An Adaptive Step Detection Algorithm for Smartwatch with Deep Learning-based Human Activity Recognition

Sohee Park
Dept. of Aerospace Engineering
Seoul National University
Seoul, Republic of Korea
annie4545@snu.ac.kr

Jae Hong Lee
Dept. of Aerospace Engineering
Seoul National University
Seoul, Republic of Korea
honglj@snu.ac.kr

Chan Gook Park
Dept. of Aerospace Engineering/ASRI
Seoul National University
Seoul, Republic of Korea
chanpark@snu.ac.kr

2024 14th International Conference on Indoor Positioning and Indoor Navigation (IPIN) | 979-8-3503-6640-2/24/$31.00 ©2024 IEEE | DOI: 10.1109/IPIN62893.2024.10786168

***Abstract*—As the use of smartwatches continues to grow, improving the accuracy of tracking for smartwatch users has become increasingly important. Smartwatches are worn on the wrist, which means the sensors may detect free and independent hand movements while walking. This creates challenges for pedestrian dead reckoning (PDR) using the inertial sensors embedded in smartwatches because the data may lose information on walking characteristics. Therefore, it is essential to identify the user's walking state first, whether moving or stationary, and then conduct precise step detection, which is a critical factor in reducing errors in PDR. In this study, we propose an adaptive step detection algorithm tailored to four categorized user activity states: walking, running, hand movements while standing, and hand movements while walking. We conducted human activity recognition (HAR) with these four categories to identify the user's state and decide the appropriate method for step detection. To enable implementation on mobile devices, we propose a lightweight deep learning-based HAR network utilizing a grouped convolutional layer. The proposed network reduces the number of trainable parameters by approximately one-third compared to conventional CNN structures while achieving 96.45% accuracy in activity classification. For adaptive step detection, we utilize gyroscope data for normal walking, accelerometer data for running, and walking frequency obtained via fast Fourier transform (FFT) from normal walking data for segments of walking interspersed with hand movements. By applying these tailored step detection methods to each walking scenario, we enhance step detection accuracy, thereby improving smartwatch user position and distance estimation in overall PDR.**

***Index Terms*—Human Activity Recognition, Smartwatch, Pedestrian Dead Reckoning, Step detection**

I. Introduction

With the evolution of sensors, wearable smart devices have become essential items for modern individuals. As smartphones have become an irreplaceable part of people's lives, other complementary wearable smart accessories, such as smartwatches, are becoming increasingly popular around the world. People wear smartwatches for a variety of reasons, but the main ones are navigation and location services, health and fitness tracking, and emergency alerts. Many people use smartwatches while exercising to monitor their calorie consumption, exercise intensity, and trajectory, expecting accurate feedback. Accurate exercise data and trajectory analysis are crucial in this context. To achieve this, it is essential to leverage the built-in inertial sensors of the smartwatch to identify the type of exercise and perform navigation precisely.

A representative method for estimating a pedestrian's position and walking distance using inertial measurement unit (IMU) data is pedestrian dead reckoning (PDR) [1], [2], [3], [4]. PDR consists of three stages: step detection, step length estimation, and heading estimation. To reduce errors in PDR, particularly for accurate tracking of distance, it is essential to first ensure precise step detection as a critical factor for successful step length estimation. Research on step detection has predominantly been conducted using smartphone-based approaches [4], [5], and similar techniques have also been applied to smartwatches [7].

However, smartwatches are worn on the wrist, where the arms have considerable freedom of movement, even when walking. This results in a wide range of data characteristics that differ significantly from those of smartphones. For instance, hand movements during walking can drastically alter the signal patterns compared to typical walking data, making traditional methods inadequate for accurate step detection and step length estimation. Additionally, since the arms can move independently of the legs, movements of the arms while stationary can be misinterpreted as walking in the data, leading to erroneous conclusions. Consequently, robust and accurate step detection algorithms specifically tailored for smartwatches are necessary.

