Exploring Movement Using Similarity Analysis

Dissertation

zur
Erlangung der naturwissenschaftlichen Doktorwürde
(Dr. sc. nat.)

vorgelegt der
Mathematisch-naturwissenschaftlichen Fakultät
der
Universität Zürich

von

Somayeh Dodge

aus dem Iran

Promotionskomitee

Prof. Dr. Robert Weibel (Vorsitz, Leitung der Dissertation)
Prof. Dr. Sara Irina Fabrikant
Dr. Patrick Laube
Dr. David Mountain

Zürich, 2011

To my dear parents,
Saeed Dodge & Shahla Darabian,
for their love, endless support, and encouragement.

Summary

Movement is a vital aspect of almost all organisms and many spatio-temporal processes. Hence it is crucial to understand movement and gain knowledge about its patterns. Recent advances in positioning technologies provide an increasing access to massive repositories of movement data and hence challenges arise to develop new exploratory tools and knowledge discovery techniques in order to extract meaningful information, discover interesting patterns, and explore the dynamic behavior of moving objects (humans, vehicles, vessels, animals) or processes (hurricanes, oil spills). Among knowledge discovery techniques, the exploration of similarities in the movement of multiple objects is a key emerging interest. Learning about movement similarities can be beneficial in the prediction, modeling and simulation of collective behavior of dynamic phenomena.

This thesis intends to contribute to GIScience's exploratory capacity to discover insights about patterns of movement as well as existing similarities between movement behaviors of different objects. Specifically, the aim is to develop concepts and methods for incorporating *movement parameters* such as speed, acceleration, or direction in the study and analysis of movement. Hence in this thesis, *movement similarity* is defined as the resemblance in the variations of movement parameters of objects over time.

This study, with a perspective on movement, undertakes a three-stage research process including the development of (*a*) a conceptual framework, (*b*) feature extraction and segmentation methods, and (*c*) similarity assessment techniques. The overall study involves an iterative research process integrating quantitative techniques from GIScience and knowledge discovery approaches in order to extract high-level information from low-level, raw movement data.

The core of the research process is presented in four scientific papers. Research Paper 1 proposes a conceptual framework for movement as well as a comprehensive classification of movement patterns. Research Paper 2 presents a segmentation technique in order to extract movement features from trajectories of moving objects. The segmentation process can be seen as a dimension reduction technique to simplify the structure of the movement data in order to facilitate knowledge discovery. Research Paper 3 proposes a novel similarity assessment approach relying on the segmentation technique. Finally, Research Paper 4 extends the dimensionality of the main approach towards the detection of relative movement patterns.

Furthermore, through a set of experiments this thesis shows that the proposed

methods can be successfully applied in conjunction with data mining techniques in order to support knowledge discovery from various movement datasets in real-world applications (e.g. transportation, meteorology). Consequently, the outcomes of this thesis can contribute to knowledge discovery from movement data where the interest is to extract or group similar behaviors of dynamic objects.

Zusammenfassung

Bewegung ist ein (lebens)notwendiger Aspekt von fast allen Organismen und vieler räumlich-zeitlicher Prozesse. Daher ist es äusserst wichtig, das Konzept *Bewegung* zu verstehen und unser Wissen über Bewegungsmuster zu erweitern.

In letzter Zeit haben Fortschritte diverser Ortungstechnologien uns den Zugang zu immens grossen Mengen von Bewegungsdaten eröffnet. Dieser Schatz von Daten stellt Forschende aber auch vor die Herausforderung, neue explorative Werkzeuge und Methoden der Erkenntnisgewinnung (Knowledge Discovery) zu entwickeln, um nützliche Informationen extrahieren, interessante Muster entdecken und dynamisches Verhalten von beweglichen Objekten (z.B. Menschen, Fahrzeuge, Schiffe, Tiere) oder von Prozessen (z.B. Hurrikane, Ölteppiche) analysieren zu können. Unter den Verfahren der Knowledge Discovery ist die Analyse von Ähnlichkeiten im Bewegungsverhalten mehrerer Objekte von grossem und wachsendem Interesse. Das Wissen über solche Bewegungsähnlichkeiten nützt der Vorhersage, der Modellierung und der Simulation von gemeinschaftlichem Verhalten dynamischer Phänomene und Objekte.

Die vorliegende Dissertation zielt darauf ab, einen Beitrag zu den exploratorischen Ansätzen der Geographischen Informationswissenschaft zu leisten, so dass diese neues Wissen über Bewegungsmuster und über Ähnlichkeiten zwischen den Bewegungsverhalten unterschiedlicher Objekte erarbeiten kann. Das spezifische Ziel der Arbeit ist die Entwicklung von Konzepten und Methoden, um *Bewegungsparameter* wie zum Beispiel Geschwindigkeit, Beschleunigung oder Richtung bei der Analyse von Bewegungen miteinzubeziehen. Daher ist der Begriff der *Bewegungsähnlichkeit* in dieser Dissertation als die Ähnlichkeit von zeitabhängigen Variationen in den genannten Bewegungsparametern von bewegten Objekten definiert.

Im Hinblick auf Bewegung als Untersuchungsgegenstand folgt die vorliegende Untersuchung einem dreiteiligen Forschungsprozess. Dieser besteht aus der Entwicklung von (a) einem konzeptionellen Rahmen, (b) Merkmalsextraktions- und Segmentierungsmethoden und (c) Techniken zur Beurteilung von Bewegungsähnlichkeit. Die Untersuchung benutzt dafür einen iterativen Forschungsprozess, der quantitative Ansätze aus der Geographischen Informationswissenschaft mit Ansätzen der Knowledge Discovery kombiniert, um aus Bewegungsrohdaten Informationen auf einem hohen konzeptionellen Niveau ableiten zu können.

Der Kern des Forschungsvorhabens wird in vier wissenschaftlichen Artikeln präsentiert. Forschungsartikel 1 stellt einen konzeptionellen Rahmen zur Be-

wegungsanalyse und eine umfassende Klassifikation von Bewegungsmustern vor. Forschungsartikel 2 führt einen Segmentierungsansatz ein, der es erlaubt Bewegungsmerkmale aus Trajektorien bewegter Objekte zu extrahieren. Der Segmentierungsprozess kann als Werkzeug zur Reduktion der Dimensionalität des Problems verstanden werden. Er vereinfacht die Struktur von Bewegungsdaten und ermöglicht dadurch die Knowledge Discovery. Forschungsartikel 3 schlägt einen neuen Ansatz zur Beurteilung von Bewegungsähnlichkeit vor, der auf dem Segmentierungsansatz aufbaut. Schliesslich erweitert Forschungsartikel 4 die Dimensionalität des Ansatzes in Richtung der Detektion von relativen Bewegungsmustern.

Zusätzlich zeigt die vorliegende Dissertation durch eine Serie von Experimenten, dass die entworfene Methodik zusammen mit Data-Mining-Techniken angewendet werden kann, um Knowledge Discovery in verschiedenartigen Bewegungsdaten (zum Beispiel aus den Gebieten der Verkehrsanalyse oder der Meteorologie) erfolgreich zu unterstützen. Die Resultate dieser Dissertation tragen daher zur Knowledge Discovery in Bewegungsdaten mit Fokus auf die Extraktion oder Gruppierung ähnlichen Bewegungsverhaltens mehrerer dynamischer Objekte bei.

