

DRAFT

Comparative Testing of Radar Signal Representations when Sensing Through Materials

Nuwan T. Attygalle^[0000-0001-8498-1923]

Matjaž Kljun^[0000-0002-6988-3046]

Una Vuletić^[0009-0001-7260-735X]

Klen Čopič Pucihar^[0000-0002-7784-1356]

Abstract In recent years, gesture recognition using miniaturised radar sensing has attracted attention in both academia and industry. The ability to sense mid-air gestures with miniaturised radars through non-conductive materials with high spatial resolution opens up new opportunities for interacting with systems embedded with such radar technology. Examples include radars embedded in wearables, cars' dashboard, smart furniture, and other objects within smart environments. However, a deeper understanding of how different occluding materials impact gesture recognition performance is needed. Previous studies have primarily focused on evaluating only one type of radar signal representation, despite the fact that several other representations exist and proved effective. One way to better understand the aforementioned impact of the occluding materials on the radar signal is thus to compare different radar signal representations. To this end, this paper conducts a comparative testing of four (4) signal representations (In-phase and Quadrature (IQ) representations in time and frequency domain, Range-Angles and Range Doppler) in order to evaluate their robustness against signal distortions caused by occluding materials. The preliminary

Nuwan T. Attygalle

Faculty of Mathematics, Natural Sciences and Information Technologies, University of Primorska, Koper, Slovenia, e-mail: nuwan.attygalle@famnit.upr.si

Matjaž Kljun

Faculty of Mathematics, Natural Sciences and Information Technologies, University of Primorska, Koper, Slovenia, and

Department of Information Science, Stellenbosch University, Stellenbosch, South Africa, e-mail: matjaz.kljun@upr.si

Una Vuletić

Faculty of Mathematics, Natural Sciences and Information Technologies, University of Primorska, Koper, Slovenia e-mail: una.vuletic@famnit.upr.si

Klen Čopič Pucihar at University of Primorska, Faculty of Mathematics, Natural Sciences and Information Technologies, Koper, Slovenia, and

Faculty of Information Studies, Novo Mesto, Slovenia, and

Department of Information Science, Stellenbosch University, Stellenbosch, South Africa, e-mail: klen.copic@famnit.upr.si

results show that the performance increases with higher transmission coefficient and that compared to IQ, Range-Doppler and Range-Angles signal representations are significantly more robust against distortions caused by occluding materials.

1 Introduction

In recent years, gesture recognition with miniaturised radar sensing has received increased attention in both academia and industry. Two factors have fuelled this trend: the emergence of low-cost, low-power radar chips and breakthroughs in deep learning, which enable highly accurate interpretation of radar signals for interaction. Furthermore, these miniaturised radars have the potential to sense mid-air gestures through non-conductive materials with high spatial resolution [17, 9]. Enhancing such interaction through materials opens up new opportunities for creating a variety of interactive systems with such radars embedded. Embedding miniaturised radars in wearables, car padding and dashboards, furniture and other objects to create smart environments will facilitate better use of computational resources around us. Despite several advances in the field of radar sensing [11, 14, 15, 10, 16] there is still a lack of deeper understanding of how different occluding materials impact gesture recognition performance and how to improve it.

When radar sensing is used for interaction through materials, the signal needs to penetrate the covering material to reach the moving target (e.g. a hand) and reflect from the target back to the radar antenna through the same material. The machine learning classifier can use the signal captured by the radar antenna for gesture classification. However, the signal is affected by the electrical properties of the covering material and the wavelengths at which the sensor operates. In prior work, Čopič Pucihar et al. [17] found that the gesture detection performance of millimetre-wave (mm-wave) radar-on-chip sensor is closely correlated to the transmission coefficient (T_c) of occluding material. However, they trained their classifiers using only one type of radar signal representation—a sequence of Range-Doppler images.

