

Leveraging Proximity Sensing to Mine the Behavior of Museum Visitors

Claudio Martella*, Armando Miraglia*, Marco Cattani†, Maarten van Steen‡

*VU University Amsterdam, †Delft University of Technology, ‡University of Twente

Email: c.martella@vu.nl, a.miraglia@student.vu.nl, m.cattani@tudelft.nl, m.r.vansteen@utwente.nl

Abstract—Face-to-face proximity has been successfully leveraged to study the relationships between individuals in various contexts, from a working place, to a conference, a museum, a fair, and a date. We spend time facing the individuals with whom we chat, discuss, work, and play. However, face-to-face proximity is not the realm of solely person-to-person relationships, but it can be used as a proxy to study person-to-object relationships as well. We face the objects with which we interact on a daily basis, like a television, the kitchen appliances, a book, including more complex objects like a stage where a concert is taking place.

In this paper, we focus on the relationship between the visitors of an art exhibition and its exhibits. We design, implement, and deploy a sensing infrastructure based on inexpensive mobile proximity sensors and a filtering pipeline that we use to measure face-to-face proximity between individuals and exhibits. Our pipeline produces an improvement in measurement accuracy of up to 64% relative to raw data. We use this data to mine the behavior of the visitors and show that group behavior can be recognized by means of data clustering and visualization.

I. INTRODUCTION

Museum staff design exhibitions to educate, engage and entertain the visitors. Yet, museums rely on surveys and expensive observational studies to collect coarse-grained information about the response of their visitors, with limits of scalability and bias. Central to the understanding of how visitors interact with an exhibition is the identification of which exhibits individuals stop at, for how long, and in which order. According to an ethnographic observational study of visiting styles at the Louvre museum, visitors can be classified into four classes according to their movements [1]. For example, one class of visitors follows a specific path spending a lot of time at almost all exhibits, while another class seems to have a specific preference for some known exhibits at which they spend a lot of time, ignoring the others. This classification has been used, based on activity logs coming from a digital museum guide, to help engage museum visitors and avoid information overload [2]. Characterizing and quantifying visitors behavior helps museum staff evaluate their curatorial decisions, reporting to stake holders and funders, and building data-driven marketing campaigns and applications.

The key to collecting this information is the fine-grained measurement of face-to-face proximity between visitors and exhibits. We call *positioning* the problem of identifying which exhibit an individual is facing at short distance. Existing approaches are not suitable for this problem. Indoor localization technologies estimate the absolute position of an individual in space, without measuring where the individual is looking at. This means that positioning an individual at the closest artwork within a given distance can produce false positives if the

individual has her back towards the exhibit or if she is facing an artwork nearby. Moreover, only very few existing state-of-the-art techniques achieve, in optimal controlled environments, an error of less than 2 meters. With such errors, positionings of the visitor at a wrong exhibit, including those at the other side of a wall, would be frequent. Finally, these techniques often require expensive investments and rely on complex setup procedures each time exhibits are re-arranged.

Contributions. In this paper, we propose a technique based on inexpensive and energy-efficient mobile proximity sensors, and a filtering pipeline to accurately position visitors at exhibits at all times. While we use radio-based proximity sensors, our filtering pipeline is not bound to any particular technology and does not require, for example, measuring distance from exhibits or expensive setups. In particular, we introduce a particle filter tailored to the problem of positioning, together with two smoothening filters that increase measurement accuracy. We use positionings to reconstruct the time spent at exhibits and the visitor path, defined as the ordered sequence of exhibits visited by the individual. We evaluate the approach with data collected from 182 volunteers during a real-world experiment. We show that by clustering this data we can identify group behavior, such as common paths and patterns of time distribution at exhibits.

The remainder of this paper is organized as follows. First, we give an overview of our system, including sensing infrastructure, filtering pipeline, and data analysis. Then, we discuss related work regarding localization systems and museum technology. After describing our model and pipeline, we evaluate the accuracy of our measurements, and apply standard data mining techniques to the data. Finally, we conclude with a discussion about limitations and future work.

II. OVERVIEW

We conducted a 5-days experiment spread across 2 weekends at the CoBrA Museum of Modern Art (CoBrA). Our data collection focused on the temporary exhibition entitled “The Hidden Picture”, a curated sample of the corporate collection of ING. The exhibition was displayed in the dedicated open space at the top floor of the museum. The space is configurable, and divider walls were used to separate the space into 6 “open rooms” dedicated to different themes. The overall space was about 100 meters long and 25 meters wide, with a ceiling reaching about 5 meters, while divider walls were some 3.5 meters high.

Rooms 1 and 2 focused on figurative art, rooms 3 and 4 mostly on abstract art, room 6 on pieces inspired by nature, for a total of 60 pieces. The pieces varied in size, style and medium, including photos, paintings, sculptures, videos, and

an installation with a cage hosting a living chameleon. None of the pieces were highly famous, and were hence appealing the visitors based on immediate reaction rather than on prior knowledge. Of the 60 pieces, we instrumented 45 exhibits with our sensing infrastructure.

A. Data Collection Architecture

We designed a system based on inexpensive radio-based proximity sensors. Our sensing solution is compliant to the Zigbee standard and it can be implemented for example through Bluetooth low energy (BLE) beaconing, available in modern smartphones. To give us freedom to investigate our solution, instead we deployed ad-hoc devices running a duty-cycled MAC protocol [3] that allows us to run our system for weeks with a single battery charge.

The sensing infrastructure comprises *mobile* devices and *anchor* points (or simply anchors). Mobile devices are sensor nodes worn by the visitors. They are attached to a lanyard worn around visitors' neck and hang on the chest. Due to the shielding effect of the visitor's body, the radio communication range is steered to the front with a controlled angle of around 90 degrees and some 2-3 meters of distance. Anchors are sensor nodes positioned at the base of each exhibit. We installed anchors inside of enclosure aluminium boxes designed to shape the communication range to approximately 60 degrees and 2-3 meters of distance. With this setup, mobile devices and anchors can communicate only when the visitor is facing an exhibit.

Every second, anchors transmit through the radio a unique anchor identifier (*AID*) that is received and timestamped by mobile devices within range. We consider the reception of an *AID* by a mobile device a *proximity detection*. Note that our sensors do not measure radio signal strength (i.e., RSSI). While it does not enable us to measure distance between points, it allows a cheaper and more energy-efficient solution. Every second, mobile devices transmit the list of detections received during the previous second together with their unique mobile-point identifier (*MID*) to a longer range of approximately 100 meters, which are received by one or more *sinks*.

