Seamless Customer Identification

Christopher Smowton[†], Jacob R. Lorch, David Molnar, Stefan Saroiu, Alec Wolman
[†]University of Cambridge, Microsoft Research

ABSTRACT

This paper proposes *seamless customer identification* (SCI), a means to identify physically present customers without any effort on customers' part beyond a one-time opt-in. With SCI, customers need not present cards or operate smartphones to convey their identities. So, stores can provide personalized shopping experiences at any time, not just at check-out. SCI uses two complementary technologies: device detection and face recognition. Device detection identifies customers by detecting their phones through low-power wireless discovery, in our case using Bluetooth Low Energy (BLE). Face recognition recognizes customers by matching pictures captured during registration with images captured by a store camera. Device detection makes face recognition feasible by limiting the number of potential customers, while face recognition provides directional information that device detection lacks. Together, these technologies provide an ordering of likely candidates to a store employee, who makes the final determination of identity. We have designed and built SCI, and demonstrated its usefulness in an application called Zero-Effort Payments (ZEP). ZEP uses SCI to let customers effortlessly make small purchases at a coffee stand. We conducted two real-world deployments of ZEP on actual customers: a two-day deployment during a technology fair and a four-month deployment in our building. Across both deployments, 274 customers made 705 purchases using ZEP. Through these deployments and other experiments, we demonstrate how our techniques make seamless customer identification feasible and practical.

1. Introduction

This paper introduces *seamless customer identification* (SCI), which lets a store identify its visiting customers without their expending effort beyond a one-time opt-in. SCI enables many different scenarios that make the shopping experience more personal and thus better for both the store and the customer. For example, a store can: (1) provide its customers on-the-fly coupons based on shopping history; (2) dispatch the "right" sales associate trained in the type of merchandise most interesting to the customer; (3) automatically bill a VIP customer's credit card if it is on file; or (4) automatically process refunds without the need for the customer to bring a receipt to validate a purchase made previously.

Customer identification is clearly useful since rudimentary approaches are used today. However, their lack of seamlessness introduces shortcomings. For example, some coffee shops offer customers a free coffee for every ten purchases, some stores give discounts to customers who swipe loyalty cards, and some business mail personalized coupons to people's homes. However, each of these relies on customers remembering to carry something to the store and to present it at checkout. More importantly, none of them allows the customer to be identified before checkout for a more personalized shopping experience. For these reasons, *seamlessness* can provide significant additional benefit to a customer identification system. Indeed, we believe SCI systems are one instance of next-generation mobile applications that take us closer to the vision of ubiquitous computing [30].

An SCI system must be more than just seamless, however: it must be fast, must not rely on expensive equipment, and must accurately identify customers and their locations. Speed lets stores provide quick service without making customers wait. Not relying on expensive hardware enables SCI to be quickly adopted at large scale. High identification accuracy is necessary to avoid frustrating stores and customers. Finally, location accuracy is useful for numerous scenarios that depend on precise customer placement. For example, a grocery store can identify customers' interests better if it knows what aisles they visit and in front of what type of merchandise they stop. Similarly, a seamless payment scheme must distinguish between two customers standing side by side, in order to charge the correct buyer.

This paper presents the design, implementation, deployment, and operation of our SCI system. To have high accuracy, our SCI system combines two indoor localization technologies: (1) wireless localization based on Bluetooth Low Energy (BLE) radios present in today's smartphones; and (2) face recognition using commodity cameras like the ones found in webcams or the Kinect. It combines these to provide a limited number of candidates to a human employee, who makes the final identification decision. This design allows our system to meet its goals: it is fast, inexepensive, and has high accuracy. As an illustration, employees never identified the wrong customer during our deployments.

We chose device identification via BLE and face recognition because they are well-suited to seamless identification and they mitigate each other's shortcomings. These shortcomings, discussed in more detail in §6, are that (1) BLE cannot distinguish between nearby people without a great deal of costly infrastructure, and (2) face recognition has low accuracy when faced with many potential identities. Fortunately, BLE narrows the set of potential identities to just those standing within wireless range, mitigating problem 2. And, face recognition provides exact placement information about its subjects, mitigating problem 1. Thus, the combination of these two technologies lets our solution meet its speed, cost, and accuracy goals.

Although many cameras and smartphones today implement face detection in hardware, their algorithms are relatively lightweight, detecting only a small number of large faces when looking straight at the camera. In contrast, SCI scenarios must detect and recognize faces present anywhere in a large scene, not looking straight at a camera, in real-time at rates of 10 frames per second (fps) or higher. An important factor enabling us to meet such a speed goal is our exploitation of the embarrassingly parallel nature of face recognition. Specifically, we divide a store's video feed into frames and dispatch each frame to a separate face-recognition "worker." We also cut frames into multiple sub-frames to further speed up face recognition. However, such frame cutting can be dangerous because it may divide a face into two parts, neither of which is identifiable by face recognition. So, in §3, we discuss a simple frame cutting technique that guarantees that every person's face will appear in at least one sub-frame.

To evaluate the SCI system, we deployed it for a scenario we call Zero-Effort Payments (ZEP). In ZEP, SCI lets a store charge customers with "zero effort" on their part. The first deployment of SCI was during a two-day technology fair with thousands of attendees. ZEP processed 102 payments for people buying espresso beverages from a coffee cart. The second deployment, in contrast, lasted over four months with participants using the system once or twice a day for several weeks. By replaying the traces collected in these deployments, we can evaluate multiple SCI configurations, not just the ones used in the deployments; we present results in §6.

The rest of the paper is structured as follows. §2 provides background on biometrics and device identification schemes useful in understanding why we chose the ones we did. §3 presents the design of our SCI system, and §4 describes its implementation and how we used it to enable ZEP. §5 evaluates the performance of our ZEP deployments. §6 describes a series of experiments that justify our design choices. Finally, we present related work (§7) and conclusions (§8).

2. Background on Identification Schemes

This section describes the background on schemes for identifying people through biometrics (§2.1) and through device identification (§2.2). This section's goal is to present the reasons that led us to choose face recognition as a biometric scheme, and BLE for device identification, in SCI.

2.1 Biometric Schemes

A *biometric* is a quality of a person that is unique to that person and can therefore be used to distinguish one person from another. We call a *biometric scheme* a set of processes that measure a biometric to determine the identity of a person in a store. A biometric scheme must meet three requirements to fit the needs of SCI.

It must be accurate. The biometric scheme must have low false positives and low false negatives. False positives lead to mis-identification, whereas false negative lead to people not being identified by the system. Both could be problematic depending on the SCI scenario.

It must be non-invasive. The biometric scheme should require little additional effort on the part of the customer. This is essential for meeting the seamless requirement of SCI.

It must resist attacks. It should be difficult for an adversary to impersonate a particular customer. In all our SCI scenarios, biometric identification is done in the presence of a sales associate or cashier. Having a "human in the loop" makes biometric schemes much harder to attack. For example, while holding a photo in front of a camera can easily fool a face-recognition algorithm, a cashier would certainly notice an attacker holding a photo in front of his or her face while making a payment.

2.1.1 Fingerprints

Fingerprints have been used as a biometric since becoming popular in police work during the 19th century. Today, low-cost fingerprint readers are integrated into a variety of laptops, phones, and other devices. The technology for searching large databases of fingerprints is mature, and fingerprint-based identification has high accuracy.

Despite these benefits, we determined that fingerprints do not meet the seamlessness requirement because they require customers to touch a fingerprint sensor. Cleanliness of the sensor is an additional issue because some people questioned to us the hygiene of such a solution. Also, their association with police work or the "mark of the beast" in Revelations makes customers in some countries reluctant to provide fingerprints on demand; the supermarket chain Piggly Wiggly discovered in 2006 that several of its customers cited these reasons for not enrolling in a fingerprint payment system [24]. There are also people, such as cooks or people who have survived a fire, that lack easily readable fingerprints.

Finally, basic fingerprint readers are not attack-resistant. It is possible to build "fake fingers" undetectable to a casual inspection by a store employee. Although sophisticated mechanisms to combat such fake readings exist, these methods increase the cost and make reading less seamless [17, 27].

