Journal of Physics: Conference Series

PAPER • **OPEN ACCESS**

A learning-based inversion strategy for passive wireless detection of crowds

To cite this article: Alessandro Polo *et al* 2020 *J. Phys.: Conf. Ser.* **1476** 012012

A learning-based inversion strategy for passive wireless detection of crowds

Alessandro Polo, Mohammad Abdul Hannan, Giorgio Gottardi, and Federico Viani

ELEDIA Research Center (ELEDIA@UniTN - University of Trento), I-38123 Trento, Italy

E-mail: alessandro.polo.1@unitn.it

Abstract. Passive indoor localization is an emerging technology for the detection and tracking of transceiver-free entities using wireless infrastructures. Customized wireless sensors are often used for the acquisition of suitable wireless signals. In this work, neither hardware customization nor dedicated deployment are introduced. The wireless signal transmitted by standard *Wi-Fi* access points is processed to solve an inverse problem for the real-time detection of crowds. Towards this end, a wavelet decomposition of the acquired data is combined with a learning-by-example (*LBE*) strategy in order to learn the complex relation between crowd presence and signal perturbations. Experimental results show the capabilities of the proposed solution in detecting people within a real *Wi-Fi* enabled test site. The obtained performance points out that a standard *Wi-Fi* network can be profitably adopted as a low-cost and scalable solution to support crowd management in many applicative fields.

1. Introduction

Passive wireless localization is based on the idea that the presence of a target moving in a radio-frequency (*RF*) environment affects the electromagnetic (*EM*) propagation, especially when dealing with frequencies whose wavelength is comparable to the target size [1][2]. Applications of passive localization are related both to military (e.g., intrusion detection, border protection) and civil (e.g., patient monitoring in smart home, advanced management of smart museums, shopping centers, airports, hospitals, etc.) contexts. Many approaches based on *RF* signal processing have been introduced [1], overcoming the limitations of traditional tracking systems based on technologies such as cameras and infrared, which are limited to line-of-sight vision and are prone to environmental conditions. The localization and tracking of individuals have received much attention [3]-[5], while the estimation of crowd characteristics has been less investigated. More recent works deal with the localization [6], counting, and activity recognition [7] of multiple subjects. These solutions often adopt dedicated wireless networks for data acquisition, installed in a way to exploit as much as possible the shadowing effects of human bodies. Such customization makes the wireless network suitable for the localization process at hand, but implicitly limits the opportunistic exploitation of existing wireless infrastructures for localization purposes. In a conventional wireless fidelity (*Wi-Fi*) scenario, mobile devices like smart-phones, tablets, laptops, are connected to wireless access points (*APs*) properly deployed to provide internet connection. The wireless links between targets and *APs* are exploited by active localization methods to accurately estimate the position of the mobile devices [8].

In the literature, to the best of authors' knowledge, *Wi-Fi* networks already deployed to provide standard wireless services have been only partially investigated for the passive detection of crowd in large indoor areas. Some studies have verified that the presence of a body in the proximity of *Wi-Fi APs* affects the Received Signal Strength Indicator (*RSSI*) stability [9], because of shadowing effects. However, the following challenges related to the application of standard *Wi-Fi* networks to the wireless localization problem have not been fully addressed:

- The network stability and the quality of service cannot be affected by the introduction of localization methods. As a consequence, no customization for the data acquisition can be applied to the wireless devices.
- The position and the density of the wireless *APs* are defined to guarantee good coverage and signal stability to the end users. Such objectives are opposite respect to that of wireless localization (i.e., deploy the system so that the target effects on the signal are maximized), and consequently the *APs* positions could be inappropriate.
- *Wi-Fi* protocols and standards consider the adaptive control of the available connection parameters (e.g., transmission power, TX/RX antenna selection, *RF* channels, selective radio activation, etc.) to maximize the system performance, taking into consideration the *RF* noise and the client mobility. Such adaptive changes lead to the stabilization of the *RF* signal, important for connection services but unsuitable for passive localization methods.

