WiHF: Gesture and User Recognition with WiFi
Chenning Li, Student Member, IEEE, Manni Liu, and Zhichao Cao, Member, IEEE

Abstract—User identified gesture recognition is a fundamental step towards ubiquitous WiFi based sensing. We propose WiHF, which first simultaneously enables cross-domain gesture recognition and user identification using commodity WiFi in a real-time manner. The basic idea of WiHF is to derive a domain-independent motion change pattern of arm gestures from WiFi signals, rendering the unique gesture characteristics and the personalized user performing styles. To extract the motion change pattern in real time, we develop an efficient method based on the seam carving algorithm. Moreover, taking as input the motion change pattern, a Deep Neural Network (DNN) is adopted for both gesture recognition and user identification tasks. In DNN, we apply splitting and splicing schemes to optimize collaborative learning for dual tasks. We implement WiHF and extensively evaluate its performance on a public dataset including 6 users and 6 gestures performed across 5 locations and 5 orientations in 3 environments. Experimental results show that WiHF achieves 97.65% and 96.74% for in-domain gesture recognition and user identification accuracy, respectively. The cross-domain gesture recognition accuracy is comparable with the state-of-the-art method, but the processing time is reduced by 30×.

Index Terms—Gesture Recognition; User Identification; Channel State Information; Commodity WiFi, Cross-domain Sensing

1 INTRODUCTION

Every time I lift my arm, it distorts a small electromagnetic field that is maintained continuously across the room. Slightly different positions of my hand and fingers produce different distortions and my robots can interpret these distortions as orders. I only use it for simple orders: Come here! Bring tea! and so on. Such an amazing scenario was described in the science fiction “The Robots of Dawn” [1] by Isaac Asimov in 1983. Nowadays, WiFi based sensing is making it happen and researchers have proposed several WiFi based systems for gesture recognition [2], [3], [4], [5], [6], [7], [8], which can improve the efficiency and quality of human living in the smart home. The fundamental principle that enables humans to naturally interact with smart devices or even robots via WiFi is that WiFi channels get distorted by arm or hand gestures as shown in Figure 1.

Besides the semantic meaning of diverse gestures, many applications usually require user identities, such as the smart factory, smart home, and VR gaming for the access control, content recommendation, and VR customization. Specifically, access control is usually required for gesture commands with different types of equipment in the smart factory. When the identity of children or parents can be known, it can recommend different contents when they are watching TV or listening to music. Besides, VR gaming can be further enhanced by associating the performed gestures with specific users, such as recording the user’s information. As shown in Figure 1, the true potential of WiFi based gesture recognition can be unleashed only when it can associate the performed gesture with a specific user simultaneously, just like HuFu used in the ancient Chinese military, which can simultaneously provide authentication (e.g., match a pair of HuFu pieces) and semantic meaning (e.g., deploy military force). Hence, the user identified gesture recognition becomes an emerging research topic.

Nevertheless, user identified gesture recognition encounters three fundamental challenges in practice. First, since the WiFi signals are usually noisy, it is difficult to derive an effective feature to represent both the unique gesture characteristics and the personalized user performing styles. Thus it cannot naturally support gesture recognition and user identification simultaneously. Second, the computation complexity of feature extraction should be low enough so that we can apply user identified gesture recognition in a real-time manner. Third, since a user may perform the same gesture in diverse locations, orientations, and environments, the recorded WiFi signals are no longer the same at all. It is challenging to keep user identified gesture recognition accurate across domains (e.g., locations, orientations, environments) [7] with low extra overhead.

State-of-the-art approaches fail to resolve all three challenges. For example, Widar3.0 [7] enables a zero-effort cross-domain gesture recognition, which means extra efforts are unnecessary in either data collection or model retraining when gestures are performed in new domains. The feature
extracted however cannot support user identification and the computation complexity of feature extraction is too high to achieve real-time. Moreover, WiID [9] achieves user identification for gestures while it must take as input the performed gesture, which can restrict the user identification accuracy significantly. Consequently, real-time efficiency and cross-domain ability are degraded.

In this paper, we propose WiHF (WiFi HuFu), a pioneering attempt to achieve cross-domain user identified gesture recognition using commodity WiFi in a real-time manner. It aims to capture the personalized motion change pattern caused by arm gestures, which includes the rhythmic velocity fluctuation and characteristic pause distribution. Moreover, it keeps consistent across domains. Then an efficient method is developed to carve the motion change pattern for computation efficiency based on the seam carving algorithm [10]. Finally, we design a collaborative dual-task DNN model to simultaneously achieve accurate user identified gesture recognition, which further applies the splitting and splicing scheme to bootstrap the cross-domain ability and collaborative learning.

We implement WiHF and evaluate it extensively on a public dataset [7]. Results demonstrate WiHF achieves 97.65% and 96.73% for in-domain gesture recognition and user identification, respectively. Moreover, WiHF demonstrates zero-effort cross-domain characteristics for gesture recognition, comparable with the state-of-the-art methods [7]. The processing time is however reduced by 30× and thus can be running in real-time. In a nutshell, the contributions of this paper are as follows:

- We design a domain-independent motion change pattern of arm gestures to reduce the deployment cost for real applications and develop several efficient algorithms for real-time operation.
- We propose a dual-task DNN framework that can recognize gestures and identify users collaboratively. And it can be further generalized to multi-task sensing using wireless signals.
- We implement WiHF and conduct extensive experiments to evaluate its performance. Evaluations demonstrate the feasibility and effectiveness of our system.

The rest of this paper is organized as follows. Related works are reviewed in Section 2 followed by the preliminary and observation in Section 3. The system design is detailed in Section 4 before the performance evaluation of Section 5. We provide the discussion and summary the conclusion in Section 6 and 7, respectively.

2 RELATED WORK

2.1 User Identification:

Prior work can be divided into passive and active user identification, depending on whether it requires extra actions of users. Passive user identification generally leverages intrinsic physiological distinctions or behavioral features, such as the gait [11], [12], [13], [14], [15], [16], [17] and location-oriented activity patterns [18], [19], [20], [21], [22]. Specifically, WiWho [11], WifiU [16], WiFi-ID [13], and Wii [17] represent user-specific gait features with statistical features from WiFi measurements. AutoID [14] proposes the $C^3SL$ shapelet learning framework to extract the unique fine-grained gait patterns. All however either require users to walk along pre-defined paths and keep their walking style while avoiding any turns or breaks. For activity based user identification, WiPIN [20] extracts 4 human biologic features and 9 statistical signal features using SVM while Shi et al. [21] leverage the statistical features and employ a deep learning model to capture the fingerprints of WiFi measurement for different users. All passive user identification methods however cannot convey extra information. Besides, they are vulnerable to domain variance such as the location, orientation and environment of users. In contrast, active user identification relies on users’ actions and conveys extra information such as users’ gesture commands. Traditional methods validate credentials actively by requiring fingerprint, iris or RFID tags in a specific position. RF-Mehndi [23] leverages the unique phase patterns of the backscatter signals caused by the user’s fingertip touching. It requires the RFID card as a credential and the system works in a distance less than 30cm. To the best of our knowledge, WiID [9] is the first scheme that can recognize gestures while identifying the user at runtime by integrating existing WiFi based gesture recognition systems as an add-on module. Moreover, the system requires data collection and model retraining when employed in a new domain.

