

IDSense: A Human Object Interaction Detection System Based on Passive UHF RFID

Hanchuan Li^{1,3}, Can Ye^{2,3}, Alanson P. Sample³

Computer Science & Engineering¹
University of Washington
hanchuan@cs.washington.edu

Electrical & Computer Engineering²
Carnegie Mellon University
yecan@cmu.edu

Disney Research, Pittsburgh³
{alanson.sample}
@disneyresearch.com

ABSTRACT

In order to enable unobtrusive human object interaction detection, we propose a minimalistic approach to instrumenting everyday objects with passive (i.e. battery-free) UHF RFID tags. By measuring the changes in the physical layer of the communication channel between the RFID tag and reader (such as RSSI, RF phase, and read rate) we are able to classify, in real-time, tag/object motion events along with two types of touch events.

Through a user study, we demonstrate that our real-time classification engine is able to simultaneously track 20 objects and identify four movement classes with 93% accuracy. To demonstrate how robust this general-purpose interaction mechanism is, we investigate three usage scenarios 1) interactive storytelling with toys 2) inference of daily activities in the home 3) identification of customer browsing habits in a retail setting.

Author Keywords

RFID; Activity Detection; Object Interaction; Touch Interface

ACM Classification Keywords

H.5.m. Information interfaces and presentation (e.g., HCI): Miscellaneous.

INTRODUCTION

Effective means of identifying people's activities in their indoor environments has the potential to enable a wide number of human-computer interaction applications [7,14,15]. One key observation is that the objects we interact with provide rich contextual information about the state of our environment and the activities that we are doing. Whether it's reading a book to a child, cooking a meal or fixing a bicycle, the objects that we use both define and reflect the activities we do in our daily lives.

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

CHI 2015, April 18 - 23 2015, Seoul, Republic of Korea
Copyright 2015 ACM 978-1-4503-3145-6/15/04...\$15.00
<http://dx.doi.org/10.1145/2702123.2702178>

The challenge is to create an unobtrusive and general purpose approach to monitoring human object interaction via computer systems. A variety of sensing approaches have been proposed and shown that activity recognition is possible based on object interaction [10,11,13]. One common approach is to instrument objects with wireless sensor nodes using accelerometers (or other sensors) to infer object interactions. This approach can provide high-fidelity streaming sensor data but due to their relatively high per unit cost, large size, and need for battery replacement these methods have found limited usage for object-based activity monitoring.

In this paper we propose IDSense, a new human object interaction detection technique which uses commercially available passive Ultra High Frequency (UHF) RFID tags and readers to detect human object interactions in the form of motion and touch. Combined with the ID information inherently provided by the RFID tags, our approach enables interaction identification for a wide variety of daily objects. This is accomplished by observing changes in the physical layer signals of the communication channel between the RFID reader and the passive tags. The key insight is that the channel parameters reported by the RFID reader, such as Received Signal Strength Indicator (RSSI), RF Phase, and Doppler shift represent a unique signature of the RF environment of each individual tag. By observing changes in

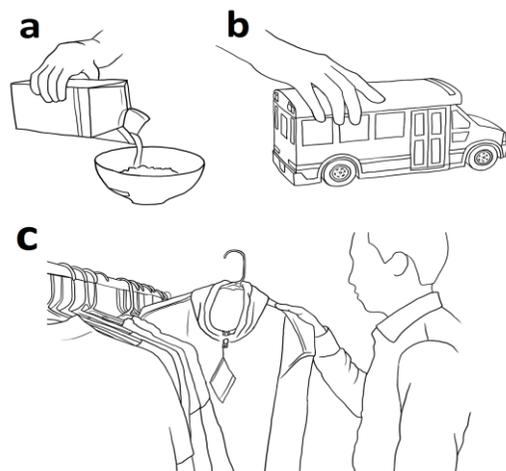


Figure 1: IDSense applications include (a) activity inferring of daily tasks (b) interactive storytelling with low cost toys (c) identification of customer browsing habits

these parameters over time, inferences can be made about the state of the tag and thus the object the tag is attached to.

IDSense is a scalable, real-time object interaction detections system that can robustly monitor large rooms and living spaces with a single RFID reader. Everyday objects can easily be retrofitted with RFID “stickers” or integrated with RFID tags by the manufacturer. Users are able to naturally interact with 10s to 100s of tagged objects and the system detects events such as object motion and tag touch. By observing these interaction events over time, it is possible to enable a wide variety of applications such as inference of daily activities in the home (Figure 1a), interactive storytelling using low-cost tangible toys with computer-based media (Figure 1b), enhanced retail experiences where interactions with tagged merchandise can be used to determine customer interests (Figure 1c). Generally speaking, IDSense makes sensing human interaction with everyday objects easy and unobtrusive, by minimally augmenting objects with low-cost and long-lived RFID tags.

Contributions:

We develop a new human object interaction detection technique based on commercially available, long-range RFID technology. This system is capable of robustly classifying tagged objects using a single RFID reader and antenna in home and office environments. The contributions of this work are as follows:

1. For the first time, we use RF Phase information along with other key low-level channel parameters, such as RSSI and read rate to create multi-path invariant features for object interaction detection
2. Create a classifier capable of identifying object motion and two types of touch events.
3. Implement a real-time, data acquisition and classification system.
4. Explore three usage scenarios: 1) interactive storytelling with toys, 2) inference of daily activities in the home, 3) identification of customer browsing habits in a retail setting.

RELATED WORKS AND APPROACHES

Automatic means of activity recognition is one of the key building blocks needed to enable new and novel human-computer interactions systems. There is a variety of approaches to accomplish this goal each with their own unique strengths and weaknesses.