2.1.2 Voice-based Identification

The state of the art in voice-based identification requires long voice samples to provide high accuracy rates. As the state of the art advances, it may become viable to do zero-effort identification of users by listening to short statements they make as a natural part of conducting a transaction.

The U.S. National Institute of Standards and Technology has run an evaluation of text-independent speaker recognition since 1996 [26]. In 2010, the trials included 50 research systems. One of the test conditions was to recognize a speaker from 10 seconds of conversation; this is the condition closest to our customer identification scenario, as we do not expect a customer to engage in extended conversation with a cashier in all cases. The systems then needed to match the speaker to one of 430 candidates, where the systems were trained with a mix of audio from phone calls and in-person conversations.

The results showed roughly 80–85% accuracy in the best case. The trials also called out a "greybeard effect" where peoples' voices change over time, and the voices of older people are different than those of younger people. The requirement for speech also means that speaker identification cannot be used to "pre-position" coupons at the selector because the identification cannot happen until the customer starts talking.

2.1.3 Iris Recognition

The human iris contains distinctive patterns that seem unique to each individual, even between identical twins. Almost two decades ago, researchers proposed a way to compute a short iris representation called an *iris code* [9]. The key requirement is that the eye be illuminated with a suitable source of infrared light, then viewed by an infrared-sensitive camera. Iris codes have been computed across populations of tens of thousands of people from different demographics with low false positive rates, and the U.S. National Institute of Standards and Technology conducts a periodic competition between different iris code implementations to measure accuracy [18].

Today, multiple companies sell iris scanners that have high accuracy at short range [23, 4]. To use one, a user must look into an eyepiece that combines illumination and a camera, a procedure far from effortless. Furthermore, recent work has shown that one system can be fooled by eye images synthesized from iris codes [10], suggesting that even iris recognition must be done with a "human in the loop" to avoid simple impersonations.

Longer-range systems have started to emerge, such as the recently announced "Iris on the Move" product aimed at airport security terminals [8]. This product consists of two large pillars, similar to a metal detector. When someone walks through the gap between the pillars, they shine infrared light on his eyes, capture the image, and compute the iris code. This removes the need for user effort, but requires placing pillars wherever people must be identified. Also, there is little publicly available information on the accuracy rates of long-range iris recognition systems.

2.1.4 Gait Identification

The advent of inexpensive depth cameras, such as the Microsoft Kinect, along with machine learning algorithms for processing depth data, enables a new avenue for biometrics. Structured light sensors work in most lighting conditions and operate at medium- to long-range, an improvement over other depth detectors like stereo RGB cameras and time-of-flight cameras [6].

One possible depth-enabled biometric is *gait identification*, which recognizes a person through idiosyncrasies in walking. This biometric is effortless since a customer would only need to walk into a section of the store covered by the depth sensor. Recent work has shown the Kinect and existing machine learning algorithms can reconstruct skeletal data using depth sensors with a 91.0% accuracy rate [21]. Although the security of this biometric is not well understood, it appears difficult to intentionally mimic.

However, we determined that gait identification is not ready for use in SCI because the technology has yet to be proven in real-life scenarios. Furthermore, the study described above only used seven subjects who all performed the same walk in the same room. Even in such an unrealistic environment, one test subject could not be recognized at all.

2.1.5 Face Recognition

Despite much research work spanning decades, face recognition has started to be incorporated in practical systems only recently. Much journalistic evidence exists that many police and law enforcement agencies use face recognition in an attempt to find criminals in large crowds or in areas with high traffic. One of the earliest known deployments was during the US SuperBowl in 2001 in which local and state police agencies scanned the faces in the stadium for known terrorists without making the public aware of the presence of such technology [32]. Although no suspects were apprehended, police forces argued that such systems help deter crime even when they do not lead to arrests.

Despite the evidence of these practical deployments, most face recognition research work is done at the algorithmical level focusing on improving the accuracy rates of the underlying algorithms, and testing them against published benchmarks [20]. Our related work section (§7) will describe this work in more detail, but for a survey of recent results, see [34]. The accuracy rates reported by this work vary widely (e.g., 50% accuracy rate in [13], and 92% accuracy rate in [35]) depending on the algorithms used, the quality of the training data, and the conditions under which testing is done, such as the degree of illumination, the variation in the subjects' posing or expressions.

Upon surveying the work on face recognition algorithms and their accuracy, two observations emerged. First, face recognition accuracy degrades rapidly as the *gallery* size increases. The gallery refers to the size of the database of identities matched against. Second, face recognition today cannot produce perfect results even under ideal conditions. While the accuracy rate can improve drastically in well-controlled experiments, it can never be guaranteed to be perfect. These two observations led us to conclude that face recognition deployments in practice can succeed only when the gallery size is not large and imperfect answers can be tolerated.

Fortunately, our SCI scenarios can overcome the shortcomings of face recognition. First, the additional use of device identification can ensure that the gallery is never too large. Second, the additional "human-in-the-loop" can correct for face recognition errors. For example, face recognition could provide a set of 4 choices for customer identification to a store associate with high accuracy rates. These observations together with the availability of mature, off-the-shelf software made us choose face recognition as the biometric for SCI. Our SCI implementation uses a recent version of Microsoft Face SDK[1].

2.2 Device Identification Schemes

Another way to identify a person is by identifying a device he or she is carrying. Three potentially useful technologies for this are Passive Radio-Frequency Identification (RFID), Bluetooth, and Bluetooth Low Energy (BLE).

Passive RFID.

RFID involves communication between a *reader* and a *tag*. The reader sends out a request for identification, and the tag responds with its identity. In passive RFID, the tag has no internal power source; it uses energy accompanying the reader's request to transmit its identity [14].

Passive RFID typically has low range, up to 3 meters [12]. Such short range would necessitate a customer to bring the device close to a reader and thus limit the seamlessness of the interaction. However, some RFID tags are claimed to have even larger range, up to 6 meters [12], and range will likely increase as technology advances. However, even if range issues are removed, passive RFID still necessitates distributing a tag to each customer.

Bluetooth.

Bluetooth, on the other hand, is a wireless protocol almost universally present in mobile devices users already carry. The most relevant element of the Bluetooth protocol is *discovery*, whereby one device can discover the presence and identify of another device [5].

Unfortunately, Bluetooth discovery is inappropriate for SCI for three reasons. First, when in discoverable mode, Bluetooth devices are typically configured to not do anything but be discovered [11]. Thus, users will not want to keep their devices in discoverable mode at all times, and will have to explicitly make them discoverable, making the process of discovery non-seamless. Second, being in a discoverable mode consumes significant power, generally about 100 mW [11]. Third, discovery times can be significant, on the order of four seconds [19].

[1] `http://research.microsoft.com/en-us/projects/facesdk/`.

BLE.

Fortunately, recent versions of the Bluetooth protocol, starting with v4.0, include a subprotocol called Bluetooth Low Energy (BLE) [5]. This protocol supports always-on discovery by allowing a device to periodically perform a low-power broadcast, called an *advertisement*, of its identity. This enables low-latency discovery at low power; in our experiments, broadcasting once a second consumes only 0.22 mW. BLE is thus an ideal technology for SCI, and the one we have selected for our prototype.

3. Design

Our SCI design has three main goals: accuracy, speed, and scalability. High accuracy ensures we do not create frustration for employees and customers due to mis-identification. Low latency is important to ensure we provide data in time for it to be useful. Finally, for the system to scale to large populations, the system must reliably identify customers even when hundreds of thousands of potential customers are registered in the system. This section describes the design chosen to satisfy these goals.

An overview of this design is presented in Figure 1. The *detector* is a computer with Bluetooth Low Energy (BLE) capability and a camera. It uses BLE to detect the presence of customers' devices (§3.1) and the camera to view customers' faces (§3.2). The detector determines which customers are present and sends this information to a *selector*, typically a store employee's tablet computer. The selector presents the customers' names and head shots to the store (§3.3), so when an employee needs to know a certain customer's identity, he or she can readily deduce it by comparing the customer's appearance with the presented head shots. The system obtains these head shots, along with the customer device information, when the customer registers with the SCI system (§3.4).

