Intelligent and Converged Networks
2023, 4(1): 15−27

ISSN 2708-6240
DOI: 10.23919/ICN.2023.0002

Research on recognition algorithm for gesture page turning based on wireless sensing

Lin Tang, Sumin Wang, Meng Zhou, Yinfan Ding, Chao Wang*, Shengbo Wang, Zhen Sun, and Jie Wu

Abstract: When a human body moves within the coverage range of Wi-Fi signals, the reflected Wi-Fi signals by the various parts of the human body change the propagation path, so analysis of the channel state data can achieve the perception of the human motion. By extracting the Channel State Information (CSI) related to human motion from the Wi-Fi signals and analyzing it with the introduced machine learning classification algorithm, the human motion in the spatial environment can be perceived. On the basis of this theory, this paper proposed an algorithm of human behavior recognition based on CSI wireless sensing to realize deviceless and over-the-air slide turning. This algorithm collects the environmental information containing upward or downward wave in a conference room scene, uses the local outlier factor detection algorithm to segment the actions, and then the time domain features are extracted to train Support Vector Machine (SVM) and eXtreme Gradient Boosting (XGBoost) classification modules. The experimental results show that the average accuracy of the XGBoost module sensing slide flipping can reach 94%, and the SVM module can reach 89%, so the module could be extended to the field of smart classroom and significantly improve speech efficiency.

Key words: Wi-Fi signal; Channel State Information (CSI); wireless sensing; human behavior recognition

• Lin Tang, Sumin Wang, Meng Zhou, Yinfan Ding, and Chao Wang are with the Key Laboratory of Specialty Fiber Optics and Optical Access Networks, Joint International Research Laboratory of Specialty Fiber Optics and Advanced Communication, Shanghai Institute for Advanced Communication and Data Science, School of Communication and Information Engineering, Shanghai University, Shanghai 200444, China. E-mail: bollytom@163.com; sumer_06@163.com; zmyshu@163.com; dingyf98@163.com; wangchao@shu.edu.cn.
• Sumin Wang is also with the School of Electronic Information Engineering, Gannan University of Science and Technology, Ganzhou 341000, China.
• Shengbo Wang is with the Affiliated High School of Peking University, Beijing 100086, China. E-mail: wangsbengbo2004@outlook.com.
• Zhen Sun is with the Peking University & National Engineering Laboratory for Big Data Analysis and Applications, Beijing 100871, China. E-mail: zhen.sun@pku.edu.cn.
• Jie Wu is with the Department of Computer and Information Sciences, Temple University, Philadelphia, PA 19122-6096, USA. E-mail: jiewu@temple.edu.
* To whom correspondence should be addressed.
Manuscript received: 2022-12-08; accepted: 2023-01-17

1 Introduction

In schools, educational institutions, lectures, and product launches, video-based presentation devices are generally used to display knowledge content or product information in the form of PowerPoint (PPT). In actual demonstrations, to synchronize the content displayed by PPT with the speech, speakers generally use the keyboard, mouse, laser pointer, remote page-turner, or other devices to control the PPT, so that the pages of PPT are turned according to their own speech speed. However, under many circumstances, there is no remote page-turning pointer or wireless mouse in hand, so the speakers are often forced to stay in front of the computer and are unable to interact effectively with the audience. Or, when these devices are missing or malfunctioning, it will also cause much inconvenience to the user and waste a lot of time. Since the human body can reflect wireless signals, and human gestures can change the transmission mode of wireless signals,

these human movements between transmitters and receivers can be identified by extracting and analyzing the wireless signal information transmitted between a pair of wireless devices (e.g., smartphones, laptops, and wireless access points) [1].

With the rapid development of wireless technology, wireless signal sensing systems have received extensive attention from research teams globally. From different perspectives, the research teams have established a variety of proof-of-concept models for human behavior perception systems with wireless signal, among which a representative is Dina Katabi's team from the Massachusetts Institute of Technology (MIT). In 2020, Yue et al. [2] proposed a sleep posture monitoring system with wireless signal called BodyCompass, which was the first system based on Radio Frequency (RF) technology, which could perform accurate sleep posture monitoring at night at the user's home, achieving a high accuracy without sacrificing privacy and sleep comfort. In the same year, Fan et al.[3] proposed a new type of person re-identification system ReID, which employed radio signals to extract long-term human identification features, such as body size and shape. It could work under occlusion and insufficient lighting, which facilitated privacy protection and could be expanded to fields related to privacy, e.g., medical care. In 2021, Katabi[4] proposed another sensing technology at an Institute of Electrical and Electronic Engineers (IEEE) conference, which tracked people's gait and movement based on the radio signals reflected by the body. By transmitting low-power wireless signals and using machine learning models to analyze their reflections, this technology could remotely monitor a person's breathing, heartbeat, and sleep quality without any physical contact with the body. By understanding people's habits and passively monitoring their vital signs, this technology could detect health problems early, improve the overall health, and realize a real smart home. Many teams in China have also proposed systems based on wireless perception. In 2018, Qian et al.[5], from Tsinghua University built a non-contact system PADS that uses the Channel State Information (CSI) to achieve mobile human body detection at dynamic speeds. In 2019, Hang et al.[6] built the first lightweight, all-day, real-time, non-contact human detection system, WiSH. Traditional systems[7–9] that use the CSI of the physical layer collected from commercial Wi-Fi devices to achieve device-less gait recognition strongly depend on the direction of walking. To solve this problem, in 2019, Zhang Daqing's team designed a Wi-Fi based gait recognition system called WiDIGR[10]. In 2021, Zhang Daqing's team also designed a low-cost, non-invasive sleep monitoring system based on Wi-Fi devices, WiFi Sleep[11], which uses commercial Wi-Fi devices to monitor and classify the four sleep stages (wake, rapid eye movement, light sleep, and deep sleep). The experimental results showed that the accuracy of WiFi-Sleep is 81.8%, which is equivalent to that of the most advanced sleep stage monitoring technology using expensive radar equipment.

