

Received April 3, 2019, accepted April 22, 2019, date of publication May 1, 2019, date of current version May 15, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2914410

Hand Pointing Gestures Based Digital Menu Board Implementation Using IR-UWB Transceivers

ASIM GHAFAR¹, (Graduate Student Member, IEEE), FAHEEM KHAN¹, (Member, IEEE), AND SUNG HO CHO¹, (Member, IEEE)

Department of Electronics and Computer Engineering, Hanyang University, Seoul 04763, South Korea

Corresponding author: Sung Ho Cho (dragon@hanyang.ac.kr)

This work was supported by the National Research Foundation (NRF) through the Bio and Medical Technology Development Program funded by the Korean Government (MIST) under Grant 2017M3A9E2064626.

ABSTRACT Digital menu boards (DMB) are convenient for customers as well as sellers. In this paper, we have implemented a DMB using IR-UWB transceivers. Unlike the traditional touch-based interfaces for menu selection, in our proposed system, users can select items from the menu without touching the screen. The screen is used to display the menu, and the users point to the specific menu item to select it. Multiple radar transceivers are used to create a virtual space divided into different grid blocks in front of the digital display. Patterns in the radar data are analyzed using a multiclass support vector machine (SVM) classifier and a histogram of oriented gradient descriptor. The system is trained at two different distances from the radar sensors in order to make it robust against distance changes. The proposed hand pointing-based DMB system was verified through different experiments, with different grid sizes, to investigate accuracy dependence on grid size. The results showed high accuracy; therefore, the system can be used in real-life scenarios.

INDEX TERMS Digital menu board (DMB), impulse radio ultrawideband, pattern analysis, gestures recognition.

I. INTRODUCTION

Gesture identification is a process of understanding and classifying specific movements by the human body, head, or hands. Hand gesture recognition is significant for human-computer interaction (HCI) and can simplify many essential applications in electronic devices, digital games, automobiles, and defense [1]. Hand gesture recognition has the potential to overcome dependency on the use of conventional HCI devices such as the mouse and keyboard [2].

The hand gesture-based digital menu board (DMB) is gaining more importance in user or human-computer interface systems. The user interface with public displays is facing challenges in both input design and user interaction. In public places, conventional input devices or touch screens are typically not widely present, for several reasons like hygiene or display location [3]. The gesture-based DMB has many advantages in different applications, such as the interaction

between displays in public places [3], selection of a menu in restaurants or cafes, and remote controls for the smart TV.

In hand gesture recognition research, vision-based sensors such as depth cameras are widely used [2]. Optical sensors are one example of high-resolution sensors that help to track and recognize finger and wrist movements. The vision-based techniques can recognize 2D, 3D and real-time hand gestures. Nonetheless, their performance depends upon the conditions of the environment [2], [4], [5]. Moreover, when using depth or vision cameras for gesture recognition, some users might have privacy issues. Radar, however, has no privacy issue. It is convenient to use for gesture recognition as it does not require physical contact. The radar signal is also least affected by outside environments such as humidity and darkness.

The main concept of the IR-UWB radar is that it transmits and receives narrow radio impulses in a wide bandwidth. Many researchers use IR-UWB radar for different applications, such as estimating vital signs by the reflection of radar waveform from the chest as well as back [6], [7]. IR-UWB radar has also been used to count people in indoor and

outdoor environments [8]. In other research work, researchers used IR-UWB radar to determine the 3D position of a moving object using an algorithm based on multiple Quasi-monostatic IR-UWB radar sensors [9]. In reference [10], the authors presented an algorithm for hand gesture recognition using IR-UWB radar. The features used for classification of gestures were Time of Arrival (TOA) and the magnitude of the reflected signal [11]. Another work presented hand gesture recognition through IR-UWB radar to control different electronic devices (sound system or radio system) inside a car, where authors used six gestures inside the car [11]. The features used for classification of gestures were magnitude variance, distance variance, and frequency of hand motion [11]. In reference [12], the authors implemented 10 gestures using FMCW radar, with an accuracy of 89%.

In this paper, we have implemented a DMB based on IR-UWB radar sensors. We used four radar sensors to create a virtual DMB, applying them in real time to evaluate the performance of the proposed system. The gestures used for this work were single click (SC), double click (DC), right swipe (RS), and left swipe (LS). After removing clutter and noise from the radar signals, we transformed them into an image and then used a Histogram of Oriented Gradients (HOG) descriptor-based SVM image classifier [13]. In this system, the user can point to the appropriate grid location to choose an item from the menu, which is displayed on the screen in front of the DMB plane. In order to evaluate the real-time performance of our proposed machine learning-based DMB system, we used a reference measurement based on the localization of the hand during pointing gestures. For this, we used TOA-based localization. Next, we used clustering to find the mid-point of the cluster i.e. centroid of the localization points, which refers to the position of the hand for pointing gestures. Then, the result of the proposed DMB was compared to the result of the selected centroid. The major contributions of this work are as follows:

1. It is the first work to implement a gesture-based menu selection board using IR-UWB radar sensors.
2. We implemented two grids of variable sizes using multiple radar sensors. In addition, we added two types of swipe gestures, i.e. right and left swipe, to change the pages, if the number of grid blocks on a single page was insufficient for the menu items. We always trained the gestures at two different locations to make the system more robust against distance and orientation of the gestures.
3. The data from multiple radar sensors was fused to construct the image for gesture classification. For the implementation of a hand pointing gestures-based DMB, we used a HOG descriptor. The computational cost of HOG was very low as compared to deep neural-based classification, and resulted in high classification accuracy of approximately 96%. Thus, HOG can be used for practical applications.

The rest of the paper is organized as follows: Section II describes the proposed method and pre-processing.

In Section III, we have discussed the algorithm and classifier of the proposed method. The experimental setup is described in Section IV. Experimental results and discussion are presented in Section V, and in Section VI we conclude this paper.

