IEEE Open Journal of Antennas and Propagation

Received XX Month, XXXX; revised XX Month, XXXX.

Pedestrian Activity Recognition and Prediction with Multi-Dimensional Point Cloud Based Deep Learning for Millimeter Wave Radar

Zihang Zhou* and Shouhei Kidera*, Senior Member, IEEE

[1]Graduate School of Informatics and Engineering, the University of Electro-Communications, Tokyo, 1828585, Japan

CORRESPONDING AUTHOR: Shouhei Kidera (e-mail: kidera@uec.ac.jp).

This research was supported by JST FOREST Program, Grant Number JPMJFR2025, Japan.

ABSTRACT

This study presents a multidimensional point cloud-based pedestrian activity recognition framework for millimeter-wave (mmW) pedestrian collision avoidance sensors. While many existing methods exploit range-velocity-angle data through Fourier-based processing along array or slow-time dimensions, such approaches face fundamental limitations including a high-frequency mmW radar systems, particularly in the 79-GHz band. These include across-range-unit (ARU) range walk due to high range resolution, a limited unambiguous velocity range, and the inherent tradeoff between temporal and velocity resolutions. To address these issues, we propose a k-decomposed weighted kernel density (WKD)-based range-Doppler velocity estimation method that directly extracts the range, Doppler velocity, direction-of-arrival (DOA) angles, and reflection intensity from point cloud data to overcome the shortcomings of short-time Fourier transform (STFT)-based analysis. Furthermore, for pedestrian motion estimation and future trajectory prediction at crosswalks, the proposed framework combines this multidimensional point cloud representation with a Transformer-based model. Specifically, it employs a 12-class classification scheme comprising three walking speed states and four trajectory patterns, and enables accurate prediction of a pedestrian ' s final position (1-2 seconds ahead) using only short-term slow-time observations.

INDEX TERMS Millimeter wave radar, Automotive radar, Human activity recognition (HAR), micro-Doppler analysis, Transformer based deep learning, weighted kernel density (WKD) Doppler estimators, Multi-dimensional point cloud data

I. Introduction

The deployment of fully autonomous driving systems and advanced driver-assistance systems (ADAS) has made reliable pedestrian collision avoidance a critical yet unresolved challenge in the field of automotive sensing. In complex urban environments, pedestrians frequently appear suddenly from behind occlusions, particularly at crosswalks. Numerous accident analyses have indicated that many pedestrian-related collisions are caused not only by missed detections but also by the inability to accurately infer the motion state and crossing intention of pedestrians sufficiently early [1], [2]. This limitation has highlighted the need for collision avoidance systems that go beyond the mere detection of pedestrians and explicitly address understanding and predicting their motion states and short-term behavior. Millimeter-wave (MMW) radar is well suited for this task because of its robustness against adverse weather and its operability in both line-of-sight (LOS) and non-line-of-sight (NLOS) environments [3]–[6]. Recent ADAS have increasingly relied on 76–81 GHz millimeter-wave radar to provide robust perception under adverse weather conditions. This technology supports applications such as adaptive cruise control, automatic emergency braking, and blind spot detection. More sophisticated systems integrate high-resolution multi-input-multi-output (MIMO) and imaging radar, along with deep learning-based methods for pedestrian detection, activity recognition, and trajectory prediction, often within sensor fusion frameworks that combine radar, cameras, and LiDAR. However, radar images are fundamentally constrained by the limited physical aperture of the radar array, which limits

the angular resolution, especially in far-range scenarios. As a result, radar image based approaches struggle to provide reliable human classification or motion interpretation [7]–[13].

To overcome this spatial resolution bottleneck, prior studies have extensively investigated micro-Doppler-based human activity recognition (HAR), which leverages motion-induced frequency modulations of individual body parts [14]–[23]. While micro-Doppler signatures provide discriminative dynamic features even at limited angular resolutions, existing approaches predominantly rely on analyzing the Doppler velocity using coherent processes such as the short-time Fourier transform (STFT) or other transform based algorithm [24]–[34]. These methods suffer from inherent tradeoffs between the temporal and velocity resolutions and are constrained by a limited unambiguous velocity range. These limitations are particularly severe for high-frequency MMW radar systems operating in the 79-GHz band. Furthermore, the ultrawide bandwidths (over 3 GHz) commonly used in this band yield a very high range resolution (under 50 mm), which introduces the across-range-unit (ARU) range walk problem in estimating the Doppler velocity for each part of human body with walking motion [35]–[39]. The resulting migration of the radar signals across range bins during the coherent processing interval disrupts coherent integration and degrades the accuracy of the Doppler velocity estimation. Consequently, conventional STFT-based micro-Doppler analysis is ill-suited for the reliable estimation of pedestrian motion states in practical automotive MMW radar systems. Moreover, in conventional HAR schemes, many approaches focus on classifying qualitative postures or motion states from range-Doppler spectrograms by using range-Doppler spectrogram via convolutional neural networks (CNNs) or transformers [24]–[34], or they utilize a point cloud-based approach derived from conventional range-azimuth data [40]–[43]. In contrast, for pedestrian collision avoidance, it is essential to estimate pedestrians' walking speed before a crosswalk, their intention to stop or proceed straight, and their eventual stopping position. However, to the best of our knowledge, there are very few studies that formulate these requirements explicitly as classification or prediction tasks. The CNN-based approaches are widely used in radar signal processing for their ability to extract local patterns via convolutional filters [24]–[26], [30], [32]. CNNs mainly capture short-term temporal correlations in time-frequency domain representations, such as range, Doppler or micro-Doppler maps. However, their limited receptive fields make it difficult to model long-term temporal dependencies without substantially increasing model complexity, which can hinder the extraction of global motion trends required for tasks such as pedestrian intent prediction. Conversely, long short-term memory (LSTM) networks explicitly model temporal dependencies through recurrent hidden-state updates which have led to their wide adoption for radar-based motion analysis [46]–[48]. Nevertheless, their performance may degrade when modeling long and non-stationary radar sequences affected by complex multipath propagation. Transformer-based models have recently attracted increasing attention for radar time-series analysis owing to their ability to model long-range temporal dependencies and complex spatiotemporal relationships. Although several studies have applied Transformer-based models to radar-based gait-level classification [49], [50], most of them rely on Doppler velocity estimates obtained through Fourier-based analysis. As a result, these approaches inherently suffer from limited temporal resolution and a limited unambiguous velocity range. In contrast, this paper leverages Doppler velocity information derived from the WKD method, which provides PRI-level temporal resolution and an effectively unbounded velocity range.

To address these challenges, this study proposes a Doppler velocity analysis framework based on weighted kernel density (WKD) estimation [44], where it converts the time-of-flight (TOF) point clouds at each slow-time frame, referred to as range-τ points, to the Doppler velocity via weighted kernel density estimation. This formulation provides a temporal resolution equivalent to the pulse repetition interval (PRI) and, importantly, alleviates the fundamental limitations in the STFT-based micro-Doppler analysis, especially in terms of unambiguous velocity range. Furthermore, leveraging the characteristics of high-frequency millimeter-wave bands, we introduce the wave-number k-decomposed WKD method with the data separation in the wave-number (k) spaces [45]. On this basis, pedestrian motion classification and prediction are achieved by using a multi-dimensional point cloud in which range-Doppler, velocity, and angle information are uniquely associated in a one-to-one manner. Additionally, as the k-decomposed WKD method yields a compact and sparse representation of pedestrian motion as multidimensional point clouds, it significantly reduces input dimensionality and improves optimization efficiency in subsequent machine learning (ML) stages. Moreover, the proposed framework analyzes multidimensional point cloud data in which range, velocity, wavenumber, and reflection intensity are uniquely associated on a one-to-one basis using a transformer architecture. This combination is highly original approach, distinguished fundamentally from existing radar-based gait analysis methods.

The main contributions of this paper are summarized as follows:

1) We introduce the k-decomposed WKD method in converting the radar data to the range-Doppler velocity-DOA angles, and slow time associated point cloud data, where it achieves PRI-level temporal resolution and overcomes key limitations of STFT-based micro-Doppler analysis, especially for limited unambiguous velocity range, and the ARU range walk problem.
2) Pedestrian dynamics are represented as sparse range-v_{d}-τ point clouds, enabling efficient learning while preserving essential spatiotemporal motion features,

This article has been accepted for publication in IEEE Open Journal of Antennas and Propagation. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/OJAP.2026.3698432

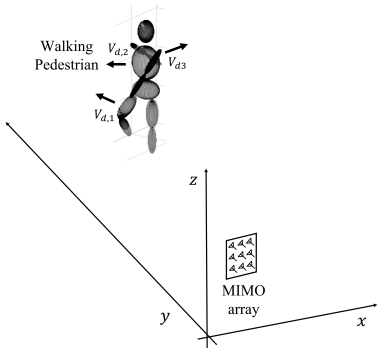


FIGURE 1: Observation model with MIMO array and walking pedestrian.

including walking acceleration and deceleration at crosswalks.

3) A Transformer-based framework is introduced to capture long-range temporal dependencies and global motion patterns within point cloud sequences, enabling robust pedestrian motion classification and short-term future position prediction from limited slow-time observations.
4) The experiments using a 79-GHz FMCW MIMO radar system in the indoor scenario demonstrate that our proposed scheme achieve reliable pedestrian motion recognition and prediction, even with relatively short observation duration.

II. Method

A. Observation Model

Fig. 1 illustrates the observation model used in this study. Multiple transmitters and and receivers form an array deployed on the plane $y = 0$. Each transmitter repeatedly emits pulses with a fixed pulse repetition interval (PRI), and the slow-time index τ is defined as the discrete temporal index of the PRI sequence. The spatial locations of each transmitter and receiver are denoted by $\boldsymbol{r}_{\mathrm{T}} = (x_T, 0, z_T)$ and $\boldsymbol{r}_{\mathrm{R}} = (x_R, 0, z_R)$, respectively. For each transmitter–receiver pair, the complex-valued electric field reflected from the target is measured and expressed as $s(\boldsymbol{r}_{\mathrm{T}}, \boldsymbol{r}_{\mathrm{R}}, R, \tau)$, where the range variable is defined as $R = ct/2$, with t being the fast-time index and c being the propagation speed of electromagnetic waves. A range extraction filter, such as a matched filter, is then applied to $s(\boldsymbol{r}_{\mathrm{T}}, \boldsymbol{r}_{\mathrm{R}}, R, \tau)$, and its output is denoted by $\tilde{s}(\boldsymbol{r}_{\mathrm{T}}, \boldsymbol{r}_{\mathrm{R}}, R, \tau)$. Once the complex-valued response $\tilde{s}(\boldsymbol{r}_{\mathrm{T}}, \boldsymbol{r}_{\mathrm{R}}, R, \tau)$ is available, the proposed framework is applicable to both pulse-Doppler radar and frequency-modulated continuous-wave (FMCW) radar systems.

B. Generation of Multidimensional Point Clouds

1) Original WKD

Here, to generate multi-dimensional point clouds from time-varying micro-Doppler analysis, we introduce a WKD-based Doppler analysis framework. The original WKD-based processing proceeds as follows. Discrete range-τ points are first extracted by identifying local maxima of $|\tilde{s}(\boldsymbol{r}_{\mathrm{T}}, \boldsymbol{r}_{\mathrm{R}}, R, \tau)|$ along the range dimension R, satisfying

$$\frac{\partial |\tilde{s}(\boldsymbol{r}_{\mathrm{T}}, \boldsymbol{r}_{\mathrm{R}}, R, \tau)|}{\partial R} = 0, \tag{1}$$

$$|\tilde{s}(\boldsymbol{r}_{\mathrm{T}}, \boldsymbol{r}_{\mathrm{R}}, R, \tau)| \geq \alpha \max_{R,\tau} |\tilde{s}(\boldsymbol{r}_{\mathrm{T}}, \boldsymbol{r}_{\mathrm{R}}, R, \tau)|, \tag{2}$$

where α is a threshold parameter satisfying $0 \leq \alpha \leq 1$. The extracted local maxima are denoted as $\boldsymbol{q}_{i,j} \equiv (\tau_i, R_{i,j})$, where $R_{i,j}$ represents the j-th range peak observed at the slow-time index τ_i.

