# Multidimensional Feature Representation and Learning for Robust Hand-Gesture Recognition on Commercial Millimeter-Wave Radar

Zhaoyang Xia, *Graduate Student Member, IEEE*, Yixiang Luomei, *Graduate Student Member, IEEE*, Chenglong Zhou, and Feng Xu, *Senior Member, IEEE*

***Abstract*— This article presents a robust hand-gesture recognition method via multidimensional feature representation and learning specifically designed for commercial frequency-modulated continuous wave (FMCW) multi-input multi-output (MIMO) millimeter-wave radar. First, the optimal configuration of the radar system parameters for the hand-gesture recognition scenario is investigated and a standard procedure to determine the system configuration is given. Then a moving scattering center model is proposed to represent the 3-D point cloud in the range–Doppler (RD)-angular multidimensional feature space. A scattering point detection and tracking algorithm is presented based on a set of motion constraints in terms of position, velocity, and acceleration. It is derived from the space-time continuity of a nonrigid target. Finally, a lightweight multichannel convolutional neural network (CNN) is designed to learn and classify multidimensional gesture features including radial RD and tangential azimuth–elevation. Extensive experiments are carried out with the developed system and a large data set is obtained to train and test the classifier. The results show that the proposed gesture recognition method can effectively distinguish gestures that are easily confused in the RD domain and achieve robust performances under various conditions.**

***Index Terms*— Detection and tracking, gesture recognition, millimeter-wave (mmw) radar, moving scattering center model, multichannel convolutional neural network (CNN), multidimensional feature.**

## I. Introduction

HAND-GESTURE interaction is one of the useful human–computer interaction technologies which has potential application in the fields of assisted driving, smart home, and consumer electronics. It can be used for vehicle equipment control [1], home equipment control [2], augmented reality–virtual reality (AR-VR) [3], game control [4], smartphone interaction [5], mobile robot control [6], sign language translation [7], and many other scenarios.

Manuscript received January 19, 2020; revised April 4, 2020 and June 25, 2020; accepted July 17, 2020. Date of publication July 31, 2020; date of current version May 21, 2021. This work was supported in part by the National Key Research and Development Program of China under Grant 2017YFA0700203, and in part by the Natural Science Foundation of China under Grant 61991422 and Grant 61822107. *(Corresponding author: Feng Xu.)*

The authors are with the Key Laboratory for Information Science of Electromagnetic Waves, Ministry of Education, Fudan University, Shanghai 200433, China, and also with the School of Information Science and Technology, Fudan University, Shanghai 200433, China (e-mail: fengxu@fudan.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/TGRS.2020.3010880

Various gesture recognition techniques ranging from wearable sensor-based, vision-based, ultrasound-based, RFID-based to Wi-Fi-based solutions have been studied. The wearable sensor-based approaches [6], [7] can get accurate motion information of gestures which can be used in challenging applications such as sign language translation. However, these methods can only work when sensors are worn by the user. The vision-based approaches [1]–[4], [8] exploit cameras or depth cameras to extract shapes, textures, and depths of the hand for gesture recognition. However, these methods cannot work in the cases of non-line of sight, poor light conditions, or being obscured. Additional issues of the camera-based approaches include privacy concern and high power consumption. The ultrasound-based approaches [5] use ultrasonic wave to sense the gesture, but are significantly affected by the propagation speed and diffraction of sound. The RFID-based [9] and Wi-Fi-based [10] approaches take advantage of the communication electromagnetic signals and use as radar mode to sense gestures. Due to the limitation of waveforms which are designed for communication purpose, these methods may perform poorly when there are interferences in the environment.

Due to the limitations of these gesture recognition solutions, very few gesture interaction systems have been deployed in real life so far. In contrast, millimeter-wave (mmw) radar-based gesture recognition solution seems to be an attractive sensing modality: high range resolution, small antenna size, not relying on light sources, and sensing through smoke, dust, and even nonmetallic materials. The mmw radar was first developed in the 1940s and then had been used in a variety of military and civilian fields in the late 1970s [11]. With the development of the monolithic microwave integrated circuit (MMIC) technology [12], the high-integrated, low-cost, low-power mmw radar chip sensors based on GaAs [13], SiGe [14], and CMOS [15] integrated circuit processes have become widely available on the market. They often use the low-complexity frequency-modulated continuous wave (FMCW) mode. In the past decades, researchers have explored the potential applications of mmw radars in many fields such as collision avoidance [16], [17], navigation [18], traffic surveillance [19], remote sensing [20], human security [21], through-the-wall surveillance [22], foreign object detection on airport runway [23], vital signs detection [24],

pedestrian detection [25], human motion recognition [26], gesture recognition [27]–[33], and so on.

Google's Soli [27] is among the first mmw radars for gesture interaction application on consumer electronics devices. Various gesture recognition methods for Soli radar have been proposed [27]–[30]. However, the existing methods for mmw radar extract gesture features from range–Doppler (RD) images [27]–[31], micro-Doppler [32] images, and range-time images [33], which only use radial features, but not tangential features. Thus, it is insufficient to represent motion gestures in 3-D space.

The key for hand-gesture recognition is to distinguish different motion gestures while preserving necessary invariances to nonessential factors such as user identity, position, and direction. It corresponds to the challenging problems of user confusion, translation confusion, and rotation confusion, respectively. These problems become severe in large field of view or long-distance (LFoV/LD) scenario such as smart home, for example, users are allowed make hand gesture from different viewing angles with respect to the radar. Intrinsic gesture features may easily get entangled with observation conditions such as user position in the conventional RD-time space [31]. It causes two types of issues.

1) Ambiguity, two different gestures may be confused as one. For example, in the RD domain, the gesture of waving from the left side to the center (waving right) versus waving from the right side to the center (waving left) would appear the same after mapping to the radial distance and radial velocity.
2) Variability, the same gesture under different conditions may be classified as two. For example, the same gesture of "waving left" occurred on the left side versus on the right side of the radar will have opposite radial features.

This article addresses the challenge in the context of multi-input multi-output (MIMO) commercial mmw radar. The keys to disentangle intrinsic gesture features and nonessential factors are to identify the hand position via MIMO beamforming and the angular dimensions should be included in the feature space. This article presents a multidimensional feature representation and learning method for robust gesture recognition with mmw radar. It mainly contains three steps. First, we detect the target points on each frame channel-average RD image (CA-RDI) and detect moving scattering centers on 2-D angular images (2-D-AIs) of all the target points. Then we track one scattering center of the target from a sequence of point clouds and extract the corresponding multidimensional gesture features. Finally, we design a multichannel convolutional neural network (CNN) to learn and classify multidimensional gesture features in a multiuser and multiposition gesture scenario. Extensive experiments are carried out where a large data set is collected for training and testing purposes. The comparative study shows that the proposed robust approach outperforms the conventional RD-domain-based methods by 3% and 16% in terms of average accuracy for four palm gestures and three finger gestures on eight training users and two nontraining users, respectively. However, limited by the poor angular resolution of the MIMO array, the approach only improves the recognition results for gestures with measurable angle changes, but does not work well for micromotion gestures.

The main contributions of this article are summarized as follows.

1) Optimal configuration of an mmw FMCW radar system for hand-gesture recognition scenarios is analyzed and a standardized procedure to determine the optimal configuration is provided. It facilitates the system setup of a commercial mmw radar for gesture recognition applications.
2) A novel approach is proposed to address the ambiguity and variability challenges of robust hand-gesture recognition in particular for the LFoV/LD scenario. MIMO beamforming is first used to detect and track the hand scatterer position and then intrinsic gesture features are disentangled in the multidimension feature space via trained CNN classifier.
3) A real-time gesture recognition demonstration system is developed (Fig. 1). Using the system, extensive experiments with ten volunteers performing seven gestures under various conditions are conducted where a large size data set is collected. Performance of robust hand-gesture recognition is quantitatively evaluated with comparative analyses.

The rest of this article is organized as follows. We first review the related works in Section II. In Section III, we present the criteria to design an optimal radar parameters' configuration for gesture recognition. In Section IV, we establish the signal model of the moving scattering centers to represent the 3-D point cloud, which is used to track a hand and obtain multidimensional spatially continuous and time-varying features of several motion gestures. Section V presents the experimental results and analyses of multidimensional features for hand-gesture recognition. In Section VI, we discuss some remaining issues and future works.

## II. Related Works

Gesture recognition based on radar has attracted a lot of attentions in recent years. The existing radar-based gesture recognition methods mainly include Doppler-radar-based [34], ultra-wideband (UWB)-radar-based [35], and FMCW-radar-based [27]–[33]. The Doppler radar has poor anti-interference ability, narrow bandwidth, and low range resolution. The IR-UWB radar uses the time difference between the target echo signal and the transmitted signal to measure the distance; however, accurate time difference measurement, signal modulation, and signal processing require high performance and thus high-cost processors. In comparison, the FMCW mmw radar indirectly measures the time difference based on the phase difference, which is simpler and has lower cost, high range resolution, high velocity resolution, and high time resolution. Thus, it is an ideal technology to sense hand gesture.

Recently, there are some progresses on motion gesture representation and recognition based on FMCW mmw radar. Lien *et al.* [27] from Google's Soli team extract various time-varying features from RD spectrum to represent micromotion gestures based on a 60-GHz mmw radar and use

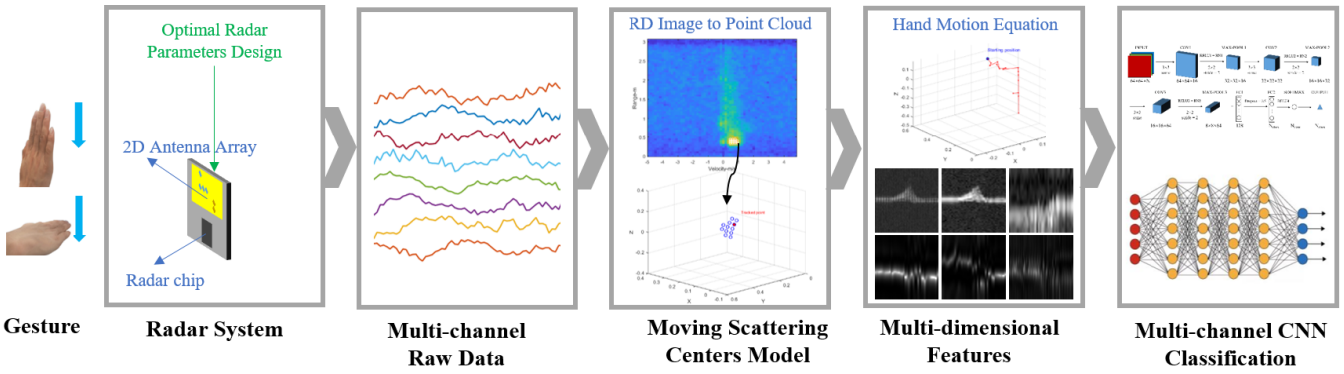


Fig. 1. Real-time hand-gesture recognition architecture.

a random forest method for gesture recognition. Hazra and Santra [29] from Soli's chip manufacturer Infineon use a long recurrent all-convolution neural network to learn automatically gesture features and classify micromotion gestures from a series of RD images. Malysa *et al.* [32] extract Doppler spectrum features for motion gesture representation based on a 77-GHz mmw radar and classify gestures based on a hidden Markov model (HMM) method. Zhang *et al.* [33] obtained the range spectrum based on an Infineon 24-GHz radar to represent radial motion gestures and then learn features and classify gestures by a recurrent 3-D CNN. In [31], we use the feature of stitched RD images to represent micromotion gestures based on a 77-GHz mmw radar and then learn features and classify gestures based on a single-channel CNN.

