

Gesture Recognition Based on Wireless Technology: Present and Future

Junran Wan*

Communication Engineering, Beijing University of Posts and Telecommunications, Beijing, China

*Corresponding author: wanjunran@bupt.edu.cn

Abstract. As a cutting-edge innovation of non-contact human-computer interaction, wireless gesture recognition technology is gradually becoming the focus of academic and industrial attention because of its significant advantages in many fields. This paper aims to make an analysis for three main wireless sensing gesture recognition technologies, including their principles of work, implementation methods and technical characteristics. By collecting the recent development status of wireless sensing gesture recognition technology, this paper discusses the application of this technology. In addition, this paper also analyzes the limitations of the technology, Wi-Fi-based gesture recognition and RFID-based Gesture Recognition, and looks forward to the future development direction. The article focuses on radar-based gesture recognition, and its use in healthcare, smart homes, and car driving. This article also points out the challenges in terms of data scarcity, system adaptability and user diversity. The authors also suggest feasible developments for multimodal fusion, continuous learning algorithms, and adaptive algorithmic direction improvement. In the future, with the continuous progress and popularization of technology, wireless gesture recognition technology is expected to be widely used in more fields.

Keywords: Gesture recognition; wireless technology; radar; Wi-Fi; RFID.

1. Introduction

With the rapid popularity of the Internet of Things (IoT) and smart devices, non-contact human-computer interaction technology has become an important hotspot in current scientific and technological research. Among the many non-contact interaction methods, gesture recognition has gradually gained widespread attention due to its simple, intuitive, natural and convenient characteristics. Traditional gesture recognition techniques, such as vision-based and sensor-based methods, although they have mature technology and wide applications, have more obvious shortcomings. Vision-based gesture recognition usually relies on optical devices such as cameras, and there are obvious shortcomings in privacy protection, as the user's private images and related video data may be recorded and stored by the camera device, thus posing the risk of privacy leakage. Sensor-based gesture recognition relies on a variety of sensor devices that the user must wear for recognition, which increases the complexity of use and the cost of recognition.

In contrast, wireless sensing technology offers significant advantages in the area of gesture recognition. The use of wireless signals, such as Wi-Fi, RFID, millimetre wave, for gesture recognition not only improves the convenience of contactless recognition, but also effectively protects user's privacy. These technologies achieve gesture recognition by analysing the changes in various eigenvalues caused by changes in gestures during the propagation of wireless signals. Since the wireless signal itself has the advantages of strong penetrability and good adaptability to the environment, it enables the gesture recognition system to work stably under complex lighting and environmental conditions. The wireless gesture recognition technology does not rely on the detection of wearable instruments, which makes the recognition of gestures more convenient. At the same time, this technology can borrow the existing wireless network signal has been built, significantly reducing the cost of equipment. Gesture recognition technology based on wireless perception has a wide range of application prospects, and can play an important role in smart home, medical domain, virtual reality, industrial control and other fields.

In the follows this paper will analyse three main wireless perception-based gesture recognition techniques and presents their applications in various fields. Based on this, the shortcomings and future development of this technology will be described.

2. Overview of Wireless Gesture Recognition Technology

Wireless gesture recognition technology is divided into three categories: radar-based gesture recognition, Wi-Fi-based gesture recognition and RFID-based gesture recognition. Each method has its own unique principle, which will be described in detail below.

2.1. Radar-based Gesture Recognition

The main process of gesture recognition by radar, as shown in Fig.1: the first step is to collect gesture data, the use of radar sensors to collect the user's dynamic gesture information, the main application of the two types of radar sensors, one type of radar sensors concentrated in the k-band and the following bands, according to the IEEE 521-2002 standard, the K-band refers to the frequency of the radio wave band in the range of 12-18GHz, is prone to interference from the surrounding environment, such as absorption by water vapour. The k-band refers to the frequency of the 12-18 GHz radio wave band, which is susceptible to interference from the environment, such as absorption by water vapour. The other category mainly uses millimetre wave, which has greater advantages than K-band, with a wider spectrum, suitable for multi-access multiplexing in conjunction with other services, and can greatly increase channel capacity. Its reliability is higher, and the higher frequency can provide good resistance to environmental influences to better provide a stable transmission channel. Then, the received radar echo signals need to be pre-processed to maximise the extraction of the target's gesture features and remove interfering signals; finally, according to the set of gesture features, a suitable gesture recognition algorithm is selected to obtain the final result.

