp

Supplemental data for this article can be accessed here. https://doi.org/10.1080/01691864.2018.1515661

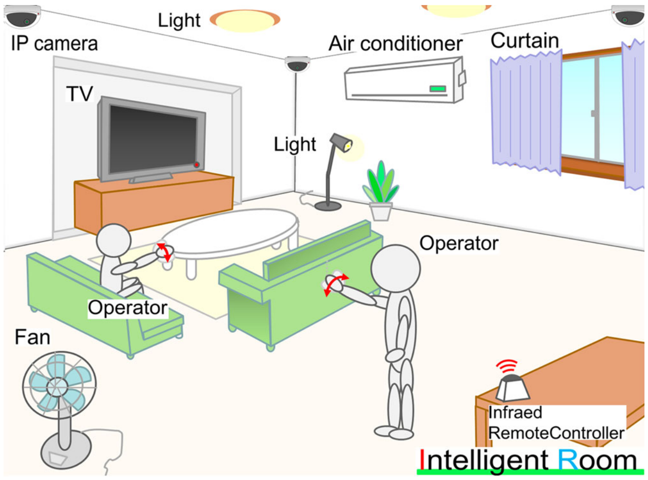

Figure 1. Conceptual diagram of the intelligent room.

[12,13] in which the new gesture interface is installed. In section 3, we explain the method of detecting hand waving in the Intelligent Room. Then, we propose the framework of hand waving in command spaces in section 4 and implement a home appliance operation system using the framework in section 5. In section 6, we verify the framework and the constructed system by experiments; finally, we conclude the paper in section 7.

2. Overview of intelligent rooms

An Intelligent Room, in which the home appliance operation system is installed, is a room where multiple cameras are installed indoors and that functions intelligently to control home appliances. The conceptual diagram of the Intelligent Room is shown in Figure 1. It is assumed to be applied in general offices and homes. As a specific function, home appliances such as television (TV) sets and lighting equipment can be operated with gestures.

In this paper, we associate commands for operating home appliance to spaces in advance, and by waving a hand in a predefined space, the command embedded in the space is retrieved, and the related home appliance is operated with the command. Note that the command spaces can be set anywhere in the room, though the room-wide detectability is restricted by fixing them. Hand-waving is detected by using the installed multiple cameras. In our implementation, the number of the cameras is four. When three or more cameras are installed, we can obtain three-dimensional (3D) coordinates of hand-waving by stereo vision using two or more cameras, even if some cameras are hidden by furniture or a person.

3. Detection of hand waving

For detecting hand waving, we adopted the method in the study by Terabayashi et al. [14] that extracts features by applying fast Fourier transform (FFT) to a time series of intensity values of each pixel. We introduce an outline of the method in this section.

3.1. Preprocessing

The color image obtained from the camera is changed to grayscale, and the resolution is lowered. The expected effects of reducing the resolution of the image are simpler calculations and less noise. In addition, the change in intensity values due to hand waving becomes smooth, and we can expect that a pattern of intensity variation closer to a sinusoidal wave can be obtained.

3.2. Extraction of frequency features

We apply FFT in the temporal direction to each pixel of the low-resolution image. In the pixel corresponding to the hand-waving region, the pixel value varies between the hand and background regions. Since this change of intensity value is periodic, FFT effectively detects frequency features.

3.3. Detection of hand-waving candidate pixels

Based on the result of FFT, hand waving is detected for each pixel by the following method.

(1) Find the maximum value of the amplitude at 3–6 [Hz].
(2) Find the difference in intensity between this intensity and 2 [Hz].
(3) Accumulate this value temporally.
(4) When the accumulated value exceeds the threshold value, it is detected as a hand-waving candidate pixel.

3.4. Calculation of hand-waving coordinates

The method in Section 3.3 is performed by multiple cameras, and pixels that satisfy the epipolar constraint between two or more cameras are identified as hand-waving pixels. After that, we derive 3D coordinates of hand-waving based on the principle of stereo vision.

