

---

This is an electronic reprint of the original article.  
This reprint may differ from the original in pagination and typographic detail.

Palipana, Sameera; Malm, Nicolas; Sigg, Stephan

## Beamsteering for Training-free Counting of Multiple Humans Performing Distinct Activities

*Published in:*

18th Annual IEEE International Conference on Pervasive Computing and Communications, PerCom 2020

*DOI:*

[10.1109/PerCom45495.2020.9127374](https://doi.org/10.1109/PerCom45495.2020.9127374)

Published: 01/03/2020

*Document Version*

Peer-reviewed accepted author manuscript, also known as Final accepted manuscript or Post-print

*Please cite the original version:*

Palipana, S., Malm, N., & Sigg, S. (2020). Beamsteering for Training-free Counting of Multiple Humans Performing Distinct Activities. In *18th Annual IEEE International Conference on Pervasive Computing and Communications, PerCom 2020* Article 9127374 IEEE. <https://doi.org/10.1109/PerCom45495.2020.9127374>

---

This material is protected by copyright and other intellectual property rights, and duplication or sale of all or part of any of the repository collections is not permitted, except that material may be duplicated by you for your research use or educational purposes in electronic or print form. You must obtain permission for any other use. Electronic or print copies may not be offered, whether for sale or otherwise to anyone who is not an authorised user.

# Beamsteering for Training-free Counting of Multiple Humans Performing Distinct Activities

Sameera Palipana, Nicolas Malm, Stephan Sigg

Department of Communications and Networking, Aalto University, Espoo, Finland

Email: `firstname.lastname@aalto.fi`

**Abstract**—Recognition of the context of humans plays an important role in pervasive applications such as intrusion detection, human density estimation for heating, ventilation and air-conditioning in smart buildings, as well as safety guarantee for workers during human-robot interaction. Radio vision is able to provide these sensing capabilities with low privacy intrusion. A common challenge though, for current radio sensing solutions is to distinguish simultaneous movement from multiple subjects. We present an approach that exploits antenna installations, for instance, found in upcoming 5G technology, to detect and extract activities from spatially scattered human targets in an ad-hoc manner in arbitrary environments and without prior training of the multi-subject detection. We perform receiver-side beamforming and beam-sweeping over different azimuth angles to detect human presence in those regions separately. We characterize the resultant fluctuations in the spatial streams due to human influence using a case study and make the traces publicly available. We demonstrate the potential of this approach through two applications: 1) By feeding the similarities of the resulting spatial streams into a clustering algorithm, we count the humans in a given area without prior training. (up to 6 people in a 22.4 m<sup>2</sup> area with an accuracy that significantly exceeds the related work). 2) We demonstrate that simultaneously conducted activities and gestures can be extracted from the spatial streams through blind source separation.

**Index Terms**—Radio sensing, Training-free crowd counting, Multi-subject recognition, Beamsteering

## I. INTRODUCTION

Our research explores passive radio sensing to analyse human presence, crowd size and to track and eventually recognize motion and interaction. Such human context can be unearthed continuously and without requiring human subjects to carry any devices. In contrast to vision sensors, privacy is ensured through unobtrusive RF-sensing that can be integrated with existing indoor WiFi/cellular installations simply through software updates. We leverage ambient (or stray) radio signal streams like those found in cellular 4G/5G and WiFi.

A common challenge in traditional, narrowband radio sensing with limited antennas is to distinguish simultaneous movement from multiple subjects [1]. Essentially, perturbations in the signal strength are typically analyzed at a receiver and interpreted as activities, gestures or other relevant motions [2], [3]. Perturbations caused by reflections and scattering off targets result in amplitude fluctuations, frequency or phase shift due to movement. In the case of multiple subjects, reflected signals are superimposed (constructively or destructively) and it is not a trivial task to distinguish individual movements apart [4].

As a solution, we present a sensing scheme that exploits the limited antennas to detect and extract activities from multiple persons simultaneously. This is achieved by steering the receiving antenna array’s directivity over the azimuth angle of the area of interest (receiver beamsteering) and collecting spatial streams representative of different regions. This is illustrated in Fig. 1. The spatial streams can have different levels of correlations (not fully orthogonal as in the figure) depending on the antenna array size, sweeping resolution and the distance of human subjects to each other. Therefore, in this work we first collect data traces from experiments where multiple persons simultaneously perform activities. We analyze these data traces and identify the correlations and anomalies in the spatial streams and explain the reasoning for such behaviour using models. Next, we transfer this knowledge for multiple human recognition. We demonstrate the potential of our approach specially with a low antenna count and narrowband signals to infer human context using two applications: 1) *training-free counting of human subjects*, and 2) *blind extraction of activities of an individual* when multiple people perform different activities simultaneously. People counting and simultaneous inference of their activities are important problems in many pervasive application domains. In surveillance systems, automatic people counting is a pre-condition to further processing [5], for instance, to estimate capacity at large public events [6] and to plan procedures during emergency situations [7]. In order to ensure workspace safety and production efficiency during human-robot cooperation in manufacturing environments, continuous and accurate perception of worker activities is required [8].

The contributions of this paper are threefold. (i) As the main contribution, we report from a case study analyzing the impact on the spatial streams from multiple subjects conducting simultaneous and distinct activities and gestures. We collect these traces from a prototype, using USRP software defined radios consisting of a transmit antenna and a phase synchronised four-antenna OFDM receiver (4x52 carriers) and run experiments in a m semi-anechoic chamber. We make available the labelled RF data together with phase configurations to steer the reception between ° to ° azimuth angles [9]. (ii) A training-free people counting algorithm that counts people when they simultaneously conduct distinct activities. Empirical results from over experiments with up to human subjects show that persons can be detected within one person error in of the time, 5 persons 43%

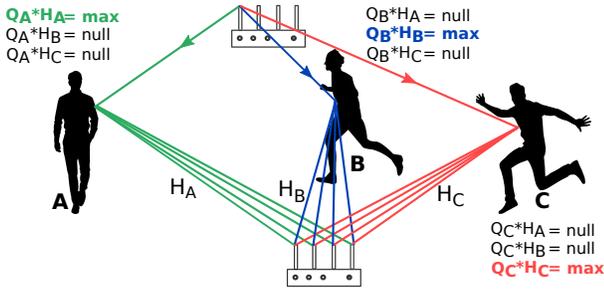


Fig. 1: Concept of multiple people counting through beamsteering. Steering matrix applied to channel amplifies the signals scattered from the person at position A, on amplifies the signals from person B and on amplifies the signals from person C.

and 6 persons 50% of the time without any training efforts. In comparison to a state-of-the-art crowd counting approach [3], we demonstrate significant human count estimation improvement. (iii) As the final contribution, we apply a blind source separation method on the spatial streams to extract 20 out of 21 human gestures from one target person when up to three persons simultaneously perform distinct activities.

## II. RELATED WORK

**Human sensing modalities.** Most prominent device free sensors used for people monitoring exploit the properties of reflectivity with time-of-flight (ToF) cameras, emissivity with thermal images or acoustic. They are, however, constrained by ranges, occlusions, environmental (smoke/fire) conditions, and privacy concerns [10], [11]. Non-image based solutions usually demand that people carry a device, yet, they limit human subjects' mobility, visibility, and communication practice [12]. Device-free solutions, which do not require people to wear any devices, are therefore more promising in Pervasive computing domains.

