

Received January 4, 2019, accepted January 24, 2019, date of publication February 4, 2019, date of current version March 4, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2897060

TS-I3D Based Hand Gesture Recognition Method With Radar Sensor

YONG WANG^{ID}, (Member, IEEE), SHASHA WANG^{ID}, MU ZHOU^{ID}, (Senior Member, IEEE), QING JIANG, AND ZENGSHAN TIAN

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

Corresponding author: Shasha Wang (1259657562@qq.com)

This work was supported in part by the National Natural Science Foundation of China under Grant 61771083 and Grant 61704015, in part by the Fundamental and Frontier Research Project of Chongqing under Grant cstc2017jcyjAX0380 and Grant cstc2015jcyjBX0065, and in part by the University Outstanding Achievement Transformation Project of Chongqing under Grant KJZH17117.

ABSTRACT Aiming at the problems of the noise impact on the parametric image of hand gestures, the difficulty of gesture feature extraction, and the inefficient utilization of continuous gesture time sequential information, we propose a time sequential inflated 3 dimensions (TS-I3D) convolutional neural network approach for hand gesture recognition based on frequency modulated continuous wave (FMCW) radar sensor. Specifically, the FMCW radar is used to acquire the hand gesture data, and the range and speed of the gesture in each frame signal are calculated by 2 dimensions fast Fourier transform. Then, the range-Doppler map (RDM) is generated based on the relationship between motion parameters and frequency. The interference in RDM caused by people and the external environment is filtered out and the peak of hand gesture in RDM is further enhanced by wavelet transform. Finally, TS-I3D network is designed to extract the range and speed change information of the continuous gestures. The experimental results show that the average recognition accuracy rate of the hand gestures of the proposed method is 96.17%.

INDEX TERMS FMCW radar, hand gesture recognition, interference filtering, deep learning, LSTM.

I. INTRODUCTION

With the development of advanced signal processing approaches and deep learning [1], [2], human-computer interaction has attracted much attention. Hand gesture recognition (HGR), as one of the most important way of human-computer interaction, gets rid of the limitations of people in traditional control devices, and has been widely used in various fields, i.e., smart home like switching TV channels, turning on the air conditioner or intelligent driving like answering a phone call.

The existing HGR methods mainly contain three stages, namely hand gesture data acquisition, feature extraction, and gesture classification.

Hand gesture data is mainly collected using optics camera [3]–[5] or radar [6]–[9]. Optical-based gesture recognition mainly applies cameras to capture gesture images, and then applies machine-learning methods [10], [11] for feature extraction and recognition. Coelho *et al.* [4] use Kinect to capture RGB and depth images of hand gestures,

The associate editor coordinating the review of this manuscript and approving it for publication was Guan Gui.

and then extract and classify picture features by Convolutional Neural Networks (CNN). Hu *et al.* [5] use HoloLens to collect multi-frame images of dynamic gestures, and design the improved 3D separable convolutional neural network for gesture recognition. However, these optical-based gesture recognition methods are easily affected by light intensity, which largely limit the applications of HGR. On the other hand, the method based on radar signals [6], [7] mainly adopts the radar device to collect the gesture signal, avoiding non-line-of-sight effects such as illumination. As a result, radar-based method of HGR has been concerned and widely applied [8], [9]. The key idea towards the radar-based HGR is to analyze the intermediate frequency (IF) signal to extract the motion parameters of the continuous gesture [6], [12]. Dekker *et al.* [6] use the terahertz radar signal to obtain the data source, and directly calculate a motion parameter map through fast fourier transform (FFT) for each gesture, and finally apply the CNN network for feature extraction. However, Zhou *et al.* [7], Kim and Toomajian [10] directly calculate a range scalar value at each moment to represent the gesture feature information, which makes the feature extraction incomplete, resulting in low accuracy of

gesture recognition. Although Zhang *et al.* [13] achieve about 96% recognition accuracy for gestures, they only use range information for the gesture recognition.¹ To address the problem of insufficient parameter features, two-dimensional Fast Fourier Transform (2D-FFT) [14]–[17] is used to process IF signals to obtain the Range-Doppler Map (RDM) images containing gesture range and speed information. After acquiring RDM, Molchanov *et al.* [14] propose deep learning network to extract its features and recognize the gesture. However, there are noises or other interferences like the arm generated by the radar, which makes the gesture information in the RDM not obvious as well as impair recognition accuracy.

In the stage of gesture feature extraction, CNN is an effective way to extract image features. Karpathy *et al.* [18] propose a CNN structure to extract the features in each frame image of the gesture, and then fuse the features to classify gestures. However, Karpathy *et al.* [18] do not consider the coherence information between frames which makes these gesture frames become irrelevant static images. To overcome the limitations of CNN, Tran *et al.* [19] propose a 3 Dimension Convolutional Neural Network (3D-CNN) in 2014 to extract information of a few consecutive frames. Such a 3D-CNN structure is widely used in various types of behavior recognition. Molchanov *et al.* [20] send multi-frame images that represent gestures to the 3D-CNN network for feature extraction, and then recognize them through the full connection layer. However, since the 3D-CNN network is shallow and its convolution kernel is single, the extracted features are not completed enough to effectively represent the whole hand gesture motion. Towards this end, Wang *et al.* [21] apply the Long Short-Term Memory (LSTM) network [22] to extract time sequential information, which makes up for the lack of 3D-CNN. However, the authors do not preprocess feature extraction to get the features in per frame hand gesture data. As a result, Carreira and Zisserman [23] combine the advantages of GoogLeNet [24] and 3D-CNN to generate a wider and deeper Inflated 3 Dimension (I3D) convolutional neural network. Hara *et al.* [25] extract the completed features to represent behavior and identify them using I3D. Unfortunately, although I3D extracts coherence information, the time sequential features are not discussed in their work.

After the gesture feature extraction, classification methods are used for gesture identification. In general, machine learning algorithms such as markov [26], Support Vector Machine (SVM) [27], [28] and k-Nearest Neighbor (KNN) [29] are adopted for classification. However, these algorithms are complex and they cannot synthesize an end-to-end network structure. Therefore, for gesture recognition, feature extraction and classification are separately carried out, resulting in inefficient applications.

¹It is worth noting that Zhang *et al.* [13] adopt only eight types of gestures with very large discrimination in their experiments. In our work, we consider less distinguishes gestures with more types of gestures.

In order to address the above shortcomings, in this paper, we propose a gesture recognition method based on Frequency Modulated Continuous Wave (FMCW) radar. We collect the signals of the continuous hand gesture by FMCW radar to generate parameter map, and make the gesture target more prominent through interference suppression and image enhancement. More importantly, we design a Time Sequential Inflated 3 Dimension (TS-I3D) convolutional neural network for feature extraction and then add a softmax layer for gesture classification. The main contributions of this paper are as follows:

- Hand gesture data collection and processing: FMCW radar signal is used as the source to collect hand gesture data. The range and speed of the continuous hand gesture are calculated by 2D-FFT [15] algorithm based on the IF signal frequency, and the RDM is generated according to the relationship among range, speed and the frequency of IF signal. Then, the peak interference including the step peak interference in single-frame RDM and the static peak interference caused by the body in multi-frame RDM are filtered. Moreover, wavelet Transform (WT) [30] is further applied for image enhancement (IE) of the gesture peak in RDM. Finally, each gesture is represented by multi-frame RDM. The hand gesture data are repeatedly collected and processed to generate a dataset.
- TS-I3D network: the I3D network is designed to extract RDM features and generate Range-Time (RT) feature sequence and Doppler-Time (DT) feature sequence. Then, two LSTM networks are employed to extract temporal gesture information from RDM. Finally, the extracted RT and DT features are concatenated and classified in softmax layer.
- Experiments verification: we repeatedly collect 10 kinds of gestures and each for 400 times. The dataset is divided into training set and testing set for training and validating the effectiveness of the designed TS-I3D network. The results show that the recognition accuracy of the proposed TS-I3D for each gesture is higher than 95%, and the average recognition accuracy is 96.17%.

