

Crowd-Centric Counting via Unsupervised Learning

Flavio Morselli*, Stefania Bartoletti*, Santiago Mazuelas†, Moe Win‡, and Andrea Conti*

*DE, CNIT at University of Ferrara, E-mail: {flavio.morselli, stefania.bartoletti, andrea.conti}@unife.it

†Basque Center for Applied Mathematics, E-mail: smazuelas@bcamath.org

‡LIDS at Massachusetts Institute of Technology, E-mail: moewin@mit.edu

Abstract—Counting targets (people or things) within a monitored area is an important task in emerging wireless applications, including those for smart environments, safety, and security. Conventional device-free radio-based systems for counting targets rely on localization and data association (i.e., individual-centric information) to infer the number of targets present in an area (i.e., crowd-centric information). However, many applications (e.g., affluence analytics) require only crowd-centric rather than individual-centric information. Moreover, individual-centric approaches may be inadequate due to the complexity of data association. This paper proposes a new technique for crowd-centric counting of device-free targets based on unsupervised learning, where the number of targets is inferred directly from a low-dimensional representation of the received waveforms. The proposed technique is validated via experimentation using an ultra-wideband sensor radar in an indoor environment.

Index terms— Crowd-centric counting, unsupervised learning, dimensionality reduction, ultra-wideband, sensor radar.

I. INTRODUCTION

Counting people and things is important for crowd sensing and behavior analysis applications, including those related to the Internet-of-Things [1], smart environments [2], social networking [3], and surveillance [4]. For counting tasks, radio-based systems are preferred to image-based systems [5]–[7], especially when privacy, implementation costs, and obstructed line-of-sight represent important limitations. Among radio-based systems, device-free systems are often preferred to systems that rely on dedicated or personal devices [8]–[12].

Device-free systems are based on networks of sensor radars (SRs) that sense the wireless environment and detect targets from signal reflections (backscattering) [13]–[18]. The presence of obstacles and other scatterers (e.g., furniture, walls, and windows) leads to clutter and multipath propagation, which have detrimental effects on the detection performance. These phenomena are particularly severe in indoor environments, where the number of scatterers is large [19]–[21].

Conventional approaches for device-free counting via SRs rely on multi-target localization or tracking [22]–[24], where each SR estimates a set of metrics (e.g., ranges or angles) associated to a single detected target (namely, individual-centric approach). Typically, this approach has a complexity that grows exponentially with the number of targets due

to data association. Individual-centric information and data association are unnecessary when the system is only interested in crowd-centric information (i.e., the number of targets). Therefore, there is a growing interest in conceiving methods that infer the number of targets without relying on localization (namely, crowd-centric approach). Existing models for relating the received waveforms to the number of targets depend on the specific environment and multipath conditions [25]–[27].

In [28], the number of targets is inferred directly from energy samples of the received waveform and relying on a statistical characterization of the wireless channel. However such statistical characterization may be unknown a priori. As a possible solution, learning techniques can be employed to directly estimate the generative model of the measurements via a training phase. Recently, unsupervised learning has been successfully applied to ranging for network localization [18], [29]. In soft range information, nodes' position is determined based on the information obtained through unsupervised machine learning techniques in the form of likelihood functions instead of single-value metrics (e.g., time-of-arrival, angle-of-arrival, and received signal strength). The use of unsupervised learning for device-free counting allows the system to skip the estimation of single-value metrics and exploit the whole received waveform to extract information about the number of targets. Moreover, the use of likelihood functions enables an efficient fusion of the information provided by each SR.

This paper proposes a new technique for crowd-centric counting of device-free targets based on unsupervised learning. We first define a low-dimensional representation of the received waveforms, based on descriptive features and principal component analysis (PCA). Then during the training phase, a joint probability density function (PDF) (i.e., a generative model) for such representation is learned and then used in the online phase to infer the number of targets. Since the number of targets is discrete, the learning process requires nonparametric estimation with both continuous and discrete data [30], [31]. Experimental results are collected in an indoor environment with an ultra-wideband (UWB) SR to validate the proposed technique in different scenarios and quantify its performance, in terms of counting error outage (CEO) and counting root-mean-square error (RMSE), in comparison with two multi-target detection algorithms. The key contributions of this paper can be summarized as follows:

- development of a crowd-centric counting method based on the energy samples of the received waveforms and

This research was supported, in part, by the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie Grant 703893, the Office of Naval Research under Grants N62909-18-1-2017 and N00014-16-1-2141, and the Spanish Ministry of Economy and Competitiveness under Ramon y Cajal Grant RYC-2016-19383.

unsupervised learning techniques;

- validation of the proposed technique via experimentation with an UWB SR in an indoor environment; and
- comparison of its performance in terms of CEO and counting RMSE with two multi-target detection algorithms.

Notations: A random variable (RV) and its realization are denoted by \mathbf{x} and x ; a random vector and its realization are denoted by \mathbf{x} and \mathbf{x} . The function $f_{\mathbf{x}}(\mathbf{x})$ and, for brevity when possible, $f(\mathbf{x})$ denote the PDF of a continuous RV \mathbf{x} ; $f_{\mathbf{x}|\mathbf{y}}(\mathbf{x}|\mathbf{y})$. A random set and its realization are denoted by \mathcal{X} and \mathcal{X} , respectively. The \emptyset denotes the empty set. For a matrix \mathbf{A} and a vector \mathbf{a} the transpose is denoted by \mathbf{A}^T and \mathbf{a}^T , respectively. The floor operator is denoted by $\lfloor \cdot \rfloor$.