These existing radar-based gesture recognition methods extract gesture classification features from RD domain and classify gestures by training with a large number of samples. However, there are differences in the same gestures made by different users, in different spatial positions, and along different moving directions, or the same gesture may appear distinct as observed under different situations. This creates the issues of ambiguity and variability which degrade the robustness of human–machine interaction via hand gesture.

The gesture recognition architecture proposed in this article aims to address the need for robust radar-based gesture recognition, that is, correct classification of gestures regardless of user identity, position, and viewing angle.

## III. Radar Principles and Optimal Configuration

### A. FMCW Radar Signal Model

A brief introduction of the principles of the FMCW radar is given here. The normalized transmitted signal of the FMCW radar is [36]

$$s_{\mathrm{T}}(t) = e^{\mathrm{j}2\pi\left(f_c t + \frac{K}{2}t^2\right)} \tag{1}$$

where $t$ denotes the fast time within a chirp (a frequency modulation period), $-(T_{\mathrm{s}}/2) \le t \le (T_{\mathrm{s}}/2)$, $f_{\mathrm{c}}$ and $K_{\mathrm{s}} = \mathrm{B}/T_{\mathrm{s}}$ denote the center frequency and the frequency slope of the chirp, where B and $T_{\mathrm{s}}$ denote the bandwidth and the time duration of the chirp, respectively.

If we assume a moving point target at range $R$, the signal received by the radar is [36]

$$s_{\mathrm{R}}(t) = \sigma \cdot \mathrm{e}^{\mathrm{j}2\pi\left((f_{\mathrm{c}}+f_{\mathrm{D}})(t-\tau)+\frac{K_{\mathrm{s}}}{2}(t-\tau)^2\right)} \tag{2}$$

where $\sigma$ is proportional to the radar cross section, antenna gain, and range attenuation, $\tau = 2R/c$ denotes the time of flight, $c$ is the speed of light, $f_{\mathrm{D}} = (2v_{\mathrm{r}}/\lambda)$ denotes the Doppler frequency shift, and $v_{\mathrm{r}}$ is the radial velocity.

Then the intermediate frequency (IF) signal after mixing the received signal and the transmitted signal and low-pass filtering is [36]

$$\begin{aligned} s_{\mathrm{IF}}(t) &= s_{\mathrm{T}}^{*}(t) s_{\mathrm{R}}(t-\tau) \\ &= \sigma \cdot \mathrm{e}^{-\mathrm{j}2\pi\left[f_{\mathrm{c}}\tau - f_{\mathrm{D}}(t-\tau)+\frac{K_{\mathrm{s}}\tau}{2}(2t-\tau)\right]}. \end{aligned} \tag{3}$$

Since $f_{\mathrm{IF}} = K_{\mathrm{s}}\tau$, where $f_{\mathrm{IF}}$ is the frequency of the IF signal or beat frequency, (3) can be rewritten as [36]

$$s_{\mathrm{IF}}(t) = \sigma \cdot e^{-\mathrm{j}2\pi\left[f_{\mathrm{c}}\tau + (f_{\mathrm{IF}}(t-\tau/2) f_{\mathrm{D}}(t-\tau))\right]} \tag{4}$$

Since $(f_{\mathrm{IF}}/\tau) = (\mathrm{B}/T_{\mathrm{s}}) = K_{\mathrm{s}}$ and $\tau = 2R/c$, the range of the target can be determined by measuring the frequency of the sampled IF signal [36]

$$R = \frac{\mathrm{c} T_{\mathrm{s}} f_{\mathrm{IF}}}{2B}. \tag{5}$$

If there are multiple targets, the ranges of them can be resolved in the frequency domain by performing range-fast Fourier transform (FFT) on the IF signals from all the targets. The Doppler shifts of them can be resolved by a chain of chirps, which consist a frame. Then we can obtain the RD images by performing range-FFT and Doppler-FFT on the sampled data of the IF signals for frame time or slow time [27].

### B. Distance and Velocity Measurement

The minimum interval between two adjacent targets that can be discriminated by the radar in the radial direction is defined as the range resolution [36]

$$r_{\mathrm{res}} = \frac{\mathrm{c}}{2B}. \tag{6}$$

To perform ADC sampling on the IF signal, the sampling frequency $F_{\mathrm{s}}$ should not be smaller than $f_{\mathrm{IF}}$, that is, $F_{\mathrm{s}} \ge f_{\mathrm{IF}}$.

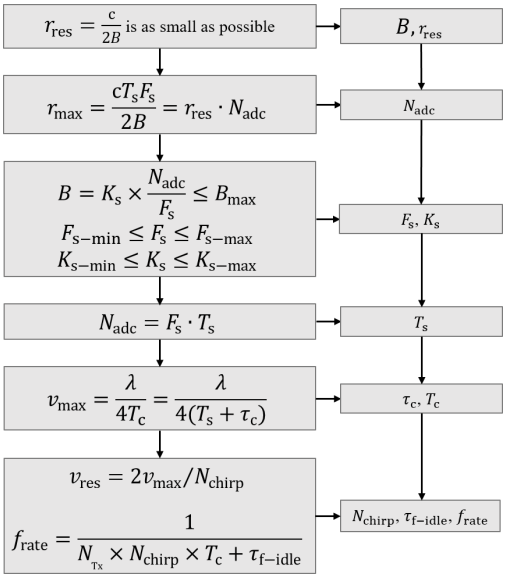


Fig. 2. Procedure to determine the optimal configuration of radar parameters for gesture recognition.

According to (5), the maximum unambiguous range measured by an FMCW radar is [36]

$$r_{\max} = \frac{cT_sF_s}{2B} = r_{res} \cdot N_{adc} \tag{7}$$

where $N_{adc} = F_s \cdot T_s$ denotes the number of samples per chirp.

If we assume that a target's radial velocity does not change during a chirp period, the phase difference in the echo caused by the target motion should be $\omega = (4\pi \Delta r)/\lambda = (4\pi v \cdot T_s)/\lambda$. To measure an unambiguous velocity, $|\omega| \le \pi$ should be guaranteed, so that the maximum unambiguous velocity is [36]

$$v_{\max} = \frac{\lambda}{4T_s} \tag{8}$$

where $\lambda$ is the wavelength corresponding to the center frequency $f_c$.

In addition, the resolution of Doppler-FFT on a chain of chirps determines the ability to distinguish the phase difference in echo $\omega_1$ and $\omega_2$, that is, $|\omega_1 - \omega_2| = (4\pi \Delta v \cdot T_s)/\lambda \ge (2\pi/N_{chirp})$, where $N_{chirp}$ denotes the number of chirps per frame, so the velocity resolution is [36]

$$v_{res} = \frac{\lambda}{2N_{chirp} \cdot T_s}. \tag{9}$$

### C. Optimal Configuration of Radar Parameters

Radar parameter configuration will certainly affect the features captured by the radar and thus affect the performance of gesture recognition. We present some important criteria to design the optimal parameters of the FMCW mmw radar for gesture recognition in this article. The procedure to determine the optimal configuration of the radar parameters is given as shown in Fig. 2.

The parameters' limitations by radar hardware must be observed in the radar parameter design. The limits of the mmw radar parameters related to gesture recognition include the maximum measurable range determined by the transmit power $r_{m-max}$, the number of transmitting antennas and receiving antennas $N_{Tx}$ and $N_{Rx}$, the number of virtual channels in the horizontal and vertical directions $N_{chan-H}$ and $N_{chan-V}$ determined by the antenna array design, the maximum sweep bandwidth $B_{max}$, the maximum and minimum sweep slope $K_{s-max}$ and $K_{s-min}$, and the maximum and minimum ADC sampling rate $F_{s-max}$ and $F_{s-min}$ determined by the radar front-end.

In application scenarios that require higher range resolution, $r_{res}$ should be as small as possible, according to (6), and $B$ should as close as possible but not greater than $B_{max}$. $B$ is determined by

$$B = K_s \times T_s = K_s \times \frac{N_{adc}}{F_s}. \tag{10}$$

Given $r_{res}$, according to (7), $N_{adc}$ can be determined by the measurable radial distance range required for the application scenario. $T_s$ is limited by (8), and then $K_s$ and $F_s$ are determined by (10), but also meet these conditions $B \le B_{max}$, $F_{s-min} \le F_s \le F_{s-max}$ and $K_{s-min} \le K_s \le K_{s-max}$.

According to (8) and (9), the maximum measurable velocity and velocity resolution of the radar system are mainly determined by the chirp cycle $T_s$. In theory, the FMCW waveform is composed of continuous chirps and the chirp cycle is equal to the sampling time. In terms of a practical radar system, after one chirp is completed, the frequency needs to reset to the start frequency of the next chirp, and the sampling section usually selects the best linear section of the chirps, so the actual chirp cycle which determines $v_{max}$ and $v_{res}$ can be expressed as

$$T_c = \frac{N_{adc}}{F_s} + \tau_c = T_s + \tau_c \tag{11}$$

where $\tau_c$ indicates the interval time other than the sampling time in the chirp cycle.

$T_s$ is usually small for close-range applications such as gesture recognition, so $v_{max}$ can be adjusted by change $\tau_c$ to meet the measurement needs of the maximum velocity of the hand. After $v_{max}$ is determined, $v_{res} = 2v_{max}/N_{chirp}$ only depends on $N_{chirp}$. The smaller the $v_{res}$, the subtler the gestures can be sensed, but the relationship between $N_{chirp}$ and the frame rate should be observed. The frame rate can be calculated by

$$f_{rate} = \frac{1}{T_f} = \frac{1}{N_{Tx} \times N_{chirp} \times T_c + \tau_{f-idle}} \tag{12}$$

where $T_f$ is the frame cycle, and $\tau_{f-idle}$ is the idle time after a set of chirps for a frame. As $N_{Tx}$ and $T_c$ are given, $N_{chirp}$ and $\tau_{f-idle}$ can be used to adjust the frame rate which determines the time resolution for gestures. The higher the frame rate, the finer the gestures can be sensed in the time domain, but the real-time capabilities of the data processing devices must also be considered.

A real case of using this procedure to determine the system parameters is given in Section V.

