Multi Human Posture Classification Using MIMO FMCW Radar Point Cloud and Deep Learning

Sohaib Abdullah, Shahzad Ahmed, Junbyung Park, Chanwoo Choi, Sung Ho Cho*
Department of Electronic Engineering, Hanyang University
Seoul, South Korea
E-mail: {engrsohaib79, shahzad1, jbp0917, choi231121, dragon}@hanyang.ac.kr
*Correspondence: Sung Ho Cho, dragon@hanyang.ac.kr

***Abstract*—Human action and pose recognition in context of health and safety has lately attracted a huge amount of attraction. Human pose recognition using radar is a challenging task since the human under consideration is static. This paper uses Multi Input Multi Output (MIMO) Frequency Modulated Continuous Wave (FMCW) radar to recognize postures of two co-located humans using Convolutional Neural Network (CNN). Two humans at different angles and arbitrary distance (in living room) are simultaneously considered for data collection. Radar-extracted spherical coordinates of posture are acquired using Fast Fourier transform (FFT) and afterwards, spatial transformation is used to convert these points into Cartesian coordinate system. The resultant image shows the posture of two persons in a single image. A clustering approach is used to classify the two postures and CNN is trained to classify each posture. Promising accuracy is achieved for one and two persons posture recognition.**

Keywords—Human Posture Recognition; FMCW Radar; Deep Learning; non-contact healthcare; CNN.

I. INTRODUCTION

The increasing life expectancy [1] and declining birth rate [2] in developed nations have raised concerns regarding the healthcare provision for aging population. Statistics suggest that elderly persons tend to either live alone or with their spouse. For example, the percentage of elderly people living alone was 36% in 2016 [3].

Continuous posture recognition in the home environment can be used to monitor human behavior for remote healthcare applications. Any abnormality in the behavior can be recognized remotely by recognizing the postures or the dynamic movements (activities).

Sensory technology to detect and a sign of danger or medical emergency based on activity and posture recognition can be a useful tool in scenarios where people (or a couple) tend to live alone. Need of having persuasive healthcare for elderly persons is increasing lately. Consequently, wireless (non-contact) human posture and activity recognition is emerging as a prominent research domain since it can provide a framework for an early detection of medical emergency. In particular, we exert our focus on multi-human posture recognition in this paper.

For posture recognition, one of the best candidate solutions could be a camera sensor however, installing camera in home environment raises several privacy related issues. A possible privacy-preserving solution for posture and activity recognition is a radio sensor such as radar. Nowadays, off the shelf (OTS) low cost and compact size radar sensors are abundantly available and have shown their usefulness in several research domains, such as gesture recognition [4] [5], vital sign monitoring [6] [7] and people counting [8].

For human posture recognition based on radar, several research works have been published in recent year [9]. Nowadays, researchers are showing interest in recognizing activities from the radar-extracted point cloud [10] [11] [12]. Activities recognized from the radar point cloud have a huge potential in recognizing the dynamic movements as well as static postures. Lee et al. [11] recognized several postures using radar extracted point clouds. However, two deep learning models were used in that study, first to extract the point cloud followed by posture and activity classification which is a computationally expensive deep learning framework. Another study presented by Singh et al. [12] extracted five exercising activities using radar point cloud followed by deep learning model. Similarly, sleep posture recognition is also a topic of consideration amongst research community [13].

In this study, we are aiming for a deep learning framework which can learn features from raw point cloud images to classify multi-human postures. The postures considered in this study are sitting, standing, lying down, and picking up something from ground. Since we are recognizing the posture of the two co-located humans, ten different scenarios (or combinations) can be visualized based on the four basic postures as shown in Figure 1. We used Multi Input Multi Output (MIMO) Frequency Modulated Continuous Wave (FMCW) radar for data capturing purposes. The cascade radar consists of 86 x 4 receivers (RX) and transmitters (TX). Although high dimensional MIMO configuration is used however, such sensor are now abundantly available in the market.

The Proposed framework is capable of recognizing the postures of two persons located at different angles and arbitrary distance within a small room. If an elderly person is standing or sitting or lying on ground, one can get a detailed insight about what the person is trying to do. For instance, if the person is lying on ground, it can be a sign of fall.

