SpeedNet: Indoor Speed Estimation with Radio Signals

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

***Abstract*—Indoor human speed estimation is critical to in-home health monitoring of elderly people since it can provide the moving status of human. Contactless indoor speed estimation with radio signals is challenging due to the complicated relationship between the speed of moving human and radio signals. In this paper, we propose an indoor speed estimation framework, SpeedNet, to estimate the speed from the radio signals. Specifically, SpeedNet first extracts the dominant path signal reflected from human through the beamforming technique. Then, SpeedNet obtains the doppler frequency shift (DFS) corresponding to the moving human through analyzing the short time Fourier transform (STFT) spectrogram of dominant path signal. Finally, SpeedNet trains a deep neural network composed of convolutional neural networks (CNN) and long short-term memory networks (LSTM) to utilize the spatial and temporal features of the DFS to estimate the speed of moving human. Experimental results show that SpeedNet can estimate the human moving speed with an average accuracy of 96.33% in a typical indoor environment, which is better than the state-of-the-art approaches.**

***Index Terms*—Wireless sensing, indoor speed estimation, deep learning, STFT, doppler frequency shift**

I. Introduction

Nowadays, the world is facing the aging of population. According to world health organization (WHO), the number of people aging 65 or above is estimated to grow from 524 million in 2010 to 1.5 billion in 2050 [1]. The increasing aging population inherently calls for the assisted-living environments, especially the in-home health monitoring technologies, for elderly people.

Since speed is an important feature of moving status, speed estimation is critical to in-home health monitoring of elderly people [2]–[4]. One straightforward approach to estimate the speed of elderly people is to use the wearable devices equipped with the accelerometer. However, the requirement of wearing a device all the time is cumbersome and may be impossible sometimes. Hence, it is hard to achieve ubiquitous speed monitoring. Another possibility is to use the optical camera together with computer vision techniques. However, this approach has some limitations including: (a) the line-of-sight (LOS) requirement, which means the presence of obstacles (wall, door, etc.) will prevent accurate monitoring; (b) the lighting requirement, which means that in the dark or dim light conditions, it is hard to monitor human; (c) privacy intrusion, i.e., the camera in room is privacy intrusion and cannot be installed in many locations such as bathroom.

In this paper, we focus on the indoor speed estimation using radio signals. The contactless and non-LOS nature of the radio signals makes it much more desirable, compared with the wearable devices and optical cameras. In a radio frequency system, the radio signals emit from the transmitter, before arriving at the receiver, are modulated by the human and the corresponding activities in the environment. By processing the modulated radio signals at the receiver, it is possible for us to analyze the effect of human movement and thus achieve the indoor speed estimation.

Nevertheless, indoor speed estimation with radio signals is non-trivial due to the complicated relationship between the speed of moving human and radio signals. There are mainly two challenges in this problem. Firstly, multipaths commonly exist in an indoor environment, due to which the signal reflected from the moving human is interfered by the signals reflected from the static objects and the LOS signal from the transmitter. How to extract the dominant path signal reflected from the moving human is very challenging. Secondly, human moves with different speeds at different time, i.e., the speed is time-varying. How to estimate the time-varying speed of moving human based on the extracted dominant path signal is also very challenging.

To resolve the above challenges, in this paper, we propose an indoor speed estimation framework, SpeedNet, to estimate the speed from the radio signals. Specifically, SpeedNet first jointly estimates the angle-of-arrival (AOA) and time-of-flight (TOF) of the moving human, and extracts the dominant path signal from the radio signals through the beamforming technique. Then, the dominant path signal is processed by the short time Fourier transform (STFT), and the doppler frequency shift (DFS) corresponding to the moving human is obtained through analyzing the spectrogram. Finally, a deep neural network composed of convolutional neural networks (CNN) and long short-term memory networks (LSTM) is utilized to extract the spatial and temporal features of the DFS to estimate the speed of the moving human. Extensive experiments are conducted to evaluate the performance of SpeedNet. Similar to that in [18], the accuracy of the moving speed estimation is evaluated by comparing the estimated distance (derived

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

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

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

from the estimated speed) with the ground-truth distance. The experimental results show that SpeedNet can estimate the human moving speed with an average accuracy of 96.33% in a typical indoor environment, which is better than the state-of-the-art approaches [17], [18], [23].

The rest of the paper is organized as follows. The related work is discussed in section II. Section III introduces the system model, while section IV describes the proposed SpeedNet framework where the three modules, dominant path extraction module, spectrum analysis module, and deep learning module, are discussed in detail. In section V, we show the extensive experimental results. Finally, conclusions are drawn in section VI.

II. Related Work

In recent years, with the development of Internet of Things (IoT) and wireless communications, contactless health monitoring has drawn more and more attention, and many techniques have been developed in the literature [5]–[10]. Generally, these techniques can be categorized into two types: model based techniques and learning based techniques.

Researchers have utilized different features of radio signals to study the human activity, including amplitude [11]–[14], phase [15]–[17], autocorrelation function (ACF) [18], AOA [19]–[21], TOF [22]–[24], and DFS [25], [26]. Hsu et al. propose a WiGait system to extract gait velocity based on the TOF of the major reflector [23]. It is shown that the coordinates of human target can be accurately localized and the moving length can be obtained. Liu et al. propose a WiRun system to process the signal amplitude with canonical polyadic (CP) decomposition and obtain the moving distance with the human gait estimation [14]. Zhang et al. design a Wispeed system to characterize the relationship between signal power ACF and velocity [18], while Wu et al. establish a WiDir model based on Fresnel Zone to obtain AOA of moving target to estimate the walking direction [19].

Learning based techniques have also been utilized for human localization and motion sensing [27]–[33]. WiWho utilizes the gait information to identify a person via training a classifier, which achieves an identification accuracy of 92% under two-participant scenario [27]. CARM builds the correlation between signal dynamics and human activity, and realizes steady human activity recognition with the hidden markov model (HMM) [30]. Zheng et al. develop a GRU-based framework to realize cross-domain gesture recognition [31], while Wang et al. leverage LSTM to process the radio signals to achieve activity recognition [32]. BiLoc builds a fingerprint-based indoor localization system with the deep auto-encoder network [33], which achieve stable localization compared with other systems.

Our system focuses on the speed estimation, and the most related work is Wispeed [18]. Different from [18], our system combines both advantages of model based techniques and learning based techniques. More specifically, model based techniques could work well in different environments without training data, but the performance may degrade when model mismatch happens. Learning based techniques are capable of extracting complex relationship between human activity and wireless signal. However, the performance may drop severely with environment change. To resolve these problems, our system first extracts the signal of interest with model based techniques. Then, we adapt a deep neural network to suppress the interference and improve the estimation accuracy. By combining both advantages of model based techniques and learning based techniques, our system achieves state-of-the-art performance for speed estimation.

TABLE I: Notation Table

Parameter	Description
λ	the wavelength of radio signal
M	the number of receiver antenna
K	the number of subcarrier
d	the space interval of uniform linear array
f_0	the center frequency of radio signal
Δf	the frequency interval of radio signal
L	the number of path
N	the number of AOA-TOF grid point
B	the number of link
H^i	the channel state information of link i
f_d^i	the frequency shift of the signal on link i
α_l^i	the complex attenuation factor of the l^{th} path on link i
d_l^i	the reflected path length of the l^{th} path on link i
ϕ_{mk}	the phase shift on the k^{th} subcarrier and the m^{th} antenna

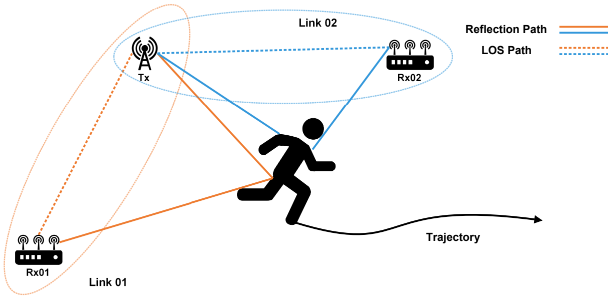


Fig. 1: System model.

III. System Model

In this paper, we consider an indoor environment with multiple links, where each link consists of one transmitter equipped with one transmitter antenna and one receiver equipped with multiple receiver antennas, as shown in Fig. 1. There is one target[1] moving in the environment[2], and the radio signals

[1] In this paper, we use target and human interchangeably.