Previous work showed that radar systems can effectively use several different *digital signal processing* techniques, resulting in a diverse set of radar signal representations that capture intricate information essential for later feature extraction and gesture recognition. These include time series of a raw signal [12, 13], range [18, 14], permittivity [14], micro-Doppler [2], Range-Doppler [7, 8, 19, 20, 21, 5, 1], Range-Angle of Arrival (AoA) [7, 22], and point cloud data [11, 10, 6].

In this paper, we perform comparative testing between four (4) different signal representations: (i) In-phase and Quadrature (IQ) representation in time domain, (ii) IQ representation in frequency domain (IQ-FFT-ABS), (iii) Range-Angles and (iv) Range Doppler. We investigate how robust are these signal representations against distortions caused by occluding materials. The results indicate that performance improves with higher transmission coefficients, aligning with observations from prior work [17]. Additionally, our preliminary evidence suggest that, compared to

IQ signal representations, Range-Doppler and Range-Angles representations are more robust against distortions caused by occluding materials.

2 Gesture recognition experiments

2.1 Gesture set, experiment system and data collection

We captured two datasets: *No-material* dataset with more than 2200 gesture recordings where no material occluded the sensor and *Material* dataset with more than 32000 gesture recordings where 14 different material were placed between the sensor and the hand performing gestures. We captured recordings of 24 different gestures (Table 2), performed by participants whilst sitting at the table. The experiment system to capture data was positioned on the table and consisted of the Texas Instruments IWR6843ISK Radar sensor and Leap Motion sensor (Fig. 1). Further details about the experiment system used are available in [4].

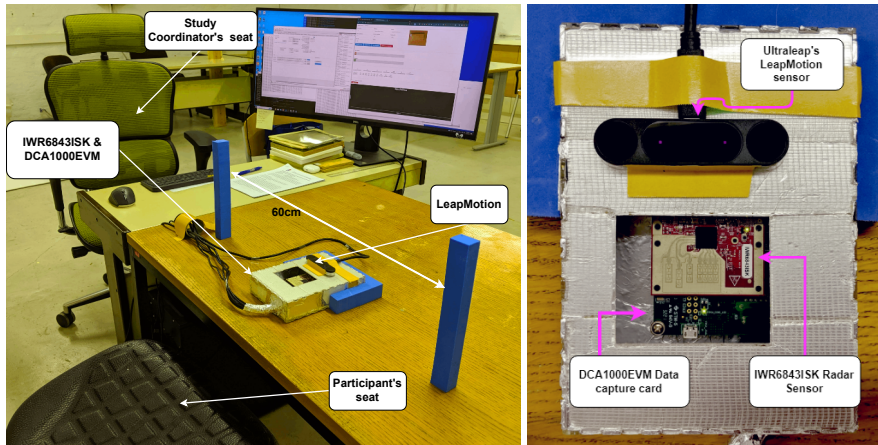


Fig. 1 Experiment system: Left—Participant sits at the desk whilst performing gestures above the capture system. Different materials are placed on the open space where the radar sensor (Texas Instruments IWR6843ISK) is located. Participants are asked to not move closer than 25 centimetres, which is indicated to them by the two vertical blue spacers on the desk. The screen indicates the state of recording, such as time, gesture and material that is being recorded. Right—Capture system with Texas Instruments IWR6843ISK radar sensor and Leap Motion sensors. Materials are placed on the open space configured to uphold the radar’s Field of View at 140 degrees azimuth and 120 degrees elevation. The radar sensor and data capture card are housed within the radome, with the Leap Motion sensor positioned on its top.

The gesture selection and sensor placement were based on a literature review, which identified horizontal sensor placement as the dominant configuration. For gesture selection, we identified all gestures proposed in the existing body of literature

and selected a subset of 24 gestures based on a heuristic approach. This approach prioritised gestures with movement trajectories that are favourable for our sensing system and sensor placement. Concretely, we selected gestures where movement trajectories induce changes in range and azimuth angle. Given our sensor’s antenna configuration and placement, our sensor has higher resolution in range and azimuth angle compared to elevation angle dimension. Therefore, we avoided gestures that rely on changes in elevation to create more favourable sensing conditions.