The key principle of these systems is based on the fundamental theory that the existence and movement of people affect the propagation path of wireless signals. By extracting the action information from the CSI data[12], the system can measure and perceive small signal changes caused by human motion, which can then be used to identify the behavior of human beings nearby. The essence of CSI is the sampling value of the channel frequency response at each subcarrier, which can effectively separate the wireless signals from different transmission paths. Therefore, CSI has the ability to distinguish the characteristics of multi-path signal transmission, evaluate the channel performance of the communication link, and describe how the signal propagates on a channel with time delay, amplitude attenuation, and phase shift. For the IEEE 802.11n standard[13] supported by common commercial Wi-Fi devices, CSI values are measured on 30 selected Orthogonal Frequency Division Multiplexing (OFDM) subcarriers. Each CSI measurement value contains 30 matrices with a dimension of $N_{Tx} \times N_{Rx}$, where N_{Tx} is is the number of transmitting antennas and N_{Rx} is the number of receiving antennas. Each element in the matrix is the channel frequency response value of a pair of antennas for any OFDM subcarrier frequency at a certain moment[14]. The time sequence of channel

frequency response values of the specified antenna pair and OFDM subcarrier is called the CSI data stream, so there are $30 \times N_{Tx} \times N_{Rx}$ CSI data streams in the time sequence of CSI values[15].

Traditional human behavior recognition systems, such as cameras, sensors, radars, and ultrasound equipment[16], have many disadvantages compared with wireless signal based behavior perception systems. For example, the human behavior recognition system based on computer vision uses high-resolution cameras to record human behavior and employs classification algorithms for image processing and effective human behavior recognition. It exhibits strong applicability, but it is also accompanied by privacy violations. Moreover, the cameras may fail to work normally due to problems, such as poor lighting and monitoring occlusion. Human behavior recognition systems based on wearable sensors[17, 18] use the acceleration or speed in the accelerator sensing space for recognition, but it is inconvenient to wear the sensors all the time. Although the non-line-of-sight and accurate radar system is very suitable for behavior recognition, it is too expensive for commercial use in ordinary civilian fields.

This paper designs a human behavior recognition system based on Wi-Fi signals, which overcomes the above shortcomings. The system is a proof-of-concept model for human behavior recognition using Wi-Fi signals, which combines digital signal processing with machine learning to analyze and process the CSI provided by the physical layer of Wi-Fi signals. Then, the human behavior can be sensed within the signal coverage, realizing the device-less and over-the-air PPT page turning. When the target user waves upward or downward in the free space covered by Wi-Fi signals, a specific data fragment is generated. When this data segment is detected and identified, PPT will automatically display the previous or next page. It can be widely used in schools, companies, and other places where PPT is used frequently. Laser pointers or mouses are no longer needed at all time, which is convenient for demonstrations and improves speech efficiency.

2 System design and implementation

2.1 System framework

The designed human behavior recognition system can be divided into four sub-modules, as displayed in Fig. 1.

(1) Data collection: Use a laptop with Intel5300

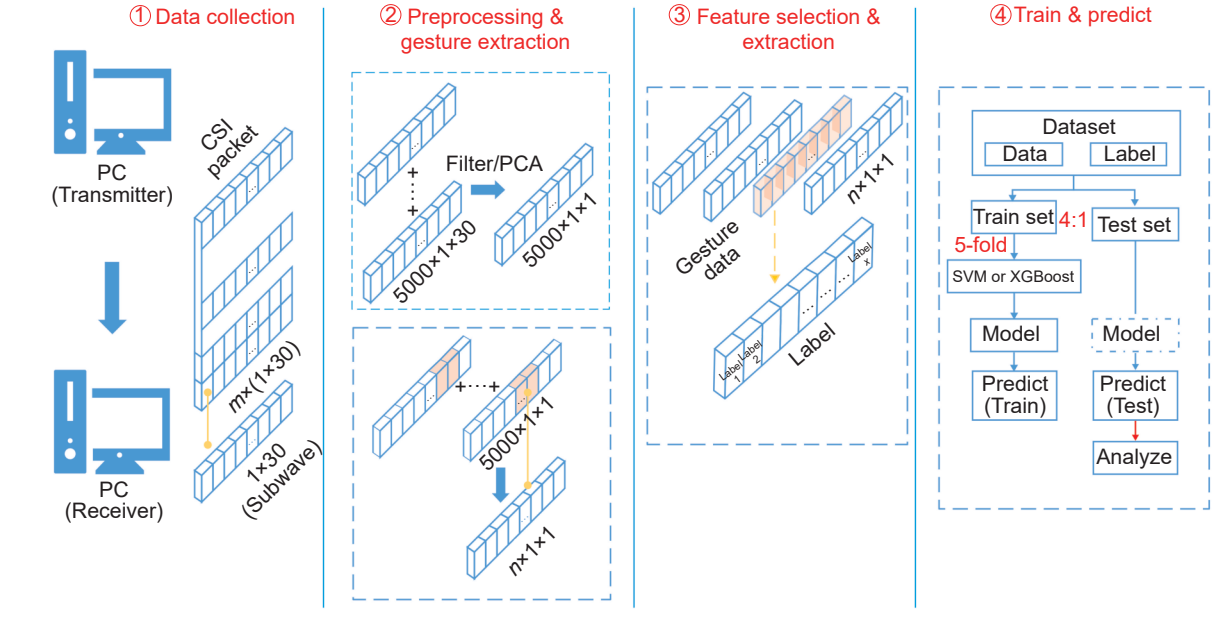

Fig. 1 Schematic diagram of overall system architecture.

wireless network card as the transmitter to send wireless signals, and use another laptop with the same configuration as the receiver to collect channel status information in the same area as the system input source.

(2) Data preprocessing and action segmentation: The collected original data are denoised by low-pass filtering, and the dimension is reduced by PCA, so as to ensure that the channel state information data related to human behavior in the measure environment can be obtained. Compared with the static state, the target user's "upward" or "downward" waving action can be considered as an abnormal segment, which can be intercepted by the exception detection algorithm.

(3) Feature extraction and selection: Because the channel state information data corresponding to different environmental states are different, eight types of time-domain features that can identify each action, namely, median absolute deviation, quartile difference, maximum, minimum, difference, average, variance, and normalized standard deviation, are extracted, so that the system can classify and identify them.

(4) Model training and testing: According to the acquired feature data of known actions, This paper uses Support Vector Machine (SVM) and eXtreme Gradient Boosting (XGBoost)[19] to conduct pattern recognition on the features obtained from the previous module, generate classification models, collect data under the same environment, and identify the unknown data, so as to know the current environmental status.

2.2 Data collection

In the wireless signal sensing system designed in this paper, the channel bandwidth of Wi-Fi signal is 20 MHz, and the data acquisition scheme is to select two subcarriers as basic grouping units, and obtain channel status information data of 30 subcarrier channels from each received packet.

In the wireless communication link between the transmitting antenna and the receiving antenna, the CSI obtained from the transmitted data packet is an $N_T \times N_R \times N_C$ complex matrix, where N_T is the number of transmitting antennas, N_R is the number of receiving antennas, and N_C is the number of OFDM subcarriers. Since the CSI exists in a complex form, the amplitude value and phase value of the channel state information can be obtained. The system only uses the amplitude value as the input of the system.

