The Wearable Radar: Sensing Gestures Through Fabrics

LUIS A. LEIVA, Aalto University, Finland
MATJAŽ KLJUN, University of Primorska, Slovenia
CHRISTIAN SANDOR, School of Creative Media, City University of Hong Kong, Hong Kong SAR
KLEN ČOPIČ PUCIHAR, University of Primorska, Slovenia

Recently, millimeter-wave radar-on-chip sensors such as Google Soli have become readily available in the mobile ecosystem. We envision such radar technology to be integrated into wearables to enable gesture-based interaction possibilities for users 'on the go', e.g. to control various devices such as phone, car infotainment system, etc. even when the sensor is occluded by some material such as fabrics. Towards achieving this vision, we developed a hybrid CNN+LSTM deep learning model, and conducted a systematic study investigating mid-air gesture recognition performance when the radar sensor was covered by three different fabrics (leather, wool, and cotton). We show that, when trained on no occluding material, the model performed worse than if trained with each of the three fabrics; however, this is only valid in the small data regime (N=20). When trained with large samples (N=200) on no occluding material, the model achieved remarkable performance also when the sensor was covered by each of the fabrics (95% avg. accuracy, 99% AUC). Our results show that sensing mid-air gestures through fabrics is both feasible and ready for practical applications, since it is not necessary to train a dedicated model for each type of fabric available in the market. We also contribute a repeatable procedure to systematically test mid-air gestures with radar technology, enabled by an experimental platform that we release with this paper.

CCS Concepts: • **Human-centered computing** → *Interaction design process and methods.*

Additional Key Words and Phrases: Radar; Wearables; Fabrics; Soli; Gestures; Deep Learning

ACM Reference Format:
Luis A. Leiva, Matjaž Kljun, Christian Sandor, and Klen Čopič Pucihar. 2020. The Wearable Radar: Sensing Gestures Through Fabrics. In *22nd International Conference on Human-Computer Interaction with Mobile Devices and Services (MobileHCI '20 Extended Abstracts), October 5–8, 2020, Oldenburg, Germany*. ACM, New York, NY, USA, 7 pages. https://doi.org/10.1145/3406324.3410720

1 INTRODUCTION AND RELATED WORK

Gesture interaction is an active research topic in Human-Computer Interaction (HCI), with a history dating back to the 1960s with Sutherland's Sketchpad project [13] and a far reaching vision of the Ultimate Display essay [14]. Gestures can be categorised into two broad groups: (1) *mid-air* gestures, used for example in consumer electronics such as gaming consoles, and (2) *stroke* gestures, used for example in touch-capable devices such as smartphones. In this paper, we focus on mid-air gesture recognition.

Radar technology has emerged as a popular medium for mid-air gesture recognition, as it offers 3D spatial information, is robust to weather conditions, does not require lighting, and can penetrate non-metallic surfaces. The Magic Carpet [10] was among the first systems using radar sensing to detect body motion, though it required excessive space and a complex setup. More recently, the miniaturised radar-on-chip millimeter-wave low-cost devices such as Google Soli [6] have democratised radar sensing in HCI and are being integrated in modern smartphones like the Google Pixel 4.

The potential of radar sensing in HCI has been demonstrated in detecting subtle, nonrigid motion gestures [6, 16], mostly articulated with hands and fingers: e.g. rubbing, pinching, or swiping. Recent research includes, among others, radar sensing for augmented reality [4], interaction with everyday objects [15], music [3, 12], activity recognition [2], and through-wall tracking [1, 18]. A radar sensor can as well distinguish various materials when placed on top of it [7, 17]. What is missing, however, is an investigation of gesture recognition performance through various materials present on and around us. In this paper, we bridge this gap by analysing mid-air gesture recognition through fabrics.

