IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received 16 January 2024, accepted 23 January 2024, date of publication 31 January 2024, date of current version 8 February 2024.

Digital Object Identifier 10.1109/ACCESS.2024.3360490

APPLIED RESEARCH

Human Activity Recognition via Wi-Fi and Inertial Sensors With Machine Learning

WEI GUO[1], (Graduate Student Member, IEEE), SHUNSEI YAMAGISHI[1], AND LEI JING[2], (Member, IEEE)

[1]Graduate School of Computer and Information Systems, The University of Aizu, Aizuwakamatsu 965-0006, Japan
[2]School of Computer Science and Engineering, The University of Aizu, Aizuwakamatsu 965-0006, Japan

Corresponding author: Lei Jing (leijing@u-aizu.ac.jp)

This work was supported in part by the Japan Society for the Promotion of Science (JSPS) KAKENHI under Grant 22K12114, in part by the JKA Foundation, and in part by the New Energy and Industrial Technology Development Organization (NEDO) Intensive Support for Young Promising Researchers under Grant JPNP20004.

This work involved human subjects or animals in its research. Approval of all ethical and experimental procedures and protocols was granted by the Ethic Review Board of The University of Aizu.

ABSTRACT Human activity recognition (HAR) plays a crucial role in human-computer interaction, smart home, health monitoring and elderly care. However, existing methods typically utilize camera, radio frequency (RF) signals or wearable devices for activity recognition. Each single-sensor modality has its inherent limitations, like camera-based methods having blind spots, Wi-Fi-based methods depending on the environment and the inconvenience of wearing Inertial Measurement Unit (IMU) devices. In this paper, we propose a HAR system that leverages three types of sensor combinations: Wi-Fi, IMU and a hybrid of Wi-Fi+IMU. We utilize the Channel State Information (CSI) provided by Wi-Fi and the accelerometer and gyroscope data from IMU devices to capture activity characteristics. Then, we employ six machine learning algorithms to recognize eight types of daily activities. These algorithms include Support Vector Machine (SVM), Multi-layer Perceptron (MLP), Decision Tree, Random Forest, Logistic Regression and k-Nearest Neighbors (kNN). Additionally, we investigate the accuracy of hand gesture recognition using different sensor combinations and analyze the calculation speed of each combination. We conduct a survey to collect user feedback on the performance of various sensor combinations in our HAR system. The results show that the combination of CSI+IMU yields the best HAR recognition accuracy, with a accuracy of 89.38%. The SVM algorithm consistently performs well across all systems, especially excelling in the CSI+IMU system supported by energy and average Fast Fourier Transform (FFT) features. However, we also find that the success of sensor fusion depends on specific algorithms and features. Fusion of CSI and IMU does not universally enhance recognition accuracy for all features and algorithms and can, in some cases, actually reduce accuracy.

INDEX TERMS Human activity recognition, channel state information, inertial measurement unit, sensor fusion, feature fusion, Wi-Fi sensing, wearable sensing, machine learning.

I. INTRODUCTION

In recent years, HAR has played an increasingly important role in human-computer interaction, health monitoring and elderly care. It employs cameras [1], [2], [3], wearable sensors [4], [5], [6], [7] and RF signals [8], [9], [10] to recognize people's daily activities, body movements and vital signs. HAR systems can be categorized into three types based on the mode of sensor carriage: device-based, device-free and multi-sensor fusion.

Device-based HAR systems require users to carry sensors on specific parts of their body, allowing the sensors to gather

The associate editor coordinating the review of this manuscript and approving it for publication was Patrizia Livreri.

human activity information. Foerster et al. [11] validate the use of accelerometers for detecting posture and motion. They demonstrate their effectiveness in both laboratory and real-life settings by comparing them with observed behaviors. Pancholi et al. [12] develop a cost-effective Electromyography (EMG) system with eight channels for accurate arm movement recognition, utilizing machine learning for enhanced performance. To minimize the discomfort and inconvenience experienced by users due to the deployment of sensors, Alevizaki et al. [13] introduce a hierarchical framework using smartwatch IMU data to learn daily living activities at varying granularity levels, employing CNN-LSTM classifiers for improved accuracy and efficiency. Jing et al. [14] design a compact wireless ring, named Magic Ring, equipped with a tri-axial accelerometer, to be worn on the finger for interaction with surrounding electronic devices, further simplifying HAR wearable devices. Wearable devices can directly acquire users' motion information for accurate activity recognition. However, in elderly care and monitoring for abnormal behaviors, requiring users to wear complex devices is impractical.

In device-free activity recognition, the camera can provide image information intuitively. Researchers use cameras to identify people's behavior and walking trajectories, lock suspicious targets, predict possible abnormal behaviors in advance, and protect assets that are vulnerable to attacks [15]. Furthermore, researchers utilize RGB-D cameras for the identification of dangerous activities, such as falls, within the daily lives of elderly individuals, thereby facilitating their independence and mitigating associated risks [16]. However, the camera-based method invades user privacy, suffers from limited recognition accuracy due to ambient light, and has issues with visual blind spots [17]. To tackle these issues, studies introduce HAR systems based on acoustic signals [18], Millimeter Wave (mmWave) [19], Ultra-Wideband (UWB) [20] and Wi-Fi devices [21]. Sound and mmWave are suitable for scenes with a smaller monitoring range, and mmWave is suitable for accurate recognition of tiny movements and the cost is higher. UWB has a wide sensing range, but the cost is also high. In order to reduce costs and expand sensing range, Wi-Fi devices widely deployed indoors have become the focus of studies. WiSee [22] and Wi-Vi [23] utilize Universal Software Radio Peripheral (USRP) devices to acquire Orthogonal Frequency-Division Multiplexing (OFDM) signals, enabling the estimation of a target's location and gesture actions, thereby establishing a groundwork for Wi-Fi-based HAR. E-eyes [24] utilizes commercial Wi-Fi devices operating under the 802.11n protocol to acquire CSI, and leverages the amplitude of CSI to identify in-place and walking activities. CARM [21] proposes two models, the CSI-Speed Model and the CSI-Activity Model: the first links changes in CSI with human movement speeds, while the second associates the movement speeds of various body parts with specific activities, enabling activity recognition by matching with CSI profiles. Despite the widespread availability of Wi-Fi devices, a change in environment or targets requires recollecting data and retraining models to ensure the effectiveness of HAR systems. To address this issue, Widar3.0 [25] proposes a zero-effort, cross-domain gesture recognition system, modeling the relationship between complex gestures and CSI dynamics, extracting domain-independent gesture velocity profiles, and employing a deep learning model to recognize gestures based on their spatial-temporal characteristics. However, it employs one transmitter and six receivers, which is impractical in real-life scenarios. WiGesture [26] achieves location-independent gesture recognition using one transmitter and two receivers, but the devices require specific placement. Conversely, Wi-Monitor [27] employs a pair of Wi-Fi devices, which do not require special placement, to recognize continuous human activities in daily life. Yet, even though Wi-Fi offers a wide sensing range and precise recognition, the CSI varies with changes in the surroundings, and the system cannot detect if the target exits the area.

To address the limitations of unimodal HAR systems, multimodal systems have garnered attention. In multimodal systems [28], activity data is collected simultaneously using two or more types of sensors. For instance, data from IMUs and pressure insoles are fused to detect falls, while IMUs and cameras are combined to address occlusion issues in visual systems. In this study, we particularly focus on the integration of IMUs with Wi-Fi devices, as they offer complementary benefits. In [29], signals from Wi-Fi and inertial sensors are fused and a Hidden Markov Model (HMM) is utilized to identify high-speed actions in tennis. Reference [30] focuses on combining Wi-Fi with data from floor-mounted accelerometers to detect falls. Similarly, WiWeHAR [31] fuses features from Wi-Fi and IMU to recognize four actions (walking, falling, sitting, and picking up an object), innovatively extracting the mean Doppler shift from CSI as a feature, thereby enhancing recognition accuracy to 99.6%-100%. Despite WiWeHAR achieving a recognition rate of over 99%, its system focuses only on four full-body movements, which are insufficient to comprehensively describe daily activities. To more clearly illustrate the differences between unimodal and multimodal HAR systems, as well as to distinguish our research from various multimodal systems, we conduct a comparative analysis in Table1. From this table, it is evident that Wi-Fi-based systems are unable to provide continuous monitoring. In contrast, the multimodal systems that do offer continuous monitoring are limited in recognizing specific tennis movements and can only identify a few simple whole-body activities. These systems fall short of comprehensively describing activities in daily life. To better depict everyday actions, this paper compares the performance and computational speed of various machine learning algorithms using data collected from different sensor combinations under distinct features. In this paper, we propose a HAR system that utilizes three types of sensor combinations: Wi-Fi, IMU,

and a hybrid of Wi-Fi+IMU. The system captures activity characteristics using CSI and accelerometer and gyroscope data from IMU devices. Six machine learning algorithms – SVM, MLP, Decision Tree, Random Forest, Logistic Regression, and kNN – are employed to recognize eight types of daily activities. Additionally, the accuracy of hand gesture recognition using different sensor combinations is investigated, along with the analysis of the calculation speed of each combination. Our main contributions are as follows:

1) We propose a HAR system that recognizes eight activities using three sensor combinations: Wi-Fi, IMU, and Wi-Fi+IMU, and compare their results.
2) Six machine learning algorithms are used for activity recognition, with a comparative analysis of recognition accuracy across different sensor combinations.
3) We invited five participants for testing, with the Wi-Fi+IMU combination achieving the best recognition accuracy of 89.38%.

The remainder of the paper is organized as follows: Section II introduces related works. In Section III, we provide an overview of the HAR system. Section IV outlines the methodology, including details on data collection, preprocessing, and feature definition. The design and implementation of experiments are presented in Section V. Section VI is the evaluation, where the HAR system is validated and results are comparatively analyzed. Finally, we summarize the conclusions.

II. RELATED WORKS

This section provides a comprehensive overview of the current state-of-the-art in HAR systems. We review various approaches, including camera-based, CSI-based, IMU-based, and the fusion of CSI and IMU techniques. Each of these methods has its unique strengths and limitations in the context of HAR. By examining these diverse methodologies, we aim to identify gaps in current research and explore how our proposed HAR system, which integrates CSI, IMU, and CSI+IMU, addresses these challenges.

A. CAMERA-BASED

Starner et al. [32] proposed two systems that interpret American Sign Language (ASL) and track unadorned hands in real-time with one color camera, using HMM. The kinds of ASL in this experiment are six pronouns, nine verbs, twenty nouns, and five adjectives. They tested two camera location, an observer of the signer and the point of view of the signer himself. The first person system adopted hat-mounted camera. The second person system adopted desk-based tracking camera. In the desk-based recognizer experiment, 478 sentences were obtained. 384 sentences were used for training, 94 sentences were used for testing. In the wearable-based recognizer experiment, 500 sentences were collected and 400 sentences were used for training, 100 sentences were used for testing.