Acknowledgments

This thesis was carried out during four years of my time as research assistant at the University of Zürich. I gratefully acknowledge the university for supporting my Ph.D. work. This thesis would not have been possible without the collaboration of several people who in one way or another contributed and extended their valuable assistance. Here, I name some of them and I inwardly thank the others.

First and foremost, I am heartily thankful to my chief supervisor, Professor Robert Weibel, whose encouragement, guidance and kind support from the initial to the final level enabled me to accomplish this research. I am deeply grateful of him for giving me the opportunity to carry out my Ph.D. in the beautiful city of Zürich, for being such a compassionate and supportive supervisor, for his time and patience, and finally for reading and correcting my sloppy manuscripts. This thesis would not have been possible without his help, and support.

I owe a huge dept of thanks to my co-supervisor, Dr. Patrick Laube, for his time, guidance, constructive criticisms, valuable inputs, and suggestions. Certainly without him, I would have had a very difficult time to complete this thesis.

I am particularly indebted to the other members of my Ph.D. committee: Sara Irina Fabrikant, professor at the GIVA division, University of Zürich, and Dr. David Mountain from the Department of Information Science at the City University of London. I am very grateful for their time for discussion, their precious suggestions and inputs in this thesis. I would also like to express my gratitude to Professor Harvey Miller from the Department of Geography at the University of Utah for reading and reviewing this thesis as external expert.

I would like to express my sincere gratitude to Dr. Ross Purves, Dr. Maike Buchin, Professor Marc van Kreveld, and Dr. Goce Trajcevski for their valuable inputs and suggestions that triggered some of the ideas in this thesis. I am indebted to Dr. Goce Trajcevski for his time for the fruitful discussions during my visit to the Department of Electrical Engineering and Compute Science at the Northwestern University.

I owe my deepest gratitude to Elisabeth Cottier, the former GIS division's secretary, for her kindness and for helping me through all the administrative work. Any time I encountered a problem, specially during the first year of my Ph.D. with all the immigration work, she was there for me with her beautiful smile. I would also like to thank Annica Mandola, our current group's administrative assistant, for her kind help during the final year of this thesis.

My time at the University of Zürich was made enjoyable mostly due to the

many friends and colleagues that became a part of my life. I am grateful of Mortiz Neun, Stefan Steiniger, Felix Hebeler, Ralph Straumann, Ronald Schmidt, Patrick Lüscher, Ramya Venkateswaran, Pia Bereuter, Arzu Çöltekin, Anna-Katharina Lautenschütz , Stefano De Sabbata, Paul Crease, Glenda García Santos, Giuseppe Feola, and Karen Budge for supporting me through my time in Zürich. Thanks for many interesting discussions and for the fun and great time that we shared together at the Irchel Campus.

I would like to express my heartfelt thanks to my cousin Dr. med. Ali Dodge-Khatami, and my hematologists Dr. med. Pietro Peghini and Dr. med. Markus Schmugge, who made my life in Switzerland possible and without any complications. I never could have made it without them!

I am grateful to my granduncle, Prof. Dr. Yadolah Dodge, for his kind help during the initial stages of this work. I probably would not have started this Ph.D. in Zürich without his encouragement and support.

Last but not least, I would like to thank my family for all their love and encouragement. For my parents, Saeed and Shahla, and my sister Sooreh, who have been supporting me in all my pursuits and went through all ups and downs of my life with me. And most of all for my loving, supportive, encouraging, and patient husband Ehsan whose faithful support during my Ph.D. is so appreciated. Thank you!

Somayeh Dodge
University of Zürich
May 2011

Contents

List of Figures

List of Abbreviations

DTW	Dynamic Time Warping	(p. 22)
ERP	Edit distance with Real Penalty	(p. 25)
GIScience	Geographical Information Science	(p. 3)
GKD	Geographic Knowledge Discovery	(p. 3)
GPS	Global Positioning System	(p. 3)
KDD	Knowledge Discovery in Databases	(p. 6)
LCSS	Longest Common Subsequence	(p. 22)
LIP	Locality In-between Polylines	(p. 32)
MP	Movement Parameter	(p. 12)
MPC	Movement Parameter Class	(p. 46)
MPO	Moving Point Object	(p. 12)
NWED	Normalized Weighted Edit Distance	(p. 47)
OWD	One Way Distance	(p. 29)
PCA	Principal Component Analysis	(p. 44)
RTSD	Rigid Transformation Similarity Distance	(p. 33)
SVM	Support Vector Machine	(p. 44)
TCSD	Temporal-Containment Similarity Distance	(p. 33)
TQuEST	Threshold Queries	(p. 27)
wDF	w-constrained Fréchet distance	(p. 33)

Part I

Synopsis

Chapter 1

Introduction

1.1 Motivation

Movement is a key element of many processes and activities. Understanding of the movement itself, as well as the patterns of movement is very important in many areas of science and technology such as ecology, meteorology, sociology, behavioral studies, transportation planning, surveillance and intelligence services. For instance, it is a crucial element in modeling and simulation of "dynamic collectives and their collective dynamics" (Galton, 2005, p.300), in diverse applications such as behavioral ecology of animals (Turchin, 1998; Nathan et al., 2008), environmental hazards (Elsner and Kara, 1999; Sinha and Mark, 2005), traffic management and human mobility studies (Mountain and Raper, 2001; Mouza and Rigaux, 2005; González et al., 2008). Equally, it is also a key factor for success in business, as the location and movement of customers, resources and products becomes a central consideration in an increasingly mobile world (MOVE, 2009).

In recent years, owing to tremendous advances in positioning and tracking technologies, massive amounts of movement data are available from diverse domains such as animal tracking using GPS collars and radar observations, monitoring hurricane movement using satellite imagery, tracking humans using GPS or smart mobile devices, to name but a few. The availability of such growing repositories of movement datasets challenges the development of new exploratory tools and knowledge discovery techniques in order to extract meaningful information, discover interesting patterns, and explore dynamic behavior of moving objects (humans, vehicles, animals) or spatio-temporal processes (hurricanes, oil spill).

Recognizing the growth in positioning and tracking technologies and movement datasets, research interest in developing computational methods for the analysis of movement data has increased significantly in the *Geographic Information Science (GIScience)* community over the past few years. Accordingly, a large number of studies have pioneered innovative approaches to exploit movement data in various aspects of *Geographic Knowledge Discovery (GKD)* processes, namely, trajectory data analysis, movement pattern mining, as well as exploratory visual analytics (Imfeld, 2000; Laube, 2005; Mountain, 2005; Giannotti and Pedreschi, 2008). To the extent that in their second edition, Miller and Han (2009) dedi-

cated five new chapters to capture the progress in developments and applications of GKD methods on spatio-temporal and moving objects datasets between 2001 (the publication year of the first edition) to 2009. Among GKD techniques for the detection of movement patterns in individual objects, the challenge of exploring similarities in the movement of multiple objects is a key emerging interest (Pelekis et al., 2007; Buchin et al., 2009; Etienne et al., 2010). This thesis intends to primarily contribute to GIScience's explorative capacity to investigate similarities in movement behaviors of multiple objects.

