Manage and Query Generic Moving Objects in SECONDO

Jianqiu Xu, Ralf Hartmut Güting
Database Systems for New Applications, Mathematics and Computer Science
FernUniversität Hagen, Germany
{jianqiu.xu,rhg}@fernuni-hagen.de

ABSTRACT

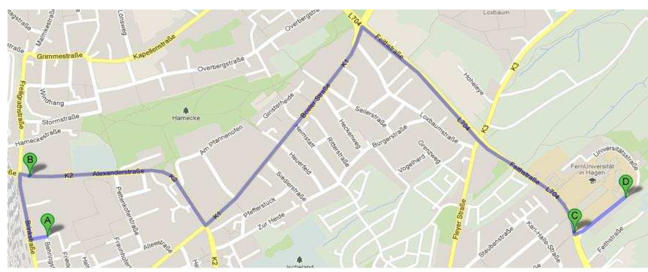
In this demonstration, we introduce a system that is able to manage moving objects in all real world environments, e.g., road network, bus network and indoor. The complete trip of a person is managed by the system such as *Walk*, *Car*, *Walk*, and *Indoor*, where the precise locations of both outdoor and indoor movements are represented. Trajectories located in several environments are integrated into the same framework. The system supports the shortest path searching for start and end locations being in different environments, for example, from a room to a bus stop. A comprehensive and scalable set of moving objects is generated to simulate human movement in practice. Optimization methods are developed to efficiently answer novel queries regarding transportation modes and mobile environments. Most of these queries are not supported by existing methods because of the limitation of data representation.

1. INTRODUCTION

Recently, the area of moving objects with different transportation modes receives a lot of attention in the literature [4, 6, 11, 7], due to novel applications on detecting transportation modes and providing advanced trip plannings or recommendations. Researchers try to infer outdoor transportation modes from raw GPS data in order to capture important characteristics and features of mobile users, as these pieces of information can help fully understand the mobility. Additionally, finding the correlation between mobile users and locations will benefit from understanding the complete trajectory with precise transportation modes, e.g., recommending friends based on location history [12]. In an advanced transportation system, it is very meaningful to provide trip plannings including choices and constraints on the modes, e.g., less than two bus transfers.

From the viewpoint of the database community, the complete trip of humans should be managed by a database system, enabling novel queries on transportation modes and underlying environments. In spite of massive research work having been conducted on the area of moving objects databases, previous works mainly deal with one environment such as free space or road network. However, in the real world a person's trip is not limited to the aforementioned two, but can include more cases such as bus network and indoor. Evidently, only managing a sub trip is not sufficient enough to fully understand the context of mobility. Consequently, existing methods only support queries on moving objects in one environment and do not concern about transportation modes. Fig. 1 depicts two example movements each of which contains several transportation modes, where the trip starts from the home place and ends at the office room. Only the outdoor movement is shown.

Trip 1: $A \xrightarrow{Walk} B \xrightarrow{Bus} C \xrightarrow{Walk} D \xrightarrow{Indoor}$ office room
Trip 2: $A \xrightarrow{Car} B \xrightarrow{Walk} C \xrightarrow{Indoor}$ office room



(a) Example Trip 1

(b) Example Trip 2

Figure 1: Trips with Multiple Transportation Modes

A data model is proposed in [4] with the aim of providing advanced trip plannings. However, they do not provide data types to represent moving objects with transportation modes but describe them abstractly. Only the outdoor environment is considered. In our system, we can provide the precise path for trip plannings in both outdoor and indoor environments. Besides trip plannings, more interesting queries can also be answered.

The major difficulty of managing the complete trip is to define a framework that is able to represent the precise location of moving objects in all available environments. Efficient query processing mechanisms should also be provided to support new queries. For trip plannings, the current techniques only support the case that start and end locations are in the same environment. However, people may start from one building and end at another where the connection route can include roads, pavement areas and indoor paths. In fact, in Fig. 1 Google map supports a sequence of places to be visited and the user can choose private vehicles, walking or public vehicles, but different modes can not be specified for sub trips.