## IV. Proposed Approach for Robust Recognition

### A. Moving Scattering Center Model

According to the geometric diffraction theory [37], when the wavelength of the incident wave is much smaller than the target size, the backscattered field from the target can be expressed as the superposition of a set of independent scattering centers. The normalized backscattered field for a given polarization can be expressed as [38]

$$E(k') = \sum_{i=1}^{N_S} A_i \left(\mathrm{j}\frac{k}{k_c}\right)^{\alpha_i} e^{-\mathrm{j}2kr_i} \tag{13}$$

where $k$ denotes the instantaneous wavenumber; $N_\mathrm{S}$ is the number of the scattering centers; $A_i$ is the amplitude coefficient of the $i$th scattering center; the normalization is conducted by a reference wavenumber $k_\mathrm{c}$; and $r_i$ is the distance from the $i$th scattering center to the antenna phase center. Note that $\alpha_i$ is the dispersive factor determined by the scattering mechanism ($\alpha_i$ has to be multiples of $(1/2)$). This factor can be ignored as the bandwidth is much smaller than the carrier frequency, that is, $k \approx k_\mathrm{c}$ and $k/k_\mathrm{c} \approx 1$.

Following from the FMCW radar principles, that is, (1)–(5), the raw signal of (13) can be focused on the range profile via fast-time FFT which can be simply written in the following form:

$$E(r) = \sum_{i=1}^{N_\mathrm{S}} A_i e^{-\mathrm{j}2kr_i} \cdot \Pi\left(\frac{r_i - r}{r_\mathrm{res}}\right) \tag{14}$$

where $\Pi(\cdot)$ denotes the impulse response function of ranging. For simplicity, it can be assumed unit rectangular window. Applying FFT to one frame of multiple chirps, it yields the 2-D RD domain, that is,

$$E(r, v, T) = \sum_{i=1}^{N_\mathrm{S}} A_i e^{-\mathrm{j}2kr_i(T)} \cdot \Pi\left(\frac{r_i(T) - r}{r_\mathrm{res}}\right) \cdot \Pi\left(\frac{v_i(T) - v}{v_\mathrm{res}}\right). \tag{15}$$

Note that we introduce the third dimension as the slow time $T$. A moving target like a hand can be represented as multiple moving scattering centers/points with different time-varying radial distances $r_i(T)$ and radial velocities $v_i(T)$. Apparently, the position and the velocity in cross-range dimensions become ambiguous. Usually, a 2-D MIMO antenna array is used to further resolve in two cross-range dimensions. Assuming the scattered waves are plane waves, which means the target is in the array's far-field [39], the signal model of the moving scattering centers for the target can be extended to the MIMO case, which can be expressed as

$$E(r, v, m, n, T) = \sum_{i=1}^{N_\mathrm{S}} A_i e^{-\mathrm{j}2kr_i(T)} \cdot e^{-\mathrm{j}2kd_{im}(T)} \cdot e^{-\mathrm{j}2kd_{in}(T)} \cdot \Pi\left(\frac{r_i(T) - r}{r_\mathrm{res}}\right) \cdot \Pi\left(\frac{v_i(T) - v}{v_\mathrm{res}}\right) \tag{16}$$

where the subscripts $m, n$ denote the $m$th Tx and the $n$th Rx, respectively. The array distance factors $d_{im}(T), d_{in}(T)$ are the projections of the $m$th Tx and the $n$th Rx antenna position offset with respect to the array center perpendicular to the radial direction toward the $i$th scattering center.

If a 2-D MIMO array is used, an imaging method such as microwave holography [40] or synthetic aperture radar (SAR) [41] can be used to obtain the target images, and a nonimaging method such as angle-FFT or beamforming can be used to obtain the target point clouds [42]. Based on the 2-D array, multiangle measurements of the backscattered field can be used to obtain the 2-D angular spectrum to resolve the moving scattering centers with the same radial distances and radial velocities. Thus, we can further convert (16) to the angular domain as

$$E(r, v, \theta, \varphi, T) = \sum_{i=1}^{N_\mathrm{S}} A_i e^{-\mathrm{j}2kr_i(T)} \cdot \Omega[\theta_i(T), \theta] \cdot \Omega[\varphi_i(T), \varphi] \cdot \Pi\left(\frac{r_i(T) - r}{r_\mathrm{res}}\right) \cdot \Pi\left(\frac{v_i(T) - v}{v_\mathrm{res}}\right) \tag{17}$$

where $\theta$ and $\varphi$ denote the elevation and the azimuth angular domain, respectively; $\theta_i(T)$ and $\varphi_i(T)$ denote the time-varying angle of arrival of the $i$th moving scatterer signal; $\Omega(\cdot)$ denotes the ambiguity function in the angular domain after beamforming. Its exact form depends on the beamforming algorithm used and is beyond the scope of this article. Due to the limited array aperture size, the angular resolution is often much lower than range and Doppler resolution. From (17), after RD-FFT and MIMO beamforming, the moving scattering centers of a target in 3-D space are mapped into the time-varying 4-D feature space.

How to detect the target points in the 4-D feature space is the key. Due to the small antenna array aperture of mmw radar, it can be assumed that the positions of the target points in the RD images of all the channels are the same. Thus, a CA-RDI with higher signal-to-noise ratio (SNR) than single channel [31] can be used to detect the moving target points to ensure the consistency of the target indices in the multichannel RD images. The range index and the Doppler index of each moving target point detected in the CA-RDI correspond to $N_\mathrm{ch}$ channels' complex amplitude distributions, which can be used to obtain the 2-D-AI by performing the 2-D angle-FFT or beamforming. Then the moving scattering centers can be detected in the 2-D-AI for all the detected target points in the CA-RDI.

The moving scattering center model presents a theoretical representation of the time-varying 4-D point cloud in the form of multichannel RD images. Furthermore, it shows that the target detection in the RD domain is insufficient, and it is necessary to combine the RD domain and the 2-D angular domain to obtain complete target feature representation.

### B. Scatterer Detection and Tracking

Based on the moving scattering center model in (17), a moving nonrigid target in 3-D space can be represented by the point cloud formed by the moving scattering centers with a set of time-varying parameters, namely, the complex amplitude $A_i$, the radial distance $r_i$, the radial velocity $v_i$, the azimuth angle $\theta_i$, and the elevation angle $\varphi_i$.

TABLE I
MOTION CONSTRAINTS OF A HAND

| Constraints | Formulas | Values |
|---|---|---|
| maximum acceleration | $a_{max-h}$ | 20 m/s$^2$ |
| maximum velocity jump | $\lvert\Delta v\rvert = a_{max-h} \cdot T_f$ | 0.5 m/s |
| maximum velocity | $v_{max-h}$ | 5 m/s |
| maximum position jump | $\lvert\Delta p\rvert = v_{max-h} \cdot T_f$ | 0.1 m |
| maximum angle jump | $\lvert\Delta\theta\rvert = \tan^{-1}(\frac{\lvert\Delta p\rvert}{r_t})$ | variable |

where $r_t$ is radial distance from one moving scattering center to the zero-phase reference point of the radar.

Apparently, for any particular scattering center, a set of motion constraints can be established among its time-varying parameters according to the law of motion. As shown in Table I, we associate the velocity with position jump and angular jump. A set of empirical boundaries of these parameters are also introduced, which are used for tracking hand scatterers.

A repeatable gesture should have a specific motion trajectory, which should follow these motion constraints. Since the number of moving scattering centers extracted from the CA-RDI and 2-D-AI are not constant with slow time, it is impractical to obtain the trajectories of all the scattering centers. In this article, we select one key moving scattering center and assume its velocity vector is constant over one frame period, and then its motion equation can be expressed as

$$\vec{r}(T) = \vec{r}_0 + \sum_{t=1}^{T} \vec{v}_t \cdot T_{\mathrm{f}} \tag{18}$$

where $\vec{r}(T)$ and $\vec{r}_0$ denote the position of the moving scattering center at slow time $T$ and the starting position of that, respectively; $\vec{v}_t$ denotes the absolute velocity vector of the moving scattering center during the $t$th frame period of the motion gestures, and the relationship between the absolute velocity and the radial velocity can be expressed as

$$v_t = \frac{v_{\mathrm{r}}}{\cos(\chi)} = \frac{v_{\mathrm{r}}}{\frac{\overrightarrow{v_{0t}} \cdot \overrightarrow{v_r}}{|\overrightarrow{v_{0t}}| \cdot v_{\mathrm{r}}}} = \frac{|\vec{r}_t - \vec{r}_{t-1}| \cdot v_{\mathrm{r}}^2}{(\vec{r}_t - \vec{r}_{t-1}) \cdot \overrightarrow{v_r}} \tag{19}$$

where $v_t$ and $v_{\mathrm{r}}$ are the absolute velocity and the radial velocity during the $t$th frame, respectively; $\chi$ denotes the intersection angle between the radial velocity direction vector and the absolute velocity direction vector. The unit direction vector of the absolute velocity is approximately defined as $\overrightarrow{v_{0t}} = (\vec{r}_t - \vec{r}_{t-1})/(|\vec{r}_t - \vec{r}_{t-1}|)$.

According to (18) and (19), the motion equation is described by the displacement vector and the radial velocity vector, so the key is to determine the target position and the radial velocity of each frame, which can be achieved in the following steps.

1) First, in the first motion frame CA-RDI, from all the moving target points that satisfy the target detection conditions [e.g., constant false-alarm rate (CFAR)], select the one with the nearest radial distance and the largest radial velocity to obtain the corresponding 2-D-AI and select the moving scattering center with largest amplitude in the 2-D-AI as the initial tracking point.
2) Estimate its motion parameters, that is, radial distance $r_1$, radial velocity $v_1$, azimuth angle $\theta_1$, and elevation angle $\varphi_1$.
3) Detect the moving target points in the $T$th motion frame CA-RDI and detect the moving scattering centers in the 2-D-AIs of all the moving target points. Then select $N_p$ moving scattering centers that satisfy the motion constraints in Table I and calculate the corresponding radial distance $r_i(T)$, radial velocity $v_i(T)$, azimuth angle $\theta_i(T)$, and elevation angle $\varphi_i(T)$. The optimal moving scattering center of the $T$th frame can be determined by the minimum displacement respect to the previous tracked scattering center, that is,

$$\begin{aligned} i^* &= \arg\min_i \big( \big\| r_i(T) \cdot \vec{a}[\theta_i(T), \varphi_i(T)] - r^*(T-1) \\ &\qquad \cdot \vec{a}\big[\theta^*(T-1), \varphi^*(T-1)\big] \big\| \big) \\ \vec{a}(\theta, \varphi) &= [\cos\varphi\cos\theta, \cos\varphi\sin\theta, \sin\varphi]^{\mathrm{T}} \end{aligned} \tag{20}$$

where $i^*$ denotes the index of the optimal scattering center for the $T$th motion frame, $i = 1, 2, 3, 4, \ldots, N_p$; $\|\cdot\|$ denotes the vector norm, and $r^*(T-1)$, $\theta^*(T-1)$, and $\varphi^*(T-1)$ denote the radial distance, the azimuth angle, and the elevation angle of the optimal scattering center for the $(T-1)$th motion frame, respectively; $\vec{a}$ is the steering vector.

Compared with the commonly used target detection method, we not only perform the scatterer detection in the RD domain but also perform the scatterer detection in the 2-D angular domain to obtain more target scatterers. Compared with the conventional tracking methods such as Kalman filtering, the proposed tracking method is more robust by incorporating *a priori* parameter constraints and more efficient for real-time application as it does not require complex velocity estimations.