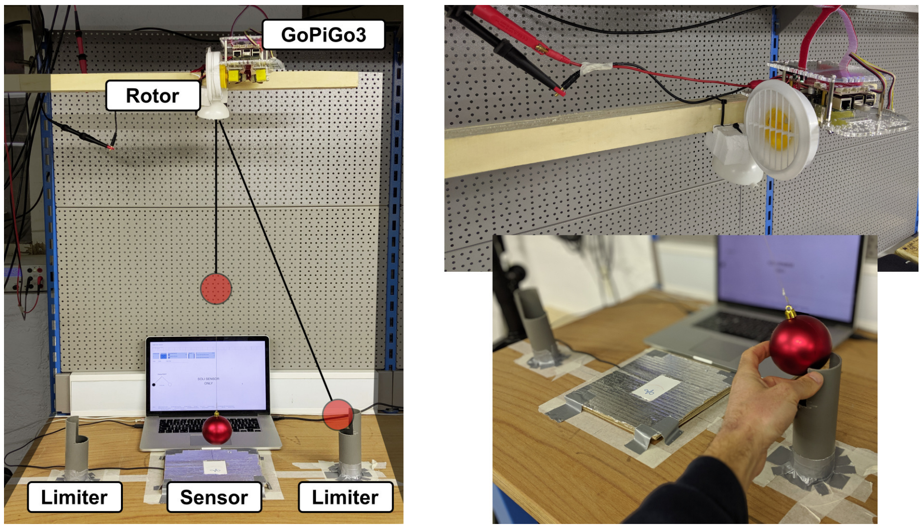

Fig. 1. Our experimental platform setup consists of: (i) an appropriated GoPiGo 3 robot to control movements in vertical direction, (ii) two limiters on the right and left side of the sensor for controlling swing movements, (iii) a frame shielded with aluminium foil to prevent the radar signal escaping around the analysed material, and (iv) the radar sensor connected to a PC for data logging.

To the best of our knowledge, we are the first to analyse how various materials covering the radar sensor influence gesture recognition performance. This is not only important for devices that may have a miniaturised radar sensor incorporated, but also for integrated sensors in various objects on and around us that we will interact with while 'on the go'. We envision such sensors integrated into everyday objects that will enable mid-air gestures interaction with the environment. This includes, for example, integration of such a sensor (i) with the bicycle's handlebar to control various functions of a phone that is stored away in our backpack; (ii) with the car's dashboard, seats, or doors to control the infotainment system and other car functions, replacing the need to touch LCD displays that now require our full sight and can thus influence ours and others' safety on the road; (iii) into furniture such as the arm rest area of an armchair to control the TV and other home systems; and (iv) into the clothing accessories such as ties, buttons, and jewellery to control various appliances and environment (e.g. the amount of light in the house). The presented use cases are just an illustration of the numerous interaction possibilities that radar sensing allows via mid-air gestures.

Our work contributes to the current body of knowledge in two ways. We present a novel and repeatable procedure to systematically test mid-air gestures with radar technology, enabled by the experimental platform that we release with this paper. We also report on preliminary results on gesture recognition performance through three fabrics: leather, wool, and cotton. The results are important for radar-enabled wearable technologies, as the internal structure of the fabrics and intertwining of the threads may impact gesture recognition. Our work opens up a new research avenue, where potentially any clothing accessory, jewellery, or clothed object can become a wearable and/or interactive device when combined with a miniaturised radar sensor.

2 EXPERIMENTAL SETUP AND DATA COLLECTION

We built a mechanical platform that allows us to accurately replicate gestures by modifying the GoPiGo 3 robot, visible in Figure 1. The robot has a ball attached with a nylon string. The platform is designed to generate two types of movements: pendulum-like (e.g. swing or swipe movements) and vertical movements along the Z axis. A pendulum movement is generated by manually releasing the ball from a limiter position, whereas the vertical movements are automatically generated by the robot. The radar sensor is placed in a frame shielded with an aluminium foil to prevent the radar signal escaping around the analysed material. The frame is placed on the table with a 5 cm gap between the sensor and the ball when the latter is in its lowest position. Various materials can then be placed on the sensor to test gesture recognition performance.