2.2 Gesture Recognition across domains:

WiFi based gesture recognition systems generally need to be adapted for new domains. Researchers have proposed to either translate features between domains [3] or generate domain-independent features [7]. Widar3.0 [7] extracts the domain-independent feature BVP from CSI but requires the accurate location of transceivers, otherwise, it may suffer due to the noise and outliers [24]. Importantly, none of aforementioned solutions can be real-time. The domain adversarial architecture [25] can also reduce the deployment cost for new domains by removing the domain-dependent components from the extracted features [26], [27].

3 PRELIMINARY AND OBSERVATION

In this section, we first discuss the preliminary principles behind the user identification based on arm gestures across domains (§3.1). Then, we demonstrate the feasibility to utilize the motion change pattern of arm gestures for user identified gesture recognition with WiFi (§3.2).

3.1 Arm Gesture based User Identification across Domains

We survey the existing works to verify two fundamental questions as follows:

3.1.1 Are the characteristics of arm gestures representative enough for user identification?

The answer is true due to the following observations. During performing an arm gesture, the potential biometric feature [9], [28], [29], [30], [31], which associates with the shape of arm and hand, is tightly coupled with the corresponding movements and brings an opportunity for...
user identification. Moreover, arm gestures, including arm sweep [30] and gesture vocabulary (e.g., drawing the line, circle, rectangle) [28], are diversely performed from user to user due to their different semantic understandings and performing styles, then have been used for user authentication. Several works have shown these unique characteristics of arm gestures can be captured by either WiFi signals or inertial sensors. Specifically, WiID [9] conducts a comprehensive measurement study to validate that the time-series of the frequencies, appearing in WiFi Channel State Information (CSI) measurements while performing a given gesture, is different from that of the same gesture performed by different users but similar to that performed by the same user in a long period. Moreover, wrist acceleration samples during performing arm gestures are collected with inertial sensors and able to provide personalized characteristics with long term stability over a month [29], [31]. Hence, the characteristics of arm gestures are indeed representative enough for user identification.

3.1.2 Can we achieve accurate user identification while avoiding extra re-training efforts across domains?

The key challenge whether a user identification system can adapt the various inputs of the same arm gesture performed by the same user across domains is to extract a domain-independent feature free from the induced variances. For example, Widar3.0 [7] derives a domain-independent feature called Body-coordinate Velocity Profile (BVP) from the Doppler Frequency Shift (DFS) spectrogram of raw CSI measurements. It summarizes the occurred relative velocity components of arm motion in the user’s body coordinate system which is irrelevant to the user’s orientations, locations, and surrounding environments. With BVP, Widar3.0 achieves cross-domain gesture recognition without the extra cost of data collection and model retraining. BVP shows the possibility of the domain-independent feature design, but it is still a challenging problem for user identification because BVP cannot preserve the personalized characteristics while performing arm gestures. Besides, BVP is too computation-intensive to be running in real-time. Hence, to solve the problem, a possible way is to derive a new domain-independent feature for both gesture recognition and user identification.

Overall, to enable user identified gesture recognition in real-time, we need to derive a new feature of arm gestures, which is domain-independent and supports both gesture recognition and user identification with efficient computation complexity.

3.2 Empirical Study of WiFi Motion Feature

We intend to investigate several feasible personalized features preserved by WiFi CSI measurements while performing arm gestures. Upon receiving the raw CSI measurements, we can derive the DFS spectrogram using Short-Term Fourier Transform (STFT) to capture the distortion of channels induced by the arm motions. By observing the DFS spectrogram, we notice that both power and temporal features (called carving paths) have the potential to indicate the personalized arm motion from velocity and rhythm aspects. Specifically, DFS spectrograms [5] can separate the movement of different arm parts when they move at different speeds since the spectrogram power varies as the reflection areas change for the certain velocity component over time. Based on the power of DFS spectrograms, two carving paths can be derived. One is Dominant Power which reflects the most dominant power in the DFS spectrograms. The other is Power Bound which profiles the dominant power area and velocity bounds [32]. Note that both are power based features. Moreover, an arm gesture can be usually divided into some atomic motions (e.g., drawing the line, arc) in the temporal order. For example, drawing the rectangle contains four lines towards four different directions. The switches between two adjacent atomic motions are called Motion Changes, which indicate motion pause/restart. We extract a carving path, called Motion Change Pattern, to represent the pattern of motion changes, namely the temporal rhythmic motion during performing the gesture.

To validate whether the three carving paths are distinguishable among different users and stable for any single user across domains, we further conduct some empirical experiments. With the collected CSI measurements of each gesture instance, we manually extract and label the three carving paths from DFS spectrograms. We have three observations as follows. Note that WiHF builds its all work with the public dataset. And details about the experimental setup can refer to Widar3.0 [7].

3.2.1 The motion change pattern of an individual user performing the same gesture stays consistent over time while different users manifest various motion change patterns for performing the same gesture, but it does not hold for power based carving paths all the time

The three top sub-figures of Figure 2 show the three carving paths of three different users while performing the same gesture (e.g., drawing the rectangle). Specifically, the black and red dashed lines denote the carving paths of dominant power and power bound. The pink dashed lines which are distributed along the axis of frequency shift indicate the motion changes. With the visualization of all carving paths, we can observe both power based features and the motion change pattern varies among different users. Intuitively, different users perform the same gesture with the personalized action understanding and performing style. Among different users, their diverse arm shapes and sizes reinforce the impacts on WiFi signals, leading diverse power based features and the motion change pattern.

