

# **Unobtrusive Activity Recognition and Position Estimation for Work Surfaces using RF-radar Sensing**

DANIEL AVRAHAMI\*, FXPAL, USA

MITESH PATEL, FXPAL, USA

YUSUKE YAMAURA, FUJI XEROX Communication Technology Laboratory, Japan

SVEN KRATZ<sup>†</sup>, FXPAL, USA

MATTHEW COOPER<sup>‡</sup>, FXPAL, USA

Activity recognition is a core component of many intelligent and context-aware systems. We present a solution for discreetly and unobtrusively recognizing common work activities above a work surface without using cameras. We demonstrate our approach, which utilizes an RF-radar sensor mounted under the work surface, in three domains; recognizing work activities at a convenience-store counter, recognizing common office deskwork activities, and estimating the position of customers in a showroom environment. Our examples illustrate potential benefits for both post-hoc business analytics and for real-time applications. Our solution was able to classify seven clerk activities with 94.9% accuracy using data collected in a lab environment and able to recognize six common deskwork activities collected in real offices with 95.3% accuracy. Using two sensors simultaneously, we demonstrate coarse position estimation around a large surface with 95.4% accuracy. We show that using multiple projections of RF signal leads to improved recognition accuracy. Finally, we show how smartwatches worn by users can be used to attribute an activity, recognized with the RF sensor, to a particular user in multi-user scenarios. We believe our solution can mitigate some of users' privacy concerns associated with cameras and is useful for a wide range of intelligent systems.

CCS Concepts: • Human-centered computing → Ubiquitous and mobile computing systems and tools;

Additional Key Words and Phrases: Activity recognition, Retail, Deskwork, Sensing, Radio Frequency Radar Sensor, IMU

## **ACM Reference Format:**

Daniel Avrahami, Mitesh Patel, Yusuke Yamaura, Sven Kratz, and Matthew Cooper. 2018. Unobtrusive Activity Recognition and Position Estimation for Work Surfaces using RF-radar Sensing. *ACM Trans. Interact. Intell. Syst.* 1, 1, Article 1 (January 2018), 27 pages.  
<https://doi.org/0000001.0000001>

<sup>\*</sup>This author's second affiliation is University of Washington.

<sup>†</sup>This author's second affiliation is Harman International.

<sup>‡</sup>This author's second affiliation is Mozilla.

---

Authors' addresses: Daniel Avrahami, FXPAL, USA, daniel@fxpal.com; Mitesh Patel, FXPAL, USA, mitesh@fxpal.com; Yusuke Yamaura, FUJI XEROX Communication Technology Laboratory, Japan, yamaura.yusuke@fujixerox.co.jp; Sven Kratz, FXPAL, USA, sven.kratz@harman.com; Matthew Cooper, FXPAL, USA, mlcooper@acm.org.

---

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

© 2018 Copyright held by the owner/author(s). Publication rights licensed to ACM.

Manuscript submitted to ACM

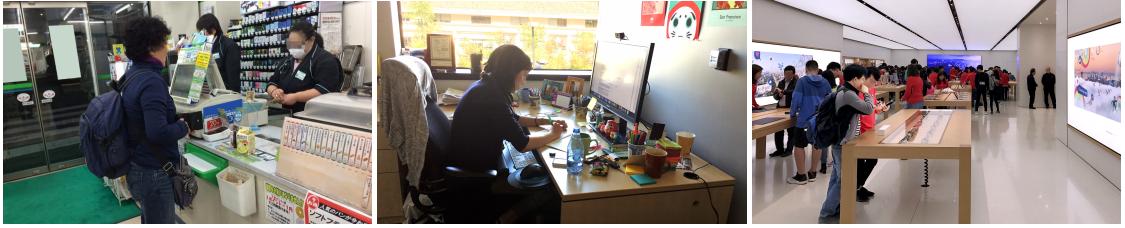


Fig. 1. Our three focus domains for recognizing activities performed above a work surface: a convenience-store checkout counter (left, ©Melissa Y.), an office desk (middle), and a retail showroom (right, ©Sharon Wswai).

## 1 INTRODUCTION

Recognizing user activity is a core component of many intelligent and context-aware systems. Over the last several decades, activity recognition has been demonstrated in a wide range of domains, including physical activity, manufacturing, cooking, activities of daily living, and many more. In the physical world, systems rely on sensors to recognize activities. This includes sensors attached to objects (such as RFID tags, shake sensors etc. cf. [7, 28]), sensors carried or worn on the user’s body (such as accelerometers or IMUs, cf. [4, 8, 13, 21, 31]), or instrumented in the environment (such as cameras, cf. [14, 19, 29, 46, 51]). Indeed, cameras have many advantages for activity recognition, including a wide view of the space in which an activity takes place, no requirement to instrument users, and, importantly, are low cost. However, intelligent systems that rely on cameras for activity recognition face a number of challenges. From a technical standpoint, cameras often require an unobstructed view of the activity and can be susceptible to changes in illumination (although a combination with non-visible light can mitigate this challenge). Furthermore, using cameras, particularly in private and semi-private spaces (such as offices and other businesses) poses potential violations (both real and perceived) of users’ privacy. Indeed, a video or even a still frame captured by a camera can often be easily interpretable by a human observer. Thus, other methods must be explored to deliver activity-recognition solutions that can overcome some users’ privacy concerns associated with cameras.

In this paper, we present an exploration of a solution for recognizing common activities performed above a work surface as well as presence and coarse position estimation around large surfaces. Our proposed solution makes use of a Radio Frequency (*RF*) radar sensor, mounted under the work surface to perform activity recognition in a discreet and unobtrusive way. The driving motivation for this work grew out of an engagement with a local convenience store chain interested in analytics of activities performed in their stores’ day-to-day operation with the goals of process optimization and improved space utilization. However, based on customer requirements, the use of cameras for capturing activities was to be avoided (other instruments such as beacons and worn sensors were acceptable). Indeed, the use of cameras can be problematic in many work environments such as office spaces and medical facilities that could benefit from accurate activity recognition for both post-hoc and real time applications. To address this constraint, we explored solutions for recognizing activities performed above the work surface (Figure 1), looking to provide a solution that would be accurate, unobtrusive, and with low instrumentation overhead.

In this paper, we focus first on activities performed at a convenience store’s checkout counter — resulting from our customer engagement — (e.g., scanning and bagging items). We then extend our work to recognition of common office deskwork activities (e.g., reading, eating, working on the computer) and to customer-engagement-estimation in a showroom environment. We demonstrate that our solution is superior to recognition based on Inertial Measuring Unit (*IMU*) sensing using smartwatches (the most common solution for camera-free recognition). However, we also

show that smartwatches, worn by the users, can still be useful for attributing a performed activity to the correct user in multi-user scenarios.

The central objective of this paper is to help establish the viability of these sensors for activity recognition in scenarios of practical interest that limit the use of cameras or user instrumentation. Our proposed solution is able to recognize a set of seven common activities performed at the counter with 94.9% accuracy and six common deskwork activities with 95.3% accuracy. Our solution is also able to perform coarse position estimation around a large surface using two sensors simultaneously with 95.4% accuracy. Since our setup does not require the use of a camera, we believe it can benefit many intelligent systems by alleviating some of the privacy concerns associated with cameras.

This work, which encompasses and extends [3], makes the following contributions:

- (1) We present a solution for recognizing activities performed above a work surface using an RF-radar sensor and show that using multiple projections of RF signal leads to improved recognition accuracy.
- (2) We demonstrate our solution in three work domains: activities performed at a convenience store's checkout counter, common deskwork activities in an office environment, and customer engagement estimation for a showroom environment.
- (3) We present a method for attributing a performed activity to the correct user in multi-user scenarios using a combination of RF-radar and IMU sensor data.

The remainder of the paper is organized as follows: we first review related work from the field of activity recognition. We then describe our solution, and present results for activity recognition in two domains. We next demonstrate using multiple sensors for covering large surfaces, focusing on position estimation in a third domain. We then present a method for identifying the user performing an activity using a combination of RF and IMU data. We conclude with a discussion and areas for future work.

## 2 RELATED WORK

Activity recognition for intelligent and context-aware systems is a broad field of research and a large body of previous work exists. Activity recognition systems have been built using sensors attached to objects, sensors carried or worn on the user's body, instrumented in the environment, or some combination.

Computer vision-based systems that take advantage of RGB images and/or depth images have been successfully used for a wide range of activity recognition tasks. Poppe *et al.* [29], Weinland *et al.* [46], Ke *et al.* [19], and Zhang *et al.* [52] provide surveys of image-based activity recognition systems and applications. Damen *et al.* [14] used a body-mounted RGB-D camera to monitor workspace activity. They were able to track individual objects in the workspace and classify a set of basic work steps, e.g., packaging goods in a box. Zhang *et al.* [51] describe a system for the recognition of daily life activities of seniors, using an RGB-D camera. Lavania *et al.* [22] use computer vision and deep neural networks to perform activity recognition for a biology laboratory. Cameras have also been used for detecting affective state (cf. [38]). Beyond several technical challenges associated with the use of cameras, they are considered particularly invasive and privacy compromising. In the papers discussed next, as well as in our proposed solution, the use of cameras was avoided.

In addition to cameras, other sensors have been used for activity recognition. Several systems (cf., [4, 5, 8, 10, 13, 21, 31, 33]) recognize physical human activities (walking, running, cycling, etc.) using a worn IMU, now a common feature of smartphones and smartwatches (for a survey on activity recognition using acceleration sensors see [21]). Ward *et al.* [44] use a combination of accelerometers and microphones for activity recognition of assembly tasks in

a workshop setting. Indeed, microphones from smartphones and smartwatches can also been used for recognizing other daily activities [39]. Many radio-based approaches have also been developed, using WiFi, RFID, ZigBee, or other technologies. These methods sense attenuation to passively detect and discriminate among activities. While some solutions, such as [35] and [45] offer device-free recognition, other solutions still require a user to wear a dedicated device (cf., [30]) or carry a phone running a dedicated application (cf., [37]). For a review, see [36].

Another important approach to human activity recognition is through the objects involved in the activity. For example, Philipose *et al.* developed a system for recognizing Activities of Daily Living (ADLs) by attaching RFID tags to household objects [28]. In [7], activity recognition from objects was expanded to include accelerometers in the tagged objects to observe not only a user's proximity to the object but the interaction itself. Marquardt *et al.* [26] developed a vision-based tool for creating interaction that utilizes the physical relationship between people, devices, and (visually tagged) non-digital objects. Recently, Laput *et al.* [20] have explored the use of a multi-sensor unit to detect user interaction with everyday household devices, intentionally avoiding using cameras. They utilized different levels of sensor hierarchy to classify higher-level human behaviors.

## 2.1 Activity Recognition for Work- and Retail Spaces

Some research exists that looked at recognizing activities within retail spaces, with much of the focus on understanding the behavior of shoppers (as opposed to store employees). Zeng *et al.* [50], for example, used changes in WiFi signals in the store to classify whether a shopper was walking or standing (and where). Radhakrishnan *et al.* [31] use a combination of a smart band and mobile phone IMU-sensing to classify customer activities in a retail scenario. In our work, we compare the performance of activity recognition using RF-radar and IMU data (and also the combination of the two) showing significantly higher accuracy using RF sensing. It must be noted that our focus is on activities of store employees, and that unlike in [31] recognition is limited to activities performed above or just around the work surface. Wimmer *et al.* [48] developed smart furniture that used networked capacitive sensors to perform activity sensing. Their CapTable is a wooden table equipped with capacitive sensors. These sensors allow tracking user hand locations and simple object manipulation tasks. Wimmer *et al.* share our goal of building a solution that is privacy-sensitive, unobtrusive and allows for implicit human-computer interaction. CapTable requires a rather complex instrumentation to cover the area of an entire work surface. In contrast, our proposed solution requires a minimal installation under an existing work surface such as a counter or a desk. Prior work has motivated sensing and understanding activities and states in an office workplace environment. Understanding a knowledge-worker's state can be useful for determining whether and when they can be disrupted (cf., [6, 16, 17, 53]). Similarly, prior work looked at a worker's activities to promote physical breaks in the workplace to combat sedentary behavior (e.g., [9, 32]). However, understanding non-digital work activities performed by a worker is important for intelligent systems to make appropriate recommendations. For example, in Züger *et al.*'s system deployment, participants complained that the system incorrectly considered them available when, in fact, they were performing an activity with a high cognitive load, just not on the computer [53]. As we show later, our solution is able to recognize activities such as writing and reading paper documents, which would alleviate some of the challenges these participants experienced.

