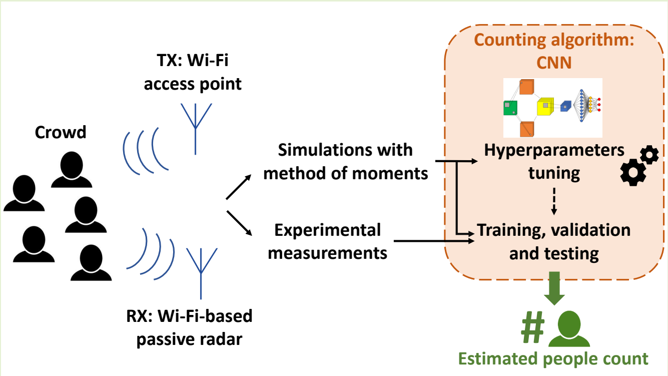
This article has been accepted for publication in the IEEE Sensors Journal. This is the author's version which has not been fully edited andcontent may change prior to final publication. Citation information: DOI 10.1109/JSEN.2024.3453070

Large-Scale Crowd Size Classification with a Wi-Fi-Based Passive Radar

Laurent Storrer, *Member, IEEE*, Hasan Can Yildirim, *Member, IEEE*, Martin Willame, *Member, IEEE*, Evert I. Pocoma Copa, Dejvi Cakoni, Sofie Pollin, *Senior Member, IEEE*, Jérôme Louveaux, *Member, IEEE*, Philippe De Doncker, *Member, IEEE*, and François Horlin, *Member, IEEE*.

***Abstract*—We investigate the problem of large-scale crowd size classification with a Wi-Fi-based passive radar for crowds of up to 100 people, with classes corresponding to intervals of numbers of people. A Convolutional Neural Network (CNN) operating on radar range-Doppler maps (RDMs) is used as a classification algorithm. We propose a crowd simulator based on the method of moments (MoM) in electromagnetics able to generate representative RDMs for large crowds. We show that these MoM simulation data can be used to design the classification algorithms and tune their hyperparameters. We also investigate the limitations of the MoM simulation data in training the classification algorithms for subsequent application on experimental data. Crowd size classification is performed with high accuracy on real-life experimental measurements of a crowd with up to 100 people, obtained by channel estimation with 802.11ax-compliant High-Efficiency Long Training Fields transmitted by a Wi-Fi-based passive radar setup featuring two Universal Software Radio Peripherals X310.**



***Index Terms*—Crowd size, counting, Wi-Fi, passive radar, method of moments, convolutional neural network.**

I. INTRODUCTION

A. Radio frequency-based crowd monitoring

Crowd monitoring is critical for public space and events supervision, to avoid overcrowding leading to deadly accidents. A central point in crowd monitoring is the ability to count in real-time the number of people in the crowd. This allows the security staff to know when the number of people is reaching critical thresholds and to react accordingly.

Various crowd monitoring methods have been developed, usually split into vision and non-vision-based methods [1]. Among non-vision-based methods, our research group has developed a MAC address fingerprinting based on Wi-Fi probe requests sniffing [2]. On the other hand, due to the quasi-ubiquitous availability of wireless signals of opportunity, Radio Frequency-based crowd monitoring is the subject of increasing interest. Metrics of the wireless channel corresponding to the scene to monitor can be exploited to derive features that are fed to a counting or size classification algorithm. Long-Term Evolution (LTE) Reference Signal Received Power (RSRP) values can be collected, and their high-order statistical moments can be exploited as counting features, for example in a k-nearest neighbour classifier with up to 5 people [3]. The Received Signal Strength (RSS) from a Wi-Fi access point can be used in a Renewal Process framework to derive the corresponding number of people in the scene with up to 20 people [4].

With the ongoing work of the 802.11 Task Group for WLAN sensing [5], a radar approach is also envisioned for people counting and crowd size classification. Wi-Fi signals of opportunity can be processed in various ways to obtain range and Doppler information, from which counting features can be extracted and fed to a classifier performing the counting or size classification task. In [6], the Doppler spectrogram is derived from Cross-Ambiguity Function processing, and derived features are given as input to a Support Vector Machine (SVM). Range-Doppler maps (RDMs) can also be the starting point for counting or crowd size classification. In [7], RDMs and Doppler spectrograms are given as input to

---

Manuscript submitted on 03-03-2024. This work was supported by the Excellence of Science (EOS) Multi-Service Wireless Network program from the Fonds de la Recherche Scientifique (F.R.S-FNRS) of Belgium.

Laurent Storrer is with the OPERA Wireless Communication Group (OPERA-WCG), Université Libre de Bruxelles (ULB), 1050 Bruxelles, and the Department of Electrical Engineering (ESAT), Katholieke Universiteit Leuven (KUL), 3000 Leuven (email: laurent.storrer@ulb.be).

Hasan Can Yildirim, Evert I. Pocoma Copa, Dejvi Cakoni, Philippe De Doncker and François Horlin are also with the OPERA-WCG (e-mails: {Hasan.Can.Yildirim, epocomac, dejvi.cakoni}@ulb.be, fhorlin,philippe.dedoncker}@ulb.ac.be).

Martin Willame and Jérôme Louveaux are with the Institute of Information and Communication Technologies, Electronics and Applied Mathematics (ICTEAM), Université Catholique de Louvain (UCL), 1348 Ottignies-Louvain-la-Neuve (e-mail: {martin.willame, jerome.louveaux}@uclouvain.be).

Sofie Pollin is also with ESAT (email: sofie.pollin@kuleuven.be).

convolutional neural networks (CNNs) for counting, but with a limited number of 4 people. In [8], features are computed via a Principal Component Analysis (PCA) followed by a Discrete Wavelet Transform on the Channel State Information matrix over multiple transmitter (TX) and receiver (RX) antenna pairs, and fed to a Long Short-Term Memory network followed by a CNN. However, the number of people is also limited to 5. Additionally, these works are based on old 802.11 standards with a limited bandwidth of 20 MHz and 40 MHz. These bandwidths yield a poor sampling time, hindering the propagation delay estimation from the signal echoes bouncing off people, hence preventing the range information from being used for counting.

In summary, most of the previously mentioned works dealt with less than 10 people. To the best of the authors' knowledge, the largest crowd handled with Wi-Fi-based passive radar approaches is the 20-people crowd from [4] with the limited 802.11g Wi-Fi standard.

B. Need for a crowd simulator

Furthermore, the algorithms and their parameters in the previously mentioned works are tested and adjusted using experimental data only. This requires considerable time and effort to install the measurement setup and to collect enough data points to be able to consistently test various algorithms and hyperparameter sets. Moreover, it is difficult to gather large numbers of people in consistent and reproducible experimental scenarios. Finally, proper positioning of the transmitter and receiver is also required, and investigating this problem is much easier by simulation. Consequently, a crowd simulator for Wi-Fi-based passive radars would be useful to reduce the effort put into measurements.

The authors of [9] developed a multistatic radar simulation environment with uniformly distributed targets to generate channel impulse responses on which counting algorithms can be trained. Nonetheless, it resorts to a geometry-based single-bounce channel model with simple path loss, hence considering point targets and neglecting electromagnetic interactions between targets. The modelling could thus be improved to integrate the scattering of an incident electromagnetic field on real bodies, to avoid point targets and to consider interactions between targets. The human body can be modelled by an infinite lossy dielectric cylinder [10] to derive a channel model using the two-dimensional Method of Moments (MoM). The Radar Cross Section (RCS) is also computed with various methods on complex shapes in [11]. However, the model has to be scaled to a crowd of people. There is little available research on scattering by multiple arbitrary dielectric bodies: it mostly uses analytical frameworks and focuses on propagation among trees. They can be modelled as finite dielectric cylinders, with randomly placed cylindrical branches and disk-shaped leaves [12], and the scattered electric field can be derived analytically by means of Foldy-Lax's multiple scattering theory [13], [14]. Scattering between trees is also investigated using the infinite cylinder approximation (ICA) [15].

Combining those various approaches, our previous work aimed at extending the knowledge about multiple human bodies interactions by using the MoM [16] with bodies modelled as dielectric cylindrical shells whose thickness is equal to the electromagnetic skin depth derived from the complex relative permittivity of the human body. Based on this work, our next step [17] was to propose a crowd simulation framework based on the two-dimensional MoM applicable to Wi-Fi-based sensing purposes and to use it to generate channel impulse responses and Doppler profiles. Those MoM profiles were used to train an SVM classifier, which was then applied to real-life experimental measurements with up to 40 people, divided into classes based on the number of people.

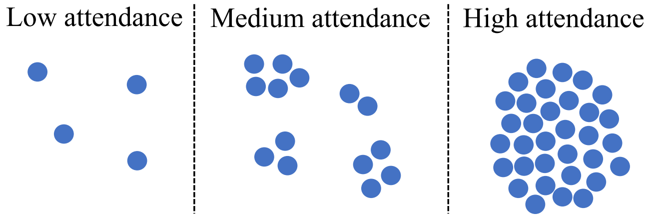


Fig. 1. Proposed crowd classes. People are represented by blue disks.

C. Crowd scenario and classes

From the above discussion, let us now define the crowd scenario to investigate in the rest of this paper. We consider mass events such as concerts or outdoor gatherings, where people join the gathering progressively.

