Indoor—A New Data Management Frontier

Christian S. Jensen† Hua Lu† Bin Yang‡

†Department of Computer Science, Aalborg University, Denmark
{csj, luhua}@cs.aau.dk
‡School of Computer Science, Fudan University, China
byang@fudan.edu.cn

Abstract

Much research has been conducted on the management of outdoor moving objects. In contrast, relatively little research has been conducted on indoor moving objects. The indoor setting differs from outdoor settings in important ways, including the following two. First, indoor spaces exhibit complex topologies. They are composed of entities that are unique to indoor settings, e.g., rooms and hallways that are connected by doors. As a result, conventional Euclidean distance and spatial network distance are inapplicable in indoor spaces. Second, accurate, GPS-like positioning is typically unavailable in indoor spaces. Rather, positioning is achieved through the use of technologies such as Bluetooth, Infrared, RFID, or Wi-Fi. This typically results in much less reliable and accurate positioning.

This paper covers some preliminary research that explicitly targets an indoor setting. Specifically, we describe a graph-based model that enables the effective and efficient tracking of indoor objects using proximity-based positioning technologies like RFID and Bluetooth. Furthermore, we categorize objects according to their position-related states, present an on-line hash-based object indexing scheme, and conduct an uncertainty analysis for indoor objects. We end by identifying several interesting and important directions for future research.

1 Introduction

During primarily the past decade, an increasingly large body of research results on moving objects has come into existence (e.g., [1, 6, 10–12]). Some of these results serve as a technology foundation for the growing location-based services (LBSs) industry. However, most moving-object research to date assumes an outdoor setting with GPS, or GPS-like, positioning. This research, unfortunately, falls short in another very important setting, namely indoor spaces.

Indoor spaces may accommodate very large populations of moving individuals. In fact, people spend large parts of their lives in indoor spaces such as private homes, office buildings, shopping malls, conference facilities, airports, and subway stations. With positioning being available in indoor spaces, it is easy to imagine that we are able to provide a wide range of indoor location-based services akin to those enabled by GPS-based positioning in outdoor settings. Example indoor services include navigation, personal security, a variety of location-based information services, and services providing insight into how and how much an indoor space is being used.

Bulletin of the IEEE Computer Society Technical Committee on Data Engineering

Much of the research on outdoor moving objects is not easily applicable in indoor settings. This can be attributed in part to two differences between indoor and outdoor settings.

First, indoor spaces are composed of entities that are unique to indoor settings: often rooms and hallways connected by doors, as exemplified in Figure 1.

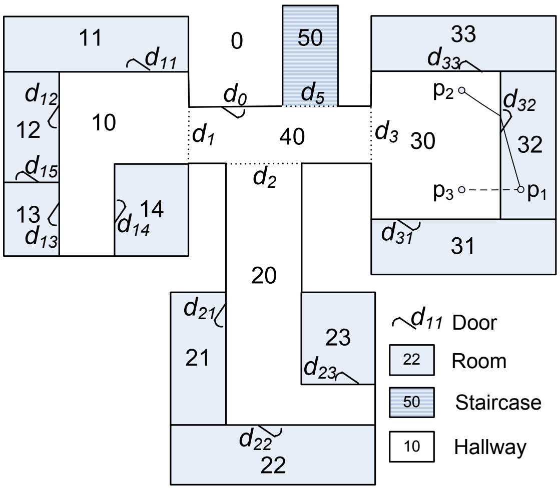

Figure 1: Example Indoor Space

These entities enable and constrain movement. In the example, a user who wishes to move from location p_1 to location p_3 must to go through door d_{32}; the wall between room 32 and room 30 blocks the direct movement.

Such constraints render the conventional Euclidean distance inapplicable in indoor settings. If disregarding the indoor topology, location p_1's (Euclidean) nearest neighbor is p_3. However, taking the indoor topology into consideration, p_1's true nearest neighbor is p_2.

