A Survey of mmWave-based Human Sensing: Technology, Platforms and Applications

Jia Zhang, *Student Member, IEEE,* Rui Xi, *Member, IEEE,* Yuan He, *Senior Member, IEEE,*
Yimiao Sun, *Student Member, IEEE,* Xiuzhen Guo, *Member, IEEE,* Weiguo Wang, *Student Member, IEEE,*
Xin Na, *Student Member, IEEE,* Yunhao Liu, *Fellow, IEEE,* Zhenguo Shi, *Member, IEEE,*
Tao Gu, *Senior Member, IEEE*

Abstract—With the rapid development of the Internet of Things (IoT) and the rise of 5G communication networks and automatic driving, millimeter wave (mmWave) sensing is emerging and starts impacting our life and workspace. mmWave sensing can sense humans and objects in a contactless way, providing fine-grained sensing ability. In the past few years, many mmWave sensing techniques have been proposed and applied in various human sensing applications (e.g., human localization, gesture recognition, and vital monitoring). We discover the need of a comprehensive survey to summarize the technology, platforms and applications of mmWave-based human sensing. In this survey, we first present the mmWave hardware platforms and some key techniques of mmWave sensing. We then provide a comprehensive review of existing mmWave-based human sensing works. Specifically, we divide existing works into four categories according to the sensing granularity: human tracking and localization, motion recognition, biometric measurement and human imaging. Finally, we discuss the potential research challenges and present future directions in this area.

Index Terms—Millimeter wave, human sensing, mmWave sensing, mmWave radar.

I. INTRODUCTION

Wireless sensing (e.g., Wi-Fi, Zigbee, RFID, and acoustic) offers contactless sensing, and it has become a new human sensing paradigm in the last decade. As the propagation of wireless signals is affected by the sensing target, the received signals contain target-related information. By analyzing the characteristic changes of wireless signals (i.e., phase, amplitude, and frequency), the target-related information, such as human gesture or respiration, can be captured and analyzed.

In recent years, millimeter wave (mmWave) sensors have been developed rapidly and attracted extensive attention, which catalyze the emergence of mmWave sensing technology. mmWave signals operate at a high frequency (30-300 GHz) with a large bandwidth. Hence it can provide high sensing sensitivity and precision. The short wavelength of mmWave signals further enables the antennas to be highly integrated, enabling beamforming and other techniques that support directional sensing capabilities. As a result, mmWave sensing has great advantages in human sensing over low-frequency sensing technologies such as Wi-Fi, UWB, and LoRa.

In addition, mmWave sensing offers many advantages over sensor-based techniques (i.e., camera, Lidar, and ultrasonic

sensor), including a wide sensing range, fine-grained and directional sensing capability, and resistance to weather and illumination conditions. Camera-based sensing techniques provide high-resolution imaging results under good illumination conditions, thereby supporting the sensing of human status such as gesture and gait. However, camera-based sensing techniques are limited by illumination conditions and concerns about privacy. In addition, a low-resolution camera may not be able to capture micro motions. Lidar-based sensing can measure micro motions but it has a narrow sensing range. They are likely to be affected by weather conditions so that their application scenarios are limited. Ultrasonic-based sensing can resist different weather conditions but may suffer from severe signal attenuation and non-directional sensing, hence it may perform poorly in multipath scenarios.

mmWave sensing has been widely explored to "see, listen, inquire and touch" the human body to provide accurate and ubiquitous human sensing capability. It can enable various human sensing tasks such as human tracking and localization, activity recognition, vital sign monitoring, sound recovery, and human imaging. For example, in 2016 Google [1], [2] designed a tiny mmWave sensing module Soli that supports finger gesture recognition, which has been integrated into its Nest Hub and Pixel 4 smartphone. Texas Instruments (TI) [3]–[7] has developed a series of mmWave radar products for human sensing. In the field of autonomous driving, mmWave radars have been widely used in object detection and classification and are being explored in human imaging. Furthermore, considering the rapid large-scale deployment of 5G and the indispensable application of mmWave signals in 5G, mmWave sensing in 5G scenarios can play a vital role in Integration Sensing and Communication (ISAC), digital twins and so on. mmWave-based human sensing is being continuously explored and deployed by researchers and application developers, and its applications range from human-computer interaction, smart medical, smart home to smart city.

mmWave-based human sensing presents unique technical challenges compared to non-human sensing. Firstly, the mobility of the human body needs to be considered. Compared with other static objects, the human's position and posture have greater variation and uncertainties, which are important factors that affect the human sensing result. Secondly, The human vital signal usually has a low signal-to-noise ratio (SNR) and is uncertain, resulting in higher noise cancellation requirements. Compared with the micro-motion of non-human objects, the
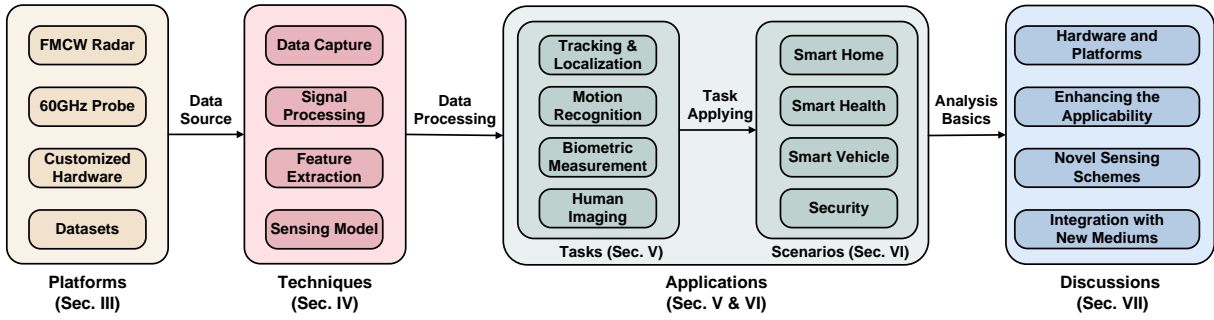
Fig. 1. Article organization.

human vital signal is more complex and easily buried by the movement-related noise. Furthermore, human vital signals are affected by various factors such as physiology, psychology, and the surrounding environment, making it difficult to analyze. Last but not least, the differences among individuals are significant. The contours, gestures, vital signals, and vocal habits of different individuals are very diverse. How to overcome these differences between individuals to achieve accurate human sensing has become a critical problem. To achieve accurate human sensing, people must carefully address these unique challenges. It is worth noting that many human sensing techniques can be extended to non-human sensing scenarios due to their attractive sensing capabilities, but this is not the focus of this article.

Existing mmWave-based human sensing works have attempted to overcome these technical challenges. However, there is still a significant gap between the performance of these works and the accurate and ubiquitous sensing capability we expect, in terms of sensing range, sensing granularity, sensing accuracy, human body constraints, etc. We wonder what factors prevent mmWave-based human sensing from this vision. To this end, we need to study the existing works so that we can summarize the common core challenges and figure out their limitations. Moreover, it is critical to point out the key challenges and development trends of mmWave-based human sensing technology so as to illuminate potential directions and improve sensing capability. Therefore, we find that there is a strong need for a comprehensive survey in mmWave-based human sensing technology, platforms, and applications.

This paper fills the gap by providing a comprehensive review of mmWave-based human sensing. The main contributions of this paper are summarized as follows:

• This paper presents a complete picture of the literature in the area of mmWave-based human sensing.

• The hardware platform and key techniques of mmWave-based human sensing are summarized in detail. Specifically, we summarize the human sensing pipeline. Then the related techniques in each module are compared and summarized.

• We comprehensively compare and summarize existing mmWave-based human sensing works into four categories according to the sensing granularity. Furthermore, we point out the core challenges and alternative solutions for each category of sensing tasks. Finally, the lessons learned about mmWave-based human sensing are given to provide readers with our suggestions.

• Following an elaborated discussion of the literature, we comprehensively discuss the key challenges and future directions related to mmWave-based human sensing, including hardware and platforms, enhancing the applicability, novel sensing schemes, and integration with new mediums. We also give a detailed discussion on the potential sensing techniques, including ISAC, THz sensing and so on.

There have been some related surveys that provide a summary of particular scopes, such as mmWave-based communication [8], [9], application scenarios [10], [11] and sensing techniques [12]–[14]. We compare and summarize them in detail in the next section. We find that none of the current surveys focuses on summarizing the existing mmWave-based human sensing works and analyzing their key challenges. As we mentioned, a comprehensive survey is necessary and important for researchers and application developers. This survey is expected to inspire researchers and developers to conduct further research in mmWave sensing and build various applications to realize accurate, ubiquitous and stable human sensing in real life.

As shown in Fig. 1, the structure of the survey is organized as follows. Sec. II summarizes the related surveys and points out the salient novelties of our paper. Sec.III discusses the mmWave hardware platform and datasets, which provide reliable data sources for subsequent sensing techniques and tasks. Sec.IV briefly introduces the key techniques of mmWave sensing from data capture to sensing model. These techniques further process the obtained data and extract corresponding human-related information for task-related analysis. Sec.V elaborates on the various mmWave-based sensing tasks from tracking and localization to human imaging. These various sensing tasks can be applied to many practical application scenarios, which are introduced in Sec.VI. Sec.V and Sec.VI together form a detailed discussion of mmWave-based human sensing applications. Based on the analysis and summary of these works, Sec.VII discusses the potential challenges and future directions. We conclude this survey in Sec.VIII.

II. RELATED WORK

There have been some related surveys that provide a mmWave-related summary of particular scopes, such as mmWave-based communication, application and sensing techniques. In this section, we first introduce these related surveys

TABLE I
SUMMARY OF RELATED SURVEYS ON MMWAVE-BASED HUMAN SENSING

Reference	Signal Sources	Application Scenarios	Sensing Tasks	Taxonomy
Wang et al. [15]	RF, Acoustic, Motion, Biometric, Optical	Wearable Application	Movement Tracking, Human-machine-interface, Vital Monitoring, Smart Hearth, Mental State Monitoring, User Authentication	Wearables' Usages within the Scope of Cross-sensing
Abdu et al. [13]	mmWave Radar Signals	Autonomous Driving	Object Detection, Classification	Learning-based Radar Data Processing for Object Detection and Classification
Shastri et al. [12]	mmWave AP Signals, mmWave Radar Signals	Indoor Localization and Sensing	Localization, Human Activity Recognition, Object Detection, Health Monitoring	Device-based Localization and Device-free Sensing
Singh et al. [10]	Radar Signals	Non-Contact Vital Sign Monitoring	Heart Rate and Respiration Rate Measurement	Research Challenges Associated with Hardware and Signal Processing
Venon et al. [14]	mmWave Radar Signals	Automotive	Motion Estimation, SLAM, Place Recognition, Detection and Recognition, Semantic Segmentation	Algorithms for Automotive
Kebe et al. [11]	Radar Signals	Vital Sign Detection	Cardio-pulmonary Signals' Detection and Monitoring	Radar Categories Used in Vital Signs Detection
This Survey	mmWave Radar Signals	Human Sensing	Tracking and Localization, Activity Recognition, Gesture Recognition, Handwriting Tracking, Gait Recognition, Vital Sensing, Sound Recognition, Human Imaging	Human Sensing Applications w.r.t. the Sensing Granularity

and summarize them in Table. I. Then we discuss the differences between the most related work and our work in detail. Finally, we point out the salient novelties of our work.

Wang et al. [15] provided a review on the cross-sensing technologies towards wearable applications. It includes sensing techniques based on multiple wireless signals, such as WiFi, mmWave and RFID, but the analysis of mmWave sensing is not enough. Abdu et al. [13] wrote a survey of deep learning approaches processing mmWave radar signals in autonomous driving applications, but their focus is on deep learning rather than on mmWave sensing. Busari et al. [8] summarized the research on mmWave massive MIMO communication and discussed research issues and future directions on mmWave massive MIMO. Although its discussions about mmWave massive MIMO communication, such as channel model and channel estimation, may be further used for mmWave sensing, the related analysis is not presented. Zhou et al. [9] conducted a review on the IEEE 802.11ay-based medium access control (MAC) layer related issues. Similarly, it mainly discussed the technical challenges, design issues and beamforming training in the IEEE 802.11ay, rather than the mmWave-based sensing issues.

The most related to our paper is [12], which summarized the state-of-the-art in device-based localization and device-free sensing using mmWave communication and radar devices. In the device-free sensing section, that paper concentrates on the relevant signal processing techniques and learning techniques. Then the application and performance of these techniques are described in terms of some sensing applications including human activity recognition, object detection and health monitoring. Compared with that work, we analyze the key techniques related to human sensing and summarize them based on the human sensing pipeline. We focus on mmWave-based human sensing tasks and the corresponding application scenarios. Furthermore, we discuss the lessons learned, key challenges and future directions of mmWave-based human sensing in detail.

There are also some related surveys that provide an analysis of particular narrow scopes, such as application scenarios [10], [11] and sensing techniques [13], [14]. We summarize these surveys in Table. I. Specifically, Wang et al. [15] provided a review on the cross-sensing techniques towards wearable applications and summarized the applied signal processing and machine learning algorithms. Abdu et al. [13] presented a survey of deep learning approaches processing radar signals to accomplish some sensing tasks in autonomous driving applications. Shastri et al. [12] focused on indoor mmWave device-based localization and device-free sensing, and provided a review of indoor localization approaches, technologies, schemes and algorithms. Singh et al. [10] focused on the research challenges of non-contact vital sign monitoring in multi-resident environments. Venon et al. [14] introduced the perception, recognition and localization techniques in automotive applications and described algorithms and applications adapted or developed for mmWave radars. Kebe et al. [11] presented a review on contactless radar-based vital signs detec-

tion and highlighted the challenges in biomedical applications. However, none of the current surveys particularly summarizes and compares existing mmWave-based human sensing works. As mmWave-based human sensing is increasingly researched in the industry and academia, a comprehensive survey is necessary and important for researchers and application developers.

Compared with these works, our paper presents new results in the following aspects: (1) Our paper is more inclusive and contains the latest advances in the field of mmWave-based human sensing. (2) We propose a comprehensive taxonomy of mmWave-based human sensing. According to the sensing granularity, our paper classifies the existing works into four categories: human tracking and localization, motion recognition, biometric measurement and human imaging. (3) We conduct a comprehensive comparison and summary of the existing works. Specifically, the hardware platform, signal form, and performance of these works are compared in detail. (4) The scope of study in our paper is broader. Besides introducing the specific mmWave-based human sensing works, we also discuss the key challenges and future directions, which may motivate more follow-up research in mmWave sensing.

III. PLATFORM & DATASETS

To meet the needs of different sensing tasks, various kinds of hardware devices have been set up to perform mmWave sensing experiments, such as commercial FMCW radars, 60GHz probes, and customized devices. In the following, we briefly introduce three hardware platforms and compare different products. Furthermore, we list some public datasets containing mmWave signals for sensing different human statuses, such as activity, vital, and 3D pose.

A. FMCW Radar

As Fig. 2 illustrates, a mmWave radar system is mainly consisting of four components: transmit (Tx)/receive (Rx), radio frequency (RF), analog components, and digital components. The mmWave signal is generated by a voltage-controlled oscillator (like LO Gen. in Fig. 2), a part of which is additionally amplified by the power amplifier (PA) and fed to the transmitting antenna, and the other part is coupled to the mixer, mixed with the received echo. The transmitter can use different types of waveforms, such as pulsed, FSK (Frequency Shift Keying), CW (Continuous Wave), and FMCW (Frequency Modulated Continuous Waveform). In this paper, we mainly focus on the FMCW radar. The transmission signal is modulated in frequency, linear changing over a defined time. The received signals by Rx are processed by LNA (Low noise amplifier) and then mixed with the transmitted signals to obtain the intermediate frequency (IF) signals, whose frequencies are equal to the frequency differences between the transmitted signals and the received signals. After analog-to-digital conversion (ADC), they are sent to the signal processor for further processing. Radar systems can either generate complex signals (i.e., I/Q signals) using I/Q quadrature mixing, or real ones (i.e., I-channel signals) by I single-channel mixing. As Fig. 2 illustrates, the complex signals contain I and Q components that have the same amplitude and frequency but
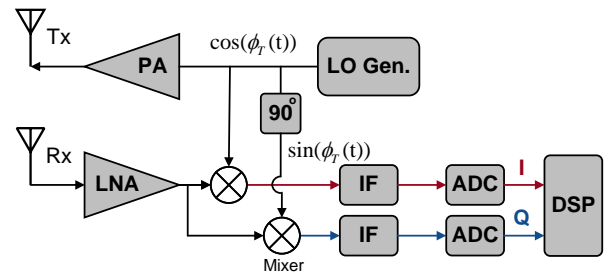

Fig. 2. The block diagram of mmWave radar system.

are shifted a quarter cycle relative to each other, which can be generated by adding a 90° phase shift difference between the two mixers. Contrarily, the I single-channel mixing only uses a mixer that provides the absolute frequency shift.
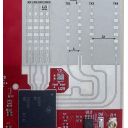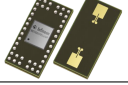
Actually, a complete radar system is much more complicated and highly integrated. Besides the above components, the manufacturer will integrate some additional subsystems in an extremely small form-factor by leveraging some advanced techniques, such as Monolithic Microwave Integrated Circuits (MMICs), Antenna in Package (AiP), Silicon-Germanium-Technology (SiGe), RFCMOS process, etc. For example, TI IWR1843 is a self-contained solution that includes a monolithic implementation of a 3TX, 4RX system with built-in PLL and ADC. It also integrates a DSP subsystem, an ARM Cortex-R4F-based processor subsystem, and a hardware accelerator (HWA), built with the low-power 45nm RFCOMS process. More details of the commercial mmWave radars can be found in TABLE II.

B. 60GHz Probe

Benefiting from the high bandwidth, high directivity, and low latency of mmWave signals, mmWave networks have enabled new applications for businesses and consumers, including enhanced telehealth and education, industrial automation, VR/AR, etc. At the 2019 World Radiocommunication Conference (WRC-19), delegates identified several mmWave frequency bands that could be used for 5G networks, including 24.25-27.5 GHz, 37-43.5 GHz, 45.5-47 GHz, 47.2-48.2 and 66-71 GHz. On the other hand, WiGig, alternatively known as 60 GHz Wi-Fi, has been promoted and released by Wireless Gigabit Alliance to support wireless communication at multi-gigabit speed. It includes the current IEEE 802.11ad standard and 802.11ay standard. These network protocols have promoted the development of mmWave-based communication probes [20], [21]. These probes can not only achieve high throughput communication with mmWave signals but also have the potential of mmWave sensing.

There are a few probe-based mmWave sensing works in the current literature. Most of them [22]–[24] utilize the Qualcomm chipset to obtain the sensing information. Qualcomm has provided a series of 60GHz WiFi solutions with the IEEE 802.11ad standard [20]. The commodity Qualcomm 802.11ad chipsets have 32 antennas assembled in a 6×6 layout with a size factor of 1.8 cm $\times$ 1.8 cm for both the transmitter and the receiver. The receiver hardware further implements

TABLE II
COMPARISON OF MMWAVE HARDWARE DEVICES

Products	Freq.	Antenna size (Tx × Rx)	ADC sampling rate (max) (MSPS)	TX power (dBm)	Package Size mm^2:(W × L)	Manufacturer	Image
AWR1xx IWR1xx	76 GHz ~ 81 GHz	3 × 4; 2 × 4 (xWR1642)	12.5 (xWR162, AWR1443); 25 (xWR1843); 37.5 (AWR1243, IWR1443)	12	10.4 × 10.4; 15 × 15 (AWR1843AOP)	TI	
AWR6xx IWR6xx	60 GHz ~ 64 GHz	3 × 4	25	12; 10 (A/IWR6843AOP)	10.4 × 10.4; 15 × 15 (A/IWR6843AOP)	TI	
AWR2xx	76 GHz ~ 81 GHz	3 × 4; 4 × 4 (AWR2944)	45 (AWR2243); 37.5 (AWR29xx)	13(AWR2243); 12(AWR29xx)	10.4 × 10.4 (AWR2243); 12 × 12 (AWR29xx)	TI	
TINYRAD24G	24 GHz ~ 24.25 GHz	2 × 4		8	85 × 85	Analog Device	
TRX_024	21.5 GHz ~ 28.7 GHz	1 × 1	2.5	6	3 × 3	Silicon Radar	
TRX_120 TRA_120	119.3 GHz ~ 125.8 GHz	4 × 4	2.5	-3	8 × 8 5 × 5	Silicon Radar	
BGT60LTR11AIP	61 GHz ~ 61.5 GHz	1 × 1	-	10	3.3 × 6.7	Infineon	
TEF82xx	76 GHz ~ 81 GHz	3 × 4	40	13.5	7.5 × 7.5	NXP	

a Golay correlator to obtain the sensing information, i.e., channel impulse response (CIR). With this information, many human sensing applications can be achieved, such as human tracking and localization, vital sensing, human imaging, etc.

C. Customized Hardware

Furthermore, many researchers intend to open up new directions for mmWave sensing and protocol development and support system-oriented research by building customized experimental platforms, such as mmWave software radio, mmWave radar, etc. We list them in TABLE III.

WiMi [35] is the first 60GHz testbed with a reconfigurable RF front-end and software-defined baseband processing modules. Based on this platform, **Openmili** [16], a software-defined mmWave network stack, is designed to span PHY layer signal processing to applications. OpenMili's most outstanding feature is a reconfigurable phased-array antenna that can switch between 16 beam patterns at microsecond granularity. The designed phased-array specifically fits the WR-15 waveguide (a standard antenna interface on 60GHz radios), ensuring it can retrofit both OpenMili's RF front-end and other commercial mmWave radios that are typically equipped with WR-15 horn antennas RF front-end.

M^3 [17] is the first mmWave massive MIMO software radio. It provides up to eight 32-element phased-arrays, at a cost that is an order of magnitude lower than existing commercial mmWave SDR solutions. The baseband module (BM or 11ad NIC) and phased array module (PM) are commercial multi-array 802.11ad nodes. The IF bridge board is a customized PCB. The baseband processing unit (BPU) can be an FPGA or an SDR, such as a WARP board or a USRP. The control FPGA is a Cmod A7 module with an adapter board. It can hijack commercial mmWave radios for any waveform transmission and real-time phased array reconfiguration.

mm-FLEX [18] supports a bandwidth of 2 GHz and is compatible with mmwave standard requirements. Moreover, it integrates a powerful FPGA-based baseband processor with full-duplex capabilities together with mm-wave RF front-ends and phased antenna arrays that are fully configurable from the processor in real time.

In [19], a customized light mmWave platform consisting of the RF board, the baseband board, and the MCU board, is also proposed. For the RF board, Rogers RT/duroid 5880, which can provide a thickness lower to 0.0096 in, is used as the substrate. Furthermore, both Tx and Rx consist of 16 antennas following the 4 × 4 layouts and the size of the microstrip patch antenna is 5 mm × 3.7 mm. The distance between Tx and Rx
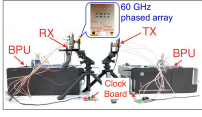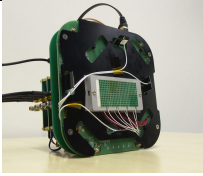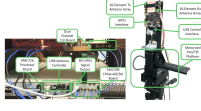
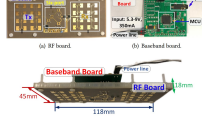TABLE III
COMPARISON OF THE CUSTOMIZED HARDWARE.
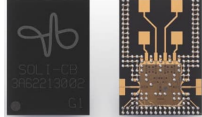
Reference	Openmili [16]	M^3 [17]	mm-Flex [18]	[19]	Soli [2]
Image					
Baseband BW	1 GHz	4 GHz	2 GHz	450 MHz	7 GHz
Carries Freq.	$57 \sim 64$ GHz	60 GHz	60 GHz	24 GHz	60 GHz
Antenna	Horn/Phased-array	Phased-array	Phased-antenna	Array	Array
Cost	$15K	below $15K	$40K	below $100	-

TABLE IV
COMPARISONS OF THE OPEN MMWAVE DATASETS.