[2] Our model can be applied into the scenario with multiple targets, where each target can be separated through the beamforming technique with the multiple receiver antennas. In the experimental results section, we will illustrate the results of such a multi-person scenario. Here, to give more insights, we assume there is only one target moving in the environment.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JIOT.2020.3022071, IEEE Internet of Things Journal

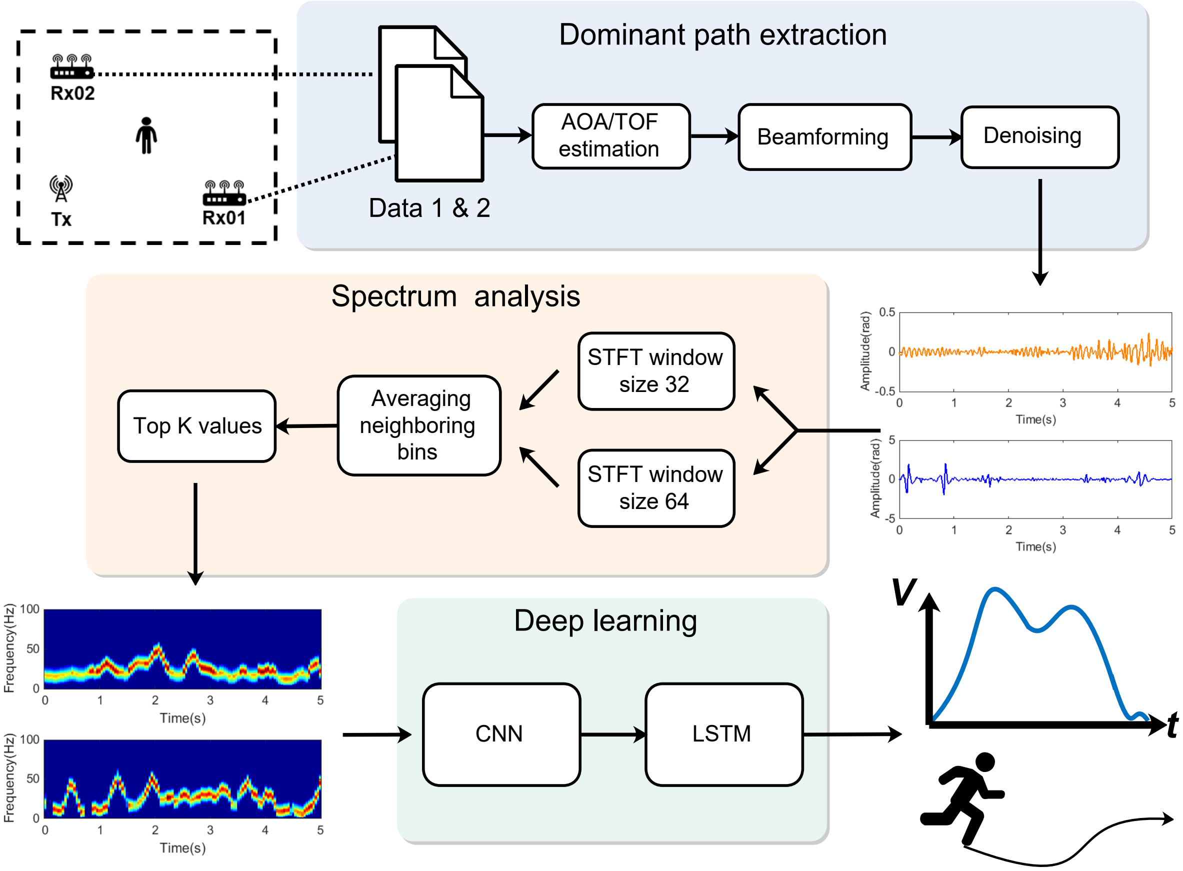


Fig. 2: The SpeedNet framework: three modules are included in the SpeedNet, which are the dominant path extraction module, the spectrum analysis module, and the deep learning module. The first module collects data and measures phase change, the middle module obtains frequency domain information and the third module extracts features to estimate the speed of the moving target.

emitted from the transmitter will propagate in the air, be reflected by target, and finally be received by the receiver. A summary of notations used in the following sections is given in Table I.

In an ideal case without the multipath and noise, the channel state information (CSI) of link i, $H^i(t)$, can be written as

$$H^i(t) = \alpha^i(t)e^{-j2\pi\frac{d^i(t)}{\lambda}}, \quad (1)$$

where t is the time index, λ denotes the wavelength, $\alpha^i(t)$ and $d^i(t)$ are the complex attenuation factor and the reflected path length of link i at time t, respectively. For detailed introduction of CSI, please refer to [14], [19], [25], [28].

From Eq(1), we can see that in the ideal case, the moving status of the target could be extracted directly from the phase of CSI, $\angle H^i(t)$. Moreover, the frequency shift of the signal on link i affected by the moving target can be written as

$$f_d^i = \frac{1}{2\pi}\frac{\mathrm{d}\angle H^i(t)}{\mathrm{d}t} = -\frac{1}{\lambda}\frac{\mathrm{d}d^i(t)}{\mathrm{d}t}. \quad (2)$$

The Eq(2) shows that there is a direct relationship between the doppler frequency shift (DFS) and the change of the reflected path length, both of which could be derived from the phase of the CSI in the ideal case. Note that the change of the reflected path length of link i is actually the speed of the moving target along a certain direction. Thus, in the ideal case, we can derive the speed of the moving target based on the change of the reflected path length of multiple links [20], [21], [24].

However, in practice, there exist multiple reflected paths in the indoor environment due to the multipath effect. Besides, the noise would be introduced at the receiver. Thus, the CSI of link i in a practical indoor environment could be written as

$$H^i(t) = \sum_{l=0}^{L}\alpha_l^i(t)e^{-j2\pi\frac{d_l^i(t)}{\lambda}} + n^i(t), \quad (3)$$

where l is the path index, L denotes the number of path, $\alpha_l^i(t)$ and $d_l^i(t)$ are the complex attenuation factor and the reflected path length of the l^{th} path on link i at time t, respectively, $n^i(t)$ is the noise on link i at time t.

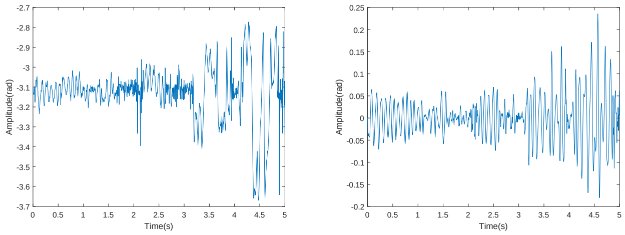
The CSI in Eq(3) involves the time-varying dominant paths which are reflected by the moving target, and the time-invariant paths which are reflected by the static objects as well as line-of-sight (LOS) path. In such a case, we cannot directly obtain the DFS as in Eq(2), which corresponds to the moving target, from the phase information of the CSI. Moreover, due to the estimation errors, we cannot directly obtain the speed of the moving target from the estimated DFS of multiple links. Therefore, to estimate the speed of the moving target from the CSI information, there are two challenges needed to be resolved: a) to extract the dominant path signal reflected by the moving target from the CSI information, and obtain the corresponding DFS of each link; b) to estimate the time-varying speed based on the extracted DFSs of different links.

IV. SpeedNet

To resolve the above challenges, in this paper, we propose an indoor speed estimation framework, SpeedNet, as shown in Fig. 2. The SpeedNet consists of three modules, which are the dominant path extraction module, the spectrum analysis module, and the deep learning module. The dominant path extraction module collects CSI data, jointly estimates the AOA-TOF through beamforming technique, and eliminates the noise through filtering. The spectrum analysis module estimates the DFS of each link with STFT, and generates training data for the deep learning module through analyzing the DFS spectrogram. Deep learning module is composed of CNN and LSTM to extract the spatial and temporal features to estimate the speed of the moving target. In the following subsections, we will introduce these three modules one-by-one in detail.

A. *Dominant path extraction*

As shown in Eq(3), due to the multipath effect, the dominant path reflected by the moving target is generally mixed with the signals from other paths. To obtain the DFS corresponding to the moving target, we need to first extract the dominant path from the received CSIs. Fig. 3 illustrates the detail of dominant path extraction module. In the following, without loss of generality, we omit the link index i for conciseness unless otherwise noted.



(a) Phase before the S-G filter

(b) Phase after the S-G filter

Fig. 4: An illustration of S-G filter.

Let θ_l and τ_l denote the AOA and TOF of the l^{th} path, respectively. We assume that M receiver antennas form a uniform linear array with space interval d, and the radio signal is transmitted on K subcarrier with center frequency f_0 and frequency interval Δf [7]. In such a case, the phase shift of the signal from θ_l on adjacent antennas is given by

$$\Phi(\theta_l) = e^{\frac{-j2\pi f_0 d \cos\theta_l}{c}}, \tag{4}$$

and the phase shift of the signal from τ_l on adjacent frequencies can be expressed as

$$\Phi(\tau_l) = e^{-j2\pi\Delta f \tau_l}. \tag{5}$$

Combining Eq(4) and Eq(5), the phase shift on the k^{th} subcarrier and the m^{th} antennas of l^{th} path can be expressed as

$$\phi_{mk}(\theta_l,\tau_l) = \exp\Bigg(-j2\pi\Big((k-1)\Delta f \tau_l + f_0\frac{(m-1)d\cos\theta_l}{c} \\ +(k-1)\Delta f\frac{(m-1)d\cos\theta_l}{c}\Big)\Bigg), \tag{6}$$

where the first term corresponds to the phase shift caused by AoA, the second one corresponds to the phase shift caused by ToF, and the third term is the cross term caused by both AoA and ToF.

