Mid-air Gesture Recognition by Ultra-Wide Band Radar Echoes

Arthur Sluyters

1Louvain Research Institute in Management and Organizations, Université catholique de Louvain, Place des Doyens 1, Louvain-la-Neuve, 1348, Belgium

Abstract
Microwave radar sensors in human-computer interaction promote several advantages over wearable and image-based sensors, such as privacy preservation, high reliability regardless of the ambient and lighting conditions, and larger field of view. However, the raw signals produced by such radars are high-dimension and very complex to process and interpret for gesture recognition. For these reasons, machine learning techniques have been mainly used for gesture recognition, but require a significant amount of gesture templates for training and calibration that are specific for each radar. To address these challenges in the context of mid-air gesture interaction, we introduce a data processing pipeline for hand gesture recognition adopting a model-based approach that combines full-wave electromagnetic modelling and inversion. Thanks to this model, gesture recognition is reduced to handling two dimensions: the hand-radar distance and the relative dielectric permittivity, which depends on the hand only (e.g., size, surface, electric properties, orientation). We are developing a software environment that accommodates the significant stages of our pipeline towards final gesture recognition. We already tested it on a dataset of 16 gesture classes with 5 templates per class recorded with the Walabot, a lightweight, off-the-shelf array radar. We are now studying how user-defined radar gestures resulting from gesture elicitation studies could be properly recognized or not by our gesture recognition engine.

Keywords
Gesture-based interfaces, Mid-air gestural interaction, Radar-based interaction

1. Context of the Problem
Two/three-dimensional (2D/3D) gesture-based User Interfaces (UIs) [1] promise a natural and intuitive interaction [2] as they rely on movements performed by the human body, which are assumed to be more natural and easier to remember than artificially-determined commands. Gesture-based UIs typically fall into three categories depending on the number of gesture dimensions and the sensor used to capture and recognize the gesture:

- **2D (nearly-)touch-based.** Gestures are performed on a sensitive surface, such as a trackpad, a touchscreen, or a touchable surface (e.g., in Gambit [3]), which constrain the movement to a series of 2D temporally-sequenced points \((x_i, y_i, t)\). Gestures performed on a spatial object, such as a tangible object, remain in this category even if the object is spatially moving. These gestures typically requires a contact-based sensor or a close-by sensor,
Figure 1: Research methodology for radar-based interaction.

when the gesture could stay close to the surface, but not touching, such as for a grazing gesture. Touch-based gestures can be augmented with a third dimension, like pressure, contact surface, or temperature.

• **3D wearable.** Gestures are captured by devices worn by the user, such as smartwatches, smart gloves, ring devices, or even smartphones. These devices often rely on accelerometers, gyroscopes, and magnetometers, usually combined into a single IMU (Internal Measuring Unit), to capture the orientation and acceleration of the device in 3D space. Other wearable devices capture motion by measuring the activity of muscles using electromyography (EMG).

• **3D non-wearable.** Gestures are captured by non-wearable devices, such as vision-based sensors or radars. Vision-based sensors are quite popular, and include the Intel RealSense cameras, the Microsoft Kinect, the Ultraleap Leap Motion Controller (LMC), and even laptop cameras. They rely on computer-vision algorithms for motion tracking. Some of these sensors (e.g., the Microsoft Kinect) can capture full-body gestures and poses, while others may be limited to some body parts, such as the LMC, which focuses on arm and hand gestures.

Existing vision-based sensors may suffer from problems inherent to image-based processing [4]: sensitivity to ambient conditions, particularly lighting, limited field of view, transient or permanent vision occlusion, and privacy concerns raised by a visible device observing the end user. Similarly, wearable sensors are more invasive, as they must be worn by the user, and may raise hygiene concerns, especially in critical environments such as hospitals. While radars have their own set of issues (e.g., radar signals can be difficult to interpret), they do not suffer from the same problems as vision-based and wearable sensors, thus making them a potential
To address the aforementioned shortcomings and challenges, we wish to investigate if the question of radar-based gesture interaction is feasible by: (1) exploring and determining the space of 3D gestures that are acceptable both from an end-user and a system perspective, (2) defining and testing a model-based approach transforming the raw signals from the radar into meaningful information for gesture recognition, (3) developing a software environment for engineering radar-based gesture interaction. To achieve this objective in a realistic and focused way, we pose the following working hypotheses: we choose to work primarily with the Walabot device, a radar that is widely commercially available for fixed or mobile interaction; we focus on hand gestures, which is considered to be among the richest and most diverse methods of interaction; we assume that the end user can add or remove gestures dynamically; we start with simple control gestures.

2. Related Work

Radar-based sensing technologies [4] are now being considered as a relevant alternative to other types of sensors for human-computer interaction. They have been successfully applied in multiple domains such as virtual reality [5], activity recognition [6, 7], material recognition [8, 4], and tangible interaction [9]. Prior works on gesture recognition using radar sensing, such as [7, 10, 11, 12], typically rely on a fixed, custom-built radar.

