

MTrack: Tracking Multi-Person Moving Trajectories and Vital Signs with Radio Signals

Dongheng Zhang, Yang Hu, and Yan Chen*, *Senior Member, IEEE*

Abstract—In this paper, we propose a human sensing system with radio signals, MTrack, for in-home healthcare, which is capable of tracking the trajectories of moving persons and vital signs of static persons under the multi-person scenarios. To achieve this, we implement a multi-antenna wideband system that can provide high-resolution angle of arrival (AoA) and time of flight (ToF). A 2D beamformer is utilized to transform the raw radio signals into the AoA-ToF domain. To track the trajectories of moving persons, we leverage the movement of persons to cancel static multipaths and propose a path selection algorithm to estimate the locations of human and suppress the interferences from dynamic multipaths. To track the vital signs of static persons, we utilize the breath of static persons to eliminate static multipaths and propose a correlation-based algorithm to eliminate dynamic multipaths. Extensive experiments show that the proposed MTrack system is capable of tracking multiple moving persons with sub-decimeter level accuracy, and can estimate the breath and heartbeat rate of static persons with median accuracy of 99.8% and 98.46%, respectively.

Index Terms—Human Tracking, Vital Sign, AoA, ToF, Multipath, Healthcare.

I. INTRODUCTION

In the past decades, the number of aged people over the world has been growing steadily [1]. The aged people suffer from various chronic diseases such as congestive heart failure, chronic obstructive pulmonary disease, etc. [2]. Due to the long period of these diseases and the limited medical resources, the treatment has been disturbing for patients and consuming for the human society. Therefore, there is an urgent need for the new solution to alleviate this problem.

In-home human monitoring systems, which could continuously monitor user information such as location and vital sign, have been attractive to provide assistance for personal healthcare, i.e., to provide healthcare professionals with rich information to understand the health conditions of users [2]–[8]. For instance, the location information is able to answer the question like “does the person spend too much time somewhere in home?” or “do the couple stay close with each

other?”; while the vital sign information can be utilized to detect unusual breath and heartbeat, which provides valuable information for medical diagnoses. All in all, in-home human monitoring systems would help doctors to discover potential health risks and provide better care for existing diseases.

In-home human monitoring systems can be generally categorized into three different types based on the sensor they adopt: vision based systems, wearable sensor based systems and radio frequency (RF) based systems. Vision based systems, which are capable of localizing persons and extracting their vital signs, suffer from low light conditions and severe privacy issues in practice [9]. Wearable sensor based systems, which require physical contact between human and sensor, tend to be left by users and thus are not suitable for long-term monitoring [10]. Due to the non-intrusive and privacy-preserving characteristic of RF signals, RF based systems have shown great potential for personal healthcare.

However, due to various practical challenges, the performance of existing RF systems are still quite limited in some aspects, e.g., only work in the single-person scenarios [11], can only detect the presence of user [12], can only track the trajectories of moving persons [13], [14], or can only monitor the breath status [15]. In this paper, we introduce a RF system that can track both the trajectories of moving persons and the vital signs of static persons in multi-person scenarios. Building such a RF system is non-trivial and there are mainly two challenges needed to be resolved: 1) signal separation from different targets; 2) alleviating the multipath interference.

Signal separation from different targets. The first challenge is to separate signals from different targets. To deal with this challenge, we have noted that with AoA and ToF, the location of a target can be determined without any ambiguity. Since different targets generally stay at different locations, their signals can be separated based on the corresponding AoA and ToF. Therefore, we implement a multi-antenna wideband transceiver system, which generates step-frequency signals with 2GHz bandwidth and receives signals through a linear 16-antenna array. We then design a beamformer to separate signals based on the corresponding AoA and ToF. With a high spatial resolution, our system could separate signals from different persons even when they stay close to each other.

Alleviating the multipath interference. The second challenge is to deal with the multipath interference. Multipath interference is a fundamental and challenging problem in wireless communications. It becomes even more challenging for the passive human sensing systems where human do not carry any active device for signal transmitting and thus the reflections from other multipaths can be stronger than that

Copyright (c) 20xx IEEE. Personal use of this material is permitted. However, permission to use this material for any other purposes must be obtained from the IEEE by sending a request to pubs-permissions@ieee.org.

*Corresponding author: Yan Chen (eeyan@ustc.edu.cn).

Dongheng Zhang is with the School of Information and Communication Engineering, University of Electronic Science and Technology of China, Chengdu, Sichuan, China, 611731. Emails: eedhzhang@std.uestc.edu.cn

Yang Hu is with the School of Information Science and Technology, University of Science and Technology of China, Hefei, Anhui, China, 230026. Email: eeyhu@ustc.edu.cn

Yan Chen is with the School of Cyberspace Security, University of Science and Technology of China, Hefei, Anhui, China, 230026. Email: eeyan@ustc.edu.cn

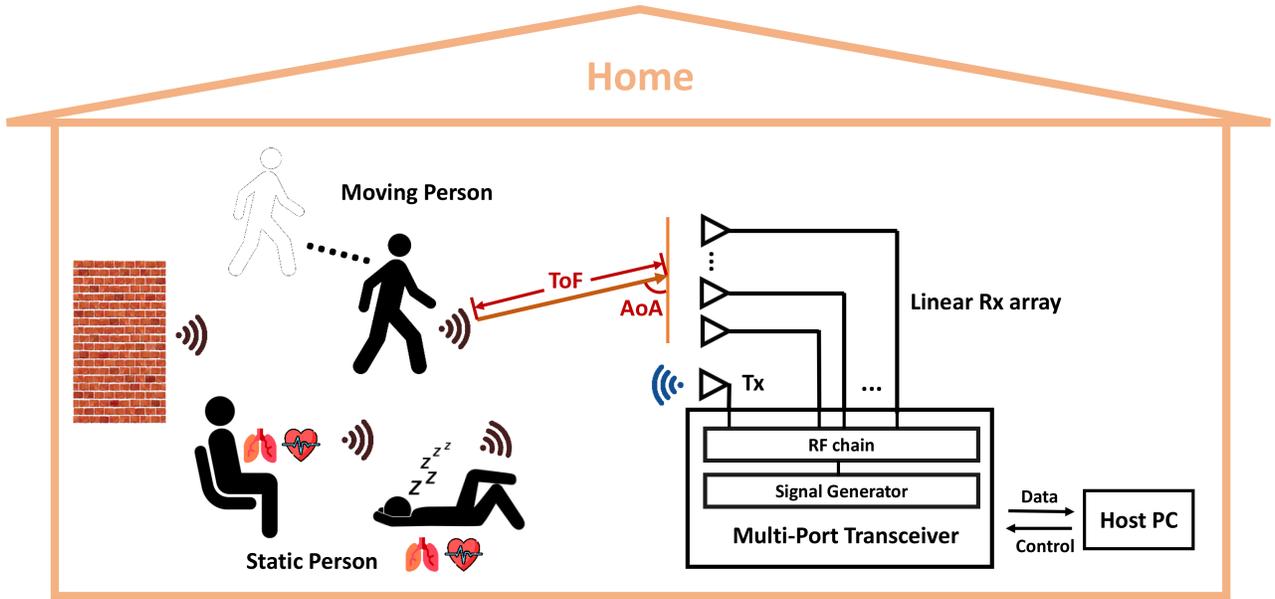


Fig. 1: System model of MTrack. Our system utilizes a step-frequency multi-port transceiver that can transmit and receive wideband signals through multiple antennas. The transmitting signal is reflected by the objects in the environment, including walls, static persons, and moving persons, and finally received by the receiver with a uniform linear antenna array.

from human. There are generally two kinds of multipath interference with different characteristics: static multipaths and dynamic multipaths [13]. We handle the multipath interference by leveraging the fact that different multipaths have different features from the signals reflected by human. Specifically, we propose a spatial-temporal path selection method to track the trajectories of moving persons based on the observations that the location of moving persons changes continuously in the time domain while the reflections of dynamic multipaths vary randomly and the reflections of static multipaths keep the same. We also propose a correlation-based method to extract the signals reflected from the static persons and track the corresponding vital signs based on the observations that the time-domain dynamic multipaths are highly correlated with the human breath signals and the static multipaths are not modulated by the breath status.

