Gesture Recognition Using LFMCW Radar and Convolutional Neural Network

Peng-hui CHEN, Yu-jing BAI and Jun WANG*
School of Electronic and Information Engineering, Beihang University, Beijing, China
*Corresponding author

Keywords: Millimeter wave radar, LFMCW, Gesture recognition, R-D map, 3D-CNN.

Abstract. With the continuous enrichment of human-computer interaction, gesture recognition technology plays an increasingly important role in various fields, especially in the non-contact and low-light conditions. In order to overcome the limitations of traditional visible optical sensors, such as low-light, occlusion and privacy, a gesture recognition system based on linear frequency modulation continuous wave (LFMCW) radar is proposed. By preprocessing the LFMCW radar intermediate frequency (IF) signals, the time series of the two-dimensional range-doppler (R-D) maps of the gesture is obtained. The R-D map series is taken as the three dimensional convolutional neural network (3D-CNN) input samples, and the gesture features are extracted via convolution layer and pooling layer, and the gesture features are classified and identified via the fully connected soft-max classifier. To explore the application of the extracted gesture features, a 77GHz radar system is applied, and six kinds of gestures are designed to verify the performance of classification and recognition. The experimental results show that the gestures can be classified and recognized effectively and the real-time recognition accuracy of the whole system can reach about 94%.

Introduction

With the development of electronic information technology, in recent years, gesture interaction is continuing to attract people's attention as an emerging form of interaction. Because gesture interaction has the advantages of fast input speed, multiple input types, compatibility with human habits, high degree of freedom, and unlimited application scenarios, many researchers are engaged in the research and development of gesture recognition related content. Especially in recent virtual reality interactions, somatosensory games, smart homes, wearable devices and other consumer-grade electronic products, low-cost, fast and efficient gesture interaction is a kind of broad prospects for human-computer interaction technology[1]. And the radar sensor has the advantages of non-contact, less interference from the external environment, penetration, embedded in the device to hide the sensor and protect user privacy to be used in gesture interaction field.

Generally, there are many researches on gesture recognition, including the traditional method of gesture recognition and the recent emerging gesture recognition combined with deep learning neural networks[1,2,3,4]. The range-doppler (R-D) maps by processing the gesture radar signal, were applied to train the long-term recurrent all-convolutional network (LRACN) to realize the radar-based gesture recognition[5], and the result can achieve 94% recognition accuracy. For the gesture radar signals, the I/Q components and the angle-of-arrival (AOA) matrix are applied to train the deep convolutional neural network (DCNN)[6]. The result was 95% accuracy for 14 gestures.

Based on the above research, a three dimensional convolutional neural network (3D-CNN) is applied to improve the efficiency of the continuous gesture recognition based on the collected gesture signals using a 77GHz linear frequency modulation continuous wave (LFMCW) radar.
**Scheme of the Procedure**

The gesture recognition system based on FMCW radar and deep convolutional neural network proposed in this paper mainly includes radar signal acquisition, radar signal preprocessing, data set construction, offline network training and real-time gesture recognition classification. Among them, the radar signal acquisition and preprocessing part realizes the time-frequency analysis of the radar signal, and obtains the 2D R-D map of the gesture radar signal, and the data set construction realizes the construction of the multi-parameter data set required for training and testing the neural network. Off-line network training and real-time gesture recognition classification realize training, classification and real-time recognition of gesture data.

**LFMCW Radar Range-Doppler map (RDM)**

The main principle is that the transmitted wave is a high-frequency continuous wave whose frequency changes with the sawtooth wave with time. The frequency of the echoes received by the radar is the same as the variation of the transmission frequency. For the targets at different locations, the delay time depends on the distances. Therefore, the intermediate frequency (IF) is proportional to the distance. The waveform is shown in Fig. 1.

![LFMCW waveform](image)

Figure 1. LFMCW waveform.