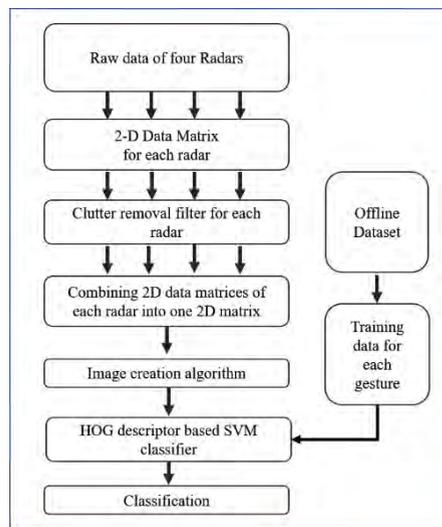


FIGURE 1. Description of the suggested framework.

II. METHODOLOGY

A. PROPOSED ALGORITHM FOR DMB

Our proposed DMB is trained to create a virtual gesture-based screen in front of a display screen. The framework for creating the conceptual digital menu grid is shown in Fig. 1. We implemented an optimized multiclass support vector machine (SVM) classifier-based HOG descriptor to classify the gestures and their position.

Signals from four IR-UWB radar sensors were initially processed to generate an individual 2D matrix for each gesture data. Then, the clutter removal filter was used to remove the ambient data. Next, the clutter removal filter output of all radar sensors was used to create single images. All four data matrices were fused to form an image file. Since the HOG descriptor-based SVM takes images as input, the complexity and computational time and cost of the HOG descriptor-based SVM operation was less than other deep-learning image classifiers. The HOG itself is a features extraction technique, so it was not necessary to extract the features from signals of radars or the combined data of all matrices. The detailed procedure for each process is described in the following sections. In this paper, we have used Xethru-x4 model IR-UWB radar.

B. GESTURES DESCRIPTION

The menu selection grid comprises four gestures, i.e. SC, DC, LS, and RS, based on the motion of the hand. If the hand points once at a grid block of the virtual 2D plane, the classifier takes it as a single click. If the hand points twice, the classifier takes it as a double click. In this paper, the main gestures are used for different purposes.

The single click was used to select an item and display details about it. The double click was used to confirm the selection of an item and the swiping gestures were used for multiple pages. The left swipe was used for the next page, and the right for the previous. In Table 1, the statement $(k \times (m \times n) + 2)$ shows the total number of gestures. The scalar k shows the number of click gestures, “ $m \times n$ ” shows the number of grids in the 2D virtual plane, and 2 represents the number of swiping gestures.

TABLE 1. Specifications of the IR-UWB radio sensor for the proposed HCI.

Parameters	Value
Output power	-12.6 dBm
Center frequency	8.748 GHz
Pulse repetition frequency	40.5 MHz
Bandwidth (-10 dB)	1.4 GHz
Range resolution	6.4 mm
Beamwidth	65°
Staggered PRF sequence length	2 ²⁰ cycles
No. of antenna arrays per radar chip	1 Tx & 1 Rx

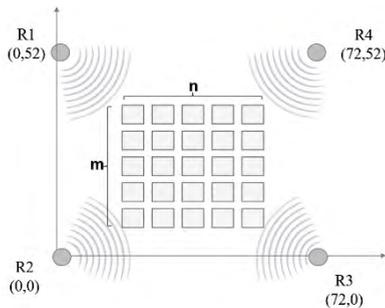


FIGURE 2. Four radars configured in a rectangular plane to create an “ $m \times n$ ” menu grid.

C. HARDWARE CONFIGURATION

The proposed radar-based digital menu selection system is one of the multi-sensory systems operating with four radar sensors. To create a virtual grid menu board, we have installed these four sensors in the 2D plane in a rectangular pattern. Since IR-UWB radar sensors usually have a directional radiation pattern, the four sensors are aligned so that their beam width is focused towards the center point of the circle. Fig.2 shows the archetype of the proposed DMB to generate $(m \times n)$ grids comprising of “ m ” rows and “ n ” columns.

Signals from multiple radars create high redundancy, providing better resolution for dividing the available space into grids as well as individual gesture recognition inside particular grids.

From the above figure, we can see that the four IR-UWB radars are transmitting signal pulses towards the virtual plane, which has been divided into grids for each pointing gesture. For the first case, we have divided the plane into 2×2 grids as shown in Fig. 3.

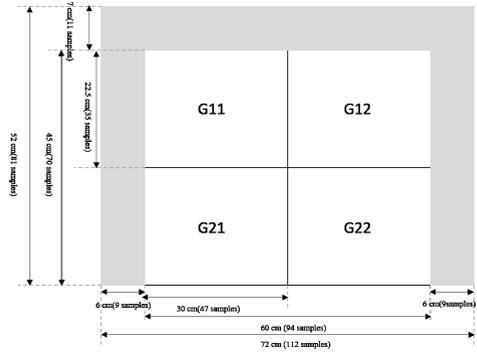


FIGURE 3. Hypothetical plane area of 2×2 grid DMB and parameters of each grid in the plane.

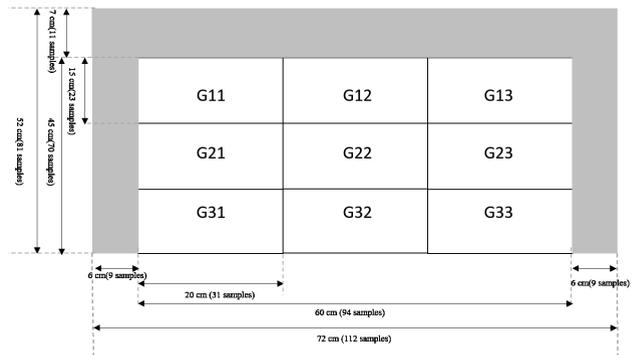


FIGURE 4. Hypothetical plane area of 3×3 grid DMB and parameters of each grid in the plane.

The total area of the hypothetical plan for the grid is A,

$$\text{Total area for } (2 \times 2) \text{ plane} \rightarrow \begin{bmatrix} G11 & G12 \\ G21 & G22 \end{bmatrix}$$

The above figure shows the measurements of length and width of each grid in “cm” as well as in sample points. Now, we will divide the area into a 3×3 grid as follows. Figure 4 shows the hypothetical plane area for 3×3 grids. For both setups, the cells (grids) are distributed uniformly.

$$\text{Total area for } (3 \times 3) \text{ plane} \rightarrow \begin{bmatrix} G11 & G12 & G13 \\ G21 & G22 & G23 \\ G31 & G32 & G33 \end{bmatrix}$$

The dimensions of the rectangular planes are shown in detail in Figs. 3 and 4. The plane size for each setup is same. However, the grid blocks in the 2×2 grid plane will be larger than the 3×3 grid plane. For both setups, the horizontal distance between two radars was 72 cm (112 samples), and the vertical distance was 52 cm (81 samples). As some of the signal samples were wasted because the distance extremely close to the radar sensors was not usable, the actual plane size is smaller compared to the plane’s physical size. The actual usable plane and the plane’s physical size are both shown in Figs. 3. and 4.