2.3 Pre-processing of original CSI data

The wireless signal is affected by environmental changes, such as temperature and pressure variations. The Butterworth low-pass filtering technology and PCA algorithm are used to pre-process the data to reduce noise. Figures 2 and 3 display the amplitude curve of the original CSI data in the case of a sudden upward wave from the static state and the CSI amplitude curves of the 6 subcarriers extracted after low-pass filtering, respectively. The correlation between the subcarriers can be observed, therefore, the approach of using the PCA algorithm to perform secondary noise removal and dimension reduction is still valid. The PCA2 curve is shown in Fig. 4.

According to Fig. 4 when the starting environment is in the steady state and the target user suddenly waves up, the PCA value changes abruptly within about 1 s of the motion. When the action is completed, the PCA returns from the dynamic state to the steady state.

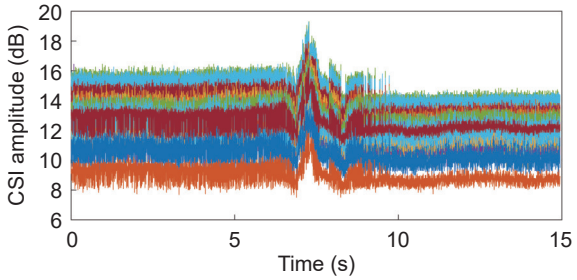

Fig. 2 Original CSI data amplitude curves of the upward wave. There are 30 subcarrier signals in total, and each color represents the signal of one subcarrier.

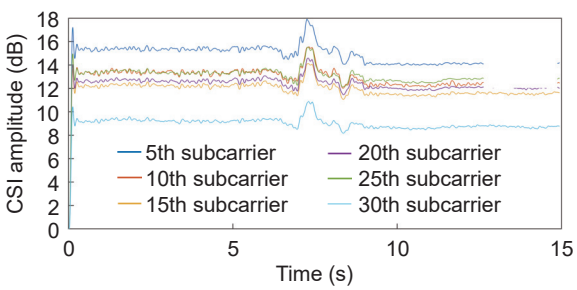

Fig. 3 CSI amplitude curves of the 6 subcarriers of the upward wave after low-pass filtering.

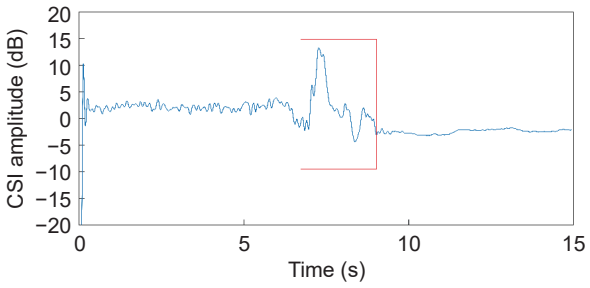

Fig. 4 PCA2 curve of the upward wave.

Therefore, this paper extracts the motion segments of upward and downward waves and compares them with the environment state when no action occurs to distinguish human behavior within the Wi-Fi signal coverage.

2.4 Local outlier factor based motion segmentation algorithm

As shown in Fig. 5, the overall spacing, density, and dispersion between points in set C_1 or set C_2 are uniform, so the points can be considered as in the same cluster. Points O_1 and O_2 are relatively isolated, so they can be considered outliers. To detect abnormal points from sets similar to C_1 or C_2 in terms of density and to achieve the universality of the algorithm, the Local Outlier Factor (LOF) detection algorithm is introduced[20]. Figure 4 shows that the CSI amplitude is stable when there is no human movement, but fluctuates significantly in case of human movement. The data from the beginning of fluctuation to the stable moment can be regarded as an outlier sequence that corresponds to the human behavior. Therefore, the corresponding human movement can be detected and extracted by the LOF detection algorithm.

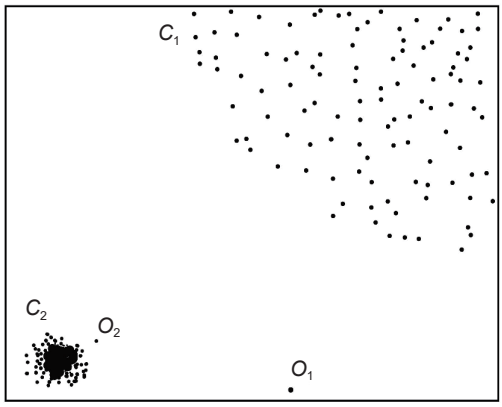

Fig. 5 Visual view of the outlier distribution.

The LOF detection algorithm judges data anomalies based on density. It compares the densities of data points and their neighbors, and then judges their isolation degrees. The LOF represents the local anomaly coefficient of different points; the larger its value, the more likely the point is an outlier.

It is supposed that k is a positive integer, S is the set of all data points, and p is a point in set S. $d_k(p)$ denotes the k-th distance of point p, and $d(p,x)$ represents the distance from point p to point x. The following conditions are satisfied:

(1) There are at least k points (excluding point p) that satisfy $d(p,x) \leqslant d_k(p)$;

(2) There are at most $k-1$ points (excluding point p) that satisfy $d(p,x) < d_k(p)$.

Define that the k-th distance neighborhood of point p is $N_k(p)$, which represents the set of points for which the distance from p does not exceed $d_k(p)$. So the number of k-th neighbor points of p is

$$|N_k(p)| \geqslant k \tag{1}$$

In set S, the k-th reachable distance from point p to point o is denoted by $Reach-Distance_k(p,o)$, then

$$Reach-Distance_k(p,o) = \max\{d_k(p), d(p,o)\} \tag{2}$$

where $d(p,o)$ represents the true distance from point p to point o. Then the local reachable distance of point p is

$$Lrd_k(p) = 1 \Bigg/ \left(\frac{\sum_{x_{N_k(p)}} Reach-Distance_k(p,x)}{|N_k(p)|} \right) \tag{3}$$

where $Lrd_k(p)$ represents the distribution of points near p. The larger $Lrd_k(p)$, the greater the density of the nearby points, and the shorter the distance from point p, that is, the smaller probability that point p is an outlier. For the local outlier factor $LOF_k(p)$ of point p, the probability that it can reflect the local anomaly of the sample points is

$$LOF_k(p) = \frac{\sum_{o \in N_k(p)} \frac{Lrd_k(o)}{Lrd_k(p)}}{|N_k(p)|} = \frac{\sum_{o \in N_k(p)} Lrd_k(o)}{|N_k(p)|} \Bigg/ Lrd_k(p) \tag{4}$$

According to Eqs. (3) and (4), the closer the $LOF_k(p)$ value is to 1, the more similar the densities of p and its

neighbor points, so p may belong to the same cluster as its neighborhood. If the value of $LOF_k(p)$ is less than 1, the density of p is higher than that of its neighbor points, so p is a dense point. If the value of $LOF_k(p)$ is greater than 1, the density of p is less than that of its neighbor points; the higher the value, the more likely that p is an outlier.