In this paper, we propose an adaptive step detection algorithm for smartwatches using a lightweight deep learning-based human activity recognition (HAR) system. The main contributions of this study are as follows:

1) We propose a lightweight HAR algorithm capable of distinguishing between various activities, including walking, running, walking with hand movement, and standing with hand movement. This classification enables us to filter scenarios accurately for step detection.

This work was supported in part by Samsung Electronics Company Ltd., through the Mobile eXperience (MX) under Grant 0418-20230050.

2) We present an adaptive step detection method tailored to each identified activity. Specifically, we use angular velocity data for normal walking and acceleration data for running. Furthermore, the walking frequency obtained from the Fast Fourier Transform (FFT) of the previous normal walking segment before the interspersed period of hand movements is used to replace the period of hand movements.

This study aims to enhance the accuracy of fundamental step detection, thereby robustly estimating the user's location and movement states using a smartwatch.

The rest of the paper is organized as follows. Section II reviews related work on HAR and step detection, and section III describes the details of the proposed HAR and adaptive step detection algorithm. Section IV presents and analyzes the experimental results, and lastly, section V concludes the paper.

II. Related Works

A. Human Activity Recognition

Human activity recognition (HAR) is a system capable of automatically recognizing and categorizing human activities, motions, and gestures based on sensor or vision data. A typical HAR process consists of data acquisition, data preprocessing, feature extraction, feature selection, and classification. In conventional methods, hand-crafted features such as time-domain [8] and frequency-domain [9] features extracted from sensor raw data are used as inputs for machine learning classification algorithms such as k-NN [10], SVM [11], etc. While the conventional method is intuitive and computationally light, it has the disadvantage of relying on manually extracted feature vectors, which capture only superficial features based on prior engineering knowledge.

As an alternative, recent research has primarily focused on deep learning-based HAR. Deep learning-based methods automatically learn and extract features, allowing them to identify hidden characteristics, resulting in more robust and higher-performing HAR. However, due to the significantly increased computational costs compared to traditional methods, recent research is being conducted to optimize and lighten deep learning algorithms for application on mobile devices. Yi et al. [12] applied a pruned and quantized CNN model and reduced the size of the network by nearly five times while maintaining accuracy. In [13], the feature maps extracted from the previous layer are divided into several groups. One group is copied in its original form, while a simple linear transformation is applied to the others, resulting in a notable reduction in floating-point operations (FLOPs). In addition, various lightweight deep learning algorithms, such as MobileNet [14] and ShuffleNet [15], can be applied to HAR.

In this study, we generated the subsequent feature maps by concatenating two groups of feature maps, applying a copy operation and a 1x1 convolutional filter, similar to the approach in [13].

B. Step Detection

Step detection is important in PDR because false or missed step detection can cause substantial errors in estimating the total walking distance. Researchers are utilizing accelerometers or gyroscopes for pedestrian step detection. In [2], researchers projected the smartphone's accelerometer data to a global coordinate system, filtered out gravity and noise, and applied a peak detection algorithm with specific thresholds to identify valid step events. In [3], a relative threshold detection scheme was utilized using a smartphone's accelerometer, identifying a step by sequentially detecting valid maxima and minima peaks within a specific interval and threshold. Ho et al. [4] applied a fast Fourier transform (FFT)-based smoother on acceleration data along with threshold methods.

Since walking motion is cyclic, gyroscopes can capture more robust information than accelerometers. Moreover, the frequency domain features are more intuitive for walking motions. Kang et al. [5] leveraged gyroscope and frequency domain features to identify and count steps irrespective of the smartphone's placement using FFT-based walking frequency estimation. In [6], angular velocity data filtered by a Butterworth low-pass filter from subjects walking at a normal pace on various terrains showed accurate gait event detection with 98% accuracy.