The motion equation describes the change over time in position, velocity, and acceleration of a hand scatterer. The motion equation can ensure spatial continuity for time-varying features. Scatterer tracking based on motion equation can enable one to track components of the nonrigid hand, such as a fingertip, and this can help us better recognize not only palm gesture but also micromotion finger gestures.

### C. Multidimensional Feature Representation

The procedure to extract spatially continuous multidimensional gesture features is summarized in Fig. 3. $S_1$ and $S_T$ denote the initial scattering center and the $T$th motion frame scattering center for $T = 2, 3, 4, \ldots, N_T$, respectively, and $N_T$ denotes the number of continuous motion frames of one gesture.

First, the CA-RDIs and 2-D-AIs are calculated frame by frame from the beginning of the gesture to detect the moving scattering centers. The CFAR method [43] is the most commonly used target detection method, but in this

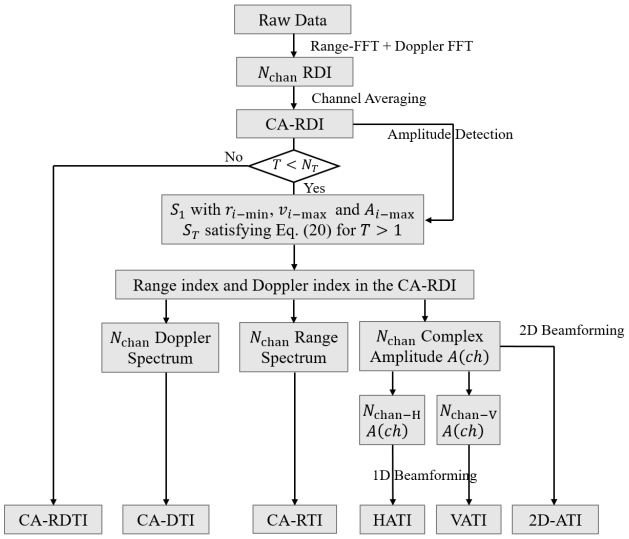


Fig. 3. Procedure to extract multidimensional gesture features.

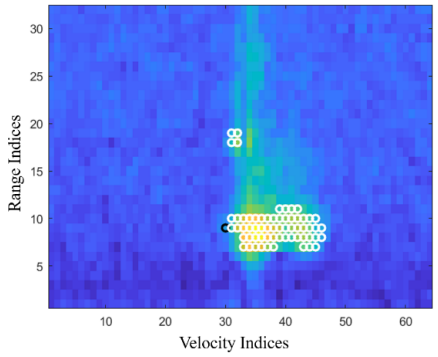


Fig. 4. Detection of the moving target points on a CA-RDI.

article we use the neighbor threshold detection method instead of the CFAR method for following reasons: 1) CFAR is designed to suppress the clutter [44]; however, the hand target appears like a clustered area occupying multiple cells in the CA-RDIs and 2-D-AIs and 2) the window-based CFAR algorithm is computationally expensive and involves complicated parameter estimations. Thus, it is not suitable for real-time implementation on a large number of target points.

For the neighbor threshold detection on the CA-RDIs, first the variable amplitude thresholds are estimated according to the SNRs. Then the points whose amplitudes exceed the amplitude thresholds are regarded as the target points, and a target point is regarded as a valid one while it has two or more neighbor target points. As shown in Fig. 4, the target points are marked by black circles, and the valid target points are marked by white circles. If there are multiple connected areas of target points, the largest one is chosen as the target area, while others come from multipath reflections or interference motions.

For the neighbor threshold detection on the 2-D-AIs, first the variable amplitude thresholds are estimated according to the maximum amplitudes. To ensure that the amplitude difference between the azimuth and the elevation angle spectrum does

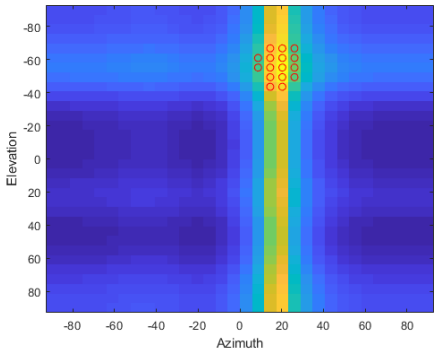


Fig. 5. Detection of the moving scattering centers on a 2-D-AI.

not affect the detection, the motion scattering centers are first detected in the azimuth dimension and then in the elevation dimension. The detection result is shown in Fig. 5, and the maximum numbers in each dimension are calculated by

$$N_{\mathrm{H}} = N_{\mathrm{V}} = \frac{2\mathrm{atan}\left(\frac{L_{\mathrm{hand}}}{2r_i}\right)}{\Delta_{\mathrm{hv}}} \tag{21}$$

where $L_{\mathrm{hand}}$ is the hand length and $\Delta_{\mathrm{hv}}$ is the angular step of beamforming. In this article, the multiple signal classification (MUSIC) algorithm is used to obtain 1-D and 2-D angle spectrum.

Generally, target detection can be implemented in the RD domain or range–angle domain. In this article, the RD domain is used rather than the range–angle domain because the Doppler resolution is high, which helps decompose a target into multiple scattering centers with different velocities. However, the angular resolution is very low for the used antenna array, which is not enough to resolve too many target scattering centers in the angular dimension. A point detected in the angular dimension cannot be uniquely mapped to one point detected in the Doppler domain.

After the moving scattering centers are detected on the 2-D-AIs for moving target points detected on each frame CA-RDI, a sequence of 3-D scattering point clouds for the moving target can be obtained, as shown in Fig. 6; the blue circles represent detected moving scattering centers, and the red dots represent the moving scattering center used for tracking the gesture. From Fig. 6, we can see that the spatial distribution of the palm changes from tilted up to down, which is consistent with the motion process of the palm down gesture.

Based on the motion equation, the tracked point in the 3-D point cloud can be uniquely identified to obtain the motion trajectory of the gesture, as shown in Fig. 7. The range index and the Doppler index of the tracked point in the CA-RDI can be determined. Then the Doppler spectrum corresponding to the range index and the range spectrum corresponding to the Doppler index can be obtained, which represent the continuous velocity distribution of multiple moving scattering centers at the same distance and the continuous distance distribution of multiple moving scattering centers with the same velocity, respectively. In addition, the range index and the Doppler index correspond to the complex amplitude distributions of $N_{\mathrm{chan-H}}$ and $N_{\mathrm{chan-V}}$ channels in the horizontal direction and

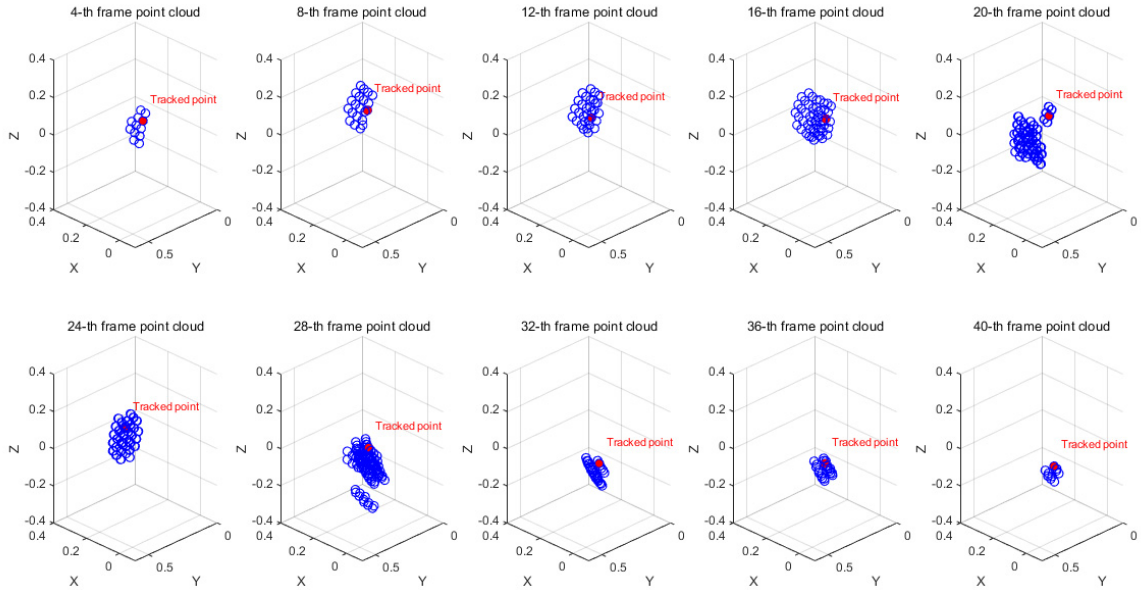


Fig. 6. Sequence of the 3-D point clouds for a palm down hand gesture; the $x$- and $z$-axes are parallel to the horizontal and vertical directions of the used 2-D antenna array, respectively, and the $y$-axis is perpendicular to the array.

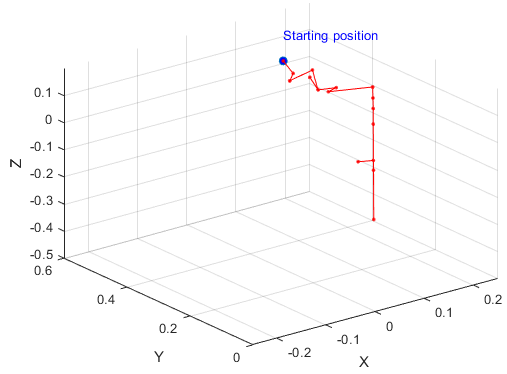


Fig. 7. Motion trajectory of a palm down hand gesture.

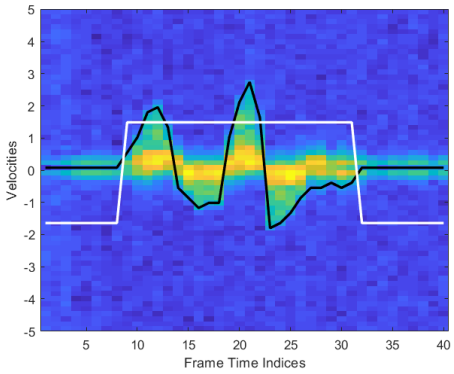


Fig. 8. Real-time segmentation and detection of a finger double-click gesture.

the vertical direction, respectively. The method of angle-FFT or beamforming can be used to obtain the horizontal and vertical angle spectrums, and the method of 2-D angle-FFT or 2-D beamforming can be performed on complex amplitude distributions to obtain the 2-D angle spectrum.

To characterize the time-varying processes of motion gestures, a sliding window with a frame length of $N_{\text{frame\_win}}$ is set and the various time-varying spectrum features are filled into the sliding window to form the channel-average range–Doppler-time-image (CA-RDTI), channel-average Doppler-time-image (CA-DTI), channel-average range-time-image (CA-RTI), horizontal-angle-time-image (HATI), vertical-angle-time-image (VATI), and 2-D-angle-time-image (2-D-ATI) and other six features, the spatial continuity of which is guaranteed by the motion equation. We choose channel-average features over single-channel features because channel averaging can significantly improve the SNR of features and thus improve the classification performance [31]. In the feature images of each channel, the noise is random and the gesture features are located at the same position. Therefore, the background noise can be suppressed and the gesture features can be enhanced by channel averaging. As verified by the two-transmitting and four-receiving TI radar used in our experiments, the channel averaging can improve the SNR by 2.5–4 times, thereby improving the average classification accuracy by more than 1%.

