Intuitive Arm-Pointing based Home-Appliance Control from Multiple Camera Views

Masae Yokota[1], Soichiro Majima[1], Sarthak Pathak[2] and Kazunori Umeda[2]

***Abstract*— The purpose of this paper is to construct and evaluate a system to operate home appliances by pointing. In Human Machine Interface (HMI) design, a natural operating method is important. Pointing is a universal gesture for selecting an object. Arm-pointing to an appliance and selecting it to perform a simple operation is a very intuitive and easy-to-use method of operation. Many studies prepare data with locations of appliances and their sizes. In this paper, we a camera-based system where the user can simply point at an appliance to select and operate it is proposed. The user's pointing direction and appliance locations are estimated automatically from image frames. This eliminates the need for any preparation beforehand and the appliances can be moved during operation. The proposed method was implemented and experimentally evaluated. It was found that the average recognition rates were about 87% and 57% when a humidifier and a TV were operated.**

I. INTRODUCTION

We spend our days surrounded by many appliances. Generally, a single home appliance is operated by a single remote control. The number of remote controls increases as the number of home appliances increases. However, to operate a home appliance with a remote control, the user must find and pick up the remote control. In addition, as appliances become more multifunctional, the buttons and letters on the remote control become smaller and more difficult to operate. Smart remote controls with voice recognition have recently become popular, but they cannot be used in noisy surroundings or in situations where the user cannot speak. Therefore, a system that can operate home appliances with natural gestures, independent of the remote control's position and the appliances' position, would be useful.

Natural gestures include hand gestures and eye gestures. Home appliance operation systems using hand gestures include those that select the home appliance to be operated by pointing it [1]-[5], those that tie gestures to each operation [6]-[12], and those that divide commands by the space in which they are performed [13][14]. There are also methods of manipulating machines using eye gestures and eye tracking [15][16].

Of the above related studies, the simplest method is to link gestures and operations on a one-to-one basis. However, if gestures are differentiated according to which home appliance is turned on, the number of gestures that must be memorized increases with the number of operations, which becomes a burden for the user. In addition, intuitive operation methods that operate appliances by pointing at or facing them all assume that the position and size of the appliance in the space is known. Therefore, whenever the position of an appliance is changed or the appliance is replaced, the user needs to update the information. Thus, it would be useful to have a system that allows users to operate home appliances with intuitive instructions even when the location or size of the home appliance is not known in advance.

The purpose of this study is to build a system that uses arm pointing to select and turn on/off appliances without prior information on the appliance's location.

II. PROPOSED METHOD

A. Environment

The system is intended to be used in a room which is assumed to be an ordinary house or office. In the room, the user can control appliances with gestures without the need for special attachments such as wristbands or gloves. As shown in Fig. 1, cameras are mounted at the four corners of the ceiling of this room, and these cameras detect the 3D positions of the users and appliances. Gestures are detected by these cameras and commands are sent to appliances through an IR transmitter placed in the room. The camera used in this system is capable of panning, tilting and zooming. The smart remote control transmits the input via Wi-Fi to the home appliance via infrared communication.

B. System Overview

The four camera images are combined one frame at a time to produce a single image. The flow of this system is shown in Fig. 2.

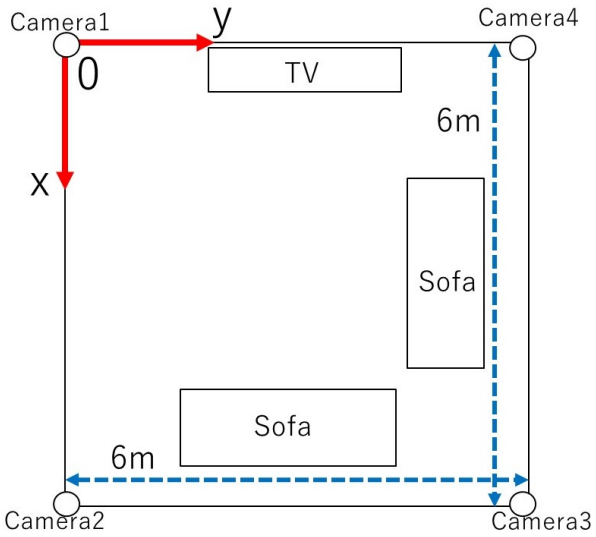
The first home appliance is detected in the 10th frame using You Only Look Once (YOLO)[17], and home appliance detection is performed every 10 frames thereafter. The frames in which the home appliances are detected are denoted $frame_{YOLO}$.