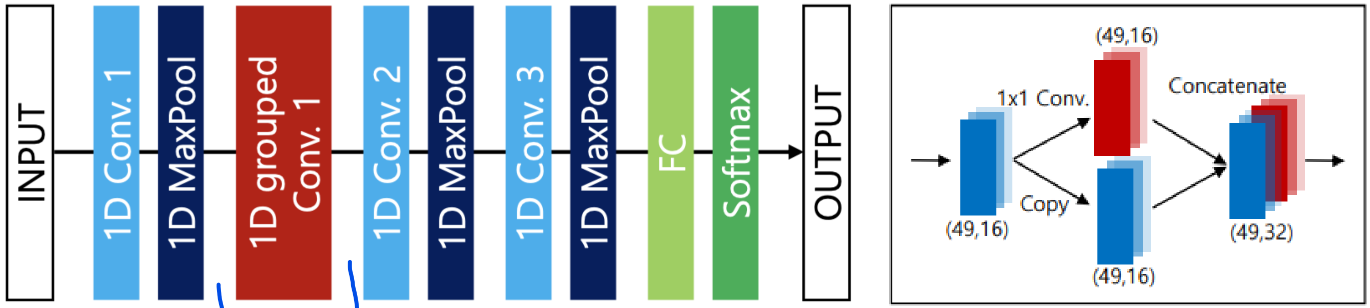
Previous studies have demonstrated the advantages of using angular velocity to analyze the periodic signals associated with walking. Consequently, we applied a low-pass filter to the angular velocity data and detected steps using a peak detection method. In addition, following the approach described in [5], we performed an FFT on the angular velocity data to determine the walking frequency.

III. Methodology

To distinguish the walking state of a smartwatch user and detect steps, data were acquired using the accelerometer and gyroscope embedded in the smartwatch. The smartwatch used in the experiment is a Samsung Galaxy Watch 5, with a sampling frequency of 100Hz.

A. Human Activity Recognition

a) Data collection: Data were collected from 10 subjects, 7 men and 3 women, across 4 activity categories. This study specifically focused on determining whether the user is walking. Therefore, experiments were conducted under four scenarios: walking, running, hand movements while standing, and hand movements while walking. To guide the subjects on hand movements, the following actions were defined: wiping sweat, checking the watch, and stretching—activities that typically occur while walking. Raw three-axis accelerometer and gyroscope data were segmented using a 1-second sliding window with a 0.5-second overlap. A total of 13,257 data samples, each with dimensions of (100, 6), were utilized in this study. The dataset was divided into training and testing sets with an 8:2 ratio.



(a) A proposed network with a grouped convolutional layer.

(b) A grouped convolutional layer structure.

Fig. 1: The proposed HAR network architecture.

b) Network architecture: To enable implementation on mobile devices such as smartwatches, we propose a lightweight CNN-based network. Inspired by the linear grouped convolution network [13], we incorporated a grouped convolutional layer into the network. This layer generates feature maps using partial group convolution and concatenation, rather than conventional convolutional layers.

Suppose the number of feature maps obtained from the previous convolutional layer is m. To obtain $2m$ feature maps from a subsequent conventional 1-d convolutional layer, $2m$ convolutional kernels are required. Let k be the convolutional kernel size, and let h and w denote the height and width of the previous feature maps, respectively. The computational cost of this layer is as follows.

$$O(m \times h \times w \times k \times 1 \times 2m) \tag{1}$$

However, if we employ the proposed method, which applies a copy operation and performs a 1x1 convolutional operation on the previous feature maps before concatenating them, the computational cost is as follows.

$$O(m \times h \times w \times 1 \times 1 \times m) \tag{2}$$

According to equations (1) and (2), the computational cost is reduced by $2k$, as when k goes to 1, $2m$ is reduced to m.