Given a reference point $\boldsymbol{q}_{i,j}$ and its neighboring points $\boldsymbol{q}_{k,l}$, a candidate Doppler velocity is defined as

$$v_{\mathrm{d}}(\boldsymbol{q}_{i,j}, \boldsymbol{q}_{k,l}) \equiv \frac{R_{i,j} - R_{k,l}}{\tau_i - \tau_k}. \tag{3}$$

For all valid combinations of $\boldsymbol{q}_{i,j}$ and $\boldsymbol{q}_{k,l}$, the optimal Doppler velocity $\hat{v}_{\mathrm{d}}(\boldsymbol{q}_{i,j})$ is estimated through a weighted kernel density maximization given by

$$\hat{v}_{\mathrm{d}}(\boldsymbol{q}_{i,j}) = \arg\max_{v_{\mathrm{d}}} \sum_{k,l} \exp\left(-\frac{\left||\tilde{s}(\boldsymbol{q}_{i,j})| - |\tilde{s}(\boldsymbol{q}_{k,l})|\right|^2}{2\sigma_s^2}\right) \times \exp\left(-\frac{|\tau_i - \tau_k|^2}{2\sigma_\tau^2}\right) \exp\left(-\frac{|v_{\mathrm{d}} - \tilde{v}_{\mathrm{d}}(\boldsymbol{q}_{i,j}, \boldsymbol{q}_{k,l})|^2}{2\sigma_{v_{\mathrm{d}}}^2}\right), \tag{4}$$

where $|\tilde{s}(\boldsymbol{q}_{i,j})|$ denotes the magnitude of the $\tilde{s}(\boldsymbol{q}_{i,j})$. The parameters σ_s, σ_τ, and $\sigma_{v_{\mathrm{d}}}$ are fixed kernel bandwidths that control the weighting with respect to signal amplitude similarity, temporal proximity, and Doppler velocity consistency, respectively; their roles and selection criteria are discussed in detail in [44]. Because the WKD estimation method obtains the Doppler velocity from range-τ point distributions by using an incoherent statistical process, it

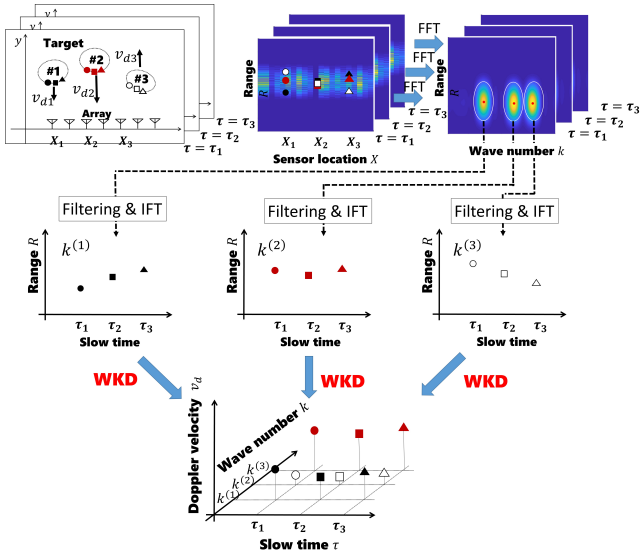


FIGURE 2: Conceptual figure of the k decomposed WKD method. respectively.

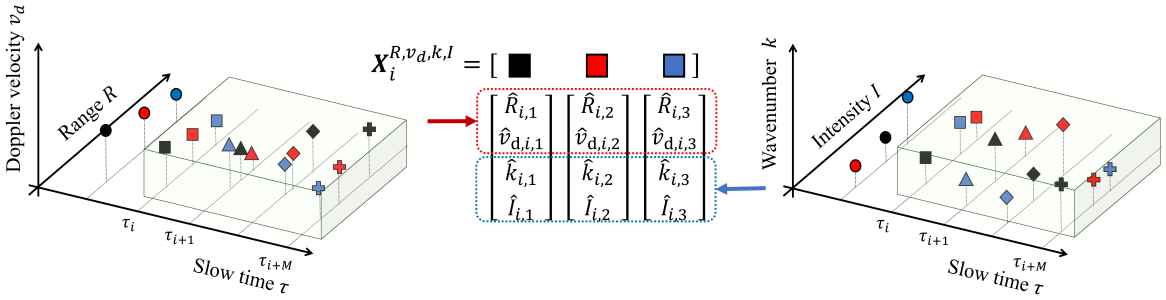


FIGURE 3: Multi-dimensional point cloud associated with R-v_{d}-k-I.

inherently avoids the across-range-unit (ARU) range walk problem that degrades conventional coherent Doppler processing. Nevertheless, the accuracy of the WKD method is fundamentally limited by the accuracy of the range-τ point representation, and interference reflections from different body parts within a range resolution cell can degrade the estimation accuracy of the Doppler velocity.

2) k Decomposed WKD

To overcome the above difficulty, this paper introduces k decomposition based WKD scheme [45] as follows. Focusing on the single-input multi-output (SIMO) or MIMO array radar model, this method assumes that the reflection responses from different parts of the human body exhibit not only different Doppler velocities when in motion but also distinct directions of arrival (DOAs) angles. Based on this assumption, the k-space decomposition enables separation of the received signals in the DOA angular domain. Under the assumption of a horizontal array, the received signal $\tilde{s}(\boldsymbol{r}_{\mathrm{T}}, \boldsymbol{r}_{\mathrm{R}}, R, \tau)$ is converted to the k-space domain via the 1D Fourier transform along the x axis as:

$$\tilde{S}(k, R, \tau) = \int_{\Omega \mathrm{A}} \tilde{s}(x, R, \tau) e^{-jkx} dx, \tag{5}$$

where Ω_{A} is the aperture area and k is the wave-number associated with the x coordinate of array elements. The actual MIMO array arrangement is converted to the equivalent SISO array. Notably, the azimuth DOA θ is associated with the center wavelength λ as $\theta = \sin^{-1}(k\lambda/2\pi)$. For each local maximum in the k space, denoted as $\boldsymbol{\zeta}^{(n)} \equiv (\tilde{k}^{(n)})$, the following filtering process and the 1-D inverse Fourier transform is applied as:

$$\hat{s}(x, R, \tau; \boldsymbol{\zeta}^{(n)}) = \int W(k; \boldsymbol{\zeta}^{(n)}) \tilde{S}(k, R, \tau) e^{jkx} dk. \tag{6}$$

where $W(k; \boldsymbol{\zeta}^{(n)})$ is the Gaussian window applied in k space. Finally, from the n-th decomposed data set $\bar{s}(R, \tau; \boldsymbol{\zeta}^{(n)})$, which is obtained by coherently averaging the $\hat{s}(x, R, \tau; \boldsymbol{\zeta}^{(n)})$ along x, is used to extracxt the range-τ points corresponding to each cluster as $\boldsymbol{q}_{i,j}^{(n)} \equiv (R_{i,j}^{(n)}, \tau_i^{(n)})$ under the following conditions:

$$\frac{\partial |\bar{s}(R, \tau; \boldsymbol{\zeta}^{(n)})|}{\partial R} = 0 \tag{7}$$

$$|\bar{s}(R, \tau; \boldsymbol{\zeta}^{(n)})| \geq \alpha \max_{R, \tau} |\bar{s}(R, \tau; \boldsymbol{\zeta}^{(n)})| \tag{8}$$

where α is a threshold parameter satisfying $0 \leq \alpha \leq 1$. The Doppler velocity value $\hat{v}_{\mathrm{d}}(\boldsymbol{q}_{i,j}^{(n)})$ corresponding to each range-τ point as $\boldsymbol{q}_{i,j}^{(n)}$ and each cluster is determined via the WKD process in similar to Eq. (4) as:

$$\hat{v}_{\mathrm{d}}(\boldsymbol{q}_{i,j}^{(n)}) = \arg \max_{v_{\mathrm{d}}} \sum_{k,l} \exp\left(-\frac{\left| |\bar{s}(\boldsymbol{q}_{i,j}^{(n)})| - |\bar{s}(\boldsymbol{q}_{k,l}^{(n)})| \right|^2}{2\sigma_s^2} \right) \times \exp\left(-\frac{|\tau_i - \tau_k|^2}{2\sigma_\tau^2} \right) \exp\left(-\frac{|v_{\mathrm{d}} - \tilde{v}_{\mathrm{d}}(\boldsymbol{q}_{i,j}^{(n)}, \boldsymbol{q}_{k,l}^{(n)})|^2}{2\sigma_{v_{\mathrm{d}}}^2} \right), \tag{9}$$

Fig. 2 illustrates the overall process of the k-decomposed WKD for each cluster. From $\bar{s}(R, \tau; \boldsymbol{\zeta}^{(n)})$, the data corresponding to each local maximum in the range-k profile are extracted and clustered, enabling the identification of range-τ points that are not resolvable under conventional range resolution. Subsequently, the WKD is applied to each cluster $\boldsymbol{q}_{i,j}^{(n)}$ to obtain k-associated range-velocity-τ points.

C. Feature Extraction with Multi-dimensional Point Cloud

The k-decomposed WKD estimation method generates multidimensional point cloud data in which clouds that jointly associate the range R, Doppler velocity v_{d}, wavenumber k, and reflection intensity I, which in turn can be used to extract features for recognizing and predicting pedestrian motion states and positions. The following four different feature vectors are defined as:

$$\boldsymbol{X}_i^R \equiv \left[\hat{R}_{i,1},\ \hat{R}_{i,2},\ \cdots,\ \hat{R}_{i,N}\right]^\top \tag{10}$$

$$\boldsymbol{X}_i^{v_{\mathrm{d}}} \equiv \left[\hat{v_{\mathrm{d}}}_{i,1},\ \hat{v_{\mathrm{d}}}_{i,2},\ \cdots,\ \hat{v_{\mathrm{d}}}_{i,N}\right]^\top \tag{11}$$

$$\boldsymbol{X}_i^k \equiv \left[\hat{k}_{i,1},\ \hat{k}_{i,2},\ \cdots,\ \hat{k}_{i,N}\right]^\top \tag{12}$$

$$\boldsymbol{X}_i^I \equiv \left[\hat{I}_{i,1},\ \hat{I}_{i,2},\ \cdots,\ \hat{I}_{i,N}\right]^\top \tag{13}$$

The rationale for selecting these features is summarized as follows:

Feature $\boldsymbol{X}_i^R$: represents the temporal variation in the range dimension and captures both the overall displacement

This article has been accepted for publication in IEEE Open Journal of Antennas and Propagation. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/OJAP.2026.3698432

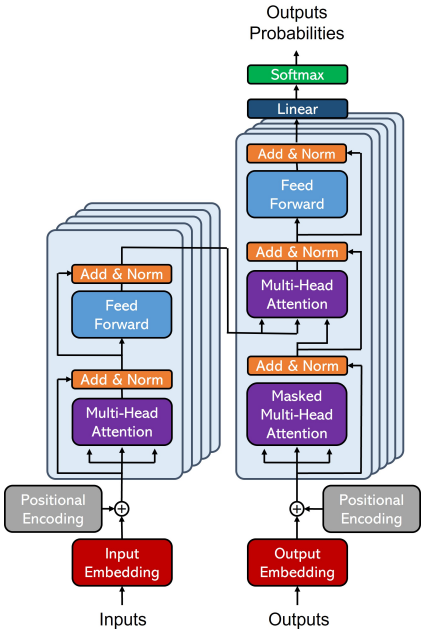


FIGURE 4: Standard structure of Transformer based DNN.

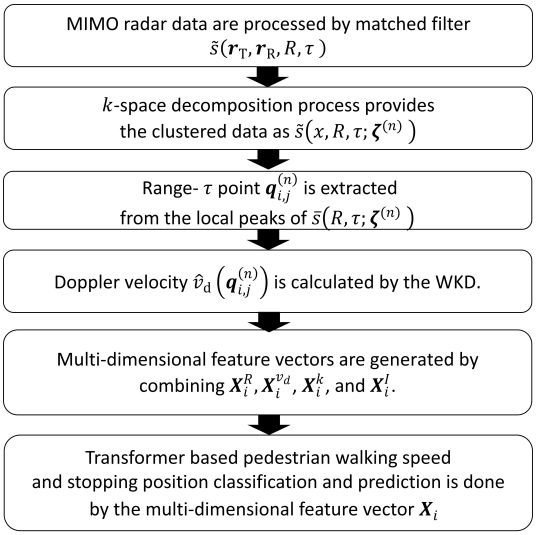


FIGURE 5: Processing flow of the proposed scheme.

of a walking pedestrian and the range variations of individual body parts along the radar LOS. It effectively encodes information related to the pedestrian ' s movement direction and position as well as the spatial relationships among different body parts.