Further, we collect three instances of the same gesture from all three users and superpose their carving paths for each user in the bottom three sub-figures of Figure 2. We can see the power based carving paths of different instances may shift along the time axis, especially for the third user. In contrast, for all users, their motion changes can be grouped into three clusters which correspond to the three pauses during drawing the rectangle. In all clusters, the largest period between two motion changes across different instances is less than 70ms (the first cluster of the third user appeared at about 500ms), which demonstrates its consistency of the motion change pattern for each user across instances. The reason behind this is the inevitable noise (e.g. multi-path, body motion) has a significant influence on the power based carving paths, but motion changes are less affected.
3.2.2 An individual user introduces similar motion changes when performing gesture across domains, but the power based carving paths vary significantly

To verify the stability of different carving paths across domains, a user performs two gestures (e.g., drawing the zigzag and triangle) at three different locations. For each gesture, the distribution of the derived carving paths is shown in Figure 3. We can see the motion change pattern is consistent across locations for both gestures. The largest period between two motion changes in a cluster is 50ms (the second cluster in the top sub-figure for drawing the zigzag). Meanwhile, the power based carving paths exhibit noticeable dynamics when a gesture is performed in different locations. The reason is the motion change pattern reflects the rhythm of the arm motion and the temporal characteristic is only related to the user performing style and gesture composition, but not where the user performs the gesture. And the diverse arm shapes and sizes of users reinforce the impacts on WiFi signals. In contrast, in the view of WiFi transceiver pairs, the velocity components of the detected arm motion vary with the location changes so that incurs DFS power dynamic since the human body is best modeled as a quasi-specular reflector [33].

Overall, in comparison with power based features, the motion change pattern is a better feature to achieve user identified gesture recognition across domains. The experimental results further verify the above intuition quantitatively (§5.5).

4 System Design

Based on the empirical observations, we propose WiHF to leverage the motion change pattern and designs a collaborative dual-task module to recognize gestures and identify users simultaneously. Figure 4 provides an overview of the architecture of WiHF. WiHF consists of three modules, from bottom to top, including data acquisition, pattern extraction, and collaborative dual-task, respectively.

Data Acquisition Module: Upon receiving raw CSI measurements from WiFi transceiver pairs, WiHF first sanitizes the CSI series using the band-pass filter and conjugate multiplication [34], [35]. Then dominant DFS spectrogram components reflected by different body parts (e.g. hand, elbow, arm) are collected using Principal Component Analysis (PCA). Thus we can remove the interference while retaining the unique gesture characteristic and the personalized user performing styles (§4.1).
Pattern Extraction Module: This module extracts the domain-independent motion change pattern [§4.2]. To derive the DFS spectrogram, we first operate the time-frequency analysis by adopting STFT on the sanitized CSI. Then, WiHF develops an efficient method based on the seam carving algorithm [10] to capture the motion change pattern readily, which is fed into the collaborative dual-task module.

Collaborative Dual-task Module: This module works for collaborative classification tasks including gesture recognition and user identification at runtime [§4.3]. First, the motion change pattern is filtered and split to the corresponding dual inputs for the dual tasks. Then a DNN of convolution based Gated Recurrent Units (GRU) [36] extracts the spatial (e.g., different body parts) and temporal (e.g., motion change pattern) features of gesture motion. Next, the gradient block layer is integrated for splicing respective features while ensuring that they do not affect each other during the back-propagation with the loss function. Finally, it outputs the predictions for gesture recognition and user identification simultaneously.

4.1 CSI Acquisition

A time series of CSI matrices characterizes MIMO channel variations from different dimensions including time (packet), frequency (subcarrier), space (transceivers). For a MIMO-OFDM channel with M transmit antennas, N receiver antennas, K subcarriers and T received packets, the 4-D CSI tensor $H \in C^{N \times M \times K \times T}$ can be formulated as follows at packet $t$, frequency $f$ and receiver antenna $a$ [37], representing amplitude attenuation and phase shift of multi-path channels:

$$H(f, t, a) = (H_s(f, t, a) + H_d(f, t, a) + N(f, t, a))e^{j\epsilon(f, t)}$$

(1)

where $\epsilon(f, t)$ is the phase offset caused by cyclic shift diversity, sampling time offset, sampling frequency offset [37]. $N$ is the complex white Gaussian noise capturing the background noise [38]. $H_s$ is the static component with zero DFS while the dynamic component $H_d$ is a superposition of vectors with time-varying phases and amplitudes [35].

Generally, we remove the phase offset $\epsilon(f, t)$ by calculating the conjugate multiplication with raw CSI measurements of two antennas on the same WiFi receiver [7], [38]. Then $H_s$ and $N$ are filtered out using the band-pass filter [39]. The remaining $H_d$ is only affected by the motion of multiple body parts. We further obtain the dominant components using PCA. The key question is how many components should be selected for feature extraction. The more components are selected, more motion features can be extracted but the processing time is sacrificed. In existing works, Widar uses the first component with filters [7], [39] while some others [3], [9] demonstrate that the third component contains the most feasible motion features without filters.

We illustrate diverse PCA components using the public dataset of Widar3.0 [7]. Figure 5 shows that different components contain diverse scales of DFS spectrograms, which is significant for retaining the unique gesture characteristic and the personalized user performing style. For example, PCA #1 shows the dominant power distribution while PCA #3 provides finer-grained motion change pattern which can represent the symmetrical arm movements for the clap gesture. Note that the first three components are selected for motion change pattern profiling while balancing the computation cost.

4.2 Motion Change Pattern Extraction

Upon receiving the dominant CSI tensor series, WiHF first applies STFT and obtains the DFS component $f_D$. And it’s incurred by the movement of the arm and associated with the velocity of different body parts in Equation (2), where $\alpha_i$ is the complex attenuation factor of the $i_{th}$ path [38] and $P_d$ is the set of dynamic paths ($f_D \neq 0$) [5], [7], [39]. Thus we derive the DFS spectrogram with the dimension as $R \times P \times F \times T$, where $R$ and $P$ are the numbers of transceiver links and PCA components. And $F$ and $T$ denote the sampling points in frequency and time domain, respectively.

$$H_d(f, t, a) = \sum_{k \in P_d} \alpha_k(f, t, a)e^{j2\pi f_k f_{D_k}(a)}du$$

(2)

As mentioned in (§3.2), users express unique personalized styles while performing the same gesture, resulting in the rhythmic increase, drop or even pause at certain instances. Thus signals reflected by diverse body parts generate consistent motion change patterns and form the corresponding DFS spectrogram sequence. The remaining challenge is how to extract the motion change pattern
efficiently. Intuitively, the rhythmic increase, drop or even pause usually induce noticeable moving velocity fluctuation detected by the DFS spectrogram. And it occurs at certain moments representing the velocity change peaks in the time domain. However, intensive computation for the derivative operation of velocity sacrifices the real-time characteristic. To retain personalized features while balancing the processing time, the motion change pattern is derived out of carving DFS spectrograms.

