Human-object Interaction Reasoning using RFID-enabled Smart Shelf

Joan Melià-Seguí and Rafael Pous
Department of Information and Communication Technologies
Universitat Pompeu Fabra
Tànger 122-140, 08018 Barcelona - Spain
{joan.melia,rafael.pous}@upf.edu

Abstract—Radio Frequency Identification (RFID)-enabled smart shelves are becoming common place in pervasive retail. These devices provide real-time information about the item's stock and location, but few efforts have been made to reliably detect human interaction with the items. We present a novel approach on real-time human-object interaction detection based on RFID using supervised machine learning techniques. By analyzing specific RFID features, we classified human interaction on a real smart shelf, achieving a performance over 84%. This work aims to provide the first method to model RFID information as a source of human activity recognition, with application to context-aware industrial infrastructure, smart environments and Internet of Things.

Index Terms—Internet of Things, RFID, Human-object Interaction Modeling, Context Awareness, Smart Environments, Data Analysis and Learning.

I. INTRODUCTION

Context-aware technologies are becoming ubiquitous in industry. Retailers, for instance, are applying these technologies in what is known as pervasive retail [1] in smart environments. By means of ubiquitous computing and context-awareness, a company can improve its operations and customer’s experiences. Real-time inventorying is one of the possible applications of pervasive retail, avoiding manual errors, and reducing the cost and processing time invested in retail operations. Sensors-enabled smart shelves are an example of the context-awareness of technology in these environments.

Radio Frequency Identification (RFID) is being widely used as a sensing platform for item identification, being a common Internet of Things (IoT) technology in retail [2]. Specifically, the Ultra High Frequency (UHF) RFID Electronic Product Code Class 1 Generation 2 [3] (EPC Gen2, for short) is the de facto standard for RFID in retail [4][5]. Smart shelves can be equipped with RFID readers and multiple antennas, identifying and locating RFID tags attached to the objects.

Smart shelves are used for items’ detection and location. In this paper we focus on a further goal: detecting human interaction with the items. The connection of human activity in brick-and-mortar stores through the IoT would allow a truly context-aware smart environment [6], for instance by issuing real-time activity-based recommendations to the user based on interacted products [7]. Adapting advertising to the current context, thanks to a better understanding of human activity in the environment, is key to improve end-users experience and operations efficiency [8].

Current approach to detect interaction in a smart shelf using RFID is to simply inventory the item. If the item is inventoried it is assumed to remain in the shelf, on the contrary, if the item disappears from the inventory it is assumed to be removed from the shelf. More complex approaches propose to use additional sensors or computer vision, but its cost makes it impossible to deploy in a real scenario.

Our goal is to allow retailers direct activity observation through an indirect approach, extracting meaning and value from the IoT. In this work we propose to use high and low-level RFID features to improve human object interaction reasoning and contextual intelligence in real scenarios, by using passive UHF RFID technology and statistical analysis.

Our research contributions can be summarized as follows:

- A study of high and low-level features from standard UHF RFID equipment to identify interaction
- A novel probabilistic model for human-object interaction for context-aware smart environments
- A real-scenario evaluation based on the proposed model, compared with current baselines

The remainder of the paper is organized as follows: Section II introduces the related work. Section III overviews RFID and smart shelf parameters. Human-object interaction modeling is described in Section IV. Section V shows the implementation and experimental results, and finally conclusion and future work are described in Section VI.

II. RELATED WORK

RFID and statistical analysis have been used before in the literature to extract information in context-aware scenarios. Some works only use RFID as identification platform, while activity information is extracted from battery enabled sensors or computer vision. In [9] the authors propose an approach to activity recognition based on detecting and analyzing a sequence of objects manipulated by the user at a kitchen or office. The authors use probabilistic methods, merging computer vision and RFID, to jointly infer the most likely activity and object labels. In [10] the authors integrate wireless sensor networks and RFID for human-object interaction detection. Activity is extracted from battery-enabled accelerometers using stream processing techniques for real-time interaction.
analysis, with preprocessing techniques to generate meaningful higher level events. In this work RFID is only used for identification purposes. A middleware to integrate sensors and identification data has been recently proposed in [11].

