This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/COMST.2019.2902824, IEEE Communications Surveys & Tutorials

Survey of Non-Image Based Approaches for Counting People

Sylvia T. Kouyoumdjieva*, Peter Danielis†, and Gunnar Karlsson*
*ACCESS Linnaeus Center, School of Electrical Engineering and Computer Science,
KTH Royal Institute of Technology, Sweden
†Faculty of Computer Science and Electrical Engineering, University of Rostock, Germany
E-mail: stkou@kth.se, peter.danielis@uni-rostock.de, gk@kth.se

***Abstract*—Accurate people count estimation, potentially in real-time, both for indoor and outdoor environments, is said to be of major importance in the smart cities of tomorrow. Application areas such as public transportation, urban analytics, building automation as well as disaster management are all expected to benefit if they were to have a better understanding of occupancy in public premises. A large body of work has been concentrated into providing people counting solutions based on images captured by surveillance cameras. However, image-based approaches are costly, as they require devoted hardware installations, and are often privacy intruding. Thus, academic and industry researchers are looking into alternative solutions for people counting. In this work we present a comprehensive study of non-image based people counting techniques. Our goal with this work is two-fold: (1) to serve as an introduction to everyone interested in gaining a better understanding on non-image based people counting techniques, and (2) to serve as a guideline to practitioners interested in implementing and testing specific solutions in their everyday practice. To this end, we provide a novel classification of available approaches, and outline the requirements they need to meet. We further discuss in detail different academic solutions, and provide comparisons between them. Furthermore, we provide a discussion on available industrial approaches and compare them to academic proposals. Finally, we discuss open challenges and future directions in the field of non-image based people counting.**

I. Introduction & Motivation

Providing accurate estimate of the number of people at a given location, also in real time, is of high importance in the context of smart cities: from planning public spaces and building automation to surveillance and crowd management in the case of natural or man-made disasters. However the techniques that we use today for estimating the number of occupants in a given area or location are often inaccurate. Let us look at a few examples of people counting inconsistencies that were reported in the last two decades:

- The Million Man March (1995). During a manifestation for civil rights of African-Americas which took place in Washington, D.C., event organizers estimated the number of participants in the march to be somewhere between 1,500,000 and 2,000,000 people. The National Park Service, a US federal government agency, however issued an estimate of around 400,000 attendees. Finally, after heated discussions, researchers at Boston University estimated the crowd size to approximately 837,000 participants, with 20% margin of error [1].
- The 25th anniversary of the Tiananmen Square massacre (2014). Reports describing the level of participation in Beijing provided inconsistent crowd estimates: from 100,000 — an estimate provided by the police, up to 180,000 people — an estimate provided by the event coordinators [2].
- The protest against the name of Republic of Macedonia (2018). Recently, during a protest in the center of Athens event organizers estimated that approximately 1,500,000 people took part in the march. However police officials estimated the attendance at 140,000 [3].

Although these large discrepancies in the number of participants may come as a result of an intentional bias introduced by each of the institutions providing the estimate, for instance in order to boost or to devalue the significance of a public gathering, it illustrates one major drawback: the lack of modern reliable technologies to produce accurate people counts that do not rely on subjective or political influences. Today we still rely on a systematic approach devised in the late 1960s by journalism professor Herbert Jacobs at the University of California, Berkley. The "Jacobs method" states that if an area is divided into b blocks, and the average number of occupants per block n is known, the average number of people in the area can be calculated by a simple multiplication $b \times n$ [4].

Following the Jacobs method, nowadays estimating the number of occupants in an area is devised automatically based on images captured by aerial cameras, often positioned at different heights. Applying automatized image-based solutions to perform large-scale counting however comes at a price in terms of installation and maintenance, as well as computation, and unfortunately is often privacy-intruding and lacking precision. Moreover, these solutions provide a relatively good estimate only when the line of sight between the camera capturing the image and the participants is unobstructed. In all of the examples above, the estimate can be distorted by participants hidden under trees or umbrellas.

The previous examples all focus on measuring the size of large crowds. However even when crowd sizes are much smaller, e.g. tens or hundreds of people, providing an accurate estimate in an automated manner is still a challenging task. Although in theory the Jacobs method could be applied to any scenario, the practical limitations we outlined above (and we discuss in more detail in Section I-C) often call for alternative

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/COMST.2019.2902824, IEEE Communications Surveys & Tutorials

solutions, which are mostly implemented on a per case basis. For instance, counting passengers on an airplane was traditionally done manually with a clicker, and has only recently been automatized by boarding pass scanners. (Note that manual people counting is labor-intensive and becomes error-prone as the size of the crowd increases; we remind the reader how passenger counting on planes was often done multiple times by different crew members equipped with clickers before the plane was ready to take off.) Another example is counting passengers in public transportation which often relies on ticket verification upon boarding a vehicle (although sensor-based solutions or light barriers are becoming more popular). A ticketing system is also used for counting museum visitors. However such solutions provide only general information about incoming and outgoing people flows, without any details of occupancy levels within the premises; they rarely operate in real-time, are not easy to scale and their applicability is limited to the particular use case. Accurately capturing incoming and outgoing people flows, and the occupancy they represent, is of highest importance when dealing with open systems, especially in the context of evacuation scenarios, where a crowd estimate needs to precisely capture people dynamics within a given area in real-time. Finally, significant efforts are made towards accurately estimating people counts in the context of smart buildings. Precise understanding on people dynamics within buildings is said to be the main factor for optimizing the operations of heating, ventilation, and air conditioning (HVAC) systems thus minimizing energy consumption indoors.

A. Definitions

In the previous section we discussed examples of people counting with large variations in the number of occupants in a given area: from hundreds to few hundred thousand. To be able to collectively describe all these use-cases, in this work we use the term *crowd* to denote any group of people who are linked together by a common geographic location or activity. In other words, the population of a city forms a crowd, just as much as the employees that occupy a firm's office form a crowd. Crowds may form sporadically (i.e., during a manifestation) or regularly (i.e., during rush hour in public transportation), and may be either static (i.e., the participants in the crowd stay the same as long as the crowd exists) or dynamic (i.e., participants come and go, thus the size of the crowd changes constantly).

Throughout this paper we use the terms *crowd counting*, *people counting* and *occupancy estimation* in order to describe the process of estimating the size of a population at a given location.

B. Types of occupancy estimation

In general, occupancy estimation can be divided into three main categories: binary, ranged and exact.

A *binary* occupancy estimate (often referred to as *occupancy detection*) has the sole purpose of evaluating whether people are present in a given location or not. Solutions range from using electricity consumption data in home environments [5,6] to relying on contextual data, such as calendars, instant messages and Wi-Fi associations in commercial buildings [7] and installing sensors at doorway passages for tracking individuals with room-level accuracy [8]. One of the main application areas for occupancy detection solutions is thus in the context of smart buildings automation, often for optimizing energy consumption in the heating, ventilation and air conditioning (HVAC) system [9].

A *ranged* occupancy estimate is suggested as a solution when the exact number of people in a space is not of importance, however, what matters is the order of magnitude of the estimate. Ranged occupancy estimation is significantly more challenging than binary occupancy estimation however it is also more feasible that estimating the exact number of people in a given location. This estimate is often represented as a histogram where each bin corresponds to a range of occupants. Our literature survey however shows that ranged occupancy estimation is rarely explored; one possible reason for this may be that most application scenarios require either a yes/no answer (i.e., binary estimation), or an exact estimation.

Finally, an *exact* occupancy estimate is preferred for those scenarios in which the accurate number of people in a given space is required, e.g., for safety or billing purposes. In the current work, we focus exclusively on surveying solutions that aim at providing exact occupancy estimates.

C. Paper scope

Crowd counting solutions can generally be divided into two main categories: *image-based* (also referred to as video-based or visual) and *non-image based* approaches. In particular, the term *image-based* refers to solutions including video and image sources that provide visuals, which can be used by an observer to recognize a person including his or her features.

Image-based crowd counting has been the preferred method for density estimation ever since Herbert Jacobs introduced the Jacobs method in the late 1960s. It was further enhanced by the introduction of closed-circuit television (CCTV, also known as video surveillance) in public locations, primarily for the purposes of fighting crime. A large body of work in academia and industry has been devoted to image-based crowd counting methods. In essence, image-based crowd counting relies on computer vision techniques to determine the number of occupants at a given location; the input is provided by one or more video sources, often surveillance cameras. Some of the main approaches used for video-based people counting include object boundary detection in videos [10], image segmentation followed by per segment counting [11], as well as object density estimation based on a random forest algorithm [12]. Over the last decade a number of works have summarized the advances in crowd analysis using computer vision techniques [13]–[16]. Recently more attention has been given to density estimation and people counting with help of surveillance cameras [17,18]. We refer the reader to any of these surveys in case she is interested in gaining more knowledge in the area of image-based crowd analysis.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/COMST.2019.2902824, IEEE Communications Surveys & Tutorials

However, we will note that image-based density estimation suffers drawbacks. Firstly, it is a costly approach as it involves camera installations in appropriate locations, as well as heavy image-processing [19]. The locations are however fixed which may lead to blind spots where people counting is performed inaccurately or not at all. Although the advent of drones equipped with high resolution cameras allows for easier hardware introduction in open spaces, especially outdoors, heavy image-processing as well as high costs still present challenges. To achieve high accuracy, a line-of-sight between cameras and the people is required in all cases, which otherwise may lead to blind areas, and thus under-estimations. Furthermore, environmental conditions (such as poorly lit spaces [20]) and overlapping objects may degrade the quality of the crowd estimate. Finally, using cameras often poses privacy concerns, as for instance stated in [21]–[23], which might be the main reason to renounce image-based solution.

Due to the existing survey work done in the field, in this work we do not include further discussion on image-based crowd counting methods. Instead, we put the emphasis on the second category, namely non-image based approaches. Broadly speaking, non-image based approaches encompass all methods for people counting that do *not* require cameras or other video sources.

Few studies incorporate limited discussions on non-image based approaches for estimating crowd sizes. For instance, Di Domenico et al. present in [24] a comparison of two of the available techniques using dedicated hardware equipment, namely crowd estimations based on received signal strength and on channel state information obtained by monitoring packets that are communicated between two radio nodes, as part of their related work section. Another recent study that discusses crowd counting from a perspective similar to the one we investigate in this work is [25], in which Irfan et al. provide a short comparative survey on crowd analysis using visual and non-visual sensors. The authors conclude the large potential of crowd counting solutions based on mobile devices in dense scenarios or evacuation scenarios. As complement, we here provide a comprehensive survey of all non-image based approaches, and do not limit ourselves only to approaches relying on mobile devices. To this end, we summarize a large body of work published in the past decade on the topic of people counting for various application scenarios. We further present in detail important findings, techniques and approaches used for people counting, and discuss the applicability of available techniques with respect to potential use-cases in the context of smart cities. To the best of our knowledge, this is the first survey of non-image based approaches for people counting.

Our goal with this work is two-fold. First, we envision that it serves as an introduction to the topic of non-image based crowd counting approaches both to experienced professionals coming from the area of image-based crowd analysis, as well as novice researchers who are interested in entering the exciting field of density estimation. To this end, we provide a comprehensive survey of available non-image based crowd counting methods, and we outline potential open questions and future directions that need to be addressed by the community. We further envision this work to serve as guidelines to practitioners who are interested in implementing and testing crowd counting solutions in their everyday practice. To this end we provide detailed information summarizing the capabilities of existing solutions, outlining their advantages as well as their disadvantages.

The main contributions are:

1) We provide a novel classification of current techniques used in the context of non-image based people counting methods, and we discuss the requirements that need to be met by the ultimate crowd counting algorithm.
2) We analyze the applicability of available people counting solutions in the context of both indoor and outdoor crowd counting, and we outline potential application areas where crowd counting will be of high importance in the future.
3) We demonstrate that there is no one-size-fits-all solution readily available neither on the market, nor from the academic research.
4) We outline the lessons learned from previous studies, and discuss open questions that need to be addressed by the community in order to provide exact crowd estimates across application scenarios.

The remainder of this paper is structured as follows: In Section II we present a taxonomy of existing non-image based people counting techniques, and discuss potential applications of people counting from academic and industrial perspective. In Sections III and IV we present a thorough study of device-free and device-based people counting methods, respectively, and outline the advantages and disadvantages of each solution. In Section V we further provide a comparison of the two categories, and discuss how suitable they are for each of the application areas outlined previously. We then discuss open challenges and future directions in Sections VII. Finally, we conclude the work in Section VIII.

II. Taxonomy, Requirements and Application Areas for Non-image based People Counting

In general, crowds can be static or characterized by dynamic and abrupt movement. We mainly focus on classifying methods for counting people in dynamic, highly mobile environments. Such methods are more demanding than counting immobile people but can still be used in static environments. Fig. 1 introduces the taxonomy of non-image based people counting techniques. As shown in this classification, techniques for non-image based people counting can be mainly categorized into *device-based* and *device-free*. An approach is considered to be device-based if the crowd estimation is done with the help of a smartphone or some wearable sensor device carried by occupants in an area; for some solutions, a dedicated application may be installed on each mobile device to facilitate data collection. Alternatively, an approach is classified as device-free if the crowd counting process is done without any intervention from the occupants in an area; instead occupancy

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/COMST.2019.2902824, IEEE Communications Surveys & Tutorials

measurements are collected via alternative channels such as signal attenuation or CO_2 concentration. To this end, the space in which occupancy is to be measured is often instrumented with additional hardware. We discuss in greater detail each of the approaches that constitute device-based and device-free non-image based people counting in the following two sections.

A. Requirements for people counting systems

In this section we discuss fundamental requirements that should be considered when designing systems and methods for people counting. Note that these requirements apply both to device-based and device-free non-image based people counting techniques.

- **Accuracy** — To be useful, a people counting system should be able to accurately determine the number of occupants at a given location (potentially in real-time). Accuracy is often expressed as a percentage value. However as there are no standardized definitions in the context of people counting, an 85% accuracy could describe either a system that achieves accurate measurements 85% of the time, or a system that could tolerate up to 15% of error in the crowd estimate at any point in time. Moreover, the required level of accuracy of the system could also vary based on the particular application at hand. For instance, for HVAC applications in smart buildings a certain estimation error may be tolerable as long as the overall occupancy trend is reflected, while for emergency evacuation applications of the same buildings a 99% accuracy may be required.
- **Automation** — To be reliable and feasible to deploy, a people counting system should not rely on manual user intervention.
- **Scalability** — To be useful in a large number of contexts, both indoors and outdoors, a people counting system should be scalable with respect to crowd size and crowd dynamics. The system should be able to perform equally well in small spaces (i.e., an office cubicle or conference room) as well as in large spaces (i.e., an open air festival). Moreover, the system should be able to adapt to churn, i.e. people entering and leaving a particular location for which the crowd counting is performed, as well as it should be able to react to the mobility of people within a location.
- **Cost** — To be easy to adopt across different application scenarios, a people counting system should incur minimal costs of hardware, installation and maintenance. This might require exploiting existing infrastructures or utilizing off-the-shelf hardware, or it might require development of innovative hardware components that could easily fit into existing infrastructures.
- **Latency** — To provide fresh information to potential customers, a people counting system should be able to report changes in the number of occupants in a timely manner. Again, the time granularity of occupancy updates may vary from one application to another, and may be a function of the speed of change of occupancy at a given location. Moreover, certain applications may require crowd estimations in real-time while others may benefit offline crowd estimations that could be used for analysis and future predictions. For instance, in the context of public transportation, crowd estimations during rush hours may require updates in real-time if crowdedness is factored in journey planning, or it may be done offline if crowdedness is considered only for network dimensioning.
- **Privacy** — To be acceptable for participants in the crowd, a people counting system should be non-intrusive from a privacy perspective. Furthermore, it should only gather information that is needed for providing the service and ideally, it should be hard to exploit the information for non-related use cases.
- **Reliability** — To be reliable, a people counting system should be tamper-proof. For instance, malicious users should not be able to manipulate the count estimate, i.e., the system should be able to avoid inconsistencies as the ones presented in Section I. Moreover, malicious users should not disable the operation of the people counting system.

B. Potential applications for people counting

There are a number of application scenarios in which people counting could add value to existing infrastructures or provide better understanding when designing new ones. Below we provide a list of application areas where people counting would be beneficial:

- **Public transportation** — In the context of intelligent transportation systems, accurate people counting could provide real-time information about passenger occupancy levels across various means of transportation and sections of the system. This information could be used by passengers during journey planning or by authorities for dimensioning the transportation system. Moreover, it is also crucial for safety purposes.
- **Urban analytics** — Crowd density estimations, both indoors (e.g., in shopping malls or conference sites) and outdoors (e.g., during open air festivals or fairs), could be provided to specific authorities and business owners (either in real-time or not) in order to improve the understanding of crowd dynamics in urban environments as well as to justify economic decisions.
- **Building automation** — Knowing the distribution of people in a building, such as number of occupants on a floor or in a room could be used for optimizing the HVAC system operation, in order to reduce energy consumption and improve the indoor climate in buildings.
- **Disaster management** — During evacuation operations having an updated understanding of the number of people in the disaster area could provide helpful information to the rescue teams.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/COMST.2019.2902824, IEEE Communications Surveys & Tutorials

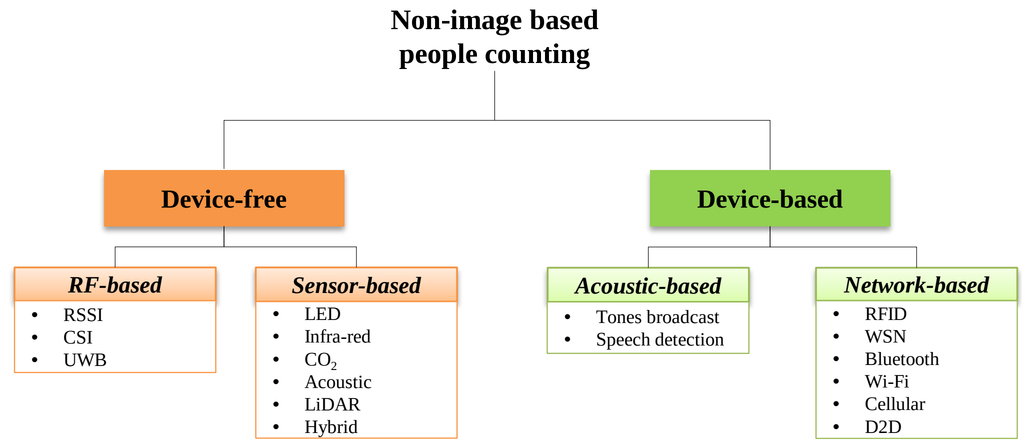

Fig. 1. A taxonomy of non-image based people counting techniques. Each technique is classified into two classes and further categorized if necessary.

III. Device-free methods

The main characteristic of device-free methods is that they do not expect participants to carry additional devices; instead these methods attempt to explore alternative means to determine the size of a crowd. Device-free methods can be broadly classified into *radio frequency (RF) based* and *sensor-based* solutions.

Figure 2 illustrates the general process of inferring occupancy estimation $\hat{y}$ from measured data $\tilde{x}$. The process often consists of a training phase followed by a test phase. During the training phase, a fixed number of people y is presented in the area under investigation. This controlled experiment setup allows the system to learn the dependency of y on $\tilde{x}$ and also to determine the accuracy of $\hat{y}$. During the test phase when the system is taken into operation, $\tilde{x}$ is measured to infer occupancy. Generally, as data measurements $\tilde{x}$ (regardless of whether they are obtained during the training or the test phase) are often noisy, they need to be preprocessed before they can be used for accurately inferring occupancy (cf. step 1 in Figure 2). Finally, a method-specific model/algorithm is used to compute the occupancy estimate $\hat{y}$ from the preprocessed data x (cf. step 2 in Figure 2).

Table I summarizes the wide variety of analytic models and algorithms used for devising occupancy estimations for device-free methods. (Observe that these models are also categorized with respect to sub-categories of RF-based and sensor-based solutions; we look into more detail in this classification later in this section.) The general concepts behind each technique are:

- Event counting is a technique in which the count number is incremented or decremented every time a new measurement value is recorded.
- Equation-based models consist of a system of linear or differential equation, which take input values and produce the occupancy estimation as a result.
- Regression analysis is a statistical modeling technique used for establishing relationships between measured data and the corresponding people count. The regression coefficients used to describe these relationships may be given either a priori or they can be obtained through machine learning.
- Probabilistic-based models can derive the people count by comparing data measurements to given or learned probability distributions, which reflect the expected occupancy estimation.
- Machine learning methods are used to classify data measurements, with each class representing a specific crowd size estimate. Machine learning methods provide both ranged [26] and exact [27,28] occupancy estimates.

