

Hand Gesture Recognition for Smart Devices by Classifying Deterministic Doppler Signals

Yi Zhang, Shuqin Dong, Chengkai Zhu, Marcel Balle, Bin Zhang, and Lixin Ran^{ID}

Abstract—Personal devices such as smartphones and tablets are rapidly becoming personal communication, information, and control centers. Apart from multitouch screens, human gestures are considered as a new interactive human–smart device interface. In this work, we propose a noncontact solution to implement hand gesture recognitions for smart devices. It is based on a continuous wave, time-division-multiplexing (TDM), single-input multiple-output (SIMO) Doppler radar sensor that can be realized by slightly modifying existing RF front ends of smart devices, and a machine-learning algorithm to recognize predefined gestures by classifying deterministic Doppler signals. An experimental setup emulating a smartphone-based radar sensor was implemented, and the experimental results verified the robustness and the accuracy of the proposed approach.

Index Terms—Classification, doppler effect, hand gesture recognition (HGR), machine learning, radar architecture.

I. INTRODUCTION

NOWADAYS, personal devices such as smartphones and tablets have become indispensable for daily human life. Integrated with wireless interfaces, sensors, and powerful computation ability, smart devices are rapidly becoming personal communication, information, and control centers [1].

So far, most interactions between humans and smart devices are based on multitouch screens [2]. With applications of smart devices extending to scenarios such as virtual reality and wireless control of televisions (TVs) and projectors, noncontact interactions based on human hand gesture recognitions (HGRs) have attracted more and more interests [3]–[6].

Originally, noncontact HGR was considered as a human–machine interface (HMI) in computer science [7]. Based on cameras and digital imaging processing algorithms, such a wireless HMI can be classified as an optical approach. Highly relying on computational resources, direct line of sights, and illuminating conditions, the application of computer vision-based HGRs is so far limited [8]. To overcome the above cost, obstacle, and illumination issues, HGRs based on microwave and millimeter-wave approaches, i.e., pulsed and continuous-wave (CW) radar approaches, have been investigated [9]–[18].

Manuscript received June 6, 2020; revised July 21, 2020 and August 12, 2020; accepted September 2, 2020. Date of publication November 4, 2020; date of current version January 5, 2021. This work was supported by NSFC under Grant 61771421 and Grant 62071417. (Corresponding author: Lixin Ran.)

The authors are with the Laboratory of Applied Research on Electromagnetics (ARE), Zhejiang University, Hangzhou 310027, China (e-mail: ranlx@zju.edu.cn).

Color versions of one or more of the figures in this article are available online at <https://ieeexplore.ieee.org>.

Digital Object Identifier 10.1109/TMTT.2020.3031619

So far, most of the reported radar-based HGRs are based on the frequency modulated continuous wave (FMCW) scheme [12]–[15], [19], [20]. Among them, the most renowned is the “Google Soli” [15], which is a 60-GHz FMCW radar module with a bandwidth up to 7 GHz. Based on classifying range-Doppler spectrograms [10], it can recognize fine gestures such as finger rubs, which can be used to control features such as the volume and music playlist of a smartphone. However, the added millimeter-wave radar chip and its peripherals increase the price, power consumption, and inner space occupation, all are scarce resources for space and resource constraint devices.

So far, two classes of recognition algorithms have been proposed in the above radar-based solutions. One is based on linear methods, trying to retrieve the displacement of targets, i.e., the moving pattern of the gestures, from the backscattered fields [4]–[6], [9]. The advantage of these methods is that the retrieved gesture patterns are meaningful, having direct correspondence with the actual gestures. The disadvantage is that due to the Mie scattering [21] of hands, the shape of the detected patterns can have significant variation due to different angles of illumination, increasing the difficulty of recognition.

Another class of algorithms is based on artificial intelligence (AI), taking advantage of the nonlinear generalization capacity of artificial neural networks (ANNs) and machine learning-based classifications [22]–[24]. These methods recognize the “signature” of the fields scattered by gestures. The advantage of these methods is that a “black-box” recognition can always be implemented to any predefined gestures, as long as such signatures are used to train the ANN-based classifiers [10], [12]–[14], [16], [17], [19], [25]. For such black-box classifications, sufficient gestures made from different people are required to train the classifiers. Unfortunately, the number of such training gestures can be up to tens of thousands, leading to high human costs and time consumption in gesture data set acquisition, data labeling, and data cleaning.

In this work, we combine the advantages of the above architectures and algorithms, proposing a systematic solution to implement HGRs for smart devices. It is based on a CW, time-division-multiplexing (TDM), single-input multiple-output (SIMO) Doppler radar sensor (DRS) and a machine-learning algorithm to recognize hand gestures by classifying deterministic Doppler signals. Such a DRS can be implemented by slightly modifying the existing RF front ends of smart devices, increasing negligible hardware cost, power consumption, and space occupation. Different from the

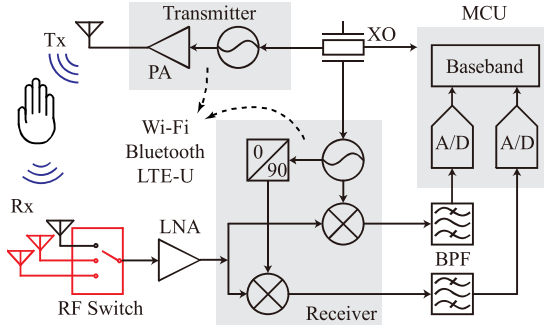


Fig. 1. Implementation of a TDM-SIMO DRS by reconfiguring the Wi-Fi and LTE-U RF channels. Devices drawn in red are the added ones.

“black-box” classifications, such deterministic Doppler signals are retrieved with clear physical and mathematical meanings, and thus a significantly reduced number of gestures made by only one person can be used to train a classifier, based on which a highly accurate HGR adaptive to different users can be obtained. Experimental demonstrations based on a setup emulating a smartphone-based DRS verified the robustness and the accuracy of the proposed HGR.

The rest of this article is organized as follows. In Section II, relevant existing DRS works are revisited, based on which the architecture of a smartphone-based TDM SIMO DRS and the algorithms dedicated for this DRS are introduced. In Section III, the design of the gestures and the algorithms for the proposed DRS are derived. Sections IV and V provide full-wave simulations and experimental demonstrations, respectively. Finally, a conclusion is drawn in Section VI.

II. ARCHITECTURE

A. Architecture

With the rapid progress of RF integrated circuits (RFICs), miniaturized CW DRSs have been developed for noncontact motion detections in the past decade. Examples of the targeted motions include human vital signs [26]–[30] and mechanical motions [31]–[33]. Originally, CW DRSs with single-input single-output (SISO) architectures were used [26], [27], [29], [30]. Assisted by algorithms capable of eliminating phase ambiguity [4], [34], [35], 1-D motions can be linearly detected. For 2-D motions, SIMO DRSs with one transmitting and two receiving channels were also implemented [4], [36], [37]. However, in all the above SIMO DRSs, the transmitting and receiving antennas were placed on a straight line and, therefore, cannot be used to detect 3-D gestures. Consisting of multiple RF channels, they are also difficult to be integrated into a smart device.

In this work, taking advantage of the ultra-narrowband nature of CW DRSs, we propose a TDM-SIMO DRS that can be realized by slightly modifying the existing RF front ends in conventional smart devices. The schematic is shown in Fig. 1.

So far, smart devices, especially smartphones and pads, have integrated multiple wireless interfaces. Apart from the RF front ends for 2G, 3G, 4G, and 5G communications, there are Wi-Fi, Bluetooth, and long term evolution (LTE)-U [38] channels working at 2.4- and 5.8-GHz bands. In most

devices, all such channels are implemented with standard digital-IF architectures sharing the same reference oscillator and baseband unit. Therefore, such channels can be flexibly chosen to implement a 2.4- or 5.8-GHz front end for HGRs.

For example, the transmitting channel of the 2.4-GHz Bluetooth and the receiving channel of the 2.4-GHz Wi-Fi can be combined to obtain a 2.4-GHz front end. Similarly, the 5.8-GHz transmitting and receiving channels of the Wi-Fi and the LTE-U can be combined to obtain a 5.8-GHz front end for HGRs. Such combinations can be accomplished by reconfiguring the related channels in baseband, without any necessity of hardware change.

To implement HGRs for 3-D gestures, as analyzed in the following, a SIMO DRS with one transmitting and at least three receiving channels are required. In this work, it is achieved by multiplexing the receiving channel. As shown in Fig. 1, a single-pole three-throw (SP3T) RF switch is used to connect three receiving antennas, realizing three TDM receiving channels. Note that this multiplexing will not impact the normal operation of Wi-Fi, Bluetooth, or LTE-U. Since human gestures are second-scale motions, even a 10-Sa/s sampling rate suffices. As an example, for an analog-to-digital converter (ADC) with a sampling frequency of 1 MHz, finishing the consecutive three samplings would only occupy a time slice of a few microseconds in each 10-ms period.

According to the above discussions, an HGR front end can be implemented by reconfiguring the existing RF channels of smart devices. The only added device is the 1P3T switch and two additional antennas.

B. Detected Signals

For the quadrature receiver shown in Fig. 1, the baseband I/Q signals digitized by the i th TDM channel can be expressed as

$$\begin{aligned} I_i[n] &= A_{I,i}[n] \cos(\Phi_i[n]) + dc_{I,i}[n] + n_{I,i}[n] \\ Q_i[n] &= A_{Q,i}[n] \sin(\Phi_i[n]) + dc_{Q,i}[n] + n_{Q,i}[n] \end{aligned} \quad (1)$$

where $A_{I,i}$ and $A_{Q,i}$ are the amplitudes of the I/Q signals, $dc_{I,i}$ and $dc_{Q,i}$ are the dc offsets of two channels, and $n_{I,i}$ and $n_{Q,i}$ are the additive noise. The dc offsets $dc_{I,i}$ and $dc_{Q,i}$ can be canceled by the gradient descent algorithm proposed in [39]. In the following discussion, we will omit the channel index i for the single channel analysis.