As shown in Fig. 1, the basic architecture of the network consists of the following: Input – 1D Conv. layer – Grouped Conv. layer – 1D Conv. layer – 1D Conv. layer – Fully Connected layer – Output. Each convolutional layer is followed by a max-pooling layer and a batch normalization layer. ReLU serves as the activation function for convolutional layers, softmax for classifying the final output classes, and Adam as the optimizer. The inputs comprise three-axis raw acceleration data and three-axis raw angular velocity data, with an input shape of (100, 6). The 1D convolutional kernels have a size of 3, and the grouped convolutional kernel size is 1. The output shape is 4, corresponding to the classes: walking, running, hand movements while standing, and hand movements while walking. The network is implemented using TensorFlow 2.

B. Adaptive Step Detection

Following the classification of activities into four classes, we implemented class-specific step detection algorithms to improve step detection accuracy and, consequently, enhance the overall performance of pedestrian dead reckoning (PDR). We conducted step detection analysis appropriate for each of the four scenarios. Fig. 2 shows the flow of the proposed adaptive step detection algorithm.

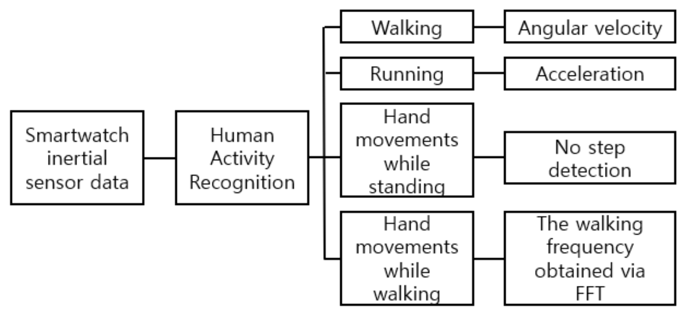


Fig. 2: The flow of the proposed adaptive step detection algorithm.

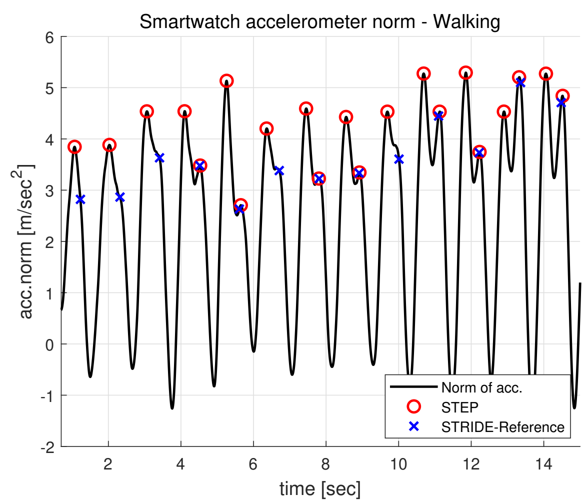
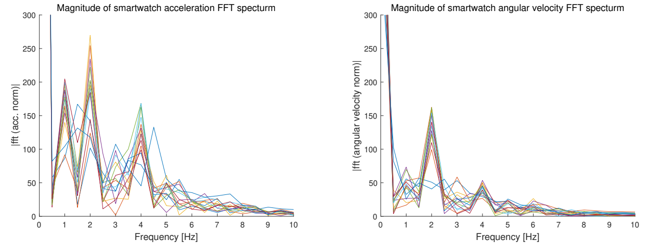


Fig. 3: False step detection case when using smartwatch acceleration magnitude.

a) Normal walking scenario: We applied a low-pass filter to the smartwatch acceleration magnitude data and performed step detection using a peak detection method. For



(a) The FFT spectrum of accelerations.

(b) The FFT spectrum of angular velocities.

Fig. 4: The FFT spectrums of normal walking data.

reference, we utilized two Xsens MTw2 sensors, which were attached to the left wrist and left ankle. The blue reference points in Fig. 3 were detected from the left ankle data utilizing the ZUPT algorithm from our previous study [16]. Since the sensor was attached to only one ankle, the reference points represent strides. Wrist data were used to synchronize the timing of the ankle data with the watch data. As shown in Fig. 3, while there were instances where step detection was successful, other cases exhibited missed steps. Specifically, we observed that when the arm wearing the smartwatch moved backward, the signal magnitude decreased, resulting in less prominent peaks.