1.1.1 Knowledge Discovery from Movement Parameters

In many applications, studying movement characteristics of objects, in terms of essential *movement parameters* (speed, acceleration, direction etc.) and the patterns they form, is more relevant than simply the movement paths. Contrary to the trajectory, which is a geometrical abstraction of the movement path over time, movement parameters convey the physical notion and can hint to the semantics of the movement of an object. Hence, such parameters can give complementary insight into the movement behavior of objects, identifying patterns *and* their causes. For instance, when a convoy of vehicles moves along the same road, the movement path is very similar for all vehicles, however, the variations of movement parameters differs in space and time for each individual, caused by influencing factors such as the transportation mode, traffic patterns, and topography of the road. Similar observations can also be obtained from the movements of vessels along the same itinerary (Etienne et al., 2010). Likewise, although the geometry of North Atlantic hurricane trajectories is approximately similar for most hurricanes and the general trend of their paths can be simplified by a parabolic sweep, the speed and acceleration patterns of hurricanes varies due to the influence of environment temperature, wind speed, and geographic latitudes. Therefore, the prediction of the exact movement of hurricanes is difficult (Elsner and Kara, 1999). The most uncertain period in the movement of hurricanes is at the time of recurvature (i.e. change to a more northerly direction), and hence, the correct anticipation of the recurvature point is very crucial in forecasting the trajectories of hurricanes (Elsner and Kara, 1999). Hence, in order to study patterns of collective behavior of such objects and processes, parameters such as speed, acceleration and turning angle (i.e. change of direction) are more important than the movement path alone. These patterns usually emerge from the effect of internal and external influencing factors such as intrinsic properties of objects and the environment (Nathan et al., 2008).

A common characteristic of the available techniques for knowledge discovery and data mining techniques from movement datasets is that they mostly rely on *positional information* of tracked objects through time, using *trajectory* representation, whereas very little consideration has been paid to other *movement*

parameters. Nowadays, with the emergence of new generations of positioning and tracking technologies and advances of in-vehicle sensors such as gyroscopes, accelerometers, it is possible to capture a variety of movement parameters in addition to position information as an object moves. These new sources of information provide a promising opportunity to explore the structure of movement data and extract relevant and useful knowledge about the movement behavior of different objects. Hence, the development of new analysis techniques that are capable of exploiting such information appears to be a logical step forward in knowledge discovery from movement datasets.

1.1.2 Similarity Search in Movement Datasets

In many applications, moving objects share similarities but also exhibit differences in their movements. For instance, homing pigeons often share certain flight patterns when they are close to their home or salient landscape features (Laube et al., 2007). Pedestrian movements exhibit certain patterns that can be very different from the patterns generated by a car. Extracting such similarities can significantly contribute to the prediction, modeling and simulation of the collective behavior of moving objects and dynamic processes. Besides, it helps to generate and test scientific hypotheses about movement datasets, and supports data mining and rule discovery tasks (Faloutsos et al., 1994). That is, the results of similarity search can substantially be employed in three important data mining tasks, including, *a*) *finding movement patterns*; *b*) *clustering movement trajectories*; *c*) *classifying movement data* (in terms of object type, behavior). However, the available similarity analysis methods quantify the similarity between the movement paths of objects, and the movement characteristics of objects have been simply disregarded so far in the similarity analysis of movement data. Hence, the development of new similarity search methods, capable of discovering similar movement patterns, is essential in many fields interested in movement.

For instance, in environmental science meteorologists are interested to study the dynamic behavior of hurricanes and extract structure in the movements of hurricanes, one of the most destructive meteorological phenomena that can cause tremendous damage to human life. The magnitude of the hurricanes' destructions can compete with major earthquakes to the extent that six out of the ten costliest weather disasters in the history of the United States were caused by them. The tropical cyclone activities over the North Atlantic basin constitute 11 % of world wide activities, and have generated the most devastating hurricanes so far (Elsner and Kara, 1999). For instance, hurricane *Andrew* in 1992 destroyed parts of southeastern Florida and caused $26.5 billion in damage in the United States. In 2005, hurricane *Katrina*, with approximately 1200 reported deaths, produced catastrophic damage at about $75 billion[1]. Therefore, it is essential

[1] http://www.nhc.noaa.gov/HAW2/english/history.shtml

for meteorologists to gain knowledge about the dynamic behavior of hurricanes over their lifeline. Discovering similarities in the movement characteristics of hurricanes would help to discover the evolution patterns of hurricanes, model the movement of future hurricanes, forecast the time and location of recurvature points, and eventually, anticipate landfalls. Thus, exploratory methods that allow to discover knowledge and extract rules from the movement characteristics of the available repository of historical tracks of North Atlantic hurricanes[2] might help to model future trends of hurricanes.

1.2 Thesis Rationale

Movement parameters enrich trajectories with additional information about the characteristics of movement. This information can be used to achieve better insight into movement behaviors of objects and understanding the processes behind movement patterns. However, while GKD methods in GIScience have been well developed, very little attention has been dedicated to movement parameters and as a result, the development of methods for extracting useful information from these derivatives is still lagging behind. Therefore, the motivation of this thesis is to contribute to the conceptual and methodological knowledge about movement parameters in support of movement behavior studies. Specifically, the aim is to exploit movement parameters in developing new movement similarity search methods, allowing knowledge discovery from the movement of objects.

1.2.1 Research Questions and Research Objectives

Research Questions: Four research questions are formulated to be investigated in this study:

1. What are the constituting components of movement, essential for defining movement patterns?
2. How can we define a classification for different types of movement patterns? To what extent are movement patterns generic to various types of moving objects?
3. How can we reduce the complexity of a trajectory, while preserving its important movement features for knowledge discovery applications?
4. How can we quantify and formalize the similarity between the movement characteristics of different objects in space and time?

Research Objectives: With the development of a conceptual and methodological framework this thesis intends to develop tools to measure and explore similarities in the movement behavior of moving objects by using methods of GIScience and knowledge discovery in databases (KDD). The aim is to

[2] http://csc-s-maps-q.csc.noaa.gov/hurricanes/

explore the similarities between multiple moving objects. The aforementioned research questions are echoed in the following research objectives:

Objective 1: This research shall develop a conceptual framework, encapsulating essential elements that characterize the movement behavior of objects.

Objective 2: This research shall establish a comprehensive classification of movement patterns. The identified movement patterns shall be defined employing the elements of the conceptual framework.

Objective 3: This research shall identify, and formalize important features characterizing the movement of objects from the parameters of movement. Quantitative methods shall be developed to extract such features from raw trajectory data, with the aim of transforming trajectories into a simpler structure, while still conveying the important movement features.

Objective 4: This research shall quantify the similarity between movement behaviors of two individuals in space and time. The similarity measure shall consider dynamic properties of the movement parameters of objects. Accordingly, a similarity assessment approach shall be developed to investigate the similarity between both movement characteristics of multiple objects and their patterns of movement.

Objective 5: The applicability of the developed methods shall be evaluated in knowledge discovery tasks such as *(a) trajectory classification*, *(b) trajectory clustering*, and *(c) movement pattern discovery* in real movement datasets.

1.2.2 Research Process and Research Papers

This research pursues a three-stage process in order to achieve the main objectives and address the research questions. The research process for this thesis is illustrated in Figure 1.1. Each stage of this process is presented in a paper or two, and is associated directly with one or two research objectives (Figure 1.1). This thesis reports on the research process successively published in peer-reviewed international scientific journals or conferences. The order of the papers corresponds to the logical sequence of the research process.