Recognition using a camera can only recognize events in the field of view of the camera. Therefore, the camera base is vulnerable to blind spots.

B. CSI-BASED

Taylor et al. [33] proposed HAR system using CSI obtained from UWRP. They used two USRPs (X310/x300 models) devices. They tested two activities, sitting down on a chair and standing from a chair. In the experiment, each sample received by the USRP contains 64 subcarriers. For each type of activities, 30 samples were selected, forming a dataset of size 3840. They tested five algorithms, Random Forest, kNN, SVM, Neural network model, and Ensemble Classifier. The accuracy (cross validation, train test split) of Random Forest (92.47%, 96.70%), kNN (88.17%, 90.71%), SVM (84.68%, 81.77%), Neural network model (90.05%, 93.40%), Ensemble Classifier (92.18%, 93.83%). They created a real-time recognition web system using Random Forest which is the best accuracy result. The other dataset was tested in this experiment using smart-phones equipped with accelerometers. They claim that, this experiment demonstrates USRP accuracy is similar to wearable sensors or higher than wearable sensors in recognition of body movements. The benchmark dataset accuracy (cross validation, train test split) of Random Forest (91.20%, 96.49%), kNN (77.06%, 92.48%), SVM (85.90%, 86.21%), Neural network model (89.21%, 96.11%), Ensemble Classifier (92.40%, 97.74%).

Shi et al. [34] propose a CSI cleaning and enhancement method (CSI-CE). The method has two stages, activity-related information extraction (ARIE) and correlation feature extraction based on principal component analysis (CFE-PCA). They use Intel 5300 network interface card (NIC) with CSI tools [35] to obtain CSI data. They recognized six activities, empty, standup, laying, walk, standing, fall with HAR-MN-EF(MN:Matching Network [36], EF:Enhanced Features), EI [37], RNN, TNNAR [38]. HAR-MN-EF is the highest accuracy. In case of HAR-MN-EF with CSI-CE, the accuracy is 75.1%, the training time is 34.5 minutes. On the other hand, HAR-MN-EF without CSI-CE, the accuracy is 58.2%, the training time is 89.7 minutes.

Li et al. [39] tested indoor localization. In corridor, 7×7 finger print, it is 4.2m $\times$ 4.2m with 49 sampling points. In conference room, there are 29 sampling points. They use TP-Link wireless router with 2.4GHz. The sampling rate is 200Hz and selected three optimal subcarriers. They changed the height of antenna, 0.2, 0.4, 0.6, 0.8, 1.0m. Under the accuracy of 0.6m, 93.9% in conference room and 96% in corridor.

In location detection with finger-print method, it is environment dependent because CSI's amplitude and phase are changed by reflection from objects, people, etc. In addtion, it changes with the size of experiment room. In HAR, only a very limited kinds of actions can be recognized because the signal of Wi-Fi change small if the action is small. The more actions we recognize, the lower the recognition rate.

TABLE 1. Comparative analysis of our system with existing studies.

Work	Device	Parameter	Method	Activity	Accuracy	Continuity
E-eyes [24]	2 Wi-Fi links	Amplitude profiles	Profile matching	Whole body activities: cooking, eating, washing dishes, studying, brushing, bathing, gaming, sleeping, watching TV and passing through doors.	96%	✗
CARM [21]	2 Wi-Fi links	Amplitude	HMM model	Whole body activities: running, walking, sitting down, opening refrigerator, falling, boxing, pushing one hand and brushing.	96%	✗
Widar3.0 [25]	6 Wi-Fi links	DFS profile and body-coordinate velocity profile	RNN+CNN	Upper limb activities: pushing, pulling, sweeping, clapping, sliding, drawing circle and drawing zigzag.	92.7%	✗
WiGesture [26]	2 Wi-Fi links	Motion navigation primitive	Threshold method	Single hand activities: sliding, tick, zig-zag, cross, drawing circle, waving and drawing number 0-9.	92%	✗
Wi-Monitor [27]	1 Wi-Fi link	Amplitude and phase	Temporal convolutional network	Whole body activities: static, walking, waving hand, boxing, lifting leg, sitting down, standing up, jumping, raising hand and picking up.	92.88%	✗
Y. Shu *et al.* [29]	1 Wi-Fi link and 1 IMU	Amplitude and quaternion	HMM model	Table tennis movements: forehand drive, forehand lob, backhand lob, backhand twist, hop step, step, and squat.	97.43%	✓
R. Ramezani *et al.* [30]	1 Wi-Fi link and 3-axes accelerometer	Amplitude and acceleration	SVM model and Adaboost classifiers	Whole body activities: falling, walking and jumping.	95%	✓
WiWeHAR [31]	1 Wi-Fi link and 1 IMU	Time-variant mean Doppler shift and magnitude of each sensor	SVM model	Whole body activities: walking, falling, picking object and sitting.	99.6%–100%	✓
Our system	1 Wi-Fi link and 3 IMUs	Amplitude, phase, accelerometers and gyroscopes data	SVM, MLP, Decision Tree, Random Forest, Logistic Regression and kNN.	Whole body activities: walking, cleaning with a vacuum, moving heavy object, sitting down, squatting, shaking hand, typing on a keyboard and writing on a whiteboard.	89.38%	✓

C. IMU-BASED

Ravi et al. [40] proposed activity recognition using x, y, z axis of a wearable sensor, CDXL04M3. They collected eight activities, standing, walking, running, climbing up stairs, climbing down stairs, sit-ups, vacuuming and brushing teeth. The window size is 256 and overlapping is 128 (50% overlap). The sampling rate is 50 per second. They extracted four features from the x, y, z axis, Mean, Standard Deviation, Energy and Correlation. They selected Decision Tables, Decision Trees, kNN, SVM and Naive Bayes. There are four experiment settings and the result of the experiment is Plurality Voting which is the best in setting 1, 2, 3, over 90% accuracy. Boosted SVM is the best in setting 4.

Jing and Cheng create recognition system of 10 daily activities and 1 accident action [41]. They use compact motion capture sensor node named WS-WiFi. It can obtain three dimension acceleration and angular velocity. The sampling rate of the sensors is 20Hz. Window size is 64 which is 3,2 seconds and overlap is the half of window size. They calculate mean, standard deviation, and energy of acceleration data. kNN is adopted and k means method to select training samples. Attach sensors on right wrist, left ankle and right side of waist of testers. The overall recognition rate is 94.3%.

Jiménez et al. [42] compared Pedestrian Dead Reckoning (PDR) algorithms using low-cost Micro Electro Mechanical Syatems (MEMS) IMU which is MTi Xsens IMU. They mounted the IMU on the instep of a right shoe. The sampling rate is 100Hz. The paper describes the result of most relevant algorithms for step detection, stride length, heading and position estimation, indoor and outdoor. They described that avoiding excessive error accumulation is the challenge in PDR algorithms in this paper.

The error accumulate in PDR with consumer level IMU. Therefore, it is difficult to tracking location with consumer level IMU.

D. CSI AND IMU FUSION

Muaaz et al. [31] combined CSI and wearable sensor data for HAR using SVM. They recognized walking, sitting, falling and picking up an object from the floor in the experiment. The IMU sensor obtains tri-axis accelerometer, tri-axis gyroscope and tri-axis magnetometer. The Wi-Fi receiver equipped Intel Wi-Fi Link 5300, which is Network Interface Card (NIC), the configuration is 5GHz band they used. Two laptops which were installed CSI-tool [35] and equipped the NICs. The testers were attached an IMU to their lower back. The sampling rate of the

IMU is 400Hz. They use time-variant Mean Doppler Shift (MDS) for recognition. Before calculate the MDS, preprocess the data. They eliminate the CSI phase distortions and execute phase calibration. They reduced dimensions by Principal Component Analysis (PCA) and filtered by low-pass filter. Finally, they calculate MDS from filtered principal component. In case of preprocessing of IMU, they compute magnitude of each sensor type, accelerometer, gyroscope and magnetometer. In their experiment, 4 unimodal classifier and 11 multimodal classifier which are bimodal, trimodal, quadmodal were tested. Multimodal approach accuracy is 99.6% to 100%. But unimodal is 91.5% to 98.1%.

E. SHORTCOMING OF THE CURRENT TECHNOLOGIES

In current technologies, camera devices have the shortcoming of a dead corner. Dead corner cannot be recognized with the camera alone. Wi-Fi devices depend on experimental environment. Therefore, it is unlikely that it can be reproduced in other environments. The consumer-level inertial sensor shortcoming is accumulation of error. Errors accumulate when operating for a long time.

Compared to other sensor fusion efforts, studies focusing on the integration of CSI with IMU for HAR are relatively scarce. Significantly, this research conducts a multifaceted evaluation of HAR by integrating three sensor combinations within a single system framework. Our study has developed a robust HAR system, undertaking extensive synthetic experimental research, thereby providing a foundational guide for various HAR applications.

III. SYSTEM OVERVIEW

In this section, we present an overview of our proposed HAR system. We present the application model and the detailed structure of the system, outlining the key components and the workflow. The system is designed to capture a comprehensive dataset from both IMU devices and Wi-Fi transceivers, ensuring precise activity recognition. We explain how these components interact within the system to process and analyze the data, utilizing machine learning algorithms for activity recognition.

A. APPLICATION MODEL

Figure 1 illustrates the application model of our HAR system. We design eight types of daily activities, such as walking, sitting, squatting, and typing on a keyboard. Testers wear IMU devices on their right wrist, right side of the waist, and left ankle, collecting data from accelerometers and gyroscopes. Simultaneously, Wi-Fi transceivers located at both ends of the room transmit signals, which are reflected by various human actions and received by Rx. The collected IMU and CSI data are then synchronized and denoised. Features are extracted from the denoised IMU and CSI data, and six machine learning algorithms are used to recognize activities in three HAR systems: CSI, IMU, and CSI+IMU.

B. SYSTEM STRUCTURE

The system architecture is shown in Figure 2. We collect IMU data using the Wonder Sense-WiFi (WS-WiFi) [41] developed in our laboratory and use two computers equipped with Intel 5300 network cards to transmit and receive Wi-Fi signals. The processing flow in this system follows these steps:

Step 1: Wi-Fi signals are transmitted from the CSI transmitter to the CSI receiver. The transmitter is equipped with 1 antenna, while the receiver has 3 antennas. The sampling rate for CSI is set at 200Hz.
Step 2: IMU data are transmitted from Wonder Sense to the IMU Hub. Three IMU sensors transmit acceleration and gyroscope data along the x, y, and z axes. The sampling rate for these sensors is set at 200Hz.
Step 3: IMU and CSI data are synchronized according to Japan Standard Time (JST).
Step 4: Utilize moving average filtering for denoising before processing the raw data.
Step 5: Calculate features.
Step 6: Test with 6 machine learning algorithms: SVM, Multi-layer Perceptron, Decision Tree, Random Forest, Logistic Regression, kNN.
Step 7: Save the model.
Step 8: Calculate the validation score.
Step 9: Test using the saved model.