In baseband, the discrete Doppler phase can be expressed as

$$\Phi[n] = \Phi[0] + \sum_{j=1}^n \Delta\Phi^j \quad (2)$$

where $\Delta\Phi^j$ is the incremental Doppler phase for the j th sampling. Previously, algorithms aiming to linearly retrieve Doppler signals, such as the extended differential and cross multiply (DACM) [35] and the arcsine algorithms [4], have been proposed. In this work, the arcsine algorithm will be used to retrieve $\Delta\Phi^j$, which is

$$\Delta\Phi^j = \arcsin \frac{I[j]Q[j+1] - I[j+1]Q[j]}{\sqrt{I^2[j+1] + Q^2[j+1]} \cdot \sqrt{I^2[j] + Q^2[j]}} \quad (3)$$

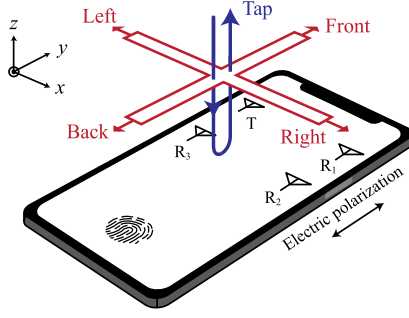
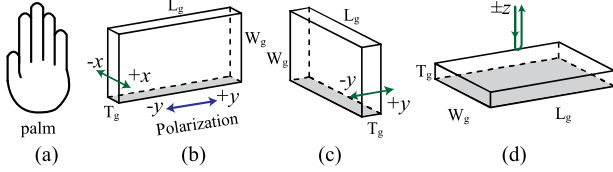


Fig. 2. Definition of the hand gestures.

Fig. 3. Regulation of the hand gestures in y-polarized electric field. (a) Palm. (b) $\pm x$ gesture. (c) $\pm y$ gesture. (d) $\pm z$ gesture.

III. THEORY

A. Definition of Gestures

Fig. 2 shows a smartphone lying on the xy plane. Assuming antennas T, R₁, R₂, and R₃ shown in Fig. 1 are placed on the sides of the smartphone. Normally, such antennas are inverted-F antennas polarized in the y -direction.

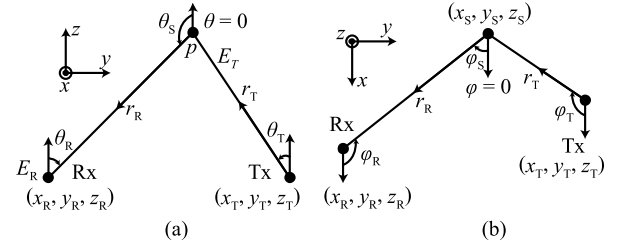
In this work, the hand gestures were defined to realize the major functions of a multitouch screen, i.e., switching between home and subscreens, scrolling pages up and down, and tap actions. For the x - y - z coordinate defined in Fig. 2, the above gestures correspond to the hand movements along the $\pm x$ -, $\pm y$ -, and $\pm z$ -directions over the screen of the smartphone, respectively. They are named “left” (L), “right” (R), “front” (F), “back” (B), and “tap” (TAP) gestures, respectively.

B. Error-Correction Mechanism

In practice, the horizontal gestures can easily deviate from the x - and y -axes, resulting in potential erroneous decisions. To solve this problem, two redundant gestures, i.e., hand movements along the diagonal lines, i.e., $y = x$ and $y = -x$, are defined to maximize the error tolerance. In practice, once a hand gesture is recognized as a diagonal movement, this gesture will be judged as an invalid gesture. Such redundant gestures eliminate the ambiguity regions between the L, R, F, and B gestures, denoted by LF, LB, RF, and RB, respectively.

C. Scattering of a Moving Palm in the Fresnel Near Field

In this work, a 5.8-GHz DRS is used to demonstrate the recognition of such gestures. At this frequency, the corresponding wavelength is 5.2 cm, the same scale as a human hand. In this case, a human palm should be considered as a Mie scatterer, and the analysis should be limited inside the Fresnel near-field range [40].

Fig. 4. Bistatic model for the scattering analysis of a moving palm. (a) yOz view. (b) xOy view.

To simplify the analysis of the Mie scattering of a human hand, we regulate that the gestures will be carried out using a flat palm, as shown in Fig. 3(a). Then, this palm can be approximated as a conductive flat cuboid, moving in the $\pm x$ -, $\pm y$ -, and $\pm z$ -directions with different orientations. In Fig. 3(b)–(d), the length, width, and thickness of the cuboid are denoted by L_g , W_g , and T_g , respectively. The shadowed sides are the ones facing the transmitting and receiving antennas. When the palm is moving in the horizontal and vertical directions, the bottom lateral side and the front side of the cuboid will scatter most of the incident field back to the receiving antennas.

In antenna theory, the scattering of a continuous surface can be considered as the secondary radiation from an aperture antenna. Conventionally, far-field equations such as the Friis equation and the radiation patterns of linear antennas can be conveniently used to analyze such scatterings. However, they are not applicable in the Fresnel near field.

Here, an *ab initio* approach is used to analyze such near-field scattering by integrating the radiation of y -polarized Hertzian dipoles on the rectangular aperture along both the x - and y -directions. In the global x - y - z coordinate shown in Fig. 4, we assume that the phase centers of the transmitting and receiving antenna are located at (x_T, y_T, z_T) and (x_R, y_R, z_R) , respectively. The vertical axes point to the $+z$ -direction. We also assume that the geometric center of the scattering aperture is located at (x_S, y_S, z_S) , with its vertical axis pointing to the $+z$ -direction.

Mathematically, the electric field illuminating the scattering aperture by the transmitting antenna can be expressed as

$$E_T(r_T, \theta_T, \varphi_T) = \hat{y} E_0 G_T(\theta_T, \varphi_T) e^{-jk r_T} / r_T \quad (4)$$

where $G_T(\theta_T, \varphi_T)$ is the radiation pattern of the transmitting antenna and r_T is the distance from transmitting antenna to the element. Similarly, we define $G_R(\theta_R, \varphi_R)$ as the radiation pattern of the receiving antenna, and r_R as the distance from the element to receiving antenna. Here, the far-field radiation patterns can be used because the far-field Fresnel range of quarter-wavelength inverted-F antennas used in smart devices is small.

According to the coordinate systems defined in Fig. 4, for a point on the scattering aperture deviated from (x_S, y_S, z_S) , denoted by $(x_S + \zeta, y_S + \xi, z_S)$, the elevation and azimuth angles and the distances in all the coordinate systems can be

determined as

$$\begin{cases} \theta_T = \arctan\left(\sqrt{(x_S + \zeta - x_T)^2 + (y_S + \zeta - y_T)^2 / (z_S - z_T)}\right) \\ \varphi_T = \arctan[(y_S + \zeta - y_T) / (x_S + \zeta - x_T)] \\ \theta_S = \arctan\left(\sqrt{(x_R - x_S - \zeta)^2 + (y_R - y_S - \zeta)^2 / (z_R - z_S)}\right) \\ \varphi_S = \arctan[(y_R - y_S - \zeta) / (x_R - x_S - \zeta)] \\ \theta_R = \arctan\left(\sqrt{(x_S + \zeta - x_R)^2 + (y_S + \zeta - y_R)^2 / (z_S - z_R)}\right) \\ \varphi_R = \arctan[(y_S + \zeta - y_R) / (x_S + \zeta - x_R)] \\ r_T = \sqrt{(x_S + \zeta - x_T)^2 + (y_S + \zeta - y_T)^2 + (z_S - z_T)^2} \\ r_R = \sqrt{(x_S + \zeta - x_R)^2 + (y_S + \zeta - y_R)^2 + (z_S - z_R)^2}. \end{cases} \quad (5)$$

Illuminated by the y -polarized incidence, the current induced on the rectangular scattering aperture satisfies a sinusoidal distribution in the y -direction [40]. It comprises infinite line currents along the x -direction, each with a surface current distribution of

$$J(\zeta, \xi) = I_0 \sin\left[k\left(\frac{l}{2} - |\xi|\right)\right] \quad (6)$$

where the amplitude $I_0 = \sigma_e E_T(r_T, \theta_T, \varphi_T)|_{\zeta=0, \xi=0}$, σ_e being an equivalent conductivity. This current can be further divided into infinite y -polarized Hertzian dipoles. For the Hertzian dipole located at $(x_S + \zeta, y_S + \zeta, z_S)$, its y -component far-field radiation can be derived as (given in the Appendix)

$$\begin{aligned} dE_S^y(\zeta, \xi, r, \theta, \varphi) \\ = \frac{-jk\eta_0 J(\zeta, \xi) e^{-jkr}}{4\pi r} (\cos^2 \theta \sin^2 \varphi + \cos^2 \varphi) d\zeta d\xi. \end{aligned} \quad (7)$$

Then, the total field scattered by the rectangular aperture can be obtained by integrating all the radiations of such Hertzian dipoles along both the x - and y -directions, and the electric field detected by the receiving antenna can be derived as

$$E_R = \int_{-l/2}^{l/2} \int_{-w/2}^{w/2} G_R(\theta_R, \varphi_R) \cdot dE_S^y(\zeta, \xi, r_R, \theta_R, \varphi_R). \quad (8)$$

Fig. 5(a)–(c) shows the θ -dependent amplitude and phase radiation patterns calculated based on (6) at a near-field Fresnel range of 20 cm for the “left/right,” “back/front,” and “tap” gestures, respectively, with L_g , W_g , and T_g being 15, 8, and 2 cm, respectively.

It is seen that the main lobes in all the amplitude patterns are wide, benefiting the coverage of all the receiving antennas. In Fig. 5(a), the phase shift along the $\pm x$ (or $\varphi = 0/180^\circ$, $\theta = 90^\circ$) directions due to the y -polarized illumination is nearly constant and immune to “left/right” motions. Similar flat phase shift regions can also be found in Fig. 5(b) and (c) for the $\pm y$ (or $\varphi = 90^\circ/-90^\circ$, $\theta = 90^\circ$) and $\pm z$ (or $\theta = 0/180^\circ$) directions, which are suitable for the “back/front” and “tap” gestures, respectively. As shown in the following, such flat phase responses will result in quasi-linear, deterministic Doppler phases, which is very important for HGRs. This is also the basis on which we defined the above gestures.