## 2.2 Prior Uses of RF Radar Sensors

Adib *et al.* [1] describe a system that uses an array of RF antennas placed behind a wall to detect humans through light materials, such as drywall. Using the sensor data, they implement 3D skeleton estimation, gestures (in-air drawing), and user identification. This involves reconstructing a human skeleton representation using multiple reflections back to the

sensor. We were inspired by Adib *et al.*'s work, and our work similarly uses RF from behind a solid material (in our case, under a work surface, for activity recognition). Also for the purpose of person identification, Lin *et al.* used a RF-based radar to continuously track cardiac motion [18]. Recovering physiological information using these sensors can require significant signal processing efforts and specialized hardware, but these methods enable unobtrusive measurement. In contrast, our approach combines a commercially available sensor and well-established machine learning methods. Ding *et al.* [15] demonstrated using changes in backscatter communication between an active RF reader and an RFID tag to classify basic approaching and departure behavior. Wang *et al.* [42] demonstrated material recognition and coarse imaging using off-the-shelf RFID reader and tags. Shangguan *et al.* [36] used objects with multiple RFID tags and a single antenna to enable gesture-based interaction with various objects. For very short-range applications, RF sensing has been used to detect gestures and materials. Lien *et al.* [23] and Wang *et al.* [43] used a Google Soli to detect small finger gestures. The RadarCat [49] project used a Soli sensor to detect different materials and user body locations. In contrast to the Walabot sensor used in our work, the Soli sensor has an operating frequency of 60 GHz, which gives it a high precision for fine details at short ranges. In long-range applications, Doppler radar range sensors were used by Liu *et al.* [25] to detect fall events of patients in long-term care facilities. Finally, a beam-scanning radar system was proposed by Wang *et al.* [41] for human location detection and detecting whether a person is sitting or standing. Our system, which we present next, extends this previous work with recognition of more detailed activities relevant for retail and workplace applications.

### 2.3 Classification Methods for RF-radar based Recognition

Adib *et al.* [1] and Lin *et al.* [24] used Support Vector Machine (SVM) classification for user identification, with Lin *et al.* also comparing the performance of k-nearest neighbor (KNN). Liu *et al.* similarly used SVM and KNN models to detect fall events of patients in long-term care facilities [25]. For material identification, Wang *et al.* [42] used KNN, while Yeo *et al.* [49] experimented with SVM and Random Forest classification approaches. Random Forests classifiers were also used by Lien *et al.* to detect small finger gestures with the Google Soli sensor [23]. Wei *et al.* [45] compared the performance of a Sparse Representation Classification (SRC) to SVM and KNN. In [43], Wang *et al.* experimented with a variety of algorithms, including Random Forests and Hidden Markov Models (HMM) before switching to deep learning with Convolutional Neural Network (CNN). Similar to much of this prior work, in this work we explore the recognition performance of a set of popular classification algorithms, including SVM, KNN, Random Forest, Naïve Bayes, Logistic Regression, and XGBoost used for activity recognition.

## 3 AN UNOBTRUSIVE ACTIVITY-SENSING SOLUTION

In this section, we introduce an unobtrusive, RF-sensor-based activity-recognition solution that can be easily deployed under a work surface. As discussed in the introduction, at this project's onset, our focus was on activity recognition for a convenience-store checkout counter, expanded later to include the applicability to an office environment, and then to a retail showroom environment. We contrast our solution to IMU-based activity recognition and investigate the potential value of combining RF sensing and a worn IMU, since in our target store environment employees wear a smartwatch as part of a different pilot deployment.

The section is organized as follows: We first describe the system components and setup used in our experiments. We next describe the construction of activity-recognition classifiers and provide analysis of how using multiple 3D projections of RF-signal can improve recognition accuracy. Finally, we demonstrate the use of IMU and RF-sensor data for attributing action to a particular user in multi-user scenarios.



Fig. 2. A Walabot Pro RF-radar sensor mounted under a work surface using a custom 3D-printed mount (left) and sample output image from the sensor (right). The sensor data are 3D projected into a Spherical coordinate system  $\varphi$ -R (thus should not be interpreted as a simple X,Y representation of the environment).

### 3.1 Implementation

For capturing RF data above and around a work surface, we use a Walabot Pro RF sensor<sup>1</sup>. The sensor operates in the frequency range of 3.3 – 10.3 GHz, which allows the sensor to “see” through different light dielectric materials such as wood, drywall, glass, etc. We mount the Walabot sensor under the work surface using a custom 3D-printed mount (see Figure 2, left). The Walabot Pro sensor uses a 72mm x 140mm array of 18 antennas to generate a 3D representation by measuring the strength of the reflected signal from the area above the antenna array. The raw point cloud is 3D-projected into a 75 x 37-pixel image in a Spherical coordinate system. That is, the intensity of each pixel denotes the reflected energy received at that point, represented in *angle* and *distance*<sup>2</sup> (see Figure 2, right). The average transmission power of the RF signal is -16 dbm. We note that based on the manufacturer’s reports, no known health risks are associated with the device<sup>3</sup>. While the Walabot Pro sensor can capture data at 30fps, we presently capture RF sensor data at a sampling rate of 2fps. As shown later, this rate provides sufficient recognition accuracy and allows keeping the size of collected data small. For comparing to activity recognition using IMU data, we use the Microsoft Band 2<sup>4</sup>. We log IMU sensor data from the wearable sensor, which consists of accelerometer and gyroscope data. The sampling rate of both these sensors is 62 Hz. Data are streamed from the Microsoft Band 2 over Bluetooth to a smartphone running Android OS and then transferred to a server for further processing. Lastly, we utilize the scikit-learn machine learning library [27] to train and test the different models presented next.

### 3.2 Activity Recognition: Checkout Counter

We first focus on recognizing activities performed at the convenience store’s checkout counter. For this purpose, we selected six common activities performed by the clerk in the checkout process (illustrated in Figure 3). The activities are: *Bowing to customer* (customary in Japan, where our project originated), *Scanning item barcodes with barcode scanner*, *Placing items into a bag*, *Passing or receiving an object on the counter* (this can be an item or payment), *Interacting with*

<sup>1</sup><https://walabot.com/>

<sup>2</sup>[https://en.wikipedia.org/wiki/Spherical\\_coordinate\\_system](https://en.wikipedia.org/wiki/Spherical_coordinate_system)

<sup>3</sup><https://walabot.com/walabot-tech-brief-416.pdf>

<sup>4</sup><https://www.microsoft.com/microsoft-band/en-us>

*the cash register, and Passing or receiving an object above the counter* (this can be an item or payment). For these six activities, both clerk and one or more customers are typically present. Finally, we also included a seventh, '*'Idle'*' activity in which the clerk and customer are not performing any of the above activities, or the clerk is standing by herself or himself at the counter. It should be noted that while we focus on these seven activities, our approach generalizes beyond the specific activities selected in this experiment. Alternate activities can also be modeled but would require the collection of corresponding training data.

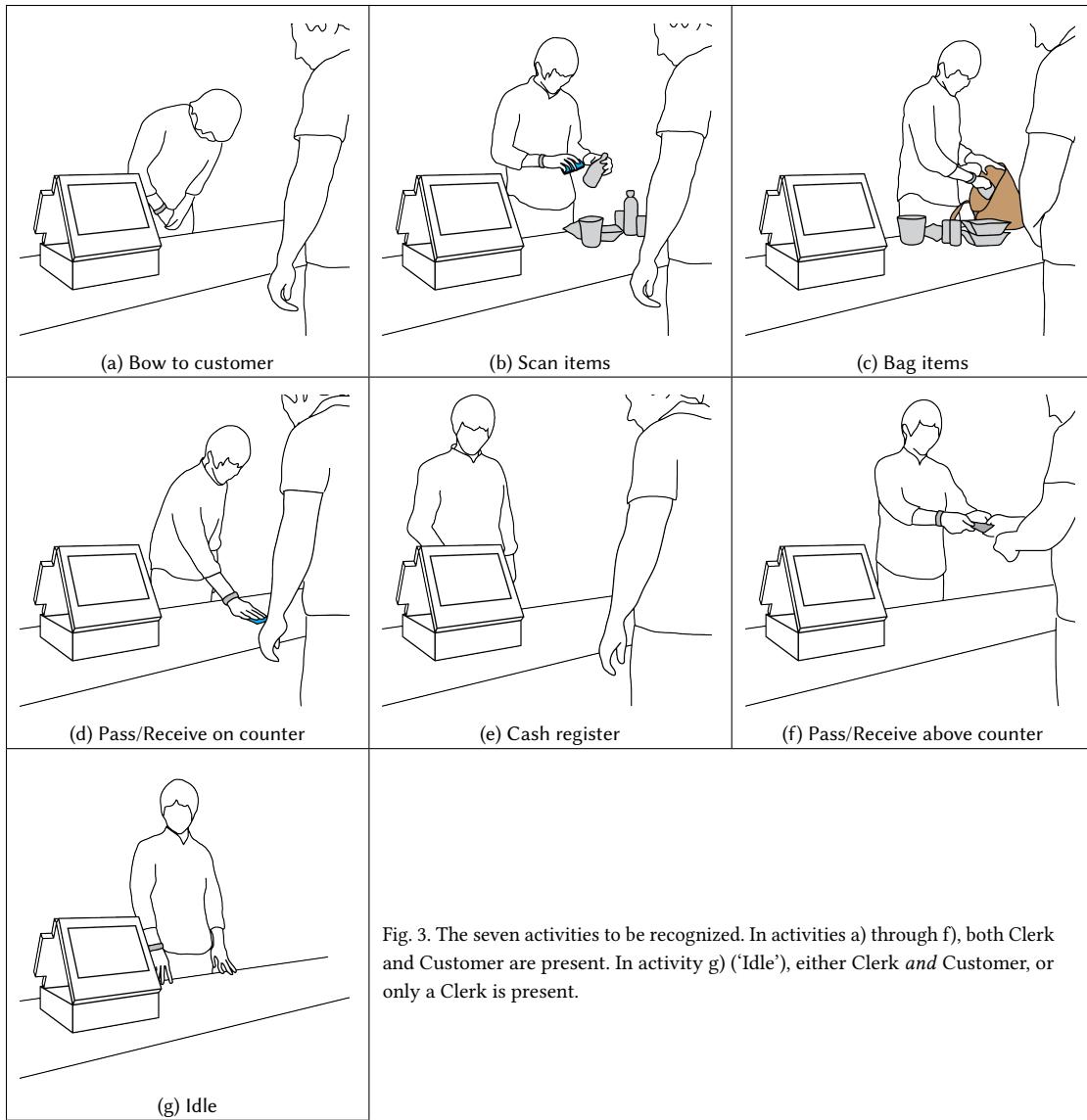


Fig. 3. The seven activities to be recognized. In activities a) through f), both Clerk and Customer are present. In activity g) ('Idle'), either Clerk and Customer, or only a Clerk is present.