To mitigate this issue, we propose using gyroscope data for step detection. The rationale behind this approach is that gyroscope data allows for more robust step detection; as the arm consistently swings during walking, the smartwatch collects data on these movements, resulting in more consistent observations of angular velocity changes compared to acceleration changes. Additionally, FFT is employed to derive the spectrum of the Euclidean norm of three-dimensional accelerations and angular velocities. As shown in Fig. 4, the spectra of angular velocities display distinct peaks compared to those of accelerations, facilitating more accurate step detection.

b) Running: While running, the movements of the hands and arms are less independent and exhibit faster arm swing speeds compared to walking, resulting in larger changes in acceleration. Therefore, we use the filtered acceleration magnitude data and perform step detection using a peak detection method. The threshold and time interval between consecutive peaks were selected empirically.

c) Hand movements while standing: If HAR indicates that the pedestrian is stationary, step detection is not performed, and the position is not updated.

d) Hand movements while walking: As depicted in Fig. 5, hand movements during walking disrupt the sensor signal, leading to incorrect step detection results. To address this issue, we leverage HAR results. When the HAR system determines that the user is walking but the signal is unsuitable for normal step detection, we extract the walking frequency from a 2-second segment of angular velocities during normal walking just before the period of hand movements using FFT. We then utilize this derived walking frequency to fill in steps during the time intervals where the signal is corrupted.

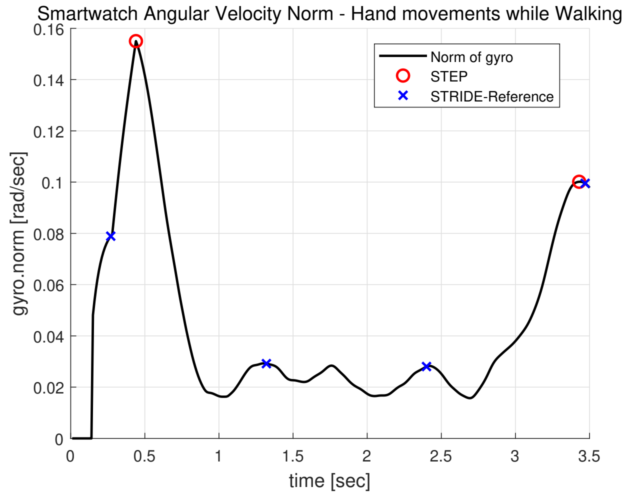


Fig. 5: Step detection with signal disruption due to hand movements during walking.

TABLE I: Comparison of HAR Results among the CNN, depthwise separable CNN, and the Proposed Network.

	CNN	***DS-CNN***	***Proposed***
Number of trainable parameters	8852	6610	6036
Model size (KB)	34.58KB	25.82KB	23.58KB
Prediction time (sec)	0.0094	0.0114	0.0092
Accuracy	0.9605	0.9650	0.9645
MACs	294400	159720	227840

IV. Experiments

A. Human Activity Recognition

To evaluate the performance of the proposed network, we used a standard CNN model as a baseline and compared it with another lightweight CNN model that employs depthwise separable convolution, as introduced in [14]. The standard CNN model consisted of four convolutional layers with a kernel size of 3. The depthwise separable convolution-based CNN (DS-CNN) model included one depthwise separable convolution layer and three regular convolutional layers with a kernel size of 3. Our proposed network comprised three convolutional layers with a kernel size of 3 and one grouped convolutional layer with a kernel size of 1. The results are presented in Table I, where we compare the number of trainable parameters, model size, prediction time, accuracy, and MACs (Multiply-ACcumulates). By utilizing four convolutional layers in all models, we observed a significant reduction in the number of parameters– approximately one-third –when replacing one convolutional layer with a grouped convolutional layer, along with a slight improvement in accuracy.