For the 2×2 grid shown in Fig. 3., each block was 22.5 cm by 30 cm (35 by 47 in samples). For the 3×3 grid, each block was decreased to 15 by 20 cm (23 by 31 in samples), as shown in Fig. 4.

D. PREPROCESSING AND PATTERN EXTRACTION

The signals transmitted from each radar are $R_1, R_2, R_3,$ and R_4 . Received signals are $r_1(t), r_2(t), r_3(t),$ and $r_4(t)$ respectively, which contain information about the environment. Received raw signal $r_k(t)$ contains details of every object. The unwanted echoes termed as clutter are removed using a simple loopback filter [11] represented by the following equations.

$$c_k(t) = \alpha c_{k-1}(t) + (1 - \alpha) r_k(t), \quad (1)$$

$$y_k(t) = r_k(t) - c_k(t), \quad (2)$$

where k is the slow time index, t is the fast time index, $c_k(t)$ is the clutter signal, and $y_k(t)$ is the background-subtracted signal from which the clutter signal is removed. Here, alpha is the weighting constant that controls the sensitivity of the clutter removal process. A high value of alpha will make the filter highly robust, and only macro motions will be passed through the loopback filter; a low value of alpha will also allow micro motions to pass through the filter. For our experiment, it was found that the value of alpha provided enough optimized allowable signal to subtract the background reflections and pass desired signals only.

The raw signal and the signal after background-subtraction are shown in Fig. 5. The signal before clutter removal has a greater magnitude around index value 60, whereas the signal after clutter removal shows that the clutter part is removed, so the signal magnitude around index value 60 is lower as compared to the motion part, which is centered at index value 118 in Fig. 5(b).

After clutter removal, the one-dimensional signal from each radar is stored for a certain slow time to make a 2D matrix, which is then fused and transformed into a single grayscale image for further image processing, as shown in Section III.

III. IMAGE TRANSFORMATION AND CLASSIFICATION

Generally, the SVM classifier is a binary classifier ($k = 2$) capable of separating the class variables into two classes. For our problem, the classification problem is multiclass, with ($k = 10$) for a 2×2 DMB and ($k = 20$) for a 3×3 DMB. The multiclass SVM uses SVM learners and a “one vs one” encoding scheme and returns a trained multiclass SVM model. For dataset training, error correcting output codes (ECOC) used multiclass SVM binary learners. To apply the multiclass SVM classifier, we combine the pre-processed signals from the four radars. The detailed procedure for training is given in Algorithm 1 as shown below.

After applying Algorithm 1, we can get the image for the gesture G_{23} as shown in Fig.6.

Fig.6 shows the transformation of the data obtained from the four radar sensors into a grayscale image. After image transformation, the HOG descriptor-based SVM classifier algorithm is applied as follows.

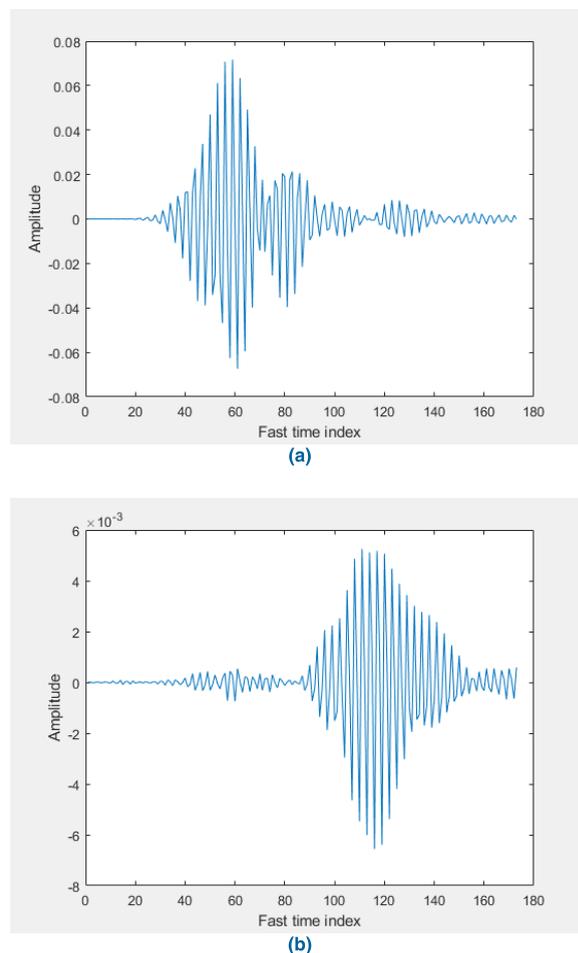


FIGURE 5. One data sample of received radio signal (a) before clutter removal and (b) after clutter removal.

Algorithm 1 Radar Signal to Image Transformation for HOG Descriptor-Based SVM Classifier

1. The input signal from four radar sensors $r(t), r_2(n), r_3(t)$ and $r_4(t)$
2. Remove the clutter from the signal as described in Sec. II-D
3. Background subtracted signals are combined into a matrix of size $k \times t$ as represented in (2)
4. Combine the 2D signal from the four radars as shown in Fig. 6(a-d)
5. Convert the four 2D matrices to the RGB of an image as shown in Fig. 6(e)
6. Convert it into the greyscale image as shown in Fig. 6(f)

A. HOG DESCRIPTOR-BASED SVM CLASSIFIER

The HOG descriptor-based SVM classifier was implemented for the DMB. Fig. 7 shows the overall schema of converting the arbitrary input image into HOG descriptors (i.e. the features [13]–[15]). First, the input image is converted into