As far as we know, all existing techniques developed for radar-based gesture recognition rely on machine/deep learning algorithms [11, 12, 13] to cope with the high dimensionality and the complexity of radar signals. For this reason, these recognizers mainly run in a stationary environment, not a mobile one. However, the Walabot device, while being commercially available and deployable in both stationary and mobile environments, introduces a new series of constraints imposed by its limited amount of antennas, its small size, and its relatively narrow bandwidth. The Google Soli radar chip [13], embedded in an Android smartphone, is able to recognize 6 classes of radar-based gestures, thus making it mobile and available, but not customizable: only the 6 gestures are recognized and cannot be changed.

In principle, radar sensing techniques allow interaction without any wearable and visible device since a radar can be operated below a surface such as a desk [6], behind a wall, and behind different materials without significantly affecting the recognition [11] (e.g., wallpaper, cardboard, and wood benefit from relatively low permittivity). Radars are also insensitive to weather and lighting conditions [14]. RadarCat [4] recognizes physical objects and materials placed on top of the sensor in real-time by extracting signals and classifying them using a random forest classifier. They successfully tested 16 transparent materials with different properties such as thickness and 10 body parts, thereby demonstrating the real potential of permittivity. Yeo et al. [9] used a radar in the context of tangible interaction for counting, ordering, and identifying objects involved in a tangible setup, for tracking their orientation, movement, and between-object distance, three variables that were originally captured by infrared. Beyond object and material detection and classification, radars start to be widely used in several domains of application, such as indoor human sensing with commodity radar [15], human activity recognition [6], human position estimation [7], motion detection and classification [11]. Pantomime [11] mounts a
fixed foot-based radar with a high-frequency and continuous bandwidth and relies on deep
learning, i.e., LSTM and Pointnet++, to recognize 21 gestures acquired from 45 participants in
terms of 3D point clouds. Wang et al. [16] require only two antennas in their radar to recognize
2D stroke gestures: their low-dimensionality, as opposed to 3D gestures, does not require more
antennas. Short-range radar-based gestures could also be recognized using 3D convolutional
neural networks [17].

3. Research Methodology

Our research methodology consists of performing the following stages (Fig. 1):

**SLR of radar-based gestures and recognition techniques (Fig. 2).** Radar-based gesture
recognition is a hot topic that attracts a lot of very recent works. Although techniques for
radar-based gesture recognition are reviewed in general [18] and for hand gestures in particu-
lar [19], we believe that there is still a need to conduct a Systematic Literature Review (SLR) to
systematically determine not only the techniques (e.g., to show that only machine/deep learning
techniques are primarily used as opposed to our model-based approach) but also the gestures
covered (e.g., to provide a compilation of recognized gestures) and the radar systems used for
capturing gestures. Our approach is inspired by the four-phase SLR method (Identification,
Screening, Eligibility and Inclusion) proposed by Liberati et al. [20] and the flow is represented
in a PRISMA diagram.

**Gesture elicitation study with radar (bottom left of Fig. 1).** To consolidate the gestures
identified in the SLR, a gesture elicitation study [22] is conducted to explore user-defined
gestures relevant to our context of use (i.e., a Walabot transforming hand gestures into control
commands in a IoT environment). The output of this study consists of computing an agreement
rate among participants for the proposed gestures. We also envision to conduct a similar
study for gestures where only a radar could operate: in dark conditions, behind a door, or in
a reflected setup. Conducting new gesture elicitation studies with radar sensors is desirable
as user gestures may not always be transferable across different devices. For instance, users
may propose different gestures for the same action performed using a radar, a vision-based, or
a wearable sensor.

**Gesture acquisition with radar (top left of Fig. 1, Fig. 3, Fig. 4).** When participants
propose gestures in an elicitation study, they behave naturally in their proposal, but they do
not necessarily focus on the gestures themselves, thus raising the need to acquire gesture datasets corresponding to the most agreed upon gestures according to a rigorous procedure. The conjunction of the results coming from this activity, along with the SLR and the gesture elicitation study, leads us to create a series of gesture datasets for radars that will be made available to the scientific community. The datasets could be recorded in various environments (e.g., in an office, behind a door) and with different types of sensors, including vision-based sensors such as the LMC for comparison purposes.

![Gesture Types](image)

**Figure 3**: The sixteen gesture types used in our first dataset.

**Radar-based gesture recognition environment (top right of Fig. 1)**. We are developing a software environment for ensuring the accurate recognition of gestures belonging to the datasets of the previous stage and for engineering a radar-based gestural user interface for an interactive software. This environment is divided into two main parts: (1) a pipeline for radar data pre-processing that removes noise and clutter from the signal and performs dimension
reduction for use with simple template-matching gesture recognition algorithms, and (2) a framework for integrating gesture recognition into an application. In this way, the developer of such an interactive software will see the process of integrating radar-based interaction facilitated by its recognition and its mapping of gestures to commands.

