Delft University of Technology

Reconstruction of Extended Target Intensity Maps and Velocity Distribution for Human Activity Classification

Kruse, N. C.; Guendel, R.G.; Fioranelli, F.; Yarovoy, A.

DOI
10.1109/TRS.2024.3509775

Publication date
2024

Document Version
Final published version

Published in
IEEE Transactions on Radar Systems

Citation (APA)
Kruse, N. C., Guendel, R. G., Fioranelli, F., & Yarovoy, A. (2024). Reconstruction of Extended Target Intensity Maps and Velocity Distribution for Human Activity Classification. *IEEE Transactions on Radar Systems*, *3 (2025)*, 14-25. https://doi.org/10.1109/TRS.2024.3509775

Important note
To cite this publication, please use the final published version (if applicable).
Please check the document version above.

Copyright
Other than for strictly personal use, it is not permitted to download, forward or distribute the text or part of it, without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license such as Creative Commons.

Takedown policy
Please contact us and provide details if you believe this document breaches copyrights.
We will remove access to the work immediately and investigate your claim.

This work is downloaded from Delft University of Technology.
For technical reasons the number of authors shown on this cover page is limited to a maximum of 10.

Green Open Access added to TU Delft Institutional Repository as part of the Taverne amendment.

More information about this copyright law amendment can be found at https://www.openaccess.nl.

Otherwise as indicated in the copyright section: the publisher is the copyright holder of this work and the author uses the Dutch legislation to make this work public.

Reconstruction of Extended Target Intensity Maps and Velocity Distribution for Human Activity Classification

Nicolas C. Kruse, *Graduate Student Member, IEEE*, Ronny G. Guendel, Francesco Fioranelli, *Senior Member, IEEE*, and Alexander Yarovoy, *Fellow, IEEE*

***Abstract*— The problem of human activity classification using a distributed network of radar sensors has been considered. A novel sensor fusion method has been proposed that processes data from a network of radar sensors and yields 3-D representations of both reflection intensity and velocity distribution. The formulated method has been verified in an experimental case study, where activity classification was performed using data collected with 14 participants moving in diverse, unconstrained trajectories and executing nine activities. The classification performance of the proposed method has been compared to alternative fusion methods on the same dataset, and a test accuracy and macro $F1$-score of, respectively, 87.4% and 81.9% have been demonstrated. A feasibility study has also been performed to demonstrate the ability of the proposed method to generate 3-D distributions of intensity and target velocity.**

***Index Terms*— Extended target, human activity classification, radar network, sensor fusion.**

I. Introduction

WITHIN the context of healthcare, radar is considered a promising sensor modality for the monitoring of patients and vulnerable individuals in their home environments. The monitoring capabilities of radar include vital sign estimation [1], [2], gait analysis [3], [4], fall detection [5], [6], gesture recognition for interaction with smart devices and automatic sign language interpretation [7], [8], [9], [10], and activity classification [11], [12], [13], [14]. The noncontact nature of radar sensing allows for monitoring in situations where wearable sensors would prove disadvantageous, such as in cases where subjects may forget or object to wearing a sensor. Additionally, radar is an active sensor and functions in complete darkness or glaring lights, with no hindrance to the user. Finally, no visual images are captured, which can be beneficial in terms of perceived privacy from the side of the end-users.

Received 12 July 2024; revised 30 September 2024 and 20 November 2024; accepted 26 November 2024. Date of publication 2 December 2024; date of current version 9 December 2024. This work was supported in part by Nederlandse Organisatie voor Wetenschappelijk Onderzoek (NWO) KLEIN Radar-aware Activity Recognition with Innovative Temporal Networks (RAD-ART) Project. *(Corresponding author: Nicolas C. Kruse.)*

This work involved human subjects or animals in its research. Approval of all ethical and experimental procedures and protocols was granted by the TU Delft HREC under Application No. 1387.

Nicolas C. Kruse, Francesco Fioranelli, and Alexander Yarovoy are with the Microwave Sensing, Signals and Systems (MS3) Group, Delft University of Technology, 2628 CD Delft, The Netherlands (e-mail: n.c.kruse@tudelft.nl; f.fioranelli@tudelft.nl; a.yarovoy@tudelft.nl).

Ronny G. Guendel is with the Netherlands Organization for Applied Scientific Research (TNO), 2509 JG The Hague, The Netherlands (e-mail: r.g.guendel@tudelft.nl).

Digital Object Identifier 10.1109/TRS.2024.3509775

To aid in activity classification, sensor fusion has been utilized to great effect in the literature. The combination of data from multiple heterogeneous sensors [15], [16] or from multiple radar sensors operating in a network [17], [18], [19] allows them to complement each other and compensate for weaknesses of individual sensors. For the case of radar in particular, the most precise position and velocity measurements are generally in the radial direction. This suggests that a network of radars can improve the perception of the location and velocity distribution of different body parts of the observed subject compared to a single radar, provided that the network is sufficiently spatially diverse.

With the proliferation of machine learning (ML) techniques, sensor fusion approaches often rely on a fused representation that constitutes a latent space in an ML model [15], [17], [19], [20]. In other words, the feature space after fusion is no longer easily human-interpretable. These fused feature spaces are optimized for a specific task and mostly do not generalize well to other applications. Other common fusion types are decision-based, where predictions from multiple classifiers are merged on the basis of, e.g., prediction confidence [14], [17]. A disadvantage of these approaches is that the individual classifiers do not have access to a representation with information from all sensors, thus potentially limiting performance. Finally, simple fusion approaches are utilized where the fused representation is a concatenation of radar data domains from multiple sensors [13], [18], [19]. In these cases, no data association between sensors is performed.

Various fusion approaches for radar networks in the literature for human activity classification have been investigated. To the best of the authors' knowledge, there is no method that aims to explicitly model and estimate both reflection intensity and velocity distribution of the target in 3-D. This capability would be important to combine interpretability and task versatility. Interpretability comes from a shared data representation from different nodes that are based on intensity and Doppler/velocity, both of which are quantities that are easily understandable and related to the kinematics of the observed activities. Task versatility is here defined as the potential to utilize the fused representations for multiple applications, for example, target tracking or activity classification.

In this work, a novel sensor fusion method is proposed that processes raw data from a network of radar sensors and yields 3-D representations of both reflection intensity and velocity distribution. Specifically, range–angle–Doppler representations of data from a network of distributed monostatic

2832-7357 © 2024 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission.
See https://www.ieee.org/publications/rights/index.html for more information.

Authorized licensed use limited to: TU Delft Library. Downloaded on August 07,2025 at 09:39:50 UTC from IEEE Xplore. Restrictions apply.

radar are processed into two 3-D fields defined in Cartesian coordinates. The first of these fields contains the reconstructed reflection intensity at each point in a 3-D spatial grid; the second is a vector field of reconstructed velocities, in the context of human activity classification related to the combined movement of different body parts. The efficacy of the proposed fusion method is evaluated through a classification case study. A challenging, publicly available dataset of continuous human activities is processed using the proposed method. The fused intensity maps and velocity fields are then used as input to a convolutional neural network-bidirectional long short-term memory (CNN-BiLSTM) architecture, tasked with discriminating nine different human activities. Additionally, an experimental feasibility study is performed to demonstrate the ability of the proposed method to yield 3-D representations of extended target shape and velocity distribution. This work is an extension of the preliminary analysis presented in [21]. Specifically, two major additions are the development of a classification pipeline to demonstrate the accuracy of the proposed fusion method, and an extension of the proposed method to yield a full 3-D representation of reflection intensity and velocity distribution. In addition, an extensive study on the number and position of radar nodes in the network is performed to evaluate their effect on classification performance.

The remainder of this work is organized as follows. In Section II, the proposed sensor fusion method for reconstruction of reflection intensity and velocity profile of human movements is explained, followed by a description of the 2-D case study used to validate the method in Section III. Results of this case study are presented and discussed in detail in Section IV. The 3-D feasibility study is presented in Section V, and conclusions follow in Section VI.

II. Proposed Method

The fusion method proposed in this work comprises two main elements: a voxelization of the observed human target into a 3-D spatial grid and a reconstruction technique of the measured reflection intensity and dominant velocity in each voxel. The method is agnostic to the angular capabilities of the radar sensors utilized and functions with any number of sensors, provided that this exceeds the amount of spatial dimensions considered for the problem. At least four spatially distributed radar nodes are needed to reconstruct intensity and velocity distribution in a 3-D grid, in the case that angular information is not available.

Inputs to the proposed fusion method are N radar data tensors from a network of N radar sensors. The tensors represent measured signal amplitude and have dimensions of range, Doppler (radial velocity), and azimuth and elevation if the radar sensors are capable of measuring them. The outputs of the method are a spatial distribution of reflection intensity as well as a vector field of reconstructed velocities.

It is assumed that the human body is a nonrigid, extended target. For the purposes of activity classification, the simplification to a point target with a single velocity is too limiting. Although the distribution of radial velocities of the human target can be measured by multiple radar systems, an association problem prevents the determination of a 3-D velocity profile through direct sensor fusion. The human body is, however, subject to kinematic constraints. The allowed joint movements are principally rotational in nature, which implies that the velocity profile along the limbs will vary smoothly. Given this implication, it follows that a volume element that is small in comparison to the human target can be assumed to contain a dominant velocity. This assumption underlies the velocity reconstruction aspect of the proposed method. It should be noted that the assumption does not hold in the cases where, e.g., two limbs pass in close proximity relative to the size of the volume element. In these cases, volume elements may contain multiple dominant velocities.