In this work, an innovative approach for the real-time estimation of crowd presence in large indoor areas covered by *Wi-Fi* signal is proposed. The approach is based on a learning-by-example (*LBE*) method [10]-[13] for the automatic learning of crowd effects caused on the *Wi-Fi* signal. The *RSSI* acquired by a *Wi-Fi* sniffer is processed by a wavelet decomposition procedure aiming at the extraction of representative features related to the crowd presence. A relation between wavelet coefficients and crowd characteristics is identified and successively exploited to train a customized support vector machine (*SVM*) classifier.

2. Passive Crowd Detection Strategy

2.1. Features Extraction from Wireless Signals

Let us consider a *Wi-Fi* wireless network composed by K *APs* deployed in known and fixed positions $\underline{r}_k$, $k = 1, ..., K$, where $\underline{r} = (x, y, z)$ is the position vector. Each *AP* is positioned and already configured to provide good wireless coverage within the considered domain Ω. Toward this end, it is assumed that the transmitting power $P_{TX}^{(k)}$, $k = 1, ..., K$, of each *AP* is controlled by the network manager and cannot be modified for localization purposes. A *Wi-Fi* sniffer is positioned within Ω in order to iteratively scan the *APs* and acquire the corresponding *RSSI* $v_k(\underline{r}_U, t)$, $k = 1, ..., K$, where $\underline{r}_U$ is the position of the receiving sniffer (Fig. 1), and t the sampling time instant. The *Wi-Fi* sniffer is a standard Access Point configured to scan the wireless channel and save the available signal quality indicators on a local database. The scanning procedure is configured to acquire the *RSSI* values of all the K wireless links between the sniffer and the *APs* (let us assume that the sniffer is within the wireless coverage of all the *APs*) at instants $t = t_0 + m\Delta t$, being t_0 the first acquisition instant, Δt a constant time interval, and $m = 0, ..., M-1$ the samples of the acquired *RSSI* time series $\{V\}_k = [v_k(\underline{r}_U, t_0), ..., v_k(\underline{r}_U, m\Delta t), ..., v_k(\underline{r}_U, (M-1)\Delta t)]$. The $\{V\}_k$, $k = 1, ..., K$ *RSSI* time series represent the data source processed to estimate the crowd presence. The crowd is represented by a variable number of individuals $0 < P \leq P_{max}$ occupying the domain Ω.

In literature, experimental results obtained with wireless sensor networks (*WSNs*) have already shown that correlation between the *RF* signal strength and the crowd presence exists [14]. However, it has to be noticed that such correlation is less evident using standard *Wi-Fi* networks, also because of the aforementioned challenges. In order to effectively highlight,

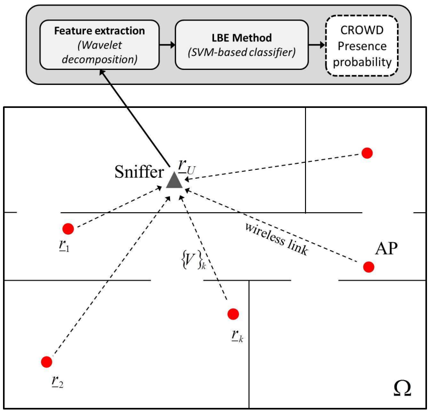


Figure 1. Block scheme of the proposed approach, from the acquisition of $RSSI$ time series $\{V\}_k$, $k = 1, ..., K$, to the estimation of the crowd presence probability.

detect, and extract the effects of crowd on the *Wi-Fi* signal, advanced processing strategies have been investigated. Among various wireless channel analysis techniques, wavelet decomposition represents an elegant tool for the estimation of wireless signal characteristics [15]. The wavelet decomposition provides a representation of the signal at different scales (also called *octaves* in the logarithmic scale). A signal can be represented by its approximation at any scale, and signal patterns can be simultaneously analyzed at different time-frequency levels. The proposed approach applies the Haar discrete wavelet decomposition to decompose the original $RSSI$ signal in successive approximations

$$\overline{A}_s^k = \overline{A}_{s-1}^k + \overline{C}_s^k \quad k = 1, ..., K \tag{1}$$

