The 14th International Conference on
Computer Science & Education (ICCSE 2019)
August 19-21, 2019. Toronto, Canada
TuesP3.15

CSIID: WiFi-based Human Identification via Deep Learning

Ding Wang, Zhiyi Zhou, Xingda Yu, Yangjie Cao*
School of Software and Applied Technology
Zhengzhou University
Zhengzhou, China
caoyj@zzu.edu.cn

***Abstract*—With the widespread popularization of commercial off-the-shelf (COTS) WiFi devices, the device-free WiFi sensing has attracted attention extensively. However, there are only a few studies on human identification by using noncontact techniques since traditional methods are facing the problem of heavy workload and low recognition accuracy. Aiming at these issues, we propose a deep learning method, named CSIID, to analyze the gait features using Channel State Information (CSI) of COTS WiFi devices In CSIID, the convolution layers are combined with long short-term memory (LSTM) layers to extract gait features automatically from CSI data and to identify persons, which effectively reduces the need for a large amount of data preprocessing by manual feature extraction. Experimental results conducted on CSI data collected from different situations indicate that the CSIID has desirable identification accuracy. The average identification accuracy of CSIID is ranging from 97.4% to 94.8% when the number of persons is from 2 to 6.**

Keywords—Contactless Human Identification, CSI, Deep Learning

I. Introduction

Human identification plays an important part in human-computer interaction. It is a normal and traditional way to use biological features, such as fingerprint, face and iris, to identify people. However, Due to the requirement of close-range or contact-based perception, the use of those features is very limited in practical applications. Therefore, long-distance human identification researches have attracted a great deal of attention from computer researchers. Gait, which is also a kind of biological features, can be perceived in long-distance situations. And it has been considered as a unique signature of humans [1]. What's more, it is contactless and difficult to hide, which helps to make it one of the hot research topics in recent years.

Numerous attempts have been made to develop methods for gait recognition by using cameras [2] and sensors [3]. However, these methods involve privacy concerns or deployment condition restriction. With the coverage of WiFi signals in indoor environment, device-free WiFi waves propagated from a transmitter, they bump into objects and go through multiple reflection, absorption, scattering or the effects caused by human or objects before reaching the receiver [4], which provides a new technical solution for identification. WiFi-based sensing helps us to achieve the goal through passive detection. Furthermore, the availability of CSI for COTS 802.11n WiFi devices further promotes the use of device-free sensing [5].

Recently, several researches have shown the potential of device-free WiFi sensing. Since specific human behavior can cause unique interference to wireless signals, researchers can use WiFi to detect human activities [6]. DeMan used human breathing patterns to detect moving and stationary people in indoor environment [7]. Wang et al. used CSI phase for angle of arrival (AOA) estimation to implement indoor localization [8]. There are also lots of studies using wireless signals to identify keystrokes [9], detect hand gestures [10], detect fall [11], and monitor sleep [12]. Each person's gait affects the surrounding WiFi signals in a unique way as people are different in shape and movement patterns. And this unique influence will be reflected on the CSI time series.

At present, there are still many challenges in WiFi-based identification. For instance, there are the complex multipath effects in the indoor environment. And a large amount of data preprocessing is often required before the manual extraction of gait features. What's more the feature selection has an important influence on the identification accuracy. The ability of extracting features automatically with deep learning can be used to solve those problems well. Deep learning has performed well in pattern recognition recently due to its powerful capability of feature extraction.

In this paper, a concise and effective deep learning method, named CSIID, is proposed to replace widely used heuristics and identify people automatically. Compared with traditional methods of manual feature extraction, CSIID use CSI time series as the input to extract gait features without a significant amount of data preprocessing as denoising and signal separation. At the stage of gait features extraction, CSIID eliminates the need for the manual feature extraction by combining Convolution layers with LSTM layers to extract features automatically. Finally, using the softmax regression classifier to make a classification and output the result.

The remaining of the paper is organized as follows. Section II provides related work of human identification efforts. The overview of CSIID method is provided in Section III. Then Section IV describes CSI data preprocessing. Section V introduces the CSIID deep learning model we proposed. Then experiment and evaluation results are presented in Section VI and conclusion showed in Section VII.

