



# Human body recognition based on the sparse point cloud data from MIMO millimeter-wave radar for smart home

Xiaohua Zhou<sup>1</sup> · Xinkai Meng<sup>1</sup> · Jianbin Zheng<sup>1</sup> · Gengfa Fang<sup>2</sup> · Tongjian Guo<sup>3</sup>

Received: 19 August 2021 / Revised: 11 December 2021 / Accepted: 23 April 2023 /  
Published online: 12 August 2023

© The Author(s), under exclusive licence to Springer Science+Business Media, LLC, part of Springer Nature 2023

## Abstract

Human body recognition is widely used in smart home. The current mainstream perception modalities, i.e., camera and wearable device, are vulnerable under challenging lighting conditions and poor convenience. On the other hand, Multi-human body recognition remains as one of the most challenging tasks in a dynamic and complex environment. In this work, we introduce the low-cost multiple-input-multiple-output (MIMO) millimeter-wave radar without exposing user's private information for human body recognition in smart home. We propose a human body recognition scheme with the clustering based on the human body tracking using the sparse point cloud data of MIMO millimeter-wave radar. Firstly, the possible position of human body is predicted based on Kalman filter. Then, the point cloud data is clustered based on the human body shape in the prediction range of the human position. Finally, label tags are used to mark the human body targets detected by each frame of the radar. We apply human body recognition to validate the effectiveness of the proposed scheme. It can achieve single-person and double-person recognition using the sparse point cloud data of MIMO millimeter-wave radar. The results show that our proposed scheme reduces the error probability by 23.4% for the single-person recognition and by 31.1% for the double-person recognition. Extensive evaluations on the application of human activity recognition well demonstrate the practicability of the proposed scheme.

**Keywords** Human body recognition · MIMO millimeter wave radar · Point cloud data · Clustering · Body track

---

✉ Tongjian Guo  
tjguo\_ciomp@hotmail.com

<sup>1</sup> College of Instrumentation & Electrical Engineering, Jilin University, Changchun, China

<sup>2</sup> Global Big Data Technologies Center, University of Technology Sydney, Sydney, Australia

<sup>3</sup> Changchun Institute of Optics, Fine Mechanics and Physics, Chinese Academy of Science, Changchun, China

## 1 Introduction

Human body recognition (HBR) becomes increasingly attractive in many applications, such as smart home, elderly care and health management [14, 15, 24]. In particular, a large number of elderly people live independently at home most of the time, and it is extremely necessary to monitor their activities, especially falls. In recent years, although great progress has been made in the development of HBR, there are still substantial challenges. For instance, privacy exposure limits the application of HBR in the real world and the complex environment with occlusion and lighting conditions might negatively impact on the recognition performance. In fact, many people reluctant to expose too much private information, but the abnormal activities have to be monitored in real-world situations. For example, people might get falling or slipping while doing house-work. Particularly, the abnormal activities might occur to the elderly at their homes. In this situation, the proactive or situated services based on HBR at home would alleviate the dangerous situation.

There are many competing devices for human body recognition, which may be categorized into two main branches: the wearable and the unwearable [3]. The wearable devices, usually the accelerometers, need be worn on the body all the time to sense the human action [9]. It is not convenient for some people, such as the elderly, to use the wearable devices in practice. On the other hand, the unwearable devices, usually the camera and the radar, are typically installed in the fixed location to detect human activities [10]. Although the camera can directly image the persons and recognize their activities from the images, it is vulnerable to the lighting conditions and often brings significant privacy concerns. In contrast, radar actively transmits electromagnetic waves and processes the reflections to sense the objects' locations and moving speed [28]. It works even under poor lighting, and it does not expose user's private information. Therefore, the radar carries great potential to be the prevailing HBR technologies in smart home.

Many HBR studies have used range, Doppler and micro-Doppler signatures [5]. For example, the pre-defined or self-learned features were extracted and classified to recognize human body. Some feature extraction methods, such as Mel-frequency Cepstral Coefficients (MFCCs), pseudo-Zernike moments, shape spectrum features and L1-norm linear discriminant analysis, were explored to differentiate the radar echo behaviors [4, 16, 17, 20]. As deep learning technique shows strong advantages in high-level deep feature extraction, many different deep-learning models had been proposed to act as feature extractors and classifiers for identifying human action from radar signatures [2]. However, these methods often face challenges such as large amount of calculation, complex model design and high hardware requirements.

Although the related work have shown many radar-based HBR methods, it is still a huge challenge when apply to the HBR in smart home. The Doppler information from human motion at the same distance in different directions will alias each other when there is only range and Doppler information, thereby increasing the probability of the HBR errors. In contrast, multiple-input-multiple-output (MIMO) millimeter-wave radar can provide not only distance and Doppler information, but also angle information [1]. Therefore, MIMO millimeter-wave radar can adapt to the scenarios of multi-person presence. Admittedly, compared with expensive lidar that can provide the dense point cloud data, the point cloud data from MIMO millimeter-wave radar is sparse, but the sparse point cloud data is enough for the HBR at indoor and home deployment. In addition, the sparse point cloud data has good real-time performance with less computation. Furthermore, the MIMO millimeter-wave radar is suitable with reasonable price for application of the health management and

elderly care in smart home. Considering these advantage, this paper explores HBR methods based on sparse point cloud data from MIMO millimeter-wave radar in smart home. Our core contributions are three folds:

- (1) We propose a HBR scheme using the point cloud data from MIMO millimeter-wave radar, which is especially suitable for the application of the monitoring equipment in smart home.
- (2) We develop the clustering method of the point cloud data based on the human body tracking, which can achieve significantly improved accuracy of the HBR in a dynamic and complex environment with low computational complexity.
- (3) We test HBR based on the sparse point cloud data from MIMO millimeter-wave radar. Extensive evaluations well demonstrate the superiority of the proposed HBR scheme for the single-person and double-person recognition.

Our work is distinctive from other works in two main points: The MIMO millimeter-wave radar is introduced for human body recognition in smart home without exposing user's private information. Compared with the vision-based detection approaches, it is not easy to be affected by lighting conditions and other responsible environments. Compared with expensive Lidar, it has the advantages of reasonable price for application of the elderly care in smart home. In addition, our proposed framework with the clustering based on the human body tracking achieves single-person and double-person recognition, which we anticipate that this would be effective in Multi-human body recognition.

The remainder of the paper is organized as follows. Section 2 describes the generation of the point cloud data from MIMO millimeter-wave radar. The proposed the HBR scheme is presented in Section 3, while the experimental results are shown in Section 4. The paper ends up with the conclusion and future research directions in Section 5.

## 2 Point cloud data generation

The radar is a multiple-input-multiple-output (MIMO) structure with three transmitters antennas (TX) and four receivers antennas (RX) collocated together [11]. In order to facilitate mathematical analysis, it is equivalent to a single-input-multiple-output (SIMO) virtual antennas array, as shown in Fig. 1. Frequency-Modulated Continuous-Wave (FMCW) millimeter-wave radar is used to detect the targets (human body) in our work. The radar continuously transmits the chirp signals at each frame to the specified airspace [6, 26, 27]. The reflection signals are obtained from different targets or different parts of the target, and they are processed into the differential frequency signals at the receiving end [19]. After 2D-FFT and 1D-FFT, the spatial coordinate position and the radial velocity corresponding to the different reflected signals can be calculated, which constitute the point cloud data of the target. Process of the point cloud data generation is shown in Fig. 2.

