

Received April 5, 2020, accepted April 20, 2020, date of publication April 29, 2020, date of current version May 15, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.2991201

A Survey of Handy See-Through Wall Technology

KANGLE MU¹, TOM H. LUAN², (Senior Member, IEEE), LINA ZHU³,
LIN X. CAI⁴, AND LONGXIANG GAO⁵, (Senior Member, IEEE)

¹School of Telecommunications Engineering, Xidian University, Shaanxi 710126, China

²School of Cyber Engineering, Xidian University, Shaanxi 710126, China

³School of Telecommunications Engineering, Xidian University, Shaanxi 710071, China

⁴Department of Electrical and Computer Engineering, Illinois Institute of Technology, Chicago, IL 60616, USA

⁵School of Information Technology, Deakin University, Geelong, VIC 3125, Australia

Corresponding author: Tom H. Luan (mailto:tom.luan@xidian.edu.cn)

This work was supported by the China NSF under Grant U1808207.

ABSTRACT The through-wall system, which applies radio technologies to detect objects behind the wall, can find many appealing applications such as public security, life detection, and medical health monitoring. While being studied for years, the recent advances in high-performance handheld computing devices and artificial intelligence have made the through-wall system more practical. In this article, we present a tutorial-like study on the fundamental radio technologies used in the through-wall system, as well as its recent advances. Different from the traditional through-wall radars, this paper mainly focuses on the handy through-wall techniques with low power, narrow bandwidth, lightweight, no contact, and civil use. Advanced through-wall systems and open research issues are also presented.

INDEX TERMS Seeing through walls, machine learning, motion capture, wireless.

I. INTRODUCTION

People always have a great interest in knowing things behind walls, which makes the through-wall technology demanding. The through-wall technology refers to the technology which has the ability to detect objects of interest inside an enclosed area, and does not require users to wear any additional devices.

Through-wall technology can find the following typical applications:

A. PUBLIC SECURITY

Law enforcement personnel can use this technology to detect how many individuals are behind the wall and locate them as well. For example, in a hostage situation, policemen can use the through-wall technology to identify the distribution of terrorists hiding in rooms and make a more effective rescue plan of hostages.

B. LIFE DETECTION

Through-wall technology provides an efficient way for life-search and detection in disaster rescue. For instance, after an earthquake, the through-wall system, installed on

The associate editor coordinating the review of this manuscript and approving it for publication was Prakasam Periasamy¹.

unmanned aerial vehicles, can be used to locate survivors trapped under the rubble.

C. SMART HOME MONITORING

A house equipped with through-wall devices is able to monitor the movement of people, e.g., the elderly and children [1], inside the house and analyze their behaviors. For instance, it would be very dangerous if the elderly fall down, or children get to dangerous positions such as windows or balconies. With the real-time continuous motion detection achieved by the through-wall technology, many dangerous events at home can be alerted and prevented.

D. MEDICAL APPLICATIONS

Through-wall technology can also be applied to measure vital signs like breathing and heart rates of human in medical health care. Typical technology for vital sign measurement requires body contact; users need to put on extra devices on their bodies which can be inconvenient and uncomfortable. The through-wall devices can track the vital signals without the skin contacts and even through obstacles. The devices can discover the slight body movement caused by breathing or heartbeat, and measure the breathing and heart rates accurately by extracting the periodicity of the received signal [2].

TABLE 1. Structure of this survey.

Section	Title	Content
II	Challenges	Signal Attenuation Signal Measurement Multipath Effect Mirror Effect
III	Categories of Through-Wall Systems/Works	WiFi-based System Radio Tomographic Imaging System (RTI) Traditional Through-Wall Radar Software-Defined Radio System (SDR)
IV	Basic Principles	Principle of Electromagnetic Wave Principle of Localization
V	Basic Methods	Distance Measurement Direction of Arrival Measurement Machine Learning
VI	Examples of Handy Through-wall Systems	Wi-Vi, WiTrack, RF-Capture, etc.

The through-wall technology mainly uses the radio frequency (RF) signal since RF can penetrate non-metallic obstacles like wooden walls and reflect off human bodies. Through-wall systems transmit electromagnetic signals and then process the received reflected signal to exploit information behind walls. In order to have a better penetrability, most proposed through-wall systems operate at the frequencies below 10GHz, which is because that wall attenuations of RF signals are reasonable below this frequency, and the higher frequency, the worse penetrability of RF signals is. In particular, the Industrial Scientific Medical (ISM) band is widely adopted in through-wall technology since most daily used wireless devices, e.g., WiFi and Bluetooth, operate at the ISM band of 2.4GHz and 5.8GHz.

The performance of through-wall systems is typically evaluated by the following key performance indicators (KPI):

- **Detection Ability** refers to the accuracy of the through-wall detection. According to specific applications, the detection ability is measured in different criteria. For through-wall imaging systems, the detection ability is evaluated by the resolution of output images; for through-wall localization systems, it is measured by the accuracy of locations of the detected targets; for through-wall human body detection systems, it is measured by the number of people they can detect.
- **Economical Efficiency** refers to the cost of a whole through-wall system.
- **Energy Efficiency** refers to the power consumption of a through-wall system.

The objective of this paper is to survey the most commonly used techniques, existing systems, and challenges in the through-wall field. Rather than the through-wall technology in a general and broad sense, this paper mainly focuses on the handy through-wall systems which are suitable for civil use (satisfy the Federal Communications Commission (FCC) regulations and in ISM band) with low transmit power (less than 1W), narrow bandwidth (less than 2GHz), lightweight and easy to deploy. The techniques for identifying, locating, or imaging the human body behind a wall are the focus of our work.

The structure of this paper is illustrated in Table 1. Section II lists several critical challenges encountered in the system design. Section III presents the four different categories of related systems or works. Section IV describes the basic principles of through-wall technology, and Section V introduces commonly used techniques to achieve through-wall detection. In Section VI, a number of existing through-wall systems are discussed with their functions, techniques, advantages, and limitations. Section VII closes the paper with concluding remarks and open research issues.

II. CHALLENGES OF THROUGH-WALL TECHNOLOGY

Although RF signals have natural advantages in the through-wall scenario, there are still fundamental engineering challenges as below.

A. SIGNAL ATTENUATION

The transmitted signal typically suffers a great attenuation, especially through dense obstacles. For example, a one-way traversal of a concrete wall can reduce WiFi signal power by 18 dB [3]. In the practical use of a through-wall system, the signal needs to traverse the wall several times, making the received signal very weak. As a result, the useful information in the received signal is prone to be buried in both the signal reflected off the wall and the direct signal from transmit antennas. The basic propagation principle is that the higher frequency, the worse penetrability is; nevertheless, the higher frequency can bring better resolution in detection. Therefore, it is a trade-off to choose an appropriate frequency in through-wall systems. A survey about through-wall radar points out that researchers tend to build through-wall radar below 3GHz [4]. Apart from that, signal strength would reduce as the propagation distance increases which has a notable effect on detecting multiple targets with different distances [5]. In multi-target scenarios, the target close to the device often has a stronger signal than the ones far from the device. In this case, the signals from those targets far from the device are prone to be buried by the stronger signals from the near targets. Therefore, setting a simple threshold to determine the existence of targets is insufficient and

through-wall systems need to process signals from different targets differently to take the distance attenuation into consideration. Besides that, it is also challenging in tiny movements detection (e.g., breathing and heart rates detection) through a wall. The fluctuation in the reflected signal caused by the tiny movements is even weaker than the reflected signal from the whole human body. To address this challenge, Vital-Radio, a wireless breathing and heart rates measurement system, separates the space into multiple buckets based on the distance from the device and extracts the periodic signal components in each bucket [2]. By separating the space into multiple buckets, the signal with useful breathing and heart rates information would be isolated from the signals reflected off furniture and walls. Then, if users keep static or do just small movements (e.g., typing, watching TV), the dominating periodic components in the received signal are mainly caused by breathing and heartbeat, which makes it easier to measure breathing and heart rates.

B. SIGNAL MEASUREMENT

An important parameter in through-wall systems is the time-of-flight (TOF), which is the time that signal takes from the transmit antenna to the receive antenna. Because the transmission speed of the RF signal is very high, it is very hard to accurately measure TOF. For example, one signal only takes around 0.1 ns to travel through 30 m. Traditional radars leverage high-powered systems with high-speed ADCs (analog-to-digital converter) to tackle this issue and they are mostly used in outdoor environments to detect objects at a distance of hundreds or thousands of meters. However, high-speed ADCs are expensive, power-consuming, and have low bit resolution [1], which make traditional through-wall radars bulky and costly. Moreover, the short distance of indoor applications impose even more challenges in accurate TOF measurements.

To address this challenge, through-wall systems mostly resort to three techniques: (1) leveraging frequency modulated continuous wave (FMCW) to measure the TOF indirectly, (2) measuring the direction of arrival (DOA) to obtain angle information instead of the distance information obtained by TOF, and (3) leveraging the pattern of received signal strength (RSS) or channel state information (CSI) to achieve localization. These techniques will be discussed in detail in Section III and V.

C. MULTIPATH EFFECT

The complicated environment of through-wall detection would cause a multipath effect, which would get worse in multi-target scenarios since more targets are accompanied by more reflection of signals. For example, in the indoor environment, through-wall systems need to deal with multipath signals such as ones reflected off walls and furniture. Such interference signals would bury the signal from intended targets. Moreover, the received signal from targets may not only be received directly by antennas but also be reflected off other objects in the environment or even other surrounding targets and then received by antennas. These indirect signals

would result in the inaccurate location estimation of the target or even create some virtual targets which do not really exist. Virtual targets caused by multipath are very similar to the scenario in which a person is standing by a mirror and an observer may observe two persons, one is the real person, the other is the virtual person in the mirror. Therefore, those virtual targets would confuse the through-wall system like a virtual image of targets in a mirror or like a ghost and some efforts have been made to alleviate the impact caused by multipath specifically in the radar field. Chen *et al.* proposed a method of multipath ghost elimination [6]. Setlur *et al.* derive a model for the multipath in an enclosed room and propose a way to associate the ghosts back to the true location [7]. Those methods of multipath elimination are highly based on the precise model of multipath signals with certain radar techniques (e.g., Synthetic Aperture Radar (SAR)). Thus, for some handy through-wall systems which do not use the same radar technique, it is hard to apply these methods of multipath elimination to these systems directly.

To address the multipath effect, different through-wall systems leverage different methods based on the techniques they use and their scenarios. For instance, Wi-Vi exploits Multiple Signal Classification (MUSIC) algorithm to extract useful information from highly correlated signals (e.g., multipath signal) [3]. WiTrack takes the assumption that the multipath signals always travel a longer distance than the directly received signal, therefore, by determining the signal with the minimum propagation distance, it can alleviate the multipath effect and achieve higher accuracy [1]. However, this assumption only holds in the one-target scenario since there are multiple direct signals reflected off multiple targets in multi-target scenarios so that the minimum propagation rule can only determine one direct signal. To address the multipath effect with multiple targets, WiTrack2.0 adds the RF heatmaps from multiple pairs of transmit and receive antennas together to strengthen the direct signals [5]. Specifically, since different antenna pairs are placed in different locations, they suffer from different patterns of multipath signals whereas similar patterns of direct signals. Then, by adding the results from multiple antenna pairs, direct signals with similar patterns would superpose each other and become more salient.

D. MIRROR EFFECT

The wavelength of RF signals is quite longer than that of the visible light; in RF, body parts would reflect the signals from the transmit antenna but not scatter them. However, scatter is always more useful because scattered signals are visible in all directions, whereas reflected signals are very directional. Such directivity could result in only part of the reflected signal being received by antennas. For instance, the signal reflected off the chest is along the direction of the receive antenna but meanwhile the signal reflected off the legs is in the direction of the floor, so that only the chest is visible at this time. In other words, different body parts act like mirrors facing different directions thus some body parts cannot reflect

the signal to the receive antennas, which is called the mirror effect [8]. In order to plot the whole human body via RF signals, one snapshot is insufficient, thus through-wall systems need extra operations. Fortunately, as the human body moves, the directions of reflected signals vary over time. Therefore, in different time periods, the visible parts of a human body are always different, which makes the capture of the whole human body possible. RF-Capture, a through-wall imaging system, leverages consecutive reflection snapshots to reconstruct the whole human figure based on the fact that different snapshots consist of different parts of the human body [9]. Likewise, RF-Pose, a through-wall imaging system with a machine learning method, needs to feed a sequence of consecutive frames of RF heatmaps into the neural network to address the mirror effect [10].

E. PRACTICAL ISSUES

One of the most practical issues is designing real-time through-wall systems which can process the received signals efficiently and render the result to users with low time delay. Handy through-wall systems are always more compact and with limited computing power, so that how to control the computation complexity is a very critical problem in system design. It is common that most through-wall systems are not able to do a perfect job to extract all kinds of information from the signals reflected off targets and most of them are designed to be an expert in extracting certain information. For instance, Wi-Vi leverages off-the-shelf WiFi devices to estimate the coarse angle information of targets behind a wall [3]. It is designed to obtain the number of people behind a wall rather than to estimate the accurate people's locations. WiTrack can only accurately locate one person behind a wall because multiple people would give rise to a more complicated environment (e.g., stronger multipath effect and distance attenuation) where more computation resources are needed to process signals [1]. RF-Capture, a through-wall imaging system, resorts to a strategy named coarse-to-fine scan to reduce the computation complexity [9]. Specifically, RF-Capture first scans a wide area with low scan resolution to find the coarse location of the target, and then scans only this target area with high resolution to achieve fine imaging. Other through-wall systems, such as through-wall breathing and heart rate measurement systems [2], do not locate the accurate location of targets and only extract useful vital information from signals. Therefore, to achieve real-time through-wall detection, designers have to make a trade-off between capacity and complexity.