II. Related Work

Techniques of using gait to implement human identification can be mainly divided into three categories: camera-based, sensor-based and WiFi-based. Deep learning methods have been widely used in camera-based and sensor-based human identification and have performed well.

Camera-based human identification first generates a sequence of human silhouettes from the video frames via separating moving targets from background, and then applies the sequence to implement recognition [13]. Wu et al. used deep CNNs to implement gait recognition via a pretty small group of labeled multi-view human walking videos [14]. In the literature, authors proposed the use of convolutional LSTM based networks to learn a video-based representation for people re-identification [15]. Human identification based on cameras has achieved high accuracy, but it brings privacy problems. Moreover, the line-of-sight (LOS) of cameras are easy blocked and they have certain requirements for lighting. Therefore, camera-based human recognition has limitations in practical applications.

Sensor-based human identification usually places the sensor device on the ground or the people's bodies to collect information and then convert it into a certain signal. Robert et al. used the ground-based sensor solution to create a user footprint model which based on the footprint profile and achieved a recognition accuracy of 93% [3]. Deep Sense employed sensors on mobile devices to identify user via integrating convolutional and recurrent neural networks [16]. However, all of the sensor-based approaches above require people to wear the sensors in a specific way, or deploy additional sensors, which tends to make users uncomfortable and cost a lot.

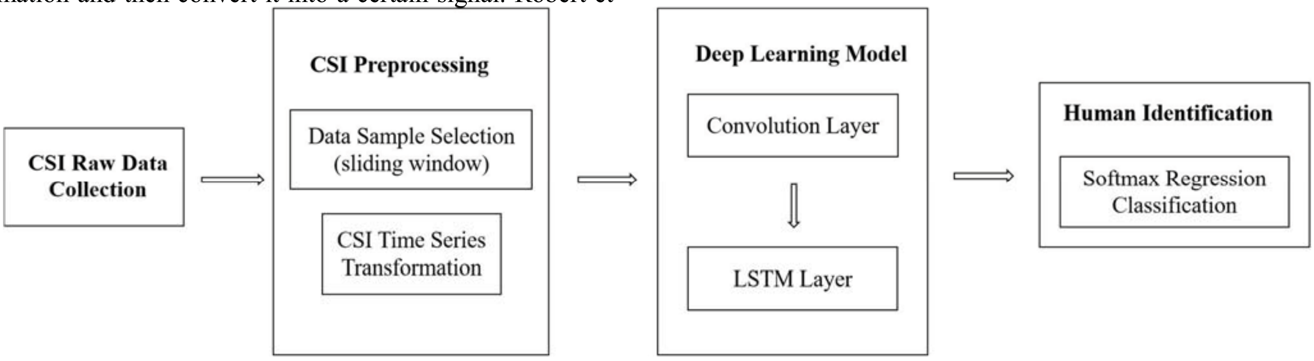

Fig.1: Overview of CSIID

Due to the high sensitivity of WiFi to environmental differences caused by moving targets, researchers began to explore the possibility of using WiFi for human recognition. In WiWho [5], after removing the long-distance multipath and high frequency noise in CSI data, WiWho extracted the step and walking features of a person from CSI samples to construct human gait signatures, and identified a person with average accuracy of 92% to 80% from a group size of 2 to 6 people respectively. Xin et al. proposed a method to extract features of LOS using Principal Component Analysis (PCA) and Dynamic Time Warping (DWT) and classify using K-nearest neighbor (KNN) classifier. It obtains 88.9% accuracy in the case of six people [17].WiFi-ID used Continuous Wave Transformation (CWT) to extract signals in different frequency bands, and used the RelieF feature selection algorithm to calculate the time-domain and frequency-domain characteristics of the select frequency band as a feature database for each person, increasing from 93% to 77% in groups of 2 to 6 people respectively [18].The above methods generally require a large amount of data preprocessing and manual extraction of gait features, which requires a large amount of workload. And the selection of features has an important impact on the recognition accuracy.

Different from those methods, our paper proposes a deep learning method for human identification, which eliminates the need for manual feature extraction and massive data preprocessing. This method can extract gait features automatically and identify them via deep learning method, and achieve desirable recognition accuracy.

