Two Dimensional Parameters Based Hand Gesture Recognition Algorithm for FMCW Radar Systems

Yong Wang(✉), Zedong Zhao, Mu Zhou, and Jinjun Wu

School of Communications and Information Engineering, Chongqing University of Posts and Telecommunications, Chongqing 400065, China
yongwang@cqupt.edu.cn

Abstract. In recent years, hand gesture recognition has increasingly become important in the field of human-computer interaction. This paper proposes a two-dimension parameter based hand gesture recognition method using frequency modulated continuous wave (FMCW) radar. Specifically, we analyze the time domain of the radar signal and estimate the radial distance and angle parameters of hand gestures, and then construct the parameter dataset. The dataset is fed into an improved convolutional neural network to extract features. Finally, the extracted features are fused and then classified by the full connection layer. Experimental results show that the recognition accuracy of the proposed approach is significantly higher than that of the single-parameter ones.

Keywords: Hand gesture recognition · FMCW radar · Convolutional neural network

1 Introduction

With the development of human-computer interaction, as one of the most important parts, hand gesture recognition influences in various fields of our life, such as home entertainment and intelligent drive. In the hand gesture recognition technology, the data sources are mainly divided into optical camera and wireless equipments. The former one uses cameras to collect the dataset, constructs a model based on the changes of the images and the motion trajectory of the hand gestures. Then, the features are extracted using machine learning methods such as neural network [1], k-Nearest Neighbor (KNN) [2] and Support Vector Machine (SVM) [3] for recognition. The latter one mainly adopts wireless equipment to collect the hand gesture signals, and then the frequency domain is analyzed through signal processing. The motion parameters are extracted and identified through clustering [4] or dynamic time regulation [5] or hidden Markov model [6]. The signal sources of the above-mentioned methods mainly include radar, ultra-wide Band (UWB) and wireless channel state information.

Motived by the above analysis, this paper proposes a Frequency Modulation Continuous Wave (FMCW) radar [7] based hand gesture recognition method. The hand gestures of the radial range and angle are calculated, and they are mapped to namely the range-time map (R-TM) and point of angle-time map (A-TM). Then, we design the

M. Jia et al. (Eds.): WiSATS 2019, LNICST 280, pp. 226–234, 2019.
https://doi.org/10.1007/978-3-030-19153-5_23

convolutional neural network to extract the features of R-TM and A-TM. The extracted features are fused and then classified using the classifier function. Finally, by dividing the dataset into training and testing ones, we use training dataset to train the confused convolutional neural network (CNN), and then use the testing dataset for hand gesture classification.

2 Signal Model of FMCW Radar

2.1 IF Signal Model

FMCW radar transmits high-frequency continuous signal, and the frequency of the transmitted triangle signal changes linearly with time. To obtain the intermediate frequency signal of the FMCW radar, the transmitting and receiving signal are input into the mixer. The high-frequency part is filtered by a low-pass filter. The intermediate frequency signal is finally obtained by sampling. The FMCW radar prototype is shown in Fig. 1.

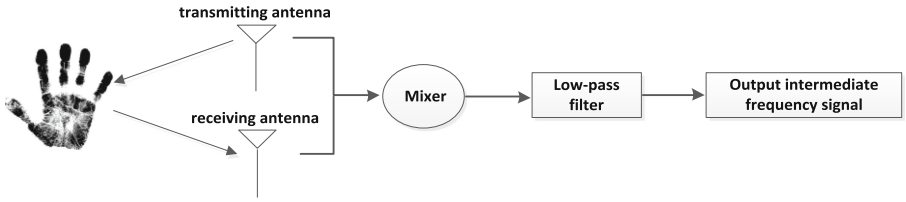

Fig. 1. FMCW radar prototype.

Both transmitting and receiving signals of radar are sawtooth wave, and there exists a fixed delay Δt_{delay} in the receiving signals. The concrete form is shown in Fig. 2.

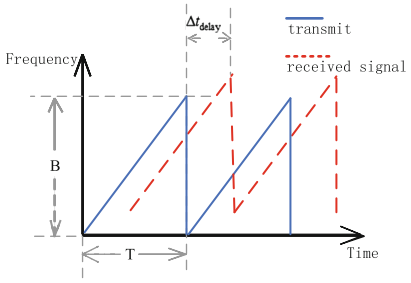

Fig. 2. Correlation curve between transmitted and received signals.

where T is the pulse width of the sawtooth signal, B is the bandwidth of the radar signal.