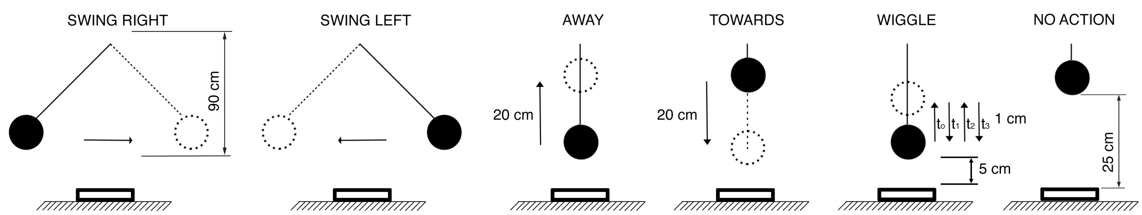

Fig. 2. Analysed gesture set. The filled and dashed circles denote, respectively, the initial and final position of the ball.

We collected 20 trials of 6 distinct gestures (see Figure 2) for each of the 3 materials under study (360 gesture recordings in total) and 200 trials for the no material condition (1200 gesture recordings in total). Our gesture set is informed by previous work [16]. The radar sensor recorded each gesture articulation as a sequence of frames, where each frame is a 32×32 px range Doppler image. We configured the Doppler range of the sensor to operate in the [-2,0] dB range and disabled the built-in adaptive clutter filter. Since Soli computes range Doppler images for each of the 4 sensor antennas, we averaged them to ensure a robust frame representation. Further, images were grayscaled and sequences were resampled to 100 Hz and padded to 300 timesteps, which is large enough to accommodate for arbitrary gesture articulations. As a reference, each of the recorded gestures took on average 1.5 seconds, or 150 timesteps.

3 MATERIALS

We used three different fabrics, as seen in Figure 3: genuine leather, wool, and cotton. The surface area of all three fabrics was 20×20 cm and the thickness was under 2 mm, wool being the thickest (just under 2 mm) and cotton the thinnest (around 0.5 mm). Notice that measuring the precise thickness of the fabric is difficult as it is easily squeezed with a caliper, therefore we report on raw measurements.

4 DEEP LEARNING MODEL

We built a hybrid deep CNN+LSTM (convolutional neural network + long short-term memory) model, inspired by previous work on human activity recognition [5, 6, 8, 9], which is depicted in Figure 4. Our model processes each frame (Doppler image) by means of a stack of $32 \times 64 \times 128$ convolutional layers with 3×3 filters to capture spatial

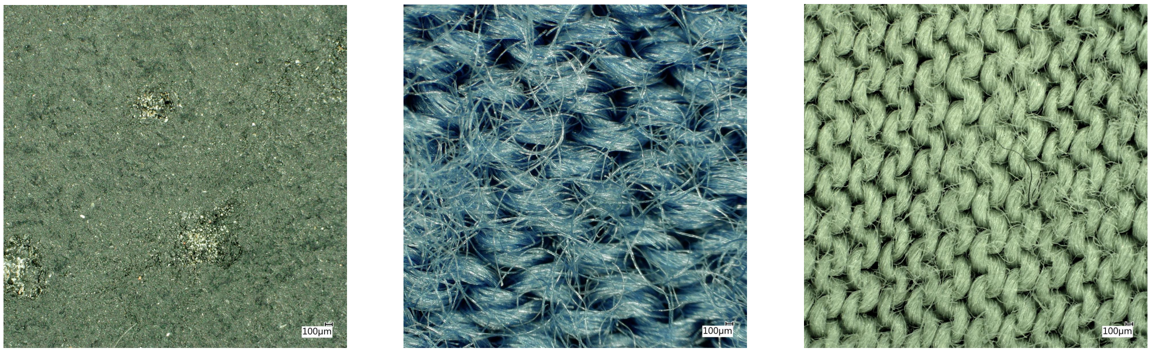

Fig. 3. We experimented with three different fabrics, shown here under a microscope at 50x magnification. From left to right: leather, wool, and cotton.

information, which is then processed in a recurrent fashion by means of an LSTM layer to capture temporal information, and eventually classified with a softmax layer.