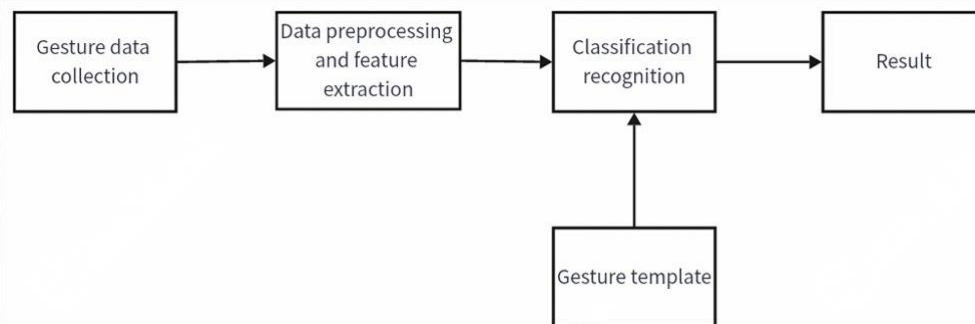


Fig. 1 Radar recognition structure diagram

Because millimetre wave radar has a greater advantage, the mainstream application of radar gesture recognition system is FMCW millimetre wave radar, in the process of gesture recognition, FMCW millimetre wave radar will be sent to the hand of the frequency of the FM continuous wave increases with time. The received and transmitted signals are filtered by a mixer and a low-pass filter and then sampled by an ADC, which generates an intermediate frequency (IF) signal. The distance, speed, and angle of the IF signal will change due to the gesture changes. The radar-based gesture recognition is intended to reflect the movement of the gesture through the changes in distance, Doppler frequency and angle. The system architecture is shown in Fig. 2 [1].

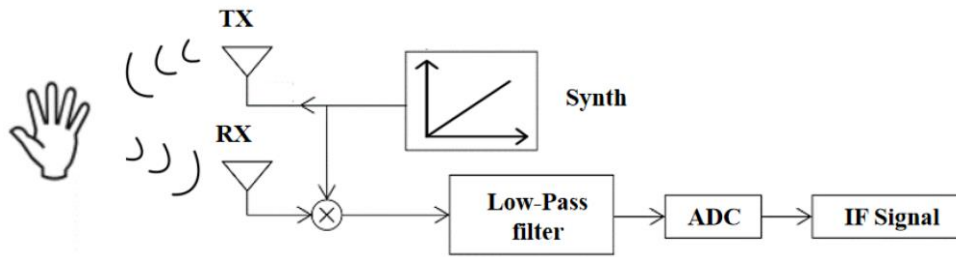


Fig. 2 Radar-based gesture recognition process[1]

Distance measurement in FMCW radar is based on the time delay between the transmitted and received signals. The distance (D) to the target can be calculated using the following formula:

$$D = \frac{\tau \cdot c}{2} = \frac{cT}{2B} f_{IF} \quad (1)$$

In formula(1), τ is the time delay between the transmitted and received signals, c is the speed of light, f_{IF} is the frequency of the IF signal, B and T are the bandwidth and sampling period of the chirp signal. This can be achieved by performing an FFT on the IF signal.

The speed of the FMCW radar is measured based on the Doppler effect, which can be obtained from the phase difference. The velocity (v) can be expressed by the following equation:

$$v = \frac{\Delta\phi\lambda}{4\pi T} \quad (2)$$

In formul(2), $\Delta\phi$ is the phase difference and λ is the wavelength of the chirp signal. This process can be achieved by using 2D-FFT on the IF signal.

