Representing, Storing and Mining Moving Objects Data

Jorge Huere Peña, and Maribel Yasmina Santos

Abstract—**Data about moving objects have been collected in huge amounts due to the proliferation of mobile devices, which capture the position of objects over time. Studies about moving objects have been developed as a specific research area of Geographic Information Systems. Those systems are designed to process traditional, static or slowly changing, geospatial data. However, moving objects have inherent a dynamism that requires different approaches to data storage and analysis. This paper presents a review of the key concepts associated to moving objects and their characteristics, as well as the approaches proposed to store data about moving objects. For the analysis of moving objects, an overview of the existing data mining techniques and some future guidelines are also presented.**

Index Terms—**Moving Objects, Moving Point, Moving Region, Spatio-temporal Data**

I. Introduction

DATA about movements of several objects, individuals, animals, virus, are being collected in growing amounts by means of current sensing and tracking technologies. A mobile object can be defined as an object that changes its geographical position as time passes. Such object integrates spatial and temporal characteristics.

The abstract physical entities for representing data associated with moving objects are the moving points and the moving regions, so this paper presents a set of details of their characteristics, nomenclature, units, among others, for the storage, management and subsequent analysis of the generated spatio-temporal databases.

Nowadays, and motivated by the abundant spatio-temporal information that is collected related about moving objects, there is a wide range of research works and publications on the perspectives, approaches and methodologies used to the storage and analysis of movement data.

Besides the characteristics of these objects and the associated data, based on the state-of-the-art study on moving objects databases, this paper also indicates possible trends and future research on this topic that is addressed by many researchers in the field of Information Systems, Computer Science, Geographic Information Systems and other related areas.

This paper is organized as follows. Section 2 presents the several types of moving objects and their characteristics. Section 3 describes the current approaches for storing and mining moving objects. Section 4 discusses some future directions in the analysis of moving objects using data mining algorithms. Section 5 concludes with some remarks about the work presented in this paper.

II. Moving Objects

This section starts with a brief description of the several types of moving objects that can be found in the literature, namely moving points and moving regions. The properties and characteristics of each moving object are presented below.

A. *Moving Objects Types*

In applications that require the modeling of single objects, three types of fundamental abstractions (as represented in Figure 1) can be used: point, line, and region [5]. Those entities (moving objects) need to be stored in a database. There are basically two different perspectives when dealing with moving objects databases: **location management perspective** and **spatio-temporal data perspective**.

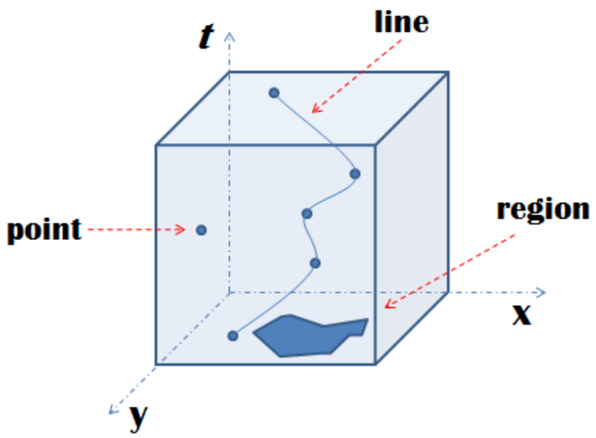


Fig. 1. The spatial data types: points, line and region (Adapted from [5]).

Manuscript received March, 23, 2011
J. H. Peña is a PhD Student at the Information Systems Department, Algoritmi Research Centre, University of Minho, Campus de Azurém, 4800-058 Guimarães, Portugal (corresponding author, phone: +351-253-510319; fax: +351-253-510300; e-mail: huerecom@hotmail.com).
M. Y. Santos is with the Information Systems Department, Algoritmi Research Centre, University of Minho, Campus de Azurém, 4800-058 Guimarães, Portugal (corresponding author, phone: +351-253-510308; fax: +351-253-510300; e-mail: maribel@dsi.uminho.pt).

