Crowd Counting Model Training with the Method of Moments in Electromagnetics

Laurent Storrer*‡, Hasan Can Yildirim*†, Martin Willame†,
Jérôme Louveaux†, Philippe De Doncker*, Sofie Pollin‡, François Horlin*

*Université Libre de Bruxelles - †Université Catholique de Louvain - ‡Katholieke Universiteit Leuven
{*laurent.storrer,hasan.can.yildirim*}*@ulb.be*, {*martin.willame,jerome.louveaux*}*@uclouvain.be*,
sofie.pollin@kuleuven.be, {*pdedonck,fhorlin*}*@ulb.ac.be*

2023 IEEE 3rd International Symposium on Joint Communications & Sensing (JC&S) | 979-8-3503-4568-1/23/$31.00 ©2023 IEEE | DOI: 10.1109/JCS57290.2023.10107513

***Abstract*—We investigate crowd counting for Wi-Fi sensing based on models trained with data from the two-dimensional method of moments for electromagnetic scattering. Those models are applied on measurements with a high number of people. We simulate a crowd with people modelled as dielectric cylindrical shells exposed to an incident electric field from a sensing-enabled Wi-Fi access point. The electric field scattered by the people is computed by the method of moments, taking into account all electromagnetic interactions between the bodies. The scattered field is used to compute channel transfer functions coefficients, from which range and Doppler profiles are derived. A set of channel features, statistical features and impulsive metrics is computed on the profiles, and reduced with Principal Component Analysis. A part of the feature set is used to train a Support Vector Machine classifier, whose output classes are ranges of values of people in the crowd. A good accuracy is reached when testing the classifier on the remaining simulation points. The classifier trained on the method of moments data is then applied on real-life measurements of a crowd, performed with Universal Software Radio Peripherals, and reaches a promising accuracy.**

***Index Terms*—Wi-Fi sensing, Method of Moments, crowd monitoring, people counting, Support Vector Machine.**

I. Introduction

Crowd management is a critical point in events and public space management to avoid overcrowding leading to injuries or even death of people, as it happened in Hillsborough, UK, in 1989 in a football stadium, at Duisburg, Germany, in 2010 during the Love Parade or more recently at the Astrowold festival in Houston, USA, in 2021 [1]. A key element in this management is crowd counting, for the organizers and security staff to know in real time when the number of people in a surveillance area is increasing beyond limits.

Due to the quasi-ubiquitous availability of wireless signals of opportunity, RF-based crowd counting is a potential solution. Moreover, with the ongoing work of the 802.11 Task Group for WLAN sensing [2], Wi-Fi-based sensing approaches for crowd counting are under the spotlights. Features can be extracted from the Doppler spectrogram obtained with a Cross-Ambiguity Function processing on Wi-Fi signals of opportunity and fed to a Support Vector Machine (SVM) classifier [3]. Range-Doppler maps and Doppler spectrograms can also be given as input to convolutional neural networks (CNNs) for counting, but with a limited number of people [4].

However, the parameters of the counting algorithms in those works are computed using measurement data only, hence requiring considerable time and effort to install the measurement setup and to collect enough data points for a robust training of the counting algorithms. Furthermore, those works rely on old versions of the 802.11 standard with a limited bandwidth of 20 MHz and 40 MHz. This prevents a reliable propagation delay estimation from the signal echoes bouncing on people in the scene due to the poor sampling time, thus hindering the use of range information for the counting. It would be useful to have a simulation framework to pre-train the counting algorithms on simulation data and to use the pre-trained models on a limited number of measurements, hence reducing drastically the effort put into measurements. This would also allow more flexibility to investigate crowd counting at other carrier frequencies and with the larger bandwidths of the latest 802.11 versions.

