2022 IEEE the 14th International Conference on Wireless Communications and Signal Processing

# Position and Orientation Independent Wireless Gesture Recognition

Yihui Wang, Yafei Tian, and Rui Peng
School of Electronics and Information Engineering, Beihang University
37 Xueyuan Road, Haidian District, Beijing 100191, P. R. China
wyh98@buaa.edu.cn; ytian@buaa.edu.cn; pengrui@buaa.edu.cn

***Abstract*—Gesture recognition through wireless communication signal is a new approach for human-computer interaction, and is an important topic for integrated sensing and communication designs in next-generation mobile communication system. Wireless gesture recognition is mainly realized by analyzing the change of propagation channel, especially the phase change of reflection path. However, the law of phase change is closely related to the position and orientation of the user, as well as the positions of transmitter and receiver. The gesture recognition system trained in a given environment usually does not work well in a new environment. In this paper we propose a position and orientation independent gesture recognition method, where the phase change under any position configuration is transformed to the gesture trajectory in an invariant body coordinate system. The transformation matrix implicitly embodies the impact of position and orientation, and is estimated by a preamble gesture each time reaching a new environment. The method is verified in a prototype system, and the recognition accuracies under different position configurations are tested.**

***Index Terms*—Channel state information, coordinate transformation, gesture recognition, position-independent, wireless sensing***

## I. Introduction

Wireless signals are not only used for high-speed and reliable transmission of information, but also have sensing ability, so that wireless gesture recognition can be a good candidate for device-free human-computer interaction means [1], [2]. On the propagation path of wireless signals, the variations of reflection, scattering and diffraction caused by gestures will affect the channel state information (CSI) [3], [4]. By analyzing the specific change modes of CSI, different gestures can be recognized.

However, many existing wireless sensing systems have position-dependency problem, since the variations of CSI are not only determined by gestures, but also highly related to the position and orientation of users, as well as the positions of transmitter and receiver [5]. If we directly use position-specific features to realize gesture recognition with end-to-end deep learning, huge amounts of CSI data in different locations and orientations needs to be collected, and the system needs to be retrained when a new environment is encountered.

Position-independent recognition methods are thus highly pursued. In [5], a translation function is proposed that can generate virtual samples of a given gesture in a new environment using the existing data, thus avoid recollecting samples. In this translation strategy, only amplitude information of the reflection signal is used, and the variation of amplitude is significant only when the hand is very close to the receiver. In [6], a deep adversarial network is designed to extract environment/subject-independent features shared by the data collected on different subjects under different environments. In [7], the body-coordinate velocity profiles (BVP) of gestures are extracted from Doppler frequency shift as unique features for recognition, but the orientation and location information of the person is a priori knowledge for the compressed-sensing based BVP extraction approach. In [8], a gesture is characterized by how the hand moves relative to its previous position and a position-independent feature is captured to describe the pattern of moving direction changes. In [9], a CSI correlation feature enhancement method is proposed to enhance the activity-dependent information and eliminate the activity-unrelated information.

In this paper, we propose a method to realize gesture recognition when the position configuration and user orientation are unknown. We use the phase change of the reflection path, and build a relationship between a speed vector in the body coordinate system and the phase change rate at the receiver. With more than two receivers, we define a phase-trajectory transformation matrix, which can transform the phase changes directly into gesture trajectory. The transformation matrix embodies the impact of position and orientation implicitly, and is estimated by a preamble gesture each time reaching a new environment. We design a specialized preamble gesture to reduce the estimation error. The recovered gesture trajectory is unrelated to the position and orientation, and can be well recognized by a pretrained neural network. This method has clear physical explanation, and we believe the trajectory recovery idea will have wider applications more than just in gesture recognitions or activity classifications.

The main contributions of this paper are in the following three aspects:

1) we study a position-independent gesture recognition method, where the gesture trajectory is recovered from the phase change of dynamic reflection path without knowing user position and orientation;
2) we design a simple but effective preamble gesture involving speed vectors in various directions, to estimate the

This work was supported by the National Natural Science Foundation of China under Grant 61971023.

phase-trajectory transformation matrix and the implicit information of position and orientation;

3) we build a prototype system based on LTE signals to recognize gestures in real-time, and test the accuracies under different position configurations.