II. SYSTEM MODEL

Consider a network composed of n_r monostatic SRs in known positions, with index set $\mathcal{N}_r = \{1, 2, \dots, n_r\}$. Each SR transmits an impulse signal $s(t)$ at each time instant kT_{pr} , where $k \in \mathbb{Z}$ and T_{pr} is the pulse repetition period. The signal propagates in the wireless environment where a random number n_t of targets indexed by the random set \mathcal{N}_t are present in unknown positions. The signal is backscattered from the targets as well as from any object in the monitored area, leading to multiple echoes. The equivalent low-pass signal received at the h -th SR can be expressed as

$$\mathbf{r}^{(h)}(t) = \mathbf{r}_t^{(h)}(t) + \mathbf{r}_m^{(h)}(t) + \mathbf{r}_c^{(h)}(t) + \mathbf{n}^{(h)}(t) \quad (1)$$

where $\mathbf{r}_t^{(h)}(t)$ contains the echoes backscattered from the targets (single bounce from the target to the radar); $\mathbf{r}_m^{(h)}(t)$ contains the echoes due to multipath propagation after target backscattering (multiple bounces after the first one); $\mathbf{r}_c^{(h)}(t)$ contains the echoes due to clutter (e.g., furniture and undesired objects in the scene); and $\mathbf{n}^{(h)}(t)$ is a zero-mean additive white Gaussian noise.

In the k -th observation frame, i.e. for $t \in (kT_{pr}, kT_{pr} + T_{obs}]$, where T_{obs} is the observation interval with $T_{obs} \leq T_{pr}$, the component related only to the targets is

$$\mathbf{r}_t^{(h)}(t) = \sum_{n \in \mathcal{N}_t} \gamma_n^{(h)} s(t - \mathbf{t}_n^{(h)}) \quad (2)$$

where $\gamma_n^{(h)}$ is the complex gain including the radiation pattern of the h -th SR antennas, the path-loss, and the n -th target radar cross section; $\mathbf{t}_n^{(h)}$ is the two-way time of flight for the h -th radar- n -th target- h -th radar path. In (1), the signal component related to multipath propagation of the target echoes is

$$\mathbf{r}_m^{(h)}(t) = \sum_{\substack{n \in \mathcal{N}_t \\ m \in \mathcal{M}_n^{(h)}}} \gamma_n^{(h)} \alpha_{n,m}^{(h)} s(t - \mathbf{t}_n^{(h)} - \tau_{n,m}^{(h)}) \quad (3)$$

where $\alpha_{n,m}^{(h)}$ is the complex gain relative to the m -th path for the n -th target and $\tau_{n,m}^{(h)}$ is the excess delay, and $\mathcal{M}_n^{(h)}$ is the random index set of the multipath components. Note that $\mathbf{r}_t^{(h)}(t)$ and $\mathbf{r}_m^{(h)}(t)$ are due to target backscattering, whereas

$\mathbf{r}_c^{(h)}(t)$ in (1) depends only on the undesired objects in the operating environment (clutter).

The signal $\mathbf{r}^{(h)}(t)$ is processed frame-by-frame using a clutter removal filter. Consider the vector $\mathbf{r}_k^{(h)}$ of $n_s = \lfloor T_{obs}/T_s \rfloor$ received signal samples obtained within the k th frame, where T_s is the sampling time. The choice of the clutter removal filter depends on the clutter dynamics. We will consider a moving average (MA) filter, which is widely adopted when the clutter is static or varies slowly with the respect to the targets. The filter output for the k -th frame results in

$$\tilde{\mathbf{r}}_k^{(h)} = \mathbf{r}_k^{(h)} - \frac{1}{n_{MA}} \sum_{i=k-1-n_{MA}}^{k-1} \mathbf{r}_i^{(h)} \quad (4)$$

where n_{MA} represents the number of frames used to estimate the clutter components. The estimate is updated every frame to capture the slow variations of the clutter. Such a clutter removal filter prevents the counting of static targets. However, due the fine time resolution of UWB signals, a person standing in a static position is not a static target due to the respiration motions. This makes the MA filter well suited for applications involving the count of persons such as affluence analytics and flow monitoring.

A number N_{int} of frames can be non-coherently accumulated at the output of an energy detector (ED) to reduce the noise effect. First, a vector $\mathbf{e}_k^{(h)}$ of $n_{bin} = \lfloor T_{obs}/T_d \rfloor$ energy bins is obtained through a quadrature integration and dump (QID) block, where T_d represents the dwell time and the j -th bin is given by

$$\mathbf{e}_{k,j}^{(h)} = \sum_{i=jN_d}^{(j+1)N_d} |\tilde{r}_{k,i}^{(h)}|^2 \quad (5)$$

in which $N_d = \lfloor T_d/T_s \rfloor$ represents the number of samples in each energy bin, and $\tilde{r}_{k,i}^{(h)}$ is the i -th element of $\tilde{\mathbf{r}}_k^{(h)}$. Then, the energy bin vectors from different frames are averaged as

$$\bar{\mathbf{e}}_k^{(h)} = \frac{1}{N_{int}} \sum_{j=(k-1)N_{int}}^{kN_{int}} \mathbf{e}_j^{(h)}. \quad (6)$$

Note that a new energy vector $\bar{\mathbf{e}}_k^{(h)}$ is obtained every N_{int} observation frames, i.e. at time $kN_{int}T_{pr}$ with $k \in \mathbb{Z}$.