Dataset	Sensors	Data Format	Setups	Types of sensing	Scenario	# of Subjects	# of Activities	# of Frames
mRI [25]	RGB,Depth,IMU, TI IWR1443	Images, Point Cloud, Inertial Signals	RGB:30 Hz, Depth:30 Hz, IMU:50 Hz, mmWave:10 Hz	Rehabilitation	Single	20	12	160K
MARS [26]	Microsoft Kinect, TI IWR1443	Point Cloud	mmWave:3.2 GHz BW, 10 Hz Camera: 30 Hz	Rehabilitation	Single	4	10	40K
mmPose [27]	TI AWR1642, Microsoft Kinect	Point Cloud	mmWave: 3.072 GHz BW, 10 Hz	Walking, Swing	Single	2	4	38K
mmBody [28]	Azure Kinect, mmWave radar	Point Clouds, RGB(D) Images	Radar: 10-30 fps, Camera: 30 fps	Skeleton, Mesh	-	20	100	200K
MMActivity [29]	TI IWR1443	Point Cloud	3.19 GHz BW, 30 fps	Boxing, Jumping, Squats, Walking	Single	2	5	16k
mmMesh [30]	TI AWR1843, VICON	Point Cloud	VICON:10 fps mmWave:3.9 GHz BW, 10 fps	Skeleton	Single	20	8	480K
M-gesture [31]	TI IWR1443	Raw signals, Point Cloud, Raw RDIs	4 GHz BW, 18.18 fps	Gestures	Single	144	5	56K
mmGait [32]	TI IWR6843, TI IWR 1443	Point Cloud	IWR6843: 3.75 GHz BW, 10 fps, IWR1443:4 GHz BW, 10 fps	Gaits	up to 5	95	-	-
mHomeGes [33]	TI IWR1443	Doppler Profiles	3.19 GHz BW, 10 fps	Arm Gestures	Single	25	10	22K
Pantomime [34]	TI IWR1443	Point Cloud	3.19 GHz BW, 30 fps	Mid-air Gestures	Single	45	21	23K

of the probe is 40 mm. Besides, it also integrates a voltage regulator, difference amplifier, on-chip 12-bit ADC and MCU together. The total size reaches 4.65 in $\times$ 4.65 in $\times$ 0.59 in with 45.4 g weights, and its power consumption is only 1.23 W.

Moreover, Google ATAP (Advanced Technologies and Projects) also designs a radar chip, termed Soli [2], implemented at mmWave RF frequencies. It can be integrated into consumer electronic devices with minimal effort. The antenna, RF front-end, baseband processing, VCO and serial communication are integrated on the chip. Similar to the commercial mmWave radar, it also operates in the 60 GHz band with 7 GHz bandwidth and requires only minor hardware modifications to be compatible with 802.11ad and Wi-Gig standards [36]. The chip uses an antenna-in-package

(AiP) patch array to form a wide 150° beam. At the same time, beamforming is realized at the receiver. It uses digital beamforming technology in SiGe-based FMCW chips instead of the analog phase shifters used on CMOS chips.

D. Datasets

To advance mmwave-based research efforts, some researchers have published their datasets covering several different types, such as gait, gesture, and rehabilitation. In the following section, we will give a detailed description and list them in TABLE IV.

MARS [26] is the first rehabilitation movement dataset using mmWave point cloud with well-labeled joints. It uses a TI IWR1443 Boost mmWave radar to collect mmWave data and a Microsoft Kinect V2 sensor to generate its respective

labels. During collections, a subject is required to perform 10 distinct rehabilitation movements. Each frame or image will be processed to compute the 3D positions of 19 joints covering the upper and lower body. Besides, **mRI** [25] also builds a high-quality and large-scale dataset, consisting of many complementary modalities, including mmWave point clouds, RGB frames, depth frames and inertial signals.

MMActivity dataset [29] is the first dataset used for human activity recognition through mmWave radars. TI IWR1443BOOST, which works in the 76 GHz to 81 GHz frequency range, is used for data collection. This dataset consists of five different activities collected from two users, including walking, jumping, jumping jacks, squats and boxing. The captured point clouds contain spatial coordinates, velocities, distances, intensities and bearing angles. The sampling rate of the radar is 30 frames per second. In addition, **mm-Pose** [27] uses two TI IWR1642, where one is rotated 90° counterclockwise with respect to the another, to collect mmWave samples from four actions. The ground truth is obtained by a Microsoft Kinect.

Furthermore, **mmBody** [28] builds a data acquisition platform using the Phoenix mmWave radar by Arbe Robotics, and two depth cameras, where the master one is right under the mmWave radar and the slave one is located on one side of the radar-body line. The subject is at a distance of 3-5 m away from the radar and the slave one is 1.5-2.3 m from the body. Furthermore, it requires 20 volunteers with different weights and heights to perform 100 motions. Additionally, it conducts the experiments under different conditions, such as poor lighting, rain, smoke, and occlusion with different materials. Similarly, **mmMesh** [30] chooses TI AWR1843BOOST to collect mmWave data reflected from 8 daily activities. VICON motion capture system [37] with a sampling rate of 10 fps is used to obtain high-precision dynamic pose information of the subject, which can be utilized to generate the ground truth human mesh.

M-gesture [31] builds the first dataset to collect mmWave radar data for gesture recognition. A total of 1357 minutes of data are collected, involving 144 people (including 64 males and 80 females). The dataset contains not merely direct sensing, but also sensing with certain blockages (e.g., paper, corrugated paper, metal board). Moreover, **mHomeGes** [33] and **Pantomime** [34] also use TI-IWR1443 to collect gesture datasets in smart home scenarios.

mmGait [32] builds the first mmWave dataset, collecting gait data from 95 volunteers (including 45 males and 50 females) using two mmWave radars, TI IWR6843 and IWR1443. their locations are 1 m away from each other. During collections, volunteers are required to walk in two modes, fixed route and freely, and it is up to 5 people for the multiple scenarios. In data processing, it further converts the point clouds, collected by two devices with different rectangular coordinate systems, into the same coordinate system.

E. Summary

To sum up, there are already many COTS mmWave devices or customized platforms to provide reliable data sources to support various sensing tasks. While different mmWave hardware and platforms have different signal bandwidths, carrier frequencies, antenna arrangements and transmission powers, which can provide different sensing ranges, sensing granularities and sensing dimensions, how to choose a suitable platform is a prerequisite for achieving a particular task. For example, vital-related motion is very tiny and its sensing requires sensing accuracy to be even as high as sub-millimeters. Therefore, the vital sensing task should use mmWave devices with high carrier frequency and large bandwidth. Moreover, human imaging requires the two-dimensional spatial information of the target, which requires the radar to provide corresponding two-dimensional sensing capabilities. In this case, these radars with the linear arrangement of the receiving antenna cannot meet the requirements.

In terms of FMCW radar, although various radars have been used for different sensing tasks, most of them are limited in the number of antennas. This leads to limited angular resolution and insufficient point cloud outputs, resulting in fragile robustness for practical applications. Therefore, developers should pay attention to these indicators in the design of future radars. More details about hardware development are discussed in Sec. VII-A.

In terms of 60GHz probe, the 60GHz devices used for sensing are still limited. With the continuous development of ISAC in 5G communication, 60GHz probes may become popular for sensing. The designers may need to pay more attention to the sensing range of the 60GHz probe due to its directivity and high attenuation. The design of the interplay between communication and sensing also needs to be considered. Sec. VII-D further discusses the challenges and opportunities of sensing combined with communication.

Moreover, customized hardware potentially provides more tailored designs that COTS devices generally do not provide, including antennas with adjustable polarization directions, customized antenna array arrangement, scene-adapted denoising hardware, etc. These customized designs can greatly help researchers to explore novel mmWave-based human sensing techniques.

Furthermore, there are some public datasets listed in this section, and different parameters of each dataset are shown. Researchers can find appropriate datasets quickly, so as to propose advanced information processing techniques and models, and promote the development of the field.

IV. KEY TECHNIQUES

In this section, we briefly introduce some key techniques related to mmWave-based human sensing. We first introduce the general working pipeline of human sensing, including data capture, signal preprocessing, feature extraction, sensing model and sensing task. Then we state the key techniques in each module separately. The sensing task will be elaborated on in the next section.

A. Human Sensing Pipeline

The general flowchart of mmWave-based human sensing is as follows: The mmWave transceiver always periodically
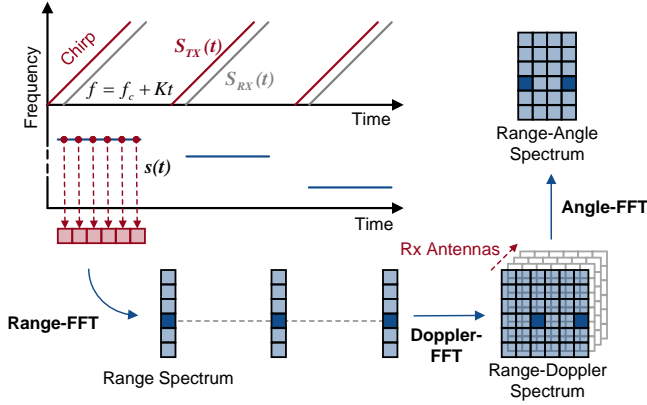
Fig. 3. Illustration of data capture based on FMCW radar.

transmits the mmWave signal and continuously receives the reflected signal. The received signal is downconverted by the mixer and sampled by the ADC. Then these samples can be captured and converted to the raw data (**Data capture**). The raw data is further processed into various signal forms to complete different sensing tasks. Some denoising techniques are performed to resist environmental interference (**Signal preprocessing**). To obtain human-related information from mmWave signals, various features can be extracted and analyzed (**Feature extraction**). Furthermore, according to different sensing tasks, the task-related features need to be fed into different models for accurate sensing (**Sensing model**). Finally, the sensing task can be achieved (**Sensing task**).

We first introduce some details of data capture in Sec. IV-B. Then we discuss the key techniques of signal preprocessing in Sec. IV-C. A brief introduction to feature extraction is presented in Sec. IV-D. Sec. IV-E described the classification of the sensing model. Finally, we elaborate on the details of the sensing task in Sec. V.

B. Data Capture

To achieve accurate mmWave-based human sensing, the raw data must be carefully captured and processed. According to the differences in data sources, we divide data capture into two categories, namely data capture based on FMCW radars and data capture based on 24/60GHz probes. We introduce them separately in the following.

1) Data capture based on FMCW radar: The mmWave radar usually periodically sends FMCW chirp signals for distance and velocity measurement. As shown in Fig. 3, the frequency difference between the transmitted signal and the received signal corresponds to the signal propagation time and can be utilized to determine the object distance. Denoting the distance between the radar and the sensing target by $R(t)$, the transmitted and the received signal can be represented as:

$$
\begin{aligned}
S_{Tx}(t) &= \exp[j(2\pi f_c t + \pi K t^2)] \\
S_{Rx}(t) &= \alpha S_{Tx}[t - 2R(t)/c]
\end{aligned} \tag{1}
$$

where α is the path loss. f_c and K are the chirp start frequency and the chirp slope of the FMCW signal, respectively. By

mixing the transmitted signal and the received signal, the *beat frequency signal* $s(t)$, namely IF signals, can be obtained as:

$$
s(t) = S_{Tx}(t)S_{Rx}(t)^* \approx \alpha \exp[j4\pi(f_c + Kt)R(t)/c] \tag{2}
$$

whose phase values indicate the distance information $R(t)$. To separate received signal components reflected from different ranges, a *Range-FFT* [38] operation is performed on the samples of $s(t)$ within a chirp for signal separation. As shown in Fig. 3, this operation maps the frequency spectrum of $s(t)$ to the range spectrum. To further measure the velocity of the target, the samples in the corresponding range bin are selected from the Range-FFT results and combined to form the *reflected signal*, which can be obtained as:

$$
s(t) \xrightarrow[\text{at object range bin}]{\text{Range-FFT}} S(t) = \alpha \exp[j4\pi f_c R(t)/c] \tag{3}
$$

Then the velocity of the object can be estimated by performing another FFT operation called *Doppler-FFT* [38] on $S(t)$. With Range-Doppler-FFT, the radar can obtain the Range-Doppler spectrum and detect the existence of the object.

To further depict the position of the object, multiple receive antennas of the mmWave radar are utilized to derive the angle between the radar and the object. By performing the third FFT, Angle-FFT [38], on these received signals, the radar can obtain the Range-Angle spectrum and detect the exact position of the object. To further improve the angular resolution, the beamforming technology [39], [40] instead of Angle-FFT can be utilized. By calculating the optimal aggregation weights of the transmit/receive antennas, they can be concentrated in the desired direction to obtain a much higher resolution.

2) Data capture based on 24/60GHz probe: Unlike the mmWave radar, which utilizes the beat frequency signal to locate sensing targets, the mmWave probe exploits the channel impulse responses (CIRs) between transmit antennas and receive antennas to determine the target's location. The time delay of the CIR indicates the signal propagation time and can be used to determine the object distance. Denoting the traveling distance of the mmWave signal reflected by the target as $d(t)$, the CIR between transmit antenna m and receive antenna n can be expressed as:

$$
h_{m,n}(t) = \alpha_{m,n} \exp(-j2\pi d_{m,n}(t)/\lambda_c) \tag{4}
$$

where $\alpha_{m,n}$ and λ_c represent the complex channel gain and the carrier wavelength, respectively. If the reflected signal falls into a certain CIR tap, the distance between the probe and the target can be determined by mapping the CIR spectrum into the range spectrum. Furthermore, by adjusting the coefficients of the transmitter/receiver antennas in sequence, the mmWave probe can perform beamforming at both transmitter and receiver to obtain the reflected signal from each angle and determine the exact position of the object.

C. Signal Preprocessing

After the data capture module, the raw data needs to be processed into various signal forms to complete different
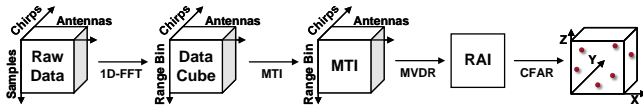
Fig. 4. The flowchart of point cloud generation.

sensing tasks. We first introduce each signal form for further human sensing. Then some denoising techniques are discussed to resist environmental interference.

1) Signal form: The signal forms of the mmWave data include the Range-Angle spectrum, Range-Doppler spectrum, phase waveform, point cloud, etc. People can choose the suitable signal form according to their sensing tasks. For example, the Range-Angle spectrum is suitable for human tracking while the phase waveform can be utilized for vital sensing. Following, we will introduce each signal form separately.

a) Range-Angle spectrum & Range-Doppler spectrum: As we introduced in Sec. IV-B, the Range-Angle spectrum can be obtained by either Range-Angle-FFT from the FMCW radar or CIR mapping and beamforming from the mmWave probe. The complex value in each Range-Angle bin in the Range-Angle spectrum indicates the reflected signal from the corresponding spatial position. The intensities can be utilized to detect the existence of objects, and the phases can further represent the micro displacement of the object.

Similarly, the Range-Doppler spectrum can be obtained by either Range-Doppler FFT from the FMCW radar or CIR mapping and phase differencing from the mmWave probe. The value in each Range-Doppler bin indicates the moving objects and their corresponding velocities. Therefore, the Range-Doppler spectrum can be utilized to separate these objects with different velocities in the same range.

b) Phase waveform: To further characterize the micro displacement of the human body, a series of phase values can be extracted from the Range-Angle spectrum or directly from the Range-FFT result to form the phase waveform. As the phases of the reflected signal indicate the reflector's micro displacement, the phase waveform can be utilized for fine-grained human sensing, such as vital sensing.

c) Point cloud: A point cloud refers to a collection of the reflection points representing the object's surface, which has been widely used in various sensing tasks such as human imaging. Similar to LiDAR and vision, mmWave radars can obtain sparse point clouds from the raw data through a series of operations. Recently, the point cloud generation method proposed by Texas Instrument Technology has been widely used. As shown in Fig. 4, the generation pipeline includes Range-FFT, Moving Target Indication (MTI) method [41] and Minimum Variance Distortionless Response algorithm (MVDR) [42]. Then the Range Angle Image (RAI) can be obtained, and the peak points can be further detected by Constant False Alarm Rate algorithm (CFAR) [43]. It is worth noting that due to the limited hardware capability, the point number in each point cloud is usually limited.

2) Signal denoise: In this section, we discuss some signal denoising techniques for each signal form mentioned before

to resist environmental interference.

a) CFAR & Spectrum subtraction: For the Range-Angle spectrum and the Range-Doppler spectrum, there is a lot of background noise from static objects and multipath interference existing in the area without humans. These noise confuses the localization of the human body and needs to be properly removed. There are many background noise elimination techniques, such as CFAR and spectrum subtraction. CFAR is a classical adaptive algorithm used to detect targets against environmental noise. It adaptively selects a noise threshold level to detect the bins with objects. Spectrum subtraction algorithm [44] is another effective method to eliminate the background noise. The main idea is to subtract the estimation of the average background noise from the noisy measurement.

b) Fitting & Filter: For the phase waveform, as the reflected signal from the area with humans includes not only the human-reflected signal but also the signal reflected from other objects in the same area, the phase values extracted from such reflected signals are distorted. Researchers have proposed some solutions to handle this problem, such as performing the circle fitting algorithm [45], [46] or the line fitting algorithm [47] on the reflected signal, applying the FIR filter or the bandpass filter to the phase waveform, and so on.

c) Clustering: For the point cloud, the sparse point clouds generated from the mmWave radar are dispersed and the noise can be significant. Some data clustering algorithms have been employed to determine which points are caused by reflections from humans. For example, DBSCAN clustering algorithm [48] and K-mean clustering algorithm [49] have been exploited to merge these points into clusters to separate these human-related points from noise.

3) Comparison: As different signal preprocessing techniques are suitable for different sensing tasks, we compare these techniques according to their applicable sensing tasks. In human localization and motion recognition, Range-Angle spectrums and point clouds are commonly used as they indicate the environment reflection, including the position and posture of the human. In this case, CFAR, spectrum subtraction and clustering can help remove background noise from static objects and multipath interference, thereby helping us to focus on the reflected signal from the human body. Whereas in biometric measurement, phase waveforms are chosen due to their unique ability to characterize the micro-motions. Compared with CFAR and clustering, fitting and filtering are more suitable for time series data. In human imaging, point clouds and Range-Angle spectrums can be selected. The former is more visible and the latter contains more spatial information. On the contrary, phase waveforms are rarely considered because there is less attention to tiny movements in human imaging.

D. Feature Extraction

According to the different sensing tasks, the various features need to be extracted from the denoised signals. Time-domain analysis and frequency-domain analysis are two categories of feature extraction methods widely used in mmWave-based human sensing. The former focuses on analyzing the periodicity,

duration, amplitude and other features of signals in the time domain, while the latter mainly extracts the frequency features, including spectrum, periodicity, power spectrum, etc.

1) Time-domain analysis: Some time-domain analysis techniques have been exploited to obtain the signal period or separate the superposition of multiple signals. For example, the template matching algorithm has been utilized in sensing human vital to obtain the heart rate and the breathing rate. Some signal decomposition techniques, such as variational modal decomposition algorithm (VMD) and empirical mode decomposition algorithm (EMD), have been exploited to separate the vital sign waveforms from the reflected signal. For example, mmHRV [50] regards the chest motion derived from the reflected mmWave signal as the linear superposition of the respiration signal, heartbeat signals and other motion signals. By taking each signal as an intrinsic mode function (IMF) component, mmHRV utilizes the VMD algorithm to realize signal decomposition and extracts a pure heartbeat signal.

2) Frequency-domain analysis: Compared with time-domain features, frequency-domain features are often more stable and distinguishable. Many feature-domain analysis techniques have been utilized to resist various noises and extract distinguishing features. For example, researchers have utilized the Fourier transform and wavelet transform techniques [51] to precisely extract the vital sign features. Furthermore, Short-time Fourier transform (STFT) [52] and some frequency-domain features, such as residual phase Cepstrum coefficients (RPCC) [53] and Mel frequency cepstral coefficient (MFCC) [54]), have been exploited to achieve sound recognition or motion recognition. For example, SPARCS [55] employs the STFT algorithm to extract the micro-Doppler spectrum of human motion. By dividing a longer reflected signal into shorter segments and computing their Fourier transform separately, SPARCS can simultaneously extract the time-frequency features and achieve accurate activity recognition.

E. Sensing Model

While the sensing tasks are application-dependent, researchers need to build a suitable model to obtain sensing results with the extracted features or mmWave signals as inputs. For simplicity, the commonly used models can be divided into two categories: domain knowledge-based and deep learning-based. The former is strongly dependent on the expert's experience. On the contrary, the latter is data-intensive and requires a large amount of data. Both of them will be described in detail below.

1) Domain knowledge-based model: For some periodic or simple predefined motions, each exhibits specific physical characteristics. Researchers can easily build task-specific models that cast above-extracted features to different motions by analyzing signals or bringing in domain-specific knowledge, such as anthropology and physics. For example, gait is a repetitive motion whose cycle can be divided into eight phases (e.g., heel strike, foot flat, midstance, mid-swing, and so on). The features of the gait, such as speed, step time, step length, etc., can be computed by peak finding, pattern matching, and other techniques. Furthermore, based on the built model,

researchers can exploit traditional machine learning techniques (e.g., random forest [56], decision tree [57] and support vector machine (SVM) [58]). However, this method has poor generalization ability and can only be used for simple or specific tasks. Moreover, its performance is limited by the researcher's domain knowledge.

2) Deep learning-based model: As opposed to the domain knowledge-based model relying on handcrafted features, due to the powerful feature representation capability, the deep learning-based model can automatically extract features directly from mmWave signals (e.g., spectrogram, point clouds) for different tasks. Amongst, Convolutional Neural Networks (CNNs) [59] and Recurrent Neural Networks (RNNs) [60] are usually exploited to learn the spatial relationship and temporal dependencies, respectively. For example, RF-SCG [61] applies a CNN-based architecture to recover SCG waveform from the reflected signals. However, these methods require a large amount of high-quality data to obtain an accurate model. To tackle it, many works choose Generative Adversarial Networks (GANs) [62] to enhance the samples by relying on the signal propagation model or task-related physical models. For example, MILLIEAR [63] directly employs a conditional generative adversarial network (cGAN) to enhance the audio components and reduce noise. Furthermore, MilliPose [64] leverages a cGAN architecture to generate a high-resolution 2D full-body silhouette image from the low-resolution 3D mmWave signals.

3) Comparison: We compare these two approaches as follows: (1) Domain knowledge-based methods can use domain knowledge and mathematical principles to derive and implement signal representations. Furthermore, they can also predict the performance limitations and robustness of the processing techniques. However, they may have difficulty dealing with non-linear, noisy, or uncertain signals that require advanced mathematical tools or assumptions, resulting in less flexibility or scalability. (2) Deep learning-based methods can learn from data and adapt to changing situations, optimize system performance, and intelligently filter signals. Moreover, they can handle complex and high-dimensional data that may be difficult for the former methods. Even though, deep learning requires large amounts of labeled data, computational resources, and human expertise to train and deploy effective models. Besides, they may also suffer from overfitting, underfitting, or bias issues if the data is not representative or sufficient.

F. Summary

As mentioned in this section, there are already many well-performing signal processing techniques from data capture to sensing models. Therefore, how to choose the appropriate techniques for each step in the sensing pipeline becomes critical to achieve a preset sensing task. Roughly speaking, point clouds and deep learning-based sensing models are often chosen when performing contour-related sensing tasks such as human imaging and gesture recognition. The reason is that deep learning-based sensing models are often used to achieve super-resolution detection or accurate multi-class classification. Furthermore, point clouds are suitable for imaging tasks due to their 3D discrete point characteristics. When performing

micro-motion-related sensing tasks such as vital sensing and sound recognition, phase waveform and frequency-domain analysis are often considered. The reason lies in the ability of phase waveform to represent tiny displacements and the requirement for frequency domain features in these sensing tasks. In summary, people are recommended to take advantage of the most suitable techniques to obtain better sensing results.

Moreover, existing sensing techniques may not meet the requirements of specific sensing tasks due to the diversity and complexity of these tasks. This prompts researchers to continuously explore novel sensing techniques, including customized data capture and signal preprocessing techniques, novel feature extraction techniques and compound sensing models. For example, mHomeGes [33] proposes a novel signal form namely Concentrated Position-Doppler Profile (CPDP) to avoid both signal noise and ambiguous angular resolution, thereby enabling accurate arm gesture recognition. mmFace [65] reconstructs the mmWave signals into 3D facial images and then extracts distance-resistant facial structure features for accurate face authentication. Such spatial features are more suitable for human imaging, compared with time-domain features or frequency-domain features. AmbiEar [66] proposes an indirect voice sensing model from surrounding objects and incorporates a deep learning model to achieve voice recognition in non-line-of-sight (NLoS) scenarios. This can inspire researchers to combine the domain knowledge-based model and deep learning-based model in a complementary manner, which may improve the interpretability and accuracy of the sensing pipeline simultaneously. These techniques can help researchers overcome specific challenges and achieve better sensing performance.

V. Sensing Task

According to the sensing granularity of the human sensing task, we divide the mmWave-based human sensing works into four categories: human tracking and localization, motion recognition, biometric measurement and human imaging. As shown in Fig. 5, the motion recognition works are further divided into three categories, including activity recognition, gesture recognition and handwriting tracking for better summary and comparison. The biometric measurement works are also divided into gait recognition, vital sensing and sound recognition according to the sensing task. We will introduce these works in the following. Finally, we summarize the lessons learned to help readers conduct their research more smoothly.

A. Tracking & Localization

Human tracking and localization is the critical task for ubiquitous human sensing. With human position and trajectory, various applications can be developed. Besides, tracking and localization are cornerstones of further human sensing, including motion recognition, biometric measurement and so on. We summarize the related works in TABLE V.