**3.2.1 Setup.** Unfortunately, to date, collecting data in the real store was not possible. Thus, a convenience-store counter was simulated in our lab using a 182cm x 91cm work-surface, as shown in Figure 4 (as described later, data for deskwork activity recognition, however, was collected in real offices). The Point-of-Sale (POS) was simulated using a large metal enclosure with a propped-up tablet. When the system is initialized, the first five seconds are used for calibration, and any objects present during those five seconds are considered background (thus, the POS is “invisible” to the system). Because in our use-case environment customers are expected or allowed only on one side of the counter, we designate one side of the counter the “Clerk” side (Figure 4, bottom), and the other side the “Customer” side (Figure 4, top). Since people stand on both sides of the counter (a clerk on one side, and customer(s) on the other), the field of view of the Walabot’s sensor was set to its maximum of 180 degrees (a hemisphere).

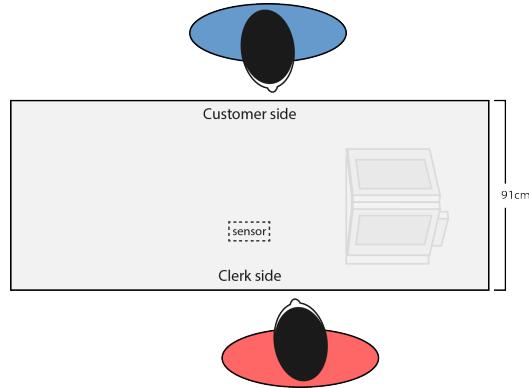


Fig. 4. The checkout counter environment simulated in our lab.

**3.2.2 Data.** Ten participants from our lab (1 woman, 9 men) volunteered to play the roles of Clerk and Customer, taking turns playing each role. When playing the role of a Clerk, participants wore a Microsoft Band 2 on the wrist of his or her dominant hand and stood on the appropriate side of the work surface. Participants then performed each of the 7 activities repeatedly for one minute (e.g., passing a grocery item back and forth). Participants were allowed to move left and right, but not move to the other side of the work surface. Each activity instance (e.g., a single bow, scanning a single item, etc.) took approximately one to three seconds (i.e., participants performed 30 bows and scanned an item 35 times).

We collected 10 sets of data. To collect data for different scenarios, in seven of the sets, we had one “clerk” with one “customer”, and in three cases, we had one “clerk” with two “customers”. Furthermore, participants had different body types (in height and weight) and different hand dominance (i.e. left handed and right handed) to ensure a varied dataset.

**3.2.3 Data Preprocessing.** A total of 17,862 RF samples were collected from 10 participants who performed the 7 activities. IMU data from the MS Band 2 were captured over Bluetooth at 62 Hz. Based on the short duration of each activity instance, data samples for learning were generated using a 1500 msec sliding window, as the mean intensity from each of 3 consecutive RF-sensor samples. Statistical features such as median, standard deviation, min, max, difference between max and min value, slope of the time series samples and a movement feature (which is extracted by taking the difference between the data sample and the mean value) were computed from the raw IMU data. These features were generated for each dimension of the IMU’s accelerometer and gyroscope data.

**3.2.4 Results.** We compare activity recognition performance when using the RF-sensor data only, the IMU data only, and when using a combination of the two. We report results from 10-fold cross validation evaluation; given the use of a sliding window in generating the data, we split the data into training and testing sets in a way that guaranteed to not have any overlapping samples in the sets.

The activity recognition performance was tested using five different classification techniques listed in Table 1. Our results show that classification accuracy using the RF sensor alone is high, much higher than that of baseline classification using the IMU data. This suggests the viability of this solution for unobtrusive activity recognition. Interestingly, we find that combining IMU and RF-sensor data does not provide meaningful accuracy gains. Figure 5 shows a confusion matrix for the seven clerk activities produced by the SVM classifier.

Table 1. Activity recognition accuracy of different classifiers using different sensors combination based on a 10-fold cross validation.

	<b>10-fold</b>	<b>Baseline (IMU Only)</b>			<b>RF</b>	<b>RF + IMU</b>
Logistic Regression		0.384			0.880	0.893
Naive Bayes		0.307			0.562	0.564
kNN		0.585			0.880	0.884
<b>SVM</b>		0.507			<b>0.905</b>	<b>0.917</b>
Random Forest		<b>0.593</b>			0.818	0.832

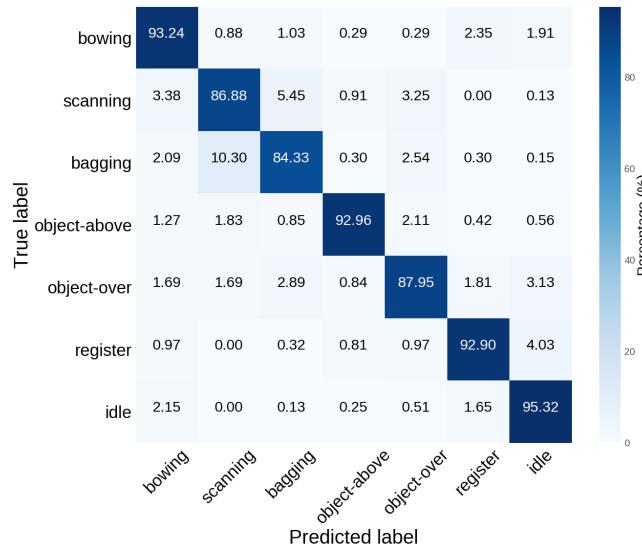


Fig. 5. Confusion matrix of activity recognition based on RF data only produced by the SVM classifier using 10-fold cross validation.

### 3.3 Multiple Projections of RF-signal for Improved Accuracy

One interesting area for investigation is whether extracting additional projections of RF data from 3D to 2D space would improve recognition accuracy further. Figure 6 illustrates the Walabot sensor's coordinate system. By default, the Walabot sensor outputs a single projection from 3D into 2D space, represented in angle  $\varphi$  and distance  $R$  (Figure 6,

left). This space is the plane corresponding to the surface on which the sensor is mounted. In Figure 6, this is the X-Y plane. It is, however, possible to generate additional projections into 2D space using angle  $\Theta$  (see Figure 6, middle). For this test, we generate 2 additional RF signal projections: A projection ( $\Theta\text{-}R$ ) that is perpendicular to the default slice ( $\varphi\text{-}R$ ), and a projection ( $\Theta\text{-}\varphi$ ) that is a cross-section of ( $\Theta\text{-}R$ ) and ( $\varphi\text{-}R$ ). The additional projections correspond to the Z-Y and Z-X planes in the Figure 6. The 3 projections are concatenated into a single image with a resolution of 187 x 37 pixels (see Figure 6, right). We evaluated this approach by performing a 10-fold cross validation on the same set of models as before.

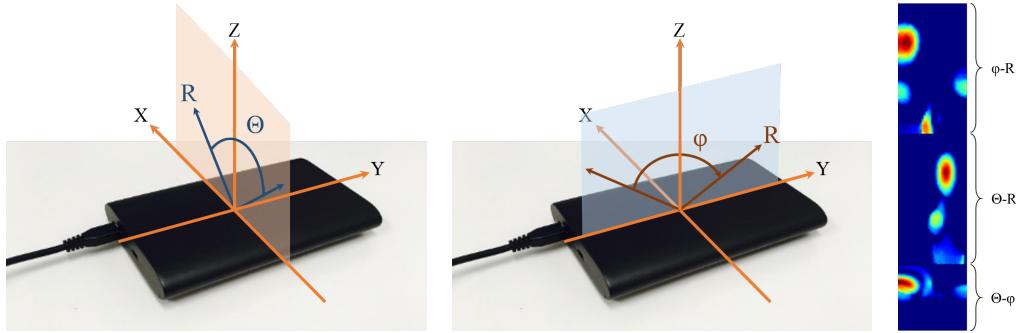


Fig. 6. The Walabot sensor coordinate system (left and middle), representing a half hemisphere area in which RF signal is transmitted and sensed. On right, the default 3D projection ( $\varphi\text{-}R$ , top) from Figure 2, with two additional extracted projections ( $\Theta\text{-}R$ , center and  $\Theta\text{-}\varphi$ , bottom) concatenated.

**3.3.1 Results.** Using the 3 projections resulted in overall improved accuracy, as can be seen in Table 2. For the best classifier (SVM), using 3 projections instead of the default projection reduced the error by 45%. Adding IMU data yielded only minor gains over the 3 projections of RF signal.

Figure 7 shows a confusion matrix for the seven activities produced by the SVM classifier. Looking more closely at performance on the different activities, we see that, of all the activities, scanning and bagging were the most likely to be confused. Still, each of these activities is recognized correctly in more than 91% of the samples.

The improvement in overall accuracy suggests that the additional 2 projections capture important information present in different directions that isn't captured using a single projection. To summarize, we are able to reach activity recognition rates of 95% for the seven activities using the RF sensor alone. As we show later, while IMU data did not generate a significant improvement in activity recognition accuracy, it can still play a role in an overall system for associating an activity with the user who performed it.

**3.3.2 User Independence Test.** Finally, to test how well the solution performs on new users, a leave-one-user-out evaluation was conducted on the retail activity dataset – recall that the simulated checkout environment remained consistent between users. This is important, for example, in the case of a new worker at the store. For each of the 10 users, the models are trained using data from 9 users and tested on the left-out user. We used the data generated with 3 projections of RF signal for this evaluation. Table 3 shows the accuracy achieved using different classification models. Highest accuracy was achieved by the Logistic Regression classifier over all 10 users, at 98.4%. This result is very positive as it suggests that a system may be able to be bootstrapped with training data and yield reasonably high recognition accuracy for new users.

Table 2. Comparing retail-activity recognition accuracy using 1 or 3 slices of RF signal, based on a 10-fold cross validation. Additional slices lead to accuracy gains.

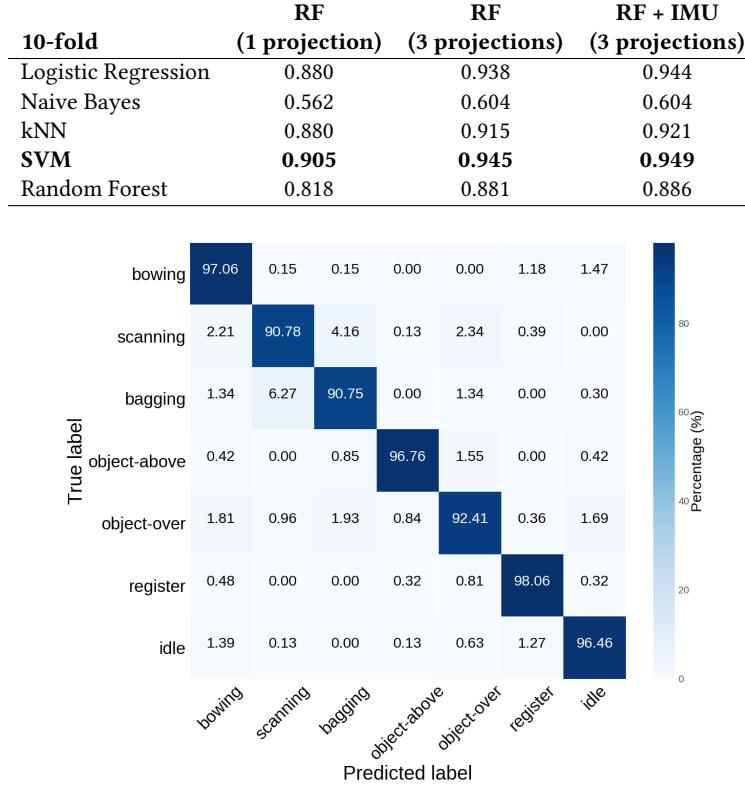


Fig. 7. Confusion matrix of activity recognition based on 3 slices of RF data only produced by the SVM classifier using 10-fold cross validation.

Table 3. Retail-activity classification accuracy using a leave-one-user-out validation.

