

Semantic Trajectories Modeling and Analysis

CHRISTINE PARENT, University of Lausanne
 STEFANO SPACCAPIETRA, Swiss Federal Institute of Technology (EPFL)
 CHIARA RENSO, ISTI-CNR
 GENNADY ANDRIENKO, Fraunhofer Institute IAIS
 NATALIA ANDRIENKO, Fraunhofer Institute IAIS
 VANIA BOGORNY, Federal University of Santa Catarina (INE/UFSC)
 MARIA LUISA DAMIANI, University of Milan
 ARIS GKOUALAS-DIVANIS, IBM Research-Zurich
 JOSE MACEDO, Federal University of Ceará
 NIKOS PELEKIS, University of Piraeus
 YANNIS THEODORIDIS, University of Piraeus
 ZHIXIAN YAN, Swiss Federal Institute of Technology (EPFL)

Focus on movement data has increased as a consequence of the larger availability of such data due to current GPS, GSM, RFID, and sensors techniques. In parallel, interest in movement has shifted from raw movement data analysis to more application-oriented ways of analyzing segments of movement suitable for the specific purposes of the application. This trend has promoted semantically rich trajectories, rather than raw movement, as the core object of interest in mobility studies. This survey provides the definitions of the basic concepts about mobility data, an analysis of the issues in mobility data management, and a survey of the approaches and techniques for i) constructing trajectories from movement tracks, ii) enriching trajectories with semantic information to enable the desired interpretations of movements, and iii) using data mining to analyze semantic trajectories and extract knowledge about their characteristics, in particular the behavioral patterns of the moving objects. Last but not least, the paper surveys the new privacy issues that rise due to the semantic aspects of trajectories.

Categories and Subject Descriptors: H. Information Systems, H.2. Database Management, H.2.0 General

General Terms: algorithms, design, legal aspects, management

Additional Key Words and Phrases: movement, mobility tracks, tracking, mobility data, trajectories, trajectory behavior, semantic enrichment, data mining, activity identification, GPS

This work is supported by the EU, FET OPEN, 2009-2012 Programme, Coordination Action type Project MODAP (Mobility, Data Mining, and Privacy) <http://www.modap.org>

Authors' addresses: Gennady and Natalia Andrienko, Fraunhofer Institute IAIS, Schloss Birlinghoven, Sankt-Augustin, D-53754 Germany, email: {gennady|natalia}.andrienko@iais.fraunhofer.de; Vania Bogorny, UFSC-CTC-INE, CEP 88040-900 - Campus Universitário Cx.P. 476, Florianópolis S.C., Brazil, vania@inf.ufsc.br; Maria Luisa Damiani: Università degli Studi di Milano, Via Comelico 39, 20135 Milano, Italy, email: mdamiani@disco.unimi.it; Aris Gkoulalas-Divanis: IBM Research GmbH, Säumerstrasse 4, CH-8803 Rüschlikon, Switzerland, agd@zurich.ibm.com; José Macedo: Campus do PICI, Computer Science Department, Fortaleza, Brazil, CEP 60.115-190, jose.macedo@lia.ufc.br; Christine Parent: Université de Lausanne HEC-ISI, 1015 Lausanne, Switzerland, email: christine.parent@unil.ch; Nikos Pelekis: Department of Statistics and Insurance Science, University of Piraeus, Karaoli-Dimitriou 80, Piraeus, GR-18534, Greece, email: npelekis@unipi.gr; Chiara Renso: ISTI-CNR, Via Moruzzi 1, 56010, Pisa, Italy, email: chiara.renso@isti.cnr.it; Stefano Spaccapietra and Zhixian Yan: EPFL-IC-LSIR, Station 14, CH-1015 Lausanne, email: {stefano.spaccapietra, zhixian.yan}@epfl.ch; Yannis Théodoridis: Department of Informatics, University of Piraeus, GR-18534 Piraeus, Greece, email: ytheod@unipi.gr

1. INTRODUCTION¹

Mobility is one of the major keywords that characterize the current development of our society. People, goods and ideas are moving faster and more frequently than ever. Ubiquitous computing facilities and location-based services have greatly supported human mobility. More recently, GPS and other positioning devices enabled capturing the evolving position of objects moving in geographical space. Massive amounts of tracking data have been created, for the benefit of novel applications that build on this movement knowledge. Researchers from the database, GIS, visualization, data mining, and knowledge extraction communities have developed models and techniques for mobility analysis. Their results represent an important step forward over previous foundational work done by e.g. Güting et al. on moving objects [Güting et al. 2000], and the EU Chorochronos project [Koubarakis et al. 2003] on spatio-temporal data.

A detailed survey of mobility research up to 2007 has been published in [Giannotti and Pedreschi 2008] as part of the EU GeoPKDD project (www.geopkdd.eu). Over the last few years the corpus of mobility techniques has greatly expanded. In particular, a new promising approach has been devised to provide applications with richer and more meaningful knowledge about movement. This is achieved by combining the raw mobility tracks (e.g. the GPS records) with related contextual data. These enriched track records are referred to as *semantic trajectories*.

This paper surveys the new ideas and techniques related to the elaboration and analysis of semantic trajectories. Some of the ideas and techniques previously developed for raw trajectories will nevertheless be mentioned for better understandability. An extensive coverage of previous techniques can be found in [Zheng and Zhou 2011]. Our paper is purposely focused on data mining techniques, as these are the most frequently used in the literature. Other techniques (e.g. visual analytics [Andrienko et al. 2011], and aggregation techniques) also play an important role in movement analysis, but they would require a full survey on their own to be properly discussed.

The sequel of the paper is organized as follows. Sections 2 and 3 define the basic concepts and terminology that support our analysis of trajectories and their behaviors. Next we discuss the three steps to convert raw data into knowledge: Trajectory reconstruction (Section 4), semantic enrichment (Section 5) and behavior knowledge extraction (Section 6). Section 7 surveys the privacy issues that characterize human trajectories. The conclusion introduces foreseeable future developments.

¹ Permission to make digital/hard copy of part of this work for personal or classroom use is granted without fee provided that the copies are not made or distributed for profit or commercial advantage, the copyright notice, the title of the publication, and its date of appear, and notice is given that copying is by permission of the ACM, Inc. To copy otherwise, to republish, to post on servers, or to redistribute to lists, requires prior specific permission and/or a fee. Permission may be requested from the Publications Dept., ACM, Inc., 2 Penn Plaza, New York, NY 11201-0701, USA, fax: +1 (212) 869-0481, permission@acm.org

2. BASIC CONCEPTS FOR A SEMANTIC VIEW ON TRAJECTORIES

The trajectory idea originates in the ability to capture the movement of an object moving in geographical space over some period of time. Movement capture results in generating for each moving object its movement track. A **movement track** basically consists in the temporal sequence of the spatio-temporal positions - i.e. (instant, point) pairs - recorded for the moving object. However, depending on the capabilities of the device, additional data, e.g. the instant speed or stillness, acceleration, direction, and rotation, may complement the (instant, point) pairs. We call **raw data** the data as captured from the device.

Raw movement tracks may be used as such for further analysis or be transformed into other kinds of representation of movement. For example, an application willing to study how movement globally spreads over a given area will opt for a continuous view of space and record movement as a field of vectors over this space. Vectors aggregate data from the individual tracks to represent, for a given instant, some characteristics – usually speed and direction – of the movements at every position in space. An application willing to globally analyze the flow of objects moving between a discrete set of points (e.g. popular places within a city) will aggregate individual movement tracks into edges between nodes of a flow network. For example, Orellana et al. use movement tracks, fields of vectors and flows for analyzing people's movements in recreational areas [Orellana et al. 2009]. There are also movement analyses that do not rely at all on movement tracks (e.g. analyses of movement of body parts such as eyes or hands). To keep our survey focused, we do not address these alternative views of movement, nor do we address deformation issues raised when considering moving objects, such as hurricanes and oil spills, that span over a changing area or volume. We focus instead on movement tracks of moving objects represented as moving points.

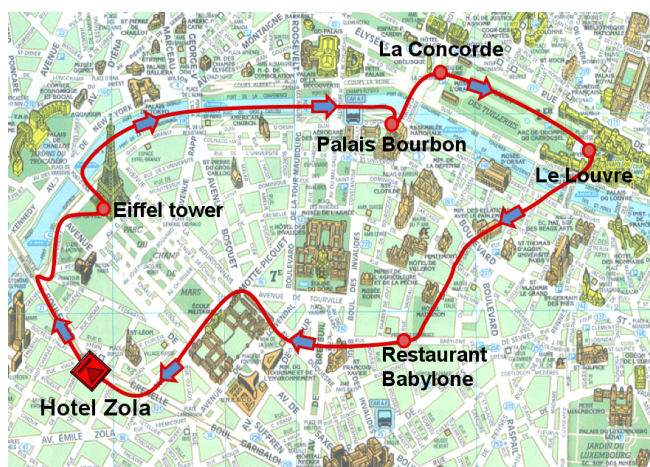


Figure 1 - 2D visualization of a one-day spatial trace left by a tourist visiting Paris – background map downloaded from Mappery.com, copyright unknown

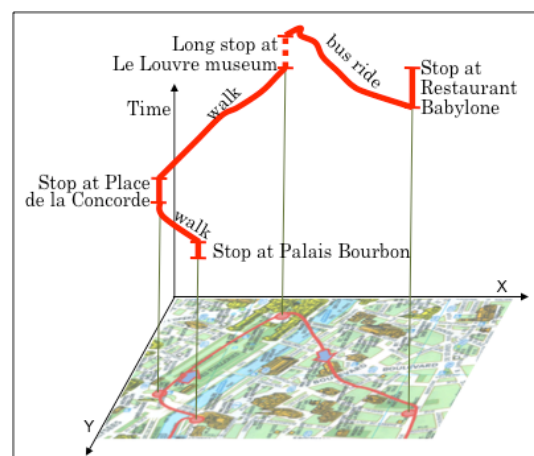


Figure 2 - A time-geography diagram showing part of the previous tourist track

Figures 1 and 2 convey possible representations of the track left by a tourist visiting Paris during one day. The three vertical segments in the trace in Figure 2 (inspired by the work of [Hägerstrand 1970]) correspond to the tourist stopping in a place for a temporal duration proportional to the length of the segment.

2.1 From Movement Tracks to Trajectories

Many applications are not interested in keeping and analyzing exhaustive 24/7 records of movement. They rather choose the segments they are interested in. We call **trajectories** the segments of the object's movement track that are of interest for a given application. In the tourist example, to globally analyze the activities performed by a tourist during his/her stay in Paris, the whole track left by the tourist in Paris will be taken as a single trajectory (spatial criterion “inside Paris”). To analyze what tourists do in one day in Paris, or what they do on specific days (e.g., on Sundays), each daily track of a tourist in Paris will be taken as a separate trajectory (see Fig.1). Figure 3 shows (as a dotted line) a section of the movement track of a moving object and, superimposed as a continuous line, two subsections identified as relevant trajectories. Each trajectory is identified by two specific spatio-temporal positions of the movement track, called the **Begin** and the **End** of the trajectory: They are the first and the last positions of the object for this trajectory [Spaccapietra et al. 2008].

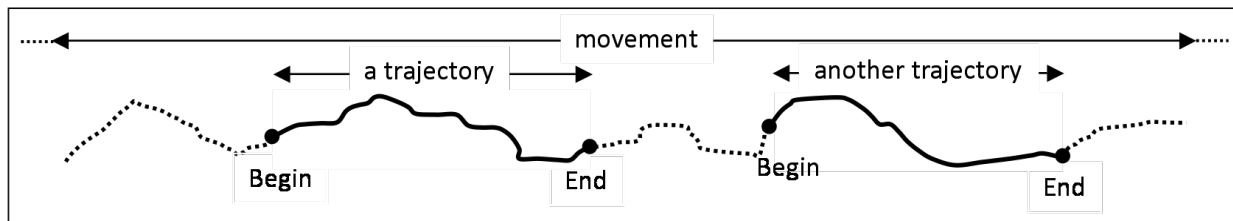


Figure 3 - Trajectories extracted from a movement track visualized as dotted line

Definition: A **raw trajectory** is a trajectory extracted from a raw movement track and containing only raw data for its Begin-End interval. It is defined as a tuple:

(trajectoryID, movingObjectID, trace: LISTOF position(instant, point, δ))

where δ denotes a possibly empty list of additional raw data (e.g. speed, direction). \square

Notice that the movement track contains only points recorded at some sampling instants while movement is continuous. Therefore it is an approximation of the real movement. Interpolation functions are frequently used to compute the likely position of the moving object for any instant between two consecutive sampled positions. The computed positions complement the captured positions, thus reconstructing continuity of movement.

It is, however, often the case that a movement track shows an abnormal (greater than the sampling rate) temporal gap between two consecutive positions. This means that over some periods of time the information on the movement of the object is missing. If this is accidental (e.g., because of device malfunction) we say there is a **hole** in the track. If this is deliberate (e.g., an employee deactivating her GPS when going for

lunch) we say there is a **semantic gap**. Semantic gaps have to be explicitly recorded in order to differentiate them from the holes. Holes may sometimes be “filled”, using e.g. linear interpolation algorithms that compute the missing positions. Semantics gaps, instead, are not to be filled, as the positions are intentionally missing [Vazirgiannis and Wolfson 2001].

To facilitate the comparative analysis of trajectories, trajectories are sometimes **synchronized**: All the trajectories are re-sampled to share the same sequence of regularly sampled instants [Gudmundsson et al. 2007, Benkert et al. 2008]. Laube et al. [Laube et al. 2005a] also follow this approach and describe the movement of n objects during m regularly sampled instants by a $n \times m$ matrix, called REMO, that contains some spatio-temporal property of the movement, like speed or direction. The matrix readily shows (i) for each object, the changes of its property value, and (ii) for each instant, the similarities or differences among the property values of the objects.