We captured data from 19 participants, where each participant made 5 gesture repetitions for each of the 24 gestures though 13 different materials (Table 1) and once without any material. Each participant thus performed $5 \times 24 \times 14$ or 1680 gestures. The data was collected over 4-6 sessions. Each session was limited to 1.5 hour. As the speed at which users managed to record gestures varied, the number of required sessions varied. After collecting data for 2 materials (30-40 minutes of capture time required), there was a compulsory 5-minute break. Participants were continually reminded they can take a break at any point during the capture if they feel tired. Materials were carefully selected to cover the whole range of transmission coefficients (Table 2). We used transmission coefficients measurements published in [17].

Despite having 24 gestures in the dataset, the evaluation reported in this paper uses only 10 gestures (see underlined gestures in Table 2). This decision was made because we were unable to build a high performance classifier for all 24 gestures in the *No-material* dataset. A poorly performing classifier cannot be used to investigate the robustness of these signal representations against distortions caused by occluding materials, as classification errors in the baseline condition (*No-material* dataset) introduced excessive noise into the signal. To address this issue, we used a subset of 10 gestures, selected based on confusion matrix results, ensuring we chose gestures with good scores.

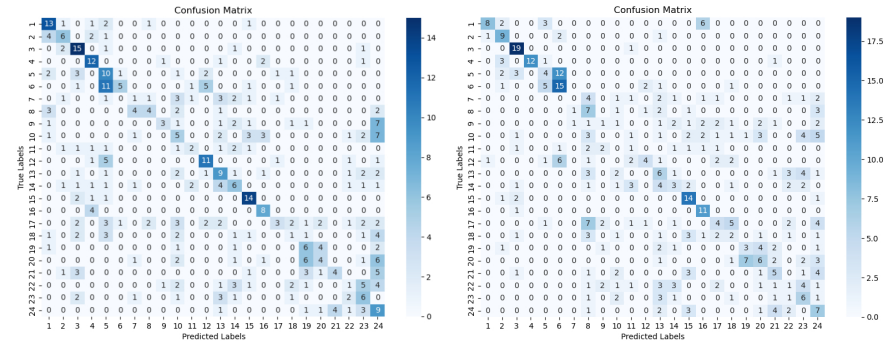


Fig. 2 Confusion matrices for Range-Angles (left) and Range-Doppler (right) based classifiers evaluated on *No-material* dataset. We selected 10 gestures favourable for our recogniser: 1, 2, 3, 4, 5, 11, 12, 14, 15 and 23 (underlined in Table 2)







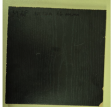



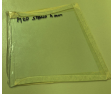


| ID | Name | Thickness | Tc | Image | ID | Name | Thickness | Tc | Image |
|----|----------------------|-----------|------|--|----|-----------------------------------|-----------|------|---|
| 12 | Drywall Plate | 12 | 0.85 |  | 30 | Acrylic Plexiglass | 10 | 0.91 |  |
| 15 | Polyethylene Foam | 15 | 0.94 |  | 37 | Paper 80G/M2 (100 sheets) | 9 | 0.63 |  |
| 17 | Styrofoam | 10 | 0.98 |  | 47 | Oriented Strand Board (OSB) | 10 | 0.54 |  |
| 18 | Chipboard | 16 | 0.31 |  | 57 | Larix Decidua | 10 | 0.66 |  |
| 19 | Ceremic | 6 | 0.82 |  | 61 | Quercus Robur Petraea | 10 | 0.33 |  |
| 20 | Glass | 4 | 0.92 |  | 66 | Ethylene-vinyl acetate (EVA foam) | 10 | 0.93 |  |
| 27 | Rubber panels tartan | 15 | 0.04 |  | | | | | |

Table 1 *Material* dataset with 13 materials selected. Materials were selected from the catalogue of materials published in [17]. For consistency purposes, we use the same IDs.

