A Novel Detection and Recognition Method for Continuous Hand Gesture Using FMCW Radar

YONG WANG, (Member, IEEE), **AIHU REN, MU ZHOU**, (Senior Member, IEEE), **WEN WANG, AND XIAOBO YANG**

School of Communications and Information Engineering, Chongqing University of Posts and Telecommunications, Chongqing 400065, China

Corresponding author: Yong Wang (yongwang@cqupt.edu.cn)

ABSTRACT In this article, a novel method for continuous hand gesture detection and recognition is proposed based on a frequency modulated continuous wave (FMCW) radar. Firstly, we adopt the 2-Dimensional Fast Fourier Transform (2D-FFT) to estimate the range and Doppler parameters of the hand gesture raw data, and construct the range-time map (RTM) and Doppler-time map (DTM). Meanwhile, we apply the Multiple Signal Classification (MUSIC) algorithm to calculate the angle and construct the angle-time map (ATM). Secondly, a hand gesture detection method is proposed to segment the continuous hand gestures using a decision threshold. Thirdly, the central time-frequency trajectory of each hand gesture spectrogram is clustered using the k-means algorithm, and then the Fusion Dynamic Time Warping (FDTW) algorithm is presented to recognize the hand gestures. Finally, experiments show that the accuracy of the proposed hand gesture detection method can reach 96.17%. The hand gesture average recognition accuracy of the proposed FDTW algorithm is 95.83%, while its time complexity is reduced by more than 50%.

INDEX TERMS FMCW radar, continuous hand gesture recognition detection, time-frequency trajectory, FDTW.

I. INTRODUCTION

With the rapid development of 5G communications [1] and artificial intelligence, human-computer interaction (HCI) has become an indispensable technology in our daily life. As an important branch, hand gesture recognition [2] is of great significance for promoting the development of HCI. Meanwhile, the development of hand gesture recognition also promotes the progress of many fields, such as smart home [3], industrial internet of things [4]–[6], sign language interaction [7], radio frequency identification [8], [9], etc. As a result, hand gesture recognition has become a research hotspot recently [9]–[13].

The research of hand gesture recognition mainly focuses on the recognition algorithms design with different data sources. The hand gesture recognition method is divided into three types: the sensor-based [10], [11], the vision-based [12] and the radar-based ones [13]. Specifically, the sensor-based hand gesture recognition method adopts sensors to collect the

hand gesture data. However, since the tester needs to wear a wearable device for a long time, the sensor-based method is usually inconvenient in terms of user experience. The vision-based method has a very high recognition accuracy because of the adopted camera images; however, this method is not friendly in terms of privacy exposure. The radar-based hand gesture recognition method applies radar to collect the hand gesture signal, and analyzes the hand gesture information through the signal processing, and then classifies the hand gestures. Compared with the two aforementioned ones the radar-based hand gesture recognition method performs in a non-contact way and brings a good user experience, as a result, it attracts extensive attention in both industry [14] and academic [15], [16]. In [15], the authors adopt the radar to measure the range, speed and angle information of the hand gestures. In the vehicle-mounted auxiliary control system [16], the authors apply the time-frequency analysis method on the beat signals of different hand gestures using a frequency modulated continuous wave (FMCW) millimeter-wave radar. Moreover, the Google Soli project [17] designs

a hand gesture recognition system using a 7 GHz bandwidth radar.

Nowadays, the methods of hand gesture recognition are usually tested using a single hand gesture [18], [19]. However, in the real-time applications of the hand gesture recognition system, the hand gestures are usually continuous and dynamic. To this end, some researchers focus on hand gesture recognition with continuous and dynamic hand gestures [20]–[23]. In [20], the authors obtain the start and end times of each single gesture by a near-zero speed detection method. However, the continuous hand gesture considered in [20] is periodical, and such a simple hand gesture type is limited to the practical applications of complex hand gestures. In [21], the authors judge the start and end of the hand gesture based on the thresholds of the calculated trigger value and ratio. However, with the number of hand gestures increases, the instability of the trigger ratio makes it difficult to detect hand gestures correctly. For the hand gesture recognition algorithms, deep learning [24]–[28] is usually used for feature extraction and hand gesture classification. Since deep learning method requires a large number of datasets for training, and the hand gesture collection time and the training time increase sharply with the increase of hand gesture types. More importantly the duration time of each hand gesture is usually different from each other, making it difficult to directly use the deep learning method for hand gesture recognition. Therefore, the design of hand gesture recognition algorithms based on the non-deep learning such as dynamic time warping (DTW) [29]–[31], Adaptive Boosting (AdaBoost) [32], [33] and Hidden Markov model (HMM) [34], [35] has received much attention. The data imbalance of AdaBoost results in the decrease of classification accuracy, and the calculation of the HMM model is complex. Since the DTW algorithm can match the similarity of two sequences with different time lengths, it is widely used in the field of hand gesture recognition. In [36], the authors propose an RGB video-based hand gesture recognition method for posture extraction, and apply the DTW algorithm for posture classification. Reference [20] adopts the DTW algorithm with the combination of the range and Doppler information for hand gestures classification, and achieves a recognition accuracy of 91.5%. In [37], a novel DTW algorithm for hand gesture recognition of wearable gloves is proposed. In [38], based on simple gesture database, the efficiency of different DTW algorithms is compared. However, the hand gesture data used in the above studies are mostly based on cameras or sensor devices. Moreover, the above works focus on the design of hand gesture recognition algorithms for a single hand gesture, and the problem of continuous hand gesture recognition is not discussed. More importantly, the time complexity of the DTW algorithm [39] is usually high because all pixels of the hand gesture spectrogram are used for matching, and the angle information of hand gesture in the above studies is still not mentioned.

Therefore, this article proposes a hand detection and recognition method for continuous and dynamic hand gestures using a FMCW radar. Specifically, we use the hand gesture raw data to obtain the intermediate frequency (IF) signal, and estimate the range-time map (RTM) and Doppler-time map (DTM). Besides, we adopt the Multiple Signal Classification (MUSIC) [40] algorithm to estimate the angle-time map (ATM) of different hand gestures. Then, the amplitude is obtained by normalizing the hand gesture spectrogram, and a threshold is set to effectively segment the continuous hand gestures. Moreover, we present a hand gesture detection range method to evaluate the segmented accuracy of the hand gestures. The central time-frequency trajectory of each hand gesture is clustered by the k-means algorithm, and the Fusion Dynamic Time Warping (FDTW) algorithm is proposed to classify the hand gestures. The experimental results show that the accuracy of the proposed gesture detection method is 96.17%, and compared to the existing alternatives, the recognition accuracy of the proposed FDTW algorithm is improved by more than 5%, and the time complexity is reduced by about 50%.

The reminder of this article is organized as follows. In Section II, we describe the FMCW radar signal processing, including the generation of IF signal and the construction of spectrogram. In Section III, we present the details of the hand gesture detection and recognition methods. Section IV presents the experiments and discussions. Finally, the conclusions are given in V.

II. FMCW RADAR SIGNAL PROCESSING
A. IF SIGNAL EXTRACTION AND RDM CONSTRUCTION
In this article, the FMCW radar is used for hand gesture data acquisition. The FMCW radar consists of a waveform generator, an antenna array with two transmitters and four receivers, a signal demodulator and an analog-to-digital converter (ADC) converter. The waveform generator transmits the chirp signal through the transmitting antenna. Then the IF signal is obtained using a low frequency filter (LPF). FIGURE 1 shows the IF signal processing.
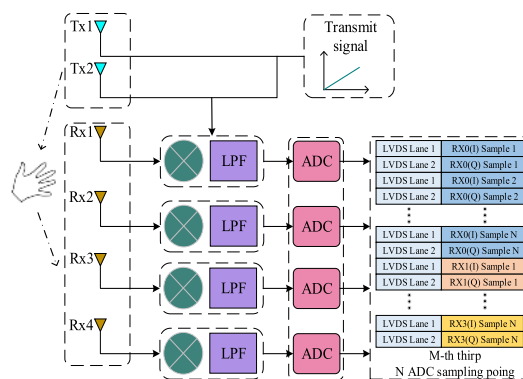

FIGURE 1. IF signal processing.