Define a single chirp signal emitted by radar as [8]:

$$s_t = e^{j\left(\omega_0 + \frac{\mu}{2}t\right)t} \quad (1)$$

Where,  $\omega_0$  is the starting frequency,  $\mu$  is the sweep rate. On the propagation path of the radar signal, the virtual antenna receives  $K$  plane wave signals reflected from the far-field position, which is defined as [21]:

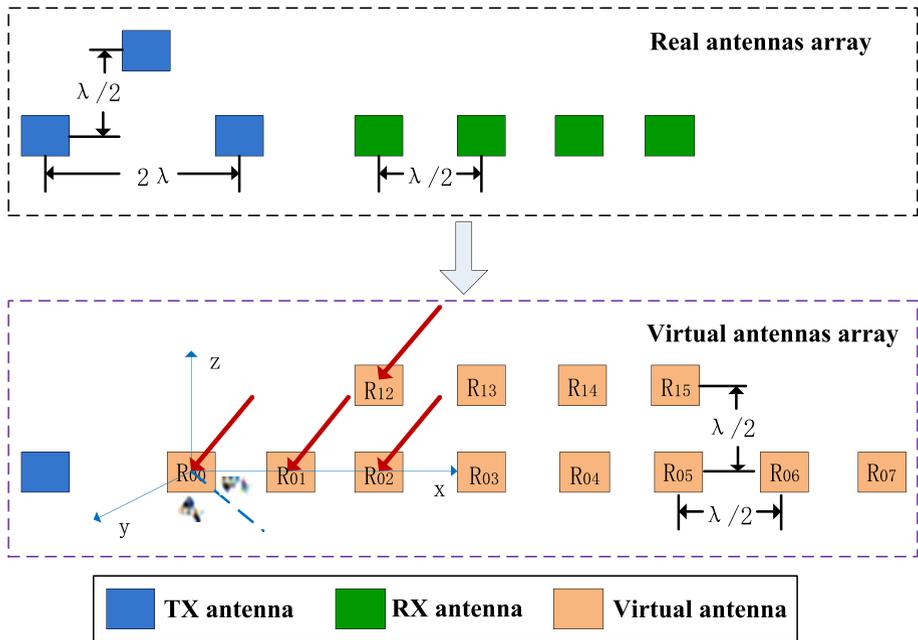


Fig. 1 Diagrammatic sketch of transformation from MIMO to SIMO

$$s_r(m, n) = \sum_{k=1}^K a_{rk} e^{j r_{mn} \Gamma_k} s_t(t - t_{hk}) \tag{2}$$

$$r_{mn} = \left[ \frac{n\lambda}{2}, 0, \frac{m\lambda}{2} \right] \tag{3}$$

$$\Gamma_k = \frac{2\pi}{\lambda} [\sin\theta_k \cos\varphi, \cos\theta_k \cos\varphi_k, \sin\varphi_k]^T \tag{4}$$

$$t_{hk} = \frac{2d_k}{c} + \frac{2v_k t}{c} = \tau_k + \eta_k t \tag{5}$$

Where,  $a_{rk}$  represents the attenuation coefficient of the signal on the path.  $r_{mn}$  is the coordinate position vector of the virtual antenna numbered  $R_{mn}$ .  $\lambda$  is the wavelength of the radar signal at the starting frequency.  $\Gamma_k$  is the wavenumber vector of the  $k$ -th plane wave.

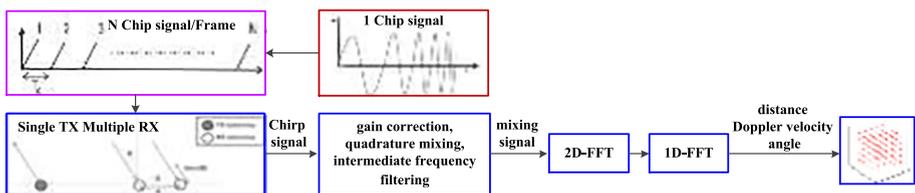


Fig. 2 Process of the point cloud data generation

$\theta_k$  and  $\varphi_k$  represent the azimuth and elevation angle of the arriving beam respectively.  $t_{hk}$  is the time delay of the  $k$ -th plane wave signal arriving at the  $R_{00}$  antenna.  $d_k$  and  $v_k$  are the starting distance and the radial velocity of the target reflecting the  $k$ -th plane wave respectively. At the receiving end, the receiving and transmitting signals are subjected to gain correction, quadrature mixing, and intermediate frequency filtering. The difference frequency signal is obtained [13]:

$$s_d(m, n) = \sum_{k=1}^K a_k e^{j r_{mn} \Gamma_k} e^{j(\omega_0 \tau_k + \omega_0 \eta_k t + \mu \tau_k t)} \tag{6}$$

Generally, the radar detects targets frame-by-frame. The radar transmits  $P$  chirp signals each frame at a period  $T$ . The starting time of each chirp signal is defined as  $t_p$  ( $p=0, 1, \dots, P-1$ ). The target radial velocity is approximately constant during the detection time of the frame. Suppose  $\tau_{kp}$  is the signal delay of the target reflecting the  $k$ -th plane wave at the starting point of the  $p$ -th transmitting signal, then [25]:

$$\tau_{kp} = \tau_{k0} + \eta_k t_p \tag{7}$$

Substituting formula (7) into formula (6), the difference frequency signal generated by the  $p$ -th transmitting signal at the receiving end can be obtained, as shown in Eq. (8). Using  $(t, t_p)$  as a binary variable to perform 2D-FFT on  $P$  difference frequency signals, as shown in formula (9)(10) (11).

$$s_d(m, n, t, t_p) = \sum_{k=1}^K a_k e^{j r_{mn} \Gamma_k} e^{j \omega_0 \tau_{k0}} e^{j \omega_0 \eta_k t_p} e^{j(\omega_0 \eta_k + \mu \tau_{k0} + \mu \eta_k t_p) t} \approx \sum_{k=1}^K a_k e^{j r_{mn} \Gamma_k} e^{j \omega_0 \tau_{k0}} e^{j(\omega_0 \eta_k + \frac{\mu \eta_k T}{2}) t_p} e^{j(\omega_0 \eta_k + \mu \tau_{k0} + \frac{\mu \eta_k P T}{2}) t} \tag{8}$$

$$F[s_d] = S(m, n, u, v) = \sum_{k=1}^K A_k e^{j r_{mn} \Gamma_k} e^{j \omega_0 \tau_{k0}} \delta(u - \omega_{dk}) \delta(v - \omega_{Dk}) \tag{9}$$

$$\omega_{dk} = \omega_0 \eta_k + \mu \tau_{k0} + \frac{\mu \eta_k P T}{2} \tag{10}$$

$$\omega_{Dk} = \omega_0 \eta_k + \frac{\mu \eta_k T}{2} \tag{11}$$

Where,  $\omega_{dk}$ ,  $\omega_{Dk}$  are the distance frequency and the Doppler frequency of the target reflecting the  $k$ -th plane wave respectively. The above formulas show that the local extreme point coordinates of the amplitude correspond to the distance frequency and the Doppler frequency of the target after performing 2D-FFT on one frame of the difference frequency signals received by each antenna. Thus, the distance and radial velocity of the target can be calculated. In order to further calculate the spatial position of the target, 2D-FFT values corresponding to the same distance and the same radial velocity of each antenna are extracted and combined into the new data as:

$$Q(m, n) = \sum_{l=1}^L A_l e^{j r_{mn} \Gamma_l} e^{j \omega_0 \tau_{l0}} \tag{12}$$

