

A Survey of Gesture Recognition Using Frequency Modulated Continuous Wave Radar

Xinran Qiu, Junhao Liu, Lulu Song, Haofei Teng, Jiaqi Zhang, Zhengjie Wang*

College of Electronic and Information Engineering, Shandong University of Science and Technology, Qingdao, China

Email: 3448288633@qq.com, 1297618536@qq.com, 3428232283@qq.com, 923367197@qq.com, 1845168981@qq.com,

*cieewangzj@163.com

How to cite this paper: Qiu, X.R., Liu, J.H., Song, L.L., Teng, H.F., Zhang, J.Q. and Wang, Z.J. (2024) A Survey of Gesture Recognition Using Frequency Modulated Continuous Wave Radar. *Journal of Computer and Communications*, 12, 115-134. <https://doi.org/10.4236/jcc.2024.126008>

Received: May 7, 2024

Accepted: June 24, 2024

Published: June 27, 2024

Copyright © 2024 by author(s) and Scientific Research Publishing Inc. This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

Abstract

With technology advances and human requirements increasing, human-computer interaction plays an important role in our daily lives. Among these interactions, gesture-based recognition offers a natural and intuitive user experience that does not require physical contact and is becoming increasingly prevalent across various fields. Gesture recognition systems based on Frequency Modulated Continuous Wave (FMCW) millimeter-wave radar are receiving widespread attention due to their ability to operate without wearable sensors, their robustness to environmental factors, and the excellent penetrative ability of radar signals. This paper first reviews the current main gesture recognition applications. Subsequently, we introduce the system of gesture recognition based on FMCW radar and provide a general framework for gesture recognition, including gesture data acquisition, data preprocessing, and classification methods. We then discuss typical applications of gesture recognition systems and summarize the performance of these systems in terms of experimental environment, signal acquisition, signal processing, and classification methods. Specifically, we focus our study on four typical gesture recognition systems, including air-writing recognition, gesture command recognition, sign language recognition, and text input recognition. Finally, this paper addresses the challenges and unresolved problems in FMCW radar-based gesture recognition and provides insights into potential future research directions.

Keywords

Millimeter-Wave Radar, Gesture Recognition, Human-Computer Interaction, Feature Extraction

1. Introduction

In recent years, rapid advancements in wireless communication and artificial intelligence have made human-computer interaction (HCI) an indispensable technology in our daily lives [1]. Gesture recognition technology, as an important part of HCI, has seen rapid development in numerous fields and has become a hot research topic. Currently, based on the signals used, gesture recognition systems can be mainly categorized into: systems based on wearable sensors, computer vision-based systems, WIFI-based systems, and systems based on FMCW radar.

Wearable Sensor-Based Gesture Recognition Systems: These systems require users to wear data gloves connected to a computer, using sensors such as accelerometers and gyroscopes [2], to capture rich hand movement information. Liang and others developed a gesture recognition system using data gloves to assist people with hearing impairments or speech disabilities [3]. Kanokoda *et al.* [4] acquired gesture data through data gloves and used artificial neural networks for real-time gesture prediction. In 2017, Andrews *et al.* proposed a gesture recognition method based on data gloves and burst detection [5] for clinical emergency communication between patients and doctors. However, wearable devices are prone to damage, functionally limited, expensive, and require long-term wear, which are greatly inconveniences to users.

Computer Vision-Based Gesture Recognition Systems: These systems use computer imaging devices (such as cameras) to collect images, and then extract hand features from the collected video images. In 2018, MR Islam and others [6] collected gesture video frames using a single camera and used Deep Convolutional Neural Networks (DCNN) and multi-class Support Vector Machines (SVMs) to extract and recognize gesture features for 26 alphabetic signs. G. Plouffe and others [7] and Y. Li and others [8] used Microsoft Kinect sensors to acquire RGB-Depth (RGB-D) data for the classification and recognition of gestures. Plouffe *et al.* [7] employed Dynamic Time Warping (DTW) to recognize 55 static and dynamic gestures, achieving a recognition rate of 92.4%. Similarly, in [8], the authors fed RGB-D data into a C3D model to extract spatiotemporal features and utilized a Support Vector Machine classifier to output the final classification results. However, computer vision-based gesture recognition methods are susceptible to the effects of lighting and visual blind spots, and they are not conducive to protecting user privacy.

WiFi-Based Gesture Recognition Systems: These systems use WiFi signals to implement recognition of human activities [9] and gestures [10] among other applications. In 2013, Pu *et al.* designed an innovative gesture recognition system, WiSee [11], which operates by calculating the Doppler frequency shifts of WiFi signals to perceive and identify human gestures within a domestic setting. Wang and colleagues employed WiFi Channel State Information (CSI) to detect gesture-induced targets within a specified area, proposing WiDG [10], a device-free gesture recognition system based on CSI and deep learning. In this sys-

tem, the authors demonstrate the ability to classify digits 0 - 9 by analyzing the variations in CSI signals caused by different hand motions. The system achieves impressive recognition accuracies of 97.2% in through-wall scenarios and 95.3% in without-wall scenarios, respectively. However, the detection range of WiFi-based gesture recognition methods is somewhat limited, restricting their widespread application.

Gesture Recognition Systems Based on Frequency-Modulated Continuous Wave (FMCW) Radar: These systems employ radar technology to collect gesture signals, which are subsequently processed using signal processing techniques. The gesture information is finally classified and recognized through the application of machine learning or deep learning methodologies. In their 2019 study, Choi *et al.* [12] used Google's 60 GHz FMCW radar, known as Soli, to collect gesture data. They successfully identified ten different gestures using a Long Short-Term Memory (LSTM) network, achieving a remarkable accuracy rate of 99.10%. Furthermore, the system demonstrated the capability to recognize the gestures of new participants with an accuracy of 98.48%. In 2021, a study [13] introduced a gesture recognition approach using a Reusable LSTM (RLSTM) network that is based on the trajectory of range-Doppler-angle and employs a 77 GHz FMCW Multiple-Input Multiple-Output (MIMO) radar. This method achieved an average precision of 99%. The advantage of FMCW radar is that it operates effectively irrespective of line-of-sight, lighting conditions, and adverse weather scenarios such as rain, snow, or smog. These systems work under a non-contact detection pattern, thus not only enhancing user experience but also significantly protecting user privacy.

To sum up, the gesture recognition system based on FMCW radar shows obvious advantages in many aspects. Compared to wearable sensors, FMCW radar does not require the user to wear a device, improving the user experience. In contrast to computer vision, FMCW radar is not affected by lights, line of sight, and bad weather. Compared to Wifi-based systems, FMCW radar's detection range is more flexible and adaptable. Therefore, gesture recognition systems based on FMCW radar perform well in terms of performance, applicability, and user experience and protect user privacy, and have important potential for development in the field of intelligent technology and human-computer interaction.

