Xu JQ, Güting RH, Zheng Y *et al.* Moving objects with transportation modes: A survey. JOURNAL OF COMPUTER SCIENCE AND TECHNOLOGY 34(4): 709–726 July 2019. DOI 10.1007/s11390-019-1938-4

Moving Objects with Transportation Modes: A Survey

Jian-Qiu Xu[1], *Member, CCF*, Ralf Hartmut Güting[2], Yu Zheng[3], *Senior Member, CCF*, and Ouri Wolfson[4], *Fellow, ACM, IEEE*

[1] *College of Computer Science and Technology, Nanjing University of Aeronautics and Astronautics Nanjing 211106, China*

[2] *Faculty of Mathematics and Computer Science, University of Hagen, Hagen 58097, Germany*

[3] *JD Intelligent City Research, Beijing 100176, China*

[4] *Department of Computer Science, University of Illinois at Chicago, Chicago 60607, U.S.A.*

E-mail: jianqiu@nuaa.edu.cn; rhg@fernuni-hagen.de; msyuzheng@outlook.com; wolfson@uic.edu

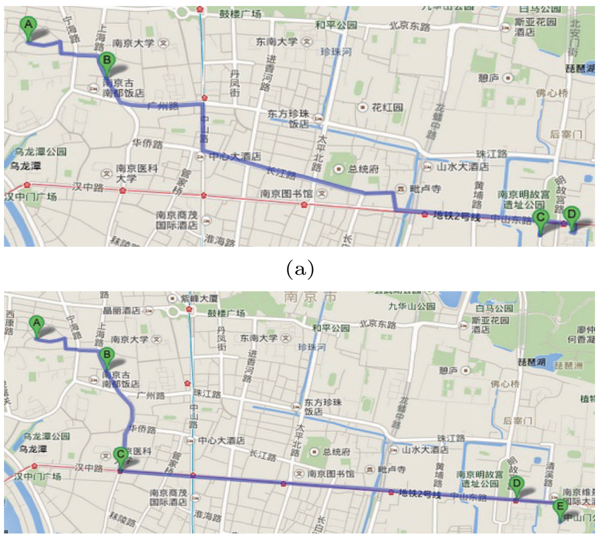
Received January 31, 2019; revised April 17, 2019.

Abstract In this article, we survey the main achievements of moving objects with transportation modes that span the past decade. As an important kind of human behavior, transportation modes reflect characteristic movement features and enrich the mobility with informative knowledge. We make explicit comparisons with closely related work that investigates moving objects by incorporating into location-dependent semantics and descriptive attributes. An exhaustive survey is offered by considering the following aspects: 1) modeling and representing mobility data with motion modes; 2) answering spatio-temporal queries with transportation modes; 3) query optimization techniques; 4) predicting transportation modes from sensor data, e.g., GPS-enabled devices. Several new and emergent issues concerning transportation modes are proposed for future research.

Keywords moving object, transportation mode, data model, performance, data generator

1 Introduction

Moving objects represent real-world objects that continuously change their locations over time[1,2], e.g., vehicles and humans. Such data have received substantial attention in both research and industry communities, and enabled a wide range of applications including traffic management and monitoring, tourist service, and mobile commerce. In addition to time-stamped locations, transportation modes reflect human behavior and play a pivotal role in understanding the mobility and contextual knowledge[3,4] and supporting advanced trip planning[5,6]. This is because time-stamped locations neither fully represent a person's state nor fully recognize the high-level intentions of complex behavior. Typically, a person's trip includes several transportation modes, e.g., *Walk*, *Car*, and *Bus*. To help understanding the issue, we provide two trips of Mike from his home to the office room, as illustrated in Fig.1.



(a)

(b)

Fig.1. Mike's trips with multiple modes. (a) O_1: $A \xrightarrow{Walk} B \xrightarrow{Car} C \xrightarrow{Walk} D$. (b) O_2: $A \xrightarrow{Walk} B \xrightarrow{Bus} C \xrightarrow{Metro} D \xrightarrow{Walk} E$.

Survey

Special Section on Spatio-Temporal Big Data Analytics

Jian-Qiu Xu is supported by the National Key Research and Development Program of China under Grant No. 2018YFB1003900 and the Fundamental Research Funds for the Central Universities of China under Grant No. NS2017073.

O_1: *Mike walks from home to the parking lot, and then drives the car to the university. After parking the car he walks to his office room.*

O_2: *Mike walks from home to a bus stop, then takes a bus to the underground train station, moves from C to D, and finally walks from the station to his office building.*

Such moving objects have two unique features. 1) *Multiple Environments*. The object moves over a range of urban areas each of which exhibits their own characters in terms of transportation infrastructures and positioning technology. 2) *A Sequence of Transportation Modes*. The value changes over time and the variation indicates the switch between different environments. We consider both outdoor and indoor movements. This imposes new challenges regarding the data management including data modeling and query processing, and the recognition of modes as they are not recorded by popular sensors.

1.1 Transportation Modes

Transportation modes provide a comprehensive picture for understanding humans' behavior. Such high-level activities capture a range of available places and build the connection between moving objects and the underlying geographical objects. The topic has received considerable attention among users and developers. Target applications include the data management of moving objects[7,8], multi-modal transportation management and traffic control[9,10], trajectory sharing and mobility data analysis[11], and physical activity and fitness monitoring[4,12], to name a few.

People's movement includes outdoor and indoor scenarios. In the former case, there are motorized modes like *Car* and *Bus* and non-motorized modes like *Walk* and *Bike*. In the latter case, people can walk, run, bike and be still, although a motorized vehicle may occur in some particular environments, e.g., airport. We summarize popular transportation modes in the current state of the art and provide a taxonomy in Table 1. Some methods do not distinguish vehicle modes *Car*, *Bus* and *Taxi* and use the term *Motor* in general[13], but some do[3]. In the literature, several words referring to the same mode are used, {*Metro*, *Subway*, *Underground Train*}, {*Bike*, *Bicycle*}, and {*Still*, *Stationary*, *Stop*, *Static*}. For the sake of consistency, we use *Metro*, *Bike* and *Still*. Notations of *Walk*, *Still* and *Run* mean outdoor modes, and are generalized as *Indoor* for people moving inside buildings. A few special modes such as Electric Vehicles are studied in particular applications[14], but are ignored in the paper.

Table 1. Summarization of Indoor and Outdoor Mode

Term	Transportation Mode
Non-motorized	*Indoor*
	Walk, Still, Run
	Bike
Motorized	*Car, Taxi, Bus*
Public	*Train, Metro, Bus*
Sharing	*Ride-sharing, Ride-hailing*

The study of Ride-Sharing[15−17] and Ride-Hailing[18,19] has attracted considerable attention in recent years. Several commercial service providers are well established, e.g., DiDi and Uber. *Ride-Sharing* refers to a transportation mode in between private car and public transportation in the sense that the ride is in a private car, but the route between the origin and the destination is indirect. For ride-hailing, there is a wait for pick-up, similar to public transportation. These modes will become increasingly important, particularly with the development of autonomous vehicles[20].

1.2 Literature Covered

Moving objects with transportation modes have been extensively studied in a variety of research communities including database[7,21], data mining[3], persuasive/ubiquitous computing[12,22], artificial intelligence[4,23,24], geographical information system[25,26], intelligent transportation systems[27], and mobile communication networks[13,28,29]. This survey summarizes recently developed techniques in those fields and classifies them into two categories: 1) database, and 2) data mining and machine learning. The two communities investigate mobility data with motion modes from different aspects.

The database community mainly focuses on modeling and representing the data to support advanced application queries such as spatio-temporal queries with transportation modes and multi-modal trip planning. Moving objects with transportation modes involve a range of environments which have different characteristics such as network space, obstructed area, and 3-D (three-dimensional) space. The management of continuously changing location data and transportation modes (environments) requires dedicated support from the underlying database system. Corresponding data models are proposed to capture the underlying environment in order to represent moving objects[7,21]. Novel queries and application analysis that involve transportation modes are studied, e.g., multi-modal trip planning[7,30], the switch between different modes[31], and range queries[8]. Query optimization techniques are proposed to improve the system performance as well. Some auxiliary tools are developed to enhance the system functionality, e.g., data simulation[32], 3D visualization and animation[33].

The communities of data mining, machine learning and geographical information system primarily aim at inferring transportation modes from low-level sensor data. Both coarse data such as GSM[34] and fine-grained data such as GPS[26,35,36] are used. Such data collection has become an important and convenient approach to investigating travel behavior due to the widespread use of mobile devices such as smartphones. To achieve the goal, a number of prediction models and classification systems are employed to recognize the travel modes for outdoor movements. The models typically consist of several widely-used baseline models such as decision-tree, random forest, SVM and Bayesian net. A set of sophisticated features is selected to train the models. Solely using GPS data is not sufficient to classify motorized transportation modes such as *Car* and *Bus* due to similar movement characteristics. Thus, multiple data sources are utilized to increase the prediction accuracy. Popular sources include GIS[5], accelerometer[25] and cellular data[37]. This line of work can be considered as a significant step in recognizing human activities.

We believe that techniques from those communities can be combined to build a comprehensive platform for managing and analyzing big mobility data, as illustrated in Fig.2. For instance, the outcome of predicting transportation modes can be used as the input data for a database system managing moving objects with transportation modes. Although the mode *Indoor* is not available from GPS data, most popular outdoor modes are covered. In this survey, we aim at offering a thorough and structured overview of the current state of the art, while how to build such a platform and fuse relevant techniques is beyond the scope of the paper.

The rest of the paper is organized as follows. We compare moving objects with transportation modes with several mostly relevant issues in Section 2. Data models are presented in Section 3. Query processing and optimization techniques are presented in Section 4 and Section 5, respectively. We introduce techniques developed to obtain moving objects with transportation modes in Section 6, followed by conclusions and future work in Section 7.