A step in this direction has already be taken in [5], where a multistatic radar environment with uniformly distributed targets is simulated and used to generate channel impulse responses on which counting algorithms can be trained. However, it considers a limited bandwidth of 20 MHz, and uses a geometry-based single-bounce channel model with simple path loss, hence considering point targets and neglecting electromagnetic interactions between targets. A more detailed model could thus be considered, integrating the scattering of an incident electromagnetic field on real bodies and not only on point targets, to take into account the interactions between targets. For scattering on multiple elements, analytical solutions exist but for trees modelling to assess their impact on communications. Trees are modelled as cylinders, and the scattering of waves among them is derived using the Foldy-Lax's multiple scattering theory [6], [7]. Numerical methods embedding human body electromagnetic properties exist to study the exposure on a single individual [8], but have not been scaled to groups of individuals.

Our previous work extended this existing knowledge to a simulation framework handling the scattering among multiple humans with numerical electromagnetics methods [9]. Taking into account the complex relative permittivity of the human body, we modelled the people as dielectric cylindrical shells whose thickness is equal to the electromagnetic skin depth, and resorted to the two-dimensional Method of Moments (MoM)

to compute the scattering of an incident field among them.

Leveraging our previous work, our contribution is threefold:

- We propose a crowd simulation framework based on the two-dimensional MoM applicable to Wi-Fi-based sensing purposes, and we use it to generate range and Doppler profiles of the simulated crowds.
- We exploit those simulated profiles to train a classifier to perform large scale crowd counting, where each class corresponds to an interval of number of people. Hence, we perform a *MoM-simulation-only training*. We carry out a performance assessment of the classifier on new profiles from a large-scale simulated crowd of 100 people.
- We apply the trained classifier on challenging *experimental measurements* of a real crowd with up to 40 people. We show that a classifier trained by MoM simulation data only has the potential to count people in a real crowd.

This article is structured as follows: in Section II, explanations are provided on the body model and the electric field computation with the MoM leading to a channel transfer function, range profile and Doppler profile. In Section III, the counting scheme and the features extracted from the profiles are explained. Finally, Section IV presents the simulation and measurement results obtained with the proposed counting framework, followed by a conclusion in Section V.

II. Method of Moments

This section explains why the MoM was chosen, and how it is used to derive the channel transfer function corresponding to an environment with several human bodies illuminated by a sensing-enabled Wi-Fi access point (AP).

A. Motivation

We choose to use the MoM over the more common ray tracing (RT) due to the fact that RT is only valid in the far field [10] [11], while in a crowd situation the bodies are not in the far field of each other. Indeed, the frontier between near and far field is conventionally defined at a distance $d_{\text{ff}} = 2L^2/\lambda$ from an emitter [11], where λ is the wavelength and L is the maximal characteristic dimension of the emitter. Here we consider a carrier frequency $f_c = 2.45$ GHz, hence a wavelength $\lambda = 0.12$ m. When an emitting antenna of a Wi-Fi AP is considered, L is small, for example $L = \lambda/2$. In this case, $d_{\text{ff}} = 0.06$ m. Targets will thus almost always be in the far field of the emitting antenna. However, when the electric field from the Wi-Fi AP antenna is incident on a body (modelled here as a cylindrical shell), the body emits a scattered field in reaction. At that moment, the body becomes the emitter, with it's diameter as characteristic dimension, here approximated as $L = 0.4$ m, much more than the characteristic dimension of an antenna. In this case, $d_{\text{ff}} = 2.61$ m. The scattered field from the body is incident on the other bodies, which in a crowd situation can be located at much smaller distances than 2.61 m from the first body. Consequently, the bodies in the crowd are not in the far field region of each other, hence diffraction coefficients computed with the approximations of RT for a single cylinder illuminated by a plane wave in far field would not be valid since the basis hypothesis of RT are violated. Furthermore, at higher frequencies, the condition becomes even more restrictive: at $f_c = 5$ GHz, $d_{\text{ff}} = 5.33$ m. To handle this complex scenario, a full-wave technique such as the MoM is required to avoid invalid approximations, since it is valid for any geometry anywhere in space.

B. Human body model

The human body is modelled as a mixture of bones, muscles, fat, organs, blood, skin and body fluids whose respective relative permittivity and conductivity are provided by [12]. The resulting total relative permeability ε_r and conductivity σ are computed at 2.45 GHz: $\epsilon_r = 33.37$ and $\sigma = 1.24$ S/m [9].