After the first home appliance is detected, the user's skeleton points are detected once every three frames using OpenPose[18]. Hereafter, the frames in which skeleton point detection is performed are called $frame_{OP}$.

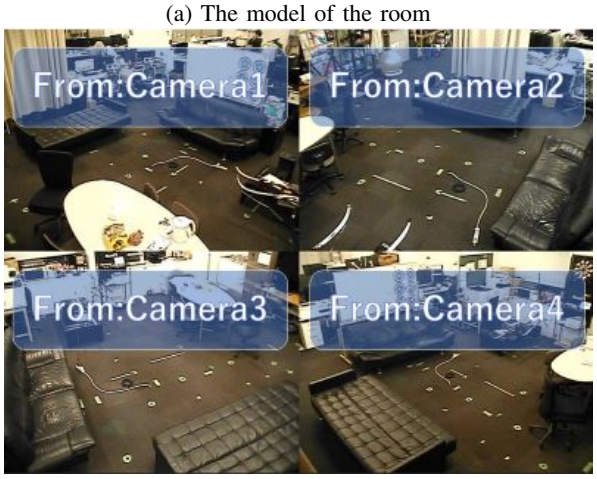
The home appliance detection and skeleton point detection are performed in $frame_{YOLO}$ and $frame_{OP}$, respectively, and triangulation is also performed at the same time. The 3D coordinates of the 3D center point C_a of the home appliance are obtained in $frame_{YOLO}$, as shown in Fig. 3a. The 3D coordinates of the skeleton points at the elbow and wrist are obtained at $frame_{OP}$, and the vector $\vec{v_p}$ from

[1]The Precision Engineering Course, Graduate School of Science and Engineering, Chuo University, 1-13-27 Kasuga, Bunkyo-ku, Tokyo, Japan. (Corresponding author: yokota@sensor.mech.chuo-u.ac.jp)

[2]The Department of Precision Mechanics, Faculty of Science and Engineering, Chuo University.



(a) The model of the room



(b) Combined images

Fig. 1: The overview of the room used in the proposed method

the elbow to the wrist is calculated. Arm pointing is detected by triangulating the 3D positions of the user's skeleton points and the appliances. When this state is established for any n consecutive frames, the home appliance can be turned on or off.

Before performing a series of operation actions, the user performs a gesture to prevent incorrect operation. This allows the user to start the appliance operation. Gestures as shown in Fig. 3b allows the user to switch between a state in which operation of the home appliance is possible and a state in which operation of the home appliance is impossible.

C. Detection of Appliances

Since the default dataset COCO[19] of YOLOv4[20] could not detect humidifier and robot vacuum cleaners, training data was prepared for this experiment to create a detector that could detect these two types of appliances. For training the detector, it is necessary to capture actual photos of the appliances to be controlled. Of course, if those categories