Leave-one-out		RF + IMU
<b>Logistic Regression</b>		<b>0.984</b>
Naive Bayes		0.711
kNN		0.956
SVM		0.974
Random Forest		0.918

### 3.4 Activity Recognition: Office Deskwork

As discussed earlier, the ability to discreetly and unobtrusively recognize activities performed above a work surface could be valuable in other domains such as office environments, to allow intelligent systems to make better estimation of workers' state and availability. To investigate the applicability of the solution to office deskwork, we conducted a pilot exploration using our solution under office desks. We focused on six common deskwork activities, illustrated in Figure 8: *Writing (on paper)*, *Reading (from paper)*, *Eating*, *Drinking*, *Computer Work*, and *Idle* (e.g., reading off the

screen). Finally, we also included a condition with the user not at their desk (either in the office, or away) that we call *Not at Desk*, as a seventh activity.

**3.4.1 Setup.** This dataset was collected in participants' personal offices. As such, the dataset represents naturalistic environments, including desk clutter present in different people's offices. Furthermore, by capturing data over time, we account for how such environments change.

Following lessons from the previous section, three projections were generated for each RF sample. Since, unlike the counter scenario, in this environment most activities are performed on one side of the work surface by one person, the field of view of the Walabot's sensor was set to 90 degrees (from -45° to 45°; refer to Figure 6 for details related to field of view).

**3.4.2 Data.** We collected data from 6 participants, all members of our lab. Data were collected in each participant's office, with the Walabot sensor mounted to the underside of the desk, coarsely under the location of the participant's keyboard. To accommodate the natural changes in people's offices and desk environments, such as bringing or removing papers, personal electronics, different food and drinks, etc., we collected the full set of activities from each participant on five different occurrences throughout the workweek. In each data-collection session, participants performed each of the activities at a time for two minutes. Participant chose in which order they wanted to perform the activities.

The full dataset contained a total of 7 hours of RF sensor data (6 participants x 5 sessions x 2 minutes/activity x 7 activities), with *one full hour* of RF data for each activity. We note that even with this small number of participants, the data include rather diverse working styles: One of the participants uses an L-shaped desk with their computer at the intersection of the L. Another participant uses a sit-stand desk. Yet another participant uses an ergonomic pullout keyboard tray. (For this participant, the Walabot was mounted further back under the desk.)

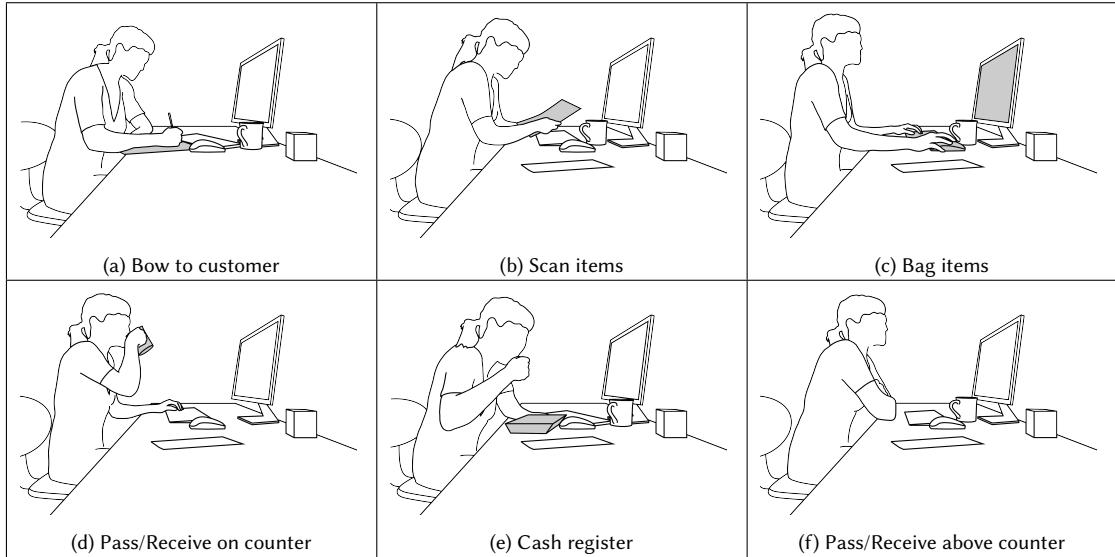


Fig. 8. The six office deskwork activities to be recognized.

Table 4. Classification accuracy for deskwork activities in participants' offices.

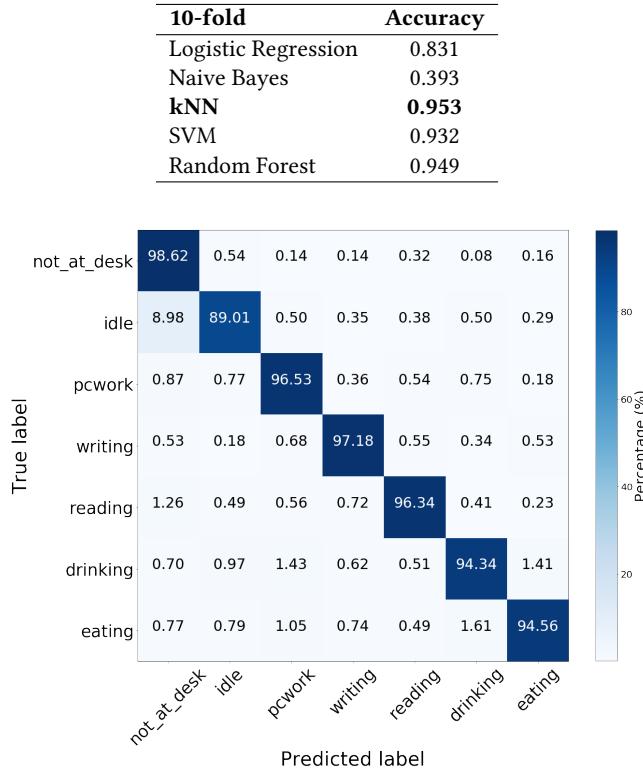


Fig. 9. Confusion matrix for office deskwork activity recognition predicted by KNN using 10-fold cross validation.

**3.4.3 Data Preprocessing.** As before, a data sample is generated for every 1,500msec window as the mean intensity for each pixel using a sliding window. From each session and each activity, we trim the first and last 1 seconds. The dataset contained a total of 23,687 samples.

**3.4.4 Results.** Table 4 demonstrates the performance of five classification approaches based on 10-fold cross validation. With the exception of Naïve Bayes, most of the models were able to correctly classify the 7 activities with accuracy in the range of 83–95%. K-nearest neighbor (KNN) and Random Forest models gave the highest accuracy. The confusion matrix for the KNN classifier is shown in Figure 9. As can be seen, in about 8–9% of the cases when a participant was idle, the models incorrectly classified them as not at their desk. It is worth noting that while detecting Computer Work can be easily done by installing a tool on the user's computer, correctly detecting that a user is reading or writing (not on their computer) can be greatly beneficial to tools that detect presence and availability, such as [6].

**3.4.5 User Independence Test.** To mirror our previous experiments, we examined how well the solution performs on new users in a new office through a leave-one-user-out evaluation. Unlike the retail activity dataset, where different participants performed activities in the same environment, in the deskwork activity dataset, participants' personal offices and personal desk setups varied greatly from one to the other. For each of the six participants, the models are

trained using data from five users and tested on the left-out user. We used the data generated with 3 projections of RF signal for this evaluation. As to be expected, given that each set was collected in a different environment, the accuracy of Leave-One-Out classification for the Deskwork activities is very low, with Naive Bayes yielding the best yet very low accuracy of 26.8%. Thus, whenever deploying the solution in a new office environment, training data should be collected and the model should be retrained or adapted.

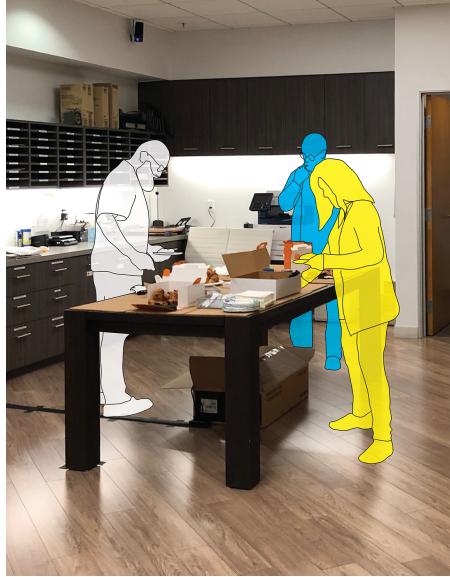


Fig. 10. Our data collection setup. Two RF-radar sensors were mounted to the underside of the table. In this picture, three lab members are at the table standing in zones 1, 3, and 6 in the Polar zone-system described in Figure 11a.

### 3.5 Coarse Position Estimation for Surfaces: Showroom Environment

In this section, we extend our proposed solution for performing presence detection and coarse position estimation around a surface. We investigate this in the work context of a retail showroom. This work context further provides the opportunity to explore the use of multiple sensors concurrently to achieve coverage over large surfaces.

Consider a retail showroom, similar to an Apple store or Microsoft store, where products are laid out on display on large tables for customers and visitors to experience. (Figure 1, right, shows customers and retail staff interacting with devices and each other in an Apple store). Understanding which products or product-categories customers spend time next to (and, even more importantly, which products receive less attention) can be useful for making important decisions about the showroom’s layout, the placement of products, and for analyzing potential a relationship between time spent by a customer and sales conversion. For example, analytics that show that customers spend very little time next to a particular product line can be used to decide to relocate the product to a more prominent location. Or analytics that show that despite an increase in time spent next to a particular product line there is no increase in sales, trained staff can be instructed to engage and assist customers in that particular showroom’s area.

In the field of Tabletop Computing, several systems have been demonstrated that aim to detect and track the presence and position of users around a tabletop computer. Tănase *et al.* [40], for example, used short-range general-purpose

optical sensors placed around the tabletop to estimate the coarse position of users. In [2], Annett *et al.* fitted the perimeter of a small tabletop computer with 34 long-range sensors spaced 3.3 cm apart. This allowed fine-grained user sensing and tracking in a small space.

Similar to [40] and [2], we investigate the viability of our solution to this application domain as an unobtrusive alternative to the use of cameras. One benefit of our approach in this context is that “observing” the scene from under the surface avoids issues of occlusion compared to, for example, mounting radar sensors or other imagers looking head on. Using only two sensors (compared to the complex setups in [2] and [40], that used 12 sensors and 138 sensors, respectively), we are able to cover a work surface much larger than those discussed in [2] and [40] thanks to the large coverage area of each sensor. This results in a greatly simplified overall setup.

**3.5.1 Test Environment.** We tested our solution using a very large table measuring 2.3-meters long by 0.95-meters deep. This table (shown in Figures 10 and 11) is similar to the tables seen in Figure 1 right, is made of solid wood, and its top surface is 4cm thick. Located in our lab’s common area in a path that receives much foot traffic, there are at least 1.3 meters on each side of the table.

Similar to display surfaces in retail showrooms (i.e., tables with products on display), our test surface is too large for a single sensor to provide sufficient coverage. Thus, to be able to cover the large surface of the table, we used two sensors simultaneously. Sensors were mounted at the center of each half of the table and oriented parallel to the long dimension of the table (see Figure 11). RF data from the sensors were recorded at a sampling rate of 1 Hz. To circumvent data drift, both sensors were recalibrated every 20 minutes (during calibration, 15 RF samples are used to estimate an empty scene).

**3.5.2 Collecting Natural Behavior Data.** In order to evaluate our approach, we chose to collect natural user behavior, with people positioning themselves naturally around the table surface. On four separate mornings, we purchased and placed food (pastries or sandwiches) as well as small plates on the table and invited all lab members using an email message. To simulate interaction with different products or product categories as in a retail showroom, on each of these mornings, we bought six different types of items and placed them strategically on the table in the six areas (or “zones”) shown in Figure 11a. The food items were intentionally kept in containers such that they are not easily accessible from across the table, thus requiring people to walk around the table to “investigate” and choose items. Figure 10 shows the table with food and three lab members. We informed lab members in email of the data being collected.