B. Adaptive Step Detection

In this study, we propose an adaptive step detection algorithm designed to robustly detect pedestrian steps using a smartwatch, tailored to the pedestrian's walking state. As

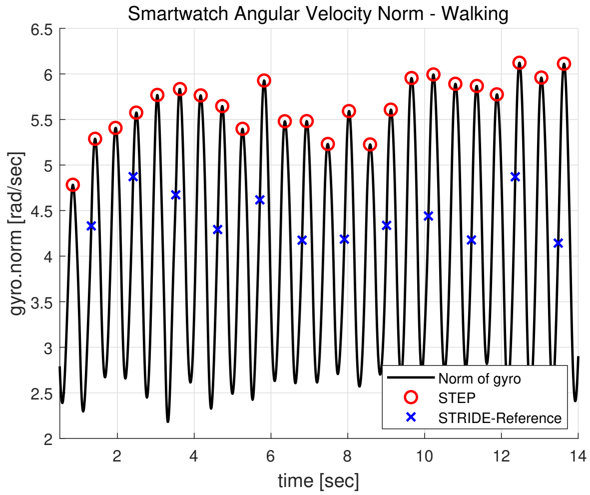


Fig. 6: Step detection using angular velocity magnitude in a normal walking scenario.

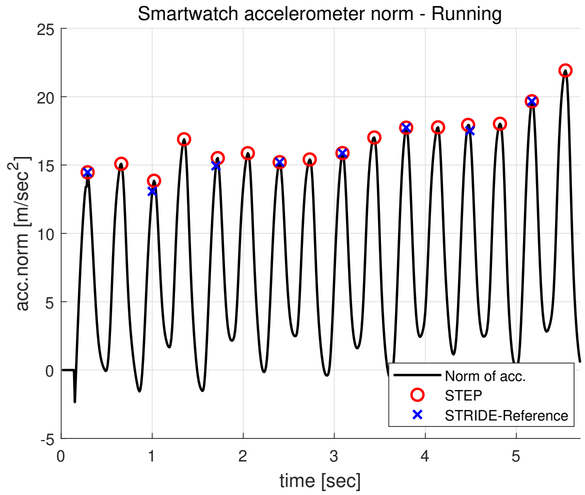


Fig. 7: Step detection using acceleration magnitude in a normal running scenario.

shown in Fig. 6, steps were accurately detected using the gyroscope during normal walking, with no missed steps. Although the reference points detected using acceleration and angular velocity data from the left ankle sensor have different detection times compared to the steps detected from the watch's angular velocity data, the total number of detected steps is consistent.

Since the acceleration of the arm changes significantly while running, as shown in Fig. 7, steps can be robustly detected based on the magnitude of acceleration.

In instances where inertial sensor data were corrupted due to upper body movements during walking, the proposed algorithm effectively tracked the actual steps. As illustrated in Fig. 9, by utilizing the 2 Hz walking frequency estimated from the angular velocity FFT during a preceding period, as shown in Fig. 8, we were able to compensate disrupted time intervals with estimated steps. In this walking scenario, a segment (highlighted in yellow) experienced signal corruption due to stretching. We performed an FFT on the angular velocity data from the 2 seconds of normal walking preceding the corrupted segment. Using this frequency, we interpolated the steps within the corrupted segment, resulting in compensated steps that closely match the reference steps. The step detection accuracy using the proposed method is 100%, although this should be considered in the context of the limited number of sequences in experimental settings like those shown in Fig 8. Since a person's walking speed generally does not change abruptly within a single sequence, and even when walking speed is significantly reduced, the walking frequency does not decrease as much as the step length, it is appropriate to use the walking frequency obtained prior to the occurrence of hand movements.

