Feature Article:

DOI. No. 10.1109/MAES.2020.3021322

High Precision Human Detection and Tracking Using Millimeter-Wave Radars

***Han Cui, Naim Dahnoun,* University of Bristol**

INTRODUCTION

Human activity recognition (HAR) is one of the most popular research topics in the world. With the recent development of sensor technologies and machine learning techniques, many HAR systems have been proposed and developed. Such systems often use one of cameras, sensors, and wearable devices or a mixture of them to analyze human behavior [13], [24]. Camera-based methods have been shown to achieve outstanding performance for various tasks, from gesture recognition [18] to posture recognition [7]. However, cameras are intrusive, and many people would be concerned about privacy. Various types of sensors, like radio frequency (RF) signal transceivers and environmental sensors, are also a popular choice for HAR. Sensors monitor and detect changes in the environment caused by human subjects, where the information can be analyzed to recognize the corresponding human activity. For example, researchers have proposed the use of ultrasonic sensors for gait estimation [16] and WiFi sensors for human localizing [23]. Most of these techniques are designed for particular use cases and only work under certain environments. This article investigates the potential use of millimeter-wave (mmWave) radars for HAR. In particular, we have selected the FMCW (frequency modulated continuous wave) mmWave radars made by Texas Instruments (TI) to carry out our experiments.

Typical mmWave sensors use frequencies from 76 to 81 GHz for automotive applications [8], or frequencies at around 60 GHz (e.g., 57 to 64 GHz following the European regulation [6]) as general purpose short range radars. The mmWave radars we selected from TI operate at 76 to 81 GHz and have a maximum of 4 GHz available bandwidth. With FMCW techniques, the high bandwidth allows object detection at a high resolution of around 4 cm. mmWave radars are nonintrusive and are able to sense in various conditions including darkness, smoke, and fog, which are crucial in many applications. They also have a relatively low cost in comparison to many of the other sensors or wearable devices. Although mmWave radars do not provide dense information as a camera would, they provide high-resolution in distance, velocity, and angle estimation of the objects in the scene, which can be potentially very useful for understanding their status and motion, as well as distinguishing the object of interest from background clutter. However, mmWave radar signals attenuate fast through the air and are prone to the effects of noise. Therefore, although existing work on mmWave radars has shown success in automotive applications, few researchers have investigated its ability in HAR.

In this article, we present a real-time human detection and tracking system using the TI mmWave radars. We present a software framework capable of communicating with multiple radars and applying a customized data processing chain. The processing is performed on a general purpose CPU (central processing unit) at 25 frames per second (fps). The system achieves constantly over 90.4% sensitivity on human detection in an indoor environment. We show that using a single mmWave radar would result in a high false-alarm rate on human detection, but the precision can be improved significantly with the use of two radars. The contribution of this article can be summarized as follows.

- We present a novel framework for human detection and tracking using mmWave radars, as an alternative technique to the traditional camera and sensor-based methods.
- We present a fast, configurable, and scalable algorithm for using the radars in different applications.
- We show that the chance of interference is low when using multiple radars concurrently.
- We show how data from multiple radars can be fused to improve the precision of the detection system significantly, from 46.9% to 98.6%.

Authors' current addresses: Merchant Venturers Building, University of Bristol, Bristol BS8 1UB, U.K. (e-mail: hc13414@bristol.ac.uk; naim.dahnoun@bristol.ac.uk).

Manuscript received January 27, 2020, revised June 1, 2020; accepted August 18, 2020, and ready for publication August 31, 2020.

Review handled by Daniel O'Hagan.

0885-8985/21/$26.00 © 2021 IEEE

This article is structured as follows. The "Related Work" section discusses related work on HAR in the literature. "mmWave Radar Preliminaries" discusses some of the preliminary knowledge of the FMCW mmWave radar we used. The "Software Framework" section introduces the software framework we designed and implemented. "Human Detection System" shows our experimental setup and the "System Evaluation" section shows the results evaluation. The article is then concluded.

RELATED WORK

HAR has been studied in depth and many systems have been proposed in the literature, especially during the past decade with the rapid development of microprocessors and computing techniques. HAR problems include, but are not limited to, detecting and recognizing the presence of humans, locating and tracking human motion, posture recognition, activity classification, and abnormal activity detection. The main hardware required for data collection in an HAR system can be categorized as cameras, sensors, and wearable devices.

Camera-based HAR has been studied in depth in the computer vision field, as reviewed by Poppe [15] and Kong and Fu [12]. For example, Dalal and Triggs [5] applied SVM (support vector machine) on HOG (histogram of oriented gradients) features for human detection. The rapid development of neural network techniques has allowed more complex tasks to be performed by computers, such as posture estimation [7]. Depth cameras, also known as three-dimensional (3-D) cameras, are specialized cameras that provide distance information from the camera to the object, in addition to the normal image. The 3-D information captured by the depth cameras led to the study of 3-D point-cloud-based HAR [11].

Sensors capture other information from the environment beyond vision. One major difference between these data and the vision data is that, while the temporal dimension in the vision data is supplementary, it has to be considered in the sensor data for a sensible interpretation. Doppler radars use various frequencies for different applications to detect the Doppler motion of objects. Researchers have used Doppler radars on many tasks, such as action classification [21] and motion detection [3]. Ultrasonic sensors detect human activities through ultrasound. For example, Qi *et al.* [16] presented a gait analysis system that uses a few passive ultrasonic sensors to determine the location of moving humans. RF sensors sense the environment with RF signals at certain frequencies, such as the WiFi signal at around 2.4 GHz. One common approach with WiFi sensors is to setup a few WLAN signal transceivers at different locations with varying signal signatures. A human presence in the environment will reflect signals with different strengths based on the distance to the transceivers. A combination of these signals gives an estimation of the location of the human [25]. Recent work has also shown the possibility of detecting the human pose [28] based on RF signals.