III. COUNTING VIA UNSUPERVISED LEARNING

Consider a fixed frame index k , i.e., $\mathbf{r}^{(h)}$ and $\bar{\mathbf{e}}^{(h)}$ are the vectors representing the received signal samples and the energy bins for the h -th SR, respectively. The counting system aims to infer the number of targets n_t from the observed data $\{\mathbf{y}^{(h)}\}_{h \in \mathcal{N}_r}$. The observed data may refer to the received signal samples $\mathbf{y}^{(h)} = \mathbf{r}^{(h)}$ and the estimate \hat{n}_t is updated frame by frame (i.e., at time kT_{pr} with $k \in \mathbb{Z}$), or can refer to the energy vectors $\mathbf{y}^{(h)} = \bar{\mathbf{e}}^{(h)}$ and the estimate \hat{n}_t is updated every N_{int} frames (i.e., at time $kN_{int}T_{pr}$ with $k \in \mathbb{Z}$).

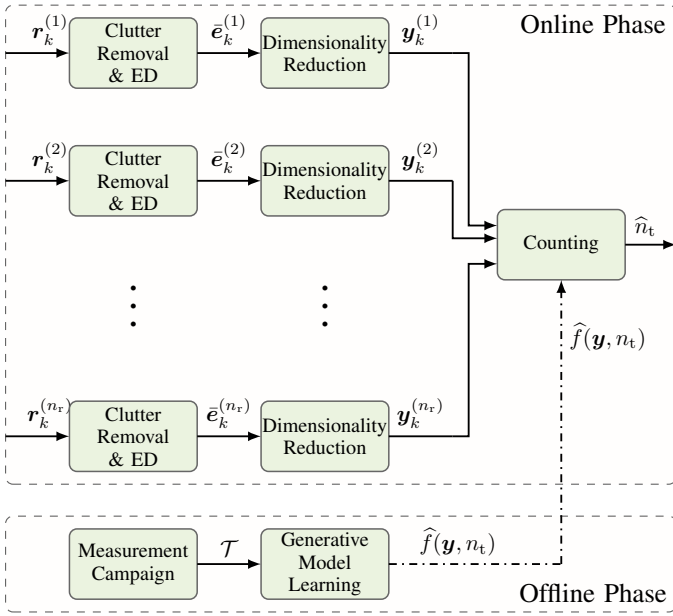


Fig. 1. Scheme of crowd-centric counting via unsupervised learning.

In crowd-centric approaches, which do not rely on target positional information, the maximum likelihood (ML) estimate of n_t is

$$\hat{n}_t = \arg \max_{n_t} \prod_{h \in \mathcal{N}_t} f_{\mathbf{y}|n_t}(\mathbf{y}^{(h)} | n_t) \quad (7)$$

where the measurement sets are considered independent and identically distributed among the different SRs for a given n_t . The random vector \mathbf{y} represents the measurement set associated to an unspecified SR.

We propose a technique for crowd-centric counting of device-free targets via unsupervised machine learning. First, a generative model (i.e., the joint PDF $f(\mathbf{y}, n_t)$) is learned during a training phase where a data set $\mathcal{T} = \{\mathbf{y}^{(l)}\}_{l \in \mathcal{N}_{tr}}$ of n_{tr} measurements indexed by \mathcal{N}_{tr} are collected. Then, $f(\mathbf{y}|n_t)$ can be obtained from $f(\mathbf{y}, n_t)$ with a priori knowledge $f(n_t)$. The dimensionality of the problem depends on the size of \mathbf{y} . In particular, the use of the received signal samples vector $\mathbf{y} = \mathbf{r}$ or of the energy vector $\mathbf{y} = \bar{\mathbf{e}}$ may lead to high computational complexity.¹ Therefore, dimensionality reduction is performed on $\bar{\mathbf{e}}$ by defining two descriptive features and PCA [32].

Consider $\mathbf{y} = [\max\{\bar{\mathbf{e}}\}, n_p, \psi(\bar{\mathbf{e}})]^T$, where $\max\{\bar{\mathbf{e}}\}$ is the global maximum for the vector $\bar{\mathbf{e}}$, n_p is the number of local maxima, and $\psi(\bar{\mathbf{e}})$ is a $1 \times d$ vector, containing the first d principal components of the $n_{bin} \times 1$ vector $\bar{\mathbf{e}}$, i.e.

$$\psi(\bar{\mathbf{e}}) = \frac{1}{\max\{\bar{\mathbf{e}}\}} \bar{\mathbf{e}}^T \mathbf{V} \quad (8)$$

where \mathbf{V} is a $n_{bin} \times d$ matrix with columns given by the eigenvectors associated to the d largest eigenvalues of Σ , which is the $n_{bin} \times n_{bin}$ sample covariance matrix of the vectors

¹Analogously to the measurement set \mathbf{y} , from now on the received signal vector \mathbf{r} and the energy vector $\bar{\mathbf{e}}$ are associated to an unspecified SR.

$\{\bar{\mathbf{e}}^{(l)} / \max\{\bar{\mathbf{e}}^{(l)}\}\}_{l \in \mathcal{N}_{tr}}$. The global maximum of the energy bin vector can be related to the presence of targets in the operating environment, while the number of peaks represents a first coarse approximation for the number of targets. Though the advantages of being easily obtainable, these two descriptive features do not encapsulate all the information carried by $\bar{\mathbf{e}}$. In contrast, PCA provides a concise representation of $\bar{\mathbf{e}}$ by projecting the data into a low-dimension linear subspace while ensuring the lowest mean-square error. The main drawback of the PCA approach is when highly nonlinear relation is present between the number of targets and $\bar{\mathbf{e}}$ as this dimensionality reduction technique fails to preserve the information carried by the measurements.