Feature $\boldsymbol{X}_i^{v_{\mathrm{d}}}$: captures the temporal variations in the Doppler velocity across body parts to provide a quantitative description of the motion of individual body parts such as relative limb velocities during walking, as well as the pedestrian ' s walking speed. In particular, when combined with feature R, it forms a discriminative feature that reflects the target ' s position and direction of movement.

Feature $\boldsymbol{X}_i^k$: represents the temporal variation of the DOA for each reflection response. By capturing the angular dynamics of the spatial relationships among body parts in the azimuthal direction, as well as the time-varying walking direction of the pedestrian, it inherently encodes information related to the target ' s direction of motion and velocity.

Feature $\boldsymbol{X}_i^I$: expresses the temporal variation in the reflection intensity I. Because reflection coefficients differ across body parts (*e.g.*, torso, arms, legs), this feature should be effective at clustering reflections corresponding to individual body parts, especially when associated with other features (R, v_{d}, and k).

All possible non-empty combinations of the four feature components are considered, resulting in a total of 15 feature vector patterns. The performance of each combination is then systematically evaluated. For example, the feature vector constructed from R, v_{d}, k and I can be defined as follows.

$$\boldsymbol{X}_i^{R,v_{\mathrm{d}},k,I} \equiv \begin{bmatrix} \hat{R}_{i,1}, & \hat{R}_{i,2}, & \cdots, & \hat{R}_{i,N} \\ \hat{v}_{\mathrm{d},i,1}, & \hat{v}_{\mathrm{d},i,2}, & \cdots, & \hat{v}_{\mathrm{d},i,N} \\ \hat{k}_{i,1}, & \hat{k}_{i,2}, & \cdots, & \hat{k}_{i,N} \\ \hat{I}_{i,1}, & \hat{I}_{i,2}, & \cdots, & \hat{I}_{i,N} \end{bmatrix}^{\top} \tag{14}$$

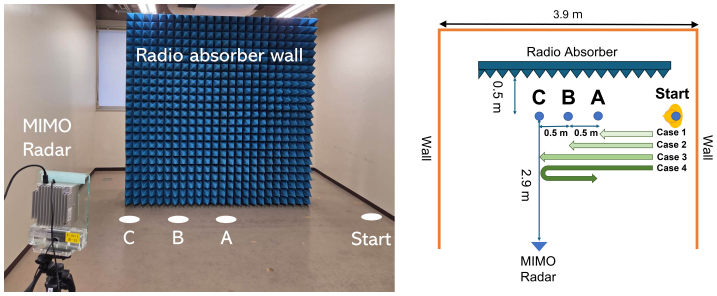
Figure 3 shows the representative illustration of the feature vector extraction for $\boldsymbol{X}_i^{R,k,v_{\mathrm{d}},I}$. The point cloud generated by the k-decomposed WKD preserves a one-to-one correspondence among the range, Doppler velocity, and slow-time index τ, which allows for precise representation of temporal variations in both the spatial and velocity domains within the single PRI. This attribute is particularly effective for identifying pedestrian motion states, including the walking speed and stopping position. Such feature integration is unique to the WKD approach and cannot be achieved using conventional STFT-based coherent integration processes.

D. Transformer Based Deep Neural Network

A transformer-based DNN is employed to estimate the pedestrian walking speed and final stopping position from the multidimensional point clouds. Given the multi-dimensional point cloud time-series input $\boldsymbol{X}_i$, self-attention is applied along the temporal axis. Each time step corresponds to an aggregated point cloud, and the following attention weights model the relevance between different τ:

$$\mathrm{Attention}(\boldsymbol{Q}, \boldsymbol{K}, \boldsymbol{V}) = \mathrm{softmax}\left(\frac{\boldsymbol{Q}\boldsymbol{K}^{\top}}{\sqrt{\mathrm{d_k}}}\right)\boldsymbol{V}, \tag{15}$$

where the projections $\boldsymbol{Q}, \boldsymbol{K}, \boldsymbol{V}$ are obtained from the selected feature vectors as $\boldsymbol{X}_i$. In this formulation, the attention coefficients $A_{t,\tau}$ quantify how strongly the point cloud features at time τ contribute to the representation at time t. This is particularly advantageous for pedestrian motion analysis, as salient motion cues may occur intermittently and



(a) Experimental scene (b) Top view

FIGURE 6: Experimental setup, geometry, and actual scene. The human model takes a stepping motion with different period.

(a) Position Start (b) Position A (c) Position B (d) Position C

(e) **LOW** speed (f) **NORMAL** speed (g) **HIGH** speed

FIGURE 7: Actual scene for each position and walking speed.

are not necessarily localized in contiguous time intervals. By treating multidimensional point clouds as tokens in a temporal sequence, the transformer-based DNN provides a unified framework for modeling radar time-series data. Figure 4 shows the processing of temporal data by the transformer-based DNN.

E. Processing Flow

Figure 5 shows the overall process flow of the proposed framework, which can be summarized as follows:

Step 1): $s(\boldsymbol{r}_{\mathrm{T}}, \boldsymbol{r}_{\mathrm{R}}, R, \tau)$ is measured for each combination of transmitter and receiver, and the matched filter is applied as $\tilde{s}(\boldsymbol{r}_{\mathrm{T}}, \boldsymbol{r}_{\mathrm{R}}, R, \tau)$

Step 2): Decomposition process in the k space is applied to $\hat{s}(x, R, \tau; \boldsymbol{\zeta}^{(n)})$ as given in Eq. 6, and the clustered range-τ points are extracted as $\boldsymbol{q}_{i,j}^{(n)}$ using Eq. 8.

Step 3): For each range-τ point $\boldsymbol{q}_{i,j}^{(n)}$, the Doppler velocity $\hat{v}_{\mathrm{d}}(\boldsymbol{q}_{i,j}^{(n)})$ is optimized by Eq. (9).

Step 4): Given N, the feature vectors $\boldsymbol{X}_i$ are extracted using arbitrary combinations of $\boldsymbol{X}_i^R$, $\boldsymbol{X}_i^{v_{\mathrm{d}}}$, $\boldsymbol{X}_i^k$, and $\boldsymbol{X}_i^I$.

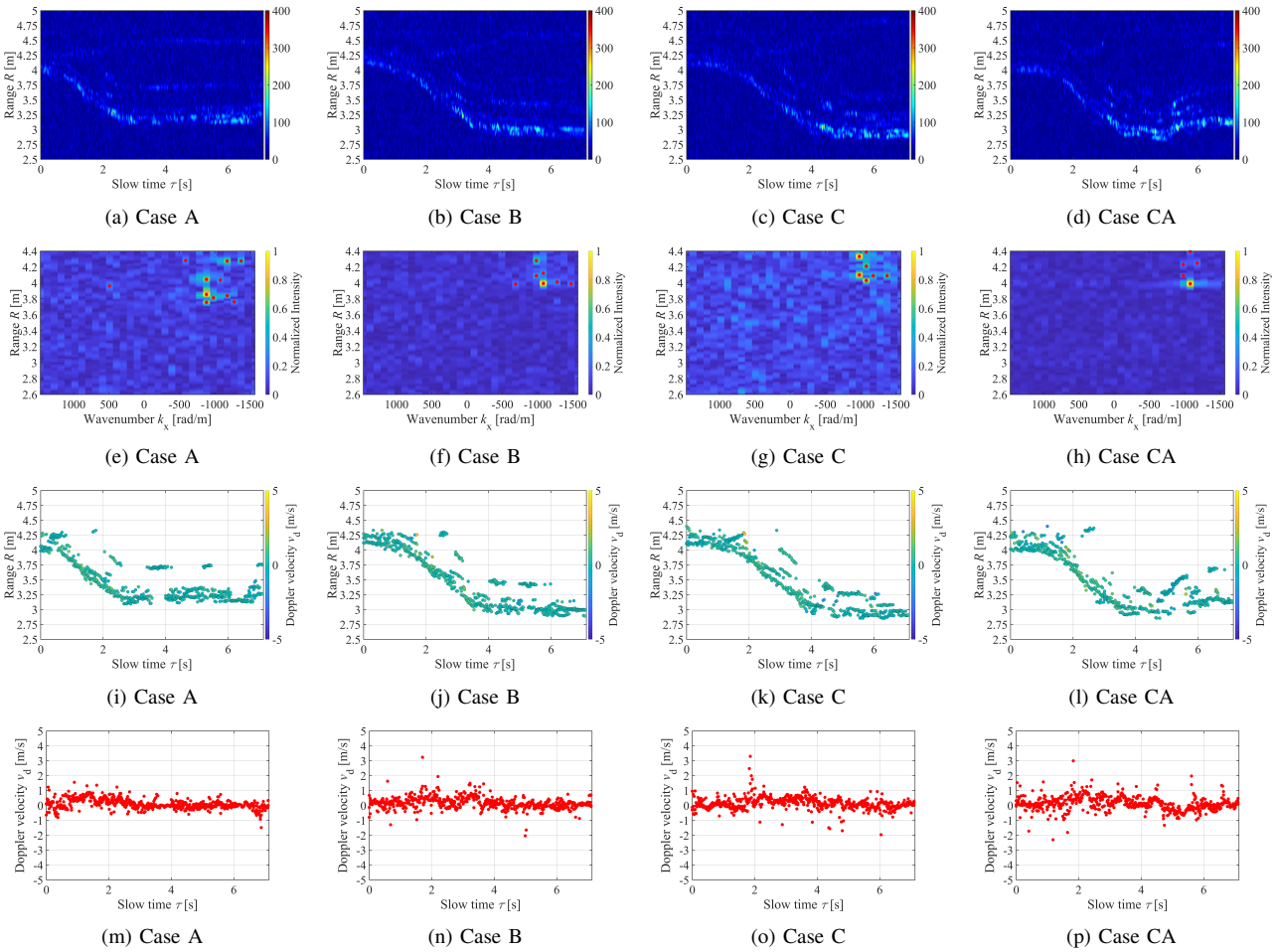
Step 5): Each feature vector is input into the transformer-based DNN to estimate the pedestrian walking speed and stopping position estimation.

III. Results in Experimental Data

A. Experimental Setup

A real-world experiment is set up to simulate a pedestrian crossing scenario and to evaluate the performance of the proposed framework regarding the classification of pedestrian motion states and prediction of the pedestrian position. A 79-GHz FMCW MIMO mmWave radar (Sakura Tech Corp.) with a 4.0-GHz bandwidth was employed comprising six transmit and eight receive channels equipped with patch antennas (output power: 10 dBm), forming a 48-element virtual array. The radar is mounted at a height of 1250 mm. Notably, the informed consent for the human subject has been obtained. Each patch antenna has vertical and horizontal 3-dB beamwidths of 10° and 45°, respectively.

This article has been accepted for publication in IEEE Open Journal of Antennas and Propagation. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/OJAP.2026.3698432



(a) Case A (b) Case B (c) Case C (d) Case CA

(e) Case A (f) Case B (g) Case C (h) Case CA

(i) Case A (j) Case B (k) Case C (l) Case CA

(m) Case A (n) Case B (o) Case C (p) Case CA

FIGURE 8: Raw data, range-k, range-τ, and Doppler velocity-τ distributions estimated by the k-decomposed WKD method in **LOW** speed case. 1st row: Raw data. 2nd row: Range-k profile at $\tau = 1$ s. 3rd row: Range-τ profile with the estimated Doppler velocity. 4th row: Doppler velocity-τ profile. The color scale in 3rd row indicates Doppler velocity estimated by k-decomposed WKD.

Assuming a target average distance of 4 m, the vertical beamwidth corresponds to an approximate coverage of ± 700 mm, which is sufficient to encompass the upper legs to the upper chest of a standing human body (approximately from 700 to 1250 mm in height). The pulse repetition interval (PRI) was set to 31 ms, yielding a maximum unambiguous Doppler velocity range of ± 0.0141 m/s under Fourier-based Doppler processing. A total of 300 pulses were transmitted, resulting in an overall observation duration of 9.3 s.