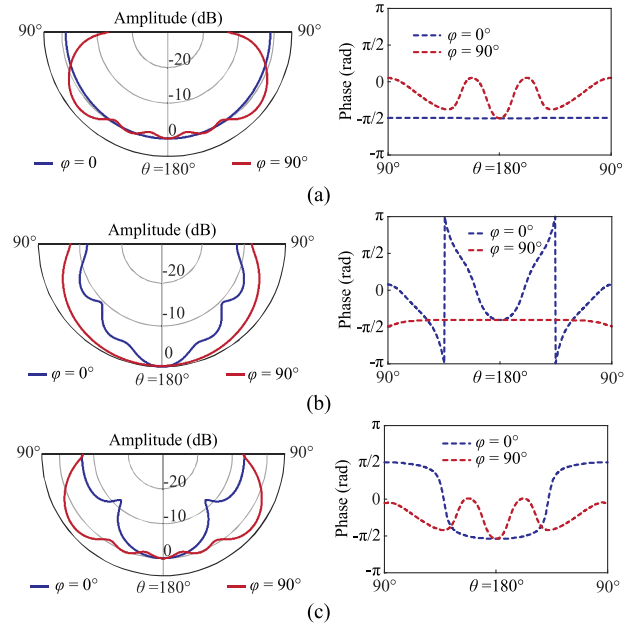


Fig. 5. Scattering patterns calculated by (6) with different aperture size. (a) $w = 2$ cm, $l = 15$ cm. (b) $w = 15$ cm, $l = 2$ cm. (c) $w = 8$ cm, $l = 15$ cm.

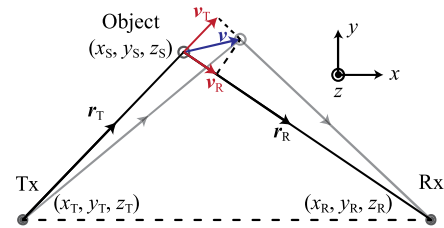


Fig. 6. Doppler phases detected by a bistatic DRS.

D. Doppler Signals Detected by a Bistatic Radar

A SIMO DRS can be treated as a combination of several bistatic radars that share the same transmitting channel. Fig. 6 illustrates a bistatic radar used to detect a moving object. An object moves from (x_S, y_S, z_S) with a velocity $\mathbf{v} = (v_x, v_y, v_z)$ and a displacement $\mathbf{p} = (p_x, p_y, p_z)$. The radial velocities v_T and v_R are the projections of \mathbf{v} onto vectors \mathbf{r}_T and \mathbf{r}_R , and thus the roundtrip distance $r(t) = |\mathbf{r}_T(t)| + |\mathbf{r}_R(t)|$.

In a short interval Δt where $\Delta \mathbf{p}(t) = (v_x(t), v_y(t), v_z(t)) \Delta t$

$$r(t + \Delta t) = \sqrt{|\mathbf{r}_T(t) + \Delta \mathbf{p}(t)|^2} + \sqrt{|\mathbf{r}_R(t) - \Delta \mathbf{p}(t)|^2}. \quad (9)$$

Then, the derivative of $r(t)$ can be calculated as (given in the Appendix)

$$\begin{aligned} \frac{dr(t)}{dt} &= \lim_{\Delta t \rightarrow 0} \frac{r(t + \Delta t) - r(t)}{\Delta t} = v(t) \cdot \hat{\mathbf{r}}_T - v(t) \cdot \hat{\mathbf{r}}_R \\ &= |v_T(t)| - |v_R(t)|. \end{aligned} \quad (10)$$

According to the Doppler formula [41], the Doppler frequency shift can be calculated as

$$f_d(t) = (|v_R(t)| - |v_T(t)|) f_c / c \quad (11)$$

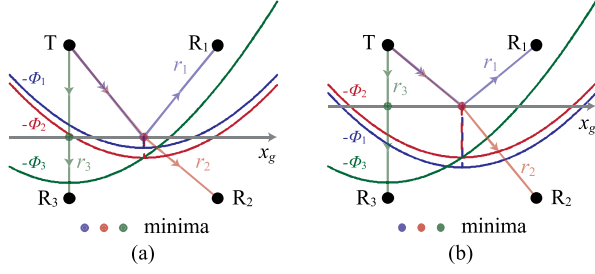


Fig. 7. Variations of Doppler signals detected by the TDM-SIMO DRS in “right” gestures. In the calculations, $a = 2.5$ cm, $h = 10$ cm, x_g represents the gesture moving from -4.5 to 4.5 cm, y_1 and y_2 are the y coordinates of “right” gestures where (a) $y_1 = -0.5$ cm and (b) $y_2 = 0.5$ cm in the examples.

where f_c is carrier frequency and c is the speed of light. Based on (10) and (11)

$$r(t) = -\frac{c}{2\pi f_c} \int 2\pi f_d(t) dt = -\frac{c}{2\pi f_c} \Phi(t). \quad (12)$$

Equation (12) shows the equivalency between the roundtrip distance $r(t)$ and the Doppler signal $-\Phi(t)$.

E. Deterministic Doppler Signals Carried by $\Phi(t)$

For antennas T, R₁, R₂, and R₃ arranged in Fig. 2, we assume that their coordinates are $(-a, a, 0)$, $(a, a, 0)$, $(a, -a, 0)$, and $(-a, -a, 0)$, respectively, and a “right” gesture moves on the $z = h$ plane. Its trajectory can be described by $X(x_g, y, h)$, where x_g is the x -coordinate of the scattering phase center of this gesture and y being the constant y -coordinate of the gesture. The roundtrip distances between the transmitter, moving gesture, and the TDM receivers can be expressed as

$$\begin{cases} r_1 = \sqrt{(x_g + a)^2 + (y - a)^2 + h^2} \\ \quad + \sqrt{(x_g - a)^2 + (y - a)^2 + h^2} \\ r_2 = \sqrt{(x_g + a)^2 + (y - a)^2 + h^2} \\ \quad + \sqrt{(x_g - a)^2 + (y + a)^2 + h^2} \\ r_3 = \sqrt{(x_g + a)^2 + (y - a)^2 + h^2} \\ \quad + \sqrt{(x_g + a)^2 + (y + a)^2 + h^2}. \end{cases} \quad (13)$$

Based on (12) and (13), the corresponding Doppler signals $-\Phi_i(t)$ can be calculated, as shown in Fig. 7, where $i = 1, 2, 3$.

It is seen that for the “right” gestures, the minimum of $-\Phi_3(t)$ appears first. Then, those of $-\Phi_1(t)$ and $-\Phi_2(t)$ follow. For antennas located on the corners of a square, the minima of $-\Phi_1(t)$ and $-\Phi_2(t)$ will appear simultaneously. Such deterministic but different Doppler signals also appear for other gestures. In this work, the temporal order of such minima of $-\Phi_i(t)$ will be used in robust gesture recognition.

In (13), h also represents the height of the gestures. When h is changed, the shape of the $-\Phi_i(t)$ curves will also be changed. However, the temporal order of the minima will remain unchanged. It means that gesture recognitions based on the temporal order of such minima will be independent on the height of gestures.

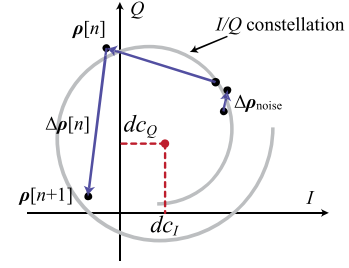


Fig. 8. Illustration of $\Delta \rho[n]$ and $\Delta \rho_{\text{noise}}$ in the I/Q constellation.

F. Segmentation of Detected Signals

To recognize the predefined gestures, the signals containing all the motion information, i.e., $\Phi[n]$, should be isolated first. This can be achieved by a dynamic power analysis to the detected I/Q signals.

Fig. 8 illustrates the vector I/Q signals at sampling n and $n + 1$. The variation of the I/Q signals during the adjacent sampling is denoted by the vector $\Delta \rho[n]$. In the presence of a moving scatterer,

$$\begin{aligned} \Delta \rho[n] &= \begin{pmatrix} A_I[n+1] \cos(\Phi[n+1]) - A_I[n] \cos(\Phi[n]) + \Delta n_I[n] \\ A_Q[n+1] \sin(\Phi[n+1]) - A_Q[n] \sin(\Phi[n]) + \Delta n_Q[n] \end{pmatrix}^T \\ &= \begin{pmatrix} A_I[n+1] \cos(\Phi[n+1]) - A_I[n] \cos(\Phi[n]) + \Delta n_I[n] \\ A_Q[n+1] \sin(\Phi[n+1]) - A_Q[n] \sin(\Phi[n]) + \Delta n_Q[n] \end{pmatrix}^T \end{aligned} \quad (14)$$

reflecting the “differential power” $\Delta P[n]$ of the I/Q signals, where $\Delta n_I[n]$ is $n_I[n+1] - n_I[n]$, so it is with $\Delta n_Q[n]$. Here, $\Delta P[n]$ can be calculated as

$$\Delta P[n] = |\Delta \rho[n]|^2. \quad (15)$$

In the absence of any motion, $\Delta \rho_{\text{noise}}[n] = (\Delta n_I[n], \Delta n_Q[n])$, representing the “differential power” of the circuit noise. Clearly, $\Delta \rho$ is always larger than $\Delta \rho_{\text{noise}}$, and thus, ΔP can be effectively used to separate the segments containing the gesture information from the detected signals.

To determine the threshold of the separation, the mathematical expectation of $\Delta \rho_{\text{noise}}$, i.e., $E(\Delta \rho_{\text{noise}})$, can be determined, based on which a value slightly larger than this expectation can be chosen as the threshold. In this work, this value is $1.1 E(\Delta \rho_{\text{noise}})$.

To ensure the reliability of the separation, an m -point moving average window can be performed to $\Delta P[n]$ at each time interval with

$$\overline{\Delta P_m}[n] \triangleq m^{-1} \sum_{j=n-m/2+1}^{n+m/2} \Delta P[j] \quad (16)$$

to reduce potential erroneous judgments due to unexpected glitches. If $\overline{\Delta P_m}[n]$ is larger than the threshold, a motion will be considered to exist inside this time interval. All such continuous time intervals correspond to a signal segment, during which a motion is considered to exist. The same operation can be performed to signals detected by all the receiving channels. The obtained three signal segments reflect the same motion detected from different orientations. The union of such segments will result in a widest segment that

containing the information of a complete motion, from which a potential gesture can be recognized.

G. Data Alignment for TDM Samplings

The utility of the TDM-SIMO architecture shown in Fig. 1 significantly reduces the complexity and cost of the DRS. However, the sampled data sequences need to be aligned.

