Chapter 3
Contactless Human Activity Analysis: An Overview of Different Modalities

Farhan Fuad Abir, Md. Ahasan Atick Faisal, Omar Shahid, and Mosabber Uddin Ahmed

Abstract Human Activity Analysis (HAA) is a prominent research field in this modern era which has enlightened us with the opportunities of monitoring regular activities or the surrounding environment as per our desire. In recent times, Contactless Human Activity Analysis (CHAA) has added a new dimension in this domain as these systems perform without any wearable device or any kind of physical contact with the user. We have analyzed different modalities of CHAA and arranged them into three major categories: RF-based, sound-based, and vision-based modalities. In this chapter, we have presented state-of-the-art modalities, frequently faced challenges with some probable solutions, and currently used applications of CHAA with future directions.

3.1 Introduction

Human Activity Analysis (HAA) is the field of research concerning the interpretation of different human activities, e.g., human motion, hand gesture, audio signal from a sequence of sensed data and their neighboring environment. Over the last few years, HAA has been one of the most prominent topics in multiple research fields including signal processing, machine learning, deep learning, computer vision, mobile

F. F. Abir · Md. A. A. Faisal · M. U. Ahmed
Department of Electrical and Electronic Engineering, University of Dhaka, Dhaka, Bangladesh
e-mail: farhanfuad.abir@gmail.com

Md. A. A. Faisal
e-mail: atickfaisal@gmail.com

M. U. Ahmed
e-mail: mosabber.ahmed@du.ac.bd

O. Shahid
Department of Robotics and Mechatronics Engineering, University of Dhaka,
Dhaka, Bangladesh
e-mail: omarshahid232@gmail.com

© Springer Nature Switzerland AG 2021
M. A. R. Ahad et al. (eds.), Contactless Human Activity Analysis,
Intelligent Systems Reference Library 200,
https://doi.org/10.1007/978-3-030-68590-4_3
computing, etc. With the development of this field, new methods of data collection have been rigorously explored everyday varying from body-worn sensors to contactless sensing. However, activity analysis based on body-attached devices sometimes becomes incompatible, even infeasible for implementation. On the contrary, with the comprehensive development of computing devices, miniaturized sensor system, and improved accuracy of wired and wireless communication networks, Contactless Human Activity Analysis (CHAA) has been prioritized and improved in different aspects.

Contactless Human Activity Analysis (CHAA) is deemed to analyze body posture or any type of physical activity of humans by means of non-wearable or contactless devices. This approach can be employed in everyday environments like marketplaces, cafes because each individual does not need separate devices. Besides, CHAA is implementable from a distance to the individual. Furthermore, CHAA is a more financially feasible way of analysis on a mass scale because only one module can analyze thousands of people in a single session whereas the wearable methods would require each individual to possess wearable hardware [1].

In recent years, there has been significant development in high-performance sensors, physiological signal detection, behavior analysis, gesture recognition systems which comfortingly leading CHAA as an impactful research sector in this domain. As wireless signal transmitter and receiver have both become very precise and gained a higher rate of accuracy, WiFi signal, RF signal, sound-based techniques are used as a powerful tool for contactless activity analysis systems. In addition to that, RGB cameras or other types of cameras like infrared cameras, depth cameras have become affordable. Therefore, vision-based activity analysis is rendering a great contribution to CHAA.

This chapter presents an introduction to the CHAA modalities. Previously, Ma et al. [2] have provided a comprehensive survey on WiFi-based activity analysis. In another work, Wang et al. [3] had reviewed a wide range of works on smartphone ultrasonic-based activity analysis. That work had stated different signal processing techniques, feature extraction methods, and various types of applications related to this field. But our study is not confined within a specific modality. We have explored the state-of-the-art approaches including the early modalities as well as the evaluation of these approaches. Therefore, this chapter will be beneficial to new researchers in this field.

The rest of this chapter is organized as follows: Sect. 3.2 has explored a detailed evolution of the contactless activity analysis system. Different modalities and techniques related to CHAA have been elaborately described in Sect. 3.3. Section 3.4 has stated various challenges and some probable solutions relevant to this field. Section 3.5 has summarized the applications of this domain. And finally, we have drawn a conclusion in Sect. 3.6.
3.2 Historical Overview

Due to the development of microelectronics and computer systems in the past two decades, there has been a massive increase in research in the field of Human Activity Recognition (HAR). Earlier works on HAR were mostly contact-based or in other words, utilized some kind of wearable sensors. In the late 90s, Foerster proposed the very first work on HAR using accelerometer sensors [4]. Although wearable sensor-based approaches are still popular these days, contactless approaches have gained significant attention in recent years.

RF-based sensing can be dated back to the 1930s when the first radar system was developed [5]. Although these systems were designed primarily to detect and track large moving objects at large distances like aircraft, they laid the foundation for the modern RF-based activity recognition methods. The use of electromagnetic waves in HAR gained attention because of some big advantages over vision-based systems; for example, no line of sight requirement, ability to work in any kind of lighting environment, and ability to pass through obstacles. Earlier, Doppler radars were developed to detect and track humans through obstacles [6]. WiFi RSSI-based indoor localization of human subjects was introduced in 2000 [7]. The capabilities of RSSI-based systems were limited due to noise and low sensitivity. To tackle this problem, many researchers utilized Universal Software Radio Peripheral (USRP) in the following years to perform signal processing for tasks like tracking the 3D movement of human subjects [8] or monitoring heart rate [9]. Activity recognition using commercial WiFi devices came to interest after Channel State Information (CSI) was introduced by Halperin et al. [10]. Although these low-frequency WiFi signal-based methods were very successful for localization and tracking, they lacked the sensitivity to detect small movements required for some activities and gesture recognition tasks. In 2002, the discovery of Micro-Doppler Signatures in Doppler radars opened a new dimension in human gait analysis [11]. Using these micro-Doppler features, it was possible to detect human body movements more precisely.

In 2016, Google developed a miniature millimeter wave radar called Soli that was capable of detecting hand gestures at submillimeter level accuracy [12].

The first ultrasound-based technology was invented by American naval architect Lewis Nixon in 1906 for detecting icebergs [13]. After proving its usability in obstacle detection, a French physicist Paul Langevin invented an ultrasound-based technology in 1917 for detecting underwater submarines [14]. After years of research and fine-tuning, the usability of ultrasound in short distance ranging was starting to come to light. In 1987, Elfes et al. [15] published their work on a mapping and navigation technique based on ultrasound technology. This was used in the navigation system of an autonomous mobile robot named Dolphin. Ultrasound was mostly used in ranging and obstacle detection back then. But after 2000, researchers tend to focus more on using the Doppler shift for activity recognition. But until this point, the research was based on standalone ultrasound devices.

The main breakthrough in sound-based activity recognition was the use of smartphones as the detecting device. This enabled the widespread use of ultrasound-based activity recognition modalities. In 2007, Peng et al. [16] presented a high accuracy
ultrasound-based ranging system using commercial off-the-shelf (COTS) mobile phones. Later in 2010, Filonenko et al. [17] explored the potential of COTS mobile phones in generating reliable ultrasound signals. As a result, in the last few years, smartphone-based active activity recognition methods as well as the passive ones have made breakthroughs.

Apart from this, the vision-based approach has become very popular in this field. In the early days of the twenty first century, the advancement of computational resources gave a revolutionary shed to the researches related to image processing. By this time, object recognition and scene analysis provided results with higher accuracy due to the development of sophisticated learning algorithms. At the beginning of vision-based activity recognition, it was still image-based [18–20]. In this case, some activities were recorded using a camera and split into frames to make datasets and later on used some handcrafted feature extraction techniques to feed them into supervised learning algorithms.

By the end of the first decade of the twenty first century, video-based activity recognition has come to light. From the viewpoint of datatypes, research on video based human activity analysis can be classified based on method of color (RGB) data [21, 22], depth data, and combination of color (RGB) and depth data (RGBD) [23, 24]. In the early period of machine learning approaches, these data were used to get handcrafted features and fed into a learning algorithm like decision tree, support vector machine, hidden Markov model, etc. There have been many proposed features for RGB data, for example, joint trajectory features [25], spatiotemporal interest point features [26], and spatiotemporal volume-based features [27]. On the contrary, depth sensors provided more stable data according to the background and the environment, therefore, it enabled real-time detection faster with pose estimation. During the last decade, innovation of the Kinect sensor gave a new insight for depth data and skeleton tracking.

However, deep learning methods work well on video-based activity analysis, unlike machine learning approaches. These methods learn features automatically from images which is more robust and convenient for human activity analysis. Deep learning networks can learn features from single-mode data as well as multimodal data. Moreover, pose estimation methods to learn skeleton features from a scene by applying deep learning methods have drawn increased attention for vision-based activity analysis.

### 3.3 Primary Techniques

Researches on Contactless Human Activity Analysis (CHAA) have been going on for quite a while now and different methods have been developed over the years. Some methods are more focused on specific goals and some are more generalized. These methods can be classified into video-based, RF-based, and ultrasonic-based approaches [2].
3.3.1 RF-Based Approaches

Wireless signals have been used quite extensively for localization, tracking, and surveillance. RF-based approaches have evolved a lot over the years and recently have made their impact in the field of Contactless Human Activity Analysis (CHAA). At a high level, these techniques rely on a radio frequency transmitter and a receiver. Wireless signals emitted from the transmitter get reflected by the surrounding environment and subject of interest and reach the receiver. Useful features are then extracted from the received signal using different tactics and fed into classification algorithms to classify different kinds of human activity.