Sinks are computers that receive mobile devices transmissions through the same type of sensor node used for anchors and mobile devices, and store the timestamped lists of detections in a central repository. Sinks are installed in various areas of the exhibition space to ensure full coverage and some degree of overlap. Note that due to the overlap of the areas covered by the sinks, mobile devices transmit their messages together with a randomly generated number that we use together with timestamps to remove duplicate detections from the database.

When a mobile device is handed out anonymously to a visitor, the visitor is assigned a unique user identifier (*UID*) that is associated to the corresponding *MID*. Each visitor check-in and check-out times are stored together with the *UID-MID* mapping. Our raw data database comprises this mapping and the list of timestamped detections collected by sinks.

B. Data Filtering Pipeline

The raw database of proximity detections is characterized by a number of shortcomings that obstacle a direct use for visitor positioning without prior filtering. Accurately computing which artwork, if any, a visitor is facing at each second based

solely on raw data is not possible for a number of reasons, related to the irregularity of wireless communications and the imperfect steering of our enclosures.

We designed a data filtering pipeline to estimate, for each second of a visit, at which artwork a visitor is positioned and the sequence of exhibits that defines such visit. The pipeline comprises three steps. First, we filter the data with a *particle filter*. We have developed a technique based on particle filters that takes into account the topology of the exhibition room, the placement and directionality of the anchors, and the movement of the visitor. Note that our technique positions the visitor at exhibits and does not compute absolute coordinates, like traditional techniques based on particle filters designed for localization.

Second, visitor-exhibit mappings are further filtered with a *density-based filtering* algorithm that corrects occasional artifacts introduced by the particle filter. While the particle filter drastically increase positioning accuracy, there are still occasions where a large number of missing detections can cause gaps in positioning data, making it appear as a visitor would return at an exhibit repeatedly in a short window of time (with some seconds in between where the visitor appears having left the exhibit either for the center of the room or, more rarely, for an exhibit nearby). The filtering algorithm fills these gaps and ignores spurious positioning data.

Third, we extract through a *majority voting* filter the sequence of *AIDs* at which the visitor was positioned. The filter scans a visitor positioning data to detect transitions from exhibit to exhibit, and disambiguating situations where a visitor may appear facing two exhibits at the same time. During this step we also reconstruct the *path* followed by the visitor, that is the sequence of stops at exhibits followed by the visitor.

C. Data Analytics Applications

Once raw data has been processed by our pipeline, we obtain for each visitor a vector r_v of N elements, each representing the number of seconds spent at each of the N anchors/exhibits, and a sequence s_v of *AIDs* to represent the path followed to visit the exhibits. If an exhibit was never visited by the volunteer, the corresponding element of r_v will contain a value of 0, and the corresponding *AID* will be missing in s_v .

We can leverage these two data structures to mine the behavior of all the visitors and discover behavioral patterns, like popular rooms and exhibits, common paths followed through the exhibition. Furthermore, if group behavior did emerge during the exhibition, one can leverage historical data of past visitors to predict the behavior of future visitors, from time spent at exhibits to their satisfaction.

III. RELATED WORK

Our work is closely related to the topic of indoor localization. A recent evaluation of 22 indoor localization mechanisms [5] provides us several insights on how state-of-the-art localization techniques could perform in our scenario.

Among all localization mechanisms, only three [6], [7], [8] achieved an error of less than 2m, while only half achieved an error of less than 3m. Moreover, localization errors increase significantly (both in terms of average and deviation) at the edges of rooms and in hallways, were most museums have

exhibits. For our application, this is quite a significant error since the distance between exhibits is usually just few meters. Differently from localization techniques, our mechanism provides a deployment density that matches the placement of exhibits. In other words, by design we tailor our positionings where needed.

Moreover, because museums often have exhibits on both sides of walls, this lack of accuracy produces an even larger error when used for positioning (e.g. positioning a visitor at a painting in the next room or even worse on the next floor). Our combination of enclosures and a tailored particle filter allows us to focus on face-to-face proximity and minimizes these errors. Furthermore, localization systems are usually evaluated in controlled settings and do not account for the variability introduced by people or by changes in the furniture setup, both typical aspects of temporary exhibitions. In more realistic conditions, the localization error of the tested indoor techniques increase by approximately 1.5m to 4m.

There are mainly two types of system. Infrastructure-free mechanisms [7], [8], [9] exploit the existing Wi-Fi access points and require a lengthy calibration phase (*fingerprinting*) that must be repeated every time the environment significantly changes. In case of a museum, this means every time the exhibition changes. Infrastructure-based approaches, on the other hand, usually require a one-time deployment of a hardware infrastructure that can be expensive and that could affect other practical issues such as aesthetics and safety certifications.

Instead of fingerprinting the radio environment, which is susceptible to changes in the environment, most infrastructure-based techniques exploit the propagation speed (*time-of-flight*) of radio [10], [11] and sound [12], [13] to estimate the distance of visitors from the anchor and triangulate their position. Interesting to note, the only localization technique able to achieve sub-meter accuracy [6] is not based on time-of-flight, but on the signal's phase offset, and took several years (5) of development to provide such accuracy. Being based on a proprietary hardware, this technique is difficult to generalize and apply, for example, to smartphone platforms.

Our technique, on the other hand, is generic and based only on face-to-face proximity, that is fuzzy and hard to predict by nature. In particular, since our mechanism does not require to sense the signal characteristic such as the strength or the phase offset, it can be applied to several protocols (e.g. ZigBee, WiFi, Bluetooth, etc.) and even different mediums (e.g. radio, sound and light).

Few sensor-based systems have been used to study the behavior of visitors of art exhibitions, in relation in particular to their movements. Early attempts made use of indoor localization systems based on Bluetooth data collected from mobile phones, to trace the movement of visitors between rooms [14]. This data can be used, for example, to support multimedia guides [15], [16], but it captures only which room (or part of it) an individual is visiting. More recently, data coming from indoor localization and physiological sensors has been used, together with entrance and exit surveys, to study the cognitive reaction and social behavior of a number of individuals in an exhibition [17].

A similar device to measure position and spatial orientation (i.e., through a compass) of the individuals has been used to study the behavior of visitor pairs in a museum. The study

presents a system to classify pairs early in the visit into one of six classes, to provide socially-aware services to the pairs, for example to increase their engagement with the exhibition [18]. This study focuses on the interaction between the individuals and does not attempt to position the pairs at exhibits.

To summarize, none of these approaches tackles the problem of fine-grained face-to-face positioning, and rely either on coarse-grained room-level positioning or on absolute localization without attempting to position visitors at exhibits.