In the demonstration proposal, we introduce how to manage moving objects with precise transportation modes in a database system. That is, a complete trip can be handled by the system, as opposed to a sub trip limited to one environment. Outdoor and indoor trips are uniformly treated, and accurate locations are represented. Compared with the outdoor environment, the location representation inside a building needs both horizontal and vertical information. As a result, a general framework is proposed to apply in all environments. Sub trips with different modes can be distinguished. Besides moving objects, the underlying environments are also managed where the location of moving objects is in fact represented by referencing to the environment. Five real world environments are considered: road network, region-based outdoor (pavement areas), bus network, metro network and indoors (a set of buildings). Due to the difficulty of getting a large amount of real data, a tool is developed in the system to produce a comprehensive and scalable set of moving objects that model human movement in practice. Moving objects with multiple transportation modes are generated and parameters can be configured in order to produce trips with different distributions in terms of locations, time and modes. We can run interesting queries on generic moving objects by proposed data access methods. A majority of these queries is not supported by previous techniques.

According to transportation modes, we can efficiently retrieve a sub trip and get its attributes, e.g., the length of a walking path, the duration of indoor movement. The relationship between moving objects and underlying environments can also be discovered. We can determine the precise location where people change the mode such as *Taxi* → *Walk* and *Walk* → *Bus*. In addition, the system supports finding an optimal route with the resulting path being located in different environments, e.g., "*find the shortest path with the minimum time from my home to the office room*". One possible result is *Walk* → *Metro* → *Walk* → *Indoor*.

In the rest of the paper, we introduce the data representation in Section 2, present the method of creating datasets in Section 3. In Section 4, we show the queries on generic moving objects. The system demonstration is reported in Section 5.

2. DATA MANAGEMENT

We consider the following real world environments for generic moving objects: Road Network, Region-based Outdoor, Bus Network, Metro Network and Indoor. These environments constitute the space for moving objects. Then, the location of moving objects is represented by referencing to the space. Figure 2 depicts an outline of the method. Correspondingly, the data manager is divided into three components responsible for: (1) infrastructure data; (2) space; (3) moving objects.

An environment (also called infrastructure in the following) consists of a set of geographic objects (GEOB) such as roads for Road Network and pavement areas for Region-based Outdoor. Since each environment has its own characteristics, different data types are proposed to represent the objects. For example, a line is used to describe a road, and a polygon defines a pavement area. New data types are also designed due to the feature of certain objects. Considering the indoor case, a building consists of a set of rooms. The data type for a room should contain (1) 2D area and (2) the height above the ground level. The information of these GEOBs is to be

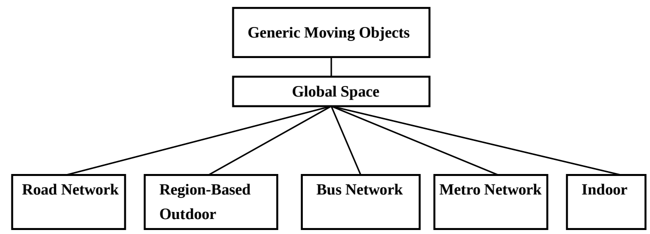


Figure 2: A Sketch of Data Model

used to identify the location for generic moving objects. There are not only static objects, but also buses and metros represented by moving objects.

Each GEOB is assigned a unique id and a set of disjoint ranges is defined for all environments, e.g., [1, 5000] for Road Network, [6000, 7000] for Bus Network. The space is built on top of these environments to be the foundation for generic moving objects, and one can determine the location environment for a given GEOB id. Afterwards, we represent the location of a moving object by two parts, defined as follows.