For M TDM receiving channels, the detected serial I -channel sequences $I_1[n]$, $I_2[n + 1/M]$, ..., $I_M[n + (M-1)/M]$, $I_1[n + 1]$, $I_2[n + 1 + 1/M]$, ... should be realigned into $I_1[n]$, $I_2[n]$, ..., $I_M[n]$, $I_1[n + 1]$, $I_2[n + 1]$, ... The difference between the detected and aligned signals for the i th channel is $(i-1)T_{\text{TDM}}/M$, where T_{TDM} is the interval between adjacent TDM sampling. In this work, a set of digital phase shift filters [42] are designed to compensate this temporal asynchrony, so that the z -domain response of this filter for i th channel can be obtained as is $H_i(z) = z^{(i-1)/M}$. Its frequency response is

$$H_i(e^{j\omega}) = H_i(z)|_{z=e^{j\omega}} = e^{j(i-1)\omega/M} \quad (17)$$

and such all-pass phase shift filter can be synthesized as

$$\begin{aligned} h_i(n) &= \frac{1}{2\pi} \int_{-\pi}^{\pi} e^{(i-1)j\omega/M} e^{j\omega k} d\omega \\ &= \frac{\sin\{\pi[n + (i-1)/M]\}}{\pi[n + (i-1)/M]}. \end{aligned} \quad (18)$$

After convoluting $I_i[n]$ with $h_i(n)$ correspondingly, the data sampled from different I -channels can be aligned. Similarly, the detected Q -channel signals can also be aligned.

H. Rescaling and Fitting of Segmented Gestures

After the segmentation and data alignment, the consecutive temporal segments, each containing a single gesture, can be obtained. For the widest gesture segment, its actual duration, or the total number of samples, varies from gesture to gesture. In addition, the detected original Doppler phase signal Φ_i can be zigzagged and, therefore, rescaling and fitting to the segmented gesture signals are needed, to obtain regulated gesture signals with the same length and smoothed shapes for further recognition.

In this work, all the widest segments are rescaled into (0, 1). To maintain the shape of $-\Phi_i$ and prevent the impact from glitches, in this work, an eighth-order polynomial fitting is adopted so that the noise on signal $-\Phi_i$ would not be fit.

After the rescaling and fitting, the moments of the minima of $-\Phi_i$ for the i th channel discussed in Section II-E can be quantified between the interval (0, 1) and denoted by $t_{m,i}$. In this work, the moments for the minima are denoted in vector form as $\mathbf{t}_m = (t_{m,1}, t_{m,2}, t_{m,3})$, which can characterize each gesture and will be the input of the classifier discussed in the following.

After this rescaling, the minima moments become independent on the speed of gestures. It plays an important role in recognizing both fast and slow gestures.

I. Classification-Based Gesture Recognition

Classification is one of the commonly used techniques in machine learning [23], aiming to categorize a set of data

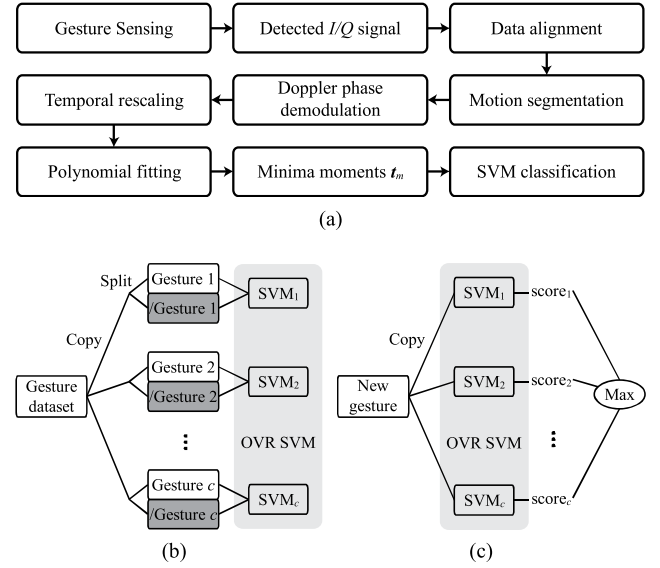


Fig. 9. Flowchart of the gesture recognition. (a) Signal processing. (b) Training of the OVR SVM. (c) Gesture recognition based on trained OVR SVM.

into classes. It has been used in image recognition, speech processing, bioinformatics, natural language processing, and so on [24].

So far, a variety of classification algorithms, or classifiers, has been proposed. Mainstream classifiers include logistic regression, linear discriminant analysis (LDA), decision trees, naïve Bayes classifier, and support vector machine (SVM) [43].

In this work, the one versus rest (OVR) SVM designed for multiclassification is chosen to recognize the predefined multiple gestures. The flowcharts of the signal processing, the training of the OVR SVM based on known gesture data set, and the recognition for unknown gestures are shown in Fig. 9(a)–(c), respectively.

To train the OVR SVM, the data set comprises known gestures can be prepared by detecting each defined gesture for N times. Then, this data set can be duplicated into c copies. Here, c is the number of the defined gestures. In the training of the i th SVM, all the i th gestures are labeled as $+1$, and all the rest are labeled as -1 .

Based on the trained SVMs, a new gesture can be recognized by comparing the “score,” or similarity, evaluated by each SVM. If the highest score was evaluated by the i th SVM, this gesture will be recognized as the i th one. Details can be found in [23].

IV. SIMULATIONS

To verify the effectiveness of the methods proposed in Section III, we ran numerical simulations based on derived equations following the steps shown in Fig. 9. In all the simulations, the coordinates of the transmitting and receiving antennas T , R_1 , R_2 , and R_3 are $(-2.5, 2.5, 0)$, $(2.5, 2.5, 0)$, $(2.5, -2.5, 0)$, and $(-2.5, -2.5, 0)$; L , W , and T of the simplified palm are 15, 8, and 2, all in cm, respectively.

A. Doppler Signals Detected of Different Palm Orientation

The defined gestures can be classified into two types. For the “left,” “right,” “front,” “back,” and “diagonal” gestures,

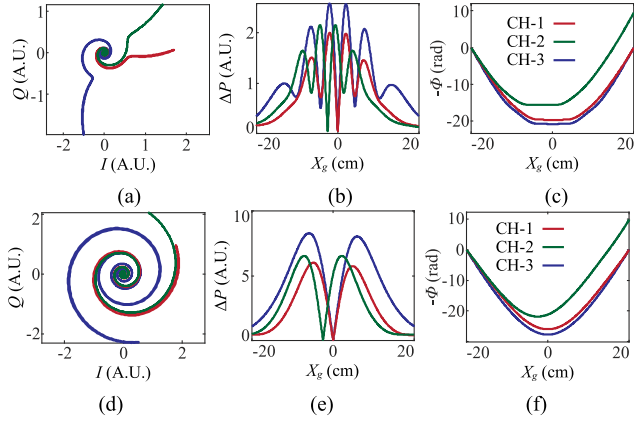


Fig. 10. Doppler signals simulated for different orientations of the palm. (a)–(f) are the constellation of the I/Q signals, the differential power for the segmentations, and the retrieved Doppler phases for the parallel and vertical palms, respectively.

the palm is vertical to the antenna plane, scattering the illuminating field with a narrow aperture. For the “tap” gesture, the palm is parallel to the antenna plane, scattering with a much wider aperture.

To explain why these two types of gestures were defined, we compared the Doppler signals calculated for “right” gestures performed with both vertical and parallel palms. In the calculation, the centers of both scattering apertures, i.e., (X_g, y, h) , move from $(-20, 0, 10)$ to $(20, 0, 10)$, respectively. Based on (2)–(8), the constellation of the I/Q signals, the differential power, $\Delta P[n]$, for the segmentations, and the retrieved Doppler phases, $-\Phi[n]$, are plotted in Fig. 10.

It can be seen that the moving vertical and parallel palms can result in very different I/Q signals and Doppler phases. Especially, for the parallel moving palm, the minima in Fig. 10(c) can hardly be determined. However, the minima in Fig. 10(f) can be easily identified. This is the consideration of the definition of gestures moving parallel above the antenna plane.

B. Simulations Following the Flowchart

In this section, numerical simulations will be performed following the steps shown in Fig. 9(a).

First, using the same method applied in calculating Fig. 10, the I/Q signals resulting from two successive “right” and “left” gestures “detected” by the R_1 , R_2 , and R_3 antenna can also be calculated. The results are shown in Fig. 11. The difference is that a white Gaussian noise is added so that the I/Q signals can be generated with a 20-dB signal-to-noise ratio (SNR).

Then, such signals are segmented according to Section III-F. Panels I and II in Fig. 12 show the demodulated Doppler phases and the 60-point moving averaged differential power $\Delta P_{60}[n]$, respectively. Choosing the threshold as 2, the gestures can be segmented in regions marked in gray. It is seen that both gestures are properly segmented. Panel III shows the same calculation as panel II under a worse 10-dB SNR condition. It is seen that the gesture segments can be hardly identified.

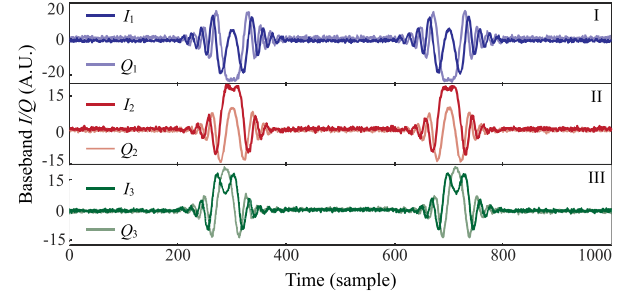


Fig. 11. Simulated I/Q signals result from two successive “right” and “left” gestures “detected” by the R_1 , R_2 , and R_3 antennas, respectively.

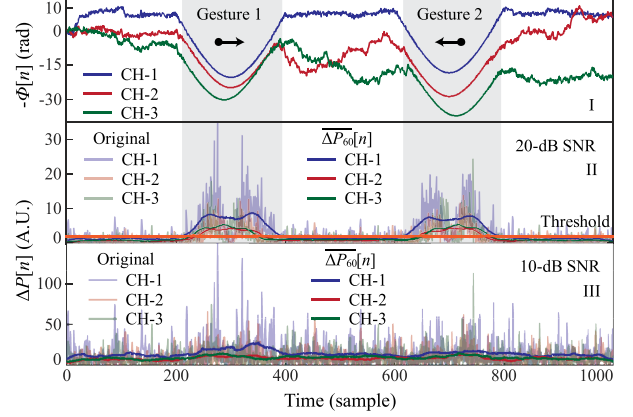


Fig. 12. Segmentations of two successive “right” and “left” gestures. Panel I, demodulated Doppler phase under 20-dB SNR. Panel II, $\Delta P[n]$ and $\Delta P_{60}[n]$ of simulated baseband signals under 20-dB SNR, as well as a threshold for segmentation. Panel III, $\Delta P[n]$ and $\Delta P_{60}[n]$ of simulated baseband signals under 10-dB SNR.