To our knowledge, multi-human posture classification using FMCW radar-extracted point cloud has not been discussed widely so far. The rest of our paper is organized as follows: Section 2 presents the methodology and Section 3 discusses the experimental setup. Afterwards, Section 4 and 5 discuss the results and conclusion respectively.

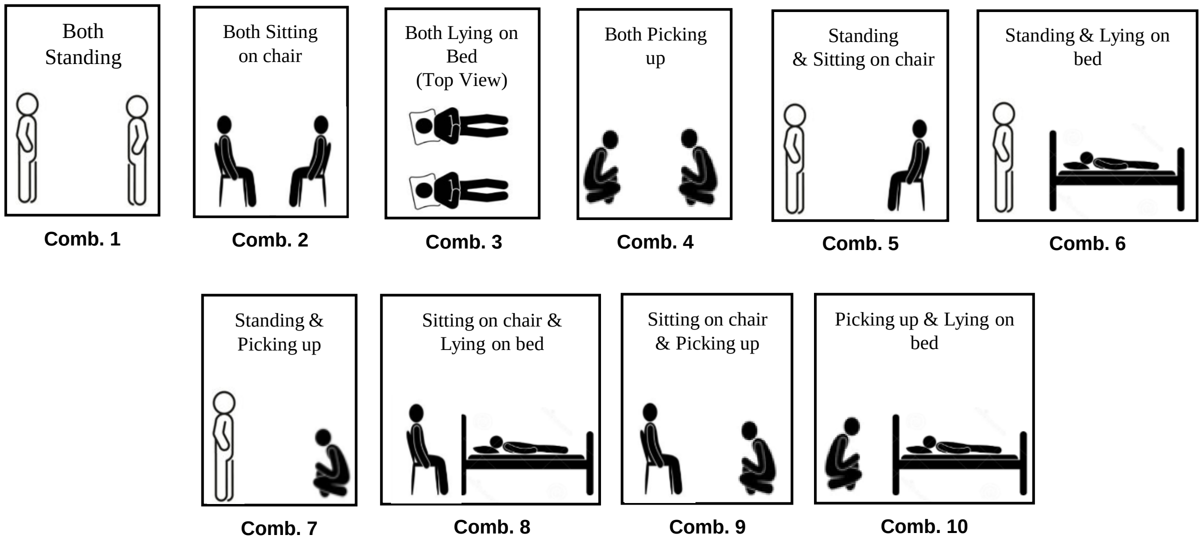


Figure 1. Ten possible combinations of postures (standing, sitting on chair, picking up something and lying down on bed) for two persons.

II. METHODOLOGY

Figure 2 shows the overall methodology of our point cloud based activity recognition framework. First the data is collected from two human participants closely located humans using FMCW radar. Afterwards, point cloud is generated using range, doppler and angle information, which will provide us the spherical coordinates of the radar returns. These spherical radar returns are further converted into three dimensional (3D) Cartesian coordinate system. Afterwards, a Deep Convolutional Neural Network (DCNN) model is trained and evaluated. Note that in our study, the extracted 3D point cloud in its raw form shows the shape of actual posture up to some extent. Next, following passages discuss each step in further detail.

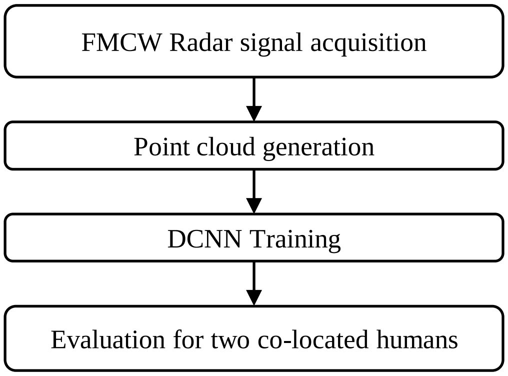


Figure 2. High level block diagram of posture recognition framework.