The main contributions of this paper are summarized as follows.

- To the best of our knowledge, MTrack is the first system that could track the trajectories of moving persons and the vital signs of static persons using a single device with impressive accuracy.
- To track the trajectories of moving persons, we propose a spatial-temporal path selection algorithm which can leverage the continuous movement of persons to suppress the random dynamic multipath reflections and the constant static multipath reflections.
- To track the vital signs of static persons, we propose a correlation-based method to remove the dynamic multipaths that are highly correlated with the human breath signals and the static multipaths that are not modulated by the breath status.

The rest of the paper is organized as follows. Section II

introduces the related work. Section III illustrates the model of our system. Section IV presents our method to localize persons and extract their vital signs. Extensive experimental results are shown in section V. Finally, conclusions are drawn in Section VI.

II. RELATED WORK

This paper is related to the human tracking and vital sign monitoring. In the following, we introduce in detail the related work on human tracking and vital sign monitoring, respectively.

A. Human Tracking

Mercuri et al. introduced a system to track the distance between transceiver and human [18], while Ram and Ling proposed to track the direction of human with an antenna array [19]. Adib et al. proposed a WiTrack system to track a single person by measuring the ToF on multiple antennas [13]. Later, an upgraded version was proposed to track multiple persons [14]. However, all these systems only obtain one location parameter of human, i.e., AoA or ToF, and then perform triangulation to estimate the location of human, which thus limits the applications of these systems. Moreover, the signals on the multiple antennas cannot be combined coherently to track the vital signs of human. Qian et al. proposed an off-the-shelf WiFi based system to track the location of human [11], [16]. The system is capable of tracking single person with only one WiFi link by jointly estimating the AoA and ToF. However, the accuracy and robustness of the system are still quite limited due to the limited number of antennas and bandwidth of WiFi. There have also been some commercial products such as [12], which however can only detect the presence of the user.

B. Vital Sign Monitoring

Adib et al. proposed a RF based vital sign estimation system [17], which can accurately estimate the breath and heartbeat rate in the single-person scenarios. However, for multi-person scenarios, it requires users to stand 2m away from each other, which may not be reasonable in practice. Zhang et al. proposed a BreathTrack system, which could track human breath status based on the off-the-shelf WiFi [15]. However, due to the limited resources of WiFi, this system also fails in the multi-person scenarios.

III. SYSTEM MODEL

As shown in Fig. 1, our MTrack system utilizes a step-frequency multi-port transceiver that can transmit and receive wideband signals through multiple antennas. Given the frequency band and number of frequency points, the signal generator generates a single-tone signal on every frequency point and the signal is transmitted through the RF chain on a transmitter antenna. The signal is then reflected by the objects in the environment, including walls, static persons, and moving persons, and finally received by the receiver with a uniform linear antenna array. Hence, the received signal can be expressed as

$$s_{m,k}(t) = \sum_{l=1}^L h_l(t) e^{-j2\pi f_k \tau_l(t)} e^{-j2\pi f_k \frac{(m-1)d \cos \theta_l(t)}{c}}, \quad (1)$$

where m , k , l , and t denote the index of receiver antenna, frequency point, propagation path, and time slot respectively, L is the number of propagation paths, $h_l(t)$ is the complex attenuation coefficient, $\theta_l(t)$ and $\tau_l(t)$ denote the AoA and ToF respectively, f_k is the signal frequency, d is the inter-element space of the antenna array, and c is the speed of the signal propagation.

Assuming that there are M receiver antennas and K frequency points, and the step-frequency signal is transmitted over T time slots, then we can put all the received signals on different antennas, different frequency points, and different time slots together into a matrix as follows

$$\mathbf{S} = [\mathbf{s}_{t_1}, \mathbf{s}_{t_2}, \dots, \mathbf{s}_{t_T}], \quad (2)$$

where \mathbf{s}_t is defined as

$$\mathbf{s}_t = [s_{1,1}(t), s_{2,1}(t), \dots, s_{M,1}(t), \dots, s_{M,K}(t)]^T. \quad (3)$$

From (1), we can see that to track the trajectories of moving persons and/or the vital signs of static persons, we need to first separate the signals reflected from different targets based on their AoAs and ToFs. Thus, accurate AoA and ToF information is an important prerequisite for our system to achieve good performance. To guarantee the accuracy of AoA and ToF, the system should have enough AoA and ToF resolution, which can be expressed as [22]

$$R_{\text{AoA}} = 2 \arcsin\left(\frac{\lambda}{M d}\right), \quad (4)$$

and

$$R_{\text{ToF}} = \frac{1}{2B}, \quad (5)$$

where B denotes the bandwidth of the signal.

According to (4) and (5), the resolution is determined by the number of antennas and the signal bandwidth. In our system, the transceiver can transmit step-frequency signals with 2GHz bandwidth and receive signals with a 16-element antenna array. Thus, the AoA and ToF resolution of our system are 16.6° and $0.25ns$, which guarantees the capability of our system for providing accurate AoA-ToF information.

IV. SIGNAL PROCESSING AND TRACKING

In this section, we first introduce how to transform the raw received signals to the AoA-ToF domain. Then, we illustrate how to track the trajectories of moving persons and vital signs of static persons, respectively.

A. Signals in the AoA-ToF Domain

Considering the signal transmitted or reflected from AoA θ and ToF τ , the relative phase shift of this signal on adjacent antennas can be expressed as

$$\Phi(\theta) = e^{-\frac{j2\pi d \cos \theta}{\lambda}}, \quad (6)$$

while the phase shift on adjacent frequencies is given by

$$\Phi(\tau) = e^{-j2\pi \Delta f \tau}, \quad (7)$$

where Δf denotes the difference between adjacent frequencies.

By compensating the phase shift and adding the signals on different antennas and frequencies, the signals from AoA θ and ToF τ would superimpose coherently while the signals from other locations would be suppressed. Hence, the signals from that AoA-ToF could be strengthened and separated. To extract the signal from a specific AoA-ToF, we design a joint AoA-ToF beamformer whose phase shift vector can be expressed as

$$\Phi_{mk}(\theta, \tau) = \exp\left(-j2\pi\left((k-1)\Delta f \tau + f_k \frac{(m-1)d \cos \theta}{c}\right)\right). \quad (8)$$

Since our system has M antennas and K frequency points, the phase shift vector, which is referred as steering vector [23], is given by

$$\Phi(\theta, \tau) = [\Phi_{1,1}(\theta, \tau), \dots, \Phi_{M,1}(\theta, \tau), \dots, \Phi_{M,K}(\theta, \tau)]^T, \quad (9)$$

With (9), the signal from a specific (θ, τ) can be extracted by

$$\mathbf{y}(\theta, \tau) = \Phi^H(\theta, \tau) \mathbf{s}, \quad (10)$$

By defining a spatial grid for candidate AoA-ToF values, we could apply (10) to extract the signals from these AoA-ToFs, which could be expressed as

$$\mathbf{y} = \mathbf{A}^H \mathbf{s}, \quad (11)$$

with \mathbf{A} being defined as follows

$$\mathbf{A} = [\Phi(\theta_1, \tau_1), \Phi(\theta_2, \tau_1), \dots, \Phi(\theta_{g_A}, \tau_1), \dots, \Phi(\theta_{g_A}, \tau_{g_D})], \quad (12)$$

where g_A and g_D denotes the number of candidate AoA and ToF, respectively. An example of such transformation is shown

