IEEE Access
Multidisciplinary ⁝ Rapid Review ⁝ Open Access Journal

Received January 1, 2021, accepted January 7, 2021, date of publication January 13, 2021, date of current version January 27, 2021.

Digital Object Identifier 10.1109/ACCESS.2021.3051253

Smart Living: An Interactive Control System for Household Appliances

MU-CHUN SU[1], JIEH-HAUR CHEN[2,3], ACHMAD MUHYIDIN ARIFAI[2], SUNG-YANG TSAI[1], AND HSI-HSIEN WEI[4]

[1]Department of Computer Science and Information Engineering, National Central University, Jhongli 32001, Taiwan
[2]Department of Civil Engineering, National Central University, Jhongli 32001, Taiwan
[3]Research Center of Smart Construction, National Central University, Jhongli 32001, Taiwan
[4]Department of Building and Real Estate, The Hong Kong Polytechnic University, Hong Kong

Corresponding author: Jieh-Haur Chen (jhchen@ncu.edu.tw)

This work was supported by the Ministry of Science and Technology (MOST), Taiwan, through the Promoting Academic Excellent of Universities Program, under Grant MOST 109-2622-E-008-018-CC2, Grant MOST 109-2221-E-008-059-MY3, Grant MOST 109-2634-F-008-007, Grant MOST 108-2221-E-008-002-MY, and Grant MOST 107-2221-E-008-084-MY2.

ABSTRACT Hand-gesture based control has enormous potential both theoretically and for practical applications due to its convenience and intuitiveness. This study presents a real-time interactive control system for household appliances. The interactive control system allowing wireless control of household appliances using a combination of 11 hand gestures and 2 waving motions is tested on hundreds of samples. It is implemented using a regular personal computer (PC) and existing digital single processing (DSP) platforms. The evaluation results show that the system performs efficiently reaching an accuracy recognition rate of 91% and spending around 30 seconds to complete the control operation for household appliances. The contributions of this work are both academic (1) successful demonstration of the integration of algorithms for solving image detection, processing, and pattern recognition, and practical (2) showing its feasibility and using commonly available hardware and software configurations for practical uses, and finally (3) establishing a mechanism for intuitively interactive control system that facilitates smart living.

INDEX TERMS Smart home, human computer interaction, hand gesture recognition, household appliance control.

I. INTRODUCTION

A smart home usually refers to a house equipped with high-tech devices, appliances, sensors, and networks. These tools increase the level of convenience to the people living in the home by allowing them to remotely access, control, and monitor home appliances through smart technologies [1], [2]. Smart home services have not been greatly expanded because the technologies needed to enable these services have not been adequately developed [3]. In addition, these services are often not familiar and/or not seen as consumer-friendly so most studies remain in the development phases. As a result, consumers have not yet embraced smart home technologies and services in order to achieve smart living. Uncertainty about these technologies and services can arise from number of concerns regarding cost, performance, and even lifestyle changes [4]. The old push-button remote controls may be convenient, but have has certain disadvantages such as the need for batter replacement and damage to devices due to battery defects. Improvement of the systems for control of household appliances to produce a more user friendly interface has become a definitive goal in multimedia and human computer interaction research [5]. In particular, the recognition of hand gestures utilizing camera-enabled devices would allow for a more intuitive interaction than traditional interfaces do. The application of embedded and ubiquitous computing in household environments should allow the human-computer interaction to be more natural, convenient, and efficient [6]. Hand gesture based interfaces can decrease the complexity of interaction between humans and computers [7], [8]. Therefore, the research objective in this work is to establish a real-time interactive control system for household appliances, integrating algorithms for image detection and processing with pattern recognition. The interactive control system provides a user-friendly interface for commonly used devices that does not require or restrict the users to wielding

The associate editor coordinating the review of this manuscript and approving it for publication was Liang-Bi Chen.

specific control devices or wearing special clothing, and presenting specific gestures or positions.

II. HAND GESTURE RECOGNITION

There are two main categories of hand gestures, static and dynamic. Static gestures refer to fixed hand gestures which can be described by an image; dynamic gestures on the other hand, are composed of multiple static gestures. The recognition methods include typical branching, inductive and image-based techniques. The former captures the subjects' movements but requires them to wear gloves equipped with acceleration sensors or latent pressure sensors. The latter adopts cameras to capture images for recognition. There are three major three image-based gesture recognition systems, depending upon the number of cameras used: the Multi-viewpoint Camera system, using two cameras (Stereo Cameras), and the single Monocular Camera system [9]–[16]. Information from multiple cameras may be used to create 3D models of the shot. Two cameras are able to yield depth information. Although recognition is better with more cameras, complex algorithms may be required, causing longer calculation times and higher costs.