A. FMCW Radar Data Pre-processing

In FMCW radar, the frequency of the transmitted signal increases linearly with time and the increasing ramp is known as chirp. Several chirps are transmitted simultaneously in a single frame. This allows the FMCW radar to extract the distance and velocity in simultaneous fashion. In addition to that, the diversity created by MIMO settings is used to extract the angle of the target as well. The transmitted signal $x(t)$ can be expressed as:

$$x(t) = exp(j2\pi(f_c t + \frac{B}{T}t^2)) \tag{1}$$

where the term B represents the bandwidth of signal, f_c denotes the operating frequency and T is pulse duration. After colliding with the body of human in sitting, standing, picking up something, or lying down settings, the received signal will be:

$$x(t) = exp(j2\pi(f_c(t-\tau) + \frac{B}{T}(t-\tau)^2)) \tag{2}$$

where τ represents the delay between transmitted and received signal. After multiplying the transmitted signal with the copy of received signal, the low frequency signal which carries the information about the target is termed as the Intermediate Frequency (IR) signal and expressed as:

$$x_{IF}(t) = exp(j2\pi(f_c\tau + \frac{B}{2T}\tau^2)) \tag{3}$$

This low frequency signal is sent to the computer and further exploited to extract target information. The FFT of this signal will give us the distance information of the targets present within the operational range of radar. The same process is repeated for all the RX channels to extract the target information at each receiver. Note that we require multiple channels to find the azimuth and elevation angles of the target.

B. Target Detection using CFAR

The acquired signal at each RX channel is passed through a Constant False Alarm Rate (CFAR) detection algorithm to detect the target and neglect the noise. A 2D-CFAR is applied

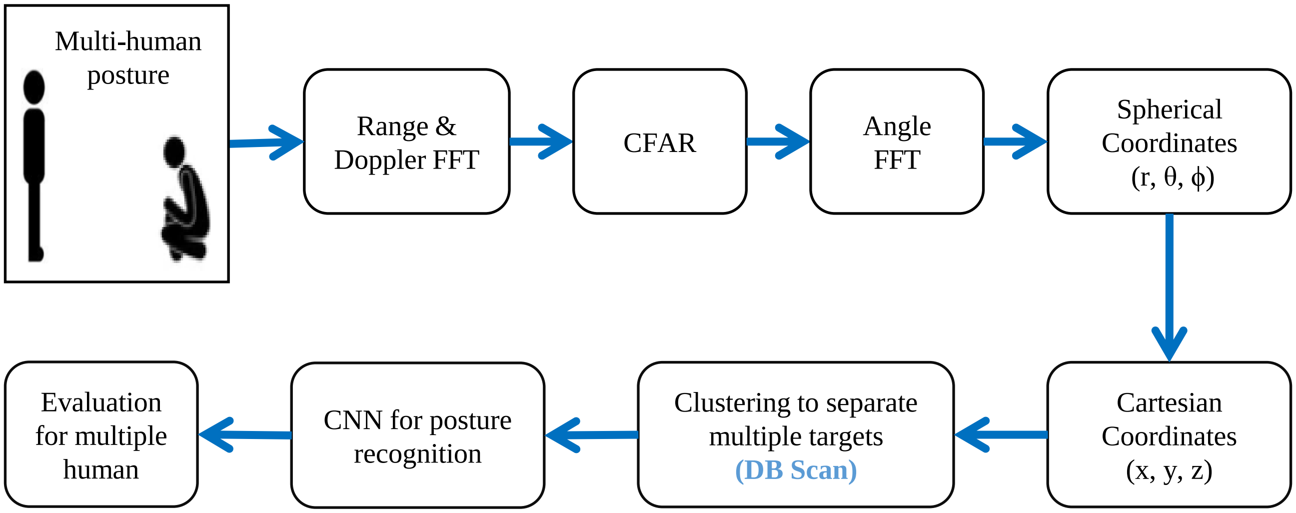


Figure 3. End-to-End framework from data capturing to classification.

on radar signals, first in the range dimension followed by doppler dimension. Specifically, Cell-Averaging Order Statistics (CASO-CFAR) is used as it leverages the benefits of both CA-CFAR and OS-CFAR, thus making robust target detection in low as well as high cluttered environments.