While sensors have to be fixed at the point of interest, have a certain range of view and do not provide any information if the human is out of the region, wearable devices are attached to the human and provide continuous information about their activity. They often require wireless transceivers to transfer the data to a central processor, or have an embedded processor for processing the data and providing real-time feedback to the user. The processing power of embedded platforms is often constrained by the power consumption and the thermal dissipation. Accelerometers and gyroscopes measure the acceleration and the orientation of the device in the x–y–z dimension. A combination of the two is also referred to as an inertial measurement unit. These sensors have shown success on various tasks, such as movement detection [20] and activities classification [2]. GPS sensors measure the geometric position of a human. Commercial GPS sensors can estimate the location of a human with a resolution of a few metres and are often used in addition to other sensors. For example, the fall detection system designed by Wu *et al.* [26] uses accelerometers for fall detection and GPS for the location of the fall. While each type of hardware gives certain information on certain aspects, fusing them together enables the full potential of HAR systems to be explored. To give a few examples, Kantoch [11] proposed a health monitoring system using a mixture of ECG sensors, temperature sensors, and accelerometers, for monitoring physiological data during different activities. Brdiczka *et al.* [4] used cameras, a set of audio sensors, and the Hidden Markov Model for HAR at home. Huang *et al.* [9]

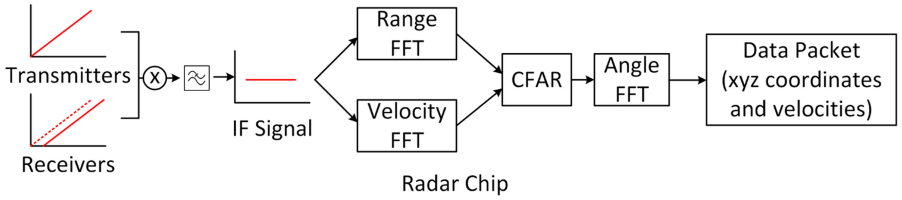

Figure 1.
Data processing chain of the TI mmWave radar.

fused mmWave radars and cameras for tracking moving objects.

While cameras provide the most detailed spatial information and are capable of very complex tasks, such as face recognition and posture recognition, their intrusive nature makes privacy a concern. Without the use of cameras, HAR requires a lot of data fusion between sensors for complex tasks, which increases the cost and the setup complexity of the system. Although the use of mmWave radars has been mentioned, they are only used as Doppler radars and their advantages of high bandwidth and short wavelength are not fully explored.

The use of mmWave is becoming increasingly popular in HAR. Yang *et al.* [27] used mmWave signals to detect the heart rate and the breath pattern of a human by analyzing the signal's RSS (received signal strength). Lien *et al.* [14] uses mmWave radars for hand gesture recognition at a close distance. Björklund *et al.* [3] used mmWave radar as a Doppler radar to detect and classify human movement. While most of the work uses mmWave radars as regular RF radars, only a few researchers, such as Zhao *et al.* [29] and Singh *et al.* [22], use mmWave radars as 3-D sensors and use neural networks for HAR and human identification. By contrast, our system focuses on the detection of a human in the scene. Since the radar image is often unstable and is prone to noise, we discuss the problem of the high false alarm rate when using an mmWave radar for human detection, and propose using two radars concurrently to address the problem.

MMWAVE RADAR PRELIMINARIES

This section gives a brief introduction to the theory of the mmWave radar. A more detailed explanation can be found in [10]. In this article, we used the TI IWR1443 FMCW mmWave radars with a frequency of 76–81 GHz. The radar has one chip consisting of three transmitters and four receivers operating concurrently, as well as integrated circuits and hardware accelerators for a complete on-chip data processing chain. The transmitters send chirp signals S_{tx} (a signal with the frequency increasing linearly with time) to detect any objects in front of the radar. When S_{tx} is reflected by the objects, the signal is received as S_{rx}. The radar combines the two signals S_{tx} and S_{rx} with a mixer and a low-pass filter to produce a mixed intermediate frequency (IF) signal. The IF signal will have a frequency and phase that is equal to the difference between the transmitted signal S_{tx} and the received signal S_{rx}. A data processing chain is then performed over the IF signal to determine the presence of any objects, including the three fast Fourier transforms (FFTs) on the range, velocity, and angle domain, and the CFAR (constant false alarm rate) algorithm to detect peaks from the FFT output, as shown in Figure 1.

DISTANCE CALCULATION

For one object, the frequency difference between the transmitted signal and received signal will be a constant value. This frequency is equal to $S \times \tau$, where S is the slope rate of the chirp and τ is the time of flight. We use d to denote the distance between the radar and the objects, and the time of flight can be expressed as $\tau = 2\ d/c$. Therefore, we can estimate d as

$$d = \frac{(f_1 - f_2)c}{2S}. \tag{1}$$

According to the Fourier transform theory, in order to separate two close frequencies, we need to have $f_1 - f_2 > \frac{1}{T}$, where f_1 and f_2 are the frequencies of the two IF signals representing the two objects and T is the length of IF signal. Therefore, the distance resolution is defined by

$$d > \frac{c}{2ST} \tag{2}$$

where d is the minimal distance required to distinguish two objects, and ST is the total bandwidth of the chirp signal. In practice, mmWave radars often use a 3–4 GHz bandwidth and have a distance resolution of around 4 cm.