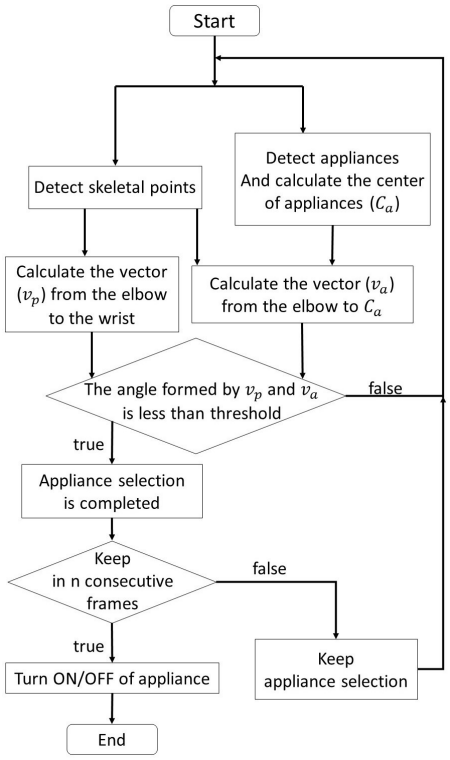


Fig. 2: Flowchart of the processing after the start gesture

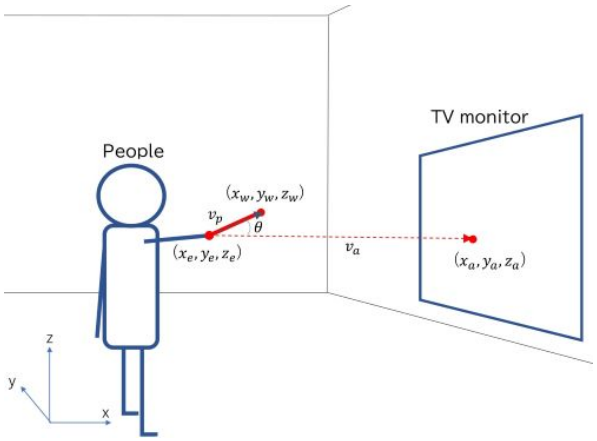
already exist in the COCO dataset, which is default for YOLOv4, there is no need to capture images again. In this paper, a hundred twenty-two photographs were taken while changing the position of the appliances in the room for training data.

After annotating the images with a sofa, a robot, a TV, and a humidifier, the data was enhanced by adding noise, changing brightness, and inverting left and right. After expansion, of the 2928 data sets, 2050 were used for training and 878 for test. The results of the test showed that Average Precision for sofa, robot, TV, and humidifier were 100%, 99.14%, 100%, and 99.12%, with a mean Average Precision (mAP) of 99.57%.

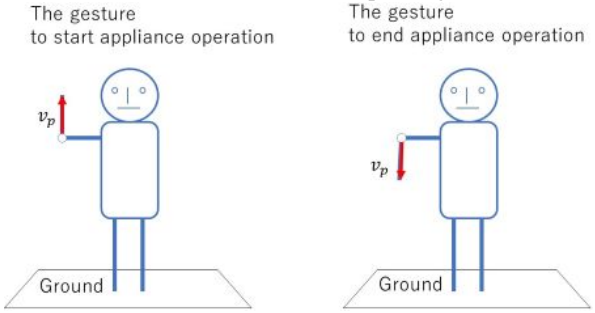
When an appliance is detected by YOLO, a bounding box surrounding the appliance is drawn as shown in Fig. 4. Therefore, the center of this bounding box is considered to be the center of the appliance.

D. Detection of Skeletal Points

Skeletal points are detected using OpenPose[18]. OpenPose is a CNN-based algorithm that performs person pose estimation by cascading heatmaps and Part Affinity Fields. This algorithm can estimate the 2D pose of a person from images. Each skeletal point is numbered in the algorithm, and the 2D coordinates of the right wrist (No. 4) and right elbow (No. 3) are used in this system. Each skeletal point is assigned a confidence level, and a confidence threshold must be set to ensure that low confidence and inaccurate coordinates are not used in the calculation. Skeleton points with a confidence level of 0.60 or higher were adopted in



(a) Estimation of arm-pointing



(b) Gestures for switching start and end appliance operation

Fig. 3: Gestures used in this method

this system.

E. Triangulation of Appliances and Skeletal Points

As shown in Fig. 4, multiple 2D coordinates of the bounding box center are obtained for the same home appliance for each camera. The 3D coordinates obtained by triangulating[21] the center points of these bounding boxes are then used as the 3D center coordinates C_a of the home appliance. In addition, the 3D coordinates of the right elbow and right wrist are calculated by triangulating using the 2D coordinates of the right elbow and right wrist on each camera image. Let (x_e, y_e, z_e) denote the 3D coordinates of the right elbow and (x_w, y_w, z_w) the 3D coordinates of the right wrist.

F. Recognition of Gestures

The vector $\vec{v_p}$ from the right elbow to the right wrist can be obtained by (1).The vector $\vec{v_a}$ from the 3D coordinates (x_e, y_e, z_e) of the right elbow to the 3D center coordinates $C_a = (x_a, y_a, z_a)$ of the appliance can be obtained by (2).