In this experiment, we define 12 representative pedestrian motion patterns relevant to collision avoidance at crosswalks, obtained by combining three walking-speed levels with four stopping-position scenarios. These patterns are designed to reflect critical situations that a vehicle must perceive and predict for safe decision-making. The pedestrian model is an adult male with a height of 1700 mm. The pedestrian's initial walking speed is categorized into three levels: HIGH (1.2 m/s), NORMAL (0.9 m/s), and LOW (0.65 m/s). In all trials, the pedestrian starts from the same initial position $(x, y) = (3.0, 2.9)$ m and moves in the x-direction, corresponding to the lateral crossing direction relative to a vehicle. The pedestrian then decelerates and comes to a stop at one of three predefined positions: A $(x, y) = (1.0, 2.9)$ m, B $(x, y) = (0.5, 2.9)$ m, or C $(x, y) = (0.0, 2.9)$ m. **Position A** represents a safe stopping location where collision avoidance by the vehicle is ensured, **Position B** corresponds to a critical boundary case in which collision avoidance is barely feasible, and **Position C** denotes a hazardous location where collision avoidance becomes difficult or unlikely. To further emulate realistic dart-out behavior that is frequently observed in traffic accidents, an additional **Position CA** pattern is introduced, in which the pedestrian initially moves toward the hazardous position C and subsequently retreats to the safe position A. Thus, four stopping-positions are considered

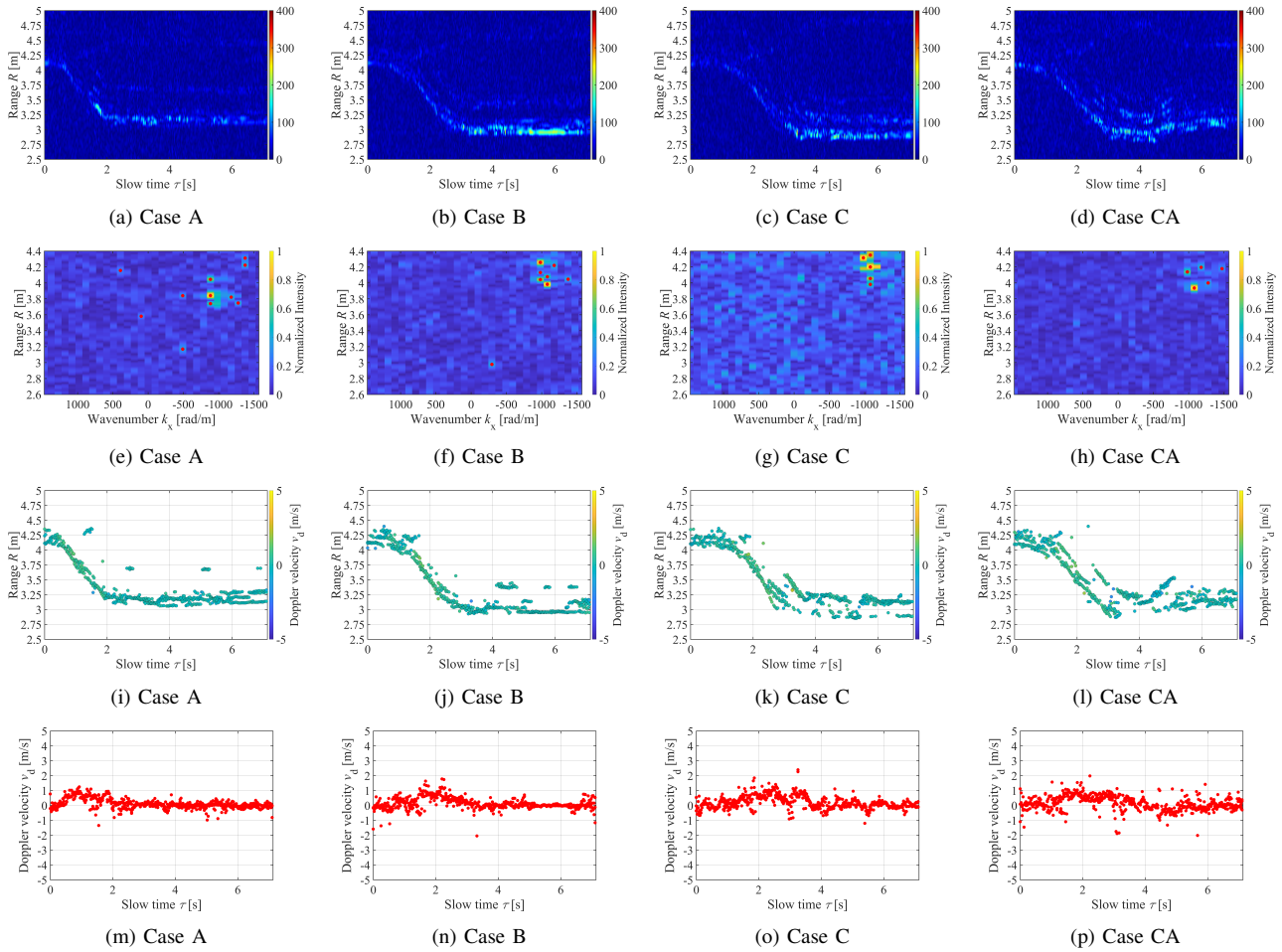


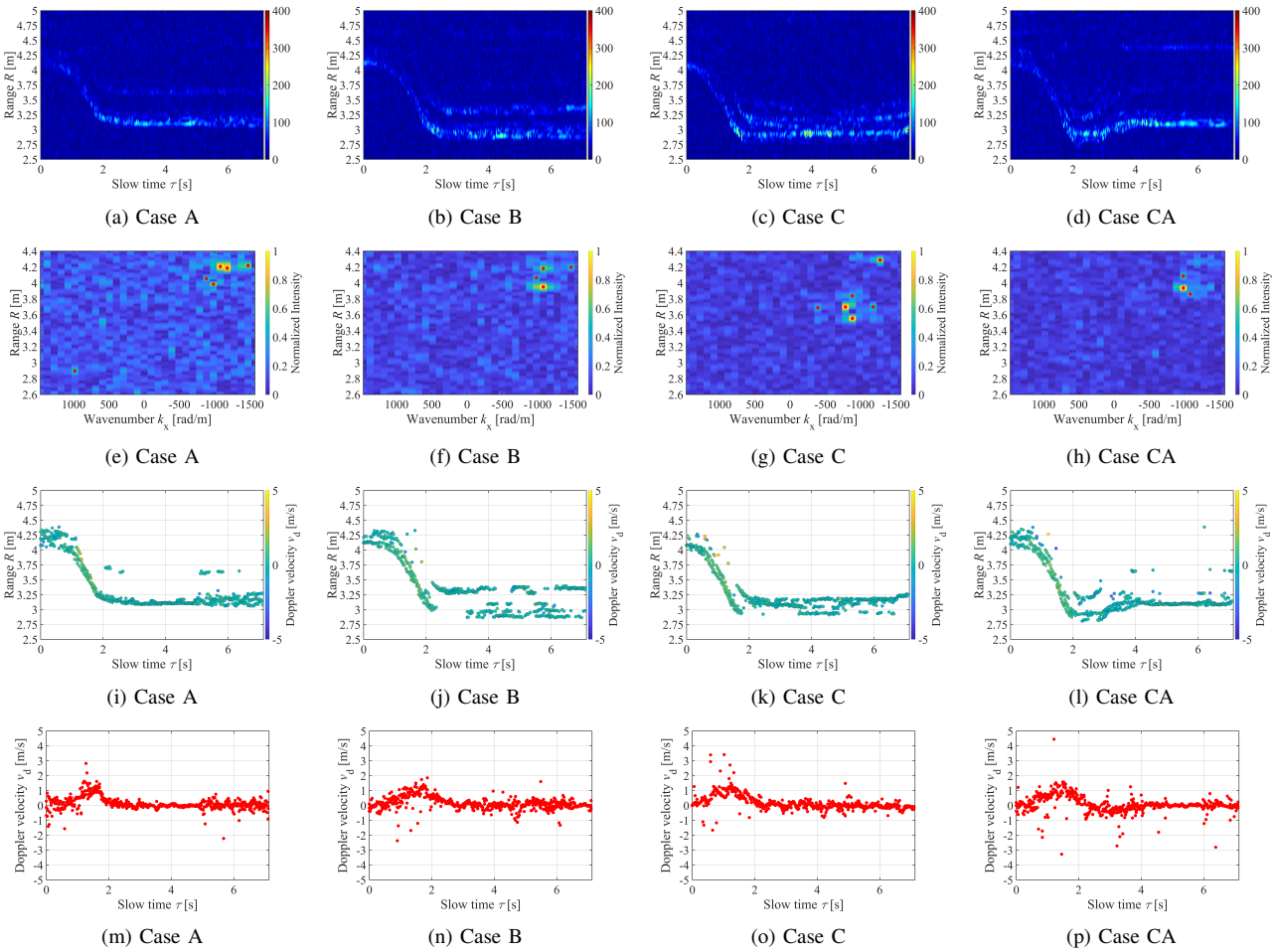
FIGURE 9: Raw data, range-k, range-τ, and Doppler velocity-τ distributions estimated by the k-decomposed WKD method in **NORMAL** speed case. 1st row: Raw data. 2nd row: Range-k profile at $\tau = 1$ s. 3rd row: Range-τ profile with the estimated Doppler velocity. 4th row: Doppler velocity-τ profile. The color scale in 3rd row indicates Doppler velocity estimated by k-decomposed WKD.

in total. Figs. 6 and 7 shows the experimental setup and the actual scenes used to represent the above stopping positions and walking speeds. Each motion pattern was repeated five times.

B. Data and Range-Velocity-τ Point Cloud

Figures 8, 9, and 10 show the reflection intensity profiles, the extracted range-τ points, range-k space ($\tau = 1$ s), and the associated Doppler velocity point clouds estimated by the k decomposed WKD method, for the motion speed as LOW, NORMAL, and HIGH, respectively. The parameters of the WKD are set as $\sigma_s = 1.0$, $\sigma_\tau = 3.0$ s, and $\sigma_{v_\mathrm{d}} = 0.1$ m/s. In this case, the start time of each motion was recorded, and radar data corresponding to 7130 ms (230 hits) were extracted from that time onward. As shown in Fig. 8, the slope of the R-τ response became smaller at lower walking speeds, the slope of the range-τ response becomes smaller, which indicated that a longer duration is required for the pedestrian to come to a stop. In the radar reflection responses, the torso contribution is dominant; however, fluctuations in the range direction induced by the motion of the arms and legs are also observable. As illustrated in the intermediate process of the range-k profile (2nd row of each figure), approximately 1 s after motion begins, the positions and intensities of peak responses vary across different walking patterns. Each walking pattern, from initial speed to braking, induces distinct body motions and limb dynamics, leading to characteristic distributions in the wavenumber k (*i.e.*, angular). In addition, even within the conventional range resolution (37.5 mm), separation in the k domain enables the isolation of distinct clusters of peak responses. When the range-τ information is extracted again for each cluster, responses from different body parts can be resolved beyond the nominal range resolution. Thus, the k-decomposed WKD can extract range-τ points from data separated in both angular and velocity domains (different

This article has been accepted for publication in IEEE Open Journal of Antennas and Propagation. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/OJAP.2026.3698432



(a) Case A (b) Case B (c) Case C (d) Case CA

(e) Case A (f) Case B (g) Case C (h) Case CA

(i) Case A (j) Case B (k) Case C (l) Case CA

(m) Case A (n) Case B (o) Case C (p) Case CA

FIGURE 10: Raw data, range-k, range-τ, and Doppler velocity-τ distributions estimated by the k-decomposed WKD method in **HIGH** speed case. 1st row: Raw data. 2nd row: Range-k profile at $\tau = 1$ s. 3rd row: Range-τ profile with the estimated Doppler velocity. 4th row: Doppler velocity-τ profile. The color scale in 3rd row indicates Doppler velocity estimated by k-decomposed WKD.

body parts), resulting in a substantial increase in the number of range-τ points, particularly during walking motion. Furthermore, in the estimation of v_{d}-τ points within the WKD framework, the k-decomposed WKD method effectively suppresses outliers and generates v_{d}-τ distributions that are consistent with the motion periodicity of individual body parts. Over the duration of a walking motion from onset to completion, the WKD scheme is able to estimate not only the average velocity variation of the whole body but also the velocities of individual body parts to some extent. Notably, in using the STFT-based analysis, the maximum velocity range is fundamentally limited to ± 0.014 m/s, which makes it difficult to accurately estimate velocity variations exceeding 1 m/s.