IV. MODEL

We consider a visitor wearing a mobile device v for a duration of T seconds. The exhibition comprises N exhibits, each instrumented with an anchor $a_i \in A = \{a_1, a_2, \dots, a_N\}$. For each visitor v we represent the set of proximity detections as a $N \times T$ matrix D_v , where $D_v(i, j) = 1$ if and only if the mobile device v detected anchor a_i at time j . $D_v(*, t)$ refers to all detections collected at any time t , and $D_v(i, *)$ to all detections of a_i . Similarly, we define a positioning matrix M_v as a $N \times T$ matrix, where $M_v(i, t) = 1$ if and only if visitor v was facing exhibit a_i at time t within a distance smaller than d (i.e., the sensor detection range, in our case 3m.). The definition of M_v is analogous to the definition of a series of proximity graphs [4], with the additional constraint of a visitor being in proximity of maximum one anchor at any time t . Note that there can be times t such that v is not positioned at any exhibit. These are the times when a visitor is walking around the museum or too far to be detected.

A. Problem definition

Our goal is to compute the positioning matrix M_v from the detection matrix D_v . Ideally, if v was facing a_l from second i to second j within distance d , we would have $D_v(l, k) = 1$ for $i \leq k \leq j$ for *only* a_l . In other words, D_v would contain detections for the whole duration of the face-to-face proximity between v and a_l , and there would be *only one* a_l for a given time k such that $D_v(l, k) = 1$. With these perfect conditions, matrix D_v would contain a continuous stream of detections between v and the anchors, and it could be directly used as positioning matrix M_v . However, in practice this is not possible for the following reasons.

First, the database is missing detections. Transmissions between anchors, mobile devices, and sinks can be lost due to many factors, such as message collisions and low signal-to-noise ratios [19], [20]. In other words, while a visitor faces an exhibit, sinks collect a *bursty* stream of detections with *gaps*.

Second, although we control the communication range of the anchors with our enclosures, it is still possible for a mobile device to detect multiple anchors at the same time. A typical scenario is when two anchors are positioned close-by and the visitor stands in an area of overlap between the two anchors' transmission ranges. Looking at the instantaneous raw data does not allow disambiguation of these occurrences. Note that this problem could occur also between anchors positioned at the two sides of a wall, causing a visitor to be in proximity of exhibits installed in two different rooms at the same time.

Third, a mobile device can detect an anchor at a distance larger than the expected transmission range, either due to a corrupted AID in the anchor message (typically caused by a collided message passing a CRC), or due to some tunneling

effect in the wireless transmission. This is perhaps the problem occurring less frequently.

We hence define our pipeline as the set of operations to filter and transform D_v into a M_v for any given visitor v .

B. Particle Filter

We designed and implemented a filter based on particle filters to tackle two challenges related to our positioning problem: (i) due to unreliable wireless communication, some detections could be missing, and (ii) due to the multipath effect, a mobile device could detect multiple anchor points, even far away and on the other side of a wall.

Particle filters have been successfully used in localization to estimate the absolute position of individuals with unreliable sensors [21], [22]. For localization, usually a mobile sensor communicates with a few anchors installed at known locations. It is assumed that the sensor can communicate with all, or a majority of, anchors from all positions and directions, and that the sensor can measure distance from these anchors, for example through signal strength or time-of-flight.

Our setup does not match these assumptions, as we deploy many anchors that communicate only directionally and at short range, without the capability to measure distance. Moreover, we would expect, and to a certain extent desire, mobile devices to detect only one anchor at each time, if any, that is the one in front of the visitor. In other words, our system was designed for the problem of positioning, therefore we need to design a particle filter that reflects the characteristics of our setup and problem.

The filter requires topology information about the exhibition. In particular, a set of anchors $A = \{a_1, a_2, \dots, a_N\}$ each characterized by a position x, y and an orientation α , and a set of walls $W = \{w_1, w_2, \dots, w_M\}$, each defined as a segment between two points, to describe the layout of the exhibition space. We define the set of particles $P = \{p_1, p_2, \dots, p_K\}$, each defined by a position x, y and a weight w (the likelihood of a particle to represent the actual visitor's position). Initially, particles are distributed uniformly at random across the layout of the exhibition space.

Given a detections matrix D_v , the particle filter comprises four steps that are executed for each time t of the visit, with $0 \leq t < T$.

- **Estimation:** We compute the likelihood of each particle's estimate (i.e., its position) given the measurement at time t , that is the set of detections in D_v contained in the t -th column. For each particle p , its weight is computed using the likelihood function $w = \Phi(p, D_v(*, t))$. More details about this function are given later in the text.
- **Positioning:** We estimate the position of v by computing the weighted average among the particle coordinates (i.e., the centroid) and find the closest anchor a_i . We then set $M_v(i, t) = 1$, unless the variance of the particle coordinates from the centroid is larger than a threshold, meaning that the confidence of the estimate is low.
- **Re-sampling:** We create a new set of particles by drawing with replacement from the current weighted set of particles. While drawing particles from the set, we favor particles proportionally to their weight (i.e., their likelihood). As a result, particles with higher likelihood are picked more often than particles with lower likelihood.

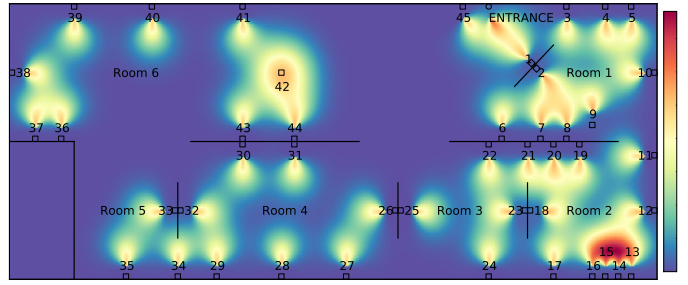


Fig. 1. The landscape of the multivariate gaussian kernel used to compute the likelihood function Φ . For displaying purposes, we assume all anchor points were detected at time t . Note that the likelihood function takes into account angle and distance from anchors, as well as overlap between detection ranges.

If all particles are improbable, we generate new ones at random.

- **Movement:** We move particles at walking speed in random directions, avoiding illegal moves, such as walking through walls.

An important component of our system is the function to compute the likelihood of a particle's estimate. Intuitively, because our sensors are steered to measure face-to-face proximity, we want a particle to have higher likelihood if it is positioned in front of a detected anchor and if it is at close distance. We define $\Phi(p, D_v(*, t))$ as a multivariate function based on a Gaussian kernel with maximum likelihood at 1 meter distance and at 0 degrees angle from the anchor (exactly in front). The likelihood is defined as 0 for distances beyond 3 meters and angles larger than 30 degrees. When multiple detections are present, we mix the likelihood functions of the involved anchors by summing their values. In those cases, the likelihood function computes a multi-iteration between the detected anchors.