IV. METHOD

We collect accelerometer and gyroscope data using three WS-WiFi devices, while simultaneously gathering CSI data with a pair of Wi-Fi devices. However, this data cannot be used directly, we need to time synchronize and denoise it first. Then, we extract features from the denoised data. Our system includes three sensor combinations: CSI, IMU, and CSI+IMU. Based on these features, we use six different machine learning algorithms to classify eight types of everyday activities for each sensor combination.

A. DATA COLLECTION

1) CSI DATA

In a narrow-band flat fading channel, the Wi-Fi OFDM system viewed in the frequency domain can be defined as [43]:

$$\vec{Y} = H\vec{X} + \vec{N} \tag{1}$$

where $\vec{Y}$ and $\vec{X}$ represent the received and transmitted signal vectors, respectively. H denotes the Channel Frequency Response (CFR) and $\vec{N}$ is the Additive White Gaussian Noise (AWGN). The CFR in time and frequency as amplitude and phase in the format of CSI, is a superposition of signals from multipath propagation. Hence, the CSI can be represented as [43]:

$$H(f_s, t) = \sum_{k=1}^{L} A_{s,k} e^{-j2\pi f_n \tau_k(t)} \tag{2}$$

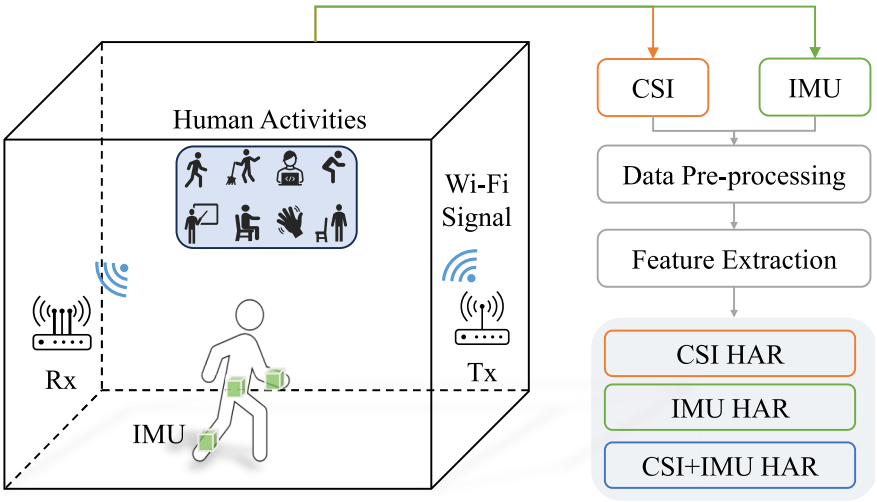


FIGURE 1. Application Model. We design eight types of daily activities, such as walking, sitting, squatting, and typing on a keyboard. Testers wear IMU devices on their right wrist, right side of the waist, and left ankle, collecting data from accelerometers and gyroscopes. Simultaneously, Wi-Fi transceivers located at both ends of the room transmit signals, which are reflected by various human actions and received by Rx. The collected IMU and CSI data are then synchronized and denoised. Features are extracted from the denoised IMU and CSI data, and six machine learning algorithms are used to recognize activities in three HAR systems: CSI, IMU, and CSI+IMU.

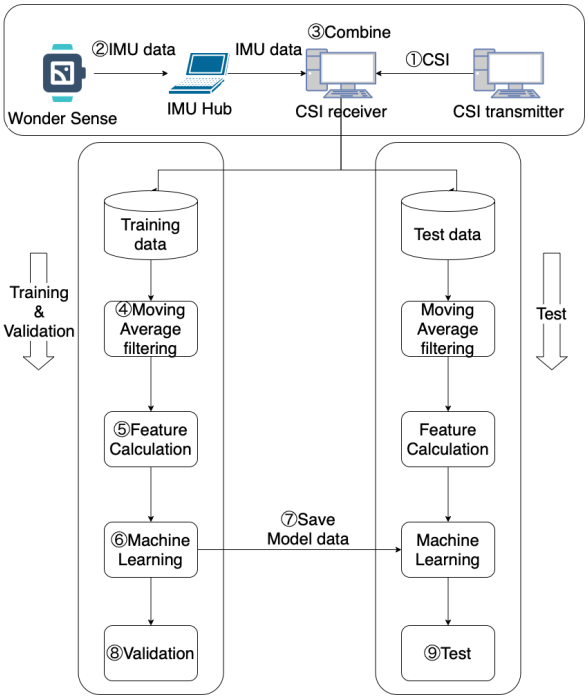


FIGURE 2. System Architecture.

where f_s is the carrier frequency of the s^{th} subcarrier, s is the index of the OFDM subcarrier, $s \in [1, 30]$. L is the number of paths, $A_{s,k}$ denotes the amplitude and $\tau_k(t)$ represents the propagation time of the k^{th} path. Moreover, the phase of CSI at carrier frequency f_s propagating along the k^{th} path can be written as $\varphi_{s,k} = 2\pi f_s \tau_k(t)$, the $\tau_k(t)$ represents the signal propagation delay [43].

The amplitude and phase have 30 subcarriers each receiver antenna. It means one antenna receive 60 dimensions information. Therefore, in three antennas, there are 180 dimensions in CSI. The changes in amplitude and phase of CSI on three antennas during walking are shown in the Figure 3. In the figure, the X-axis represents the number of CSI packets, the Y-axis represents the subcarrier index, and the Z-axis represents the amplitude and phase values. It can be observed that the amplitude and phase of CSI change with the movement of a person walking.

2) IMU DATA

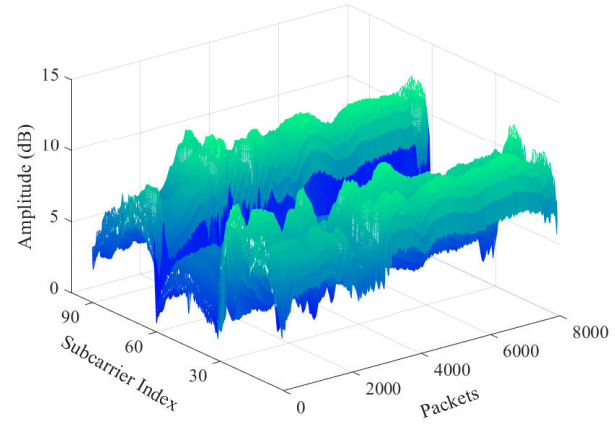
We utilize three WS-WiFi devices to obtain IMU data. Acceleration values in the x, y, and z directions are acquired from the triaxial accelerometer, while angular velocity values in the x, y, and z directions are obtained from the triaxial gyroscope. Therefore, we have 18 dimensions of IMU data. The data from six sensors when a person is waving is shown in the Figure 4. The figure shows data collected by six sensors placed on different parts of the body. As the person continues to wave, the accelerometers on the left ankle and waist show minimal changes, while the accelerometer on the right wrist exhibits significant variations. The data from all three gyroscopes also show notable changes.

B. DATA PRE-PROCESSING

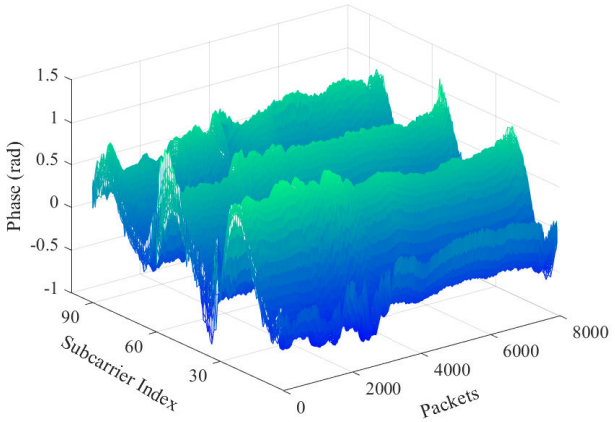
1) DATA SYNCHRONIZATION

Since the data is obtained from two types of sensors, WiFi and WS-WiFi, it is necessary to synchronize the data. Figure 5 shows the time synchronization process between each sensor data.

Step 1: Synchronize two PC time, IMU Hub and CSI receiver with JST.
Step 2: Save the starting time of CSI data acquisition time.
Step 3: Save the starting time of IMU data acquisition time.



(a) Amplitude of CSI



(b) Phase of CSI

FIGURE 3. Amplitude and Phase Data of Walking.

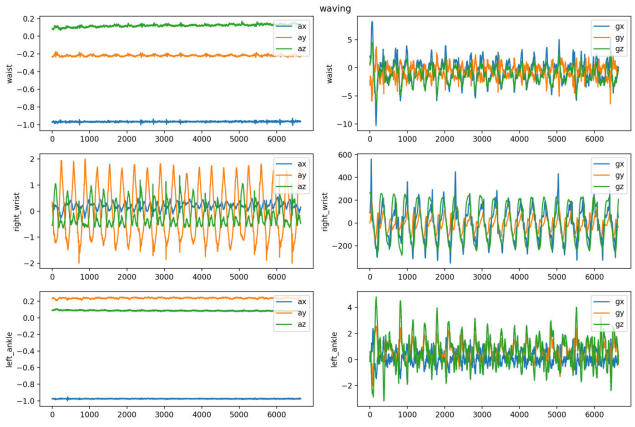


FIGURE 4. Accelerometers and Gyroscopes Data of Waving. The figures depict variations in accelerometer and gyroscope readings at three body locations during the waving action. The sensors have a sampling rate of 200Hz, with the x-axis representing packet numbers and the y-axis showing sensor values.

Step 4: Save the starting time of Camera shooting time.

Step 5: Synchronize each data by aligning each time using the three sensor time.

2) MOVING AVERAGE FILTERING

Moving average filtering filtering is one of the method of noise reduction. The noise of training data may affect to recognition rate. Machine learning models learn features related to the training data from noise reduced data. The MA formula is following,

$$y_i = \frac{1}{N} \sum_{j=i-(N-1)}^{i} x_j, \quad i \geq N \tag{3}$$

where N and i are the number of tap and the start of the data.