The bottom right part of Fig. 1 depicts the range of radar-based gestures in terms of two dimensions: the X-axis structures gestures according to their agreement rate obtained in the gesture elicitation study and the Y-axis structures gestures according to their recognition rate obtained from our recognition environment. Some user-defined gestures may receive a high agreement rate among participants, but be hardly recognized: to preserve the experimental conditions, participants are untold about the restrictions imposed by the technology and the recognition environment. Conversely, some system-defined gestures may benefit from a high accuracy resulting from our environment, but receive a low agreement rate among participants. Therefore, the target area, depicted in dark green on the top right portion of the space, is the most desirable area where gestures receive both a high agreement rate and a high recognition accuracy.

4. Current status

Based on the research methodology (Fig. 1), an SLR has been completed (Fig. 2), in which we identified 118 relevant papers from a set of 1515 references. Its results are now available, but not yet published. A new gesture elicitation study has been conducted with the Walabot device and its results are being analyzed. Two gesture datasets have already been created: one with 16 gesture classes having 5 templates each (Fig. 3) and another one with 20 gesture classes having 50 templates each (Fig. 3 and 4). Based on radar electromagnetic modeling and inversion [23, 24], a very first version of the radar signal processing pipeline has been developed and tested on the first dataset, already delivering encouraging results [25]. We developed QuantumLeap, a framework for developing gesture-based applications that has been applied to the Leap Motion Controller and evaluated with seven developers [26]. LUI, a gesture-based application for manipulating multimedia content, has been developed with the QuantumLeap framework and the LMC [27]. In the future, both QuantumLeap and the LUI application should be adapted to support radar-based gestures.
5. Challenges

In this section, we discuss the main challenges of this thesis that have been identified so far:

**Gesture recognition accuracy.** The system should be able to recognize gestures with sufficient accuracy to not hinder user interaction with the system. Inaccurate gesture recognition could create a lot of frustration for users, as it would regularly force them to perform the same gesture twice, and could cause the system to incorrectly react to user gestures. Reaching \( >90\% \) accuracy, especially between different users and in various contexts, will be a major challenge to solve.

**Online radar-based gesture recognition.** The performance of the gesture recognition environment should be high enough to enable real-time execution. Some stages of the current radar signal processing pipeline, such as the “inversion” stage, should be optimized to meet this objective. In addition, appropriate gesture segmentation techniques will be required to accurately identify gestures performed by the user from the continuous stream of radar data.

![Figure 5: Variation of hand size across humans.](image)

**Data normalization across users.** Physical differences between two persons, such as their hand size or arm length (Fig. 5), may prevent a system trained with data from one user to accurately recognize gestures performed by another user. For instance, for the same gesture, the amplitude of the reflected radar signal varies depending on hand size (a larger hand reflects more signal). Similarly, users’ arm length and body size will impact the reflected signal and thus could result in lower accuracy between users. One solution would be to integrate a data normalization stage into the radar signal processing pipeline that would remove the effect of users’ physical characteristics on the signal after a small calibration step.

**Two-handed and multi-user interaction.** In its current state, the radar-based gesture recognition environment only supports one-handed gestures performed by a single user. Supporting multiple users and/or two-handed gestures will require major changes to the radar signal processing pipeline, in particular the inversion stage.

**Privacy concerns.** In contrast to vision-based sensors, radars such as the Walabot do not capture clear images of the users. However, they possess other features that may raise privacy concerns, such as their ability to see through some materials, and thus to identify user motion while being completely hidden, or their ability to function in dark environments. Further research is required to identify the privacy concerns of radars and to evaluate how they are perceived by end-users.
6. Expected contributions

The expected contributions of the doctoral thesis are as follows:

1. One SLR related to radar-based gestural interaction highlighting the existing techniques for gesture recognition, the radar systems used, and a classification of all gestures involved in these works to initiate an inventory of radar-based gestures.

2. Various gesture elicitation studies in the context of radar-based gestural interaction. The studies will cover different different types of radars, including Google Soli [28] and the Walabot. One study will focus on controlling IoT devices with a Walabot, while other studies will focus on radar gestures with the Walabot in extreme conditions (e.g., in a dark environment).

3. A consolidation of all radar-based gestures found in the literature and resulting from the above studies into a repository in order to acquire radar gesture sets in original format or by transformation [29].

4. A radar-based gesture recognition environment that ensures both the recognition process and the engineering process for integrating radar-based gestures into a gestural interface of an interactive software.

5. A validation of this environment on selected use cases answering the initial research question.

Acknowledgments

The author of this paper acknowledges the support of the MIT-Belgium MISTI Program under grant COUHES n°1902675706 and of the "Fonds de la Recherche Scientifique - FNRS" under Grant n° 40001931.

References