In addition, indoor movement is less constrained than outdoor spatial-network constrained movement, where the position of an object is constrained to a position on a polyline. Consequently, symbolic models rather than geometric models are often used for modeling indoor spaces [3].

Second, GPS-like positioning is typically unavailable in indoor spaces. Rather, other positioning technologies are deployed in indoor settings that differ fundamentally from GPS-like positioning. Specifically, technologies that have been proposed for short-range communication, such as RFID [13], Bluetooth [4], and Infrared, can be exploited for indoor positioning. However, unlike GPS that is able to report continuously positions and velocities of moving objects with varying accuracies, such technologies often rely on proximity analysis [7] and are unable to report velocities or accurate locations.

In particular, an indoor object is detected only when it enters the activation range of a positioning device, e.g., an RFID reader or a Bluetooth base station. Depending on the deployment of devices, such detections occur more or less frequently. As a result, the indoor positioning technologies create much more uncertain tracking data in indoor spaces when compared to outdoor settings.

The differences between the outdoor and indoor settings call for new research on indoor moving objects. This paper covers some of the background and results of such ongoing research. The paper presents a graph-based model for the effective and efficient tracking of indoor moving objects with proximity-based positioning technologies like RFID and Bluetooth. It presents an indexing scheme for on-line indoor moving objects. It also conducts a brief analysis on the inherent uncertainty of indoor moving objects. Finally, it suggests several interesting and important research directions.

The rest of this paper is organized as follows. Section 2 presents a graph-based model for indoor tracking. Section 3 presents foundations for the management of indoor objects, presenting a hashing-based indexing scheme and covering also object location uncertainty. Section 4 concludes and offers research directions.

2 Tracking Indoor Moving Objects

2.1 Symbolic Indoor Positioning

In our setting, each deployed positioning device detects and reports the objects that enter its range at a relatively high sampling rate. For example, in an RFID-based positioning system, an RFID reader can detect objects with passive tags attached. Or in a Bluetooth-based system, a base station can detect objects equipped with a Bluetooth-enabled device. A raw reading of the form $\langle deviceID,\ objectID,\ t\rangle$ states that device $deviceID$ detected object $objectID$ at time t.

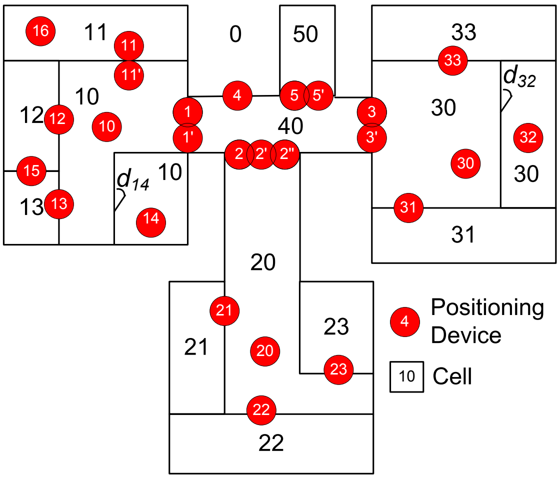

Figure 2: Device Deployment

Figure 2 shows a possible positioning device deployment, where the numbered circles indicate the positioning devices and their activation ranges. For positioning devices with overlapping ranges, we treat the intersections activation ranges of new, virtual positioning devices. Thus, the intersection of $device_1$ and $device_{1'}$ is assigned to a virtual device $device_{1'1}$. An object seen by $device_1$, but not $device_{1'}$, is then in the non-intersecting part of the range of $device_1$.

We also accommodate so-called paired devices (covered in Section 2.2) that detect movement direction, e.g., the entry into or exit from a room.