Common gesture recognition methods include the hidden Markov model, dynamic time warping neural network and finite-state machine techniques [12]–[21]. In one study researchers placed four cameras to capture hand images from different directions, followed by processing of the finger link structure and 3D surface structure. The experimental results showed that 69% of the connection errors were less than 1 cm [11]. Another study captured hand images with a single camera, resulting in reduced dimension feature vectors [14]. A recent study applied hand gesture recognition for the operation of a TV. Skin color detection was used to extract the shape of the hand and a moment invariant method applied to describe the characteristics. The results showed that this multi-layer perceptron method could identify six hand shapes with 100% accuracy [22].

III. METHODOLOGY

The structure of the interactive control system for household appliances is shown in Figure 1. The system structure starts with determining whether any subject (e.g. arm) is detected, followed by checking whether there are enough skin pixels in the region of interest (ROI). The procedure goes back to skin color detection for image recognition and morphological calculations if no subject is detected. The labeling algorithm removes small objects in order to identify the arm part. After segmentation of the arm image, the contour of the hand is revealed, followed by measurement of the distance curve from the contour to the centroid using the fast Fourier transform frequency feature. Finally, identification of the hand gesture can be achieved using k-nearest neighbor (kNN) classification and the indices from the decision tree method.

The interactive control system can be set up for use with a both personal computer (PC) and demand-side platform (DSP). The PC version uses a general home network

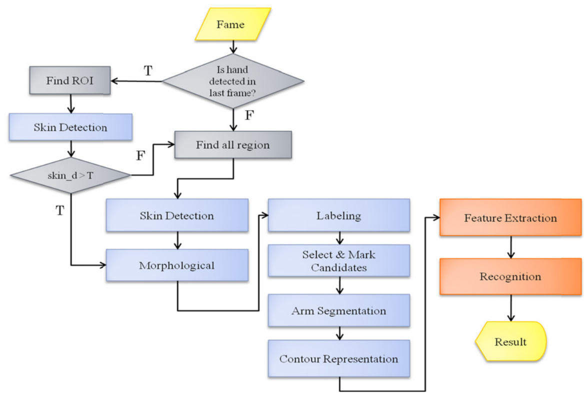


FIGURE 1. System structure.

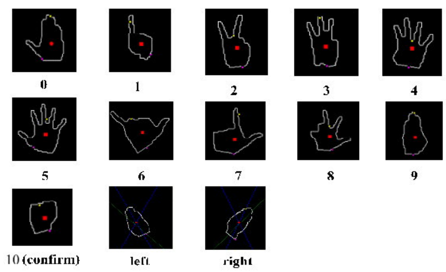


FIGURE 2. 11 hand shapes and two waving gestures.

camera and the 320×240 RGB image input format. The DSP version needs to be equipped with separate-video (S-video), a signaling standard for standard definition video, or composite terminal special connected cameras and 1024×768 coordinate transformation associated RGB color space with a 4:2:2 (YCbCr422) video input format. The 10 most common types of hand gestures (0 - 9) identified by the system are shown in Figure 2. Each shape represents a specific gesture for outputting a digital hand shape, plus two left and right waving gestures.

Starting with hand shape detection, the YCbCr color space produces better recognition results for skin color [23]–[26] as follows:

$$Skin = \begin{cases} 1, & if \begin{cases} 77 < Cb < 127 \\ 133 < Cr < 173 \\ 190 < Cb + 0.6Cr < 215 \end{cases} \\ 0, & else \end{cases} \tag{1}$$

$$Skin = \begin{cases} 1, & if \begin{cases} 60 < Y < 255 \\ -25 < Cb < 0 \\ 10 < Cr < 45 \end{cases} \\ 0, & else \end{cases} \tag{2}$$

$$Skin = \begin{cases} 1, & if \begin{cases} 60 \leq Y \leq 250 \\ 90 \leq Cb \leq 135 \\ 135 \leq Cr \leq 170 \end{cases} \\ 0, & else, \end{cases} \tag{3}$$

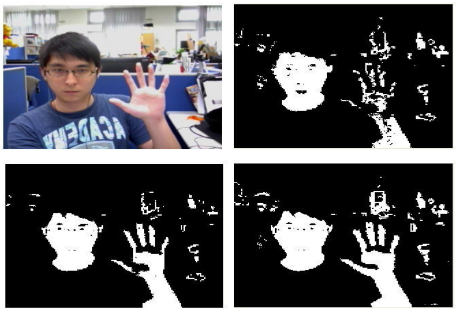


FIGURE 3. Hand detection: upper right from Equation (1); lower left from Equation (2); lower right from Equation (3).

where Y is the luma component, Cb represents the blue-difference component, and Cr stands for the red-difference component. As can be seen in Figure 3 shows Equation (3) yields the best results for hand shape that is adopted for the interactive control system. For better processing of images, the dilation and opening algorithms with 5×5 closing and opening are then adopted to make the images smoother.