**3.5.3 Ground truth.** In order to verify our system’s performance against ground truth, we deployed a Microsoft Kinect 2 depth camera<sup>5</sup> mounted close to the ceiling and with view of the table and surrounding area. The Kinect was used to perform skeleton tracking at 30fps. We recorded the 3D coordinates of the top of the head of the skeleton’s joints model. When synchronizing RF data (captured once a second) and 3D ground-truth data (captured 30 times per second), we average the 3D position recorded. We then transform 3D ground-truth data to a 2D coordinate system as follows: We first captured the head-joint 3D coordinates for a person standing at each of the table’s four corners. From these coordinates, we compute rotation and transformation matrices. We then scale the data such that the four points (standing at the table’s corners) correspond to coordinates (0,0), (0,1), (1,0) and (1,1). We then use the translation and rotation matrices and scaling parameters to project all ground-truth data to the table’s 2D plane.

---

<sup>5</sup><https://www.microsoft.com/en-us/store/d/kinect-sensor-for-xbox-one/91hq5578vksc>

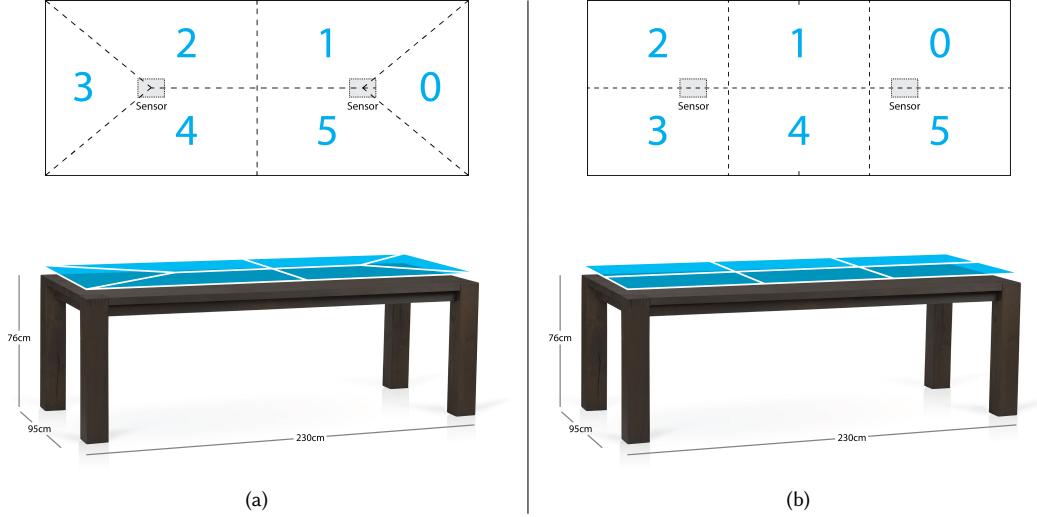


Fig. 11. The test environment: a 230cm x 95cm table with two mounted sensors. The two 6-zone systems used for coarse position estimation shown. A Polar zone system on left, and a Grid zone system on right.

**3.5.4 Zone Systems.** As mentioned above, the goal in this section is for a system to provide data on which products or product lines in a showroom, customers spend time next to. As such, in our experiments, we focus on estimating the presence and number of customers next to six zones or regions of the surface. We note that in our tests we only attempt to estimate the presence of customers and not their physical orientation (i.e., which direction they are facing). In our data collection, we focused on the zone system depicted in Figure 11a, which we refer to as the "*Polar*" zone system, with two zones on each side of the table's long dimension and one zone at each table's end. In our experiments, we also examine position estimation using the "*Grid*" zone system as shown in Figure 11b. We note that the Polar zone-system, which considers positions at the table's ends, goes beyond what would be needed in many real-world showroom configurations – at Apple's stores, for example, tables showcase products only along the table's long dimension.

**3.5.5 Defining the Interactive Area.** We consider the "*Interactive Area*" to be the area around the table within which a person is assumed to be next to the table and possibly looking at, or interacting with items on the table. The interactive area was estimated empirically for our table as follows: in addition to the four coordinates representing standing at the table's four corners, capture a second set of four positions where a person is standing just *beyond arm's reach* of the table. These points represent the closest positions "outside" the interactive area. In our setup, we estimated this boundary to be 35 centimeters on each side of the table.

**3.5.6 Data overview.** During each of the four data-collection days, data were collected starting when the food items were placed on the table until they were gone. This resulted in a total of 5 hours and 24 minutes of RF-sensor data and ground truth data – 19,413 RF data samples in total. In 32% of the data (6,236 of the samples – 1 hour and 44 minutes) at least one person was observed by the Kinect sensor. With the boundary of the Interactive Area set at 35cm, 4,890 of the 6,236 samples have at least one person *inside* the Interactive Area. Figure 12 shows the recorded ground-truth positions of people around the table across the full dataset. The table is represented as a yellow rectangle. A blue dot

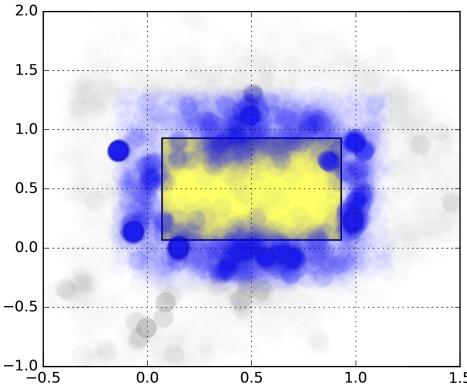


Fig. 12. A scatter plot showing ground-truth positions of people around the table in our data based on captured 3D coordinates of a person's head. The table is illustrated as a yellow rectangle. Samples inside the interactive area are shown in blue. Samples outside the interactive shown in gray. The interactive area boundary was set to 35cm on each side.

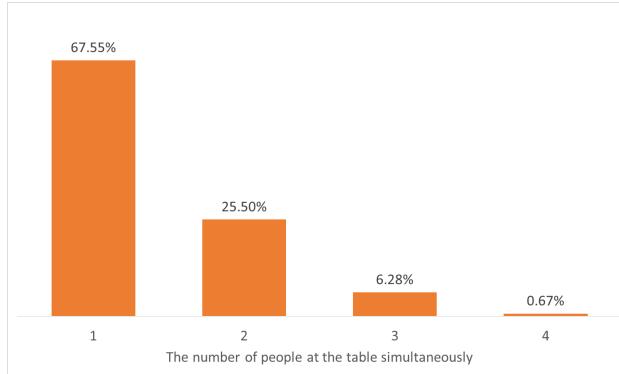


Fig. 13. The frequency of multiple people simultaneously inside the Interactive Area based on a 35cm boundary around the table.

represents a person standing within the interactive area and a gray dot represents a person outside the interactive area (with the boundary set to 35cm).

Of the 4,890 sample (in blue in Figure 12), at most 4 people were inside the Interactive Area at the same time. This occurred in less than 1% of cases. Figure 13 shows the distribution of multiple people inside the interactive area at the same time. Indeed, a single person at the table was most common (68% of cases). Furthermore, in our data, we observed only 95 total samples (1.4% of the data) in which more than one person occupied *the same zone*. Thus, since two people in the same zone was so rare, the position-estimation experiments described below do not attempt to estimate the number of people inside each zone, rather only estimate whether a zone is occupied or not.

While a key benefit of collecting natural behavior data is that the data are realistic, one drawback is that some of the possible configurations of people standing at the table may not have been observed. For our six-zone systems, the total number of possible configurations of people occupying different zones is  $2^6 = 64$ . In our data, however, only 41 of these 64 configurations are represented. Thus, instead of attempting to train the 64-class classifier, we instead train a

Table 5. Accuracy of binary classifiers for each of the six zones of the Polar zone system, comparing six learning methods and the prior probability.

Zone	Prior	Random		Logistic		Naive	
		Forest	XGBoost	SVM	Regression	kNN	Bayes
0	0.822	0.966	0.968	0.966	0.967	0.963	0.836
1	0.852	0.930	0.943	0.913	0.941	0.929	0.817
2	0.837	0.934	0.938	0.921	0.937	0.936	0.830
3	0.872	0.980	0.981	0.981	0.978	0.975	0.904
4	0.809	0.950	0.956	0.954	0.954	0.955	0.876
5	0.839	0.939	0.939	0.933	0.942	0.931	0.822
<b>Average</b>	<b>0.839</b>	<b>0.950</b>	<b>0.954</b>	<b>0.944</b>	<b>0.953</b>	<b>0.948</b>	<b>0.847</b>

Table 6. Accuracy of binary classifiers for each of the six zones of the Grid zone system, comparing six learning methods and the prior probability.

Zone	Prior	Random		Logistic		Naive	
		Forest	XGBoost	SVM	Regression	kNN	Bayes
0	0.826	0.948	0.959	0.939	0.947	0.941	0.828
1	0.839	0.951	0.957	0.945	0.951	0.948	0.822
2	0.845	0.950	0.957	0.953	0.948	0.955	0.832
3	0.861	0.954	0.956	0.953	0.954	0.953	0.875
4	0.826	0.951	0.955	0.944	0.952	0.952	0.844
5	0.829	0.947	0.948	0.953	0.956	0.947	0.848
<b>Average</b>	<b>0.838</b>	<b>0.950</b>	<b>0.955</b>	<b>0.948</b>	<b>0.951</b>	<b>0.949</b>	<b>0.841</b>

dedicated classifier for each of the six zones predicting whether that zone is occupied or not. The predictions from the 6 binary classification models can then be aggregated to produce a final estimate.

**3.5.7 Classifier training.** Binary classifiers for every zone were trained on subsets of 5,000 samples, then tested on a subset of 1,660 samples. Because the collected data was heavily skewed towards samples with no person present (around 70% of the samples), we subsampled the data when splitting the training and test sets. The training and test sets were generated as follows:

Each test set was constructed by first randomly sampling 100 positive examples for each zone, then adding 1,560 randomly-sampled negative examples. Negative examples include 460 samples with people standing in other zones, 200 samples with people outside of the Interactive Area, and 900 samples with no person present at all. It was essential to have all types of negative samples represented in the test sets. The training set was composed of all remaining positive samples as well as a balanced, randomly sampled set of negative data for the other three categories. We note that for classifier training, balancing these classes in a training set greatly improves accuracy for all available classes on the representative test sets.

Finally, to reduce the feature dimensionality of the combined RF-sensor samples, we applied principal components analysis (PCA) to the image data, to project the images to 90 dimensions. The projected data was scaled according to the corresponding singular values to generate unit variance features. We experimented with computing PCA over the full training sets, as well as over the training subset for Present Inside the Interactive Area. In the end, we computed 90 dimensional PCA subspaces for both, and concatenated the projections to reduce the feature dimensionality from 13,246 to 180. These projections were then applied to the test set for classification.

**3.5.8 Results.** We first report position estimation accuracy for the two 6-zone systems (Polar vs. Grid) using 6 methods: Random Forest, Extreme Gradient Boosting (XGBoost) [11], Support Vector Machine (SVM), Logistic Regression, kNN, and Naive Bayes. We then present an evaluation of accuracy as a factor of the size of the Interactive Area.

For the Polar zone system, Table 5 shows the classification performance for each zone classifier as well as the prior probability (the probability of being right when always guessing the most frequent class). As can be seen, all the zone classifiers except for Naive Bayes were able to estimate the presence or absence of a person in the zone with over 91% accuracy. Comparing to the prior probability – 84% on average – the zone classifiers perform significantly better. Classification performance for the Grid zone system, shown in Table 6, is similarly high.

As an illustration of how our classifiers can be used to provide an overview of users' interaction with items on the table, Figure 14 visualizes the frequency of each of the 6 zones classified as occupied in the Polar zone system. While we do not currently track the state of the zones between frames (as [2] do), this feature could be added easily.

