Consumer Wi-Fi device based action quality recognition: An illustrative example of seated dumbbell press action

Yuxi Qin | Su Pan | Weiwei Zhou | Duowei Pan | Zibo Li

1 Key Lab of Broadband Wireless Communication and Sensor Network Technology, Ministry of Education, Nanjing University of Posts and Telecommunications, Nanjing, China
2 College of Letters & Science, University of California, Berkeley, California, USA

Abstract
A system called WiSDP, which is based on Wi-Fi signals, to detect whether a Seated Dumbbell Press action is standard by using inexpensive consumer Wi-Fi devices is proposed. Compared with the scheme based on high speed cameras and wearable sensors, Wi-Fi devices are insensitive to light and colour, do not need wear any device, and decrease the risk of disclosing privacy. WiSDP senses environment changes through the Channel State Information which is fine-grained physical layer information comparing to frequently used Received Signal Strength Indicator. Compared to the action recognition, action quality recognition depends on slight differences between a non-standard action and standard actions, which makes it challenging. The authors propose an improved sliding window algorithm calculating action energy to extract Seated Dumbbell Press actions from the Channel State Information streams, estimate action quality by choosing an appropriate classifier and use Principal Component Analysis and Butterworth low-pass filter to remove noise. The authors conduct experiments in two different scenarios and the average true positive rate of WiSDP are 94.66% and 95.11%, respectively.

1 INTRODUCTION
Action quality recognition is very important in many scenes, such as sports competition, rehabilitation medicine, and daily fitness. The objective evaluation of action quality recognition based on machine learning can reduce the subjective factors of judgment and make people get professional guidance easily. This paper takes daily fitness as an example to study action quality identification. Fitness is closely related to each of us, but not everyone has the opportunity to get the guidance of a professional coach [1]. The right action to get fit can help us shaping and stay healthy, while the wrong action to get fit can make it less efficient and cause damage to our bodies in the long run. Seated Dumbbell Press (SDP) action, one of the most common body building actions, is used as an example to study action quality recognition.

Action quality recognition is a kind of action recognition. Action recognition is to classify actions according to the feature differences between different actions. The feature differences between different actions are large and there are less feature differences between the same actions. Action quality recognition is to evaluate whether the same action meets the requirements including accuracy, coordination, elegance and economy. This paper mainly studies the accuracy and coordination of action quality. Accuracy refers to whether the technical action conforms to the specification requirements. For example, a user should raise dumbbells with his arms directly above his body. Coordination refers to whether the technical movement is coherent, the rhythm is reasonable, and all parts of the body coordinate and cooperate in the process of completing the movement. For example, a user cannot shake his body when lifting dumbbells. From the perspective of data acquisition, action recognition can be divided into three categories: sensor-based, vision-based and wireless-based. Sensor-based scheme use wearable sensing devices to capture body movements and postures [2, 3], but it requires users to wear sensors on body, which makes users feel uncomfortable. Vision-based scheme use high-speed video recorder to record motion detail information through frames [4, 5]. However, the recognition accuracy is easily affected by the light intensity and the contrast between the foreground and background colours. Most of all, the video contains a lot of sensitive information that is not relevant to the...
target to be detected, which is likely to result in the disclosure of users’ privacy. Wireless-based scheme utilise Wi-Fi signals affected by human actions to distinguish between different actions [6, 7], which is a device-free recognition technology and can protect users’ privacy.

Action quality recognition needs more fine-grained features [8, 9]. Most researchers use a wearable sensor-based approach except WiQ [10]. To estimate the driving behaviour quality, Lv et al. present a system called WiQ based on Wi-Fi signals. In the system, two Universal Software Radio Peripheral (USRP) devices are used for sending and receiving Wi-Fi signals and Received Signal Strength Indicator (RSSI) is used for sensing. WiQ adopts an activity-based fusion policy which use at least three actions to estimate the quality of a driving activity, which means WiQ cannot effectively recognise quality for a single driving action. Beyond that, compared to a inexpensive Wireless Local Area Network (WLAN) card, USRP is a dedicated wireless device and can obtain more precise RSSI data than commercial WLAN card with a shortcoming: expensive. Radio Frequency Identification Devices (RFID) is used to detect drowsy driving actions in [11]. In order to get the movement information of the driver’s head, the driver needs to wear a hat with RFID tag. RFID-based way requires the driver to wear a special device, and can only detect simple movements of head, not suitable for detecting more complex body movements. Fine-grained Channel State Information (CSI) obtained from commercial WLAN card can get more sensitive information. CSI can reflect the channel gain of each subcarrier of OFDM signals used in Wi-Fi system while RSSI detected by inexpensive Wi-Fi devices represents the received power of all Orthogonal Frequency Division Multiplexing (OFDM) subcarriers. Based on CSI, Chen et al. use Bi-directional Long Short-Term Memory (BLSTM) to recognise six common daily activities of ‘Lie down, Fall, Walk, Run, Sit down, Stand up’ [12]. There are big differences between these actions, and we can easily distinguish different actions in various ways.

We design and implement a wireless-based recognition system called WiSDP which use inexpensive consumer WLAN card to obtain CSI and use machine learning algorithm to recognise action quality. The system allows users to be aware of whether there are any errors in their SDP actions without the supervision of a professional fitness trainer and wearing any special sensors. With such a device-free action quality recognition system, users can easily do exercise and get guidance at home. To the best of my knowledge, WiSDP is the first system using fine-grained CSI to estimate quality of body building actions. In WiSDP, we have solved several challenges.