Face identification is CPU-intensive and can incur high latency. To combat this, the detector offloads tasks to *workers*, i.e., server-class machines on the premises or in the cloud (§3.2.1). Face identification can also raise privacy concerns, so we need to inform customers and give them control over their private data (§3.2.2). To provide more transparency in what data our system is gathering, we installed a CCTV-like monitor that shows customers what the camera is capturing.

3.1 Device Identification

Each customer registered with our system must carry a device, such as a modern smartphone, equipped with a BLE radio. The customer configures his or her device to be always discoverable, perhaps via an application. Because BLE is designed for low-power discovery, such configuration does not significantly impact the device's battery life.

BLE includes various protocols for use in discovery, formalized as a set of *roles* a participant can fill [5]. The ones we use are the *broadcaster* role, which periodically broadcasts an advertisement, and the *observer* role, which watches

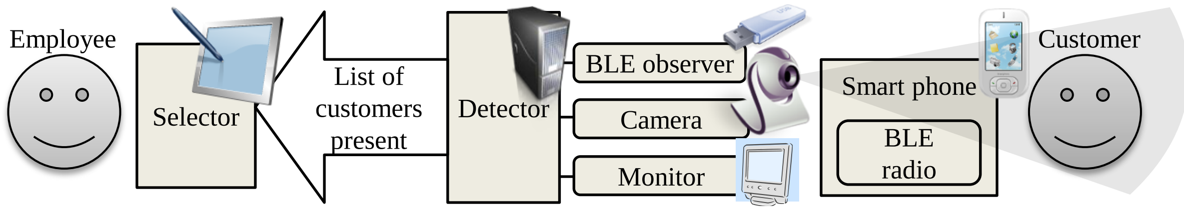

Figure 1: Overview of SCI design.

for advertisements. The customer's device acts as the broadcaster and the detector's device acts as the observer. We selected one second as the broadcaster's period to ensure we do not have to wait long for a customer to be detectable, and yet not consume much power on the customer's device.

The observer role in BLE, which our detector uses, also has an associated period: the scan period. During each scan period, BLE's observer role ignores multiple advertisements from the same broadcaster. Therefore, to receive as many advertisements as possible, we should not set the observer period to be more than the broadcaster's period. Additionally, the receiver may miss incoming broadcasts when the receiver's radio transitions from one scan period to the next. Thus, setting the observer period exactly equal to the broadcaster's period could cause these blind periods to precisely line up with a particular broadcaster's transmissions, making the broadcaster invisible to the detector. We thus set the observer's scan period to 0.9 sec, slightly less than the broadcaster's period.

The detector decides a customer's device is present if an advertisement from that device was observed in the last five seconds. The use of a period much longer than the broadcast period allows for limited amounts of packet loss.

3.2 Face Identification

To determine which customers are in a certain location of interest, we point the detector's camera at that location. By comparing the faces appearing in the camera's video with those of customers whose devices are nearby, the detector decides which customers are not just close to, but actually in, the area of interest.

To identify customers' faces, the detector uses Microsoft FaceSDK, an off-the-shelf library for detecting and identifying faces in digital pictures. For each video frame the camera captures, it passes the frame to FaceSDK. It also provides FaceSDK with a corpus of potential customers to choose from, namely those whose devices have recently been observed. It describes this set of customers to FaceSDK as a set of *profiles*, each of which contains a set of face images of a certain customer. These are images we collect from the customer during registration, as we will discuss in §3.4.

FaceSDK provides a ranking of each customer profile for each face detected in a frame. These rankings can be combined in many ways to produce face recognition scores for a final identification ranking during payment. During our deployments, we calculated a *score* for each customer using a weighted average. However, our evaluation section §6 will explore the effectiveness of different rankings.

As we will discuss in §6, face identification cannot guarantee perfect accuracy, even when identifying a person out of a small set of choices. Even if only one of the nearby customers is in view of the camera, that customer is not necessarily the top-scoring one. Therefore, the ultimate decision to identify a customer is never made by the detector alone. As we will discuss in §3.3, the detector's responsibility is to present the scoring information to a selector held by a human employee responsible for the final identification.

3.2.1 Reducing Face Identification Latency

Face identification can be CPU-intensive, so unless the detector is highly provisioned it may become overloaded, causing face identification tasks to queue and experience high latency. One strategy to avoid this is to have the detector drop frames instead of queueing them; after all, there is a lot of redundancy in consecutive frames. However, a preferred strategy is to offload work to *workers*, if they are available on the premises or in the cloud.

Workers can offer large amounts of computing power for the process of face identification. To leverage more than one worker, we can readily parallelize the work since each video frame's processing is independent. We can thus leverage multiple workers by sending different frames to different workers.

Although such frame-by-frame parallelization reduces the latency of face identification, it cannot reduce it below the time to process a single frame. Unfortunately, though, this time can be significant. An average server takes about 2–3 seconds on a 1280x768 frame that contains a few faces. When many faces are present, as was often the case during our technology-fair deployment, a frame can take as long as seven seconds. Indeed, even a frame with no faces can take up to one second.

To solve this problem, we designed another technique to further reduce the latency of face recognition. We divide each frame into sub-frames, and send each sub-frame to a separate worker. To ensure we do not leave out any faces by cutting them into unrecognizable halves, we cut the frames in a way that *guarantees* that each face will appear in full in at least one sub-frame. To do this, we determine the maximum length any human's face could possibly occupy in any

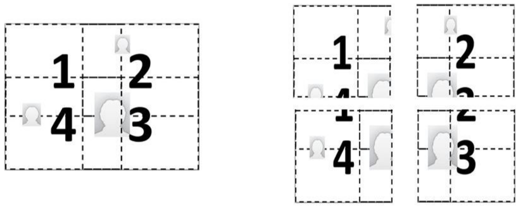

Figure 2: Dividing frames into smaller sub-frames to speed up face recognition. The sub-frames must overlap to ensure that one face appears in at least one sub-frame. For example, the largest face above, although spanning all sub-frames, appears in sub-frame #3 in full and thus can be processed by face recognition.

dimension from the camera's perspective. We then choose sub-frames so that they overlap in at least this amount. Figure 2 shows an illustrative four-way cut in which any two adjacent sub-frames overlap, but note that our technique can use subframe counts other than four. As the overlapping region's size is equal to the maximum facial length, each face appears in full in at least one sub-frame. We experimented with this technique during our system evaluation but we omit presenting the results.

3.2.2 *Protecting Privacy*

SCI raises serious privacy concerns. In particular, some people become uncomfortable knowing their faces are subject to an automatic face recognition process. In our experience, we ran into many people who thought technology has reached a point today where it is able to identify and track them even when not registered with our system. The discrepancy between what some people think face recognition technology can do and what it in fact does is quite large and appears to stem from technologically-implausible scenarios prevalent in some of today's popular movies and TV shows. To alleviate these concerns, we have taken several steps that all aim at making our SCI process as transparent as possible.

First, in all our SCI deployments, we made heavy use of signage indicating a face recognition system is deployed in the area. Also, the area covered by our camera was clearly delimited in the carpet by a dark region. We wanted to guarantee no identification was possible for anyone standing outside of the dark-region carpet. For this, we oriented the camera downward at an angle that ensured we would not capture even the feet of a person standing outside the area. One side effect of such placement was that the camera was not shooting horizontally straight at the customer's face, but rather at an approximately 30° angle facing down. We suspect this orientation hurts the face recognition accuracy although we have not determined the exact accuracy loss.

The dark-region area was sprinkled with many signs describing our process and also listing three ways in which customers could ensure that our data would not capture their faces: by not stepping in the well-delimited area, by asking the operators to turn off the camera, or by sending us e-mail requesting that we manually remove all frames capturing their faces accidentally. Over the course of our deployments, there were several times when the camera was turned off due to a customer request, and we also received e-mail requesting manual removal twice. Another lesson learned from these deployments is that there is a need for a universally-known signage for face recognition, the same way there is well-understood signage for police investigation areas, or CCTV cameras.

Additionally, we mounted a CCTV-like monitor to show the camera feed to customers entering the dark-region area. Our intuition is that people feel more at ease if they can directly view what the camera is capturing. This monitor also showed when the camera was off, serving as a visual cue for people who requested the camera be turned off.