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

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

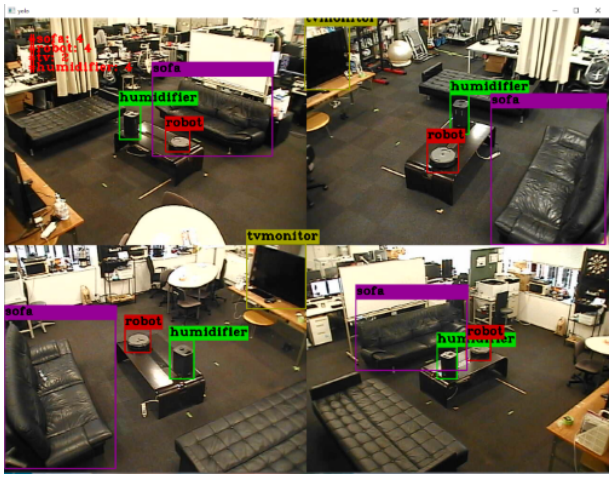


Fig. 4: Object detection results: bounding boxes

The angle θ between $\vec{v_p}$ and $\vec{v_a}$ is used to judge the home appliance operation. If θ is less than or equal to the threshold value θ_{th}, the instruction has been made, and if the instruction state is maintained for n consecutive frames, the home appliance is turned on. The θ is obtained by (3).

$$\theta = \arccos\left(\frac{\vec{v_p} \cdot \vec{v_a}}{|\vec{v_p}||\vec{v_a}|}\right) \tag{3}$$

Gestures as shown in Fig. 3b are provided to switch the start and end of home appliance operation, so that the arm pointer is recognized only when the user wants to operate the home appliance. Home appliance operation starts when the gesture to raise the arm so that the angle between $\vec{v_p}$ and the ground is between 1.25 radians and 1.58 radians and the wrist is above the elbow. The gesture terminates the home appliance operation when the angle between $\vec{v_p}$ and the ground is between 1.25 radians and 1.58 radians and the wrist is below the elbow.

The greater the distance from the appliance, the more the appliance can be approximated as a point by the user's eye. Therefore, the threshold value θ_{th} is varied according to the value $|\vec{v_a}|$.

Assuming that the relationship between the distance to the appliance and the threshold can be expressed as a linear equation, the coefficients were set for each appliance to be arm-pointed. The coefficients in these equations were obtained empirically from θ data when the appliance was kept pointing by one person. Equation (4) and (5) show the relationship between them that determines the choice when the TV or humidifier is indicated.

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

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

III. EVALUATION

Experiments on recognition accuracy and usability were conducted to evaluate the method. We conducted experiments

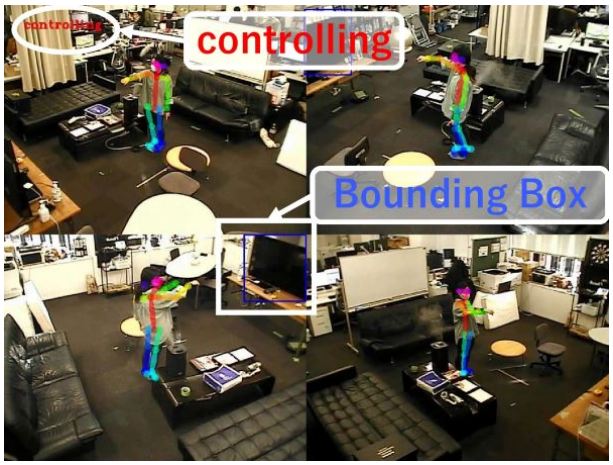


Fig. 5: Feedback screen

on recognition accuracy and usability on 10 subjects, who participated in the experiments after a brief explanation of the method.

In Experiment 1, we evaluated the recognition rate of gestures for switching the start and end of home appliance operation. In Experiment 2, we evaluated the recognition accuracy of operations to turn on and off the home appliance by pointing it with the arm. Moreover, the number of frames in which the selection state of the home appliance must be maintained continuously for the power operation of the home appliance was set to 6 frames. In both experiments, subjects were presented with a feedback screen as shown in Fig. 5. Finally, we evaluated the usability of this system using SUS (System Usability Scale[22]) in experiment 3.

Discussion from the obtained results is given in the next section.

A. Experiment 1

The combination of operations that switch the start and end of home appliance operations and their actions are as shown in Table I.