2.2 Data pre-processing

Each gesture recording is 1.2 seconds long. The radar sensor was configured to capture data at 100 frames per second. We used an antenna array configuration of 4Rx and 3Tx antennas in Multiple Input Multiple Output (MIMO) configuration with Time Domain Multiplexing (TDM). Chirps were generated using Frequency Modulated Continuous Waveform (FMCW) with 96 chirps per frame operating within the frequency range of approx 60-64GHz. Detailed radar sensor configurations are provided in appendix.

For comparative testing, we pre-processed four different radar signal representations. Each frame is represented as: IQ and IQ-FFT-ABS radar cube, or as a set of Range-Angle and Range-Doppler images. The data was pre-processed using Dis-ProTool. For more details regarding the digital signal processing pipelines, see [3].

DRAFT

| ID | Name | Time | ID | Name | Time |
|----|--------------------------|------|----|---------------------------------|------|
| 1 | <u>Swipe left</u> | | 13 | Hand opening | |
| 2 | <u>Swipe right</u> | | 14 | <u>Clenching (Hand closing)</u> | |
| 3 | <u>Swipe up</u> | | 15 | <u>Pull</u> | |
| 4 | <u>Swipe down</u> | | 16 | Push | |
| 5 | <u>Left cross</u> | | 17 | Clockwise circle | |
| 6 | <u>Right cross</u> | | 18 | Counter clockwise circle | |
| 7 | <u>Pinch index</u> | | 19 | Palm hold | |
| 8 | <u>Star</u> | | 20 | Fist hold | |
| 9 | <u>Finger slide up</u> | | 21 | Thumbs up | |
| 10 | <u>Finger slide down</u> | | 22 | V | |
| 11 | <u>Palm tilt</u> | | 23 | <u>L</u> | |
| 12 | <u>Wiggle 3 times</u> | | 24 | Pointing index | |

Table 2 Gesture set. Ten underlined gestures were selected for the evaluation based on the confusion matrix.

2.3 Model architecture

We used Conv3d layer based classifier. However, hyperparameters were optimised for each signal representation. For IQ signal, the input is in the format of (120, 96, 32, 4), for Range-Doppler (500, 32, 32, 1), and for Range-Angles (400, 32, 32, 1)

For IQ signal representations, each 1.2 second gesture recorded at 100 Hz is represented in the format of 120 frames \times 96 chirps \times 4 antennas \times 256 fast time samples. Radar was configured to sense the maximum distance of 10 meters; however, we performed micro gestures within 1 m range. For this reason, we reduced the fast time samples (range bins) to 32 bins.

For Range-Doppler signal representation, each 1.2 second gesture recorded at 100 Hz is represented as a set of 32×32 Range-Doppler images. Out of 12 available virtual antennas we chose 4 (i.e. 4, 5, 10 and 11). As a result, each gesture contains 480 Range-Doppler images.

For Range-Angles signal representation, each 1.2 second gesture recorded at 100 Hz is represented as a set of 32×32 Range-Angle images. One set along the azimuth and another along the elevation angles. This resulted, $2 \times 120 = 240$ (32×32) Range-Angle images.

The input data was then fed into the first convolutional layer (Conv3D) This layer is immediately followed by a 3D max pooling layer. This is repeated 4 times. Each time we used different configuration of kernel size and number of filters (see appendix for details). The output from these layers is then flattened and passed through a fully connected layer, leading to a final dense layer of 10 neurons with softmax activation (see appendix for details about number of neurons).

2.4 Data partitioning and training

All models were trained on *No-material* dataset only and we consider 10 out of 24 gestures (see Table 2) as explained earlier. The dataset was partitioned as follows: 70%, 20% and 10%, for training, validation and testing respectively. In case of *Material* dataset, no partitioning was done. The whole dataset can be considered as unseen data and is used to investigate how robust are different signal representations against distortions caused by occluding materials.