We note that certain solutions may leverage more than one algorithm or model when calculating occupancy. For instance, often machine learning methods are applied to determine parameters for regression analysis and probabilistic-based models [27,28].

In the sections to follow, we discuss in detail the state-of-the-art solutions that fall into each of the categories, and in Table II we provide an overview of their advantages and disadvantages.

A. Radio frequency-based solutions

In the following subsections, we divide the device-free RF-based solutions into the following three categories with respect to the radio signal characteristics they utilize for deriving occupancy measurements, namely solutions based on (1) received signal strength indicator (RSSI); (2) channel state information (CSI); and (3) ultra-wide band signals (UWB).

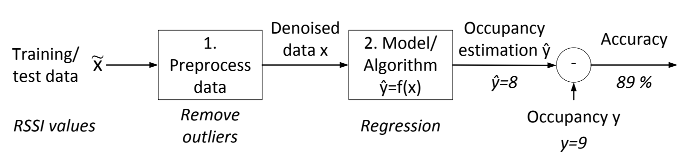

Fig. 2. Step model of how to infer occupancy from measured data for device-free methods. An example (shown in italics below each step) is based on [27].

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/COMST.2019.2902824, IEEE Communications Surveys & Tutorials

TABLE I
ANALYTIC MODELS USED FOR OCCUPANCY ESTIMATION VIA DEVICE-FREE METHODS. ASSIGNMENT OF INVESTIGATED WORKS TO CORRESPONDING CLASS OF MODELS.

	Class	Algorithm/Model				
		Event counting	**Equation-based**	**Regression**	**Probabilistic-based**	**Machine learning**
RF-based	RSSI-based			[27,29]	[26,28,30,31]	[21,26]–[33]
	CSI-based				[22,24]	[22,24]
	UWB-based	[34,35]		[36]	[37]–[39]	[36,38,39]
Sensor-based	PIR-based	[40,41]		[42]	[40,43]	[42,43]
	LED-based			[44,45]		[45]
	Acoustic-based			[46]	[20,47]	[46,47]
	CO_2-based		[48]–[51]	[52,53]		[52,53]
	LiDAR-based	[54]			[55,56]	
	Hybrid				[57]	[58]–[60]

1) Received signal strength indicator based solutions: RSSI is a measurement of the power level that a radio device receives from a radio infrastructure node (such as a ZigBee or a WLAN access point) in a given location and at a given time. In ideal environments, the RSSI value is constant when there is an unobstructed line of sight between the sender and the receiver, and it decreases whenever an obstacle is introduced between them, also when the obstacle is a person. Thus, occupancy in an area can be measured based on the different RSSI values measured at the receiving device. The interested reader is referred to [30] for more information.

Usually, in the training phase RSSI values are measured over some time period, typically in a specific, often controlled, environment. These measurements are then used to establish a model that describes the dependency of the people count on the measured RSSI values (cf. step 2 in Fig. 2). In the simplest case, for instance, the model may assume a linear relationship between RSSI measurements and the people count. In the more complex case, RSSI measurements are classified into different classes representing different numbers of people. In the testing/monitoring phase, the number of people is estimated by feeding measured RSSI values into the model and taking the output as an estimate.

Ideally, the measured RSSI value would result only from the signal received through line-of-sight propagation. However, especially indoors multipath propagation falsifies the measured data. Therefore, the measurements usually have to be pre-processed to mitigate the resulting errors by, e.g., removing outliers (cf. step 1 in Fig. 2).

The feasibility of using RSSI for estimating crowd density is first demonstrated in [29] in 2008. Nakatsuka et al. developed a prototype to measure RSSI values between two ZigBee sensors and between two WLAN access points over a period between 15 s and 60 s while introducing different number of subjects in the observed area. The authors use linear regression to develop a model for computing crowd density estimates based on the average and variation of RSSI measurements; the computations are done on an external server. The authors show that the RSSI between two radio nodes decreases with

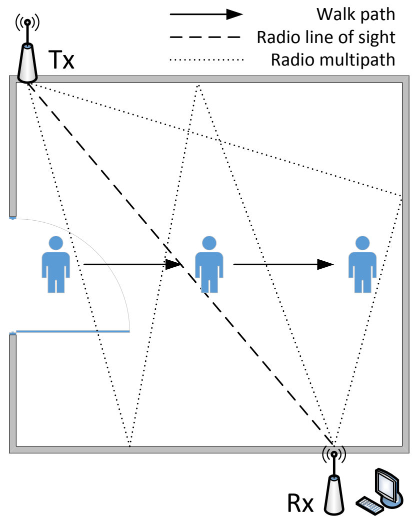

Fig. 3. For RSSI-based crowd counting, a sending node (Tx) such as a ZigBee or WLAN AP sends a packet and a receiving node (Rx) measures the RSSI value of the received signal. People moving between Tx and Rx change the RSSI value compared to the value measured in an empty area. From the different values, occupancy can be estimated. However, multipath propagation falsifies the measurements and may lead to wrongful estimations.

an increasing number of subjects located between these nodes, and that it does not make a difference whether ZigBee or WLAN is used as communication technology. The proposed linear model is able to count up to 29 people.

In [30] Xu et al. present a fingerprinting-based method using probabilistic classification approaches for subject localization. The 16 radio nodes, eight for sending and eight for receiving packets in the experiments, contain a Chipcon radio transceiver and a 16-bit Silicon Laboratories microprocessor. Assuming the resulting number of 64 independent radio links and a room divided into 32 cells, each 0.75 m $\times$ 0.75 m in size, in the training phase RSSI values are measured for each link and cell. First, RSSI values are measured with no person present and then one subject is placed in the room. This one subject enters each cell in turn, stands in the middle of each cell, and spins. Finally, the person re-enters each cell one at a time and moves randomly in each cell before entering the next cell. The data is sent to a centralized system for data collection and analysis. The change of RSSI values between these two cases, referred to as RSSI footprint, is stored for each cell and link and a classifier is built based on the RSSI footprints. In the testing phase, a multiclass classifier technique is used to estimate whether a subject is located in which cell. This technique mitigates errors caused by the multipath effect, especially indoors, and improves localization accuracy while reducing the training overhead. However, taking RSSI fingerprints for a site survey is time-consuming, labor-intensive, and environmental dynamics change the fingerprint so updates are continuously required.

Xu et al. extend their localization approach from [30] in [31] where they propose an algorithm that is able to count multiple subject in addition to localizing them. In the training phase, they still use a single subject's training data to keep the training overhead low. In the testing phase, multiple subjects appear randomly. First, they are counted in rounds to remove the influence of each subject one after the other. Second, the subjects are localized. The authors show that they can count and localize up to four subjects with an average accuracy of 86% in typical indoor environments (150 m^2 office cubicles, 400 m^2 open floor plan). However, this accuracy comes at a price of additional hardware installation; more than 20 devices are introduced in the premises in which localization is performed. More precisely, the office environment is divided into 37 cells and 13 radio transmitters and nine radio receivers are utilized. The open floor space is partitioned into 56 cells and 12 transmitters and eight receivers are used. Moroever, the scalability of the approach is limited by the number of cells, in which the room is divided, as per cell solely one subject can be counted. Counting people located at the edge of two cells also presents a challenge, which may deteriorate the accuracy of the estimate.

Yuan et al. propose a crowd density estimation method for sensor networks consisting of three phases, which are constantly repeated [32]. In a training phase, a fingerprinting database is established from RSSI values, which are measured by TelosB nodes, each periodically broadcasting beacon messages and listening to them, and reported to some external device(s). In the monitoring phase, a K-means clustering algorithm is applied to the measured data to cluster different types of crowd density. To improve their RSSI-based method, the work utilizes an iterative calibration mechanism in the third phase to reduce the estimation error in the measurements. Experiments are carried out in a 7.2 m $\times$ 7.2 m room with 16 TelosB nodes and one or more persons moving to different positions. Furthermore, large-scale simulations with 400 nodes deployed on a 20 m $\times$ 20 m grid are run achieving estimation results with an accuracy of 81%. In [21], they extend their work to a simulation network of 400 nodes on a 100 m $\times$ 100 m grid and people moving randomly. The accuracy of a correct estimation reaches an average of 83% even if 5% to 10% of all RSSI measurements are incorrect.

In [27], Yoshida and Taniguchi propose a regression-based approach for counting people. One AP in a room periodically broadcasts beacons every 100 ms and RSSI measurements are collected by 10 receiving nodes such as laptops, televisions, printers, and tablets. Data is sent to a centralized people counting computer every 300 s. It first removes outliers and then estimates the people count by means of linear or support vector regression (SVR) for the measurement interval of 300 s. The linear regression weights the RSSI median measured at each node with some coefficient to estimate the number of people. SVR is a non-linear regression technique based on one of the most common machine learning techniques, namely support vector machine. It is assumed that the respective coefficients of the regression formulas have been learned in a training phase from measurement data of three days. However, the authors do not describe exactly how the coefficients are extracted from the data, which makes the traceability of the experiments more difficult. The authors show that the SVR-based approach outperforms linear regression, counting up to seven people with an accuracy of at most 77%.

In [28], Depatla et al. estimate crowd sizes based on RSSI measurements obtained from one single stationary WLAN transmitter and one single stationary WLAN receiver. They develop an analytic model describing the probability distribution of the received signal amplitude as a function of the total number of occupants. This probability distribution function is then compared to the one obtained from experiments using the Kullback-Leibler divergence; this metric shows how much two probability distributions diverge from one another. The argument that minimizes this metric is taken as the estimate of the number of occupants. The approach studies the achievable accuracy in an outdoor and an indoor site with up to nine persons for the cases of using 802.11 b/g equipment with either directional or omnidirectional antennas. RSSI measurements are collected for 300 s. In the outdoor site, the estimation error does never exceed two persons when using directional antennas and does occasionally exceed two persons when applying omnidirectional antennas. The indoor environment experiences stronger multipath effects due to the reflection from static objects. Thus, using omnidirectional antennas decreases the accuracy in indoor environments, with the system being able to operate with an estimation error of two persons or less only 63% of the time; Depatla et al. show that introducing directional antennas is able to improve performance. However a solution consisting of directional antennas presents a number of practical limitations due to the fact that additional hardware installations may be required. As the solution is only tested with a maximum of nine participants, its scalability is not assured.

In [33] Li et al. propose a smartphone-based people counting system, Wi-Counter, that leverages existing Wi-Fi infrastructure in the form of WLAN APs. The system operates in three phases. In the first phase, crowdsourcing, a smartphone-based application collects Wi-Fi RSSI values from the WLAN APs in a student laboratory room every 20 s. The current number of the people in the room is entered manually by the operator of the mobile device. In the second phase, offline training, the noisy crowdsourced data is smoothed via a Wiener filter and then used to train a five-layer neural network. Finally, the third phase, online people counting, uses the trained neural network to estimate the number of people (up to 50) at a given location and achieves an accuracy of up to 93%. Wi-Counter is demonstrated to have better performance in terms of estimation accuracy than the linear approach in [29] and the sequential counting approach in [31].

In [26], Fadhlullah and Ismail develop a prototype of a crowd density estimation system based on statistical analysis and RSSI measurements. Certain members of the crowd carry one or more sensor tags, equipped with a ZigBee chipset and microcontrollers. Data is processed externally on a coordinating node. The authors first apply a statistical analysis to evaluate the significance of mobile and stationary crowds on the signal attenuation, and conclude that the effect is statistically the same. However, signal attenuation is shown to be affected by crowd size, number of used sensor tags, and crowd patterns (scattered or lumped). The exact dependencies are obtained with the help of a Design of Experiment analysis and then are used to train the crowd density estimation algorithm, which classifies the measurement into low density crowds (five people) and medium density crowds (10 or 15 people). For all experiments the area under consideration is 5 m $\times$ 35 m. With the help of a signal path loss propagation model, a prediction model is used to estimate the actual crowd size from the classification. It can then be applied in the testing phase to estimate the crowd size from the measured RSSI values, which are reported by the tags to the coordinator. A low crowd density (five people) can be detected with an accuracy of 75% while the accuracy for a medium crowd density (10 or 15 people) is 70%.

Pros and cons: There is a wide variety of device-free works based on RSSI measurements. These measurements often are obtained from available WLAN infrastructure, so there is no need for installing additional hardware. However, while this may be true for certain types of indoor environments, such as offices or campus laboratories, in outdoor scenarios this assumption may not be valid. Moreover, measurements are distorted due to multipath effects and may require preprocessing to infer an occupancy estimate with a reasonable accuracy. Furthermore, the requirement for extensive data collection and training before the actual online people counting phase may be a hindrance when implementing such a system in multiple locations. Finally, most works except [21,32] do not investigate the scalability of this method. This is partially due to the fact that it is not trivial to run large-scale experiments both in terms of recruiting participants for the experiment, but also in terms of hardware setup, and system training. Moreover, simulations may not be reliable, as simulation models need to take into account propagation models and signal attenuation by human bodies, as well as potential realistic human mobility models. To conclude, crowd counting methods based on RSSI measurements have limited applicability in real-life scenarios with respect to exact occupancy measurements.

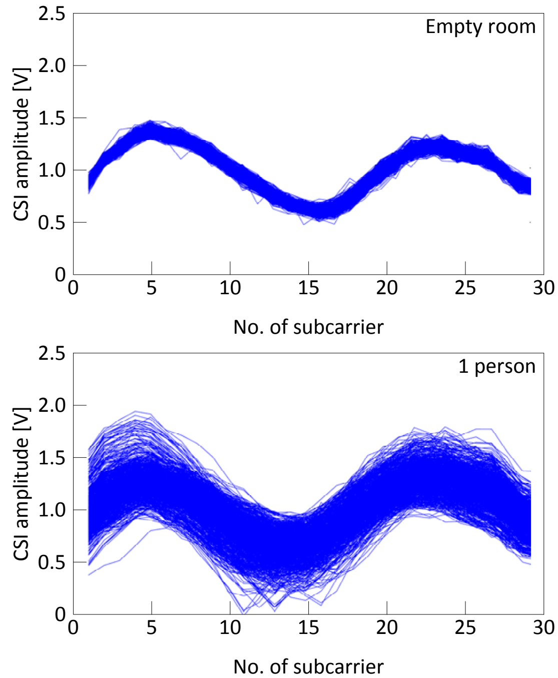

Fig. 4. CSI amplitude values of several measurements for a set of 30 subcarriers, which have been obtained from a packet received by an OFDM-based 802.11n device in [24]. For an empty room, low CSI variations can be observed while one moving person introduces higher multipath variations, which lead to higher CSI variations.

2) Channel state information based solutions: Recent approaches conduct people counting based on Channel State Information (CSI). While the general procedure is similar to RSSI-based approaches (see Section III-A1), CSI is a fine-grained value derived from the physical layers as opposed to RSSI values. It describes how a signal propagates from the transmitter to the receiver and contains, among other things, information on the joint effect of scattering, fading, and power decay with distance. Hence, CSI reflects environmental variance resulting from moving subjects more precisely than RSSI. Figure 4 shows several measurements of CSI amplitude values for 30 subcarriers, which have been obtained from a packet received by an OFDM-based 802.11n device in [24]. In an empty room, CSI variations are low and solely result from thermal noise and interference. However, as soon as people start to move in the room, higher multipath variations are introduced as a consequence of a higher number of moving scatterers leading to higher CSI variations. These variations are then exploited in CSI-based people counting solutions to infer the correct number of occupants in a given space. For more information, see [24].

In [22], Xi et al. propose a device-free crowd counting system based on a monotonic relationship between CSI variations and the number of moving people in a specific area. To indicate the crowd size, the authors introduce a metric that adapts to the changes in CSI and increases with the increase of subjects in the area; however this increase is only permitted to a certain threshold, which limits the scalability of the proposed solution. By applying a probabilistic model,

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/COMST.2019.2902824, IEEE Communications Surveys & Tutorials

the number of people can be directly estimated from the introduced metric without fingerprinting. The operation of the proposed system passes through a training and a monitoring phase. In the training phase, the system learns CSI values when no or a few people are inside the area of interest. In the monitoring phase, the system collects CSI measurements from monitoring points, estimates the number of people, and updates the training profile. The test environment consists of four laptops with Intel 5300 NICs, with one of them broadcasting beacon messages and the other three receiving them. The solution is suitable for counting up to 30 moving people with an estimation error of less than two persons in 98% (indoor) to 70% (outdoor) of all cases. Interestingly and in contrast to RSSI-based solutions, multipath effects, which have a higher impact indoors than outdoors, actually help to increase the accuracy of the occupancy estimate. This is due the fact that CSI is much more sensitive to the change of the environment than RSSI since it can reveal the effect of scattering, fading, and power decay with distance more precisely. Moving people between transmitter and receiver lead to distinct CSI amplitude variations on each sub-carrier in the frequency domain, from which occupancy can be inferred.

In [24], Di Domenico et al. introduce a people counting method based on differential measurements of CSI. As opposed to previous works in the field, their method does not require a training phase exactly in the same environment where the system is also tested as differential measurements are less sensitive to environment characteristics than absolute measurements. The proposed people counting system is based on the measurement of how the CSI varies over time, which mainly depends on the number of people moving in the environment rather than the characteristic of the environment itself. The system solely uses one single Wi-Fi transmitter with two antennas and one single receiver with three antennas leading to six available RF channels. The used firmware provides 30 CSI values (one value for each of the 30 subcarriers supported by the firmware) for each received packet and each channel, i.e., 180 CSI values per measurement are obtained. The authors repeat the measurement multiple times over time and show that the more persons in the environment, the more the CSI measurements differ from each other. The difference between two CSI measurements is expressed with the help of the Euclidean distance. To further reduce the dependence on a specific environment, CSI measurements of a time sample and the Euclidean distance between two CSI measurements are normalized. The distances of the six different channels are then combined. Subsequently, a statistical analysis of the combined distances, determined from packets received during a sliding window of 10 s, is performed to compute four different features. These features are calculated from first and second order statistical moments using both the distance between consecutive CSI measurements and between all pairs of CSI measurements within the sliding window. In order to select the combination of these features with the best classification accuracy, the authors apply machine learning. Finally, the number of people can be estimated using a linear classifier. An experimental evaluation shows that the system can count up to seven people in indoor environments with an error less than or equal to two subjects 81% of the time in a large room (6 m $\times$ 12.5 m) to 91% in a small room (5 m $\times$ 6 m).

Pros and cons: CSI-based approaches are relatively new as compared to RSSI-based people counting solutions, and as such have not been explored to their fullest extent. However, initial works reveal that CSI may be a better candidate than RSSI for performing people counting in indoor environments. Similar to RSSI-based approaches, CSI-based approaches often rely on WLAN infrastructure that is already installed, thus no further installations are required if these solutions are to be deployed in real-life scenarios. As CSI is more sensitive to environmental variances than RSSI, CSI-based solutions are considered to be more suitable for environments with high levels of mobility. However, it is likely that counting immobile people in a room is more difficult than if they move. This is because non-moving people cause less variations in CSI values, and thus their number can be less clearly inferred. In addition, if people change their position in the test phase compared to the training phase the variations make it even less clear to conclude on the number of people. Furthermore, the scalability of the method has not been investigated, and based only on current results it is not possible to conclude whether CSI-based solutions can be deployed in larger areas or in dense scenarios.

3) Ultra-wide band signal based solutions: Ultra-wide band (UWB) signal based solutions conduct people counting based on the principle of multi-target detection via radar networks [61,62]. Thereby, an extremely wide bandwidth impulse signal is transmitted, several signals are backscattered by targets, and received to detect targets within the radar range. Due to the inverse relationship between bandwidth and duration of a signal, UWB systems are characterized by very short duration pulses with a low duty cycle, typically on the order of one nanosecond. That is, the ratio between the pulse transmission timing and the average time between two consecutive transmissions is usually kept small. Such a pulse-based UWB signaling scheme is referred to as impulse radio (IR) UWB [63,64]. In IR UWB systems, received pulse signals are first converted to digital signals and, if necessary, filtered by a digital bandpass filter. The received signals can be assigned to a distance index according to their delay, as illustrated in Figure 5. This distance index corresponds to the distance of the objects, from which the signals were reflected, to the receiver. Since the received signals are reflected by every object in the environment, i.e. a person as well as a floor or ceiling, unwanted signals eventually need to be removed by background substraction methods [39]. Finally, humans can be counted by either trying to detect the signal for each individual human in the waveform or to deduce the number based on the pattern of the waveform of the received signals. More details are given in [39].