Each convolutional layer automatically extracts feature maps from the input frames that are further processed by max pooling and spatial dropout layers. The max pooling layers (pool size of 2) downsample the feature maps by taking the largest value of the map patches, resulting in a local translation invariance.

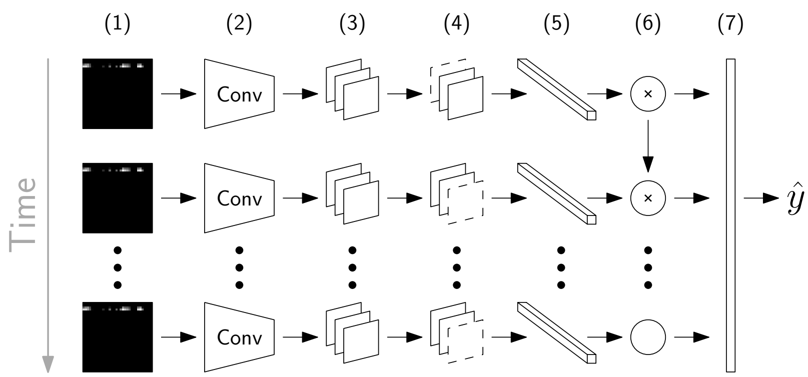

Fig. 4. Model architecture. Range Doppler images (1) are processed with a CNN (2) that extracts feature maps followed by max pooling (3) and spatial dropout (4) layers. Then, a fully connected layer (5) creates the feature vectors for the recurrent LSTM layer (6) and finally a softmax layer (7) outputs the model prediction ($\hat{y}$).

Crucially, the spatial dropout layer (drop rate of 0.25) removes entire feature maps at random, instead of individual neurons (as it happens in regular dropout layers), which promotes independence between feature maps, thus improving performance. Finally, the LSTM layer (embedding size of 128) uses the computed feature maps to classify the frame sequences in a recurrent fashion, i.e., the information embedded at each frame depends on the previous frame. The LSTM layer uses both a dropout rate and a recurrent dropout rate of 0.25. The softmax layer has dimensionality of 6, since we have 6 gesture classes.

We created random splits comprising 60% of the data for model training, 10% for model validation, and the remaining 30% for model testing. The test data are held out as a separate partition, which simulates unseen data. We also used a simple data augmentation technique: each training sequence was randomly left- and right-trimmed to 10% of the original sequence length, dropping both initial and ending frames.

We used the Adam optimiser with learning rate $\eta = 0.0005$ and decay rates $\beta_1 = 0.9, \beta_2 = 0.999$. The model was trained in batches of 10 sequences each using categorical cross-entropy as loss function. The maximum number of epochs was set to 200, but we also set an early stopping criteria of 20 epochs; i.e. training stopped if the validation loss did not improve after 20 consecutive epochs, and the best model weights obtained up to that moment were retained.

5 RESULTS

We trained our model on several fabrics combinations and sample sizes, including the 'no material' condition where there was no fabric covering the sensor. Together with classification accuracy, we report the F_1 score (weighted mean of precision and recall) and the area under the ROC curve (AUC), which informs about the discriminative power of any classifier [11]. Figure 5 shows the results of our experiments, which we discuss as follows:

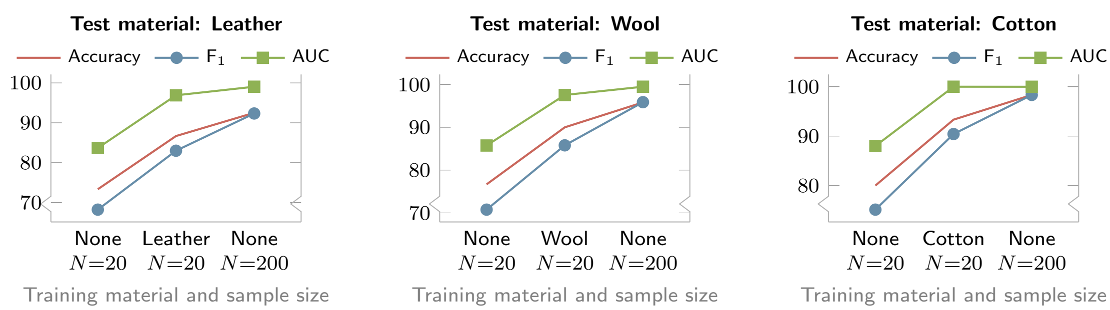

Fig. 5. The effect of train and test materials on gesture recognition. None means 'no material covering the sensor'.