Geometrically, the human body can be modelled as an infinitely high cylindrical shell along the z dimension, following the infinite cylinder approximation [13], [14], with a radius $r = 0.2$ m. The cross section of each cylinder in the xy plane is discretized into circular cells whose radius is $\lambda_{in}/(5\sqrt{\pi})$, where λ_{in} is the wavelength inside the dielectric [15]. More details on this model are available in [9].

C. Electric field and transfer function

We consider an arbitrary number P of bodies, *i.e.* people, in the observation scene. The set of discretized cells at positions (x, y) in space belonging to all the different bodies is denoted $\mathcal{B}$. The antenna of the sensing device TX emits a vertically polarized frequency-dependent electric field at any point (x, y) in space. It is expressed as

$$E^i_{(x,y)}(f) = \frac{E_0}{\sqrt{d_{\text{TX}}/d_0}} e^{-j2\pi f d_{\text{TX}}/c}, \tag{1}$$

where f is the frequency, $d_{\text{TX}} = \sqrt{(x - x_{\text{TX}})^2 + (y - y_{\text{TX}})^2}$ is the distance between a point (x, y) and the coordinates $(x_{\text{TX}}, y_{\text{TX}})$ of TX, E_0 is a reference field, d_0 is a reference distance 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 discretized cells in $\mathcal{B}$, *i.e.* inside the P bodies, are denoted $\boldsymbol{E^i}$ and $\boldsymbol{E}$, respectively. From Maxwell's equations, the MoM framework is used to derive coefficients linking these incident and total fields in all bodies, in a matrix $\boldsymbol{C}$, yielding the following equation [15]:

$$\boldsymbol{C}\boldsymbol{E} = \boldsymbol{E^i}. \tag{2}$$

Diagonal blocks of $\boldsymbol{C}$ model the electromagnetic interactions between cells of one body, while non-diagonal blocks model the interactions between cells from different bodies. In other words, (2) models the interactions between all bodies illuminated by the incident field from TX. It has to be solved for $\boldsymbol{E}$, yielding the total field in all bodies cells. 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 $(x, y) \notin \mathcal{B}$.

The channel transfer function (CTF) can be computed from the frequency-dependent scattered field $E^s_{(x,y)}(f)$ evaluated on a bandwidth B around the carrier frequency f_c, by running

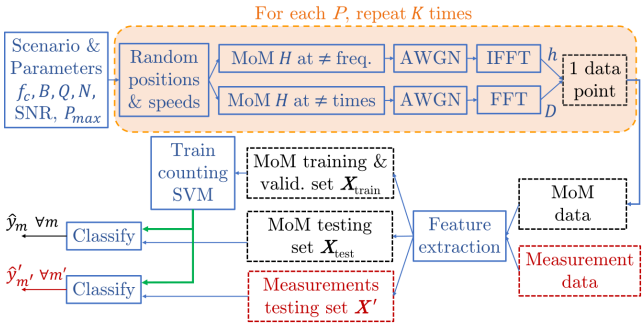

Fig. 1. Simulation data generation, measurement data, and classification scheme. Elements in red correspond to experimental measurement data.

the MoM on Q discrete frequencies $f_q = f_c + q\frac{B}{Q}$, with $q \in \mathbb{Z} \cap [-\frac{Q}{2}, \frac{Q}{2} - 1]$. From the scattered field at RX E^s_{RX}, and assuming that the incident field is perfectly cancelled at RX, we compute a normalized CTF as [9]

$$H(f_q) = \frac{E^s_{\text{RX}}(f_q)}{E_0}. \tag{3}$$

The field E^s_{RX} contains the electromagnetic interactions between all bodies, hence so does $H(f)$. Additive White Gaussian Noise (AWGN) is also added to $H(f)$ for a chosen value of SNR. The channel impulse response $h(\tau)$, also called range profile, is obtained through an Inverse Fast Fourier Transform (I-FFT) on $H(f)$. It is dependent on the propagation delay τ of the echoes created by the scattering on the bodies and among them. The delay τ corresponds to a range $d = c\tau/2$.