The models were trained using categorical cross-entropy as the loss function and are optimised with the Adam optimiser, featuring an adjusting learning rate of 0.001 and a weight decay of 0.0002. Training was done with 200 epochs and patience of 50.

3 Results

We performed model evaluation on two datasets. *No-material* dataset to see how good are models in general, and *material* dataset to investigate how robust are different signal representations against distortions caused by occluding materials.

3.1 Model evaluation on *No-material* dataset

The results in Table 3 show accuracy, Area Under the Curve (AUC) and $Performance = (ACC + AUC)/2$ scores for *No-material* dataset. The training and evaluation processes were only conducted once, so no statistical testing comparing model performance is possible. The highest performance was achieved by model trained with IQ-FFT-ABS data representation. The limitation of this testing is a very small test partition (only 10% of data was used for testing). This decision was made due to small size of *No-material* dataset.

| Data representation | ACC | AUC | (ACC+AUC)/2 |
|---------------------|------|------|-------------|
| IQ-FFT | 0.85 | 0.99 | 0.92 |
| IQ-FFT-ABS | 0.91 | 1.00 | 0.95 |
| Range-Angle | 0.86 | 0.99 | 0.92 |
| Range-Doppler | 0.80 | 0.97 | 0.89 |

Table 3 Classification performance on *No-material* dataset.

3.2 Model Evaluation on *Material* dataset

The results in Fig. 3 show how classification performance of 4 different models changes with transmission coefficient. In total, the graph represents performance through 13 different materials (Table 1). We can observe performance increase as transmission coefficient increases. This is expected and in line with what was observed within prior work [17].

The results in Fig. 4 compare overall performance of 4 different models on *Material* dataset (left) and performance drop, defined as the difference between performance in *No-material* and *Material* datasets. The worst performing model is IQ model, which also experiences the highest performance drop. Statistical testing was done using non-parametric tests (Friedman with Dubin-Conover pairwise comparison) as the data was not normally distributed. The results show that both IQ data representation performed significantly worst compared to Range-Doppler and Range-Angles. This suggests that Range-Doppler and Range-Angles signal repre-

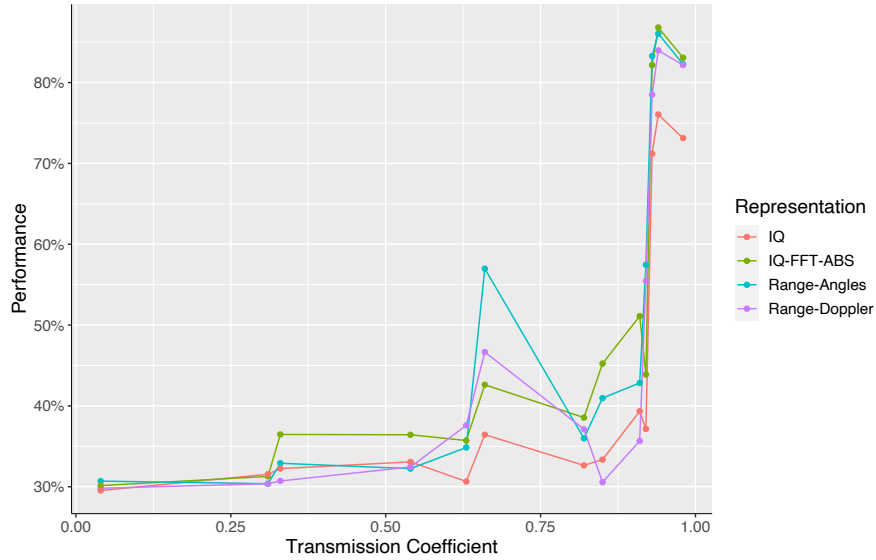


Fig. 3 Recogniser performance in relation to transmission coefficient of materials.

representations are more robust against distortions caused by occluding materials. In the case of performance, the minimum effect size where significance was reported is $d = 0.11$. This resulted in statistical power of 0.57. In the case of performance drop, minimum effect size where significance was reported is $d = 0.19$. This resulted in statistical power of 0.68.

