WiFi Vision: Sensing, Recognition, and Detection With Commodity MIMO-OFDM WiFi

Ying He, *Graduate Student Member, IEEE*, Yan Chen, *Senior Member, IEEE*, Yang Hu, and Bing Zeng, *Fellow, IEEE*

***Abstract*—Indoor human sensing, recognition, and detection, as key enablers of building smart environments, such as smart home, smart retail, and smart museum, have gained tremendous attention in recent years. Compared with traditional vision-based and wearable sensor-based solutions, radio-frequency (RF)-based approaches are more desirable with the contactless and nonline-of-sight nature. Among all RF-based approaches, WiFi-based approaches have been the focus of many researchers because of the ubiquitous availability and cost efficiency. In this article, we present a survey of recent advances in WiFi vision problems, i.e., sensing, recognition, and detection by utilizing the channel state information (CSI) of the commodity WiFi devices. We focus on nine key applications of smart environments, including WiFi imaging, vital sign monitoring, human identification, gesture recognition, gait recognition, daily activity recognition, fall detection, human detection, and indoor positioning. Such a survey can help readers have an overall understanding of sensing, recognition, and detection with commodity WiFi, and thus expedite the development of smart environments.**

***Index Terms*—Channel state information (CSI), detection, recognition, sensing, WiFi vision.**

I. Introduction

INDOOR human sensing, recognition, and detection are key enablers of building smart environments that can analyze and react to our daily activities. Smart environments include a smart home that can detect the location and movement of human for optimizing energy consumption as well as monitor vital sign for healthcare, smart retail with shelves that can measure consumers' interest level of products, and smart museum that can guide tourists. Due to the wide application areas, human sensing, recognition, and detection have gained a lot of attention recently.

Manuscript received January 6, 2020; revised March 24, 2020; accepted April 18, 2020. Date of publication April 22, 2020; date of current version September 15, 2020. This work was supported in part by the National Natural Science Foundation of China under Grant 61672137, and in part by the 111 Project under Grant B17008. The work of Yan Chen was supported in part by the Thousand Youth Talents Program of China. *(Corresponding author: Yan Chen.)*

Ying He, Yang Hu, and Bing Zeng are with the School of Information and Communication Engineering, University of Electronic Science and Technology of China, Chengdu 611731, China (e-mail: heying@std.uestc.edu.cn; yanghu@uestc.edu.cn; eezeng@uestc.edu.cn).

Yan Chen was with the School of Information and Communication Engineering, University of Electronic Science and Technology of China, Chengdu 611731, China. He is now with the School of Cyberspace Security, University of Science and Technology of China, Hefei 230026, China (e-mail: eecyan@ustc.edu.cn).

Digital Object Identifier 10.1109/JIOT.2020.2989426

Existing solutions for human sensing, recognition, and detection can be classified into three categories: 1) vision-based approaches; 2) wearable sensor-based approaches; and 3) radio-frequency (RF)-based approaches. Traditional vision-based approaches can be used for passive human sensing, recognition, and detection, which rely on cameras in the area of interest. The cameras gather snapshots of human, and then computer vision and machine learning techniques are used for understanding human activities. However, this kind of approaches has some limitations, including the Line-of-Sight (LOS) requirement, which means the presence of obstacles (wall, door, etc.) will prevent accurate detection, the lighting requirement, which means that in the dark or dim environment, it is hard to sense human, and privacy intrusion, i.e., the camera in the room is privacy intrusion and cannot be employed in many positions such as the bathroom.

Wearable sensor-based approaches are commonly used as an alternative to assess human activities. In these systems, the devices have to be attached to the individual. Inner sensors of these devices, such as accelerometers, magnetometers, or gyroscopes record data that can reveal the motion of individuals. Sensor data can be processed for gait recognition or activity classification [1]–[3]. Unfortunately, the requirement of wearing a device all the time is cumbersome and may be impossible sometimes. Hence, it is hard to achieve ubiquitous human monitoring.

The contactless and non-LOS nature make the RF-based approaches much more desirable. In RF-based systems, the wireless signals emitted from the transmitter, before arriving at the receiver, are modulated by the human and the corresponding activities in the environment. By analyzing the modulated wireless signals at the receiver, it is able to achieve human sensing, recognition, and detection. The past two decades have witnessed extensive RF-based techniques, including RFID, Bluetooth, UWB, and WiFi.

RFID is a means of storing and retrieving data through the electromagnetic transmission to an RF compatible integrated circuit [4]. An RFID system has several basic components, including a number of RFID readers, RFID tags, and the communication between them. A number of indoor localization, tracking, and activity recognition systems are developed by utilizing RFID [5]–[7]. For example, LANDMRC [8] is an indoor localization system using active tags. The tags include reference tags with known locations and tracking tags attached to the target. Signal strengths of the reference tags and tracking

TABLE I
THREE DIFFERENT CATEGORIES OF NINE WIFI VISION APPLICATIONS

Category	Application	Reference	Description
Sensing	WiFi imaging	[23] [24] [25] [26] [27]	Image human or other objects
	Vital sign monitoring	[28] [29] [30] [31] [32] [33] [34]	Monitor breath or heartbeat
Recognition	Human identification	[35] [36] [37] [38]	Identify people without LOS condition
	Gesture recognition	[39] [40] [41] [42] [43] [44]	Gesture recognition for human-computer interaction
	Gait recognition	[45] [46]	Estimate stride length and rate to recognize individuals
	Daily activity recognition	[47] [48] [20] [49] [19] [21] [50] [22] [51] [52]	Recognize standing,cooking, smoking, etc.
Detection	Fall detection	[53] [54] [55] [56]	Detect fall in time to avoid serious injuries
	Human detection	[57] [58] [59] [60] [61] [62] [63]	Detect human presence in the area of interest
	Indoor positioning	[64] [65] [66] [67] [68] [69] [70] [71] [72] [73] [74] [75] [76] [77] [78] [79] [80]	Localize target in an indoor environment

tags received on RFID readers are compared with each other to determine the target location according to the reference points. The 90% accuracy of this system are around 5.93–5.97 m, depending on the number of tags and placement.

The Bluetooth is devised as an open specification for low power, short-range wireless data, and voice connections [9]. It can also be used for localization and tracking people. Bluetooth tags as small transceivers have their unique IDs, which can be used for localization. Localization systems based on Bluetooth utilize the dependence between the received signal strength and distance to calculate the distance with meter-level accuracy [10], [11]. In comparison with WLAN, the gross bit rate of Bluetooth is lower and the range is shorter (typically 10–15 m) [12]. The short-range feature makes it suffer from the dense deployment for localization.

UWB is defined as an RF signal occupying a portion of the frequency spectrum that is greater than 20% of the center carrier frequency or with a bandwidth greater than 500 MHz. The resolution of the system in the time domain is a function of the total bandwidth. Since the wider bandwidth leads to higher time resolution, UWB is considered as a promising technology for indoor localization and tracking [11], [13]–[15]. However, UWB would continuously occupy a wide band and may impose interference on other devices that share the same frequency.

Compared with RFID, Bluetooth, and UWB, WiFi is much more widely used in our daily lives. Besides communications, it is found that WiFi signals are sensitive to environmental dynamics caused by human motion. Therefore, WiFi provides a desirable solution for indoor human sensing, recognition, and detection. The main benefits are twofold: 1) ubiquitously available due to the widely deployed WiFi infrastructure and 2) cost effective without hardware modification and redeployment.

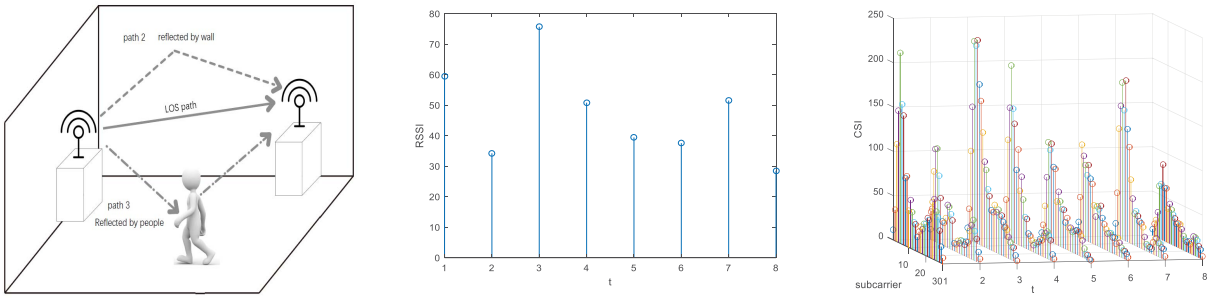
In this article, we would like to conduct a survey about state-of-the-art human sensing, recognition, and detection systems based on the channel state information (CSI) of the commodity WiFi devices. We notice that there have been some surveys about human activity recognition [16]–[18]. Specifically, in [17], WiFi radar systems are discussed and divided into two main streams: 1) data driven and 2) model-based approaches. General system architecture and applications corresponding to the two streams are introduced [17]. However, no more than four systems belonging to each stream are discussed, and the detailed processing steps are not discussed and compared. Yousefi *et al.* [16] only focused on data-driven approaches of activity recognition. Four WiFi-based activity recognition systems are surveyed and they are divided into two categories according to the extracted features: 1) histogram [19] and 2) short-time Fourier transform (STFT) features [20]–[22]. Moreover, the systems surveyed in [16] and [17] are only related to human daily activities recognition and gesture recognition. Jiang *et al.* [18] surveyed more related works in four categories: 1) health monitoring; 2) gesture recognition; 3) contextual information acquisition; and 4) authentication, but miss some important applications.

Compared to the existing surveys, this article is much more comprehensive with the coverage of nine kinds of applications, i.e., WiFi imaging, vital sign monitoring, human identification, gesture recognition, gait recognition, daily activity recognition, fall detection, human detection, and indoor positioning. Some applications that are not discussed in the existing surveys are also of vital importance for building a smart environment. For example, human detection to determine whether there are people in the area of interest is a key enabler for intrusion detection and emergency response. WiFi imaging helps us see things even in dark conditions or with obstacles. In addition, for the systems in each category, we review their methodologies and performance in detail one by one and provide a summary of the general framework of each category as well as a comparison among different systems in the same category. Notice that such comparison and summarization can help readers have an overall understanding of how to realize a specific application, but cannot be found in the existing surveys [16]–[18].

The remainder of this article is organized as follows. In Section II, we introduce the concept of CSI. In Section III, we discuss the detailed preprocessing steps to obtain the calibrated CSI information. In Section IV, we survey nine different applications related to sensing, recognition, and detection, and summarize them in Table I. Finally, the conclusions are drawn in Section V.

II. CHANNEL STATE INFORMATION

WiFi is one kind of wireless local area network technologies based on the family of IEEE 802.11 standard protocols (e.g., IEEE 802.11 a/b/g/n/ac). With a WiFi access point (AP), wireless data transmission among smart devices can be realized. With the advantages of providing online services with high data rates, great mobility, and broad coverage [17], WiFi has been utilized to connect devices, such as computers, smartphones, TVs, game consoles, surveillance



(a) (b) (c)

Fig. 1. Variations of the environment are modulated in the WiFi signals, more specifically in the CSI. (a) Multipath. (b) RSSI. (c) CSI.

cameras, refrigerators, loudspeakers, etc., to home networks and Internet. This provides a larger number of WiFi links, some of which use a constant beacon. With CSI tools, the CSI, which represents the channel properties of the communication link between the transmitter and the receiver, can be extracted from the signals of the commodity WiFi devices equipped with network interface card (NIC) conveniently [81]. More specifically, the CSI reveals how the signal propagates through the channel. The impairments that the signal undergoes due to the multipath, shadowing, and scattering, are embedded in the CSI, as shown in Fig. 1.

Two important techniques adopted in the 802.11 protocol are orthogonal frequency-division multiplexing (OFDM) and multiple-input–multiple-output (MIMO), which enable WiFi signals to provide more fine-grained channel measurements. With the OFDM technique, the signal is transmitted across multiple orthogonal subcarriers at different subcarrier frequencies, which means that the CSI is measured at the subcarrier level. With the MIMO technique, the signal is transmitted with multiple antennas to exploit the spatial diversity to increase the diversity gain, array gain, and multiplexing gain [82]. Therefore, the signal received at the receiver can be written as

$$\mathbf{y}_i = \mathbf{H}_i \mathbf{x}_i + \mathbf{n}_i \tag{1}$$

where i is the subcarrier index, $\mathbf{x}_i \in R^{N_T}$ and $\mathbf{y}_i \in R^{N_R}$ are the transmitted signal and the received signal, respectively, N_T is the number of transmitter antennas, N_R is the number of receiver antennas, $\mathbf{n}_i$ is the noise vector, and $\mathbf{H}_i \in C^{N_R \times N_T}$ denotes the CSI matrix of the subcarrier i

$$\mathbf{H}_i = \begin{pmatrix} h_i^{11} & h_i^{12} & \cdots & h_i^{N_T} \\ h_i^{21} & h_i^{22} & \cdots & h_i^{2N_T} \\ \vdots & \vdots & \ddots & \vdots \\ h_i^{N_R 1} & h_i^{N_R 2} & \cdots & h_i^{N_R N_T} \end{pmatrix} \tag{2}$$

where h_i^{mn} is the CSI of the ith subcarrier for the link between the mth receiver antenna and the nth transmitter antenna.

h_i^{mn} is a complex value, which can be represented as

$$h_i^{mn} = |h_i^{mn}| e^{j \angle h_i^{mn}} \tag{3}$$

where $|h_i^{mn}|$ and $\angle h_i^{mn}$ denote the amplitude and phase, respectively.

The CSI in (2) captures the status of the environment. As shown in Fig. 1, the WiFi signals received at the receiver are the superposition of signals from different paths, which are caused by the objects in the environment. Therefore, the environment information is modulated in the WiFi signals and can be captured by the CSI. By analyzing the CSI, we can extract the variations of the environment and accomplish the vision tasks. Notice that the RSSI is the total received power at the receiver [83], which is upper bounded by the sum of the power of each element in the CSI matrix (2). Therefore, compared with the RSSI, the CSI is a fine-grained characterization of the environment that can be utilized to sense, recognize, and detect the variations of the environment.

III. CSI Preprocessing

Due to various reasons, such as the hardware imperfection of the commodity WiFi devices and the interference from nearby devices, the CSI measurements may be corrupted. To obtain useful information from the CSI, the CSI sanitization is necessary. In general, the CSI sanitization includes denoising, outlier removal, interpolation, and phase calibration, as follows.

A. Denoising

In general, the speed of human motion, such as gesture, breath, and walking is low. The variations of CSI caused by these low-speed motions lie in the low-frequency spectrum, while the noise introduced by the propagation environment and hardware imperfection has a relatively high frequency. Therefore, the Butterworth low-pass filter is a natural choice to eliminate noise in such a situation [19], [21], [40], [55], [84]. The Butterworth filter has a flat frequency response in the passband and rolls off toward zero in the stopband, which ensures that the information in the signal will not be significantly distorted when removing the out-band noise. However, the internal state transitions in the transmitter and the receiver, including the change of power and adaptation of transmission rate, introduce a high amplitude impulse and burst noises in the CSI streams, and the low-pass filters do not perform well under such a circumstance.

Due to the high energy and large bandwidth of impulse noises, the passband of the low-pass filter usually needs

to be smaller than one-twentieth of the sampling rate to suppress the residual energy of the noise [20]. When the cut-off frequency is low enough to remove the noise, the useful signal will be attenuated. The principal component analysis (PCA)-based denoising algorithm is adopted to remove this kind of noise [20], [22], [46]. The variations of CSI caused by human motion are correlated across streams, PCA can discover the correlations and extract principal components of the CSI streams. Noise caused by the internal state transitions presents in all CSI streams. Due to the high correlation, it is captured in the first principal component. Discarding the first component can eliminate the noise. Besides, PCA removes the uncorrelated noise by discarding those noisy components. Liu *et al.* [28], [29] argued that Butterworth low-pass filters are not appropriate to remove the high-frequency noise in CSI since conventional filters not only filter out noise but also blur the rising/falling edges that possibly appear in the CSI signal. They apply wavelet filters instead since it can preserve extremely well the sharp transitions in the CSI.