RFID data as the main source for information extraction is also present in the literature. In [12] and [13] the false positive reads problem is explored, statistically analyzing whether reading patterns from different objects and antennas comply with predefined retail operations. Huiting et al. exploit phase measurements of EPC Gen2 tags for localization and movement detection [14]. Finally, Keller et al. [15] propose a machine learning based approach that makes use of the low-level RFID reader data to detect false positive reads in a scenario with static labeled objects such as a factory or a distribution center. Finally, Kriara et al. propose in [16] a method to detect gesture recognition based on HF RFID.

Opposite to [9] and [10], our approach to analyze human-object interaction is only based on RFID measurements, which simplifies and reduces its cost in a real implementation. Besides the actual high-level inventory information, we use multiple low-level RFID features such as received signal strength indicator (RSSI) and RF Phase (RFP) to improve resolution on the human-object interaction. Our approach differs to single features utilization as in [12] and [15]. However, we follow a similar approach as [15] to select relevant features from which to extract meaningful activity information.

III. RFID-ENABLED SMART SHELF

Next, the main characteristics of UHF EPC Gen2 RFID performance are described. The integration of this technology in commercial shelves, using time-multiplexed antennas, allows to further detect interaction of humans with objects.

A. Items Identification Using UHF EPC Gen2 RFID

EPC Gen2 is a low-cost passive RFID technology performing on the UHF band. It is mainly composed by a reader and antenna (also named interrogator), one to several tags (also named electronic labels) and an information systems back-end. It works following the Interrogator-Talks-First basis, where the reader provides energy to the tags through RF waves, and also sends interrogation commands to which the tags answer. The identification process can be divided in three phases. On a first stage, the reader looks for how many tags are in its communication range based on an Aloha-like Medium Access Control (MAC) protocol [3]. This stage is called the Select operation. On a second stage the reader individually identifies all tags in the field, what is known as the Inventory operation. Finally, if the reader wants to modify the tag’s memory or access the tag’s reserved memory, the Access operation is performed.

The time to inventory the items in UHF RFID depends on the number of tagged items, the environmental RF properties and the reader configuration. The larger the number of tags, the slower the Select stage. However, EPC Gen2 is designed to identify large populations of tags, being able to identify hundreds of tags per second [3]. The RF environment affects

the EPC Gen2 performance, since the signal codification and collisions can delay the communication. Finally, depending on the number of tags to be identified, the RFID reader read time can be adjusted to speed up the inventorying process, being the typical read time in the range of a few hundreds of milliseconds for common real scenarios.

The above mentioned parameters affect the inventorying time, for a single-antenna reader. Next, we discuss the implications of using multiple antennas.

B. Antenna Multiplexation

The aim of a smart shelf is to track the objects placed on it, to obtain real-time information about a store’s stock. Figure 1(a) shows a typical smart shelf implementation, measuring about 2 m height and 1 m width. There exists two main options to cover with RFID signal all the smart shelf: using a single reader/antenna covering the whole smart shelf, or using more antennas placed along all of the smart shelf surface.

The first option, although simpler in terms of implementation, is not able to retrieve relevant information for the store such as the location of objects. Implementing different single-antenna readers in a smart shelf quickly increases the cost, besides provoking RF interferences because of the dense reader environment (RF channels management in a reduced area) [3]. The current approach to solve this issue is antenna-multiplexing using a single reader.