**Device-free radio sensing.** Device-free localization, tracking and activity recognition using radio signals has shown advantages over camera or wearable device-based solutions due to low intrusion of privacy and convenience [13], [14]. Research has demonstrated the use of various radio signal measurements for human sensing including time delay [2], phase [15], Doppler [16], and signal strength [17]. Those have been used for purposes like vital sign monitoring [18], activity and gesture recognition [19], [20], localization [21], gait identification [22] and fall detection [23].

Even though beamforming and beamsteering has been previously used for device-based localization, it is not prominently used in device-free radio sensing. Wihear [24] used beamforming to focus a beam on a person's mouth to mitigate multipath effects from the environment for lip reading and speech recognition using SDRs and a WiFi physical layer. mtrack [25] leverages electronically steerable antennas to track small objects like pens using GHz millimeter wave radios. Our previous work explored the extraction of the direction of arrival (DoA) of a human target through a beamsweeping

method [26]. In contrast to these works, we use receiver-end beamsteering to sense *full body motions, count people and extract their activities*.

**People counting.** The feasibility of people counting from RF-signals has been demonstrated in previous work [27], [28], [29], [3]. One straightforward approach for people counting is the utilization of dispersion in the received signal strength (RSS) [17], [30] or CSI [27], which is loosely correlated to the number of people present in a space. It is well known that RSS variation in commercial transceivers as a signal descriptor is not stable to count people due to low granularity and multipath effects. In order to distinguish spatially separated people, [17] uses transceivers that are spatially distributed. Other approaches for crowd counting have explored multiple antennas in both transmitter and receiver to achieve spatial diversity [28], information theory based methods like transfer learning to minimize the environment impact on people counting features [3], as well as exploitation of additional frequency-domain features such as the shape of the Doppler spectrum which is correlated to the number of people moving in the monitored environment [29]. *We are different to these works both in the approach, beamsteering using limited antennas at the receiver, and application, count and extracting activities of spatially distributed humans and has the advantage of not requiring prior training.*

## III. SPATIAL FILTERING USING OFDM BEACONS

Here, we implement a delay and sum beamformer using OFDM beacons. The beamformer consists of a Uniformly spaced Linear antenna Array (ULA) as shown in Fig. 2a. The narrowband signal received at time instant by the ULA is modeled as

$$(1)$$

where , and are  $\mathbb{C}^{\times 1}$  vectors representing the line of sight (LoS) signal, signal scattered off a person and noise respectively. Unlike in device-based localization where is the desired signal, is the desired signal in our case and it can be amplified by finding the correct steering vector. A signal arriving from direction at arrival angle , induces a phase shift of in antenna element relative to the element where is the wavelength. A phase shifter connected to the array element changes the phase of the input signal by . By setting the phase shift

$$(2)$$

and summing the shifted signals, the received power corresponding to the angle can be maximized. This way the beam can be steered to the desired direction and forms the steering vector for that direction. Since radio waves arrive at the receiver over multiple paths due to reflection and scattering from human subjects in the environment, and these signals are superimposed in the line of sight signal, the steering vector that maximizes the signal amplitudes corresponding to the human is unknown. Therefore we generate steering vectors for azimuth angles using Eq. 2 so that signals

arriving from direction  $\theta$  are amplified by the  $m$ th steering vector.

In what follows, we explain how this is achieved using the OFDM symbols. In OFDM, a channel is divided into multiple subcarriers and data is modulated in each subcarrier. In our testbed (detailed in Sec. VII-A), a 12.48 MHz channel is divided into 52 sub-carriers each with a spacing of 240 kHz. The transmitter sends a beacon with a known sequence  $\mathbf{s}$  each subframe. IQ samples  $\mathbf{y}_m$  are received by antenna element  $m$ .  $\mathbf{y}_m$  can be represented as

$$\mathbf{y}_m = \mathbf{h}_m \mathbf{s} + \mathbf{n}_m \quad (3)$$

where  $\mathbf{h}_m$  is the transmitted beacon and  $\mathbf{n}_m$  is the wireless propagation channel. A signal arriving from direction  $\theta$  introduces a phase rotation of  $\frac{2\pi}{\lambda} \mathbf{r}_m \sin \theta$  due to the time delay  $\frac{\mathbf{r}_m \sin \theta}{c}$  to arrive at antenna element  $m$  where  $c$  is the speed of light. We estimate  $\mathbf{h}_m$  by computing the cross-correlation

and  $\hat{\mathbf{h}}_m$  maximizes the correlation between transmitted and received beacons:

$$\hat{\mathbf{h}}_m = \frac{\mathbf{y}_m \mathbf{s}^H}{\|\mathbf{y}_m\|} \quad (5)$$

to obtain spatial stream  $\mathbf{y}_m$  for a total duration of  $T$  with  $N$  where

#### IV. CHARACTERIZATION OF SPATIAL STREAMS FROM A PRACTICAL EXPERIMENT

In this section, we collect data traces from a practical experiment wherein up to five persons perform distinct in-place activities at five different locations. We apply spatial filtering according to Sec. III and analyse the behaviour of the resulting spatial streams in the directions of the human occupants.

##### A. Experiments

We conducted a case study (Fig. 2) in a semi-anechoic chamber of size  $5.6\text{m} \times 4\text{m}$ . A signal continuously emitted by a single antenna transmitter was captured by a receiver with a ULA of 4 phase-synchronized elements. The antenna elements have a spacing of  $\frac{\lambda}{2}$  where  $\lambda$  is the wavelength at a carrier frequency of  $2.45\text{GHz}$ . We performed five measurement campaigns having one person at B, two persons at D & E, three persons at B, C, & D, four persons at B, C, D & E and 5 persons at A, B, C, D, & E performing distinct in-place activities simultaneously while IQ samples were collected.

The steering vectors for directions A, B, C, D & E were derived from 5 other measurement campaigns by placing a single antenna transmitter at positions A, B, C, D and E in a human-free environment and recording the IQ samples.

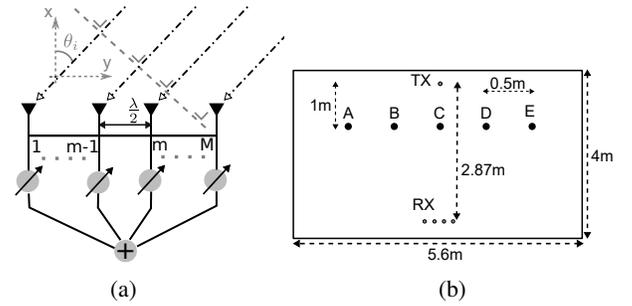


Fig. 2: (a) The phased array system that maximizes signals arriving from direction  $\theta$  by steering the beam. (b) Experiment Environment. # of RX antennas: 4, # of TX antennas: 1, antenna heights from ground: RX 2.87m, TX 0.5m, room size: 5.6m x 4m. A, B, C, D and E are human locations

Then, we computed the azimuth angles for those five positions by minimizing the mean square error between the measured steering vectors and the modeled steering vectors for azimuth angles  $\theta$  using Eq 2 [26]. The azimuth angles that yielded were  $0^\circ$ ,  $15^\circ$ ,  $30^\circ$ ,  $45^\circ$ ,  $60^\circ$ . Learning the azimuth angles and the corresponding steering vectors helps to understand directions of the considered positions and generate steering vectors for the nearby azimuth angles.