Among them, the data with different numbers  $m$  and the same  $n$  is used to calculate the elevation angle. The data with the same number  $m$  and different  $n$  is used to calculate the azimuth. According to the virtual antenna position in Fig. 1, there are two valid data for solving the elevation angle. At this time, it is impossible to distinguish the difference in the

elevation direction of the target, that is, the target with the same distance and radial velocity has the same elevation angle and corresponding space. The  $z$  coordinate value is:

$$z = d \sin \varphi = d \frac{\angle Q(1, n) - \angle Q(0, n)}{\pi} \quad (13)$$

Where,  $d$  is the distance to the target. The symbol  $\angle$  represents the phase angle of the data. There are eight valid data to calculate the azimuth. At this time, it is possible to further distinguish the difference in azimuth between targets with the same distance and radial velocity. For  $Q(0, n)$ , the new data is:

$$Q(0, n) = \sum_{l=1}^L A_l e^{j\omega_0 \tau_{l0}} e^{jn\pi \sin \theta_l \cos \varphi_l} \quad (14)$$

The 1D-FFT is performed with  $n$  as a parameter, there is:

$$F[Q(0, n)] = Q_F(w) = \sum_{l=1}^L A_l e^{j\omega_0 \tau_{l0}} \delta(w - \pi \sin \theta_l \cos \varphi_l) \quad (15)$$

The local extreme point coordinate of the amplitude after 1D-FFT is:

$$\omega_{al} = \pi \sin \theta_l \cos \varphi_l \quad (16)$$

The spatial coordinate values of the target are obtained as follows:

$$x_1 = d \frac{\omega_{al}}{\pi} \quad (17)$$

$$y_1 = \sqrt{d^2 - x^2 - z^2} \quad (18)$$

In this way, the point cloud data including the distance, Doppler velocity and angle information can be obtained.

### 3 Human body recognition

The proposed HBR scheme consists of four components: Coordinate transformation, Projection of human body position, Point cloud data clustering and Label the human body. The architecture of the scheme is illustrated in Fig. 3. The point cloud data are converted through coordinate transformation. The processed data are then passed through projection of human body position to describe the motion process. Generally, the point cloud data obtained by millimeter-wave radar are reflections from different targets in space. In order to use accurately the point cloud data for the HBR, the point cloud data reflected by different targets need to be clustered separately. After clustering the point cloud data, the tags are used to mark the human body detected by radar in each frame. The final result of HBR is decided by the tags.

Specifically, because the point cloud data collected by MIMO millimeter-wave radar is based on the radar antenna coordinate system, in order to eliminate the influence of the installation position of the radar antenna, firstly, the point cloud data from the radar is converted to the ground coordinate system. Then, the possible position of the human body is predicted based on the Kalman filter method. Furthermore, the point cloud data

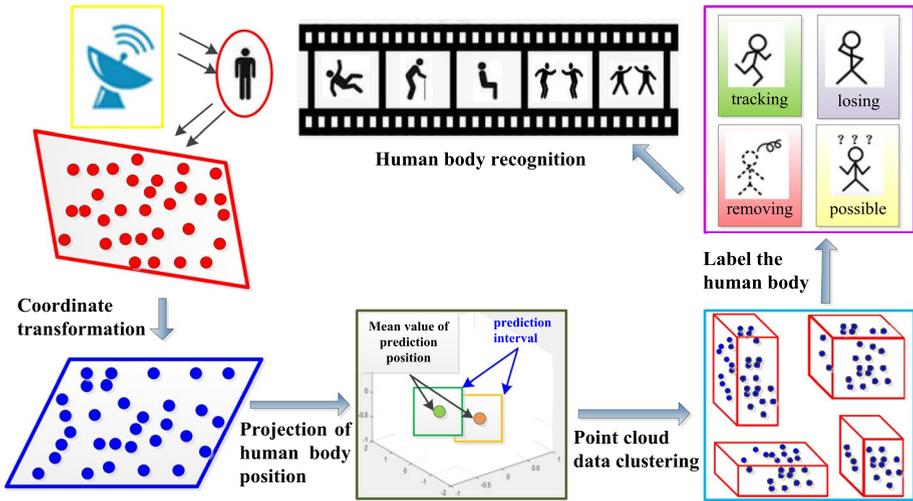


Fig. 3 Process of the HBR

conforming to human body characteristics is clustered based on human body shape in the prediction interval of human body position. Finally, tags are used to mark the human targets detected by the radar in each frame with tracking, losing, removing and possible. For the target with ‘tracking’ tag, the target state is updated. The clustering result can be used for further human action recognition and other applications [18, 22].

### 3.1 Coordinate transformation

In order to eliminate the influence of the radar antenna installation position, it is necessary to convert the point cloud data from the radar to ground coordinate system. The formula of the coordinate transformation is [7]:

$$X_g = RX_s + T \tag{19}$$

$$v_g = \frac{X_g \cdot X_s}{\|X_g\| \|X_s\|} v_s \tag{20}$$

Where,  $X_g$  and  $X_s$  are the space position in the ground coordinate system and the antenna coordinate system respectively.  $R$  is the coordinate rotation matrix.  $T$  is the coordinate translation matrix.  $v_g$  and  $v_s$  represent the radial velocity in the ground coordinate system and the antenna coordinates respectively.

$$X \triangleq [x \ y \ z]^T \tag{21}$$

$$R = \begin{bmatrix} \cos\gamma & \sin\gamma & 0 \\ -\sin\gamma & \cos\gamma & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \cos\beta & 0 & -\sin\beta \\ 0 & 1 & 0 \\ \sin\beta & 0 & \cos\beta \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos\alpha & \sin\alpha \\ 0 & -\sin\alpha & \cos\alpha \end{bmatrix} \tag{22}$$

$$T = [x_0 \ y_0 \ z_0]^T \quad (23)$$

Where,  $\alpha$ ,  $\beta$  and  $\gamma$  respectively represent the rotation angles required for the antenna coordinate system to be parallel to the coordinate axis of the ground coordinate system after clockwise rotation.  $(x_0, y_0, z_0)$  is the spatial position of the antenna in the ground coordinate system.

### 3.2 Projection of human body position

Within the range of radar measurement, although human body motion is a complex process, it can still be equivalent to a particle. In this case, the Kalman model can be used to describe the motion process [23]. The motion model is as follows:

$$x_{k+1} = Ax_k + Bw \quad (24)$$

$$y_k = Cx_k + v \quad (25)$$

Where,  $x_k$  is the state vector at  $k$  time,  $y_k$  is the measured value at  $k$  time, and it is the position in this paper.  $A$  and  $C$  represent the state transition matrix and the observation matrix respectively.  $w$  and  $v$  describe the process noise and observation noise respectively which is independent of each other.  $\sigma_w^2$  and  $\sigma_v^2$  are the variances respectively. Suppose  $\bar{x}_0$  and  $P_0$  are the mean value and variance at the initial state. Using Kalman recursive estimation method, the mean value and variance of the state vector  $x_k$  at  $k$  time are estimated to be  $\bar{x}_k$  and  $P_k$  respectively according to the observed value  $y_k$ . Then the mean value  $\tilde{y}$  and variance  $\tilde{P}_y$  of  $\hat{y}_{k+1|k}$  at  $k + 1$  time are predicted to be:

$$\tilde{y} = CAx_k \quad (26)$$

$$\tilde{P}_y = CAP_k A^T C^T + CBB^T C^T \sigma_w^2 + \sigma_v^2 \quad (27)$$

Obviously, most of the point cloud data reflected by the target at  $k + 1$  time is included within the range of  $\tilde{y} \pm r\sqrt{\tilde{P}_y}$ . There are many such intervals in the case of multiple human bodies. In this way, the clustering is to search point data in this interval according to certain criteria and associate with the corresponding human body. Here, it is considered that the human body only moves on the ground, and the uniform motion model is adopted.