The proposed fusion method takes radar data tensors $\mathcal{D}(r, \theta, \phi, v)$ from a network of radar sensors as input, where azimuth θ and elevation ϕ are optional, depending on the front-end architecture of the considered radar. Dimensions r and v represent the range and Doppler, respectively. As this information can be provided by a variety of radar systems with varying processing approaches, no specific architecture or typology of radar is assumed, provided that Doppler shifts and so radial velocity components can be measured. In the specific case of a pulsed single-channel system, with a complex-valued fast-time/slow-time matrix $\mathcal{S}(t_f, t_s)$, the radar data tensor is acquired through the application of a discrete Fourier transform (DFT) along the slow-time dimension

$$\mathcal{D}(r, v) = \mathcal{F}_{t_s}\left[\mathcal{S}\left(t_f, t_s\right)\right]. \tag{1}$$

Here, $\mathcal{F}_{t_s}$ is the DFT operation along the slow-time dimension t_s.

The set of radar sensors will be indicated by $\mathcal{N}$, and the cardinality of this set is denoted by $N = |\mathcal{N}|$. In the remainder of this section, vectors and unit vectors will be indicated with arrows ($\vec{p}$) and hat operators ($\hat{p}$), respectively. Thus, $\vec{p} = p\hat{p}$ and $p = \vec{p} \cdot \hat{p}$. An estimated vector is indicated with a tilde ($\tilde{p}$).

Ethics approval for this article was granted by the TU Delft HREC (ID 1387).

A. Intensity Maps

As a first step to the proposed method, 3-D space in the field of view of the radar network is uniformly voxelized in Cartesian coordinates, resulting in a grid with a spacing that can be adjusted based on, e.g., the resolution characteristics of the radar sensors utilized. For each volume element in the grid, the range to each sensor is computed, as well as the azimuth and elevation angles if available for the radar sensors used. These computations are achieved through a coordinate transfer mapping for each individual sensor.

Consider the center of a volume element in Cartesian coordinates $\vec{x} = [x, y, z]$. The vector is first translated to place the sensor n in the origin and is rotated such that the sensor boresight is in the positive x-direction, as

$$\vec{x}'_n = R_n^{-1}\left(\vec{x} - \vec{x}_n\right) \tag{2}$$

where $\vec{x}_n$ is the position of sensor n and

$$\begin{aligned} R_n &= R(\theta_{Az})R(\phi_{El}) \\ &= \begin{bmatrix} \cos\theta_{Az}\cos\phi_{El} & -\sin\theta_{Az} & \cos\theta_{Az}\sin\phi_{El} \\ \sin\theta_{Az}\cos\phi_{El} & \cos\theta_{Az} & \sin\theta_{Az}\sin\phi_{El} \\ -\sin\phi_{El} & 0 & \cos\phi_{El} \end{bmatrix}. \end{aligned} \tag{3}$$

Authorized licensed use limited to: TU Delft Library. Downloaded on August 07,2025 at 09:39:50 UTC from IEEE Xplore. Restrictions apply.

Here, the rotation matrix R_n for the sensor n is defined as a function of the boresight elevation ϕ_{El} and azimuth angle θ_{Az} with respect to the positive x-axis. Subsequently, a mapping $f : \mathbb{R}^3 \rightarrow \mathbb{R}^3$ is employed to transform to a spherical coordinate system: $f(\vec{x}) = \vec{r} = [r, \theta, \phi]$. The transformation is given by

$$r = ||\vec{x}|| \tag{4}$$

$$\theta = \arctan \frac{y}{x}, \quad x > 0 \tag{5}$$

$$\phi = \arcsin \frac{z}{||\vec{x}||}, \quad x > 0. \tag{6}$$

Note that the angles ϕ and θ are limited to the hemisphere in the positive x-direction.

After transforming the grid to the frames of the individual sensors as in (2) and applying the aforementioned mapping into spherical coordinates, the respective radar data tensors can be evaluated to yield the measured signal amplitude at each volume element. As the transformed grid may not align with the range–angle map of a sensor, the signal amplitude of a volume element is taken to be that of the closest data point in the sensors' range–angle map. The individual tensors are of dimension $\mathcal{D}(r, \theta, \phi, v)$, and reflection intensity for each volume element $\vec{x}$ is determined by summing received signal amplitude over the radial velocity dimension v. Contributions from the set of radar sensors $\mathcal{N}$ are then summed up, as

$$I(\vec{x}) = \sum_{n \in \mathcal{N}} \sum_{v} \mathcal{D}\big(f(\vec{x}'_n), v\big). \tag{7}$$

By repeating this process for each volume element in the grid, an intensity map I can be constructed.

B. Velocity Field Reconstruction

The reconstruction of the velocity field for the extended target starts with the same grid definition and transformation as outlined in Section II-A. In contrast to the intensity map computation, the dominant Doppler/velocity component is recorded for each voxel in the grid, and for each radar sensor in the network. For this work, the dominant Doppler component is considered to be the Doppler index with the highest amplitude, as

$$v_n = v_n(r, \theta, \phi) = \arg\max_v(\mathcal{D}(r, \theta, \phi, v)). \tag{8}$$

The resultant set of N Doppler components from the set of radar sensors is assumed to originate from a dominant velocity in the volume element. In order to reconstruct this full 3-D velocity from the N projections, i.e., the projections on the line of sight of each radar, the process of orthogonal projection is inverted by means of a minimization problem.

First, consider the generic plane defined by the vectors $\vec{p}$ for which the following holds:

$$\vec{p} \cdot \hat{n} + c_i = 0. \tag{9}$$

Here, $\hat{n}$ is a unit vector orthogonal to the plane and c_i a constant. For the specific case of a Doppler/velocity projection seen by a radar sensor, the true target velocity $\vec{u}$ projects onto the sensor's line of sight $\hat{x}'_n$, yielding $\vec{v}_n$ as shown in Fig. 1. It follows that all possible $\vec{u}$ are given by the vectors $\vec{u} - \vec{v}_n$

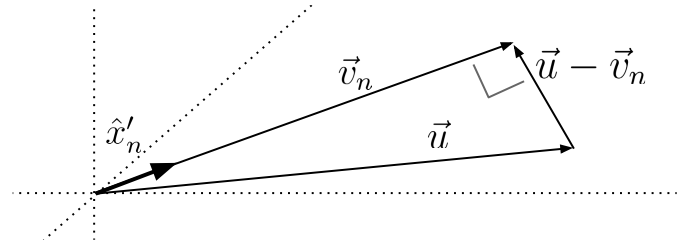


Fig. 1. Vector representation of quantities relevant for velocity reconstruction. $\hat{x}'_n$: line-of-sight unit vector from radar n to voxel at $\vec{x}$; $\vec{u}$: true velocity in voxel at $\vec{x}$; and $\vec{v}_n$: radial velocity measured by sensor n.

that are in the plane that is orthogonal to $\hat{x}'_n$. Since v_n is known and $v_n = \vec{v}_n \cdot \hat{x}'_n$, $\vec{v}_n$ can be substituted for $\vec{p}$ and $\hat{x}'_n$ for $\hat{n}$ in (9) to solve for the constant $c_{i,n}$

$$\vec{v}_n \cdot \hat{x}'_n + c_{i,n} = 0 \tag{10}$$

$$c_{i,n} = -v_n. \tag{11}$$

A set of N planes can now be defined that, in an ideal case, would intersect in a single point $\vec{u}$

$$\vec{p} \cdot \hat{x}'_n - v_n = 0 \tag{12}$$

where $\vec{p} \in \mathbb{R}^3$ and $n \in \mathcal{N}$.

In the nonideal case of real targets, an optimization problem can be set up to find the point with minimum distance to the set of N planes, i.e., to determine the best fitting intersection point $\tilde{u}$. First, a cost function is defined as the sum of squared distances of a point $\vec{p}$ to the set of planes as per (12). This is formulated as follows:

$$E(\vec{p}) = \sum_{n \in \mathcal{N}} \big(\vec{p} \cdot \hat{x}'_n - v_n\big)^2 \tag{13a}$$

$$= \sum_{n \in \mathcal{N}} \Big(\big(\vec{p} \cdot \hat{x}'_n\big)^2 - 2\vec{p} \cdot v_n \hat{x}'_n + v_n^2\Big) \tag{13b}$$

$$= \sum_{n \in \mathcal{N}} \big(\big(\vec{p}^{\mathrm{T}} \hat{x}'_n \hat{x}'^{\mathrm{T}}_n \vec{p}\big) - 2\vec{p}^T v_n \hat{x}'_n + v_n^2\big). \tag{13c}$$

Matrix notation is used from (13c) onward for the sake of clarity. The derivative of the cost function with respect to $\vec{p}$ is given by

$$\frac{dE}{d\vec{p}} = 2 \sum_{n \in \mathcal{N}} \hat{x}'_n \hat{x}'^{\mathrm{T}}_n \vec{p} - 2 \sum_{n \in \mathcal{N}} v_n \hat{x}'_n \tag{14}$$

and setting this derivative to 0 yields the best fitting intersection point, namely, the vector $\tilde{u}$, as

$$\sum_{n \in \mathcal{N}} \hat{x}'_n \hat{x}'^{\mathrm{T}}_n \tilde{u} - \sum_{n \in \mathcal{N}} \vec{v}_n = 0 \tag{15a}$$

$$\tilde{u} = \left(\sum_{n \in \mathcal{N}} \hat{x}'_n \hat{x}'^{\mathrm{T}}_n\right)^{-1} \sum_{n \in \mathcal{N}} \vec{v}_n. \tag{15b}$$

The resulting vector $\tilde{u}$ is the assumed dominant velocity in the voxel. This velocity reconstruction can be applied to all voxels in the considered volume of interest of the radar network, thus yielding a vector field $\tilde{u}(\vec{x})$. The velocity field can in principle be computed for all voxels, including those with only noise. Hence, it can be decided to only compute, or visualize, the field at those voxels where the corresponding intensity map exceeds a user-defined threshold, i.e., $I(\vec{x}) > \epsilon$.