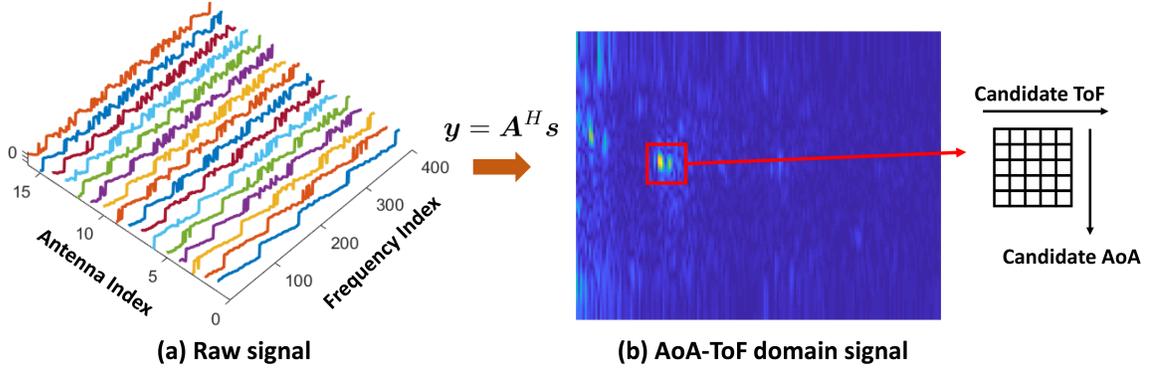


Fig. 2: Transformation to AoA-ToF domain. We design a beamformer to transform raw signal to AoA-ToF domain. The rows in (b) correspond to the candidate AoA while the columns correspond to the candidate ToF. The yellow elements denote strong reflection paths while the blue ones denote weak reflection paths.

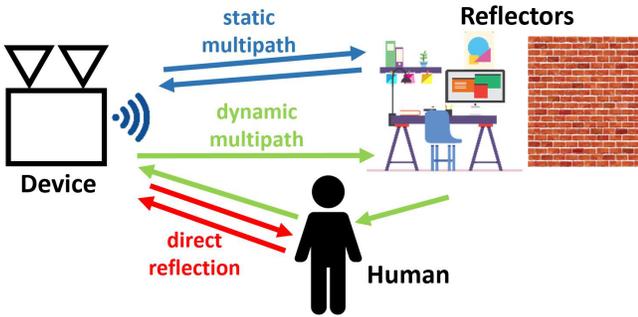


Fig. 3: Illustration of multipath interference.

in Fig. 2. Raw signals are transformed with (11) and reshaped into a matrix. The rows correspond to the candidate AoA and columns correspond to the candidate ToF. The yellow elements denote strong reflection paths while the blue ones denote weak reflection paths.

As shown in Fig. 2(b), each element in \mathbf{y} corresponds to the signal from a specific AoA-ToF at a particular time index. Hence, we call \mathbf{y} the signal in the AoA-ToF domain. With such a transformation, the raw signals are transformed into AoA-ToF domain for further processing. Since our system continuously transmits and receives the RF signals, the AoA-ToF domain signals at all time slots can be obtained by

$$\mathbf{Y} = \mathbf{A}^H \mathbf{S}, \quad (13)$$

where \mathbf{Y} is defined as

$$\mathbf{Y} = \begin{bmatrix} y_1(\theta_1, \tau_1) & y_2(\theta_1, \tau_1) & \dots & y_T(\theta_1, \tau_1) \\ y_1(\theta_2, \tau_1) & y_2(\theta_2, \tau_1) & \dots & y_T(\theta_2, \tau_1) \\ \dots & \dots & \dots & \dots \\ y_1(\theta_{g_A}, \tau_1) & y_2(\theta_{g_A}, \tau_1) & \dots & y_T(\theta_{g_A}, \tau_1) \\ \dots & \dots & \dots & \dots \\ y_1(\theta_{g_A}, \tau_{g_D}) & y_2(\theta_{g_A}, \tau_{g_D}) & \dots & y_T(\theta_{g_A}, \tau_{g_D}) \end{bmatrix}.$$

The columns of \mathbf{Y} correspond to the signals from different (θ, τ) , the rows correspond to the signals from a specific (θ, τ) at different time slots. We use $\mathbf{Y}_{*\mathbf{t}}$ and \mathbf{Y}_{s*} to denote the columns and rows in the rest of this paper to avoid confusion.

B. Track the Trajectories of Moving Persons

In this subsection, we will discuss how to track the trajectories of moving persons. For the simplicity of analysis, we first consider the single-person scenarios. Then, we extend our discussions to the multi-person scenarios.

In the previous subsection, we have discussed how to transform the raw received signals into the AoA-ToF domain, where the amplitude denotes the reflected signal strength from a specific (θ, τ) . Hence, the amplitude of the AoA-ToF domain signal can be utilized to indicate the likelihood that a target exists at that location. In such a case, to localize a target, a straightforward approach is to find the (θ, τ) with the maximum amplitude in the AoA-ToF domain. However, due to the multipath interference, this approach may lead to wrong estimation in practice.

As discussed in the introduction section, there are two kinds of multipaths: static multipaths and dynamic multipaths [13]. As shown in Fig. 3, static multipaths denote the signals reflected from static objects in the environment, e.g., furniture and wall. With the existence of various static objects, the energy of static multipaths is generally very strong. Thus, if we directly choose the (θ, τ) in the AoA-ToF domain with the maximum amplitude, the estimated AoA-ToF would correspond to the location of static objects rather than the moving persons. Fig. 4(a) shows an example of raw AoA-ToF data in practical measurement. All elements with strong reflections are actually static multipaths while human reflections are totally invisible.

To remove static multipaths, we leverage the fact that the reflections from moving persons and static objects have different characteristics in the time domain. More specifically, the reflected signals from moving persons would change over time due to the movement of persons, while the static multipath reflections keep the same over time. Thus, by simply subtracting the signals in consecutive measurements, signals from static objects would be removed while signals from moving persons still persist, i.e.,

$$\mathbf{Y}'_{*\mathbf{t}_i} = \mathbf{Y}_{*\mathbf{t}_{i+1}} - \mathbf{Y}_{*\mathbf{t}_i}. \quad (14)$$

With (14), the static multipaths can be removed, due to

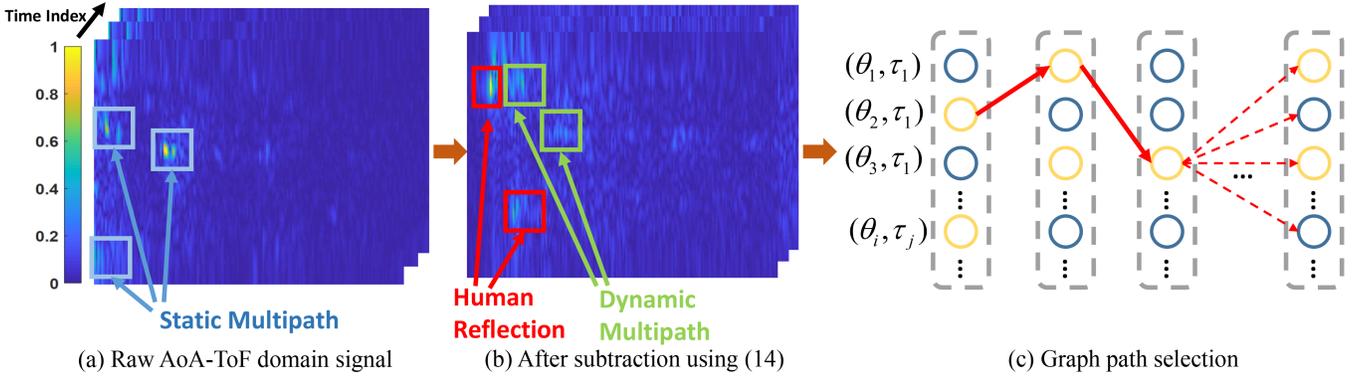


Fig. 4: Procedures of localizing moving persons. We first subtract the signals in consecutive measurements to eliminate static multipath. Then, we formulate and solve the graph path selection problem to suppress dynamic multipath.

which one may localize the moving persons at time t_i by

$$(\hat{\theta}, \hat{\tau})_{t_i} = \arg \max_{\theta, \tau} |\mathbf{Y}'_{*t_i}| \quad (15)$$

where $|\cdot|$ denotes the absolute value.