For displaying purposes, Figure 1 shows the landscape of Φ assuming that all anchors are detected at the same point in time. In case only one anchor is detected, one can imagine the landscape to have values larger than 0 only in front of the anchor. Note that artwork 42 was a cage, positioned far from the walls, of which visitors could walk around. For this reason, we placed the sensor inside an enclosure that allowed detections from any angle, still within the same distance as defined for the other artworks. The function models directionality and range of both mobile devices and anchors, leveraging the radio directionality due to the enclosures.

The filter keeps positioning a visitor at the last exhibit during gaps of missing detections. As soon as particles spread too much due to their random movements, the filter assumes the visitor is far from any exhibit. Similarly, spurious detections from anchors far away or at the other side of a wall are ignored as it takes time for particles to reach them (e.g. without crossing walls), and hence for their likelihood to be affected by those anchors.

C. Density-based Filter

The particle filter computes the positionings of a visitor regardless of missing, wrong, and ambiguous detections. The resulting matrix M_v contains for each anchor a series of bits that tells us whenever a visitor was facing an anchor at a certain time. While smoothed by the particle filter, the series in M_v can still present gaps, for example, during short periods

where particle confidence was too low (wrongly, due to many missed detections), or when a visitor was wrongly associated to a nearby exhibit. Even though they are rare, we want to remove these artifacts by further smoothing M_v through a density-based filter. In principle, the density-based filter acts analogously to a low-pass filter implemented through a sliding window, but it is able to compute the optimal values for the parameters corresponding to window size and threshold.

The density-based filter [23] first analyses all the series in M_v to compute the k-nearest neighbors statistics for each positioning, and it then uses these statistics to automatically identify bursts of bits through the density-based clustering algorithm DBSCAN [24]. Each identified cluster is effectively a period of time when the visitor was continuously facing an exhibit. Once clusters are identified, we can fill the gaps within the clusters. Note that, at the same time, the clustering algorithm classifies positionings that are isolated and not part of any cluster as noise. The result is a new positioning matrix M'_v that is effectively a smoothed version of M_v .

Note that, differently from M_v , in M'_v visitor v can be positioned at multiple anchor points for the same time t , as series are filtered independently. We solve this problem in the next, and last, step of the pipeline.

D. Majority-voting Filter

We define a majority-voting filter to disambiguate those times in M'_v when we position v at more than one anchor (i.e., those columns of M'_v that have more than one row with a value of 1). The majority-voting filter looks at a window of duration of L seconds ahead of, and including, t to decide at which anchor to position v . The filter decides by choosing the anchor with the largest number of positionings in that window of time. Formally, we let N windows $w_{j,L,a_i} = \langle M'(i, j), M'(i, j+1), \dots, M'(i, j+L) \rangle$ slide in parallel over the series (i.e., rows) in M'_v , with $0 \leq j < T-L$. We break ties by picking the anchor chosen at the previous slide, or at random when no anchor was chosen at the previous slide (i.e., when all windows contain all zeros). While we filter M'_v we create a new positioning matrix M''_v and a path sequence s_v of anchor AIDs that represents the order used to visit the exhibits. We consider v to have transitioned to a new exhibit when the filter positions v at an anchor different from the previous.

To summarize, for each visitor v , the pipeline outputs (i) the positioning matrix M''_v representing at which anchor v was positioned, if any, at each time t , and (ii) the sequence s_v of anchors AIDs that represents in which order v visited the exhibits. Note that not necessarily all anchor are present in s_v and also that an anchor AID can appear multiple times, though not consequently, in s_v . Finally, we can compute a vector r_v of length N , that represents the number of seconds spent by the visitor facing each exhibit. More precisely, we compute the values of r_v as $r_v[i] = \sum_{t=0}^T M''_v(i, t)$, for $0 < i < N$. We will use r_v and s_v in Section VI to identify group behavior.

V. EVALUATION

A. Methodology

We organized two experiments at CoBrA to evaluate our model and infrastructure. In a *controlled* experiment, we scripted 28 visits and asked volunteers to follow the instructions through the script with a timer. The script defined a visit

as a sequence of stops at exhibits, each characterized by a time of arrival and a time of departure from each exhibit. Volunteers were asked to stay at some 1-2 meters of distance and facing the exhibit for the whole duration of each stop. For this experiment, we focused on Room 4 only. While we focused on a subset of the exhibits, we kept all anchors on at all times, including those at the other side of the walls.

In a *real-world* experiment, we asked the visitors of the museum to volunteer in the experiment by wearing one of our sensors during their visit. Volunteers were not instructed or scripted in any way, and could move freely in the exhibition space for the whole duration of their visit. A total of 182 volunteers decided to participate, spread over the 5 days of duration of the real-world experiment. Two human observers collected *ground truth* positionings for 19 volunteers, by annotating arrival and departure times at each exhibit (corresponding to M''_v), and the order used to visit exhibits (corresponding to s_v).

Setups. At the end of the experiments, we processed the data collected by the sinks in the central repository. We utilized the same sensing infrastructure, i.e., the sensors and the enclosures, as well as the same filtering pipeline for both experiments. For the particle filter, we utilized 1000 particles and set the particle speed to 1m/s. Density-based filter parameters were chosen as described in [23] by choosing the knee point of the k-distance plot, corresponding to $\epsilon = 15$ for $minPts = 2$. We chose a window size $L = 10$ that we chose empirically as it would maximize accuracy of both positionings and paths, though results did not vary significantly in the interval [5,30].

Metrics. To measure the performance of our solution at the task of positioning visitors at the exhibits we compute the number of:

- *False positives* (FP): positionings that are present in the measurement but not in the ground truth.
- *False negatives* (FN): positionings that are present in the ground truth but not in the measurement.
- *True positives* (TP): positionings that are present in the measurement and in the ground truth.
- *True negatives* (TN): positionings that are missing in the measurement and in the ground truth.

We compute these values by comparing M''_v and the ground truth for each annotated (or scripted) volunteer. We use these tests to compute two statistical measures of performance for binary classification tests. Sensitivity can be used to measure the ability of the test to identify positive results and is defined as $sensitivity = \frac{TP}{TP+FN}$. Specificity can be used to measure the ability of the test to identify negative results and is defined as $specificity = \frac{TN}{TN+FP}$. Intuitively, they measure the ability to correctly estimate positioning and its absence, respectively. Balanced accuracy, used in cases of unbalanced classes such as ours (where true negatives are much more frequent), is defined as the arithmetic mean of sensitivity and specificity.

To measure the performance of our solution at the task of computing paths, we compared paths extracted through our method with paths extracted from the ground truth. To this end, we use two types of metrics: *sequence-based* and *coordinate-based* metrics.

