Data in Brief 59 (2025) 111316

Contents lists available at ScienceDirect

Data in Brief

journal homepage: www.elsevier.com/locate/dib

Data Article

mmPrivPose3D: A dataset for pose estimation and gesture command recognition in human-robot collaboration using frequency modulated continuous wave 60Hhz RaDAR

Nima Roshandel[a,b,e,*], Constantin Scholz[a,b], Hoang-Long Cao[a,c,*], Milan Amighi[a,b], Hamed Firouzipouyaei[a,b], Aleksander Burkiewicz[a,b], Sebastien Menet[a,b], Felipe Ballen-Moreno[a,c], Dylan Warawout Sisavath[a,b], Emil Imrith[a], Antonio Paolillo[a,f], Jan Genoe[d,e], Bram Vanderborght[a,e]

[a] *Brubotics, Vrije Universiteit Brussel, Brussels, Belgium*
[b] *imec-IMS-VUB, Belgium*
[c] *Flanders Make, Brussels, Belgium*
[d] *imec-SAT, Belgium*
[e] *KU Leuven, Belgium*
[f] *SOFT Languages Lab, Vrije Universiteit Brussel, Belgium*

ARTICLE INFO

Article history:
Received 21 August 2024
Revised 6 January 2025
Accepted 13 January 2025
Available online 16 January 2025

Dataset link: mmPrivPose3D: DataSet (Original data)

Keywords:
Human-robot collaboration
IWR6843AOPEVM
RaDAR
Pose estimation
Gesture command recognition

ABSTRACT

3D pose estimation and gesture command recognition are crucial for ensuring safety and improving human-robot interaction. While RGB-D cameras are commonly used for these tasks, they often raise privacy concerns due to their ability to capture detailed visual data of human operators. In contrast, using RaDAR sensors offers a privacy-preserving alternative, as they can output point-cloud data rather than images. We introduce mmPrivPose3D, a dataset of 3D RaDAR point-cloud data that captures human movements and gestures using a single IWR6843AOPEVM RaDAR sensor with a frequency of 10 Hz synchronized with 19 corresponding 3D skeleton keypoints as the ground truth. These keypoints were extracted from RGB-D images captured by an Intel RealSense

* Corresponding authors.
E-mail addresses: nima.roshandel@vub.be (N. Roshandel), hoang.long.cao@vub.be (H.-L. Cao).
Social media: @BramVDBorght (B. Vanderborght)

https://doi.org/10.1016/j.dib.2025.111316

camera recorded at 30 frames per second using the Nuitrack SDK, and labeled with gestures. The dataset was collected from $n = 15$ participants. Our dataset serves as a fundamental resource for developing machine learning algorithms to improve the accuracy of pose estimation and gesture recognition using RaDAR data.

Specifications Table

Subject	Manufacturing Engineering
Specific subject area	3D point-cloud data of human movements and gestures captured by a RaDAR sensor synchronized with RGB-D images installed on the base of a robot manipulator.
Type of data	*3D point-cloud, 3D skeleton keypoints* *Raw, Labelled*
Data collection	We collected 3D point-clouds of humans from an IWR6843AOPEVM RaDAR with 10 Hz data frequency and 3D keypoints extracted by the Nuitrack SDK from RGB-D image from an Intel RealSense L515 camera at 30 frame rate per second as the ground truth in real-time, we synchronized the frames of both sensors together. An offset between the sensors (3.9 cm in both horizontal and vertical directions) and a 90-degree rotation around the local x-axis were applied to the camera data for spatial calibration and alignment with the RaDAR's coordinate system.
Data source location	Institution: Brubotics, Vrije Universiteit Brussels District/City: Elsene, Brussels Country: Belgium
Data accessibility	Repository name: Mendeley Data Data identification number: 10.17632/pmdr5rgn8c.1 Direct URL to data: https://data.mendeley.com/datasets/pmdr5rgn8c/1
Related research Article	Not yet available (Under Submission)