With antenna multiplexing, the RF channel is time multiplexed along different antennas, that is, each antenna from a single reader performs an identification process sequentially. Antenna multiplexing provides spatial resolution, also lowering the implementation cost. Nevertheless, antenna multiplexing adds some temporal uncertainty due to the time it takes to cycle through all of the antennas. Since only one antenna can be active simultaneously for t seconds, the time gap without signal for an antenna increases to t(n − 1) seconds, where n is the number of multiplexed antennas. This represents a drawback for the detection of human-object interaction on a smart shelf. Since most of the time the antennas are inactive, the system can miss interactions on the smart shelf. However, dividing the identification space reduces the number of objects to inventory per antenna, allowing short read times, and thus,
fast multiplexing. Next, we propose a novel method to infer human-object interactions on a RFID smart shelf, overcoming the antenna-multiplexing issue.

IV. HUMAN-OBJECT INTERACTION USING RFID

We define an interaction as a customer taking an item off the shelf, and eventually returning the item to the shelf. In this section we describe the RFID high and low-level features analysis to model human-object interaction on a smart shelf, using standard UHF EPC Gen2 RFID technology.

A. High and Low-level RFID Features

High-level RFID information typically includes the 96-bit identification code, a time stamp, a read count (the number of times a tag has been read in a reading cycle), and an identifier of which antenna detected the tag. The 96-bit identification code is implicit to the inventoried object, and the time stamp is implicit to each measured sample. In the smart shelf scenario the read time is adjusted to the minimum allowing the identification of all objects (cf. Section III), giving the read count a typical value of one. The next high-level feature we consider is what antenna detects each tag. Since the antennas are slightly directive (∼100° beamwidth), this feature brings an approximate location of the object. Due to the antennas’ radiation pattern, a tag close to the smart shelf (i.e. close to the antennas) will be inventoried typically by one antenna. On the contrary, if the tag moves away from the smart shelf it may be detected by more antennas (up to n).

Besides these high-level RFID parameters, most commercial readers offer the capability of sensing low-level RFID information, such as the RSSI or RFP. The RSSI is modeled by the two-way radar equation for a monostatic transmitter. The reader provides an approximate RFP value which is a combination of the roundtrip distance between the reader’s antenna and the tag, plus the phase rotation introduced in transmission, reception, and at the tag itself.

Figure 2 depicts the RSSI and RFP information from a tagged object, placed on a four-antenna smart shelf, for about ten minutes. During that time, a human-object interaction is produced. The interaction event in this example is denoted by two vertical dashed lines. The tag backscatters the reader’s signal, which is received by the reader at around -40 dBm if the object is placed on the shelf (thus, close to an antenna), and drops under -60 dBm if the object is on customer’s hand. During interaction, the RFP rotates up to 60°.

The smart shelf will inventory the RFID tag population also obtaining the low-level information aforementioned. Both RSSI and RFP are indicators of the relative distance of the object to the antennas. Similarly to [15], besides considering the values as measured by the reader in each sample (α), we also use the absolute difference |Δα| values of each sample with its predecessor from the same antenna, to obtain a sense of the object’s movement. For instance, an object can be slightly moved reporting a similar RSSI, but a high RFP difference. Hence, both low-level measurements add information to the system.

B. Interaction Modeling and Reasoning

A simplistic baseline method to detect interaction using RFID is to track the presence/absence of identification for each specific tag on the shelves. However, this method presents a main drawback that makes it unreliable despite of its simplicity: a tag can still be inventoried even on customer’s hands, hence, missing the interaction.

In the proposed scenario, for each inventory round r, n antennas $J = \{j_1 : j_n\}$ are sequentially activated t seconds, eventually returning samples $S$ with timestamp $p$ from the tag population $K = \{k_1 : k_m\}$. We propose to detect the human-object interaction using a combination of low-level and high-level RFID events. For both RSSI and RFP, we consider the raw value ($\alpha$) as received from the reader, and the absolute difference with its predecessor from the same antenna ($|\Delta \alpha_r|_{r-1}$). The number of antennas (#Ant) detecting the same tag is the high-level RFID event considered in the analysis. Since the inventory process uses sequential-antenna measurements, it is not possible to simultaneously detect the tags from different antennas. Hence, we should consider the number of different antennas identifying an object within some period to the current sample. Equation 1 defines the samples format, where $\{x_1 : x_5\}$ are the five high and low-level RFID events obtained from the reader as features.