Learning the generative model for \mathbf{y} is a density estimation problem with both continuous and discrete RVs. The jittering technique is employed to transform the discrete RVs into continuous RVs so classical density estimation techniques (that are suitable for continuous RVs only) can be employed [30], [31]. Consider $\mathbf{x} = [\tilde{\mathbf{x}}, \check{\mathbf{x}}]^T$ where $\tilde{\mathbf{x}} = [\max\{\bar{\mathbf{e}}\}, \psi(\bar{\mathbf{e}})]$ represents the continuous RVs; and $\check{\mathbf{x}} = [n_p, n_t] + \mathbf{w}$ represents the discrete RVs after adding the jitter $\mathbf{w} \in \mathbb{R}^2$, i.e. a random vector representing a noise term. The new vector \mathbf{x} of continuous RVs is such that

$$f_{\mathbf{y}, n_t}(\mathbf{y}, n_t) = f_{\mathbf{x}}(\mathbf{x}) \quad (9)$$

holds $\forall \mathbf{w}$, $f_{\mathbf{w}}(\mathbf{0}) = 1$, $f_{\mathbf{w}}(\mathbf{w}) = 0 \forall \mathbf{w} \in \{\mathbb{Z} \setminus \{0\}\}^2$.

A linear transformation, namely data sphering, is then applied to the jittered training data $\mathbf{x}^{(l)}$ so that the covariance matrix becomes the identity matrix [33]

$$\mathbf{z}^{(l)} = \mathbf{A}^{-\frac{1}{2}} \mathbf{A}^T (\mathbf{x}^{(l)} - \bar{\mathbf{x}}) = [\tilde{\mathbf{z}}^{(l)}, \check{\mathbf{z}}^{(l)}]^T \quad (10)$$

where \mathbf{A} is the diagonal matrix formed by the eigenvalues of the sample covariance matrix relative to $\{\mathbf{x}^{(l)}\}_{l \in \mathcal{N}_{tr}}$, \mathbf{A} is the matrix formed by its corresponding eigenvectors, and $\bar{\mathbf{x}}$ is the sample mean vector from the training data.

The density estimation is carried out on the sphered training set $\{\mathbf{z}^{(l)}\}_{l \in \mathcal{N}_{tr}}$ and then the PDF in the original domain is obtained as $|\det(\mathbf{A}^{-\frac{1}{2}} \mathbf{A}^T)| f_{\mathbf{z}}(\mathbf{A}^{-\frac{1}{2}} \mathbf{A}^T (\mathbf{x} - \bar{\mathbf{x}}))$. In particular, we consider two density estimation techniques based on the Fisher-Wald (FW) and the kernel density estimator (KDE) [29]. In the FW setting, the density function $f_{\mathbf{z}}(\mathbf{z})$ is approximated with a mixture of m Gaussian distributions as

$$\hat{f}(\mathbf{z}) = \sum_{i=1}^m \beta_i \varphi(\mathbf{z}; \boldsymbol{\mu}_i; \boldsymbol{\Xi}_i) \quad (11)$$

where $\varphi(\cdot; \boldsymbol{\mu}_i; \boldsymbol{\Xi}_i)$ is the Gaussian PDF; the weights β_i , the mean vectors $\boldsymbol{\mu}_i$, and the covariance matrices $\boldsymbol{\Xi}_i$ are obtained through the minimization of the empirical risk in the FW setting [34]. The solution of the minimization problem is obtained via ML estimator or its approximation (e.g., the solution provided by the expectation maximization algorithm).

In the KDE, the density function $f_{\mathbf{z}}(\mathbf{z})$ is approximated as

$$\hat{f}(\mathbf{z}) = \frac{1}{n_{tr}} \sum_{l \in \mathcal{N}_{tr}} K(\tilde{\mathbf{z}} - \tilde{\mathbf{z}}^{(l)}; \tilde{\mathbf{H}}) K(\check{\mathbf{z}} - \check{\mathbf{z}}^{(l)}; \check{\mathbf{H}}) \quad (12)$$



Fig. 2. Measurement campaign environment.

where $K(\tilde{z}; \tilde{\mathbf{H}})$, $K(\check{z}; \check{\mathbf{H}})$ are product kernels, parametrized by the matrices $\tilde{\mathbf{H}}$, $\check{\mathbf{H}}$, called bandwidth matrices. We consider Gaussian kernels $K(\tilde{z} - \tilde{z}^{(l)}; \tilde{\mathbf{H}}) = \varphi(\tilde{z}; \tilde{z}^{(l)}; \tilde{\mathbf{H}})$, $K(\check{z} - \check{z}^{(l)}; \check{\mathbf{H}}) = \varphi(\check{z}; \check{z}^{(l)}; \check{\mathbf{H}})$. The bandwidth matrices are obtained through the normal reference rule [35].