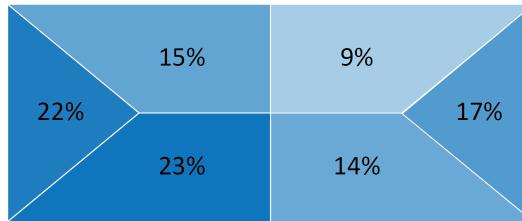


Fig. 14. The frequency of each of the table's zones classified as occupied using the six binary-classifiers approach.

We next examined more carefully our solution's ability to distinguish between people inside or outside of the Interactive Area. Specifically, we wanted to get a better understanding of the limits of the information captured by the sensors.

**3.5.9 Sensor Limits and the Interactive Area.** To understand how people's distance from the table impacts the system's ability to perform classification, we generated a set of 3-way classifiers, each predicting whether any person is Present Inside the Interactive Area, Present Outside the Interactive Area, or No Person is present. We then examine prediction accuracy as we vary the bounds of the Interactive Area in 5 centimeter increments, from 0cm to 35cm.

Table 7 and Figure 15 show the relationship between the size of the Interactive Area and presence classification performance. Here we discuss the precision, recall, and F1-score for detection of *Present Inside* the Interactive Area

Table 7. The relationship between the size of the Interactive Area (in boundary size) and presence-classification performance. Showing measures for "Inside the Interactive Area" class.

Boundary size	Precision	Recall	F1-score	Overall accuracy
0cm	0.88	0.95	0.91	0.81
5cm	0.85	0.93	0.89	0.81
10cm	0.82	0.94	0.88	0.79
15cm	0.76	0.92	0.83	0.78
20cm	0.75	0.93	0.83	0.78
25cm	0.72	0.91	0.81	0.75
30cm	0.72	0.92	0.81	0.76
35cm	0.70	0.89	0.78	0.73

samples using gradient boosting. As can be seen, the F1-scores demonstrate that overall detection performance decays slowly as the size of Interactive Area grows. However, the precision of detecting that a person is present inside the zone degrades noticeably as the area increases in both the Grid Zone and Polar Zone systems. The decrease in recall is less pronounced. This suggests that false positive detection errors account for the overall loss in performance. In practical terms, the more the Interactive Area extends from the table, more samples from the Outside the Interactive Area and No Person Present are misclassified as a person present at the table. Precision may thus be a useful metric for best defining the bounds of the Interactive Area to suit specific deployments and analytic objectives.

To summarize, our results show the potential usefulness of our solution to a retail showroom use-case. However, the results further suggest that accurately estimating that a person is sufficiently close the table to be considered potentially interacting with items on it can be challenging. One potential solution could be to include temporal smoothing. This would potentially further help determine that a user is actually engaged with items on display.

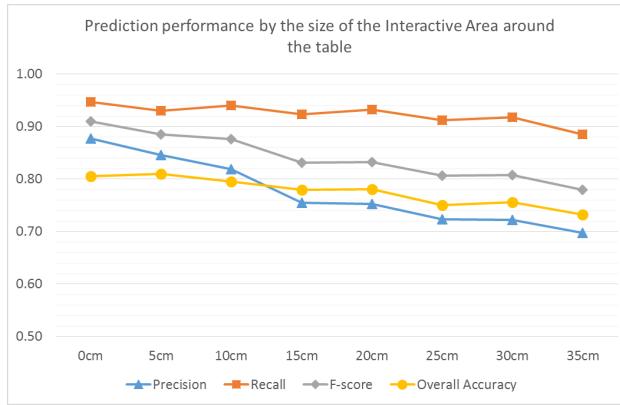


Fig. 15. Presence-classification performance as a factor of the size of the Interactive Area.

#### 4 ATTRIBUTING ACTION IN MULTI-USER SCENARIOS

One challenge with using RF sensing for activity recognition is that when more than one user occupies the workspace (for example, more than one clerk is working at the store), the system cannot tell which user is the one performing the activity that was recognized. In this section, we present our experiments for identifying the user performing an activity above the workspace.

One approach to attempting to identify the user performing an activity in a multi-user scenario would be to train a “person classifier” using the RF sensor, similar to [1]. In [1], an antenna mounted behind a wall is used to identify human body parts from RF snapshots across time to distinguish between users. In our setup, however, the sensor is mounted under the work surface, getting only limited view of the users’ body. Another approach is to use additional sensors to identify users, such as IMU data from smartwatches worn by users. As such, the system tries to identify which smartwatch was worn by the user performing the activity. This is analogous to work by Cho *et al.* [12] and Wilson and Benko [47] that correlated images from cameras and IMU data.

#### 4.1 Data

To simulate a multi-user scenario in which two users each wear a smartwatch, we used the RF-sensor and IMU dataset collected earlier to generate the following: Each RF-data sample (representing a 1,500msec window) in the dataset was paired both with IMU data captured at the same time (i.e. “*Matching*”), but also with IMU data chosen randomly from a different activity and user (i.e. “*Mismatching*”), simulating the activity of a different user. By way of illustration, consider RF data sample captured while a user was scanning an item, paired with its corresponding IMU data as well as with mismatching IMU data recorded during a “*Bowing*” activity. The system’s goal is to correctly link the RF data and its matching IMU data.

For this dataset, we used 3,930 RF data samples. For each sample, we generated a matching pair using the corresponding IMU data, as well as six additional pairs with IMU data from each of the remaining six activities. The full dataset included 27,510 pairs. A balanced set was also generated by randomly picking for each RF sample just one of the mismatching pairs. The balanced set included 7,860 pairs.

We report two experiments, each using a different method for determining the match between IMU and RF samples.

#### 4.2 Method I: Direct Learning

In a first method, supervised learning is used to directly learn the match between RF and IMU data. Specifically, each pair of matching RF and IMU data is treated as a single sample and labeled ‘*Matching*’. Each pair of mismatched RF and IMU data are treated as a single sample labeled ‘*Mismatching*’. We used the balanced dataset with equal number of *Matching* and *Mismatching* samples, 7,860 in total.

We trained classifiers using five different classification techniques used in the previous sections and evaluated them using 10-fold cross validation. Classification results are shown in Table 8.

As can be seen, classification accuracy is modest, with some not performing better than random. It is quite possible that this low accuracy corresponds directly to the low activity recognition accuracy using only IMU data that we observed earlier. We note that on the imbalanced dataset, F-score for the classifiers is especially low, ranging from 0.0 to 0.37. This is not surprising given the baseline accuracy of 0.86 when always guessing that a pair is mismatching.

We next explored a different method that relies on the high accuracy of the activity recognition models presented earlier, and using the agreement between classifiers for determining whether sensor data pairs represent the same activity or not.

Table 8. Recognizing matching vs. mismatching pairs of RF and IMU data, based on a 10-fold cross validation. Trained on a balanced dataset (i.e. baseline accuracy is 0.5).

<b>10-Fold</b>	<b>Accuracy</b>	<b>F-score</b>	<b>F-score on imbalanced set</b>
Logistic Regression	0.500	0.52	0.00
Naive Bayes	0.501	0.54	0.00
k-NN	0.661	0.67	<b>0.37</b>
SVM	<b>0.688</b>	<b>0.70</b>	0.04
Random Forest	0.500	0.52	0.00

### 4.3 Method II: Classification Matching

As described above, our attempt to directly learn whether sensor-data pairs match or not did not succeed. We thus decided to pursue a different approach that takes advantage of the high classification accuracy of our activity recognition models. Specifically, we implemented a 4-step process for matching RF and IMU data using activity recognition classification:

In step 1, the system performs activity recognition on the RF data alone, using an SVM classifier (this classifier was used as it provided the highest recognition accuracy). The activity classified with the highest likelihood is recorded, as denoted in equation 1:

$$\text{Activity} = \arg \max(p(\text{activity}|\text{RF})) \quad (1)$$

In step 2, the system uses a Random Forest model to perform activity recognition on both matching and mismatching IMU data samples (the Random Forest classifier was chosen as it had the highest accuracy for IMU-only data). Recall that the system does not know which IMU sample is the matching one. For each IMU sample, the likelihood score for the Activity classified on the RF data is recorded, as denoted in equation 2 (continuing with our illustration, the likelihood of “Scanning”, returned by each IMU data).

$$M_1 = p(\text{Activity}|\text{IMU}_1), M_2 = p(\text{Activity}|\text{IMU}_2) \quad (2)$$

In step 3, the system returns the IMU sample that produced a higher likelihood score as “Matching”, as denoted in equation 3. In case the likelihood produced for both IMU samples is identical, the system returns “*Matching IMU Unknown*”:

$$\text{Matching} = \arg \max_{\text{IMU}_1, \text{IMU}_2}, \text{IMU}(M_1, M_2) \quad (3)$$

In the final step, the system returns a confidence score, denoted in equation 4, calculated as the absolute difference between the likelihoods returned for each IMU data sample. As we show later, this score is a good measure for the confidence in the system’s choice between the two IMU samples.

$$\text{Confidence} = |M_1 - M_2| \quad (4)$$

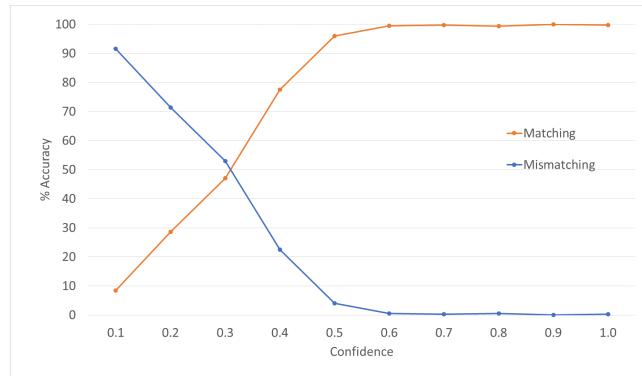


Fig. 16. Distribution of prediction accuracy by prediction confidence, calculated as the absolute difference in likelihood score between both IMU samples.

**4.3.1 Results.** We report the results of a 10-fold cross-validation evaluation. Our system correctly selected the matching IMU data in 64% of cases. Figure 16 shows the classification accuracy at different classification-confidence bands. It is apparent that when the confidence score returned by the system is above 0.4, the classification is likely to be correct. In fact, overall classification accuracy is 88% for all samples with a confidence greater than 0.2 (true for 66% of the data). This result shows our approach performs well on a large segment of the data. This result further shows that while the overall accuracy of Method I and Method II is comparable, the confidence score provided by Method II is a highly valuable piece of information. Still, however, for 34% of the samples the system had only low confidence in its prediction, suggesting that further investigation is needed.

## 5 DISCUSSION & FUTURE WORK

The results, presented in the previous section, illustrate how activities performed above a work surface can be recognized discreetly and unobtrusively, without requiring a camera. We further show that additional projection of RF signal greatly improves recognition accuracy. We also show that multiple sensors can be used to produce good coarse-position estimation around large surfaces. Finally, we also demonstrate the use of IMU data from a worn sensor for identifying the person performing the recognized activity in multi-user scenarios. We note, however, that asking users to wear a smartwatch with an IMU would reduce the level of the system’s unobtrusiveness.

One of the key areas for consideration and discussion is the notion of unobtrusive and discreet sensing and recognition. As described in the introduction, we investigated the use of RF-radar sensing as a potential replacement for cameras for activity recognition. Indeed, the use of cameras can be prohibitive in many environments, since the data they produce (stills and videos) is so readily interpretable by humans. However, continuous sensing and continuous activity recognition is only as unobtrusive as the intended use of the data. That is, users are likely to feel more comfortable with data captured by an RF sensor than being observed by a camera (particularly if shown the output of the RF sensor itself). However, particularly in private and semi-private environments, people must still be made aware that a system continuously monitors activities, aware of the system’ effective sensing area, and how the system’s output will be used.