We apply pre-processing to the raw readings in order to support subsequent on-line and off-line applications. An on-line record is of form $\langle deviceID,\ objectID,\ t,\ flag\rangle$, where $flag = ENTER$ indicates that the object is entering the device's activation range, and $flag = LEAVE$ indicates the object is leaving the range. Note that such records cam be emitted when an object enters or leaves the range of a device with a delay not exceeding the sampling frequency. In contrast, an off-line record is of the form $\langle deviceID,\ objectID,\ t_s, t_e\rangle$, which indicates the presence of the object within the device's activation range during the time interval $[t_s, t_e]$. The details of this pre-processing can be found elsewhere [8].

2.2 Positioning Device Deployment Graph

In a deployment, a subset of the devices, the so-called *partitioning devices*, partition the indoor space into cells in the sense that an object cannot move from one cell to another without being observed. An example is a device deployed by the single door of a room. *Undirected partitioning devices* (*UP*) cannot detect movement directions between cells. In Figure 2, $device_{21}$ cannot tell whether an observed object enters or leaves cell c_{21}.

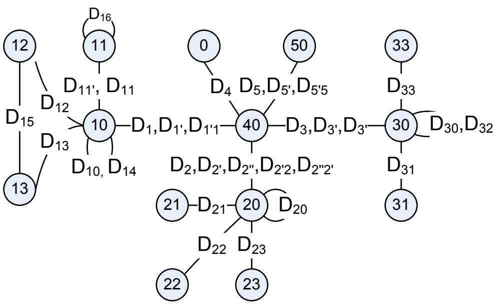

Figure 3: Deployment Graph

Note that $device_1$, $device_{1'}$, and $device_{1'1}$ are also undirected. In contrast, *directed partitioning devices* (*DP*) consist of entry/exit pairs of sensor that enables the movement direction of an object to be inferred by the reading sequence, e.g., $device_{11}$ and $device_{11'}$ in Figure 2. Finally, *presence devices* (*PR*) simply sense the presences of objects in their ranges, but do not contribute to the space partitioning. Device $device_{10}$ in Figure 2 is a presence device.

To facilitate tracking and querying moving objects, a deployment graph is created based on the topological relationship of the floor plan and the device deployment [8]. Formally, a deployment graph is a labeled graph $G = \langle C, E, \Sigma_{devices}, \ell_E\rangle$, where:

(1) C is a set of vertices corresponding to cells.

(2) E is a set of edges. Each edge is an unordered pair of vertices, indicating that the two cells are connected.

(3) ℓ_E: $E \rightarrow 2^{\Sigma_{devices}}$ assigns a set of devices to an edge. A non-loop edge is labeled by the partitioning device(s) that partition its two cells, and a loop edge captures the presence device(s) in the edge's cell.

Figure 3 shows the deployment graph corresponding to Figure 2; label D_i indicates a positioning device $device_i$.

2.3 Graph Model Based Indoor Tracking

The goal of indoor tracking is to capture the position of an object at any point in time. We propose techniques for both on-line and off-line tracking. By exploiting the indoor floor plan, the deployment graph, and maximum speeds of objects, we try to minimize the possible region(s) an object can be in at a particular time [8]. In doing so, we exploit the deployment graph, which captures the indoor topology that constrains the movements of indoor objects. For example, an object can only move from a graph vertex (a cell) to an adjacent vertex (another cell connected with some partitioning devices).

Given a set of off-line records in the form of $\langle deviceID,\ objectID,\ t_s, t_e\rangle$, off-line tracking of an indoor moving object is conducted in three steps. Step one augments each reading record with corresponding deployment graph elements (vertices or edges) during the time interval $[t_s, t_e]$. Pre-defined mappings between positioning devices and relevant graph vertices (cells) are also used in this step.

Step two identifies cells that an object can possibly be in during its *vacant time intervals*, which are the intervals during which no tracking record exists for the object. Specifically, step one tells where (graph elements) the object is before and after a vacant time interval; its position during the vacant time interval is constrained to the graph elements that connect the before and the after parts. So, by intersecting the graph elements before and after the vacant time interval, we identify the cells the object can be in during the vacant time interval.