User-centric systems require that individuals wear sensor nodes [1] or a smartphone [6] that continually collect data throughout the day. These types of systems show good results for determining the state of the body (i.e. running, walking, and sitting), but it is difficult to determine higher level activities such as cooking meals or reading a book. In order to gain a better understanding of user activity, Ren et al. [12] used body worn cameras to identify handheld objects

This approach provides a rich amount of data but like all computer vision systems, cameras require line of sight, significant amounts of computational resources, and raises privacy concerns. Furthermore, all user-centric systems require that the user actively participates at all times by wearing a device and maintaining its battery level which is not applicable for passive and/or intermittent users.

The work presented in this paper focuses on the alternative paradigm where inexpensive sensors are densely distributed throughout the environment thereby freeing the individual to go about their daily routine. A straightforward approach is to instrument objects with wireless sensor nodes that use accelerometers (or other sensors) to infer object interactions [13] This approach can provide high-fidelity streaming sensor data but due to their relatively high per unit cost (tens of dollars), large size, and need for battery replacement, these methods have found limited usage for object based activity monitoring. Thus, many research efforts have focused on lowering cost and improving the lifetime of battery based sensing systems.

At the extreme end of the hardware spectrum are RFID tags, which are battery-free (passive), fixed-function devices that cost 10 to 20 cents and report their unique ID when energized and interrogated by an RFID reader. Depending on the RFID technology used, the read range can be a 30cm for near-field systems or 10m for far-field UHF RFID systems. Since typical RFID systems only report the binary information that a tag is within range of a reader or not, researchers have focused on developing new types of tags and new ways of inferring tag activity.

Philipose et al. [11] developed a wrist worn near-field RFID reader system that could identify objects tagged with button size RFID tags. Although this system did require the user to wear an RFID reader bracelet, this work demonstrated the feasibility of identifying daily activities solely from tracking human object interaction events. Buettner et al. [2] used the Wireless Identification and Sensing Platform (WISP), which is a battery-free, long range RFID tag enhanced with an accelerometer to detect movement of a tagged object. RFID readers were placed in the ceiling of a living room environment and the WISPs reported move events along with their ID. This system was capable of inferring 12 daily activities in a home setting. Since the WISP is still a research platform its relatively high pre-unit cost of more than \$100 USD does not make it feasible yet for large-scale deployments. Both of these examples demonstrate that object interaction events can be utilized to reliably infer activities.

Early work by Fishkin et al [5] demonstrates the feasibility of passive UHF RFID based motion detection by measuring changes in tag read rate to infer object motion. They showed good results for detecting object rotation but, as the authors stated, their system was “nearly unable to detect translation-only movement”. Furthermore, this system required multiple RFID reader antennas and multiple tags on each object.

The closest prior work to our approach has been done by Parlak et al. [10] who presented a passive UHF RFID detection system specially designed for a trauma resuscitation scenario. The authors focused on constant rotation and linear tag movement at speeds of 1m/s and showed an average of 90% accuracy in motion detection based on a binary classifier in four lab controlled scenarios, and approximately 80% accuracy for real-world scenarios. The motion detection classifier described in their system is based solely on RSSI data and thus requires multiple antennas to detect tags that move laterally to reader antenna at a constant distance.

It should be noted that the IDSense system presented here does not require multiple RFID reader antennas or multiple RFID tags on a given object, while still achieving high accuracy motion detection with low false positives. In fact, all the applications shown are done with a single antenna in real-world multipath environments. Additionally, we demonstrate two types of touch not previously reported. Thus, this system provides a flexible solution for object interaction detection, which can enable a wide number of human-computer interaction scenarios.

SYSTEM OVERVIEW AND IMPLEMENTATION

The goal of this project is to use minimalistic hardware in the form of commercially available RFID tags to provide enough sensing capability to robustly detect basic human object interactions. Figure 2 shows an example of a toy ambulance augmented with a UHF RFID tag on its hood; the copper antenna is most visible in panel C. Good RF engineering practices should be observed when choosing the type of tag and it’s placement on the object. There is a wide variety of tag shapes and sizes to choose from, as well as tags specially designed for glass and metal object. Generally speaking, it was not difficult to find good tag locations but some trial and error can be expected.

Three different object states are investigated as well as two different touch events. The primary state is “object still”, meaning no interaction with tagged objects. The second state

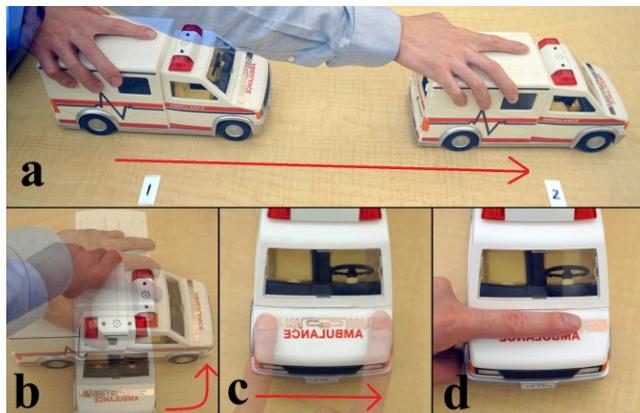


Figure 2: Types of human object interactions a) translation b) rotation c) swipe touch d) cover touch

is object translation as shown in Figure 2a, where an object translation is defined as movements of greater than 10cm within 2 seconds. The third state is rotation as depicted in Figure 2b) which consists of a 90° rotation around one of the objects axis. Swipe touch (shown in Figure 2c) consist of the user swiping their finger across the tag antenna within 2 seconds. Finally, a cover touch is when a user touches more than half the tag antenna for a minimum of 1 second. It should be noted that small and/or extremely short duration movements were not specifically studied and only natural human interactions are explored.