When dealing with large-scale crowd monitoring in an event, precise counting and localisation of each person in the crowd is not required for the people monitoring the event. Instead, a few categories corresponding to intervals of numbers of people should be considered. Indeed, in a mass event, danger occurs when the number of people over a given area crosses a certain threshold. Therefore, the basic version of a crowd monitoring scheme would be a binary classification between a small crowd and a large crowd. It is interesting to refine this information by adding an intermediate class, ending up with a 3-class view: low, medium, and high attendance. This is the classification design chosen here. Moreover, it was noted that the crowd spatial configuration, *i.e.* the pattern of people's position in the crowd, evolves during the event and that it could be linked to each class. The three classes are defined as follows:

- **Class 1 - Low attendance:** the first class represents a situation where only a few people are passing by at the beginning of the event. Spatially, the people are mainly spread around the event area.
- **Class 2 - Medium attendance**: the second class represents an intermediate state where the number of people has increased. Spatially, people discuss with each other in small groups, *i.e.* they form spatial clusters.
- **Class 3 - High attendance**: the third and last class represents the state where the event attendance is increasing quickly and reaches a high number of people. Spatially, the event reaches its final structure, with people all gathered in a central zone to attend a show, *i.e.* in one large cluster.

An illustration of the classes is presented schematically in Fig. 1. This class division framework allows one to have clear

This article has been accepted for publication in the IEEE Sensors Journal. This is the author's version which has not been fully edited andcontent may change prior to final publication. Citation information: DOI 10.1109/JSEN.2024.3453070

information about what is happening in the crowd, and can be generalised to different crowd scales, *i.e.* different maximal numbers of people and different area sizes.

D. Contributions

Based on the state-of-the-art and the scenario presented above, our main contribution in this paper is to demonstrate experimentally that a Wi-Fi-based passive radar using preambles from the recent 802.11ax standard is able to perform large crowd size classification considering gatherings of up to 100 people. Our passive radar leverages the increased bandwidth from the recent 802.11ax standard for a finer range resolution and reaches more than 85% classification accuracy using a reference classifier (here a CNN). It is meant to be a proof of concept of the capabilities of 802.11ax-based passive radars for crowd monitoring. Achieving this goal also relies on the following supporting contributions to obtain simulated and experimental data:

- We are the first to propose a MoM-based crowd simulator allowing the generation of RDMs for large-scale crowds. This is much more computationally intensive than generating decoupled range and Doppler profiles like in [17], and is achieved through a thorough computational optimisation compared to [17]. This simulator is suited to generate synthetic data for Wi-Fi-based passive radar crowd size classification. We show that the classification algorithm hyperparameters can be tuned using MoM simulation data.
- We tackle the challenging task of collecting a set of 1140 measurements, *i.e.* 1140 RDMs, at a real event with up to 100 people, using a Wi-Fi-based passive radar setup featuring two Universal Software Radio Peripherals X310 and transmitting 802.11ax-compliant High-Efficiency Long Training Fields. These measurements are used to train, validate, and test the crowd size classifier. We also analyse the limits of simulation-only training for subsequent application on experimental data, yet showing the relevance of this approach in a simplified case.

E. Paper structure

The rest of this article is structured as follows: In Section II, the system model and radar processing framework to obtain RDMs are described. In Section III, our Wi-Fi-based passive radar experimental setup is presented, along with a description of the outdoor gathering that was analysed to collect real-life measurements of a crowd. In Section IV, the Method of Moments is developed for the simulation counterpart of the crowd scenario. In Section V, we detail the algorithm performing the crowd size classification task based on the simulated and measured data. The algorithm hyperparameters tuning procedure is also explained, along with the algorithm training approaches. Finally, Section VI presents the hyperparameters tuning results and the classification results on simulation data and real-life crowd measurements. Section VII concludes this article. Throughout this paper, bold notations refer to vectors or matrices (*e.g.* $\boldsymbol{X}$), and bold notations with a tilde refer to tensors (*e.g.* $\widetilde{\boldsymbol{X}}$). The $\hat{}$ symbol refers to estimated quantities (*e.g.* $\hat{y}$), and the $'$ symbol indicates measurements-related quantities (*e.g.* y').

II. RADAR PROCESSING

A. Signal model

To monitor the crowd, we consider a continuous stream of 802.11ax-compliant High Efficiency-Long Training Fields (HE-LTF) Orthogonal Frequency-Division Multiplexing (OFDM) symbols, transmitted from TX [18]. Considering a signal bandwidth B, the Q OFDM subcarriers are indexed with $q = -Q/2, \ldots, Q/2-1$ corresponding to a frequency $f_q = f_c + qB/Q$ with f_c the carrier frequency. In the time domain, each HE-LTF contains $Q + L_{cp}$ samples, where L_{cp} is the Cyclic Prefix length. Each HE-LTF is sampled with a sampling time $T_s = 1/B$, also denoted as fast time, and indexed with $i = 0, \ldots, Q-1$. The time between the reception of two HE-LTFs is $T = (Q + L_{cp})T_s$, also known as the slow time. It is indexed with an index $k = 0, \ldots, N-1$, corresponding to slow time instants $t_k = kT$.

The OFDM signals are received at RX after scattering on people. The echoes from scattering are embedded in a Channel Transfer Function (CTF) in the frequency domain, at each slow time instant t_k. With one antenna at TX and at RX, it can be expressed with a matrix $\boldsymbol{H}$ whose elements are $H[q,k]$ [1].

B. Range-Doppler Map

At RX, the CTF at each slow time instant can be estimated through an element-wise division of the received signal and the known transmitted signal (HE-LTF) in the frequency domain. This is a so-called frequency domain least-squares channel estimation [19] yielding an estimate $\hat{\boldsymbol{H}}$ of $\boldsymbol{H}$. The Channel Impulse Responses (CIRs) at all slow time instants t_k, also called range profiles, can be obtained through an Inverse Fast Fourier Transform (IFFT) along the first dimension of $\hat{\boldsymbol{H}}$. This yields a matrix $\hat{\boldsymbol{h}}$ with elements $\hat{h}[i,k]$, called the range-slow time map. Each column of $\hat{\boldsymbol{h}}$ is one CIR at a slow time instant t_k. This IFFT is often referred to as *range IFFT*. The fast time index i discretises the scattered echoes propagation delays at instants iT_s, or equivalently at ranges $ciT_s/2$, where c is the speed of light in vacuum.

Computing the Fast Fourier Transform (FFT) on the second dimension of $\hat{\boldsymbol{h}}$ yields the usual range-Doppler map (RDM) $\hat{\boldsymbol{D}}$ with elements $\hat{D}[i,n]$, where $n = -N/2, \ldots, N/2-1$ is the Doppler frequency bin index. This operation is often referred to as *Doppler FFT*. The processing to obtain $\hat{\boldsymbol{D}}$ from the estimated CTF matrix $\hat{H}$ can be summarised as

$$\hat{D}[i,n] = \frac{1}{\sqrt{QN}} \sum_{k=0}^{N-1} \sum_{q=-Q/2}^{Q/2-1} \hat{H}[q,k] e^{j2\pi\frac{qi}{Q}} e^{-j2\pi\frac{kn}{N}}. \qquad (1)$$

---

[1] For example, a basic approach could be to consider a geometry-based single bounce model, which writes $H[q,k] = \sum_{r=0} a_r e^{-j2\pi f_q \tau_r} e^{j2\pi f_r t_k}$ with the speed of light in vacuum c [19], [20]. The index r corresponds to each echo, with a complex amplitude a_r, a propagation range d_r yielding a propagation delay $\tau_r = 2d_r/c$, and a Doppler frequency shift f_r. The index $r = 0$ refers to the direct path between TX and RX.

This article has been accepted for publication in the IEEE Sensors Journal. This is the author's version which has not been fully edited andcontent may change prior to final publication. Citation information: DOI 10.1109/JSEN.2024.3453070

The RDM $\hat{\boldsymbol{D}}$ is the starting point for the crowd classification detailed later. This radar processing can be applied to any estimated CTF matrix, whether it is obtained from measurements or simulations. Note that no static clutter removal is applied, as it would also remove static people in the scene. Robustness to the environment is achieved by training the classifier on a sufficiently representative set of measurement data, for instance, obtained by rotating the directive TX/RX antennas, as explained in Section III. Let us now see how the measurements are performed to obtain a set of real-life measured RDMs. Afterwards, in Section IV, the Method of Moments will be developed to obtain a set of simulated RDMs.

III. EXPERIMENTAL DATASET

Our measurement setup consists of two Universal Software Radio Peripherals (USRPs) X310, connected to 2.4-2.5 GHz directional panel antennas L-com RE09P via Sucoflex 126E cables with SMA connectors. The directional antennas have a gain of 8 dBi and a horizontal beamwidth of 75°. They are used to cancel the direct incident field between TX and RX. The first and second USRPs are each connected to a different antenna, acting as TX and RX respectively. The USRPs are connected via 10 Gigabit Ethernet cables to one single computer equipped with two 10Gtek X520-10G-2S-X8 10-Gigabit Ethernet cards and one 12-core AMD Ryzen 9 3900X CPU clocked at 3.8 GHz. TX and RX are placed in a quasi-monostatic configuration to ease the setup manipulation from one single controlling computer. The clocks of TX and RX are shared to avoid non-idealities such as carrier frequency offset since it is not the focus of this work. Measurements were performed at the carrier frequency $f_c = 2.45$ GHz. Although not standard compliant, a bandwidth $B = 100$ MHz was chosen as a middle ground between the different high bandwidth values of the 802.11ax standard to leverage the improved bandwidth that it allows. The dwell time of each measurement was 0.417 s. All signal parameters are summarised later in Table II.