Another practical issue is the relatively low detection resolution of RF signals comparing to the visible light. This issue actually limits the performance of today's through-wall systems in many aspects. For gesture classification, due to the low resolution of RF signals, most through-wall systems can only recognize large movements such as walking forward or backward, standing up or sitting down, and raising an arm in a certain direction. Moreover, tracking the trajectory of certain body parts is still challenging, researchers point

out that due to the mirror effect, some body parts cannot reflect the signal to the receive antenna efficiently in certain directions or places [9]. Though some remarkable work has achieved fine-grained gesture and motion recognition via CSI, these systems cannot work well in through-wall scenarios and nearly have no ability to locate targets behind a wall [11]–[13].

For through-wall systems leveraging off-the-shelf devices such as WiFi and Radio Frequency Identification (RFID), band interference is also a crucial factor affecting the system performance. Most WiFi-based systems extract useful information from measured CSI or RSS. Take CSI as an example, in most experiments, CSI is obtained by an open-source CSI Tool which measures channel matrices for 30 subcarriers in the orthogonal frequency division multiplexing (OFDM) WiFi system [14]. Fine-grained gesture and motion recognition is achieved by analyzing the CSI pattern of different gestures and motions. However, once there are other WiFi devices operating within the same WiFi channel, signals from other WiFi devices would contaminate the measured CSI and then influence the performance of these systems. WiFi-based through-wall systems face the same situation that the ubiquitous WiFi devices may cause interference to them. Therefore, the robustness of such through-wall systems is influenced by band interference.

Apart from the above challenges, some researchers also focus on the properties of walls that could impact the performance of through-wall systems. When penetrating walls, RF signal would behave differently due to various electromagnetic parameters of walls. Therefore, by operating the estimation of wall properties, such distortion can be corrected. Solimene *et al.* propose a method to estimate the wall transmission coefficient by MIMO which is both time-saving and resource-saving [15]. Vishwakarma *et al.* propose a machine-learning-based method to mitigate wall interference effects [16].

III. CATEGORIES OF THROUGH-WALL SYSTEMS OR WORKS

Through-wall systems or systems with certain through-wall detection ability can be divided into four categories: (1) WiFi-based system, (2) radio tomographic imaging (RTI) system, (3) traditional through-wall radar, (4) and software-defined radio (SDR) system. The general setup of the four categories is shown in Fig. 1.

A. WiFi-BASED SYSTEM

Nowadays, nearly everywhere indoor is covered by WiFi signals and WiFi devices. WiFi-based systems are defined as the systems which leverage WiFi devices as a platform and detect things of interest by measuring the received WiFi signal. There are mainly three approaches for WiFi-based systems: (1) analyzing the Received Signal Strength (RSS) from the MAC layer [17]–[21], (2) analyzing the Channel State Information (CSI) from the physical layer [22]–[25], and (3) analyzing WiFi signals through software radio technology.

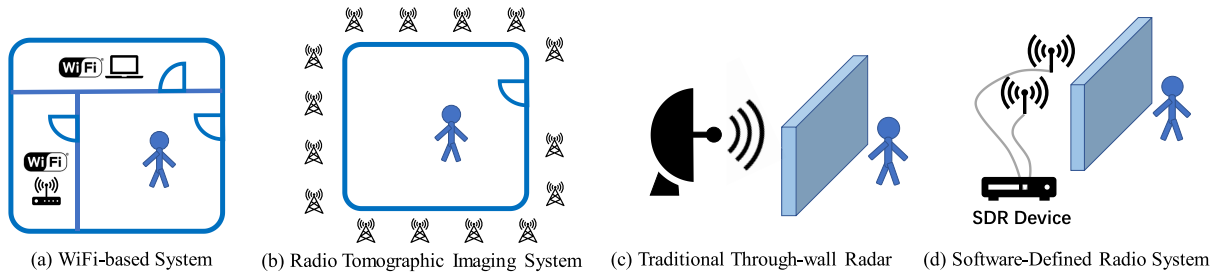


FIGURE 1. A brief illustration of the setup of the four categories of through-wall systems. (a) shows the general setup of WiFi-based systems. Most WiFi-based systems measure the channel fluctuation caused by human movement between multiple WiFi devices within a WiFi network. Other WiFi-based systems just leverage a single WiFi device to transmit signals, like radar, and then extract human body information from the received signals. (b) shows the conception of radio tomographic imaging (RTI). RTI systems measure the signal strength between each pair of sensors in a dense sensor network to capture human body information. (c) is the classic scenario of traditional through-wall radar. (d) shows the simplest setup of the software-defined radio (SDR) system. Such systems usually use several antennas to transmit and receive signals and then process received signals by SDR devices, e.g., USRP and personal computers.

RSS represents the received signal power or the signal-to-noise ratio (SNR) which can be measured within one or more packets transmitted in WiFi. RSS-based systems analyze the pattern of varying RSS, i.e., the received signal strength at the receiver, to extract information of obstacles (e.g., furniture, walls, etc.); the presence or movement of a human body would also cause an influence on the RSS. Nuzzer, proposed by Seifeldin *et al.*, utilizes changes in RSS measured by wireless networks to locate users directly [17]. EZ, proposed by Chintalapudi *et al.*, determines the location of users' mobile devices by measuring the RSS at different access points (APs) [18]. It is worth noting that EZ locates users by locating their mobile devices and therefore is not device-free. WiGest, proposed by Abdelnasser *et al.*, utilizes changes in RSS to sense hand gestures which can be used to control a video player [19]. Mostofi *et al.* achieve through-wall imaging based on mobile platforms such as unmanned aerial vehicles (UAVs) [20] and robots [21].

Different from RSS which records the collective signal strength at the receiver, CSI between one pair of transmit-receive antennas contains the signal strength and the phase information of each OFDM sub-channel. As the signal strength and the phase information of one sub-channel reflect the properties of this channel, such as the information of signal propagation, and the effect of path loss, scattering, fading, etc., CSI describes more detailed information about the channel compared to RSS, and accordingly the processing of CSI is more complicated but also more reliable and accurate. E-eyes, proposed by Wang *et al.*, leverages CSI measured from a WiFi network to track activities of one user [22]. Wang *et al.* show that CSI is more fine-grained than RSS by detecting the same activity with CSI and RSS, respectively, and comparing their results. FarSense, proposed by Zeng *et al.*, can monitor human respiration by CSI even in through-wall scenarios with commodity WiFi devices [23]. FarSense employs the ratio of CSI readings from two antennas to reduce noise and the ratio enables the use of phase information to extract respiration information. WiSpy, proposed by Hanif *et al.*, can sense the movement of persons

behind a wall by analyzing CSI and predict their number as well by applying machine learning algorithms to CSI data [24]. WiGeR, proposed by Al-qaness *et al.*, can achieve gesture recognition based on CSI extracted from any common WiFi router [25].

The software radio scheme applies the universal software radio peripheral (USRP) [3], [11] to process WiFi signals, or wireless open access research platform (WARP) to design customized WiFi systems [26]. Note that this type of system also meets the definition of Software-Defined Radio (SDR) system, as discussed in Section III-D. Specifically, USRP is an integrated hardware platform for software radio which allows users to configure and process the radio through software, such as modulation, coding, and signal processing. In other words, USRP provides more freedom in adapting WiFi radios than standard WiFi protocols and thus can provide and enables the deployment of radar techniques and complicated frequency analysis for more accurate measurement at the cost of system cost and complexity. Wi-Vi, proposed by Adib *et al.*, can detect the moving people behind a wall [3]. It leverages the MIMO of the WiFi device to address the flash effect, which would be discussed in detail in Section VI. USRP is then employed to process WiFi signals with radar methods. WiSee, proposed by Pu *et al.*, is a gesture recognition system leveraging USRP and can work in through-wall scenarios [11]. WARP has been developed for the algorithm or system design and is compatible with the commercial WiFi standard. WiDeo, proposed by Joshi *et al.*, is a through-wall device-free motion tracing system with customized algorithms to process WiFi signals on WARP [26].

In conclusion, RSS is the most accessible information since you can obtain the received signal power or SNR easily on nearly every wireless system. Compared to RSS, CSI can provide details about each sub-channel (i.e., the strength and phase information) instead of the rough power provided by RSS. Though CSI is supported in 802.11n protocols, not every network interface controller (NIC) provides access to CSI data from the physical layer. The method most commonly used by researchers is CSI Tool with Intel 5300 NIC [14].

Therefore, it is harder to obtain and more expensive to process CSI data. USRP is a peripheral device independent of WiFi whereas we can obtain RSS and CSI merely through WiFi devices. The processing flexibility and ability of USRP outweigh WiFi devices but the cost of the USRP hardware platform is much higher than off-the-shelf WiFi devices.

The most significant advantage of WiFi-based systems is that they can exploit off-the-shelf WiFi devices, and therefore they are relatively cheaper and easier to implement. Besides that, the transmit power of WiFi devices is usually less than 50 mW which is lower than that of traditional through-wall radars. But the disadvantage is that the performance of such systems would be limited by WiFi devices (e.g., bandwidth, sampling rate, etc.) since they are designed for low-cost wireless communications.

B. RADIO TOMOGRAPHIC IMAGING (RTI) SYSTEM

RTI is an emerging technique [27]–[31] which reconstructs the spatial loss field of the environment by a dense wireless sensor network [30]. In specific, by placing a large number of sensors around the target area, RTI systems analyze the spatial accumulation effect of shadowing loss of the RSS to image area of interest [28]. RTI is an effective technique for through-wall detection and its sensor nodes are low-cost and low-power [29]. In order to have better accuracy, most RTI systems need to place a mass of sensors to enclose the whole area of interest. This, however, makes it impractical in some scenarios like hostage rescue.

C. TRADITIONAL THROUGH-WALL RADAR

Traditional through-wall Radar is a well-studied area and has achieved great success, especially in military applications. Radar-based methods can be roughly divided into three categories: ultra-wideband (UWB) radar [32]–[41], Doppler radar [42], [43], and Frequency Modulated Continuous Wave (FMCW) radar [44], [45]. With radar devices, some critical challenges in through-wall scenarios (e.g., flash effect [3]) can be easily solved. The most widely used one is UWB radar because of its high range resolution and good penetrability [32]. However, most traditional through-wall radars are costly and bulky since they need high-performance hardware, higher transmit power, and large antenna arrays which are the key features that make them different from the other three categories. For instance, though some SDR through-wall systems (will be discussed in Section III-D) also adopt techniques from radar fields, such as FMCW, their transmit power is less than 1mW, which is far less than that of the FMCW radars mentioned above (e.g., the transmit power is 1W in [45]). The high transmit power makes it hard for traditional through-wall radars to meet the power limitation in the ISM band, and costly and bulky hardware also prevents traditional through-wall radars from civil use.

D. SOFTWARE-DEFINED RADIO (SDR) SYSTEM

The SDR system refers to the RF system where most hardware components are replaced by software methods.

It requires professional RF devices (e.g., special antennas, radio frequency identification (RFID) devices) and special programmable software platforms (e.g., USRP). The flexibility of software provides SDR systems with more algorithm options (from radar to machine learning). Besides that, SDR systems controlled by software could achieve reasonable resolution with less transmit signal power than WiFi (even less than 1mW) since they can break the limits of WiFi devices (e.g., bandwidth, sampling rate, etc.) and focus only on through-wall purpose. Table 5 summarizes the power consumption of some systems. SDR-based systems have achieved through-wall localization [1], [5], [46], through-wall imaging [9], [10], [47], [48], and human vital signals detection [2], [49]. The advantage of such systems is that signal processing methods are relatively flexible but the disadvantage is the high cost of special devices like USRP.

WiTrack [1] and WiTrack2.0 [5], proposed by Adib *et al.*, are designed for through-wall localization. They leverage USRP to generate FMCW signals to measure the distance from antennas to targets. RF-Capture, proposed by Adib *et al.*, leverages FMCW and USRP to achieve through-wall imaging [9]. Tadar, proposed by Yang *et al.*, leverages an RFID reader and multiple RFID tags to locate people behind a wall [46]. RF-Pose [10], RF-Pose3D [47], and RF-Avatar [48], proposed by Zhao *et al.*, can reconstruct the 2D or 3D model of a human body through RF signals. These three systems combine RF devices and camera devices, and leverage machine learning techniques to train RF systems supervised by vision systems. Besides that, Vital-Radio, proposed by Adib *et al.*, can achieve breathing and heart rate detection with FMCW signals even in through-wall scenarios, which shows the great potential of SDR systems [2].

Table 2 summarizes the systems mentioned above. The reference number with a superscript indicates that the corresponding paper meets the definition of handy through-wall technology in Section I and they are the focus of this paper. The basic principles and methods discussed in Section IV and V are mainly based on these systems.

IV. BASIC PRINCIPLE OF THROUGH-WALL TECHNOLOGY

Through-wall systems transmit electromagnetic waves to detect the space and then receive the reflected signals to obtain the location of targets. Some basic principles of the electromagnetic wave and localization are as follows.

A. PRINCIPLE OF ELECTROMAGNETIC WAVE

Since electromagnetic wave was discovered, it has been applied in numerous fields. Because electromagnetic wave has the ability of transmission and reflection, it is widely used in the through-wall technology with basic principles as follows.

1) TRANSMISSION

Electromagnetic signals can penetrate non-metallic obstacles, which means that these signals can allow us to get valuable information from things behind walls. The transmission

TABLE 2. Related systems/works summary.