The rest of this paper is organized as follows. In Section 2, we calculate the IF signal to obtain RDM of the gesture by 2D-FFT algorithm, and process the RDM by peak interference filtering (PIF) and IE. The designed TS-I3D network is described, and the temporal information extraction of range and speed are also given in Section 3. We present the experiments and discussions in Section 4. Finally, Section 5 gives the conclusions and future work.

II. FMCW RADAR SIGNAL PROCESSING

The mapping of each gesture signal to the final RDM mainly contains the following three stages. In stage I, the intermediate frequency signal of each radar frame is calculated by mixing the transmitted and received signals. In stage II, 2D-FFT is applied to calculate the corresponding RDM for each frame. In stage III, the step peak interference in the

single RDM, the static peak interference caused by the human body in the continuous frames are filtered out, and the gesture information in RDM image is enhanced.

A. IF SIGNAL EXTRACTION

The FMCW radar used in this paper emits a zigzag frequency-modulated continuous wave. The transmitted signal is mixed with the received signal reflected by the hand, and then the high-frequency part is filtered by a low-pass filter. After ADC sampling, an IF signal is generated, and the whole process for IF signal generation is shown in Fig. 1.

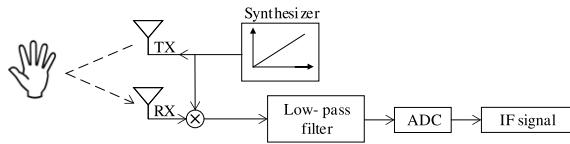


FIGURE 1. IF signal processing.

The transmitted signal used in this paper is a high frequency continuous sawtooth wave whose frequency varies with time. The transmit signal of FMCW [31] is

$$\begin{aligned} S_{TX}(t) &= A_{TX} \cos \left(2\pi f_{ct} t + 2\pi \int_0^t f_T(\tau) d\tau \right) \\ &= A_{TX} \cos \left(2\pi f_{ct} t + 2\pi \times \frac{1}{2} \times \frac{B}{T} t^2 \right), \end{aligned} \quad (1)$$

where t is the time of the transmitted signal, $f_T(\tau) = \frac{B}{T} \times \tau$ is a linear function of the frequency of the transmitted signal, f_c is the carrier frequency, B is the bandwidth, A_{TX} is the amplitude of the transmitted signal, and T is the signal period.

Since the frequency change of the received signal is consistent with the transmitted signal, the final form of the received signal is

$$S_{RX}(t) = A_{RX} \cos \left(2\pi f_{ct} t + \pi \frac{B}{T} (t^2 + t_d^2 - 2t_d t) + \Delta\phi \right), \quad (2)$$

where $t_d = 2 \times \frac{R_0+vt}{c}$ is the time delay of the transmitted signal caused by gesture movement, $\Delta\phi = \frac{4\pi v \times t}{\lambda}$ is the phase shift, v is the radial motion speed of the gesture relative to the radar, R_0 is the range from the gesture to the radar at time $t=0$, c is the speed of light, and A_{RX} is the amplitude of the received signal.

The transmitted signal and the received signal are mixed, and a low-pass filter is used to obtain a low-frequency component in the mixed signal. The IF signal is calculated as follows

$$S_{IF}(t) = \frac{1}{2} A_{IF} \cos \left\{ 2\pi \cdot \frac{B}{T} \left(\frac{1}{2} t_d^2 - t_d t \right) - \Delta\phi \right\}. \quad (3)$$

Since t_d is very small in the actual measurement, t_d^2 is negligible, and A_{IF} is the amplitude of the IF signal. Then, the frequency of IF signal can be approximated as

$$f_{IF} = \frac{B}{T} \times 2 \frac{R_0+vt}{c} + \frac{\Delta\phi}{2\pi t}. \quad (4)$$

B. 2D-FFT BASED RDM CALCULATION

As can be seen from Eq. (4), the range from hand to the radar is proportional to the frequency of the IF signal. So, the range d could be calculated by the following equation

$$d = \frac{c \times f_{IF} \times T}{2B}. \quad (5)$$

Each frame of the radar transmitted in this paper contains 128 pulses. The IF signal for each pulse is sampled at 64 points with f_s as the sampling frequency, and then the Range-FFT [32] can be performed. The thus generated 64 spectra represent the frequency components of the IF signal. It is known from Eq. (5) that the range is proportional to the frequency of the IF signal. In this paper, the maximum range is

$$d_{max} = \frac{c \times f_s \times T}{2B}, \quad (6)$$

where $c = 3 \times 10^8 \text{ m/s}$, $T = 3.8 \times 10^{-5} \text{ s}$, $B = 4 \times 10^9 \text{ Hz}$, $f_s = 105.202 \text{ MHz/us}$, the range difference between every two sampling points is $\Delta d = \frac{d_{max}}{64}$.

When the transmitted signal encounters a stationary hand, there is no frequency shift in the individual spectra of the continuous pulses. It is noted that moved hand gestures cause frequency shifts in the various spectra of successive pulses due to the Doppler effect. Since the Range-FFT has extracted the respective frequency components of the IF signal in the continuous pulse, it is only necessary to perform Doppler-FFT [32] on the signals composed of the respective frequency components, and the doppler frequency shift f_{FFT} of the continuous pulse is obtained. It is noted that speed is equivalent to Doppler speed in the rest of the paper. Since the wavelength is λ , the speed of the gesture is obtained from the frequency of the IF signal as follows

$$v = \frac{\lambda \Delta\phi}{4\pi t} = \frac{\lambda f_{FFT}}{4T}. \quad (7)$$

Since each frame signal contains 128 pulses, the signal composed of each frequency component contains 128 sampling points. Therefore, the maximum speed calculated by the Doppler-FFT is $v_{max} = \frac{\lambda \times f_s}{4}$. Since the speed has a direction, the minimum speed is $v_{min} = -\frac{\lambda \times f_s}{4}$, and the speed difference between every two sampling points is $\Delta v = \frac{v_{max}-v_{min}}{128}$.

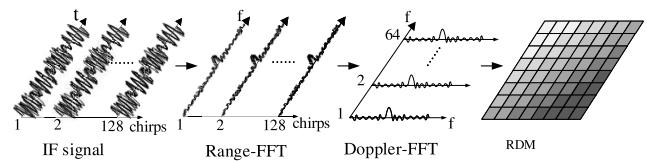


FIGURE 2. IF signal generation RDM process.

RDM of the gesture is calculated using 2D-FFT, and the whole calculation process is shown in Fig. 2. Firstly, the IF signal of each pulse is subjected to Range-FFT to calculate the corresponding spectrum, and then the components of the same frequency of each pulse are extracted for Doppler-FFT.

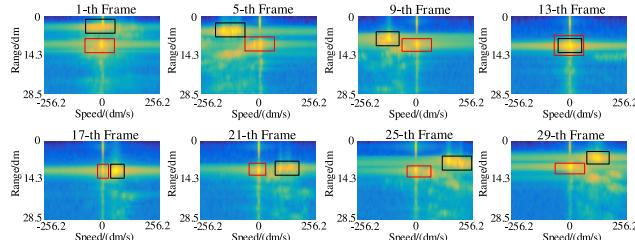


FIGURE 3. RDM based on 2D-FFT.

It is noted that the vertical axis of the RDM indicates the speed of hand gesture, and the horizontal axis indicates the range.

