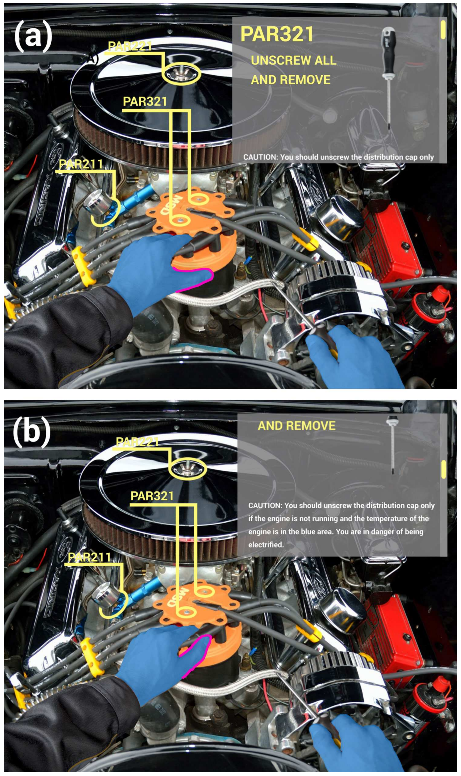


**Application scenario and envisioned interaction examples. A mechanic who is following AR instructions (a) can swipe on the nearby cable to (b) scroll down the instructions view.**

# The Missing Interface: Micro-Gestures on Augmented Objects

**Klen Čopič Pucihar***
University of Primorska
klen.copic@famnit.upr.si

**Christian Sandor**
NAIST
sandor@is.naist.jp

**Matjaž Kljun***
University of Primorska
matjaz.kljun@upr.si

**Wolfgang Huerst**
Utrecht University
huerst@uu.nl

**Alexander Plopski**
NAIST
plopski@is.naist.jp

**Takafumi Taketomi**
NAIST
takafumi-t@is.naist.jp

**Hirokazu Kato**
NAIST
kato@is.naist.jp

**Luis A. Leiva**
Aalto University
name.surname@aalto.fi

*Also with Faculty of information studies, Novo Mesto, Slovenia

## ABSTRACT

Augmenting arbitrary physical objects with digital content leads to *the missing interface* problem, because those objects were never designed to incorporate such a digital content and so they lack a user interface. A review of related work reveals that current approaches fail due to limited detection fidelity and spatial resolution. Our proposal, based on Google Soli's radar sensing technology, is designed to detect micro-gestures on objects with sub-millimeter precision. Preliminary results with a custom gesture set show that Soli's core features and traditional machine learning models (Random Forest and Support Vector Machine) do not lead to robust recognition accuracy, and so more advanced techniques should be used instead, possibly incorporating additional sensor features.



*CHI'19 Extended Abstracts, May 4–9, 2019, Glasgow, Scotland UK*



ACM ISBN 978-1-4503-5971-9/19/05.
https://doi.org/10.1145/3290607.3312986

**KEYWORDS**

Augmented Reality; Google Soli; Millimeter-wave Radar; Micro-gesture Recognition

***The missing interface*** *problem occurs when physical objects are augmented with digital elements, because those objects were never designed to incorporate digital content and thus they do not provide a proper user interface to allow for interaction.*

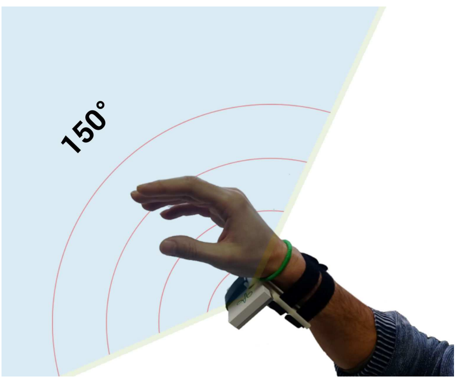


**Figure 1: We used a wrist-mounted Google Soli for sensing micro-gestures on physical objects.**

[1]https://docs.fcc.gov/public/attachments/DA-18-1308A1.pdf

## INTRODUCTION

With the development of Augmented Reality (AR) technologies, we have adopted tools that are able to manipulate our visual system and modify our perspective of reality, e.g. adding or removing realistic or abstract objects from our field of view. Even though AR technology is still far from reaching the holy grail in which the augmented content can no longer be distinguished from reality [11], there are plenty of use cases where this (conscious) sense of augmentation is required. Some prominent examples include e.g. remote guidance, instruction manuals, and training systems (see teaser Figure). Such systems attempt to convey spatially complex information to the user by projecting additional information onto physical objects. This is in fact an important advantage of AR interfaces over conventional interaction systems [5, 6].