Some RF-based approaches do not need additional device. These approaches rely on physical (PHY) layer properties like the Received Signal Strength Indicator (RSSI) and the Channel State Information (CSI) of the WiFi. These measurements can be easily extracted from commercially available WiFi Network Interface Cards (NIC). The Intel 5300 [10] and the Atheors 9580 [28] are two examples of such NICs. Some RF-based frameworks require custom hardware setup such as the Universal Software Radio Peripheral (USRP) to work with the Frequency Modulated Carrier Wave (FMCW) signals [8]. Other methods use a similar setup to measure the Doppler shift of the RF signal [29]. Apart from these, many other signal properties can be extracted from both commodity and customized hardware-based devices. But the most common signal properties that can be used in human activity analysis are signal power (RSSI), channel information (CSI), Doppler shift, and time of flight (ToF). These properties are discussed below in detail.

3.3.1.1 Received Signal Strength Indicator (RSSI)

Wireless signals are electromagnetic waves that propagate through the air at the speed of light. As the signal propagates further from the transmitter, its power decays exponentially with the distance. This power distance relationship can be represented as [30]:

$$P_r = \frac{P_t G_t G_r \lambda^2}{(4\pi d)^\gamma},$$

where $P_r$ is the received power at distance $d$, $P_t$ is the transmitted power, $G_t$ and $G_r$ are transmitting and receiving antenna gain respectively, $\lambda$ is the wavelength, and $\gamma$ is the environmental attenuation factor. This formula is the basis of power-based ranging using wireless signals. RSSI is simply the received power and in terms of received voltage $V_r$, it is denoted as [31]:

$$RSSI (dB) = 20log(|V_r|^2)$$

Because of its availability in mainstream wireless signal measurements, RSSI has been adopted by numerous localization and activity recognition methods. But characterizing the signal strength by the above equation does not work in a real-
Fig. 3.1 Effects of shadowing and multipath on RSSI

Life scenario. In practical cases, the received signal is a superposition of signals taking multiple reflection paths (multipath effect) and also deviates from the expected value due to obstacles (shadowing effect) as shown in Fig. 3.1. Seidel et al. took the shadowing effect into consideration and proposed the Log-normal Distance Path Loss (LDPL) model which can be written as [32]:

$$PL(d)_{dB} = PL(d_0)_{dB} + 10\gamma log\left(\frac{d}{d_0}\right) + X_{\sigma},$$

where $PL(d)$ denotes the path loss at distance $d$, $P(d_0)$ denotes the path loss at distance $d_0$, $\gamma$ is the path loss exponent and $X_{\sigma}$ reflects the effect of shadowing.

It has been observed that the RSSI value changes dramatically in the presence of a human subject and the movement of the subject results in fluctuations of the RSSI value [33]. Useful features can be extracted from these RSSI fluctuations and used for activity recognition. Although RSSI measurement has been adopted in many RF-based activity recognition frameworks, it is unstable even in indoor conditions [34].

3.3.1.2 Channel State Information (CSI)

Wireless signals can take multiple reflection paths while traveling through the channel. Since the human body is a good reflector of wireless signals, the presence of a human subject adds more reflection paths to the environment. In the case of a moving subject, at a particular instance, some of these paths might interfere constructively or destructively. This creates fluctuations in RSSI value. The biggest drawback of RSSI-based measurement is its inability to capture this multipath effect. It gives a summed up view of the situation and fails to capture the bigger picture.

Adopted in IEEE 802.11 a/g/n, Channel State Information (CSI) captures the effect of multiple path propagation and presents us with a broader view of the wireless channel. To take the multipath effect into consideration, the wireless channel can be considered as a temporal linear filter. This filter is called Channel Impulse Response (CIR) and it is related to transmitted and received signal as follows [35]:
\[ y(t) = x(t) * h(t), \]

where \( x(t) \) is the transmitted signal, \( y(t) \) is the received signal, and \( h(t) \) is the Channel Impulse Response (CIR). The Channel Frequency Response (CFR) is simply the Fourier Transform of CIR and it is given by the ratio of received and transmitted signal in the frequency domain. The CSI monitored by WiFi NICs is characterized by the CFR value of the wireless channel. Complex valued CFR can be expressed in terms of amplitude and phase as follows [36]:

\[ H(f, t) = \sum_{i=1}^{N} a_i(f, t) e^{-j2\pi f \tau_i(t)}, \]

where \( a_i(f, t) \) represents the attenuation and initial phase shift of the \( i\)th path, and \( e^{-j2\pi f \tau_i(t)} \) is the phase shift that is related to the \( i\)th path with a propagation delay of \( \tau_i(t) \).

Using the above equation, it is possible to obtain the phase shift of a particular path taken by the wireless signal and perform activity analysis based on that. But, unfortunately, due to the mismatch of hardware between the transmitter and the receiver device, there will always be some non-negligible Carrier Frequency Offset (CFO). This CFO can result in a phase shift as large as \( 50\pi \) according to the IEEE 802.11n standards which will overshadow the phase shift resulting from the body movement [37].

Wang et al. [37] proposed a CSI-Speed model that leverages CSI power to resolve the effect of CFO. This model considers the CFR as a combination of static and dynamic components:

\[ H(f, t) = e^{-j2\pi \Delta f t} (H_s(f) + \sum_{i \in P_d} a_i(f, t) e^{-j2\pi \frac{d_i(t)}{\lambda}}), \]

where \( H_s(f) \) is the static CFR component, \( d_i(t) \) is the path length of the \( i\)th path, \( \lambda \) is the wavelength, and \( e^{-j2\pi \Delta f t} \) represents the phase shift for frequency offset \( \Delta f \).

According to the CSI-Speed model, CFR power changes when the subject of interest is in motion. The instantaneous CFR power can be expressed as follows:

\[
|H(f, t)|^2 = \sum_{i \in P_d} 2|H_s(f)a_i(f, t)| \cos \left( \frac{2\pi v_i t}{\lambda} + \frac{2\pi d_i(0)}{\lambda} + \phi_{si} \right) \\
+ \sum_{i, k \in P_d} 2|a_i(f, t)a_j(f, t)| \cos \left( \frac{2\pi (v_i - v_j)t}{\lambda} + \frac{2\pi d_i(0) - d_j(0)}{\lambda} + \phi_{ik} \right) \\
+ \sum_{i \in P_d} |a_i(f, t)|^2 + |H_s(f)|^2,
\]
where $\frac{2\pi d_i(0)}{\lambda} + \phi_{si}$ and $\frac{2\pi d_i(0) - d_i(0)}{\lambda} + \phi_{ik}$ are constant values that represent the initial phase offsets. This equation shows that the total CFR power is a combination of a static offset and a set of sinusoidal oscillations. The frequencies of these oscillations are directly related to the speed of movements that create changes in the path length. Thus, the CSI speed model provides a quantitative way of relating CSI power with human activity.

### 3.3.1.3 Doppler Shift

Doppler shift is characterized by the frequency shift of wireless signals when the transmitter and the receiver are moving relative to each other. Doppler radars work on this principle of Doppler shift to track a moving object. This technique has been adopted by many security surveillance, object tracking, and navigation systems. Recently Doppler radars have caught the attention of many researchers for the purpose of human detection and activity recognition. Since the human body reflects wireless signals, it can be thought of as a wireless signal transmitter [38]. Thus, human activities that involve motion, create Doppler shifts in the wireless signals. By measuring these frequency shifts, it is possible to detect the activity or gesture. When the subject is moving towards the receiver, the resulting Doppler shift is positive and when the subject is moving away from the receiver, the resulting Doppler shift is negative. More generally, when a point object is moving at an angle of $\phi$ from the receiver with velocity $v$, the resulting Doppler shift is expressed as [39]:

$$\Delta f = \frac{2vcos(\phi)}{c} f,$$

where $f$ is the carrier frequency of the wireless signal and $c$ is the speed of light. From this equation, we can see that a human activity involving a motion of 0.5 m/sec will create a maximum frequency shift of 17 Hz for a 5 GHz wireless signal which is really small compared to the typical wireless signal (e.g. WiFi) bandwidth of 20 MHz. WiSee [38], a whole-home gesture recognition system, solves this issue by splitting the received signal into narrowband pulses of only a few Hertz.

Human activity recognition requires not only capturing the whole-body motion but also capturing the limb motion. It has been observed that limb motion adds Micro-Doppler-Signatures in the received signal [11]. These signatures are considered to be the key feature for activity and gesture recognition using a Doppler radar-based framework. Figure 3.2 shows a hypothetical example of a Doppler spectrogram resulting from walking activity. In the figure, we can differentiate the Micro-Doppler signatures due to limb motion apart from the Doppler shift due to body motion. Useful features can be handcrafted from these spectrograms and used in classical machine learning algorithms [29] or they can be used directly as the input for Deep Convolutional Neural Networks (DCNN) [40] for activity classification.

