IDCam: Precise Item Identification for AR Enhanced Object Interactions

Hanchuan Li
CSE, University of Washington
Microsoft Corporation
Seattle, Washington
hanchuan.li@microsoft.com

Eric Whitmire
CSE, University of Washington
Seattle, Washington
emwhit@cs.washington.edu

Alex Mariakakis
CSE, University of Washington
Seattle, Washington
atm15@cs.washington.edu

Victor Chan
Qualcomm Research
San Diego, USA
victor.chan@kelzal.com

Alanson P. Sample
CSE, University of Michigan
Ann Arbor, USA
apsample@umich.edu

Shwetak N. Patel
CSE, University of Washington
Seattle, Washington
shwetak@cs.washington.edu

***Abstract*—Augmented reality (AR) promises to revolutionize the way people interact with their surroundings by seamlessly overlaying virtual information onto the physical world. To improve the quality of such information, AR systems need to identify the object with which the user is interacting. AR systems today heavily rely on computer vision for object identification; however, state-of-the-art computer vision systems can only identify the general object categories, rather than their precise identity. In this work, we propose IDCam, a system that fuses RFID and computer vision for precise item identification in AR object-oriented interactions. IDCam simultaneously tracks users' hands using a depth camera and generates motion traces for RFID-tagged objects. The system then correlates traces from vision and RFID to match item identities with user interactions. We tested our system through a simulated retail scenario where 5 participants interacted with a clothing rack simultaneously. In our evaluation study deployed in a lab environment, IDCam identified item interactions with an accuracy of 82.0% within 2 seconds.**

***Index Terms*—Augmented Reality, RFID, Object Recognition, Sensor Fusion, Object Interaction.**

I. Introduction

Augmented reality (AR) enables applications to seamlessly overlay virtual information onto the physical world, immersing users with interactive digital contents. Recent advances have brought dramatic changes to user experiences in many application domains, including retail, gaming, education, and remote collaboration [1], [2], [24], [30]. In retail, for example, AR could enable systems to display customizable information such as online reviews or video tutorials when users interact with products (Figure 1a). With permission, an AR system could also remind customers of potential discounts or provide relevant product recommendations. In a residential setting, AR could be used to provide real-time instructions as a user builds furniture or performs repairs. In an industry setting, AR could allow remote collaboration, delivering expert knowledge to workers when and where it is needed. These AR scenarios merge digital information with the physical world to improve users' experience and productivity as they perform daily tasks. However, none of these applications are possible without a precise understanding of the objects with which the user is interacting and the task at hand.

Fig. 1. **(a)** IDCam enables information to be automatically displayed when a customer wearing an AR headset interacts with a product; **(b)** In addition to the RFID infrastructure, IDCam prototype requires a HoloLens headset with a Leap Motion controller mounted on top.

One way to automatically detect interactions with an object would be through an AR headset's outward-facing camera. State-of-the-art systems have recently surpassed human performance at identifying objects [5], [6], yet these systems require excessive training and are limited to classifying item categories (*e.g.*, mugs, bowls, shirts) without being able to identify the specific instance of objects within each category. Real-world object-oriented scenarios require an AR system to recognize the specific identity of an object, so that they can be linked to corresponding digital augmentation. Tasks like recognizing a specific model number for a repair job or differentiating similar mugs owned by different people pose serious challenges for systems solely based on computer vision.

We propose IDCam, a hybrid system that combines two sensing modalities, computer vision and radio frequency iden-

978-1-7281-1210-7/19/$31.00 ©2019 IEEE

tification (RFID), to achieve accurate item identification for seamless information augmentation. We assume that each item is instrumented with an inexpensive RFID tag and that there are long-range RFID readers installed in the ambient environment to query these tags in real-time. We believe such a deployment is plausible for retail environments in the near future, given ongoing adoption effort by major retailers like Macy's [26]. As people interact with RFID-tagged objects, they disturb the RF signal that is reflected back to the reader from the tag. Using insights from prior work [12], these disruptions manifest in the low-level channel parameters measured by an RFID reader. IDCam captures these changes to determine which items are being manipulated and how fast they are moving. Separately, the motion of the customer's hands is tracked via their AR headset using computer vision. When these traces correlate to one another, users can be matched to the objects with which they are interacting.