According to the waveform in Fig. 4, because of the device controlling by the experimenter, some human-related actions may be collected shortly after data acquisition. Therefore, the system intercepts one segment of the stable CSI data for anomaly detection by the LOF algorithm, as shown in Fig. 6. The analysis of the waveform inside the red box reveals that the duration of the extracted upward and downward waves is about 1 s, so the CSI data within this period can be interpreted as abnormal segments. When the system obtains the stable data, the LOF algorithm is used for detection. After experimentation, the system manually sets the LOF threshold to 1.5 and the time window to 3 s for segmenting the upward wave. A piece of independent data about human movement segmented from the time sequence of the original CSI data is shown in Fig. 7.

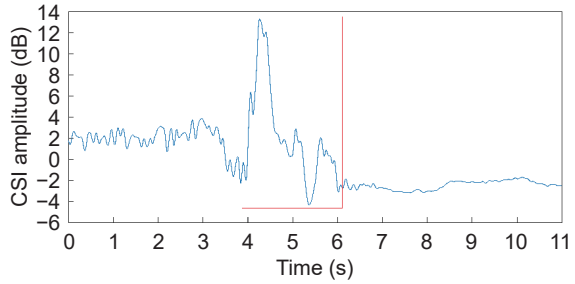

Fig. 6 Amplitude curve extracted from a steady data segment.

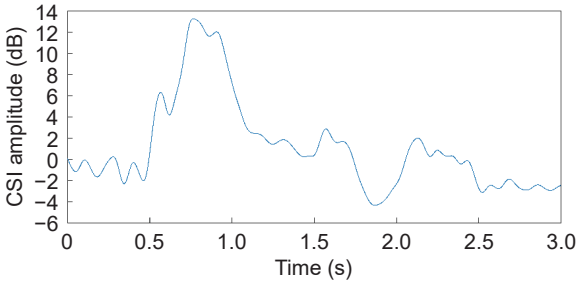

Fig. 7 Upward wave data segmented from the original CSI time sequence.

2.5 Time-domain feature extraction and selection

To distinguish the upward wave from the downward wave in a static environment for PPT page turning detection, it is still necessary to extract unique features that can identify the two states. The following 8 features are selected to represent actions: absolute median deviation, quartile deviation, maximum, minimum, calculation range, mean, variance, and normalized standard deviation. The 8 eigenvalues from three environment states form the input of the classification algorithm. The 8 time-domain characteristics of a sudden wave in a static environment are shown in Figs. 8−15. The eigenvectors numbered 1–30 belong to the upward wave, the eigenvectors numbered 31–60 belong to the downward

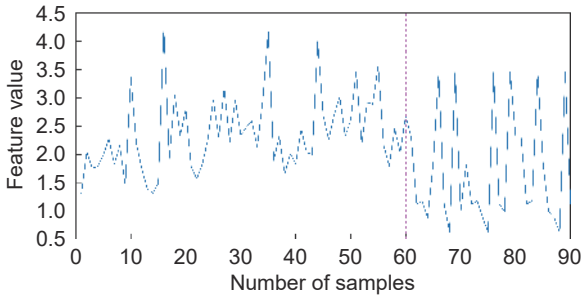

Fig. 8 Median absolute deviation of the three environment states.

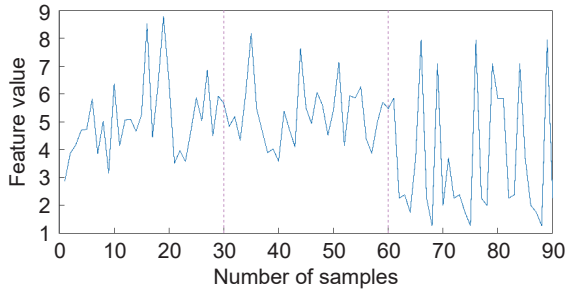

Fig. 9 Quartile deviation of the three environment states.

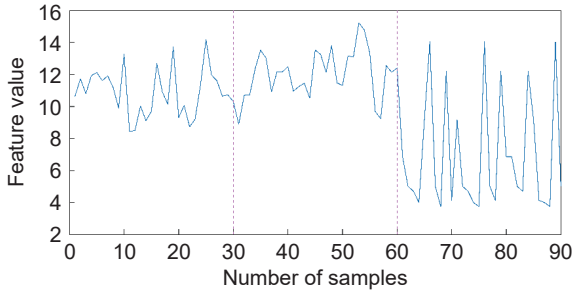

Fig. 10 Maximum value of the three environment states.

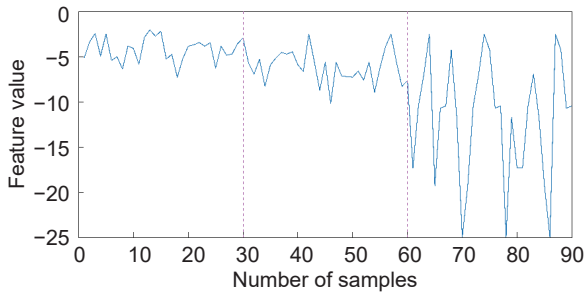

Fig. 11 Minimum value of the three environment states.

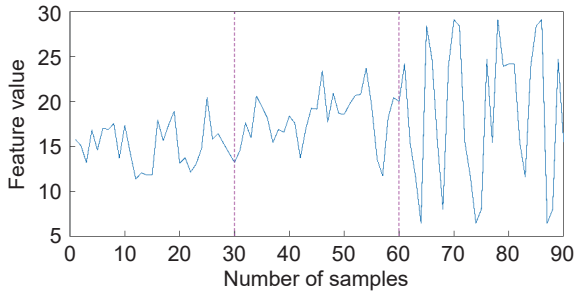

Fig. 12 Calculation range of the three environment states.

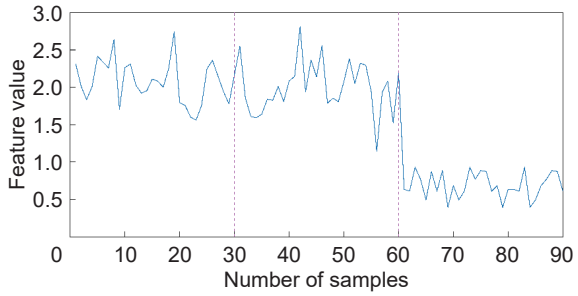

Fig. 13 Mean value of the three environment states.