2.2 From Raw Trajectories to Semantic Trajectories

Raw trajectories are well fitted for applications that aim only at locating some moving object (e.g., where was Mary at 8am on March 8, 2012?) or computing statistics on the spatio-temporal characteristics of trajectories (e.g., which percentage of trajectories show an average speed over 10km/h?). However, most application analyses require complementing raw data with additional information from the application context. For example, interpreting trajectories of persons within a city requires some knowledge about the features of the city (e.g. map, points of interest). Thanks to city information, spatio-temporal coordinates can be replaced with street and crossing names, or with names of places of interest, such as shops, restaurants, and museums. Information about ongoing events (e.g. concerts, fairs) enables e.g. traffic monitoring applications to differentiate normal traffic conditions from exceptional traffic conditions. We generically call **contextual data repository** the external data sources (e.g., application databases, geo databases, web pages) that provide contextual data.

Inputting contextual knowledge into application analysis processes can be done in two ways: (i) by dynamically computing the links to the desired contextual objects during the analyses, or (ii) by preparing in advance the trajectories for the analyses, i.e. computing the links to the contextual objects and adding data from these objects to the raw trajectories. Adding knowledge to raw trajectories is known as a **semantic enrichment** process (see Section 5). Enrichment conveys the idea that existing data is complemented with additional data, called annotations. An **annotation** here is any additional data that is attached either to a trajectory as a whole or to some of its subparts. For example, recording the goal of a person's trip to Paris (e.g., business, tourism) is an annotation at the trajectory level (i.e., holding one value per trajectory). Instead, recording the transportation means used by a person at each position is an annotation that may be recorded at the position level (i.e., holding one value per position).

An annotation value is an attribute value (e.g. the string “bus” is a possible value for a *TransportationMeans* annotation) or a link to an object in the contextual data repository (e.g. the key of the tuple “bus_line_13” in the Bus table may be a value for a

more precise *TransportationMeans* annotation). It may also be a complex value composed of attribute values and links to objects. Annotation values may be captured by observers (e.g., ecologists manually capturing the activity of the observed animals) or by sensors (e.g. instant speed). They may be computed from raw data (e.g., distance from Begin), extracted from contextual data (e.g., local weather), and inferred by reasoning (e.g., the transportation means may be inferred from the velocity and acceleration data, combined with contextual knowledge about the transport and road networks) [Guc et al. 2008].

Another way of enhancing the knowledge on trajectories is to superimpose a structure of homogeneous segments (with respect to given segmentation criteria) that are meaningful for the application. Such homogeneous segments are called ***episodes*** [Mountain and Raper 2001], where an episode is defined as a maximal sub-sequence of a trajectory such that all its spatio-temporal positions comply with a given predicate. The predicate bears on the spatio-temporal coordinates of the positions, their annotations and/or their spatio-temporal relationships to contextual objects. For example, a person's trajectory may be segmented into episodes defined by the value of the *TransportationMeans* annotation. Thus, a trajectory may consist of a first episode where *TransportationMeans*="walk", followed by a second episode where *TransportationMean*="bus", and ending with a third episode where *TransportationMeans*="walk" [Zheng et al. 2010]. A trajectory may be segmented in various ways corresponding to whatever segmentation criteria are of interest to the application at hand. A popular segmentation criterion is stillness versus movement, which generates two kinds of alternating episodes, called *stops* and *moves*. Section 5.1 provides more details on trajectory segmentation.

Definition: A ***semantic trajectory*** is a trajectory that has been enhanced with annotations and/or one or several complementary segmentations. It is defined as a tuple (a full explanation is given in [Spaccapietra and Parent 2011]):

(trajectoryID, movingObjectID, trajectoryAnnotations,
 trace: LISTOF position (instant, point, δ , positionAnnotations),
 semanticGaps: LISTOF gap (t_1 , t_2),
 segmentations: SETOF segmentation (segmentationID,
 episodes: LISTOF episode (t_3 , t_4 , definingAnnotation, episodeAnnotations))) □

The definition shows that each episode is characterized by its starting/ending instants (t_3 , t_4) and the value of the annotation of the segmentation criterion that defines this episode (definingAnnotation). Episodes may also bear other annotations (episodeAnnotations). For instance, in several works each stop episode is annotated with the reference to the closest place of interest (see Section 5.2). The definition covers the trajectories made up of stops and moves, as well as trajectories that are sequences of episodes defined by e.g. the transportation means or the region traversed. So sequences of regions, sequences of road segments (for network constrained trajectories), sequences of events are semantic trajectories.

A peculiar kind of segmented trajectories are ***semantic map-matched trajectories***. This representation is specific to trajectories of objects whose movements are

constrained by a network, like cars or trains. First the raw spatio-temporal positions are snapped onto the nodes and edges of the network (see Section 4.2); Then they are transformed into relative positions on the linear geo-objects that correspond to the edges. Each trajectory becomes a sequence of (instant, streetId, distance) where distance is the distance of the moving object to the beginning of the street [Cao and Wolfson 2005, Ding and Deng 2011].

Segmented trajectories may convey only a partial representation of the movement. For instance, for a space partitioned into zones, [Stewart-Hornsby and Cole 2007] describes trajectories as sequences of events that say when the object enters a new zone, while [du Mouza and Rigaux 2005] keeps sequences of couples (zoneID, duration) that say how much time the moving object spent in crossing the zone.

Knowing what is in a trajectory, we can now look at how to analyze trajectories to infer further characterization of the information they convey.

3. BASIC CONCEPTS FOR TRAJECTORY-RELATED BEHAVIORS

There are many ways to use trajectory data. The simplest one is by querying the data to find facts about some moving object. For example, a trajectory database in an express delivery company allows customers to find out where their shipment is at any point in time. More frequently, applications focus on analytical querying, whose ultimate goal is to detect and analyze trends extracted from the trajectory database. For example, the trajectories of tourists visiting Paris may be analyzed for building tourist profiles and suggesting personalized tours and services, regulating the flows of tourists in the attractions, tuning the facilities... All that relies on analyzing the similarities and dissimilarities among the trajectories, classifying the trajectories into groups of similar trajectories, identifying outliers, extracting the common characteristics that distinguish one group from another. A set of distinguishing characteristics forms a summary description of the group of trajectories. These summary descriptions are called patterns or behaviors [Laube 2009].

Definition: A *trajectory behavior* (behavior, in short) is a set of characteristics that identifies a peculiar bearing of a moving object or of a set of moving objects. The behavior is defined by a predicate that says if a given trajectory (or a given set of trajectories) shows the behavior. □

Behavior examples for the daily trajectories of persons moving around within Paris with a GPS include the *Tourist* behavior:

Tourist behavior: A daily trajectory shows the *Tourist* behavior if: Its Begin point P1 is a place of kind “Accommodation”, it makes at least one stop in a place of kind ‘Museum’ or “TouristAttraction”, it makes one stop in a place of kind “EatingPlace”, and its End point is in the same P1 place as its Begin point. □

The characteristic on which this definition is based is the “kind of place” characteristic associated to trajectory stops, Begin and End. An example of a more generic, application independent behavior, is the *Meet* behavior that characterizes a set of trajectories [Dodge et al. 2008]:

Meet: A set of trajectories shows the *Meet* behavior if: Every trajectory of the set roughly ends at the same point and at the same instant. \square

The predicate that defines a behavior may rely on the spatio-temporal characteristics of the trajectories, e.g. speed for the *Speeding* behavior that characterizes vehicles speeding above the speed limit, or the Begin and End positions for the *Meet* behavior). It may also rely on the semantic information conveyed via the annotations and the contextual data that are linked through spatio-temporal relationships to the trajectories. This is the case, for example, for the *Tourist* behavior.

A trajectory can show several behaviors. For instance a trajectory can show both the *Speeding* and the *Tourist* behaviors, and simultaneously be part of a group of trajectories showing the *Meet* behavior. For each behavior the predicate relies on different characteristics.

Given the lack of a standardized taxonomy or ontology of trajectory behaviors, it is essential for a mobility application to precisely define the predicates for the trajectory behaviors that are relevant to its goals. Existing taxonomic studies include [Laube and Imfeld 2002, Dodge et al. 2008, Thériault et al. 1999] for spatio-temporal behaviors. More recent proposals include [Andrienko and Andrienko 2007] which addresses semantic behaviors that take into consideration terrain features (e.g. obstacles) as well as some semantic annotations (e.g. means of transportation), and [Wood and Galton 2009b] which develops a deeper investigation into concepts for collective behaviors. An important obstacle to an agreed upon taxonomy is the absence of a community body having the authority to establish definitions. Currently each research paper tends to set its own definitions even for the most well known behaviors, as discussed in [Laube 2009].

3.1 Individual Behavior versus Collective Behavior

An important characteristics of trajectory behaviors is whether they apply to single moving objects (individual behaviors), or to groups of moving objects (collective behaviors).

Definitions: An *individual trajectory behavior* is a trajectory behavior whose predicate $p(T)$ bears on a trajectory (T).

A *collective trajectory behavior* is a trajectory behavior whose predicate $p(S)$ bears on a non empty set of trajectories (S). \square

For instance, the *Speeding* and *Tourist* behaviors are individual behaviors that characterize individual trajectories. On the opposite, the *Meet* behavior is collective. Another well-known example of collective behavior is the *Flock* behavior that characterizes a group of trajectories that stay not far from each other during some minimal duration [Benkert et al. 2008, Laube et al. 2005b]:

Flock behavior: A set of trajectories shows the *Flock* behavior during a given time interval if: At each instant t of the time interval there is a circle such that 1) its radius is smaller than a given threshold, and 2) it contains the positions at t of all the trajectories.

Collective behaviors can be defined for:

- known pre-existing groups. E.g. a platoon of soldiers, a herd of sheep shepherded by a dog may both show the *Flock* behavior, i.e. moving together like a flock.
- groups that are identified as such by the evaluation of the behavior predicate. E.g. a protest march may also show the *Flock* behavior.

Wood and Galton have done significant work on an ontological analysis of the group (called “collective” in their approach) and group motion concepts (see e.g., [Wood and Galton 2009a, 2009b, 2010]). In [Wood and Galton 2009a] they propose a taxonomy of collectives that stems from a characterization of the different features of collectives and of the relationships between a collective and its members. [Wood and Galton 2009b] complements the taxonomy of collectives with a taxonomy of collective movement patterns. [Wood and Galton 2010] provides an analysis of spatial and temporal granularity issues in the observation of groups and their movement.

In between individual and collective behaviors are behaviors that characterize a trajectory that has a peculiar bearing with respect to a group of trajectories or to some other trajectories. A typical example is the *Leadership* behavior of e.g. the alpha wolf of a pack of wolves [Andersson et al. 2008]:

Leadership: Let S be a set of trajectories showing the *Flock* behavior. A trajectory T of S shows the *Leadership* behavior during some given time interval if: During the time interval, each time the flock S moves, T is ahead of the other trajectories of the flock S .

3.2 Other Behavior Classifications

Trajectory behaviors can also be classified according to other criteria. These other classifications are independent from the individual/collective one. Based on which characteristics of the trajectories are involved in the predicate defining a behavior, behaviors can for example be classified as spatio-temporal or semantic.

Definitions: A *spatio-temporal behavior* is a trajectory behavior whose predicate bears only on the spatial and/or temporal data (no contextual data).

A *semantic behavior* is a trajectory behavior whose predicate bears on some contextual data and possibly on some spatial and/or temporal data. □

The taxonomy by Dodge et al. [Dodge et al. 2008] describes many spatio-temporal behaviors. Semantic behaviors can characterize semantic trajectories as well as raw trajectories if the latter can be linked to some contextual semantic objects.

Finally, a popular class of behaviors is the *Sequence* behavior, whose predicate involves a temporally ordered scan of the trajectory.

Definition: A *sequence behavior* is a trajectory behavior whose predicate specifies a sequence of component predicates that have to be satisfied in a specified temporal order. □

Typically a sequence predicate uses sequence operators such as THEN_NEXT (i.e., the next episode must comply with the predicate), THEN_LATER [d] (there must be, (some duration d) later within the trajectory, an episode (or a spatio-temporal position)

that complies with the predicate). An example of sequence behavior is: Begin in place P1, next cross area A2, and later stop in place P3. Several sequence definition languages have been proposed or used, see e.g. [Hadjileftheriou et al. 2005], [Mokhtar and Su 2005], [du Mouza and Rigaux 2005], [Gomez et al. 2008], [Laube et al. 2005a], [Sakr and Güting 2011].

3.3 Discovered versus Predefined Behaviors

Two types of research are relevant. On the one hand we have researches that, given a set of trajectories, aim at identifying which behaviors can be found in the set. No a priori behavior or application-defined behavior is assumed. Techniques for this kind of research typically include data mining, machine learning, and knowledge extraction in general.

On the other hand we have many researches that for a given problem domain use a set of behaviors that have been predefined by the experts in the domain. For example, Carey et al. [Carey et al. 2010] analyze behavior of flies in a cage by looking for six predefined behaviors: walking, moving, flying, feeding, drinking, and resting. Issues in these approaches basically relate to detection, quantification-evaluation, and visualization of the considered behaviors.

4. FROM GPS FEEDS TO “SOUND” RAW TRAJECTORIES

This section provides a generic view of the reconstruction process whose global aim is to turn the imperfect raw movement data into a trajectory data set that is correct and manageable from the viewpoint of the targeted application. More specifically, we elaborate on various approaches for trajectory reconstruction as the means for identifying *clean* (i.e. without/less noise), *accurate* (i.e. map-matched), and *compressed* (i.e. compact) trajectories, from the original sequence of spatio-temporal records (e.g. GPS records) of the moving objects. Let us name these trajectories “*sound*” raw trajectories. In this context, this section presents reconstruction methods, including techniques and algorithms, for data cleaning, map-matching and trajectory compression.

4.1 Trajectory Data Cleaning

In the literature of moving object databases, there is a common assumption that the spatio-temporal positions of moving objects can be precisely provided. Most trajectory query processing and indexing techniques are built upon this assumption. However, real-life trajectory data is far from being reliable enough for applications. Datasets collected by mobile sensors are often imprecise and incorrect due to noise [Yan et al. 2010]. This is mainly caused by the limitations of positioning systems (e.g. inaccurate GPS measurement and sampling errors, indoor signal loss, smart phone battery outage) or to concerns about the users' privacy protection (see Section 7). Therefore, trajectory cleaning cannot be overlooked when constructing sound trajectories from the GPS feeds.