Choi et al. propose in [34] a IR UWB radar system, which transmits a wide bandwidth impulse signal and receives the backscattered signals to infer the number of targets from it.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/COMST.2019.2902824, IEEE Communications Surveys & Tutorials

TABLE II
SUMMARY OF ADVANTAGES, DISADVANTAGES AND APPLICABILITY FOR DIFFERENT CATEGORIES OF DEVICE-FREE PEOPLE COUNTING TECHNIQUES.

Category	Advantages	Disadvantages	Applicability
RSSI-based [21, 26]–[33]	Utilizes existing WLAN infrastructure, no further installations needed.	Low quality measurements so that processing of them is necessary.	Indoor and outdoor environments
CSI-based [22,24]	Utilizes existing WLAN infrastructure, no further installations needed. More accurate than RSSI.	Scalability not investigated.	Indoor and outdoor environments
UWB-based [34]–[39]	More accurate than RSSI and CSI.	Requires installation of additional infrastructure. Scalability has solely been investigated indoors.	Indoor and outdoor environments
PIR-based [40]–[43]	Proven solution for counting people in public transport systems.	Utilization in outdoor environment infeasible.	Indoor environments
LED-based [44, 45]	Achieves reasonable accuracy for a small number of people.	Requires availability of LED lighting infrastructure. Utilization in outdoor environment infeasible.	Indoor environments
Acoustic-based [20,46,47]	Achieves reasonable accuracy for a small number of people.	Requires installation of additional infrastructure.	Indoor environments
CO_2-based [48]–[53]	Coarse-grained estimates due to low accuracy might be useful for HVAC systems. CO_2 sensors are often part of air conditioning systems.	Reacts slowly to dynamics. Utilization in outdoor environment infeasible.	Indoor environments
LiDAR [54]–[56]	Achieves high accuracy.	Requires installation of additional infrastructure. Rather suitable for object identification, which might pose privacy concerns.	Outdoor environments
Hybrid [57]–[60]	Combination of sensors can increase accuracy.	Unclear if the increase of accuracy justifies the increased costs for the installation of multiple sensors.	Indoor environments

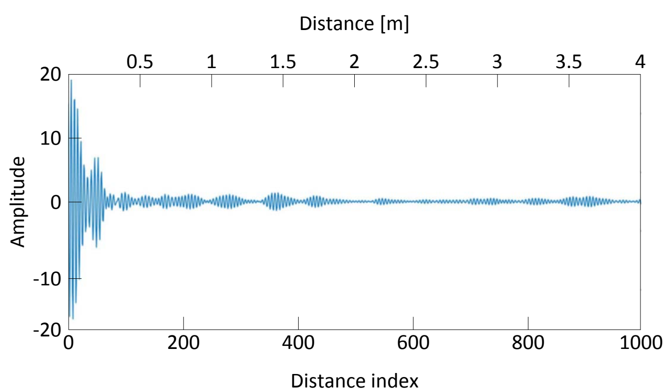

Fig. 5. Example of backscattered signals received by an IR UWB receiver after their digitization by means of an A/D converter [39]. The amplitude of the waveform is depicted over the distance index, i.e., the distance between an object and a receiver.

The counting algorithm performs an iterative search for the local maximum of the power profile of the received signals. As soon as this local maximum exceeds a threshold value, the algorithm assumes that a target has been detected and then deletes a set of samples within a window around the local maximum. Experiments in an 6 m × 6 m indoor environment were carried out whereby up to three targets moved randomly. The error rates is up to 8% when detecting three targets.

The authors of [34] extend their work in [37] and introduce a people counting system based on IR UWB radar sensors in indoor environments such as a subway station. A system module consists of two radar sensors spaced by a certain distance and is equipped with antennas, which transmit impulse signals and receive multiple reflected signals from multiple humans and objects. The correlation between the signals of both radar sensors are used to generate numerical indicators by means of segmentation, data reduction, and selecting representative values from intervals. The people count can then be inferred from the numerical indicators. During the evaluation, measurements are taken for two minutes by four modules immediately after the subway arrived. The first and second modules are installed at the entrance of the stairway to the subway platform just after passing through the ticket gate in the direction of the up line and the third and fourth modules are in the down line direction. Traffic card data that people tag as they pass the ticket gate is recorded as benchmark. Data is collected over seven days whereby up to 98 people for two minutes and up to 8881 passengers for one day could be counted with an error of less than 10%. Errors are especially caused by the fact that the counting module has difficulties to count the people who are stuck on the stairs because of the high number of people coming up from the subway at the same time during rush hours and the bottleneck due to the ticket gate.

Quan et al. introduce a method to count occupants entering and exiting an area by utilizing a UWB radar at the area's entrance [35]. It is stated that two people can be counted when people enter or leave the room by passing through a gate field, in which they change the state. A state change is detected by threshold crossing. No information on obtained results is given in this paper.

Another IR UWB system for people counting is proposed in [36]. A low-power pulsed radar is utilized to transmit impulse signals and then to infer occupancy from received waveforms within the 10 m range of the radar in indoor room settings. In a training phase, a regression map is first learned using SVM for nonlinear regression to establish a relationship between people count and time and frequency domain features extracted from the received waveforms. In the test phase, 750 minutes of radar data, obtained from multiple radars, with known people count is collected in four rooms with the number of people ranging from 0 to 43. Estimate numbers exhibit an error of 2.17 people, by which they deviate from the ground truth.

Bartoletti et al. propose an IR UWB based counting system, which relies on energy detection (ED) [38]. The paper proposes a low-complexity technique based on energy samples extracted from the wireless waveforms received after target backscattering. A tractable theoretical model is established for estimating the target number from the energy samples. It has a complexity that is independent of the number of targets and is therefore supposed to scale. They compare the ED approach to three other algorithms from [34] and [36]. For up to 30 people randomly located in a 10 m $\times$ 10 m indoor environment, their algorithm is shown to be superior in terms of accuracy with an counting error of below two persons in 88% of cases.

In [39], a people counting algorithm using an IR UWB radar sensor is proposed. Instead of detecting the reflected signal of each individual human among the received signals, and finally to conclude on their number, the course of the signal amplitudes is examined. During the test phase, the course of the received signal amplitudes is compared with a previously learned probability distribution function of the amplitude values to deduce the number of people within range of the radar. The authors claim that even in densely populated indoor environments with many unwanted received signal reflections, the number of people can be estimated with high accuracy; even if many people are equidistant from the radar. Experiments are performed within the 5 m $\times$ 5 m radius of an IR UWB radar located in a room as well as in a 2 m $\times$ 1.7 m $\times$ 2.4 m elevator lined with metal. The first measurements are used as training data before real-time operation begins. The maximum number of people in the room is 10; in the elevator up to nine people are located. Despite unwanted reflections from objects in the room and the metal in the elevator, the estimation error is on average one person. However, the results show that UWB-based solutions achieve an accuracy and scalability as low as RSSI and CSI-based approaches if many people are standing in close space, which leads to many unwanted signal reflections and therefore the pattern of the course of all signals has to be investigated. Thus, UWB-based solutions are better suited for more open environments with fewer objects, which cause unwanted signal reflections, so that the reflected signal can be detected for each human to be able to infer occupancy.

Pros and cons: Sensor radars based on UWB signals are more suitable for people counting than RSSI and CSI based methods since from the signal reflections the number of people can directly be inferred as long as not too many people are next to each other in close space. However, the installation of sensors radars is necessary and the scalability of the approach remains mostly unclear, especially outdoors. Even though the authors of [38] claim that their approach is independent of the people to be counted, they only count up to 30 people.

B. Sensor-based solutions

Sensor-based solutions can be further categorized into the following categories with respect to the type of sensors used for collecting data for occupancy estimation: (1) passive infrared (PIR) sensors; (2) light-emitting diode (LED) sensors; (3) acoustic sensors; (4) carbon dioxide (CO_2) sensors; (5) light detection and ranging (LiDAR) sensors; as well as (6) hybrid solutions utilizing more than one of the sensors above. The advantages and disadvantages per category are summarized in Table II.

1) Passive infrared sensor based solutions: Passive infrared (PIR) sensors are simple motion sensors that are usually used to deduce movements in the sensors' field of view from changing temperatures caused by moving object. When an object such as a person passes at a point in the sensor's field of view, the temperature changes, which triggers the sensor. Thus, in PIR-based solutions sensors are mounted at a specific position, such as at entry and exit locations of a room or building, to infer occupancy. Although PIR sensors can detect movements in their field of view very well, the accuracy is subject to some limitations. Thus, after the detection of a movement, a certain amount of time passes, called the masking time, in which no further movement can be detected. The work of Raykov et al. provides more details about this technique [42].

In [40], Wahl et al. propose a method for people counting in indoor environments motivated by one of the major building automation problems, namely the lack of information about utilization of indoor spaces that leads to non-optimal energy usage. To provide accurate people count, the authors suggest the installation of a network of distributed passive infrared sensors strategically placed at passages. Each passage is equipped with a pair of passive infrared sensors, an inward-facing one and an outward-facing one, equipped with a wireless communication unit in order to facilitate estimating movement directionality. The authors present and compare the performance of two algorithms: a simplistic direction-based people counting and a more complex probabilistic distance-based algorithm. They implement an infrared sensor prototype to validate their approach on a small-scale testbed encompassing a single office space, and then simulate both algorithms on a larger floor-wide setting with up to six persons. The results show that the probabilistic distance-based algorithm outperforms the direction-based one across all configurations as it compensates for errors introduced due to infrared masking effects. During masking time after registering a movement, sensors are unable to detect further movements, which could lead to wrongful counting if multiple people pass through a passage in a short amount of time.

In [43], PIR sensors are used to track and identify persons in partially observed environments, in which sensors are installed only in certain areas but do not cover the whole space. To address this problem, the authors apply probabilistic plan recognition based on action libraries. These libraries contain existing prior knowledge about the structure of human

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/COMST.2019.2902824, IEEE Communications Surveys & Tutorials

behavior and its environmental dependencies. Training data is required to a lesser extent and it is possible to abstract from scenario specifics. The authors conduct experiments in typical office building with up to five persons moving around in nine rooms and a corridor. The corridor of this building is equipped with PIR sensors. Moreover, data is recorded from light switches in some rooms as well. However, using this data in addition to PIR sensor data only has marginal influence on the results. Interestingly, the authors mention disaster management in buildings as related application scenario. They compare recognition results obtained when using a machine-learned model and a model based on prior knowledge about sensors or movement speed. They state that up to nine rooms need to be checked to find a certain person with 100% probability when four other person are involved in the setting. Then, the authors present the accuracy when estimating the number of persons per room. Generally, the mean squared error of estimating the number persons (up to five) is lower in case of using the machine learning approach and equals 0.18 for five persons.

Suthokumar et al. propose a sensor-based wireless distributed system for people counting and activity detection in [41]. They conduct people counting by means of a PIR sensor mounted in a laboratory. If its digital output is low, some kind of motion is assumed and a person is counted. On average, the accuracy amounts to 89% on average for up to 39 people.

Raykov et al. introduce an approach for occupancy estimation from the measurements of a single PIR sensor [42]. Raw training data obtained from the PIR sensor is clustered by using the infinite hidden Markov model machine learning method to extract motion patterns. The authors use statistical regression methods to infer the people count from these motion patterns. The estimate can be updated every 30 seconds to reflect changes. They test their system in more than 50 office meetings in seven different conference rooms of an office building with a data acquisition board including a PIR sensor placed on a table positioned approximately in the middle of the room next to the wall. The height of the table varies from room to room. Their estimates deviate by ± 1 individual from the ground truth for at most 14 occupants. The results do not appear to be affected by the height of the table, on which the data acquisition board was positioned, but the placement of the sensor on a particular side of the room caused occupants to conceal each other during larger meetings (typically with eight or more people). This problem can be solved by testing different positions of the PIR sensor; a promising start would be the ceiling of the room. This will likely result in a consistent improvement in accuracy due to the clear, unobstructed view of all occupants.

Pros and cons: PIR-based systems provide a promising solution for people counting in indoor environments by passively monitoring temperature changes without the need to actively send radio signals. Simple low-cost PIR sensors can be readily obtained for less than \$10 and are easy to deploy and maintain. In addition, PIR sensors are already being used in commercial buildings for various tasks related to human motion detection such as controlling light switches and triggering burglar alarms [42]. Even though academic works have not investigated the scalability of PIR-based systems, the industrial product IRMA [65] has shown its ability of being able to count large crowds in public transport systems with an accuracy of at least 95% (cf. Section VI). In this regard, the industrial solutions are ahead academic approaches, however, the academic work in [43] looks at disaster management as application scenario, which none of other cited works does. The disadvantage of PIR-based systems is that a realization in open outdoor environments is infeasible due to short transmission ranges. Disturbances caused by sunlight and heating also affect the precision of PIR-based systems. Moreover, PIR sensors suffer from further limitation: stationary occupants are often not detected since they do not trigger the PIR sensor. In addition, the positioning of the PIR sensors plays an important role in the achievable accuracy as this is determined by the field of view of the sensors. Only in the field of view, the sensors are able to detect movements and hence can count people only there.

2) LED-based solutions: Lighting in general can be assumed to be available in any human occupied indoor space. Illumination sources such as LED luminaries emit visible light into the indoor space. Visible light can be detected using simple and inexpensive photodiodes or solar cells, which requires minimal processing effort on the detection device. Thus, visible light provides a significant advantage for sensor applications over media such as radio frequency signals, which require complex processing, and radios operating in a licensed spectrum. In an indoor space, the emitted light is reflected. This diffuse light reflection, which mainly results from reflections from the floor and furniture, is perturbed by human presence in this space. These perturbations can be sensed to infer occupancy. A more detailed description of this technique is available in [45].

In [44], Wang et al. propose a framework for estimating occupancy distribution in a room using color-controllable LED lights (the same lights that simultaneously provide illumination) and sparsely distributed non-imaging color sensors. By modulating randomly generated pertubation patterns onto the driver signals of the LED lights and measuring the changes in the color sensor responses, a light transport model for the room is established. Two approaches are proposed to estimate the spatial distribution of occupancy based on a light blocking model and a light reflection model, respectively. These two approaches, which can be combined, estimate the occupancy of the room while preserving the privacy of its occupants. For the experimental evaluation, in a 2.2-m-wide, 3.4-m-long and 2.2-m-high room, twelve color-controllable LED lights are fixed to the ceiling and twelve wireless optical light sensors are mounted either on the walls or on the ceiling. The room is divided into six regions and in nine scenarios one or two regions are occupied by people and furniture. The accuracy is only good enough for the light control module to determine which part of the room is occupied and which type of light to deliver. The proposed system is thus applicable only to the control of the LED lights of an intelligent lighting system for the purpose of energy efficiency, productivity, and human

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/COMST.2019.2902824, IEEE Communications Surveys & Tutorials

comfort.

In [45] Yang et al. present CeilingSee, an occupancy inference system that builds upon existing LED lightning systems. One of the main advantages of such a system is that it does not require user involvement. The main idea behind CeilingSee is that ceiling-mounted LED luminaires are converted to act as sensors, sensing the variences in diffused reflection that is caused by the number of occupants in a given indoor area. Sensing information is gathered by multiple LED sensors over time, whereby sensing information obtained during a specific point in time is referred to as snapshot. During an offline training phase, for input snapshots the corresponding occupant counts are learned. Periodically, the system is adapted to environmental dynamics online (normally every week). From the findings of the training phase, the authors derive a function to fit the relationship between snapshots and occupant counts by formulating the learning problem as a regularized regression problem, specifically by means of support vector regression (SVR). CeilingSee is deployed in a small room-size testbed, and it is shown to have an accuracy of occupancy estimation of at least 90% when the miscount tolerance level is set to one person. The authors show that the system performs better when the participants are static in the room (and located right below each sensor installation) than when they are roaming around the room.

Pros and cons: LED-based approaches have hardly been studied so far and the only visible work that gives precise information on the accuracy shows that they can achieve a reasonable accuracy of at least 90% in indoor environment for a small number of people. However, they depend on available sensing infrastructure. Their scalability has not been investigated so far.

3) Acoustic-based solutions: Basically, acoustic-based solutions transmit ultrasound waves and receive the waves reflected by occupants. When the reverbation of the transmitted waves is received, from its properties such as the receive time or the the signal decay the number of occupants is inferred. One basic idea of this technique is that the more people are in the room, the faster the signal decays, and thus the reverberation time can be used as a feature to measure occupancy. Another idea is to measure the energy of the acoustic signal over time to conclude on the people count. For more information about this technique, we refer the reader to [47].

One of the first works on acoustic-based device-free people counting is [20] in which Chen et al. introduce an approach for counting moving objects using ultrasonic sensor networks (MOCUS). Multiple sensor clusters, each consisting of one ultrasound transmitting node and two receiving nodes, work together and form a wireless sensor network. By applying a temporal correlation algorithm for intra-cluster analysis and an inter-cluster cooperation algorithm, a counting estimate can be computed. MOCUS is evaluated experimentally on a small-scale testbed consisting of three sensor clusters mounted on a corridor ceiling. We note that although the authors claim that MOCUS can be utilized also in open environments, they do not discuss sensor placement in such scenarios. Scalability in terms of counting large groups of simultaneously moving subjects is also not discussed. The authors only evaluate the system with one or two moving subjects at a time, reporting accuracy of at most 90% for long periods of observation; however individual detection accuracy is often as low as 50%. Thus, in its current state MOCUS is mostly suitable for moving object detection in limited spaces, such as entry points or corridors.

In [46], Shih et al. present an active sensing technique that makes use of the change in acoustic properties in a room in order to estimate the number of occupants. The system consists of an omni-directional ultrasonic treble speaker, also referred to as tweeter, with a co-located microphone; the tweeter first transmits an ultrasonic chirp (a sinusoidal signal that linearly increases in frequency), and then measures the response over time as the signal decays. The main idea behind this technique is that the more people there are in the room, the more rapidly the signal decays and thus the reverberation time can be used as a feature to measure occupancy. The occupancy estimation algorithm consists of two parts: (1) categorizing collected data into clusters, and (2) building a regression model based on two training points and the clustering result. The authors evaluate their proposed algorithm in three different settings: a small conference room, a medium-sized classroom and an auditorium of capacity 150 people. The experimental results with up to 50 people show that clustering performs well as long as the occupancy does not reach the maximum occupancy of the space. Moreover, the larger the space in which measurements are taken, the larger the estimation error of the algorithm.

In [47], Huang et al. utilize audio processing techniques such as speaker recognition and background audio energy estimation to determine the indoor occupancy in the context of smart buildings. The estimation is based on audio recordings through microphones under the assumption that the indoor sound is predominantly human speech and the sound level from neighboring rooms or outdoors is negligible. The authors investigate two scenarios: when people speak separately, and when several persons speak simultaneously. In the first scenario, people are simply classified according to their speech characteristics, and the total number of people in the room is estimated as a sum across all identified speakers. To estimate the number of people in a given location in the second scenario, the authors use the short-time energy (STE) feature of speech; STE describes the signal energy over an interval of time. STE is dependent on the size of the room, the position of the microphone as well as the position of participants with respect to the microphone. However, if all parameters are known, occupancy can be estimated. Although the authors achieve estimation accuracy of 94% for 80 speakers in a room (via simulations), the requirements for STE may prevent large deployments, as it is often unrealistic to be aware of the position of occupants with respect to the microphone. Furthermore, STE assumes that all participants in the room are speakers in order to estimate their count; thus silent occupants

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/COMST.2019.2902824, IEEE Communications Surveys & Tutorials

may further reduce the accuracy of the approach. Finally, identifying speakers may raise privacy concerns.

Pros and cons: Acoustic-based solutions are shown to perform best for small indoor spaces and small crowds. Scalability is however limited as accuracy decreases significantly both with increase in space and in occupancy. Moreover, in most experimental and simulation studies the effect of objects present in the space is not well-studied. However in real-life deployments, sound wave absorption in objects may distort the estimate. For instance, seats in an auditorium are sometimes designed to absorb as much sound as a person. Hence, the acoustics may be invariant to the occupancy level. Finally, additional sensing infrastructure often needs to be installed.

4) Carbon dioxide based solutions: Carbon dioxide based solutions utilize CO_2 sensors to estimate the occupancy in an indoor environment from the CO_2 concentration at one point in the room. Since every person in a room produces CO_2, the CO_2 concentration can be used to deduce the number of people. However, many other factors play a role in the attainable accuracy, as, for example, venting mechanisms continuously reduce the CO_2 concentration and thus these aspects must be taken into account. CO_2 sensors are often part of the room air-conditioning infrastructure and thus obviate the installation of additional hardware. More information is given in [50].