Higher frequency signals produce a higher Doppler shift which is essential for capturing small movements. Google’s project Soli [12] leverages a 60GHz radar to
achieve high temporal resolution and classify hand gestures with submillimeter accuracy. This method is only useful when the subject of interest is in motion. Because, if the subject is stationary, there will be no Doppler shift.

3.3.1.4 Time of Flight (ToF)

Time of Flight (ToF) refers to the time it takes for the transmitted signal to reach the receiver. ToF provides a way to measure the distance of the subject from the sensing device; thus, can be a useful measurement for human activity analysis. Since the wireless signals propagate at the speed of light, it is very difficult to measure the ToF directly. Frequency Modulated Carrier Wave (FMCW) radar sweeps the carrier frequency in a sawtooth wave fashion. The reflected wave captures the ToF in the form of frequency shift of the carrier signal [41].

\[
\Delta t = \frac{\Delta f}{m},
\]

where \( m \) is the slope of the frequency sweep. Unlike direct measurement, we can easily measure the \( \Delta f \) from FMCW radar and thus calculate the ToF [42]. Figure 3.3 demonstrates this property of FMCW operation.

A number of activity recognition methods have been developed by utilizing the FMCW concept. WiTrack [8] is a 3D human tracking and activity recognition framework which leverages the FMCW technique. Implemented on USRP hardware, WiTrack can track a human subject with centimeter-level accuracy. Vital-Radio [9] utilizes FMCW-based distance measurement to monitor breathing and heart rate wirelessly.
All the signal properties for human activity analysis discussed above, come with their own set of advantages and drawbacks. Although RSSI-based frameworks are the simplest, they suffer from instability and low accuracy. CSI provides more detailed information about human activity but processing the CSI stream can be difficult. Doppler and FMCW-based frameworks require custom hardware and cannot be implemented using Commercial-Off-The-Shelf (COTS) devices. The choice of frequency also plays a vital role in RF-based systems. Higher frequency signals provide better sensitivity but their range is small compared to low-frequency signals. A robust RF-based human activity recognition framework should take these matters into consideration.

3.3.2 Sound-Based Approaches

Human Activity Recognition (HAR) using sound signal has recently been explored quite extensively along with the other approaches. Based on the range of usable sound frequency for HAR, we have divided the study into two sub groups—ultrasound signal-based and audible signal-based approaches. Though the ultrasound range is widely used for activity analysis and recognition, study of the audible sound in this field has recently become a topic of interest. Based on the necessary components, we can divide the sound-based approaches of contactless human activity analysis into two major categories—active detection method and passive detection method. The active method consists of a sound generator and a receiver while the passive method consists of the receiver only. In most applications, ultrasound-based active methods are used because it does not hamper the day to day activities of humans. On the other hand, audible range is mostly used with the passive methods.

Sound-based human activity analysis can be accomplished with different modalities and implementation techniques. The choice of modality depends mostly on the application and the hardware. Different modalities along with their applications are discussed in the following part.
3.3.2.1 Time of Flight (ToF)

ToF is a very trivial technique for ultrasound-based ranging. In last decade, a good number of researchers have explored the potential of this method for detecting small changes in distance of an object up to millimeters. The increased accuracy in ranging has opened new dimensions for activity recognition research.

The basic principle of ToF is similar in both RF-based ToF and sound-based ToF but there are some differences in the measurement and analysis techniques. Compared to the RF-based ToF, sound-based ToF measurement is more direct. The transmitter transmits ultrasound pulses which get reflected back by a nearby human in the line of sight of the signal propagation. The propagation velocity of sound wave is very low compared to that of the RF signal. Hence, in contrast to the RF-based ToF, the processing unit associated with the receiver in sound-based ToF, computes the ToF directly by comparing the transmitted and received signal. This process is based on the following equation [43]:

\[ D = \frac{1}{2} v_s \Delta t, \]

where \( D, \Delta t \) and \( v_s \) denotes the total distance, the time of flight (ToF), and the propagation velocity of sound respectively. Small change in distance occurs during any physical activity. Based on the small change, a certain pattern of the signal can be found associated with a certain activity.

Different features can be extracted from the ToF signal like Coefficient of Variation (COV) [44, 45], Range [45] and Linear Regression Zero-crossing Quotient (LRZQ) [44]. Afterward, the activities can be classified using classification models. Griffith et al. [45] used Support Vector Machine (SVM) and Biswas et al. [44] compared Nave Bayes, Logistic Regression and Sequential Minimal Optimizer (SMO) to classify human locomotor activities. Al-Naji et al. [43] explored a different domain where they used ToF to measure the displacement of thorax during inhale and exhale motion. Based on that, breathing rate (BR) was measured. Afterwards, comparing the BR of healthy subjects and respiratory patients, they identified different respiratory problems. Researchers have used this technique in a wide range of contactless human activity analysis applications.

3.3.2.2 Doppler Effect

Doppler shift is another well-known activity analysis technique which can is used both in RF-based and sound-based modalities. In both cases, the principle is the same—frequency shift due to the movement direction and velocity of an object relative to the transmitter. The differences lie in the type of required-hardware, analysis techniques and applications. With the development of sophisticated Doppler sensors and efficient processing components, nowadays very small change in frequency can
be detected. Which enabled this modality to analyze human activities ranging from walking pattern to the smallest finger movement.

Like ToF, the sensing part of the system consists of a transmitter and a receiver. In some cases, the time domain signal is frequency demodulated before the frequency domain conversion [46–49]. Most researchers have used Fast Fourier Transform (FFT) for this purpose. Moreover, Discrete Cosine Transform (DCT) [48, 49] is also used which only consists of the real components unlike the Fourier Transform. Kalgaonkar et al. [47] used Goertzel algorithm to get the energy information from a narrow bandwidth (50–120 Hz) which is more efficient than FFT for narrow bandwidth. Spectrogram is a very useful tool to visualize the Doppler effect in frequency domain and analyze the correlation between the activities and the frequency information. Figure 3.4 shows a sample spectrogram with the frequency shift due to walking.

The raw frequency domain signal is very difficult to work with and needs to be processed further. Firstly, a calibration method is often used in order to adjust the frequency response of the device. In case of two or more transmitters, the frequency information can be transformed into Euclidean space to get a velocity vector using the angles of transmitters [50]. A few challenges in this step are attenuating the noise, eliminating the multi-path echo, and adjusting device diversity. For these purposes, FFT normalization [51–53], Squared Continuous Frame Subtraction [52], Gaussian Smoothing [52], and Low Pass Filter [46–49] are commonly used. The preprocessed frequency domain signal is then segmented using different windowing techniques. 512-point [49], 1024-point [46–48], and 2048-point [52, 54] Hamming windows are used in most studies. The increased point per window provides more accuracy but takes more time to process. Features are extracted subsequently from each frequency
bin for classification. Some common features are direction, duration, average speed, spatial range [52, 54], and Power Spectral Density (PSD) [49, 55]. Afterward, the feature data is passed through classification algorithms such as Gaussian Mixture Model (GMM) [46, 48, 49], Naive Bayes (NB) [51], Random forest (RnF) [50, 51], Support Vector Machine (SVM) [51] and so on. Fu et al. [51] also implemented Convolutional Neural Network (CNN) which achieved a higher accuracy over RnF, NB and SVM. Moreover, some researchers used their own mathematical model [52, 54, 55] for classification which also showed good results.

3.3.2.3 Phase Shift

Phase of a wave is another information which can be used in activity recognition. The phase changes if it is reflected upon a moving object in its propagation path. Path length calculation based on the phase shift shows better ranging accuracy than ToF-based methods. Compared to Doppler shift calculation, this method shows low latency.

Wang et al. [56] used a trajectory-tracking method named Low Latency Acoustic Phase (LLAP) which analyzes the change in phase of the ultrasound signal reflected by moving fingers. The In-phase (I) and Quadrature (Q) components of the received signal change with the hand movements. The complex signal consists of two vectors: the static vector represents the signal for no hand movement and the dynamic vector represents the signal corresponded to the hand movement. The received signal is transformed into two versions. One is multiplied with \( \cos(2\pi ft) \) and the other is multiplied with the phase shifted version \( -\sin(2\pi ft) \). Then each signal is passed through a Cascaded Integrator Comb (CIC) filter which removes the high frequency components and provides the In-phase (I) and Quadrature (Q) components of the received signal. They used a novel algorithm named Local Extreme Value Detection (LEVD) which takes the I/Q components separately and estimates the real and the imaginary part of the static vector. Afterward, to derive the dynamic vector \( \phi(t) \), the static vector is subtracted from the baseband signal. The path length can be calculated from the equation below [56]:

\[
d(t) - d(0) = -\frac{\phi_d(t) - \phi_d(0)}{2\pi} \lambda,
\]

where the left side of the equation denotes the total path length in \( t \) time, \( \phi(0) \) and \( \phi(t) \) denotes the initial signal phase and the phase at time \( t \). This gesture tracking method provides average accuracy up to 3.5 mm and 4.6 mm for 1-D and 2-D hand movements respectively. Moreover, in smartphones, processing latency of LLAP is about 15 ms [56].