Equation (4) eliminates all irrelevant binary pixels and calculates the height, width, and area of the detected subject [25].

$$Arm = \begin{cases} 1, & else \\ 0, & \begin{cases} area < 1800 or area > 60000 \\ \dfrac{height}{width} < 0.2 or \dfrac{height}{width} > 5 \\ height < 30 or width < 30 \end{cases} \end{cases} \tag{4}$$

where *Arm* stands for the recognition results for the human arm(s) or any other object; *height, width,* and *area* represent the object's height, width, and area to be recognized. It is possible that the subject might not be an arm or misclassification could occur. The interactive control system is comprised of two steps to prevent just such situations which are shown as follows:

Step 1:

$$Hand = MAX(\bar{y}_1, \bar{y}_2) \tag{5}$$

Step 2:

$$Face = \begin{cases} 1, & if \begin{cases} 31 \le \overline{Y} \le 69 \\ 123 \le \overline{Cb} \le 133 \\ 123 \le \overline{Cr} \le 133 \end{cases} \\ 0, & else \end{cases} \tag{6}$$

where *Hand* and *Face* are the target objects for recognition; $\overline{y_1}$ and $\overline{y_2}$ are the y-axes for the palm of the hand on a plane. A face is detected if the $\overline{Y}, \overline{Cb}$, and $\overline{Cr}$ values are within the specified ranges where $\overline{Y}$ is the value range for luma component; $\overline{Cb}$ represents the value range for the blue-difference component; $\overline{Cr}$ stands for the value range for the red-difference component. These steps are efficient enough to distinguish an arm from a face or other subject using the threshold values in Equation (6).

Next, we develop an algorithm which is designed to remove the arm part of the image(s) but keep the image section for the palm(s) for the next recognition stage, as in Equations (7)-(10). In order to cut the arm images, use Equations (7) and (8) to obtain the raw and central moments for the block direction and Equation (9) for the image coordinates:

$$M_{ij} = \sum_x \sum_y x^i y^j f(x, y) \tag{7}$$

$$\begin{aligned} \mu_{20} &= \sum_x \sum_y (x - \bar{x})^2 f(x, y) \\ \mu_{11} &= \sum_x \sum_y (x - \bar{x})(y - \bar{y}) f(x, y) \\ \mu_{02} &= \sum_x \sum_y (y - \bar{y})^2 f(x, y), \end{aligned} \tag{8}$$

$$where\ \bar{x} = \frac{M_{10}}{M_{00}}, \bar{y} = \frac{M_{01}}{M_{00}}$$

$$\emptyset = (1/2) \tan^{-1}\left(\frac{2\mu_{11}}{\mu_{20} - \mu_{02}}\right), \tag{9}$$

where M_{ij} is the raw moment of a multivariable probability function $f(x,y)$; μ stands for the central moment using the inverse binomial transform; $\emptyset$ represents the image (x,y) coordinates. The edge capture component needs to determine whether a pixel is an edge by defining a pixel P_0 (i, j) and its 4 neighbors P_1 $(i\text{-}1, j)$ at the up side, P_2 $(i, j+1)$ at the right side, P_3 $(i+1, j)$ at the down side, and P_4 $(i, j\text{-}1)$ at the left side; and by assuming that the previous background pixel value is 1 and the current background pixel value is 0. Equation (10) expresses whether P_0 is an edge.

$$P_0 = \begin{cases} 1, & if\ P_1 * P_2 * P_3 * P_4 = 0 \\ 0, & else \end{cases} \tag{10}$$

The image outline is determined following the 4 steps below. Having obtained the result from edge capture for hand images and selecting any point on the edge that has the coordinates (i, j), then we can: (1) select one edge point as $(start_i, start_j)$ and $(step_i, step_j)$; (2) search for neighbors starting with vector$\overrightarrow{dir}_{1-8}$ in a clockwise direction and update $(step_i, step_j)$ when meeting any edge point; (3) let $\overrightarrow{dir}+5$ be the next initial search point; (4) repeat (2) until $(start_i, start_j) = (step_i, step_j)$ or no adjacent edge point is found.

Feature extraction is carried out utilizing the distance from the edge point to the centroid as a feature which is normalized to a 512-point feature length. This feature is scale rotation invariant. By converting it to the frequency domain and using the Fast Fourier Transform (FFT), we sample all 512 points. Since the number of human fingers per palm is always less than six, only the first six frequencies (F_1-F_6) need to be considered and described as in Equations (11-14):

$$F = \frac{\omega \cdot F_s}{2\pi} \tag{11}$$

$$\omega = \frac{2\pi \cdot F}{F_s} = \delta \frac{2\pi}{N} \tag{12}$$

where $\omega = \frac{2\pi \cdot \frac{6}{512}}{1}$ due to $\delta = \frac{\omega \cdot N}{2\pi} = \frac{\frac{12\pi}{512} \cdot 512}{2\pi} = 6$.

Here, F, F_s, and ω are the analog frequency, sample frequency, and digital frequency; δ stands for the FFT hierarchical number and N is the frame size. This significantly reduces the computing time for comparison among peaks (fingers).