One of the earliest studies on estimating actual occupancy in indoor spaces is presented in [48] by Wang and Jin in 1998. It works by measuring CO_2 concentration of the return air, used by return air vents to maintain the air pressure in a room and filter out debris, and the flow rate of the outdoor air, provided for ventilation in order to maintain acceptable indoor air quality. The authors formulate analytically three occupancy detection methods: a steady-state detection and dynamic detection, using an approximate or an exact solution respectively. The methods are then validated via simulations of three different occupancy scenarios consisting of 10 to 80 people. Results show that across all scenarios both dynamic methods exhibit quick response with changes in the occupancy level and achieve at most 5% deviation from the ground truth. On the other hand, the steady-state approach is slow and deviates as much as 25% from the ground truth. The authors further use these findings to design a demand-controlled ventilation system.

Following Wang and Jin's findings, in 2004 Mumma presents an experimental testbed for estimating real-time occupancy in indoor spaces through CO_2 emissions in [49]. The author assumes a constant CO_2 generation rate per person, and relies on measurements of CO_2 concentration in the space and the outdoor air, as well as the outdoor air flow rate. The proposed method uses both steady-state and transient equations to solve for occupancy, with the latter causing oscillations in the estimated occupancy, and thus requiring a special damping to compensate for these fluctuations. The method achieves an occupancy estimate, which agrees with the actual occupancy count of up to 25 people within two occupants and hence achieves an accuracy of up to 92%.

In [50] Sun et al. present an in-situ implementation of an adaptive demand-controlled ventilation strategy which aims to optimize the air flow rate with the help of a CO_2-based dynamic occupancy estimation. The authors derive an analytic model for estimating occupancy online and then adjust on demand the outdoor air flows with respect to the estimate; the analytic model consists of a system of differential equations that take as input values CO_2 and air flow measurements and produce results for the whole indoor space as well as for predefined zones within the space. For testing purposes the premises of a single floor is instrumented with a total of 22 CO_2 sensors, as well as air flow meters which are further calibrated to increase measurement accuracy. Test data from a typical working day are analyzed, and the predicted occupancy estimation is shown to agree well with the actual occupancy on the floor, determined with the help of central CCTV. The results show that with the help of CO_2 sensors the authors are able to simultaneously detect up to 80 occupants on the floor. However, as occupancy estimation is not the main goal of the work, details on achievable accuracy are not provided. Although the approach seems to scale well with the number of occupants, the relationship between maximum number of occupants and number and positioning of CO_2 sensors needs further investigation.

In [51] Jin et al. describe an occupancy detection algorithm using indoor CO_2 concentration based on the sensing by proxy. A link model is proposed based on equations that capture the spatial and temporal features of the system and link unobserved human emissions to proxy measurements of CO_2 concentrations. Both controlled and field experiments are carried out in a conference room equipped with a full ventilation system including an air return vent and an air supply vent. CO_2 sensors are installed at both vents. In addition to field tests, simulations are further performed. The authors show that estimating the number of occupants in the room based on CO_2 measurements at the air return and air supply vents by sensing by proxy outperforms a range of machine learning algorithms. Moreover, the authors show the robustness of the proposed method towards external influences such as emission rate and physique of occupants, as well as noise added to the measurements by ventilation rate or activities within the space. However the maximum number of occupants across all experiments does not exceed seven persons. Thus, the scalability of the proposed algorithm requires further investigation.

In [52] Basu et al. present PerCCS, a person-counting algorithm based on CO_2 emissions in indoor spaces. PerCCS works in two stages. In the first stage, it denoises the CO_2 data and learns a low-dimensional representation of it. In the second stage, this denoised data is used as a predictor variable for estimating occupancy via ensemble least square regression. Thereby, several least square regression models are trained to map occupancy as a linear function of the low dimensional representation of CO_2. The PerCCS algorithm is tested in a classroom of maximum occupancy of 42 people; data was collected in the course of 13 days (out of which data from

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/COMST.2019.2902824, IEEE Communications Surveys & Tutorials

the first nine days was used for training). Predictions are made with a latency of 15 minutes, which is due to the slow CO_2 dynamics. Results show that zero-occupancy is detected in 91% of all cases, while the actual number of people in the room is detected in 15% of all cases. Depending on the application at hand (e.g., efficient heating-cooling operations) these values may be sufficient; in other cases (e.g., emergency evacuation) other approaches which are able to achieve higher occupancy estimation accuracy may be preferred.

In [53], Arief-Ang et al. present the Carbon Dioxide-Human Occupancy Counter (CD-HOC) system, which estimates the number of people within closed spaces from a single CO_2 sensor. CO_2 data is factorized by means of seasonal-trend decomposition based on moving average or the non-parametric least square regression called Loess. Then, different regression algorithms are used to predict the occupancy. The trained models are run in an academic staff room with up to four persons and a cinema theater with up to 300 people. The achieved accuracy amounts to 94% for the staff room and 77% for the cinema, which is still better than the baseline method based on SVR.

Pros and cons: CO_2-based solutions achieve a low accuracy for a high number of people but might be useful in cases where a coarse-grained estimate of the number of people is enough (e.g., for the control of HVAC systems better than solely turning it on or off). CO_2 sensors are often part of indoor air conditioning systems and hence no additional sensing infrastructure needs to be installed. This approach however reacts slowly to dynamics as the CO_2 concentration needs some time to change if occupants enter or leave the room. If people that have entered leave the room so quickly that the CO_2 concentration in the room barely changes, their number may not be considered in the occupancy estimate, which could be important if the time course of the estimate is relevant. Moreover, depending on the ventilation, CO_2 concentration may differ from room to room and it might be impossible to find a generalized model for inferring occupancy from measured CO_2 levels.

5) Light detection and ranging (LiDAR): LiDAR-based systems use laser scanners to capture images with distance information, in which objects can be detected. LiDAR-based systems are less sensitive to varying lighting conditions and require less data processing time compared to image-based systems. Systems that can be found in the literature first and foremost aim at tracking, i.e., estimating the motion of objects including their distances and speeds from laser-scanned images. Such systems are characterized by high resolution images and thus fall into the category of image-based solutions. However, LiDAR-based systems may also capture images with low or medium resolution. In this case, their characteristics allow for performing people counting by simply counting the occupied areas in the laser-scanned image, and can therefore be considered as non-image based solution for occupancy estimation. In what follows, we only aim at describing the functionality that a system would require in order to count people. The work of Kanaki et al. [56] provides a good overview of this technique.

In [55], a group of mobile robots finds moving objects such as people, cars, and bicycles in its own laser-scanned images using a binarized occupancy-grid-based method. The laser image is divided into a grid consisting of cells and moving objects are put into the corresponding cell where it is detected. Such a cell is called occupied. The robots send their measurements to a central server and occupancy can be estimated. The accuracy of the approach depends on both the number of robots used for object tracking and the number and position of the object to be tracked. The authors conduct an experiment with 270 scans and a total duration of 27 s, in which two robots track a motorcycle, a bicycle, and two persons. Cooperative tracking using two robots provides better tracking accuracy compared to the case if only one robot tracks moving objects.

Hashimoto et al. introduce a method for laser-based tracking of people in a group [54]. In sparse environments, laser-based tracking achieves good accuracy. However, its accuracy decreases in cluttered or crowded environments, in which many occlusions occur. Therefore, the proposed method groups people with similar motions to improve tracking performance. Two laser scanners set at different heights are used to detect people's heads and knees, respectively. The authors conduct experiments, in which they track 11 people in a 8 m $\times$ 5 m hallway by means of the laser measurements. Although the tracking of up to three persons temporarily fails during the experiments, this does not necessarily mean that the number of counted people is incorrect. However, no further information on the accuracy of the current people count is given.

The author further extend their work in [56] by using multi-layer laser scanners (MLSs), which provide richer information than previously used single-layer scanners and thereby enable an accurate recognition of the surrounding environment. In the experimental system, two mobile sensor nodes equipped with a wheel encoder to measure the wheel's velocity and gyro sensor as well as MLSs and GPS sense their surrounding environment. They detect moving objects with binarized occupancy-grid-based method like in [55] and upload this information to a central server. Already now, moving objects can be counted. In the experiments, a car and two pedestriants were tracked. By cooperation of the two mobile nodes, the authors state that the tracking accuracy increase compared to solely using a single mobile node for tracking. No further detailed results on the accuracy are however given.

Pros and cons: A LiDAR-based system can provide high accuracy but it heavily depends on the infrastructure that scans the objects to be counted. The approach seems more suitable for scenarios, in which it is necessary to determine the object types (pedestrians, bicycle, car, etc.) and their motions. Since this allows objects to be identified, privacy concerns may arise.

6) Hybrid solutions: Oftentimes, buildings accommodate various kinds of sensors or other infrastructure, which can be exploited to infer occupancy estimation. They can comprise sensors for a multitude of environmental parameters such as CO, CO_2, temperature, humidity, acoustics, motion and illumination to mention only a few. The estimate can subsequently be used for human-centered environmental control, security, and energy-efficient and sustainable green buildings. The accuracy of the estimation depends on the combination of used sensors whereby an intelligent combination could increase the precision. We refer the reader to [60] for more details about this technique.

In [58] Lam et al. analyze the correlation between occupancy levels and environmental data captured from three different sensor networks: a wired gas detection sensor network, a wireless ambient sensor network and an independent CO_2 sensor network. The authors deploy a testbed consisting of state-of-the-art IT systems as well as distributed sensors for a variety of environmental parameters such as CO, CO_2, temperature, humidity, acoustics, motion and illumination; the test bed is deployed in an open-plan indoor environment. As part of the testbed, the authors also deploy cameras, and further use the captured images to determine ground truth for the occupancy levels at any time. It is shown that the parameters of CO_2 and acoustic sensors, as well as passive infrared (PIR) sensors, have the largest correlation with the number of occupants in the space. These findings are further used for evaluating three machine learning techniques, namely support vector machine, artificial neural network and hidden Markov model. The results show that a hidden Markov model performs best and is able to realistically capture quick changes in occupancy levels, as well as to maintain a constant level during static occupancy periods. An accuracy of at least 60% for up to three occupants is reported across daily and weekly results and across different areas on the floor of the open-space office building. However, it is unclear whether increasing the number of occupants would lead to similar results.

In [59], Yang et al. propose an occupancy estimation model for demand-driven HVAC operations that takes into account readings from a number of non-intrusive sensors, such as a light sensor, a sound sensor, a motion sensor, a CO_2 sensor, a temperature sensor, a humidity sensor, and a PIR sensor. Sensor data is processed in real time using a neural network in order to estimate the number of occupants in a space at a given point in time. Four tests with up to nine occupants are carried out in two different laboratory settings: the tests are based either on self-estimation (i.e., data from the same laboratory is used for training and validation as well as testing) or on cross-estimation (i.e., data from different laboratories is used for the training and validation and the testing phase). The authors show that self-estimation provides better results in terms of accuracy than cross-estimation (when the error tolerance level is set to one person, self-estimation achieves approximately 86% accuracy as compared to only 64% accuracy in the cross-estimation case). This signals that a number of constraints (such as differences in environmental settings as well as sensor calibration) may become limiting when attempting to devise a universal occupancy estimation model based on multi-sensor readings. Moreover, some of the applied sensors could be used to identify individuals, which might pose privacy concerns.

In [57], Ebadat et al. exploit the statistical correlations between a given occupancy level and a collection of CO_2 concentration, room temperature, and ventilation actuation signals to develop a dynamic model. They formulate an estimation problem as regularized deconvolution problem, for which the estimated occupancy is the input such that the currently measured CO_2 levels are best explained. The model consists of a training phase and a test phase. In the training phase all of the above environmental parameters are measured, together with the observed occupancy levels, and this information is used for deriving an linear time-invariant estimator. In the test phase the occupancy levels are determined based on the measured data and the estimator by solving the inverse deconvolution problem. Both online and offline estimators are proposed whereby the latter is shown to perform better than other data-based building occupancy estimators such as a neural network or a support vector machine. Experimental results for up to four people obtained in a fully instrumented facility at a university campus building show that the proposed model reports the correct occupancy level in more than 88% of the times. As the model dynamics are assumed to be linear, the solution is said to perform well independent of space or occupancy levels. Further investigation is however required to verify this statement.

In [60], Yang et al. investigate whether available sensory information in homes and commercial buildings could be used for inferring binary, range and exact occupancy. The authors explore two different sources of information: motion sensors (PIRs) installed in a three-person 10-room home, and smart electric meters installed in a 12-person university laboratory. Motion sensor data is first pre-processed and then processed with machine-learning techniques (conditional hidden Markov model, conditional random fields, hidden-Markov support vector machine) or rule-based methods, which do not require any training data. Data from smart electric meters is pre-processed, features are extracted, and finally classifiers, which have been trained before, are utilized to derive binary or ranged occupancy. The results demonstrate that based solely on the motion sensor data, occupancy estimation on a home-level is achievable with 19% deviation from the ground truth, and it is possible to differentiate and track individual occupants based on their movement patterns, which might pose privacy concerns. For the laboratory setting, both binary and range occupancy inferences are shown to be feasible, with binary occupancy performing better. The work aims to demonstrate that undesirable inferences about occupancy are feasible; the authors point out that with advances in machine learning techniques, the estimation error is expected to further decrease, making it possible for providers as well as adversaries to infer occupancy information from public sources.

Pros and cons: By applying multiple sensors, the obtained parameters can show a correlation with the number of occupants, which might be exploited to increase the accuracy. However, there is no comprehensive study, which combination of sensors increases accuracy to such an extent that it justifies the increased costs for the installation of additional sensing infrastructure. Some of the sensors used might moreover pose

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/COMST.2019.2902824, IEEE Communications Surveys & Tutorials

privacy issues.

C. Summary of device-free approaches

In summary, there are different development trends in the RF-based and the sensor-based domain. In the sensor-based domain, over the years several solutions have been proposed from time to time with no clear advance to one or the other. This is due to the fact that most of the approaches show similar accuracy that can be achieved with comparable costs for installing additional sensing infrastructure. However, most of the solutions may be applicable for a small number of occupants if moderate accuracy is sufficient; especially in indoor environments. Although LiDAR technology could achieve higher accuracy, the associated costs for equipment are high and may prevent large-scale deployments. PIR-based solutions represent an exception. Even though there is no academic approach, which shows all of its benefits, there are proven industrial solutions for counting people in public transport systems, which have been shown to achieve both high accuracy and scalability (cf. Section VI). This results, among other things, from the fact that the installation of PIR sensors is cost-effective and simple. Moreover, as apparent from Table I, a simple counting algorithm is sufficient to calculate the people count. The other solutions base their estimations on diverse algorithms and models, respectively, which often require a training phase to learn the characteristics of certain environments, which might be infeasible if crowds form sporadically, e.g., during concerts, public meetings, fairs, etc. Hence, they are often limited to the application in environments, in which they have been trained.

Contrary, there seems to be an evolution towards UWB-based approaches in the RF-based domain. The development started with solutions based on RSSI, which can achieve moderate accuracy for a small number of occupants without the need for installing additional sensing infrastructure. RSSI measurements however need preprocessing due to low quality resulting from multipath effects. Hence, often complex regression- or probabilistic-based, and machine learning algorithms and models are necessary to infer occupancy from RSSI measurements. Newer WLAN infrastructure offers the possibility to implement people counting based on CSI measurements. They provide better quality, which improves accuracy, especially in dynamic environments but is still limited in the number of occupants. Finally, most recent works show the ability of UWB-based radar systems to achieve high accuracy while being scalable. In [37], a people counting system based on IR-UWB radar sensors for subway stations is introduced, which is able to count up to 98 people for two minutes and up to 8881 passengers for one day with an error of less than 10%. Such solutions however require the utilization of additional radar sensors and it is not yet clear if they are applicable to all kinds of indoor and outdoor environments. More field trials are needed here to investigate the achievable performance under realistic circumstances. As opposed to sensor-based systems, all RF-based solutions have in common that such systems can be used even in the case of fire, smoky, and lightless condition.

Finally, it should be noted that most device-free systems pose no privacy issues for the individuals to be counted. However, we note that solutions that record movement patterns of persons [60] or their speech characteristics [47], may reveal the identity of individuals and thus violate privacy. LiDAR-based approaches are an exception among device-free solutions because laser scanners are used to capture images. Depending on how the system is constructed, objects and persons can be identified. However, such functionality is not necessary for use in people counting, and the system is easier to set up without the need for image analysis functionality.

IV. Device-based Methods

Device-based occupancy estimation comprises approaches where users are expected to carry devices to facilitate the people counting process. A direct consequence of this definition is that providing a correct crowd estimate is highly dependent on the willingness of participants to carry these devices with them on a daily basis (or at least during crowd measurement campaigns). Since incentives for equipping people with additional hardware are often not easy to justify, devices are mostly considered to be smartphones (possibly with some pre-installed software for data collection and processing) due to their ubiquity, however wearable sensors are also an alternative. We note that within certain contexts other types of devices may be as prevalent as smartphones. For instance, access keycards are common among employees within an organization or commuters inside a city's transportation system. However, such devices often pertain only to a particular segment of users, and even if used for people counting, the granularity of the estimate would be rather coarse (i.e., one may know how many people have entered a particular corporate building but not on which floor there are more occupants at any point in time).

We differentiate between two main categories of device-based methods for crowd counting: *acoustic-based* and *network-based*. In the following sections we present an in-depth discussion on state-of-the-art solutions comprising each of these categories. Table III summarizes the advantages and disadvantages of the most prominent methods in each category.

A. Acoustic-based solutions

Acoustic-based solutions perform occupancy estimation by evaluating audio signals that are either transmitted by smartphones or produced by speaking people.

In the past years there have been two main approaches in the area of device-based acoustic-based solutions for people counting. The first approach, referred to as tones broadcast, is presented by Kannan et al. in [66] who propose a people counting solution based on audio tones that leverages the microphone and speakerphones commonly available in current mobile devices. The main idea behind the system design is as follows: Each device initially chooses a bit pattern where every bit corresponds to a simple tone. (The number of audio frequencies available is pre-defined.) Devices periodically broadcast the tones corresponding to their bit pattern to nodes

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/COMST.2019.2902824, IEEE Communications Surveys & Tutorials

TABLE III
SUMMARY OF ADVANTAGES, DISADVANTAGES AND APPLICABILITY FOR DIFFERENT CATEGORIES OF DEVICE-BASED PEOPLE COUNTING TECHNIQUES.

Category	Advantages	Disadvantages	Applicability
Acoustic-based [66]–[69]	Utilizes existing hardware on mobile devices.	Counting latency increases with number of devices. Counting accuracy decreases with silent occupants.	Indoor environments
RFID-based [70], [71]	Proven solution for counting people in public transport systems.	Requires installation of additional infrastructure and people to carry additional hardware on them.	Indoor and outdoor environments
WSN-based [72, 73]	Cost-effective solution, as hardware components are rather inexpensive.	Requires incentives for people to carry additional hardware on them. Counting accuracy in highly dynamic scenarios not studied.	Indoor environments
Wi-Fi-based [74]–[79]	Utilizes existing infrastructure, no further installations needed.	Counting accuracy depends on assumption that Wi-Fi interface is turned on on devices. Occupancy estimation is rather coarse (zone/room level).	Indoor environments
Bluetooth-based [80]–[84]	Utilizes existing hardware on mobile devices.	Counting accuracy depends on the number of devices in discoverable mode. Requires installation of additional hardware.	Indoor and outdoor environments
Cellular-based [85]	Utilizes existing infrastructure, no further installations needed.	Counting accuracy depends on potential collaboration between network operators.	Indoor and outdoor environments
Collaborative [86,87]	Utilizes existing hardware on mobile devices.	Counting accuracy depends on potential collaboration between participants. Technical and implementation challenges not yet addressed.	Indoor and outdoor environments

in their vicinity. As nodes receive audio samples from other devices in their communication range, they incorporate the received bit pattern into their initial bit pattern. Two tone-counting algorithms are proposed: a simple uniform hashing and a more involved geometric hashing. In the case of uniform hashing, exactly one bit in the initial bit pattern is set to one; in the more complex case of geometric hashing, more than one bit can be set to 1, which increases reliability. The proposed solution is evaluated via simulations, as well as by extensive experimental evaluations performed both in indoor and outdoor environment of a university campus. The experiments encompass up to 25 Android devices, deployed statically in each experimental scenario. Results show that people counting using audio tones may achieve up to 90% accuracy in certain experimental configurations, however other scenarios may experience more than 50% estimation errors. Furthermore, counting latency increases significantly as the number of devices participating in the counting increases both in a single-hop and in a multi-hop scenario. This indicates that the solution may not be appropriate as-is for highly dynamic scenarios in which nodes join and leave the system while counting is in progress. Energy consumption is also evaluated, and it is shown that using tone counting is able to reduce the energy consumption of a mobile device by up to 80% as compared to simply using a Wi-Fi or 3G radio interface to report presence to a server which in turn performs the counting in a centralized manner. Although scalability is said to be feasible, it is limited by the number of counting frequencies used in the system. Moreover, practical challenges such as support by devices, impact of clothing, and environmental noise levels strongly affect the parameters that need to be preset before initiating the people counting process.