III. Overview of CSIID

CSIID considers the indoor environment as the operational setting. Experimental devices consist of a transmitter and a receiver. The transmitter periodically sends packets, while the receiver captures packets and extracts CSI information. Subjects to be identified are asked to walk along a straight line on a predetermined path to construct the training set and test set. In the training phase, we input the processed CSI data into the neural network to train parameters so as to build gait feature database. In the testing phase, the collected CSI data is preprocessed in the same way, which is then input into the trained model to uniquely identify the subject.

The outline of CSIID is showed in Fig.1. CSIID consists of two parts: CSI preprocessing and gait features extraction with deep learning model. In the data preprocessing, we extract the CSI amplitude and ignore the CSI phase to remove the influence of the carrier frequency offset (CFO) and construct our dataset by expressing the amplitude as CSI time series. Furthermore, in order to learn the relationships better in CSI time series, we use the sliding window method to select data samples as input to our model. In the stage of gait features extraction in CSI time series, we select the CNN and LSTM as the hidden layers of the network to learning CSI time series features. Then we use the features as the input of softmax regression features to classify and determine the targets' identity.

IV. CSI DATE PREPROCESSING

A. *CSI Time Series Transformation*

Many WiFi devices support the IEEE 802.11n/ac standard and have multiple antennas for Multiple-Input Multiple-Output (MIMO) communications. These devices use Orthogonal Frequency Division Multiplexing (OFDM) at the physical layer. In OFDM, the channel is divided into multiple subcarriers, and CSI information represents the amplitude and phase information of the OFDM subcarriers [19]. Unlike Received Signal Strength Indication (RSSI) that represents the total received signal strength at the receiver, CSI contains information of each subcarrier between each pair of transmitting and receiving antennas. Therefore, CSI can describe the channel characteristics of the communication link and distinguish multipath propagation at the subcarrier level.

With COTS WiFi devices, K set of subcarriers in CSI can be collected for each antenna pair, and each CSI contains amplitude and phase information of subcarriers which can be expressed as:

$$\mathrm{H}\left(f_k\right) = \| H(f_k) \| e^{j\angle \mathrm{H}(f_k)}, k \in [1, K]. \tag{1}$$

Where $\mathrm{H}\left(f_k\right)$ is the k-th subcarrier of CSI, which is a complex value. $\| H(f_k) \|$ and $\angle H\left(f_k\right)$ indicate the amplitude and phase, respectively. In our work, the amplitude of CSI is used to extract gait features, let $H_k = \| H(f_k) \|$, so the CSI sequence for each antenna pair can be expressed as:

$$H = [H_1, H_2, ..., H_k]. \tag{2}$$

In MIMO systems, N_t and N_r represent the number of transmitting and receiving antennas respectively, which form $N_t \times N_r$ antennas pairs. Let $N = N_t \times N_r$ CSI sequences of the n-th antenna pair can be represented as:

$$H^n = [H_1^n, H_2^n, ...H_k^n], n \in [1, N], \tag{3}$$

where H_k^n represents the k-th subcarrier of CSI sequence between the n-th antenna pair. In order to identify the person in the detection area, we need to collect continuous CSI data measured at regular intervals and take the collected M CSI sequences in a period of time as a group of data, where $H' = [H_k^1, H_k^2, ...H_k^n]$ represents the set of k-th subcarrier for all antennas, CSI time series can be repressed as:

$$\begin{bmatrix} H'_{11} & H'_{12} & \cdots & H'_{1k} \\ H'_{21} & H'_{22} & \cdots & H'_{2k} \\ \vdots & \vdots & \ddots & \vdots \\ H'_{m1} & H'_{m2} & \cdots & H'_{mk} \end{bmatrix}, m \in (1, M), k \in (1, K),$$

where H'_{mk} indicates the set of k-th subcarrier at the m-th time.

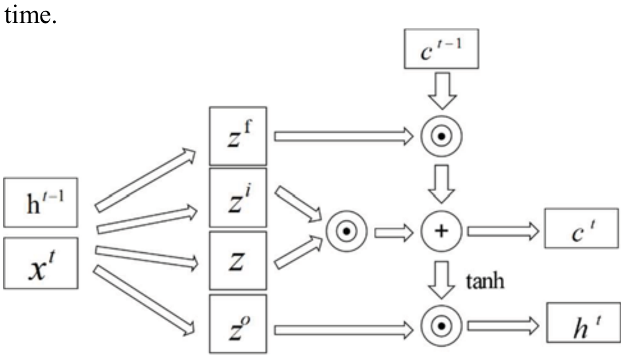

Fig.2 The Structure of LSTM

B. Data Sample Selection

Each group of CSI time series data contains gait features. In order to facilitate feature extraction of neural network and study the time relationship better in CSI time series, CSIID uses the sliding window processing method to select the data sample as the input of neural network.