Despite these systems achieving impressive gesture recognition performance, comprehensive reviews on gesture recognition utilizing millimeter-wave radar are still scarce. While some reviews [14] exist, their analytical perspective differs from that of this paper. This paper investigates the latest applications, presents the typical system architecture, provides the signal flows and analyzes innovative applications. Besides, this paper also first proposes application category methods from four different aspects, provides a detailed analysis, and offers some insights for future developments in the field.

This paper is composed of five sections. Section 1: The development and ap-

plication of gesture recognition technology based on different systems are described, with emphasis on the advantages of the system based on FMCW radar. Section 2: The overall framework and process of the system are described in detail, including gesture data acquisition, data preparation, classification algorithm, and gesture recognition. Section 3: Four typical gesture recognition application systems based on FMCW radar are introduced, including air writing, gesture command recognition, sign language recognition, and text input, and their performance is analyzed and summarized. Section 4: The limitations and future research directions of FMCW radar gesture recognition are pointed out. Section 5: The conclusion is presented.

2. The Structure of a Frequency-Modulated Continuous-Wave Radar Gesture Recognition System

The basic principle of gesture recognition based on FMCW radar is that the radar generates a linear frequency-modulated pulse through the radio frequency module [15]. When the transmitted signal encounters obstacles (such as hands, walls, etc.), it is reflected, and the receiver (RX) captures the echo signal. The mixer combines the transmitted and received signals, producing an intermediate frequency (IF) signal that contains the target information at a lower frequency. This IF signal can then be analyzed and processed using either dedicated digital processing circuitry integrated into the radar chip, or by feeding the signal to a PC running specialized software for more advanced signal processing and analysis tasks.

By setting the register parameters of the millimeter wave radar chip, the parameters of the FMCW transmission signal can be configured. The general framework of gesture recognition based on FMCW radar mainly includes four key stages, as shown in **Figure 1**.

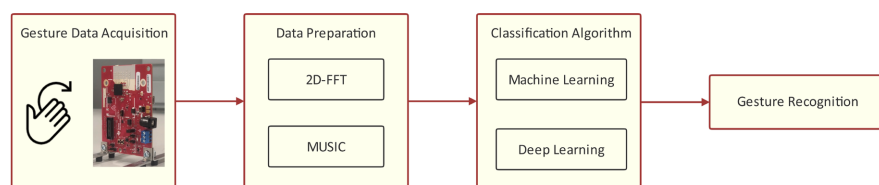


Figure 1. The architect of hand recognition of using FMCW.

1) **Gesture Data Acquisition:** In this stage, the FMCW radar system is used to collect raw data from hand gestures. The radar transmits frequency-modulated continuous waves and receives the reflected signals from the moving hand.

2) **Data Preparation:** The acquired raw data is preprocessed to remove noise, clutter, and other unwanted components. This stage may involve techniques such as filtering, signal transformation, and feature extraction to obtain a clean and informative representation of the gesture data.

3) **Classification Algorithm:** The preprocessed gesture data is fed into a classification algorithm, which is trained to recognize and differentiate between dif-

ferent types of gestures. Various machine learning algorithms can be employed for this purpose.

4) Gesture Recognition: Finally, the trained classification algorithm is used to recognize and classify new, unseen gesture data. The output of this stage is the predicted gesture label or class, which can be used to trigger corresponding actions or commands in the target application.

These four stages form the general framework for gesture recognition using FMCW radar technology, enabling the development of intuitive and contactless human-machine interaction systems.

2.1. Gesture Data Acquisition

In gesture recognition based on FMCW radar, experiments are usually done in a laboratory setting. Participants wave their hands or arms to make specific gestures in the area covered by radar signals to obtain radar echo signals and complete the collection of gesture data. Some typical systems for collecting data usually include types of gestures, feature information, data shape, and the number of samples, as shown in **Table 1**.

2.2. Signal Preparation

After collecting the echo signals, it's necessary to preprocess the received signals, which include useful FMCW radar gesture signals, to remove noise from the environment and hardware, thereby improving the accuracy of gesture recognition. Data preprocessing can use two-dimensional Fast Fourier Transform

Table 1. The typical data collection applications.

Paper	Gesture type	Feature	Data shape	Sample number
Wang [16]	Clockwise rotation, counterclockwise rotation, palm moving up and down, sliding left, sliding right	distance, speed, phase	100*3	1500
Zhang [17]	Sliding left, sliding right, pulling, pushing, knocking, waving, swinging up and down, patting	spatiotemporal features	8 sub-spectrograms, 64*64*8	3200
Li [18]	Sliding left and right, waving, and other 5 types of gestures	distance, Doppler, angle	---	3500
Zheng [13]	Sliding left, sliding right, moving up and down, clockwise rotation, and other 5 types of gestures	Distance, Doppler-angle, trajectory	32*32	4500
Wang [19]	Pushing, pulling, clockwise rotation, counterclockwise rotation, and other 6 types of gestures; distance and speed	RT and DT sequences	256*8192	4000
Zhang [20]	Sliding left and right, pushing, pulling, tapping, and other 4 types of gestures	Distance, speed	256*256	3200

(2D-FFT) and Multi-Signal Classification (MUSIC) algorithms to remove interference information and extract gesture features to the maximum extent to obtain the distance, speed, angle, and other features of target gestures.

1) 2D-FFT:

2D-FFT is a common preprocessing step in radar signal processing for gesture recognition. It transforms signals from the time domain to the frequency domain to extract frequency domain features of the signal. Usually, the intermediate frequency (IF) signals undergo processing via 2D-FFT. Leveraging the connection between distance, speed, and the frequency of the IF signal, a Range-Doppler Map (RDM) is derived. RDM offers abundant feature information, including trajectory, distance, speed, and shape variations of hand movements. In 2018, Ryu *et al.* [21] used 2D-FFT to obtain RDM images containing gesture distance and speed information to achieve gesture classification and recognition. In 2021, a gesture recognition system [12] was proposed that uses 2D-FFT to generate RDM maps, employs Constant False Alarm Rate (CFAR) detection for gestures, and inputs the RDM sequence from the detection region into an LSTM to extract temporal features for gesture classification and recognition.