V. Conclusion

In this paper, we proposed a lightweight CNN-based human activity recognition (HAR) network by replacing one convolutional layer with a grouped convolutional layer, significantly

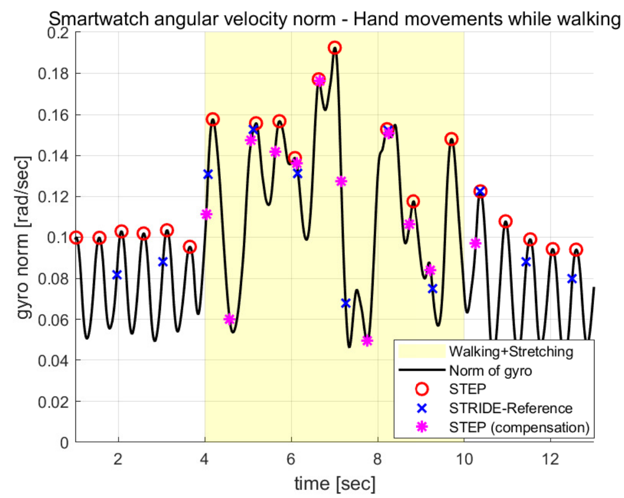


Fig. 8: Adaptive step detection results when there is a segment with arm movements while walking.

reducing the number of parameters and computational load. With a HAR accuracy of 96.45%, we demonstrated that the network effectively distinguishes between four activity classes. Generally, since the arm consistently swings while walking, causing distinct changes in angular velocity, we use gyroscope data to detect steps during normal walking. For running, where changes in acceleration are more pronounced, we rely on accelerometer data for step detection. When HAR detects a stationary state, no steps are detected. If HAR identifies a walking state but the sensor signals are disrupted by arm movements, we estimate the steps using the walking frequency derived from the FFT of walking data just before the disruption period. This approach allows us to correct the step count during periods of unusable signals.

In future research, we plan to further refine and diversify the activities classified by HAR by defining a wider range

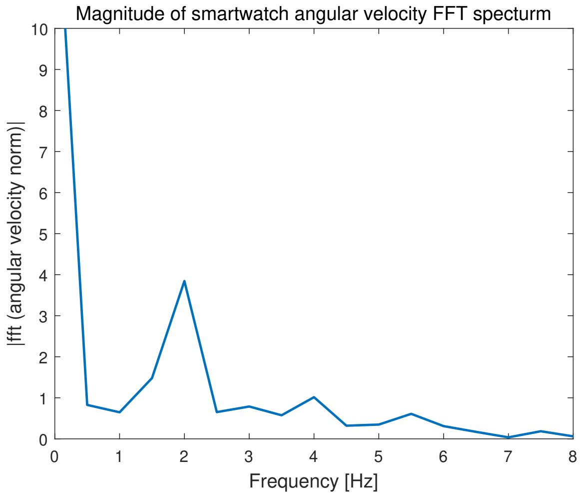


Fig. 9: The FFT spectrum of angular velocities from the 2 seconds of normal walking before the corrupted segment.

of complex hand movements or activities to improve the classification capabilities for different scenarios experienced by smartwatch users. In addition, we will use the robustly detected steps to estimate the user's step length and position. Finally, we will modify and improve the algorithm to enable real-time processing on actual smartwatch platforms.

References

[1] S. Beauregard and H. Haas, "Pedestrian Dead Reckoning: A Basis for Personal Positioning," *Proc. 3rd Workshop Positioning, Navig. and Commun.* 2006, pp. 27-35.

[2] W. Kang and Y. Han, "SmartPDR: Smartphone-Based Pedestrian Dead Reckoning for Indoor Localization," *IEEE Sensors J.* vol. 15, no. 5, pp. 2906-2916, May 2015, doi: 10.1109/JSEN.2014.2382568.

[3] A. R. Pratama, Widyawen, and R. Hidayat, "Smartphone-based Pedestrian Dead Reckoning as an indoor positioning system," *2012 Int. Conf. Sys. Eng. Technol. (ICSET)*, Bandung, Indonesia, 2012, pp. 1-6, doi: 10.1109/ICSEngT.2012.6339316.