Authorized licensed use limited to: TU Delft Library. Downloaded on August 07,2025 at 09:39:50 UTC from IEEE Xplore. Restrictions apply.

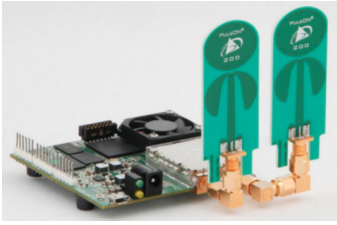


Fig. 2. Single humatics PulsON P410 radar sensor with antennas.

After the computation of the scalar reflection intensity map and velocity vector field at all voxels, the final representation generated by the proposed method is given by both $I(\vec{x})$ and $\tilde{u}(\vec{x})$. This representation constitutes a fusion of the data from the N sensors into three spatial dimensions and can be subsequently used for tasks such as classification of activities.

III. Experimental Case Study in 2-D

To demonstrate the efficacy of the proposed method, an experimental case study is conducted. The objective of this study is to gauge the accuracy of the reconstructed velocity distribution and reflection intensity maps for human activity classification tasks. Specifically, the intensity map and velocity field are computed according to the method described in Section II and used as input to a deep learning model with the aim of classifying different human activities. The dataset used for this case study is publicly available [22] and features 120-s sequences of nine different human activities of unconstrained duration and direction. The data are captured with a network of five pulsed ultra-wideband (UWB) radar sensors, operating as distributed monostatic nodes. For classification, a hybrid CNN-BiLSTM architecture is employed to process the fused representations of intensity maps and velocity fields.

A. Radar Operating Characteristics

The radar system utilized for data capture is a set of five time-domain PulsON P410 pulsed UWB sensors [23], as shown in Fig. 2. For interference mitigation, the sensors feature a coded, pseudorandom pulse repetition interval (PRI) variation, paired with coherent pulse integration [23], [24]. Returns from other nodes, thus, only represent an elevated noise floor. This coding allows simultaneous operation of all sensor nodes, essentially operating the radar network as an ensemble of distributed monostatic nodes. The nodes operate at a center frequency of 4.3 GHz and feature a bandwidth of 2.2 GHz, resulting in a range resolution of approximately 6.8 cm. A pulse repetition frequency (PRF) of 122 Hz is maintained, which gives a maximum unambiguous velocity of $\pm 2.13\,\mathrm{m{\cdot}s^{-1}}$. All nodes are single-input-single-output (SISO) and feature an antenna pattern that is symmetric in azimuth. As such, no information is available on the angle of arrival of target reflections.

B. Sensor Geometry

The location of the five radar sensors is shown in Figs. 3 and 4. All sensors are placed on a semicircle of diameter 6.38 m, at regular 45° intervals. The measurement area is the concentric circle of diameter 4.38 m.

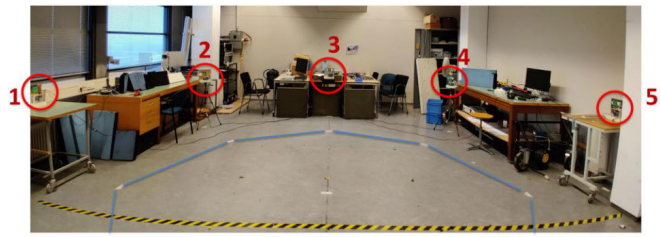


Fig. 3. Experimental setup used in the case study. Five radar sensors are arranged in a semicircle at 45° intervals.

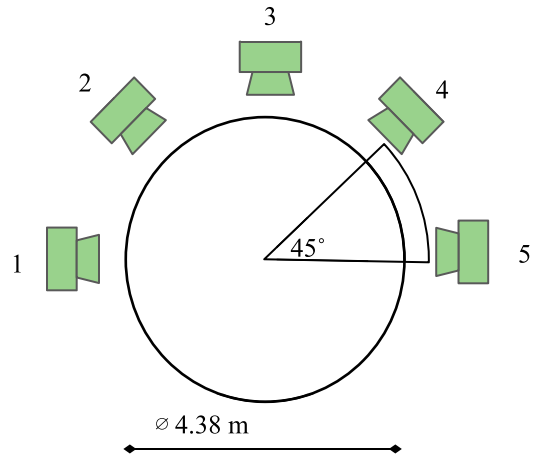


Fig. 4. Schematic representation of the experimental setup used in the case study, with the network of five radar nodes.

Since the proposed method is inherently based on the fusion of data from different radar nodes, various sensor network geometries are evaluated. As the dataset has been collected with the fixed five-node geometry from Fig. 4, only subsets of these nodes can be evaluated. These subsets can be realized by leaving data from specific nodes out during the processing. Due to computational constraints, not all subsets that are possible with five nodes have been tested experimentally. Hence, subsets that are in the same rotational symmetry group, or that exhibit a bilateral symmetry along a horizontal axis, are represented in this study by a single member of the respective symmetry group. This is done under the assumption that the activities in the experimental scene were performed in fully random trajectories so that there is minimal bias in activity location or orientation. As an example, the configuration $(1, 2, 4)$ is rotationally symmetric to the configuration $(2, 3, 5)$, and only the former is included in the experiment. As an exception to this, configurations $(1, 3)$ and $(2, 4)$ are both included with the aim of gauging any bias in the location and direction of activities in the dataset. All the geometries that are evaluated in this work are displayed in Fig. 5.

C. Sequences of Human Activities

The dataset contains sequences of continuous human activities of unconstrained duration. A total of 14 participants are recorded performing 30 sequences of 120 s each, yielding 840 min of data in 420 sequences. In total, nine different activities are performed, as listed in Table I. The sequence types are diverse in the order and combination of the nine activities, including walking around the measurement area

Authorized licensed use limited to: TU Delft Library. Downloaded on August 07,2025 at 09:39:50 UTC from IEEE Xplore. Restrictions apply.

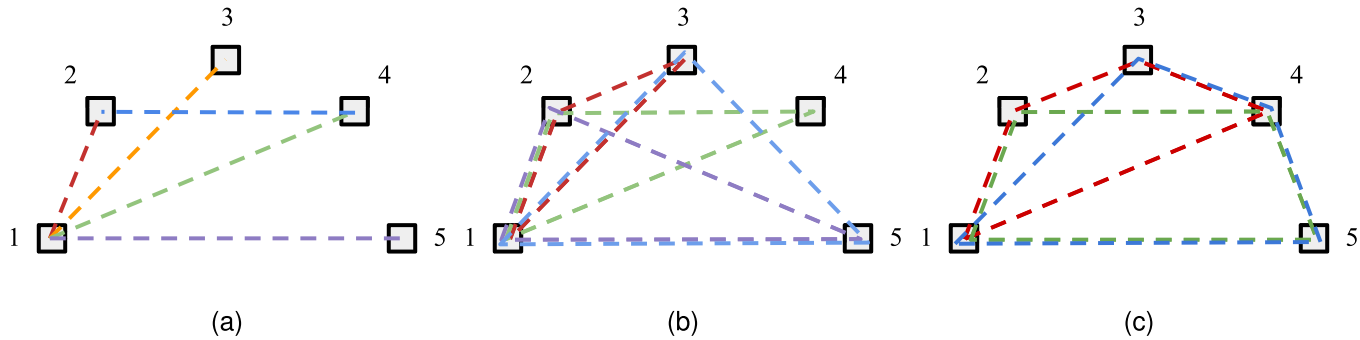


Fig. 5. Schematic representation of all the evaluated sensor network geometries with the five available radar nodes in the network. Each considered geometry is color-coded, and the constituent sensors are connected by dashed lines. (a) Two-node geometries. (b) Three-node geometries. (c) Four-node geometries.