##### B. Characterization of the spatial streams.

We divide the analysis of the spatial streams into four categories based on the number of human subjects: *single person*, *two persons*, *three persons* and *four and five persons*. Fig. 3 summarizes these cases. In all figures, the amplitudes are centered to have zero mean and eliminate the LoS effect.

1) *A single person*: Fig. 3a shows the fluctuations in the spatial streams when there is a person in the direction of B. Comparing all the streams, the spatial stream in the direction of the person, stream B in Fig. 3a, is clearly excited and has the highest fluctuation, i.e., stream B has the highest variance among all streams. However, the adjacent streams also experience correlated fluctuations to the spatial streams that a human occupies, e.g. streams A and C in Fig. 3a. The adjacent streams, A and C, despite having similar patterns to the human occupied stream, have distortions like local scaling and noise. Other streams, D & E show fluctuations at similar time instants yet with a higher attenuation: in Fig. 3a spatial streams D and E show correlated fluctuations with high attenuation, however, the majority of the streams is dominated by Gaussian noise.

We further verify this behavior by constructing the array response of the ULA by placing a unit source at direction B. The array response can be constructed by multiplying the steering vector of direction B by the steering vectors of azimuth angles  $\theta$ . Fig. 4 shows the array responses of directions A, B, C, D & E and the response of B is specially highlighted in red. Given that A, B and C have azimuth angles of  $0^\circ$ ,  $15^\circ$  and  $30^\circ$ , the amplitude of B has about 0.9 influence on C and 0.8 influence on

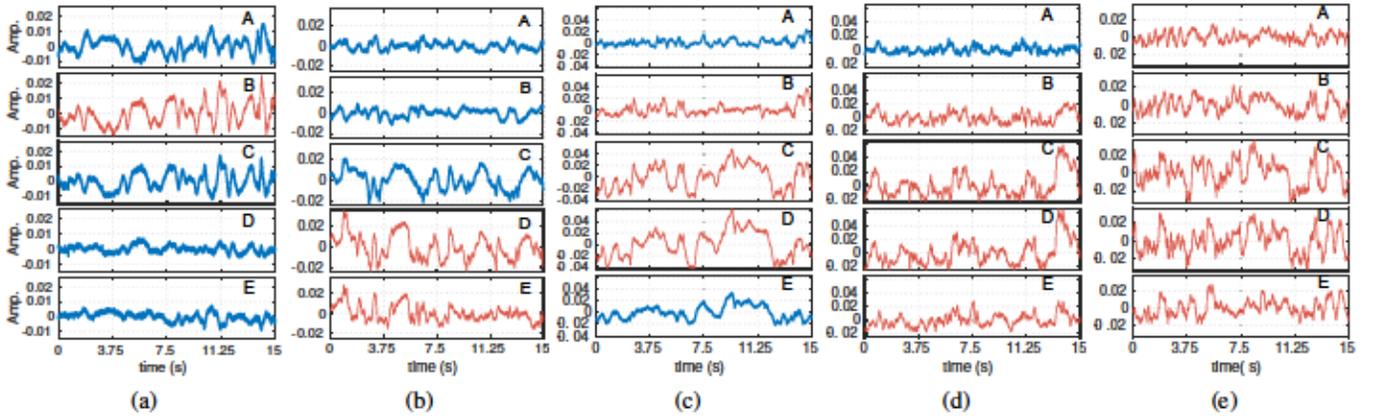


Fig. 3: (a) One human subject at B, (b) Two human subjects at D & E, (c) three human subjects at B, C & D, (d) four human subjects at B, C, D & E, (e) 5 humans subjects at A, B, C, D & E.

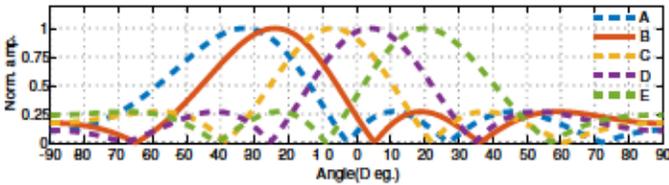


Fig. 4: The array responses for directions A, B, C, D & E interfering each other.

A. Whereas the influence on D and E are below 25% which explains the high amount of noise in D and E.

2) *Two persons*: Fig. 3b corresponds to the spatial streams when ‘two’ humans reside in directions ‘D’ and ‘E’ (we denote this as 2DE). We notice that in certain subsequences, similarity between D and E can be observed, especially between 0 and 7.5 s. However, the adjacent pattern C, where a human does not exist is also similar. We quantify these similarities using constraint derivative dynamic time warping (cDDTW) distance  $d_{i,j}$  where  $i, j \in \{A, B, C, D, E\}$  such that  $0 \leq d_{i,j} < 1$  (Table I, cf and a lower value for  $d_{i,j}$  indicates a high similarity. Sec. V-B2). Accordingly, for 2DE,  $d_{C,D} = 0.38$  is less than  $d_{D,E} = 0.5$ . This behaviour can be explained as follows. As stream C corresponds to a human unoccupied link, adjacent to human occupied stream D, C is only interfered by D. However, as both D & E are occupied by humans, they are mutually interfered from each other, causing a lower similarity (and a higher distance for  $d_{D,E}$ ). Again, the spatial streams beyond stream C witness high attenuation similar to the single human cases in Fig 3a.

3) *Three persons*: Fig. 3c corresponds to the case where three humans reside in the directions of the spatial streams B, C and D (denoted 3BCD). Again, the streams B, C and D show similarity intermittently in subsequences. Especially, streams C and D show the highest similarity which is confirmed by  $d_{C,D} = 0.27$  for 3BCD in Table I.

Stream B indicates low fluctuations unlike C or D. Interestingly, stream A indicates a noise dominated stream showing little to no influence from B. This is also indicated by  $d_{A,B} = 0.58$ . This behaviour indicates that the person at

	1B	2DE	3BCD	4ABCD	5ABCDE
A-B	0.58	-	0.58	0.38	0.43
B-C	0.28	-	0.48	0.47	0.5
C-D	-	0.38	0.27	0.33	0.3
D-E	-	0.5	0.55	0.47	0.5

TABLE I: Similarity between human occupied stream and the adjacent stream quantified using cDDTW for experiments having a single human (left table) and multiple humans (right table). 1A, 1B etc. in the tables denote the number of humans in the experiment and their positions.

B is not involved in large body movements. Given that the major pattern in C, D and E are not leaked to B, we can conjecture that the most influential human motion is at D. This behavior can be attributed to the type of activity that the person performs in respective streams. As an example, if a person is involved in full body motion at stream D, the body can disrupt the multipath heavily causing spatial streams C and E to have high interference from D.

4) *Four and five persons*: Fig. 3d illustrates the case where 4 humans reside in spatial streams B, C, D and E and Fig. 3e illustrates the case for 5 humans in streams A, B, C, D and E. The trends among the streams where humans are present are consistent to what was observed in the three human subjects case where there is similarity in adjacent streams with local dissimilarities. In the single human subject case,  $d_{i,j} = [0.13, 0.64]$  among adjacent streams, whereas here,  $d_{i,j} = [0.3, 0.5]$  among AB, BC, CD and DE. This indicates that the adjacent streams do not have drastic differences when 4 or 5 humans occupy the streams. However, the frequency of dissimilarities in subsequences (local dissimilarities) between adjacent streams tend to increase as more humans occupy the adjacent streams compared to the cases with two or three humans. This can be expected as nearby streams tend to interfere more with each other. When more humans occupy adjacent streams, more interference among them can be expected.