On the technical side, in our current work, samples were created as snapshots from the IMU and RF sensor. However, many of the activities we are interested in do have a temporal component, especially in the convenience-store counter environment (e.g., a scanning motion, a bagging motion). Thus, while the accuracy of our classifiers was very high, future similar solutions may need to explore features contained in the time domain.

One interesting question that came up during this work but is not currently addressed is that of activities that are performed above the work surface then transition off it (or start off the work surface and continue above it). For example, moving a box or a bag on and off a counter. It is possible that for such activities, data from the worn IMU sensor could provide a significant contribution. This would allow the system to more accurately determine the beginnings and completions of activities.

### 5.1 Future Research Directions

A potential extension for the work could involve attempting to include knowledge of the materials used or placed on the work surface. This could be done, for example, in a method similar to that used by [49]. Furthermore, the physical relationship between people and objects could be used to enhance both recognition and interaction, as in [26]. This information could then be supplied to the activity-recognition model as additional information.

Beyond the office desk and checkout counter use-cases, we expect that this solution could prove useful for other intelligent systems that rely on activity recognition but where the use of cameras may be prohibited. For meeting

rooms, for example, detecting position and movement of meeting attendees could be used for augmenting meetings. It could also assist remote meeting participants to direct their attention appropriately. Also in work environments, this system could be used to monitor hot-desking utilization – understanding which desks or cubicles are occupied without relying on cameras.

One aspect that needs to be considered by a person wishing to deploy RF-radar-based solutions is the important topic of calibration in a dynamic environment. In the convenience-store domain, for example, the physical environment is for the most part static – a counter, a fixed POS, and no other furniture – making calibration simpler. By contrast, offices and meeting-room environments contain physical elements that can be moved at any time and that will greatly affect any RF sensor data. Consider, for example, initiating an RF sensor under a meeting-room table. During calibration, a sensor will treat any present signals as background. However, if a chair that was present during calibration is moved, or a computer that was on the table, the system may have to be re-calibrated. One possible approach is to train a dedicated classifier for detecting that no person is present, and use that knowledge to initiate re-calibration.

## 6 LIMITATIONS

A limitation of our dataset for convenience-store activities is that, unlike the deskwork dataset, which captured more diverse environments, we collected data in only a single simulated checkout environment. Furthermore, our dataset was limited in the number of "customers" present (not more than 2) and limited in the number of items that were interacted with – scanned, bagged, or exchanged. We expect that deploying the sensors in a new environment would require the collection of new training data. Still, our data includes data from participants of different height and weight and conditions of more than one "customer". In the future, we hope to deploy our system in a real store and to better understand the effect of different work surface materials on system performance.

Our deskwork dataset included activities performed in participants' real offices and accounted for daily variations in objects, clothing, and postures. However, in this pilot dataset of deskwork activities, activities were performed separately from one another. Whereas activities at the convenience store counter are rarely interleaved, at an office desk, multiple activities often occur at the same time (for example, working on the computer and drinking). Our work has also focused on a set of common activities, but more activities could be recognized. As such, a larger set, with more and interleaved activities needs to be collected, and an approach for recognition, such as in [34], can be attempted.

Additionally, our experiments show that each user's desk area included unique environmental characteristics that need to be modeled separately. In other words, collecting training data may be required for each new office environment. This may also be an opportunity to explore domain adaptation approaches to reusing data across users in future work. Nonetheless, we believe our pilot office-desk activity dataset serves to demonstrate the potential applicability of our solution to this work domain. Our position-estimation solution focused on a set of discrete positions that correspond to semantic placement of products on the surface. As discussed earlier, collecting natural user behavior implies that the dataset is imbalanced and not all possible configurations of people around the table were observed. Furthermore, if a fine-grained position resolution is required, a different, continuous estimation approach would be needed.

## 7 CONCLUSIONS

We presented a solution for accurate, yet discreet and unobtrusive activity recognition solution that utilizes an RF-radar sensor mounted under a work surface. We further show how the use of the RF sensor in combination with a wrist-worn IMU can be used for attributing a recognized action to the correct user in multi-user scenarios. We demonstrated our solution in three domains: a convenience-store counter, an office environment, and a retail showroom. Our experiments

reported in this paper show that recognition could be improved by extracting additional projections of RF data. We further show the inverse correlation between recognition accuracy and distance from the sensor. In conclusion, we believe this paper places a useful tool in the toolbox of developers and designers of intelligent context-aware systems.

## 8 ACKNOWLEDGEMENTS

We thank Yulius Tjahjadi and Chelhwon Kim for their help and thank Jenn Marlow and Thi Avrahami for their help with the paper. Finally, we thank our participants for volunteering their time.

## REFERENCES

- [1] Fadel Adib, Chen-Yu Hsu, Hongzi Mao, Dina Katabi, and Frédéric Durand. 2015. Capturing the Human Figure Through a Wall. *ACM Trans. Graph.* 34, 6 (Oct. 2015), 219:1–219:13. <https://doi.org/10.1145/2816795.2818072>
- [2] Michelle Annett, Tovi Grossman, Daniel Wigdor, and George Fitzmaurice. 2011. Medusa: A Proximity-aware Multi-touch Tabletop. In *Proceedings of the 24th Annual ACM Symposium on User Interface Software and Technology (UIST '11)*. ACM, New York, NY, USA, 337–346. <https://doi.org/10.1145/2047196.2047240>
- [3] Daniel Avrahami, Mitesh Patel, Yusuke Yamaura, and Sven Kratz. 2018. Below the Surface: Unobtrusive Activity Recognition for Work Surfaces Using RF-radar Sensing. In *23rd International Conference on Intelligent User Interfaces (IUI '18)*. ACM, New York, NY, USA, 439–451. <https://doi.org/10.1145/3172944.3172962>
- [4] Ling Bao and Stephen S. Intille. 2004. Activity Recognition from User-Annotated Acceleration Data. In *Pervasive Computing*. Springer, Berlin, Heidelberg, 1–17. [https://doi.org/10.1007/978-3-540-24646-6\\_1](https://doi.org/10.1007/978-3-540-24646-6_1)
- [5] Oresti Baños, Miguel Damas, Héctor Pomares, Ignacio Rojas, Máté Attila Tóth, and Oliver Amft. 2012. A Benchmark Dataset to Evaluate Sensor Displacement in Activity Recognition. In *Proceedings of the 2012 ACM Conference on Ubiquitous Computing (UbiComp '12)*. ACM, New York, NY, USA, 1026–1035. <https://doi.org/10.1145/2370216.2370437>
- [6] James "Bo" Begole, Nicholas E. Matsakis, and John C. Tang. 2004. Lilsys: Sensing Unavailability. In *Proceedings of the 2004 ACM Conference on Computer Supported Cooperative Work (CSCW'04)*. ACM, New York, NY, USA, 511–514. <https://doi.org/10.1145/1031607.1031691>
- [7] Michael Buettner, Richa Prasad, Matthai Philipose, and David Wetherall. 2009. Recognizing Daily Activities with RFID-based Sensors. In *Proceedings of the 11th International Conference on Ubiquitous Computing (UbiComp '09)*. ACM, New York, NY, USA, 51–60. <https://doi.org/10.1145/1620545.1620553>
- [8] Andreas Bulling, Ulf Blanke, and Bernt Schiele. 2014. A Tutorial on Human Activity Recognition Using Body-worn Inertial Sensors. *ACM Comput. Surv.* 46, 3 (Jan. 2014), 33:1–33:33. <https://doi.org/10.1145/2499621>
- [9] Scott A. Cambo, Daniel Avrahami, and Matthew L. Lee. 2017. BreakSense: Combining Physiological and Location Sensing to Promote Mobility During Work-Breaks. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems (CHI'17)*. ACM, New York, NY, USA, 3595–3607. <https://doi.org/10.1145/3025453.3026021>
- [10] Ricardo Chavarriaga, Hesam Saghaf, Alberto Calatrone, Sundara Tejaswi Digumarti, Gerhard Tröster, José Del R. Millán, and Daniel Roggen. 2013. The Opportunity Challenge: A Benchmark Database for On-body Sensor-based Activity Recognition. *Pattern Recogn. Lett.* 34, 15 (Nov. 2013), 2033–2042. <https://doi.org/10.1016/j.patrec.2012.12.014>
- [11] Tianqi Chen and Carlos Guestrin. 2016. XGBoost: A Scalable Tree Boosting System. In *Proceedings of the 22Nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD '16)*. ACM, New York, NY, USA, 785–794. <https://doi.org/10.1145/2939672.2939785>
- [12] Yongwon Cho, Yunyoung Nam, Yoo-Joo Choi, and We-Duke Cho. 2008. SmartBuckle: Human Activity Recognition Using a 3-axis Accelerometer and a Wearable Camera. In *Proceedings of the 2nd International Workshop on Systems and Networking Support for Health Care and Assisted Living Environments (HealthNet'08)*. ACM, New York, NY, USA, 7:1–7:3. <https://doi.org/10.1145/1515747.1515757>
- [13] Tanzeem Choudhury, Gaetano Borriello, Sunny Consolvo, Dirk Haehnel, Beverly Harrison, Bruce Hemingway, Jeffrey Hightower, Predrag "Pedja" Klasnja, Karl Koscher, Anthony LaMarca, James A. Landay, Louis LeGrand, Jonathan Lester, Ali Rahimi, Adam Rea, and Danny Wyatt. 2008. The Mobile Sensing Platform: An Embedded Activity Recognition System. *IEEE Pervasive Computing* 7, 2 (April 2008), 32–41. <https://doi.org/10.1109/MPRV.2008.39>
- [14] Dima Damen, Andrew Gee, Walterio Mayol-Cuevas, and Andrew Calway. 2012. Egocentric real-time workspace monitoring using an rgb-d camera. In *2012 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 1029–1036. <https://doi.org/10.1109/IROS.2012.6385829>
- [15] Han Ding, Chen Qian, Jinsong Han, Ge Wang, Zhiping Jiang, Jizhong Zhao, and Wei Xi. 2016. Device-free Detection of Approach and Departure Behaviors Using Backscatter Communication. In *Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing (UbiComp'16)*. ACM, New York, NY, USA, 167–177. <https://doi.org/10.1145/2971648.2971699>
- [16] James Fogarty, Scott E. Hudson, Christopher G. Atkeson, Daniel Avrahami, Jodi Forlizzi, Sara Kiesler, Johnny C. Lee, and Jie Yang. 2005. Predicting Human Interruption with Sensors. *ACM Trans. Comput.-Hum. Interact.* 12, 1 (March 2005), 119–146. <https://doi.org/10.1145/1057237.1057243>
- [17] Eric Horvitz, Paul Koch, and Johnson Apacible. 2004. BusyBody: Creating and Fielding Personalized Models of the Cost of Interruption. In *Proceedings of the 2004 ACM Conference on Computer Supported Cooperative Work (CSCW'04)*. ACM, New York, NY, USA, 507–510. <https://doi.org/10.1145/1031607.1031690>