B. Outliers Removing

Sometimes there are some anomalous measurements, outliers, that may be caused by the internal state transition. These outliers are some abrupt changes of the CSI amplitude which look like impulses. The Hampel identifier is often utilized to eliminate these outliers. It declares that any points falling out of a certain interval $[\mu - \gamma * \sigma, \mu + \gamma * \sigma]$ are outliers, where μ and σ are median and median absolute deviation of the data sequence, respectively, and the standard threshold γ can be determined by some preliminary measurements [28], [30], [31], [40].

C. Interpolation

For some applications, such as respiration detection and gait recognition, an accurate time index of each received packet is needed. However, even the packets are transmitted with a fixed transmission rate, due to the packet loss, transmission delay, and other processing delays, the received packets may be nonuniformly distributed, which leads to distorted time index information. In such a case, interpolation is necessary to ensure that the CSIs are evenly spaced in time [28], [50], [55].

D. Phase Calibration

It is shown by experiments that the phase of CSI is more sensitive to the environment change than the amplitude of CSI [47]. However, the phase of CSI obtained by the commodity WiFi devices is usually distorted by various offsets, including carrier frequency offset (CFO) caused by the unsynchronized frequency oscillator between the transmitter and the receiver, sampling frequency offset (SFO) generated by the frequency mismatch between analog-to-digital (A/D) converter of the transmitter and the receiver, packet detection delay (PDD) generated when the receiver does not detect the packet in time, and the phase-locked-loop (PLL) initial phase due to different antennas using different RF links. Because of these offsets, the raw phase becomes randomly distributed, making it inapplicable for the recognition and detection tasks. In an ideal situation, the CSI of the kth subcarrier for the mth antenna at time t can be written as follows:

$$C(m,k,t) = \sum_{l=1}^{L} a_l(t) e^{-j2\pi f_k \tau_l^m(t)} \tag{4}$$

where $a_l(t)$ is the reflection coefficient of the lth path at time t, $\tau_l^m(t)$ is the time delay experienced by the signal propagating along the lth path to the mth antenna, and f_k is the signal frequency of the kth subcarrier. Under the influence of the aforementioned four phase offsets, the actual CSI measured at the receiver can be written as follows:

$$C_M(m,k,t) = e^{-j2\pi(f_{\text{CFO}}t + k\triangle f(\tau_{\text{SFO}}(t) + \tau_{\text{PDD}}(t)))} e^{-j\phi_m} \sum_{l=1}^{L} a_l(t) e^{-j2\pi f_k \tau_l^m(t)} \tag{5}$$

where f_{CFO}, τ_{SFO}, τ_{PDD}, and ϕ are CFO, SFO, PDD, and PLL, respectively.

The four offsets can be divided into two categories: 1) time-varying offsets (CFO, SFO, and PDD) and 2) constant offset (PLL). The methods for removing the time-varying offsets depend on the application scenarios. When accurate phase information is necessary, we can connect the transmitter and the receiver with coaxial directly to obtain the reference signal [85], [86]. The signal received by other antennas and the reference signal have the same time-varying offsets, which can be extracted from the reference signal. However, such a direct connection is only possible when the transmitter antenna and the receiver antenna are close together and under control. When accurate phase information is not necessary, based on the observation that there are linear correlations among different subcarriers, linear transformation on the raw phase can be utilized to remove the significant component of random phase offsets [58]. Wang *et al.* [55] proposed to utilize the phase difference rather than the raw phase according to the fact that the CFO among different antennas on the same NIC is the same. The phase difference is obtained by computing the conjugation of CSI of different antennas. According to (5), the phase difference among different antennas is not affected by CFO, SFO, and PDD. Note that such phase comparison is only possible when the phases come from the same packet. The constant offset PLL is estimated in [85], where the underlying idea is that correct angle-of-arrival (AOA) estimation can be obtained when the PLL is correctly estimated.

IV. Sensing, Recognition, and Detection Based on WiFi CSI

With the aforementioned advantages of CSI, the indoor applications based on WiFi CSI have gained much attention recently and various applications related to WiFi sensing, recognition, and detection have been proposed. Sensing is to acquire information about the environment, including the static and dynamic objects. Recognition is to recognize an event from a set of events, such as human identity and daily activities. Detection is to determine whether an event happens, such as whether anyone exists or any behavior happens in the area of interest. According to the above

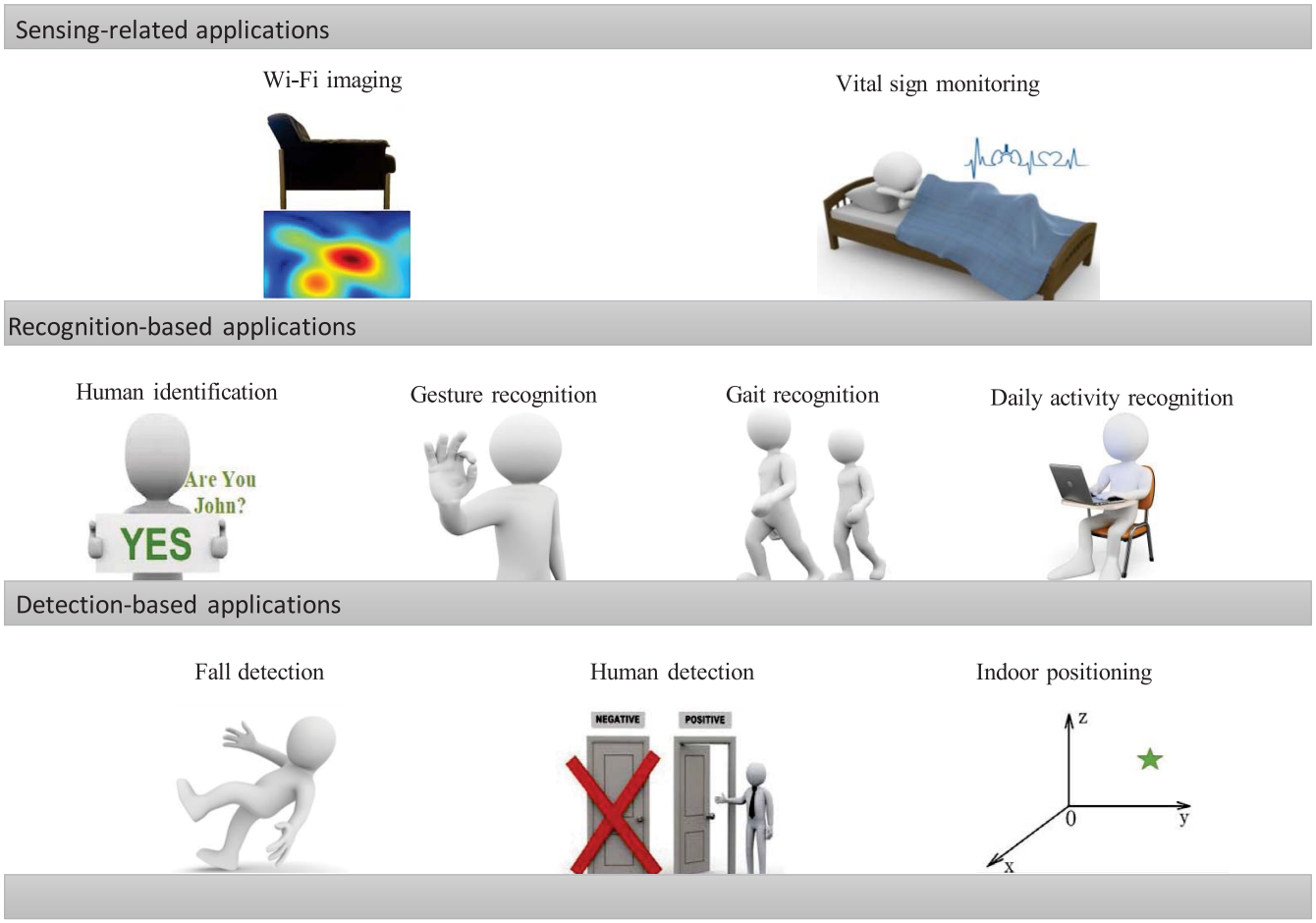


Fig. 2. Nine different categories of WiFi vision applications: WiFi imaging, vital sign monitoring, human identification, gesture recognition, gait recognition, daily activity recognition, fall detection, human detection, and indoor positioning.

descriptions of sensing, recognition, and detection, the nine applications can be divided into three categories as summarized in Table I: 1) sensing-related applications, including WiFi imaging and vital sign monitoring; 2) recognition-based applications, including human identification, gesture recognition, gait recognition, and daily activity recognition; and 3) detection-based applications, including fall detection, human detection, and indoor positioning. An illustration of these nine applications is shown in Fig. 2. The motivations of including these nine applications in our survey are as follows.

1) WiFi imaging enables us to "see" the area of interest without LOS and/or lighting requirements.
2) Vital sign monitoring is a device-free way to estimate the respiration and heartbeat rate which are important for healthcare.
3) Human identification can trigger customized settings of a smart home.
4) Gesture recognition is a key enabler for contactless human–computer interaction.
5) Gait recognition estimates walking parameters, such as stride length and stride rate, which are important parameters for accurate human identification and health condition detection.
6) Daily activity recognition helps us to recognize the user activity and further understand the user behavior.
7) Fall is the major source of serious injuries, such as fractures and head trauma, and thus the fall detection is an emerging and important topic in the healthcare industry.
8) Human detection, aiming to determine the human presence, is a key enabler for a range of indoor services, such as intrusion detection and healthcare for children and elders.
9) Indoor localization is the basis of navigation and tracking.

The general frameworks and scenarios of sensing-related applications, recognition-related applications, and detection-related applications, are shown in Figs. 3–5, respectively. For the sensing-related applications shown in Fig. 3, with a pair of WiFi devices, the systems first obtain the CSI information, then utilize the data preprocessing step to calibrate the collected CSI information, and finally, design the sensing algorithm to generate the sensing results. The recognition-related applications shown in Fig. 4 often contain two steps: 1) the offline step to extract features and train the recognition models and 2) the online step where the CSI measurements are input into the model to achieve recognition. For detection-related applications, the human presence or fall will cause unique CSI patterns, which can be utilized to realize detection-based activities. As shown in Fig. 5, the general framework of detection-related applications is similar to that

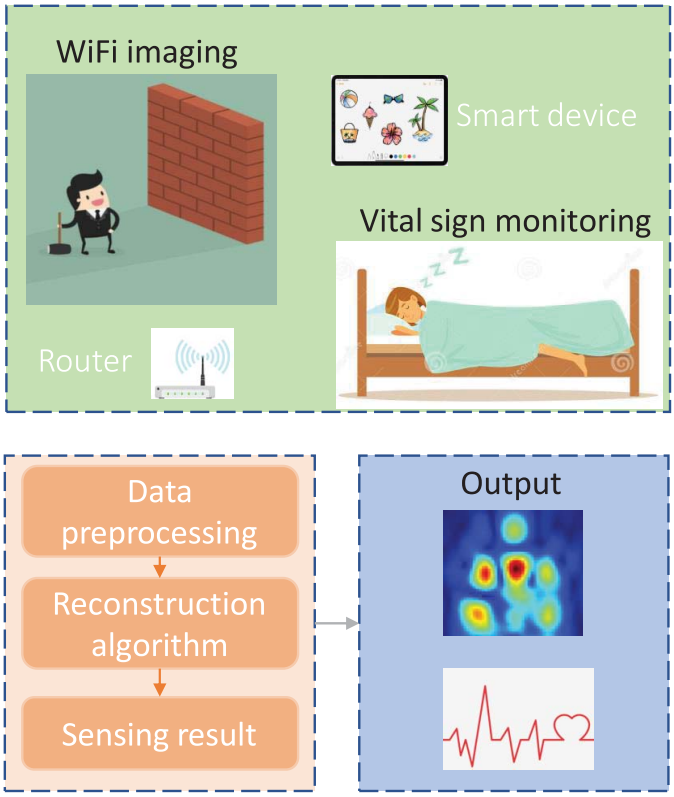


Fig. 3. Sensing-related applications: scenarios and general framework.

of sensing-related applications, i.e., including data calibration, system design, and result generation. In the following sections, we will introduce the nine applications in detail, respectively.

A. Sensing-Related Applications

1) WiFi Imaging: Traditional RF imaging systems are often expensive due to the wide bandwidth and special modulated transmission. Imaging with WiFi devices is an alternative low-cost solution to RF imaging. WiFi devices emit wireless signals which propagate in space, resulting in signal encoding the view of all objects in the environment. WiFi imaging is the process that recovers the 2-D or 3-D views of objects with the received WiFi signals.

Wi-Vi [23] is an imaging system that can achieve through-wall moving objects tracking with WiFi signals. Since the flash effect caused by the strong reflection of the wall is much stronger than the signal reflected from objects behind the wall, Wi-Vi introduces MIMO-based interference nulling to remove the flash effect. The MIMO system that consists of only two transmitter antennas and one receiver antenna, leverages the inverse synthetic aperture radar (ISAR) to exploit the moving target to emulate an antenna array. However, while removing the wall's reflection, the system also removes reflections received from other stationary objects. Therefore, Wi-Vi can only track moving objects. The system achieves good resolution by measuring the signal along with different spatial directions and time slots. For multihuman scenarios, the system utilizes the smoothed MUSIC algorithm to detect humans. The prototype of Wi-Vi is built using USRP N210 with SBX daughter boards. The experimental results show that Wi-Vi can see moving objects through walls and identify simple gestures. However, limited by 20-MHz bandwidth of the WiFi channel, Wi-Vi fails to detect humans behind concrete walls thicker than 8.

By leveraging the subcarriers as sensors, Poudel *et al.* [25] proposed a WiFi imaging system based on the CSI information. The CSI matrix can be formulated as the product of a transfer matrix and a scattering grid vector. The transfer matrix is the total electric field throughout the measurement plane without any object. It can be calculated by assuming the Born approximation with a general coherent imaging forward model. The entity of the vector indicates whether there is an object at each grid. By reconstructing the vector, the imaging of the object can be obtained. By utilizing an Intel 5300 NIC with three antennas as the receiver, and a router with two antennas as the transmitter, a proof-of-concept system for imaging human-sized objects is provided. It is also claimed that increasing the number of channels in the future will make the imaging algorithm even compelling and feasible.

Wision [24] is an imaging system that explores the feasibility and limit of leveraging WiFi signals for static objects imaging. Wision is similar to the optical imaging systems where imaging is typically formed by measuring the intensity of signal reflected from different directions. Therefore, the key problem is to differentiate the azimuth and elevation angles. With a 2-D antenna array, the phase of the received signal at each antenna should be adjusted to ensure that only signals from the right azimuth and elevation angles would be combined coherently, which can be efficiently implemented with a 2-D Fourier transform. Furthermore, the depth information along the same direction can be extracted by performing beamforming at the transmitter, i.e., the depth of the object can be identified by determining whether the signal is directed at an object when focusing the transmitter's signal at different depth of the receiver. Wision replicates a WiFi OFDM receiver on the USRP to extract CSI. By moving a USRP-N210 on a linear actuator to emulate a 2-D antenna array, Wision can obtain the imaging of objects, such as leather couches and metallic shapes with 10-MHz bandwidth. However, only relatively large objects or objects with good reflective properties within the radiation pattern of the transmitter can be imaged.