Angle measurement in FMCW radars are usually implemented using a multi-cell antenna array. The radar can estimate the angle of the target based on the phase difference between the signals at different antennas. The angle (θ) of the target is expressed as follows:

$$\theta = \sin^{-1}\left(\frac{\Delta\phi\lambda}{2\pi L}\right) \quad (3)$$

In which, L is the distance between adjacent antenna units.

By varying these three variables, the radar can achieve recognition of the given variation of the gesture.

In 2020, S. Franceschini's team combined radar gesture recognition with convolutional neural network(CNN) and implemented a three-layer CNN for gesture classification [2]. They used the CNN to receive a spectrogram of the gesture as the input and output the labels associated with the gesture classification. Compared to the shortcomings of other methods that require the design of a extra feature collection step, the use of CNN networks has the advantage of being able to extract useful features on its own, allowing the model to be evaluated for optimal weights very quickly after the initial training process. The main disadvantage of CNN is the large amount of data used in the training step, but its computational burden can be reduced by the high-performance graphics processing unit(GPU), which allows CNN training to be completed in a feasible time frame. S. Franceschini's team has experimented and concluded that the network is very effective in classifying an internal experimental dataset, with an accuracy of 97% [2].

2.2. Wi-Fi-based Gesture Recognition

Most of the existing researches on Wi-Fi based gesture recognition use the amplitude and phase of the channel state information (CSI) as features to distinguish between different human activities. The wireless signals are transmitted from the transmitter to the receiver through multiple propagation paths as shown in Fig. 3 [3]. The propagation of the wireless signal is blocked when hands are moving. At this point the signal received by the receiver exhibits some significant changes. CSI is commonly used to describe the changes in the received signal.

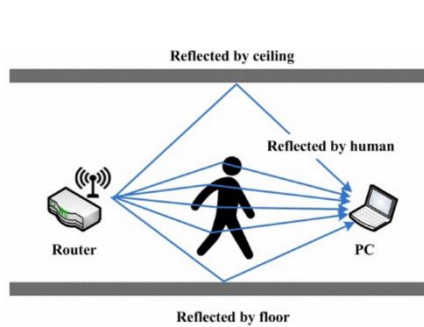


Fig. 3 Signal path schematic[3]

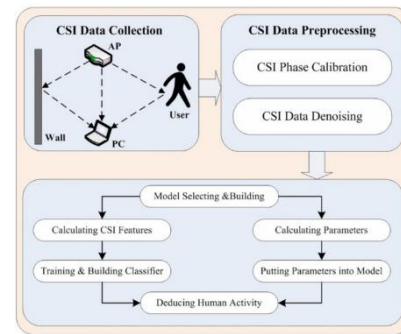


Fig. 4 Basic flowchart[3]

The flow of most models used in practice is shown in Figure 4, which consists of three parts: CSI data collection, signal preprocessing and human activity inference modeling [3]. The first step is signal acquisition, CSI data can be collected by different devices, but in raw CSI data usually contains a lot of noise caused by environmental factors and hardware devices. Therefore, a preprocessing step is necessary to remove the noise. In order to match the changes in the CSI signals with the gestures, it is necessary to describe the changes in the CSI signals using an appropriate physical model. For simple activities, some systems use the model to directly compute parameters to determine the user's position or activity, as shown in the right-hand flow in Fig.4. For complex activities, most applications first use the physical model to extract CSI features and then use classifiers trained on these features to complete human activity recognition which is shown in the flowchart on the left side of Fig.4.

There are following two approaches in collecting the changes of gesture, AoA model and Fresnel model. AoA model is usually used for human localization and tracking, while the Fresnel model is more sensitive to small movements and used for more accurate movement detection.