The main concept behind the second approach for device-based acoustic-based people counting, which we denote as speech detection, is illustrated in Fig. 6. To estimate the number of active speakers in a group, a three-step process based on unsupervised machine learning techniques is devised: speech detection and segmentation, feature extraction and counting. This approach was first introduced by Xu et al. in [67,68]. The authors demonstrate a mobile application, Crowd++, that accurately estimates the number of people talking in a certain place. Crowd++ is shown to perform well in small-scale scenarios, such as counting the number of participants in a conference room. As any technique, dependent on audio sampling, a number of external factors — such as smartphone location and surrounding audio context — may affect audio quality. Thus, Crowd++ is only applicable when the exact count estimation is not required. Furthermore, it is not straight-forward whether Crowd++ could be used for estimating crowds of sizes larger than 10. Given the fact that the application collects and processes data only based on inputs from the phone's vicinity, it may be non-trivial to scale the concept for large-scale formations.

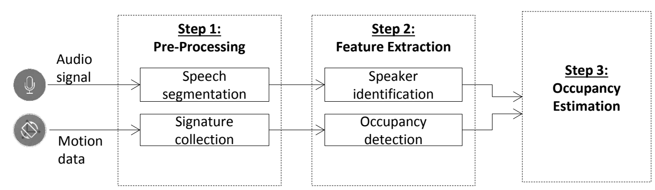

Fig. 6. Acoustic-based people counting with the help of mobile devices often implements the following generic three-step process: speech detection and segmentation, feature extraction and actual counting [67,68]. Additional locomotive models aimed at improving estimation accuracy pertain to the same three-step process [69]. Note that the whole analysis may be carried out locally at the mobile end or remotely on a dedicated server.

One of the main drawbacks with purely microphone-based

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/COMST.2019.2902824, IEEE Communications Surveys & Tutorials

solutions such as Crowd++ is that they fail in the absence of any conversational data from the surrounding environment. To address this issue, in [69] Khan et al. propose SensePresence, an opportunistic collaborative sensing system that exploits the acoustic and motion sensors built in smartphones to infer the number of occupants in a crowded environment. The system comprises two subsystems: one deployed on a mobile device (for sample collection) and another deployed in a server (for data analysis and people counting). The system has two distinct modules, that are able to infer occupancy either via acoustic or motion data, and are also able to work jointly in complex environments where not all participants are conversing simultaneously. The authors validate the system on a testbed and show that their approach performs better than purely acoustic approaches for groups of up to 10 people. However, it is not clear how such a solution scales with the number of participants in a room. Since samples are sent to a centralized unit for analysis and classification, user privacy may be compromised. Moreover, the system seems to exploit the underlying assumption that participants are mobile while they remain silent; this is a strong assumption and may not be applicable in all contexts.

Pros and cons: The idea of using audio-based methods for people counting (either relying solely on acoustic data or on a combination of acoustic data and data obtained by other sensors available on mobile devices) is interesting, however still in its infancy with very limited applicability at the current stage of existing solutions. One of the main disadvantages with such approaches is the vulnerability of the audio signal, which is easily affected by various external factors. At the moment, scalability is also a challenge which may potentially be resolved by introducing different (potentially radio frequency based) techniques on top of existing acoustic-based approaches. Although in [66] Kannan et al. test their solution in indoor and outdoor environments, currently the predominant applicability of acoustic-based people counting remains in the context of small-scale indoor occupancy estimation. Finally, collecting and processing audio samples that uniquely identify participants in a crowd, raises privacy concerns.

B. Network-based solutions

In the following subsections we divide existing wireless network-based solutions into five subcategories with respect to the technology they employ: (1) radio-frequency identification (RFID); (2) wireless sensor networks (WSN); (3) Wi-Fi in infrastructure mode; (4) Bluetooth; (5) cellular infrastructure; and (6) collaborative solutions adopting some sort of device-to-device communication.

1) RFID-based solutions: Radio frequency identification (RFID) is a technology that utilizes radio waves to identify and track electronically stored information from digital tags attached to physical objects. Typically a RFID system consists of three components: one or more tags, a reader and a scanning antenna. A tag, in turn, consists of microchip, memory and antenna. Whenever a reader and a tag are in direct communication range, the tag is able to transmit the information stored on it to the reader. Thus, object counting is possible whenever data stored on each RFID tag is unique to its corresponding object.

In [71] Li et al. propose an occupancy detection system to support demand-driven HVAC operations. The system is intended for detecting, counting as well as tracking individuals in thermal zones in an indoor environment; a thermal zone is a space served by a dedicated HVAC subsystem. To this end, the proposed system consists of six tracking RFID tags each attached to one of the occupants in the space, 25 reference RFID tags deployed strategically in a total of 13 zones in the environment to report known locations, as well as four readers with two antennas each and a server. The testbed is deployed on a floor of a university building. Different tests with varying number of occupants, stationary or mobile, are performed. Occupancy estimation, although being a byproduct of system operation, achieves 100% accuracy across all configurations. The authors also show that the detection rate of occupant's location is higher for stationary occupants than for mobile ones. This is mainly due to the sub-optimal placement of reference RFID tags. Moreover the scalability of the approach is not investigated. Although the results are promising, such approach raises significant concerns with respect to user privacy as the system is able to track individuals based on predefined mapping of an RFID tag and a person. Moreover, the occupancy estimate is based on the willingness of participants to wear additional devices on them. While this may be incentivized if the RFID tags are integrated within access cards for regular office occupants, it may still impossible to capture the behavior of external visitors within the space. Whether such errors may be tolerated depends on the application at hand.

In [70] Weaver et al. investigate the possibility of using RFID technology to monitor large crowds in outdoor environments. The proposed system consists of RFID tags, a RFID reader, and a single antenna. The focus is on dimensioning the optimal height off the ground and the angle of inclination of the antenna for the purpose of tracking moving individuals equipped with passive or battery-assisted passive RFID tags when the read range of the antenna is 100 meters. The authors show that such read range corresponds to a maximum coverage area of 4,000 m^2 and could theoretically count up to 5,000 people. Coverage area could be extended by adding multiple antennas strategically positioned in space such that they form a mesh or cell configuration. However, the read rate of the reader is shown to be the limiting factor: for capturing 5,000 people it takes approximately 50 s. Thus, such solution may be more appropriate for sparsely populated areas, in order to be able to provide an occupancy estimate in a reasonable time, or for densely populated areas with static crowds.

Pros and cons: RFID-based systems represent a feasible solution for counting people both in indoor and outdoor environments. Both academic and industrial solutions have shown that respective system can achieve high scalability. The achievable accuracy is not specified but as the industrial solution (cf. Section VI) is supposed to be applied in public transport systems, it is assumedly above 95%. However, extra infrastructure needs to be installed and people need to be incentivized to carry RFID tags.

2) WSN-based solutions: It is feasible to devise the size of a crowd by utilizing density estimation approaches from the wireless sensor networks domain if all participants in the crowd are equipped with sensor devices. However density estimation in wireless sensor networks often assumes static deployments in which nodes attempt to determine the number of neighbors they are surrounded by [88,89]. Furthermore, often the network topology is assumed to be known in advance [90] and no node failures are considered, i.e. nodes are not allowed to join or leave the system while counting is performed [91]. In the context of people counting in mobile environments, such assumptions would be unrealistic. Therefore, we here only focus on solutions that offer full support for node churn and node mobility.

In [73] Kamra et al. present CountTorrent, a distributed algorithm that enables fast estimation of aggregate queries, including counting, in both static and mobile wireless sensor networks. CountTorrent operates in two phases: label assignment and information swapping. (i) During the label assignment phase each node receives a unique identifier derived from the identifier of another node, a parent, thus forming an abstract tree of identifiers. In static environments, parent-children relationships are established only in the beginning of the protocol execution, however in mobile environments the process is performed continuously as nodes join, leave or move through the network. (ii) During the information swapping phase nodes exchange crowd measurements with one another when in direct communication range, and merge their local measurements with the received measurements in order to obtain a correct estimate of the node count. The way in which data is merged depends on the identifiers assigned in the label assignment phase. Different heuristics are applied to minimize traffic overhead and increase convergence speed. CountTorrent is evaluated via simulations and experiments, and is shown to achieve between 80% (in mobile environments) and 100% (in static environments) accuracy in the count estimate. The maximum number of nodes evaluated is 100 and 500 in mobile and static environments, respectively; for the testbed the authors perform experiments with 15 static MICAz motes. We note however that the mobility assumptions adopted in the evaluation are rather simplistic (i.e., random waypoint models), and the network density is rather low. It is therefore not straightforward to judge whether CountTorrent has potential to operate in highly populated scenarios or scenarios with different types of mobility; however one potential application of CountTorrent may be in performing indoor occupancy estimations in semi-static environments such as counting participants in a meeting.

In [72] Cattani et al. present Estreme, a fast, asynchronous and lightweight node count estimator in dense wireless sensor networks. In order to estimate the number of neighboring devices in a network Estreme makes use of the times between periodic but random events (such as beacons transmissions) and defines an estimator based on the mean time to observe all events. Two versions of the estimator (a spatial and a temporal one) are implemented on top of a low-power listening MAC protocol for the Contiki OS, and the performance is compared to state-of-the-art benchmarks. For a testbed of 100 sensor nodes equipped with an MSP430 processor and a CC1101 radio, Estreme is shown to estimate the number of neighbors with an error of approximately 10% for static scenarios. The estimator is also evaluated in the presence of three mobile nodes however mobility is rather instrumented. Thus, it is not straight-forward how Estreme would behave in scenarios where all nodes are mobile. Moreover, as Estreme is designed to estimate the number of neighbors, a global view of the device count across the whole network is not achievable with the current state of the work.

Pros and cons: As wireless sensor networks have traditionally been considered to predominantly constitute of static nodes, very few studies explore the problem of neighbor counting under the assumption of node mobility. When node mobility is evaluated, the mobility patterns are often rather unrealistic and scenarios are sparsely populated. Moreover, initial proposals suggest that mobility significantly reduces the accuracy of the estimate. It is therefore not possible to easily asses whether these solutions would fit in the context of people counting where the system experiences both high dynamics as well as high densities.

3) Wi-Fi-based solutions: Due to the ubiquitious deployment of Wi-Fi access points in homes, offices and public spaces, monitoring and analyzing Wi-Fi *probe-request frames* has been suggested as a means for inferring information such as social relationships in crowds [92] or trajectories of mobile devices [93]. The basic idea is that WLAN-enabled mobile devices periodically send special frames, called probe-requests, across all channels as part of their IEEE 802.11 protocol operation to detect available access points in their vicinity. If an appropriate network monitoring hardware is present (which is in most cases integrated within an access point, but could potentially be deployed as a stand-alone equipment), it is possible to overhear these probe-requests and thus accurately count all devices within communication range.

In [74] Handte et al. present a system for real-time crowd density estimations based on Wi-Fi probe-request frames in the context of people counting in public transportation vehicles. The system is validated on a prototype and is evaluated experimentally on three buses in the city of Madrid. The authors find that the proposed solution has around 20% accuracy with respect to the ground truth; the actual bus occupancy is gathered manually. Whether this approach provides sufficient accuracy or not depends strongly on the application at hand. However, the study is indicative that people often do not use the WLAN interface of their mobile devices when on-the-go, and that other communication technologies may be preferred during their daily journeys.

In [75] Yaik et al. further investigate the correlation between counting Wi-Fi probe-request frames and actual number of people in a crowd. They perform a crowd-counting campaign

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/COMST.2019.2902824, IEEE Communications Surveys & Tutorials

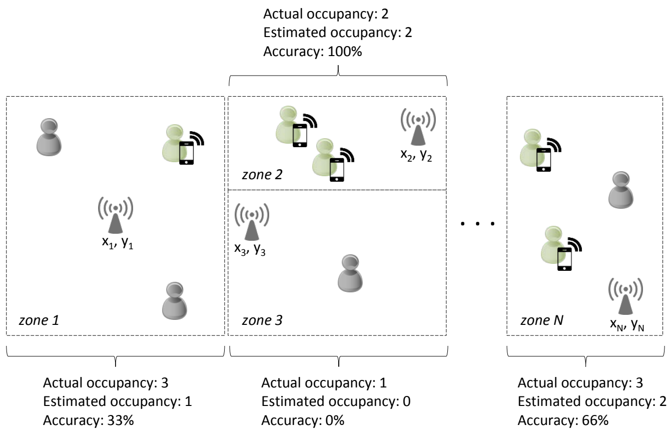

Fig. 7. In Wi-Fi-based people counting the space is often divided into a number of areas, or zones, with a Wi-Fi access point associated with each zone [77,78]. The occupancy estimate is strongly dependent on the number of people that actually carry a Wi-Fi-enabled mobile device while inside a given zone.

during an open day at the university, in which a total of 850 people (guests and staff) participate. Volunteers are recruited to perform the manual crowd counting on an hourly basis which then serves as a benchmark for comparison. The results are promising, showing correlation of up to 89%, however as the outcome reflects only a single experiment in an academic environment (where people are more prone to have the Wi-Fi interface of their mobile device turned on), it may not be applicable to other scenarios. Moreover, a general problem with crowd counting via Wi-Fi probe-request frames is the implicit requirement of availability of Wi-Fi infrastructure. Thus, while such an approach may be appropriate for certain indoor scenarios, its applicability in outdoor contexts is limited.

In [76] Fierro et al. adopt a passive localization technique in the context of real-time occupancy counting for building automation. Given the coordinates of existing Wi-Fi access points in an area, the authors localize a mobile device based on the centroid of these coordinates and the average detected signal strength with a zone-level precision. Although the level of precision is low as compared to other localization techniques, the approach is considered suitable for the purposes of counting people in HVAC or lighting zone scenarios.

In [77] Balaji et al. present Sentinel, a system that estimates occupancy in commercial buildings based on existing Wi-Fi infrastructure as well as mobile devices carried by people. Sentinel is deployed in a five-story building and during a 10-day test it is able to accurately infer binary occupancy 86% of the time. The space in the building is divided in two main categories: private areas (i.e., offices) and shared areas (i.e., corridors, meeting rooms, etc.), and occupancy is only inferred for the private areas. The occupancy estimation is rather coarse: if a mobile device of a particular user is attached to an access point which covers the user's office space then the user is said to be inside the office. However, such coarse granularity is considered appropriate (albeit conservative) in the context of HVAC regulation, as it only requires to be able to determine whether to turn heating on and off in a particular office space.

In [78] Lu et al. present WiPin, a system that accurately infers occupancy information with the help of existing Wi-Fi infrastructure and minimum user effort. The system is evaluated on a hardware testbed deployed in a workspace with 25 occupants for six weeks. In this work, the authors do no expect users to download any specific software on their mobile devices. Instead, the system resides in two modules: a front-end module (on the Wi-Fi router) for snapshot occupancy estimation performed in real-time and a back-end module for continuous occupancy estimation taking into account historical occupancy data in order to cope with challenging events, such as user mobility. The front-end module can work in isolation or together with the back-end module, depending on the level of accuracy required for occupancy inference. To achieve this, the indoor space is divided into zones (see Fig. 7), and each zone is assigned a single Wi-Fi access point with a well-known location; such a setup allows the authors to achieve room-level occupancy estimation. In the front-end snapshot estimation, users are assigned to zones with respect to the RSSI measurements; in the back-end, continuous estimation of temporal correlation is exploited. In comparison to other approaches, WiPin is shown to reduce the mean squared estimation error by up to 37% in the case of snapshot estimation, and to further improve accuracy when the back-end is activated.

In [79], Melfi et al. undertake a different approach to Wi-Fi based people counting and introduce the concept of implicit occupancy estimation in which they try to infer occupancy based on associations of mobile devices with existing wireless infrastructure in a university campus building. The counting method simply measures the number of hosts associated with each access point indicated by DHCP leases, and is thus intended to replace or to complement dedicated sensor-based solutions. To validate their approach, the authors use an experimental testbed consisting of eight access points in two different locations. They further compare the obtained results to a PIR sensor-based solution and the ground truth, which is determined by visually counting the occupants, and conclude that implicit occupancy achieves an accuracy of around 90% for tests with up to 60 people. The authors thus claim that implicit occupancy estimation might be sufficient for HVAC applications, however as the tests are conducted in university campus where people are more prone to use devices within the Wi-Fi network, it is not straight forward whether such levels of accuracy could be achieved in other scenarios.

Pros and cons: The ubiquitous deployment of Wi-Fi access points in both public and private indoor spaces makes Wi-Fi-based people counting techniques significantly more attractive than other approaches due to their potentially faster adaptation. However Wi-Fi deployments in outdoor spaces are still relatively sparse, often serving mainly touristic areas of capital cities. Moreover, as Wi-Fi coverage is limited, some collaborative algorithms for people counting on a larger scale need to be implemented in the existing infrastructure.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/COMST.2019.2902824, IEEE Communications Surveys & Tutorials

Another major challenge with the applicability of Wi-Fi-based techniques is the lack of systematic approach for inferring actual occupancy levels out of only a subset of mobile devices. As shown in some of the studies in this subsection, more often than not people tend to have the Wi-Fi interface on their devices turned off, thus counting solutions often end up underestimating the total population. The provided estimate may be sufficient in smart building scenarios for optimizing HVAC operations, however they may become inapplicable to critical scenarios such as evacuation. Providing appropriate incentives to users for turning on the Wi-Fi interface on their devices or developing complex predictive algorithms may be the two main roads to go in this case.

4) Bluetooth-based solutions: Initially, Bluetooth-based solutions for crowd counting were introduced in the context of intelligent transportation systems with the objective to help better understand passengers behaviors and potentially provide relevant information to those on board. One of the first works in this domain is by Kostakos et al. [80] where the authors describe an inexpensive system that uses off-the-shelf Bluetooth hardware to derive passenger origin/destination matrices. An origin/destination matrix is a typical means for describing the flow of passengers between various points in the transportation network. The system consists of a Bluetooth scanner mounted on the ceiling of a bus, which periodically scans for discoverable Bluetooth devices in its range. The system is implemented in one bus traversing 19 different routes in the city of Funchal, Portugal, in the course of four weeks, and during the trial data from more than 1000 unique Bluetooth devices is collected. This data is then correlated to bus schedule information in order to derive device trips, i.e., duration from the moment a Bluetooth-enabled mobile device is detected onboard the bus until the moment it no longer responds to the scanner's requests. Data is collected during bus operation hours, and then post-processed offline. Although this work does not specifically evaluate crowdedness in the transportation system, bus occupancy is determined based on the collected sample size.

In [81] Versichele et al. evaluate proximity-based Bluetooth tracking as a method for analyzing the spatio-temporal dynamic of mobile users during mass events. The authors install static Bluetooth scanners in strategic locations throughout the center of the town of Ghent during a 10-day festival that attracts approximately 1.5 million visitors. The main goal of the study is to develop quantitative understanding of visitor's behavior during the festival, including but not limited to the number of visitors at any given moment. A data set of more than 100 000 uniquely detected Bluetooth devices is collected and analyzed. (Observe that this data set is much smaller than the anticipated number of visitors at the event.) The authors study different aspects of human movement dynamics, including visit duration, crowdedness and flow analysis. However all analysis is performed offline, after the final day of the event.

In [82] and [83] Weppner and Lukowicz present a collaborative technique for crowd density estimation based on detecting discoverable Bluetooth devices of mobile users. Here, the term "collaborative" refers to the fact that the crowd estimate is calculated based on measurements collected by multiple devices in a given location. In two consecutive studies the authors investigate different feature sets for crowd counting, and evaluate these sets during two extensive experiments during the Munich Octoberfest festival and a public gathering during the European soccer championship. Both experiments show that using collaborative features instead of simply counting discoverable devices achieves up to 30% improvements in the crowd estimate (as compared to a ground truth, obtained by camera footage). However, the approach has one major disadvantage: it requires the recruitment of volunteers equipped with mobile devices with Bluetooth scanning software who should either stay static in popular locations or move according a specific mobility pattern in order to meet as many participants in the observed area as possible. The discussion on how many volunteers are needed and what mobility patterns should be followed is left for further investigation.