This work focuses on common indoor environments such as the home and office with no special considerations given to building materials, room selection or furniture placement. Figure 3 shows an image of the lab environment where initial testing and validation of the IDSense system was done. The RFID reader antenna is placed on top of a ceiling panel (highlighted in pink) pointed downwards. A coax cable leading back to the Impinj Speedway Revolution reader and host computer is visible. The read distance of UHF RFID tags in free space is +10 meters. In this example, the reader coverage zone extends from the gray workbench on the left to the gray workbench on the right and includes the wooden table and floor in the middle of the frame. In this region per tag read rate (or sampling rate) is between 15-40 reads per second depending on the size of the tag population. As a point of reference, it is possible to read tags on the far white shelves in the background but read rate is typically below 10 reads per second. As will be discussed later in the paper, proper RFID reader antenna placement is important to achieve good performance and reduce false detection caused by human activity, and a detailed discussion of room coverage can be found in [4].

Physical Layer Signals of UHF RFID Systems

The commercially available UHF RFID system used in this work is capable of reporting low-level channel parameters such as Received Signal Strength Indicator (RSSI), RF Phase, and Doppler shift as well as the unique identification number of each tag. The RFID reader interrogates tags within

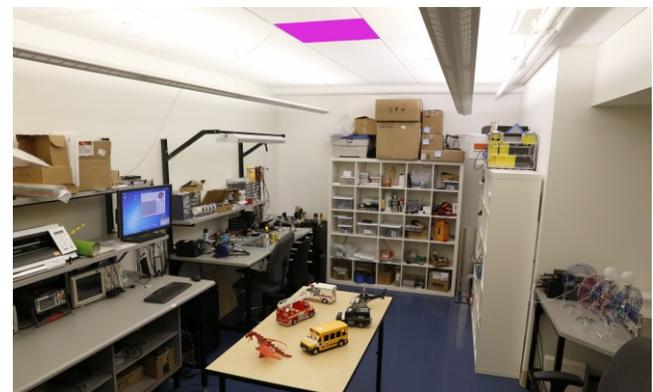


Figure 3: Image of the lab environment used for initial testing. The RFID reader antenna is placed on top of a ceiling panel (highlighted in pink) pointing downwards.

its range according to the ISO-18000-6C specification, based on the Slotted Aloha protocol, and has a maximum theoretical read rate of 1,200 tags/sec. Given proper reader settings, it has been observed that an individual tag can be read at ~ 90 reads/second and a population of 10 tags can be read at ~ 330 reads/sec with individuals reading at 30 reads/sec. Using a naïve approach a large population will begin to saturate the system. For instance, 60 tags had a total read rate of 616 reads/sec and thus the individual rate dropped to around 10 reads/sec. To overcome this limitation, it is possible to programmatically mask sub-populations of tags to increase read rate. This technique was not needed for this work and as a rule of thumb, the typical per tag read rate was between 15-40 reads per second for all experiments.

In RFID systems, RSSI is a measurement of the signal power received at the reader and is predominantly affected by large changes in the distance between the tag and the reader. RF phase is a measure of the phase angle between the RF carrier transmitted by the reader and the return signal from the tag. Phase is dominated by small changes in distance and/or in carrier frequency and repeats every wavelength. Finally, Doppler is the frequency shift between the transmitted and reflected signals caused by quickly moving objects. Each time a tag is read, the RFID reader measures these physical layer channel parameters and reports them along with the tag ID and the transmit frequency to our real-time host application. To retrieve the low-level data streams, we implemented a reader communication software in C# using Octane SDK provided by Impinj.

A plot showing 60 seconds of raw RSSI and Phase data for a single tag is depicted in Figure 4, panel a. The tag is “still” during the first 20 seconds, next the tag is “moved” for 20 seconds, and for the remaining 20 seconds it is “still”. This sequence of events can be inferred from the RSSI data, but the phase data does not show a discernable trend. This is due to FCC regulations which require RFID readers in the 915MHz ISM band to pseudo-randomly change their transmit frequency in order to minimize interference with other devices. The result is that the RFID reader must “frequency hop” across 50 channels from 902MHz to 928MHz (in the USA) at an interval of approximately 0.2 seconds. This causes significant discontinuities in the RF phase reported by the reader as a function of time (see Figure 4a), which makes detecting tag movements particularly difficult. However, the RFID reader also reports which channel (aka frequency) was used when a tag is read. Thus, re-mapping the window of the RF phase data from time into transmitted frequency (as shown in Figure 4c) reveals well-defined structures that can be used to build classification features.

One of the key insights is that these low-level channel parameters represent a snapshot of the RF environment that is unique to each tag. Each tag’s RF environment is comprised of, the far-field signal path from the reader to the tag (including all multipath elements), as well as the objects

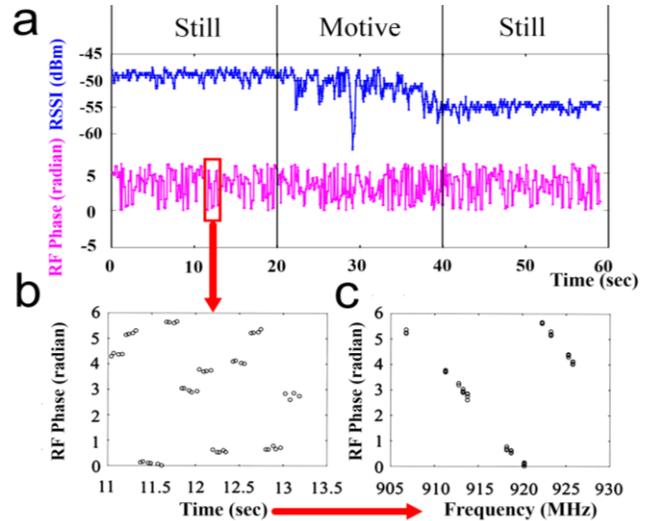


Figure 4: Raw RSSI and Phase data for a single RFID tag undergoing an interaction event (panel a). Due to pseudo-random frequency hopping a two second windows of phase data (panel b) must be sorted by channel (i.e. frequency) to reveal the tag expected phase behavior (panel c).