3.3 Customer Identification

As discussed earlier, face recognition is not perfect, so the detector cannot identify customers alone. Instead, it provides guidance to a human employee, and that employee makes the ultimate customer identification. It provides this guidance via the selector, which operates as follows. Every second, it requests a list of potentially present customers from the detector. The detector sends the selector the list of customers whose devices are present, sorted in decreasing order of score. The selector then presents these to the employee to aid in his or her identifications.

To present the customer data, the selector displays the head shots and names of the customers in order. In our current version, only four head shots are visible at a time in the selector. Since the employee may want to consider customers with even lower scores than these four, the selector provides a way to scroll to lower-scored customers.

An earlier iteration of our selector's user interface posed a problem for employees. Whenever the set of highest-scoring customers changed, our UI immediately refreshed the display. Thus, frequently, an employee attempting to click on a particular customer's face to get more information would be frustrated by it changing during that attempt. In some cases, the employee would not even realize the screen had refreshed, and would become confused by the screen showing a different face than the one selected. To fix this, we consulted a UI expert, who recommended the sliding-tile motif we now use: The selector's display is a tile of static information; when it needs to be changed, the tile visibly slides off the screen as another tile with new content slides in to replace it. While obvious in retrospect, the following UI principle guided the later design of all user-facing components: UI refreshes must be made in a gradual manner and not instantaneously.

3.4 Registration

A customer must register with our SCI system before it can identify that customer. During registration, we record the customer's BLE MAC address and a short video of the

Figure 3: Texas Instruments CC2540 Mini Development Kit includes, from left to right, a debugger for programming BLE devices, a key fob BLE device, and a USB dongle BLE device.

customer's face. We produce a FaceSDK profile from all the frames of this video, and we tell the customer to select a single one of the frames as if he were selecting one for a picture ID. We use this single selected frame as his head shot, i.e., the picture we present to an employee trying to find a match for a physically-present customer's face.

4. Implementation

This section starts by presenting our concrete implementation of SCI, and then describes how we integrated it into a Zero-Effort Payments (ZEP) system allowing customers to make purchases at a coffee stand. According to SLOCCount [31], our implementation consists of 23,626 lines of C# code excluding the face detection libraries.

4.1 BLE Devices

Many mobile devices support BLE, such as the iPhone 4S, the iPhone 5, and many Android smartphones. However, as BLE is fairly new, it is not yet well exposed to developers on these platforms. For instance, the iPhone does not currently allow applications to use the broadcaster role, or to maintain BLE discoverability while the phone is asleep. This, combined with the fact that phones with BLE were not generally possessed by our customers at the time our deployment, made use an alternative BLE device.

We used Texas Instruments CC2540 BLE Mini Development Kits. Each kit includes a USB dongle and battery-powered key fob, as depicted in Figure 3. We used the USB dongle as the detector's BLE device, and we provided each customer with a key fob that simulates a future smartphone with better BLE support. We programmed each BLE fob to use the broadcaster role with a one-second period, and the dongle to use the observer role with a 1.1-second period.

Each BLE fob uses a CR2032 battery, which provides 200 mAh at 3 V. By connecting a power meter to the fob, we determined that it consumes an average of 0.22 mW. This suggests the battery should last slightly under four months, which is consistent with our experience. Furthermore, since consuming 0.22 mW for 24 hours would use only about 0.1% of an iPhone 4S's battery capacity, we expect customers will not mind running BLE continuously.

4.2 Store Devices

We ran the detector on an HP Z210 station equipped with 16GB of RAM and an 8-core Intel Xeon E21245 CPU running at 3.3GHz. We used a Microsoft LifeCam Studio camera which currently retails for US$100, and a CC2540 USB dongle as the BLE scanner. We ran the workers on 2–5 servers, with 8 workers per server, one per core. The selector ran on an HP 2740P EliteBook tablet.

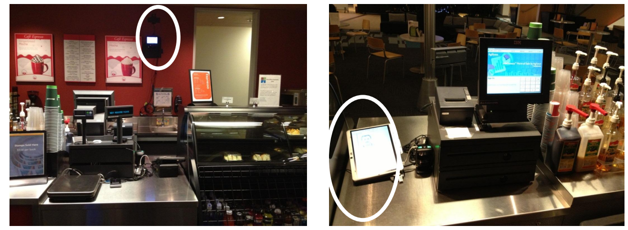

Figure 4: ZEP deployment in our building's cafeteria. On the left, the camera and CCTV-like monitor are placed on the backwall. On the right, the tablet-based selector is placed next to the POS.

	Deployment #1	Deployment #2
Duration	2 days (March 7th—8th, 2012)	20 weeks (May 14th – Sept. 28th, 2012)
# of registered users	255	19
# of payments made	102	603
# of frames gathered	256,831	33,453,300

Table 1: High-level statistics of our two deployments.

4.3 Zero-Effort Payments

ZEP relies on the SCI system to help a cashier identify a customer wishing to make a payment. When a customer asks to use ZEP, the cashier consults the selector, clicks the customer's face, then pushes a button to confirm the payment. ZEP then passes the customer's payment identifier to the cash register to complete the payment. This payment identifier, in our case a meal card number used by our corporate cafeteria system, is obtained from the customer during registration. Figure 4 shows the deployment of ZEP in our building's cafeteria.

The presence of the camera in our system allowed ZEP to offer an additional feature, video receipts. Each time a transaction occurs, ZEP sends a receipt by e-mail to the customer

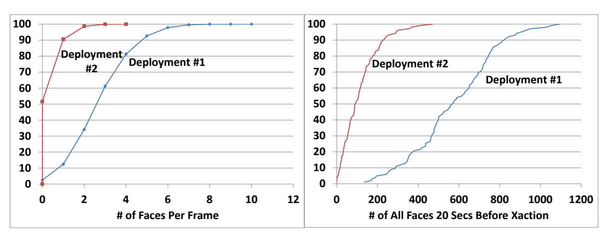

Figure 5: Distribution of number of faces in the two deployments. On the left, the graph displays the number of faces per frame. On the right, the number of all faces identified in all frames 20 seconds prior to a transaction.

identified in the transaction, with a link to a video of the ten seconds surrounding the payment. The original goal of these receipts was to help with fraud: If an attacker were to impersonate a victim, the victim would receive the video of the transaction and could use it to dispute the charge. Although there were no disputes during any of our deployments, we found customers appreciated these video receipts. The few times the system stopped sending receipts because of bugs, we immediately received e-mails from customers pointing out the lack of receipts.

4.4 Implementation Issues

The detector is a single point of failure. For this reason, we decided to pursue a design that minimizes the recovery time for the detector and uses a crash-restart model: On any error, the detector crashes and quickly restarts afresh. Note that a detector failure does not immediately affect the UI tablet (i.e., the selector) which continues to function and show the identification matches. Each second, the tablet contacts the detector; if down, the tablet eventually times out (5 seconds) and shows no more identification matches. Until the time out fires, the tablet remains functional and can be used to conduct transactions. This design makes the detector's restarts transparent to employees and customers.

Unfortunately, the behavior of our hardware made us abandon a crash-restart model for the detector. First, our camera took an average of two seconds to initialize, a behavior consistent with inexpensive Web cameras. We believe a lower bound of two seconds of downtime on recovery was unacceptable for the detector's availability needs. Also, the camera driver would sometimes return an error to an initialization request if the camera was recently running. Thus, sometimes it would take 4–5 seconds of repeated tries for the camera to initialize successfully.

Second, we originally relied on BLE packet capture software written by TI, the manufacturers of our BLE hardware. This software would capture any incoming BLE packets and relay them over UDP to the detector. As this software was designed for short-term debugging rather than long-term operation, it often froze without reporting an error. So, about two months into our second deployment, we re-wrote the firmware of our BLE sniffer device and eliminated the need to run the TI software.

Both these hardware issues made us reconsider a crash-restart model for the detector. Instead, we decided to try to make the detector as robust as possible by offloading as much functionality as possible from it. The detector ended up being quite lean; its roles were to capture the frames, write one copy to the disk and feed one copy to the face recognition workers, and offer a live feed of the face recognition scores to the UI tablet. Despite relatively little functionality, making the detector robust turned out to be challenging: At best, our detector could run for a week without crashing. To overcome this issue, we restarted the detector manually at the end of each working day.