First of all, existing action segmentation algorithms [10, 13, 14] are not applicable to our system in which the interval between the actions is short and the start time of the action is uncertain. For example, Ren et al. extracted the target action from CSI waveform with a specific frequency range, which will not work when the frequency of CSI waveform caused by the action changes significantly [15]. Therefore, we propose the concept of ‘action energy’ which is the average value of the difference between adjacent packets in a data stream segment. By calculating the action energy scheme, we do not need to estimate the frequency range caused by actions because the changes of CSI waveforms frequency caused by the speed of actions have little influence on the action energy. We propose a sliding window algorithm with variable step size to calculate the action energy. We use a large step size to search for the start and end ranges of the action, and a small step size to accurately search the range determined by the large step size. In this way, we can remove the data when there is no action, transient actions and small actions, which do not belong to SDP.

Second, to solve the problem of high noise caused by cheap WiFi devices, we propose to use a method based on Principal Component Analysis (PCA) to reduce noise. We calculate the eigenvalue and the eigenmatrix of the CSI matrix and remove the eigenmatrix corresponding to the maximum eigenvalue, because random noise is proved usually distributed in the first dimension which correspond to maximum eigenvalue [16]. The denoised data can be obtained by the new eigenmatrix.

Third, to select the appropriate classifier for the action quality recognition, we test various machine learning algorithms and find that the Support vector Machine (SVM) has a good performance in the action quality recognition. Although the classifiers based on neural network can achieve a better recognition accuracy, it requires much more training samples and long training time than the SVM. Therefore, based on SVM, we can easily migrate WiSDP to be suitable for other action quality recognition scenarios.

In summary, the main contributions of this paper can be summarised as follows:

- We propose and implement WiSDP, a system based on cheap consumer Wi-Fi devices, to recognise action quality and take seated dumbbell press actions as an example.
- We propose to use action energy to accurately reflect the action change, which has good robustness. The improved sliding window algorithm which can dynamically adjust the search step size is proposed for extracting actions accurately.
- To overcome the shortcoming caused by inexpensive Wi-Fi devices, we propose to use PCA to remove the first-dimensional feature where noise mainly distribute.
- After trying many popular machine learning algorithms, we find a classification algorithm with high recognition accuracy, short training time and short running time. Thus, WiSDP can be easily adapted to recognise action quality in different scenarios.

The rest of this paper is organised as follows. We first introduce the related work in Section 2. Then in Section 3, we present the principle and structure of WiSDP system, followed by describing implementation and evaluation in Section 4. Finally, we give our discussions and conclusions of WiSDP in Sections 5 and 6, respectively.

2 | RELATED WORK

With the development of artificial intelligence technology, coarse-grained action recognition has achieved good results in
many fields, but the evaluation of action quality still needs further study. In this section, we introduce the relevant technologies of action quality recognition from three schemes: sensor-based, vision-based and WiFi-based.

### 2.1 Sensor-based action quality recognition

Sensor-based scheme have been widely used in action quality recognition [17, 18]. Increasing the number of sensors can effectively improve the recognition accuracy, which makes it have good robustness. For recognising lifting action quality, Velloso et al. used four 9 degrees of freedom razor inertial measurement units to record data [19]. Three sensors were worn by the user and one was attached to the dumbbell. Giorgis et al. used signals of many wearable inertial sensors to detect relevant peaks of acceleration of limbs (arms and legs) [20]. These data were used to classify the level of skills of athletes performing karate kata. The results judged by the system were almost the same as those judged by the experts. However, sensors must be tightly attached to the athletes, which is likely to affect performance of athletes.

### 2.2 Vision-based action quality recognition

Vision-based scheme records the action information by video, which turns the original three-dimensional information into two-dimensional information [21, 22]. To get richer visual information, Elkholy et al. used depth cameras to detect abnormality in common daily action performance [23]. They used extracted medically features from three-dimensional skeletal data provided by a depth camera to train a probabilistic normalcy model and build a linear regression model to assess action quality. Depth cameras are susceptible to angles of view and environment, so the combination of vision-based scheme and sensor-based scheme is often used to evaluate the action quality. Niewiadomski et al. using video and MoCap tested two karate actions to evaluate action quality, and the results showed that the action quality evaluated by the scheme was highly similar to the action quality evaluated by human experts [24].

### 2.3 WiFi-based action quality recognition

For WiFi-based scheme, prior work mainly focuses on how to distinguish different actions, and few researchers pay attention to how to judge whether an action is standard or not. Based on Wi-Fi signal, there are two main sensing schemes: RSSI-based and CSI-based. RSSI measures the superposition effect of multi-path signals propagation and cannot distinguish multiple signals propagation paths one by one. Therefore, RSSI is often used for coarse-grained indoor positioning and rarely used for fine action recognition [25, 26]. For example, Haseeb et al. custom seven hand gestures to operate the media player using RSSI signals and the results show the system can achieve an average accuracy of 87.5% [27]. It is difficult to obtain high recognition accuracy by RSSI-based scheme even if the movement is very simple. For CSI-based scheme, there is no system for detecting action quality, but researchers have proved that CSI-based scheme can detect even small body actions [28]. Ali et al. proposed a system called WiKey to recognise the keyboard typing [29]. The detection accuracy of WiKey was more than 97.5%. However, the high accuracy of WiKey is because the receiver is only 30 cm away from the keyboard and use a high-performance enterprise router as the sending device which is too expensive for ordinary consumers. Zhang et al. presented a system called WiriFi, in which users wrote 26 letters in a size of 25 cm × 25 cm air area [30]. The hidden Markov model was established in the system by dividing each letter according to the stroke and the average recognition accuracy could reach 86.75% in the experimental environment. In practice, it is difficult to separate each stroke of a continuous letter, so the system must require the writer to have an unnatural pause.

### 3 WiSDP SYSTEM MODEL

WiSDP is a wireless sensor system working on cheap consumer Wi-Fi devices, obtaining fine-grained CSI data from the physical layer and using feature-based machine learning technologies to recognise human action quality in indoor scenarios. In the overall design, WiSDP is divided into action sensor module, action detection module and action quality recognition module. Figure 1(a) is the sensor module in the system and Figure 1(b) is the workflow of the whole system.