## II. Methodology

### A. Channel Model

Wireless signals propagate through multiple paths in the environment. During hand movement, the channel is composed of static paths and dynamic path, as shown in Fig. 1. The static paths include the line-of-sight (LoS) propagation from the transmitter to the receiver, and reflections or scatterings by other static objects. The dynamic path is caused by the reflection of a moving object, which is the hand we considered in this paper.

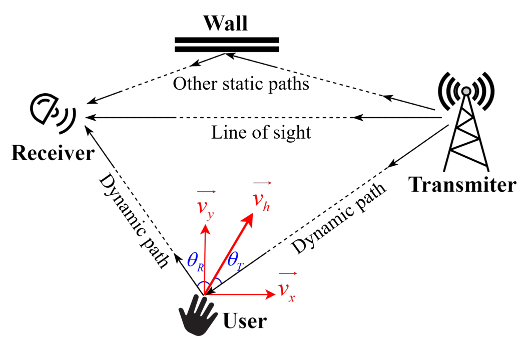


Fig. 1. Multipath propagation of wireless signal and hand moving speed decomposition in body coordinate system.

The channel response is time-varying, and can be expressed as

$$\begin{aligned} h(\tau,t) = &\underbrace{a_{\mathrm{LoS}} e^{-j\frac{2\pi d_{\mathrm{LoS}}}{\lambda}} \delta\left(\tau - \frac{d_{\mathrm{LoS}}}{c}\right) + r_{\text{static}}^{\text{others}}(\tau)}_{\text{static path}} \\ &\underbrace{+ a_r(t) e^{-j\left(\frac{2\pi d_r(t)}{\lambda}+\pi\right)} \delta\left(\tau - \frac{d_r(t)}{c}\right)}_{\text{dynamic path}}, \end{aligned} \tag{1}$$

where $a_{\mathrm{LoS}}$ and $d_{\mathrm{LoS}}$ are the attenuation and path length of the LoS component respectively, $\lambda$ is the wavelength of wireless signal, $c$ is the speed of light, $a_r(t)$ and $d_r(t)$ are the attenuation and path length of the dynamic component respectively. As the hand moves, the change of $d_r(t)$ leads to the change of the reflection path phase, which is

$$p_h = \frac{2\pi}{\lambda} d_r(t) = \frac{2\pi}{\lambda}\left(|\overrightarrow{UR}| + |\overrightarrow{UT}|\right), \tag{2}$$

where $\overrightarrow{UR}$ and $\overrightarrow{UT}$ are the vectors from user to receiver and transmitter, respectively. Since the distance variation in a gesture is in the scale of decimeters, the signal delay variation is in the scale of nanoseconds, which is indistinguishable for Wi-Fi or LTE signals. The only significant change we can observe is the phase of dynamic path, which exactly depends on the gesture trajectory and the position configurations.

In general, the distances among the transmitter, receiver and user are one or two orders of magnitude longer than the hand movement distance, so that the geometry relation among them can be considered as fixed during a gesture. Define the unit vectors of $\overrightarrow{UR}$ and $\overrightarrow{UT}$ as $\vec{R}$ and $\vec{T}$, respectively. At a certain moment, assuming that the speed vector of hand movement is $\vec{v_h}$, the instantaneous phase change rate is

$$\begin{aligned} \frac{dp_h}{dt} &= \frac{2\pi}{\lambda}\left(\frac{d|\overrightarrow{UR}|}{dt} + \frac{d|\overrightarrow{UT}|}{dt}\right) \\ &= \frac{2\pi}{\lambda}\left(|\vec{v_h}|\cos\theta_R + |\vec{v_h}|\cos\theta_T\right) \\ &= \frac{2\pi}{\lambda}\left(\vec{v_h}\cdot\vec{R} + \vec{v_h}\cdot\vec{T}\right). \end{aligned} \tag{3}$$

We can see that the phase change not only depends on the transmitter-user-receiver geometry relation, which is implied in $\vec{R}$ and $\vec{T}$, but also depends on the user orientation, which causes different $\vec{v_h}$ even for the same gesture. To recognize a given gesture with any possible user positions and orientations, we must find a decoupling relation to resolve the trajectory from the phase changes.