- [18] M. C. Huang, J. J. Liu, W. Xu, C. Gu, C. Li, and M. Sarrafzadeh. 2016. A Self-Calibrating Radar Sensor System for Measuring Vital Signs. *IEEE Transactions on Biomedical Circuits and Systems* 10, 2 (April 2016), 352–363. <https://doi.org/10.1109/TBCAS.2015.2411732>
- [19] Shian-Ru Ke, Hoang Le Uyen Thuc, Yong-Jin Lee, Jenq-Neng Hwang, Jang-Hee Yoo, and Kyoung-Ho Choi. 2013. A Review on Video-Based Human Activity Recognition. *Computers* 2, 2 (June 2013), 88–131. <https://doi.org/10.3390/computers2020088>
- [20] Gierad Laput, Yang Zhang, and Chris Harrison. 2017. Synthetic Sensors: Towards General-Purpose Sensing. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems (CHI '17)*. ACM, 3986–3999. <https://doi.org/10.1145/3025453.3025773>
- [21] Oscar D. Lara and Miguel A. Labrador. 2013. A survey on human activity recognition using wearable sensors. *IEEE Communications Surveys and Tutorials* 15, 3 (2013), 1192–1209. <https://doi.org/10.1109/SURV.2012.110112.00192>
- [22] Chandrashekhar Lavania, Sunil Thulasidasan, Anthony LaMarca, Jeffrey Scofield, and Jeff Bilmes. 2016. A Weakly Supervised Activity Recognition Framework for Real-time Synthetic Biology Laboratory Assistance. In *Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing (UbiComp '16)*. ACM, New York, NY, USA, 37–48. <https://doi.org/10.1145/2971648.2971716>
- [23] Jaime Lien, Nicholas Gillian, M. Emre Karagozler, Patrick Amihood, Carsten Schwesig, Erik Olson, Hakim Raja, and Ivan Poupyrev. 2016. Soli: Ubiquitous gesture sensing with millimeter wave radar. *ACM Transactions on Graphics (TOG)* 35, 4 (2016), 142. <http://dl.acm.org/citation.cfm?id=2925953>
- [24] Feng Lin, Chen Song, Yan Zhuang, Wenyao Xu, Changzhi Li, and Kui Ren. 2017. Cardiac Scan: A Non-contact and Continuous Heart-based User Authentication System. In *Proceedings of the 23rd Annual International Conference on Mobile Computing and Networking (MobiCom '17)*. ACM, New York, NY, USA, 315–328. <https://doi.org/10.1145/3117811.3117839>
- [25] Liang Liu, Mihail Popescu, Marjorie Skubic, Marilyn Rantz, Tarik Yardibi, and Paul Cudihy. 2011. Automatic fall detection based on Doppler radar motion signature. In *2011 5th International Conference on Pervasive Computing Technologies for Healthcare (PervasiveHealth) and Workshops*. IEEE, 222–225. <https://doi.org/10.4108/icst.pervasivehealth.2011.245993>
- [26] Nicolai Marquardt, Robert Diaz-Marino, Sebastian Boring, and Saul Greenberg. 2011. The Proximity Toolkit: Prototyping Proxemic Interactions in Ubiquitous Computing Ecologies. In *Proceedings of the 24th Annual ACM Symposium on User Interface Software and Technology (UIST '11)*. ACM, New York, NY, USA, 315–326. <https://doi.org/10.1145/2047196.2047238>
- [27] Fabian Pedregosa, Gaël Varoquaux, Alexandre Gramfort, Vincent Michel, Bertrand Thirion, Olivier Grisel, Mathieu Blondel, Peter Prettenhofer, Ron Weiss, Vincent Dubourg, Jake Vanderplas, Alexandre Passos, David Cournapeau, Matthieu Brucher, Matthieu Perrot, and Édouard Duchesnay. 2011. Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research* 12 (Oct. 2011), 28252830. <http://jmlr.csail.mit.edu/papers/v12/pedregosa11a.html>
- [28] M. Philipose, K. P. Fishkin, M. Perkowitz, D. J. Patterson, D. Fox, H. Kautz, and D. Hahnel. 2004. Inferring activities from interactions with objects. *IEEE Pervasive Computing* 3, 4 (Oct. 2004), 50–57. <https://doi.org/10.1109/MPRV.2004.7>
- [29] Ronald Poppe. 2010. A survey on vision-based human action recognition. *Image and vision computing* 28, 6 (2010), 976–990. <http://www.sciencedirect.com/science/article/pii/S0262885609002704>
- [30] X. Qi, G. Zhou, Y. Li, and G. Peng. 2012. RadioSense: Exploiting Wireless Communication Patterns for Body Sensor Network Activity Recognition. In *2012 IEEE 33rd Real-Time Systems Symposium*. 95–104. <https://doi.org/10.1109/RTSS.2012.62>
- [31] Meera Radhakrishnan, Sharanya Eswaran, Archan Misra, Deepthi Chander, and Koustuv Dasgupta. 2016. Iris: Tapping wearable sensing to capture in-store retail insights on shoppers. In *Pervasive Computing and Communications (PerCom), 2016 IEEE International Conference on*. IEEE, 1–8. <https://doi.org/10.1109/PERCOM.2016.7456526>
- [32] Derek Reilly, Emma Westcott, David Parker, Samuel Perreault, Derek Neil, Nathan Lapierre, Kate Hartman, and Harjot Bal. 2013. Design-driven Research for Workplace Exergames: The Limber Case Study. In *Proceedings of the First International Conference on Gameful Design, Research, and Applications (Gamification '13)*. ACM, New York, NY, USA, 123–126. <https://doi.org/10.1145/2583008.2583030>
- [33] Attila Reiss and Didier Stricker. 2012. Creating and Benchmarking a New Dataset for Physical Activity Monitoring. In *Proceedings of the 5th International Conference on PErvasive Technologies Related to Assistive Environments (PETRA '12)*. ACM, New York, NY, USA, 40:1–40:8. <https://doi.org/10.1145/2413097.2413148>
- [34] Daniele Riboni, Timo Szytler, Gabriele Civitarese, and Heiner Stuckenschmidt. 2016. Unsupervised Recognition of Interleaved Activities of Daily Living Through Ontological and Probabilistic Reasoning. In *Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing (UbiComp '16)*. ACM, New York, NY, USA, 1–12. <https://doi.org/10.1145/2971648.2971691>
- [35] Markus Scholz, Till Riedel, Mario Hock, and Michael Beigl. 2013. Device-free and Device-bound Activity Recognition Using Radio Signal Strength. In *Proceedings of the 4th Augmented Human International Conference (AH '13)*. ACM, New York, NY, USA, 100–107. <https://doi.org/10.1145/2459254>
- [36] Longfei Shangguan, Zimu Zhou, and Kyle Jamieson. 2017. Enabling Gesture-based Interactions with Objects. In *Proceedings of the 15th Annual International Conference on Mobile Systems, Applications, and Services (MobiSys '17)*. ACM, New York, NY, USA, 239–251. <https://doi.org/10.1145/3081333.3081364>
- [37] S. Sigg, U. Blaner, and G. Tröster. 2014. The telepathic phone: Frictionless activity recognition from WiFi-RSSI. In *2014 IEEE International Conference on Pervasive Computing and Communications (PerCom)*. 148–155. <https://doi.org/10.1109/PerCom.2014.6813955>
- [38] Chiew Seng Sean Tan, Johannes Schöning, Kris Luyten, and Karin Coninx. 2013. Informing Intelligent User Interfaces by Inferring Affective States from Body Postures in Ubiquitous Computing Environments. In *Proceedings of the 2013 International Conference on Intelligent User Interfaces (IUI '13)*. ACM, New York, NY, USA, 235–246. <https://doi.org/10.1145/2449396.2449427>

- [39] Edison Thomaz, Cheng Zhang, Irfan Essa, and Gregory D. Abowd. 2015. Inferring Meal Eating Activities in Real World Settings from Ambient Sounds: A Feasibility Study. In *Proceedings of the 20th International Conference on Intelligent User Interfaces (IUI '15)*. ACM, New York, NY, USA, 427–431. <https://doi.org/10.1145/2678025.2701405>
- [40] Andy Cristian Tănase, Radu-Daniel Vatavu, Ștefan-Gheorghe Pentiuc, and Adrian Graur. 2008. Detecting and Tracking Multiple Users in the Proximity of Interactive Tabletops. *Advances in Electrical and Computer Engineering* 8, 2 (June 2008), 61–64. <https://doi.org/10.4316/AECE.2008.02011>
- [41] Guochao Wang, Changzhan Gu, Takao Inoue, and Changzhi Li. 2014. A hybrid FMCW-interferometry radar for indoor precise positioning and versatile life activity monitoring. *IEEE Transactions on Microwave Theory and Techniques* 62, 11 (2014), 2812–2822. <https://doi.org/10.1109/TMTT.2014.2358572>
- [42] Ju Wang, Jie Xiong, Xiaojiang Chen, Hongbo Jiang, Rajesh Krishna Balan, and Dingyi Fang. 2017. TagScan: Simultaneous Target Imaging and Material Identification with Commodity RFID Devices. In *Proceedings of the 23rd Annual International Conference on Mobile Computing and Networking (MobiCom '17)*. ACM, New York, NY, USA, 288–300. <https://doi.org/10.1145/3117811.3117830>
- [43] 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 *Proceedings of the 29th Annual Symposium on User Interface Software and Technology (UIST '16)*. ACM, New York, NY, USA, 851–860. <https://doi.org/10.1145/2984511.2984565>
- [44] Jamie A. Ward, Paul Lukowicz, Gerhard Troster, and Thad E. Starner. 2006. Activity recognition of assembly tasks using body-worn microphones and accelerometers. *IEEE transactions on pattern analysis and machine intelligence* 28, 10 (2006), 1553–1567. <https://doi.org/10.1109/TPAMI.2006.197>
- [45] Bo Wei, Wen Hu, Mingrui Yang, and Chun Tung Chou. 2015. Radio-based Device-free Activity Recognition with Radio Frequency Interference. In *Proceedings of the 14th International Conference on Information Processing in Sensor Networks (IPSN '15)*. ACM, New York, NY, USA, 154–165. <https://doi.org/10.1145/2737095.2737117>
- [46] Daniel Weinland, Remi Ronfard, and Edmond Boyer. 2011. A survey of vision-based methods for action representation, segmentation and recognition. *Computer vision and image understanding* 115, 2 (2011), 224–241. <http://www.sciencedirect.com/science/article/pii/S1077314210002171>
- [47] Andrew D. Wilson and Hrvoje Benko. 2014. CrossMotion: Fusing Device and Image Motion for User Identification, Tracking and Device Association. In *Proceedings of the 16th International Conference on Multimodal Interaction (ICMI '14)*. ACM, New York, NY, USA, 216–223. <https://doi.org/10.1145/2663204.2663270>
- [48] Raphael Wimmer, Matthias Kranz, Sebastian Boring, and Albrecht Schmidt. 2007. CapTable and CapShelf: unobtrusive activity recognition using networked capacitive sensors. In *Networked Sensing Systems, 2007. INSS'07. Fourth International Conference on*. IEEE, 85–88. <https://doi.org/10.1109/INSS.2007.4297395>
- [49] Hui-Shyong Yeo, Gergely Flamich, Patrick Schrempf, David Harris-Birtill, and Aaron Quigley. 2016. RadarCat: Radar Categorization for Input & Interaction. In *Proceedings of the 29th Annual Symposium on User Interface Software and Technology (UIST '16)*. ACM, New York, NY, USA, 833–841. <https://doi.org/10.1145/2984511.2984515>
- [50] Yunze Zeng, Parth H. Pathak, and Prasant Mohapatra. 2015. Analyzing Shopper's Behavior Through WiFi Signals. In *Proceedings of the 2Nd Workshop on Workshop on Physical Analytics (WPA '15)*. ACM, New York, NY, USA, 13–18. <https://doi.org/10.1145/2753497.2753508>
- [51] Chenyang Zhang and Yingli Tian. 2012. RGB-D camera-based daily living activity recognition. *Journal of Computer Vision and Image Processing* 2, 4 (2012), 12. <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.669.9558&rep=rep1&type=pdf>
- [52] Shugang Zhang, Zhiqiang Wei, Jie Nie, Lei Huang, Shuang Wang, and Zhen Li. 2017. A Review on Human Activity Recognition Using Vision-Based Method. <https://doi.org/10.1155/2017/3090343>
- [53] Manuela Züger, Christopher Corley, André N. Meyer, Boyang Li, Thomas Fritz, David Shepherd, Vinay Augustine, Patrick Francis, Nicholas Kraft, and Will Snipes. 2017. Reducing Interruptions at Work: A Large-Scale Field Study of FlowLight. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems (CHI '17)*. ACM, New York, NY, USA, 61–72. <https://doi.org/10.1145/3025453.3025662>