Primary Classification	Secondary Classification	Main Function	Ref.
WiFi-based System	RSS	Indoor Localization Gesture Recognition Through-wall Imaging	[17], [18] [19] [20]*, [21]*
	CSI	Indoor Activity Recognition Respiration Detection Through-wall Moving Target Detection People Counting Gesture Recognition	[22] [23] [24]* [25]
	USRP, MIMO	Through-wall Moving Target Detection Gesture Recognition	[3]* [^] [11]* [^]
	WARP, MIMO	Through-wall Motion Recognition	[26]* [^]
Traditional Through-wall Radar	UWB Radar	Through-Wall Target Detection Through-wall Localization Through-wall Activities Classification Building Layout Reconstruction Through-wall Moving Target Detection	[32], [39] [33], [34], [35], [38] [36] [37] [41]
	Doppler Radar	Through-wall Target Detection Respiration Rate Detection	[42] [43]
	FMCW Radar	Through-wall Localization	[44][45]
RTI		Indoor/Through-wall Localization Through-wall Localization	[27], [28], [30], [50], [51] [29]
SDR	FMCW, USRP	Through-wall Localization Through-wall Imaging	[1]*, [5]* [9]*
	FMCW	Respiration Detection	[2]*
	FMCW, Machine Learning	Through-wall 2D Pose Estimation Through-wall 3D Pose Estimation Through-wall 3D Mesh Model Estimation Indoor Motion Recognition	[10]* [47]* [48]* [52]
	RFID	Through-wall Localization	[46]*

*The corresponding papers meet the definition of handy through-wall technology in Section I and some of them are discussed in detail in Section VI.

[^]The corresponding papers also belong to SDR systems.

TABLE 3. Attenuation with different frequencies [54].

Frequency (GHz)	Loss (dB)
1.01	2.06
1.91	3.02
3.11	4.30
4.01	5.27
5.21	6.55

Material: Bricks, 8.7 cm × 19.8 cm × 5.8 cm

performance varies by frequency. Specifically, the higher the frequency, the worse the transmission performance is, which results in a low-pass filtering effect when penetrating an obstacle [53]. For example, visible light is in the band of 3.9×10^5 – 8.6×10^5 GHz, whereas the RF signal is in the band of 3–300 GHz. Consequently, visible light has an extremely poor penetrability but RF signals can penetrate obstacles well or just bypass them like FM (frequency modulated) signals used in radio systems. Apart from that, the performance of transmission is influenced significantly by the material of obstacles. Materials like wood, drywall, and styrofoam have small loss effects whereas brick walls and blocks have significant loss effects [53]. Table 3 shows the attenuation of signals with different frequencies. Table 4 shows the attenuation of signals through different materials.

2) REFLECTION

Transmitted signals are reflected once encounter obstacles such as human bodies and furniture. Then, by analyzing these

TABLE 4. Attenuation through different materials [54].

Material	Dimension (cm)	Loss (dB)
Wallboard	1.2 × 121.8 × 196.9	0.45
Cloth Partition	5.9 × 140.7 × 153.1	2.55
Structure Wood	2.1 × 121.5 × 197.8	1.35
Styrofoam	9.9 × 121.8 × 197.7	0.10
Concrete Block	19.45 × 39.7 × 19.5	13.6

Frequency of the transmitted signal: 5 GHz

reflected signals, through-wall systems can calculate the distance or angle of objects. However, not all the reflected signals are valuable such as ones reflected off walls. Therefore, through-wall systems always need an extra filtering process to filter out unwanted reflections. In addition, there is a great difference between the reflection of RF signals and that of visible signals. Since the wavelength of RF signals is much longer than that of visible light, the human body acts as a reflector rather than a scatterer.

B. PRINCIPLE OF LOCALIZATION

Localization is a core function of through-wall technology. For easy understanding, we assume that the location of the target is in the 2D plane. Through-wall systems leverage the following principles to locate.

First, a widely used method for locating is achieved by the measurement of distance. Specifically, the round-trip distance, from the transmit antenna to the target and then back

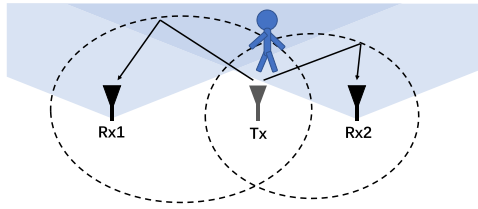


FIGURE 2. Two pairs of directional antennas can uniquely locate the location of a person.

to the receive antenna, is used to calculate the location of the target. Suppose we know a round-trip distance between a pair of antennas, according to math principles, we can plot an ellipse and the sum of distances to the two antennas is equal to the round-trip distance we know. In other words, the couple of antennas is a couple of foci of an ellipse; the location of the target must fall on the periphery of the ellipse. Using one pair of antennas, however, is insufficient to locate the target, and we need another pair of antennas. Similarly, the second pair of antennas with known round-trip distance enables us to plot another ellipse on the plane. The foci of the new ellipse are two other antennas. Since the correct location is on both ellipses, one of the intersection points is the location of the target.

To express mathematically, assuming that the round-trip distance is d_1 and the locations of the pair of antennas are $(x_1, 0)$ and $(x_2, 0)$ in the (x, y) plane, respectively. The standard parameters of this ellipse are

$$2a = d_1 \tag{1}$$

$$2c = |x_2 - x_1|, \tag{2}$$

where a is the length of the semi-major axis and c is the distance between a focus and the center of the ellipse. Then we can derive the equation of this ellipse as

$$\frac{(x - \frac{x_1+x_2}{2})^2}{(\frac{d_1}{2})^2} + \frac{y^2}{(\frac{d_1}{2})^2 - (\frac{|x_2-x_1|}{2})^2} = 1. \tag{3}$$

Similarly, assuming the round-trip from another pair of antennas is d_2 , and the location of this pair of antennas are $(x_3, 0)$ and $(x_4, 0)$, respectively. Then we can derive the equation of this ellipse as

$$\frac{(x - \frac{x_3+x_4}{2})^2}{(\frac{d_2}{2})^2} + \frac{y^2}{(\frac{d_2}{2})^2 - (\frac{|x_3-x_4|}{2})^2} = 1. \tag{4}$$

Now we have the equations of two ellipses on which the location of the target falls. Therefore, by solving (3) and (4), the solution is the possible location of the target.

If the antennas are directional, it's very easy to rule out the invalid intersection, as shown in Fig. 2. If there are more than two pairs of antennas, the location can be uniquely determined. It's easy to generalize the argument to a 3D space. In the 3D space, ellipses established by pairs of antennas would become ellipsoids and 3 ellipsoids can determine one intersection (with directional antennas).

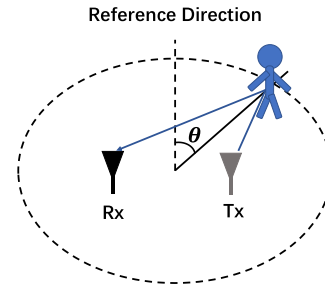


FIGURE 3. Localization can be achieved with a round-trip distance and an angle from a reference direction.

Another way to locate is by leveraging both the distance and the angle. It is similar to the localization in the polar coordinate system. Take a 2D situation for example, each point in the polar coordinate system is determined by a distance from a reference point and an angle from a reference direction. Once we know the distance and the angle of the target, we can locate it accurately. In antenna systems, we can only get the round-trip distance rather than the direct distance to the target, hence like the above statement, we leverage the round-trip distance to plot an ellipse rather than a circle. Then, if we know the angle of the object, we can plot a ray and there will be only one intersection point which indicates the location of the object, as shown in Fig. 3. It is straightforward to generalize the argument to 3D space. In a 3D space, like a point determined by (r, θ, φ) in a spherical coordinate system, the location of an object can be determined by the round-trip distance and two spacial angles.

V. BASIC METHOD OF THROUGH-WALL TECHNOLOGY

Most through-wall systems leverage RF techniques to obtain distances and angles (namely direction of arrival (DOA)) of targets. Nowadays, in radar systems, there are many approaches to measure the distance and DOA, and most of them are based on the following ideas. Since through-wall technology has a strong relationship with radar technology, we can find that some through-wall techniques are similar to radar techniques.

A. OBTAINING DISTANCE

We usually leverage RF signals to measure the distance between two objects. Specifically, we know that RF signals travel at the speed of light so once we obtain the time of flight (TOF), which the signal takes to travel from its transmitter to receiver, we can calculate the propagation distance. The direct and indirect measurements of TOF are as follows.

1) PULSE METHOD

The pulse method is a direct solution to measure the TOF. First, the antenna transmits a very short pulse. Then, the receive antenna receives the echo of the pulse and calculates the time delay. The time delay is exactly the TOF we want. Finally, the distance from the transmit antenna to the

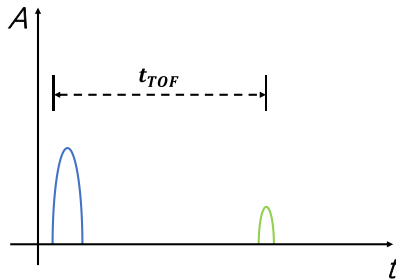


FIGURE 4. The blue one is the transmitted pulse signal and the green one is the echo from the object. The time delay between the two signals is TOF.

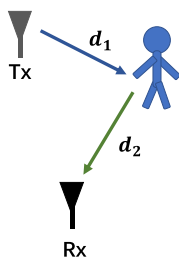


FIGURE 5. d_1 is the distance between the transmit antenna and the target and d_2 is the distance between the target and the receive antenna. $d_1 + d_2$ is the round-trip distance.

target and back to the receive antenna is

$$d_1 + d_2 = c \times t_{TOF}, \tag{5}$$

where d_1 is the distance between the transmit antenna and the target, d_2 is the distance between the target and the receive antenna, c is the propagation speed, and t_{TOF} is the time delay. Fig. 4 and Fig. 5 illustrate this situation.

The advantage of this method is that the measurement is simple and does not need complex algorithms. However, the propagation speed is so fast that TOF at short range is extremely hard to measure. Specifically, in order to obtain such a short TOF, the through-wall system must operate at a very high speed to sample the signals at subnanosecond intervals. Therefore, high-speed ADCs that operate at multi-GS/s are necessary [1]. Such ADCs are power-consuming, expensive, and not practical in small-size systems.

2) FMCW

Frequency modulated continuous wave (FMCW) technique is an indirect method to measure the TOF. FMCW transmits a periodic signal whose carrier frequency changes linearly with time. The chirp reflected off objects and then received by antennas has both time delay and frequency shift. Instead of measuring the delay directly, FMCW leverages the frequency shift to calculate the delay. Since the carrier frequency is changing linearly in time, the delay and the frequency shift are linearly related. In Fig. 6 (a), the blue line is the transmitted signal whose frequency is changing linearly with time. The green line is the received signal reflected off an object. Because the received signal delays in time, there is a

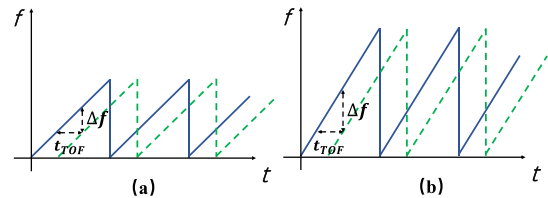


FIGURE 6. (a) shows the transmitted signal in blue and the received signal in green. The received signal has both the time delay (t_{TOF}) and the frequency shift (Δf) and these two differences are linearly related. (b) shows that with a higher rate of change in frequency, the Δf would be more notable.

frequency shift between the transmitted and received signals at the same time. Obviously, the TOF and round-trip distance can be calculated by the equations as

$$t_{TOF} = \frac{\Delta f}{k} \tag{6}$$

$$d = c \times t_{TOF}, \tag{7}$$

where k is the slope of the frequency.

Besides that, to show the advantage of this method, we assume that the object is static, i.e., the distance between the object and antennas is constant. Hence, the time delay of the received signal is constant. According to (6), Δf only depends on k which is the rate of the frequency change. Therefore, the higher the rate of change, the larger the Δf is, as shown in Fig. 6 (b). Because the TOF is always very short in the short-range situation, we can obtain the TOF much easier by measuring the Δf which can be more than billions of times larger than TOF. For example, if the carrier frequency changes from 5 GHz to 7 GHz per second, theoretically, the frequency shift would be 2×10^9 times TOF, hence the easier measurement of TOF.

When applying FMCW to a real system, system designers need to make sure that the carrier frequency of the transmit signal changes linearly with time. To achieve this goal, the linearity of the hardware of the transmitter within the bandwidth is very important and any nonlinear components in the transmitter would affect the estimation of TOF. Take a voltage-controlled oscillator (VCO) as an example, if the control signal is out of the linearity tuning range, the output frequency of VCO would not satisfy the linear relationship with the control signal. Moreover, the selection of the bandwidth of FMCW affects not only the linearity of the output frequency, but also the detection resolution of FMCW. Typically, the resolution of FMCW, i.e., the minimum measurable change in location, is:

$$R = \frac{c}{2B}, \tag{8}$$

where B is the bandwidth of the FMCW signal [1]. Therefore, in order to achieve a sub-meter accuracy, FMCW-based systems always have the bandwidth of multiple GHz, as shown in Table 5. Note that the bandwidths of such through-wall systems are selected around 1.7GHz to satisfy the FCC regulations for civil use.

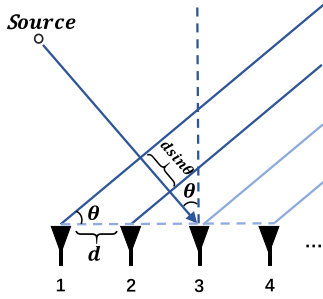


FIGURE 7. The estimation of DOA (θ) can be achieved by an antenna array which is placed uniformly in a straight line and the distance between two adjacent antennas is d . With the assumption that the wave arriving at the array is considered to be planar, the distance that the signal propagates between two adjacent sensors can be calculated as $d \sin \theta$. Therefore corresponding time delay τ is $\frac{d \sin \theta}{c}$ [55].

