Analysis and Research on Radar Gesture Micro-motion Feature Extraction Method

Yaoyao Dong\textsuperscript{1*}, Wei Qu\textsuperscript{2}, Tianhao Gao\textsuperscript{3}, Pengda Wang\textsuperscript{4} and Haohao Jiang\textsuperscript{5}

\textsuperscript{1,3,4,5} Graduate School, Space Engineering University, Beijing, 101400, China
\textsuperscript{2} Department of Electronic and Optical Engineering, Space Engineering University, Beijing, 101400, China
\textsuperscript{*}Corresponding author’s e-mail: hgddyy@dingtalk.com

Abstract. Radar gesture recognition technology is widely used in daily life, and the micro motion feature is the key information to distinguish the feature quantity of different gestures. The extraction of micro motion feature is the key step of radar gesture recognition. Joint time-frequency analysis is a common method for radar signal micro motion feature extraction. Firstly, this paper makes a theoretical analysis of the common joint time-frequency analysis methods; Then radar gesture model is completed, and the performance of different JTFA is compared by extracting the time-frequency features of the motion gesture echo; Finally, the measured data are collected under the condition of microwave anechoic chamber. After analysis, the theoretical simulation results are consistent with the measured data, which verifies the reliability of the simulation.

1. Introduction

As a relatively new way of human-computer interaction, there is no doubt about the convenience and practicability of gesture recognition technology. Gesture recognition is well reflected in application scenarios such as smart home\cite{1}, assisted driving\cite{2}\cite{3}, virtual reality\cite{4}, sign language translation\cite{5}, game entertainment\cite{6}. Early gesture recognition methods mainly include gesture recognition based on wearable sensors, visual images such as depth camera\cite{7}, wireless communication signals such as RFID\cite{8}, Wi Fi\cite{9}, and radar sensors. Wearable sensor technology is cumbersome and expensive; Vision sensor gesture recognition technology is largely limited by light conditions, but also has the security problem of privacy disclosure. Wireless communication signal gesture recognition technology has low resolution and is difficult to filter out background interference. In addition to the advantages of all-weather, all-time and strong flexibility, radar sensor can also effectively protect users' privacy information in terms of privacy security.

The extraction of hand motion feature is the key step of radar gesture recognition. The micro motion feature is the key information to distinguish the feature quantity of different gestures. Fretting characteristics are reflected in radar signal processing, especially micro Doppler shift. Micro Doppler shift is a time-frequency shift. In order to analyze time-varying frequency characteristics, Fourier transform is not applicable because it can not provide time-dependent frequency information. The common analysis methods to describe signals in time domain and frequency domain are instantaneous frequency analysis and joint time-frequency analysis (JTFA).

As a classical method for analyzing time-varying spectrum, joint time-frequency distribution is designed to interpret information for the energy distribution of a given signal in two-dimensional time-
frequency domain. It is very suitable not only for single component signals, but also for multi-component signals. Based on the radar gesture simulation, this paper mathematically deduces the common JTFA methods, and then verifies the simulation results and mathematical essence through the radar measured data.

2. Theory of micro motion feature extraction

Human motion is divided into two categories: macro motion and micro motion. Big sport refers to the movement in which the position of the human body changes such as running and walking; Small movement, also known as micro movement, refers to the movement in which the human body's position remains unchanged, such as the movement of limbs and head in situ, as well as the breathing and heartbeat in a static state. The signal to clutter and noise ratio (SCNR) of hand micro motion target echo is very low, so the micro motion feature extraction has become the key of target classification.

According to the movement of human hand, the radar echo has the following characteristics: the micro Doppler frequency changes nonlinearly with time; The physiological structure of human hand makes the target echo generally multi-component signal; At the same time, the signal component may have amplitude modulation. The traditional Fourier transform analysis can not reflect the time-varying characteristics of frequency. Therefore, at present, the more common processing method is to use JTFA to analyze the micro doppler characteristics, and then complete the extraction of micro motion features. JTFA include linear time-frequency distribution and quadratic time-frequency distribution, such as short-time Fourier transform (STFT) \cite{10} \cite{11}, Wigner Ville distribution (WVD), pseudo Wigner Ville distribution (PWVD), smooth pseudo Wigner Ville distribution (SPWVD) etc.

STFT is the initial attempt to introduce time-domain information based on Fourier analysis. Through reasonable selection of window function, time-frequency analysis can be realized to a certain extent. For a given non-stationary signal \( s(t) \in \mathcal{L}^2(R) \), its STFT is defined as:

\[
\text{STFT}_h(t, \omega) = \int_{-\infty}^{\infty} s(\tau) h(\tau - t) \cdot \exp(-j\omega\tau) \, d\tau
\]

Among them, \( h(t) \) is called window function. STFT has the highest calculation efficiency and high spectral quality. It is the most commonly used time-domain spectral calculation method. However, this method also has some limitations, that is, because the time resolution and frequency resolution will be limited by the width of window function, it can not achieve the best effect at the same time.