ANGLE CALCULATION

The angular position of the object can be calculated by comparing phase differences between neighboring receivers. Given that the phase of any sine wave after traveling along a distance d is $2\pi \cdot \frac{d}{\lambda}$, the phase of the IF signal at any receiver will be $4\pi \cdot \frac{d}{\lambda}$. Assuming there are a number of receivers separated by a distance of $l = \lambda/2$, we

can calculate the angle-of-arrival of a signal through a trigonometric approximation as

$$\theta = \frac{\lambda \triangle \phi}{2\pi l}. \tag{3}$$

Signals from subsequent antennas will form a linear progression in terms of phase, and an estimation of θ can be made with another FFT (known as the angle-FFT). The angular resolution depends on the number of samples we have for the angle-FFT, which is determined by the number of antennas. With N_{tx} TX and N_{rx} RX antennas, we can generate a virtual antenna array of $N_{tx} \times N_{rx}$ with MIMO techniques [17], and the angular resolution can be written as

$$\theta_{\text{res}} = \frac{\lambda}{l \cdot \cos(\theta) \cdot N_{rx} \cdot N_{tx}}. \tag{4}$$

The IWR1443 radar we used has three transmitters arranged in a triangular layout. Therefore, it is able to differentiate objects both horizontally and vertically.

VELOCITY CALCULATION

In order to measure velocity, the radar transmits two chirps separated by time T_c and compares the phase difference between the two received signals. If the object is moving at velocity v, it can be calculated from the phase difference

$$\triangle \phi = 2\pi \frac{2 \cdot T_c \cdot v}{\lambda} \Rightarrow v = \frac{\lambda \triangle \phi}{4\pi T_c}. \tag{5}$$

To get an accurate velocity estimation, the radar sends multiple successive chirps to form a chirp frame, and performs a Doppler-FFT over the phases received from these chirps to find the velocity.

DATA FORMAT

There are two ways to read data from a radar: reading raw data directly from the ADC or reading the processed data from the serial port. The on-chip hardware processor on the radar provides a complete data processing chain to process the ADC data, and therefore it is much easier for the user to use the on-chip processors and only capture the processed data. Communication to the radar is made through the use of two serial ports—one configuration port and one data port. The configuration port allows the PC to interact with the radar and send commands, such as configuring the antennas and switching ON/OFF the radar. The data port is read-only from the PC side, where the radar will start dumping the processed data to this port once it starts operating.

The on-chip data processing chain is user-programmable, and, by using the out-of-box image provided by TI, the processed data can be captured in the form of data messages. The most important data packets are those stating the presence of any object in front of the radar, which will be reported with its *x*–*y*–*z* coordinates, its velocity, and signal strength. We use the term "frame" throughout the rest of this article to define the collection of data points (or data cloud) detected by one or multiple radars.

RADAR AND ANTENNA CONFIGURATION

With the mmWave SDK (software development kit) provided by TI, the user can configure the chirp signal of the radar to fit their use cases. The configuration is defined in the form of configuration files. The file will be transferred to the radar and processed by the on-chip ARM processors upon startup, and the processors will configure the radar subsystem accordingly. The main properties that need to be configured include the number of transmitters and receivers to use, the characterization of the chirp signal, and parameters of the postprocessing algorithms.

Throughout our experiment, we used a radar configuration tuned for indoor environments, with a maximum range of 8 m, a range resolution of 4 cm, a maximum velocity of 1 m/s, and a velocity resolution of 0.1 m/s. The time of each chirp is 125 us, with 10 us idle time (for resetting the chirp) and 115 us chirp ramp time. With a slope rate of 35 MHz/us, we utilize the full 4 GHz bandwidth available for the radar. As our target use case is HAR, we set the CFAR threshold to a relatively low value so that we can receive enough data for postprocessing.

SOFTWARE FRAMEWORK

We implemented a new software framework for managing the radar and performing postprocessing to the data. The system is written in Python and has the following main modules.

- *Radar Handler*: Connects to the radar through the serial ports, loads, and sends the configuration files, receives detection results, and packs them into data matrices.
- *Frame Processor*: Takes data matrices as input, performs customized data processing tasks, and outputs data matrices with the same format.
- *Visualizer*: Manages a number of frame processors for a data processing chain, and displays the final output in 2-D or 3-D formats. The module allows cameras to be connected to the system and to interact with the frame processors, which will be discussed in the "Ground Truth From Cameras" section.

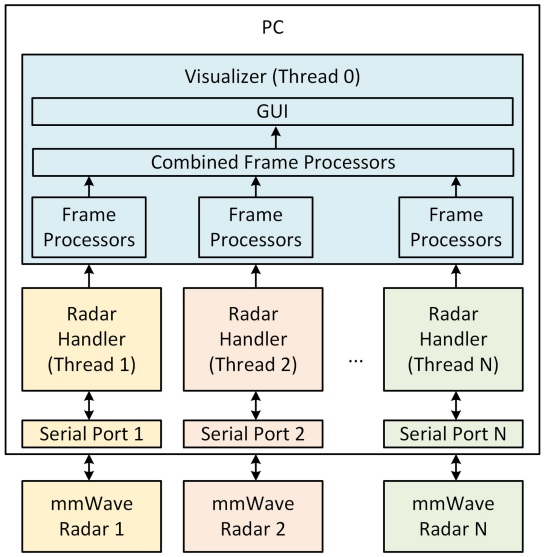

Figure 2.
Software framework for managing multiple radars and applying customised processing chain.