Stage I: Theoretical Framework Development The first stage aims to provide the underlying knowledge about movement parameters and movement patterns. In this stage, the principal elements of movement are identified and described in a conceptual framework. The developed framework provides the elementary concepts for this research. A classification of movement patterns is developed based on the review of the relevant literature. This stage is presented in depth in Research Paper 1. Correspondingly, the first paper

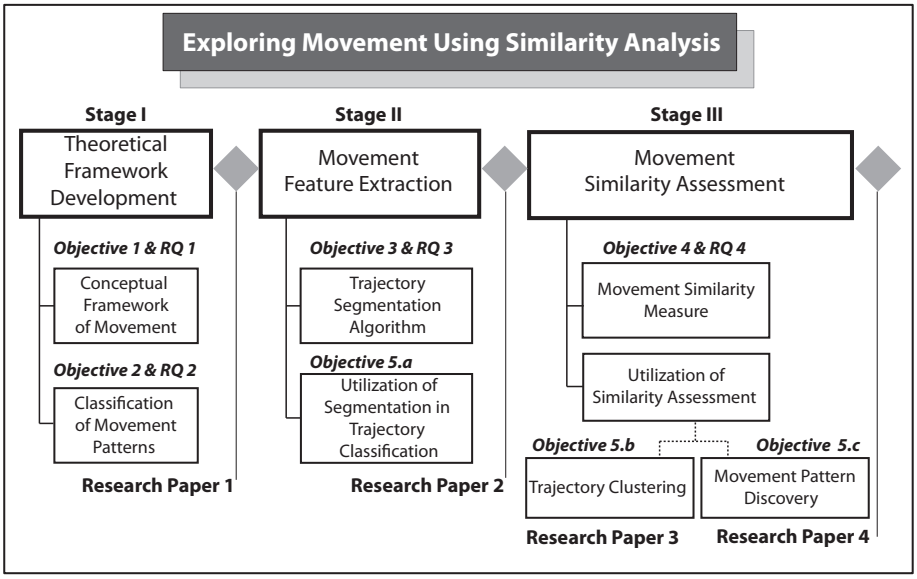


Figure 1.1: Research process of the thesis

covers Objectives 1 and 2 and addresses the associated research questions 1 and 2.

➤ Research Paper 1:

Dodge, S., Weibel, R. and Lautenschütz, A.K. (2008). Towards a Taxonomy of Movement Patterns. *Journal of Information Visualization*, Vol. 7, pp. 240 – 252.

Stage II: Movement Feature Extraction This stage provides methods that allow extracting movement features from the trajectories of different types of moving objects. *Movement features* are quantified using the properties of variations of movement parameters and hence describe the important dynamic properties of movement parameters of objects. These features are essential for reconstructing, modeling and analyzing trajectories of moving objects in applications such as trajectory classification, simulation and extraction of movement patterns. In this stage, a segmentation technique is developed for trajectories of moving objects that transforms trajectories into sections of *homogeneous movement properties*. Research Paper

2 reflects the developed methods and the outcomes of this stage. Consequently, it treats Objective 3 and the associated Research Question 3 of this research. In response to Objective $5_{(a)}$, this paper assesses the applicability of the developed methods in the domain of *trajectory classification*, as a case study. For that, distinct movement datasets from different domains such as eye-tracking and transportation are analyzed.

➤ **Research Paper 2:**

Dodge, S., Weibel, R. and Forootan, E. (2009). Revealing the physics of movement: Comparing the similarity of movement characteristics of different types of moving objects. *Computers, Environment and Urban Systems*, Volume 33, Issue 6, November 2009, pages 419 – 434.

Stage III: Movement Similarity Assessment This stage intends to define a similarity measure between the movement characteristics of two objects. Accordingly, two different movement similarity assessment approaches are proposed in Research Papers 3 and 4. The proposed techniques aim at finding structures in movement datasets by seeking trajectories that exhibit common patterns in the variation of their movement parameters. Thereby, this stage and its corresponding papers address the Objectives 4 and $5_{(b,c)}$ and accordingly respond to the Research Questions 4 of this thesis (see Figure 1.1).

Research Paper 3 presents the first approach as a novel trajectory similarity assessment technique that relies on the trajectory segmentation that was developed in the second stage. Further, the applicability of the technique is investigated in *trajectory clustering*. For this purpose, two clustering strategies are proposed and examined using distinct types of movement data from two different application domains (i.e. meteorology and transportation). The outcomes of the developed methods are assessed by comparison to a related geometric similarity search technique, using the hurricane dataset.

As the second approach, documented in paper 4, this stage introduces an alternative similarity assessment method that relies on two or more movement parameters. This method involves additional dimensions in comparison to the previous approach, where only one single movement parameter is exploited in similarity assessment at a time. Finally, the applicability of this method is evaluated in the *discovery of movement patterns* such as *coincidence* and *concurrence* patterns using North Atlantic hurricane data.

➤ Research Paper 3:

Dodge, S., Laube, P., and Weibel, R. (in revision, 2011). Movement Similarity Assessment Using Symbolic Representation of Trajectories. *International Journal of Geographic Information Science.*

➤ Research Paper 4:

Dodge, S., Weibel, R., and Laube, P. (2011). Trajectory Similarity Analysis in Movement Parameter Space. *GISRUK 2011*, April 27-29, 2011, University of Portsmouth, UK, pages 270 – 279, Short paper.

1.3 Structure of the Thesis

The content of this thesis is presented in two parts. In the first part (Synopsis) the above research papers are embedded in the scientific context. In the second part (Research Papers), the papers are presented with the content and format as they were submitted or published. In addition, a summary and key findings of each of these research papers can be found in chapter 3.

After this introductory chapter, the Synopsis continues with the theoretical background and a review of the state of the art (chapter 2). First, some principles of the movement model in mobility data mining and visual analytics disciplines are introduced. The chapter continues with a review to applications of similarity analysis in GKD in movement data. Following this, the concept of similarity for movement data is introduced and relevant techniques are reviewed. The Synopsis proceeds with an introduction to the research process, the methods used in this research, as well as the most relevant results of this study in chapter 3. Afterwards, in chapter 4 the individual contributions of the four papers is put into a broader context and the substances of the papers are discussed. Finally, the synopsis ends with conclusions and an outlook on future research (chapter 5).

Chapter 2

Theoretical Background and State of the Art

The overall aim of this chapter is fourfold: First, to illustrate the concept of movement itself in a generic model and describe the main components of movement model that are central to this research (section 2.1). Second, to capture the scope of movement research in a broad perspective (section 2.2) and give an overview of the previous and ongoing trends of research studies on movement (section 2.3). Third, to introduce the process of *knowledge discovery* in movement datasets (section 2.4). And finally, to introduce the *movement similarity analysis problem* and report on the relevant literature (section 2.5). However, first of all key terms of this thesis need to be explained:

Movement is "a fundamental characteristic of *life*, driven by processes that act across multiple spatial and temporal scales". Movement for the purpose of this research is defined as "a *change* in the spatial location of the whole individual in time", that is, as whole-body movement (Nathan et al., 2008, p.19052).

Frank (2001, p.22) identifies two forms of *change* in general: "change of the objects of interest" (appear, disappear, merge, and split); and "change in the position or geometric form of these objects" (move). Specifically, temporal change in the life of an individual (i.e. humans, animals) can take place in three forms of "*birth*, *death*, and *movement*" (Turchin, 1998, p.2). This thesis only deals with *movement* based on the Lagrangian approach, which in contrast to other forms of change, has at least two dimensions, namely, *temporal* and *spatial* (Turchin, 1998). Figure 2.1 represents the movement of an ant in a schematic way.