Finally, based on their previous studies, Weppner et al. describe in [84] a system that leverages users who voluntarily have their smartphones scan the environment for discoverable Bluetooth devices in order to analyze the crowd conditions in urban environments. The system is experimentally tested during a three-day city-wide festival in Zurich, including a real-life data set of 1000 scanning devices and ground truth data set (in the form of a collection of GPS traces) from nearly 30 000 users (of a total of 2 million participants). The system is implemented in a mobile app released for the purposes of the festival. Data is first collected, and analyzed at a later stage, i.e., the crowd estimation process is not done in real-time. The authors define a 12-dimensional spatio-temporal feature vector and use regression analysis to compare their crowd estimations to the ground truth. Similar to their previous studies, the evaluation demonstrates that using a complex vector space than simply counting number of (uniquely) discovered Bluetooth devices results into smaller estimation error. The authors go a step further and also demonstrate the general feasibility to detect crowd flows. However, the main questions as to how the number of volunteers affects crowd estimation, how to incentivize people to participate outside of specific events and how to enable real-time crowd estimation remain without answer.

Pros and cons: Similar to the Wi-Fi based solutions in the previous section, one of the main drawbacks of Bluetooth-based solutions is the lack of a systematic approach for determining the actual crowd size out of a computed sample, potentially in real-time. Different studies report a ratio of detected Bluetooth devices from 7% ([94] in 2006) to 11% ([81] in 2012) however these numbers are experiment-specific and cannot be considered reliable for future extrapolations. What is worth noting, though, is the seemingly increasing tendency of the number of discoverable Bluetooth devices over the last years. On the one side, such tendency may be considered indicative for the potential of Bluetooth-based crowd counting solutions. On the other side, with the introduction of wearables and other Bluetooth-enabled devices, Bluetooth-

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/COMST.2019.2902824, IEEE Communications Surveys & Tutorials

based solutions may produce results with higher bias as one individual is more prone to carry multiple devices, and thus be considered multiple times in the crowd counting process – a phenomenon which has not been considered in current studies.

5) Cellular-based solutions: The only known work on cellular-based crowd estimation is proposed in [85] by Ramachandran. The author discusses the addition of a simple hardware and software component, a crowd size analyzer, to the cellular network infrastructure which collects and analyzes information from cell tower associations in order to estimate the crowdedness in a given location.

Pros and cons: One of the main advantages of cellular-based solutions is that they are able to perform people counting in any location: from indoor environments (via picocells) to large outdoor areas at a city or even country level (via macrocells). However, as different mobile users are subscribed to different operators, the accuracy of the crowd size estimate is only limited to a subset of users belonging to a particular mobile operator. Thus, a potential collaboration among mobile operators is required for assessing the overall crowdedness in any location. Additional external factors, e.g., unreliability of the wireless channel in crowded locations, may further deteriorate the crowd estimation. Privacy issues may also arise, if data is not properly anonymized prior to providing crowd estimates to interested parties.

6) Collaborative device-to-device solutions: Recently crowd counting methods based on opportunistic device-to-device communication between mobile devices in direct range have been proposed as an alternative to the infrastructure-based solutions discussed in the previous subsections.

In [86] Pajevic and Karlsson propose an application for crowd size estimation based on epidemic spreading of small messages among users in direct communication range. The authors define a stochastic model to study the performance of the application. However, the model is shown to be unable to capture precisely changes in scenarios with bursty arrivals and departures, as it assumes constant arrival rates.

In [87] Danielis et al. propose UrbanCount, a fully distributed crowd counting protocol for scenarios with high crowd densities. The protocol relies on mobile device-to-device communication to perform crowd estimation. Each node collects crowd size estimates from other participants in the system whenever in communication range. The objective of UrbanCount is to produce a precise mapping of the local estimate to the anticipated global result while preserving node privacy. The protocol is evaluated via extensive simulations and it is shown to perform well in dense scenarios with crowd sizes up to 2400 people, achieving more than 90% accuracy. However counting accuracy decreases significantly for sparsely populated scenarios.

Pros and cons: Collaborative device-to-device communication solutions appear to be a promising approach for enabling real-time distributed crowd estimation both for indoor and outdoor environments in a scalable and privacy-preserving manner. However, the main disadvantage of existing proposals is the lack of prototype implementations that demonstrate the feasibility of the concept on actual devices. This is partially due to the fact that the most suitable communication technology for collaborative device-to-device communication has not been determined yet. Moreover, a challenge lies also in the recruitment of sufficient number of participants for conducting measurement campaigns at large in the open.

C. Summary of device-based approaches

In summary, the overall development of device-based approaches remains relatively segmented with respect to applicability. Solutions are often targeted towards a specific scenario (sometimes even a specific event occurrence), and replication of experimental setups is hard to achieve. As outlined in Table III each approach has its own advantages and disadvantages, however currently there is clearly no one-size-fits-all solution that significantly outperforms the rest of the proposals in the literature.

In the acoustic-based domain, there is a general sense of continuity as solutions move away from purely acoustic-based towards hybrid approaches which exploit both acoustic and other sensors for better occupancy estimation. Albeit these advances, acoustic-based approaches are still in their infancy, and further investigations are needed in terms of applicability, scalability as well as privacy protection.

Looking at research trends in the network-based domain, one can clearly notice two large categories of solutions: (1) those that require nodes to be equipped with external devices (WSN-based and RFID-based methods); and (2) those that require nodes to be equipped with their own mobile devices (Wi-Fi and Bluetooth-based methods, either centralized or distributed). There is a clear trend towards using solutions that fall into the second category, which is indicated by the larger body of work that appears in this area. This is mainly due to two factors. Firstly, incentivizing users to carry additional devices (sensors or tags) is a challenging task. Given the ubiquity of mobile devices in our day-to-day life, utilizing what is already present decreases the adaptation barrier for potential go-to-market solutions. Secondly, in the context of infrastructure-based solutions, one can rely on existing deployments, thus decreasing installation and maintenance costs. We would like to note that the initial findings in the context of collaborative network-based solutions [86,87] demonstrate a promising solution for large-scale scenarios.

It is also worth noting that network-based solutions which make use of existing wireless infrastructure are not expected to pose further requirements on the network side in terms of increased traffic volumes, additional overhead or contention for the wireless medium. As devices already associate and exchange control data with access points [77,78] or base stations [85], it is only a matter of interpreting this data in order to accurately infer the people count. Whether the estimate is done at the network edge or in the cloud depends on the chosen system design. Any limitations in terms of maximum number of devices that can be served from a given access point/base station are related to network dimensioning, not to the particular application at hand.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/COMST.2019.2902824, IEEE Communications Surveys & Tutorials

Finally, we would like to note that in certain cases device-based approaches may be considered as privacy-intruding if data obtained from devices is not processed correctly, i.e. anonymized. In the acoustic-based domain, whenever speech characteristics are collected, persons could be individually identified. In the network-based domain, individual identification may be performed if there is a well-known (predefined) mapping device IDs to the persons who carry those devices. This is particularly true for solutions that collect data individually from each and every participant in the crowd. However, this issue does not pertain only to the domain of people counting. On the contrary, privacy concerns (in the context of tracing individuals or learning about their usage patterns) may arise when using any application that establishes a one-to-one communication between a device and a network. Thus, one can make use of the same privacy preserving techniques that are currently applied in wireless networking when designing people counting solutions that rely on existing infrastructure. Collaborative device-to-device solutions are privacy preserving by design, as nodes exchange aggregate data for the crowd estimate and data is presented only to trusted authorities.

V. Device-free vs. Device-based Methods: A Comparison

Now that we detailed the advantages and disadvantages of solutions belonging to the device-free and device-based category, in this section we provide a general comparison of device-free and device-based methods, and we discuss their suitability with respect to scalability and to the application areas discussed in Section II-B.

On the one hand, device-free approaches do not require people to carry a device so that they represent non-invasive solutions. While RSSI- and CSI-based systems can even leverage commercial off-the-shelf WLAN infrastructure, UWB-based radar and sensor-based approaches require the installation of an additional sensing infrastructure. RSSI-, CSI-, and most sensor-based systems solely achieve moderate accuracy for a small to moderate number of occupants. This might be sufficient for smart buildings or urban analytics but it is definitely not feasible for disaster management. Moreover, as apparent from Table I device-free solution often need complex algorithm and models, respectively, involving a learning phase in a specific environment to infer occupancy from measurements. PIR-based approaches have not been investigated in detail in academic works but have been shown to be able to achieve high accuracy and scalability in industrial products such as IRMA [65]. Moreover, there is a trend from RSSI- and CSI-based systems towards UWB-based radar systems, which can probably achieve high accuracy and high scalability but their full potential is still subject to further investigations. Most of the device-free solutions preserve privacy due to their inability to identify individuals. However, some solutions like LiDAR-based systems and other sensors can be used to track and identify persons based on their shapes and movement patterns so that a few solutions pose privacy concerns.

Device-based approaches, on the other hand, are a promising solution for crowd counting of larger populations, both indoors and outdoors. One of the most attractive features of device-based solutions is that people are often already equipped with at least a single device (a smartphone) which can be utilized for performing distributed counting tasks. Thus, no or minimal installation costs may be required, which makes the introduction barrier on the market of device-based solutions rather low. We note that this may not hold if users are expected to purchase and carry on them separate devices, as in the case of current RFID-based solutions. Furthermore, most studies exhibit relatively high accuracy in the crowd estimate, even when solely applying simple counting algorithm and only a subset of the population is considered. However, this also indicates one of the major challenges in front of device-based solutions which may severely hinder their ubiquitous deployment: the current lack of incentives for participating in the crowd-counting process. As shown by Handte et al. [74] as well as by Weppner et al. [84] among others, the participation rate of people in the crowd counting process can be rather low (not more than 20%) as people are often concerned with battery consumption on their devices, as well as potential privacy infringements. We note that it is possible for different applications to have different requirements with respect to privacy.

We note that the scalability of various methods, which is in turn tightly coupled to their achievable accuracy, is highly dependent on the crowd density at any given moment. On the one hand, device-free RF-based approaches with adequate transmission range are well-suited for counting people in low-density scenarios, in which people do not stand close to one another. Such scenarios allow for each person to be individually distinguished during the counting process, and thus avoid incorrect estimations [39]. Alternatively, device-free as well as device-based acoustic-based solutions can also be applied to scenarios with low density. However, as shown previously, scalability becomes an issue when the number of people increases. On the other hand, device-based collaborative solutions that exchange count estimates using device-to-device communication are shown to perform poorly when applied to scenarios with low density, however demonstrate their potential in denser environments [87].

Finally, in Table IV we summarize the suitability of each subcategory of device-based and device-free approaches with respect to the application areas listed in Section II-B. It can be seen that RF-based solutions are expected to perform better than acoustic-based across all application areas in the context of device-based people counting. This is mostly due to the fact that the accuracy of acoustic-based solutions is highly dependent on the vulnerability of the audio signal which in turn is easily affected by external factors. This also poses an issue with the potential scalability of such solutions. Network-based solutions, on the other hand, present a more reliable approach for people counting across various application areas, mainly due to their availability in terms of existing infrastructure (centralized or decentralized). However they should not

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/COMST.2019.2902824, IEEE Communications Surveys & Tutorials

be considered as a one-size-fits-all solution either. Although it is expected that network-based solutions may have a greater potential in the context of urban analytics and disaster management, especially in the context of collaborative device-to-device communication solutions, they might require additional support in the context of public transportation and building automation, i.e. in the form of one or more complementary technologies in order to increase accuracy.

On the contrary, for device-free approaches sensor-based solutions are expected to be overall more suitable than RF-based approaches. PIR-based systems are a proven solution for people counting in public transport with very high accuracy and have therefore been rated as very suitable for this purpose. This applies equally to the use of PIR-based systems for urban analytics, since they can be used for counting people in shopping halls, but also at festivals. After all, PIR-based solutions as well as LED-based approaches are also suitable for counting people in smart buildings, for example, to control heating, ventilation and air conditioning. Although sensor-based systems, in particular PIR-based approaches, provide very high accuracy, they still cannot achieve 99.999%, which is required for disaster management scenarios. Here, one could use the combination with LiDAR-based systems and apply a hybrid solution to achieve the required accuracy. Although LiDAR-based systems are expensive, these costs are justified for disaster management.

All RF-based approaches have in common that they are only suitable for counting a few people with moderate accuracy. Therefore, they were rated low in terms of their suitability for public transport, urban analytics and disaster management. Only for use in smart buildings, they are conditionally suitable, since this scenario does not depend so much on a very high accuracy and the advantage of the low cost of using these systems outweighs the disadvantage of low accuracy.

Ultimately, the choice of solution is dependent on the particular application at hand and its specific requirements, so this classification should be considered only as a general guideline. The reader should always keep in mind that there may be exceptions from this general rule.

VI. Industrial Solutions for People Counting

A number of industrial products have appeared over the past years in the context of people counting in different application areas. Although many of these solutions are predominantly video-based (e.g., *Eurotech*'s PCN-1001 optical passenger counter for boarding and leaving public transportation vehicles [95] and *Flir*'s Brickstream®2D and 3D vision sensors [96] and *ViNotion*'s ViSense visual camera sensor system [97] for retail analytics and security), we here focus the discussion around some of the non-image based commercial solutions for people counting[1].

Both device-based and device-free approaches for people counting in public transportation have been commercialized. The companies *Trapeze* and *Albis Technologies* developed a ticketing system called ComfoAccess®[98]. Passengers carrying a ComfoAccess®RFID card are automatically registered in the vehicles they board by two antennas with different transmission range to ensure that the passenger is really on-board. The collected data is sent by the on-board computer to a remote cloud where the information is processed, and passengers are counted in real-time. One of the global market leaders *Iris* developed IRMA (InfraRed Motion Analyzer), a device-free people counting system that comprises infrared sensors and analyzers [65]; the system has already been implemented in more than 50,000 vehicles. IRMA sensors are specifically designed for reliable use in vehicles and are able to operate even in rough conditions (temperature, humidity, vibration). Sensors are generally mounted above the door, and people counting occurs whenever a passenger enters the vehicle. IRMA promises more than 95% accuracy for occupancy estimation.

Eco-compteur [99] provides a number of device-free solutions for people counting in the context of urban planning. The main component of the solution is the PYRO-Box, a passive infrared sensor that can be installed permanently or temporarily in urban environments to count both pedestrians and cyclists by detecting their body temperature. The solution is envisioned to be applicable in a number of areas of urban planning, from active transportation to downtown management and dimensioning of recreation areas. As of current, the solution is mostly used for monitoring pedestrian traffic in city centers in some of the largest cities around the globe.

Irisys (InfraRed Integrated Systems Ltd.) [100] provides a hybrid device-free solution based on thermal imaging and sensors for estimating occupancy in next generation smart buildings. The solution works both in real-time as well as offline and presents rich data analysis to be utilized for reducing operational costs in buildings by gauging the energy consumption with respect to the number of people located in it at any time of the day.

Xandar Technologies [101] is a spin-off from a research group at the Hanyang University in South Korea which aims at providing a people counting solution for public spaces with high user mobility, such as coffee shops or subway stations. Xander utilizes IR UWB technology, and consists of antennas which scan from a distance of 1 meter to 20 meters depending on the antenna strength; counting through walls is also possible. The solutions is able to achieve accuracy of up to 98% and has a very quick response to changes in occupancy levels.

Acorel [102] is a French company specialized in automatic people counting solutions and people flow analysis. Their system provides an automatic solution for recording the number of people that enter and leave a restricted area, and is thus well suited for application in the public transportation sector. The company provides three different sensor-based solutions: an infrared solution consisting of both passive and active IR sensors, a LiDAR solution and a stereoscopic 3D sensor (image-based) solution. Of the three approaches, the IR-based

[1]We note that all accuracy levels cited in this section are provided by the manufacturers.

TABLE IV
SUITABILITY OF DEVICE-BASED AND DEVICE-FREE APPROACHES. ONE STAR DENOTES THAT A PROPOSED TECHNIQUE IS ILL-SUITED FOR A PARTICULAR APPLICATION AREA, THREE STARS DENOTE A GOOD FIT.

Application area	**People-counting technique**			
	Device-based		**Device-free**	
	Network-based	**Acoustic-based**	**RF-based**	**Sensor-based**
Public transportation	**	*	*	***
Urban analytics	***	*	*	***
Building automation	**	**	**	***
Disaster management	***	*	*	**

solution provides the lowest accuracy (95%) and exhibits the highest sensitivity towards change of light and temperature, as well as people stopping beneath the sensor.

Dilax [103] is a Swiss company that focuses on automatic passenger counting as well as people counting in retail and airports. Their public transportation system comprises proprietary sensors. Moreover, apart from device-free counting, they use device-based (smartphone) tracking to determine mobility data and provide richer information with respect to people flows. For the purposes of counting visitors in retail or at airports, Dilax relies on a combination of cameras and 3D sensors.

Table V compares the application areas based on existing solutions for people counting via non-image based techniques proposed by academia and industry. Although there are commercial solutions in almost every application area, their number is limited. On the other hand, academic solutions present an abundance of alternatives for people counting, either with respect to device-based or device-free techniques. Since there are already proven device-free industrial solutions such PIR-based system for use in public transport with sufficient accuracy, there is no incentive for the further exploration of such an approach in the academic field. In particular, the RF-based device-free solutions, however, still have disadvantages especially in terms of accuracy, which is why several academic papers here investigate the applications urban analytics and building automation. Furthermore, it is interesting to note that there is no overlap between the technical approaches undertaken by industrial and academic solutions even when they target the same application areas. This is most visible in the context of device-based solutions, where there is a significant gap between what is available in academic research and as ready-for-market solutions. An exception is the field of public transportation, where network-based solutions can be found both in academic and industrial efforts. However, we note that even in this case, the device-based approach used in industry [103] is only a complementary solution which provides additional data, and is not a full-fledged system to allow people counting in public transportation. This poses a number of questions as to whether current academic solutions live up to the expectations or requirements of industry as well as whether they provide sufficient ease of deployment such that they could be converted into a product. It also signals the need for closer collaborations between industry and academia when building the next generation of solutions for people counting. Finally, following recent events, such as the Tohoku earthquake and tsunami in 2011 which took the lives of more than 15 000 people, it comes as a surprise that there are no prominent industrial solutions for people counting in the context of disaster management (see Table V). Fedhululllah and Ismail acknowledge this fact and stress the importance of human crowd monitoring to address the issue of disaster management and prevention [26]. On the one hand, from the viewpoint of industry this outcome is not striking: after all, developing solutions in application areas other than disaster management is more prone to bring revenues to a company in the short run. On the other hand, it is harder to explain the lack of comprehensive academic involvement when it comes to solutions for people counting during or immediately after disasters. The work of Krüger et al. at least investigates disaster management as related application scenario in an office building with up to five moving people [43]. The inability to test prototypes at large scale in realistic environments could be considered as a reason for this research gap.

Crowd counting in adjacent fields

We here provide a short overview of industrial counting and monitoring solutions introduced in adjacent fields to people counting. Namely, the focus of this subsection is livestock monitoring and management. It is interesting to note that as opposed to people counting, in the field of livestock counting non-image based solutions generally outnumber image-based solutions (for instance, a drone-based solution is presented by *Airborn Robotics* [104]). Moreover, the predominant type of livestock counting and monitoring solutions are device-based. Cattle is equipped with motion sensors with communication capabilities. Information is collected via the cellular or satellite network (in the case of *CattleWatch* [105]) or via a proprietary WLAN (in the case of *SenseTime* [106]), and is further delivered to the farmer.

VII. OPEN CHALLENGES AND FUTURE DIRECTIONS

Table VI and VII summarize the available device-based and device-free solutions for people counting and illustrates the extent to which each solution covers the requirements

TABLE V
APPLICATION AREAS OF NON-IMAGE BASED ACADEMIC AND INDUSTRIAL SOLUTIONS FOR PEOPLE COUNTING. A ✓ DENOTES THAT A SOLUTION IS EITHER PRESENTED IN AN ACADEMIC WORK, OR IS READILY AVAILABLE ON THE MARKET.