\[
S^k_{r,j,p} = \{x_1 : x_5\}
\]

(1)

Using supervised machine learning techniques, the above features are obtained to model a Bayesian Network. First, we classify the feature values into interacted or static objects using Information Gain analysis [17]. It can be interpreted as a measure of how well the two tag types can be separated using a specific value as threshold [15]. Second, we apply this supervised learning to a Bayesian Noisy-OR Network [18] to infer interaction. Equation 2 details the Noisy-OR model where $y$ denotes interaction, $X'_i \in X$ is a subset of the z features described in Equation 1 better classifying the interaction, $\lambda_0$ is the leak probability [18], and $\lambda_i$ are the interaction probabilities associated to each feature $X_i'$.

\[
P(y|X'_i) = 1 - \left(1 - \lambda_0 \right) \prod_{i=1}^{z} (1 - \lambda_i)
\]

(2)

Each of these features become a parent in the noisy-OR model. We select the Noisy-OR network due to the independence of causal influence of the selected parameters generating an interaction, and the implicit noise addition in the passive RFID measurements, besides its simplicity for large-scale deployments. Thus, the combined influence of the smart shelf measurements return an interaction probability $P(y|X'_i)$ conditioned by the aforementioned features.

V. EXPERIMENTAL SETUP AND RESULTS

Next, we detail the experimental setup used to validate the theoretic model described in Section IV-B, and the empirical results obtained from our experiments. The RFID devices used
in this research are the Advantenna-P11 and the AdvanReader-100 [19] (using the M6e 4-port UHF RFID module [20]), compatible with EPC Gen2 [3] with ETSI regulations (European region [21]). The reader transmits 31.5 dBm power, with a receiving sensibility of -80 dBm. RFP is measured modulo 180 (due to modulation constraints). The RFID tags used in this research are the AK UHF tags [22], depicted in Figure 1(b), although any commercial tag could be used.

A. Experimental Setup

To validate our human-object interaction model we deployed a real Smart Shelf with labeled objects (cf. Figure 1), in a laboratory emulating a real scenario with a single reader to avoid the dense reader environment [3]. One of the shelves was inventoried by a row of four antennas (n = 4), connected to the RFID commercial equipment described above (Section V). The RFID reader used in our experiment can inventory up to 400 tags per second [19]. For the sake of simplicity in the experiment, five tagged books \( k_1, k_2, \ldots, k_5 \) were placed on the shelf. We requested volunteers to take the RFID labeled books from the shelf as they would do in a book store (i.e. take a look at the book). The interaction time for each book was recorded to be used as ground truth data to enable supervised learning. An example of these interactions is shown in Figure 2.

The RFID reader used in our experiment can reliably inventory this tag population in \( t = 200 \text{ ms} \), plus approximately 50 ms for processing antenna switching. Since four antennas were used, an inventory round \( r \) was completed in approximately one second, being a reasonable time to measure an interaction. We used the same time period to accumulate the number of antennas \#Ant\( r , j = 1:3 \) measuring a tag \( k \). A summary of the experimental setup is described in Table I.

The reader was configured with one sequential multiplexing level, and monostatic configuration (each single antenna

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Reader</td>
<td>AdvanReader-100 [19] (with M6e module [20])</td>
</tr>
<tr>
<td>Tx Power</td>
<td>31.5 dBm</td>
</tr>
<tr>
<td>Sensitivity</td>
<td>-80 dBm</td>
</tr>
<tr>
<td>Inventory</td>
<td>Sequential</td>
</tr>
<tr>
<td>Q value</td>
<td>Dynamic</td>
</tr>
<tr>
<td>Read time</td>
<td>200 ms + 50 ms processing</td>
</tr>
</tbody>
</table>

| Antennas  | 4 Advantenna-P11 [19]                      |
| Directivity | ~100 degree                                |
| Gain       | 3.2 dBi                                    |
transmits the RFID commands and receives the responses). Phase ambiguity due to channel selection in ETSI regulation [21] (≈ T°) was ignored since its effect was not significant. However, it may be considered for other regions [14]. Ten interactions were performed at regular intervals on the objects. A dataset containing about 20,000 samples (spanning 20 minutes of inventorying time) were used for training and evaluation using 10-fold cross-validation.