An outdoor gathering corresponding to the scenario description from Section I-C was analysed. During this event, measurements were collected with our Wi-Fi-based passive radar experimental setup, and processed to obtain RDMs. People joined the gathering progressively, allowing one to take measurements with numbers of people P between 1 and $P_{max} = 100$ people. The number of classes is denoted N_c, with $N_c = 3$ as per the scenario definition in Section I-C. Each class has a class label $y = n_c$, $n_c = 1, \dots, N_c$. The ground truth for the number of people was obtained by manual counting with pictures.

The interval of number of people P in each class was defined by visual analysis during the event. It is given in Table I. In other words, in the rest of this paper, each RDM obtained from a measurement or a simulation with a given number of people P has a label y, assigned based on the value of P, following Table I. Pictures of the scene belonging to each class in this event are shown in Fig. 2. Additionally, it was noticed that people's movements become slower as the crowd gets larger.

TABLE I
CLASSES SUMMARY

Class name	Label	Nb. of people
Low attendance	$y = 1$	$P \in [1, 10]$
Medium attendance	$y = 2$	$P \in [11, 30]$
High attendance	$y = 3$	$P \in [31, 100]$

Fig. 2. Pictures of the measurement scene with the different classes. Class 1 - Low attendance (top left), Class 2 - Medium attendance (top right), Class 3 - High attendance (bottom).

During the measurement session, the bulk of the crowd was moving around the whole area. Hence, the directive antennas of TX and RX were rotated several times to point at it. This presents the advantage of collecting a diverse dataset, preventing it from being tied to one specific scene and one static environment. In total, 1140 measurements were collected, *i.e.* 1140 CTF matrices were estimated, at different time instants during the event, covering the 3 classes described above. Hence, applying the radar processing described in Section II-B yielded a set of 1140 measurement RDMs. They are denoted as $\hat{\boldsymbol{D}}'$, each with a label y' following Table I. The $'$ superscript indicates that they are derived from measurements.

Examples of measured RDMs for each class are given in Fig. 3. The RDMs are normalised by their maximal value in these examples. These examples illustrate the fact that the level of congestion of the RDM, *i.e.* the number of amplitude echoes in the RDM, is not proportional to the level of attendance, which can seem counter-intuitive. This is because when fewer people are present, they are more free to move and there is less masking between them. Hence, it gives rise to higher-speed echoes, to the presence of echoes from people's limbs, and echoes at further range as in the top right plot with 15 people. When the number of people is high, there is more masking, hence there are fewer echoes at high distances. These examples prefigure the difficulty of the crowd size classification task with passive radars and contribute to the way we choose the inputs to the classification algorithm, as discussed in Section V.

This article has been accepted for publication in the IEEE Sensors Journal. This is the author's version which has not been fully edited andcontent may change prior to final publication. Citation information: DOI 10.1109/JSEN.2024.3453070

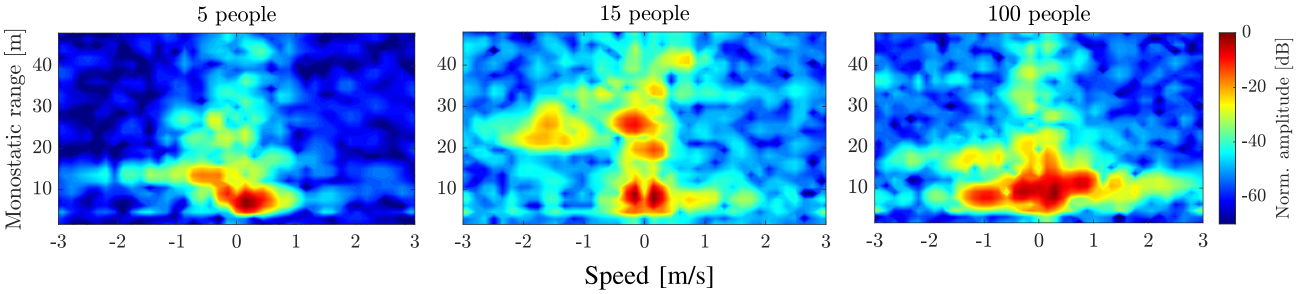


Fig. 3. Example of RDMs from experimental measurements from different classes: class 1 with 5 people (left), class 2 with 15 people (middle), and class 3 with 100 people (right). Clutter removal was applied for visualisation purposes in this example.

IV. METHOD OF MOMENTS FOR SIMULATED RDMS

The simulation counterpart of the crowd scenario is the generation of a large set of RDMs from the MoM. This Section details how to simulate the CTF matrices introduced in Section II-A in the crowd scenario.

A. MoM motivation

We choose to use the MoM over more common ray tracing (RT) techniques for the following reason: RT is only valid in the far field [21] [22], and in the crowd situations considered here, the bodies are not in the far field from each other. Hence diffraction coefficients computed with the approximations of ray tracing would not be valid since the basis hypothesis of RT is violated. To handle this complex scenario, a full-wave technique such as the MoM is required to avoid invalid approximations.

The main drawback of the MoM compared to RT is its computation time, due to the resolution of the large linear system resulting from the discretisation of Maxwell's equations into a large interaction matrix. To tackle this problem, the MoM is accelerated by exploiting the block structure of its interaction matrix, as explained in IV-D.

B. Human Body Model

As already introduced in our previous articles [16], [17], the human body is modelled as a mixture of bones, muscles, fat, organs, blood, skin and humors, yielding a relative permeability $\epsilon_r = 33.37$ and a conductivity $\sigma = 1.24$ S/m at 2.45 GHz. Geometrically, the body is modelled as a dielectric cylindrical shell whose height is infinite following the aforementioned infinite cylinder approximation [15], [23]. The shell thickness is equal to $\Delta + 2\delta$ with δ the electromagnetic skin depth and Δ the MoM discretisation step [16], [17].

C. From electric field to RDM

We consider a given number P of bodies, *i.e.* people, in the observation scene, each moving at their respective speed. The set of all the discretised cells at positions (x, y) in space belonging to all bodies is denoted $\mathcal{B}$. The antenna of the TX device transmitting signals of opportunity emits a vertically polarised frequency-dependent electric field. At any point (x, y) in space, the incident electric field is expressed as

$$E^i_{(x,y)}(f) = \frac{E_0}{\sqrt{d_{\mathrm{TX}}/d_0}} e^{-j2\pi f(d_{\mathrm{TX}}-d_0)/c}, \tag{2}$$

where f is the frequency, $d_{\mathrm{TX}} = \sqrt{(x - x_{\mathrm{TX}})^2 + (y - y_{\mathrm{TX}})^2}$ is the distance between a point (x, y) and the coordinates $(x_{\mathrm{TX}}, y_{\mathrm{TX}})$ of the TX, E_0 is a reference field at a reference distance d_0, and c is the speed of light in vacuum.

In reaction to this incident field, the bodies emit a scattered field $E^s_{(x,y)}(f)$. The total field is $E_{(x,y)}(f) = E^i_{(x,y)}(f) + E^s_{(x,y)}(f)$. The vectors of incident and total fields in the discretised cells in $\mathcal{B}$, *i.e.* inside the P bodies, are denoted $\boldsymbol{E^i}$ and $\boldsymbol{E}$, respectively. From Maxwell's equations and the MoM framework, we have [16], [17], [24]

$$\boldsymbol{C}\boldsymbol{E} = \boldsymbol{E^i}, \tag{3}$$

where the matrix $\boldsymbol{C}$ models the interactions between all bodies illuminated by the incident field from TX [16], [24]. Eq. (3) is solved for $\boldsymbol{E}$. From $\boldsymbol{E}$, Maxwell's equations are used again to obtain the scattered field $E^s_{(x,y)}(f)$ outside the bodies, *i.e.* at the points $(x, y) \notin \mathcal{B}$. A CTF can be computed from the frequency-dependent scattered field $E^s_{(x,y)}(f)$ evaluated by running the MoM at the Q OFDM subcarrier frequencies f_q. Assuming that the incident field is perfectly cancelled at RX and considering the position of the bodies at the slow time instant t_k, we compute the CTF matrix $\boldsymbol{H}$ from the scattered field at RX E^s_{RX,t_k} as [16], [17]

$$H[q, k] = \frac{E^s_{\mathrm{RX},t_k}(f_q)}{E_0}. \tag{4}$$

The field E^s_{RX,t_k} accounts for the electromagnetic interactions between all bodies at instant t_k, hence so does $\boldsymbol{H}$. Additive White Gaussian Noise (AWGN) is also added to $\boldsymbol{H}$ for a chosen value of Signal-to-Noise Ratio (SNR), yielding the simulated version of the estimated CTF, $\hat{\boldsymbol{H}}$. Then, the steps described in Section II-B can be applied to $\hat{\boldsymbol{H}}$ to obtain the RDM $\hat{\boldsymbol{D}}$.

As mentioned before, the RDM $\hat{\boldsymbol{D}}$ is the starting point for the crowd classification presented in Section V. The RDM generation process can be repeated K times for a given number of people P, with K different realisations of the body positions in the crowd, as described below in Section IV-E.