According to [7], the emission signal of FMCW radar can be expressed as

$$s_T(t) = A_T \cos 2\pi \left[f_c t + \int_0^t f_T(\tau) d\tau \right] \tag{1}$$

where f_c is the central frequency of the carrier, $f_T(\tau)$ denotes the frequency of the transmitted signal in a period of time with length T and T is the pulse width of the sawtooth signal, A_T denotes the amplitude of the transmitting signal.

Due to the influence of flight delay of radar echo signal and doppler frequency shift of the hand gesture, the frequency of radar echo signal is

$$f_R(t) = s(t - \Delta t_{\text{delay}}) + \Delta f_{doppler} \tag{2}$$

where Δt_{delay} is a flight delay between the sending signal and the receiving echo signal, $\Delta f_{\text{doppler}}$ denotes the Doppler shift. The echo signal can be obtained by substituting Eqs. (2) into (1):

$$s_R(t) = A_R \cos 2\pi \left[f_c(t - \Delta t_{\text{delay}}) + \int_0^t f_R(\tau) d\tau \right] \tag{3}$$

where A_R is the amplitude of an echo signal. By mixing $s_R(t)$ and $s_T(t)$, we get the intermediate frequency signal $s_{IF}(t)$ using a low frequency filter:

$$s_{IF}(t) = f_{LPF}\{s_T(t) s_R(t)\} = \frac{1}{2} A_T A_R \cos\ \varphi \tag{4}$$

where φ is the phase.

2.2 Range and Angle Dataset Establishment

When the gesture is not moving, the intermediate frequency (IF) signal should be a sinusoidal signal with a constant frequency. Otherwise, the frequency of the IF signals change with the range between the hand gesture and the radar. In this paper, the intermediate frequency of each frame contains 128 pulses. Because Δt_{delay} is very small in actual measurement, so:

$$\frac{B}{T} = \frac{f_{IF}}{\Delta t_{\text{delay}}} \tag{5}$$

where B is the bandwidth of the radar signal, f_{IF} is frequency point of IF signal.

The corresponding relation between range estimation R and frequency point f_{IF} of IF signal is obtained

$$R = \frac{cT}{2B} f_{IF} \tag{6}$$

The distance-doppler map was obtained by 2D-FFT analysis based on one frame. Therefore, the range can be obtained according to the relationship between the range estimation and the frequency point of the intermediate frequency signal.

Assuming that K hand gesture exists in front of the FMCW radar, the IF signal of the radar is expressed as

$$s_{IF}(t) = \sum_{k=1}^{K} A^{(k)} e^{j2\pi \left[f_c \Delta t_{\text{delay}}^{(k)} + \left(f_{IF}^{(k)} - \Delta f_{\text{doppler}}^{(k)} \right) t \right]} \tag{7}$$

where k represents the target corresponding to the k-th distance unit in the FMCW radar range.

In this paper, the radar has $N_T = 2$ transmitting antennas and $N_R = 4$ receiving antennas. There are 8 virtual receiving antennas. Considering that the radar signals are affected by noise, the signal model is

$$s(m,t) = s_{IF}(m,t) + n(m,t) \tag{8}$$

where $m = 1, 2, \ldots, 8$ is different antenna arrays, $s_{IF}(m,t)$ and $n(m,t)$ represents the signal component and noise component of route m. According to Eqs. (7) and (8), discrete signal $s(m,l)$ is obtained after sampling $s(m,t)$ with the sampling rate F_s, and it can be expressed as

$$s(m,l) = \sum_{k=1}^{K} A^{(k)} e^{j2\pi \left[f_c \cdot \Delta t_{delay}^{(k)} - \frac{l}{F_s} \Delta f_{doppler}^{(k)} + f_{IF}^{(k)} \right]} + n(l) \tag{9}$$

where $l = 0, 1, 2, \ldots, L-1$ satisfies the relations $L{=}T \cdot F_s$. Then, we can construct the signal vector matrix S as

$$S = \begin{Bmatrix} s(1,1) & \cdots & s(1,l) & \cdots & s(1,L) \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ s(m,1) & \cdots & s(m,l) & \cdots & s(m,L) \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ s(N_T N_R, 1) & \cdots & s(N_T N_R, l) & \cdots & s(N_T N_R, L) \end{Bmatrix} \tag{10}$$

The corresponding angle is then obtained by searching the spectral peak of the spatial spectral function.