The core challenge of human tracking and localization lies in how to distinguish the human from surrounding static objects. As the reflected signal of the human body is likely to
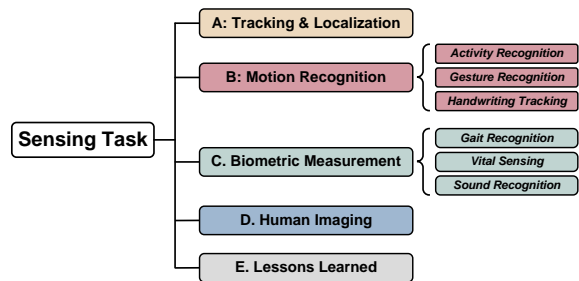

Fig. 5. The structure of sensing task.

be confused with that of other objects, specific human-related features need to be mined to distinguish people. Furthermore, when multi-person localization is required, continuous tracking is also a fundamental problem.

Most of the existing works mainly detect humans by mining the dynamics of the human body. Some works focus on human motion dynamics and utilize the signal variance to detect humans, while other works achieve human recognition by observing the dynamics of the vital signs. There are also some works that utilize the human body contour to enable human detection. For multi-person tracking and localization, researchers usually consider it as the association between the detection results and existing trajectories. It can be translated into a bipartite graph-matching problem and can be well-solved by many classical algorithms, such as the Hungarian algorithm.

mmSense [67] enables a multi-person detection and identification system with a single 60GHz mmWave radio. It is implemented based on a customized 60GHz sensing platform, which contains a Keysight signal generator equipped with the Vubiq 60GHz front-end [70]. mmSense first segments the detection region into multiple areas and constructs the RSS fingerprints with and without human presence. It then utilizes a human detection model based on long short term memory (LSTM) to determine whether there are people in each area. In this way, mmSense can detect the presence of humans and localize their positions. mmSense further utilizes multiple features to identify humans, including body surface boundary, body surface curvature and vital signs. Results show that mmSense achieves a human presence classification accuracy of 98.75% and a human identification accuracy of 93%.

Different from mmSense, **mmTrack** [23] utilizes the spatial spectrum rather than the RSS fingerprints to achieve multi-person localization. mmTrack is implemented based on a Qualcomm 60GHz 802.11ad chipset with an additional antenna array. The co-located transmitter and receiver arrays are both equipped with 32 elements. As illustrated in Fig. 6, it first extracts the CIR measurements from the received signal and employs the beamforming algorithm to obtain the spatial spectrum. Considering the human motion dynamics, mmTrack calculates the variation of the energy distribution of the spatial spectrum at each range to detect targets. After transforming the points in the Spherical coordinate system to 3D locations in Euclidean space, mmTrack clusters these points via K-means clustering algorithms [71] to estimate the human lo-

TABLE V
COMPARISON OF TRACKING & LOCALIZATION WORKS

| Method	Signal Form	Algorithm	Hardware Platform	Identify Accuracy	Localization Error
mmSense [67]	RSS Fingerprint	LSTM	Customized 60GHz Platform	93%	-
mmTrack [23]	CSI	Beamforming, K-means Clustering, Weighted Bipartite Graph Matching	Qualcomm 60GHz 802.11ad Chipset	97.8%	16.24 cm
mID [68]	Point Cloud	DBSCAN Clustering, Hungarian Algorithm, BiLSTM	TI IWR1443boost	89%	16 cm
PALMAR [69]	Point Cloud	Adaptive Order Hidden Markov Model, Crossover Path Disambiguation	TI 79-81GHz mmWave Sensor	-	6.9 cm of Single Person 11.5 cm of Multiple Persons
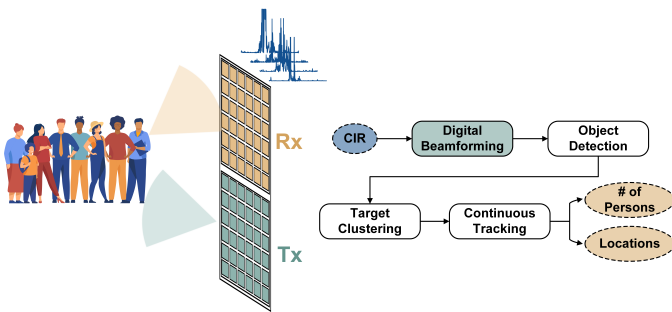
Fig. 6. The overview of mmTrack.

cations. mmTrack further leverages the geometric properties of the detected targets to design post-validation techniques for clustering robustness improvement. Finally, mmTrack models continuous human tracking as a weighted bipartite graph matching problem to perform the association between the detection results and existing trajectories. It achieves a median location error of 16.24 cm with moving and static users.

The spatial distribution of point clouds indicates the contour of the human body, which can also be used for human detection. **mID** [68] demonstrates the feasibility of human tracking and identifying with mmWave radars. It utilizes the sparse point clouds generated from mmWave radars to track and identify multiple people. The sparse point clouds show the 3D location of the reflection points and indicate the people's information. mID first merges the point clouds into clusters by DBSCAN [48] clustering algorithm. Considering that the distribution of the people's point cloud on the z-axis is more dispersed than that on the x-axis and y-axis, mID modifies the Euclidean distance to place less weight on the contribution from the z-axis. After that, mID employs the Hungarian algorithm [72] to associate between each cluster at the current timestamp and the track obtained before, which is a classic many-to-many assignment problem. mID further utilizes a Kalman filter [73] to correct tracking errors. Finally, the points of potential human objects are voxelized by a fixed-size bounding box to form an occupancy grid. mID passes it into a bi-directional LSTM network [74] for final people identification. mID is developed on top of a COTS mmWave

radar, TI IWR1443boost [3]. The experimental results show that mID achieves a median position error of 0.16 m and identification accuracy of 89% for 12 people.

When people are too close to be distinguished by mmWave radar, the trajectory crossover problem arises and degrades the tracking accuracy. Some works have attempted to handle this problem. **PALMAR** [69] applies the Adaptive Order Hidden Markov Model (AO-HMM) and the Crossover Path Disambiguation Algorithm (CPDA) to handle multiperson path ambiguity and trajectory crossover [75]. It achieves an overall error of 11.5 cm in tracking multiple persons and outperforms mID with an overall 57.4% improvement.

In addition to tracking and locating the human body, there are also some works to locate the human parts directly. **WaveEar** [76] utilizes a customized 24GHz mmWave probe with 16 antennas to scan in all directions. It extracts all the time domain features to detect the throat's location. **RadioMic** [47] leverages the symmetry of the Doppler shift caused by sound vibration to locate the sound source directly. **RF-SCG** [61] estimates the optimal 3D location of the heart by combining the estimated heart rate with the time-domain spectral properties extracted by FFT and beamforming techniques.

B. Motion Recognition

Human motion recognition plays an important role in a wide range of real-world applications. As the reflected mmWave signals usually carry substantial information about the human object, human motion can be well recognized with mmWave sensing. Not only can human activities or gestures be distinguished, but the handwriting can be accurately tracked. We summarize the related works in TABLE VI.

1) Activity recognition: Human activity recognition (HAR) has received continuous attention due to its wide applicability in actual scenes. Based on the high spatial resolution of mmWave signals, many researchers have begun to explore mmWave-based human activity recognition.

The core challenge of activity recognition is to find appropriate and environmental-independent features to represent activity-related information in reflected signals. The CIR frequency response, micro-Doppler spectrum and voxelized point clouds have been exploited as activity-related features in the

TABLE VI
COMPARISON OF MOTION RECOGNITION WORKS

Category	Method	Signal Form	Algorithm	Hardware Platform	Performance
Activity Recognition	**EI** [77]	CIR	CNN, Unsupervised Domain Adversarial Training	COTS 60GHz mmWave Transceiver	65% Accuracy
	SPARCS [55]	CIR	Joint Probabilistic Data Association Filter, Iterative Hard Thresholding	mm-FLEX Platform	90% Accuracy
	RadHAR [29]	Point Cloud	Time-distributed CNN, BiLSTM	TI IWR1443BOOST	90.47% Accuracy
	m-Activity [78]	Point Cloud	Fast-calculation CNN, Less-parameter RNN	TI IWR1443BOOST	93.25% of Off-line 91.52% of Real-time
	PALMAR [69]	Point Cloud	DBSCAN and BIRCH Clustering, Adaptive Order Hidden Markov Model, Variational Autoencoder	79GHz mmWave Radar	91.88% Accuracy
Gesture Recognition	**Soli** [2]	Range-Doppler, Spectrum	Bayesian Filter	Soli Chip	92.1% Accuracy
	RFWash [79]	Range-Doppler, Spectrum	BiLSTM	TI IWR1443BOOST	92.59% Accuracy
	mmASL [80]	Doppler Spectrum	STFT and Log Normalization, CNN, Multitask Learning	NI+SiBeam 60GHz FPGA Software Radio	87% Accuracy
	mmPose [27]	Point Cloud	3D Heatmap Conversion Forked CNN	TI AWR1642BOOST	3.2 cm,7.5 cm,2.7 cm Error in Depth, Elevation and Azimuth
	mHomeGes [33]	Point Cloud	CNN, Ghost Image Separation, Hidden Markov Model-based Voting Mechanism	TI IWR1443BOOST	97.96% Accuracy
	Pantomime [34]	Point Cloud	Pointnet++, LSTM	TI IWR1443BOOST	95% Accuracy
Handwriting Tracking	**mTrack** [35]	RSS/Phase	Discrete Beam Steering, Dual-differential Background Removal	Customized 60GHz Software-radio Testbed	90-percentile Tracking Error below 8 mm
	mmWrite [81]	CIR	Background Subtraction, Discrete Cosine Transform	Qualcomm 60GHz 802.11ad Chipset	2.8 mm Median Error
	mmKey [82]	CIR	Adaptive Background Cancellation, MUSIC	Qualcomm 60GHz 802.11ad Chipset	95% Accuracy for Single-key 90% Accuracy for Multi-key

existing works. Moreover, learning-based techniques such as domain adversarial technique and domain adaptation technique have been applied to resist the impact of the environment in feature extraction and classification.

EI [77] proposes an environment-independent activity recognition framework. It utilizes a COTS 60GHz mmWave transceiver system equipped with a 24-element phased antenna array to collect multi-user activity data from different environments. EI first obtains the CIR measurements and transforms each CIR measurement into a frequency response sample. These frequency responses in a period of time then form an input matrix. EI employs CNN to extract activity features from the input matrix. After that, a fully-connected layer and a softmax layer are used to obtain the activity probability vector. In particular, EI adopts the unsupervised domain adversarial training technique [83] with unlabeled data to produce environmental-independent activity features. It further proposes three constraints to tackle the overfitting problem and improve the recognition performance, including confident control constraint, smoothing constraint and balance constraint. The experimental results show that EI achieves a

recognition accuracy of about 65% in different environments.

When people move, the different movements of the body parts cause different frequency modulation on the reflected signal, called the micro-Doppler effect. This feature is highly related to human movement and can be exploited to recognize the human activity. **SPARCS** [55] explores the micro-Doppler spectrum extracted from CIR samples for activity recognition. It reuses existing communication traffic and provides an integrated human sensing and communication (ISAC) solution.

After obtaining the CIR estimation from communication traffic, SPARCS first tracks each person's distance and angular position by performing peak detection, angle of arrival (AoA) estimation and Joint Probabilistic Data Association Filter (JPDAF) [84]. Due to the bursty and irregular traffic pattern, the CIR samples are sparse and irregular. In this case, the micro-Doppler spectrum can not be directly obtained by performing STFT, which needs uniform samples. To recover the micro-Doppler spectrum from such irregular and sparse CIR samples, SPARCS resamples the random CIR samples to obtain the regularly spaced samples and missing values with a fixed interval. Then the recovery of the micro-Doppler spec-
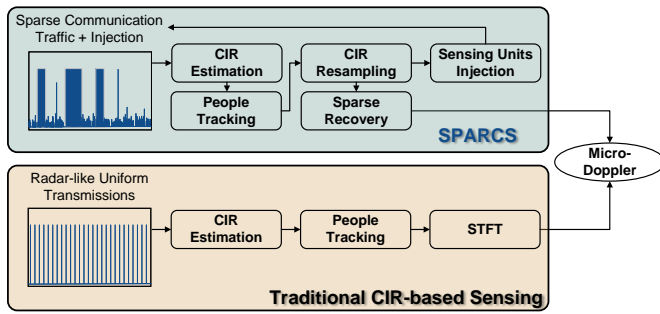
Fig. 7. Comparison between the traditional CIR-based human sensing and SPARCS.

Fig. 8. Soli: the first mmWave radar gesture sensing system.

trum from such incomplete CIR samples can be formulated as a sparse recovery problem and solved with the Iterative Hard Thresholding (IHT) algorithm [85]. Furthermore, when the communication traffic is too sparse to recover the accurate micro-Doppler spectrum, SPARSE injects short sensing units to provide enough CIR measurement. SPARSE is developed on top of the open-source mm-FLEX platform [18] with 60GHz mmWave RF front-ends. The experimental results show that SPARCS reaches over 0.9 F1 scores on all four activities.

In addition to the frequency response and the micro-Doppler spectrum, point clouds are also exploited to achieve activity recognition. **RadHAR** [29] explores how to perform accurate HAR using point clouds generated through a mmWave radar. RadHAR first captures the point clouds which contain spatial coordinates, velocities, ranges, intensities and angles. The point clouds are then converted into 3D voxels to tackle the non-uniformity of points number in each frame. The value of each voxel is the number of data points within its boundaries. After that, RadHAR utilizes a sliding time window to accumulate point clouds to overcome the sparsity of points number. Finally, RadHAR evaluates different classifiers and selects the time-distributed convolutional neural network (CNN) + bidirectional LSTM classifier as the best-performing classifier. The experimental results show that RadHAR achieves a recognition accuracy of 90.47% with five different activities.

Considering the applicability of activity recognition in the real world, **m-Activity** [78] enables a real-time activity recognition system in noisy environments. It removes the noise points based on the density of point clouds and extracts the point clouds of human activity. Then the 3D data is accumulated as an integral temporal flow to reduce recognition cost. m-Activity further combines a fast-calculation CNN with a less-parameter recurrent neural network (RNN) to classify different activities. It achieves an off-line activity accuracy of 93.25% and real-time activity accuracy of 91.52%.

PALMAR [69] further exploits the point clouds to track and recognize multi-person activities simultaneously. It first collects the point clouds generated from a 79GHz mmWave radar and represents them in voxel format. Then PALMAR employs DBSCAN and BIRCH [86] clustering algorithm to cluster these voxels of each people. PALMAR further applies the Adaptive Order Hidden Markov Model (AO-HMM) to track multiple targets and the Crossover Path Disambiguation Algorithm (CPDA) to handle multiperson path ambiguity and
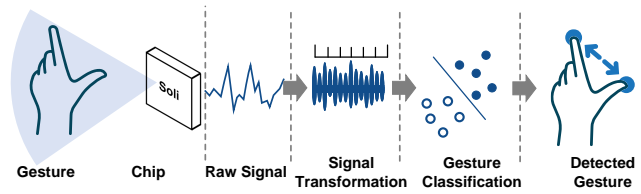
trajectory crossover [75]. Finally, PALMAR proposes a deep domain adaptation model based on a variational autoencoder to improve activity recognition accuracy and domain adaptation. It achieves a recognition accuracy of 91.88% with domain adaptation in multi-user scenarios.

2) Gesture recognition: "In air" gesture recognition using mmWave signals has shown its potential in human-computer interaction and health monitoring. Various gesture recognition-based applications have been developed, including fine gesture interaction, hand hygiene monitoring, sign language communication, skeletal posture estimation and so on.

Similar to activity recognition, gesture recognition requires extracting appropriate features from the received signals. To do so, researchers have explored the potential of Range-Doppler spectrums, Doppler spreads, point clouds and concentrated position-Doppler profiles. Furthermore, as gesture recognition requires more fine-grained analysis and sensing, compared with activity recognition, various feature extraction algorithms and classification approaches have been proposed, including bidirectional LSTM, forked CNN, hidden Markov model-based voting mechanism and so on.

Soli [2] is the first mmWave radar gesture recognition system developed by Google. As shown in Fig. 8, it achieves ubiquitous fine gesture interaction and has been integrated into smart devices such as Google Pixel 4 [1]. Soli customizes its own radar chips in 12×12 mm and 9×9 mm scales. Each chip operates at the 60 GHz band with 7 GHz bandwidth. It also has two transmit and four receive antenna elements to support digital beamforming.

To enable ubiquitous gesture recognition, Soli first proposes the scattering center model of the human hand. It models the RF response of the hand as a superposition of the response of discrete scattering centers on the hand. Then Soli performs slow-time processing and fast-time processing on the received signal to produce the Range-Doppler spectrum. It further projects the Range-Doppler spectrum into the range profile and the Doppler profile for feature extraction. In the feature extraction phase, Soli extracts a variety of features, including explicit scattering center tracking features, low-level descriptors of physical RF measurement and data-centric machine learning features. Finally, different machine learning classifiers can be used to identify the user's gesture. The experimental results demonstrate that Soli can achieve a recognition accuracy of 92.1% with a Bayesian filter.

RFWash [79] proposes a mmWave-based gesture sensing approach to monitor the nine-step Alcohol-Based Hand Rub technique. As shown in Fig. 9, unlike popular gesture recognition approaches which contain the detection/segmentation
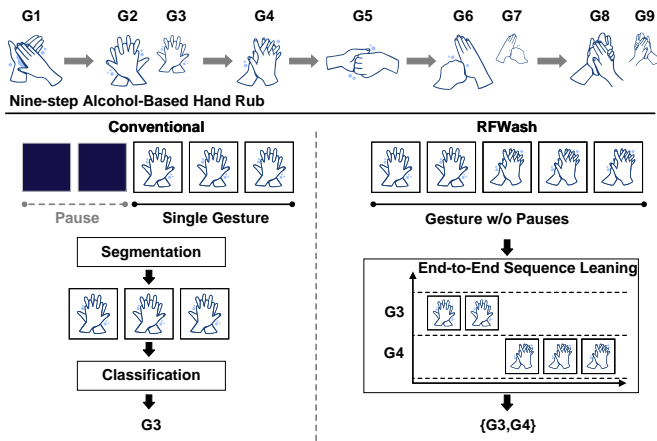
Fig. 9. Gesture sequence recognition of RFWash.

step and the recognition step, RFWash demonstrates that data segmentation is a tough task and seriously affects gesture classification accuracy. Inspired by speech and handwriting recognition approaches, RFWash proposes a segmentation-free approach to accurately recognize hand hygiene gestures without segmentation.

RFWash mounts the mmWave radar on a soap dispenser and extracts the Range-Doppler spectrum at each time instant. It further limits the range to less than 1m to remove interference from other people. To predict the gesture sequence from the Range-Doppler spectrums, RFWash first employs bidirectional recurrent layers with LSTM cell type (BiLSTM) [87] to extract the gesture-related spatiotemporal features. Then it adopts temporal alignment learning [88] to map the output of BiLSTM to the corresponding gesture sequence. RFWash further employs order-preserving concatenation to augment training data and extends the data size quadratically. The results show that RFWash achieves a gesture error rate of less than 8% and reduces manual labeling overhead by about 67%.

Sign language communication is another major application of gesture recognition. To serve the Deaf and Hard-of-Hearing (DHH) community, **mmASL** [80] proposes a home assistant system that can recognize American Sign Language (ASL) using 60GHz mmWave signals. mmASL utilizes a sinusoid of 1MHz as the baseband signal and extracts the corresponding Doppler spread of the received signal, which is caused by the movements of the body parts. After obtaining the received signal, a low-pass 1KHz filter and a high-pass 10Hz filter are first applied to the received signal for denoising. Then the spectrums are plotted by STFT and enhanced by log normalization. mmASL then utilizes these spectrums to achieve wake-word recognition and ASL sign recognition. Specifically, mmASL employs beam scanning through a set of beam sectors and performs STFT for each sector's data. These results are concatenated to form the spatial spectrograms. Then mmASL employs a CNN-based machine learning model to recognize the wake word quickly. It further leverages a multitask learning network to extract and learn the ASL-related features to achieve ASL recognition. mmASL is implemented on NI+SiBeam 60GHz multi-FPGA software radio platform

[89] with a phased antenna array. The experimental results demonstrate that mmASL can detect wake-word with an average accuracy of 94% and achieves an average sign recognition accuracy of 87%.

mmPose [27] exploits the point clouds generated from the mmWave signals to achieve real-time human skeletal posture estimation. It uses TI AWR1642BOOST board [7] to collect the reflected signal and build the 3D point clouds. Then mmPose assigns an RGB weighted pixel value to each point based on its reflection intensity, resulting in a 3D heatmap. mmPose further compresses this heatmap into two low-resolution planes, the depth-azimuth plane and the depth-elevation plane, to overcome the sparsity of the heatmap and reduce unnecessary computing costs. Finally, a forked CNN architecture is used to output the skeletal joint coordinates of the human. The experimental results show that it accurately predicts human motion with four different gestures.

It is desirable to perform gesture recognition in smart home scenarios. However, the surrounding interference can prevent the mmWave radar from discovering users and thus make gesture recognition unavailable. **mHomeGes** [33] attempts to overcome this challenge and achieves a real-time arm gesture recognition system. mHomeGes first captures a series of point clouds with a fixed-length sliding window. Then it extracts the concentrated position-Doppler profile (CPDP) that compresses the intensity of each point into the distance dimension and the Doppler dimension. It can be regarded as the sum in the time dimension of the Range-Doppler spectrums. Then CPDP is used as the input of a customized CNN recognition model to recognize fine-grained gestures. To eliminate the multipath effect caused by the surrounding reflectors, mHomeGes proposes a novel ghost image separation algorithm using the velocity synchrony between the user and the corresponding ghost image. Finally, mHomeGes employs a hidden Markov model-based voting mechanism (HMM-VM) to achieve continuous gesture recognition. mHomeGes is implemented based on a commercial mmWave radar, TI IWR1443 board [5]. Results show that mHomeGes achieves a recognition accuracy of 97.96% for 25 volunteers across five home scenes.

Another challenge of gesture recognition using point clouds is the sparsity of the point clouds. It makes gesture-related signals challenging to be distinguished. **Pantomime** [34] proposes a novel hybrid model to achieve accurate gesture recognition with sparse point clouds. Pantomime first aggregates the point clouds of a gesture segment to obtain more points in a single frame. Then the aggregated point cloud is resampled to meet a fixed point number with upsampling and downsampling algorithms. To further extract the spatio-temporal features from the aggregated point clouds, Pantomime employs a hybrid architecture combining the Pointnet++ architecture [90] and the LSTM modules. Pointnet++ is designed to extract spatial features from 3D point clouds, and the LSTM modules continue to extract temporal features from obtained features. Finally, these two feature vectors are concatenated to form the final feature vector and classified by a fully connected layer. The experimental results show that Pantomime achieves 95% accuracy and 99% the area under the ROC curve (AUC) with 21 gestures.
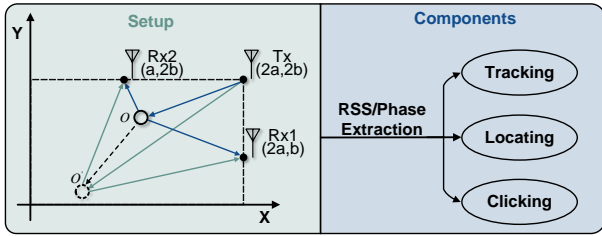
Fig. 10. mTrack setup and system components.

3) Handwriting tracking: Passive handwriting tracking is a critical task in performing ubiquitous human-computer interaction. Benefiting from the short wavelength and large bandwidth of mmWave signals, researchers have proposed mmWave-based handwriting tracking approaches that can achieve mm-level tracking accuracy.

Different from activity recognition and gesture recognition, handwriting tracking requires accurate quantitative tracking of the writing object, such as a finger or a pen. The micro displacement measurement based on mmWave signals relies heavily on the phase changes of the received signal. However, the mmWave signals also suffer from background interference that distorts the phases of the mmWave signals. How to extract the phase changes caused by micro displacement becomes the key to achieving accurate handwriting tracking. Researchers have exploited beamforming algorithms and signal-denoising techniques to eliminate background interference and extract handwriting-related phase changes.

mTrack [35] proposes a high-precision handwriting tracking with 60GHz mmWave radios. The mTrack setup and system components are shown in Fig. 10. The transmitter is equipped with a quasi-omni beam pattern to illuminate the tracking region, and the receiver adopts highly directional antennas to boost the target-reflected signals. With three well-designed modules, namely the anchor point acquisition module, the phase tracking module and the touch detection module, mTrack can achieve handwriting detection and tracking.

mTrack extracts the received signal strength (RSS) and phase changes from each receiver antenna to locate and track the writing object. It first performs discrete beam steering techniques to obtain the discrete RSS samples and captures the relative angle to each receiver antenna with the interpolated maximum RSS. Then mTrack continuously tracks the writing object by extracting the phase changes. As the phase changes of the received signals are distorted by background interference, directing the conversion of the phase changes to moving distance will result in incorrect tracking results. mTrack proposes a dual-differential background removal (DDBR) algorithm to remove such an impact. Specifically, it differentiates the phase of three consecutive samples to obtain a single-phase shift sample. mTrack also designs a phase counting and regeneration algorithm to complement the DDBR algorithm. Furthermore, mTrack utilizes the RSS measurement and the phase variance to detect the touch events. mTrack is implemented on a customized 60GHz software-radio testbed with the Vubiq 60GHz RF front-end. Results show a 90-percentile tracking error below 8 mm.