1. Value of the Data

- The RaDAR sensor approach offers a privacy-preserving method for pose estimation and gesture recognition compared to visual data (e.g. images and video), which is particularly relevant for compliance with GDPR and other data protection regulations.
- This dataset provides 24,005 samples of 3D point-cloud data synchronized with 3D skeleton keypoints extracted from RGB-D images, which are essential for developing and training machine learning models that improve pose estimation and gesture recognition using RaDAR sensors.
- This dataset uses a single frontally facing IWR6843AOPEVM RaDAR sensor compared to multiple RaDAR sensors used in the other datasets.
- Researchers can use this dataset to benchmark new algorithms for pose estimation and gesture recognition, and validate the performance of their models for safety in human-robot collaborative manufacturing or industrial automation. For example, the predicted human operator's body keypoints can be used in a speed and separation monitoring mode according to ISO 15,066 so that the robot slows down or stops depending on the operator's proximity.

2. Background

Collaborative robots face challenges in ensuring safety at high speeds, and their rapid operation can compromise human safety [1–3]. In this context, accurate 3D pose estimation of human operators and their gesture commands is necessary to ensure safety and facilitate human-robot interaction. Sensor-based solutions have been implemented to improve safety while maintaining efficiency. For example, light-based sensors such as RGB-D cameras are widely used but often struggle with environmental conditions and raise privacy concerns [4,5]. To address these limitations, RaDAR sensors offer a promising alternative, providing reliable detection even in harsh environments [6]. For example, the MARS dataset contains approximately 40,000 frames of upper/lower limb extension, front/side lunge, and squat as activities [7]. However, this dataset is collected from only four individuals, leading to a lack of diversity among participants. Additionally, it used a 77 GHz IWR1443BOAST which is not authorized for industry purposes. Another dataset is mmpose-NLP which contains 16,200 RaDAR point-cloud samples of walking back and forth, and arm swing as activities [8]. This dataset has the same disadvantages as the previous dataset by using two RaDAR sensors without industrial authorized frequency and with only two participants. Finally, the HuPR dataset contains 141,000 samples performing hand waving, standing, and walking [9]. This dataset was collected from six participants and has stored range-doppler heatmaps as the data. It used two 77 GHz IWR1843BOAST RaDAR sensors which is not authorized for industry purposes. Using heatmap instead of point-clouds enables more advanced applications while increasing the data size and making the dataset more dependent on the hardware type being exactly the dataset sensor.

3. Data Description

The data was collected in a public environment at the Brubotics Lab, Vrije Universiteit Brussel (Belgium). The mmPrivPose3D dataset is organized into two folders: **pose_estimation** and **gesture_recogtion**, see Fig. 1. Each folder includes CSV files of 3D point-cloud and 3D skeleton keypoints labeled with gesture commands. The names of 19 skeleton keypoints are illustrated in Fig. 2. The format of each CSV file is detailed in Table 1. This dataset contains 24,005 samples for the pose estimation model from participants walking and 86,483 samples of participants hand-waving. Free walking and left/right-hand waving were chosen based on their popularity in human-robot collaboration applications and relevant datasets ([10,11,9,12,13]). The purpose of gathering more samples from hand-wavings is that these movements are from smaller body parts (hands and wrists), which reflect less power than larger ones, as mentioned in various papers, including RFPose3D [14]. This results in one RaDAR point-cloud set from a waving human that does not have enough points to enable the classifier to distinguish between two motions. As a result, to gather more discriminative features between the point cloud data from the two motions and classify them correctly, the number of samples was increased. In total, the size of this dataset is 195.8 MB.

Table 1
Format of the CSV file.

Elapsed Time	Timestamp	Sorted Point-cloud	Keypoints 3D	Label
[Timestamp (ms)]	[Date Time]	(x_i, y_i, z_i) for $i = 1,...,n$	(x_i, y_i, z_i) for $i = 1,...,19$	None|Left-hand wave|Right-hand wave
62.0303154	2024-06-24 11:28:09.589416	[-1.87980679 3.52237416 -0.37596134],...	[-0.0235302 3.74955- 0.20174967],...,	None
60.85205078	38:06.0	[-2.07987979 5.63128322 0.75631991],...	[0.05464058 3.1844783 -0.1878637],...	Left-hand wave
59.96990204	48:40.8	[-1.78086966 6.06348627 0.3957488],...	[0.0709745 3.7584932 -0.2031855],...	Right-hand wave