There are 11 hand shapes and 2 waving gestures recognized in this study (Figure 2). For the waving gestures, Equation (13) is applied to examine to which side the wave image belongs and Equation (14) is used to detect whether the arm is moving. Equation (15) is a combination of Equations (13) and (14) expressed as follows:

$$Motion_n = \begin{cases} Left, & if\ m_H < m_L\ and\ m_H > 0 \\ Right, & if\ m_H > m_R\ and\ m_H < 0 \\ Stand, & else \end{cases} \tag{13}$$

$$Wave = \begin{cases} True, & if\, Motion_{n-1} = Stand\ and\ Motion_n! = Stand \\ False, & esle \end{cases} \tag{14}$$

$$Waving\ Gesture = \begin{cases} WaveLeft, & if\, Motion_n = Left\ and\ Wave = True \\ WaveRight, & if\, Motion_n = RightandWave = True \end{cases} \tag{15}$$

where $Motion_n$ is the n^{th} arm motion detected by the system; m_H represents the axis and slope for the arm; m_L and m_R are the values for an arm moving toward the left or right. The interactive control system utilizes the kNN enhanced decision tree method for analysis of the 11 hand shapes, as illustrated in Figure 4. The k in kNN represents the number of neighbors to be referred to while the neighbors are the first k samples nearest to them, using the general Euclidean distance. Among these k neighbors, the class with the highest number is regarded as the type to which the current sample belongs. For example, given k = 10, we find the closest distance to the first ten feature distances for the sample and then determine the class with the highest number among these.

The kNN is a simple and effective tool to detect the fingers of the human palm; however, it may be insufficient, leading misclassification of the shape of the palm due to multiple-axes from the center of the palm to the fingers. The decision tree (DT) technique is another simple and effective method to resolve the afore-mentioned problem. As a result, the kNN method is used only for distinguishing peaks/fingers (left side in Figure 4) and then DT can find the palm shapes/gesture based on the peak number yielded from kNN approach. We design the mechanism similar to "if…then…else" mechanism but gestures 0-10 is determined pair by pair (except for gesture 10) as shown in Figure 4, which basically comes from the DT concept. The steps in the process for finding the palm shapes using the DT method are illustrated on the lower right of Figure 4 and demonstrated below. For example, the process starts with Equation (16) where the input (independent variable) is set to the normalized distance measured from edge to center and the output (dependable variable) is set to Gesture 0 or 6. For the rest equations (17-20), the DT mechanism is similar and the process stops when all inputs are recognized.

$$Identify0\_6(d) = \begin{cases} gesture\ 0, & if\ d \le 3.5 \\ gesture\ 6, & \text{else} \\ where d = |p_1 - \bar{p}| + |p_2 - \bar{p}| \\ \quad + |p_3 - \bar{p}| \\ \bar{p} = \dfrac{p_1 + p_2 + p_3}{3} \end{cases} \tag{16}$$

$$Identify2\_7(\theta_{12}) = \begin{cases} gesture\ 2, & if\ \theta_{12} < 55 \\ gesture\ 7, & if\ 55 < \theta_{12} < 105 \\ unknown, & else \end{cases} \tag{17}$$

$$Identify3\_8(\hat{\theta}) = \begin{cases} gesture\ 3, & if\ 0 \le \hat{\theta} < 80 \\ gesture\ 8, & if\ 80 \le \hat{\theta} < 120 \\ unknown, & else \end{cases} \tag{18}$$

$$p_6 > D\,(\text{i}\,(p_6) - 20)\ \text{ and } p_6 > D\,(\text{i}\,(p_6) + 20) \tag{19}$$

$$Identify\ 4\_5(c(p)) = \begin{cases} gesture\ 4, & if\ c(p) = 5 \\ gesture\ 5, & if\ c(p) = 6 \end{cases} \tag{20}$$

where d is the difference; $p_{1\text{-}6}$ indicates the 6 peak values for the fingers; θ stands for the angle between two fingers; $\hat{\theta}$ = MAX(θ_{12}, θ_{23}, θ_{13}); D represents the distance curve; $i(p_6)$ indicates the offset of the 6^{th} peak in D; $c(p)$ means the peak count. For example, the peak number for gesture 4 or 5 must be greater than or equal to 5. Calculating the distances for the sixth highest peak, we plug the result into Equation (19) that is set to determine gesture 4 or 5. If p_6 satisfies Equation (19) which is designed to detect any further significant peak, the gesture can be 4 or 5. Then, Equation (20) performs the peak count to determine the final gesture type.

In order to make the system more portable and easy-to-install in most houses, it is implemented based on an embedded system with a digital signal processor. The specific hardware and instructional design has the advantages of being easily programmable, with high accuracy and can be easily integrated with a SoC. The interactive control system offers better performance than the traditional single-chip processor.