Summarizing, main observations of this section are:

- 1) Spatial streams fall into three categories: a) streams having direct influence from a person, b) unoccupied streams

adjacent to a human occupied stream showing correlated fluctuations, and *c*) streams that are unoccupied and dominated by noise.

- 2) Presence of a person has most influence on the stream in that person’s direction followed by both adjacent streams.
- 3) If two adjacent streams are occupied by two persons, both streams can show correlated fluctuations with intermittent discontinuities (occlusions).
- 4) Distortions between two streams can occur irrespective of whether those streams are occupied by persons or not
  - a*) amplitude offset, *b*) local scaling: subsequences have different amplifications, *c*) complexity: subsequences have Gaussian noise, and *d*) patterns are similar but inverted.

We now use these observations to develop an algorithm to count the number of people in spatial streams.

## V. HUMAN SUBJECT COUNTING

Estimation of a single person can be achieved by using a beamscanning technique where the direction of arrival (DoA) of that person forms a Gaussian distribution over time [26]. However, when there are more than one person, this method fails as the DoA distributions often look like a mixture of Gaussians due to the superposition of the responses of two or more people. Therefore, along with the above method, here we propose an algorithm to estimate more than one person based on the observations in section IV. As shown in Fig. 5 the solution is divided into three processes: *i*) *Divide the IQ samples from the receiver using spatial filtering*. The spatial filtering executes the functionality as mentioned in Sec. III where it applies the steering vectors of directions on the IQ samples of the receiver to obtain spatial streams. *ii*) *Detection of the absence of a person in a spatial stream and discarding those streams*, and *iii*) *human count estimation using the remaining streams*.

### A. Detecting vacant spatial streams

According to Sec. IV, the detection of fluctuation in a spatial stream is not a necessary indication of the presence of a human as interference from nearby spatial streams can also cause fluctuations. Therefore, we filter the human-free spatial streams by comparing the distribution of those to a Gaussian distribution to increase confidence that no human exists in a given stream. We model the distribution of human free data traces as a Gaussian distribution based on the assumption that streams having no influence of the human are dominated by Gaussian noise.

To detect that no human exists in stream, we first center the stream to have zero mean to remove any amplitude

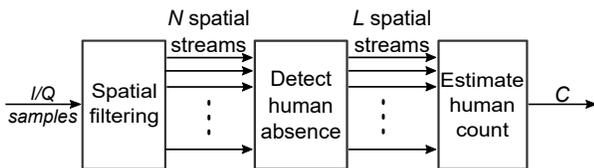


Fig. 5: Modularized architecture for human counting.

offsets due to LoS and apply *Kullback-Leibler divergence* between the probability mass function (PMF) of random variable from amplitudes of a Gaussian distribution modeling the human-free data traces and the PMF of stream.

$$\log \frac{P(x)}{Q(x)} \quad (6)$$

When such that is the maximum Kullback-Leibler divergence corresponding to the Gaussian distributed traces, we consider stream to be free of a person. Accordingly, the streams that indicate a human presence are exploited for human count estimation.

### B. Human Count Estimation

To estimate the human count when more than one person exist, from the selected spatial streams, we utilize the observations made in Sec. IV-B. Here, we work on the premise that the amount of similarity of two spatial streams is an indication of the human count. As an example, if two adjacent streams have a high similarity, this indicates the influence of one person, and if two streams have motions with low similarity, this indicates the influence of two or more persons in the direction of those two streams. First, we capture these similarities using a distance measure and feed them to a clustering algorithm so that the number of clusters in the spatial streams indicate the number of people in those streams.

1) *Segmentation to subsequences*: We divide each spatial stream into subsequences and estimate the human count within each subsequence. As observed in Sec. IV-B, the sequences can have dissimilarities within certain time intervals due to occlusions. Segmentation helps to capture these effects in a high granularity in time. Additionally, the computation time in subsequent steps, e.g. during similarity measurement of the subsequences, can also be reduced.

2) *Measuring similarity of the spatial streams*: As the distance measure selection is domain dependant [31], we first study the types of distortions that exist in our time series data. We treat the spatial streams as different time series.

The types of distortions that are unique to any two streams in our data as mentioned in Sec.IV are *i*) *amplitude offset*, *ii*) *local scaling*, *iii*) *occlusion*, *iv*) *complexity* and *v*) *inverted similar patterns*. From these distortions, invariance to amplitude offset is achieved by z-normalization, and complexity invariance is achieved by noise filtering. Occlusions in our case are a desired effect, therefore, we select a distance measure that is sensitive to occlusions in the data. Additionally, the distance measure should be invariant to local scaling. The most commonly used distance measure in the literature that achieves local scaling invariance is dynamic time warping (DTW) which allows a local non-linear alignment between the sequences [32]. However, this may yield unnatural alignments where a single point on one subsequence is mapped onto a large subsection of another subsequence. Therefore, we apply *constrained derivative dynamic time warping* (cDDTW) [33]. To achieve invariance to pattern inversion, we apply cDDTW two times for each pair of subsequences, once for

the original subsequence and another time for the inverted subsequence. Then, we obtain the minimum distance from those two instances.

In cDDTW, the DTW algorithm is applied to minimize the warping distance between the first derivative of the time sequences. This way, it is easier to capture the shape-based similarity by mitigating distortions in the amplitude. On the other hand, by constraining the warping path to visit a subset of the cells, an accurate distance between the sequences can be found as too much flexibility can result in poor discrimination between the sequences [34].

cDDTW is sensitive to noise in the spatial streams as the first derivative of each time sequence is required, furthermore, the derivative operation introduces additional noise. Therefore, we apply a 4<sup>th</sup> order Butterworth low pass filter with cut-off frequency at 30 Hz to isolate the effects of human motion both before and after applying the derivative operation. We apply cDDTW among all pairs of  $L$  spatial streams and obtain an  $L \times L$  distance matrix  $\mathbf{D}$

$$\mathbf{D} = \begin{bmatrix} 0 & d_{1,2}^p & \cdots & d_{1,L}^p \\ d_{2,1}^p & 0 & \cdots & d_{2,L}^p \\ \vdots & \vdots & \ddots & \vdots \\ d_{L,1}^p & d_{L,2}^p & \cdots & 0 \end{bmatrix} \quad (7)$$

where  $d_{i,j}^p$  is the cDDTW distance between spatial streams  $i$  and  $j$  at subsequence  $p$ .

3) *Clustering and human count estimation*: We cluster spatial streams into groups conditioned on their similarity to each other. *Hierarchical agglomerative clustering* (HAC) is chosen as the clustering algorithm because of its inherent capability of finding natural clusters within data [35]. Here we feed the distance matrix  $\mathbf{D}$  directly to the clustering algorithm and construct a cluster tree. Then we extract the median linkage [36] between each cluster in the hierarchy. Median linkage measures the Euclidean distance between weighted centroids of two clusters. We find natural clusters within the spatial streams similar to an approach on finding the knee of an error curve [37]. More specifically, the median linkage with maximum separation between any two clusters is selected as the cut off point to select the number of natural clusters  $C_p$  for subsequence  $p$ . As an example, Fig. 6a illustrates a hierarchical cluster tree structure of five spatial streams A, B, C, D & E while three humans are present in the directions of streams B, C & D. HAC correctly identifies the similarities of A, B and D, E and groups A & B into cluster 1, D & E into cluster 2 and C into cluster 3. As the distance between links 2 and 3 is the highest distance among all the links in Fig 6b, the cut-off height lies between links 2 and 3 which results in 3 clusters.