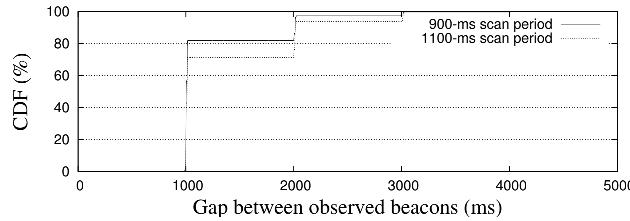

Figure 6: CDF of gap between observed beacons from a BLE broadcaster at location 1 in Figure 8.

5. Evaluation of Our Deployments

We deployed ZEP in two environments. The first was Microsoft TechFest, a two-day technology fair with thousands of attendees, and the second was a long-term installation at a coffee stand in our corporate cafeteria. Table 1 summarizes high-level statistics of the data gathered in each of the two deployments. During TechFest, we only gathered data during the second deployment day.

The two deployments were quite different. During TechFest, many people visited our booth and coffee cart. Thus, most gathered frames contain several faces. In contrast, the frames gathered during the long-term deployment in our corporate cafeteria have much fewer frames on average. On the left, Figure 5 illustrates the distribution of the number of faces per frame during each deployment. On the right, Figure 5 displays the distribution of all faces found within 20 seconds before a transaction occurred. For deployment #2, many frames only had one face; since our camera's shutter speed was 10 fps, only a few transactions (15%) had more than 200 face images in the 20 preceding seconds. In contrast, during deployment #1, more than 200 face images were discovered in those 20 seconds 95% of the time.

During both deployments, we never learned about any mis-identification for any transaction. No customer ever reported not being charged properly, or being charged on behalf of someone else. In both our deployments, the ZEP selector showed up to four identities as potential matches on its screen. A simple interface allowed the cashier to scroll down for the next four matches. During TechFest, ZEP displayed the correct identity of the paying customer on the first screen, i.e., in the top four, 80% of the time. If one considers the second screen as well, then ZEP was perfect: All customers appeared on the top two screens. In fact, the correct identity was in the top five matches 92% of the time. In our second deployment, ZEP was always perfect and showed the correct identity on the first UI screen.

5.1 BLE Detection Latency

To measure how long it takes to detect a BLE device, we performed an experiment while the coffee stand was closed and only one device was present. We placed this device at the same location where customers would stand in our cafeteria, and recorded the time every time the detector observed a beacon from it. We performed this experiment twice, once with a scan period of 0.9 sec and once with a scan period of 1.1 sec. Each experiment ran for 30 minutes. Figure 6 shows the resulting CDFs of time between beacons.

This experiment shows us two things. First, we see that using an observer scan period just over the 1-sec broadcaster period causes more broadcast beacons to be missed than using a scan period just under 1 second. This is because a long scan period can sometimes contain more than one beacon and thereby cause repeats to be ignored. Second, we see that some packet loss occurs that is not attributable to an overly long scan period. However, this loss generally only causes one or two consecutive beacons to be missed: Only 0.06% of gaps were four seconds or more and no gaps exceeded five seconds.

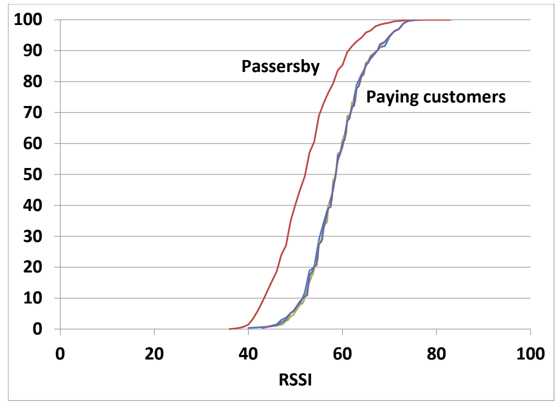

Figure 7: The BLE RSSI distributions for paying customers versus passersby.

6. Evaluation of SCI

This section presents the evaluation of our seamless customer identification scheme. We start by presenting our results on the accuracy of BLE-based identification alone, and face recognition alone. We then show that the combination of these two technologies improves accuracy to a level that makes SCI practical.

6.1 Methodology

During each ZEP transaction, the cashier selected the person paying at the register to enable payment. Since this triggered the sending of a video receipt, and since these receipts never triggered complaints about misdirected payments, we consider the cashier's selection to be "ground truth" for whose face is depicted in the central position in the transaction video.

We use this information to experimentally evaluate how well face identification would have worked under various conditions. For instance, we can simulate an arbitrary set of customers being device-present by passing their profiles to the face identification algorithm. We can also simulate an arbitrary parameter setting or algorithm by re-running face identification on the recorded frames.

Various factors can affect the performance of device detection, such as where the device's holder is standing, where and how the holder holds the device, and the remaining battery charge level of the device since it affects the voltage the battery supplies. Unfortunately, we could not control or even measure all these properties during our deployments. Thus, to evaluate their effects, we manually modify these conditions and measure their effects while the coffee stand is closed.

The face recognition results use the entire trace gathered during the first deployment, but only a two-week subset of the second trace. We are currently running all experiments on the full second trace, and we anticipate to have the results ready in several weeks. We expect none of our conclusions to change.

6.2 Performance of BLE Alone

As described earlier, many SCI scenarios require a high degree of accuracy when identifying a customer's location. For payments, it is important to detect which customer is standing exactly at the head of the checkout line. For grocery stores, it is important detect in front of what product the customer has stopped. Thus, this section investigates whether device identification using BLE alone would be adequate for determining a customer's location. For this, we analyze signal-strength measurements performed over two datasets: one gathered during our long-term ZEP deployment, and one we gathered systematically during off hours but using the same BLE infrastructure.

In indoor environments, RF signal strength can vary due to many factors including multipath interference, physical obstructions (including people), and interference from other wireless networks. Nevertheless, previous research has shown that RF signal strength can be used as an approximate measure of the distance between wireless devices [3]. The goal of our experiments is to determine, in the environment of our ZEP deployment, whether our sole BLE receiver can distinguish between the different locations just using signal strength to approximate distance.

6.2.1 BLE RSSI Data Gathered During Our Second Deployment

Our system logged the RSSI value of each BLE beacon received from one of our participants' fobs. We split this data in two datasets. The first includes the RSSI values of all customers within the last 5 seconds before making a purchase. The second include the remaining RSSI values corresponding to customers either not making any purchases (passersby) or too early to tell whether they will make a purchase. If RSSI values present in the two datasets would be disjoint, this would suggest that BLE alone could be used to identify the customer making a purchase.

Figure 7 shows the cumulative distribution function of these two datasets. For the paying customers, we used three different ways to compute their RSSI values, by averaging,

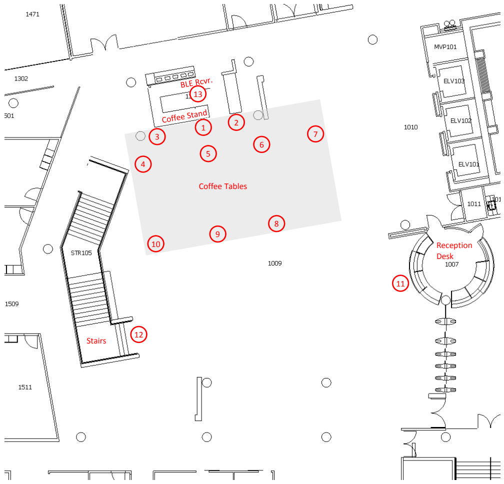

Figure 8: Diagram of the coffee stand area where the ZEP long-term deployment took place. Locations 1 through 13 indicate the locations where we recorded the signal strength of a customer's BLE radio.

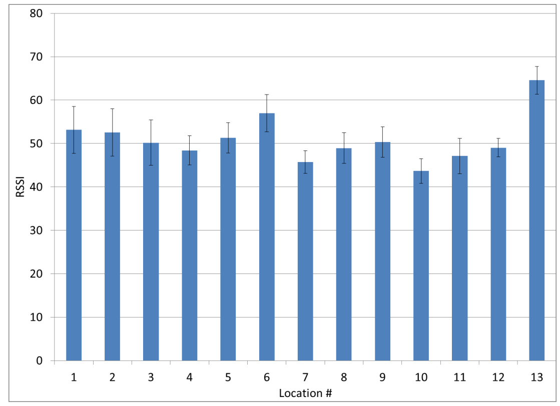

Figure 9: Mean signal strength (RSSI) and standard deviation, at locations 1 through 13.

taking the median, and taking the maximum RSSI reading within the last 5 seconds before making the purchase. While the Figure displays all these three distributions, they are very similar and thus we label them with a single label – "paying customers".