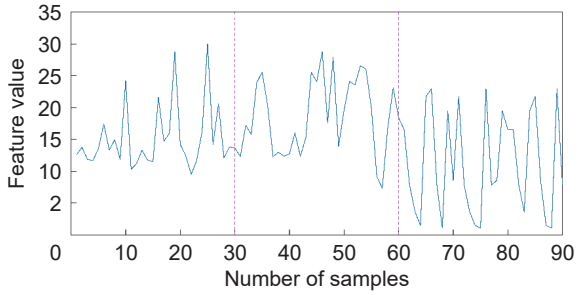

Fig. 14 Variance figure of the three environment states.

wave, and the feature sequence numbered 61–90 is for the static environment.

According to Figs. 8−15, the eigenvectors for the upward or downward wave and the static environment differ significantly, as divided by the red dashed lines, while the difference between the upward wave and the downward wave is not obvious, because the limb

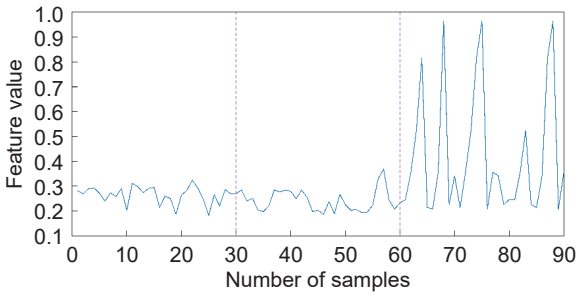

Fig. 15 Normalized standard deviation of the three environment states.

changes of the two states are similar. Therefore, the system has a higher accuracy when distinguishing a human movement from the static environment than when distinguishing the two movements.

2.6 Action recognition

After feature extraction, a 90 × 8 feature matrix F is generated. Meanwhile, a label matrix *Label* corresponding to the feature matrix is generated. Specifically, the sequence elements corresponding to the upward wave are set to 0, the sequence elements corresponding to the downward wave are set to 1, and the sequence elements corresponding to the static environment are set to 2. The feature matrix and the label matrix constitute the dataset of the classification model, which is divided into the training set to train the classification and recognition model and the test set to assess model performance.

The key to realize the function of PPT page turning detection is to accurately determine whether the feature data are for an upward or downward wave or a static environment, so as to know the current status. The machine learning classification algorithm can be used to train the data feature classifier. Based on the above classification ideas, this paper adopts SVM classification algorithm and XGBoost algorithm. At the same time, the characteristics of channel state information are trained differently, and finally two groups of trained classification models are obtained.

2.6.1 SVM algorithm classification and recognition

The system uses LIBSVM toolbox[21] to complete the work of data classification and recognition based on SVM classification algorithm, and the software package also provides Cross Validation (CV) function.

The practical application of SVM classification algorithm in wireless signal sensing system can be divided into training stage and testing stage.

LIBSVM selects the Radial Basis Function (RBF) kernel function, and its decision function is

$$predict_{label} = \mathrm{sgn}\left(\sum_{i=1}^{n} w_i \mathrm{e}^{-gamma\|x_2-x\|^2} + b\right).$$

where $\mathrm{e}^{-gamma\|x_2-x\|^2}$ represents the kernel function used by the SVM algorithm, and *gamma* is the pamameter of the radial basis kernel function; x_2 is the support vector in the training dataset; x is the support vector sample of the tag to be predicted; w_i is the coefficient of the support vector in the decision function, and b is the opposite number of the constant terms in the decision function. $predict_{label}$ is the decision result. Therefore, the system can obtain the labels of the sample data to be tested as long as the parameters of the decision function are determined, so as to identify the current state of the environment.

The training calculation complexity of SVM is between $O(N_{sv}^3 + L\times N_{sv}^2 + d\times L\times N_{sv})$ and $O(d\times L^2)$, where N_{sv} is the number of support vectors, L is the number of training set samples, and d is the dimension of each sample, that is, the original dimension, and the dimension before mapping to the high-dimensional space.

There are two parameters in RBF kernel function: *Cost* (loss function parameter, hereinafter referred to as abbreviation *C*) and *gamma* (hereinafter referred to as abbreviation *g*). Parameter *C* is a tradeoff between misclassified samples and the simplicity of the interface. A high *C* value selects more support vectors by increasing the freedom of the model to ensure that all samples can be correctly classified. The lower the *C* value, the smoother the interface. The parameter *g* is the reciprocal of the influence radius of the support vector sample, which represents the influence size of a single training sample. The smaller the *g* value is, the greater the influence is. The system uses the best parameters *C* and *g* as the parameters of the SVM algorithm to generate a training model, which is suitable for the current test environment. In the testing stage of the system, aiming at a group of unknown characteristic matrix data from the same room, the system can identify whether this group of data represents gesture behavior based on this training model, thus realizing the gesture recognition and detection function.

2.6.2 XGBoost classification algorithm recognition

Compared with previous learning algorithms, XGBoost's loss function introduces regularization terms to control the complexity of the model and prevent the model from overfitting. At the same time, the loss function is expanded by second-order Taylor formula to make the results more accurate. At the same time, it supports parallel reduction of model calculation and saves model training time. Therefore, this paper synchronously uses XGBoost to train the characteristics of the channel state information, and finally gets the trained classification model, and compares the test results with the results obtained from the previous SVM algorithm model.

XGBoost uses an integration strategy on the basis of decision tree and adopts the gradient boosting algorithm to reduce the loss of previously generated decision trees. New trees are also generated in the model to ensure the reliability of the final decision. The classification model of XGBoost is represented by

$$\widehat{y_i} = \sum_{k=1}^{K} f_k(x_i) \tag{5}$$

where K is the total number of trees, f_k ( ) denotes the *k*-th tree, x_i represents all eigenvectors, $\widehat{y_i}$ is the predicted result of sample x_i. The objective function established from each sub-tree is[22]

$$obj = \sum_{i=1}^{m} l(y_i, \widehat{y_i}) + \sum_{k=1}^{K} \Omega(f_k) \tag{6}$$

where y_i represents the true value, m is the number of samples, $L = \sum_{i=1}^{m} l(y_i, \widehat{y_i})$ is the loss function representing the model deviation, which is the regularization term used to prevent overfitting.