within the near-field region of the tag, which has an effective radius around the tag of a half wavelength (~ 16 cm). Thus, any changes in distance and/or tag orientation will result in altering the signal paths and will be reported as changes in RSSI and/or RF phase. By watching the change in these parameters over time, the state of an individually tagged object can be inferred. Furthermore, changes in the near-field region of the tag (such as hand touch) will alter its resonant frequency and/or the impedance match between the RFID IC and the antenna. Both of these effects will be reported as changes in RSSI and RF phase as reported by the reader.

Feature Selection and Machine Learning

Through experimentation, it was determined that 2 seconds is approximately the upper bound needed for our participants to complete translation, rotation, swipe touch, and cover touch interaction in a natural fashion. Thus, a 2-second sliding window (which is advanced each second) was employed to segment the RSSI and RF phase data stream to generate features as inputs for the object interaction classifier. A longer window could be used to identify longer object interaction, but it would also increase the latency of the real-time system.

Figure 5 shows examples of the raw RSSI and RF Phase signals for the same object undergoing four different types of interactions. Figure 5a shows a still tag (i.e., no human interaction) with the RSSI vs time plot on the top and the RF phase vs. transmit frequency plot on the bottom. The RSSI vs time plot is relatively stable for a two second time window and the RF phase vs. frequency plot shows RF phase decreasing at a constant slope. Translation (Figure 2b) and rotation (Figure 2c) has a major influence on RF Phase variation while swipe touch (Figure 2d) has a major influence on RSSI variation.

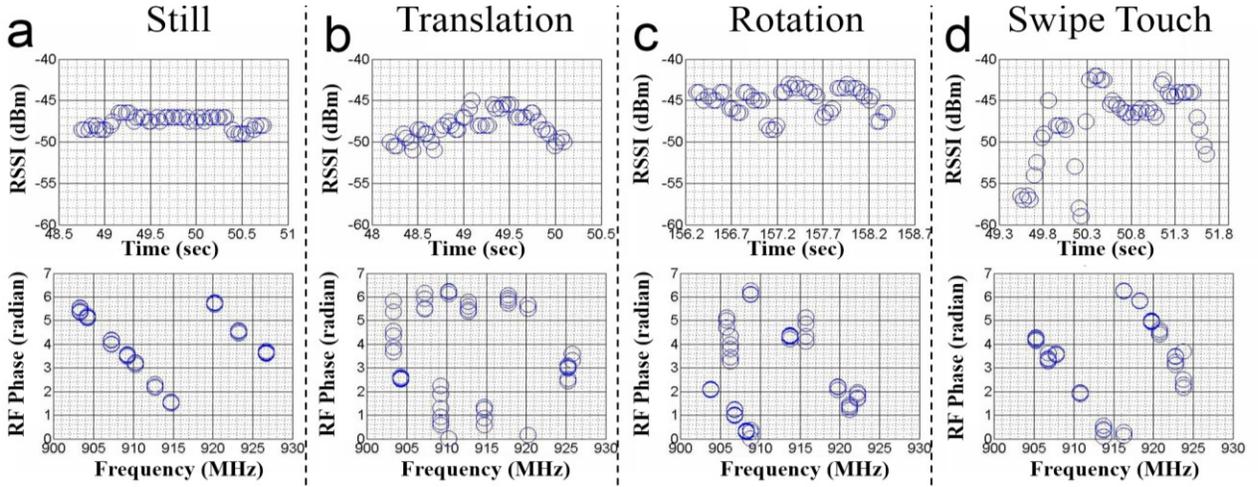


Figure 5: Tag RSSI vs time and phase vs frequency for Still, Translation, Rotation, and Swipe Touch over a 2 second window

To minimize the influence of RF signal multipath effect, the features are based on differentials rather than absolute values of RSSI and RF Phase. Eight features have been implemented from RSSI, RF Phase, as well as read rate.

RSSI Features

Generally speaking, changes in RSSI are predominantly caused by changes in the distance between the reader and the tag as well as the orientation of the tag antenna. However, it is well known that multipath effects can cause unpredictable variations in signal strength between a transmitter and receiver. In real-world settings, multipath increases the spatial variation in RSSI and thus providing a greater likelihood of detecting motion events. To identify these changes the following features have been selected.

1. Standard Deviation of RSSI
2. Mean of RSSI Standard Deviation within each frequency.
3. Mean of difference between neighboring RSSI

RF Phase Features

RF phase is sensitive to smaller changes in distance between the tag and reader and is particularly useful for detecting translations. Additionally in Figure 4, the frequency hopping effect demonstrates RF Phase dependency on channel frequency, which results in the phase related features being divided into two subgroups: the Constant Frequency Phase Rate (CFPR) and the Variable Frequency Phase Rate (VFPR). Since the RFID reader performs many tag reads on a single frequency before hopping to the next channel, changes in the phase are a good indicator of an interaction event. Here we define the Constant Frequency Phase Rate as:

$$CFPR = Phase[i + 1] - Phase[i] \quad (1)$$

Where $Phase[i+1]$ and $Phase[i]$ are neighboring RF phase measurements at the same frequency (for a given time window). We use the following three features to represent variations of Constant Frequency Phase Rate caused by human tag interaction.