According to the speed of light $c$, the distance can be obtained as

$$ d = \frac{c \tau}{2}. \quad (1) $$

Let the slope of the sweep chirp be, which is the ratio of the sweep bandwidth $B$ to the sweep period $T$, i.e.

$$ S = \frac{B}{T}. \quad (2) $$

Then, the intermediate frequency signal obtained by mixing the echo signal and the transmitted signal is

$$ f_{\text{IF}} = S \tau. \quad (3) $$

The final radar output is the IF signal, and the intermediate frequency signal contains the distance information of the measured target. The distance to the target can be obtained as

$$ d = \frac{c f_{\text{IF}}}{2S}. \quad (4) $$

where $c$ is the speed of light and $S$ is the setting of the radar parameters, and the sweep slope is obtained, so that the target distance can be obtained by the IF signal output by the radar.

When multiple targets are at the same distance from the radar, and it is impossible to distinguish multiple targets by distance alone, it is necessary to distinguish multiple targets by different speeds of different targets. At small scales, the speed at which multiple targets are resolved by doppler
frequency is not obvious, but the phase changes are very sensitive, so the different phases of the echo signals can be reflected by different targets to obtain velocity information of different targets.

When the phase information of multiple targets is mixed in the echo signal, the single-cycle echo signal cannot distinguish multiple targets, so it is necessary to accumulate the signals of multiple echo periods and perform fast Fourier transformation (FFT) to the frequency domain. Then the speed information of multiple targets is obtained.

According to the FFT-converted signal, phase information of a plurality of targets can be obtained, and phase information can be used to calculate speed information of a plurality of targets.

\[ v = \frac{\lambda \omega}{4\pi T_c}. \]  

where \( \lambda \) is the wavelength of the radar signal, \( \omega \) is the target phase information obtained by the FFT processing of the echo signal, and \( T_c \) is the period of the radar sweep Chirp. A group of signals generally consisting of N Chirp periods is a frame.

According to the previous derivation, it can be known that the processing of the radar echo signal can obtain the R-D map of the radar echo signal. A frame of radar echo signals consisting of N Chirps is stored in a matrix, and firstly, a 1D-FFT transform is performed in the matrix row direction to obtain information about the distance of the radar echo signals. And then, 1D-FFT conversion is performed in the matrix column direction to obtain an R-D map of the radar echo signal. Through the RDM diagram, characteristic information such as the distance and velocity of the target can be characterized, and the approximate position and velocity state of the target can be obtained. The procedure of this preprocessing is illustrated in Fig. 2.

\[ \text{Figure 2. Preprocessing of range-doppler map.} \]

In the gesture recognition application, the R-D map of the radar echo signal is mainly used for feature extraction and classification to obtain the result of different gestures.

**Gesture Action Data Set Construction**

This article mainly sets six kinds of gestures for classification and recognition which are the palms vertically move close to the radar, the palms vertically move away from the radar, the fingers pinch, the fingers open, the palms vertically swept left, and the palms vertically swept right. Each gesture action lasts approximately 1 second. The setting of the gesture action range is limited to 0~40cm, so it is necessary to intercept the information at 0~40cm from the matrix after 1D-FFT, and then perform 1D-FFT in the velocity direction on the intercepted matrix to obtain the finally required gesture R-D map. To acquire the gesture series, continuous calculation of 12 R-D maps indicate a single complete gesture. The size of the resulted time series R-D map matrix is 32×120×12. The three dimensions of a complete gesture motion data represent the distance, the speed, and the frame number, respectively.

Since the value of the R-D map has the large fluctuations, it is necessary to normalize each gesture motion data to meet the convolutional neural network requirement.

**Gesture Recognition Based on 3D Convolutional Neural Network**

In this paper, the three dimensional convolutional neural network is used to extract the feature motion data, and the fully connected soft-max classifier is used to classify the gestures based on the extracted gestures.
The three-dimensional convolutional neural network in this paper consists of two convolutional layers which have the structure of 5×5×5 and 3×3×3 respectively, two pooling layers which have a structure of 2×2×2, the pooling layer uses the max-pooling function and three fully connected layers. To prevent overfitting, add two dropout layers and use a soft-max linear classifier at the output. The specific network structure is shown in Fig. 3.

