University of Zurich[UZH]

Bruno Rodrigues, Ile Cepilov, Muriel Franco, Eder J. Scheid,
Christian Killer, Jan von der Assen, Burkhard Stiller

CCount: Correlating RFID and Camera Data for High Precision Indoor Tracking

TECHNICAL REPORT – No. IFI-2022.02

January 2022

University of Zurich
Department of Informatics (IFI)
Binzmühlestrasse 14, CH-8050 Zürich, Switzerland

ifi

CCount: Correlating RFID and Camera Data for High Precision Indoor Tracking

Bruno Rodrigues, Ile Cepilov, Muriel Franco, Eder J. Scheid,
Christian Killer, Jan von der Assen, Burkhard Stiller
Communication Systems Group CSG, Department of Informatics IfI, University of Zürich UZH
Binzmühlestrasse 14, CH—8050 Zürich, Switzerland
E-mail: [rodrigues¦franco¦scheid¦killer¦vonderassen¦stiller]@ifi.uzh.ch, ile.cepilov@uzh.ch

***Abstract*—During acute health crises, such as the COVID-19 pandemic, social distancing and safe occupancy rates are paramount. Thus, the analysis of movement, capacity, and engagement is of critical importance in the context of strategic planning of live marketing initiatives for retail spaces, trade shows, and promotional events.**

The approach combines different sources of data, thus, it becomes possible to increase the tracking precision. Hence, this paper proposes Cloud Counter (CCount), a solution that expands such analysis by integrating existing passive sensing approaches from 3D cameras with input of RFID (Radio Frequency Identification) tags. In an experimental proof-of-concept, people walked in a controlled environment to show the effectiveness of this method in real scenarios. Results show that the solution designed and implemented is capable of identifying the number of unique visitors at any given time, since more than 90% of RFID tags are captured, with a correlation rate around 70-80% between 3D cameras and RFID readings.

***Index Terms*—Indoor Tracking, RFID Tag Data, 3D Camera Data, Correlation**

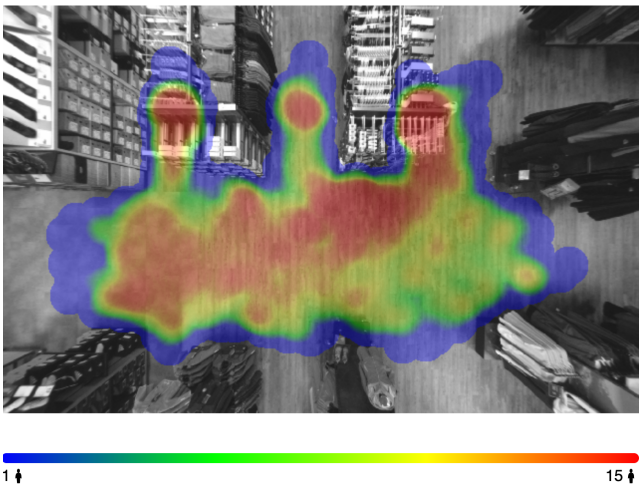

Fig. 1: Data Collected from the Evaluation Scenario with a 3D Camera

I. Introduction

Analyzing consumer behavior is one of the most important methods of identifying positive and negative features in products and services. To improve their services, a company needs to identify, for example, insights on the movement, capacity, and engagement of their customers. Such tasks are essential to planning proactive marketing initiatives in retail stores, trade shows, and promotional activities [18]. Also, such insights are key during an acute health crises, such as COVID-19, where social distancing and safe occupancy rates are paramount.

Unlike in e-commerce, implementing customer journey analytics is not straightforward in physical retail settings. It is essential to avoid tracking tools and methods that require consumers to interact with a product or service actively. Once their identity (*e.g.*, documents or face) could be profiled, their privacy would not be preserved [10]. Similarly, the accuracy (or the lack thereof) to estimate customers' positions is relevant in indoor tracking. For instance, measurements based on a single source of tracking may not provide enough detailed information about the device's position [18]. Besides, mobile devices typically employ Media Access Control (MAC) address randomization [17] that periodically generate a new MAC address on the software level, rendering the tracking ineffective in WiFi-based protocols (*i.e.*, based on IEEE 802.11).

In this regard, combining different tracking approaches, such as cameras [25], Bluetooth trackers [18], RFID (Radio-Frequency IDentification) readers [11], and data from other sources is available [19], do provide an interesting alternative to increase indoor tracking precision by correlating their measurements. There are non-invasive ways of visualizing an audience's mobility. For example, tracking the flow of people over a defined area, such as a public transport system [16] or a university campus [14], [21], [23], could provide essential insights into the capacity of transport systems and the flow of people during specific periods (*e.g.*, rush hours). Such information can be used for crowd control, city planners [14], emergency services, lowering the energy consumption of buildings [21], and also for business owners who gather marketing information used to tailor an advertisement to future customers [2].

Nonetheless, tracking approaches based on wireless signals (active or passive) typically present low accuracy due to various effects, such as multi-path fading, path loss, and MAC address randomization [15]. In the current scenario where data science and machine learning techniques gain

relevance in several network management aspects, the combination and correlation of multiple tracking sources simultaneously allow for increasing tracking accuracy and minimizing single-source inefficiencies.

Thus, this paper proposes Cloud Counter (CCount), an approach to collect, store, process, and match data from 3D cameras by expanding such tracking with data from RFID readers. In this regard, cameras are based on the Xovis PC2 [25], which offers four privacy protection levels (detailed in Section II) being compatible with General Data Protection Regulation (GDPR) [22] standards enforced in Europe (*cf.* Figure 1). While 3D cameras provide absolute positioning of tracked individuals, they have a limited field of view that restricts their use in large events. By combining their use with RFID readers [11] and tags, it is possible to expand the tracking area at a relatively low cost (than using more 3D cameras), while the accuracy can be increased by correlating the data on the backend in near real-time.

Evaluation results in a real-world controlled setting presented good (*i.e.*, high accuracy) and bad (*i.e.*, low accuracy) scenarios. On one hand, the CCount's prototype does uniquely identify individuals with 92.9% accuracy, identifying 13 out of 14 tags. Although the RFID read could not always provide accurate data, 3D cameras regularly captured the position of tags, resulting in a data correlation around 70-80% between the two tracking sources. On the other hand, the tracking of tag paths with more than ten tags in a high-density scenario (*e.g.*, many people walking close to each other) became difficult with respect to the tag's identification, yielding a 28% accuracy only. The major contribution of CCount is the modular design of the correlation approach that can incorporate different tracking sources, such as IEEE 802.11b or 802.11g (2.4 GHz) or 802.11a (5 GHz) sources.

The remainder of this paper is organized as follows. Section II overviews fundamentals. While Section III describes the design and implementation, Section IV details the evaluation. Section V summarizes final considerations.

II. Background and Related Work

While a comprehensive overview on indoor tracking is surveyed for instance in [15], [3], [8], fundamental concepts are essential for RFID and camera data correlations, offering privacy and similarity algorithms.