The core idea of AoA technology is that sensing the angle of arrival of signals sent by other devices through hardware devices and calculating the relative direction of the receiving node and the sending node. The angle of arrival is usually obtained using antenna array technology. In a multi-antenna array, there is a time difference for signals arriving at the antenna array from different angles, and this time difference corresponds to the arrival angle of the signal. There are two common methods for calculating the time difference between the arrival of signals at different antennas: the first method is the Multiple Signal Classification (MUSIC) algorithm; the second is the beamforming technique, which amplifies the signals in different directions and detects the signal strength information in different directions to determine the angle of arrival.

The Fresnel model consists of a set of Fresnel zones. A Fresnel zone is a series of concentric ellipsoids as shown in Fig.5 [3]. The transmitter and receiver are located at the two focus of the concentric ellipsoids, P1 and P2. The innermost ellipsoid with these two points as the focus is called the first Fresnel zone, and the second Fresnel zone corresponds to the ellipsoid ring space between the first and second ellipsoids. Similarly, the n^{th} Fresnel zone corresponds to the ellipsoid ring space between the $(n - 1)^{th}$ and n^{th} ellipsoid. When the human body is in the Fresnel region, the wireless signal is very sensitive to the movement of the human activities. The boundary between two adjacent Fresnel regions is defined as the Fresnel boundary. When the human body crosses the Fresnel boundary, the amplitude and phase of the wireless signal will change accordingly. This phase change can be calculated from the change in the length of the reflection path caused by the movement of the human body. Changes in human gestures can be detected by performing the phase change.

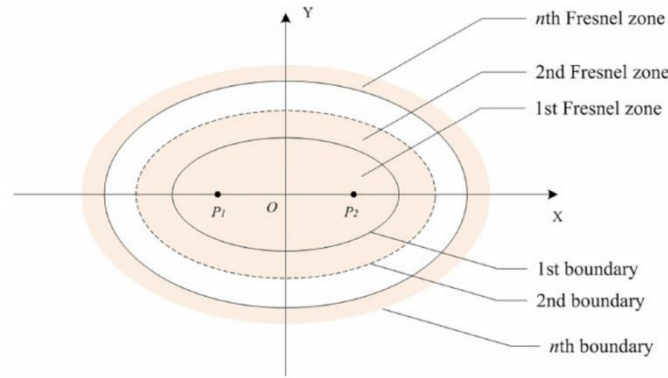


Fig. 5 The Fresnel model[3]

Y.Gu's team has built a system called WiGRUNT which uses a dual attention network, and the framework of the system is shown in Fig. 6 [4]. In this system, the spatio-temporal attention module A focuses directly on the basic cues for gesture recognition in both temporal and spatial dimensions by assigning weights to different pixels in the input image. Meanwhile, the spatio-temporal attention module B focuses indirectly on the essential information in both temporal and spatial dimensions by assigning weights to each channel in the feature map. The simultaneous use of two spatio-temporal attention modules achieves better performance than a single module. Both attention modules in WiGRUNT can provide important cues and these cues can complement each other, allowing for more accuracy in gesture recognition.

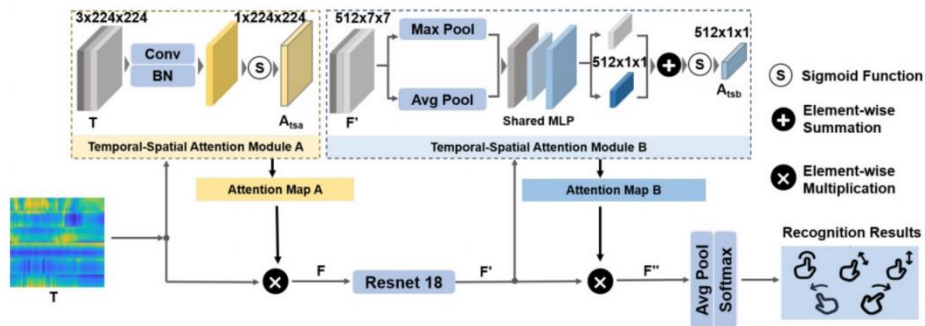


Fig. 6 WiGRUNT system architecture[4]

2.3. RFID-based Gesture Recognition

Radiofrequency identification(RFID) technology contains a variety of sensors, passive sensors have the advantages of low power consumption, low cost, small size, light weight, moderate reading distance, and no electromagnetic interference problems. Therefore, the majority of RFID-based gesture recognition systems use passive RFID systems.