IV. EVALUATION AND DISCUSSION

The evaluation was carried out in three phases: (1) analysis of recognition rates with the first-stage kNN method; (2) analysis of recognition rates with the integrated kNN decision tree method; and (3) implementation with household appliances. Considering device accessibility and availability in certain regions where this may be difficult, we set up a basic

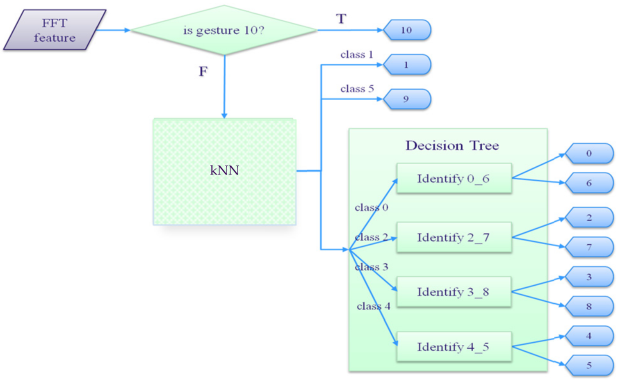


FIGURE 4. Identification process.

TABLE 1. Phase I evaluation results for kNN recognition rates.

Set #	k =	1	3	4	5	6
100	Correct/total	96/100	95/100	95/100	94/100	94/100
	Successful recognition rate	96%	95%	95%	94%	94%
400	Correct/total	383/400	381/400	382/400	379/400	378/400
	Successful recognition rate	96%	95%	96%	95%	95%

prototype configuration including a board with 512Kb flash, 1K SRAM, 2 by 8-bit timers, 512Kb EEPROM (546 channels), 16-bit timers,10-bit A/D single-ended, 20 MHz F. max, 1.8-5.5V Vcc, AM24L01BS-U RF transmission module with max. 30 bit packets, etc. The evaluation started with data collection of 100 random sets containing hand characteristics for evaluation in Phase I; 400 sets of hand gestures for evaluation in Phase II. For faster computation with Equations (11) and (12), k is set to 6, since the number of fingers per palm is always less than 6. Table 1 summarizes the Phase 1 results. The recognition success rates are all greater than 94% where the selection of the k value does not influence the outcomes. Although, as can be seen in Table 1, the best results are obtained when $k = 1$, the other results for $k = 1$-6 show that kNN is an effective method. It is meaningful and effective. The selection of the k value can be simplified so we set the value to 3 for quicker computation in Phase II.

Phase II also requires data collection of 440 random sets of hand gestures with 2/3 of the total being right hand gestures. Table 2 shows the results of the Phase II evaluation. The kNN-decision tree system has a slightly higher successful recognition rate for right hand gestures, 92%, than for left hand gestures, 89%. The average recognition rate accuracy is 91% with the highest being 100% and the lowest 82.5%. Figure 5 summarizes the correct classification and misclassification results for hand gestures. Gesture 3 has the highest misclassification rate due to its similarity to Gestures 0 and 2.

TABLE 2. Phase II evaluation result for the kNN-decision tree.

		Gesture											
		0	1	2	3	4	5	6	7	8	9	10	Accuracy rate
Gesture	0	**37**						2	1				92.5%
	1		**37**					3					92.5%
	2			**40**									100%
	3	4		1	**33**					2			82.5%
	4				1	**36**	2			1			90%
	5					3	**36**			1			90%
	6	5						**35**					87.5%
	7	1						4	**35**				87.5%
	8				1				1	**38**			95%
	9							1			**39**		97.5%
	10										6	**34**	85%
Correct/total		400/440											91%

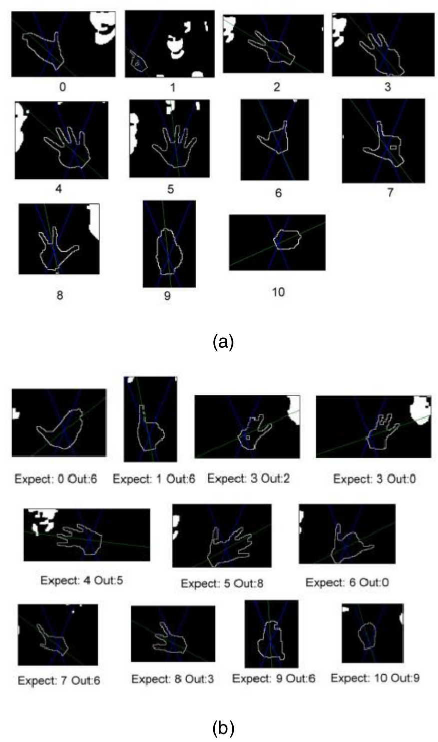


FIGURE 5. Illustration of hand gestures: (a) correct classification and (b) misclassification.

Gesture 10 also has a relatively weak recognition rate because of its similarity to Gesture 9.