Through the configuration file, the user can specify the number of radars, the model of each radar, the serial port number, and the antenna configuration. The framework utilizes a multithreaded environment. A number of threads will be spawned at startup. Each radar in use will have an independent thread spawned. These threads will each execute a Radar Handler module, connect to the serial ports and handle the communication between the host and the radar. In addition, one visualization thread will be spawned with a visualizer module, and a number of frame processors to achieve the customized postprocessing on the received data. A number of data queues will be created for each of the radar threads and be shared with the visualization thread. The radar threads read data from the serial ports continuously and parse them into an appropriate format, but only push the result into the shared queue if the queue is empty. The visualization thread fetches the data from each queue, performs the user-defined postprocessing tasks on each one of them, displays the combined results and then fetches the next batch of data. The system is designed in a way such that the radar threads only push data once the visualization thread has finished the last frame, to avoid out-of-synchronization caused by different processing speeds of threads. The performance bottleneck of the system will be either the transmitting speed of the radars or the processing speed of all the frame processors, whichever is slower. An overview of the software framework is shown in Figure 2.

The system works best on multicore CPUs when each thread can utilize one physical CPU core, but it can also work on single core machines with reduced performance. The following sections provide a detailed discussion of each module.

RADAR HANDLER

As discussed in the "Data Format" section, the radar has two serial ports that can be accessed by the PC, one for configuring the radar and the other for transmitting the results. The Radar Handler does the following tasks: opens up the two serial ports as specified by the system configuration file, loads the commands from the antenna configuration file, writes the commands to the configuration port, checks the response of each command, and starts listening to the data port upon success. When decoding data from the data port, the radar first searches for the data packet header, filters out the unused packets and extracts the detected object in the frame. The data will be re-arranged into an N by 3 matrix, where N is the number of detected objects and 3 is the x–y–z coordinates. The thread then checks the status of the shared queue, pushes the matrix into the queue if it is empty, and continues searching for the next data packet.

VISUALIZER

The visualizer is responsible for loading the data matrices from all the radar threads, applying user-defined frame processors, combining them into a single frame, and displaying the final output. While combining the data, it applies appropriate rotation and translation to the coordinates from different radars, so that radars at different locations will have a consistent view of the scene. The display can be configured to be 2-D, 3-D, or both and provides a convenient way for interpreting the result.

FRAME PROCESSOR

Frame processors define the operations to be performed on each radar frame. In this article, we introduce three types of frame processors.

- *FIFO Queue*: This module stores the frames using a FIFO (first-in–first-out) queue. During experiments, we found that stacking data in the temporal domain can help to stabilize detection, as data points from real objects will be emphasised but the noise will not.
- *Clustering*: This module groups data points in one frame into clusters according to their distance and filters out small clusters with low numbers of points. We use the DBSCAN (density-based spatial clustering of applications with noise) algorithm for clustering, which does not require prior knowledge of the scene and can extract all qualified clusters. This module helps significantly in reducing the noise.
- *Foreground Extraction*: This module attempts to learn the environment during the first few frames

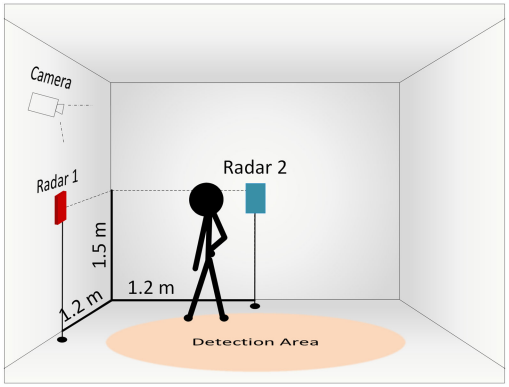

Figure 3.
Hardware setup of the two radars for human detection.

(e.g., frames collected during the first minute when the system starts). Once the collection is finished, it performs the DBSCAN clustering algorithm on these frames and records the detected objects in a local database as clutter. Then, for new frames, the module will compare any new clusters with the clutter in the database and filter out those with a similar size and location. This module can be useful when irrelevant static objects are presented in the area and should be removed.

The frame processor module provides a standard interface also for any other customized operations. All modules work independently and can be loaded as per user requirement, and additional functionality can be easily integrated into the system with new modules, which allows the system to be adapted and deployed for different use cases.

HUMAN DETECTION SYSTEM

SYSTEM SETUP

We built a novel system for locating a human in a room using the two mmWave radars. The hardware setup of our system is shown in Figure 3. We put two radars at different perspectives and put the camera on the top of one radar. The camera is used only to provide the ground truth for the system and is not involved in the detection process. Both of the radars are the IWR1443 model with the same antenna configuration, and the detection area is defined as the intersection area in the sight of both radars.

The radars are calibrated offline, where a rotation matrix and a translation matrix are generated for each radar based on their orientations and locations. The metrics are recorded in a configuration file. They will be loaded into the frame manager module at runtime and be used to translate the detection results into one coordinate system.

HUMAN DETECTION AND TRACKING

The full detection and tracking procedure can be divided into three stages: the two radars sense the scene independently and pass the data to a central processor on a computer; the central processor fuses the data from the two radars and detects the presence of people; the processor invokes the tracking module to verify the detection and refine the results.

INDIVIDUAL DETECTION

As introduced in "mmWave Radar Preliminaries," the radar has a complete on-chip data processing chain to process the analogue mmWave signal and output objects in the form of a data cloud with x–y–z coordinates. This data will be transmitted to the central processor and be processed by the frame processor module independently. The frames will be stacked along the temporal domain using the FIFO queue module. We stacked 10 frames every time which gives a few hundreds points for each subject, at a cost of around 0.4 s processing delay. The data will then be clustered using the DBSCAN algorithm, which examines all the detected points and groups them based on their Euclidean distances between each other, where points within 15 cm will be classified into one cluster. Clusters with a low population will be treated as noise and be discarded. The foreground extraction module can be loaded here to remove static objects in the area. It is considered as an optional module depending on the environment. The resulting clusters from each radar will then be passed to the central frame processor for data fusion.