B. OBTAINING DIRECTION OF ARRIVAL

For localization, sometimes the distance information is insufficient. There is another technique that allows us to obtain the direction of arrived signals (DOA) and can be used to detect the angles of objects. This technique leverages antenna arrays to achieve this goal.

First of all, assume we have an antenna array containing m identical antennas which are placed uniformly in a straight line and the distance between two adjacent antennas is d , as shown in Fig. 7. After the demodulation, the output of antenna k ($k = 1, 2, \dots, m$) can be written as the following equation [55]:

$$y_k(t) = H_k(\omega_c) e^{-j\omega_c(k-1)\frac{d \sin \theta}{c}} s(t) + e_k(t), \quad (9)$$

where θ is DOA, c is the propagation speed of signals, ω_c is the carrier frequency, $s(t)$ is the baseband signal, $H_k(\omega)$ is the frequency response of k th antenna within the band of $s(t)$, and $e_k(t)$ is the noise received at the k th antenna.

Then, with m sensors, we will obtain

$$y(t) = a(\theta)s(t) + e(t), \quad (10)$$

where

$$\begin{aligned} y(t) &= [y_1(t) \dots y_m(t)]^T \\ a(\theta) &= [H_1(\omega_c), H_2(\omega_c) e^{-j\omega_c \frac{d \sin \theta}{c}} \\ &\dots H_m(\omega_c) e^{-j\omega_c(m-1)\frac{d \sin \theta}{c}}]^T \\ e(t) &= [e_1(t) \dots e_m(t)]^T. \end{aligned}$$

By (10) with known $y_k(t)$, $s(t)$, and $H_k(\omega_c)$ ($k = 1, \dots, m$), we can obtain θ (DOA) approximately. The details of derivation are illustrated in the appendix.

In the real-world scenario which contains more than one source, however, the environment would become more complicated since more sources could bring more noises and interferences. Some advanced algorithms were proposed to estimate multiple DOAs. The most famous algorithm is Multiple Signal Classification (MUSIC) algorithm, which performs well in multi-DOA estimation [56]–[58].

C. MACHINE LEARNING

To achieve through-wall localization with higher precision and more fine-grained through-wall motion classification, system designers need to formulate more sophisticated models to fit the complicated environment caused by minor movements of the human body and fine-grained motions. Therefore, researchers resort to machine learning, which is a powerful technology to find the hidden relationship in a complicated environment. The most used machine learning method is the convolutional neural network (CNN) which has a good performance in image processing. The reason that CNN is widely adopted in through-wall systems is that the formats of RF signals (e.g., RF heatmaps which indicate the signal strength in different positions) are quite similar to images. In the through-wall imaging scenario, each snapshot consists of lots of pixels representing the spacial information of RF signals. However, feeding the whole snapshot to a fully connected neural network is intractable since such a dense neural network would contain a huge number of parameters which are hard to train and easy to overfit. Therefore, CNN is used to reduce the complexity and extract only important features from the input RF signals and then uses these features to predict the positions of targets [10], [47], [48].

The traditional structure of CNN is shown in Fig. 8. There are three kinds of layers in CNN: Convolutional Layer, Pooling Layer, and Fully Connected Layer. Convolutional layers are used to extract features from the input data. Each neuron in a convolutional layer is connected with a group of neurons in the previous layer, and the relationship between the neurons in two layers is determined by a matrix which contains the weights of connection between neurons in two layers. This matrix is called Kernel or Filter. Specifically, a neuron in one layer is calculated by the convolution operation between the previous layer and a Kernel. Different Kernels can capture different features of data from the previous layers. Then, the pooling layer is used to reduce the spatial size of the convolved feature from convolutional layers. The main purpose of this layer is to identify dominant features and ignore the trivial ones in order to reduce the computation complexity. After multiple layers of convolution and pooling, the extracted features are put into the following fully connected layers, which are used for classification. In short, convolutional and pooling layers are responsible for feature extraction, and fully connected layers are responsible for mapping different combinations of features with different outputs.

The structure of the neural network in CNN-based through-wall systems varies from system to system but the ideas of extracting features from RF signals are quite similar. Specifically, a common way to process RF signals is called the spatio-temporal convolution, which applies convolution operations to both the spatial dimension and the temporal dimension. The motivation is that, in order to locate different body parts, the neural network should not only focus on the information in one snapshot (spatial dimension) but also take a series of consecutive snapshots (temporal dimension)

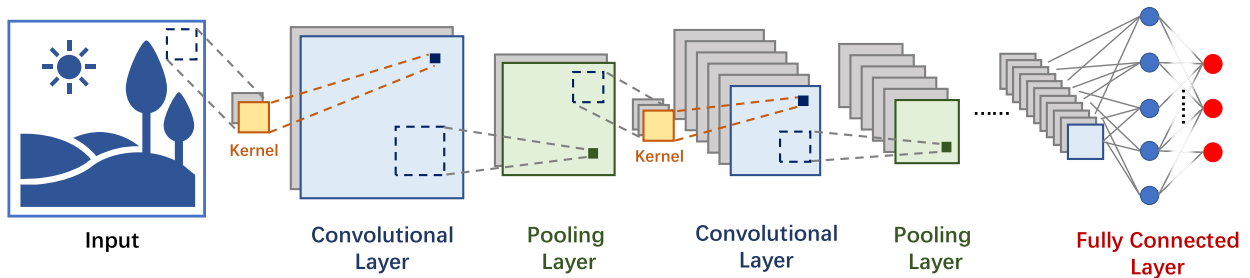


FIGURE 8. Structure of CNN. The convolutional layers are used for extracting features, and the more convolutional layers, the more high-level features (e.g., faces and eyes) can be extracted. The pooling layers are designed to extracting dominant features after convolutional layers. The last layer is the fully connected layer, which works as a classifier to match different features with corresponding objects.

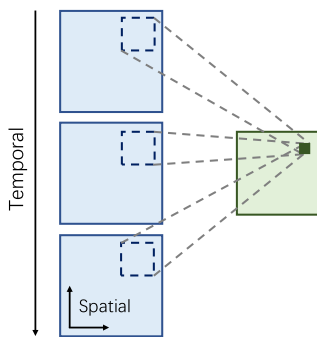


FIGURE 9. Spatio-temporal convolution. The convolution operation is applied to a series of consecutive RF snapshots to extract features from both spatial and temporal dimensions [59].

into consideration, since the motions of the human body are continuous and highly correlated over time. Based on this motivation, the common input of CNN is a series of consecutive RF snapshots and the convolution operation is applied to both spatial and temporal dimensions, as shown in Fig. 9. For instance, in the 2-dimensional through-wall imaging scenario, we need 3-dimensional (2 spatial dimensions and 1 temporal dimension) convolution operations to process RF data; in the 3-dimensional through-wall imaging scenario, we need 4-dimensional (3 spatial dimensions and 1 temporal dimension) convolution operations to process RF data. Note that the higher the dimension of convolution operations, the higher the computational complexity is. Applying 4-dimensional convolution operations to extract features from RF signals directly is time-consuming and even mainstream open-source machine learning libraries (e.g., Tensorflow) only support up to 3-dimensional convolution operations. Therefore, Zhao *et al.* proposed a tensor decomposition technique to reduce the convolution dimension and leveraged auxiliary neural networks to help CNN focus on the regions with targets to reduce the computational complexity [47], [48].

In through-wall systems, machine learning is often used in cross-modal supervision. This is because the training set, i.e., RF data labeled with true human body positions, is insufficient in the through-wall field. Thus, systems with machine learning leverage vision-based human body tracking systems to supervise the training. Specifically, in the training

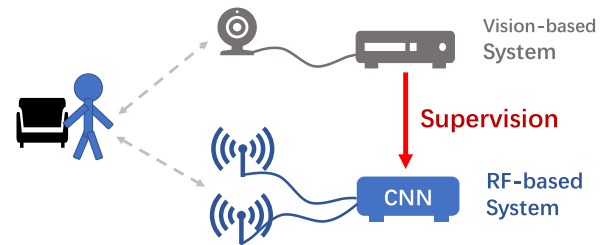


FIGURE 10. Through-wall systems with machine learning methods are usually based on the structure of cross-modal supervision. The vision-based system is used to supervise the RF-based system.

stage, the RF-based system and vision-based system work together to extract human bodies respectively. The results of the vision-based system are then used to amend the neural network in the RF-based system. Most often, the vision-based system is based on another well-trained neural network. After the training, the RF-based system can work alone. The cross-modal supervision structure is shown in Fig. 10.

VI. EXAMPLE OF HANDY THROUGH-WALL SYSTEMS

In recent years, there is a rapid rise of handy through-wall systems with the functions of identification, localization, or imaging. These systems have low power, narrow bandwidth, and lightweight. Besides that, different from some traditional monitoring devices, these systems do not need extra devices equipped on the human body. Some radar techniques (e.g., FMCW and ISAR) and machine learning techniques (e.g., CNN) have been used in the existing through-wall systems. Some of these systems are explained in detail as follows and crucial implementation characters of these systems are summarized in Table 5, and their performance is summarized in Table 6.

A. Wi-Vi

Wi-Vi is a wireless device which was proposed by Adib *et al.* in 2013 [3]. It belongs to both the WiFi-based system defined in Section III-A and the SDR system defined in Section III-D. The main function of Wi-Vi is to detect the movement of people walking behind the wall. Wi-Vi leverages the MIMO technique, ISAR, and MUSIC algorithm to work in the through-wall scenario. Furthermore, another function of

TABLE 5. Handy through-wall systems implementation summary.

System	Frequency	Antenna	Power	Main Method	Function	Ref.
Wi-Vi	2.4 GHz	2 Tx 1 Rx	20 mW	WiFi USRP MIMO ISAR	Wi-Vi can detect and count moving people and allow users to transmit information by walking forward and backward without any carried devices.	[3]
WiTrack	5.46–7.25 GHz	1 Tx 3 Rx	0.75 mW	USRP FMCW	WiTrack can track the 3D position of a single moving person, estimate the pointing angle of an arm, and detect a fall.	[1]
WiTrack2.0	5.46–7.25 GHz	5 Tx 5 Rx	≤ 1 mW	USRP FMCW	WiTrack2.0 can track the 3D position of up to 5 users even when they are static.	[5]
RF-Capture	5.46–7.24 GHz	4 Tx 16 Rx	0.07 mW	USRP FMCW	RF-Capture can capture the human figure, identify and track the trajectory of certain body parts.	[9]
Vital-Radio	5.46–7.25 GHz	1 Tx 1 Rx	≤ 1 mW	FMCW	Vital-Radio can monitor breathing and heart rate without body contact.	[2]
RF-Pose	-	-	≤ 1 mW	CNN FMCW	RF-Pose can track the 2D human pose through RF signals with machine learning methods.	[10]
RF-Pose3D	5.4–7.2 GHz	-	≤ 1 mW	CNN FMCW	RF-Pose3D can track the 3D human pose through RF signals with machine learning methods.	[47]
Tadar	920–926 MHz	1 reader 45 tags	≤ 1 W ^a	RFID	Tadar can capture the motion of a mobile object behind a wall with RFID tags attached on the outer wall.	[46]

^aThe RFID reader that Tadar uses is Impinj Speedway R420. Its maximum transmit power is 1 W.

TABLE 6. Handy through-wall systems performance summary.

System	Evaluation Criterion	Performance	Range ^a	Detection Ability ^b	Ref.
Wi-Vi	Accuracy of detection of human	100% with 0, 1 human 85% with 2 humans 90% with 3 human	≤ 10 m	3 humans	[3]
WiTrack	Median location error	13.1 cm along x dimension 10.3 cm along y dimension 21.0 cm along z dimension	3–11 m	1 human	[1]
WiTrack2.0	Median location error	8.4–16.1 cm along x dimension 7.1–10.5 cm along y dimension	≤ 10 m	5 humans	[5]
RF-Capture	Accuracy of identification of body part	99.13% 76.4%	3 m 8 m	1 human	[9]
Vital-Radio	Accuracy of breathing rate Accuracy of heart rate	99.2% 90.1%	4 m ^c	1 human ^d	[2]
RF-Pose	Average Precision	58.1%	≤ 12 m ^e	3 humans ^f	[10]
RF-Pose3D	Average location error	5.2 cm along x dimension 3.7 cm along y dimension 4.7 cm along z dimension	1–10 m	5 humans	[47]
Tadar	Median location error	7.8 cm along x dimension 20.0 cm along y dimension	≤ 7.5 m	1 human	[46]

^aThe ranges of the above systems indicate the maximum detection range of the corresponding system or the maximum range at which the system is tested.

^bThe detection ability of the above systems indicates the maximum number of humans that a system can detect or the maximum number of people during the test.

^cThe maximum tested range of Vital-Radio in through-wall scenarios is 4 m whereas the maximum range is 8 m in line-of-sight scenarios.

^dThe number of tested humans is 1 in the through-wall scenario whereas the maximum detection ability is 3 humans in line-of-sight scenarios.

^eThe maximum range stated in the original paper is 40 feet, and we convert that to m in this table for easier comparison.

^fThe maximum number of humans in each frame in the through-wall testing set is 3.

Wi-Vi is gesture encoding by tracking the motion of the human body. Wi-Vi allows a person to send bit information as “1” and “0” by moving forward and backward without any communication device.

Wi-Vi is a MIMO device with 2 transmit antennas and 1 receive antenna and leverages WiFi OFDM signals in the ISM band (at 2.4 GHz). Besides, Wi-Vi is based on the typical WiFi hardware which makes it relatively low-power, low-cost, and low-bandwidth. Since it has the standard WiFi structure, Wi-Vi might be implemented on WiFi devices and operated by the public in the future.