Similarly, a Doppler profile $D(f_D)$, function of the Doppler frequency f_D of the bodies echoes, is computed by evaluating $H(f_c)$ on the carrier frequency f_c at N subsequent slow-time instants spaced by a time T, and applying an FFT [9].

D. Computational efficiency

The main drawback of the MoM wrt. RT is its computation time, due to the resolution of the large linear system in (2). To tackle this, the MoM is accelerated by exploiting the block structure of its interaction matrix $\boldsymbol{C}$. We resort to a generalized minimal residual iterative solver [16] with a block-Jacobi preconditioner [17] and a block-by-block matrix-vector product. The low-rank structure of off-diagonal blocks is exploited to compress them using Adaptive Cross Approximation [18].

III. Counting: Data, Features and Algorithm

This section explains how our dataset for people counting is built, how the real crowd measurements are performed and how the counting is achieved. We generate range and Doppler profiles as explained in Section II, for different numbers P of people or bodies in K different random positions configurations. From these profiles, we compute representative features. Those features are fed as inputs to a Support Vector Machine (SVM) classifier [19]. The SVM outputs classes labels corresponding to ranges of values of people. The SVM trained on MoM data is then applied on measurements. The whole process is summarized in Fig. 1 and explained herebelow.

A. MoM range and doppler profiles generation

The first step of the process is to generate range and Doppler profiles from the MoM as follows:

1) Definition of the parameters: carrier frequency f_c, bandwidth B, number of frequencies Q, number of slow-time instants N, slow-time interval T, maximal number of people P_{max}, number of realizations K.
2) For each $P \in \{0, 1, ..., P_{max}\}$, repetition K times of:
   - Initialization of the positions of the people in the crowd, *i.e.* the centers of the cylindrical shells modelling them. Those positions are generated following a random binomial point process in the defined area.
   - Initialization of all discretized body cells around those centers, *i.e.* the set of $(x, y) \in \mathcal{B}$.
   - Computation of the range and Doppler profiles h and D with the MoM as described in Section II-C.
   - Saving each (h,D) pair as one "data point". For each P there are thus K realizations.

B. Processing of the experimental measurements

Apart from the h and D from the MoM, measurements are performed to obtain real-life range and Doppler profiles with Orthogonal Frequency-Division Multiplexing (OFDM) Wi-Fi signals. A continuous stream of 802.11ax-compliant High Efficiency-Long Training Fields (HE-LTF) OFDM symbols is sent from TX, sampled at a frequency $F_s = B$. The OFDM signals are received at RX after scattering on people. FFT-based range and Doppler processing are applied on them to obtain estimates of the range and Doppler profiles, $\hat{h}$ and $\hat{D}$, along with averaging to improve the SNR [20].

C. Features

Once the set of range and Doppler profiles are obtained from simulations and measurements, representative features are extracted from them. The approach for the features selection is the following. With the 802.11ac and 11ax Wi-Fi standards, higher bandwidth values are reachable, enabling a finer range resolution than with the previous versions of the standard, allowing to have more meaningful range profiles. For example by considering $B = 100$ MHz (cf. Section IV), the range resolution is 1.5 m. However, in a large crowd, people are spaced by a distance that is still lower than this range resolution. Similarly, the people have speed differences that can be inferior to the speed resolution, which is equal here to 0.15 m/s for a dwell time of 0.417 s. This means that a lot of of targets cannot be resolved by the radar. Moreover, some people might be partially or totally masked by others [9]. Those reasons imply that we cannot rely only on intuitive features like counting amplitude peaks in the range and Doppler profiles after a thresholding. Therefore our approach is to select a wide variety of features to capture all possible variations of the crowd configuration and movement. The features are listed in Table I [21], [22]. They are all computed on h, D, $\mathcal{F}(h)$ and $\mathcal{F}(D)$ for the simulations and $\hat{h}$, $\hat{D}$, $\mathcal{F}(\hat{h})$ and $\mathcal{F}(\hat{D})$ for the measurements, where $\mathcal{F}$ denotes the FFT operation. Exceptions are the delay and Doppler

