People Counting Using IR-UWB Radar Sensor in a Wide Area

Jae-Ho Choi, Ji-Eun Kim, and Kyung-Tae Kim, *Member, IEEE*

***Abstract*— Conventional radar-based people counting systems are designed mainly for dense spatial distributions in a small region of interest (ROI). Therefore, a system with only conventional energy-based features, which are effective for a small ROI with a limited spatial distribution of individuals, generally fails to cope with the diverse and complex spatial distributions that arise from the freer movements of individuals as the ROI widens. To address this problem, a novel approach that achieves robust people counting in both wide and small ROIs is presented in this study. The proposed technique incorporates modified CLEAN-based features in the range domain and energy-based features in the frequency domain to efficiently address both dense and dispersed distributions of individuals. Subsequently, principal component analysis and an appropriate normalization of the proposed features are performed for improving the people counting system further. Based on several experiments in practical environments with wide ROIs and severe multipath effects, we observed that the proposed approach yields significantly improved performance compared with traditional people counting systems.**

***Index Terms*— Feature extraction, impulse radio ultra-wideband (IR-UWB) radar, indoor sensing, modified CLEAN (MD-CLEAN), people counting.**

I. Introduction

SENSOR technology plays an important role in a variety of fields as it enables systems to automatically acquire information on the surroundings [1]. Owing to the development of the Internet of Things (IoT) technology, numerous applications that use the context information from sensors have been proposed [2]–[10]. For example, in the medical field, it is possible to monitor an individual's biological activity through sensors and, thus, prevent disease [2]–[4]. In the environmental monitoring field, the automatic detection of variations in the surroundings enables economical crop management, geological surveys, and road surveillance [5]–[7]. In future smart cities, information on the number of individuals in a certain area can facilitate efficient resource management and safety control [8]–[10].

With regard to sensing the density information of crowds, numerous people counting solutions have been studied. Current people counting systems can be categorized into device-based counting and device-free counting systems [11]. Device-based counting systems utilize the radio frequency (RF) signals from each individual's dedicated device, such as Zigbee, Bluetooth, WiFi, and radio frequency identification tags [12]-[18]. In device-based systems, people counting is conducted by analyzing the signals transmitted from each device and estimating the number of signal sources. Device-based approaches generally show superior counting performance than device-free counting approaches, but the individuals are required to carry dedicated devices, which causes significant inconvenience in everyday life [11].

Device-free counting approaches have been introduced to overcome the drawbacks of device-based counting approaches [19]–[31]. In device-free systems, indoor sensors such as cameras [19]–[23] and radars [24]–[31] are installed in the region of interest (ROI). People counting is then accomplished by analyzing the patterns of sensor signals, which are modulated by the movements of humans. In the majority of cases, device-free systems are based on vision sensors. In vision-based counting, the individuals do not need to carry any dedicated device. In addition, the informative optical reflections from individuals enable the vision-based systems to achieve satisfactory people counting accuracies. However, the intrinsic property of vision sensors raises concerns regarding privacy infringement and performance variation caused by surrounding illumination [25].

In contrast to vision sensors, low-power impulse radio ultra-wideband (IR-UWB) radar maintains robust performance under various weather and lighting conditions, while being free from privacy problems. Furthermore, it has the capability to provide fine-range resolution. These attributes make IR-UWB radar a potential candidate sensor to be used in people counting applications.

Nevertheless, there exist several challenging factors in adopting the IR-UWB radar for people counting. Because of regulations and size limitations [32], IR-UWB radar keeps low-power transmit signals, which eventually cause a low signal-to-noise ratio (SNR) at the radar receiver. However, the radar cross-section (RCS) of a human tends to maintain at a low level of approximately 0.5 m^2 [33], which is significantly smaller than that of surrounding objects, such as walls and furniture. Furthermore, it severely fluctuates depending on the individual's movement and relative orientation. Consequently, the weak SNR and RCS fluctuation issues prevent the credible separation of human signals from clutters and noise. Another

This research was supported by Energy Cloud R&D Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Science, ICT (NRF-2019M3F2A1073402).

The authors are with the Department of Electrical Engineering, Pohang University of Science and Technology, Pohang 790-784, South Korea (e-mail: jhchoi93@postech.ac.kr; jieun7@postech.ac.kr; kkt@postech.ac.kr).

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/JIOT.2020.3032710, IEEE Internet of Things Journal

difficulty is attributable to the ghost signals resulting from multipath propagation. Multipath signals have amplitude levels similar to those of human signals and, therefore, entail a severe over-counting problem. Moreover, the superposition of human signals occurs when over two individuals are positioned in the same range from the radar, resulting in an under-counting problem. In summary, it is essential to consider the clutter, multipath, and noise components in the reflected echoes, as well as the overlaps of individuals, to utilize the IR-UWB radar as a credible people counter.

Several methods have been proposed to employ the IR-UWB radar for people counting in a small ROI [27]-[31]. In a small ROI, it is likely that the returned echo in a certain range bin originates from the overlapped echoes of over two individuals because the movement of each individual is commonly restricted with a dense distribution. Therefore, conventional people counting methods have been based on the fact that the number of people is proportional to the reflected energy at each range bin, which has facilitated the use of energy-based features for people counting tasks. Meanwhile, for the case of scattered individuals in a wider area, the probability of occurrence of overlapped individuals in the same range bin would largely decrease. In addition, we observed that even if the superposition of human signals from the overlapped individuals occurs, as the distance increases, the energy of the returned echo is no longer proportional to the number of overlapped individuals. This causes conventional people counting methods, based only on energy-based features, to perform unreliably in a wider area. Furthermore, when the ROI widens, individuals are likely to wander to range bins far away from the radar, resulting in the degradation of the SNR. Besides, the probability that the reflected RCS of each person fluctuates increases because the individuals can have significantly freer movements, such as trots, sudden swings, or even runs. Consequently, these problems make people counting tasks in a wide area highly challenging, and this case has rarely been addressed in the available literature.

In this study, we provide effective solutions for the problems described previously, ensuring successful counting even in a wider area. It is necessary to consider a variety of spatial distributions of individuals, including scattered and dense distributions, to achieve reliable performance in a wider area. Accordingly, the proposed scheme devises energy-based features in the frequency domain as well as new ones in the range domain, which are based on modified CLEAN (MD-CLEAN), in order to mitigate the energy dependency of reflected echoes. In the range domain, the MD-CLEAN algorithm [28], [29] is used to identify peaks from individuals against noise and multipath ghosts. In particular, we introduce a novel multi-threshold (MT) scheme to enable the MD-CLEAN-based features to contain multiple sensitivities with respect to noise and multipath effects. That is, the MT scheme for MD-CLEAN yields a higher feature dimension, which enhances the separability of the classifier in the feature space of the given training templates by increasing the amount of information corresponding to the number of individuals. In the frequency domain, new energy-based features using the frequency spectrums of the received echoes are formed, but note that these are also based on MT, which results in multiple sensitivities of the features. The MT scheme employed in the range and frequency domains allows an improved separability of the classifier at the cost of an increased feature dimension. However, it also causes inflation of redundant information, which deteriorates computational efficiency and generalization performance. To prevent such problems and improve the generalization capability, we further carry out compression of the proposed features in the range and frequency domain via principal component analysis (PCA), followed by normalization. Consequently, the proposed MT features can better address the scattered (using MD-CLEAN-based features), as well as dense distributions (using energy-based features) in a wider area. The experimental results show that the proposed people counting system provides reliable predictions even in a wide ROI and significantly outperforms the conventional methods.

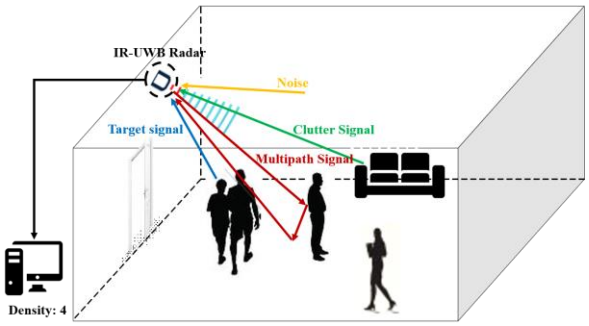

Fig. 1. Conceptual diagram of the people counting system using IR-UWB radar.

The remainder of this paper is organized as follows: In Section II, the effect of noise and multipath on the people counting system is analyzed. Then, we investigate the limitations of the conventional energy-based features in detail. In Section III, we discuss the proposed people counting system, which is based on the newly defined features. In Section IV, we demonstrate the performance of the proposed people counting solution based on the experimental results conducted in indoor environments. Finally, in Section V, we present the conclusions of the study and future work.

II. Motivation

A. Noise and Multipath Effect on MD-CLEAN

An overall conceptual diagram of the IR-UWB radar-based people counting system is illustrated in Fig. 1. As shown in the figure, the reflected signals of an IR-UWB radar can be regarded as a linear combination of human, multipath, clutter, and noise signals. In general, indoor clutters such as walls, pillars, and furniture have almost no movement. Therefore, the reflected signals from clutters have significantly smaller variance than human signals along the time domain. Based on this property, it is straightforward to suppress the clutter components in the received signals [34]–[37]. Meanwhile, ghost signals owing to multipath effects are highly similar to

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/JIOT.2020.3032710, IEEE Internet of Things Journal

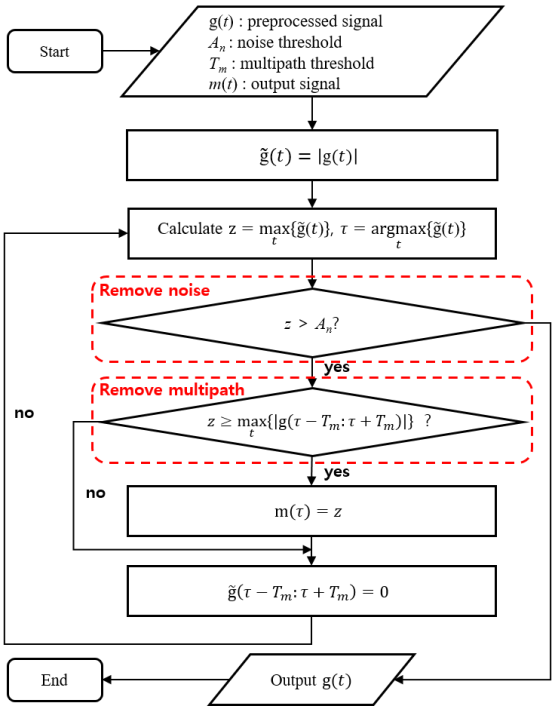

Fig. 2. Overall flow chart of the MD-CLEAN algorithm.

human signals, which hinders the separation of human peaks from multipath ghosts. In particular, single-bounced multipath components, which are reflected indirectly along both the targets and floor, have comparable SNR levels to those of human signals, resulting in severe overcounting problems. In addition, as individuals shift farther away from the IR-UWB radar, the resulting reflected signals with lower SNR also influence the accuracy of the people counting system adversely. Therefore, discrimination of human signals from multipath and noise components is essential for robust people counting capability, particularly for a wide area.