Considering the system with M antennas and K subcarriers, the phase shift vector is then given by

$$a(\theta_l,\tau_l) = [1,\ \cdots,\ \phi_{mk}(\theta_l,\tau_l),\ \cdots,\ \phi_{MK}(\theta_l,\tau_l)]^T. \tag{7}$$

Assuming there are L paths in the environment, the steering matrix could be expressed as

$$A = [a(\theta_1,\tau_1),\ a(\theta_2,\tau_2),\ \cdots,\ a(\theta_L,\tau_L)]. \tag{8}$$

Thus, the measured CSI in practice could be written as

$$H(t) = As(t) + n(t), \tag{9}$$

where $s(t)$ is the transmitted signal, $n(t)$ is the noise introduced by the receiver, and $H(t)$ is the measured CSI which contains both amplitude and phase information.

To obtain the dominant path signal, we could first compensate the phase shift caused by AOA and TOF, and then add the signals on different antennas and frequencies. In this case, the signal from specific (θ_l,τ_l) would superimpose coherently while other signals would be averaged. To find the (θ,τ) of the dominant path, we select a set of candidate AOA-TOF values, and the steering matrix of these candidate AOA-TOFs can be expressed as

$$\hat{A} = [a(\hat{\theta}_1,\hat{\tau}_1),\ a(\hat{\theta}_2,\hat{\tau}_2),\ \cdots,\ a(\hat{\theta}_N,\hat{\tau}_N)], \tag{10}$$

where N denotes the number of candidate AOA-TOFs.

Then, the AOA-TOF spectrum can be obtained by applying $\hat{A}$ to the CSI data $H(t)$, i.e.,

$$P = \hat{A}^H H(t) = \hat{A}^H As(t) + \hat{n}(t), \tag{11}$$

where $A \in C^{MK\times L}$, $\hat{A} \in C^{MK\times N^2}$, $s \in C^{L\times 1}$, $\hat{n}(t) = \hat{A}^H n(t)$, P denotes the spectrum of AOA-TOF, and the amplitude in P corresponds to the signal strength from the candidate AOA-TOF. Hence, the amplitude of candidate AOA-TOF would be large if it is consistent with the AOA-TOF of the dominant path. The AOA-TOF of the dominant path is the one with the maximum amplitude by excluding the LOS path, i.e.,

$$(\theta^*,\tau^*) = \arg\max_{(\hat{\theta}_i,\hat{\tau}_i)\neq(\theta_0,\tau_0)} |P| = |\hat{A}^H H(t)|, \tag{12}$$

where (θ_0,τ_0) is the AOA-TOF of the LOS path. Then, the signal from the dominant path can be extracted as

$$\tilde{s}_d(t) = a(\theta^*,\tau^*)^H H(t). \tag{13}$$

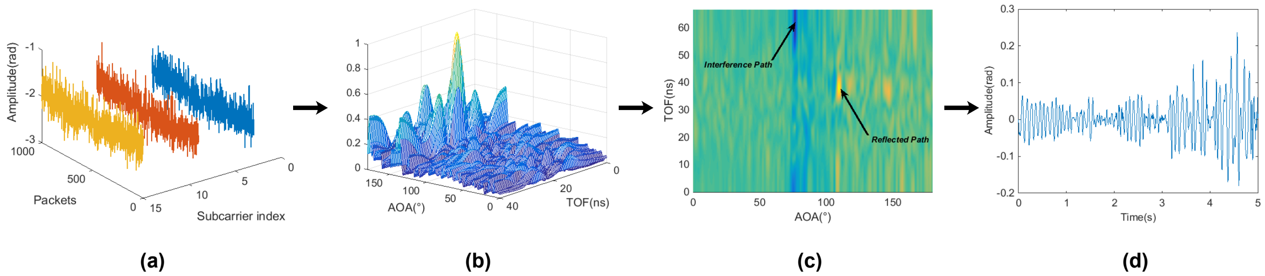


Fig. 3: The structure of dominant path extraction module: (a) the raw phase of radio signal; (b) the results of AOA and TOF estimation, where the maximum value is the location of the received signal; (c) beamforming with the estimated AOA-TOF, where the reflected path will be enhanced and the interference path will be restrained; (d) the phase of the dominant path signal after denoising.

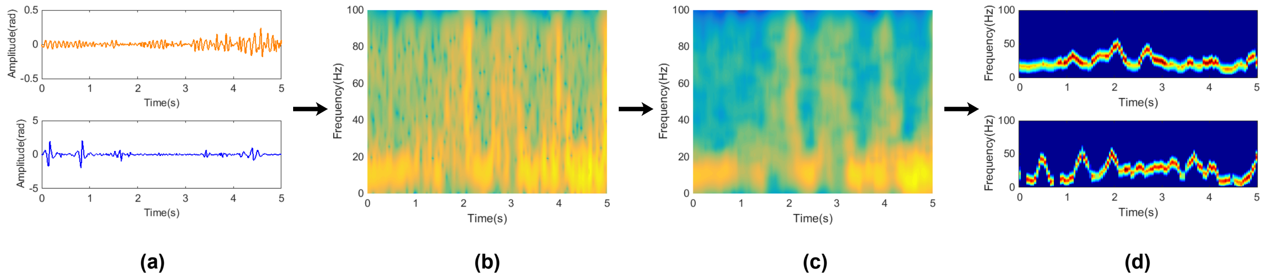


Fig. 5: The structure of spectrum analysis module: (a) two links phase data after denoising; (b) the STFT spectrogram of signal phase, where the window size is 32 and the FFT point is 512; (c) the spectrogram after averaging neighboring bins; (d) top-k values in the spectrogram, where the data have been filtered. Considering the limitation of time and frequency resolution, the maximum value may erroneous in some cases.

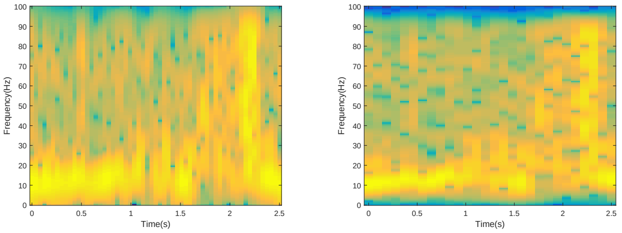
Since the DFS of the moving target lies within a certain range of frequency, we further utilize a S-G filter to remove the DC and high frequency component of phase variation. S-G filter, which is based on local polynomial least square to fit the data in time domain [34], is commonly used in data stream smoothing and denoising. An illustration of the S-G filtering process is shown in Fig. 4, where Fig. 4(a) and Fig. 4(b) are the phase before and after the S-G filter, respectively. We can see that the S-G filter can indeed remove the DC and high frequency component of the phase variation. The output of the dominant path extraction module, i.e., the extracted dominant path signal $\hat{s}_d(t)$, can be written as follows

$$\hat{s}_d(t) = SG\left(\tilde{s}_d(t)\right) = SG\left(a(\theta^*, \tau^*)^H H(t)\right), \tag{14}$$

where $SG(.)$ stands for S-G filtering.

B. Spectrum analysis module

As shown in Eq(2), there is a relationship between the DFS and the change of the path length reflected from the moving target. Therefore, we would first obtain the DFS corresponding to the moving target. To do so, we analyze the spectrum of the dominant path signal by applying the STFT to the time-domain signal obtained from the dominant path extraction module. Fig. 5 illustrates the detail of spectrum analysis module. To perform STFT, we need to choose a window size, which is used to balance the time resolution and frequency resolution of the spectrogram. As shown in Fig. 6, we illustrate the spectrogram with the window size 32 and 64, where we can see that a smaller window size leads to a better time resolution while a larger window size leads to a better frequency resolution. Thus, to take advantages of better resolution in both time and frequency domains, similar to [35], we perform STFT with several different window sizes to obtain different resolution spectrogram, and fuse them to obtain a better resolution spectrogram in both time and frequency domains.



(a) window size 32

(b) window size 64

Fig. 6: The spectrogram obtained using STFT with different window sizes.