The main focus of trajectory data cleaning is to remove GPS errors. [Jun et al. 2006] identifies two types of GPS errors: *systematic* errors, due to system's limitations, and *random* errors, due to external reasons. Systematic errors are totally invalid GPS positions, caused by horizontal dilution of position due to the low number of available satellites, while random errors are small errors up to ± 15 meters caused by the satellite orbit, atmospheric/ionospheric effects, and receiver issues. Both types of errors refer to the spatial positions. The temporal positioning is precise thanks to the high calibration clocks embedded in the satellites. For systematic errors researchers may resort to visual inspection, which obviously is impractical for large datasets. Usually they design automatic filtering methods to remove them. In this context, [Markatos et al. 2008] proposes an online approach that filters noisy positions by using the maximum speed of the moving object.

On the other hand, random errors are small distortions from the true values. Their influence can be reduced by smoothing methods (e.g. [Schussler and Axhausen 2009]). Such techniques can be categorized based on the statistical methods used. A first approach uses the least squares spline approximation that aims at minimizing overall errors. More specifically, this approach minimizes the sum of squared residuals; it is similar to regression-based smoothing such as the local polynomial regression, cubic fits, exponential smoothing, and time series approximation models. A second category relies on a kernel-based smoothing method, which adjusts the positioning of the moving object. This approach builds on the idea of nearest neighbor smoothing and locally weighted regression models. [Yan et al. 2010] proposes a Gaussian kernel based local regression model to smooth out GPS feeds. In this approach, a smoothed spatial position is the weighted local regression based on past and future positions within a sliding time window, using a Gaussian kernel weight. The last category smoothes the positions by recursively modifying error values using a Kalman filter. The filter uses measurements observed over time (the positions coming in the GPS receiver), and predicts positions that tend to be closer to the true values of the measurements [Lashley and Bevly 2007]. Eventually, the Kalman filter smoothes a position by computing a weighted average of the predicted position and the measured position. The objective of data cleaning by smoothing algorithms (e.g., Gaussian smoothing, Kalman filtering) is to decrease GPS errors in general, but the smoothing techniques cannot guarantee they can completely remove errors for every GPS point. In some cases, the smoothing algorithms may bring new random errors to the returned locations, especially when the trajectory data is sparse. Nevertheless, in general, the objective of data cleaning is to minimize the total number of errors for all GPS records.

4.2 Trajectory Map-Matching

The previous trajectory cleaning methods are designed for objects moving freely, without any constraint. However, in many applications moving objects are, by their nature, restricted to move within a given network that is represented as a graph [Güting et al. 2006]. For instance vehicles move on the road network, trains on the railway network. This type of constraints is spatial. On the other hand, frequent spatio-temporal constraints are speed constraints, e.g. a pedestrian cannot walk at a

speed above a certain limit. Temporal constraints are more common for animals' trajectories, e.g. usually bats don't fly during daytime. Constraints may bear not on the whole trajectory, but on specific types of episodes. For instance, a bus may stop only at locations registered as bus stops. On the opposite, constraints may bear on several trajectories, as it is the case for airplane trajectories that cannot intersect.

For network-constrained trajectories, map-matching methods may be used for two reconstruction levels: (i) cleaning the trajectory by replacing each position of the trajectory by the point on the network that is the most likely position of the moving object [Qudus et al. 2007] – this is for building “sound” raw trajectories, or (ii) transforming the raw trajectory into a semantic map-matched trajectory (i.e. a sequence of road segments) as defined in Section 2 – this is for semantic trajectory compression (see Section 4.3 for some details). From a computational point of view, map matching can be performed either online (the algorithm should be fast enough as new positions flow in real time) or offline (the algorithm is run when all positions are available). Both groups can be further classified as *geometric* or *topological* methods. Geometric methods utilize the underlying road network and various distance measures to determine the actual traveled roads. Distance measurements can be point-to-point (e.g. Euclidian distance), point-to-curve (e.g. perpendicular distance), or curve-to-curve (e.g. Fréchet distance). For instance, [Yin and Wolfson 2004] applies Dijkstra's shortest path algorithm [Douglas and Peucker 1973] to determine the distance between a trajectory and a sequence of arcs on a map. The route with the smallest distance from the initial trajectory is taken as the map-matched trajectory.

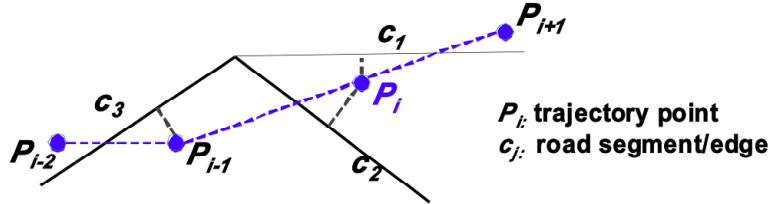


Figure 4 - The map-matching algorithm by [Brakatsoulas et al 2005]

Different from geometric approaches, the topological approaches avoid switching between unconnected road segments by using also the *connectivity* and *contiguity* of the road network. The method in [Yu et al. 2006] uses the correlations among the GPS heading, the road topology (e.g. U-turns, connections), and the road geometry (curvature). In the same context, [Brakatsoulas et al. 2005] proposes the following methodology: for every point P_i , given that point P_{i-1} has already been matched to an edge c_3 , the adjacent edges to this edge are the candidate edges to be matched to P_i . In Figure 4, P_{i-1} is matched to edge c_3 , hence c_1 , c_2 and c_3 , are the candidate edges for point P_i . Two measures are used for choosing among the candidate edges, applying *similarity* and *orientation* criteria, respectively. The higher the sum of these measures is, the better the match to this edge is. If the projection of the current point on the candidate edges does not lie on any of these edges, the algorithm does not proceed to the next point. Instead, the nearest edge of the candidates is set as part of the trajectory and then the next set of candidate edges is evaluated.

Recently, map-matching methods deal with the problematic case of GPS data with low sampling rate (e.g. one point every two minutes) [Lou et al. 2009] and high noise [Newson and Krumm 2009]. Such methods employ both distance and topology, for a better global algorithm. A global algorithm aligns an entire trajectory with the road network. Global approaches achieve better accuracy at a higher computational cost. In some cases, additional methods are used like probabilistic approaches. [Newson and Krumm 2009] uses Hidden Markov Model (HMM) to find the most likely road route corresponding to a sequence of positions. These advanced hybrid map-matching proposals usually include several post-processing techniques to calibrate and correct the initial matching results. Obviously this worsens the algorithm’s cost and efficiency. This important issue should be addressed by future researches.

4.3 Compression of Trajectory Data

Trajectory data in applications grow progressively and intensively as the tracking time goes by. For example, a taxi-tracking system (with 5,000 cabs in San Francisco) collects 7.2 millions GPS positions each day [Li et al. 2010b]. Such enormous amounts of data can sooner or latter lead to storage, transmission, computation, and display challenges. Therefore, trajectory data compression is an essential task of trajectory reconstruction. According to [Meratnia and de By 2004], the objectives of trajectory compression are: (i) to reduce the size of the dataset, (ii) to allow computations with acceptable/low complexity, and (iii) to support low deviation (by a given threshold) between the reduced trajectory and the original one.

From a geometric perspective, compression techniques exploit on line simplification that removes positions from a trajectory without warping the trend of the trajectory or distorting the data. In general, trajectory compression algorithms can be classified into four categories: *top-down*, *bottom-up*, *sliding window*, and *opening window*. The top-down algorithms recursively split the sequence of positions and only keep the key (“representative”) positions in each sub-sequence. A classical top-down method is the Douglas-Peucker (DP) algorithm [Douglas and Peucker 1973], with many subsequent extensions. The bottom-up algorithms start from the finest possible representation, and merge the successive points until some halting conditions are met. Sliding window methods compress data in a fixed window size; whilst open window methods use a dynamic and flexible data segment size. For instance, Meratnia and de By propose the Top-Down Time Ratio (TD-TR) and OPen Window Time Ratio (OPW-TR) algorithms for compressing spatio-temporal data [Meratnia and de By 2004]. The TD-TR approach uses the DP algorithm and, moreover, takes time into account. In particular, it replaces the Euclidean distance used in DP by a time-aware one, called Synchronous Euclidean Distance (SED), as illustrated in Figure 5. In this example, let P_b be the currently examined point against line P_1P_n . The DP approach uses the perpendicular distance of P_b to P_1P_n , while the TD-TR uses the distance of P_b to P_b' (i.e. the SED). The coordinates of the projected point P_b' are calculated using linear interpolation. OPW-TR applies the same time-ratio measurement but in an open window strategy.

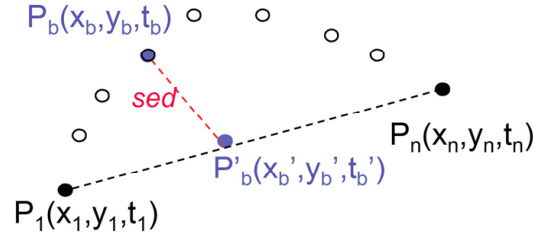


Figure 5 - Synchronous Euclidean Distance (SED) by [Meratnia and de By 2004]

[Potamias et al. 2006] proposes two online trajectory compression algorithms (*Thresholds* and *STTrace*). They use the coordinates, speed, and orientation of a position in order to calculate a safe area where the next position might be located. If the next incoming position lies in the calculated safe area, it can be ignored. There are two options for the definition of the safe area. It is either calculated by using the last position, whether it has been previously ignored or not, or by using the last chosen position. To achieve better results, the two algorithms can be combined: both areas are calculated, but only their intersection is defined as the safe area.

These trajectory compression approaches, which are primarily based on the extension of geometric methods, may be successfully applied to trajectories that move freely in space, but they are not suitable for network-constrained trajectories. Therefore, recent works like [Schmid et al. 2009, Kellaris et al. 2009] design another kind of trajectory compression models that make use of the underlying road network. Map-matching is applied for transforming the raw trajectories into semantic map-matched trajectories that are sequences of road segments. Therefore there is no more need for keeping the numerous original movement points. For instance, [Schmid et al. 2009] introduces a semantic-based representation model that characterizes significant trajectory points within a network. Key points are selected according to a combination of network, velocity, and direction criteria. In the same context, [Kellaris et al. 2009] replaces certain episodes of a trajectory by selected shortest paths between the beginning and ending positions of these episodes. Such semantic replacement of the raw GPS positions by specific road segments achieves significant data reduction.

5. SEMANTIC ENRICHMENT OF TRAJECTORIES

Adding application context data to raw trajectories, a process called semantic enrichment, is key for supporting mobility and behavioral analyses that are of interest to the application at hand. The mobility literature is rich in proposals for semantic enrichment approaches and techniques, each one built on the background of some popular application domain and focusing on the acquisition of knowledge specific to the domain. Particularly, human urban mobility is one of the domains that captured a lot of attention, given the current easiness to collect GPS data from devices installed in vehicles and in personal cell phones. Moreover, public concerns about city and traffic management largely motivate research in this domain.

A typical trajectory semantic enrichment process receives as input a set of sound raw trajectories, a contextual data repository and produces as output a set of semantically

annotated trajectories (in short, semantic trajectories). Trajectories can be annotated at different levels of detail: trajectory, episodes (meaningful parts of trajectories), and positions. Annotating trajectory positions is not efficient as it may generate a large number of repetitive annotations. Therefore annotation is rather performed at the episode and trajectory levels. Annotating episodes requires first identifying the episodes within the trajectory. The process that identifies the episodes is called *trajectory segmentation*.

5.1 Trajectory Segmentation into Episodes

Trajectory segmentation is driven by application dependent criteria. Whenever trajectory trips are not one shot (going straight from Begin to End), a very natural way of segmenting a trajectory is to split its path into periods of time when the object is considered as stationary and periods where the object is indeed moving. The former periods are denoted as *stops* while the latter are denoted as *moves*. According to this view, a trajectory is a sequence of alternating stop and move episodes, always starting with a move and having zero or more stops.

A popular challenge for researchers has therefore been to elaborate ways of identifying stops. The characterization of stop and move episodes depends on the application requirements, like for any trajectory segmentation. For example, in the tourist application no movement at all during some minimal length of time will catch the stops made while sleeping, eating, or seeing a show in a theater, while movement confined inside a small area during some minimal duration will also catch the stops made while visiting an exhibition. In Figure 6 the stops have been found by looking for slow movement in a small area that intersects a place of interest (POI). The following stops have been found: Eiffel tower, Palais Bourbon, Place de la Concorde, Le Louvre museum, restaurant Babylone. In fact, the characterization of a stop can imply no movement at all, slow speed, movement within constrained area, or proximity to some POI, to name a few.

5.1.1 Finding the stops

At first, stop identification has been performed on raw trajectories. A first approach simply associates stops with the parts of a raw car trajectory where either there is absence of signal (the GPS is off or the car entered a garage) or the velocity is zero for a given temporal interval (e.g. the car is parked) (see e.g., [Ashbrook et al. 2003]). The limitation of this approach lies in the possibility that an actual stop be not identified because, due to signal errors, the measured speed is slightly above zero. Later approaches use somehow more sophisticated techniques. An example is provided by the criteria used in [Krumm et al. 2006] to split a car movement track into trajectories. These criteria can similarly identify stops within a trajectory: 1) a five-minute gap indicating that the car was not moving, and 2) at least five minutes of very low speed (below two miles per hour), suggesting car stillness while the GPS has kept sending signals giving the same position with some approximation due to noise. The approach of [Andrienko et al. 2007] uses a larger temporal threshold (2 hours) to identify the important places visited by a moving person. The assumption is that the more time is

spent in a place, the more important is the place for the moving person. A similar assumption is used in [Zheng et al. 2011]: stops are computed as sequences of consecutive GPS positions such that their spatial distance is below a threshold while the temporal duration is above another threshold. CB-SMoT [Palma et al. 2008] proposes another approach: Stop episodes are computed based on the variation of the speed of the trajectory. The stops are those segments of the trajectory in which the speed is lower than the average speed of the trajectory. Another criterion to detect stop is the change of direction, as introduced in [Rocha et al. 2010]. This is the case of the fishing boats, where a stop for fishing is characterized by a sudden change of direction in contrast to the move where the boat goes straight.