IDCam is capable of identifying objects and sensing interactions even when multiple users are interacting with similar objects in close proximity to each other. To support this claim, we evaluated IDCam in a simulated shopping scenario. Two groups of five participants were asked to interact with 20 clothing items on the same rack simultaneously, which is beyond the typical density of interactions at a normal retail store. IDCam was able to match user interactions with object identities 82.0% of the time within 2 seconds.

Our contributions are the following:

1) A technique that correlates velocity calculations from RFID and computer vision with different coordinates systems to match object interactions with users, and
2) An evaluation study in a simulated shopping scenario to quantify the performance of our system.

II. Related Work

A. Object Identification in Augmented Reality

Real-time object recognition is quickly being realized thanks to advances in the field of computer vision. Large-scale open-source image databases like ImageNet [3] provide a variety of images with which researchers can train and test their systems. State-of-the art deep learning models have demonstrated promising results when given large amounts of images for training [6], [22]. However, the granularity of datasets like ImageNet is categorical; ImageNet may contain multiple brands of jeans, but the labels for all of them is the same. In a retail setting, identifying an object by its brand and size is critical. Knowing the exact identity of an object is still very challenging for computer vision given the visual resemblance of similar products (*e.g.*, the same jeans in slightly different sizes). New products can be registered in an existing model, but that would require new pictures of them taken from different angles and environments, thus incurring a large deployment cost which limits scalability.

Objects can be easily recognized if they are instrumented with visual identifiers. Prior work has looked at retrieving item information through bar codes [10], [20] and QR codes [9]. However, these interactions are not as seamless as they may sound. Acquiring a picture of a visual code requires that the code is easily accessible to the customer and that the customer can properly frame the code within the camera's view. In addition, visual code can be obtrusive since it alters the visual appearance of objects. An important advantage of IDCam is its ability to seamlessly handle identification as users interact with objects without explicit actions.

B. RF Sensing

Commercial solutions can achieve coarse-grained RFID localization using phased array antennas. For example, Impinj's Xarray [28] can achieve a localization accuracy of 1.5 m or less with 85% confidence. This level of accuracy is not sufficient for detecting object interactions; however, this could localize items by shelf in a retail setting, reducing the potential set of users who could be interacting with objects in that area. To increase localization resolution of tags, prior work has explored using synthetic aperture antennas on the RFID reader. These approaches were able to achieve centimeter level accuracy. For example, Miesen *et al.* [16], [17] used an antenna on a linear actuator to create a synthetic aperture. However, this requires that tagged objects in the environment remain static while scanning occurs. Wang *et al.* [25] used a spinning antenna to create a synthetic aperture, but their system needs densely spaced marker tags placed throughout the environment to disambiguate tag motions from reader antenna motions. Yang *et al.* [29] demonstrated the use of multiple RFID reader antennas that can locate various moving tags in harsh multipath environments. In their system, tags can be located while traveling at a constant velocity along a known trajectory (*i.e.*, on a conveyor belt). In general, RFID localization techniques has demonstrated promising results showing where objects are located in physical environments, however, prior work in RFID localization could not determine which user is interaction with what items.

Researchers have also explored how RFID systems can be used to detect user interactions with everyday object. IDSense [13] applied machine learning techniques on RF channel parameters to determine when objects were being touched or moved. PaperID [12] extended this work by providing a technique to estimate the radial velocity of tags relative to the RFID reader. Although these systems can track objects and infer interactions, they can not detect the users involved. In other words, a solely RFID-based system can sense that two objects are being manipulated in an environment, but not whether there are one or two people responsible for moving those objects or who those people are. In this work, we leverage techniques presented in prior work to extract a motion trace for each object instrumented with an RFID sticker and then correlate these traces with hand motion traces generated from computer vision to determine which user is interacting with what items.

C. Sensor Fusion

Sensor fusion has been explored through a number of different sensor combinations for identification proposes. Fusion between motion data and computer vision has been explored by the CrossMotion project [27], where visible users can be correlated with their smartphone's acceleration data as they walk through a room. Such a system is feasible for identifying people since they carry their smartphones with them at all times, but it is not practical to instrument everyday objects with IMUs. Researchers have also explored the combination of computer vision and RFID sensing for user tracking [4], [14] and item localization [19], [21]. In those studies, the computer vision and RFID sensing systems are collocated and installed at relatively fixed locations. For AR scenarios, however, augmentation happens from a first-person angle, whether through a smartphone or smart glasses. Because most RFID sensing systems are bulky, it is impractical for them to be collocated with the wearable computer vision system, meaning that they must be decoupled. In this work, we demonstrate that the two sensing systems can be decoupled and still match users with item interactions. This is made possible by a real-time coordinate transformation and trace matching system. In addition, prior systems require a few meters of user motion to correlate the motion traces from RFID and computer vision. In contrast, we demonstrate that item identity association can happen within tens of centimeters.