In a passive RFID system, when the reader transmits RF signal to interrogate the tag, the passive tag within the reading range receives energy from the signal to respond, and the reader receives the time-varying backward scattering signal $S(t)$, $S(t)$ can be calculated as followed:

$$S(t) = x(t)e^{-j\theta(t)} \quad (4)$$

In this expression, $x(t)$ and $\theta(t)$ are the amplitude and phase of the received signal, and j is an imaginary unit.

As the user's hand moves, the propagation path of some of the signals remain steady, while others change with the movement of the hand, as shown in Fig. 7[5]. The signals which remain steady are known as static paths, and the change signals are known as dynamic paths. In Fig. 7, the dashed line is the dynamic path and the solid line is the static paths.

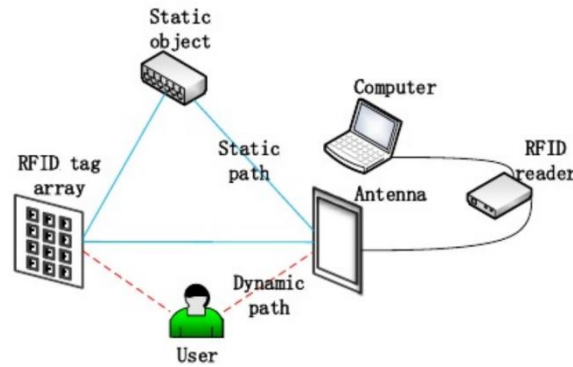


Fig. 7 RFID signal propagation diagram[5]

For a more intuitive presentation, this paper uses the complex plane to describe the signal vectors, resulting in Fig. 8[5]. S_s and S_d in the figure represent the signal vectors for static and dynamic paths.

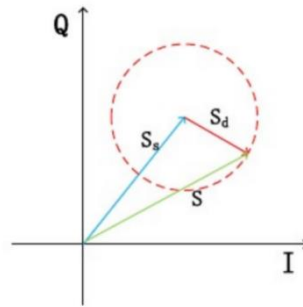


Fig. 8 Description of signals in the complex plane[5]

While the gesture is moving, S_s remains constant and the amplitude of S_d (the modulus of the S_d vector) can also be considered as a constant which unchange with time, but the phase changes a lot. The phase variation results in a rotation that is performed with the end point as the centre of the circle. In this way, the magnitude ($|S|$) and phase ($\angle S$) of the vector of the signal S , resultant from the static and dynamic path vectors, will also change with the gesture.

Thus, $S(t)$ can be expressed as:

$$S(t) = S_s(t) + S_d(t) = x_s e^{-j\theta_s} + x_d(t) e^{-j\theta_d(t)} \quad (5)$$

The changes in gestures are identified by analysing the changes in the S signal.

In 2020, a multimodal RFID data fusion was used in the RFree-GR model proposed by CaoDian and his team [5]. This model uses multimodal CNN to integrate the phase and received signal strength(RSS) streams which are reported by RFID readers. Multimodal has higher detection accuracy compared to unimodal, overcoming the perception of unimodal modality, multimodal RFID data fusion performs better and increases the accuracy by 2.37%, 2.46% and 1.88% in the test with one user left, new setup and new location respectively [5].