Thus, each frame of the radar signal generates an RDM map with a frame size of 64×128 . It is noted that the vertical axis represents the range from 0 to d_{\max} , and the horizontal axis represents the speed from v_{\min} to v_{\max} . Each pixel in the RDM represents the precision of Δd and Δv . In this paper $\Delta d = 0.44$ m and $\Delta v = 4$ dm/s. The value in the RDM corresponds to the amplitude of the 2D-FFT, and a larger amplitude indicates that the position corresponding to the range and speed has a target. In this paper, each frame signal lasts 40 ms and 32 frames lasts for 1.28 s. Gestures generally lasts from 0.5 s to 1 s, and 32 frames contains a gesture. Thus, we use 32 frames of RDM to represent a gesture. From the 32 RDM frames of pull-push gesture, one frame is selected every 4 frames, and eight frames of RDM are obtained as an example, shown in Fig. 3. Since the gesture is pull in the first 16 frames, we can see from the 1st, 5th, 9th and 13th frames in Fig. 3 that the range of the gesture is increasing while the speed increases first and then slows down. From the 17th, 21st, 25th and 29th frames, the range of the gesture is decreasing and the speed slows down similarly in the last 16 frames. Therefore, Fig. 3 confirms that the first 16 frames represent the pull action, and the last 16 frames represent the push action. It is noted that when different people make gestures, they may have different starting positions and gestures of different magnitudes. However, the movement parameters of each gesture are restrained in the same way. For example, the gestures of push-pull and push-right are complex and similar. The range and speed in the first 16 frames are the same, but the range of push-pull becomes smaller, while the range of push-left stays almost the same in the last 16 frames. Therefore, the RDM generated by the different gestures are different.

As can be seen from Fig. 3 that in multi-frame RDM, a step peak appears in all range rows with speed of 0, which is due to interference from antenna equipment. In addition, there are two peaks in red and black boxes in most RDM maps, as shown in Fig. 3. It is observed that both the ordinates and abscissa of the peak value marked in red color remain unchanged. From the above RDM calculation process, it can be seen that the RDM ordinate represents range and the abscissa represents speed. Thus, the red box corresponding to range remains unchanged which is the arms and the body

since the people sit in front of the radar. While the peak marked by black box corresponds to the gesture target.

C. PEAK INTERFERENCE FILTERING IN RDM

Since the antennas of the radar used in this paper are microstrip array antennas, there may be interference due to the closed distance between the transmitted antennas and the received antennas. The column of RDM indicates the speed of 0 in each frame RDM has a step peak. Since the human body appears in the radar monitoring range, several lines in each frame RDM with certain range may have peaks (see Fig. 3). To address the above challenges, we propose the following methods for PIF.

1) STEP PEAK FILTERING

Since the peak of the gesture is attributed to the human arms and the body, the peak in the RDM is a coherence rather than a step. The interference in RDM caused by transmitted antennas and received antennas is step, as shown in Fig. 4(a). Based on this characteristic, the step peak caused by interference can be filtered out.

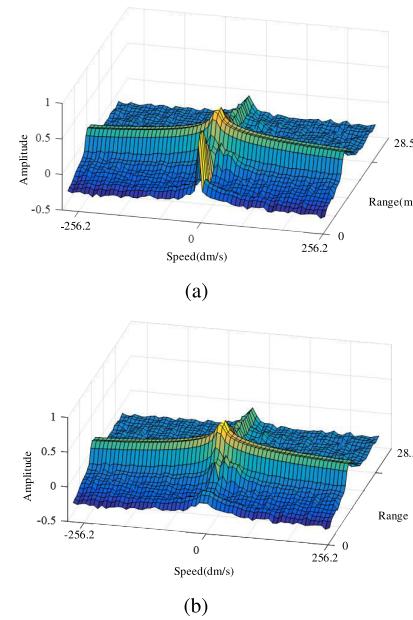


FIGURE 4. Step peak filtering comparison. (a) Before step peak filtering. (b) After step peak filtering.

The gradient matrix is obtained by calculating the gradient of the growth of each speed unit in the speed direction of per RDM frame. Then the maximum value of each column of gradient matrix is calculated, and the gradient row vector representing the change of the amplitude of each column of RDM is generated. The gradient threshold [33] of each frame is chosen as

$$th = \mu + K\delta, \quad (8)$$

where μ is the mean of the gradient row, δ is the variance, K is a constant coefficient, and we set $K = 3$ [33].

We calculate the gradient of each column in RDM, and select the column whose gradient value exceeds the threshold value. Then, the first order exponential smoothing [34] is applied to replace the values of the selected columns. The step peak filtering result is shown in Fig. 4(b). As we can see from Fig. 4(b), the step peaks caused by transmitted antennas and received antennas are filtered out.

2) STATIC PEAK INTERFERENCE SUPPRESSION

Since the human body is stationary in front of the radar, the peak value of RDM in 32 frames does not change in the ordinate of range, and always appears at the position of 0 in the abscissa of speed.

This paper determines whether the corresponding range of a gesture in 32 frame RDM is disturbed. A gesture corresponding to IF signal is processed by 2D-FFT [15], and then the signal intensity with speed of 0 at different range in each frame is extracted according to Eq. (5) and Eq. (7). And the range-time matrix with horizontal axis representing the sequence of signal frames and vertical axis representing different range is generated. Eq. (8) calculates the threshold of each row of the matrix, and then determines whether the corresponding 32 frame RDM represents static peak interference for each row of the range. As shown in Fig. 5, the range-time matrix comparison is made the presence and the absence of static peak interference.

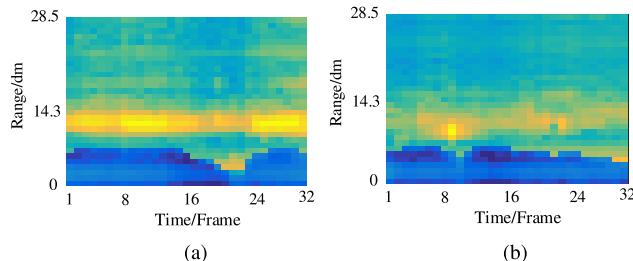


FIGURE 5. Comparison of range-time matrix with static peak interference. (a) Range-time matrix with static interference. (b) Range-time matrix without static peak interference.

Since the peak value appears in the columns representing the minimum speed, this paper deals with the middle n columns. As the peak value of gesture and body may coincide, RDM with two or more peaks per frame is screened. The peak value of the middle n column is replaced by first-order exponential smoothing [34] for each row from the forward and reverse speed directions, respectively. Eq. (7) shows that each column in RDM represents the interval of 0.4 m/s, so $n = 10$ is chosen in this paper. The effect of static peak interference suppression is shown in Fig. 6. After interference suppression, the peak marked by red box of each frame (see Fig. 3) is weakened. The peak marked by black box representing gesture is clearer, thus avoiding the network judging other peaks as gesture targets. In frame 13, the peak of gesture motion coincides with the peak of human body. However, it appears only one peak, and only one step peak is filtered out.

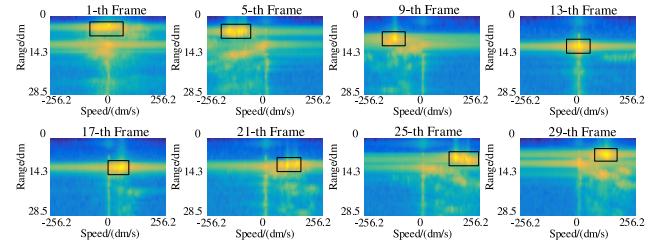


FIGURE 6. RDM after static peak interference suppression.

D. IMAGE ENHANCEMENT

After filtering out the obvious peak interference in RDM, there are still some small targets in the RDM image such as external noise and arm, which make the RDM image blurred and distorted. Wavelet transform is a time-frequency analysis method with the characteristics of multi-resolution analysis. By decomposing the image into different frequency subbands, the approximation and details at different scales are highlighted to enhance the image. Therefore, this paper uses wavelet decomposition and reconstruction to enhance the RDM image, while suppressing the impact of small targets such as noise.