In sensor module, WiFi-based sensor is mainly composed of two commercial devices: an Access Point (AP) (such as a router) and a Detection Point (DP) (such as a laptop). To let DP sense the change of environment, we constantly send data to DP to maintain intensive communication. If there is any action between AP and DP, it will change the propagation path of the signals, resulting in different CSI which can be sensed by DP.

In detection module, we first pre-process the obtained original CSI signals to remove outliers and high-frequency noises. Then we detect the signals that contains the target actions and separate it from the original signals.

In action quality recognition module, we first select some representative characteristics from samples to describe the action quality. Then we train the SVM model using the characteristics. Finally, trained SVM model can be used to recognise quality of samples.

#### 3.1 Data collection

#### 3.1.1 SDP actions

Seated dumbbell press action, one of the most popular body building actions, is chosen to conduct our experiments. In the experiments, we use a stool with no backrest, which will cause more wrong actions by the users and will bring more challenges to the action quality recognition. We list nine SDP actions, including one standard action and eight actions with different
errors, as shown in Table 1. The SDP1 is standard action, and the action details is as follows:

1. User sit on a stool with no backrest and stabilise body with legs, hold the bell in both hands on both sides of head and keep the angle between upper and lower arms about 90°. Keep palms facing the front.
2. While inhaling deeply, push the dumbbells up from the sides of your body to the top of your head. Two dumbbells do not knock against one another.
3. Exhale as the dumbbells slowly fall and arms fall to the starting position.

The differences of action recognition and action quality recognition can be illustrated by Table 1. Action recognition is to detect different actions, and the CSI waveforms caused by different actions are usually quite different. Therefore, different actions can be easily recognised by analysing the CSI waveforms. Action quality recognition is to detect the same actions, and the same actions are usually only slightly different, so it is more challenging to recognise action quality. For example, compared to the SDP1 (standard action) in Table 1, the SDP2 is identical to the SDP1 except that the user lowers his head slightly when pushing up. Compared to the SDP actions, slight lowering of the head is a small action. To recognise the action quality, we have to find out the characteristics of these small actions on the CSI waveforms.

### 3.1.2 CSI measurement

In PHY layer, to improve the transmission efficiency of data, OFDM converts high-speed data signals into parallel low-speed sub-data streams by dividing channels into several orthogonal subcarriers and modulating them into each subcarriers for transmission. OFDM has been applied in IEEE 802.11 a/g/n and is used in proposed WiSDP.

In the wireless channel, the received signals matrix $Y$ can be expressed as Equation (1):

$$Y = HX + N,$$

where $X$ is the transmit signals matrix, $H$ is the channel gain matrix and $N$ is the Gaussian white noise, so the approximation of $H$ can be expressed as Equation (2):

$$h = \frac{Y}{X}.$$

In Equation (2), $h$ is CSI actually obtained, which contain the channel gain of the signals in each subcarrier. Because subcarriers have different frequencies, each subcarriers have different channel gains. Compared with CSI, RSSI measured by Intel 5300 Network Interface Card (NIC) cannot distinguish multiple signals one by one, which makes the information reflected by RSSI less than CSI [31]. Therefore, CSI can be more sensitive to the changes of wireless channels and can perceive subtle differences between actions.

The data structure of a CSI packet is $q \times s$, where $q$ is the number of sending and receiving antenna pairs and $s$ is the number of subcarriers in a sending and receiving antenna pair. In a CSI packet, we use $h^{(j)}_{i}$ to express channel gain of $i$th subcarrier in $j$th sending and receiving antenna pair. $h_{i}$ is channel gain matrix of $i$th subcarrier in all sending and receiving antenna pairs. $h^{(j)}_{i}$ is channel gain matrix of all subcarriers in $j$th sending and receiving antenna pairs. Channel gain $h$ can be denoted as...
The time interval of received CSI packets follows:

\[ h = \begin{bmatrix} h_1 \\ h_2 \\ \vdots \\ h_s \end{bmatrix}^T = \begin{bmatrix} h^{(1)}_1 & h^{(1)}_2 & \cdots & h^{(1)}_s \\ h^{(2)}_1 & h^{(2)}_2 & \cdots & h^{(2)}_s \\ \vdots & \vdots & \ddots & \vdots \\ h^{(q)}_1 & h^{(q)}_2 & \cdots & h^{(q)}_s \end{bmatrix} \cdot \tag{3} \]

We use \( h[k] \) as channel gain matrix of \( k \)th packet and we have

\[ h[1 : n] = [h[1] \ h[2] \ \cdots \ h[n]] \cdot \tag{4} \]

In a sending and receiving antenna pair, we can only obtain one value if sensing RSSI and we can obtain \( s \) values if sensing CSI.

### 3.2 Data pre-processing

Many raw CSI are affected by interference and environmental noise, so we need to remove interference and noise to obtain data related to human activities.

#### 3.2.1 Channel gain amplitude outlier removal

We remove outliers from the CSI streams by Hampel filter [28, 32]. Outliers are the points falling out of the closed interval \([\mu - \lambda \sigma, \mu + \lambda \sigma]\), where \( \mu \) is the median, \( \sigma \) is median absolute deviation and \( \lambda \) is an adjustable coefficient relying on the data sequence. Finding outliers and replacing them with the average of the data are helpful to get more reliable CSI data.

#### 3.2.2 Time outlier removal

Due to the instability of the receiving system constructed by inexpensive devices, the problem of missing data occasionally occurs in the received packets. To ensure the uniform data format, we set the missing value as the mean value of the two adjacent packets. Not only that, the inexpensive devices we used cannot precisely control the time interval of received packets. As shown in Figure 2, the receiving interval can sometimes be as much as five times the sending interval. We used the linear interpolation algorithm [33] to recalculate CSI strictly according to the time interval.