The basic idea is to extract the motion change pattern comparable with acceleration as biometrics for dominant body parts, such as wrist, elbow, arm. However, we are facing three challenges. First, DFS only demonstrates the power value directly for the specific velocity component over time. It cannot provide the accurate fine-grained acceleration information corresponding to various body parts due to the superimposition of velocity components at the receiver. Second, the motion change pattern requires the derivative calculation of high dimension data, which is computation-intensive and cannot be running in real time. Third, the DFS spectrogram contains excessive irrelevant interference, resulting in unnecessary computation and memory.

For the challenge for the superimposition of velocity components, we propose a model to fill the gap between the power distribution of the DFS spectrogram and the body part acceleration information, leading to the motion change pattern. As CARM [5] elaborates, the power distribution of spectrograms changes as reflection areas $S$ vary for specific Doppler shift frequency at instant $t$. Thus the power $P_{ds}$ for the DFS Spectrogram can be defined with the scaling factor $c$ due to propagation loss as the Equation (3a). Assuming $K$ body parts are dominant to define the gesture, the relation between $P_{ds}$ and the superposed reflected area $S$ of body parts can be modeled as Equation (3b):

$$P_{ds}(f_D, t) = c \cdot S(f_D, t)$$

$$P_{ds}(f_D, t) = c \cdot \sum_{k=1}^{K} R{ef}(k, t) \cdot \mathbb{1}(f_{dfs}(k, t) = f_D)$$

where $R{ef}(\bullet)$ and $f_{dfs}(\bullet; \bullet)$ denote the individual reflection area and Doppler shift frequency at time $t$ for the $k$-th body part. And $\mathbb{1}(\bullet)$ represents the indicator function while $c$ and $f_D$ denotes the scaling factor and the DFS component, respectively.

However, accurate $P_{ds}$ is non-reachable due to the resolution of the WiFi signals and the approximate estimation for body parts $K$. Thus we can get the experimental approximation $\hat{P}_{ds}$ with acceptable computing error $\varepsilon$ and attenuation factor $a_{at}$ as (4a). For accessibility and derivability, Gaussian distributions is applied to model the superimposition for movements of body parts as (4b):

$$\hat{P}_{ds}(f_D, t) \approx c \cdot \sum_{k=1}^{K} R{ef}(k, t)a_{at}R{ect}(\frac{f_{dfs}(k, t) - f_D}{2\varepsilon})$$

$$\hat{P}_{ds}(f_D, t) \approx c \cdot \sum_{k=1}^{K} R{ef}(k, t) \cdot e^{-\frac{(f_{dfs}(k, t) - f_D)^2}{4(\varepsilon/2)^2}}$$

Since $f_{dfs}$ demonstrates the moving velocity $v$ with $v = f_{dfs} \times \frac{\lambda}{2}$ [38], [39], the corresponding acceleration $a_k$ of each body part can be denoted as $\frac{\partial}{\partial t} f_{dfs}(k, t) - f_{D0}$ for a fixed $f_{D0}$. Assuming the $R{ef}(\bullet)$ variance can be omitted compared with the superimposition effects on $\hat{P}_{ds}$ between consecutive DFS spectrogram, power change rate can be derived as Equation (5a). With the limitation of rigid body part of human and acceleration continuation, we can properly decimate the derivative of power as Equation (5b) to simplify the relationship between the DFS power change and acceleration $a_k$:

$$\frac{\partial \hat{P}_{ds}(f_{D0}, t)}{\partial t} \approx c \cdot \sum_{k=1}^{N} R{ef}(k, t) \cdot \frac{9\Delta_k}{\varepsilon^2} e^{-\frac{9\Delta_k^2}{2\varepsilon^2}} \cdot a_k$$

$$\Delta_k = |f_{dfs}(k, t) - f_{D0}|, a_k = \frac{\partial \hat{P}_{ds}(f_{D0}, t)}{\partial t}$$

Thus we find the power change rate increases as $a_k$ rises for all $K$ body parts. The personalized acceleration information as biometrics over time [28], [29], [30], [31] can be detected when users perform gestures by computing the derivative of DFS spectrograms.

Further, the remaining challenges are the efficiency of the derivative calculation and the interference of redundant data. And we design the algorithm based on the seam carving problem in computer graphics for content-aware image resizing [10]. First, we filter the redundant interference with edge detection methods while optimizing derivative calculation using the difference scheme with the convolution operator. Then we develop an efficient method based on the seam carving algorithm to generate multiple dominant carving paths mentioned in (§3.2) as the motion change pattern for each DFS spectrogram of PCA components. Suppose estimations of $K$ dominant carving paths are considered and each path demonstrates the most significant motion change pattern over time. We use $w_{i,j} \in [0, 1]$ to denote the weight of the $\frac{\partial \hat{P}_{ds}(f_D, t)}{\partial t}_{i,j}$ for the $i$-th frequency bin at $j$-th packet. Thus, the optimal Motion Change Pattern (MCP) along the frequency axis, as the function of indices of timestamps, can be defined as:

$$MCP_{opt} = MCP(\text{argmax}_{i} \sum_{j=1}^{F_D} w_{i,j} \frac{\partial \hat{P}_{ds}(f_D, t)}{\partial t}_{i,j} \)$$

where $F_D$ denotes the numbers of frequency bins with STFT.

For computation efficiency, the Sobel operator for the time axis [40] is applied on DFS spectrogram $\hat{P}_{ds}(f_D, t)$ thus we can get the temporal gradient matrix for each DFS spectrogram.