After the clustering of the point cloud, the center of gravity for the point cloud is calculated, and it is used as the center of gravity for the human body to update the human motion state [12]. The state update process is:

$$\hat{x}_{k+1} = \hat{x}_{k+1|k} + G_{k+1}(y_{k+1} - \hat{y}_{k+1|k}) \quad (28)$$

$$G_{k+1} = P_{k+1|k} C^T (C P_{k+1|k} C^T + \sigma)^{-1} \quad (29)$$

Where,  $G_{k+1}$  is the gain matrix of the Kalman.

### 3.3 Point cloud data clustering

In order to cluster the point cloud data conforming to human characteristics, the clustering method of the point cloud data based on human volume is proposed. To be specific, a

cuboid with a certain volume is equivalent to a human body, and the cuboid is moved in the human body position prediction interval to search the most point cloud data contained in the cuboid. The location area with the highest density of the point cloud data under the fixed volume is regarded as the location area of human body. The center point of the location area is regarded as the location of human body, and it is regarded as the update value of human motion position prediction. Thus, the target clustering and location generation are completed. The steps of the cloud data clustering are shown in Table 1.

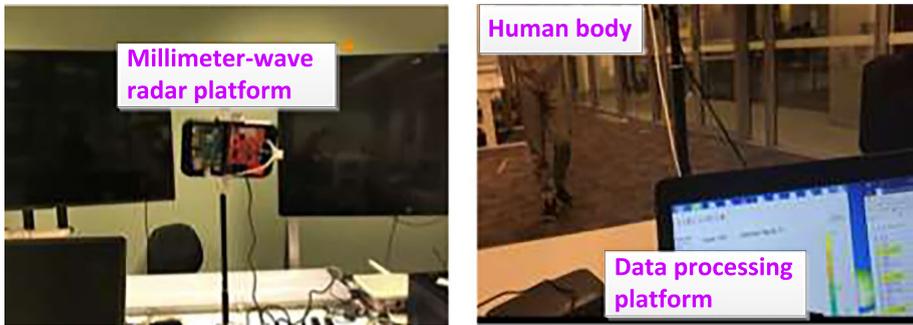
### 3.4 Label the human body

After clustering the point cloud data, four types of the tags are used to mark the human body detected by radar in each frame. They are tracking, losing, removing and possible. To be specific, if the current frame of the target is assigned a specified number of point cloud data, and the previous frame is marked as tracking or possible. The target is marked as ‘tracking’. If the current frame of the target is not assigned a specified number of point cloud data, and the previous frame is marked as tracking or losing. The target is marked as ‘losing’. For the ‘losing’ target, the position prediction of the target is maintained, but the target motion state is not updated. When previous continuous frames of the target are marked as losing, the target is marked as removing. Once the target is marked as removing, the target position prediction is no longer maintained. When the radar detects a new target, it is marked as possible.

In particular, for the possible target, it is necessary to establish the possible trajectory. The process is as follows:

**Table 1** Steps of the cloud data clustering

<b>Input:</b>	Dataset, Prediction information $P_i = \{(x_i, y_i), e_i\}$ , Minimum number of point cloud for the target $M$ , Human body shape $L = \{L_1 \square L_2 \square \dots \square L_k \dots \square L_n\}$ .
<b>Output:</b>	Serial number of the point cloud assigned human body $I_d$ .
<pre> begin   for <math>P_i = \{(x_i, y_i), e_i\}</math>     for <math>x = x_i - e_i &lt; x_i &lt; x_i + e_i</math>       for <math>y = y_i - e_i &lt; y_i &lt; y_i + e_i</math>         <math>N \leftarrow</math> Maximum number of point cloud included in the           human body shape <math>L</math>         <math>q = \{N, L, I_d, (x, y)\}</math>       end for     end for     Minimum distance between <math>(x, y)</math> and <math>(x_i, y_i)</math>     <math>P_i \leftarrow q</math>   end for   if <math>N &gt; M</math>     Return <math>I_d</math>   end if end                 </pre>	



**Fig. 4** Experimental setup

Step1: Get the possible target position. The radar detection interval is gridded. The point cloud data assigned to the target is deleted. The number of the point cloud data in each grid is calculated. If the grid with the largest number of the point cloud data exceeds the threshold, it is considered that there is a possible target in the grid. The center of the grid is regarded as the possible target position.

Step 2: Initialize the motion state of the possible target. The velocity value is obtained by the difference between the possible target position of the current frame and the possible target position of the front and rear frames. The velocity value is taken as the mean value of the initial state vector of the possible target motion, and the variance of the initial state vector is calculated according to the double grid error.

Step 3: Update the motion state of the possible target. After the Step 2, the subsequent new possible target is judged according to the above method described in Section 3.2 to determine whether it falls within the position prediction interval and to update the state. If the possible target of the continuous frames falls in the position prediction interval, it is marked as tracking. Otherwise, the initialization process of the possible target state is carried out again according to Step2.

## 4 Experiments

### 4.1 Datasets

Data are collected through a millimeter-wave radar platform with three-TX and four-RX antennas. The radar installation height is about 1.8 m. The pitch angle is about  $10^\circ$ . The active test area of  $4.5 \text{ m} \times 4.5 \text{ m}$  is marked at 1.1 m in front of the radar. At the same time, the video is used to mark the human body corresponding to the frame of the radar. The frame period of the radar signal is 100ms. 48 chirp signals are sent out in each frame period. The chirp start frequency is 77 GHz. The chirp sweep bandwidth is 3.2 GHz. The chirp sweep rate is 100 MHz/us. The experimental setup is shown in Fig. 4.

During the experiment, ten groups of data were collected, including six single-person groups and four double-person groups. The data collection duration of each group was about 10 min. There were five different activities involved in the groups, including standing, walking, falling, lying and rising. Different activities were completed randomly and

**Table 2** Data statistics of the single-person groups

Group	Human Activity	Times of the activity
Single-person	Standing	246
	Walking	120
	Falling	240
	Lying	240
	Rising	240

**Table 3** Data statistics of the double-person groups

	P1	Standing	Walking	Falling	Lying	Rising
P2						
Standing	41	30	30	30	30	30
Walking	15	20	15	15	15	15
Falling	30	30	40	30	30	30
Lying	30	30	30	40	30	30
Rising	30	30	30	30	30	40

collected by millimeter-wave radar. Meanwhile, the video was used to mark the human body corresponding to the frame data of the radar for the result verification of the HBR. The data set are shown in Tables 2 and 3.