[4] N. H. Ho, P. H. Truong, and G. M. Jeong, "Step-Detection and Adaptive Step-Length Estimation for Pedestrian Dead-Reckoning at Various Walking Speeds Using a Smartphone," *Sensors*, vol. 16, no. 9, 1423, Sep. 2016, doi: 10.3390/s16091423.

[5] X. Kang, B. Huang, and G. Qi, "A Novel Walking Detection and Step Counting Algorithm Using Unconstrained Smartphones," *Sensors*, vol. 18, no. 1, 297, Jan. 2018, doi: 10.3390/s18010297.

[6] P. Catalfamo, S. Ghoussayni, and D. Ewins, "Gait Event Detection on Level Ground and Incline Walking Using a Rate Gyroscope," *Sensors*, vol. 10, no. 6, pp. 5683-5702, June 2010, doi: 10.3390/s100605683.

[7] V. Genovese, A. Mannini and A. M. Sabatini, "A Smartwatch Step Counter for Slow and Intermittent Ambulation," *IEEE Access*, vol. 5, pp. 13028-13037, 2017, doi: 10.1109/ACCESS.2017.2702066.

[8] S.J. Preece, J. Y. Goulermas, L. P. J. Kenney, D. Howard, K. Meijer, and R. Crompton, "Activity identification using body-mounted sensors–a review of classification techniques," *Physiol Meas.*, vol. 30, no. 4, pp. R1-33, Apr. 2009, doi: 10.1088/0967-3334/30/4/R01.

[9] L. Bao and S. S. Intille, "Activity Recognition from User-Annotated Acceleration Data," in *Pervasive Computing*, A. Ferscha and F. Mattern, Eds. Berlin, Heidelberg: Springer, 2004, vol. 3001, Lecture Notes in Computer Science, pp. 1-17.

[10] S. J. Preece, J. Y. Goulermas, L. P. J. Kenney and D. Howard, "A Comparison of Feature Extraction Methods for the Classification of Dynamic Activities From Accelerometer Data," *IEEE Trans. Biomed. Eng.*, vol. 56, no. 3, pp. 871-879, March 2009, doi: 10.1109/TBME.2008.2006190.

[11] T. Huynh and B. Schiele, "Towards Less Supervision in Activity Recognition from Wearable Sensors," *2006 10th IEEE Int. Symp. Wearable Comput.*, Montreux, Switzerland, 2006, pp. 3-10, doi: 10.1109/ISWC.2006.286336.

[12] M. -K. Yi, W. -K. Lee and S. O. Hwang, "A Human Activity Recognition Method Based on Lightweight Feature Extraction Combined With Pruned and Quantized CNN for Wearable Device," *IEEE Trans. Consum. Electron.*, vol. 69, no. 3, pp. 657-670, Aug. 2023, doi: 10.1109/TCE.2023.3266506.

[13] T. Liu, S. Wang, Y. Liu, W. Quan, and L. Zhang, "A lightweight neural network framework using linear grouped convolution for human activity recognition on mobile devices," *J. Supercomput.*, vol. 78, pp. 6696-6716, doi: 10.1007/s11227-021-04140-5.

[14] A. G. Howard *et al.*, "Mobilenets: Efficient convolutional neural networks for mobile vision applications," 2017, *arXiv:1704.04861*.

[15] X. Zhang, X. Zhou, M. Lin, and J. Sun, "ShuffleNet: An Extremely Efficient Convolutional Neural Network for Mobile Devices," 2017, *arXiv:1707.01083*.

[16] H. Ju and C. G. Park, "A Pedestrian Dead Reckoning System Using a Foot Kinematic Constraint and Shoe Modeling for Various Motions," *Sensors and Actuators A*, vol. 284, pp. 135-144, Dec. 2018, doi: 10.1016/j.sna.2018.09.043.