C. Data to Point Cloud Conversion

Once targets are detected, a MIMO FMCW radar allows the extraction of distance (r), azimuth angle (θ), and the elevation angle (ϕ) of the targets. These spherical values of each back-scattered reflection shows where the target is located in the spherical coordinate system. These values are further converted into the Cartesian coordinate system by using below set of identities:

$$x = r\sin(\theta)cos(\phi) \tag{4}$$

$$y = r\cos(\theta)cos(\phi) \tag{5}$$

$$z = rsin(\phi) \tag{6}$$

where the pairs x, y, z represents the Cartesian points of the target corresponding to the spherical points r, θ, ϕ.

D. Recognition Framework

1) Radar Data to Image Conversion

The overall end-to-end recognition framework is shown in Figure 3. As discussed earlier, taking the FFT of each received frame shown in (3), provides the distance of back-scattered signal. Taking another FFT across each chirp within a frame provides the velocity FFT. Prior to the extraction of angle, CFAR detection is applied to find human targets. Afterwards, taking FFT of range-Doppler across each receiving channel will provide the angles of the target denoted as r, θ, ϕ.

After extracting the Cartesian coordinate points of the target, the point where human is located appears as a cluster of co-located points. These patterns are saved as a 2-D images for Deep learning model training purposes.

2) Model Training

In this work, the model is first trained to recognize each individual posture. A four-class classifier using DCNN model named as shuffle-net [14] is used for that purpose. Image containing a single human posture is first labeled, since the supervised learning approach requires dataset labeling [15]. In comparison to the other deep variants of CNN such as AlexNet and GoogleNet [4], Shuffle-net is extremely efficient in terms of computation [14].

Several density-based region selection methods exist for recognition of multiple objects in an image such as region based rCNN and fast rCNN [16]. We opted a simple clustering approach to detect multiple objects in an image. DBSCAN algorithm is opted to find multiple (two) clusters in an image. DBSCAN joins the point based on regional density, rather than the distances. It has an extensive usage in radar-based target localization applications [17]. In this way, multiple images are generated from a single image containing posture of two persons, and multiple deep learning classifiers are used to find the two postures within one image.

III. EXPERIMENTAL SETUP

The experimental setup and equipment to capture radar point cloud is shown in Figure 4. Two participants were in front of radar with a slightly distinct angle. Two commercial radars named AWR2243 CASCADE and IWR6843ISK-ODS were used to show the posture quality of a high and low virtual antenna array. Both radars are manufactured by Texas Instrument (TI), United States. CASCADE radar offers 86 x 4 horizontal and vertical arrays for azimuth and elevation calculation whereas ODS radar consists of 4 x 4 azimuth

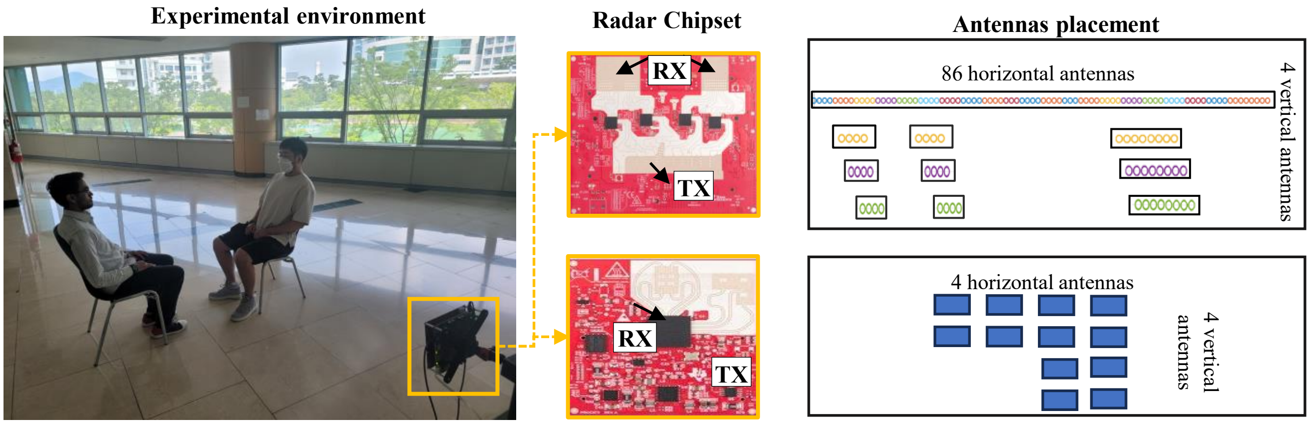
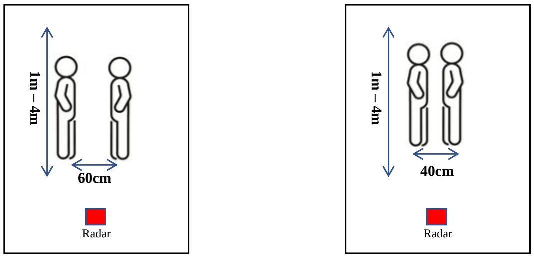