TABLE I: Combination of the operation and the gesture for switching

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

This operation was performed 10 times in the upright state at position A and B in Fig. 6, respectively, and if it was not recognized after at least 5 seconds, the operation was considered to have failed. The recognition rate of each operation for each position is shown in Table II. Significant figures should be two digits.

For the operation start gesture, the recognition rate at position A was 89%, at position B was 73%, and 81% on average. For the operation end gesture, the recognition rate at position A was 87%, at position B was 70%, and 79% on average.

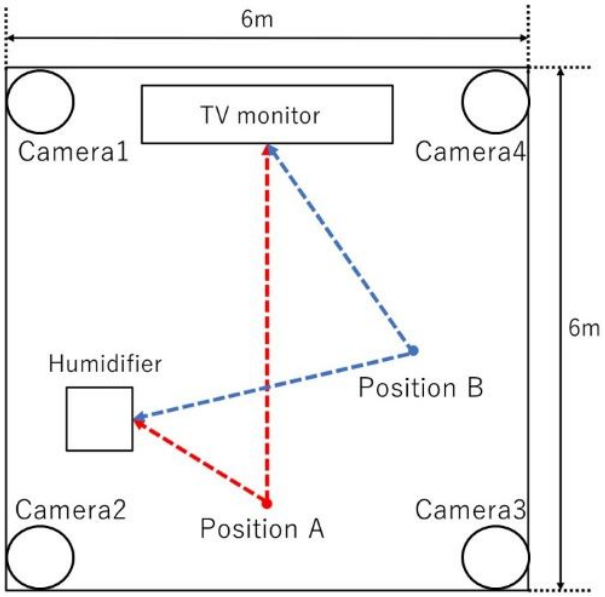


Fig. 6: Experimental environment viewed from above

TABLE II: The recognition rate for operation start and end switching operations [%]

	Position A	Position B	Average
Start	89	73	81
End	87	70	78.5

B. Experiment 2

The user pointed with his/her arm at the home appliance in both standing and seated positions, and if the home appliance was not recognized after at least 5 seconds, the operation was considered to have failed. This was done at position A and position B. The average recognition rate of each operation for each position is shown in Table III. Significant figures should be two digits.

TABLE III: The recognition rate of each operation for each position[%]

Status	Appliance	Position A	Position B	Average
Standing	TV	57	75	66
	Humidifier	95	72	83.5
Sitting	TV	34	61	47.5
	Humidifier	92	88	90

The recognition rate was 57% when the user operated the TV while standing at position A, and 75% at position B. The average recognition rate was 66%. The recognition rate was 34% when the user operated the TV while seated at position A, and 61% when the user operated the TV at position B. The average recognition rate was 48%. When the user operated the humidifier while standing at position A, the average recognition rate was 95%, and 72% at position B. The average recognition rate was 84%. Next, when the user operated the humidifier while seated at position A, the recognition rate was 92%, and 88% at position B. The average recognition rate was 90%. The average recognition

TABLE IV: 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

rate exceeded 50 percent in all three situations except for the TV operation while seated.

From these results, we found that the average recognition rate for the humidifier operation was 87%, while the average recognition rate for the TV operation was 57%.

C. Experiment 3

By analyzing the results of the questionnaire survey using SUS, the user's subjective satisfaction with the system or product can be evaluated and an overall satisfaction assessment can be made.

The questionnaire used a 5-point Likert scale[23], with 1 being "not at all disagree" and 5 being "completely agree." Subjects were asked to indicate which 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 5 from the response score 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 IV. The SUS score of this system is about 72.8.

IV. DISCUSSION

The results of the experiments are discussed in the following section. Experiments 1 and 2 are collectively referred to as the gesture recognition rate evaluation experiment, and Experiment 3 evaluates as the usability of the system.

A. Evaluation of Recognition Gestures

Results from experiment 1 suggest that the operation to switch the start and end of home appliance operation is correctly recognized with a high probability, which can prevent erroneous operation. On the other hand, the reason why the recognition rate at position B is slightly lower than that at position A is considered to be that the right forearm is sometimes hidden by the user's head depending on the height of the user as shown in Fig. 6 from Camera3, and is not recognized properly. This is thought to be because the right forearm was hidden by the user's head depending on the user's height.