In [28], [29], and [34], the MD-CLEAN technique has been exploited to separate peaks of human signals from multipath ghosts and noise signals. This technique enables the detection of human peaks in an iterative manner while removing the peaks induced from noise and multipath ghosts to the extent possible. Specifically, the elimination of noise and multipath can be achieved based on the fact that multipath ghosts arise near each individual's signal and the noise level is generally lower than that of the human signal. For clear illustration, Fig. 2 shows the overall flowchart of the MD-CLEAN algorithm, in which human peaks can be extracted by removing both noise and multipath components from the preprocessed signal $g(t)$ via hard thresholding (based on the noise threshold A_n) and iterative windowing (based on the multipath threshold T_m), respectively.

It is noteworthy that one of the most crucial factors in the MD-CLEAN algorithm is the determination of the optimal thresholds A_n and T_m, which consequently affects the separation capability of human signals from noise and multipath ghosts. Fig. 3 shows the IR-UWB radar echo signal from four individuals after the removal of the clutter components using a running average (RA) filter [37]. As shown in Fig. 3, it is exceptionally challenging to determine the

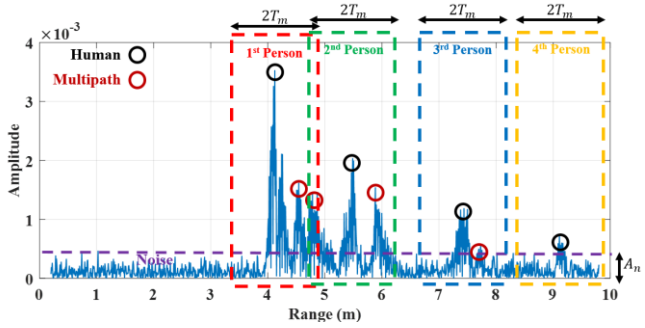

Fig. 3. Absolute levels of IR-UWB radar signal reflected from four individuals after clutter elimination using RA filter: the black circle in each dashed box denotes the human peak from each individual; the red circle in each dashed dox denotes the corresponding multipath peak.

optimum thresholds for completely reducing noise and multipath while preserving only the reflections from individuals. In other words, if A_n increases (i.e., the purple and horizontal dashed line in Fig. 3 becomes higher), the algorithm would yield a lower false alarm rate due to noise, though the detection rate of humans would also decrease, which results in the underestimation of the number of individuals. Meanwhile, if A_n decreases, the probability of detecting people increases, but the probability of false alarms by noise also tends to increase. This leads to overestimation of the number of individuals.

With regard to the removal of multipath ghosts, the MD-CLEAN selects only the single and local maximum peak as the human response within certain bounds T_m, as presented in Fig. 2. Accordingly, as T_m increases (i.e., each dashed box in Fig. 3 becomes wider), the algorithm can eliminate multipath ghosts more effectively. However, it cannot separate the reflected peaks from adjacent individuals, thereby underestimating the number of individuals. On the other hand, the separation capability of two adjacent individuals is improved when T_m decreases, but there is also an increasing possibility of detecting multipath ghosts, leading to overestimation problems.

Considering the abovementioned conflicting relationships that arise when the two MD-CLEAN thresholds are either increased or decreased, it is nearly impossible to determine the optimum thresholds, A_n and T_m, in the framework of the people counting task. Nevertheless, conventional people counting approaches based on MD-CLEAN attempted to compute a single optimum threshold by utilizing empirical or statistical solutions [28], [29], [34], [38]. This yielded limited performance owing to their features based on an imperfect single-threshold value for MD-CLEAN. Accordingly, there is a demand for a new strategy to handle the limitations of conventional schemes and, thus, design more advanced people counting systems. Inspired by this, we develop a novel paradigm suitable for people counting, which is based on the fusion of new features generated by MD-CLEAN with MT.

B. Limitations of Using Only Energy-Based Features

In addition to suppressing nuisance components, such as noise, clutter, and multipath ghosts, it is also essential to design appropriate features that contain only the information corresponding to the number of people to achieve credible people counting. In [29], the authors observed that the reflected

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/JIOT.2020.3032710, IEEE Internet of Things Journal

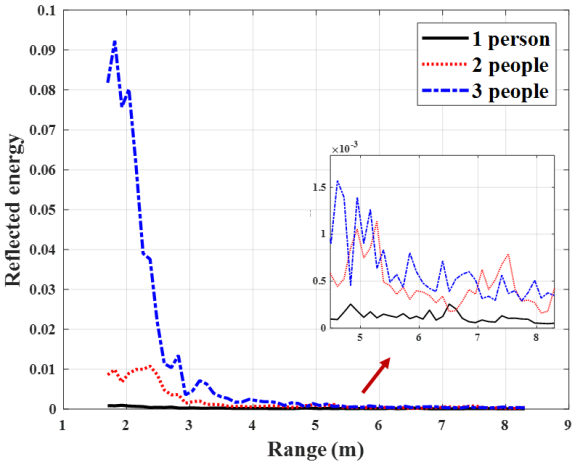

Fig. 4. Reflected energy when one, two, or three persons are located at a certain range from the radar.

signal peak in each range bin is approximately proportional to the number of people. Correspondingly, they performed stochastic fitting to the reflected energy level in each range bin, and then predicted the number of individuals using a maximum likelihood (ML) estimator based on the preformed probabilistic models. In [31], energy coefficients based on the curvelet domain and the range domain were deployed to obtain improved features, which are invariant from the moving direction of each individual, demonstrating remarkable improvements in dense people counting in a small ROI.

Although a few methodological differences exist, the conventional people counting systems share the common principle that the reflected signal energy would be proportional to the number of individuals, which facilitates the development of energy-based features. However, as the system coverage widens, each individual is likely to move around more freely in the expanded space. Hence, it would be less likely for the individuals to be located in the same range. In addition, even if individuals are at an equal distance, the differences in energy levels according to the number of people were found to be saturated abruptly as they shift farther from the radar.

To provide a typical example for illustration, Fig. 4 shows the average energy levels of the reflected IR-UWB radar signals when one, two, or three individuals are located at the same radar line-of-sight (RLOS) range bin. The reflected energy levels reveal substantial distinctions between 0 and 4.5 m, and they are proportional to the number of individuals. However, as the individuals are beyond 4.5 m, the reflected energies become saturated. Therefore, it would be exceptionally challenging to discriminate the number of people based only on the energy levels when they are far from the radar.

Considering the above results, it is evident that energy-based features alone cannot guarantee stable performance in a wide ROI although they are highly useful in a small ROI. Accordingly, to implement a reliable people counting system in a wider ROI, it is necessary to devise new and suitable features that are capable of discriminating the number of individuals even at a large distance.

Motivated by the previous discussion, in this study, we propose a new feature extraction scheme to overcome the drawbacks of energy-based features in a wide ROI. Specifically, MD-CLEAN-based features are newly devised for scattered distributions of individuals in a wider ROI, and energy-based features are further adopted in the frequency domain for the dense distributions. In the next section, we present a detailed explanation of the proposed people counting approach.

III. Methodology

A. Preprocessing

As described in Section II-*A*, the echoes received from an IR-UWB radar contain several undesirable components that preclude accurate counting of individuals. Thus, it is crucial to preprocess the raw IR-UWB radar signals to suppress the clutter and noise components to the extent possible.

The proposed preprocessing scheme consists of four steps: frame formation, clutter cancellation, matched filtering (MF), and adaptive gain control (AGC). These are discussed in detail below:

1) Frame Formation: For the first step, a sequential series of radar-received pulses is collected to form a frame. This process enables the people counting system to produce each prediction on a frame-by-frame basis rather than a pulse-by-pulse basis. By combining information from multiple pulses in a frame to reach a decision on the number of individuals, robust features can be generated compared to the case using only a single pulse.

After digitization using an analog-to-digital converter (ADC), the j-th received IR-UWB radar pulse $\mathbf{s}_j$ can be represented as

$$\mathbf{s}_j = [s_j(t_1), s_j(t_2), \ldots, s_j(t_i), \ldots, s_j(t_{N_r})]^T, \quad (1)$$

where $s_j(t)$ is the normalized amplitude of the j-th received pulse at the fast time t, $[t_1, t_2, \ldots, t_{N_r}]$ is the sampling time instant from the ADC, and N_r is the total number of data samples in the fast time domain. The j-th frame $\mathbf{S}[j]$ is then formed by collecting N_p radar pulses as follows:

$$\mathbf{S}[j] = \left[\mathbf{s}_j\, \mathbf{s}_{j+1} \cdots \mathbf{s}_{j+N_p-1}\right] = \begin{bmatrix} s_j(t_1) & s_{j+1}(t_1) & \cdots & s_{j+N_p-1}(t_1) \\ s_j(t_2) & s_{j+1}(t_2) & \cdots & s_{j+N_p-1}(t_2) \\ \vdots & \vdots & \ddots & \vdots \\ s_j(t_{N_r}) & s_{j+1}(t_{N_r}) & \cdots & s_{j+N_p-1}(t_{N_r}) \end{bmatrix}. \quad (2)$$

2) Clutter Cancellation: In indoor environments, the reflected signal energies arise mostly from clutter objects, and thus, the relatively weak signals of individuals are obscured. In general, most clutters in indoor environments remain stationary, whereas individuals show non-stationary movements. Even when a person stays still, his or her skin displays certain vibrations originating from breathing or heartbeats [39]. Thus, clutter signals can be eliminated effectively by suppressing all the stationary components in the received signals.

If the background signals can be measured *a priori* from an empty room, a simple method for suppressing clutter components is to simply subtract the background signals 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/JIOT.2020.3032710, IEEE Internet of Things Journal

the received signals. This is called the reference method (RM) [34], [35]. If the information on the background signal is unavailable, several techniques, such as singular value decomposition [36] and running average (RA) [37], can be adopted for clutter cancellation. These are capable of adaptively updating current clutter components from the radar echoes. We employ RA for clutter suppression in the proposed people counting system, owing to its property wherein it does not require pre-measurements and is computationally efficient. The RA filter continuously updates the clutter component for each pulse based on the current signal $\mathbf{s}_j$, so that the pulse-to-pulse variations in the clutter signal can be considered. The clutter-suppressed signal $\mathbf{u}_j$ can be obtained by applying the RA filter to each frame as follows:

$$\hat{\mathbf{e}}_j = \begin{cases} \dfrac{1}{N_p}\sum_{k=1}^{N_p}\mathbf{s}_k, & \text{if } j=1 \\ \gamma\hat{\mathbf{e}}_{j-1}+(1-\gamma)\mathbf{s}_j, & \text{if } j>1 \end{cases}$$

$$\mathbf{u}_j = \mathbf{s}_j - \hat{\mathbf{e}}_j, \tag{3}$$

where $\hat{\mathbf{e}}_j$ is the weighted average between the previously estimated clutter component $\hat{\mathbf{e}}_{j-1}$ and current echo $\mathbf{s}_j$. γ is the weighting factor between 0 and 1, which adjusts the amount of current echo signal $\mathbf{s}_j$ in the estimation of $\hat{\mathbf{e}}_j$ by comparing with the previously estimated $\hat{\mathbf{e}}_{j-1}$ in a recursive manner.