On one hand the distributions are distinct showing that paying customers have stronger RSSI values than passersby. On the other hand, there is a large area of overlap between the two distributions. In fact, a quarter of passersby have RSSI values strictly stronger than half the customers. This suggests that RSSI values alone could not have been used for identification in our deployment.

6.2.2 BLE RSSI Data Gathered Systematically

Experiment Design.

Figure 8 shows a diagram of the building lobby and coffee stand where our long-term deployment took place. Locations 1 through 13 in the diagram represent locations where we placed a customer's BLE radio, and then recorded the signal strength as reported by the BLE receiver. Location 1 in the diagram is where a customer at the head of the line would stand when making a purchase, directly across from the cash register. Locations 2 through 12 correspond to other possible locations of people in the lobby area. Location 13 is a location behind the counter, that is as close as possible to the BLE receiver (approximately 1 foot away from the receiver).

We performed experiments at each location. For certain locations we varied the position of the BLE radio on the person performing the experiment, and we varied the battery charge on the BLE radio. Each result shown in the graphs below shows the signal strength data as recorded by the BLE receiver over a two minute period. We show both the mean signal strength and the standard deviation for each result.

Results.

In the environment shown in Figure 8, location 1 is where the current paying customer is likely to stand, and it also happens to be the nearest location to our BLE receiver besides its own location 13. Figure 9 shows the signal strength from all 13 locations, where each location is identified on the x-axis. In all 13 locations, the BLE radio was in the front pocket of the person performing the experiment.

In this figure, we see a rough correlation of distance and signal strength, but with significant variation. For example, the mean RSSI at location 6 is larger than that at locations 1, 2, and 5. However, 1, 2, and 5 are all closer to the BLE receiver than location 6. In this figure, the only location that stands out as being significantly different than the others is location 13 (the closest location to the BLE receiver).

This suggests that if we were to relocate the BLE receiver to be very close to the cash register, then this might allow to us to determine which customer is standing at the register.

Figure 10 presents the effects on signal strength of where the BLE radio is located on the person carrying it. For each of five positions, we show four bars, which represent 1) in the person's front pocket, 2) in the person's back pocket, 3) in the person's right hand, and 4) inside a laptop shoulder bag that is zipped shut. From this graph, we see that the location of the transmitter on the person who is carrying it can have a large effect on the signal strength: in location 1, the in-hand signal strength is more than 12 dB larger than the back-pocket signal strength.

Finally, Figure 11 presents the effects of battery charge of the BLE transmitter on BLE signal strength at the BLE receiver. As with the previous graph, we show three bars for each of five locations. Each bar type corresponds to a different remaining-charge level, and thus a different voltage level, for the battery that powers the BLE radio. We used the same BLE transmitter, but replaced the battery to perform these experiments. In this graph, once again we find that variations in battery charge can lead to measurable differ-

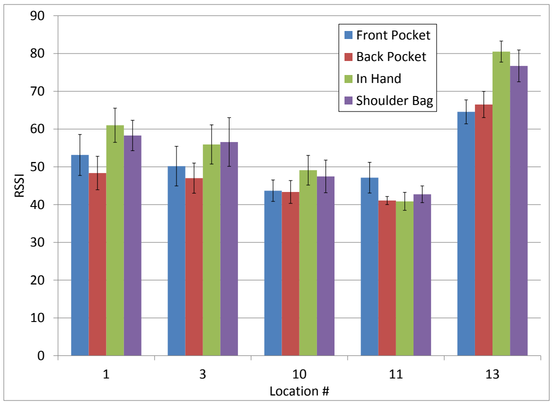

Figure 10: Comparison of BLE signal strength at different positions on a person (front pocket, back pocket, in hand, and in shoulder bag).

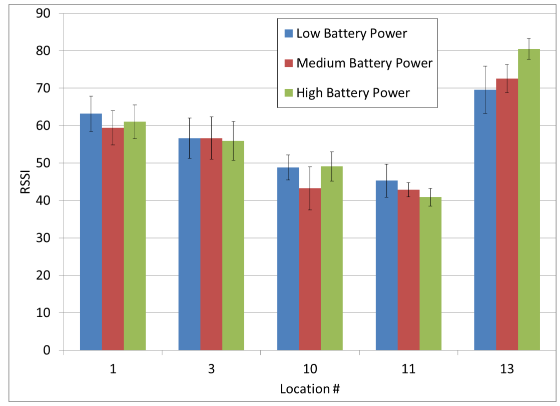

Figure 11: Comparison of different battery charges on the BLE transmitter (Low = 1.75V, Medium = 2.10V, High = 2.8V).

ences in the receiver mean signal strength. We see the most pronounced effect at location 13, which is very close to the BLE receiver, where the high charge battery has a mean signal strength that is more than 10 db larger than the battery with a low charge.

Our findings from Figures 10 and 11 indicate that even if we colocated the BLE receiver with the cash register, it would still be difficult to say with certainty which person was currently standing in front of the register. For example, in Figure 11, we see that the bottom of the standard deviation bar for location 13 with low battery power overlaps with the top of the standard deviation bar for location 1 with low battery power. As a result, we conclude that in our setting, using BLE by itself is inadequate. It is possible that if we used multiple BLE receivers and we performed RF environment profiling [2], this could provide the accuracy we need. However, such an approach would increase the deployment cost and installation overhead.

6.3 Performance of Face Recognition Alone

We now evaluate the accuracy of customer identification based on face recognition alone. Such a scenario is very ambitious. As our related work section will show (§7), face recognition accuracy in uncontrolled environments, i.e., in the wild, is far from perfect. Our use of face recognition is done frame-by-frame; more sophisticated techniques also exist. For example, face tracking is the ability to track a face across frames; face tracking could improve accuracy rates because it treats a collection of faces spanning multiple frames as one single identity. We plan to implement and investigate more sophisticated techniques like this in future work.

Our accuracy evaluation compares the rankings produced by face recognition with the "ground truth". In particular, we compute the rank of the paying customer when face recognition alone is used for identification; a rank of "1" would be a perfect match. However, in our system deployment, any rank between "1" and "4" guarantees that the customer's face immediately appears on the selector UI facing the cashier. For ranks higher than "4", the cashier would need to scroll down through the UI to find the identify of the paying customer.

Since face recognition produces a ranking of customers for every face image found in every frame immediately preceding a transaction, these separate rankings must be aggregated together to produce a single full ranking. Many separate schemes and heuristics can be used to aggregate these rankings. Based on our experience, we selected the following aggregation schemes, and used them on all frames gathered up to 20 seconds before a purchase was made:

1. Average/Median of all rankings. This scheme computes, for each customer, the average or median of all rankings that person attains for every face image in the transaction. Customers are then ranked by this measure.

2. Best ranking. This scheme computes, for each customer, the best ranking he achieves for all faces in the transaction. Customers are then ranked by this measure.

3. Largest face. This scheme chooses the largest face in a frame and uses its ranking. The intuition behind this heuristic is that the paying customer is likely to be closest to our camera. However, note that this is not perfect because (1) people have different face sizes, and (2) several people can all stand in front of the cashier even though only one is the true paying customer.

4. Top k. This scheme considers the top k matches for each ranking for each face. These rankings are then summed. An additional boost is added to the top-ranked customer for each face image. The intuition behind this top-k filtering is that beyond the first k results, face-recognition results are probably very noisy and should be discarded.