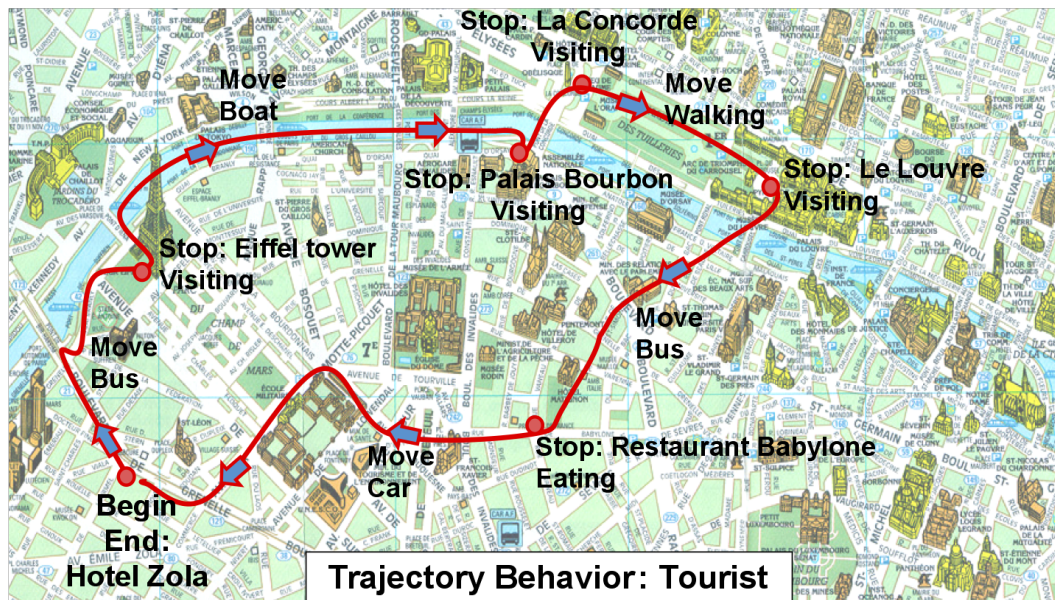


Figure 6 - A Semantic tourist trajectory with annotated stops and moves (map from Mappery.com, copyright unknown)

Other methods identify stops by using a combination of raw data, contextual geographic information and application information. For example, in the SMoT method [Alvares et al. 2007] a stop is a position where a trajectory holds on for a minimal time duration (or longer) and that corresponds to the position of one of the POIs defined by the application. Each stop is annotated by the corresponding POI. Rather than focusing on a single technique, Yan et al. developed a computing platform that supports the progressive construction of semantic trajectories from raw GPS data [Yan et al. 2011]. Raw trajectories are first preprocessed for cleaning and compression. The second platform layer supports the segmentation of the trajectories. Stops and moves can be computed by taking into account several spatio-temporal criteria like position density, velocity, and direction. The third layer is the semantic annotation layer with contextual data extracted from application sources and from geographical sources.

5.1.2 Other types of segmentation

Trajectories can be segmented with other criteria than stops and moves. For example, in [Xie et al. 2009] the authors use an enhanced geographical context to segment the trajectories according to the proximity to POIs of kind points. They define a trajectory episode as the segment of the trajectory whose positions are “influenced” by a given POI. A POI poi influences a position p of the trajectory if the distance between p and poi is smaller than the distance between p and any other POI. Each episode is annotated with the corresponding POI.

Another very popular criterion for segmenting a trajectory into episodes is the means of transportation used by the moving object, here a person. This knowledge is important for all planning applications in city transportation management. While different techniques are proposed for this segmentation, the common background is a characterization of each mode of transportation, basically in terms of typical speed (e.g. walking speed is less than 5km per hour), motion continuity (e.g., a bus makes stops while a taxi doesn’t), and direction and route constraints (e.g., the bus network uses only some defined routes while a tram network uses only rail tracks).

For example, [Liao et al. 2007] uses a Gaussian mixture model based on the speed, which is classified into three speed ranges: walking, high speed and low speed. Therefore, a new episode is created whenever a switch in the speed range is detected. A similar approach is developed in [Zheng et al. 2010]. Using speed, acceleration and speed change rate, the authors first detect the positions where the movement switches between walking and non-walking. In a second step, they refine the non-walking segments into segments characterized by the other transportation modes: bicycle, bus, and driving. They use a combination of techniques, from supervised learning to decision tree inference, and add a post-processing step to improve the accuracy of the segmentation. The post-processing step relies on a graph that contains commonsense constraints about the real world and typical human behaviors.

Beyond using raw data and its derived features (e.g. velocity, acceleration), the approach in [Yan et al. 2011] uses also semantic information such as road categories (e.g., using a bike path indicates that the transportation mode is either bike or walking, definitely not car, bus, or train) and public transport networks (e.g., the location of bus stops and train stations can determine whether a trajectory segment is traveled by bus or by train).

5.2 Episodes Annotation

As already stated, usually each episode is annotated with its defining annotation. For instance, Stop/Move episodes are annotated as “stop” or “move” episodes, transportation mode episodes are annotated as “walk”, “bus”, “bike”, or “car”. POI-influence episodes are annotated with the identifier of the corresponding POI. However, episodes may also bear many other annotations to fulfill application requirements. A frequently quoted example is stop episodes, which may be additionally annotated with application contextual information, such as the identifier of the most likely POI where the moving object stopped, or its type (e.g. home, workplace, restaurant), or some other application specific annotation (e.g. the activities

pursued during the stop). For instance in Figure 6, the stop episodes have been annotated with two annotations, POI and activity, and the move episodes with a transportation means annotation.

5.2.1 Annotating stop episodes with places of interest (POIs)

Stopping during a trip means that there is something of interest to do at or nearby the stop. The existence of POIs is therefore inherent to trajectory analyses. Analysts want to know which POI motivated the stop.

Ashbrook et al. annotate stops with the POIs that are “near” the stop, i.e. within a given radius defining a spatial buffer around the stop [Ashbrook et al. 2003]. The use of the buffer relies on the assumption that the GPS signal is subject to measurement errors, therefore the visited POI is found by spatially generalizing the actual stop position. Instead, Palma et al. annotate a stop with the POI located at the stop [Palma et al. 2008]. Stop episodes that do not intersect any POI are labeled as unknown stops. Spinsanti et al. propose a more sophisticated method that relies on additional information about the POIs including common sense constraints [Spinsanti et al. 2010]. They annotate each stop episode with the list of POIs likely visited by the person. POIs are selected based on two criteria: they must be at reasonable distance (e.g., 15 minutes walking) from the stop (assumed to represent the position where the moving person parked her car), and the duration of the stop episode must comply with the standard duration specified for a visit to the POI (e.g., for a restaurant POI a 1-2 hours duration is compatible at lunch or dinner time, a museum POI requires at least 1 hour for a visit).

In [Yan et al. 2010] the authors extend their stop annotation approach to a more general technique that supports the annotation of stops with regions of interest and the annotation of move episodes with the streets that correspond to the path of the moving object. Yan et al.’s technique is in fact a set of methods that allow annotating episodes with three kinds of geo-objects, regions, points and lines. The methods are based on spatial joins and also use direction, distance, and topological spatial relations such as intersection. Annotating with regions relies on the identification of Regions of Interest (ROIs). These can be defined a priori by the application or derived from an analysis of trajectory. For example, [Uddin et al. 2011, Giannotti et al. 2007] detect ROIs defined as the more visited - or passed through – areas. However, while these works detect find ROIs they do not use them for trajectory annotation.

Andrienko et al. do not annotate the episodes of a trajectory but focus on extracting and analyzing their Begin and End in view of characterizing individual movement behavior [Andrienko et al. 2007]. Trajectory Begin and End are annotated with the identifier or type of the POI that corresponds to the places visited by the user of the moving car. The authors discover the POI or its type by analyzing the temporal profiles of visits and using background map and knowledge about the territory.

5.2.2 Annotating stop/move episodes with activities

In [Liao et al. 2005] the authors annotate stop episodes with the activities carried out (e.g. AtHome, Shopping) and the transportation mode for the move episodes. The

inference of activities is based on their characteristic features: (1) temporal patterns, since different activities have different temporal durations; (2) POIs, which play a central role since the location of the stop can be related to a number of business activities; (3) transition relations, since one activity may or may not be followed by another (i.e. “stay at home” is typically followed by “being at work”); (4) spatial constraints, since activities in the same place are often similar, and finally (5) global features representing soft constraints on people activities, such as the number of “lunch” activities possible per day.

The literature is rich in publications dealing with activity annotations. In particular, many ecology studies proceed through manual annotation of activity episodes as directly observed by the scientists in the field (see for example [Cagnacci et al. 2010]). The set of possible activities is predefined and the study is about the frequency distribution of these activities among the observed animals. Activity models are mainly application-specific.

5.3 Annotating the Trajectory

Annotating the trajectory as a whole assumes that all the information in the trajectory (in particular, all its episodes) is somehow synthesized into a single label that characterizes the whole trajectory. For instance in Figure 6, the trajectory has been annotated with a Behavior annotation whose value “Tourist” was inferred from the value of the activity annotation of its stops.

The work of [Spinsanti et al. 2010] proposes a series of rules for the annotation of the trajectory with the overall activity expressed by the trajectory data. Considering a trajectory whose stop episodes are annotated with possible activities and common sense IF-THEN rules, the authors annotate the trajectory with the most probable global activity. For example, if a trajectory is annotated with home and work stops, the entire trajectory is annotated as *home-to-work*. Similarly, trajectories may be annotated as *tourism* or *shopping* depending on the fact that most stops happen in tourist places or in shopping malls. An interesting feature of this approach is that it uses a number of intuitive common sense rules to infer a possible behavior annotation for the trajectory.

A similar but richer (because it considers both stops and moves) approach is represented by the work of Baglioni et al. [Baglioni et al. 2012]. This work automatically annotates trajectories with activities (e.g., tourist, home-to-work) by using an ontology and an inference engine. The trajectories are semantic trajectories with stop and move episodes. The ontology formally defines the concepts of trajectory, stop, move, trajectory activity, as well as contextual information (e.g. TouristAttraction, AccommodationPlace). The various activities are defined by ontology axioms, i.e. logical expressions built over trajectory features. For example, a “tourist” activity may be defined as a trajectory stopping at accommodation places at night and at tourist attractions during the day. The technique used for annotating trajectories with activities is also used to enrich the knowledge about the moving object itself. For example, a trajectory frequently stopping at monuments, museums and restaurant is probably characterizing a tourist. In this case, the annotation is twofold. On the one

hand, the trajectory itself shows a *tourist* activity; on the other hand the moving object may be annotated as a “tourist” person.

6. EXTRACTING BEHAVIORAL KNOWLEDGE FROM TRAJECTORIES

Knowledge discovery from trajectories aims at identifying behaviors, either among individual trajectories or groups of trajectories. In a data mining perspective, contributing techniques include: (i) clustering trajectories sharing similar characteristics as shape, speed, direction, or being positioned at short distance one from the other; (ii) classifying trajectories into predefined classes (e.g. fishing ship trajectories or cargo ship trajectories), (iii) discovering common sequences of elementary movements (e.g. going from A to B then to C), and (iv) identifying trajectories of objects whose movement is somehow related to each other, in particular, objects moving together (e.g. a platoon of soldiers, flocks of animals).

Most works for extracting behavioral knowledge from trajectories use clustering techniques (see sub-section 6.1), and operate on raw trajectories. Some of these works were presented in [Nanni et al 2008], that surveys the general spatio-temporal data mining literature (not only trajectories) up to 2007. In this section we first recall the main aspects of clustering and classification relevant to movement analysis. Next we recall the most significant results of traditional techniques for trajectory knowledge discovery. In sub-section 6.3 we move into knowledge discovery for semantic trajectories, the newest research challenge that has received significant attention. Several workshops and research projects have emerged on this topic, e.g. the SEEK - SEmantic Enrichment of trajectory Knowledge - European project [SEEK 2012].

6.1 General Characteristics for Trajectory Knowledge Discovery

The discovery of common behaviors among either individual trajectories or groups of trajectories is based on similarity evaluations that allow defining meaningful classes of trajectories. Two machine-learning techniques are very popular in this context: clustering and classification.

Clustering is a well-known generic unsupervised learning technique that is applied to determine relevant groups of data (trajectories in our case) from the analysis of a large data set, without any a priori knowledge of the targeted grouping. The groups are computed based on some characteristics that should be shared by the trajectories either along the whole trajectories or along some sub-parts of the trajectories. Most commonly, the computation of similarity is based on a **distance** measure. In spatial clustering and in temporal clustering, the usual distance measures are spatial distance and temporal distance. In trajectory clustering, these distances may be applied as well. For example, if a trajectory *A* occurred in the morning and a trajectory *B* occurred in the afternoon, these trajectories may be considered temporally distant, while two trajectories occurring in the same hour may be considered temporally close. However, the real challenge is the definition of a spatio-temporal distance that would better correspond to the spatio-temporal nature of trajectories. For instance, [Pelekis

et al. 2011a] defines a method that groups trajectories using various distance functions based on motion properties such as spatial location, speed, acceleration and direction. The method evaluates distances over the whole trajectories. In contrast, Lee et al.'s clustering algorithm looks for similarity between segments of trajectories [Lee et al. 2007]. They do not require that similarity hold for the whole trajectory. The advantage of using segment similarity is the possibility to fine-tune the grouping of trajectories. For example, a group of tourists in Paris spend the morning together: They meet at Place de la Concorde, walk up to Le Louvre, stop there for visiting and lunch. After that they split, each one going on his own. Their trajectories are considered similar based on the similarity of the morning segments.

When the trajectories are not free but constrained by a network, like a road network for cars' trajectories, new density and distance measures have to take into account the specificity of the network. Usually, distance measures are no longer based on the Euclidian distance but on algorithms for finding the shortest path in a graph [Tiakas et al. 2009]. In [Han et al. 2012] the authors find clusters of segments of cars' trajectories by looking for sequences of contiguous road segments that are followed by continuous traffic flows.

Trajectory clustering has been used in several contexts as trajectory searching and querying [Panagiotakis et al. 2012], trajectory visualization [Rinzivillo et al. 2008], and processing of trajectory uncertainty (e.g. missing points) [Pelekis et al. 2011b].