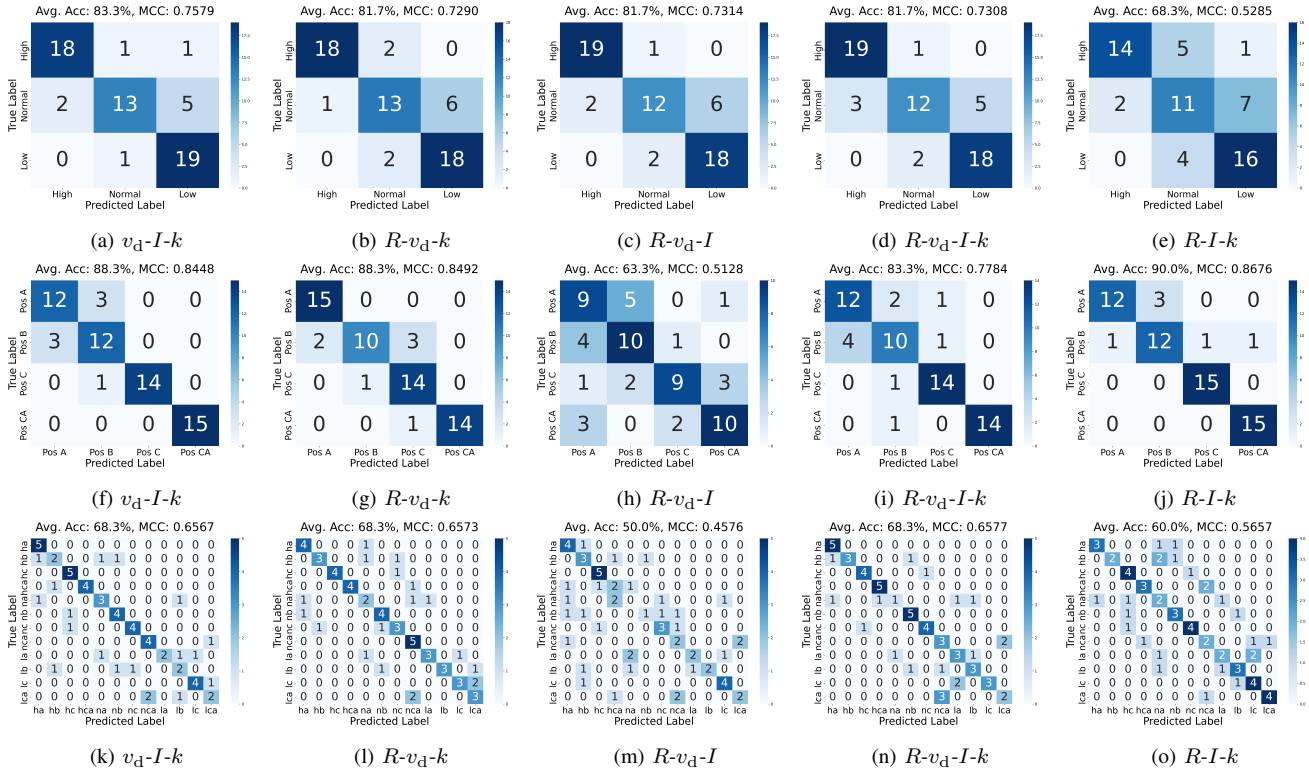
C. Classification Performance

The classification performance is evaluated in terms of the walking speed and stopping position. Because five trials were conducted for each motion pattern, five-fold cross-validation is used for the evaluation. Fig. 11 shows the classification results for only the walking speed (three classes), only the stopping position (four classes), and the joint speed-position patterns (12 classes). Five representative input vectors were selected from combinations of range (R), Doppler velocity (v_{d}), wavenumber (k), and intensity (I) in the multi-dimensional point-cloud vectors defined in Section C.

The classification criteria are accuracy and the Matthew's correlation coefficient (MCC) [52]. The accuracy of a multi-class classification represents the proportion of samples that is correctly predicted and is mathematically defined as

$$\text{Accuracy} \equiv \frac{N_{\mathrm{CC}}}{N_{\mathrm{Total}}} \tag{16}$$

where N_{CC} and N_{Total} denote the total number of correctly classified samples and the total number of samples in the test dataset, respectively. This is equivalent to dividing the sum of the diagonal elements of the confusion matrix (*i.e.*, numbers



(a) v_{d}-I-k (b) R-v_{d}-k (c) R-v_{d}-I (d) R-v_{d}-I-k (e) R-I-k

(f) v_{d}-I-k (g) R-v_{d}-k (h) R-v_{d}-I (i) R-v_{d}-I-k (j) R-I-k

(k) v_{d}-I-k (l) R-v_{d}-k (m) R-v_{d}-I (n) R-v_{d}-I-k (o) R-I-k

FIGURE 11: Confusion matrices using each feature combination of R, v_{d}, k, and I using the proposed method. 1st row: Classification of 3 different speeds. 2nd row: Classification of 4 different stopping position. 3rd row: Classification of 12 different classes (3 speeds $\times$ 4 stopping positions).

of correct predictions) by the total number of samples. The MCC [52] of a multiclass classification can be defined as follows by using the confusion matrix $\boldsymbol{C}$ with K classes:

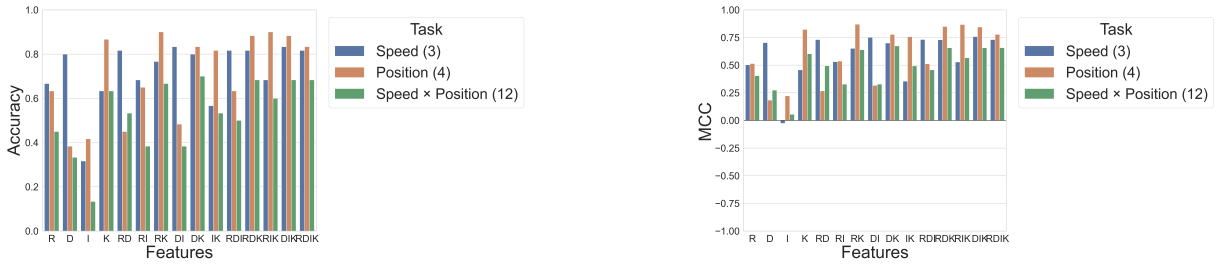
$$\mathrm{MCC} = \frac{c\,s - \sum_{k=1}^{K} p_k t_k}{\sqrt{\left(s^2 - \sum_{k=1}^{K} p_k^2\right)\left(s^2 - \sum_{k=1}^{K} t_k^2\right)}}. \tag{17}$$

where s denotes the total number of samples, c is the total number of correctly predicted samples (*i.e.*, the sum of the diagonal elements of the confusion matrix), t_k is the number of true occurrences of class k (ground-truth labels), and p_k denotes the number of samples predicted as class k. Fig. 12 shows the results the accuracy and MCC for all combinations of feature values for R, v_{d}, k, and I (15 patterns in total). The classification results for only the walking speed show that the combination of v_{d}-k-I yields the highest MCC of 0.758. In contrast, for the position classification and joint position-speed classification, the combinations k or R-k consistently achieve higher MCC values. The classification results for only the walking speed show that the accuracy was reduced at the normal speed. This is likely because the medium speed was between the low and high speeds, which made it more difficult to discriminate, and the reproducibility of the walking speed was not perfectly controlled in the experimental setting. The classification results for only the stopping position show that the accuracy was slightly lower for positions A and B than for position C. This may be because position C lay near the center of the radar beam, which resulted in a high signal-to-noise ratio (SNR). In contrast, positions A and B did not pass through the center of the radar beam, which led to a lower SNR and potentially reduced the reliability of the data. For position CA, where the pedestrian returned from position C to A, the trajectory passed through position C, which yielded a higher SNR yielding higher SNR and a characteristic range variation, that should improve the classification performance. Nevertheless, the random classification would yield accuracies of approximately 33% for three-class classification, 25% for four-class classification, and 8.3% for twelve-class classification. The results demonstrate that the proposed method achieves classification accuracies that are significantly higher than these chance levels.

D. Prediction Performance

Next, we describe the prediction performance of the proposed method. In this evaluation, the task is to predict the final stopping position of a pedestrian using radar data acquired before the walking motion is completed, where the

This article has been accepted for publication in IEEE Open Journal of Antennas and Propagation. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/OJAP.2026.3698432



(a) Accuracy (b) MCC

FIGURE 12: Accuracy and MCC for all combinations of multi-dimensional point cloud of R, v_{d}, k, and I. Label "R", "D", "K", and "I" denote the features of R, v_{d}, k, and I, respectively. Blue, orange, and green bars denote the results in classifying of "Speed (3 classes)", "Position (4 classes)", and "Speed $\times$ Position (12 classes)", respectively.

TABLE 1: Parameters used in the Transfomer.

Parameter	Value
Number of epochs	80
Mini-batch size	32
Initial learning rate	0.006
Embedding dimension	64
Number of multi-head attention heads	2
Number of stacked encoder layers	3

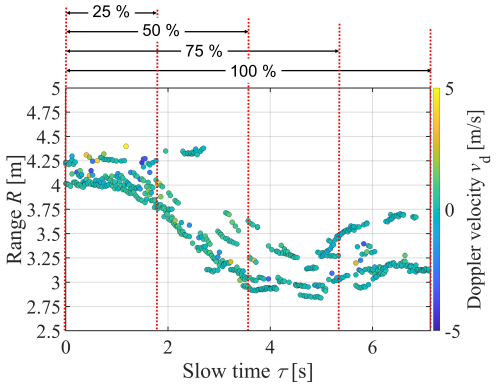
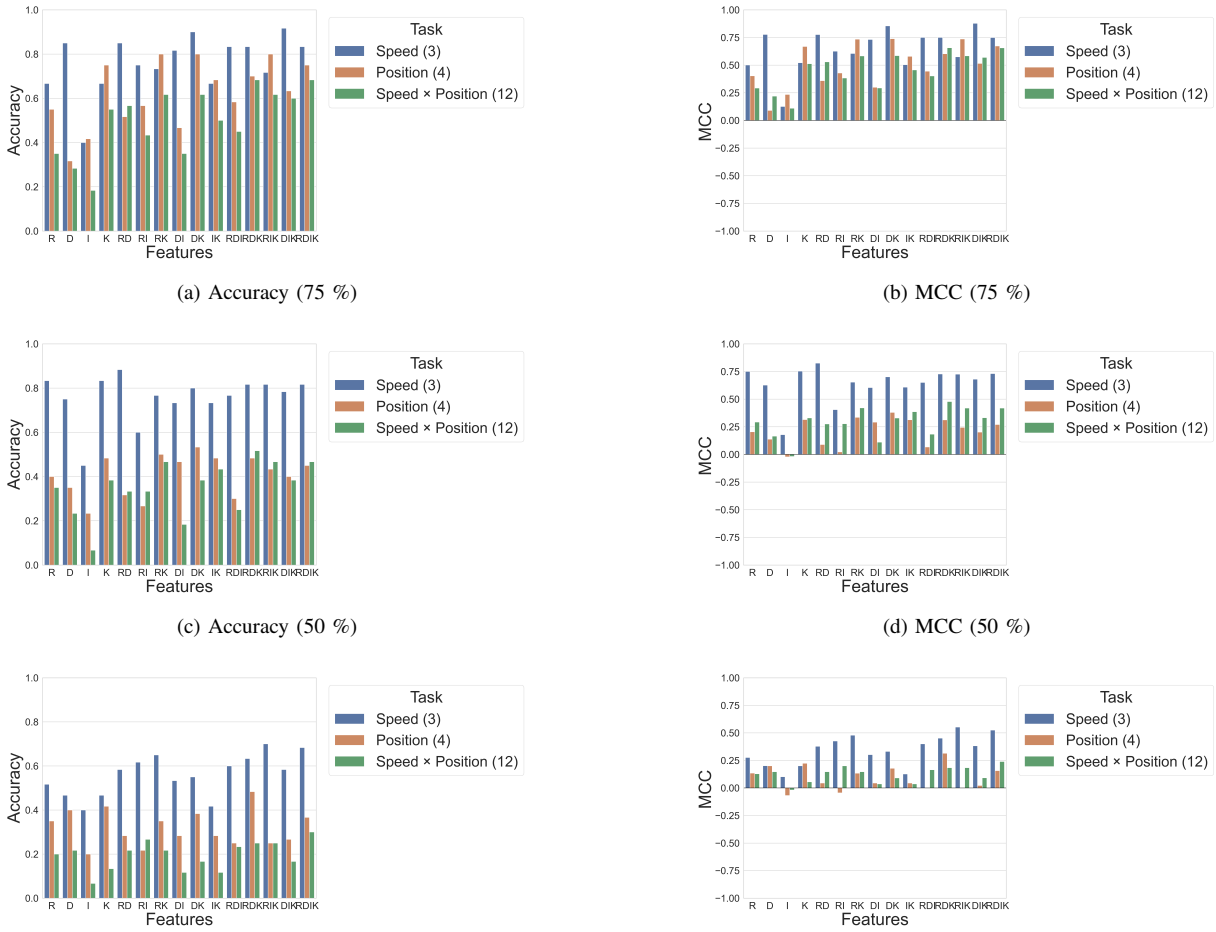