3) Matched Filtering: Subsequently, the MF technique is applied to the clutter-suppressed signal $\mathbf{u}_j$ to improve the SNR of the received signal:

$$\mathbf{a}_j = \mathbf{u}_j * \mathbf{r}, \tag{4}$$

where $\mathbf{a}_j$ is the filtered output signal from the MF process. $\mathbf{r}$ is a digitized format of the transmitted reference signal $r(t)$, which is modeled as a sinusoidal short pulse with a finite pulse width τ_{width}: [40]

$$r(t) = \cos(\omega_{center}t)e^{-\left(\frac{t}{\tau_{width}}\right)^2}, \tag{5}$$

where ω_{center} is the center frequency of the IR-UWB radar.

4) Adaptive Gain Control: In general, the degree of attenuation in the radar-received signals increases gradually as each individual shifts farther from the radar. This tendency causes the output features to be dependent upon the spatial distribution of individuals, namely, the feature values may fluctuate depending on whether the individuals are close or far from the radar. To address this problem, we applied the AGC process to the MF output $\mathbf{a}_j$. Formally, the AGC can be performed by multiplying different weight factors along the RLOS ranges, as follows:

$$\mathbf{g}_j = \left[a_j(t_1)\left(\frac{ct_1}{2}\right)^{\beta}, a_j(t_2)\left(\frac{ct_2}{2}\right)^{\beta}, \ldots, a_j(t_{N_r})\left(\frac{ct_{N_r}}{2}\right)^{\beta} \right], \tag{6}$$

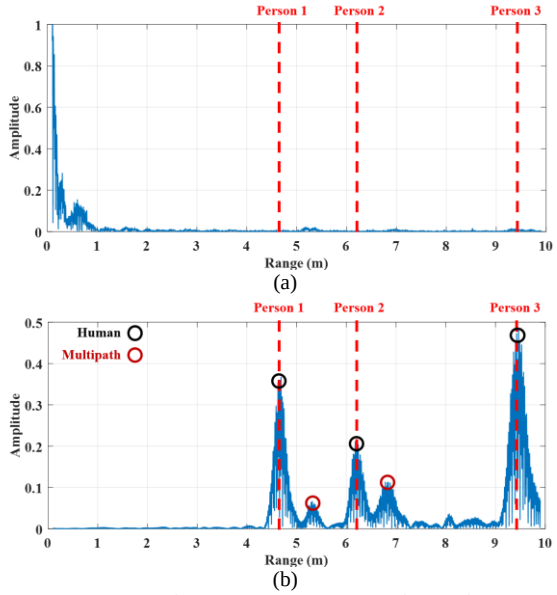

Fig. 5. Absolute values of IR-UWB radar signals reflected from three people, before and after preprocessing. (a) Reflected radar signal before preprocessing. (b) Reflected radar signal after preprocessing.

where $\mathbf{g}_j$ is the j-th output of the AGC process and c is the speed of light. Further, β can be selected as a value near two because the amplitude of a received radar signal is inversely proportional to the square of the distance. Finally, we can convert the j-th raw frame $\mathbf{S}[j]$ to the j-th preprocessed frame $\mathbf{G}[j]$:

$$\mathbf{G}[j] = \left[\mathbf{g}_j \, \mathbf{g}_{j+1} \cdots \mathbf{g}_{j+N_p-1} \right], \tag{7}$$

Fig. 5 shows the raw and preprocessed echoes when three individuals are located at 4.6, 6.2, and 9.5 m from the IR-UWB radar. From the raw signal in Fig. 5(a), it is impossible to identify the presence of individuals because the leakage components near 0 m, as well as the clutter components, conceal human signals. Meanwhile, the preprocessed signal in Fig. 5(b) clearly reveals three high peaks from the three individuals through the clutter cancellation process. Fig. 5(b) also shows that the AGC mitigates the attenuation problem of the individual signal, which is proportional to the distance from the radar, resulting in distinctive three peaks corresponding to each individual.

Still, it should be noted that the AGC process also increases the power of the peaks from the imperfectly removed clutter components as well as noise components, thereby yielding several small peaks at approximately 8 m in Fig. 5(b). Additionally, the preprocessing stage consisting of clutter cancellation, MF, and AGC is not capable of handling the undesirable high peaks owing to the multipath ghosts at approximately 5.3 m and 6.8 m. In the next section, we introduce a novel feature formation scheme based on the preprocessed signals, which is aimed at addressing the issues discussed above.

B. Feature Formation

Since the proposed system determines the number of people from each frame, the preprocessed frame **G** must be converted to a single feature vector. In this subsection, the detailed

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/JIOT.2020.3032710, IEEE Internet of Things Journal

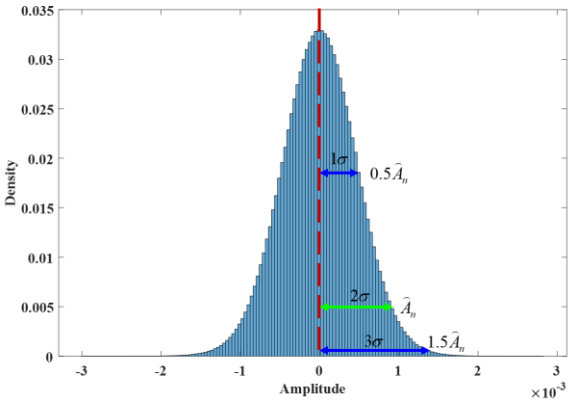

Fig. 6. Normalized histogram of the 5,000 radar signals measured from an empty ROI, where each signal is preprocessed with only two steps: clutter suppression and MF without AGC.

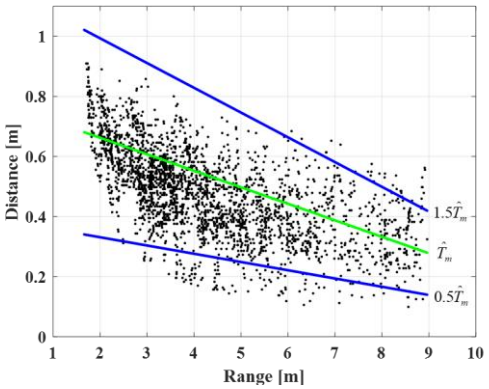

Fig. 7. Distance between each human peak and its corresponding multipath peak with respect to the RLOS ranges of individuals.

methodology of constituting a feature vector based on **G** is formulated. The proposed features can be established in two domains: the range domain and frequency domain. First, we devise a new framework for designing features in the range domain.

1) Features in the Range Domain: Note that the feature formation process in the range domain employs the MD-CLEAN algorithm (Fig. 2) on each preprocessed radar pulse (i.e., each column of **G**), in order to find only the desired human peaks. Thus, one of the most significant things in forming the range domain features is the extraction of only human-induced information while excluding the information from spurious components, such as noise and multipath ghosts. However, as discussed in Section II-*A*, the MD-CLEAN algorithm has critical shortcomings in terms of the selection of the optimal threshold values of A_n (for noise cancellation) and T_m (for multipath cancellation). Specifically, a tradeoff relationship exists in increasing or decreasing the thresholds, so it is substantially challenging to determine a single optimal value for A_n and T_m. Consequently, the incomplete, single MD-CLEAN output is likely to be biased only toward certain distributions, and thus, cannot guarantee the feasibility of robust people counting on the diverse spatial distributions in a wide ROI.

Strongly motivated by this, we devise a new feature formation approach in which the MD-CLEAN algorithm is coupled with the MT scheme. The proposed algorithm does not require a strictly optimum threshold value with respect to A_n and T_m, but rather attempts to exploit MT values, and then computes novel features based on those MT. That is, when the MD-CLEAN algorithm is applied to each preprocessed signal in a frame, i.e., each column of **G**, human peaks are extracted via MT candidates, not a single optimized value.

Before applying the MT scheme, we first compute the sub-optimal thresholds for the noise $\hat{A}_n$ and multipath $\hat{T}_m$, which would be used as baseline (reference) values, because it is exceptionally challenging to acquire perfectly optimal thresholds. For the noise threshold value $\hat{A}_n$, we exploited the received signals from the IR-UWB radar measured from an empty space. We collected 5,000 pulses from the empty ROI, and then applied only two preprocessing steps: clutter suppression (i.e., RA filtering) and MF without AGC. Assuming that the RA removed most of the clutter components, the resulting signals were mainly composed of noise, whose normalized histogram is illustrated in Fig. 6. From the figure, it can be easily noticed that the noise components after clutter suppression and MF approximately follow the Gaussian distribution with mean $\mu = 0$ and standard deviation $\sigma = 4.765 \times 10^{-4}$. Correspondingly, assuming that the background signals after clutter elimination are likely to follow the Gaussian distribution $\mathcal{N}(\mu,\sigma^2)$, the reference threshold for noise is set to 2σ level, which implies that approximately 95% of the noise components in a received pulse can be eliminated using the threshold. Because an additional step of AGC is also performed before MD-CLEAN, the baseline value of the amplitude threshold $\hat{A}_n$ is scaled and selected as

$$\hat{A}_n = 2\sigma\left(\frac{ct_{N_r}}{2}\right)^{\beta}. \tag{8}$$

Note that $\hat{A}_n$ can be utilized regardless of whether the signal has a positive or negative value, since the noise distribution in Fig. 6 is a nearly even function. Because $\hat{A}_n$ cannot serve as a perfectly optimal threshold, for each pulse **g**, MD-CLEAN is performed five times along five different amplitude thresholds $\mathbf{C}^A$, by increasing or decreasing it.

$$\begin{aligned}\mathbf{C}^A &= \left\{c_1^A, c_2^A, c_3^A, c_4^A, c_5^A\right\} \\ &= \left\{0.5\hat{A}_n, 0.75\hat{A}_n, \hat{A}_n, 1.25\hat{A}_n, 1.5\hat{A}_n\right\},\end{aligned} \tag{9}$$

where $\hat{A}_n$ is the baseline threshold. In particular, it should be noted that $\mathbf{C}^A$ covers from the 1σ level (i.e., approximately 68% of the noise components can be removed, and the human peaks with low SNR are more likely to be detected) to the 3σ level (i.e., approximately 99% of the noise components can be removed, but the human peaks with low SNR are less likely to be detected). This strategy that combines MD-CLEAN with the MT enables the features to retain multiple amplitude sensitivities, facilitating the generation of many augmented data for improving the generalization performance of a classifier, and the improved management of the conflict between noise and human peaks with low SNR.