C. FEATURE DEFINITION

The range of each sensor data is different. For example, if the data range of CSI is 15 to 25, IMU acceleration range is -1.0 to 1.0, gyroscope range is -100 to 100, the difference in the range of data is too large. To fix the problem, this experiment standardizes sensors data. The standardization equation is

$$x_{std}^{(i)} = \frac{x^{(i)} - \mu}{\sigma} \tag{4}$$

where μ and σ are mean and standard deviation. The feature values are calculated by dividing the time-series data into the window size by the sliding window method. Let n, x_i be the window size, the time-series data in the i-th sampling point, respectively. In this paper, the feature values in the time domain are defined by the following formulas.

1) STANDARD DEVIATION (σ)

$$\sigma = \sqrt{\frac{1}{n} \sum_{x_i \in \text{window}} (x_i - \bar{x})^2} \tag{5}$$

2) MEAN ($\bar{X}$)

$$\bar{x} = \frac{1}{n} \sum_{x_i \in \text{window}} x_i \tag{6}$$

3) VARIANCE (σ^2)

$$\sigma^2 = \frac{1}{n} \sum_{x_i \in \text{window}} (x_i - \bar{x})^2 \tag{7}$$

Then, we define the feature values in the frequency domain. The time-series data in the window is transformed by the Discrete Fourier Transform (DFT). The DFT is defined as

$$DFT[k] = \sum_{x_i \in \text{window}} e^{-2\pi j \frac{ki}{n}} x_i \tag{8}$$

The amplitudes of $DFT[k]$ are normalized by the following formula.

$$|DFT[k]| \leftarrow \frac{1}{\lfloor \frac{n}{2} \rfloor} |DFT[k]| \tag{9}$$

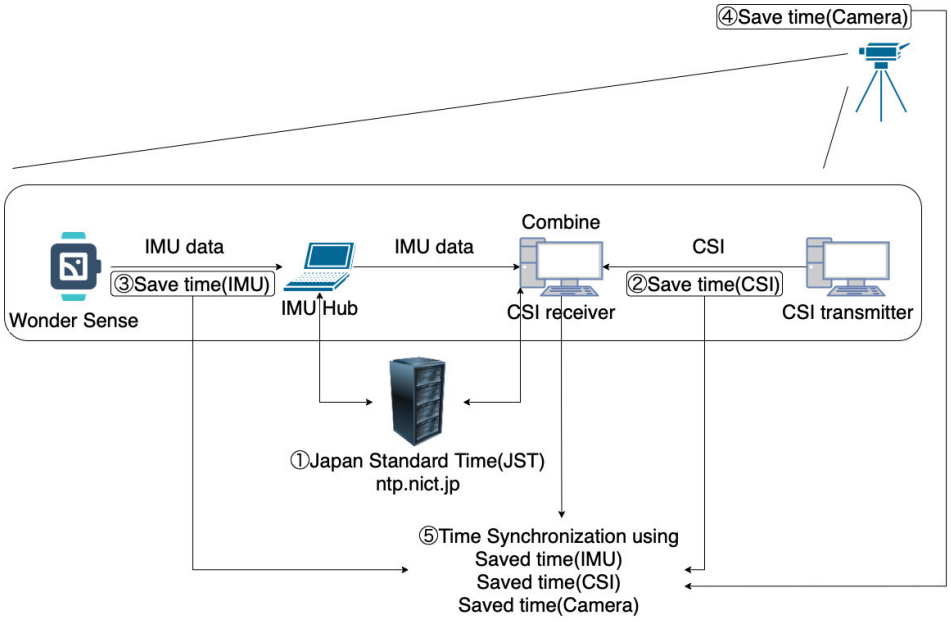


FIGURE 5. Time Synchronization Process.

The formula (9) means that $\frac{1}{\lfloor \frac{n}{2} \rfloor}|DFT[k]|$ is substituted into $|DFT[k]|$. $\lfloor \cdot \rfloor$ denotes the floor function. The feature values in the frequency domain are defined by the following formulas.

4) ENERGY (E)

$$E = \frac{1}{\lfloor \frac{n}{2} \rfloor} \sum_{k=0}^{\lfloor n/2 \rfloor - 1} |DFT[k]|^2 \tag{10}$$

5) ENTROPY (H)

$$\hat{P}(\omega_i) = \frac{1}{\lfloor \frac{n}{2} \rfloor}|DFT[i]|^2 \tag{11}$$

$$p_i = \frac{\hat{P}(\omega_i)}{\sum \hat{P}(\omega_i)} \tag{12}$$

$$H = -\sum_{i=0}^{\lfloor n/2 \rfloor - 1} p_i ln(p_i) \tag{13}$$

6) THE MEAN OF THE FREQUENCY-DOMAIN ($\overline{|DFT|}$)

$$\overline{|DFT|} = \frac{1}{\lfloor \frac{n}{2} \rfloor} \sum_{k=0}^{\lfloor n/2 \rfloor - 1} |DFT[k]| \tag{14}$$

7) THE VARIANCE OF THE FREQUENCY-DOMAIN ($(\sigma_F)^2$)

$$(\sigma_f)^2 = \frac{1}{\lfloor \frac{n}{2} \rfloor} \sum_{k=0}^{\lfloor n/2 \rfloor - 1} (|DFT[k]| - \overline{|DFT|})^2 \tag{15}$$

8) THE MAXIMUM VALUES OF THE FREQUENCY-DOMAIN (M_F)

$$m_f = max\ |DFT(k)| \tag{16}$$

D. DATA SPLITTING OF TRAINING, VALIDATION AND TEST

From the data collected through CSI and IMU, we use 80% as training data and 20% as test data. The machine learning models are trained using 5-fold cross-validation.

V. EXPERIMENTAL

A. EXPERIMENTAL PURPOSE

In existing works, the fusion of CSI and IMU data for HAR has been relatively limited. Previous studies have either focused on recognizing specific movements in sports like tennis or identified a few isolated full-body activities. Consequently, our research sets out to conduct an experimental comparison of HAR using three distinct sensor combinations: CSI, IMU, and CSI+IMU. We aim to evaluate the accuracy of each sensor combination by recognizing eight daily activities. Moreover, we utilize six different machine learning algorithms with various features and feature combinations to implement HAR. This involves testing the performance of different sensor combinations, machine learning algorithms, and feature sets. This comprehensive experimental effort is designed to enhance the understanding and application of HAR in various contexts, serving as a foundational guide for diverse HAR applications.

B. EXPERIMENTAL DESIGN

Figure 6 is the system structure in this experiment. The explanation of steps of Figure 6 are following:

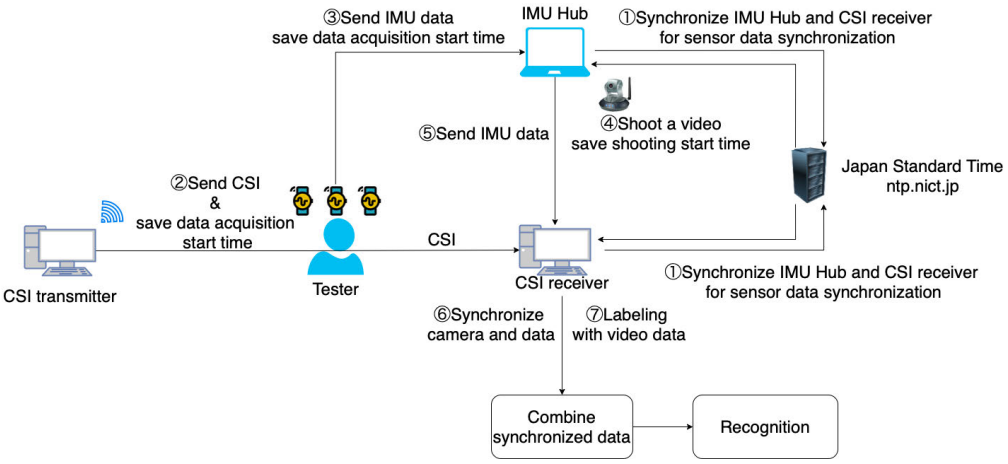


FIGURE 6. Experimental System Structure.

1) Synchronize IMU Hub and CSI receiver
   - Get JST from ntp.nict.jp by using NTP.
   - Synchronize IMU Hub clock and CSI receiver clock with JST.
2) Send CSI and save data acquisition start time
   - Send CSI packets and save the data acquisition start time for synchronization. CSI receiver receives CSI packets.
3) Send IMU data and save data acquisition start time
   - Send IMU data and save the data acquisition start time for synchronization. IMU Hub receives IMU data.
4) Shoot a video and save shooting start time
   - Shoot a video and save shooting start time for synchronization.
5) Send IMU data from IMU Hub to CSI receiver
   - Send IMU data from IMU Hub to CSI receiver
6) Synchronize camera and data
   - Synchronize CSI, IMU and Camera for processing.
7) Labelling with video data
   - Teacher data labelling with video data.

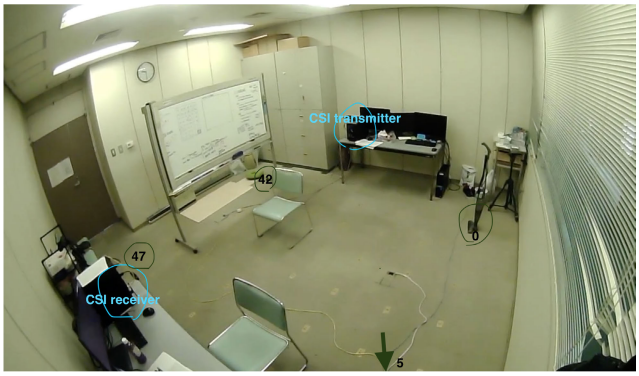


FIGURE 7. Photo of The Experiment Environment.

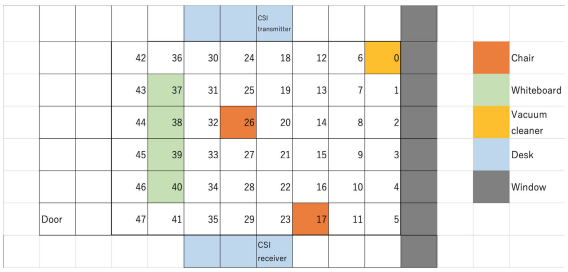


FIGURE 8. Grid Division of The Experiment Area.

C. IMPLEMENTATION

The experimental scenarios are illustrated in Figure 7 and Figure 8. Figure 7 shows a real photograph of the indoor environment, while Figure 8 presents our grid division of the experimental area. In our experimental setup, each tile in the laboratory is assigned a number, with each tile measuring 50cm × 50cm. There are two red tiles, number 17 and 26, in the figure which show chair on the tiles. A whiteboard on green tile number 37 to 40. There is a vacuum cleaner on orange tile number 0.

We utilize two computers equipped with Intel 5300 network cards as the transmitting and receiving devices for Wi-Fi. The transmitter is equipped with one antenna, while the receiver has three antennas forming an antenna array. Our computers operate on the Ubuntu 18.04 system, using the CSI Tool [35] to parse the signals and obtain CSI. The sampling rate of CSI is 200Hz.