From FIGURE 1 we learn the storage format of radar data. For the receiving antennas, two Low Voltage Differential Signaling (LVDS) channels are used, and the first channel sends odd samples while the other one sends even samples,
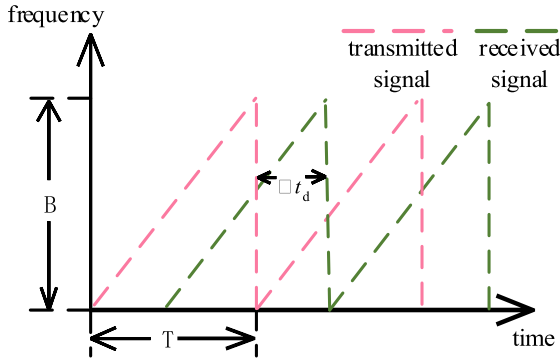
FIGURE 2. The frequencies of the signal waveform.

FIGURE 3. The relationship between IF and delay time.
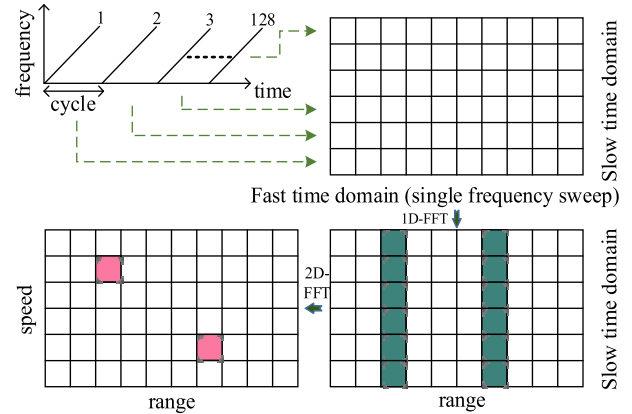
FIGURE 4. IF signal generation RDM process.

and each LVDS channel sends one real and one imaginary part signal. For each chirp signal, the data is stored starting from the lowest receiver to the highest one. With known the storage principle, the data of the I/Q signals can be obtained.

Since we use the sawtooth wave as the transmitted signals, the received signals are also sawtooth wave, and the frequencies of the transmitted and received signals are shown in FIGURE 2. The transmitted signal is expressed as

$$s_T(t) = A_T con2\pi(f_c t + \int_0^t f_T(\tau)d\tau), \qquad (1)$$

where A_T is the amplitude of the transmission signal, f_c is the central frequency of the carrier signal, $f_T(\tau) = \tau B/T$ is the frequency of the transmitted signal and it increases linearly with time τ during the period of T, and T is the width of the pulse signal, B is the bandwidth of the signal

The received signal can be expressed as

$$s_R(t) = A_R \cos 2\pi(f_c(t - \Delta t_d) + \int_0^t f_R(\tau)d\tau), \qquad (2)$$

where A_R is the amplitude of the received signal, Δt_d is the delayed time, $f_R(\tau)$ is the frequency of the received signal.

By mixing the transmitted and the received signals we get the mixed signal $S_M(t) = s_T(t) \cdot s_R(t)$, and an LPF is used to filter out the high frequency part and obtain the IF signal $s_{IF}(t)$.

$$s_{IF}(t) = \frac{1}{2}A_T A_R \cos\left\{2\pi\left[f_c\Delta t_d + (\frac{B}{T}\Delta t_d - \Delta f_{doppler})t\right]\right\}, \qquad (3)$$

where $\Delta f_{doppler}$ is the Doppler shift caused by the movement of hand gestures.

Assume that the range between the hand gesture and radar is R. The relationship between Δt_d and the R is

$$\Delta t_d = \frac{2R}{c} \qquad (4)$$

where c is the speed of the light.

Then, we can measure the delayed time Δt_d to obtain the range R. Since Δt_d is very small and it is difficult to evaluate in practical. The frequency f_{IF} of the IF signal is generally
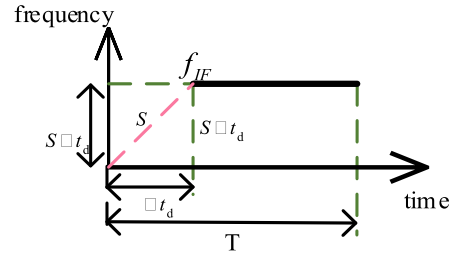
used for the estimation of Δt_d, and the relationship between f_{IF} and Δt_d is shown in FIGURE 3.

From FIGURE 3, we learn that

$$\frac{f_{IF}}{\Delta t_d} = \frac{B}{T}. \qquad (5)$$

Then, substituting Eq. (4) into Eq. (5), we have

$$R = f_{IF}\frac{cT}{2B}. \qquad (6)$$

For the dynamic hand gestures, the phase of the signal changes with the range of the hand gesture and the phase difference between each chirp can be expressed as

$$\Delta\phi = 2\pi f_c\Delta t_d = \frac{4\pi\Delta d}{\lambda}\frac{4\pi vT_c}{\lambda}. \qquad (7)$$

The Doppler frequency shift of IF signal on each chirp can be obtained by the 2D-FFT. The speed of the hand gesture can be obtained using the frequency of IF signal and Eq. (8) given by

$$v = \frac{\lambda\Delta\phi}{4\pi T_c} = \frac{\lambda\Delta f_{dopper}}{4T_c}. \qquad (8)$$

Combining Eq. (6) and Eq. (8), we know that the range is proportional to the frequency of the IF signal, and the speed of the hand gesture is proportional to the Doppler shift. In the following, a range-Doppler map (RDM) of the hand gesture is generated, and the construction process of RDM is shown in FIGURE 4. It can be seen from FIGURE 4 that in the fasttime domain (each single chirp), the frequency spectrum

of the IF signal is obtained by performing FFT on each chirp signal, and the frequency corresponding to the hand gesture can be obtained by a spectral peak search approach. The speed (Doppler) estimation needs to accumulate the estimated spectrum results and performs FFT on multiple chirps (namely a frame). Then, the coupling RDM is obtained by cumulating the slow-time domain of the range and speed. Since the range and Doppler frequency shift of the hand gesture can be extracted from RDM on each frame, RTM and DTM are obtained using multi-frame cumulation.

B. MUSIC BASED ATM

Although RTM and DTM represent the features of the hand gestures, the angle parameters provides more features of the hand gestures. In FMCW radar, the phase difference of multiple antennas on the echo signal is used to measure the angle information, shown in FIGURE 5. In the following, the mathematical model is established to calculate the angle of hand gestures.
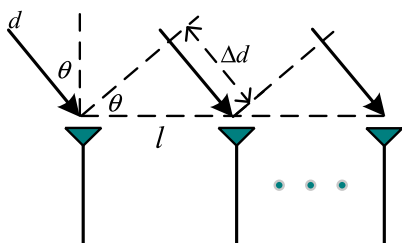
FIGURE 5. The principle of angle estimation.