where $\overline{A}_s^k = (a_{s,e})^k$ is the series approximation at the decomposition scale $s = 1, ..., S$, and $\overline{C}_s^k = (c_{s,e})^k$ are the corresponding coefficients (or details) sets, $e = 1, ..., N_s$, being the coefficient index at each s-th scale, and N_s, $s = 1, ..., S$ the total number of coefficient at each s-th scale. For the sake of clarity, the first-level approximation is $\overline{A}_1^k = \overline{A}_0^k + \overline{C}_1^k$, the second-level is $\overline{A}_2^k = \overline{A}_1^k + \overline{C}_2^k$, up to the last level $\overline{A}_S^k = \overline{A}_{S-1}^k + \overline{C}_S^k$. This formulation, according to the basic Haar wavelet theory, illustrates that the original signal can be represented with a low-resolution approximation $\overline{A}_0^k$, which is the time series bias, added with multi-resolution details, as follows

$$\{V\}_k = \overline{A}_0^k + \sum_{s=1}^{S} \overline{C}_s^k \quad k = 1, ..., K. \tag{2}$$

Increasing the scale index is equivalent to enhance the time resolution of the details, up to the highest scale $S = \log_2(M)$ (M being the samples of the considered $RSSI$ time series). The multi-resolution property of the wavelet coefficients is exploited to analyze the $RSSI$ properties at different scales, aiming at highlighting the effects of the crowd presence hidden in the complex pattern of the time series at hand. Toward this end, the terms

$$\delta_s^k = \frac{1}{N_s} \sum_{e=1}^{N_s} \left|(c_{s,e})^k\right|^2 \quad s = 1, ..., S;\ k = 1, ..., K \tag{3}$$

have been calculated for each k-th wireless link as simple indicators of the changes between two successive approximations $\overline{A}_s^k$ and $\overline{A}_{s-1}^k$, $s = 1, ..., S$. In presence of an unconventional perturbation in the *RSSI* time series, it is expected that such indicators are higher in correspondence to the time-frequency contributions of the perturbation itself. According to this rule, the trend of the indicators $\overline{\delta} = \left[\delta_s^k;\ s = 1, ..., S;\ k = 1, ..., K\right]$ is representative of the perturbation caused by the crowd presence. The general behavior of the decomposed *RSSI* could be summarized with *"the higher the perturbation caused by crowd presence, the higher the values of indicators* $\overline{\delta}$". However, the relation between such indicators and the crowd is not trivial, especially in the transition from *empty* to *crowded* conditions when the *RSSI* perturbations are negligible. Such an ambiguity introduces high complexity in the exact estimation of the crowd presence, which can not be solved by applying simple thresholding strategies. On the contrary, the proposed approach aims at learning the behavior of the considered indicators respect to the crowd presence and to correctly detect even small crowd. Toward this end, the evaluated indicators $\overline{\delta}$ are processed by the *LBE* strategy described in Sect. 2.2.

2.2. Learning Strategy for Crowd Wireless Detection

A customized *LBE* method based on a *SVM* classifier [10] has been defined to learn the complex and non-linear behavior of the indicators δ_s^k, $s = 1, ..., S$, $k = 1, ..., K$, and successively estimate the crowd absence and presence. More in detail, the training set

$$\left\{\left(\delta_s^k, \chi_n;\ s = 1, ..., S;\ k = 1, ..., K\right)^{(n)};\ n = 1, ..., N\right\} \tag{4}$$

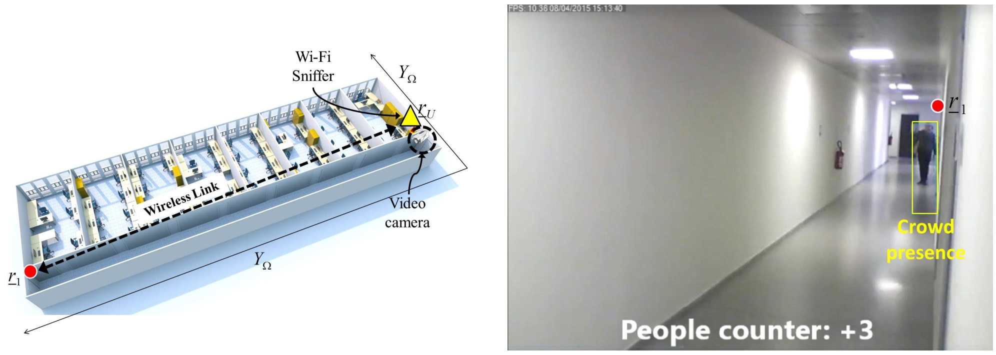
has been defined as the collection of N training samples. $\chi_n = \pm 1$, $n = 1, ..., N$, is the binary class index that represents the a-priori information about crowd absence ($\chi = -1$) and presence ($\chi = +1$). Each *input-output* pair of the training set has been generated according to the following guidelines:

