The Great Indoors: A Data Management Frontier

Christian S. Jensen Hua Lu
Department of Computer Science, Aalborg University, Denmark

1 Introduction

Much of the research on data management for moving objects has assumed an outdoor setting in which objects move in Euclidean space (possibly constrained) or some form of spatial network and in which GPS or GPS-like positioning is assumed explicitly or implicitly. That body of research provides part of an enabling foundation for the growing Location-Based Services industry.

However, we lead large parts of our lives in indoor spaces: homes, office buildings, shopping and leisure facilities, and collective transportation infrastructures. The latter may be large: For example, each day in 2009, London Heathrow Airport, UK had on average 180,000 passengers, and the Tokyo Subway (Tokyo Metro and Toei Subway), Japan delivered a daily average of 8.7 million passenger rides in 2008. Tokyo's Shinjuku Station alone was used by an average of 3.64 million passengers per day in 2007.

Indoor differs from outdoor in important ways and thus calls for new research. The remainder of this paper covers selected differences between indoor and outdoor and discusses the implications for research.

2 Indoor Versus Outdoor

Indoor Services 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, and a variety of location-based information services, as well as services providing insight into how an indoor space is being used.

Symbolic Indoor Space Indoor spaces are composed of delineated and crisp entities, e.g., rooms and hallways bounded by walls and connected by doors, that are unique to indoor settings. These entities enable as well as constrain movement. Thus, when an individual wishes to move from one room to another in an office building, the doors and connecting hallways must be used—the walls block direct movement.

The implications are several. Rather than using (latitude, longitude) coordinates pairs as positions in outdoor Euclidean spaces, users may be positioned in terms of the discrete indoor entities or partitions of these.

Next, the conventional Euclidean distance is inaccurate and thus not generally applicable in indoor spaces. The true travel distance between two indoor locations may differ significantly from the Euclidean distance between the locations. 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. Specifically, indoor space may be modeled using a graph model where the vertices are symbolic locations such as rooms and edges indicate accessibility between locations.

Proximity-Based Indoor Positioning The positioning available in indoor settings may differ fundamentally from GPS-like positioning. Technologies that have been proposed for short-range communication, such as

RFID [6] and Bluetooth [1], can be exploited for indoor positioning. Unlike GPS that is able to report the positions and velocities of moving objects continuously (at 1 Hz) and with relatively high accuracies, such indoor technologies often rely on proximity analysis [2] 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 particular device deployment, such detections occur more or less frequently. As a result, the indoor positioning technologies yield location data that is much more uncertain [8] when compared to outdoor, GPS-like positioning.

Following a post-processing step, the positioning system is then capable of delivering records of the form $\langle deviceID,\ objectID,\ t,\ \{enter|leave\}\rangle$ in real time or records 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 $[t_s, t_e]$, in off-line settings [3].

In a positioning device 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. In contrast, *presence devices* simply sense the presences of objects in their ranges, but do not contribute to the space partitioning.

To facilitate the tracking and querying of moving objects, a deployment graph is created that captures the deployment of positioning devices. This graph has a vertex for each cell, and its edges then indicate connectivity between the cells. The edges are labeled by the devices that capture movement between the cells.

3 Example Ongoing Research

The differences between the outdoor and indoor settings call for new research on indoor moving objects. To illustrate the differences between indoor and outdoor, we consider briefly the tracking and indexing of indoor moving objects.

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 [3]. 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. In doing so, we exploit the deployment graph, which captures the indoor topology that constrains the movements of indoor objects.

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. First, we augment each reading record with corresponding deployment graph elements (vertices or edges) during the time interval $[t_s, t_e]$.

Then we identify the 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. We have already determined where 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 after parts. Thus, we intersect the graph elements before and after the vacant time interval to identify the cells the object can be in. Finally, we make use of the maximum speeds of the objects and that way reduce the possible cells obtained in step two to smaller regions.

The resulting representation of an object's past movement should be contrasted with the polyline representation that is typically used in outdoor space.

Indexing of Indoor Moving Objects Because of the discrete nature of indoor space, hashing may be applied for indoor indexing. 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. The consequent partitioning of objects can be exploited in a hashing-based indexing scheme. Specifically, we can create hash functions that:

- Map devices to the active objects in their ranges.
- Map cells to the deterministic objects they contain.
- Map cells to the non-deterministic objects they contain.
- Map objects to the cell or cells they are or can be located in.

The update of these hash functions and their use in query processing are covered elsewhere [7]. Also, it is possible to extend the R-tree to index large volumes of historical indoor tracking data [4].

4 Research Directions

There are many research opportunities in data management for indoor spaces [5]. Here, we mention but a few.

- It is of interest to integrate different types of positioning technologies in order to improve positioning accuracy. Notably, integration with fingerprinting-based technologies like Wi-Fi is called for.
- Integration with outdoor positioning will enable services that cross the indoor/outdoor boundary.
- 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, which 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 tracking accuracy and 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.

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] 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.

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

[3] C. S. Jensen, H. Lu, and B. Yang. Graph Model Based Indoor Tracking. In *Proc. MDM*, pp. 122–131, 2009.

[4] 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.

[5] C. S. Jensen H. Lu, and B. Yang. Indoor–A New Data Management Frontier. In *IEEE Data Engineering Bulletin*, 33(2), 2010, to appear.

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

[7] 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.

[8] 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.