III. System Implementation

In this section, we will describe the implementation of IDCam through a AR retail example where IDCam recognizes customer interactions with products in a retail store and provides them with relevant information. Our description includes the hardware requirements to setup IDCam and the software processing that is necessary on the different data streams.

A. Hardware

Users who wish to leverage IDCam must wear an AR headset so that they can receive real-time information about products as an overlay. In addition, the self-localization capabilities of the headset is crucial towards our trace matching algorithm correlating users with items, which we will discuss later in this paper. In our implementation, we use Microsoft HoloLens [7] (Figure 1b). Even though the HoloLens utilized in our work (Baraboo version released 2016) is equipped with hand tracking hardware (IR cameras), hand-tracking can only be triggered by specific hand gestures. To track the user's precise hand location continuously, we mounted a Leap Motion controller [18] on top of the HoloLens (Figure 1b).

To setup a suitable environment for RFID sensing, RFID readers [8] are mounted on the ceiling to provide good coverage of the space. An RFID tag [23] is placed on each object to associate them with a unique identity. RFID tags are inexpensive, thin, and battery-free, making them as easy to use as the price tags that are already attached on products.

We implemented IDCam as a HoloLens UWP app using the Unity 3D game engine. It streams information about the user's head position and orientation to a MATLAB server running on a desktop PC. The server is responsible for matching this information with data from the Leap Motion and RFID antenna, computing the RFID and hand velocities, and detecting whether an object is being held by the user.

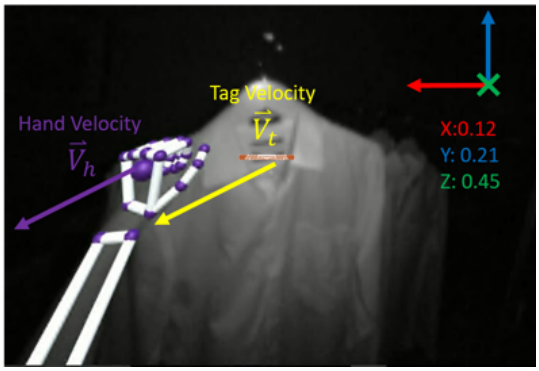

Fig. 2. A frame showing the output of the Leap Motion controller overlaid with velocity vectors for the hand ($\vec{V_h}$) and the RFID tag ($\vec{V_t}$)

B. Approach

Consider a case where a user picks up a shirt instrumented with an RFID tag while wearing an AR headset (Figure 2). Assuming the position of the RFID tag is fixed relative to the customer's hand as they take the shirt off the rack, the velocity vectors of their hand ($\vec{V_h}$) and the RFID tag ($\vec{V_t}$) at any given moment should be approximately equal, so is the trajectory of the hand $\vec{T_h}$ and the tag $\vec{T_t}$. As a result, even when there are other customers interacting with nearby items, it is unlikely that all of the customers will move items in the same way, especially with respect to the RFID antenna mounted on the ceiling. Based on this observation, users can be matched with item interactions through a correlation of velocity vectors between the RFID tag motion and user hand motion tracked by computer vision.

The hand position (denoted as X, Y, Z in Figure 2) can be accessed using the Leap Motion controller's API. Computing the change in hand position over time leads to measurement of $\vec{V_h}$ in the Leap Motion's coordinate space, which moves with the HoloLens headset. As the user moves throughout the environment, the inside-out tracking capabilities of the HoloLens report the user's position. We perform a coordinate transformation to compute the hand position in the world's coordinate system using the head position and orientation tracking APIs built into the HoloLens. We can then computing the distance between the hand and the RFID reader, which has a known, fixed location in the world coordinate system. Meanwhile, the RFID reader can report the phase angle difference between the transmitted signal and the backscattered signal. Prior work has demonstrated that the motion trace of a tag relative to the antenna can be reconstructed by utilizing the phase information retrieved from multiple communication channels [12], thus yielding a measurement of $\vec{V_t}$. Given the physical co-location of the hand and the item, the two

velocities and the corresponding motion traces should be correlated in space.