Sequence-based. We used two metrics designed to compute *similarity* between sequences. The first metric is the Jaro-Winkler (Jaro) [25] similarity metric, which is used to compute

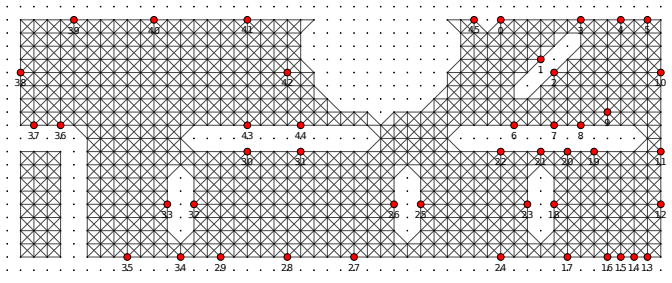


Fig. 2. The mesh used as layout to compute paths between paintings. The planimetry is split in cells, each one mapping 1 square meter of the layout space. Red dots labeled with a number represent artworks. Walls are represented as disconnected dots in the mesh. The circular gap in the upper-central part of the mesh is due to a non-walkable atrium.

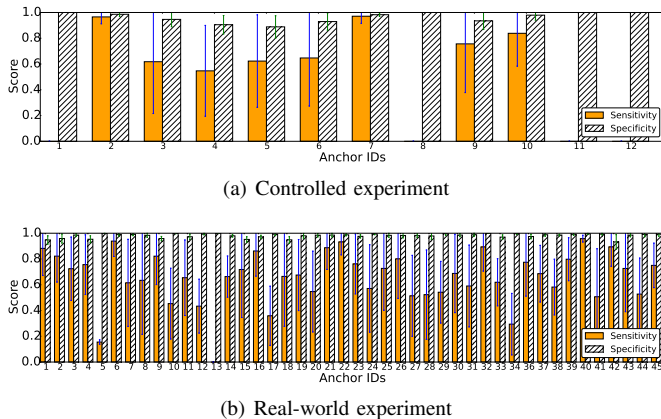


Fig. 3. Anchors sensitivity and specificity values for the two experiments across annotated/scripted visitors. For the controlled experiment, some anchors have no sensitivity because no stops were scripted at those anchors (they were positioned at the other side of walls of anchors involved in the experiment).

string similarity. Jaro is a type of string-edit distance that considers explicitly the mismatch in the order in which elements appear in two sequences (an operation called transposition) and how far mismatching elements are in the sequences. Intuitively, inverting in s_v the order of two exhibits that are nearby in the sequence is less penalized than inverting two exhibits far from each other in the sequence (e.g. $jaro("ABCDE", "ABCE") > jaro("ABCDE", "AECDB")$). The second metric is the Ratcliff-Obershelp (Seqencematcher) [26] matching metric, which is used for pattern recognition and is less forgiving when inverting nearby elements. Seqencematcher matches recursively the elements in the longest-common-subsequences between two given sequences. Both metrics compute a value between 0 and 1, with 1 representing perfect similarity and 0 representing no match.

Coordinate-based. We used a metric designed to compute the *distance* between two sets of coordinates. Hausdorff distance measures how far two subsets of a metrics space are from each other. Intuitively, two sets of coordinates are close if every point in one set is close to some point in the other set. More precisely, Hausdorff is defined as the longest distance from a point in one set to the closest point in the other set. While Hausdorff distance is originally defined on the longest distance, often the mean and median distance are also used to gain a better picture of the distance between the two sets. We generate a set of coordinates from a sequence s_v as follows.

We create a mesh by splitting the layout of the exhibition space in $1m^2$ cells. Mesh edges represent cell edges, and mesh vertices represent cell vertices. In addition to cell edges, we also connect vertices through edges representing cell diagonals. We do not allow edges crossing walls. Every vertex in the mesh has an associated coordinate that depends on its position when overlaid on the exhibition space layout (with the $(0,0)$ vertex being positioned at the top-left of the layout). Figure 2 shows the mesh. We compute the shortest paths between each pair of adjacent elements in s_v and we concatenate the list of coordinates associated with the vertices in the shortest paths. This way, we obtain a representation of the path in space that visitor v would have followed while performing s_v , had she followed the shortest one.

While we do not expect the visitor to have walked precisely through those coordinates, we used also a coordinate-based metric to better measure the impact of errors in s_v . A missing or wrongly added anchor in s_v may result in a single string edit operation of Jaro that does not capture whether the visitor was positioned far away in space from the correct anchor, for example, on the other side of a wall compared to the anchor 1 meter away. Introducing a coordinate-based metric allows us to understand the impact of our errors in relation to the placement of the anchors in the exhibition space. Moreover, this metric allows us to understand the accuracy of path visualizations, as two subsequences with low Hausdorff distance (e.g. if we miss an anchor by positioning the visitor at the anchor 1 meter away) look very similar, often enough for a visualization, regardless of having potentially a low sequence-based similarity.

B. Results

Positioning. Figure 3 presents the mean values and standard deviations of sensitivity and specificity for each anchor. The average sensitivity across the anchors for the controlled and real-world experiments was 0.73 (std 0.02) and 0.61 (std 0.04) respectively, while the average specificity was respectively 0.944 (std 0.002) and 0.981 (std 0.001), for a balanced accuracy of respectively 0.84 and 0.79. Note that the pipeline increases balanced accuracy by 42% in the case of the real-world experiment, and by 63% in the case of the controlled experiment, compared to using raw detections as positionings.

We can notice that most errors are caused by false negatives, as the values of specificity are very close to the maximum value of 1.0, while sensitivity values are smaller and vary more. This can be expected to be caused by missing detections, which are more likely than false positives. Moreover, reaching very high sensitivity may be extremely difficult due to the method used to collect ground truth. As visitors face exhibits for short periods of time (on average around 15 seconds), a small annotation error can impact substantially the sensitivity metric. Observers were asked to start timing a positioning at an anchor when the visitor was facing the exhibit from a distance of some 2-3 meters, but it is difficult for the observer to identify the exact moment the visitor is at range and facing the exhibit. An error of 1-2 seconds about arrival and departure times by the observers can mean missing 10 – 20% of the true positives.