IV. CASE STUDY

The proposed technique is validated via experimental results in terms of CEO and counting RMSE. The CEO is defined as the probability that the counting error $|\hat{n}_t - n_t|$ is above a given target value n^*

$$P_{\text{ceo}}(n^*) = \mathbb{P}\{|\hat{n}_t - n_t| > n^*\} \quad (13)$$

while the counting RMSE is defined as ²

$$e_{\text{rms}} = \sqrt{\mathbb{E}\{|\hat{n}_t - n_t|^2\}}. \quad (14)$$

The measurement campaign was carried out at the Department of Engineering, University of Ferrara (see Fig. 2). The size of the monitored area is 6×4 m² and the presence of a table, chairs, panels, walls, and other objects leads to a large number of potential scatterers (i.e., clutter and multipath). A network of $n_r = 3$ SRs in monostatic configuration is emulated recording the measurements with a single SR in different locations. The measurements are then properly combined in the online stage. The SR is based on Novelda's UWB SR "X4M03" chip [36]. Tab. I summarizes the SR settings and the parameters values for the proposed technique. The number of targets (persons) n_t in the monitored area varies between 0 and $n_{\text{max}} = 3$. For each value of $n_t > 0$, $n_c = 10$ possible spatial arrangements of the targets in the monitored area are considered (i.e., spatial configurations), with index set $\mathcal{N}_c = \{1, 2, \dots, n_c\}$. Each configuration $\mathcal{P}_{n_t, j}$, with $j \in \mathcal{N}_c$ corresponds to a different n_t -tuple of target positions, i.e. $\mathcal{P}_{n_t, j} = \{\mathbf{p}_{n, j}\}_{n \in \mathcal{N}_t}$, where $\mathcal{N}_t = \{1, 2, \dots, n_t\}$ is the index

²The CEO is approximated with the rate and the expectation in the RMSE is approximated with the sample mean due to the finite set of experimental data.

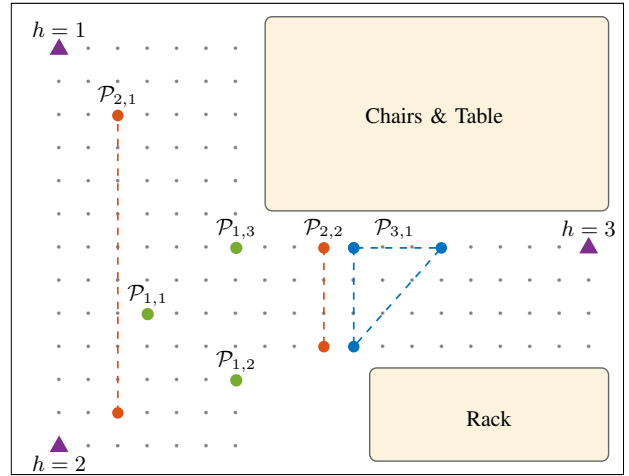


Fig. 3. Map of the hall with the configurations $\mathcal{P}_{1,1}, \mathcal{P}_{1,2}, \mathcal{P}_{1,3}$ for $n_t = 1$, $\mathcal{P}_{2,1}, \mathcal{P}_{2,2}$ for $n_t = 2$, and $\mathcal{P}_{3,1}$ for $n_t = 3$. The violet triangles represent the SRs positions. The grid spacing is 0.3 m.

TABLE I
SR SETTING AND SIGNAL PROCESSING PARAMETERS VALUES.

Parameter (Symbol)	Value
SR carrier frequency	7.26 GHz
SR pulse bandwidth	1.4 GHz
SR energy per pulse	2.6 pJ
Repetition interval (T_{pr})	40 ms
Sample time (T_s)	42.8 ps
Number of samples per frame (n_s)	1079
Observation time (T_{obs})	43.4 ns
Window length MA filter (n_{MA})	30
Dwell time (T_d)	2 ns
Number of samples per energy bin (N_d)	47
Length energy bin vector (n_{bin})	23
Number of energy bin vectors integrated (N_{int})	20
Number of principal components PCA (d)	9
Number of components in FW (m)	9

set relative to the targets, $\mathbf{p}_{n, j}$ is the position of the n -th target for the j -th configuration associated to n_t . For $n_t = 0$, only $n_c = 3$ configurations (i.e., measurements of the empty room at different time instants) are considered due to the absence of targets. Example configurations are depicted in Fig. 3. For each configuration $\mathcal{P}_{n_t, j}$, one thousand observation frames per SR are collected. After the clutter removal and energy detection, the number of energy vectors $\mathbf{e}_k^{(h)}$ are 50 for each radar and each configuration (i.e., $k \in \{1, 2, \dots, 50\}$).

The measurements are used as input of an iterative procedure to validate the proposed technique. At each run, the training (offline) and validation (online) phases are performed by picking a set of configurations \mathcal{T} for training, and a set of configurations \mathcal{V} for validating the generative model, following the leave- p -out cross-validation criteria [37].³ This procedure ensures that the validation of the proposed technique is carried

³In particular, $p = 4$ so the validation set \mathcal{V} has cardinality $|\mathcal{V}| = 4$ and contains one configuration for each possible value of n_t at each iteration. The training set \mathcal{T} , with $|\mathcal{T}| = 29$, contains all the other configurations (i.e., $\mathcal{T} \cap \mathcal{V} = \emptyset$ with 27 configurations relative to the case $n_t > 0$ and 2 relative to the case $n_t = 0$).