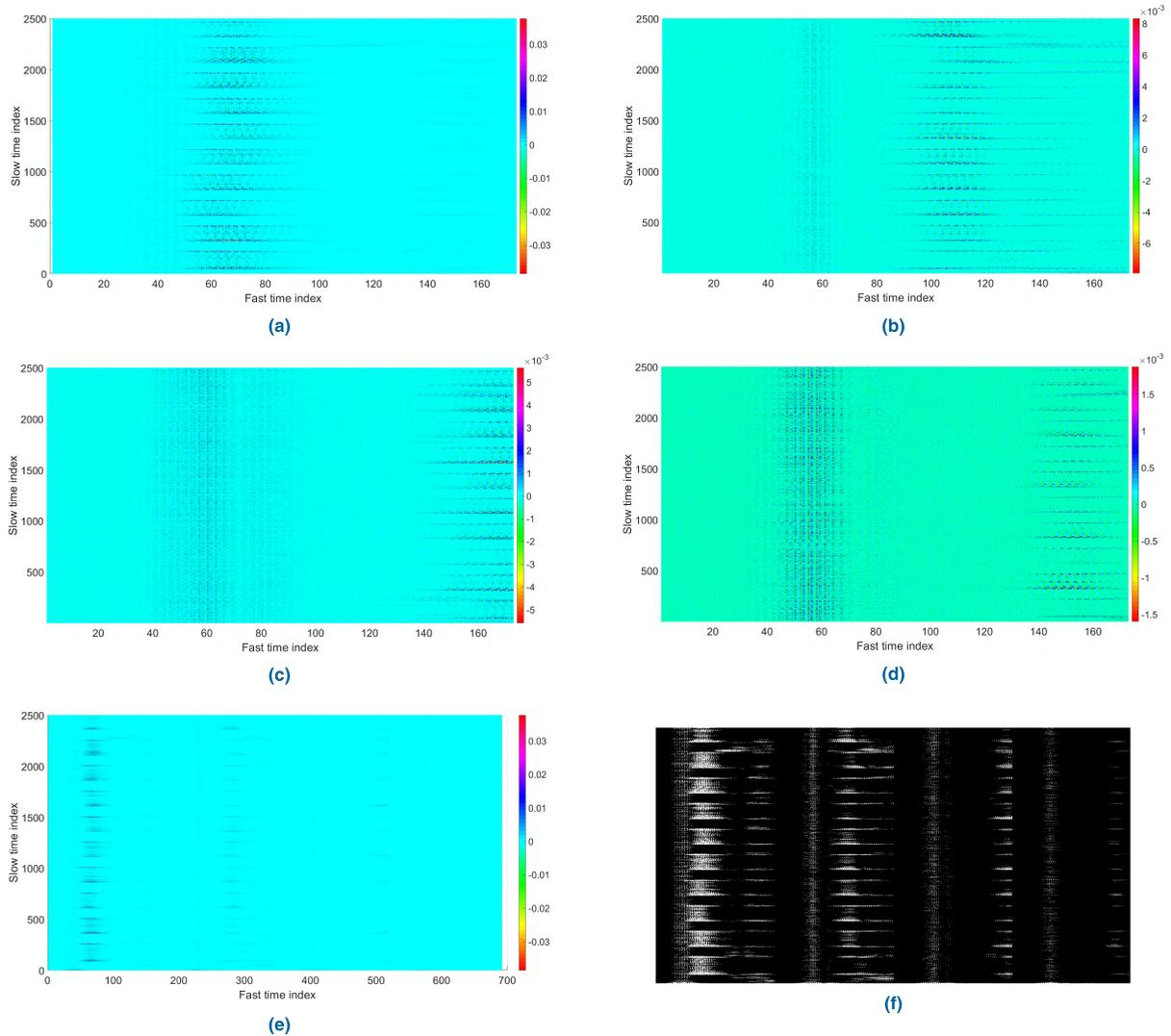


FIGURE 6. (a) Background subtracted data received from radar R1. (b) Background subtracted data received from radar R2. (c) Background subtracted data received from radar R3. (d) Background subtracted data received from radar R4. (e) Combined Background subtracted data of the four radars. (f) Greyscale image of the Background subtracted data from four radars for pointing G_{23} .

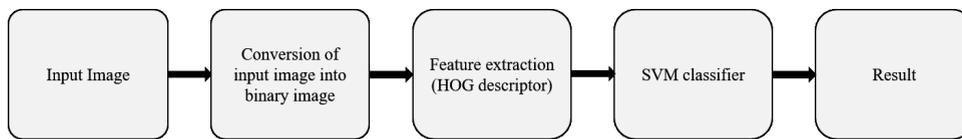


FIGURE 7. Steps of HOG descriptor based multiclass SVM.

binary images. Then, a HOG descriptor has to be computed to calculate the horizontal and vertical gradients, which can be accomplished by filtering the image through kernel filters as shown in (3) and (4) [14]:

$$G_x = [-1, 0, 1] \tag{3}$$

$$G_y = [-1, 0, 1]^T \tag{4}$$

After the calculation of gradients, we need to calculate their magnitude and direction. The magnitude “G” and

direction “D” of the gradients was found by the equations given below.

$$D = \tan^{-1} \frac{G_y}{G_x} \tag{5}$$

$$G = \sqrt{G_x^2 + G_y^2} \tag{6}$$

The magnitude of gradients triggers where a sharp change in illumination occurs. In the gradient image, a lot of clutter was removed. For feature calculation, we divided the image

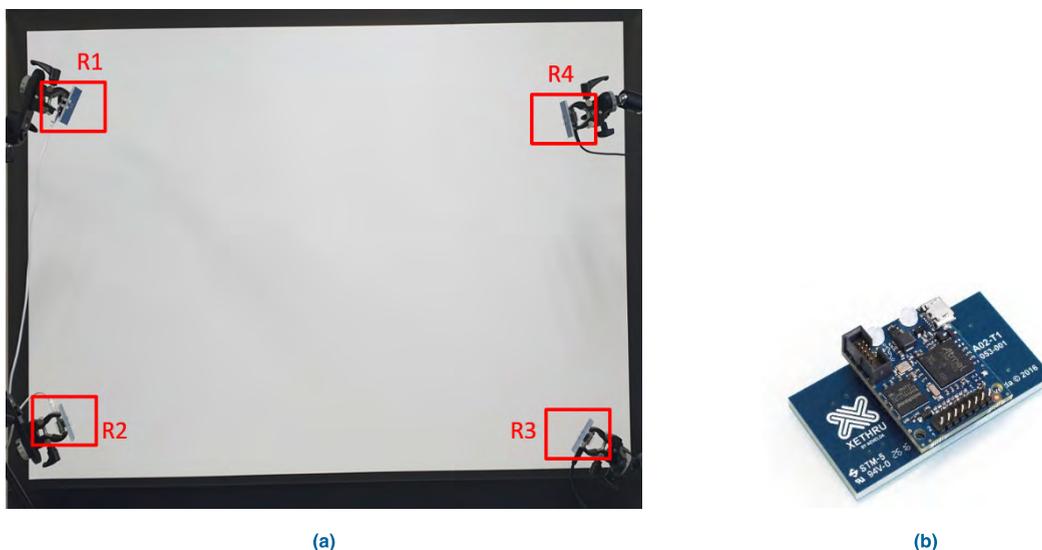


FIGURE 8. (a) Front view of our actual experimental setup. (b) IR-UWB radar module.