## 4.2 Clustering test

In order to test the effect of the HBR, DBSCAN was used as a comparative analysis method with our clustering method based on human tracking model (CBBT). The single-person groups and double-person groups with different activities in data set of Tables 2 and 3 were selected to perform clustering test. Clustering parameters of different methods are shown in Table 4. DBSCAN algorithm needs to be given two parameters. One of the parameters is radius (EPS), which represents the range of circular neighborhood centered on a given point. Another parameter is the number of minimum points (MinPts). Within the radius

**Table 4** Clustering parameters of different methods

Method	Parameter	Value	Unit
DBSCAN	Neighbor radius(EPS)	0.4	m
	Minimum number of core point (MinPts)	5	-
CBBT	Number of activity types	5	-
	Number of clustering cuboid types	6	-
	Search step size	0.05	m
	Search radius	0.6	m
	Target label category	Tracking,Losing, Removing,possible.	-

**Fig. 5** Single-person clustering results of three-dimensional point cloud data (Red and green dots represent objects. Blue dots represent backgrounds. Pink dots represent possible objects. Black solid lines represent cubes used by CBBT clustering. Red dashed lines represent cube boundaries of possible objects. Light blue box points represent losing objects)

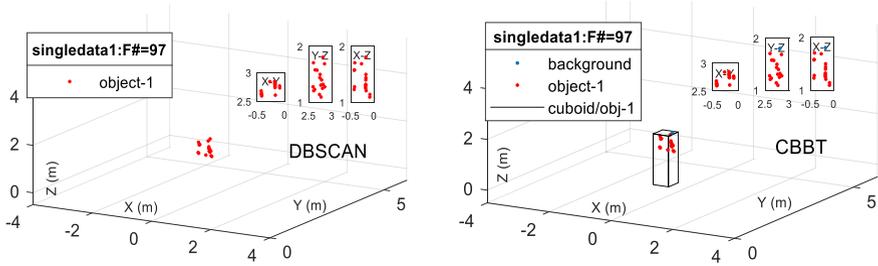
of a neighborhood, the neighborhood of MinPts number is considered as a cluster. If the distance between data points is less than or equal to the specified EPS, they will be of the same class.

#### 4.2.1 Single-person clustering

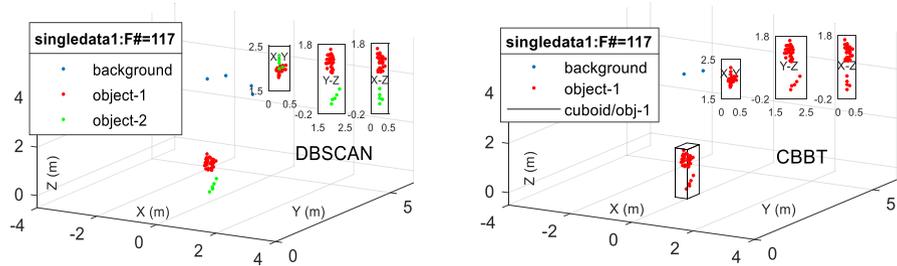
In the single-person groups, the point cloud data were used to cluster by DBSCAN and CBBT. Only one person with different activities was recorded in different frames of the millimeter-wave radar. The video was used to mark the human body responding to each frame data of the radar. Here, we present the clustering results corresponding to four different frames, as shown in Fig. 5. We can see that two methods in single-person clustering results at the 97th Frame are consistent. However, the clustering result based on DBSCAN is incorrect at 117th Frame and the 349th Frame, while the clustering based on CBBT can still keep correct results. Notably, one losing object has emerged at the 1664th Frame for CBBT because of changes of human action. This shows indirectly that CBBT can indicate changes of human action.

To analyze the performance of different methods, we provide statistical results of the HBR in the single-person groups. As can be seen from Fig. 6, the absolute error probability of CBBT on the clustering objects number (Object = 1) is lower than that of DBSCAN for the same point cloud data of the single-person groups. In addition, the joint probability distribution of the clustering objects is shown in Table 5. It shows that the error probability of DBSCAN with 30.5% is higher than that of CBBT with 7.1% in this experiment, which means our proposed method reduces the error probability by 23.4%. Furthermore, the consistency of the two clustering methods in the case of single-person is compared. We extract the point cloud data of single-person clustered by DBSCAN and CBBT and calculate the proportion of the same data in the total cluster points. The statistical result of the proportion of the same cluster points in the whole data is shown in Fig. 7. It can be seen that the same data accounts for 90% in the cluster points with the proportion of more than 80%, which means that the consistency of the cluster points is good.

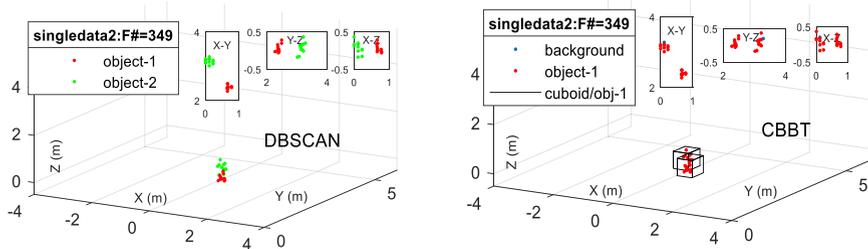
To further evaluate the performance of our proposed HBR scheme, we apply the clustering results to human activity recognition (HAR) in different frames of the millimeter-wave radar. The video was used to mark the human activity responding to each frame data of the radar. We processed the point cloud data with the above clustering results of one object for the HAR. The statistical characteristics of the point cloud data were extracted as the feature parameters of the behavior recognition. Gaussian kernel SVM was adopted as the classifier of the behavior recognition. The confusion matrix on feature dataset for different activities is shown in Fig. 8. The HAR result in different frames of the millimeter-wave radar is provided in Fig. 9. Obviously, different frames can be very good corresponding to human activity. For instance, human activity is standing at the 97th Frame, walking at the 117th Frame, rising at the 349th Frame, and falling at the 1664th Frame respectively, which is well corresponding to the clustering objects and the video record.



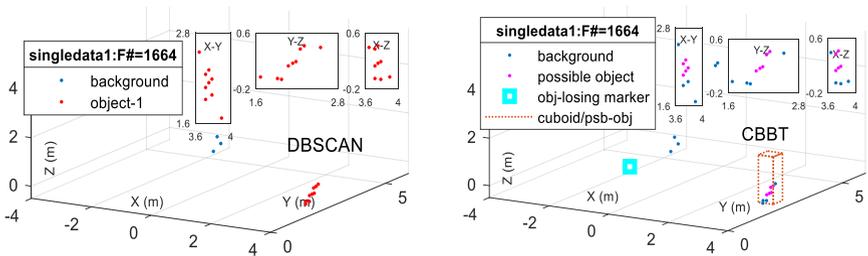
(a) Single-person clustering at 97th Frame (DBSCAN : 1 object. CBBT : 1 object)



(b) Single-person clustering at 117th Frame (DBSCAN : 2 objects. CBBT : 1 object)



(c) Single-person clustering at 349th Frame (DBSCAN : 2 objects. CBBT : 1 object)



(d) Single-person clustering at 1664th Frame (DBSCAN : 1 object. CBBT : 1 possible object, 1 losing object)