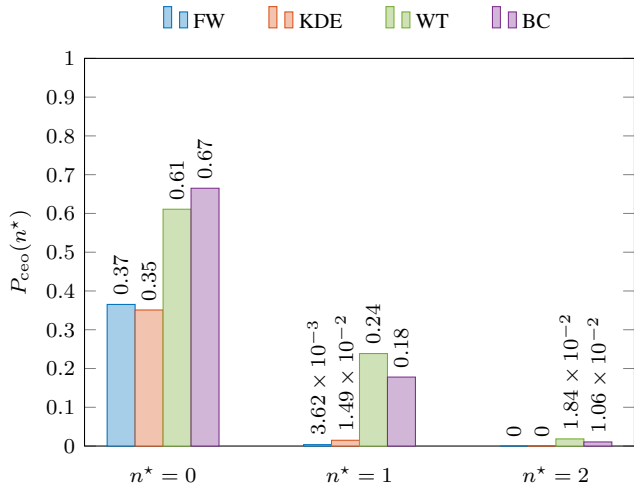


Fig. 4. CEO as a function of the target counting error n^* for FW, KDE, WT, and BC algorithms.

out on configurations relatively different from the ones used to learn the generative model.

Numerical results are provided for FW and KDE algorithms with ML estimation. The jitter \mathbf{w} is uniformly distributed over the interval $[-0.5, 0.5]^2$ in order to meet the constraints in (9). Results are compared with two multi-target detection algorithms, namely window threshold (WT) [38] and binary clustering (BC) [17], with CA-CFAR detection. The WT threshold value and the CA-CFAR window length are optimized such that the counting RMSE is minimized at each offline phase. The WT left and right guard lengths are 1 and 2 bins, respectively. The CA-CFAR guard length is 1 bin and the target false alarm probability is 0.01. The BC threshold distance is set to 0.45 m (approximately one and an half of the distance associated to an energy bin). In the multi-target detection cases, each SR provides an estimate of the number of targets and the final estimate \hat{n}_t is obtained as the mode of the estimates provided by each SR.

Fig. 4 shows the CEO as a function of the target counting error n^* . It can be observed that the proposed algorithms outperform the multi-target detection algorithms. The error is zero in the 63% of the cases for both KDE and FW (the CEO is below 37% for $n^* = 0$), whereas the error is zero only in the 39% of the cases for the WT (the CEO is below 61% for $n^* = 0$) and 33% for the BC (the CEO is below 67% for $n^* = 0$). This can be attributed to the fact that even though the clutter removal filter eliminates most of the contributes from static reflectors, the multipath generated after target backscattering remains unaltered at the filter output. This detrimental phenomena can be in part mitigated by the crowd-centric approach, while it severely affects the performance of multi-target detection algorithms. It can also be observed that the CEO is decreasing with n^* more rapidly for the proposed algorithms with respect to the multi-target detection ones. For example, the CEO goes from 0.35 with $n^* = 0$ to 0.003 with $n^* = 1$ for the FW, whereas it goes from 0.61 to 0.24 with

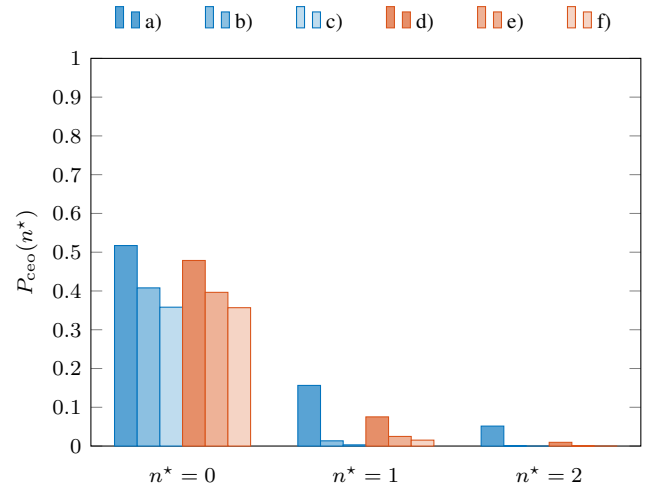


Fig. 5. CEO as a function of n^* for FW and KDE algorithms: a) FW with $|\mathcal{T}| = 4$; b) FW with $|\mathcal{T}| = 17$; c) FW with $|\mathcal{T}| = 29$; d) KDE with $|\mathcal{T}| = 4$; e) KDE with $|\mathcal{T}| = 17$; and f) KDE with $|\mathcal{T}| = 29$.

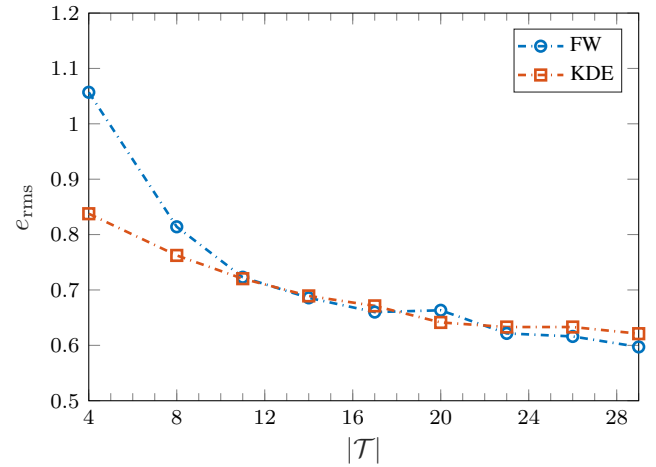


Fig. 6. Counting RMSE as a function of $|\mathcal{T}|$ for FW and KDE algorithms.

$n^* = 1$ for WT. Therefore, even in the presence of a counting error, its magnitude is smaller for the proposed algorithms.