## VI. EXTRACTION OF INDIVIDUAL ACTIVITIES

In this section we implement an algorithm to extract the activity patterns when multiple people are present in an area. We use  $L$  spatial streams computed for the azimuth angles  $-90 \leq \theta_t \leq 90$  with 1 degree resolution. As already discussed in Sec.IV, these streams are not completely orthogonal to each other and activity patterns of each individual are mixed

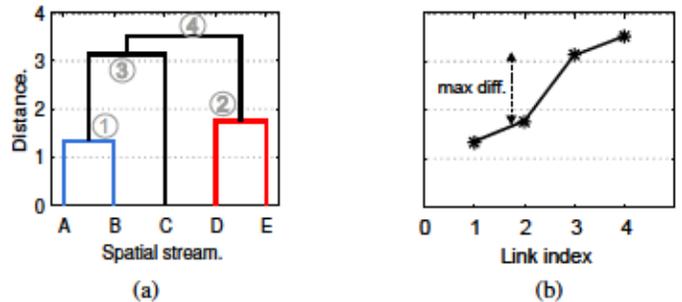


Fig. 6: Finding natural clusters from a hierarchical cluster tree using the maximum median linkage among clusters. (a) Hierarchical cluster tree of 5 spatial streams, (b) median linkage among the clusters.

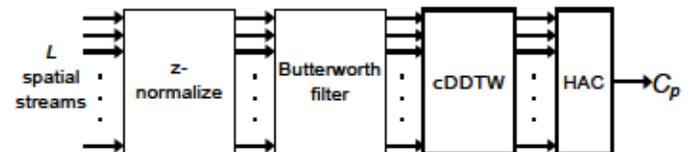


Fig. 7: Human count estimation process within a subsequence.

in different proportions depending on the positions of the persons. Therefore, we use a blind source separation approach to extract these activities.

In the literature, different types of blind source separation algorithms exist with independent component analysis (ICA) being most popular [38]. These algorithms try to exploit certain statistical properties or temporal structures of the latent sources. Here, we cannot assume that the latent sources, i.e, the signal perturbations caused by human activities, have a certain temporal structure due to the complexity of human behaviour. Therefore, we use a variant of ICA, the *joint approximate diagonalization of eigen-matrices* (JADE) algorithm [39] which assumes that the latent sources are non-Gaussian. We observed in Sec IV how the spatial streams impacted by a person lost the noisy Gaussian structure, which motivated the use of this approach. In JADE, the spatial streams are transformed to the latent source signals by exploiting fourth order moments of the spatial streams which is a measure of Gaussianity and independence of the sources.

One issue with JADE is that the number of mixtures should be equal to the number of sources, i.e, we should have an estimate of the people count, and the dimensions of the spatial streams should be reduced from  $L$  to the people count. As a solution, we use *principal component analysis* (PCA) [40] to reduce the dimensions of the spatial streams and whiten the data before applying JADE.

## VII. IMPLEMENTATION

### A. Testbed Description

The measurements were obtained using our software-defined radio platform [41]. The system used in these measurements has two parts: a receiver and a transmitter. The transmitter is comprised of a host computer running Ubuntu 16.04 and a Universal Software Radio Peripheral (USRP) X300 series

TABLE II: Main parameters of the testbed.

Center frequency	BW	Useful subcarriers	Samples/s	Antennas
GHz	MHz	52	5408	

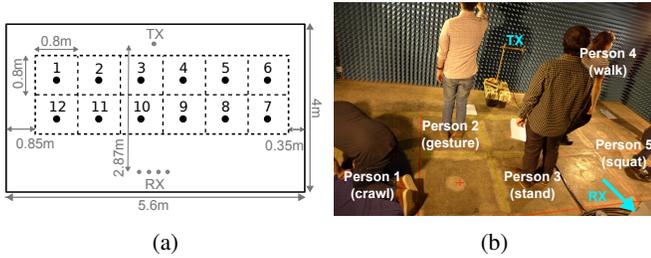


Fig. 8: (a) Experiment Environment (semi-anechoic chamber) of the second campaign. (b) Measurement setup for steering vector calculation for 12 directions e.g. steering vector of the direction of location 4 is estimated by placing the TX at 4 and getting IQ samples when no human is present. (c) 5 people performing in-place activities in the experiment environment.

with UBX-160 and SBX RF-daughterboards as the radio front-ends. The receiver was comprised of three USRPs, of which two served to collect the signal from the other high-level node (= transmitter). The third device is the reference USRP. It operates by synchronizing over the air with the primary two USRPs and then transmitting a reference signal to calibrate the starting offset between them. A four element ULA with – inter-element spacing was connected to the data collection USRPs. Dipoles were used as array elements.

Phase coherence between the two receive USRPs was accomplished using a clock distribution system. It provides both a pulse-per-second signal and a 10 MHz reference signal to discipline the local oscillators of the USRPs. A reference transmitter was however also necessary due to the effects reported in [41]: inter-device random phase offset at start-up and component variable in each RF chain. Additionally, the presence of the reference transmitter helps ensure that phase coherence, and therefore beamforming performance, is maintained throughout the measurements.

Table II summarises the main parameters used in our testbed. The air-interface configuration utilised was an OFDM frame structure at a carrier frequency of 3.42 GHz. The total bandwidth of the system was 15.36 MHz with 52 useful subcarriers yielding a useful bandwidth of 12.48 MHz. Each USRP stream IQ samples at a rate of 16.66 megasamples per second. The time domain subdivides into subframes of 3082 samples in length, yielding approximately 5408 subframes per second. This rate is essentially the sampling rate of the IQ samples used as input to the spatial filtering module.

### B. Experiment Setup

We performed experiments in two measurement campaigns. **CAMPAIGN-I.** We used the experiment setup in Fig. 2 detailed in Sec. IV-A. Even though the experiment area is an anechoic

chamber, ground reflections are still possible as the ground was not insulated with radio wave absorbents. Therefore, it is essentially a semi-anechoic chamber. Table III shows the number of persons that participated in each experiment and their position (A,B,C,D or E) in the room. The maximum number of participants in any given experiment in this campaign did not exceed 5. The distance between any two participants in this campaign belongs in the interval  $m$   $m$ . We collected measurements for 26 types of experiments, each lasting for  $s$  and each experiment was repeated again to collect a total of 52 measurements. Experiments with indices 1-22 were in-place random activities while in experiments 23-26, the participants were walking in the experiment area.

**CAMPAIGN-II.** For this 2nd campaign, we used the experiment setup in Fig. 8a with following settings. Number of RX antennas:  $n$ , the number of TX antennas:  $n$ , antenna heights from ground: RX  $m$ , TX  $m$ , room size:  $m$   $m$   $m$ . Locations marked  $n$  are to be occupied by humans. We increased the number of positions in the room which humans can potentially occupy to 12. The steering vectors for the directions were also computed as described in Sec III. Table III illustrates the number of persons that participated in each experiment in this campaign and their position (this time  $n$ ) in the room. The distance between any two participants in this campaign belongs in the interval  $m$   $m$ . Here, we collected measurements for 36 types of experiments each lasting for  $s$ . Experiments with indices 1-24 and 33-36 were repeated again and experiments having the indices 25-32 were only performed once. Therefore the total number of experiments were 64 in this campaign. Experiments with indices 1-22 were in-place activities where the participants remained in a  $m$   $m$  square they were assigned to as shown in Fig. 8a. In experiments 23-24, the participants were circling the experiment area, walking through all the positions marked in the figure. The maximum amount of participants for any given experiment was 6. The types of in-place activities that were performed by the participants include *squatting*, *walking inside the assigned square*, *standing*, *jumping*, *crawling* and *hand gestures*.