We found that the last two schemes produce the highest accuracy, so we omit presenting results for the other schemes. Figure 12 show the distribution of accuracy for each of our two traces. During TechFest, our database had 255 identities. With our off-the-shelf face recognition library

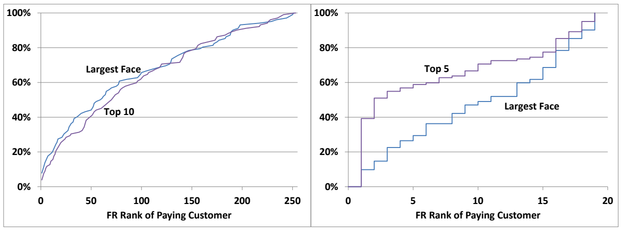

Figure 12: The accuracy of face recognition alone for the TechFest trace, on the left, and the coffee stand trace, on the right.

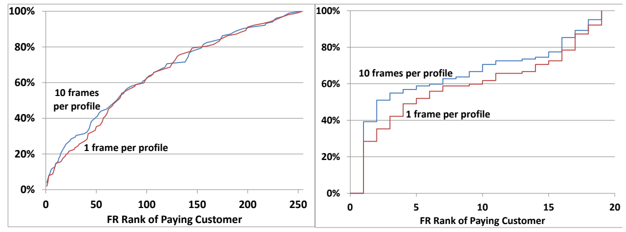

Figure 13: The accuracy of face recognition with 1 frame versus 10 frames per profile. The TechFest trace is shown on the left, whereas the coffee stand is on the right.

alone, a SCI system would rank the identity of the true customer in the top four matches, i.e., without the need for the cashier to scroll down searching, only 10–15% of the time depending on the heuristic used. Even worse, the identity of the customer would be in the bottom half of the ranking 25% of the time.

We manually inspected many of these transactions and we discovered two reasons for this lack of accuracy. First, for some transactions, the paying customer is accompanied by one of our project members during TechFest. Our project members tended to look at the camera much more often than the customer did. On such frames, face recognition alone is likely to mistake our project members for the true customer. Second, some customers faced downward because they were looking at our selector, a featured piece of equipment in our demonstration. Face recognition would struggle to find an accurate match for such frames.

During our long-term deployment, the accuracy rates improve significantly. This is unsurprising, because the database size is much smaller. While for the previous trace we used $k = 10$ for our top k heuristic, here we used $k = 5$ due to the smaller-sized database. This heuristic alone would find the true identity of the customer on the first screen of the selector, i.e., among the top four matches, more than half the time. However, in some cases, the true identity of the customer is still ranked very low. As with the previous deployment, this often occurred when the customer never looked up at the camera. Moreover, in this second deployment at the coffee stand, we had to position the camera eight feet above ground looking down at a $30°$ angle due to our privacy requirements, as described in §3.2.2. This positioning made it even less likely for customers to look "up" at the camera.

6.3.1 Richer Profiles Improve Accuracy

We also experiment with the size of the customer profile gathered at registration time. During registration, we captured a short video with the customer's face. We instructed the customer to "act naturally", and most customers looked straight at the camera for about 3–5 seconds, during which we gathered 30–70 frames. We deliberately did not instruct the customers to show different poses or do anything sophisticated to register because we envision such SCI registrations need to be as simple as possible to be viable in practice.

From this short video, we extracted a set of frames and used them as a person's "profile" during face recognition. Figure 13 shows the accuracy of face recognition alone with 10 frames per profile versus 1 frame per profile for both traces. Although the accuracy of face recognition improves with richer profiles, the improvement is not drastic. This suggests that customer registration in SCI could be done with a single face-frontal frame, similar to the single camera shot taken during a driver's-license registration, without much loss in face recognition accuracy.

6.4 Combining BLE with Face Recognition

The previous section presented the accuracy of face recognition over the entire database of SCI registrants. BLE allows SCI to drastically reduce the size of the database because the candidates for face recognition would only be people discoverable by BLE. For example, the database of our TechFest trace had 255 registrants, whereas with BLE, we rarely discovered more 10 people in the vicinity of our booth.

We systematically investigated SCI by selecting different samples of people assumed to be "nearby" due to BLE discovery. Figure 14 shows the probability of displaying the correct identity of the paying customer on the top screen (i.e., top 4 matches) and top two screens (i.e., top 8 matches) as a function of the number of people nearby. On the left, the data from deployment #1 shows that when 10 people are nearby, the correct match is found on the top UI screen 65% of the time, and on the first or second UI screen 81% of the time. On the right, the data from deployment #2 (the x-axis is using a log-scale) shows that when 10 people are nearby, the correct match is found on the top UI screen 60% of the time, and on the top two screens 90% of the time.

Putting these results in perspective, SCI systems can identify the correct identity of their VIP customers 80–90% of the time with minimal effort on behalf of the store's employee. These results assume that having more than 10 VIP customers present nearby is rare in many SCI scenarios.

7. Related Work

The techniques used for seamless customer identification

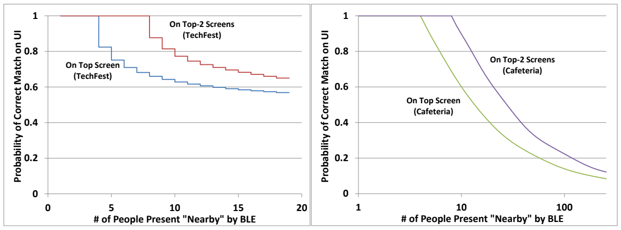

Figure 14: Probability of correct matching the paying customer on the top and top 2 UI screens as a function of the number of people nearby. On the left, deployment #1 (TechFest). On the right, deployment #2 (cafeteria); the x-axis is displayed in log-scale to zoom in on the head of the curve.

draw on previous work in wireless localization, face recognition, and mobile systems that make use of computer vision.

7.1 Wireless Localization

Over the past decade, there has been much work on using wireless radios for localization. One of the earliest projects was Radar [2], which built an indoor positioning system based on Wi-Fi signal strength. However, research projects have used a variety of types of wireless radio including Wi-Fi, RFIDs, Bluetooth, cellular, and ZigBee, to locate people indoors; excellent overviews can be found in two recent books [15, 33].

Despite all this work, wireless-based indoor localization is not a solved research problem. While wireless-based outdoor localization is now a mainstream technology present in smartphones and laptops, indoor localization has yet to become ubiquitous despite much effort.

7.2 Face Recognition

As mentioned in §2, most research work in face recognition has focused on designing new algorithms and improving their accuracy rates. In contrast, there is much less published work on the challenges facing the deployment of face recognition systems in practice. In the US, the National Institute for Standards and Technology (NIST) has put together a benchmark called the *face recognition grand challenge* (FRGC). While researchers are measuring their algorithms' accuracy against benchmarks like FRGC, these benchmarks are far from the conditions systems experience in practice. For example, each person in the FRGC dataset of profiles appears in exactly seven frames: four frontal controlled shots taken in a studio under various lighting sessions; two frontal uncontrolled shots where the subjects appear in real life; and one 3D image of the subject [20].

In contrast, our experience with SCI differs in myriad ways from this research due to the needs of application to practice. For instance, frames do not necessarily capture front shots of people; indeed, some people never look at the camera. Also, the lighting can drastically change over time, e.g., a lightbulb may stop working for a day then get replaced with a newer, much brighter bulb. Additionally, for cost and logistical reasons, we used a non-professional-grade camera.

Nevertheless, the face recognition literature [28, 35, 34] contains several projects focused on evaluating the accuracy of face recognition in more realistic scenarios. One project evaluated the accuracy of recognizing a set of 35 celebrities in videos stored on YouTube; it reported a 60–70% accuracy rate depending on the algorithm used [13]. To achieve this, the identification techniques relied on face tracking, which identifies the same person across multiple consecutive frames. Our SCI deployment did not use face tracking.

Another related project conducted an accuracy evaluation of several face recognition techniques using footage of lower quality [7]. The accuracy is evaluated using a metric called the *half-error total rate*, which is the average of false positive and false negative rates. While measuring accuracy is similar to measuring the false positive rate, a higher accuracy came at a higher false negative rate, which means that many frames reported no faces detected. Examining the results, most algorithms achieved an 80% accuracy rate only by admitting a 25–50% false-negative rate, i.e., by accepting no faces are found1 in a quarter to half of all frames.

Finally, a few other projects report high accuracy rates for face recognition in uncontrolled environments, specifically 86.3% [25] and 92% [35]. However, these results are obtained by constraining subject poses to be either front-facing [25] or constant across frames [35].