The idea of the holography is exploited to achieve 3-D imaging based on the commodity WiFi devices [26]. The holographic wave is generated at a plane with a specific depth, called a recorded plane. With a scanning antenna moving across the recorded plane and a stationary reference antenna, the holographic data are recorded by a homodyne scheme, which normalizes the data from the scanning antenna by the data from the reference antenna. Then, the 3-D view of the object is reconstructed by feeding the holographic data into a reconstruction algorithm, where the object at an arbitrary depth can be recovered from the holographic data at the recorded plane according to the angular-spectrum relation. By using a dipole emitter producing a single wave packet with a bandwidth of 22 MHz centered around the carrier frequency in accordance with the 802.11 standard, it is shown that 3-D

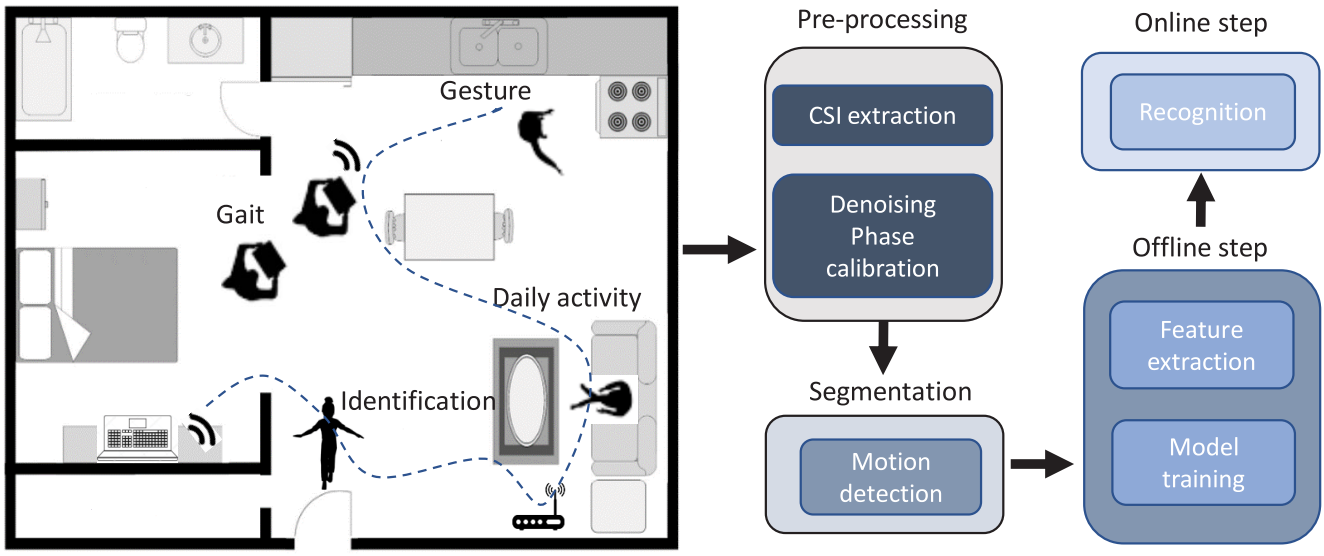


Fig. 4. Recognition-related applications: scenarios and general framework.

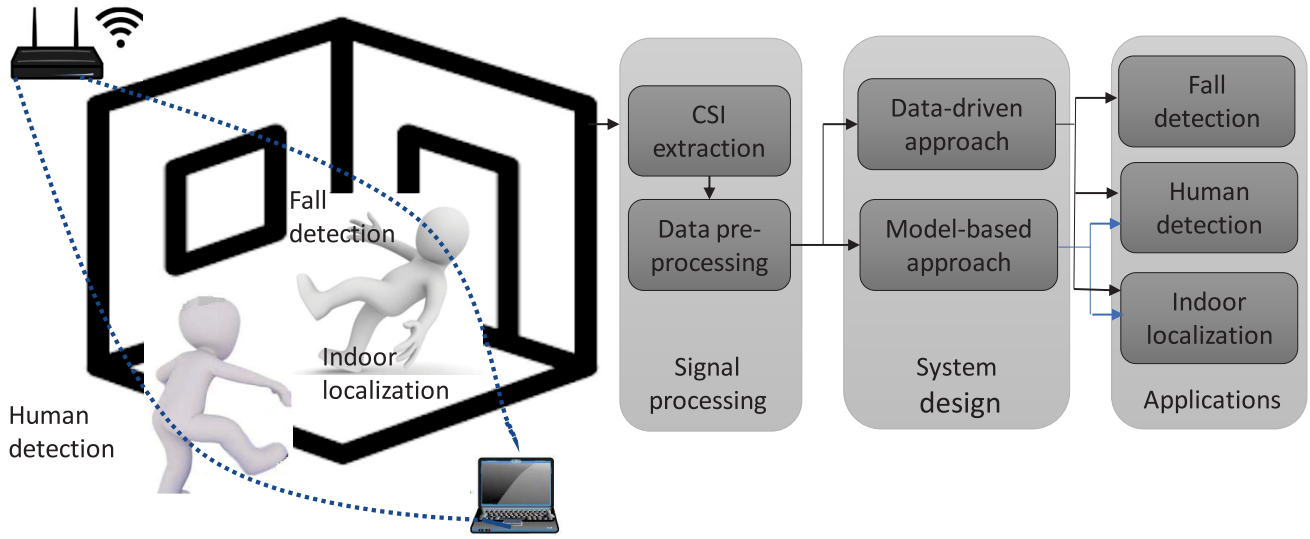


Fig. 5. Detection-related applications: scenarios and general framework.

tomography can be feasible for a cross and a 10-m sized building.

Vakalis *et al.* [27] studied the imaging by capturing signals emitted from transmitters. The captured signals measure the spatial-frequency information which is the 2-D Fourier transform of the spatial scene intensity that can build the imaging result. It is demonstrated that the signals emitted from the wireless transmitters are sufficiently incoherent. Then, according to the Van Cittert–Zernike theorem, the spatial scene intensity can be reconstructed from the measured samples through the inverse Fourier transform. By using a Keysight M8190A Arbitrary Waveform Generator to generate pseudo-random 16-QAM signals modulated on a carrier frequency of 5.5 GHz with 25-MHz bandwidth, an X-shaped target can be reconstructed with an observable contour.

Remarks: In Table II, we summarize the hardware device, bandwidth, antenna array, and imaging algorithm of the aforementioned methods. We can see that the existing imaging systems exploit either the spatial diversity (antenna array [24], [26], [27], the moving target as a virtual antenna array [23]), or the frequency diversity (different subcarriers [25]) to achieve the imaging. However, since the number of antennas and the bandwidth of the WiFi system is limited, the imaging resolution is generally not good enough. To improve the imaging resolution, one may need to increase either the number of antennas or the bandwidth. Fortunately, this is possible in the near future with the standardization of the IEEE 802.11ad where the bandwidth at the 60 GHz is much wider as well as the development of massive MIMO where the number of antennas can be large. Moreover, imaging reconstruction from the CSI information is generally ill-posed and highly nonconvex. Similar to the reconstruction of natural images, one may employ the deep learning techniques to resolve the imaging reconstruction from the CSI information. It would be interesting to investigate whether the exploration of supervised learning with deep networks can

TABLE II
WIFI IMAGING

System	Hardware device	Bandwidth	Antenna array	Imaging algorithm
[23]	USPR	20 MHz	Moving target as a virtual antenna array	Spatial angle estimation based on intensity computation
[25]	Intel 5300 NIC	20MHz/40MHz	Three receiver and two transmitter antennas	Indicative vector reconstruction based on least square minimization
[24]	USRP	10 MHz	Two dimensional synthetic receiver antenna array	Back projection based on FFT
[26]	Dipole emitter	22 MHz	Synthetic antenna array with a reference antenna	Holographic wave reconstruction
[27]	Keysight M8190A	25 MHz	T-structured 2D receiver antenna array	Inverse Fourier transform

break the resolution limitation due to the number of antennas and bandwidth. Some other future directions have also been mentioned in the aforementioned works, including using a stationary 2-D array rather than SAR to fasten imaging [26], picking antennas with better radiation patterns, or optimizing the antenna position to maximize the reach, leveraging transmissions from multiple WiFi devices in the environment [24]. It is worth pointing out that there are wide gaps in realizing high-resolution WiFi imaging, since the existing solutions only provide proof-of-concept prototypes [23]–[27].

2) Vital Sign Monitoring: Vital signs, such as breath and heartbeat statuses are important factors of health condition. Wearable devices are commonly used for daily respiration rate and heartbeat rate estimation. However, carrying the devices may be cumbersome. Nonintrusive respiration and heartbeat monitoring based on the ubiquitous WiFi devices are a favorable alternative. The basic idea of the vital sign monitoring based on the WiFi devices is that when a person breaths, the chest will have periodic movements and such movements are embedded in the signals reflected by the person.

Wi-Sleep [28] is one of such systems that use the commodity WiFi device to monitor the respiratory. It is based on the observation that the amplitude of CSI has a ripple pattern due to the human respiration. It is also pointed out in Wi-Sleep that the CSI amplitude is sensitive to the position of the human, i.e., the performance of the respiration rate estimation is affected by the position of humans and antenna pairs. Based on Wi-Sleep, Liu *et al.* [29] illustrated in detail how to obtain the respiration rate with the CSI amplitude. The motion segmentation is first performed to extract the segment of CSI that includes the respiration information. Then, the FFT is applied on the segment where only the peak and the corresponding two adjacent bins are preserved as in [87]. Finally, an inverse FFT is applied to obtain the complex time-domain signal, whose phase is linear and corresponding to the respiration rate. The system utilizes two transmitters and two receivers equipped with Intel 5300 NIC. The accuracy of respiration rate estimation is measured as $RR = (m/N)$, where m is the number of the segments where the respiration rate is successfully identified and N is the total number of segments. Under a controlled environment, the accuracy is higher than 85%.

By capturing the sinusoidal-like periodic patterns embedded in the CSI amplitude, the system proposed in [30] can estimate both the breath and heartbeat rates. It is based on the fact that the breath and heartbeat rates lie in different frequency ranges. The lower frequency information of the CSI amplitude is processed for the breath rate estimation by calculating the peak-to-peak intervals. Since the minor movements of the heartbeat do not produce the observable periodic pattern of the CSI amplitude, bandpass filters are used to remove the frequency band corresponding to the breath to facilitate the heartbeat rate estimation. This system works well in the two-person scenario where the two people breathe at different rates. With a transmitter and a receiver placed at two sides of the bed with a distance of 3 m, the system achieves the performance where over 90% estimation errors of a single person are smaller than 0.2 bpm. For the two-person scenario, 90% estimation errors are smaller than 16 bpm. When the transmitter is equipped with directional antennas, over 90% of estimation errors of heartbeat rate estimation of a single person are smaller than 4 bpm.

PhaseBeat [31] proposes to utilize the CSI phase difference to estimate the respiration rate. The DWT is employed to generate a time-frequency representation of phase difference and to remove the high-frequency noise in the phase difference. With the DWT, the phase difference is decomposed into an approximated coefficient vector and a detailed coefficient vector. PhaseBeat first uses the sliding window method to identify the peak of the approximated coefficient vector and then averages all the peak-to-peak intervals to obtain the breath rate. For the multiperson scenario, PhaseBeat leverages all CSI subcarrier phase difference series to build the correlation matrix and incorporates the standard root-MUSIC method [88] to obtain breath frequencies of different persons. By utilizing the detailed coefficient vector obtained from the DWT, PhaseBeat can also estimate the heartbeat rate. With off-the-shelf WiFi devices working at 5 GHz, PhaseBeat achieves the breath rate estimation performance with 90% of the estimation errors under 0.5 bpm in more complicated environments than those in [30]. With the directional antennas, 80% of the heartbeat rate estimation errors are under 2.5 bpm.

TR-Breath [32] proposes to combine the time-reversal technique with the root-MUSIC algorithm to estimate the respiration rates of multiple people with off-the-shelf WiFi devices. The key idea is to first leverage the time-reversal technique to coherently utilize the multipath to construct a correlation matrix and then apply the root-MUSIC algorithm on the correlation matrix to extract features for breath rate estimation for multiple persons. The average accuracy for the single-person LOS scenario is 99.56% when the experiment is conducted in a 5 m $\times$ 5 m office with two off-the-shelf WiFi devices placed at two sides of the room.

FarSense [33] aims to address the problem that existing respiration approaches only work well when the target is close to the transceiver. By the division operation, it obtains the ratio of CSI from two antennas, whose noise and phase offsets

TABLE III
VITAL SIGN MONITORING

System	Base Signal	Pre-processing	Respiration	Heartbeat	Accuracy	Hardware device	Carrier frequency	Bandwidth
[28]& [29]	Amplitude	Outlier remove, interpolation, wavelet filter	STFT, peak filter, IFFT, slope extraction	\	accuracy $\geq$ 85%	Intel 5300 NIC	2.4 GHz	20 MHz /40 MHz
[30]	Amplitude	Outlier remove, subcarrier selection	Single person: peak-to-peak interval Two people: FFT, peak frequency	Band pass filter, peak frequency	Respiration: error $<$ 0.2 bpm Heartbeat: error $<$ 4 bpm	Intel 5300 NIC	2.4 GHz	20 MHz
PhaseBeat [31]	Phase	Phase calibration, DWT	Single person: peak-to-peak interval Multiple people: root-MUSIC	DWT, peak filter, IFFT, slope extraction	Respiration: error $<$ 0.5 bpm Heartbeat: error $<$ 2.5 bpm	Intel 5300 NIC	5 GHz	20 MHz
TR-Breath [32]	Phase	Phase calibration	Root-MUSIC based on TRRS	\	Respiration: accuracy 99.56%	WiFi NIC	5.765 GHz	40 MHz
FarSense [33]	Amplitude, phase	CSI ratio	FFT, autocorrelation	\	Respiration: detection rate 95%	Intel 5300 NIC	5.24 GHz	20 MHz

are mostly canceled out to significantly increase the sensing range. FarSense combines the real part and imaginary part of the ratio of CSI to extract respiration pattern with FFT, and the respiration rate is estimated by applying autocorrelation on the respiration pattern. With a pair of GIGABYTE mini-PCs equipped with Intel 5300 NIC working at a central frequency of 5.24 GHz with 20-MHz bandwidth, FarSense achieves a detection rate higher than 95% when the person is 5.7-m away from the transceivers.

Theoretical studies about why and when the respiration can be sensed with WiFi signals have also been conducted in the literature. The Fresnel zone model is utilized to showed that there exist some blind spots where respiration detection may experience poor performance [89]. Zeng *et al.* [34] showed that the amplitude and phase of CSI are perfectly complementary to each other, i.e., a bad location for amplitude sensing turns out to be a good position for phase sensing and *vice versa*. Thus, they utilize the phase and amplitude information together to design a respiration detection scheme that enables full coverage without blind spot.

Remarks: From the above discussions, we can see that the vital sign monitoring generally involves the base signal selection, preprocessing of the base signal, and vital sign estimation. In Table III, we summarize these processes for the aforementioned methods. The basic idea of all these methods is exploiting the sinusoidal-like periodic patterns of either amplitude or phase caused by breath and heartbeat to achieve the breath rate or heartbeat rate estimation. To improve the vital sign monitoring performance, more elaborated preprocessing could be one future direction. Especially, there is still room for the improvement of phase calibration considering the components of phase errors mentioned in Section III. Besides, adopting more sophisticated methods such as time delay correlation may improve the frequency estimation accuracy of the sinusoidal-like signal for the single-person scenarios, while considering the location information to extract the vital sign caused by different people may facilitate multiperson scenarios. Other future directions include improving accuracy and robustness performance under more realistic, uncontrolled, and complicated scenarios such as the transceivers are not in the same room with the target [33], [34].