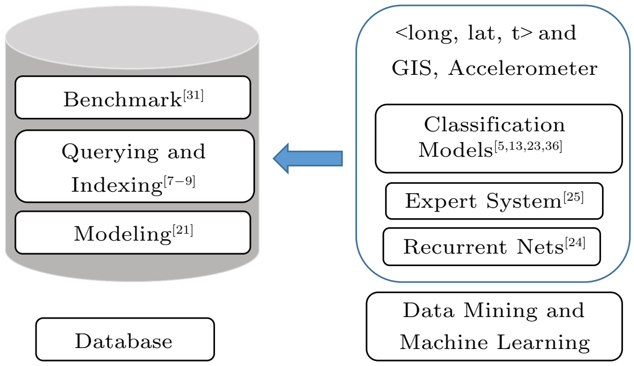


Fig.2. Connection between two communities. long: longitude; lat: latitude; t: time.

2 Semantic, Symbolic and Multi-Attribute Trajectories

Recently, extending the knowledge about mobility data has attracted considerable attention due to the fact that the location information is not sufficient to fully understand people's behavior and answer queries involving semantic data. Existing work can be classified into three categories: 1) semantic and activity trajectories, 2) symbolic trajectories, and 3) multi-attribute trajectories.

Semantic and Activity Trajectories[38−42]. A semantic trajectory enriches a spatio-temporal trajectory by attaching a semantic label to a location. Usually, the locations are points of interest at which people perform actions and activities such as *hotel* and *restaurant*. Using the semantic trajectory representation, Mike's trip O_2 in Fig.1 will be

$$\begin{aligned}&Semantic(O_2)\\ =&<(t_1, A, \text{"}home\text{"}), (t_2, B, \text{"}bus\ stop\text{"}),\\ &(t_3, C, \text{"}metro\ stop\text{"}), (t_4, D, \text{"}metro\ stop\text{"}),\\ &(t_5, E, \text{"}university\text{"})>.\end{aligned}$$

Such a trajectory enables queries and analytics considering semantic interests and location preferences. For example, users may search for relevant trajectories that pass B and D and contain keywords "*bus*" and "*metro*". Given such a trajectory, semantic locations are sparsely defined because among a person's trip only a few locations have semantics. In the example, semantic locations between A and B are not defined because the semantics is not known. Consequently, semantic trajectories mainly deal with ranked and top-k

queries but cannot support continuous queries. As a further step, a comprehensive knowledge graph for urban movement is constructed in which nodes and edges are represented in latent semantic space. One can apply the graph to predict the extent of user attention paid to different locations in a city[43]. In addition to extracting semantics from outdoor trajectories, motion trajectories of indoor scenarios are utilized to estimate and iteratively refine the underlying route network[44]. The method assumes that the underlying route network is unknown because in many cases digital models are not given or easily obtained. Outlier movement detection, self-healing analysis and trajectory cleaning are supported.

Symbolic Trajectories[45,46]. A symbolic trajectory is represented as a sequence of pairs (t, l), in which t is a time interval and l is a label (short character string) describing certain aspects of a trajectory. The symbolic information is computed from the movement itself or obtained from the geographical environment. Typical examples include names of roads, activities and transportation modes. The goal is to provide a simple and flexible model for any kind of semantic information, while geometric locations are not defined. If transportation modes are considered, the symbolic trajectory for Mike's trip O_2 is denoted by

$$\begin{aligned}&Symbolic(O_2)\\=\ &<([t_1,t_2], Walk), ([t_2,t_3], Bus), ([t_3,t_4], Metro),\\&([t_4,t_5], Walk), ([t_5,t_6], Indoor)>.\end{aligned}$$

Multi-Attribute Trajectories[47,48]. In practice, objects or entities are naturally of multiple attributes in addition to spatial and temporal aspects, amenable to diverse types of analysis. This kind of trajectories combines spatio-temporal trajectories and characteristic attributes. That is, a multi-attribute trajectory consists of a sequence of time-stamped locations and a set of attributes. The attribute domain depends on real applications. For example, the management of urban vehicle trajectories defines two attributes COLOR = {SILVER, RED} and BRAND = {VW, BMW, BENZ}. Consider Mike's trip O_1. Assuming that he drives a SILVER VW from the home to the university, the multi-attribute trajectory is

$$\begin{aligned}&MultiAtt(O_1)\\=\ &(<(t_1, loc_1), ..., (t_n, loc_n)>, (SILVER, VW)),\end{aligned}$$

in which the spatio-temporal trajectory and the attribute model are integrated into one framework. An important point is that location-independent attributes are primarily tackled in order to represent the object from diverse aspects. This allows users to issue queries containing both spatio-temporal and attribute constraints, e.g., "did any (SILVER, VW) pass the university between [7am, 8am]?" A time-dependent attribute could be defined to represent transportation modes, but the solution processes transportation modes and locations separately. That is, we are aware of the modes, but do not know the place at which the mode occurs or changes.

The three kinds of trajectories above follow the same direction as moving objects with transportation modes that explore spatio-temporal trajectories with additional information, but there are some significant differences between them. 1) Semantic trajectories add semantic data by attaching labels to locations, which are points of interest in general. A data pre-processing procedure is involved to enrich trajectories with semantic annotations. Transportation modes are not supported because such data are related to pieces of movements rather than individual locations. 2) Symbolic trajectories deal with generic semantic information including transportation modes and users' activities, but geographical locations are not defined. 3) Multi-attribute trajectories aim at providing a full picture of moving objects by considering a range of aspects, in particular, location-independent data. This is orthogonal to moving objects with transportation modes because motion modes are related to locations.

3 Data Models

3.1 Moving Objects in a Single Environment

Researchers start investigating moving objects in free space in which there is no limitation about the movement. In practice, most outdoor and indoor motions are constrained due to the fixed underlying structure such as the presence of obstacles and floor plans. In the literature, there has been a large body of work on modeling and querying moving objects, which can be classified into four categories according to the underlying environment, 1) free space[1,49–51], 2) road (spatial) network[52–56], 3) obstacle space[57–60], and 4) indoor[61–65].

Free space is an environment in which objects move without any constraint. Practically, objects usually move on a pre-defined set of paths as specified by the underlying environment, e.g., roads and highways. The obstacle space defines an area that inhibits the movement connected by a straight line as obstacles block the connection. In comparison with road network, there is no pre-defined path in the obstacle space but the path between two points should take into account obstacles. The main difference between free space and constrained environments lies in location representation and distance computation. The coordinate representation (i.e., longitude and latitude) is typically used in free space but is not an optimal choice for constrained environments. This is because the method does not provide the referenced target by the moving object, e.g., the road/street. To solve the problem, locations are represented by first referencing to a geographical object and then recording the relative location according to that object. Regarding the distance computation, two arbitrary points are directly reachable and connected by a straight line in free space. However, such a line may not exist in a constraint space due to the underlying network structure or be blocked by an obstructed area. A shortest path query is performed to determine the distance in a constrained space, which is a costly procedure.

In addition to outdoor space, indoor data management has attracted considerable attention in the last decade. Such an environment has two unique features.

1) *Space Constraint.* This is characterized by entities such as rooms, hallways and staircases that enable and constrain the movement. Indoor movement is surely not within free space but is less constrained than the movement in a spatial network. Indoor travels are bound by a building infrastructure, typically incurring a short distance and time period compared with outdoor travels.

2) *Positioning Technology.* GPS signal is not reliably available in indoor settings. Indoor moving objects are typically monitored by proximity-based indoor positioning technologies such as Wi-Fi, RFID and bluetooth. Symbolic models are often used for indoor movement[66,67].

Public transportation network is an environment in which pre-defined paths are offered for moving objects. The movement is not only limited to fixed routes but also under a specified schedule. Furthermore, passengers can only start and end their trips at stops and stations. There are some special features, e.g., the waiting time and the number of transfers. GTFS (General Transit Feed Specification)① defines a format for public transportation schedules and associated geographic information. Each file records a list of items with the format (*name*, *value*). Locations of a bus traveler are a set of items recording bus stops as well as arrival and departure time at each stop. People can use the specification to provide schedules and geographic information to Google Maps.

Data models for a single environment are not appropriate for moving objects with transportation modes. On the one hand, they do not consider a range of available environments such that the system is only aware of a sub-trip instead of the complete movement. The relationship/interaction between different environments is not handled, for example, the places at which people switch transportation modes such as bus stops and building entrances. On the other hand, data models above propose different techniques for location representation. Specifically, moving objects in a road network are represented by referencing to a road, and indoor trips are represented by referencing to an office room or using the symbolic method. In order to manage the complete trip like O_1 or O_2, the system needs to maintain several pieces of trips each of which has a particular model only for one environment. This incurs much system overhead and significantly complicates the query processing when multiple modes are involved. As a consequence, a robust and general location representation is essentially needed that can be applied in all cases. Meanwhile, the method should be consistent with the well-established work for an individual environment in order to provide a systematic solution.

3.2 Modeling Multiple Environments

The urban transportation is a complex system involving a number of components such as infrastructures, networks, and schedules. There are static objects such as routes and stops, and also dynamic objects such as vehicles and traffic condition. To fully represent the system, a multi-dimensional data model is required to consider a range of factors including spatial and temporal domains, topology and relationship, granularity and hierarchy. Consensus-based functional requirements for multi-dimensional transportation data management are presented[68]. The fundamental issue is to develop a comprehensive transportation location reference system such that objects are represented as they occur in the real world including 1-D, 2-D and 3-D models. The data model includes a number of components for managing multi-dimensional data, e.g., navigation, multi-dimensional location referencing and multi-scale representations. A multi-dimensional data model is pro-

①http://code.google.com/transit/spec/transit_feed_specification.html, Oct. 2011.

posed to provide a foundation for capturing and querying complex transportation infrastructures[69]. The model captures important transportation infrastructure concepts such as roads, road parts, lanes and the relationships among them. Each individual lane is captured due to different road characteristics, and the relationship containment is captured among segments at different levels. Different contents are attached to specific points and road sections, e.g., traffic accidents, gas stations, and speed limits. Later, Jensen *et al.*[70] extended the representation dimension by introducing three new relations on dimension values to capture direction, traffic exchange and lane change relationships between road segments. Some properties of the three relations are defined such as transitivity and propagation of direction. Each representation of the transportation infrastructure is modeled as a separate dimension hierarchy. Two categories of queries are considered: transportation infrastructure and dynamic. However, those techniques focus on modeling and representing transportation infrastructures, but do not involve moving objects.