#### 3.2.3 Noise removal

CSI can be obtained in each subcarrier, which can provide more information than RSSI. In order to demonstrate the denoising effect, we chose to display the CSI of a subcarrier. Figure 3(a) shows the original CSI waveform of an SDP action. Compared with the expensive USRP used in [10, 13], the CSI data collected by the cheap Wi-Fi devices contain more noise which has random intensity in time domain and are uncorrelated with human actions [34, 35]. Therefore, we use PCA-based scheme to remove random noise in CSI streams. First, we use Z-Score standard algorithm to normalise the matrix to mean 0 and variance 1. Second, we calculate the covariance matrix and the eigenvalue, and then order the eigenvalues from large to small. Then we remove the first eigenvalue and calculate the
eigenmatrix corresponding to the eigenvalues, because random noise is proved usually distributed in the first dimension [16]. Finally, we obtain the denoised CSI matrix by inverse transformation. Figure 3(b) displays the CSI waveform denoised by PCA. We can see that although there is still a lot of high-frequency noise in the data, many noises have been removed and the characteristics of the original data have been retained.

Our experimental results show that the frequency of CSI waveforms changes caused by arm and body movements is less than 20 Hz. Therefore, we use Butterworth low-pass filter to remove high-frequency noise and the filter frequency response can be expressed as Equation (5)

\[ |F(j\omega)|^2 = \frac{1}{1 + (\frac{\omega}{\omega_c})^{2l}}, \quad (5) \]

where \( l \) is the order of filter and \( \omega_c \) is cut-off frequency. In Figure 3(c), we can see that most of the noise has been removed and only the effective information is left.

3.3 SDP action detection

The CSI of adjacent subcarriers are very similar [29]. Thus we divide \( s \) subcarriers into \( g \) groups. We merge the adjacent \( \xi \) subcarriers together to get average channel gain of \( g \) groups, denoted as \( \bar{h} \) which is expressed in Equation (6), where \( \lambda_i \) is the weight of corresponding combined subcarriers, \( \xi \times i \leq s \) and \( i \in \mathbb{Z}^+ \). The value of \( \xi \) is 3 in WiSDP.

\[ \bar{h}_i = \frac{\lambda_i (h_{\xi(i-1)} + h_{\xi(i-1)+1} + \cdots + h_{\xi i})}{\xi}. \quad (6) \]

According to Equations (3) and (6), a packet merged the adjacent \( \xi \) subcarriers in CSI a stream that can be expressed as follows:

\[ \bar{h} = [\bar{h}_1 \bar{h}_2 \cdots \bar{h}_g]. \quad (7) \]

Rapid actions tend to make CSI waveforms changes more drastic [30, 36]. We can learn from Figure 3(c) that there is an SDP action between 0.5 s and 3.5 s and the CSI waveform changes relatively slowly when there is no action. In related studies, there are many techniques to split target actions from CSI streams [14, 30, 36]. However, these techniques are not suitable for WiSDP, which needs to detect actions with small time intervals and actions can start at any time. We define the action energy to reflect the intensity of the waveform change in a CSI stream. Mean Absolute Difference (MAD) algorithm, a common way to compare the similarity of two pictures, is used to calculate the difference between two adjacent packets in this paper. In Equation (8), \( \nabla h[i] \) is instantaneous action energy between \( i \) and \( i - m \) packets and \( m \) is a number that can be adjusted by...
Because actions start in different time, we measure the average energy of a CSI stream over time. In Equation (9), we sum up the instantaneous energy of all the antenna pairs and calculate the average action energy $E[1:n]$ of $n$ packets, where $\nabla \hat{h}[j][i]$ is the value of $\nabla \hat{h}[i]$ in $i$th subcarrier and $j$th antenna pair. Figure 4 displays the denoised energy of the SDP action.

$$E[1:n] = \frac{1}{g \times q \times n} \sum_{i=1}^{g} \sum_{j=1}^{q} \sum_{k=1}^{n} \nabla \hat{h}[j][i] [k]. \quad (9)$$

Figure 5 shows the action energy of different actions. Energy of sitting, walking and SDP are close and the energy of drinking is relatively lower. Since there is a big difference in energy between a action and no action, we can use action energy to estimate whether there are human actions or not and extract human actions from CSI streams.

In a real-time CSI stream, we do not know when user started an SDP action or when it ended. Before recognising quality of an SDP action, we must split it from the CSI stream by checking the action energy. To look for possible SDP actions in a real-time CSI stream, we propose to use the slide window algorithm with variable step size to search a possible SDP action by timeline which is shown in Algorithm 1. Through the algorithm, we use action energy to remove invalid data from the CSI stream, such as no action, instantaneous actions and small actions information. To improve the retrieval efficiency and ensure the accuracy of the data, we use big steps to detect the start and end ranges of an action, where the small steps are used to accurately retrieve the start or the end points. The effect of our algorithm mainly depends on the length of the sliding window $a$, the step length of the sliding window $b$ and the threshold of minimum energy of sliding window $E_m$. $\nabla \hat{h}[1:n]$ is the pre-processed CSI packets by Equation (8), $[E_l, E_h]$ is the energy range of an SDP action packet, $[T_l, T_h]$ is the time range of an SDP action and $\lambda$ is a coefficient for adjusting step of accurate search. In order to ensure the efficiency of the search and as accuracy as possible to obtain the start and end of the action, we first use bigger step $b$ to roughly estimate the range of SDP actions and then use the smaller step $b \lambda$ for accurate detection. In lines 6 and 16, we first find the start point and end point, respectively. Then we take a step back and continue with a smaller step in line 27. Finally, we get the action range $[\text{start} : \text{end}]$ in line 19. We can learn from Figure 5 that SDP actions cannot be effectively distinguished only by action energy. We use AdaBoost classifier with a maximum depth of five to find the SDP actions from other indoor actions.