Augmenting arbitrary physical objects with digital elements leads to the *missing interface* problem, because those objects were never designed to incorporate digital content and thus they do not provide a proper user interface to allow for interaction. For example, imagine a maintenance task where instruction annotations are projected directly onto the objects in need of repair. How do we browse the instructions? How do we query for additional information? How do we provide feedback in order to communicate a problem to the remote expert? How do we effectively curate new digital instructions if we are the authors of such instruction manuals?

In this paper, we explore the possibilities of a “Swiss Army knife”-like interface that would fill in the missing interface void. We do this by (i) reviewing current approaches and similar interaction methods and (ii) proposing and evaluating a novel gesture detection system based on Google Soli’s millimeter-wave radar sensing technology. This new sensing technology has been very recently approved by the Federal Communications Commission (FCC),[1] suggesting it will soon become widespread to developers and practitioners. This makes technology evaluation, such as the one presented in this paper, an important and timely contribution.

## RELATED WORK

To overcome the missing interface problem, AR interactive systems follow two different patterns: (i) *remote* interaction, where interaction is dislocated from the augmented object (e.g. mid-air gestures to place holograms around the room) and (ii) *direct* interaction, where interaction is performed on the augmented object (e.g. tapping the object with annotation to reveal more information).

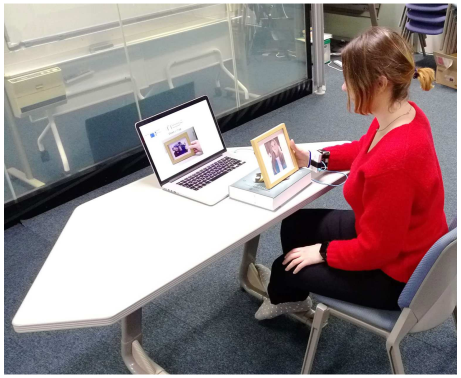

**Figure 2: In our user study, participants sat at a table while interacting with a picture frame.**

### Remote Interaction

Interfaces that enable remote interaction can be implemented through speech recognition, using additional handheld controllers, or even by performing mid-air gestures with hands and body. These systems work reasonably well in controlled environments, particularly when used in multimodal setups, however they face some restrictions. For example, handheld controllers require the user to hold a device and gesture recognition systems are commonly based on captured RGB or RGB-D data streams [13] imposing the following limitations: (i) gestures need to be performed within the camera's field of view; (ii) gestures require a reasonably large hand or finger movements, cf. HoloLens gestures;[2] (iii) the hand and fingers performing the gesture must not be occluded.

To address the lack of line of sight when detecting mid-air gestures, previous research looked at alternative sensing abilities, such as electric field sensing for full-body gestures [2], fingertip tracking [8], audio-based doppler shift sensing [7], and capacitive sensing [10]. Other solutions have been explicitly designed as wearables, such as acceleration-sensitive finger rings [4] or a wristband device that recognizes hand gestures and forearm movements [10].

[2]https://fologram.com/docs/articles/360000919873

### Direct Interaction

Building interfaces that enable direct interaction is more challenging, since the interface needs to be integrated with the object being augmented. Moreover, objects add noise to the gesture detection pipeline and increase the difficulty of hand segmentation. To overcome these problems, previous research utilized infrared proximity sensors, allowing for e.g. multi-touch interaction around small devices [1], capacitive sensing techniques enabling detection of touch events on humans, screens, liquids, and everyday objects [12], and even electromyograpy systems that measure muscle tension [3].

Irrespective of recent advances in gesture detection systems and sensing methods, current solutions suffer from limited gesture detection fidelity and spatial resolution, especially when gestures involve subtle hand/finger movements. Therefore, current gesture detection systems fail to support micro-gestures on objects that otherwise would enable inconspicuous, precise, and flexible object oriented interactivity. This is important when a direct interaction pattern is more appropriate for the task at hand, a common case in AR-supported remote guidance, instruction manuals, and training systems.