The phase difference of the signals received by two adjacent receiving antennas is

$$\Delta\varphi = \frac{2\pi\,\Delta d}{\lambda}, \qquad (9)$$

where $\Delta d = l \cdot \sin\theta$ is the distance difference, l is the distance difference between the two antennas, θ is the arrival angle of signal. Then, the angle θ can be obtained

$$\theta = \sin^{-1}\left(\frac{\lambda\Delta\varphi}{2\pi l}\right). \qquad (10)$$

Based on the above analysis, this article adopts the Multiple Signal Classification (MUSIC) algorithm [40] to estimate the angle of the hand gestures. The MUSIC algorithm is based on signal subspace decomposition, and the signal and noise subspace can be obtained by characteristic decomposition of the covariance matrix. The spatial spectrum function is constructed, and then the angle is detected by a spectral peak search approach. When the kth target is found the phase difference between the first receiving antenna and the m receiving antenna is

$$\Delta\varphi = (m-1)\frac{2\pi \cdot \sin\theta_k}{\lambda}. \qquad (11)$$

Therefore, the received signal of the m-th antenna can be expressed as

$$x_m(t) = \sum_k^K a_k(t)e^{-j\cdot\Delta\varphi_k} + n_m(t), \qquad (12)$$

where $a_k(t)$ is the amplitude of the k-th target, $n_m(t)$ is the noise received by the m-th antenna. We define the guide vector $S = [s_m(\theta_k)]$, where

$$s_m(\theta_k) = e^{-j\cdot\Delta\varphi_k} = e^{-j(m-1)\frac{2\pi l \cdot \sin\theta_k}{\lambda}}. \qquad (13)$$

Thus, the signal of the m-th receiving antenna can be further expressed as

$$x_m(t) = \sum_k^K a_k(t)s_m(\theta_k) + n_m(t). \qquad (14)$$

The received signals of all antennas is

$$X = OS + J, \qquad (15)$$

where O is the amplitude matrix, S is the guide vector matrix, J is the noise matrix, and these variables can be further expressed as

$$X = [x_1(t), x_2(t), \ldots x_M(t)]^T, \qquad (16)$$

$$O = [a_1(t), a_2(t), \ldots a_K(t)], \qquad (17)$$

$$S = [s(\theta_1), s(\theta_2), \ldots s(\theta_K)]^T$$
$$= \begin{bmatrix} 1 & e^{-j\Delta\varphi_1} & \ldots & e^{-j(M-1)\Delta\varphi_1} \\ 1 & e^{-j\Delta\varphi_2} & \ldots & e^{-j(M-1)\Delta\varphi_2} \\ \ldots & \ldots & \ldots & \ldots \\ 1 & e^{-j\Delta\varphi_K} & \ldots & e^{-j(M-1)\Delta\varphi_K} \end{bmatrix}, \qquad (18)$$

$$J = [n_1(t), n_2(t), \ldots a_M(t)]^T. \qquad (19)$$

Then, the covariance matrix of radar signal is

$$\begin{aligned} R_X &= E[XX^H] \\ &= OE[SS^H]O^H + E[JJ^H] \\ &= OR_S O + R_J, \end{aligned} \qquad (20)$$

where $R_S = E[SS^H]$ is the correlation matrix of the signal and $R_J = \sigma^2 I$ is the noise, where σ^2 is noise power and I is the identity matrix.

By applying the decomposition of the covariance matrix, M eigenvalues of the matrix can be obtained, and the eigenvalues can be obtained by sorting the eigenvalues as

$$0 < \lambda_M \leq \ldots \leq \lambda_2 \leq \lambda_1. \qquad (21)$$

It is assumed that there are K hand gestures in front of the radar, the signal is corresponded to the K larger eigenvalues, while the noise is corresponded to $M-K$ smaller eigenvalues. We also assume that the eigenvector of the i-th eigenvalue λ_i is ζ_i then the noise matrix can be constructed from the $M-K$ eigenvectors

$$E_J = [\zeta_{K+1}, \zeta_{K+2}, \ldots \zeta_M]. \qquad (22)$$

The spatial spectral function can be constructed by using the noise matrix E_J and guidance vector $s(\theta)$

$$P(\theta) = \frac{1}{s^H(\theta)E_J E_J^H s(\theta)} = \frac{1}{\left\| E_J^H s(\theta) \right\|^2}. \quad (23)$$

When the inner product of signal vector $s(\theta)$ and noise E_J approaches 0, the spectral function has peak value. Therefore, the angle of the hand gesture can be found by searching the angle of the spectral function, and then using the spatial spectrum to search the spectral peak. Finally, we can obtain an ATM by accumulating the estimated angles with multiple frames.

III. PROPOSED DETECTION AND RECOGNITION METHOD

In order to achieve recognition with dynamic and continuous hand gestures, this article proposes a two-step scheme, illustrated in FIGURE 6. In step one, the continuous hand gestures are detected and segmented using the proposed amplitude based detection method, and the detected hand gesture is recognized by the proposed FDTW method in step two. The FDTW method contains the offline and online stages. In offline stage, the k-means algorithm is employed to cluster the central trajectory of all hand gesture spectrums, and the FDTW is applied for matched distance training. In online stage, the testing data of the hand gestures are clustered and input into the FDTW model for matching distance calculation and hand gesture recognition.
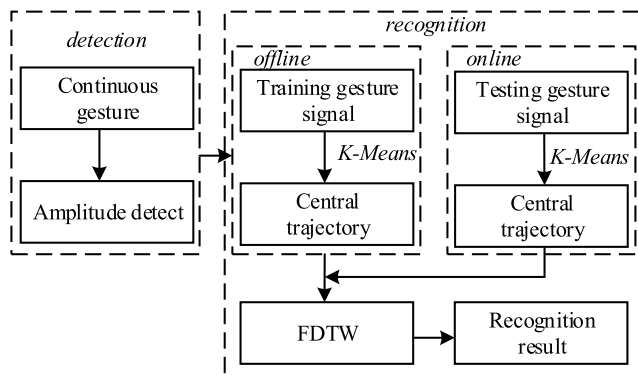

FIGURE 6. Detection and recognition methods framework.

A. HAND GESTURE DETECTION METHOD

In real-time applications, the hand gesture is usually continuous and dynamic. Therefore, the first step for hand gesture recognition is how to detect and segment each gesture from the continuous and dynamic real-time hand gesture data. In this article, we propose a simple but effective continuous hand gesture detection method based on the amplitude of hand gesture spectrograms. The detailed steps are as follows.

Firstly, the hand gesture spectrograms of RTM, DTM and ATM are respectively normalized using the

following equation.

$$A_{ij} = \frac{R_{ij} - \sum_{i}^{n'} \sum_{j}^{m'} \frac{R_{ij}}{n' \cdot m'}}{\left\| \max_{i \in (1,n'), j \in (1,m')} (R_{ij}) - \min_{i \in (1,n'), j \in (1,m')} (R_{ij}) \right\|}, \quad (24)$$

where A_{ij} is the normalized spectrograms R_{ij} is value of the original hand gesture spectrograms m' and n' are the column and row of the hand gesture spectrograms, respectively.

Then, we sum the amplitude of each frame of hand gesture spectrogram

$$Y_{sumi} = \sum_{j=1}^{m'} A_{ij}, \quad (25)$$

where Y_{sumi} is the summation amplitude of the normalized spectrogram on the i-th frame

Finally, we set a threshold, when the amplitude is first larger than the threshold, we mark it as the starting frame of the hand gesture. The hand gesture is ended when the Y_{sumi} is no larger than the threshold. The hand gesture area is

$$G_{area} = Gesture\ frame\ Y_{sumi} > threshold, \quad (26)$$

where G_{area} is the gesture frame area, i is the frame number. In this article, we found that the value of Y_{sumi} is usually larger than zero when there is a hand gesture, otherwise it is less than zero. Therefore, the threshold used for hand gesture segmentation is zero.
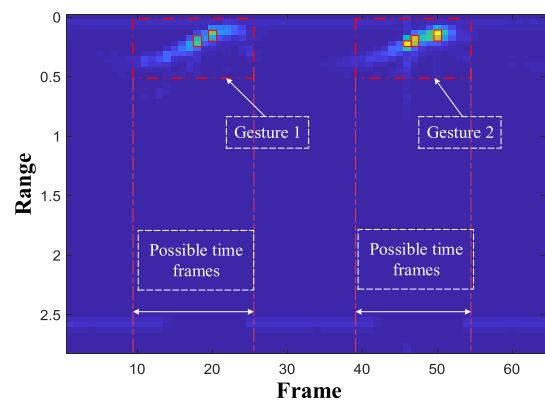
FIGURE 7. The original RTM.

The RTM of two continuous push hand gestures before amplitude calculation is shown in FIGURE 7, and the corresponding results of the amplitude based hand gesture detection method is given in FIGURE 8. We learn from FIGURE 7 that the starting and ending frames of the gestures are not clear. In FIGURE 8, by applying the amplitude based detection method, the duration time of the first hand gesture is 16 frames (from the 10-th frame to the 25-th frame), and the second hand gesture starts from the 39-th frame to the 54-th frame.

FIGURE 8. The result of amplitude.