The RDM image is decomposed by three-level multi-scale wavelet, and the sym4 mother wavelet was selected as the basis function. The scaling coefficients and wavelet coefficients are obtained as follows. and the scale coefficients and wavelet coefficients are obtained as

$$\begin{cases} c_{k;n,m} = \sum_{l,j} a_{l-2n,j-2m} c_{k+1;l,j} \\ d_{k;n,m}^i = \sum_{l,j} b_{l-2n,j-2m}^i c_{k+1;l,j}, i = 1, 2, 3 \end{cases} \quad (9)$$

where $c_{k;n,m}$ is the scale coefficients of RDM decomposition, $d_{k;n,m}^1$, $d_{k;n,m}^2$ and $d_{k;n,m}^3$ represent the the wavelet coefficients of horizontal, vertical and diagonal directions, respectively.

For scale coefficients $c_{k;n,m}$, it represents the low-frequency decomposition coefficients of the image that corresponds to the contours of the RDM image. As a result, we have not processed these coefficients. Part of the wavelet coefficients in $d_{k;n,m}^1$, $d_{k;n,m}^2$ and $d_{k;n,m}^3$ that correspond to the background details of noise and arm are smaller, while the wavelet coefficients related to the peak of gesture are larger. Therefore, the wavelet coefficients are shrunk by a threshold [35], and the larger coefficients are retained to reduce the influence of noise in the RDM image. Finally, the RDM is reconstructed using the processed wavelet coefficients. The results are shown in Fig. 7. It is shown in Fig. 7 that IE makes the peak value of gesture more obvious since the noise and arm interference in RDM are suppressed.

III. TS-I3D NETWORK MODEL CONSTRUCTION

After the 32-frame RDM is mapped from a continuous hand gesture, it is input to a TS-I3D network for feature extraction and classification. Firstly, the 32-frame RDM is input to the designed I3D network in this paper for RDM feature extraction. Then, according to the characteristics of the data

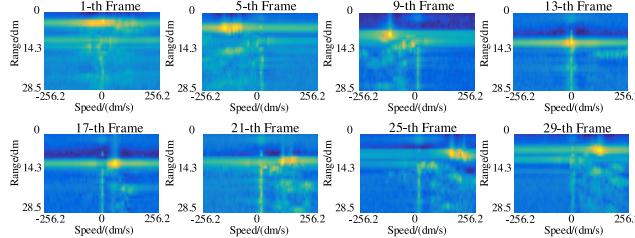


FIGURE 7. RDM after PIF and IE.

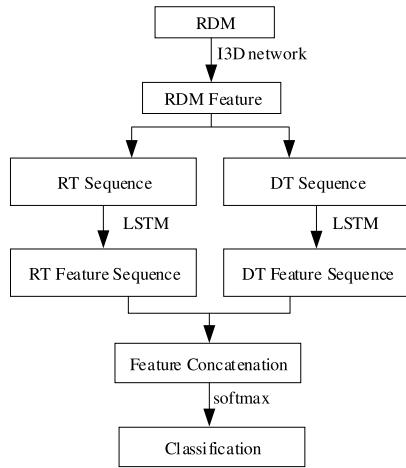


FIGURE 8. TS-I3D network model.

structure, the extracted RDM features are reorganized into RT and DT feature sequences, and the generated feature sequences are input into the LSTM network for time sequential feature extraction. Finally, the extracted time sequential features are concatenated and classified by the softmax layer. The flow chart is shown in Fig. 8.

A. RDM FEATURE EXTRACTION

The I3D network is based on GoogLeNet, inheriting the Inception module [24], and extracting the depth feature of the network by deepening the depth and width of the network model. According to the data structure and the complexity of the image, we simplify the I3D network [23]. The specific I3D network structure is shown in Fig. 9. It is noted that the format of the convolution kernel is NCHW [19].

In the classical CNN network with VGG-16 architecture [36], there are five convolution segments, two or three convolution layers in each segment, and a maximum pooling layer is connected at the end of each segment to reduce the size of the feature map. Motivated by this idea, the shallow features in the RDM are extracted using two common convolution pooling layers in our I3D network. Then, we add 3 Inception modules and in each Inception. In each Inception, we change the convolution kernel size of the second and third convolution extraction features from the usual $1 \times 1 \times 1$ to $3 \times 3 \times 1$ and $3 \times 1 \times 3$. Since the continuous gesture behavior used in this paper, the convolution depth of the

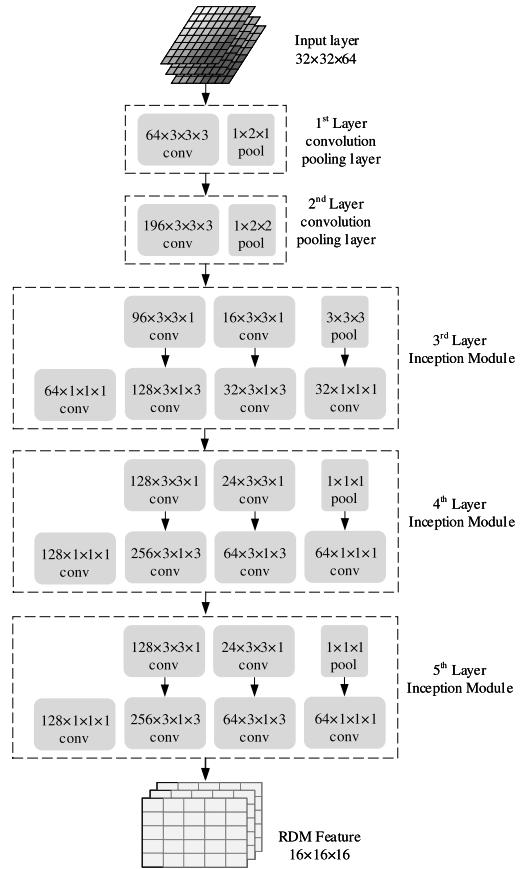


FIGURE 9. I3D network structure.

convolution kernel on RDM sequence is 3 instead of 1 for a static image. For each instantaneous gesture image, the sizes of 1×3 and 3×1 are used to extract the different features of a single frame image. By extracting RDM features from different dimensions and integrating them, the features are enriched. The resultant RDM features are of size $16 \times 16 \times 16$ after passing through the various layers of the I3D network.

B. RT AND DT TIME SEQUENTIAL FEATURE EXTRACTION

In this paper, the feature extraction of RDM is carried out by I3D network, and the 3D convolution mode in the network retains the RDM data structure. The 3D convolution method in the I3D network can only process up to 3 consecutive RDMs at the same time. Therefore, the generated RDM feature only includes the consistency of several frames of RDM before and after the gesture action, and it does not represent the time sequential feature of a complete gesture. In order to extract the time sequential information in the RDM features, the RDM features are reorganized and then input to the LSTM network, and the process is given in Fig. 10.

As shown in Fig. 10, since RDM features retain the structural characteristics of RDM, the horizontal axis of the RDM feature represents the range, the vertical axis represents the speed, and the order of the features represents the time. The rows that indicates the same range in each RDM feature are

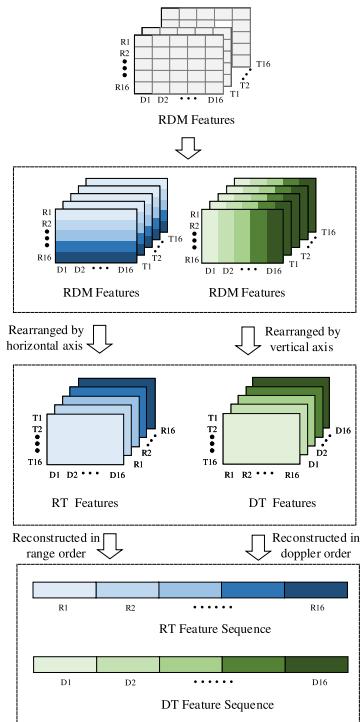


FIGURE 10. RDM feature reorganization.

extracted and recombined to generate a new feature map. In this way, the RDM features are rearranged by range. Similarly, the columns indicating speed in each RDM feature are processed. Finally, a two-dimensional RT feature sequence is reconstructed in order of range. Therefore, the RT sequence represents the range change information of a gesture. Similarly, DT feature sequences representing gesture speed change information are generated according to the same speed column in RDM feature.