D. Computational efficiency

The main drawback of the MoM w.r.t. RT is its computation time. Indeed, in (3), there is one equation per body discretisation cell, yielding a large linear system to solve. It is accelerated by exploiting the block structure of $\boldsymbol{C}$. A

This article has been accepted for publication in the IEEE Sensors Journal. This is the author's version which has not been fully edited andcontent may change prior to final publication. Citation information: DOI 10.1109/JSEN.2024.3453070

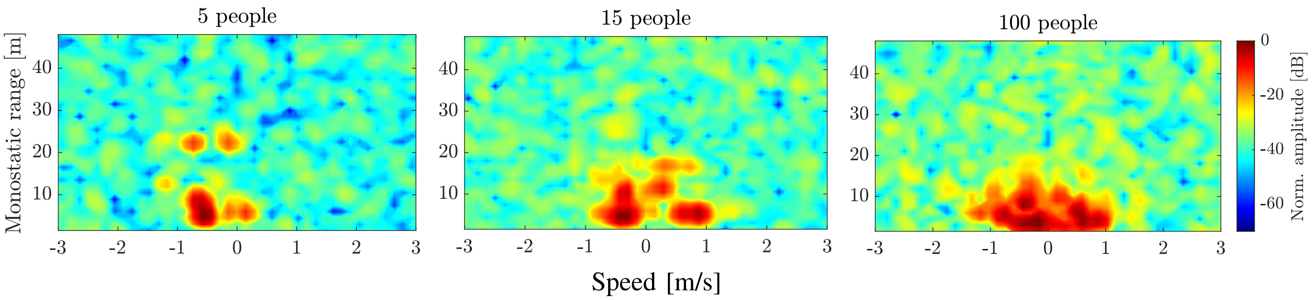


Fig. 4. Examples of RDMs from our MoM-based crowd simulator from different classes: class 1 with 5 people (left), class 2 with 15 people (middle), and class 3 with 100 people (right). Clutter removal was applied for visualisation purposes in this example.

generalised minimal residual (GMRES) iterative solver is used [25], where the matrix-vector product is implemented block by block by compressing non-diagonal blocks with Adaptive Cross Approximation (ACA) [26] and exploiting the symmetry of $\boldsymbol{C}$. A block-Jacobi preconditioner is used to fasten the convergence of GMRES [27]. It is computed from the LU factorisation of the diagonal blocks of $\boldsymbol{C}$.

Moreover, for one given crowd position realisation, (3) has to be solved for each frequency f_q at each slow time instant k. We implement this in a distributed way on a multi-CPU computing cluster, exploiting the fact that the resolutions of (3) are decoupled. Each CPU handles all the N slow time instants for a given frequency f_q, for a given subset of 10 crowd position realisations. This spreads the MoM computation on hundreds of CPUs to efficiently generate RDMs for many crowd position realisations.

TABLE II
SIGNAL & MOM COMPONENTS

Symbol	Name	Value
P_{max}	Maximum number of bodies (people)	100
P	Number of bodies (people)	$1, \ldots, P_{max}$
f_c	Carrier frequency	2.45 GHz
B	Bandwidth	100 MHz
Q	Number of discrete frequencies	32
	Number of fast time instants	32
N	Number of slow time instants	128
	Number of Doppler frequency bins	128
T_s	Fast (sampling) time	1 μs
T	Slow time	33 ms
K	Number of crowd realisations per P	324
E_0	MoM reference electric field	1 V/m
d_0	MoM reference distance	1 m
$\mathcal{B}$	Set of body points	
E^i	Incident electric field	
E^s	Scattered electric field	
E	Total electric field	
$\boldsymbol{C}$	Interaction matrix	
$\boldsymbol{H}$	Channel transfer function matrix	
$\hat{\boldsymbol{H}}$	Estimated channel transfer function matrix	
$\hat{\boldsymbol{h}}$	Range-slow time map	
$\hat{\boldsymbol{D}}$	Range-Doppler map (RDM)	
f	Frequency	
i	Fast time index	
q	Discrete frequency index	
k	Slow time index	
n	Doppler bin index	

E. MoM simulation

With the MoM RDM computation process in mind, the generation of a large amount of simulated RDMs can now be described. The position configuration of the people is generated differently per class, leveraging the spatial configuration of the crowd defined in Section I-C. The values of the signal parameters are given in Table II. The RDM generation works as follows. For each $P = 1, ..., P_{max}$, the following steps are repeated K times:

- Initialisation of the positions of the people in the crowd, *i.e.* the centres of the cylindrical shells modelling them. Those positions are generated depending on the crowd class:
  - Class 1: positions are generated according to a random Binomial Point Process (Binomial PP) in the whole surveillance area [28]. That process models the dispersed spatial configuration of the crowd.
  - Class 2: positions are generated according to a random Matérn cluster PP [28]. This is a classical and intuitive technique to model random clusters. It is a random PP where cluster centres are defined according to the classical Poisson PP, and where for each cluster centre, a random number of points (people) following a Poisson distribution are defined in a disk around the centre. This models the clustered behaviour of the crowd in this class.
  - Class 3: positions are generated according to random Binomial PP in a disk at the centre of the surveillance area. This models the fact that people are all gathered together in one single group.
- Initialisation of the speeds of the people in the crowd. The speeds along the x-axis and the y-axis follow a Gaussian mixture distribution with two components having a mean of -0.15 m/s and 0.15 m/s respectively, and both having a standard deviation of 0.60 m/s, 0.53 m/s, 0.42 m/s for classes 1, 2 and 3 respectively. The two components model the positive or negative speeds. The mean and standard deviation values are empirically set to model the fact that people mainly move slowly while in a group and that their movements become slower as the crowd gets larger, as observed visually and on measured RDMs. Infinite speed values are avoided by removing speed values that create collisions between people during their movement.
- Initialisation of all discretised body cells around those positions, *i.e.* initialisation of the set of $(x, y) \in \mathcal{B}$.
- Computation of the RDM $\hat{\boldsymbol{D}}$ with the MoM as described

This article has been accepted for publication in the IEEE Sensors Journal. This is the author's version which has not been fully edited andcontent may change prior to final publication. Citation information: DOI 10.1109/JSEN.2024.3453070

in Section IV-C. For each P there are K realisations, hence K RDMs.
- Generating the class label y based on the number the number of people in the simulated scene, as defined in Table I.

Examples of RDMs generated by our MoM-based crowd simulator are given in Fig. 4, for the different classes in a low SNR case. They follow identical patterns as the examples of measurement RDMs from Fig. 3, although they contain fewer echoes since the MoM does not model body limbs nor ground reflections, and is two-dimensional.

V. CLASSIFICATION

A. RDM sets & parameters summary

After the measurement session and the simulation data generation, two sets of raw RDMs are thus available. The first is the set of simulated RDMs $\hat{\boldsymbol{D}}_m$, with $m = 1, ..., M$ the simulation RDM index and M the total number of simulated RDMs, here $M = P_{max}K$. For each simulated RDM there is an associated class label y_m. The second is the set of RDMs measured in real life, $\hat{\boldsymbol{D}}'_{m'}$, with the measured RDM index $m' = 1, ..., M'$ and M' the number of measurements performed. As said in Section III, 1140 measurements were performed, hence $M' = 1140$. For each measurement RDM there is an associated class label $y'_{m'}$. Below, we explain the algorithm for the crowd size classification, and how these raw RDMs are converted as input datasets for this algorithm.

As mentioned above, a summary of the signal parameters value and MoM elements is provided in Table II. It is worth noting that the value of Q is equal to 1024 for the measurements, but equal to 32 for the MoM to keep a realistic computation time while still covering a sufficiently long distance to contain all targets in the scene. Hence, the first dimension of the measurement-based RDM is cut to the 32 first samples to ensure that the MoM and measurement-based RDMs have the same dimension.

B. Convolutional Neural Network

The approach for the classification is the following. With the 802.11ac and 11ax Wi-Fi standards, higher bandwidth values are reachable, yielding a finer range resolution, *e.g.* 1.5 m with $B = 100$ MHz (cf. Section III). However, in a large crowd, people are spaced by a distance that is still lower than this range resolution. The same principle holds for the speed dimension. Moreover, some people might be partially or totally masked by others [16]. Hence, we cannot simply only rely on intuitive features like counting amplitude peaks in the RDM after a thresholding.

In our previous work [17], [29], we considered the use of an SVM, but it suffered from two limitations: it operated on hand-designed features, hence possibly non-exhaustive, and the features originated from range and Doppler profiles, hence from one-dimensional profiles only. In the present work, we resort to a CNN [29]–[31] operating on the RDMs to address these two problems, *i.e.* to build a set of more exhaustive synthetic features and to exploit a two-dimensional target separation. A classical CNN architecture is implemented here [29]–[31], with a filter concatenation [30], [32]: each convolution block contains two sets of filters. The first set contains large filters to capture large-scale generic features from the RDM. The second set contains small filters to extract details and fine features from the RDM. The RDM is convolved with both sets of filters, and both resulting feature maps are concatenated along the third dimension, *i.e.* the channel/depth dimension. This is followed by a ReLU non-linear activation [30] and max pooling. For the rest of this article, a convolutional block indicates the group of layers containing the pair of filter sets (large and small), depth concatenation, ReLU and pooling.

To handle the complex values of the RDM, its real and imaginary parts are treated as separate channels [30], *i.e.* the RDM is formatted as a 3D tensor $\widetilde{\boldsymbol{D}}$ where the first frontal slab is the real part of $\hat{\boldsymbol{D}}$ and the second frontal slab is the imaginary part of $\hat{\boldsymbol{D}}$:

$$\widetilde{D}[i,n,1] = \text{Re}\left(\hat{D}[i,n]\right), \quad \widetilde{D}[i,n,2] = \text{Im}\left(\hat{D}[i,n]\right). \quad (5)$$

The size of $\widetilde{\boldsymbol{D}}$ is thus $Q \times N \times 2$. The tensor $\widetilde{\boldsymbol{D}}'$ can be formed similarly from the measurement RDM $\hat{\boldsymbol{D}}'$. The frontal slabs of $\widetilde{\boldsymbol{D}}$ and $\widetilde{\boldsymbol{D}}'$, *i.e.* the real and imaginary parts of $\hat{\boldsymbol{D}}$ and $\hat{\boldsymbol{D}}'$, undergo an instance normalisation separately [31]. They are standardised so that their values have a mean of 0 and a standard deviation of 1. Each formatted RDM $\widetilde{\boldsymbol{D}}_m$ or $\widetilde{\boldsymbol{D}}'_{m'}$ is one data point to be given as input to the CNN, with a class label y_m or $y'_{m'}$ respectively.