Application area	Academic		Industrial	
	Device-based	Device-free	Device-based	Device-free
Public transportation	✓	—	✓	✓
Urban analytics	✓	✓	—	✓
Building automation	✓	✓	—	✓
Disaster management	—	✓	—	—

outlined in Section II-A. At a glance it is easy to see that the problem of providing accurate, timely and automated occupancy estimates at a low cost while preserving privacy and supporting scalability remains unsolved in general. Only few solutions [37,66,74] do consider all of the requirements (except reliability), however their applicability is limited to specific use-cases and the accuracy of the proposed solutions is relatively low.

In this section we outline the steps that need to be undertaken towards developing the next generation of people counting algorithms.

A. Introduction of benchmarks

One of the main drawbacks of current solutions is the lack of common ground to facilitate the comparison between different approaches. As shown in Table VI and VII most solutions are evaluated via some sort of an experimental setup, with a scale ranging from less than ten to more than 100,000 nodes. Unfortunately, traces and blueprints are often kept proprietary, thus preventing the reproduction of the setup. This creates segmentation in the field, and further raises questions with respect to the validity of the results. Ultimately, in place should be common benchmarks for both device-based and device-free approaches, however a good first step would be to create separate benchmarks for each of the categories. For instance, a benchmark for evaluating indoor device-free and device-based approaches may include a selection of floor plan blueprints, and a subset of mobility patterns that occur in the space. More particular, a benchmark for evaluating device-free works in indoor environments could be a specific setup in a room with some pre-installed devices as reference. However, it remains unclear how to ensure comparability, especially since tests may be influenced by subjectivity in case of involved volunteers. By taking into account the human factor, measurements can be combined with subjective tests, as done in the video domain, to better consider personal perceptions. A benchmark for estimating outdoor occupancy, on the other hand, may consider an open environment with free flows and different churn rates. A valid question is however how a benchmark should be chosen and as apparent from the above reflections this is not an easy task. One alternative is to take inspiration from the annual Microsoft Indoor Localization Challenge[2] where teams from different academic institutions and industries have the opportunity to evaluate their localization algorithms in a common environment. Another alternative is to develop a common database, similar to Crawdad[3], where datasets from different evaluations can be collected and reused by researchers in the field.

B. Introduction of thresholds

What does an accuracy of 85% mean? Is it too low, or is it too high? Ultimately, one would expect the accuracy of the crowd estimate to approach five nines, however this requirement may come at a price either in terms of latency [70] or in terms of equipment costs [30,31], as well as privacy intrusion. Thus, there is a need for introducing application-based thresholds that determine whether the performance of an algorithm is satisfactory or not. For instance, the expected accuracy in the context of a building automation system may be significantly lower than that in the context of disaster management, while at the same time tolerating higher latency.

C. Introduction of metrics

With benchmarks and thresholds in place, the next step is developing a systematic approach for evaluation. As of today, different works often choose to evaluate specific metrics, mostly with respect to the evaluation methodology that has been applied. As shown in Table I, a number of different algorithms and models, respectively, have been applied over the years, especially in the context of device-free approaches. While diversity of methodologies is a merit, as it encourages researchers to explore the applicability of existing methods in new contexts, as well as to create new methods, there is a strong need for a subset of common metrics against which each solution should be evaluated, to ensure ease of comparison.

D. Introduction of hybrid solutions

Although hybrid solutions have been explored both for device-based [69] and device-free [58]–[60] approaches, there is no complete study that explores all possible sensor combinations and outlines their common performance. Thus, current hybrid solutions may only serve as a demonstration of the potential of combining multiple technologies for people counting into a single solution. However, based on existing studies, it is not possible to determine whether there is a superior

[2]https://www.microsoft.com/en-us/research/event/microsoft-indoor-localization-competition-ipsn-2017/

[3]http://crawdad.org/

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/COMST.2019.2902824, IEEE Communications Surveys & Tutorials

combination of sensors and measurements to meet specific performance requirements, possibly on a per application basis. What is completely missing so far, is the combination of device-free and device-based technologies into a combined hybrid solution. The two different technologies can mutually compensate for each others disadvantages. For example, on the one hand not all people in an area may carry a device so device-based methods are only able to count a subset of all occupants but potentially with high accuracy. On the other hand, device-free technologies might be able to capture the overall occupancy but with lower accuracy, which can be corrected by the accurate partial knowledge about occupancy obtained from the device-based method, i.e., by means of a kind of extrapolation.

We note that there are emerging solutions combining image and non-image-based methods to improve the accuracy of occupancy monitoring in smart buildings. However, most of these works do not aim at providing a crowd estimate. For instance, in [107] Meyn et al. aim at detecting occupancy by analyzing information from CO_2-sensors, digital video cameras, and PIR detectors. In [108] Umetani et al. introduce a system which locates people based on fused information from wireless base stations and cameras. For an extensive survey of such fusion approaches, we refer the interested reader to the work of Akkaya et al. [109].

E. Introduction of tier models

Existing solutions mostly pertain to estimating occupancy in a limited space (either indoors or outdoors), and as such can be considered as representatives of micro-models. However, a number of scenarios (e.g., disaster management or crowd estimation for a city-wide public transportation system) may require a system view crowd count based on the local estimates calculated over different subareas. Thus, macro-models need to be in place, and methods need to be devised for transitioning between micro and macro estimates. These methods have further to be mapped to the specific context to provide the required context-specific accuracy even for the overall macro estimate obtained from the combination of micro estimates.

F. Introduction of scalable solutions

Existing solutions are mostly evaluated in scenarios with limited number of participants, often less than 50 (see Table VI and VII). On the one hand, such small scale evaluation setups may be sufficient for some applications in the context of building automation. On the other hand, for a people counting solution to be widely adopted across various application areas, as well as within a single application area, it should be proven scalable for larger crowds. Such evaluations are however hard to implement as they often require recruitment of larger crowds. One possible approach for researchers to perform evaluation of their proposed solutions is to make use of large public events and gatherings such as events attracting more than 100,000 people [81]. In such cases, the design and reproducibility of experiments should be taken into account. Alternatively, solutions that cannot be tested at a larger scale, should be evaluated via simulations using realistic representations of mobility of participants, and their environment [87].

G. Introduction of reliable solutions

History shows that oftentimes people-count estimates have been distorted due to the introduction of intentional bias (recall the examples in the beginning of this paper). In order to assure that the estimated people count is representative of the real occupancy level at any given location, crowd counting solutions should be reliable and tamper-proof by design. However, none of the existing academic solutions takes reliability into consideration (see Table VI and VII) as they all assume that all participants in the system are not malicious. Thus, extra measures should be taken into account on top of existing counting systems to provide better authentication and to deal with potential attackers.

VIII. Conclusion

Providing accurate people-count estimation, potentially in real-time, both for indoor and outdoor environments, is expected to be a game-changer for many application areas within the context of smart cities, such as public transportation, urban analytics, building automation as well as disaster management. Traditionally, significant efforts have been made towards providing people counting solutions based on images captured by surveillance cameras. However, image-based approaches are costly, as they require devoted hardware installations, and are often privacy intruding. Thus, academic and industry researchers have been looking into alternative solutions for people counting.

In this work we presented a comprehensive study of non-image based people counting techniques. We identified key requirements for providing accurate and reliable solutions for people counting in the context of smart cities. Moreover, we devised an extensive novel classification of existing approaches and discussed in great detail advantages and downsides of state-of-the-art solutions, demonstrating that as of today there is no ubiquitous solution for non-image based people counting. We further presented a comparison between various academic approaches and showed that the domain of people counting is rather siloed with respect to potential applications that make use of the crowd size estimation. Next, we discussed available industrial solutions, and concluded that the counting techniques favoured by the industry often differ from solutions proposed by academia. Finally, we outlined open challenges and future directions that could help advance the field of non-image based people counting.

Acknowledgment

The authors would like to thank the Stockholm County Council (SLL) (research grant LS2016-1423) for their financial support.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/COMST.2019.2902824, IEEE Communications Surveys & Tutorials

TABLE VI

OVERVIEW IF RESPECTIVE WORKS INVESTIGATE THE REQUIREMENTS STATED IN SECTION II-A. THE TYPE (SIMULATION, EXPERIMENT OR PATENT) AND SCALE OF EVALUATION ARE GIVEN. THE REQUIREMENTS ARE EITHER MARKED AS INVESTIGATED (✓) OR NOT INVESTIGATED (—). THE FOLLOWING LEGEND APPLIES TO THE REQUIREMENTS: AC = ACCURACY, S = SCALABILITY, C = COST, L = LATENCY, AU = AUTOMATION, P = PRIVACY, R = RELIABILITY.

	Class	**Category**	**Paper**	**Evaluation**		**Requirements**						
				Type	Scale	Ac	S	C	L	Au	P	R
Device-free	RF	RSSI	[29]	Exp.	23-29	—	—	✓	✓	✓	✓	—
			[30,31]	Exp.	4	✓	—	✓	✓	✓	✓	—
			[21,32]	Sim. & Exp.	400	—	✓	✓	—	✓	✓	—
			[27]	Exp.	7	✓	—	✓	✓	✓	✓	—
			[28]	Exp.	9	✓	—	✓	✓	✓	✓	—
			[33]	Exp.	50	✓	—	✓	✓	✓	✓	—
			[26]	Exp.	15	—	—	✓	—	✓	✓	—
		CSI	[22]	Exp.	30	✓	—	✓	—	✓	✓	—
			[24]	Exp.	7	✓	—	✓	✓	✓	✓	—
		UWB	[34]	Exp.	3	✓	—	✓	—	✓	✓	—
			[35]	—	—	—	✓	✓	—	✓	✓	—
			[36]	Exp.	43	✓	—	✓	✓	✓	✓	—
			[38]	Sim.	30	✓	✓	✓	—	✓	✓	—
			[37]	Exp.	8881	✓	✓	✓	✓	✓	✓	—
			[39]	Exp.	10	✓	—	✓	✓	✓	✓	—
	Sensor	PIR	[40]	Sim. & Exp.	6	—	—	✓	✓	✓	✓	—
			[43]	Exp.	5	✓	—	✓	✓	✓	—	—
			[41]	Exp.	39	✓	—	✓	✓	✓	✓	—
			[42]	Exp.	14	✓	—	✓	✓	✓	✓	—
		LED	[44]	Exp.	2	✓	—	✓	✓	✓	✓	—
			[45]	Exp.	20	✓	—	✓	✓	✓	✓	—
		Acoustic	[20]	Exp.	2	✓	—	✓	✓	✓	✓	—
			[46]	Exp.	50	✓	—	✓	✓	✓	✓	—
			[47]	Sim.	200	✓	—	✓	✓	✓	—	—
		CO_2	[48]	Sim.	10-80	✓	✓	—	✓	—	—	—
			[49]	Exp.	25	✓	—	✓	✓	✓	✓	—
			[50]	Exp.	80	—	—	—	—	✓	—	—
			[51]	Sim. & Exp.	7	✓	—	—	✓	✓	—	—
			[52]	Exp.	42	✓	—	✓	✓	✓	✓	—
			[53]	Exp.	300	✓	—	✓	✓	✓	✓	—
		LiDAR	[55]	Exp.	4	✓	—	✓	✓	✓	—	—
			[54]	Exp.	11	—	—	✓	✓	✓	✓	—
			[56]	Exp.	3	✓	—	✓	✓	✓	✓	—
		Hybrid	[58]	Exp.	3	✓	—	✓	✓	✓	✓	—
			[59]	Exp.	9	✓	—	✓	✓	✓	—	—
			[57]	Exp.	4	✓	—	✓	✓	✓	—	—
			[60]	Exp.	12	✓	—	✓	✓	✓	—	—

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/COMST.2019.2902824, IEEE Communications Surveys & Tutorials

TABLE VII

(CONTINUED.) OVERVIEW IF RESPECTIVE WORKS INVESTIGATE THE REQUIREMENTS STATED IN SECTION II-A. THE TYPE (SIMULATION, EXPERIMENT OR PATENT) AND SCALE OF EVALUATION ARE GIVEN. THE REQUIREMENTS ARE EITHER MARKED AS INVESTIGATED (✓) OR NOT INVESTIGATED (—). THE FOLLOWING LEGEND APPLIES TO THE REQUIREMENTS: AC = ACCURACY, S = SCALABILITY, C = COST, L = LATENCY, AU = AUTOMATION, P = PRIVACY, R = RELIABILITY.

	Class	Category	Paper	Evaluation		Requirements						
				Type	Scale	Ac	S	C	L	Au	P	R
Device-based	Acoustic	Pure	[66]	Sim. & Exp.	≤ 25	✓	✓	✓	✓	✓	✓	—
			[67,68]	Exp.	≤ 10	✓	—	—	—	✓	✓	—
		Hybrid	[69]	Exp.	≤ 10	✓	—	—	—	✓	—	—
	Network	RFID	[70]	Sim.	5,000	—	✓	✓	✓	✓	✓	—
			[71]	Exp.	6	✓	—	—	—	—	—	—
		WSN	[72]	Exp.	100	✓	—	—	✓	✓	—	—
			[73]	Sim. & Exp.	15-500	✓	—	—	✓	✓	—	—
		Wi-Fi	[74]	Exp.	≤ 52	✓	✓	✓	✓	✓	✓	—
			[75]	Exp.	850	✓	—	✓	—	✓	—	—
			[76]	Exp.	n/a	✓	—	—	✓	✓	✓	—
			[77]	Exp.	116	✓	—	—	✓	✓	✓	—
			[78]	Exp.	25	✓	—	—	✓	✓	—	—
			[79]	Exp.	60	✓	—	✓	✓	✓	—	—
		Bluetooth	[80]	Exp.	1000	—	✓	✓	—	—	✓	—
			[81]	Exp.	100,000	—	✓	✓	—	✓	✓	—
			[82,83]	Exp.	≤ 5200	✓	✓	—	—	—	—	—
			[84]	Exp.	20,000	✓	—	✓	—	—	✓	—
		Cellular	[85]	Patent	—	—	—	✓	—	✓	—	—
		Collaborative	[86]	Sim.	2400	—	✓	—	—	—	—	—
			[87]	Sim.	2400	✓	✓	—	✓	✓	✓	—

REFERENCES

[1] Wikipedia, "Million man march." [Online]. Available: https://en.wikipedia.org/wiki/Million_Man_March

[2] BBC News, "Tiananmen anniversary marked at huge hong kong vigil," 2014. [Online]. Available: http://www.bbc.com/news/world-asia-china-27702206

[3] Guardian, "Hundreds of thousands join macedonia name protest in athens," 2018. [Online]. Available: https://www.theguardian.com/world/2018/feb/04/hundreds-of-thousands-join-macedonia-name-protest-in-athens

[4] TIME Magazine, "Reporting: The perils of crowd counting," 1967. [Online]. Available: http://content.time.com/time/magazine/article/0,9171,843533,00.html

[5] W. Kleiminger, C. Beckel, T. Staake, and S. Santini, "Occupancy detection from electricity consumption data," in *Proceedings of the 5th ACM Workshop on Embedded Systems For Energy-Efficient Buildings*, ser. BuildSys'13. New York, NY, USA: ACM, 2013, pp. 10:1–10:8.

[6] D. Chen, S. Barker, A. Subbaswamy, D. Irwin, and P. Shenoy, "Non-intrusive occupancy monitoring using smart meters," in *Proceedings of the 5th ACM Workshop on Embedded Systems For Energy-Efficient Buildings*, ser. BuildSys'13. New York, NY, USA: ACM, 2013, pp. 9:1–9:8.

[7] S. K. Ghai, L. V. Thanayankizil, D. P. Seetharam, and D. Chakraborty, "Occupancy detection in commercial buildings using opportunistic context sources," in *2012 IEEE International Conference on Pervasive Computing and Communications Workshops*, March 2012, pp. 463–466.

[8] T. W. Hnat, E. Griffiths, R. Dawson, and K. Whitehouse, "Doorjamb: Unobtrusive room-level tracking of people in homes using doorway sensors," in *Proceedings of the 10th ACM Conference on Embedded Network Sensor Systems*, ser. SenSys '12. New York, NY, USA: ACM, 2012, pp. 309–322.

[9] Y. Agarwal, B. Balaji, R. Gupta, J. Lyles, M. Wei, and T. Weng, "Occupancy-driven energy management for smart building automation," in *Proceedings of the 2Nd ACM Workshop on Embedded Sensing Systems for Energy-Efficiency in Building*, ser. BuildSys '10. New York, NY, USA: ACM, 2010, pp. 1–6. [Online]. Available: http://doi.acm.org/10.1145/1878431.1878433

[10] P. Karpagavalli and A. V. Ramprasad, "Estimating the density of the people and counting the number of people in a crowd environment for human safety," in *Communications and Signal Processing (ICCSP), 2013 International Conference on*, April 2013, pp. 663–667.

[11] M. Deng, Y. Xu, P. Jiang, and X. Yang, "Real time and scene invariant crowd counting: Across a line or inside a region," in *Multimedia Signal Processing (MMSP), 2015 IEEE 17th International Workshop on*, Oct 2015, pp. 1–6.

[12] V. Q. Pham, T. Kozakaya, O. Yamaguchi, and R. Okada, "Count forest: Co-voting uncertain number of targets using random forest for crowd density estimation," in *2015 IEEE International Conference on Computer Vision (ICCV)*, Dec 2015, pp. 3253–3261.

[13] B. Zhan, D. N. Monekosso, P. Remagnino, S. A. Velastin, and L.-Q. Xu, "Crowd analysis: a survey," *Machine Vision and Applications*, vol. 19, no. 5, pp. 345–357, 2008.

[14] J. C. S. J. Junior, S. R. Musse, and C. R. Jung, "Crowd analysis using computer vision techniques," *IEEE Signal Processing Magazine*, vol. 27, no. 5, pp. 66–77, Sept 2010.

[15] T. Li, H. Chang, M. Wang, B. Ni, R. Hong, and S. Yan, "Crowded scene analysis: A survey," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 25, no. 3, pp. 367–386, March 2015.

[16] V. Murino, M. Cristani, S. Shah, and S. Savarese, *Group and Crowd Behavior for Computer Vision*, 1st ed. Academic Press, Elsevier, Apr 2017.

[17] S. A. M. Saleh, S. A. Suandi, and H. Ibrahim, "Recent survey on crowd density estimation and counting for visual surveillance," *Engineering Applications of Artificial Intelligence*, vol. 41, pp. 103 – 114, 2015.

[18] A. Kowcika and S. Sridhar, "A literature survey on crowd (people)

counting with the help of surveillance videos," *International Journal of Information Technology and Decision Making*, vol. 3, no. 4, pp. 2353–2361, June 2015.

[19] D. Ryan, S. Denman, S. Sridharan, and C. Fookes, "An evaluation of crowd counting methods, features and regression models," *Computer Vision and Image Understanding*, vol. 130, pp. 1 – 17, 2015.

[20] Q. Chen, M. Gao, J. Ma, D. Zhang, L. Ni, and Y. Liu, "Mocus: moving object counting using ultrasonic sensor networks," *International Journal of Sensor Networks*, vol. 3, no. 1, pp. 55–65, 2008.

[21] Y. Yuan, J. Zhao, C. Qiu, and W. Xi, "Estimating crowd density in an rf-based dynamic environment," *IEEE Sensors Journal*, vol. 13, no. 10, pp. 3837–3845, Oct 2013.

[22] W. Xi, J. Zhao, X.-Y. Li, K. Zhao, S. Tang, X. Liu, and Z. Jiang, "Electronic frog eye: Counting crowd using wifi," in *INFOCOM, 2014 Proceedings IEEE*, April 2014, pp. 361–369.

[23] T. Xin, B. Guo, Z. Wang, M. Li, and Z. Yu, "Freesense: Indoor human identification with wifi signals," *CoRR*, vol. abs/1608.03430, 2016. [Online]. Available: http://arxiv.org/abs/1608.03430

[24] S. Di Domenico, M. De Sanctis, E. Cianca, and G. Bianchi, "A trained-once crowd counting method using differential wifi channel state information," in *Proceedings of the 3rd International on Workshop on Physical Analytics*, ser. WPA '16. New York, NY, USA: ACM, 2016, pp. 37–42. [Online]. Available: http://doi.acm.org/10.1145/2935651.2935657

[25] M. Irfan, L. Marcenaro, and L. Tokarchuk, "Crowd analysis using visual and non-visual sensors, a survey," in *2016 IEEE Global Conference on Signal and Information Processing (GlobalSIP)*, Dec 2016, pp. 1249–1254.

[26] S. Y. Fadhlullah and W. Ismail, "A statistical approach in designing an rf-based human crowd density estimation system," *International Journal of Distributed Sensor Networks*, 2016.

[27] T. Yoshida and Y. Taniguchi, "Estimating the number of people using existing wifi access point in indoor environment," in *Proceedings of the 6th European Conference of Computer Science (ECCS15)*, 2015, pp. 46–53.