A conceptual data model presented in [7] integrates moving objects databases and graph-based databases to support trip planning with several transportation modes in urban transportation networks, e.g., *Bus* → *Walk* → *Train*. Each vertex in the graph corresponds to a place in a transportation network, which contains a name and a geometric object, e.g., point or region. Each edge is associated with a transportation mode. Edges with different modes can be incident on the same vertex indicating that a transfer between different modes can happen. The model supports returning the shortest path with multiple transportation modes as well as constraints and choices, e.g., different motion modes, and the number of transfers. Such a higher-order model enables the creation of new computing services that respond to customized requirement and support accurate predictions about future behavior.

To model moving objects across all available environments, a comprehensive location representation system should encompass all available elements and abstractions for transportation-based objects. Employing the referencing method, a generic data model is designed to manage moving objects in outdoor and indoor environments[21]. The idea is to conceptually partition the space into a set of so-called infrastructures, each of which 1) corresponds to an environment with particular transportation modes and 2) consists of a set of geographical objects defining available places for moving objects. Five infrastructures are included in total, {Free Space, Road Network, Public Transportation Network, Region-Based Outdoor and Indoor}, as illustrated in Fig.3. Public Transportation Network includes bus and metro networks in which bus and metro routes are geographical objects, Region-Based Outdoor defines the areas for pedestrians, and Indoor is a set of public buildings such as hotels, office buildings, and hospitals. Private buildings are discarded because usually simple movements occur and the dataset in this setting is rather difficult to obtain due to the privacy issue. The reference model has the advantage that a range of geographical objects are incorporated including 2-D objects (roads, regions), 3-D objects (buildings) and dynamic objects (buses).

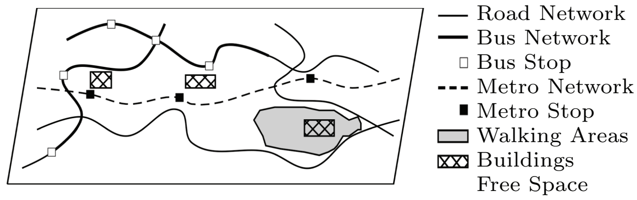
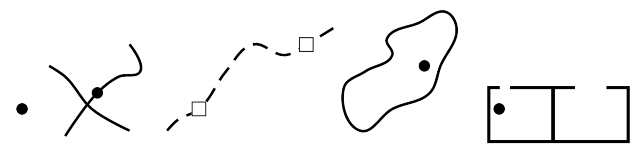


Fig.3. Partitioning the space into infrastructures.

A full list of data types is provided to represent geographical objects, each of which consists of a unique ID, a label for the data type and the value. All data types are embedded into a relational interface in order to exchange the information in a consistent fashion. The location model first maps the location of a moving object to a geographical object and then records the relative position according to that object, as illustrated in Fig.4. Moving objects with transportation modes are represented by a sequence of movements, each of which corresponds to a sub-trip with one mode. Both precise and approximate locations are supported in order to provide a flexible representation. A group of operators is designed to manipulate the data to answer queries with transportation modes.



(a) (b) (c) (d) (e)

Fig.4. Referenced location representation. (a) Free space. (b) Road network. (c) Bus network. (d) Pavement area. (e) Office rooms.

A unified pseudograph model of outdoor and indoor spaces is proposed[71]. Two essential elements are captured, topology and dynamics, in which the former defines geometric properties and the latter represents the changes in motion. The model is for receptor-based systems, e.g., RFID readers or wireless sensor networks.

Motivated by the fact that getting a taxi in highly congested areas such as airports and railway stations is time-consuming and expensive, a model of multi-modal ride-sharing is designed for public transportation hubs[72]. The method combines in a unique way three mechanisms: virtual queues, walking for the purpose of ride-sharing, and multiple-drop-off ride-sharing. A trip is defined as a triplet including destination, the number of travelers, and constraints (maximum walking time and maximum delay). Two modes are included, *Walk* and *Car*.

3.3 Dynamic Transportation Networks

Transportation network attributes like the travel time are associated with particular time instances due to dynamic conditions such as rush hour, road construction, accidents and events[73,74]. Ding and Güting[75] proposed a data model for moving objects on dynamic transportation networks. A dynamic graph is defined by associating a temporal attribute to each edge or node to express traffic jams and blockages caused by temporary constructions and the insertion and deletion of junctions or routes. Under these circumstances, finding an optimal route is a critical problem for vehicle navigation because both the traffic condition and the location of the traveler change over time. The traveling route may be modified from time to time. A novel dynamic shortest path algorithm is developed to compute the dynamic shortest path between a moving object and the destination in which traffic conditions are updated in real time[76]. The methods[77,78] integrate time dependent traffic into the model to provide users more accurate travel time and better sequenced routes. A standard relational DBMS is employed such that the tables, joins and sorting algorithms can be leveraged. The task is to model the dynamic transportation network for trip planning services instead of representing moving objects with different modes.

4 Query Workload

4.1 Spatio-Temporal Queries with Transportation Mode Constraints

We are primarily interested in queries containing spatio-temporal parameters and transportation modes as solely investigating modes is a simple task. One can ask queries where and when the mode changes or extract a sub-trip according to the mode as the followings.

*Q*1. "Where and when did Mike switch from *Bus* to *Metro*?"

*Q*2. "How long and where does Mike walk?"

To answer the queries, the database system should manage time-stamped locations as well as transportation modes and determine the relationship between moving objects and referenced objects. A list of representative queries is proposed and formulated by an SQL-like language[21]. GMOBench[31], a benchmark for moving objects with multiple transportation modes, is developed to evaluate and optimize a prototype database system. The benchmark targets measuring the cost of common operator constellations and access patterns. Queries are classified into two categories: four infrastructure queries and 17 trajectory queries, in which the former deals with geographical objects such as "which streets does Bus No. 12 pass by" and the latter deals with moving objects such as "who arrived by taxi at the university on Friday". In comparison with spatio-temporal queries, a wide range of geographical objects and the relationships between different environments and modes characterize the queries (see the following examples).

*Q*3. "How many people take the same bus as Mike?"

*Q*4. "Did Mike spend more than 15 minutes on waiting for the bus?"

Complementary to the benchmark workload, three types of range queries are investigated to return trips that intersect a spatio-temporal window and contain specified transportation modes[8]. For example, one can search for trips that intersect the city center and contain modes $\{Bus, Metro\}$. Transportation modes may follow a particular order like $Bus \rightarrow Metro$.

4.2 Trip Planning with Modes

In a public transportation system, the journey planning typically involves *Bus*, *Metro* and *Walk*. The service aims at providing the itinerary between an origin-destination pair that optimizes a range of costs such as the travel time, the number of transfers, and the walking/waiting time. Route planning in such a system considers spatio-temporal constraints such as the time required to wait for a bus at a station, or the feasibility of transferring from one bus to another given their respective schedules.

Graph models are widely used to represent the transportation system. A data model is proposed to conceptually and abstractly provide a multi-modal trip[7]. The graph model connects places in a transportation network. The proposed framework is able to return the shortest path connecting the origin and the destination with mode constraints and choices, e.g., less than two bus transfers. A labeling approach is proposed to model a public transportation network as a timetable graph, in which each node represents a station and each directed edge is associated with a timetable recording the departure time of each vehicle at the station[79]. Although an efficient method is developed, the graph is built on a single mode transportation network without supporting the switch between different services (e.g., bus, subways).

A dynamic programming-based algorithm is developed to solve the itinerary planning problem[9]. The method well formulates the issue as a shortest path problem with time windows on a multi-modal time-schedule network and optimizes a set of criteria such as total travel time, the number of transfers, and total walking and waiting time. An interesting operator called isochrones is studied in multi-modal and schedule-based transport networks[80]. The goal is to find the set of points on a road network, from which a specific point of interest can be reached within a given time span including transportation modes *Walk* and *Bus*. In an intelligent transportation system, the multi-modal interconnectivity between different transportation modes represents the quality of offering transportation mode options to users. For example, the amount of waiting time spent on changing transportation modes (e.g., *Bus*$\rightarrow$ *Walk* $\rightarrow$ *Metro*) is used to evaluate how well the interconnectivity is. Optimization methods are proposed to reduce the waiting time[30].

Bike is usually a convenient option for the first and last miles because bus and metro stations are often outside the walkable range. Tang *et al.*[81] considered a multi-modal public transportation system including shared bicycles for the first and last miles, and optimized the size of bicycle pools. The method guarantees bicycle availability with a high probability and reduces the number of bicycles per customer from 2 to 1.25. A flexible mini-shuttle like transportation system is developed to provide a transportation mode between *Bus* and *Taxi* because buses and metros are slow due to many stops, and taxies cause greater traffic congestion and pollution[82]. Effective routing algorithms are designed based on mining combinable trips from taxi GPS trajectories.

Ride-sharing is an environmental-friendly mode of commute that a group of travelers with similar itineraries and time schedules share a vehicle for their trips. This reduces the travel cost including fuel, tolls, and parking fees. Dynamic sharing for a large number of taxies is studied in order to quickly retrieve candidate taxies satisfying a user query. An experimental platform is built to produce taxi ride queries conforming to the real query distribution and generate ride-sharing schedules reducing the total travel distance[83]. To offer travelers multiple options, a dynamic ride-sharing solution is proposed and considers both the pick-up time and the price[84].

An open source software called OpenTripPlanner is published to provide passenger information and transportation network analysis services②. The core component finds itineraries combining transit, pedestrian, bicycle, and car segments through transportation networks built from OpenStreetMap③ and GTFS data.