The output of the CNN, denoted as *class output*, is the probability of the RDM belonging to one of the N_c classes, computed by a softmax regression function [29]. The classification output, *i.e.* the estimated class label $\hat{y}$, is attributed to the highest probability class. All quantities and parameters related to classification are summarised in Table III, and discussed further in the hyperparameters tuning phase in Section V-D.

C. Training approaches

CNN training approaches using different combinations of the MoM and measurement datasets are listed in Fig. 5. The loss function to minimise during all training approaches is the cross-entropy loss [29]. Approach 1 is used to verify if the classification algorithm and its tuned hyperparameters (cf. Section V-D) is functioning when trained, validated, and tested on simulations data only, *i.e.* in a simpler case than in real life. Approach 2 aims to analyse the performance and limits of the algorithm when trained on simulation only, and validated and tested on experimental data. A first analysis of this approach was carried out in our previous work [17] with a medium-scale crowd of 40 people. This is a very challenging task that will present limits in accuracy. Hence, Approach 3 aims to assess if the system is functional in a more classical training approach, *i.e.* how the algorithm performs with experimental data for training, validation, and testing.

For Approach 1, the training, validation and testing datasets are split as 70%, 20% and 10% of the total MoM dataset. For Approach 2, the training dataset consists of 90% of the total

This article has been accepted for publication in the IEEE Sensors Journal. This is the author's version which has not been fully edited andcontent may change prior to final publication. Citation information: DOI 10.1109/JSEN.2024.3453070

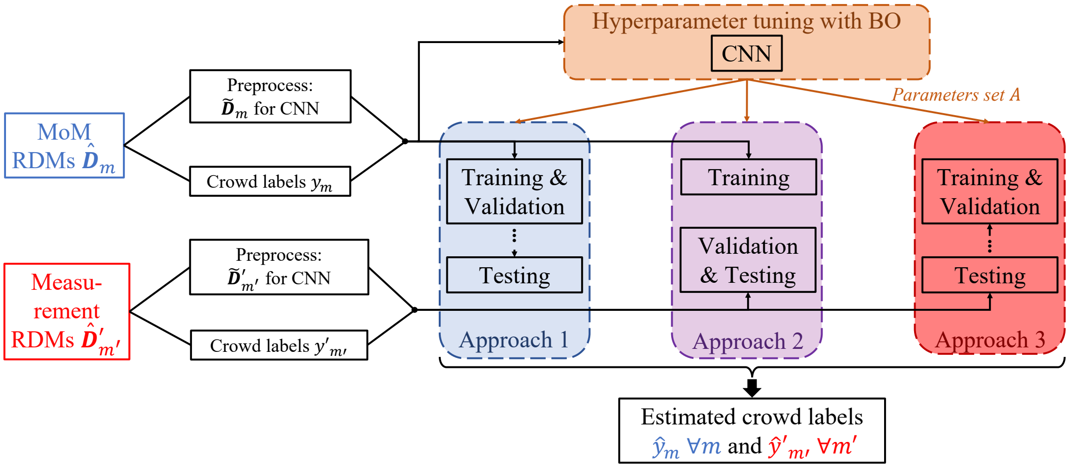


Fig. 5. Hyperparameter tuning data & Training, validation, and testing approaches.

TABLE III
CLASSIFICATION & CNN NOTATIONS

Symbol	Name
$\tilde{\boldsymbol{D}}$	Formatted RDM as CNN input
y	Data point class label
m	Data point index
M	Number of data points
$\ldots'$	Measurement corresponding element
N_{b_1}	Nb. of convolutional blocks in CNN
N_{b_2}	Nb. of fully connected layers in CNN
N_f	Nb. of filters in one set of filters in CNN
N_{FC}	Nb. of neurons in one FC layer in CNN
a	Large filter kernel size
b	Small filter kernel size

MoM dataset [2]. The validation and testing datasets each consist of 50% of the total measurements dataset. For Approach 3, the training, validation and testing datasets are split as 70%, 20% and 10% of the total measurement dataset.

The performance of each approach is evaluated through the confusion matrix computed on the testing set. It is a square matrix where diagonal elements contain the percentage of correct classifications per class, while each off-diagonal element contains the percentage of incorrect classifications between the true class corresponding to the element row index and the wrong class corresponding to the element column index [33]. Hence, one wants to obtain the highest possible diagonal values and the lowest possible off-diagonal values. Multiple realisations of the training, validation, and testing datasets are drawn: for each realisation, data points are randomly assigned to each set. The average and best confusion matrix are computed on the testing set in all realisations and are used as performance metrics.

D. Hyperparameter tuning procedure

Before evaluating the three training approaches, the CNN has hyperparameters to tune, *i.e.* structural parameters of the algorithms. This step is critical, as these parameters affect drastically the classification capabilities of the algorithms. The hyperparameters are:

- In the feature extraction part, *i.e.* the convolutional blocks:
  - The number of convolutional blocks N_{b_1}.
  - The number of filters N_f per convolution filter set.
  - The kernel sizes a and b of the large and small convolution filters respectively.
- In the classification part, *i.e.* the fully connected layers:
  - The number of fully connected layers N_{b_2}.
  - The number of neurons N_{FC} in each layer.

This set of CNN hyperparameters is noted $A = \{N_{b_1}, N_f, a, b, N_{b_2}, N_{FC}\}$. The convolution stride and the pooling stride are set at the classical values of [1 1] and 2 respectively [30], they were not tuned to avoid enlarging unnecessarily the hyperparameters search space.

Due to the computational complexity of training CNNs with multiple parameter sets, and the absence of a deterministic cost function, a Bayesian Optimisation (BO) strategy is chosen here [34]–[36]. It tackles an optimisation problem of the form $\hat{A} = \text{argmin}_A\ f(A)$ where $f(A)$ is the classification error rate (CER) computed on the validation data for a CNN with a hyperparameter set A. This error rate is the percentage of misclassifications of the CNN, *i.e.* the sum of the off-diagonal elements of the confusion matrix divided by the number of classes N_c. The optimisation problem is tackled iteratively. Each iteration tests one hyperparameter set and consists of two parts: firstly, the CNN is trained with the hyperparameter set under test by minimising the cross-entropy loss on the training data. Secondly, the BO objective function, *i.e.* the classification error rate $f(A)$, is evaluated on the validation data with the trained CNN [34]. BO performs a given number of starting iterations. Then, based on the obtained objective function values, it models the objective function as a Gaussian process, and tries to pick the hyperparameter sets for the next iterations in the good direction to reach a local minimum. This direction is chosen based on an *acquisition function* that is less costly to evaluate than the objective function [34]. To reduce the dependency on the initial random combinations of

---

[2] The remaining 10% are left out to enable diversity among the different realisations of the datasets shuffle described below.

This article has been accepted for publication in the IEEE Sensors Journal. This is the author's version which has not been fully edited andcontent may change prior to final publication. Citation information: DOI 10.1109/JSEN.2024.3453070

hyperparameter values used by the algorithm for its estimation, and to explore multiple local minima, BO is run multiple times. The set $\hat{A}$ corresponding to the lowest classification error rate among the different runs is selected as the final set to use.

This hyperparameter tuning procedure is performed using only MoM simulation data, as summarised in the top part of Fig. 5, to highlight a very helpful use of simulations in that type of classification problem. Indeed, hyperparameter tuning is a long and difficult procedure, hence being able to perform it with simulation data only prior to a measurement session is critical to save time and resources. In other words, for all three training approaches, the CNN hyperparameters are the ones derived from the BO procedure on MoM simulation data only. This will show that hyperparameters tuned with simulation data can be used for crowd size classification on experimental measurement data.

The BO results for the CNN hyperparameter tuning are discussed in Section VI-A. The obtained hyperparameter set is thus saved and subsequently used for all training approaches. Before discussing the results, a block diagram summarising the system is provided in Fig. 5. The algorithms described in this section are implemented using Matlab and its Statistics & Machine Learning and Deep Learning toolboxes.

VI. CLASSIFICATION RESULTS

A. Hyperparameter tuning results

The range of values considered by the BO for the CNN tuning is given in the first line of Table IV. It was chosen to set a low maximal possible number of convolutional blocks and FC layers to keep the number of parameters to train at reasonable values w.r.t. the limited number of training data points, to avoid overfitting [30]. The result of the best BO run for the CNN hyperparameter tuning is displayed in blue in the top plot of Fig. 6. The best hyperparameter set is highlighted in red and given in the second line of Table IV. A comparison with a run of BO applied on measurement data is displayed in the bottom plot of Fig. 6 in light grey.

From the BO blue curve (MoM data) in Fig. 6, it is noticed that using only $N_{b_1} = 1$ convolutional layer (cf. first point of the curve) is not enough to obtain a decent accuracy. This means that several convolutional layers are necessary to extract meaningful features for the crowd size classification from the RDMs. The BO reaches low error values for either $N_{b_1} = 3$ convolutional layers with $N_{b_2} = 2$ FC layers or $N_{b_1} = 2$ convolutional layers with only $N_{b_2} = 1$ FC layer. In both cases, the number of neurons N_{FC} has to be kept high, close to the allowed maximum $N_{FC_{max}} = 200$, to model the separation between classes with enough accuracy. The best large filter size a is the largest possible, *i.e.* $a = 7$. The small filter size b evolves during the BO but is finally set as the smallest possible, *i.e.* $b = 2$. These values of a and b mean that it is better for the classification to have a wider diversity of features through significantly different filter sizes.