We employ the WS-WiFi wearable motion capture sensor node described in [41]. This system comprises IMU sensors that relay acceleration and gyroscope data along the x, y, and z axes to an IMU Hub, a MacBook Pro (Retina, 15-inch, Mid 2015). For the experiment, three IMU sensors are attached to the right waist, right wrist, and the tip of the left foot, each with a sampling rate of 200Hz.

D. DATASET

We design eight activities to comprehensively describe daily life activities, as shown in Table 2. The activities are divided

TABLE 2. Human daily activities.

Whole body	Upper half of body
Squat	
Walking	Writing on a Whiteboard
Move Heavy Object	Typing on a Keyboard
Clean with Vacuum	Hand-shaking
Sitting-down	

into two categories: whole body activities and upper body activities. The former includes squatting, walking, moving heavy objects, cleaning with a vacuum cleaner, and sitting down. The latter encompasses writing on a whiteboard, typing on a keyboard, and hand-shaking.

We invite five participants to perform each of the eight activities listed in Table 2 within a 3m × 3m rectangular area, with vertices marked as 0, 5, 30, 35, as shown in Figure 8. Each activity is performed for a duration of five minutes. When collecting each set of data, we first start the CSI receiving program, followed by the IMU receiving program. At this stage, CSI and IMU are out of sync. However, during data preprocessing, we synchronize the CSI and IMU data in terms of JST.

Specifically, when collecting data for the sitting-down activity, we use a camera to record the movement, as sitting down is not a continuous action. Unlike other activities, sitting-down cannot be repeatedly performed in a sequence, as it inherently involves alternating between standing up and sitting down. The sequence involves standing up first, sitting down second, and then standing up again, with the sitting-down data captured between the first and third standing-up actions. Therefore, we only film this activity during data collection. For activities other than sitting down, we remove the first 10 seconds of the collected 5-minute data. From the remaining 4 minutes and 50 seconds, we extract the last 4 minutes as training data and the first 1 minute as test data.

The specific implementation of each activity is as follows:

1) Squat: squat on a tile which tile number is 14. The points are described in Figure 8.
2) Walking: walk a rectangle route which is a line connecting vertices 4, 1, 31, 34.
3) Move heavy object: repeat the sequence, move a chair from 17 to 12, and move from 12 to 17.
4) Clean with vacuum: clean the rectangle area with point 12, 16, 22, 18 as vertices.
5) Sitting-down: repeat sit down and stand up on 27 with a chair on point 26.
6) Writing on a whiteboard: write on a whiteboard while standing on 33.
7) Typing on a keyboard: sit down on 17 and type on a keyboard.
8) Hand shaking: standing on 14 and hand shake.

VI. EVALUATION

We utilize the collected data to validate the recognition accuracy of the HAR system. Initially, we verify the accuracy of recognizing eight types of activities by the CSI, IMU, and CSI+IMU HAR systems. Seven data features are processed through six machine learning algorithms to evaluate the performance of CSI, IMU, and CSI+IMU integrated systems. Concurrently, the performance enhancement from the fusion of CSI and IMU sensors is analyzed. The combined use of Energy and Mean FFT features quantifies each system's performance, aiding in comparative analysis of individual variability and the capacity to detect tiny activities. Additionally, computational time of each system is examined. Finally, we conduct a survey among the participants to gather their feedback on the use of different systems.

A. RECOGNITION RATE OF HAR WITH CSI, IMU AND CSI+IMU

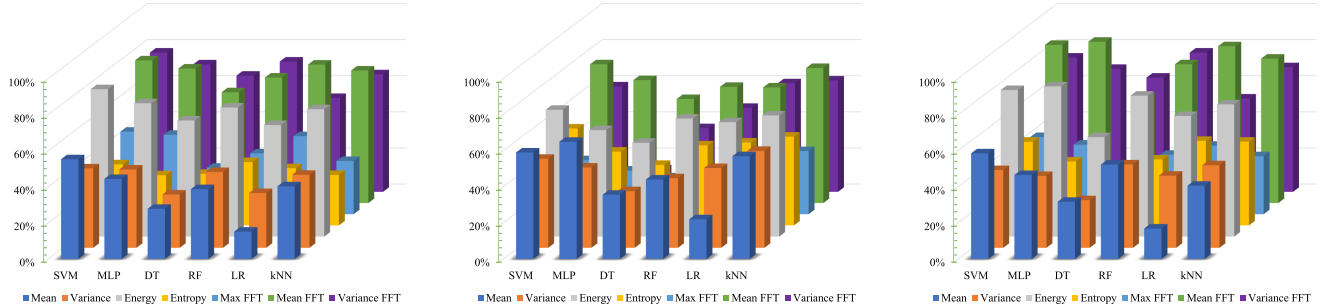
We utilize six machine learning algorithms and seven features to identify eight types of activities under three systems: CSI, IMU, and CSI+IMU. The recognition results are shown in Table 3, Table 4 and Table 5, respectively. To display the results more intuitively, we use Figure 9 to compare the recognition results of different features, algorithms, and sensor combinations.

From Table 3, it is observed that the SVM and random forest algorithms exhibit relatively high performance in CSI HAR system, achieving average recognition accuracies of 59.58% and 51.86% respectively. The Energy and Variance FFT features notably enhance the recognition rate, particularly in the SVM and random forest algorithms, where they contribute to recognition accuracies of 81.72% and 77.19% in SVM, respectively.

Table 4 reveals that the SVM and kNN algorithms demonstrate more effective performance in IMU HAR system, with average recognition accuracies of 56.83% and 56.96%, respectively. The Energy feature significantly influences the recognition rate, particularly in SVM and kNN, where it contributes to recognition accuracies of 70.31% and 67.19%, respectively.

Table 5 clearly illustrates that the CSI+IMU HAR system exhibits enhanced performance in the SVM and random forest algorithms, with average recognition accuracies of 62.03% and 57.19%, respectively. Notably, the Mean FFT and Energy features significantly enhance the system's effectiveness, especially in the SVM algorithm, where they lead to recognition accuracies of 87.66% and 81.25%, respectively.

Comparing the three systems, it is observed that the CSI+IMU system generally surpasses the individual CSI and IMU systems in performance, underscoring the effectiveness of sensor fusion in HAR. The SVM algorithm consistently shows superior performance across all three systems, indicating its suitability for HAR tasks involving these sensors. In terms of features, Energy and Mean FFT play a pivotal role in all systems, particularly in enhancing the recognition rates in the CSI+IMU system.



(a) CSI based HAR with one tester (b) IMU based HAR with one tester (c) CSI+IMU based HAR with one tester

FIGURE 9. Recognition Rates of the Three Sensor Combinations.

TABLE 3. Results of CSI-based HAR across various features and methods using single tester data.

	Mean	Variance	Energy	Entropy	Max FFT	Mean FFT	Variance FFT	Average
SVM	55.31%	44.06%	81.72%	33.91%	45.78%	79.06%	77.19%	59.58%
Multi-layer Perceptron	44.38%	43.44%	73.91%	27.81%	44.06%	74.53%	70.63%	54.11%
Decision Tree	27.81%	29.53%	64.38%	28.59%	25.78%	61.41%	64.38%	43.13%
Random Forest	38.75%	42.03%	71.56%	35.16%	33.75%	69.55%	72.19%	51.86%
Logistic Regression	15.17%	30.36%	61.88%	31.72%	43.28%	76.72%	52.03%	44.45%
kNN	40.31%	40.47%	70.63%	27.97%	29.53%	73.44%	65.16%	49.64%
Average	36.96%	38.32%	70.68%	30.86%	37.03%	72.45%	66.93%	

TABLE 4. Results of IMU-Based HAR across various features and methods using single tester data.

	Mean	Variance	Energy	Entropy	Max FFT	Mean FFT	Variance FFT	Average
SVM	59.06%	49.38%	70.31%	53.75%	30.00%	76.88%	58.44%	56.83%
Multi-layer Perceptron	65.00%	44.53%	59.06%	40.94%	24.22%	68.13%	12.50%	44.91%
Decision Tree	35.63%	31.41%	52.03%	33.59%	20.47%	57.66%	35.31%	38.01%
Random Forest	44.06%	38.59%	65.47%	44.38%	25.31%	64.38%	46.56%	46.96%
Logistic Regression	22.03%	44.22%	63.44%	45.94%	22.03%	64.06%	60.16%	45.98%
kNN	57.03%	53.75%	67.19%	49.22%	35.00%	74.84%	61.72%	56.96%
Average	47.14%	43.65%	62.92%	44.64%	26.17%	67.66%	45.78%	

TABLE 5. Results of CSI+IMU-based HAR across various features and methods using single tester data.

	Mean	Variance	Energy	Entropy	Max FFT	Mean FFT	Variance FFT	Average
SVM	58.59%	43.13%	81.25%	46.41%	42.81%	87.66%	74.38%	62.03%
Multi-layer Perceptron	46.56%	39.84%	83.28%	35.47%	38.59%	89.38%	68.13%	57.32%
Decision Tree	31.72%	26.41%	55.16%	29.38%	25.16%	58.59%	63.28%	41.39%
Random Forest	52.34%	46.25%	78.13%	36.56%	32.97%	76.88%	77.19%	57.19%
Logistic Regression	16.88%	40.00%	66.88%	46.88%	37.97%	86.88%	51.72%	49.60%
kNN	40.63%	45.78%	73.28%	46.41%	32.19%	80.00%	69.06%	55.34%
Average	41.12%	40.24%	73.00%	40.19%	34.95%	79.90%	67.29%	

TABLE 6. CSI and CSI+IMU system recognition accuracy improvement comparison.

	Mean	Variance	Energy	Entropy	Max FFT	Mean FFT	Variance FFT	Average
SVM	5.93%	-2.11%	-0.58%	36.86%	-6.49%	10.88%	-3.64%	5.84%
Multi-layer Perceptron	4.91%	-8.29%	12.68%	27.54%	-12.41%	19.92%	-3.54%	5.83%
Decision Tree	14.06%	-10.57%	-14.32%	2.76%	-2.40%	-4.59%	-1.71%	-2.40%
Random Forest	35.07%	10.04%	9.18%	3.98%	-2.31%	10.54%	6.93%	10.49%
Logistic Regression	11.27%	31.75%	8.08%	47.79%	-12.27%	13.24%	-0.60%	14.18%
kNN	0.79%	13.12%	3.75%	65.93%	9.01%	8.93%	5.99%	15.36%
Average	12.00%	5.66%	3.13%	30.81%	-4.48%	9.82%	0.57%	

The comparison between CSI and CSI+IMU systems, as depicted in Table 6, indicates that sensor fusion generally enhances recognition accuracy, especially in SVM and Random Forest algorithms, where the average improvement is 5.84% and 10.49%, respectively. The Mean FFT and Energy features significantly contribute to this increase, particularly in SVM, with accuracy improvements of 10.88% and 36.86%. Fusion seems most beneficial when complex movements are involved or when maintaining high recognition rates under varied environmental conditions.