Table III shows that the recognition rate of the TV power operation is considerably lower than that of the humidifie power operation. This may be because the center of the TV was more difficult for the user to indicate than the center of the humidifier. During the experiment, the subject frequently changed the direction of the TV when pointing at the TV. Therefore, it is thought that the recognition rate decreased.

In both standing and sitting situations, the recognition rate of the TV instructions was lower when the user was seated than when the user was standing. This may be because the vector $\vec{v_p}$ from the wrist to the elbow and the vector $\vec{v_a}$ from the elbow to the home appliance fluctuates, and the angle θ is out of the instruction judgment range, even if the user intends to indicate the same point by moving the eye line up and down.

In the future, it is necessary to set judgment conditions for appliance instructions that are less affected by changes in line of sight, such as standing or sitting. In addition, it is necessary to devise how to avoid misidentification of appliance selection and how to perform correction operations when the misidentification occurs. A possible method for avoiding misrecognition of home appliance selection is to estimate the operations the user wants to perform based on the user's body posture and the scene, and select home appliances accordingly. Furthermore, a possible method for correcting misrecognition of home appliance selection is to shift the selection state to the nearest neighboring home appliance after canceling the selection state by hand gestures.

Finally, we believe that in order to improve recognition accuracy, it is important not only to improve the operation method, but also to improve the feedback presented to the user during operation to make it easier for the user to understand the operation status. This could be achieved not only by showing a feedback screen as is currently done, but also by changing the color of the smart remote control between the home appliance selection state and the state in which the power operation is recognized.

B. Evaluation of System's Usability

Sauro derived an average SUS score of 68.1 points from data revealing the relationship between SUS and percentiles based on more than 5,000 SUS score measurements[24]. The SUS score for this method was found to be 72.8, which is higher than the 68.1 score, indicating that the system has excellent usability.

The only home appliance operation that can be performed with this system is the power supply operation. It is easy to imagine that users will feel annoyed if the number of gesture types and procedures increases in order to perform more diverse operations in the future. Therefore, in order to maintain high usability while using many functions, it is necessary to introduce gestures that are more natural and easy to continue to perform, and operation methods that are

easy to understand for users who are using the system for the first time.

V. CONCLUSIONS

In this study, we proposed a system to control home appliances that uses simple gestures that are easy to remember, and does not require prior information on the location and size of the appliances. The system allows the user to turn home appliances on and off by continuously indicating the appliance with his/her forearm. The system also can switch the start and end of home appliance operation to prevent accidental operation. The proposed method was evaluated through experiments.

These days, robot vacuum cleaners that clean automatically and robots that are designed to care for personal needs of users to reduce their physical burden have been developed and are gradually becoming popular. Therefore, this method can be used to interact with robots. In this study, the target of manipulation was home appliances at arbitrary stationary positions, but in the future, we would like to study a teleoperation method using physical instructions for machines that move continuously.

REFERENCES

[1] J. R. B. Bodollo, J. Daniel V. Cortez, E. R. P. Maraya, E. V. Navarro, R. Q. L. Saquing and R. E. Tolentino, "Selection of Appliance Using Skeletal Tracking and 3D Face Tracking for Gesture Control Home Automation," 2019 1st International Conference on Advanced Technologies in Intelligent Control, Environment, Computing & Communication Engineering (ICATIECE), Bangalore, India, pp. 1-7, 2019.

[2] H. Asano, T. Nagayasu, T. Orimo, K. Terabayashi, M. Ohta and K. Umeda, "Recognition of finger-pointing direction using color clustering and image segmentation," The SICE Annual Conference 2013, Nagoya, Japan, pp. 2029-2034, 2013.

[3] A. I. D. Viaje, P. S. Bernardo, K. N. Manuel, G. M. Pacheco, K. -R. C. Barroma and R. E. Tolentino, "Selection of Appliance Using Skeletal Tracking of Hand to Hand-tip for a Gesture Controlled Home Automation," 2020 International Conference on Electronics and Sustainable Communication Systems (ICESC), Coimbatore, India, pp. 575-580, 2020.

[4] 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 Symposium (SAS), Zadar, Croatia, pp. 1-6, 2015.

[5] 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 International Conference on Informatics, Electronics and Vision (ICIEV), Dhaka, Bangladesh, pp. 891-896, 2016.