4. Median of the CFPR
5. Sum of the absolute values of CFPR
6. Standard Deviation of the CFPR

When the RFID reader does frequency hop from one channel to another, the change in frequency adds an additional dimension of information to infer human tag interaction. Equation (2) shows that the distance between the reader and tag is proportional to the partial derivative of the phase with respect to the derivative of frequency as described in [9].

$$d = -\frac{c}{4\pi} \frac{\partial \phi}{\partial f} \quad (2)$$

Therefore, the Variable Frequency Phase Rate is defined in equation (3) as the incremental change in phase divided by the incremental change in frequency.

$$VFPR = \frac{Phase[i + 1] - Phase[i]}{Frequency[i + 1] - Frequency[i]} \quad (3)$$

Finally, since VFPR is proportional to the distance of the tag to the reader, the standard deviation of VFPR is used to determine tag motion.

7. Standard Deviation of the VFPR

Read Rate per Tag

A cover touch event on a tag will dramatically weaken the received signal strength, which results in decreased read rate. Read rate of an uncovered tag ranges from 15 to 40 per second while a covered tag (half covered or fully covered) usually has a read rate less than 10 reads per second.

8. Read Rate: Number of packets received from each RFID tag per second.

Doppler Features

Doppler shift is used in a number of radio sensing scenarios to infer the relative motion of two radio systems. We developed several features using the Doppler information reported by the reader; unfortunately due to the relatively low speeds of human motion and the large amount of noise in the signal, these features did not prove to be expressive enough when compared to the other eight features. It should be noted that RFID-based Doppler shift features would be useful in fast moving scenarios such as outdoor sporting activities and automotive settings.

Motion and Touch Event Classifier

In a pilot study conducted in the lab setting depicted in Figure 3, we determined that our system is able to detect horizontal translation greater than 20 centimeters in distance, vertical translation greater than 10 centimeters, and rotation of more than 45 degrees; all in a 2 second window. One limitation of the system is that when objects are moved very slowly, there may not be a significant RF signal change (in the two second sliding window) to be classified as an interaction event. At the other extreme, it could be possible to start and complete an object moment so quickly that the data would only appear as a brief impulse, thus making classification difficult. This edge condition was difficult to produce and was not observed in practice.

In a pilot study, 600 instances of interaction events were recorded with one participant interacting with one tagged toy doing 5 types of interactions: still, translation, rotation, swipe touch and cover touch. For translation, the toy is moved by 20 centimeters and in the rotation class, the toy is rotated by 90 degrees. Interactions are conducted on a table top with the RFID reader antenna mounted on the ceiling facing downwards at a distance of approximately 3 meters. A real-time classifier using Support Vector Machine (SVM) [3] with Radial Basis Function (RBF) kernel was implemented in Matlab, which received streaming RFID read events from a Java application over TCP/IP.

We trained a 5-class classifier based on the 600 annotated instances and tuned parameters in the RBF kernel by 10 fold cross validation. This approach was able to achieve an 86.0% accuracy. However, the major classification confusion occurred between the translation class and the rotation class, which could only be distinguished with a 74.7% accuracy. It is believed that using multiple tags on a single object could improve these results in the future.

Since the goal of this work is to robustly identify human object interactions, we combined the rotation and translation classes into one “motion class”. The final classifier detects human tag interaction including still, motion, swipe touch and cover touch. The 10-fold cross-validation results in a 4-class classifier, which shows an improved accuracy of 95.7%. It should be noted that our classifier is invariant to which object was used to train the classifier. Thus annotated data from all objects are being used to train a uniformed model. In later sections, we demonstrate that this classifier can be used by different participants, without the need for retraining. This makes sense, since the system is detecting that an object is being interacted with, not a gesture or action that is unique to an individual.

The real-time classifier reports results once per second for each tag, which is accomplished by sliding the two-second window over the data stream in one-second intervals. The following sections investigate several applications using the IDSense system and are implemented using the real-time classifier.

Classified as →	Still	Trans	Rotate	Swipe	Cover
Still	97.5%	1.7%	0.0%	0.8%	0.0%
Trans	0.8%	68.3%	28.3%	2.5%	0.0%
Rotate	0.8%	20.0%	74.2%	4.2%	0.8%
Swipe	0.0%	3.3%	3.3%	93.3%	0.0%
Cover	0.0%	0.8%	0.8%	1.7%	96.7%

Table 1: 10 fold cross-validation result for 5-class classifier

IDSense APPLICATION AND EVALUATION

One of the key tradeoffs of this approach is the we are inherently sacrificing rich sensing data of an object’s state, for a low-cost method of instrumenting that object with a passive RFID tag. The previous section demonstrated that the system is fundamentally capable of identifying basic human object interaction in a normal office/lab environment. To gain a better understanding of the capabilities of the system, it is useful to focus on a few application spaces.

The following sections explore three application scenarios and provide a deeper analysis of system performance. It should also be noted that human subject approval was obtained for all studies in this paper and the RFID equipment used is commercially available and meets the FCC regulations for health and safety as well as radio interference.

Interactive Storytelling with Physical Toys

An interactive storytelling application is shown in Figure 6 where a stuffed toy lion is enhanced with a low-cost RFID collar that communicates with a higher cost game console (RFID reader) connected to a computer or TV. When a child plays with the real toy, interaction events are recorded by the IDSense system which triggers actions by a virtual character on a computer screen. For instance, a swipe touch near the collar is interpreted as petting the lion while a cover touch triggers the character to take a nap. Likewise shaking the lion causes the digital character to dance. Any of these actions can advance the plot line of the story and be used to trigger visual and audio feedback. Since RFID tags inherently provides unique identification information, multiple toys can be used simultaneously to create complex and dynamic stories. Additionally, each toy can be personalized based on previous story lines or user’s preferences using a database or the writeable memory in the RFID tag. Ultimately IDSense offers an unobtrusive way to bridge interactive digital media with real-world toys and objects.