4. Operation of home appliances using command spaces

We propose a new framework for operating home appliances: hand waving in ‘command spaces.’ Home appliances are operated by waving a hand in the space associated with the operation of a specific home appliance. In addition, we introduce the multiplexing of command space modules and the integration of multiple commands

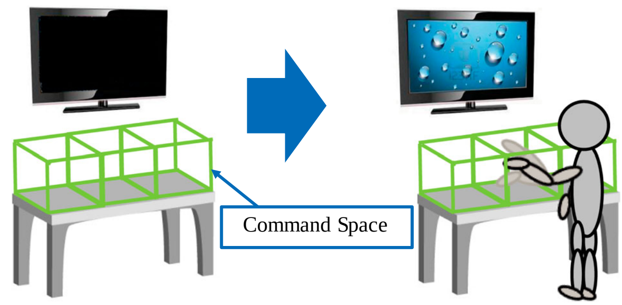

Figure 2. Conceptual diagram of the command space.

to realize the operation of a complicated home appliance such as scheduling the recording of a TV program.

4.1. Command spaces

Using the idea of Niitsuma's Spatial Memory [15] that embeds information in space, we associate a command to operate a home appliance with a particular space. By performing a predetermined gesture in the space (command space) associated with operation of the appliance, it is possible to perform the associated home appliance operation. In this paper, we use hand waving only as a gesture to operate home appliances. A conceptual diagram of the command space in this paper is shown in Figure 2. Figure 2 illustrates that the command space in the middle is associated with the power ON/OFF of the TV set, and by waving the hand in the command space, the command is retrieved and carried out, and, consequently, the TV set is turned on.

4.2. Multiplexing of command space modules

When performing complex operations of a home appliance, it is necessary to sequentially perform multiple types of operations. Therefore, when one home appliance operation is associated with one command space, as many command spaces as there are types of operations are required. Therefore, we introduce the multiplexing of command space modules. An arbitrary number of modules can be set. A selected module is changed when a command space associated with changing modules or performing specific home appliance operations is selected. When the selected module is changed, it should be indicated to the operator. Audio guidance or a display can be used for this purpose.

4.3. Integration of commands

By integrating multiple home appliance operation commands into one command space, complicated home appliance operation is possible with fewer operations.

5. Constructed home appliance operation system

5.1. System overview

We constructed a home appliance operation system in a room as shown in Figure 3. The room size is 6 [m] × 6 [m], and a TV set SONY KJ-55X9350D was set near the wall. An operator's hand waving is recognized by cameras installed in the four corners of the room. The cameras are Axis 233D network dome camera. The angle of view is 55.8 [°] × 43.3 [°] (maximum), the resolution is 640 × 480 [pixel], and the frame rate is 30 [fps]. The PC to implement the system is a Windows PC with Intel Core i7 3.33 GHz, 6GB RAM, and NVIDIA GeForce GTX 560. We used Visual Studio (C# and C++) as the development environment and used some OpenCV functions. By waving a hand in the command space installed in the room, it is possible to operate the corresponding home appliance. The command space we implemented is a cubic space of 0.3 [m] on one side and is arranged on a table, as shown in Figure 4. There are seven command spaces. If the operator remembers the position of the command spaces and the relation between commands and command spaces, no other interface is necessary. If that is not the case, some interface to support the operator should be added, such as a display. In our implementation, we use sheets of paper with written commands placed on the table to indicate the position of the command space and the operation associated with each command space, as shown in Figure 5. Each command space is set just above the sheet. Note that multiple commands are written on each sheet according to the multiplexing of the command space modules as described in the next section. The number of modules in this case is three.

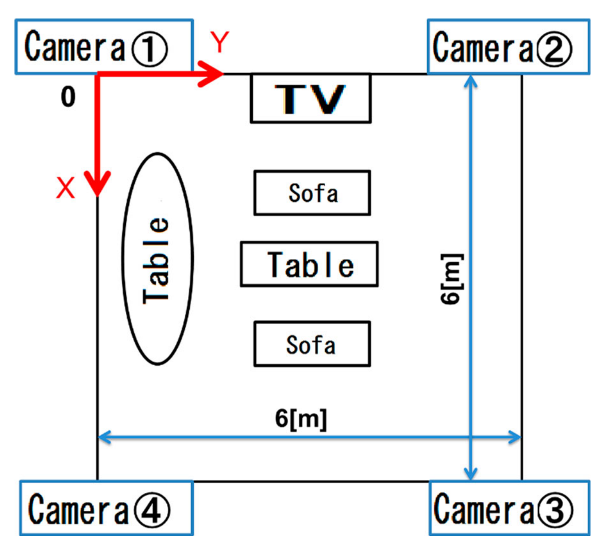

Figure 3. Schematic diagram of a home appliance operation system.