FIGURE 13: Data splits scheme for the prediction task (25 %, 50 %, 75 %, and 100 %).

observation window is set to 7.1 s from the start of motion. The problem is formulated as a four-class classification task corresponding to the stopping positions, *i.e,* Case 1, 2, 3, and 4 in Fig. 7. To assess early prediction capability, the data length used for prediction is varied as a fraction of the total duration from motion onset to arrival at the final stopping position. Specifically, predictions are performed by the first 25 %, 50 %, and 75 % of the data, where 100% corresponds to the full duration until arrival. Fig. 13 illustrates the data splits for the prediction task. Fig. 14 presents the prediction performance in terms of the accuracy and MCC according to the walking speed and each data ratio for each data split. Notably, the average arrival times differ depending on the walking speed: 4.6 s for the **LOW**-speed condition, 3.3 s for the **Normal**-speed condition, and 2.5 s for the **HIGH**-speed condition. When the available data are limited, *i.e.*, even when temporally shorter data segments are used, both the position estimation accuracy (orange bars) and the MCC consistently degrade across all feature types. In particular, when only 50 % of the data are used, the MCC drops below 0.5 in many cases, which is substantial decline compared with the results obtained using the full dataset. This clearly indicates that relying only on the early portion of the observation data is insufficient for accurate prediction of the final position, especially in scenario with position CA. These results highlight that reliable prediction of the final stopping position requires a sufficiently long temporal context. Estimation of the walking speed showed no notable differences in performance at data splits of 100 %, 75 %, and 50 %, and a notable degradation in performance appeared only at a data split of 25 %. Even in this severe case, the accuracy remains at approximately 70 %. This robustness suggests that the walking speed can be inferred primarily from the slope information of the range-τ profile, which allows meaningful estimation even with a relatively short observation duration. In the 12-class classification task, when 75 % of the data are used, the MCC remain around 0.5 when employing the RDIK features. Given that random prediction yields an accuracy of only 8.3 %, this result is significantly better than chance and indicates that the proposed framework retains predictive capability even under reduced data availability. These results indicate that, while accurate prediction of the stopping position demands longer temporal observation, the proposed framework can still provide useful and reliable estimates of dynamic attributes such as the speed and maintain a reasonable multiclass classification accuracy with moderate data lengths. These characteristics are particularly important for real-world applications, where early prediction from limited observations is often required.

Here, we evaluate the computational complexity of each process of the proposed method, including the computational



(a) Accuracy (75 %)

(b) MCC (75 %)

(c) Accuracy (50 %)

(d) MCC (50 %)

(e) Accuracy (25 %)

(f) MCC (25 %)

FIGURE 14: Accuracy and MCC for all combinations of multi-dimensional point cloud of R, v_{d}, k, and I. Label "R", "D", "K", and "I" denote the features of R, v_{d}, k, and I, respectively. 1st, 2nd, and 3rd columns express the cases when using 75 %, 50 %, and 25 % data for prediction scenario, respectively. Blue, orange, and green bars denote the results in classifying of "Speed (3 classes)", "Position (4 classes)", and "Speed $\times$ Position (12 classes)", respectively.

time required for the k-decomposed WKD method based on the observed data, and the data length required for classification, defined as the product of the PRI and the number of pulse hits. Table 2 lists the computational times of the k-decomposed WKD, for each data set as 25 %, 50 %, 75 %, and 100 %, using CPU Intel Xeon Gold 6330 x2 processor with 1024GB RAM. The computational complexity of the k-decomposed WKD is proportional to the number of extracted range-τ points N_{RP}; thus using a smaller dataset significantly reduces computational costs. However, because the RP data are processed in segments along the slow-time dimension, the full computational time for the 100 % dataset is not always necessary. By contrast, if the number of range-τ points is insufficient, the velocity estimation accuracy of the WKD will decrease. Therefore, determining the appropriate slow-time length (*i.e.*, number of pulse hits) is essential for balancing these considerations. Furthermore, because the current implementation has not been optimized for computational efficiency, further code optimization will be needed for faster processing in practical applications.

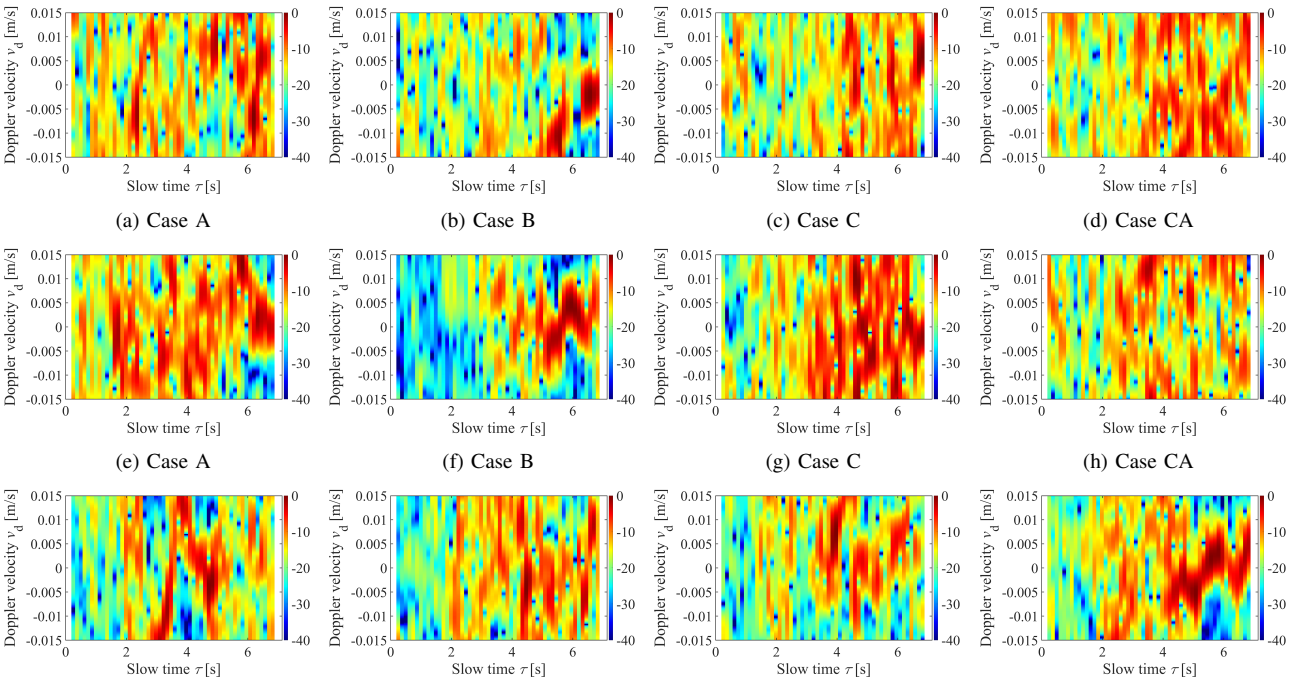
TABLE 2: Computational times for each process in the proposed method.

	25 %	50 %	75 %	100 %
Steps 1 and 2	43 s	91 s	139 s	187 s
Step 3	35 s	27 s	68 s	107 s
Steps 4 and 5	$>$ 0.1 s	$>$ 0.1 s	$>$ 0.1 s	$>$ 0.1 s

E. Comparison to STFT-Based Classifiers

To validate the effectiveness of the proposed method, we compare it with conventional deep-learning-based benchmark approaches using STFT-based inputs, specifically CNN- and Vision Transformer (ViT)-based methods. CNNs

This article has been accepted for publication in IEEE Open Journal of Antennas and Propagation. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/OJAP.2026.3698432



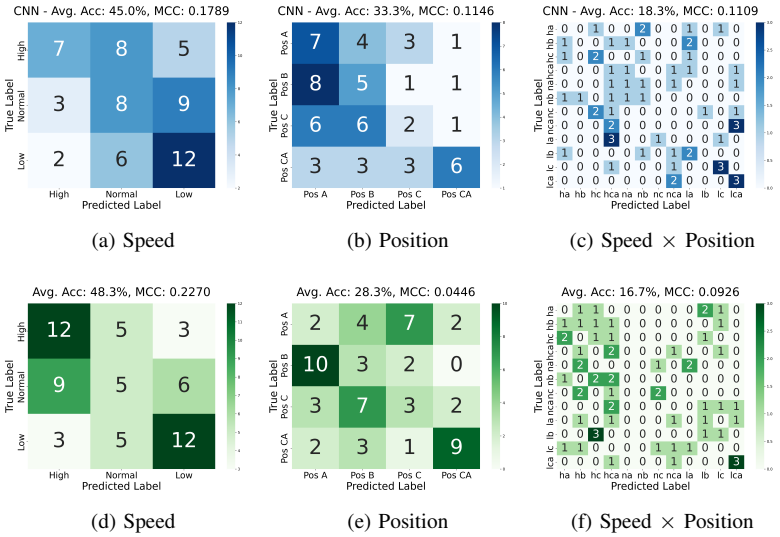
(a) Case A (b) Case B (c) Case C (d) Case CA

(e) Case A (f) Case B (g) Case C (h) Case CA

(i) Case A (j) Case B (k) Case C (l) Case CA

FIGURE 15: STFT profiles for each case, which are input to the post-deep learning scheme for CNN or ViT. 1st row: **Low** speed. 2nd row: **Normal** speed. 3rd row: **High** speed.

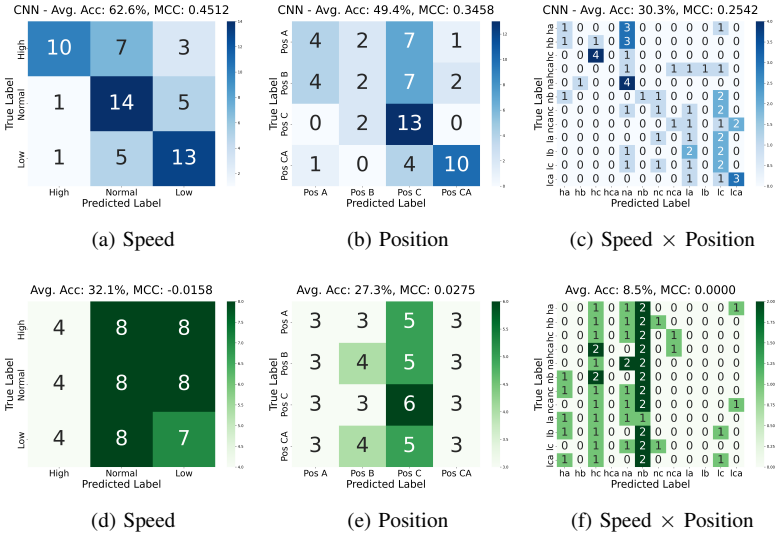
with STFT inputs have been extensively evaluated in the literature (*e.g.*, [24]–[34]) and have demonstrated reasonable performance even with limited training data due to their image compression capabilities. By contrast, ViT extends the Transformer architecture to image recognition, emphasizing global shape features more than CNNs, thereby providing a perception mechanism akin to human cognition. Both approaches are considered as state-of-the-art benchmarks in this field. However, the above conventional STFT-based methods assume that the target range remains approximately constant, with variations on the order of the range resolution. In contrast, in the present experimental model, the target range varies by approximately 2 m over a few seconds, and the range variations of the human body and its individual parts are significantly larger than the range resolution (37 mm). As a result, it is inherently difficult to directly transform the range-τ data into the STFT domain, leading to the so-called ARU problem. Thus, in this section, to address this issue, an $R - v_{\mathrm{d}}$ map is generated at each τ with STFT, and the range $R_{\max}$ corresponding to the maximum energy along the v_{d} direction is extracted. This process is repeated for all τ, and the resulting $R_{\max}$-v_{d} maps are sequentially arranged along the τ axis to construct an STFT input that is consistent with the above conditions. Fig. 15 shows the STFT processed profiles for each case. As shown in results, it is confirmed that distinct profiles are obtained for each stopping position and velocity. However, since the maximum unambiguous velocity range of the STFT is limited to ± 0.015 m/s, it prevents accurate extraction of the motion velocities of the torso and individual body parts. Consequently, severe velocity aliasing occurs, resulting in multiple fold-over artifacts and rendering the data highly ambiguous and difficult to interpret. Figure 16 shows the confusion matrices for three scenarios: (i) velocity only (three patterns), (ii) stopping position only (four patterns), and (iii) joint classification of velocity and stopping position (12 patterns), using CNN and ViT with STFT inputs. The confusion matrices indicate that, compared with the proposed method in Fig. 11, both CNN and ViT fail to achieve satisfactory classification accuracy for velocity, position, and their combination. This performance degradation is mainly attributed to the limited velocity range of the STFT and overfitting caused by high-dimensional image inputs. Furthermore, Figure 17 shows the results when the range-τ profiles in Figs. 8, 9, and 10, are directly used as inputs to CNN and ViT instead of STFT representations. In this case, the classification accuracy, especially for velocity, is improved, likely because velocity can be inferred from features such as the slope of the range-τ profiles. However, the position accuracy remains lower than that of the proposed method. This is presumably due to overfitting and the inherent limitation that range-only information does not provide angular information. Tables 3 and 4 provide comparisons of accuracy and MCC, respectively, between the proposed method and these benchmark models. The proposed method achieves an accuracy of approximately 0.6 and an MCC exceeding 0.5 for the 12-class problem