Classification is a complementary supervised learning technique that determines the appropriate grouping when the targeted classes are defined a priori. For example, classification may be used to group trajectories of ships into three predefined classes: trajectories of *fishing* ships, trajectories of *passenger* ships, and trajectories of *cargo* ships. Lee et al. proposed one of the first works applying classification to trajectories [Lee et al. 2008]. As discussed in the previous paragraph, the authors work on trajectories partitioned into segments. After using clustering to find the interesting segments, they define a class for each cluster and apply classification to populate the classes. Their example about ship trajectories is shown in Figure 7. All trajectories go from Port A to Port B. The classes defined via clustering are: container ships, tankers and fishing boats. At this point, it is not known to which class each trajectory belongs. By analyzing specific parts of the trajectories (those segments showing trajectory dissimilarity) the method is able to classify the trajectories into the chosen classes. The first class holds trajectories going from Port A to Port B by making a detour (similar segments) through the container port, the second class holds trajectories making a detour through the refinery, and the third class holds trajectories making a detour for roaming in the fishery area. This approach does not take time into account, only space.

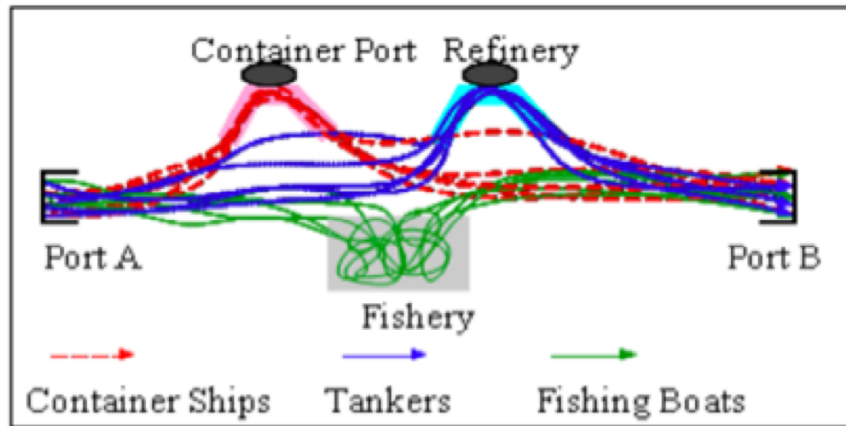


Figure 7 - Example of trajectory classification [Lee et al. 2008]

6.2 Non-Semantic Trajectory Knowledge Discovery

A critical review of the literature on trajectory behavior analysis is available in [Laube 2009], while [Wood and Galton 2009b] provides a detailed analysis of collective behaviors.

Two classes of trajectory behaviors have been frequently investigated. On the one hand researchers focused on the dynamics of single trajectories in terms of which regions they traversed. This leads to the definition of Sequence behaviors. On the other hand, researchers focused on characterizing behaviors implying the participation of several moving objects and showing specific interactions or coordination. A typical example is the Flock behavior.

For the first class, works mainly focused on the sequence of regions traversed by the trajectories. Cao et al. first partition the raw movement of an object into trajectories of the same duration [Cao et al. 2007]. Then they analyze the resulting set of trajectories in order to find behaviors of kind Sequence, which are defined as ordered lists of regions that are frequently repeated in the trajectories. The authors call these behaviors "periodic patterns". A region can be a spatial location defined by the user or a location computed by the algorithm. For a region to be part of a Sequence behavior it has to be a dense area, i.e., an area that contains a minimal number of trajectory positions and the object must visit this area several times. This number of times is application dependent (e.g. in a criminal application, visiting a region two or three times may be enough for being considered as a suspect, while in a urban planning application a moving object is considered a worker only if he visits the same region at least four times a week).

Giannotti et al. follow the same goal, identifying Sequence behaviors, but the trajectories they consider do not need to have the same duration and belong to different moving objects [Giannotti et al. 2007]. The algorithm they propose for extracting Sequence behaviors from groups of trajectories computes a sequence of regions, frequently visited in a specified order and with similar transition times. The interesting regions can be either defined by the user or computed by the method.

Interesting regions (ROIs) are computed based on the density of trajectory positions in space, not in time. This step uses clustering techniques based on density. Having identified the ROIs, the method filters the trajectories that intersect these regions and then computes the travel time duration of these trajectories between the regions. A Sequence behavior is generated if the sequence of regions is visited by at least a minimal number of trajectories. [Monreale et al. 2009] extends Giannotti et al.'s work by using the discovered Sequence behaviors for predicting the next locations that the moving objects will visit.

[Li et al. 2010a] proposes an algorithm to find periodic behaviors of moving objects. A periodic behavior is loosely defined as “the repeating activities at certain locations with regular time intervals”. Their basic assumption is that periodic behaviors are associated with a corresponding frequently visited ROI. ROIs are semantically meaningful locations, e.g., home, office, etc. Interesting ROIs are identified based on density of trajectory points. By confronting the ROIs with the trajectories, they detect the time periods spent by moving objects within the ROI. They use clustering to identify significant periods and finally derive the periodicity that characterizes periodic behaviors over these periods.

Regarding behaviors of groups of objects, a frequently quoted representative is the work developed by Laube et al. over a series of papers, since [Laube et al. 2005a]. The authors proposed the collective trajectory behaviors Flock, Leadership, Convergence and Encounter. These behaviors are well explained in [Nanni et al. 2008]. A variant of the Flock behavior is the Moving Cluster behavior [Kalnis et al. 2005]. It denotes a group of objects that move together, but, contrarily to the flock, the membership into the group is variable [Wood and Galton 2009a]: Members may leave the group and new members may join the group.

As the behavior discovery approaches that deal with raw trajectories in general look for common behavior in dense regions, the discovered behaviors are not as rich and meaningful as those extracted from semantic trajectories.

6.3 Semantic-based Trajectory Knowledge Discovery

In every application domain, using contextual and semantic information enriches the meaning of the behaviors, helps understanding the behaviors, and may reveal behaviors that would be difficult to identify without using semantics. For example, a behavior of a group of objects defined as “converge in the morning to the city centre” bears more information than a convergence behavior expressing that a group of objects “converge at the same time to the same place”. Behaviors as the former can only be discovered when geographic information or domain knowledge is considered. Semantics-based behavior discovery techniques can be divided in two main groups: (i) approaches searching for common behaviors that are previously unknown, and (ii) approaches looking for specific behaviors.

6.3.1 Discovering Unknown Behaviors

For knowledge discovery in a domain where location is inherent to the data, as in trajectories, the integration with contextual geographic information plays an essential

role towards the discovery of easily understandable behaviors. Considering the semantics of space (e.g. the place where a trajectory has passed) and the semantics of time (e.g. morning, afternoon or weekday instead of a timestamp) gives more meaning to a behavior.

Figure 8 (Left) shows an example of analyzing raw trajectories and discovering the spatio-temporal behavior *Meet*: A group of trajectories end at the same region C. If we use semantic trajectories with stops annotated with POIs (school A, school B, ..., cinema Lux), as in Figure 8 (Right), we can discover the frequent semantic behavior “Going from school to cinema on Wednesday afternoon”, which corresponds to the fact that this cinema offers special price tickets for students on Wednesday afternoons.

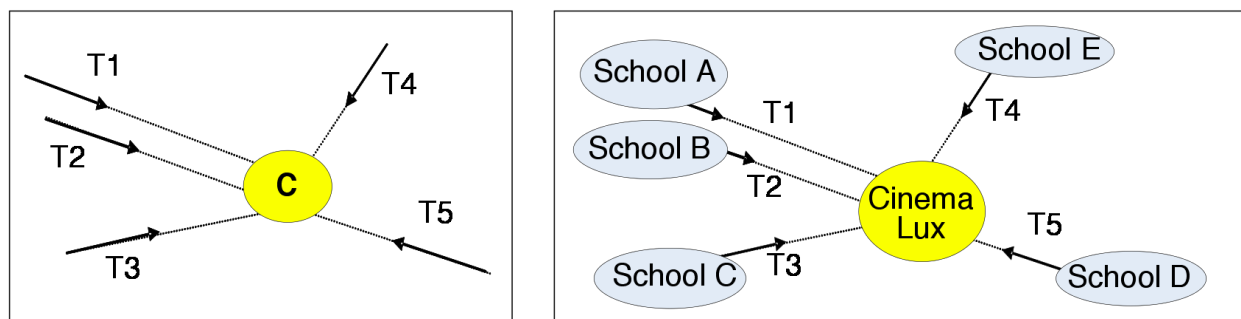


Figure 8 - (Left) Spatio-temporal behavior *Meet*. (Right) Semantic behavior *Going from school to cinema on Wednesday afternoon*

For discovering a semantic behavior on a specific day (e.g. *Wednesday*) or period (e.g. *rush hour*), the original temporal information (a time interval) and the spatial one (here a POI instance) have to be transformed into the appropriate granularities (e.g. $\langle \text{Wednesday afternoon, school} \rangle$) [Bogorny et al. 2009]. For example, the schools could be represented by instances of school (e.g., school A or school B) or by kinds of school (e.g., high school or primary school, or, at a more generic level, teaching institution). Multi-levels (or multiple granularity) mining and granularity transformation are well-known issues in knowledge discovery. Without considering the semantic information of space and time at different granularities, the semantic behavior *Going from school to cinema on Wednesday afternoon* would not emerge.

It is important to observe that the integration of trajectories with contextual and semantic, spatial and temporal, information is vital for the discovery of meaningful behaviors. Behaviors extracted from semantic trajectories cannot be obtained from raw data only. For instance, in the example in Figure 8, approaches dealing with raw trajectories would find that trajectories meet to C, and annotating a posteriori would just find C as a *cinema*. Annotating trajectories before mining allow revealing the movement behavior *from schools to cinema at a specific weekday and period*.

The Semantic Trajectory Data Mining Query Language (ST-DMQL) [Bogorny et al. 2009] allows users to specify semantic enrichment of trajectories with contextual domain information, as presented in Section 5. It also allows users to specify semantic behaviors that are to be mined in the set of semantic trajectories. The users specify the behaviors by defining at some chosen granularity level the characteristics that the

stops must show: the POI instance (School A, Le Louvre museum) or the POI type (e.g. school, museum) that annotates the stop, and the temporal interval of the stop or one of its generalizations (e.g. weekday-morning or weekday). Three kinds of semantic behaviors can be mined: Visiting sequences of places, visiting the same places with a certain frequency but in any order, and visiting associations of places (e.g. trajectories that visit museums do also visit religious places). The language is implemented in Weka-STPM [Bogorny et al. 2011], an extended version of the Weka data mining toolkit and the first toolkit for multi-level mining of semantic trajectories. It is a free and open source tool that also provides spatial visualization of the semantic trajectories and behaviors.

Another tool that analyses semantic trajectories to infer behavior is M-ATLAS [Giannotti et al. 2011]. This system provides support for both raw and semantic trajectories, and is organized in a plug-and-play architecture that allows the easy integration of different mining algorithms, from clustering to classification techniques. The algorithms can be combined using a data mining query language, thus enabling the discovery of several kinds of behavior. Furthermore, the integration with the Athena system [Baglioni et al. 2012] allows annotating trajectories with a behavior such as home-to-work, commuter or tourist.

6.3.2 Discovering Specific Behaviors

Two recent works focus on defining new kinds of individual trajectory behaviors, *Chasing* and *Avoidance* [Alvares et al. 2011]. Even if in fact these behaviors rely only on spatio-temporal criteria, we present them in this section 6.3 because they describe well-known complex "semantic" activities performed by human or animals. [Siqueira and Bogorny 2011] formally defines the *Chasing* behavior and provides an algorithm that evaluates if an individual (a person or animal) called the *stalker* intentionally follows another individual called the *target*. The *stalker* must follow the *target* for a certain time period, and during this period, the movement of the two individuals must remain with similar speed and direction. Moreover the target must always be in front of the stalker.

[Alvares et al. 2011] defines the *Avoidance* behavior and presents an algorithm for identifying the trajectories that avoids a static object. For example, when analyzing human trajectories an avoidance of street cameras may reveal a suspect behavior; in animal monitoring, an animal avoiding a given area may suggest a territoriality behavior. A trajectory shows an *Avoidance* behavior when it moves towards a target geographic object, turns around this object without intersecting it, and after avoiding the target object, the trajectory returns to its original path. This approach distinguishes between intentional (e.g. a suspect avoiding a security camera) and natural (i.e. a blocked street) avoidance.

The work of [Cao et al. 2010] presents an interesting method for mining behaviors from semantic trajectories. The first step finds the stay points, a concept very similar to the concept of stops. Stay points are the gaps in trajectories where, for instance, a GPS is turned off. The last position of the trajectory before the GPS is turned off and the first position when the GPS is turned on represent the starting and ending

positions of a stay. The temporal duration of the gap must be longer than a given threshold to be a stay point. In the second step, the location of the starting position of the gap is mapped (using Google) to a street address and, by looking in the yellow pages, the street address is transformed into a POI. Then the stay point is annotated with the POI (e.g., if the address of the trajectory position corresponds to a hotel, this stay point will be annotated with the name of the hotel). According to Cao, the yellow pages service returns a list of semantic locations that are near the queried street address, and one of them may match exactly the given street address of the stay point of the trajectory. The last step generates the list of places visited by the trajectories, based on the duration of the visits and the frequency of the visits. This list is called the location history.

7. PRIVACY

Whenever the trajectory positions refer to the personal sphere of individuals, e.g., where individuals live and work, the places they visit and the people they meet, they are to be considered as personal information and protected against unauthorized disclosure. Several countries worldwide restrict by norms and laws the collection and use of personal information. Privacy regulations, however, cannot protect personal information against malicious parties, i.e. parties deliberately trying to access protected data. Specific techniques, known as *privacy-enhancing technologies* (PETs) have been developed to block malicious accesses and avoid violation of individuals' privacy.

This section overviews recent research on PETs that is relevant for the protection of privacy in semantic trajectories. Semantic trajectories magnify the risk for privacy because behavior information on individuals is explicitly extracted and represented in a machine-readable form, therefore can be used within information processing applications and easily unfolded to third parties. For this reason, semantic trajectories and privacy clash.