### B. Phase-Trajectory Transforming

For a gesture, the trajectory in the body coordinate system is invariant. For example, define the body center as the coordinate origin, the right direction as x-axis, and the forward direction as y-axis, as shown in Fig. 1. If we can find the transformation relation between the speed vector and the phase change rate in the body coordinate system, we can recover the gesture trajectory without knowing user position and orientation. Then the gesture can be easily recognized. Given a random user position and orientation, define the unit speed vector in x-axis and y-axis as $\vec{v_x}$ and $\vec{v_y}$, respectively. Then the corresponding phase change rates are

$$\begin{aligned} \frac{dp_x}{dt} &= \frac{2\pi}{\lambda}\left(\vec{v_x}\cdot\vec{R} + \vec{v_x}\cdot\vec{T}\right), \\ \frac{dp_y}{dt} &= \frac{2\pi}{\lambda}\left(\vec{v_y}\cdot\vec{R} + \vec{v_y}\cdot\vec{T}\right). \end{aligned} \tag{4}$$

Note that any speed vector $\vec{v_h}$ can be represented by $\vec{v_x}$ and $\vec{v_y}$ as

$$\vec{v_h} = k_x\,\vec{v_x} + k_y\,\vec{v_y}, \tag{5}$$

where $k_x$ and $k_y$ are weighting coefficients. The phase change rate can also be decomposed into two affecting factors, i.e.,

$$\begin{aligned} \frac{dp_h}{dt} &= \frac{2\pi}{\lambda}\left(\vec{v_h}\cdot\vec{R} + \vec{v_h}\cdot\vec{T}\right) \\ &= \frac{2\pi}{\lambda}\left((k_x\,\vec{v_x} + k_y\,\vec{v_y})\cdot\vec{R} + (k_x\,\vec{v_x} + k_y\,\vec{v_y})\cdot\vec{T}\right) \\ &= k_x\frac{2\pi}{\lambda}\left(\vec{v_x}\cdot\vec{R} + \vec{v_x}\cdot\vec{T}\right) + k_y\frac{2\pi}{\lambda}\left(\vec{v_y}\cdot\vec{R} + \vec{v_y}\cdot\vec{T}\right) \\ &= k_x\frac{dp_x}{dt} + k_y\frac{dp_y}{dt}. \end{aligned} \tag{6}$$

From (5) and (6) we can see that, the speed vector and the phase change rate have the same decomposition coefficients. Since $\overrightarrow{v_x}$ and $\overrightarrow{v_y}$ are orthogonal, their contributions to the phase change rate are linearly added. In the gesture recognition system, the speed vector $\overrightarrow{v_h}$ is the intrinsic variable, and the phase change rate $\frac{dp_h}{dt}$ is the observable variable. Using preamble gestures, we can have estimations of $\frac{dp_x}{dt}$ and $\frac{dp_y}{dt}$. Then if we know $k_x$ and $k_y$ from $\frac{dp_h}{dt}$, we can synthesize $\overrightarrow{v_h}$.

To solve $k_x$ and $k_y$, we need at least two groups of phase equations. That means at least two receivers (or two transmitters) are required to construct multiple groups of transmitter-user-receiver geometry relations. Due to the lack of space, we consider that there are only two receivers, the case of more receivers can be easily extended. Using similar terminologies, we have

$$\begin{aligned} \begin{bmatrix} \frac{dp_{h1}}{dt} & \frac{dp_{h2}}{dt} \end{bmatrix} &= \begin{bmatrix} k_x & k_y \end{bmatrix} \begin{bmatrix} \frac{dp_{x1}}{dt} & \frac{dp_{x2}}{dt} \\ \frac{dp_{y1}}{dt} & \frac{dp_{y2}}{dt} \end{bmatrix} \\ &= \begin{bmatrix} k_x & k_y \end{bmatrix} \boldsymbol{H}_{\text{tran}}, \end{aligned} \tag{7}$$

and

$$\begin{bmatrix} k_x & k_y \end{bmatrix} = \begin{bmatrix} \frac{dp_{h1}}{dt} & \frac{dp_{h2}}{dt} \end{bmatrix} \boldsymbol{H}_{\text{tran}}^{-1}. \tag{8}$$

The transformation matrix $\boldsymbol{H}_{\text{tran}}$ implies the impact of unit speed vector $\overrightarrow{v_x}$ and $\overrightarrow{v_y}$ on the phase change rates at receiver 1 and 2, and it involves the impact of transmitter-user-receiver geometry relation and user orientation. It is invariant as long as the user position and orientation are fixed. However, $\boldsymbol{H}_{\text{tran}}$ might be ill-conditioned if the phase change rates at two receivers are highly correlated.