Meanwhile, the determination of $\hat{T}_m$ is more complicated than that of $\hat{A}_n$ because the distance between each human peak and its corresponding multipath peak decreases as the individual shifts away from the radar. That is, the multipath threshold $\hat{T}_m$ is dependent on the individual's RLOS ranges. To address this difficulty and enhance the capability of eliminating multipath peaks, we set $\hat{T}_m$ as a function of R, unlike the conventional MD-CLEAN [28], [29], [34], [38]. Considering that the single-bounced components are dominant in multipath ghosts for indoor IR-UWB radar signals [41], the distance between a human peak and its corresponding single-bounced multipath peak has been measured along each RLOS range, as shown in Fig. 7. In this figure, each black point denotes the distance of a human peak and its corresponding multipath peak according to his or her RLOS range. Notably, it can be seen that there is a trend of decreasing distance (i.e., the human and multipath peaks become closer) as the individual shifts away from the radar. Based on these observations, the least square fitting is applied to determine the coefficients of the first-order polynomial for the baseline (reference) of the multipath threshold: $\hat{T}_m(R) = -0.1040R + 0.8714$, which is illustrated as the green solid line in Fig. 7.

Similar to the case of $\hat{A}_n$, MD-CLEAN is performed five times for each pulse along five different multipath thresholds $\mathbf{C}^T$ as follows:

$$\begin{aligned}\mathbf{C}^T &= \left\{c_1^T,\, c_2^T,\, c_3^T,\, c_4^T,\, c_5^T\right\} \\ &= \left\{0.5\hat{T}_m,\, 0.75\hat{T}_m,\, \hat{T}_m,\, 1.25\hat{T}_m,\, 1.5\hat{T}_m\right\},\end{aligned} \tag{10}$$

where $\hat{T}_m$ is the baseline threshold. As shown by the uppermost blue solid line ($1.5\hat{T}_m$ of Fig. 7), most of the multipath peaks can be eliminated via $1.5\hat{T}_m$, but the adjacent human peaks are difficult to be distinguished (The points under the $1.5\hat{T}_m$ line are considered as multipath ghosts, so they would be removed.) Meanwhile, the MD-CLEAN with the bottom-most $0.5\hat{T}_m$ line is inefficient for removing multipath ghosts but beneficial for separating the adjacent human peaks. Thus, by varying the threshold values in the range domain from $0.5\hat{T}_m$ to $1.5\hat{T}_m$, we can better handle the tradeoff between the removal of multipath ghosts and the detection of the adjacent human peaks. Consequently, this also enables the MD-CLEAN to extract peaks with multiple sensitivities in the range domain and produces many augmented training data to improve the generalization.

Finally, the thresholds for noise and multipath can be selected from the candidate sets $\mathbf{C}^A$ and $\mathbf{C}^T$, respectively, allowing the MD-CLEAN to retain multiple sensitivities in both amplitude and range domains. By combining each element of $\mathbf{C}^A$ and $\mathbf{C}^T$, we can form a threshold set consisting of 25 (c_x^A, c_y^T)'s with ($x, y = 1, 2, \ldots, 5$) for each pulse $\mathbf{g}$. Accordingly, MD-CLEAN is performed 25 times for each preprocessed pulse $\mathbf{g}$ according to 25 (c_x^A, c_y^T) values. Let $\mathbf{M} = \left\{\mathbf{M}^{(c_x^A, c_y^T)} \,|\, x, y = 1, 2, \ldots, 5\right\}$ be a 25-length set of MD-CLEAN outcomes, where each matrix $\mathbf{M}^{(c_x^A, c_y^T)} \in \mathbb{R}^{N_r \times N_p}$ denotes an output of MD-CLEAN from the input $\mathbf{G}$ with a specific noise threshold c_x^A and specific multipath threshold c_y^T:

$$\mathbf{M}^{(c_x^A, c_y^T)} = \begin{bmatrix} m_{1,j}^{(c_x^A, c_y^T)} & m_{1,j+1}^{(c_x^A, c_y^T)} & \cdots & m_{1,j+N_p-1}^{(c_x^A, c_y^T)} \\ m_{2,j}^{(c_x^A, c_y^T)} & m_{2,j+1}^{(c_x^A, c_y^T)} & \cdots & m_{2,j+N_p-1}^{(c_x^A, c_y^T)} \\ \vdots & \vdots & \ddots & \vdots \\ m_{N_r,j}^{(c_x^A, c_y^T)} & m_{N_r,j+1}^{(c_x^A, c_y^T)} & \cdots & m_{N_r,j+N_p-1}^{(c_x^A, c_y^T)} \end{bmatrix}. \tag{11}$$

Then, its element $m_{i,j}^{(c_x^A, c_y^T)}$ depends on whether an object exists at a certain fast time instant t_i by the following relationship:

$$m_{i,j}^{(c_x^A, c_y^T)} = \begin{cases} 0, & \text{no humans on } t_i \\ \left|g_j(t_i)\right|, & \text{otherwise} \end{cases}. \tag{12}$$

As a result, for a single frame, we can form 25 matrices of $\mathbf{M}^{(c_x^A, c_y^T)}$ as the MD-CLEAN output.

Based on each MD-CLEAN output $\mathbf{M}^{(c_x^A, c_y^T)}$, two types of range domain features are proposed, which are likely to be proportional to the number of individuals. Considering that the number of peaks extracted from MD-CLEAN increases gradually as the number of individuals increases, the first type of feature $f_1^{(c_x^A, c_y^T)}$ is defined as the overall count of the extracted peaks in $\mathbf{M}^{(c_x^A, c_y^T)}$:

$$\begin{aligned} f_1^{(c_x^A, c_y^T)} &= \underset{i,j}{\text{Count}}\left\{m_{i,j}^{(c_x^A, c_y^T)} > 0\right\} \\ &(i = 1, 2, \ldots, N_r,\, j = 1, 2, \ldots, N_p). \end{aligned} \tag{13}$$

Next, based on the fact that the location of the extracted peaks using MD-CLEAN is more likely to be scattered as the number of individuals increases, the second type of feature $f_2^{(c_x^A, c_y^T)}$ is defined as follows:

$$f_2^{(c_x^A, c_y^T)} = \frac{\sum_{i=1}^{N_r} \underset{j}{E}\left(m_{i,j}^{(c_x^A, c_y^T)}\right)}{N_r}. \tag{14}$$

In (14), $E(\cdot)$ denotes the Shannon entropy which measures the degree of uncertainty of the underlying distribution, and can be formulated as

$$\underset{j}{E}\left(m_{i,j}^{(c_x^A, c_y^T)}\right) = -\sum_{j=1}^{N_p} p_j \log(p_j), \tag{15}$$

where

$$p_j = \frac{m_{i,j}^{(c_x^A, c_y^T)}}{\sum_{j=1}^{N_p} m_{i,j}^{(c_x^A, c_y^T)}}.$$

Note that because there are 25 MD-CLEAN outputs $\mathbf{M}^{(c_x^A, c_y^T)}$, either $f_1^{(c_x^A, c_y^T)}$ or $f_2^{(c_x^A, c_y^T)}$ also has 25 elements ($x, y = 1, 2, \ldots, 5$).

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/JIOT.2020.3032710, IEEE Internet of Things Journal

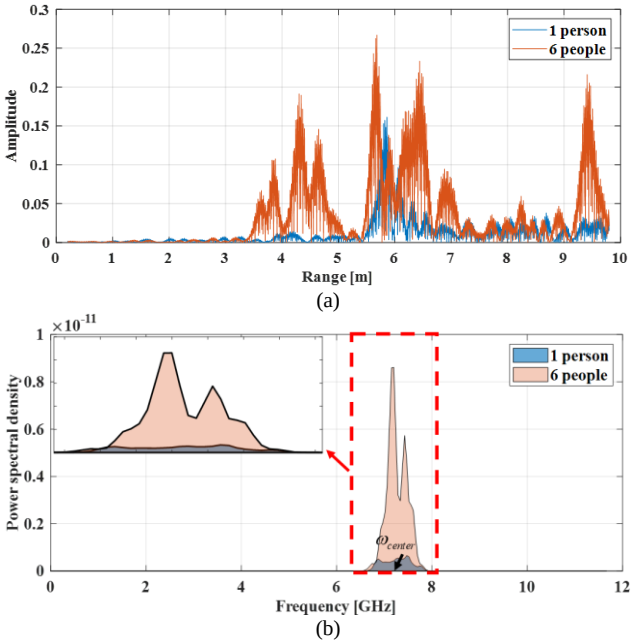

Fig. 8. Absolute values of preprocessed echo and corresponding PSD when there are one or six individuals in the ROI. (a) Preprocessed radar signal in range domain. (b) Corresponding PSD in frequency domain.

Assuming that the conventional MD-CLEAN with single-threshold scheme is utilized, one must select from one of two conflicting requirements: a high threshold for eliminating undesired components, such as noise and multipath at the expense of weak and indistinguishable human signals reflected from distant or adjacent individuals, and a low threshold for detecting weak and indistinguishable human signals while involving undesired noise and multipath signals. On the other hand, the proposed scheme with MT contemplates both the pros and cons of the two tradeoff requirements, while at the same time alleviating the disadvantage of detecting false alarms from low thresholds through the subsequent PCA. Moreover, the proposed MD-CLEAN-based features $f_1^{(c_x^A, c_y^T)}$ and $f_2^{(c_x^A, c_y^T)}$ can minimize the dependency on the amplitude (i.e., energy) levels of the extracted peaks, which cause saturation problems at large distances (as discussed in Section II-*B*). Consequently, the proposed features coupled with MT scheme enable us to develop a people counting system that can handle the complex spatial distributions of individuals, so that the entire detectable range of the radar can be utilized.

2) Features in the Frequency Domain: The two types of range-domain features $f_1^{(c_x^A, c_y^T)}$ and $f_2^{(c_x^A, c_y^T)}$ are highly effective when the individuals are spread over a wide area. However, these become vulnerable in cases where individuals are highly grouped in a small region. This is attributable to the intrinsic property of MD-CLEAN that generally extracts only one output peak even when there are several individuals in the same range bin. That is, individuals in the same range bin are likely to be indistinguishable based only on the range domain features. To overcome this problem, we propose new energy-based features in the frequency domain, which are effective particularly for dense cases. Thus, the proposed people counting system is capable of maintaining robust performance on both scattered and dense distributions by combining the MD-CLEAN-based features in the range domain and the energy-based features in the frequency domain together.