TABLE I
DISTRIBUTION OF SAMPLES AMONG THE NINE ACTIVITY CLASSES

Activity	Fraction
Walk	43.12%
Stationary	11.52%
Sit Down on Chair	4.33%
Stand Up (From sitting)	3.96%
Bending (Sitting)	9.87%
Bending (From Standing)	10.51%
Falling (From Walking)	2.81%
Falling (From Standing)	4.39%
Standing Up (After Fall)	9.48%
Total	100%

and falling at various locations and orientations. "Mixed" sequences containing all nine activities performed in random order and locations are also included. The ground-truth labels are created by the participants themselves by means of a handheld remote control. A full description of all sequence types is included with the dataset [22].

D. Preprocessing

Preprocessing is performed on the data of each radar sensor prior to the implementation of the proposed method described in Section II. First, the real-valued vector output by each of the sensors is Hilbert-transformed, and a fast-time/slow-time matrix with complex values is constructed, which is essentially considered as a range–time matrix. The slow-time indices corresponding to one coherent processing interval T_{CPI} are then selected from this larger matrix for Doppler processing. Finally, fast Fourier transform (FFT) along the slow-time dimension is performed to obtain a range–Doppler representation. This process is repeated along the sequence of original data with constant T_{CPI} to generate a sequence of range–Doppler matrices. As no angular information can be extracted from the data of a single radar, the tensor defined in Section II is in this case only represented in two dimensions, as: $\mathcal{D}(r, \theta, \phi, v) = \mathcal{D}(r, v)$. Hence, the proposed method will operate in this case on bidimensional range–velocity matrices $\mathcal{D}(r, v)$.

E. Method Parameters and Classification Approach

Since the radar sensors offer no angular information and are located in the same horizontal plane, it is decided to consider only the horizontal xy plane for this study. The vertical z-direction is, thus, effectively projected onto the horizontal plane. It is hypothesized that in the recorded scenarios enough movement occurs in this horizontal plane in order to effectively perform activity classification, i.e., the activities can be distinguished by their horizontal velocity and reflection intensity distributions.

For the case study, the method parameters employed for obtaining the intensity map and velocity field are specified as follows. The coherent processing interval T_{CPI} is set to 0.26 s, which at the PRF of 122 Hz corresponds to 32 slow-time samples. This value is based on proven effectiveness in [25] and balances Doppler and time resolution for the highly dynamic nature of human motions. In order to process the full 120-s sequences, a series of intensity maps and velocity fields are computed. These are spaced at 32 slow-time sample intervals, i.e., one per coherent processing interval without overlap. This interval will hereafter be referred to as a "time step." The measurement area is divided into a grid of 15 × 15 cm cells. As mentioned before, the range resolution of each sensor is about 7 cm and the grid size is selected as a compromise between reconstruction accuracy and computation time. Finally, the velocity field that is computed at each step is masked using the intensity map. This ensures that only velocities associated with the presence of a detected target are used as input to the classification pipeline. Specifically, the velocity vectors that are kept are those at the coordinates where the intensity values exceed the 95th percentile of the intensity values distribution for that time step. The size of the extended target after applying a threshold between the 90th and 99th percentiles remains largely unchanged, due to a relatively steep intensity boundary.

The classification approach aims to utilize a hybrid CNN-BiLSTM model. The CNN first extracts features from the 2-D intensity map and velocity field for each time step, i.e., each CPI. Subsequently, the BiLSTM network performs the prediction step with access to the time series of feature vectors of the full sequence. The CNN-BiLSTM architecture benefits from the translation equivariance of the CNN to effectively extract features from a spatial domain. In the time domain, the BiLSTM can process the extracted features as a time series spanning the full duration of the activity sequence. Such hybrid spatial–temporal architectures are used effectively in the literature [7], [26], [27], and the model for this study is

Authorized licensed use limited to: TU Delft Library. Downloaded on August 07,2025 at 09:39:50 UTC from IEEE Xplore. Restrictions apply.

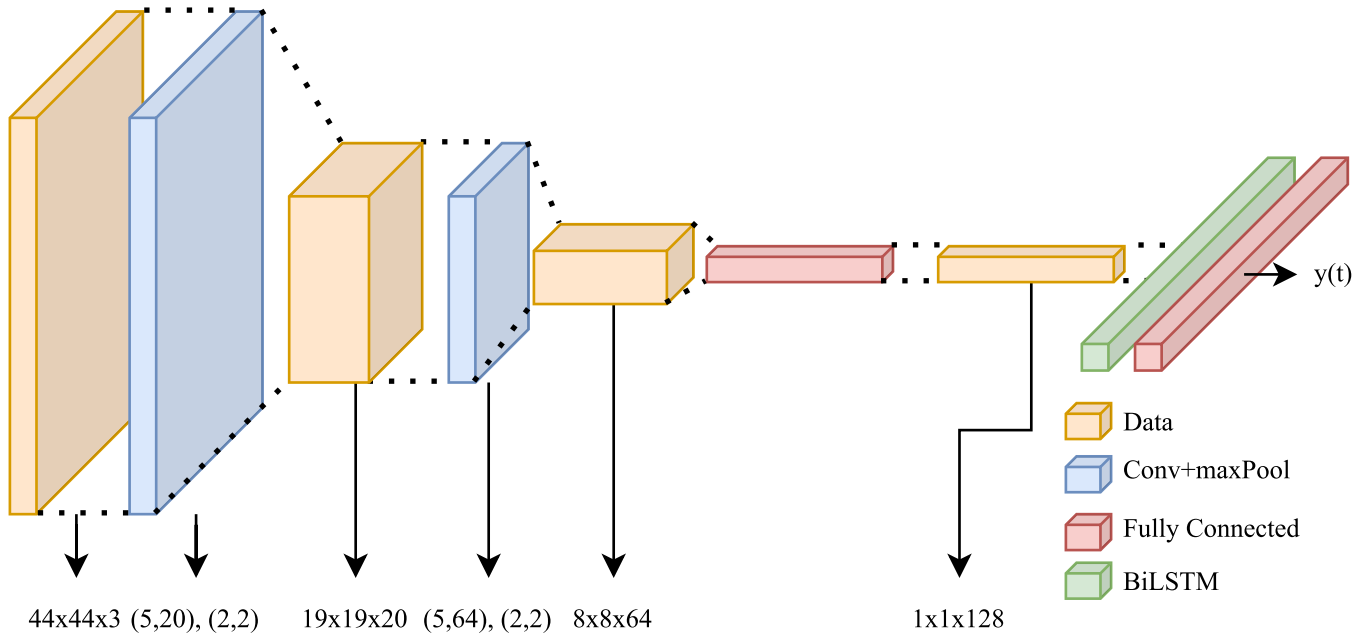


Fig. 6. Schematic representation of the hybrid CNN-BiLSTM network utilized in the case study. Data are represented as 3-D tensors with dimension Width × Height × Channels/Depth. The size of convolutional and maxpooling layers is shown as (CNN kernel size, channels) and (maxpooling kernel size, maxpooling stride size). The output after the last fully connected layer is a time series of class predictions $y(t)$.

inspired by Zhu et al. [20]. The hybrid architecture is shown schematically in Fig. 6.

The velocity vector field generated by the proposed method at each time step is rewritten as two scalar 2-D fields of magnitude and angle, the latter confined in the interval $[-\pi, \pi]$. Together with the corresponding intensity map, they are stacked, forming a 44 × 44 ×3 input array. This input size is specifically the result of the chosen 15 × 15 cm grid cell size, and the 6.38-m baseline of the sensor network. The activities in the dataset are performed in all directions and at random locations. To further mitigate potential bias in the orientation of the movement, all inputs are randomly rotated in the horizontal plane by integer multiples of 90°. Within the network, two sequential convolutional blocks extract spatial features and are paired with maxpooling layers for dimensionality reduction. The hyperparameters for the two convolutional blocks are shown in Fig. 6 and are based on the LeNet-5 architecture [28]. A fully connected layer subsequently yields a feature vector of 128 elements for this time step. Finally, a BiLSTM layer with 168 hidden units processes all feature vectors and yields a class prediction at each time step. The number of hidden units and the feature vector length are based on [18] and [20], respectively.

For training and validation, a leave-one-person-out (L1PO) scheme is adopted. Data from one participant are kept as a testing set, using the remaining participants' data for training the model. This process is repeated for all participants.

IV. Experimental Results and Discussion

In this section, the results of the case study will be discussed, namely: an example of a typical intensity map and velocity field, the L1PO benchmark result, and the results of the study on sensor geometry. Despite the lack of angular capabilities of the individual radar sensors, favorable results are achieved, thanks to the proposed method that leverages data from the radar network. The results are then compared to alternative methods in the literature in Section IV-B. Finally, the results are further discussed in the form of a classification error analysis, and computational considerations will be outlined.

In Fig. 7, a typical intensity map and velocity vector field output are shown for the proposed method. The five sensors are marked with red squares, and the color scale is normalized to the maximum intensity value. The subject can be seen at the experimental area center point and is in this CPI moving in the negative y-direction. At half of the maximum intensity, it can be seen that the target takes up a space of approximately 60 ×30 cm. Based on the target orientation, these measurements correspond to expectation, i.e., approximately 60-cm shoulder width and 30-cm torso depth. The reconstructed velocities also correspond to the motion performed at this specific moment in time, with the vector field representing the bulk torso motion in the negative y-direction. A processing artifact can be seen in the form of a ring centered around middle sensor 3, with a section of this ring contained in the dashed ellipse. This artifact is the result of the relatively strong contribution of sensor node 3 in this particular situation, paired with the inability of differentiating returns from multiple azimuth directions.