4 Discussion and Future Work

4.1 Is choice of signal representation important when sensing through materials?

Several systems are focused on radar-based human-computer interaction. These systems may differ in *digital signal processing* techniques used, resulting in a diverse set of radar signal representations that capture intricate information essential for later feature extraction and gesture recognition. For example, reliable interaction systems in the literature use various signal representations, including raw signal [12, 13], range [18, 14], permittivity [14], micro-Doppler [2], Range-Doppler [7, 8, 19, 20, 21, 5, 1], Range-Angle of Arrival (AoA) [7, 22], and point cloud data [11, 10, 6]. In this work, we experimented with four (4) different signal representations: IQ, IQ-FFT-ABS, Range-Angles and Range-Doppler. The results showed significant difference among these representations, particularly between IQ and more complex

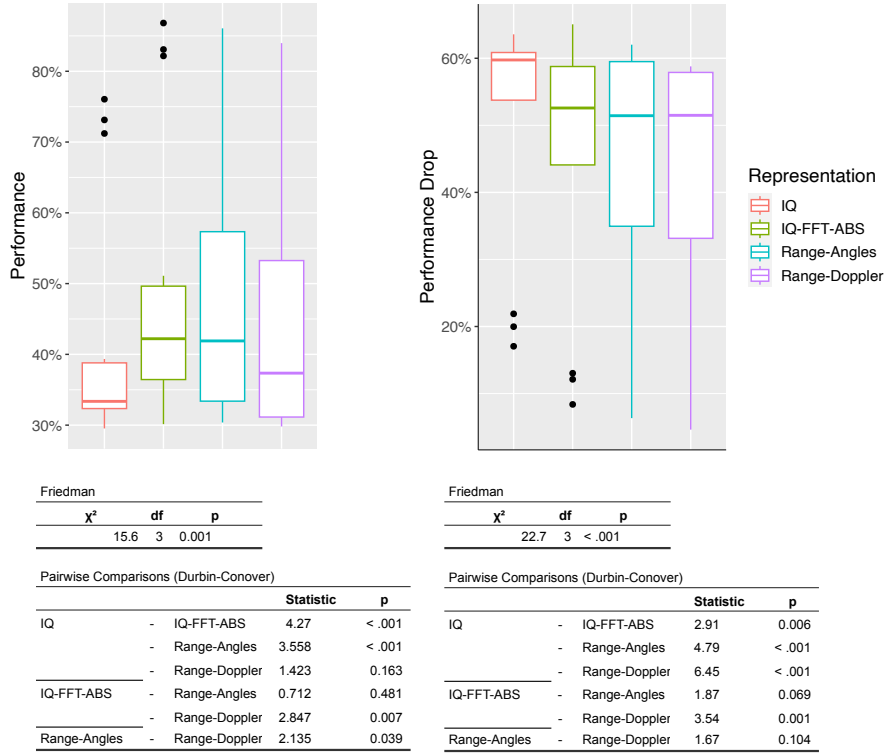


Fig. 4 Comparison of Overall performance and Performance drop for 4 different signal representations. Top—box-lots of Performance and Performance Drop. Bottom—Results of Friedman test with Durbin-Conover pairwise comparison for Performance and Performance Drop. NOTE: $Performance = (Accuracy + AUC)/2$ and $Performance_{Drop} = Performance_{NoMaterial} - Performance_{Material}$

Range-Angles and Range-Doppler representations. The latter proved to be more robust against distortions caused by occluding materials.

Having said this, it is important to highlight several limitations of this comparative testing. While our focus is on signal representation, we also use different model architectures, which might influence the obtained results. This decision was necessary because tuning network architecture for the task at hand is crucially affected by the type of input provided to the network. This interdependence makes it practically impossible to decouple network architecture from signal representation. To mitigate this effect, we used the same type of deep neural network and only change hyperparameters, such as: kernel size, number of filters, number of neurones, learning rate and decay. Additionally, We ensured that the model architectures are optimised to achieve comparable classification accuracy on baseline condition (e.g. *No-material* dataset).