(a) Speed (b) Position (c) Speed × Position

(d) Speed (e) Position (f) Speed × Position

FIGURE 16: Confusion matrices by STFT-based input with CNN and ViT models. 1st row: STFT + CNN. 2nd row: STFT + ViT.



(a) Speed (b) Position (c) Speed × Position

(d) Speed (e) Position (f) Speed × Position

FIGURE 17: Confusion matrices by Range-tau-based input with CNN and ViT models. 1st row: Range-τ + CNN. 2nd row: Range-τ + ViT.

(using the combination of speed and position), which are significantly superior to all the benchmark model.

One major reason for the inaccuracy of the benchmark model is that models such as ViT require considerable training data to handle high-dimensional inputs such as STFT or range-τ representations. Thus, their performance is limited when utilizing only a moderately sized dataset (5 samples). By contrast, the WKD-based method enables velocity estimation beyond the limitations of STFT and provides time resolution at the PRI level, enabling more accurate and higher temporal resolution range velocity information. Furthermore, the proposed method directly feeds the multidimensional point cloud generated by the k-decomposed WKD, where velocity, range, wavenumber, and intensity are jointly represented, into a Transformer model. This preserves the physical relationships among multiple features while reducing redundant dimensionality. This results in higher classification accuracy than STFT-based or range-τ based high-dimensional input models, even with limited training data.

F. Limitations and Discussions

1) Dependency on Pedestrian Size, Motion Type and Location

Here, we describe the training data limitation in the proposed scheme, in terms of size, location and walk-

TABLE 3: Accuracy for each classification type and model.

	STFT+CNN	STFT+ViT	Range-τ+CNN	Range-τ+ViT	WKD+Transformer (Proposed)
Speed	45.0 %	48.3 %	62.6 %	32.1 %	81.7 %
Position	33.3 %	28.3 %	49.4 %	27.3 %	83.3 %
Speed $\times$ Position	18.3 %	16.7 %	30.3 %	8.5 %	68.3 %

TABLE 4: MCC for each classification type and model.

	STFT+CNN	STFT+ViT	Range-τ+CNN	Range-τ+ViT	WKD+Transformer (Proposed)
Speed	0.179	0.227	0.451	-0.016	0.731
Position	0.115	0.045	0.346	0.02	0.779
Speed $\times$ Position	0.111	0.093	0.254	0.00	0.658

ing directions of the pedestrians, including multi-pedestrian scenarios. In the experimental evaluation presented in this study, walking speed and stopping-position estimations were conducted for single pedestrians at a distance of 4.0 m from the radar. Variations in pedestrian posture, height, and particularly limb length may introduce fluctuations in the observed range and Doppler velocity profiles, even for identical walking speeds.

Although **pedestrian size** can affect extracted features, because the radar primarily captures reflections from body parts within its limited vertical beamwidth, the height variation among adult pedestrians is relatively small (*e.g.*, approximately 170-180 cm, with limb length differences of a few centimeters). This leads to generally consistent range-velocity-τ patterns across individuals. With respect to **different motion patterns**, the proposed method is expected to perform reliably within the range of walking speeds and stopping positions considered in the training data. However, for more complex behaviors, such as pedestrians passing without deceleration, additional training data are necessary to assess the performance of the proposed method.

Because this study focuses on pedestrian scenarios relevant to collision avoidance, particularly pedestrians crossing in front of vehicles (*e.g.*, at crosswalks), the current dataset is limited to crossing motions. Although pedestrians moving in other directions may occur in real environments, these instances are less likely to represent road-crossing behavior and can often be excluded from classification targets through preprocessing of radar images or related filtering techniques. In addition, as long as a pedestrian ' s direction of motion remains consistent, the range resolution and Doppler velocity distribution should yield profiles that are largely independent of the absolute distance between the radar and pedestrian. However, variations in signal-to-noise ratio (SNR) and line-of-sight displacement due to distance may introduce some degree of performance variations.

For **multiple pedestrians**, only those crossing the road can be identified using radar imaging-based clustering and tracking techniques. The proposed k-decomposed WKD processing effectively separates targets in both the range and angular domains. Once targets are adequately resolved in both range and angle, the resulting multidimensional point cloud allows for individual analysis of each target's motion.

2) Effect from Clutter

To manage **background clutter** such as poles, vehicles, and guardrails, separation can be achieved through both angular discrimination and Doppler velocity discrimination. This allows for the distinction between static objects and moving pedestrians [53].

Although this study does not address scenarios involving fully occluded targets (*i.e.*, NLOS conditions), the proposed method may still be applicable in situations where reflected signals are accessible through multipath propagation, as discussed in [54].

Our focus is on pedestrian motion analysis within a forward-looking automotive radar context. Therefore, we designed our experiments to simulate a quasi-outdoor road environment, minimizing strong indoor multipath components such as wall reflections. However, real outdoor environments can present challenges due to static clutter and dynamic multipath effects from additional scatterers, such as vehicles, roadside objects, and ground reflections. In practical automotive radar systems, target localization is first performed using range-angle estimation or radar imaging. Based on this spatial information, a region of interest can be defined for the pedestrian. This enables the suppression of delayed or geometrically inconsistent multipath responses, including ghost targets created by secondary reflections.

3) Effect by Mobile Radar Platform

On the other hand, while the current experiments assume a stationary radar, practical deployment targets vehicle-mounted radar systems, where the radar platform itself moves with a certain velocity toward the target direction. In such cases, it is expected that the range and velocity estimates obtained via the WKD can be compensated using the radar platform ' s own velocity information, enabling classification using training data acquired under stationary conditions. In general, the ego-vehicle velocity can be estimated with high accuracy, and its position can be tracked over time. Therefore, ego-motion effects can be compensated for in the range-τ and Doppler (velocity) domains. For instance, when methods such as [53] are applied, radar images can be aligned across time frames based on their correlation, which enables the estimation of the ego-motion trajectory and adaptation to a moving radar framework.

In our previous work [53], we proposed a classification scheme for mobile radar platforms based on radar image intensity correlation, which enables discrimination between human targets and other objects such as vehicles and other man-made structures (*e.g.*, poles, guardrails). By first identifying human targets using this approach, we can apply the proposed method to analyze their motion in detail, thereby enhancing collision avoidance performance. Further validation of this integrated framework remains an important direction for future research. Overall, since the proposed method does not rely on radar image representations, it enables the estimation of pedestrian walking speed and stopping position without being constrained by horizontal angular resolution, which typically degrades with increasing distance. Compared with conventional radar image-based recognition and classification approaches, this characteristic is expected to significantly broaden the applicability of the proposed method.

IV. Conclusion

This study proposed a multi-dimensional point cloud-based pedestrian activity recognition framework tailored for millimeter-wave (mmW) collision avoidance radar systems. Motivated by the fundamental limitations of conventional Fourier-based range-velocity-angle processing, including ARU range walk issues caused by high range resolution, limited unambiguous velocity ranges, and intrinsic tradeoffs between temporal and velocity resolutions, we introduced a k-decomposed WKD estimator for point cloud based conversion from range-τ to the Doppler velocity-τ points. Through k-v_{d} data decomposition (coherent process), the proposed method effectively mitigates inter-body-part interference and significantly increases the number of reliable point cloud associated with range, velocity, wave-number k, and τ, particularly during walking motions. Furthermore, the transformer-based DL scheme to capture temporal dependencies in pedestrian motion has been introduced with those multi-dimensional point cloud inputs.

Under the experimental validations via 79 GHz MIMO FMCW radar and real pedestrian subject with walking motion, our framework adopts a 12-class classification scheme combining three walking speed states and four trajectory patterns. These results demonstrated that our proposed scheme could achieve accurate estimation of pedestrian walking speed (81.7 %) as well as prediction of the pedestrian's final stopping position (90.0 %) by selecting the appropriate feature vectors. Furthermore, the additional validations also indicated that it is possible to predict the pedestrian's stopping position with velocity even with several seconds ahead using only short-term slow-time observations. Future work will include evaluations with multiple pedestrians of diverse physical characteristics, as well as extensions to vehicle-mounted radar scenarios, where compensation for self-motion should be implemented.

REFERENCES

[1] R. Quintero, I. Parra, D. Fernández-Llorca, and M. A. Sotelo, "Pedestrian path, pose and intention prediction through Gaussian process dynamical models and pedestrian activity recognition", *arXiv preprint*. arXiv:2004.14747, (2020).

[2] J. A. Abbasi, N. M. Imran, L. C. Das, and M. Won, "WatchPed: Pedestrian crossing intention prediction using embedded sensors of smartwatch", *arXiv preprint*. arXiv:2208.07441, (2022).

[3] M. Gustafsson, Å. Andersson, T. Johansson, *et al.,* "Extraction of human micro-doppler signature in an urban environment using a "sensing-behind-the-corner" radar," *IEEE Geosci. Remote Sens. Lett.*, vol. 13, no. 2, pp. 187-191, Feb., 2016.

[4] Q. Wang, Z. Li, H. Zhang, Y. Yang and X. Meng, "An indoor UWB NLOS correction positioning method based on anchor LOS/NLOS map," *IEEE Sensors J.*, vol. 23, no. 24, pp. 30739-30750, Dec., 2023.

[5] S. Li, G. Cui, S. Guo, *et al.*, "On the electromagnetic diffraction propagation model and applications," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 13, pp. 884-895, 2020.

[6] J. He, S. Terashima, H. Yamada and S. Kidera, "Diffraction signal-based human recognition in non-line-of-sight (NLOS) situation for millimeter wave radar," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 14, pp. 4370-4380, 2021.

[7] B. Zhang et al., "Multichannel enhanced millimeter-wave SAR imaging via low-rank tensor-train decomposition," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 18, pp. 1551-1561, 2025.

[8] Y. Arima and A. Hirose, "Millimeter-wave coherent imaging of moving targets by using complex-valued self-organizing map and auto-encoder," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 13, pp. 1784-1797, 2020

[9] D. A. Andrews, S. W. Harmer, N. J. Bowring, N. D. Rezgui, and M. J. Southgate, "Active millimeter wave sensor for standoff concealed threat detection", *IEEE Sensors J.*, Vol. 13, No. 12, pp. 4948–4955, Dec., 2013.

[10] H. Sun, C. Gao, Z. Zhang, X. Liao, X. Wang and J. Yang, "High-resolution anisotropic prestack Kirchhoff dynamic focused beam migration," *IEEE Sensors J.*, vol. 20, no. 20, pp. 11753-11760, 15 Oct., 2020

[11] W. Zhang, P. Wang, N. He and Z. He, "Super resolution DOA based on relative motion for FMCW automotive radar," *IEEE Trans. Veh. Technol.*, vol. 69, no. 8, pp. 8698-8709, Aug. 2020.