According to experimental requirements, we select 500ms as the size of the sliding window and set 200ms as the step size. In this way, two adjacent samples can maintain a certain overlap ratio and a continuous CSI time series can be divided into overlapping short time windows. Then, we independently extract features and perform identification for each overlapping short time window. Ultimately, these short time windows' recognition results are combined to produce the final result of the entire CSI time series.

V. Deep Learning Model of CSIID

The deep learning model proposed in this paper is shown in Fig.1. The basic structure includes the Convolution layers and LSTM layers. The gait features in the CSI time series are extracted through the two layers. Then softmax regression is employed to train the classifier. The following sections will describe how each layer works in detail.

A. Convolution layers

CSIID uses three convolutional layers, which are used to extract the space features in CSI signals. Using the sliding window to select CSI sequences collected in a certain period of time as the input of the model:

$$\begin{aligned} S(p,q) &= (X * C)(p,q) \\ &= \sum_j \sum_k X(j,k)C(p-j, q-k). \end{aligned} \tag{4}$$

The input is a 500 by 90 matrix in the first convolution layer. In order to extract features from each subcarrier of different antenna pairs, 30 100*3 convolution kernels are used to convolute the input data. And the step length of time dimension (vertical direction) and subcarrier dimension (horizontal direction) are set as 1 and 3, so as to obtain 30 feature graphs. Similar processing methods are applied to the second and third convolution layers, and 30 feature graphs with the size of 203*22 are finally obtained. Then rectified linear unit (ReLU) is used as the activation function to train the network. Moreover, in order to speed up the training of the network, batch normalization (BN) processing is added to the input of each convolutional layer.

B. LSTM layers

The traditional RNN model suffers from the problem of gradient exploding or vanishing when dealing with long-term dependency problems. Addressing this issue, many improved structures have been proposed, among which LSTM, as a classic improved model, performs well in many tasks. In this paper, LSTM neurons are used to extract features of the CSI time series.

The LSTM unit structure is presented in Fig.2. The activation value h_t in the LSTM unit at time t is linear interpolation of the activation value h_{t-1} at the previous time. After splicing vectors timing the weight matrix, the values of z^f, z^i and z^o are between 0 and 1 through the sigmoid function, as the gate states. Z converted the result into a value between -1 and 1 using the tanh activation function. σ is the sigmoid function. $\odot$ is the multiplication operation of the corresponding element number.

$$z = \tanh(W \cdot [x^t, h^{t-1}]), \tag{5}$$

$$z^i = \sigma(W^i \cdot [x^t, h^{t-1}]), \tag{6}$$

$$z^f = \sigma(W^f f \cdot [x^t, h^{t-1}]), \tag{7}$$

$$z^o = \sigma(W^o \cdot [x^t, h^{t-1}]), \tag{8}$$

$$c^t = z^f \odot c^{t-1} + z^i \odot z, \tag{9}$$

$$h^t = z^o \odot \tanh(c^t). \tag{10}$$

The output of the convolution layers is a tensor of 203*22*30, where 203 corresponds to the time dimension information, so the input of the first LSTM layer is a 660 dimensional vector. Finally, the result of the classification is output via the softmax regression classifier.

VI. EXPERIMENT AND EVALUATION

A. Experimental Environment and Settings

We choose the indoor environment for the experiment. The environment is shown in Fig. 3. In CSIID, WiFi devices are used to collect raw CSI data. The prototype consists of two devices: a ThinkPad x201 notebook with Intel 5300 802.11n WiFi NIC and a WiFi access point (AP) TP_LINK AC1750 router, which act as the receiver and the transmitter. The distance between the two devices is about 3 meters and the height from the two devices is 1 meter. We install Ubuntu 14.04 with custom Intel NIC driver [20] in the laptop to collect the raw CSI data. The number of transmitting and receiving antennas is 1 and 3 respectively and each antenna pair has 30 subcarriers, so the CSI stream consists of 30*1*3=90 subcarriers.