B. Recognition-Related Applications

1) Human Identification: Human identification aims to associate a person with an identity, which is the key enabler for some applications such as customization of a smart home [37]. It is found that the measured CSI from the commodity WiFi can be used to differentiate people since a different person has a different influence pattern on surrounding WiFi signals.

FreeSense [35] is a nonintrusive and privacy-preserving human identification method based on the WiFi signals. To maximize the performance, FreeSense utilizes the waveform segment corresponding to the LOS path crossing the moments and applies the dynamic time warping (DTW) technique to find the minimum distance alignment between two waveform segments. With a transmitter operating in the 802.11n AP mode and a laptop equipped with Intel 5300 NIC, which is 2.5-m away, as the receiver, FreeSense achieves the identification accuracy from 94.5% to 88.9% when the number of persons changes from 2 to 6 in a typical home environment.

WFID [36] proposes to identify humans by learning the feature pattern of the CSI. It extracts the subcarrier amplitude frequency as features and utilizes the support vector machine (SVM) classifier to differentiate different people. By using a router with one antenna as the transmitter, and a computer equipped with Intel 5300 NIC and three antennas, 3.6-m away, as the receiver, WFID achieves the human identification performance with an accuracy of 93.1% and 91.9% for the group of 6 and 9 people in two indoor scenarios (laboratory and corridor), respectively.

WiWho [37] is a system that can identify a person from a small group of people by extracting the gait profile of these people from the CSI. The system first constructs a step cycle with a peak-valley detection algorithm and extracts

TABLE IV
HUMAN IDENTIFICATION

System	Pre-processing	Feature Extraction	Classification	Accuracy	Hardware device	Carrier frequency	Bandwidth
FreeScen [35]	Low-pass filter, LOS waveform segmentation	Shape of waveform	K-NN	88.9% over 6 users	Intel 5300 NIC	2.4GHz	20 MHz /40 MHz
WFID [36]	Walking waveform segmentation	Amplitude frequency	SVM	93.1% over 6 users	Intel 5300 NIC	\	20 MHz /40 MHz
WiWho [37]	\	Gait profile	SVM	80% over 6 users	Intel 5300 NIC	\	20 MHz /40 MHz
Radio Biometrics [38]	CSI calibration, background subtraction	CSI measurements of individual	TR based similarity	98.78% over 11 users	WiFi NIC	5.845 GHz	40 MHz

the time-domain statistic features of a single step. Then, the system performs frequency domain analysis to the entire walk segment and extracts frequency domain features. All these features are combined to build up the gait profiles of people. With a pair of off-the-shelf WiFi devices equipped with three antennas, WiWho achieves an identification accuracy of 92%–80% for a group size of 2 to 6, respectively, when the walking path is in parallel to the LOS link at a distance of 1 m in office rooms and classrooms.

Radio biometrics [38] proposes a human identification system based on the time-reversal technique. It utilizes the time-reversal technique to transform the high-dimensional complex CSI multipath profiles into a real scalar similarity matrix. The radio biometric profile for each individual is defined as the CSI matrix measured when the human is presently subtracting the background CSI measured with a static environment. To realize the human identification, the system compares the measured CSI with the stored profile by computing the strength of time-reversal spatial–temporal resonance. A prototype of the TR human identification system using standard WiFi chipsets with 3×3 MIMO transmission is built and evaluated in a commercial office building. It is shown that the system can achieve an average identification rate as 98.78% when the number of participants is 11 under the through-wall scenario.

Remarks: The general processes of the aforementioned human identification systems are summarized in Table IV, including the preprocessing, feature extraction, and classification. From the above discussions, we can see that the aforementioned human identification approaches are all data driven, i.e., first, build a CSI profile for the presence of different people and then, compare the measured CSI against the stored profiles to identify different people. The handcrafted features, such as the shape of CSI waveform and amplitude frequency is used, which are dominated by the multipath effect and not robust to the environmental change. Therefore, these systems can achieve high identification accuracy only for a very small group of people. One possible way to improve performance is to leverage neural networks to extract the feature automatically. Moreover, as mentioned in [38], more transceiver pairs and larger bandwidth will make it possible to extract fine-grained biometric characteristics for human identification. Another possible way is to utilize behavioral characteristics such as gait and a daily habit for human identification. For example, WiWho [37] demonstrates the feasibility of realizing human identification by exploiting gait profiles. Other biological features, such as the muscle mass index and the body temperature can also be utilized to enhance the uniqueness of human radio biometrics [38]. Again, it is worth pointing out that there are still wide gaps in realizing human identification in uncontrolled environments in reality.

2) Gesture Recognition: Gesture recognition based on radio signals is a key enabler for the smart home. With such a technique, for example, one can change the lighting intensity in the dark with a simple hand gesture instead of looking for the switch.

WiG [39] is a gesture recognition system that attempts to recognize gestures based on the CSI of WiFi. The system first utilizes the local outlier factor (LOF) algorithm [90] to detect the anomaly segment. Then, the mean value, standard deviation, median absolute deviation, and max value of anomaly segments are chosen as features, and SVM is adopted to classify four predefined gestures. With Intel 5300 NIC working in the 2.4 GHz frequency, WiG can classify four gestures with an average accuracy of 92% in the LOS scenario and 88% in the NLOS scenario, respectively.

Wifinger [40] is a fine-grained gesture recognition system that can recognize finger movements to represent numbers. The number gestures are commonly used in our daily communication and have vital importance in developing the human–computer interface. To extract the information corresponding to the finger movements, the correlation matrix of the CSI amplitudes is obtained and the eigendecomposition is utilized to find the occurrence of finger motion. To fully represent the differential features of each finger motion, the system combines 30 subcarriers by averaging every six subcarriers and then concatenates them to form a feature vector. Since directly using the feature vectors obtained at a different time leads to a high computational complexity, the system utilizes DWT to compress the feature vector and classifies different finger motions by K-NN with the similarity metrics calculated by DTW. By setting the transmitter and the receiver with a distance of 50 cm and the user performing finger gestures in the middle of the LOS path, Wifinger achieves an average recognition accuracy of 90.4% for the 9-digit finger gestures. To further improve the recognition performance, the system proposed in [41] utilizes both the amplitude and phase information of CSI. The system explores the sensitivity of both the amplitude and phase information in the action transitions, and then uses the amplitude difference variance for motion detection. For feature extraction, the system adopts deep learning to train features. The experimental results show that this

TABLE V
GESTURE RECOGNITION

System	Pre-processing	Motion Detection	Feature Extraction	Classification	Accuracy	Hardware device	Carrier frequency	Bandwidth
WiG [39]	Bridge-massart filter	LOF	Statistic features	SVM	92% of 4 gestures	Intel 5300 NIC	2.4 GHz	20 MHz /40 MHz
WiFinger [40]	Outlier remove, low-pass filter	Motion indicator calculation	Feature vector by subcarriers averaging	K-NN	90.4% of 9 gestures	Intel 5300 NIC	5.745 GHz	20 MHz /40 MHz
DeNum [41]	Band-pass filter, phase calibration	Amplitude difference variance	Deap learning features extraction	SVM	94% of 10 gestures	Intel 5300 NIC	5 GHz	20 MHz /40 MHz
WiSign [42]	Subcarriers selection, median filter	\	Statistic features	Semi-supervised SVM	87.38% of 3 gestures	Intel 5300 NIC	2.4 GHz	20 MHz /40 MHz
WiAG [43]	PCA	First oder difference of CSI stream	DWT	Configuration based K-NN	91.4% of 6 gestures	Intel 5300 NIC	2.4 GHz	20 MHz
[44]	phase difference computation	\	2D CSI data	Siamese recurrent CNN	92.8% of 6 gestures	Atheros NIC	5 GHz	40 MHz

approach can identify ten gestures performed in the LOS path with an average accuracy of 94%, in an office scenario with the transmitter and the receiver spacing about 3 m.

The system in [42] proposes to utilize the CSI information to identify the simple sign languages, including yes, no, and thanks. The system extracts eight statistics of the CSI amplitude as features, similar to the WiFall system in [53], to build the feature profiles. It is found that not all the eight features but only the average of amplitude and average of median absolute difference (MAD) are useful for classification. To reduce the effort of generating labeled profiles, the system proposes to build the profiles semiautomatically. Specifically, for the new user with a few labeled profiles, the system trains an SVM model and a K-NN model, respectively, and classifies unlabeled data with the two trained models. If the classification results of two classifiers are the same, the new classified data can be used as new labeled data. The system also considers transfer learning to utilize the labeled profiles from other users. Since traditional SVM classifiers only support two labels, a two-stage classification with three classifiers are designed to recognize the three sign languages. By using a TP-Link TL-WR1043ND WiFi router in the AP mode at 2.4 GHz as a transmitter, and a Lenovo X100e laptop 0.2-m away as a receiver, WiSign achieves a mean prediction accuracy of 87.38%.

Most gesture recognition systems require that the position/orientation of a user does not change significantly compared to that when the user provides training samples. To resolve such a challenge, Virmani and Shahzad [43] proposed a WiAG system that can recognize the gestures of a user regardless of the position/orientation. The basic idea of WiAG is to generate virtual samples for the gestures in all possible position/orientation by applying a translation function on the training samples. The K-NN model is adopted in WiAG for the classification. Using an off-the-shelf WiFi device with two antennas as the transmitter, and an off-the-shelf WiFi device with three antennas as the receiver, WiAG improves the gesture recognition accuracy significantly from 51.4% to 91.4% when the position/orientation of a user is not the same as that at the time of collecting training samples.

Yang *et al.* [44] proposed a deep learning model to learn gesture-related spatiotemporal features. The convolutional neural network (CNN) together with a long short-term memory (LSTM) is utilized to extract the spatial and temporal features of gestures from the CSI phase difference. To make the system work in a new environment with minimal tuning, they propose a Siamese architecture with the transferable pairwise loss to retain only the features relevant to the gesture recognition task. Using two WiFi routers equipped with three antennas, this system achieves the average test accuracy of 92.8% when the training and testing data come from the same person in the same environment.

Remarks: The aforementioned gesture recognition systems share a similar framework: CSI preprocessing to eliminate noise, motion detection to extract segments corresponding to gestures, feature extraction to obtain features that can stand for different gestures, and classification to train a classifier for different gestures. The detailed components used in different systems are summarized in Table V. The challenges of gesture recognition are twofold: one is that the CSI change caused by the gesture is small; and the other is that the CSI change caused by gesture shares similar patterns. To improve the performance, extracting the information corresponding to gestures can help facilitate the micromovement gesture recognition. Moreover, deep learning can be utilized to obtain more fine-grained features [41]. Another future direction is to improve the accuracy and robustness of existing gesture recognition approaches in the multiperson scenario [41].

3) Gait Recognition: Walking is the most common activity we perform every day. Specific walking patterns of people can be robust features for people's identification. Therefore, there is a certain need for a walking monitoring system that can estimate the stride rate and stride length [46].

Maheshwari and Tiwari [45] utilized the CSI amplitude to estimate the walking stride rate based on the observations that different walks show different CSI statistical features. The system averages the CSI amplitude of all subcarriers and extracts the cumulative moving standard deviation and the mean signal velocity as features. Based on the classification tree, the system can obtain a coarse estimation of the walking stride rate to differentiate among different stride rate categories. By carrying a Dell Latitude E4300 Laptop equipped with Intel 5300 NIC on the back as the receiver, and two 802.11n wireless routers 2 inches above the ankle as

TABLE VI
GAIT RECOGNITION

System	Base Signal	Preprocessing	Estimation Algorithm	Accuracy	Hardware device	Carrier frequency	Bandwidth
[45]	Amplitude	Average among all subcarriers	Estimate stride rate range based on classification	>90%	Intel 5300 NIC	\	20 MHz /40 MHz
WifiU [46]	Amplitude	Time-frequency analysis, Spectrum enhancement	Autocorrelation to estimate cycle time Percentile method to estimate gait speed	93.05%	Intel 5300 NIC	5 GHz	20 MHz

transmitters, the system can classify walking and running with accuracy above 90%.

Another gait recognition system based on the CSI information is proposed in [46]. Since different body parts move at different speeds, the system applies the STFT to the CSI to separate the CSI information corresponding to the different body parts. Based on the feature of different body parts, the system learns the walking state from the CSI spectrum. For example, the signal reflected from the torso, the biggest part of the body, is the strongest component in the spectrum. The system first obtains a torso contour to reflect the walking procedure. Then, based on the period of walking, the system utilizes the autocorrelation of the torso contour to determine the gait cycle. The walking speed is estimated with the percentile method. By setting the transmitter and the receiver with a distance of 1.6 m, and users walking in a straight line on the LOS path, this system can estimate the gait circle and walking speed to identify human with an accuracy of 93.05%.

Remarks: Based on the CSI changes caused by human walking, gait recognition systems leverage various methods (machine learning [45] and Doppler shift [46]) to estimate gait parameters and further utilize these parameters as features to realize the human identification. The detailed components used in different systems are summarized in Table VI. However, these systems can have good performance only under restricted settings. To estimate gait parameters in a more general environment, data preprocessing needs to be improved to obtain more calibrated amplitude and phase information. Moreover, the gait recognition under a multiperson scenario is still under investigation. Finally, while the current gait systems mainly focus on the gait of a healthy human, the gait recognition for people with walking disabilities should also be considered.

4) Daily Activity Recognition: Daily activities can be divided into two categories: 1) atomic activities and 2) combined activities. Atomic activities are simple activities, such as running, sitting, and hand motion, which can be combined to form more sophisticated daily activities. This kind of activity can generally cause a significant fluctuation of the measured CSI and thus can be detected and recognized. Combined activities are complex activities, consisting of multiple atomic activities, such as cooking, bathing, and smoking. They can reveal the living condition of people. By tracking a sequence of meaningful activities and generating statistics for a person, it is possible to monitor the well being and suggest behavioral changes.

Atomic Activity Recognition: APsense [47] is a hand motion recognition system, which utilizes the features of the CSI amplitude for motion classification. The features include the statistics of the CSI amplitude of each subcarrier, as well as the first and second eigenvalues of the correlation matrices among the CSI amplitude of different subcarriers at different time instances. The features are fed into the naive Bayes classifier and decision tree to classify different hand motions. By setting the user staying on the LOS link facing the transmitter, APsense can classify four motions with an accuracy as high as 90%.

Wi-chase [48] utilizes both the amplitude and phase information of CSI to differentiate the atomic activities, including running, walking, and hand moving. Based on the observation that there are correlated variations on subcarriers due to the reflections caused by human activities, the system uses the variance among subcarriers to detect the human motion. By using a Linksys EA4500 Dual-Band router as the transmitter working at 2.4 GHz, and a laptop equipped with an Intel 5300 NIC as the receiver, Wi-chase achieves an average accuracy of 94.2%, 89.2%, and 97.3% for running, walking, and hand moving activities, respectively.

CARM [20] is an activity recognition system built on the CSI-speed model [91]. It is based on the observation that human activities can be characterized by their specific moving speed patterns. For example, walking can be characterized by a constant moving speed of around 1 m/s and falling can be characterized by a sudden speeding up within 0.5 s. The frequency components in the CSI time series are related to the moving speeds of the body, which can be separated in the frequency domain. CARM applies DWT on the denoised CSI to extract both moving duration and speed as features. The hidden Markov model (HMM) is utilized to build CSI-activity relationship of eight different activities. The hardware configuration is similar to that of Wi-chase except working at a 5-GHz band with a shorter wavelength, which leads to better distance resolution and less interference. It is shown that the average recognition accuracy of 96% can be achieved with CARM.