TABLE I
FEATURES FOR CROWD COUNTING

Features type	Features names
Channel features	Delay spread, Doppler spread
Statistical features	Mean, std, moments 3&4, median, kurtosis, entropy, spectral entropy
Impulsive metrics	CFAR peaks, min, max, max-min, impulse factor, crest factor, clearance factor

spread that are only computed on h,$\hat{h}$ and D,$\hat{D}$ respectively. Subsequently, the dimensionality of the dataset can be reduced with a 95%-variance Principal Component Analysis (PCA) [23]. The output of this feature extraction forms two datasets:

- The MoM dataset $\boldsymbol{X}$ for all P people and all K realizations, where each row m is the features vector corresponding to one data point (*i.e.* one h and one D), with $m = 1, ..., (P_{max}+1)K$. It is split into the training & validation set, and the testing set. The training & validation set will be used to train the SVM classifier and to make a first evaluation of its performance by cross-validation [23], [24]. The performance of the SVM on new data will be evaluated by using the testing set.
- The measurement testing dataset $\boldsymbol{X}'$, where each row m' is the features vector corresponding to one measurement (*i.e.* one $\hat{h}$ and one $\hat{D}$), with $m' = 1, ..., M'$ and M' the number of measurements performed. This measurement set will be directly classified by the SVM classifier trained on the MoM training set, which outputs the crowd counting labels, *i.e.* the class labels for the measurements.

D. Class labels and Classifier

To each data point m and m', *i.e.* each row $\boldsymbol{X}[m,:]$ and $\boldsymbol{X}[m',:]$, corresponds a class label $y_m \in \{1, ..., N_c\}$ or $y'_m \in \{1, ..., N_c\}$, respectively, where N_c is the number of crowd counting classes. Each class corresponds to a range of values of the number of people in the scene. It corresponds to $\frac{P_{max}}{N_c}$ possible numbers of people: $y_m = n_c$ corresponds to a range of number of people in $\{(n_c-1)\frac{P_{max}}{N_c}, ..., n_c\frac{P_{max}}{N_c}-1\}$. For example, for $P_{max} = 100$ people and $N_c = 2$, $y_m = 0$ corresponds to a number of people $P \in \{0, ..., 49\}$, and $y_m = 1$ to $P \in \{50, ..., 100\}$.

The classifier used in this work is an SVM with a quadratic kernel and a box constraint of 0.19 [19], [23]. We chose the SVM for its flexibility both in handling potentially high dimension features vectors and in the choice of kernel functions to find non-linear decision boundaries between classes. It is extended to multiclass classification by resorting to the Error-Correcting Output Code method [23]. The SVM is thus trained on the training part of the simulation set $\boldsymbol{X}$, first tested on the testing part of $\boldsymbol{X}$, and then tested on the measurement set $\boldsymbol{X}'$.

This scheme allows to train the SVM on generic simulation data and to directly apply the obtained SVM model on new measurement data. The training phase is thus much simpler and cheaper than in the common approach consisting in training on measurement data, because it does not require to go on site to perfom training measurements.

TABLE II
SIMULATION PARAMETERS

f_c	B	N	Q	T	P_{max}	K	SNR
2.45 GHz	100 MHz	128	32	3.26 ms	100	242	20 dB

Fig. 2. Crowd counting scenario (left) and measurements scene (right), with TX and RX seen from behind.

IV. RESULTS

A. Scenario & parameters

The simulated crowd scenario with the MoM is illustrated on the left of Fig. 2. A crowd of up to $P_{max} = 100$ people is located in a rectangular 11.25 x 15 m area. The sensing-enabled WiFi AP acting as monostatic TX/RX is located at 3.75 m on the right of that area. The simulation parameters are summarized in Table II. Although not standard compliant, a bandwidth $B = 100$ MHz was chosen as a middleground between the different high bandwidth values of the 802.11ax standard to leverage the improved bandwidth that it allows.