[12] A. Y. Nashashibi, T. J. Douglas, M. Kashanianfard, S. W. Decker and K. Sarabandi, "High-resolution polarimetric radar for autonomous vehicle research at W-band frequencies," *IEEE Transactions on Radar Systems*, vol. 2, pp. 632-644, 2024.

[13] M. Farhadi, R. Feger, J. Fink, T. Wagner and A. Stelzer, "Combining MIMO DBF with automotive synthetic aperture radar imaging and phase error correction," *IEEE Access*, vol. 12, pp. 31944-31959, 2024.

[14] Y. Kim, S. Ha, J. Kwon, "Human detection using doppler radar based on physical characteristics of targets", *IEEE Geosci. Remote Sens. Lett.* vol. 12, no. 2, pp. 289 - 293, Feb., 2015.

[15] X. Shi, X. Yao, X. Bai, *et al.*, "Radar echoes simulation of human movements based on MOCAP data and EM calculation", *IEEE Geosci. Remote Sens. Lett.* , vol. 16, no. 6, pp. 859 - 863, Jun., 2019.

[16] Y. Wang and Y. Zheng, "An FMCW radar transceiver chip for object positioning and human limb motion detection", *IEEE Sensors J..,* vol. 17, no. 2, pp. 236-237, Jan., 2017.

[17] X. Li, Y. He, F. Fioranelli, X. Jing, A. Yarovoy and Y. Yang, "Human motion recognition with limited radar micro-doppler signatures," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 8, pp. 6586-6599, Aug. 2021.

[18] G. Li and P. K. Varshney, "Micro-Doppler parameter estimation via parametric sparse representation and pruned orthogonal matching pursuit," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 7, no. 12, pp. 4937-4948, Dec. 2014.

[19] L. Du, B. Wang, P. Wang, Y. Ma and H. Liu, "Noise reduction method based on principal component analysis with beta process for micro-Doppler radar signatures," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 8, no. 8, pp. 4028-4040, Aug. 2015.

[20] P. Khomchuk, I. Stainvas and I. Bilik, "Pedestrian motion direction estimation using simulated automotive MIMO radar," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 52, no. 3, pp. 1132-1145, June, 2016.

[21] S. Zhu, R. G. Guendel, A. Yarovoy and F. Fioranelli, "Continuous human activity recognition with distributed radar sensor networks and CNN–RNN architectures," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, pp. 1-15, 2022.

[22] Y. Kim, I. Alnujaim and D. Oh, "Human activity classification based on point clouds measured by millimeter wave MIMO radar with deep recurrent neural networks," *IEEE Sensors J.*, vol. 21, no. 12, pp. 13522-13529, June, 2021.

[23] F. Qi et al., "Position-information-indexed classifier for improved through-wall detection and classification of human activities using UWB bio-radar," *IEEE Antennas Wirel. Propag. Lett.*, vol. 18, no. 3, pp. 437-441, March 2019.

[24] K. K. Mishra and R. B. Pachori, "Human activity recognition from radar signals based on FBSE-EWT and quantum convolution neural network," *IEEE Sensors Letters*, vol. 9, no. 9, pp. 1-4, Sept. 2025, Art no. 7004604.

[25] R. Lu, Y. Tang, S. Zhang, Y. Li and J. Wang, "Through-wall human activity recognition with compact MIMO ultrawideband radar," *IEEE Internet Things J.*, vol. 12, no. 19, pp. 39762-39776, 1 Oct.1, 2025,

[26] S. Waqar and M. Pätzold, "A simulation-based framework for the design of human activity recognition systems using radar sensors," *IEEE Internet Things J.*, vol. 11, no. 8, pp. 14494-14507, 15 April15, 2024

[27] X. Feng, P. Chen, Y. Weng and H. Zheng, "CMDN: Continuous human activity recognition based on multidomain radar data fusion," *IEEE Sensors J.*, vol. 25, no. 6, pp. 10432-10443, 15 March15, 2025,

[28] R. G. Guendel, N. C. Kruse, F. Fioranelli and A. Yarovoy, "Multipath exploitation for human activity recognition using a radar network," *IEEE Trans. Geosci. Remote Sens.*, vol. 62, pp. 1-13, 2024, Art no. 5103013,

[29] C. Ding, S. Guo, G. Cui, X. Yang, C. Jia and J. Gu, "A non-line-of-sight human activity recognition method based on radar multispectrogram," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 61, no. 5, pp. 13647-13661, Oct. 2025.

[30] D. Wang, C. Zhao, Y. Song and T. Jin, "Crucial region search and feature discrimination for radar-based human activity recognition," *IEEE Internet Things J.*, vol. 12, no. 17, pp. 36141-36154, 1 Sept.1, 2025.

[31] Y. Yang, D. Zhao, X. Yang, B. Li, X. Wang and Y. Lang, "Open-scenario-oriented human gait recognition using radar micro-Doppler signatures," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 60, no. 5, pp. 6420-6432, Oct. 2024.

[32] M. Ding, P. Lv, Y. Peng, G. Dongye and Y. Ding, "A stable gait recognition algorithm under multiview and multiwear using millimeter-wave radar," *IEEE Sensors J.*, vol. 24, no. 22, pp. 38135-38143, 15 Nov.15, 2024.

[33] L. Qu, J. Cong, T. Yang and L. Zhang, "Human tangential activity recognition based on Swin transformer and supervised contrastive learning using interferometric radar," *IEEE Sensors J.*, vol. 25, no. 15, pp. 29189-29200, 1 Aug.1, 2025.

[34] Y. Wang, J. Lv, K. Wako and M. Kim, "Performance enhancement of human activity recognition using millimeter-wave multi-link channels," *IEEE Open J. Antennas Propag.*, vol. 7, no. 1, pp. 170-183, Feb. 2026

[35] C. Pang, S. Liu and Y. Han, "Coherent detection algorithm for radar maneuvering targets based on discrete polynomial-phase transform," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 12, no. 9, pp. 3412-3422, Sept. 2019.

[36] T. Zeng, C. Mao, C. Hu and W. Tian, "Ground-based SAR wide view angle full-field imaging algorithm based on keystone formatting," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 9, no. 6, pp. 2160-2170, June 2016.

[37] X. Chen, J. Guan, Y. Huang, N. Liu and Y. He, "Radon-linear canonical ambiguity function-based detection and estimation method for marine target with micromotion," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 4, pp. 2225-2240, April 2015,

[38] W. Wu, G. H. Wang and J. P. Sun, "Polynomial Radon-polynomial Fourier transform for near space hypersonic maneuvering target detection," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 54, no. 3, pp. 1306-1322, June 2018

[39] W. Wu, D. Liu, J. Sun and G. Wang, "Multihypothesis multimodel elliptic Radon transform for low-observability maneuvering range-ambiguity target detection," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 60, no. 4, pp. 4088-4104, Aug. 2024

[40] X. Peng, Y. Hu, T. Liu, Y. Wu, T. Saito and T. Toda, "Stability-enhanced human activity recognition with a compact few-channel mm-wave FMCW Radar," *IEEE Transactions on Radar Systems*, vol. 3, pp. 360-378, 2025.

[41] Y. Hu, X. Yang, Z. Xia and F. Xu, "Human activity recognition trained on simulated millimeter-wave radar data with domain adaptation," *IEEE Trans. Instrum. Meas.*, vol. 74, pp. 1-13, 2025.

[42] F. Zhang, H. Sun, J. Peng and H. Wang, "LPBS-net: a lightweight network for human activity recognition from sparse millimeter-wave radar point clouds," *IEEE Sensors Letters*, vol. 9, no. 12, pp. 1-4, Dec. 2025, Art no. 6012704.

[43] Y. Wang et al., "Multi-human activity recognition based on sequential 4D point clouds using frequency-modulated continuous wave radar," *IEEE Trans. Veh. Technol.*, vol. 74, no. 10, pp. 15589-15602, Oct. 2025.

[44] M. Setsu, T. Hayashi, J. He and S. Kidera, "Super-resolution doppler velocity estimation by kernel-based range- τ point conversions for UWB short-range radars", *IEEE Trans. Geosci. Remote Sens.*, vol. 58, no. 4, pp. 2430-2443, April 2020.

[45] T. Ando and S. Kidera, "Accurate micro-Doppler analysis by Doppler and k -space decomposition for millimeter wave short-range radar," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 15, pp. 2503-2518, 2022.

[46] G. Gao, Q. Liu, W. Wang, Z. Yu and X. Liu, "Human activity recognition based on 4D millimeter-wave radar," *2025 40th Youth Academic Annual Conference of Chinese Association of Automation (YAC)*, Zhengzhou, China, 2025, pp. 1822-1828

[47] A. Shrestha, H. Li, J. Le Kernec, and F. Fioranelli, "Continuous human activity classification from FMCW radar with Bi-LSTM networks", *IEEE Sensors J.*, vol. 20, no. 22, pp. 13607–13619, Nov. 2020.

[48] H. Li, A. Shrestha, H. Heidari, J. Le Kernec, and F. Fioranelli, "Bi-LSTM network for multimodal continuous human activity recognition and fall detection", *IEEE Sensors J.*, vol. 20, no. 3, pp. 1191–1201, 2020.

[49] S. Chen, W. He, J. Ren, X. Jiang, "Attention-based dual-stream vision transformer for radar gait recognition" *2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2022.

[50] J. Bai, L. Zheng, S. Li, B. Tan, S. Chen, L. Huang, "Radar transformer: an object classification network based on 4D MMW imaging radar", *Sensors*. 21(11):3854, 2021.

[51] V. U. Zavorotny and A. G. Voronovich, "Comparison of geometric optics and diffraction effects in radar scattering from steep and breaking waves", *Proc. IEEE, IEEE International Geoscience and Remote Sensing Symposium, IGARSS 2007*., July, 2007.

[52] B. W. Matthews, "Comparison of the predicted and observed secondary structure of T4 phage lysozyme", *Biochimica et Biophysica Acta (BBA)-Protein Struct.*, vol. 405, no. 2, pp. 442-451, Oct. 1975.

[53] S. Tanaka and S. Kidera, "Radar Image Based Pedestrian Detection Scheme in Mobile Radar Platform for Millimeter Wave Automotive Radar," *IEEE Sensors J.,* 2026.

[54] J. Okazaki, T. Matsuoka, J. He and S. Kidera, "Non-Line-of-Sight Pedestrian Recognition Approach Using Multiple Reflections for Millimeter Wave Radar Sensing Applications," *IEICE Transactions on*

Communications, vol. E108-B, no. 11, pp. 1338-1348, November 2025,

Zihang Zhou received the B.E. degree in electronic information engineering from Changchun University of Science and Technology, Changchun, China, in 2024. He is currently pursuing the M.E. degree with the Graduate School of Informatics and Engineering, University of Electro-Communications, Tokyo, Japan. His research interests include millimeter-wave radar signal processing and multi-dimensional point cloud deep learning for advanced driver assistance systems.

Shouhei Kidera received his B.E. degree in Electrical and Electronic Engineering from Kyoto University in 2003 and M.I. and Ph.D. degrees in Informatics from Kyoto University, Kyoto, Japan, in 2005 and 2007, respectively. In 2009, he joined as an Assistant Professor with the University of Electro-Communications, Tokyo, Japan, where he is currently a full Professor with Graduate School of Informatics and Engineering in the University of Electro-Communications, Tokyo, Japan. His current research interest is in advanced radar signal processing or electromagnetic inverse scattering issue for ultra wideband (UWB) three-dimensional sensor or bio-medical applications. He has been stayed at the Cross-Disciplinary Electromagnetics Laboratory in the University of Wisconsin Madison as the visiting researcher in 2016. He has been a Principal Investigator of the PRESTO Program of Japan Science and Technology Agency (JST) from 2017 to 2021. He was a recipient of the 2012 Ando Incentive Prize for the Study of Electronics, 2013 Young Scientist's Prize by the Japanese Minister of Education, Culture, Sports, Science and Technology (MEXT), and 2014 Funai Achievement Award, 2022 KDDI Foundation Award, Contribution Award, and 2023 RIEC Award. He is a senior member of the Institute of Electronics, Information, and Communication Engineers of Japan (IEICE), and the International Union of Radio Science (Union Radio-Scientifique Internationale, URSI), and a member of the Institute of Electrical Engineering of Japan (IEEJ), and the Japan Society of Applied Physics (JSAP).