Step three makes use of the maximum speeds of the objects and reduces the possible cells obtained in step two to smaller regions. If the object moves at its maximum speed V_{max} from any point inside the activation range of a device and its trajectory is a straight line, its position at time t_x will be bounded by circles centered at all possible start points in the activation range and with radius $V_{max} \cdot \Delta t_1$. Applying this constraint to two consecutive tracking records, the possible region of the object during a vacant time interval can be simplified to an speed-constrained ellipse.

Given a set of records of the form $\langle deviceID,\ objectID,\ t,\ flag\rangle$, on-line tracking treats the cases where $flag$ is either $ENTER$ or $LEAVE$ differently. Details can be found elsewhere [8].

3 Management of Indoor Moving Objects with Inherent Uncertainty

3.1 Indexing of Indoor Moving Objects

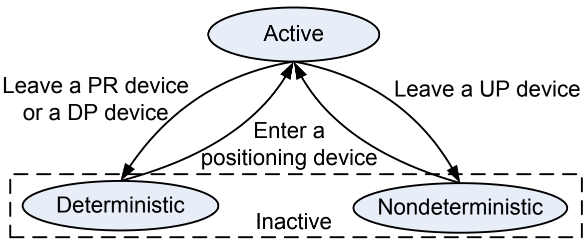

Figure 4: Object State Transition Diagram

An object may be active or inactive. An *active object* is currently seen by at least one positioning device, while an *inactive object* is currently not seen by any positioning device. The latter are further divided into *deterministic objects* that must be in one specific cell and *nondeterministic objects* that may be in more than one cell. An object can change state according to the diagram shown in Figure 4.

The consequent partitioning of objects can be exploited in a hashing-based object-location indexing technique. Let O_{indoor} be the set of all the moving objects in the indoor space of interest. A *Device Hash Table (DHT)* maps each positioning device, identified by $deviceID$, to the set of active objects in its range:

$$DHT[deviceID] = O_A; \quad deviceID \in \Sigma_{devices}, O_A \subseteq O_{indoor}$$

Next, a *Cell Deterministic Hash Table (CDHT)* maps each cell, identified by $cellID$, to the set of deterministic objects in it:

$$CDHT[cellID] = O_D; \quad cellID \in C, O_D \subseteq O_{indoor}$$

Similarly, a *Cell Nondeterministic Hash Table (CNHT)* maps a cell to the set of nondeterministic objects in it:

$$CNHT[cellID] = O_N; \quad cellID \in C, O_N \subseteq O_{indoor}$$

Finally, an *Object Hash Table (OHT)* captures the states of all objects:

$$OHT[objectID] = (STATE,\ t,\ IDSet); objectID \in O_{indoor}$$

Here $STATE$ denotes the object's current state and t is the start time of the state. If the object's state is active, $IDSet$ is a singleton set of a device identifier. If the state is deterministic, $IDSet$ is a singleton set of a cell identifier. If the state is nondeterministic, $IDSet$ is a set of cell identifiers.

The update of these hash tables and their use in query processing are covered elsewhere [14]. Also, it is possible to extend the R-tree to index large volumes of historical indoor tracking data [9].

3.2 Uncertainty Analysis for Indoor Moving Objects

As for outdoor moving objects [5], the uncertainty region of an indoor object o at time t, denoted by $UR(o, t)$, is a region such that o must be in this region at time t. The uncertainty region of an active object is the activation range of the corresponding device, while the uncertainty region of an inactive object is the cell or cells that the object can belong to.

If the object's maximum speed V_{max} is given, its uncertainty region can be captured at a finer granularity. The uncertainty region of a deterministic object is refined as the intersection between the object's cell and its *maximum-speed constrained circle*. For a nondeterministic object, the region is the union of the intersections between each cell and the circle.

Let the last $LEAVE$ observation of object o be from device dev at time t and let the duration from t to the current time be $\Delta t = t_{now} - t$. The longest possible distance o can move away from the boundary of dev's activation range is $o.V_{max} \cdot \Delta t$. Formally, the maximum-speed constrained circle $C_{MSC}(o, dev, t)$ of o is defined as the circle centered at dev's deployment location and with radius $o.V_{max} \cdot \Delta t$ plus the radius of dev's activation range. We also exclude the activation range of dev from the circle.