Fig. 5 shows the CEO for FW and KDE as a function of the target counting error n^* for different sizes of the set \mathcal{T} . The values are derived by averaging the performance over 200 randomly chosen validation sets \mathcal{V} obtained from the iterative procedure. Given \mathcal{V} , the generative model is validated multiple times adding each time one configuration for each value of n_t to \mathcal{T} . It can be observed that the CEO decreases when the number of configurations available during the training phase increases. For $n^* = 0$, the CEO decreases from about 51% to 35% for the FW and decreases from 47% to 35% for the KDE when the cardinality increases from $|\mathcal{T}| = 4$ to $|\mathcal{T}| = 29$. This represents an increase of about 12% for the probability of estimate the correct number of targets. Moreover, the decrement can be appreciated for all the values of n^* . This can be attributed to the fact that we add to \mathcal{T} one configuration for each value of n_t in each iteration, so the

generative model is not biased toward any specific value of n_t .

Fig. 6 shows the counting RMSE as a function of the size of the training set \mathcal{T} for FW and KDE algorithms. The values are obtained with the same procedure explained for Fig. 5. It can be observed that the performance improves when the number of configurations used in the training phase increases. Incrementing the cardinality, decreases the e_{rms} from 1.05 to 0.59 for the FW and from 0.83 to 0.62 for the KDE (the counting RMSE with $|\mathcal{T}| = 29$ is $e_{\text{rms}} = 1.19$ for WT and $e_{\text{rms}} = 1.13$ for BC). This can be attributed to the fact that the generative model becomes more accurate when the amount of data available during the learning process increases.

V. FINAL REMARK

This paper has proposed a crowd-centric technique based on unsupervised learning for counting device-free targets. Based on a low-dimensional representation of the received waveforms, the number of targets present in a monitored area is inferred through a generative model learned during a training phase. This technique avoids data association complexity of the individual-centric methods and does not leverage on any specific assumption on the operating environment. The proposed technique has been validated via network experimentation in a real environment, where the crowd-centric approach based on unsupervised learning outperforms classical multi-target detection algorithms. The use of unsupervised learning is a good candidate solution for counting device-free targets for applications where crowd-centric information is required.