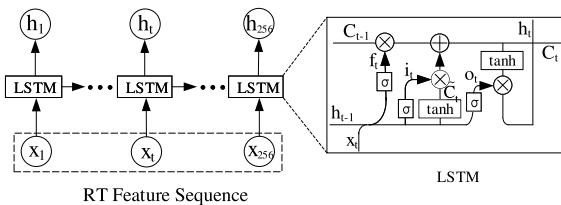


FIGURE 11. LSTM extracts time sequential information for RT.

After the RDM features are reorganized, the generated RT feature sequence dimension is 256×8192 . It is noted that 256 is the step size, and the 8192 at each step contains the speed information at the range unit and the convolutional feature information through different dimensions of the I3D network. In order to extract the gesture change information of the range and the speed in the sequence, RT and DT feature sequence are respectively input into the LSTM network, and Fig. 11 gives RT as an example. It is observed from Fig. 11 that the feature information of each step determines part of C_{t-1} from the unit state by the forgetting gate f_t , and then

determines the information \tilde{C}_t stored in the unit state by the input gate i_t , as follows

$$\begin{cases} f_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f) \\ i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i) \\ \tilde{C}_t = \tanh(W_C \cdot [h_{t-1}, x_t] + b_C) \\ C_t = f_t \cdot C_{t-1} + i_t \cdot \tilde{C}_t, \end{cases} \quad (10)$$

where $\sigma(\cdot)$ is the sigmoid function, $\sigma(x) = \frac{1}{1+e^{-x}}$, W_f , W_i and W_C are the weights in the LSTM unit, and b_f , b_i and b_C are the corresponding biases.

The results of the input gate are passed through the output gate to obtain the hidden layer state h_t and the output information o_t , using the following equation

$$\begin{cases} o_t = \sigma(W_o \cdot [h_{t-1}, x_t] + b_o) \\ h_t = o_t \cdot \tanh(C_t). \end{cases} \quad (11)$$

C. FEATURE CONCATENATION AND CLASSIFICATION

After the RT sequential feature and the DT sequential feature are respectively generated using the LSTM network, the feature dimensions of both the RT and DT sequential feature are 512×1 . The two time sequential features are concatenated in sequential to the feature of 1024×1 , which contains change information of various gestures in range and speed. Then the features that are highly abstracted after multiple convolutions are normalized. Finally, the feature vectors generated by the fully connected layer are input into the following normalized exponential function [37], [38]

$$\text{softmax}(z) = \frac{\exp(\theta_i^T z_i)}{\sum_{j=1}^k \exp(\theta_j^T z_j)}, \quad (12)$$

where i is the i -th gesture and k is the gesture type. In this paper, $k=10$, z_i is the i -th element of the feature vector, and θ_i is the weight corresponding to z_i .

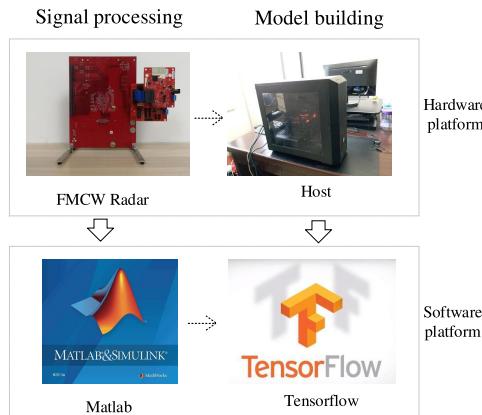
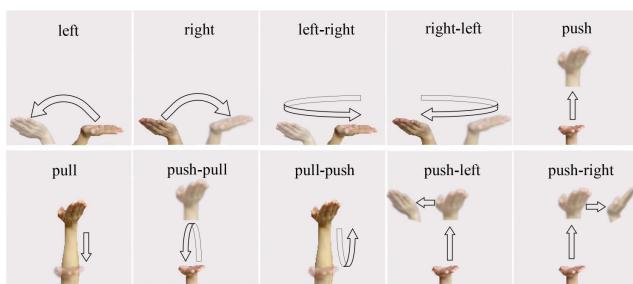
IV. EXPERIMENTAL RESULTS AND ANALYSIS

A. EXPERIMENTAL PLATFORM

The radar platform of this paper is the Texas Instruments AWR1642 single-chip FMCW radar sensor. The radar is equipped with 2 rounds of 4 microstrip antennas. The slope of the transmitted sawtooth signal is 105MHz/us and the bandwidth is 4GHz. The host used is configured with Intel-6700K processor and NVIDIA-GTX1080 graphics card. We select a spacious indoor environment for data collection. The process flow is shown in Fig. 12. The testers sit in front of the radar with their hands slipping making gestures. We collect hand gesture by FMCW radar, and process the radar signals by Matlab to generate the gesture RDM dataset. Then, the TS-I3D is built under the Tensorflow framework in the host for training and testing.

B. EXPERIMENTAL DATA

We design 10 types of hand gestures, which are left(L), right(R), left-right(L-R), right-left(R-L), push(PS), pull(PL),

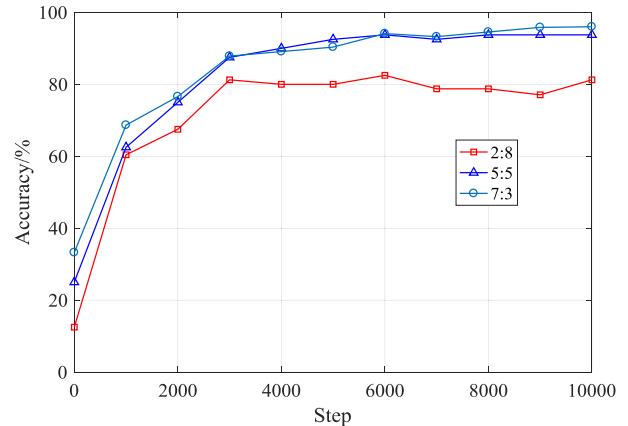
**FIGURE 12.** Experiment platform.**FIGURE 13.** Gesture type.

push-pull(PS-PL), pull-push(PL-PS), push-left(PS-L) and push-right(PS-R), shown in Fig. 13. Each type of gesture contains 400 data and each data contains a complete gesture. The total sample size of the 10 types hand gestures is 4000. In the experimental environment, the radar transmits 32 frames at a time, and during the signal transmission, a tester makes gestures. To enhance the robustness of the proposed method, gestures of different testers are collected to generate dataset. Then dataset is divided according to a certain proportion. In this paper, a simple cross-validation is used. In order to prevent gestures of the same tester from gathering together, the gestures of each type of hand are randomly disrupted. In order to verify the effectiveness of interference suppression and feature extraction of the TS-I3D network, we input the training dataset into the designed network for training, and then uses the test dataset for gesture recognition.