B. HAND GESTURE RECOGNITION ALGORITHM

In this subsection, we propose a FDTW hand gesture recognition algorithm, which applies RTM, DTM and ATM to improve hand gesture recognition accuracy. Since the spectrograms of the three parameters contains too many pixels, it may cost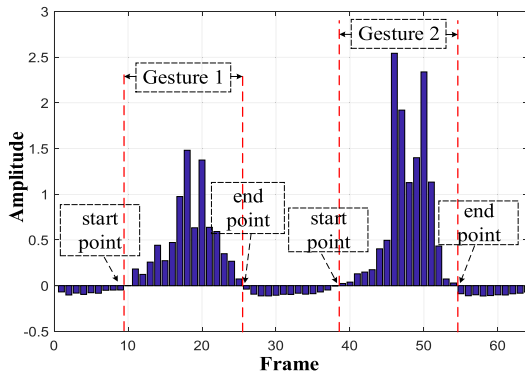 much time by calculating cosine similarity point by point. Therefore, in the offline stage, to reduce the redundant input data of hand gesture spectrograms, the k-means algorithm is presented to cluster the hand gesture central trajectory. Then, the cosine similarity of the three parameters are respectively calculated as the distances. Finally, the three distances are fused and classified by the FDTW algorithm. The block diagram of proposed algorithm is shown in FIGURE 9.
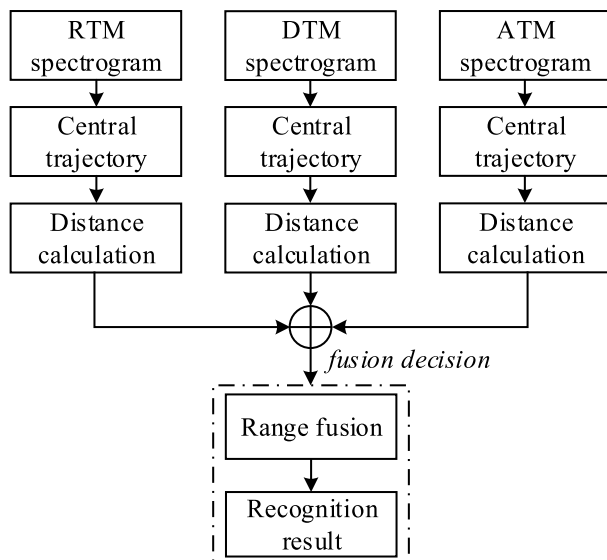

FIGURE 9. The proposed FDTW algorithm.

1) HAND GESTURE CLUSTERING

The k-means algorithm is applied to extract the center trajectory of the gesture spectrogram, and the clustered points describe the features of the hand gestures. In this article,

the spectrograms of the three parameters (RTM, DTM and ATM) are two-dimensional matrixes. We assume that there are n gesture features in each spectrogram, then we cluster the hand gesture features of each spectrogram as hand gesture central trajectory (contains k ($k <= n$) points). The specific steps of the k-means algorithm for hand gesture spectrogram clustering is given as follows.

Step 1, we randomly select k hand gesture feature samples of RTM, DTM and ATM, respectively, as the initial central trajectory points, record as $u_1, u_2, \ldots, u_k$ for each hand gesture spectrogram. In this article, we choose the value of k as 10.

Step 2, we define the cost function as

$$\Psi(c, \mu) = \min \frac{1}{\mu c} \sum_{i=1}^{M'} \left\| x_i - \mu_{c_i} \right\|^2, \quad (27)$$

where x_i is i−th hand gesture sample, c_i is the cluster of x_i, u_{c_i} is the central trajectory point of cluster, and M is total number of hand gesture samples.

Step 3, we define $t = 0$ as the iteration number, and repeat the following process until the value of the cost function $\Psi_{(c,\mu)}^t \leq 10^{-5}$

a) Assign x_i to the nearest cluster

$$c_i^{(t+1)} = \arg \min \frac{\left\| x_i - \mu_{c_i}^{(t)} \right\|^2}{k}, \quad (28)$$

b) Let $t = t + 1$. For each type of cluster, we recalculate the central trajectory point of the cluster

$$\Psi_{(c,\mu)}^t = \min \frac{1}{\mu c} \sum_{i=1}^{M'} \left\| x_i - \mu_{c_i} \right\|^2. \quad (29)$$

Therefore, the features in the hand gesture spectrogram of RTM, DTM and ATM are, respectively, extracted and represented by the k clustered points.
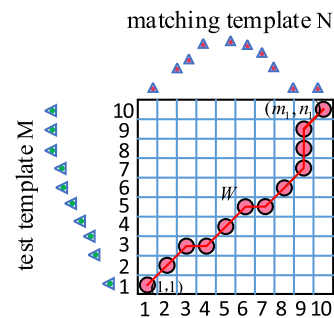

FIGURE 10. The DTW model.

2) FDTW ALGORITHM FOR GESTURE RECOGNITION

The DTW model for hang gesture recognition is shown in FIGURE 10, which gives the relationship between the test template and the matching template, and there are $m_1 = n_1 = 10$ points. The horizontal and vertical axes represent the hand gesture frames, and each small cell in the

grid represents the intersection of the test template and the matching template. The similarity between the two templates of different lengths can be measured by finding the shortest path.

To apply the DTW algorithm for hand gesture recognition, the following three conditions should be satisfied.

a) Monotonicity: if the shortest path is $W_k = (i, j)_k$ then the next point W_{k+1} in the path must be one of $(i + 1, j)_{k+1}, (i, j + 1)_{k+1}$, or $(i + 1, j + 1)_{k+1}$

b) Continuity: continuity ensures that every coordinate in sequence N and M appears in path W.

c) Starting and ending constraints: the boundary condition guarantees that the path starts from $W(1, 1)$ to the end of the gesture $W(m_1, n_1)$.

Fortunately, the three conditions are satisfied since the clustered central trajectory points are sequences. In order to improve the accuracy of hand gesture recognition, this article proposes the FDTW recognition algorithm based on different hand gesture features of RTM, DTM and ATM.

First, we calculate the matching distance between the RTM, DTM and ATM of hand gestures by using the cosine similarity, which is defined as

$$d = \frac{\sum_{l=1}^{k} (s_l \cdot t_l)}{\sqrt{\sum_{l=1}^{k} (s_l)^2} \cdot \sqrt{\sum_{l=1}^{k} (t_l)^2}}, \qquad (30)$$

where l is the lth point of the matching hand gesture template and the test hand gesture template.

Second, we define $\alpha(c, v)$ as the matching distance of RTM, which is the shortest distance. Similarly, $\beta(g, h)$ and $\gamma(e, j)$ are respectively the shortest matching distances of DTM and ATM, where c is the training samples of RTM, v is the test samples of RTM, g is the training samples of DTM, h is the test samples of DTM, e is the training samples of ATM, j is the test samples of ATM. Then, the fusion matching distance d_f is

$$d_f = \alpha(c, v) + \beta(g, h) + \gamma(e, j), \qquad (31)$$

where α, β and γ are the coefficients of the range, speed and angle, respectively. In our experiments we found that we can get highest recognition accuracy when the coefficients equal to one. In the online stage, the k-means algorithm is applied to the tested hand gesture data, and fusion matching distance d_f is calculated. Then, the shortest fusion distance can be obtained

$$R_{shortest\,fusion} = \min(d_{f1}, d_{f2}, d_{f3}, \dots, d_{f\eta}), \qquad (32)$$

where η is the head gesture type.

Finally, we record the hand gesture with the shortest fusion distance as the recognized hand gesture type.

IV. EXPERIMENTAL ANALYSIS AND RESULTS
A. PALTFORM AND RADAR CONFIGURATION
The experimental environment and the radar equipment chosen for data acquisition is shown in FIGURE 11. The millimeter wave radar is Texas Instruments (TI) AWR1642 radar
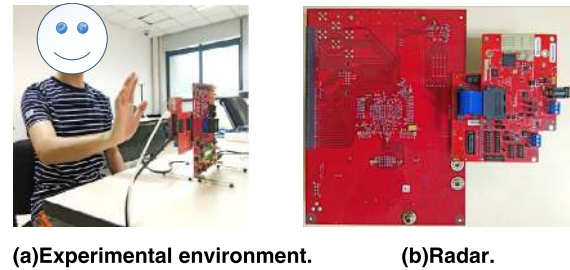

(a)Experimental environment.   (b)Radar.