3. Application

3.1. Medical Treatment

In the medical field, wireless gesture-based recognition technology can be used on devices in sterile environments. Doctors and nurses can use gestures to control medical devices, such as viewing patient images and operating instruments which can avoid direct contact with the device and prevent users from infection. Miller. Elishiah team has applied radar-based gesture recognition for healthcare

workers to view medical images in 2019, with the aim of helping healthcare workers to view patient medical images aseptically to reduce the possibility of infection [6].

3.2. Smart Home

In a smart home environment, wireless gesture recognition technology can be used to control a wide range of home appliances such as lights, televisions and air conditioners. Users do not need to touch the devices but only need to make certain gestures to control them. In addition, wireless gesture recognition can be linked to other devices in the smart home system, allowing users to control multiple devices simultaneously with a simple gesture to achieve more intelligent home management. In the Liu Haipeng team's research, the millimetre wave gesture recognition was applied to the field of smart homes, which made mHomeGes are able to achieve a real-time high recognition accuracy rate of more than 95.30% in a variety of smart home situations [7][8].

3.3. Automobile

Wireless gesture recognition technology can also be used in the field of driving, the driver has noneed to leave the steering wheel, just make the corresponding gestures to complete the operation, the driver can maintain focus and reduce distraction, thus improving driving safety. In 2018, WiFi-based gesture recognition has been applied to the automobile system 'WiBot!' [9].

Wireless gesture recognition is already blended in in many aspects of our lives and will continue to expand its application in other areas.

4. Challenges and Future Expectation

4.1. Challenges

To have a more accurate result of gesture recognition, researchers not only need to collect data from different groups of people, but also need to collect the impact of different shelter and different background noise on the signal, which greatly increases the complexity of data collection. Collecting a huge amount of available data is a time-consuming and labour-intensive task, which requires a lot of time and resources for human. In addition, the variety and complexity of gesture movements means that datasets need to be constantly updated and extended to accommodate emerging gestures and user behavioural changes.

Gesture recognition systems also need to meet a high standard of real-time to ensure a smooth user experience and timely response. Real-time is a key requirement in the application of gesture recognition technology, especially in the fields of smart home, medical where delays may significantly affect the use of the device. In order to meet the real-time requirement, team must optimize the computational efficiency of the algorithm. This measure can quickly process and respond to the user's gesture input. Gesture recognition systems can achieve low-latency and highly responsive performance with its optimized software and hardware which can meet user's expectations for real-time interaction.

There are significant differences in the gesture movements of different users. Users can have different gesture habits, speeds, strengths and ranges of movement. different cultures also influenc gesture movements. These all require the recognition system to have the ability to deal with these differences in order to accurately recognise the gesture movements of different users. Algorithms can be used to allow the whole system to learn autonomously and adapt to different kinds of gesture processing with small amounts of data.

4.2. Future Expectation

Multimodal fusion technique refers to the integration of data from different types of sensors to enhance the performance of the gesture recognition system. In the future development of gesture recognition technology, the fusion of vision, sound and sensors can be used for gesture recognition

based on the original method. Multimodal fusion can provide complementary information, reduce the impact of the shortcomings of a single technology in the combination of different technologies, and provide richer and more accurate gesture features. In 2023, radar-based gesture recognition was combined with vision-based gesture recognition in a study by H. Liu, and the method effectively improved the adaptability of the classifier to complex environments and demonstrated satisfactory robustness [10].

Continuous learning algorithms refer to the ability of a system to continuously learn and adapt to new data and environmental changes during operation. This dynamic learning method allows the gesture recognition system to remain efficient in the face of new gesture movements and user behaviours. Traditional gesture recognition systems generally use pre-trained static models, so it is difficult to recognise unrecorded gestures. Online learning allows for greater flexibility and adaptability by recognising based on original recognition methods. In published research, continuous learning has been used in mobile gesture recognition and it is believed that it can be well used in wireless perception based gesture recognition [11].