DATA FUSION

The central frame processor will be triggered once both radar results are ready. It will first transform all the data into one coordinate system by using the calibration parameters. Then, based on the size and the location of the clusters, it will calculate the eigenvectors of each cluster, estimate the distance and the overlapping region between every pair of the clusters, and only keep them if their centroids are close and the majority of the areas overlap. An illustration of the procedure is shown in Figure 4. The raw data from the two radars can be clustered into six candidate subjects (①-⑥), but only ② and ⑥ are overlapping and are considered as one candidate.

A candidate human model will be constructed based on each verified cluster pair and the underlying point cloud data, which contains the estimation of the person's position, height, and volume. While these properties are not expected to be an accurate representation of the real subject, they provide essential information for these candidates to be compared and distinguished. These candidates

will be passed to the tracking module to be correlated with previous frames.

TRACKING

The tracking module records all the candidates at each timestamp and exploits the temporal relationship between them. The concept is similar to a Kalman filter, where we use prior information about an object to estimate the probability distribution of its new position and then verify it. The system will take a 25-frame temporal window, compare the new candidate with each detected object from the previous frames and look for the best match using the candidate properties. If a match is found, i.e., the new candidate is close to a detected object and has a similar size, then it is considered to be the same object being detected again. The decision thresholds are learned during a training stage with a person moving at different speeds and along different paths, to model the possible variation of the parameters. If a match is not found, then the candidate is recorded as a potential new subject and the module waits for further frames to verify it.

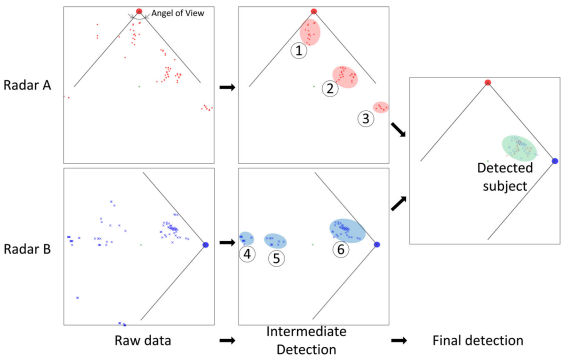

Figure 4.
Workflow of the human detection system, with one person presenting in the area (top-down view).

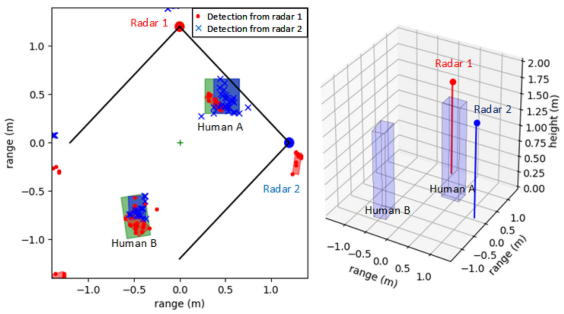

Figure 5.
Example detection when two people are present in the area, from a top-down view (left) and a 3-D view (right).

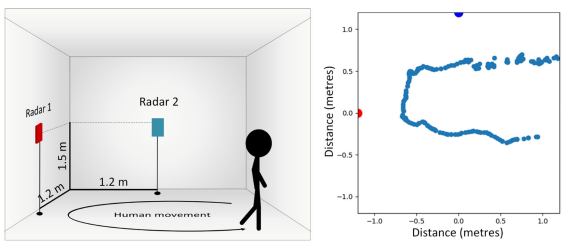

Figure 6.
Example of human tracking. Left: room setup and the movement of the human. Right: human location tracked by the radars.

The module keeps records of the live time of each detected subject and will only report the presence of a subject if the presence has lasted for more than a second, to avoid any phantom effect caused by signal noises. Meanwhile, the position of the subject will be smoothed over the past second to provide a more accurate estimation and reduce outlier effects, taking the assumption that the person will not move at a high speed in an indoor environment and the position should not vary too much within a second. The system is able to resolve multiple people in the area, as the detection process for each subject is independent. An example detection is shown in Figure 5 where two people are presented in the scene and have been detected successfully. An example of human tracking is shown in Figure 6. The current system uses the estimated properties of the human subject (the position, height, and volume) only to correlate them in the temporal domain. However, it is possible that this information can be further exploited for other tasks, such as object classification, human identification, and posture analyze, which we leave for future work.

The system requires a low memory usage and has a low computational cost, allowing the entire process to be performed in real-time. When running on an Intel i7-6700 CPU, the system can achieve 25 fps with only 10% average CPU utilization. The processing speed is only limited by the data processing and transmission speed of the radar. We avoided computationally expensive algorithms, like neural networks on vision-based methods, which would require additional graphic processing units and a much higher cost and power consumption. Therefore, it is possible to port the proposed system onto low power consumption platforms and embedded processors. The system also benefits from its high configurability due to the Frame Processor module, which allows customized functionality to be incorporated into the system based on the use case. For example, the foreground extraction module would be useful when the monitored area has clutter that needs to be removed prior to performing human detection. When using multiple radars, the independent detection stage and the calibration stage mean that the system does not have any restriction on the position or the orientation of the radars, nor the number of radars being used. While in this article we used two radars in a short area, it would be possible to extend the range of view by using more radars without modifying the framework.