2) MUSIC Algorithm:

The MUSIC algorithm provides high resolution, estimation accuracy, and stability under specific conditions. Its basic idea is to perform eigen decomposition on the covariance matrix of the output data from an arbitrary array to obtain the signal subspace corresponding to signal classification and the noise subspace orthogonal to the signal components. Then, by exploiting the orthogonality of these two subspaces, a spatial spectrum function is constructed. Through peak searching in the spectrum, the Direction Of Arrival (DOA) of the gesture target is obtained. Subsequently, the angle estimation spectrum of the gesture target is explored. By accumulating multiple frames of the angle estimation spectrum, an Angle-Time Map (ATM) is generated, furnishing angle information regarding the gesture target. In the literature [22], the authors utilized 2D-FFT to extract the original gesture data's distance and Doppler parameters, while employing the MUSIC algorithm to compute the angle and construct the ATM. They utilized the Fusion Dynamic Time Warping (FDTW) algorithm for gesture recognition. In 2021, Zheng *et al.* used the Discrete Fourier Transform (DFT), MUSIC algorithm, and Kalman filtering to extract distance-Doppler-angle trajectories and proposed an LSTM network that reuses the forward propagation method to recognize gestures, achieving an average accuracy of 99.4% [13].

2.3. Gesture Recognition Algorithm

For the preprocessed gesture data, feature extraction is performed, and the extracted features are used for gesture classification or recognition. The classification and recognition algorithm is an important step in radar gesture recognition research. Contemporary machine learning techniques excel not only in data processing, target classification, and model prediction but also hold great prom-

ise for advancing gesture recognition. Presently, widely adopted gesture recognition algorithms encompass both machine learning and deep learning methodologies.

1) Gesture recognition algorithms based on machine learning

Machine learning, which uses computers and networks as platforms, data as research objects, and algorithms as the core, aims to predict and analyze data. Among the radar gesture recognition algorithms based on machine learning, the more widely used ones are Dynamic Time Warping, Hidden Markov Models, K-Nearest Neighbors (KNN), and Random Forest.

a) Dynamic Time Warping (DTW)

Dynamic Time Warping (DTW) adopts the idea of dynamic programming and uses a warping function to calculate the temporal similarity between test data and reference templates, thereby obtaining the similarity between two time-related sequences. When processing radar data with DTW, it is first necessary to construct a set of reference templates. By comparing the similarity between test data and reference templates, the gesture data with the smallest difference is computed as the output result. In the paper [22], Wang *et al.* used the AWR1642 millimeter-wave radar and TSW140 high-speed data acquisition card to capture gesture signals. The Fusion Dynamic Time Warping (FDTW) algorithm was employed, achieving a recognition rate of 95.83% for six types of gestures within a range of 10-70 centimeters from the radar. In gesture recognition systems, DTW achieved fewer training samples and has high recognition accuracy, but it has high computational complexity and poor stability.

b) Hidden Markov Model (HMM)

The Hidden Markov Model (HMM) is a probabilistic model over time that describes a sequence of unobservable states randomly generated by a hidden Markov chain. When using HMM for FMCW radar gesture recognition, it is necessary to construct an HMM model for each gesture separately. The probability of the test gesture for each HMM model is calculated, and the gesture corresponding to the HMM model with the highest probability is taken as the model output result. Greg and others proposed a gesture recognition system combining Range-Doppler and Hidden Markov Model (HMM) based on 77 GHz FMCW radar [23], which achieved a recognition rate of 83.3% for four types of gestures. This method can effectively improve recognition accuracy, but the large amount of computation during algorithm execution limits the recognition speed.

c) K-Nearest Neighbors (KNN) Classifier

The KNN algorithm is a basic method for classification and regression. Based on a training dataset of gestures, for a new input gesture test instance, the algorithm finds the k instances in the training dataset that are closest to the test instance. The class that has the majority among these k instances is then predicted as the class for the new input gesture instance. For instance, Wan *et al.* proposed a gesture recognition system utilizing a portable smart radar [24]. The system

utilizes Principal Component Analysis (PCA) to extract spatial features and analyze the time-frequency characteristics of radar signals. By using amplitude difference and Doppler frequency shift calculations and experiments, the KNN classification algorithm achieved a classification accuracy of over 95%. The KNN method is simple and easy to understand, but it requires large storage space and has high time complexity.

d) Random Forest

The basic unit of a random forest is a decision tree, which is an algorithm that integrates multiple decision trees through the concept of ensemble learning. To explain from the perspective of gesture classification and recognition, each decision tree acts as a classifier. For an input gesture sample, N trees will produce N classification results. The random forest integrates all the classification votes and designates the category with the most votes as the final output result. Researchers like Lien [25] conducted gesture recognition experiments using millimeter-wave radar and implemented a random forest classifier to recognize four types of gestures with an accuracy of 92.1%. In gesture recognition, random forests are fast to train, can handle high-dimensional data, and are simple to implement, but they can overfit, and the classification results may not be as good with small or low-dimensional datasets.

2) Gesture recognition algorithms based on deep learning

Deep learning algorithms primarily involve training deep neural network models to classify data. These models consist of an input layer, multiple hidden layers, and an output layer. Compared to traditional neural network models, deep neural networks contain multiple hidden layers composed of a large number of simple, interconnected neurons. In radar gesture recognition, the algorithms related to deep learning mainly include Convolutional Neural Networks (CNNs) and Long Short-Term Memory networks (LSTMs).

a) CNN

CNNs are the most common neural network models with deep structures and convolutional computations. They train model parameters using backpropagation and use the SoftMax function for classification. In the literature [19], Wang *et al.* combined a 60GHz FMCW radar with convolutional and recurrent neural networks to achieve gesture classification, with an 87% recognition rate for 11 different gestures from 10 participants. Dekker *et al.* [26] collected gesture information using a 24 GHz FMCW radar and used a CNN for feature extraction. The network was trained and tested using micro-Doppler spectrograms, achieving a 99% recognition rate for three different gestures on the test set. In 2020, Liu proposed an end-to-end time-series convolutional neural network (TS-CNN) method for gesture recognition based on FMCW radar signals, with a classification accuracy of 93% for seven gestures [27]. CNNs have strong generalization capabilities, can automatically extract features, and share weights, making them widely applied in various fields such as behavior recognition and pose estimation.

b) LSTM

Compared to CNNs, Recurrent Neural Networks (RNNs) have a stronger ability to capture the sequential information of series, yielding more accurate results. RNNs are recursive neural networks used for processing sequence data and recursively evolving in the direction of the sequence. The most widely used type of RNN is the Long Short-Term Memory network (LSTM). In the paper [28], the authors proposed a Long Recurrent All-Convolutional Network (LRACN) to recognize five types of gestures. In the article [12], the authors used an adaptive background model based on Gaussian Mixture Models (GMM) to remove noise, and then LSTM was used to realize real-time gesture recognition. Compared to traditional RNNs, LSTMs are not limited to fixed-length inputs or outputs; they can also handle data of varying sequence lengths, and all network units are connected in a chain with powerful memory capabilities.