C. Coordinate Transformation

The different coordinate systems of the HoloLens, Leap Motion, and RFID reader present a challenge in correlating motion from the RFID tags and hand. In this section, we describe our approach to align data from these different systems.

First, we define the reference frames relevant to computation:

- **W**: The world reference frame, defined by the position of the HoloLens during a one-shot calibration procedure. The RFID reader is specified in this frame based on the geometry of the environment,
- **U**: The Unity world frame, defined by the placement of the app in the environment when launched,
- **C**: The HoloLens camera frame, defined within **U** by the tracking system of the HoloLens, and
- **L**: The Leap Motion frame, defined by the physical placement of the Leap Motion on the HoloLens.

We use quaternions to represent all orientations and rotation of objects.To explain the transformation, we adopt the notation from [15]. ${}^{A}_{B}\hat{q}$ represents the rotation of frame B relative to frame A, and ${}^{A}_{C}p$ represents the position of object C in frame A. Our objective is to compute ${}^{W}_{H}p$, the position of the hand in world space.

On the HoloLens, we compute the position and orientation of the HoloLens in the Unity frame, ${}^{U}_{C}p$ and ${}^{U}_{C}\hat{q}$ respectively, and stream this to the MATLAB server for further analysis. To align the Unity and world reference frames, **U** and **W** respectively, we first position the HoloLens at the origin of our desired world coordinates, pointed toward the floor. We save these values as ${}^{U}_{C_0}\hat{q}$ and ${}^{U}_{C_0}p$. We now define the HoloLens camera frame during calibration, C_0, as the world frame **W**. This allows us to treat ${}^{U}_{W}\hat{q}$ as ${}^{U}_{C_0}\hat{q}$

We can then transform new positions and orientations to compute the HoloLens position in world space.

$$ {}^{W}_{C}\hat{q} = {}^{U}_{C}\hat{q} \otimes {}^{U}_{C_0}\hat{q}^{\star} \qquad \text{Head orientation} \qquad (1) $$

$$ {}^{W}_{C}p = {}^{U}_{W}\hat{q} \otimes \left({}^{U}_{C}p - {}^{U}_{C_0}p\right) \otimes {}^{U}_{W}\hat{q}^{\star} \qquad \text{Head position} \qquad (2) $$

To include data from the Leap Motion hand tracker, we transform the raw coordinates of the hand in the **L** frame, ${}^{L}_{H}p$, to the HoloLens camera frame. We do this by transforming the right-handed coordinates of the Leap Motion data to approximately match the left-handed Unity coordinates by Equation 3.

$$ {}^{U}_{H}p' = \begin{bmatrix} -1 & 0 & 0 \\ 0 & 0 & -1 \\ 0 & 1 & 0 \end{bmatrix} \times {}^{L}_{H}p \qquad (3) $$

The LeapMotion was physically tilted downwards by 20 degrees relative to the HoloLens, so we account for that in the calculation by rotating estimated hand coordinates in the opposite direction to obtain an estimate of hand position in the camera frame, ${}^{C}_{H}p$. Finally, we obtain the hand position in world coordinates by applying the HoloLens rotation and position quaternions:

$$ {}^{W}_{H}p = \left({}^{W}_{C}\hat{q}^{\star} \otimes {}^{C}_{H}p \otimes {}^{W}_{C}\hat{q}\right) + {}^{W}_{C}p \qquad (4) $$

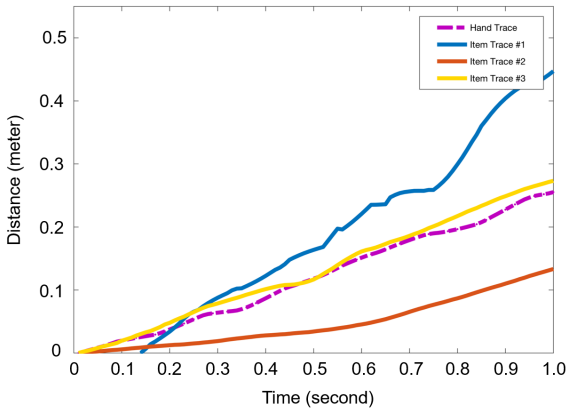

Fig. 3. A one-second segment of the motion trace tracked of hand T_h when compared to 3 different RFID traces T_t1, T_t2, T_t3.