B. Counting accuracy on simulation data

We compute the overal cross-validation and testing accuracies of the classifier for a varying number of classes N_c, and compare them to the accuracy $1/N_c$ of a random guess classifier. We analyze the confusion matrices corresponding to different values of N_c. The global cross-validation and testing accuracy curves are displayed in Fig. 3 with and without PCA.

In Fig. 3, the accuracy for $N_c = 2$ is 78.43% in the case without PCA. It decreases smoothly when N_c increases, as there are more classes to discriminate, hence more possible errors. At $N_c = 10$, it reaches 31.5%, which is still acceptable w.r.t. the random guess classifier having an accuracy of 10%. We also notice that applying PCA does not change the accuracy significantly: the difference in accuracy between the case with and without PCA is less than 3%. Finally, it is observed that the validation and testing accuracies are very close in both cases, meaning that the SVM can classify new data correctly, hence it is not overfitting the training data.

The confusion matrices for $N_c = 2, 3, 5$, and 10, without PCA, are displayed in Fig. 4. We see that with $N_c = 2$ the percentage of correct classification for class 1 and 2 are 70.9% and 85.8% respectively, *i.e.* that the classifier tends to classify slightly more towards class 2, hence towards high number of people. However this is not a strong trend, since for $N_c = 3$ the classification accuracy is higher for class 1 than for the other classes. For $N_c = 3$, class 2 is a bit more challenging since its feature space is close to the one of class 1 and class 3 simultaneously. Nevertheless, the classifier still reaches an acceptable accuracy w.r.t. the random guess classifier. We also

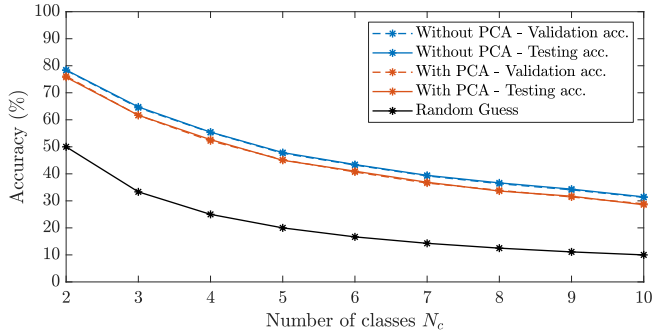

Fig. 3. Simulation data only - Accuracy in function of the number of classes

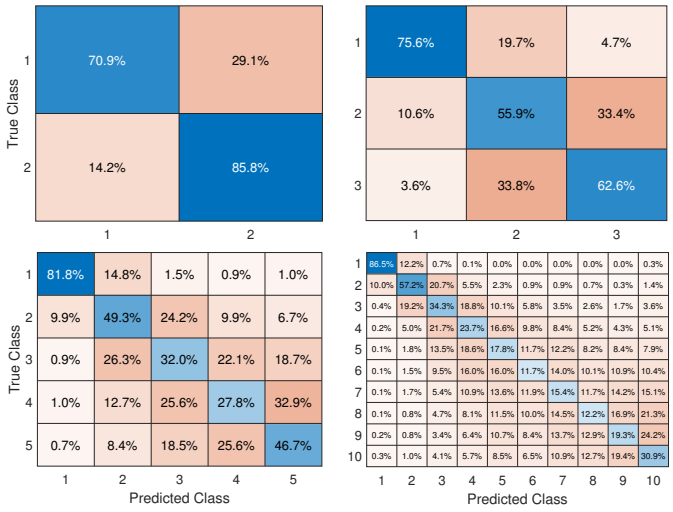

Fig. 4. Simulation data only, confusion matrices: $N_c = 2$ (top left) $N_c = 3$ (top right), $N_c = 5$ (bottom left), $N_c = 10$ (bottom right)

see that there are almost no misclassifications between the opposite classes 1 and 3, meaning that the classifier manages to separate the opposite cases well and makes almost only mistakes between neighbouring classes.