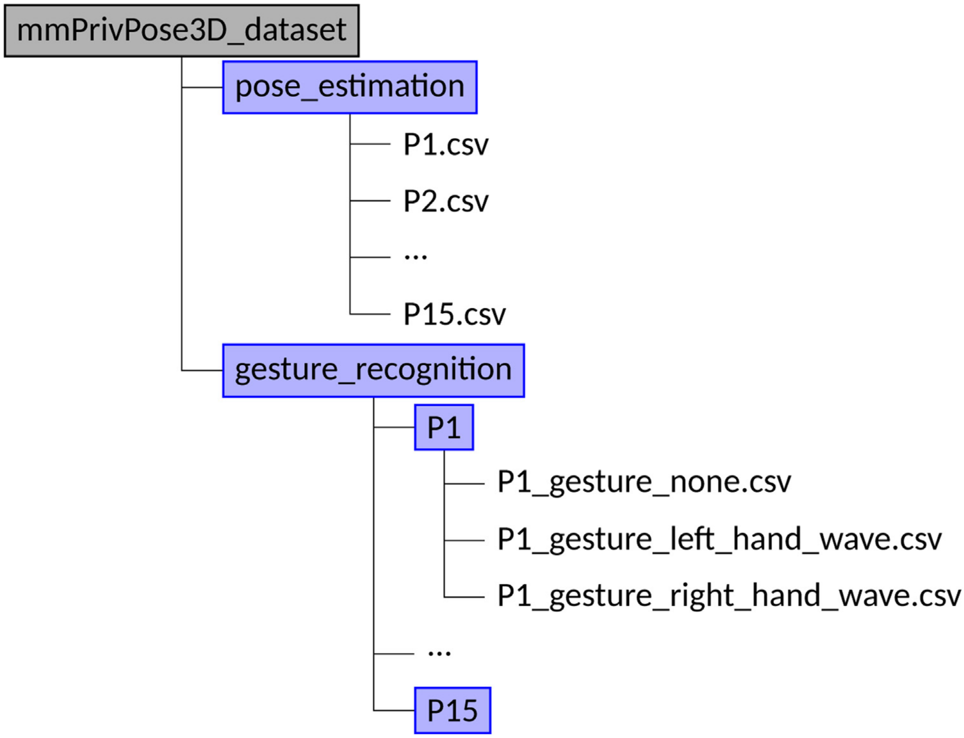

Fig. 1. The dataset structure of the mmPrivPose3D_dataset in Mendeley data.

4. Experimental Design, Materials and Methods

We recruited a diverse group of 15 participants, including 14 males and 1 female, aged between 24 and 37 years, with heights ranging from 1.7 to 1.8 m and varying body sizes. It should be noted that the sensors used in this work are independent of demographic variables such as gender and age and they only capture the mechanical aspects of movement. The sensors are also not sensitive to height. While this range may not cover all possible heights globally, it is a reasonable height range in Europe. This reflects our practical and logistical conditions rather than intentional exclusion. RaDAR is resilient to environmental variations and therefore supports the reliability of this dataset across environmental conditions [6]. The data collection received ethical approval from the Ethics Committee for Human Sciences at the Vrije Universiteit Brussel on 26th April 2024 (ECHW_511). Informed consent was obtained from all participants for data collection and processing.

The experimental setup is illustrated in Fig. 3A–B. We collected 3D point-clouds of a single operator in front of the robot from an IWR6843AOPEVM RaDAR installed on the robot base with a 10 Hz data frequency. At the same time in parallel, 3D skeleton keypoints are extracted from RGB-D images captured by an Intel RealSense L515 camera with a 30 frame rate per second (fps) using the Nuitrack SDK, which then are synchronized. During the experiments, the operator moved in an area where their distance to both sensors can vary between two to four meters. The choice of a minimum distance of two meters was to ensure that human safety was maintained while human-robot interaction as it is above the reach of long collaborative robots such as UR10e (1.3 m). In addition, this minimum distance ensures capturing the full body of a human that can even have a height of 2.21 m within the field of view of both sensors (RaDAR and Intel RealSense L515 used as ground truth). The experiment was performed in a controlled environment free of