FIGURE 11. Experimental environment and equipment.

TABLE 1. Radar system parameter configuration.

Radar Configuration	value
Start Frequency (GHz)	77
Frequency Slope (MHz/μs)	105
ADC Samples	64
Sample Rate (ksps)	2000
Ramp End Time (μs)	38
Bandwidth (MHz)	3997
Antenna Configuration	2TX,4RX

(a)PS   (b)PL   (c)LF   (d)RG   (e)P-P   (f)U-D
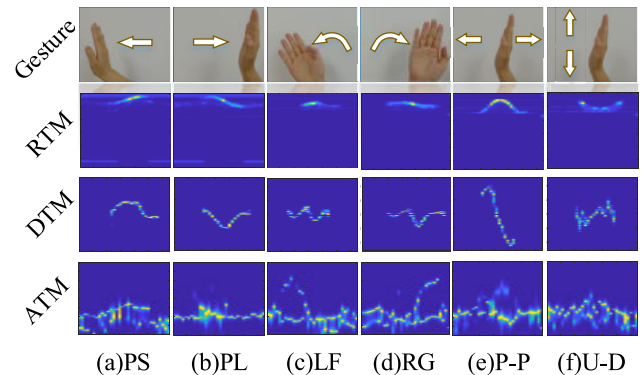
FIGURE 12. Gesture type.

board, and the high-speed data acquisition card is TSW1400. The radar system parameter configuration used in this article is shown in TABLE 1. The starting frequency of the radar is 77 GHz, the bandwidth is 4 GHz, and the radar has 2 transmitting and 4 receiving antennas. Each frame of radar data contains 128 chirps, and the period of each chirp is 38μs. The tester sits in front of the radar with the dynamic hand gesture range from 10 cm to 70 cm. We first collect the radar signal and construct the spectrograms. Then, the continuous hand gestures are effectively segmented, and the FDTW algorithm is proposed to recognize the hand gestures.

B. EXPERIMENTAL DATA
In this article, we design six types of hand gestures, which include push(PS), pull(PL), left(LF), right(RG), push-pull(P-P), up-down(U-D), shown in FIGURE 12. We collect

(a) Detection offset of PS.

(b) Detection offset of PL.

(c) Detection offset of LF.

(d) Detection offset of RG.

(e) Detection offset of P-P.

(f) Detection offset of U-D.

FIGURE 13. The detection offset.

200 sample data for each type of hand gesture, and the total number of hand gesture sample is 1200. In FIGURE 12, we give a completed hand gesture diagrams, an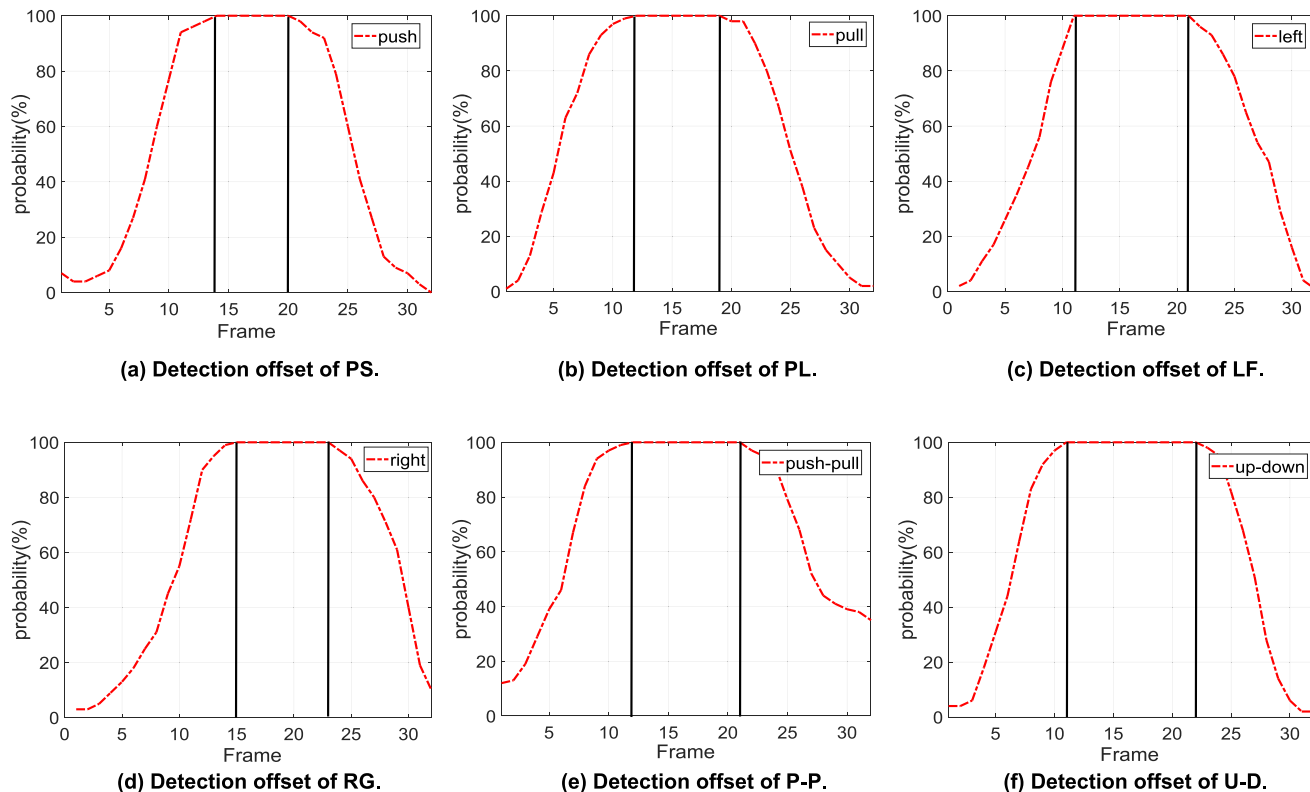d the corresponding three spectrograms (RTM, DTM and ATM) of each hand gesture. It should be noted that the duration time of a completed hand gesture is less than one second, and Figure 12 shows the results of 32 frames (equally 1.28 seconds). We learn from FIGURE 12 that different hand gestures have different RTM, DTM, as well as ATM. FIGURE 12 (a) shows an sample of a push hand gesture, and the range of the hand gesture becomes smaller while speed of the gesture (see FIGURE 12 (b)) first increases and then reduces to zero when the gesture finally stop in front of the radar. It should be noted that in order to show the dynamic and continuous hand gestures, each collected data sample contains multiple hand gestures.

C. ANALYSIS OF HAND GESTURE DEDETION RESULTS

Before investigating the recognition performance of continuous hand gesture, we first analyze the segmentation accuracy of the proposed amplitude-based detection method.

To describe hand gesture segmentation accuracy, the detection range is applied in this article.

1) HAND GESTURE DETECTION RANGE

Through the experimental test, the detection offset distribution map of each hand gesture can be obtained. The period

TABLE 2. Detection range of each hand gesture.

Hand gesture type	Detection range [frame]		
	5% drop	3% drop	1% drop
PS	11~23	12~22	13~21
PL	8~23	9~22	11~20
LF	10~24	11~23	11~22
RG	12~26	13~25	14~24
P-P	8~24	9~23	10~22
U-D	8~24	9~23	10~23

with the highest occurrence probability in the distribution map is defined as the detection range. In the actual commercial system, the accuracy of recognition system is more than 95%, that means there will be an error (drop) rate of 5% [21]. As a result, the detection range built in this article can be reduced to 5% drop.