- *Input*: the input features δ_s^k, $s = 1, ..., S$, $k = 1, ..., K$, are evaluated starting from the *RSSI* samples acquired at time instants $t = t_0 + m\Delta t$, $m = 0, ..., M - 1$. For the collection of N training samples, a sliding window is applied on the *RSSI* data stream in order to successively process multiple time periods starting at time $t_0^n = t_0^{n-1} + \Delta t$, $n = 2, ..., N$.
- *Output*: the a-priori information χ_n, $n = 1, ..., N$, is associated with the input features according to a simple and unsupervised approach: predefined crowd absence/presence patterns have been extracted from the a-priori knowledge of the actual crowd presence within the monitored domain. The class index $\chi = -1$ has been assigned to data acquired during *absence* time periods (e.g., during the night-time when Ω is empty), while $\chi = +1$ to the data measured during the day when Ω is crowded. This simplification in the labeling of the training data has enabled the automatic and online update of the training set simply checking the daily hour and setting the class index accordingly.

Once the training set is generated, the problem is formulated as the definition of a suitable decision function separating the two classes $\chi = \pm 1$. Toward this end, the *SVM* classifier defines a linear discriminant function $\Psi(\cdot)$ in a higher dimensional feature space, where the original data are mapped through a non-linear operator φ. In the feature space, the goal of the training procedure is the definition of the hyperplane that maximizes the separating margin between the training data belonging to the two classes [10]. Real-time inversions are then yielded during the successive on-line test phase by processing previously-unseen input samples.

3. Preliminary Experimental Validation

A preliminary experimental result is reported and analyzed in order to assess the effectiveness and the current limitations of the proposed *LBE* inversion approach for crowd detection. With



(a) (b)

Figure 2. Wireless link within the indoor domain Ω for experimental validation (a) and a sample snapshot of the crowd presence ground truth taken by the surveillance system (b).

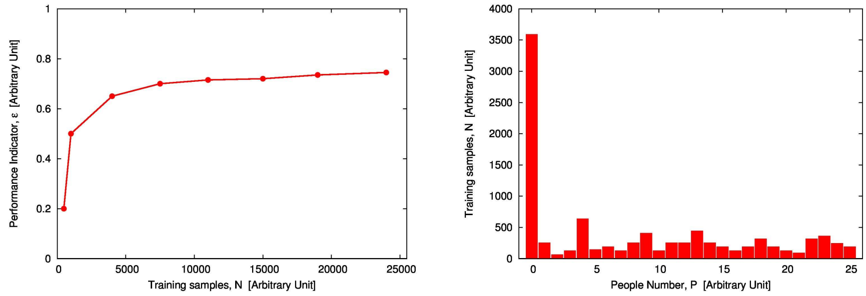
reference to Fig. 2(a), the considered domain is a real indoor test-site within the laboratories of the ELEDIA Research Center at the University of Trento, Italy.
The considered domain has geometrical size $X_\Omega = 15$ $[m]$, $Y_\Omega = 45$ $[m]$, and ceiling height $Z_\Omega = 2.8$ $[m]$. A *Wi-Fi* AP compliant to the well-known communication standard IEEE 802.11a/b/g located in $\underline{r}_1 = (4.5\,[m]\,,\,43.0\,[m]\,,\,2.6\,[m])$ and with transmitting power $P^1_{TX} = 18\,dBm$ has been considered the reference transmitter for the experiments at hand. The *Wi-Fi* sniffer has been positioned in $\underline{r}_U = (5.0\,[m]\,,\,1.2\,[m]\,,\,2.6\,[m])$ and the arising wireless link is 41.8 $[m]$ long (a single wireless link has been considered in this work to preliminary show the potentialities of the proposed method). The sniffer has been configured to acquire *RSSI* values $v_k\,(\underline{r}_U, m\Delta t)$ with $\Delta t = 0.6$ $[sec]$ at the working frequency $f = 2.4$ $[GHz]$. The acquired data are processed in slots of $M = 64$ time samples in order to apply a wavelet decomposition up to scale $S = 6$. The indicators $\delta^1_1, ..., \delta^1_6$ have been computed and used as input features of the learning strategy. The training set size N has been calibrated in order to maximize the system performance in terms of crowd detection capability. Toward this end, the following performance indicator has been defined