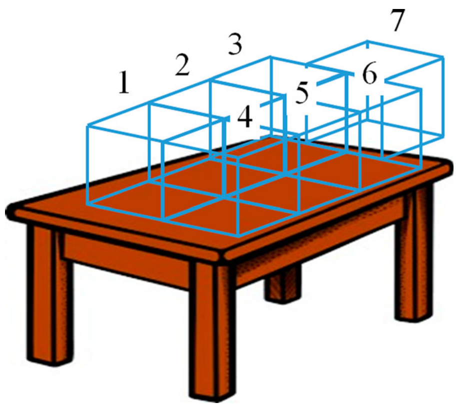

Figure 4. Arrangement of the command spaces.

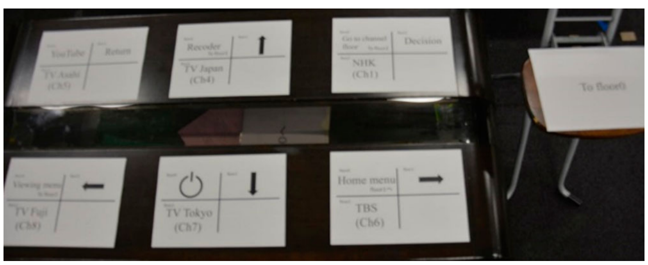

Figure 5. Sheets of paper to show home appliance operations.

5.2. *Multiplexing of command space modules*

Figure 6 shows the multiplexing structure of the constructed home appliance operation system. The system starts from module 0. In module 0, operations that should be initially performed, such as 'TV ON,''viewing menu,' and "to the channel module," are associated. In module 1, operations for selecting an item such as a four-way controller, 'decision,' and 'return' are associated. Module 2 is for channel selection, and each command space is associated with a different TV channel. Note that command space #7 is associated with a return to module 0 and can be returned to the initial state. When selecting 'recorder' (space #2), 'viewing menu' (space #4), or 'home menu' (space #6) in module 0, the selected module is changed to module 1, since it is expected to perform a cross operation for the next operations. When selecting 'to the channel module' (space #3) in module 0, the selected module is changed to module 2.

5.3. *Integrated command spaces*

'YouTube' (space #1) and 'recorder' (space #2) are integrated command spaces. 'YouTube' is the integration of 'home,''↓,''↓,''↓,' and 'decision,' and 'recorder' is the integration of 'viewing menu,''→,''→,' and 'decision.' By selecting these spaces, multiple operations can be performed sequentially at once.

5.4. *Audio guidance*

When changing the module, since there is no change in the home appliance, etc., it is difficult for an operator to confirm the change of module, and some interface should be given to improve system operability. In our implementation, we use audio guidance such as 'changed to module 1' when the module is changed.