FIGURE 13 shows the detection offset for the six types of hand gestures. The detection offset indicates that the detection range will increase with the appearance of the error frame, as a result, the hand gesture recognition performance will decrease. We define the error frame as 1 frame unit. TABLE 2 shows the details of the hand gesture detection range based on different performance drop of 5%, 3% and 1%, respectively. It can be seen from TABLE 2 that from
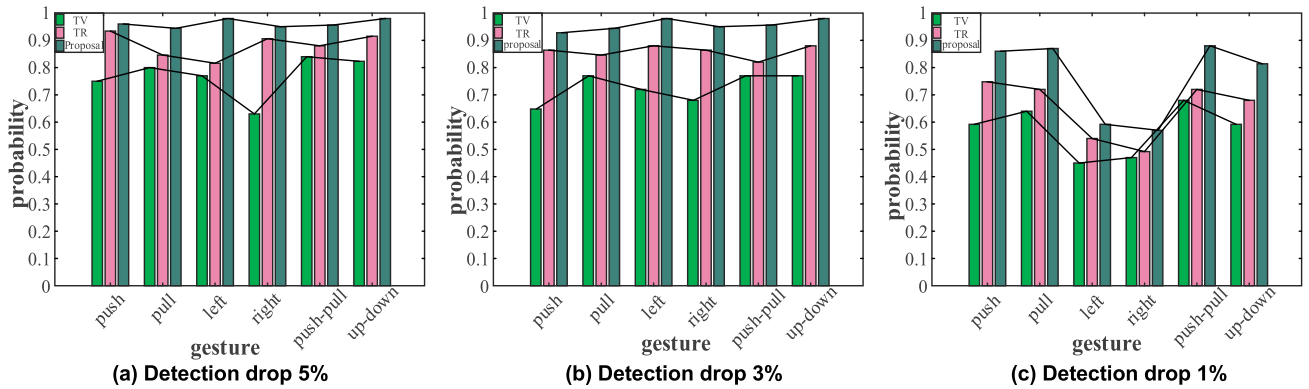
(a) Detection drop 5%   (b) Detection drop 3%   (c) Detection drop 1%

FIGURE 14. Detection probability of each method.

14 to 20 frames of the PS gesture, no error frames appear and the hand gesture detection performance is 100%. When the performance drops by 1%, the hand gesture detection range is expanded to 13 to 21 frames. Similarly, the performance drops by 3%, the gesture detection range is expanded to 12 to 22 frames, and the performance drops by 5%, the hand gesture detection range becomes larger. This change results in the drop of hand gesture detection performance and the increase of gesture detection range.

2) DETECTION PROBABILITY

In this part, we evaluate the accuracy of the segmented hand gesture, and get the detection probability of the hand gestures. FIGURE 14 shows the detection probability of hand gestures after segmentation by using trigger value (TV) [21], trigger ratio (TR) [21] and the amplitude-based detection method proposed in this article under different performance drop. We can see from FIGURE 14 that the detection probability of the proposed amplitude-based detection method is higher than that of TV and TR. With the decrease of the detection offset, the detection range of the hand gesture increases, making the improvement of detection probability of the hand gestures. The average detection probability is shown in TABLE 3.

TABLE 3. Average detection probability.

offset	TV [21]	TR [21]	proposal
1% drop	72.85%	73.75%	83.50%
3% drop	76.20%	86.25%	89.50%
5% drop	88.70%	93.75%	96.17%

We learn from TABLE 3 that the detection probability of the amplitude-based detection method is higher than that of the TV and TR. The average detection probability of the proposed method is 83.50% when the performance drops by 1%, which is about 10% higher than the other two alternatives. When the performance drops by 5%, the range of hand

gesture detection is enlarged, making the detection probability increase to 96.17%, which is 3% higher than the existing alternatives. Therefore, we apply 5% performance drop for subsequent experiments.

D. HAND GESTURE CLUSTERING RESULTS

The *k*-means algorithm is used to cluster the central time-frequency trajectory of the hand gesture spectrogram. Since the duration of a single hand gesture after continuous gesture segmentation is different, the observation duration of the segmented hand gesture is uniformly expanded to the original 32 frames. FIGURE 15 shows the extended segmented spectrogram of RTM, DTM and ATM of PS, and their corresponding central trajectory after clustering. It is observed from FIGURE 15 that the time-frequency trajectories represent the main features of the gesture spectrograms.

E. HAND GESTURE RECOGNITION RESULTS

To prove the effectiveness of the proposed FDTW algorithm, we use several different datasets: RTM, DTM, ATM and multi-feature dataset (contains RTM, DTM and ATM). The K-Nearest Neighbor (KNN) algorithm and the traditional DTW algorithm are taken for comparison. The recognition accuracies are shown in TABLE 4. As can be seen from TABLE 4 that the average recognition accuracy of DTW algorithm is higher than the KNN algorithm. This is attributed to the fact that the KNN algorithm uses the simple Euclidean Distance as the distance, while DTW applies the cosine similarity, which is improved to be more efficient [41]. Moreover, the recognition accuracies of single-parameter datasets for both KNN algorithm and DTW algorithm are always less than 90%. However, the proposed FDTW algorithm makes full use of RTM, DTM and ATM features, making the hand gesture recognition rate (95.83%) higher than that of the existing alternatives. The accuracy of proposed FDTW algorithm is even 1.33% higher than using the DTW algorithm with RTM and DTW, which proves the effectiveness of using the fusion data.
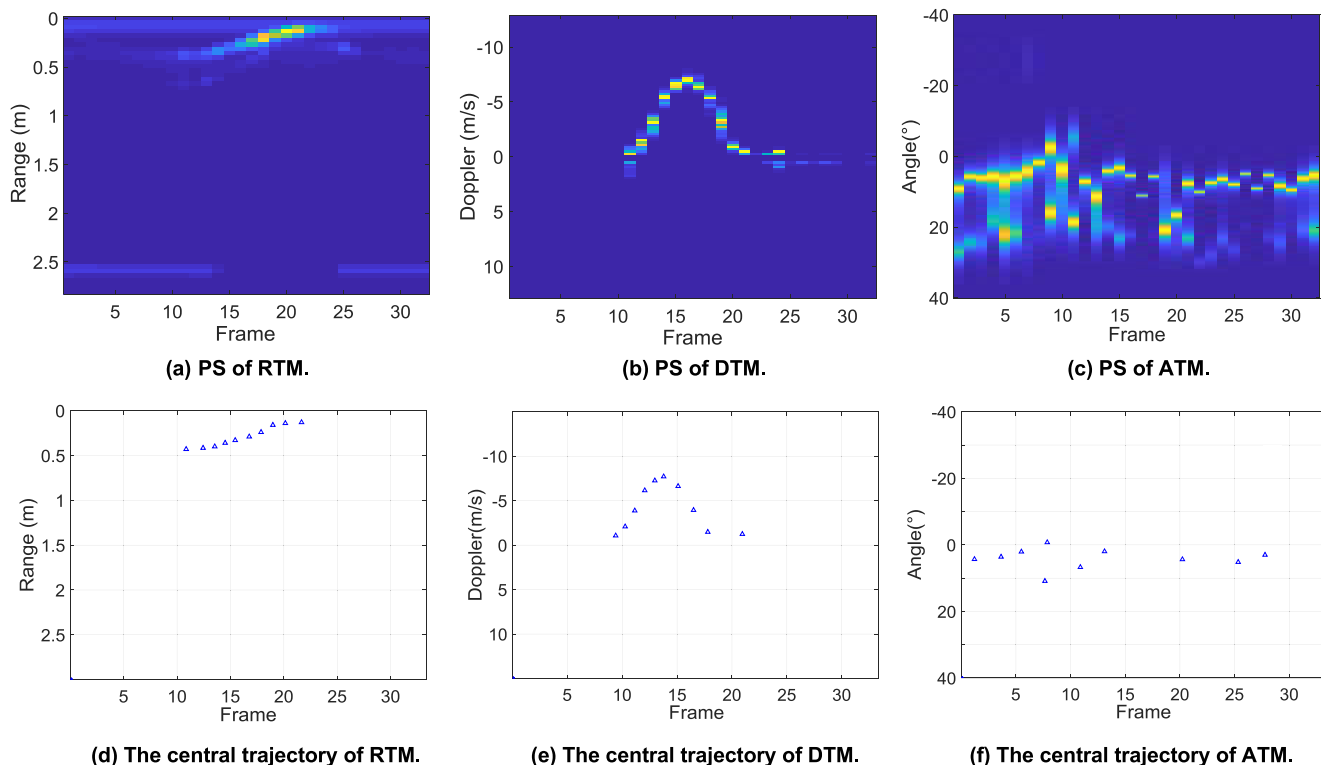
(a) PS of RTM.