Figure 4. Experimental setup and equipment. Left: Experimental environment; middle: OTS FMCW radars; right: Antennas placement configuration.



(a) Case-I

(b) Case-II

Figure 5. Experimental setup showcasing target range and separation.

and elevation antennas. For CASCADE radar, two cases were taken into study as shown in Figure 5. In Case-I, the range of targets from radar was within one to four meters, while keeping a fixed 60cm separation between both participants at all times. In Case-II, the separation between participants was further reduced to 40cm. For ODS radar, separation of participants was set to 60cm at 2.5m range. A Total of 600 samples were collected for all the ten combinations shown in Figure 1 and data from two different participants was collected.

The rest of the hardware parameters of both radars are shown in Tables I.

TABLE I. Radar Sensor Parameters.

Parameter	CASCADE Radar	ODS Radar
Start Frequency	77 GHz	60 GHz
Bandwidth	3.3 GHz	4 GHz
Number of Chirps	32	32
Number of ADC samples	256	256
Frame Rate	20 FPS	20 FPS
Number of Frames	20	20
Number of TX antennas	12	3
Number of RX antennas	16	4
Antenna Array (TX x RX)	86 x 4	4 x 4
Range Resolution	4.5 cm	3.76 cm
Azimuth Angle Resolution	1.4^o	29^o
Elevation Angle Resolution	18^o	29^o

IV. RESULTS

This Section discusses the results obtained using CASCADE and ODS FMCW radars and compare their efficacies for co-located multi-person posture recognition.

A. Point Cloud Visualization with Cascade Radar

1) Case-I: 60cm Separation between Targets and Arbitrary Range

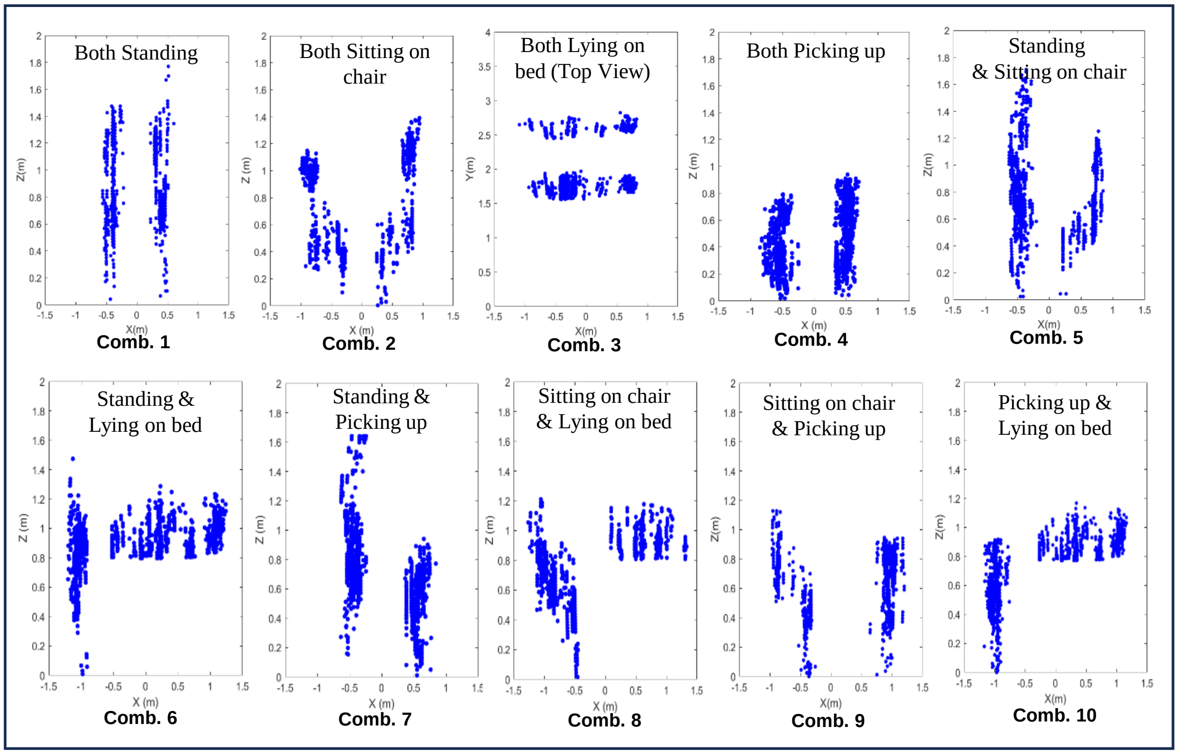
Figure 6 shows the clustered point-cloud of all the ten combinations of postures using high resolution cascade radar. The first four images are the cases when both the participants are having same posture, that is to say, both participants are standing, sitting, picking up something from ground, or lying down. Afterwards, the individual postures are combined in such a way that there is no repetition of posture in the patterns. As shown in Figure 6, the cascade radar is capable of extracting the exact postures of two participants with high precision.