With $\boldsymbol{H}_{\text{tran}}^{-1}$ in hand, we can estimate the speed vector $\overrightarrow{v_h}\ (t)$ at any moment,

$$\begin{aligned} \overrightarrow{v_h}(t) &= \begin{bmatrix} k_x(t) & k_y(t) \end{bmatrix} \begin{bmatrix} \overrightarrow{v_x} \\ \overrightarrow{v_y} \end{bmatrix} \\ &= \begin{bmatrix} \frac{dp_{h1}(t)}{dt} & \frac{dp_{h2}(t)}{dt} \end{bmatrix} \boldsymbol{H}_{\text{tran}}^{-1} \begin{bmatrix} \overrightarrow{v_x} \\ \overrightarrow{v_y} \end{bmatrix}, \end{aligned} \tag{9}$$

and the trajectory can be calculated as the integration of $\overrightarrow{v_h}\ (t)$,

$$\begin{aligned} \overrightarrow{s}\ (t) &= \int \overrightarrow{v_h}\ (t)dt \\ &= \int \begin{bmatrix} \frac{dp_{h1}(t)}{dt} & \frac{dp_{h2}(t)}{dt} \end{bmatrix} dt \cdot \boldsymbol{H}_{\text{tran}}^{-1} \begin{bmatrix} \overrightarrow{v_x} \\ \overrightarrow{v_y} \end{bmatrix} \\ &= \begin{bmatrix} p_{h1}(t) & p_{h2}(t) \end{bmatrix} \boldsymbol{H}_{\text{tran}}^{-1} \begin{bmatrix} \overrightarrow{v_x} \\ \overrightarrow{v_y} \end{bmatrix}. \end{aligned} \tag{10}$$

In this way, we can recover the gesture trajectory from the observed phases independent of the user position and orientation, and $\boldsymbol{H}_{\text{tran}}^{-1}$ is thus called the phase-trajectory transformation matrix.

There are two sources of error that might deteriorate the recovered trajectory. The first is the estimation error in the reflection path phase, which not only affect the estimation of transformation matrix $\boldsymbol{H}_{\text{tran}}$, but also affect the observed phase $p_{h1}(t)$ and $p_{h2}(t)$ in gesture recognition. When the moving speed is slow, or the moving direction is parallel to the border of Fresnel zone, the separation of dynamic path and the estimation of its phase are more fragile to noise. The second is the possible ill-condition of $\boldsymbol{H}_{\text{tran}}$ even there is no estimation error. The bad geometry relation will further amplify the impact of phase estimation error.

### C. Estimation of the Transformation Matrix

From (4) we know that, in ideal condition the transformation matrix $\boldsymbol{H}_{\text{tran}}$ can be obtained by two simple preamble gestures, pushing straight to the right and to the forward. However, in some configurations, the direction of $\overrightarrow{v_x}$ or $\overrightarrow{v_y}$ is close to the tangent of the Fresnel zone, and this will result in an unreliable $\boldsymbol{H}_{\text{tran}}$. Furthermore, a pushing operation is actually a variable acceleration process, it is hard to confirm which part in the motion period is the unit speed vector.

Therefore, we design a preamble gesture named "pre-O", which involves two continuous circles, to ensure that there are speed vectors in various directions and the velocities sampled at any moment are uniformly consistent. Since the start and end of a gesture may have accelerations, we can pick the speed vector samples in the intermediate part. Two circles are thus enough for us to have samples in every directions.

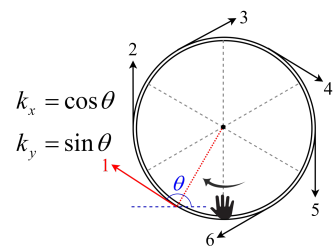


Fig. 2. Every speed sample is a linear composition of basis vectors.