Consider Figure 5 and assume that object o left $device_{16}$ at time t. Its maximum-speed constrained circle $C_{MSC}(o, device_{16}, t)$ is then indicated by R_1 in the figure. Since $device_{16}$ is a presence device, after leaving $device_{16}$ the inactive object o must be in the cell c_{11} (according to $G.\ell_E^{-1}(device_{16})$). Due to the two constraints, object o's uncertainty region is the intersection of cell c_{11} and circle R_1, i.e., the shaded region in the top-left part of Figure 5.

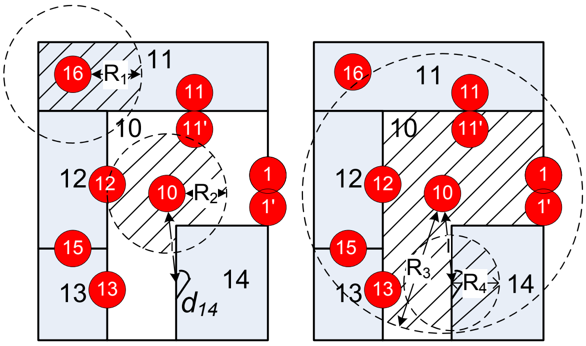

Figure 5: Uncertainty Regions

If the cell where the deterministic object resides has more than one room, e.g., cell c_{10} contains rooms 10 and 14, the determination of the uncertainty region is more complicated. Suppose object o left $device_{10}$ at time t. According to $G.\ell_E^{-1}(device_{10})$, o should be in cell c_{10} after leaving $device_{10}$. From predefined mappings that capture the deployments of devices [15], it follows that $device_{10}$ resides in room 10 and that the distance from $device_{10}$ to door d_{14} is l. If the maximum speed constraint guarantees that o cannot have gone through door d_{14}, i.e., $o.V_{max} \cdot (t_{now} - t) < l$, the object o must remain in room 10. Thus, the uncertainty region is the intersection between room 10 and $C_{MSC}(o, device_{10}, t)$, which is indicated by R_2 to the left in Figure 5.

On the other hand, referring to the right part of Figure 5, if $o.V_{max} \cdot (t_{now} - t) \geq l$, object o may have entered room 14. Its uncertainty region therefore contains two parts: the intersection between room 10 and $C_{MSC}(o, device_{10}, t)$ (indicated by R_3); and the intersection between room 14 and the circle with door d_{14} as the center and $R_4 = o.V_{max} \cdot (t_{now} - t - l/o.V_{max})$ as the radius.

The uncertainty region of an active object can also be refined in the similar way [14, 15].

The online indexing scheme and the uncertainty analysis have been used for processing queries on indoor moving objects, e.g., indoor range monitoring [14] and indoor k nearest neighbor queries [15].

4 Conclusion and Future Work

Indoor spaces differ substantially from outdoor spaces and are not modeled well by Euclidean spaces or spatial networks. Further, indoor positioning may be accomplished by presence-sensing technologies rather than the GPS-like positioning that is often assumed in research targeting outdoor settings. Due to these and other factors, "indoor" offers new research challenges.

This paper offers a glimpse of selected aspects of the foundations for ongoing research on data management for indoor moving objects. It touches upon graph model based indoor tracking, indoor moving-object indexing, and the capture of the uncertainty of indoor moving objects.

There are many research opportunities in data management for indoor spaces. Here, we mention but a few.