### 3.4 Feature extraction

After obtaining the data corresponding to the SDP actions from the CSI stream, we extract features from denoised CSI matrix. First, we use PCA to extract the features from each SDP action. Although PCA is based on unsupervised learning to extract
features, it can often achieve better results in the case of few samples. By mapping the \( n \)-dimensional feature to the orthogonal \( k \)-dimension, the correlation between variables is eliminated. Through experiments, we found that using the features corresponding to the top 20 eigenvalues can well express the overview of SDP actions. Then, we divide an SDP action into two parts: push up and pull down, and extracted the features which is caused by different sites on the body and different speeds. We extract the profile for each part and choose three features as the profile: root mean square, variance and peak. Therefore, the number of extracted features is 26.

### 3.5 Classification

Although the classifiers based on neural network can achieve high accuracy, it needs a large number of samples, and collecting samples will take a lot of time. Once the category changes, it needs to be retrained and the training time is long. We have tested SVM, K-NearestNeighbor (KNN), Gradient Boosting Decison Tree (GBDT), Random Forest (RF) and Convolutional Neural Networks (CNN) algorithms. Finally, we select SVM model based on Radial Basis Function (RBF) to evaluate the SDP action quality. The reason is that it needs much less samples to perform well comparing to neural network, and training speed of the model is fast. SVM recognise action quality by finding the decision surface with the largest spacing between various samples. SVM can distinguish two categories, but we need to distinguish nine actions to evaluate SDP action quality. We thereby established \( \frac{k(k-1)}{2} \) SVM classifiers by One-Versus-One (OVO) way. In the process of classification, the category that receives the most votes through all classifiers is the predicted category of a sample. Compared with One-Versus-Rest (OVR), OVO do not need to retrain all classifiers when adding categories, and we can also avoid the problem that the data of negative samples is much larger than that of positive samples. The performance of the SVM models is related to the parameters. To find the optimal parameters, grid search which use result of cross-validation for score is used to traverse the main parameters.

### 4 Implementation and Evaluation

#### 4.1 Implementation

In WiSDP, we use the TL-WDR6300 router as the AP. The router has four antennas, two of which work in 2.4 GHz frequency band and the other work in 5 GHz frequency band. We use 5 GHz in WiSDP, which can reduce the interference between signals, and get higher receiving packet rate. We use Lenovo G480 laptop as DP, which is equipped with Intel i5-3210m processor, 8G DDR3 1600 MHz memory and Intel 5300 NIC. We modify the BIOS white list of the laptop to ensure compatibility between devices. WiSDP works on the laptop installed with OS Ubuntu 14.04.4 LTS and Python 3.7.3. We use ping command on the DP to make a request to the AP at a
Evaluation setup

Impact of the improved sliding SDP detection analysis

SDP recognition analysis

ALGORITHM 1 SDP Action Detection and Extraction

Input: Pre-processed CSI packets, $V_{k1 - 1}$; SDP action packet energy range, $[E_i, E_{i+1}]$; Timespan of an SDP action, $[T_i, T_{i+1}]$; Length of the sliding window, $a$; Step length of sliding window, $b$; Minimum energy of slide window, $E_m$; Exact search step size coefficient, $\lambda$.

Output: A possible SDP action packets range, $[start, end]$.

1: Initialise variable $i ← start ← end ← 0$; $isStart ← isEnd ← false$;
2: while $i + a ≤ n$ do
3: Calculate $E[i + a]$ by Equation (9);
4: if $E[i + a] > E_m$ then
5: if $isStart = 0$ then
6: $isStart ← true$; $i ← i + b$;
7: else
8: if $start = 0$ then
9: $start ← i$;
10: end if
11: end if
12: else
13: if $start > 0$ then
14: if $i ← start ∈ [T_i, T_{i+1}]$ and $\frac{E[start + b]}{E[start]} ∈ [E_i, E_{i+1}]$ then
15: if $isEnd = 0$ then
16: $isEnd ← true$; $i ← i + b$;
17: else
18: $end ← i + a$;
19: Return $[start, end]$;
20: end if
21: else
22: $isStart ← start ← 0$;
23: end if
24: end if
25: end if
26: if $(start = 0$ and $isStart = true$) or $(end = 0$ and $isEnd = true$) then
27: $i ← i + \frac{b}{\lambda}$;
28: else
29: $i ← i + b$;
30: end if
31: end while

rate of 200 times per second, and the response CSI streams can be obtained through the Intel 5300 NIC. The router and laptop we used are cheap and commercially available, which makes WiSDP available in practice.

4.2 Evaluation setup

In our experiments, AP and DP are installed at a height of 1 m and separated by 2.8 m. The height of the backless stool is 0.45 m. Four male volunteers who are 170–175 cm tall and weigh 57–70 kg took part in our experiments. In the line of sight environment, users sit in the middle of AP and DP, and repeat 60 times for each SDP action. In the experiment, we require that the layout of the room remain unchanged and that only one volunteer is allowed in the room. We make one-key sampling program based on CSI Tool to let Intel 5300 NIC receive the CSI packets. As the program runs, the computer camera begins to record the user’s actions for training model. With the timeline of the video, we can tag the CSI data for each SDP action.

4.3 SDP detection analysis

To evaluate the accuracy of the SDP detection, we introduce four basic indicators: True Positive (TP), False Positive (FP), True Negative (TN) and False Negative (FN), which are based on the real category and prediction category. True Positive Rate (TPR), False Positive Rate (FAR) and Accuracy can be express as follows:

$$TPR = \frac{TP}{TP + FN}, \quad (10)$$
$$FAR = \frac{FP}{FP + TN}, \quad (11)$$
$$Accuracy = \frac{TP + TN}{TP + FN + FP + TN}. \quad (12)$$

We invite four volunteers to carry out experiments involving five indoor actions (motionless, sitting, drinking, walking and SDP). For motionless, sitting, drinking and walking, we require user to repeat each action 90 times. For SDP, we collect 10 samples per user for each action in Table 1. Figure 6 is the gray scale of the confusion matrix for five indoor actions. We use cross-validation to get average value. From the confusion matrix, we can know that WiSDP using action energy can distinguish motionless scenario from other four scenarios without any errors. We use AdaBoost classifier to distinguish SDP from sitting, drinking, and walking with a TP rate of 99%. We can infer from the confusion matrix that the FAR of WiSDP is 0. In conclusion, the average detection TPR of the five indoor scenarios is 98.6%.