The results for the L1PO testing approach are summarized in Fig. 8. Test accuracy and macro $F1$-score are shown for each participant in the dataset, averaged over all sequences. The average test accuracy and macro $F1$-score over all participants are also indicated with horizontal lines. Two notable

Authorized licensed use limited to: TU Delft Library. Downloaded on August 07,2025 at 09:39:50 UTC from IEEE Xplore. Restrictions apply.

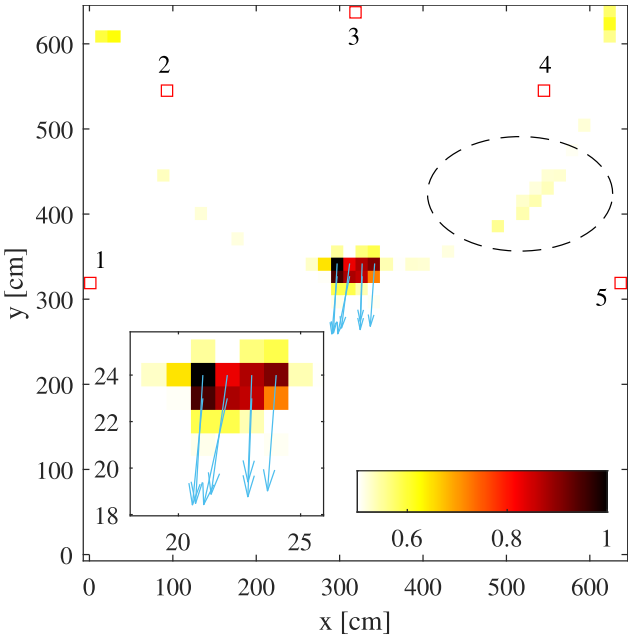


Fig. 7. Typical intensity map and velocity vector field output with zoomed-in view inset on the human target area. Sensors are numbered and marked with red squares. The human subject is located in the center and is moving in the negative y-direction at this specific CPI. Velocity vectors are only shown where reflection intensity exceeds 80% of the maximum. A dashed ellipse is used to show processing artifacts.

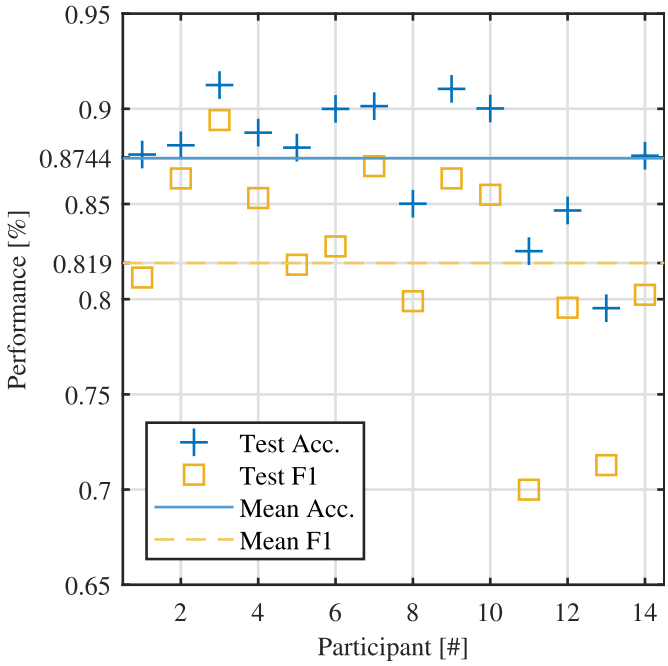


Fig. 8. Results for the classification task, evaluated using an L1PO testing scheme. Test accuracy and macro $F1$-score results are shown for each participant. Averages are indicated with horizontal lines.

outliers with low $F1$-score can be identified in Fig. 8, namely, participants 11 and 13, with respect to the other participants. Inspection of the test results for these participants reveals that the excess errors primarily stem from transition events between, e.g., walking and a subsequent fall. As the ground truth is recorded by the participants themselves by clicking on a remote controller, there is variability in when a transition is indicated. For the two outliers, it is noted that this variation is relatively large compared to the remaining 12 participants, leading to increased ambiguity. Despite the outliers, the average performance metrics show the suitability of the method for classification. Section IV-B compares the attained result with alternative methods in the literature on the same dataset, demonstrating improved classification performance of the proposed method.

TABLE II
RESULTS OF THE STUDY ON SENSOR NETWORK GEOMETRY. THE UTILIZED TWO-, THREE-, AND FOUR-NODE GEOMETRIES CAN BE SEEN IN FIG. 5(A)–(C), RESPECTIVELY

Nodes	Test Accuracy	Macro F1-score
(1,2)	0.758	0.631
(1,3)	0.768	0.647
(1,4)	0.791	0.677
(1,5)	0.797	0.691
(2,4)	0.773	0.643
Mean	**0.777±0.016**	**0.658±0.025**
(1,2,3)	0.804	0.731
(1,2,4)	0.812	0.709
(1,2,5)	0.821	0.723
(1,3,5)	0.834	0.756
Mean	**0.822±0.011**	**0.729±0.024**
(1,2,3,4)	0.834	0.748
(1,2,4,5)	0.848	0.775
(1,3,4,5)	0.843	0.767
Mean	**0.842±0.007**	**0.763±0.014**
(1,2,3,4,5)	0.866	0.804

A. Sensor Geometry Evaluation

A summary of the experiments pertaining to the sensor network geometry is shown in Table II, where different utilized geometries are shown in Fig. 5. For each amount of radar nodes used to reconstruct intensity and velocity fields, the average test accuracy and macro $F1$-score over all unique geometries are displayed, along with the standard deviation for both metrics. For every individual geometry experiment, the L1PO scheme is used. As such, $14 \cdot (5+4+3+1) = 182$ models are trained in total.

It is noted that performance increases with a larger amount of radar sensors used for the proposed data fusion method. This can be explained by two main factors.

1) Increased spatial diversity within the sensor network provides a greater range of observation angles, mitigating the effects of (partial) occlusion of the human target and individual body parts. This enhanced spatial diversity improves target illumination, resulting in a stronger intensity distribution following the application of the proposed method. In contrast with the surrounding empty space, this enhanced distribution enables the application of the threshold to more accurately identify the target volume. As a result, a more accurate selection of velocity vectors originating from the target is achieved, providing the classification model with the most reliable information regarding both the spatial extent of the target and its velocity distribution.
2) More sensors result in more instances of radial velocity measurement. This overdetermined system, paired with

Authorized licensed use limited to: TU Delft Library. Downloaded on August 07,2025 at 09:39:50 UTC from IEEE Xplore. Restrictions apply.

the least square error (LSE)-based approach used for the velocity reconstruction, results in mitigation of, e.g., measurement error in individual Doppler components. This, in turn, increases the accuracy of the velocity reconstruction in a grid cell.

In the case of two node geometries, it is hypothesized that a geometry that provides orthogonal measurements would give the highest performance. As the sensors provide no azimuth information, an orthogonal pair of range measurements will yield the most localized intensity profile. Inspection of the results in Table II, however, reveals the opposite, with geometries $(1, 2)$ and $(1, 3)$ yielding a performance that is almost two standard deviations lower than geometries $(1, 4)$ and $(1, 5)$. A possible explanation for this is the higher importance of larger differences in observation angles, allowing the target to be more completely illuminated.

To gauge potential bias in the data toward specific orientations or areas, geometries $(1, 3)$ and $(2, 4)$ are compared. The attained accuracy is 76.8% and 77.3%, respectively, within the two-sensor standard deviation of 1.6%.

For the three- and four-node geometries, the increase in performance is accompanied by a decrease in the standard deviation of the results. This lower effect of sensor positioning may, in this case, be indicative of partial redundancy. Of interest is that for three- and four-node geometries, the best performance is again achieved for those geometries that include the sensor pair $(1, 5)$. It is, thus, concluded that a large azimuthal opening angle of the sensor geometry is beneficial for the application of the proposed method for activity classification.

Summarizing the study on sensor network geometry reveals a roughly linear trend in performance increase per added sensor. At the maximum of five sensors, no saturation of performance is observed. This seems to indicate that further expansion of the sensor network can increase the classification performance. Based on the error analysis in Section IV-C, promising locations for additional sensors would be those that offer boresight axes that are not in the horizontal plane. These additional sensors would be able to provide information pertaining to the vertical velocity distribution. A secondary note on the geometry study is the relatively low effect of sensor geometry in comparison to the amount of sensors. The former yields performance variations on the order of 2%, whereas the latter brings about changes in excess of 5%. This may be due to the range of studied geometries all being relatively restricted to a half-circle, as opposed to, e.g., on a sphere or more complex topology. Finally, not all possible network configurations are evaluated, as mentioned in Section III. Under the assumption that the dataset is unbiased in activity location and orientation, this should have a limited impact. However, individual nodes observe different clutters in the laboratory space which potentially affects their contribution to the fusion process.