Nandakumar et al. [57] proposed another phase shift based novel system, FingerIO, which can detect small changes in finger movement. This method implemented the properties of Orthogonal Frequency Division Multiplexing (OFDM) to detect the small shift in the reflected signal more precisely. It is shown in Fig. 3.5a. The system
generates OFDM signals within 18–20 kHz range. The generated signal consists of 84 samples of OFMD symbol and 200 samples of silence as cyclic suffix. This duration of silence is enough for any reflected signal within 1m range to get back to the receiver.

To generate the echo profile, the received signal is correlated with the transmitted signal. A sample echo profile is shown in Fig. 3.5b. Here, each peak denotes the start of an echo. The distance of the reflecting object can be interpreted from this echo profile. This process gives a decent distance measurement error of 1.4–2.1 cm. The finger movement is identified by comparing two consecutive echo profiles. When a finger movement occurs, the peak of the echo profile is shifted. Comparing with a threshold value, the finger movement is detected. To fine-tune the measurement, FFT is implemented in the approximate region of the peak change in the echo profile. Then the linear phase shift is used in the FFT output to estimate the beginning of the echo with accuracy up to 0.8 cm.

### 3.3.2.4 Passive Signal Processing

The principle of this detection method is to receive and analyze the acoustic signals produced by human activity, i.e., keystroke. The framework only consists of an acoustic signal receiver. Analyzing the received signal in both time and frequency domain, useful features can be derived which can further be used in the classification of certain activities. This method of human activity analysis has been explored in the last few years mainly due to the improved classification tools.

UbiK is a text entry method proposed by Wang et al. [58] which detects the keystrokes on a surface. Taking advantage of the dual microphone in mobile phones, this system extracts multi-path features from the recorder audio and locates the stroke distance from a mapped keyboard outline. A sample setup is shown in Fig. 3.6a. The
Passive method for keystroke detection

Fig. 3.6 Sample setup and ASD profiles of an acoustic-based keystroke detection method using printed keyboard on wooden surface

user taps all the keys on the printed keyboard which is recorded as the training strokes. Afterward, the keystrokes can be detected with their novel keystroke localization algorithm based on the distinct acoustic features of each keystroke. But this framework has two basic conditions to work properly: the surface must make audible sound when the keystroke occurs and the position of the printed keyboard and the mobile phone cannot be changed. The Amplitude Spectrum Density (ASD) profiles of the acoustic signals provide distinct information about the tapped key location. The difference in ASD profiles for two different keystrokes are presented in Fig. 3.6b.

Chen et al. [59] proposed another system named Ipanel which records, analyzes, and classifies the acoustic signals from the finger sliding on to the nearby surface to the recorder. This system can classify surface writing of 26 alphabets, 10 numbers and 7 gestures with overall recognition accuracy above 85%. Both time and frequency domain features have been extracted from the acoustic signals. Firstly, ASD is implemented as a frequency domain feature which did not show overall good accuracy with most machine learning classifiers. Afterward, the time domain Mel Frequency Cepstral Coefficients (MFCC) information with K-Nearest Neighbors (KNN) classifier has been implemented and the system achieved 92% classification accuracy only for the gestures. Lastly, the spectrograms of the acoustic signals have been fed into Convolutional Neural Network (CNN) as images. This technique gave the best overall accuracy among the three. In another study, Du et al. [60] have proposed a deep-learning based handwriting recognition system named WordRecorder. This system detects the words written using pen and paper based on the acoustic signals made during writing. The acoustic signal is recorded by a smartwatch (Huawei smartwatch 1) and then sent to the smartphone. The smartphone preprocesses the acoustic signal using DC component removal, zero padding and Low Pass Filtering. Then the processed signals are passed through CNN which detects letter by letter. With the help of an extra word suggestion module, the predicted letters are converted into words.
Among the above-mentioned sound-based modalities of CHAA, the first three modalities—ToF, Doppler Shift, and Phase Shift are active modalities. In order to minimize the interference from the environmental audible sound and avoid the continuous audible chirp, ultrasound range is preferred over the audible range. On the other hand, the passive signal processing modality is based on the nearby audible signal produced by human activity and the receiver must be in a close range of the activity performed. Moreover, the ToF and Doppler Shift can be employed based on both RF and sound signal. The selection of signal type is based on application scope, available hardware, and environmental dynamics. All the four modalities can be implemented by specialized hardware or general-purpose devices—smartphones, laptops, tablets, etc. To increase the usability, the recent research is more focused to develop the modalities using general-purpose devices.

3.3.3 Vision-Based Approach

This approach has been very popular and powerful tool for contactless activity recognition over the last decade for its usability and accessibility. Vision-based approaches can be classified into two types depending on data type and they are: video data and skeleton data.

3.3.3.1 Video-Based Activity Recognition

Video is a sequence of frames. So, to understand the video, we need to extract frames from the video first. Afterward, different methods are applied on extracted frames to identify where and what is shown in the frames. Figure 3.7 shows a simple working diagram of video-based activity recognition.

Vrigkas et al. [61] presented a comprehensive survey of existing research works that are relevant to vision-based approaches for activity analysis and classified them into two main categories: unimodal and multi-modal approaches. Unimodal methods use data through a single modality are further classified into space-time based, rule-based, shape-based, and stochastic methods. On the other hand, multimodal approaches use different sources for data and group them into three types: social-networking, behavioral, and effective methods. And the type of activity recognition based on video is mostly dependent on viewpoints. Single viewpoint datasets are less challenging for classification as it deals with less environmental noise. On the other hand, multi-view datasets are more challenging, robust, and realistic.

In vision-based HAR, the first task is to segment all background objects using different types of object segmentation algorithms. Following the object segmentation, distinct characteristics of the scene are extracted as feature set [62] and used as input for classification. Moreover, video object segmentation methods are two types based on its environment—background construction-based methods and foreground extraction-based methods. In case of background construction-based methods, the
camera setup remains static relative to the background, hence all information related to the background is obtained in advance and the developed model perform well for object segmentation. On another note, if videos are captured through moving camera, then the scene might have variability and dynamic features. Hence, the background model cannot be built in advance [63], and it is recommended to obtain the model online.

Experimental setup provides a huge contribution on how the methodology should be designed to have a proper classification technique. Background setup, illumination, static and dynamic features of environment, camera view—every point is important. Figure 3.8 shows an ideal setup for multi-view video based HAR.

Light dependency is one of the challenges faced by the traditional cameras as most of them are not illumination invariant. The development of depth cameras have a huge contribution on night vision. For surveillance purpose, it has become one of the most popular techniques. Liu et al. [64] have reviewed different approaches of activity recognition using RGB-D (RGB+Depth) data.

### 3.3.3.2 Skeleton Data-Based Activity Recognition

Skeleton data simply refers to the joint coordinate of human body. Every joint has a position which can be described by three coordinate values. Existing skeleton-based action recognition approaches can be categorized into two types based on the feature extraction techniques: joint-based approaches and body part-based approaches. Joint-based approaches recognize the human skeleton simply as a set of coordinates and its been inspired by the experiment of Johanson [65] which is known by the classic moving lights display experiment. These approaches intend to predict the motion
through a model using individual joint coordinate or combinations of various joints with various features like joint angle with positions, joint orientations with respect to a particular axis, pair-wise relative joint positions, etc. Besides, body part-based approaches consider the skeleton as a connected set of rigid segments. However, these approaches either try to predict the temporal sequence of individual body parts or focus on the connected pairs of rigid segments and predict the temporal sequence of joint angles.

In Fig. 3.9, we have classified the methods for 3D action representation into three major categories: joint-based representations, that capture the relation between body joints through extraction of alternative features; mined joint-based descriptors, that are useful to distinguish distinct characteristics among actions by learning what body parts are involved; and dynamics-based descriptors, that are used in developing the dynamic sequences of 3D trajectories. On another note, the joint-based representations can be categorized into spatial descriptors, geometric descriptors, and key pose-based descriptors depending on the characteristics of extracted skeleton sequences to the corresponding descriptor.

Background setup and camera viewpoint is very important in case of collecting skeleton data. Generally, skeleton data is collected using Microsoft Kinect sensor or any other depth measuring camera. Here, illumination acts as a big factor to track skeleton. Huge amount of light is not appreciable in case of collecting skeleton data. In general case, experimental setup is free from background complexity, dynamic feature or any other disturbance. This kind of environment is possible to create in experimental studio or laboratory. But in real world scenario, there are problems like background complexity, dynamic feature, variable subject or objects interruption,
etc. Those problems have different solutions to deal with. In laboratory experimental setup, only ideal skeleton data is considered.

Despite of the lack of vision-based open datasets, there are many publicly available datasets for research purposes, which have been recorded in constrained environments. For example, RGBD-HuDaAct [66], Hollywood [67], Hollywood-2 [68], UCF sports [69] are some of datasets consist of interactive actions between humans. These are all RGB datasets. On the contrary, Weizmann [70], MSRC-Gesture [71], NTU-RGBD [72] represents some popular dataset of highly primitive actions of daily activities. Furthermore, there are a few datasets of medical activities that are accessible for research purpose, such as USC CRCNS [73], HMD [74], UI-PRMD [75].