Moreover, often visitors moved during their time at exhibits, getting close and further from the exhibit, temporarily facing somewhere else to discuss with another visitor or to approach an information sheet attached to a wall nearby the

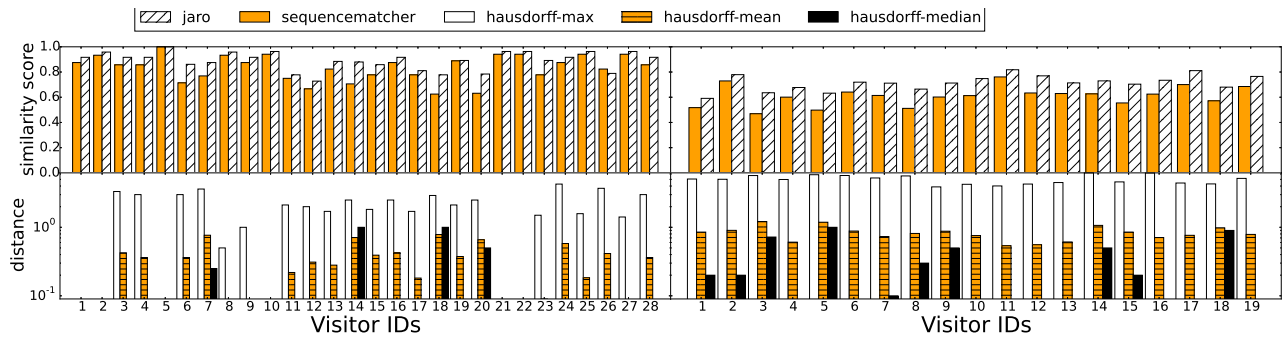


Fig. 4. Sequence-based similarity values (top) and coordinate-based distance (bottom) for the controlled experiment (left) and the real-world experiment (right). We plot coordinate-based distance in log-scale.

exhibit. It is almost impossible for the observers to include all these fine-grained elements in their annotations. This hypothesis is confirmed by the increased accuracy obtained with the controlled experiment, though also in the case of scripted visits it is difficult for visitors to consistently approach and depart the detection range correctly with second-level precision as defined by the script.

Some sensors in denser areas reach often either very low or very high sensitivity, while variance is lower in areas of lower density. This can be an expected effect of the last stages of the pipeline, where smoothing filters favor certain anchors in a winner-takes-all fashion, perhaps due to a favoring conditions of the enclosures position and orientation (e.g. exhibits 5, 13 and 34). This happens less frequently between anchors positioned more distantly from each other.

To investigate the impact of the variability in the sensitivity we ranked the artworks by the amount of time spent in total by all annotated visitors facing them. We ranked exhibits based on the measurement and on the annotations. Intuitively, we would expect the two rankings to be similar if the measurement was accurate enough regardless of variance in missed positionings (i.e., we miss some positionings but we can still order the exhibits by time correctly). We computed Spearman rank correlation between the two rankings to see whether we can capture the relative relationships between anchors. The correlation values for the whole ranking was 0.413 with a p-value of 0.01. We then computed correlations for the top N exhibits for $N = 5, 10, 15, 20$, yielding respectively correlations values $0.996(p = 0.0002), 0.768(p = 0.009), 0.65(p = 0.008), 0.544(p = 0.005)$. These results show that errors are accumulated at the tail of the ranking, where time spent at exhibits are more similar and hence easier to mistake. We would expect to converge to better results with a sample larger than the 19 annotated individuals.

Paths. Figure 4 shows sequence-based similarity and coordinate-based distance values for both experiments, with values for each annotated or scripted visitor. The average Jaro similarity across visitors for the controlled and real-world experiments was 0.879 (std 0.08) and 0.716 (std 0.05) respectively, while Sequencematcher similarity was 0.814 (std 0.09) and 0.613 (std 0.05) respectively. The average sequence lengths were respectively 8.147 for the controlled experiment and 36.316 for real-world experiment. The measures show that we are able to accurately reconstruct path sequences, even when they are long. Moreover, higher values of Jaro similarity confirm that sometimes we invert nearby anchor points in

sequences.

Regarding coordinate-based distance, the average (max, mean, median) values of Hausdorff distance for the controlled experiment and the real-world experiment were $(1.186, 0.317, 0.137)$ and $(5.002, 0.825, 0.245)$ respectively. Mean and median submeter error show that we consistently position the visitor correctly (or less frequently at an anchor nearby), while a 5 meters maximum distance suggests that worst errors cause a positioning of a visitor at a “walking” distance of 5 meters.

VI. APPLICATION

In this section we present the application of data mining techniques applied to visitor data.

A. Methodology

We used the set of r_v vectors and s_v path sequences from all 182 volunteers as dataset for two clustering tasks: (i) identifying common paths chosen by visitors when visiting the museum, and (ii) identifying patterns in the distribution of visiting time across rooms and exhibits. For both tasks, we utilize *Hierarchical Agglomerative Clustering* (HAC) [27], a bottom-up clustering algorithm where items initially form individual clusters and are merged in a series of iterations based on a clustering criterion. We chose the Ward method [28] as a criterion for choosing clusters to merge at each step, which focuses on minimizing total within-cluster variance.

The input of the algorithm is the distance matrix between all items in a dataset. To identify common paths, we compare all s_v sequences with Jaro (and use $1 - jaro(a, b)$ as Jaro computes similarity while HAC requires distance). For the task of identifying patterns of time distribution, we compute the Euclidean distance between all r_v vectors. Before computing distances, we pre-process r_v vectors as follows. First, we use a threshold such that r_v contains only elements larger than 15 seconds (that is we consider for each visitor only the exhibits where she spent more than 15 seconds), and then we scale and center each adapted r_v . In other words, we transform r_v vectors into vectors describing how visitors distributed their time, among those exhibits where they spent more than 15 seconds. We then fed both datasets to the same clustering algorithm.

B. Results

Time distribution. In Figure 5(a) we show the distance matrix between the r_v vectors, which are organized according