5 Query Optimization

5.1 Index Structures and Algorithms

A number of spatio-temporal indexes are proposed to well manage time-stamped locations for historical data and on-line updating. However, they are not sufficient for answering queries with transportation modes because one cannot use the index to prune the search space on modes. The performance significantly deteriorates for large datasets as one needs to iteratively evaluate the trajectory. An index structure called TM-RTree (Transportation Mode R-Tree) is developed to manage spatial, temporal data as well as transportation modes[8]. The structure, based on a 3-D R-tree, is built on trajectories each of which contains only one transportation mode. For example, the trip O_1 will be decomposed into pieces of movements, as shown in Fig.5. All possible transportation modes are collected including two forms: a single mode and a pair of modes. In particular, a pair of modes are of the form $X \rightarrow Walk$ or $Walk \rightarrow X$. This is based on the observation that the mode switch often includes *Walk*. An integer is integrated into the tree node to indicate transportation modes contained by trajectories in the sub-tree. The TM-RTree is primarily used for answering spatio-temporal range queries with transportation modes.

Application queries are interested in not only searching for trips with particular modes but also determining the relationship between moving objects and referenced geographical objects. For example, to an-

②http://www.opentripplanner.org/, Jan. 2019.

③http://www.openstreetmap.org/, Aug. 2018.

swer the query Q_3 in Subsection 4.1, the system requires to identify all travelers taking the same bus as Mike. The TM-RTree is able to find all bus trips but cannot establish which bus for the trip. To solve the problem, a two-level structure called Mode-RTree is proposed[31]. The upper level contains a list of pairs with the form (*mode*, *record*), in which the root node of an extended 3-D R-tree is maintained for each mode. The lower level consists of a set of extended 3-D R-trees by integrating an integer into each node to manage referenced objects for moving objects such as roads, pavements, buses and rooms. Mode-RTree takes in a set of trips with a single mode and inserts each trip into a sub-tree according to the transportation mode. Fig.6 depicts the two structures.

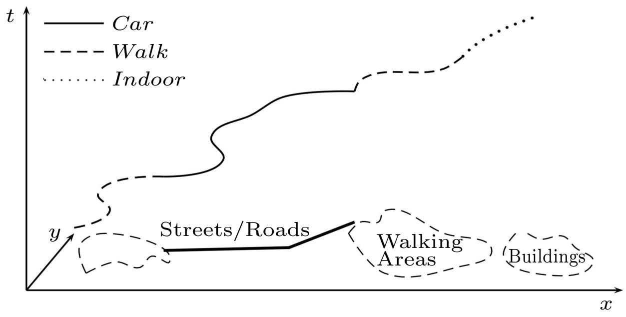


Fig.5. Partitioning O_1 according to the mode.

Corresponding query algorithms are developed over TM-RTree and Mode-RTree. Employing TM-RTree, three algorithms are proposed for queries containing a single mode, multiple modes and a sequence of modes. In particular, the algorithm dealing with a sequence of modes combines two modes as a pair to check the mode existence. This significantly improves the pruning ability in comparison with checking an individual mode. Algorithms running over Mode-RTree first determine the sub-tree according to the mode and then access the structure to collect qualified trajectories. If the query requests multiple modes, an intersection is performed on candidates received from sub-trees.

5.2 Prototype System

A prototype database system is developed to manage moving objects over a range of real-world environments[85]. The implementation is based on an extensible database system SECONDO[86] by incorporating into a number of modules including data storage and representation, operators, data generators, index structures and algorithms. Regarding the indoor environment, the query interface supports 3-D visualization of floor plans and animation of indoor movements, and is later extended to support displaying outdoor movements and the R-tree structure in a 3-D viewer[33]. When complex queries are executed, e.g., *joins*, the I/O communication usually becomes a bottleneck. A tool is developed to monitor database files at execution time[87], enabling us to better understand the query progress and perform analysis on the system.

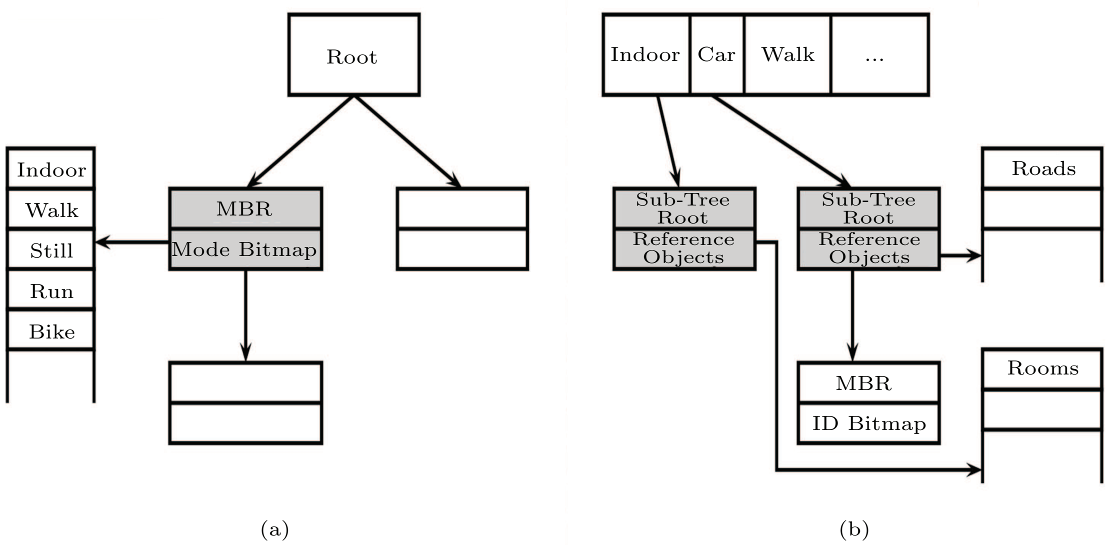


Fig.6. Comparison between two structures. (a) TM-RTree. (b) Mode-RTree.

6 Trips with Motion Modes

Large real datasets of moving objects with transportation modes are difficult to obtain due to too much effort from humans because people have to explicitly attach labels to mark transportation modes. Users will be easily bored if making additional data-labeling jobs. Furthermore, only the mode information is not sufficient for most applications as the underlying geographical objects are to be established as well. To solve the problem, there are two alternative solutions. One is to infer transportation modes from sensor data such as GPS and GSM, and the other is to develop data generators.

6.1 Transportation Modes Detection

An important issue in ubiquitous computing and GIS applications is to build rich models to predict human behavior from low-level sensor data[35]. For example, a healthcare assistant informs the user if he/she runs for too long and estimates the number of burnt calories by analyzing transportation modes from the travel diary[4]. Many artificial intelligence techniques have been recently applied to detect transportation modes. They attempt to extract good features from various sensor data and build powerful classifiers based on machine learning models such as SVM, Hidden Markov Models and Decision Tree[27]. GPS is the primary sensor data, but other sensors are utilized to increase the accuracy. We summarize the methods of predicting outdoor transportation modes in Table 2.

Some preliminary work is done to learn a unified model of transportation modes in an unsupervised manner[35,88]. The method[88] not only infers a user's mode of transportation but also predicts when and where he/she will change the mode. However, few transportation modes are established, the experimental dataset is limited (GPS logs from one person are used), and the data quality is not high.

Zheng *et al.*[3,22,23] developed a systematic method based on supervised learning to automatically infer users' outdoor transportation modes including *Car*, *Walk*, *Bus* and *Bike*. The goal is to provide more contextual information and enrich a user's mobility with informative knowledge because raw GPS logs are limited in understanding users' mobility. The method includes off-line learning and on-line reference. We illustrate the off-line learning in Fig.7. This phase 1) partitions trajectories into segments by utilizing some commonsense knowledge, e.g., *Walk* is a transition between different modes, and 2) extracts features such as distance, average velocity and heading change rate. An

Table 2. Summary of Approaches for Predicting Outdoor Transportation Modes

Sensor	Prediction Method (Ground Truth Comparison and Accuracy Calculation)	Accuracy (%)	Transportation Mode
GPS	Bayesian model[35] (hand-labeled modes, cross-validation)	84	{*Car*, *Walk*, *Bus*}
	Hierarchical Markov model[88] (activities in the historical data)	84	{*Car*, *Walk*, *Bus*}
	Supervised learning[3] (visualize traces on a map)	71	{*Car*, *Walk*, *Bus*, *Bike*}
	Tree-based ensemble classifiers[26] (labeled modes, K-fold cross-validation)	91	{*Car*, *Walk*, *Bike*, *Metro*, *Train*}
	Random Forest classifier[36] (particular rules)	93	{*Car*, *Walk*, *Bike*, *Metro*, *Bus*}
GSM	Statistical classification and boosting[34] (custom diary application)	85	{*Car*, *Walk*, *Still*}
GPS+ accelerometer	Classification system[13] (10-fold cross validation)	94	{*Motor*, *Walk*, *Bike*, *Still*, *Run*}
	Recurrent nets[24] (signal logs containing modes)	93	{*Car*, *Walk*, *Bus*, *Bike*, *Train*, *Tram*, *Subway*}
	HMM[89] (annotate transport modes)	76	{*Car*, *Walk*, *Bus*, *Bike*, *Train*, *Tram*, *Metro*, *Motocycle*}
GPS+GIS	Transportation network + classification system[5] (labeled modes, web application)	94	{*Car*, *Walk*, *Still*, *Bus*, *Bike*, *Train*}
	ArcGIS[90] (GPS logger, travel diary)	83	{*Car*, *Walk*, *Bus*, *Metro*, *Tram*}
	Expert system + OpenStreetMap[25] (manual classification)	92	{*Car*, *Walk*, *Bus*, *Bike*, *Train*, *Tram*, *Metro*, *Ferry*, *Boat*, *Aircraft*}

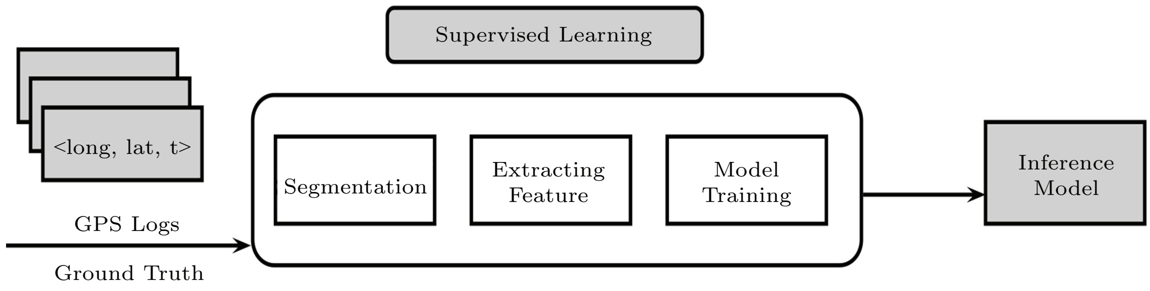


Fig.7. Predicting transportation modes.

inference model is built at the off-line learning. At the on-line phase, the model probabilistically predicts the transportation mode of each trajectory segment and a post-processing algorithm improves the inference accuracy. The mode with the maximum probability is used.