into 8×8 cells. Using a large cell size can cause loss of useful information because of squeeze in image size. Conversely, if we use small cells in a high-resolution study, some useless information may be used in the image. For a classifier, it is difficult to discriminate between useful and useless information. Therefore, in this paper, we used an optimum 8×8 cell size.

IV. EXPERIMENTAL SETUP

We used four IR-UWB radars as shown in Fig. 8(a), to obtain a mid-air digital virtual menu board (DMB). Four radars were used for implementation of the DMB because the transceivers have a low beam width (around 65 degree), which makes it hard to cover the whole plane with two, or even three, radar sensors. When we tested two and three radar sensors, there were many inaccurate results as the hand pointing gestures occasionally did not occur in the beam width of the transceivers and, hence, resulted in decreased RCS values. Four sensors had some diversity; thus, the accuracy of recognition was improved. The IR-UWB radar module used was Xethru X4 (Novelda, Norway), as shown in Fig 8(b). In this paper, we used two different setups to create a DMB. In the first setup, we formed a 2×2 grid mid-air DMB. The 2×2 plane had a total of four grids, each trained by two gestures. Thus, it had eight click gestures (i.e. SC and DC); with two more swiping gestures, the total number became 10. The total number of gestures for any grid size may be calculated from Table 2. In addition, we trained the gestures at multiple distances to make the gesture recognition robust against distance between pointing hand location and virtual plane.

After the first setup, we implemented the second, where we increased the number of grids from 2×2 to 3×3 to include more item slots for selection. The setup had nine grids and 20 total gestures. Fig. 8(a). shows the actual experimental setup.

TABLE 2. Gesture description.

Item	Quantity	Description
Click gestures	2	Single click (SC) and double click (DC)
Swiping gestures	2	Left swipe (LS) and right swipe (RS)
Grid size	$m \times n$	Division of screen
Number of gestures per grid	$k \times (m \times n)$	Two gestures per grid
Total number of gestures	$k \times (m \times n) + 2$	All gestures to be implemented

V. EXPERIMENTAL RESULTS AND DISCUSSION

To create diversity in the training set, the training data used in this research for the proposed DMB was collected from four male volunteers aged between 24 and 31. A total of 100 samples was collected for each gesture, of which 80 were used for training and 20 for evaluation of the classifier. The classifier automatically extracts the features of a gesture performed in the grid plane and compares its main features with trained features, hence classifying it. We first studied gesture recognition efficiency with offline data when the HOG descriptor-based SVM classifier was trained. After offline data training and evaluation, we evaluated the trained classifier in real time. The performance of our proposed virtual DMB was efficient for the proposed application.

A. RAW DATA PATTERN FOR DIFFERENT GRIDS

In order to show how data differed in both cases (e.g. two extreme cases), we presented an example involving two single click gestures at two different grids as follows. Fig. 9(a). shows the data obtained when a pointing gesture was performed at grid position G_{11} (top left). Fig. 9(b). shows the pattern of grid position G_{33} . Visibly, the patterns in the data obtained from pointing gestures at two different locations vary significantly; hence, they can be classified using

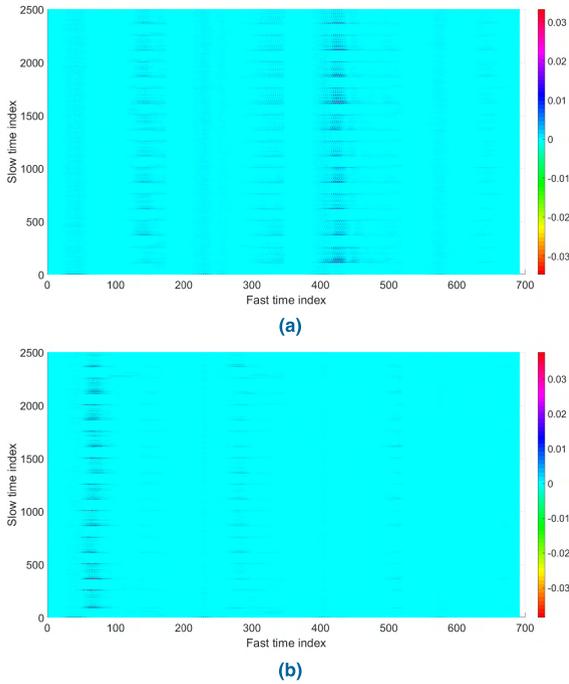


FIGURE 9. (a) Radar data plot for gesture $G_{11}C_1$ and (b) $G_{33}C_1$.

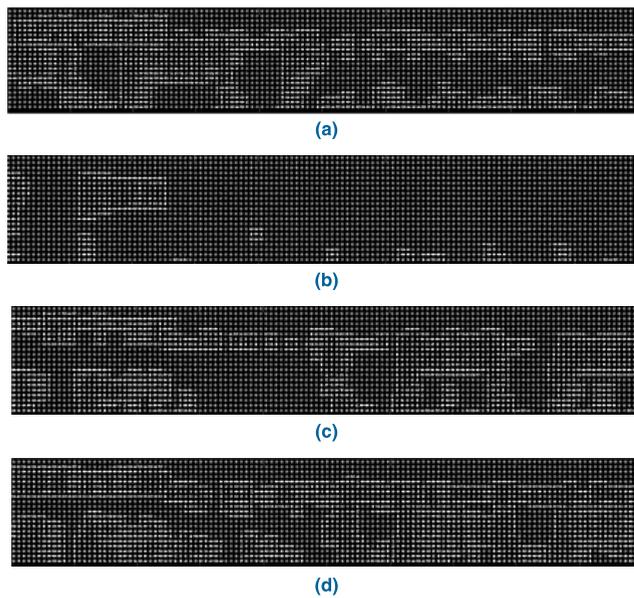


FIGURE 10. Binary image of SC (single click) of 2×2 DMB (a) is the binary image of $G_{11}C_1$, (b) is of $G_{12}C_1$ (c) is of $G_{21}C_1$ and (d) is of $G_{22}C_1$.

an image classification algorithm as illustrated in the next section.