Another potential confounding variable is the method used for handling data from a virtual antenna array. For IQ data representations, the signal was not decoupled using MIMO TDM. In contrast, for the Range-Doppler representation, we used data from only 4 of 12 virtual antennas, which are represented in 4 Range-Doppler images. Meanwhile, the data in the Range-Angles representation utilised all virtual antenna pairs, but was represented only in 2 Range-Angles images. Further experimentation is needed to verify the existence and significance of these potential confounding variables.

4.2 Limitations and future work

In this work, we ran comparative testing for only 4 different signal representations. There are still several other promising signal representations methods that should be tested in the future, such as for example: point cloud, micro-Doppler, combination of range and permittivity as done in [14] or [16].

The other limitation is the size of *No-material* dataset. K-fold cross validation should be done in the future, or we should work on increasing the size of *No-material* dataset. This would also allow us to do proper model evaluation for baseline condition (e.g. *No-material* dataset).

5 Conclusion

In this paper, we performed cooperative testing between four (4) different signal representations: In-phase and Quadrature (IQ) representations in time and frequency domain, Range-Angles and Range Doppler. We investigated how robust are these signal representations against distortions caused by occluding materials. The results show that performance increases with higher transmission coefficient. This is expected and in line with what was observed within prior work [17]. We also found preliminary evidence that compared to IQ signal representations, Range-Doppler and Range-Angles representations are more robust against distortions caused by occluding materials. This pioneering study introduces the first example of cooperative testing, evaluating different signal representations in scenarios where sensors are occluded by materials. It offers substantial potential for future research, which makes it a valuable contribution to the community.

Appendix A: Radar sensor configuration

Appendix B: Detailed model architectures



Fig. 5 Model architectures for Range-Angles —left, Range-Doppler —centre and IQ —right

References

1. M. Altmann, P. Ott, N. C. Stache, and C. Waldschmidt. Multi-modal cross learning for an fmcw radar assisted by thermal and rgb cameras to monitor gestures and cooking processes. *IEEE Access*, 9:22295–22303, 2021.
2. M. G. Amin, Z. Zeng, and T. Shan. Hand gesture recognition based on radar micro-doppler signature envelopes. In *2019 IEEE Radar Conference*, pages 1–6, 2019.
3. N. T. Attygalle, M. Kljun, U. Vuletić, and K. Čopič Pucihar. Disipro: Digital signal processing tool for radar-based human-computer interaction. *TBD*, 1(1):1–10, 2024.