A statistical method is deployed to generate global features and extract several local features by employing tree-based ensemble models such as Random Forest and XGBoost, among which XGBoost achieves the best performance[26]. A random forest classifier combined with a rule-based method is developed to detect transportation modes[36]. Seven GPS-related variables are selected as the feature set. The rule-based method, which is more appropriate than the advanced classifier in the scenario without GPS signal and with incomplete GPS signal, is able to identify subway trips at a high accuracy (up to 98%).

Accelerometer-based techniques are widely used in conjunction with GPS data for transportation mode detection on mobile phones due to the advantages of low power consumption, directly measuring users' movements and highly detailed information. A classification system is built to determine the transportation mode of an individual when it is outside by making use of a mobile phone with a built-in GPS receiver and an accelerometer[13]. The overall classification system consists of a decision tree followed by a first-order discrete hidden Markov model. The method does not distinguish between various modes under motorized transportation, such as *Car* versus *Bus*. Transportation mode detection is hierarchically decomposed into subtasks based on a novel set of accelerometer features[28]. Coarse-grained GSM data from mobile phones are used to recognize high-level properties of user mobility[34]. Statistical classification and boosting techniques are employed, but a few transportation modes are distinguished (*Walk*, *Car*, *Still*) and the accuracy is not high. An algorithm is developed to estimate the gravity component of the accelerometer measurements, and novel accelerometer features such as spectrum peak position and stationary duration are extracted to capture key characteristics of vehicular movement patterns. A trip analysis system is developed based on smartphones and mobile apps to identify both the travel mode and the purpose (e.g., home-based work, home-based shopping)[11].

Distinguishing between motorized and non-motorized modes is not difficult, but the problem becomes challenging when classifying modes like *Car*, *Bus* and *Train*. They have similar GPS-related features and accelerometer readings. To increase the accuracy, GIS data are utilized to create discriminative features. The knowledge of the underlying transportation network such as bus stop locations and railway lines is considered to distinguish between motorized modes[5]. Novel features related to transportation network information are identified and derived to improve classification effectiveness, e.g., average bus closeness, and average rail closeness. The proposed approach treats above-ground train as a transportation mode and achieves the accuracy up to 93.5%. To build a multi-modal transportation network for mode detection, a number of GIS layers are cleaned and edited such as streets, bus routes and stops, subway lines and stations[90]. A deep learning model is built to work directly with raw signals from an embedded accelerometer. Different types of recurrent neural networks are used including the typical recurrent net and the two-layer versions[24]. Relying on fuzzy concepts found in expert systems and OpenStreetMap data, up to 10 transportation modes are distinguished and classified into three categories {*Land*, *Water*, *Air*}. The method handles data with signal shortages and noise[25]. The cellular data of mobile phone service providers are utilized to know public transportation modes and crowd density estimation[37].

Mobile devices involve continuous sensing of GPS or acceleration modules to infer transportation modes.

This decreases the battery lifetime. Using cellular network information as a priori knowledge, a battery-efficient method is developed to minimize power consumption and maximize detection accuracy[29]. A low-power classifier is designed to address the requirement of collecting more training data and achieving low computational complexity[4]. The power consumption is drastically reduced by 99%, while competitive mode-detection accuracy is maintained and the complexity is independent of the data size.

Indoor detection differs drastically from the outdoor scenario due to the physical environment as well as the availability and reliability of sensing resources. Indoor route networks are small in comparison with road networks, resulting in a short traveling distance and time interval. GPS is a popular data source for outdoor mode detection but is not reliably available in indoor settings. Indoor transportation scenarios are explored to achieve a conceptual model by utilizing Wi-Fi and accelerometer data collected through smartphones in a hospital[14]. Detected indoor modes include general modes {*Still*, *Bike*, *Walk*} and hospital-specific electric vehicles {*e-bus*, *bedpusher*}. Extracted features from the sensor data can be applied for both indoor and outdoor scenarios such as signal strength based features and position based features. Popular techniques for indoor mode detection include smart phone sensors[91], depth camera models[92], and Wi-Fi fingerprints with user information[93].

6.2 Data Generators

Large real datasets are often hard to come by as most people do not want to publish their movements, in particular, time-stamped locations with transportation modes. There are several published datasets, e.g., GeoLife④, NYC⑤, and DiDi⑥. These are taxi trips in general and not comprehensive enough to perform queries with transportation modes and evaluate the system in consideration. Although machine learning techniques are able to accurately infer most outdoor transportation modes, particular geographical objects do not receive adequate attention, that is, at which road the vehicle is located, which bus the passenger takes and in which area the pedestrian walks. These objects are essentially important when representing the location of a moving object using the referencing method. Map matching is a procedure that estimates the route traveled by vehicles or people by using observed coordinates[94,95], but does not obtain walking areas and moving buses. Furthermore, one needs all available environments in a consistent space. That is, road networks, public transportation systems, pavement areas and buildings are within the same city. This motivates researchers to develop data generators to produce datasets with variable sizes in a simulated scenario.

A mini-world generator (MWGen) is developed to build a range of infrastructures and generate moving objects with both indoor and outdoor transportation modes[32]. The data generator works in a two-step process. At the first step, the tool takes in a set of roads and public floor plans such as library and hotel, and defines some parameters to build available environments for moving objects. At the second step, based on popular movement rules, the system performs trip planning across different environments to connect origins and destinations and generates moving objects. The workflow of MWGen is illustrated in Fig.8. The tool is able to simulate popular human movements and scale the data size for performance evaluation. In addition, 3-D visualization and animation of indoor moving objects are supported and the shortest path query in an obstructed space can be answered.

Traffic simulators have been extensively studied in the literature[96,97]. In particular, public transportation systems play an essential role in the process of urbanization. A spatial interaction coverage model is used to model the relationship between demand points and bus stops and remove redundant bus stops[98]. The model considers the attractiveness of stops and the distance decay. Such an ability provides better bus routes for the transportation agency. SMARTS (Scalable Microscopic Adaptive Road Traffic Simulator)[99] is built with a distributed architecture for fast large-scale simulations. The route of public transportation (e.g., buses and trams) can be imported by reading route information from OpenStreetMap data.

There are also data simulators developed to create moving objects in a single environment, mainly including free space, road network and indoor.

Free Space. GSTD[100], a widely used spatio-temporal generator, defines a set of parameters to control the generated trajectories. The generator is later

④http://research.microsoft.com/en-us/projects/geolife, Jan. 2019.
⑤http://www.nyc.gov/html/tlc/html/technology/data.shtml, Jan. 2019.
⑥https://outreach.didichuxing.com/app-vue/personal, Dec. 2018.

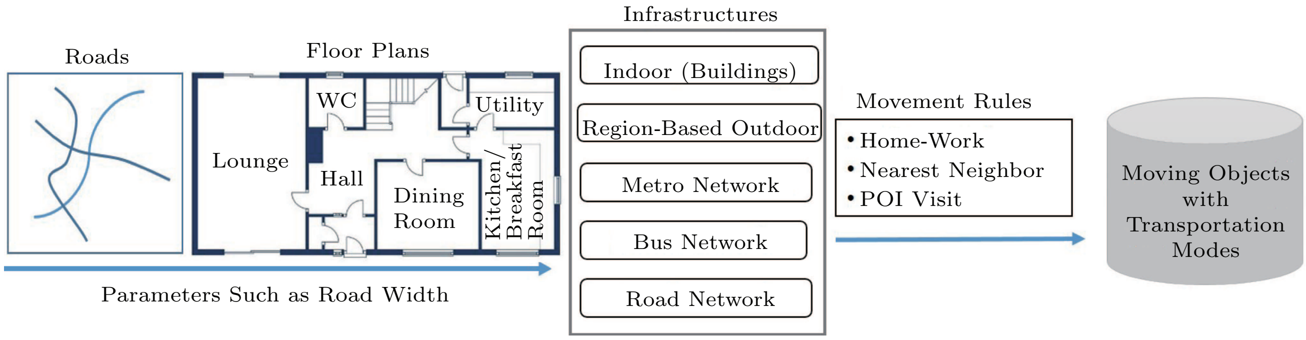


Fig.8. Outline of MWGen.

extended to produce more realistic moving behaviors such as group movement and obstructed movement[101]. Two methods are used for setting positions and velocities for moving objects generation[102], uniform distribution and skewed distribution. BerlinMOD[103] provides a benchmark by generating moving objects based on trip planning, e.g., from home regions to work regions. Users can set a range of parameters to simulate a person's everyday trips including the number of observation days and the way of selecting the initial and final positions.