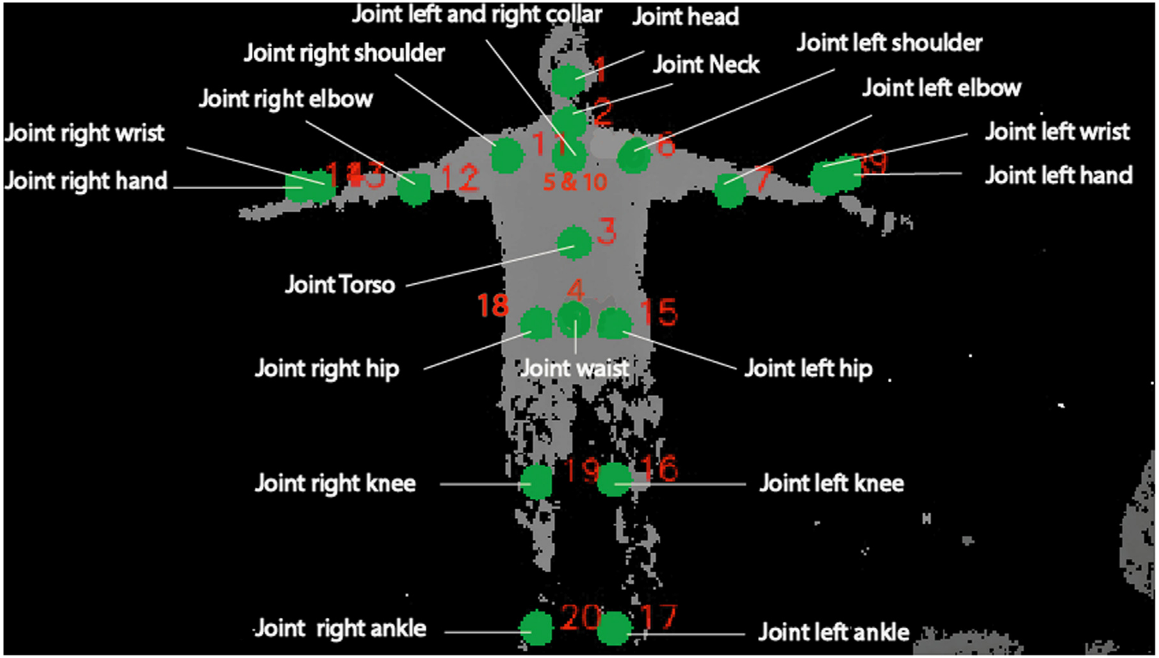

Fig. 2. Arrangement of the 19 3D skeleton keypoints extracted from RDG-D images.