## MICO-GESTURES ON AUGMENTED PHYSICAL OBJECTS

Millimeter-wave radar sensing technology has recently shown promising results in the context of detecting micro-gestures [9, 15]. This technology is based on the principle of emitting electromagnetic waves at high frequencies (57–64 GHz) and capturing their reflections using a radar antenna. These signals are later processed to produce data features that can be used to detect user fine-grained interactions using machine learning techniques.

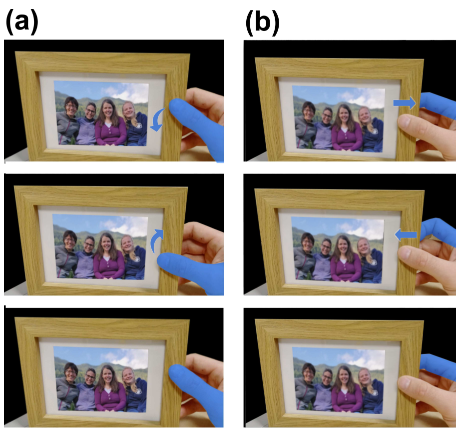


**Figure 3: Thumb moving up&down (a) and Scratch (b) gesture examples.**

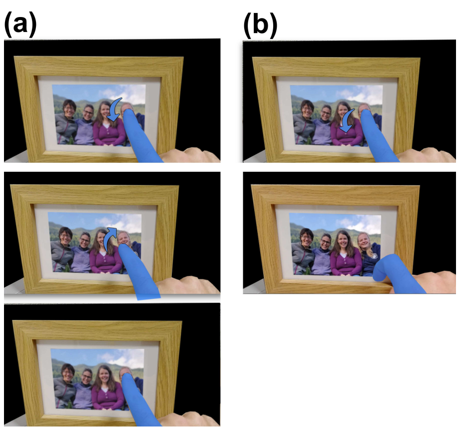


**Figure 4: Tickle (a) and Swipe (b) gestures.**

Due to the nature of radar waves (e.g. can penetrate certain materials) and the proof that micro-gesture detection is possible with this technology [9], we hypothesize that *millimeter-wave radar sensing technology is capable of detecting micro-gestures on physical objects without having to instrument the augmented object.* We test our hypothesis using Google Soli, which has shown that radar-based interaction technology can be used to perform "robust, high-resolution, low-power, miniature mid-air gesture sensing" [9]. In addition, the Soli sensor has been shown to be successful at classifying everyday objects and materials [16–18].

We designed and 3D-printed a hand-mounted wristband for the Google Soli sensor (see Figure 1). The sensor was positioned to illuminate the fingers, maximizing the amount of captured reflected signal caused by hand and finger movements. The sensor was connected to a laptop computer with a USB cable. Current form factor is likely to miniaturize as the technology improves making the approach viable for wearables.

Next, we conducted a user study (Figure 2) and recorded ~3.2K labeled instances (4 gestures, 10 users, 20 gesture repetitions, 3–5 sessions). The four gestures were Thumb, Scratch, Tickle, and Swipe; see Figures 3 and 4. These gestures were designed based on the analysis of finger-based gesture sets that have been used in previous work [9, 15]. Participants (all students, 6 male, 4 female, aged 23–50) were sitting at a table whilst performing the gestures on a picture frame. An image of the gesture with its name was shown to each participant on the nearby laptop. After a beep sound, the participant had to perform the gesture. The order of gestures was randomized and after each round the sensor was reset (the clutter map was rebuilt). Participants could perform each gesture in a warm-up condition before starting the experiment. Participants were asked to repeat the gesture if the one they performed did not match the gesture shown on the image (e.g. the user performed the Scratch gesture instead of the Thumb gesture).

## EVALUATION

We used the standard gesture detection pipeline proposed by Google Soli's SDK. We created feature vectors using the 9 built-in core features of Google Soli sensor recommended by the SDK: Acceleration, Energy Moving, Energy Total, Movement index, Range, Spatial dispersion, Velocity, Velocity Dispersion. We computed 4 meta-features for each core feature: mean, standard deviation, sum, and absolute sum. Considering that each of our gestures is approximately 1 second long, we logged sensor data at 100 Hz. As a result, our feature vectors have a dimensionality of $(9 + 9 \cdot 4) \cdot 100 = 4500$. We used these feature vectors to train two popular machine learning models: Random Forest (RF, with forest size of 200 and depth of 10) and Support Vector Machine (SVM, with RBF kernel and regularization). The hyperparameters of each model were tuned via grid search.