Due to the limited time and frequency resolution of spectrogram, a continuous movement of target could result in the change of multiple neighboring time-frequency bins in spectrogram. Thus, for each time-frequency bin of the spectrogram, we propose to use the average of neighboring 0.1s and 1Hz bins to represent it. Moreover, since human generally move slowly in an indoor environment, the corresponding DFS lies within $5 \sim 95$Hz [29]. All the bins larger than

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JIOT.2020.3022071, IEEE Internet of Things Journal

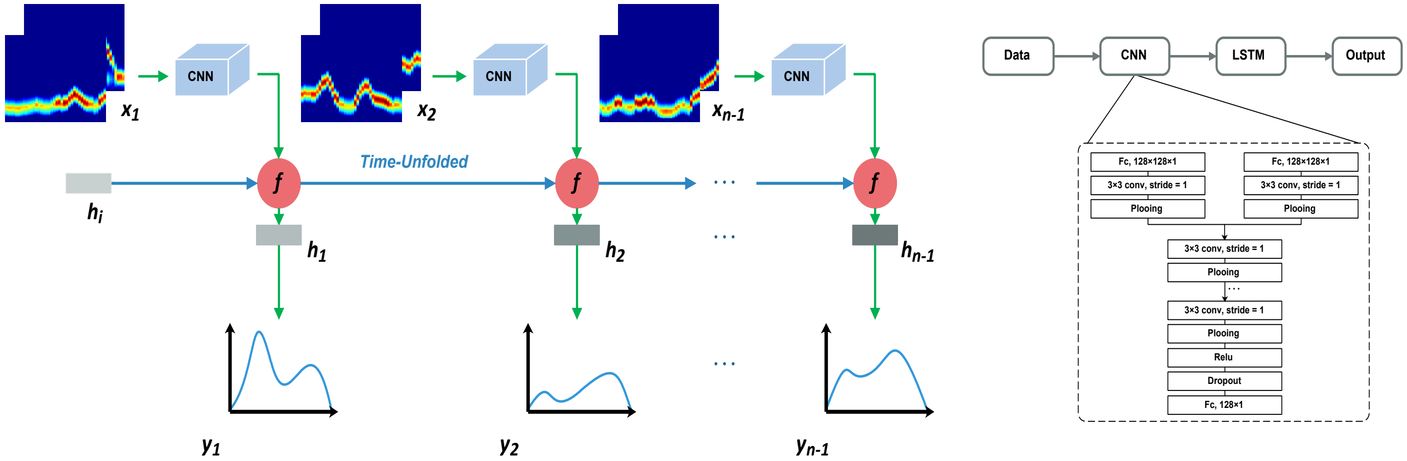


Fig. 7: The architecture of deep learning module: on the left is an expansion diagram of the deep learning module, where two links data will enter the network at each moment. The x_i and y_i represent the input of network and the speed, respectively. The f denotes the LSTM and h is the hidden layer. On the right is the detailed structure of CNN.

100Hz are filtered out. Due to the limited frequency resolution and random noise, the value with the highest amplitude in spectrogram could not provide full information of the human speed. Nevertheless, the useful information is generally hidden around the signal with the highest amplitude. Hence, we choose top-k values in the spectrogram to represent DFS. Larger k can provide better information of human speed while it would also increase the computation complexity. The output of the spectrum analysis module, i.e., the DFS spectrogram of link i, $\hat{F}_d^i(t)$, can be expressed as follows

$$\hat{F}_d^i(t) = SAM\left(\hat{s}_d(t)\right), \tag{15}$$

where $SAM(.)$ stands for the whole procedures of the spectrum analysis module.

C. Deep learning module

With the spectrum analysis module, we can derive the DFS spectrogram corresponding to the moving target. However, due to the estimation errors, we cannot directly obtain the speed of the moving target from the estimated DFS spectrogram. In this subsection, we propose to utilize a deep learning module, composed of CNN and LSTM as shown in Fig. 7, to extract the spatial and temporal features to estimate the speed of the moving target from the DFS spectrogram, i.e.,

$$\hat{v}(t) = D\left(\hat{F}_d^1(t), ..., \hat{F}_d^B(t)\right), \tag{16}$$

where $\hat{v}(t)$ is the estimated speed of the moving target, $D(.)$ is the whole procedures of the deep learning module, B is the number of links. Since the DFS of one link only represents part of the actual speed, we need to utilize the DFS of multi-links to jointly estimate the actual speed.

CNN has been shown to be very effective in extracting spatial feature from images for various computer vision tasks. In this paper, we utilize CNN to extract the spatial feature from the DFS spectrogram. As shown in Fig. 7, we use a network architecture similar to Visual Geometry Group (VGG) architecture [36]. With deeper network structure and smaller convolution kernel, VGG can not only guarantee the perception field, but also reduce the parameters of the convolutional (Conv) layer. Specifically, the DFS spectrogram on each link is first passed through one fully-connected (Fc) layer, one Conv layer and one pooling layer, and then combined together to go through the rest of the network. The network has included convolutional and pooling layers in each branch, which realizes feature extraction and dimension reduction at the same time. The components are connected to a dropout layer to further extract spatial feature from the merged maps. We choose ReLu as the activation function. The input size of CNN is 128 × 128 × 1 and the output of CNN network is then passed into the LSTM network to exploit the temporal feature of DFS spectrogram to estimate the speed. We use the standard LSTM architecture with "many to many" case. To obtain more temporal feature, we utilize three layout and set the hidden size as 512.

We consider an end-to-end training for both the CNN and LSTM network by minimizing the following loss function

$$\mathcal{L} = \mathcal{L}_s(v(t), \bar{v}(t)) + \beta \mathcal{L}_{TV}(v(t)), \tag{17}$$

where $\bar{v}(t)$ is the ground-truth speed, $\mathcal{L}_s(v(t), \bar{v}(t))$ is the smooth L_1 loss defined as follows

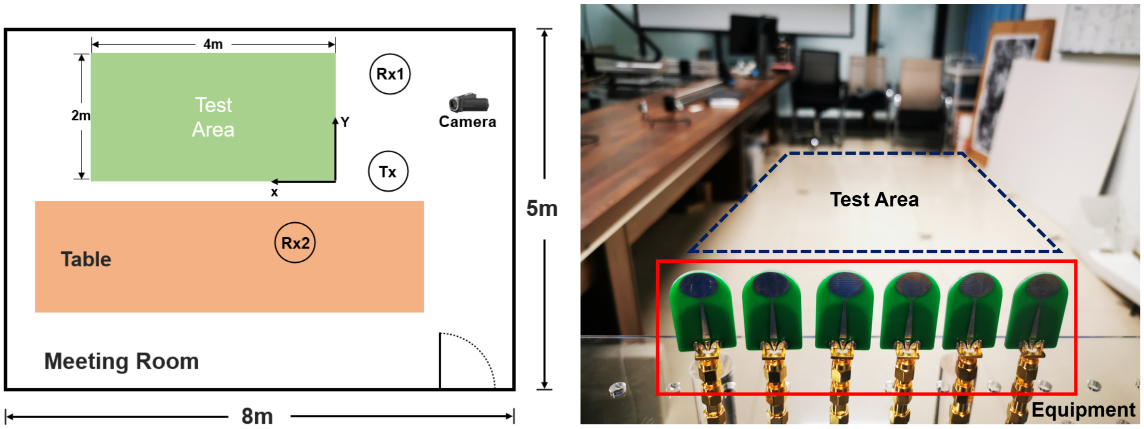
$$\begin{aligned}\mathcal{L}_s(v(t), \bar{v}(t)) = &\sum_{t=1}^{T} \Big[\mathbb{I}(|v(t) - \bar{v}(t)| \geq 1)\left(|v(t) - \bar{v}(t)| - 0.5\right) \\ &+ (1 - \mathbb{I}(|v(t) - \bar{v}(t)| \geq 1)) \frac{1}{2}|v(t) - \bar{v}(t)|^2\Big]\end{aligned} \tag{18}$$

where $\mathbb{I}$ is an indicator function. The smooth L_1 loss combines the advantage of L_1 loss and L_2 loss, which is insensitive to outlier and easy to control the magnitude of the gradient. $\mathcal{L}_{TV}(v(t))$ is the total variation of the speed defined as follows

$$\mathcal{L}_{TV}(v(t)) = \sum_{t=1}^{T-1} |v(t+1) - v(t)|, \tag{19}$$

and β denotes a regularization coefficient used to restrain the strength of $\mathcal{L}_{TV}(v(t))$.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JIOT.2020.3022071, IEEE Internet of Things Journal



(a) The layout of meeting room

(b) The equipment and test area

Fig. 8: The experimental setup and environment.

V. Experimental Results

In this section, we conduct extensive experiments to evaluate the performance of the proposed SpeedNet. To illustrate the superiority of the proposed framework, we compare it with the state-of-the-art methods.