[28] S. Depatla, A. Muralidharan, and Y. Mostofi, "Occupancy estimation using only wifi power measurements," *IEEE Journal on Selected Areas in Communications*, vol. 33, no. 7, pp. 1381–1393, July 2015.

[29] M. Nakatsuka, H. Iwatani, and J. Katto, "A study on passive crowd density estimation using wireless sensors," in *The 4th Intl. Conf. on Mobile Computing and Ubiquitous Networking (ICMU 2008)*. Citeseer, 2008.

[30] C. Xu, B. Firner, Y. Zhang, R. Howard, J. Li, and X. Lin, "Improving rf-based device-free passive localization in cluttered indoor environments through probabilistic classification methods," in *Proceedings of the 11th International Conference on Information Processing in Sensor Networks*, ser. IPSN '12. New York, NY, USA: ACM, 2012, pp. 209–220.

[31] C. Xu, B. Firner, R. S. Moore, Y. Zhang, W. Trappe, R. Howard, F. Zhang, and N. An, "Scpl: Indoor device-free multi-subject counting and localization using radio signal strength," in *2013 ACM/IEEE International Conference on Information Processing in Sensor Networks (IPSN)*, April 2013, pp. 79–90.

[32] Y. Yuan, C. Qiu, W. Xi, and J. Zhao, "Crowd density estimation using wireless sensor networks," in *Mobile Ad-hoc and Sensor Networks (MSN), 2011 Seventh International Conference on*. IEEE, 2011, pp. 138–145.

[33] 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 Transactions on Human-Machine Systems*, vol. 45, no. 4, pp. 442–452, Aug 2015.

[34] J. W. Choi, J. H. Kim, and S. H. Cho, "A counting algorithm for multiple objects using an ir-uwb radar system," in *Network Infrastructure and Digital Content (IC-NIDC), 2012 3rd IEEE International Conference on*. IEEE, 2012, pp. 591–595.

[35] 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 *Electronics, Information and Communications (ICEIC), 2014 International Conference on*. IEEE, 2014, pp. 1–2.

[36] J. He and A. Arora, "A regression-based radar-mote system for people counting," in *Pervasive Computing and Communications (PerCom), 2014 IEEE International Conference on*. IEEE, 2014, pp. 95–102.

[37] J. W. Choi, X. Quan, and S. H. Cho, "Bi-directional passing people counting system based on ir-uwb radar sensors," *IEEE Internet of Things Journal*, 2017.

[38] S. Bartoletti, A. Conti, and M. Z. Win, "Device-free counting via wideband signals," *IEEE Journal on Selected Areas in Communications*, vol. PP, no. 99, pp. 1–1, 2017.

[39] J. W. Choi, D. H. Yim, and S. H. Cho, "People counting based on an ir-uwb radar sensor," *IEEE Sensors Journal*, vol. 17, no. 17, pp. 5717–5727, Sept 2017.

[40] F. Wahl, M. Milenkovic, and O. Amft, "A distributed pir-based approach for estimating people count in office environments," in *2012 IEEE 15th International Conference on Computational Science and Engineering*, Dec 2012, pp. 640–647.

[41] G. Suthokumar, J. Tharmakulasingam, A. Johnnirajh, P. Subendran, and D. Dias, "Smart building — towards the insight remotely," in *2016 International Conference on Internet of Things and Applications (IOTA)*, Jan 2016, pp. 395–400.

[42] Y. P. Raykov, E. Ozer, G. Dasika, A. Boukouvalas, and M. A. Little, "Predicting room occupancy with a single passive infrared (pir) sensor through behavior extraction," in *Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing*, ser. UbiComp '16. New York, NY, USA: ACM, 2016, pp. 1016–1027. [Online]. Available: http://doi.acm.org/10.1145/2971648.2971746

[43] F. Krüger, M. Kasparick, T. Mundt, and T. Kirste, "Where are my colleagues and why? tracking multiple persons in indoor environments," in *2014 International Conference on Intelligent Environments*, June 2014, pp. 190–197.

[44] Q. Wang, X. Zhang, and K. L. Boyer, "Occupancy distribution estimation for smart light delivery with perturbation-modulated light sensing," *Journal of solid state lighting*, vol. 1, no. 1, p. 17, 2014.

[45] Y. Yang, J. Hao, J. Luo, and S. J. Pan, "Ceilingsee: Device-free occupancy inference through lighting infrastructure based led sensing," in *Proceedings of the IEEE International Conference on Pervasive Computing and Communications*, ser. PerCom '17, 2017.

[46] O. Shih and A. Rowe, "Occupancy estimation using ultrasonic chirps," in *Proceedings of the ACM/IEEE Sixth International Conference on Cyber-Physical Systems*, ser. ICCPS '15. New York, NY, USA: ACM, 2015, pp. 149–158.

[47] Q. Huang, Z. Ge, and C. Lu, "Occupancy estimation in smart buildings using audio-processing techniques," *arXiv preprint arXiv:1602.08507*, 2016.

[48] S. Wang and X. Jin, "Co_2-based occupancy detection for on-line outdoor air flow control," *Indoor + Built Environment*, vol. 7, pp. 165–181, 1998.

[49] S. A. Mumma, "Transient occupancy ventilation by monitoring co_2," in *ASHRAE IAQ 2004: Critical Operations: Supporting the Healing Environment Through IAQ Performance Standards*, 2004, pp. 21–23.

[50] Z. Sun, S. Wang, and Z. Ma, "In-situ implementation and validation of a co2-based adaptive demand-controlled ventilation strategy in a multi-zone office building," *Building and Environment*, vol. 46, no. 1, pp. 124 – 133, 2011. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0360132310002131

[51] M. Jin, N. Bekiaris-Liberis, K. Weekly, C. Spanos, and A. Bayen, "Sensing by proxy: Occupancy detection based on indoor co2 concentration," *UBICOMM 2015*, vol. 14, 2015.

[52] C. Basu, C. Koehler, K. Das, and A. K. Dey, "Perccs: Person-count from carbon dioxide using sparse non-negative matrix factorization," in *Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing*, ser. UbiComp '15. New York, NY, USA: ACM, 2015, pp. 987–998.

[53] I. B. Arief-Ang, F. D. Salim, and M. Hamilton, "Cd-hoc: Indoor human occupancy counting using carbon dioxide sensor data," *arXiv preprint arXiv:1706.05286*, 2017.

[54] M. Hashimoto, A. Tsuji, A. Nishio, and K. Takahashi, "Laser-based tracking of groups of people with sudden changes in motion," in *2015 IEEE International Conference on Industrial Technology (ICIT)*, March 2015, pp. 315–320.

[55] M. Hashimoto, R. Izumi, Y. Tamura, and K. Takahashi, "Laser-based tracking of people and vehicles by multiple mobile robots," in *Informatics in Control, Automation and Robotics (ICINCO), 2014 11th International Conference on*, vol. 2. IEEE, 2014, pp. 522–527.

[56] S. Kanaki, K. Fujishita, M. Hashimoto, R. Murabayashi, K. Inui, and K. Takahashi, "Cooperative moving-object tracking with multiple

mobile sensor nodes — size and posture estimation of moving objects using in-vehicle multilayer laser scanner," in *2016 IEEE International Conference on Industrial Technology (ICIT)*, March 2016, pp. 59–64.

[57] A. Ebadat, G. Bottegal, D. Varagnolo, B. Wahlberg, and K. H. Johansson, "Estimation of building occupancy levels through environmental signals deconvolution," in *Proceedings of the 5th ACM Workshop on Embedded Systems For Energy-Efficient Buildings*, ser. BuildSys'13. New York, NY, USA: ACM, 2013, pp. 8:1–8:8.

[58] K. P. Lam, M. Hynck, B. Dong, B. Andrews, Y. shang Chiou, D. Benitez, and J. Choi, "Occupancy detection through an extensive environmental sensor network in an open-plan office building," in *Proc. of Building Simulation 09, an IBPSA Conference*, 2009.

[59] Z. Yang, N. Li, B. Becerik-Gerber, and M. Orosz, "A multi-sensor based occupancy estimation model for supporting demand driven hvac operations," in *Proceedings of the 2012 Symposium on Simulation for Architecture and Urban Design*, ser. SimAUD '12. San Diego, CA, USA: Society for Computer Simulation International, 2012, pp. 2:1–2:8.

[60] L. Yang, K. Ting, and M. B. Srivastava, "Inferring occupancy from opportunistically available sensor data," in *2014 IEEE International Conference on Pervasive Computing and Communications (PerCom)*, March 2014, pp. 60–68.

[61] R. Niu and P. K. Varshney, "Target location estimation in sensor networks with quantized data," *IEEE Transactions on Signal Processing*, vol. 54, no. 12, pp. 4519–4528, Dec 2006.

[62] 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 Transactions on Aerospace and Electronic Systems*, vol. 48, no. 2, pp. 1466–1489, APRIL 2012.

[63] M. Z. Win and R. A. Scholtz, "Impulse radio: how it works," *IEEE Communications Letters*, vol. 2, no. 2, pp. 36–38, Feb 1998.

[64] S. Gezici and H. V. Poor, "Position estimation via ultra-wide-band signals," *Proceedings of the IEEE*, vol. 97, no. 2, pp. 386–403, 2009.

[65] iris (Infrared Intelligent Sensors), "Irma (infrared motion analyzer) for automatic passenger counting," 2017. [Online]. Available: http://www.irisgmbh.de/en/products/

[66] P. G. Kannan, S. P. Venkatagiri, M. C. Chan, A. L. Ananda, and L.-S. Peh, "Low cost crowd counting using audio tones," in *Proceedings of the 10th ACM Conference on Embedded Network Sensor Systems*, ser. SenSys '12. New York, NY, USA: ACM, 2012, pp. 155–168. [Online]. Available: http://doi.acm.org/10.1145/2426656.2426673

[67] C. Xu, S. Li, G. Liu, Y. Zhang, E. Miluzzo, Y.-F. Chen, J. Li, and B. Firner, "Crowd++: Unsupervised speaker count with smartphones," in *Proceedings of the 2013 ACM International Joint Conference on Pervasive and Ubiquitous Computing*, ser. UbiComp '13. New York, NY, USA: ACM, 2013, pp. 43–52. [Online]. Available: http://doi.acm.org/10.1145/2493432.2493435

[68] C. Xu, S. Li, Y. Zhang, E. Miluzzo, and Y. f. Chen, "Crowdsensing the speaker count in the wild: implications and applications," *IEEE Communications Magazine*, vol. 52, no. 10, pp. 92–99, October 2014.

[69] M. A. A. H. Khan, H. M. S. Hossain, and N. Roy, "Sensepresence: Infrastructure-less occupancy detection for opportunistic sensing applications," in *Mobile Data Management (MDM), 2015 16th IEEE International Conference on*, vol. 2, June 2015, pp. 56–61.

[70] J. Weaver, F. Dawood, and A. Alkhalidi, "Modelling of radiation field of patch antenna for use in long range rfid crowd monitoring," in *Computer Applications Technology (ICCAT), 2013 International Conference on*. IEEE, 2013, pp. 1–4.

[71] N. Li, G. Calis, and B. Becerik-Gerber, "Measuring and monitoring occupancy with an rfid based system for demand-driven hvac operations," *Automation in Construction*, vol. 24, pp. 89 – 99, 2012. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0926580512000283

[72] M. Cattani, M. Zuniga, A. Loukas, and K. Langendoen, "Lightweight neighborhood cardinality estimation in dynamic wireless networks," in *Proceedings of the 13th international symposium on Information processing in sensor networks*. IEEE Press, 2014, pp. 179–189.

[73] A. Kamra, V. Misra, and D. Rubenstein, "Counttorrent: Ubiquitous access to query aggregates in dynamic and mobile sensor networks," in *Proceedings of the 5th International Conference on Embedded Networked Sensor Systems*, ser. SenSys '07. New York, NY, USA: ACM, 2007, pp. 43–57.

[74] M. Handte, M. U. Iqbal, S. Wagner, W. Apolinarski, P. J. Marrn, E. M. Muoz, S. Martinez, I. Y. Desarrollo, S. Izquierdo, M. Gonzlez, and P. Europeos, "Crowd density estimation for public transport vehicles," in *Workshop Proceedings of the EDBT/ICDT 2014 Joint Conference*, 2014.

[75] O. B. Yaik, K. Z. Wai, I. K.T.Tan, and O. B. Sheng, "Measuring the accuracy of crowd counting using wi-fi probe-request-frame counting technique," *Journal of Telecommunication Electronic and Computer Engineering*, vol. 8, no. 2, 2016.

[76] G. Fierro, O. Rehmane, A. Krioukov, and D. Culler, "Zone-level occupancy counting with existing infrastructure," in *Proceedings of the Fourth ACM Workshop on Embedded Sensing Systems for Energy-Efficiency in Buildings*, ser. BuildSys '12. New York, NY, USA: ACM, 2012, pp. 205–206.

[77] B. Balaji, J. Xu, A. Nwokafor, R. Gupta, and Y. Agarwal, "Sentinel: Occupancy based hvac actuation using existing wifi infrastructure within commercial buildings," in *Proceedings of the 11th ACM Conference on Embedded Networked Sensor Systems*, ser. SenSys '13. New York, NY, USA: ACM, 2013, pp. 17:1–17:14.

[78] X. Lu, H. Wen, H. Zou, H. Jiang, L. Xie, and N. Trigoni, "Robust occupancy inference with commodity wifi," in *2016 IEEE 12th International Conference on Wireless and Mobile Computing, Networking and Communications (WiMob)*, Oct 2016, pp. 1–8.

[79] R. Melfi, B. Rosenblum, B. Nordman, and K. Christensen, "Measuring building occupancy using existing network infrastructure," in *2011 International Green Computing Conference and Workshops*, July 2011, pp. 1–8.

[80] V. Kostakos, T. Camacho, and C. Mantero, "Wireless detection of end-to-end passenger trips on public transport buses," in *13th International IEEE Conference on Intelligent Transportation Systems*, Sept 2010, pp. 1795–1800.

[81] M. Versichele, T. Neutens, M. Delafontaine, and N. V. de Weghe, "The use of bluetooth for analysing spatiotemporal dynamics of human movement at mass events: A case study of the ghent festivities," *Applied Geography*, vol. 32, no. 2, pp. 208 – 220, 2012.

[82] J. Weppner and P. Lukowicz, "Collaborative crowd density estimation with mobile phones," *Proc. of ACM PhoneSense*, 2011.

[83] ——, "Bluetooth based collaborative crowd density estimation with mobile phones," in *2013 IEEE International Conference on Pervasive Computing and Communications (PerCom)*, March 2013, pp. 193–200.

[84] J. Weppner, P. Lukowicz, U. Blanke, and G. Trster, "Participatory bluetooth scans serving as urban crowd probes," *IEEE Sensors Journal*, vol. 14, no. 12, pp. 4196–4206, Dec 2014.

[85] J. Ramachandran, "Systems, methods, and computer program products for estimating crowd sizes using information collected from mobile devices in a wireless communications network," Dec. 1 2011, uS Patent App. 12/791,463. [Online]. Available: https://www.google.com/patents/US20110295577

[86] L. Pajevic and G. Karlsson, "Characterizing opportunistic communication with churn for crowd-counting," in *World of Wireless, Mobile and Multimedia Networks (WoWMoM), 2015 IEEE 16th International Symposium on a*, June 2015, pp. 1–6.

[87] P. Danielis, S. T. Kouyoumdjieva, and G. Karlsson, "Urbancount: Mobile crowd counting in urban environments," in *2017 8th IEEE Annual Information Technology, Electronics and Mobile Communication Conference (IEMCON)*, Oct 2017, pp. 640–648.

[88] N. Riga, I. Matta, and A. Bestavros, "Dip: Density inference protocol for wireless sensor networks and its application to density-unbiased statistics," Boston University, Tech. Rep. BUCS-TR-2004-023, May 2004.

[89] O. Landsiedel, F. Ferrari, and M. Zimmerling, "Chaos: Versatile and efficient all-to-all data sharing and in-network processing at scale," in *Proceedings of the 11th ACM Conference on Embedded Networked Sensor Systems*. ACM, 2013, p. 1.

[90] N. Malpani, Y. Chen, N. H. Vaidya, and J. L. Welch, "Distributed token circulation in mobile ad hoc networks," *IEEE Transactions on Mobile Computing*, vol. 4, no. 2, pp. 154–165, March 2005.

[91] K. P. Hatzis, G. P. Pentaris, P. G. Spirakis, and B. Tampakas, "Counting in mobile networks: Theory and experimentation," in *Proceedings of the 3rd International Workshop on Algorithm Engineering*, ser. WAE '99. London, UK, UK: Springer-Verlag, 1999, pp. 95–109.

[92] M. V. Barbera, A. Epasto, A. Mei, V. C. Perta, and J. Stefa, "Signals from the crowd: Uncovering social relationships through smartphone probes," in *Proceedings of the 2013 Conference on Internet Measurement Conference*, ser. IMC '13. New

York, NY, USA: ACM, 2013, pp. 265–276. [Online]. Available: http://doi.acm.org/10.1145/2504730.2504742

[93] A. B. M. Musa and J. Eriksson, "Tracking unmodified smartphones using wi-fi monitors," in *Proceedings of the 10th ACM Conference on Embedded Network Sensor Systems*, ser. SenSys '12. New York, NY, USA: ACM, 2012, pp. 281–294. [Online]. Available: http://doi.acm.org/10.1145/2426656.2426685

[94] E. O'Neill, V. Kostakos, T. Kindberg, A. F. g. Schiek, A. Penn, D. S. Fraser, and T. Jones, "Instrumenting the city: Developing methods for observing and understanding the digital cityscape," in *Proceedings of the 8th International Conference on Ubiquitous Computing*, ser. UbiComp'06. Berlin, Heidelberg: Springer-Verlag, 2006, pp. 315–332.

[95] Eurotech, "Pcn-1001 optical passenger counter," 2017. [Online]. Available: https://www.eurotech.com/en/products/PCN-1001

[96] Flir, "Brickstream®people counting and tracking," 2017. [Online]. Available: http://www.brickstream.com/

[97] ViNotion, "Visense: Real-time statistics of people and traffic," 2017. [Online]. Available: https://www.visense.eu

[98] Trapeze, "Trapeze comfoaccess," 2017. [Online]. Available: http://www.trapezegroup.eu/trapeze-be-in-be-out-comfoaccess

[99] Eco Compteur, "Eco-compteur pyro-box solution," 2017. [Online]. Available: http://www.eco-compteur.com/en/products/pyro-range/pyro-box

[100] InfraRed Integrated Systems Ltd., "Smart building technology," 2017. [Online]. Available: http://www.irisys.net/smart-building-technology

[101] Xandar, "Xandar solutions ir-uwb based systems," 2017. [Online]. Available: http://www.xandartech.com/solutions/

[102] Acorel, "High precision sensors," 2018. [Online]. Available: https://acorel.com/en/products-and-services/high-precision-sensors/

[103] Dilax, "Automatic passenger counting," 2018. [Online]. Available: https://www.dilax.com/public-mobility/portfolio/automatic-passenger-counting/

[104] Airborn Robotics, "Counting software for plants and livstock," 2018. [Online]. Available: http://www.airborne-robotics.com/en/counting-software-plants-and-livestock

[105] Cattle Watch. [Online]. Available: http://www.cattle-watch.com/

[106] SCR by Allflex, "Dairy cow monitoring and herd management solutions," 2013. [Online]. Available: http://www.scrdairy.com/

[107] S. Meyn, A. Surana, Y. Lin, S. M. Oggianu, S. Narayanan, and T. A. Frewen, "A sensor-utility-network method for estimation of occupancy in buildings," in *Decision and Control, 2009 held jointly with the 2009 28th Chinese Control Conference. CDC/CCC 2009. Proceedings of the 48th IEEE Conference on*. IEEE, 2009, pp. 1494–1500.

[108] T. Umetani, S. Yamane, and Y. Tamura, "Mobile client localization using data fusion of wireless lan signals and three-dimensional camera image sequences," in *Ubiquitous Robots and Ambient Intelligence (URAI), 2012 9th International Conference on*. IEEE, 2012, pp. 240–245.

[109] K. Akkaya, I. Guvenc, R. Aygun, N. Pala, and A. Kadri, "Iot-based occupancy monitoring techniques for energy-efficient smart buildings," in *Wireless Communications and Networking Conference Workshops (WCNCW), 2015 IEEE*. IEEE, 2015, pp. 58–63.