Adaptive algorithm means that the system can dynamically adjust the parameters of the system according to the changes of the environment and the user to achieve more accurate recognition. In the actual application environment, compared with the laboratory environment, the application needs to deal with a variety of interference factors, such as signal strength changes, noise, user posture and distance. Adaptive algorithms can rely on the identification of environmental factors to optimize the model structure and parameter configuration, so as to improve the robustness of the system and have a more accurate recognition capability. The DANGER algorithm researched by Z. Han's team has a better recognition ability in different environments, and the algorithm parameters are adjusted for the difference of CSI signals in different environments [12].

5. Conclusion

As an emerging gesture recognition method, gesture recognition based on wireless sensing technology has shown its huge advantages and broad application prospects, making it widely noticed and recognised by the society. With the progress of the times, this type of gesture recognition technology will continually improve and gradually overcome the existing challenges. In the future, this technology will lead a new trend, promoting the emergence of more methods in the field and advancing social progress.

References

- [1] Tang G, Wu T, Li C. Dynamic Gesture Recognition Based on FMCW Millimeter Wave Radar: Review of Methodologies and Results. *Sensors*. 2023, 23(17):7478.
- [2] S. Franceschini et al., Hand Gesture Recognition via Radar Sensors and Convolutional Neural Networks, 2020 IEEE Radar Conference (RadarConf20), Florence, Italy, 2020: 1-5.
- [3] Zhengjie Wang, Zehua Huang, Chengming Zhang, Wenwen Dou, Yinjing Guo, Da Chen, CSI-based human sensing using model-based approaches: a survey, *Journal of Computational Design and Engineering*, 2021,8: 510–523.
- [4] Y. Gu et al., "WiGRUNT: WiFi-Enabled Gesture Recognition Using Dual-Attention Network," in *IEEE Transactions on Human-Machine Systems*, 2022, 52(4): 736-746.
- [5] Cao Dian, Dong Wang, Qian Zhang, Run Zhao, and Yinggang Yu. 2020. Towards Domain-independent Complex and Fine-grained Gesture Recognition with RFID. *Proc. ACM Hum.-Comput. Interact.* 2020.
- [6] Miller, Elishiah & Li, Zheng & Mentis, Helena & Park, Adrian & Zhu, Ting & Banerjee, Nilanjan. RadSense: Enabling one hand and no hands interaction for sterile manipulation of medical images using Doppler radar. *Smart Health*, 2019.

- [7] Haipeng Liu, Yuheng Wang, Anfu Zhou, Hanyue He, Wei Wang, Kunpeng Wang, Peilin Pan, Yixuan Lu, Liang Liu, and Huadong Ma. 2020. Real-time Arm Gesture Recognition in Smart Home Scenarios via Millimeter Wave Sensing. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 2020, 4: 28.
- [8] P. Wang et al., A Gesture Air-Writing Tracking Method that Uses 24 GHz SIMO Radar SoC, in *IEEE Access*, 2020, 8: 152728-152741.
- [9] Ahmed, Shahzad & Khan, Faheem & Ghaffar, Asim & Hussain, Farhan & Cho, Sung Ho. Finger-Counting-Based Gesture Recognition within Cars Using Impulse Radar with Convolutional Neural Network. *Sensors*. 2019.
- [10] H. Liu and Z. Liu, A Multimodal Dynamic Hand Gesture Recognition Based on Radar–Vision Fusion, in *IEEE Transactions on Instrumentation and Measurement*, 2023, 72: 1-15.
- [11] H. -S. Park, J. Park, D. -W. Kim and J. Lee, CGRS: Continual Gesture Recognition System, 2024 IEEE International Conference on Consumer Electronics (ICCE), Las Vegas, NV, USA, 2024: 1-6.
- [12] Z.Han, L. Guo, Z. Lu, X. Wen and W. Zheng, Deep Adaptation Networks Based Gesture Recognition using Commodity WiFi, 2020 IEEE Wireless Communications and Networking Conference (WCNC), Seoul, Korea (South), 2020: 1-7.