3. Typical Applications

In this section, we present several typical application systems for gesture recognition based on Frequency-Modulated Continuous-Wave (FMCW) radar and analyze their system performance from multiple perspectives. We categorize these studies based on the objectives of the research as follows. Air writing primarily focuses on the trajectory changes and tracking of gestures to form letters or words. Gesture command recognition mainly concentrates on the variations of gestures, enabling the recognition of different gestures to achieve the purpose of controlling devices. Sign language recognition is primarily aimed at recognizing gestures as sign language, with the research content and scope being determined by the sign language representations. Text input considers the input of text or letters, and due to the diversity of text input, it typically encompasses a wide variety of input types.

3.1. Air-Writing Tracking

The application primarily aims to implement air writing tracking, with the main research focus on how to obtain continuous hand movement positions to derive a continuous trajectory, thereby forming continuous text.

Regani *et al.* proposed the mmWrite system [29], which utilizes a 60 GHz signal large phased array to transform any planar area into an interactive writing surface, supporting millimeter-precision handwriting tracking. The system comprises several components, as shown in **Figure 2**. Initially, background subtraction is performed to reduce the impact of static objects in the environment on the received signal. Subsequently, digital beamforming technology is employed to obtain spatial features, which are then transformed into the Doppler domain to differentiate moving/writing actions from other static objects in the environment. Target detection is achieved using 3D-CFAR and clutter mapping techniques, with further identification of targets within the detected space. Finally, the calculated trajectory points are compared with the original movement trajectory, and the handwriting trajectory is obtained through smoothing techniques.

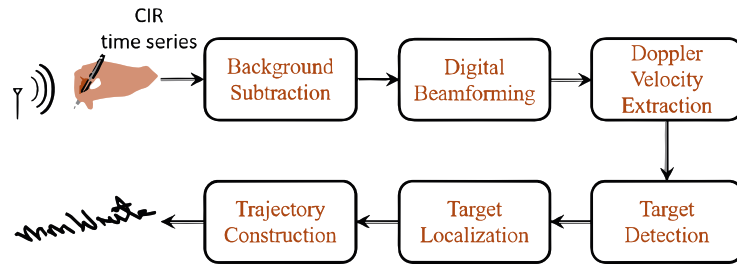


Figure 2. The component of mmWrite [29].

The mmWrite can achieve a median error tracking of 2.8 mm for fingers/pens, thus enabling the completion of handwritten characters of 1 cm × 1 cm. The system can provide a convenient and high-precision handwriting tracking system for the field of human-computer interaction.

Regani *et al.* [29] tested the character recognition accuracy when characters of different distances and proportions were written. They found that the accuracy of character recognition decreased as the distance from the device increased, while the accuracy increased as the handwriting grew larger. For example, at distances of 20 cm and 30 cm, the accuracy of character recognition at 3 cm × 3 cm is 80% and 72%, respectively. Similarly, when the distance is 20 cm, the accuracy of 3 cm × 3 cm and 5 cm × 5 cm is 80% and 82%, respectively. In addition, the mmWrite system has not been extended to multiple moving targets and needs further study.

In 2020, Wang *et al.* [16] proposed a gesture air-writing recognition system based on a 24 GHz FMCW radar. The basic concept of the system is to calculate the range-Doppler graph (RDM) from the radar echo signal and extract the range, velocity, and phase features from the RDM of each frame. The authors stored the extracted feature information in a 100 × 3 vector as a sample and collected 1500 samples for five types of gestures, of which 1200 samples were used for the training set and 300 for the test set, employing an LSTM network to distinguish five gestures. During the testing phase, extensive experiments were conducted to evaluate the system's performance. The experiments demonstrated that the system could recognize five types of gestures with an average accuracy of 97.6%. Additionally, the authors recognize 10 digital numbers and 9 alphabetic characters.

Wang *et al.* [16] tested the classification effect of three different classification algorithms, SVM, 5NN and LSTM, on different gestures. The results show that compared with the SVM and 5NN algorithm, the LSTM network shows a higher average recognition rate and smaller variance of recognition rate, indicating that gestures are easier to recognize by the LSTM network. However, despite the excellent performance of LSTM networks in gesture recognition, the characters generated by this system cannot be accurately recognized. Therefore, future research will focus on improving recognition algorithms for generated characters.

3.2. Gesture Command Recognition

The primary objective of gesture recognition is to categorize the direction and velocity of hand movements, thereby enabling human-computer interaction and application control.

Liu *et al.* proposed the M-Gesture gesture recognition system [30], which facilitates control of media players and cameras, as depicted in **Figure 3**. The system is comprised of four main components: (a) signal conversion, (b) gesture modeling, (c) gesture recognition, and (d) system response. This system integrates a pseudo-representative model (PRM) with a custom neural model to characterize and extract intrinsic gesture features, employing a System Status Transition (SST) to filter out non-predefined motions. The system is capable of recognizing five routine gestures, achieving high-precision recognition for untrained new users on a limited dataset, with an impressive accuracy of 99% and a rapid response time within 25 milliseconds.

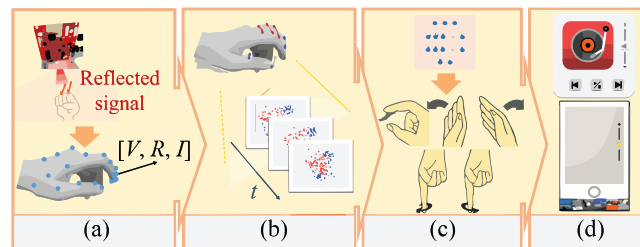


Figure 3. The component of M-Gesture [30].

In the context of remotely controllable household devices, Liu *et al.* [30] found that when extending the detection range of m-Gesture from the individual domain (about 0.5 m) to longer distances (greater than 3 m), extracting long-distance gestures becomes more challenging because these gestures are more easily confused with other body movements (such as walking). Therefore, future research needs to focus on the development of in-air gesture HCI interfaces applicable to more scenarios.