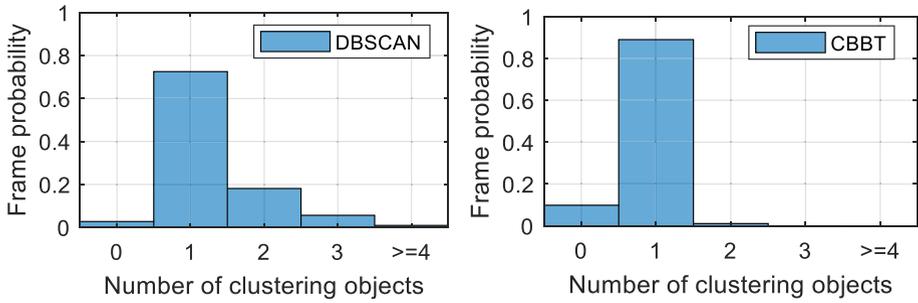


Fig. 6 Histogram of the number of objects for the single-person groups clustering

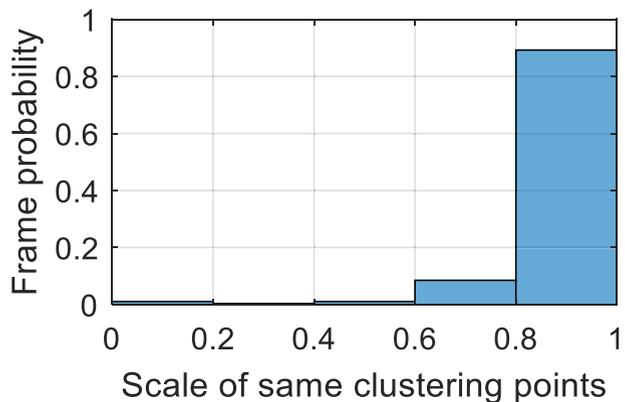
Table 5 Joint probability distribution of the clustering objects

Method/Objects		CBBT				
		0	1	2	3	≥4
DBSCAN	0	0.02	0.009	0	0	0
	1	0.06	0.66	0.002	0	0
	2	0.014	0.16	0.007	0	0
	3	0.002	0.05	0.006	0.	0
	≥4	0	0.01	0	0	0

### 4.2.2 Double-person clustering

In the double-person group, we also use DBSCAN and CBBT to cluster the point cloud data. Two persons with different activities were recorded in different frames of the millimeter-wave radar. Here, we present the clustering results corresponding to four different frames, as shown in Fig. 10. We can see that two methods in double-person clustering results at 833th Frame are consistent. However, the clustering result based on DBSCAN is incorrect at 1178th Frame, 1272th Frame and 1275th Frame, while the clustering based on CBBT can still keep correct results. Notably, one losing object at 1275th Frame has

Fig. 7 Proportional histogram of the same cluster points for DBSCAN and CBBT (Single-person groups clustering)



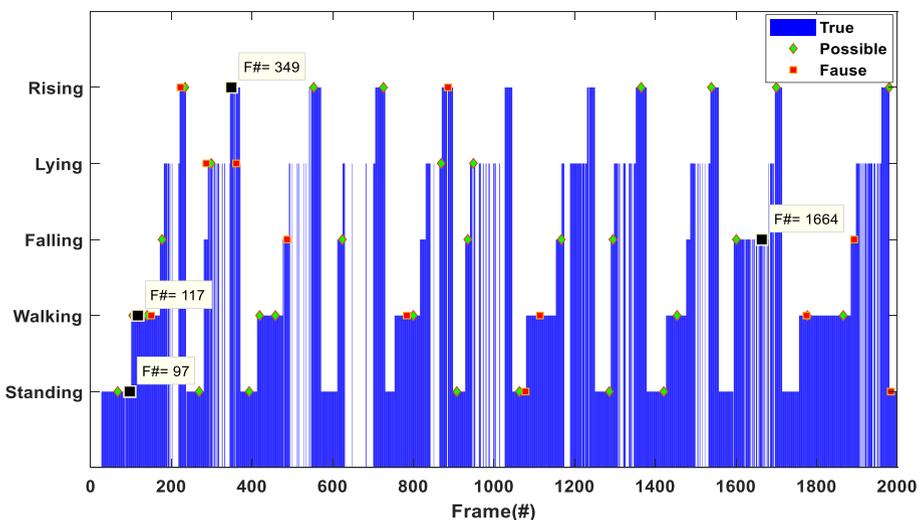
**Fig. 8** Confusion matrix on feature dataset using 5-fold cross-validation

Accuracy: 90.04%

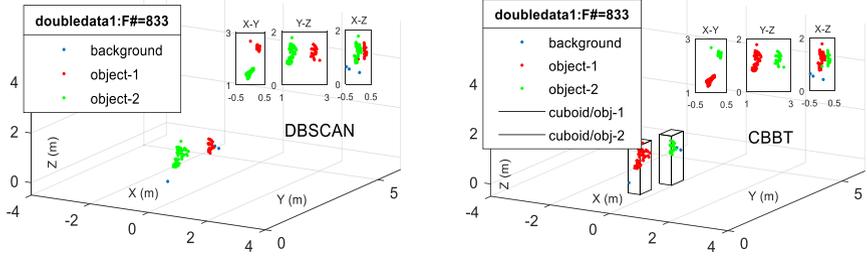
True Class	Predicted Class				
	standing	walking	falling	lying	rising
standing	92.5% 5590	2.7% 162	1.2% 71	0.3% 20	3.3% 199
walking	5.3% 250	92.4% 4327	1.5% 69	0.8% 36	0.0% 1
falling	1.9% 56	3.0% 89	88.5% 2620	6.3% 187	0.3% 9
lying	0.1% 4	0.1% 6	5.8% 232	88.6% 3561	5.4% 217
rising	5.8% 262	0.0% 0	0.1% 5	7.5% 337	86.6% 3897

emerged for CBBT because of changes of human action. This shows further that CBBT can indicate changes of human action.

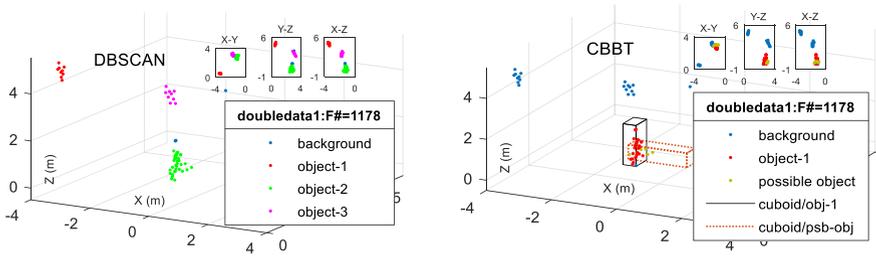
Similarly, we also provide statistical results of the HBR in the double-person groups. As can be seen from Fig. 11, the absolute error probability of CBBT on clustering objects number (Objects=2) is lower than that of DBSCAN for the same point cloud data of the double-person groups. In addition, the joint probability distribution of the clustering objects is shown in Table 6. It shows that the error probability of DBSCAN with 58.6% is higher than that of CBBT with 27.5% in this experiment, which means our proposed method reduces the error probability by 31.1%. Furthermore, the consistency of the two



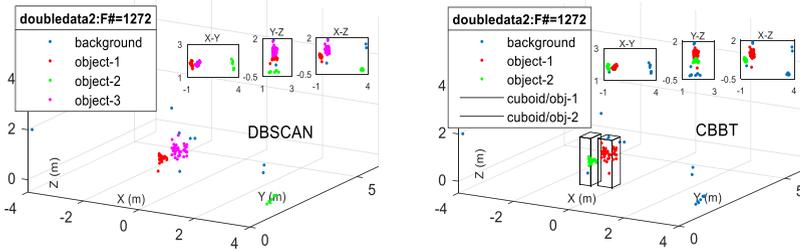
**Fig. 9** Different actions corresponding to different frames