Definition 2.1 *Generic Location*
$D_{\underline{genloc}} = \{(oid, (loc_1, loc_2)) | oid \in D_{\underline{int}}, loc_1, loc_2 \in D_{\underline{real}}\}$

The first part is an identifier corresponding to the underlying object and the second stands for the relative position according to that object, described by (loc_1, loc_2). The referenced object contains its own geographical information so that the range of a location can be determined. With the second value, a precise location can be identified. This method is able to represent the location in all cases. Based on the location representation, a data type for generic moving objects is proposed. Basically, the movement is described by a sequence of temporal units ordered by time where each unit defines the movement during a time interval. Let $mo = < u_1, u_2, ..., u_n >$ denote a generic moving object. We have u_i = $(i,\ gl_1,\ gl_2,\ m)$ $(gl_1, gl_2 \in D_{\underline{genloc}} \wedge gl_1.oid = gl_2.oid$, $m \in \underline{int})$ where i denotes the time interval, gl_1, gl_2 are the start and end locations, respectively, and m is the transportation mode such as *Walk*, *Car*, and *Indoor*. We assume that the object moves linearly during i so that the positions between gl_1 and gl_2 are calculated by a linear function.

Consider the example Trip2 described by a sequence of transportation modes: *Car* → *Walk* → *Indoor*. The value of mo is to be a set of units recording ids for (1) roads; (2) pavements; and (3) rooms. In each unit, the precise location is identified by gl_1 (gl_2). Take the indoor movement as an example. gl_1 (gl_2) stores the relative location inside the room where the origin point is defined to be the left lower point of the 2D bounding box on the room.

In the system implementation, each environment is developed to be a set containing (1) a symbol to show the environment type; (2) all its GEOBs and (3) some indices to efficiently access the data, e.g., B-trees and R-trees. The space is implemented as a list of items each of which (i) records a reference representation to the corresponding environment and (ii) stores the integer range of GEOBs in that environment. We use a relation to store moving objects where the data type is embedded as an attribute in a tuple, thus leveraging the power of relational operators, e.g., select.

3. DATASETS

Due to the difficulty of getting the comprehensive and realistic data of moving objects with transportation modes, we develop a tool in the system to create the data. Besides, we also create all GEOBs based on some input data and defined parameters because it is also not easy to get the real data for all of them in a consistent environment, e.g., pavements and buildings. The procedure of generating datasets consists of three phases: (1) creating infrastructure data; (2) building the space; (3) producing moving objects.

To represent the location of generic moving objects, all infrastructures have to be managed in the database system. We developed a tool MWGen [8, 10] to create all real world environments based on roads and floor plans. Since the input data have a lot of public resources, the tool can be used by other researchers to create their own data. We provide a relational interface for infrastructure data where relations are used to store GEOBs of an environment.

We extended MWGen to create a scalable set of generic moving objects to model human movements in practice. A set of movement rules that reflect the characteristics of human mobility is defined to produce trips. For example, there are *regular* trips between home and work places. During the office time, people may travel to another place for a conference or business meeting. On the weekend, people may go shopping or visit friends. There are also trips produced by the nearest neighbor query, e.g., a pedestrian searches the closest restaurant. The result might be *Walk* $\rightarrow$ *Bus* $\rightarrow$ *Walk*.

Table 1 summarizes all possible transportation modes. Modes *Indoor* and *Walk* are included for each generated trip. The reason is two-fold: (1) usually people start and end their movements inside buildings; (2) in practice people do not directly change from *Bus* to *Car*, or from *Indoor* to *Taxi*. A short distance walking is required where the mode *Walk* is the transition between different transportation modes. This is consistent with the result in [11] where the authors infer transportation modes from raw GPS data and find that the walk segment is up to 99%. Parameters such as location and time are defined with the aim of configuring the settings to create the data with different distributions. The system is also able to create the trip with a single mode, e.g., moving inside the building, walking around in the city center.

Table 1: Transportation Modes for Moving Objects

$$\textit{Indoor} + \textit{Walk} + \begin{cases} \textit{Bike} \\ \textit{Car} \\ \textit{Bus} \\ \textit{Metro} \\ \textit{Taxi} \end{cases}$$

4. QUERIES

A group of queries on transportation modes and mobile environments can be formulated by an SQL-like language using proposed operators and the relational interface managing moving objects and infrastructure data. Some query examples are given below. More queries and the formulation refer to [9]. We can execute the queries in the database system (in SECONDO's executable language).