ISBN: 978-988-19251-5-2
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)

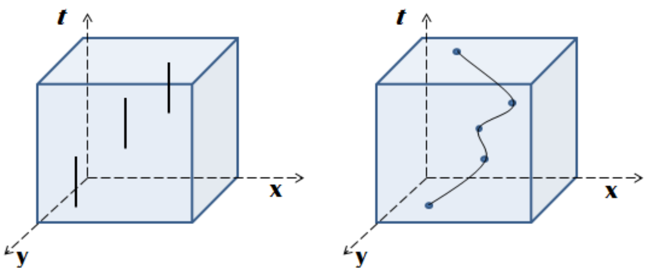
The Location Management Perspective involves the storage in the database of the moved object instantly and manages the current position. This approach is not concerned about the dynamics of movement because it does not record all the history of the moving object. The Spatio-Temporal Data Perspective considers the storage in a (static) spatial database of movement data, with an interest in observing the variations in time to understand and analyze the history of the moving objects under study [5].

1) Moving Point

A Moving Point (*MPoint*) is the basic abstraction of a physical object moving around in the plane or a higher-dimensional space, for which only the position, but not the extent, is relevant. It is defined either as a continuous function from time into the 2D plane, or as a polyline in the three-dimensional (2D + time) space [6]. When talking about the motion of humans or animals in space, that movement can easily be modeled and handled as a series of observations of moving points, represented as tuples of t, x and y coordinates [7]. Examples of moving point entities are cars, aircraft, ships, mobile phone users, terrorists, or polar bears [8].

The existing research on spatio-temporal databases allows us to introduce a classification of time-dependent point data. These data can be viewed in a natural way as being embedded in a space that is the cross-product of the original spatial domain and of time, 2D space and restricting the attention to a single time dimension, namely the valid time. In this context data are represented in a 3D space (as illustrated in Figure 2). In the Time domain, the time can be viewed as discrete, dense, or continuous, for practical reasons temporal database models often use discrete representations of time. In contrast, continuous models are more appropriate for dealing with moving objects [5].

.



(b) (b)

Fig 2. a) Discretely changing point (b) Continuously changing point (Adapted from [5])

The model of moving points is centered on the concept of trajectory, obtained from the continuous movement, using the algebra for moving objects that represent the motion of a moving point in a discrete form. This uses instants of time associated with the unit motion of a moving point in the x, y plane as a function of time (Figure 3).

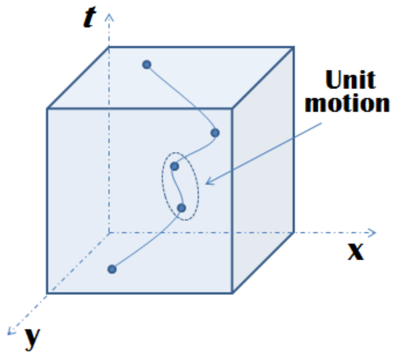


Fig 3. Discrete representation of motion of a moving point (Adapted from [9])

Thus, trajectories of moving points are often defined as sequences of (x, y, t) tuples:

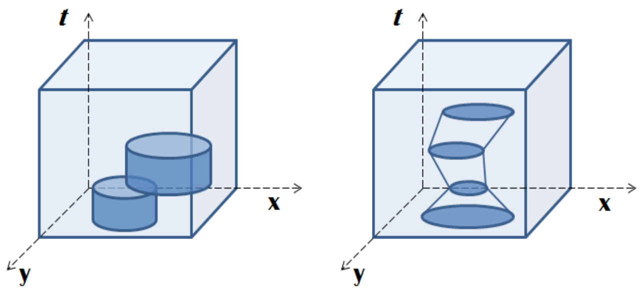
$T = \{(x_1, y_1, t_1), (x_2, y_2, t_2), \ldots, (x_n, y_n, t_n)\}$,
Where
$x_i \in R$ and $t_1 < t_2 < \ldots < t_n$,

Considering the temporal data types: Instant and Interval, type *UPoint* define a unit motion of a moving point in a time interval, consisting of: a starting point, an end point and a time interval. *MPoint* type defines a set of occurrences of *UPoint* types, representing the entire movement of a moving point [9].