SIGNAL INTERFERENCE BETWEEN MULTIPLE RADARS

When using multiple radars, it is important to ensure that they do not interfere with each other. Assuming that we are measuring a maximum distance of 6 m, then the time-of-flight of a round trip would be 0.04 us. With a 35 MHz/us slope rate, this time period gives a frequency change of around 1.4 MHz, as shown in Figure 7. Assuming that there are two radars working simultaneously, we can represent the transmitter signal and the receiver signal of the two radars in (6) to (9), respectively (the amplitude and the phase of the signal can be ignored in this session)

$$S_{tx1}(t) = \sin(2\pi f_1 t) \tag{6}$$

$$S_{rx1}(t) = \sin(2\pi(f_1 - 1.4)t) \tag{7}$$

$$S_{tx2}(t) = \sin(2\pi f_2 t) \tag{8}$$

$$S_{rx2}(t) = \sin(2\pi(f_2 - 1.4)t). \tag{9}$$

Assuming the signals $S_{tx2}(t)$ and $S_{rx2}(t)$ are also detected by the first radar, then the mixer will produce a combination of sinusoidal signals with six different frequency components

$$\begin{aligned} S_{\text{mix}}(t) = {} & \sin(2\pi \cdot 1.4t) + \sin(2\pi|2 \cdot f_1 - 1.4|t) \\ & + \sin(2\pi|f_1 + f_2|t) + \sin(2\pi|f_1 - f_2|t) \\ & + \sin(2\pi|f_1 + f_2 - 1.4|t) \\ & + \sin(2\pi|f_1 - f_2 + 1.4|t). \end{aligned} \tag{10}$$

Since both f_1 and f_2 are within 77 to 81 GHz, the summation frequencies will be very high and, therefore, will be filtered out by the low-pass filter, leaving the other three terms

$$\begin{aligned} S_{\text{filtered}}(t) = {} & \sin(2\pi \cdot 1.4t) + \sin(2\pi|f_1 - f_2|t) \\ & + \sin(2\pi|f_1 - f_2 + 1.4|t). \end{aligned} \tag{11}$$

The first term is the desired result, whereas the other two are the possible interference signals. By configuring the ADC sampling rate and with the help of the built-in digital filter, frequencies beyond 1.4 MHz could be filtered out. In other words, the radar will only keep the detection within the 0.04 us period (the 6-m range). Assuming the cutoff frequency of the radar is set to 1.4 MHz, then the two extra terms in (11) will only stay if $|f_1 - f_2| < 1.4$ MHz or $|f_1 - f_2 + 1.4| < 1.4$ MHz, which evaluates to

$$-2.8\,\text{MHz} < f_1 - f_2 < 1.4\,\text{MHz}. \tag{12}$$

This means that the two radars will only interfere with each other if their frequency difference falls into the 4.2 MHz range. With a 4 GHz bandwidth, this is a probability of around 0.1%, assuming that the radars are switched ON at a random time.

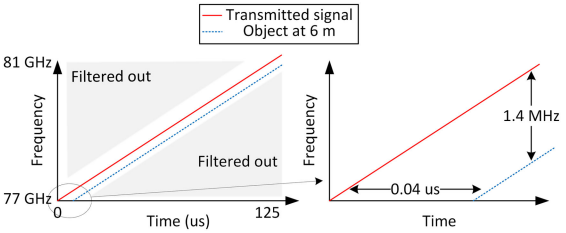

Figure 7.
Received signal strength (and the standard deviation represented by the coloured area) at zero-Doppler domain from the main radar, when the interference radar is placed at a close distance.

As an experiment, we placed two radars at a close distance and pointed them toward the same scene from different angles, kept one of them switched ON (referred to as the main radar) and kept switching ON/OFF the other one periodically (referred to as the interference radar). The scene is setup with static objects placed between 0.5 and 5 m and kept unchanged at all times. We recorded and analyzed the FFT results in the range domain from the main radar.

The experiment was carried out multiple times with different radar locations and lengths of recording. The average variances of the main radar's detection results were recorded and are shown in Table 1. It can be shown that, in all cases, the variances are very similar for the entire scene within the 6 m range, regardless of the status of the interference radar. When paying particular attention to the detection within 3 m (in line with our experimental setup), or the detection with signal strength greater than -3 dB (when the signals are strong enough to be identified), the variances are even lower. Therefore, we conclude that the probability of interference is very low when using two radars concurrently.

One example of the experiment results is shown in Figure 8. The red plot shows the detection result of the main radar when the interference radar was switched OFF, and the blue plot shows the result when the interference radar was switched ON/OFF every three seconds. The results shown were recorded and averaged over a 5-min period (3000 frames). It can be seen that, as the two plots are overlapping, they do not have any significant differences and the variances are low most of the time.

Table 1.

Average Variances of the Main Radar's Detection on Static Objects

	Interference radar active	Interference radar inactive
All detection	0.23	0.20
Detection within 3-m	0.08	0.07
Detection with signal strength >-3 dB	0.06	0.07

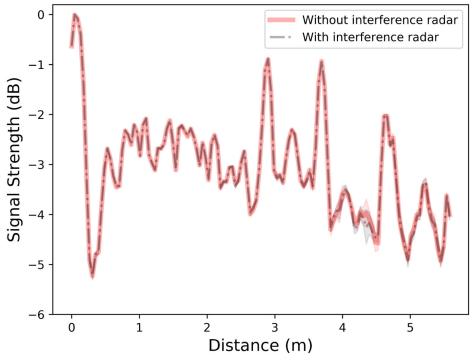

Figure 8.
Transmitted and received signals when detecting an object at 6 m.