For high values $N_c = 5$ and $N_c = 10$, the accuracy for the classes with low number of people, *i.e.* the first classes, remains very satisfying: the accuracy for class 1 for $N_c = 5$ and $N_c = 10$ is 81.8% and 86.5% respectively, and the accuracy of class 2 is 49.3% and 57.2% respectively. Compared to the first classes, the middle classes yield a lower precision, with misclassifications among them. It shows that the classes with higher number of people are more difficult to discriminate. This is because when a lot of people are present in an area, their contribution to the range and Doppler profiles become more difficult to resolve even with generic features, and the masking between people is important, preventing the radar to see the complete scene. Nevertheless, we see that most misclassifications happen between the middle classes, *i.e.* that the proportion of misclassifications between distant classes is limited. Finally, the last class, corresponding to the highest number of people, exhibits an increased accuracy compared to the middle classes, due to the fact that it has less neighbouring classes in the features space.

C. Results from the experimental measurements

The results obtained with the measurements on which we perform the classification based on the SVM model trained with the MoM simulation data are described here. Our measurement setup consists of two USRPs X310, connected to 2.4-2.5 GHz directional panel antennas L-com RE09P via Sucoflex 126E cables with SMA connectors. 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. 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 same frequency and bandwidth as the simulations, $f_c = 2.45$ GHz and $B = 100$ MHz. The dwell time of each measurement was 0.417 s.

The measurements took place outdoors at a gathering. The geometrical configuration of the event (crowd area and TX/RX position w.r.t. the crowd) was similar to the simulation configuration, although static elements such as trees and tables were not modelled. People joined the gathering progressively, allowing to take measurements with numbers of people P between 2 and $P_{max} = 40$ people. The ground truth for the number of people was obtained by manual counting. A picture of the scene is displayed on the right of Fig. 2.

A lot of measurements were taken when the number of people was lower than 40 people, hence the measurements dataset has more realizations in the low classes, meaning that a global accuracy analysis with high numbers of classes as done in Section IV-B is not relevant. Instead, we focus on a more realistic case in which we consider that we might want to know when the number of people crosses a given threshold, here $P_{max}/2$. It corresponds to a classification with $N_c = 2$, in which class 1 corresponds to $\{0, ..., 19\}$ people and class 2 to $\{20, ..., 40\}$. We also consider a similar case but with an intermediate class, *i.e.* a situation in which we want to know if the number of people is low, average or high. It corresponds to a classification with $N_c = 3$, in which class 1 corresponds to $\{0, ..., 13\}$ people, class 2 to $\{14, ..., 26\}$ people and class 3 to $\{27, 40\}$ people. The SVM classifier was thus trained on the MoM dataset for $P_{max} = 40$ for both cases and applied on the measurements. It was noticed that the dimensionality reduction offered by PCA was beneficial here to increase the accuracy slightly, *i.e.* that simpler and smaller feature vectors were more suited to train separation boundary properly generalizable to measurements than large features vectors containing all features in Table I. Similarly, using a low SNR value (10 dB) on the MoM training data helped slightly increasing the accuracy on measurements, due to the fact that a low SNR value allowed to train more loose separation boundaries. The confusion matrices for the two studied cases ($N_c = 2$ and $N_c = 3$), with PCA and training SNR of 10 dB, are displayed in Fig. 5.

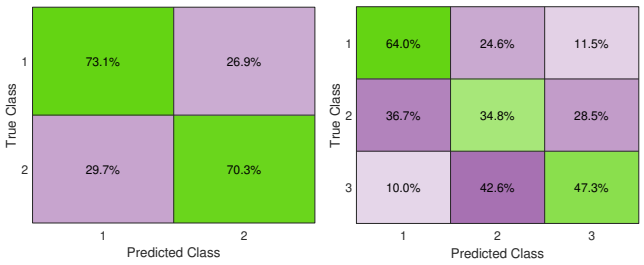

Fig. 5. Measurement data, confusion matrices: $N_c = 2$ (left), $N_c = 3$ (right). A modified color scheme is used to emphasize the fact that these confusion matrices are computed on experimental measurement data.