Instead of using multiple separate receiver antennas, **mmWrite** [81] enables a high-precision handwriting tracking system with a single mmWave radio. mmWrite is implemented based on a Qualcomm 60GHz 802.11ad chipset with an additional antenna array. mmWrite exploits the CIR measurement from the receiver for handwriting tracking. Specifically, mmWrite first employs the background subtraction algorithm to the recorded CIRs to eliminate the impact of other static objects. Then the digital beamforming technique is performed to determine the range and direction of the reflected signals. mmWrite further extracts the Doppler velocity by STFT and performs 3D constant false alarm rate (CFAR) [43] target detection based on the Doppler power. This series of target locations construct an estimated trajectory of the target. Finally, mmWrite proposes a smoothing technique based on Discrete Cosine Transform (DCT) to enhance the estimated trajectory. The experimental results show that it achieves a median tracking error of 2.8 mm.

Similarly to mmWrite, **mmKey** [82] enables a universal virtual keyboard with a single mmWave radio. mmKey first utilizes the differential amplitude of the CIR to detect the presence of motions. Then mmKey employs a novel motion filter and adaptive background cancelation to extract only the finger-related reflections. It further performs the multiple signal classification algorithm (MUSIC) [91] to enable precise keystroke localization. Results show that mmKey can achieve a keystroke recognition accuracy of more than 95% for single-key cases and more than 90% for multi-key scenarios.

C. Biometric Measurement

Human biometric measurement is an important step to achieve ubiquitous human sensing. Through mmWave-based sensing of gait, vital signs or sounds, people can achieve health monitoring, human-computer interaction and other applications. We summarize the related works in TABLE VII.

1) Gait recognition: Gait-based human identification is a key requirement for secure and smart applications. Compared with visual-based solutions, mmWave-based identification systems are more acceptable, as the concerns about privacy security can be eliminated. With the strong directivity of mmWave signal, they are capable of tracking and identifying multiple targets simultaneously.

The key problem of gait recognition is to define appropriate feature forms to effectively extract gait-related features from the raw data. Considering that gait-related features include the temporal, spatial and velocity variation from the reflected signal. Point clouds and Range-Doppler spectrums are widely used because they maintain rich gait-related information. Recent works further propose to use the micro-Doppler spectrum to detailedly describe the human gait.

mmGait [32] proposes a deep-learning driven method to achieve mmWave gait recognition. mmGait uses two mmWave devices, TI IWR1443 and IWR6843, to build a mmWave gait dataset from 95 volunteers. It first employs DBSCAN algorithm and Hungarian algorithm to obtain the point clouds corresponding to every single human. it further proposes a customized neural network structure, called mmGaitNet, to

TABLE VII
COMPARISON OF BIOMETRIC MEASUREMENT WORKS

| Category	Method	Signal Form	Algorithm	Hardware Platform	Performance
Gait Recognition	**mmGait** [32]	Point Cloud	DBSCAN Clustering Hungarian Algorithm, DNN	TI IWR1443BOOST TI IWR6843BOOST	90% Accuracy for Single-person 88% Accuracy for Five-person
	MU-ID [92]	range-Doppler Spectrum	CNN	TI AWR1642BOOST	97% Accuracy for Single-person 92% Accuracy for Four-person
	GaiCube [93]	micro-Doppler Spectrum	Log Spectrum Analysis CNN	TI IWR1443BOOST	98.3% Accuracy
Vital Sensing	**mmVital** [94]	RSS	Band-pass Filters	Vubiq Platform	0.43 Bpm Error 2.15 bpm Error
	ViMo [22]	CIR	2D-CFAR Autocorrelation	Commercial 60-GHz WiFi	0.19 Bpm Error 0.92 bpm Error
	mBeats [95]	Phase Waveform	Biquad Cascade IIR Filter DNN	TI IWR6843BOOST	95.26% Accuracy
	RF-SCG [61]	Phase Waveform	4D Cardiac Beamformer CNN Translator	TI IWR1443BOOST	0.72 Correlation Coefficient
	mmHRV [50]	Phase Waveform	Modified VMD	TI IWR1443BOOST	97.96% Accuracy
	mmECG [96]	Phase Waveform	Hierarchy VMD Template-based Optimization	TI AWR1642BOOST	0.37 bpm Error
	HeartPrint [97]	Phase Waveform	WPT	TI IWR1443BOOST	96.16% Accuracy
	[98]	Phase Waveform	LSTM	TI AWR1243Boost	5.57 Bpm Error for Moving 3.32 bpm Error for Moving
	Movi-Fi [99]	Phase Waveform	Deep Contrastive Learning Encoder-decoder Model	TI IWR1443BOOST	about 2% Bpm Error about 1% Bpm Error
	VED [100]	Phase Waveform	Variational Encoder-decoder	TI IWR1843BOOST	above 0.92 Similarity
	CardiacWave [101]	Phase Waveform	Mask Filter DNN	TI IWR1443BOOST	97.96%
	mmBP [102]	Phase Waveform	Functional Link Adaptive Filter	TI IWR1843BOOST	0.87 mmHg SBP Error 1.55 mmHg DBP Error
Sound Recognition	**WaveEar** [76]	Phase Waveform	DNN Griffin-Lim Phase Reconstruction	Customized 24GHz mmWave Probe	Lower than 6% WER
	VocalPrint [103]	Phase Waveform	Resilience-aware Clutter Removal	TI AWR1642BOOST	96% Accuracy
	mmPhone [104]	Phase Waveform	DNN Training-free Harmonic Extension	TI AWR1843BOOST	93% Accuracy
	Wavedropper [105]	Phase Waveform	Wavelet-based Analysis RNN	TI IWR1642BOOST	91.3% Accuracy
	MILLIEAR [63]	Phase Waveform	Conditional GAN	TI IWR1642BOOST	3.68 MCD
	Wavoice [106]	Phase Waveform	Multi-modal Signals Fusion	TI IWR1642BOOST	0.69% CER 1.72 WER
	AmbiEar [66]	Phase Waveform	Modified MVDR End-to-end Network	TI IWR1642BOOST	87.21% Accuracy for NLoS
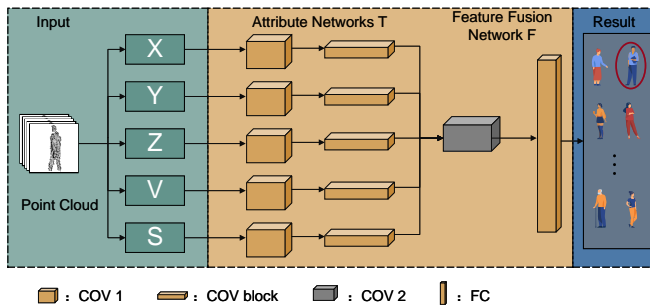
Fig. 11. The overview of mmGaitNet.

distinguish the point cloud. The network structure is shown in Fig. 11. The inputs of mmGaitNet are point clouds' five attributes X, Y, Z, V and S, where X, Y, Z represent the spatial location of each point, V and S denote the correspond-

ing radial speed and signal strength, respectively. The attribute features are first extracted by the attribute network, then they are input into the feature fusion network to fuse the features. Finally, the final fully connected layer output the class score. The experimental results show that mmGait achieves 90% identification accuracy for single-person scenarios and 88% accuracy for five co-existing persons.

In addition to using the distribution of point clouds to recognize different humans, researchers also capture the signal frequency changes caused by human motion to achieve gait recognition. **MU-ID** [92] focuses on the gait feature from the lower limb motion in the spatial-temporal domain to achieve multi-user identification. MU-ID first converts the raw data into the Range-Doppler spectrum and removes the stationary interference. Then the Range-Doppler spectrum is compressed along the velocity dimension. A sequence of the compressed result is further arranged in the order of time to extract

the lower limb motion features. When multiple users appear simultaneously, their features are separated based on AoA difference and segmented based on silhouette analysis. Finally, MU-ID develops a deep CNN-based classifier to output the classification probability distribution. The experimental results show that MU-ID achieves 97% single-person identification accuracy and over 92% accuracy for four people.

Compared with velocity analysis, micro-Doppler spectrum indicates more detail local motion information and can be used for gait recognition. Inspired by this observation, **GaitCube** [93] proposes to extract a joint-feature representation of micro-Doppler and micro-range signatures over time, called gait data cube to embody the physical relative features of human gait. Then GaitCube can achieve accurate gait classification with the gait cubes. Specifically, GaitCube first obtains the time-range spectrum from raw data and chooses the maximum-variance trace to detect a subject. Then the human's walking phase can be detected with speed extraction and threshold comparison. After that, GaitCube estimates the gait cycle by extracting the Time-Doppler log spectrum energy and segments each gait cycle. The micro-Doppler spectrums over the range in each gait cycle are further downsampled and aligned to form the gait cubes. Assisted by several additional features such as trace length and cycle duration, the gait cube is fed into a CNN-based classification for gait recognition. The experimental results show that GaitCube achieves an accuracy of 96.1% with a single gait cycle with one RX antenna and the accuracy increases to 98.3% with all RX antennas.

2) Vital sensing: Non-contact continuous monitoring of human vital signs is one of the important applications of mmWave sensing. Through the continuous in-depth analysis of the vital information contained in the mmWave reflected signals, not only the vital feature can be extracted but also the vital waveform can be recovered.

The core challenge of vital sensing lies in the fine-grained characterization of the time-domain vital-related signal changes and exact vital signal extraction from such signals. When the sensing target is vital features such as breathing and heart rate, frequency domain analysis or learning-based feature extraction methods are widely adopted. When the sensing target spreads to the vital waveform, learning-based waveform extraction methods or signal decomposition technologies begin to play an important role.

mmVital [94] first explores how to enable continuous monitoring of human breathing and heart rates with 60GHz mmWave signals. mmVital locates the users based on the reflection loss of the human body, which is different from other objects. Then mmVital adjusts Tx and Rx angles to transmit the signal to the human body and extract the receiver signal strength (RSS) of the reflected signal. The RSS signal is sensitive to periodic movement of the human body, resulting in a peak in the frequency domain. To achieve better accuracy, mmVital detects the corresponding frequency peaks with band-pass filters to estimate the breathing and heart rates. mmVital utilizes a mmWave development platform provided by Vubiq [70] for transceiver angle adjustment and RSS value collection. The experimental results show that mmVital provides reliable vital sensing with a mean estimation error of 0.43 breaths per

minute (Bpm) and 2.15 beats per minute (bpm).

As the channel impulse response (CIR) can obtain more detailed vital-related information than the RSS, **ViMo** [22] try to extract the vital features with the CIR time-series. It enables multiperson vital sign monitoring using a commercial 60GHz WiFi. Due to the extremely high frequency and large bandwidth of the mmWave signal, 60GHz WiFi can offer high directionality with large phased arrays in small size. When the mmWave signal is reflected by the human body, the phase of the CIR measurement changes periodically due to periodic human motions, i.e. breath and heartbeat. ViMo first employs 1D-CFAR in range dimension and 2D-CFAR in angle dimension to detect reflecting objects. Then humans can be distinguished from the static reflecting objects according to the variation of their corresponding phases. ViMo estimates the breathing rate by applying the autocorrelation function to the CIR phase and finding the peak location. After that, ViMo adopts a smooth spline to estimate and eliminate the breathing signal. Finally, the enhanced heartbeat signal is utilized to continuously estimate the heart rate with dynamic programming. ViMo is embedded in a commodity 60GHz WiFi [20], which has 32 antennas assembled in a 6×6 layout. The median errors are 0.19 Bpm and 0.92 bpm in the case of a single user, and the mean accuracy of both breathing rate and heart rate is over 92.8% in the case of 3 users.

Besides frequency domain analysis, learning-based feature extraction methods are also applied to extract vital features. **mBeats** [95] provides a heart rate monitoring system that is insensitive to user orientation and position. mBeats utilizes a domestic service robot equipped with a mmWave radar to track the human's orientation and position and perform vital sensing. As the service robots are typically low in height, mBeats measures the heart rate in a user's leg. The measured phase variation is caused by the micro displacement of the user's skin but filled with noise. To accurately estimate the heart rate, mBeats first leverages a biquad cascade IIR filter to extract heartbeat waveforms. Then the heart rate estimation problem is formulated as a regression problem and solved by a customized deep neural network (DNN) predictor. mBeats utilize a commercial radar, TI IWR6843ISK board [4], to send mmWave signals and extract the phase variation. The experimental results show that mBeats provides a heart rate estimation accuracy of 95.26% under 8 different user poses.

mmWave sensing can capture the subtle displacement of the human chest, which is mainly caused by the superposition of the heartbeat signal and respiration signal. Inspired by this observation, researchers begin to explore how to recover the vital waveform rather than just extract vital features based on mmWave sensing.

RF-SCG [61] introduces how to utilize the reflection of the mmWave radar signal to reconstruct the seimocardiogram (SCG) waveform, which captures fine-grained cardiovascular events. To sense the cardiac micro-vibrations, RF-SCG first needs to exactly capture the reflected mmWave signal from the apex of the heart. RF-SCG designs a 4D Cardiac Beamformer component to discover the 3D location of the heart. The detailed architecture is shown in Fig. 12. RF-SCG first formulates the heart rate estimation problem as
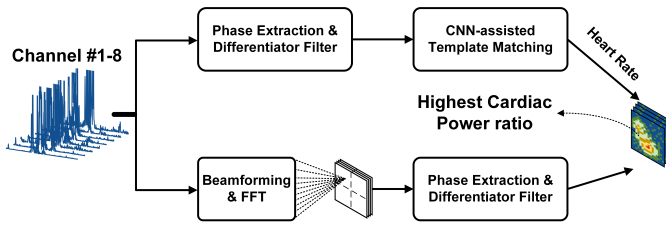
Fig. 12. The architecture of 4D Cardiac Beamformer.

a 1D matching problem and solves it with a CNN-assisted template matching technical. Then RF-SCG applies FFT and beamforming techniques to the received signal to extract the corresponding time-domain signals in 3D bins. Combining the estimated heart rate with the time-domain spectral properties, RF-SCG can estimate the optimal 3D location of the heart.

After obtaining the reflection signal from the heart, RF-SCG learns a CNN-based RF-to-SCG translator to recover SCG waveforms from the mmWave reflections. Moreover, RF-SCG modifies the Unet architecture [107] to automatically label five fiducial points, including aortic valve opening, aortic valve closing, mitral valve opening, mitral valve closing, and isovolumetric contraction. RF-SCG is implemented based on a commercial mmWave radar, TI IWR1443BOOST board [3]. Results show that RF-SCG has more than 0.72 correlation coefficient between each subject's ground truth and the recovered waveform. It also can robustly time five key cardiovascular events with a median error between 0.26%-1.29%.

In addition to directly recovering the vital waveform from the reflected signal, researchers explore extracting the vital signals with different signal decomposition technologies. **mmHRV** [50] utilizes a modified VMD algorithm to extract the heartbeat signal from the reflected signal and monitor the heart rate variability (HRV). Similarly, **mmECG** [96] designs a hierarchy VMD approach to estimate the heart movement in mmWave signals. It further leverages a template-based optimization method to reconstruct the cardiac cycle in driving environments. **HeartPrint** [97] analyzes the minor changes in heartbeat waveform with a 3-level wavelet packet transform (WPT), which can conduct multi-resolution analysis in multiple frequency domains. [98] designs customized LSTM models to estimate a baseline of heart rate and respiration rate separately only when the human is static. It further predicts these features on top of the baseline when the human moves.

The fundamental assumption of the signal decomposition algorithm is the linear superposition of signals. However, **MoVi-Fi** [99] states a different point of view: The reflected signals caused by vital signs are composited with other motion-incurred reflections in a nonlinear manner. To solve such a nonlinear blind source separation problem, MoVi-Fi employs deep contrastive learning to reverse the nonlinear composition between body movement and vital signs. Then an encoder-decoder model is utilized to recover the heartbeat waveform and the breathing waveform. The experimental results show that MoVi-Fi can recover fine-grained vital signs waveforms under severe body movements. **VED** [100] also claims that the composition between respiration and heartbeat can be highly nonlinear. VED explores the ability of the variational encoder-decoder, a novel deep neural network, to convert the raw data to the heartbeat waveform. It achieves a heart rate estimation median error less than 2.4% and a median waveform cosine similarity higher than 0.92.

Different from these works, **CardiacWave** [101] recovers vital signs by analyzing the electromagnetic (EM) field changes rather than the chest movements brought by the heart. CardiacWave presents a novel and interesting observation: the electromagnetic field induced by cardiac electrical activity will modulate the chest-scattered mmWave signals, called Cardiac-mmWave scattering effect (CaSE). Specifically, the frequency spectrums of the intermediate frequency (IF) signal are different when the heat activity changes. CardiacWave employs a mask filter with learnable coefficients to extract the CaSE feature from the IF signal. Then a DNN-based profiler is developed to obtain the ECG-like signals and perceive the cardiac events. Results show that the recovered waveforms have a high positive correlation with the ground truth and contain high-fidelity heart clinical characteristics.

Apart from the aforementioned applications, mmWave sensing also has great potential in blood pressure (BP) monitoring. To measure BP values, mmWave reflections are processed to reconstruct fine-grained pulse waves which contain unique BP-related features, e.g., minimum value, peak value, and first inflection. However, due to high frequency and short wavelength, mmWave signals in the time domain are susceptible to noise, degrading the performance of pulse waveform reconstruction and BP measurement. Moreover, the tiny body motion is another concern for mmWave-based BP monitoring, which would cause severe non-linear signal distortions and lead to a huge performance drop in BP monitoring.

To address the above challenges, **mmBP** [102] develops a contact-free BP monitoring system to achieve high accuracy and tiny-motion robustness. To reduce noise, mmBP first converts the received mmWave signals from time domain to the delay-Doppler (DD) domain, and then retains the pulse-related information but filters out noise by leveraging on their different properties in the DD domain. To alleviate the impact of tiny motion, mmBP develops a functional link adaptive filter (FLAF) by exploiting the periodic property and correlation character of the pulse signals. mmBP is implemented with an off-the-shelf mmWave radar (TI IWR1843BOOST), and results demonstrate that mmBP achieves the mean errors of 0.87 mmHg and 1.55 mmHg for systolic blood pressure (SBP) and diastolic blood pressure (DBP), respectively, and the standard deviation errors of 5.01 mmHg and 5.27 mmHg for SBP and DBP, respectively. Results meet the acceptable error range specified in the Association for the Advancement of Medical Instruments (AAMI) standard and rank Grade A in the Britain Hypertension Society (BHS) standard.

3) Sound recognition: Sound recognition is another important application field of mmWave sensing. Conventional solutions based on microphone(s) often fail in noisy environments or outside soundproof scenes. Thanks to the penetrability and high directivity of mmWave signals, people can directionally capture the voice-related reflected signals by mmWave sensing to achieve sound recognition or eavesdropping.
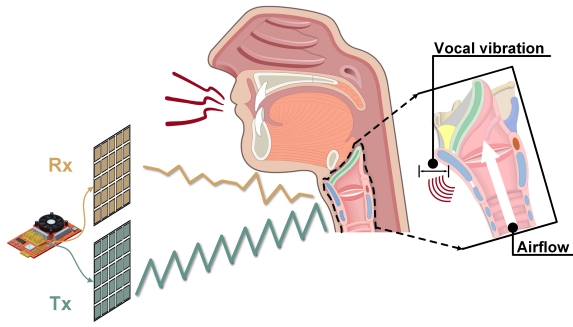
Fig. 13. The reflected mmWave signal indicates speech information.

Different from gait recognition or vital sensing, sound recognition mainly extracts the time-frequency spectrum from the sound-related signal to reconstruct the sound. This means that sound recognition is more sensitive to the complex noise in the frequency domain, which may come from the surrounding acoustic field, electromagnetic interference or even imperfect hardware. Moreover, as the sensing target is the tiny near-throat region, how accurately locate this vocal part becomes a key problem.

For denoising in the frequency domain, learning-based denoising methods are often applied as the complex noise overlaps with the speech frequency band. These methods can also help recover the severely distorted high-frequency part. To accurately locate the speaker's throat, the signal variance brought by throat vibration are often considered. Some works also solve this problem with the help of microphones or the surrounding environment.

WaveEar [76] first explores how to utilize the mmWave signal to achieve noise-resistant speech sensing. Its observation is that the vibration of the speaker's vocal cord contains speech information and can be measured on the throat's surface. As shown in Fig. 13, the skin-reflected mmWave signal has a strong correlation with human speech. WaveEar designs a customized 24GHz mmWave probe, where both Tx and Rx consist of 16 antennas following the 4×4 layout. With the help of such a powerful probe, WaveEar can easily scan in all directions and extract the time domain features to detect the throat's location. After that, the mmWave reflected signal is converted to a series of spectrograms and fed into a novel DNN to obtain the corresponding sound spectrograms. Finally, the Griffin-Lim phase reconstruction approach [108] is employed to recover the sound. Results show that WaveEar can achieve a mean Mel-Cepstral distortion (MCD) [109] lower than 1.5 and a word error rate (WER) lower than 6%.

With the directivity and small range resolution of mmWave sensing, Many noises can be eliminated. **VocalPrint** [103] enables a secure voice authentication system by exploiting the unique characteristics of the vibrating near-throat region. As the vocal cords and vocal tract vary from person to person, the voice pronunciation process can be regarded as a unique feature, just like the voice itself. VocalPrint first applies a resilience-aware clutter removal scheme on the Range-Doppler matrix (RDM) calculated from the reflected mmWave signals. In this way, the complex clutters can be moved from RDM and

the vocal biometric properties can be preserved. Furthermore, VocalPrint extracts a series of text-independent features (e.g., residual phase Cepstrum coefficients (RPCC) [53] and Mel frequency cepstral coefficient (MFCC) [54]) and feeds them into a classifier for user authentication. VocalPrint is implemented based on a commercial mmWave radar, TI AWR1642BOOST [7]. The experimental results show that it achieves over 96% authentication accuracy under different conditions.

The work on [110] also proposes a speaker verification system based on mmWave sensing. It utilizes the mmWave radar to capture both vocal cord vibration (VCV) and lip motion (LM) as multimodel biometrics for user verification. MFCC features and fuzzy WPT techniques are separately applied to extract VCV features and LM features. These features are further fed into a CNN for user verification.

Sound recognition based on mmWave sensing can be utilized not only for authentication but also for illegal eavesdropping. **mmPhone** [104] utilizes piezoelectric films and a mmWave radar to eavesdrop on the loudspeaker speech protected by soundproof environments. The properties of piezoelectric films can change with sound pressure and can be captured by the mmWave radar. In this case, the films can be converted to a passive "microphone" as long as they are in the sound field. To improve the signal-to-noise ratio (SNR) of the received signal, mmPhone employs a DNN-based denoising method to handle non-stationary noise. Finally, a training-free harmonic extension scheme is utilized to improve speech intelligibility. Different from mmPhone, **Wavedropper** [105] eavesdrops on speech contents by directly sensing the speaker's throat vibration in the obstructed condition. It applies wavelet-based analysis to extract the voice-related signal from the hybrid signal and an RNN-based classifier for speech recognition. The experimental results show that Wavedropper can achieve 91.3% accuracy for 57-word recognition.

As the reflected mmWave signal is seriously affected by various noises, such as environmental acoustic noise and electromagnetic noise, the SNR of the received signal constrains the recovered sound quality. To resist the impact of complex noises, **RadioMic** [47] first utilizes a line fitting algorithm to extract the phase change caused by the sound. Then RadioMic employs a selection combining scheme on these signals from multiple antennas and multipath components. Finally, a sound enhancement scheme via deep learning is proposed for neural bandwidth expansion and denoising. **MILLIEAR** [63] directly employs a conditional generative adversarial network (cGAN) [111] to enhance the audio components and reduce noise. It achieves a mean MCD of less than 4, which implies that the reconstructed audio is highly similar to the original speech.

Furthermore, mmWave-based sound recognition is constrained by another thorny problem: Due to the dynamics of the speaker's position and posture, the speaker's throat is hard to detect and locate in real-world scenarios, making the voice-related signal can not be accurately extracted and analyzed. **Wavoice** [106] explores the inherent correlation between the reflected signal from a mmWave radar and the audio signals collected from a microphone to achieve noise-resistant speech recognition. The collected audio signals, although noisy, can guide mmWave radars to detect the vibrating throat and
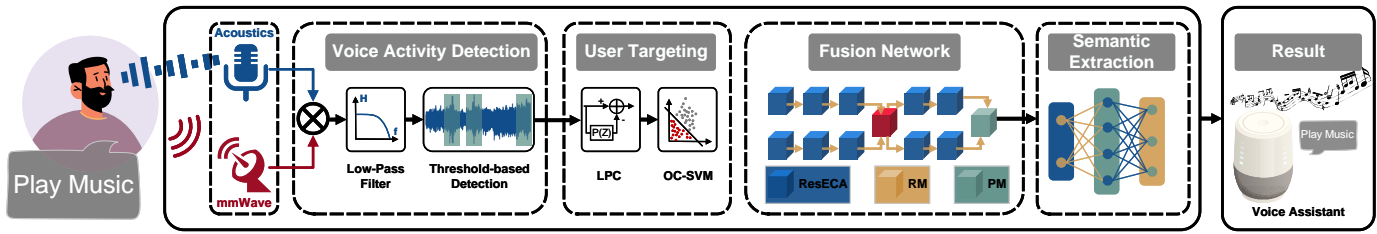
Fig. 14. The multi-model speech recognition system.

compensate for information loss in mmWave signals.