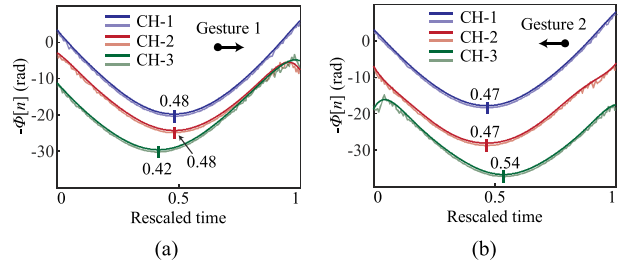


Fig. 13. Determination of Doppler minima in the rescaled and fit Doppler phases for the successive “right” and “left” gestures. Dark lines: eighth-order polynomial fit Doppler phase. Light lines: original Doppler phase with noises. The quantified minima moments for (a) gesture 1 are $t_m = (0.48, 0.48, 0.42)$, and the that for (b) gesture 2 is $t_m = (0.47, 0.47, 0.54)$.

Next, the segmented gestures are rescaled based on the discussion in Section III-H with an eighth-order polynomial fitting. After the rescaling and fitting, as shown in Fig. 13, the minima’s moments of the rescaled Doppler phases can be clearly extracted, based on which the temporal sequential order of these minima can also be determined. In this work, such deterministic, quantified minima in the order of channels, e.g., $t_m = (0.48, 0.48, 0.42)$ and $t_m = (0.47, 0.47, 0.54)$ in this figure, are the inputs of the OVR SVM classifier, instead of the I/Q constellations or the Doppler signatures of used in the “black-box” gesture classifications.

Fig. 14 shows the Doppler phases segmented from I/Q signals simulated for all the defined gestures. In the simulation,

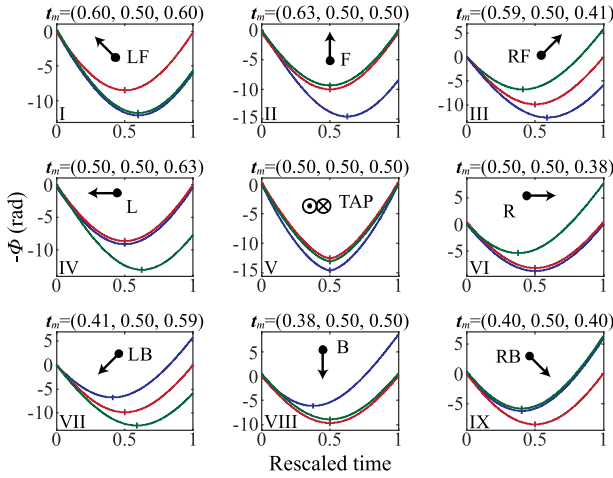


Fig. 14. Simulated Doppler phases of defined gestures detected by different receiving channels.

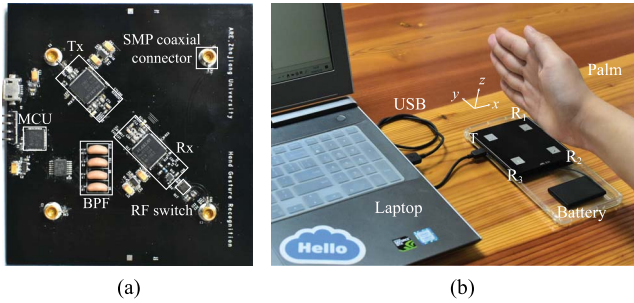


Fig. 15. TDM-SIMO DRS implemented with a pair of Wi-Fi chips. (a) RF board. (b) Battery powered DRS and the experimental environment.

for all the gestures moving in horizontal directions, the distance between the bottom side of the hand and the antenna plane is set as 10 cm, and for the tap gesture, the shortest distance between the palm and the antenna plane is set as 5 cm. It is seen that the temporal sequential orders of minima for each segmented gesture is deterministic.

V. EXPERIMENTS

A. Implementation of the TDM-SIMO DRS

Fig. 15 shows the TDM-SIMO DRS implemented according to the architecture proposed in Fig. 1. In this work, a pair of Wi-Fi chips, Maxim Integrated MAX2829s, were used to emulate the 5.8-GHz LTE-U and Wi-Fi channels, and an RF switch, HMC7992, was used to switch the receiving channel to antennas R_1 , R_2 , and R_3 , respectively. Similar to [44], the L-shaped arrangement of receiving antennas was used to detect 3-D gestures. Without losing any generality, such antennas were designed as standard patch antennas. The antenna and the RF boards were connected by four sub-miniature push-on connector (SMP) coaxial connectors. A standard lithium battery was used to supply power to the DRS. An STM32F103T8U6 micro-controller unit (MCU) was used for Wi-Fi chips configuration and I/Q signals acquisition by the MCU with its built-in ADCs. The sampled data were finally transferred to a laptop via an universal serial bus (USB)-universal asynchronous receiver/transmitter (UART) bridge chip, Silicon Labs CP2102, mounted on the RF board.

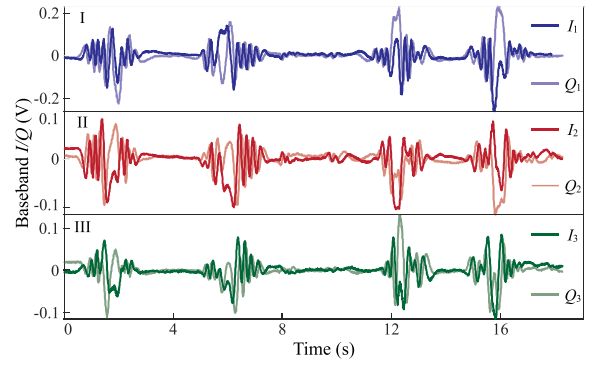


Fig. 16. Experimental I/Q signals result from four successive “front,” “back,” “left,” and “right” gestures detected by antennas R_1 , R_2 , and R_3 , respectively.

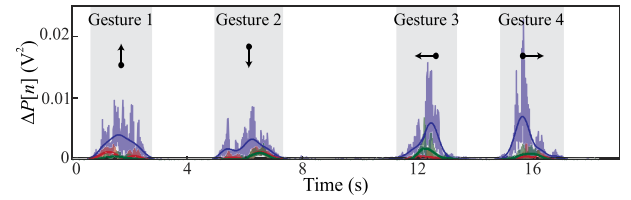


Fig. 17. Segmentation process on the signals detected in Fig. 16 based on the differential power $\Delta P[n]$ as well as moving averaged $\Delta P_{60}[n]$.

B. Processing of Experimental Gesture Signals

Corresponding to Figs. 11–13, the panels I–III in Fig. 16 show the baseband I/Q signals detected by antennas R_1 , R_2 , and R_3 , respectively, for successive “front,” “back,” “left,” and “right” gestures, respectively. Using the same approach for calculating Figs. 12 and 17 shows the segmented gestures detected in Fig. 16. Fig. 18 shows the demodulated Doppler phases and their minima moments t_m for further classification.

It is seen that for experimental data, taking the “right” gesture as the example, the minima of $-\Phi_1(t)$ and $-\Phi_2(t)$ can appear at slightly different times due to the nonplanar scattering aperture of the actual hand. However, if the data set used to train the classifier included such deviated minima, it has few impacts on the correction of the classification.

Corresponding to Fig. 14, the same nine gestures were performed with the same distances to the antenna plane, and the constellations of I/Q signals detected for such gestures by the TDM-SIMO DRS are shown in Fig. 19, respectively. It can be noticed that both the amplitude and the shape of the “tap” gesture are significantly different from the others due to the vertical movements and the larger scattering area of the palm.

Fig. 20 shows the $-\Phi_i$ retrieved from the constellations plotted in Fig. 19. It is seen that the sequential orders of the minima moments t_m are consistent with those in Fig. 14.

C. Gesture Classifications

To classify the defined gestures, we collected 500 samples for different gestures as the data set, 50–60 samples for each gesture. Among them, 80% in the 500 samples (around 40–50 samples for each gesture) were randomly

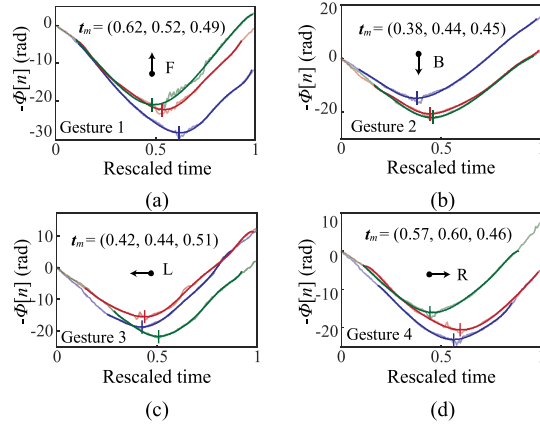


Fig. 18. Determination of Doppler minima of the rescaled and fit Doppler phases for the four successive gestures in Fig. 16. Dark lines: eighth-order polynomial fit Doppler phase curves. Light lines: original Doppler phase curves. The quantified minima moments for the four gestures are (0.62, 0.52, 0.49), (0.38, 0.44, 0.45), (0.42, 0.44, 0.51), and (0.57, 0.60, 0.46), respectively.

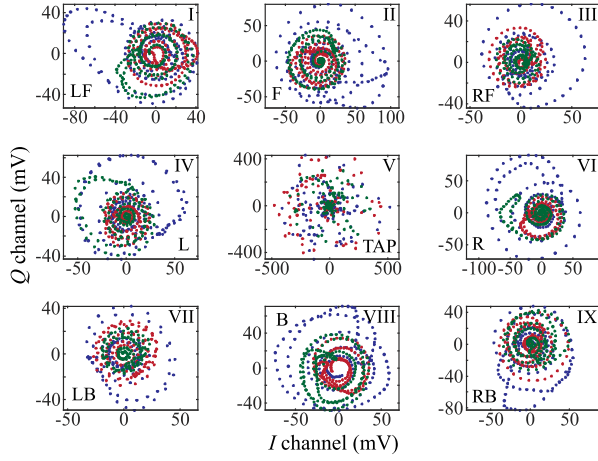


Fig. 19. Experimental constellations of performed gestures.