For $N_c = 2$, the correct classification percentage is above 70% for both classes. It means that the classifier can detect when people cross a certain threshold in the scene with a decent accuracy, which makes it a good application for crowd monitoring where preventive actions can be taken when the number of people exceeds a certain value. This accuracy is a promising result considering that the classifier is trained on simulation data only and that counting people at this large scale with radar data is a challenging task. Nevertheless, we see that there is still room for improvement in the accuracy, which can be investigated in future work by enlarging the training dataset, mixing simulation and measurement data for training, refining the features, or computing the features on other functions than the range and Doppler profiles.

For $N_c = 3$, the correct classification percentage has to be compared with the random guess classifier which would yield 33.33% on average. As expected from the simulation analysis in Section IV-B, the intermediate class (class 2) is the most challenging for the classifier. For this class, the correct classification percentage is close to the random guess, with misclassifications to neighbouring classes happening almost equally. We see that the classifier performs very well for class 1: it reaches a high accuracy compared to the random guess, and the majority of its errors are misclassifications to the neighbouring class and not to the opposite class (class 3). For class 3, the classifier achieves a decent precision compared to a random guess classifier, but still struggles to avoid misclassifications to class 2. However, it manages well to avoid misclassifications to the opposite class (class 1).

V. Conclusions

We presented a crowd counting framework for Wi-Fi sensing revolving around the MoM for electromagnetic scattering. We showed that we can use simulation data from the MoM accounting for electromagnetic interactions between bodies to train a counting classifier that can subsequently be applied on real crowd measurements. Promising counting accuracies were obtained, both on simulations and measurements testing sets.

References

[1] G. K. Still, *Introduction to crowd science*. CRC Press, 2014.

[2] IEEE 802.11, "Wi-Fi sensing," 2019, Last consult.: Feb. 2022. [Online]. Available: https://mentor.ieee.org/802.11/documents?is_dcn=DCN%2C%20Title%2C%20Author%20or%20Affiliation&is_group=SENS

[3] 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.

[4] 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.

[5] 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.

[6] 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.

[7] 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

[8] F. J. Meyer, D. B. Davidson, U. Jakobus, and M. A. Stuchly, "Human exposure assessment in the near field of gsm base-station antennas using a hybrid finite element/method of moments technique," *IEEE Trans. Biomed. Eng.*, vol. 50, no. 2, pp. 224–233, Feb. 2003.

[9] 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 International Symposium on Joint Communications & Sensing (JC&S)*. IEEE, 2022, pp. 1–6.

[10] 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.

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

[12] D. Andreuccetti, "An Internet resource for the calculation of the dielectric properties of body tissues in the frequency range 10 Hz-100 GHz," *http://niremf.ifac.cnr.it/tissprop/*, 2012.

[13] 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.

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

[15] 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.

[16] 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.

[17] 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.

[18] 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.

[19] C. Cortes and V. Vapnik, "Support-vector networks," *Machine learning*, vol. 20, no. 3, pp. 273–297, 1995.

[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 802.11 ax wi-fi-based multi-antenna passive radar," *IEEE Sensors Journal*, vol. 21, no. 18, pp. 20 462–20 474, Sept. 2021.

[21] S. Di Domenico, G. Pecoraro, E. Cianca, and M. De Sanctis, "Trained-once device-free crowd counting and occupancy estimation using wifi: A doppler spectrum based approach," in *2016 IEEE 12th Int. Conf. Wireless Mobile Comput., Netw. Commun.* IEEE, 2016, pp. 1–8.

[22] H. Zou, Y. Zhou, J. Yang, W. Gu, L. Xie, and C. Spanos, "Freecount: Device-free crowd counting with commodity wifi," in *2017 IEEE Global Communications Conference*. IEEE, 2017, pp. 1–6.

[23] 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.

[24] Mathworks, "Train support vector machine (SVM) classifier for one-class and binary classification." Last modification: 2021. [Online]. Available: https://nl.mathworks.com/help/stats/fitcsvm.html