B. HOG FEATURE EXTRACTOR RESULTS

The images obtained were passed to the HOG feature extractor, whose results are presented in this section. The feature vector of an image has a size of “89280” features and a cell size of 8×8 . For comparison, the visualized images of the HOG of different gestures in a 2×2 DMB are shown in Fig. 10. The visualized images show the dominant direction

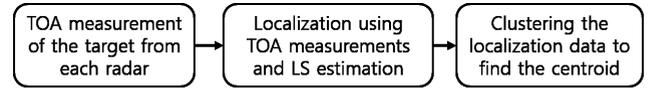


FIGURE 11. Block diagram for localization based gesture detection.

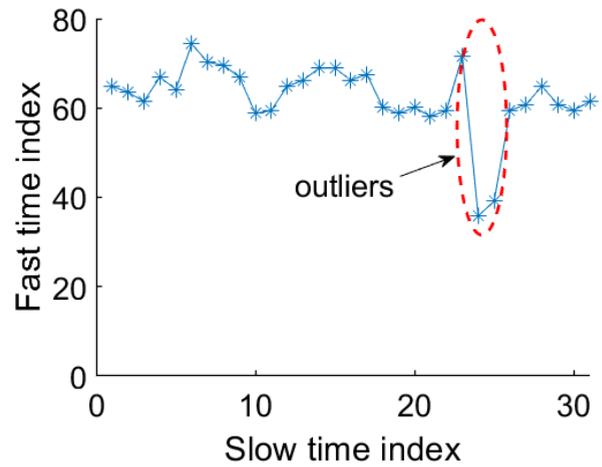


FIGURE 12. TOA result of radar sensor R1 for the pointing gesture G_{23} .

of the histogram that captures the shape of performed gestures. Figure 10 shows that there is a clear difference in magnitude and direction of HOG for different gestures of the 2×2 DMB.

C. CLASSIFICATION PERFORMANCE

1) REFERENCE MEASUREMENTS

To find reference measurement for our algorithm evaluation, we chose localization-based gesture detection as shown in [16]. Authors in [16] detected the touchless menu selection using a Kinect sensor with an accuracy of 97.3%, which can be used as a standard for evaluation of our algorithm. Although this localization-based detection was not convenient as it needs manual tuning of sensor coordinates and other parameters, we only used it for comparison with our machine learning-based algorithm, which does not need manual tuning. In our proposed technique, the features were automatically extracted by the algorithm during the training phase. In order to evaluate the performance of our proposed system, we used the least square (LS) technique to find the location of the hand at each slow time and then cluster all these points for the whole gesture duration. In this way, the center of the cluster gave us the precise location of the hand as a reference point for comparison with our proposed HOG-based technique.

The TOA algorithm was very simple, as we used the magnitude of the signal as a criterion for hand localization. The result of the TOA of radar sensor R1 for the menu position of G_{23} is shown in the following figure.

Due to the narrow beam width and clutter in the background, the TOA measurements may result in some outliers such as the samples at slow time index “24” and “25” in Fig. 12. We used a median filter to remove these outliers from the measurement data.

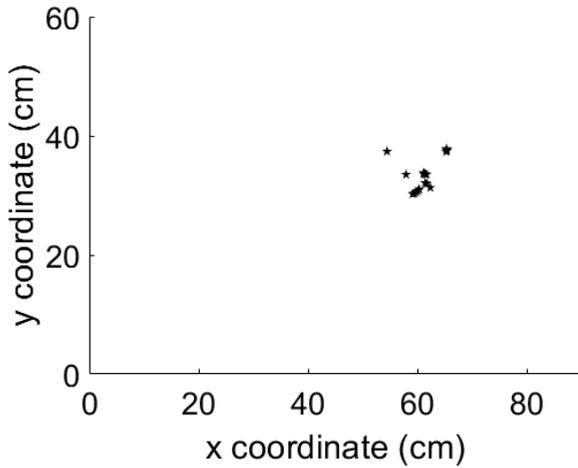


FIGURE 13. Localization result for the pointing gesture G_{23} .

After the TOA estimation for the four radar sensors, we needed to estimate the location of the hand in the 2D space. In order to get the localization data, we used the Least Squares (LS) estimation technique. The LS technique was developed in [17], where the TOA measurements of each sensor are used to find the location of the object in the 2D plane. The relation between the distances and coordinates of the target and radars is given by (7). LS was applied to estimate the position of the hand [17].

$$d_i = \sqrt{(x - x_i)^2 + (y - y_i)^2} \quad i = 1, \dots, 4 \quad (7)$$

d_i is the distance from the i^{th} radar to the hand and (x_i, y_i) is the position of the radar. The detailed solution to this range equation is given in [17]. The result of the localization for grid position G_{23} is shown in Fig.13 below:

After the localization of the hand during the pointing gestures, we computed the center of the centroid of all these localization points to determine which grid the pointing gesture belonged to. To that end, we used an unsupervised learning algorithm, i.e. k-means clustering, which is fast, robust,

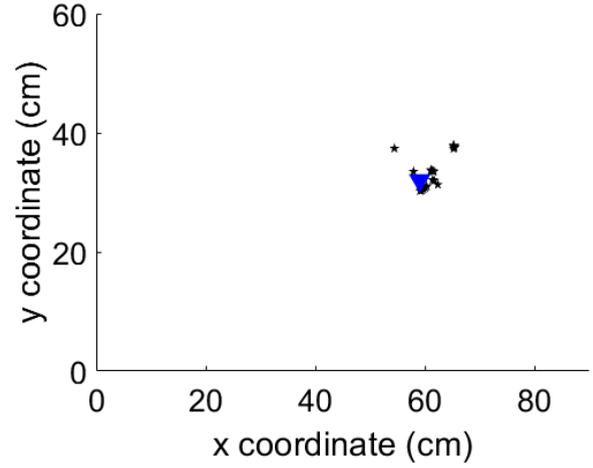


FIGURE 14. Clustering of localization points to find the centroid as shown in blue.

and very simple [11]. The cost function is given by (8).

$$J(V) = \sum_{i=1}^c \sum_{j=1}^{c_i} (\|x_i - v_j\|)^2 \quad (8)$$

where, $\|x_i - v_j\|$ is the Euclidean distance between x_i and v_j and ' c_i ' is the number of data points in i^{th} cluster while " c " is the number of cluster centers.