7.3 Mobile Systems and Computer Vision

Recent work has started to use computer vision in mobile systems. One application is localizing distant objects, such as buildings, by looking at them through a smartphone [16]. The combination of GPS-based localization with computer-vision processing of images gathered by a smartphone shows promising results in accurately pinpointing an object's location. Another application is cloud-based face recognition, such as that done by Google Picasa, to automatically tag photos taken by a smartphone [22]. Another project implements an indoor localization scheme based on ambience fingerprinting by observing that most stores have very distinct photo-acoustic signatures [1]. Finally, a recent workshop paper demonstrates the practicality of identifying people based on the patterns and colors of their clothes [29].

8. Conclusions

This paper introduced the notion of seamless customer identification (SCI), and showed how it could be practically built and deployed. Our key observation is that no single identification technology is sufficient, but combining device detection and face recognition lets each mitigate the other's shortcomings. Device detection can make quite accurate determinations about whose devices are nearby, since packet loss is rare. But, it cannot easily distinguish customers in different locations since signal strength is poorly correlated with location. Face recognition, on the other hand, can indicate exactly where faces are. But, it cannot determine who those faces belong to if it must choose among all registered customers rather than just those in wireless range. Together, they can provide better information than either alone to aid store employees in identifying customers.

We further demonstrated the real-world usefulness of SCI by using it to deploy Zero-Effort Payments (ZEP). ZEP customers need expend no effort to pay, since cashiers can use SCI to determine who they are and directly bill them. Across our two deployments, 274 customers made 705 purchases, and we received no complaints about the wrong customer being billed.

9. References

[1] M. Azizyan, I. Constandache, and R. R. Choudhury. SurroundSense: Mobile Phone Localization Via Ambience Fingerprinting. In *Proc. of Mobicomm*, 2009.

[2] P. Bahl and V. N. Padmanabhan. RADAR: An In-Building RF-Based User Location and Tracking System. In *Proc. of IEEE INFOCOM*, Mar. 2000.

[3] N. Banerjee, S. Agarwal, P. Bahl, R. Chandra, A. Wolman, and M. Corner. Virtual compass: Relative positioning to sense mobile social interactions. In *Proc. of the 8th International Conference on Pervasive Computing*, May 2010.

[4] Bayometric. Crossmatch Retinal Scan 2 Iris Scanner, 2012. `http://www.bayometric.com/products/i-scan-2-usb-iris-scanner-iris-scanning.htm`.

[5] Bluetooth SIG. Specification of the Bluetooth System, Core Version 4.0. Available at http://www.bluetooth.org, 2010.

[6] Canesta Corporation. Canesta, 2012. `http://en.wikipedia.org/wiki/Canesta`.

[7] S. Chen, S. Mau, M. T. Harandi, C. Sanderson, A. Bigdeli, and B. C. Lowell. Face Recognition from Still Images to Video Sequences: A Local-Feature-Based Framework. *EURASIP Journal on Image and Video Processing*, 2011(790598), 2010.

[8] S. Corporation. Iris-on-the-move, 2012.

[9] J. Daugman. High confidence visual recognition of persons by a test of statistical independence. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 15(11):1148–1161, 1993.

[10] J. Galbally, A. Ross, M. Gomez-Barrero, J. Fierrez, and J. Ortega-Garcia. From the iriscode to the iris: A new vulnerability of iris recognition systems. In *Black Hat Briefings USA*, 2012.

[11] O. Kasten and M. Langheinrich. First Experiences with Bluetooth in the Smart-Its Distributed Sensor Network. In *Workshop on Ubiquitous Computing and Communications*, 2001.

[12] M. Kaur, M. Sandhu, N. Mohan, and P. S. Sandhu. RFID Technology Principles, Advantages, Limitations & Its Applications. *International Journal of Computer and Electrical Engineering*, 3(1), Feb. 2011.

[13] M. Kim, S. Kumar, V. Pavlovic, and H. A. Rowley. Face Tracking and Recognition with Visual Constraints in Real-World Videos. In *IEEE Computer Vision and Pattern Recognition (CVPR)*, 2008.

[14] A. Koelle, S. Depp, and R. Freyman. Short-range radio-telemetry for electronic identification, using modulated RF backscatter. In *Proceedings of the IEEE*, Aug. 1973.

[15] A. LaMarca and E. de Lara. *Location Systems: An Introduction to the Technology behind Location Awareness*. Morgan and Claypool, 2008.

[16] J. Manweiler, P. Jain, and R. R. Choudhury. Satellites in Our Pockets: An Object Positioning System Using Smartphones. In *Proc. of MobiSys*, 2012.

[17] T. Matsumoto, H. Matsumoto, K. Yamada, and S. Hoshino. Impact of Artificial "Gummy" Fingers on Fingerprint Systems. In *SPIE Volume #4677 Optical Security and Counterfeit Deterrence Techniques IV*, 2002. `http://www.cryptome.org/gummy.htm`.

[18] NIST. IREX III evaluation of one-to-many iris identification algorithms, 2012.

[19] B. S. Peterson, R. O. Baldwin, and J. P. Kharoufeh. Bluetooth Inquiry Time Characterization and Selection. *IEEE Transactions on Mobile Computing*, 5(9), Sept. 2006.

[20] P. J. Phillips, P. J. Flynn, T. Scruggs, K. W. Bowyer, J. Chang, K. Hoffman, J. Marques, J. Min, and W. Worek. Overview of the Face Recognition Grand Challenge. In *IEEE Computer Vision and Pattern Recognition (CVPR)*, pages 947–954, 2005.

[21] J. Preis, M. Kessel, M. Werner, and C. Linnhoff-Popien. Gait Recognition with Kinect. In *First Workshop on Kinect in Pervasive Computing, in conjunction with Pervasive 2012*, 2012. `http://noggnogg.com/pervasivekinect`.

[22] C. Qin, X. Bao, , R. R. Choudhury, and S. Nelakuditi. TagSense: A Smartphone based Approach to Automatic Image Tagging. In *Proc. of MobiSys*, 2011.

[23] Sarnoff Corporation. IOM N-Glance, 2012. `http://www.sri.com/engage/products-solutions/iom-n-glance-modular-system`.

[24] E. Schuman. eWeek: Biometric Alternatives to Fingerprints, 2006. `http://www.eweek.com/c/a/Enterprise-Applications/Consumers-Resist-Retail-Biometrics/1/`.

[25] X. Tan and B. Triggs. Enhanced Local Texture Feature Sets for Face Recognition under Difficult Lighting Conditions. 19:1635–1650, 2009.

[26] US National Institute of Standards and Technology. Speaker Recognition Evaluation (SRE) Results, 2010. `http://www.nist.gov/itl/iad/mig/sre10results.cfm`.

[27] T. van der Putte and J. Keuning. Biometrical fingerpring recognition: Don't get your fingers burned. In *IFIP TC8/WG8.8 Fourth Working Conference on Smart Card Research and Advanced Applications*, 2000.

[28] W. Zhao and R. Chellappa and P. J. Phillips and A. Rosenfeld. Face Recognition: A Literature Survey. *ACM Computing Surveys*, 35(4):299.

[29] H. Wang, X. Bao, , R. R. Choudhury, and S. Nelakuditi. Recognizing Humans without Face Recognition. In *Proc. of HotMobile*, 2013.

[30] M. Weiser. The Computer for the 21st Century. *Scientific American*, 1991.

[31] D. A. Wheeler. Linux kernel 2.6: It's worth more! `http://www.dwheeler.com/essays/linux-kernel-cost.html`, 2004.

[32] J. D. Woodward. *Biometrics: Identity Assurance in the Information Age*. Osborne McGraw Hill, 2003.

[33] J. Yang, Y. Chen, R. Martin, W. Trappe, and M. Gruteser. *On the Performance of Wireless Indoor Localization Using Received Signal Strength – in Handbook of Position Location: Theory, Practice, and Advances*. Wiley, 2012.

[34] C. Zhang and Z. Zhang. A Survey of Recent Advances in Face Detection. Technical Report MSR-TR-2010-66, Microsoft Research Technical Report, 2010.

[35] X. Zhang and Y. Gao. Face recognition across pose: A review. *Pattern Recognition*, (42):2876–2896, 2009.