Fig. 3 **Experimental environment**

Our experiments use 5GHz frequency band, which can reduce the interference of 2.4GHz WiFi networks. CSI data is recorded with sampling rate of 1KHz.There are 6 participants that include 3 males and 3 females between 17 and 27 years old in our experiment. At the stage of data collection, each person walks along a predetermined path and walks back and forth through the LOS path. Each walking time is 5 seconds, and walking cycles 100 times. According to experimental requirements, we use participant's name as the label to input into our model, and then train our model via supervised learning.

We evaluate the performance of our deep learning model using TensorFlow in Python. The Adam optimizer is used for loss minimization and learning rate is set to 0.0001. We use the 70% of data set for training, and then use the other 30% data set as a test set to evaluate the performance of the model.

B. Experiment Evaluation

1) ***Comparing among different group sizes:*** When human identification group size from 2 to 6, the accuracy of CSIID in the indoor environment reaches 97.4% and 94.8%, respectively. Fig. 4 shows results of our model.

In Fig. 5, we show confusion matrix of the identification experiments when the number of participants is 6. Each row represents the real identity given in the label, and each column represents the identity tested by the model. A, B, C, D, E and F in the confusion matrix shows details of identification of different people. Obviously, different people have different recognition accuracy.

Fig. 4 Experimental results

The experimental result shows that the accuracy of identification will gradually decline as more people are involved in the environment. This can be explained by the similarities of body shape and behavior pattern between different individuals. The increasing number of people will minimize the specific difference of gait features, thus leading to the increasing rate of incorrect classification.

		Predicate Label					
		A	B	C	D	E	F
True Lable	A	230	0	4	3	1	2
	B	0	226	0	6	5	3
	C	4	2	224	6	4	0
	D	0	3	0	230	2	5
	E	4	0	3	1	232	0
	F	6	4	5	0	2	223

Fig.5: Confusion matrix of six people identification

2) ***Comparing among different models:*** In order to evaluate the performance of our model better, we conduct comparison experiments among CNN, LSTM and CSIID in the indoor environment of six participants. Experimental results are depicted in Table I. It can be observed that the accuracy of CSIID is highest.

Table I: Comparison Between Different Models

Models	Accuracy
CNN	90.6%
LSTM	91.4%
CSIID	94.8%

3) ***Changing wearing and accessories:*** We collect additional data from the participant to verify whether the changing of wearing will influence accuracy of CSIID. The result is still 94.8% even if the participant changes the clothes. And the accuracies are 93.9% when the participant carrying a bag during the experiment. In short, our model still maintains a stable identification accuracy rate in the case of human body decoration changes.

4) ***Deviating from the predetermined path:*** We ask the participant to walk straight 30 times in the path which is from the predetermined training path 1 m. The result of the test is 92.7% which indicates that features extracted by our deep learning method are robust to the deviation from path that predetermined during training.

VII. CONCLUSION

The traditional methods of human identification, facing a great deal of data preprocessing work, can easily lead to low accuracy. In this paper, the CSIID, an effective deep learning method, is proposed to human identification by using gait information detected via COTS WiFi devices. Combining convolution layers with LSTM layers, CSIID not only eliminates the need for the manual feature extraction, but also achieves high accuracy. Our model is ranging from 97.4% to 94.8% when the number of persons is from 2 to 6. In the future, we will collect more data to improve the accuracy of our algorithms. In addition, we will further study the situation of more people in the monitoring area, which is much more challenging for performing identification.

REFERENCES

[1] Connor, P. , & Ross, A. . (2018). Biometric recognition by gait: a survey of modalities and features. *Computer Vision and Image Understanding*, S1077314218300079.

[2] Castro, F. M. , Manuel J. Marín-Jiménez, Nicolás Guil, & Nicolás Pérez de la Blanca. (2017). Automatic Learning of Gait Signatures for People Identification. International Work-conference on Artificial Neural Networks. Springer, Cham.

[3] R. J. Orr and G. D. Abowd, "The smart floor: a mechanism for natural user identification and tracking," in CHI '00 Extended Abstracts on Human Factors in Computing Systems, 2000, pp. 275–276.