Road Network. Brinkhoff is the earliest attempt to generate moving objects over a road network[104,105]. To simulate the traffic scenario, objects are created in a random way and appear and disappear when their destinations are reached. SUMO[106] is an open source generator that simulates the movements of private and public transports. Vehicle movements are represented in a network that does not allow conflicts. An important feature is that the proper interaction between moving objects is considered. GAMM[107] generates cellular network trajectories and symbolic trajectories to simulate real-life mobility patterns and constraints. The tool Hermoupolis[108] takes a road network, points of interest and mobility patterns as input to generate trajectories conforming to the mobility pattern requirements. An extensible web-based road network traffic generator is developed to produce traffic data at any arbitrary road network[109]. The tool is shipped with different traffic generators including Brinkhoff and BerlinMOD as well as various road network sources such as U.S.A. Tiger files⑦ and OpenStreetMap.

Indoor. To generate indoor moving objects, a floor plan, showing the relationships between rooms, spaces, and other physical features at one level of a structure, plays an essential role. The SLAM-like (simultaneous localization and mapping) approaches are proposed for floor plan reconstruction through observations of devices carried by humans[91]. Yang *et al.* created indoor moving objects based on floor plans and some pre-defined movement rules, e.g., an object in a room can move to the hallway[66,67]. IndoorSTG generates semantic-based trajectories in a simulated indoor environment including rooms, doors, corridors, stairs, elevators and virtual position devices[110]. A toolkit named Vita is developed to generate indoor mobility data for real-world buildings[111]. The tool produces the desired data in a three-layer pipeline: infrastructure, moving object and position. The moving object layer offers the functionality of defining objects or trajectories, with configurable indoor moving patterns, distribution models, and sampling frequencies.

7 Conclusions

The article summarized a number of research results related to moving objects with transportation modes. The covered topics mainly include modeling and representing the data, answering spatio-temporal queries with motion modes, and predicting travel modes from sensor data. Although the survey aims at including as many relevant issues as possible, some are left out due to the space limitation. We hope that the survey may serve as a basic guide for researchers and engineers who are interested in studying the topic and making further contribution.

One important future work is to study the online update as most existing techniques deal with the past movement. Infrastructure objects may be updated, e.g., construction of roads and buildings, and new bus schedule. Moving objects are updated in terms of location, speed, direction and transportation

⑦http://www.census.gov/geo/maps-data/data/tiger-line.html, Nov. 2018.

modes. In particular, updating transportation modes incurs the change of the environment and the referenced object. Another interesting issue is to discover interesting movement patterns regarding different environments and transportation modes. Meanwhile, the knowledge graph for urban movement can be built to extract meaningful relationships between humans and further enhance innovative mobile applications.

References

[1] Sistla P, Wolfson O, Chamberlain S, Dao S. Modeling and querying moving objects. In *Proc. the 13th International Conference on Data Engineering*, April 1997, pp.422-432.

[2] Güting R H, Schneider M. Moving Objects Databases. Morgan Kaufmann, 2005.

[3] Zheng Y, Chen Y, Li Q, Xie X, Ma W Y. Understanding transportation mode based on GPS data for web application. *ACM Transaction on the Web*, 2010, 4(1): Article No. 1.

[4] Yu M, Yu T, Wang S, Lin C, Chang E Y. Big data small footprint: The design of a low-power classifier for detecting transportation modes. *Proceedings of the VLDB Endowment*, 2014, 7(13): 1429-1440.

[5] Stenneth L, Wolfson O, Yu P, Xu B. Transportation mode detection using mobile phones and GIS information. In *Proc. the 19th ACM SIGSPATIAL International Symposium on Advances in Geographic Information Systems*, November 2011, pp.54-63.

[6] Tong Y, Zeng Y, Zhou Z, Chen L, Ye J, Xu K. A unified approach to route planning for shared mobility. *Proceedings of the VLDB Endowment*, 2018, 11(11): 1633-1646.

[7] Booth J, Sistla P, Wolfson O, Cruz I F. A data model for trip planning in multimodal transportation systems. In *Proc. the 12th International Conference on Extending Database Technology*, March 2009, pp.994-1005.

[8] Xu J, Güting R H, Zheng Y. The TM-RTree: An index on generic moving objects for range queries. *GeoInformatica*, 2015, 19(3): 487-524.

[9] Zografos K G, Androutsopoulos K N. Algorithms for itinerary planning in multimodal transportation networks. *IEEE Trans. Intelligent Transportation Systems*, 2008, 9(1): 175-184.

[10] Chaturvedi M, Srivastava S. Multi-modal design of an intelligent transportation system. *IEEE Trans. Intelligent Transportation Systems*, 2017, 18(8): 2017-2027.

[11] Li M, Dai J, Sahu S, Naphade M R. Trip analyzer through smartphone apps. In *Proc. the 19th ACM SIGSPATIAL International Symposium on Advances in Geographic Information Systems*, November 2011, pp.537-540.

[12] Froehlich J, Dillahunt T, Klasnja P V, Mankoff J, Consolvo S, Harrison B L, Landay J A. UbiGreen: Investigating a mobile tool for tracking and supporting green transportation habits. In *Proc. the 27th International Conference on Human Factors in Computing Systems*, April 2009, pp.1043-1052.

[13] Reddy S, Mun M, Burke J, Estrin D, Hansen M H, Srivastava M B. Using mobile phones to determine transportation modes. *ACM Transactions on Sensor Networks*, 2010, 6(2): Article No. 13.

[14] Prentow T S, Blunck H, Kjærgaard M B, Stisen A. Towards indoor transportation mode detection using mobile sensing. In *Proc. the 7th International Conference on Mobile Computing, Applications, and Services*, November 2015, pp.259-279.

[15] Asghari M, Shahabi C. An on-line truthful and individually rational pricing mechanism for ride-sharing. In *Proc. the 25th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems*, November 2017, Article No. 7.

[16] Cheng P, Xin H, Chen L. Utility-aware ridesharing on road networks. In *Proc. the 2017 ACM International Conference on Management of Data*, May 2017, pp.1197-1210.

[17] Wolfson O, Lin J. Fairness versus optimality in ridesharing. In *Proc. the 18th IEEE International Conference on Mobile Data Management*, May 2017, pp.118-123.

[18] Chiang M, Lim E, Lee W, Hoang T. Inferring trip occupancies in the rise of ride-hailing services. In *Proc. the 27th ACM International Conference on Information and Knowledge Management*, October 2018, pp.2097-2105.

[19] Xu Z, Li Z, Guan Q *et al*. Large-scale order dispatch in on-demand ride-hailing platforms: A learning and planning approach. In *Proc. the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, August 2018, pp.905-913.

[20] Agatz N A H, Bazzan A L C, Kutadinata R J, Mattfeld D C, Sester M, Winter S, Wolfson O. Autonomous car and ride sharing: Flexible road trains: (vision paper). In *Proc. the 24th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems*, October 2016, Article No. 10.

[21] Xu J, Güting R H. A generic data model for moving objects. *GeoInformatica*, 2013, 17(1): 125-172.

[22] Zheng Y, Li Q, Chen Y, Xie X, Ma W Y. Understanding mobility based on GPS data. In *Proc. the 10th International Conference on Ubiquitous Computing*, September 2008, pp.312-321.

[23] Zheng Y, Liu L, Wang L, Xie X. Learning transportation mode from raw GPS data for geographic applications on the web. In *Proc. the 17th International Conference on World Wide Web*, April 2008, pp.247-256.

[24] Vu T H, Dung L, Wang J. Transportation mode detection on mobile devices using recurrent nets. In *Proc. the 2016 ACM Conference on Multimedia Conference*, October 2016, pp.392-396.

[25] Biljecki F, Ledoux H, van Oosterom P. Transportation mode-based segmentation and classification of movement trajectories. *International Journal of Geographical Information Science*, 2013, 27(2): 385-407.

[26] Xiao Z, Wang Y, Fu K, Wu F. Identifying different transportation modes from trajectory data using tree-based ensemble classifiers. *ISPRS International Journal of Geo-Information*, 2017, 6(2): Article No. 57.

[27] Ashqar H I, Almannaa M H, Elhenawy M, Rakha H A, House L. Smartphone transportation mode recognition using a hierarchical machine learning classifier and pooled features from time and frequency domains. *IEEE Trans. Intelligent Transportation Systems*, 2019, 20(1): 244-252.

[28] Hemminki S, Nurmi P, Tarkoma S. Accelerometer-based transportation mode detection on smartphones. In *Proc. the 11th ACM Conference on Embedded Network Sensor Systems*, November 2013, Article No. 13.

[29] Bloch A, Erdin R, Meyer S, Keller T, de Spindler A. Battery-efficient transportation mode detection on mobile devices. In *Proc. the 16th IEEE International Conference on Mobile Data Management*, June 2015, pp.185-190.

[30] Coffey C, Nair R, Pinelli F, Pozdnoukhov A, Calabrese F. Missed connections: Quantifying and optimizing multimodal interconnectivity in cities. In *Proc. the 16th IEEE International Conference on Mobile Data Management*, June 2012, pp.26-32.

[31] Xu J, Güting R H, Qin X. GMOBench: Benchmarking generic moving objects. *GeoInformatica*, 2015, 19(2): 227-276.

[32] Xu J, Güting R H. MWGen: A mini world generator. In *Proc. the 13th IEEE International Conference on Mobile Data Management*, July 2012, pp.258-267.

[33] Wang W, Xu J. A tool for 3D visualizing moving objects. In *Proc. the 1st International Joint Conference on Web and Big Data*, July 2017, pp.353-357.

[34] Sohn T, Varshavsky A, LaMarca A, Chen M Y, Choudhury T, Smith I E, Consolvo S, Hightower J, Griswold W G, de Lara E. Mobility detection using everyday GSM traces. In *Proc. the 8th International Conference on Ubiquitous Computing*, September 2006, pp.212-224.

[35] Patterson D J, Liao L, Fox D, Kautz H A. Inferring high level behavior from low-level sensors. In *Proc. the 5th International Conference on Ubiquitous Computing*, October 2003, pp.73-89.