- Training with small samples ($N = 20$) on no material is not as beneficial as training on any of the tested fabrics. This was expected, since the model had no knowledge about the impact that each fabric would have on the acquired sensor signal.
- Training with large samples ($N = 200$) on no material improves recognition performance and in fact achieves better results than training on small samples with any of the tested fabrics. This was unexpected, since the radar signal is not attenuated by the fabric covering the sensor at training time. However, this can be explained by the fact that the model has more training data to derive a robust representation of each gesture articulation.
- Some fabrics allow for better gesture recognition than others. For example, wool and cotton surpassed 90% of classification accuracy when the model was trained on both materials with small samples, and achieved 95% and 98% accuracy when trained on no material with large samples. As a reference, a random classifier achieved 1/6=16% accuracy on this task.
- Leather fabrics attenuate the signal slightly more than the other two fabrics tested, lowering classification accuracy by 3 and 7 points and AUC by 2 and 5 points, respectively. This can be explained by the nature of the leather fibers and the way they are intertwined, which have a noticeable impact on gesture recognition performance.

6 DISCUSSION AND LIMITATIONS

Overall, some degradation of the radar signal is expected when covering the sensor with fabrics. However, our deep learning model trained with no material with large samples is robust enough to sense gestures through different fabrics. This result, we believe, is of key importance since training on every fabric or textile available is impractical. However, with our experimental platform it is possible to acquire training data and create a very competitive classifier. We conclude therefore that sensing mid-air gestures through fabrics is both feasible and ready for practical applications. As a matter of fact, when the model is both trained and tested on no material, it is extremely discriminative (99% AUC) and achieves 99% of accuracy and F_1 score.

One limitation of our model is that is has been trained on machine-generated movements with a plastic ball, in order to make our experimental platform as replicable as possible. Therefore, in order to test the generalisation capabilities of our model, an evaluation with real users is needed. Another limitation worth mentioning is that the textiles we have tested were laying flat on top of the radar sensor. In a real-world setup, however, textiles would have folds and the sensor device might move.

In the context of wearable technology, our work enables novel interaction possibilities with a technologically supported mobile world. We have provided some use case examples in section 1, such as placing the sensor in accessories that are commonly covered with clothes, as well as in clothed objects in our environment that could be used to control various appliances.

7 CONCLUSION AND FUTURE WORK

We have presented a systematic study of mid-air gesture recognition through fabrics using millimeter-wave radar technology and a repeatable procedure enabled by our experimental platform. We have tested how three different fabrics commonly used in clothes, clothing furniture (e.g. car interior, doors, seats), or that could cover accessories (e.g. tie clips, buttons, jewellery), may impact mid-air gesture recognition performance.

We are currently investigating the practicability of sensing through various materials that are common in our environment and used in everyday objects, through systematic analyses like the one presented in this paper. In addition, we plan to investigate if other data representations besides Doppler images are also feasible, such as Soli DSP core features (e.g. energy or acceleration). Our model and instructions to build our experimental platform are publicly available at https://dist.famnit.upr.si/en/HICUP/hicup-projects/google-soli.

ACKNOWLEDGMENTS

The authors acknowledge the European Commission for funding the InnoRenew CoE project (Grant Agreement 739574) under the Horizon2020 Widespread-Teaming program and the Republic of Slovenia (Investment funding of the Republic of Slovenia and the European Union of the European Regional Development Fund). We also acknowledge support from the Slovenian research agency ARRS (program no. P1-0383, J1-9186, J1-1715, J5-1796, and J1-1692), Google ATAP, and the Finnish Center for Artificial Intelligence (FCAI).