Wi-Vi has to address several challenges and the toughest one is called flash effect. In order to detect things behind walls, signals need to traverse the wall at least twice which would lead to significant signal attenuations. Therefore the reflections from the objects of interest are buried in the reflections off the wall as well as the signal directly received from the transmit antenna. These two kinds of signals are so strong that overwhelm the ADC of the standard WiFi device (the overflow in ADC means background subtraction is not desirable). Wi-Vi utilizes the MIMO technique to eliminate the flash effect.

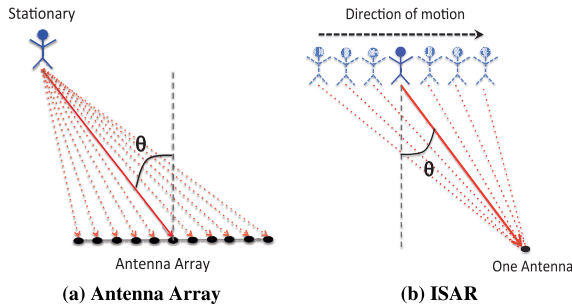


FIGURE 11. Generally speaking, DOA is obtained by an antenna array, as shown in (a). Wi-Vi with only one receive antenna, as shown in (b), exploits ISAR to simulate a virtual antenna array by the movement of the target (figure from [3]).

First of all, to eliminate the flash effect, Wi-Vi estimates a special channel between the first transmit antenna and the receive antenna and another channel between the second transmit antenna and the receive antenna. Then Wi-Vi lets the signals from two transmit antennas interfere with each other at the receive antenna. In this case, Wi-Vi can eliminate the direct signals and all the reflections off the static objects such as walls and furniture. After the flash is removed, Wi-Vi can receive only signals reflected from moving objects. The next step is localization. Because there is only one receive antenna, Wi-Vi employs a technique called Inverse Synthetic Aperture Radar (ISAR). ISAR leverages the movement of the target to emulate an antenna array. As mentioned above, to obtain the DOA of signals, an antenna array is necessary. Hence ISAR provides Wi-Vi a virtual antenna array to obtain the DOA. Fig. 11 illustrates ISAR used in localization. Comparing with a real antenna array, ISAR with only one antenna is cheaper and simpler. In the scenario of tracking multiple people, since the received signal is a superposition of multiple moving humans with significant noise and correlates in time, Wi-Vi leverages the smoothed MUSIC algorithm to process the received signals. Finally, data of DOA are the only straightforward output of Wi-Vi and all the other functions (e.g., number identification and gesture-based communication) are based on the information of DOA.

1) EXPERIMENTAL EVALUATION

Wi-Vi can distinguish between 0, 1, 2, and 3 moving humans with an accuracy of 100%, 100%, 85%, and 90% respectively at a room with 6-inch hollow walls supported by steel frames. Wi-Vi correctly decodes all the messages within 5 meters. And the decoding accuracy decreases to 75% at 8 meters. Besides, Wi-Vi cannot detect gestures beyond 9 meters.

2) LIMITATIONS

First, Wi-Vi can only track moving objects, i.e., if a person keeps stationary, its existence will not be detected. Second, the detectable number of individuals is limited because more individuals lead to more complicated situations. Third, the virtual antenna array is affected by the velocity of moving

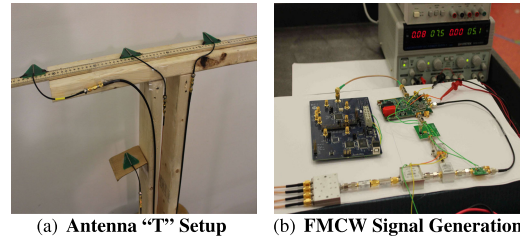


FIGURE 12. WiTrack implementation (figure from [1]). (a) shows WiTrack's antenna setup. One receive antenna is in the middle of the upper three antennas and the remaining ones are 3 transmit antennas. (b) shows the hardware of the FMCW signal generator.

objects. However both the value and direction of the velocity are unknown, the DOA calculated in this antenna model is relatively rough.

B. WiTrack

WiTrack is a wireless system which was proposed by Adib *et al.* in 2014 [1]. It belongs to the software-defined radio system defined in Section III-D. The main function of WiTrack is 3D motion tracking in both line-of-sight and through-wall scenarios. WiTrack leverages FMCW technique and its hardware is shown in Fig. 12(b). Besides, WiTrack can also detect the movement of body parts. Based on the above two functions, WiTrack can be applied to elderly fall detection and gesture-based appliance control.

WiTrack has 4 antennas: 1 transmit antenna and 3 receive antennas which are arranged in a "T" shape. Fig. 12(a) illustrates the setup of antennas. By analyzing the reflected signals from a human body, WiTrack calculates the round-trip distances from the transmit antenna to the body then back to each receive antennas. As known in Section IV-B, the 3D location of the body can be uniquely determined by 3 round-trip distances. Detailed steps are as follows:

The first step is to obtain the TOF. WiTrack exploits FMCW, as illustrated in Section V-A.2, to obtain TOF and then calculates each round-trip distances. Specifically, WiTrack transmits a periodic signal whose carrier frequency changes linearly with time. Then, by measuring the frequency shift in the received signal, WiTrack can obtain the TOF easily without high-speed ADCs which are expensive and high power. At last, round-trip distances can be calculated directly from TOFs.

The second step is to extract the information of the human body. WiTrack also needs to face the flash effect and the multi-path effect. This is because the objects in the room would reflect a huge amount of signal and reflected signal would bounce off other objects before reaching the receive antenna. Therefore, useful information is buried in these interference signals and WiTrack needs extra operations to extract the useful parts. To remove the reflections from all the static objects, WiTrack leverages the fact that the frequency shift of the signal reflected off static objects is constant over time. Therefore, by extracting the changing component, WiTrack can distinguish body (moving) reflections from static reflections and this process is called background

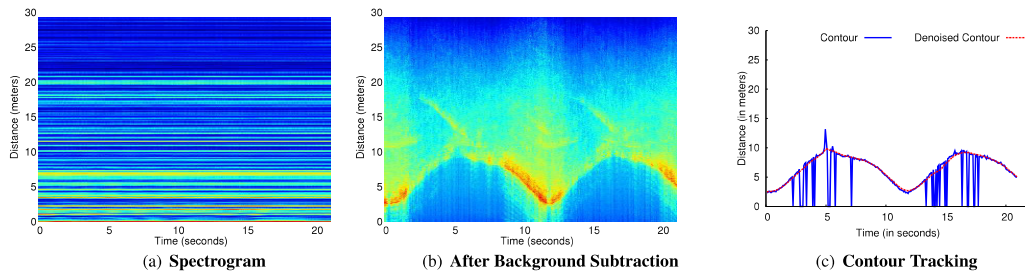


FIGURE 13. Distance extraction (figure from [1]). (a) shows the spectrogram of distance information. Different colors represent different power at the corresponding distance. (b) is the result of background subtraction. (c) shows the useful information extracted from (b) and the denoised result.

subtraction. Besides that, to address the multi-path effect, WiTrack extracts the valid path according to the assumption that the direct signal reflected from the human would have the smallest frequency shift (the distance is the shortest). The noise, however, would affect the judgment on the smallest frequency shift. WiTrack averages the data over five consecutive sweeps (the period of FMCW) to reduce the effects of noise. Because the noise is random and hence adds up incoherently, the useful information would become more significant after averaging. Then WiTrack finds the local maximum value with the shortest distance as a valid path. Finally, WiTrack leverages outlier rejection and Kalman Filter to obtain a better result. Fig. 13 illustrates this step. A noteworthy point is that, due to the background subtraction, WiTrack cannot track the person who stops moving. Therefore, WiTrack interpolates the latest location to supplement the missing data when a person stops moving.

The third step is the 3D localization. After getting the 3 round-trip distances between antennas and the body, WiTrack can localize the position in 3D space. Based on this 3D position, WiTrack has the ability to detect falls and the direction of a pointing hand.

1) EXPERIMENTAL EVALUATION

WiTrack’s median location error is 9.9 cm, 8.6 cm, and 17.7 cm along the x, y, and z dimensions in line-of-sight scenarios respectively. The median location error is 13.1 cm, 10.25 cm, and 21.0 cm along the x, y, and z dimensions in through-wall scenarios respectively.

2) LIMITATIONS

First, WiTrack can track only one person. Second, WiTrack cannot identify still people, i.e., in order to be detected, the user needs to move. Third, the detection of body parts is relatively coarse. Only the motions of large parts such as legs and arms can be tracked. Besides that, WiTrack cannot identify which part of the body has moved.

C. RF-CAPTURE

RF-Capture is an advanced wireless device which was proposed by Adib et al. in 2015 [9]. It belongs to the software-

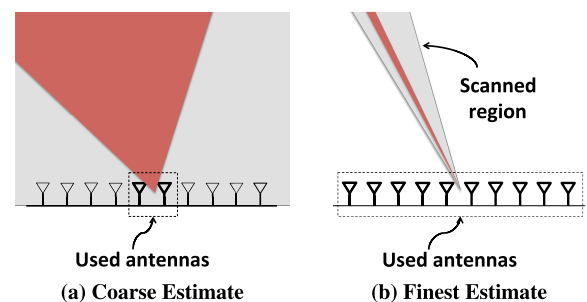


FIGURE 14. Coarse-to-fine angular scan (figure from [9]). (a) shows the coarse angle scan with a small number of antennas. (b) shows the finer scan with more antennas but within the limited region.

defined radio system defined in Section III-D. The main function is capturing the human figure through a wall. Specifically, RF-Capture can capture a coarse human skeleton with low-power RF signals (1/1000 the power of WiFi), classify different subjects, as well as identify and track body parts through a wall.

The antenna array of RF-Capture is “T” shaped and contains 4 transmit antennas and 16 receive antennas. The scan of the environment is realized by a special coarse-to-fine algorithm based on the FMCW technique and antenna array technique. Besides, the figure capture is achieved by several identification and classification algorithms. Two key components are as follows.

1) COARSE-TO-FINE 3D SCAN

To capture a figure of a human, RF-Capture needs to scan every voxel in the surrounding space. RF-Capture exploits the spherical coordinate (r, θ, ϕ) to determine each voxel. The depth r is obtained through FMCW and two angles θ, ϕ are obtained through the “T” shaped antenna array. Calculating the position of each voxel, however, would take a long time. Therefore, RF-Capture leverages a coarse-to-fine algorithm to reduce algorithmic complexity. The coarse-to-fine algorithm contains two components: coarse-to-fine angular scan and coarse-to-fine depth scan. Coarse-to-fine angular scan exploits the fact that the larger an array, the finer its spatial resolution. Hence, as shown in Fig. 14, RF-Capture starts with a small array to scan all the space with a relatively low resolution and then gradually adds more antennas to the

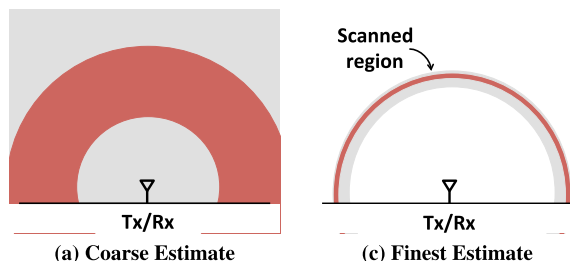


FIGURE 15. Coarse-to-fine depth scan (figure from [9]). (a) shows the coarse depth scan with a small chunk of bandwidth. (b) shows the finer scan with more bandwidth but within the limited region.

array to scan the region with high reflection power. In this case, the region with more objects would have a higher resolution, which can save lots of computing resources in empty regions. Similar to the angular scan, the coarse-to-fine depth scan changes the depth resolution by changing the bandwidth of FMCW. First, RF-Capture uses a small chunk of bandwidth to obtain a coarse region of the human body. Then, by adding more bandwidth, RF-Capture rescans the region of interest with higher resolution, as shown in Fig. 15.

2) MOTION-BASED FIGURE CAPTURE

One challenge RF-Capture has to face is the mirror effect, as discussed in Section II. To address this challenge, RF-Capture leverages consecutive reflection snapshots to reconstruct the whole human figure. The process contains four steps: Compensation for Depth, Compensation for Swaying, Body Part Segmentation, and Skeletal Stitching.

3) EXPERIMENTAL EVALUATION

In the 5 users' condition, the classification accuracy is 95.7%. And in the 15 users' condition, the classification accuracy is 88.2%. Besides, the through-wall body part identification accuracy is 99.13% when the user is 3 m away and 76.4% when the user is 8 m away. Finally, RF-Capture can track the palm of a user within a couple of centimeters.

4) LIMITATIONS

First, the algorithm of motion-based figure capture is based on the assumption that the user starts by walking towards the device. Hence, it doesn't work with other motions. Second, though the resolution of RF-Capture has a great leap comparing to other through-wall devices, it can just capture a relative coarse human skeleton. Besides that, some functions such as body part tracking can only be applied to certain motions at certain positions since not all the reflected signals can be detected by the device. Finally, because of the use of background subtraction, RF-Capture cannot detect the static object.

D. RF-POSE

RF-Pose is a wireless device which was proposed by Zhao *et al.* in 2018 [10]. It belongs to the software-defined

radio system defined in Section III-D. RF-Pose is a neural network system that reconstructs an accurate 2D human pose even in through-wall scenarios.

RF-Pose is relatively special because it follows a teacher-student design based on a deep neural network. RF-Pose has two parts: visual supervision and RF-based pose estimation, as shown in Fig. 16. The visual part has the ability to extract 2D pose through a camera, which provides cross-modal supervision for RF-based pose estimation. The RF-based part is used to detect reflections of RF signals and perform pose estimation which is trained by the visual part. Specifically, RF-based pose estimation leverages FMCW to obtain depth information with two antenna arrays: vertical and horizontal. Therefore, as shown in Fig. 17, RF-Pose generates two RF heatmaps and an RGB image (recorded at the same time) as input to the deep neural network. Then the teacher network predicts keypoint (parts of a human body) confidence map based on the image and exploits this keypoint map to supervise the student network. Once the student network is trained, RF-Pose can work only with RF signals.