Algorithm 1 elaborates on the motion change pattern carving algorithm. Upon receiving the complete DFS spectrograms from each receiver, the mean compressing method is applied for retaining the unique patterns while avoiding redundant computation. Assigning an appropriate value to the number of segmentation $T_s$ is crucial since if $T_s$ is too large, the segmentation may be instantaneous and cannot guarantee the robust feature extracted by carving paths with a small sliding window. In contrast, with a small $T_s$, carving paths can become similar because the unique motion change pattern of individual users get too averaged out to be distinguished with a large sliding window. Moreover, an adaptive
Algorithm 1 Motion Change Pattern Extraction

Input: \([M_{\text{Pow}}, D_F \times T_s, W_{D_F \times T_s}, \text{ker}_{\text{sobel}}, K_{\text{path}}, T_s]\)
Output: \(MCP_{\text{VelocityBins}} \times T_s\)

1: \([M_{\text{Pow}}, D_F \times T_s] = \text{meanCompressing}(M_{\text{Pow}}, D_F \times T_s);\)
2: for \(n = 1\) to \(K_{\text{path}}\) do
3: initialize weight matrix \(W_{D_F \times T_s};\)
4: \(P_{ds} = W \odot M_{\text{Pow}}; M_{\text{gradient}} = \text{Conv}(P_{ds}, \text{ker}_{\text{sobel}});\)
5: \(M_{\text{sum}}(1, 1 : T_s) = P_{ds}(1, 1 : T_s); M_{\text{index}}(1, 1 : T_s) = 1;\)
6: for \(i = 2\) to \(D_F\) do
7: for \(j = 1\) to \(T_s\) do
8: \([\text{Val}, \text{Index}] = \max[M_{\text{sum}}(i - 1, \max(j - 1, 1) : \min(j + 1, T_s));\]
9: \(M_{\text{index}}(i, j) = (\text{Index} - \min(2, j)) + M_{\text{index}}(i, j);\)
10: \(M_{\text{sum}}(i, j) = \text{Val} + M_{\text{sum}}(i, j);\)
11: end for
12: end for
13: \([\cdot; \text{IndexTail}] = \max(M_{\text{sum}}(D_F, 1 : T_s));\)
14: \([\text{Index}, \text{IndexTail}] = \text{BackTrack}(M_{\text{index}}, \text{IndexTail});\)
15: \(W = \text{UpdateWeight}(W; \text{Index});\)
16: end for
17: \(MCP_{\text{VelocityBins}} \times T_s = \text{VelocityMapping}(\text{Index});\)

Fig. 6. Visualization examples of DFS spectrograms and motion change patterns using Algorithm 1 when two users draw the rectangle.

The motion change pattern extraction algorithm is too computation-intensive to be real-time. We set \(T_s\) with a constant value 60. Therefore, the duration for each segmentation is restricted between 35ms \(\sim 70ms\) considering the total sample length, which is demonstrated as appropriate segmentation guideline range [9] for DFS spectrograms. Besides, the empirical study of motion change pattern (§3.2) also shows that the most evident difference between adjacent carving paths is less than 70ms. Note that we update \(W_{D_F \times T_s}\) for each path to avoid the overlapping of carving paths. Thus it can carve the DFS spectrogram comprehensively. We further apply the velocity mapping function to alleviate the error for the resolution of the WiFi signals and the grid estimation for body part since the real moving velocity component may be projected to the DFS bin adjacent to the true one [7]. We set the 0.16 m/s resolution within \([-1.6, 1.6]\) velocity range to achieve 20 velocity bins. \([MCP_{\text{VelocityBins}}]_{T_s \times P}\) is derived out of each sample and fed into the Dual-task Module, where the number of Velocity Bins, the segmentation number \(T_s\), the number of receivers \(R\) as well as the number PCA components \(P\).

To evaluate the effectiveness of Algorithm 1 on extracting the motion change pattern from DFS spectrograms, we illustrate two examples in Figure 6 when two users draw the rectangle, respectively. First, we can observe the consistent motion change pattern with the preliminary, which can capture the motion changes as the temporal rhythm motion during performing the gesture. Second, motion change patterns are distinguishable between two users for the same gesture, representing the personalized performing styles. We further evaluate the Algorithm 1 separately in gesture recognition and user identification with the ablation study (§5.4).

4.3 Dual-task Module

Data Adaptation and Feature Extraction: Upon receiving the motion change pattern, it’s necessary to adapt the motion change pattern before being fed into DNN. Note that WiFi signals need a specific DNN architecture and data modification for CSI measurements since it contains a lot of noises and is super-sensitive to environmental changes [37]. We reshape the motion change pattern as dimensions \([MCP_{(V_B \times R) \times P}]_{T_s}\). Thus it is similar to a digital image with the spatial resolution of \(V_B \times R\) and \(P\) as color channels. The underlying principle of the modification is that signals across receivers convey the Angle-of-Arrival (AoA) information [37], [38] while velocity bins contain body part movements [7], [32]. In contrast, PCA components are set as the color channel of images with information at different scales.

Inspired by Widar3.0 [7], we first extract spatial features from the individual motion change pattern and then profile the temporal dependencies of the whole feature sequence. To do this, a DNN of the Convolutional Neural Network (CNN) based GRU is adopted with the input tensor \([MCP_{(V_B \times R) \times P}]_{T_s}\). For each \(t\) \(th\) sampling component, the matrix \([MCP_{(V_B \times R) \times P}]_{T_s}\) is fed into a CNN, which contains 16 \(3 \times 3\) filters and two 64-unit dense layers sequentially. ReLU function and the flatten layer are further applied for non-linear feature mapping and dimension reshape, resulting in the final output for CNN characterizing the \(t\) \(th\) sampling component. Then the spatial feature series is fed into the following GRU for temporal profiling. Empirically, we adopt the 128-unit single-layer GRUs to profile the temporal relationships. Next, a dropout layer is added for avoiding over-fitting followed by the Softmax layer with cross-entropy loss for dual-task prediction. Note that the early stopping scheme is utilized to halt the training at the right time with the patience epochs 30 for value loss [41].

Splitting and Splicing Scheme: As illustrated in Figure 7, the dual-task module requires to splice individual unique features for bootstrapping performance of dual tasks collaboratively.
First, we split $MCP \times T_x$ evenly along the time axis. Thus we can avoid the issue of vanishing gradients with too long time series ($T_x \sim 60$). On the other hand, splitting the input generates dual inputs for the module and the correlation between them can enhance the performance of the user identification. Then the gradient block layer is tailored for feature splicing inspired by Multi-Task Learning (MTL). However, different from traditional MTL, the dual tasks here are expected to only slice respective features for collaborative learning while avoiding impacts of loss from each other during the back-propagation process. Take the gesture recognition task as an example, we utilize the output of its own CNN based GRU module as the superior feature while the feature extracted by the CNN_GRU module for user identification is taken as an inferior one. Then both are spliced together and fed into the final layer to predict the gesture. The key point is we do not back-propagate the gesture prediction loss to the CNN based GRU for user identification. In other words, we keep the CNN based GRU of user identification from being influenced by gesture predictions. Thus we introduce the Gradient Reversal Layer (GRL) [25] and adapt it into the gradient block layer by setting the splicing factor with zero, which can be used for the generative adversarial network with a positive factor while MTL with a negative one as normal back-propagation.