A. Experiment Setup

We implement a multi-antenna ultra-wideband (UWB) transceiver system to conduct real-world experiments. Our system consists of one transmitter and two receivers. The transmitter is equipped with one omni-directional antenna while each receiver is equipped with a uniform linear antenna array which has 6 antenna elements. The two receivers receive the signal from the same transmitter simultaneously. In other words, we use 2 links in the experiments. The signals are transmitted on 201 subcarriers with 5.5GHz center frequency and 1GHz bandwidth, which corresponds to the frequency interval of 5MHz. The transceiver transmits 200 packets per second. In the dominant path extraction module, the AOA grid scans from 0° to 180° with $N_\theta = 128$ and the TOF grid scans from $0ns$ to $70ns$ with $N_\tau = 128$. In the spectrum analysis module, we perform STFT with window size 32 and 64 to take advantages of high resolution in both time and frequency domains. Besides, we choose k to be 5 to provide better information of human speed in spectrogram. We implement the deep learning module with Pytorch and train the deep neural network with Nvidia 1080Ti.

We conduct experiments in a meeting room with the layout shown in Fig. 8(a), where Tx and Rx denote the transmitter and receiver, respectively. One receiver is put on the table while the other is fixed on the holder, and participants move within the $2m \times 4m$ green area. Several chairs and tables are put in our meeting room to simulate an actual in-home environment. We ask participants to walk in the area to ensure that they could be recorded by the camera. Fig. 8(b) is the photograph of experiment environment.

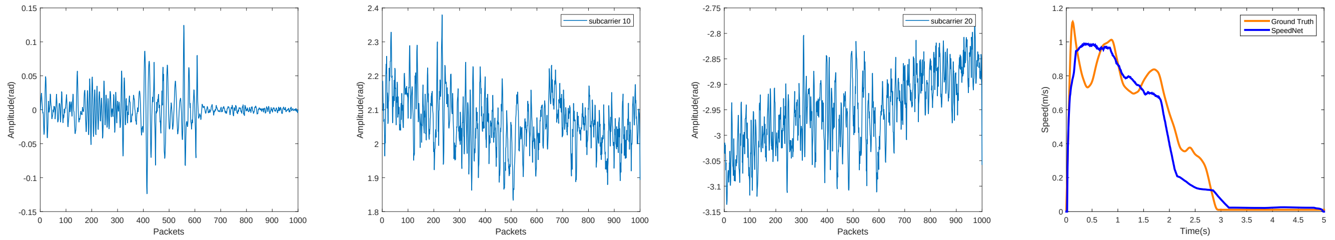
To obtain the groundtruth human speed, we install a digital camera to record the video of the moving human with resolution 1920×1080 at a rate of 30 fps, and then we adapt a visual tracking algorithm to extract the moving speed as ground-truth speed [38]. At the same time, we manually measure the walking distance of the human as ground-truth moving distance. In our system, our transceiver would display the start of signal transmitting on a monitor. We use the camera to simultaneously record the start of human movement and the start of signal transmitting to synchronize the time series.

During the experiment, one person controls the equipment and asks participants to walk and stop. We record 5 seconds of human moving data in each experiment. In total, we collect data over 200 times of experiment. These data cover different moving status of participants, including continuous walk, walk for a distance and then stop, stop randomly during the movement then continue walking. We select 80% of the data as the training dataset and 20% as the test dataset. The learning rate is set as 0.1 at the beginning, and it would be reduced after a certain number of iterations. We adapt SGD as the optimization algorithm and the dropout rate is set as 0.5 in CNN. Model parameters are initialized by the Kaiming initializer [37]. We train the data on training dataset and observe the value of loss function. We verify the model on the test dataset once the value of loss becomes stable.

B. Performance of SpeedNet

We first consider two scenarios with a single person walking: in the first scenario, the participant is asked to walk normally first and then stop at a random position; in the second scenario, the participant is asked to stop at a starting position first, and then walk normally. The results[3] are shown in Fig. 9 and Fig. 10, respectively. Since the DFS highly depends on the quality of the phase information, we present the phase of the signals obtained from the dominant path extraction module. For comparison, we also show the phase information on some specific subcarriers. As shown in Fig. 9(a) and Fig. 10(a),

[3] Due to errors existing in the measure of human speed in practice, it is inappropriate to compare the speed accuracy only. In our experiments, distance could be measured precisely, which could also be derived from speed. As the result, we compare both speed and moving distances in our experiments. The accuracy is for distance if not specified.



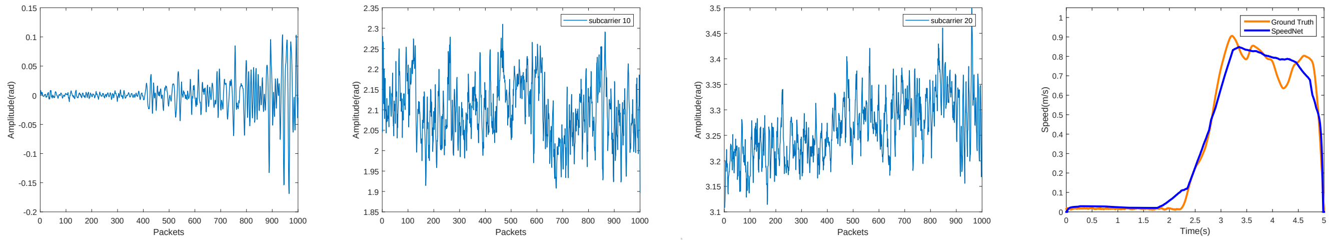
(a) Raw phase extracted from the dominant path

(b) Raw phase of the signal on subcarrier 10

(c) Raw phase of the signal on subcarrier 20

(d) The speed curve in this case

Fig. 9: The case when one participant moves normally first and then stops at a random position.



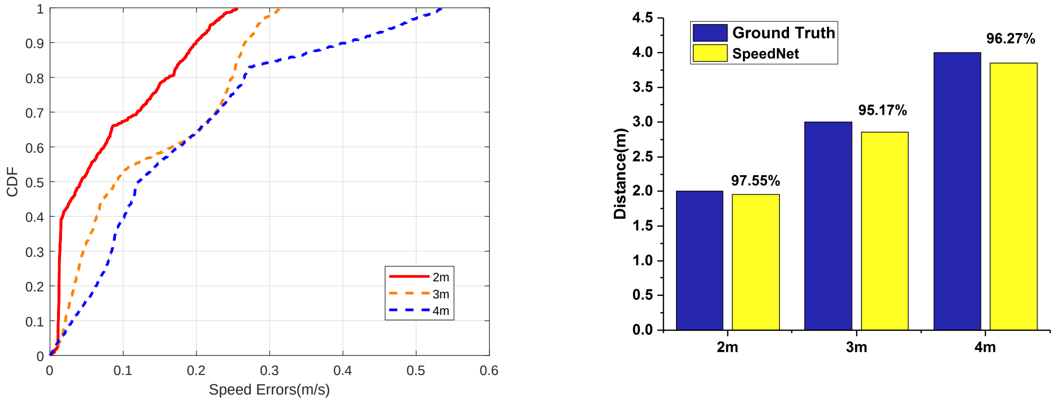
(a) Raw phase extracted from the dominant path

(b) Raw phase of the signal on subcarrier 10

(c) Raw phase of the signal on subcarrier 20

(d) The speed curve in this case

Fig. 10: The case when one participant stops at start point first and then moves normally.



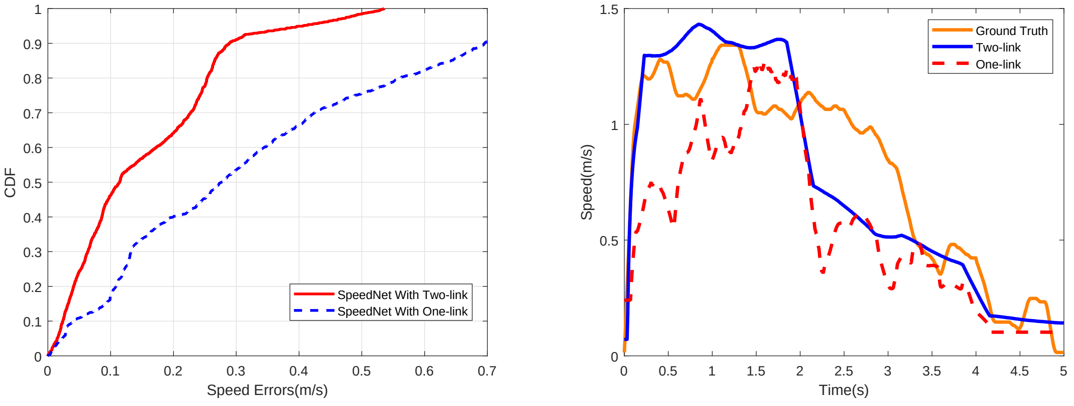
(a) Speed estimation errors

(b) Distance estimation errors

Fig. 11: Speed estimation accuracy with different moving distance.

with the dominant path extraction module, the phase variation matches well with the action of the participant, i.e., the phase variation is small when the participant stops, and becomes large when the participant moves. On the other hand, without the dominant path extraction module, the phase variation on different subcarriers does not match with the action of the participant, as shown in Fig. 9(b)-(c) and Fig. 10(b)-(c). This is because in practical indoor environment, the receivers suffer from severe multipath interference and random noise, which makes the raw signals on different subcarriers random and noisy; while the proposed method could deal with this problem by extracting signals from human based on the AOA and TOF, which could suppress the interference effectively. The estimated speed curves and the ground-truth are shown in Fig.9(d) and Fig.10(d). We can see that the proposed SpeedNet can well estimate the speed of the moving participant. We calculate the human moving distance based on the estimated human speed. In these two scenarios, the estimated moving distances are 1.87m and 1.82m, respectively, where the ground-truth distances are 1.94m and 1.77m, respectively.