Select $N$ motion segments at fixed moments as speed samples and we know the decomposition coefficients at each moment. For example, as shown in Fig. 2, taking 6 samples with uniform time interval, we have

$$\begin{bmatrix} \overrightarrow{v_o}\ (t_1) \\ \overrightarrow{v_o}\ (t_2) \\ \vdots \\ \overrightarrow{v_o}\ (t_N) \end{bmatrix} = \boldsymbol{K} \begin{bmatrix} \overrightarrow{v_x} \\ \overrightarrow{v_y} \end{bmatrix}, \quad \boldsymbol{K} = \begin{bmatrix} -0.866 & 0.5 \\ 0 & 1 \\ 0.866 & 0.5 \\ 0.866 & -0.5 \\ 0 & -1 \\ -0.866 & -0.5 \end{bmatrix}. \tag{11}$$

Fig. 3 shows the "pre-O" gesture trajectory and the corresponding dynamic phases at two receivers in a certain position configuration. The bold fragments stands for the sampled moments in trajectory and observed phases. Combining the observed phase change rates at $N$ moments, we can have a more accurate and robust estimation of the transformation matrix, i.e.,

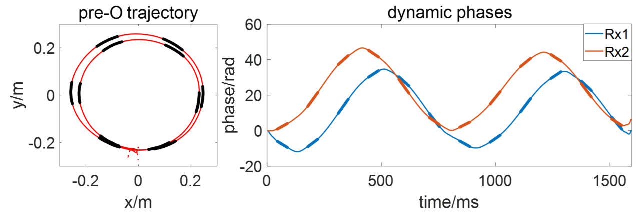


Fig. 3. Preamble gesture trajectory and the corresponding dynamic phases at two receivers.

$$\begin{bmatrix} \frac{dp_1(t_1)}{dt} & \frac{dp_2(t_1)}{dt} \\ \vdots & \vdots \\ \frac{dp_1(t_N)}{dt} & \frac{dp_2(t_N)}{dt} \end{bmatrix} = \boldsymbol{K} \begin{bmatrix} \frac{dp_{x1}}{dt} & \frac{dp_{x2}}{dt} \\ \frac{dp_{y1}}{dt} & \frac{dp_{y2}}{dt} \end{bmatrix}, \tag{12}$$

and,

$$\boldsymbol{H}_{\text{tran}} = \left(\boldsymbol{K}^T\boldsymbol{K}\right)^{-1}\boldsymbol{K}^T \begin{bmatrix} \frac{dp_1(t_1)}{dt} & \frac{dp_2(t_1)}{dt} \\ \vdots & \vdots \\ \frac{dp_1(t_N)}{dt} & \frac{dp_2(t_N)}{dt} \end{bmatrix}. \tag{13}$$

In practice, the "pre-O" gesture may not be drawn perfectly. The circle may become ellipse and the speed cannot be kept constant, which will bring estimation error to $\boldsymbol{H}_{\text{tran}}$. Dividing the circle into more segments, such as 8 or 12, or carrying the preamble with more rounds can help reduce this error. However, comparing with the possible ill-condition of $\boldsymbol{H}_{\text{tran}}$ incurred by bad geometry relations, we believe that the influence of imperfect circle is small.

### *D. Gesture Recognition*

We design six gestures to be recognized as shown in Fig. 4, and the hand moves from the blue end to the yellow end. The hand naturally moves in front of the body, and the movement range is roughly equivalent to the length of arm. A convolutional neural network (CNN) with 5 layers is designed to recognize the gesture trajectory, as shown in Fig. 5, where Conv1d_1, Conv1d_2 and Dense_1 are followed by the ReLU activation function, and Dense_2 is followed by the softmax activation function. The input layer has dimension of (850, 2), where 850 refers to gesture duration of 850 ms with CSI sampling interval 1 ms, and 2 refers to the X and Y coordinates. According to our experiments, generally the gesture completion time of a gesture is within 850 ms, thus we cut of 850 ms segment of the separated dynamic path to process. Since CNN has translation invariance, the segment involving gesture is not necessarily at the central, small amount of offset is allowed.

Since it is hard to collect a large amount of real experiment datas, we have simulated 9000 trajectory datas to train the neural network. But these simulation data will not be used for test, in the experiments, we will use the CSI obtained from real receivers to test the performance of gesture recognition.