Recent research attempts to mitigate this conflict, by forestalling the extraction of patterns conveying sensitive information such as visits to hospitals and religious places, and presence at specific locations revealing data about for example sex life. We refer to such patterns as sensitive stops. The approaches addressing the protection of sensitive stops can be classified in two categories based on whether techniques are applied before stops are recognized, therefore during the process of semantic trajectory construction, or, alternatively, after the semantic trajectories have been built. In order to better characterize those techniques especially with respect to existing solutions, we first introduce a general framework for the classification of position PETs. Subsequently, we discuss the technical challenges posed by the protection of sensitive stops and illustrate recent research directions.

7.1 Characterizing PETs for Position Privacy

Research on privacy of position data took off with the emergence of mobile applications based on stored people's tracks [Gruteser and Grunwald 2003, Beresford and Stajano 2003]. The identification of basic characteristics of position PETs enables easier

analysis and comparison of PET's methods and the issues they address. The following are helpful criteria.

1. **Data model.** Some PETs are designed to protect single trajectory positions (as in streaming, see below), while other PETs (as in data publishing) deal with whole (raw or semantic) trajectories. In PETs caring about POIs, POIs are equipped with more or less semantic information. PETs may indeed consider POI's spatial location only, or their location and their type, or their location and type combined with other semantic information (e.g. population density).

2. The **privacy goal** defines the purpose of the privacy protection mechanism. The following classification of goals is inspired by [Jensen et al. 2009]:

- i. *Identity privacy.* The position of a moving person can help revealing the his/her identity in situations in which persons should be anonymous. For example, if one can recognize that a certain position is the person's home, the person's identity can be easily discovered.
- ii. *Location privacy.* The position of a moving person may represent sensitive information. It then belongs to the set of sensitive attributes. For example, disclosing that the person is at a certain position in the evening may endanger the physical safety of the individual.
- iii. *Both identity and location privacy.* For example, revealing that one is at home not only can reveal the identity of the person but also the fact that he/she is at home (instead of being at work, for example).

3. The **application context** in which the position must be protected. We emphasize two application contexts:

- i. *Data publishing.* The data publisher is commonly trustworthy. People therefore are willing to provide their position data to the data publisher. The trusted data publisher has full knowledge of persons' positions and is allowed to modify the data in order to guarantee privacy before publishing the data. In this case the privacy goal is primarily identity privacy. For example, a geo-marketing company sells data about the routes of vehicles. To comply with privacy regulations the company anonymizes the data before publication.
- ii. *Streaming.* An untrusted party (e.g., a LBS provider, a geo-referenced social network) gathers trajectories of persons according to a client-server model. The protection is enforced on the fly while the stream of positions is sent to the untrusted party. Specialized components, running on clients or trusted third parties ensure the privacy of the stream of positions by transforming each position, one by one as they flow in, into a privacy-preserving position (see e.g. [Gruteser and Grunwald 2003], [Chow et al. 2009]). Two characteristics of streaming approaches are: The PET has a very incomplete knowledge of the person's trajectory (usually only the current position), and the technique has to be very efficient.

In the light of this classification, techniques designed to forestall the extraction of sensitive stops, present the following characteristics: They rely on privacy data models based on POIs, i.e. the presence of a person in a certain POI is interpreted as sensitive stop; the privacy goal includes location privacy. In the following, we survey existing solutions in streaming and data publishing applications, respectively.

7.2 The protection of sensitive stops in streaming applications

Conventional position PETs in streaming applications are not capable of preventing the disclosure of sensitive stops. For example, policy-based techniques allow their users to define enforceable rules specifying which positions are to be disclosed to position recipients, when and to whom (e.g., "John's position can be only disclosed to Alice in daily hours"). These techniques have their roots in access control policies, and in the bulk of work developed at the end of the nineties for privacy protection in e-commerce applications. Yet, these techniques do not offer any protection against untrustworthy position recipients, e.g. LBS providers, which can jeopardize the confidentiality of position data.

A different approach is to mask the association between the identity and the position of a person (i.e. to achieve identity privacy). For example, *location k-anonymity* (i.e. an extension of the k-anonymity paradigm [Sweeney 2002]), postulates that the exact position of a person must be replaced by a coarse, *cloaked* region containing k persons. This way, the third party, specifically the LBS provider, cannot univocally identify persons exclusively based on position information. While this technique is effective in protecting the association between person's identity and LBS request, it does not prevent the disclosure of sensitive stops. In fact, the k persons can be grouped together within a cloaked region falling inside a place denoting a sensitive stop, e.g. hospital.

Moreover, the techniques that supplement location *k-anonymity* with *location l-diversity* (i.e., an extension of the concept of l-diversity [Machanavajjhala et al. 2006]) are not helpful [Bamba et al. 2008, Wang and Liu 2009, Xue et al. 2009]. Location *l-diversity* requires that every cloaked region containing *k* individuals contains at least *l* different POIs. The degree of diversity is measured by counting the number of POIs occurrences or types inside the cloaked region, while the different degree of sensitivity of those places is ignored. In essence, an l-diverse region "blurs in the crowd" the POIs where the person can stay, independently from the meaning of those places. The result is an excessive loss of position accuracy. Moreover, there are situations in which individuals are not anonymous [Duckham and Kulik 2005], in particular in the context of geo-social applications [Ruiz Vicente et al. 2011], therefore these methods cannot be applied.

A first step towards an effective protection of sensitive stops is offered by the *semantic location cloaking* paradigm [Damiani et al. 2011]. This paradigm is grounded on the idea that positions may exhibit a different degree of sensitivity depending on the context, in particular the POI (or place) in which the moving person is located. For example, being on a street is likely to be less sensitive than being in a hospital. Accordingly, only those positions that are perceived as sensitive are protected by

degrading the accuracy of position information while the others that are not sensitive are disclosed at the finest granularity. For semantic location cloaking to be sound it must satisfy the following requirements:

- **Semantic diversity.** The person's position cannot be blurred exclusively when it is inside a sensitive place because in that case one could immediately infer the actual POI. Rather, a cloaked region must include POIs of diverse types, both sensitive and innocuous. That way, the place in which the person is located remains uncertain.
- **Independence** of the cloaking method from the position of the person. This condition prevents the discovery of the correlation between the cloaked region and the true position. Such information could be leveraged to infer where the user is located.

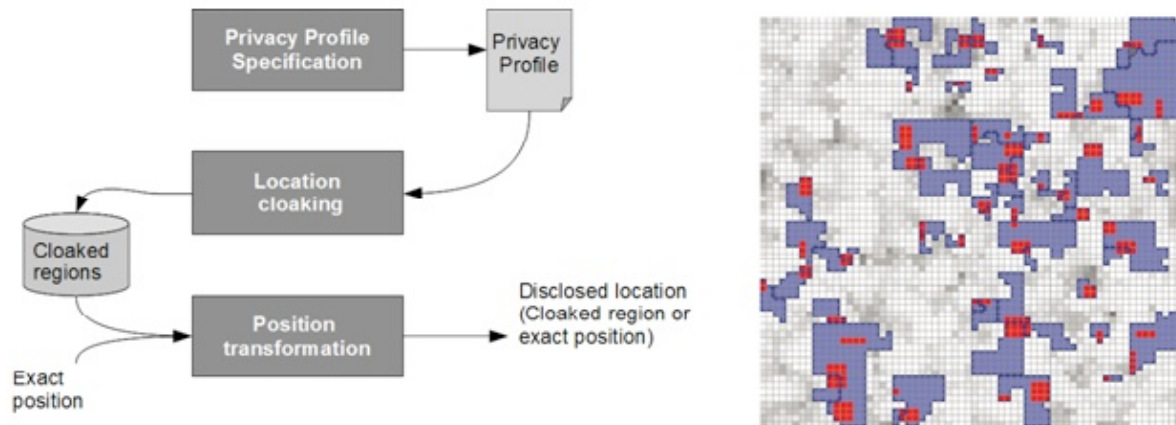


Figure 9 - The Probe system: (Left) The workflow. (Right) Obfuscated map: The blue polygons represent cloaked regions, the red rectangles sensitive POIs; the grey background the distribution of population in space

These guidelines have been embodied in the privacy-preserving framework called Probe (Privacy-aware Obfuscation Environment) [Damiani et al. 2010]. Figure 9 (Left) illustrates the workflow of the position cloaking process. Users first specify in a privacy profile which categories of POIs are sensitive (selecting for example from a pre-defined list, e.g. hospitals, religious buildings) along with the degree of privacy desired for each of those categories. For example a privacy degree of 0.1 assigned to hospitals means that the probability of locating the user inside a hospital must be less than 0.1. Next, cloaking algorithms generate coarse regions satisfying the privacy preferences, independently from the persons' positions, in order to prevent possible inferences on their reciprocal positions. Background knowledge on positions distribution is considered. A sample set of cloaked regions is shown in Figure 9 (Right). Finally, at runtime if the user's position falls inside one of the coarse regions, that region is delivered instead of the exact position. Recent research extends semantic location cloaking methods to the case in which users' movement is confined to road

network [Yigitoglu et al. 2012]. In this case the cloaked region takes the form of a subgraph of a semantically annotated graph representing the urban setting.

7.3 Anonymization of Trajectories in Data Publishing

Trajectory anonymization approaches aim at the protection of the identity and/or the position of moving persons in published personal trajectories. They commonly assume the following publishing framework: a trusted data owner, such as a telecom operator, collects the trajectories for a large number of persons. The data owner wishes to share these trajectories with potentially untrusted parties in order to support various analytic tasks. To achieve this in a privacy-preserving way, the data owner applies some trajectory anonymization technique to the original trajectories to effectively conceal the identity and/or the positions of the persons prior to releasing the data. Several approaches have been recently proposed to anonymize trajectories. These approaches can be partitioned into two broad categories: (i) those that anonymize each trajectory as a whole, without assuming specific background knowledge of the malicious observers (i.e., attackers) [Abul et al. 2008, Mohammed et al. 2009, Nergiz et al. 2008], and (ii) those that consider attackers who can effectively link specific (location, time) pairs to persons, in order to uniquely re-identify them [Terrovitis and Mamoulis 2008, Yarovoy et al. 2009]. Approaches in the first category offer increased privacy as they can forestall more powerful attacks at the expense, however, of data utility.

These approaches operate by grouping the trajectories into clusters of k members so that each trajectory within a cluster is indistinguishable from the other trajectories in the same cluster. When a potential attacker asks for a trajectory, the PET answers by returning the whole cluster instead of the trajectory alone. For example, NWA [Abul et al. 2008] anonymizes trajectories by generating cylindrical volumes that contain at least k trajectories.

The second category of approaches is specific to semantic trajectories. It considers attackers who have background knowledge about ordered sequences of POIs that were visited by specific persons, whom they wish to re-identify. Terrovitis and Mamoulis [Terrovitis and Mamoulis 2008] consider trajectories of POIs and protect the identity of the moving persons from these attackers. Figure 10a shows a set of trajectories $t_1 \dots t_8$ that need to be anonymized and Figure 10b shows the assumed knowledge of an attacker. As one can notice, the attacker has the knowledge to associate persons to POIs a_1 , a_2 and a_3 . Using his knowledge, the attacker can, for example, uniquely re-identify t_5 because t_5 is the only trajectory in this database that passed from a_1 to a_3 . The goal of the proposed anonymization approach is to prohibit the attacker from associating a sequence of POIs he knows to less than k individuals. To achieve that, the authors propose a suppression technique that aims at removing the least number of POIs from the persons' trajectories so that the remaining trajectories are k -anonymous with respect to the knowledge of each considered attacker.

(a)

t_{id}	trajectory
t_1	$a_1 \rightarrow b_1 \rightarrow a_2$
t_2	$a_1 \rightarrow b_1 \rightarrow a_2 \rightarrow b_3$
t_3	$a_1 \rightarrow b_2 \rightarrow a_2$
t_4	$a_1 \rightarrow a_2 \rightarrow b_2$
t_5	$a_1 \rightarrow a_3 \rightarrow b_1$
t_6	$a_3 \rightarrow b_1$
t_7	$a_3 \rightarrow b_2$
t_8	$a_3 \rightarrow b_2 \rightarrow b_3$

(b)

t_{id}	trajectory
t_1^A	$a_1 \rightarrow a_2$
t_2^A	$a_1 \rightarrow a_2$
t_3^A	$a_1 \rightarrow a_2$
t_4^A	$a_1 \rightarrow a_2$
t_5^A	$a_1 \rightarrow a_3$
t_6^A	a_3
t_7^A	a_3
t_8^A	a_3

Figure 10 - A trajectory database (a) and the assumed knowledge of an attacker (b).

A more recent anonymization approach extends the l -diversity concept to the case of persons' trajectories [Monreale et al. 2011]. In this paper, the authors aim to protect moving persons from adversaries who can infer that they have visited a sensitive POI (e.g., a hospital), when the adversaries know a set of non-sensitive POIs (e.g., a park and a café) that the persons have visited. The main characteristic of this work is that the anonymization takes place at the semantic level instead of the spatial level: Instead of replacing points by cloaked areas, a hierarchical ontology of POIs allows replacing the POIs of the stops by more generic concepts, e.g. "Eiffel Tower" is replaced by "Tourist Place", assuming that "Tourist Place" is a generalization of "Eiffel Tower" in the POIs taxonomy. This allows preserving some level of semantics in trajectories although losing the geographic information.

[Yarovoy et al. 2009] also offer k -anonymity for semantic trajectories by making groups of at least k indistinguishable trajectories. The novelty is that it also supports privacy personalization, like for streaming. It allows each user to specify his/her own preferences in terms of privacy protection for his/her trajectories. Instead of considering that users have a common quasi-identifier, as in [Terrovitis and Mamoulis 2008], this work assumes that each user defines the POIs and times for which he/she requires protection, thereby enabling each trajectory to be protected differently.

An alternative to data publishing in mobility data sharing is introduced in [Gkoulalas-Divanis et al. 2008] and pursued in [Pelekis et al. 2011c]. The approach targets organizations where their trajectory data must reside in-house and yet be made accessible to external untrusted users. A privacy enhanced trajectory query engine allows subscribed users to gain restricted access to the database and accomplish various analytic tasks. The query engine supports range, distance, k -nearest neighbors, and spatio-temporal queries, while ensuring privacy by auditing the queries and blocking potential attacks to user privacy.