Study Design Overview

In this study, we evaluate the performance of the system at classifying events necessary to support the interactive storytelling scenario. These events include the toy being: still, in motion, swipe touched, and cover touched. In this case, the “motion class” includes both toy translation (Figure 2a) and rotation (Figure 2b).

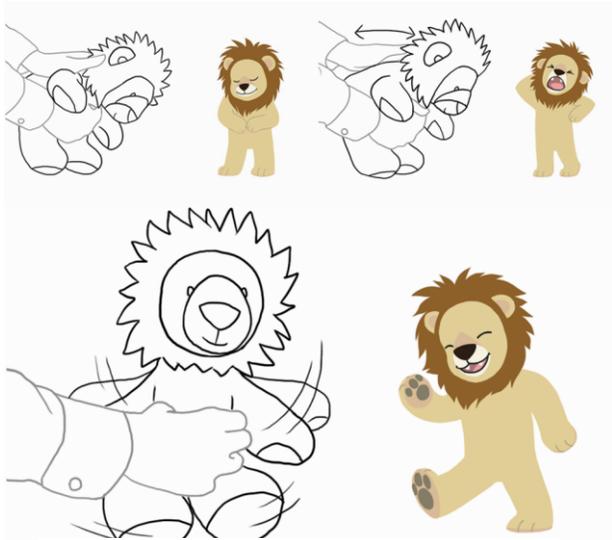


Figure 6: An example of interactive storytelling where interaction with a physical toy lion, such as petting, holding, and shaking triggers digital character actions and plot events.

Five tagged toys were placed on a table measuring 140cm x 70cm as shown in Figure 7, which is in the same location as shown in Figure 3. The table was divided into 6 sections marked with number 1-6. For move events, the toys were moved from one section to another, distance between neighboring sections was approximately 20 to 35 centimeters, and non-neighboring sections approximately 40 to 90 centimeters.

Study Procedures

We recruited 11 participants including 7 males and 4 females, with a mean age 25.7 years. Each participant finished the study independently, 11 studies spanned over a period of two weeks. During the study, each participant was asked to follow visual instructions on a monitor to perform 10 instances of each of the following interactions on 2 randomly selected toys. Instructions were given once every 5 seconds. The instruction script is used as ground truth and user's mistakes are manually annotated. We also monitor interaction records on the 3 unselected toys to test false alarms triggered by interaction with nearby objects.

1. Translation: Translate 2 toys simultaneously between 6 sections for 10 instances (Figure 2a, Figure 7). 5 translations to neighboring sections and 5 translations to non-neighboring sections. Note that 2 toys were following different paths during translations.
2. Rotation: Rotate 2 toys simultaneously by approximately 90 degrees for 10 instances (Figure 2b)
3. Swipe touch: Perform 10 swipe touches on 2 toys simultaneously. (Figure 2c)
4. Cover touch: Perform 10 cover touches on 2 toys simultaneously. (Figure 2d)
5. Still: Pauses between two interactions are recorded as still instances



Figure 7: Study setting for toy interaction

Training & Testing

We trained our 4-class interaction classifiers based on annotated data from one participant and tested the classifier for all other 10 participants. Classification results were reported in real time (and recorded for post processing), but were not made visible to the participants during testing. The script with annotated mistakes was then compared to data collected by the RFID reader.

Evaluation Results:

The system achieved an average of 93.7% (SD=1.0%) classification accuracy for 5 toys on 1600 instances across 4 classes collected from 10 participants, and a 2.8% (2400 instances) false alarm rate on the 3 still reference toys. Table 2 shows detailed classification results.

These results show that the system is capable of classifying multiple tagged objects even when simultaneous interaction events occur. During natural interactions with toys, participants would have to reach over "still toys" in order to pick up adjacent objects when prompted. This could cause a change in the RF signature of the "still toy" since the arm of the participant would partially blocked some of the RF signals from the reader, potentially causing a false positive. However, our training session included this type of interference and the results show that it did not cause a significant false positive rate.

We also tested the system in the same setting with only one tagged toy to get an upper bounds on performance. This accuracy is only slightly higher than the 5 tag scenario with an accuracy of 95.3% (SD=4.0%).

Classified as→	Still	Motion	Swipe	Cover
Still	94.5%	2.5%	3.0%	0.0%
Motion	2.3%	92.3%	5.5%	0.0%
Swipe	1.0%	5.5%	93.5%	0.0%
Cover	0.0%	4.0%	1.5%	94.5%

Table 2: Classification results for user toy interaction events. Five toys were interacted with in total and at any one time two toys were being interacted with.

Interaction Detection of Daily Objects for Activity Inferencing

Activity inferencing of daily tasks and events in the home and office environment has long been an important capability for ubiquitous computing and smart environment applications. Using human object interaction in the form of “move” events has been shown as a reliable method for inferring daily activities [11,13]. For testing purposes, a home living room environment was setup in our office space and objects were augmented with RFID tags as shown in Figure 8. In this scenario, IDSense shows that it can achieve results similar to previous activity inferencing projects based on RFID without the need for expensive custom hardware or for the users to wear an RFID reader.

Study Design Overview

In this study, we evaluate the performance of our system in classifying events necessary to support activity inferencing scenarios, based on human object interaction. Ten commonly used items in a living room/kitchen setting are retrofitted with tags including: 1) drinking glass, 2) milk container, 3) cereal box, 4) bowl, 5) glasses case, 6) book, 7) TV remote control, 8) vitamin box, 9) window cleaner, 10) toothbrush.

The items are placed on 3 separate tables/counter tops, which occupy approximately a 4m x 4m space. The RFID reader antenna is placed near the ceiling on a tall tripod and is pointed downwards towards the floor so that it covers the tables and counter.