4.4 SDP recognition analysis

4.4.1 Impact of the improved sliding window algorithm on recognition accuracy

The improved sliding window algorithm can reduce redundant data. In order to verify that the improved sliding window algorithm can improve the recognition accuracy, we compare the algorithm with the general sliding window algorithm in WiSDP. To ensure the reliability of the experimental results, we use...
a data set collected from different volunteers to evaluate the algorithms. The results can be seen in Figure 7. By comparing recognition accuracy of the two algorithms, we can find that the improved sliding window algorithm can improve the recognition accuracy in WiSDP, in which the recognition accuracy of user1, user3 and user4 is obviously improved and the recognition accuracy of all users is improved by 0.53% on average. In theory, our algorithm can only improve the recognition accuracy slightly, because each person has different behavioural characteristics, which sometimes make the recognition accuracy less obvious. For example, our algorithm has not achieved good results for user2. Although the recognition accuracy is not improved, the recognition accuracy is not reduced. In general, the improved sliding window algorithm is helpful to improve the recognition accuracy.

### 4.4.2 Impact of classifier on recognition accuracy

We compare the performance of various machine learning algorithms on our data set, including SVM, KNN, GBDT, RF and CNN. For each SDP in Table 1, we collect 240 samples from four volunteers. We randomly extract 20% of the entire data set as a test data set, and then use a fourfold cross-validation on the remaining data set to adjust and optimise model parameters. In Table 3, we present the main parameters of the algorithms in Table 2 in our experimental environments. We can know from Table 2 that SVM is significantly better than other algorithms with a recognition rate of 95.11%. The specific role of these parameters can be referred to the scikit-learn machine learning library [37]. Training time and running time of KNN, GBDT and RF are close to SVM, but recognition accuracy of SVM is much higher than others. Although the training time and running time of GBDT are the least, the training time and recognition accuracy of GBDT are greatly affected by the learning_rate parameter. We have tested three CNN-based schemes: 1D-CNN, 2D-CNN and 3D-CNN, and 1D-CNN have better performances in recognition accuracy and training time. The archi-
TABLE 3  The main parameters of the algorithms in TABLE 2

| Algorithm | Parameter            | Value                  |
|-----------|----------------------|------------------------|
| SVM       | C                    | 12                     |
|           | decision_function_shape | ovo                   |
|           | kernel                | rbf                    |
|           | gamma                 | 0.02                   |
| KNN       | p                    | 2                      |
|           | n_neighbours         | 8                      |
|           | weights               | distance               |
|           | algorithm             | ball_tree              |
| GBDT      | learning_rate        | 0.08                   |
|           | n_estimators          | 200                    |
|           | criterion             | friedman_mse           |
|           | loss                  | deviance               |
| RF        | n_estimators          | 200                    |
|           | max_depth             | 7                      |
|           | criterion             | entropy                |
|           | max_features          | 10                     |
| CNN       | optimiser             | adagrad                |
|           | loss                  | sparse_categorical_crossentropy |
|           | metrics               | accuracy               |

TABLE 4  The struct of 1D-CNN in TABLE 2

| Layer index | Layer (size) | Output shape | Activation |
|-------------|--------------|--------------|------------|
| 1 (input)   | Dense(180*70) | (180, 70)    | —          |
| 2           | Conv1D(32*3)  | (178, 32)    | relu       |
| 3           | MaxPooling1D(3) | (59, 32)    | —          |
| 4           | Conv1D(64*3)  | (57, 64)     | relu       |
| 5           | MaxPooling1D(2) | (28, 64)   | —          |
| 6           | Conv1D(32*3)  | (26, 32)     | relu       |
| 7           | Flatten       | (832)        | —          |
| 8 (output)  | Dense(9)      | (9)          | softmax    |

Architecture of 1D-CNN is illustrated in Table 4. Conv1D layers are used to perceive the local information of the CSI waveforms. MaxPooling1D layers are used to reduce dimensions of feature, compress the number of data and parameters, reduce the overfitting, and improve the fault tolerance of the algorithm. Flatten layer is used to convert multi-dimensional inputs into one-dimensional outputs and connect the layer 6 and layer 8. Although CNN and SVM have similar recognition accuracy, the training time and running time of CNN are longer than non-neural network algorithms. In addition, improper network structure will lead to overfitting or underfitting, and network structure needs to be adjusted accordingly to achieve a better recognition accuracy. In summary, CNN is more suitable for large training data sets. Acquiring a large number of training samples in a practical application will take a significant amount of time, which is not conducive to WiSDP applications in a variety of scenarios. Compared with CNN, SVM can achieve an ideal recognition accuracy by training fewer parameters.

4.5  | Robustness analysis

We evaluate the system performance with four factors that affect the action recognition result into consideration: the receiving packet rate, the number of antennas, the sample number and the environment.

4.5.1  | Impact of receiving packet rate on recognition accuracy

We simulate different receiving packet rates with different time intervals. We collect 50 samples generated by each SDP action with one sending and receiving antenna pair from four volunteers. A sample contains all the CSI packets from the beginning to the end of an action. We use the fourfold cross-validation to get more convincing results in our experiments. We can learn from Figure 8(a) that increasing the receiving packet rate does not necessarily improve the recognition accuracy of SDP actions. When the receiving packet rate is from 20 to 200 packets/s, there is little difference in recognition accuracy. The recognition accuracy begins to decline obviously when the receiving packet rate is lower than 20 packets/s. Too high receiving packet rate will increase the time of data processing, and too low receiving packet rate will not achieve a good recognition rate. Thus the optimal receiving packet rate is 20 packets/s based on the SDP actions recognition rate of four users experiments.