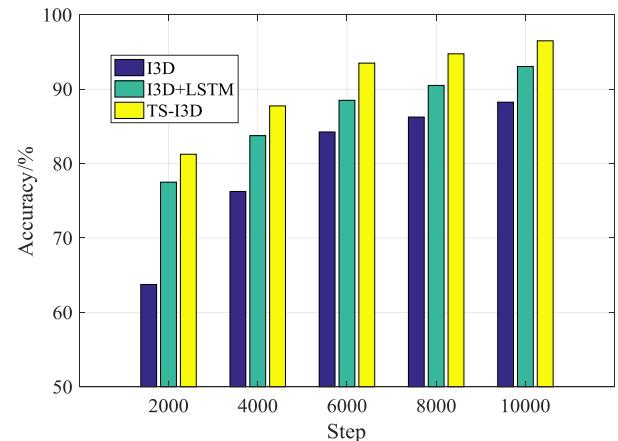
C. NETWORK STRUCTURE AND TRAINING PROGRESS

1) IMPACT OF THE TRAINING TO TESTING DATASET RATIO

In this experiment, the performance of the proposed method is analyzed under different ratios of training to testing dataset. The ratios are set as 2:8, 5:5, and 7:3. The results of accuracy varying with the number of steps are given in Fig. 14. We can see that when the ratio is 2:8, the generalization ability of the model trained with the training dataset is poor, while the data volume of the test dataset is large. As a result of the poor fitting ability of the small training dataset, the TS-I3D network cannot classify the gesture in testing dataset that is very different from the training dataset, so the recognition

**FIGURE 14.** Accuracy of different dataset proportions.

result is poor. At this time, the workload of data acquisition needs to be greatly increased to illustrate the problem. By increasing ratio to 5:5 and 7:3, the training set already contains a large number of training data, which makes the I3D network generalization ability of the training office higher and the recognition accuracy of the test dataset higher. In the following analysis, we carried out experiments at a ratio of 7:3.

**FIGURE 15.** Accuracy of each network structure.

2) SUBSTRUCTURE IMPACT OF TS-I3D NETWORK

In order to analyze the rationality of the TS-I3D network structure, the full connection layer is added to the I3D network for directly classification. In addition, in order to verify effectiveness of LSTM for the temporal information extraction of the range and speed changes of gestures, the RDM features extracted from I3D network are directly input into LSTM network in time order without reconstruction process. Thus, the I3D+LSTM network is formed compared with our TS-I3D network, and the results are given in Fig. 15. In the first 4000 steps of iterative training, the weights of the three networks are randomly initialized and updated, and the accuracy of I3D, I3D+LSTM and TS-I3D are 76%, 84% and 88%, respectively. From 4000 to 8000 steps, the updated gradient

of the weights becomes smaller, the updating speed is slow, and the accuracy reaches 86%, 91% and 95% respectively. When the training is finished, the final accuracy of I3D network is about 88%, which shows that I3D network can extract the features of RDM. However, compared with the other two networks, the accuracy of I3D network is the worst one because the extracted features are incomplete. Since LSTM network is added to the I3D network to obtain the variation characteristics of the mixed feature information of range and speed, the accuracy of I3D+LSTM works better and reaches about 93%. In TS-I3D, two LSTM are respectively employed to extract the range change information and the speed change information in RDM, which extract more detailed gesture features. As a result, the final accuracy rate is higher than 96%.

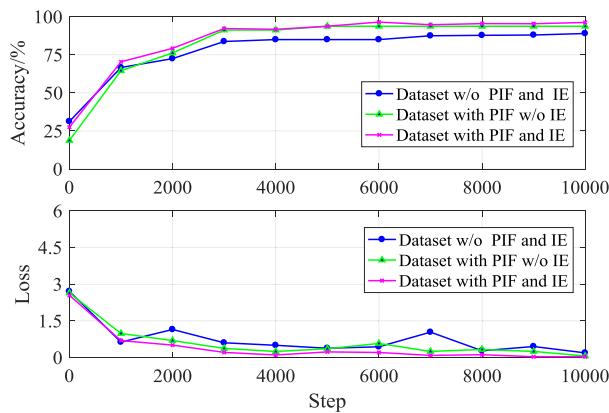


FIGURE 16. Accuracy and loss function of various data.

3) IMPACT OF PIF AND IE

In order to compare the effect of PIF and IE, we input the unprocessed gesture data, the data processed with PIF and the data processed with both PIF and IE into the network for testing. Fig. 16 shows the recognition accuracy and the loss function value comparison of the TS-I3D under different training iteration steps. It is noted that the symbol ‘w/o’ indicates ‘without’ in Fig. 16. As can be seen from the Fig. 16, the accuracy and loss function are largely improved by PIF and IE. Moreover, it is also verified that IE further improve both the accuracy and loss function of the TS-I3D network, and the convergence performance as well as the recognition accuracy are enhanced.

4) IMPACT OF LEARNING RATE

Learning rate is an important hyperparameter for deep learning, and it can adjust the weights updating gradient of network by loss function. We analyze the impact of different learning rates of TS-I3D on accuracy, and the results are shown in Fig. 17. It is noted that when the step is 0, the weights of TS-I3D network are randomly initialized and have not been trained iteratively yet, and the accuracy is very low. We can see that learning rate controls the speed of iteration training. The smaller the learning rate is, the smaller the gradient of weight updating is. Therefore, the network training takes

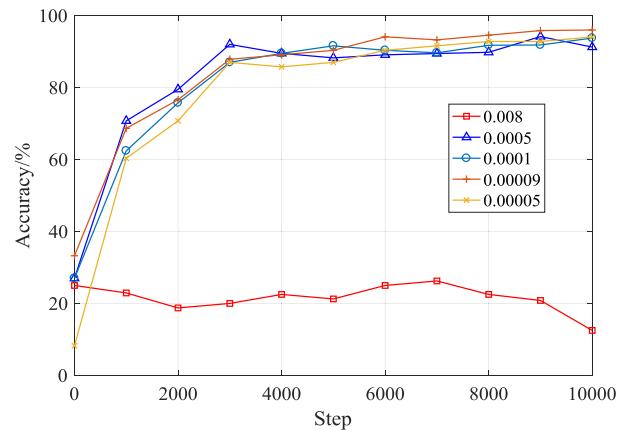


FIGURE 17. Accuracy of model under different learning rates.

more time and it converges more slowly. If the learning rate is too large, such as 0.008 and 0.0005, the TS-I3D network is not converge or falls into local optimum. However, when the learning rate is too low (0.00005), the weight update of TS-I3D is too slow and consumes much time. When the learning rate is set to 0.00009, the recognition effect is the best and we use this learning rate for the rest of the experiments.

D. ACCURACY COMPARISON WITH STATES OF THE ARTS

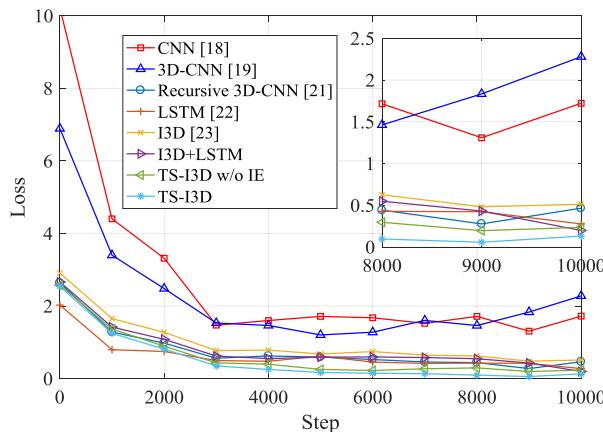
To further verify the effectiveness of the proposed method, five behavioral recognition deep learning networks are selected for comparison: CNN [18], 3D-CNN [19], recursive 3D-CNN [21], LSTM [22], I3D network [23], I3D+LSTM and TS-I3D without (w/o) IE. CNN uses five layers of network, and the size of kernel is 3×3 . The number of CNN kernel in each layer is 64, 128, 256, 512 and 512. Based on CNN, 3D-CNN modifies kernel to $3 \times 3 \times 3$. Recursive 3D-CNN inputs the final RDM features into the LSTM network. I3D network uses the RDM feature extraction part directly. For fair comparison, the parameters of all the methods in this subsection are set as the proposed ones. Specifically, the batch size of each input samples is set to 16, the iteration number for training is 10000, and the initial learning rate is set to 0.00009.