(b) PS of DTM.

(c) PS of ATM.

(d) The central trajectory of RTM.

(e) The central trajectory of DTM.

(f) The central trajectory of ATM.

FIGURE 15. The original hand gesture spectrogram and the corresponding central trajectory.

TABLE 4. Accuracy of gesture recognition (%).

Methods and Data	Hand Gesture Classification						
	PS	PL	LF	RG	P-P	U-G	Average
KNN+RTM	57	65	89	88	100	98	82.83
KNN+DTM	100	77	88	85	61	100	85.17
KNN+ATM	78	80	97	93	79	65	82.00
DTW+RTM	92	76	93	77	100	99	89.50
DTW+DTM	94	92	98	87	69	99	89.83
DTW+ATM	89	90	97	91	83	81	88.50
DTW+RTM+DTM	98	92	93	96	93	95	94.50
FDTW	**99**	**92**	**94**	**98**	**95**	**97**	**95.83**

F. TIME COMPLEXITY ANALYSIS

Since the central time-frequency trajectories of RTM, DTM and ATM are clustered, we will show the time efficiency in this subsection. We test the running time of different methods before clustering (BC) and after clustering (AC) respectively.

The running time is shown in TABLE 5. We learn from TABLE 5 that the running time of KNN algorithm under the single-parameter datasets before clustering is 2.12 ms, and the running time after clustering is 1.13 ms. Moreover, the running time proposed FDTW algorithm after clustering

TABLE 5. Actual running time (ms).

Methods		Sample								
		1	2	3	4	5	6	7	8	Average
BC	KNN+RTM/DTM/ATM	2.37	1.92	2.42	2.14	2.11	1.87	1.68	2.42	2.12
	DTW+RTM/DTM/ATM	6.67	6.33	5.81	6.19	5.81	5.78	8.23	6.37	6.44
	DTW+RTM+DTM	8.81	8.12	7.90	9.58	8.34	7.80	8.67	9.49	8.58
	FDTW	**11.03**	**12.50**	**11.34**	**11.25**	**10.24**	**10.09**	**11.72**	**10.83**	**11.13**
AC	KNN+RTM/DTM/ATM	1.33	1.12	1.08	1.15	1.13	1.09	1.05	1.12	1.13
	DTW+RTM/DTM/ATM	2.08	1.43	2.04	3.63	0.90	0.80	1.51	1.08	1.61
	DTW+RTM+DTM	3.90	3.56	3.45	4.29	3.67	3.4	3.83	4.24	3.79
	FDTW	**5.18**	**4.28**	**4.54**	**5.25**	**6.16**	**5.36**	**4.08**	**5.19**	**5.01**

is about 5 ms, which is reduced by more than 50% compared to the FDTW algorithm without time-frequency trajectory clustering.

V. CONCLUSION

In this article, a continuous hand gesture detection and recognition method was proposed. Firstly, we collected the raw data of the radar to obtain the IF signal, and used the IF signal to estimate the RTM, DTM and ATM of hand gestures. Then, we proposed an amplitude-based detection method, which applied the amplitude of the normalized hand gesture spectrogram in conjunction with a threshold. Finally, the k-means algorithm was adopted to cluster the center time-frequency trajectories of RTM, DTM and ATM, and FDTW algorithm was proposed to recognize the hand gestures. Experimental results showed that the segmentation accuracy of the proposed hand gesture detection method could reach 96.17%, and the average recognition accuracy of the proposed FDTW algorithm for six types of hand gestures was improved by more than 5% while saving half of the running time.

REFERENCES

[1] X. Liu, M. Jia, X. Zhang, and W. Lu, "A novel multichannel Internet of Things based on dynamic spectrum sharing in 5G communication," *IEEE Internet Things J.*, vol. 6, no. 4, pp. 5962–5970, Aug. 2019.

[2] 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.

[3] A. Purushothaman and S. Palaniswamy, "Development of smart home using gesture recognition for elderly and disabled," *J. Comput. Theor. Nanosci.*, vol. 17, no. 1, pp. 177–181, Jan. 2020.

[4] X. Liu and X. Zhang, "NOMA-based resource allocation for cluster-based cognitive industrial Internet of Things," *IEEE Trans. Ind. Informat.*, vol. 16, no. 8, pp. 5379–5388, Aug. 2020.

[5] X. Liu, X. Zhai, W. Lu, and C. Wu, "QoS-guarantee resource allocation for multibeam satellite industrial Internet of Things with NOMA," *IEEE Trans. Ind. Informat.*, early access, Nov. 6, 2019, doi: 10.1109/TII.2019.2951728.

[6] J. Su, Z. Sheng, A. X. Liu, Y. Han, and Y. Chen, "A group-based binary splitting algorithm for UHF RFID anti-collision systems," *IEEE Trans. Commun.*, vol. 68, no. 2, pp. 998–1012, Feb. 2020.

[7] Q. Xiao, Y. Zhao, and W. Huan, "Multi-sensor data fusion for sign language recognition based on dynamic Bayesian network and convolutional neural network," *Multimedia Tools Appl.*, vol. 78, no. 11, pp. 15335–15352, Jun. 2019.

[8] J. Su, Y. Chen, Z. Sheng, Z. Huang, and A. X. Liu, "From M-ary query to bit query: A new strategy for efficient large-scale RFID identification," *IEEE Trans. Commun.*, vol. 68, no. 4, pp. 2381–2393, Apr. 2020.

[9] J. Su, Z. Sheng, A. X. Liu, Z. Fu, and Y. Chen, "A time and energy saving-based frame adjustment strategy (TES-FAS) tag identification algorithm for UHF RFID systems," *IEEE Trans. Wireless Commun.*, vol. 19, no. 5, pp. 2974–2986, May 2020.

[10] M. L. Gavrilova, Y. Wang, F. Ahmed, and P. P. Paul, "Kinect sensor gesture and activity recognition: New applications for consumer cognitive systems," *IEEE Consum. Electron. Mag.*, vol. 7, no. 1, pp. 88–94, Jan. 2018.

[11] X. Yang, Z. Liu, W. Nie, W. He, and Q. Pu, "AP optimization for Wi-Fi indoor positioning-based on RSS feature fuzzy mapping and clustering," *IEEE Access*, vol. 8, pp. 153599–153609, 2020, doi: 10.1109/ACCESS.2020.3018147.

[12] B. K. Chakraborty, D. Sarma, M. K. Bhuyan, and K. F. MacDorman, "Review of constraints on vision-based gesture recognition for human–computer interaction," *IET Comput. Vis.*, vol. 12, no. 1, pp. 3–15, Feb. 2018.

[13] G. Li, S. Zhang, F. Fioranelli, and H. Griffiths, "Effect of sparsity-aware time–frequency analysis on dynamic hand gesture classification with radar micro-Doppler signatures," *IET Radar, Sonar Navigat.*, vol. 12, no. 8, pp. 815–820, Aug. 2018.

[14] A. Hyodo, S. Oho, and T. Nagasaku, "Single-chip 77 GHz radar sensor and its automotive applications," *SAE Int. J. Passenger Cars-Electron. Electr. Syst.*, vol. 5, no. 1, pp. 272–279, Apr. 2012.

[15] M. Q. Nguyen, A. Flores-Nigaglioni, and C. Li, "Range-gating technology for millimeter-wave radar remote gesture control in IoT applications," in *IEEE MTT-S Int. Microw. Symp. Dig.*, Chengdu, China, May 2018, pp. 1–4.

[16] E. Hyun and Y. S. Jin, "Doppler-spectrum feature-based human-vehicle classification scheme using machine learning for an FMCW radar sensor," *Sensors*, vol. 20, no. 7, pp. 2001–2015, Apr. 2020.