- It is of interest to integrate different types of positioning technologies in order to improve tracking accuracy. For example, we may combine proximity analysis based on RFID and Bluetooth with fingerprinting-based technologies like Wi-Fi [2]. This may yield an augmented graph model [8] of indoor space.
- In addition to relying on symbolic locations for co-location queries, it is of interest to accommodate distances in indoor models. This may enable distance-aware queries. For example, it becomes possible to monitor closest pairs of indoor moving objects. As another example, given a distance value e, an e-distance join returns those pairs of objects whose distance is smaller than e. Distance-aware queries may have security and social-network applications.
- Given large volumes of real tracking data, it is interesting to mine patterns or association rules. This may enable on-line prediction of aggregate and individual movements, which in turn may improve on-line tracking accuracy. It may also serve to improve query processing efficiency.
- While initial research has assumed that objects move independently, it is of relevance to consider more advanced models of object movement. For example, it is relevant to conduct probabilistic analyses that assume Gaussian distributions.
- It is worth developing benchmarks for indoor moving object data management that enable the comparison of competing techniques. Relevant aspects include, but are not limited to, indoor space and positioning device configuration, object movement workload generation, and query workload generation.

Acknowledgments

This research was partially supported by the Indoor Spatial Awareness project of the Korean Land Spatialization Group and BK21 program. C. S. Jensen is an adjunct professor at University of Agder, Norway.

References

[1] P. K. Agarwal, L. Arge, and J. Erickson. Indexing Moving Points. In *Proc. PODS*, pp. 175–186, 2000.

[2] P. Bahl and V. N. Padmanabhan. RADAR: An In-Building RF-Based User Location and Tracking System. In *Proc. INFOCOM*, pp. 775–784, 2000.

[3] C. Becker and F. Dürr. On Location Models for Ubiquitous Computing. *Personal Ubiquitous Computing*, 9(1):20–31, 2005.

[4] S. Feldmann, K. Kyamakya, A. Zapater, and Z. Lue. An Indoor Bluetooth-Based Positioning System: Concept, Implementation and Experimental Evaluation. In *Proc. ICWN*, pp. 109–113, 2003.

[5] R. Cheng, D. V. Kalashnikov, and S. Prabhakar. Querying imprecise data in moving object environments. *IEEE TKDE*, 16(9):1112–1127, 2004.

[6] R. H. Güting, M. H. Böhlen, M. Erwig, C. S. Jensen, N. A. Lorentzos, M. Schneider, and M. Vazirgiannis. A Foundation for Representing and Quering Moving Objects. *ACM TODS*, 25(1):1–42, 2000.

[7] J. Hightower and G. Borriello. Location Systems for Ubiquitous Computing. *IEEE Computer*, 34(8):57–66, 2001.

[8] C. S. Jensen, H. Lu, and B. Yang. Graph model based indoor tracking. In *Proc. MDM*, pp. 122–131, 2009.

[9] C. S. Jensen, H. Lu, and B. Yang. Indexing the trajectories of moving objects in symbolic indoor space. In *Proc. SSTD*, pp. 208–227, 2009.

[10] G. Kollios, D. Gunopulos, and V. J. Tsotras. On Indexing Mobile Objects. In *Proc. PODS*, pp. 261–272, 1999.

[11] S. Šaltenis, C. S. Jensen, S. T. Leutenegger, and M. A. Lopez. Indexing the Positions of Continuously Moving Objects. In *Proc. SIGMOD*, pp. 331–342, 2000.

[12] M. Pelanis, S. Šaltenis, and C. S. Jensen. Indexing the Past, Present, and Anticipated Future Positions of Moving Objects. *ACM TODS*, 31(1):255–298, 2006.

[13] R. Want. RFID Explained: A Primer on Radio Frequency Identification Technologies. *Synthesis Lectures on Mobile and Pervasive Computing*, 1(1):1–94, 2006.

[14] B. Yang, H. Lu, and C. S. Jensen. Scalable continuous range monitoring of moving objects in symbolic indoor space. In *Proc. CIKM*, pp. 671–680, 2009.

[15] B. Yang, H. Lu, and C. S. Jensen. Probabilistic threshold k nearest neighbor queries over moving objects in symbolic indoor space. In *Proc. EDBT*, pp. 335–346, 2010.