The architecture of Wavoice is shown in Fig. 14. Wavoice first detects the voice activity by comparing the audio signals and mmWave signals in each range bin. It multiplies these two signals and inputs the result into a low-pass filter. If these two signals have the same or similar frequency components, there will be an energy peak in the output of the low-pass filter. Then Wavoice proposes two learning-based modules for multi-modal signals fusion. One module exchanges valid features while the other module projects respective information into a joint feature space. Finally, the semantic information can be extracted from the joint feature space with a typical speech-to-text translation system. Results show that Wavoice outperforms existing methods with the character recognition error rate (CER) below 1% in a range of 7 m.

However, the voice detection problem still exists in non-line-of-sight (NLoS) scenarios, where the line-of-sight (LoS) path between the throat and the radar is hard to find or even does not exist. **AmbiEar** [66] pinpoints that although the near-throat region may not be sensed, surrounding objects around the speaker can provide voice-related information. Its insight is that sound propagates as a mechanical wave. The speaker's voice causes similar vibrations of the surrounding objects, which are highly correlated with the voice. To detect the surrounding objects around the speaker, AmbiEar proposes a variance-based human trajectory detection algorithm to detect the human's trajectory and locate surrounding objects. Then a modified minimum variance distortionless response (MVDR) [42] algorithm is applied to these signals from surrounding objects to extract their common components and enhance each signal. Finally, these signals are superimposed and fed into an end-to-end network for voice recognition. AmbiEar provides an effective voice recognition approach in NLoS scenarios with a word recognition accuracy of 87.21%.

D. Human Imaging

mmWave imaging system emits artificially generated signals to illuminate the target object, and 2D/3D images can be reconstructed by the reflected signals from the target according to the reflectance distribution. Different from the existing vision, IR, and X-ray imaging systems, mmWave signals can penetrate clothes and work in low visibility conditions. Due to its higher privacy and millimeter-scale ranging resolution, mmWave-based imaging has been widely pursued for pose/posture tracking [30] [112], automatic driving [114], security concerns [65] [113] [119], etc. We summarize the related works in TABLE VIII.
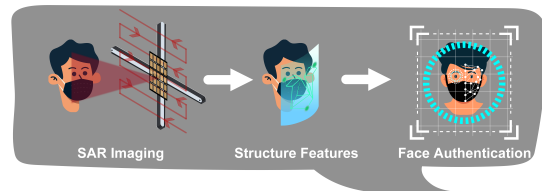

Fig. 15. The flowchart of mmFace.

The biggest challenge of mmWave imaging system is the data sparsity. The small size of the antenna array of mmWave radar device and its specularity lead to a poor spatial resolution. In general, the imaging resolution of a radar system is defined by $resolution \propto wavelength \times distance/aperture$. For an antenna array size of 1.8 cm $\times$ 1.8 cm, its imaging resolution can only reach up to 28 cm at 1 m distance. To enhance resolution, a universal technique is the Synthetic Aperture Radar (SAR) [120], in which the mmWave device needs to move a pre-determined trajectory and collect reflections from uniformly and densely grid locations. Recent works further propose to use the deep learning algorithm to generate images from mmWave signals.

mmFace [65] proposes a anti-spoofing face authentication (FA) system using a COTS mmWave radar (TI IWR1642-BOOST [5]). As Fig. 15 illustrates, mmFace emulates a large aperture planar antenna array by moving a radar along a specific 2D slide rail, and collects the reflected signals from the human face. Firstly, as the reflection coefficient only depends on the material of the facial surface [121], mmFace designs a novel algorithm to extract biometric features using the amplitudes of the reflected signals for liveness detection. With the reflected signals, a Range Migration algorithm (RMA) [122] is applied to construct a 3D facial image. However, it is distance-dependent and requires the authentication distance and registration distance to be identical, which is impractical. To further obtain distance-resistant features, mmFace utilizes the contour of the bright area on the facial image, which corresponds to the user's facial region and its surface curvature is relatively stable. This feature is utilized to calculate Fourier descriptors as the facial structure features for similarity-based user matching. Furthermore, to mitigate the registration overhead, mmFace also proposes a VRS (Virtual Registration Signals) generation method rather than directly collecting mmWave signals. Given three 2D facial photos taken from different perspectives, mmFace can reconstruct a virtual scene and generate the reflected mmWave signals by building a

TABLE VIII
COMPARISON OF HUMAN IMAGING WORKS

Method	Signal Form	Algorithm	Hardware Platform	Performance
mmFace [65]	Amplitude, Phase	Range Migration Algorithm, Similarity-based Matching, Ray-tracing	TI IWR1642BOOST	96% ASR lower than 5% EER
SquiggleMilli [112]	Phase, Waveform	Non-linear Motion Compensation, Compressed Sensing, Unconstrained Basis Pursuit De-noising, cGAN	TI IWR1443BOOST	0.85-0.95 Similarity 90% Classification Accuracy
milliCam [113]	Phase, Waveform	SAR, Motion Compensation Algorithm	Customized 60GHz Testbed	20% Error for 5×5 cm^2 4% Error for 20×20 cm^2
MILLIPOINT [114]	Point Cloud	Dynamic Programming, Automatic Multi-focusing Mechanism, Capon Algorithm	TI MMWCAS-DSP-EVM	1.2% Cumulative Tracking Error
mmMesh [30]	Point Cloud	Attention Mechanism, Skinned Multi-Person Linear Model	TI AWR1843BOOST	2.47 cm Error
mmFER [115]	Phase, Waveform	Gaussian Mixture Model, Cross-domain Transfer, Autoencoder	TI IWR1843BOOST	80.57% Accuracy
RPM [116]	Point Cloud	Attention Module, Deformable Multi-stage Convolution Module	Two FMCW Radars	5.71 cm Error
Hawkeye [117]	Phase, Waveform	cGAN, Ray-Tracing, SAR	Customized 60GHz Testbed	30 cm Ranging Error 27° Orientation Error
mmEye [24]	Point Cloud	MUSIC, Joint Transmitter Smoothing, Background and Noise Cancellation	Qualcomm 60GHz 802.11ad Chipset	7.6 cm Keypoint Precision
m3Track [118]	Doppler Spectrum	MVDR, Two-steam Deep Learning Architecture, K-means Algorithm	TI AWR1443BOOST	42.4 mm Tracking Error

theoretical model of mmWave signal propagation. With the generated VRSes, the user's corresponding facial structure features are stored as templates. The experimental results that mmFace can achieve an average authentication success rate (ASR) of 96% and an average equal error rate (EER) lower than 5%.

Unfortunately, the pre-determined trajectory and uniform sampling make it extremely challenging to emulate SAR on a hand-held device. To this end, **SquiggleMilli** [112] uses a non-linear motion compensation and a compressed sensing-based framework to generate uniformly and densely spaced grid locations, as the traditional SAR imaging system does. As a user freely moves the device in the air, SquiggleMilli maps the measured sample to the nearest point on the uniform grid according to its phase change. To avoid shape aliasing from the missing grid samples, SquiggleMilli estimates the missing by combining several measurements around its corresponding location. The recovery procedure then can be represented as an L1-norm minimization problem and solved by the unconstrained basis pursuit de-noising method [123]. Then it applies voxel segmentation to extract the 3D shape of objects at different depths. To enhance the perceptibility of the reconstructed 3D shape, SquiggleMilli leverages a pre-trained cGAN [111] model to recover the high spatial frequencies in the object to generate an accurate 2D shape with the partial 3D mmWave shape. It further designs two networks to predict its 3D features and category. The experimental results show that SquiggleMilli can reconstruct 2D shapes with a similarity score ranging from 0.85 to 0.95. For the 3D feature, it can predict the mean depth and rotation angle with less than 1% error and 1.5° error for the 90-th percentile. Meanwhile, it achieves more than 90% accuracy for classifying the objects.

Different from SquiggleMilli, **milliCam** [113] proposes a mmWave-based scanning system for capturing a high-resolution shape of a small metallic object by emulating the SAR imaging principle and swiping the hand-held mmWave device over the air. As SAR requires stringent linear motion and mm-scale localization accuracy, milliCam first leverages the co-located camera to compute the position and tractory of the device by measuring the translation and rotation. The camera has sub-mm pixel resolution and can achieve a 7mm localization error. Considering the target-scene is very close-by, such localization errors still cause significant errors. milliCam observes that most of the target-scene in the mmWave signals are sparse, due to the specular reflection. It re-designs a novel motion compensation algorithm based on the airborne SAR compensation by fusing a clustering algorithm and a bicubic interpolation. Furthermore, to resist the varied reflectivity and noise reflections, milliCam iteratively segments and corrects the defocused image using Otsu's method and squint correction, separately. After that, in order to ensure the phase coherency across the emulated apertures, milliCam calibrates the incoherent phases relying on the line-of-sight signals from Tx to Rx. Results show that the measurement errors are less than 20% and 4% for 5×5 cm^2 and 20×20 cm^2, respectively.

However, autonomous driving cars move fast and their trajectories are uncertain, making it very challenging for precise localization at the scale of the signal wavelength and uniform sampling, which are the requirements of emulating the SAR technique. To this end, **MILLIPOINT** [114] enables SAR imaging on low-cost commodity radar to generate 3D point clouds. First, MILLIPOINT observes that different Tx/Rx pairs may experience similar channel responses with a lagging effect, and the delay depends on the spacing between
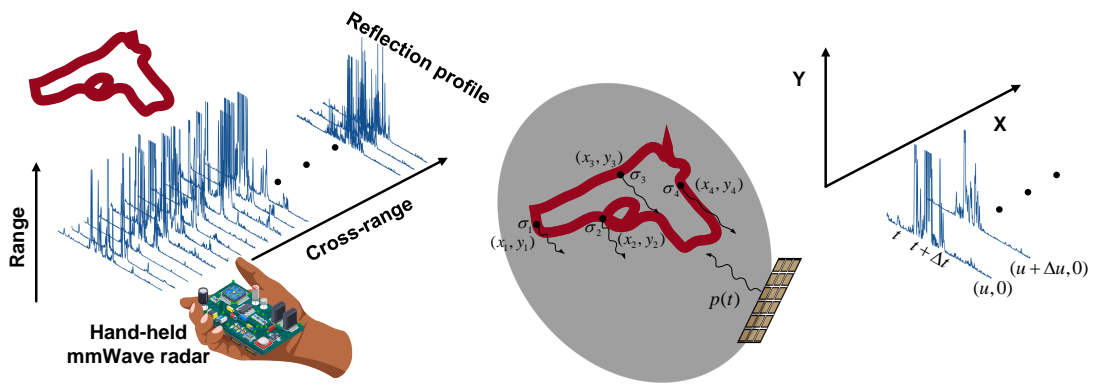
Fig. 16. The flowchart of milliCam.

antenna pairs. It characterizes the radar's cross-movement by correlating the received signals. The correlation reaches the maximum at the range of half a spacing of two antenna pairs. In addition, MILLIPOINT uses multiple antenna pairs with different spacing to overcome the influence of side peaks and achieves robust self-tracking by applying dynamic programming.

Furthermore, due to the existence of specular reflection, only reflections around the normal direction of the target can be received by the radar, resulting in a very small effective aperture. To synthesize a full image of the 3D environment, MILLIPOINT proposes an automatic multi-focusing mechanism to focus on each target separately and then synthesize them. To achieve it, MILLIPOINT first compensates for the antenna gain and the attenuation in the cross-range spectrums. It further identifies targets as the peaks of frequency spectrums along the range direction by an empirical threshold. The corresponding peak widths can be regarded as the target's effective aperture lengths. Then the classical SAR imaging algorithm is applied and synthesizes a 2D image by selecting pixels with the maximum value across all images. To extend it with height information, MILLIPOINT models the relationship between height and phase response of pixel, according to the idea of AOA estimation with antenna array [124]. The height is estimated with the Capon algorithm [39]. Through evaluation of the commercial mmWave Radar with 6 Tx and 8 Rx antennas, results show that MILLIPOINT achieves the median cumulative tracking error of 1.2% and has a precision rating of sub-mm.

As the SAR-based imaging methods require the target to be stationary or relatively stationary, they are not suitable for tracking the target's posture that dynamically changes. However, the spatial relationships and temporal dependencies of the joint skeletons are concealed in the sequence of mmWave reflected signals. This information can be utilized for pose tracking. **MilliPose** [64] proposes a machine-learning model for pose tracking and solves the problems of poor resolution, specular reflection, and varied reflectivity. Firstly, MilliPose leverages cGAN to generate a high-resolution 2D full-body silhouette image from the low-resolution 3D mmWave signals. Meanwhile, Considering that the degrees of freedom (DoF) and the ranges of body joints are very limited, MilliPose de-

signs an RNN, consisting of a two-layered GRU network [125] and a Structured Prediction Layer (SPL) [126], to learn these rules and predict skeleton joint of the next pose. Then it feedbacks the predicted pose to cGAN and can generate a high-quality body shape. Results show that the median error of joint location prediction is only 2.1 cm.

Instead of applying SAR imaging, **mmFER** [115] develops a mmWave radar-based system to recognize facial expressions by extracting and exploring subtle facial muscle movements from raw mmWave signals. To this end, mmFER first leverages the MIMO technique to localize the facial areas of the target and eliminate ambient noise. Then it utilizes Gaussian Mixture Model (GMM) to mitigate the impact of body motions on facial localization. mmFER further proposes a cross-domain transfer pipeline to guarantee an effective and reliable model knowledge transformation from image to mmWave. In particular, a hybrid learning loss function is designed to effectively address the over-fitting issue caused by small-scale training datasets. Moreover, an autoencoder is explored to learn the transition of latent features to reshape mmWave data, thereby eliminating the influence of data heterogeneity. mmFER is fully implemented using a commercial mmWave radar (TI IWR1843BOOST), and results show that mmFER obtains an accuracy of at least 80.57% when the distance between the subject and radar is within the range of 0.3-2.5 m.

Different from generating 2D silhouette images, **mmMesh** [30] proposes a deep learning framework to reconstruct the dynamic 3D human mesh from the mmWave signals. To alleviate the influence of ambient noise and multipath effects, mmMesh uses an attention mechanism to discriminate the point cloud and the points reflected by the subject tend to have higher qualities. Then a rough estimation of the shape and pose can be generated. To correctly align the points and make the human mesh more accurate, mmMesh dynamically chooses some "virtual locations" near the subject as anchor points that are each related to a part of the human body. According to the anchor locations, mmMesh dynamically groups the 3D point cloud into several subsets, and each one corresponds to a different body segment. Meanwhile, the local structure and associations can be learned. Finally, mmMesh incorporates the Skinned Multi-Person Linear (SMPL) [127] model as an additional constraint. It allows us to represent the whole 3D
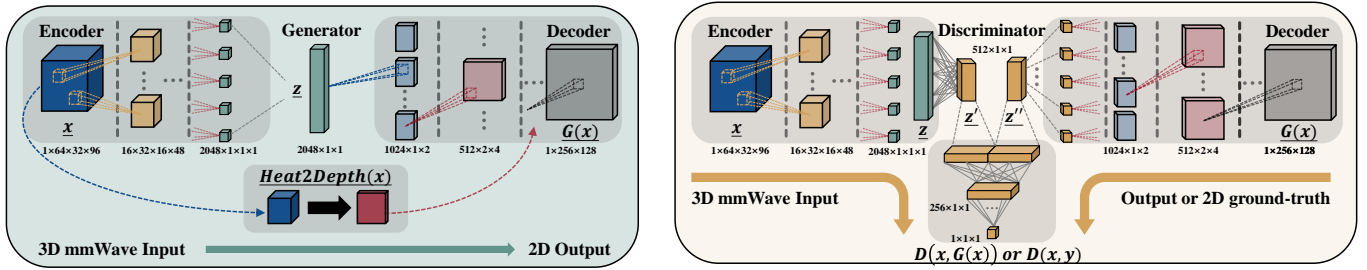
Fig. 17. The architecture of HawkEye.

human mesh with 86 parameters and does not need to calculate the positions of thousands of vertices. The experimental results show that mmMesh can achieve an average vertex error of 2.47 cm and a mesh localization error of 1.27 cm. Besides, the average joint localization and rotation errors are 2.18 cm and 3.8 degrees, respectively.

Besides, **RPM** [116] designs a multi-dimensional feature fusion backbone network for human pose estimation. It consists of two blocks: one is a channel fusion block that can effectively fuse horizontal and vertical RF features based on the channels' correlation. The other one is a deformable multi-stage convolution module that maintains scale-insensitive feature representations. Due to the high directivity of mmWave radar, it is possible to receive the reflected signal of a subset of limbs at each RF snapshot. Therefore, a spatial attention module is applied to recover the remaining body parts from a single RF snapshot by modeling non-local skeleton relationships. A temporal attention module is further applied to model the temporal dependencies across the 3D skeleton sequences. Finally, the complete 3D skeleton can be obtained. The experimental results reveal that RPM can achieve a Mean Per Joint Position Error (MPJPE) of 5.71 cm.

Although deep learning-based methods are proven to be effective, the procedure of collecting real-world signals for training is very time-consuming and labor-intensive. Benefiting from the success of the GAN technique and its robust resistance to noise, **HawkEye** [117] uses the GAN technique to generate the mmWave signals for training. In this work, a cGAN-based architecture is designed to predict high-resolution depth maps with a low-resolution mmWave heatmap input. HawkEye utilizes an encode-decode framework as the generator network to generate a 2D depth image. Meanwhile, the skip connection mechanism [128] is applied to retain the high-frequency details in depth. The discriminator employs a two-stream architecture to classify real vs. generated samples. Furthermore, to avoid collecting real-world mmWave data, HawkEye proposes a data synthesizer to simulate the propagation of the mmWave signals in the real world using ray-tracing [129] and generate realistic mmWave heatmaps.

Different from the use of radar devices, **mmEye** [24] is the pioneering work of attempting to achieve mmWave imaging on a single commodity 60GHz WiFi device. To break through the resolution limited by the physical aperture, mmEye proposes to reconstruct the spatial spectrum by applying the Multiple Signal Classification (MUSIC) [91] with Akaike information criterion (AIC) [130], a well-known information-theoretic approach for model order selection. However, The rank of MUSIC is likely much less than the number of incoming signals, making its performance greatly deteriorate or even fail. To overcome the rank deficiency problem without loss in precision, mmEye designs a novel 2D spatial smoothing termed *Joint transmitter smoothing (JIT)*. This technique reuses all the Rxs for each Tx rather than dividing the receive array. mmEye further proposes a background and noise cancellation (BANC) algorithm to remove interfering signals and a minimum mean square error (MMSE) [131] estimator is applied to compute the background CIR. Additionally, mmEye searches the point of interest (PoI) based on the variance of the energy distribution of the spatial spectrum. Finally, mmEye transforms the PoIs into plain images by solving a weighted least absolute deviation problem. The experimental results show that it achieves a median silhouette (shape) difference of 27.2% and a median boundary keypoint precision of 7.6 cm.

Although commodity RF signals have been utilized to realize 3D human posture tracking, these studies cannot handle cases where multiple users are in the same space. Towards this end, m^3**Track** [118] proposes a 3D posture tracking system. Based on the varying Doppler responses along with the speed of movement, m^3Track performs a convolution operation on the Range-Doppler-Profile to detect all users with their corresponding ranges. In the meanwhile, it utilizes minimum variance distortionless response (MVDR) [132] to separate all users from different angles. Given the range and the angle, m^3Track extracts the mmWave profiles of each user based on a 3-cylinder model, relating to the head-neck, chest-arm and lap-leg, respectively. Each profile contains two Range-Angle-Profiles (spatial features) and one Range-Doppler-Profile (temporal features). Then, m^3Track designs a two-steam deep learning architecture. It takes the spatial features and temporal features as inputs to extract the global body shape features and local body motion features. By concatenating these features, a regression model is applied to predict the 3D coordinates of skeleton joints and reconstruct the 3D body posture. For tracking posture in multi-user scenarios, it is necessary to map the reconstructed posture into real-world 3D space accurately. m^3Track leverages the K-means algorithm [49] to separate the point clouds into multiple clusters. Each cluster center corresponds to a posture joint, and the optimal mapping relation can be obtained by searching the minimal mapping error. Considering that the above computed 3D coordinate

is in the polar coordinate system, m^3Track further designs a coordinate-correlated extended Kalman filter to calculate the user's position in the Cartesian coordinate system. With the position trajectories, m^3Track realizes continuous 3D posture tracking in real-world multi-user scenarios. With TI AWR1443 [6], the experimental results show that the overall joint tracking error is 42.4 mm in a 4-user scenario.

E. Lessons Learned

The current mmWave-based human sensing has covered a variety of sensing tasks and corresponding solutions. When people try to utilize mmWave sensing to achieve a certain sensing task, they should first choose the appropriate hardware platform and processing signal form. For example, the point cloud is more suitable for contour-related sensing tasks such as activity recognition or human imaging, while the phase waveform is more suitable for fine-grained quantitative sensing tasks such as handwriting tracking or biometric measurement.

As we mentioned before, different sensing tasks have different technical frameworks and thus different challenges. For example, tracking and localization works are still plagued by trajectory crossing and multipath effect, which leave significant room for improvement in indoor multi-person localization. In terms of motion recognition, continuous motion recognition and dynamic human orientation changes are still open problems. Considering the unique requirements of biometric measurements for localization and sensing of human body parts (chest, throat, etc.), precise body parts localization and enhancement of the weak biometric signals are worthy of research, due to the uncertainty of human body position and posture. In human imaging, since SAR technique and deep learning models are widely used, how to improve the human imaging speed and model generalization is worthy of further study.

People then are encouraged to explore the core technical challenges of the corresponding sensing task, which determine the detailed technical targets of the algorithms to be adopted. For example, in recognition tasks such as activity recognition, gesture recognition, and voice recognition, considering the data differences (duration, intensity, variation, etc.) brought by different individuals, the learning-based classification methods often show better recognition abilities. Moreover, in measurement tasks such as handwriting tracking and vital sensing, the complex noise has a significant impact on sensing results, due to the requirement for fine-grained quantitative measurements. Therefore, appropriate noise reduction methods need to be carefully designed.

Furthermore, the sensing scenes should be carefully observed to assist sensing methods and achieve better sensing results. For example, human orientation and radar deployment need to be carefully considered to reduce potential signal distortion and environmental noise. On the other hand, multiple reflections between human and environmental reflectors and the impact of the human body on the environment may contain additional human-related information, which can help to achieve more general or accurate sensing.

Finally, people are encouraged to deploy their systems in various scenarios to verify their performance in detail.

Considering the increasing demand for fine-grained sensing in complex scenarios, such as in smart cockpits and smart homes, multi-scenario experimental verification can not only explore the potential capabilities but also promote the deployment of mmWave-based human sensing in actual scenarios.

VI. APPLICATION SCENARIOS

Various human sensing tasks can be applied to many actual scenarios to realize attractive applications. In this section, we briefly introduce some application scenarios of mmWave-based human sensing, including smart home, smart health, smart vehicle and security.

A. Smart Home

Smart home is very attractive in our daily scenarios as people can directly interact with devices through gestures or voices to achieve various applications. Moreover, smart home can automatically monitor our daily activities to provide customized services and protect our safety [133]–[142]. As mmWave-based human sensing techniques can capture human-related activity information in real-time, they can be well adapted to smart homes.

As the mmWave radar can capture people's commands through activity recognition [29], [69], [77], gesture recognition [33], [34], [80] and sound recognition [66], [76], [106], people can only give preset actions or voices to directly interact with devices, such as opening curtains and turning off lights. For example, mHomeGes [33] enables real-time arm gesture recognition in smart home scenarios to provide natural user experiences. Furthermore, mmWave-based gesture recognition can also help Deaf and Hard of Hearing (DHH) people communicate with others. mmASL [80] provides a home assistant system for DHH users that can perform American Sign Language (ASL) recognition with a single mmWave radar. Considering complex home scenarios, signal denoising techniques need to be adopted and deep learning-based sensing models are highly recommended.

Smart home is also expected to automatically monitor people's daily activities to provide more intelligent services. The tracking and localization of people can help to identify the location of people to provide more accurate services [23], [68]. mmTrack [23] proposes a passive indoor localization system and enables various applications that demand ubiquitous human localization, such as intelligently rotating the TV's viewing angle with the user's location and turning on/off the TV when users are coming/leaving. Furthermore, the mmWave-based vital sensing [22], [61] can monitor the sleep state of people, so as to choose the appropriate light intensity and temperature. Considering the non-invasive sensing capability of mmWave radar, it is more comfortable and suitable for long-term use, compared with wearable sensor-based sensing.