A. RFID Technology

RFID (Radio Frequency Identification) refers to a technology, in which digital data is encoded in RFID tags and captured by one or multiple readers through radio waves. RFID belongs to the group of auto-ID technologies and it enhances them by allowing tags to be read without a line-of-sight. Depending on the type of RFID the reading range is up to or slightly above 20 m [6]. Since data acquisition can be performed without any human intervention, RFID tags' operation is efficient. Due to no line-of-sight requirements, the placement of tags on objects can be performed with fewer limitations, thus, improving the degree of flexibility. In general, RFID tags are categorized as follows [12]:

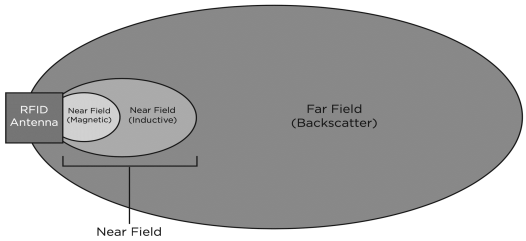

Fig. 2: RFID Far-Field and Near-Field

- **Active Tags** receive energy from their power supply in order to operate, they can autonomously transmit data to the reader, and they cover greater distances than passive tags.
- **Passive Tags** do not operate on an energy source, but receive power from the signal originating from the reader. They consist of a chip (with a unique identifier and memory), an antenna, and a support or container. Once a reader passes the tag, the respective radio frequency activates the microchip inside the tag to generate energy needed to operate.
- **Semi-passive/Active Tags** are equipped with an energy source, which is used to power the microchip or even another device, such as sensors, but this does not power the transmitter. To be able to transmit information, the tag still must be within range of the reader/antenna.

Passive tags can be divided into *Near-Field RFID* and *Far-Field RFID* types, depending on the frequency band used to communicate and the Electromagnetic (EM) spectrum. (*cf.* Figure 2). In the near-field region, the interaction between components is dominated by the magnetic field generated by the antenna, which induces an electric current in the tag by inductive coupling and allows the chip to be activated. Tags of this type are part of the Low Frequency (LF) and High Frequency (HF) classes. In contrast, in the far-field region type the interaction of components is dominated by the EM field created by the antenna. The RFID tag resonates with the frequency of the EM field and the current generated activates the chip. Tags of this type are part of the Ultra High Frequency (UHF) class.

B. 3D Camera

In the context of the analysis of movements, capacity, and engagement of people, a Xovis PC2 Camera [25] is utilized. Additionally, Xovis 3D cameras are equipped with a 3D sensor and two wide-angle lenses, which perceive the scene from different perspectives, therefore, achieving a precise depth map or 3D image of the entire scene. Both hardware and software components of this technology have to be compliant with the EU General Data Protection Regulation (GDPR) [22] and additional data protection laws,

such as Federal Act on Data Protection (FADP) [5], which are applicable in countries of the European Union. Thus, the camera in use offers four levels of privacy protection: While level 0 of privacy has no restrictions, only a still image is shown in level 1 with insights on a person's path. Level 2 disables the video stream's functionality, allowing only for path tracking of visitors, without revealing their identities. In level 3, no tracked person paths' are visible anymore which makes it compelling for use in places where strict privacy protections regulations apply.

C. Similarity Algorithms

The correlation of multiple data sources is a fundamental aspect, which is based on the position reported by different input vectors, including their respective timestamps. Correlating tracking sources based on known similarity algorithms are summarized in Table I.

TABLE I: Overview of Similarity Algorithms, based on [7]

Algorithm	Short Description	Equation
Euclidean	Distance between 2 points in a multidimensional space. Takes the square root of the sum of the squared pair-wise distances of every dimension	$\sqrt{\sum_{i=1}^{n}(x_i - y_i)^2}$ (1)
Manhattan	Similar to the Euclidean distance but the distance is calculated based on grid lines, taking the shortest diagonal path	$\sum_{i=1}^{n}\lvert x_i - y_i\rvert$ (2)
Cosine	Measures the orientation of two sample vectors irrespective to their magnitude. Calculated by the dot product of two numeric vectors	$\frac{x \cdot y}{\lvert\lvert x\rvert\rvert \cdot \lvert\lvert y\rvert\rvert}$ (3)
Pearson	Obtained via a Least-Squares fit and a value of 1 represents a positive relation, -1 a negative relation, and 0 indicates the absence of relation	$\frac{\sum_{i=1}^{n}(x_i - \overline{x})(y_i - \overline{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \overline{x})^2 (y_i - \overline{y})^2}}$ (4)

Equation 1 defines the Euclidean distance, which is calculated by taking the square root of the sum of the squared pair-wise distances of every dimension. Distances (x, y) reported by different tracking sources within a rolling time window (based on the timestamp of each source) are compared in a simple Cartesian plane. The smaller the distance, the larger the correlation is. Equation 2 is similar to the Euclidean distance, but calculates the distance between two vectors by considering the diagonal path ("beeline") between two points, *i.e.*, considering a grid line.

As with a slightly different approach, the cosine similarity measures the orientation of two n-dimensional sample vectors irrespective of their magnitude. It is calculated by the dot product of two numeric vectors, and it is normalized by the product of the vector lengths such that output values close to 1 indicate high similarity. The Pearson correlation coefficient (*cf.* 4) is the most widely used measure for linear relationships between two standard distributed variables and, thus, often just called "correlation coefficient". Usually, the Pearson coefficient is obtained via a least-squares fit, and a value of 1 represents a perfect positive relationship, -1 defines a perfect negative relationship, and 0 indicates the absence of a relationship between these variables.

III. CCount's Design

The application scenario considers 3D cameras and RFID readers mounted on the ceiling. Individuals within an exhibition wear or hold a low-cost UHF RFID tag in the form of a badge. As individuals walk within the view of the 3D camera and RFID reader, RFID tags are continuously read and the 3D Camera tracks the participant's position. In such a scenario, the CCount must correctly match data originating from 3D cameras with the data collected by RFID readers to uniquely count people in a scene (*cf.* Figure 3).

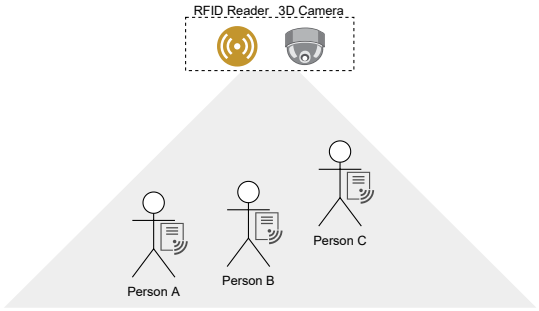

Fig. 3: Illustration of the Application Scenario

The challenge is identifying RFID tags belonging to IDs assigned by the 3D camera and is, thus, associated with an individual. If a person who left the scene returns later, CCount has to recognize and prevent it from being counted twice. To accomplish this, a method for correlating the RFID ID of the 3D cameras with the Electronic Product Code (EPC) of the RFID tags is required. For that, a frontend is developed to handle the interaction with the user, encompassing the selection of the desired camera, the associated event date, and a graphical representation of the dashboard associated with the selected event. The data is then requested from the backend, which is hosted on Amazon Web Services (AWS) and based on lambda functions (*cf.* Figure 4). It is built based on a serverless architecture built upon microservices to scale components while maintaining extensibility dynamically.