B. Comparison to Alternative Approaches

A comparison is made with alternative fusion-classification approaches. Results are obtained on the same dataset, and employing the same L1PO testing scheme. It should be noted that some of the reference methods performed their evaluation on a constrained five-class problem, where activities like "Falling from walking" and "Falling from standing" are merged into "Falling." Where this is the case, the results for the proposed method are also evaluated for the five-class problem. Table III shows the nine-class comparison, and Table IV shows the five-class comparison.

TABLE III

TEST ACCURACY AND MACRO $F1$-SCORE RESULTS FOR THE PROPOSED CLASSIFICATION METHOD VERSUS TWO REFERENCE METHODS. ALL RESULTS ARE BASED ON THE SAME DATASET [22] AND THE SAME L1PO VALIDATION SCHEME. THE RESULTS IN THIS TABLE ARE FOR THE FULL NINE-CLASS CLASSIFICATION PROBLEM. *: [14] AND †: [20]

Classifier	Test Accuracy	Macro F1-score
Proposed	**0.874**	**0.819**
Point Transformer*	0.869	0.787
CNN-BiGRU†	0.851	-

TABLE IV

TEST ACCURACY AND MACRO $F1$-SCORE RESULTS FOR THE PROPOSED CLASSIFICATION METHOD VERSUS TWO REFERENCE METHODS. ALL RESULTS ARE BASED ON THE SAME DATASET [22] AND THE SAME L1PO VALIDATION SCHEME. THE RESULTS IN THIS TABLE ARE FOR A FIVE-CLASS CLASSIFICATION PROBLEM. *: [14], †: SIGNAL FUSION [18], AND ‡: FEATURE FUSION [18]

Classifier	Test Accuracy	Macro F1-score
Proposed	**0.930**	**0.898**
Point Transformer*	-	0.862
GRU†	0.909	0.778
LSTM†	0.910	0.769
bi-GRU†	0.933	0.844
BiLSTM†	0.931	0.836
BiLSTM‡	0.924	0.840

For the full nine-class classification task, the reference fusion-classification methods comprise one based on the point Transformer architecture [14], and one on a hybrid CNN-Bi-gated recurrent unit (GRU) architecture [20]. For the former, radar data are processed into a point cloud representation with dimensions range, radial velocity, time, and reflection intensity. Point clouds from various sensors are fused by simply adding the points from all individual sensors into a single point cloud object. No data association is performed, or an attempt to estimate a more comprehensive representation of reflection intensity and/or velocity distribution, as in the proposed method. In [20], the fusion of data from N sensors is achieved through a maxpooling operation over N feature maps. The feature maps are the result of the CNN module processing spectrogram representations of the sensor data. Both reference methods feature a fused representation that is at a higher abstraction level than the proposed method with respect to the kinematics of the observed experimental scene. Because of this, the versatility of the proposed fused representation is assumed to be greater in terms of applicability to other tasks, such as for instance tracking or specific fall detection methods, such as the one proposed in [21]. In addition, the classification performance of the proposed method is improved with respect to the reference methods.

Authorized licensed use limited to: TU Delft Library. Downloaded on August 07,2025 at 09:39:50 UTC from IEEE Xplore. Restrictions apply.

True Class \ Predicted Class	Bending (sitting)	Bending (standing)	Falling (standing)	Falling (walking)	Sitting Down	Standing up (ground)	Standing up (sitting)	Stationary	Walking
Bending (sitting)	70.2%				13.5%		13.9%	2.4%	
Bending (standing)		89.4%			6.7%		2.9%	1.0%	
Falling (standing)	3.1%		92.2%	0.7%		0.7%		1.4%	1.9%
Falling (walking)				68.0%					32.0%
Sitting Down		16.2%	0.5%	0.5%	73.9%		4.9%	3.8%	0.2%
Standing up (ground)	1.5%		2.3%			93.2%		3.0%	
Standing up (sitting)	0.3%	22.6%			3.6%	4.5%	66.5%	2.5%	
Stationary	0.2%	0.1%		0.1%		6.0%		84.0%	9.6%
Walking		0.1%			0.0%			4.7%	95.2%

Fig. 9. Confusion matrix for the test result for a single participant.

In the case of the constrained five-class classification problem, favorable results are also achieved with the proposed method. The reference methods are again based on the point Transformer architecture [14], and a selection of RNNs [18]. Two types of fusion are employed in [18]. First, signal fusion entails elementwise addition of N complex-valued range–time representations from N sensors. A single spectrogram is then computed based on the fused range–time matrix. The second type is feature fusion and comprises the concatenation of N spectrogram representations along the Doppler dimension. In terms of macro $F1$-score, improvements of almost 4 %point are achieved when utilizing the proposed method. The test accuracy is within 0.4 %point of the best-performing model.

Finally, the benefits of the velocity reconstruction are illustrated by training a model using only the intensity map data, and without the accompanying velocity vector fields. Sequences are randomly distributed following an 80%/20% ratio into a training set and a testing set. This approach yields an accuracy of 50.04% and a macro $F1$-score of 48.45%, clearly indicating the superiority of the proposed method.

C. Error Analysis and Computational Considerations

Fig. 9 shows a representative confusion matrix for a single L1PO result, i.e., with test data from a single participant. The five largest errors in this matrix will be discussed. Most prevalent among the classification errors is the confusion between "Falling (Walking)" and "Walking." A likely explanation for this error is the ambiguous transition point between these activities. Since predictions are made at every time step (0.26 s), the ambiguous time between walking and falling constitutes a relatively large amount of equally ambiguous predictions. This ambiguity is compounded by variability in the ground truth due to each participant indicating transition points themselves by clicking on a remote controller. The next four most frequent errors are between activities that all share an initial forward rotation of the torso, including standing up from sitting. Since the case study is constrained to the 2-D horizontal plane, these errors can be attributed to a lack of information on both the vertical posture of the subject, as well as a lack of a detailed vertical velocity profile. It is noted that standing up from the ground is kinematically distinct to the other activities with a strong vertical component and as such exhibits fewer errors during classification.

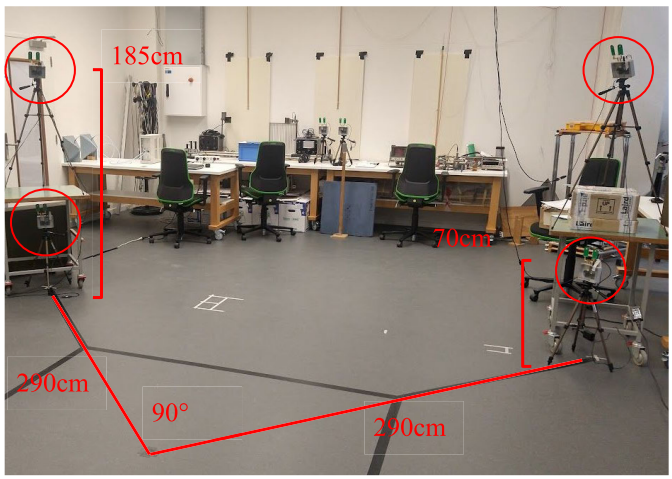


Fig. 10. Experimental setup used in the 3-D feasibility study. Sensors are placed at the base of an isosceles right triangle with sides of 290 cm. The lower sensors are located at a height of 70 cm and the upper sensors at 185 cm. Targets are positioned at the apex of the triangle.

Typical computational requirements, based on the classification case study and available hardware, are reported here. Fusion processing of a full 120-s sequence on a single core of a 3.40-GHz i5 CPU takes on average 51 s. However, since time steps are independent, multicore processing yields significant improvements. Typical CNN-BiLSTM models are on the order of 10 MB, and inference time for a full sequence is less than 0.1 s on an NVIDIA Tesla V100S.

V. Experimental Feasibility Study in 3-D

To demonstrate the ability of the proposed method to yield 3-D representations of extended targets and motions, a feasibility study is performed. For this study, data are captured with

Authorized licensed use limited to: TU Delft Library. Downloaded on August 07,2025 at 09:39:50 UTC from IEEE Xplore. Restrictions apply.