[17] J. Lien, N. Gillian, M. F. Karagozler, and P. Amihood, "Soli: Ubiquitous gesture sensing with millimeter wave radad," *ACM Trans. Graph.*, vol. 35, no. 4, pp. 1–19, Jul. 2016.

[18] S. Saha, M. Pal, and A. Konar, "Triangular membership function based real-time gesture monitoring system for physical disorder detection," *Comput. Vis. Sci.*, vol. 22, nos. 1–4, pp. 1–14, Dec. 2019.

[19] 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 Sens. Lett.*, vol. 2, no. 3, pp. 1–4, Sep. 2018.

[20] Z. Zhou, Z. J. Gao, and Y. M. Pi, "Dynamic gesture recognition with a terahertz radar based on range profile sequences and Doppler signatures," *Sensors*, vol. 18, no. 2, pp. 10–18, Dec. 2017.

[21] M. Yu, N. Kim, Y. Jung, and S. Lee, "A frame detection method for real-time hand gesture recognition systems using CW-radar," *Sensors*, vol. 20, no. 8, p. 2321, Apr. 2020.

[22] S. Nasri, A. Behrad, and F. Razzazi, "A novel approach for dynamic hand gesture recognition using contour-based similarity images," *Int. J. Comput. Math.*, vol. 92, no. 4, pp. 662–685, Apr. 2015.

[23] K. Tripathi, N. Baranwal, and G. C. Nandi, "Continuous dynamic indian sign language gesture recognition with invariant backgrounds," in *Proc. Int. Conf. Adv. Comput., Commun. Informat. (ICACCI)*, Aug. 2015, pp. 2211–2216.

[24] J.-H. Kim, G.-S. Hong, B.-G. Kim, and D. P. Dogra, "DeepGesture: Deep learning-based gesture recognition scheme using motion sensors," *Displays*, vol. 55, pp. 38–45, Dec. 2018.

[25] D. Wu, L. Pigou, P.-J. Kindermans, N. D.-H. Le, L. Shao, J. Dambre, and J.-M. Odobez, "Deep dynamic neural networks for multimodal gesture segmentation and recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 38, no. 8, pp. 1583–1597, Aug. 2016.

[26] S. Hussain, R. Saxena, X. Han, J. A. Khan, and H. Shin, "Hand gesture recognition using deep learning," in *Proc. Int. SoC Design Conf. (ISOCC)*, Seoul, South Korea, Nov. 2017, pp. 48–49.

[27] Y. Wang, S. Wang, M. Zhou, Q. Jiang, and Z. Tian, "TS-I3D based hand gesture recognition method with radar sensor," *IEEE Access*, vol. 7, pp. 22902–22913, Mar. 2019.

[28] J. Su, R. Xu, S. Yu, B. Wang, and J. Wang, "Redundant rule detection for software-defined networking," *KSII Trans. Internet Inf. Syst.*, vol. 14, no. 6, pp. 2735–2751, 2020.

[29] B. Hartmann and N. Link, "Gesture recognition with inertial sensors and optimized DTW prototypes," in *Proc. IEEE Int. Conf. Syst., Man Cybern.*, Istanbul, Turkey, Oct. 2010, pp. 2102–2109.

[30] M. Pfitscher, D. Welfer, M. A. de Souza Leite Cuadros, and D. F. L. G. Gamarra, "Activity gesture recognition on kinect sensor using convolutional neural networks and fast DTW for the MSRC-12 dataset," in *Proc. Int. Conf. Intell. Syst. Design Appl.* Cham, Switzerland: Springer, Dec. 2018, pp. 230–239.

[31] J. Wu and R. Jafari, "Orientation independent activity/gesture recognition using wearable motion sensors," *IEEE Internet Things J.*, vol. 6, no. 2, pp. 1427–1437, Apr. 2019.

[32] Y.-T. Li and J. P. Wachs, "HEGM: A hierarchical elastic graph matching for hand gesture recognition," *Pattern Recognit.*, vol. 47, no. 1, pp. 80–88, Jan. 2014.

[33] J.-B. Wen, Y.-S. Xiong, and S.-L. Wang, "A novel two-stage weak classifier selection approach for adaptive boosting for cascade face detector," *Neurocomputing*, vol. 116, pp. 122–135, Sep. 2013.

[34] K. M. Sagayam and D. J. Hemanth, "A probabilistic model for state sequence analysis in hidden Markov model for hand gesture recognition," *Comput. Intell.*, vol. 35, no. 1, pp. 59–81, Feb. 2019.

[35] X. L. Guo and T. T. Yang, "Gesture recognition based on HMM-FNN model using a kinect," *J. Multimodal User Interfaces*, vol. 11, no. 1, pp. 1–7, Mar. 2016.

[36] Y. Li, Q. Miao, X. Qi, Z. Ma, and W. Ouyang, "A spatiotemporal attention-based ResC3D model for large-scale gesture recognition," *Mach. Vis. Appl.*, vol. 30, no. 5, pp. 875–888, Jul. 2019.

[37] Z. Zhou, Y. Dai, and W. Li, "Gesture recognition based on global template DTW for Chinese sign language," *J. Intell. Fuzzy Syst.*, vol. 35, no. 2, pp. 1969–1978, Aug. 2018.

[38] I. Assent, M. Wichterich, R. Krieger, H. Kremer, and T. Seidl, "Anticipatory DTW for efficient similarity search in time series databases," *Proc. VLDB Endowment*, vol. 2, no. 1, pp. 826–837, Aug. 2009.

[39] I.-J. Ding and C.-W. Chang, "Feature design scheme for kinect-based DTW human gesture recognition," *Multimedia Tools Appl.*, vol. 75, no. 16, pp. 9669–9684, Aug. 2016.

[40] A. Desoki, J.-I. Takada, and I. Hagiwara, "Efficient array design algorithm for wide-band application of the multiple signal classification algorithm," *Acoust. Sci. Technol.*, vol. 30, no. 3, pp. 187–198, 2009.

[41] A. F. R. Hernandez and N. Y. G. Garcia, "Distributed processing using cosine similarity for mapping big data in Hadoop," *IEEE Latin Amer. Trans.*, vol. 14, no. 6, pp. 2857–2861, Jun. 2016.

YONG WANG (Member, IEEE) received the B.S., M.S., and Ph.D. degrees from the Harbin Institute of Technology, Harbin, China, in 2010, 2012, and 2018, respectively. From January 2014 to June 2015, he was a visiting Ph.D. Student with the Department of Electrical and Computer Engineering, University of Toronto, Canada. He is currently a Lecturer with the Chongqing University of Posts and Telecommunications, Chongqing, China. His research interests include human–computer interaction (HCI), radar signal processing, resource allocation in cooperative networks, deep learning, and WIFI localization.

AIHU REN received the B.S. degree in telecommunication engineering from Shangqiu Normal University, China, in 2018, where he is currently pursuing the M.S. degree. His current research interests include indoor localization, machine learning, and deep neural networks.

MU ZHOU (Senior Member, IEEE) received the Ph.D. degree in communication and information systems from the Harbin Institute of Technology (HIT), China, in 2012. He was a joint-cultivated Ph.D. Student with the University of Pittsburgh (PITT), USA. He was supported by the Chongqing Municipal Program of Top-Notch Young Professionals for Special Support of Eminent Professionals and awarded with the Outstanding Cooperation Project Award from Huawei Technologies Company Ltd. Over the past five years, he was engaged in five national projects, nine provincial and ministerial projects (including two major projects), and seven enterprise projects.

WEN WANG received the B.S. degree from Nanchang Hangkong University, Jiangxi, China, in 2018, where he is currently pursuing the degree. His research interests include deep learning and radar signal processing.

XIAOBO YANG received the B.S. degree from the Chengdu Institute of Meteorology, Chengdu, China, in 2000, and the M.S. degree from the Chongqing University of Posts and Telecommunications, Chongqing, China, in 2008. From January 2016 to July 2016, she was a Visiting Scholar with the University of California at San Diego, San Diego, USA. She is currently a Lecturer with the Chongqing University of Posts and Telecommunications. Her research interests include indoor localization, machine learning, and deep neural networks.

• • •