As a comparison, the hyperparameter set obtained from a BO run applied on measurements (light grey curve) is $\{2, 58, 7, 4, 1, 193\}$. It is very close to that obtained with the MoM data, the main difference being the larger size of the small filter ($b = 4$) compared to that obtained with the MoM data ($b = 2$). This proximity of the MoM-based hyperparameters tuning with what would be obtained with measurements shows the relevance of tuning the hyperparameters with MoM data.

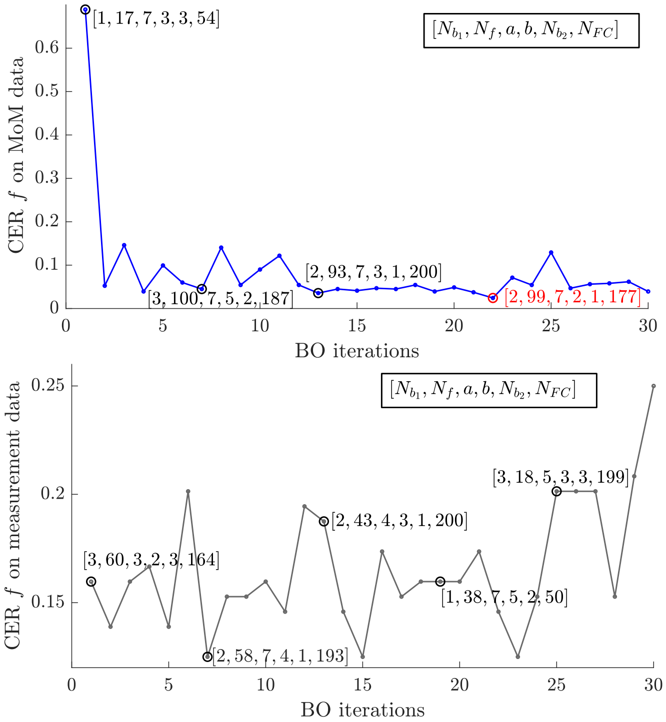


Fig. 6. Top plot, blue line: CNN hyperparameter tuning on MoM data with BO, classification error rate for the best run. The set of hyperparameters corresponding to the minimum of the classification error function is highlighted in red. Bottom plot, grey line: comparison run of the BO applied on measurement data. The set of hyperparameters used in the following Sections VI-B, VI-C, and VI-D is that from the MoM, *i.e.* in red.

TABLE IV
CNN HYPERPARAMETER TUNING RESULTS

	N_{b_1}	N_f	a	b	N_{b_2}	N_{FC}
Value range	$[1, 3]$	$[10, 100]$	$[3, 7]$	$[2, 6]$	$[1, 3]$	$[50, 200]$
BOs result	2	99	7	2	1	177

B. Training Approach 1: training, validation, and testing on MoM

With the tuned CNN, Approach 1 can be considered, *i.e.* training, validating and testing the CNN on MoM data. The averaged (left) and best (right) confusion matrices on the testing set on 25 realisations of the dataset shuffling are displayed in Fig. 7. The classification accuracy is above 95% for each class. There is no misclassification between extreme classes, which is critical since that kind of misclassification corresponds to a high error in the number of people. This successful Approach 1 illustrates that the proposed method and its tuned hyperparameters set are consistent and that it can yield high performance on ideal data from simulations. It opens the way for Approach 2 with real-life experimental measurement data.

This article has been accepted for publication in the IEEE Sensors Journal. This is the author's version which has not been fully edited andcontent may change prior to final publication. Citation information: DOI 10.1109/JSEN.2024.3453070

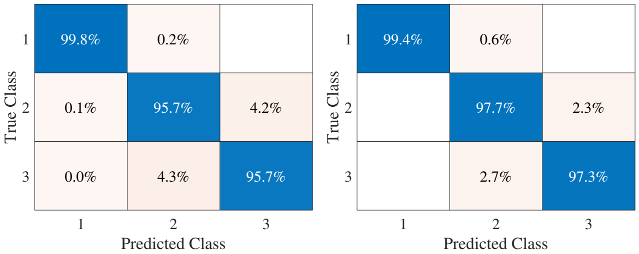


Fig. 7. Approach 1 (training, validation, and testing on MoM data) - Averaged (left) and best (right) confusion matrices for CNN-based crowd size classification.

C. Training Approach 2: training and validation on MoM, testing on measurements

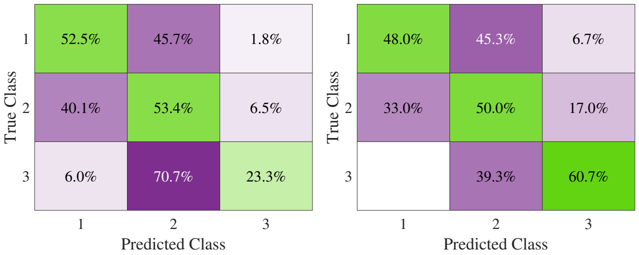


Fig. 8. Approach 2 (training on MoM data, validation and testing on measurement data) - Averaged (left) and best (right) confusion matrices for CNN-based crowd size classification. A modified colour scheme is used to emphasise the fact that these confusion matrices are computed on experimental measurement data.

The confusion matrices are displayed in Fig. 8 for Approach 2, *i.e.* when the training is performed on MoM simulation data and the validation and testing on measurement data. The accuracy obtained for Approach 2 is limited. Namely, the CNN manages on average (left plot) to discriminate between 2 classes, although with limited accuracies. The majority of points in class 3 are however not correctly classified. In the best case (right plot), the three classes are discriminated, but with limited accuracies. However, the misclassifications between extreme classes 1 and 3 are limited, which is positive.

These low-accuracy results are expected due to the limitations of the MoM simulation data: body limbs and ground reflections are not modelled, and the simulation is two-dimensional. Hence, the feature space of features extracted from the MoM RDMs only matches partially the feature space from experimental measurements. Thus, by training the algorithms on MoM data only, they are trained to extract features that are not always relevant and to classify based on them. This partial feature match yields poor accuracy on average but allows intermediate accuracies in some training realisations. In our previous work [17] with a medium-scale crowd of 40 people and a simple SVM, this approach yielded acceptable performances (72.15% accuracy on average) for $N_c = 2$ but a limited accuracy (52.62%) for $N_c = 3$. This previous work showed that in a simplified 2-classes case, MoM-based training can yield acceptable results.

Let us follow the same philosophy here, and see how Approach 2 performs on the 100-people crowd in a simplified case. Coming back to the scenario, we consider a case where the people monitoring the event would need to know when the attendance level crosses a given threshold and transitions from low/medium to high, which could lead to a dangerous situation. This translates into a binary classification with class $n_c = 1$ corresponding to $P \in [1, 30]$ and class $n_c = 2$ corresponding to $P \in [31, 100]$. The obtained averaged and best confusion matrices are given on the left and right of Fig. 9. The accuracy is much higher in this simplified case than in the 3-class case: both classes exhibit close accuracies, and the average and best mean accuracy are 76.40% and 89.17% respectively. With Approach 2, we thus see that MoM data contains the required features to classify the crowd size with simulation-only training in a simplified binary classification case.

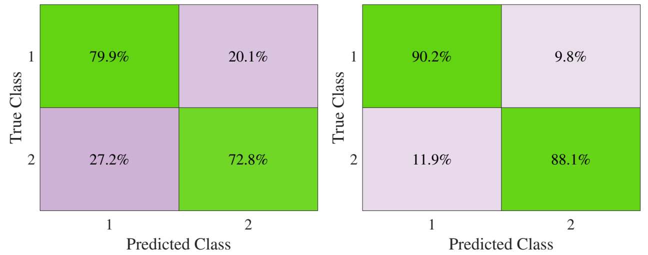


Fig. 9. Approach 2 (training on MoM data, validation and testing on measurement data) in a simplified 2-class case with class $\boldsymbol{n_c} = \mathbf{1}$ corresponding to $\boldsymbol{P} \in [\mathbf{1}, \mathbf{30}]$ and class $\boldsymbol{n_c} = \mathbf{2}$ corresponding to $\boldsymbol{P} \in [\mathbf{31}, \mathbf{100}]$ - Averaged (left) and best (right) confusion matrices for CNN-based crowd size classification.

D. Training Approach 3: training, validation, and testing on measurements

For Approach 3, the confusion matrices (computed on real-life experimental data, cf. Fig. 5), are displayed in Fig. 10. The CNN exhibits an accuracy of 85.7% on all 3 classes for the averaged confusion matrix. The accuracy of class 2, *i.e.* the intermediate class, is slightly lower than that of the other classes: the CNN yields 80.1% accuracy on average for this class, and misclassifications are spread between classes 1 and 3. This is expected since the feature space of class 2 is close to the one of class 1 and class 3 simultaneously, as previously highlighted in our past work with $P_{max} = 40$ people [17]. Moreover, class 2 corresponds to the point where there starts to be more people in the scene, meaning that the different people's contributions to the RDM become more difficult to resolve with the limited radar resolution. With this increased number of people, the electromagnetic masking effects are also more important. Those 3 effects are responsible for class 2 to be sometimes confused with its neighbouring classes. The accuracy of class 1 reaches a higher value, 84.2% on average, and the CNN achieves a very high accuracy for class 3, 92.5% on average. There is around 1% misclassification between those classes, hence there is almost no case with high size classification error. In the best case, *i.e.* in the confusion matrix on the right of Fig. 10, class 3 can achieve a 100% accuracy. Class 3 has the highest accuracy, as in the averaged case. Class 2 remains the one yielding the lowest accuracy, but still achieves 87%.