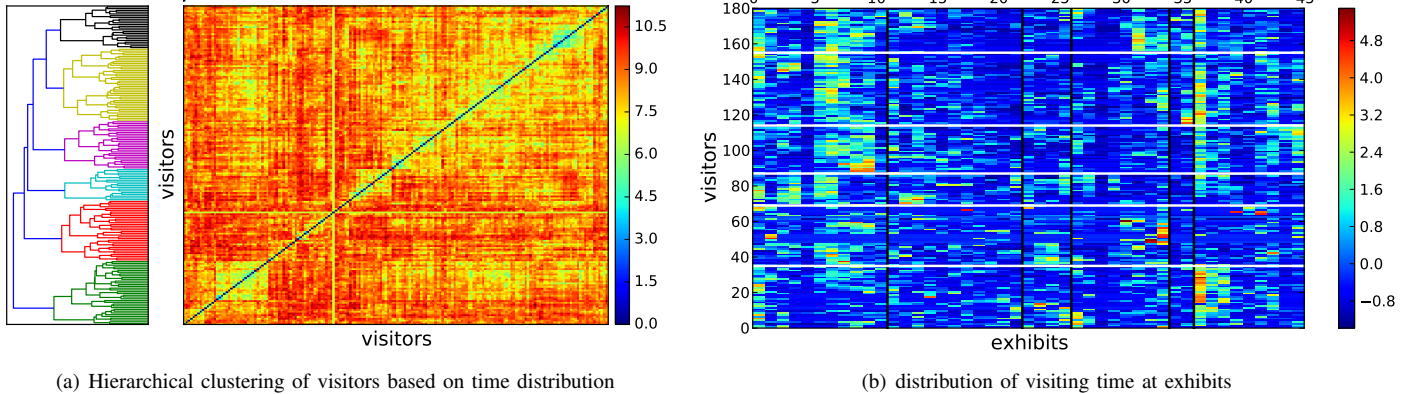


Fig. 5. Hierarchical agglomerative clustering of visitor time distribution vectors. The clustering algorithm identifies one small group (green in the dendrogram) that spent time mostly in room 6, and another major group (top) that is further clustered. Horizontal white lines show cluster divisions and black vertical lines room divisions.

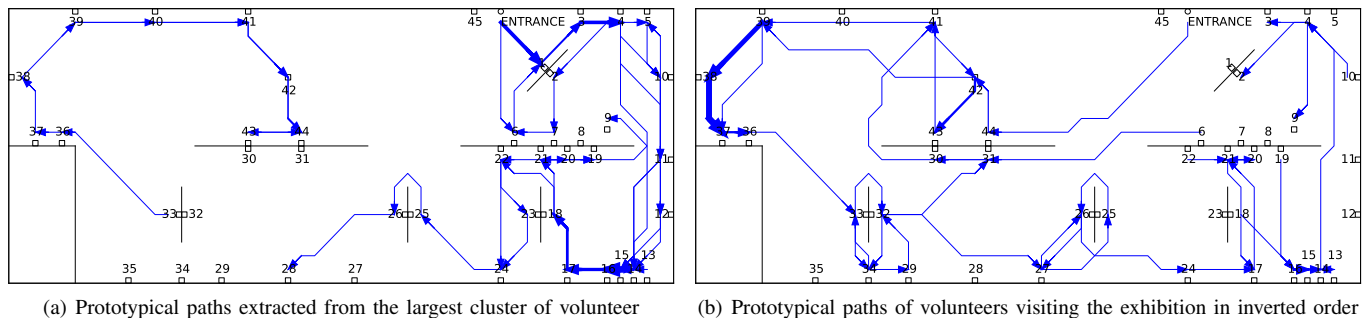


Fig. 6. Visualizations of two groups of common paths identified through hierarchical agglomerative clustering of visitor paths, one (a) representing the common trend among the largest group of visitors, and one (b) representing the 10% of visitors that visited the exhibition starting from the end.

to the result of the agglomerative clustering displayed in the linkage dendrogram (the dendrogram shows how visitors are grouped into clusters and the hierarchy of clusters), and in Figure 5(b) we show the set of pre-processed r_v vectors grouped by the result of the clustering algorithm (vertical black lines show room divisions and horizontal white lines show cluster division). The results show that the clustering algorithm identifies two major clusters. The first cluster includes the bottom visitors (1-38, green in the dendrogram) for their particular interest in 3 artworks in room 6 and little interest in room 1 except for exhibit 1. The second cluster includes all visitors between (65-182) who spent time in room 1 and in front of exhibit 36. This second cluster is further clustered in other clusters, for example, (i) (85-118, violet) due to specific interest in the last 3 exhibits, (ii) (124-160, yellow) due to specific interest in exhibits 1, (iii) (160-182, black) due to some specific interest in exhibits 31, 32, 33 and 34.

Paths. In Figure 6 we show two of the clusters identified with HAC. We construct the visualization of a set of visitor paths s_v by computing the statistics of the longest-common-subsequences (LCS) by comparing all sequences in the set, and choosing the 20 most frequent ones. We visualize these LCSs as paths by computing the shortest paths between the exhibits in the LCSs on the mesh grid, as described previously with respect to coordinate-based metrics (the line width is proportional to the frequency of that LCS). One such visualization summarizes the most frequent path choices made by the visitors in the set. Note that the visualization does not contain time information, but only frequency, and hence frequent LCSs

do not imply that a lot of time was spent at those exhibits.

The first cluster includes the largest group of visitors and shows the most common behavior. HAC identified further groups within this major cluster, for example, splitting visitors turning right towards exhibit 6 from visitors turning left towards exhibit 3 after exhibit 1 at the entrance. The second cluster in Figure 6(b) shows a cluster with 10% of the visitors who visited the exhibition space in “inverted order”, starting from room 6. Perhaps these visitors did not understand what was expected from them by the curators. Similarly, through HAC we identified another group of visitors who decided to visit room 6 first, only to return right after back to room 1 and continued from there, perhaps after realizing their mistake. In general, path visualizations outlined different trends of choices, with some individuals skipping rooms completely, with the most common behavior consisting in following the perimeter of the room along the outer wall.

Furthermore, we can notice that both groups make more different choices of paths while visiting the first rooms, whereas the last rooms are characterized by one common path where visitors scan the room sequentially along the outer walls. This phenomenon was dubbed “museum shuffle” by the staff of CoBrA and was associated to a decrease in attention after around 30 minutes in the visit.

VII. DISCUSSION AND FUTURE WORK

The proposed solution is able to reconstruct the behavior of a group of visitors with a pipeline of software filters. Through a simple and inexpensive sensing infrastructure, we were able

to accurately position visitors at paintings regardless of data loss and noise caused by our enclosures and the absence of distance estimation from anchors.

We plan to investigate how estimating distance from anchors could impact the sensitivity of the measurement, in particular in areas with higher density of anchors where our technique produces more ambiguous positioning. Similarly, we would expect substituting our enclosures with tailor-made directional antennas to produce more robust and reliable communication ranges, and hence more accurate measurements.

With the widespread use of wearable sensors, like smart watches, glasses, and bracelets, pervasive and ubiquitous sensing capabilities will extend further than those provided by smart phones only. A fusion of different data sources should allow us to better quantify the quality of the proximity relationships between visitors and artworks.

We plan to repeat our measurements in conditions of much higher density and scale. We expect such challenging settings to increase data loss and noise, but they also represent an opportunity to study how positioning data can help tackling high-density conditions, which are positioned by museum staff at the top of their list of undesirable circumstances.