WVD makes up for the deficiency of Fourier transform to a certain extent and has high time-frequency resolution. WVD of signal is defined as:

\[
\text{WVD}_s(t, \omega) = \int_{-\infty}^{\infty} s(t + \frac{\tau}{2}) \cdot s^\ast(t - \frac{\tau}{2}) \cdot \exp(-j\omega\tau) \, d\tau
\]

Where, \( s^\ast(t - \frac{\tau}{2}) \) is the conjugate form of \( s(t - \frac{\tau}{2}) \).However, there is a serious short board in the application of WVD, that is, there is a cross term interference (CTI), which limits the application to a certain extent. Assuming the signal \( s(t) = s_1(t) + s_2(t) \), the WVD of the signal brought into the formula is:

\[
\text{WVD}_s(t, \omega) = \int_{-\infty}^{\infty} \left[ s_1(t + \frac{\tau}{2}) + s_2(t + \frac{\tau}{2}) \right] \cdot \left[ s_1^\ast(t - \frac{\tau}{2}) + s_2^\ast(t - \frac{\tau}{2}) \right] \cdot \exp(-j\omega\tau) \, d\tau
\]

\[
= \text{WVD}_h(t, \omega) + \text{WVD}_{s_1}(t, \omega) + 2 \cdot \text{Re} \left[ \text{WVD}_{s_1, s_2}(t, \omega) \right]
\]

It can be seen from the formula that the WVD distribution of the two signals is the sum of each individual signal WVD, and the cross WVD between the two signals is added at the same time. In this way, it will seriously interfere with the time-frequency analysis results, thus drowning the information of the target signal. In order to eliminate the disadvantages brought by the cross term, the local characteristics of the signal can be analyzed by the WVD of the truncated signal by windowing the
signal in the frequency domain like STFT. At this time, the pseudo Wigner Weill distribution is obtained. PWVD is defined as:

$$\text{PWVD}_s(t, \omega) = \int_{-\infty}^{\infty} s_h(t + \frac{\tau}{2}) \cdot s_h^*(t - \frac{\tau}{2}) \cdot \exp(-j\omega \tau) d\tau$$

(4)

For multiple components, PWVD can effectively reduce the influence of cross terms. Continue to smooth PWVD with time-domain window function to obtain SPWVD. SPWVD is defined as:

$$\text{SPWVD}_s(t, \omega) = g(u) \cdot \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} s_h(t + \frac{\tau}{2}) \cdot s_h^*(t - \frac{\tau}{2}) \cdot \exp(-j\omega \tau) d\tau du$$

(5)

Where, $g(u)$ is a real even function, and $g(0) = 1$. SPWVD further reduces the interference caused by cross terms, but the smoothing result also reduces the time-frequency resolution. Therefore, there is a trade-off between smoothness and resolution.

3. Radar gesture simulation and experimental verification

3.1. Radar gesture modeling

The geometric layout of the radar human motion gesture simulation scene is shown in Figure 1. The millimeter wave radar with the wavelength of 3.89mm is located in the $[2,0,1]$ radar fixed coordinate system, the corresponding radar transmission frequency is 77GHz, the radar bandwidth is 4GHz, and the corresponding range resolution is 3.75cm. The duration of hand movement is set to, and the number of sampling points of radar echo is 2048. The simulation adopts the form of "target complex echo superposition", that is, the whole human hand echo is composed of the target echo of each finger. Because the thumb is not reflected in the gesture studied in this paper, and the target of interphalangeal joint is smaller than that of phalanx, the target complex echo mainly includes the phalanx of index finger, middle finger, ring finger and little finger. The proximal phalanx of the index finger is set as 2.4cm, and the middle man's wrist bone and metacarpal bone are simplified into a unified cylinder target, and this position is taken as the starting point $[0,0,1]$ of hand movement. The reciprocating motion of the human hand along the Z axis of the radar fixed coordinate system is defined as the up and down hand gesture. Let the motion speed of the hand be 0.5m/s and the acceleration be -0.31m/s².

![Radar Fixed Coordinate System (X,Y,Z)](image)

Figure 1. Motion gesture simulation scene

3.2. JTFA performance simulation comparison

JTFA method is used to compare the pure gesture echo data. STFT, WVD, PWVD and SPWVD methods are used for simulation analysis. The Hamming window of default length is used for analysis. Taking the hand up and down gesture as an example, the corresponding time-frequency image is shown in Figure 2. It can be seen from the figure that STFT can clearly reflect the doppler characteristics of up and down gestures; Although the resolution performance of WVD method has been improved, the disadvantages of cross terms caused by analysis can not be avoided; Although PWVD smoothes the cross terms, there is still some interference in the analysis results; Spwvd further suppresses the cross
term, which results in the reduction of frequency domain resolution and damages the frequency properties of the original WVD.