The underlying rationale comes from both theoretical analysis and experimental validation. First, the dual tasks are defined as a sub-type (user-defined gesture vs gesture-defined user) task instead of a main-type task (gesture vs user), required by the preliminary (§3.2). That means the features for the main-type task can assist the sub-type task as a superior indicator while it should not be influenced by the loss of the inferior feature for the sub-type tasks. On the other hand, it has been validated that the user-specific features have noise to gesture recognition as domain information [7]. Previous work [25], [26], [27] applies GRL with a negative factor to eliminate domain noise and extract the cross-domain feature while sacrificing the performance of the predictor. Since the motion change pattern already contains cross-domain knowledge, we no longer need to apply GRL to eliminate noise at the expense of predictor performance. Experiments demonstrate the efficiency of the zero splicing factor for gradient blocking in (§5.6).

## 5 IMPLEMENTATION AND EVALUATION

We implement WiHF and evaluate its performance through extensive experiments. The detailed settings are illustrated as follows:

**Dataset:** We evaluate WiHF on the public dataset from Widar3.0 [7], which contains 9 gestures of 16 users collected from 5 locations and 5 orientations in 3 environments. And we adopt the dataset including 9 users and randomly select 6 users for the overall performance evaluation since the selected users share the 6 same gestures across 75 same domains (5 positions × 5 orientations × 3 environments) while the distribution is non-uniform for the other users, making it not equivalent with the adopted ones for the cross-domain testing. Besides, it’s sufficient for WiHF to work with 6 users given the smart home scenario. To make the comparison study convinced while fully extracting the motion change pattern, we make three feature datasets using the Pattern-Carving Module, shown in Table 1. First, HuFuM (HuFu Mini) feature dataset is collected using the original dataset for the comparison with Widar3.0. To fully extract the characteristics representing the personalized motion change pattern, we concatenate HuFuM across instances and gestures respectively, delivering the feature dataset HuFuE (HuFu Extend) with the doubled average sample length of 3.238s and HuFu feature dataset. The rationale is the motion change pattern is impacted by the duration and complexity of movements. For example, drawing the rectangle is more sophisticated than sweeping since the former has a much larger influential coverage area and more explicit rhythmic velocity fluctuation.

**Metrics:** To characterize the WiHF’s performance, prediction ACCuracy (ACC) and latency are the two main metrics for both of gesture recognition and user identification. And the former is the measure of the confidence of prediction for each instance. Besides, False Authorized Rate (FAR) and False Unauthorized Rate (FUR) are adopted for user identification as [23]. And FAR measures the likelihood that WiHF incorrectly accepts a gesture attempt by an unauthorized user while FUR evaluates the likelihood that WiHF incorrectly rejects the gesture performed by an authorized user. All the metrics can be calculated with the confusion matrix as:

$$ACC = \frac{TruePositive}{TruePositive + FalsePositive}$$  \hspace{1cm} (7)

$$FAR = \frac{FalsePositive}{FalsePositive + TrueNegative}$$  \hspace{1cm} (8)

$$FUR = \frac{FalseNegative}{FalseNegative + TruePositive}$$  \hspace{1cm} (9)

### 5.1 Overall Accuracy

We first evaluate the performance of distinguishing gestures and users on the HuFu feature dataset. Taking all domain
for the cross-domain gesture recognition, the average accuracy of WiHF with HuFuM feature dataset slightly increases across locations and environments but drops for orientation. Specifically, WiHF achieves the average accuracy of 90.32%, 79.14%, and 89.67% for gesture across locations, orientations, and environments, comparable with the state-of-the-art [7] with 88.78%, 82.97%, and 88.95% respectively. The performance of the worst instance is only 68.19% with edge orientations (orientation#1 and #5) [7] as the target orientation. Such an accuracy decrease can also be observed for Widar3.0, which is 73.26%. And the reason behind this is that gestures might be shadowed by human body parts in edge orientations, resulting in unrecoverable signal loss. WiHF drops more since the shadowing effect can induce detailed information loss for the fine-grained motion change pattern. To evaluate the impact of motion details, we increase the gesture duration by repeating the gesture in HuFuE dataset. The performance of HuFuE demonstrates the longer gesture instances can improve the overall cross-domain recognition and compensate for the pattern missing due to body shadow. For example, we find the gesture of drawing the rectangle demonstrates more resilient to the cross-domain scenarios. Thus we can provide additional resilience to the shadowing effect by optimizing the gesture design.

Then we evaluate the performance of cross-domain dual tasks on HuFu feature dataset. Table 2 shows that the cross-domain performance for gesture recognition remains above 85% for orientation 2,3,4 and all the locations while declines by over 10% at edge orientation 1,5, which demonstrates consistent and better accuracy with Widar3.0, HuFuM and HuFuE. And it verifies the efficiency of designing new gestures with more complexity and duration. Nevertheless, the decrease also results from the body part shadow which blocks effective wireless signal sources for the motion change pattern.

For user identification, Table 2 shows an apparent decrease compared with gesture recognition and even descends to the worst 57.17% when testing at orientation 1. Generally, WiHF achieves 75.31% and 69.52% across locations and orientations, respectively. The reason for the performance decrease is that the collaborative dual-task module identifies users with finer-grained information than gestures. It is weaker due to the cross-domain noise, especially considering the short and simple gestures. We can alleviate the decrease across domains by extracting more PCA components and carving paths. On the other hand, the performance loss can be minimized by designing more

### Table 2

<table>
<thead>
<tr>
<th>Target Label</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gesture</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Location</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Orientation</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>User</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Location</td>
<td>73.83%</td>
<td>66.69%</td>
<td>73.44%</td>
<td>76.22%</td>
<td>86.33%</td>
</tr>
<tr>
<td>Orientation</td>
<td>57.17%</td>
<td>77.33%</td>
<td>73.17%</td>
<td>77.39%</td>
<td>59.00%</td>
</tr>
</tbody>
</table>

*The target label denotes the data for testing while others for training.*

**Cross-Domain Evaluation**

As for cross-domain characteristics of motion change pattern, we first compare it with Widar3.0 [7] using BVP and HuFuM feature dataset. Then HuFuE feature dataset is adopted for exploring the impact of the gesture duration on the motion change pattern. And we calculate the average accuracy across domains by using one out of all domain instances for testing while others for training. Note that the other domain components keep unchanged for evaluation on the specific domain component. By doing this, we can evaluate its zero-effort cross-domain ability as Widar3.0 does. Figure 9 plots the accuracy distribution for each domain component.