4.5.2  | Impact of the antenna number on recognition accuracy

Based on receiving packet rates experiments before, we set the receiving packet rate to 20 packets/s. We have two transmitting antennas and three receiving antennas, so there are six sending and receiving antenna pairs. We increase the number of antenna pairs one by one and use cross-validation to test the accuracy of the recognition. As depicted in Figure 8(b), increasing the number of antenna pairs is conducive to improving the recognition accuracy. Compared with using only one antenna pair, the recognition accuracy can be improved by 5.38% and 8.37% on average when using two antenna pairs and six antenna pairs, respectively. The more sending and receiving antenna pairs work, the more CSI data we can obtain. The large amount of CSI data allows us to analyse the details of the SDP action more comprehensively and get higher recognition accuracy.

4.5.3  | Impact of sample number on recognition accuracy

In this experiments, the receiving packet rate is 20 packets/s and the number of antenna pairs is 6. We collect 60 samples
for each SDP action to train the model and a number of samples are randomly selected from each action category to form a new data set. We conduct three groups of experiments to evaluate the relationship between sample number and recognition accuracy. We rebuild SVM model with 40, 50 and 60 samples for each action. Figure 8(c) shows the comparison of the Leave-One-Out cross-validation results for nine SDP actions. With the increase of sample number, the recognition accuracy is increasing. Using 40 samples to build model, we can get the recognition accuracy of 93.67%. When the number of samples for each SDP increases from 40 to 50, the average accuracy of four users increases to 95.07%. Compared with 50 training samples, the recognition accuracy increased by 1.11% on average when the number of samples increased to 60. When the sample number is small, there may be overfitting in the SVM model, which will lead to a low recognition rate. With the increase of the sample number, the generalisation performance of our model is enhanced and the recognition accuracy is improved, while SDP3, SDP4, SDP5 and SDP7 are more sensitive to training sample number than other actions.

4.5.4 Impact of the environment on recognition TPR

We conduct our experiments in two different scenarios to assess the impact of the environment on the recognition TPR. With the same experimental parameters, we carried out the experiments in the student dormitory and family living room. Figure 9 shows the room layout for the two scenarios. Figure 9(a) is
an ordinary student dormitory, in which there are beds, tables, chairs and some daily furnishings. Figure 9(b) is an ordinary family living room with a TV set, a dining table and a sofa. In the experiments, we do not need to move any furniture in the house, only to avoid irrelevant people into the experimental environment. In each scenario, we collected 60 CSI samples for each actions from four volunteers. The receiving packet rate is 20 packets/s and six antenna pairs is used for collecting data. Figure 10 displays the recognition result of nine SDP actions in two scenarios. We can learn from the confusion matrix that WiSDP can recognise the nine SDP action with an average TPR of 94.66% and 95.11% in student dormitory and family living room. The TPR of SDP6, SDP7 and SDP8 is lower than average value and the misclassification among them is the main reason for the low recognition accuracy. The other six SDP actions have good TPR in two scenarios. On the whole, WiSDP can work well in different room layouts.

5 | DISCUSSION

Although WiSDP have achieved a good recognition accuracy for quality of SDP action, there are three limitations in our system:

1. Unrelated users walking in the room will reduce recognition accuracy. We know that CSI is easily disturbed by the environment, and the movement of other users will introduce a lot of interference, so that it cannot effectively recognise the action quality.
2. WiSDP is currently only working within the line of sight area. In the line of sight area, CSI is more sensitive to human movement, which helps to catch subtle differences.
3. WiSDP can only identify a single error in an SDP action. When there is more than one error in an SDP action, the system cannot recognise all the errors. Because the various wrong are superimposed on each other, the characteristics will be different from the original.

6 | CONCLUSION

We propose a fine-grained action quality recognition system taking SDP action as an example. In the system, two cheap consumer devices, a laptop and a router are used for collecting CSI of Wi-Fi signals. We use PCA and Butterworth low-pass filter to denoise the collected CSI. We propose the concept of action energy to estimate the intensity of the waveform change, which can be used to distinguish some different actions by thresholds. The action energy detection algorithm based on sliding adaptive size window is proposed for obtaining CSI streams segments in which only the CSI packets related with SDP actions are contained. We test different algorithms and choose SVM which can reach high accuracy of recognising action quality with small sample number. The experiment results show that we can achieve an average TPR of 94.66% and 95.11% in two different scenarios.