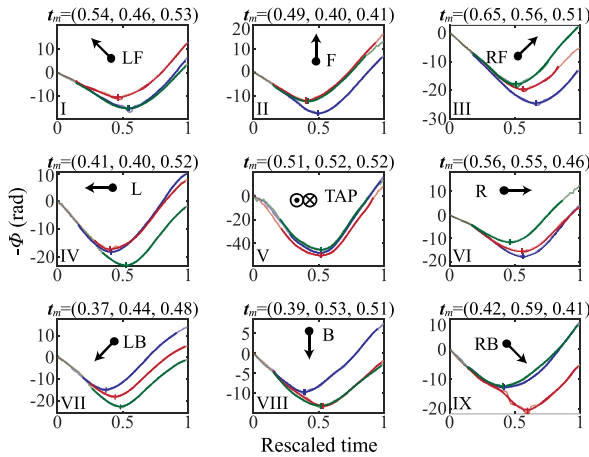


Fig. 20. Experimental Doppler phase variations and minima moments of performed gestures.

chosen as the training set. The rest 20% were chosen as the testing set, including unknown gestures to be recognized. To visualize the minima moments t_m in the training and testing sets, a rectangular coordinate is established, whose axes

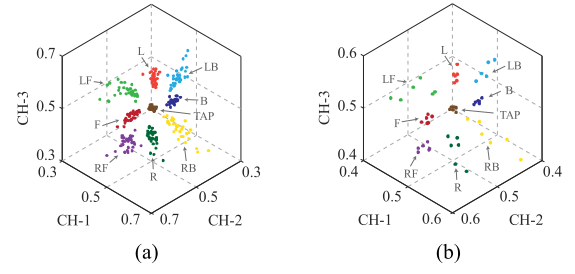


Fig. 21. Mapping of Doppler phase minima moments t_m of all kind of gesture into the 3-D feature coordinate. The axes represent the rescaled time for channel 1–3. (a) Training set. (b) Testing set.

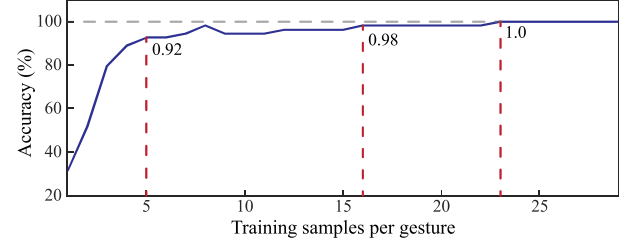


Fig. 22. Learning curve of the OVR SVM evaluated by the unknown samples.

correspond to the minima moments detected by each receiving channel. As shown in Fig. 21, the minima moments of each kind of gesture are clustered and clearly separated from other clusters in both data sets, verifying the determinacy of gestures represented by the minima moments. It also implies that using the OVR SVM trained by the training set, the gestures in the testing set can be accurately classified.

To evaluate the robustness of the minima moments-based classification, we calculate the learning curve [23] by classifying 50 unknown gestures based on the OVR SVM trained by different subsets of the data set. The result is shown in Fig. 22. It is seen that a 98% recognition accuracy can be recognized by the OVR SVM trained by only six samples, and the 100% accuracy can be obtained when the training samples were increased to 23. The training only costed less than 3 s using a regular personal computer.

To verify the generalizability of this classifier, a one-subject-leave-out experiment [12] was conducted by inviting four additional subjects to perform the same gestures. Each gesture is repeated for 10 times. Before performing, they were aware of the moving direction of each gesture, the posture of the hand, and the sensitive region. Fig. 23 shows the mapping of the data sets consisting of the minima moments detected from these subjects. It is seen that compared with the mapping of the training set shown in Fig. 21(a), these minima moments are also clustered and clearly separated. It indicates that although differences exist between the gestures performed by different subjects, the rescaled minima moments exhibit sufficient similarity for the OVR SVM classifier.

Using the classifier trained by the data set detected from the original subject, the minima moments detected from additional subjects shown in Fig. 23 were classified. The recognition accuracies 96.67%, 100%, 97.78%, and 96.67% can be obtained by calculating the confusion matrix for

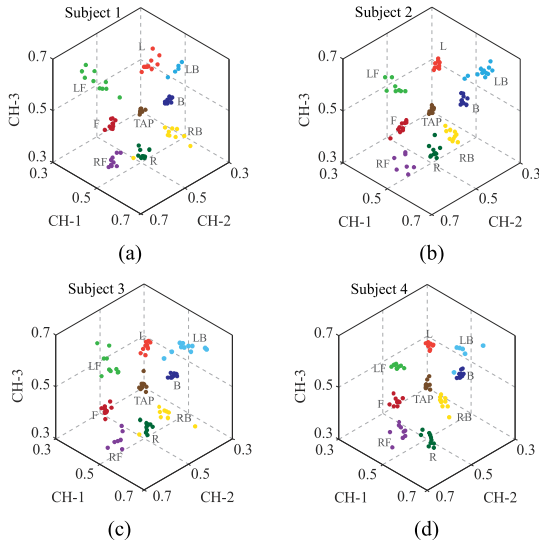


Fig. 23. Mapping of the minima moments detected from the new subjects.

Tr \ Pr	F	B	RF	RB	R	LF	LB	L	TAP
F	24	0	0	0	0	0	0	0	0
B	0	24	0	0	0	0	0	0	0
RF	0	0	22	0	0	0	0	0	2
RB	0	0	0	24	0	0	0	0	0
R	0	0	0	0	24	0	0	0	0
LF	0	0	0	0	0	24	0	0	0
LB	0	0	0	0	0	0	24	0	0
L	0	0	0	0	0	0	0	23	1
TAP	0	0	0	0	0	0	0	0	24

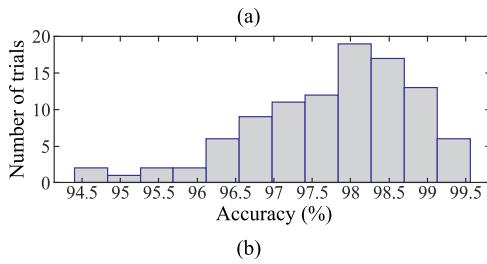


Fig. 24. Accuracy evaluation on the trained OVR SVM. (a) Confusion matrix, in which Tr and Pr stand for “truth” and “prediction,” respectively. (b) Histogram of cross-validation accuracy in 100 trials.

each new subject, respectively. This result verifies that the classifier trained by a certain subject is applicable for unknown subjects.

Finally, the cross-validation [23] experiment was conducted, in which the gestures detected from all the five subjects were mixed. Among the mixed gestures, 75% gestures were randomly chosen as the training set, and the remaining 25% ones were used as testing set. For each gesture, the number of gestures in the testing set is 24. Fig. 24 shows the confusion matrix calculated for gestures performed by all the five subjects attended the one-subject-leave-out experiment. It can be calculated that the total recognition accuracy is up to 98.6%. Also, Fig. 24(b) shows the histogram of cross-validation accuracies in 100 trials. It is calculated that the maximum, minimum, and average accuracies are 99.54%, 94.44%, and 97.83%, respectively, with a standard deviation of 1.15%.

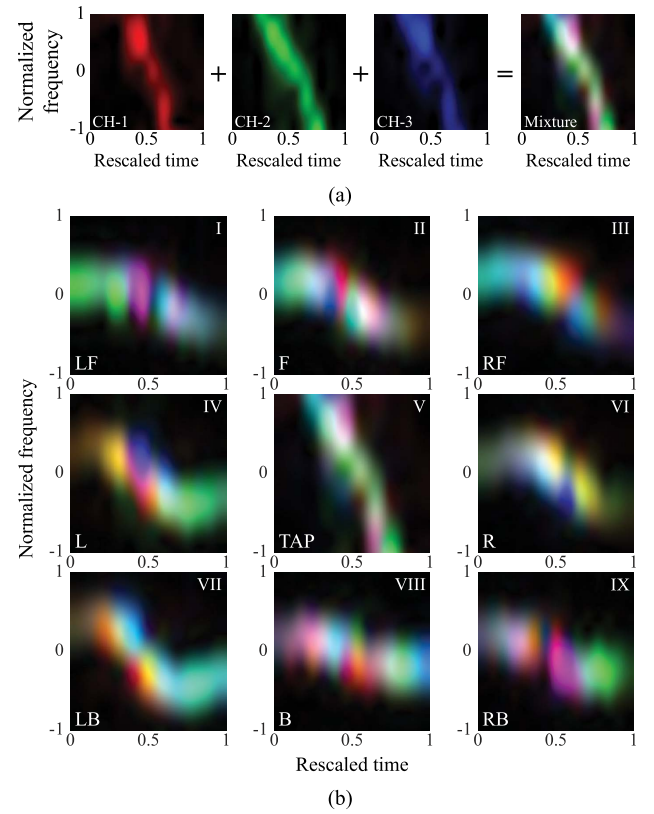


Fig. 25. Time-frequency spectrograms resulted from different gestures. (a) Monochromatic and panchromatic spectrograms. (b) Panchromatic Doppler spectrograms used for DCNN-based classification.

D. Discussion

The above results verified the effectiveness and accuracy of the approach proposed for smart devices based on detecting and processing of deterministic, multichannel Doppler signals.

As the consumer products, smart devices, especially smartphones, are highly integrated, compact, and extremely concern about battery life, product size, and cost. Working at a single frequency, this CW solution can be implemented by reusing the existing RF resources of smart devices, increasing negligible cost, power consumption, and space occupation.

A unique characteristic resulted from the determinacy of the Doppler signals is the high recognition accuracy achieved by the highly efficient OVR SVM classifier trained by a very small data set. In comparison, much more training samples would be needed in the gesture recognition based on classifying the time-frequency spectra of the Doppler signals.

For comparison, Fig. 25(a) shows the short-time Fourier transform (STFT) of the Doppler phases detected by each channel of the SIMO DRS. In the calculation of this STFT spectrogram, a 20-point Hanning sliding window is used for each channel, and the number of overlapping points is 10. Since the temporal duration of each gesture is different, the spectrogram is resized into 400×400 pixels. Then, the spectrogram intensity of each channel is normalized into (0, 1).