- **Q1.** At 8am on Monday, who sits in the bus No. 32?

This query deals with travelers who take the bus from a specific route at the given time. We do not deal with a particular bus or metro, but all available buses moving on the route at the query time.

- **Q2.** Who arrived by taxi at the university on Friday?
- **Q3.** Who entered bus No. 3 at bus stop "University" on Tuesday afternoon?

To answer the above two queries, one should find the place where people change transportation modes.

- **Q4.** Did anyone who was on floor H-5 of the office building between 2pm and 5pm take a bus to the train station on Friday?
- **Q5.** Did bus No. 35 pass by any bicycle traveler on Monday?

Q5 considers the distance between two moving objects with different transportation modes. Discovering such a relationship is interesting as the two objects move in different environments. The bus parameter means a group of buses which all belong to the route No. 35 but with different departure time, instead of a particular bus trip.

- **Q6.** Find out all people staying at room 154 in the university for more than one hour on Thursday.

Besides the outdoor trips, indoor moving objects can also be traced.

- **Q7.** Find the top k bus routes with high passenger flow for all workdays.
- **Q8.** Find the top k road segments with high traffic during the rush hour for all workdays.

By analyzing the histories of movements, routes with high passenger flow can be discovered and the schedule can be adjusted to improve the transportation. The traffic value of a road segment is set as the number of moving objects passing by during the rush hour ([7:00, 9:00] $\cup$ [16:00, 18:00]) where the following transportation modes are considered: $\{$*Car*, *Taxi*, *Bike*, *Bus*$\}$.

Optimization techniques are developed to improve the query efficiency. The involved transportation modes of a moving object are represented in a compact way with the aim of improving the mode existence examination, as the case of determining a given mode in a trip occurs frequently. An index is built on the trip units to accelerate the procedure of accessing the data.

Besides queries on moving objects, the system supports requesting data from infrastructures. Although infrastructure objects are located in different environments and represented by various data types, comparing them in the same context is still available.

- **Q9.** Given a building, find all bus stops that are within a radius of 300 meters.
- **Q10.** Which streets does Bus No. 12 pass by?

5. DEMONSTRATION

The demonstration is performed in an extensible database system SECONDO [5] which supports new technology to be embedded in a database system. A new module is developed in the system to manage generic moving objects. We are able to generate a comprehensive and scalable set of moving objects. In the demonstration, 200,000 generic moving objects are created to model people moving around inside a metropolitan area. We take the roads of Berlin [1] and some floor plans [2, 3] as input and create pavement areas, bus network, metro network and a set of buildings to simulate a city environment. The overall time period for moving objects

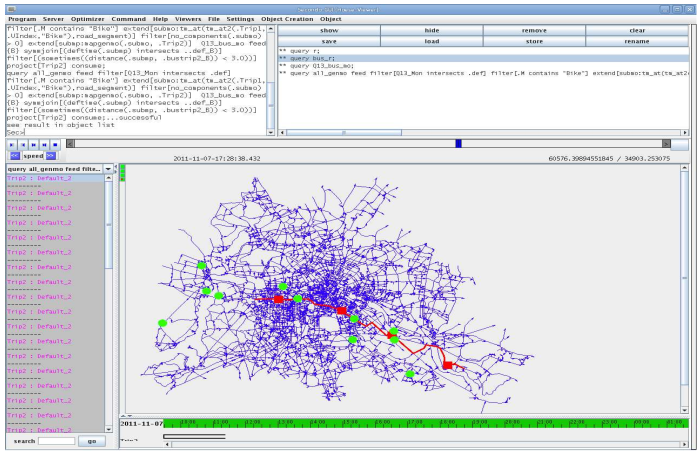
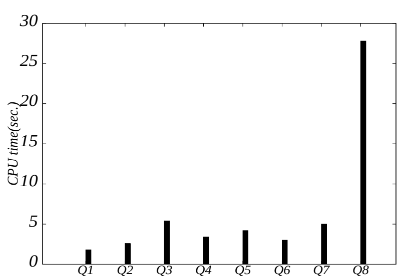