In the XGBoost algorithm, each prediction result is related to the last, as expressed in the following:

$$\widehat{y_i}^{(t)} = \sum_{k=1}^{t} f_k(x_i) = \widehat{y_i}^{(t-1)} + f_t(x_i) \tag{7}$$

where, $\widehat{y_i}^{(t)}$ represents the predicted result of sample *i*

after the t-th iteration, $\widehat{y}_i^{(t-1)}$ is the result of sample i after the first $t-1$ iterations, and $f_t(x_i)$ is the tree model after the t-th iteration. The structures of the first $t-1$ trees have been determined, which means the complexity is fixed, so the objective function is changed to

$$obj^{(t)} = \sum_{i=1}^{m} l(y_i, \widehat{y}_i^{(t-1)} + f_t(x_i)) + \Omega(f_t) + \text{constant} \quad (8)$$

where constant is a constant term, and the Taylor expansion of the objective function is

$$obj^{(t)} \cong \sum_{i=1}^{m} [l(y_i, \widehat{y}_i^{(t-1)}) + g_i f_t(x_i) + \frac{1}{2} h_i f_t^2(x_i)] + \Omega(f_t) + \text{constant} \quad (9)$$

where g_i and h_i are the first and second derivatives of the loss function with respect to $\widehat{y}_i^{(t-1)}$. For the t-th iteration, $\widehat{y}_i^{(t-1)}$ is known, so $l(y_i, \widehat{y}_i^{(t-1)})$ is a constant. After optimizing all the constant terms, the objective function becomes

$$obj^{(t)} \cong \sum_{j=1}^{T} [G_j \omega_j + \frac{1}{2}(H_j + \lambda)\omega_j^2] + \gamma T \quad (10)$$

where $G_j = \sum_{i \in I_j} g_i$, $H_j = \sum_{i \in I_j} h_i$, γ and λ are complexity parameters that can ensure the control of model complexity during training, T is the total number of leaf nodes, ω is the core of the leaf node, and $\frac{1}{2}\lambda \sum_{j=1}^{T} \omega_j^2$ is the L_2 modulus square of ω. Equation (10) shows that the target minors of the leaf nodes are independent, that is, when the minor of each leaf node reaches the extreme point, the whole objective function $obj^{(t)}$ reaches the extreme point as well. G_j and H_j are the results obtained from the first $t-1$ steps and can be considered as constants, so the derivative of ω_j is solved. By setting the derivative of ω_j to 0, the optimal solution of ω_j can be obtained, which can be substituted into the final objective function,

$$obj^{(t)} = -\frac{1}{2} \sum_{j=1}^{T} \frac{G_j^2}{H_j + \lambda} + \gamma T \quad (11)$$

For XGBoost algorithm, the number of data samples is N, the number of features is M, the number of trees is K, the tree depth is D, and the row sampling and column sampling are not considered. The time complexity analysis is as follows:

(1) Pre-sorting of global features: Because of global sorting, the global sorting information can be reused in the later node re-splitting without re-sorting, so the sorting complexity is $O(M \times N \times \log(N))$.

(2) Complexity of building a single tree: Because XGBoost implementation is based on level-wise, the time complexity of each layer is $O(M \times N)$, and the complexity of K trees is $O(K \times M \times N \times D)$.

(3) The final time complexity: $O(M \times N \times \log(N)) + O(K \times M \times N \times D)$.

In this paper, when the XGBoost learning algorithm is used to train the recognition model, a grid search method is used to optimize its parameters to get an optimal learning model, which can improve the recognition accuracy and achieve the monitoring effect. Grid search is a parameter tuning tool, and its principle is to search exhaustively, that is, to try every possibility through cyclical traversal in the selection of all the candidate parameters, so the best performing parameter is the final result. Eventually, the system selects the following key parameters to optimize:

(1) learning_rate: It denotes the step size of each iteration. Adjusting this parameter makes the model more robust.

(2) max_depth: It represents the maximum depth of a tree, which is used to control the overfitting. The larger the value of max_depth, the more specifically the model learns.

(3) min_child_weight: It defines the minimum sum of weights to be observed and can be used to control overfitting.

(4) colsample_bytree: It is used to control the proportion of columns in each randomly sampled tree (each column is a feature).

(5) subsample: It controls the proportion of random sampling for each tree. Reducing this value will make the algorithm more conservative and avoid overfitting. However, if this value is set excessively small, it may result in under-fitting.

(6) n_estimators: It is the maximum number of trees generated and also the maximum number of iterations.

After tuning the parameters using the grid search method, the following parameters are selected:

learning_rate = 0.6, max_depth = 6, min_child_weight = 8, colsample_bytree = 1, subsample = 1, and n_estimators = 200. Under these parameters, the training learning model is optimized, and the best recognition effect is achieved, thus achieving the sensing and detection function of PPT page turning.

3 Analysis of experimental results

3.1 Experimental environment

The experimental environment for Wi-Fi based PPT page turning detection was a common conference room. It was an ideal indoor free space where there were some simple tables and chairs, which could be evaluated as a scene for frequent PPT use, e.g., classrooms or corporate conference rooms. The experimental settings for the actual scene are displayed in Fig. 16. In addition to a user under test, there was another still person in the room to control the data acquisition device. The experiment aimed to detect whether the user's hand was waving upward or downward. There were three kinds of states for the user, the first was to remain still, the second was to wave upward s from the static state, and the third was to wave downward from the static state. The data were collected for 15 s at a time, and 30 groups of data were acquired for each state.

3.2 Experimental result

The PPT page turning detection function was validated according to the experimental requirements. Table 1 lists the specific values of system recognition accuracy for different motions, and Fig. 17 displays the corresponding histogram.

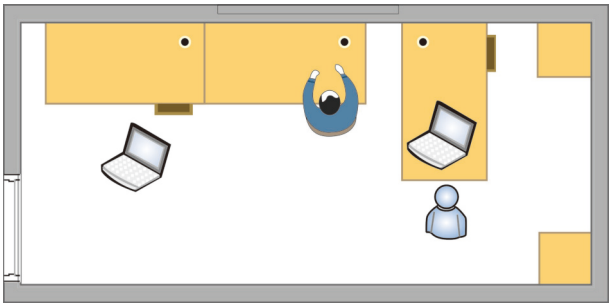

Fig. 16 Plan of experimental setup for PPT page turning detection.

Table 1 System recognition accuracies of different motions.