Zhang and colleagues introduced a novel hand gesture recognition (HGR) system, Riddle [31], which employs millimeter-wave radar for real-time hand depiction and human-computer interaction. The central premise is that when the millimeter-wave radar sensor captures hand motion, distance information is observable in the spectral domain. By integrating deep neural networks with Connectionist Temporal Classification (CTC) algorithms, the system achieves real-time recognition of diverse hand gestures. Through training, the deep neural network is capable of effectively extracting gesture features and class boundaries. Moreover, the research team developed an architecture that combines 3D Convolutional Neural Networks (3D-CNN) with LSTM networks. The 3D-CNN is utilized for short-term spatio-temporal modeling, the LSTM for global temporal feature extraction, and the CTC layer for real-time gesture classification. Additionally, the CTC algorithm allows for the recognition of zero-latency or

negative-latency gestures from unsegmented input data. This network architecture is illustrated in **Figure 4**. In the experiments, the researchers designed six types of gestures and invited four volunteers to repeat these gestures 100 times in front of the radar antenna, collecting data for each gesture within one minute. Riddle was compared with state-of-the-art methods such as Support Vector Nucleus (SVN), Convolutional 3D (C3D), and Hidden Markov Model (HMM)-based approaches, outperforming them with a recognition accuracy of 96%.

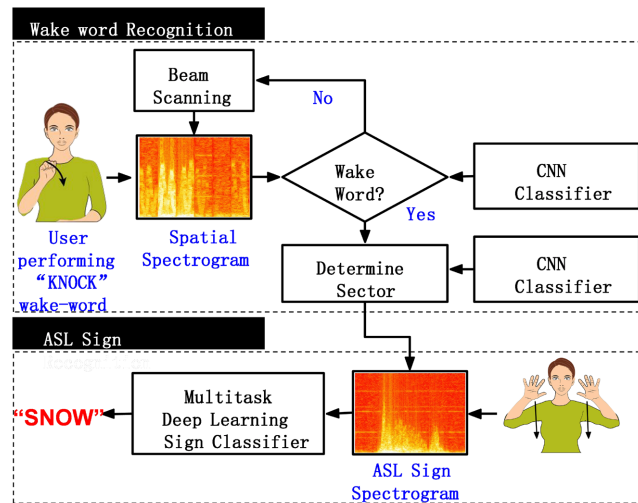


Figure 4. The component of mmASL [33].

Zhang *et al.* [31] compared different training architectures of 2D-CNN and 3D-CNN on radar data and found that 3D-CNN performed better. This shows that 3D-CNN can extract spatio-temporal information more effectively, especially for Human Gesture Recognition (HGR). However, their system employs a more complex 3D-CNN and LSTM network structure. Future research can explore learning the spatio-temporal features of dynamic gestures through a single network and integrating 3D operations into the LSTM network to simplify the model and improve performance.

3.3. Sign Language Recognition

Santhalingam *et al.* introduced the mmASL system [33], which utilizes a 60 GHz millimeter-wave platform for sign language recognition, as depicted in **Figure 4**. The system comprises two main components: (1) wake word recognition and (2) ASL symbol recognition. mmASL operates by continuously scanning a set of predefined areas to create spatial spectrograms, which are used to detect the presence of a wake word based on a CNN-based model. Upon detecting a wake word, a CNN-based classifier determines the user's current position. Subsequently, ASL gestures are captured in the form of ASL symbol spectrograms and recognized using a multi-task deep learning model, culminating in the generation of the final results. The system was evaluated using data from 65 participants and over 12K samples, demonstrating robust resistance to interference and

insensitivity to positional and environmental changes. Comparing mmASL with ASL recognition systems based on Kinect and RGB cameras, it achieves comparable performance, with an average recognition accuracy of 87%.

Santhalingam *et al.* used WGAN and MBGAN to train CAE models and test them with synthetic data. On local ASL samples, the classification accuracy of WGAN and MBGAN was 74.28% and 80.82%, respectively. By fine-tuning 30% of the original sample, the performance improved to 86% and 91.30%. However, the performance was poor when the fine-tuning sample was small, showing that there was still a significant difference between the local sample and the synthetic local sample even when the data was transformed using CycleGAN. Only fine-tuning with more local ASL samples will improve performance.

Rahman [32] *et al.* explored the efficacy of radio frequency sensors in supporting human-computer interaction for deaf or hard-of-hearing individuals through word-level recognition of American Sign Language (ASL). They proposed a scheme for ASL recognition using Wi-Fi and low-cost, short-range radar systems. The study utilized RF data of ASL (ASL-R) acquired by the TI AWR1642BOOST 77 GHz FMCW transceiver, measuring the native sign language of deaf individuals or Children of Deaf Adults (CODAs) who are fluent in ASL, as well as the mimicked sign language from hearing participants. The researchers collected data under different bandwidths and found that operating the TI 77 GHz FMCW transceiver with a bandwidth of 4 GHz and 255 chirps per CPI was the optimal setting for capturing ASL. A major challenge of the system was training deep neural networks due to the difficulty in acquiring native ASL sign language data. The authors employed adversarial domain adaptation techniques to bridge the physical/kinematic differences between the mimicked sign language of hearing individuals (repeating gestures after watching videos) and the proficient sign language of deaf signers. They compared the results of domain adaptation with those obtained directly using Generative Adversarial Networks (GANs) to synthesize ASL sign language. Ultimately, they achieved a word-level classification accuracy of 91.3% for 20 ASL words.

Rahman *et al.* [32] compared the performance of mmASL with Kinect and RGB cameras, and the results show that mmASL can achieve accurate gesture recognition in a variety of practical scenarios such as the presence of other interfering users, environmental changes, and different user positions. However, in classification, mmASL cannot accurately distinguish between gestures with similar hand movements but different hand shapes (e.g., NIGHT and TIME).

3.4. Text Input

Hu *et al.* proposed mmKey [34], a system designed to implement a universal virtual keyboard interface, as illustrated in **Figure 5**. This system employs a signal processing pipeline for the detection, segmentation, and recognition of both single and multi-finger keystrokes. Specifically, the initial stage involves motion detection. This is followed by motion distinction, which serves to discern

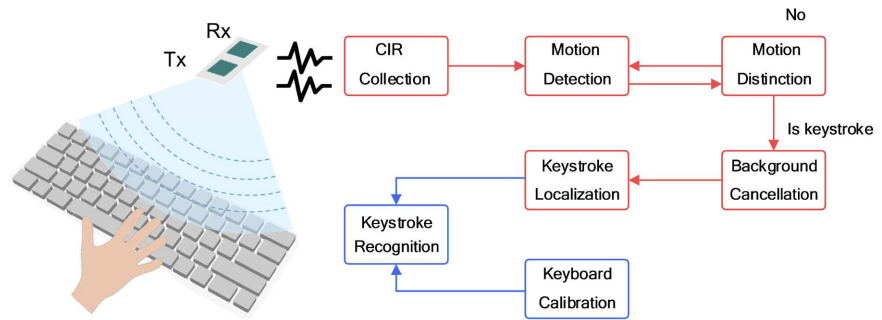


Figure 5. The component of mmKey [34].

whether the keystrokes originate from extraneous movements. Subsequently, adaptive background cancellation is executed. The final stage encompasses keystroke localization and calibration, facilitating key-location mapping for accurate keystroke recognition. The system is capable of transforming any flat surface into an effective input medium, supports various keyboard layouts, and requires no training. Experimental results from ten participants demonstrated that the system achieves a single-key recognition accuracy of over 95%, and a multi-key recognition accuracy of over 90%, leading to a word recognition accuracy of greater than 97%.

