Below the Surface: Unobtrusive Activity Recognition for Work Surfaces using RF-radar sensing

Daniel Avrahami  
FXPAL  
Palo Alto, USA  
daniel@fxpal.com

Mitesh Patel  
FXPAL  
Palo Alto, USA  
mitesh@fxpal.com

Yusuke Yamaura  
FUJI XEROX  
Yokohama, Japan  
yamaura.yusuke@fujixerox.co.jp

Sven Kratz*  
Harman International  
Mountain View, USA  
sven.kratz@harman.com

ABSTRACT

Activity recognition is a core component of many intelligent and context-aware systems. In this paper, 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 two work domains; recognizing work activities at a convenience-store counter (useful for post-hoc analytics) and recognizing common office deskwork activities (useful for real-time applications). We classify seven clerk activities with 94.9% accuracy using data collected in a lab environment, and recognize six common deskwork activities collected in real offices with 95.3% 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.

Author Keywords
Activity recognition; Retail; Deskwork; Sensing; Radio Frequency Radar Sensor; IMU.

ACM Classification Keywords
Human-centered computing → Ubiquitous and mobile computing systems and tools

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. c.f. [5, 23]), sensors carried or worn on the user’s body (such as accelerometers or IMUs, c.f. [3, 6, 10, 17, 25]), or instrumented in the environment (such as cameras, c.f. [11, 15, 24, 34, 39]). Indeed, cameras have many advantages for activity recognition, including a wide view of the space in which an activity takes place, not requiring instrumenting 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 of users’ privacy concerns associated with cameras.

Background

The work described in this paper originated from an engagement with a local convenience store chain interested in analytics of activities performed in the day-to-day operation of stores with the goals of process optimization and improved space utilization. Based on discussions, however, the use of cameras for capturing activities in the environment was to be avoided (other instruments, such as beacons and worn sensors were acceptable). To address this constraint, we explored solutions for recognizing activities performed above a work surface (Figure 1). Indeed, the use of cameras can be problematic in many other work environments, such as office spaces and medical facilities that could benefit from accurate activity recognition for both post-hoc and real time applications.

In this paper, we present our exploration of a solution for recognizing common activities performed above a work surface, focusing first on activities performed at a convenience-store’s checkout counter and later on common office deskwork activities. Our proposed solution makes use of an RF-radar sensor, mounted under the work surface.

* - The work was performed while the author was at FXPAL
to perform activity recognition in a discreet and unobtrusive way. We show how smartwatches, worn by the users, can be used to attribute a performed activity to the correct user in multi-user scenarios. 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. 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 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. Demonstrate our solution in two work domains: activities performed at a convenience-store’s checkout counter, and common deskwork activities in an office environment.
3. 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 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.

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 that make use of 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. [24], Weinland et al. [34], and Ke et al. [15] provide surveys of image-based activity recognition systems and applications.

Damen et al. [11] 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. [39] describe a system for the recognition of daily life activities of seniors, using an RGB-D camera. Lavania et al. [18] use computer vision and deep neural networks to perform activity recognition for a biology laboratory. Cameras have also been used for detecting affective state (c.f. [29]). 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 (c.f., [2, 3, 6, 8, 10, 17, 25, 27]) demonstrated recognizing 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 [17]). Ward et al. [33] 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 be used for recognizing other daily activities [30].

Another important approach for human activity recognition is to recognize the activity through the objects that are used. For example, Philipose et al. developed a system for recognizing Activities of Daily Living (ADLs) by attaching RFID tags to household objects [23]. In [5], 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. [21] 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. [16] 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.

**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. [38], for example, used changes in WiFi signals in the store to classify whether a shopper was walking or standing (and where).

Radhakrishnan et al. [25] 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 [25] recognition is limited to activities performed above or just around the work surface.

Wimmer et al. [36] 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 discussed applications for 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 (c.f., [4, 13, 14, 40]). Similarly, prior work looked at a worker’s activities to promote physical breaks in the workplace to combat sedentary behavior (e.g., [7, 26]). 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 [40]. 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.

**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. 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).

Ding et al. [12] demonstrated using changes in backscatter communication between an active RF reader and an RFID tag to classify basic approaching and departure behavior.