Although (15) is able to localize the moving persons, the accuracy and robustness would be very limited due to the existence of the dynamic multipaths in \mathbf{Y}'_{*t_i} . As shown in Fig. 3, dynamic multipaths are caused by the fact that moving persons not only reflect signal themselves but also change some reflection paths of static objects, e.g., the moving person may block the wall and thus the reflection from the wall would be changed. Since the dynamic multipath signals change with time, similar to the signals from the moving persons, they still persist after the removal of static multipaths. As shown in Fig. 4(b), the signal strength of dynamic multipaths caused by one person is even stronger compared with the direct reflection from another person. In the single-person scenarios, one can leverage the fact that the direct signal reflected from the moving person to the receiver has a smaller ToF, to remove the dynamic multipaths [13]. However, this method is not applicable to the multi-person scenarios since the ToF of dynamic multipaths caused one person may be smaller than the direct reflection from another person.

To resolve this problem, we observe that the locations of moving persons change continuously due to the limited speed of human movement, while the locations of dynamic multipaths change fast and randomly due to the random distribution of the multipath reflectors in the environment. Based on the observations, we propose to jointly estimate the continuous trajectory of the moving person (θ, τ) rather than simply pick the peak amplitudes of \mathbf{Y}' at every time index.

More specifically, assuming that the AoA-ToF estimation at time index i is a and the AoA-ToF estimation at time index $i+1$ is b , then we can define the cost function at time index i as follows

$$c_i = -|\mathbf{Y}'_{ai}| - |\mathbf{Y}'_{bi+1}| + \omega \alpha(a, b), \quad (16)$$

where $|\mathbf{Y}'_{ai}|$ and $|\mathbf{Y}'_{bi+1}|$ denote the amplitude of signal at AoA-ToF a in time i and AoA-ToF b in time $i+1$, ω is the weighting factor, and $\alpha(a, b)$ denotes the location constraint

in time domain as follows

$$\alpha(a, b) = \omega_\theta \|\theta_a - \theta_b\|_2 + \omega_\tau \|\tau_a - \tau_b\|_2, \quad (17)$$

where ω_θ and ω_τ are the weighting factors.

From (17), we can see that for the signals reflected from moving human, their locations change continuously, which makes $\alpha(a, b)$ small. On the contrary, for dynamic multipaths, their locations change fast and randomly, which makes $\alpha(a, b)$ big. Thus, by adding the location constraint in (17), we could extract the trajectory of moving person by minimizing the following total cost function

$$(\hat{\theta}, \hat{\tau}) = \arg \min_{\theta, \tau} C = \sum_{i=1}^{T-1} c_i. \quad (18)$$

The optimization problem in (18) can be re-formulated as a graph path selection problem. As shown in Fig. 4(c), the vertices of the graph are the candidate AoA-ToFs, and the vertices are connected to all vertices in the last time stamp and the next time stamp with the weight of edges being the cost function in (16). Suppose that we have already known the optimal path from t_0 to every vertices \mathbf{Y}'_{at_i} at t_i , to obtain the optimal path from t_0 to $\mathbf{Y}'_{bt_{i+1}}$ at t_{i+1} , we have

$$C(\mathbf{Y}'_{*t_0} \rightarrow \mathbf{Y}'_{bt_{i+1}}) = \min_{a \in G_{ALL}} \left\{ C(\mathbf{Y}'_{*t_0} \rightarrow \mathbf{Y}'_{at_i}) + c(\mathbf{Y}'_{at_i} \rightarrow \mathbf{Y}'_{bt_{i+1}}) \right\} \quad (19)$$

where G_{ALL} denotes the AoA-ToF grid.

From (19), we can see that the optimization problem in (18) can be efficiently solved by the dynamic programming with value iteration algorithm [21], [24], i.e., the optimal trajectory can be obtained with dynamic programming.

To extend our algorithm to the multi-person scenarios, we propose an iterative estimating and cancellation method to track the trajectories of multiple persons. More specifically, we first perform the aforementioned algorithms once by solving the optimization problem in (18) to extract the trajectory corresponding to the moving human with the maximum signal amplitude. Due to the near-far effect, the signal amplitude of one person may overwhelm the others. Thus, we need to cancel the signal of the first person to track the second one [14]. Hence, with the estimated trajectory, we trace back to \mathbf{Y}' to

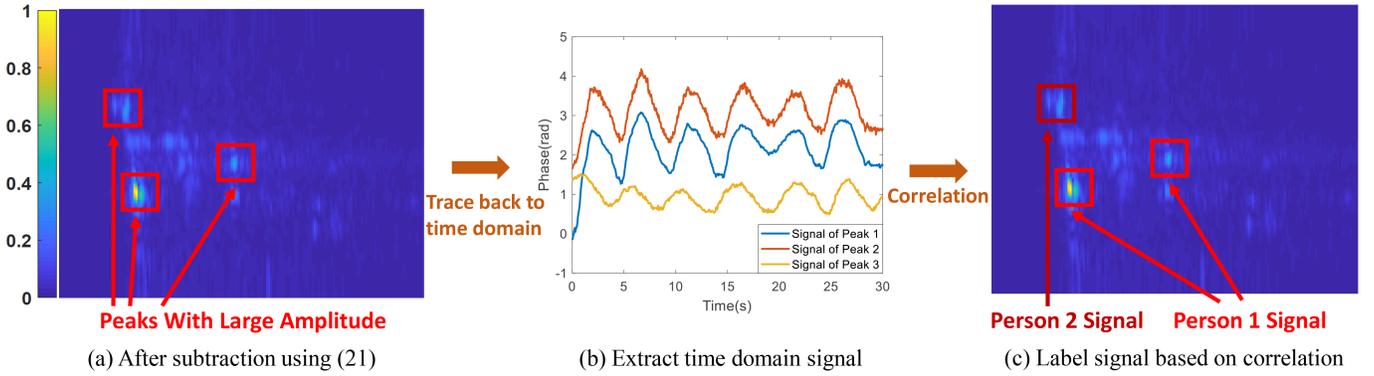


Fig. 5: Procedures of extracting static person vital signs. We first subtract the signals after and before a time window to eliminate static multipath. Then, we trace back to the time domain signal which has large amplitude in AoA-ToF domain. Finally, we label these signals based on the correlation.

eliminate the signals from the first person. Note that since human body is not a point reflector, we need to eliminate the spread of reflected signals in all candidate AoA-ToF points. To achieve this, we set a square mask centered at the location of human body and zero all signals in the mask. We repeat the above procedures until the estimated C is lower than the empirical noise floor to obtain the trajectories of all moving persons.

With the estimated AoA-ToFs, we can obtain the human locations (x, y) through a simple transformation

$$x = \frac{c\tau}{2} \cos \theta \quad y = \frac{c\tau}{2} \sin \theta \quad (20)$$

where τ is divided by 2 since the ToF is the round-trip distance between human and transceiver.

C. Track Vital Signs of Static Persons

Since signals from different targets have been separated in the AoA-ToF domain, given the location of a static person, we could trace back to \mathbf{Y} to extract the signals from that person. Hence, to track the vital signs, i.e., breath and heartbeat status, of static persons, we need to first estimate their locations.