D. Trace Correlation

There can be thousands of RFID-tagged objects and hundreds of shoppers in retail stores, so it would be infeasible to compute the velocity correlation between all possible pairs. IDCam first narrows down the set of items the user could be interacting with by only considering the items that are seen by the closest antenna to the user. The sensing area of each RFID antenna is approximately 200 ft^2, so there may still be many tags that need to be checked. IDCam further reduces the search space by filtering out tags with insignificant motion. When a person takes a shirt off of a rack, they are likely to jostle many of the nearby shirts. We observed that these interactions usually result in a tag speed of less than 10 cm/s, so IDCam uses this threshold to ignore items with no significant motion. Combining the location information of items and the speed threshold, IDCam narrows the list of potential items that the user could be handling by a significant margin.

1) Trace Correlation: We illustrate the principle behind our velocity correlation approach with an example. Assume that three people grab tagged objects under the same RFID reader. Figure 3 shows a one-second segment of position data from one user's Leap Motion (pink dotted curve) and the position data of the three tags from the RFID system (blue, orange, yellow solid curves). The position in both domains was calculated by integrating the velocity measurements. Given that the antenna can only infer the relative position change of items, all of the position traces are aligned at the origin for easy comparison. By overlaying all four trajectories, it is clear to see that the motion of Item #3 (yellow) was most correlated with the participant's hand (pink). IDCam computes trace similarity using the Pearson correlation coefficient (PCC) over 1-second windows with 80% overlap. The object that

leads to the highest correlation coefficient is likely the one that the user is manipulating, so its information is displayed on the AR headset.

Using PCC alone for matching is not enough since the calculation can be noisy. Figure 4 shows an example data set when a user picks up five different items within 30 seconds in the vicinity of other active users. Each rising edge indicates when the user takes an item off the rack (and away from the RFID reader), while each falling edge indicates putting items back onto the rack. Green sections represent when the user was correctly matched with the item they were handling. Red sections indicate mismatches, which happen when interactions conducted by other users are falsely matched with the target user. Black sections indicate that IDCam was uncertain because no significant motion was detected from the RFID tags or there was poor correlation between the two systems. Figure 4A shows an example of how well the described matching algorithm worked for this example 30 second dataset. Using the PCC correlation, IDCam has a matching accuracy of 84.0%. Further assumptions can be made in order to remove brief, spurious matching errors. It is highly unlikely that a user will handle one item for a few seconds, switch to a second item for very short period of time and then go back to the original item; it is more likely that the second item was mismatched with the customer. IDCam handles these cases by building a 1-second buffer of correlation results and then applying a majority vote to smooth out the matching results. Using this method, matching accuracy in the example scenario improves to 86.9% (Figure 4B). Nearest neighbor matching can also be used to reduce brief moments when no object motion is detected (*e.g.*, the customer is holding the item, but not moving it). Using this method, matching accuracy of the example data is improved to 89.5% (Figure 4C).

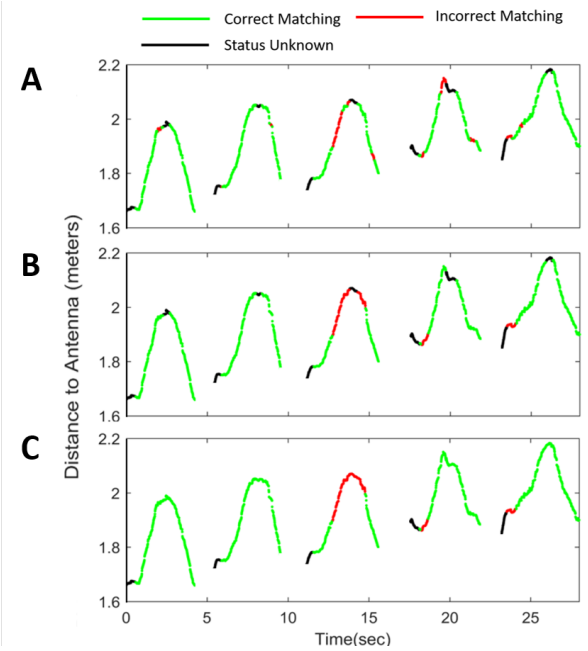

Fig. 4. A test dataset including 30 seconds of RFID/Leap motion tracking data where IDCam identifies five sequential user-item interactions. **(A)** Using PCC alone for matching leads to moments when IDCam is uncertain of of the interaction; **(B)** a 1-second voting buffer is applied to smooth the results; **(C)** nearest neighbor algorithm is applied for filling in motion gaps.