### D. Trigger and Gesture Segmentation

In practice, a real-time gesture recognition system needs to be robustly triggered so that it can correctly identify and segment the gesture feature data. Here, we use the micro-Doppler features as trigger which can represent millimeter-level miniature motion. The CA-DTI in the sliding window is detected to determine whether a data frame is a motion frame and obtain the length of motion frames. As shown in Fig. 8, the black line corresponds to the time-varying velocities of the tracked scattering center in the sequence of 3-D point clouds, and the white line corresponds to the gesture segmentation result. If the width of the white line is greater than a preset threshold,

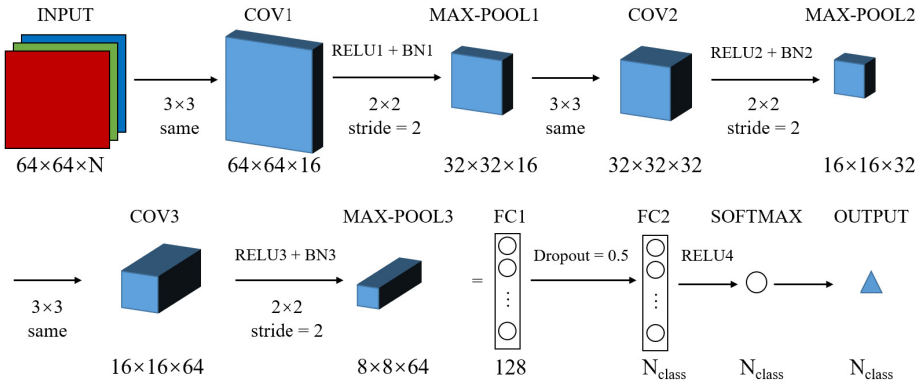


Fig. 9. Multichannel CNN architecture for real-time hand-gesture recognition, where $N$ is the number of channels of the input feature images, $N_{\text{class}}$ is the number of classes of the gestures, and COV, RELU, BN, MAX-POOL, FC, and SOFTMAX denote the convolutional layer, activation function layer, batch normalization layer, max pooling layer, fully connected layer, and softmax layer, respectively.

it is determined that a gesture is observed and the system is triggered. Then we locate the key frame with the largest average amplitude. The frame window near the key frame is segmented to obtain multiple gesture features, which are input into the CNN for subsequent learning and classification.

### *E. CNN-Based Classifier*

To be able to extract deeper features from gesture feature images, we design a lightweight CNN architecture as shown in Fig. 9, the key layers of which include only three 2-D convolutional layers and two 1-D fully connected layers. The input image size of the CNN is $64 \times 64 \times N$, where $N$ is set according to the number of gesture feature types being input to the CNN at the same time.

Because of the small pixel sizes and simple contour features, the CNN is designed with a small convolution kernel 3 in the convolutional layers, a small stride 2 in the max pooling layers, and a small number of neurons in the fully connected layers. The output dimension of the softmax layer is $N_{\text{class}} \times 1$, which corresponds to the probability density distribution classified into various gestures. The model size of the designed CNN classifier is only about 2 MB, and it takes only about 15 ms to classify a single sample on the experimental computer.

To compensate for the horizontal position difference in feature regions in the gesture feature images, data enhancement is set for translating from 5 pixels left to 5 pixels right. We use a momentum gradient descent optimizer, a loss function of cross entropy, a constant learning rate of 0.001, a batch size of 64, and iterate epochs of 50 for training.

Note that we choose CNN over other learning methods mainly because 1) the multidimensional features used for classification are 2-D images which fit best to the CNN and 2) deep neural networks are known good at learning universal features from a large number of data samples, which helps enhance the generalization for multiple users and multiple positions.

## V. Experiments and Analyses

Our real-time hand-gesture recognition architecture consists of five parts: real-time radar system, raw data processing, moving scattering center model, multidimensional feature extraction, and multichannel CNN classification. As shown in Fig. 1, first, the all-channel raw data from the real-time radar system are processed to obtain CA-RDI based on the range-FFT and Doppler-FFT. Then multiple moving scattering centers are detected on 2-D-AIs of all the target points detected on the CA-RDI and estimated target parameters such as amplitudes, positions, velocities, and accelerations frame by frame. By selecting and tracking the optimal moving scattering center which is used to represent the gestures, multidimensional gesture features are extracted and input into the multichannel CNN to learn and classify.

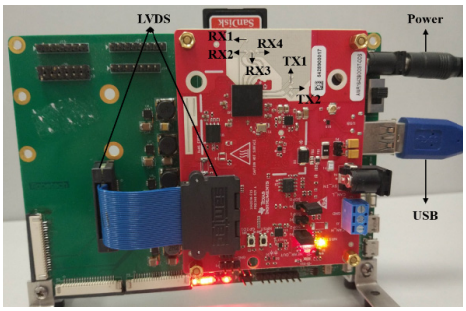


Fig. 10. TI-enabled FMCW radar experimental system.

### *A. Experimental Setup*

To evaluate our hand-gesture recognition architecture, we develop a real-time mmw radar system. As shown in Fig. 10, the system consists of two functional modules: an mmw radar evaluation board (Texas Instruments AWR1642BOOST-ODS) and a real-time high-speed data-capture adapter. The data-capture adapter captures raw ADC data from the radar chip by the low-voltage differential signaling (LVDS) interface and outputs raw data to a computer for further processing by a USB3.0 interface. The computer controls the radar system and performs the remaining feature extraction and CNN classification. Based on the time division multiplexed multi-input multi-output (TDM-MIMO) scheme, two transmitting antennas and four receiving antennas of

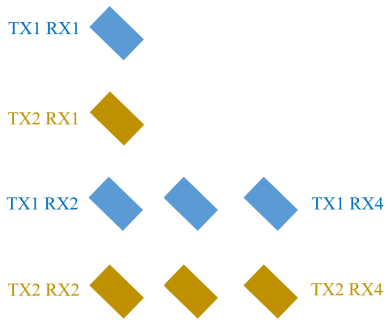


Fig. 11. MIMO virtual array configuration of the system shown in Fig. 10.

TABLE II
PARAMETERS LIMITED BY THE RADAR HARDWARE

| Parameters | Values |
|---|---|
| $N_{\mathrm{Tx}}$ | 2 |
| $N_{\mathrm{Rx}}$ | 4 |
| maximum sweep frequency range | 77-81 GHz |
| $B_{\max}$ | 4 GHz |
| $r_{\mathrm{m-max}}$ | 20 m (for RCS~10 $\mathrm{m}^2$) |
| $K_{\mathrm{s-max}}$ | 150 MHz/μs |
| $F_{\mathrm{s-max}}$ | 15 MHz |
| $F_{\mathrm{s-min}}$ | 2 MHz |

the radar system form a 2-D virtual antenna array, which includes eight data channels, as shown in Fig. 11. There are up to three and four virtual channels in the horizontal and vertical directions, corresponding to angular resolutions of 38° and 29°, respectively. The parameters limited by the radar system hardware are listed in Table II and the configuration parameters for the radar system are listed in Table III, which is calculated based on the procedure to determine the optimal configuration in Fig. 2. The sliding window length of gesture features is set to 40 frames or 1 s.

### B. Data Set

We recruit ten volunteers (six males and four females, age: 22–28 years) to perform seven gestures as shown in Fig. 12 and one nongesture as negative samples at different spatial positions within the radar's field of view and within 1 m distance relative to the radar, 50 groups for each class. The seven gestures contain palm up, palm down, palm left, palm right, finger double-click, finger circle, and forefinger-thumb open-close, where the first four classes are 3-D motion gestures of palm with the overall hand movements, and the latter three classes are micromotion gestures with only finger movements. We obtain six gesture feature images of CA-RDTI, CA-DTI, CA-RTI, HATI, VATI and 2-D-ATI as shown in Fig. 12, and all the image sizes have been scaled to 64 × 64. Finally, we obtain six gesture feature data sets consisting of 10 people × 8 classes × 50 images.

TABLE III
RADAR CONFIGURATION PARAMETERS FOR HAND-GESTURE RECOGNITION

| Parameters | Values |
|---|---|
| $f_c$ | 79.215 GHz |
| $B$ | 3.0976 GHz |
| $r_{\mathrm{res}}$ | 0.0484 m |
| $N_{\mathrm{adc}}$ | 64 |
| $r_{\max}$ | 3.0973 m |
| $F_s$ | 2 MHz |
| $K_s$ | 96.8009 MHz/μs |
| $T_s$ | 32 μs |
| $\lambda$ | 3.8 mm |
| $\tau_c$ | 160 μs |
| $T_c$ | 192 μs |
| $N_{\mathrm{chirp}}$ | 64 |
| $v_{\max}$ | 4.9282 m/s |
| $v_{\mathrm{res}}$ | 0.1540 m/s |
| $\tau_{\mathrm{f-idle}}$ | 424 μs |
| $f_{\mathrm{rate}}$ | 40 Hz |

TABLE IV
COMPARISON OF CLASSIFICATION RESULTS BASED ON SIX GESTURE FEATURES

| Feature type | Classification method | Average accuracy ( % ) | | |
|---|---|---|---|---|
| | | 8 Training User | Test User A | Test User B |
| CA-RDTI | Single-channel CNN | 95.5 | 60.8 | 69.5 |
| CA-DTI | Single-channel CNN | 95.3 | 63.0 | 73.0 |
| CA-RTI | Single-channel CNN | 85.2 | 55.8 | 49.0 |
| HATI | Single-channel CNN | 73.0 | 54.5 | 50.0 |
| VATI | Single-channel CNN | 77.5 | 51.7 | 53.0 |
| 2D-ATI | Single-channel CNN | 83.8 | 63.2 | 65.8 |

### C. Performance Evaluation

To evaluate the gesture representation capabilities of different features, CA-RDTI, CA-DTI, CA-RTI, HATI, VATI, and 2-D-ATI are used as single gesture feature, and the corresponding gesture data sets are input into the CNN for training and test, respectively.