REFERENCES

[1] F. Adib, C.-Y. Hsu, H. Mao, D. Katabi, and F. Durand. 2015. Capturing the Human Figure through a Wall. *ACM Trans. Graphics* 34, 6 (2015).

[2] D. Avrahami, M. Patel, Y. Yamaura, and S. Kratz. 2018. Below the Surface: Unobtrusive Activity Recognition for Work Surfaces Using RF-Radar Sensing. In *Proc. IUI*. 439–451.

[3] F. Bernardo, N. Arner, and P. Batchelor. 2017. O soli mio: exploring millimeter wave radar for musical interaction. In *Proc. NIME*. 283–286.

[4] B. Ens, A. Quigley, H.-S. Yeo, P. Irani, T. Piumsomboon, and M. Billinghurst. 2017. Exploring Mixed-Scale Gesture Interaction. In *SIGGRAPH Asia Posters*.

[5] N. Y. Hammerla, S. Halloran, and T. Plotz. 2016. Deep, Convolutional, and Recurrent Models for Human Activity Recognition Using Wearables. In *Proc. IJCAI*. 1533–1540.

[6] J. Lien, N. Gillian, M. E. Karagozler, P. Amihood, C. Schwesig, E. Olson, H. Raja, and I. Poupyrev. 2016. Soli: Ubiquitous gesture sensing with millimeter wave radar. *ACM Trans. Graphics* 35, 4 (2016), 1–19.

[7] J. McIntosh, M. Fraser, P. Worgan, and A. Marzo. 2017. DeskWave: Desktop Interactions Using Low-Cost Microwave Doppler Arrays. In *Proc. CHI EA*. 1885–1892.

[8] J. Ng, M. Hausknecht, S. Vijayanarasimhan, O. Vinyals, R. Monga, and G. Toderici. 2015. *Beyond short snippets: Deep networks for video classification.* Technical Report arXiv:1503.08909. Cornell University.

[9] F. J. Ordóñez and D. Roggen. 2016. Deep Convolutional and LSTM Recurrent Neural Networks for Multimodal Wearable Activity Recognition. *Sensors* 16, 1 (2016).

[10] J. Paradiso, C. Abler, K.-y. Hsiao, and M. Reynolds. 1997. The Magic Carpet: Physical Sensing for Immersive Environments. In *Proc. CHI EA*. 277–278.

[11] D. Powers. 2011. Evaluation: from Precision, Recall and F-measure to ROC, Informedness, Markedness and Correlation. *J. Mach. Learn. Technol.* 2, 1 (2011), 37–63.

[12] C. Sandor and H. Nakamura. 2018. SoliScratch: A Radar Interface for Scratch DJs. In *Proc. ISMAR-Adjunct*. 427–427.

[13] I. E. Sutherland. 1963. *Sketchpad: A man-machine graphical communication system.* Technical Report 296. Lincoln Laboratory, MIT.

[14] I. E. Sutherland. 1965. The ultimate display. In *Proc. IFIP Congress*. 506–508.

[15] K. Čopič Pucihar, C. Sandor, M. Kljun, W. Huerst, A. Plopski, T. Taketomi, H. Kato, and L. A. Leiva. 2019. The Missing Interface: Micro-Gestures on Augmented Objects. In *Proc. CHI EA*.

[16] S. Wang, J. Song, J. Lien, I. Poupyrev, and O. Hilliges. 2016. Interacting with Soli: Exploring fine-grained dynamic gesture recognition in the radio-frequency spectrum. In *Proc. UIST*. 851–860.

[17] H.-S. Yeo, G. Flamich, P. Schrempf, D. Harris-Birtill, and A. Quigley. 2016. RadarCat: Radar Categorization for Input & Interaction. In *Proc. UIST*. 833–841.

[18] M. Zhao, T. Li, M. A. Alsheikh, Y. Tian, H. Zhao, A. Torralba, and D. Katabi. 2018. Through-Wall Human Pose Estimation Using Radio Signals. In *Proc. CVPR*. 7356–7365.