8. CONCLUSIONS

Mobility data management and analysis have emerged in the last decade as a very active research domain, with several dedicated conferences and even Google sponsoring a mobility data contest. While previous research mainly focused on processing the raw data received from sensors, GPS devices and alike, recent research rather focuses on methods to enrich a movement track with more semantic,

application-oriented information. Adding semantics to movement data conveys a huge potential for innovative applications benefiting from the new capabilities of running far reaching analyses of mobility-related phenomena.

This survey focused on semantic trajectories, whose goal is to convey semantic information about mobility. We first introduced the concepts and definitions that enable a clear understanding of semantic trajectories. We extended this initial task to cover moving objects' behaviors, whose discovery represents one of the most popular uses of mobility data and possibly the ultimate goal of trajectory analyses. Understanding why and how people and animals move, which places they visit and for which purposes, what are their activities, and which resources they use, is of tantamount importance for all kinds of decision makers, in particular public authorities in charge of managing societal resources. The relative novelty of the domain leaves many avenues for future work open.

For a broader scope, several complementary types of movement remain to be investigated, including movement of large and deforming objects (e.g. oil spills, diseases), raster representations of movement, constrained movements, relative movement, and collective movement for any kind of collections of objects.

Existing techniques have to be reconsidered taking a more global approach. For example, map-matching will be improved by taking into account the semantic aspects (e.g. purpose of stops). More sophisticated analyses can be enabled via careful tuning (e.g., tuning stop identification and interpretation to make it efficient even for short stops), and via consistency enforcement for multiple, correlated annotations and segmentations. A set of precise and agreed-upon definitions of trajectory behaviors is still missing, in particular for semantic patterns. Privacy-preserving solutions for the protection of sensitive data remains to be extended to behaviors.

From a pragmatic perspective, ground truth benchmarks need to be developed to create better possibilities to assess the value of inference algorithms. In terms of technical support, database and data warehouse engines still have to reach beyond the prototyping step and all involved tools and facilities will have to reach online availability to enable continuous analysis of and feedback for ongoing trajectories. New computational models for semantic trajectories will be needed when really huge trajectory data sets become available. Dynamic systems shall support the self-improvement of the knowledge at hand (e.g., contextual repositories).

All of the above know-how exists in a research-oriented, preliminary assessment perspective. More systematic exploration and experimentation is necessary to consolidate theories and tools leading to a sound framework for application development in multiple application domains. Finally, innovative research is expected via the combination of traditional knowledge extraction techniques and visual approaches (a trend that already is on its way to effective tools construction) as well as via the integration of trajectory data with the geographically correlated data extracted from social networks interactions.

9. REFERENCES

- ABUL, O., BONCHI, F., AND NANNI, M. 2008. Never Walk Alone: Uncertainty for anonymity in moving objects databases. *Proc. of the 24th IEEE Int. Conf. on Data Engineering (ICDE)*
- ALVARES, L.O., BOGORNY, V., KUIJPERS, B., DE MACEDO, J.A.F., MOELANS, B., AND VAISMAN, A. 2007. A model for enriching trajectories with semantic geographical information. *Proc. ACM-GIS*, pp. 162–169, New York, NY, USA. ACM Press
- ALVARES, L.O., LOY, A. M., RENSO, C. AND BOGORNY, V. 2011. An Algorithm to Identify Avoidance Behavior in Moving Object Trajectories. *Journal of the Brazilian Computer Society*, 17(3): 193-203
- ANDERSSON, M., GUDMUNDSSON, J., LAUBE, P., AND WOLLE, T. 2008. Reporting leaders and followers among trajectories of moving point objects. *GeoInformatica* 12(4), 497-528
- ANDRIENKO, G., ANDRIENKO, N., AND WROBEL, S. 2007. Visual Analytics Tools for Analysis of Movement Data, *ACM SIGKDD Explorations*, 2007, v.9 (2): 38-46
- ANDRIENKO, N., AND ANDRIENKO, G. 2007. Designing visual analytics methods for massive collections of movement data. *Cartographica*, 42(2), (March 2007), 117-138
- ANDRIENKO, G., ANDRIENKO, N., BAK, P., KEIM, D., KISILEVICH, S., AND WROBEL, S. 2011. A conceptual framework and taxonomy of techniques for analyzing movement, *Journal of Visual Languages and Computing* 22(3): 213-232
- ASHBROOK, D., AND STARNER, T. 2003. Using GPS to learn significant locations and predict movement across multiple users, *Personal Ubiquitous Computing*, 7, 275–286
- BAGLIONI, M., DE MACEDO, J.A.F., RENSO, C., TRASARTI, R., AND WACHOWICZ, M. 2012. How you move reveals who you are: understanding human behavior by analyzing trajectory data. *Knowledge and Information System Journal (KAIS)*, special issue on Behavior Computing. To appear.
- BAMBA, B., LIU, L., PESTI, P., AND WANG, T. 2008. Supporting Anonymous Location Queries in Mobile Environments with PrivacyGrid. *Proc. of 17th Int. World Wide Web Conf.*
- BENKERT, M., GUDMUNDSSON, J., HÜBNER, F., AND WOLLE, T. 2008. Reporting flock patterns. *Computational Geometry* 41, 111-125
- BERESFORD, A.R., AND STAJANO, F. 2003. Location Privacy in Pervasive Computing, *IEEE Pervasive Computing*, 2003, 2, 46-55
- BOGORNY, V., KUIJPERS, B., AND ALVARES, L.O. 2009. ST-DMQL: a semantic trajectory data mining query language. *Int. Journal of Geographical Information Science*, 23:1245-1276.
- BOGORNY, V., AVANCINI, H., DE PAULA, B.L., KUPLISH, C.R., AND ALVARES, L.O. 2011. Weka-STPM: a Software Architecture and Prototype for Semantic Trajectory Data Mining. *Transactions in GIS*, 15:(2) 227-248
- BRAKATSOULAS, S., PFOSE, D., SALAS, R., AND WENK, C. 2005. On map-matching vehicle tracking data. *In the Proc. of the 31st Int. Conf. on VLDB (VLDB '05)*, 853-864
- CAGNACCI, F., BOITANI, L., POWELL, R.A., AND BOYCE, M.S. (eds.) 2010. Challenges and opportunities of using GPS location data in animal ecology. *Philosophical Transactions of the Royal Society of London: Biological Science*, Vol. 365, Issue 1550
- CAO, H., MAMOULIS, N., AND CHEUNG, D.W. 2007. Discovery of periodic patterns in spatiotemporal sequences. *IEEE Transactions on Knowledge and Data Engineering*, 19(4):453–467.
- CAO, H., WOLSON, O. 2005. Nonmaterialized Motion Information in Transport Networks, *Proc. of the Int. Conf. on Database Theory (ICDT)*, Springer LNCS 3363:173-188
- CAO, X., CONG, G., AND JENSEN, C.S. 2010. Mining Significant Semantic Locations From GPS Data. *PVLDB* 3(1): 1009-1020
- CAREY, J.R. et al. 2010. A high-resolution system for recording the daily and lifetime behavioral and movement patterns of individual tephritid fruit flies. *Proc. of Measuring Behavior 2010*, Eindhoven, The Netherlands, August 24-27
- CHOW, C.Y., MOKBEL, M., AND AREF, W. 2009. Casper*: Query Processing for Location Services without Compromising Privacy, *ACM Transactions on Database Systems (ACM TODS)*, Vol. 34, No. 4, pages 24:1-24:48
- DAMIANI, M.L., BERTINO, E., AND SILVESTRI C. 2010. The PROBE framework for the personalized cloaking of sensitive positions. *Transactions on Data Privacy*. Vol. 3, No. 2, pages 123-148

- DAMIANI, M.L., SILVESTRI, C., AND BERTINO, E. 2011. Fine-grained cloaking of sensitive positions in location sharing applications. *IEEE Pervasive Computing* Vol.10 (4): 64-72
- DING, Z., AND DENG, K. 2011. Collecting and Managing Network-Matched Trajectories of Moving Objects in Databases, *Proc. of the Int. Conf. on Database and Expert Systems Applications (DEXA)*, Springer LNCS 6860: 270-279
- DODGE, S., WEIBEL, R., AND LAUTENSCHÜTZ, A.K. 2008. Taking a systematic look at movement: Developing a taxonomy of movement patterns. *The AGILE workshop on GeoVisualization of Dynamics, Movement and Change*, Girona, Spain, May 5
- DOUGLAS, D. AND PEUCKER, T. 1973. Algorithms for the reduction of the number of points required to represent a digitized line or its caricature. *The Canadian Cartographer*, 10(2):112-122
- DUCKHAM, M., AND KULIK L. 2005. A Formal Model of Obfuscation and Negotiation for Location Privacy. *Pervasive Computing*, Springer, 2005, 152-170
- du MOUZA, C., AND RIGAUX, P. 2005. Mobility patterns. *Geoinformatica* 9, 4 (Dec. 2005), 297-319
- GIANNOTTI, F., NANNI, M., PEDRESCHI, D., PINELLI, F., RENSO, C., RINZIVILLO, S., AND TRASARTI, R. 2011. Unveiling the complexity of human mobility by querying and mining massive trajectory data. *VLDB Journal Special issue on Data Management for Mobile Services*, Vol.20(5): 695-719
- GIANNOTTI, F., NANNI, M., PINELLI, F., AND PEDRESCHI, D. 2007. Trajectory pattern mining. In *Proc. KDD '07*, pp. 330-339. ACM
- GIANNOTTI F. AND PEDRESCHI D., Eds. 2008, *Mobility, Data Mining and Privacy*, Springer
- GKOUALAS-DIVANIS, A., AND VERYKIOS, V. 2008. A privacy-aware trajectory tracking query engine. *SIGKDD Explorations*, 10(1): pp. 40-49
- GOMEZ, L., KUIJPERS, B., AND VAISMAN, A. 2008. Aggregation languages for moving object and place of interest, *Proc. SAC'08*, Fortaleza, Ceara, 16-20 March (2008) Brazil
- GRUTESER, M., AND GRUNWALD, D. 2003. Anonymous Usage of Location-Based Services Through Spatial and Temporal Cloaking. *Proc. of the 1st Int. Conf. on Mobile systems, Applications and Services*
- GUC, B., MAY, M., SAYGIN Y., AND KORNER, C. 2008. Semantic annotation of GPS trajectories, *11th AGILE Int. Conf. on Geographic Information Science*, Gerona, Spain
- GUDEMUNDSSON, J., VAN KREVELD, M., AND SPECKMANN, B. 2007. Efficient Detection of Patterns in 2D Trajectories of Moving Points, *Geoinformatica* 11:195-215
- GÜTING, R.H., DE ALMEIDA, V.T., AND DING, Z. 2006. Modeling and querying moving objects in networks. *The VLDB Journal*, 15(2): 165-190
- GÜTING, R.H., BÖHLEN, M.H., ERWIG, M., JENSEN, C.S., LORENTZOS, N.A., SCHNEIDER, M., AND VAZIRGIANNIS, M. 2000. A foundation for representing and querying moving objects. *ACM Transactions on Database Systems* 25(1) 1-42
- HADJILEFTHARIOU, M., KOLLIOS, G., BALAKOV, P., AND TSOTRAS, V. 2005. Complex spatio-temporal pattern queries, *Proc. of VLDB 2005*, Trondheim, Norway, August 30 - September 2
- HÄGERSTRAND, T. 1970. What about People in Regional Science. *Papers of the Regional Science Association* 24 (1970), 6-21
- HAN, B., LIU, L., AND OMIECINSKI, E. 2012. NEAT: Road Network Aware Trajectory Clustering, *Proc. of the 32nd Int. Conf. on Distributed Computing Systems (ICDS)*, Macau, China, June 18-21, 2012, To appear
- JENSEN, C.S., LU, H., AND YIU, M. 2009. Location Privacy Techniques in Client-Server Architectures. *Privacy in Location-Based Applications: Research Issues and Emerging Trends*, Springer-Verlag, 2009
- JUN, J., GUENSLER, R., AND OGLE, J. 2006. Smoothing Methods to Minimize Impact of Global Positioning System Random Error on Travel Distance, Speed, and Acceleration Profile Estimates. *Transportation Research Record: Journal of the Transportation Research Board*, 1972(1):141-150
- KALNIS, P., MAMOULIS, N. AND BAKIRAS, S. 2005. On discovering moving clusters in spatio-temporal data. *Proc. of 9th Int. Symposium on Spatial and Temporal Databases*, pages 364-381. Springer
- KELLARIS, G., PELEKIS, N., AND THEODORIDIS, Y. 2009. Trajectory Compression under Network Constraints. *Proc. of the 11th Int. Symposium Advances in Spatial and Temporal Databases (SSTD '09)*, 392-398
- KOUBARAKIS, M. et al, Eds, 2003. *Spatio-Temporal Databases: The CHOROCHRONOS Approach*. LNCS 2520, Springer
- KRUMM, J., AND HORVITZ, E. 2006. Predestination: Inferring Destinations from Partial Trajectories. *Ubicomp*: 243-260