Hu *et al.* [34] contrasted mmKey with conventional beamforming techniques, including conventional beamforming (CBF) and minimum variance distortion response (MVDR) beamforming. It was found that only the MUSIC algorithm was able to detect two adjacent sources, while MVDR or CBF had difficulty distinguishing multiple keystrokes. After applying the three estimators in different scenarios, the results show that MUSIC-based mmKey exhibits the highest accuracy in all cases, especially in the case of double keys, with an overall accuracy of more than 90%. However, the mmKey system requires a single mmwave radio device, which may limit its popularity and application range.

Wei *et al.* proposed the IndexPen [35], a text input system operated by two fingers. The basic working principle of the system involves measuring the minute angular and velocity changes produced by the movements of two fingers. Utilizing a neural network, the system identifies letters and gestures to facilitate content input. The IndexPen is capable of accurately recognizing 30 distinct gestures, representing the letters A-Z, as well as space, backspace, enter, and a special activation gesture designed to prevent inadvertent inputs. The system comprises a radio frequency (RF) processing pipeline, a classification model, and a real-time detection algorithm. It was trained using data collected over more than ten days, with five participants achieving a cross-validation accuracy of 95.89% across 31 categories (including noise). The accuracy for users typing sentences with IndexPen was 86.2%, measured by string similarity.

Wei *et al.* [35] conducted experiments on character gestures and users to evaluate the overall validation accuracy of IndexPen and found that the best accuracy without clutter removal was only 3.4% lower than the model with clutter

removal. However, the training process is more stable after clutter removal. In addition, IndexPen provides an innovative method of touch-free text entry, and it may take some time for users to adapt to this new way of interaction, so there is a certain learning curve.

3.5. Summary

Herein, we provide a concise summary of several prevalent applications as depicted in **Table 2**. We present the objectives and key elements of four distinct applications.

Table 2. Key features of four typical application categories.

Paper	Application type	Aim	Highlight
[16] [29]	Air-Writing Tracking	Recognize specific number and alphabet	Track hand location
[30] [31]	Gesture Recognition	Control device or application	Recognize hand motion direction and speed
[32] [33]	Sign Language Recognition	Communication with the other person	Recognize sign symbol
[34] [35]	Text Input	Input data to device	Recognize text and sentence

The Air-Writing Tracking application predominantly involves gestural movements in mid-air, capable of recognizing a limited set of letters or numbers. Its essence lies in the continuous tracking of hand positions to ascertain trajectories, which, based on their shapes, facilitate the identification of characters or numerals.

The Gesture Command Recognition application typically discerns the direction and velocity of hand movements to manipulate devices or software. The key to this application is a predefined set of gestures, making it a categorical application where recognition of specific velocities and directions suffices.

The Sign Language Recognition application aims to facilitate communication. Its core challenge is the depiction of everyday communicative acts through gestural motions. Given the complexity of daily communication, sign language recognition applications have a broader scope and inherently higher difficulty and complexity.

The Text Input application focuses on the input of textual information. Its key is the generation of input content, which may range from specific numeric texts to more intricate material. Consequently, the implementation difficulty of these applications varies, primarily determined by the complexity of the content to be inputted.

4. Discussion

This section provides a summary of key challenges and corresponding solutions

encountered in FMCW radar-based gesture recognition applications. The challenges in FMCW radar gesture recognition primarily encompass the following aspects.

4.1. Complex Environments

Current research experiments are conducted in relatively simple scenarios where the FMCW radar performs gesture detection with few interfering objects, meaning the background environment is ideal. For instance, in Latern [17] and u-DeepHand [36], participants were instructed to maintain stillness during the experiments. However, in real-world applications, the detected gesture signal samples are frequently contaminated with various types of noise due to environmental uncertainties. Moreover, as the authors mentioned in the article [36], the issue of recognizing multiple gestures, and current scenarios is mostly for single users [13]. The recognition of radar gestures in multi-user scenarios not only needs to consider the extraction of different gesture features and the construction of gesture models but also needs to respond quickly and accurately to different participants in the same scene. Therefore, future research work should consider setting up more complex experimental scenarios [36] to ensure that the gesture recognition system maintains good performance in more complex application scenarios.

4.2. Micro-Motion Gestures

In the existing experimental environment, researchers often define a larger range of gestures to obtain more detailed motion information for better experimental results. In literature [16], the authors defined the range of gesture writing within a rectangle 20 cm long and 15 cm wide. Due to the bandwidth limitations of radar equipment, the accuracy of gesture recognition decreases as the range of gesture activity becomes smaller. For instance, in a study [36], the impact of gesture scale on classification accuracy was investigated by setting dimensions of 0.5 m, 0.3 m, and 0.1 m, resulting in accuracies of 95%, 87%, and 76%, respectively. Micro-motion gestures are characterized by small movement amplitudes and slow speeds. Reducing the motion scale is equivalent to decreasing the radial movement distance. When the radial movement distance falls below the distance resolution, the gesture trajectory cannot be recognized. In the future, micro-motion gestures could be identified by employing radar platforms with higher bandwidth and distinguishing features.

4.3. Neural Networks

The algorithms for radar gesture recognition have high requirements for the quality and quantity of training samples. Effective sample data is the foundation and prerequisite for enhancing the accuracy of gesture recognition. For example, in their experiments, Li *et al.* [18] collected 3500 samples, while Wang *et al.* [22] gathered 1200 samples. Although the classification algorithms they chose achieved

good results, it has not been verified whether these algorithms will still perform well with the increased complexity of the environment and the number of dataset samples. As the volume of data samples increases, the requirements for neural network models also increase. Therefore, optimizing neural network models that are robust and have shorter training times becomes a priority for gesture recognition systems. In the future, consideration should be given to designing lightweight models, reducing model depth, and optimizing model parameters to enhance system robustness.