- The **API Gateway** collects and aggregates data coming from 3D cameras and RFID readers in time windows.
- **Data Transformation**: Data received at the API is typically encoded in different standards (*e.g.*, Cameras use Protobuf and RFID readers use JavaScript Object Notation, JSON), and in this sense, the data transformation standardizes them for database storage.
- The **Data Correlation** retrieves data in pre-defined time windows from the database and uses the Pearson correlation to determine, whether there exists statistical evidence for a linear relationship among variable pairs, such as coordinates.
- **Process Metrics and KPIs (Key Performance Indicators)** handle the localization and filtering of data and minimize errors by removing outliers. Once localiza-

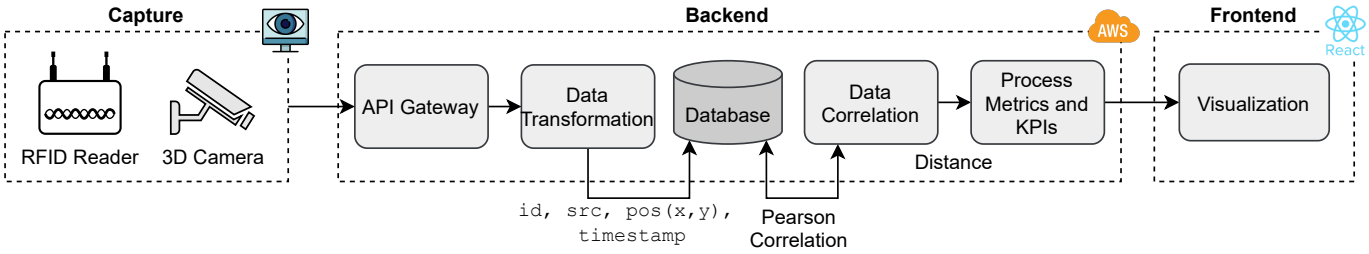

Fig. 4: CCount Workflow

tion data is processed, metrics, such as the unique count of individuals, dwell times, or visit duration, are calculated.

- The **Visualization** provides a graphical frontend, where it is possible to visualize metrics and KPIs in near real-time, and a REST API, where data can be retrieved.

A. Cameras Implementation Details

A camera sends data to the API Gateway, which forwards to the Data transformation module. This module implements a lambda function that comprises a list of objects, each containing the following attributes: *timestamp*, *coordinates*, and *numberOfPeople*. The next step involves the preparation of the retrieved coordinates for the actual heatmap creation process. Figure 8 shows an example of a heatmap during the testing stage in the laboratory cafeteria (3D Camera deployed in the ceiling).

Heatmaps were dynamically generated based on the predefined time-windows. The idea is to iterate over elements in the coordinates array and uniquely count the occurrences of every *(x,y)* pair with the timestamp laying within the slider values. This way, it is possible to ensure that every time the slider is adjusted, the data is filtered and the heatmap is updated accordingly.

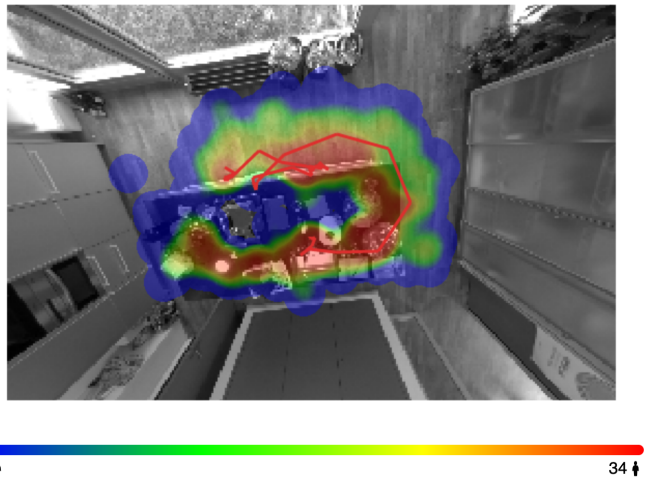

Fig. 5: Adjusting tracking and heatmap based on Camera data

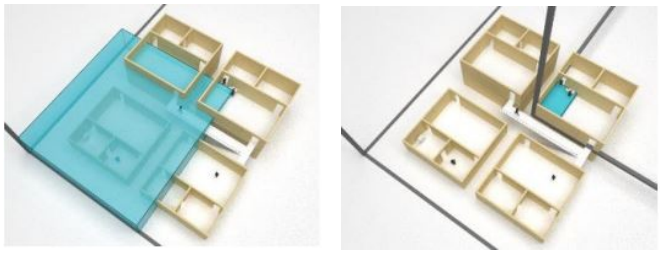

Fig. 6: Absolute (Left) vs. Relative (Right) Positioning [20]

B. Data Transformation

Before the correlation and metrics as well as KPIs are applied, measurement data needs to be transformed into the correct format. Three steps are necessary to correlate data: *(1)* transform relative into absolute positions and *(2)* ensure that the time difference (Δt) is within a predetermined offset for a given time window. Lastly, *(3)* initial filtering of data and erroneous readings from RFID and cameras is performed.

1. Relative and Absolute Positioning. For *(1)* different sensors operate on coordinate systems that differ (*i.e.*, are relative) from the absolute position in the monitoring environment. The necessary input is the internal mapping of the environment in which the position of each sensor (here cameras and RFID readers) and its field of view (*cf.* Figure 6) are determined. For example, cameras have a limited field of view relative to the RFID reader and are typically installed at critical points, such as entrances, exits, or checkout lines, where individuals tend to crowd. Conversely, RFID readers have a more extended range for tag reading, coupled with a reduced accuracy, especially at crowded points.

2. Clock Synchronization. An essential aspect of distributed systems is the need to ensure that different nodes are synchronized, and in the context of using different sensors, the problem applies. At this step *(2)* RFID reader and cameras feature a built-in clock that keeps track of the date and time information (*i.e.*, received data is time stamped from its source). Since the clock is subject to drift, it may need to be synchronized periodically via the Network Time Protocol (NTP) or manually adjusted. Also, when receiving data in a time window, duplicates with the same ID and

time stamp, but in different locations, are eliminated.

3. Data Filtering. The Kalman filter in step *(3)* is used to smooth RFID readings and allows for the increased accuracy of values estimated in the previous time window. It requires two central equations that require specification and parameterization to fit a particular context. The first equation makes a forward projection about the state s based on the previous state, an (optional) control input and an error term [13], [24], $s_k = As_{k-1} + Bu_{k-1} + w_{k-1}$.