Fig. 5. Each article of clothing in the study had an RFID tag placed near the price tag.

IV. Evaluation

We designed a small-scale shopping simulation to determine how well IDCam could work in a dense space.

A. Study Procedures

A single rack of clothing 1.5 m long was was placed in the middle of a laboratory space along with an RFID reader. The rack held 20 articles of clothing that had an RFID tag placed on the center of their price tag (Figure 5). The proximity of the clothing was close enough that hangers could jostle with one another to create possible false positives. An RFID reader was placed 2 m away from the shelf on the ceiling pointing down. We recruited 10 undergraduate and graduate students from a public university (6 male, 4 female). The participants were separated into 2 groups, each with 5 participants. We believe the density of users and clothing make our simulation even more challenging than many real-world retail scenarios.

For each study session, one participant was asked to wear a HoloLens with a Leap Motion controller. Participants were asked to pick up articles of clothing by their hanger, read the product information on the price tag, and then put the hanger back on the rack. The participants not wearing the HoloLens were allowed to interact with items as they pleased, but the participant wearing the HoloLens was asked to pick up hangers in a sequence to facilitate ground truth labeling. Furthermore, that participant was also asked to extend their hands in front of the Leap Motion controller as an initialization gesture. This is because the Leap Motion controller has difficulty tracking a person's hand with a cluttered background; our contribution is not meant to be one of computer vision, but rather one of sensor fusion, so this was an acceptable limitation for us. Each of the five participants took turns wearing the HoloLens and performing the procedure mentioned above.

The start and finish time for each interaction by the person wearing the HoloLens was annotated using a video recording. There were 893 item interactions in total generated by all of

Fig. 6. The simulated shopping experience used to evaluate IDCam involved five participants interacting with clothing on a rack. One of the five participants wore a HoloLens with a Leap Motion controller mounted on top

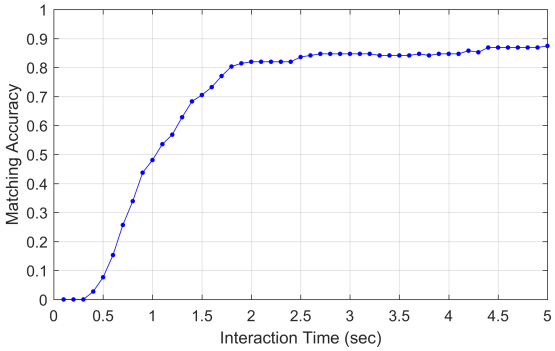

Fig. 7. The accuracy of item identification improves with longer duration of item interactions.

our participants, 200 interactions of which were generated by users wearing the HoloLens. Each session took 5.1 minutes on average. Each item interaction took 13.7 seconds on average, which means that four tagged items were typically in motion simultaneously.

B. Study Evaluation

Despite the initialization gesture we asked participants to use for better Leap Motion tracking, there were 17 interactions among the 200 collected in the user study that were not properly tracked. For the 183 item interactions that were fully or partially successfully tracked, the hand motion traces were transformed into the RFID antenna's coordinate space and compared them to RFID tags.

Figure 7 is a cumulative distribution function showing IDCam's interaction matching accuracy as a function of the time duration of the interaction traces. After 2 seconds, IDCam achieved an accuracy of 82.0%. The accuracy improved to 88.0% after 5 seconds of user interactions. At the beginning of each interaction, it is hard for our system to make accurate decisions due to the lack of motion. As users generate longer motion traces, IDCam makes better decisions using the sliding window and voting buffer approach mentioned previously.

V. Discussion

Our goal was to develop a precise object identification method for augmented reality that would be easy to deploy and scale to large number of new objects. We addressed this goal by proposing IDCam, a system that fuses velocity information from computer vision and RFID to identify the precise object-of-interest for a user. IDCam is relevant to a number of augmented reality applications, but we focused our evaluation on a simulated shopping scenario where we challenge IDCam with five simultaneous participants on the same clothing rack. In that study, IDCam correctly matched users with object interactions 82.0% of the time within 2 seconds.

