Featured Based Segmentation Method for Building Millimeter Wave Radar Gesture Recognition Data Sets

Xiao-ying ZHAO, Yi-lin ZHAO, Peng-hui CHEN and Jun WANG
School of Electronic and Information Engineering, Beihang University, Beijing, China
*Corresponding author

Keywords: Gesture segmentation, LFMCW, Gesture recognition, R-D map.

Abstract. Gesture recognition based on artificial neural network is an important application of the millimeter wave radar. In addition to extracting gesture features and constructing neural networks, the establishment of effective dynamic gesture data sets is also the direction worth paying attention to in gesture recognition research. In order to solve some problems caused by fixed radar observation time and multiple measurements of single gesture during the establishment of gesture data set, six gestures were tested in a single radar observation. Combining the relationship between actual gesture motion and speed, the data file containing six gestures is divided into six sets containing only one specific gesture which will be used in the gesture recognition system of millimeter wave radar based on neural network finally.

Introduction

The establishment of gesture data set is an important step in the process of gesture recognition based on neural network. Preliminary work on this gesture recognition system based on linear frequency modulation continuous wave (LFMCW) has been completed includes using millimeter wave radar to measure the original gesture data and obtaining the high range resolution profile (HRRP) of each group of gesture data through radar echo data preprocessing algorithm[1,2,3,4].

According to the gesture recognition system, radar observation include 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. Usually, when the millimeter wave radar observes gestures, it sets a fixed observation time and a single gesture. However, due to the different speed of hand movement of different test subjects, the fixed observation time may lead to incomplete gesture movement or excessive redundant time. In order to build a sufficiently large set of gesture data, repeated testing is often required. In fact, gesture is a process of coming and going. For example, in order to get the palm close to the radar repeatedly, after one close action, the palm has to generate a distance action to return to the original test distance. In order to separate the two movements, the radar needs to be manually controlled during observation. This also adds some complexity to the testing process.

In order to ensure enough observation time to obtain the complete gesture and reduce the redundant time as much as possible, this paper chooses to complete the gesture acquisition under the condition that a radar observation does not stop, that is, the gesture data obtained by a radar observation is continuous in time. That means six gesture data will be stored in one file. However, neural network test set and training set in gesture recognition system are data sets containing only one specific gesture. The main work of this paper is to segment the continuous gestures by using the speed characteristics of gesture motion based on high-resolution distance image, and finally obtain six sets of different gesture data sets.

Radar Model

In this paper, LFWCW millimeter-wave radar is adopted. The transmitted wave is frequency modulated continuous wave, whose frequency varies with time according to sawtooth wave law.
The radar transmitted signal and echo signal are

\[ s_f(t) = A_f \cos 2\pi \left[ f_c t + \int_0^t f_r(\tau) d\tau \right], \]

and

\[ s_R(t) = A_R \cos 2\pi \left[ f_c (t - \Delta t_d) + \int_0^t f_r(\tau) d\tau \right], \]

respectively, where \( A_f \) and \( A_R \) is the amplitude of the transmitted signal and the echo signal, \( f_c \) is the center frequency of radar, \( \Delta t_d \) is time delay, \( \gamma \) is the slope of the sweep chirp, \( \Delta f_d \) is doppler frequency shift, \( f_r(\tau) \) and \( f_r(\tau) \) is the frequency of the transmitted signal and the echo signal, i.e.

\[ f_r(\tau) = \gamma \tau, \quad f_r(\tau) = (\tau - \Delta t_d) \gamma + \Delta f_d. \]

Intermediate frequency (IF) signals are obtained after mixing and filtering of radar transmitted signals and echo signals, i.e.

\[ s_{IF}(t) = \frac{1}{2} A_f A_R \cdot \cos 2\pi \left[ f_c \Delta t_d + (\gamma \Delta t_d - \Delta f_d) t \right], \]

where \( f_{IF} = \gamma \Delta t_d \) is the IF signal frequency.

The relative radar distance of the gesture target can be calculated by light speed and echo delay, that is

\[ R = \frac{1}{2} c \cdot \Delta t_d = \frac{1}{2} c \cdot \frac{f_{IF}}{\gamma}. \]

Therefore, the phase of IF signals contains the distance and doppler information of the target, which can be estimated by frequency domain analysis.

**Gesture Segmentation Method**

The basic principle of this gesture segmentation method can be explained as follows. There will be a time when the velocity is zero during the process of switching from the previous gesture to the next so that the segmentation of different gestures can be achieved by calculating the speed distribution of the dynamic gesture and selecting the speed zero as the segmentation point of gesture data.

Since the velocity information of multiple dynamic gestures is mixed in the radar echo signal, which cannot be distinguished only by the echo signal of a single period, the radar signals of multiple echo periods are accumulated in time and converted to the frequency domain through fast Fourier transformation (FFT) for analysis, so as to obtain the velocity information of dynamic gestures. By applying FFT and maximum projection to each frame of the HRRP containing a group of six gestures, the velocity of various gestures can be calculated

\[ v = \max \left( f_r(S(r,t_1), S(r,t_2), ..., S(r,t_6)) \right), \]

where \( v \) is the velocity of dynamic gesture, \( S(r,t) \) means HRRP of a period radar echo. The HRRP of radar echo data of six gestures is stored in a matrix, and each line represents a one-dimensional range profile of radar sequence echo. \( f_r \) means FFT that is applied to each column of the matrix, \( \max \) means maximum projection on doppler frequency.