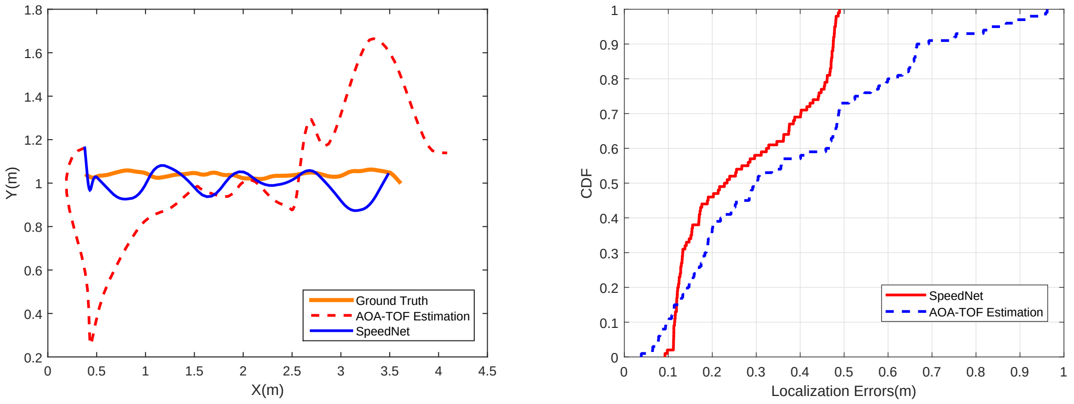
We then evaluate the effect of the walking distance on the estimation performance, where the participant is asked to start from the same position and walk 2m, 3m, and 4m, respectively. With the proposed SpeedNet, the speed at each time index is estimated, through which the walking distance is estimated. The speed estimation accuracy is shown in Fig. 11. Fig. 11(a) presents the cumulative distribution function (CDF) of the speed estimation errors and 11(b) shows the walking distance



(a) Speed estimation errors

(b) Speed curves

Fig. 12: Impact of using multi-links.



(a) Localization points

(b) Localization errors

Fig. 13: The performance of two different methods in localization.

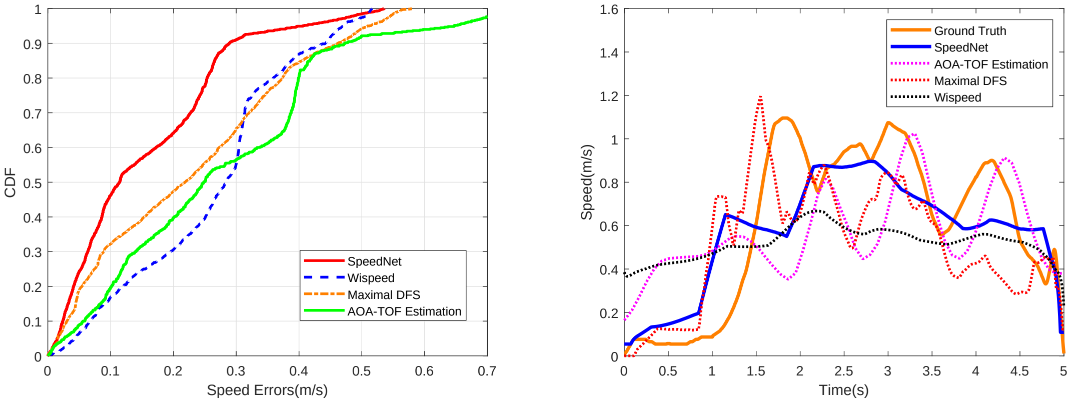
estimation errors. The median speed estimation errors in three scenarios are 0.05m/s, 0.09m/s and 0.13m/s, respectively. With the accurate human speed, the walking distance can be estimated with average accuracy of 96.33% in three scenarios.

Next, we evaluate the advantages of multiple links by comparing the one-link results with two-link results, and the results are shown in Fig. 12, where Fig. 12(a) shows the CDF of speed estimation errors and Fig. 12(b) shows the speed curves. From Fig. 12(a), we can see that with two links, the median speed error reduces from 0.28m/s to 0.12m/s, i.e., multiple links can significantly improve the performance of system. This is mainly because multiple links can provide more spatial diversity than one link, which improves the speed estimation accuracy. From Fig. 12(b), we can also see that the speed estimation with two-link radio signals is much better than that with one-link radio signals. The estimated moving distances are 3.84m and 2.82m for the two-link case and one-link case, respectively, where the ground-truth distance is 4m. This clearly demonstrates the advantages of multiple links.

C. Performance Comparison

In this subsection, we compare the proposed SpeedNet with three other methods [17], [18], [23]. The first one is “AOA-TOF estimation”, which directly localizes the moving participant by estimating the AOA-TOF at every time index and then calculates the moving speed based on the localization results. The second one is “maximal DFS”, which directly estimates the speed by finding the maximal DFS at every time index. This approach is actually the proposed SpeedNet without the deep learning module. The third one is “Wispeed”, which utilizes the relationship between the ACF of the power of the received electric field and the speed of motions.

We first compare SpeedNet with “AOA-TOF estimation”, and the results are shown in Fig. 13 and Fig. 14. In our system, the signal is transmitted with 1GHz bandwidth and received with 6 antennas, which makes it be able to estimate AOA-TOF accurately in the ideal case. However, due to the severe multipath interference in an indoor environment, the practical localization accuracy is not comparable with the ideal case. Fig. 13(a) shows the localization results, where the orange curve is the ground truth, while the blue curve is the results with SpeedNet, and the red dash curve is the results



(a) Speed estimation errors

(b) Speed curves

Fig. 14: The performance of four different methods in speed estimation.

with “AOA-TOF estimation”. The corresponding x direction and y direction are drawn in Fig. 8(a) and all the points have been transferred to the corresponding region. We can see that the variance of SpeedNet is much smaller than that of “AOA-TOF estimation”. Fig. 13(b) shows the CDF of the localization errors. We can see that SpeedNet achieves smaller localization error with median error 0.24m than “AOA-TOF estimation” with median error 0.29m. Fig. 14(a) illustrates the CDF curves of speed estimation errors. The median estimation error of “AOA-TOF estimation” is 0.25m/s while the SpeedNet achieves a much smaller error. The speed curve obtained by ”AOA-TOF estimation” is shown in Fig. 14(b), which suffers from severe multipath interference and random noise.

We then compare SpeedNet with “maximal DFS”, and the results are still shown in Fig. 14. From Fig. 14(a) we can observe that SpeedNet achieves much better estimation accuracy with a median error of 0.12m/s while the median error of “maximal DFS” is 0.21m/s. The estimated speed at each time index is shown in Fig. 14(b), where we can see that the estimated speed with SpeedNet matches better than that with “maximal DFS”. As for the walking distance, the ground-truth distance is 3m, while the estimated walking distances with SpeedNet and “maximal DFS” are 2.85m and 2.57m, respectively. These results clearly demonstrate the advantages of the deep learning network module.

Finally, we compare SpeedNet with Wispeed, and the results are also shown in Fig. 14(a). From the figure, we can see that the median speed estimation errors of SpeedNet and Wispeed are 0.12m/s and 0.28m/s, respectively. Fig. 14(b) shows the speed curve of Wispeed. In this scenario, the ground-truth walking distance is 3m, the estimated distance errors of SpeedNet and Wispeed are 0.14m and 0.38m, respectively. Therefore, by better exploiting both the spatial and temporal features in the radio signals, SpeedNet can achieve better performance than Wispeed.

D. Multi-person Scenario

The SpeedNet can also work in the multi-person scenario when different participants are separated by a certain distance.