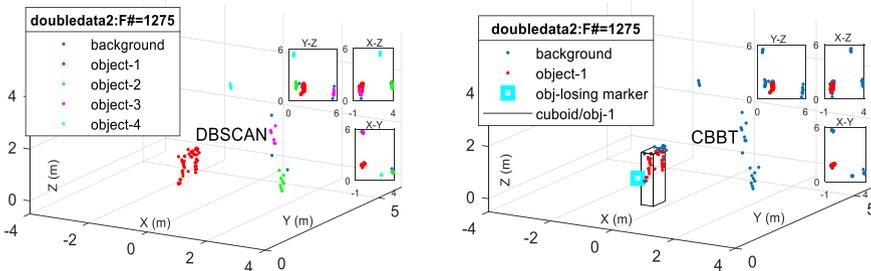
(a) Double-person clustering at 833th Frame (DBSCAN = 2 objects. CBBT = 2 objects)



(b) Double-person clustering at 1178th Frame (DBSCAN = 3 objects. CBBT = 1 object, 1 possible object)



(c) Double-person clustering at 1272th Frame (DBSCAN = 3 objects. CBBT = 2 objects)



(d) Double-person clustering at 1275th Frame (DBSCAN = 3 objects. CBBT = 1 object, 1 losing object)

**Fig. 10** Double-person clustering results of three-dimensional point cloud data

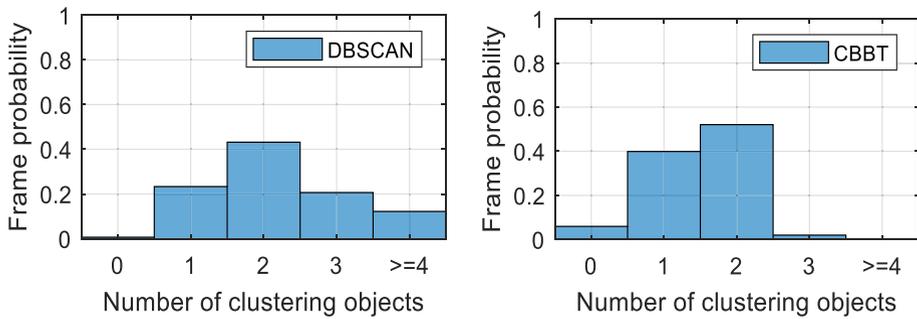


Fig. 11 Target number histogram of clustering

clustering methods in the case of double-person is compared. We extract the point cloud data of double-person clustered by DBSCAN and CBBT and calculate the proportion of the same data in the total clustered data. The statistical result of the same clustered data proportion in the total cluster points is shown in Fig. 12. It can be seen that the same data accounts for 65% in the cluster points with the proportion of more than 80%. It is true that the consistency of double-person is lower than that of single-person, but it still meets the requirement of the clustering consistency.

Likewise, we processed the point cloud data with the above clustering results of two objects for the HAR. The feature extraction method, classifier and confusion matrix are all consistent with the single-person HAR. The HAR result in different frames of the millimeter-wave radar is provided in Fig. 13. For instance, the two persons are both standing at the 833th Frame and walking at the 1272th Frame. One person is standing and the other is lying at the 1178th Frame. One person is walking and the other is falling at the 1275th Frame. Obviously, the different frames can also be very good corresponding to the clustering results and video record. This further illustrates the applicability of our proposed scheme.

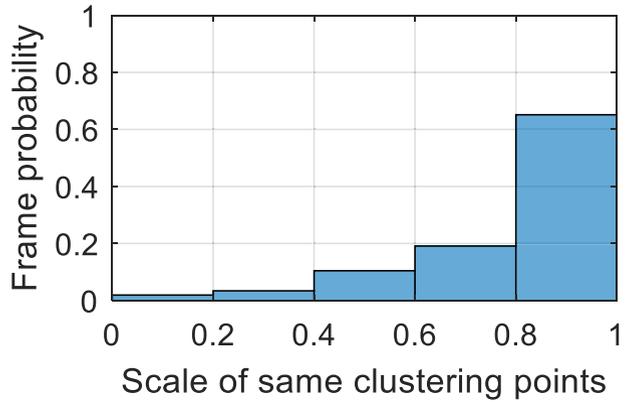
### 5 Conclusion

The MIMO millimeter-wave radar has been adopted in automotive electronics, intelligent transportation, and security applications. In this paper, we mark the important step in HBR based on the sparse point cloud data from MIMO millimeter-wave radar for smart home. The MIMO millimeter-wave radar is used to avoid exposing personal privacy information

Table 6 Joint probability distribution of clustering objects

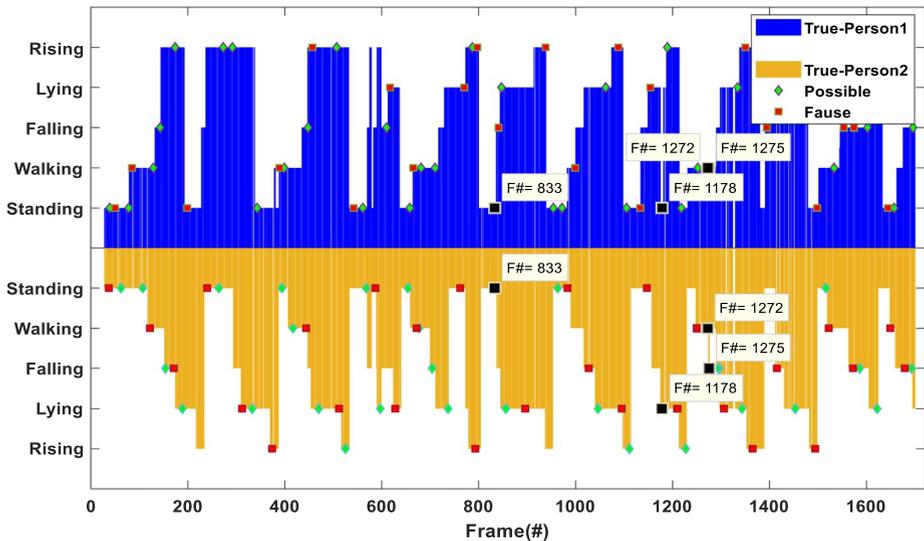
Method/Objects	CBBT					
	0	1	2	3	≥4	
DBSCAN	0	0.005	0.001	0	0	0
	1	0.028	0.14	0.06	0.001	0
	2	0.02	0.16	0.25	0.005	0
	3	0.006	0.065	0.13	0.005	0
	≥4	0.002	0.03	0.084	0.008	0

**Fig. 12** Proportional histogram of the same cluster points for DBSCAN and CBBT (Double-person groups clustering)



and suitable with low cost for smart home. We develop the clustering method of the point cloud data based on the human body tracking, which realizes the single-person and double-person recognition.