1) LOSS FUNCTION COMPARISON

The loss function measures the difference between the predicted value and the real one, which indicates the degree of the fitting data. At the same time, it also indicates complexity of the proposed model for the current data. Comparing the loss function of this method with other ones, we can verify the efficiency of the proposed model structure and the corresponding convergence behavior. Therefore, we verify the loss function via inputting the training dataset into the model under different steps. Fig. 18 shows the loss function of each network during the training process. Since CNN and 3D-CNN only use 2D convolution and 3D convolution, respectively, to extract gesture feature and build a very deep network, their convergence rates are worse than the other methods. Based

TABLE 1. Recognition accuracy comparison for each gesture(%).

Methods	L	R	L-R	R-L	PS	PL	PS-PL	PL-PS	PS-L	PS-R	Ave.
CNN [18]	88.57	89.56	87.17	87.44	78.57	81.25	74.29	85.37	78.57	76.92	82.77
3D-CNN [19]	90.00	91.54	94.12	91.17	85.71	85.36	85.71	90.63	82.14	84.38	88.07
Recursive 3D-CNN [21]	93.10	92.30	93.42	92.14	94.83	93.33	94.83	92.68	92.85	91.42	93.09
LSTM [22]	92.85	90.63	91.67	88.57	87.09	87.80	94.29	90.00	90.00	90.63	90.35
I3D [23]	92.03	93.75	93.33	91.42	89.28	87.18	85.30	91.67	84.46	85.29	89.37
I3D+LSTM	92.83	92.36	93.94	92.47	94.55	92.85	93.75	91.84	93.10	92.78	93.05
TS-I3D w/o IE	93.75	95.83	95.67	94.12	95.04	96.55	95.12	94.29	93.54	93.33	94.72
TS-I3D	96.67	97.43	95.67	95.23	97.45	96.55	95.12	95.50	95.67	96.12	96.17

**FIGURE 18.** Training process comparison of each networks.

on 3D-CNN, I3D combines 3D convolution of different sizes to extract the coherence information between frames in the hand gesture data. Therefore, the convergence effect of I3D is better than CNN and 3D-CNN. However, it ignores the time sequential information in the continuous hand gestures. To this end, in recursive 3D-CNN, a LSTM is used to enhance the model on the basis of 3D convolution, and it extracts the time sequential information in gestures with a simple model. However, the convolution kernel of the recursive 3D-CNN network is single, and the extracted features are insufficient. Besides, we compare the convergence of I3D+LSTM network and TS-I3D without IE, and the convergence effect of I3D+LSTM network is similar to that of recursive 3D-CNN. Moreover, the convergence rate of the dataset without IE is obviously worse than the dataset with IE. This is attributed to the fact that IE highlights the similarities of the same kind of gesture data, resulting in better gesture feature extraction. In TS-I3D, a improved I3D network is proposed to extract the preliminary feature firstly, and LSTM is used to extract the time sequential features in gestures. As a result, the loss function of TS-I3D network is the lowest and the convergence effect is the best among the considered methods in Fig. 18.

2) RECOGNITION ACCURACY COMPARISON

After the model training process is completed, the test dataset is input to each network for gesture recognition. Table 1 shows the recognition accuracy of each gesture over our self-built dataset. It is worth noting that Ave. in Table 1 stands for average. It can be seen from the

Table 1 that compared with CNN, 3D-CNN can extract better features in terms of hand gesture recognition (accuracy rate increase by 6.4%). Under the premise of using three-dimensional convolution, the I3D network uses a different size convolution method, and it has a wider and deeper network than 3D-CNN. As a result, the recognition accuracy of I3D is about 1.5% higher than 3D-CNN. Moreover, since recursive 3D-CNN adds LSTM to extract time sequential features after 3D convolutional network, its recognition accuracy is higher than 3D-CNN for some complex hand gestures, such as PS-L and PS-R. For I3D+LSTM, its accuracy for each type of gesture recognition is higher than 91%, and the final average accuracy is 93.05%. Since the 3D-CNN network is shallow, we use a variety of convolution kernels to enrich the gesture features. Therefore, the final recognition accuracy is almost the same as that of recursive 3D-CNN. To verify the performance of image enhancement, we input data without IE into TS-I3D network for recognition. Since IE suppresses noises and highlights the peak of the gesture, TS-I3D can extract features of gesture motion more easily in RDM. The average recognition accuracy of TS-I3D is increased by 1.5% compared to the TS-I3D without IE. Finally, by applying an I3D network to extract the depth feature and using LSTM to extract the time sequential information, the recognition accuracy of TS-I3D of each gesture is higher than other networks, and its average recognition accuracy is 96.17% (improved by 3.3% compared to the existing recursive 3D-CNN).

V. CONCLUSION AND FUTURE WORK

In this paper, a complete HGR process named TS-I3D was presented. Using FMCW radar for gesture data acquisition, we firstly calculated IF signal of FMCW radar and obtained RDM using 2D-FFT algorithm. Then, we filtered out the interference based on the characteristics of peak interference, and WT was further applied for RDM image enhancement. We proposed a TS-I3D network to extract the depth features in RDM, and reconstructed the RDM features according to range and speed by the data structure characteristics. LSTM was applied to extract the temporal features of gestures from different angles. We validated TS-I3D on self-built gesture dataset. The results showed that the average recognition accuracy was 96.17%, which was increased by 3.3% compared with the state of the arts.

In the future, we try to extract the spatial-temporal features of dynamic hand gestures by a single network via adding 3D operation into LSTM network. Moreover, we are designing

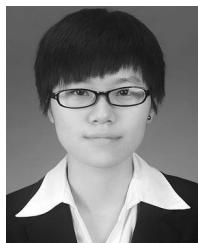
a real-time HGR system by replanting the deep learning network into smart terminals.