Classification of human movements	SVM classification and recognition accuracy (%)	XGBoost classification and recognition accuracy (%)
Upward versus static	100.00	100.00
Downward versus static	100.00	100.00
Upward versus downward	83.33	91.67
Upward, downward, and static	88.89	94.44

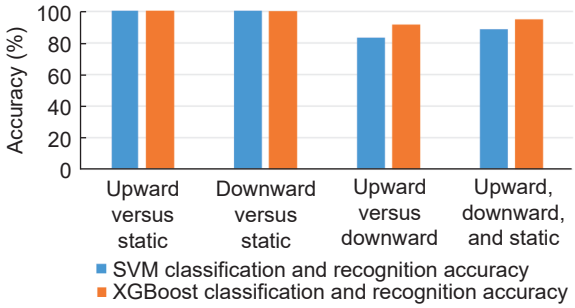

Fig. 17 Accuracy histogram of the system based on SVM and XGBoost in different actions.

According to the experimental result data in Table 1, no matter which algorithm is used, upward and static or downward and static can be accurately distinguished, and when distinguishing upward and downward waving actions, the recognition accuracy decreases significantly. This is because in the static state, sudden actions will lead to more obvious changes in the signals reflected by the human body, while upward and downward waves are both a dynamic environment, and the speed and mode of changes in the human body's limbs are not much different. The feature vectors are calculated from the channel state information carried by the received signals of human reflexes. The difference between these feature vectors is slightly small, which may affect the training of classification models and lead to a decrease in recognition accuracy. By analyzing the trend of the histogram, we can see that in the upward, downward, and upward actions, as well as downward and static waving actions environments, the system recognition accuracy using XGBoost algorithm is better than that of using the SVM algorithm. This is because the SVM

algorithm uses the kernel function to project the separation hyperplane into a curved surface, which can solve the nonlinear classification problem, and requires grid search during training. The trained model performance is better than that of XGBoost algorithm, but the training accuracy is lower than XGBoost algorithm. In the future use of actual scenarios, the system can flexibly adopt one of the classification algorithms according to specific needs, which will be more reliable and practical, and can further train the optimization model during non use without affecting the user experience.

4 Conclusion

The designed human behavior recognition system based on Wi-Fi signals extracts effective feature information related to human motion from wireless signals and uses the digital signal processing algorithm and XGBoost classification algorithm to realize the detection of PPT page turning. Compared with traditional human behavior recognition systems, the proposed system does not require worn hardware during the detection process, so the users will not feel the existence of the system and be affected by the detection. In addition, unlike cameras, this system is non-invasive and protects privacy. Moreover, because Wi-Fi devices are widely deployed, the system is universal, convenient, and inexpensive. As the research continues deepening, the following issues need to be explored before putting the system into practical applications.

(1) Due to the limited time and experiment site, the designed Wi-Fi signal based human behavior recognition system was merely tested in a limited and controlled space. To ensure the robustness and accuracy of the system, more types of environments need to be selected for validation.

(2) In this paper, only upward and downward waves in a static environment were sensed for PPT page turning. Further research should involve dynamic environments and more human motion types.

(3) The LOF detection algorithm was used for motion segmentation in this paper; some better motion segmentation algorithms can be explored to enhance the accuracy of behavior detection.

The designed human behavior recognition system based on wireless perception has preliminarily established its proof-of-concept model, which was experimentally tested in a small and controlled space in this paper. Considering the achievements in the field of wireless perception, the future application prospects of the proposed system are very broad. Wireless perception will adapt to the current background of artificial intelligence era, and wireless perception based human behavior recognition systems will have higher commercial values.

Acknowledgment

This work was supported by the Special Zone Project of National Defense Innovation.

References

[1] C. Wang, S. Chen, Y. Yang, F. Hu, F. Liu, and J. Wu, Literature review on wireless sensing-Wi-Fi signal-based recognition of human activities, *Tsinghua Science and Technology*, vol. 23, no. 2, pp. 203–222, 2018.

[2] S. Yue, Y. Yang, H. Wang, H. Rahul, and D. Katabi, BodyCompass: Monitoring sleep posture with wireless signals, *Proc. ACM Interact. Mob. Wearable Ubiquit. Technol.*, vol. 4, no. 2, pp. 1–25, 2020.

[3] L. Fan, T. Li, R. Fang, R. Hristov, Y. Yuan, and D. Katabi, Learning longterm representations for person re-identification using radio signals, in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, Virtual Event, 2020, pp. 10699−10709.

[4] D. Katabi, Keynote: Monitoring people and their vital signs using radio signals and machine learning, in *Proc. of 2021 IEEE International Conference on Pervasive Computing and Communications* (*PerCom*), Kassel, Germany, 2021, p. 1.

[5] K. Qian, C. Wu, Z. Yang, Y. Liu, F. He, and T. Xing, Enabling contactless detection of moving humans with dynamic speeds using CSI, *ACM Trans. Embedd. Comput. Syst.*, vol. 17, no. 2, pp. 1–18, 2018.

[6] T. Hang, Y. Zheng, K. Qian, C. Wu, Z. Yang, X. Zhou, Y. Liu, and G. Chen, WiSH: WiFi-based real-time human detection, *Tsinghua Science and Technology*, vol. 24, no. 5, pp. 615–629, 2019.

[7] W. Wang, A. X. Liu, and M. Shahzad, Gait recognition using wifi signals, in *Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing*, Heidelberg, Germany, 2016, pp. 363−373.

[8] J. Zhang, B. Wei, W. Hu, and S. S. Kanhere, Wifi-id: Human identification using wifi signal, in *Proc. of 2016 International Conference on Distributed Computing in Sensor Systems* (*DCOSS*), Washington, DC, USA, 2016, pp. 75−82.

[9] T. Xin, B. Guo, Z. Wang, M. Li, Z. Yu, and X. Zhou, Freesense: Indoor human identification with Wi-Fi signals, in *Proc. of 2016 IEEE Global Communications Conference* (*GLOBECOM*), Washington, DC, USA, 2016, pp. 1−7.

[10] L. Zhang, C. Wang, M. Ma, and D. Zhang, WiDIGR: Direction-independent gait recognition system using commercial Wi-Fi devices, *IEEE Intern. Things J.*, vol. 7, no. 2, pp. 1178–1191, 2019.

[11] B. Yu, Y. Wang, K. Niu, Y. Zeng, T. Gu, L. Wang, C. Guan, and D. Zhang, WiFi-Sleep: Sleep stage monitoring using commodity Wi-Fi devices, *IEEE Intern. Things J.*, vol. 8, no. 18, pp. 13900–13913, 2021.

[12] Y. Zeng, P. H. Pathak, and P. Mohapatra, WiWho: WiFi-based person identification in smart spaces, in *Proc. of 2016 15th ACM/IEEE International Conference on Information Processing in Sensor Networks* (*IPSN*), Vienna, Austria, 2016, pp. 1−12.