The matrix A relates the state s at the previous time window $(k-1)$ to the current time window (k). Matrix B relates the optional control input to the state s. w determines the process noise, assuming that it is Gaussian noise with a mean of 0 and variance of Q, $p(w) \sim N(0, Q)$. The second equation relates the state in time k to the measurement in time window k with a measurement error $z_k = Hs_k + v_k$. Therefore, the matrix H is called the transition matrix. To specify the matrices A and B correctly, it is important to recall the underlying process. The positions x and y in time window k received from RFID data rely on positions of x and y in the previous time window $k-1$ as well as on velocities $\dot{x}$ and $\dot{y}$ at time $k-1$. It is possible to take into account accelerations $\ddot{x}$ and $\ddot{y}$, too.

The model defined assumes a constant velocity for individuals, *i.e.*, the velocity is constant between time windows (Δt) and the accelerations $\ddot{x}$ and $\ddot{y}$ are both 0. While this reads as a simplifying assumption, it is justifiable in this situation, since the time windows are relatively small (1 s). Moreover, variations in speed (*i.e.*, deceleration and acceleration) are considered in the process' noise covariance matrix Q. Therefore, a control matrix B is not specified. The transition matrix is obtained as follows:

$$A = \begin{bmatrix} 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

The Kalman filter also requires parameterizing R and Q. While it is possible to optimize the transition or covariance matrix, instead of providing them, underlying algorithms for this are computationally expensive and not suitable in the context of near real-time applications. One of the parameters needed is the matrix of measurement noises R.

$$R = \begin{bmatrix} \sigma_x^2 & 0 \\ 0 & \sigma_y^2 \end{bmatrix}$$

This matrix contains measurement noises for x and y and their covariance. The covariance is assumed to be zero, so it is needed to obtain the variance of x and y. The RFID manufacturer states that in 66% of cases the real measurement will be within 1 m and in 85% of cases the real measurement will be within 1.5 m [11]. Assuming that the normal distribution holds for these errors, it is possible to fit a normal distribution to the values provided. By adjusting the distribution shape (*i.e.*, the parameter σ^2, since μ is assumed to be 0), the area under their curve is equal to 0.66

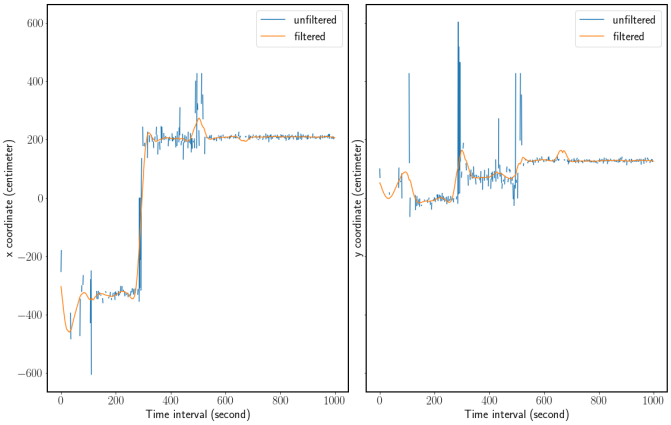

Fig. 7: Comparison of Unfiltered and (Kalman) Filtered Values for x and y Coordinates with Static RFID Tags

(with z values of -0.5 and 0.5). Thus, it is possible to state that 85% of these observations are within 1.5 m, yielding a standard deviation of around 0.52 m. Further, it is assumed that measurement errors for x and y are independent (*i.e.*, $\sigma_{xy} = 0$).

Figure 7 displays the values measured (blue) versus the values filtered (orange). It is visible that large swings are smoothed out (*e.g.*, for x in the interval [0:200]), unless there is a sustainable movement in position (*e.g.*, the one that occurs for x in the interval [200:400]). This is more apparent when looking at measures of y, since there are a few upward swings, all of which are smoothed out by the Kalman filter. While the unfiltered values are relatively more variable, the filtered values show a smaller variation at the beginning of each interval. The scenario used in this example considered a static tag in the time interval [600:1000].

C. Data Correlation

While cameras provide the actual position of individuals detected, they are not able to distinguish between individuals. The camera assigns a new ID to every individual detected in the frame, but different IDs may refer to the same individual leading to inaccurate measurements. The Pearson correlation [1] was used to determine statistical evidence for a linear relationship among variable pairs, such as time stamped coordinates originating from the camera and the RFID reader. The Pearson coefficient **R** in Equation 5 should be at least 0.5 with a tolerance of 1 second between time stamps, which means that for each pair of coordinates $\forall(x, y) \in CAM$ of cameras, and each pair of coordinates $\forall(x', y') \in RFID$ of RFID tags, should have at least 50% of correlation (*i.e.*, moderate) with a max difference of one second between their time stamps. Considering that both RFID and Camera are synchronized, correlation R-values less than 50% and within the time window are discarded. A common results considering possible distortions in the RFID signal especially when multiple tags are grouped during a peak time, for instance.

$$\mathbf{R} = \frac{\sum_{i=1}^{n}(x_i - \overline{x})(y_i - \overline{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \overline{x})^2 (y_i - \overline{y})^2}} \tag{5}$$

In order to normalize the data for RFID readers, boundaries of the 3D camera (*i.e.*, line-of-sight) need to be manually mapped into the RFID reader's coordinate system. Whenever a change is made to the configuration of the RFID reader (*e.g.*, changing the antenna power or moving the reader), and these values must be reassessed. In addition, the matching algorithm has to read values for each RFID reader and apply the normalization. Furthermore, to smooth the location data and improve accuracy a filter is required. In this regard, a Kalman filter is one of the most suitable ones for near real-time applications: it does not require a large memory of past observations, being recursive and relatively fast [3]. Conversely, a moving average would require a certain number of rolling past observations to be kept in memory for smoothing.

D. Process Metrics and KPIs

In the context of real-world applications (*e.g.*, shopping, sports events, and concerts), it is often helpful to create aggregated views of data to understand, for example, the overall trend and make it accessible for a more comprehensive (non-technical) audience. Thus, these data can be used for different purposes, such as to inform policy decisions or evaluate the effectiveness of marketing measures. The scenario of a trade fair (or other exhibitions), for example, is valid for a merchant or Point-of-Reference (PoR) to understand how visitors behave around a booth. In this context, three pieces of information seem of particular interest for practical RFID tracking applications: *(1)* How many of the visitors spot a particular booth? *(2)* How many of the ones, who spot the booth, move closer to see what it is about? *(3)* How many of the ones that inform themselves about the booth move on to interact with the PoR? In an ideal setting, KPI are easy to understand and compare, which is often achieved by using a ratio:

$$\textit{Visibility} = \frac{\text{Number of Views}}{\text{Number of Opportunities}}$$

$$\textit{Engagement Rate} = \frac{\text{Number of Visits}}{\text{Number of Views}}$$

$$\textit{Interaction Rate} = \frac{\text{Number of Interactions}}{\text{Number of Visits}}$$

In order to understand the number of visitors in a place and its behaviors, it is possible to use the concept of *Views* (or impressions). This concept is defined as a count within the views zone of interest, describing the number of times the visit zone was viewed. In a similar direction, the *dwell time* consists of a minimum and maximum time interval (in seconds), which is different for each zone of interest. In this context, the dwell time is used to associate targets to specific areas of interest. Indeed, if a target is within the dwell time range and its location is inside a particular zone of interest, it is classified and assigned to that zone.