The obtained time-frequency spectra were plotted into monochromatic red, green, and blue colors. Then, they are mixed to obtained a panchromatic spectrogram. Fig. 25(b)

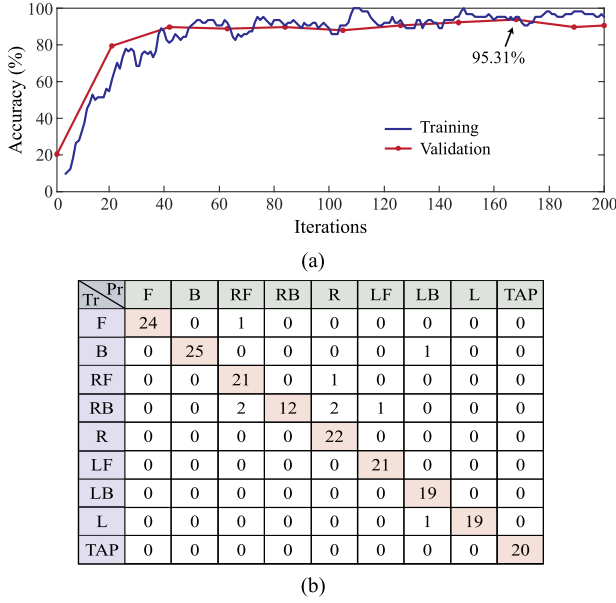


Fig. 26. Comparison to the DCNN-based classification. (a) Training of the DCNN classifier. (b) Confusion matrix.

shows the spectrograms of the nine different gestures calculated based on the signals shown in Fig. 19. In previous Doppler radar-based recognitions [10], [12], [14], [16], [17], [25], such “images” were used to train the classifiers, normally a deep convolutional neural network (DCNN), and then unknown gestures can be recognized by the trained DCNN.

As an example, 560 spectrograms of detected gestures were used to train the “MobileNets” DCNN intended for mobile and embedded devices [45], and used a mini-batch size of 8 to avoid potential overfitting. Such spectrograms come from the mixture of all the five subjects. Then, 192 unknown spectrograms were classified by the trained DCNN so that the recognition accuracy with respect to the iteration time can be obtained.

As seen in Fig. 26(a), the recognition accuracy increases overall with the increased iteration number. The final recognition accuracy reaches 95.31%. Using a regular computer, the training of the DCNN based on the spectrograms costed around 8 min, much longer than the 3 s costed in the training of the OVR SVM based on the sequenced minima moments. While the DCNN would require a much larger training set, it is difficult to reach a 100% recognition accuracy. Based on the trained DCNN, the confusion matrix can be calculated, as shown in Fig. 26(b).

Another challenge for practical gesture recognitions is to deal with interfering random gestures and motions. In this work, while the OVR SVM can help to filter out some of them, this problem is solved during the signal segmentation. As shown in Fig. 20, for each qualified gesture, since the segmented curve of the Doppler phase detected by each receiving antenna is convex, there is only one minimum moment. Therefore, the random gestures or motions can be excluded by simply examining the number of minima moments in segmented signals.

Finally, the gestures defined in this work correspond to the major functions of multitouch screens. The purpose is to provide additional wireless gesture control functions for smart devices. The Doppler phase curves of such gestures are convex functions. Our method can also be extended to bidirectional and multidirectional gestures, whose Doppler phase curves consist of connected convex functions. Such gestures can be a combination of gestures performed in the horizontal or vertical plane. However, the existing method does not apply for gestures from which sequenced minima moments cannot be retrieved. For these gestures such as finger rubs, micro-Doppler-based analyses may be indispensable.

VI. CONCLUSION

In conclusion, we theoretically analyzed and experimentally demonstrated HGR based on a SIMO-DRS that can be implemented in smart devices by slightly modifying their existing RF front ends. Based on detecting and processing deterministic Doppler signals detected by the SIMO radar system, a small data set can be used to train an OVR SVM and realize millisecond recognition of nine predefined gestures. We envision this robust, highly efficient approach can also be extended to a variety of applications.

APPENDIX

A. Derivation of (7)

For a y-polarized Hertzian dipole, denote vector potential as

$$\mathbf{A} = \hat{\mathbf{y}} \mu I \Delta l e^{-jkr} / 4\pi r \quad (\text{A.1})$$

where Δl and I are an infinitesimal length and a linear current intensity, respectively. In an r - θ - φ spherical coordinate

$$\hat{\mathbf{y}} = \hat{\mathbf{r}} \sin \theta \sin \varphi + \hat{\boldsymbol{\theta}} \cos \theta \sin \varphi + \hat{\boldsymbol{\phi}} \cos \varphi. \quad (\text{A.2})$$

Then, the electric field

$$\begin{aligned} \mathbf{E} &= \frac{1}{j\omega\epsilon} \nabla \times \mathbf{H} = \frac{1}{j\omega\epsilon} \nabla \times \left(\frac{1}{\mu} \nabla \times \mathbf{A} \right) \\ &= \frac{1}{j\omega\epsilon} \frac{I \Delta l}{4\pi} \left[\frac{2}{r^3} (1 + jkr) \sin \theta \sin \varphi e^{-jkr} \hat{\mathbf{r}} \right. \\ &\quad \left. + \left(\frac{k^2}{r} - \frac{1}{r^3} - \frac{jk}{r^2} \right) \cos \theta \sin \varphi e^{-jkr} \hat{\boldsymbol{\theta}} \right. \\ &\quad \left. + \left(\frac{k^2}{r} - \frac{1}{r^3} - \frac{jk}{r^2} \right) \cos \varphi e^{-jkr} \hat{\boldsymbol{\phi}} \right]. \quad (\text{A.3}) \end{aligned}$$

In the far-field range

$$\begin{aligned} \lim_{r \rightarrow \infty} \mathbf{E} &= \frac{1}{j\omega\epsilon} \frac{I \Delta l k^2}{4\pi r} e^{-jkr} (\hat{\boldsymbol{\theta}} \cos \theta \sin \varphi + \hat{\boldsymbol{\phi}} \cos \varphi) \\ &= \frac{-jk\eta_0 I \Delta l e^{-jkr}}{4\pi r} (\hat{\boldsymbol{\theta}} \cos \theta \sin \varphi + \hat{\boldsymbol{\phi}} \cos \varphi). \quad (\text{A.4}) \end{aligned}$$

In cartesian coordinate

$$\begin{aligned} \hat{\boldsymbol{\theta}} &= \hat{\mathbf{x}} \cos \theta \cos \varphi + \hat{\mathbf{y}} \cos \theta \sin \varphi - \hat{\mathbf{z}} \sin \theta \\ \hat{\boldsymbol{\phi}} &= -\hat{\mathbf{x}} \sin \varphi + \hat{\mathbf{y}} \cos \varphi. \quad (\text{A.5}) \end{aligned}$$

Then, the y-component of the electric field scattered by the Hertzian dipole is

$$E_{\text{dipole}}^y = \frac{-jk\eta I \Delta l e^{-jkr}}{4\pi r} (\cos^2 \theta \sin^2 \varphi + \cos^2 \varphi). \quad (\text{A.6})$$

Replacing $I \Delta l$ by the actual surface current element $J(\zeta, \xi) d\zeta d\xi$, the far-field electric field radiated from this Hertzian dipole can be derived as

$$\begin{aligned} dE_S^y(\zeta, \xi, r, \theta, \varphi) \\ = \frac{-jk\eta_0 J(\zeta, \xi) e^{-jkr}}{4\pi r} (\cos^2 \theta \sin^2 \varphi + \cos^2 \varphi) d\zeta d\xi. \end{aligned} \quad (\text{A.7})$$

B. Derivation of (9) and (10)

For the bistatic radar described in Fig. 6, the transient displacement $\mathbf{p} = (p_x, p_y, p_z)$ and the 3-D transient velocity $\mathbf{v} = (v_x, v_y, v_z)$ satisfy $\mathbf{v} = d\mathbf{p}/dt$, whose differential form is $\Delta \mathbf{p}(t) = \mathbf{v}(t) \Delta t$. Therefore, the roundtrip distances at the time t and $t + \Delta t$ can be calculated as $r(t) = |\mathbf{r}_T(t)| + |\mathbf{r}_R(t)|$ and

$$r(t + \Delta t) = \sqrt{|\mathbf{r}_T(t) + \Delta \mathbf{p}(t)|^2} + \sqrt{|\mathbf{r}_R(t) - \Delta \mathbf{p}(t)|^2} \quad (\text{A.8})$$

respectively. For Δt approaching to 0

$$\begin{aligned} \lim_{\Delta t \rightarrow 0} r(t + \Delta t) \\ = \lim_{\Delta t \rightarrow 0} \sqrt{|\mathbf{r}_T(t)|^2 + 2\mathbf{r}_T(t) \cdot \Delta \mathbf{p}(t) + |\Delta \mathbf{p}(t)|^2} \\ + \sqrt{|\mathbf{r}_R(t)|^2 - 2\mathbf{r}_R(t) \cdot \Delta \mathbf{p}(t) + |\Delta \mathbf{p}(t)|^2} \\ \approx \lim_{\Delta t \rightarrow 0} \sqrt{|\mathbf{r}_T(t)|^2 + 2\mathbf{r}_T(t) \cdot \Delta \mathbf{p}(t)} \\ + \sqrt{|\mathbf{r}_R(t)|^2 - 2\mathbf{r}_R(t) \cdot \Delta \mathbf{p}(t)} \\ \approx \lim_{\Delta t \rightarrow 0} |\mathbf{r}_T(t)| + \Delta \mathbf{p}(t) \cdot \hat{\mathbf{r}}_T(t) + |\mathbf{r}_R(t)| - \Delta \mathbf{p}(t) \cdot \hat{\mathbf{r}}_R(t). \end{aligned} \quad (\text{A.9})$$

Note that in the first approximation, $|\Delta \mathbf{p}(t)|^2$ has been omitted. In the second approximation, the second-order Taylor expansion $\lim_{\alpha \rightarrow 0} (1 + \alpha)^{1/2} \approx 1 + \alpha/2$ is applied. Thus, the derivative of the roundtrip distance can be derived as

$$\begin{aligned} \frac{dr(t)}{dt} &= \lim_{\Delta t \rightarrow 0} \frac{r(t + \Delta t) - r(t)}{\Delta t} \\ &= (\Delta \mathbf{p}(t) \cdot \hat{\mathbf{r}}_T - \Delta \mathbf{p}(t) \cdot \hat{\mathbf{r}}_R) / \Delta t \\ &= \mathbf{v}(t) \cdot \hat{\mathbf{r}}_T - \mathbf{v}(t) \cdot \hat{\mathbf{r}}_R \\ &= |v_T(t)| - |v_R(t)|. \end{aligned} \quad (\text{A.10})$$