### 3.4 Frequent Challenges

Three of the most widely used contactless activity recognition modalities have been discussed in the previous sections. One modality may show the greatest accuracy in a controlled environment, i.e., lab-environment but experience performance loss in real-life scenarios. Another one might work best in real life scenario but the computational cost is not simply feasible. According to Table 3.1, we can see each modality
Table 3.1  Challenges of different modalities

| Challenges                        | RF-based modalities | Sound-based modalities | Vision-based modalities |
|-----------------------------------|---------------------|------------------------|-------------------------|
| Line of sight                     | Moderate            | Crucial                | Crucial                 |
| Ambient condition                 | Moderate            | Moderate               | Crucial                 |
| Multipath effect                  | Crucial             | Crucial                | Not applicable          |
| System robustness                 | Variable            | Variable               | Variable                |
| Range constraint                  | Crucial             | Crucial                | Moderate                |
| Noise cancellation                | Crucial             | Crucial                | Moderate                |
| Device dependency                 | Moderate            | Moderate               | Minimal                 |
| Intra and inter class variability | Present             | Present                | Present                 |
| Dataset unavailability            | Crucial             | Crucial                | Moderate                |

has its own advantages and disadvantages over others. The frequent challenges that come with these approaches are discussed in the following discussion.

**Line-of-Sight (LoS) Requirement**: Some contactless activity recognition methods require the sensing device and the subject to be in line of sight of each other to work properly. Vision and ultrasound-based methods only work in LoS condition. Although some RF-based approaches can perform even in non-line-of-sight (NLoS), their accuracy drops drastically in NLoS conditions [8].

**Ambient Condition**: Video-based approaches are mostly dependent on the background and illumination conditions of the environment. Below a threshold level of illumination, the whole framework may fail [76]. Variations in ambient condition need to be solved by algorithms so that it can be illumination invariant.

**Multipath Effect**: RF and sound-based human localization and activity recognition systems rely on analysis of the received signal. The transmitted signal can take multiple paths to reach the receiver. Some of these alternate paths are created due to the reflection from stationary objects like walls or roofs and some of them are due to the reflection from the actual human subject. Separating the signals reflecting back from the subject can be challenging. Presence of Multiple humans makes thing more complicated.

**System Robustness**: It is very difficult to detect activity while tracking a subject. Usability and accuracy of the systems largely depend on the distance and velocity of the subject with respect to the sensing module [59]. Some existing works on gesture recognition and vital sign measurement require the human subject to stay stationary to perform activity analysis [9, 43].

**Range Constraint**: Millimeter wave-based systems are very successful at detecting small movements or gestures. But the range of these high frequency signals are comparatively low. Existing works require the subject to be in close proximity to the transmitter. Any obstacle between the transmitter and the subject may cause
system failure [77]. WiFi-based frameworks usually work in indoor situations and limited by the range of the WiFi router. On the other hand, sound-based approaches also face range issues. The active ultrasound-based detection methods work up to 4 m distance from the device [16]. But in applications for micro hand gesture recognition, the working range is below 30 cm from the device [56].

Noise: Removing the noise from the received signal is a key part of these activity recognition methods. The noise can be both internal and external, and it is impossible to remove them completely. Electromagnetic waves in the environment appear as noise for RF-based approaches. Selecting the proper noise threshold can be very challenging. In case of sound-based approaches, sound signals from different equipment can be at the same frequency range which may cause performance degradation [60].

Device Dependency: Some WiFi-based activity recognition methods relying on the RSSI or CSI, can be completely device-free. Doppler and ToF-based approaches require custom hardware like the Universal Software Radio Peripheral (USRP) [42]. Sound-based approaches use both COTS devices like laptops, smartphones, tablets as well as custom made devices. The custom made devices mostly use around 40 kHz of ultrasound signals [45–49]. On the other hand, the smartphone or other COTS device-based methods use 18–22 kHz range [52, 53, 58, 60]. The device is chosen based on the application. Vision-based approaches need a camera device whether its RGB camera or depth camera depending on the application.

Intra and Inter Class Variability: In vision-based approach, an action may have huge variance in its own class. For example, if we capture sign language data, there might be some signs that may have various types of scene or may have very rare amount of similarity in between them. These types of variability are so much challenging to detect [78]. On the contrary, different classes of action may have similarities which make them complex to classify.

Dataset Unavailability: One of the most common problems of this research domain is the lack of standardized dataset. In most cases, datasets are being recorded using personal research instruments and they are not open for all to use, specially for RF and sound-based CHAA. Furthermore, there are some privacy issues related to giving access of video or image data as they contain personal information of the participant. Mostly, the challenge become severe in case of medical activity analysis [79]. Therefore, medical data are very rare to have open access.

Due to the diverse challenges of each modality, any researcher or developer must evaluate the application criteria carefully at the very beginning and choose the optimal modality accordingly.

3.5 Application Domain

The above mentioned modalities can be applied for different purposes. With the development of smart appliances and IoT, the field of Contactless Human Activity
Analysis (CHAA) has been widened in last few years. At the same time application-based CHAA studies have come into focus. Different application fields of the three CHAA modalities are discussed in this section.

**Safety Surveillance:** Surveillance cameras can detect any type of unusual activity by vision-based activity analysis modalities. The advancements of computer vision have given us the opportunity to track walking pattern, eye movement which are significant tools for surveillance. For example, someone is carrying a gun in a crowd or running from the scene by stealing something-all can be detected with vision-based approaches. Furthermore, Doppler radars are widely used for intrusion detection [6, 7]. Some WiFi frameworks do not require LoS condition and can work through walls and obstacles which can be very useful for security surveillance systems [80–82]. RASID [83] uses standard WiFi hardware to detect intrusion. Combined with its ability to adapt to the environment, RASID can achieve a F-measure of 0.93. Moreover, Ding et al. [84] proposed a device-free intrusion detection system using WiFi CSI with an average accuracy of 93%.

**Daily Activity Monitoring:** Daily activity monitoring is one of the basic applications of any CHAA modality. Ultrasound-based approaches have shown promise in monitoring daily activities like sitting, walking and sleep-movements [85]. Kim et al. [40] used Micro-Doppler spectrograms as input to a deep convolutional neural network to classify 4 activities with 90.9% accuracy. E-eyes [86] is a device-free activity recognition system that uses WiFi-CSI to classify different daily activities. WiReader [87] uses energy features from WiFi-CSI along with an LSTM network to recognize handwriting. Moreover, the researches have shown the usability of ultrasound-based approaches for detecting talking faces [48], voice activity [47], and even gait analysis [47]. On the other hand, advancement of computer vision allows us to monitor our daily activities with higher precision than other approaches. Our daily activities like walking, running, standing, jumping, swimming, hand-shaking [88, 89] as well as micro activities like cutting food, drinking water can also be monitored using computer vision [90].

**Elderly and Patient Care:** In case of elderly care, vision-based approaches play a significant role. It can detect whether something unusual has happened to the subject or not. For example, fall detection [91] is one of the prime applications for elderly care. Although static human subjects do not affect WiFi CSI in time domain, a sudden fall results in a sharp change in CSI. WiFi-based fall detection techniques, i.e., WiFall [92] and RT-Fall [93] have used this concept and utilized different machine learning algorithms to distinguish falling from other activities with relatively high accuracy. Compressed features from Ultra-Wideband (UBW) have also been used for this purpose [94]. Moreover, vision-based approaches play a vital role in patient rehabilitation. Extracting body joints from RGB images using pose net has been very popular in case of fitness monitoring or rehabilitation. Pose net method extracts exact coordinates of body joints with high accuracy. Those joint coordinates can be used to monitor exercise [95]. For example, if someone needs to move his hand or other parts of the body with a certain amount of angle, this approach can evaluate the exercise if it is going perfectly or not.
**Digital Well-being:** It involves vital signals, i.e., heartbeats and breathing rate monitoring, sleep tracking and checking fitness activities. Some RF-based frameworks provide a contactless approach for heart rate and breathing rate estimation. Some of these frameworks utilize RSSI measurement [96, 97], some use fine-grained WiFi CSI [98, 99] and in another study doppler radar [100] is used. Adib et al. [9] proposed an FMCW-based approach that measures chest motion during inhalation and exhalation to calculate breathing rate. EQ-Radio [101] converts FMCW-based breathing rate measurement into human emotions using machine learning algorithms. SleepPoseNet [102] utilizes Ultra-Wideband (UBW) radar for Sleep Postural Transition (SPT) recognition which can be used to detect sleep disorders. Ultrasound-based breathing rate monitoring study based on the human thorax movement has shown capability to monitor abnormal breathing activities [43]. Vision-based sleep or drowsiness detection is a significant application in monitoring the working environment. For example, monitoring a driver’s condition plays an important role to evaluate how much that person is fit at duty time. EZ-Sleep [103] provides a contactless insomnia tracking system that utilizes RF signal leveraging RF-based localization to find bed position and an FMCW-based approach to keep track of sleeping schedule. Day to day fitness activities like bicycle, toe-touch and squat can be detected and monitored using ultrasound-based system [51]. Moreover, totally contactless methods of acquiring Electrocardiogram (ECG) signal and detecting Myocardial Infarction (MI) have also been explored [104, 105].