DFLAR [49] reorganizes the CSI information on different subcarriers at different time instances into a 2-D radio image and utilizes the image processing techniques to recognize the activities. During the offline phase, the system generates training data and extracts the color and texture features of the radio images. A deep-learning-based framework is used for extracting discriminative features from radio images and learning the parameters of the classification module. During the online phase, the learned feature generation module and the classification module are used for the measured CSI data to determine the activity and location of the target. It is shown that DFLAR classifies eight activities at twelve locations with an accuracy of 90% using commodity WiFi devices.

Combined Activity Recognition: E-eyes [19] is a device-free in-home activity identification system based on the CSI

information available in commodity WiFi devices. The basic idea of E-eyes is matching the measured CSI patterns to the CSI profiles corresponding to different activities. Since people have some specific activities at a specific location, e.g., they usually cook in the kitchen, only a few CSI profiles are needed at each location. The system first identifies walking activities and in-place activities. Walking activities cause a large cumulative moving variance of CSI samples across different subcarriers, whereas small variance indicates the presence of in-place activities such as watching TV on a sofa. The histograms of CSI amplitude are used as features, the Earth mover's distance (EMD) is adopted to measure similarities for in-place activities, and the multidimensional DTW (MD-DTW) is further used for the similarity measure of walking activities. E-eye conducts experiments in an 802.11n network with three off-the-shelf WiFi devices in two different apartments, where the bandwidth is 20 MHz. A total of nine typical daily in-place activities and eight walking activities (passing through four doorways) with different walking speeds is performed by four male adults in both apartments. For the one-bedroom apartment, the average accuracy of identifying in-place activities is 97% while in the two-bedroom apartment, the average accuracy of identifying in-place activities is 97.38%. The major limitation of E-eyes is that there is only a single occupant without pets or other movements at home, i.e., E-eyes relies on a relatively stable environment.

WiHear [21] is a word recognition system by monitoring the motion of mouth with WiFi devices. The first challenge of WiHear is to construct the mouth motion profile. To extract the reflected signal from the mouth only, WiHear exploits beamforming to find the direction of mouth to obtain maximal concentrated signal power from the mouth reflection and applies bandpass filters on the received signal to remove the out-band interference. Then, the system uses the peak-to-peak value on each subcarrier within a sliding window as features to build the mouth motion profiles, and DWT is applied on the profiles to facilitate the subsequent segmentation and classification. Another challenge of WiHear is how to recognize lip reading and translate the extracted signal features into words. After segmentation, WiHear obtains wavelet profiles for different pronunciations, each with 16 fourth-order subwaveforms from high frequency to low-frequency components, and applies a multicluster/class feature selection (MCFS) scheme in [92] to extract representative features from the wavelet profiles. Finally, DTW is utilized to classify the same word spoken at different speeds into the same group. With CSI from different receivers in multiple angles of views, WiHear can achieve recognition accuracy of 91% for a single user speaking no more than six words and up to 74% for hearing from no more than three users simultaneously.

Smokey [50] is a rhythmic smoke recognition system based on the CSI information. To extract the subtle motion caused by smoking, Smokey proposes a motion extraction method that is similar to the foreground detection in image processing. After motion detection, Smokey detects smoke based on the unique respiration pattern of the smoking period. During five workdays, two volunteers live in their own apartment and work in the same office as usual. Smokey uses commodity WiFi devices as the receivers, which operate in the IEEE 802.11n mode at 2.4 GHz. With one receiver in the office and three receivers in the apartment, Smokey successfully detects 92.8% of the smoking activities and misjudges 2.3% of the normal activities.

WiKey [22] is a keystroke recognition system based on the CSI information. The system utilizes PCA to extract the information related to the keystroke from continuously measured CSIs. The shape of the CSI waveform obtained by DWT is chosen as features, and the K-NN classifier and DTW-based similarity metric are used to differentiate different keystrokes. By placing a keyboard on the LOS path with the distances to the transmitter and the receiver being 4 and 0.5 m, respectively, WiKey achieves an average keystroke recognition accuracy of 96.4% for 37 typed keys and 93.47% for typed sentences from 80 training samples. However, it is worth pointing out that the experimental environments are relatively stable and controlled, and the performance may be affected by variations in the environments.

Besense [51] leverages commodity WiFi to recognize microgestures, such as keystrokes and mouse movements. These microgestures are small in range and short in time, and thus difficult to capture. Therefore, Besense builds a Fresnel zone-based model to guide the antenna deployment and designs a lightweight segment algorithm for motion detection to extract the segments of microgestures. The variance and time duration are used as features to determine whether the microgesture is typing or mouse moving. Three kinds of behaviors, i.e., surfing the Internet, gaming, and working, are composites of these microgestures that can be recognized through HMM. Besense achieves 93.3% recognition accuracy when ten volunteers perform these three behaviors in an office environment.

WiSDAR [52] shows that, depending on the placement of the transmitter and the receiver, there are effective areas and ineffective areas for activity recognition. Recognition algorithms only achieve reasonably good recognition results within the effective area. WiSDAR exploits the spatial diversity to resolve this issue by constructing multiple transmitter–receiver pairs and developing an algorithm to obtain the CSI from the effective area. Then, a deep learning-based framework is adopted to extract both the spatial and temporal features. With two Intel 5300 NICs, WiSDAR forms six antenna pairs and recognizes eight activities with 96% accuracy.

Remarks: Both the atomic activity recognition and combined activity recognition can be generally divided into four main steps: 1) preprocessing step; 2) motion detection step; 3) feature extraction step; and 4) classification step. The techniques used in the aforementioned daily activity recognition systems are summarized in Table VII. The features extracted by these approaches are often shallow features referring to statistic information, and the features are extracted via a handcraft way heavily relying on the experience of humans. In such a case, these approaches can achieve promising results only in controlled environments with a few labeled activities. Thus, the future directions include using deep learning to learn high-level features automatically and developing algorithms that can recognize the activities of multiple people simultaneously.

TABLE VII
DAILY ACTIVITY RECOGNITION

Category	System	Base Signal	Pre-processing	Motion Detection	Feature Extraction	Classification	Accuracy
Atomic Activity	APsense [47]	Amplitude	\	\	Statistic features, eigenvalues	Naive bayes decision tree	90%
	Wi-Chase [48]	Amplitude &phase	Low-pass filter, MCS index filter	Distribution of variance among subcarriers	Statistic features	SVM	97%
	CARM [20]	Amplitude &phase	PCA	Motion indicator calculation	Moving speed and duration calculation	HMM	96%
	DFLAR [49]	Amplitude &phase	Phase calibration, CSI to image transform	\	Radio image features, deep leaning features	Softmax regression algorithm	90%
Combined Activity	E-eyes [19]	Amplitude	Low-pass filter, MCS index filter	Coarse activity determination	Histogram of CSI amplitude	DTW based similarity calculation	96%
	WiHear [21]	Amplitude &phase	Beamforming, band-pass filter, multipath remove	Peak To peak value within sliding window	DWT based wavelet profile, multi-class feature selection	DTW or SRDA	91%
	Smokey [50]	Amplitude	Interpolation	Foreground detection	Periodicity analysis	Threshold based recognition	97.6%
	WiKey [22]	Amplitude	PCA	Moving window to detect CSI change rate	DWT based waveform shape extraction	DTW based K-NN	96.4%
	Besense [51]	Amplitude	Subcarrier selection, Butterworth filter	Variance based segmentation	Variance, slope. duration	SVM, KNN, Random Forest	93.3%
	WiSDAR [52]	Amplitude	PCA, target area dertermination	Amplitude and duration threshold	frequency components	CNN-LSTM	96%

C. Detection-Related Applications

1) Fall Detection: Fall is defined as the physical activity that a person loses the balance and falls on the ground. Fall is the major source of serious injuries, such as fractures and head trauma, leading to hospitalization and even death. Therefore, fall detection is an emerging and important topic in the healthcare industry and has attracted a lot of attention from academia in the past two decades.

The Wi-Fall in [53] analyzes the relationship between the CSI amplitude and human motion based on the radio propagation model and finds that the CSI amplitude is a robust indicator of fall. Seven features of the CSI amplitude are chosen to characterize the activity, including the normalized standard deviation, the offset of signal strength, the period of the motion, the median absolute deviation, the interquartile range, the signal entropy, and the velocity of signal change. Then, all these features are fed into SVM to differentiate the fall from the sit and the stand. By using two NICs and two routers equipped with three antennas to form two 3×3 MIMO systems, Wi-Fall realizes 87% detection rate with 18% false alarm rate in laboratory and dormitory environments.

While Wi-Fall only uses the amplitude information of CSI, Anti-Fall in [54] utilizes both the amplitude and phase information of CSI to detect fall. Due to the CFO and SFO, the raw phase of CSI is randomly distributed. To tackle this challenge, Anti-Fall [54] exploits the phase difference over two antennas and utilizes the variance of the CSI phase difference to extract the fall and fall-like segment from the measured CSI streams. Then, the same features as [53] are extracted from both CSI amplitude and phase difference information. With a pair of commercial devices operating in the 802.11n mode at 5 GHz, Anti-Fall realizes 88% detection rate with 12% false alarm rate in an office room and a meeting room.

RT-Fall [55] is another fall detection system utilizing the CSI phase information. While the CSI phase is calibrated in the same way as Anti-Fall [54], theoretical analysis about why the CSI phase difference among antennas can remove the linear distortion caused by the CFO is studied in RT-Fall [55]. Moreover, RT-Fall conducts real experiments to analyze the sensitivity of the CSI amplitude and phase difference to the human motion. The experimental results reveal that the CSI phase information, compared with the CSI amplitude information, is more suitable for fall detection. By utilizing both the power profiles and phase difference variances, the fall and fall-like activities can be differentiated from all other activities, such as eating and walking. It is worth pointing out that RT-Fall is one of the earliest works that can detect the fall with commodity WiFi devices under the practical settings, i.e., detect the fall under the condition that numerous daily activities are performed naturally and continuously. By using a router with one antenna as the transmitter, and an Intel 5300 NIC with two antennas as the receiver, RT-Fall achieves 91% detection accuracy in four different environments with three students performing fall and fall-like activities.

For the same activity, the CSI can be very different when the activity is performed in a different environment. Thus, a fall detection model trained in one specific environment may not work well in other environments. To address this challenge, TL-Fall proposes to use the transfer-learning technique for fall detection [56]. TL-Fall uses DWT to extract frequency components as features and builds an SVM model based on the transfer-learning theory to solve the adaptation problem of scenario changes. With two laptops equipped with Intel 5300 NIC as transceivers, TL-Fall achieves 86.83% detection accuracy when a single person performs fall and nonfall activities in three indoor environments.

TABLE VIII
FALL DETECTION

System	Base Signal	Pre-processing	Motion Detection	Feature	Classification	Detection Rate	Hardware device	Carrier frequency	Bandwidth
Wi-Fall [53]	Amplitude	Aggregation among subcarriers, weighted moving average along time	LOF	Statistic features	SVM	87%	Intel 5300 NIC	5 GHz	20 MHz /40 MHz
Anti-Fall [54]	Amplitude &phase	Low-pass filter	Phase difference variance	Statistic features	SVM	88%	Intel 5300 NIC	5 GHz	20 MHz /40 MHz
RT-Fall [55]	Amplitude &phase	Interpolation, band-pass filter	Phase difference variance, power profile	Statistic features	SVM	91%	Intel 5300 NIC	5 GHz	20 MHz /40 MHz
TL-Fall [56]	Amplitude	Band-pass filter, weighted moving average	PCA, duration detection	Frequency component	SVM	86.83%	Intel 5300 NIC	5.825 GHz	20 MHz /40 MHz

Remarks: The aforementioned fall detection systems [53]–[56] share a similar framework, which mainly contains four steps: 1) CSI preprocessing; 2) motion detection for extracting segments corresponding to fall or fall-like activities; 3) feature extraction to obtain features that can differentiate fall from other activities; and 4) classification to determine fall by machine learning techniques. Detailed comparisons among the aforementioned three fall detection systems are summarized in Table VIII. From the table, we can see that all three systems adopt the statistic features of either phase or amplitude to differentiate fall from other activities with the SVM classifier. When multiple people are in the area of interest, their movements will cause mutual influence to the signal propagation. As a consequence, statistic features will change because of the change of signal propagation, which makes it difficult to identify fall among multiple people. To address this issue, one may combine the location information and activities together. The performance can also be improved with more sophisticated machine learning technologies, such as decision tree, Gaussian mixture model, and deep neural networks.

2) Human Detection: The basic idea of human detection is to determine whether there is a human in an area of interest. It is a key enabler for a range of indoor services, such as intrusion detection, healthcare for children and elders, as well as emergency responses. The CSI-based human detection scheme can overcome the shortcomings of existing solutions, including specialized hardware, LOS, and illumination constraints. The existing CSI-based human detection methods can be classified into two categories as follows: 1) the feature-based methods and 2) the fingerprint-based methods.

Feature-Based Methods: The feature-based methods extract features that can distinguish the signal pattern under the static or dynamic environment. The most widely used feature is the eigenvalue of the CSI correlation matrix obtained at different times. The reason is that with the human presence, the correlation among different time is lower, i.e., the eigenvalue decreases to a small value. FIMD [57] is one of such indoor human detection systems. It exploits the correlation of CSI amplitude obtained within a time window and finds that the eigenvalue of the correlation matrix can reveal the variation of the CSI along time. Preliminary experiments show that the eigenvalue of the correlation matrix in the static status is maximum and becomes smaller in a dynamic environment. By utilizing the density-based classification algorithm DBSACN [93], with a pair of off-the-shelf WiFi devices, FIMD achieves 90% human detection rate when false positive (FP) rate is around 9% in laboratory and corridor environments.

PADS [58] and DeMan [59] leverage both the phase and amplitude information of CSI. Since the phase may be distorted by various offsets, such as CFO and SFO, phase correction is needed before it can be used for human detection. PADS [58] and DeMan [59] utilize linear transform to eliminate the phase distortion based on the assumption of linear correlation among subcarriers. Then, the maximal and second maximal eigenvalue of the phase and amplitude correlation matrix are obtained as features, and the SVM classifier is utilized to detect the presence of the human. With a single-antenna TP-link wireless router as the transmitter and a three-antenna mini PC as the receiver, PADs achieves 97% detection accuracy under three different environments where a volunteer walks through.

Fingerprint-Based Methods: The fingerprint-based methods operate in two steps, i.e., the offline calibration step and the online monitoring step. During the offline step, the CSIs measured under the scenario where no one exists in the monitored area is stored to build a normal profile. During the online step, the measured CSIs are compared with the normal profile. Once the similarity between the measured CSI during the online step and the normal profile is below a predefined threshold, the system announces that the human is present. Pilot [60] is one of such fingerprint-based human detection systems. With a pair of commercial WiFi devices as transceivers, Pilot achieves 90% detection accuracy in laboratory and lobby environments when the FP rate is 30%. By making the threshold adaptive to the environment, the method proposed in [61] is more robust to the environment change.

The off-the-shelf approaches can only have an ellipse coverage whose foci are the transmitter and the receiver. Omni-PHD [62], [63] utilizes the histogram of the CSI amplitude as features and uses the fingerprint-based algorithm to achieve omnidirectional detection coverage by building the normal profile of points outside the ellipse coverage. The EMD is adopted to measure the similarity between the measured histogram and the normal profile. By setting the person 0.5/1 m

TABLE IX
HUMAN DETECTION

System	Base Signal	Feature	Decision Method	Detection Rate	Hardware device	Carrier frequency	Bandwidth
FIMD [57]	Amplitude	Eigenvalue	DBSACN	90%	Intel 5300 NIC	2.4 GHz	20 MHz /40 MHz
PADs [58]	Amplitude &phase	Eigenvalue	SVM	97%	WiFi NIC	\	\
Pilot [60]	Amplitude &phase	CSI measurement	Crosscorrelation	90%	Intel 5300 NIC	\	20 MHz /40 MHz
Omi-PHD [62], [63]	Amplitude	Histogram of amplitude	EMD	93%	Intel 5300 NIC	2.4GHz	20 MHz

away from the receiver, Omni-PHD achieves 90% detection rate in the hall and lab environments.