3 Proposed Fusion Neural Network Based Gesture Recognition

3.1 Range and Angle Feature Extraction

The work of this chapter is to extract feature values by using improved neural network. In R-TM and A-TM, the aim is to extract the continuous position change information in

the sequence according to the dimension information of the sequence. According to the network structure of literature [8] and VGG-16 [9], the network is setting as 5 convolution pooling layers. Since the R-TM and A-TM of gesture movement have great differences, the network is a single-parameter network improved by VGG-16-Net. The entire network consists of 5 convolution pooling layers and 2 full connection layers. In the first and second convolutional pooling layers, the input was convoluted twice and pooled once, and the latter three convoluted and pooled once. After the last pooling of the full connection layer, the last input softmax [10] layer was classified, as shown in Fig. 3.

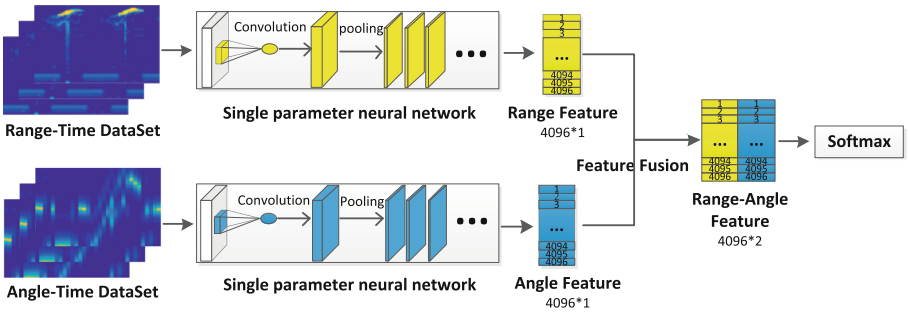

Fig. 3. R-AM schematic.

By using convolution and pooling for R-TM and A-TM, the continuous frames of feature graph are obtained. Then, the range and angle feature vectors are obtained through the full connection layer of the two layers. In this model, there are 5 convolution and pooling modules. The size of the eigenvalue output to the full connection layer through the single-parameter network is 4096×1, and the size of the fused eigenvalue is 4096×2.

3.2 Training Model

After extracting the features of the hand gesture, we add the full connection layer to map the gestures to the specimen marker space. As the final fusion feature is generated by different networks, normalization is carried out before the full connection layer. Finally, the eigenvector $\mathbf{z}$ input into the normalized exponential function for classification, as:

$$\text{softmax}(\mathbf{z}) = \frac{\exp(\theta_i^{\mathrm{T}} z_i)}{\sum_{j=1}^{k} \exp(\theta_j^{\mathrm{T}} z_j)} \tag{11}$$

where i represents the gesture of class i, k represents the type of hand gestures, in this paper, $k = 6$ represents the i-th element of feature vector $\mathbf{z}$, and θ_i represents the corresponding weight of z_i.

In the training, the initial learning rate is setting to a fixed value 0.009. With the iterative tuning of the model, the error is gradually reduced. When the parameter is close to the optimal value, too much update could make the parameter jitter near the optimal value, and too little update will slow down the learning rate. Therefore, in this paper, the exponential decay method is adopted to select a larger learning rate in the initial stage of training, so that the loss value of the network converges to a smaller value more quickly, and the loss value of the network model gradually becomes stable with the exponential reduction of the learning rate.

4 Experimental Results and Analysis

4.1 Experimental Platform

In this paper, the radar platform is a single chip FMCW sensor of Texas instruments AWR1642, which is equipped with two transmitting antennas and four receiving antennas. The slope of the FMCW sawtooth wave signal is 105 MHz/us, and the bandwidth is 4 GHz. In the experiment, hand gesture data from the radar sensor is collected and transmitted to PC, and signal processing is carried out using Matlab software. Then we use Tensorflow deep learning framework for training on the server configured with Intel-6700K processor and NVIDIA-GTX1080 graphics card.