Localizing static persons is considered as a more challenging problem compared with localizing moving persons. This is because static persons do not perform significant movements to change the received signals, which makes it difficult to eliminate static multipaths as we did in the last subsection. Nevertheless, we observe that although the static persons locations keep unchanged, their chests move slightly and periodically due to the breath. In one period of breath, the inhale and exhale make the person's chest move sub-centimeter distance, which would modulate the reflected signals. By selecting an appropriate time window, i.e., half of a breath period, subtracting the signals after and before the time window, the static multipath reflections can be removed while the signals reflected from static persons still persist as below

$$\mathbf{Y}_{*t_i}'' = \mathbf{Y}_{*t_i+W} - \mathbf{Y}_{*t_i}, \quad (21)$$

where W denotes the window length.

When persons keep static in the environment, the dynamic multipath reflections, which are affected by the locations of persons, also keep unchanged. In such a case, the optimization algorithm in (18) could not be applied to suppress the dynamic multipaths anymore. To resolve this problem, we have noted that the dynamic multipaths are caused by the inter modulation between human breath and static multipaths as shown in Fig. 3. Thus, the received signals of direct reflection from static persons and dynamic multipaths can be expressed as

$$s_h(t) = e^{-j2\pi f(\tau_h + \tau_b(t))}, \quad (22)$$

and

$$s_w(t) = e^{-j2\pi f(\tau_w + \tau_b(t))}, \quad (23)$$

where τ_h and τ_w denote the constant ToF between transceiver and the human as well as the dynamic multipath reflector, respectively, $\tau_b(t)$ denotes the time-varying ToF caused by human breath. Since $\tau_b(t)$ is the only term which varies with time, the time variation of the signals reflected from human and dynamic multipaths would be highly correlated. Fig. 5(a)(b) show an example of this phenomenon in practical measurement. The signal from person 1 has the strongest amplitude and the corresponding dynamic multipath is even stronger than the direct reflection from person 2. From the time domain signals, we can observe that the first and second curves are highly correlated because they are actually modulated by the same person, i.e., person 1, while the third curve corresponds to the signal from another person which is very different.

To localize multiple static persons, we first subtract the signals using window W in (21). Since the locations of static persons do not change, we average \mathbf{Y}_{*t_i}'' in time domain to suppress random noise, which can be expressed as

$$\overline{\mathbf{Y}_{*t_i}''} = \frac{1}{T} \sum_{i=1}^T |\mathbf{Y}_{*t_i}''|. \quad (24)$$

Then, we detect the first candidate location of static persons $(\hat{\theta}, \hat{\tau})$ by simply picking the peak with the maximum amplitude in $\overline{\mathbf{Y}_{*t_i}''}$. Then, we eliminate the signals around $(\hat{\theta}, \hat{\tau})$ as we did when localizing moving persons. We repeat

the aforementioned procedures until the amplitude of the peak is below the noise floor.

Next, we define two sets: labeled set \mathbb{D} and unlabeled set \mathbb{R} . The labeled set contains the signals which has been analyzed and labeled, while the unlabeled set contains the signals to be analyzed. Initially, labeled set \mathbb{D} only contains the first detected $(\hat{\theta}_1, \hat{\tau}_1)$, while unlabeled set contains the other $(\hat{\theta}, \hat{\tau})$ to be analyzed, i.e.,

$$\begin{aligned} \mathbb{D} &= \{ \{(\hat{\theta}_1, \hat{\tau}_1)\}_1 \} \\ \mathbb{R} &= \{ (\hat{\theta}_2, \hat{\tau}_2), (\hat{\theta}_3, \hat{\tau}_3), \dots, (\hat{\theta}_N, \hat{\tau}_N) \}, \end{aligned} \quad (25)$$

where N denotes the number of AoA-ToF candidates detected.

To determine whether the signal from $(\hat{\theta}_2, \hat{\tau}_2)$ belongs to another person or belongs to person 1, we trace back to \mathbf{Y} and calculate the correlation coefficient between \mathbf{Y}_{1*} and \mathbf{Y}_{2*} , which is given by

$$\rho_{12} = \frac{\text{cov}(\mathbf{Y}_{1*}, \mathbf{Y}_{2*})}{\sigma_{\mathbf{Y}_{1*}} \sigma_{\mathbf{Y}_{2*}}}. \quad (26)$$

If ρ_{12} in (26) is higher than a pre-defined threshold, we conclude that \mathbf{Y}_{1*} and \mathbf{Y}_{2*} are modulated by the same person and move \mathbf{Y}_{2*} to the detected set with label 1. Then the detected set is given by

$$\mathbb{D} = \{ \{(\hat{\theta}_1, \hat{\tau}_1), (\hat{\theta}_2, \hat{\tau}_2)\}_1 \}. \quad (27)$$

On the other hand, if ρ_{12} in (26) is lower than the pre-defined threshold, we conclude that \mathbf{Y}_{2*} is the signal corresponding to another person. We move \mathbf{Y}_{2*} to the detected set with a new label 2. Then the detected set is given by

$$\mathbb{D} = \{ \{(\hat{\theta}_1, \hat{\tau}_1)\}_1, \{(\hat{\theta}_2, \hat{\tau}_2)\}_2 \}. \quad (28)$$

We iteratively repeat this procedure and compare the signals in the unlabeled set with all signals in the labeled set. If the correlation is lower than the threshold, we add a new label in the labeled set, or the signals in the unlabeled set would be labeled with the same label of which has the maximum correlation. After this procedure, we have classified the detected signals with different labels. The signals with the same label correspond to the signal modulated by the same person as shown in Fig. 5(c). One of them is the direct reflection, while the others are dynamic multipath reflections. Since the signals from different persons have been separated in the AoA-ToF domain, we could leverage the solution for single-person tracking to eliminate the dynamic multipath reflections [13]. More specifically, leveraging the fact that in the single-person scenarios, the direct reflection from human has the minimum ToF compared with other multipath reflections, we pick the AoA-ToF pairs with the minimum ToF as our output.

Once we have known the (θ, τ) of static persons, we trace back to the original \mathbf{Y} to obtain the signals from them. Note that the phase variations of signals correspond to the breath and heartbeat directly as follows

$$\phi(t) = 4\pi \frac{d_b(t) + d_h(t)}{\lambda}, \quad (29)$$

where $d_b(t)$ and $d_h(t)$ are the vibrations caused by breath and heartbeat defined as follows

$$\begin{aligned} d_b(t) &= \alpha_b \cos(2\pi f_b t), \\ d_h(t) &= \alpha_h \cos(2\pi f_h t), \end{aligned} \quad (30)$$



(a) Antenna Configuration (b) Static Human Configuration

Fig. 6: Experiment settings.

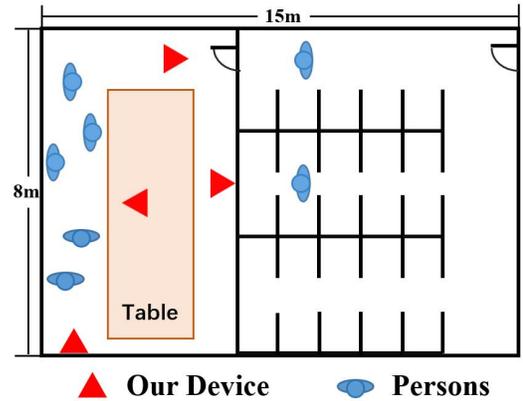


Fig. 7: Experiment environment layout.

where f_b and f_h are the breath rate and heartbeat rate, respectively. α_b and α_h are the corresponding amplitudes, respectively.

To estimate the breath rate, we first perform fast Fourier transform (FFT) on $\phi(t)$. Due to the fact that human breath rate lies between 0.1Hz and 0.4Hz, we find the frequency which has the maximal amplitude between 0.1Hz and 0.4Hz as our estimate, which can be expressed as

$$\hat{f}_b = \arg \max_{f \in [0.1, 0.4]} \Phi(f) = |\text{FFT}(\phi(t))|. \quad (31)$$

Compared with breath signals, heartbeat signals are much weaker. To estimate heartbeat rate, we first pass $\phi(t)$ through a bandpass FIR filter with the passband 1Hz-2Hz, which is the frequency range of human heartbeat. Then, we find the peak from 1Hz to 2Hz with the maximum amplitude as the heartbeat rate estimation.