One challenge RF-Pose has to face is the mirror effect, as discussed in Section II. Since the human body reflects RF signal like a mirror not a scatter, one snapshot cannot capture all parts of the human body. To address this challenge, RF-Pose packages a sequence of RF snapshots as the input to the neural network, since multiple snapshots contain more information about the human body and motion.

1) EXPERIMENTAL EVALUATION

In line-of-sight scenarios, the average precision (AP) of RF-Pose is 62.4 whereas the AP of the vision-based system is 68.8. In through-wall scenarios, the AP of RF-Pose is 58.1 whereas the vision-based system doesn't work.

2) LIMITATIONS

The RF signal of RF-Pose can traverse walls, however, it cannot traverse the human body. Therefore, inter-person occlusion is a limitation of RF-Pose.

E. OTHER SYSTEMS

There are some other handy through-wall systems with advanced or novel functions. All of these systems have features of low power, low bandwidth, and compaction. Besides, these systems do not need users to carry extra devices and can be applied to non-military use as well.

WiTrack2.0, proposed by Adib *et al.* in 2015, is an updated vision of WiTrack [5]. WiTrack2.0 belongs to the software-defined radio system defined in Section III-D. WiTrack2.0 has the ability to locate multiple people (up to five people) and even static ones. There are two challenges that WiTrack2.0 has to face: The first one is the significant multipath effect since the movement of multiple people would result in more reflections, which influence signals a lot. The second one is the near-far problem. Since the power of near reflections is much higher than that of distant reflections,

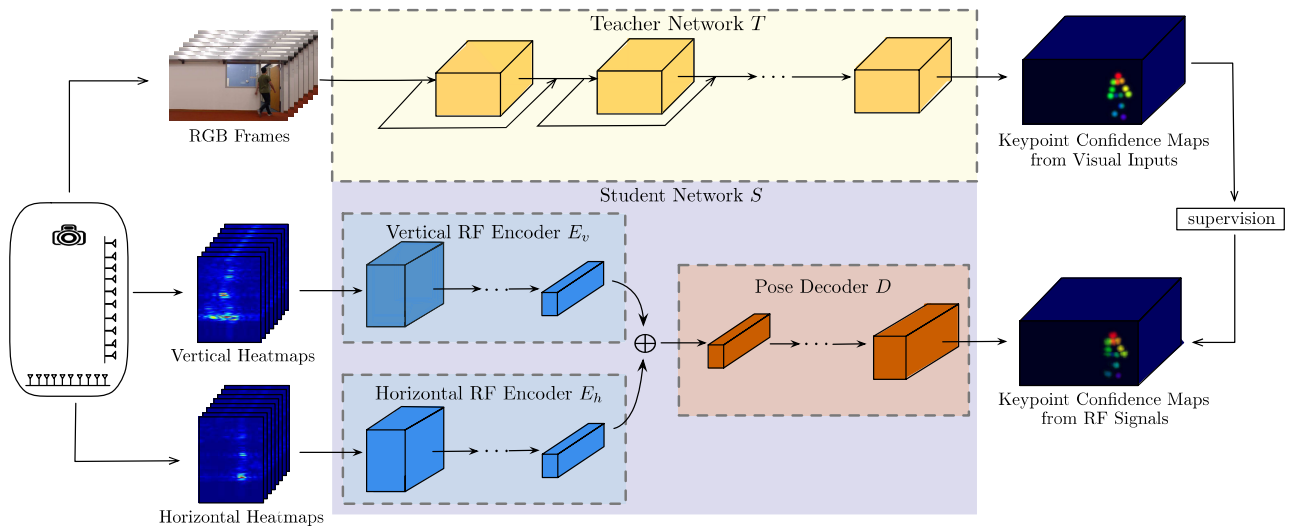


FIGURE 16. RF-Pose is based on a teacher-student network (figure from [10]).

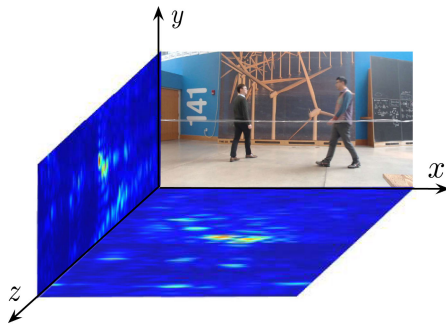


FIGURE 17. Input of RF-Pose's neural network (figure from [10]). Vertical and horizontal RF heatmaps and corresponding RGB image.

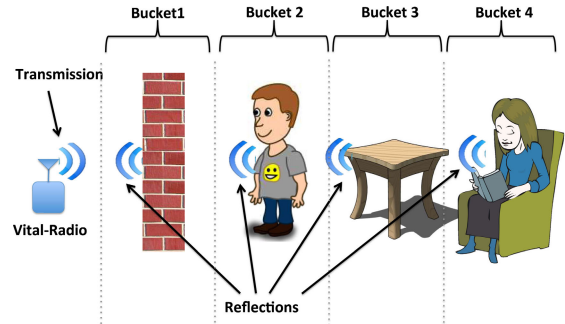


FIGURE 18. Vital-Radio leverages FMCW to separate reflectors into different buckets (figure from [2]).

it is difficult to detect further people. WiTrack2.0 leverages multi-shift FMCW to address the multipath effect and successive silhouette cancellation algorithm to face the near-far problem. Besides that, the detection of static people is based on their breathing, i.e., when a person is breathing, his/her chest moves within a sub-centimeter. Finally, WiTrack2.0 has a median accuracy of 11.7 cm in each of x and y dimensions.

Vital-Radio, proposed by Adib *et al.* in 2015, is a wireless monitoring device [2]. It belongs to the software-defined radio system defined in Section III-D. Vital-Radio can track breathing and heartbeats without physical contact even through a wall. Besides that, it can monitor the vital signs of multiple people simultaneously. The main idea of detection is to identify the minor movement of the human body caused by the breathing and heartbeat. There are three steps in the operation: First, Vital-Radio leverages FMCW to separate the space into different buckets depending on the distance from the device, as shown in Fig. 18. In this case, Vital-Radio can isolate reflections from different users and eliminate reflections off furniture and walls since they are in different buckets. Then, Vital-Radio identifies the

reflections involving the breathing and heartbeat by analyzing the periodicity in each bucket. Since the breathing and heartbeat are periodic, the reflections in the buckets where the dominant motion is breathing and heartbeat would be periodic. Finally, leveraging FFT and linear regression on phase, Vital-Radio can extract the breathing and heart rates from the received signals. In the experimental evaluation, Vital-Radio can track users' breathing and heart rates with a median accuracy of 99% when a person is 8 m from the device.

RF-Pose3D, proposed by Zhao *et al.* in 2018, is an updated vision of RF-Pose [47]. RF-Pose3D belongs to the software-defined radio system defined in Section III-D and is the first system that extracts full dynamic 3D skeletons of people (including the head, arms, shoulders, hip, legs, etc.) from RF signals. Besides, RF-Pose3D works with multiple people even in through-wall and occlusive scenarios. RF-Pose3D is based on CNN, which takes the RF signal captured by multi-antenna FMCW and 3D video of skeletons as inputs, as shown in Fig. 19. There are three components in the architecture: First, sensing the 3D skeleton. RF-Pose3D obtains the reflected signal and then leverages CNN to extract skeletons.

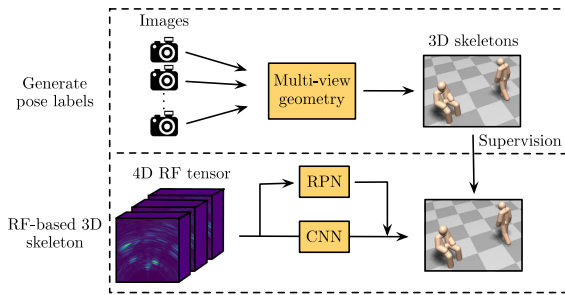


FIGURE 19. Architecture of RF-Pose3D (figure from [47]). The upper half shows the visual-based skeleton identification system which provides supervision for the RF-based system. The lower half is the RF-based system.

Second, scaling to multiple people. RF-Pose3D exploits a deep neural network to learn to detect people and focus on people. Third, training. The training data are obtained by a set of 2D visual-based skeleton identification systems. Since training data should be the 3D skeleton information, the team of RF-Pose3D develops a system containing 12 such 2D systems to obtain 3D training data. Once the network is trained, RF-Pose3D can extract 3D skeletons just from RF signals. In the experimental evaluation, the average errors of tracking each keypoint on the human body are 5.2 cm, 3.7 cm, and 4.7 cm along the x, y, and z dimensions, respectively (in through-wall scenarios).

Tadar, proposed by Yang *et al.* in 2015, is a through-wall system achieved by radio frequency identification (RFID) technique [46]. It belongs to the software-defined radio system defined in Section III-D. RFID is a popular technique in wireless identification fields. Typically, an RFID system is made of a reader and several tags. In the progress of obtaining information from one tag, the reader first transmits a continuous wave signal to activate a tag. The widely used tags are passive which means that there is no battery in such tags and they cannot transmit signals actively. Therefore, tags need to be activated by a reader, that is, tags leverage the energy of the received signal (from the reader) to respond to the reader and the identity information is carried by the responding signal, or called the backscatter signal. The design of Tadar is novel. Tadar contains one reader and 45 tags attached on the outer wall, constituting a virtual antenna array, as shown in Fig. 20. The transmitted signal from the reader would reflect off walls, furniture, human bodies, and other things inside the room. Thus tags would be activated by both the reflected signal and the signal directly from the reader and then generate the backscatter signal. By extracting the reflection of the human body from the backscatter signal, Tadar can obtain the location of the human body. One advantage of leveraging RFID is the simplicity of passive tags. Passive tags do not need batteries and other complicated structures so that the cost of each tag is low and tags are easy to be deployed. However, the disadvantage of Tadar is that the transmit power is higher than the systems we have discussed above by 3 orders of magnitude, as shown in Table 5. This is because the human body reflections are relatively weaker than other reflections

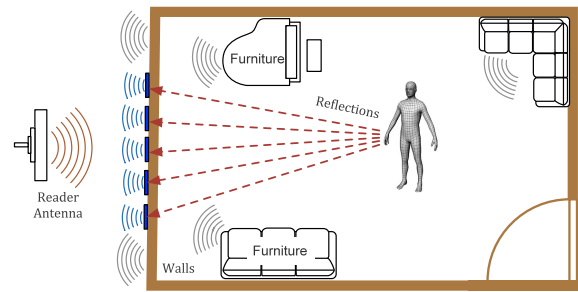


FIGURE 20. Design of Tadar (figure from [46]). Tadar contains 1 RFID reader and 45 tags attached on the outer wall. The tags serve as a virtual antenna array to receive the reflected signal from the room for the reader.

so that the backscatter signals from passive RFID tags activated by these weak human body reflections would be much weaker. Therefore, to capture such backscatter signals, higher transmit power is necessary. In the experimental evaluation, the median of the tracking error of Tadar is 7.8 cm and 20 cm along the x and y dimensions, respectively.

VII. CONCLUSION AND OPEN ISSUE

This paper surveys the handy through-wall techniques which have low power, narrow bandwidth, lightweight, without extra devices equipped on the human body and can work in non-military scenarios. The main functions of systems based on these techniques are identifying, locating, or imaging the human body behind a wall. By applying the radar technologies and artificial intelligence (AI) in off-the-shelf mobile wireless systems, through-wall systems are becoming more practical, accurate and efficient to be applied in civil applications.

Though through-wall systems have made a great achievement in research community, several open issues need to be solved in the future:

A. FEASIBILITY

Most handy through-wall systems are only tested in experimental environments and are still facing lots of problems when applied to real-world scenarios. Note that the real-world environment is much more complicated than the lab environment. For example, for through-wall systems used for human body localization, the presence of pets in the house may influence the performance of devices. The fast movement of pets would make a distortion of signals and interfere with the detection of human bodies. Besides that, through-wall systems also need to be tested in outdoor environments. Specifically, dense urban areas always have complicated electromagnetic interference. For example, the intensive coverage of WiFi signals in city centers is a huge challenge to WiFi-based systems. In this case, any WiFi device from shops or pedestrians is the interference to these systems. The ability of through-wall localization is also limited by the number of people. If there is a meeting in a room that contains dozens of people, most through-wall systems cannot work with such a massive density of people. How to further improve the

performance in such challenging scenarios deserves further study.

B. MOBILITY

While the paper investigates the handy through-wall system, the real-world portable through-wall systems is still challenging; most the through-wall devices need to be fixed at a certain position to make the signal processing much easier, and the bulky hardware devices (e.g., USRP) and real-time processing make it challenging to be portable. It is also challenging yet appealing to apply the through-wall technology in a highly mobile environment. For example, by equipping autonomous cars with through-wall devices, the cars can identify potential risks at the street corner that camera-based visual systems cannot work.

C. INTELLIGENT SIGNAL PROCESSING METHOD

Nowadays, AI has become a very powerful method to process signals. The combination of AI and through-wall technology could be the mainstream of future development. As several systems mentioned above [10], [47], [48], [52], AI has been proved to have the ability to extract fine-grained information of human bodies from coarse RF signals. How to combine AI and through-wall technology much closer is a challenge to future through-wall systems. Supervised machine learning is now widely used in through-wall scenarios to interpret complicated RF information. Supervised learning, however, requires a huge amount of training set which is scarce in the through-wall detection field. For example, in the absence of labeled RF data of walking people, today’s systems turn to the visual-based system for help. However, how to teach through-wall systems to recognize dancing people instead of walking ones, a new training set is required. Due to such challenges in supervised learning, unsupervised learning may be an alternative choice. Therefore, combining AI and through-wall technology effectively is a big deal for future research.