The phase extraction and phase-trajectory transforming may induce twisting, stretching and various kinds of distortion. It is not enough to only add noise on the trajectory to generate the training data. We established a simulation system involving the whole procedure occurred in real system, which mainly includes a trajectory generation module, a CSI generation module, a transformation matrix solving module, and a trajectory recovery module. The trajectory generation module simulates the random start position, speed, movement range, and the acceleration adjustment at the beginning and end stages. The CSI generation module simulates the multipath propagation and channel noise, and separates the dynamic path from the static paths. The transformation matrix solving module calculates the phase-trajectory transformation matrix and the trajectory recovery module reconstructs the trajectory from the observed phases of the dynamic paths. In the simulation, 6 position configurations are used, and in each configuration 1500 gestures are randomly generated. The recognition rate on validation set is 99%.

In the experiments, given a fixed user position and orientation, the preamble gesture is first performed and $\boldsymbol{H}_{\text{tran}}$ is estimated. Then the recovered trajectory is sent to the CNN to classify. Since the simulation system considered main affecting factors, the CNN trained by simulation data performs well in real experiment datas.

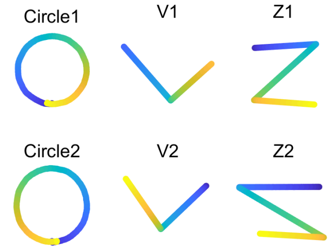


Fig. 4. Illustration of 6 gestures.

| | | |
|---|---|---|
| **InputLayer** | Input: | (None,850,2) |
| **Conv1d_1** | Input: | (None,850,2) |
| | Output: | (None,284,32) |
| **MaxPooling1D** | Input: | (None,284,32) |
| | Output: | (None,141,32) |
| **Conv1d_2** | Input: | (None,141,32) |
| | Output: | (None,141,64) |
| **Flatten** | Input: | (None,141,64) |
| | Output: | (None,9024) |
| **Dense_1** | Input: | (None,9024) |
| | Output: | (None,64) |
| **Dense_2** | Input: | (None,64) |
| | Output: | (None,6) |

Fig. 5. The structure of CNN.

## III. Experiments

We build a prototype system based on LTE signals to verify the proposed method. To ensure no other movement in the environment, we are not using the signal from commercial base station (BS), yet the LTE signal is transmitted by a software radio. The carrier frequency is set as 2.27 GHz and the corresponding wavelength is 13 cm. The maximum transmit power is 10 dBmW, it has good coverage in a laboratory and cause no interference to other rooms. We have measured that the received signal power is similar with that coming from a commercial micro-BS, which has a maximum transmit power of 30 dBmW.

In Rx side, a software radio platform YunSDR Y550s is used, which has four channels of RF links and can be used as two receivers. Each receiver has two antennas, and CSI ratio between two antennas are used as the CSI obtained in this receiver, to eliminate the impact of phase noise [10]. The software radio platform implements down-conversion and sampling, and all baseband processing are implemented in the host computer. The whole system is implemented in real-time, and we can observe the experiment results without delay.

The experiment scenario is shown in Fig. 6, the system automatically intercepts the CSI segment with dynamic reflection paths, and displays the CSI and the reconstructed trajectory on the screen. The obtained CSI ratios and phase changes in one example are shown in Fig. 7, where the user is standing in front of the midpoint of Rx1 and Rx2, and performs Z1 gesture.

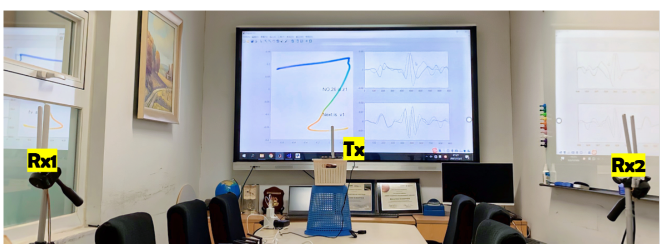


Fig. 6. Photo of one experiment room.

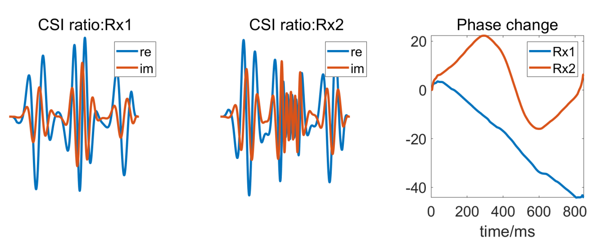


Fig. 7. CSI ratios obtained at two receivers, and the extracted phases of reflection paths.

At the beginning of each test experiment, user needs to perform the "pre-O" gesture, and the system automatically intercepts the effective segments, calculates the transformation matrix $\boldsymbol{H}_{\text{tran}}$. Then the user performs a set of gestures according to 60 action commands which are randomly provided by the system.