- LASHLEY, M., AND BEVLY, D. 2007. Analysis of Discriminator Based Vector Tracking Algorithms. *Proc. of the 2007 National Technical Meeting of The Institute of Navigation*, San Diego, CA, 570-576.
- LAUBE, P. 2009. Progress in Movement Analysis. In H. K. Aghajan Ed., *Behaviour Monitoring and Interpretation - BMI - Smart Environments*, IOS Press, 43-71
- LAUBE, P., AND IMFELD, S. 2002. Analyzing relative motion within groups of trackable moving point objects. In M.J. Egenhofer, and D.M. Mark, Eds., *Geographic Information Science*, GIScience, Lecture Notes in Computer Science, 2478, 132-144, Springer, Heidelberg
- LAUBE, P., VAN KREVELD, M., AND IMFELD, S. 2005a. Discovering relative motion patterns in groups of moving point objects. *Int. Journal of Geographical Information Science*, Vol. 19 (6), July 2005, 639-668
- LAUBE, P., VAN KREVELD, M., AND IMFELD, S. 2005b. Finding REMO - Detecting relative motion patterns in geospatial lifelines. In *Developments in Spatial Data Handling*, Springer, Part 5, 201-215
- LAUBE, P. 2009. Progress in Movement Analysis. In H. K. Aghajan Ed., *Behaviour Monitoring and Interpretation - BMI - Smart Environments*, IOS Press, 43-71
- LEE, J.-G., HAN, J., AND WHANG, K.-Y. 2007. Trajectory clustering: a partition-and-group framework. In *Proc. SIGMOD Conf.*, 593-604. ACM
- LEE, J.G., HAN, J., LI, X., AND GONZALEZ, H. 2008. TraClass: Trajectory Classification Using Hierarchical Region-Based and Trajectory-Based Clustering, *Proc. of the Int. Conf. on Very Large Data Base (VLDB'08)*, Auckland, New Zealand, 779-790.
- LI, Z., DING, B., HAN, J., KAYS, R., AND NYE, P. 2010a. Mining periodic behaviors for moving objects. *KDD 2010*: 1099-1108.
- LI, Z., LEE, J.-G., LI, X., AND HAN, J. 2010b. Incremental Clustering for Trajectories. In the *Proc. of the 2010 Database Systems For Advanced Applications (DASFAA'10)*, 32-46
- LIAO, L., FOX, D., AND KAUTZ, H. 2005. Location-based Activity Recognition using Relational Markov Networks. In *Proc. of the 9th Int. Conf. on Artificial Intelligence, IJCAI'05*.
- LOU, Y., ZHANG, C., ZHENG, Y., XIE, X., WANG, W., AND HUANG, Y. 2009. Map-matching for low-sampling-rate GPS trajectories. *IGIS '09: Proc. of 17th ACM SIGSPATIAL Int. Conf. on Advances in Geographic Information Systems*, 352-361
- MACHANAVAJJHALA, A., GEHRKE, J., AND VENTIKASUBRAMANIAM, D.K. 2006. L-Diversity: Privacy Beyond k-Anonymity, *Proc. 22nd Int. Conf. on Data Engineering*
- MARKETOS, G., FRENTZOS, E., NTOUTSI, I., PELEKIS, N., RAFFAETA, A. AND THEODORIDIS, Y. 2008. Building Real-World Trajectory Warehouses. *Proc. of the 7th Int. ACM SIGMOD Workshop on Data Engineering for Wireless and Mobile Access (MobiDE'08)*, 8-15
- MERATNIA, N. AND DE BY, R.A. 2004. Spatiotemporal Compression Techniques for Moving Point Objects. *Proc. of the 9th Int. Conf. on Extending Database Technology (EDBT '04)*, 765-782
- MOHAMMED, N., FUNG, B.C., AND DEBBABI, M., 2009. Walking in the crowd: anonymizing trajectory data for pattern analysis. *Proc. of the 18th ACM Conf. on Information and Knowledge Management*
- MOKHTAR, H.M.O., AND SU, J. 2005. A query language for moving object trajectories, *Proc. of the 17th Int. Conf. on Scientific and Statistical Database Management*, Santa Barbara, CA
- MONREALE, A., PINELLI, F., TRASARTI, R., AND GIANNOTTI, F. 2009. WhereNext: a location predictor on trajectory pattern mining. *Proc. of the 15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining - KDD 2009*: 637-646
- MONREALE, A., TRASARTI, R., PEDRESCHI, D., RENSO, C., BOGORNY, V. 2011. Data Privacy for the Semantic Trajectories of Human Mobility. *Transactions on Data Privacy*, 2011 4(2): 73-101.
- MOUNTAIN, D., AND RAPER, J.F. 2001. Modelling human spatio-temporal behaviour: a challenge for location-based services, *Proc. of the 6th Int. Conf. on GeoComputation*, Brisbane, Australia, September 24-26, 2001
- NANNI, M., KUIJPERS, B., KÖRNER, C., MAY, M., AND PEDRESCHI, D. 2008: Spatiotemporal Data Mining. In *Mobility, Data Mining and Privacy*: 267-296
- NERGIZ, E, ATZORI, M., AND SAYGIN Y., 2008. Towards Trajectory Anonymization: A Generalization-based Approach. *Proc. of the ACM GIS Workshop on Security and Privacy in GIS and LBS*

- NEWSON, P. AND KRUMM, J. 2009. Hidden Markov Map Matching Through Noise and Sparseness. *Proc. of 17th ACM SIGSPATIAL Int. Symposium on Advances in Geographic Information Systems*, 336–343
- ORELLANA, D., WACHOWICZ, M., ANDRIENKO, N., ANDRIENKO, G. 2009. Uncovering Interaction Patterns in Mobile Outdoor Gaming, *Proc. of the Int. Conf. on Advanced Geographic Information Systems & Web Services (GEOWS)*, February 1-7, Cancun, Mexico
- PALMA, A.T., BOGORNY, V., KUIJPERS, B., AND ALVARES, L.O. 2008. A clustering-based approach for discovering interesting places in trajectories. *ACMSAC*, pp. 863–868, New York, NY, USA. ACM Press.
- PANAGIOTAKIS, C., PELEKIS, N., KOPANAKIS, I., RAMASSO, E., AND THEODORIDIS, Y. 2012. Segmentation and Sampling of Moving Object Trajectories based on Representativeness. *IEEE Transactions on Knowledge and Data Engineering*. To appear
- PELEKIS, N., ANDRIENKO, G., ANDRIENKO, N., KOPANAKIS, I., MARKETOS, G., THEODORIDIS, Y. 2011a. Visually Exploring Movement Data via Similarity-based Analysis. *Journal of Intelligent Information Systems*: 129-140
- PELEKIS, N., KOPANAKIS, I., KOTSIFAKOS, E.E., FRENTZOS, E. AND THEODORIDIS, Y. 2011b. Clustering uncertain trajectories. *Knowledge and Information Systems* 28(1): 117-147
- PELEKIS, N., GKOUALAS-DIVANIS, A., VODAS, M., KOPANAKIS, D., AND THEODORIDIS, Y. 2011c. A Privacy-Aware Query Engine for Sensitive Trajectory Data. *Proc. of the 20th ACM Conf. on Information and Knowledge Management (CIKM)*, October 24-28, 2011
- POTAMIAS, M., PATROUMPAS, K., AND SELLIS, T. 2006. Sampling Trajectory Streams with Spatiotemporal Criteria. *Proc. of the 18th Int. Conf. on Scientific and Statistical Database Management (SSDBM '06)*, 275–284
- QUDDUS, M.A., OCHIENG, W.Y., AND NOLAND, R.B. 2007. Current Map-Matching Algorithms for Transport Applications: State-Of-The Art and Future Research Directions. *Transportation Research Part C: Emerging Technologies*, 15(5): 312–328
- RINZIVILLO, S., PEDRESCHI, D., NANNI, M., GIANNOTTI, F., ANDRIENKO, N., AND ADRIENKO, G. 2008. Visually-driven analysis of movement data by progressive clustering, *Information Visualization*, 7(3/4): 225-239
- ROCHA, J.A.M., TIMES, V.C., OLIVEIRA, G., ALVARES, L.O., AND BOGORNY, V. 2010. DB-SMoT: A direction-based spatio-temporal clustering method. *IEEE Conf. of Intelligent Systems 2010*, 114-119
- RUIZ VICENTE, C., FRENI, D., BETTINI, C., AND JENSEN, C.S. 2011. Location-Related Privacy in Geo-Social Networks, *IEEE Internet Computing*, 15(3), pp. 20-27
- SAKR, M.A., AND GÜTING, R.H. 2011. Spatiotemporal pattern queries, *Geoinformatica*, 15: 497-540
- SCHMID, F., RICHTER, K.-F., AND LAUBE, P. 2009. Semantic Trajectory Compression. *Proc. of the 11th Int. Symposium on Advances in Spatial and Temporal Databases*, 411–416.
- SCHUSSLER, N. AND AXHAUSEN, K.W. 2009. Processing Raw Data from Global Positioning Systems Without Additional Information. *Transportation Research Record: Journal of the Transportation Research Board*, 8:28–36
- SEEK, 2012. *SEmantic Enrichment of trajectory Knowledge*, EU FP7–PEOPLE Marie Curie Actions <http://www.seek-project.eu>
- SIQUEIRA, F.L. AND BOGORNY, V. 2011. Discovering Chasing Behavior in Moving Object Trajectories. *Transactions in GIS*, 15 (5): 667-688
- SPACCAPIETRA, S., PARENT, C., DAMIANI, M.L., MACEDO, J.A., PORTO, F., AND VANGENOT, C. 2008. A conceptual view on trajectories. *Data & Knowledge Engineering* 65: 126–146
- SPACCAPIETRA, S. AND PARENT, C. 2011. Adding meaning to your steps, *Proc. of the ER 2011 Int. Conf.*, Brussels, October 31-November 3, 2011, LNCS Springer
- SPINSANTI, L., CELLI, F., AND RENSO, C. 2010. Where you stop is who you are: understanding peoples' activities. in *5th BMI'10, User Behaviour Modelling*, Karlsruhe, Germany
- STEWART-HORNSBY, K., COLE, S. 2007. Modeling Moving Geospatial Objects from an Event-based Perspective, *Transactions in GIS*, 11(4): 555–573
- SWEENEY L. 2002. Achieving k-anonymity privacy protection using generalization and suppression. *Int. Journal of Uncertainty, Fuzziness, and Knowledge-Based Systems*, World Scientific Publishing Co., Inc., 2002, 10, 571-588
- TERROVITIS M., AND MAMOULIS N. 2008. Privacy preservation in the publication of trajectories. *Proc. of the 9th Int. Conf. on Mobile Data Management*

- THÉRIAULT, M., CLARAMUNT, C., AND VILLENEUVE, P. 1999. A spatio-temporal taxonomy for the representation of spatial set behaviours. M.H. BÖHLEN, C.S. JENSEN & M. SCHOLL, Eds., *Spatio-Temporal Database Management, Int. Workshop STDBM'99*, Edinburg, Scotland, Springer LNCS, 1-18
- TIAKAS, E., PAPADOPOULOS, A.N., NANOPOULOS, A., MANOLOPOULOS, Y., STOJANOVIC, D., AND DJORDJEVIC-KAJAN, S. 2009. Searching for similar trajectories in spatial networks. *The Journal of Systems and Software*, 82(5), May 2009
- UDDIN, M.R., RAVISHANKAR, C., TSOTRAS, V.J. 2011. Finding Regions of Interest from Trajectory Data, *Proc. of the 12th IEEE Int. Conf. on Mobile Data Management (MDM)*, pages 39-48
- VAZIRGIANNIS, M., WOLFSON, O. 2001. A Spatiotemporal Model and Language for Moving Objects on Road Networks, *Proc. of the 7th Int. Symposium on Advances in Spatial and Temporal Databases (SSTD)*, Springer LNCS 2121: 20-35
- WANG T., AND LIU L. 2009. Privacy-aware mobile services over road networks. *Proc. VLDB Endow.* 2, 1 (August 2009), 1042-1053.
- WOOD, Z., AND GALTON, A. 2009a. A taxonomy of collective phenomena, *Applied Ontology* 4 (2009) 267-292
- WOOD, Z., AND GALTON, A. 2009b. Classifying Collective Motion. In B. GOTTFRIED AND H. AGHAJAN, Eds, *Behavior Monitoring and Interpretation – BMI*, IOS Press, 2009, 129-155
- WOOD, Z., AND GALTON, A. 2010. Zooming in on collective motion. In M. BHATT, H. GUESGEN AND S. HAZARIKA, Eds, *Spatio-temporal dynamics, Proc. of Workshop 21, 19th European Conf. on Artificial Intelligence, ECAI 2010*, August 16-20, 2010, Lisbon, Portugal, pp. 25-30
- XIE, K., DENG, K., AND ZHOU, X. 2009. From trajectories to activities: a spatio-temporal join approach. *GIS-LBSN 2009*: 25-32
- XUE, M., KALNIS, P., AND PUNG, H. 2009. Location Diversity: Enhanced Privacy Protection in Location Based Services. *Proc. of the 4th Int. Symposium on Location and Context Awareness*
- YAN, Z., CHAKRABORTY, D., PARENT, C., SPACCAPIETRA, S., AND ABERER, K. 2011. SeMiTri: A Framework for Semantic Annotation of Heterogeneous Trajectories. *Proc. of the 14th Int. Conf. on Extending Database Technology (EDBT '11)*, 259-270
- YAN, Z., PARENT, C., SPACCAPIETRA, S., AND CHAKRABORTY, D. 2010. A Hybrid Model and Computing Platform for Spatio-Semantic Trajectories. *Proc. of the 7th Extended Semantic Web Conf. (ESWC'10)*, 60–75
- YAROVY, R., BONCHI, F., LAKSHMANAN, V.S., AND WANG, W.H. 2009. Anonymizing moving objects: How to hide a MOB in a crowd? *Proc. of the 12th Int. Conf. on Extending Database Technology (EDBT)*
- YIGITOGU, E., DAMIANI, M.L., ABUL, O., AND SILVESTRI, C. 2012. Privacy-preserving sharing of sensitive semantic locations under road network constraints. *Proc. of the IEEE 13th International Conference of Mobile Data Management (MDM)*
- YIN, H. AND WOLFSON, O. 2004. A weight-based map matching method in moving objects databases. In *Proc. of the 16th Scientific and Statistical Database Management (SSDBM)*, 437-438
- YU, M., LI, Z., CHEN, Y., AND CHEN, W. 2006. Improving Integrity and Reliability of Map Matching Techniques, *Journal of Global Positioning Systems*, 5(1-2):40-46
- ZHENG, Y., CHEN, Y., XIE, X., AND MA, W.-Y. 2010. Understanding transportation modes based on GPS data for Web applications, *ACM Transaction on the Web*, 4(1), January 2010
- ZHENG, Y., ZHANG, L., MA, Z., XIE, X., AND MA, W.-Y. 2011. Recommending Friends and Locations Based on Individual Location History, *Journal ACM Transactions on the Web (TWEB)*, 5(1): 1559-1131
- ZHENG, Y., AND ZHOU, X., Eds. 2011. *Computing with Spatial Trajectories*, Springer