2) Case-II: 40cm Separation between Targets and Arbitrary Range

The cascade radar can clearly differentiate two closely located humans, due to very high azimuth angle resolution (see Tables I). Figure 7 shows some combinations of postures in which both participants are in very close proximity. In this case, feet of both participants were in touch with each other. For both sitting scenario, point cloud for left participant has some empty area because of very small separation between them. These two figures suggest that a high resolution antenna diversity radar is required for multiple person human posture recognition problem. Figure 6 and Figure 7 suggest that if the two participants are slightly distinct in terms of horizontal angle, the both the postures can be visualized using the framework presented in this paper. In the end, samples from both cases were merged for final classification.

B. Point Cloud Visualization with ODS Radar

The point cloud extracted using the radar consisting of fewer TX and RX antennas was not as accurate as the one shown in Figure 6. Hence, only the point cloud images extracted



Comb. 1 **Comb. 2** **Comb. 3** **Comb. 4** **Comb. 5**

Comb. 6 **Comb. 7** **Comb. 8** **Comb. 9** **Comb. 10**

Figure 6. Point cloud for all the ten combinations of postures using cascade FMCW radar (for Case-I).

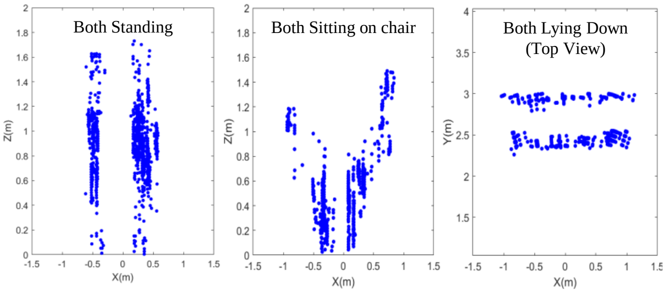


Figure 7. Point cloud of two persons extracted using cascade FMCW radar (for Case-II).

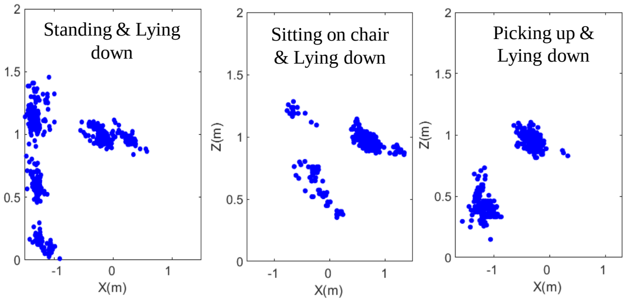


Figure 8. Point cloud of two persons, extracted using a low (angle) resolution (ODS) FMCW radar.