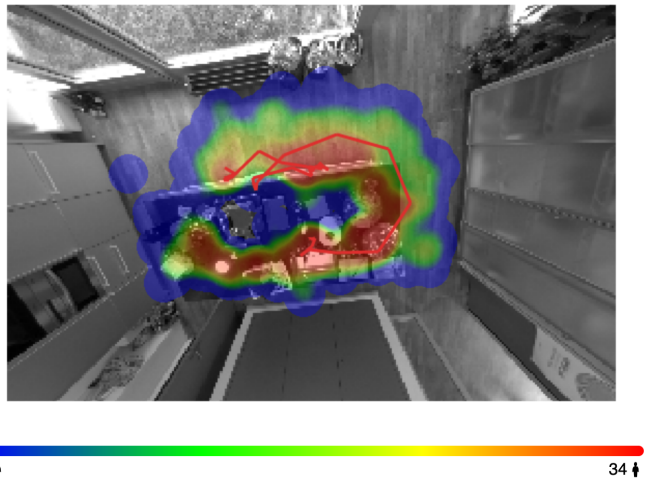

Fig. 8: Heat Map/Path Visualizations based on Camera Data

There are three zones (one for interaction, one for visitors, and one for view) with others designated as opportunities. Given that *(i)* the reader can read 6 m in each direction, *(ii)* the location accuracy is approximately at 50 cm; *(iii)* based on simulating multiple PoRs, *(iv)* KPIs should be sufficiently different for later visualization. While the zone radius was defined in 100 cm steps, the interaction zone shows a radius of 100 cm, the visiting zone has a radius of 200 cm, and the view zone has a radius of 300 cm.

This definition enables a straightforward interpretation (and comparability) in practice, since these values are percentages and, thus, ranging between 0 and 100, because any count in the interaction zone is also a count in other zones. Another consideration is comparability. While the above definition enables the reader to understand the number quickly, it does not tell how the number compares. One can base the assessment on how good a particular value is, which is necessarily between 0 and 100. This, however, is not practical, since it is not known if an interaction rate of 100% is practically achievable. Thus, two benchmarks are used. The first benchmark is an in-group benchmark, comparing the KPI of a particular period to the minimum and maximum of that KPI during a specified time range. The second benchmark does the same, but instead of using in-group comparison, it uses an out-of-group comparison.

Lastly, to determine whether an individual is within any of the relevant zones, the distance between each observation and PoR is calculated by the Euclidean distance, where p denotes the observation of a tag consisting of x and y coordinates and M denotes the PoR position with x and y coordinates: $d(p,M) = \sqrt{(p_x - M_x)^2 + (p_y - M_y)^2}$. Afterwards, a binary variable is created for each PoR's zone. These variables equal 1, if an observation falls within a defined zone, otherwise 0. This ensures that an observation can be simultaneously in multiple zones.

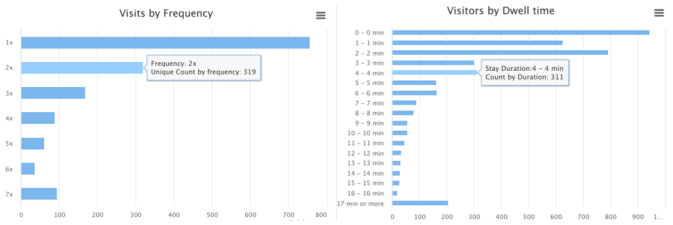

Fig. 9: Visualization of Visitors and Dwell Time

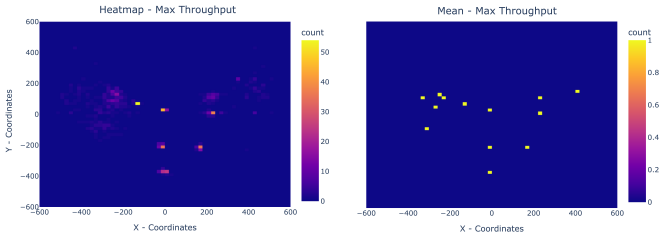

Fig. 10: Heat Map of Max (Left) and Mean (Right) Throughput

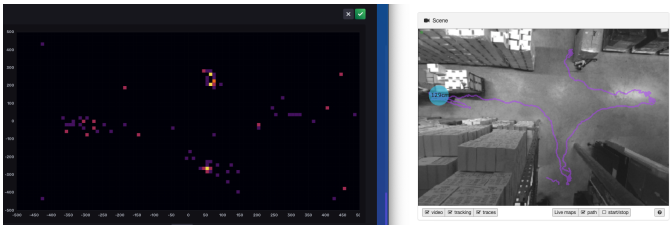

Fig. 11: Recording Frame Used to Correlate RFID Readings

E. Visualization

CCount extensively uses tables and data visualization techniques, such as heat maps, bar charts, and line charts, which rely on data that may be subject to change over time and needs to be shared among different components. The heat map generation process is triggered as soon as the frontend receives the necessary data, in this case, person coordinates, from the backend (*cf.* Figure 8). Heat maps and other KPIs (*e.g.*, number of unique visitors and dwell times as depicted in Figure 9) are dynamically generated based on the predefined time windows. This allows for the iteration over elements in coordinates array and uniquely counts occurrences of every *(x,y)* pair with the time stamp laying within the slider values. This way, it is possible to ensure that the slider is adjusted, data is filtered, and graphical elements are updated accordingly.

IV. Experimental Evaluation

The evaluation of CCount combines and matches data collected by RFID readers with those of 3D cameras. A comparison of different RFID reader (Impinj xArray R680 reader) configurations (*cf.* Table II) demonstrates the results from a practical example. Selecting the adequate RFID reader mode based on the environment is key to accurately calculate the tags' location. The test consisted of fourteen passive UHF RFID tags arranged in different locations within a warehouse and an RFID reader, and a 3D camera mounted on the ceiling. After running each mode for a minute, the collected data was analyzed.