[36] Wang B, Gao L, Juan Z. Travel mode detection using GPS data and socioeconomic attributes based on a random forest classifier. *IEEE Trans. Intelligent Transportation Systems*, 2018, 19(5): 1547-1558.

[37] Li G, Chen C, Huang S, Chou A, Gou X, Peng W, Yi C. Public transportation mode detection from cellular data. In *Proc. the 2017 ACM Conference on Information and Knowledge Management*, November 2017, pp.2499-2502.

[38] Yan Z, Chakraborty D, Parent C, Spaccapietra S, Aberer K. SeMiTri: A framework for semantic annotation of heterogeneous trajectories. In *Proc. the 14th International Conference on Extending Database Technology*, March 2011, pp.259-270.

[39] Zheng K, Shang S, Yuan N J, Yang Y. Towards efficient search for activity trajectories. In *Proc. the 29th IEEE International Conference on Data Engineering*, April 2013, pp.230-241.

[40] Zheng B, Yuan N J, Zheng K, Xie X, Sadiq S W, Zhou X. Approximate keyword search in semantic trajectory database. In *Proc. the 31st IEEE International Conference on Data Engineering*, April 2015, pp.975-986.

[41] Choi D W, Pei J, Heinis T. Efficient mining of regional movement patterns in semantic trajectories. *Proceedings of the VLDB Endowment*, 2017, 10(13): 2073-2084.

[42] Wang S, Bao Z, Culpepper J S, Sellis T, Sanderson M, Qin X. Answering top-k exemplar trajectory queries. In *Proc. the 33rd IEEE International Conference on Data Engineering*, April 2017, pp.597-608.

[43] Zhuang C, Yuan N J, Song R, Xie X, Ma Q. Understanding people lifestyles: Construction of urban movement knowledge graph from GPS trajectory. In *Proc. the 26th International Joint Conference on Artificial Intelligence*, August 2017, pp.3616-3623.

[44] Prentow T S, Thom A, Blunck H, Vahrenhold J. Making sense of trajectory data in indoor spaces. In *Proc. the 16th IEEE International Conference on Mobile Data Management*, June 2015, pp.116-121.

[45] Valdés F, Güting R H. Index-supported pattern matching on symbolic trajectories. In *Proc. the 22nd ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems*, November 2014, pp.53-62.

[46] Güting R H, Valdés F, Damiani M L. Symbolic trajectories. *ACM Transactions on Spatial Algorithms and Systems*, 2015, 1(2): Article No. 7.

[47] Xu J, Lu H, Güting R H. Range queries on multi-attribute trajectories. *IEEE Transactions on Knowledge and Data Engineering*, 2018, 30(6): 1206-1211.

[48] Xu J, Güting R H, Gao Y. Continuous k nearest neighbor queries over large multi-attribute trajectories: A systematic approach. *GeoInformatica*, 2018, 22(4): 723-766.

[49] Wolfson O, Xu B, Chamberlain S, Jiang L. Moving objects databases: Issues and solutions. In *Proc. the 10th International Conference on Scientific and Statistical Database Management*, July 1998, pp.111-122.

[50] Forlizzi L, Güting R H, Nardelli E, Schneider M. A data model and data structures for moving objects databases. In *Proc. the 2000 ACM SIGMOD International Conference on Management of Data*, May 2000, pp.319-330.

[51] Güting R H, Böhlen M H, Erwig M, Jensen C S, Lorentzos N A, Schneider M, Vazirgiannis M. A foundation for representing and querying moving objects. *ACM Transactions on Database Systems*, 2000, 25(1): 1-42.

[52] Vazirgiannis M, Wolfson O. A spatiotemporal model and language for moving objects on road networks. In *Proc. the 7th International Symposium on Spatial and Temporal Databases*, July 2001, pp.20-35.

[53] Speicys L, Jensen C S, Kligys A. Computational data modeling for network-constrained moving objects. In *Proc. the 11th ACM International Symposium on Advances in Geographic Information Systems*, November 2003, pp.118-125.

[54] Hage C, Jensen C S, Pedersen T B, Speicys L, Timko I. Integrated data management for mobile services in the real world. In *Proc. the 29th International Conference on Very Large Data Bases*, September 2003, pp.1019-1030.

[55] Güting R H, de Almeida V T, Ding Z M. Modeling and querying moving objects in networks. *The VLDB Journal*, 2006, 15(2): 165-190.

[56] Speičys L, Jensen C S. Enabling location-based services — Multi-graph representation of transportation networks. *GeoInformatica*, 2008, 12(2): 219-253.

[57] Park S H, Lee J, Kim D. Spatial clustering based on moving distance in the presence of obstacles. In *Proc. the 12th International Conference on Database Systems for Advanced Applications*, April 2007, pp.1024-1027.

[58] Gao Y, Zheng B, Chen G, Chen C, Li Q. Continuous nearest-neighbor search in the presence of obstacles. *ACM Transactions on Database Systems*, 2011, 36(2): Article No. 9.

[59] Gao Y, Zheng B, Chen G, Li Q, Guo X. Continuous visible nearest neighbor query processing in spatial databases. *The VLDB Journal*, 2011, 20(3): 371-396.

[60] Gao Y, Liu Q, Miao X, Yang J. Reverse k-nearest neighbor search in the presence of obstacles. *Information Science*, 2016, 330: 274-292.

[61] Jensen C S, Lu H, Yang B. Graph model based indoor tracking. In *Proc. the 10th International Conference on Mobile Data Management*, May 2009, pp.122-131.

[62] Jensen C S, Lu H, Yang B. Indoor — A new data management frontier. *IEEE Data Eng. Bull.*, 2010, 33(2): 12-17.

[63] Xie X, Lu H, Pedersen T B. Efficient distance-aware query evaluation on indoor moving objects. In *Proc. the 29th IEEE International Conference on Data Engineering*, April 2013, pp.434-445.

[64] Xie X, Lu H, Pedersen T B. Distance-aware join for indoor moving objects. *IEEE Transactions on Knowledge and Data Engineering*, 2015, 27(2): 428-442.

[65] Lu H, Cheema M A. Indoor data management. In *Proc. the 32nd IEEE International Conference on Data Engineering*, May 2016, pp.1414-1417.

[66] Jensen C S, Lu H, Yang B. Indexing the trajectories of moving objects in symbolic indoor space. In *Proc. the 11th International Symposium on Spatial and Temporal Databases*, July 2009, pp.208-227.

[67] Yang B, Lu H, Jensen C S. Probabilistic threshold k nearest neighbor queries over moving objects in symbolic indoor space. In *Proc. the 13th International Conference on Extending Database Technology*, March 2010, pp.335-346.

[68] Koncz N A, Adams T M. A data model for multi-dimensional transportation applications. *International Journal of Geographical Information Science*, 2002, 16(6): 551-569.

[69] Timko I, Pedersen T B. Capturing complex multidimensional data in location-based data warehouses. In *Proc. the 12th ACM International Workshop on Geographic Information Systems*, November 2004, pp.147-156.

[70] Jensen C S, Kligys A, Pedersen T B, Timko I. Multidimensional data modeling for location-based services. *The VLDB Journal*, 2004, 13: 1-21.

[71] Hussein S H, Lu H, Pedersen T B. Towards a unified model of outdoor and indoor spaces. In *Proc. the 2012 ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems*, November 2012, pp.522-525.

[72] Lin J, Sasidharan S, Ma S, Wolfson O. A model of multimodal ridesharing and its analysis. In *Proc. the 17th IEEE International Conference on Mobile Data Management*, June 2016, pp.164-173.

[73] Yang Y, Gao H, Yu J X, Li J. Finding the cost-optimal path with time constraint over time-dependent graphs. *Proceedings of the VLDB Endowment*, 2014, 7(9): 673-684.

[74] Li L, Hua W, Du X, Zhou X. Minimal on-road time route scheduling on time-dependent graphs. *Proceedings of the VLDB Endowment*, 2017, 10(11): 1274-1285.

[75] Ding Z, Güting R H. Managing moving objects on dynamic transportation networks. In *Proc. the 16th International Conference on Scientific and Statistical Database Management*, June 2004, pp.287-296.

[76] Huang B, Wu Q, Zhan F B. A shortest path algorithm with novel heuristics for dynamic transportation networks. *International Journal of Geographical Information Science*, 2007, 21(6): 625-644.

[77] Bakalov P, Hoel E G, Heng W, Tsotras V J. Maintaining connectivity in dynamic multimodal network models. In *Proc. the 24th International Conference on Data Engineering*, April 2008, pp.1267-1276.

[78] Bakalov P, Hoel E G, Heng W. Time dependent transportation network models. In *Proc. the 31st International Conference on Data Engineering*, April 2015, pp.1364-1375.

[79] Wang S, Lin W, Yang Y, Xiao X, Zhou S. Efficient route planning on public transportation networks: A labelling approach. In *Proc. the 2015 ACM SIGMOD International Conference on Management of Data*, May 2015, pp.967-982.

[80] Bauer V, Gamper J, Loperfido R, Profanter S, Putzer S, Timko I. Computing isochrones in multi-modal, schedule-based transport networks. In *Proc. the 16th ACM SIGSPATIAL International Symposium on Advances in Geographic Information Systems*, November 2008, Article No. 78.

[81] Tang G, Keshav S, Golab L, Wu K. Bikeshare pool sizing for bike-and-ride multimodal transit. *IEEE Trans. Intelligent Transportation Systems*, 2018, 19(7): 2279-2289.

[82] Bastani F, Xie X, Huang Y, Powell J W. A greener transportation mode: Flexible routes discovery from GPS trajectory data. In *Proc. the 19th ACM SIGSPATIAL International Symposium on Advances in Geographic Information Systems*, November 2011, pp.405-408.

[83] Ma S, Zheng Y, Wolfson O. T-share: A large-scale dynamic taxi ridesharing service. In *Proc. the 29th IEEE International Conference on Data Engineering*, April 2013, pp.410-421.