The last phase is to implement the system for the control of household appliances, selecting the most common appliances and operation modes. The feasibility of the interactive control system is evaluated and measured based on the action time. A TV and fan were selected for Phase III evaluation. The common scenarios for controlling the TV and the fan are described in Figure 6. The designed gestures are universal for household appliances. "Gesture 0, waving to the right" represents "Turn on" and "Gesture 0, waving to the left"

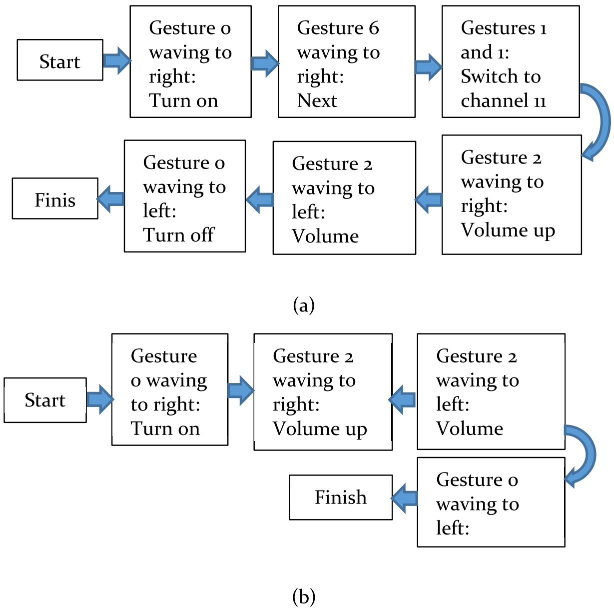
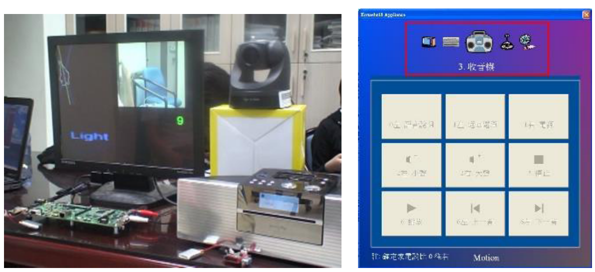


FIGURE 6. Illustration of the household appliance controls for: (a) TV and (b) fan.



(a) (b)

FIGURE 7. Hardware (a) and interface (b) for the system.

represents the command "Turn off". Gesture 2, waving to the right for "up" and to left for "down" is used for volume control. Simple hand gestures are designed for channel numbers.

Figure 7 shows the (a) hardware used for evaluation and (b) a sample of the interface for the interactive control system. Table 3 records the time taken to complete the operation illustrated in Figure 6. The average operation time for both appliances is similar, around 30 seconds, even though the evaluation scenario for controlling the TV has two more steps than for the fan. Compared with recent work which usually examines only parts of the overall process, especially in terms of concept, interface, cost, image capture, and image process [20]–[26], the interactive control system is innovative since there are 20 equations in the study that combines the concepts of image processing, Fast Fourier Transform, motion detection, kNN, and modified DT from YCbCr color recognition for human skin to ultimate gesture recognition. Its successful demonstration of the integration of algorithms needed for solving the image detection and pattern recognition problems is beneficial to academic. The other practicable contribution lies in the illustration for hardware and software configuration need for practical applications from device specifications and setups to the interactive interface.

TABLE 3. Phase III evaluation results.

Subject	TV (seconds)	Fan (seconds)
1	24	24
2	24	37
3	36	25
4	33	23
5	39	30
6	30	20
Average	31	26.5

V. CONCLUSION

A gesture-based human-computer interaction system is developed for the control of household appliances using 11 gestures and waving to the right and to the left. The system utilizes real-time computation using both a PC and DSP. Testing is carried out with hundreds of samples and evaluation scenarios. The system performs efficiently reaching a recognition rate accuracy of 91% and spending around 30 seconds to complete the operation for household appliances. Even though the test samples and subjects are similar in shape and body language, the system functions well with commonly available hardware and software settings. The contributions of this study are clear including a (1) successful demonstration of the integration of algorithms needed for solving the image detection, processing, and pattern recognition problems; (2) a demonstration of the feasibility of the system using commonly available hardware and software configuration for practical use; and (3) establishing a mechanism for an intuitively interactive control system that facilitates smart living.

The interactive control system requires only basic extra devices to work with household appliances. It can effectively eliminate the need for several remote controls in home. Although these devices are widely available, there is room for improvement. It is suggested that future studies can be carried out to upgrade the interactive control system to a Universal Plug and Play (UPnP) network protocol that may require a few more steps to construct. The settings for the prototype configuration need to be upgraded due to improvements and availability of the technology. In addition, the use of gestures with a higher accuracy rate and functions with a higher frequency is suggested for more effective control.

REFERENCES

[1] N. Balta-Ozkan, R. Davidson, M. Bicket, and L. Whitmarsh, "Social barriers to the adoption of smart homes," *Energy Policy*, vol. 63, pp. 363–374, Dec. 2013.