Contrastingly, Table 7's comparison between IMU and CSI+IMU systems shows mixed results. While there is a notable performance improvement in SVM and Multi-layer

TABLE 7. IMU and CSI+IMU system recognition accuracy improvement comparison.

	Mean	Variance	Energy	Entropy	Max FFT	Mean FFT	Variance FFT	Average
SVM	-0.80%	-12.66%	15.56%	-13.66%	42.70%	14.02%	27.28%	10.35%
Multi-layer Perceptron	-28.37%	-10.53%	41.01%	-13.36%	59.33%	31.19%	445.04%	74.90%
Decision Tree	-10.97%	-15.92%	6.02%	-12.53%	22.91%	1.61%	79.21%	10.05%
Random Forest	18.79%	19.85%	19.34%	-17.62%	30.26%	19.42%	65.79%	22.26%
Logistic Regression	-23.38%	-9.54%	5.42%	2.05%	72.36%	35.62%	-14.03%	9.79%
kNN	-28.76%	-14.83%	9.06%	-5.71%	-8.03%	6.89%	11.89%	-4.21%
Average	-12.25%	-7.27%	16.07%	-10.14%	36.59%	18.13%	102.53%	

TABLE 8. Comparison of CSI, IMU, and CSI+IMU based HAR with one tester using energy and mean FFT features.

	CSI	IMU	CSI+IMU
SVM	79.84%	78.59%	87.50%
Multi-layer Perceptron	78.28%	67.19%	89.38%
Decision Tree	66.25%	56.88%	60.16%
Random Forest	70.16%	62.03%	78.13%
Logistic Regression	78.75%	64.53%	85.94%
kNN	72.81%	75.16%	79.06%
Average	74.35%	67.40%	80.03%

TABLE 9. Comparison of CSI, IMU, and CSI+IMU based HAR with five testers using energy and mean FFT Features.

	CSI	IMU	CSI+IMU
SVM	75.28%	62.38%	80.00%
Multi-layer Perceptron	74.59%	60.19%	79.13%
Decision Tree	45.94%	30.38%	40.41%
Random Forest	60.19%	43.81%	53.94%
Logistic Regression	55.78%	39.41%	64.66%
kNN	63.72%	55.16%	67.75%
Average	62.58%	48.56%	64.32%

Perceptron algorithms (10.35% and 74.90% respectively), a decrease in performance is observed in kNN by 4.21%. This suggests that sensor fusion might not always yield expected improvements, particularly with certain algorithms like kNN.

In conclusion, while sensor fusion in HAR systems can lead to significant performance enhancements, its effectiveness is highly dependent on the chosen algorithms and features. It is crucial to consider these factors when deciding on a fusion strategy, as it may not always result in performance gains, particularly in algorithms like kNN where individual sensor systems might suffice.

B. RECOGNITION RATE COMPARISON OF THREE TYPES OF SENSORS UNDER ENERGY + MEAN FFT FEATURE FUSION

The experimental results show that the Energy and Mean FFT features consistently provide superior performance in activity recognition tasks. Therefore, we fuse these two features to enable a comparative analysis of recognition accuracy among various algorithms and sensors. The experimental results are shown in Table 8.

The table shows that the SVM and multi-layer perceptron algorithms are highly effective in all systems, particularly in the CSI+IMU system where they reach the highest accuracies of 87.50% and 89.38% respectively. The CSI+IMU system stands out with an average accuracy of 80.03%, demonstrating that combining different sensors can greatly improve results. Merging the energy and mean FFT features also significantly boosts accuracy. These results underline the major benefits of using both feature and sensor fusion in HAR systems. Algorithms like SVM and multi-layer perceptron are really good at making sense of complex data from combined sensors.

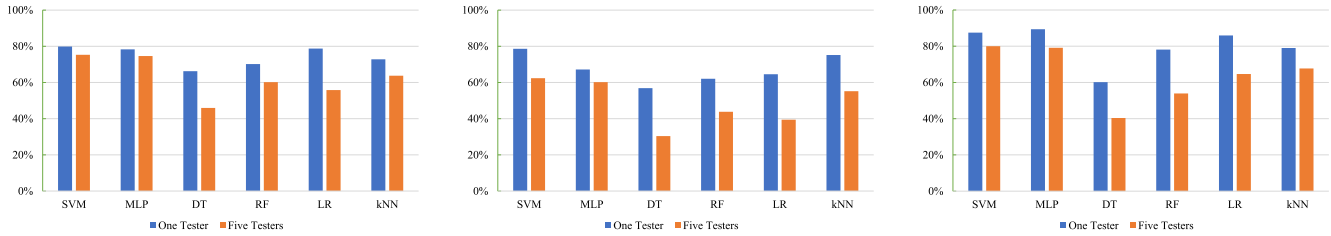
C. COMPARISON OF RECOGNITION RATES ACROSS THREE TYPES OF SENSORS FOR ONE TESTER AND FIVE TESTERS

In order to test the impact of individual variability on the recognition accuracy of CSI, IMU and CSI+IMU HAR system, we fed the data of 5 people together to the machine learning algorithms, and the results are shown in Table 9. In order to compare the recognition accuracy of one tester and five tester more intuitively, we use histograms to compare the results as shown in Figure 10.

In the one tester scenario, where each person's data is individually fed into the machine learning algorithms, the systems generally exhibit higher recognition accuracies. This is evident from the average accuracies of 74.35%, 67.40%, and 80.03% for the CSI, IMU, and CSI+IMU systems, respectively. However, in the five testers scenario, where data from all five individuals are combined and fed into the algorithms, a noticeable drop in accuracy is observed across all systems. This drop, from 74.35% to 62.58% for CSI, from 67.40% to 48.56% for IMU, and from 80.03% to 64.32% for CSI+IMU, underscores the challenge posed by individual differences. Despite this, the CSI+IMU system demonstrates a comparatively lesser decline in performance, suggesting its better handling of individual variability. This result shows that considering individual variability is important when designing HAR systems. It also shows that sensor fusion can help lessen the impact of such variability.

D. RECOGNITION RATE OF TINY ACTIVITIES WITH CSI, IMU AND CSI+IMU

In the context of this experiment, 'tiny activities' are defined as activities involving only hand movements, such as hand-shaking, typing on a keyboard, and writing on a whiteboard. In this experiment, we simultaneously feed the data collected



(a) CSI (b) IMU (c) CSI+IMU

FIGURE 10. Impact of Individual Variability on The Recognition Accuracy of CSI, IMU and CSI+IMU HAR System.

TABLE 10. Performance metrics of CSI-based HAR system for each class.

Activities	Accuracy	Precision	Recall	F1 Score
Clean-with-vacuum	72.50%	66.21%	72.50%	69.21%
Hand-shaking	75.75%	67.18%	75.75%	71.21%
Move-heavy-object	90.25%	84.94%	90.25%	87.52%
Sitting-down	34.75%	83.23%	34.75%	49.03%
Squat	77.75%	54.85%	77.75%	64.32%
Typing-on-a-keyboard	74.75%	94.92%	74.75%	83.64%
Walking	99.75%	98.03%	99.75%	98.88%
Writing-on-whiteboard	76.75%	71.40%	76.75%	73.98%

TABLE 11. Performance metrics of IMU-based HAR system for each class.

Activities	Accuracy	Precision	Recall	F1 Score
Clean-with-vacuum	67.50%	58.57%	67.50%	62.72%
Hand-shaking	64.25%	65.56%	64.25%	64.90%
Move-heavy-object	64.91%	65.74%	64.91%	65.32%
Sitting-down	83.25%	86.05%	83.25%	84.63%
Squat	45.75%	53.98%	45.75%	49.53%
Typing-on-a-keyboard	38.50%	44.51%	38.50%	41.29%
Walking	94.25%	65.45%	94.25%	77.25%
Writing-on-whiteboard	40.75%	53.62%	40.75%	46.31%

TABLE 12. Performance metrics of CSI+IMU-based HAR system for each class.

Activities	Accuracy	Precision	Recall	F1 Score
Clean-with-vacuum	84.25%	75.22%	84.25%	79.48%
Hand-shaking	67.00%	69.07%	67.00%	68.02%
Move-heavy-object	84.00%	80.77%	84.00%	82.35%
Sitting-down	68.50%	96.48%	68.50%	80.12%
Squat	77.25%	77.44%	77.25%	77.35%
Typing-on-a-keyboard	78.00%	96.89%	78.00%	86.43%
Walking	100.00%	93.68%	100.00%	96.74%
Writing-on-whiteboard	81.00%	62.79%	81.00%	70.74%

from five individuals into the system, employing a fusion of two key features: energy and mean FFT. The confusion matrix of the recognition results are shown in Figure 11. At the same time, we calculated the accuracy, precision, recall and F1 score of each system.

In Table 10, the 'Hand-shaking', 'Typing-on-a-keyboard', and 'Writing-on-whiteboard' activities in the CSI system demonstrate accuracies of 75.75%, 74.75%, and 76.75% respectively. These activities, particularly 'Typing-on-a-keyboard', showcase a moderate level of performance, suggesting reasonable effectiveness of the CSI system in detecting these subtle activities.

In the IMU system, the same activities exhibit accuracies of 64.25%, 38.50%, and 40.75%, as shown in Table 11. Notably, the performance for 'Typing-on-a-keyboard' and 'Writing-on-whiteboard' drops significantly, indicating challenges faced by the IMU system in accurately recognizing these finer activities.

In Table 12, the CSI+IMU system shows enhanced accuracies for these activities: 67.00% for 'Hand-shaking', 78.00% for 'Typing-on-a-keyboard', and 81.00% for 'Writing-on-whiteboard'. This improvement, especially in 'Typing-on-a-keyboard' and 'Writing-on-whiteboard', highlights the superior capability of the CSI+IMU system in detecting tiny activities.

Comparing the three systems, it is clear that the CSI+IMU system outperforms the CSI and IMU systems alone in identifying tiny activities. The integration of CSI and IMU sensors significantly enhances the ability to accurately detect subtle movements and actions, as shown by the improved accuracy of 'Typing-on-a-keyboard' and 'Writing-on-whiteboard' IMU systems, while effective for more obvious activity, show limitations in detecting finer movements, which are better captured by a combined CSI+IMU system. However, in terms of overall accuracy, there is room for enhancement in recognizing small movements across all three systems. This improvement could be achieved by increasing the number of IMUs, expanding the number of Wi-Fi device links, employing deep learning algorithms, and augmenting the sample size.