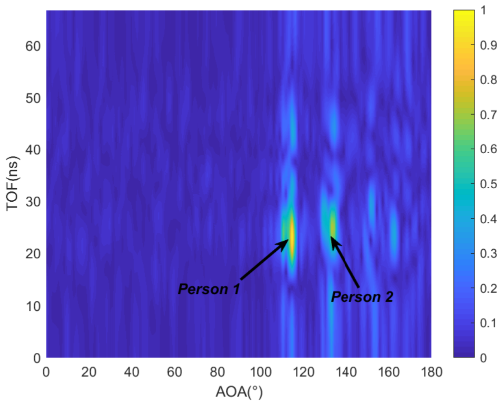
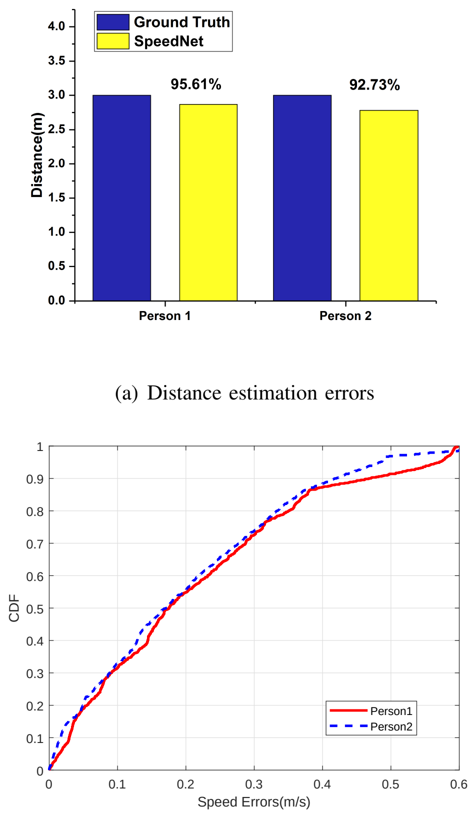


Fig. 15: The AOA-TOF spectrum under the two-participant scenario.

Due to the fact that signals from different targets have been separated in the dominant path extraction module, we do not need to train the neural net with different number of people. Fig. 15 illustrates the AOA-TOF spectrum in this scenario. Two participants are asked to walk the same distance (3m in this experiment) with different speeds. The performance of the moving distance estimation for the two participants are shown in Fig. 16(a). From the figure, we can see that the estimation of person 1 achieves 95.61 % accuracy, while the estimation of person 2 is 92.73 % accuracy. And the CDF of speed errors is shown in Fig. 16(b). We notice that these two curves are similar and the median errors are both 0.17m/s. The performance under the multi-person scenario is slightly worse than that under the single-person scenario, which is mainly due to the severer multipath effect in multi-person scenario.

E. Impact of Sampling Rate

In Fig. 17, we illustrate SpeedNet’s estimation accuracy under four different sampling rates. Due to the hardware limitation, the highest sampling rate we can achieve is 200Hz. We vary the sampling rate from 50Hz to 200Hz. From the figure, we can see that the estimation accuracy increases as



(a) Distance estimation errors

(b) Speed estimation errors

Fig. 16: Speed estimation accuracy under the two-participant scenario.

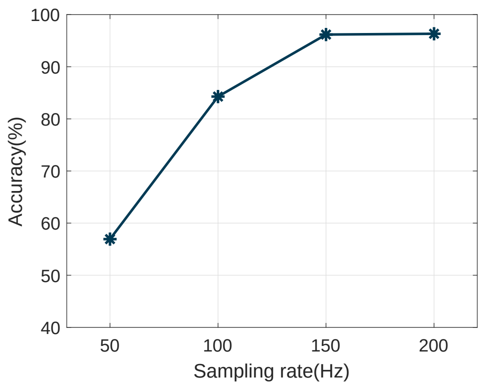


Fig. 17: Impact of sampling rate.

the sampling rate increases. Generally, human moves relatively slow in practical indoor environment and the DFS of human movement is less than 75Hz. According to the Nyquist Sampling Theorem, 150Hz sampling rate in time domain could sample the signal without aliasing. Hence, the performance of estimation under 150Hz is the same as that under 200Hz, which is much better than those under 50Hz and 100Hz.

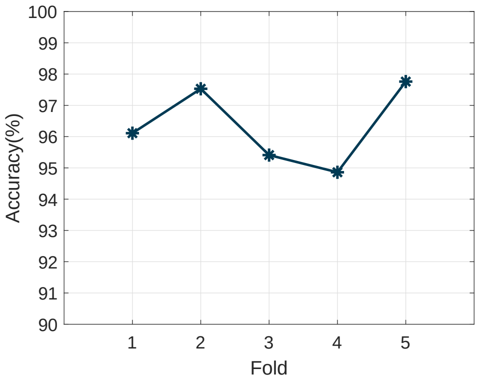


Fig. 18: The accuracy of speed estimation in each fold.

F. Cross Validation Results

In order to obtain stable model and avoid over-fitting, a five fold cross validation has been performed in model evaluation. Specifically, the initial data is divided into five folds randomly. Then, we select one fold as the data for model validation, and the rest is used for training. Next, we change the fold for validation and repeat this procedure five times until each fold has been utilized for validation once. The final accuracy is averaged over different folds. Fig. 18 shows the accuracy of speed estimation in each fold. We can observe consistent results among different folds.

G. Discussion on Combination of Traditional and Learning based Techniques

The combination of traditional techniques and learning based techniques is non-trivial. To demonstrate this, we compare the proposed SpeedNet with other three methods. The first one is a traditional method, named "AOA-TOF estimation", which directly localizes moving persons by estimating the AOA-TOF at every time index and then calculates the moving speed based on the localization results. The second one is a learning based method, named "Received Signal", which trains a deep neural network with the raw received signal as input. The third one is a combination of traditional and learning based techniques, named "Original AOA-TOF Spectrum", which transforms the raw signal into AOA-TOF domain first and then trains a deep neural network with the AOA-TOF domain signal as input. The deep neural networks in the second and third methods have the same architecture with the deep learning module in SpeedNet. The only difference is the number of parameters due to the different size of input.

The cumulative distribution function (CDF) of speed estimation error with these three methods and the proposed SpeedNet are shown in Fig. 19. As we can see, the proposed SpeedNet outperforms other three methods with significant performance improvement. For "AOA-TOF estimation" method, the traditional technique could localize moving person robustly. However, the accuracy is limited due to the multipath and many other interference in practical environment. For "Received Signal" method, although learning based techniques could extract complex relationship between wireless signal and

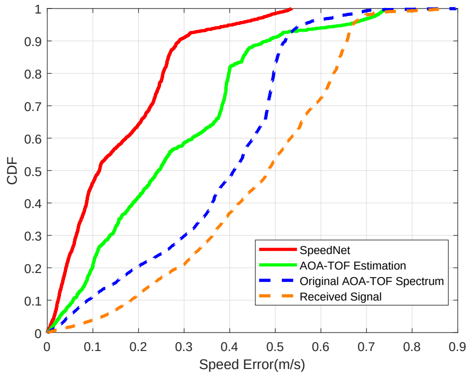


Fig. 19: Performance comparisons between different techniques.

human moving speed, adopting raw signal as input is not efficient which makes it difficult for the neural network to achieve accurate estimation. The ”Original AOA-TOF Spectrum” method combines the traditional and learning based techniques. However, its performance is not comparable with SpeedNet. This is because the ”Original AOA-TOF Spectrum” method takes the whole AOA-TOF domain signal as input, which contains much redundant information. On the contrary, SpeedNet focuses on the dominant path signal, which only contains valuable information for human speed estimation. It is worth noting that although ”Original AOA-TOF Spectrum” method combines traditional and learning based techniques, the performance is worse compared with “AOA-TOF estimation” which is a simple traditional method. This is because the amount of data for training the deep neural network is limited in practical experiments. Hence, it is difficult for the neural network to achieve impressive performance without appropriate design. In such a case, introducing deep neural network without proper design may even cause negative effect on the performance of the system.

VI. Limitations and Discussions

A. Direction Detection

Human speed is a vector with its direction. In our current system, we only obtain the magnitude of speed with the neural network, and estimate the speed direction by localizing the human target with AOA-TOF information, of which the accuracy is limited. In the future, we will study the deep neural network which can simultaneously predict the speed magnitude and direction.

B. Multi-person Detection

Sensing multiple targets simultaneously has been an important topic for wireless sensing. The proposed SpeedNet system could work in the multi-person scenario without additional training. However, it is limited by two facts. First, since human target is not a point reflector in practical environment, the reflected signal would exist in several AOA-TOF bins. Second, the signal separation performance with AOA-TOF is determined by the number of antennas and the width of signal bandwidth, which is limited in our system. As a result, we require participants to be separated by at least one meter in the experiments, which may not be true in some cases. We consider building a system with more antennas and larger bandwidth in our future work to alleviate this problem.

VII. Conclusion

In this paper, we have proposed a contactless indoor speed estimation framework, SpeedNet, to estimate the speed from the radio signals. We first adopted beamforming technique to extract the dominant path signal reflected from human, and then analyzed the STFT spectrogram of dominant path signal to obtain the DFS corresponding to the moving human. Finally, we utilized a deep neural network composed of CNN and LSTM to estimate the human moving speed from the spatial-temporal DFS. Experimental results showed that, compared with the state-of-the-art approaches, SpeedNet performs better in estimating the human moving speed with an average accuracy of 96.33% in a typical indoor environment.