2) Moving Region

A region is defined as a connected subset of the plane with non-empty interior, or as a polygon with polygonal holes [6]. A Moving Region (*MRegion*) is the basic abstraction of a physical object moving around in the plane or in a higher-dimensional space. The *MRegions* are moving entities with an extent and time-dependent position in space, as for example, hurricanes, forest fires, oil spills, armies, tribes of people in history, the spread of vegetation or of an illness, epidemic diseases, animal herds, and so forth [10].

A *MRegion* is represented graphically in a 3D space in different forms depending on the time domain, either discretely or continuously changing [5], as shown in Figure 4.



(a) (b)

Fig 4: a) Discretely changing region (b) Continuously changing region (Adapted from [5])

ISBN: 978-988-19251-5-2
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)

3) Non-Temporal Types

Table I shows a classification of Non-Temporal types, in a one-dimensional space (1D Space) and in two dimensional spaces (2D Spaces). Although the focus of this study is on temporal types and moving objects, this table allows a clarification on non-temporal types, namely the instant and periods data types [11].

TABLE I
CLASSIFICATION OF NON-TEMPORAL TYPES (Adapter from [11])

	1D Space					2D Space
	Discrete		Continuous			
	Integer	Boolean	String	Real	Time	2D
Point	Int	Bool	String	Real	Instant	Point
Point set	Range(int)	Range (bool)	Range (string)	Range (real)	Periods	Point, Line and Region

B. Moving Objects Characteristics

In this section, it is presented some of the characteristics of moving objects, based on the current research in this field, which allows the storage, the modeling and the querying of movement data. Moving objects data are usually stored in spatio-temporal database systems, providing those systems a language for posing queries.

The characteristics of Moving Objects, according to Güting & Schneider, with respect to their "shape" in 3D space, can be summarized as follows[5]:

Points

- Events in space and time: <point, instant>;
- Locations valid for a certain period of time: <point, period>;
- Set of location events–sequence of: <point, instant>;
- Stepwise constant locations–sequence of <point, period>; and,
- Moving entities: moving point.

Region

- Region events in space and time: <region, instant>;
- Regions valid for some period of time: <region, period>;
- Set of region events: – sequence of <region, instant>;
- Stepwise constant regions: – sequence of <region, period>; and,
- Moving entities with extent: moving region.

For the representation of the movement of moving objects in Static Maps the following characteristics are proposed: Identity (moving object), Path Traveled, Time, Speed and Duration of resting state, determining for each feature their respective variables [12], as shown in Table II.

TABLE II
COMPONENTS OF VISUAL VARIABLES (Adapter from [12])

Component	Level	form of presentation	visual variable
Identity (moving object)	Selective	Linear	Color
Path Traveled	Space	Linear	Location
Time	Ordinal	Linear	Saturation
Speed	Ordinal	Linear	Size
Duration of resting state	Ordinal	Punctual	Size

Another feature of moving objects is that they can implement monitoring applications for modeling and querying moving objects, using methodologies and techniques such as the implementation of data model within algebra modules that offers a set of type constructors and a set of operators, developed by R. Güting et al. The field of moving objects databases appeared characterized by two perspectives, the location management and the spatio-temporal database, as already mentioned [8].

C. Storing, Querying and Mining Moving Objects Data: Current Approaches

Moving objects database have been studied for about 15 years. During this period hundreds of works have been done. However, few of these resulted in research prototypes or commercial products. The work on this area is ranging from data models and query languages to implementation aspects, such as efficient indexing, query processing and optimization techniques, turning the field to a certain extent theoretical[4]. This is an open area of research integrating aspects related with databases supporting time dependent and continuously changing geometries, oriented to practical applications [4].

As mentioned previously, the work of R. Güting et al. defines a comprehensive framework for the abstract data types of moving objects intended for use on a conceptual basis for the representation and query of spatio-temporal data. The paper, *"A Foundation for Representing and Querying Moving Objects"* was published in January 2000[11]. This work allows the integration of abstract data types involving: base types, spatial types, time types, temporal and spatio temporal types. Embedding this in a Database Manager System (DBMS) query language allows the use of a query language for spatio-temporal data and moving objects.