5. Conclusion

This paper aims to provide a comprehensive review of the current state of FMCW radar-based gesture recognition applications. First, we reviewed some common gesture recognition systems and categorized them into wearable device-based, computer vision-based, WIFI-based, and FMCW radar-based systems. Next, we introduced related work in radar-based gesture recognition. We then focused on FMCW radar-based gesture recognition systems, providing a general framework for gesture recognition. After that, we discussed some processing techniques used in gesture recognition, including gesture signal collection, gesture echo signal preprocessing, feature extraction, and gesture recognition algorithms. Subsequently, we surveyed recent research on FMCW radar-based gesture recognition, dividing it into four typical systems: air-writing recognition, gesture command recognition, sign language recognition, and text input recognition, and explained the implementation process and performance evaluation of various gesture recognition systems. Finally, we discussed the limitations and existing issues in FMCW radar-based gesture recognition, considered current research trends, and provided insights for future research directions. This paper can serve as a reference for researchers' subsequent studies and assist other authors in developing more suitable gesture recognition applications.

Fund

The work is funded by the foundation of the Innovation and Entrepreneurship Training Program for College Students (202310424202) and Exploration and Practice of Programming Course Teaching Models for Electronic Information Engineering Major in the Context of New Engineering Education (QX2022M39).

Conflicts of Interest

The authors declare no conflicts of interest regarding the publication of this paper.

References

- [1] Zhou, Y., Jiang, G. and Lin, Y. (2016) A Novel Finger and Hand Pose Estimation Technique for Real-Time Hand Gesture Recognition. *Pattern Recognition*, **49**, 102-114. <https://doi.org/10.1016/j.patcog.2015.07.014>
- [2] Gupta, H.P., Chudgar, H.S., Mukherjee, S., Dutta, T. and Sharma, K. (2016) A Con-

- tinuous Hand Gestures Recognition Technique for Human-Machine Interaction Using Accelerometer and Gyroscope Sensors. *IEEE Sensors Journal*, **16**, 6425-6432. <https://doi.org/10.1109/jsen.2016.2581023>
- [3] Liang, R. and Ouhyoung, M. (1995) A Real-Time Continuous Alphabetic Sign Language to Speech Conversion VR System. *Computer Graphics Forum*, **14**, 67-76. https://doi.org/10.1111/j.1467-8659.1995.cgf143_0067.x
- [4] Kanokoda, T., Kushitani, Y., Shimada, M. and Shirakashi, J. (2019) Gesture Prediction Using Wearable Sensing Systems with Neural Networks for Temporal Data Analysis. *Sensors*, **19**, Article 710. <https://doi.org/10.3390/s19030710>
- [5] Andrews, S. and Karthik, S. (2017) Burst Detection in Gesture Recognition for Emergency Communication Using Wearable Data Gloves. *Proceedings of the 2017 Fourth HCT Information Technology Trends (ITT)*, Al Ain, 25-26 October 2017, 9-12. <https://doi.org/10.1109/ctit.2017.8259558>
- [6] Islam, M.R., Mitu, U.K., Bhuiyan, R.A. and Shin, J. (2018) Hand Gesture Feature Extraction Using Deep Convolutional Neural Network for Recognizing American Sign Language. *Proceedings of the 2018 4th International Conference on Frontiers of Signal Processing (ICFSP)*, Poitiers, 24-27 September 2018, 115-119. <https://doi.org/10.1109/icfsp.2018.8552044>
- [7] Plouffe, G. and Cretu, A.-M. (2016) Static and Dynamic Hand Gesture Recognition in Depth Data Using Dynamic Time Warping. *IEEE Transactions on Instrumentation and Measurement*, **65**, 305-316. <https://doi.org/10.1109/tim.2015.2498560>
- [8] Li, Y., Miao, Q., Tian, K., Fan, Y., Xu, X., Li, R., et al. (2018) Large-Scale Gesture Recognition with a Fusion of RGB-D Data Based on Saliency Theory and C3D Model. *IEEE Transactions on Circuits and Systems for Video Technology*, **28**, 2956-2964. <https://doi.org/10.1109/tcsvt.2017.2749509>
- [9] Wang, Z., Jiang, K., Hou, Y., Dou, W., Zhang, C., Huang, Z., et al. (2019) A Survey on Human Behavior Recognition Using Channel State Information. *IEEE Access*, **7**, 155986-156024. <https://doi.org/10.1109/access.2019.2949123>
- [10] Wang, Z., Song, X., Fan, J., Chen, F., Zhou, N., Guo, Y., et al. (2021) WiDG: An Air Hand Gesture Recognition System Based on CSI and Deep Learning. *Proceedings of the 2021 33rd Chinese Control and Decision Conference (CCDC)*, Kunming, 22-24 May 2021, 1243-1248. <https://doi.org/10.1109/ccdc52312.2021.9602438>
- [11] Pu, Q., Gupta, S., Gollakota, S. and Patel, S. (2013) Whole-Home Gesture Recognition Using Wireless Signals. *Proceedings of the 19th Annual International Conference on Mobile Computing & Networking*, Miami, 30 September-4 October 2013, 27-38. <https://doi.org/10.1145/2500423.2500436>
- [12] Choi, J.-W., Ryu, S.-J. and Kim, J.-H. (2019) Short-Range Radar Based Real-Time Hand Gesture Recognition Using LSTM Encoder. *IEEE Access*, **7**, 33610-33618. <https://doi.org/10.1109/access.2019.2903586>
- [13] Yang, Z. and Zheng, X. (2021) Hand Gesture Recognition Based on Trajectories Features and Computation-Efficient Reused LSTM Network. *IEEE Sensors Journal*, **21**, 16945-16960. <https://doi.org/10.1109/jsen.2021.3079564>
- [14] Wang, Z., Liu, F., Li, X., Ma, M., Feng, X. and Guo, Y. (2022) A Survey of Hand Gesture Recognition Based on FMCW Radar. *Proceedings of the 8th International Conference on Communication and Information Processing*, Beijing, 3-5 November 2022, 73-79. <https://doi.org/10.1145/3571662.3571674>
- [15] Ullmann, I., Guendel, R.G., Kruse, N.C., Fioranelli, F. and Yarovoy, A. (2023) A Survey on Radar-Based Continuous Human Activity Recognition. *IEEE Journal of Microwaves*, **3**, 938-950. <https://doi.org/10.1109/jmw.2023.3264494>