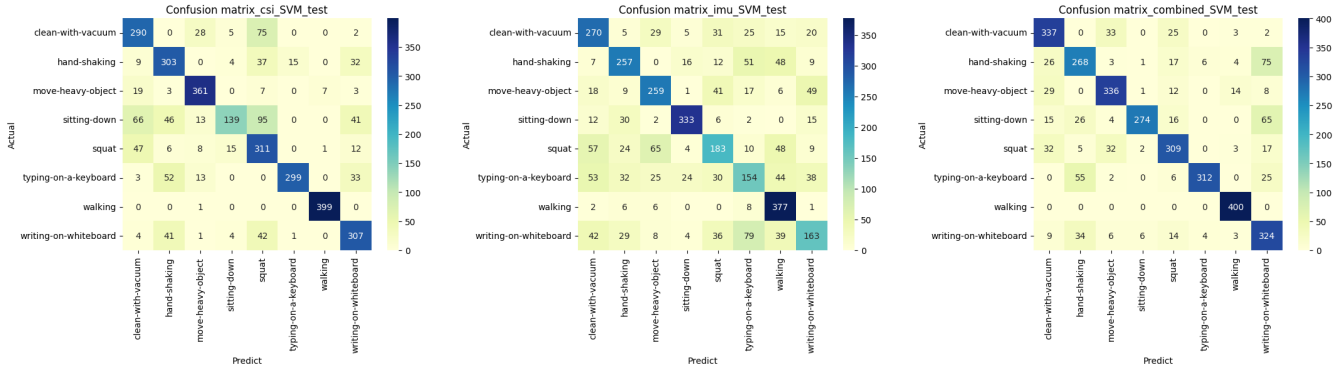
E. CALCULATION SPEED OF CSI, IMU AND CSI+IMU

For comparison of calculation speed of each sensor, this experiment measured time cost of training of machine learning models. There are 180 dimensions including 90-dimensional amplitude + 90-dimensional phase in CSI. In this measurement, use Energy and Mean FFT. Therefore, Energy:180 dimensions + Mean FFT:180 dimensions = 360 dimensions. IMU also use Energy and Mean FFT. Therefore, 18 dimensions + 18 dimensions = 36 dimensions. CSI+IMU dimensions is 360 + 36 = 396 dimensions.

Table 13 shows the time cost of each sensor. There is not much difference between CSI and CSI+IMU. Only IMU is much faster than the other two sensors. In multi-layer perceptron and kNN, amount of data increases, the amount of calculation becomes enormous.

F. A QUESTIONNAIRE ABOUT TROUBLESOME OF IMU

Figure 12 presents the layout of the questionnaire. It includes sections assessing physical effort, frustration

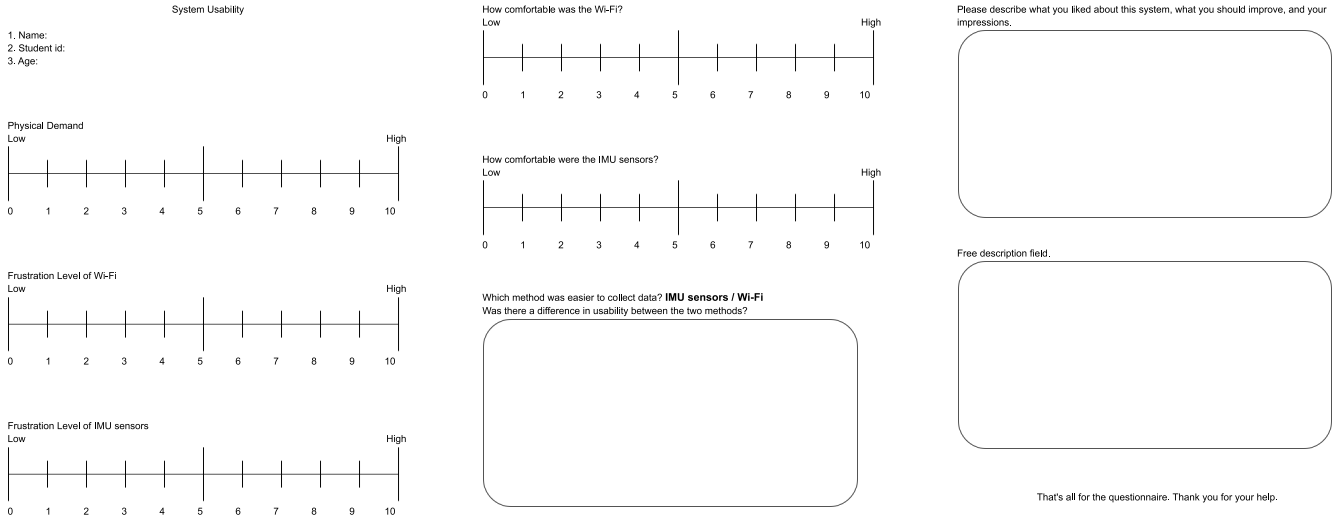


(a) CSI (b) IMU (c) CSI+IMU

FIGURE 11. Confusion Matrices of SVM Classifier Using Data from Five Testers.

TABLE 13. The comparison of calculation speed of each sensor.

	CSI(360 dimensions)	IMU(36 dimensions)	CSI+IMU(396 dimensions)
SVM	3576.20s	559.65s	3935.43s
Multi-layer perceptron	8653.77s	4896.55s	6890.62s
Decision tree	39.53s	4.92s	39.11s
Random forest	111.35s	36.33s	114.80s
Logistic regression	51.50s	18.26s	54.09s
kNN	8024.20s	1214.29s	8677.85s
Average	3409.42s	1121.67s	3285.32s



(a) The questionnaire page 1 (b) The questionnaire page 2 (c) The questionnaire page 3

FIGURE 12. The questionnaire form.

TABLE 14. The questionnaire about this experiment.

	Physical demand	Frustration level of Wi-Fi	Frustration level of IMU	Wi-Fi ease of use	IMU ease of use
Tester 1	7	0	0	10	7
Tester 2	5	4	4	9	6
Tester 3	8	0	8	10	7
Tester 4	7	0	3	10	5
Average	6.75	1	3.75	9.75	6.25

levels associated with Wi-Fi and IMU usage, ease of use for both Wi-Fi and IMU, along with three fields for additional descriptions. Table 14 shows the result of questionnaire of this experiment. About physical effort and frustration level,

lower is better. About Wi-Fi and IMU ease of use, higher is better. The lowest is zero and the highest is ten. Therefore, the evaluation of usability is 11-grade evaluation.

Notably, the data indicates that the physical effort required during the data collection step was relatively high, as evidenced by an average score of 6.75. This suggests that participants experienced a notable level of exertion while engaging in the experimental tasks. In contrast, the frustration levels for Wi-Fi were remarkably low, averaging at 1, indicating a smooth and user-friendly experience. However, the frustration level for the IMU was higher, averaging at 3.75, which could imply some usability challenges with this device.

Regarding the ease of use, Wi-Fi was rated highly, with an average score of 9.75, reflecting its user-friendly interface and functionality in the context of this experiment. The IMU's ease of use, with an average score of 6.25, suggests that there are opportunities for improvement to enhance its intuitiveness and user experience.

Overall, these results point towards a favorable experience with Wi-Fi in terms of usability and a moderate level of physical effort required in interacting with the IMU. The findings underscore the need for further enhancements in IMU's usability to reduce physical effort and improve the overall user experience.

VII. CONCLUSION

In this paper, we propose a HAR system that uses CSI, IMU, and CSI+IMU to identify eight daily activities. Our experimental results demonstrate that the CSI+IMU system outperforms the independent CSI and IMU systems, showing the effectiveness of sensor fusion in improving recognition accuracy. The SVM algorithm consistently performs well across all systems, especially excelling in the CSI+IMU system supported by energy and average FFT features. However, it is important to note that the success of sensor fusion depends on specific algorithms and features. Fusion of CSI and IMU does not universally enhance recognition accuracy for all features and algorithms and can, in some cases, actually reduce accuracy. In experiments considering individual differences, the CSI+IMU system exhibits better performance.

Although the overall performance of the CSI+IMU is superior to that of the standalone CSI and IMU HAR systems, the CSI+IMU system still reveals room for improvement in identifying small activities. Enhancements can be achieved by increasing the number of IMUs, expanding Wi-Fi connectivity, employing advanced algorithms, and enlarging the sample size.

Regarding calculation speed, there is no significant difference between CSI and CSI+IMU. Only the IMU system is much faster than the others. If real-time processing is required, it is appropriate to reduce dimensions using methods like PCA or to solely use IMU sensors. Future work will focus on computational speed, feature selection, and dimensionality reduction applied to sensor data.

Additionally, our user experience analysis indicates that the availability of Wi-Fi and manageable physical interaction with the IMU is well received. These findings highlight the need for further enhancements in the usability of IMUs to improve the overall user experience in HAR applications.

ACKNOWLEDGMENT

Special thanks are extended to Yuto Shoji for his invaluable assistance in experimental design, data collection, and analysis. His contributions are vital to this research.

REFERENCES

[1] J. M. Chaquet, E. J. Carmona, and A. Fernández-Caballero, "A survey of video datasets for human action and activity recognition," *Comput. Vis. Image Understand.*, vol. 117, no. 6, pp. 633–659, Jun. 2013, doi: 10.1016/j.cviu.2013.01.013.

[2] I. Alrashdi, M. H. Siddiqi, Y. Alhwaiti, M. Alruwaili, and M. Azad, "Maximum entropy Markov model for human activity recognition using depth camera," *IEEE Access*, vol. 9, pp. 160635–160645, 2021.

[3] X. Yang and Y. Tian, "Super normal vector for human activity recognition with depth cameras," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 5, pp. 1028–1039, May 2017.

[4] N. Dua, S. N. Singh, and V. B. Semwal, "Multi-input CNN-GRU based human activity recognition using wearable sensors," *Computing*, vol. 103, no. 7, pp. 1461–1478, Jul. 2021.

[5] O. D. Lara and M. A. Labrador, "A survey on human activity recognition using wearable sensors," *IEEE Commun. Surveys Tuts.*, vol. 15, no. 3, pp. 1192–1209, 3rd Quart., 2013.

[6] E. Ramanujam, T. Perumal, and S. Padmavathi, "Human activity recognition with smartphone and wearable sensors using deep learning techniques: A review," *IEEE Sensors J.*, vol. 21, no. 12, pp. 13029–13040, Jun. 2021.

[7] W. Gomaa and M. A. Khamis, "A perspective on human activity recognition from inertial motion data," *Neural Comput. Appl.*, vol. 35, no. 28, pp. 20463–20568, Oct. 2023. [Online]. Available: https://api.semanticscholar.org/CorpusID:260408553

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

[9] W. Wang, A. X. Liu, M. Shahzad, K. Ling, and S. Lu, "Device-free human activity recognition using commercial WiFi devices," *IEEE J. Sel. Areas Commun.*, vol. 35, no. 5, pp. 1118–1131, May 2017.