This article has been accepted for publication in the IEEE Sensors Journal. This is the author's version which has not been fully edited andcontent may change prior to final publication. Citation information: DOI 10.1109/JSEN.2024.3453070

The mean accuracy on all 3 classes is 93.15% for the best case. These high accuracy values demonstrate the feasibility of our crowd size classification classification scheme on real-life experimental data, with algorithm hyperparameters tuned on MoM simulation data only.

E. Comparison with hyperparameter set obtained from measurements

As a comparison, all Approaches are also run with the hyperparameter set obtained with the BO run on the measurement data, *i.e.* from the bottom plot of Fig. 6. In Fig. 11, we show the global accuracy over all classes, *i.e.* the sum of the diagonal elements of the confusion matrix divided by the number of classes, for the classification task with hyperparameters from the MoM-based BO run (from Fig. 7, 8, 9, and 10) and with new tests using hyperparameters from the measurement-based BO run. It can be noticed that for all Approaches, the accuracy difference between the results obtained with these different hyperparameter sets is negligible. In other words, this shows the relevance of performing the hyperparameter tuning only with MoM data without requiring measurement data.

F. Classification results discussion

Regarding the different approaches, in Approach 1 we have seen that training and testing the proposed classification algorithms on simulation-only data yielded a very high accuracy, verifying the feasibility of both methods. In Approach 2, we have seen in [17] that transferring models trained on MoM simulations only to experimental measurements is functional for average-scale crowd scenarios. For large-scale crowds, in this paper, the classification accuracy was low due to the limitations of the simulation data compared to reality. The CNN could still perform moderately in some cases but with limited accuracy. However, with a simplified binary classification scheme for the large-scale crowd, Approach 2 yielded a much higher accuracy, showing that MoM data still contain the required features to classify the crowd size in a simplified case. Finally, Approach 3, with the large-scale crowd and the three classes, showed that the proposed classification scheme was functional with experimental measurement data, reaching more than 85% accuracy on average. Hence, in a real-life deployment:

- For a binary classification: MoM data can be used for the hyperparameter tuning and for the training before even deploying the measurement setup and classifying the measured crowd size.
- For a 3-classes classification scenario: MoM data are useful for the hyperparameter tuning before deploying the measurement setup, and the training is left to the measurement data collected after the deployment.

To elaborate further on the results of Approach 3, we can compare the performances obtained in this work with the results of the other works mentioned in Section I. This comparison is summarised in Table V. For each work, the considered number of people, the number of classes, the main algorithm, and the mean accuracy over all classes are given in the Table. For the counting in [4], there are no classes since it features an exact counting algorithm and not a classification algorithm. In the second last line of the Table, we also propose a comparison with the SVM from our previous work [17] which relied on hand-designed features from range and Doppler profiles, that we extended and re-tuned to the present crowd scenario. The last line of the Table corresponds to the present work following Approach 3, with the average (85.7%) and best (93.15%) accuracies. It can be noticed that the classification accuracy reached by other works is around 10% above our average accuracy, and close to our best accuracy. As can be seen in the second column of the Table, and as already mentioned in Section I, the number of people considered here is much larger than in the other works. That is, we managed to perform crowd size classification with accuracies comparable to other works but with a larger crowd size.

By comparing the two last lines of Table V, we also note that the CNN proposed here performs much better than the SVM from our previous work [17] that used hand-designed features computed on range and Doppler profiles. This is

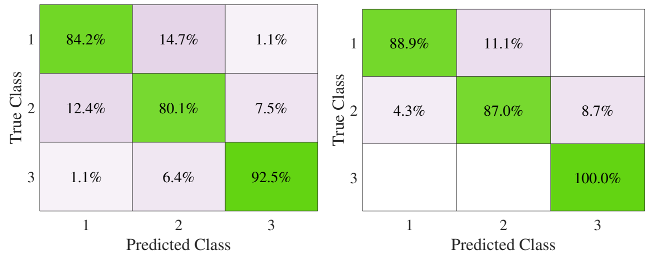


Fig. 10. Approach 3 (training, validation, and testing on measurement data) - Averaged (left) and best (right) confusion matrices for CNN-based crowd size classification. A modified colour scheme is used to emphasise the fact that these confusion matrices are computed on experimental measurement data.

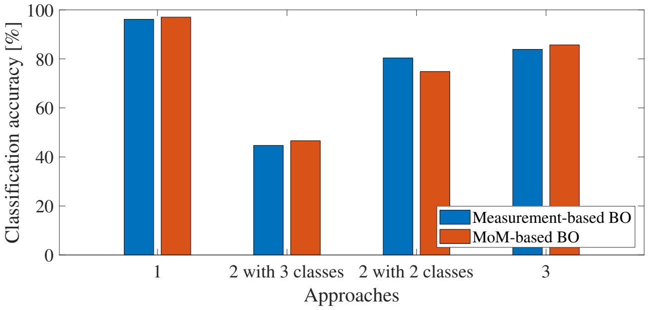


Fig. 11. Global classification accuracy comparison between the hyperparameter set from the measurement-based and MoM-based BO.

TABLE V
ACCURACY COMPARISON FOR APPROACH 3

Work	#people	#classes	Algorithm	Accuracy
Event classification in [6]	1	5	SVM	98%
Detection part in [7]	4	2	CNN	96%
Counting in [7]	4	5	CNN	93%
Counting in [8]	5	5	LSTM-CNN	86%-90%
Counting in [4]	20	/	Renewal Proc.	94%
Our SVM [17] extended	100	3	SVM	70%
Present work (approach 3)	100	3	CNN	86%-93%

because the hand-designed features for the SVM cannot be guaranteed to be exhaustive for the considered classification problem since the classification part is decoupled from the feature computation part. The CNN builds its own synthetic features that are optimised to minimise the classification error so it does not suffer from this problem. The difference in performance is also caused by the use of one-dimensional profiles with the SVM, not leveraging the separation of the targets in two dimensions and the spatial structure of the RDM.

VII. CONCLUSIONS

In conclusion, we showed that a Wi-Fi-based passive radar using preambles from the recent 802.11ax standard is able to perform crowd counting with experimental measurements of a large crowd containing up to 100 people during a real-life event. We proposed the first crowd simulator based on MoM to generate RDMs for large crowds. We tackled the counting as a classification problem with the RDMs as inputs, and demonstrated that we could tune the hyperparameters of a reference classifier, a CNN here, with MoM simulation data only, and even train the classifier on these data in a simplified case. We collected experimental measurements with a Wi-Fi-based passive radar setup based on USRPs, and obtained high classification accuracies on these measurements.

Future work might investigate a finer or more flexible class subdivision. The CNN output would be expanded to more classes, and the CNN architecture would be re-tuned to address this new classification problem. A regression model could also be implemented to try to reach an actual count. A refinement of the classification or counting could also be performed over time with multiple successive measurements.

ACKNOWLEDGEMENTS

Computational resources have been provided by the Consortium des Équipements de Calcul Intensif (CÉCI), funded by the Fonds de la Recherche Scientifique de Belgique (F.R.S.-FNRS) under Grant No. 2.5020.11 and by the Walloon Region.

REFERENCES

[1] U. Singh, J.-F. Determe, F. Horlin, and P. De Doncker, "Crowd monitoring: State-of-the-art and future directions," *IETE Technical Review*, vol. 38, no. 6, pp. 578–594, 2021.

[2] J.-F. Determe, U. Singh, F. Horlin, and P. De Doncker, "Forecasting crowd counts with wi-fi systems: Univariate, non-seasonal models," *IEEE Trans. Intell. Transp. Syst.*, vol. 22, no. 10, pp. 6407–6419, 2020.

[3] S. Di Domenico, M. De Sanctis, E. Cianca, P. Colucci, and G. Bianchi, "Lte-based passive device-free crowd density estimation," in *2017 IEEE Int. Conf. on Commun. (ICC)*. IEEE, 2017, pp. 1–6.

[4] S. Depatla and Y. Mostofi, "Crowd counting through walls using wifi," in *2018 IEEE international conference on pervasive computing and communications (PerCom)*. IEEE, 2018, pp. 1–10.

[5] IEEE 802.11, "Wi-Fi sensing," 2019, Last consult.: Feb. 2023. [Online]. Available: https://mentor.ieee.org/802.11/documents?is_dcn=DCN%2C%20Title%2C%20Author%20or%20Affiliation&is_group=SENS

[6] W. Li, B. Tan, and R. J. Piechocki, "Wifi-based passive sensing system for human presence and activity event classification," *IET Wireless Sensor Systems*, vol. 8, no. 6, pp. 276–283, Dec. 2018.

[7] C. Tang, W. Li, S. Vishwakarma, K. Chetty, S. Julier, and K. Woodbridge, "Occupancy detection and people counting using wifi passive radar," in *2020 IEEE Radar Conference*. IEEE, 2020, pp. 1–6.

[8] S. Liu, Y. Zhao, F. Xue, B. Chen, and X. Chen, "Deepcount: Crowd counting with wifi via deep learning," *arXiv preprint arXiv:1903.05316*, 2019.