The current capabilities of hand-tracking imposes a number of limitations on IDCam sensing accuracy. There were 17 cases of item interactions where the Leap Motion controller failed to track the user's hand. This was primarily due to the hand being outside of its field-of-view. There were also additional 10 cases when the hand was only temporarily tracked. We acknowledge that the current hand-tracking speed and accuracy is not ideal, but we hope that advancements in computer vision will lead to more improvements in the near future. This trend should dramatically improve IDCam's accuracy and robustness. We also hope that continuous hand-tracking APIs will be exposed to developers so that IDCam can be realized as a self-contained system.

Natural and unconstrained user behavior can be challenging to evaluate and predict. However, we believe that IDCam can better handle potential edge cases by providing users with instructions. As an extreme example, IDCam could suggest that a user shakes an item, create longer matching traces if the product is not being detected for some reason.

The reader we used to build our IDCam prototype (Impinj R420) allows for a maximum read rate around 1000 reads per second for all tags. IDCam requires a minimum sampling rate of 10 reads per second per tag to achieve a good velocity measurement. In cases where each reader has more than 100 tags within the read range, the reduced read rate per tag will limit the velocity measurement and reducing matching accuracy. Our evaluation results are generated from a small-scale lab evaluation study. A more through evaluation in real-world environments is required to fully understand the limitations of our system in terms of scalability. The methods presented in this paper repurely based on signal processing. In future work, We plan to investigate using machine learning algorithms such as SVM or CNN to help boost our trace matching accuracy. In addition, we also want to explore the possibility of a customized battery-powered wearable RFID readers with a small footprint and reduced reading range (*i.e.*, arm's reach), thereby reducing the number of tags visible to a single reader and simplifying the correlation problem.

VI. Conclusion

IDCam fuses long-range RFID with computer vision for linking objects in the physical world with digital contents. IDCam matches object identity stored in the RFID tag to user

interactions by correlating the motion traces tracked by the two systems. IDCam is able to differentiate visually similar items and scale to a large number of new items without requiring retraining. We plan to deploy and further evaluate IDCam in a real-world scenario for enhancing costumer shopping experiences. We believe IDCam has many other applications in fields such as remote collaboration, AR guidance, and education. We intend to explore and evaluate these applications in the future work.

ACKNOWLEDGEMENT

We thank Qualcomm Research for their generous support of this work. Part of the materials contained in this paper were previously included in the first author's PhD thesis [11]

REFERENCES

[1] Mark Billinghurst and Hirokazu Kato. 2002. Collaborative augmented reality. *Commun. ACM* 45, 7 (2002), 64–70.

[2] Julie Carmigniani, Borko Furht, Marco Anisetti, Paolo Ceravolo, Ernesto Damiani, and Misa Ivkovic. 2011. Augmented reality technologies, systems and applications. *Multimedia tools and applications* 51, 1 (2011), 341–377.

[3] Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei. 2009. Imagenet: A large-scale hierarchical image database. In *Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on*. IEEE, 248–255.

[4] Thierry Germa, Frédéric Lerasle, Noureddine Ouadah, and Viviane Cadenat. 2010. Vision and RFID data fusion for tracking people in crowds by a mobile robot. *Computer Vision and Image Understanding* 114, 6 (2010), 641–651.

[5] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. 2015. Delving deep into rectifiers: Surpassing human-level performance on imagenet classification. In *Proceedings of the IEEE international conference on computer vision*. 1026–1034.

[6] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. 2016. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*. 770–778.

[7] Hololens. 2018. *https://www.microsoft.com/en-us/hololens* (2018).

[8] Impinj Inc. 2018. SPEEDWAY R420 RAIN RFID READER. https://www.impinj.com/platform/connectivity/speedway-r420/. (2018). [Online; accessed 10-Feb-2019].

[9] Tai-Wei Kan, Chin-Hung Teng, and Wen-Shou Chou. 2009. Applying QR code in augmented reality applications. In *Proceedings of the 8th International Conference on Virtual Reality Continuum and its Applications in Industry*. ACM, 253–257.

[10] Hiroko Kato and Keng T Tan. 2007. Pervasive 2D barcodes for camera phone applications. *IEEE Pervasive Computing* 6, 4 (2007).