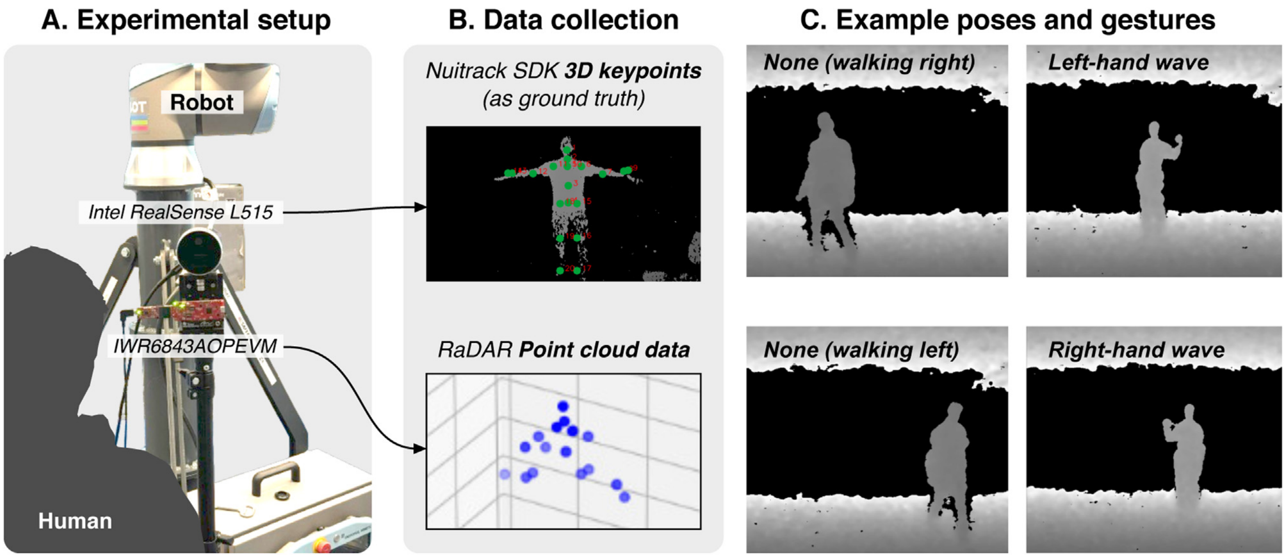

Fig. 3. The experimental setup. A RaDAR sensor is used to collect a 3D point-cloud while an RGB-D camera is used to capture images for 3D keypoints extraction.

objects other than humans in front of the RaDAR to ensure minimum reflections that might result in outliers. A statistical outlier removal algorithm was also placed to remove any possible outliers. The 3D keypoints serve as the ground truth in real time. An offset between the sensors (3.9 cm in both horizontal and vertical directions) and a 90-degree rotation around the local x-axis were applied to the camera data for spatial calibration and alignment with the RaDAR's coordinate system. The RaDAR sensor was mounted 1.15 m above the ground, chosen based on the 120° field of view of the sensor in azimuth and elevation to ensure full human body coverage between two to four meters from the sensor. To account for potential misalignments or drift, the experimental setup was regularly inspected to verify that both alignment and mounting distance remained consistent. Periodic recalibration checks were performed using known reference points to maintain accurate sensor alignment and address any shifts or drift. Some examples of poses, left-hand wave and right-hand wave gestures can be seen in Fig. 3C.

This dataset can be used to develop machine learning algorithms for pose estimation and gesture command recognition using RaDAR data. Researchers and developers can develop different models based on their targeted applications. As **an example**, we have used a 3D Convolutional Neural Network (CNN) architecture to extract 19 human skeleton keypoints. Separately, we have used a random forest classifier for gesture command recognition to overcome the issue with the low reflectivity [14]. For pose estimation, we achieved the lowest overall MPJPE (Mean Per Joint Position Error) of 48.3 mm, outperforming RPM 2.0 [15] at 57.5 mm, RPM [13] at 59.2 mm, and RFPose3D [14] at 134.1 mm. For gesture command recognition, we achieved an inference accuracy of 96.2 %, outperforming the CNN model (three layers of convolution, pooling, and fully connected layers) at 94.9 %, the Set-Transformer (utilizing an encoder-decoder framework with multi-head self-attention) at 94.3 %, and a three-layer LSTM model at 85.7 % on the same dataset.

Limitations

To keep the dataset lightweight, we do not provide the raw RaDAR data but the point-cloud output from the RaDAR's DSP based on-RaDAR chip range bin processing.

We relied on the Nutritrack SDK which does not account for the orientation of the human relative to the radar sensor. As a result, when the human turns 180°, the system incorrectly maps the left foot to the right foot (backward) and vice versa, as the skeleton model does not rotate with the subject.

The single RaDAR sensor in this dataset has a limited field of view (FOV). While additional RaDAR sensors could expand coverage, this dataset focuses on a single FOV for a compact hardware installation.

Ethics Statement

The data collection received ethical approval from the Ethics Committee for Human Sciences at the Vrije Universiteit Brussel on 26th April 2024 (ECHW_511). Informed consent was obtained from all participants for data collection and processing.

CRediT Author Statement

All authors contributed to the study's conception and design. Material preparation, data collection, and analysis were performed by Nima Roshandel. The first draft of the manuscript was written by Nima Roshandel, Constantin Scholz and Hoang-Long Cao. All authors commented on previous versions of the manuscript. All authors have read and approved the final manuscript. The funding was acquired by Constantin Scholz, Bram Vanderborght and Jan Genoe.

Data Availability

mmPrivPose3D: DataSet (Original data) (Mendeley Data).