D. PRIVACY

The through-wall technology, if inappropriately used, would severely violate the privacy of people. Therefore, privacy-preserving technology is necessary before the wide civil adoption of the through wall system.

**APPENDIX A
DERIVATION OF DOA**

To make the task easier and reduce the difficulty of calculation, some assumptions are made as follows [55]:

- The propagation medium is homogeneous which means that the waves arriving at the array can be considered planar.
- The locations of every sensor (antennas array) are known and uniform.
- Sensors can be modeled as linear time-invariant (LTI) systems.

- The received signals are narrow-band and the sensor frequency response is flat over the pass-band.
- Sensors are identical and omnidirectional over the DOA range of interest which means the sensor frequency responses are independent of DOA.

Above all, we use the transmitted signal $x(t)$ at a reference point to describe the signal received by sensors(antenna array). The reference point can be one of the sensors or any other point placed near enough to the sensors. Assume that the number of sensors is m . Then the output of sensor k ($k = 1, 2, \dots, m$) can be written as

$$\bar{y}_k(t) = \bar{h}_k(t) * x(t - \tau_k) + \bar{e}_k(t), \tag{11}$$

where $\bar{h}_k(t)$ is the impulse response of the k th sensor, τ_k is the time that signals take to travel from the reference point to the k th sensor, and $\bar{e}_k(t)$ is additive noise. Similarly, the output of sensor k in the complex frequency domain is

$$\bar{Y}_k(\omega) = \bar{H}_k(\omega)X(\omega)e^{-j\omega\tau_k} + \bar{E}_k(\omega), \tag{12}$$

where $\bar{Y}_k(\omega)$, $\bar{H}_k(\omega)$, $X(\omega)$, and $\bar{E}_k(\omega)$ denote the Fourier transform of $\bar{y}_k(t)$, $\bar{h}_k(t)$, $x(t)$, and $\bar{e}_k(t)$, respectively.

Let $s(t)$ denote the baseband signal associated with $x(t)$. Then $X(\omega)$ can be expressed with $S(\omega)$ (Fourier transform of $s(t)$) through real modulation process:

$$X(\omega) = S(\omega - \omega_c) + S^*(-(\omega + \omega_c)), \tag{13}$$

where ω_c is the carrier frequency of the modulation. According to (13), (12) can be written as

$$\bar{Y}_k(\omega) = \bar{H}_k(\omega) [S(\omega - \omega_c) + S^*(-\omega - \omega_c)] e^{-j\omega\tau_k} + \bar{E}_k(\omega). \tag{14}$$

After the modulated signal ($\bar{Y}_k(\omega)$) has been received by a sensor, the next the procedure is demodulation. Let $\tilde{y}_k(t)$ and $\tilde{Y}_k(\omega)$ denote the demodulated signal. Then we can obtain the demodulated signal by translating in frequency to the left by ω_c :

$$\begin{aligned} \tilde{Y}_k(\omega) &= \bar{Y}_k(\omega + \omega_c) \\ &= \bar{H}_k(\omega + \omega_c) [S(\omega) + S^*(-\omega - 2\omega_c)] \\ &\quad \times e^{-j(\omega + \omega_c)\tau_k} + \bar{E}_k(\omega + \omega_c). \end{aligned} \tag{15}$$

The next step is to extract the baseband signal from the demodulated signal ($\tilde{y}_k(t)$) by passing through a lowpass filter whose bandwidth is equal to $s(t)$. Let $y_k(t)$ denote the output of the lowpass filter and $Y_k(\omega)$ is its Fourier transform. Then

$$\begin{aligned} Y_k(\omega) &= H_k(\omega + \omega_c)S(\omega)e^{-j(\omega + \omega_c)\tau_k} \\ &\quad + E_k(\omega + \omega_c) \end{aligned} \tag{16}$$

$$\begin{aligned} &\approx H_k(\omega_c)S(\omega)e^{-j\omega_c\tau_k} \\ &\quad + E_k(\omega + \omega_c), \end{aligned} \tag{17}$$

where $H_k(\omega)$ and $E_k(\omega)$ denote the parts of $\bar{H}_k(\omega)$ and $\bar{E}_k(\omega)$ within the band of the lowpass filter. The approximation in (17) is based on the above assumption that the received

signals are narrowband. Therefore, in time domain, (17) can be transformed into

$$y_k(t) = H_k(\omega_c)e^{-j\omega_c\tau_k}s(t) + e_k(t), \quad (18)$$

where $y_k(t)$ and $e_k(t)$ are the inverse Fourier transforms of $Y_k(\omega)$ and $E_k(\omega + \omega_c)$.

In order to obtain the DOA by detecting the signals, we need to obtain the relationship between $y_k(t)$ and $\text{DOA}(\theta)$. Assume that the reference point is sensor 1 (i.e., $\tau_1 = 0$) and let d denote the distance between sensors, as show in Fig. 7. Then τ_k and θ meet the following equation:

$$\tau_k = (k - 1)\frac{d \sin \theta}{c}. \quad (19)$$

Then (18) can be written as

$$y_k(t) = H_k(\omega_c)e^{-j\omega_c(k-1)\frac{d \sin \theta}{c}}s(t) + e_k(t). \quad (20)$$

Therefore, with m sensors, we will obtain a matrix equation:

$$y(t) = a(\theta)s(t) + e(t), \quad (21)$$

where

$$\begin{aligned} y(t) &= [y_1(t) \dots y_m(t)]^T \\ a(\theta) &= [H_1(\omega_c), H_2(\omega_c)e^{-j\omega_c\frac{d \sin \theta}{c}} \\ &\quad \dots H_m(\omega_c)e^{-j\omega_c(m-1)\frac{d \sin \theta}{c}}]^T \\ e(t) &= [e_1(t) \dots e_m(t)]^T. \end{aligned}$$

Leveraging (21) with known $y_k(t)$, $s(t)$, and $H_k(\omega_c)$ ($k = 1, \dots, m$), we can obtain θ (DOA) approximately.

REFERENCES

- [1] F. Adib, Z. Kabelac, D. Katabi, and R. C. Miller, "3D tracking via body radio reflections," in *Proc. 11th USENIX Conf. Netw. Syst. Design Implement.*, Seattle, WA, USA, Apr. 2014, pp. 317–329. [Online]. Available: <https://www.usenix.org/conference/nsdi14/technical-sessions/presentation/adib>
- [2] 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. Hum. Factors Comput. Syst. (CHI)*, New York, NY, USA, 2015, pp. 837–846, doi: [10.1145/2702123.2702200](https://doi.org/10.1145/2702123.2702200).
- [3] F. Adib and D. Katabi, "See through walls with WiFi!," in *Proc. ACM SIGCOMM Conf.*, New York, NY, USA, 2013, pp. 75–86, doi: [10.1145/2486001.2486039](https://doi.org/10.1145/2486001.2486039).
- [4] P. K. M. Nkwari, S. Sinha, and H. C. Ferreira, "Through-the-wall radar imaging: A review," *IETE Tech. Rev.*, vol. 35, no. 6, pp. 631–639, Nov. 2018, doi: [10.1080/02564602.2017.1364146](https://doi.org/10.1080/02564602.2017.1364146).
- [5] F. Adib, Z. Kabelac, and D. Katabi, "Multi-person localization via RF body reflections," in *Proc. 12th USENIX Symp. Netw. Syst. Design Implement. (NSDI)*, Oakland, CA, USA, May 2015, pp. 279–292. [Online]. Available: <https://www.usenix.org/conference/nsdi15/technical-sessions/presentation/adib>
- [6] X. Chen, "Multipath ghost elimination for through-wall radar imaging," *IET Radar, Sonar Navigat.*, vol. 10, no. 11, pp. 299–310, Feb. 2016. [Online]. Available: <https://digital-library.theiet.org/content/journals/10.1049/iet-rsn.2015.0171>
- [7] P. Setlur, M. Amin, and F. Ahmad, "Multipath model and exploitation in through-the-wall and urban radar sensing," *IEEE Trans. Geosci. Remote Sens.*, vol. 49, no. 10, pp. 4021–4034, Oct. 2011.
- [8] P. Beckmann and A. Spizzichino, *The Scattering of Electromagnetic Waves From Rough Surfaces*. SAO/NASA Astrophysics Data System, 1987. [Online]. Available: <https://ui.adsabs.harvard.edu/abs/1987ah.book...B>
- [9] F. Adib, C.-Y. Hsu, H. Mao, D. Katabi, and F. Durand, "Capturing the human figure through a wall," *ACM Trans. Graph.*, vol. 34, no. 6, pp. 1–13, Nov. 2015, doi: [10.1145/2816795.2818072](https://doi.org/10.1145/2816795.2818072).
- [10] M. Zhao, T. Li, M. A. Alsheikh, Y. Tian, H. Zhao, A. Torralba, and D. Katabi, "Through-wall human pose estimation using radio signals," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 7356–7365.
- [11] Q. Pu, S. Gupta, S. Gollakota, and S. Patel, "Whole-home gesture recognition using wireless signals," in *Proc. 19th Annu. Int. Conf. Mobile Comput. Netw. (MobiCom)*, 2013, pp. 27–38, doi: [10.1145/2500423.2500436](https://doi.org/10.1145/2500423.2500436).
- [12] H. Li, W. Yang, J. Wang, Y. Xu, and L. Huang, "WiFinger: Talk to your smart devices with finger-grained gesture," in *Proc. ACM Int. Joint Conf. Pervasive Ubiquitous Comput.*, New York, NY, USA, Sep. 2016, pp. 250–261, doi: [10.1145/2971648.2971738](https://doi.org/10.1145/2971648.2971738).
- [13] K. Ali, A. X. Liu, W. Wang, and M. Shahzad, "Keystroke recognition using WiFi signals," in *Proc. 21st Annu. Int. Conf. Mobile Comput. Netw. (MobiCom)*, New York, NY, USA, 2015, pp. 90–102, doi: [10.1145/2789168.2790109](https://doi.org/10.1145/2789168.2790109).
- [14] 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, Jan. 2011.
- [15] R. Solimene, T. Rajvanshi, G. Buonanno, and A. dell'Aversano, "Microwave imaging through an unknown wall by a MIMO configuration and SVD approach," in *Multimodal Sensing: Technologies and Applications*, vol. 11059, E. Stella, Ed. Munich, Germany: SPIE, Jun. 2019, pp. 11–18, doi: [10.1117/12.2527472](https://doi.org/10.1117/12.2527472).
- [16] S. Vishwakarma, V. Ummalaneni, M. S. Iqbal, A. Majumdar, and S. S. Ram, "Mitigation of through-wall interference in radar images using denoising autoencoders," in *Proc. IEEE Radar Conf. (RadarConf)*, Apr. 2018, pp. 1543–1548.
- [17] M. Seifeldin, A. Saeed, A. E. Kosba, A. El-Keyi, and M. Youssef, "Nuzzer: A large-scale device-free passive localization system for wireless environments," *IEEE Trans. Mobile Comput.*, vol. 12, no. 7, pp. 1321–1334, Jul. 2013.
- [18] K. Chintalapudi, A. Padmanabha Iyer, and V. N. Padmanabhan, "Indoor localization without the pain," in *Proc. 16th Annu. Int. Conf. Mobile Comput. Netw. (MobiCom)*, New York, NY, USA, 2010, pp. 173–184, doi: [10.1145/1859995.1860016](https://doi.org/10.1145/1859995.1860016).
- [19] H. Abdelnasser, M. Youssef, and K. A. Harras, "WiGest: A ubiquitous WiFi-based gesture recognition system," in *Proc. IEEE Conf. Comput. Commun. (INFOCOM)*, Apr. 2015, pp. 1472–1480.
- [20] C. R. Karanam and Y. Mostofi, "3D through-wall imaging with unmanned aerial vehicles using WiFi," in *Proc. 16th ACM/IEEE Int. Conf. Inf. Process. Sensor Netw. (IPSN)*, Apr. 2017, pp. 131–142.
- [21] S. Depatla, L. Buckland, and Y. Mostofi, "X-ray vision with only WiFi power measurements using ryto wave models," *IEEE Trans. Veh. Technol.*, vol. 64, no. 4, pp. 1376–1387, Apr. 2015.
- [22] 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. (MobiCom)*, New York, NY, USA, 2014, pp. 617–628, doi: [10.1145/2639108.2639143](https://doi.org/10.1145/2639108.2639143).
- [23] 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," in *Proc. ACM Interact. Mobile Wearable Ubiquitous Technol.*, vol. 3, no. 3, Sep. 2019, doi: [10.1145/3351279](https://doi.org/10.1145/3351279).
- [24] A. Hanif, M. Iqbal, and F. Munir, "WiSpy: through-wall movement sensing and person counting using commodity WiFi signals," in *Proc. IEEE SENSORS*, Oct. 2018, pp. 1–4.
- [25] M. Al-Qaness and F. Li, "WiGeR: WiFi-based gesture recognition system," *ISPRS Int. J. Geo-Inf.*, vol. 5, no. 6, p. 92, 2016. [Online]. Available: <https://www.mdpi.com/2220-9964/5/6/92>
- [26] K. Joshi, D. Bharadia, M. Kotaru, and S. Katti, "WiDeo: Fine-grained device-free motion tracing using RF backscatter," in *Proc. 12th USENIX Conf. Netw. Syst. Design Implement. (NSDI)*, 2015, pp. 189–204.
- [27] S. Nannuru, Y. Li, Y. Zeng, M. Coates, and B. Yang, "Radio-frequency tomography for passive indoor multitarget tracking," *IEEE Trans. Mobile Comput.*, vol. 12, no. 12, pp. 2322–2333, Dec. 2013.
- [28] T. Liu, X. Luo, and Z. Liang, "Enhanced sparse representation-based device-free localization with radio tomography networks," *J. Sensor Actuator Netw.*, vol. 7, no. 1, p. 7, 2018. [Online]. Available: <https://www.mdpi.com/2224-2708/7/1/7>