Trajectory In this thesis, the movement of an individual is represented by its *trajectory*, also called *geospatial lifeline*, as a time-ordered set of positions (Laube et al., 2007; Spaccapietra et al., 2008) as shown in Figure 2.1. For practical reasons of measuring or observing the positions of a moving object over time a *trajectory* consists of a series of *discrete* space-time observations. In this thesis, an observation point along a trajectory is referred to as a *trajectory fix* (or *fix*), like the ones shown in Figure 2.1.

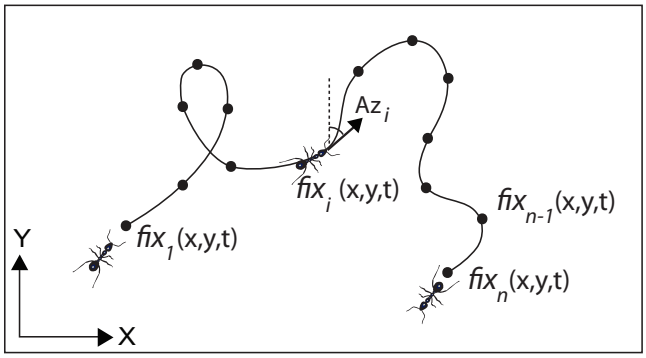


Figure 2.1: Trajectory of a moving object (an ant), representative of its movement path over time.

Moving object (also called *mobile object* or *dynamic object*) is defined as an entity whose position changes over time (e.g. the ant in Figure 2.1). In this thesis, moving objects are conceptualized by moving points, called *Moving Point Object (MPO)* (i.e. as $0 - D$ geometric entities). That is, the location of the object in time is considered to be more important than its dimension. Location is usually indicated using geographic coordinates (ϕ, λ) or Cartesian coordinates $(x, y, (z))$. Accordingly, in this thesis the location of a moving object at time t is specified by a tuple $(x, y(, z), t)$ of coordinates. We consider two categories of moving objects: *a*) georeferenced (i.e dynamic objects or processes that move about in geographic space, such as animals, vehicles, humans, and hurricanes); and *b*) non-georeferenced dynamic objects (i.e. dynamic phenomena that move in a non-geographic space such as gaze point movements of eyes) .

Movement Parameters (MP) comprise the measurable quantities of movements, that can be observed along objects' geospatial lifelines, and their derivatives. Movement parameters are divided into two types of *instantaneous* parameters (i.e. detectable at individual moments) such as position, speed, acceleration, direction (e.g. as represented by the azimuth Az_i at fix_i in Figure 2.1) and *relative* parameters (i.e. measurable over time intervals) such as relative speed, turning angle, and path sinuosity (Laube et al., 2007; Giannotti and Pedreschi, 2008).

Movement Features are properties of movement parameters of an object, essential to characterize its movement. In this thesis, movement features are identified as the amplitude and frequency of the variations of movement parameters.

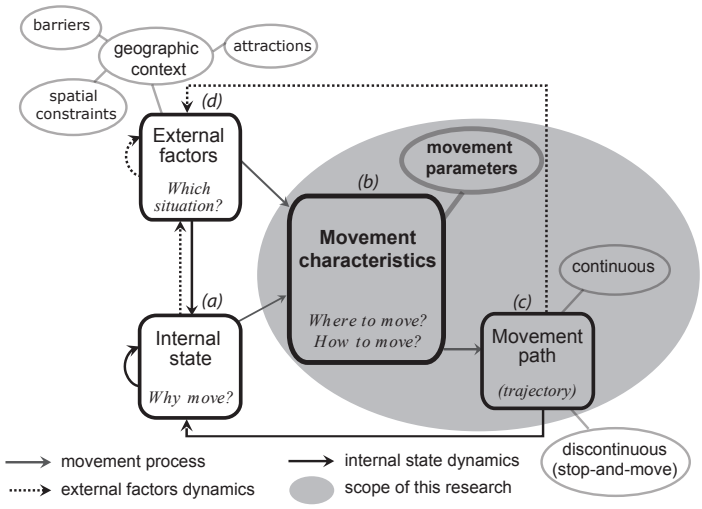


Figure 2.2: Conceptual model of movement (an extended version of the framework by Nathan et al., 2008)

Dynamic behavior (also called *movement behavior*) of a moving object is referred to as the dynamic *movement characteristics* of an object and the way that the object moves during the whole duration of observation or an episode (Dykes and Mountain, 2003) of its geospatial lifeline. In this thesis, the dynamic behavior gives an indication of variations or trends of movement parameters of the moving object over time.

2.1 Conceptual Model of Movement

The basis of this research is a theoretical model of movement, illustrated in Figure 2.2. This model is an extended version of the conceptual framework introduced in Nathan et al. (2008). The model was originally proposed for movement ecology, in order to explore the causes, mechanisms, and patterns of movements of organisms (Nathan et al., 2008).

The model is composed of four major components, including (*a*) *internal state*; (*b*) *movement characteristics*; (*c*) *movement path*; and (*d*) *external factors* (Figure 2.2). The first three components are related to the focal individual (i.e. the moving object under study). The latter is related to the environment within which the movement takes place. The focal individual is specified by the object's particular intrinsic physical and behavioral movement properties.

Internal state: The internal state of the individual is denoted by its physiological and, where appropriate, its psychological specifications such as the motivation of an individual for movement activity, readiness to move, and ability to execute and orient its movement.

Movement characteristics: In the original version of this model by Nathan et al. (2008) the movement characteristics are divided into motion capacity and navigation capacity, which are related to the focal individual. Similarly, here the movement characteristics encompass both capacities, where appropriate, as well as *positional* and *temporal* information of movement, and the physical *movement parameters* of the individual. These parameters can be derived from the trajectory of the focal object or can be captured directly from sensors.

Movement path: The movement process generates a movement path, which can take two forms: Namely, *continuous path* (i.e. curvilinear path), exemplified by the trajectory of a hurricane or a pedestrian, and *discontinuous path* (i.e. steps) with a series of stops-and-moves (e.g. saccadic movement of eyes, trajectory of a butterfly or a bee between flowers). In the original framework, the movement path is considered as a product of movement. Here, we take it as one of the principal elements of the movement model, since for some objects the movement path is constrained and predefined by the geographic context such as the road network or barriers (e.g. rivers).

External factors: The external factors represent the surrounding environment and the context of the movement activity as well as factors constraining or triggering the movement (e.g. spatial constraints, barriers, and attractions, weather condition etc.).

The relationships between the four components are shown with arrows in Figure 2.2. The arrows indicate the direction of influence of the components on each other. The movement process is performed when the individual is triggered by an internal motivation (i.e. readiness to start an activity). However, the external factors affect the way the individual performs the movement process (i.e. movement behavior). And finally, the movement characteristics of the individual have an impact on the geometry of the movement path.

The scope of this research extends over those aspects of the focal individual that are related to its movement characteristics and movement path. The main focus is on movement parameters and the patterns of their variation over time. This research mainly deals with the continuous type of movement paths. However, a comparison between movement characteristics of objects generating the two different forms of movement paths is carried out in the second stage of this thesis, which is presented in Research Paper 2 (page 97). In this thesis, the *trajectory* approach is employed in order to represent the continuous movement path of objects 2.1.