B. Empirical Results

Table II summarizes the measured RFID features from the supervised learning classification analysis based on Information Gain. Each evaluated feature returns the samples measurement range, the optimal classification threshold as defined by the maximum Information Gain value, the maximum Information Gain value, and its individual classification accuracy (λᵢ) in either interaction or non-interaction samples. Results return better performance from the RSSI feature (x₁) and the absolute RFP difference (x₇) within the low-level features tested. The high-level number of antennas (x₅) also returns good performance. Figure 3 depicts the distribution and Information Gain analysis of the three best features being used to model the Noisy-OR network.

To evaluate the performance metrics of our proposed method, different analysis methods have been compared: presence/absence (p/a) inventory detection baseline, single-feature (e.g. RSSI monitoring) baselines, pair-wise models, and the proposed Noisy-OR network model (cf. Section IV-B). All baselines use one second as interaction time, for better feature (e.g. RSSI monitoring) baselines, pair-wise models, presence/absence (p/a) inventory detection baseline, single-feature (e.g. RSSI monitoring) baselines, pair-wise models, and the proposed Noisy-OR network model (cf. Section IV-B). All baselines use one second as interaction time, for better comparison with the proposed model.

The evaluation metrics measure how well interactions are classified from within all measured samples. True Positives are assigned when correctly detecting an interacted sample, and True Negatives are assigned to actual static samples (i.e. objects not being interacted). False Positives and False Negatives happen when incorrectly classifying either labels. The above measurements generate the following evaluation metrics:

- **Precision**: Measures the fraction of actual interactions with respect to all measurements detected as interactions.
- **Recall**: Measures the fraction of actual interactions with respect to all actual interactions (whether these samples were correctly classified or not). From all metrics this is the most relevant in a context-aware scenario since it measures how reliable is the interaction detection.
- **F-Score**: Measures the weighted harmonic mean of Precision and Recall, returning a measure of the correctness of both metrics together, being also a good evaluation metric for the correctness of the system.
- **Accuracy**: Measures the fraction of correctly classified samples out of the total amount of samples. In our scenario the number of True Negatives is larger than the rest of samples (interaction time in retail is usually smaller than static time), thus, we scale this measure in accordance with the number of samples.

Table III shows the performance results from the evaluated baselines and methods. We can observe that the presence/absence baseline performance is very poor. Single-feature baselines perform better than presence/absence, with special mention to RSSI raw value (x₁). The proposed three-features model (x₁, x₄, x₅) provides the best overall classification metrics (F-Score = 84.1% and Accuracy = 81.6%), due to its correctness on detecting actual interactions within the samples (Recall). Actually, the proposed three-features model (x₁, x₄, x₅) improves the detection of actual interactions by 7% over measuring RSSI alone. Correctly classifying interactions within all detected interactions (Precision) is better modeled by ignoring RSSI from the model (pair-wise x₄, x₅). Nevertheless, human-object interaction modeling is better described by Recall metrics (97.1% in our experiments), since reliably detecting an actual interaction in a real retail scenario can lead to much more accurate decisions in smart environments [8], like placing a targeted advertisement on time, or suggesting a recommendation through the users’ smartphone.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Range</th>
<th>Threshold</th>
<th>max(IG)</th>
<th>λᵢ</th>
</tr>
</thead>
<tbody>
<tr>
<td>x₁: RSSI</td>
<td>-76:-36</td>
<td>-54</td>
<td>0.26</td>
<td>0.78</td>
</tr>
<tr>
<td>x₂: ΔRSSI</td>
<td>0.27</td>
<td>9</td>
<td>0.03</td>
<td>0.43</td>
</tr>
<tr>
<td>x₃: RFP</td>
<td>0.177</td>
<td>168</td>
<td>0.04</td>
<td>0.50</td>
</tr>
<tr>
<td>x₄: ΔRFP</td>
<td>0.84</td>
<td>18</td>
<td>0.14</td>
<td>0.64</td>
</tr>
<tr>
<td>x₅: #Ant</td>
<td>1:4</td>
<td>3</td>
<td>0.13</td>
<td>0.74</td>
</tr>
</tbody>
</table>
TABLE III
THE PROPOSED MODEL PROVIDES THE BEST OVERALL CLASSIFICATION FOR THE HUMAN-OBJECT INTERACTIONS