The ten experimental subjects are divided into eight training users and two test users, and the data sets corresponding to the test users do not participate in training. The gesture feature data sets of eight training users are randomly shuffled and divided into training sets, verification sets, and test sets at a ratio of 6: 2: 2. The number of training samples, verification samples, and test samples in each category are 240, 80,

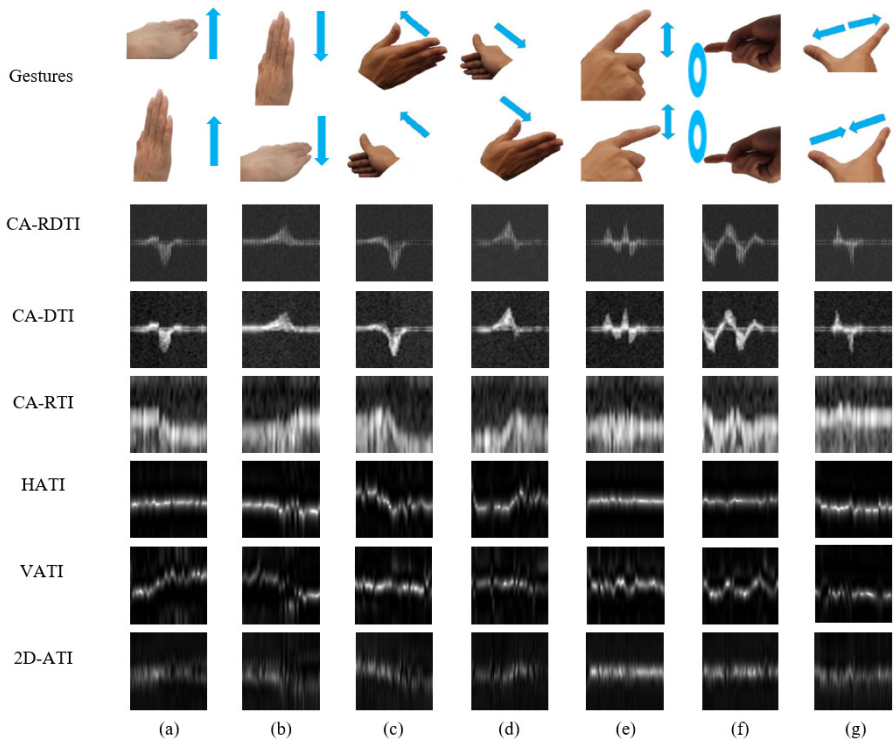


Fig. 12. Gesture diagram and feature images. (a) Palm up. (b) Palm down. (c) Palm left. (d) Palm right. (eFinger double-click. (f) Finger circle. (g) Forefinger-thumb open-close.

TABLE V
CLASSIFICATION CONFUSION MATRIX OF EIGHT TRAINING USERS BASED ON CA-DTI

| | | True category | | | | | | | | Accuracy ( % ) |
|---|---|---|---|---|---|---|---|---|---|---|
| | | Palm up | Palm down | Palm left | Palm right | Finger double-click | Fingers circle | Forefinger-thumb open-close | Non-gesture | |
| Predicted category | Palm up | 77 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 98.7 |
| | Palm down | 1 | 74 | 2 | 4 | 1 | 0 | 0 | 0 | 90.2 |
| | Palm left | 1 | 4 | 77 | 4 | 0 | 0 | 0 | 1 | 88.5 |
| | Palm right | 1 | 1 | 1 | 69 | 0 | 0 | 0 | 0 | 95.8 |
| | Finger double-click | 0 | 0 | 0 | 0 | 79 | 2 | 1 | 2 | 94 |
| | Finger circle | 0 | 0 | 0 | 0 | 0 | 78 | 0 | 0 | 100 |
| | Forefinger-thumb open-close | 0 | 0 | 0 | 0 | 0 | 0 | 79 | 0 | 100 |
| | Non-gesture | 0 | 0 | 0 | 3 | 0 | 0 | 0 | 77 | 96.3 |
| | Accuracy ( % ) | 96.3 | 92.5 | 96.3 | 86.3 | 98.8 | 97.5 | 98.8 | 96.3 | 95.3 |

and 80, respectively. First, the training samples and the verification samples are input into the CNN for training to obtain the classification models. After training, the generated CNN models are called to classify the test sets of eight training users and two test users for testing. The classification accuracy results of eight training users and two test users based on six single gesture features are shown in Table IV.

It can be seen in Table IV that the classification performances based on the CA-DTI feature and the CA-RDTI feature are close on eight training users, and the average classification accuracy corresponding to the CA-DTI is 2% higher than that corresponding to the CA-RDTI on the two test users, indicating that a single radial Doppler feature is sufficient to accurately represent the radial changes in seven gestures and has better generalization ability. Comparing the classification results corresponding to CA-DTI, CA-RTI, HATI, VATI, and 2-D-ATI features, it can be seen that the radial Doppler feature has significantly better gesture representation capability than

TABLE VI
CLASSIFICATION CONFUSION MATRIX OF THE TEST USER A BASED ON CA-DTI

| | | True category | | | | | | | | Accuracy ( % ) |
|---|---|---|---|---|---|---|---|---|---|---|
| | | Palm up | Palm down | Palm left | Palm right | Finger double-click | Fingers circle | Forefinger-thumb open-close | Non-gesture | |
| Predicted category | Palm up | 15 | 1 | 6 | 0 | 0 | 0 | 0 | 0 | 68.2 |
| | Palm down | 6 | 41 | 11 | 8 | 0 | 0 | 0 | 0 | 62.1 |
| | Palm left | 3 | 5 | 25 | 17 | 0 | 0 | 0 | 0 | 50.0 |
| | Palm right | 15 | 3 | 7 | 23 | 0 | 0 | 0 | 0 | 47.9 |
| | Finger double-click | 1 | 0 | 0 | 0 | 23 | 16 | 0 | 2 | 54.8 |
| | Finger circle | 0 | 0 | 1 | 0 | 27 | 34 | 0 | 3 | 52.3 |
| | Forefinger-thumb open-close | 0 | 0 | 0 | 0 | 0 | 0 | 46 | 0 | 100 |
| | Non-gesture | 10 | 0 | 0 | 2 | 0 | 0 | 4 | 45 | 73.8 |
| | Accuracy ( % ) | 30 | 41 | 50 | 46 | 46 | 68 | 92 | 96.3 | 63.0 |

TABLE VII
CLASSIFICATION CONFUSION MATRIX OF THE TEST USER B BASED ON CA-DTI

| | | True category | | | | | | | | Accuracy ( % ) |
|---|---|---|---|---|---|---|---|---|---|---|
| | | Palm up | Palm down | Palm left | Palm right | Finger double-click | Fingers circle | Forefinger-thumb open-close | Non-gesture | |
| Predicted category | Palm up | 32 | 0 | 10 | 2 | 0 | 0 | 0 | 0 | 72.7 |
| | Palm down | 3 | 44 | 19 | 1 | 4 | 1 | 0 | 0 | 61.1 |
| | Palm left | 7 | 0 | 14 | 19 | 0 | 0 | 0 | 0 | 35.0 |
| | Palm right | 6 | 6 | 4 | 25 | 0 | 0 | 0 | 0 | 61.0 |
| | Finger double-click | 0 | 0 | 0 | 0 | 33 | 2 | 0 | 0 | 94.3 |
| | Finger circle | 0 | 0 | 3 | 0 | 13 | 47 | 0 | 0 | 74.6 |
| | Forefinger-thumb open-close | 0 | 0 | 0 | 0 | 0 | 0 | 47 | 0 | 100 |
| | Non-gesture | 2 | 0 | 0 | 3 | 0 | 0 | 3 | 50 | 86.2 |
| | Accuracy ( % ) | 64 | 88 | 28 | 50 | 66 | 94 | 94 | 100 | 73.0 |

the radial distance feature and tangential angle features. This is because the angular resolution is low and only centimeter-level radial motion can be resolved by the radial distance; as a contrast, the radial Doppler frequency shifts are obtained by measuring the phase differences in two adjacent chirps, and thus submillimeter-level fine-grained motion can be distinguished by the radial Doppler feature [31]. Furthermore, comparing the classification results of HATI, VATI, and 2-D-ATI features, we can find that the 2-D angle feature is significantly better than 1-D to represent gestures which contain 2-D angle variations, and it fully illustrates the importance of multidimensional representation.

To evaluate the difference in gesture classification performance between the training users and the test users, the classification results of the CA-DTI feature with the best performance are further analyzed. The classification confusion matrices corresponding to eight training users and two test users are shown in Tables V–VII. Comparing the classification results of eight training users and two test users based on six single gesture features, it can be seen that although the radial CA-DTI and CA-RDTI features can obtain approximately good classification performance on the training users,

TABLE VIII
COMPARISON OF CLASSIFICATION RESULTS OF THREE MULTIDIMENSIONAL FEATURE COMBINATION STRATEGIES

| Feature type | Classification method | Average accuracy ( % ) | | |
|---|---|---|---|---|
| | | 8 Training Users | Test User A | Test User B |
| CA-RDTI + HATI + VATI | 3-channel CNN | 98.4 | 77.5 | 77.5 |
| CA-DTI + HATI + VATI | 3-channel CNN | 98.9 | 79.8 | 89.3 |
| CA-DTI + CA-RTI + 2D-ATI | 3-channel CNN | 97.2 | 67.3 | 72.3 |
| CA-DTI + CA-RTI + HATI + VATI | 4-channel CNN | 98.9 | 73.5 | 81.5 |

TABLE IX
CLASSIFICATION CONFUSION MATRIX FOR EIGHT TRAINING USERS BASED ON CA-DTI + HATI + VATI

| | | True category | | | | | | | | Accuracy ( % ) |
|---|---|---|---|---|---|---|---|---|---|---|
| | | Palm up | Palm down | Palm left | Palm right | Finger double-click | Fingers circle | Forefinger-thumb open-close | Non-gesture | |
| Predicted category | Palm up | 77 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 100 |
| | Palm down | 2 | 80 | 1 | 0 | 0 | 0 | 0 | 0 | 96.4 |
| | Palm left | 0 | 0 | 79 | 0 | 0 | 0 | 0 | 0 | 100 |
| | Palm right | 0 | 0 | 0 | 80 | 0 | 0 | 0 | 0 | 100 |
| | Finger double-click | 1 | 0 | 0 | 0 | 80 | 1 | 0 | 2 | 95.2 |
| | Finger circle | 0 | 0 | 0 | 0 | 0 | 79 | 0 | 0 | 100 |
| | Forefinger-thumb open-close | 0 | 0 | 0 | 0 | 0 | 0 | 80 | 0 | 100 |
| | Non-gesture | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 78 | 100 |
| | Accuracy ( % ) | 96.3 | 100 | 98.8 | 100 | 100 | 98.8 | 100 | 97.5 | 98.9 |

the accuracies of classification on untrained users drop sharply, especially on the four palm gestures with obvious angle changes, indicating a single radial feature cannot accurately represent a variety of 3-D motion gestures, and have poor generalization ability.

We believe that due to the limitation of the spatial dimension, the radial features are insufficient to represent 3-D motion gestures. The RDTI, DTI, and RTI features can only represent radial movements of the gestures, that is, only have a 1-D representation capability. As shown in Fig. 12, the three radial features of the palm up and palm left gestures are very similar and confusing when they are preformed keeping move away from radar, and a similar situation occurs with the palm down and palm right gestures. In addition, when the same gestures are performed at different starting positions, the radial features are different and even correspond to completely opposite radial changes. As shown in Fig. 13, for the case (a), the palm left moving at the left side of the radar is a gesture away from the radar, but for the case (e), the palm left moving at the right side of the radar is a gesture close to the radar. As a comparison, we can see consistent tangential angle features for the same palm left gestures at different spatial positions in Fig. 13.