From the clustering experiments, it can be seen that CBBT has good consistency with DBSCAN for the point cloud data with the same number of objects from MIMO millimeter-wave radar, which can be considered that most of the clustering data are from the same object reflection. Especially, for the radar data with the wrong clustering results of DBSCAN and the correct clustering results of CBBT, it is reasonable to believe that most of the data clustered by CBBT are still from the points reflected by the object. It is worth noting that DBSCAN is easy to cluster some sudden scene disturbances (such as slight movement of objects) into objects, which will improve the false alarm rate in the scene of multi-person body recognition. While CBBT takes



**Fig. 13** Different actions corresponding to different frames

into account the correlation of human body tracking, and has the ability to resist scene disturbance. Therefore, compared with DBSCAN, the clustering accuracy of CBBT is higher, and the clustering effect of CBBT is better than that of DBSCAN in the scene disturbance. In summary, our proposed HBR scheme can achieve single-person and double-person recognition using the sparse point cloud data of MIMO millimeter-wave radar with satisfactory recognition accuracy.

Though HBR based on the sparse point cloud data from MIMO millimeter-wave radar has achieved good experimental results, there is still the work to develop the research in future. This research aims at HBR of smart home application, the environment is relatively simple. In general, human body activities are complex and diverse. How to combine situational and environmental information for higher semantic behavior recognition is also a further research. Currently, there is a growing demand for body part estimation (e.g. finger joints, facial parts). MIMO millimeter-wave radar can become an alternative or complementary solution to the camera or wearable devices applied in smart home, and it can be better developed by combining machine learning and the vision-based detection approaches. The HBR scheme may also inspire new human activity recognition algorithms for the multi-human activity recognition in a dynamic and complex environment.

**Acknowledgements** This work is granted by the Science and Technology Department of Jilin Province, China [Grant No. 20220101153JC]. Some or all data during the study are available from the corresponding author by request (Email: tjguo\_ciomp@hotmail.com). Authors express their gratitude to IoT Innovation Lab at UTS in Australia for their generously provided the experimental equipment. Authors would also like to thank anonymous reviewers and editors provided many helpful comments on the manuscript.

## Declarations

**Conflict of interest** The authors declare that they have no conflict of interest.

## References

1. Abramovich YI, Frazer GJ, Johnson BA (2013) Principles of modeSelective MIMO OTHR. *IEEE Trans Aerosp Electron Syst* 46(3):1839–1868
2. Banan A, Nasiri A, Taheri-Garavand A (2020) Deep learning-based appearance features extraction for automated carp species identification. *Aquac Eng* 89:102053–102053
3. Chalabi NE, Attia A, Bouziane A, Akhtar Z (2021) Particle swarm optimization based block feature selection in face recognition system. *Multimedia Tools Appl* 80:33257–33273
4. Clemente C, Pallotta L, Maio AD, Soraghan JJ, Farina A (2015) A novel algorithm for radar classification based on doppler characteristics exploiting orthogonal Pseudo-Zernike polynomials. *IEEE Trans Aerosp Electron Syst* 51(1):417–430
5. Erol B, Amin MG, Boashash B, Ahmad F, Zhang YD (2016) Wideband radar based fall motion detection for a generic elderly. 50th Asilomar Conference on Signals, Systems and Computers
6. Fan Y, Bao J, Aljumaily MSL, Li H (2019) Communications via frequency-modulated continuous-wave radar in millimeter wave band. *IEEE Global Communications Conference*
7. Gurbuz SZ, Clemente C, Balleri A, Soraghan JJ (2017) Micro-Doppler-based in-home aided and unaided walking recognition with multiple radar and sonar systems. *IET Radar Sonar Navig* 11(1):107–115
8. Hai D, Himed B (2009) A Virtual Antenna Beamforming (VAB) approach for radar systems by using orthogonal coding waveforms. *IEEE Trans Antennas Propag* 57(2):425–435
9. Ji Y, Yang Y, Xu X, Shen HT (2018) One-shot learning based pattern transition map for action early recognition. *Sig Process* 143(1):364–370
10. Ji XP, Cheng J, Feng W, Tao DP (2018) Skeleton embedded motion body partition for human action recognition using depth sequences. *Sig Process* 143(1):56–58

11. Jin F, Sengupta A, Cao SY, Wu YJ (2020) MmWave radar point cloud segmentation using GMM in multimodal traffic monitoring. *IEEE International Radar Conference*
12. Jiang X, Zhang Y, Yang Q, Deng B, Wang H (2020) Millimeter-wave array radar-based human gait recognition using multi-channel threedimensional convolutional neural network. *Sensors* 20(19):5466–5466
13. Jokanović B, Amin M (2018) Fall detection using deep learning in range-doppler radars. *IEEE Trans Aerosp Electron Syst* 54(1):180–189
14. Laura G, Jiménez JM, Miran T, Jaime L (2018) Wireless technologies for IoT in smart cities. *Netw Protoc Algorithm* 10(1):23–64
15. Laura G, Lorena P, Miran T, Jaime L (2018) System for detection of emergency situations in Smart City Environments Employing Smartphones. In: 2018 International Conference on Advances in Computing, Communications and Informatics (ICACCI), pp 266–272
16. Liang L, Popescu M, Skubic M, Rantz M, Cuddihy P (2011) Automatic fall detection based on Doppler radar motion signature. In: 2011 5th International Conference on Pervasive Computing Technologies for Healthcare, pp 222–225
17. Markopoulos PP, Zlotnikov S, Ahmad F (2019) Adaptive radar-based human activity recognition with L1-Norm linear discriminant analysis. *IEEE J Electromagn Rf Microwaves Med Biol* 3(2):120–126
18. Pham C, Nguyen L, Nguyen A, Nguyen N, Nguyen VT (2021) Combining skeleton and accelerometer data for human fine-grained activity recognition and abnormal behaviour detection with deep temporal convolutional networks. *Multimedia Tools Appl* 80(19):28919–28940
19. Qian K, He ZY, Zhang XY (2020) 3D Point cloud generation with millimeter-wave radar. *Proc ACM Interact Mob Wearable Ubiquitous Technol* 4(4):1–23
20. Ricci R, Balleri A (2015) Recognition of humans based on radar micro-doppler shape spectrum features. *Sonar Navig IET Radar* 9(9):1216–1223
21. Schumann O, Hahn M, Dickmann J, Wöhler C (2018) Semantic segmentation on radar point clouds. 21st International Conference on Information Fusion, pp 2179–2186
22. Varshney N, Bakariya B, Kushwaha AKS (2022) Human activity recognition using deep transfer learning of cross position sensor based on vertical distribution of data. *Multimedia Tools Appl* 81:22307–22322
23. Wan EA, Van der Merwe R, Nelson AT (2000) Dual estimation and the unscented transformation. *Advances in Neural Information Processing Systems*. NIPS 1999, pp 666–672
24. Wang M, Zhang YD, Cui G (2019) Human motion recognition exploiting radar with stacked recurrent neural network. *Digit Signal Proc* 87(2):125–131
25. Wang M, Cui G, Yang X, Kong L (2018) Human body and limb motion recognition via stacked gated recurrent units network. *IET Radar Sonar Navig* 12(9):1046–1051
26. Wei L, Yang X, Chao Z, Shuai L, Ning L (2016) DRFM range false target cancellation method based on slope-varying LFM chirp signal. In: 2016 IEEE 13th International Conference on Signal Processing (ICSP)
27. Zhai W, Zhang Y (2014) Motion parameters estimation of high-speed moving target for stepped frequency chirp signal. In: 2014 International Radar Conference
28. Zhu SY, Xu JH, Guo HQ, Liu QW, Wu SE, Wang HG (2018) Indoor human activity recognition based on ambient radar with signal processing and machine learning. In: 2018 IEEE International Conference on Communications, pp 1–6

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.