V. EXPERIMENTS

In this section, we conduct extensive experiments in different scenarios to verify the effectiveness of our MTrack system. We first introduce the implementation details of our system. Then, we present the accuracy of localizing moving persons. Finally, we illustrate the accuracy of localizing static persons and vital signs estimation.

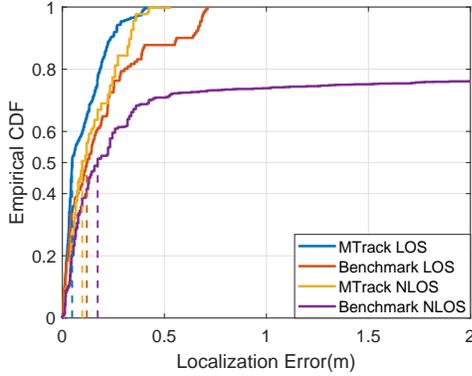


Fig. 8: Localization accuracy of moving persons.

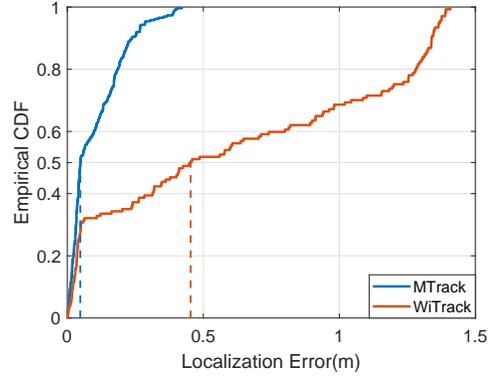


Fig. 9: Accuracy comparison between MTrack and WiTrack.

A. Implementation Details

Our system has one transmitter antenna and sixteen receiver antennas as shown in Fig. 6(a). All antennas are omnidirectional to increase the coverage area of the system. The receiver antennas are arranged on a plastic platform to form a linear antenna array. The inter-element space is 2.6cm, which is set approximately half wavelength to avoid grating lobe. The transceiver generates signal from 4GHz to 6GHz with 401 frequency points, which corresponds to the frequency step of 5MHz. The system performs 16.66 whole frequency step periods in one second, which corresponds to the sampling frequency of 16.66Hz in time domain. Since typical human breath and heartbeat rate are under 2Hz, the system could satisfy the Nyquist frequency to extract breath and heartbeat signals without aliasing. The transmitting signal power is set as -10dBm. The transceiver is controlled by the host PC with a TCP/IP connection. Once the signal acquisition finished, the transceiver transmits raw data to host PC through TCP/IP for further processing. The algorithms proposed in this paper are performed on the host PC. The AoA grid spans $[1, 180]^\circ$ with $g_A = 180$ and ToF grid spans $[0, 0.066]\mu s$ with $g_D = 201$.

To evaluate the performance of our system, we conduct experiments in our lab with twelve participants: four females and eight males. Experiments are conducted in a meeting room and a busy office area as shown in Fig. 7. The meeting room and office are separated by a wall made of double-layered glasses and metal frame with a wooden door. We conduct experiments in both Line-of-Sight (LOS) scenario and Non-Line-of-Sight(NLOS) scenario. In the LOS experiments, participants stay in the meeting room. In the NLOS experiments, participants stay in the office area while the device is still deployed in the meeting room.

B. Accuracy of Tracking Moving Persons

In this subsection, we report the accuracy for tracking moving persons. In all experiments, we ask participants to walk in the experiment area. To obtain the ground-truth trajectories, we ask the participants to walk following the predefined trajectory labeled on the floor. The localization error is defined as the distance between the ground-truth location and the estimated location. In the LOS case, we ask

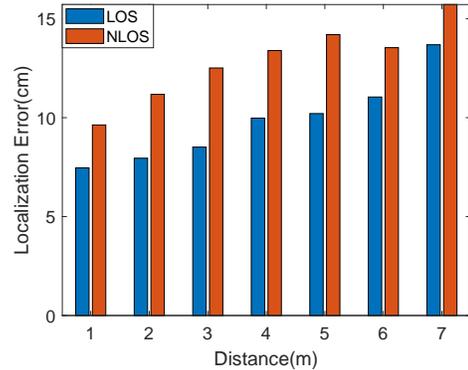


Fig. 10: Localization accuracy of moving persons versus distance.

up to 3 persons to walk simultaneously in the meeting room. In the NLOS case, due to the strong signal attenuation and multipath reflection caused by the wall, we only consider the single-person scenarios. To demonstrate the effectiveness of our system, we compare the performance of our system with a benchmark algorithm, which estimates human locations on different time slots independently as (15).

The localization accuracy in the LOS and NLOS scenarios are shown in Fig. 8. We adopt the cumulative distribution function(CDF) of localization error to illustrate the accuracy of our system. From the figure, we can see that our MTrack system could achieve sub-decimeter level median accuracy in both LOS and NLOS scenarios. In both scenarios, MTrack outperforms the benchmark, and in the NLOS scenario the improvement is more remarkable. This is because in the NLOS scenario, the stronger signal attenuation makes the system suffer from severe multipath and random noise. The benchmark method usually mistakes the human signal and the interference, while MTrack fully utilizes the characteristics of human signal to suppress the interference.

To further illustrate the advantages of the proposed MTrack system, we implement WiTrack in [13], [14] with our antenna configuration and the localization accuracy comparison in LOS scenario is shown in Fig. 9. The accuracy of WiTrack with our antenna configuration is not comparable with the results reported in [13], [14] due to the fact that WiTrack requires the

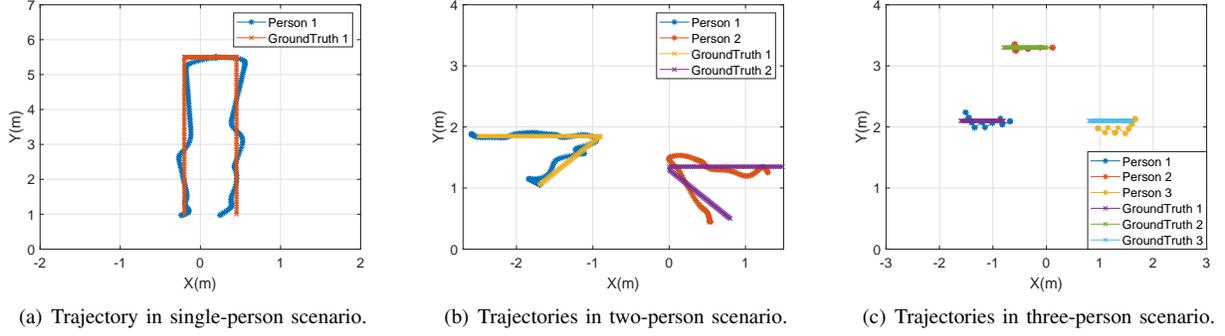


Fig. 11: Extracted human trajectories in different scenarios.

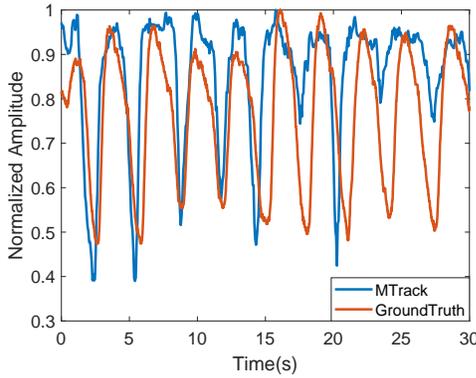


Fig. 12: Comparison of extracted human signal and ground-truth breath signal.

deployment of antennas on a huge platform to achieve accurate trilateration. On the contrary, in our system, 16 antennas are arranged on a platform with only 0.5m width, which is compact but not adequate for WiTrack to achieve accurate trilateration.