For very short-range applications, RF sensing has been used to detect gestures and materials. Lien et al. [19] and Wang et al. [32] used a Google Soli to detect small finger gestures via a deep learning models. The RadarCat [37] 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 with Support Vector Machine (SVM) and K-nearest neighbor (KNN) models were used by Liu et al. [20] to detect fall events of patients in long-term care facilities. Finally, a beam-scanning radar system was proposed by Wang et al. [31] 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.

**AN UNOBTRUSTIVE 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 of the solution to an office 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 the use of multiple 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.

**Implementation**

For capturing RF data above and around a work surface, we use a Walabot Pro RF sensor\(^1\). 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 point cloud is projected into a 75 x 37 pixel image. The intensity of each pixel denotes the reflected energy received at that point, represented in angle and distance (see Figure 2, right). The average transmission power of RF signal is -16 dbm. We note that based on the

---

\(^1\) [https://walabot.com/](https://walabot.com/)
manufacturer’s reports, no known health risks are associated with the device\(^2\).

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 \(^2\). 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 \([22]\) to train and test the different models presented next.

**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 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 our proposed system is generalizable and not constrained to the activities selected in this experiment.

<table>
<thead>
<tr>
<th>10-fold</th>
<th>Baseline (IMU only)</th>
<th>RF</th>
<th>RF + IMU</th>
</tr>
</thead>
<tbody>
<tr>
<td>Logistic Regression</td>
<td>0.384</td>
<td>0.880</td>
<td>0.893</td>
</tr>
<tr>
<td>Naïve Bayes</td>
<td>0.307</td>
<td>0.562</td>
<td>0.564</td>
</tr>
<tr>
<td>KNN</td>
<td>0.585</td>
<td>0.880</td>
<td>0.884</td>
</tr>
<tr>
<td>SVM</td>
<td>0.507</td>
<td><strong>0.905</strong></td>
<td><strong>0.917</strong></td>
</tr>
<tr>
<td>Random Forest</td>
<td><strong>0.593</strong></td>
<td>0.818</td>
<td>0.832</td>
</tr>
</tbody>
</table>

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

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

**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 × 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 180 degrees.

\(^2\) [https://walabot.com/walabot-tech-brief-416.pdf](https://walabot.com/walabot-tech-brief-416.pdf)

\(^3\) [https://www.microsoft.com/microsoft-band/en-us](https://www.microsoft.com/microsoft-band/en-us)
Data
10 participants from our lab 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).

To collect data for different scenarios, seven of the participants performed the activities with one “customer”, and three participants performed the activities 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.

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 1500msec 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.

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.

<table>
<thead>
<tr>
<th>10-fold</th>
<th>RF (1 projection)</th>
<th>RF (3 projections)</th>
<th>RF (3 projections) + IMU</th>
</tr>
</thead>
<tbody>
<tr>
<td>Logistic Regression</td>
<td>0.880</td>
<td>0.938</td>
<td>0.944</td>
</tr>
<tr>
<td>Naïve Bayes</td>
<td>0.562</td>
<td>0.604</td>
<td>0.604</td>
</tr>
<tr>
<td>KNN</td>
<td>0.880</td>
<td>0.915</td>
<td>0.921</td>
</tr>
<tr>
<td>SVM</td>
<td>0.905</td>
<td>0.945</td>
<td>0.949</td>
</tr>
<tr>
<td>Random Forest</td>
<td>0.818</td>
<td>0.881</td>
<td>0.886</td>
</tr>
</tbody>
</table>

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.

Figure 6. The Walabot sensor coordinate system (left), representing a half hemisphere area in which RF signal is transmitted and sensed. On right, the default projection (top), with two additional extracted projections (center and bottom) concatenated.

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

**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 $\Phi$ and distance $R$ (Figure 6, top). It is, however, possible to generate additional projections into 2D space using angle $\Theta$ (see Figure 6, bottom). For this test we generate 2 additional RF signal projections: A projection ($\Theta$-$R$) that is perpendicular to the default slice ($\Phi$-$R$), and a projection ($\Theta$-$\Phi$) that is a cross-section of ($\Theta$-$R$) and ($\Phi$-$R$). 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.

**Results**

Even with recognition rates already high with a single projection, 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.