Furthermore, mmWave-based human sensing can help smart home achieve intrusion detection. by locating the intruder's location [68], [69] and identifying its gait [92], [93] or voiceprint [103], [110], smart home can judge whether the intruder's identity is legal and take corresponding warning

measures. Compared with existing WiFi-based intrusion detection systems [143], [144], mmWave sensing is expected to achieve more accurate detection results, thanks to its high spatial resolution and fine-grained sensing capabilities.

B. Smart Health

Smart health is another attractive application scenario [10], [11], [145]–[148]. Through the continuous non-intrusive sensing of the user's physical state, the mmWave devices can provide the user with daily health monitoring and timely anomaly detection.

When the user suffers from certain chronic diseases and needs real-time monitoring, mmWave-based vital sensing can provide the users with continuous health monitoring and assist doctors by providing sufficient auxiliary data, such as heart rate [22], [61], breathing rate [95], [99], blood pressure changes [102], etc. Especially, thanks to the contactless fine-grained sensing capability of mmWave radar, users do not have to remain still to obtain the vital sensing results. Movi-Fi [99] enables a motion-robust vital signs monitoring system and can recover vital signs waveform under severe body movements. mmECG [96] utilizes the mmWave radar to monitor the heart movements of drivers in moving vehicles.

When some special groups (e.g., children, the elderly, patients in convalescence, etc.) cannot guarantee their own safety during activities, the mmWave devices can monitor their activities [55] and timely detect abnormal activities [149], such as falls, bumps, chokes, etc. In this case, body-harmful actions can be detected and dealt with in time. Furthermore, mmWave-based activity recognition can also help people adopt recommended habits and stay healthy. For example, RF-Wash [79] attempts to track hand hygiene by monitoring hand washing activities and is expected to be applied in future healthcare facilities.

C. Smart Vehicle

With the rapid development of autonomous driving techniques and the Internet of Vehicles, smart vehicles are moving from a beautiful vision toward reality [150]–[154]. Smart vehicles can monitor road conditions and provide drivers with comprehensive assisted driving services.

When driving in scenes with limited vision such as heavy fog and rainy days, it is difficult for the driver to judge pedestrians and cyclists beside the car, which leads to potential driving accidents. Thanks to the penetrability of mmWave in these scenarios, the mmWave devices can identify potential pedestrians and cyclists through human imaging [114], [117] or activity recognition [29], [78] to avoid driving accidents. Moreover, smart vehicles can predict the trajectories of pedestrians by tracking and localizing pedestrians [69], [118], thereby providing safer assisted driving information. Researchers have emphasized that being able to recognize, detect and track objects from afar is essential in smart vehicles to prevent traffic accidents [13], [14]. Moreover, the radar manufacturers, such as Texas Instruments, have provided detailed guidance [155] on the application of mmWave radar in automotive applications, which can help to further explore the application potential of mmWave radar in smart vehicles.

D. Security

Security has been widely concerned by people as they are closely related to our property and personal information [15], [156], [157]. On the one hand, mmWave-based identity authentication can provide people with high security and protect people from troubles brought by attackers. On the other hand, the sniffing of private information by mmWave devices may also bring new challenges to existing security services.

mmWave-based identity authentication is being gradually developed and adopted. Through sensing tasks such as gait recognition [32], [93], sound sensing [103], [110] and human imaging [65], mmWave devices can identify and authenticate the human's physical characteristics, including gait, voiceprint and facial features. As mmWave-based identity authentication relies on changes in mmWave signals, it is hard to attack and can work with other authentication methods to further enhance authentication security. mmFace [65] points out that the camera-based face authentication approaches are vulnerable to emerging 3D spoofing attacks. With the material sensitivity of RF signals, the mmWave radar can be utilized to distinguish real human faces from other materials and defend against 3D spoofing attacks. In this way, combining cameras and radars appears to be a more secure authentication method.

Conversely, mmWave-based human sensing can capture subtle changes in human physiological signs, such as throat vibrations [76], [106] and gesture variance [81], [82]. This means that mmWave devices can sniff sensitive private information, such as private voice and manually entered passwords, thereby threatening existing security services. For example, mmKey [82] enables a virtual keyboard to sense the position and action of human fingers. This approach, in turn, can be utilized by attackers for illegal eavesdropping. Moreover, AmbiEar [66] validates the feasibility of recovering human voice from surrounding objects. This approach may inspire rethinking the security issues of acoustic systems.

VII. CHALLENGES & FUTURE DIRECTIONS

According to the discussion in the previous section, we can see that there is still a big gap between the state of the arts and the ideal vision of mmWave-based human sensing, with respect to the capacity, applicability, accuracy, and ubiquity. Filling this gap requires continuous innovations in all the dimensions around this technology, such as hardware, algorithm, sensing mediums, and application designs. Therefore, our discussion in this section will accordingly focus on four main research directions, namely hardware and platforms, enhancing the applicability, novel sensing schemes, and integration with new mediums.

A. Hardware and Platforms

1) Current limitations: The current mmWave-based sensing technology still deeply suffers from limited hardware capabilities. The insufficient number of antennas, limited transmission power and hardware denoise capability hinder the further development of mmWave-based sensing.

Taking mmWave radar based sensing as an example, the COTS mmWave radar only has several antennas (e.g., 3 TX

antennas and 4 RX antennas in IWR1443 and IWR6843ISK-ODS), which leads to very limited angle resolution (15°). The reason is that these radars can only provide $3 * 4 = 12$ virtual antennas by letting 3 Tx antennas take turns transmitting signals and 4 Rx antennas receive signals simultaneously. Then the Angle-FFT operation can obtain an angle resolution of $\frac{180°}{3*4} = 15°$ [46]. Such angle resolution is insufficient to enable many fine-grained sensing tasks, such as face imaging. Researchers have attempted to perform Rx beamforming [45], [46] and synthetic aperture radar (SAR) technique [65], [114] to expand the radar's angle resolution. However, they either have limited improvements or need a customized slide rail, causing the current solutions still be unsatisfactory. To further enhance the angle resolution, it's necessary to increase the number of antennas and exploit novel signal enhancement techniques. Texas Instruments has launched a powerful radar, TI MMWCAS-RF-EVM [158], which has 12 Tx antennas and 16 Rx antennas and has a much better angle resolution. However, its four-radar cascaded array implementation makes its form factor multiply. Moreover, its power consumption (higher than 20 W [158]) is much higher than that of those radars mentioned before (less than 2 W [3], [4]). These factors limit the practical deployment of this radar, especially in mobile scenarios where energy consumption needs to be carefully considered. In fact, the impact of its high energy consumption has been discussed in the autonomous driving community. Moreover, some signal local resolution techniques, such as Chirp-Z transform techniques [159], can be explored.

The sensing range is another important technical indicator. Although the mmWave radar can sense vehicle targets from tens of meters, the range of human sensing is very limited. For example, mmTrack [23] can only track users within 5 m range. WaveEar [76] localizes the position of the speaker and performs speech recognition within 2 m range. The reason is that the extremely high frequency of the mmWave causes its rapid attenuation, which makes the human-related signal features blurred a few meters away. To overcome the limitation of sensing range, more powerful equipment is needed. They can be radars with higher transmission power or some auxiliary equipment that can enhance the reflected signals, such as retro-reflective tags [151], [154], [160].

To obtain more accurate sensing results, careful signal noise reduction is necessary to resist hardware deficiency and environmental interference. However, the existing hardware denoise capability cannot fully meet our needs. For example, mmSpy [161] models the noise from the ramp/settle times in the radar's oscillator. It further employs statistical error correction techniques to subtract it from the phase waveform. These imperfect hardware features may affect our sensing results and must be carefully modeled and handled.

2) Potential directions: Considering that the current limited hardware capabilities seriously constrain the development of sensing techniques, we hope to discuss the development trend of the future mmWave hardware to provide some ideas and reasonable imagination for researchers in industry and academia.

We think the future mmWave hardware will have the characteristics of miniaturization, an increased number of

antennas and flexible customization, etc. With the continuous development of the radio-frequency integrated circuit (RFIC) technique, the RF front-end and computing units of the radar have been able to be integrated into a small-size chip. In this way, the mmWave hardware can be easily integrated into mobile phones, smartwatches and other devices to achieve more ubiquitous sensing. In the past few years, Soli [2] has made this attempt but most radars still have not been integrated. Nevertheless, we believe that miniaturization will be a major trend of mmWave hardware.

Furthermore, since the number of antennas determines the angle resolution and directional sensing capability of the mmWave hardware, more antennas will definitely be welcomed in industry and academia. Texas Instruments has noticed this extensive demand and launched a more powerful radar, TI MMWCAS-RF-EVM [158], which forms a 12×16 virtual antenna array. Compared with mmWave radars, mmWave probes usually have more antennas, such as the 32 Tx antennas and 32 Rx antennas of Qualcomm 802.11ad 60GHz WiFi. However, these probes cannot easily form a larger virtual antenna array. In addition, the sensing pipeline of transmitting and receiving in turn also limits their sensing frequency.

Finally, we believe that more flexible customization is attractive and fascinating. The current mmWave hardware platform has achieved a lot of signal collection and processing functions. However, due to the high integration of hardware, it is difficult for developers to modify the hardware to make their attempts, such as antenna layout and polarization direction modification. We believe that more flexible customization can certainly stimulate the community creatively.

B. Enhancing the Applicability

1) Challenges in real deployments: In addition to limited hardware capabilities, complex deployment environments also bring various critical challenges to mmWave-based human sensing. NLoS scenarios and multipath effects are the most common challenge in these complex environments.

Due to the human's position and posture dynamics and the blockage between the radar and the target, the human-related signal sometimes cannot be obtained directly, especially in NLoS scenarios. In this case, we can obtain the human-related signal with environmental reflection [46] or realize indirect sensing through the influence of the human on surrounding objects [66]. Some works also try to utilize the penetrability of the mmWave signal to achieve human sensing in NLoS scenarios [104], [105], [162]. However, their application scenarios are limited due to insufficient penetrability.

Multipath effects also challenge the actual deployment of mmWave-based human sensing, especially in dense environments. Due to the reflection of mmWave signals on the surface of environmental objects, the received signal is the superposition of multiple signals with different time delays and attenuation. Researchers have attempted to eliminate multipath interference by either deploying multiple radars [163] or exploiting the consistency of the dynamic between the target and the ghost images [33]. Some works begin to

utilize rather than eliminate the multipath effect. For example, mmReliable [164] utilizes multiple beams to achieve more reliable communication links. However, the exploitation of multipath effects in human sensing remains to be explored.

In addition, large-scale ground truth data is indispensable to the actual deployment of mmWave-based human sensing. Although the existing works have provided a lot of datasets, as mentioned in Sec. III-D. Datasets that contain quantitative sensing results are lacking. For example, the actual measurement dataset of human respiration and heartbeat is significant to develop the relevant mmWave-based sensing techniques.

2) Promising directions: To resist the complex deployment environment and provide accurate sensing results, researchers have made many various efforts [33], [45], [46], [66], [154]. Based on those works, we discuss the promising directions that appear to be effective in enhancing the applicability of mmWave-based human sensing. Retracing the continuous enrichment of mmWave-based human sensing applications in the past few years, we think that the future mmWave sensing applications will be more quantitative, ubiquitous and fine-grained, etc.

Although mmWave-based human sensing has realized a variety of sensing applications, varying from tracking and localization to vital sign sensing, the quantitative sensing of the human body still needs to be developed [45], [46]. Taking gesture recognition as an example, the existing works can well distinguish the gesture types. However, the quantitative indicators such as the amplitude of each gesture are difficult to analyze. To further promote the development of mmWave sensing techniques, accurate quantization results are needed. Moreover, quantitative sensing is a key step to realizing the practical application of sensing techniques. Without quantitative indicators that meet the error standard, it is difficult for the public to pay for these sensing techniques.

Another development trend of mmWave sensing applications lies in ubiquitous sensing, including wider sensing ranges, more universal sensing scenarios and richer sensing capabilities. As we mentioned before, due to the complex deployment environment, most of the existing works are the realization of single sensing tasks within the limited sensing range of specific scenarios. How to achieve more ubiquitous sensing is still an open problem. In the actual scenes, many factors such as range, occlusion, multipath and so on will lead to the limited application of mmWave sensing. Some works have tried to expand the sensing application scenarios [33], [66], [154]. However, more efforts are needed for this vision.

Finally, with the deepening of mmWave sensing techniques, more and more fine-grained sensing applications have been developed, including vital sensing and sound recognition, etc. However, we believe that the potential of mmWave sensing techniques goes beyond this. With directional and fine-grained sensing capabilities, mmWave sensing techniques are expected to enable more sophisticated sensing applications, such as skin disease detection, blink recognition, etc.

C. Novel Sensing Schemes

To further enrich the mmWave-based sensing capabilities, researchers have begun to find novel sensing schemes, such as fusion sensing and multitask sensing. Furthermore, researchers have noticed the indirect sensing ability of mmWave sensing and tried to realize enhanced sensing through "sensing side channels".

1) Fusion sensing: After realizing the limitation of existing mmWave-based human sensing techniques, some researchers begin to explore the feasibility of fusion sensing. By combining mmWave signal with other sensing media, such as vision, inertial measurement unit (IMU) and acoustic, fusion sensing is expected to provide more accurate and robust performance.

On the one hand, the sensing efficiency and ability can be increased by fusion sensing. The sensing techniques based on other media can share a part of the sensing task for mmWave sensing. For example, [162] utilizes a mmWave + camera multimodal sensing system to track the box's position and the relative position of fragile products in the box. The vision-based sensing algorithm is employed to locate the box along the conveyor belt, making the mmWave sensing can focus on computing the product's location.

On the other hand, by mutual correction with the sensing results based on other media, fusion sensing can further improve the sensing accuracy. [165] employs a mmWave radar based simultaneous localization and mapping (SLAM) solution assisted by IMU. The IMU data is utilized to combine continuous radar scanning point clouds into "multi-scan" to achieve accurate and robust SLAM results. Wavoice [106] explores the inherent correlation between the audio sensing result from a mmWave radar and a microphone. By combining the noisy audio signals and the aimless mmWave signal, Wavoice can achieve noise-resistant speech recognition. Considering the complexity of human sensing, fusion sensing is an attractive and effective way.

2) Multitask sensing: The complexity of human sensing is also reflected in the interlacing of human sensing tasks. For example, both gesture recognition and vital sensing are important components of health monitoring in smart home. If we can sense these two tasks simultaneously with a single mmWave radar, the deployment and implementation of smart home can be simplified, and the constraints on people will be further released.

However, multitask sensing is not simply the superposition of multiple single-task sensing. For example, human gestures and vital signals are closely related and are difficult to be separated. The relation between multiple human sensing tasks is a challenge to achieve multitask sensing, which makes it difficult to describe and analyze the reflected signals. Conversely, it is also an opportunity as mutual correlation and improvement can be performed. In this case, the multitask learning techniques [166], which have made great progress in the past ten years, may be able to show unique advantages in multitask sensing.

3) Side channel for human sensing: In addition to directly sensing the human body, there are many potential "sensing side channels" that can be used to obtain human-related information. By analyzing the influence of the human body on other objects, we can deduce the characteristics of the human body itself.

The current sensing side channel is most commonly used in sound recognition especially in eavesdropping. mmEve

[167] and mmSpy [161] attempt to recover speech emitted from smartphone earpieces. They both utilizes the fact that the reflected mmWave signals from the smartphone's rear are highly correlated with the soundwaves emitted from the smartphone's earpiece. Similarly, mmPhone [104] utilizes the changes of piezoelectric films with sound pressure to decode the speech through soundproof obstacles. There are also some works to eavesdrop devices with "sensing side channels". For example, SpiralSpy [168] utilizes a malicious software to encode data into the fan control signals. Then the fan motion status can be sensed and decoded by a mmWave radar.

Considering that the influence of the human body on the surrounding environment does not only exist in voice but also in many other physical signs, researchers can further explore the sensing side channel in many sensing tasks, such as keyboard sensing and gait recognition.

D. Integration with New Mediums

Innovations in the wireless medium is the extra boost of mmWave-based sensing. Emerging wireless communication techniques are deemed to provide ideal sensing mediums. With the continuous development and innovation of wireless communication techniques, many of them have been applied to sensing tasks to achieve ubiquitous sensing, including integration sensing and communication, backscatter, intelligent reflecting surface and THz sensing.

1) Integration sensing and communication: With the increasing demand for human sensing and the number of mmWave network techniques such as 3GPP 5G-NR and IEEE 802.11ad/ay, Integration sensing and communication (ISAC) has received extensive attention from academia and industry.

Given the growing popularity of mmWave networks, human sensing based on communication devices seems to be more suitable for life scenes. Chenshu Wu *et al.* propose a series of works based on 60GHz probe, such as mmWrite [81], ViMo [22] and mmTrack [23]. However, these works all need to adjust the probe to a radar-like mode to give up the communication function. Therefore, such a design can only support the rotation of sensing and communication but can not achieve ISAC. On the contrary, SPARCS [55] proposes an IEEE 802.11ay based ISAC solution. It utilizes the CIR measurements from sparse communication data packages to analyze human movement without interfering with communication. Considering the popularity of 5G network and WiZig, ISAC has great potential to be applied to daily scenarios.

Moreover, some researchers have begun to endow the mmWave radar, which is customized for sensing, with additional communication capabilities. For example, people have changed the frequency, transmission time and interval of the mmWave signal to encode information. We will not expand as this is beyond the scope of this survey.

there are several critical technical challenges in implementing ISAC, both communication-assisted sensing and sensing-assisted communication. Firstly, people need to carefully consider the trade-off between information delivery and channel state estimation. The latter usually serves as the basis for sensing. More power allocated to information delivery means

less sensing capability, and vice versa. Since the ratio of communication and sensing is different in different scenarios, people should analyze the trade-off by actual measurement and modeling to achieve the desired target. Secondly, due to hardware limitations and environmental interference (e.g., LoS path blockage), the coverage of ISAC is often limited, which hinders the application of ISAC in many scenarios. To solve this problem, new auxiliary devices such as Intelligent Reflecting Surface (IRS) can be considered to be deployed around the main ISAC device to further expand the coverage of ISAC. For example, Mosiac [169] utilizes multiple curved reflectors to provide comprehensive coverage to achieve omnidirectional around-corner automotive radars. Finally, the interference from other devices will be stronger in ISAC compared to pure wireless sensing scenarios. The interfering signals can cause the ISAC signals to be corrupted, delayed or even lost. To improve the robustness of ISAC in scenes full of interference, people can either adopt compensatory operation similar to SPARCS [55] or protect ISAC signals from interference through scheduled measures, such as frequency division multiplexing and time division multiplexing.

2) Sensing over backscatter: Backscatter technology [170]–[174] has been developed rapidly in the past few years to provide communication capability for low-end devices. Compared with active radios, backscatter [175]–[180] promises to be extremely low power, smaller and cheaper alternative. As the reflected signal of the backscatter tag is affected by both the modulation information and the channel state, we can sense the target state by attaching the tag to the target and analyzing the change of the reflected signal.

Backscatter technology has been used to improve the sensing range of mmWave sensing. For example, Millimetro [154] exploits a customized backscatter tag to realize accurate localization in a long range. The tag is formed by Van Atta retro-reflectors, which can reflect the carrier signals back in the direction of arrival regardless of the incident angle and the tag's mobility. With the retroreflected tag, The attached target can be localized at high accuracy (centimeter-level) over extended distances (over 100 m). Such a design can enhance the mmWave reflected signal and expand the sensing range of mmWave sensing.

The mmWave sensing capability can also be enhanced by combining mmWave sensing and backscatter technology. RoS [151] utilizes multiple configurable Van Atta arrays to build radar readable road signs. The arrays are formed as customized geometrical layouts and can be distinguished by an automotive radar. With the continuous development of mmWave backscatter technology [160], [181], we believe that there will be more mmWave sensing works combined with backscatter technology adapting to various sensing tasks.

3) Sensing with intelligent reflecting surface: As a cost-effective technology, Intelligent Reflecting surface (IRS) [182]–[187] has been an attractive subject to improve the performance of communication and sensing systems by reconfiguring the wireless propagation environments smartly. Generally, IRS is a metasurface consisting of a large number of passive reflecting elements. Each reflector induces an

adjustable phase shift and amplitude variation to the incident signals. With the ability to reconfigure the direction of mmWave signals, IRS is expected to play an important role in enhancing the coverage and energy efficiency of mmWave systems.

Due to the strong directivity and rapid attenuation of mmWave signals, the coverage of mmWave sensing is very limited. As IRS can reconfigure the mmWave signal direction, it can guide the mmWave signal to the originally inaccessible positions to enhance the sensing coverage. For example, MilliMirror [188] designs a passive metasurface prototype which expands the coverage of mmWave radios to blind spots by redirecting and reshaping mmWave signals to any anomalous directions. mmWall [189] proposes another tunable smart surface made of metamaterial, which enables a fast mmWave beam relay through the wall and redirects the beam power to another direction when a human body blocks a line-of-sight path. In addition, mmWall also supports splitting the incoming signal into multiple beams and concurrently steering the multi-armed beams. In this way, "through-wall" sensing or even multi-person sensing can be performed.

With such powerful capability, however, IRS has the potential to be utilized for illegal attacks. MeSS [190] has achieved illegal eavesdropping based on mmWave metasurface. By leveraging two degrees of freedom (space and time) in reconfigurable surfaces, MeSS generates and steers a concealed directional sideband toward the eavesdropper while maintaining the direction of the mainband toward the legitimate client. Joonas Kokkoniemi *et al.* [191] also discuss the channel modeling of the phased array type reconfigurable IRSs in the mmWave band.

To sum up, we believe the research of IRS has the opportunity to open a novel direction for solving the pain points of mmWave sensing, such as NLoS, mobility, and so on. Meanwhile, it also brings new challenges to our privacy protection.

4) From mmWave to THz: With the rapid increase of smart applications and devices, it is foreseeable that 5G network is hard to meet the ever-increasing network traffic. According to ITU-T's estimation [192], the global mobile traffic will grow to 5016 EB per month in 2030, compared to 62 EB per month in 2020. To solve this problem, the 6G network has been discussed and researched by industry and academia. mmWave signal and THz signal are expected to be used in 6G because of their wide unexplored spectrum. At present, mmWave communication and THz communication have reached 1 terabyte/s and 206.25 Gbit/s data rate in laboratory environments [193], [194].

On the other hand, due to the lack of corresponding standards, we make some predictions about the requirement of 6G in human sensing from application scenarios. The 6G network is expected to be human-centric and provide comprehensive human sensing. Therefore, 6G will pay more attention to various human-related applications, such as holographic-type communications, digital twins and tactile internet. Taking the tactile internet as an example, human sensing techniques may be used to recognize human body movements and the 6G network is expected to meet the low transmission delay

of 1 millisecond or less reaction time [195]. Based on our summary of mmWave sensing, we give some discussions on THz sensing to complement blueprint for 6G.

THz signals share many similarities with mmWave signals, including high frequency, large bandwidth and extremely high attenuation. With the continuous development of wireless sensing, THz signals have also begun to enter the field of vision of researchers and are used to realize some sensing tasks, such as indoor positioning [196]. Since THz signal frequency (0.3-3 THz) is much higher than that of mmWave signal (0.03-0.3 THz), THz sensing faces many unique challenges and has attractive potential capabilities.

The high-frequency characteristics of THz signals will cause extremely higher attenuation, resulting in very limited sensing coverage and low-SNR received signal. The limited sensing coverage may make it difficult to apply THz sensing to outdoor human sensing. Furthermore, how to reliably extract the human-related phases from the low-SNR received signal needs to be further explored. Novel low-noise electronics may be demanded. In addition, considering the separated transceiver device, tight synchronization is needed to extract human-related information, which will be more challenging in higher frequencies.

Conversely, the higher frequency and larger bandwidth of THz signals enable THz sensing to have finer sensing capability and higher spatial resolution, which means that more fine-grained sensing tasks can be achieved, such as fine-grained medical imaging, fingerprint detection, skin texture detection, etc. Considering the frequency limitation of mmWave signals, THz sensing is more suitable for achieving such fine-grained sensing tasks.

VIII. CONCLUSION

mmWave sensing has become more and more popular with the development of the 5G network and automatic driving. This survey focuses on recent research advances in mmWave-based human sensing. This survey presents various hardware and the key techniques of mmWave sensing, reviews existing mmWave sensing works based on different sensing tasks, i.e., human tracking and localization, motion recognition, biometric measurement and human imaging. We further discuss potential challenges and future directions, including hardware and platforms, enhancing the applicability, novel sensing schemes, and integration with new mediums.