[9] S. Bartoletti, A. Conti, and M. Z. Win, "Device-free counting via OFDM signals of opportunity," in *2018 IEEE International Conference on Communications Workshops*. IEEE, 2018, pp. 1–5.

[10] A. M. Eid and J. W. Wallace, "Accurate modeling of body area network channels using surface-based method of moments," *IEEE Transactions on Antennas and Propagation*, vol. 59, no. 8, pp. 3022–3030, Aug. 2011.

[11] R. Neely, K. Naishadham, A. Sharma, and K. Bing, "Electromagnetic human body modeling with physiological motion for radar applications," in *2012 IEEE Radar Conference*. IEEE, 2012, pp. 0818–0823.

[12] M. Cheffena and F. Pérez-Fontán, "Channel simulator for land mobile satellite channel along roadside trees," *IEEE Transactions on antennas and propagation*, vol. 59, no. 5, pp. 1699–1706, May 2011.

[13] L. L. Foldy, "The multiple scattering of waves. i. general theory of isotropic scattering by randomly distributed scatterers," *Phys. Rev.*, vol. 67, pp. 107–119, Feb. 1945. [Online]. Available: https://link.aps.org/doi/10.1103/PhysRev.67.107

[14] M. Lax, "Multiple scattering of waves," *Reviews of Modern Physics*, vol. 23, no. 4, p. 287, Oct. 1951.

[15] H. Huang, L. Tsang, A. Colliander, and S. H. Yueh, "Propagation of waves in randomly distributed cylinders using three-dimensional vector cylindrical wave expansions in foldy–lax equations," *IEEE J. Multiscale Multiphysics Comput. Tech.*, vol. 4, pp. 214–226, Dec. 2019.

[16] L. Storrer, H. C. Yildirim, J. Louveaux, P. De Doncker, S. Pollin, and F. Horlin, "Impact of inter-body scattering on people counting with wi-fi sensing," in *2022 2nd IEEE Int. Sympos. Joint Commun. & Sensing (JC&S)*. IEEE, 2022, pp. 1–6.

[17] L. Storrer, H. C. Yildirim, M. Willame, J. Louveaux, P. De Doncker, S. Pollin, and F. Horlin, "Crowd counting model training with the method of moments in electromagnetics," in *2023 3rd IEEE Int. Sympos. Joint Commun. & Sensing (JC&S)*. IEEE, 2023, pp. 1–6.

[18] Aruba Networks, "White paper: 802.11ax," 2019, Last consultation: February 2020. [Online]. Available: https://www.arubanetworks.com/assets/wp/WP_802.11AX.pdf

[19] Y. S. Cho, J. Kim, W. Y. Yang, and C. G. Kang, *MIMO-OFDM Wireless Communications with MATLAB*. Wiley Publishing, 2010.

[20] L. Storrer, H. C. Yildirim, M. Crauwels, E. I. P. Copa, S. Pollin, J. Louveaux, P. De Doncker, and F. Horlin, "Indoor tracking of multiple individuals with an 802.11ax wi-fi-based multi-antenna passive radar," *IEEE Sensors Journal*, vol. 21, no. 18, pp. 20 462–20 474, Sept. 2021.

[21] P. Pathak, W. Burnside, and R. Marhefka, "A uniform gtd analysis of the diffraction of electromagnetic waves by a smooth convex surface," *IEEE Trans. Antennas Propag.*, vol. 28, no. 5, pp. 631–642, 1980.

[22] C. A. Balanis, *Advanced engineering electromagnetics*. John Wiley & Sons, 2012.

[23] L. Tsang, J. A. Kong, and R. T. Shin, "Theory of microwave remote sensing," 1985.

[24] J. Richmond, "Scattering by a dielectric cylinder of arbitrary cross section shape," *IEEE Trans. Antennas Propag.*, vol. 13, no. 3, pp. 334–341, May 1965.

[25] Y. Saad and M. H. Schultz, "GMRES: A generalized minimal residual algorithm for solving nonsymmetric linear systems," *SIAM Journal on scientific and statistical computing*, vol. 7, no. 3, pp. 856–869, 1986.

[26] K. Zhao, M. N. Vouvakis, and J.-F. Lee, "The adaptive cross approximation algorithm for accelerated method of moments computations of emc problems," *IEEE Trans. Electromagn. Compat.*, vol. 47, no. 4, pp. 763–773, Nov. 2005.

[27] G. Burger, H.-D. Bruns, and H. Singer, "Advanced method of moments based on iterative equation system solvers," in *IEEE 1997, EMC, Austin Style. IEEE 1997 Int. Symp. Electromagn. Compat. Symposium Record (Cat. No. 97CH36113)*. IEEE, 1997, pp. 236–241.

[28] Q. Gontier, C. Tsigros, F. Horlin, J. J. Wiart, C. Oestges, and P. De Doncker, "Modeling the spatial distributions of macro base stations with homogeneous density: theory and application to real networks," in *Proc. of the European Cooperation in Science and Technology*, 2022.

[29] T. Hastie, R. Tibshirani, J. H. Friedman, and J. H. Friedman, *The elements of statistical learning: data mining, inference, and prediction*. Springer, 2009, vol. 2.

[30] M. S. Seyfioğlu, A. M. Özbayoğlu, and S. Z. Gürbüz, "Deep convolutional autoencoder for radar-based classification of similar aided and unaided human activities," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 54, no. 4, pp. 1709–1723, 2018.

[31] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*. MIT Press, 2016, http://www.deeplearningbook.org.

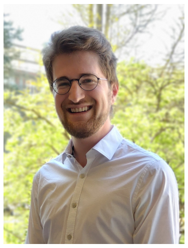
[32] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich, "Going deeper with convolutions," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2015, pp. 1–9.

[33] S. Theodoridis and K. Koutroumbas, *Pattern recognition*. Elsevier, 2006.

[34] J. Snoek, H. Larochelle, and R. P. Adams, “Practical bayesian optimization of machine learning algorithms,” *Advances in neural information processing systems*, vol. 25, 2012.

[35] M. I. Sameen, B. Pradhan, and S. Lee, “Application of convolutional neural networks featuring bayesian optimization for landslide susceptibility assessment,” *Catena*, vol. 186, p. 104249, 2020.

[36] H. Cho, Y. Kim, E. Lee, D. Choi, Y. Lee, and W. Rhee, “Basic enhancement strategies when using bayesian optimization for hyperparameter tuning of deep neural networks,” *IEEE access*, vol. 8, pp. 52 588–52 608, 2020.



Laurent Storrer (M’2019) received his PhD in electrical engineering from the Université Libre de Bruxelles (ULB), Brussels, Belgium & the Katholieke Universiteit Leuven (KULeuven) in 2023. He is working on Wi-Fi-based passive radars, FMCW radars, target tracking, and classification for crowd monitoring. In 2019, he carried out an internship at the Inter-university Micro-Electronics Centre (IMEC) on Wi-Fi-based sensing. He is currently a postdoctoral researcher at ULB and Macq.

Hasan Can Yildirim was born in Turkey in 1991. He received his PhD from ULB in 2022. Since October 2022, he has been a postdoctoral researcher at the OPERA department. Currently, he is working on passive radars based on emerging Wi-Fi systems.

Martin Willame received the electrical engineering degree from the Université catholique de Louvain (UCLouvain), Louvain-la-Neuve, Belgium in 2021. Since October 2021, he has been a PhD student at UCLouvain and ULB with a scholarship from F.N.R.S.-E.O.S. He is working on Wi-Fi-based passive radars and multistatic systems.

Evert I. Pocoma Copa was born in 1990 in El Alto, Bolivia. He received a scholarship from the Foundation Simon I. Patiño that helped him to receive the M.Sc. degree in electrical engineering from ULB & VUB in 2019. Since September 2019, he has been working toward the Ph.D degree with the OPERA-WCG, ULB. He is currently working on distributed localisation and tracking techniques in wireless sensor networks.

Dejvi Cakoni was born in Albania in 1998. He received his MSc in electrical engineering from the joint program between ULB and VUB in 2022. In March 2022, he joined the OPERA Wireless Communication Group as a PhD student. He is currently working on target counting and classification based on micro-Doppler signatures captured by active radars.”

Sofie Pollin (IEEE Senior Member) obtained her PhD degree at KULeuven in 2006. From 2006-2008 she continued her research on wireless communication, energy-efficient networks, cross-layer design, coexistence and cognitive radio at UC Berkeley. In November 2008 she returned to imec to become a principal scientist in the green radio team. Currently, she is an associate professor at the electrical engineering department at KU Leuven. Her research centres around constrained Networked Systems.

Jérôme Louveaux received the Ph. D. degree from UCLouvain in 2000. In 2000-2001, he was a visiting scholar at Stanford University, CA. In 2004-2005, he was a postdoctoral researcher at the Delft University of Technology, Netherlands. Since 2006, he has been a Professor in the ICTEAM institute at UCLouvain. His research interests are in signal processing for digital communications, and in particular: multicarrier modulations, xDSL systems, resource allocation, synchronisation and estimation.

Philippe De Doncker received the M.Sc. degree in physics engineering and the Ph.D. degree in science engineering from ULB, in 1996 and 2001, respectively. He is currently a Professor with ULB, where he also leads the research activities on wireless channel modelling and electromagnetics.

François Horlin received the Ph.D. degree from UCLouvain in 2002. He specialised in the field of signal processing for digital communications. His Ph.D. research aimed at optimising the multi-access for 3G cellular communications. He joined IMEC in 2006 as a senior scientist. In 2007, he became professor at ULB. He is supervising a research team working on communication systems. Recently, the team focused on Wi-Fi-based passive radars for crowd monitoring.