REFERENCES

- [1] S. D'Oro, L. Galluccio, G. Morabito, and S. Palazzo, "Exploiting object group localization in the internet of things: Performance analysis," *IEEE Trans. Veh. Technol.*, vol. 64, no. 8, pp. 3645–3656, Aug 2015.
- [2] G. Cardone, L. Foschini, P. Bellavista, A. Corradi, C. Borcea, M. Talasila, and R. Curtmola, "Fostering participation in smart cities: a geo-social crowdsensing platform," *IEEE Commun. Mag.*, vol. 51, no. 6, pp. 112–119, Jun. 2013.
- [3] L. Evans, "Maps as deep: Reading the code of location-based social networks," *IEEE Technol. Soc. Mag.*, vol. 33, no. 1, pp. 73–80, Mar. 2014.
- [4] S. M. George, W. Zhou, H. Chenji, M. Won, Y. O. Lee, A. Pazarloglou, R. Stoleru, and P. Baroah, "Distressnet: A wireless ad hoc and sensor network architecture for situation management in disaster response," *IEEE Commun. Mag.*, vol. 48, no. 3, pp. 128–136, Mar. 2010.
- [5] N. C. Tang, Y. Y. Lin, M. F. Weng, and H. Y. M. Liao, "Cross-camera knowledge transfer for multiview people counting," *IEEE Trans. Image Process.*, vol. 24, no. 1, pp. 80–93, Jan 2015.
- [6] B. K. Dan, Y. S. Kim, Suryanto, J. Y. Jung, and S. J. Ko, "Robust people counting system based on sensor fusion," *IEEE Trans. Consum. Electron.*, vol. 58, no. 3, pp. 1013–1021, August 2012.
- [7] A. B. Chan and N. Vasconcelos, "Counting people with low-level features and Bayesian regression," *IEEE Trans. Image Process.*, vol. 21, no. 4, pp. 2160–2177, April 2012.
- [8] E. Cianca, M. D. Sanctis, and S. D. Domenico, "Radios as Sensors," *IEEE Internet of Things J.*, vol. 4, no. 2, pp. 363–373, Apr. 2017.
- [9] H. Li, E. C. L. Chan, X. Guo, J. Xiao, K. Wu, and L. M. Ni, "Wi-counter: Smartphone-based people counter using crowdsourced Wi-Fi signal data," *IEEE Trans. Human-Mach. Syst.*, vol. 45, no. 4, pp. 442–452, Aug 2015.
- [10] S. Depatla, A. Muralidharan, and Y. Mostofi, "Occupancy estimation using only WiFi power measurements," *IEEE J. Sel. Areas Commun.*, vol. 33, no. 7, pp. 1381–1393, July 2015.
- [11] K. K. Mada, H.-C. Wu, and S. S. Iyengar, "Efficient and robust EM algorithm for multiple wideband source localization," *IEEE Trans. Veh. Technol.*, vol. 58, no. 6, pp. 3071–3075, Jul. 2009.
- [12] L. Lu, H. Zhang, and H.-C. Wu, "Novel energy-based localization technique for multiple sources," *IEEE Syst. J.*, vol. 8, no. 1, pp. 142–150, Mar. 2014.
- [13] M. I. Skolnik, *Radar Handbook*, 3rd ed. New York, NY 10020: McGraw-Hill, 1970.
- [14] S. Bartoletti, A. Conti, A. Giorgetti, and M. Z. Win, "Sensor radar networks for indoor tracking," *IEEE Wireless Commun. Lett.*, vol. 3, no. 2, pp. 157–160, Apr. 2014.
- [15] S. Bartoletti, W. Dai, A. Conti, and M. Win, "Threshold profiling for wideband ranging," *IEEE Signal Process. Lett.*, vol. 25, no. 6, pp. 873–877, Jun. 2018.
- [16] S. Bartoletti, A. Giorgetti, M. Z. Win, and A. Conti, "Blind selection of representative observations for sensor radar networks," *IEEE Trans. Veh. Technol.*, vol. 64, no. 4, pp. 1388–1400, Apr. 2015.
- [17] M. Chiani, A. Giorgetti, and E. Paolini, "Sensor radar for object tracking," *Proc. IEEE*, vol. 106, no. 6, pp. 1022–1041, Jun. 2018.
- [18] A. Conti, S. Mazuelas, S. Bartoletti, W. C. Lindsey, and M. Z. Win, "Soft information for localization-of-things," *Proc. IEEE*, to appear.
- [19] M. Z. Win, W. Dai, Y. Shen, G. Chrisikos, and H. V. Poor, "Network operation strategies for efficient localization and navigation," *Proc. IEEE*, vol. 106, no. 7, pp. 1224–1254, Jul. 2018, special issue on *Foundations and Trends in Localization Technologies*.
- [20] S. Bartoletti, W. Dai, A. Conti, and M. Z. Win, "A mathematical model for wideband ranging," *IEEE J. Sel. Topics Signal Process.*, vol. 9, no. 2, pp. 216–228, Mar. 2015.
- [21] M. Z. Win, F. Meyer, Z. Liu, W. Dai, S. Bartoletti, and A. Conti, "Efficient multi-sensor localization for the Internet-of-Things," *IEEE Signal Process. Mag.*, vol. 35, Sep. 2018, to appear.
- [22] R. Niu, R. S. Blum, P. K. Varshney, and A. L. Drozd, "Target localization and tracking in noncoherent multiple-input multiple-output radar systems," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 48, no. 2, pp. 1466–1489, Apr. 2012.
- [23] R. Niu and P. K. Varshney, "Target location estimation in sensor networks with quantized data," *IEEE Trans. Signal Process.*, vol. 54, no. 12, pp. 4519–4528, Dec. 2006.
- [24] F. Meyer, T. Kropfreiter, J. L. Williams, R. A. Lau, F. Hlawatsch, P. Braca, and M. Z. Win, "Message passing algorithms for scalable multitarget tracking," *Proc. IEEE*, vol. 106, no. 2, pp. 221–259, Feb. 2018.
- [25] X. Quan, J. W. Choi, and S. H. Cho, "In-bound/out-bound detection of people's movements using an IR-UWB radar system," in *Int. Conf. on Electronics, Information and Communications (ICEIC)*, Kota Kinabalu, Malaysia, Jan. 2014, pp. 1–2.
- [26] J. W. Choi, J. H. Kim, and S. H. Cho, "A counting algorithm for multiple objects using an IR-UWB radar system," in *IEEE Int. Conf. on Net. Infrastr. and Digital Content*, Beijing, China, Sep. 2012, pp. 591–595.
- [27] J. He and A. Arora, "A regression-based radar-mote system for people counting," in *2014 IEEE Int. Conf. on Pervasive Computing and Communications (PerCom)*, Louis Missouri, USA, Mar. 2014, p. 95102.
- [28] S. Bartoletti, A. Conti, and M. Z. Win, "Device-free counting via wideband signals," *IEEE J. Sel. Areas Commun.*, vol. 35, no. 5, pp. 1163–1174, May 2017.
- [29] S. Mazuelas, A. Conti, J. C. Allen, and M. Z. Win, "Soft range information for network localization," *IEEE Trans. Signal Process.*, vol. 66, no. 12, pp. 3155–3168, Jun. 2018.
- [30] T. Nagler, "A generic approach to nonparametric function estimation with mixed data," *Statistics and Prob. Lett.*, vol. 137, no. C, pp. 326–330, Apr. 2017.
- [31] —, "Asymptotic analysis of the jittering kernel density estimator," *Mathem. Meth. of Statistics*, vol. 27, no. 1, pp. 32–46, Jan. 2018.
- [32] I. Jolliffe, *Principal Component Analysis*. John Wiley & Sons, 2005.
- [33] J. Klemelä, *Smoothing of Multivariate Data: Density Estimation and Visualization*. John Wiley & Sons, 2009.
- [34] V. Vapnik, *Statistical learning theory*. Wiley, New York, 1998, vol. 3.
- [35] D. W. Scott, *Multivariate Density Estimation: Theory, Practice, and Visualization*. John Wiley & Sons, 2015.
- [36] Novelda, "X4 datasheet," <https://www.xethru.com/community/resources/x4-datasheet.106/>, [Online].
- [37] S. Arlot and A. Celisse, "A survey of cross-validation procedures for model selection," *Statistics Surveys*, vol. 4, pp. 40–79, 2010.
- [38] J. W. Choi, S. S. Nam, and S. H. Cho, "Multi-human detection algorithm based on an impulse radio ultra-wideband radar system," *IEEE Access*, vol. 4, pp. 10 300–10 309, 2016.