[4] Chang, J. Y. , Lee, K. Y. , Wei, Y. L. , Lin, C. J. , & Hsu, W. . (2016). We can "see" you via wi-fi - wifi action recognition via vision-based methods.

[5] Zeng, Y. , Pathak, P. H. , & Mohapatra, P. . (2016). [IEEE 2016 15th ACM/IEEE International Conference on Information Processing in Sensor Networks (IPSN) - Vienna, Austria (2016.4.11-2016.4.14)] 2016 15th ACM/IEEE International Conference on Information Processing in Sensor Networks (IPSN) - WiWho: WiFi-Based Person Identification in Smart Spaces. *2016 15th ACM/IEEE International Conference on Information Processing in Sensor Networks (IPSN)* (pp.1-12). IEEE.

[6] Wang, Y. , Liu, J. , Chen, Y. , Gruteser, M. , & Liu, H. . (2014). *E-eyes: Device-free location-oriented activity identification using fine-grained WiFi signatures*.

[7] Wu, C. , Member, S. , IEEE, Yang, Z. , & Member. (2015). Non-invasive detection of moving and stationary human with wifi. *IEEE Journal on Selected Areas in Communications, 33*(11), 1-1.

[8] Wang, X. , Wang, X. , & Mao, S. . (2017). [IEEE ICC 2017 - 2017 IEEE International Conference on Communications - Paris, France (2017.5.21-2017.5.25)] 2017 IEEE International Conference on Communications (ICC) - CiFi: Deep convolutional neural networks for indoor localization with 5 GHz Wi-Fi. *IEEE International Conference on Communications*(pp.1-6). IEEE..

[9] Ali, K. , Liu, A. X. , Wang, W. , & Shahzad, M. . (2015). Keystroke recognition using wifi signals.

[10] Abdelnasser, H. , Youssef, M. , & Harras, K. A. . (2015). [ieee ieee infocom 2015 - ieee conference on computer communications - kowloon, hong kong (2015.4.26-2015.5.1)] 2015 ieee conference on computer communications (infocom) - wigest: a ubiquitous wifi-based gesture recognition system. 1472-1480.

[11] Y. Wang, K. Wu, and L. M. Ni, "Wifall: Device-free fall detection by wireless networks," IEEE Transactions on Mobile Computing, vol. PP, no. 99, pp. 1–1, 2017

[12]X. Liu, J. Cao, S. Tang, and J. Wen, "Wi-sleep: Contactless sleep monitoring via wifi signals," in Real-Time Systems Symposium, 2015, pp. 346–355

[13] Nixon, M. S., & Carter, J. N. (2007). Automatic recognition by gait. *Proceedings of the IEEE, 94*(11), 2013-2024.

[14] Wu, Z. , Huang, Y. , Wang, L. , Wang, X. , & Tan, T. . (2016). A comprehensive study on cross-view gait based human identification with deep cnns. *IEEE Transactions on Pattern Analysis & Machine Intelligence, 39*(2), 209-226.

[15] L. Wu, C. Shen, and A. V. D. Hengel, "Convolutional lstm networks for video-based person re-identification," 2016.

[16] Yao, S. , Hu, S. , Zhao, Y. , Zhang, A. , & Abdelzaher, T. . (2016). Deepsense: a unified deep learning framework for time-series mobile sensing data processing.

[17]T. Xin, B. Guo, Z. Wang, M. Li, Z. Yu, and X. Zhou, "Freesense: Indoor human identification with wi-fi signals," in Global Communications Conference, 2017, pp. 1–7.

[18] Zhang, J. , Wei, B. , Hu, W. , & Kanhere, S. S. . (2016). WiFi-ID: Human Identification Using WiFi Signal. *2016 International Conference on Distributed Computing in Sensor Systems (DCOSS)*. IEEE Computer Society.

[19] Zheng, Y., Zhou, Z., & Liu, Y. (2013). From rssi to csi:indoor localization via channel response. *Acm Computing Surveys, 46*(2), 1-32.

[20] Halperin, D. , Hu, W. , Sheth, A. , & Wetherall, D. . (2011). Tool release: gathering 802.11n traces with channel state information. *Acm Sigcomm Computer Communication Review,41*(1), 53-53.