[6] A. Tsagaris, S. Manitsaris, E. Hatzikos and A. Manitsaris, "Methodology for finger gesture control of mechatronic systems," Proceedings of 15th International Conference MECHATRONIKA, Prague, Czech Republic, pp. 1-6, 2012.

[7] M. S. Verdadero, C. O. Martinez-Ojeda and J. C. D. Cruz, "Hand Gesture Recognition System as an Alternative Interface for Remote Controlled Home Appliances," 2018 IEEE 10th International Conference on Humanoid, Nanotechnology, Information Technology,Communication and Control, Environment and Management (HNICEM), Baguio City, Philippines, pp. 1-5, 2018.

[8] S. Kaur, Anuranjana and N. Nair, "Electronic Device Control Using Hand Gesture Recognition System for Differently Abled," 2018 8th International Conference on Cloud Computing, Data Science & Engineering (Confluence), Noida, India, pp. 371-375, 2018.

[9] X. Zhang and X. Wu, "Robotic Control of Dynamic and Static Gesture Recognition," 2019 2nd World Conference on Mechanical Engineering and Intelligent Manufacturing (WCMEIM), Shanghai, China, pp. 474-478, 2019.

[10] Y. Muranaka, M. Al-Sada and T. Nakajima, "A Home Appliance Control System with Hand Gesture based on Pose Estimation," 2020 IEEE 9th Global Conference on Consumer Electronics (GCCE), Kobe, Japan, pp. 752-755, 2020.

[11] T. -H. Tsai, Y. -J. Luo and W. -C. Wan, "Live Demonstration: Home Appliance Control System with Dynamic Hand Gesture Recognition base on 3D Hand Skeletons," 2022 IEEE 4th International Conference on Artificial Intelligence Circuits and Systems (AICAS), Incheon, Korea, Republic of, pp. 503-503, 2022.

[12] R. A. Urmee, N. S. Prome and T. Ahmed, "Hand Gesture-Based Home Automation System," TENCON 2022 - 2022 IEEE Region 10 Conference (TENCON), Hong Kong, Hong Kong, pp. 1-5, 2022.

[13] 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 International Symposium on System Integration (SII), Honolulu, HI, USA, pp. 366-370, 2020.

[14] M. Niitsuma, H. Kobayashi and A. Shiraishi, "Enhancement of Spatial Memory for Applying to Sequential Activity", Journal of Advanced Sciences, vol. 9, no. 1, pp. 121-137, 2012.

[15] A. Dongre, R. Pinto, A. Patkar and M. Lopes, "Computer Cursor Control Using Eye and Face Gestures," 2020 11th International Conference on Computing, Communication and Networking Technologies (ICCCNT), Kharagpur, India, pp. 1-6, 2020.

[16] B. Singh, N. Kandru and M. Chandra, "Application control using eye motion," 2014 International Conference on Medical Imaging, m-Health and Emerging Communication Systems (MedCom), Greater Noida, India, pp. 206-210, 2014.

[17] J. Redmon, S. Divvala, R. Girshick and A. Farhadi, "You Only Look Once: Unified, Real-Time Object Detection," 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, pp. 779-788, 2016.

[18] Z. Cao, G. Hidalgo, T. Simon, S. -E. Wei and Y. Sheikh, "OpenPose: Realtime Multi-Person 2D Pose Estimation Using Part Affinity Fields," in IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 43, no. 1, pp. 172-186, 2021.

[19] T. Lin, M. Maire, S. Belongie, L. Bourdev, R. Girshick, J. Hays, P. Perona, D. Ramanan, C. L. Zitnick, P. Dollár, "Microsoft COCO: Common Objects in Context," 2014.

[20] A. Bochkovskiy, C. Wang, H. Mark Liao, "YOLOv4: Optimal Speed and Accuracy of Object Detection," arXiv preprint arXiv:2004.10934, 2004.

[21] R. Hartley, A. Zisserman, "Multiple View Geometry in Computer Vision," Cambridge University Press, second edition, 2004.

[22] J. Brooke, "SUS–a quick and dirty usability scale," Usability Evaluation in Industry, pp. 189-194, 1996.

[23] R. Likert, "A Technique for the Measurement of Attitudes" Archives of Psychology,22 140, 55, 1932.

[24] J. Sauro, "A Practical Guide to the System Usability Scale," Measuring Usability LLC, 2011.