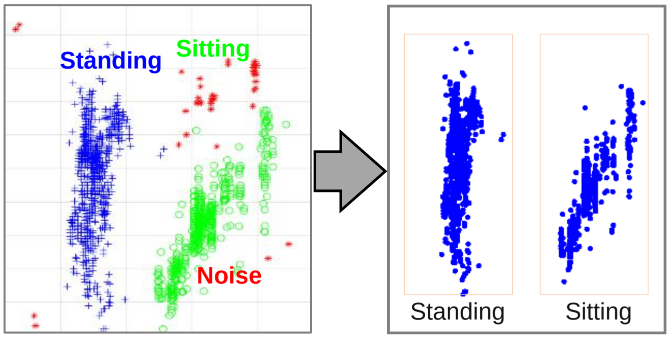
using CASCADE FMCW radar are used in this study. A few scenarios of two persons point cloud extracted using a 4x4 low (angle) resolution FMCW radar are shown in Figure 8. In order to be properly detected as two distinct targets, the separation between two targets was set to 60cm but the point cloud generated using 4x4 radar is not clear in comparison to the point cloud generated using cascade radar.

C. *Classification Accuracy*

Although we performed classification using both the radars, the confusion matrix of Cascade radar is included in the paper since the accuracy of ODS radar was very low. Consequently, we can conclude that a radar with high angle resolution can be used to extract point cloud based posture of multiple human subjects consecutively.

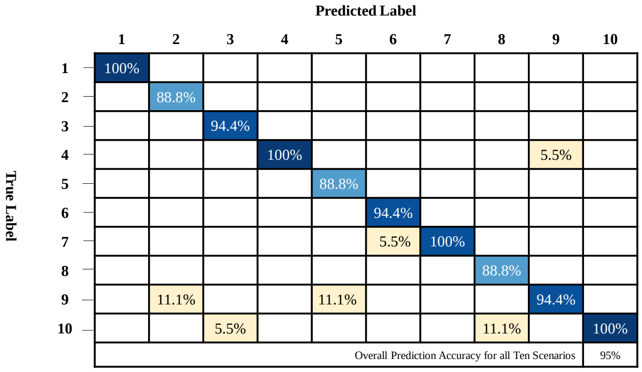
The results of the DBSCAN based image cropping are shown in Figure 9. The image on the left side shows separate clusters being formed for standing and sitting postures. In addition to that, noise is also shown as a third cluster. For the captured dataset, DBSCAN clustering approach was able to divide the input point cloud image having two postures into two separate images.

The confusion matrix presented in Figure 10 shows the classification accuracy of all the ten scenarios using Cascade radar, consisting of 86 RX horizontal and 4 vertical antennas (see Figure 4). It can be seen that, 2nd, 5th and 8th combination showed highest classification error followed by 3rd, 6th and 9th combination as expressed in the confusion matrix shown in Figure 10. Although DBSCAN allows robust target segregation while rejecting noise, some information of the human body can also be lost, if not properly detected by the radar. This is why, sitting on chair and lying down was confused with picking up in some cases when head and feet were not properly detected by the radar. Nevertheless, the algorithm was able to classify the two persons scenario with an overall success rate of 95%.



Point cloud of two postures **Separated images**

Figure 9. DBSCAN based posture segregation approach.



Predicted Label

True Label	1	2	3	4	5	6	7	8	9	10
1	100%									
2		88.8%								
3			94.4%							
4				100%					5.5%	
5					88.8%					
6						94.4%				
7						5.5%	100%			
8								88.8%		
9		11.1%			11.1%				94.4%	
10			5.5%					11.1%		100%
Overall Prediction Accuracy for all Ten Scenarios										95%

Figure 10. Confusion matrix for Cascade radar.

V. CONCLUSION AND FUTURE WORK

This paper presents a framework to recognize the posture of two humans located at different angles. A point cloud map showing both the person is first constructed using FMCW radar, and DBSCAN clustering is used to segregate the postures of the two humans. Afterwards, deep learning is used to find the postures of each human. A considerable amount of accuracy is observed while using a high- resolution antenna array.

This framework enables non-intrusive surveillance of elderly couples by providing an assessment of their current position, like lying down. This allows medical professionals and caregivers to intervene promptly in case of emergencies. In addition to static poses, the framework can be extended to detect subtle changes in movement patterns providing more insight into potential risks, like fall detection and prevention.