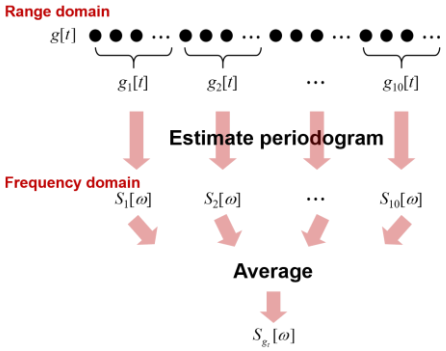

Fig. 9. Detailed process of the Bartlett method.

Note that the preprocessed signals of the IR-UWB radar can be considered as a linear combination of delayed transmit signals. Namely, let $g[t]$ represent a preprocessed signal with respect to t, then it can be expressed as

$$g[t] = \sum_k \alpha_k s\left[t - \frac{2R_k}{c}\right] + n[t], \tag{16}$$

where α_k and R_k denote the reflectivity and RLOS range, respectively, of each individual. Further, $s[t]$ and $n[t]$ are the transmit pulse and noise, respectively, with respect to t. Then, taking the Fourier transform (FT) along t, the output signal $\mathcal{F}\{g[t]\}$ is represented as follows:

$$\mathcal{F}\{g[t]\} = \sum_k \alpha_k \mathcal{F}\{s[t]\} e^{-j2R_k\omega/c} + \mathcal{F}\{n[t]\}, \tag{17}$$

where $\mathcal{F}\{\cdot\}$ denotes the FT operator. Based on (17), $|\mathcal{F}\{g[t]\}|$ can be considered as a weighted sum of $|\mathcal{F}\{s[t]\}|$ in the transmit frequency band. In other words, all the energies reflected from the individuals in the range domain are concentrated in a transmit band in the frequency domain regardless of their distances from the radar, thereby yielding translational invariance. Hence, according to Parseval's relation [42] and (17), as the number of individuals increases, there would be a simultaneous increase of the energy level of $g[t]$ in the range domain and that of the corresponding spectrum $|\mathcal{F}\{g[t]\}|$ in the transmitting frequency band.

To provide a clear illustration, Fig. 8 presents the absolute values of the preprocessed radar echo signals (Fig. 8(a)) and its power spectral density (PSD) (Fig. 8(b)) when one or six individuals exist in the area. The figure reveals that the wide distribution of the reflected energy from each individual in the range domain becomes highly concentrated in the transmitting frequency band between 6.54 GHz and 8.04 GHz. Motivated by this property of the signals in the frequency domain, we devise a new energy-based feature f_3 as the average energy level of each preprocessed pulse in a frame, which can simply be calculated by integrating the PSD of each preprocessed range profile within a certain bandwidth around the center frequency ω_{center}, and then averaging the N_p energy coefficients:

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/JIOT.2020.3032710, IEEE Internet of Things Journal

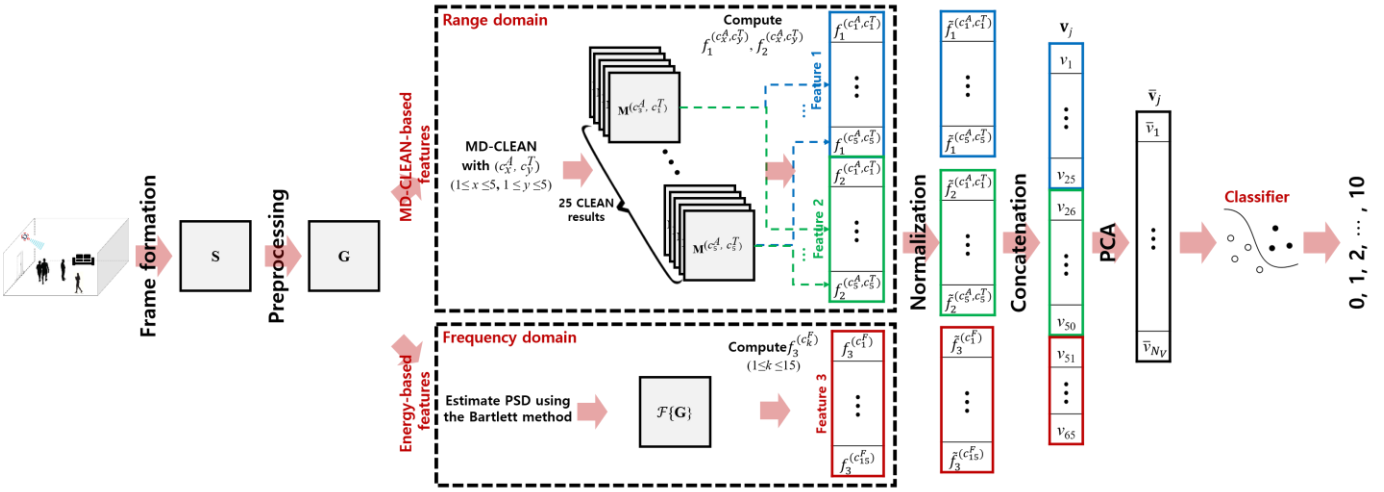

Fig. 10. Overall framework of the proposed people counting method.

$$f_3^{(c_k^F)} = \frac{\sum_{j=1}^{N_p} \sum_{\omega=\omega_{center}-c_k^F/2}^{\omega=\omega_{center}+c_k^F/2} S_{\mathbf{g}_j}[\omega]}{N_p}, \tag{18}$$

where $S_{\mathbf{g}_j}[\omega]$ is the estimated PSD of each preprocessed signal $\mathbf{g}_j$. c_k^F determines the cutoff frequency for the integration of the PSD. For PSD estimation, we adopt the Bartlett method (Fig. 9), which is superior in terms of variance, compared to periodogram [43]. c_k^F determines the cutoff frequency for the integration of the PSD. It is noteworthy that because the AGC step in the preprocessing stage equalizes the received signal levels at different RLOS ranges, the devised frequency-domain feature $f_3^{(c_k^F)}$ is likely to be invariant to variations in energy levels caused by the distance variations of individuals.

It should be noted that an increase or decrease in c_k^F generates a conflicting relationship as in c_x^A and c_y^T, hindering the selection of the optimal c_k^F. Specifically, as c_k^F increases, most of the energy caused by individuals is likely to be included in $f_3^{(c_k^F)}$, whereas the influence of noise on $f_3^{(c_k^F)}$ also becomes pronounced. In contrast, as c_k^F decreases, we can exclude noise components in $f_3^{(c_k^F)}$ more effectively but there is an increasing possibility of losing weak signals from individuals. To solve this tradeoff issue in terms of the determination of the optimal c_k^F, we apply the MT scheme again as in the case of the range domain features. Considering the congregation of reflected energies in the frequency domain within the transmitting bandwidth, the MT candidates for the cutoff bound c_k^F is set by dividing the transmitting bandwidth into 15 steps. Let ω_B denote the transmitting bandwidth of the radar. The candidate set $\mathbf{C}^F$ for c_k^F is formed as

$$\mathbf{C}^F = \left\{ c_k^F \;\middle|\; k = 1, 2, \ldots, 15 \right\}$$

$$= \left\{ \frac{k\omega_B}{15} \;\middle|\; k = 1, 2, \ldots, 15 \right\}, \tag{19}$$

Consequently, the frequency domain features $f_3^{(c_k^F)}$ with $k = 1, 2, \ldots, 15$ can achieve multiple sensitivity as in the case of the range domain features.

3) Feature Fusion: The final feature vector is formed by fusing the three types of features defined in the range and frequency domains. Namely, each of $f_1^{(c_x^A, c_y^T)}$ and $f_2^{(c_x^A, c_y^T)}$ in the range domain that provides a feature with a dimension of 25 respectively, and $f_3^{(c_k^F)}$ in the frequency domain that provides a feature with a dimension of 15, are first normalized to hold similar dynamic range level and statistical characteristics to each other, and then concatenated to develop a single feature vector. Let $\tilde{f}$ denote the normalized element of each feature. Then, the fused feature vector $\mathbf{v} \in \mathbb{R}^{65\times 1}$ is formulated as follows:

$$\mathbf{v} = \left[v_1, v_2, \ldots, v_{65} \right]^T, \tag{20}$$

where

$$\begin{aligned}
v_1 &= \tilde{f}_1^{(c_1^A, c_1^T)}, \\
v_2 &= \tilde{f}_1^{(c_1^A, c_2^T)}, \\
&\vdots \\
v_{25} &= \tilde{f}_1^{(c_5^A, c_5^T)}, \\
v_{26} &= \tilde{f}_2^{(c_1^A, c_1^T)}, \\
&\vdots \\
v_{50} &= \tilde{f}_2^{(c_5^A, c_5^T)}, \\
v_{51} &= \tilde{f}_3^{(c_1^F)}, \\
&\vdots \\
v_{65} &= \tilde{f}_3^{(c_{15}^F)}.
\end{aligned}$$

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/JIOT.2020.3032710, IEEE Internet of Things Journal

With the proposed MT strategy, the feature vector **v** attains the advantages of multiple sensitivities along the amplitude, range, and frequency domains, which enables the system to consider a variety of spatial distributions of freely moving individuals. However, the MT scheme has the problem of involving redundant information in **v**, leading the classifier to show degraded capability in terms of computational complexity and generalization. In addition, the proposed feature **v** is also likely to contain a few undesired components from the noise and multipath because weak signals from individuals generally compete with those signals. That is, **v** becomes dependent not only on the number of people but also on certain undesired components. Thus, to mitigate the influence of the undesired signals and minimize the redundancy, we adopt the PCA to project the information in **v** into a lower dimensional space where the feature variance is maximized. In terms of feature fusion, this can also be regarded as a process of generating a compound feature set by assigning a different weight to each element of **v** according to its importance (i.e., feature variance) [44]. PCA enables **v** to be compressed effectively while suppressing insignificant information. Consequently, each 65-dimensional feature vector $\mathbf{v} \in \mathbb{R}^{65\times 1}$ corresponding to each frame is compressed into the N_V-dimensional feature vector $\bar{\mathbf{v}} \in \mathbb{R}^{N_V\times 1}$ $(65 > N_V)$ using PCA, making **v** insensitive to marginal variations caused by undesired components. Considering both the people counting performance and compression ratio, N_V was empirically selected as 15, where 95% of the original data variance can be retained after PCA.

4) People Counting Based on Classifier: The proposed people counting system employs the machine learning framework, in which a linear or non-linear classifier is trained with feature vector data and corresponding labels (i.e., the number of people), enabling the classifier to predict the number of individuals from newly received signals. Let a set of the training database is given as $\left\{\left(\bar{\mathbf{v}}_1, y_1\right), \cdots, \left(\bar{\mathbf{v}}_{N_T}, y_{N_T}\right)\right\}$, where $\bar{\mathbf{v}}$ is the input feature vector of dimension N_V, $y \in \{0, 1, \cdots, P-1\}$ denotes the output label of total P classes ($P-1$: the maximum number of countable individuals), and N_T is the total number of frames in the training database.