[11] Hanchuan Li. 2018. *Enabling Novel Sensing and Interaction with Everyday Objects using Commercial RFID Systems*. Ph.D. Dissertation.

[12] Hanchuan Li, Eric Brockmeyer, Elizabeth J Carter, Josh Fromm, Scott E Hudson, Shwetak N Patel, and Alanson Sample. 2016. Paperid: A technique for drawing functional battery-free wireless interfaces on paper. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*. ACM, 5885–5896.

[13] Hanchuan Li, Can Ye, and Alanson P Sample. 2015. IDSense: A human object interaction detection system based on passive UHF RFID. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*. ACM, 2555–2564.

[14] Hanchuan Li, Peijin Zhang, Samer Al Moubayed, Shwetak N Patel, and Alanson P Sample. 2016. Id-match: a hybrid computer vision and rfid system for recognizing individuals in groups. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*. ACM, 4933–4944.

[15] Sebastian Madgwick. 2010. An efficient orientation filter for inertial and inertial/magnetic sensor arrays. *Report x-io and University of Bristol (UK)* 25 (2010).

[16] Robert Miesen, Fabian Kirsch, and Martin Vossiek. 2011. Holographic localization of passive UHF RFID transponders. In *RFID (RFID), 2011 IEEE International Conference on*. IEEE, 32–37.

[17] Robert Miesen, Fabian Kirsch, and Martin Vossiek. 2013. UHF RFID localization based on synthetic apertures. *IEEE Transactions on Automation Science and Engineering* 10, 3 (2013), 807–815.

[18] Leap Motion. 2018. *https://www.leapmotion.com/* (2018).

[19] Theresa Nick, Sebastian Cordes, Jurgen Gotze, and Werner John. 2012. Camera-assisted localization of passive rfid labels. In *Indoor Positioning and Indoor Navigation (IPIN), 2012 International Conference on*. IEEE, 1–8.

[20] Harry E Pence. 2010. Smartphones, smart objects, and augmented reality. *The Reference Librarian* 52, 1-2 (2010), 136–145.

[21] Alanson P Sample, Craig Macomber, Liang-Ting Jiang, and Joshua R Smith. 2012. Optical localization of passive UHF RFID tags with integrated LEDs. In *RFID (RFID), 2012 IEEE International Conference on*. IEEE, 116–123.

[22] Yaniv Taigman, Ming Yang, Marc'Aurelio Ranzato, and Lior Wolf. 2014. Deepface: Closing the gap to human-level performance in face verification. In *Proceedings of the IEEE conference on computer vision and pattern recognition*. 1701–1708.

[23] Alien Technology. 2016. ALN-9740 Squiggle Inlay (Higgs™ 4). http://www.alientechnology.com/wp-content/uploads/Alien-Technology-Higgs-4-ALN-9740-Squiggle.pdf. (2016). [Online; accessed 10-Feb-2019].

[24] Bruce H Thomas. 2012. A survey of visual, mixed, and augmented reality gaming. *Computers in Entertainment (CIE)* 10, 1 (2012), 3.

[25] Jue Wang and Dina Katabi. 2013. Dude, where's my card?: RFID positioning that works with multipath and non-line of sight. In *ACM SIGCOMM Computer Communication Review*, Vol. 43. ACM, 51–62.

[26] Macy's Inventory will be 100 percent RFID-tagged by 2017 Supply Chain Dive. 2017. *http://www.supplychaindive.com/news/Macys-RFID-inventory-tracking/428937/* (2017).

[27] Andrew D Wilson and Hrvoje Benko. 2014. Crossmotion: fusing device and image motion for user identification, tracking and device association. In *Proceedings of the 16th International Conference on Multimodal Interaction*. ACM, 216–223.

[28] Impinj Xarray. 2018. *https://www.impinj.com/platform/connectivity/xarray/* (2018).

[29] Lei Yang, Yekui Chen, Xiang-Yang Li, Chaowei Xiao, Mo Li, and Yunhao Liu. 2014. Tagoram: Real-time tracking of mobile RFID tags to high precision using COTS devices. In *Proceedings of the 20th annual international conference on Mobile computing and networking*. ACM, 237–248.

[30] Steve Chi-Yin Yuen, Gallayanee Yaoyuneyong, and Erik Johnson. 2011. Augmented reality: An overview and five directions for AR in education. *Journal of Educational Technology Development and Exchange (JETDE)* 4, 1 (2011), 11.