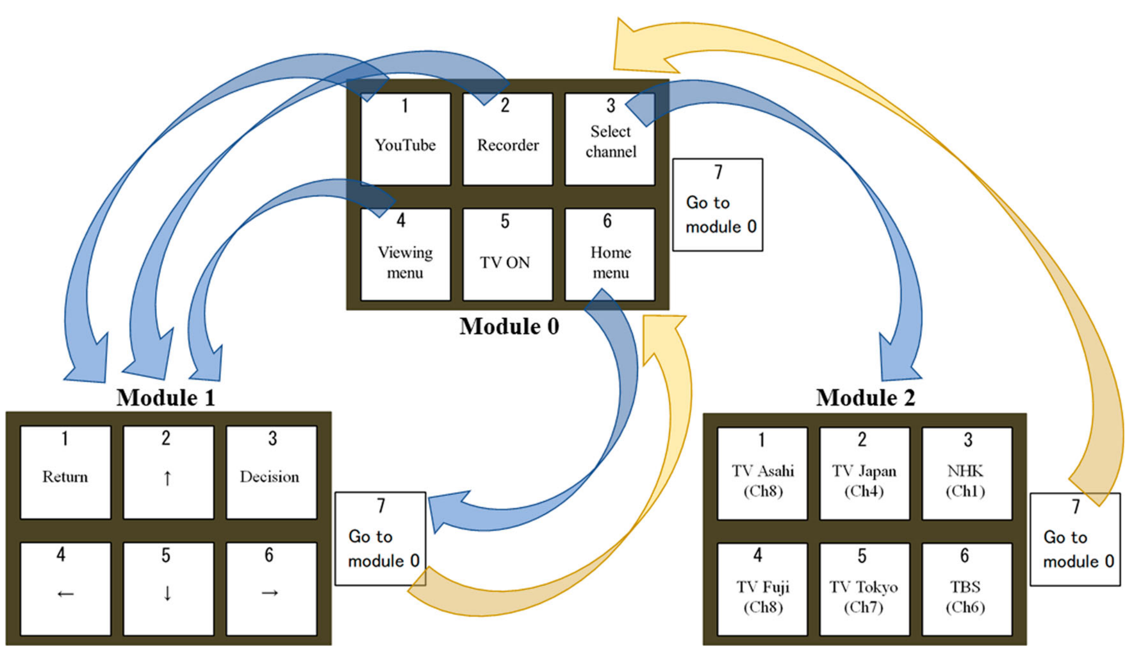

Figure 6. Constructed command spaces with a multiplexing structure.

6. Experiments

We conducted experiments to evaluate the constructed home appliance operation system with the proposed hand-waving interface.

6.1. Experiments on command space selection

First, we conducted experiments to show that it is possible to select a command space by hand-waving recognition. The subjects were three males in their twenties who sat on the sofa across the table as viewed from the TV set in Figure 3 and waved their hands. Each subject waved a hand 10 times in each of the 7 command spaces and investigated the recognition rate.

Results of the experiments are shown in Table 1. The table row is the number of the command space in which hands were waved, and the column is the number of the command space in which hand waving was recognized. The recognition rate was 94% on average. In command spaces 1, 2, and 3, which are far from the operator, the recognition rate was as low as 80% on average, whereas in other command spaces, the recognition rate was 100%. The reason for the recognition errors in the far command spaces is thought to be that other parts of the arm also moved periodically in accordance with hand waving, and their motions were erroneously detected.

Table 1. Recognition results of each command space (row: performed; column: recognized).

	1	2	3	4	5	6	7	Recognition rate (%)
1	26	0	0	4	0	0	0	87
2	0	23	0	3	4	0	0	77
3	0	0	23	0	2	5	0	77
4	0	0	0	30	0	0	0	100
5	0	0	0	0	30	0	0	100
6	0	0	0	0	0	30	0	100
7	0	0	0	0	0	0	30	100

6.2. Experiment on home appliance operation

We conducted experiments to confirm that practical home appliance operation is possible with the constructed system. We chose scheduling the recording of a TV program with a specified date and time as a practical and complex home appliance operation. The number of subjects was 20, and they were all in their 20s.

The situation of the experiment is shown in Figure 7. The left four images are captured by the four cameras in the Intelligent Room, and the right image is the TV set to be operated.

The procedures of the experiment are shown below.

- Start the TV set with command space 5.
- Enter the menu for the recording with command space 4 (move to module 1).
- Select the recording menu by operating the four-way controller with command spaces 2, 4, 5, and 6.
- Use the four-way controller to select the scheduling menu with the specified date and time.
- Use the four-way controller to specify the date and time.
- Select scheduling confirmation with the four-way controller, and confirm with command space 3.

As a result of the experiments, all subjects succeeded in scheduling the recording of a TV program with a specified date and time. Table 2 shows the time required for the task by each subject, which was 112.4 [s] on average. On the other hand, completing the same operation with the