[2] T. Hatuka and H. Zur, "From smart cities to smart social urbanism: A framework for shaping the socio-technological ecosystems in cities," *Telematics Informat.*, vol. 55, Dec. 2020, Art. no. 101430.

[3] A. Benlian, J. Klumpe, and O. Hinz, "Mitigating the intrusive effects of smart home assistants by using anthropomorphic design features: A multimethod investigation," *Inf. Syst. J.*, vol. 30, no. 6, pp. 1010–1042, Nov. 2020.

[4] J. Hong, J. Shin, and D. Lee, "Strategic management of next-generation connected life: Focusing on smart key and car–home connectivity," *Technol. Forecasting Social Change*, vol. 103, pp. 11–20, Feb. 2016.

[5] H. Dong, A. Danesh, N. Figueroa, and A. E. Saddik, "An elicitation study on gesture preferences and memorability toward a practical hand-gesture vocabulary for smart televisions," *IEEE Access*, vol. 3, pp. 543–555, 2015.

[6] P. Mertala, "Young children's perceptions of ubiquitous computing and the Internet of Things," *Brit. J. Educ. Technol.*, vol. 51, no. 1, pp. 84–102, Jan. 2020.

[7] T. Vuletic, A. Duffy, L. Hay, C. McTeague, G. Campbell, and M. Grealy, "Systematic literature review of hand gestures used in human computer interaction interfaces," *Int. J. Hum.-Comput. Stud.*, vol. 129, pp. 74–94, Sep. 2019.

[8] T. Wang, "Intelligent employment rate prediction model based on a neural computing framework and human–computer interaction platform," *Neural Comput. Appl.*, vol. 32, no. 21, pp. 16413–16426, Nov. 2020.

[9] H. V. P. Singh and Q. H. Mahmoud, "ViDAQ: A computer vision based remote data acquisition system for reading multi-dial gauges," *J. Ind. Inf. Integr.*, vol. 15, pp. 29–41, Sep. 2019.

[10] B. Li, J. P. Munoz, X. Rong, Q. Chen, J. Xiao, Y. Tian, A. Arditi, and M. Yousuf, "Vision-based mobile indoor assistive navigation aid for blind people," *IEEE Trans. Mobile Comput.*, vol. 18, no. 3, pp. 702–714, Mar. 2019.

[11] Y. Kim and D. Kim, "A CNN-based 3D human pose estimation based on projection of depth and ridge data," *Pattern Recognit.*, vol. 106, Oct. 2020, Art. no. 107462.

[12] E. J. E. Cardenas and G. C. Chavez, "Multimodal hand gesture recognition combining temporal and pose information based on CNN descriptors and histogram of cumulative magnitudes," *J. Vis. Commun. Image Represent.*, vol. 71, Aug. 2020, Art. no. 102772.

[13] P. Szczuko, "Deep neural networks for human pose estimation from a very low resolution depth image," *Multimedia Tools Appl.*, vol. 78, no. 20, pp. 29357–29377, Oct. 2019.

[14] W. Yang, D. Yang, Y. Liu, and H. Liu, "EMG pattern recognition using convolutional neural network with different scale signal/spectra input," *Int. J. Humanoid Robot.*, vol. 16, no. 4, Aug. 2019, Art. no. 1950013.

[15] S. F. Chevtchenko, R. F. Vale, and V. Macario, "Multi-objective optimization for hand posture recognition," *Expert Syst. Appl.*, vol. 92, pp. 170–181, Feb. 2018.

[16] W. Wang and A. C. den Brinker, "Modified RGB cameras for infrared remote-PPG," *IEEE Trans. Biomed. Eng.*, vol. 67, no. 10, pp. 2893–2904, Oct. 2020.

[17] J. Kong, B. Zan, and M. Jiang, "Human action recognition using depth motion maps pyramid and discriminative collaborative representation classifier," *J. Electron. Imag.*, vol. 27, no. 3, May 2018, Art. no. 033027.

[18] Y. Zhang, L. Shi, Y. Wu, K. Cheng, J. Cheng, and H. Lu, "Gesture recognition based on deep deformable 3D convolutional neural networks," *Pattern Recognit.*, vol. 107, Nov. 2020, Art. no. 107416.

[19] S. Ameur, A. B. Khalifa, and M. S. Bouhlel, "A novel hybrid bidirectional unidirectional LSTM network for dynamic hand gesture recognition with leap motion," *Entertainment Comput.*, vol. 35, Aug. 2020, Art. no. 100373.

[20] B. Vandersmissen, N. Knudde, A. Jalalvand, I. Couckuyt, T. Dhaene, and W. De Neve, "Indoor human activity recognition using high-dimensional sensors and deep neural networks," *Neural Comput. Appl.*, vol. 32, no. 16, pp. 12295–12309, Aug. 2020.

[21] Z. Zhang, C. He, and K. Yang, "A novel surface electromyographic signal-based hand gesture prediction using a recurrent neural network," *Sensors*, vol. 20, no. 14, p. 3994, Jul. 2020.