**Parameter Selection.** As our test-bed provides samples at  $n$  samples/s, we downsample them to  $n$  samples/s. The main motivation for selecting this sampling rate is to reduce the computational time in the cDDTW algorithm. Therefore, each experiment lasting for  $n$  s produces 2700 samples at  $n$  Hz. As mentioned in Sec. V-B1, these samples are divided into  $n$  subsequences having a length of  $n$  samples and each subsequence has an overlap of  $n$  samples with the neighbouring subsequence.  $n$  and  $n$  are empirically identified to minimize the human count error by varying  $n$  samples and  $n$  samples. The human count error is measured by finding the absolute difference between estimated human count ( $n$ ) and the ground truth human count  $n$ .

TABLE III: Experiment parameters of the two measurement campaigns. Experiment index, the position of each person in the environment and the number of participants in each experiment are illustrated here.

CAMPAIGN-I	Index	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
	Position	-	A	B	C	D	E	AB	BC	CD	DE	ABC	BCD	CDE	ABCD	BCDE
Count	0	1	1	1	1	1	1	2	2	2	2	3	3	4	4	
CAMPAIGN-II	Index	16	17	18	19	20	21	22	23	24	25	26				
	Position	ABCDE	-	A	AB	ABC	ABCD	ABCDE	-	-	-	-				
Count	5	0	1	2	3	4	5	2	3	4	5					
CAMPAIGN-II	Index	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
	Position	1,3,4,6	1,2,5,6	3,4,9,10	3,4,8,11	2,5,9,10	1,6,7,12	2,3,4,5	3,9,11	2,4,10	2,3,4	2,5	3,4	7,12	2,10	2,10
Count	4	4	4	4	4	4	4	4	3	3	2	2	2	2	2	
CAMPAIGN-II	Index	16	17	18	19	20	21	22	23	24	25-32	33	34	25	36	
	Position	10	2	2,3,4,5,6	2,4,6,8,10,12	1,3,5,7,9,11	1,2,3,4,5,6	2,3,4,5,9,10	-	-	-	2,10	3,4	2,5	7,12	2
Count	1	1	5	6	6	6	6	6	2	1	0	2	2	2	2	

Actual ( $H_c$ )	Estimated ( $\hat{H}_c$ )						
	0	1	2	3	4	5	6
0	12	0	0	0	0	0	0
1	0	16	2	0	0	0	0
2	0	1	14	13	1	0	0
3	0	0	1	16	1	0	0
4	0	0	3	9	4	3	0
5	0	0	6	2	6	0	0
6	0	0	0	2	2	3	1

Actual (%)	Estimated (%)						
	0	1	2	3	4	5	6
0	100	0	0	0	0	0	0
1	0	89	11	0	0	0	0
2	0	3	47	47	3	0	0
3	0	0	1	55	89	5.5	0
4	0	0	0	16	47	21	16
5	0	0	0	43	14	43	0
6	0	0	0	25	25	37.5	12.5

TABLE IV: People counting confusion matrices. Actual vs Estimated count (left) and Actual (%) vs Estimated (%) (right)

## VIII. PERFORMANCE EVALUATION

### A. People Counting Confusion Matrix

We evaluate the performance of the proposed training-free human counting algorithm using two data sets CAMPAIGN-I and CAMPAIGN-II by comparing the actual human count  $H_c$  against estimated  $\hat{H}_c$ . Table IV shows the confusion matrix for both campaigns where the maximum number of participants was 6, and the total number of measurements was 116. When the actual human count,  $H_c$ , is between 0 – 4, the algorithm estimates it well up to one person error where 0-1: 100%, 2: 99%, 3: 100% and 4: 84%. When  $H_c$  exceeds more than 4 humans, the algorithm underestimates the count within 1 – 3 humans with accuracies of 5: 43% and 6: 50% within one person error. Accuracy degradation when the number of humans increase can be explained as follows. For the experiments, only a 4 antenna ULA is used at the receiver and the minimum distance between two nearby human subjects is only 0.8 m in a 22.4 m<sup>2</sup> room. In the current setup, the adjacent spatial streams experience mutual interference. It is well known that, as the number of antenna elements increase in the linear array, the beamwidth can further be reduced so that the interference on adjacent spatial streams can be further attenuated. We also note that when  $H_c = 0$ , in both campaigns, the human count is estimated correctly in 100% of the cases. This is due to the fact that the human vacancy detection algorithm in Sec V-A distinguishes human occupied streams from the unoccupied accurately.

### B. Accuracy vs the Number of Antennas

The use of multiple antennas enables beam steering in different directions. During the experiments we used up to 4 antennas to steer the beams. Here we analyse the influence of the number of antennas on the performance of people counting by reducing the number of antennas for the same data set e.g. beamsteering by using 3, 2 or 1 ULAs. As beamsteering is

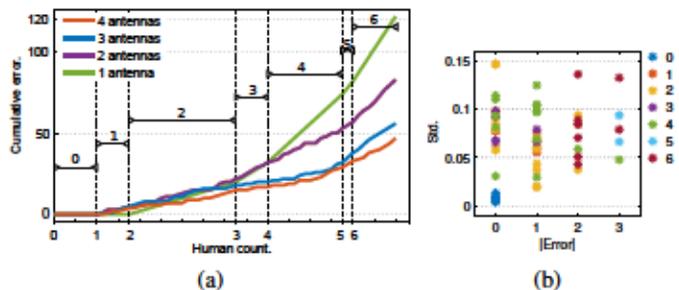


Fig. 9: (a) The effect of the number of antennas on human count estimation error. (b) The relationship between dispersion in the data and human count estimation error.

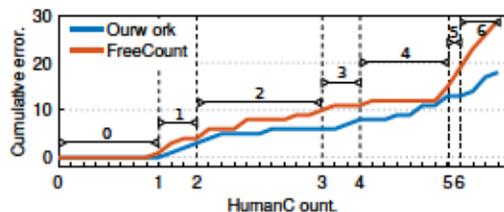


Fig. 10: Comparison of the cumulative error of FreeCount vs our work.

done offline, the effect of ULA size can be evaluated for the same data set.

Fig. 9a illustrates the cumulative human count estimation error of both CAMPAIGN-I and CAMPAIGN-II against the increasing order of actual human count  $H_c$ . As expected, 4 antennas provide the lowest cumulative error over all experiments. Comparatively, 3 antennas provide a lower error compared to 2 and 1 antennas. The difference between the cumulative error of 1 & 2 antennas compared to 3 & 4 antennas begin to widen from  $H_c \geq 3$ . Whereas, the difference between the cumulative errors of 3 and 4 antennas widen from  $H_c \geq 5$ . This provides an estimate on when the system starts to break in terms of estimating the human count when different number of antennas are used.