4.2 Experimental Data

Because there are few samples in the gesture dataset of radar signals, the self-built gesture signal data set is verified in the study of establishing deep learning network for feature extraction. This paper designs six types of gestures, including scroll left, scroll right, push forward, pull backward, scroll left-right and push-pull. Each type of gesture was repeated 200 times, with a total of 1200 gesture radar data. In this paper, the platform is placed in a relatively empty indoor environment. The gesture testers repeat the dynamic gestures continuously when the radar transmitted signals to ensure sufficient experimental data collection.

In this paper, each data acquisition contains 32 frames, and each frame contains 128 sawtooth pulses. The parameter graphs are obtained by calculating the range and angle. After data processing, datasets are divided into training sets and test sets, which input into neural network for training and testing.

4.3 Experimental Results

4.3.1 Network Training

In order to compare the performance of the proposed network, this paper carries out model training under different initial learning rates. Figure 4 shows the accuracy rate of different initial learning rates. According to the test results, the network does not converge when the initial learning rate is 0.3. When the initial learning rate is 0.03 and 0.006, each update range is too large to obtain the global optimal solution. When the initial learning rates are 0.001 and 0.0009, the network weight update is too slow.

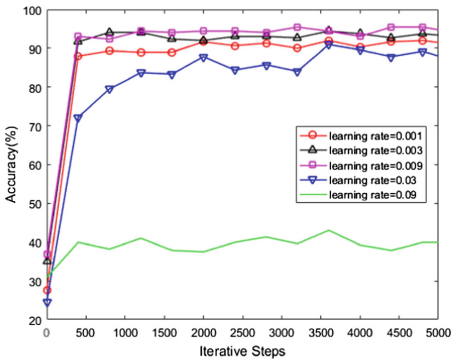

Fig. 4. Accuracy of R-AM under different initial learning rates.

When the learning rate was 0.009, the recognition effect was the best, and the accuracy rate reaches to 95%.

4.3.2 Analysis of Gesture Recognition Results

The accuracy curves of convolutional neural network training reflect the variation of network model accuracy with iterative steps. In this paper, the R-TM and A-TM training and test sets have the same data sources for single-parameter network validation. Each dataset contains 900 and 300 samples respectively. Batch_size is the amount of data required for each iteration, and it is set to 32. This model adopts the optimization strategy of stochastic gradient descent, the maximum number of iterative steps is 5000 and the initial learning rate is 0.003. The study rate attenuation strategy adopts the exponential attenuation method. The attenuation interval is 3 epochs and the attenuation rate is 0.99.

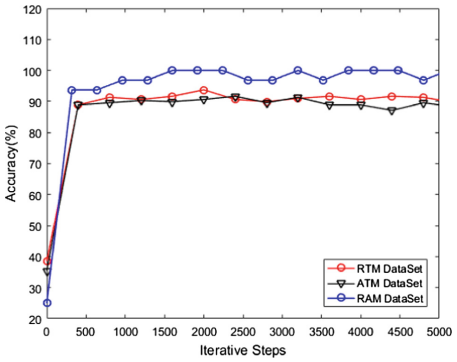

Fig. 5. R-TM, A-TM and R-AM accuracy curve.

Figure 5 compares the accuracy curve of R-AM and single-parameter network when separately trained on R-TM and D-TM data sets. The results show that compared

with the single-parameter gesture recognition approach, the proposed one has significantly improved the recognition accuracy by 5%.

This paper designs six types of gestures, including push forward, pull backward, scroll left, scroll right, push-pull and scroll left-right, the test set of each gesture has 50 data, the correct number of data predicted is 46, 49, 48, 47, 46 and 48 in R-AM's confusion matrix. In R-TM and A-TM confusion matrix, the test set of each gesture has 50 data, too. Figure 6 shows the confusion matrix of R-AM, R-TM and A-TM. It can be seen from the figure that the accuracy of R-AM network on the test set reaches 94.7%. The average accuracy rates of R-TM and A-TM were 91.8% and 90.0%. In Table 1 compared with the single-parameter network and the single-parameter data set, the method in this paper has improved the accuracy of gesture recognition by about 5%.