**Infection Prevention and Control:** Due to the outbreak of COVID-19, the researchers have explored different rigorous approaches for Infection Prevention and Control (IPC) [106]. In order to limit the spread of the disease in a mass scale, the vision-based modalities have shown the most potential. CCTV footage of the public spaces, e.g., offices, markets, streets can be analyzed for monitoring social-distancing [107], detecting face-masks [108, 109], detecting high body temperature [110, 111], etc. In very short period of time, some of the researches have been converted into commercial products. NVIDIA’s Transfer Learning Toolkit (TLT) and DeepStream SDK have been used to develop a real-time face-mask detection system [112, 113]. Stereolabs has shown usability of their 3D cameras in real-time social distancing monitoring [114]. Chiu et al. [115] presented a thermography technique which was used on a mass scale during previous SARS breakout. The system was used on 72,327 patients and visitors entering Taipei Medical University—Wan Fang Hospital in Taiwan for 1 month. Negishi et al. [116] proposed a vision-based framework which can monitor respiration and heart-rate along with skin temperature.

**Gesture Control:** Nowadays with the widespread use of Internet of Things (IoT) and availability of smartphones and smartwatches, the contactless device navigation as well as smart home appliances control have opened new dimensions of hand gesture recognition applications. Sound-based CHAA researches explored the area of gesture control extensively using smartphones [53, 56] over last decade. Moreover, systems like AudioGest [52], SoundWave [54], MultiWave [50] have explored the possibility of using laptops and desktop computers. These gesture control methods include both static and dynamic hand gestures as well as small finger gestures. Systems like
FingerIO [57] have presented air-writing recognition capability of CHAA. On the other hand, UbiK [58] and Ipanel [59] have explored surface-writing detection using fingers. WiFinger [117] is a WiFi-based gesture recognition system that leveraged the fine-grained CSI to recognize 9 digits with an accuracy of 90.4%. WiFinger can be used as a number text input with devices like smartphones and laptops. Googles project Soli [12] uses 60GHz millimeter wave radar to detect finger gestures with very high accuracy. In computer vision domain, the invention of the Kinect sensor has brought about drastic changes in the entertainment industry. It had made easier to make 3D environment game and real-time interaction with the gaming environment [118].

Thus, with the increase research on CHAA modalities, new application scopes are being discovered every now and then.

### 3.6 Conclusion

Human activity recognition (HAR) is a vast area of research with various categories. While contact-based methods have been there for a long time, from application point of view, they have considerable limitations. Therefore, Contactless Human Activity Analysis (CHAA) has been a very popular approach by leveraging the properties of wireless signals, ultrasound, video, and skeleton data. It can be used in a wide range of application fields with easy techniques and less complexity. These modalities have been discussed in the previous sections. Some of these modalities can be more effective than others depending on the requirements and applications. Video-based approaches require proper ambient lighting and LoS conditions to work properly. In case the environment does not satisfy these conditions, RF or sound-based approaches might be more suitable options. RF-based approaches do not need LoS but sound-based approaches have greater usability in day to day life due to the widespread use of smartphones. On the other hand, with proper environment, video-based approaches can achieve more accuracy compared to other methods. Flexibility, feasibility, and hardware requirements should also be given a good amount of thought before implementing one of these methods. In this chapter, we have given an overview of the evolution and the current state of all these approaches which will be a beneficiary guide for the new researchers in this field.

### References

1. Hussain, Z., Sheng, M., Zhang, W.E.: Different approaches for human activity recognition: a survey. arXiv preprint arXiv:1906.05074 (2019)
2. Ma, J., Wang, H., Zhang, D., Wang, Y., Wang, Y.: A survey on wi-fi based contactless activity recognition. In: Intl IEEE Conferences on Ubiquitous Intelligence & Computing, Advanced and Trusted Computing, Scalable Computing and Communica-
3 Contactless Human Activity Analysis: An Overview of Different Modalities

1. A survey on human behavior recognition using smartphone-based ultrasonic signal. IEEE Access 7, 100 581–100 604 (2019)
2. Detection of posture and motion by accelerometry: a validation study in ambulatory monitoring. Comput. Human Behav. 15(5), 571–583 (1999)
3. Radar in war and in peace (1945)
4. Radar surveillance through solid materials. In: Command, Control, Communications, and Intelligence Systems for Law Enforcement, vol. 2938. International Society for Optics and Photonics, pp. 139–146 (1997)
5. Radar: an in-building rf-based user location and tracking system. In: Proceedings IEEE INFOCOM 2000. Conference on Computer Communications. Nineteenth Annual Joint Conference of the IEEE Computer and Communications Societies (Cat. No. 00CH37064), vol. 2, pp. 775–784. IEEE (2000)
6. 3d tracking via body radio reflections. In: 11th {USENIX} Symposium on Networked Systems Design and Implementation ({NSDI} 14), pp. 317–329 (2014)
7. Smart homes that monitor breathing and heart rate. In: Proceedings of the 33rd annual ACM Conference on Human Factors in Computing Systems, pp. 837–846 (2015)
8. Tool release: gathering 802.11 n traces with channel state information. ACM SIGCOMM Comput. Commun. Rev. 41(1), 53 (2011)
9. High-resolution doppler model of the human gait. In: Radar Sensor Technology and Data Visualization, vol. 4744. International Society for Optics and Photonics, pp. 8–18 (2002)
10. Soli: Ubiquitous gesture sensing with millimeter wave radar. ACM Trans. Graph. (TOG) 35(4), 1–19 (2016)
11. Study of object detection in sonar image using image segmentation and edge detection methods. Indian J. Sci. Technol. 9(42) (2016)
12. Who knew piezoelectricity? rutherford and langevin on submarine detection and the invention of sonar. Notes and Records of the Royal Society 66(2), 141–157 (2012)
13. Sonar-based real-world mapping and navigation. IEEE J. Robot. Autom. 3(3), 249–265 (1987)
14. Beepeep: a high accuracy acoustic ranging system using cots mobile devices. In: Proceedings of the 5th International Conference on Embedded Networked Sensor Systems, pp. 1–14 (2007)
15. Investigating ultrasonic positioning on mobile phones. In: 2010 International Conference on Indoor Positioning and Indoor Navigation, pp. 1–8. IEEE (2010)
16. Human motion analysis: a review. Comput. Vis. Image Underst. 70(3), 428–440 (1999)
17. The visual analysis of human movement: a survey. Computer Vis. Image Underst. 73(1), 82–98 (1999)
18. The meaning of action: a review on action recognition and mapping. Adv. Robot. 21(13), 1473–1501 (2007)
19. Benchmarking a multimodal and multiview and interactive dataset for human action recognition. IEEE Trans. Cybern. 47(7), 1781–1794 (2016)
20. Hierarchical clustering multi-task learning for joint human action grouping and recognition. IEEE Trans. Pattern Anal. Mach. Intell. 39(1), 102–114 (2016)
23. Yang, X., Tian, Y.: Super normal vector for activity recognition using depth sequences. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 804–811 (2014)

24. Li, M., Leung, H., Shum, H.P.: Human action recognition via skeletal and depth based feature fusion. In: Proceedings of the 9th International Conference on Motion in Games, pp. 123–132 (2016)

25. Burghouts, G., Schutte, K., ten Hove, R.-M., van den Broek, S., Baan, J., Rajadell, O., van Huis, J., van Rest, J., Hanckmann, P., Bouma, H., et al.: Instantaneous threat detection based on a semantic representation of activities, zones and trajectories. Signal Image Video Process. 8(1), 191–200 (2014)

26. Dawn, D.D., Shaikh, S.H.: A comprehensive survey of human action recognition with spatio-temporal interest point (stip) detector. Visual Comput. 32(3), 289–306 (2016)

27. Nguyen, T.V., Song, Z., Yan, S.: Stap: Spatial-temporal attention-aware pooling for action recognition. IEEE Trans. Circuits Syst. Video Technol. 25(1), 77–86 (2014)

28. Xie, Y., Li, Z., Li, M.: Precise power delay profiling with commodity wi-fi. IEEE Trans. Mobile Comput. 18(6), 1342–1355 (2018)

29. Kim, Y., Ling, H.: Human activity classification based on micro-doppler signatures using a support vector machine. IEEE Trans. Geosci. Remote Sens. 47(5), 1328–1337 (2009)

30. Rappaport, T.S., et al.: Wireless communications: principles and practice 2 (1996)

31. Patwari, N., Wilson, J.: Spatial models for human motion-induced signal strength variance on static links. IEEE Trans. Inform. Forensics Secur. 6(3), 791–802 (2011)

32. Seidel, S.Y., Rappaport, T.S.: 914 mhz path loss prediction models for indoor wireless communications in multifloored buildings. IEEE Trans. Antennas Propagation 40(2), 207–217 (1992)

33. Yuan, Y., Zhao, J., Qiu, C., Xi, W.: Estimating crowd density in an rf-based dynamic environment. IEEE Sensors J. 13(10), 3837–3845 (2013)

34. Wu, K., Xiao, J., Yi, Y., Gao, M., Ni, L.M.: Fila: Fine-grained indoor localization. In: Proceedings IEEE INFOCOM, pp. 2210–2218. IEEE (2012)

35. Yang, Z., Zhou, Z., Liu, Y.: From rssi to csi: indoor localization via channel response. ACM Comput. Surv. (CSUR) 46(2), 1–32 (2013)

36. Tse, D., Viswanath, P.: Fundamentals of Wireless Communication. Cambridge University Press (2005)

37. Wang, W., Liu, A.X., Shahzad, M., Ling, K., Lu, S.: Understanding and modeling of wifi signal based human activity recognition. In: Proceedings of the 21st Annual International Conference on Mobile Computing and Networking, pp. 65–76 (2015)