Fig.1 shows the velocity of the motions of a palm moving close to and away from the radar, and sweeps left and sweeps right. Comparing the gesture observation process with the velocity curve, it can be seen that the point where the velocity is zero is the moment when the two gestures switch. Therefore, according to the velocity curve, the zero point of velocity is manually selected as the
segmentation point of the HRRP of the dynamic gesture. Although it is not convenient to select points manually, some invalid gestures caused by subjective reasons of test objects can be deleted during the process of manually selecting points. In addition, in the actual process, we found that the above method also has some disadvantages. When the hand movement of the test object is slow or the changes of hand movements were relatively small, the curve of the velocity distribution diagram is not clear enough. This leads to a large error when manually selecting the velocity zero point. Even the velocity curve of some gestures does not have obvious fluctuation, so it is difficult to find the point where the velocity is zero.

![Figure 1: Velocity curves: (a) a palm moves close to the radar and moves away from the radar; (b) a palm sweeps left and sweeps right.](image)

**Improved Differential-Velocity Segmentation**

In order to improve the accuracy of gesture segmentation, this paper suggests that the HRRP can be firstly differentiated in time to remove the background clutter interference. Radar echoes can be regarded as composed of dynamic gesture echo and static background clutter, i.e.

\[
\begin{align*}
    S_1(r,t_i) &= B + G(r,t_i) \\
    S_2(r,t_2) &= B + G(r,t_2) \\
    &
    \vdots \\
    S_N(r,t_N) &= B + G(r,t_N)
\end{align*}
\]

where \( B \) is static background clutter, and \( G(r,t_i), (i=1,2,...,N) \) is the one-dimensional range profile of radar echo generated by gesture.

To restrain the effects of the background on the micro motion feature extraction of gesture echoes, the background can be removed by the coherence cancellation in time domain, which can be expressed as

\[
\begin{align*}
    \Delta S_1 &= S(r,t_1) - S(r,t_1) = G(r,t_1) - G(r,t_1) \\
    \Delta S_2 &= S(r,t_2) - S(r,t_2) = G(r,t_2) - G(r,t_2) \\
    &
    \vdots \\
    \Delta S_{N-1} &= S(r,t_{N-1}) - S(r,t_{N-1}) = G(r,t_{N-1}) - G(r,t_{N-1})
\end{align*}
\]

According to the relevant knowledge of the Fourier transform, the time shift is equivalent to the phase change in the frequency domain, so the velocity of the gesture calculated by using the fast Fourier transform and the maximum projection can be expressed as
\[ v = \max \left( f_i(\Delta S_1, \Delta S_2, \ldots, \Delta S_{N-1}) \right) \]
\[ = \max \left\{ f_i \left[ G(r, t_1, t_3, \ldots, t_N) - G(r, t_1, t_2, \ldots, t_{N-1}) \right] \right\} \]
\[ = \max \left\{ f_i \left[ G(r, t_1 + \Delta t, t_2 + \Delta t, \ldots, t_{N-1} + \Delta t) - G(r, t_1, t_2, \ldots, t_{N-1}) \right] \right\} \]
\[ = \max \left\{ (e^{j(\omega \Delta t)} - 1) \cdot f_i \left[ G(r, t_1, t_2, \ldots, t_{N-1}) \right] \right\} \] (9)

In this way, the velocity distribution of gesture motion can be obtained after the maximum projection and will not be affected by the background echo.

Fig.2 illustrates the effect of the background. Compared with the simple use of one-dimensional range profile to calculate the velocity distribution, the velocity curve obtained after differential operation is more clear, which can effectively improve the accuracy of manually selecting the velocity zero point.

![Velocity curve without background clutter: (a) a palm moves close to the radar and moves away from the radar; (b) a palm sweeps left and sweeps right.](image)

**Results**

In order to verify the result of gesture segmentation and establish the gesture recognition data set, this paper calculates the range-doppler (R-D) maps of each gesture based on the segmented gesture data. Fig.3 shows the results of a palm moving closes to the radar. When calculating the R-D map of a gesture with a long duration, it should be divided into several R-D maps evolving with time, and the step size in time should be selected according to certain rules. In this paper, a series of R-D maps with a size of 32×128×12 are obtained by one gesture finally.

**Summary**

This algorithm combines the speed change in the process of hand movement and frequency domain analysis, and realizes the segmentation of a group of continuous gestures by looking for the moment when the speed of dynamic gesture is zero in the movement. It not only simplifies the process of gesture data measurement, but also adjusts the signal time so that the redundant time can be reduced on the basis of ensuring the complete gesture movement, even if the gesture movement time of different observers is not uniform.
Figure 3. Segmentation of a palm moving close to the radar. (a) HRRP; (b) the R-D map for the neural network; (c) 12 R-D maps evolving over time of one gesture which will be used in the neural network.

References

[1] S. Ryu, J. Suh, S. Baek, etc. “Feature-Based Hand Gesture Recognition Using an FMCW Radar and its Temporal Feature Analysis,” IEEE Sensors Journal, vol. 18, no. 18, pp. 7593-7602, 2018.

[2] G. Li, S. Zhang, F. Fioranelli, etc. “Effect of sparsity-aware time–frequency analysis on dynamic hand gesture classification with radar micro-Doppler signatures,” IET Radar, Sonar & Navigation, vol. 12, no. 8, pp. 815-820, 2018.

[3] T. Sakamoto, X. Gao, E. Yavari, etc. “Hand Gesture Recognition Using a Radar Echo I–Q Plot and a Convolutional Neural Network,” IEEE Sensors Letters, vol. 2, no. 3, pp. 1-4, 2018.

[4] Y. Sang, L. Shi and Y. Liu. “Micro Hand Gesture Recognition System Using Ultrasonic Active Sensing,” IEEE Access, vol. 6, pp. 49339-49347, 2018.