Remarks: The general procedures of the aforementioned human detection systems are summarized in Table IX, which include the base signal selection, the feature extraction, and decision method. For the above discussions, we can see that human detection systems aim at extracting the CSI variations along time. This is because when the human intrusion happens, it will cause a significant variation of CSI measurements. To better exploit the variation over different times, the LSTM can be adopted. The fingerprint-based methods enjoy a high detection rate to distinguish static status and dynamic status. However, due to the multipath effect, the minor change of the environment will introduce a significant change to the normal profiles collected in the static environment, which will lead to performance degradation. To resolve this problem, environment-change-resistant features should be extracted to generate the normal profiles instead of directly adopting the measured CSI as normal profiles. We note that most existing human detection methods are evaluated under controlled environments. The robustness under more realistic environments needs to be further investigated.

3) Indoor Localization: The general indoor localization methods can be classified into two catalogs: 1) triangulation and 2) fingerprinting [12]. Triangulation-based localization approaches determine the location of the target based on the geometric properties of triangles, while the fingerprint-based localization approaches consist of an offline phase and an online phase. During the offline phase, the system measures the CSI profiles at locations of interest to build the fingerprint database. During the online phase, the measured CSI is compared with the fingerprint database to determine the location of the target.

Triangulation: There are two possible ways to perform triangulation: 1) angulation and 2) lateration. Angulation estimates the location of an object by measuring its direction, i.e., AOA, relative to the reference points. Lateration estimates the position of an object by measuring its distance, i.e., the time of flight (TOF), from the reference points.

a) Angulation: ArrayTrack [64] is an indoor localization system that exploits the increasing number of antennas at commodity WiFi APs. To have sufficient antennas, hardware modification is needed for ArrayTrack. The system is based on the observation that the phase shift of the received signals between antennas is a function of AOA and thus, utilizes the MUSIC algorithm [94] to estimate the AOA. With three APs, each equipped with 16 antennas, ArrayTrack can localize clients to a median 57 cm and mean 1 m accuracy. With six APs, ArrayTrack achieves a median 23 cm and a mean 31 cm location accuracy, localizing 95% of clients within 90 cm.

Unlike ArrayTrack, SpotFi [65] can achieve accurate indoor localization by the CSI of commodity WiFi devices without any hardware modification. The basic idea is that subcarriers can also be used as sensors when considering the TOF. The system considers the phase shift between subcarriers as a function of TOF to reformulate the steering matrix. By treating both antennas and subcarriers as sensors, SpotFi utilizes the MUSIC algorithm to estimate AOA. By using six off-the-shelf Intel 5300 NICs as APs, which operate at a 5-GHz band with 40-MHz bandwidth, Spotfi achieves median localization error of 0.4 m in an office room.

In general, the number of dominant paths in a typical indoor environment is small and the measured CSI matrix at the antenna array follows the superposition principle. Based on these observations, ROArray [66] casts the AoA estimation into a sparse recovery problem by linearizing the AOA and TOF into a grid. With six Intel 5300 NICs as APs, ROArray achieves 0.91-m median accuracy in the low-SNR scenario, whereas the median accuracies of SpotFi and ArrayTrack degrade to 2.61 and 3.52 m, respectively.

MaTrack [67] is a device-free passive localization approach. Different from the active methods which focus on the direct path between the transmitter and the receiver, the device-free passive methods care about the signal reflected from the target. It is observed that the moving objects of interest cause dynamic multipath whereas the static paths are merged into one on the spectrum due to the signal coherence. Such an observation significantly reduces the number of antennas required to capture the dynamic paths. Moreover, the system leverages the method proposed in [65] to use the 30 subcarriers available on the commodity WiFi to increase the number of effective sensors without any hardware modification. By using Intel 5300 NICs with three antennas as the transceivers, which operate at the 5-GHz frequency with 40-MHz bandwidth, MaTrack achieves a median localization accuracy of 52 and 62 cm when the target is spinning and walking, respectively.

Ubicarse [68] utilizes the idea of synthetic aperture radar (SAR) to design an indoor localization system that enables mobile devices to emulate an antenna array. In other words, the moving device collects signal snapshots as it moves along its trajectory and jointly processes these snapshots to emulate an antenna array to determine the direction of AP. However, with SAR, the position of the moving antenna at each time slot must be known exactly. Fortunately, the mobile device

TABLE X
ANGULATION

System	Algorithm	Median accuracy	Hardware device	Carrier frequency	Bandwidth
ArrayTrack [64]	Music	23cm	WARP&USRP	2.4 GHz	40 MHz
SpotFi [65]	Music	40cm	Intel 5300 NIC	5 GHz	40MHz
ROArray [66]	Sparse recovery	52cm	Intel 5300 NIC	5 GHz	40MHz
MaTrack [67]	Music	60cm	Intel 5300 NIC	5 GHz	40MHz
Ubicarse [68]	Power	39cm	Intel 5300 NIC	5 GHz	\

can provide accurate orientation information by the gyroscope. Since the conjugate product of CSIs on two different antennas is independent of transition and only depends on rotation, by using the relative channel with a conjugate product, Ubicarse is translation resilient and can achieve tens of centimeters of localization accuracy. With a tablet equipped with Intel 5300 NIC and a Yei technology motion sensor mounted on the tablet, Ubicarse achieves a median localization error of 39 cm in LOS scenarios in a library.

Remarks: The comparisons of the five aforementioned indoor localization systems in terms of algorithm, accuracy, and hardware are summarized in Table X. From the table, we can see that existing AOA estimation systems exploit the spatial diversity (phase difference between different sensors) to achieve AOA estimation. Therefore, the main factor that limits the AOA estimation accuracy is the number of sensors. In the near future with the standardization of the IEEE 802.11ad, there will be a larger number of antennas in the massive MIMO systems. Hence, high-resolution AOA estimation can be realized.

b) Lateration: FILA [69] develops a refined indoor propagation model to represent the relationship between the efficient CSI and the propagation distance to estimate the location of a target. The system consists of three steps: 1) data preprocessing; 2) calibration; and 3) location determination. During CSI preprocessing, the paths except the LOS and the NLOS close to the LOS are removed, and the effective CSI is computed to compensate for the fading in the frequency domain. Given a packet with 30 subcarriers, the effective CSI is calculated as the weighted average of CSI amplitude among all subcarriers with the weight being the normalized frequency of each subcarrier. The calibration step derives the CSI-propagation model which indicates the relationship between the distance and the effective CSI. The parameter of the model is determined by a fast training algorithm to obtain a propagation model that is reasonable for the indoor environment. During the localization step, the system applies the simplest trilateration method to locate the target based on the distance. FILA utilizes three fixed APs to localize a laptop and achieves a median accuracy of 0.45 and 1.2 m in laboratory and lecture theater, respectively.

LiFS [70] is a model-based device-free target localization system, which involves multiple APs with known locations as transmitters and multiple mobile clients as receivers. Depending on the location of the target, i.e., LOS, NLOS but in the first Fresnel zone, or outside the first Fresnel zone, the different preprocessing method is used for the CSI measurement. With the power fading model, the localization problem is formulated as a set of over-determined equations, which can be solved by the gradient descent and genetic algorithm hybrid method. With four Intel 5300 NICs as transmitters and seven Intel 5300 NICs as receivers, LiFS achieves a median accuracy of 0.7 m in the home environment.

Chronos [71] enables two WiFi nodes to localize each other without additional infrastructure support. The small bandwidth of the WiFi signal limits the TOF resolution. To resolve this challenge, Chronos [71] collects CSI measurements on multiple WiFi frequency bands by frequency hopping and stitches them together to estimate subnanosecond TOF. Under a single path condition, the accumulated phase shift at each band can be directly used for the TOF estimation, and different bands can be combined based on the Chinese remainder theorem. Under a multipath condition, a nonuniform discrete Fourier transform (NDFT) is utilized to reconstruct the accurate TOF. The experimental results show that Chronos achieves a median TOF error of 0.47 ns in LOS and 0.69 ns in NLOS settings, corresponding to a physical distance accuracy of 14.1 and 20.7 cm, respectively.

Another TOF estimation system based on frequency hopping is ToneTrack [72], which estimates TOF with the MUSIC algorithm by leveraging the fact that the phase shift across subcarriers is a linear function of TOF. Since the bandwidth of the WiFi signal limits the estimation resolution, ToneTrack proposes to combine multiple frequency-agile transmissions from the client to form a virtual wider bandwidth transmission. In an office building, ToneTrack achieves a 90-cm localization accuracy with four APs overhearing just three packets transmitted over three adjacent 20-MHz bandwidth channels.

Remarks: The comparisons of the four aforementioned indoor localization systems in terms of algorithm, accuracy, and hardware are summarized in Table XI. The key component of lateration-based localization methods is the TOF estimation. These methods rely on the monotonicity relationship between the CSI amplitude or phase and the transmitting distance. However, due to the multipath fading and shadowing in an indoor environment, the monotonicity may not hold. To mitigate the multipath fading and shadowing, the signal path should be differentiated from other paths. Unfortunately, the typical bandwidth of 802.11n is 20 MHz, which means that the distance resolution provided by commodity WiFi is 15 m, which is too low to distinguish different paths in the indoor environment. To improve indoor localization performance, frequency hopping approaches are proposed to obtain larger effective bandwidth [71], [72]. The standardization of the IEEE 802.11ad with much wider bandwidth can introduce finer TOF estimation.

TABLE XI
LATERATION

System	Algorithm	Median accuracy	Hardware device	Carrier frequency	Bandwidth
FILA [69]	CSI-based propagation model	0.45m	Intel 5300 NIC	\	20 MHz/40 MHz
LiFS [70]	GD and GA hybrid method	0.7m	Intel 5300 NIC	\	20 MHz/40 MHz
Chronos [71]	Frequency hopping, NDFT	0.141m	Intel 5300 NIC	2.4 GHz and 5 GHz	multi-GHz
ToneTrack [72]	Frequency-agile	0.9m	WARP	2.4 GHz	60MHz

Fingerprinting: There are two different fingerprint-based localization methods: 1) the deterministic and 2) probabilistic methods [95]. The deterministic methods use different similarity metrics such as the Euclidean distance, cosine similarity to compare the measured data with fingerprints and determine the location of the target as the one with the best similarity metric [73]. The probabilistic methods are based on the statistical inference between the observed data and the stored fingerprint and thus can handle the uncertainty in the CSI measurements [95]. The information about signal distribution at reference locations is stored and the probabilistic techniques based on ML, MAP, or MMSE criterion are used to estimate the user location.

c) Deterministic: The CSI-MIMO [73] utilizes both the amplitude and phase information of CSI to construct fingerprint. The CSI is first aggregated over different MIMO streams and then subtracted by the sequential subcarrier to obtain the amplitude and the phase difference as the location fingerprint. CSI-MIMO uses two APs mounted at a fixed location as transmitters, and two mobile laptops equipped with Intel 5300 NIC as receivers. The mobile devices collect data at 19 different locations, where the distance between reference locations ranges from 1.5 to 2 m. By using K-NN to localize a mobile device, CSI-MIMO achieves the accuracy as high as 0.95 m.

Chen *et al.* [74] proposed a time-reversal fingerprint-based indoor localization system which can achieve centimeter accuracy. To achieve such high accuracy, the system exploits frequency diversity via frequency hopping, i.e., concatenating the CSI of different channels together to augment the effective bandwidth. In the offline phase, CSIs are first processed by sanitizing, sifting, and averaging to mitigate the synchronization errors as well as interference from other WiFi networks. The system collects concatenated CSI at locations of interest based on frequency hopping to form location fingerprints. During the online phase, CSI measurements are compared with fingerprints in the database via time-reversal resonating strength (TRRS) to determine the location.

Instead of exploiting frequency diversity as in [74], the method in [75] utilizes the spatial diversity of the MIMO system to achieve high localization accuracy. The system proposes to combine the CSIs over different pairs of antennas to augment the virtual bandwidth. During the offline stage, the system obtains CSIs of multiple antenna links at the reference locations and then combines them into location fingerprints. During the online phase, the system captures CSIs and matches them with the fingerprints based on TRRS to determine the location.

CiFi [79] employs the deep convolution neural network for indoor localization with commodity WiFi devices. Instead of using CSI waveform, CiFi utilizes an AOA image as a fingerprint. For a training location, CiFi collects 960 packets to construct 16 images with size 60×60, where the column and row of the images correspond to the packet index and AOA value, respectively. During the offline phase, the constructed images from all locations are used to train the network. In the online phase, the target location is determined based on the constructed images of newly collected CSI. With a pair of off-the-shelf WiFi devices as transceivers, CiFi achieves an error under 3 m for 87% of locations, and an error under 1 m for 40% of locations.

Remarks: The aforementioned four deterministic fingerprint-based localization systems adopt deterministic methods to match the captured data to fingerprint. Different forms of fingerprint and deterministic methods lead to different localization performance, as summarized in Table XII. Even though these methods achieve high accuracy, generating fingerprint is an elaborated process.

d) Probabilistic: Pinloc [76] is a probabilistic localization system based on the CSI information. The system is based on two observations: 1) the CSIs at each location are not entirely random but exhibit a defined structure across time and 2) the CSI structure of a given location is different from the structures of all other locations. At the training phase, the system models the data collected at each reference point as a Gaussian distribution and estimates its parameters. During the online phase, the system uses log likelihood as a distance metric to express how likely the observed CSI belongs to a particular cluster from the training database. With 50 spots of 1 m $\times$ 1 m in size, PinLoc can localize users to a spot with 90% mean accuracy by utilizing off-the-shelf WiFi devices.

In [69], the total power of the CSI over 30 subcarriers corresponding to each AP is computed and the distribution is derived. During the online stage, the system uses a probabilistic approach to determine the location. Specifically, given a signal power vector of K APs, the localization problem is equivalent to find the position maximizing the posterior probability, which can be calculated by the prior probability and the conditional probability. The prior probability can be computed based on the similarity between the observed data and the stored CSI data. The conditional probability can be computed

TABLE XII
DETERMINISTIC FINGERPRINTING

System	Fingerprint	Deterministic method	Accuracy	Hardware device	Carrier frequency	Bandwidth
CSI-MIMO [73]	Phase and amplitude difference of aggregated CSI	K-NN	95cm	Intel 5300 NIC	\	20 MHz /40 MHz
[74]	Concatenated CFRs of different channels	TRRS	1-2cm	USRP	5.4 GHz	1GHz
[75]	Combined CFRs of different Links	TRRS	1-2cm	WiFi NIC	5.24 GHz	178 MHz
CiFi [79]	AOA images	DCNN	1m	Intel 5300 NIC	5 GHz	20 MHz /40 MHz

TABLE XIII
PROBABILISTIC FINGERPRINTING

System	Collected Data	Distribution	Probabilistic Criterion	Accuracy	Hardware device	Carrier frequency	Bandwidth
Pinloc [76]	CSI of a cluster	Gaussian	Log-likelihood	90% within 1m	Intel 5300 NIC	\	20 MHz /40 MHz
[69]	Power of a packet	Gaussian	Maximum posterior probability	Median 0.65m	Intel 5300 NIC	2.4 GHz	20MHz
MonoPHY [77]	CSI amplitude of a link	Gaussian	Maximum posterior probability	Median 1.36m	Intel 5300 NIC	2.4 GHz	20MHz
PhaseFi [78]	Weights of the deep network	Radial basis function	Maximum posterior probability	Median 1.08m	Intel 5300 NIC	2.4 GHz	20 MHz /40 MHz
DS-3DCNN [80]	3D CSI-MIMO matrices	Output of 3DCNN	MMSE	Mean error 0.984m	Intel 5300 NIC	2.4 GHz	40MHz

based on the derived distribution of the total power of the CSI. To further enhance the localization accuracy, the final estimation is the weighted average over reference points with the posterior probability as weight. By utilizing six off-the-shelf WiFi devices as APs, the system can achieve the median accuracy of 0.65 m in the laboratory and mean accuracy of 1.07 m in the corridor, respectively.