VIII. CONCLUSIONS

We presented a method to measure person-to-object relationships via face-to-face proximity. The positionings extracted from our sensors can be processed through data mining techniques that identify group behavior. The approach is inexpensive and requires little setup. The method provides a reliable measurement of positioning that is not bounded to the particular technology used to sense face-to-face proximity, and can hence be used in a variety of scenarios and applications. Our approach is not specific to our devices or to a particular medium, and it applies to any technology that can map onto the proposed binary model.

ACKNOWLEDGEMENTS

This publication was supported by the Dutch national program COMMIT.

REFERENCES

- [1] E. Verón and M. Levasseur, *Ethnographie de l'exposition: l'espace, le corps et le sens*. Centre Georges Pompidou, Bibliothèque publique d'information, 1989.
- [2] T. Kuffik, Z. Boger, and M. Zancanaro, "Analysis and prediction of museum visitors behavioral pattern types," in *Ubiquitous Display Environments*. Springer, 2012, pp. 161–176.
- [3] M. Dobson, S. Voulgaris, and M. van Steen, "Merging ultra-low duty cycle networks," in *Proceedings of the 41st International Conference on Dependable Systems & Networks (DSN 2011)*, 2011.
- [4] C. Martella, M. Van Steen, A. Halteren, C. Conrado, and J. Li, "Crowd textures as proximity graphs," *Communications Magazine, IEEE*, vol. 52, no. 1, 2014.
- [5] D. Lymberopoulos, J. Liu, X. Yang, R. R. Choudhury, V. Handziski, and S. Sen, "A realistic evaluation and comparison of indoor location technologies: experiences and lessons learned," in *Proceedings of the 14th International Conference on Information Processing in Sensor Networks*. ACM, 2015.
- [6] R. Reimann, A. Bestmann, and M. Ernst, "Locating technology for aal applications with direction finding and distance measurement by narrow bandwidth phase analysis," in *Evaluating AAL Systems Through Competitive Benchmarking*. Springer, 2013.
- [7] C. Beder and M. Klepal, "Fingerprinting based localisation revisited: A rigorous approach for comparing rssi measurements coping with missed access points and differing antenna attenuations," in *Indoor Positioning and Indoor Navigation (IPIN), 2012 International Conference on*. IEEE, 2012.
- [8] C.-L. Li, C. Laoudias, G. Larkou, Y.-K. Tsai, D. Zeinalipour-Yazti, and C. G. Panayiotou, "Indoor geolocation on multi-sensor smartphones," in *Proceeding of the 11th annual international conference on Mobile systems, applications, and services*. ACM, 2013.
- [9] H. Zou, L. Xie, Q.-S. Jia, and H. Wang, "Platform and algorithm development for a rfid-based indoor positioning system," *Unmanned Systems*, vol. 2, no. 03, 2014.
- [10] S. Adler, S. Schmitt, Y. Yang, Y. Zhao, and M. Kyas, "Fubloc: Accurate range-based indoor localization and tracking," Technical report, Microsoft Indoor Localization Competition, Tech. Rep., 2014.
- [11] A. Ashok, C. Xu, T. Vu, M. Gruteser, R. Howard, Y. Zhang, N. Mandayam, W. Yuan, and K. Dana, "Bifocus: Using radio-optical beacons for an augmented reality search application," in *Proceeding of the 11th annual international conference on Mobile systems, applications, and services*. ACM, 2013.
- [12] P. Lazik and A. Rowe, "Indoor pseudo-ranging of mobile devices using ultrasonic chirps," in *Proceedings of the 10th ACM Conference on Embedded Network Sensor Systems*. ACM, 2012.
- [13] Z. Jiangy, W. Xiy, X.-Y. Li, J. Zhaoy, and J. Hany, "Hiloc: A tdoa-fingerprint hybrid indoor localization system," Technical report, Microsoft Indoor Localization Competition, Tech. Rep., 2014.
- [14] Y. Yoshimura, F. Girardin, J. P. Carrascal, C. Ratti, and J. Blat, "New tools for studying visitor behaviours in museums: a case study at the louvre," *19th International Conference on Information Technology and Travel & Tourism*, pp. 1–13, 2012.
- [15] G. Wilson, "Multimedia tour programme at tate modern," in *Museums and the Web*, vol. 3, 2004.
- [16] E. Bruns, B. Brombach, T. Zeidler, and O. Bimber, "Enabling mobile phones to support large-scale museum guidance," *IEEE multimedia*, no. 2, pp. 16–25, 2007.
- [17] V. Kirchberg and M. Tröndle, "The museum experience: Mapping the experience of fine art," *Curator: The Museum Journal*, vol. 58, no. 2, 2015.
- [18] E. Dim and T. Kuffik, "Automatic detection of social behavior of museum visitor pairs," *ACM Transactions on Interactive Intelligent Systems (TiiS)*, vol. 4, no. 4, p. 17, 2014.
- [19] G. Zhou, T. He, S. Krishnamurthy, and J. A. Stankovic, "Impact of radio irregularity on wireless sensor networks," in *Proceedings of the 2nd international conference on Mobile systems, applications, and services*, ser. MobiSys '04, 2004.
- [20] N. Ababneh, "Radio irregularity problem in wireless sensor networks: new experimental results," in *Proceedings of the 32nd international conference on Sarnoff symposium*, ser. SARNOFF'09, 2009.
- [21] F. Evennou, F. Marx, and E. Novakov, "Map-aided indoor mobile positioning system using particle filter," in *Wireless Communications and Networking Conference, 2005 IEEE*. IEEE, 2005.
- [22] F. Gustafsson, F. Gunnarsson, N. Bergman, U. Forssell, J. Jansson, R. Karlsson, and P.-J. Nordlund, "Particle filters for positioning, navigation, and tracking," *Signal Processing, IEEE Transactions on*, 2002.
- [23] C. Martella, M. Dobson, A. van Halteren, and M. van Steen, "From proximity sensing to spatio-temporal social graphs," in *Pervasive Computing and Communications (PerCom), 2014 IEEE International Conference on*. IEEE, 2014.
- [24] M. Ester, H. P. Kriegel, J. S, and X. Xu, "A density-based algorithm for discovering clusters in large spatial databases with noise," in *KDD, 1996*.
- [25] W. E. Winkler, "String comparator metrics and enhanced decision rules in the fellegi-sunter model of record linkage," 1990.
- [26] J. W. Ratcliff and D. E. Metzener, "Pattern-matching-the gestalt approach," *Dr Dobbs Journal*, vol. 13, no. 7, p. 46, 1988.
- [27] O. Maimon and L. Rokach, *Data mining and knowledge discovery handbook*. Springer, 2005, vol. 2.
- [28] J. H. Ward Jr, "Hierarchical grouping to optimize an objective function," *Journal of the American statistical association*, vol. 58, no. 301, 1963.