### C. Accuracy vs Dispersion in the Data

We analyse how the amount of dispersion in the measurements affects the human count accuracy. Fig. 9b illustrates the relationship of absolute error  $|H_c - \hat{H}_c|$  and dispersion for all experiments in CAMPAIGN-II. We use standard deviation as the measure of dispersion and apply the standard deviation on

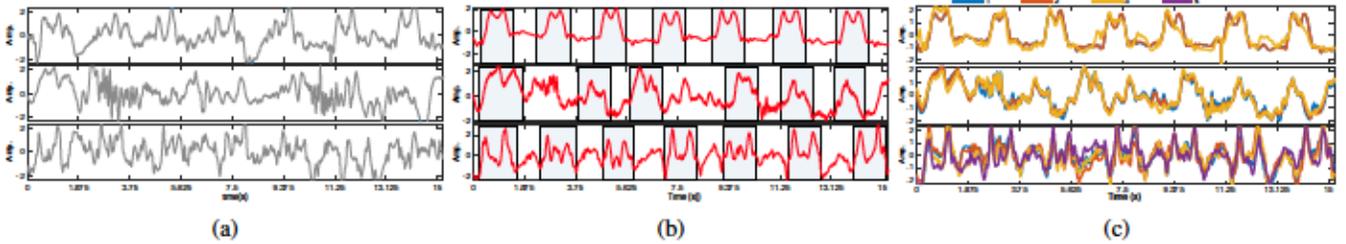


Fig. 11: Recovering activity patterns of a person through blind source separation, (a) mixed patterns, (b) recovered patterns when  $\hat{H}_c = H_c$  and (c) pattern recovery at  $\hat{H}_c = 1, 2, 3,$  or  $4$  when  $H_c = 2$  (top),  $H_c = 2$  (middle) and  $H_c = 3$  (bottom).

$R(\theta_m)$  (we do not apply spatial filtering here). As expected, the standard deviation is lowest when  $H_c = 0$ . Even though  $\max(|H_c - \hat{H}_c|) = 3$  occurs when  $H_c = 4, 5$  or  $6$ , the standard deviation is  $< 0.1$  in 4 out of 5 occasions. The maximum standard deviation occurs at  $H_c = 2$  which highlights the fact that our algorithm performs irrespective of the amount of dispersion in the data.

#### D. Comparison with the State of the Art

We compare the performance of our human count estimation with an estimation of human count via the the statistical features utilized by FreeCount [3], a state of the art crowd counting algorithm for CAMPAIGN-II. The features we extracted from the correlation amplitudes ( $|R(\theta_m)|$ ) at Eq. 3 (we do not apply spatial filtering here) are mean, min, max, standard deviation, median, entropy, 1<sup>st</sup>, 2<sup>nd</sup>, 3<sup>rd</sup> and 4<sup>th</sup> order moments. The features are extracted from 64 experiments in CAMPAIGN-II. As each experiment (except the experiments indexed 25-32) is repeated once, we use the original set as the training set and the repeated set as the testing set in a multiclass classification implementation.

Fig. 10 compares the performance of our work against FreeCount. We plot the cumulative absolute error (y axis) against the human count ( $H_c$ ) (x axis) in 32 experiments. As depicted, when  $H_c = 0$ , both approaches estimate the human count accurately. However, as  $H_c$  increases from 1 to 3 we begin to see the gap between the curves widening. When  $H_c > 4$ , the gap further widens resulting in a maximum cumulative error difference of 12. This indicates the robustness of our algorithm as the human count increases compared to FreeCount. We also note that FreeCount uses a supervised learning approach whereas we use non-supervised learning for human count estimation which can be considered as a competitive advantage of our approach.

#### E. Device-free Activity Extraction

In this section, we evaluate the performance of our approach to extract activities. We use three position configurations in Campaign II with increasing difficulty. In Exp. A, two persons are present at locations 2 and 10. The person at position 2, performs a *push-pull* hand gesture with 7 repetitions and the person at 10 performs squatting. In Exp. B, two persons are placed at 2 and 5. The person at 2 performs gestures with 7 repetitions and the person at 5 crawls in the area

continuously. In Exp. C, three persons are present at locations 2,3 and 4. The person at 2 performs jumping, the person at 3 performs squatting and the person at 4 performs the gesture with 7 repetitions. Fig. 11a illustrates the amplitude fluctuations detected from a single antenna (single antenna is shown for space limitations) for the three experiments (Exp. A: top, Exp. B: middle and Exp. C: bottom). Fig. 11b illustrates the extracted gestures from the three experiments highlighted in the grey area. The 7 gestures from Exp. A are clearly detected in the top figure with all 7 of them having the same pattern. From Exp. B, 6 out of the 7 gestures can be extracted as shown in the middle plot. From Exp. C, again the 7 patterns can be observed. From the three experiment setups, Exp. A gestures are the most consistent, while Exp. B gesture patterns have the most inconsistencies. Exp. C gestures can also be identified but not as clearly as in Exp. A. Here we conjecture that, Exp. B gestures have been difficult to extract because the other person is performing continuous activities. If the corresponding raw amplitudes are analyzed from Fig. 11a, we can observe that those are mostly high frequency motion (e.g. 1.875 s - 3.75 s and 9.375 s - 11.25 s show rapid fluctuations) compared to the other two experiments. Finally, Fig 11c shows the effect on the extracted patterns when the  $\hat{H}_c$  has an error. Here the patterns are extracted by varying the estimated count up to  $\hat{H}_c = [1, 3]$  in top and middle figures while it is varied till  $\hat{H}_c = [1, 4]$  in the bottom figure. However, the results do not show significant deviation from the correct pattern at  $H_c = \hat{H}_c$  when  $\hat{H}_c$  is varied.

## IX. DISCUSSION

The proposed human counting and activity recognition solution relies on the beamsteering capabilities of the receiver which relies on accurate steering vector calculation. Even though, new wireless standards such as IEEE 802.11ac, ax and 5G support such capabilities, currently there is no commodity hardware that can be utilized for our purpose, thus, we use USRP hardware. With the current expansion of research on device-free human sensing, we expect chipset manufacturers to expose such capabilities in a wide range of products.

The experiments were conducted with prototype radio equipment in a controlled lab environment. Up to 6 people were differently positioned in the same room (but in predefined spots for which antennas have been calibrated). When this is realized in real environments, the main factors degrading

the accuracy would be the multipath effects and coverage problems. We also note that the accuracy of our crowd counting technique degraded when the number of human subjects exceeded four. As the receiver ULA size has a correlation with people counting accuracy, we will, in further investigations, study settings with 8 and 16 antennas in order to increase the directional resolution and the count of people. If two persons happen to be in a straight line, e.g. in direct line of sight of TX and RX, in the current configuration, the system detects them as a single person. We note that this type of situations can be mitigated by using spatially separated multiple receivers. Another limitation is that the human count is estimated based on the motion of people. If a person is completely static, this could imply that the respective stream is vacant. This can be solved by using respiration detection techniques.

## X. CONCLUSION

We have reported on a device-free multiple human subject recognition approach exploiting limited antenna, narrowband reception devices using beamsteering capabilities. Leveraging observations from practical experiments, we develop methods to count people and extract individual activities in an ad-hoc manner. The people counting approach has been compared to a state-of-the-art algorithm on the same dataset and clearly outperforms the prior approach. This, in particular, is because our approach is less susceptible to interference due to movements from spatially distributed multiple subjects in the environment. Finally, we could demonstrate that the extraction of activity patterns is possible through blind source separation. The data we recorded in our study is openly available to ease comparison of future algorithms with our approach.

## ACKNOWLEDGMENT

The authors would like to thank Sanaz Kianoush at IEIT, Italy for helping with the experiments. This work has been funded by CHIST-ERA III, Radiosense project.