38. Pu, Q., Gupta, S., Gollakota, S., Patel, S.: Whole-home gesture recognition using wireless signals. In: Proceedings of the 19th Annual International Conference on Mobile Computing & Networking. pp. 27–38 (2013)

39. Soumekh, M.: Synthetic Aperture Radar Signal Processing. Wiley, New York, vol. 7 (1999)

40. Kim, Y., Moon, T.: Human detection and activity classification based on micro-doppler communications using deep convolutional neural networks. IEEE Geosci. Remote Sens. Lett. 13(1), 8–12 (2015)

41. Griffiths, H.: New ideas in fm radar. Electron. Commun. Eng. J. 2(5), 185–194 (1990)

42. Liu, J., Liu, H., Chen, Y., Wang, Y., Wang, C.: Wireless sensing for human activity: a survey. IEEE Commun. Surv, Tutorials (2019)

43. Al-Naji, A., Al-Askery, A.J., Gharghan, S.K., Chahl, J.: A system for monitoring breathing activity using an ultrasonic radar detection with low power consumption. J. Sensor Actuator Netw. 8(2), 32 (2019)

44. Biswas, S., Harrington, B., Hajiaghajani, F., Wang, R.: Contact-less indoor activity analysis using first-reflection echolocation. In: 2016 IEEE International Conference on Communications (ICC), pp. 1–6. IEEE (2016)

45. Griffith, H., Hajiaghajani, F., Biswas, S.: Office activity classification using first-reflection ultrasonic echolocation. In: 39th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC). IEEE 2017, 4451–4454 (2017)
46. Kalgaonkar, K., Raj, B.: Acoustic doppler sonar for gait recognition. In: 2007 IEEE Conference on Advanced Video and Signal Based Surveillance, pp. 27–32. IEEE (2007)
47. Kalgaonkar, K., Hu, R., Raj, B.: Ultrasonic doppler sensor for voice activity detection. IEEE Signal Process. Lett. 14(10), 754–757 (2007)
48. Kalgaonkar, K., Raj, B.: Recognizing talking faces from acoustic doppler reflections. In: 2008 8th IEEE International Conference on Automatic Face & Gesture Recognition, pp. 1–6. IEEE (2008)
49. Kalgaonkar, K., Raj, B.: One-handed gesture recognition using ultrasonic doppler sonar. In: 2009 IEEE International Conference on Acoustics, Speech and Signal Processing, pp. 1889–1892. IEEE (2009)
50. Pittman, C.R., LaViola, J.J.: Multiwave: Complex hand gesture recognition using the doppler effect. Graphics Interface, pp. 97–106 (2017)
51. Fu, B., Kirchbuchner, F., Kuijper, A., Braun, A., Vaithyalingam Gangatharan, D.: Fitness activity recognition on smartphones using doppler measurements. In: Informatics, vol. 5, no. 2. Multidisciplinary Digital Publishing Institute, p. 24 (2018)
52. Ruan, W., Sheng, Q.Z., Yang, L., Gu, L., Xu, P., Shangguan, L.: Audiogest: enabling fine-grained hand gesture detection by decoding echo signal. In: Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing, pp. 474–485 (2016)
53. Qi Fan, Y., Hao, T., Xuebing, Z., Yin, L., Sanfeng, Z.: Dolphin: ultrasonic-based gesture recognition on smartphone platform. In: 2014 IEEE 17th International Conference on Computational Science and Engineering, pp. 1461–1468. IEEE (2014)
54. Gupta, S., Morris, D., Patel, S., Tan, D.: Soundwave: using the doppler effect to sense gestures. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, pp. 1911–1914 (2012)
55. Wang, T., Zhang, D., Wang, L., Zheng, Y., Gu, T., Dorizzi, B., Zhou, X.: Contactless respiration monitoring using ultrasound signal with off-the-shelf audio devices. IEEE Internet Things J. 6(2), 2959–2973 (2018)
56. Wang, W., Liu, A.X., Sun, K.: Device-free gesture tracking using acoustic signals. In: Proceedings of the 22nd Annual International Conference on Mobile Computing and Networking, pp. 82–94 (2016)
57. Nandakumar, R., Iyer, V., Tan, D., Gollakota, S.: Fingerio: using active sonar for fine-grained finger tracking. In: Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems, pp. 1515–1525 (2016)
58. Wang, J., Zhao, K., Zhang, X., Peng, C.: Ubiquitous keyboard for small mobile devices: harnessing multipath fading for fine-grained keystroke localization. In: Proceedings of the 12th Annual International Conference on Mobile Systems, Applications, and Services, pp. 14–27 (2014)
59. Chen, M., Yang, P., Xiong, J., Zhang, M., Lee, Y., Xiang, C., Tian, C.: Your table can be an input panel: Acoustic-based device-free interaction recognition. Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies 3(1), 1–21 (2019)
60. Du, H., Li, P., Zhou, H., Gong, W., Luo, G., Yang, P.: Wordrecorder: accurate acoustic-based handwriting recognition using deep learning. In: IEEE INFOCOM 2018-IEEE Conference on Computer Communications, pp. 1448–1456. IEEE (2018)
61. Vrigkas, M., Nikou, C., Kakadiaris, I.A.: A review of human activity recognition methods. Front. Robot. AI 2, 28 (2015)
62. Jalal, A., Kamal, S., Kim, D.: Shape and motion features approach for activity tracking and recognition from kinect video camera. In: 2015 IEEE 29th International Conference on Advanced Information Networking and Applications Workshops, pp. 445–450. IEEE (2015)
63. Lin, W., Sun, M.-T., Poovandran, R., Zhang, Z.: Human activity recognition for video surveillance. In: IEEE International Symposium on Circuits and Systems. IEEE 2008, 2737–2740 (2008)
64. Liu, B., Cai, H., Ju, Z., Liu, H.: Rgb-d sensing based human action and interaction analysis: a survey. Pattern Recogn. 94, 1–12 (2019)
65. Nie, Q., Wang, J., Wang, X., Liu, Y.: View-invariant human action recognition based on a 3d bio-constrained skeleton model. IEEE Trans. Image Process. 28(8), 3959–3972 (2019)
66. Ni, B., Wang, G., Moulin, P.: Rgbd-hudaact: a color-depth video database for human daily activity recognition. In: 2011 IEEE International Conference on Computer Vision Workshops (ICCV workshops). IEEE, pp. 1147–1153 (2011)
67. Laptev, I., Marszalek, M., Schmid, C., Rozenfeld, B.: Learning realistic human actions from movies. In: 2008 IEEE Conference on Computer Vision and Pattern Recognition, pp. 1–8. IEEE (2008)
68. Marszalek, M., Laptev, I., Schmid, C.: Actions in context. In: 2009 IEEE Conference on Computer Vision and Pattern Recognition, pp. 2929–2936. IEEE (2009)
69. Rodriguez, M.D., Ahmed, J., Shah, M.: Action mach a spatio-temporal maximum average correlation height filter for action recognition. In: IEEE Conference on Computer Vision and Pattern Recognition. IEEE 2008, pp. 1–8 (2008)
70. Blank, M., Gorelick, L., Shechtman, E., Irani, M., Basri, R.: Actions as space-time shapes. In: Tenth IEEE International Conference on Computer Vision (ICCV’05) Volume 1, vol. 2, pp. 1395–1402. IEEE (2005)
71. Fothergill, S., Mentis, H., Kohli, P., Nowozin, S.: Instructing people for training gestural interactive systems. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, pp. 1737–1746 (2012)
72. Liu, J., Shahroudy, A., Perez, M.L., Wang, G., Duan, L.-Y., Chichung, A.K.: Ntu rgb+ d 120: A large-scale benchmark for 3d human activity understanding. IEEE Trans. Pattern Anal. Mach, Intell (2019)
73. Carmi, R., Itti, L.: The role of memory in guiding attention during natural vision. J. Vis. 6(9), 4 (2006)
74. Corbillon, X., De Simone, F., Simon, G.: 360-degree video head movement dataset. In: Proceedings of the 8th ACM on Multimedia Systems Conference, pp. 199–204 (2017)
75. Vakanski, A., Jun, H.-P., Paul, D., Baker, R.: A data set of human body movements for physical rehabilitation exercises. Data 3(1), 2 (2018)
76. Ramanathan, M., Yau, W.-Y., Teoh, E.K.: Human action recognition with video data: research and evaluation challenges. IEEE Trans. Human-Mach. Syst. 44(5), 650–663 (2014)
77. Wang, S., Song, J., Lien, J., Poupyrev, I., Hilliges, O.: Interacting with soli: exploring fine-grained dynamic gesture recognition in the radio-frequency spectrum. In: Proceedings of the 29th Annual Symposium on User Interface Software and Technology, pp. 851–860 (2016)
78. Bulling, A., Blanke, U., Schiele, B.: A tutorial on human activity recognition using body-worn inertial sensors. ACM Comput. Surv. (CSUR) 46(3), 1–33 (2014)
79. Ahad, M.A.R., Antar, A.D., Shahid, O.: Vision-based action understanding for assistive healthcare: a short review. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops 2019, 1–11 (2019)
80. Wilson, J., Patwari, N.: See-through walls: Motion tracking using variance-based radio tomography networks. IEEE Trans. Mobile Comput. 10(5), 612–621 (2010)
81. Adib, F., Katabi, D.: See through walls with wifi!. In: Proceedings of the ACM SIGCOMM 2013 Conference on SIGCOMM, pp. 75–86 (2013)
82. Chetty, K., Smith, G.E., Woodbridge, K.: Through-the-wall sensing of personnel using passive bistatic wifi radar at standoff distances. IEEE Trans. Geosci. Remote Sens. 50(4), 1218–1226 (2011)
83. Kosba, A.E., Saeed, A., Youssef, M.: Rasid: a robust wlan device-free passive motion detection system. In: 2012 IEEE International Conference on Pervasive Computing and Communications, pp. 180–189. IEEE (2012)
84. Ding, E., Li, X., Zhao, T., Zhang, L., Hu, Y.: A robust passive intrusion detection system with commodity wifi devices. J. Sens. 2018, (2018)
85. Fu, B., Karolus, J., Grosse-Puppendahl, T., Hermann, J., Kuijper, A.: Opportunities for activity recognition using ultrasound doppler sensing on unmodified mobile phones. In: Proceedings of the 2nd International Workshop on Sensor-Based Activity Recognition and Interaction, pp. 1–10 (2015)
3 Contactless Human Activity Analysis: An Overview of Different Modalities