$$\varepsilon = 1 - |H\,(P) - Pr\,\{\chi = +1\}| \tag{5}$$

where P is the actual number of people within Ω during the evaluation of the probability $Pr\,\{\chi = +1\}$, and $H\,(\cdot)$ is the step function

$$H\,(P) = \begin{cases} 1 & if\,P > 0 \\ 0 & otherwise \end{cases}. \tag{6}$$

The ground truth of the crowd absence/presence has been collected using a surveillance video camera installed for validation purposes [the view of the domain from the camera is shown in Fig. 2(b)]. The results of the crowd detection versus the training set size, changed in the range $200 \leq N \leq 24000$, are reported in Fig. 3(a).
As expected, the indicator $\varepsilon \in [0 \div 1]$ is lower with small training sets and converges to higher values ($\varepsilon > 0.7$) when N increases, since an higher a-priori information is provided during the training phase. According to the obtained results, the size of the training data set has been calibrated to $N = 7500$, which is the best trade-off between the minimization of the training set size (aiming at avoiding the well-known overfitting problem) and the maximization of the performance indicator ε. The distribution of the selected training samples respect to the number of people P occupying Ω during the training periods has been reported in Fig. 3(b). As it can be observed, the samples are uniformly distributed between $P_{min} = 1$ and $P_{max} = 25$.



(a) (b)

Figure 3. (a) Performance indicator ε versus training set size $200 \leq N \leq 24000$ and (b) distribution of $N = 7500$ training samples respect to the number of people $0 \leq P \leq P_{max}$, $P_{max} = 25$ occupying Ω during the training phase.

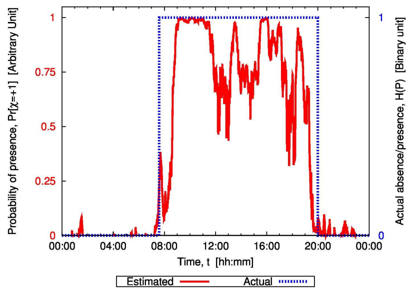


Figure 4. (a) Comparison between actual crowd absence/presence $H(P)$ and the estimated probability $Pr\{\chi = +1\}$ during a *24*-hours time period; (b) long-term performance assessment of daily crowd detection including the weekend time period.

The samples acquired when $P = 0$ are associated to the class $\chi = -1$. Such a distribution of training samples has been chosen to guarantee well-balanced SVM classes ($N|_{\chi_n=-1} = 3550$ and $N|_{\chi_n=+1} = 3950$).
The illustrated experiment aims at validating the detection capabilities during a *24*-hours time period ($M = 144 \times 10^3$, $\Delta t = 0.6$ $[sec]$) characterized by a highly time varying crowd conditions. The method has been verified using unknown test data acquired the day after the collection of the selected training set ($N|_{\chi_n=-1} = 3550$ have been collected during the night-time from time *01:00* [*hh:mm*] to *01:35* [*hh:mm*], while the remaining samples $N|_{\chi_n=+1} = 3950$ during the day-time with different crowd conditions). Figure 4 shows the daily estimation of the crowd presence probability $Pr\{\chi = +1\}$ compared with the binary step function $H(P)$. The obtained results clearly show the correct estimation of the crowd absence/presence. As expected, empty area during the night-time caused $Pr\{\chi = +1\} = 0$ with only negligible and isolated spikes due to the intrinsic instability of the $RSSI$ behavior, while crowd presence in the day-time has been correctly estimated pointing out higher probabilities during the whole crowded time periods (from *07:15* [*hh:mm*] to *20:10* [*hh:mm*]. In particular, it has to be noticed that the first entrance in the monitored area has been correctly detected at 07:15 [*hh:mm*], even if only one target individual ($P = 1$) was moving along the corridor shown in Fig. 2. The quality of the detection estimation reported in Fig. 4 has been quantified by the performance indicator ε, which turns out to be equal to $\varepsilon = 0.78$.