[10] J. Liu, H. Liu, Y. Chen, Y. Wang, and C. Wang, "Wireless sensing for human activity: A survey," *IEEE Commun. Surveys Tuts.*, vol. 22, no. 3, pp. 1629–1645, 3rd Quart., 2020.

[11] F. Foerster, M. Smeja, and J. Fahrenberg, "Detection of posture and motion by accelerometry: A validation study in ambulatory monitoring," *Comput. Hum. Behav.*, vol. 15, no. 5, pp. 571–583, Sep. 1999.

[12] S. Pancholi and R. Agarwal, "Development of low cost EMG data acquisition system for arm activities recognition," in *Proc. Int. Conf. Adv. Comput., Commun. Informat. (ICACCI)*, Sep. 2016, pp. 2465–2469.

[13] A. Alevizaki, N. Pham, and N. Trigoni, "Invited paper: Hierarchical activity recognition with smartwatch IMU," in *Proc. 24th Int. Conf. Distrib. Comput. Netw.*, Jan. 2023, pp. 48–57.

[14] L. Jing, Y. Zhou, Z. Cheng, and T. Huang, "Magic ring: A finger-worn device for multiple appliances control using static finger gestures," *Sensors*, vol. 12, no. 5, pp. 5775–5790, May 2012.

[15] R. Bodor, B. Jackson, and N. Papanikolopoulos, "Vision-based human tracking and activity recognition," in *Proc. 11th Medit. Conf. Control Automat.*, vol. 1. Citeseer, 2003, pp. 1–6.

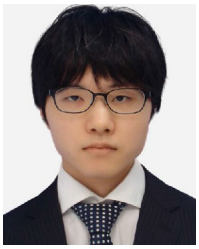
[16] C. Zhang and Y. Tian, "RGB-D camera-based daily living activity recognition," *J. Comput. Vis. Image Process.*, vol. 2, no. 4, p. 12, 2012.

[17] J. Nan, G. Wei, Y. Yuling, and W. Lin, "Leveraging TPDV to direction estimation with a single pair of commodity devices," *IEEE Access*, vol. 8, pp. 62248–62260, 2020.

[18] A. G. Putrada, M. Abdurohman, D. Perdana, and H. H. Nuha, "Machine learning methods in smart lighting toward achieving user comfort: A survey," *IEEE Access*, vol. 10, pp. 45137–45178, 2022.

[19] Y. Xie, R. Jiang, X. Guo, Y. Wang, J. Cheng, and Y. Chen, "Universal targeted attacks against mmWave-based human activity recognition system," in *Proc. 20th Annu. Int. Conf. Mobile Syst., Appl. Services*, 2022, pp. 541–542.
[20] Z. Chen, C. Cai, T. Zheng, J. Luo, J. Xiong, and X. Wang, "RF-based human activity recognition using signal adapted convolutional neural network," *IEEE Trans. Mobile Comput.*, vol. 22, no. 1, pp. 487–499, Jan. 2023.
[21] W. Wang, A. X. Liu, M. Shahzad, K. Ling, and S. Lu, "Understanding and modeling of WiFi signal based human activity recognition," in *Proc. 21st Annu. Int. Conf. Mobile Comput. Netw.*, Sep. 2015, pp. 65–76.
[22] Q. Pu, S. Gupta, S. Gollakota, and S. Patel, "Whole-home gesture recognition using wireless signals," in *Proc. 19th Annu. Int. Conf. Mobile Comput. Netw. (MobiCom)*, 2013, pp. 27–38.
[23] F. Adib and D. Katabi, "See through walls with WiFi!" in *Proc. ACM SIGCOMM Conf. (SIGCOMM)*, Aug. 2013, pp. 75–86.
[24] Y. Wang, J. Liu, Y. Chen, M. Gruteser, J. Yang, and H. Liu, "E-eyes: Device-free location-oriented activity identification using fine-grained WiFi signatures," in *Proc. 20th Annu. Int. Conf. Mobile Comput. Netw.*, Sep. 2014, pp. 617–628.
[25] Y. Zheng, Y. Zhang, K. Qian, G. Zhang, Y. Liu, C. Wu, and Z. Yang, "Zero-effort cross-domain gesture recognition with Wi-Fi," in *Proc. 17th Annu. Int. Conf. Mobile Syst., Appl., Services*, Jun. 2019, pp. 313–325.
[26] R. Gao, "Towards position-independent sensing for gesture recognition with Wi-Fi," *Proc. ACM Interact., Mobile, Wearable Ubiquitous Technol.*, vol. 5, no. 2, pp. 1–28, Jun. 2021.
[27] S. Zhou, L. Guo, Z. Lu, X. Wen, and Z. Han, "Wi-monitor: Daily activity monitoring using commodity Wi-Fi," *IEEE Internet Things J.*, vol. 10, no. 2, pp. 1588–1604, Jan. 2023.
[28] A. Sharshar, A. H. A. Eitta, A. Fayez, M. A. Khamis, A. B. Zaky, and W. Gomaa, "Camera coach: Activity recognition and assessment using thermal and RGB videos," in *Proc. Int. Joint Conf. Neural Netw. (IJCNN)*, Jun. 2023, pp. 1–8. [Online]. Available: https://api.semanticscholar.org/CorpusID:260387310
[29] Y. Shu, C. Chen, K.-I. Shu, and H. Zhang, "Research on human motion recognition based on Wi-Fi and inertial sensor signal fusion," in *Proc. IEEE SmartWorld, Ubiquitous Intell. Comput., Adv. Trusted Comput., Scalable Comput. Commun., Cloud Big Data Comput., Internet People Smart City Innov. (SmartWorld/SCALCOM/UIC/ATC/CBDCom/IOP/SCI)*, Oct. 2018, pp. 496–504.
[30] R. Ramezani, Y. Xiao, and A. Naeim, "Sensing-fi: Wi-Fi CSI and accelerometer fusion system for fall detection," in *Proc. IEEE EMBS Int. Conf. Biomed. Health Informat. (BHI)*, Mar. 2018, pp. 402–405.
[31] M. Muaaz, A. Chelli, A. A. Abdelgawwad, A. C. Mallofré, and M. Pätzold, "WiWeHAR: Multimodal human activity recognition using Wi-Fi and wearable sensing modalities," *IEEE Access*, vol. 8, pp. 164453–164470, 2020.
[32] T. Starner, J. Weaver, and A. Pentland, "Real-time American sign language recognition using Des. and wearable computer based video," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 20, no. 12, pp. 1371–1375, Dec. 1998.
[33] W. Taylor, S. A. Shah, K. Dashtipour, A. Zahid, Q. H. Abbasi, and M. A. Imran, "An intelligent non-invasive real-time human activity recognition system for next-generation healthcare," *Sensors*, vol. 20, no. 9, p. 2653, 2020.
[34] Z. Shi, J. A. Zhang, R. Xu, Q. Cheng, and A. Pearce, "Towards environment-independent human activity recognition using deep learning and enhanced CSI," in *Proc. IEEE Global Commun. Conf.*, Dec. 2020, pp. 1–6.
[35] D. Halperin, W. Hu, A. Sheth, and D. Wetherall, "Tool release: Gathering 802.11 n traces with channel state information," *ACM SIGCOMM Comput. Commun. Rev.*, vol. 41, no. 1, p. 53, Jan. 2011.
[36] O. Vinyals, C. Blundell, T. Lillicrap, and D. Wierstra, "Matching networks for one shot learning," in *Proc. Adv. Neural Inf. Process. Syst. (NIPS)*, 2016, pp. 3630–3638.
[37] W. Jiang, "Towards environment independent device free human activity recognition," in *Proc. 24th Annu. Int. Conf. Mobile Comput. Netw.*, 2018, pp. 289–304.
[38] J. Wang, V. W. Zheng, Y. Chen, and M. Huang, "Deep transfer learning for cross-domain activity recognition," in *Proc. 3rd Int. Conf. Crowd Sci. Eng.*, Jul. 2018, pp. 1–8.
[39] T. Li, H. Wang, Y. Shao, and Q. Niu, "Channel state information–based multi-level fingerprinting for indoor localization with deep learning," *Int. J. Distrib. Sensor Netw.*, vol. 14, no. 10, Oct. 2018, Art. no. 155014771880671.
[40] N. Ravi, N. Dandekar, P. Mysore, and M. L. Littman, "Activity recognition from accelerometer data," in *Proc. 20th Nat. Conf. Artif. Intell. 17th Innov. Appl. Artif. Intell. Conf.*, 2005, pp. 1541–1546.
[41] L. Jing and Z. Cheng, "Recognition of daily routines and accidental event with multipoint wearable inertial sensing for seniors home care," in *Proc. IEEE Int. Conf. Syst. Man, Cybern. (SMC)*, Oct. 2017, pp. 2324–2389.
[42] A. R. Jimenez, F. Seco, C. Prieto, and J. Guevara, "A comparison of pedestrian dead-reckoning algorithms using a low-cost MEMS IMU," in *Proc. IEEE Int. Symp. Intell. Signal Process.*, Aug. 2009, pp. 37–42.
[43] W. Guo and L. Jing, "Toward low-cost passive motion tracking with one pair of commodity Wi-Fi devices," *IEEE J. Indoor Seamless Positioning Navigat.*, vol. 1, pp. 39–52, 2023.

WEI GUO (Graduate Student Member, IEEE) received the B.S. degree in communication engineering and the M.S. degree in electronic science and technology from Yanshan University, in 2016 and 2019, respectively. She is currently pursuing the Ph.D. degree with the Graduate School of Computer Science and Engineering, The University of Aizu, Japan. Her research interests include wireless sensing systems, signal processing, and sensor fusion for mobile computing.



SHUNSEI YAMAGISHI received the B.S. degree in computer science and engineering from The University of Aizu, Japan, in 2022, where he is currently pursuing the M.S. degree with the Graduate School of Computer Science and Engineering. His research interests include the algorithms of attitude estimation for the inertial measurement units, Pedestian dead reckoning, and sensor fusion methods.

LEI JING (Member, IEEE) received the B.A. degree in mechatronics from the Dalian University of Technology, China, and the Ph.D. degree from the Graduate School of Computer Science, The University of Aizu, Japan. He is a Senior Associate Professor with the Department of Computer Science and Engineering, The University of Aizu. His research goal is to seamlessly connect human actions with the ubiquitous networks. He is particularly interested in developing novel motion capture methods and systems by applying multiple discipline knowledge, including intelligent sensing, signal processing, wireless sensor networks, artificial intelligence, electronics, and biomechanics. He is a member of ACM, IPSJ, and IEICE.