4. N. T. Attygalle, U. Vuletic, M. Kljun, and K. Č. Pucihar. Towards hand gesture recognition prototype using the iwr6843isk radar sensor and leap motion. In *Proceedings of the 8th Human-Computer Interaction Slovenia (HCI SI) Conference 2023*, pages 78–88, 2024.
5. E. Hayashi, J. Lien, N. Gillian, L. Giusti, D. Weber, J. Yamanaka, L. Bedal, and I. Poupyrev. Radarnet: Efficient gesture recognition technique utilizing a miniature radar sensor. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*. Association for Computing Machinery, 2021.
6. S. Hazra, H. Feng, G. N. Kiprit, M. Stephan, L. Servadei, R. Wille, R. Weigel, and A. Santra. Cross-modal learning of graph representations using radar point cloud for long-range gesture recognition. In *2022 IEEE 12th Sensor Array and Multichannel Signal Processing Workshop*, pages 350–354, 2022.
7. S. Hazra and A. Santra. Radar gesture recognition system in presence of interference using self-attention neural network. In *2019 18th IEEE International Conference On Machine Learning And Applications*, pages 1409–1414, 2019.
8. H. R. Lee, J. Park, and Y.-J. Suh. Improving classification accuracy of hand gesture recognition based on 60 ghz fmcw radar with deep learning domain adaptation. *Electronics*, 9(12), 2020.
9. L. A. Leiva, M. Kljun, C. Sandor, and K. Copic Pucihar. The wearable radar: Sensing gestures through fabrics. In *22nd International Conference on Human-Computer Interaction with Mobile Devices and Services*. Association for Computing Machinery, 2021.
10. H. Liu, Y. Wang, A. Zhou, H. He, W. Wang, K. Wang, P. Pan, Y. Lu, L. Liu, and H. Ma. Real-time arm gesture recognition in smart home scenarios via millimeter wave sensing. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.*, 4(4), 2020.
11. S. Palipana, D. Salami, L. A. Leiva, and S. Sigg. Pantomime: Mid-air gesture recognition with sparse millimeter-wave radar point clouds. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.*, 5(1), 2021.
12. J. Park, J. Jang, G. Lee, H. Koh, C. Kim, and T. W. Kim. A time domain artificial intelligence radar system using 33-ghz direct sampling for hand gesture recognition. *IEEE Journal of Solid-State Circuits*, 55(4):879–888, 2020.
13. T. Sakamoto, X. Gao, E. Yavari, A. Rahman, O. Boric-Lubecke, and V. M. Lubecke. Hand gesture recognition using a radar echo i-q plot and a convolutional neural network. *IEEE Sensors Letters*, 2(3):1–4, 2018.
14. A. Sluÿters, S. Lambot, and J. Vanderdonckt. Hand gesture recognition for an off-the-shelf radar by electromagnetic modeling and inversion. In *Proceedings of the 27th International Conference on Intelligent User Interfaces*, page 506–522. Association for Computing Machinery, 2022.
15. A. Sluÿters, S. Lambot, J. Vanderdonckt, and R.-D. Vatavu. Radarsense: Accurate recognition of mid-air hand gestures with radar sensing and few training examples. *ACM Transactions on Interactive Intelligent Systems*, 2023.
16. A. Sluÿters, S. Lambot, J. Vanderdonckt, and S. Villarreal-Narvaez. Analysis of user-defined radar-based hand gestures sensed through multiple materials. *IEEE Access*, 12:27895–27917, 2024.
17. K. Čopič Pucihar, N. T. Attygalle, M. Kljun, C. Sandor, and L. A. Leiva. Solids on soli: Millimetre-wave radar sensing through materials. *Association for Computing Machinery*, 6(EICS), 2022.
18. L. Wang, Z. Cui, Z. Cao, S. Xu, and R. Min. Fine-grained gesture recognition based on high resolution range profiles of terahertz radar. In *IGARSS 2019 - 2019 IEEE International Geoscience and Remote Sensing Symposium*, pages 1470–1473, 2019.
19. P. Wang, J. Lin, F. Wang, J. Xiu, Y. Lin, N. Yan, and H. Xu. A gesture air-writing tracking method that uses 24 ghz simo radar soc. *IEEE Access*, 8:152728–152741, 2020.
20. S. Wang, J. Song, J. Lien, I. Poupyrev, and O. Hilliges. Interacting with soli: Exploring fine-grained dynamic gesture recognition in the radio-frequency spectrum. In *Proceedings of the 29th Annual Symposium on User Interface Software and Technology*, pages 851–860. Association for Computing Machinery, 2016.

DRAFT

14

Nuwan T. Attygalle Matjaž Kljun Una Vuletić Klen Čopič Pucihar

21. G. Zhang, S. Lan, K. Zhang, and L. Ye. Temporal-range-doppler features interpretation and recognition of hand gestures using mmw fmcw radar sensors. In *2020 14th European Conference on Antennas and Propagation*, pages 1–4, 2020.
22. L. Zheng, J. Bai, X. Zhu, L. Huang, C. Shan, Q. Wu, and L. Zhang. Dynamic hand gesture recognition in in-vehicle environment based on fmcw radar and transformer. *Sensors*, 21(19):6368, 2021.