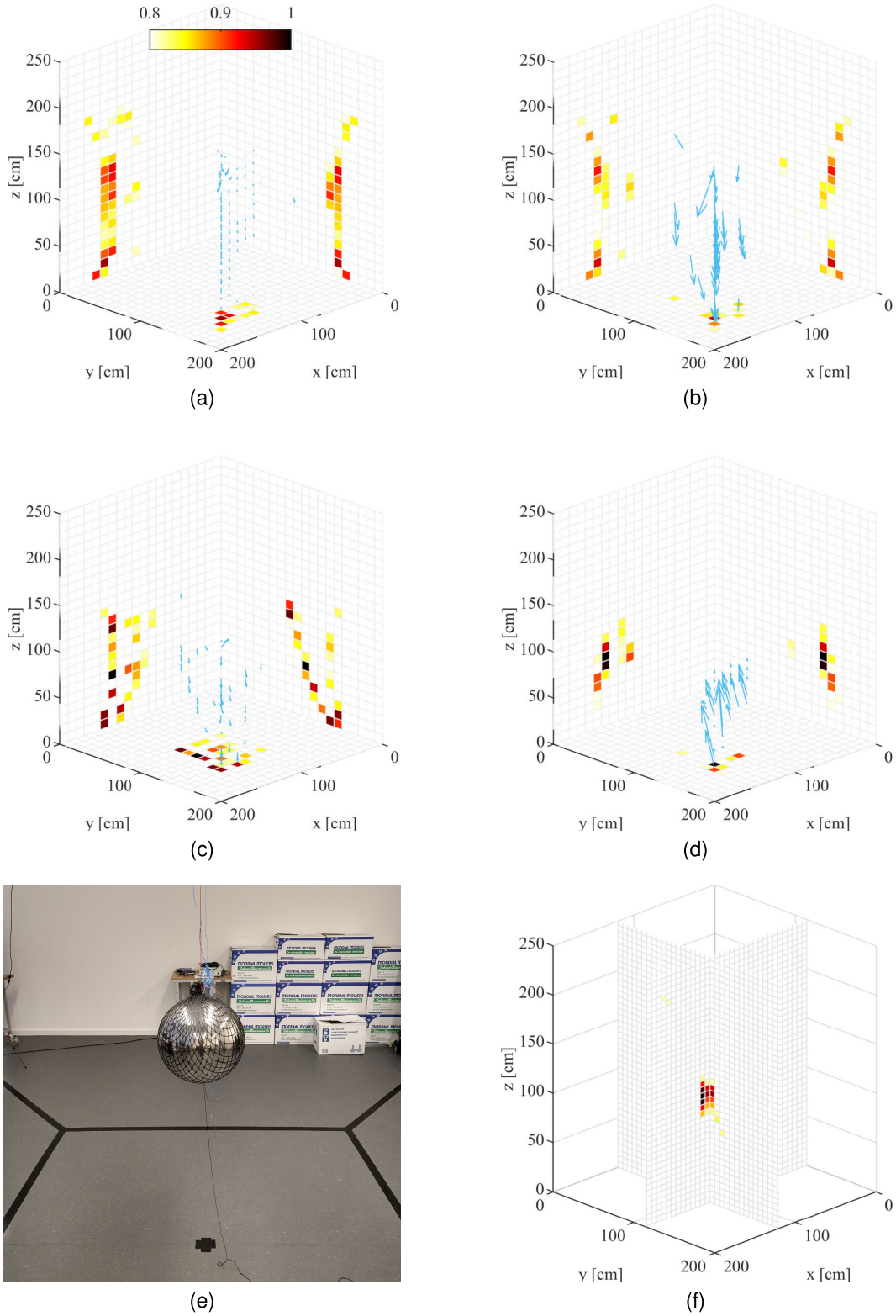


Fig. 11. Results for the 3-D feasibility study. (a)–(d) 3-D intensity and velocity fields are represented with the maximum intensity of the thresholded 3-D matrix projected on three orthogonal axes. (a) Human subject standing upright. (b) Human subject moving down into squatting position. (c) Human subject sitting in squat. (d) Human subject moving back up to the standing position. (e) Metal sphere used for feasibility study. (f) Intensity distribution of metal sphere, represented with two 2-D slices of full 3-D intensity matrix.

a four-sensor network that extends not only in the horizontal plane but in the vertical direction as well. The measurement setup is shown in Fig. 10, and the sensors are the same as those used in the 2-D case study. Two scenes are captured: a metal sphere of diameter 30 cm suspended 120 cm from the ground, and a human subject performing a series of squats, thus exhibiting movements toward the upward and downward directions periodically.

Authorized licensed use limited to: TU Delft Library. Downloaded on August 07,2025 at 09:39:50 UTC from IEEE Xplore. Restrictions apply.

The data are processed following the proposed method, and the results are shown in Fig. 11. Fig. 11(a)–(d) represents 3-D intensity and velocity fields at key points during the squatting activity, with the maximum intensity of the thresholded 3-D matrix projected on three orthogonal axes to facilitate visualization. The four states of standing upright (a), moving down into squatting position (b), sitting in squat (c), and moving back up to the standing position (d), feature intensity and velocity distributions that correspond to the physical expectation of the human movement. Noise levels in this experiment are, however, higher, indicating that an alternative sensor geometry or additional preprocessing may be required to fully utilize the proposed method in 3-D for subsequent classification purposes. Fig. 11(e) and 11(f) contains the metal sphere and corresponding 3-D intensity distribution, respectively. The distribution is visualized by means of two 2-D slices of the 3-D intensity matrix. The sphere is seen at the correct height and surface area.

VI. Conclusion

This work proposes a novel sensor fusion method that processes data from a network of radar sensors and yields 3-D representations of both reflection intensity and velocity distribution. The fused representation can be obtained regardless of the capacity of the individual radars to determine angle of arrival. The fused data representation is easily linked to the kinematic of the observed target and allows for versatile application to various tasks, from tracking to human activity classification.

The proposed method is evaluated in a case study, where it is applied to a human activity classification problem. For this task, sequences of human activities from a publicly available dataset are processed using the method. Subsequently, classification is performed by means of a hybrid CNN-BiLSTM model. For an L1PO testing scheme, a test accuracy and macro $F1$-score of 87.4% and 81.9% are achieved, outperforming alternative fusion methods on the same dataset. An extensive study on the number and position of radar nodes in the network is performed to evaluate their effect on classification performance. It is found that for the five-node geometry utilized in the case study, additional nodes can likely further increase prediction accuracy. Most notably, error analysis indicates that additional sensors that are not in the horizontal plane are expected to improve classification most substantially. An experimental feasibility study is also conducted, successfully demonstrating the proposed method's capability to produce 3-D representations of extended target shapes and velocity distributions.

In future work, a network of multiple-input multiple-output (MIMO) sensors will be utilized to fully exploit the capabilities of the method. Specifically, the inclusion of velocity components in the vertical direction is expected to mitigate classification errors between activities that feature distinctive vertical motions. Additionally, an extension of the method to allow for the classification of multiple people in the experimental scene is planned, leveraging on the multiple perspectives from different radars to separate the signatures of each individual.

Acknowledgment

The authors are grateful to all the volunteers who participated in the data collection. They thank the anonymous reviewers for their constructive comments and suggestions in the review process.

References

[1] X. Dang, Z. Chen, and Z. Hao, "Emotion recognition method using millimetre wave radar based on deep learning," *IET Radar, Sonar Navigat.*, vol. 16, no. 11, pp. 1796–1808, Nov. 2022.

[2] S. K. Koul and R. Bharadwaj, "UWB and 60 GHz radar technology for vital sign monitoring, activity classification and detection," in *Wearable Antennas and Body Centric Communication*, 1st ed., Singapore: Springer, 2021, ch. 7, pp. 191–215.

[3] M. M. Rahman, D. Martelli, and S. Z. Gurbuz, "Gait variability analysis with multi-channel FMCW radar for fall risk assessment," in *Proc. IEEE 12th Sensor Array Multichannel Signal Process. Workshop (SAM)*, Jun. 2022, pp. 345–349.

[4] R. Zavorka, R. Marsalek, J. Vychodil, E. Zochmann, G. Ghiaasi, and J. Blumenstein, "Human activity classification via Doppler shift estimation from 60 GHz OFDM transmission," in *Proc. 32nd Int. Conf. Radioelektronika (RADIOELEKTRONIKA)*, Apr. 2022, pp. 1–5.

[5] Y. Yao et al., "Fall detection system using millimeter-wave radar based on neural network and information fusion," *IEEE Internet Things J.*, vol. 9, no. 21, pp. 21038–21050, Nov. 2022.

[6] S. A. Shah, A. Tahir, J. Le Kernec, A. Zoha, and F. Fioranelli, "Data portability for activities of daily living and fall detection in different environments using radar micro-doppler," *Neural Comput. Appl.*, vol. 34, no. 10, pp. 7933–7953, 2022.

[7] C. Wang, X. Zhao, and Z. Li, "DCS-CTN: Subtle gesture recognition based on TD-CNN-Transformer via millimeter-wave radar," *IEEE Internet Things J.*, vol. 10, no. 20, pp. 17680–17693, Oct. 2023.

[8] L. Wang, Z. Cui, Y. Pi, C. Cao, and Z. Cao, "Adaptive framework towards radar-based diversity gesture recognition with range-Doppler signatures," *IET Radar, Sonar Navigat.*, vol. 16, no. 9, pp. 1538–1553, Sep. 2022.

[9] E. Kurtoglu, A. C. Gurbuz, E. A. Malaia, D. Griffin, C. Crawford, and S. Z. Gurbuz, "ASL trigger recognition in mixed activity/signing sequences for RF sensor-based user interfaces," *IEEE Trans. Human-Mach. Syst.*, vol. 52, no. 4, pp. 699–712, Aug. 2022.

[10] S. Z. Gurbuz et al., "Multi-frequency RF sensor fusion for word-level fluent ASL recognition," *IEEE Sensors J.*, vol. 22, no. 12, pp. 11373–11381, Jun. 2022.

[11] L. Werthen-Brabants, G. Bhavanasi, I. Couckuyt, T. Dhaene, and D. Deschrijver, "Split BiRNN for real-time activity recognition using radar and deep learning," *Sci. Rep.*, vol. 12, no. 1, pp. 1–12, May 2022.