To investigate the effective range of our system, we ask participants to walk at different distances to the transceiver. The average localization errors in the LOS and NLOS scenarios are shown in Fig. 10. We can see that our system could localize human targets even when the human is 7m away in the NLOS scenario without severe performance degradation, which makes it be capable of covering typical in-home areas.

To illustrate the accuracy of MTrack intuitively, the human trajectories extracted by MTrack are shown in Fig. 11. We can see that the difference between our estimation and ground-truth is very small even in the three-person scenario.

C. Accuracy of Tracking Static Person Vital Signs

In this subsection, we report the accuracy of our MTrack system for tracking vital signs of static persons. Since human vital signs are too weak to capture, we first conduct single-person experiments to verify the capability of our system, and then perform multi-person experiments. To obtain the ground-truth breath and heartbeat rate, participants wear breath sensor and heartbeat sensor as shown in Fig. 6(b).

In the single-person experiments, participants stay 3m away

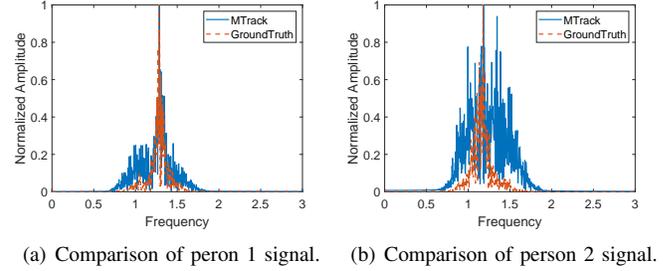


Fig. 13: Comparison of extracted human heartbeat signal and ground-truth in frequency domain.

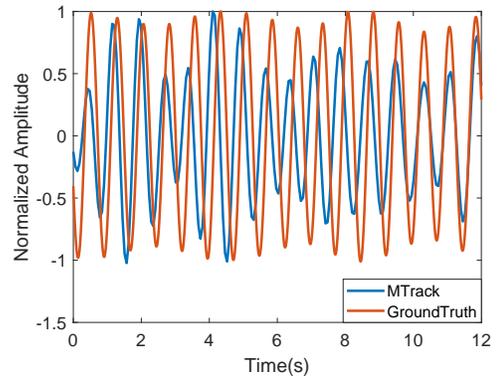


Fig. 14: Comparison of extracted human heartbeat signal and ground-truth heartbeat signal.

from the device, and the monitoring time varies from 1 minute to 5 minutes. During experiments, we ask them to sit down and watch movies. We collect over 4 hours data from all 12 participants. In these experiments, the average localization error for static persons is within 5cm. Human breath rate and heartbeat rate can be estimated with the average accuracy of 99.8% and 98.46%, respectively. To illustrate the effectiveness of our system, the extracted human signals and the ground-truth human breath signals are shown in Fig. 12. As we can see, these two waveforms have the same periods in time domain. The ripples on the extracted human signals actually correspond to the heartbeats. The frequency-domain heartbeat signals of two persons are shown in Fig. 13. The location of the peak with maximum amplitude, which corresponds to heart-

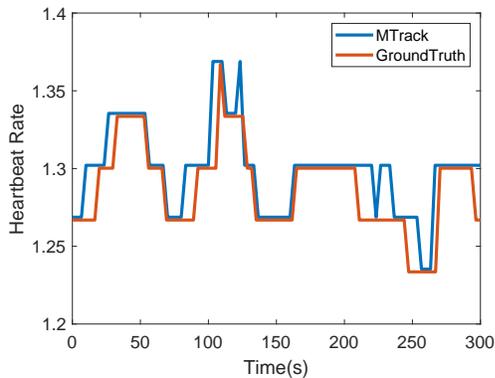


Fig. 15: Comparison of estimated heartbeat rate variation and the ground-truth.

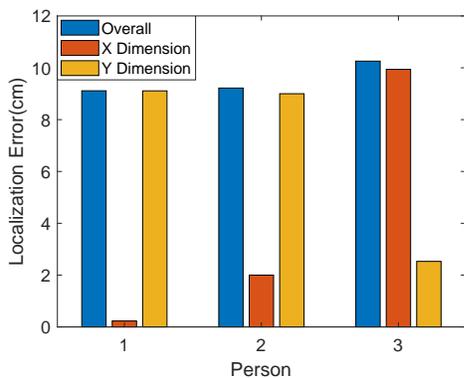


Fig. 16: Localization accuracy of static persons.

beat frequency, is approximately the same for the estimated and groundtruth spectrum, which demonstrates that our system could achieve accurate heartbeat frequency estimation. We also show the time-domain heartbeat signals in Fig. 14. We can see that the extracted heartbeat signals are highly correlated with the ground-truth ones, which demonstrates the effectiveness of our system. To further verify the capability of our system for tracking human heartbeat rate, we let one person sit down and watch movies for 5 minutes. We continuously estimate the heartbeat rate and the results are shown in Fig. 15. Our system could track the frequency variation of the heartbeat accurately, which confirms that we do obtain the heartbeat signals rather than the noise from the environment.

Finally, we evaluate the performance of our system on tracking the vital signs under the multi-person scenarios, where we ask 3 participants to sit in the meeting room. We require the participants to be separated for at least 0.8m away to make the system be capable of separating signals from different persons. During experiments, we ask participants to select their locations freely as long as they are separated for 0.8m. We let them sit down to watch videos and start monitoring for 1 minute. Then, we let them change their locations and repeat the experiment. The accuracy is taken on average of 19 minutes data. The distances between transceiver and participants vary from 2m to 4m in these experiments. The localization in this scenario is still very accurate as shown in Fig. 16. The

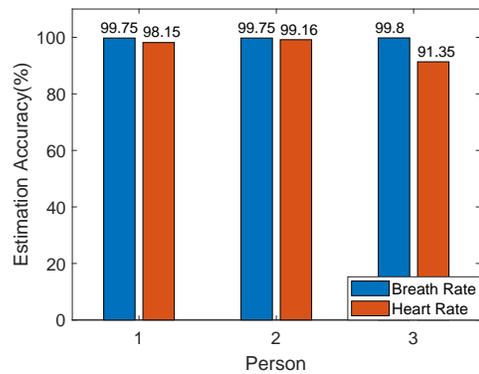


Fig. 17: Vital signs accuracy of static persons.

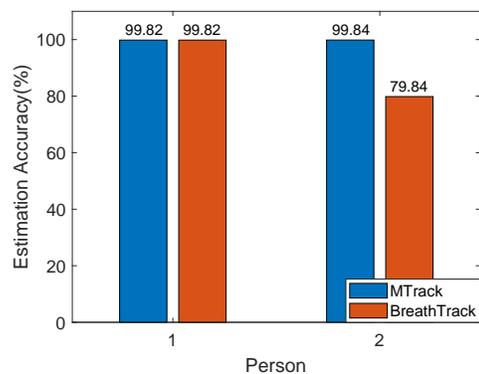


Fig. 18: Accuracy comparison between MTrack and BreathTrack.

breath rate and heartbeat rate accuracy are shown in Fig. 17. We can see that the breath rates of all three persons can be accurately estimated. However, the heartbeat rate estimation accuracy of one participant drops significantly. This is due to the fact that although we could separate signals from different persons based on their AoA-ToF, the separation is not perfect. Since heartbeat signal is too weak to capture, it tends to be distorted by the signals from other persons, which yields the performance degradation.

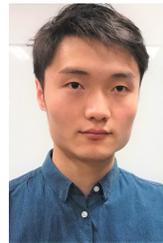
To further illustrate the superiority of the proposed method, we compare MTrack with BreathTrack [15], which simply regards signals with large amplitudes as human signals. The comparison of their breath rate estimation accuracy in a 2-person experiment is shown in Fig. 18. Both methods could achieve high accuracy to estimate the breath rate of person 1, which has the maximum signal amplitude. However, the accuracy of BreathTrack for person 2 drops dramatically while MTrack is still accurate. This is because the dynamic multipath caused by person 1 is stronger than the signal from person 2, which makes BreathTrack mistake the dynamic multipath for person 2, which yields large estimation error; while MTrack could handle this problem with the correlation-based method to distinguish human signal and dynamic multipath.