- [16] Wang, P., Lin, J., Wang, F., Xiu, J., Lin, Y., Yan, N., et al. (2020) A Gesture Air-Writing Tracking Method that Uses 24 GHz SIMO Radar SoC. *IEEE Access*, **8**, 152728-152741. <https://doi.org/10.1109/access.2020.3017869>
- [17] Zhang, Z., Tian, Z. and Zhou, M. (2018) Latern: Dynamic Continuous Hand Gesture Recognition Using FMCW Radar Sensor. *IEEE Sensors Journal*, **18**, 3278-3289. <https://doi.org/10.1109/jsen.2018.2808688>
- [18] Li, G., Zhang, R., Ritchie, M. and Griffiths, H. (2018) Sparsity-Driven Micro-Doppler Feature Extraction for Dynamic Hand Gesture Recognition. *IEEE Transactions on Aerospace and Electronic Systems*, **54**, 655-665. <https://doi.org/10.1109/taes.2017.2761229>
- [19] Wang, S., Song, J., Lien, J., Poupyrev, I. and Hilliges, O. (2016) Interacting with Soli: Exploring Fine-Grained Dynamic Gesture Recognition in the Radio-Frequency Spectrum. *Proceedings of the 29th Annual Symposium on User Interface Software and Technology*, Tokyo, 16-19 October 2016, 851-860. <https://doi.org/10.1145/2984511.2984565>
- [20] Wang, Y., Wang, S., Zhou, M., Jiang, Q. and Tian, Z. (2019) TS-I3D Based Hand Gesture Recognition Method with Radar Sensor. *IEEE Access*, **7**, 22902-22913. <https://doi.org/10.1109/access.2019.2897060>
- [21] Ryu, S.-J., Suh, J.-S., Baek, S.-H., Hong, S. and Kim, J.-H. (2018) Feature-Based Hand Gesture Recognition Using an FMCW Radar and Its Temporal Feature Analysis. *IEEE Sensors Journal*, **18**, 7593-7602. <https://doi.org/10.1109/jsen.2018.2859815>
- [22] Wang, Y., Ren, A., Zhou, M., Wang, W. and Yang, X. (2020) A Novel Detection and Recognition Method for Continuous Hand Gesture Using FMCW Radar. *IEEE Access*, **8**, 167264-167275. <https://doi.org/10.1109/access.2020.3023187>
- [23] Malysa, G., Wang, D., Netsch, L. and Ali, M. (2016) Hidden Markov Model-Based Gesture Recognition with FMCW Radar. *Proceedings of the 2016 IEEE Global Conference on Signal and Information Processing (GlobalSIP)*, Washington DC, 7-9 December 2016, 1017-1021. <https://doi.org/10.1109/globalsip.2016.7905995>
- [24] Wan, Q., Li, Y., Li, C. and Pal, R. (2014) Gesture Recognition for Smart Home Applications Using Portable Radar Sensors. *Proceedings of the 2014 36th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, Chicago, 26-30 August 2014, 6414-6417.
- [25] Lien, J., Gillian, N., Karagozler, M.E., Amihoud, P., Schwesig, C., Olson, E., et al. (2016) Soli: Ubiquitous Gesture Sensing with Millimeter Wave Radar. *ACM Transactions on Graphics*, **35**, 1-19. <https://doi.org/10.1145/2897824.2925953>
- [26] Dekker, B., Jacobs, S., Kossen, A.S., Kruihof, M.C., Huizing, A.G. and Geurts, M. (2017) Gesture Recognition with a Low Power FMCW Radar and a Deep Convolutional Neural Network. *Proceedings of the 2017 European Radar Conference (EURAD)*, Nuremberg, 11-13 October 2017, 163-166. <https://doi.org/10.23919/eurad.2017.8249172>
- [27] Liu, D. and Meng, H. (2020) Application of FMCW Radar for the Recongnition of Hand Gesture Using Time Series Convolutional Neural Networks. *Proceedings of the 2020 International Conference on Microwave and Millimeter Wave Technology (ICMMT)*, Shanghai, 20-23 September 2020, 1-3. <https://doi.org/10.1109/icmmt49418.2020.9386392>
- [28] Hazra, S. and Santra, A. (2018) Robust Gesture Recognition Using Millimetric-Wave Radar System. *IEEE Sensors Letters*, **2**, 1-4. <https://doi.org/10.1109/lens.2018.2882642>

- [29] Regani, S.D., Wu, C., Wang, B., Wu, M. and Liu, K.J.R. (2021) mmWrite: Passive Handwriting Tracking Using a Single Millimeter-Wave Radio. *IEEE Internet of Things Journal*, **8**, 13291-13305. <https://doi.org/10.1109/jiot.2021.3066507>
- [30] Liu, H., Zhou, A., Dong, Z., Sun, Y., Zhang, J., Liu, L., *et al.* (2022) M-Gesture: Person-Independent Real-Time In-Air Gesture Recognition Using Commodity Millimeter Wave Radar. *IEEE Internet of Things Journal*, **9**, 3397-3415. <https://doi.org/10.1109/jiot.2021.3098338>
- [31] Zhang, Z., Tian, Z., Zhou, M., Nie, W. and Li, Z. (2018) Riddle: Real-Time Interacting with Hand Description via Millimeter-Wave Sensor. *Proceedings of the 2018 IEEE International Conference on Communications (ICC)*, Kansas City, 20-24 May 2018, 1-6. <https://doi.org/10.1109/icc.2018.8422765>
- [32] Rahman, M.M., Mdrafi, R., Gurbuz, A.C., Malaia, E., Crawford, C., Griffin, D., *et al.* (2021) Word-Level Sign Language Recognition Using Linguistic Adaptation of 77 GHz FMCW Radar Data. *Proceedings of the 2021 IEEE Radar Conference (RadarConf21)*, Atlanta, 7-14 May 2021, 1-6. <https://doi.org/10.1109/radarconf2147009.2021.9455190>
- [33] Santhalingam, P.S., Hosain, A.A., Zhang, D., Pathak, P., Rangwala, H. and Kushalnagar, R. (2020) mmASL: Environment-Independent ASL Gesture Recognition Using 60 GHz Millimeter-Wave Signals. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, **4**, 1-30. <https://doi.org/10.1145/3381010>
- [34] Hu, Y., Wang, B., Wu, C. and Liu, K.J.R. (2022) mmKey: Universal Virtual Keyboard Using a Single Millimeter-Wave Radio. *IEEE Internet of Things Journal*, **9**, 510-524. <https://doi.org/10.1109/jiot.2021.3084560>
- [35] Wei, H., Li, Z., Galvan, A.D., Su, Z., Zhang, X., Pahlavan, K., *et al.* (2022) IndexPen: Two-Finger Text Input with Millimeter-Wave Radar. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, **6**, 1-39. <https://doi.org/10.1145/3534601>
- [36] Zhang, Z., Tian, Z., Zhang, Y., Zhou, M. and Wang, B. (2019) u-DeepHand: FMCW Radar-Based Unsupervised Hand Gesture Feature Learning Using Deep Convolutional Auto-Encoder Network. *IEEE Sensors Journal*, **19**, 6811-6821. <https://doi.org/10.1109/jsen.2019.2910810>