MonoPHY [77] constructs the fingerprint without combining all subcarriers. The system observes that the CSI amplitude on each subcarrier over different packets can be mapped into no more than three clusters with Gaussian distributions. The fingerprint at each reference point is the mean and variance of the Gaussian distribution of each cluster. The online phase uses the probabilistic approach to estimate the location by maximizing the posterior probability. MonoPHY is evaluated in a typical apartment with an area of approximately 100 m^2. The fingerprint is constructed for 35 different locations, uniformly distributed over the area. With a pair of commercial WiFi devices, MonoPHY achieves 1.36-m median distance error.

PhaseFi [78] utilizes the CSI phase information for localization. To remove the phase errors due to the CFO and SFO, a linear transform is applied to calibrate the phase information. In the offline training stage, PhaseFi employs a deep network with three hidden layers to train the calibrated phase to maximize the marginal distribution of the calibrated phase. This approach is different from the traditional methods that directly store the measurement data as fingerprints, which may be influenced by the complex indoor propagation environment. Besides, a large number of weights in the deep network are used as feature-based fingerprints, which effectively represent the characteristics of the calibrated phases for each position. In the online stage, a probabilistic method based on the radial basis function is used to estimate the target location, where the underlying idea is to measure the similarity between the calibrated phase data and calibrated phase reconstructed by the deep network. PhaseFi achieves a mean location error of 1.08 m and a standard deviation of 0.4046 m in a 4 m $\times$ 7 m living room with 32 reference locations.

Jing *et al.* [80] proposed a dual-stream 3-D CNN (DS-3DCNN) for indoor localization. The DS-3DCNN consists of two parallel subnetworks with several convolutions, batch normalization (BN) [32], max-pooling, and fully connected layers. During the offline phase, the amplitude and phase are input into DC-3DCNN to train the network. In the online phase, with the measured CSI, the network outputs the probabilities of locations, with which the final localization is estimated based on the MMSE criterion. By evaluating in a university laboratory with 26 reference points and a long corridor with 16 reference points, the system achieves mean error of localization 0.984 and 1.909 m, respectively.

Remarks: The detailed comparisons among the aforementioned probabilistic localization approaches are summarized in Table XIII, from which we can see that with different features and criterion, different localization performances are achieved. To achieve high localization accuracy, these systems need a comprehensive site survey to build detailed fingerprints, and they require updating fingerprints once the environment changes. These two requirements make them impractical for real indoor environment deployment. Crowdsourcing individual user's measurements to build up the fingerprint database may be one possible solution.

V. CONCLUSION

In this article, a survey of the state-of-the-art indoor applications using the CSI information of the commodity WiFi devices has been conducted. We discussed the distortions of

the CSI obtained from the commodity WiFi devices and summarized the preprocessing steps for the CSI sanitization. We reviewed recent advances in nine key applications of WiFi vision problems, including WiFi imaging, vital sign monitoring, human identification, gesture recognition, gait recognition, daily activity recognition, fall detection, human detection, and indoor positioning. All these works demonstrated that WiFi is a desirable solution for building a smart environment. There are still several challenges that need to be addressed in the future work, including how to obtain calibrated phase information of CSI, how to realize real time and ubiquitous motion segmentation, and how to build a theoretical model to analyze the relationship between the activity and the CSI measurement.

REFERENCES

[1] M. Cornacchia, K. Ozcan, Y. Zheng, and S. Velipasalar, "A survey on activity detection and classification using wearable sensors," *IEEE Sens. J.*, vol. 17, no. 2, pp. 386–403, Jan. 2017.

[2] N. Hegde *et al.*, "The pediatric smartshoe: Wearable sensor system for ambulatory monitoring of physical activity and gait," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 26, no. 2, pp. 477–486, Feb. 2018.

[3] O. Politi, I. Mporas, and V. Megalooikonomou, "Human motion detection in daily activity tasks using wearable sensors," in *Proc. IEEE 22nd Eur. Signal Process. Conf. (EUSIPCO)*, 2014, pp. 2315–2319.

[4] J. L. M. Flores, S. S. Srikant, B. Sareen, and A. Vagga, "Performance of RFID tags in near and far field," in *Proc. IEEE Int. Conf. Pers. Wireless Commun. (ICPWC)*, 2005, pp. 353–357.

[5] J. R. Smith *et al.*, "RFID-based techniques for human-activity detection," *Commun. ACM*, vol. 48, no. 9, pp. 39–44, 2005.

[6] L. Yang, Y. Chen, X.-Y. Li, C. Xiao, M. Li, and Y. Liu, "Tagoram: Real-time tracking of mobile RFID tags to high precision using cots devices," in *Proc. ACM 20th Annu. Int. Conf. Mobile Comput. Netw.*, 2014, pp. 237–248.

[7] J. Wang, D. Vasisht, and D. Katabi, "RF-IDraw: Virtual touch screen in the air using RF signals," in *Proc. ACM SIGCOMM Comput. Commun. Rev.*, vol. 44, 2014, pp. 235–246.

[8] L. M. Ni, Y. Liu, Y. C. Lau, and A. P. Patil, "LANDMARC: Indoor location sensing using active RFID," *Wireless Netw.*, vol. 10, no. 6, pp. 701–710, 2004.

[9] F. Naya, H. Noma, R. Ohmura, and K. Kogure, "Bluetooth-based indoor proximity sensing for nursing context awareness," in *Proc. 9th IEEE Int. Symp. Wearable Comput.*, 2005, pp. 212–213.

[10] H. J. P. Iglesias, V. Barral, and C. J. Escudero, "Indoor person localization system through RSSI Bluetooth fingerprinting," in *Proc. IEEE 19th Int. Conf. Syst. Signals Image Process. (IWSSIP)*, 2012, pp. 40–43.

[11] D. Dardari, A. Conti, U. Ferner, A. Giorgetti, and M. Z. Win, "Ranging with ultrawide bandwidth signals in multipath environments," *Proc. IEEE*, vol. 97, no. 2, pp. 404–426, Feb. 2009.

[12] H. Liu, H. Darabi, P. Banerjee, and J. Liu, "Survey of wireless indoor positioning techniques and systems," *IEEE Trans. Syst., Man, Cybern. C, Appl. Rev.*, vol. 37, no. 6, pp. 1067–1080, Nov. 2007.

[13] Z. Farid, R. Nordin, and M. Ismail, "Recent advances in wireless indoor localization techniques and system," *J. Comput. Netw. Commun.*, vol. 2013, Sep. 2013, Art. no. 185138.

[14] F. Adib, Z. Kabelac, D. Katabi, and R. C. Miller, "3D tracking via body radio reflections," in *Proc. NSDI*, vol. 14, 2014, pp. 317–329.

[15] Q. Pu, S. Gupta, S. Gollakota, and S. Patel, "Whole-home gesture recognition using wireless signals," in *Proc. ACM 19th Annu. Int. Conf. Mobile Comput. Netw.*, 2013, pp. 27–38.

[16] S. Yousefi, H. Narui, S. Dayal, S. Ermon, and S. Valaee, "A survey on behavior recognition using WiFi channel state information," *IEEE Commun. Mag.*, vol. 55, no. 10, pp. 98–104, Oct. 2017.

[17] Y. Zou, W. Liu, K. Wu, and L. M. Ni, "Wi-Fi radar: Recognizing human behavior with commodity Wi-Fi," *IEEE Commun. Mag.*, vol. 55, no. 10, pp. 105–111, Oct. 2017.

[18] H. Jiang, C. Cai, X. Ma, Y. Yang, and J. Liu, "Smart home based on WiFi sensing: A survey," *IEEE Access*, vol. 6, pp. 13317–13325, 2018.

[19] Y. Wang, J. Liu, Y. Chen, M. Gruteser, J. Yang, and H. Liu, "E-eyes: Device-free location-oriented activity identification using fine-grained WiFi signatures," in *Proc. 20th Annu. Int. Conf. Mobile Comput. Netw.*, 2014, pp. 617–628.

[20] W. Wang, A. X. Liu, M. Shahzad, K. Ling, and S. Lu, "Device-free human activity recognition using commercial WiFi devices," *IEEE J. Sel. Areas Commun.*, vol. 35, no. 5, pp. 1118–1131, May 2017.

[21] G. Wang, Y. Zou, Z. Zhou, K. Wu, and L. M Ni, "We can hear you with Wi-Fi!" *IEEE Trans. Mobile Comput.*, vol. 15, no. 11, pp. 2907–2920, Nov. 2016.

[22] K. Ali, A. X. Liu, W. Wang, and M. Shahzad, "Recognizing keystrokes using WiFi devices," *IEEE J. Sel. Areas Commun.*, vol. 35, no. 5, pp. 1175–1190, May 2017.

[23] F. Adib and D. Katabi, "See through walls with WiFi!" in *Proc. ACM SIGCOMM Conf. SIGCOMM*, vol. 43, 2013, pp. 75–86.

[24] D. Huang, R. Nandakumar, and S. Gollakota, "Feasibility and limits of Wi-Fi imaging," in *Proc. 12th ACM Conf. Embedded Netw. Sensor Syst.*, 2014, pp. 266–279.

[25] K. N. Poudel, D. Schurig, and N. Patwari, "Spatial imaging using a communication system's channel state information," in *Proc. USNC-URSI Radio Sci. Meeting*, 2016, pp. 41–42.

[26] P. M. Holl and F. Reinhard, "Holography of Wi-Fi radiation," *Phys. Rev. Lett.*, vol. 118, May 2017, Art. no. 183901.

[27] S. Vakalis, L. Gong, and J. A. Nanzer, "Imaging with WiFi," *IEEE Access*, vol. 7, pp. 28616–28624, 2019.

[28] X. Liu, J. Cao, S. Tang, and J. Wen, "Wi-Sleep: Contactless sleep monitoring via WiFi signals," in *Proc. Real Time Syst. Symp. (RTSS)*, 2014, pp. 346–355.

[29] X. Liu, J. Cao, S. Tang, J. Wen, and P. Guo, "Contactless respiration monitoring via off-the-shelf WiFi devices," *IEEE Trans. Mobile Comput.*, vol. 15, no. 10, pp. 2466–2479, Oct. 2016.

[30] J. Liu, Y. Wang, Y. Chen, J. Yang, X. Chen, and J. Cheng, "Tracking vital signs during sleep leveraging off-the-shelf WiFi," in *Proc. 16th ACM Int. Symp. Mobile Ad Hoc Netw. Comput.*, 2015, pp. 267–276.

[31] X. Wang, C. Yang, and S. Mao, "PhaseBeat: Exploiting CSI phase data for vital sign monitoring with commodity WiFi devices," in *Proc. IEEE 37th Int. Conf. Distrib. Comput. Syst. (ICDCS)*, 2017, pp. 1230–1239.

[32] C. Chen *et al.*, "TR-BREATH: Time-reversal breathing rate estimation and detection," *IEEE Trans. Biomed. Eng.*, vol. 65, no. 3, pp. 489–501, Mar. 2018.

[33] Y. Zeng, D. Wu, J. Xiong, E. Yi, R. Gao, and D. Zhang, "FarSense: Pushing the range limit of WiFi-based respiration sensing with CSI ratio of two antennas," *Proc. ACM Interact. Mobile Wearable Ubiquitous Technol.*, vol. 3, no. 3, pp. 1–26, 2019.

[34] Y. Zeng, D. Wu, R. Gao, T. Gu, and D. Zhang, "FullBreathe: Full human respiration detection exploiting complementarity of CSI phase and amplitude of WiFi signals," *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.*, vol. 2, no. 3, p. 148, Sep. 2018.

[35] T. Xin, B. Guo, Z. Wang, M. Li, Z. Yu, and X. Zhou, "FreeSense: Indoor human identification with Wi-Fi signals," in *Proc. IEEE Global Commun. Conf. (GLOBECOM)*, 2016, pp. 1–7.

[36] F. Hong, X. Wang, Y. Yang, Y. Zong, Y. Zhang, and Z. Guo, "WFID: Passive device-free human identification using WiFi signal," in *Proc. 13th Int. Conf. Mobile Ubiquitous Syst. Comput. Netw. Services*, 2016, pp. 47–56.

[37] Y. Zeng, P. H. Pathak, and P. Mohapatra, "WiWHO: WiFi-based person identification in smart spaces," in *Proc. 15th Int. Conf. Inf. Process. Sensor Netw.*, 2016, p. 4.

[38] Q. Xu, Y. Chen, B. B. Wang, and K. J. R. Liu, "Radio biometrics: Human recognition through a wall," *IEEE Trans. Inf. Forensics Security*, vol. 12, no. 5, pp. 1141–1155, May 2017.

[39] W. He, K. Wu, Y. Zou, and Z. Ming, "WIG: WiFi-based gesture recognition system," in *Proc. 24th Int. Conf. Comput. Commun. Netw. (ICCCN)*, 2015, pp. 1–7.

[40] H. Li, W. Yang, J. Wang, Y. Xu, and L. Huang, "WiFinger: Talk to your smart devices with finger-grained gesture," in *Proc. 2016 ACM Int. Joint Conf. Pervasive Ubiquitous Comput.*, 2016, pp. 250–261.

[41] Q. Zhou, J. Xing, J. Li, and Q. Yang, "A device-free number gesture recognition approach based on deep learning," in *Proc. IEEE 12th Int. Conf. Comput. Intell. Security (CIS)*, 2016, pp. 57–63.

[42] J. Shang and J. Wu, "A robust sign language recognition system with sparsely labeled instances using Wi-Fi signals," in *Proc. IEEE 14th Int. Conf. Mobile Ad Hoc Sensor Syst. (MASS)*, 2017, pp. 99–107.

[43] A. Virmani and M. Shahzad, "Position and orientation agnostic gesture recognition using WiFi," in *Proc. 15th Annu. Int. Conf. Mobile Syst. Appl. Services*, 2017, pp. 252–264.

[44] J. Yang, H. Zou, Y. Zhou, and L. Xie, "Learning gestures from WiFi: A Siamese recurrent convolutional architecture," *IEEE Internet Things J.*, vol. 6, no. 6, pp. 10763–10772, Dec. 2019.

[45] S. Maheshwari and A. K. Tiwari, "Walking parameters estimation through channel state information preliminary results," in *Proc. IEEE 9th Int. Conf. Signal Process. Commun. Syst. (ICSPCS)*, 2015, pp. 1–8.

[46] W. Wang, A. X. Liu, and M. Shahzad, "Gait recognition using WiFi signals," in *Proc. ACM Int. Joint Conf. Pervasive Ubiquitous Comput.*, 2016, pp. 363–373.

[47] Y. Zeng, P. H. Pathak, C. Xu, and P. Mohapatra, "Your AP knows how you move: Fine-grained device motion recognition through WiFi," in *Proc. 1st ACM Workshop Hot Topics wireless*, 2014, pp. 49–54.

[48] S. Arshad *et al.*, "Wi-Chase: A WiFi based human activity recognition system for sensorless environments," in *Proc. IEEE 18th Int. Symp. World Wireless Mobile Multimedia Netw. (WoWMoM)*, 2017, pp. 1–6.