In terms of tangential angle features, the HATI and VATI features can only represent the 1-D horizontal or vertical angle changes in 3-D motion gestures, respectively. Because of the low angular resolution, it is impossible to use only 1-D angle features to distinguish 3-D gestures with multiple freedoms of movement. The 2-D-ATI feature can help distinguish four palm gestures with significant angle changes but have limited contribution to three finger gestures with insignificant angle changes because of poor angular resolution.

To improve the ability to represent 3-D motion gestures, we propose a multidimensional feature representation method that combines radial and tangential features. The specific feature combination strategies include RDTI + HATI + VATI, DTI + HATI + VATI, DTI + RTI + 2-D-ATI, and DTI + RTI + HATI + VATI, which are based on the facts that the tangential angle features in the azimuth dimension and the elevation dimension and one radial feature are essential to represent 3-D motion information, and the radial Doppler feature is superior to the radial distance feature. The classification accuracy results of the four combination strategies are shown in Table VIII.

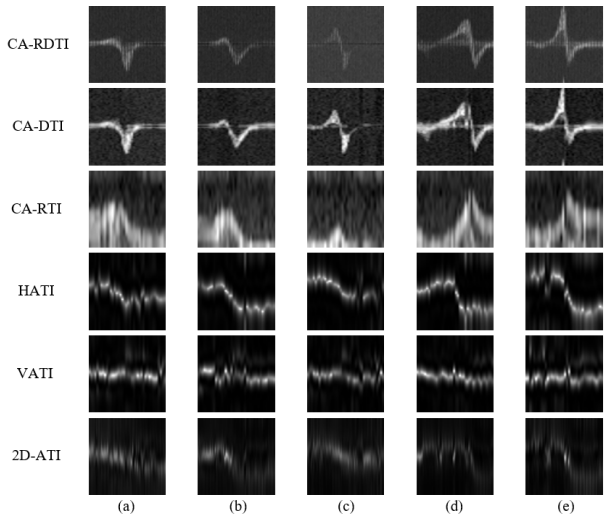


Fig. 13. Six feature images of the palm left gestures moving from different starting positions at (a) 1cm (b) 5cm (c) 10cm (d) 15cm (e) 20cm.

It can be seen in Table VIII that the feature combination strategy for CA-DTI + HATI + VATI is superior to CA-RDTI + HATI + VATI and CA-DTI + CA-RTI + 2-D-ATI, especially for two test users, which show that the latter two have poorer generalization ability. This is because the CA-RDTI feature is obtained by stitching multiframe RD images, which affects the time continuity of the radial

TABLE X
CLASSIFICATION CONFUSION MATRIX FOR THE TEST USER A BASED ON CA-DTI + HATI + VATI

| | | True category | | | | | | | | Accuracy ( % ) |
|---|---|---|---|---|---|---|---|---|---|---|
| | | Palm up | Palm down | Palm left | Palm right | Finger double-click | Fingers circle | Forefinger-thumb open-close | Non-gesture | |
| Predicted category | Palm up | 35 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 100 |
| | Palm down | 2 | 47 | 0 | 0 | 0 | 0 | 0 | 0 | 95.9 |
| | Palm left | 0 | 1 | 50 | 0 | 0 | 0 | 0 | 0 | 98.0 |
| | Palm right | 0 | 0 | 0 | 49 | 0 | 0 | 0 | 0 | 100 |
| | Finger double-click | 5 | 0 | 0 | 0 | 22 | 24 | 0 | 3 | 40 |
| | Finger circle | 0 | 0 | 0 | 0 | 27 | 26 | 0 | 0 | 49.1 |
| | Forefinger-thumb open-close | 1 | 1 | 0 | 0 | 1 | 0 | 43 | 0 | 93.5 |
| | Non-gesture | 7 | 1 | 0 | 1 | 0 | 0 | 7 | 47 | 75.8 |
| | Accuracy ( % ) | 70.0 | 94.0 | 100 | 98.0 | 44.0 | 52.0 | 86.0 | 94.0 | 79.8 |

TABLE XI
CLASSIFICATION CONFUSION MATRIX FOR THE TEST USER B BASED ON CA-DTI + HATI + VATI

| | | True category | | | | | | | | Accuracy ( % ) |
|---|---|---|---|---|---|---|---|---|---|---|
| | | Palm up | Palm down | Palm left | Palm right | Finger double-click | Fingers circle | Forefinger-thumb open-close | Non-gesture | |
| Predicted category | Palm up | 40 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 100 |
| | Palm down | 3 | 48 | 0 | 2 | 2 | 0 | 0 | 0 | 87.3 |
| | Palm left | 1 | 0 | 50 | 0 | 0 | 0 | 0 | 0 | 98.0 |
| | Palm right | 3 | 2 | 0 | 47 | 0 | 0 | 0 | 0 | 90.4 |
| | Finger double-click | 3 | 0 | 0 | 0 | 37 | 6 | 1 | 0 | 78.7 |
| | Finger circle | 0 | 0 | 0 | 0 | 10 | 41 | 0 | 0 | 80.4 |
| | Forefinger-thumb open-close | 0 | 0 | 0 | 0 | 1 | 0 | 44 | 0 | 97.8 |
| | Non-gesture | 0 | 0 | 0 | 1 | 0 | 3 | 5 | 50 | 84.7 |
| | Accuracy ( % ) | 80.0 | 96.0 | 100 | 94.0 | 74 | 82.0 | 88.0 | 100 | 89.3 |

Doppler feature, and the 2-D-ATI feature is obtained by stitching multiframe 2-D angle images, which affects the time continuity of the angle features.

Comparing the two feature combination strategies of CA-DTI + HATI + VATI and CA-DTI + CA-RTI + HATI + VATI in Table VIII, it can be seen that the classification accuracy is the same for the eight training users, but the former performs better for the two test users. This means that in the case of existing radial Doppler features, increasing the radial distance feature will reduce the classification performance. All things considered, CA-DTI + HATI + VATI is an optimal feature combination strategy of the four. The classification confusion matrices of eight training users and two test users based on that are shown in Tables IX–XI.

From Tables V–VII and IX–XI, it can be seen that compared with the single CA-DTI feature, the classification performance based on the CA-DTI + HATI + VATI representation is significantly improved for the four palm gestures; however, that becomes worse for the three finger gestures. It indicates that when the angle resolution is not high, the representation ability of the angle features for micromotion gesture is not strong.

## VI. DISCUSSION AND CONCLUSION

We believe there are two major reasons that caused recognition errors. The first reason is that the nongesture features are sometimes similar to gesture features during gesture switching. For this type of mis-classification, one can explore motion associations between different gestures to ensure that the nongesture motions between the two gestures are not captured or recognized by mistake. The second reason is the poor angle resolution, which leads to wrong results of a small quantity of palm gestures with unobvious angle changes and a large

number of finger gestures. It can be improved by increasing the angle resolution or reducing the proportion of angle feature in multidimensional feature combination.

Although the gesture recognition architecture is designed for a single-user scene, we also conduct some test experiments to see its applicability for multiuser scenario, that is, one person performs gestures with another person in background also waving around and find that the gesture recognition process is hardly affected as long as the interference target and the measured target can be separated by the radial distances, or the velocity of the interference target is not larger than the gesture velocity. Therefore, the multiuser applications can be implemented by separating multiple users with different radial distances or different radial velocities.

In the future, it is hoped that more scattering and motion characteristics of a hand can be extracted for more accurate hand-gesture recognition. We use the radar data to establish the scattering center motion model to extract the classification features of the hand, which is helpful for subsequent researchers to establish a more accurate motion model for describing nonrigid targets.

Although the proposed hand-gesture recognition method can solve the problem of translation confusion in the RD domain to ensure the translation invariance in 3-D space, it is not suitable for micromotion gestures without measurable angle changes. In addition, the rotation confusion of gestures in 3-D space is also an unsolved problem in the field of radar gesture recognition. These are the remaining issues to be studied.

In summary, a real-time gesture recognition system with mmw radar is developed, where an optimal parameter configuration procedure is first proposed. A motion equation is proposed to ensure spatial continuity and used to track scattering centers in the multidimension feature space. A multichannel CNN is designed to learn the multidimensional gesture features from data obtained via extensive experiments. The optimal multidimensional feature combination strategy is found to be DTI + HATI + VATI which shows significantly better classification performance. The gesture recognition architecture based on multidimensional feature representation and multichannel learning outperforms the existing single feature representation and single-channel learning by 3% and 16% in terms of average accuracy for four palm gestures and three finger gestures on eight training users and two nontraining users, respectively. The proposed method can effectively solve the problem of translation confusion in the conventional RD domain for the palm gestures with obvious angle changes.

## ACKNOWLEDGMENT

The authors would like to acknowledge Junyu Jie, Xiangfeng Wang, and Tao Zhou, for their help in experimental data collection.

## REFERENCES

[1] E. Ohn-Bar and M. M. Trivedi, "Hand gesture recognition in real time for automotive interfaces: A multimodal vision-based approach and evaluations," *IEEE Trans. Intell. Transp. Syst.*, vol. 15, no. 6, pp. 2368–2377, Dec. 2014.

[2] M. R. Abid, E. M. Petriu, and E. Amjadian, "Dynamic sign language recognition for smart home interactive application using stochastic linear formal grammar," *IEEE Trans. Instrum. Meas.*, vol. 64, no. 3, pp. 596–605, Mar. 2015.

[3] Y. Jang, I. Jeon, T.-K. Kim, and W. Woo, "Metaphoric hand gestures for orientation-aware VR object manipulation with an egocentric viewpoint," *IEEE Trans. Human-Mach. Syst.*, vol. 47, no. 1, pp. 113–127, Feb. 2017.

[4] N. H. Dardas and N. D. Georganas, "Real-time hand gesture detection and recognition using bag-of-features and support vector machine techniques," *IEEE Trans. Instrum. Meas.*, vol. 60, no. 11, pp. 3592–3607, Nov. 2011.

[5] Y. Qifan, T. Hao, Z. Xuebing, L. Yin, and Z. Sanfeng, "Dolphin: Ultrasonic-based gesture recognition on smartphone platform," in *Proc. IEEE 17th Int. Conf. Comput. Sci. Eng.*, Dec. 2014, pp. 1461–1468.

[6] M. A. Simao, O. Gibaru, and P. Neto, "Online recognition of incomplete gesture data to interface collaborative robots," *IEEE Trans. Ind. Electron.*, vol. 66, no. 12, pp. 9372–9382, Dec. 2019.

[7] B. G. Lee and S. M. Lee, "Smart wearable hand device for sign language interpretation system with sensors fusion," *IEEE Sensors J.*, vol. 18, no. 3, pp. 1224–1232, Feb. 2018.

[8] C. Wang, Z. Liu, and S.-C. Chan, "Superpixel-based hand gesture recognition with Kinect depth camera," *IEEE Trans. Multimedia*, vol. 17, no. 1, pp. 29–39, Jan. 2015.