## REFERENCES

- [1] J. Wang *et al.*, "E-HIPA: An energy-efficient framework for high-precision multi-target-adaptive device-free localization," *IEEE Trans. on Mobile Comput.*, vol. 16, no. 3, pp. 716–729, 2016.
- [2] F. Adib *et al.*, "Smart Homes That Monitor Breathing and Heart Rate," in *Proc. of CHI*, 2015.
- [3] H. Zou *et al.*, "FreeCount: Device-free crowd counting with commodity wifi," in *Proc. of GLOBECOMM*. IEEE, 2017.
- [4] S. Sigg *et al.*, "Rf-based device-free recognition of simultaneously conducted activities," in *Proc. of Percom (Adjunct)*. ACM, 2013.
- [5] E. Bondi *et al.*, "Real-time people counting from depth imagery of crowded environments," in *Proc. of AVSS*. IEEE, 2014.
- [6] T. Wang, "Learning to do more with less: Adapting Campus security for lean times," 2012. [Online]. Available: <https://www.asmag.com/showpost/12628.aspx>
- [7] H. H. Cetinkaya and M. Akcay, "People counting at campuses," *Procedia-Social and Behavioral Sciences*, vol. 182, pp. 732–736, 2015.
- [8] "ISO/TS 15066:2016: Robots and robotic devices," Int. Organiz. for Standardization, 2016.
- [9] S. Palipana, N. Malm, and S. Sigg, "1x4 radio channel data at 3.42 GHz for device-free human sensing," Jul. 2019. [Online]. Available: <https://doi.org/10.5281/zenodo.3264887>
- [10] J. Kruger, T. Lien, and A. Verl, "Cooperation of human and machines in assembly lines," *CIRP Annals - Manufacturing Technology*, vol. 58, no. 2, pp. 628 – 646, 2009.

- [11] J. Fryman and B. Matthias, "Safety of industrial robots: From conventional to collaborative applications," in *Proc. of the German Conf. on Robotics*, 2012, pp. 1–5.
- [12] T. Higuchi *et al.*, "Mobile devices as an infrastructure: A survey of opportunistic sensing technology," *Journ. of Information Processing*, vol. 23, no. 2, pp. 94–104, 2015.
- [13] S. Savazzi *et al.*, "Device-free radio vision for assisted living: Leveraging wireless channel quality information for human sensing," *IEEE Signal Processing Magazine*, vol. 33, no. 2, pp. 45–58, 2016.
- [14] M. Seifeldin *et al.*, "Nuzzer: A large-scale device-free passive localization system for wireless environments," *IEEE Trans. on Mobile Comput.*, vol. 12, no. 7, pp. 1321–1334, 2012.
- [15] M. Kotaru *et al.*, "SpotFi: Decimeter Level Localization Using WiFi," in *Proc. of SIGCOMM*. ACM, 2015.
- [16] Q. Pu *et al.*, "Whole-home Gesture Recognition Using Wireless Signals," in *Proc. of Mobicom*, 2013.
- [17] C. Xu *et al.*, "SCPL: Indoor device-free multi-subject counting and localization using radio signal strength," in *Proc. of IPSN*. ACM, 2013.
- [18] H. Wang *et al.*, "Human Respiration Detection with Commodity WiFi Devices: Do User Location and Body Orientation Matter?" in *Proc. of Ubicomp*. ACM, 2016.
- [19] S. Sigg, U. Blanke, and G. Tröster, "The telepathic phone: Frictionless activity recognition from wifi-rssi," in *Proc. of PerCom*. IEEE, 2014.
- [20] P. Melgarejo *et al.*, "Leveraging Directional Antenna Capabilities for Fine-grained Gesture Recognition," in *Proc. of Ubicomp*, 2014.
- [21] S. Shi *et al.*, "Accurate location tracking from CSI-based passive device-free probabilistic fingerprinting," *IEEE Trans. on Veh. Technol.*, vol. 67, no. 6, pp. 5217–5230, 2018.
- [22] W. Wang *et al.*, "Gait recognition using wifi signals," in *Proc. of Ubicomp*. ACM, 2016.
- [23] S. Palipana *et al.*, "FallDeFi: Ubiquitous fall detection using commodity Wi-Fi devices," *Proc. of IMWUT*, vol. 1, no. 4, p. 155, 2018.
- [24] G. Wang *et al.*, "We can hear you with wi-fi!" *IEEE Trans. on Mobile Comput.*, vol. 15, no. 11, pp. 2907–2920, 2016.
- [25] T. Wei and X. Zhang, "mtrack: High-precision passive tracking using millimeter wave radios," in *Proc. of Mobicom*. ACM, 2015.
- [26] S. Palipana and S. Sigg, "Extracting Human Context Through Receiver-End Beamforming," *IEEE Access*, vol. 7, pp. 154 535–154 545, 2019.
- [27] W. Xi *et al.*, "Electronic frog eye: Counting crowd using wifi," in *In Proc. of INFOCOM*. IEEE, 2014.
- [28] F. Adib, Z. Kabelac, and D. Katabi, "Multi-Person Localization via RF Body Reflections," in *Proc. of NSDI*. USENIX, 2015.
- [29] S. Di Domenico *et al.*, "Trained-once device-free crowd counting and occupancy estimation using WiFi: A Doppler spectrum based approach," in *Proc. of WiMob*. IEEE, 2016.
- [30] S. Depatla, A. Muralidharan, and Y. Mostofi, "Occupancy estimation using only WiFi power measurements," *IEEE J. Sel. Areas Commun.*, vol. 33, no. 7, pp. 1381–1393, 2015.
- [31] G. Batista *et al.*, "A complexity-invariant distance measure for time series," in *Proc. of the Int. Conf. on Data Mining*. SIAM, 2011.
- [32] J. Paparrizos and L. Gravano, "k-shape: Efficient and accurate clustering of time series," in *Proc. of SIGMOD*. ACM, 2015.
- [33] E. J. Keogh and M. J. Pazzani, "Derivative dynamic time warping," in *Proc. of the Int. Conf. on Data Mining*. SIAM, 2001.
- [34] H. Sakoe and S. Chiba, "Dynamic programming algorithm optimization for spoken word recognition," *Readings in speech recognition*, vol. 159, p. 224, 1990.
- [35] N. Sharma *et al.*, "Comparison the various clustering algorithms of weka tools," *Facilities*, vol. 4, no. 7, pp. 78–80, 2012.
- [36] R. R. Sokal, "A statistical method for evaluating systematic relationship," *University of Kansas science bulletin*, vol. 28, pp. 1409–1438, 1958.
- [37] S. Salvador and P. Chan, "Determining the number of clusters/segments in hierarchical clustering/segmentation algorithms," in *Proc. of the Int. Conf. on Tools with Artificial Intelligence*. IEEE, 2004.
- [38] A. Hyvärinen and E. Oja, "Independent component analysis: algorithms and applications," *Neural networks*, vol. 13, no. 4-5, pp. 411–430, 2000.
- [39] J.-F. Cardoso and A. Souloumiac, "Blind beamforming for non-Gaussian signals," in *Proc. Radar and signal process.*, vol. 140, no. 6. IET, 1993, pp. 362–370.
- [40] R. Bro and A. K. Smilde, "Principal component analysis," *Analytical Methods*, vol. 6, no. 9, pp. 2812–2831, 2014.
- [41] N. Malm *et al.*, "User localization enabled ultra-dense network testbed," in *Proc. of 5GWF*. IEEE, 2018.