- [29] Y. Zheng and A. Men, "Through-wall tracking with radio tomography networks using foreground detection," in *Proc. IEEE Wireless Commun. Netw. Conf. (WCNC)*, Apr. 2012, pp. 3278–3283.
- [30] H. Yigitler, R. Jantti, O. Kallio, and N. Patwari, "Detector based radio tomographic imaging," *IEEE Trans. Mobile Comput.*, vol. 17, no. 1, pp. 58–71, Jan. 2018.
- [31] J. Wilson and N. Patwari, "A fade-level skew-laplace signal strength model for device-free localization with wireless networks," *IEEE Trans. Mobile Comput.*, vol. 11, no. 6, pp. 947–958, Jun. 2012.
- [32] Y. Xu, S. Wu, C. Chen, J. Chen, and G. Fang, "A novel method for automatic detection of trapped victims by ultrawideband radar," *IEEE Trans. Geosci. Remote Sens.*, vol. 50, no. 8, pp. 3132–3142, Aug. 2012.
- [33] T. S. Ralston, G. L. Charvat, and J. E. Peabody, "Real-time through-wall imaging using an ultrawideband multiple-input multiple-output (MIMO) phased array radar system," in *Proc. IEEE Int. Symp. Phased Array Syst. Technol.*, Oct. 2010, pp. 551–558.
- [34] T. Mu and Y. Song, "Through wall imaging based on the space-frequency time reversal," in *Proc. 14th IEEE Int. Conf. Signal Process. (ICSP)*, Aug. 2018, pp. 884–888.
- [35] X. Liang, H. Zhang, G. Fang, S. Ye, and T. A. Gulliver, "An improved algorithm for through-wall target detection using ultra-wideband impulse radar," *IEEE Access*, vol. 5, pp. 22101–22118, 2017.
- [36] F. Qi, F. Liang, M. Liu, H. Lv, P. Wang, H. Xue, and J. Wang, "Position-information-indexed classifier for improved through-wall detection and classification of human activities using UWB bio-radar," *IEEE Antennas Wireless Propag. Lett.*, vol. 18, no. 3, pp. 437–441, Mar. 2019.
- [37] Y. Song, J. Hu, N. Chu, T. Jin, J. Zhang, and Z. Zhou, "Building layout reconstruction in concealed human target sensing via UWB MIMO through-wall imaging radar," *IEEE Geosci. Remote Sens. Lett.*, vol. 15, no. 8, pp. 1199–1203, Aug. 2018.
- [38] S. Bahrami, "Through the wall imaging using ultrawideband frequency domain sampling method," *IETE J. Res.*, pp. 1–9, Oct. 2018, doi: [10.1080/03772063.2018.1536525](https://doi.org/10.1080/03772063.2018.1536525).
- [39] J. Li, Z. Zeng, J. Sun, and F. Liu, "Through-wall detection of human Being's movement by UWB radar," *IEEE Geosci. Remote Sens. Lett.*, vol. 9, no. 6, pp. 1079–1083, Nov. 2012.
- [40] Z. Zeng, J. Li, L. Huang, and F. Liu, "Improvement of target imaging quality by multi-polarization MIMO GPR," in *Proc. 14th Int. Conf. Ground Penetrating Radar (GPR)*, Jun. 2012, pp. 119–124.
- [41] K. Wang, Z. Zeng, and J. Sun, "Through-wall detection of the moving paths and vital signs of human beings," *IEEE Geosci. Remote Sens. Lett.*, vol. 16, no. 5, pp. 717–721, May 2019.
- [42] G. Gennarelli, G. Ludeno, and F. Soldovieri, "Real-time through-wall situation awareness using a microwave Doppler radar sensor," *Remote Sens.*, vol. 8, no. 8, p. 621, 2016. [Online]. Available: <https://www.mdpi.com/2072-4292/8/8/621>
- [43] A. Basu, M. P. Abegaonkar, and S. K. Koul, "Through the wall respiration rate detection of multiple human subjects using Hilbert vibrational decomposition," *Prog. Electromagn. Res.*, vol. 80, pp. 83–91, Mar. 2019.
- [44] F.-K. Wang, T.-S. Horng, K.-C. Peng, J.-K. Jau, J.-Y. Li, and C.-C. Chen, "Detection of concealed individuals based on their vital signs by using a see-through-wall imaging system with a self-injection-locked radar," *IEEE Trans. Microw. Theory Techn.*, vol. 61, no. 1, pp. 696–704, Jan. 2013.
- [45] Y.-C. Li, D. Oh, S. Kim, and J.-W. Chong, "Dual channel S-band frequency modulated continuous wave through-wall radar imaging," *Sensors*, vol. 18, no. 2, p. 311, 2018. [Online]. Available: <https://www.mdpi.com/1424-8220/18/1/311>
- [46] L. Yang, Q. Lin, X. Li, T. Liu, and Y. Liu, "See through walls with COTS RFID system!," in *Proc. 21st Annu. Int. Conf. Mobile Comput. Netw. (MobiCom)*, New York, NY, USA, 2015, pp. 487–499, doi: [10.1145/2789168.2790100](https://doi.org/10.1145/2789168.2790100).
- [47] M. Zhao, Y. Tian, H. Zhao, M. A. Alsheikh, T. Li, R. Hristov, Z. Kabelac, D. Katabi, and A. Torralba, "RF-based 3D skeletons," in *Proc. Conf. ACM Special Interest Group Data Commun.*, New York, NY, USA, Aug. 2018, pp. 267–281, doi: [10.1145/3230543.3230579](https://doi.org/10.1145/3230543.3230579).
- [48] M. Zhao, Y. Liu, A. Raghu, H. Zhao, T. Li, A. Torralba, and D. Katabi, "Through-wall human mesh recovery using radio signals," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2019, pp. 10113–10122.
- [49] C.-Y. Hsu, A. Ahuja, S. Yue, R. Hristov, Z. Kabelac, and D. Katabi, "Zero-effort in-home sleep and insomnia monitoring using radio signals," *ACM Interact., Mobile, Wearable Ubiquitous Technol.*, vol. 1, no. 3, pp. 1–18, Sep. 2017, doi: [10.1145/3130924](https://doi.org/10.1145/3130924).
- [50] D. De, W.-Z. Song, M. Xu, C.-L. Wang, D. Cook, and X. Huo, "FindingHuMo: Real-time tracking of motion trajectories from anonymous binary sensing in smart environments," in *Proc. IEEE 32nd Int. Conf. Distrib. Comput. Syst.*, Jun. 2012, pp. 163–172.
- [51] K. Kleisouris, B. Firner, R. Howard, Y. Zhang, and R. P. Martin, "Detecting intra-room mobility with signal strength descriptors," in *Proc. 11th ACM Int. Symp. Mobile Hoc Netw. Comput. (MobiHoc)*, New York, NY, USA, 2010, pp. 71–80, doi: [10.1145/1860093.1860104](https://doi.org/10.1145/1860093.1860104).
- [52] C.-Y. Hsu, R. Hristov, G.-H. Lee, M. Zhao, and D. Katabi, "Enabling identification and behavioral sensing in homes using radio reflections," in *Proc. CHI Conf. Hum. Factors Comput. Syst. (CHI)*, New York, NY, USA, 2019, doi: [10.1145/3290605.3300778](https://doi.org/10.1145/3290605.3300778).
- [53] A. M. Attiya, A. Bayram, A. Safaai-Jazi, and S. M. Riad, "UWB applications for through-wall detection," in *Proc. IEEE Antennas Propag. Soc. Symp.*, vol. 3, Jun. 2004, pp. 3079–3082.
- [54] A. Safaai-Jazi, S. M. Riad, A. Muqabel, and A. Bayram, "Ultra-wideband propagation measurements and channel modeling," in *Proc. Rep. Through-Wall Propag. Mater. Characterization*, 2002, pp. 1–75.
- [55] P. Stoica and R. L. Moses, *Spectral Analysis of Signals*. Upper Saddle River, NJ, USA: Prentice-Hall, 2005.
- [56] R. Schmidt, "Multiple emitter location and signal parameter estimation," *IEEE Trans. Antennas Propag.*, vol. AP-34, no. 3, pp. 276–280, Mar. 1986.
- [57] T.-J. Shan, M. Wax, and T. Kailath, "On spatial smoothing for direction-of-arrival estimation of coherent signals," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. ASSP-33, no. 4, pp. 806–811, Aug. 1985.
- [58] 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.*, New York, NY, USA, Sep. 2016, pp. 196–207, doi: [10.1145/2971648.2971665](https://doi.org/10.1145/2971648.2971665).
- [59] S. Ji, W. Xu, M. Yang, and K. Yu, "3D convolutional neural networks for human action recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 1, pp. 221–231, Jan. 2013.



KANGLE MU is currently pursuing the B.E. degree from the School of Telecommunication Engineering, Xidian University, China.

His research interests include wireless multimedia network, signal processing, and machine learning.

Mr. Mu was a recipient of the National Scholarship from Ministry of Education of PRC, and the First Prize of the Contemporary Undergraduate Mathematical Contest in Modeling.



TOM H. LUAN (Senior Member, IEEE) received the B.Sc. degree from Xi'an Jiaotong University, China, in 2004, the M.Phil. degree from the Hong Kong University of Science and Technology in 2007, and the Ph.D. degree from the University of Waterloo, in 2012.

He is currently with Xidian University, as a Professor. He has published around 80 technical articles in journals and conferences proceedings, including the IEEE TRANSACTIONS

ON MOBILE COMPUTING (TMC), the IEEE JOURNAL ON SELECTED AREAS IN COMMUNICATIONS (JSAC), the IEEE TRANSACTIONS ON VEHICULAR TECHNOLOGY (TVT), the *IEEE International Conference on Computer Communications (Infocom)*, the *IEEE International Symposium on Quality of Service (IWQoS)*, and *SECON*. His research mainly focuses on edge computing, vehicular networking, wireless content distribution, blockchain, and distributed machine learning. He has been awarded the 2017 IEEE VTS Best Land Transportation Paper Award and the 2018 IEEE ICCS Best Paper Award.



LINA ZHU received the B.E. degree from the Suzhou University of Science and Technology, China, in 2009, and the Ph.D. degree in communication and information system from Xidian University, China, in 2015.

She is currently a Lecturer with the State Key Laboratory of Integrated Services Networks, Xidian University. Her current research interests include mobility model, trust forwarding, routing, and MAC protocols in vehicular networks.



LIN X. CAI received the M.A.Sc. and Ph.D. degrees in electrical and computer engineering from the University of Waterloo, Waterloo, ON, Canada, in 2005 and 2010, respectively.

She was a Postdoctoral Research Fellow with the Electrical Engineering Department, Princeton University, in 2011, before she joined the Futurewei Wireless Research and Development Center, as a Senior Engineer, in 2012. She has been an Assistant Professor with the Department

of Electrical and Computer Engineering, Illinois Institute of Technology, Chicago, IL, USA, since August 2014. Her research interests include green communication and networking, broadband multimedia services, and radio resource and mobility management.

Dr. Cai received a Postdoctoral Fellowship Award from the Natural Sciences and Engineering Research Council of Canada (NSERC), in 2010, the Best Paper Award from the IEEE Globecom, in 2011, and the NSF Career Award, in 2016. She is also an Associated Editor of the IEEE TRANSACTION ON WIRELESS COMMUNICATIONS, the *IEEE Network Magazine*, and the Co-Chair of the IEEE conferences.



LONGXIANG GAO (Senior Member, IEEE) received the Ph.D. degree in computer science from Deakin University, Australia.

He was a Postdoctoral Research Fellow with IBM Research and Development, Australia. He is currently a Senior Lecturer and the Co-founder of the Deakin Blockchain Innovation Laboratory, School of Information Technology, Deakin University. He has more than 50 publications, including patent, monograph, book chapter, journal, and

conference papers. Some of his publications have been published in the top venue, such as the IEEE TRANSACTIONS ON MOBILE COMPUTING (TMC), the IEEE INTERNET OF THINGS JOURNAL (IoTJ), the IEEE TRANSACTIONS ON DEPENDABLE AND SECURE COMPUTING (TDSC), the IEEE TRANSACTIONS ON VEHICULAR TECHNOLOGY (TVT), the IEEE TRANSACTIONS ON COMPUTATIONAL SOCIAL SYSTEMS (TCSS), and the IEEE TRANSACTIONS ON NETWORK SCIENCE AND ENGINEERING (TNSE). His research interests include fog/edge computing, blockchain, data analysis, and privacy protection. He has being the Chief Investigator (CI) for more than 20 research projects (the total awarded amount is \$4.6million), from pure research project to contracted industry research.

Dr. Gao is an active in the IEEE Communication Society. He has served as the TPC Co-Chair, the Publicity Co-Chair, the Organization Chair, and TPC Member for many international conferences. He is also an Associate Editor of IEEE ACCESS and the Assessor of Australian Research Council (ARC) projects. Meanwhile, he has served as the President of Chinese Association of Professionals and Scholars, Australia, which is the premier organization promoting and facilitating education, industry-relevant research and business opportunities between Australia and China with more than 2000 members of Chinese professionals, Ph.D. students, and young scholars living and working in Australia. He is currently serving as Executive President of advisory committee, Chinese Association of Professionals and Scholars, Australia.

• • •