We verify the system performance under two position configurations, as shown in Fig. 8, where the user is at the coordinate origin and the receivers are located at different positions. These two position configurations have not been shown up in the training stage, and there are dense static reflections in the experiment rooms. Since the user orientation is relative to the direction of transmitter and receivers, the user orientations under these two configurations are actually different.

We also draw the Fresnel zones along with Tx-Rx1 and Tx-Rx2 pairs. Given the same gesture, for example V1 shown in the figure, the obtained CSIs are different for different position configurations. The phase of reflection path only changes when the hand crosses Fresnel zones. If the hand moves along the tangent line of the Fresnel zone, the phase change is slow and the reconstructed trajectory will have severe distortion.

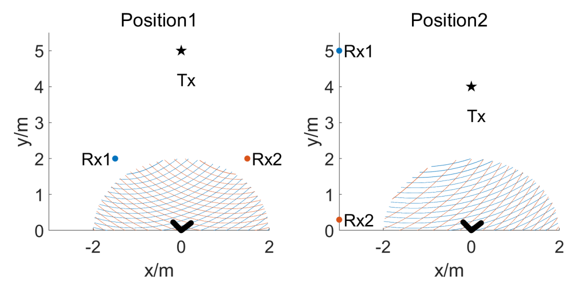


Fig. 8. Two kinds of position configurations used in experiments, and the corresponding Fresnel zones.

Fig. 9 and 10 show some reconstruction results of six gestures under two position configurations. Although the moving ranges and speeds in each test are slightly different, the results of position 1 are quite satisfactory. The reconstructed trajectories are clear and similar with the designed gestures. It only has small offset at the beginning, end segments and turning points due to slow and tiny movements. The results of positions 2 have some distortions, especially for V1, V2 and Z1. The right half of V does not cut the Fresnel zone obviously, so the reconstructed right segments of V1 and V2 as well as the middle segments of Z1 have obvious offset. Thus V1 and V2 are easily confused, and some reconstructed Z1 looks like V1.

The statistical results of recognition accuracies and confusion matrix are shown in Table I and Fig. 11, where five groups of experiments are conducted in each position configuration. The experiment results show that the proposed method is feasible to reconstruct and recognize gestures in different environments without any priori knowledge of the position and orientation. They also show that in some position configurations, due to ill-condition of the transformation matrix or the weak response of specific movements, the recognition accuracy is to be improved.

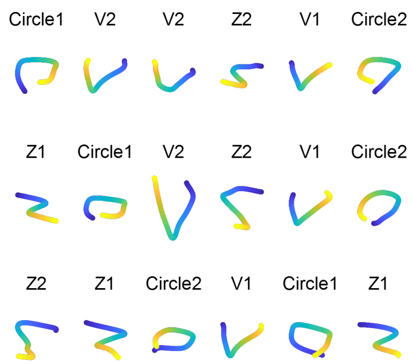


Fig. 9. Reconstructed gesture trajectories under position configuration 1.

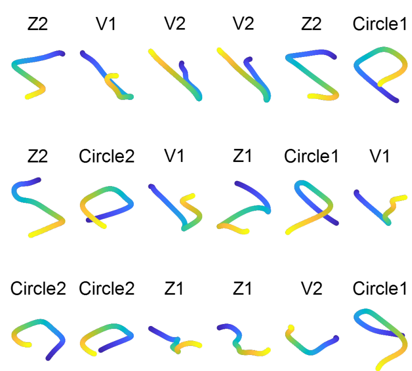


Fig. 10. Reconstructed gesture trajectories under position configuration 2.

The most direct method to improve the accuracy under any possible position configurations is by adding more Tx-Rx pairs. So that at least two pairs of transceivers can satisfy the requirements that the signal is not blocked by the human body and has obvious reflection path phase changes. In real applications, it would be effective to increase the available Tx-Rx pairs by listening several neighboring base stations simultaneously. In addition, the current method can only serve in the environment where only one moving object exists. To inhibit the disturbance of other random activities, or to pursue multi-user gesture recognition, we believe that it would be better to improve the delay resolution and spatial resolution with larger bandwidth and more antennas [11], [12].