<table>
<thead>
<tr>
<th>Leave-one-out</th>
<th>RF + IMU</th>
</tr>
</thead>
<tbody>
<tr>
<td>Logistic Regression</td>
<td>0.984</td>
</tr>
<tr>
<td>Naïve Bayes</td>
<td>0.711</td>
</tr>
<tr>
<td>KNN</td>
<td>0.956</td>
</tr>
<tr>
<td>SVM</td>
<td>0.974</td>
</tr>
<tr>
<td>Random Forest</td>
<td>0.918</td>
</tr>
</tbody>
</table>

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

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 very high activity recognition rates for our 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.

Figure 8. The six office deskwork activities to be recognized.
**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. 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 projection 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.

**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.

**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 angles ($\Theta$ and $\phi$) of the Walabot’s sensor were set to 90 degrees (Refer to Figure 6 for details related to field of view).

**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 5 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.

<table>
<thead>
<tr>
<th>10-fold</th>
<th>Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>Logistic Regression</td>
<td>0.831</td>
</tr>
<tr>
<td>Naïve Bayes</td>
<td>0.393</td>
</tr>
<tr>
<td>KNN</td>
<td><strong>0.953</strong></td>
</tr>
<tr>
<td>SVM</td>
<td>0.932</td>
</tr>
<tr>
<td>Random Forest</td>
<td>0.949</td>
</tr>
</tbody>
</table>

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

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.)

**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.

**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 [4, 40].

**ATTRIBUTING ACTION IN MULTI-USER SCENARIOS**

One challenge with using RF sensing for activity recognition is that when more than one user occupies the work space (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 work space.

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. [9] and Wilson and Benko [35] that correlated images from cameras and IMU data.

**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 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.

**Process**

We implemented the following 4-step process for matching RF and IMU data:

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.

\[
1) \quad Activity = \text{argmax} \{ p(\text{activity}|RF) \}
\]

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).

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

In step 3, the system selects 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”.

\[
3) \quad \text{Matching} = \text{argmax}(M_1,M_2)
\]

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.

\[
4) \quad \text{Confidence} = |M_1-M_2|
\]

**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 10 shows the classification accuracy at different classification-confidence bands. It is apparent that classification returned by the system with a high confidence value likely corresponds to correct classification. Specifically, for all samples with a confidence greater than 0.2 (true for 66% of the data), classification accuracy is 88%. This result shows our

![Figure 10. Distribution of prediction accuracy by prediction confidence, calculated as the absolute difference in likelihood score between both IMU samples.](image)
approach performs well on a large segment of the data, and that the confidence metric provided by the system is a good indicator for the reliability of the classification. However, the large number of samples with a low confidence score suggests that further investigation is needed.

LIMITATIONS
A limitation of our dataset for convenience-store activities is that data were collected in a simulated checkout environment. Still, our data includes conditions of more than one customer, and included participants of different height and weight. In the future, we hope to deploy our system in a real store. 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 [28], can be attempted. Nonetheless, we believe our pilot office-desk activity dataset serves to demonstrate the potential applicability of our solution to this work domain.

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. 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.

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.

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 [37]. Furthermore, the physical relationship between people and objects could be used to enhance both recognition and interaction, as in [21]. This information could then be supplied to the activity-recognition model as additional information.

Beyond the office desk and check-out 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. In retail environments, our solution could be used to monitor customer position and interaction with products in a showroom. For such use-cases, the RF sensor could be used to detect the position of movement around a product display and would allow businesses to observe correlations between physical placement of products and sales.

Finally, as described in the Introduction, we chose to investigate the use of RF radar sensing as a potential replacement for the use of cameras for activity recognition. This was particularly important for environments where installing a camera may be prohibitive. Yet, it is an open question whether users would be more comfortable with the RF sensor mounted under the work surface compared to being observed by a camera. We speculate that it would be important to show users the output of the RF sensor as a way of highlighting the lack of humanly interpretable information in that data.

CONCLUSIONS
We present 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 both specialized and general domains: a convenience-store counter and an office environment, and believe that this work has potential application in many other domains. We believe this paper places a useful tool in the toolbox of developers and designers of intelligent context-aware systems.

ACKNOWLEDGMENTS
We thank Matt Cooper for his comments and suggestions. We thank Jenn Marlow and Thi Avrahami for their help with the paper. Finally, we thank our participants for volunteering their time.
REFERENCES


19. Jaime Lien, Nicholas Gillian, M. Emre Karagozler, Patrick Amihood, Carsten Schwegel, Erik Olson,