ACKNOWLEDGMENT

LIST OF ABBREVIATIONS

ADC   Analog-to-digital converter
AiP   Antenna in package
AoA   Angle of arrival

ASL	American sign language
ASR	Authentication success rate
AUC	Area under the ROC curve
BANC	background and noise cancellation
CER	Character error rate
CFAR	Constant false alarm rate
CGAN	Conditional generative adversarial nets
CIR	Channel impulse response
CNN	Convolutional neural network
COTS	Commercial off-the-shelf
CPDA	Crossover path disambiguation algorithm
CPDP	Concentrated position-Doppler profile

DBSCAN Density-based spatial clustering of applications with noise
DCT Discrete cosine transformd
DDBR Dual-differential backgound removal
DHH Deaf of hard-of-hearing
DNN Deep neural network
DoF Degrees of freedom

EER Equal error rate
EKF Extended Kalman filter
EMD Empirical mode decomposition algorithm
EM Electromagnetic

FA Face authentication
FFT Fast fourier transform
FMCW Frequency modulated continuous wave
FSK Frequency shift keying

GAN Generative adversarial network
GRU Gated recurrent network

HAR Human activity recognition
HMM-VM Hidden markov model-based voting mechanism
HRV Heart rate variability
HWA Hardware accelerator block

IF Intermediate frequency
IHT Iterative hard thresholding
IMF Intrinsic mode function
IMU Inertial measurement unit
IRS Intelligent reflective surface
ISAC Integrated sensing and communication

JIT Joint transmitter smoothing

LoS Line-of-sight
LSTM Long short term memory

MCD Mel-Cepstral distortion
MFCC Mel frequency cepstral coefficient
MMSE Minimum mean square error
MTI Moving target indication
MUSIC Multiple signal classification

MVDR Minimum variance distortionless response

NLoS Non-line-of-sight

PoI Point of interest

RAI Range angle image
RDM Range-Doppler matrix
RFIC Radio-frequency integrated circuit
RF Radio frequency
RNN Residual neural network
RPCC Residual phase cepstrum coefficients
RSS Receiver signal strength

SAR Synthetic aperture radar
SLAM Simultaneous localization and mapping
SNR Signal-to-noise ratio
STFT Short-time Fourier transform

VCV Vocal cord vibration
VMD Variational modal decomposition

WER Word error rate
WPT Wavelet packet transform

REFERENCES

[1] Google, "Soli: You are the only interface you need," https://atap.google.com/soli/, 2022.

[2] J. Lien, N. Gillian, M. E. Karagozler, P. Amihood, C. Schwesig, E. Olson, H. Raja, and I. Poupyrev, "Soli: Ubiquitous gesture sensing with millimeter wave radar," ACM Transactions on Graphics (TOG), vol. 35, no. 4, pp. 1–19, 2016.

[3] T. I. Incorporated, "Iwr1443boost: Iwr1443 single-chip 76-ghz to 81-ghz mmwave sensor evaluation module," https://www.ti.com/tool/IWR1443BOOST, 2022.

[4] T. I. Incorporeted, "Iwr6843isk-ods: Iwr6843 intelligent mmwave overhead detection sensor (ods) antenna plug-in module," https://www.ti.com/tool/IWR6843ISK-ODS, 2022.

[5] T. I. Incorporated, "Iwr1642: Single-chip 76-ghz to 81-ghz mmwave sensor integrating dsp and mcu," https://www.ti.com/product/IWR1642, 2022.

[6] T. I. Incorporeted, "Single-chip 76-ghz to 81-ghz automotive radar sensor integrating mcu and hardware accelerator," https://www.ti.com/product/AWR1443, 2022.

[7] T. I. Incorporated, "Awr1642: Single-chip 76-ghz to 81-ghz automotive radar sensor integrating dsp and mcu," https://www.ti.com/product/AWR1642, 2022.

[8] S. A. Busari, K. M. S. Huq, S. Mumtaz, L. Dai, and J. Rodriguez, "Millimeter-wave massive mimo communication for future wireless systems: A survey," IEEE Communications Surveys & Tutorials, vol. 20, no. 2, pp. 836–869, 2017.

[9] P. Zhou, K. Cheng, X. Han, X. Fang, Y. Fang, R. He, Y. Long, and Y. Liu, "Ieee 802.11 ay-based mmwave wlans: Design challenges and solutions," IEEE Communications Surveys & Tutorials, vol. 20, no. 3, pp. 1654–1681, 2018.

[10] A. Singh, S. U. Rehman, S. Yongchareon, and P. H. J. Chong, "Multi-resident non-contact vital sign monitoring using radar: A review," IEEE Sensors Journal, vol. 21, no. 4, pp. 4061–4084, 2020.

[11] M. Kebe, R. Gadhafi, B. Mohammad, M. Sanduleanu, H. Saleh, and M. Al-Qutayri, "Human vital signs detection methods and potential using radars: A review," Sensors, vol. 20, no. 5, p. 1454, 2020.

[12] A. Shastri, N. Valecha, E. Bashirov, H. Tataria, M. Lentmaier, F. Tufvesson, M. Rossi, and P. Casari, "A review of millimeter wave device-based localization and device-free sensing technologies and applications," IEEE Communications Surveys & Tutorials, 2022.

[13] F. J. Abdu, Y. Zhang, M. Fu, Y. Li, and Z. Deng, "Application of deep learning on millimeter-wave radar signals: A review," Sensors, vol. 21, no. 6, p. 1951, 2021.

[14] A. Venon, Y. Dupuis, P. Vasseur, and P. Merriaux, "Millimeter wave fmcw radars for perception, recognition and localization in automotive applications: A survey," IEEE Transactions on Intelligent Vehicles, vol. 7, no. 3, pp. 533–555, 2022.

[15] Z. Wang, J. Li, Y. Jin, J. Wang, F. Yang, G. Li, X. Ni, and W. Ding, "Sensing beyond itself: Multi-functional use of ubiquitous signals towards wearable applications," Digital Signal Processing, vol. 116, p. 103091, 2021.

[16] J. Zhang, X. Zhang, P. Kulkarni, and P. Ramanathan, "Openmili: A 60 ghz software radio platform with a reconfigurable phased-array antenna," in Proceedings of the 22nd Annual International Conference on Mobile Computing and Networking, 2016, pp. 162–175.

[17] R. Zhao, T. Woodford, T. Wei, K. Qian, and X. Zhang, "M-cube: A millimeter-wave massive mimo software radio," in Proceedings of the 26th Annual International Conference on Mobile Computing and Networking, 2020, pp. 1–14.

[18] J. O. Lacruz, D. Garcia, P. J. Mateo, J. Palacios, and J. Widmer, "mm-flex: an open platform for millimeter-wave mobile full-bandwidth experimentation," in Proceedings of the 18th International Conference on Mobile Systems, Applications, and Services, 2020, pp. 1–13.

[19] Z. Li, Z. Yang, C. Song, C. Li, Z. Peng, and W. Xu, "E-eye: Hidden electronics recognition through mmwave nonlinear effects," in Proceedings of the 16th ACM Conference on Embedded Networked Sensor Systems, 2018, pp. 68–81.

[20] Qualcomm, "802.11ad 60ghz wifi," https://www.qualcomm.com/products/features/80211ad, 2022.

[21] TP-LINK, "Talon ad7200 multi-band wi-fi router," https://www.tp-link.com/us/home-networking/wifi-router/ad7200/, 2022.

[22] F. Wang, F. Zhang, C. Wu, B. Wang, and K. R. Liu, "Vimo: Multiperson vital sign monitoring using commodity millimeter-wave radio," IEEE Internet of Things Journal, vol. 8, no. 3, pp. 1294–1307, 2020.

[23] C. Wu, F. Zhang, B. Wang, and K. R. Liu, "mmtrack: Passive multi-person localization using commodity millimeter wave radio," in IEEE INFOCOM 2020-IEEE Conference on Computer Communications. IEEE, 2020, pp. 2400–2409.

[24] F. Zhang, C. Wu, B. Wang, and K. R. Liu, "mmeye: Super-resolution millimeter wave imaging," IEEE Internet of Things Journal, vol. 8, no. 8, pp. 6995–7008, 2020.

[25] S. An, Y. Li, and U. Ogras, "mRI: Multi-modal 3d human pose estimation dataset using mmwave, RGB-d, and inertial sensors," in Thirty-sixth Conference on Neural Information Processing Systems Datasets and Benchmarks Track, 2022. [Online]. Available: https://openreview.net/forum?id=Oa2-cdfBxun

[26] S. An and U. Y. Ogras, "Mars: mmwave-based assistive rehabilitation system for smart healthcare," ACM Transactions on Embedded Computing Systems (TECS), vol. 20, no. 5s, pp. 1–22, 2021.

[27] A. Sengupta, F. Jin, R. Zhang, and S. Cao, "mm-pose: Real-time human skeletal posture estimation using mmwave radars and cnns," IEEE Sensors Journal, vol. 20, no. 17, pp. 10032–10044, 2020.

[28] A. Chen, X. Wang, S. Zhu, Y. Li, J. Chen, and Q. Ye, "mmbody benchmark: 3d body reconstruction dataset and analysis for millimeter wave radar," in Proceedings of the 30th ACM International Conference on Multimedia, 2022, pp. 3501–3510.

[29] A. D. Singh, S. S. Sandha, L. Garcia, and M. Srivastava, "Radhar: Human activity recognition from point clouds generated through a millimeter-wave radar," in Proceedings of the 3rd ACM Workshop on Millimeter-wave Networks and Sensing Systems, 2019, pp. 51–56.

[30] H. Xue, Y. Ju, C. Miao, Y. Wang, S. Wang, A. Zhang, and L. Su, "mmmesh: Towards 3d real-time dynamic human mesh construction using millimeter-wave," in Proceedings of the 19th Annual International Conference on Mobile Systems, Applications, and Services, 2021, pp. 269–282.

[31] H. Liu, A. Zhou, Z. Dong, Y. Sun, J. Zhang, L. Liu, H. Ma, J. Liu, and N. Yang, "M-gesture: Person-independent real-time in-air gesture recognition using commodity millimeter wave radar," IEEE Internet of Things Journal, vol. 9, no. 5, pp. 3397–3415, 2021.

[32] Z. Meng, S. Fu, J. Yan, H. Liang, A. Zhou, S. Zhu, H. Ma, J. Liu, and N. Yang, "Gait recognition for co-existing multiple people using millimeter wave sensing," in Proceedings of the AAAI Conference on Artificial Intelligence, vol. 34, no. 01, 2020, pp. 849–856.

[33] H. Liu, Y. Wang, A. Zhou, H. He, W. Wang, K. Wang, P. Pan, Y. Lu, L. Liu, and H. Ma, "Real-time arm gesture recognition in smart home scenarios via millimeter wave sensing," Proceedings of the ACM on interactive, mobile, wearable and ubiquitous technologies, vol. 4, no. 4, pp. 1–28, 2020.

[34] S. Palipana, D. Salami, L. A. Leiva, and S. Sigg, "Pantomime: Mid-air gesture recognition with sparse millimeter-wave radar point clouds," Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies, vol. 5, no. 1, pp. 1–27, 2021.

[35] T. Wei and X. Zhang, "mtrack: High-precision passive tracking using millimeter wave radios," in Proceedings of the 21st Annual International Conference on Mobile Computing and Networking, 2015, pp. 117–129.

[36] C. J. Hansen, "Wigig: Multi-gigabit wireless communications in the 60 ghz band," IEEE Wireless Communications, vol. 18, no. 6, pp. 6–7, 2011.

[37] VICON], "Vicon motion systems." https://www.vicon.com, 2022.

[38] C. Iovescu and S. Rao, "The fundamentals of millimeter wave sensors," Texas Instruments, pp. 1–8, 2017.

[39] J. Capon, "High-resolution frequency-wavenumber spectrum analysis," Proceedings of the IEEE, vol. 57, no. 8, pp. 1408–1418, 1969.

[40] M. A. Richards, Fundamentals of radar signal processing. McGraw-Hill Education, 2014.

[41] I. S. Reed, R. M. Gagliardi, and L. B. Stotts, "A recursive moving-target-indication algorithm for optical image sequences," IEEE transactions on aerospace and electronic systems, vol. 26, no. 3, pp. 434–440, 1990.

[42] J. Benesty, J. Chen, and Y. Huang, "A generalized mvdr spectrum," IEEE Signal Processing Letters, vol. 12, no. 12, pp. 827–830, 2005.

[43] H. Rohling, "Radar cfar thresholding in clutter and multiple target situations," IEEE transactions on aerospace and electronic systems, no. 4, pp. 608–621, 1983.

[44] S. V. Vaseghi, Advanced digital signal processing and noise reduction. John Wiley & Sons, 2008.

[45] C. Jiang, J. Guo, Y. He, M. Jin, S. Li, and Y. Liu, "mmvib: micrometer-level vibration measurement with mmwave radar," in Proceedings of the 26th Annual International Conference on Mobile Computing and Networking, 2020, pp. 1–13.

[46] J. Guo, M. Jin, Y. He, W. Wang, and Y. Liu, "Dancing waltz with ghosts: Measuring sub-mm-level 2d rotor orbit with a single mmwave radar," in Proceedings of the 20th International Conference on Information Processing in Sensor Networks (co-located with CPS-IoT Week 2021), 2021, pp. 77–92.

[47] M. Z. Ozturk, C. Wu, B. Wang, and K. Liu, "Radiomic: Sound sensing via mmwave signals," arXiv preprint arXiv:2108.03164, 2021.

[48] M. Ester, H.-P. Kriegel, J. Sander, and X. Xu, "Density-based spatial clustering of applications with noise," in Int. Conf. Knowledge Discovery and Data Mining, vol. 240, no. 6, 1996.

[49] A. Likas, N. Vlassis, and J. J. Verbeek, "The global k-means clustering algorithm," Pattern recognition, vol. 36, no. 2, pp. 451–461, 2003.

[50] F. Wang, X. Zeng, C. Wu, B. Wang, and K. R. Liu, "mmhrv: Contactless heart rate variability monitoring using millimeter-wave radio," IEEE Internet of Things Journal, vol. 8, no. 22, pp. 16623–16636, 2021.

[51] A. Graps, "An introduction to wavelets," IEEE computational science and engineering, vol. 2, no. 2, pp. 50–61, 1995.

[52] J. Allen, "Short term spectral analysis, synthesis, and modification by discrete fourier transform," IEEE Transactions on Acoustics, Speech, and Signal Processing, vol. 25, no. 3, pp. 235–238, 1977.

[53] J. Wang and M. Johnson, "Residual phase cepstrum coefficients with application to cross-lingual speaker verification," in Thirteenth Annual Conference of the International Speech Communication Association, 2012.

[54] K. S. R. Murty and B. Yegnanarayana, "Combining evidence from residual phase and mfcc features for speaker recognition," IEEE signal processing letters, vol. 13, no. 1, pp. 52–55, 2005.

[55] J. Pegoraro, J. O. Lacruz, M. Rossi, and J. Widmer, "Sparcs: A sparse recovery approach for integrated communication and human sensing in mmwave systems," in 2022 21st ACM/IEEE International Conference on Information Processing in Sensor Networks (IPSN). IEEE, 2022, pp. 79–91.

[56] T. K. Ho, "Random decision forests," in Proceedings of 3rd international conference on document analysis and recognition, vol. 1. IEEE, 1995, pp. 278–282.

[57] J. R. Quinlan, "Simplifying decision trees," International journal of man-machine studies, vol. 27, no. 3, pp. 221–234, 1987.

[58] W. S. Noble, "What is a support vector machine?" Nature biotechnology, vol. 24, no. 12, pp. 1565–1567, 2006.

[59] J. Gu, Z. Wang, J. Kuen, L. Ma, A. Shahroudy, B. Shuai, T. Liu, X. Wang, G. Wang, J. Cai et al., "Recent advances in convolutional neural networks," Pattern recognition, vol. 77, pp. 354–377, 2018.

[60] W. Zaremba, I. Sutskever, and O. Vinyals, "Recurrent neural network regularization," arXiv preprint arXiv:1409.2329, 2014.

[61] U. Ha, S. Assana, and F. Adib, "Contactless seismocardiography via deep learning radars," in Proceedings of the 26th Annual International Conference on Mobile Computing and Networking, 2020, pp. 1–14.

[62] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, "Generative adversarial networks," Communications of the ACM, vol. 63, no. 11, pp. 139–144, 2020.

[63] P. Hu, Y. Ma, P. S. Santhalingam, P. H. Pathak, and X. Cheng, "Milliear: Millimeter-wave acoustic eavesdropping with unconstrained vocabulary," in IEEE INFOCOM 2022-IEEE Conference on Computer Communications. IEEE, 2022, pp. 11–20.

[64] A. Adhikari and S. Sur, "Millipose: Facilitating full body silhouette imaging from millimeter-wave device," in Adjunct Proceedings of the 2021 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2021 ACM International Symposium on Wearable Computers, 2021, pp. 1–3.

[65] W. Xu, W. Song, J. Liu, Y. Liu, X. Cui, Y. Zheng, J. Han, X. Wang, and K. Ren, "Mask does not matter: anti-spoofing face authentication using mmwave without on-site registration," in Proceedings of the 28th Annual International Conference on Mobile Computing And Networking, 2022, pp. 310–323.

[66] J. Zhang, Y. Zhou, R. Xi, S. Li, J. Guo, and Y. He, "Ambiear: mmwave based voice recognition in nlos scenarios," Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies, vol. 6, no. 3, pp. 1–25, 2022.

[67] T. Gu, Z. Fang, Z. Yang, P. Hu, and P. Mohapatra, "Mmsense: Multi-person detection and identification via mmwave sensing," in Proceedings of the 3rd ACM Workshop on Millimeter-wave Networks and Sensing Systems, 2019, pp. 45–50.

[68] P. Zhao, C. X. Lu, J. Wang, C. Chen, W. Wang, N. Trigoni, and A. Markham, "mid: Tracking and identifying people with millimeter wave radar," in 2019 15th International Conference on Distributed Computing in Sensor Systems (DCOSS). IEEE, 2019, pp. 33–40.

[69] M. A. U. Alam, M. M. Rahman, and J. Q. Widberg, "Palmar: Towards adaptive multi-inhabitant activity recognition in point-cloud technology," in IEEE INFOCOM 2021-IEEE Conference on Computer Communications. IEEE, 2021, pp. 1–10.

[70] Pasternack, "60 ghz transmit/receive (tx/rx) development system," https://www.pasternack.com/60-ghz-test-development-system-pem003-kit-p.aspx, 2022.

[71] J. A. Hartigan and M. A. Wong, "Algorithm as 136: A k-means clustering algorithm," Journal of the royal statistical society. series c (applied statistics), vol. 28, no. 1, pp. 100–108, 1979.

[72] J. Edmonds, "Paths, trees, and flowers," Canadian Journal of mathematics, vol. 17, pp. 449–467, 1965.

[73] G. Welch, G. Bishop et al., "An introduction to the kalman filter," 1995.

[74] S. Hochreiter and J. Schmidhuber, "Long short-term memory," Neural computation, vol. 9, no. 8, pp. 1735–1780, 1997.

[75] D. De, W.-Z. Song, M. Xu, C.-L. Wang, D. Cook, and X. Huo, "Findinghumo: Real-time tracking of motion trajectories from anonymous binary sensing in smart environments," in 2012 IEEE 32nd international conference on distributed computing systems. IEEE, 2012, pp. 163–172.

[76] C. Xu, Z. Li, H. Zhang, A. S. Rathore, H. Li, C. Song, K. Wang, and W. Xu, "Waveear: Exploring a mmwave-based noise-resistant speech sensing for voice-user interface," in Proceedings of the 17th Annual International Conference on Mobile Systems, Applications, and Services, 2019, pp. 14–26.

[77] W. Jiang, C. Miao, F. Ma, S. Yao, Y. Wang, Y. Yuan, H. Xue, C. Song, X. Ma, D. Koutsonikolas et al., "Towards environment independent device free human activity recognition," in Proceedings of the 24th annual international conference on mobile computing and networking, 2018, pp. 289–304.

[78] Y. Wang, H. Liu, K. Cui, A. Zhou, W. Li, and H. Ma, "m-activity: Accurate and real-time human activity recognition via millimeter wave radar," in ICASSP 2021-2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). IEEE, 2021, pp. 8298–8302.

[79] A. Khamis, B. Kusy, C. T. Chou, M.-L. McLaws, and W. Hu, "Rfwash: a weakly supervised tracking of hand hygiene technique," in Proceedings of the 18th Conference on Embedded Networked Sensor Systems, 2020, pp. 572–584.

[80] P. S. Santhalingam, A. A. Hosain, D. Zhang, P. Pathak, H. Rangwala, and R. Kushalnagar, "mmasl: Environment-independent asl gesture recognition using 60 ghz millimeter-wave signals," Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies, vol. 4, no. 1, pp. 1–30, 2020.

[81] S. D. Regani, C. Wu, B. Wang, M. Wu, and K. R. Liu, "mmwrite: passive handwriting tracking using a single millimeter-wave radio," IEEE Internet of Things Journal, vol. 8, no. 17, pp. 13 291–13 305, 2021.

[82] Y. Hu, B. Wang, C. Wu, and K. R. Liu, "mmkey: Universal virtual keyboard using a single millimeter-wave radio," IEEE Internet of Things Journal, vol. 9, no. 1, pp. 510–524, 2021.

[83] Y. Ganin and V. Lempitsky, "Unsupervised domain adaptation by backpropagation," in International conference on machine learning. PMLR, 2015, pp. 1180–1189.

[84] Y. Bar-Shalom, F. Daum, and J. Huang, "The probabilistic data association filter," IEEE Control Systems Magazine, vol. 29, no. 6, pp. 82–100, 2009.

[85] Y. C. Eldar and G. Kutyniok, Compressed sensing: theory and applications. Cambridge university press, 2012.

[86] T. Zhang, R. Ramakrishnan, and M. Livny, "Birch: an efficient data clustering method for very large databases," ACM sigmod record, vol. 25, no. 2, pp. 103–114, 1996.

[87] M. Schuster and K. K. Paliwal, "Bidirectional recurrent neural networks," IEEE transactions on Signal Processing, vol. 45, no. 11, pp. 2673–2681, 1997.

[88] A. Graves, S. Fernández, F. Gomez, and J. Schmidhuber, "Connectionist temporal classification: labelling unsegmented sequence data with recurrent neural networks," in Proceedings of the 23rd international conference on Machine learning, 2006, pp. 369–376.

[89] Wikipedia, "Sibeam," https://en.wikipedia.org/wiki/SiBeam, 2022.

[90] C. R. Qi, L. Yi, H. Su, and L. J. Guibas, "Pointnet++: Deep hierarchical feature learning on point sets in a metric space," Advances in neural information processing systems, vol. 30, 2017.

[91] R. Schmidt, "Multiple emitter location and signal parameter estimation," IEEE transactions on antennas and propagation, vol. 34, no. 3, pp. 276–280, 1986.

[92] X. Yang, J. Liu, Y. Chen, X. Guo, and Y. Xie, "Mu-id: Multi-user identification through gaits using millimeter wave radios," in IEEE INFOCOM 2020-IEEE Conference on Computer Communications. IEEE, 2020, pp. 2589–2598.

[93] M. Z. Ozturk, C. Wu, B. Wang, and K. R. Liu, "Gaitcube: Deep data cube learning for human recognition with millimeter-wave radio," IEEE Internet of Things Journal, vol. 9, no. 1, pp. 546–557, 2021.

[94] Z. Yang, P. H. Pathak, Y. Zeng, X. Liran, and P. Mohapatra, "Monitoring vital signs using millimeter wave," in Proceedings of the 17th ACM international symposium on mobile ad hoc networking and computing, 2016, pp. 211–220.

[95] P. Zhao, C. X. Lu, B. Wang, C. Chen, L. Xie, M. Wang, N. Trigoni, and A. Markham, "Heart rate sensing with a robot mounted mmwave radar," in 2020 IEEE International Conference on Robotics and Automation (ICRA). IEEE, 2020, pp. 2812–2818.

[96] X. Xu, J. Yu, C. Ma, Y. Ren, H. Liu, Y. Zhu, Y.-C. Chen, and F. Tang, "mmecg: Monitoring human cardiac cycle in driving environments leveraging millimeter wave," in IEEE INFOCOM 2022-IEEE Conference on Computer Communications. IEEE, 2022, pp. 90–99.

[97] Y. Wang, T. Gu, T. H. Luan, M. Lyu, and Y. Li, "Heartprint: Exploring a heartbeat-based multi-user authentication with single mmwave radar," IEEE Internet of Things Journal, 2022.

[98] J. Gong, X. Zhang, K. Lin, J. Ren, Y. Zhang, and W. Qiu, "Rf vital sign sensing under free body movement," Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies, vol. 5, no. 3, pp. 1–22, 2021.

[99] Z. Chen, T. Zheng, C. Cai, and J. Luo, "Movi-fi: Motion-robust vital signs waveform recovery via deep interpreted rf sensing," in Proceedings of the 27th Annual International Conference on Mobile Computing and Networking, 2021, pp. 392–405.

[100] S. Zhang, T. Zheng, Z. Chen, and J. Luo, "Can we obtain fine-grained heartbeat waveform via contact-free rf-sensing?" in IEEE INFOCOM 2022-IEEE Conference on Computer Communications. IEEE, 2022, pp. 1759–1768.

[101] C. Xu, H. Li, Z. Li, H. Zhang, A. S. Rathore, X. Chen, K. Wang, M.-c. Huang, and W. Xu, "Cardiacwave: A mmwave-based scheme of non-contact and high-definition heart activity computing," Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies, vol. 5, no. 3, pp. 1–26, 2021.

[102] Z. Shi, T. Gu, Y. Zhang, and X. Zhang, "mmbp: Contact-free millimetre-wave radar based approach to blood pressure measurement," in Proceedings of the 20th ACM Conference on Embedded Networked Sensor Systems, 2022.