Acknowledgements

The work leading to these results received funding from euROBIN under Grant Agreement No. 101070596, and as part of the SAFEBOT program through funding provided by imec.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

[1] C. Scholz, H.-L. Cao, E. Imrith, N. Roshandel, H. Firouzipouyaei, A. Burkiewicz, M. Amighi, S. Menet, D.W. Sisavath, A. Paolillo, X. Rottenberg, P. Gerets, D. Cheyns, M. Dahlem, I. Ocket, J. Genoe, K. Philips, B. Stoffelen, J. Van den Bosch, S. Latre, B. Vanderborght, Sensor-enabled safety systems for human–robot collaboration: a review, IEEe Sens. J. 25 (1) (2025) 65–88.
[2] C. Scholz, H.-L. Cao, I. El Makrini, S. Niehaus, M. Kaufmann, D. Cheyns, N. Roshandel, A. Burkiewicz, M. Shhaitly, E. Imrith, et al., Improving robot-to-human communication using flexible display technology as a robotic-skin-interface: a co-design study, Int. J. Intell. Robot. Appl. (2024) 1–18.
[3] E. Matheson, R. Minto, E.G. Zampieri, M. Faccio, G. Rosati, Human–robot collaboration in manufacturing applications: a review, Robotics 8 (4) (2019) 100.
[4] D. Pascual-Hernández, N.O. de Frutos, I. Mora-Jiménez, J.M. Canas-Plaza, Efficient 3d human pose estimation from rgbd sensors, Displays 74 (2022) 102225.
[5] W. Ma, K. Wang, J. Li, S.X. Yang, J. Li, L. Song, Q. Li, Infrared and visible image fusion technology and application: a review, Sensors 23 (2) (2023) 599.
[6] C. Zhu, Z. Zhao, Z. Shan, L. Yang, S. Ji, Z. Yang, Z. Zhang, Robust target detection of intelligent integrated optical camera and mmwave radar system, Digit. Signal. Process. 145 (2024) 104336.
[7] S. An, U.Y. Ogras, Mars: mmwave-based assistive rehabilitation system for smart healthcare, ACM Trans. Embedded Comput. Syst. (TECS) 20 (5s) (2021) 1–22.
[8] A. Sengupta, S. Cao, mmpose-nlp: a natural language processing approach to precise skeletal pose estimation using mmwave radars, IEEe Trans. Neural Netw. Learn. Syst. 34 (11) (2023) 8418–8429.
[9] S.-P. Lee, N.P. Kini, W.-H. Peng, C.-W. Ma, J.-N. Hwang, Hupr: a benchmark for human pose estimation using millimeter wave radar, in: Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision, 2023, pp. 5715–5724.
[10] A. Sengupta, F. Jin, R. Zhang, S. Cao, mm-pose: real-time human skeletal posture estimation using mmwave radars and cnns, IEEe Sens. J. 20 (17) (2020) 10032–10044.
[11] A. Sengupta, S. Cao, mmpose-nlp: a natural language processing approach to precise skeletal pose estimation using mmwave radars, IEEe Trans. Neural Netw. Learn. Syst. (2022).
[12] C. Shi, L. Lu, J. Liu, Y. Wang, Y. Chen, J. Yu, mpose: environment-and subject-agnostic 3d skeleton posture reconstruction leveraging a single mmwave device, Smart Health 23 (2022) 100228.
[13] C. Xie, D. Zhang, Z. Wu, C. Yu, Y. Hu, Y. Chen, Rpm: rf-based pose machines, IEEe Trans. Multimedia (2023).
[14] M. Zhao, Y. Tian, H. Zhao, M.A. Alsheikh, T. Li, R. Hristov, Z. Kabelac, D. Katabi, A. Torralba, Rf-based 3d skeletons, in: Proceedings of the 2018 Conference of the ACM Special Interest Group on Data Communication, 2018, pp. 267–281.
[15] C. Xie, D. Zhang, Z. Wu, C. Yu, Y. Hu, Y. Chen, Rpm 2.0: rf-based pose machines for multi-person 3d pose estimation, IEEE Transactions on Circuits and Systems for Video Technology, 2023.