Following this framework several models and prototypes have been proposed in the literature. From them, we point out in this paper two Moving Objects Database (MOD) engines: the HERMES engine and the SECONDO prototype.

1) HERMES engine

HERMES is a prototype system based on a powerful query language for trajectory databases, which enables the support of aggregative Location-Based Services. Its main goal is to support modeling and querying of continuously moving objects [1].

HERMES prototype system intends to take advantage of Moving Object Databases (MOD) that are emerging through the explosion of mobile devices and positioning technologies. Current database models have limited applicability in real-world applications, since proper estimations about future positions should involve past positions as well.

HERMES provides spatiotemporal functionalities handling objects that change location, shape and size, either discretely or continuously in time. It makes available a collection of data types and their corresponding operations, which were developed and provided as an Oracle data cartridge, called HERMES Moving Data Cartridge (HERMES-MDC), which is the core component of the HERMES system architecture, as shown in Figure 5 [1].

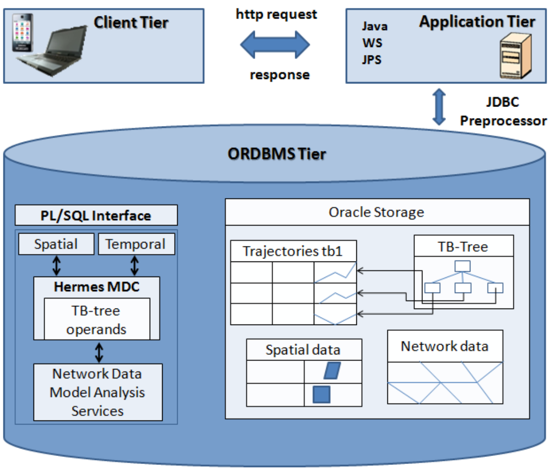


Fig 5. HERMES architecture (From [1])

HERMES project is developing Audio and Visual Processing Components. Taking for instance in data mining for example is available: *Data logging and mining for reminder generation* that is used for person and location specific reminders.

2) SECONDO prototype

SECONDO is an extensible database system, which DBMS prototype has been developed at the University of Hagen since about 1995. It runs on Windows, Linux, and MacOS X platforms and is freely available as open source software. The main design goals were a clean extensible architecture and support for spatial and spatio-temporal applications. The extensible architecture of SECONDO is shown in Figure 6. It consists of three major components: the Graphical user interface (GUI), the kernel and the optimizer [4].

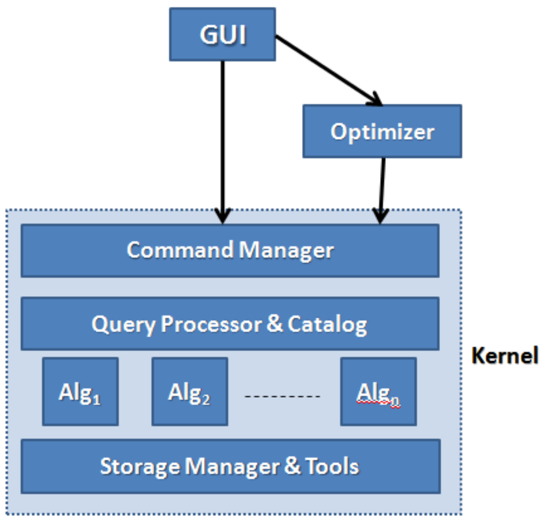


Fig 6. SECONDOs architecture (From [4])

The moving objects implementations that are available are the BerlinMOD benchmark and the SECONDO plugin facility.

Nowadays, many applications use the BerlinMOD tool, as the foundation for spatio-temporal database management systems (STDBMS) used to compare the performance and details on queries employing moving point data [13].