Study procedures

We recruited 11 participants including 6 males and 5 females, with a mean age of 25.3 years. Each participant was given audio instructions to perform three randomly ordered sets of the following eight activities, resulting in 24 activity instances: 1) Drink milk 2) Make cereal 3) Wear glasses 4) Read book 5) Watch TV 6) Take vitamin 7) Clean window 8) Brush teeth. Each activity lasted approximately 20 seconds.

An instruction script was used as ground truth and user’s mistakes were manually annotated. Some activities were not



Figure 8: Activity inferencing study setting in a mock living room environment

exactly the same as in real life (i.e. participants were not required to consume food), but participants were asked to perform object interaction with tagged objects as they would do in a real-life setting.

Training & Testing

In this scenario, the goal is not to simply quantify IDSense’s ability to identify individual move and touch events which were demonstrated in the last two sections. Instead, the time series output of the IDSense motion event classification engine is sent to a second activity inferencing classifier. The goal of the activity inferencing engine is to identify higher order activities based on the series of lower level object move events.

We trained our lower level motion event classifier based on 50 annotated motion events, of 10 objects manipulated by one participant. For the activity classifier, we set our sliding classification window length to 20 seconds, which was advanced on a one-second interval. Object motion events are mapped into activities and used to identify the other 10 participants’ activities. Results are post-processed based on the ground truth script with annotated user mistakes.

Evaluation Results:

IDSense reported an average of 96.9% (SD=2.5%) precision and 95.8% (SD=4.3%) recall, for inferring eight activities given a total of 240 instances across 10 participants. Table 3 shows classification details on each activity. These promising results showcase the adaptability of our low-cost, commercially available RFID-based system in a home setting for activity detection and inferencing.

Product Interaction Tracking for Customer Interest Monitoring.

RFID systems have already shown potential for enhancing traditional retail stores while improving inventory management [8]. IDSense offers the potential to determine which items or merchandise displays are most appealing to consumers based on interaction events. Retailers can use such information to provide consumers with customized shopping experiences. One of the challenges in a retail environment is that the number of RFID tags visible by an individual reader may be significantly larger than that in a

Activity	Precision	Recall
Drink Milk	93.5%	96.7%
Make cereal	98.1%	86.7%
Wear glasses	96.8%	100.0%
Read book	100.0%	96.7%
Turn on/off TV	96.5%	93.3%
Take vitamin	100.0%	96.7%
Clean window	96.7%	96.7%
Brush teeth	93.7%	100.0%
Totals	96.9%	95.8%

Table 3 Activity detection results for eight common tasks in the home

home setting. As the population of tags increases, the per-tag read rate (or the sampling rate of the tag) can decrease. Thankfully there is not an inversely proportional correlation as the ISO-18000c standard which is based on the Slotted Aloha protocol is specially designed to quickly inventory very large population of tags. Additionally, with densely packed clothes on hangers, it is likely that the movement of one item may cause motion on nearby objects as well. On the other hand, since shoppers will spend more than a few seconds with an object that they are interested in, multiple move events can be registered in terms of higher frequency and duration to determine the items of interest.

Study Design Overview

In this study, 20 tagged clothes are displayed on two clothing racks (Figure 9). The RFID reader antenna is located on the ceiling pointed towards the racks. Each participant is asked to browse the selection of clothing and select items of interest. Since the RFID tags are densely packed together on hanging racks, small movements of one piece of clothing can move other shirts resulting in spurious move events. Therefore, we created a second classifier that monitored the data stream of move events into two shopping activities: “browsing” and “item of interest”. Browsing indicates which rack of clothing a shopper is looking through. The “item of interest” classifier detects that a single garment tag is being interacted over an extended period of time.

Study procedures

We recruited 11 participants including 7 males and 4 females, with a mean age of 24.2 years. Each participant was asked to browse through the racks of clothes and then choose one piece to try on. We recorded browsing and clothing choice ground truth manually. The clothing interaction classifier reported interaction events once per second. The action of browsing was determined by matching clusters of interaction events to a particular clothing rack. Items of interest are determined by monitoring interaction frequency. In a moving window of 30 seconds, if more than 10 interaction events are detected on one tag, the tagged clothes will be classified as an *item of interest*.



Figure 9: Retail shopping scenario study setting

Training & Testing

We trained our interaction classifier based on annotated data from one participant doing 50 instances of annotated motions on clothes. Since there are only motion and still events in this scenario, the classifier is binary, separating motion class from still class. Objects motive events and their frequency in a 30-second sliding window (which is advanced each second) was used to determine browsing and item of interest. Classification results were reported in real-time but were not made visible to the participants during testing.

Evaluation Results:

A total of 10 participants made 50 browsing events and 50 clothing choices. The system was able to detect 49 out of 50 “browsing” events with 4 false alarms and 48 out of 50 “item of interest” clothing choice events with 1 false alarm. Thus, yielding 92.5% precision, 98.0% recall on browsing detection and 98.0% precision and 96.0% recall on identifying clothes choices. These results show the potential to apply the IDSense system in the retail environment to identify consumer shopping habits and help provide better consumer experiences.

DISCUSSION & LIMITATIONS

There are many possible methods for implementing human object interaction detection systems and IDSense focuses on the paradigm of minimalistic instrumentation of objects with ultra-low-cost radio sensing capability in the form of RFID tags. While other approaches can provide streaming sensor data at higher data rates, this paper demonstrates that simple *move* and *touch* interaction detection can enable compelling ubiquitous sensing and human object interaction applications.