Classification performance was based on offline and online data. The system was trained based on data collected, and real-time testing was also performed to check the accuracy in the actual environment. Initially, the 2×2 DMB system was trained for one distance. A total of 100 samples were collected for each gesture. After training and testing the system at the same distance, we obtained an overall accuracy of 100%. Detailed results are shown in Table 3. However, the system was trained at one distance and tested at two different distances, which gives us an overall accuracy of around 75%. The accuracy decreased because of the magnitude of the reflected signal; hence, the signal pattern changed with the changing distance. To increase the robustness of the system, the 2×2 DMB was trained at two distances.

TABLE 3. Classification results of 2×2 virtual DMB for one distance dataset.

Actual/predicted	LS	RS	$G_{11}C_1$	$G_{11}C_2$	$G_{12}C_1$	$G_{12}C_2$	$G_{21}C_1$	$G_{21}C_2$	$G_{22}C_1$	$G_{22}C_2$
LS	100	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
RS	0.0	100	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
$G_{11}C_1$	0.0	0.0	100	0.0	0.0	0.0	0.0	0.0	0.0	0.0
$G_{11}C_2$	0.0	0.0	0.0	100	0.0	0.0	0.0	0.0	0.0	0.0
$G_{12}C_1$	0.0	0.0	0.0	0.0	100	0.0	0.0	0.0	0.0	0.0
$G_{12}C_2$	0.0	0.0	0.0	0.0	0.0	100	0.0	0.0	0.0	0.0
$G_{21}C_1$	0.0	0.0	0.0	0.0	0.0	0.0	100	0.0	0.0	0.0
$G_{21}C_2$	0.0	0.0	0.0	0.0	0.0	0.0	0.0	100	0.0	0.0
$G_{22}C_1$	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	100	0.0
$G_{22}C_2$	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	100

TABLE 4. Classification results of 2 × 2 virtual DMB for two distance datasets.

Actual/predicted	LS	RS	G ₁₁ C ₁	G ₁₁ C ₂	G ₁₂ C ₁	G ₁₂ C ₂	G ₂₁ C ₁	G ₂₁ C ₂	G ₂₂ C ₁	G ₂₂ C ₂
LS	100	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
RS	0.0	100	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
G ₁₁ C ₁	0.0	0.0	100	0.0	0.0	0.0	0.0	0.0	0.0	0.0
G ₁₁ C ₂	0.0	0.0	15.0	85.0	0.0	0.0	0.0	0.0	0.0	0.0
G ₁₂ C ₁	0.0	0.0	0.0	0.0	90.0	10.0	0.0	0.0	0.0	0.0
G ₁₂ C ₂	0.0	0.0	0.0	0.0	15.0	85.0	0.0	0.0	0.0	0.0
G ₂₁ C ₁	0.0	0.0	0.0	0.0	0.0	0.0	100	0.0	0.0	0.0
G ₂₁ C ₂	0.0	0.0	0.0	0.0	0.0	0.0	0.0	100	0.0	0.0
G ₂₂ C ₁	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	100	0.0
G ₂₂ C ₂	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	100

TABLE 5. Classification results of 3 × 3 virtual DMB.

Gesture	Average accuracy
LS	100
RS	100
G ₁₁ C ₁	100
G ₁₁ C ₂	100
G ₁₂ C ₁	100
G ₁₂ C ₂	90
G ₁₃ C ₁	100
G ₁₃ C ₂	100
G ₂₁ C ₁	100
G ₂₁ C ₂	95
G ₂₂ C ₁	100
G ₂₂ C ₂	100
G ₂₃ C ₁	100
G ₂₃ C ₂	85
G ₃₁ C ₁	100
G ₃₁ C ₂	85
G ₃₂ C ₁	100
G ₃₂ C ₂	80
G ₃₃ C ₁	100
G ₃₃ C ₂	100

After obtaining all data of all gestures for the 2 × 2 DMB at two different locations, we evaluated the accuracy of the gesture recognition in real time as shown in Table 4. The results were tabulated in the form of a confusion matrix. Hence, training at multiple distances made the classification robust to distance change. The classification results of the 3 × 3 DMB are given in Table 5. The results of the training and evaluation process were determined by a multiclass SVM. The overall accuracy of gesture recognition in our proposed DMB was distinct for both setups. The overall accuracy for both setups was calculated by (9). Tables 3, 4, and 5 are based on the

individual accuracy of each gesture. From Table 4, we can see that all gestures showed perfect accuracy of 100%, except in a few cases: G₁₁C₂ had 85% accuracy, G₁₂C₁ had 90%, G₁₂C₂ had 85%, and the overall accuracy was 96%. From Table 5, we can see that all gestures showed perfect accuracy of 100%, except in a few cases: G₁₂C₂ had an accuracy of 90%, G₂₁C₂ had 95%, G₂₃C₂ had 85%, G₃₂C₂ had 80%, and overall accuracy was 96.75%. In the 3 × 3 DMB setup, we used only one distance to obtain the gestures, due to which the accuracy was higher as compared to the 2 × 2 DMB setup in Table 4.