[84] Chen L, Zhong Q, Xiao X, Gao Y, Jin P, Jensen C S. Price-and-time-aware dynamic ridesharing. In *Proc. the 34th IEEE International Conference on Data Engineering*, April 2018, pp.1061-1072.

[85] Xu J, Güting R H. Manage and query generic moving objects in SECONDO. *Proceedings of the VLDB Endowment*, 2012, 5(12): 2002-2005.

[86] Güting R H, Behr T, Düntgen C. SECONDO: A platform for moving objects database research and for publishing and integrating research implementations. *IEEE Data Eng. Bull.*, 2010, 33(2): 56-63.

[87] Wei X, Xu J. MDBF: A tool for monitoring database files. In *Proc. ER*, October 2018, pp.54-58.

[88] Liao L, Patterson D J, Fox D, Kautz H A. Learning and inferring transportation routines. *Artif. Intell.*, 2007, 171(5/6): 311-331.

[89] Widhalm P, Nitsche P, Brändle N. Transport mode detection with realistic smartphone sensor data. In *Proc. the 21st International Conference on Pattern Recognition*, November 2012, pp.573-576.

[90] Gong H, Chen C, Bialostozky E, Lawson C T. A GPS/GIS method for travel mode detection in New York City. *Computers, Environment and Urban Systems*, 2012, 36(2): 131-139.

[91] Asmar D C, Zelek J S, Abdallah S M. SmartSLAM: Localization and mapping across multi-environments. In *Proc. the 2004 IEEE International Conference on Systems, Man and Cybernetics*, October 2004, pp.5240-5245.

[92] Du H, Henry P, Ren X, Cheng M, Goldman D B, Seitz S M, Fox D. Interactive 3D modeling of indoor environments with a consumer depth camera. In *Proc. the 13th International Conference on Ubiquitous Computing*, September 2011, pp.75-84.

[93] Jiang Y, Yun X, Pan X, Li K, Lv Q, Dick R P, Shang L, Hannigan M. Hallway based automatic indoor floorplan construction using room fingerprints. In *Proc. the 2013 ACM International Joint Conference on Pervasive and Ubiquitous Computing*, September 2013, pp.315-324.

[94] Srivatsa M, Ganti R K, Wang J, Kolar V. Map matching: Facts and myths. In *Proc. the 21st SIGSPATIAL International Conference on Advances in Geographic Information Systems*, November 2013, pp.474-477.

[95] Taguchi S, Koide S, Yoshimura T. Online map matching with route prediction. *IEEE Trans. Intelligent Transportation Systems*, 2019, 20(1): 338-347.

[96] AlDwyish A, Xie H, Tanin E, Karunasekera S, Ramamohanarao K. Using a traffic simulator for navigation service. In *Proc. the 25th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems*, November 2017, Article No. 78.

[97] Keler A, Kaths J, Chucholowski F, Chucholowski M, Grigoropoulos G, Spangler M, Kaths H, Busch F. A bicycle simulator for experiencing microscopic traffic flow simulation in urban environments. In *Proc. the 21st International Conference on Intelligent Transportation Systems*, November 2018, pp.3020-3023.

[98] Liang Y, Gao S, Wu T, Wang S, Wu Y. Optimizing bus stop spacing using the simulated annealing algorithm with spatial interaction coverage model. In *Proc. the 11th ACM SIGSPATIAL International Workshop on Computational Transportation Science*, November 2018, pp.53-59.

[99] Xie H, Tanin E, Karunasekera S, Kulik L, Zhang R, Qi J, Ramamohanarao K. Studying transportation problems with the SMARTS simulator (demo paper). In *Proc. the 26th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems*, November 2018, pp.580-583.

[100] Theodoridis Y, Silva J R O, Nascimento M A. On the generation of spatiotemporal datasets. In *Proc. the 6th International Symposium on Spatial Databases*, July 1999, pp.147-164.

[101] Pfoser D, Theodoridis Y. Generating semantics-based trajectories of moving objects. *Computers, Environment and Urban Systems*, 2003, 27(3): 243-263.

[102] Saltenis S, Jensen C S, Leutenegger S T, López M A. Indexing the positions of continuously moving objects. In *Proc. the 2000 ACM SIGMOD International Conference on Management of Data*, May 2000, pp.331-342.

[103] Düntgen C, Behr T, Güting R H. BerlinMOD: A benchmark for moving objects database. *The VLDB Journal*, 2009, 18(6): 1335-1368.

[104] Brinkhoff T. Generating network-based moving objects. In *Proc. the 12th International Conference on Scientific and Statistical Database Management*, July 2000, pp.253-255.

[105] Brinkhoff T. A framework for generating network-based moving objects. *GeoInformatica*, 2002, 6(2): 153-180.

[106] Krajzewicz D, Hertkorn G, Rössel C, Wagner P. Sumo (simulation of urban mobility): An open-source traffic simulation. In *Proc. the 4th Middle East Symposium on Simulation and Modelling*, September 2002, pp.183-187.

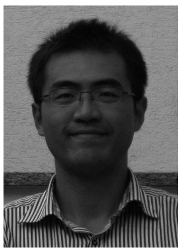
[107] Hu H, Lee D L. GAMMA: A framework for moving object simulation. In *Proc. the 9th International Symposium on Spatial and Temporal Databases*, August 2005, pp.37-54.

[108] Pelekis N, Ntrigkogias C, Tampakis P, Sideridis S, Theodoridis Y. Hermoupolis: A trajectory generator for simulating generalized mobility patterns. In *Proc. the 2013 European Conference on Machine Learning and Knowledge Discovery in Databases*, September 2013, pp.659-662.

[109] Mokbel M F, Alarabi L, Bao J, Eldawy A, Magdy A, Sarwat M, Waytas E, Yackel S. A demonstration of MNTG — A web-based road network traffic generator. In *Proc. the 30th IEEE International Conference on Data Engineering*, March 2014, pp.1246-1249.

[110] Huang C, Jin P, Wang H, Wang N, Wan S, Yue L. IndoorSTG: A flexible tool to generate trajectory data for indoor moving objects. In *Proc. the 14th International Conference on Mobile Data Management*, June 2013, pp.341-343.

[111] Li H, Lu H, Chen X, Chen G, Chen K, Shou L. Vita: A versatile toolkit for generating indoor mobility data for real-world buildings. *Proceedings of the VLDB Endowment*, 2016, 9(13): 1453-1456.



Jian-Qiu Xu got his Bachelor's and Master's degrees in computer science from Nanjing University of Aeronautics and Astronautics, Nanjing, in 2005 and 2008, respectively. Then, he got his Ph.D. degree supervised by Prof. Dr. Ralf Hartmut Güting in 2008–2012 from University of Hagen, Hagen, focusing on moving objects databases and spatial databases. In 2013, he joined Nanjing University of Aeronautics and Astronautics, Nanjing, as an assistant professor. His research has been funded by NSFC (National Natural Science Foundation of China). He has published a book on moving objects with transportation modes and more than 20 papers in journals and conferences such as TKDE, Geoinformatica, Information System and ICDE. He won the Best Demo Paper in APWeb/WAIM 2017.

Ralf Hartmut Güting has been a full professor in computer science at the University of Hagen, Hagen, since 1989. He received his Diploma and Dr. rer. nat. degrees from the University of Dortmund, Dortmund, in 1980 and 1983, respectively, and became a professor at that university in 1987. From 1981 to 1984 his main research area was computational geometry. After a one-year stay at the IBM Almaden Research Center in 1985, extensible and spatial database systems became his major research interests; more recently, also spatio-temporal or moving objects databases. He has been an associate editor of ACM Transactions on Database Systems and an editor of VLDB Journal, and is on the Editorial Board of GeoInformatica. He has published two German text books on data structures and algorithms and on compilers, respectively, and an English text book on moving objects databases, as well as around 80 journal and conference articles. His group has built prototypes of extensible and spatio-temporal database systems, the Gral system and the SECONDO system.

Yu Zheng is the vice-president of JD. He received the award of ACM Distinguished Scientist in 2016. His research interests include location-based services, spatio-temporal data mining, ubiquitous computing, and mobile social applications. He has published over 50 referred papers as a leading author at high-quality international conferences and journals, such as SIGMOD, SIGKDD, AAAI, ICDE, WWW, Ubicomp, IEEE TKDE, and ACM TWEB. He has received three Best Paper Awards from UIC 10, ACM SIGSPATIAL GIS 11, and ADMA 11 as well as one Best Paper Nominee from Ubicomp 11. He is also on the editorial board of four international journals, such as GeoInformatica and Journal of Location Based Services. In 2008, he was recognized as the Microsoft Golden Star. He joined MSRA (Microsoft Research Asia), Beijing, in July 2006 right after he received his Ph.D. degree in computer science from Southwest Jiaotong University, Chengdu, and joined JD as the vice-president for leading the team of urban computing in 2018.

Ouri Wolfson received his B.A. degree in mathematics, and his Ph.D. degree in computer science from Courant Institute of Mathematical Sciences, New York University. His main research interests are in database systems, distributed systems, and mobile/pervasive computing. He is currently the Richard and Loan Hill Professor of Computer Science at the University of Illinois at Chicago, and an affiliate professor in the Department of Computer Science at the University of Illinois at Urbana Champaign. He is the founder of Mobitrac, a high-tech startup that was acquired by Fluensee Co. in 2006, and the founder and president of Pirouette Software Inc. which specializes in mobile data management. He served as a consultant to Argonne National Laboratory, US Army Research Laboratories, DARPA, and NASA. Before joining the University of Illinois he has been on the computer science faculty at the Technion and Columbia University, and a member of Technical Staff at Bell Laboratories. Ouri Wolfson authored over 200 publications, and holds seven patents. He is a fellow of the Association of Computing Machinery (ACM), a fellow of the American Association for the Advancement of Science (AAAS), a fellow of the Institute of Electrical and Electronics Engineers (IEEE), and a University of Illinois Scholar for 2009. He co-authored six award winning papers, and served as a distinguished lecturer for the Association of Computing Machinery during 2001–2003.