The gesture recognition process in this paper mainly includes two aspects. On the one hand, offline training using the prepared gesture data set, and iterative process through forward propagation and backward propagation on the server. On the other hand, online training is performed using trained network parameters. In this case, iterative training is not required, only the trained network parameters need to be loaded, and the data collected online is subjected to forward propagation calculation, thereby obtaining the result of classification identification.

Experimental Results

Experimental Environment

The experimental platform is based on a 77GHz millimeter-wave radar with a bandwidth of 4GHz and a sweep period of 420us. The sampling rate of the IF radar echo signal is 10MHz, and the coherent accumulation time is 26.88ms, which means that 64 consecutive Chirps in the same distance unit are used as one frame to process with FFT. At the same time, the gesture action of 12 frames can be obtained for about 1 second.

The training of the 3D-CNN is based on the Tensorflow deep learning framework[7], which is implemented in a Python environment and is calculated using an 8-core Intel Core i5 processor and 8 GB of memory.

Gesture Recognition Results Based on Convolutional Neural Networks

In this paper, the radar gesture signal is acquired by the 77GHz radar system, and the R-D map sequence matrix of the training recognition in the input neural network is obtained. A sequence of R-D map patterns including six gestures of approaching, moving away, grasping, opening, left pendulum, and right pendulum. Fig. 4 shows a schematic diagram of each gesture and a corresponding radar R-D map with a size of 32×120.

The R-D map time series is first input into the 3D-CNN for offline training. The training data set contains 480 units of samples, and the test set contains 120 units of samples.
For convolutional neural networks, the network accuracy rate can be obtained from the training accuracy curve varying with the number of iteration steps. The result is shown in Fig. 5. It can be seen that the offline recognition accuracy of the 3D-CNN is about 95% after training.

After obtaining the offline training result, the trained neural network is tested using the gesture data collected online to verify the recognition effect of the neural network. In the online verification stage, 20 groups of six gestures were collected, and a total of 120 units of samples were tested. The confusion matrix is shown in Table 1, where the relations between the abbreviation and each gesture are as follows: A–Approaching, M–Moving away, G–Grasping, O–Opening, L–Left, R–Right.

Table 1. Confusion matrix results.

<table>
<thead>
<tr>
<th>Gesture</th>
<th>Train set</th>
<th>Test set</th>
<th>A</th>
<th>M</th>
<th>G</th>
<th>O</th>
<th>L</th>
<th>R</th>
<th>acc(%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>80</td>
<td>20</td>
<td>18</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>90</td>
</tr>
<tr>
<td>M</td>
<td>80</td>
<td>20</td>
<td>0</td>
<td>19</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>95</td>
</tr>
<tr>
<td>G</td>
<td>80</td>
<td>20</td>
<td>1</td>
<td>0</td>
<td>19</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>95</td>
</tr>
<tr>
<td>O</td>
<td>80</td>
<td>20</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>19</td>
<td>0</td>
<td>0</td>
<td>95</td>
</tr>
<tr>
<td>L</td>
<td>80</td>
<td>20</td>
<td>0</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>18</td>
<td>0</td>
<td>90</td>
</tr>
<tr>
<td>R</td>
<td>80</td>
<td>20</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>20</td>
<td>100</td>
</tr>
</tbody>
</table>

The results show that the online recognition accuracy can reach about 94% on average, and it has a good classification and recognition result, which can clearly distinguish the data characteristics of the six gestures.

**Conclusion**

A gesture recognition procedure based on LFMCW radar and 3D-CNN is proposed. Firstly, the radar R-D maps which are used for feature extraction and network training are obtained by analyzing and preprocessing the radar signal, and a complete radar gesture R-D map sequence by 32×120×12 is obtained. Secondly, by designing a 3D-CNN neural network and using the collected data sets for offline training, a trained neural network is obtained. Finally, through the use of trained neural network for online identification verification, experiments show that the system's online gesture recognition accuracy is about 94%. And this system achieved the same accuracy of 94% with less data set for LRACN[5]. A higher recognition accuracy can be achieved by improving the training data set so as to implement in several practical applications.

**Acknowledgement**

The authors give thanks to Minglang Qiao for offering the CNN code.

**References**