[12] S. Z. Gurbuz, M. M. Rahman, E. Kurtoglu, and D. Martelli, "Continuous human activity recognition and step-time variability analysis with FMCW radar," in *Proc. IEEE-EMBS Int. Conf. Biomed. Health Informat. (BHI)*, Sep. 2022, pp. 1–4.

[13] Z. Li, Y. Liu, B. Liu, J. Le Kernec, and S. Yang, "A holistic human activity recognition optimisation using AI techniques," *IET Radar, Sonar Navigat.*, vol. 18, no. 2, pp. 256–265, Sep. 2023.

[14] N. C. Kruse, F. Fioranelli, and A. Yarovoy, "Radar point cloud processing methods for human activity classification with point transformer networks," *IEEE Trans. Radar Syst.*, vol. 2, pp. 1–12, 2024.

[15] Z. Yu et al., "An intelligent implementation of multi-sensing data fusion with neuromorphic computing for human activity recognition," *IEEE Internet Things J.*, vol. 10, no. 2, pp. 1124–1133, Jan. 2023.

[16] M. J. Bocus et al., "OPERAnet, a multimodal activity recognition dataset acquired from radio frequency and vision-based sensors," *Sci. Data*, vol. 9, no. 1, p. 474, 2022.

[17] H. Li, A. Mehul, J. Le Kernec, S. Z. Gurbuz, and F. Fioranelli, "Sequential human gait classification with distributed radar sensor fusion," *IEEE Sensors J.*, vol. 21, no. 6, pp. 7590–7603, Mar. 2021.

[18] R. G. Guendel, F. Fioranelli, and A. Yarovoy, "Distributed radar fusion and recurrent networks for classification of continuous human activities," *IET Radar, Sonar Navigat.*, vol. 16, no. 7, pp. 1144–1161, Apr. 2022.

[19] A. Gorji, T. Gielen, M. Bauduin, H. Sahli, and A. Bourdoux, "A multi-radar architecture for human activity recognition in indoor kitchen environments," in *Proc. IEEE Radar Conf. (RadarConf)*, May 2021, pp. 1–6.

Authorized licensed use limited to: TU Delft Library. Downloaded on August 07,2025 at 09:39:50 UTC from IEEE Xplore. Restrictions apply.

[20] 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, 2022, Art. no. 5115215.

[21] N. Kruse, R. Guendel, F. Fioranelli, and A. Yarovoy, "Distributed radar fusion for extended target location and velocity reconstruction," in *Proc. IEEE Radar Conf. (RadarConf)*, May 2024, pp. 1–6.

[22] R. G. Guendel, M. Unterhorst, F. Fioranelli, and A. Yarovoy. (Nov. 2021). *Dataset of Continuous Human Activities Performed in Arbitrary Directions Collected With a Distributed Radar Network of Five Nodes*. [Online]. Available: https://data.4tu.nl/articles/dataset/Dataset_of_continuous_human_activities_performed_in_arbitrary_directions_collected_with_a_distributed_radar_network_of_five_nodes/16691500

[23] *PulsON410*. Time Domain. Accessed: Jun. 20, 2024. [Online]. Available: https://www.ion.org/gnss/upload/files/904_P410%20Ranging%20Radio.pdf

[24] M. Unterhorst, "Distributed radars for unconstrained direction classification of human activities," M.S. thesis, Dept. Eng., Università Politecnica delle Marche, Ancona, Italy, 2020.

[25] N. C. Kruse, F. Fioranelli, and A. Yarovoy, "Continuous human activity classification with radar point clouds and point transformer networks," in *Proc. 20th Eur. Radar Conf. (EuRAD)*, Sep. 2023, pp. 302–305.

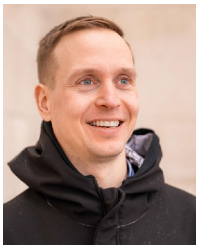
[26] Z. Zhang, Z. Tian, and M. Zhou, "Latern: Dynamic continuous hand gesture recognition using FMCW radar sensor," *IEEE Sensors J.*, vol. 18, no. 8, pp. 3278–3289, Apr. 2018.

[27] Z. Yang, H. Wang, P. Ni, P. Wang, Q. Cao, and L. Fang, "Real-time human activity classification from radar with CNN-LSTM network," in *Proc. IEEE 16th Conf. Ind. Electron. Appl. (ICIEA)*, Aug. 2021, pp. 50–55.

[28] Y. Lecun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," *Proc. IEEE*, vol. 86, no. 11, pp. 2278–2324, Nov. 1998.

Nicolas C. Kruse (Graduate Student Member, IEEE) received the B.Sc. degree in applied physics from Delft University of Technology (TU Delft), Delft, The Netherlands, in 2017, and the M.Sc. degree in physics from the University of Groningen, Groningen, The Netherlands, in 2020.

He joined the Microwave Sensing, Signals and Systems (MS3) Group, Faculty of Electrical Engineering, TU Delft, in March 2021, where he is currently researching classification algorithms for continuous human activity sequences through micro-Doppler signatures.



Ronny G. Guendel received the Diploma-Engineering degree [Dipl.-Ing. (FH)] in automotive electronics from the University of Applied Sciences Zwickau, Zwickau, Germany, in 2017, the M.Sc. degree from Villanova University, Villanova, PA, USA, in 2019, and the Ph.D. degree from Delft University of Technology (TU Delft), Delft, The Netherlands, in November 2023, promoted by Dr. F. Fioranelli and Prof. A. Yarovoy.

From 2017 to 2018, he was a Fulbright Scholar of electrical engineering with Villanova University, where he was a Researcher with the Center for Advanced Communications under Prof. M. G. Amin, focusing on continuous activity classification, from 2018 to 2019. Since 2024, he has been working as a Radar Scientist with The Netherlands Organization for Applied Scientific Research (TNO), Hague, The Netherlands. He focused on monitoring continuous human activities beyond micro-Doppler using distributed radar networks and analyzing human signatures for healthcare and security applications. His research interests include distributed radar systems, radar signal processing with machine learning/deep learning integration, and combining radar with other sensing modalities.

Francesco Fioranelli (Senior Member, IEEE) received the Ph.D. degree from Durham University, Durham, U.K., in 2014.

He was a Research Associate with University College London, London, U.K., from 2014 to 2016, and an Assistant Professor with the University of Glasgow, Glasgow, U.K., from 2016 to 2019. He is currently an Associate Professor with Delft University of Technology (TU Delft), Delft, The Netherlands. He has authored over 190 peer-reviewed publications and edited the books "Micro-Doppler Radar and Its Applications" and "Radar Countermeasures for Unmanned Aerial Vehicles" (IET-Scitech, 2020). His research interests include the development of radar systems and automatic classification for human signature analysis in healthcare and security, drones and unmanned aerial vehicle (UAV) detection and classification, and automotive radar.

Dr. Fioranelli received four best paper awards and the IEEE AESS Fred Nathanson Memorial Radar Award in 2024.

Alexander Yarovoy (Fellow, IEEE) received the Diploma degree (Hons.) in radiophysics and electronics and the Candidate Physics and Mathematics Science and Doctor Physics and Mathematics Science degrees in radiophysics from Kharkov State University, Kharkov, Ukraine, in 1984, 1987, and 1994, respectively.

In 1987, he joined the Department of Radiophysics, Kharkov State University, as a Researcher, and became a Full Professor in 1997. From September 1994 to 1996, he was a Visiting Researcher with the Technical University of Ilmenau, Ilmenau, Germany. Since 1999, he has been with Delft University of Technology (TU Delft), Delft, The Netherlands, where he has been the Chair of the Microwave Sensing, Signals and Systems Group, since 2009. He has authored or co-authored more than 600 scientific or technical articles and 14 book chapters, and holds 11 patents. His main research interests are in high-resolution radar, microwave imaging, and applied electromagnetics (in particular, ultra-wideband (UWB) antennas).

Prof. Yarovoy was a recipient of the European Microwave Week Radar Award for the article that best advances the state of the art in radar technology in 2001 (together with L. P. Ligthart and P. van Genderen) and 2012 (together with T. Savelyev). In 2023, together with Dr. I. Ullmann, N. C. Kruse, R. G. Guendel, and Dr. F. Fioranelli, he received the Best Paper Award at IEEE Sensor Conference. In 2010 together with D. Caratelli, he received the Best Paper Award of the Applied Computational Electromagnetic Society (ACES). From 2008 to 2017, he served as the Director for the European Microwave Association (EuMA). He is and has been serving on various editorial boards such as that of IEEE TRANSACTIONS ON RADAR SYSTEMS. From 2011 to 2018, he served as an Associate Editor for the *International Journal of Microwave and Wireless Technologies*. He has been a member of numerous conference steering and technical program committees. He served as the General TPC Chair for the 2020 European Microwave Week (EuMW'20), the Chair and a TPC Chair for the 5th European Radar Conference (EuRAD'08), as well as the Secretary for the 1st European Radar Conference (EuRAD'04). He also served as the Co-Chair and a TPC Chair for the Xth International Conference on GPR (GPR2004).

Authorized licensed use limited to: TU Delft Library. Downloaded on August 07,2025 at 09:39:50 UTC from IEEE Xplore. Restrictions apply.