TABLE II: Average Count of Entries and Number of Electronic Product Codes (EPC) Found for Each Reader Mode

Algorithm	Average Count	Number of EPCs found
Auto Set Custom	59.538462	13
Auto Set Dense Reader Deepscan	62.153846	13
Auto Static Dense Reader	62.384615	13
Auto Static Fast	60.692308	13
Dense reader M4	60.769231	13
Dense reader M8	61.384615	13
Hybrid M2	61.307692	13
Max Miller	60.076923	13
Max Throughput	61.461538	13

Despite the use of multiple reader modes, 13 tags out of 14 were captured, *i.e.*, with a success rate of *92.9%*. The "missing" tag was not isolated from the others, instead, it was placed in a dispersed location, where the signal could not be successfully captured. As part of the matching algorithm between 3D cameras and RFID tags, a critical metric is the number of times a particular RFID tag can be captured, *i.e.*, how many data points a reader can deliver within a certain period. As a result of changing several parameters in these configurations, the different RFID reader modes delivered approximately 60 entries within 1 min, which corresponded to the granularity of a 3D camera of one data point per second. Thus, the accuracy of the various algorithms can be compared.

Figure 10 (left side) represents the heat map for RFID tags captured within a minute, whereas the right graph shows the approximate position of data points captured within this minute for each RFID tag. In regions, where tags are more distant, and therefore, the density is low, approximated positions are similar across the maps. However, the RFID reader performs poorly, when tags are clustered together closely, resulting in higher density. As a result, the heat map is unclear and it is impossible to distinguish between the various tags. This behavior results in a large amount of noise in the data, making it more difficult for the matching algorithm to correctly identify individuals wearing an RFID tag as they appear to be moving.

Figure 11 shows a snapshot of a recorded video, which was used as the focal point on the path, where participants walked carrying RFID tags. Thus, it was possible to correlate RFID readings outside the camera's focal point with objects detected by the camera to increase the accuracy of the track. Moreover, this setup was used to assess the system's functionality in real-time successfully.

A. Single RFID Tag

This scenario evaluates the accuracy and precision of CCount, when tracking a single individual wearing an RFID tag under the scene; *cf.* Figure 13 (top) for the correct path of a single individual wearing this tag.. The algorithm

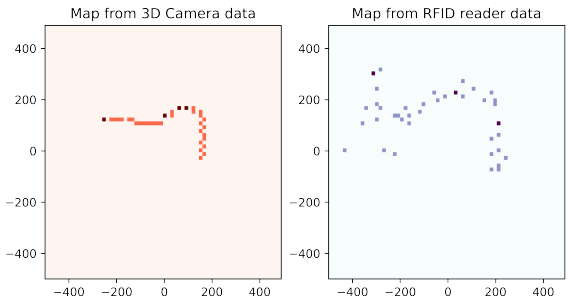

Fig. 12: Maps from a 3D camera and a RFID reader of a single person wearing a RFID tag walking between a group of people.

identifies the individual with a 100% accuracy, when the RFID reader captures the RFID tag's position regularly. The success rate of the correlation was higher than 70-80%. Figure 13 (top) shows the correct path of the single individual wearing the RFID tag.

Since the CCount system assumes that each individual is assigned a personal and unique RFID tag, the scenario considering people randomly walking, but only a single individual carrying a tag should never occur. However, it was conducted for testing the algorithm with arbitrary noise in the data. Several persons were randomly walking under the camera, but only one was carrying an RFID tag. Since these individuals were moving in different directions, thus, making it easier to find the suitable correlation for the algorithm, the tag was successfully assigned to the correct individual. A false-positive rate of 0% was observed.

B. Multiple RFID Tags

A group of individuals wearing RFID tags in a crowded environment was tracked to assess the system's accuracy and precision. The results indicated that the algorithm barely managed to assign tags correctly. While every individual in this variant wore an RFID tag, only one individual was walking. It has been previously demonstrated that, with the presence of multiple tags, the inaccuracy of the RFID reader increases substantially, making it more challenging to match the ID of the 3D camera with the RFID tags.

However, on selected occasions, the algorithm was able to find a match successfully despite the noise. Figure 13 shows a path taken that was correctly assigned. As seen, noise is present in the data coming from the RFID reader. As for the previous scenario, a case considering individuals wearing an RFID tag and standing still did not yield precise results. The algorithm is designed to take only the correlation into account, which resulted in erroneous results as shown in Figure 13 (bottom). A different metric, *e.g.*, the *Euclidian distance*, can prevent this wrong assignment from occurring.

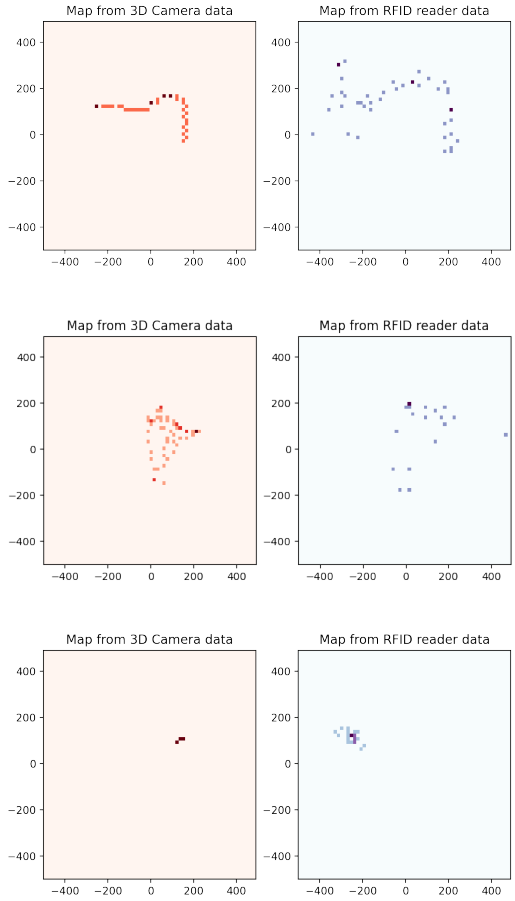

Fig. 13: **Top**: Single Person Walking; **Middle**: Single Person Walking Between a Group (Multiple Tags); **Bottom**: Multiple Persons (Tags) Standing Still in a Group

C. Visitor Counts: 3D Camera vs. RFID Reader

Data collected during these experiments can, however, be analyzed independently from each other and compared. In order to conduct this comparison, visitor counts are grouped according to 1-min intervals (*cf.* Figure 15).

In contrast to the 3D camera, which assigns different and random IDs to people entering and leaving the scene, RFID tags are always uniquely identifiable by the RFID reader, due to their EPC. Thus, data collected from RFID readers can determine the unique number of visitors present at any given point in time. In this scenario, two participants were physically present, but the 3D camera has assigned up to 13 different identities within five minutes (when using a grouping based on 5 min intervals, different IDs were found to be 35). Note that the number of RFID tags did not exceed the actual count of RFID tags present, *i.e.*, 14.