To validate this tool in a practical way BerlinMOD used simulation models of cars moving through the network of streets in the capital of Germany, Berlin, and the behavior of workers when moving between home and work and traveling in their free day. An interesting aspect of this tool is the fact of being open source software. Researchers with other data structures or other algorithms can use this approach to compare their solutions [14].

Also in this context, we mention the recent work done by Düntgen et al. which proposes three different approaches to represent the history of moving objects in databases: the Compact Representation, the Unit Representation and the Hybrid Representation, based upon the BerlinMOD benchmark [15].

3) Other models and prototypes

Following different approach, other systems and prototypes enable the discovery of interesting moving object clusters with relaxed temporal constraint. Zhenhui et al. raises two interesting works. The first, MoveMine, that is designed for sophisticated moving object data mining by integrating several attractive functions including moving object pattern mining and trajectory mining. The system has an user-friendly interface, as shown in Figure 7, to facilitate interactive exploration of mining results [3].

Fig 7. MoveMine: user-friendly interface (From [3])

Another interesting work is "Swarm". It proposes a new pattern of movements, whose goal is to find all discriminative swarms, namely closed swarm. While the search space for closed swarms is prohibitively huge, they designed a method named Object Growth, to efficiently retrieve the result [2]. The key observation is that the moving objects in a cluster may actually diverge temporarily and converge at certain timestamps (Figure 8).

ISBN: 978-988-19251-5-2
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)

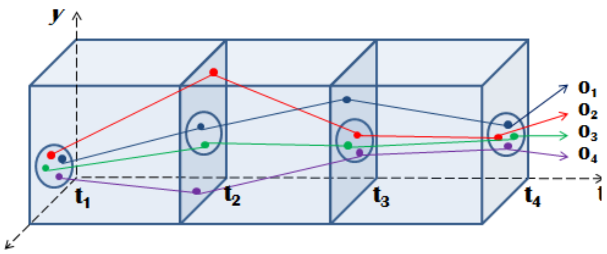


Fig 8. Loss of interesting moving object clusters in the definition of moving cluster, flock and convoy (Adapted from [2]).

4) Mining MOD

Data mining is commonly used as a synonym for knowledge discovery in databases (KDD). However in the literature, Data mining is one step of the KDD process. Data mining involves the application of techniques for distilling data into information or facts implied by the data [16].

The mining process in Moving Objects Databases intends to extract new knowledge from data using data mining algorithms. This new knowledge or patterns, useful in the decision making process, involve different data positions in the time dimension.

Research in Mining Moving Object Data has not yet produced a theoretical framework for Data Mining on spatio-temporal data. The available approaches mainly focus on the spatial component or try to adapt classical local patterns to this specific domain [17].

It is possible to find different perspectives on data mining functions, with three being the most used: *Characterization*, *Clustering* and *Association of data*. In some systems these data mining functions are implemented with the help of a mining language called *Spatial Mining Language* (SML).

Characterization adds a new attribute to an object class, based on the values of other attributes. Its formal syntax is MINE CHARACTERIZATION in the SML.

Clustering creates groups of objects based on the values of some attributes. The formal syntax is MINE CLUSTERING in the SML. *Association* relates objects, based on their attribute values. Its formal syntax is MINE ASSOCIATION [18].

Regarding research work on *Mining Moving Objects Data* we find the work of G. Marketos who proposed a framework for *Mobility Data Warehousing and Mining* (MDWM) that takes into consideration the complete flow of tasks required for the development of a *Trajectory Data Warehouse* (TDW) and the application of trajectory-inspired mining algorithms to extract traffic patterns[19]. An example of the results that can be achieved using the TDW is the called typical trajectory that describes the trend of movement within a cell. This is a demanding issue since it must be defined how to derive the typical trajectory of a cell based on the trajectory of a cell based on the typical trajectories of its sub-cells

About moving regions, Junghans & Gertz present an approach to model and predict the evolution of the moving regions. The approach is based on modeling moving heterogeneous regions using enclosing boxes. To evaluate this approach, the authors use two real data sets, related with fires. The data sets contain the perimeters (regions) of wildfires at different times and were taken from the geospatial data archive created by the Northern Rockies Coordination Group during the 2003 fire season [20].