REFERENCES

1. Yuan, X., et al.: Performance analysis of IEEE 802.15.6-based coexisting mobile WBANS with prioritized traffic and dynamic interference. IEEE Trans. Wireless Commun. 17 (8), 5637–5652 (2018)
2. Dawar, N., Kehtarnavaz, N.: Action detection and recognition in continuous action streams by deep learning-based sensing fusion. IEEE Sens. J. 18 (23), 9660–9668 (2018)
3. Ghojogh, B. et al.: Fisherposes for human action recognition using kinect sensor data. IEEE Sens. J. 18 (4), 1612–1627 (2018)
4. Zhang, J., et al.: Semi-supervised image-to-video adaptation for video action recognition. IEEE Trans. Cybern. 47 (4), 960–973 (2017)
5. Tu, Z., et al.: Action-stage emphasized spatiotemporal VLAD for video action recognition. IEEE Trans. Image Process. 28 (6), 2799–2812 (2019)
6. Chen, Z., et al.: Wi6 CSI based passive human activity recognition using attention based BLSTM. IEEE Trans. Mob. Comput. 18 (11), 2714–2724 (2019)
7. Wang, Z., et al.: A survey on CSI-based human behavior recognition in through-the-wall scenario. IEEE Access 7, 78772–78793 (2019)
8. Chen, C. et al.: A real-time human action recognition system using depth and inertial sensor fusion. IEEE Sensors J. 16 (3), 773–781 (2016)
9. Khan, P., et al.: Differential channel state information based human activity recognition in IoT networks. IEEE Internet Things J. 7, 11290–11302 (2020)
10. Lv, S., et al.: Qualitative action recognition by wireless radio signals in human-machine systems. IEEE Trans. Hum.-Mach. Syst. 47 (6), 789–800 (2017)
11. Yang, C. et al.: Unsupervised drowsy driving detection with RFID. IEEE Trans. Veh. Technol. 69 (8), 8151–8163 (2020)
12. Chen, Z., et al.: Wifi CSI based passive human activity recognition using attention based BLSTM. IEEE Trans. Mob. Comput. 18 (11), 2714–2724 (2019)
13. Zhang, T., et al.: Wigrus: a wifi-based gesture recognition system using software-defined radio. IEEE Access 7, 92758–92767 (2019)
14. Haseeb, M.A.A., Parasuraman, R.: Wisture: touch-less hand gesture classification in unmodified smartphones using wi-fi signals. IEEE Sens. J. 19 (1), 257–267 (2019)
15. Ren, S., et al.: Intelligent contactless gesture recognition using WLAN physical layer information. IEEE Access 7, 92758–92767 (2019)
16. Cao, X. et al.: Wi-vire: fine-grained writing recognition using wi-fi signals. 2016 IEEE Trustcom/BigDataSE/ISPA, August 2016, pp. 1366–1373. IEEE, Piscataway, NJ (2016)
17. Volta, E., Volpe, G.: Automated analysis of postural and movement qualities of violin players. 2019 International Workshop on Multilayer Music Representation and Processing (MMRP), January 2019, pp. 56–59. IEEE, Piscataway, NJ (2019)
18. Głowinski, D., et al.: Analyzing expressive styles and functions of body movement in violinist performance. MOCO ’14: Proceedings of the 2014 International Workshop on Movement and Computing, pp. 154–155. ACM, New York, NY (2014)
19. Velloso, E., et al.: Qualitative activity recognition of weight lifting exercises. AH ’13: Proceedings of the 4th Augmented Human International Conference, pp. 116–123. ACM, New York, NY (2013)
20. De Giorgis, N., et al.: Evaluating movement quality through intrapersonal synchronization. IEEE Trans. Human-Mach. Syst. 49 (4), 304–313 (2019)
21. Berkati, O., Srifi, M.N.: Predict driver fatigue using facial features. 2018 International Symposium on Advanced Electrical and Communication Technologies (ISAECT), November 2018, pp. 1–5. IEEE, Piscataway, NJ (2018)
22. Dornakia, F., et al.: Driver drowsiness detection in facial images. 2018 Eighth International Conference on Image Processing Theory, Tools and Applications (IPTA), November 2018, pp. 1–6. IEEE, Piscataway, NJ (2018)
23. Elkholy, A., et al.: Efficient and robust skeleton-based quality assessment and abnormality detection in human action performance. IEEE J. Biomed. Health Inf. 24 (1), 280–291 (2020)
24. Niewiadomski, R., et al.: Analysis of movement quality in full-body physical activities. ACM Trans. Interact. Intell. Syst. 9 (1), 1–20 (2019)
25. Xue, W., et al.: Improved wi-fi RSSI measurement for indoor localization. IEEE Sensors J. 17 (7), 2224–2230 (2017)
26. Xue, W., et al.: A new algorithm for indoor RSSI radio map reconstruction. IEEE Access 6, 76118–76125 (2018)
27. Abdelnasser, H., et al.: A ubiquitous wifi-based fine-grained gesture recognition system. IEEE Trans. Mobile Comput. 18 (11), 2474–2487 (2019)
28. Liu, J., et al.: Monitoring vital signs and postures during sleep using wi-fi signals. IEEE Internet Things J. 5 (3), 2071–2084 (2018)
29. Ali, K., et al.: Recognizing keystrokes using wi-fi devices. IEEE J. Sel. Areas Commun. 35 (5), 1175–1190 (2017)
30. Fu, Z., et al.: Writing in the air with wi-fi signals for virtual reality devices. IEEE Trans. Mob. Comput. 18 (2), 473–484 (2019)
31. Halperin, D., et al.: Predictable 802.11 packet delivery from wireless channel measurements. SIGCOMM’10: Proceedings of the ACM SIGCOMM 2010 Conference, pp. 159–170. ACM, New York, NY (2010)
32. Davies, L., Garther, U.: The identification of multiple outliers. J. Am. Stat. Assoc. 88 (423), 782–792 (1993)
33. Nandakumar, R. et al.: Wi-fi gesture recognition on existing devices. Comput. Sci. 3 (2), 17–22 (2014)
34. Wang, W., et al.: Understanding and modeling of wifi signal based human activity recognition. MobiCom ’15: Proceedings of the 21st Annual International Conference on Mobile Computing and Networking, pp. 65–76. ACM, New York, NY (2015)
35. Zhu, H., et al.: R-ttwd: Robust device-free through-the-wall detection of moving human with wifi. IEEE J. Sel. Areas Commun. 35 (5), 1090–1103 (2017)
36. Wang, Y. et al.: Wifall: Device-free fall detection by wireless networks. IEEE Trans. Mobile Comput. 16 (2), 581–594 (2017)
37. Pedregosa, F., et al.: Scikit-learn: Machine learning in python. J. Mach. Learn. Res. 12, 2825–2830 (2011)

How to cite this article: Qin Y, Pan S, Zhou W, Pan D, Li Z. Consumer Wi-Fi device based action quality recognition: An illustrative example of seated dumbbell press action. IET Commun. 2021;15:613–626. https://doi.org/10.1049/cmu2.12093