We performed an 80/20 split of the data, where 80% of the data is used for training and the remaining 20% is used for testing. The results are presented in Tables 1–3. As can be observed, RF outperforms

| | Full set | Reduced set |
|---|---|---|
| No. classes | 4 | 3 |
| No. samples | 3156 | 2367 |
| RF accuracy (%) | 55 | 68 |
| SVM accuracy (%) | 50 | 61 |

**Table 1: Classification results.**

| | Thumb | Scratch | Tickle | Swipe |
|---|---|---|---|---|
| Thumb | **59%** | 19% | 14% | 8% |
| Scratch | 23% | **54%** | 15% | 7% |
| Tickle | 13% | 16% | **55%** | 16% |
| Swipe | 16% | 11% | 20% | **53%** |

**Table 2: Confusion matrix for RF. Full gesture set.**

| | Thumb | Scratch | Swipe |
|---|---|---|---|
| Thumb | **71%** | 14% | 15% |
| Scratch | 9% | **70%** | 21% |
| Swipe | 11% | 25% | **64%** |

**Table 3: Confusion matrix for RF. Reduced gesture set.**

[3] https://docs.microsoft.com/en-us/windows/mixed-reality/gestures

the SVM model, however if we consider our original set of 4 gestures it only manages to achieve an accuracy of 55%, which is insuficient for a production-ready system. Looking at the confusion matrix presented in Table 2, we can see that Thumb/Scratch gestures and Tickle/Swipe gestures are often misrecognized. If we remove the Tickle gesture gesture from our recognition pipeline, the RF model achieves an overall accuracy of 68%, which represents an increment of 23%.

## DISCUSSION AND FUTURE WORK

Our work attempts to fill in the void of *the missing interface* which is present in physical objects that are augmented with digital content, since those objects were never designed to support digital augmentation and so do not provide the necessary interface. Our preliminary results show that built-in features (provided by the Soli sensor) and traditional machine learning methods (Random Forest and Support Vector Machine) do not lead to robust recognition when using our gestures; e.g. the overall accuracy for a set of 4 gestures is 56%, whereas for a set of 3 gestures is 68%. Nonetheless, further improvements should be possible to attain with e.g. additional sensor features and more advanced machine learning methods. This is further supported by previous work from Wang et al. [14] that evaluated mid-air gesture detection using Range-Doppler images and deep neural networks architectures (convolutional and recurrent neural networks) and achieved very good results; e.g. 85% accuracy with a set of 11 gestures.

Our experiment had a number of limitations that will be addressed in future work, such as: (i) statistical testing for feature selection: we used the feature set proposed in the Google Soli SDK, however a feature selection technique should be considered to improve accuracy; (ii) cross-validation training: this should help to improve classification accuracy as well, particularly in the case of small datasets. Nevertheless, our preliminary results have important implications, as we provide insights into how effective are Google Soli core features for micro-gesture detection. To our knowledge, no such evaluation has been reported before and thus may be of interest to developers and practitioners.

Besides experimenting with the aforementioned changes to the gesture detection pipeline, we plan to evaluate the proposed system on an extended gesture set and integrate it into a final application. Due to the fact that objects introduce noise to the gesture detection pipeline, we also plan to evaluate how resilient is our system when gestures are performed on different object materials of different shapes and sizes as well as mid-air vs. on-object versions of the same gesture. Another interesting experiment would be to compare micro-gestures interaction against other current gesture sets, such as those used in HoloLens.[3]

## 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). This research was also supported by Slovenian research agency ARRS (program no. P1-0383 and J1-9186), Google ATAP, and the Academy of Finland (BAD project). We also acknowledge the support of NVIDIA Corporation with the donation of the Titan Xp GPU used for this research.

## REFERENCES

[1] Alex Butler, Shahram Izadi, and Steve Hodges. 2008. SideSight: Multi-“touch” Interaction Around Small Devices. In *Proc. UIST*, Vol. 23. 201.

[2] Gabe Cohn, Daniel Morris, Shwetak Patel, and Desney Tan. 2012. Humantenna: Using the Body as an Antenna for Real-Time Whole-Body Interaction. In *Proc. CHI*. 1901.