ACKNOWLEDGMENT

This study was supported by National Research Foundation (NRF) of South Korea (NRF-2022R1A2C2008783). Author 1 and 2 contributed equally as co-first authors.

REFERENCES

[1] L. F. Berkman and B. C. Truesdale, “Working longer and population aging in the us: Why delayed retirement isn’ta practical solution for many,” *The Journal of the Economics of Ageing*, vol. 24, p. 100438, 2023.

[2] O. Okpareke, A. Lakhanpal, and S. Chattopadhyay, “The decline in us birthrates in recent years is indicative of cultural and economic changes,” 2022.

[3] T.-H. Tan *et al.*, “Binary sensors-based privacy-preserved activity recognition of elderly living alone using an rnn,” *Sensors*, vol. 21, no. 16, p. 5371, 2021.

[4] S. Ahmed and S. H. Cho, “Hand gesture recognition using an ir-uwb radar with an inception module-based classifier,” *Sensors*, vol. 20, no. 2, p. 564, 2020.

[5] S. Ahmed, K. D. Kallu, S. Ahmed, and S. H. Cho, “Hand gestures recognition using radar sensors for human-computer-interaction: A review,” *Remote Sensing*, vol. 13, no. 3, p. 527, 2021.

[6] M. Kebe *et al.*, “Human vital signs detection methods and potential using radars: A review,” *Sensors*, vol. 20, no. 5, p. 1454, 2020.

[7] S. Ahmed, J. Park, and S. H. Cho, “Effects of receiver beamforming for vital sign measurements using fmcw radar at various distances and angles,” *Sensors*, vol. 22, no. 18, p. 6877, 2022.

[8] C. Y. Aydogdu, S. Hazra, A. Santra, and R. Weigel, “Multi-modal cross learning for improved people counting using short-range fmcw radar,” in *2020 IEEE International Radar Conference (RADAR)*. IEEE, 2020, pp. 250–255.

[9] S. Ahmed, J. Park, and S. H. Cho, “Fmcw radar sensor based human activity recognition using deep learning,” in *2022 International Conference on Electronics, Information, and Communication (ICEIC)*. IEEE, 2022, pp. 1–5.

[10] Y. Huang *et al.*, “Activity recognition based on millimeter-wave radar by fusing point cloud and range–doppler information,” *Signals*, vol. 3, no. 2, pp. 266–283, 2022.

[11] G. Lee and J. Kim, “Improving human activity recognition for sparse radar point clouds: A graph neural network model with pre-trained 3d human-joint coordinates,” *Applied Sciences*, vol. 12, no. 4, p. 2168, 2022.

[12] A. D. Singh, S. S. Sandha, L. Garcia, and M. Srivastava, “Radhar: Human activity recognition from point clouds generated through a millimeter-wave radar,” in *Proceedings of the 3rd ACM Workshop on Millimeter-wave Networks and Sensing Systems*. Association for Computing Machinery, 2019, pp. 51–56.

[13] J. E. Kiriazi, S. M. Islam, O. Borić-Lubecke, and V. M. Lubecke, “Sleep posture recognition with a dual-frequency cardiopulmonary doppler radar,” *IEEE Access*, vol. 9, pp. 36 181–36 194, 2021.

[14] X. Zhang, X. Zhou, M. Lin, and J. Sun, “Shufflenet: An extremely efficient convolutional neural network for mobile devices,” in *Proceedings of the IEEE conference on computer vision and pattern recognition*. IEEE, 2018, pp. 6848–6856.

[15] S. Asghar, J. Choi, D. Yoon, and J. Byun, “Spatial pseudo-labeling for semi-supervised facies classification,” *Journal of Petroleum Science and Engineering*, vol. 195, p. 107834, 2020.

[16] R. Girshick, “Fast r-cnn,” in *Proceedings of the IEEE international conference on computer vision*. IEEE, 2015, pp. 1440–1448.

[17] L. Y. Chan, D. Genschow, and U. T. Schwarz, “Combining delta-phi velocity measurement and dbscan clustering to localize slowly moving objects in short ranges with limited slow-time radar data,” in *2023 24th International Radar Symposium (IRS)*. IEEE, 2023, pp. 1–11.