A variety of classifiers, such as naïve Bayes, random forest, support vector machine (SVM), and multi-layer perceptron (MLP), can be trained using the normalized feature database. Finally, the classifier can estimate the target class (i.e., the number of individuals) from an arbitrary input feature vector. When newly reflected signals are received from the radar, a feature vector at each frame is formed through the aforementioned preprocessing and feature formation scheme. Then, the pre-trained classifier can accomplish automatic people counting based on the new input feature vector. The overall procedure of the proposed people counting method is illustrated in Fig. 10.

IV. Experimental Results

In this section, several experimental results based on real data and their evaluations are presented to verify the effectiveness of the proposed people counting approach.

TABLE I
Specifications of the X4M03 IR-UWB Radar

Carrier frequency	7.29 GHz
Frequency bandwidth	1.5 GHz
Pulse repetition frequency	50 Hz
Tx peak pulse power	6.3 dBm >
Rx noise figure	6.7 dB
Rx sampling rate	23.328 GHz
Pulse width	65.8 ns
Interface	USB

(a)

(b)

Fig. 11. Experimental environments: (a) Environment-I: open indoor lobby with no walls and a high ceiling; (b) Environment-II: closed indoor hall with two sides blocked by walls and with a low ceiling.

A. Experimental Setup

In the experiments, we used a commercial IR-UWB radar (X4M03, Novelda Inc.) to collect the radar signals reflected from individuals in real indoor environments. As listed in Table I, the X4M03 transmits impulse-like UWB signals with a center frequency of 7.29 GHz and a bandwidth of 1.5 GHz, resulting in a range resolution of 10 cm. For further details, the datasheet on the X4M03 radar is publicly provided [40], which includes its overall block diagram and hardware configurations.

The experiments were performed in two indoor environments: one is a lobby (referred to as Environment-I) that is an open indoor environment with no nearby walls and a high ceiling (see Fig. 11a), and the other is a hall (referred to as Environment-II) that is a rather closed environment with two sides blocked by walls and a low ceiling (see Fig. 11b). Therefore, the echo signal from the latter environment is more likely to be suffered

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/JIOT.2020.3032710, IEEE Internet of Things Journal

TABLE II
PHYSICAL CONDITIONS OF THE 16 PARTICIPANTS

Individual	A	B	C	D	E	F
Gender	W	M	M	M	W	M
Height (cm)	167	187	186	178	159	179
Weight (kg)	60	84	73	83	48	75
Individual	G	H	I	J	K	L
Gender	M	M	M	M	W	M
Height (cm)	173	176	192	174	170	169
Weight (kg)	69	69	93	73	68	60
Individual	M	N	O	P		
Gender	M	M	W	M		
Height (cm)	181	175	163	180		
Weight (kg)	76	72	57	80		

from multipath fading. In each indoor environment, up to ten people were permitted to move around freely in a fan-shaped space. Then, we installed the IR-UWB radar sensor at a height of 2 m near the apex of each space to collect the reflected signals of the individuals inside.

Based on the above experimental setup, we constructed three types of training and test datasets by altering the measurement spaces and environmental conditions. The first (Dataset-1) and second datasets (Dataset-2) were obtained from Environment-I and Environment-II, respectively, within a fan-shaped region with a radius of 10 m and a field of view (FOV) of 80°. The third dataset (Dataset-3) was collected again from Environment-II, but in this case, within a fan-shaped region with a radius of 5 m and a FOV of 80°. This configuration yielded more restricted movements of each individual, compared with the first and second datasets. Under each condition, the reflected signals for training were collected for 10 min per number of individuals (from 0 to 10). Since X4M03 can transmit and receive 50 pulses per second, the constructed training dataset contains approximately 30,000 reflected pulses for each number of individuals. After a few days, we reinstalled the IR-UWB radar at a position near where it had previously been placed for training. Then, the reflected signals were recorded again for approximately 5 min (i.e., approximately 15,000 pulses) for each number of individuals, to obtain test dataset. Note that the individuals for each measurement were selected randomly from 16 participants (the physical information of these participants is summarized in Table II). Thus, we intended to evaluate the people counting system against time variations, marginal alterations in the radar's installed position, and variations in the appearance of individuals. The three types of constructed datasets are summarized in Table III.

To the best of our knowledge, our experimental spaces for Dataset-1 and Dataset-2 are significantly wider than those of previous radar-based people counting works (maximum length of 5 m) [29]-[31]. Namely, in our experiments, each individual is provided at least four times as much space as in the previous works. Accordingly, the individuals in these spaces performed significantly freer actions such as trots, abrupt swings, or even runs, rather than slow walks or standings. Consequently, a variety of different spatial distributions of individuals, such as dense ones (including the case all of them are almost close to each other, i.e., a density of about four people per square meter), widespread ones (including the case they are all separated from each other, i.e., a density of not more than one person per square meter), those with individuals near the radar, and those with individuals far from the radar, could be considered during the measurements. However, our scenario does not include some challenging situations, such as people sitting still in chairs or lying down in beds. In such cases, even their micro-motions arisen from respiratory motions become restricted, making it highly difficult to distinguish them from static clutter, which is beyond our study.

For processing the received signals, we used MATLAB R2019 software running in Windows 10 on an Intel i7-9700K CPU processor and an Nvidia Titan RTX GPU processor.The number of pulses N_p in each frame was selected as 50 in the people counting experiments in this study. Therefore, the system can declare the number of individuals every 1 s.

B. Performance Analysis

In this subsection, we present extensive investigations on the performance of the proposed people counting system from several different perspectives.

1) Proposed Features Combined with Different Classifiers: First, we compared the people counting accuracies when the proposed features were combined with several classifiers. Specifically, we applied the proposed features as the input to four classifiers: naïve Bayes, random forest, SVM, and MLP classifiers. The random forest classifier was created with 200 decision trees, and the SVM classifier was tuned using a five-fold cross-validation technique. For MLP, the network was built with one input layer with 15 neurons, two hidden layers with 15 neurons each, and one output layer with 11 neurons.

Fig. 12 summarizes the outcomes of the four classifiers based on Dataset-1. In the figure, the blue bars represent the probability of correct classifications in terms of strict accuracy, i.e., the number of individuals is counted correctly with no errors. The red bars represent the probability of classifications in terms of rather relaxed accuracy, permitting counting errors of ±1. The outcomes of the red bars (i.e., relaxed accuracy) show that the proposed feature can perform nearly accurate people counting for up to 10 individuals irrespective of the classifier type. However, the blue bars (disallowing counting errors of ±1) display that the four classifiers produce some incorrect estimations. Nevertheless, the random forest classifier outperforms the other classifiers. Therefore, the random forest will be combined with the proposed features for the remaining people counting experiments in this study.

2) Efficacy of the Multi-threshold Scheme: The performances in terms of test accuracy between the two people counting systems (one implemented with the proposed MT scheme and the other implemented without the MT scheme) were compared to verify the efficacy of the proposed MT scheme when applied in the range and frequency domain features. For the system without the MT scheme, the optimum single threshold for each proposed feature f_1, f_2 (MD-CLEAN-based features in the range domain), and f_3 (energy-based feature in the frequency domain) were determined empirically by carrying out numerous people counting experiments.

Fig. 13 summarizes the test accuracies of the two people counting systems based on Dataset-1. In the figure, the two bars

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/JIOT.2020.3032710, IEEE Internet of Things Journal

TABLE III
DESCRIPTION OF THE THREE COLLECTED DATASETS

	Dataset-1	Dataset-2	Dataset-3
Training (0–10 people)	30,000 pulses for each no. of individuals	30,000 pulses for each no. of individuals	30,000 pulses for each no. of individuals
Test (0–10 people)	15,000 pulses per each no. of individuals (different days & individuals)	15,000 pulses for each no. of individuals (different days & individuals)	15,000 pulses per each no. of individuals (different days & individuals)
Experimental Area	Environment-I (Fan-shaped space with radius of 10 m and FOV of 80°)	Environment-II (Fan-shaped space with radius of 10 m and FOV 80°)	Environment-II (Fan-shaped space with radius of 5 m and FOV of 80°)

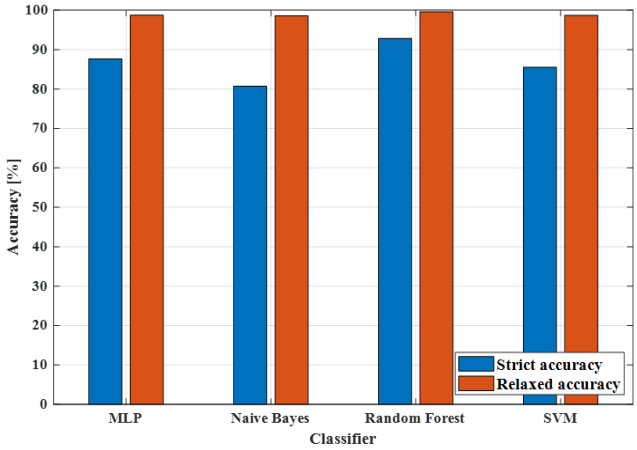

Fig. 12. People counting performance (test accuracy based on Dataset-1, %) of the proposed features with respect to four classifiers.

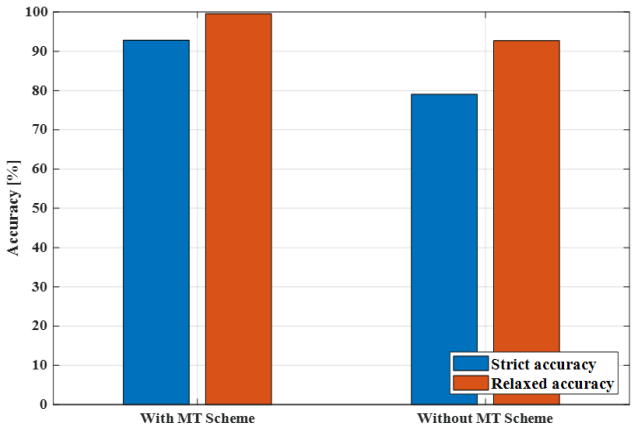

Fig. 13. Performance (test accuracy based on Dataset-1, %) comparison of the two people counting systems: one implemented based on MT scheme and the other implemented without MT scheme.

on the left represent the strict and relaxed accuracies of the proposed people counting system with the MT scheme. The two bars on the right represent the results of the people counting system without the MT scheme. The results illustrate that the proposed MT strategy can improve the people counting performance substantially (by 13.78% and 6.89% in terms of strict and relaxed accuracy, respectively), compared with the one without the MT strategy.

Based on the results, it can be concluded that the MT strategy lets each feature be multi-sensitive and facilitates a better consideration of severe fluctuations resulting from the complicated spatial distributions, thereby achieving a robust people counting even in a wide ROI.

3) Comparison with State-of-the-Art Radar-Based People Counting Systems: In this sub-section, a comparative analysis is described between the proposed people counting method and three state-of-the-art people counting methods: probabilistic model (PM)-based people counting [29], convolutional neural network (CNN)-based people counting [30], and curvelet transform (CT)-based people counting [31].