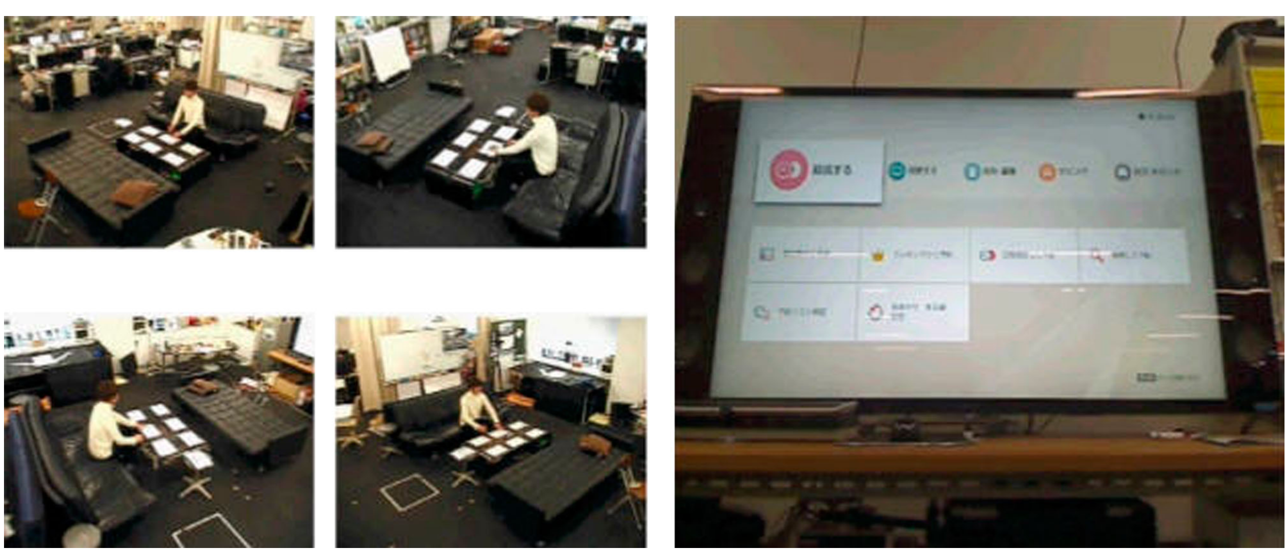

Figure 7. Experimental situation. Left: images captured by four cameras; right: TV set to be operated.

Table 2. Time required for scheduling the recording of a TV program with a specified date and time.

Subject	1	2	3	4	5	6	7	8	9	10
Time (s)	125	93	174	97	115	108	143	81	146	82

11	12	13	14	15	16	17	18	19	20	Average	S.D.
106	108	60	246	111	66	66	59	130	131	112.4	43.0

remote control was 35.3 [s], which is an average of eleven subjects (subject 10–20 in Table 2), with a standard deviation (S.D.) of 5.3 [s]. Note that 'Start the TV set' and 'enter the menu for the recording' correspond to 'power' and 'option' buttons of a remote control respectively. Therefore, it is difficult to say that the constructed system is an easy-to-use interface regarding time. The reason the system is time-consuming is that it takes time to recognize hand waving, and by repeating hand waving multiple times, the total operation time accumulated and became long.

6.3. Evaluation of usability

After the above-mentioned experiments, we conducted questionnaires using Brooke's System Usability Scale (SUS) [16] to evaluate the usability of the constructed system. Ten items of the SUS are shown below.

(1) I think that I would like to use this system frequently.
(2) I found the system unnecessarily complex.
(3) I thought the system was easy to use.
(4) I think that I would need the support of a technical person to be able to use this system.
(5) I found the various functions in this system were well integrated.
(6) I thought there was too much inconsistency in this system.
(7) I would imagine that most people would learn to use this system very quickly.
(8) I found the system very cumbersome to use.
(9) I felt very confident using the system.
(10) I needed to learn a lot of things before I could get going with this system.

The method of calculating usability evaluations is shown below.

- For the 10 items above, ask a subject to evaluate in 5 stages from 1 (completely disagree) to 5 (totally agree).
- If the item's number is odd, subtract 1 from the score.
- If the item's number is even, subtract the score from 5.
- Add up the scores for the 10 indicators and multiply the sum by 2.5. The total score is 0 (unusable) to 100 (very easy to use).

Table 3 shows the usability evaluation for each subject. As a result of the questionnaire, the scores for 20 subjects ranged from 40 to 90, and their average was 66.8. This is nearly 68, the average score of SUS, and scores of 10 subjects exceed the average score, in spite of the long operation time mentioned above. This is thought to be because the subjects felt it was fun to operate the television by hand waving, which is an unusual method of operation; consequently, the impression of the system was relatively positive, and a good score was obtained.