[103] H. Li, C. Xu, A. S. Rathore, Z. Li, H. Zhang, C. Song, K. Wang, L. Su, F. Lin, K. Ren et al., "Vocalprint: exploring a resilient and secure voice

authentication via mmwave biometric interrogation," in Proceedings of the 18th Conference on Embedded Networked Sensor Systems, 2020, pp. 312–325.

[104] C. Wang, F. Lin, T. Liu, Z. Liu, Y. Shen, Z. Ba, L. Lu, W. Xu, and K. Ren, "mmphone: Acoustic eavesdropping on loudspeakers via mmwave-characterized piezoelectric effect," in IEEE INFOCOM 2022-IEEE Conference on Computer Communications. IEEE, 2022, pp. 820–829.

[105] C. Wang, F. Lin, Z. Ba, F. Zhang, W. Xu, and K. Ren, "Wavesdropper: Through-wall word detection of human speech via commercial mmwave devices," Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies, vol. 6, no. 2, pp. 1–26, 2022.

[106] T. Liu, M. Gao, F. Lin, C. Wang, Z. Ba, J. Han, W. Xu, and K. Ren, "Wavoice: A noise-resistant multi-modal speech recognition system fusing mmwave and audio signals," in Proceedings of the 19th ACM Conference on Embedded Networked Sensor Systems, 2021, pp. 97–110.

[107] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in International Conference on Medical image computing and computer-assisted intervention. Springer, 2015, pp. 234–241.

[108] D. Griffin and J. Lim, "Signal estimation from modified short-time fourier transform," IEEE Transactions on acoustics, speech, and signal processing, vol. 32, no. 2, pp. 236–243, 1984.

[109] R. Kubichek, "Mel-cepstral distance measure for objective speech quality assessment," in Proceedings of IEEE pacific rim conference on communications computers and signal processing, vol. 1. IEEE, 1993, pp. 125–128.

[110] Y. Dong and Y.-D. Yao, "Secure mmwave-radar-based speaker verification for iot smart home," IEEE Internet of Things Journal, vol. 8, no. 5, pp. 3500–3511, 2020.

[111] M. Mirza and S. Osindero, "Conditional generative adversarial nets," arXiv preprint arXiv:1411.1784, 2014.

[112] H. Regmi, M. S. Saadat, S. Sur, and S. Nelakuditi, "Squigglemilli: Approximating sar imaging on mobile millimeter-wave devices," Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies, vol. 5, no. 3, pp. 1–26, 2021.

[113] M. S. Saadat, S. Sur, S. Nelakuditi, and P. Ramanathan, "Millicam: Hand-held millimeter-wave imaging," in 2020 29th International Conference on Computer Communications and Networks (ICCCN). IEEE, 2020, pp. 1–9.

[114] K. Qian, Z. He, and X. Zhang, "3d point cloud generation with millimeter-wave radar," Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies, vol. 4, no. 4, pp. 1–23, 2020.

[115] X. Zhang, Y. Zhang, T. Gu, and Z. Shi, "Millimetre-wave radar based facial expression recognition for multimedia iot applications," in Proceedings of the 29th Annual International Conference on Mobile Computing and Networking, 2023.

[116] C. Xie, D. Zhang, Z. Wu, C. Yu, Y. Hu, Q. Sun, and Y. Chen, "Accurate human pose estimation using rf signals."

[117] J. Guan, S. Madani, S. Jog, and H. Hassanieh, "High resolution millimeter wave imaging for self-driving cars," arXiv preprint arXiv:1912.09579, 2019.

[118] H. Kong, X. Xu, J. Yu, Q. Chen, C. Ma, Y. Chen, Y.-C. Chen, and L. Kong, "m3track: mmwave-based multi-user 3d posture tracking," in Proceedings of the 20th Annual International Conference on Mobile Systems, Applications and Services, 2022, pp. 491–503.

[119] S. Wei, Z. Zhou, M. Wang, J. Wei, S. Liu, J. Shi, X. Zhang, and F. Fan, "3dried: A high-resolution 3-d millimeter-wave radar dataset dedicated to imaging and evaluation," Remote Sensing, vol. 13, no. 17, p. 3366, 2021.

[120] M. Soumekh, Synthetic aperture radar signal processing. New York: Wiley, 1999, vol. 7.

[121] C. Wu, F. Zhang, B. Wang, and K. R. Liu, "msense: Towards mobile material sensing with a single millimeter-wave radio," Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies, vol. 4, no. 3, pp. 1–20, 2020.

[122] M. E. Yanik, D. Wang, and M. Torlak, "3-d mimo-sar imaging using multi-chip cascaded millimeter-wave sensors," in 2019 IEEE Global Conference on Signal and Information Processing (GlobalSIP). IEEE, 2019, pp. 1–5.

[123] J. Yang and Y. Zhang, "Alternating direction algorithms for \ell_1-problems in compressive sensing," SIAM journal on scientific computing, vol. 33, no. 1, pp. 250–278, 2011.

[124] J. Xiong and K. Jamieson, "Arraytrack: A fine-grained indoor location system," in 10th USENIX Symposium on Networked Systems Design and Implementation (NSDI 13), 2013, pp. 71–84.

[125] K. Cho, B. Van Merriënboer, D. Bahdanau, and Y. Bengio, "On the properties of neural machine translation: Encoder-decoder approaches," arXiv preprint arXiv:1409.1259, 2014.

[126] E. Aksan, M. Kaufmann, and O. Hilliges, "Structured prediction helps 3d human motion modelling," in Proceedings of the IEEE/CVF International Conference on Computer Vision, 2019, pp. 7144–7153.

[127] M. Loper, N. Mahmood, J. Romero, G. Pons-Moll, and M. J. Black, "Smpl: A skinned multi-person linear model," ACM transactions on graphics (TOG), vol. 34, no. 6, pp. 1–16, 2015.

[128] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in International Conference on Medical image computing and computer-assisted intervention. Springer, 2015, pp. 234–241.

[129] V. Degli-Esposti, F. Fuschini, E. M. Vitucci, M. Barbiroli, M. Zoli, L. Tian, X. Yin, D. A. Dupleich, R. Müller, C. Schneider et al., "Ray-tracing-based mm-wave beamforming assessment," IEEE Access, vol. 2, pp. 1314–1325, 2014.

[130] A. P. Liavas and P. A. Regalia, "On the behavior of information theoretic criteria for model order selection," IEEE Transactions on Signal Processing, vol. 49, no. 8, pp. 1689–1695, 2001.

[131] B. Zhou and Q. Chen, "A tutorial on minimum mean square error estimation," Southwest Jiaotong Univ., Sichuan, China, Tech. Rep, 2015.

[132] E. A. P. Habets, J. Benesty, I. Cohen, S. Gannot, and J. Dmochowski, "New insights into the mvdr beamformer in room acoustics," IEEE Transactions on Audio, Speech, and Language Processing, vol. 18, no. 1, pp. 158–170, 2009.

[133] W. Wang, J. Li, Y. He, X. Guo, and Y. Liu, "Motorbeat: Acoustic communication for home appliances via variable pulse width modulation," pp. 31:1–31:24, 2022.

[134] M. Wang, W. Sun, and L. Qiu, "MAVL: multiresolution analysis of voice localization," in Proceedings of the 18th USENIX Symposium on Networked Systems Design and Implementation, 2021.

[135] W. Wang, J. Li, Y. He, and Y. Liu, "Symphony: Localizing multiple acoustic sources with a single microphone array," in Proceedings of the 18th ACM Conference on Embedded Networked Sensor Systems, 2020.

[136] S. Shen, D. Chen, Y. Wei, Z. Yang, and R. R. Choudhury, "Voice localization using nearby wall reflections," in Proceedings of the 26th ACM Annual International Conference on Mobile Computing and Networking, 2020.

[137] W. Wang, J. Li, Y. He, and Y. Liu, "Localizing multiple acoustic sources with a single microphone array," IEEE Transactions on Mobile Computing, 2022.

[138] W. Wang, J. Li, M. Jin, and Y. He, "Chordmics: Acoustic signal purification with distributed microphones," in Proceedings of the 29th IEEE International Conference on Computer Communications and Networks, 2020, pp. 1–9.

[139] C. Aranzazu-Suescun and M. Cardei, "Distributed algorithms for event reporting in mobile-sink wsns for internet of things," Tsinghua Science and Technology, 2017.

[140] L. Wang, T. Gu, H. Chen, X. Tao, and J. Lu, "Real-time activity recognition in wireless body sensor networks: From simple gestures to complex activities," 2010 IEEE 16th International Conference on Embedded and Real-Time Computing Systems and Applications, pp. 43–52, 2010.

[141] T. Gu, L. Wang, H. Chen, X. Tao, and J. Lu, "Recognizing multiuser activities using wireless body sensor networks," IEEE Transactions on Mobile Computing, vol. 10, no. 11, pp. 1618–1631, 2011.

[142] K. Gai, Q. Xiao, M. Qiu, G. Zhang, J. Chen, Y. Wei, and Y. Zhang, "Digital twin-enabled ai enhancement in smart critical infrastructures for 5g," ACM Transactions on Sensor Networks (TOSN), vol. 18, no. 3, pp. 1–20, 2022.

[143] S. Li, Z. Liu, Y. Zhang, Q. Lv, X. Niu, L. Wang, and D. Zhang, "Wiborder: Precise wi-fi based boundary sensing via through-wall discrimination," Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies, vol. 4, no. 3, pp. 1–30, 2020.

[144] T. Wang, D. Yang, S. Zhang, Y. Wu, and S. Xu, "Wi-alarm: Low-cost passive intrusion detection using wifi," Sensors, vol. 19, no. 10, p. 2335, 2019.

[145] W. Z. Khan, Y. Xiang, M. Y. Aalsalem, and Q. Arshad, "Mobile phone sensing systems: A survey," IEEE Communications Surveys & Tutorials, 2013.

[146] A. Wang, J. E. Sunshine, and S. Gollakota, "Contactless infant monitoring using white noise," in Proceedings of the 25th ACM Annual International Conference on Mobile Computing and Networking, 2019.

[147] B. Yu, Y. Wang, K. Niu, Y. Zeng, T. Gu, L. Wang, C. Guan, and D. Zhang, "Wifi-sleep: Sleep stage monitoring using commodity wi-fi devices," IEEE Internet of Things Journal, vol. 8, no. 18, pp. 13 900–13 913, 2021.

[148] Z. Yang, P. H. Pathak, Y. Zeng, X. Liran, and P. Mohapatra, "Vital sign and sleep monitoring using millimeter wave," ACM Transactions on Sensor Networks (TOSN), 2017.

[149] Y. Wang, K. Wu, and L. M. Ni, "Wifall: Device-free fall detection by wireless networks," IEEE Transactions on Mobile Computing, vol. 16, no. 2, pp. 581–594, 2016.

[150] W. Wang, L. Mottola, Y. He, J. Li, Y. Sun, S. Li, H. Jing, and Y. Wang, "Micnest: Long-range instant acoustic localization of drones in precise landing," in Proceedings of the 20th ACM Conference on Embedded Networked Sensor Systems, 2022.

[151] J. Nolan, K. Qian, and X. Zhang, "Ros: Passive smart surface for roadside-to-vehicle communication," in Proceedings of ACM SIGCOMM, Virtual Event, USA, August 23-27, 2021.

[152] Y. Sun, W. Wang, L. Mottola, R. Wang, and Y. He, "Aim: Acoustic inertial measurement for indoor drone localization and tracking," in Proceedings of the 20th ACM Conference on Embedded Networked Sensor Systems, 2022.

[153] Y. Wang, J. Cao, W. Li, T. Gu, and W. Shi, "Exploring traffic congestion correlation from multiple data sources," Pervasive and Mobile Computing, vol. 41, pp. 470–483, 2017.

[154] E. Soltanaghaei, A. Prabhakara, A. Balanuta, M. Anderson, J. M. Rabaey, S. Kumar, and R. Anthony, "Millimetro: mmwave retro-reflective tags for accurate, long range localization," in Proceedings of ACM MobiCom, New Orleans, LA, USA, October 25-29, 2021.

[155] K. Ramasubramanian, "mmwave radar for automotive and industrial applications," Texas Instruments, 2017.

[156] M. U. Hassan, M. H. Rehmani, and J. Chen, "Differential privacy techniques for cyber physical systems: A survey," IEEE Communications Surveys & Tutorials, 2020.

[157] F. Hussain, R. Hussain, S. A. Hassan, and E. Hossain, "Machine learning in iot security: Current solutions and future challenges," IEEE Communications Surveys & Tutorials, 2020.

[158] T. I. Incorporated, "Mmwcas-rf-evm: mmwave cascade imaging radar rf evaluation module," https://www.ti.com/tool/MMWCAS-RF-EVM?keyMatch=MMWCAS-RF-EVM, 2022.

[159] L. Rabiner, R. W. Schafer, and C. Rader, "The chirp z-transform algorithm," IEEE transactions on audio and electroacoustics, vol. 17, no. 2, pp. 86–92, 1969.

[160] M. H. Mazaheri, A. Chen, and O. Abari, "mmtag: A millimeterwave backscatter network," in Proceedings of ACM SIGCOMM, Virtual Event, USA, August 23-27, 2021.

[161] S. Basak and M. Gowda, "mmspy: Spying phone calls using mmwave radars," in 2022 IEEE Symposium on Security and Privacy (SP). IEEE, 2022, pp. 1211–1228.

[162] A. Gadre, D. Vasisht, N. Raghuvanshi, B. Priyantha, M. Kotaru, S. Kumar, and R. Chandra, "Milton: Sensing product integrity without opening the box using non-invasive acoustic vibrometry," in 2022 21st ACM/IEEE International Conference on Information Processing in Sensor Networks (IPSN). IEEE, 2022, pp. 390–402.

[163] K. Bansal, K. Rungta, S. Zhu, and D. Bharadia, "Pointillism: Accurate 3d bounding box estimation with multi-radars," in Proceedings of the 18th Conference on Embedded Networked Sensor Systems, 2020, pp. 340–353.

[164] I. K. Jain, R. Subbaraman, and D. Bharadia, "Two beams are better than one: towards reliable and high throughput mmwave links," in Proceedings of the 2021 ACM SIGCOMM 2021 Conference, 2021, pp. 488–502.

[165] Y. Li, Y. Liu, Y. Wang, Y. Lin, and W. Shen, "The millimeter-wave radar slam assisted by the rcs feature of the target and imu," Sensors, vol. 20, no. 18, p. 5421, 2020.

[166] R. Caruana, "Multitask learning," Machine learning, vol. 28, no. 1, pp. 41–75, 1997.

[167] C. Wang, F. Lin, T. Liu, K. Zheng, Z. Wang, Z. Li, M.-C. Huang, W. Xu, and K. Ren, "mmeve: eavesdropping on smartphone's earpiece via cots mmwave device," in Proceedings of the 28th Annual International Conference on Mobile Computing And Networking, 2022, pp. 338–351.

[168] Z. Li, B. Chen, X. Chen, H. Li, C. Xu, F. Lin, C. X. Lu, K. Ren, and W. Xu, "Spiralspy: Exploring a stealthy and practical covert channel to attack air-gapped computing devices via mmwave sensing," in The 29th Network and Distributed System Security (NDSS) Symposium 2022. The Internet Society, 2022.

[169] T. Woodford, X. Zhang, E. Chai, and K. Sundaresan, "Mosaic: leveraging diverse reflector geometries for omnidirectional around-corner automotive radar," in Proceedings of the 20th Annual International Conference on Mobile Systems, Applications and Services, 2022, pp. 155–167.

[170] J. Wang, H. Hassanieh, D. Katabi, and P. Indyk, "Efficient and reliable low-power backscatter networks," in Proceedings of ACM SIGCOMM, Helsinki, Finland, August 13-17, 2012.

[171] X. Na, X. Guo, Z. Yu, J. Zhang, Y. He, and Y. Liu, "Leggiero: Analog wifi backscatter with payload transparency," in Proceedings of the 21th Annual International Conference on Mobile Systems, Applications and Services, 2023.

[172] X. Guo, L. Shangguan, Y. He, J. Zhang, H. Jiang, A. A. Siddiqi, and Y. Liu, "Aloba: Rethinking on-off keying modulation for ambient lora backscatter," in Proceedings of ACM SenSys, Virtual Event Japan, November 16-19, 2020.

[173] V. Talla, M. Hessar, B. Kellogg, A. Najafi, J. R. Smith, and S. Gollakota, "LoRa backscatter: Enabling the vision of ubiquitous connectivity," in Proceedings of ACM UbiComp, Maui, HI, USA, September 11-15, 2017.

[174] X. Guo, Y. He, Z. Yu, J. Zhang, Y. Liu, and L. Shangguan, "Rf-transformer: A unified backscatter radio hardware abstraction," in Proceedings of ACM MobiCom, Sydney, Australia, October 17-21, 2022.

[175] X. Guo, L. Shangguan, Y. He, N. Jing, J. Zhang, H. Jiang, and Y. Liu, "Saiyan: Design and implementation of a low-power demodulator for lora backscatter systems," in Proceedings of USENIX NSDI, Renton, WA, USA, April 4-6, 2022.

[176] H. Jiang, J. Zhang, X. Guo, and Y. He, "Sense me on the ride: Accurate mobile sensing over a lora backscatter channel," in Proceedings of ACM SenSys, Coimbra, Portugal, November 15-17, 2021.

[177] V. Liu, A. Parks, V. Talla, S. Gollakota, D. Wetherall, and J. R. Smith, "Ambient backscatter: Wireless communication out of thin air," in Proceedings of ACM SIGCOMM, Hong Kong, China, August 12-16, 2013.

[178] M. Katanbaf, A. Weinand, and V. Talla, "Simplifying backscatter deployment: Full-duplex lora backscatter," in Proceedings of USENIX NSDI, virtual, April 12-14, 2021.

[179] Z. Chi, X. Liu, W. Wang, Y. Yao, and T. Zhu, "Leveraging ambient lte traffic for ubiquitous passive communication," in Proceedings of ACM SIGCOMM, Virtual Event, USA, August 10-14, 2020.

[180] X. Guo, L. Shangguan, Y. He, J. Zhang, H. Jiang, A. A. Siddiqi, and Y. Liu, "Efficient ambient lora backscatter with on-off keying modulation," IEEE/ACM Transactions on Networking, vol. 30, no. 2, pp. 641–654, 2022.

[181] Y. Chae, K. Bae, P. Pathak, and S. Kim, "On the feasibility of millimeter-wave backscatter using commodity 802.11ad 60 ghz radios," in Proceedings of ACM WiNTECH, London, United Kingdom, September 21, 2020.

[182] Y. Zhang, X. Wang, C. Feng, X. Li, Y. Cai, Y. Ren, F. Wang, and K. Li, "Pushing the limits of respiration sensing with reconfigurable metasurface," in Proceedings of ACM SenSys, Coimbra, Portugal, November 15-17, 2021.

[183] C. Feng, X. Li, Y. Zhang, X. Wang, L. Chang, F. Wang, X. Zhang, and X. Chen, "Rflens: Metasurface-enabled beamforming for iot communication and sensing," in Proceedings of ACM MobiCom, New Orleans, LA, USA, January 31-February 4, 2021.

[184] D. Xie, X. Wang, and A. Tang, "Metasight: Localizing blocked rfid objects by modulating nlos signals via metasurfaces," in Proceedings of ACM MobiSys, Portland, OR, USA, June 25-July 1, 2022.

[185] M. Dunna, C. Zhang, D. Sievenpiper, and D. Bharadia, "Scattermimo: Enabling virtual mimo with smart surfaces," in Proceedings of ACM MobiCom, London, United Kingdom, September 21-25, 2020.

[186] R. I. Zelaya, R. Ma, and W. Hu, "Towards 6g and beyond: Smarten everything with metamorphic surfaces," in Proceedings of ACM HotNets, Virtual Event, UK, November 10-12, 2021.

[187] J. Xu, C. Yuen, C. Huang, N. Ul Hassan, G. C. Alexandropoulos, M. Di Renzo, and M. Debbah, "Reconfiguring wireless environments via intelligent surfaces for 6g: reflection, modulation, and security," Science China Information Sciences, vol. 66, no. 3, p. 130304, 2023.

[188] K. Qian and X. Zhang, "Fully passive 3d printed reflecting surface for millimeter-wave coverage expansion," in Proceedings of ACM MobiSys, Portland, OR, USA, June 25-July 1, 2022.

[189] K. W. Cho, M. H. Mazaheri, J. Gummeson, O. Abari, and K. Jamieson, "mmwall: A reconfigurable metamaterial surface for mmwave networks," in Proceedings of ACM HotMobile, United Kingdom, February 24-26, 2021.

[190] H. Chen and Y. Ghasempour, "Malicious mmwave reconfigurable surface: Eavesdropping through harmonic steering," in Proceedings of ACM HotMobile, Tempe, AZ, USA, March 9-10, 2022.

[191] M. Kokkoniemi, Joonas andJuntti, "Channel modeling for reflective phased array type riss in mmwave networks," in Proceedings of ACM mmNets, New Orleans, LA, USA, January 31-February 4, 2021.

[192] I. Union, "Imt traffic estimates for the years 2020 to 2030," Report ITU, vol. 2370, 2015.

[193] E. T. Desk, "China claims new 'world record' in data streaming using next-gen wireless communication," 2022, https://eurasiantimes.com/100-times-faster-china-claims-new-world-record-in-data-streaming/.

[194] N. Kumar, "Which country has a 6g network?" 2022, https://techbyte.co.in/6g-network-what-is-6g-in-india-application/.

[195] G. P. Fettweis, "The tactile internet: Applications and challenges," IEEE Vehicular Technology Magazine, vol. 9, no. 1, pp. 64–70, 2014.

[196] A. Kludze, R. Shrestha, C. Miftah, E. Knightly, D. Mittleman, and Y. Ghasempour, "Quasi-optical 3d localization using asymmetric signatures above 100 ghz," in Proceedings of the 28th Annual International Conference on Mobile Computing And Networking, 2022, pp. 120–132.

Jia Zhang is currently a PhD. student in the School of Software and BNRist of Tsinghua University. He received his B.E. degree in Tsinghua University in 2019. His research interests include Internet of Things and wireless sensing.

Rui Xi is currently an Assistant Researcher at School of Computer Science and Engineering in University of Electronic Science Technology of China (UESTC) since August 2022. During 2019.07 - 2022.07, he was a postdoctoral researcher at Tsinghua University, worked with Prof. Yuan He. He obtained the M.S. and Ph.D. degrees in Computer Science from UESTC, Chengdu, China, in 2014 and 2019, respectively. His main research interests include Internet of Things, wireless sensing and AI.

Yuan He is an associate professor in the School of Software and BNRist of Tsinghua University. He received his B.E. degree in the University of Science and Technology of China, his M.E. degree in the Institute of Software, Chinese Academy of Sciences, and his PhD degree in Hong Kong University of Science and Technology. His research interests include wireless networks, Internet of Things, pervasive and mobile computing. He is a senior member of IEEE and a member of ACM.

Yimiao Sun is currently a PhD. student in Tsinghua University. He received his B.E. degree in the University of Electronic Science and Technology of China (UESTC) in 2021. His research interests include Internet of Things and wireless sensing.

Xiuzhen Guo is a postdoc in the School of Software and BNRist of Tsinghua University. She received her B.E. degree in Southwest University, and her PhD degree in Tsinghua University. Her research interests include wireless networks and Internet of Things.

Weiguo Wang is currently a PhD. student in Tsinghua University. He received his B.E. degree in the University of Electronic Science and Technology of China (UESTC). His research interests include acoustic sensing and mobile computing.

Xin Na is currently a PhD. student in Tsinghua University. He received his B.E. degree in Tsinghua University in 2020. His research interests include wireless networks and Internet of Things.

Yunhao Liu received his B.S. degree in Automation Department from Tsinghua University, and an M.A. degree in Beijing Foreign Studies University, China. He received an M.S. and a Ph.D. degree in Computer Science and Engineering in Michigan State University, USA. He is now a professor at Automation Department and Dean of the GIX in Tsinghua University, China. He is a Fellow of IEEE and ACM.

Zhenguo Shi received his M.S.(in 2011) degree and PhD (in 2016) from Harbin Institute of Technology (HIT), China. He received his second PhD degree (in 2022) from University of Technology Sydney, Australia. He was a visiting student at the University of Leeds (2012-2013) and a visiting scholar at the University of Technology Sydney, Australia (2016-2017). He is currently working as a Postdoctoral Research Associate in the School of Computing at the Macquarie University, Australia. His research interests include Wireless sensing, Human activity recognition, IoT, Deep learning, Cognitive radio, Interference alignment, Massive MIMO and wireless communication.

Tao Gu is currently a Professor at Macquarie University, and he is an expert in the fields of Internet of Things, Embedded/Edge AI, Mobile and Ubiquitous Computing. His publications typically appear in conferences including MobiCom, SenSys, IPSN, UbiComp and INFOCOM. He has also served a leadership role in many conferences, including General Co-Chair of MobiCom 2022 and TPC Co-Chair of IoTDI 2021.