Additional Applications

Since the IDSense system is easy to train and deploy, we believe the potential application space is not limited to the three scenarios described in the previous sections. In fact, everyday objects provide a board space for IDSense to enable novel interaction detection applications. Other applications we have explored include seat occupancy detection and gross posture estimation. For example, tags on a seat surface can be utilized to detect occupancy by sensing cover touch. Tags on the seat back and armrest can be used to sense a leaning back posture and resting arm posture, enabling gross posture detection.

A second class of compelling application space is infrastructure monitoring, where RFID tags can infer the state of the environment. For example, RFID tags placed on fixed infrastructure, such as doors, floors, and windows, can be used for motion tracking for security scenarios. Furthermore, tags can be integrated into objects as sensors, where the motive or still states of mechanical methods can be inferred by monitoring RSSI and RF Phase features.

Human Interference

IDSense demonstrates good performance for multi-tag applications and the usage scenario should be designed to mitigate object-to-object and object-to-human interference.

However, due to the diversity of human behavior, unintended interactions can be recorded. For instance, when humans walk between the reader antenna and tagged objects (or block most of the RF paths to the target tag) it is possible to create false interaction events. Thus, choosing good antenna placement can improve system performance. We found that reader antennas placed in the ceiling provide a good balance between reader coverage area and reduction of human signal blocking events. Other approaches, such as multiple reader antennas and the use of fixed reference tags that help calibrate the system may provide additional improvements.

Training

One of the advantages of only classifying move and touch events is that the classifier can be generalized to multiple users, and the same classifier can be used for multiple objects. In our evaluation study, the training process required a single participant to record training data for no more than 10 minutes. When a single interaction classifier is applied to multiple usage scenarios in different environments, it can only achieve limited performance. So we chose to boost performance by having one participant retrain the classifier for each usage scenario. Since each of the three studies was done in very different locations, it is believed that the environment had some effect even though the features are based on signal differentials. In future work, we want to explore the machine learning space to develop a unified model to automatically adapt to dramatically different circumstances to reduce training and enable human object interaction in even more sophisticated environments.

CONCLUSIONS

This paper presents a robust method for enabling human object interaction by using minimalistic instrumentation in the form of passive UHF RFID tags. Even though RFID tags can only provide their ID when read by an RFID reader; we have shown that it is possible to infer tag motion and touch events by measuring changes in the RF communication channel between the tag and reader with 93% to 97% accuracy.

To demonstrate the versatility of our system, we investigated three application scenarios and conducted user studies to show the viability of our approach. First an interactive storytelling scenario was demonstrated where toys were tagged so that they could be augmented by digital media. Second, we showed for the first time that low-cost, commercially available RFID tags could be used to infer daily activities in a home setting without the need to wear a reader. Lastly, we showed that implementing this system in a retail scenario can enable the identification of real-time shopping behaviors. Ultimately, IDSense makes sensing human interaction with everyday objects easy and unobtrusive, by minimally augmenting objects with low-cost and long-lived RFID tags.

REFERENCES

1. Bao, L. and Intille, S.S. Activity Recognition from User-Annotated Acceleration Data. *Pervasive Computing*, (2004), 1–17.

2. Buettner, M., Prasad, R., Philipose, M., and Wetherall, D. Recognizing Daily Activities with RFID-Based Sensors. *UbiComp*, (2009), 51–60.
3. Chang, C.-C. and Lin, C.-J. Libsvm. *ACM Transactions on Intelligent Systems and Technology* 2, 3 (2011), 1–27.
4. Dimitriou, A.G., Bletsas, A., Polycarpou, A.C., and Sahalos, J.N. On Efficient UHF RFID Coverage inside a Room. *2010 Proceedings of the Fourth European Conference on Antennas and Propagation (EuCAP)*, (2010), 1-5
5. Fishkin, K.P., Jiang, B., Philipose, M., and Roy, S. I Sense a Disturbance in the Force: Unobtrusive Detection of Interactions with RFID-tagged Objects. *UbiComp*, (2004), 268–282.
6. Gaëtanne Haché, Edward D. Lemaire, N.B. Wearable mobility monitoring using a multimedia smartphone platform. *IEEE Transactions on Instrumentation and Measurement* 60, 9 (2011), 3153–3161.
7. Hodges, S., Thorne, A., Mallinson, H., and Floerkemeier, C. Assessing and optimizing the range of UHF RFID to enable real-world pervasive computing applications. *Pervasive Computing*, (2007), 280–297.
8. Meli, J. and Morenza-cinos, M. Enhancing the shopping experience through RFID in an actual retail store. *UbiComp*, (2013), 1029–1035.
9. Nikitin, P. V., Martinez, R., Ramamurthy, S., Leland, H., Spiess, G., and Rao, K.V.S. Phase based spatial identification of UHF RFID tags. *RFID 2010: International IEEE Conference on RFID*, (2010), 102–109.
10. Parlak, S. and Marsic, I. Detecting Object Motion Using Passive RFID: A Trauma Resuscitation Case Study. *IEEE Transactions on Instrumentation and Measurement* 62, 9 (2013), 2430–2437.
11. Philipose, M. "Large-scale human activity recognition using ultra-dense sensing. *The Bridge, National Academy of Engineering* 35, 4 (2005).
12. Ren, X. Egocentric recognition of handled objects: Benchmark and analysis. *2009 IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops*, (2009), 1–8.
13. Tapia, E.M., Intille, S.S., and Larson, K. Activity Recognition in the Home Using Simple and Ubiquitous Sensors. *Springer Berlin Heidelberg*, (2004), 158–175.
14. Ward, J. a, Lukowicz, P., Tröster, G., and Starner, T.E. Activity recognition of assembly tasks using body-worn microphones and accelerometers. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 28, 10 (2006), 1553–1567.
15. Weiser, M. The Computer for the Twenty-First Century. *Scientific American* 265, 3 (1991), 94–104.