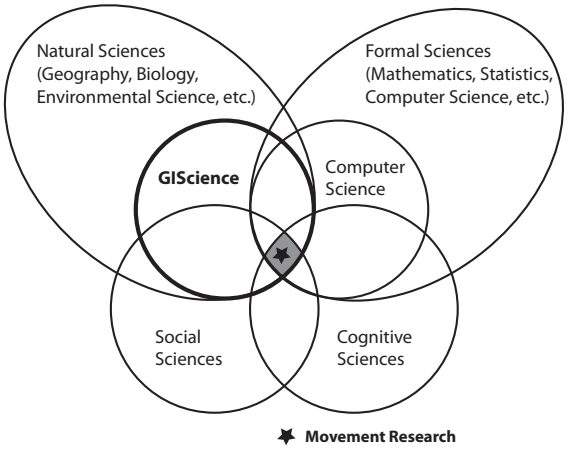


Figure 2.3: Movement is a multidisciplinary research area

2.2 The Scope of Movement Research

Movement is essential to almost all organisms and spatio-temporal processes (even plants show seed dispersal, although most do not move themselves). In fact, movement research is a multidisciplinary/interdisciplinary field, which deals with dynamic aspects of moving entities and processes as well as the collective behavior of moving objects. As a consequence, the study of movement is a key interest of many disciplines in science and technology, as evidenced by the vast amount of literature published on the subject during the past decade. Accordingly, movement research applies to many fields, such as movement ecology, behavioral studies, transportation, or disaster management, to name but a few. For instance, in the domain of movement ecology alone, as observed by Holyoak et al. (2008), nearly 26000 published articles referred to movement of organisms and tackled issues such as *measuring* and *describing* movement, and *testing hypotheses* about it. Similarly, recent years have witnessed almost an explosion of research activities on movement datasets, triggered by the advent of cheap and ubiquitous positioning technologies, in many disciplines such as GIScience, computer science, environmental science, social science, and cognitive science.

Figure 2.3 illustrates the multidisciplinary nature of movement research on a broad scale. Also, it highlights the position of *movement research*, and accordingly the current research, with respect to the related disciplines. As it can be seen in the figure, movement research crosses many disciplinary boundaries, such as environmental sciences (e.g. biology, meteorology, geography), formal sciences

(e.g. mathematics, statistics, computer science), social sciences, and cognitive sciences. That is, movement research focuses on problems that cross the boundaries of two or more disciplines. For instance, knowledge discovery and analytical methods that are developed in GIScience or computer science, can be applied in movement ecology to study the movement behavior of animals.

The remainder of this section briefly summarizes the state of the art of movement studies from the GIScience perspective. The Research Papers in Part II will review further literature, tailored to the scope of the paper.

2.3 Movement Research in GIScience

The importance of temporal aspects of movement has attracted a range of studies in GIScience and related disciplines, including investigations of space-time settings (i.e. space-time path, prism, and station) (Hornsby and Egenhofer, 2002; Miller, 2005), modeling moving objects and their collective dynamics (Erwig et al., 1999; Galton, 2005; Laube et al., 2007), development of new analytical methods for movement pattern discovery (Imfeld, 2000; Laube, 2005), exploratory data analysis, and visual analytics techniques for movement (Kraak, 1988; Andrienko and Andrienko, 1999, 2009; Dykes and Mountain, 2003).

Figure 2.4 shows the interdisciplinary research areas related to movement studies in GIScience. Recognizing the broad scope of movement research in GIScience, Mountain (2005) illustrated the scope of movement-related studies in GIScience and described how different aspects of movement have been studied in various contexts such as *time geography*, *modeling and prediction*, *geographic data mining and knowledge discovery*, *geovisualization*, *information retrieval*, and *mobile computing*. Further, in a comprehensive literature review, Laube (2005) documented the progress of movement-related research, until 2005, in the aforementioned disciplines in different interrelated tasks such as *capturing*, *quantifying*, *modeling*, *formalizing*, *querying*, *visualizing*, *analyzing*, and *simulating* movement. Later, Giannotti and Pedreschi (2008) reported the recent advance of related studies, specifically after 2005, with more focus on the latest research efforts on *privacy and security* issues, as well as *"trajectory data mining"* (the term was first coined by Smyth 2001) techniques.

The above publications document significant progress of movement research in GIScience over the past few years. Since knowledge discovery forms the main focus of this work, we introduce the knowledge discovery process from movement data and highlight the related research advances in the following section.

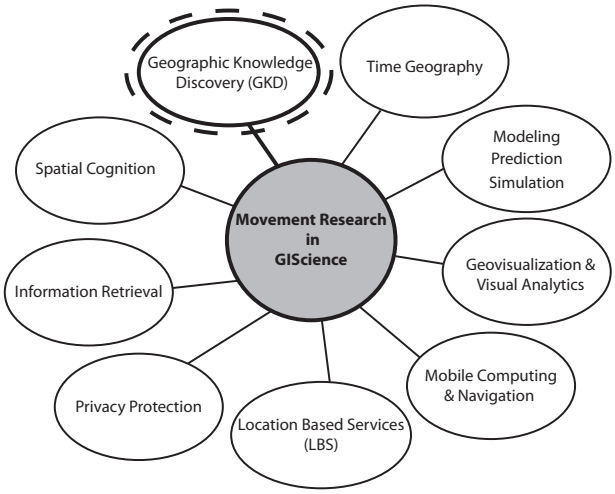


Figure 2.4: Interdisciplinary research on movement in GIScience

2.4 Geographic Knowledge Discovery in Movement Data

2.4.1 Overview

Knowledge Discovery in Databases (KDD) is defined as a "process of obtaining information through data mining and distilling this information into knowledge through interpretation of information and integration with existing knowledge" (Miller and Han, 2009, p.3). According to this definition, the KDD process consists of several steps, namely, data selection, data preprocessing, data enrichment, data reduction and projection, data mining and pattern recognition, and reporting. Data mining is a step of the KDD process that refers to the application of low-level techniques for revealing hidden information and patterns in a database (Fayyad et al., 1996). Data mining tasks include: segmentation or clustering, classification, association, deviations, trends and regression analysis, generalizations (Miller and Han, 2009).

Geographic Knowledge Discovery (GKD) is a special case of KDD that deals with geographic information. Geographic information possesses distinct properties such as *high dimensionality*, *topology*, *geometry*, *spatial dependency*, and *spatial heterogeneity*, and hence, demands careful consideration in comparison to the other datasets (Miller and Han, 2009). In addition, spatio-temporal data, including moving objects and processes, are usually more complex than other geographic data. Hence, in the context of GKD proper analytical techniques need to be developed to capture different aspects of spatio-temporal data (Im-

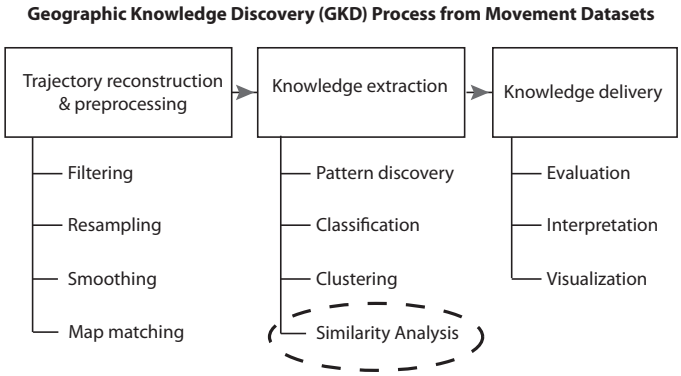


Figure 2.5: GKD process in movement databases (extending the textual descriptions of Giannotti and Pedreschi, 2008)

feld, 2000). Sophisticated knowledge discovery and data mining techniques are required to adequately analyze such data.