We compared the usability of the constructed system with remote control. We utilized the SUS questionnaire and asked each subject who tested both the constructed system and the remote control that which of them fit the SUS 10 items. The answers were either of them and mostly same. And when the constructed system is chosen, 1 is added for the odd item and subtracted for the even item. Tables 4 and 5 show the results for each subject and each item respectively. In Table 4, subject numbers correspond to the ones in Tables 2 and 3. It is indicated that the usability of the constructed system is inferior to the remote control, though it is not statistically significant. However, the constructed system is superior for items 6 and 10, and even for item 7. It suggests that the constructed system has consistency and is easy to learn.

6.4. Evaluation for middle-aged subjects

We conducted another experiment to evaluate the effect of subjects' ages. The number of subjects is 6; 3 are in their 60s, and 3 are in their 50s, i.e. all subjects in this experiment are middle-aged.

Table 6 shows the experimental results of both time and evaluation score of usability. By comparing with

Table 3. Evaluation score of the usability of the constructed system.

Subject	1	2	3	4	5	6	7	8	9	10
Evaluation score	57.5	55	72.5	90	65	62.5	75	72.5	62.5	80

11	12	13	14	15	16	17	18	19	20	Average	S.D.
70	75	60	57.5	60	60	72.5	40	75	72.5	66.8	10.6

Table 4. Comparison of usability between the constructed system and remote control.

Subject	10	11	12	13	14	15	16	17	18	19	20	Average	S.D.
Score	−2	7	−6	−5	−4	−3	2	0	−5	−3	−5	−2.2	3.7

Table 5. Comparison of usability between the constructed system and remote control for each item.

Item	1	2	3	4	5	6	7	8	9	10	Average	S.D.
Score	−6	−3	−3	−3	−5	3	0	−3	−6	2	−2.4	3.0

Table 6. Time and evaluation score of usability for middle-aged subjects.

Subject	1	2	3	4	5	6	Average	S.D.
Age	60s	60s	60s	50s	50s	50s	–	–
Time (s)	85	111	83	87	124	80	95.0	16.5
Evaluation score	65	57.5	70	60	80	45	62.9	10.8

Tables 2 and 3, it is shown that both their times and scores are not different much from the young subjects' ones; the p-values of two-tailed t-test are 0.36 for time and 0.47 for evaluation score, though the average time is shorter and the average score is lower. It can be said so far that middle-aged people can use the constructed system. Experimental evaluation with more subjects and older subjects should be done in the future.

6.5. Discussion

Results of the experiments confirmed that the constructed system can perform complicated home appliance operations. However, these operations took time, which is an issue to be overcome. It is also necessary to improve the recognition rate of the command spaces. To do this, it may help to change the arrangement of the command spaces. In addition, we are now using sheets of paper to indicate the position of the command space and the operation associated with each command space. This is expected to make the system more flexible and easier to use by presenting information about the current command space module on a display.

7. Conclusion

In this paper, we proposed a framework to utilize hand waving to operate home appliances, based on an Intelligent Room. A command space that is associated with home appliance operation is introduced, and the operation is retrieved by waving a hand in the command space. We then realized a home appliance operation system with the function of multiplexing command space modules and integrating multiple commands. With the proposed framework and the system, operating a home appliance is possible by simply waving hand, a gesture intuitive for humans. Experiments were conducted, and it was verified that complex home appliance operations, such as scheduling the recording of a TV program, can be performed with the system. We also evaluated the usability of the constructed system using the SUS. The average SUS score was 66.8, which indicates that the subjects had relatively positive impression on the system in spite of the long operation time.

In the future, we aim to improve the system by reducing the time needed to operate a home appliance and by presenting command space information more flexibly. So far the feedback to the user is just simple audio guidance. Using a wearable glass to display a virtual 3D interface [17] may be a candidate for the feedback. In addition, we will try to improve the system so that an operator can set the command spaces flexibly by hand waving anytime and anywhere.

Acknowledgements

We thank Shixun Yan for his support of experiments.

Disclosure statement

No potential conflict of interest was reported by the authors.

Notes on contributors

Takumi Kano B.Eng. degree in precision mechanics and M.Eng. degree in precision engineering from Chuo University, Japan, in 2016 and 2018, respectively. He currently works at Duplo Corporation. His research interests include human interface using vision and range image processing.

Takuya Kawamura received B.Eng. degree in precision mechanics from Chuo University, Japan, in 2017. He currently works at Kiriu Corporation. His research interests include human interface using vision.