TABLE I
TEST RESULTS

| Accuracy(%) \ Test / Position | 1 | 2 | 3 | 4 | 5 |
|---|---|---|---|---|---|
| Pos1 | 98.3 | 95 | 95 | 96.7 | 95 |
| Pos2 | 85 | 86.7 | 88.3 | 83.3 | 85 |

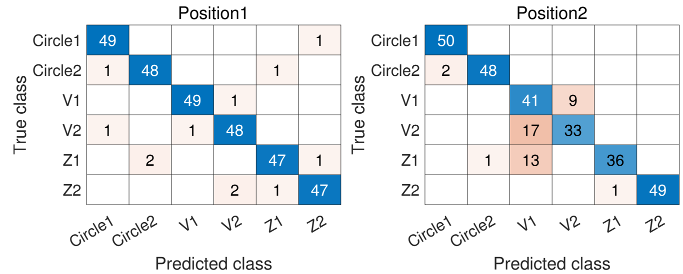


Fig. 11. Confusion matrices of 300 gestures tested in two position configurations. Accuracy rates are 96% and 85.7% in position 1 and position 2, respectively.

## IV. Conclusion

In this paper, we proposed a position and orientation independent gesture recognition method. With the help of preamble gesture, a phase-trajectory transformation matrix is first estimated, and then the gesture trajectory can be recovered and recognized without retraining in a new environment. We analyzed the source of distortions, and designed a "pre-O" gesture to better estimate the transformation matrix. We built a real-time prototype system to verify the feasibility and test the performance under different position configurations. Robust recognition under any circumstances will be assured with more Tx-Rx pairs.

## References


[1] S. Xu and Y. Tian, "Device-free motion detection via on-the-air LTE signals," *IEEE Commun. Lett.*, vol. 22, no. 9, pp. 1934–1937, Sep. 2018.

[2] I. Nirmal, A. Khamis, M. Hassan, W. Hu, and X. Zhu, "Deep learning for radio-based human sensing: Recent advances and future directions," *IEEE Commun. Surveys Tuts.*, vol. 23, no. 2, pp. 995–1019, Second Quarter 2021.

[3] D. Zhang, H. Wang, and D. Wu, "Toward centimeter-scale human activity sensing with Wi-Fi signals," *Computer*, vol. 50, no. 1, pp. 48–57, Jan. 2017.

[4] Y. Tian, Y. He, and H. Duan, "Passive localization through channel estimation of on-the-air LTE signals," *IEEE Access*, vol. 7, pp. 160 029–160 042, 2019.

[5] A. Virmani and M. Shahzad, "Position and orientation agnostic gesture recognition using WiFi," in *Proc. ACM MobiSys 2017*, pp. 252–264.

[6] W. Jiang, C. Miao, F. Ma, and et al., "Towards environment independent device free human activity recognition," in *Proc. ACM MobiCom 2018*, pp. 289–304.

[7] Y. Zheng, Y. Zhang, K. Qian, and et al., "Zero-effort cross-domain gesture recognition with Wi-Fi," in *Proc. ACM MobiSys 2019*, pp. 313–325.

[8] R. Gao, M. Zhang, J. Zhang, and et al., "Towards position-independent sensing for gesture recognition with Wi-Fi," *Proc. ACM Interact. Mobile Wearable Ubiquitous Technol.*, vol. 5, no. 2, pp. 1–28, Jun. 2021.

[9] Z. Shi, J. A. Zhang, R. Y. Xu, and Q. Cheng, "Environment-robust device-free human activity recognition with channel-state-information enhancement and one-shot learning," *IEEE Trans. Mobile Comput.*, vol. 21, no. 2, pp. 540–554, Feb. 2022.

[10] Y. Zeng, D. Wu, J. Xiong, and et al., "FarSense: Pushing the range limit of Wi-Fi-based respiration sensing with CSI ratio of two antennas," *Proc. ACM Interact. Mobile Wearable Ubiquitous Technol.*, vol. 3, no. 3, pp. 1–26, Sep. 2019.

[11] R. Peng, Y. Zhao, Y. Tian, and S. Han, "Device free wireless gesture recognition by 5G-NR signal," in *Proc. IEEE/CIC ICCC 2022*, pp. 1–6.

[12] R. Peng, Y. Tian, and S. Han, "Multiuser wireless hand gesture recognition by spatial beamforming," *IEEE Trans. Veh. Technol.*, early access.