![Time frequency comparison diagram of up and down gestures](image)

**Figure 2. Time frequency comparison diagram of up and down gestures**

In addition to analyzing different time-frequency images, this paper also compares the computational complexity of different JTFA methods. Taking the analysis of hand up and down gestures as an example, the calculation time of each method under the same conditions is shown in Table 2. It can be seen from the table that the calculation efficiency of STFT under the same conditions is 8.04 times that of SPWVD.

| JTFA method | STFT    | WVD     | PWVD    | SPWVD   |
|-------------|---------|---------|---------|---------|
| Time Consuming (s) | 0.207800 | 0.211701 | 0.198414 | 1.670352 |

### 3.3. Verification of radar measured data

The microwave anechoic chamber built with special microwave absorbing materials and metal shields can provide a relatively pure electromagnetic environment for testing. Collecting the relevant radar gesture echo in the darkroom can avoid the interference of background clutter to the greatest extent, improve the accuracy and efficiency of the measured gesture, facilitate the verification of simulation data and ensure the reliability of verification. In this section, awr1642 millimeter wave radar system is used to detect gesture targets. Figure 3 (a) is the schematic diagram of the experimental scene, and figure 3 (b) is the real scene of gesture acquisition in the microwave darkroom.
Under the same analysis conditions, the gesture echo data collected in the darkroom are compared, the STFT, WVD, PWVD and SPWVD method is used for actual measurement and analysis respectively. The corresponding time-frequency image is shown in Figure 4. It can be seen from the figure that with the continuous increase of radar signal components, STFT has great advantages in time-frequency resolution. The cross term interference of WVD is too serious. Although PWVD and SPWVD have improved in suppressing cross terms, the lost frequency information can not be made up.
Table 2. Comparison of calculation time of measured JTFA method.

|JTFA method| STFT  | WVD    | PWVD   | SPWVD  |
|-----------|-------|--------|--------|--------|
|Time Consuming (s)| 1.514003| 1.692675| 1.412631| 29.334233|

Under the same conditions, the calculation time of each method is shown in table 2. It can be seen from the table that the calculation efficiency of STFT under the same conditions is 19.37 times that of SPWVD.

4. Conclusion

By completing the simulation analysis of radar moving hand echo, benchmarking the theoretical basis, close to the practical application scene, this paper makes a further demonstration of the theory from the result level, and also gives the time-frequency analysis results of different JTFA, which lays a theoretical basis for the analysis and research of radar gesture measured data in the future.

References

[1]. Rahman M A, Hossain M S. A Gesture-Based Smart Home-Oriented Health Monitoring Service for People with Physical Impairments[C]// International Conference on Smart Homes and Health Telematics. Springer International Publishing, 2016.

[2]. Joseph J, Divya D S. Hand Gesture Interface for Smart Operation Theatre Lighting[J]. International Journal of Engineering & Technology, 2018, 7(2):20-23.

[3]. Faheem K, Seong L, Sung C. Hand-Based Gesture Recognition for Vehicular Applications Using IR-UWB Radar[J]. Sensors, 2017, 17(4):833-.

[4]. Xia Z, Sang X, Chen D, et al. An interactive VR system based on full-body tracking and gesture recognition[C]// Optical Design & Testing VII. International Society for Optics and Photonics, 2016.

[5]. Orihuela-Espina, Isabel F, Palafox L, et al. Neural reorganization accompanying upper limb motor rehabilitation from stroke with virtual reality-based gesture therapy.[J]. Topics in Stroke Rehabilitation, 2013, 20(3):197-209.

[6]. Divkar A, Bailkar R, Pawar C S. Gesture Based Real-time Indian Sign Language Interpreter[J]. International Journal of Scientific Research in Computer Science Engineering and Information Technology, 2021:387-394.

[7]. Plouffe G, Cretu A M. Static and Dynamic Hand Gesture Recognition in Depth Data Using Dynamic Time Warping[J]. IEEE Transactions on Instrumentation and Measurement, 2016,65(2):305-316.

[8]. Zou Y, Jiang X, Han J, et al. GRfid: A Device-Free RFID-Based Gesture Recognition System[J]. IEEE Transactions on Mobile Computing, 2017, 16(2):381-393.

[9]. Tang Z, Liu Q, Wu M, et al. WiFi CSI gesture recognition based on parallel LSTM-FCN deep space-time neural network[J]. China Communications, 2021, 18(3):205-215.

[10]. WANG S W, SONG J. Interacting with Soli: Exploring Fine-Grained Dynamic Gesture Recognition in the Radio-Frequency Spectrum [A]. Proceedings of the 29th Annual Symposium on User Interface Software and Technology [C]. Tokyo: ACM, 2016:851−860.

[11]. ZHANG S, RITCHIE M A, FIORANELLI F, et al. Dynamic Hand Gesture Classification Based on Radar Micro-Doppler Signatures[C]// 2016 CIE INTERNATIONAL CONFERENCE ON RADAR. IEEE, 2016.