Hidetsugu Asano received B.Eng. and M.Eng. degrees from Yokohama National University, Japan, in 2000 and 2002 respectively, and Ph.D. degree in medicine from Tokyo Women's Medical University, Japan, in 2017. He is a researcher in Pioneer Corporation since 2002. His research interests include image processing, machine learning, human interface using vision,

and medicine. He is a member of the Japan Neurosurgical Society.

Takeshi Nagayasu received B.Eng. degree in precision mechanics and M.Eng. degree in precision engineering from Chuo University, Japan, in 2010 and 2012, respectively. He currently works at Paramount Bed Co., Ltd. His research interests include machine learning, automatic driving, speech recognition, and image recognition.

Kazunori Umeda received B.Eng., M.Eng., and Ph.D. degrees in precision machinery engineering from the University of Tokyo, Japan, in 1989, 1991 and 1994, respectively. He became a Lecturer of Precision Mechanics at Chuo University, Japan in 1994, and is currently a Professor since 2006. He was a visiting worker at National Research Council of Canada from 2003 to 2004. His research interests include robot vision, 3D vision, and human interface using vision. He is a member of RSJ, IEEE, JSPE, JSME, SICE, IEICE, etc.

References

[1] Pavlovic VI, Sharma R, Huang TS. Visual interpretation of hand gestures for human-computer interaction: a review. IEEE Trans PAMI. 1997;19(7):677–695.
[2] Dominguez SM, Keaton T, Sayed AH. A robust finger tracking method for multimodal wearable computer interfacing. IEEE Trans Multimedia. 2006;8(5):956–972.
[3] Bayazit M, Couture-Beil A, Mori G. Real-time motion-based gesture recognition using the GPU. Iapr conference on MVA 2009; 2009:9-12.
[4] Li H, Greenspan M. Multi-scale gesture recognition from time-varying contours. Proceedings of the tenth IEEE international conference on computer vision (ICCV2005); 2005.
[5] Peshkova E, Hitz M, Kaufmann B. Natural interaction techniques for an unmanned aerial vehicle system. IEEE Pervasive Computing. 2017;16(1):34–42.
[6] Liu X, Fujimura K. Hand gesture recognition using depth data. Ieee international conference on automatic face and gesture recognition; 2004:529–534.
[7] Hirsch M, Lanman D, Raskar R, et al. Bidi screen: depth and lighting aware interaction and display. Acm SIGGRAPH Asia 2009; 2009:159.
[8] Plouffe G, Cretu AM. Static and dynamic hand gesture recognition in depth data using dynamic time warping. Trans Instrument Measure. 2016;65(2):305–316.
[9] Vogler C, Metaxas D. Asl recognition based on a coupling between HMMs and 3D motion analysis. Proceedings of the IEEE international conference on computer vision; 1998:363–369.
[10] Fukui R, Watanabe M, Shimosaka M, et al. Development of wrist contour measuring device for an interface using hand shape recognition. Advanced Robotics. 2013;27(7):481–492.
[11] Baraldi L, Paci F, Serra G, et al. Gesture recognition using wearable vision sensors to enhance visitors' museum experiences. IEEE Sens J. 2015;15(5):2705–2714.
[12] Irie K, Wakamura N, Umeda K. Construction of an intelligent room based on gesture recognition—operation of electric appliances with hand gestures. Proceedings of the IEEE/RSJ international conference on intelligent robots and systems; 2004:193–198.
[13] Nagayasu T, Asano H, Takahashi M, et al. Improvement of an intelligent room that detects hand waving motion for operation of home appliances. Proceedings of the SICE annual conference 2011; 2011:821–826.
[14] Terabayashi K, Asano H, Nagayasu T, et al. Detection of small-waving hand by distributed camera system. Proc. of the ninth international conference on networked sensing systems (INSS2012); 2012.
[15] Niitsuma M, Kobayashi H, Shiraishi A. Enhancement of spatial memory for applying to sequential activity. J Advanc Sci. 2012;9(1):121–137.
[16] Brooke J. SUS–a quick and dirty usability scale. In: Jordan PW, Thomas B, McClelland IL, Weerdmeester B, editors. Usability evaluation in industry. London: CRC Press; 1996. p. 189–194.
[17] Zhang S, Wang J, Gong Y, et al. Free-hand gesture control with 'touchable' virtual interface for human-3DTV interaction. Proceedings of the 3DTV-conference; 2015.