VI. CONCLUSIONS

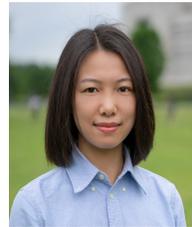
In this paper, we reported a RF system, MTrack, which is capable of tracking the trajectories of moving persons and the vital signs of static persons under multi-person scenarios. To achieve this, we implemented a multi-antenna wideband RF sensing system to separate signals from different persons based on their AoA and ToF. Experimental results showed that the proposed MTrack system can track multiple moving persons with sub-decimeter level accuracy, and estimate the breath and heartbeat rate of static persons with median accuracy of 99.8% and 98.46%, respectively. It is worth to be noted that MTrack is also applicable in other RF-based sensing applications, such as activity recognition, gesture recognition, etc.

REFERENCES

- [1] L. Fontana, B. K. Kennedy, V. D. Longo, D. Seals and S. Melov, "Medical research: Treat ageing", *Nature*, vol. 511, pp. 405-407, Jul. 2014.
- [2] C. Hsu, Y. Liu, Z. Kabelac, R. Hristov, D. Katabi and C. Liu, "Extracting Gait Velocity and Stride Length from Surrounding Radio Signals", in *Proc. ACM CHI'17*, 2017.
- [3] C. Hsu., R. Hristov, G. Lee, M. Zhao and D. Katabi, "Enabling Identification and Behavioral Sensing in Homes using Radio Reflections", in *Proc. ACM CHI'19*, 2019.
- [4] B. Odeh, R. Kayyali, S. Nabhani-Gebara and N. Philip, "Optimizing cancer care through mobile health", *Supportive Care in Cancer*, vol. 23, no. 7, pp. 2183-2188, Jul. 2015.
- [5] B. M. C. Silva, Joel J. P. C. Rodrigues, Isabel. d. I. T. Diez, M. Lopez-Coronado and K. Saleem, "Mobile-health: A review of current state in 2015", *Journal of Biomedical Informatics*, vol. 56, pp. 265-273, Aug. 2015.
- [6] M. S. Mahmud, H. Wang, A. M. Esfar-E-Alam and H. Fang, "A Wireless Health Monitoring System Using Mobile Phone Accessories", *IEEE Internet Things J.*, vol. 4, no. 6, pp. 2009-2018, Dec. 2017.
- [7] B. Wang, Q. Xu, C. Chen, F. Zhang and K. J. R. Liu, "The Promise of Radio Analytics: A Future Paradigm of Wireless Positioning, Tracking, and Sensing", *IEEE Signal Process. Mag.*, vol. 35, no. 3, pp. 59-80, May. 2018.
- [8] F. Zhang, C. Wu, B. Wang, M. Wu, D. Bugos, H. Zhang and K. J. R. Liu, "SMARS: Sleep Monitoring via Ambient Radio Signals", *IEEE Trans. Mobile Comput.*, to be published. doi:10.1109/TMC.2019.2939791
- [9] L. Ren, L. Kong, F. Foroughian, H. Wang, P. Theilmann and A. F. Fathy, "Comparison Study of Noncontact Vital Signs Detection Using a Doppler Stepped-Frequency Continuous-Wave Radar and Camera-Based Imaging Photoplethysmography", *IEEE Trans. Microw. Theory Techn.*, vol. 65, no. 9, pp. 3519-3529, Sep. 2017.
- [10] R. Steele, A. Lo, C. Secombe and Y. K. Wong, "Elderly persons perception and acceptance of using wireless sensor networks to assist healthcare", *International Journal of Medical Informatics*, vol. 78, no. 12, pp. 788-801, Dec. 2009.
- [11] K. Qian, C. Wu, Y. Zhang, G. Zhang, Z. Yang and Y. Liu, "Widar2.0: Passive Human Tracking with a Single WiFi Link", in *Proc. ACM MobiSys'18*, 2018.
- [12] Novelda presence sensor. <https://novelda.com/>.
- [13] F. Adib, Z. Kabelac and D. Katabi, "3D Tracking via Body Radio Reflections", in *Proc. Usenix NSDI'14*, 2014.
- [14] F. Adib, Z. Kabelac and D. Katabi, "Multi-Person Localization via RF Body Reflections", in *Proc. Usenix NSDI'15*, 2015.
- [15] D. Zhang, Y. Hu, Y. Chen and B. Zeng, "BreathTrack: Tracking Indoor Human Breath Status via Commodity WiFi", *IEEE Internet Things J.*, vol. 6, no. 2, pp. 3899-3911, Apr. 2019.
- [16] D. Halperin, W. Hu, A. Sheth, and D. Wetherall, "Tool release: Gathering 802.11n traces with channel state information", *ACM SIGCOMM Comput. Commun. Rev.*, vol. 41, no. 1, pp. 53, Jan. 2011.
- [17] F. Adib, H. Zhao, Z. Kabelac, D. Katabi and R. C. Miller, "Smart Homes that Monitor Breathing and Heart Rate", in *Proc. ACM CHI'15*, 2015.
- [18] M. Mercuri, I. R. Lorato, Y. Liu, F. Wieringa, C. V. Hoof and T. Torfs, "Vital-sign monitoring and spatial tracking of multiple people using a contactless radar-based sensor", *Nature Electronics*, vol. 2, pp. 252-262, Jun. 2019.
- [19] S. S. Ram and H. Ling, "Through-Wall Tracking of Human Movers Using Joint Doppler and Array Processing", *IEEE Geosci. Remote Sens. Lett.*, vol. 5, no. 3, pp. 537-541, Jul. 2018.
- [20] A. Goldsmith, *Wireless Communication*, Cambridge University Press, 2005.
- [21] C. Wu, F. Zhang, Y. Fan and K. J. R. Liu, "RF-based inertial measurement", in *Proc. ACM SIGCOMM'19*, 2019.
- [22] B. R. Mahafza, "Radar systems analysis and design using MATLAB". CRC Press, 2002.
- [23] H. Krim and M. Viberg, "Two decades of array signal processing research: the parametric approach," *IEEE Signal Process. Mag.*, vol. 13, no. 4, pp. 67-94, Jul. 1996.
- [24] T. H. Cormen, C. E. Leiserson, R. L. Rivest and C. Stein, *Introduction to Algorithms*, MIT Press, 2009.



Dongheng Zhang received the B.S. degree from the School of Electronic and Engineering, University of Electronic Science and Technology of China, Chengdu, China, in 2017. He is currently pursuing the Ph.D. degree at the School of Information and Communication Engineering, University of Electronic Science and Technology of China. His research interests are in signal processing, wireless communications and networking.



Yang Hu received the B.S. and Ph.D. degrees in electrical engineering from the University of Science and Technology of China, Hefei, China, in 2004 and 2009 respectively. She was with the University of Maryland Institute for Advanced Computer Studies as a research associate from 2010 to 2015. She is currently an associate professor with the School of Information Science and Technology at the University of Science and Technology of China. Her current research interests include computer vision, machine learning and multimedia signal processing.



Yan Chen (SM'14) received the bachelor degree from the University of Science and Technology of China in 2004, the M.Phil. degree from the Hong Kong University of Science and Technology in 2007, and the Ph.D. degree from the University of Maryland, College Park, MD, USA, in 2011. He was with Origin Wireless Inc. as a Founding Principal Technologist. From Sept. 2015 to Feb. 2020, he was a Professor with the School of Information and Communication Engineering at the University of Electronic Science and Technology of China. He is currently a Professor with the School of Cyberspace Security at the University of Science and Technology of China. His research interests include wireless sensing and imaging, multimedia, and signal processing.

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