REFERENCES

- [1] M. R. Alam, M. B. I. Reaz, and M. A. M. Ali, "A review of smart homes—past, present, and future," *IEEE Trans. Syst., Man, Cybern., C, Appl. Rev.*, vol. 42, no. 6, pp. 1190–1203, Nov. 2012.
- [2] S. Kim *et al.*, "A highly sensitive capacitive touch sensor integrated on a thin-film-encapsulated active-matrix OLED for ultrathin displays," *IEEE Trans. Electron Devices*, vol. 58, no. 10, pp. 3609–3615, Oct. 2011.
- [3] J. Han, L. Shao, D. Xu, and J. Shotton, "Enhanced computer vision with microsoft kinect sensor: A review," *IEEE Trans. Cybern.*, vol. 43, no. 5, pp. 1318–1334, Oct. 2013.
- [4] T. Fan *et al.*, "Wireless hand gesture recognition based on continuous-wave Doppler radar sensors," *IEEE Trans. Microw. Theory Techn.*, vol. 64, no. 11, pp. 4012–4020, Nov. 2016.
- [5] Y. Zhang *et al.*, "Hand gesture recognition based on SIMO Doppler radar sensors," in *Short-Range Micro-Motion Sensing with Radar Technology*. Edison, NJ, USA: IET, 2019.
- [6] Y. Yao and Y. Fu, "Contour model-based hand-gesture recognition using the Kinect sensor," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 24, no. 11, pp. 1935–1944, Nov. 2014.
- [7] V. I. Pavlovic, R. Sharma, and T. S. Huang, "Visual interpretation of hand gestures for human-computer interaction: A review," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 19, no. 7, pp. 677–695, Jul. 1997.
- [8] S. Berman and H. Stern, "Sensors for gesture recognition systems," *IEEE Trans. Syst., Man, Cybern. C, Appl. Rev.*, vol. 42, no. 3, pp. 277–290, Dec. 2011, doi: 10.1109/tsmcc.2011.2161077.
- [9] S. Heunisch, L. O. Fhager, and L.-E. Wernersson, "Millimeter-wave pulse radar scattering measurements on the human hand," *IEEE Antennas Wireless Propag. Lett.*, vol. 18, no. 7, pp. 1377–1380, Jul. 2019.
- [10] L. O. Fhager, S. Heunisch, H. Dahlberg, A. Evertsson, and L.-E. Wernersson, "Pulsed millimeter wave radar for hand gesture sensing and classification," *IEEE Sensors Lett.*, vol. 3, no. 12, pp. 1–4, Dec. 2019.
- [11] S.-J. Ryu, J.-S. Suh, S.-H. Baek, S. Hong, and J.-H. Kim, "Feature-based hand gesture recognition using an FMCW radar and its temporal feature analysis," *IEEE Sensors J.*, vol. 18, no. 18, pp. 7593–7602, Jul. 2018.
- [12] Z. Zhang, Z. Tian, and M. Zhou, "Latern: Dynamic continuous hand gesture recognition using FMCW radar sensor," *IEEE Sensors J.*, vol. 18, no. 8, pp. 3278–3289, Apr. 2018.
- [13] Z. Zhang, Z. Tian, Y. Zhang, M. Zhou, and B. Wang, "U-DeepHand: FMCW radar-based unsupervised hand gesture feature learning using deep convolutional auto-encoder network," *IEEE Sensors J.*, vol. 19, no. 16, pp. 6811–6821, Aug. 2019.
- [14] S. Hazra and A. Santra, "Robust gesture recognition using millimetric-wave radar system," *IEEE Sensors Lett.*, vol. 2, no. 4, pp. 1–4, Dec. 2018.
- [15] J. Lien *et al.*, "Soli: Ubiquitous gesture sensing with millimeter wave radar," *ACM Trans. Graph.*, vol. 35, no. 4, p. 142, 2016.
- [16] T. Sakamoto, X. Gao, E. Yavari, A. Rahman, O. Boric-Lubecke, and V. M. Lubecke, "Hand gesture recognition using a radar echo I-Q plot and a convolutional neural network," *IEEE Sensors Lett.*, vol. 2, no. 3, pp. 1–4, Sep. 2018.
- [17] S. Skaria, A. Al-Hourani, M. Lech, and R. J. Evans, "Hand-gesture recognition using two-antenna Doppler radar with deep convolutional neural networks," *IEEE Sensors J.*, vol. 19, no. 8, pp. 3041–3048, Apr. 2019.
- [18] G. Li, R. Zhang, M. Ritchie, and H. Griffiths, "Sparsity-driven micro-Doppler feature extraction for dynamic hand gesture recognition," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 54, no. 2, pp. 655–665, Apr. 2018.
- [19] J.-W. Choi, S.-J. Ryu, and J.-H. Kim, "Short-range radar based real-time hand gesture recognition using LSTM encoder," *IEEE Access*, vol. 7, pp. 33610–33618, 2019.
- [20] C. Liu, Y. Li, D. Ao, and H. Tian, "Spectrum-based hand gesture recognition using millimeter-wave radar parameter measurements," *IEEE Access*, vol. 7, pp. 79147–79158, 2019.
- [21] L. Tsang, J. A. Kong, and K.-H. Ding, *Scattering of Electromagnetic Waves: Theories and Applications*. Hoboken, NJ, USA: Wiley, 2004.
- [22] H. B. Demuth, M. H. Beale, O. De Jess, and M. T. Hagan, *Neural Network Design*. Stillwater, OK, USA: Martin Hagan, 2014.
- [23] C. M. Bishop, *Pattern Recognition and Machine Learning*. Cham, Switzerland: Springer, 2006.
- [24] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, pp. 436–444, May 2015.
- [25] Y. Kim and B. Toomajian, "Hand gesture recognition using micro-Doppler signatures with convolutional neural network," *IEEE Access*, vol. 4, pp. 7125–7130, 2016.
- [26] S. Dong *et al.*, "Doppler cardiogram: A remote detection of human heart activities," *IEEE Trans. Microw. Theory Techn.*, vol. 68, no. 3, pp. 1132–1141, Nov. 2019.
- [27] Q. Lv *et al.*, "Doppler vital signs detection in the presence of large-scale random body movements," *IEEE Trans. Microw. Theory Techn.*, vol. 66, no. 9, pp. 4261–4270, Sep. 2018.
- [28] Y. Xiao, J. Lin, O. Boric-Lubecke, and M. Lubecke, "Frequency-tuning technique for remote detection of heartbeat and respiration using low-power double-sideband transmission in the Ka-band," *IEEE Trans. Microw. Theory Techn.*, vol. 54, no. 5, pp. 2023–2032, May 2006.
- [29] M. Mercuri *et al.*, "Analysis of an indoor biomedical radar-based system for health monitoring," *IEEE Trans. Microw. Theory Techn.*, vol. 61, no. 5, pp. 2061–2068, May 2013.

- [30] C.-S. Lin, S.-F. Chang, C.-C. Chang, and C.-C. Lin, "Microwave human vocal vibration signal detection based on Doppler radar technology," *IEEE Trans. Microw. Theory Techn.*, vol. 58, no. 8, pp. 2299–2306, Aug. 2010.
- [31] A. Montuori *et al.*, "The interferometric use of radar sensors for the urban monitoring of structural vibrations and surface displacements," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 9, no. 8, pp. 3761–3776, Aug. 2016.
- [32] G. Grazzini, M. Pieraccini, D. Dei, and C. Atzeni, "Simple microwave sensor for remote detection of structural vibration," *Electron. Lett.*, vol. 45, no. 11, pp. 567–569, May 2009.
- [33] C.-M. Nieh, C. Wei, and J. Lin, "Concurrent detection of vibration and distance using unmodulated CW Doppler vibration radar with an adaptive beam-steering antenna," *IEEE Trans. Microw. Theory Techn.*, vol. 63, no. 6, pp. 2069–2078, Jun. 2015.
- [34] A. D. Droitcour, O. Boric-Lubecke, V. M. Lubecke, J. Lin, and G. T. A. Kovacs, "Range correlation and I/Q performance benefits in single-chip silicon Doppler radars for noncontact cardiopulmonary monitoring," *IEEE Trans. Microw. Theory Techn.*, vol. 52, no. 3, pp. 838–848, Mar. 2004.
- [35] J. Wang, X. Wang, L. Chen, J. Huangfu, C. Li, and L. Ran, "Noncontact distance and amplitude-independent vibration measurement based on an extended DACM algorithm," *IEEE Trans. Instrum. Meas.*, vol. 63, no. 1, pp. 145–153, Jan. 2014.
- [36] Z. Gu *et al.*, "Blind separation of Doppler human gesture signals based on continuous-wave radar sensors," *IEEE Trans. Instrum. Meas.*, vol. 68, no. 7, pp. 2659–2661, Jul. 2019.
- [37] Z. Gu *et al.*, "Remote blind motion separation using a single-tone SIMO Doppler radar sensor," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 1, pp. 462–472, Jan. 2019.
- [38] R. Zhang, M. Wang, L. X. Cai, Z. Zheng, X. Shen, and L.-L. Xie, "LTE-unlicensed: The future of spectrum aggregation for cellular networks," *IEEE Wireless Commun.*, vol. 22, no. 3, pp. 150–159, Jun. 2015.
- [39] Q. Lv *et al.*, "High dynamic-range motion imaging based on linearized Doppler radar sensor," *IEEE Trans. Microw. Theory Techn.*, vol. 62, no. 9, pp. 1837–1846, Sep. 2014.
- [40] S. Silver, *Microwave Antenna Theory and Design*, no. 19. Edison, NJ, USA: IET, 1984.
- [41] Y. Kalkan, "Frequency-only radars and other frequency-based systems with applications," *IEEE Microw. Mag.*, vol. 17, no. 7, pp. 26–52, Jul. 2016.
- [42] H. Johansson and A. Eghbali, "Two polynomial FIR filter structures with variable fractional delay and phase shift," *IEEE Trans. Circuits Syst. I, Reg. Papers*, vol. 61, no. 5, pp. 1355–1365, May 2014.
- [43] C. Cortes and V. Vapnik, "Support-vector networks," *Mach. Learn.*, vol. 20, no. 3, pp. 273–297, 1995.
- [44] Y. Sun, T. Fei, X. Li, A. Warnecke, E. Warsitz, and N. Pohl, "Real-time radar-based gesture detection and recognition built in an edge-computing platform," *IEEE Sensors J.*, vol. 20, no. 18, pp. 10706–10716, Sep. 2020.
- [45] A. G. Howard *et al.*, "MobileNets: Efficient convolutional neural networks for mobile vision applications," 2017, *arXiv:1704.04861*. [Online]. Available: <http://arxiv.org/abs/1704.04861>