86. Wang, Y., Liu, J., Chen, Y., Gruteser, M., Yang, J., Liu, H.: E-eyes: device-free location-oriented activity identification using fine-grained wifi signatures. In: Proceedings of the 20th Annual International Conference on Mobile Computing and Networking, pp. 617–628 (2014)

87. Guo, Z., Xiao, F., Sheng, B., Fei, H., Yu, S.: Wireader: adaptive air handwriting recognition based on commercial wi-fi signal. IEEE Internet Things J. (2020)

88. Reddy, K.K., Shah, M.: Recognizing 50 human action categories of web videos. Mach. Vis. Appl. 24(5), 971–981 (2013)

89. Soomro, K., Zamir, A.R., Shah, M.: A dataset of 101 human action classes from videos in the wild. Center Res. Comput. Vis. 2 (2012)

90. Jhuang, H., Garrote, H., Poggio, E., Serre, T., Hmdb, T.: A large video database for human motion recognition. In: Proceedings of IEEE International Conference on Computer Vision, vol. 4, no. 5, 2011, p. 6

91. Mubashir, M., Shao, L., Seed, L.: A survey on fall detection: principles approaches. Neurocomputing 100, 144–152 (2013)

92. Wang, Y., Wu, K., Ni, L.M.: Wifall: Device-free fall detection by wireless networks. IEEE Trans. Mobile Comput. 16(2), 581–594 (2016)

93. Wang, H., Zhang, D., Wang, Y., Ma, J., Wang, Y., Li, S.: Rt-fall: A real-time and contactless fall detection system with commodity wifi devices. IEEE Trans. Mobile Comput. 16(2), 511–526 (2016)

94. Sadreazami, H., Mitra, D., Bolic, M., Rajan, S.: Compressed domain contactless fall incident detection using uwb radar signals. In: 18th IEEE International New Circuits and Systems Conference (NEWCAS). IEEE 2020, pp. 90–93 (2020)

95. Kendall, A., Grimes, M., Cipolla, R.: Posenet: a convolutional network for real-time 6-dof camera relocalization. In: Proceedings of the IEEE International Conference on Computer Vision, pp. 2938–2946 (2015)

96. Patwari, N., Brewer, L., Tate, Q., Kaltiokallio, O., Bocca, M.: Breathfinding: a wireless network that monitors and locates breathing in a home. IEEE J. Selected Topics Signal Process. 8(1), 30–42 (2013)

97. Abdelnasser, H., Harras, K.A., Youssef, M.: Ubibreathe: a ubiquitous non-invasive wifi-based breathing estimator. In: Proceedings of the 16th ACM International Symposium on Mobile Ad Hoc Networking and Computing, pp. 277–286 (2015)

98. Liu, J., Chen, Y., Wang, Y., Chen, X., Cheng, J., Yang, J.: Monitoring vital signs and postures during sleep using wifi signals. IEEE Internet Things J. 5(3), 2071–2084 (2018)

99. Wang, X., Yang, C., Mao, S.: Phasebeat: exploiting csi phase data for vital sign monitoring with commodity wifi devices. In: 2017 IEEE 37th International Conference on Distributed Computing Systems (ICDCS), pp. 1230–1239. IEEE (2017)

100. Islam, S.M., Boric-Lubecke, O., Lubecke, V.M.: Concurrent respiration monitoring of multiple subjects by phase-comparison monopulse radar using independent component analysis (ica) with jade algorithm and direction of arrival (doa). IEEE Access 8, 73 558–73 569 (2020)

101. Zhao, M., Adib, F., Katabi, D.: Emotion recognition using wireless signals. In: Proceedings of the 22nd Annual International Conference on Mobile Computing and Networking, pp. 95–108 (2016)

102. Piriyajitakonkij, M., Warin, P., Lakhan, P., Leelaarporn, P., Kumchaiseemak, N., Suwajanakorn, S., Pianpanit, T., Niparnan, N., Mukhopadhyay, S.C., Wilaiarpitporn, T.: Sleepposenet: multi-view learning for sleep postural transition recognition using uwb. IEEE J, Biomedical Health Inform (2020)

103. Hsu, C.-Y., Ahuja, A., Yue, S., Hristov, R., Kabelac, Z., Katabi, D.: Zero-effort in-home sleep and insomnia monitoring using radio signals. Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies 1(3), 1–18 (2017)

104. Weeks, J., Elsaadany, M., Lessard-Tremblay, M., Targino, L., Lianini, M., Gagnon, G.: A novel sensor-array system for contactless electrocardiogram acquisition. In: 42nd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC), pp. 4122–4125. IEEE (2020)
105. Zhang, J., Chen, Y., Chen, T., et al.: Health-radio: towards contactless myocardial infarction detection using radio signals. IEEE Trans, Mobile Comput (2020)
106. Ulhaq, A., Khan, A., Gomes, D., Pau, M.: Computer vision for covid-19 control: a survey. arXiv preprint arXiv:2004.09420 (2020)
107. Yang, D., Yurtsever, E., Renganathan, V., Redmill, K., Özgüner, U.: a vision-based social distancing and critical density detection system for covid-19. Image video Process, DOI (2020)
108. Jiang, M., Fan, X.: Retinamask: a face mask detector. arXiv preprint arXiv:2005.03950 (2020)
109. Ge, S., Li, J., Ye, Q., Luo, Z.: Detecting masked faces in the wild with lle-cnns. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 2682–2690 (2017)
110. Lahiri, B., Bagavathiapan, S., Jayakumar, T., Philip, J.: Medical applications of infrared thermography: a review. Infrared Phys. Technol. 55(4), 221–235 (2012)
111. Somboonkaew, A., Prempree, P., Vuttivong, S., Wetcharungsri, J., Porntheeraphat, S., Chanhorn, S., Pongsoon, P., Amarit, R., Intaravanne, Y., Chaitavon, K.: Mobile-platform for automatic fever screening system based on infrared forehead temperature. In: Opto-Electronics and Communications Conference (OECC) and Photonics Global Conference (PGC). IEEE 2017, pp. 1–4 (2017)
112. Github—nvidia-ai-iot/face-mask-detection: Face mask detection using nvidia transfer learning toolkit (tlt) and deepstream for covid-19. https://github.com/NVIDIA-AI-IOT/face-mask-detection. Accessed 10 Oct 2020
113. Implementing a real-time, ai-based, face mask detector application for covid-19 l nvidia developer blog. https://developer.nvidia.com/blog/implementing-a-real-time-ai-based-face-mask-detector-application-for-covid-19/. Accessed 10 Oct 2020
114. Using 3d cameras to monitor social distancing stereolabs. https://www.stereolabs.com/blog/using-3d-cameras-to-monitor-social-distancing/. Accessed 10 Oct 2020
115. Chiu, W., Lin, P., Chiu, H., Lee, W., Lee, C., Yang, Y., Lee, H., Hsieh, M., Hu, C., Ho, Y., et al.: Infrared thermography to mass-screen suspected sars patients with fever. Asia Pacific J. Public Health 17(1), 26–28 (2005)
116. Negishi, T., Sun, G., Sato, S., Liu, H., Matsui, T., Abe, S., Nishimura, H., Kirimoto, T., "Infection screening system using thermography and ccd camera with good stability and swiftness for non-contact vital-signs measurement by feature matching and music algorithm. In: 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), pp. 3183–3186. IEEE (2019)
117. Li, H., Yang, W., Wang, J., Xu, Y., Huang, L.: Wifinger: talk to your smart devices with finger-grained gesture. In: Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing, pp. 250–261 (2016)
118. Altanis, G., Boloudakis, M., Retalis, S., Nikou, N.: Children with motor impairments play a kinect learning game: first findings from a pilot case in an authentic classroom environment. Interaction Design and Architecture (s) J.-IxD&A, vol. 19, no. 19, pp. 91–104 (2013)