4. Conclusions

In this work, the problem of indoor passive detection of crowd has been formulated as an inverse problem and it has been addressed exploiting the *RSSI* information opportunistically acquired from existing *Wi-Fi APs*. The *RSSI* time series has been decomposed by the Haar wavelet to exploit its time-frequency multi-resolution properties, and to profitably extract the signal patterns caused by crowd presence. A suitable *SVM*-based classifier has been customized to learn the unknown relation between the crowd presence and the values of the wavelet coefficients. The capabilities of the proposed approach to estimate the crowd presence have been experimentally verified in a real indoor test-site. The obtained preliminary results have confirmed the ability of the method to detect the crowd.

Acknowledgments

This work benefited from the networking activities carried out within the Project "SMARTOUR - Piattaforma Intelligente per il Turismo" (Grant no. SCN_00166) funded by the Italian Ministry of Education, University, and Research within the Program "Smart cities and communities and Social Innovation", and the Project "CYBER-PHYSICAL ELECTROMAGNETIC VISION: Context-Aware Electromagnetic Sensing and Smart Reaction (EMvisioning)" funded by the Italian Ministry of Education, University, and Research within the PRIN2017 Program.

References

[1] Patwari N and Wilson J 2010 RF sensor networks for device-free localization: Measurements, models, and algorithms *Proc. IEEE* **98** 1961-1973

[2] Liu C et *al.* 2016 RSS Distribution-based passive localization and its application in sensor networks *IEEE Trans. Wireless Commun.* **15** 2883-2895

[3] Viani F, Migliore M D, Polo A, Salucci M and Massa A 2018 Iterative classification strategy for multi-resolution wireless sensing of passive targets *Electron. Lett.* **54** 101-103

[4] Ahmadi H, Polo A, Moriyama T, Salucci M and Viani F 2016 Semantic wireless localization of WiFi terminals in smart buildings *Radio Sci.* **51** 876-892

[5] Wilson J and Patwari N 2012 A fade-level skew-Laplace signal strength model for device-free localization with wireless networks *IEEE Trans. Mobile Comput.* **11** 947-959

[6] Nannuru S, Li Y, Zeng Y, Coates M and Yang B 2013 Radio frequency tomography for passive indoor multi-target tracking *IEEE Trans. Mobile Comput.* **12** 2322 - 2333

[7] Sigg S, Scholz M, Shi S, Ji Y and Beigl M 2014 RF-sensing of activities from non-cooperative subjects in device-free recognition systems using ambient and local signals *IEEE Trans. Mobile Comput.* **13** 907-920

[8] Feng C, Anthea W S, Valaee A and Tan Z 2012 Received-signal-strength-based indoor positioning using compressive sensing *IEEE Trans. Mobile Comput.* **11** 1983-1994

[9] Wyne S, Santos T, Singh A P, Tufvesson F and Molish A F 2010 Characterisation of a time-variant wireless propagation channel for outdoor short-range sensor networks *IET Commuications* **4** 253-264

[10] Massa A, Oliveri G, Salucci M, Anselmi N and Rocca P 2018 Learning-by-examples techniques as applied to electromagnetics *J. Electromagn. Waves Appl.* **32** 516-541

[11] Salucci M, Anselmi N, Oliveri G, Calmon P, Miorelli R, Reboud C and Massa A 2016 Real-time NDT-NDE through an innovative adaptive partial least squares SVR inversion approach *IEEE Trans. Geosci. Remote Sens.* **54** 6818-6832

[12] Salucci M, Vrba J, Merunka I and Massa A 2017 Real-time brain stroke detection through a learning-by-examples technique - an experimental assessment *Microw. Opt. Technol. Lett.* **59** 2796-2799

[13] Salucci M, Tenuti L, Oliveri G and Massa A 2018 Efficient prediction of the EM response of reflectarray antenna elements by an advanced statistical learning method *IEEE Trans. Antennas Propag.* **66** 3995-4007

[14] Yuan Y, Zhao J, Qiu C and Xi W Estimating crowd density in an RF-based dynamic environment *IEEE Sensors J.* **13** 3837-3846

[15] Hess-Nielsen N, Wickerhauser M V 1996 Wavelets and time-frequency analysis *Proc. IEEE* **84** 523-540