<table>
<thead>
<tr>
<th>Method</th>
<th>Precision</th>
<th>Recall</th>
<th>F-Score</th>
<th>Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>Baseline (p/a)</td>
<td>0.032</td>
<td>0.060</td>
<td>0.042</td>
<td>0.319</td>
</tr>
<tr>
<td>Baseline (_x_1)</td>
<td>0.740</td>
<td>0.901</td>
<td>0.813</td>
<td>0.792</td>
</tr>
<tr>
<td>Baseline (_x_4)</td>
<td>0.837</td>
<td>0.242</td>
<td>0.375</td>
<td>0.597</td>
</tr>
<tr>
<td>Baseline (_x_5)</td>
<td>0.851</td>
<td>0.565</td>
<td>0.683</td>
<td>0.737</td>
</tr>
<tr>
<td>Pair-wise (<em>x</em>{1,4})</td>
<td>0.738</td>
<td>0.925</td>
<td>0.821</td>
<td>0.798</td>
</tr>
<tr>
<td>Pair-wise (<em>x</em>{4,5})</td>
<td>0.857</td>
<td>0.650</td>
<td>0.739</td>
<td>0.771</td>
</tr>
<tr>
<td>Pair-wise (<em>x</em>{1,5})</td>
<td>0.723</td>
<td>0.972</td>
<td>0.829</td>
<td>0.799</td>
</tr>
</tbody>
</table>

Our model \(_x_{1,4,5}\) 0.742 0.972 0.841 0.817

VI. CONCLUSION AND FUTURE WORK
Context-aware Internet of Things (IoT) is opening new ways to enhance operations and user satisfaction in pervasive and smart environments. Radio Frequency Identification (RFID)-enabled shelves are an example of these pervasive technologies in industrial retail. We presented a novel method to detect real-time human-object interaction, modeled by a combination of three RFID features. By means of machine learning techniques and following a probabilistic approach, we defined for the first time a model for classifying human-object interaction by only using standard EPC Gen2 RFID technology with time-multiplexed antennas. Empirical results of our human-object interaction detection model achieve 84.1% performance in a real environment.

The aim of our proposed method is to serve as the basis for further context-aware applications to connect human activity in physical smart spaces to online applications. Our experience with this work has enlightened us to a number of ideas:
- To improve the model by adding new features or refining the existing ones. For instance, assigning different weights to samples from each antenna, which can improve the accuracy of interaction.
- To test or adapt the model features with a larger set of antennas, also considering phase ambiguity due to channel selection or frequency-hopping spread spectrum.
- To experiment with multi-platform IoT. Contextual intelligence applications like coupons or recommendations delivery in a smartphone application, driven by human-object interaction detection in a brick-and-mortar smart environment.

ACKNOWLEDGMENT
The authors would like to thank the anonymous reviewers for their valuable comments and suggestions. The authors also thank Ausias Vives and Marc Morenza-Cinos for their help an support in this project. This work was partly funded by the Spanish Government through project TIN2012-34965 PIGALL, Obra Social "la Caixa"-ACUP through project 2011ACUP00261, and Keonn Technologies S.L.

REFERENCES