2.4.2 GKD Process in Movement Databases

Recognizing the importance and complexity of moving object data, the most remarkable advances in GKD have taken place in spatio-temporal and moving object databases alongside the increased production of such datasets using positioning technologies, and geo-sensor networks (Miller and Han, 2009). In this context, Giannotti and Pedreschi (2008) introduced three main steps for knowledge discovery from movement databases, including, *trajectory reconstruction*, *knowledge extraction*, and *knowledge delivery*, as shown in Figure 2.5.

Trajectory reconstruction and preprocessing

In the first step, the subject movement datasets shall be processed to obtain trajectories of individual moving objects. The preprocessing stage involves all or some of the following processes depending on the purpose of analysis. The order may vary and some steps may even be revisited according to the application domain.

- *Filtering*, to detect and remove outliers using statistical approaches.
- *Resampling*, to obtain regularly sampled trajectories (fixed time granularity) using interpolation techniques.
- *Smoothing*, to remove the effect of noise from tracking data (e.g. GPS data) using approximation techniques such as Kalman filtering, moving average,

or Kernel-based smoothing (Jun et al., 2006).

- *Map matching*, to match the position of the individual with the actual map; mostly relevant in transportation and navigation applications (Bernstein and Kornhauser, 1996; Brakatsoulas et al., 2005).

Knowledge extraction

This stage refers to the process of exploiting knowledge discovery and data mining techniques, in order to discover *patterns* and *structure* in movement data and acquire useful knowledge about the behavior of moving objects. The knowledge extraction process can be carried out using the main data mining techniques such as *pattern discovery*, *classification*, *clustering*, and *similarity analysis* (Miller and Han, 2009; Giannotti and Pedreschi, 2008) (Figure 2.5). These techniques are briefly expanded on in the following:

a) Movement pattern discovery:
Movement pattern discovery is referred to as the process of finding *interesting patterns* in a large movement dataset by applying data mining methods such as exploratory data analysis, descriptive and predictive modeling, mining association rules, and other pattern recognition techniques. In their definition of KDD process, Fayyad et al. (1996) relate *pattern extraction* to fitting a model to data, finding *structure*, or making any *high-level description* from data.

Definition. *A **pattern** reflects the behavior of a subset of data (Andrienko and Andrienko, 2007), and is defined as non-random properties and relationships that are valid, novel (i.e. nontrivial, unexpected), useful, understandable (i.e. simple, interpretable), and interesting (Fayyad et al., 1996; Laube and Purves, 2006).*

Research Paper 1 (page 81) provides a broad overview and a classification of different types of movement patterns from the related literature. Recently, in a comprehensive review Laube (2009) documented the research progress on the development of techniques to formalize, discover, and understand movement patterns.

b) Trajectory classification:
In KDD, *classification* is denoted as "finding rules or methods to assign data items to pre-existing classes" (Miller and Han, 2009, p.7). Accordingly, trajectory classification is defined as the process of applying model construction, segmentation, and recognition algorithms for identifying the class labels (i.e type) of moving objects based on their movement trajectories (Lee et al., 2008). Trajectory classification is very important in real world applications. For instance, extraction of information about the mode of transport (e.g., bicycle, car, train, and boat) from a movement dataset is essential for domains

such as travel behavior research, transportation planning, and traffic management. A number of studies applied classification techniques in modeling and differentiating moving object trajectories in imagery and video surveillance databases (Fraile and Maybank, 1998), recognition of object activities (Bashir et al., 2007), and behavior studies of individuals (Blythe and Miller, 1996; Bay and Pazzani, 2001), to name but a few.

c) ***Trajectory clustering:***
Trajectory clustering is one of the exploratory data mining techniques that facilitate studying movement data and understanding its structure by reducing its complexity. In general, *clustering* is defined as the process of grouping a set of objects into classes of *similar* objects. Trajectory clustering is a process of grouping moving object trajectories based on their spatial and/or temporal similarity. It can be applied to identify typical trends in datasets; and hence, supports *deviation analysis* to detect outliers and anomalies in data (Miller and Han, 2009). Furthermore, trajectory clustering can support *data aggregation* in empirical user studies to gain a better understanding of dynamic cognitive processes and for evaluation purposes (Fabrikant et al., 2008; Çöltekin et al., 2010).
Miller and Han (2009) and Kisilevich et al. (2010) provided a survey of the recent progress in the development of trajectory clustering techniques. Overall, trajectory clustering techniques can be classified into two main categories:

i) *distance-based clustering approaches*, such as hierarchical and K-means clustering (Miller and Han, 2009), where a distance function is required to compute the distance (i.e. dissimilarity) between trajectories in space, or in space and time.

ii) *density-based clustering approaches*, such as DBSCAN (Ester et al., 1996) and OPTIC (Ankerst et al., 1999), where clusters are identified as a dense region in space based on a density threshold.

A large number of proposed trajectory clustering approaches rely on the similarity of the geometric shapes (Fu et al., 2005; Lee et al., 2007; Rinzivillo et al., 2008; Giannotti and Pedreschi, 2008; Miller and Han, 2009; Li et al., 2010). Geometric clustering techniques proposed so far do not necessarily capture spatio-temporal similarity between the movements of objects. Additional information is required to cluster trajectory data according to the spatio-temporal characteristics of moving objects. In this respect, recent work has focused on developing spatio-temporal clustering techniques for trajectory data (Nanni and Pedreschi, 2006; Etienne et al., 2010). However, this problem has not been fully addressed so far and effective techniques still need to be developed.

d) ***Movement similarity analysis:***
Movement similarity analysis, also called *movement similarity assessment*, is referred to as the process of finding similarities in a large dataset, and is a key task in knowledge discovery. In fact, similarity analysis can also be seen as a low-level knowledge extraction technique, since its outcomes can substantially be exploited in the aforementioned data mining techniques (i.e. pattern discovery, classification, and clustering). For instance, most movement patterns, such as *flocking* and *concurrency* (Laube, 2005), emerge from *similarity* in one or several movement parameters. Also, clustering and classification processes rely on existing *similarities* among objects in datasets. Specifically, similarity assessment is a prerequisite for the first group of clustering approaches (i.e. *distance-based clustering*). Therefore, it is crucial to develop effective approaches to assess and extract similarities in movement data.

Considering that the major focus of this thesis is on *similarity analysis of movement data*, section 2.5 gives a detailed state of the art of the related literature.

Knowledge delivery

After the knowledge extraction process, it is essential to reason about the detected patterns, and evaluate the reliability, meaningfulness, and interestingness of the outcomes. Effective visualization techniques are required in order to appropriately present the results, support suitable interpretation of the results, and eventually deliver the appropriate knowledge about the subject movement dataset (Giannotti and Pedreschi, 2008).