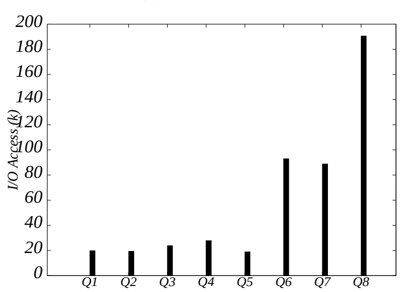
Figure 3: A Screenshot of Buses and Bike Travelers (Q5)

is one week. Each trip may contain several transportation modes, listed in Table 1.

We show how to access the data of generic moving objects by the proposed operators. Given a generic moving object, we can project its movement into the space where the trajectory is located in different environments. The precise location of each sub path is represented. We can restrict the movement to a certain transportation mode, e.g., an indoor trip. The movement according to a specific referenced GEOB can also be determined. For example, a sub trip by bus or the movement on a particular road. Additionally, the system is able to find a precise path where the start and end positions are located in different environments. Sub paths with different modes can be distinguished. We can return a path between two indoor locations inside different buildings, leading to both indoor and outdoor movement. For routing in a bus network, a short distance walking may be involved for transfer.



(a) CPU Time



(b) I/O Accesses

Figure 4: Query Cost

Figure 4 shows the query cost in terms of the CPU Time and I/O accesses for queries on moving objects. More queries are also available. The results demonstrate the efficiency of proposed methods where **Q8** takes more time than the others due to the costly procedure of mapping bus routes into road segments. We demonstrate the result of **Q5** at a time instant in Figure 3 showing buses (red rectangles) moving on the route and some bike travelers.

Acknowledgment

The first author gratefully acknowledges the financial support by Chinese Scholarship Council.

6. REFERENCES

[1] http://www.bbbike.de/cgi-bin/bbbike.cgi (2011).

[2] http://www.greenhosp.org/floor_plans.asp (2011).

[3] http://www.modulargenius.com/default.aspx (2011).

[4] J. Booth, P. Sistla, O. Wolfson, and I.F. Cruz. A data model for trip planning in multimodal transportation systems. In *EDBT*, pages 994–1005, 2009.

[5] R. H. Güting, V. Almedia, D. Ansorge, T. Behr, Z. Ding, T. Höse, F. Hoffmann, and M. Spiekermann. SECONDO: An extensible dbms platform for research prototyping and teaching. In *ICDE, Demo Paper*, pages 1115–1116, 2005.

[6] S. Reddy, M. Mun, J. Burke, D. Estrin, M. H. Hansen, and M.B. Srivastava. Using mobile phones to determine transportation modes. *TOSN*, 6(2), 2010.

[7] L. Stenneth, O. Wolfson, P. Yu, and B. Xu. Transportation mode detection using mobile devices and gis information. In *ACM SIGSPATIAL*, pages 54–63, 2011.

[8] J. Xu and R. H. Güting. Infrastructures for research on multimodal moving objects. In *MDM, Demo Paper*, pages 329–332, 2011.

[9] J. Xu and R. H. Güting. A generic data model for moving objects. Informatik-report 360, Fernuniversität Hagen, 2011. http://dna.fernuni-hagen.de/articles2010.html.

[10] J. Xu and R. H. Güting. MWGen: A Mini World Generator. In *MDM, To Appear*, 2012.

[11] Y. Zheng, Y. Chen, X. Xie, and W. Y. Ma. Understanding transportation mode based on gps data for web application. *ACM Transaction on the Web*, 4(1):1–36, 2010.

[12] Y. Zheng, L. Zhang, Z. Ma, X. Xie, and W. Y. Ma. Recommending friends and locations based on individual location history. *TWEB*, 5(1):5, 2011.