References

[1] World report on ageing and health, https://www.who.int/

[2] T. Teixeira, D. Jung, G. Dublon and A. Savvides, ”PEM-ID: Identifying people by gait-matching using cameras and wearable accelerometers”, *IEEE ICDSC*, 2009.

[3] S. Pan, N. Wang, Y. Qian, I. Velibeyoglu, H. Y. Noh and P. Zhang, ”Indoor person identification through footstep induced structural vibration”, *ACM HotMobile*, 2015.

[4] S. E. Schaefer, C. C. Ching, H. Breen and J. B. German, ”Wearing, thinking, and moving: Testing the feasibility of fitness tracking with urban youth”, *American Journal of Health Education*, vol. 47, no. 1, pp.8-16, 2016.

[5] M. Khan, B. N. Silva and K. Han, ”Internet of Things based energy aware smart home control system”, *IEEE Access*, vol. 4, pp. 7556-7566, 2016.

[6] F. Adib, H. Mao, Z. Kabelac, D. Katabi and R. C. Miller, ”Smart homes that monitor breathing and heart rate”, *ACM CHI*, 2015.

[7] D. Zhang, Y. Hu, Y. Chen and B. Zeng, ”BreathTrack: Tracking Indoor Human Breath Status via Commodity WiFi”, *IEEE Internet of Things Journal*, vol. 6, no. 2, pp. 3899-3911, 2019.

[8] K. Qian, C. Wu, Z. Zhou, Y. Zheng, Z. Yang and Y. Liu, ”Inferring motion direction using commodity wi-fi for interactive exergames”, *ACM CHI*, 2017

[9] Q. Xu, Y. Chen, B. Wang and K. J. R. Liu, ”TRIEDS: Wireless events detection through the wall”, *IEEE Internet Things Journal*, vol. 4, no. 3, pp. 723-735, 2017.

[10] I. Bisio, A. Delfino, A. Grattarola, F. Lavagetto and A. Sciarrone, ”Ultrasounds-based Context Sensing Method and Applications over the Internet of Things”, *IEEE Internet of Things Journal*, vol. 5, no. 5, pp. 3876-3890, 2018.

[11] Z. Wu, Q. Xu, J. Li, C. Fu and Q. Xuan, ”Passive Indoor Localization Based on CSI and Naive Bayes Classification”, *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, vol. 48, no. 9, pp. 1566-1577, 2017.

[12] W. Wang, A. X. Liu and M. Shahzad, ”Gait recognition using wifi signals”, *ACM UbiComp*, 2016.

[13] C. Wu, Z. Yang, Z. Zhou, X. Liu, Y. Liu and J. Cao, ”Non-Invasive Detection of Moving and Stationary Human With WiFi”, *IEEE Journal on Selected Areas in Communications*, vol. 33, no. 11, pp. 2329-2342, 2015.

[14] M. Liu, L. Zhang, P. Yang, L. Lu and L. Gong, ”Wi-Run: Device-free step estimation system with commodity Wi-Fi”, *Journal of Network and Computer Applications*, vol. 143, pp. 77-88, 2019.

[15] S. Sen, B. Radunovic, R. R. Choudhury, and T. Minka, ”You are facing the Mona Lisa: spot localization using PHY layer information”, *ACM MobiSys*, 2012.

[16] M. Kotaru, K. Joshi, D. Bharadia, and S. Katti, ”Spotfi: Decimeter level localization using wifi”, *ACM SIGCOMM*, 2015.

[17] J. Wang, J. Tong, Q. Gao, Z. Wu, S. Bi and H. Wang, ”Device-Free Vehicle Speed Estimation With WiFi”, *IEEE Transactions on Vehicular Technology*, vol. 67, no. 9, pp. 8205-8214, 2018.

[18] F. Zhang, C. Chen, B. Wang and K. J. R. Liu, ”WiSpeed: A Statistical Electromagnetic Approach for Device-Free Indoor Speed Estimation”, *IEEE Internet Things Journal*, vol. 5, no. 3, pp. 2163-2177, 2018.

[19] D. Wu, D. Zhang, C. Xu, Y. Wang and H. Wang, ”WiDir: walking direction estimation using wireless signals”, *ACM UbiComp*, 2016.

[20] K. Qian, C. Wu, Y. Zhang, G. Zhang, Z. Yang and Y. Liu, ”Widar2.0: Passive human tracking with a single wi-fi link”, *ACM MobiSys*, 2018.

[21] Y. Xie, J. Xiong, M. Li and K. Jamieson, ”mD-Track: Leveraging multi-dimensionality for passive indoor Wi-Fi tracking”, *ACM MobiCom*, 2019.

[22] F. Adib, Z. Kabelac, D. Katabi and R. C. Miller, ”3D tracking via body radio reflections”, *USENIX NSDI*, 2014

[23] C. Y. Hsu, Y. Liu, Z. Kabelac, R. Hristov, D. Katabi and C. Liu, ”Extracting gait velocity and stride length from surrounding radio signals”, *ACM CHI*, 2017.

[24] X. Li, D. Zhang, Q. Lv, J. Xiong, S. Li, Y. Zhang and H. Mei, ”IndoTrack: Device-free indoor human tracking with commodity Wi-Fi”, *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, vol. 1, no. 3, pp. 72, 2017.

[25] K. Qian, C. Wu, Z. Yang, Y. Liu and K. Jamieson, ”Widar: Decimeter-level passive tracking via velocity monitoring with commodity Wi-Fi”, *ACM MobiHoc*, 2017.

[26] Q. Pu, S. Gupta, S. Gollakota and S. Patel, ”Whole-home gesture recognition using wireless signals”, *ACM MobiCom*, 2013.

[27] Y. Zeng, P. H. Pathak and P. Mohapatra, ”WiWho: wifi-based person identification in smart spaces”, *IEEE IPSN*, 2016.

[28] Y. Wang, K. Wu and L. M. Ni, ”Wifall: Device-free fall detection by wireless networks”, *IEEE Transactions on Mobile Computing*, vol. 16, no. 2, pp. 581-594, 2016.

[29] W. Wang, Alex. X. Liu, M. Shahzad, K. Ling and S. Lu, ”Understanding and modeling of wifi signal based human activity recognition”, *ACM MobiCom*, 2015.

[30] W. Wang, Alex. X. Liu, M. Shahazad, K. Ling and S. Lu, ”Device-Free Human Activity Recognition Using Commercial WiFi Devices”, *IEEE Journal on Selected Areas in Communications*, vol. 35, no. 5, pp. 1118-1131, 2017.

[31] Y. Zheng, Y. Zhang, K. Qian, G. Zhang, Y. Liu, C. Wu and Z. Yang, ”Zero-Effort Cross-Domain Gesture Recognition with Wi-Fi”, *ACM MobiSys*, 2019.

[32] F. Wang, W. Gong and J. Liu, ”On Spatial Diversity in WiFi-Based Human Activity Recognition: A Deep Learning-Based Approach”, *IEEE Internet of Things Journal*, vol. 6, no. 2, pp. 2035-2047, 2018.

[33] X. Wang, L. Gao and S. Mao, ”BiLoc: Bi-Modal Deep Learning for Indoor Localization With Commodity 5GHz WiFi”, *IEEE Access*, vol. 5, pp.4209-4220, 2017.

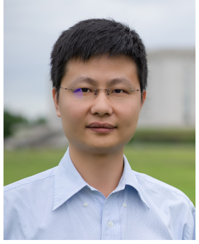
[34] A. Savitzky and M. J. E. Golay, ”Smoothing and differentiation of data by simplified least squares procedures”, *Analytical chemistry*, vol. 36, no. 8, pp. 1627-1639, 1964.

[35] T. Kurashima, T. Althoff, and J. Leskovec, ”Stfnets: Learning sensing signals from the time-frequency perspective with short-time fourier neural networks”, *ACM WWW*, 2019.

[36] K. Simonyan, A. Zisserman, ”Very deep convolutional networks for large-scale image recognition”, *arXiv:1409.1556*, 2014.

[37] K. He, X. Zhang, S. Ren and J. Sun, ”Delving Deep into Rectifiers: Surpassing Human-Level Performance on ImageNet Classification”, *ACM ICCV*, 2015.

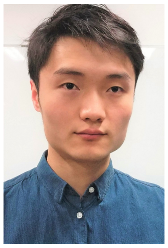
[38] L. Bertinetto, J. Valmadre, S. Golodetz, O. Miksik and P. H. S. Torr, ”Staple: Complementary learners for real-time tracking”, *ACM CVPR*, 2016.



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.

Hongyu Deng received the B.S. degree from the School of Information and Communication Engineering, University of Electronic Science and Technology of China, Chengdu, China, in 2018. He is currently pursuing the M.S. degree at the School of Information and Communication Engineering, University of Electronic Science and Technology of China. His current research interests include wireless sensing, networking and signal processing



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.