The chance of interference can increase if we plan to use more than two radars. When having N radars picking random 4.2 MHz frequency bands in the 4 GHz band, the probability of interference is the probability that any two of the radars pick the same frequency, which is

$$P(N) = 1 - \prod_{i=1}^{N} \frac{4000 - 4.2 \cdot (i-1)}{4000}. \qquad (13)$$

The probability of interference is generally low (less than 1% with four radars and less than 5% with ten radars). This figure will be higher with more than ten radars, which will then require explicit synchronization between radars or an interference detection algorithm.

SYSTEM EVALUATION

GROUND TRUTH FROM CAMERAS

In order to evaluate the performance of our system, we need to have an accurate ground truth on the presence of a human. Since camera-based human recognition has been studied in depth and a lot of successful systems have been developed, we can use them for calculating the ground truth and providing a baseline for evaluation. We used the Yolo-v3 model [19] for human detection.

When the system starts, the visualizer thread reads in the camera data, applies the neural network to the image, obtains the coordinates of the bounding boxes around the humans, and approximates the 3-D areas accordingly. Meanwhile, the radar frame is clustered by the frame processors and each cluster is verified with the 3-D areas. The 3-D areas are estimated using trigonometry and have sector-shapes, and the system will validate a radar-detected-object only if it fits closely in the sector. More specifically, the centroid of the radar detection and the camera detection needs to be within 0.25 m and have at least 70% overlapping area. This provides a low-cost and real-time approach for verifying radar detection and has the potential to allow more complex data labeling for future work.

Table 2.

Performance Evaluation of the System		
	Sensitivity	Precision
One radar	96.4%	46.9%
Two radars	90.4%	98.6%

EVALUATION RESULT

We use the following metrics for evaluating our human detection system.

- *Positives* (P): Humans presented in the detection area.
- *True Positives* (TP): Humans in the detection area that are successfully detected by the radar, with the position verified by the camera detection.
- *False Positives* (FP): Noise or other objects in the detection area that are falsely detected as human, or if the detection is too far from the camera detection.
- *Sensitivity* (TP/P): The ability to detect humans when they are presented in the detection area.
- *Precision* (TP/(TP+FP)): The ability to distinguish humans from false detection.

An ideal system should have both a high sensitivity and a high precision. All the experiments were carried out in a 2.4-m-by-2.4-m region in our laboratory under daily conditions. The system was run for two days and data was collected when at least one human was present in the area. During 56.8% of the time there was only one person in the area, 12.1% with two people, 19.6% with three people, and the rest with more than three people. The results are shown in Table 2.

The high sensitivity in both cases indicates that, whenever a human is present in the area, the system has a very high probability of detecting it. However, with one radar, the 46.9% precision indicates that more than half of the detections would be false detections. With two radars, the system sensitivity was reduced slightly, but the precision improved significantly to 98.6%. In other words, when the one-radar setup detects an object, there is over half the chance that it is a false detection, whereas with two radars the system can be very confident in its detection.

When detecting with one radar, the system reports a large number of false alarms due to noise and flicking of the results. The flicking is observed because of the FFT

process and the peak detection algorithm, where a small change in the signal, once it comes through the FFT, can result in a change in the FFT bins and hence a few centimetres displacement on the object coordinates. This effect will be enlarged when carried over to the angle-FFT, where a displacement in the angle will result in a much larger displacement in the 3-D space. On the other hand, when using two radars, the system has access to two independent detections and can verify the results from each other. As a result, the false alarm rate was reduced significantly (represented by the rise in precision) with only a tiny reduction in the sensitivity.

One limitation of our system is the ability to distinguish multiple people at short distances. The issue is not significant under daily conditions when people are often separated by more than a metre. However, the performance of the system will drop in certain situations, such as counting people in a queue. When there are three or more people and people are occluded by others, the system can only confidently report people in the front, which results in a loss of sensitivity. The occlusion can potentially be solved by using more radars to cover the scene from more angles. As discussed in the "Human Detection and Tracking" section and the "Signal Interference Between Multiple Radars" section, it is possible to adapt more radars into the system without modifying it much. Therefore, the system can be easily adapted to fit different use cases if necessary. Similarly, although we carried out all experiments in a 2.4-m-by-2.4-m region, as we found that the radar's sensitivity to stationary target drops significantly beyond 2.5 m, the range of detection can also be extended by incorporating more radars into the system. We leave the study of using a different number of radars or radar arrays for future work.

CONCLUSION

In this article, we have presented a real-time human detection and tracking system using two mmWave radars. We selected the mmWave radars due to their high resolution, nonintrusive nature, and ability to sense in various view conditions. We introduced preliminary knowledge of mmWave radars in the context of human detection and the typical data processing chain of processing radar data. We presented our software framework for managing multiple radars in a multithreaded environment and applied a customized detection and tracking algorithm. We used lightweight algorithms for real-time processing at 25 fps on a general purpose CPU, making it possible to port the system onto low power consumption platforms. We showed that our system is able to detect humans in indoor environments with over 90% sensitivity. We have discussed the problems of high false alarm rates with a single radar and showed that the precision can be improved from 46.9% to 98.6% with a two-radar setup. We have shown mathematically and empirically that using two mmWave radars will have a very low chance of interference. Since the system is highly configurable, it is possible to incorporate more radars into the system if the application requires a higher detection range or to resolve occlusion. The success in human detection and tracking opens future research opportunities on more complex HAR tasks using mmWave radars.

ACKNOWLEDGMENTS

Special thanks to J. Brand, G. Peake, and A. Thomson from Texas Instruments for equipment support.