[13] IEEE, IEEE standard for information technology--local and metropolitan area networks--specific requirements--Part 11: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) specifications. Amendment 5: Enhancements for higher throughput, 802.11n-2009, https://standards.ieee.org/ieee/802.11n/3952/, 2009.

[14] H. Abdel-Nasser, R. Samir, I. Sabek, and M. Youssef, MonoPHY: Mono-stream-based device-free WLAN localization via physical layer information, in *Proc. of 2013 IEEE Wireless Communications and Networking Conference* (*WCNC*), Shanghai, China, 2013, pp. 4546−4551.

[15] H. Chen, Y. Zhang, W. Li, X. Tao, and P. Zhang, ConFi: Convolutional neural networks based indoor Wi-Fi localization using channel state information, *IEEE Access*, vol. 5, pp. 18066–18074, 2017.

[16] J. Gu, J. Wang, L. Zhang, Z. Yu, X. Xin, and Y. Liu, Spotlight: Hot target discovery and localization with crowdsourced photos, *Tsinghua Science and Technology*, vol. 25, no. 1, pp. 68–80, 2019.

[17] Z. Song, Z. Cao, Z. Li, J. Wang, and Y. Liu, Inertial motion tracking on mobile and wearable devices: Recent advancements and challenges, *Tsinghua Science and Technology*, vol. 26, no. 5, pp. 692–705, 2021.

[18] Z. Zhang, X. Cong, W. Feng, H. Zhang, G. Fu, and J. Chen, WAEAS: An optimization scheme of EAS scheduler for wearable applications, *Tsinghua Science and Technology*, vol. 26, no. 1, pp. 72–84, 2020.

[19] X. Zhang, C. Xiu, Y. Wang, and D. Yang, High-precision WiFi indoor localization algorithm based on CSI-XGBoost, *J. Beijing Univ. Aeronaut. Astronaut.*, vol. 44, no. 12, pp. 2536–2544, 2018.

[20] M. M. Breunig, H. -P. Kriegel, R. T. Ng, and J. Sander, LOF: Identifying density-based local outliers, in *Proceedings of the 2000 ACM SIGMOD International Conference on Management of Data*, Dallas, Texas, USA, 2000, pp. 93−104.

[21] C. C. Chang and C. J. Lin, LIBSVM: A library for support vector machines, *ACM Trans. Intell. Syst. Technol.*, vol. 2, no. 3, pp. 1–27, 2011.

[22] T. Chen and C. Guestrin, Xgboost: A scalable tree boosting system, in *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, San Francisco, CA, USA, 2016, pp. 785−794.

Lin Tang received the MEng degree from Shanghai University, Shanghai, China in 2011. Currently, he is a PhD candidate at the School of Communication and Information Engineering, Shanghai University. His main research interest is wireless sensing technology.

Meng Zhou received the BEng degree from Yangtze University, China in 2019. She is currently a master student at the School of Communication and Information Engineering, Shanghai University. Her main research interest is wireless sensing technology.

Yinfan Ding received the BEng degree from Guilin University of Electronic Technology, China in 2020. He is currently a master student at the School of Communication and Information Engineering, Shanghai University. His main research interest is wireless sensing technology.

Shengbo Wang is an undergraduate student at the Affiliated High School of Peking University and participates in the joint project of Peking University National Engineering Laboratory for Big Data Analysis and Application and IEEE Shanghai Computer Chapter. His main research interest is intelligent transportation.

Sumin Wang received the MEng degree from Jiangxi University of Science and Technology, China in 2010. Currently, she is a PhD candidate at Key Laboratory of Specialty Fiber Optics and Optical Access Networks, Joint International Research Laboratory of Specialty Fiber Optics and Advanced Communication, Shanghai Institute for Advanced Communication and Data Science, Shanghai University. She is also with the School of Electronic Information Engineering, Gannan University of Science and Technology. Her main research interests include intelligent transportation and quantum computing.

Chao Wang received the PhD degree from Tongji University, China in 1999. He is a professor at the Key Laboratory of Specialty Fiber Optics and Optical Access Networks, Joint International Research Laboratory of Specialty Fiber Optics and Advanced Communication, Shanghai Institute for Advanced Communication and Data Science, Shanghai University. He is also a senior member of CCF, an IEEE senior member, the vice chair of IEEE China Council, the council member of China Institute of Electronic, the council member of China Association of AI, the deputy director of Information Security Experts Committee (China Institute of Electronic), the vice chair of IEEE Shanghai Computer Chapter, and the committeeman of the Sixth Shanghai Expert Committee for Informatization. His research interests include AI, network information security, wireless sensing, and quantum computing cryptography.

Zhen Sun received the BEng degree from Shandong University of Technology, China in 2010. He is an engineer at Peking University & National Engineering Laboratory for Big Data Analysis and Applications. His main research interests include data transaction, industrial big data, industrial Internet, and intelligent manufacturing.

Jie Wu received the PhD degree from Florida Atlantic University, USA in 1989. He is currently the director of the Center for Networked Computing and a Laura H. Carnell professor at Temple University. He also serves as the director of International Affairs at the College of Science and Technology. His current research interests include mobile computing and wireless networks, routing protocols, network trust and security, distributed algorithms, applied machine learning, and cloud computing.

He served as the chair of the Department of Computer and Information Sciences from Summer 2009 to Summer 2016, and an associate vice provost for international affairs from fall 2015 to summer 2017. Prior to joining Temple University, he was the program director of the National Science Foundation of USA. He was elected a member of the European Academy of Sciences in 2022. He regularly publishes in scholarly journals, conference proceedings, and books. He serves on several editorial boards, including *IEEE Transations on Mobile Computing*, *IEEE Transations on Service Computing*, *Journal of Parallel and Distributed Computing*, and *Journal of Computer Science and Technology*. He was the general chair/co-chair for IEEE IPDPS' 08, IEEE DCOSS'09, IEEE ICDCS'13, ACM MobiHoc'14, ICPP'16, IEEE CNS'16, WiOpt'21, and ICDCN'22, and the program chair/co-chair for IEEE MASS'04, IEEE INFO-COM' 11, CCF CNCC'13, and ICCCN'20. He was an IEEE Computer Society Distinguished Visitor, an ACM Distinguished Speaker, and the chair of the IEEE Technical Committee on Distributed Processing (TCDP). He was a recipient of the 2011 China Computer Federation (CCF) Overseas Outstanding Achievement Award. He is a fellow of the AAAS. He is an IEEE fellow.