[9] Y. Zou, J. Xiao, J. Han, K. Wu, Y. Li, and L. M. Ni, "GRfid: A device-free RFID-based gesture recognition system," *IEEE Trans. Mobile Comput.*, vol. 16, no. 2, pp. 381–393, Feb. 2017.

[10] M. A. A. Haseeb and R. Parasuraman, "Wisture: Touch-less hand gesture classification in unmodified smartphones using Wi-Fi signals," *IEEE Sensors J.*, vol. 19, no. 1, pp. 257–267, Jan. 2019.

[11] D. M. Vavriv, O. O. Bezvesilniy, V. A. Volkov, A. A. Kravtsov, and E. V. Bulakh, "Recent advances in millimeter-wave radars," in *Proc. Int. Conf. Antenna Theory Techn. (ICATT)*, Kharkiv, Ukraine, Apr. 2015, pp. 1–6.

[12] B. W. Battershall and S. P. Emmons, "Optimization of diode structures for monolithic integrated microwave circuits," *IEEE Trans. Microw. Theory Techn.*, vol. MTT-16, no. 7, pp. 445–450, Jul. 1968.

[13] H. Kusamitsu, Y. Morishita, K. Maruhasi, M. Ito, and K. Ohata, "The flip-chip bump interconnection for millimeter-wave GaAs MMIC," *IEEE Trans. Electron. Packag. Manuf.*, vol. 22, no. 1, pp. 23–28, Jan. 1999.

[14] S. T. Nicolson, P. Chevalier, B. Sautreuil, and S. P. Voinigescu, "Single-chip W-band SiGe HBT transceivers and receivers for Doppler radar and millimeter-wave imaging," *IEEE J. Solid-State Circuits*, vol. 43, no. 10, pp. 2206–2217, Oct. 2008.

[15] B. Razavi, "A millimeter-wave CMOS heterodyne receiver with on-chip LO and divider," *IEEE J. Solid-State Circuits*, vol. 43, no. 2, pp. 477–485, 2008.

[16] M. Steinhauer, H.-O. Ruoss, H. Irion, and W. Menzel, "Millimeter-wave-radar sensor based on a transceiver array for automotive applications," *IEEE Trans. Microw. Theory Techn.*, vol. 56, no. 2, pp. 261–269, Feb. 2008.

[17] V. Ziegler, F. Schubert, B. Schulte, A. Giere, R. Koerber, and T. Waanders, "Helicopter near-field obstacle warning system based on low-cost millimeter-wave radar technology," *IEEE Trans. Microw. Theory Techn.*, vol. 61, no. 1, pp. 658–665, Jan. 2013.

[18] M. Moallem and K. Sarabandi, "Polarimetric study of MMW imaging radars for indoor navigation and mapping," *IEEE Trans. Antennas Propag.*, vol. 62, no. 1, pp. 500–504, Jan. 2014.

[19] J. M. Munoz-Ferreras, F. Perez-Martinez, J. Calvo-Gallego, A. Asensio-Lopez, B. P. Dorta-Naranjo, and A. Blanco-del-Campo, "Traffic surveillance system based on a high-resolution radar," *IEEE Trans. Geosci. Remote Sens.*, vol. 46, no. 6, pp. 1624–1633, Jun. 2008.

[20] W.-Q. Wang, Q. Peng, and J. Cai, "Waveform-diversity-based millimeter-wave UAV SAR remote sensing," *IEEE Trans. Geosci. Remote Sens.*, vol. 47, no. 3, pp. 691–700, Mar. 2009.

[21] S. Bertl and J. Detlefsen, "Effects of a reflecting background on the results of active MMW SAR imaging of concealed objects," *IEEE Trans. Geosci. Remote Sens.*, vol. 49, no. 10, pp. 3745–3752, Oct. 2011.

[22] J.-T. Gonzalez-Partida, P. Almorox-Gonzalez, M. Burgos-Garcia, B.-P. Dorta-Naranjo, and J. I. Alonso, "Through-the-wall surveillance with millimeter-wave LFMCW radars," *IEEE Trans. Geosci. Remote Sens.*, vol. 47, no. 6, pp. 1796–1805, Jun. 2009.

[23] M. Ferri, G. Giunta, A. Banelli, and D. Neri, "Millimetre wave radar applications to airport surface movement control and foreign object detection," in *Proc. Eur. Radar Conf. (EuRAD)*, Rome, Italy, 2009, pp. 437–440.

[24] A. Prat, S. Blanch, A. Aguasca, J. Romeu, and A. Broquetas, "Collimated beam FMCW radar for vital sign patient monitoring," *IEEE Trans. Antennas Propag.*, vol. 67, no. 8, pp. 5073–5080, Aug. 2019.

[25] M. Vahidpour and K. Sarabandi, "Millimeter-wave Doppler spectrum and polarimetric response of walking bodies," *IEEE Trans. Geosci. Remote Sens.*, vol. 50, no. 7, pp. 2866–2879, Jul. 2012.

[26] X. Bai, Y. Hui, L. Wang, and F. Zhou, "Radar-based human gait recognition using dual-channel deep convolutional neural network," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 12, pp. 9767–9778, Dec. 2019.

[27] J. Lien *et al.*, "Soli: Ubiquitous gesture sensing with millimeter wave radar," *ACM Trans. Graph.*, vol. 35, no. 4, pp. 1–19, 2016.

[28] S. Wang, J. Song, J. Lien, I. Poupyrev, and O. Hilliges, "Interacting with soli: Exploring fine-grained dynamic gesture recognition in the radio-frequency spectrum," in *Proc. 29th Annu. Symp. ACM*, 2016, pp. 851–860.

[29] 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.

[30] K. A. Smith, C. Csech, D. Murdoch, and G. Shaker, "Gesture recognition using mm-Wave sensor for human-car interface," *IEEE Sensors Lett.*, vol. 2, no. 2, pp. 1–4, Jun. 2018.

[31] Z. Xia, C. Zhou, J. Jie, T. Zhou, X. Wang, and F. Xu, "Micro-motion gesture recognition based on multi-channel frequency modulated continuous wave millimeter wave radar," *J. Electron. Inf. Technol.*, vol. 42, no. 1, pp. 164–172.

[32] G. Malysa, D. Wang, L. Netsch, and M. Ali, "Hidden Markov model-based gesture recognition with FMCW radar," in *Proc. IEEE Global Conf. Signal Inf. Process. (GlobalSIP)*, Washington, DC, USA, Dec. 2016, pp. 1017–1021.

[33] 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.

[34] 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.

[35] A. Ghaffar, F. Khan, and S. H. Cho, "Hand pointing gestures based digital menu board implementation using IR-UWB transceivers," *IEEE Access*, vol. 7, pp. 58148–58157, 2019.

[36] M. Jankiraman, *FMCW Radar Design*. London, U.K.: Artech House, 2018.

[37] J. B. Keller, "Geometrical theory of diffraction," *J. Opt. Soc. Amer.*, vol. 52, no. 2, pp. 16–130, 1962.

[38] L. C. Potter, D.-M. Chiang, R. Carriere, and M. J. Gerry, "A GTD-based parametric model for radar scattering," *IEEE Trans. Antennas Propag.*, vol. 43, no. 10, pp. 1058–1067, 1995.

[39] R. Feger, C. Wagner, S. Schuster, S. Scheiblhofer, H. Jager, and A. Stelzer, "A 77-GHz FMCW MIMO radar based on an SiGe single-chip transceiver," *IEEE Trans. Microw. Theory Techn.*, vol. 57, no. 5, pp. 1020–1035, May 2009.

[40] R. K. Amineh, A. Khalatpour, and N. K. Nikolova, "Three-dimensional microwave holographic imaging using co- and cross-polarized data," *IEEE Trans. Antennas Propag.*, vol. 60, no. 7, pp. 3526–3531, Jul. 2012.

[41] S.-H. Jung, Y.-S. Cho, R.-S. Park, J.-M. Kim, H.-K. Jung, and Y.-S. Chung, "High-resolution millimeter-wave ground-based SAR imaging via compressed sensing," *IEEE Trans. Magn.*, vol. 54, no. 3, pp. 1–4, Mar. 2018.

[42] Z. Zhao *et al.*, "Point cloud features-based kernel SVM for human-vehicle classification in millimeter wave radar," *IEEE Access*, vol. 8, pp. 26012–26021, 2020.

[43] F. C. Robey, D. R. Fuhrmann, E. J. Kelly, and R. Nitzberg, "A CFAR adaptive matched filter detector," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 28, no. 1, pp. 208–216, Jan. 1992.

[44] T. Wagner, R. Feger, and A. Stelzer, "Radar signal processing for jointly estimating tracks and micro-Doppler signatures," *IEEE Access*, vol. 5, pp. 1220–1238, 2017.

**Zhaoyang Xia** (Graduate Student Member, IEEE) received the B.E. degree in physics and the M.S. degree in optics from Zhengzhou University, Zhengzhou, China, in 2016 and 2018, respectively. He is pursuing the Ph.D. degree in electronics and information with the School of Information Science and Technology, Fudan University, Shanghai, China.

His research interests include radar signal processing and target recognition.

**Yixiang Luomei** (Graduate Student Member, IEEE) received the B.E. and M.E. degrees in electronic and communication engineering from Xidian University, Xi'an, China, in 2014 and 2018, respectively. She is pursuing the Ph.D. degree in electronics and information with the School of Information Science and Technology, Fudan University, Shanghai, China.

Her research interests include multi-input multi-output (MIMO) radar signal processing and unmanned aerial vehicle (UAV) synthetic aperture radar (SAR) imaging.

**Chenglong Zhou** received the B.E. degree in electronic information engineering from the Dalian University of Technology, Dalian, China, in 2018. He is pursuing the M.S. degree in electronics and communication engineering with the School of Information Science and Technology, Fudan University, Shanghai, China.

His research interests include radar signal processing and posture recognition.

**Feng Xu** (Senior Member, IEEE) received the B.E. degree (Hons.) in information engineering from Southeast University, Nanjing, China, in 2003, and the Ph.D. degree (Hons.) in electronic engineering from Fudan University, Shanghai, China, in 2008.

From 2008 to 2010, he was a Post-Doctoral Fellow with the NOAA Center for Satellite Application and Research (STAR), Camp Springs, MD, USA. From 2010 to 2013, he was a Research Scientist with Intelligent Automation Inc., Rockville, MD. Since 2013, he has been a Professor with the School of Information Science and Technology, Fudan University, where he serves as the Vice-Dean and the Director of the Key Laboratory for Information Science of Electromagnetic Waves, Ministry of Education (MoE). He has authored more than 60 articles in peer-reviewed journals and coauthored three books, among many conference articles and patents. His research interests include electromagnetic scattering theory, synthetic aperture radar (SAR) information retrieval, and advanced radar systems.

Dr. Xu was a recipient of the Second-Class National Nature Science Award of China in 2011, the 2014 Early Career Award of the IEEE Geoscience and Remote Sensing Society, and the 2007 SUMMA Graduate Fellowship in the advanced electromagnetics area. He serves as an Associate Editor for IEEE GEOSCIENCE AND REMOTE SENSING LETTERS. He is the Founding Chair of the IEEE GRSS Shanghai Chapter.