We can find WiHF achieves the median accuracy of 93.92% for in-domain recognition using HuFuM while it reaches 90.58% for Widar3.0 [7]. Compared with Widar3.0

**Fig. 8.** Overall performance for gesture recognition and user identification on HuFu feature dataset.

**Fig. 9.** Comparison with Widar3.0 for gesture recognition across domains.
sophisticated gestures with more complexity and duration, which can minimize the body shadow effect. Besides, more receivers can be added to increase the coverage so that it can reduce the possibility of signals missing.

To conclude, WiHF achieves average performance of 92.07% and 82.38% for gesture across locations and orientations. It shows 75.31% for all locations and 75.96% for centering orientations in user identification simultaneously.

5.3 Latency

In practice, the time consumption of WiHF mainly comes from feature extraction for Pattern-carving Module and recognition as well as identification for Dual-task Module. Table 3 shows the time consumption distribution. Note that results are all computed assuming it runs in parallel across receivers since the feature $[MCP_{V_B \times R \times P}]$, can be derived from various receivers individually and concatenated together. We can find WiHF spends more time on signal processing due to STFT operation for more PCA components. Widar3.0 demands on 70.61s for feature extraction while WiHF takes 69.93s less than Widar3.0. The total time consumption for the HuFu feature extraction is 2.488s with the average gesture duration 3.669s. We believe it is sufficient for most user identified gesture recognition application scenarios. The remarkable improvement on time consumption lies in two folds. First, Widar3.0 derives the BVP in body part coordinate system using the $l_0$ optimization problem with respect to Earth Mover’s Distance (EMD) metric. And it estimates the high dimensional BVP as the square of velocity bins resolution ($20^2$ variables) and becomes computation-intensive although Widar3.0 controls the estimated variables using sparsity coefficients. Second, WiHF designs the optimization problems individually for each receiver concerning the low dimensional estimated carving path as the resolution of velocity bins (20 variables). Moreover, it leverages the efficient method based on the seam carving algorithm instead of the constrained nonlinear multi-variable function as Widar3.0.

In a nutshell, WiHF demonstrates comparable cross-domain characteristics using the motion change pattern with Widar3.0 for gesture recognition. However, the processing time is reduced by 30×.

5.4 Ablation Study

To evaluate the motion change pattern extracted by Algorithm 1 separately, we remove the dual-task network and adopt the simple classifier of Widar3.0 [7], which takes as input the motion change pattern. Table 4 shows that it can reach 90.67% and 91.16% for gesture recognition and user identification, respectively. And both are slightly dropped without the dual-task network when the dual tasks are divided into two separate tasks. It demonstrates the effectiveness of the collaborative dual-task network on fully enhancing the spatial and temporal features for motion change pattern.

5.5 Comparative Study

We compare WiHF with WiID [9] to demonstrate its superiority. On one hand, WiID utilizes the motion contour of body parts as the power based feature for user identification while WiHF leverages the motion change pattern to represent both the unique gesture characteristic and the personal-ized user performing styles (§3.2). On the other hand, WiHF bootsraps the dual-task learning collaboratively while WiID can only identify users using the known gesture recognized by existing gesture recognition systems.

We implement WiID to extract the motion contour using MatLab and compare it with WiHF using the HuFu feature dataset. Note that, WiID requires 50 instances for each specific domain when training with the simple classifier [9] while only 10 instances can be provided with the HuFu feature dataset.

As shown in Table 4, WiHF achieves 96.74% accuracy for user identification on the HuFu feature dataset, much higher than the one for WiID with only 68.95%. Besides, we also evaluate the performance of user identification across domains. We can find the cross-location accuracy of WiHF is 30% higher than WiID, shown in Figure 10. And WiHF achieves 20% improvement for the cross-orientation testing for the three centering orientations. The reason behind the poor performance of WiID has two folds. First, the power based feature cannot be utilized for user identification a robust feature, especially in the cross-domain scenarios. And the evaluations are consistent with our preliminary
and observations (§3.2). Second, WiID is data-hungry and requires 50 instances for training [9]. Thus it suffers using only 10 instances for each specific domain in the HuFu feature dataset. Note that we over-estimate the performance of WiID since we assume the gesture is truly known. However, the state-of-the-art methods only achieve 92% accuracy of the gesture recognition [7], which can affect the performance of user identification of WiID.

5.6 Parameter Study

5.6.1 Impact of Numbers of PCA Components:
As shown in Table 5, the accuracy gradually drops for dual tasks of the in-domain, cross-location, and cross-orientation scenarios as the numbers of PCA components reduce from 3 to 1. The rationale is more PCA components can provide more fine-grained motion change patterns at different scales for the Pattern-carving Module, illustrated in Figure 5. And it can reduce the information loss induced by the shadowing effect and supplement detailed motion of different body parts.

<table>
<thead>
<tr>
<th># PCA components</th>
<th>1</th>
<th>2</th>
<th>3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gesture Recognition</td>
<td>In-domain 90.63% 94.33% 97.65%</td>
<td>Location 79.00% 88.33% 92.07%</td>
<td>Orientation 69.50% 76.47% 82.38%</td>
</tr>
<tr>
<td>User Identification</td>
<td>In-domain 89.33% 94.07% 96.74%</td>
<td>Location 64.43% 68.53% 75.31%</td>
<td>Orientation 59.93% 65.15% 69.52%</td>
</tr>
</tbody>
</table>

Table 5: Impact of the number of PCA components

5.6.2 Impact of Gesture and User Type Numbers:
To study the impact of the number of gestures and subjects, various sets of the HuFu feature dataset are used for different total type numbers (the default type number is 6 for both tasks). Table 6 shows that the in-domain accuracy remains above 92% though the number of gestures or users increases to 9. As far as the accuracy of WiHF is concerned, it’s not significantly affected by the number of gestures and users. Moreover, we can find that the respective types of dual tasks cannot influence each other. That means performance for gesture recognition stays consistent as the number of users varies even across domains, resulting from the effect of the gradient block layer.

5.6.3 Impact of Slicing Schema:
We analyze the effectiveness of feature slicing schema in predicting gestures and users collaboratively. Different factors for the gradient layer are adopted to validate the relationship between the features of gestures and users. Figure 11 shows that WiHF achieves the best performance in both in-domain and cross-domain scenarios with the zero splicing factor. Besides, the in-domain performance declines with a positive factor while the existence of GRL cannot improve its cross-domain performance as expected. The reason may be that noisy signal from the domain features suppresses the target feature with a fixed factor [25]. The performance decreases with negative factor as MTL shows that the loss function of the inferior feature for the sub-type tasks can induce noises for the superior feature of the main-type task. Thus it should be blocked using the gradient block layer with the zero splicing factor.