The overall accuracy was calculated by equation (9).

$$\text{Overall accuracy} = \frac{\text{Correctly predicted gestures}}{\text{Total no. of gestures}} \times 100 \quad (9)$$

VI. CONCLUSION

In this paper, we applied a virtual DMB method using IR-UWB radar sensors and machine learning. In our proposed method, we focused on a pointing gesture-based HCI. We used a multiclass SVM classifier for gesture classification and a HOG descriptor for feature extraction. We implemented two DMBs of different grid sizes to show the scalability of our proposed system. Moreover, we trained our proposed algorithm for pointing gestures at multiple locations to enhance its performance and make it robust against distance change. The gesture classification results showed high accuracy, which means that this system may be feasible for practical applications. The main concept of our work was to build an idea toward use of hand gestures in modern technologies instead of touch-based public display screens. The major benefit of our proposed DMB is that it is touchless, and significantly more user-friendly in any environment. This method can work in modern restaurants or cafes, which use touch-based display screens for menu selection. It can also work for smart TVs and other such applications. In the future, touchless screens could be merged in homes, restaurants, etc.

ACKNOWLEDGMENT

(Asim Ghaffar and Faheem Khan are co-first authors.)

REFERENCES

- [1] Y. Kim and B. Toomajian, "Hand gesture recognition using micro-Doppler signatures with convolutional neural network," *IEEE Access*, vol. 4, pp. 7125–7130, 2016.
- [2] J. Wang and S. Payandeh, "Hand motion and posture recognition in a network of calibrated cameras," *Adv. Multimedia*, vol. 29, Apr. 2017, Art. no. 2162078.
- [3] R. Walter, G. Bailly, N. Valkanova, and J. Müller, "Cuenesics: Using mid-air gestures to select items on interactive public displays," in *Proc. 16th Int. Conf. Hum.-Comput. Interact. Mobile Devices Services*, Sep. 2014, pp. 299–308.
- [4] C. Tran and M. M. Trivedi, "3-D posture and gesture recognition for interactivity in smart spaces," *IEEE Trans. Ind. Informat.*, vol. 8, no. 1, pp. 178–187, Feb. 2012.
- [5] M. H. Yang, N. Ahuja, and M. Tabb, "Extraction of 2D motion trajectories and its application to hand gesture recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 24, no. 8, pp. 1061–1074, Aug. 2002.
- [6] F. Khan and S. H. Cho, "A detailed algorithm for vital sign monitoring of a stationary/non-stationary human through IR-UWB radar," *Sensors*, vol. 17, no. 2, p. 290, 2017.
- [7] S. K. Leem, F. Khan, and S. H. Cho, "Vital sign monitoring and mobile phone usage detection using IR-UWB radar for intended use in car crash prevention," *Sensors*, vol. 17, no. 6, p. 1240, 2017.
- [8] J. W. Choi, S. H. Cho, Y. S. Kim, N. J. Kim, S. S. Kwon, and J. S. Shim, "A counting sensor for inbound and outbound people using IR-UWB radar sensors," in *Proc. IEEE Sensors Appl. Symp. (SAS)*, Apr. 2016, pp. 1–5.
- [9] C. J. Woo and S. H. Cho, "3D positioning algorithm based on multiple quasi-monostatic IR-UWB radar sensors," in *Proc. IEEE Radar Conf. (RadarConf)*, May 2017, pp. 1531–1535.
- [10] R. Nan, X. Quan, and S. H. Cho, "Algorithm for gesture recognition using an IR-UWB radar sensor," *J. Comput. Commun.*, vol. 4, no. 3, p. 95, 2015.
- [11] F. Khan, S. K. Leem, and S. H. Cho, "Hand-based gesture recognition for vehicular applications using IR-UWB radar," *Sensors*, vol. 17, no. 4, p. 833, Apr. 2017.
- [12] L. Gang, S. Zhang, F. Fioranelli, and H. Griffiths, "Effect of sparsity-aware time-frequency analysis on dynamic hand gesture classification with radar micro-Doppler signatures," *IET Radar, Sonar Navigat.*, vol. 12, no. 8, pp. 815–820, Apr. 2018.
- [13] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, Jun. 2005, pp. 886–893.
- [14] K. Silanon, "Thai finger-spelling recognition using a cascaded classifier based on histogram of orientation gradient features," *Comput. Intell. Neurosci.*, vol. 24, no. 5, Apr. 2017, Art. no. 9026375.
- [15] A. Sophian and D. Aini, "Fingertip detection using histogram of gradients and support vector machine," *J. Elect. Eng. Comput. Sci.*, vol. 8, no. 2, pp. 482–486, 2017.
- [16] D. Chattopadhyay and D. Bolchini, "Touchless circular menus: Toward an intuitive UI for touchless interactions with large displays," in *Proc. Int. Working Conf. Adv. Vis. Interfaces*, May 2014, pp. 33–40.
- [17] K. W. Cheung, H. C. So, W.-K. Ma, and Y. T. Chan, "Least squares algorithms for time-of-arrival-based mobile location," *IEEE Trans. Signal Process.*, vol. 52, no. 4, pp. 1121–1130, Apr. 2004.



ASIM GHAFFAR (GS'19) was born in Khuzdar, Baluchistan, Pakistan, in 1991. He received the B.E. degree in electrical engineering from the Baluchistan University of Engineering and Technology Khuzdar (BUETK). He is currently pursuing the M.S. and Ph.D. degrees in electronics, computer and communication with Seoul Hanyang University, Seoul, South Korea.

His research interests include signal processing, image processing, and embedded system implementation, especially about the IR-UWB radar-based gesture recognition, detection, and positioning.



FAHEEM KHAN (M'19) was born in Bannu, Pakistan, in 1988. He received the Ph.D. degree in electronics and computer engineering from Hanyang University, South Korea, in 2018.

He is currently a Postdoctoral Researcher with Hanyang University. He has authored several conference papers and articles in reputed journals. His research interests include radar signal processing, sensors-based localization, and tracking and gestures recognition.



SUNG HO CHO (S'86–M'88) received the Ph.D. degree in electrical and computer engineering from The University of Utah, Salt Lake City, USA, in 1989. From 1989 to 1992, he was a Senior Member of Technical Staff with the Electronics and Telecommunications Research Institute, Daejeon, South Korea. He joined the Department of Electronic Engineering, Hanyang University, Seoul, South Korea, in 1992, where he is currently a Professor. His research interests include applied signal

processing, machine learning for signal processing, context-aware computing, radar sensors, and smart space and wireless networks. Since 2008, he has been a recipient of the High-Level Foreign Experts Fellowship at the Beijing University of Posts and Telecommunications, Beijing, China, led by the Ministry of Education of China.

...