Table IV summarizes the strict and relaxed accuracies of each people counting method with respect to the three types of constructed datasets (Dataset-I, Dataset-II, and Dataset-III). It should be noted that the outcomes shown in the table were evaluated based on the test datasets measured against different days, different individuals, and marginally different radar positions in comparison with the training datasets, reflecting more realistic situations. It can be observed that among the three existing methods (top three rows of Table IV), the CT-based method in [31] outperforms the other two methods [29], [30] in terms of both strict and relaxed accuracies. In addition, it is noteworthy that the CNN-based people counting method in [30] yielded the worst outcomes: this failure originates from the property of the deep neural network wherein the large-size architecture directly learns clutter-specific information as well as the human-specific one from the received signals. This, in turn, causes the people counting system to be substantially sensitive to even marginal variations in the clutter environments, thereby resulting in severe overfitting problems.

Meanwhile, the outcomes of the proposed method (the last row in Table IV) reveal a highly improved performance compared with those of the three conventional methods, regardless of the environmental conditions. Specifically, in terms of strict accuracy, the proposed method yields 11.89% higher people counting accuracy for Dataset-1 (a wide ROI in Environment-I), 12.27% for Dataset-2 (a wide ROI in Environment-II), and 0.85% for Dataset-3 (a small ROI in Environment-II), than the most outstanding one out of the existing techniques (i.e., CT-based people counting). In terms of relaxed accuracy, the proposed people counting method has slightly improved performance for Dataset-2 and Dataset-3, and a similar one for Dataset-1.

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/JIOT.2020.3032710, IEEE Internet of Things Journal

TABLE IV
PERFORMANCE (TEST ACCURACY, %) COMPARISON OF THE THREE CONVENTIONAL PEOPLE COUNTING METHODS AND THE PROPOSED MODEL

	Dataset-1 (Wide ROI in Environment-I)		**Dataset-2 (Wide ROI in Environment-II)**		**Dataset-3 (Small ROI in Environment-II)**	
	Strict accuracy [%]	Relaxed accuracy [%]	Strict accuracy [%]	Relaxed accuracy [%]	Strict accuracy [%]	Relaxed accuracy [%]
PM-based [29]	69.90	92.67	58.74	90.38	66.79	91.11
CNN-based [30]	20.83	45.55	9.72	31.96	24.67	43.84
CT-based [31]	80.94	**99.59**	69.58	95.29	80.78	95.65
Proposed	**92.83**	99.56	**81.85**	**96.70**	**81.63**	**99.38**

Notably, the overall outcomes indicate that the proposed method provides more accurate predictions than the others, especially in the wide ROIs. In particular, it is evident from the outcomes in Dataset-2 and Dataset-3 (i.e., the outcomes with respect to different ROI sizes in the same environment) that the existing approaches undergo severe performance degradation as the experimental area widens, implying that the people counting systems based on the conventional features are likely to be applicable only to small areas. This degradation is attributable to the fact that the previous approaches were designed mainly for dense distribution of individuals. Therefore, the resulting systems became significantly dependent on energy-based features, whose separability tends to be saturated when the individuals are scattered over a wide area. In contrast, the proposed people counting system devises new features based on MD-CLEAN as well as energy-based features, to consider both scattered and dense distributions of individuals. Furthermore, as verified above, the MT strategy enables the proposed features to remain multi-sensitive, so that the signal fluctuations originating from diverse spatial distributions of individuals can be efficiently addressed. These novel strategies yielded a substantially enhanced performance when the ROI was expanded. Therefore, it can be concluded that the proposed method was capable of robust people counting even in a wider ROI and, hence, extended the detection coverage of the system significantly.

Meanwhile, considering that all people counting techniques require a training phase in advance, we performed additional comparisons in terms of the computational burden for training. Table V summarizes the computation time for training and feature (or template) dimension of each people counting method. Here, it should be noted that the training time was computed by averaging the time taken for a total of 150 independent training processes (50 times each for Dataset-1, 2, and 3) for statistical reliability. From the outcomes, it can be observed that the PM-based method takes the least time for training as it only requires the process of estimating probability density functions, without classifier learning based on feature vectors. In contrast, the CNN-based method must learn high-dimensional raw data directly (i.e., a dimension of $50 \times 1535 = 76{,}750$ per frame), leading to a heavy computational burden for training despite the use of GPU parallel computing. Our method demands a much smaller amount of training time compared with the CNN-based approach, which would further facilitate the utilization of online learning in real applications.

TABLE V
PERFORMANCE COMPARISON OF THE THREE CONVENTIONAL METHODS AND THE PROPOSED MODEL IN TERMS OF TRAINING BURDEN

	Training time [s]	Feature dimension
PM-based [29]	**86.90**	-
CNN-based [30]	1816.51	$N_t \times N_p(50) \times N_r(1535)$
CT-based [31]	193.05	$N_t \times 300$
Proposed	124.87	$N_t \times N_v(15)$

4) Analysis of the Proposed Method Based on Confusion Matrix: To offer a detailed analysis, the outcomes of the proposed method are presented in the form of normalized confusion matrices, all of which are illustrated in Fig. 14 (for Dataset-1), Fig. 15 (for Dataset-2), and Fig. 16 (for Dataset-3). In each confusion matrix, the probabilities of correct counting are highlighted in blue, those of counting with errors of one individual are in yellow, and those of counting errors of two or more individuals are in red. For all the datasets, the proposed people counting method succeeded in predicting the exact number of individuals in most cases. Even in the cases where it was incorrect, the errors were almost within one or two counts. Moreover, our system shows perfect predictions for zero individual (i.e., when there is no one in the ROI) regardless of the environmental conditions, which implies that the system can also be utilized as a reliable system for detecting the presence of individuals in indoor environments.

Meanwhile, comparing the confusion matrices of Dataset-1 (Fig. 14) and Dataset-2 (Fig. 15) (i.e., the outcomes from different environments and identical ROI size), the proposed people counting system clearly provides more precise estimations in Environment-I than in Environment-II. Furthermore, the counting errors in Environment-I tend to be smaller than those in Environment-II. Considering that the reflected echoes from Environment-II were more likely to be unfavorably affected by the clutter and multipath, this result emphasizes the importance of extracting only human-specific reflections from the raw received signals, for designing a robust people counting system.

V. CONCLUSION AND FUTURE WORK

To ensure robust people counting in a wide ROI, a variety of spatial distributions of people must be considered because each individual is provided a larger space to move. However, conventional people counting approaches generally consider

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/JIOT.2020.3032710, IEEE Internet of Things Journal

Predicted number of people \ True number of people	0	1	2	3	4	5	6	7	8	9	10
0	100	0	0	0	0	0	0	0	0	0	0
1	0	99	0	0	0	0	0	0	0	0	0
2	0	1	100	0	0	0	0	0	0	0	0
3	0	0	0	100	1	0	0	0	0	0	0
4	0	0	0	0	94.7	4.7	0	0	0	0	0
5	0	0	0	0	4.3	86.7	1.7	1.7	0	0	0
6	0	0	0	0	0	6.3	82.3	9.7	0	0	0
7	0	0	0	0	0	2.3	15	85	4.7	0	0
8	0	0	0	0	0	0	1	3.7	86.7	7.7	0
9	0	0	0	0	0	0	0	0	8.7	92	5.3
10	0	0	0	0	0	0	0	0	0	0.3	94.7

Fig. 14. Confusion matrix of the proposed people counting system under Dataset-1 (wide ROI in Environment-I).

Predicted number of people \ True number of people	0	1	2	3	4	5	6	7	8	9	10
0	100	0	0	0	0	0	0	0	0	0	0
1	0	87.7	3.3	0.3	0	0	0	0	0	0	0
2	0	11.7	85.3	4	0	0	0	0	0	0	0
3	0	0.7	9.7	80.3	9	1	0	0	0	0	0
4	0	0	1.7	14.7	82.3	9.7	3	0	0.3	0	0
5	0	0	0	0.7	5.7	78.3	14.3	1.3	1	0.7	0
6	0	0	0	0	3	8.7	69.7	12	2.3	0	0
7	0	0	0	0	0	1.7	10	76.7	6	2.7	1.3
8	0	0	0	0	0	0.7	1.3	4.3	80	3.3	3
9	0	0	0	0	0	0	1.7	3	8	79	14.7
10	0	0	0	0	0	0	0	2.7	2.3	14.3	81

Fig. 15. Confusion matrix of the proposed people counting system under Dataset-2 (wide ROI in Environment-II).

Predicted number of people \ True number of people	0	1	2	3	4	5	6	7	8	9	10
0	100	0	0	0	0	0	0	0	0	0	0
1	0	84.3	1.3	0	0	0	0	0	0	0	0
2	0	15.7	91	3.3	0	0	0	0	0	0	0
3	0	0	7.7	95.3	0	0	0	0	0	0	0
4	0	0	0	1.3	82.3	1	0	0	0	0	0
5	0	0	0	0	17.7	77.3	6	0	0	0	0
6	0	0	0	0	0	21.7	80.3	9.7	3	0	0
7	0	0	0	0	0	0	13.7	57.7	9.3	0	0
8	0	0	0	0	0	0	0	30	70.3	16	1
9	0	0	0	0	0	0	0	2.7	17.3	78.7	18.3
10	0	0	0	0	0	0	0	0	0	5.3	80.7

Fig. 16. Confusion matrix of the proposed people counting system under Dataset-3 (small ROI in Environment-II).

only limited spatial distributions in dense environments, resulting in significant performance degradation when applied in a wide area. A novel approach based on a new feature extraction scheme is proposed in this study to address this problem and design a reliable people counting system that is effective even in an expanded ROI. Specifically, we developed MD-CLEAN-based features in the range domain, which are significantly favorable for scattered distributions of individuals as well as energy-based features in the frequency domain for dense distributions. In particular, a newly devised MT scheme was applied in forming both the range and frequency domain features, which is followed by PCA and normalization. Consequently, the final features can attain multi-sensitivity property, enabling the people counting classifier based on those features to cope with various and complex distributions of individuals in a wider space. Several experiments using real measured data revealed that the proposed method can automatically predict the number of individuals with higher reliability even in a wide area, compared with the existing methods. Although the proposed method achieves a satisfactory performance level, there still remains room for further improvement in terms of practicality. One limitation of this study is that the performance of the proposed people counting system has been verified only for a restricted number of individuals up to 10. To verify the performance for more number of individuals, future works will construct expanded datasets by collecting additional radar signals reflected from more number of people (e.g., 11, 12, and more). Also, super-

resolution techniques and simulation-based training scheme will be integrated to facilitate robust people counting on such crowded environments. Meanwhile, another limitation is the hardware complexity caused by high sampling frequency, so additional research on sparse signal recovery from sub-Nyquist sampled IR-UWB radar data will be considered to further relieve the system burden.