D. Mining Moving Objects Data: Future Directions

This section briefly presents some open issues in the field of mining moving objects data.

Current research approaches use huge amounts of time-stamped locations, associated with moving objects. Additionally, data associated to with the direction or speed of moving objects could be considered. These data can be collected or calculated from positions readings.

Considering that most of the Spatio-Temporal research of moving objects is aimed at determining location patterns as well as trajectory models, an interesting research topic would be to create or improve their profiles, focusing on the classification of moving objects for detecting possible misbehaviors.

Due to the existence of big amount of data of moving objects in various applications, it is necessary to develop real-time applications for integration of information contained in different sources, such as tourist information, advertising, traffic monitoring services in land, air, sea among others.

The time dimension of moving objects must also be investigated in depth, especially regarding time intervals, as entities may lose some features if time intervals are very wide. So, it is necessary to define an appropriate reading frequency that is necessary for further investigation.

Researches on moving objects usually work in a given period of time, due to the characteristics and methodologies of the proposed research. So, it is also necessary to verify the implications of working with historical data thus considering a longer period of time. For this, a spatial data warehouse could be useful.

The analysis of movement data has allowed the identification of trajectories of objects moving in space and the identification of different types of movement, either from pedestrians, bicycles, motorcycles, cars, trucks, and so on. However, an emerging area for the analysis of movement data is associated with the identification of flows, where there is the need to identify groups of objects moving together in space.

The use of *Data Mining Query Language* (DMQL) has great difficulties because of the complexities and peculiarity of the moving objects in the spatio-temporal domain.

Another area is associated with the visualization of the obtained patterns, where the analyst needs appropriate visualization capabilities for presenting the models, patterns or trends found in data.

Existing techniques and algorithms developed for *Spatial Data Mining* would need new developments to be able to deal with spatio-temporal data. This line of work also requires future research.

From this broad list of open issues, the authors will focus on the identification of flows of moving objects evolving in a synchronous way in space. For the concretization of this goal, the first step will be associated to the identification of data sets that facilitate the emergence of groups or the collection of data over a certain period of time in a space that facilitate the emergence of such groups. This space could be a building, a shopping center, a supermarket, a park, a road network, an open space, depending on the moving objects under consideration.

After data availability, the analysis of the collected data will allow the identification of flows of movement. We propose the use of clustering algorithms, namely density-based clustering algorithms, for which the distance function needs to be tuned in order to correctly identify flows of movement. Besides the definition of the distance function that will be used in the clustering process, and for which we foresee that variables like position, speed, bearing and time will be of great importance, the weight of them in the distance function also needs to be set. Different weights will lead to the identification of different types of movement flows (Figure 9)

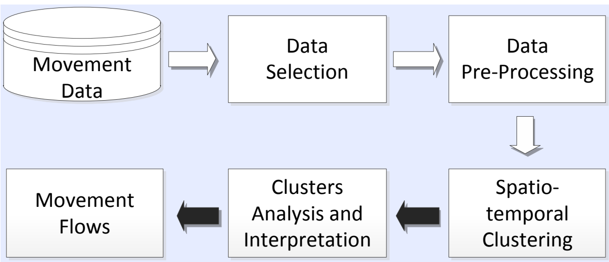


Fig 9. Proposed approach

Looking at Figure 9, the proposed approach will store the available data in a spatial database that supports the fundamental querying functionalities. The Data Selection phase will ensure the selection of the attributes needed for the clustering process. Those attribute need to be pre-processed, putting them in the appropriate format for data analysis. In this step, mechanisms for evaluating the quality of the data under analysis will also be considered. In the spatio-temporal clustering phase, clustering algorithms will be applied to identify groups of objects that move together. The obtained results need to be analyzed in order to classify clusters that represent flows of movement from clusters that represent, for example, suspension of movement. This classification will lead to the identification of flows of movement and to the definition of the characteristics of such flows.