In the period of *10:15-11:05 hours*, these experiments were performed using two different RFID tags and participants entered and left the camera's field of view constantly, resulting in a large discrepancy between the two measure-

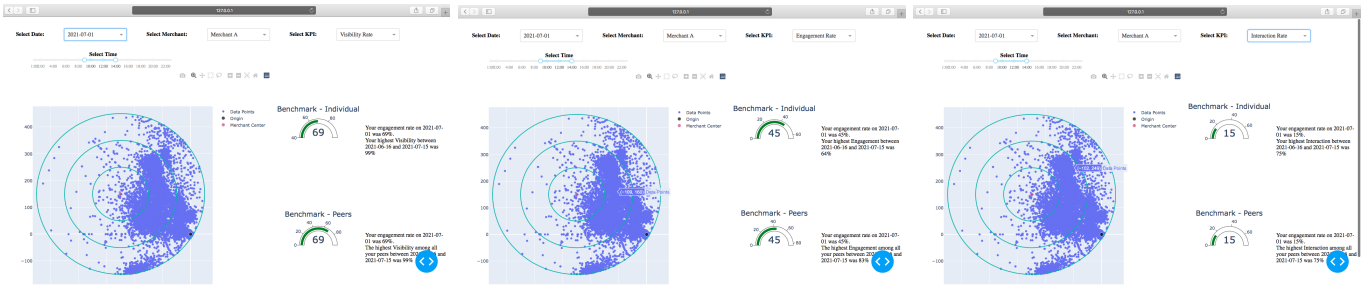

Fig. 14: Dashboard centered as merchant A on July 1 between 09.00 - 14.00
(A) Visibility Rate (B) Engagement Rate (C) Interaction Rate

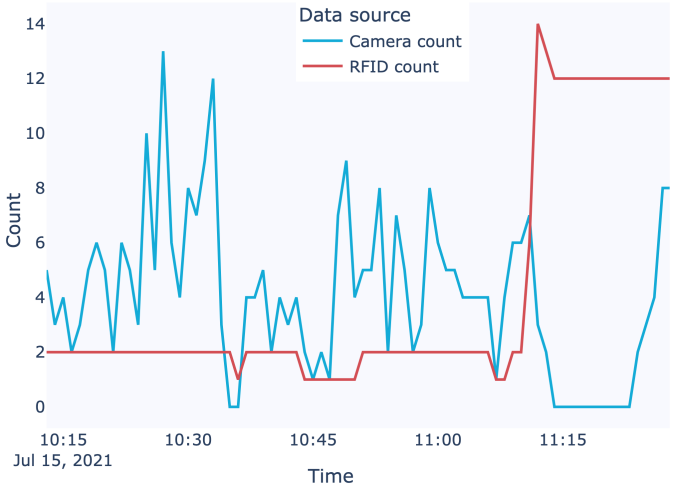

Fig. 15: Visitor Count From 3D Cameras and RFID Readers

ments. In the period of *11:10-11:20 hours*, different reader modes of the RFID reader were evaluated. In this scenario, no participants were present under the scene. Instead, RFID tags were statically placed under the camera. Therefore, the camera's count equals 0, whereas the RFID count is almost constant at 13. Consequently, the number of visitor counts taken from RFID readers data can be directly interpreted as the valid number.

D. Data Visualization

The KPIs defined in the section above are incorporated into suitable visualization and displayed in the dashboard. There are a range of filters embedded to facilitate the analytical use of the board. Figure 14 show how the dashboard changes with different filters.

Marketing campaign statistics for merchant A on July 1, 2021 is shown in Figure 14. It is clear that merchant A has visibility rate at 69%, engagement rate at 45% and interaction rate at 15%. Among the 69% of opportunities that appeared also in the visit zone, less than half of them converts to a visit and less than a fifth of those visitors interacted with merchant representatives. Moreover, one can see that the dials used for benchmarking allows for a quick assessment of the KPI in the broader context. For merchant A, the interaction rate is clearly below its own and peer group averages, while the visibility and engagement rate are both at or slightly above average.

As the web prototype presented in the thesis can be seen as an information-oriented website, the Web Usability Index can be used for evaluating the usability of this web application [9]. This evaluation will go through the checklist by the Index's five categories because the Web Usability Index consists of a checklist with as many as 150 questions [4].

Starting off from the first category: "Navigation and Orientation: consistency of navigation, color of links, etc". The author believes the web prototype has consistent navigation. It is clear that the dashboard has all the filter options located on the top panel of the website. Three filter options are presented as a dropdown menu and one time frame filter choice is done with a timeline plus two adjustable handles.

In terms of "Interaction and Information-Exchange: availability of a homepage, skip functionality for intros, etc", the web prototype does encourage users to engage on the dashboard by providing responsive display depending on the choice of filters a user has made despite the dashboard having only one page.

For the categories "Being up-to-date and Quality: marking of texts with author and date, absence of spelling mistakes, etc" and " Ease of Access and Accessibility: connection between URL and website, availability of high- and low-tech variants of the website, etc", the prototype has been reviewed by the friend of the author. He has confirmed that the website is available on the localhost and free from context errors.

Coming down to the last category on the list "Information- and Text-Design: size of the font, expressiveness if icons, etc", the three visual charts are intentional choices of the author. The zone of interest scatter plot on the left side aims to present the overview of crowd during the selected time range. The two dial plots are meant to provide prompt comparisons across merchants and/or over time. The text next to the dial is an aid for the users and additionally provides a written explanation and the exact analytical numbers.

V. Summary and Future Work

This paper presented Cloud Counter (CCount), a system for collecting, storing, processing, and matching movement data from 3D cameras from third parties with data from RFID readers. CCount allows to tangibly explain the effectiveness of data correlation to improve accuracy in indoor tracking in cases of occlusions and perspective distortions of RFID reader processing.

Experimental results presented both good (*i.e.*, high accuracy) and bad (*i.e.*, low accuracy) scenarios. In the worst-case scenario, especially in situations with several people grouped together being out of the camera's line of sight, the data collected were insufficient to distinguish between single visitors. Nonetheless, even without matching data from 3D cameras, the system was able to identify at any given time the number of unique visitors, since more than 90% of RFID tags were captured. In non-peak situations, the filtering and correlation algorithm resulted in a high accuracy of single individual counts, identifying 13 tags out of 14 that were captured, with a success rate of 92.9%.

Adding 3D cameras is the most efficient alternative to increase accuracy within CCount, but it is generally associated with a considerably higher cost/benefit ratio than using and correlating wireless technologies. 3D cameras are relatively expensive, have a limited line-of-sight in contrast to other range of technologies, such as RFID and IEEE 802.11-based protocols. Lastly, they represent a higher computational cost for image processing at the backend.

The read rate of an RFID tag is generally determined primarily by the amount of Radio Frequency (RF) power that it receives. As a result, RFID tags with a low reading rate are often located at the edge of the reading zone, which lies beyond the field of view of the 3D camera. As a consequence, such tags can be excluded automatically, since they do not qualify as matches.