[49] Q. Gao, J. Wang, X. Ma, X. Feng, and H. Wang, "CSI-based device-free wireless localization and activity recognition using radio image features," *IEEE Trans. Veh. Technol.*, vol. 66, no. 11, pp. 10346–10356, Nov. 2017.

[50] X. Zheng, J. Wang, L. Shangguan, Z. Zhou, and Y. Liu, "SMOKEY: Ubiquitous smoking detection with commercial WiFi infrastructures," in *Proc. 35th Annu. IEEE Int. Conf. Comput. Commun. (INFOCOM)*, 2016, pp. 1–9.

[51] Y. Gu, X. Zhang, Z. Liu, and F. Ren, "BeSense: Leveraging WiFi channel data and computational intelligence for behavior analysis," *IEEE Comput. Intell. Mag.*, vol. 14, no. 4, pp. 31–41, Nov. 2019.

[52] F. Wang, W. Gong, and J. Liu, "On spatial diversity in WiFi-based human activity recognition: A deep learning-based approach," *IEEE Internet Things J.*, vol. 6, no. 2, pp. 2035–2047, Apr. 2019.

[53] Y. Wang, K. Wu, and L. M. Ni, "WiFall: Device-free fall detection by wireless networks," *IEEE Trans. Mobile Comput.*, vol. 16, no. 2, pp. 581–594, Feb. 2017.

[54] D. Zhang, H. Wang, Y. Wang, and J. Ma, "ANTI-fall: A non-intrusive and real-time fall detector leveraging CSI from commodity WiFi devices," in *Proc. Int. Conf. Smart Homes Health Telematics*, 2015, pp. 181–193.

[55] H. Wang, D. Zhang, Y. Wang, J. Ma, Y. Wang, and S. Li, "RT-Fall: A real-time and contactless fall detection system with commodity WiFi devices," *IEEE Trans. Mobile Comput.*, vol. 16, no. 2, pp. 511–526, Feb. 2017.

[56] L. Zhang, Z. Wang, and L. Yang, "Commercial Wi-Fi based fall detection with environment influence mitigation," in *Proc. 16th Annu. IEEE Int. Conf. Sens. Commun. Netw. (SECON)*, 2019, pp. 1–9.

[57] J. Xiao, K. Wu, Y. Yi, L. Wang, and L. M. Ni, "FIMD: Fine-grained device-free motion detection," in *Proc. IEEE 18th Int. Conf. Parallel Distrib. Syst. (ICPADS)*, 2012, pp. 229–235.

[58] K. Qian, C. Wu, Z. Yang, Y. Liu, and Z. Zhou, "PADS: Passive detection of moving targets with dynamic speed using PHY layer information," in *Proc. 20th IEEE Int. Conf. Parallel Distrib. Syst. (ICPADS)*, 2014, pp. 1–8.

[59] C. Wu, Z. Yang, Z. Zhou, X. Liu, Y. Liu, and J. Cao, "Non-invasive detection of moving and stationary human with WiFi," *IEEE J. Sel. Areas Commun.*, vol. 33, no. 11, pp. 2329–2342, Nov. 2015.

[60] J. Xiao, K. Wu, Y. Yi, L. Wang, and L. M. Ni, "PILOT: Passive device-free indoor localization using channel state information," in *Proc. IEEE 33rd Int. Conf. Distrib. Comput. Syst. (ICDCS)*, 2013, pp. 236–245.

[61] L. Gong *et al.*, "An adaptive wireless passive human detection via fine-grained physical layer information," *Ad Hoc Netw.*, vol. 38, pp. 38–50, Mar. 2016.

[62] Z. Zhou, Z. Yang, C. Wu, L. Shangguan, and Y. Liu, "Towards omni-directional passive human detection," in *Proc. IEEE INFOCOM*, 2013, pp. 3057–3065.

[63] Z. Zhou, Z. Yang, C. Wu, L. Shangguan, and Y. Liu, "Omnidirectional coverage for device-free passive human detection," *IEEE Trans. Parallel Distrib. Syst.*, vol. 25, no. 7, pp. 1819–1829, Jul. 2014.

[64] J. Xiong and K. Jamieson, *ArrayTrack: A Fine-Grained Indoor Location System*, USENIX, Berkeley, CA, USA, 2013.

[65] M. Kotaru, K. Joshi, D. Bharadia, and S. Katti, "SpotFI: Decimeter level localization using WiFi," in *Proc. ACM SIGCOMM Comput. Commun. Rev.*, vol. 45, 2015, pp. 269–282.

[66] W. Gong and J. Liu, "Robust indoor wireless localization using sparse recovery," in *Proc. IEEE 37th Int. Conf. Distrib. Comput. Syst. (ICDCS)*, 2017, pp. 847–856.

[67] X. Li, S. Li, D. Zhang, J. Xiong, Y. Wang, and H. Mei, "Dynamic-MUSIC: Accurate device-free indoor localization," in *Proc. ACM Int. Joint Conf. Pervasive Ubiquitous Comput.*, 2016, pp. 196–207.

[68] S. Kumar, S. Gil, D. Katabi, and D. Rus, "Accurate indoor localization with zero start-up cost," in *Proc. 20th Annu. Int. Conf. Mobile Comput. Netw.*, 2014, pp. 483–494.

[69] K. Wu, J. Xiao, Y. Yi, D. Chen, X. Luo, and L. M. Ni, "CSI-based indoor localization," *IEEE Trans. Parallel Distrib. Syst.*, vol. 24, no. 7, pp. 1300–1309, Jul. 2013.

[70] J. Wang *et al.*, "LIFS: Low human-effort, device-free localization with fine-grained subcarrier information," in *Proc. ACM 22nd Annu. Int. Conf. Mobile Comput. Netw.*, 2016, pp. 243–256.

[71] D. Vasisht, S. Kumar, and D. Katabi, "Decimeter-level localization with a single WiFi access point," in *Proc. NSDI*, vol. 16, 2016, pp. 165–178.

[72] J. Xiong, K. Sundaresan, and K. Jamieson, "ToneTrack: Leveraging frequency-agile radios for time-based indoor wireless localization," in *Proc. 21st Annu. Int. Conf. Mobile Comput. Netw.*, 2015, pp. 537–549.

[73] Y. Chapre, A. Ignjatovic, A. Seneviratne, and S. Jha, "CSI-MIMO: Indoor Wi-Fi fingerprinting system," in *Proc. IEEE 39th Conf. Local Comput. Netw. (LCN)*, 2014, pp. 202–209.

[74] C. Chen, Y. Chen, Y. Han, H. Lai, and K. J. R. Liu, "Achieving centimeter-accuracy indoor localization on WiFi platforms: A frequency hopping approach," *IEEE Internet Things J.*, vol. 4, no. 1, pp. 111–121, Feb. 2017.

[75] C. Chen, Y. Chen, Y. Han, H. Lai, F. Zhang, and K. J. R. Liu, "Achieving centimeter-accuracy indoor localization on WiFi platforms: A multi-antenna approach," *IEEE Internet Things J.*, vol. 4, no. 1, pp. 122–134, Feb. 2017.

[76] S. Sen, R. R. Choudhury, B. Radunovic, and T. Minka, "Precise indoor localization using PHY layer information," in *Proc. 10th ACM Workshop Hot Topics Netw.*, 2011, p. 18.

[77] H. Abdel-Nasser, R. Samir, I. Sabek, and M. Y. Monophy, "Mono-stream-based device-free WLAN localization via physical layer information," in *Proc. Wireless Commun. Netw. Conf. (WCNC)*, 2013, pp. 4546–4551.

[78] X. Wang, L. Gao, and S. Mao, "PhaseFI: Phase fingerprinting for indoor localization with a deep learning approach," in *Proc. IEEE Global Commun. Conf. (GLOBECOM)*, 2015, pp. 1–6.

[79] X. Wang, X. Wang, and S. Mao, "Deep convolutional neural networks for indoor localization with CSI images," *IEEE Trans. Netw. Sci. Eng.*, vol. 7, no. 1, pp. 316–327, Jan. 2020.

[80] Y. Jing, J. Hao, and P. Li, "Learning spatiotemporal features of CSI for indoor localization with dual-stream 3D convolutional neural networks," *IEEE Access*, vol. 7, pp. 147571–147585, 2019.

[81] D. Halperin, W. Hu, A. Sheth, and D. Wetherall, "Tool release: Gathering 802.11n traces with channel state information," *ACM SIGCOMM Comput. Commun. Rev.*, vol. 41, no. 1, p. 53, 2011.

[82] E. H. Ong, J. Kneckt, O. Alanen, Z. Chang, T. Huovinen, and T. Nihtilä, "IEEE 802.11ac: Enhancements for very high throughput WLANs," in *Proc. IEEE 22nd Int. Symp. Pers. Indoor Mobile Radio Commun.*, Sep. 2011, pp. 849–853.

[83] Z. Yang, Z. Zhou, and Y. Liu, "From RSSI to CSI: Indoor localization via channel response," *ACM Comput. Surveys*, vol. 46, no. 2, p. 25, 2013.

[84] S. Maheshwari and A. K. Tiwari, "Ubiquitous fall detection through wireless channel state information," in *Proc. Int. Conf. Comput. Netw. Commun. (CoCoNet)*, 2015, pp. 1–9.

[85] D. Zhang, Y. Hu, Y. Chen, and B. Zeng, "Calibrating phase offsets for commodity WiFi," *IEEE Syst. J.*, vol. 14, no. 1, pp. 661–664, Mar. 2020.

[86] D. Zhang, Y. Hu, Y. Chen, and B. Zeng, "BreathTrack: Tracking indoor human breath status via commodity WiFi," *IEEE Internet Things J.*, vol. 6, no. 2, pp. 3899–3911, Apr. 2019.

[87] F. Adib, H. Mao, Z. Kabelac, D. Katabi, and R. C. Miller, "Smart homes that monitor breathing and heart rate," in *Proc. 33rd Annu. ACM Conf. Human Factors Comput. Syst.*, 2015, pp. 837–846.

[88] B. D. Rao and K. V. S. Hari, "Performance analysis of Root-Music," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 37, no. 12, pp. 1939–1949, Dec. 1989.

[89] H. Wang *et al.*, "Human respiration detection with commodity WiFi devices: Do user location and body orientation matter?" in *Proc. ACM Int. Joint Conf. Pervasive Ubiquitous Comput.*, 2016, pp. 25–26.

[90] M. M. Breunig, H.-P. Kriegel, R. T. Ng, and J. Sander, "LOF: Identifying density-based local outliers," in *Proc. ACM SIGMOD Rec.*, vol. 29, 2000, pp. 93–104.

[91] W. Wang, A. X. Liu, M. Shahzad, K. Ling, and S. Lu, "Understanding and modeling of WiFi signal based human activity recognition," in *Proc. 21st Annu. Int. Conf. Mobile Comput. Netw.*, 2015, pp. 65–76.

[92] D. Cai, C. Zhang, and X. He, "Unsupervised feature selection for multi-cluster data," in *Proc. 16th ACM SIGKDD Int. Conf. Knowl. Disc. Data Min.*, 2010, pp. 333–342.

[93] M. Ester, H.-P. Kriegel, J. Sander, and X. Xu, "A density-based algorithm for discovering clusters in large spatial databases with noise," in *Proc. KDD*, vol. 96, 1996, pp. 226–231.

[94] R. Schmidt, "Multiple emitter location and signal parameter estimation," *IEEE Trans. Antennas Propag.*, vol. AP-34, no. 3, pp. 276–280, Mar. 1986.
[95] S. He and S.-H. G. Chan, "Wi-Fi fingerprint-based indoor positioning: Recent advances and comparisons," *IEEE Commun. Surveys Tuts.*, vol. 18, no. 1, pp. 466–490, 1st Quart., 2016.

Ying He (Graduate Student Member, IEEE) received the B.S. degree from the School of Electronic and Engineering, University of Electronic Science and Technology of China, Chengdu, China, in 2017, where she is currently pursuing the Ph.D. degree with the School of Information and Communication Engineering.

Her research interests include wireless communications and signal processing.

Yan Chen (Senior Member, IEEE) received the bachelor's degree from the University of Science and Technology of China, Hefei, China, in 2004, the M.Phil. degree from Hong Kong University of Science and Technology, Hong Kong, in 2007, and the Ph.D. degree from the University of Maryland at College Park, College Park, MD, USA, in 2011.

He was with Origin Wireless Inc., Greenbelt, MD, USA, as a Founding Principal Technologist. From September 2015 to February 2020, he was a Full Professor with the School of Information and Communication Engineering, University of Electronic Science and Technology of China, Chengdu, China. He is currently a Full Professor with the School of Cyberspace Security, University of Science and Technology of China. His research interests include multimedia, signal processing, game theory, and wireless communications.

Prof. Chen was a recipient of multiple honors and awards, including the Best Student Paper Award at PCM in 2017, the Best Student Paper Award at IEEE ICASSP in 2016, the Best Paper Award at IEEE GLOBECOM in 2013, the Future Faculty Fellowship and Distinguished Dissertation Fellowship Honorable Mention from the Department of Electrical and Computer Engineering in 2010 and 2011, the Finalist of the Dean's Doctoral Research Award from the A. James Clark School of Engineering, the University of Maryland in 2011, and the Chinese Government Award for outstanding students abroad in 2010.

Yang Hu received the B.S. and Ph.D. degrees in electrical engineering from the University of Science and Technology of China, Hefei, China, in 2004 and 2009, respectively.

She was with the University of Maryland Institute for Advanced Computer Studies, College Park, MD, USA, as a Research Associate from 2010 to 2015. She is currently an Associate Professor with the School of Information and Communication Engineering, University of Electronic Science and Technology of China, Chengdu, China. Her current research interests include computer vision, machine learning, and multimedia signal processing.

Bing Zeng (Fellow, IEEE) received the B.Eng. and M.Eng. degrees in electronic engineering from the University of Electronic Science and Technology of China (UESTC), Chengdu, China, in 1983 and 1986, respectively, and the Ph.D. degree in electrical engineering from Tampere University of Technology, Tampere, Finland, in 1991.

He worked as a Postdoctoral Fellow with the University of Toronto, Toronto, ON, Canada, from September 1991 to July 1992 and as a Researcher with Concordia University, Montreal, QC, Canada, from August 1992 to January 1993. He then joined Hong Kong University of Science and Technology (HKUST), Hong Kong. After 20 years of service with HKUST, he returned to UESTC in the summer of 2013, through the China 1000-Talent-Scheme. At UESTC, he leads the Institute of Image Processing to work on image and video processing, 3-D and multiview video technology, and visual big data. During his tenure with HKUST and UESTC, he graduated more than 30 Master and Ph.D. students, received about 20 research grants, filed eight international patents, and published more than 260 papers. Three representing works are as follows: one paper on fast block motion estimation, published in IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS FOR VIDEO TECHNOLOGY in 1994, has so far been SCI-cited more than 1000 times (Google cited more than 2200 times) and currently stands at the 8th position among all papers published in this transactions; one paper on smart padding for arbitrarily shaped image blocks, published in IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS FOR VIDEO TECHNOLOGY in 2001, leads to a patent that has been successfully licensed to companies; and one paper on directional discrete cosine transform, published in IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS FOR VIDEO TECHNOLOGY in 2008, receives the 2011 IEEE CSVT Transactions Best Paper Award.

Dr. Zeng received the Best Paper Award at ChinaCom three times (2009 Xi'an, 2010 Beijing, and 2012 Kunming). He received the 2nd Class Natural Science Award (the first recipient) from the Chinese Ministry of Education in 2014 and was elected as an IEEE Fellow in 2016 for contributions to image and video coding. He served as an Associate Editor for the IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS FOR VIDEO TECHNOLOGY for eight years and received the Best Associate Editor Award in 2011. He was the General Co-Chair of VCIP-2016 and PCM-2017.