III. Conclusion

This paper presented the particular case of moving points and moving regions as abstract features to represent, store and analyze data about moving objects. Moving points are used to characterize objects for which only position in space is relevant, whereas moving regions are used to characterize objects for which position and extension is relevant. For both, the time dimension is considered.

For the storage and analysis of movement data, this paper also summarized the main developments in systems or prototypes like Hermes, SECONDO, MoveMine and Swarm.

This paper mentions several open issues in the analysis of movement data using data mining techniques. For one of these open issues, it is proposed the use of spatial data mining algorithms with add-ons able to deal with the time dimension in order to mine movement data and identify flows of objects moving together in the geographical space.

Reference

[1] N. Pelekis, E. Frentzos, N. Giatrakos *et al.*, “HERMES: Aggregative LBS via a Trajectory DB Engine,” *ACM*, pp. 9-11, 2008.
[2] Z. Li, B. Ding, J. Han *et al.*, “Swarm: Mining Relaxed Temporal Moving Object Clusters,” *VLDB*, pp 723-734, 2010.
[3] Z. Li, M. Ji, J.-g. Lee *et al.*, “MoveMine: Mining Moving Object Databases,” *ACM*, pp. 1203-1206, 2010.
[4] R. Güting, and T. Behr, “SECONDO: A Platform for Moving Objects Database Research and for Publishing and Integrating Research Implementations,” *DBMS*, pp. 1-8, 2010.
[5] R. H. Güting, and M. Schneider, *Moving Objects Databases*, Elsevier Inc., 2005.
[6] L. Forlizzi, R. H. Güting, E. Nardelli *et al.*, “A Data Model and Data Structures for Moving Objects Databases*,” *ACM*, pp. 319-330, 2000.
[7] P. Laube, and S. Imfeld, “Analyzing Relative Motion within Groups of Trackable Moving Point Objects,” *Information Systems*, 2002.
[8] R. H. Güting, V. A. Teixeira, and Z. Ding, “Modeling and querying moving objects in networks,” *VLDB*, pp. 165-190, 2006.
[9] V. Ferreira, and M. Véras, “Uma implementação da Álgebra para Pontos Móveis usando Postgresql (in Portuguese),” *CONNEPI2010*, 2010.
[10] C. Junghans, and M. Gertz, “Modeling and Prediction of Moving Region Trajectories,” *ACM 978*, 2010.
[11] R. H. Guting, M. H. Bohlen, M. Erwig *et al.*, “A Foundation for Representing and Querying Moving Objects,” *Language*, pp. 1-37, 2000.
[12] D. S. Cotrim, and J. Campos, “Representação das Características do Movimento de Objetos Móveis em Mapas Estáticos (in Portuguese),” *Symposium A Quarterly Journal In Modern Foreign Literatures*, 2007.
[13] "BerlinMOD Site," 2011. [Online]. Available: http://dna.fernuni-hagen.de/secondo/BerlinMOD/BerlinMOD.html/.
[14] "SECONDO Site," 2011. [Online]. Available: http://dna.fernuni-hagen.de/Secondo.html/.
[15] C. Düntgen, T. Behr, and R. Hartmut Güting, "Assessing Representations for Moving Object," *INFORMATIK BERICHTE*, 2010
[16] F. Giannotti, and P. D., *Mobility, Data Mining and Privacy: A Vision of Convergence*, Berlin Heidelberg: Springer-Verlag, 2008.
[17] M. Nanni, B. Kuijpers, C. Körner *et al.*, *Spatiotemporal Data Mining*, Berlin Heidelberg: Springer-Verlag, 2008.
[18] S. Brakatsoulas, D. Pfoser, and N. Tryfona, “Modeling, Storing and Mining Moving Object Databases,” *Research Academic Computer Technology Institute*, 2004.
[19] G. Marketos, "Mobility Data Warehousing and Mining," *VLDB ’09 PhD Workshop*, 2009.
[20] C. Junghans, and C. Gertz, "Modeling and Prediction of Moving Region Trajectories." *IWGS ’10*, 2010.

ISBN: 978-988-19251-5-2
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)