With regards to these tracking applications, the web-based Dash prototype could be deployed on the AWS and be fully integrated with InfluxDB so that the dashboard could run KPI calculation on near real-time data. The beneficial effect of adapting Dash has been shown: firstly, the code for data processing can be leveraged in the implementation as Dash is also based on Python. Secondly, Dash has a strong focus on data analytics, which aligns with the core of this thesis. Thirdly, the initial work spent on react javascript is not wasted because Dash inherits partially from React. Lastly, Dash combines both front-end and back-end, which makes it ideal for flexible and agile prototype development.

Therefore, future studies in the context of CCount should examine the benefits and costs associated with incorporating more data pre-processing steps into the pipeline and employing additional solutions for determining statistical dependencies. Lastly, different technologies, such as thermal or infrared sensors (*e.g.*, LIDAR), can be incorporated into CCount once they become privacy-preserving sources. This potentially improves the accuracy of measurements and the flexibility of the overall approach to gain new insights into privacy-preserving visitor flow and behavior.

Acknowledgments

This work has been supported by Innosuisse, the Swiss Agency for Innovation Demands under Grant No. 42193.1 IP-ICT. Also, the authors would also like to thank Simon Tuck, Livealytics AG, who supported the research and development during those experiments in Corona times of limited public accessibility and with many discussions on different setups and interpretations.

References

[1] J. Benesty, J. Chen, Y. Huang, and I. Cohen, "Pearson Correlation Coefficient," in *Noise Reduction in Speech Processing*. Springer, 2009, pp. 1–4.

[2] J. H. Betzing, "Beacon-based customer tracking across the high street: Perspectives for location-based smart services in retail," *White Paper, Smart Market Square*, 2018.

[3] S. Bhatti and J. Xu, "Survey of Target Tracking Protocols Using Wireless Sensor Network," in *5th International Conference on Wireless and Mobile Communications (WMC 2009)*. Cannes, France: IEEE, 2009, pp. 110–115.

[4] T. Creutzenberg, "Prototyping: An Overview of Current Trends, Developments, and Research in Prototyping," pp. 40–48. [Online]. Available: http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.227.7242&rep=rep1&type=pdf#page=47

[5] Fedlex, "Federal Act on Data Protection (FADP)," https://www.fedlex.admin.ch/eli/cc/1993/1945_1945_1945/en, 2021, Accessed: 2021-10-17.

[6] K. Finkenzeller, *RFID Handbook: Fundamentals and Applications in Contactless Smart Cards, Radio Frequency Identification and Near-field Communication*. John wiley & sons, 2010.

[7] M. F. Franco, B. Rodrigues, and B. Stiller, "MENTOR: the Design and Evaluation of a Protection Services Recommender System," in *2019 15th international conference on network and service management (CNSM)*. IEEE, 2019, pp. 1–7.

[8] R. Harle, "A Survey of Indoor Inertial Positioning Systems for Pedestrians," *IEEE Communications Surveys & Tutorials*, Vol. 15, No. 3, pp. 1281–1293, 2013.

[9] W. S. Harms and J. Strobel, "Usability Evaluation Von WebAngeboten Mit Dem Web Usability Index." DGI, pp. 283–292.

[10] S. Holcer, J. Torres-Sospedra, M. Gould, and I. Remolar, "Privacy in Indoor Positioning Systems: A Systematic Review," in *2020 International Conference on Localization and GNSS (ICL-GNSS)*. IEEE, 2020, pp. 1–6. [Online]. Available: https://ieeexplore.ieee.org/document/9115496/

[11] Impinj, "Speedway Connect Capabilities," https://www.impinj.com/products/software/speedway-connect, 2021, Accessed: 2021-09-25.

[12] A. Juels, "RFID Security and Privacy: A Research Survey," *IEEE Journal on Selected Areas in Communications*, Vol. 24, No. 2, pp. 381–394, 2006.

[13] R. Kalman, "A New Approach to Linear Filtering and Prediction Problems," *The American Society of Mechanical Engineers- ASME*, Vol. 82, pp. 35–45, 1960.

[14] E. Kalogianni, M. Lam, K. Zhou, and M. Van der Ham, "Passive WiFi Monitoring of the Rhythm of the Campus," in *Proceedings of The 18th AGILE International Conference on Geographic Information Science*, 2005.

[15] L. Mainetti, L. Patrono, and I. Sergi, "A Survey on Indoor Positioning Systems," in *22nd International Conference on Software, Telecommunications, and Computer Networks (SoftCOM 2014)*. Split, Croatia: IEEE, 2014, pp. 111–120.

[16] T. A. Myrvoll, J. E. Håkegård, T. Matsui, and F. Septier, "Counting Public Transport Passenger using WiFi Signatures of Mobile Devices," in *2017 IEEE 20th International Conference on Intelligent Transportation Systems (ITSC)*. IEEE, 2017, pp. 1–6.

[17] R. H. Ribeiro, B. B. Rodrigues, C. Killer, L. Baumann, M. F. Franco, E. J. Scheid, and B. Stiller, "ASIMOV: a Fully Passive WiFi Device Tracking," in *2021 IFIP Networking Conference (IFIP Networking)*. IEEE, 2021, pp. 1–3.

[18] B. Rodrigues, C. Halter, M. Franco, E. J. Scheid, C. Killer, and B. Stiller, "BluePIL: a Bluetooth-based PassIve Localization Method," in *2021 IFIP/IEEE International Symposium on Integrated Network Management (IM)*. IEEE, 2021, pp. 28–36.

[19] B. Rodrigues, L. Müller, E. J. Scheid, M. F. Franco, C. Killer, and B. Stiller, "LaFlector: a Privacy-preserving LiDAR-based Approach for Accurate Indoor Tracking," in *2021 IEEE 46th Conference on Local Computer Networks (LCN)*. IEEE, 2021, pp. 367–370.

[20] G. Sithole and S. Zlatanova, "Position, Location, Place and Area: An indoor Perspective," *ISPRS Ann. Photogramm. Remote Sens. Spat. Inf. Sci*, Vol. 3, No. 4, pp. 89–96, 2016.

[21] A. Vaccari, S. Samouhos, L. Glicksman, and C. Ratti, "MIT Enernet: Correlating WiFi Activity to Human Occupancy," in *Proceedings of Healthy Buildings*, 2009.

[22] P. Voigt and A. Von dem Bussche, "The EU General Data Protection Regulation (GDPR)," *A Practical Guide, 1st Ed., Cham: Springer International Publishing*, Vol. 10, p. 3152676, 2017.

[23] L. Vu, K. Nahrstedt, S. Retika, and I. Gupta, "Joint Bluetooth/WiFi Scanning Framework for Characterizing and Leveraging People Movement in University Campus," in *Proceedings of the 13th ACM international conference on Modeling, analysis, and simulation of wireless and mobile systems*, 2010, pp. 257–265.

[24] G. Welch and G. Bishop, "An Introduction to the Kalman Filter," *Chapel Hill, NC, USA*, p. 16, 2006.

[25] Xovis, "Guaranteed Data Privacy," https://www.xovis.com/de/technology/detail/guaranteed-data-privacy/, 2021, Accessed: 2021-09-25.