REFERENCES

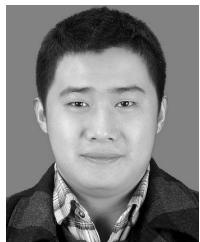
- [1] Y. Lin, Y. Li, X. Yin, and Z. Dou, "Multisensor fault diagnosis modeling based on the evidence theory," *IEEE Trans. Rel.*, vol. 67, no. 2, pp. 513–521, Jun. 2018.
- [2] G. Gui, H. Huang, Y. Song, and H. Sari, "Deep learning for an effective nonorthogonal multiple access scheme," *IEEE Trans. Veh. Technol.*, vol. 67, no. 9, pp. 8440–8450, Jun. 2018.
- [3] Y. Li, X. Wang, W. Liu, and B. Feng, "Deep attention network for joint hand gesture localization and recognition using static RGB-D images," *Inf. Sci.*, vol. 441, pp. 66–78, May 2018.
- [4] Y. L. Coelho, J. M. Salomao, and H. R. Kulitz, "Intelligent hand posture recognition system integrated to process control," *IEEE Latin Amer. Trans.*, vol. 15, no. 6, pp. 1144–1153, Jun. 2017.
- [5] Z. Hu, Y. Hu, J. Liu, B. Wu, D. Han, and T. Kurfess, "3D separable convolutional neural network for dynamic hand gesture recognition," *Neurocomputing*, vol. 318, pp. 151–161, Nov. 2018.
- [6] B. Dekker, S. Jacobs, A. S. Kossen, M. C. Kruithof, A. G. Huizing, and M. Geurts, "Gesture recognition with a low power FMCW radar and a deep convolutional neural network," in *Proc. Radar Conf. (EURAD)*, Oct. 2017, pp. 163–166.
- [7] Z. Zhou, Z. Cao, and Y. Pi, "Dynamic gesture recognition with a terahertz radar based on range profile sequences and Doppler signatures," *Sensors*, vol. 18, no. 1, p. 10, Dec. 2017.
- [8] S.-J. Ryu, J.-S. Suh, S.-H. Baek, S. Hong, and J. H. Kim, "Feature-based hand gesture recognition using an FMCW radar and its temporal feature analysis," *IEEE Sensors J.*, vol. 18, no. 18, pp. 7593–7602, Sep. 2018.
- [9] Z. Zhang, Z. Tian, Z. Mu, and Y. Liu, "Application of FMCW radar for dynamic continuous hand gesture recognition," in *Proc. MOBIMEDIA*, Beijing, China, Sep. 2018, pp. 298–303.
- [10] Y. Kim and B. Toomajian, "Hand gesture recognition using micro-Doppler signatures with convolutional neural network," *IEEE Access*, no. 4, pp. 7125–7130, 2016.
- [11] Y. Li, X. Cheng, and G. Gui, "Co-robust-ADMM-net: Joint ADMM framework and DNN for robust sparse composite regularization," *IEEE Access*, vol. 6, pp. 47943–47952, 2018.
- [12] H. Huang, J. Yang, H. Huang, Y. Song, and G. Gui, "Deep learning for super-resolution channel estimation and DOA estimation based massive MIMO system," *IEEE Trans. Veh. Tech.*, vol. 67, no. 9, pp. 8549–8560, Jun. 2018.
- [13] 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, Feb. 2018.
- [14] P. Molchanov, S. Gupta, K. Kim, and K. Pulli, "Multi-sensor system for driver's hand-gesture recognition," in *Proc. 11th IEEE Int. Conf. Workshops Autom. Face Gesture Recognit. (FG)*, May 2015, pp. 1–8.
- [15] S. Jardak, S. Ahmed, and M.-S. Alouini, "Low complexity moving target parameter estimation for MIMO radar using 2D-FFT," *IEEE Trans. Signal Process.*, vol. 65, no. 18, pp. 4745–4755, Sep. 2017.
- [16] P. Molchanov, S. Gupta, K. Kim, and K. Pulli, "Short-range FMCW monopulse radar for hand-gesture sensing," in *Proc. IEEE Radar Conf.*, May 2015, pp. 1491–1496.
- [17] S.-T. Huang and C.-H. Tseng, "Hand-gesture sensing Doppler radar with metamaterial-based leaky-wave antennas," in *Proc. ICMIM IEEE MTT-S Int. Conf.*, Mar. 2017, pp. 49–52.
- [18] A. Karpathy, G. Toderici, T. Leung, R. Sukthankar, and S. Shetty, "Large-scale video classification with convolutional neural networks," in *Proc. CVPR*, 2014, pp. 1725–1732.
- [19] D. Tran, L. Bourdev, R. Fergus, L. Torresani, and M. Paluri, "C3D: Generic features for video analysis," *CoRR*, vol. 2, no. 7, p. 8, 2014.
- [20] P. Molchanov, S. Gupta, K. Kim, and J. Kautz, "Hand gesture recognition with 3D convolutional neural networks," in *Proc. CVPR*, 2015, pp. 1–7.
- [21] X. Wang, L. Gao, J. Song, and H. Shen, "Beyond frame-level CNN: Saliency-aware 3-D CNN with LSTM for video action recognition," *IEEE Signal Process. Lett.*, vol. 24, no. 4, pp. 510–514, Apr. 2017.
- [22] D. Tran, L. Bourdev, R. Fergus, L. Torresani, and M. Paluri, "Learning spatiotemporal features with 3D convolutional networks," in *Proc. ICCV*, 2015, pp. 4489–4497.
- [23] J. Carreira and A. Zisserman, "Quo vadis, action recognition? A new model and the kinetics dataset," in *Proc. CVPR*, 2017, pp. 4724–4733.
- [24] C. Szegedy et al., "Going deeper with convolutions," in *Proc. CVPR*, 2015, pp. 1–9.
- [25] K. Hara, H. Kataoka, and Y. Satoh, "Learning spatio-temporal features with 3D residual networks for action recognition," in *Proc. ICCV*, vol. 2, no. 3, 2017, p. 4.
- [26] W. Wang, A. X. Liu, M. Shahzad, K. Ling, and S. Lu, "Device-free human activity recognition using commercial WiFi devices," *IEEE J. Sel. Area Commun.*, vol. 35, no. 5, pp. 1118–1131, Mar. 2017.
- [27] H. Wu, Q. Huang, D. Wang, and L. Gao, "A CNN–SVM combined model for pattern recognition of knee motion using mechanomyography signals," *J. Electromogr. Kinesiol.*, vol. 42, pp. 136–142, Oct. 2018.
- [28] X.-X. Niu and C. Y. Suen, "A novel hybrid CNN–SVM classifier for recognizing handwritten digits," *Pattern Recognit.*, vol. 45, no. 4, pp. 1318–1325, Apr. 2012.
- [29] A.-J. Gallego, A. Pertusa, and P. Gil, "Automatic ship classification from optical aerial images with convolutional neural networks," *Remote Sens.*, vol. 10, no. 4, p. 511, Mar. 2018.
- [30] I. Daubechies, "The wavelet transform, time-frequency localization and signal analysis," *IEEE Trans. Inf. Theory*, vol. 36, no. 5, pp. 961–1005, Sep. 1990.
- [31] V. Winkler, "Range Doppler detection for automotive FMCW radars," in *Proc. Radar Conf. Eur.*, 2007, pp. 166–169.
- [32] E. Hyun, Y.-S. Jin, and J.-H. Lee, "A pedestrian detection scheme using a coherent phase difference method based on 2D range-Doppler FMCW radar," *Sensors*, vol. 16, no. 1, p. 124, Jan. 2016.
- [33] J. Li, Y. Li, Y. Li, and Z. Qian, "Downhole microseismic signal denoising via empirical wavelet transform and adaptive thresholding," *J. Geophys. Eng.*, vol. 15, no. 6, p. 2469, Dec. 2018.
- [34] D. S. Moore, G. P. McCabe, and B. A. Craig, *Introduction to the Practice of Statistics*: W/Student CD San Francisco, CA, USA: Freeman, 2012.
- [35] S. Saravani, R. Shad, and M. Ghaemi, "Iterative adaptive despeckling SAR image using anisotropic diffusion filter and Bayesian estimation denoising in wavelet domain," *Multimedia Tools Appl.*, vol. 77, no. 23, pp. 31469–31486, Jun. 2018.
- [36] K. Simonyan and A. Zisserman. (2014). "Very deep convolutional networks for large-scale image recognition." [Online]. Available: <https://arxiv.org/abs/1409.1556>
- [37] Y. Yu and F. Liu, "Dense connectivity based two-stream deep feature fusion framework for aerial scene classification," *Remote Sens.*, vol. 10, no. 7, p. 1158, Jul. 2018.
- [38] F. Zhu, X. Kong, H. Fu, and Q. Tian, "A novel two-stream saliency image fusion CNN architecture for person re-identification," *Multimedia Syst.*, vol. 24, no. 5, pp. 569–582, Oct. 2018.



YONG WANG (M'18) 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 2014 to 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 resource allocation and signal processing in cooperative networks, deep learning, radar signal processing, and Wi-Fi localization.



SHASHA WANG received the B.S. degree from the Chongqing University of Posts and Telecommunications, Chongqing, China, in 2016, where she is currently a graduate student. Her research interests include deep learning and radar signal processing.



MU ZHOU was born in 1984. He received the Ph.D. degree from the Harbin Institute of Technology, in 2012. He is currently a Professor with the Chongqing University of Posts and Telecommunications. His current research interests include wireless positioning and navigation technology, signal reconnaissance and detection technology, convex optimization, and depth learning theory.



ZENGSHAN TIAN received the Ph.D. degree from the University of Electronic Science and Technology of China, in 2002. He is currently a Full Professor with the Chongqing University of Posts and Telecommunications. His current research interests include personal communication, precise localization and attitude measure, and data fusion.

• • •



QING JIANG is currently a Professor and the Director of the Communication and Information Foundation Teaching Department, Communication College, Chongqing University of Posts and Telecommunications, China. Her current research interest includes broadband network technologies.