[3] Ulysse Côté-Allard, Cheikh Latyr Fall, Alexandre Drouin, Alexandre Campeau-Lecours, Clément Gosselin, Kyrre Glette, François Laviolette, and Benoit Gosselin. 2018. Deep Learning for Electromyographic Hand Gesture Signal Classification by Leveraging Transfer Learning. *arxiv.org* (2018), 1–15. arXiv:1801.07756

[4] Masaaki Fukumoto and Yoshinobu Tonomura. 1997. “Body Coupled FingerRing”: Wireless Wearable Keyboard. In *Proc. CHI*. 147–154.

[5] Steffen Gauglitz, Benjamin Nuernberger, Matthew Turk, and Tobias Höllerer. 2014. In Touch With The Remote World: Remote Collaboration with Augmented Reality Drawings and Virtual Navigation. In *Proc. VRST*. 197–205.

[6] Steffen Gauglitz, Benjamin Nuernberger, Matthew Turk, and Tobias Höllerer. 2014. World-stabilized Annotations and Virtual Scene Navigation for Remote Collaboration. In *Proc. UIST*. 449–459.

[7] Sidhant Gupta, Daniel Morris, Shwetak Patel, and Desney Tan. 2012. SoundWave: Using the Doppler Effect to Sense Gestures. In *Proc. CHI*. 1911–1914.

[8] Mathieu Le Goc, Stuart Taylor, Shahram Izadi, and Cem Keskin. 2014. A Low-cost Transparent Electric Field Sensor for 3D Interaction on Mobile Devices. In *Proc. CHI*. 3167–3170.

[9] Jaime Lien, Nicholas Gillian, M Emre Karagozler, Patrick Amihood, Carsten Schwesig, Erik Olson, Hakim Raja, Ivan Poupyrev, and Google Atap. 2016. Soli: Ubiquitous Gesture Sensing with Millimeter Wave Radar. *ACM Transactions on Graphics (TOG)* 35, 10 (2016), 1–19.

[10] Jun Rekimoto. 2001. GestureWrist and GesturePad: Unobtrusive Wearable Interaction Devices. In *Proc. ISWC*. 21–27. arXiv:0-7695-1318-2

[11] Christian Sandor, Martin Fuchs, Alvaro Cassinelli, Hao Li, Richard Newcombe, Goshiro Yamamoto, and Steven Feiner. 2015. Breaking the Barriers to True Augmented Reality. *arxiv.org* December (2015), 1–13. arXiv:1512.05471

[12] Munehiko Sato, Ivan Poupyrev, and Chris Harrison. 2012. Touché: : Enhancing Touch Interaction on Humans, Screens, Liquids, and Everyday Objects. In *Proc. CHI*. 483. arXiv:1-58113-734-6

[13] Toby Sharp, Yichen Wei, Daniel Freedman, Pushmeet Kohli, Eyal Krupka, Andrew Fitzgibbon, Shahram Izadi, Cem Keskin, Duncan Robertson, Jonathan Taylor, Jamie Shotton, David Kim, Christoph Rhemann, Ido Leichter, and Alon Vinnikov. 2015. Accurate, Robust, and Flexible Real-time Hand Tracking. In *Proc. CHI*. 3633–3642.

[14] Quan Wang, Yu Sang, Quan Wang, and Yimin Liu. 2016. Micro Hand Gesture Recognition System Using Ultrasonic Active Sensing Method. *arxiv.org* May (2016), 1–8. arXiv:arXiv:1712.00216v1

[15] Saiwen Wang, Jie Song, Jaime Lien, Ivan Poupyrev, and Otmar Hilliges. 2016. Interacting With Soli: Exploring Fine-Grained Dynamic Gesture Recognition in the Radio-Frequency Spectrum. In *Proc. UIST*. 851–860.

[16] HS Yeo, A Quigley Interactions, and Undefined 2017. 2017. Radar Sensing in Human-computer Interaction. *Interactions* 25, 1 (2017), 70–73.

[17] Hui-Shyong Yeo, Barrett Ens, and Aaron Quigley. 2017. Tangible UI by Object and Material Classification with Radar. In *Proc. SIGGRAPH Asia*. 1–2.

[18] Hui-Shyong Yeo, Gergely Flamich, Patrick Schrempf, David Harris-Birtill, and Aaron Quigley. 2016. RadarCat: Radar Categorization for Input & Interaction. In *Proc. UIST*. 833–841.