REFERENCES

[1] P. K. D. Pramanik, B. K. Upadhyaya, S. Pal, and T. Pal, "Internet of things, smart sensors, and pervasive systems: Enabling connected and pervasive healthcare," in *Healthcare Data Analytics and Management*, Elsevier, 2019, pp. 1–58.

[2] H. Abdelnasser, K. Harras, and M. Youssef, "UbiBreathe: A ubiquitous noninvasive WiFi-based breathing estimator," in *Proc. MobiHoc*, Hangzhou, China, 2015, pp. 277–286.

[3] C. Li, V. M. Lubecke, O. Boric-Lubecke, and J. Lin, "A Review on recent advances in Doppler radar sensors for noncontact healthcare monitoring," *IEEE Trans. Microw. Theory Tech.*, vol. 61, no. 5, pp. 2046–2060, May 2013.

[4] C. Li *et al.*, "A review on recent progress of portable short-range noncontact microwave radar systems," *IEEE Trans. Microw. Theory Tech.*, vol. 65, no. 5, pp. 1692–1706, May 2017.

[5] D. Kasampalis, T. Alexandridis, C. Deva, A. Challinor, D. Moshou, and G. Zalidis, "Contribution of remote sensing on crop models: A review," *J. Imaging*, vol. 4, no. 4, p. 52, Mar. 2018.

[6] J. Geng, X. Ma, X. Zhou, and H. Wang, "Saliency-guided deep neural networks for SAR image change detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 10, pp. 7365–7377, Oct. 2019.

[7] J. Guerrero-Ibáñez, S. Zeadally, and J. Contreras-Castillo, "Sensor technologies for intelligent transportation systems," *Sensors*, vol. 18, no. 4, pp. 1-24, Apr. 2018.

[8] J. Li, Z. Zeng, J. Sun, and F. Liu, "Through-wall detection of human being's movement by UWB radar," *IEEE Geosci. Remote Sens. Lett.*, vol. 9, no. 6, pp. 1079–1083, Nov. 2012.

[9] Y. W. Kim, S. J. Ha, and J. H. Kwon, "Human detection using doppler radar based on physical characteristics of targets," *IEEE Geosci. Remote Sens. Lett.*, vol. 12, no. 2, pp. 289–293, Feb. 2015.

[10] T. A. Nguyen and M. Aiello, "Energy intelligent buildings based on user activity: A survey," *Energy Build.*, vol. 56, pp. 244–257, Jan. 2013.

[11] S. Bartoletti, A. Conti, and M. Z. Win, "Device-free counting via wideband signals," *IEEE J. Sel. Areas Commun.*, vol. 35, no. 5, pp. 1163–1174, 2017.

[12] B. Mrazovac, M. Bjelica, D. Kukolj, B. Todorovic, and D. Samardzija, "A human detection method for residential smart energy systems based on Zigbee RSSI changes," *IEEE Trans. Consum. Electron.*, vol. 58, no. 3, pp. 819–824, Aug. 2012.

[13] J. Weppner and P. Lukowicz, "Bluetooth based collaborative crowd density estimation with mobile phones," in *2013 IEEE Int. Conf. Pervas. Comput. Commun. (PerCom)*, San Diego, CA, USA, 2013, pp. 193–200.

[14] H. Li, E. C. L. Chan, X. Guo, J. Xiao, K. Wu, and L. M. Ni, "Wi-Counter: Smartphone-based people counter using crowdsourced Wi-Fi signal data," *IEEE Trans. Human-Machine Syst.*, vol. 45, no. 4, pp. 442–452, Aug. 2015.

[15] 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, Jul. 2015.

[16] E. Vattapparamban, B. S. Ciftler, I. Guvenc, K. Akkaya, and A. Kadri, "Indoor occupancy tracking in smart buildings using passive sniffing of probe requests," in *2016 IEEE Int. Conf. Communic. Workshops (ICC)*, Kuala Lumpur, Malaysia, 2016, pp. 38–44.

[17] F. Guidi, N. Decarli, S. Bartoletti, A. Conti, and D. Dardari, "Detection of multiple tags based on impulsive backscattered signals," *IEEE Trans. Commun.*, vol. 62, no. 11, pp. 3918–3930, Nov. 2014.

[18] N. Ahmed, A. Ghose, A. K. Agrawal, C. Bhaumik, V. Chandel, and A. Kumar, "SmartEvacTrak: A people counting and coarse-level localization solution for efficient evacuation of large buildings," in *2015 IEEE Int. Conf. Pervas. Comput. Commun. Workshops (PerCom Workshops)*, St. Louis, MO, USA, 2015, pp. 372–377.

[19] J. Garcia, A. Gardel, I. Bravo, J. L. Lazaro, M. Martinez, and D. Rodriguez, "Directional people counter based on head tracking," *IEEE Trans. Ind. Electron.*, vol. 60, no. 9, pp. 3991–4000, 2013.

[20] C. Wang, H. Zhang, L. Yang, S. Liu, and X. Cao, "Deep people counting in extremely dense crowds," in *Proc. 23rd ACM Int. Conf. Multim. - MM '15*, Brisbane, Australia, 2015, pp. 1299–1302.

[21] J. Cao, L. Sun, M. G. Odoom, F. Luan, and X. Song, "Counting people by using a single camera without calibration," in *2016 Chinese Control and Decision Conference (CCDC)*, Yinchuan, China, 2016, no. 1, pp. 2048–2051.

[22] S. I. Cho and S. J. Kang, "Real-time people counting system for customer movement analysis," *IEEE Access*, vol. 6, pp. 55264–55272, 2018.

[23] S. Sun, N. Akhtar, H. Song, C. Zhang, J. Li, and A. Mian, "Benchmark data and method for real-time people counting in cluttered scenes using depth sensors," *IEEE Trans. Intell. Transp. Syst.*, vol. 20, no. 10, pp. 3599–3612, Oct. 2019.

[24] X. Quan, J. W. Choi, and S. H. Cho, "Direction recognition of moving targets using an IR-UWB radar system," in *2014 4th IEEE Int. Conf. Netw. Infrastruct. Digit. Content*, Beijing, China, 2014, pp. 485–489.

[25] J. W. Choi, X. Quan, and S. H. Cho, "Bi-directional passing people counting system based on IR-UWB radar sensors," *IEEE Internet Things J.*, vol. 5, no. 2, pp. 512–522, Apr. 2018.

[26] J. W. Choi, J. H. Kim, and S. H. Cho, "A counting algorithm for multiple objects using an IR-UWB radar system," in *2012 3rd IEEE Int. Conf. Netw. Infrastruct. Digit. Content*, Beijing, China, 2012, no. l, pp. 591–595.

[27] J. W. Choi and S. H. Cho, "A new multi-human detection algorithm using an IR-UWB radar system," in *3rd Int. Conf. Innov. Comput. Technol. (INTECH 2013)*, London, UK, 2013, no. 1, pp. 467–472.

[28] J. W. Choi, S. S. Nam, and S. H. Cho, "Multi-human detection algorithm based on an impulse radio ultra-wideband radar system," *IEEE Access*, vol. 4, pp. 10300–10309, 2016.

[29] J. W. Choi, D. H. Yim, and S. H. Cho, "People counting based on an IR-UWB radar sensor," *IEEE Sens. J.*, vol. 17, no. 17, pp. 5717–5727, Sep. 2017.

[30] X. Yang, W. Yin, and L. Zhang, "People counting based on CNN using IR-UWB radar," in *2017 IEEE/CIC Int. Conf. Commun. China (ICCC)*, Qingdao, China, 2017, pp. 1–5.

[31] X. Yang, W. Yin, L. Li, and L. Zhang, "Dense people counting using IR-UWB radar with a hybrid feature extraction method," *IEEE Geosci. Remote Sens. Lett.*, vol. 16, no. 1, pp. 30–34, Jan. 2019.

[32] V. Yajnanarayana, S. Dwivedi, and P. Handel, "IR-UWB detection and fusion strategies using multiple detector types," in *2016 IEEE Wirel. Commun. Netw. Conf.*, Doha, Qatar, 2016, pp. 1–6.

[33] B. Vandersmissen *et al.*, "Indoor person identification using a low-power FMCW radar," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 7, pp. 3941–3952, Jul. 2018.

[34] V. H. Nguyen and J. Y. Pyun, "Location detection and tracking of moving targets by a 2D IR-UWB radar system," *Sensors*, vol. 15, no. 3, pp. 6740–6762, Mar. 2015.

[35] M. Cristani, M. Farenzena, D. Bloisi, and V. Murino, "Background subtraction for automated multisensor surveillance: A comprehensive review," *EURASIP J. Adv. Signal Process.*, vol. 2010, no. 1, pp. 1-24, Dec. 2010.

[36] A. Nezirovic, A. G. Yarovoy, and L. P. Ligthart, "Signal processing for improved detection of trapped victims using UWB radar," *IEEE Trans. Geosci. Remote Sens.*, vol. 48, no. 4, pp. 2005–2014, Apr. 2010.

[37] B. Lee, S. Lee, Y. J. Yoon, K. M. Park, and S. C. Kim, "Adaptive clutter suppression algorithm for human detection using IR-UWB radar," in *2017 IEEE Sensors*, Glasgow, UK, 2017, pp. 1–3.

[38] S. H. Chang, N. Mitsumoto, and J. W. Burdick, "An algorithm for UWB radar-based human detection," in *2009 IEEE Radar Conf.*, Pasadena, CA, USA, 2009, pp. 1–6.

[39] A. Lazaro, D. Girbau, and R. Villarino, "Analysis of vital signs monitoring using an IR-UWB radar," *Prog. Electromagn. Res.*, vol. 100, pp. 265–284, 2010.

[40] Novelda, Norway. *X4 – Datasheet.* (2020). [Online]. Available: https://novelda.com/fo/themes/default/img/contents/x4_datasheet_revF.pdf

[41] Suiyan Geng, S. Ranvier, Xiongwen Zhao, J. Kivinen and P. Vainikainen, "Multipath propagation characterization of ultra-wideband indoor radio channels," in *2005 IEEE International Conference on Ultra-Wideband*, Zurich, Switzerland, 2005, pp. 11–15.

[42] A. V. Oppenheim, R. W. Schafer, and J. R. Buck, *Discrete Time Signal Processing*, Pearson Education India, 1999.

[43] P. Stoica and R. L. Moses, *Spectral Analysis of Signals*, Upper Saddle River, NJ, USA: Prentice-Hall, 2005.

[44] T. Loutas, N. Eleftheroglou, G. Georgoulas, P. Loukopoulos, D. Mba, and I. Bennett, "Valve Failure Prognostics in Reciprocating Compressors Utilizing Temperature Measurements, PCA-Based Data Fusion, and Probabilistic Algorithms," *IEEE Trans. Ind. Electron.*, vol. 67, no. 6, pp. 5022–5029, 2020.