[22] M. Tan, J. Zhou, K. Xu, Z. Peng, and Z. Ma, "Static hand gesture recognition with electromagnetic scattered field via complex attention convolutional neural network," *IEEE Antennas Wireless Propag. Lett.*, vol. 19, no. 4, pp. 705–709, Apr. 2020.

[23] S. R. Bose and V. S. Kumar, "Efficient inception v2 based deep convolutional neural network for real-time hand action recognition," *IET Image Process.*, vol. 14, no. 4, pp. 688–696, Mar. 2020.

[24] C. Huang, Y. Li, C. C. Loy, and X. Tang, "Deep imbalanced learning for face recognition and attribute prediction," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 42, no. 11, pp. 2781–2794, Nov. 2020.

[25] Y. Liu, J. Chen, and Y. Qiu, "Joint multi-patch and multi-task CNNs for robust face recognition," *IEICE Trans. Inf. Syst.*, vol. E103.D, no. 10, pp. 2178–2187, Oct. 2020.

[26] Y.-Z. Xiao, Z.-Q. Tian, J.-C. Yu, Y.-S. Zhang, S. Liu, S.-Y. Du, and X.-G. Lan, "A review of object detection based on deep learning," *Multimedia Tools Appl.*, vol. 79, no. 33, pp. 23729–23791, Sep. 2020.

MU-CHUN SU received the B.S. degree in electronics engineering from National Chiao Tung University, Taiwan, in 1986, and the M.S. and Ph.D. degrees in electrical engineering from the University of Maryland, College Park, in 1990 and 1993, respectively. He is currently a Professor of computer science and information engineering with National Central University, Taiwan. He has authored more than 100 journals and refereed conference papers. His current research interests include computational intelligence, neural networks, fuzzy systems, swarm intelligence, affective computing, human–computer interaction, robotics, pattern recognition, biomedical signal processing, image processing, and rehabilitation technology. He is a Senior Member of the IEEE Computational Intelligence Society and Systems, Man, and Cybernetics Society. He is also an IET Fellow. He was the IEEE Franklin V. Taylor Award recipient in 1991. He serves as an Associate Editor for many journals.

JIEH-HAUR CHEN received the B.S. degree in civil engineering from National Central University (NCU), Taiwan, in 1995, the M.S. degree in project management from Northwestern University, Evanston, IL, USA, in 1999, and the Ph.D. degree in civil engineering from the University of Wisconsin–Madison, WI, USA, in 2003. Since 2004, he has been with the Department of Civil Engineering, NCU. He is the author of more than 200 articles. His research interests include computational intelligence, managerial finance in engineering, smart city applications, and property management. He is an Editor of major engineering-related journals and a panel member of the National Science Council, the Ministry of Education, the Ministry of National Defense, the Ministry of Labor, and the Ministry of Internal Affairs, Taiwan. He has received the permanent Distinguished Professor title from NCU, since 2013. His awards and honors include the Extraordinary Military Service Award in 1997, the College Outstanding Service Award in 2009, the NCU Outstanding Researcher Award of the Year in 2011 and 2012, the Special Contribution in Technical Transferring Award in 2015 and 2016, and the NCU Excellence in Service Award in 2017.

ACHMAD MUHYIDIN ARIFAI received the B.S. degree in civil engineering from Sebelas Maret University (UNS), Indonesia, in 2007, and the M.S. degree in project management from the University of Indonesia, in 2016. He is currently pursuing the Ph.D. degree in construction engineering and management with National Central University, Taiwan. Since 2018, he has been a Lecturer with the Trisakti Institute of Transportation and Logistics, Indonesia. His research interests include computational intelligence, project management, value engineering, sustainable development, and transit oriented development. In 2017, he was awarded the Third Place in research competition from the Ministry of Transportation Indonesia.

SUNG-YANG TSAI received the B.S. degree in computer science and information engineering from National Chung Cheng University (NCCU), Taiwan, in 2009, and the M.S. degree in computer science and information engineering from National Central University (NCU), Taiwan, in 2011.

He has nine-year experience in the software industry, including three years of engineering management. He has served in several companies, which are CyberLink, Honestbee, and Carousell. The business domain of the companies covered from multimedia, and online grocery and food delivery service to C2C e-commerce. He has expertise in software development from frontend to backend, collaborating with business stakeholders, people management, and leading cross-functional teams in the startups.

HSI-HSIEN WEI received the B.S. degree in civil engineering from Tamkang University, Taiwan, and the Ph.D. degree in civil and environmental engineering from the University of Maryland, College Park. He was a Senior Research Fellow with the Department of Structural Engineering, Ben-Gurion University, Israel, from 2014 to 2015. He has been an Assistant Professor with the Department of Building and Real Estate, The Hong Kong Polytechnic University, since 2016. His research interests include digital construction and construction technology, resilience and sustainability assessment of the built environment, and interdependent infrastructure systems.