6 LIMITATION & DISCUSSION

Although our proposed WiHF system can recognize gestures and identify users simultaneously across domains, it still has some limitations.

Background interference: First, WiFi based sensing systems generally suffer from common issues, such as the background interference induced by other moving objects or APs. And it can introduce the uninterested amplitude attenuation and phase shift to CSI measurements, affecting the distinctness and stability of the designed feature. To compare the impact of the interference statistically, we adopt data from three environments with various interference (density of the APs and background objects) [7]. As the interference gets stronger for three environments, the gesture recognition achieves 97.65%, 96.54%, and 96.19% while the user identification for 96.74%, 95.37%, and 93.64%, respectively. Evaluations show that a clearer environment can introduce less interference from the background objects and other APs, achieving a better performance for dual tasks. WiHF however renders the acceptable performance with the limited degradation due to the adopted algorithms for noise reduction.

To alleviate the interference, PCA [4], [5], [16], [39] and the band-pass filter [7], [39] are employed to extract princi-
ple components of raw measurements and filtering out uninterested frequency components. Besides, we optimize the deployment for CSI acquisition. For example, transceivers can be set to work on channel 165 at 5.825 GHz where there are few interfering radios [7], [42]. WiHF utilize the CSI collected from a relatively clean channel and employs the PCA and filter for interference cancellation.

Behavior dynamics: Moreover, humans cannot always act regularly, rendering the unacceptable shift for the DFS spectrogram and the motion change pattern (See Section 2). Thus our system can suffer due to variances of the extracted gesture characteristics and the personalized performing style in some cases. For example, the gesture can be misclassified or the authorized user can be rejected sometimes with the irregular behavior. WiHF can alleviate this issue with online adaption. Specifically, given the application scenarios such as the smart home and VR gaming, users are usually in a relatively relaxing status and seldom change the personalized performing style abruptly, which has been evaluated through long-term experiments [9], [29], [31]. To provide additional resilience to the variance of the performing style, we can also adapt the deep learning model to capture the variance pattern of users by re-training the dual-task network using users’ gestures over some time.

Deployment cost: Besides, DNN has some common issues on deployment cost, such as cumbersome data collection and model retraining, especially when we employ the model in a new environment. Existing DNN based works [27], [43] adopt machine learning techniques, such as transfer learning and adversarial learning to improve the cross-domain generalization ability for gesture recognition. Specifically, El [27] designs an adversarial architecture with specific loss functions to exploit characteristics of unlabeled data in the target domain while CrossSense [43] employs an ANN-based roaming model to translate features from the source domain to target domain. Both however require extra training efforts in either data collection or model re-training each time a new target domain is added into the recognition model [7].

Given the domain-independent feature of our system, evaluations in section 5.2 demonstrate that WiHF can reduce the deployment cost in cross-domain scenarios. Specifically, we train the model with data from four domains and test it in a new one, say different locations and orientations. And we also train the model in one environment and test it in another environment directly. WiHF achieves the average accuracy of 90.32%, 79.14%, and 89.67% for gesture across locations, orientations, and environments, comparable with the state-of-the-art [7] with 88.78%, 82.97%, and 88.95% respectively. As for the user identification, one user is only required to do the gesture in some locations, orientations, and environments. And it can achieve acceptable performance when adopted in a new location, orientation, or environment. From this view, it can reduce the deployment cost significantly compared with similar user identification works, such as WiFiU [16] and WiID [9] in Section 5.5.

Shadowing effect: Further, the cross-orientation evaluation suffers more than others. The rationale is that the shadowing effect of body parts can induce motion information loss by blocking reflected WiFi signals. One possible solution is to utilize more antennas and optimize the deployment of receivers, delivering more vantage views for the target person. Besides, some specific gestures can also be designed for WiHF to fully represent the personalized performing styles and reduce the information loss from shadowing. Evaluations in Section 5.2 demonstrate that gestures with longer duration can improve the recognition accuracy for gesture recognition, shown in Figure 9. Note that the processing overhead cannot increase significantly since we can extract the motion change pattern from antennas in parallel. Besides, samples can be segmented for feature extraction with complex gestures, delivering a real-time system.

Future work: Dedicated algorithms and experimental setup can be designed to further improve the robustness and accuracy of WiHF. For example, we can employ the background subtraction [34], [44], [45] to subtract the initial background interference. To provide additional resilience to the noise, we can incorporate multi-dimensional features [34], [38] to capture more individual information of users, such as Doppler shift induced by the moving arm [4], [5] and BVP [7]. Besides, we can also adapt our system over some time to capture the performing variances of humans. Furthermore, some specific gestures can be designed to fully represent the personalized performing style. And we leave them as our future work.

7 CONCLUSIONS

To unleash the potential of the WiFi based gesture recognition, we propose WiHF to enable gesture recognition and user identification simultaneously in real time. WiHF proposes to derive a domain-independent motion change pattern of arm gestures from WiFi signals, rendering the unique gesture characteristics and the personalized user performing styles. To be real-time, we carve the dominant motion change patterns and develop an efficient method based on the seam carving algorithm. Taking the carving path for motion change pattern as input, a collaborative dual-task DNN with the splitting and splicing schemes is adopted. We implement WiHF and evaluate its performance on a public dataset. Experimental results show that WiHF achieves 97.65% and 96.74% for in-domain gesture recognition and user identification accuracy, respectively. And the cross-domain gesture recognition performance is comparable with the state-of-the-art methods, but the processing time is reduced by 30x.

REFERENCES


Chennig Li (S’20) received the B.Eng. degree from Tsinghua University, Beijing, China, in 2018. He is currently working toward the Ph.D. degree in the Department of Computer Science and Engineering, Michigan State University, USA. His primary research interests include wireless sensing, indoor localization, mobile computing, and artificial intelligence of things.

Manni Liu received the B.Eng. degree in the Department of Mathematics and Applied Mathematics at Sun Yat-sen University. She is currently pursuing the Ph.D. degree in the Department of Computer Science and Engineering, Michigan State University, USA. Her research interests include internet of things, mobile computing, and indoor localization.
Zhichao Cao (M’12) received the BE degree in the Department of Computer Science and Technology at Tsinghua University, and his Ph.D. degree in the Department of Computer Science and Engineering at Hong Kong University of Science and Technology. He worked as an assistant professor in the School of Software, Tsinghua University. He is currently with Department of Computer Science and Engineering, Michigan State University. His research interests include IoT system, mobile and edge computing. He is a member of the IEEE.