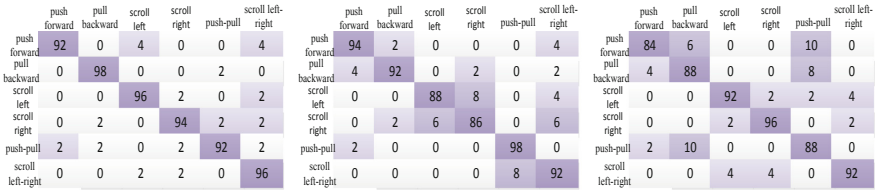

Fig. 6. R-AM, R-TM and A-TM confusion matrix.

Table 1. Comparison results of accuracy of different methods.

Network structure and data sets	Accuracy (%)
One parameter network+R-TM	91.7
One parameter network+A-TM	90.0
This article's fusion network+R-TM/A-TM	94.7

5 Conclusion

This paper proposed an improved convolutional neural network gesture recognition method based on two-parameter estimation of FMCW radar. Firstly, the radial distance and angle parameters of the forward hand gestures were obtained. Secondly, range and angle information were accumulated in time domain to construct the datasets of R-TM and A-TM. Then, the dataset were sent to the proposed CNN based hand gesture recognition system for training and testing. The results showed that compared with the single-parameter gesture recognition approach, the proposed one has significantly improved the recognition accuracy by about 5%.

Acknowledgments. This work was supported in part by the National Natural Science Foundation of China (61771083, 61704015), Program for Changjiang Scholars and Innovative Research Team in University (IRT1299), Special Fund of Chongqing Key Laboratory (CSTC),

Fundamental and Frontier Research Project of Chongqing (cstc2017jcyjAX0380, cstc2015jcyjBX0065), University Outstanding Achievement Transformation Project of Chongqing (KJZH17117), and Postgraduate Scienti_cResearch and Innovation Project of Chongqing (CYS17221).

References

1. Coelho, Y.L., Salomao, J.M., Kulitz, H.R.: Intelligent hand posture recognition system integrated to process control. IEEE Lat. Am. Trans. **15**(6), 1144–1153 (2017)
2. Salunke, T.P., Bharkad, S.D.: Power point control using hand gesture recognition based on hog feature extraction and K-NN classification. In: 2017 International Conference on Computing Methodologies and Communication (ICCMC), pp. 1151–1155. IEEE (2017)
3. Pai, N.S., Hong, J.H., Chen, P.Y., et al.: Application of design of image tracking by combining SURF and TLD and SVM-based posture recognition system in robbery pre-alert system. Multimed. Tools Appl. **76**(23), 25321–25342 (2017)
4. Park, J., Cho, S.H.: IR-UWB radar sensor for human gesture recognition by using machine learning. In: IEEE, International Conference on High PERFORMANCE Computing and Communications; IEEE, International Conference on Smart City; IEEE, International Conference on Data Science and Systems. pp. 1246–1249. IEEE (2017)
5. Zhou, Z., Cao, Z., Pi, Y.: Dynamic gesture recognition with a terahertz radar based on range profile sequences and doppler signatures. Sensors **18**(1), 1–15 (2018)
6. Wang, W., Liu, A.X., Shahzad, M., et al.: Device-free human activity recognition using commercial WiFi devices. IEEE J. Sel. Areas Commun. **35**(5), 1118–1131 (2017)
7. Li, G., Zhang, R., Ritchie, M., et al.: Sparsity-based dynamic hand gesture recognition using micro-Doppler signatures. In: 2017 IEEE Radar Conference (RadarConf), pp. 0928–0931. IEEE (2017)
8. Winkler, V.: Range Doppler detection for automotive FMCW radars. In: Microwave Conference, European, pp. 1445–1448. IEEE (2007)
9. Pan, H., Zhang, F., Shi, C., et al.: High-precision frequency estimation for frequency modulated continuous wave laser ranging using the multiple signal classification method. Appl. Opt. **56**(24), 6956–6961 (2017)
10. Vamplew, P., Dazeley, R., Foale, C.: Softmax exploration strategies for multiobjective reinforcement learning. Neurocomputing **263**, 74–86 (2017)