REFERENCES

[1] J. K. Aggarwal and L. Xia, "Human activity recognition from 3d data: A review," *Pattern Recognit. Lett.*, vol. 48, pp. 70–80, 2014.

[2] K. Altun and B. Barshan, "Human activity recognition using inertial/magnetic sensor units," in *Proc. Int. Workshop Human Behav. Understanding*, 2010, pp. 38–51.

[3] S. Björklund, H. Petersson, A. Nezirovic, M. B. Guldogan, and F. Gustafsson, "Millimeter-wave radar micro-doppler signatures of human motion," in *Proc. IEEE 12th Int. Radar Symp.*, 2011, pp. 167–174.

[4] O. Brdiczka, M. Langet, J. Maisonnasse, and J. L. Crowley, "Detecting human behavior models from multimodal observation in a smart home," *IEEE Trans. Autom. Sci. Eng.*, vol. 6, no. 4, pp. 588–597, Oct. 2009.

[5] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, 2005, pp. 886–893.

[6] ETSI, "Short Range Devices (SRD); Radio equipment to be used in the 40 GHz to 246 GHz frequency range; Harmonised Standard for access to radio spectrum," 2017.

[7] H.-S. Fang, S. Xie, Y.-W. Tai, and C. Lu, "RMPE: Regional multi-person pose estimation," in *Proc. ICCV*, 2017.

[8] FCC, "Equipment authorization guidance for 76-81 GHz radar devices," 2019.

[9] W. Huang, Z. Zhang, W. Li, and J. Tian, "Moving object tracking based on millimeter-wave radar and vision sensor," *J. Appl. Sci. Eng.*, vol. 21, no. 4, pp. 609–614, 2018.

[10] C. Iovescu and S. Rao, "The fundamentals of millimeter wave sensors," Texas Instruments, 2017.

[11] E. Kantoch, "Technical verification of applying wearable physiological sensors in ubiquitous health monitoring," in *Proc. IEEE Comput. Cardiology*, 2013, pp. 269–272.

[12] Y. Kong and Y. Fu, "Human action recognition and prediction: A survey," 2018, *arXiv:1806.11230*[cs.CV].

[13] O. D. Lara and M. A. Labrador, "A survey on human activity recognition using wearable sensors," *IEEE Commun. Surv. Tut.*, vol. 15, no. 3, pp. 1192–1209, Third Quarter 2013.

[14] J. Lien *et al.*, "Soli: Ubiquitous gesture sensing with millimeter wave radar," *ACM Trans. Graph.*, vol. 35, no. 4, 2016, Art. no. 142.

[15] R. Poppe, "A survey on vision-based human action recognition," *Image Vis. Comput.*, vol. 28, no. 6, pp. 976–990, 2010.

[16] Y. Qi, C. B. Soh, E. Gunawan, K.-S. Low, and R. Thomas, "Assessment of foot trajectory for human gait phase detection using wireless ultrasonic sensor network," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 24, no. 1, pp. 88–97, Jan. 2016.

[17] S. Rao, "MIMO radar," Texas Instruments, 2017.

[18] S. S. Rautaray and A. Agrawal, "Vision based hand gesture recognition for human computer interaction: A survey," *Artif. Intell. Rev.*, vol. 43, no. 1, pp. 1–54, 2015.

[19] J. Redmon and A. Farhadi, "YOLOv3: An incremental improvement," 2018, *arXiv:1804.02767*[cs.CV].

[20] A. M. Sabatini, C. Martelloni, S. Scapellato, and F. Cavallo, "Assessment of walking features from foot inertial sensing," *IEEE Trans. Biomed. Eng.*, vol. 52, no. 3, pp. 486–494, Mar. 2005.

[21] M. S. Seyfioğlu, A. M. Özbayoğlu, and S. Z. Gürbüz, "Deep convolutional autoencoder for radar-based classification of similar aided and unaided human activities," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 54, no. 4, pp. 1709–1723, Aug. 2018.

[22] A. D. Singh, S. S. Sandha, L. Garcia, and M. Srivastava," RadHAR: Human activity recognition from point clouds generated through a millimeter-wave radar," in *Proc. 3rd ACM Workshop Millimeter-Wave Netw. Sens. Syst.*, 2019, pp. 51–56.

[23] B. Wang, S. Zhou, L. T. Yang, and Y. Mo, "Indoor positioning via subarea fingerprinting and surface fitting with received signal strength," *Pervasive Mobile Comput.*, vol. 23, pp. 43–58, 2015.

[24] J. Wang, Y. Chen, S. Hao, X. Peng, and L. Hu, "Deep learning for sensor-based activity recognition: A survey," *Pattern Recognit. Lett.*, vol. 119, pp. 3–11, 2019.

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

[26] F. Wu, H. Zhao, Y. Zhao, and H. Zhong, "Development of a wearable-sensor-based fall detection system," *Int. J. Telemedicine Appl.*, vol. 2015, p. 2, 2015, Art. no. 576364.

[27] Z. Yang, P. H. Pathak, Y. Zeng, X. Liran, and P. Mohapatra, "Vital sign and sleep monitoring using millimeter wave," *ACM Trans. Sensor Netw.*, vol. 13, no. 2, 2017, Art. no. 14.

[28] M. Zhao *et al.*, "Through-wall human pose estimation using radio signals," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2018, pp. 7356–7365.

[29] P. Zhao *et al.*, "mID: Tracking and identifying people with millimeter wave radar," in *Proc. IEEE 15th Int. Conf. Distrib. Comput. Sensor Syst.*, 2019, pp. 33–40.