2.5 Similarity Analysis of Movement Data

2.5.1 Overview

Similarity analysis is important as an exploratory tool in many data mining and knowledge discovery applications. The concept of similarity is domain specific and may even vary according to the purpose of queries within the same application (Faloutsos et al., 1997). In general, the *similarity analysis* is defined as in Alt and Guibas (2000):

Definition. *Given two objects T_1 and T_2, the aim is to determine how much the two objects resemble each other.*

According to this definition, similarity analysis problems can be classified into two categories (Agrawal et al., 1993; Alt and Guibas, 2000):

a) *complete similarity*, also called *whole matching* and *complete matching*, where complete objects, each as a single unit, are to be compared to each other.

b) *partial similarity*, also called *partial matching* or *subsequence matching*, where some parts of objects that best match to each other are to be sought. Partial similarity is further classified into two groups of: *i*) *whole-to-part*, and *ii*) *part-to-part* similarity.

In the data mining literature, *similarity* between two objects T_1 and T_2 has been often measured by the inverse of *dissimilarity*. *Dissimilarity* is quantified as the cost of transforming one entity into another, or the distance between the two objects (Faloutsos et al., 1997). The distance function, hereinafter denoted by $D(T_1, T_2)$, which quantifies the dissimilarity between objects, provides the basis of a similarity analysis technique and is called *similarity measure*. So far, a variety of *similarity measures* have been developed in order to address various aspects of similarity analysis problems in different application domains, such as pattern matching in images, text, and video datasets; sequence matching; and geometric shape matching, to name but a few.

The overall goal of this section is to review the similarity analysis techniques suggested in the literature from movement-related disciplines. In this literature review, first a short overview of the previous similarity search efforts in (*1*) *time series analysis*, and (*2*) *computational geometry* are presented. The reason is that these domains have a very close connection to movement research studies. Moreover, most of the similarity measures proposed so far for movement data have been inspired from such research. This is due to the fact that the trajectory of a moving object either can be modeled as a sequence of time-ordered coordinates (3D or 4D) of points, or it can be considered as a set of points representative of a static geometric shape, often as a curve, ignoring the time dimension. Therefore, most of the similarity analysis techniques developed for time series analysis or geometric shapes can be adapted to movement data mining applications.

2.5.2 Time Series Similarity Measures

A large body of research has investigated similarity analysis techniques on time series data such as financial, marketing and production time series (e.g. stock prices) as well as scientific time series data (e.g. weather data, sea level data etc.) in order to predict future trends, and test hypotheses (Agrawal et al., 1993; Faloutsos et al., 1994; Ding et al., 2008b). Generally speaking, almost every proposed similarity measure for time series data can be identified as a variation of one of the following basic groups of similarity measures:

- Minkowski distance (L_p-Norm family)
- Dynamic Time Warping (DTW)
- Edit distance
- Longest Common Subsequence (LCSS)
- Distances based on local features

Minkowski distance (L_p-Norm family)

The Minkowski distance of order p ($p \geq 0$) defines a series of metrics to compute the distance between two entities (T_1 and T_2) in vector space, as:

$$L_p(T_1, T_2) = \sqrt[p]{\sum_{k=1}^{n} (T_1^i - T_2^i)^p}, \quad |T_1| = |T_2| = n \tag{2.5.1}$$

In time series analysis, the Minkowski distance is usually used with $p = 1$ (i.e. *Manhattan* distance) or $p = 2$ (i.e. *Euclidean* distance) (Chen and Ng, 2004).

The **Euclidean distance** has been widely employed as a distance function in time series similarity analysis techniques on real sequences or other representations of sequence data (Agrawal et al., 1993; Faloutsos et al., 1994, 1997). It is defined as the sum of squared differences at each point of sequences of the same length, with

$$D_{euclidean}(T_1, T_2) = L_2 = \sqrt{\sum_{k=1}^{n} (T_1^i - T_2^i)^2}, \quad |T_1| = |T_2| = n \tag{2.5.2}$$

Agrawal et al. (1993) pioneered applying the *Euclidean distance* in time series similarity analysis. In their study, a whole matching method on the *Discrete Fourier Transform* representation of time series is introduced. As an extension of that work, Faloutsos et al. (1994) used Fourier transforms to map each sequence into a small set of multidimensional rectangles in a feature space. Euclidean distance is then applied to measure the similarity between the points in the feature space. Later, Faloutsos et al. (1997) introduced a *signature based technique* to search for similar signatures (*shrank* sequences) instead of the real sequences to speed up the matching process. Accordingly, in their study a different similarity measure is proposed as the combination of the *Euclidean distance*, as the base cost function, and the *Edit distance* (described later in this section, page 24), as the cost of transforming one signature to another. Subsequent work has mostly focused on developing new dimension reduction techniques, while using the Euclidean distance as the similarity measure. Examples include: *Singular Value Decomposition*, *Discrete Wavelet Transform*, and *Piecewise Aggregate Approximation* (Keogh et al., 2001).

Dynamic Time Warping (DTW)

DTW as a classic speech recognition tool, was first introduced by Berndt and Clifford (1994) for the purpose of pattern matching in time series. The method allows a time series to be *stretched* or *compressed*, and hence provides a better

match between time series. In fact, DTW is a transformation that temporally warps a sequence in order to minimize the distance between two sequences. Such transformation between two sequences $T_1[1 \ldots n]$ and $T_2[1 \ldots m]$ can be computed using a *dynamic programming algorithm* with complexity of $O(n \times m)$. The DTW distance between sequences T_1 and T_2 is defined as follows (Berndt and Clifford, 1994):

$$\begin{aligned} &D_{dtw}(T_1, T_2) = f(n, m) \\ &f(i, j) = ||T_1^i - T_2^j|| + min \begin{cases} f(i, j-1) \\ f(i-1, j) \\ f(i-1, j-1) \end{cases} \end{aligned} \tag{2.5.3}$$

where $i = 1, \ldots, n$ and $j = 1, \ldots, m$.

DTW allows local acceleration and deceleration in the rate of the time series, hence, it does not preserve the natural features of a time series (Yi et al., 1998). To overcome this problem, Yi et al. (1998) proposed an approach called *FastMap*, to map sequences to a set of points in a k-dimensional space prior to applying DTW. This approach preserves the original distances between the sequences. However, since DTW does not satisfy *triangle inequality* (which is a property required from a *metric* distance), *FastMap* may introduce false dismissals.

Another major challenge of DTW is that it incurs a heavy computation cost. In this regard, several studies suggested various strategies in order to speed up similarity search using DTW (Ding et al., 2008b). As an example, Sakurai et al. (2005) proposed a fast search method for time warping, called *Fast Time Warping*, which reduces the search cost. The proposed method prunes the search candidates using a lower bounding distance measure and generates no false dismissals in contrast to the previous approach (Sakurai et al., 2005).

Edit distance

The *edit distance* was originally introduced for pattern matching in alphanumeric datasets. The *Levenshtein distance* is one of the most famous edit distances which has been widely used in string matching and sequence analysis. The Levenshtein (or edit distance) is defined as the minimum number of operations that are needed to convert a query text to a pattern text (Levenshtein, 1966). Three operations, *deletion*, *insertion*, and *substitution* are considered in the conversion process, whereas the cost of each operation equates to 1 (Crochemore and Rytter, 1994; Bozkaya et al., 1997).

The edit distance between two sequences $T_1[1...n]$ and $T_2[1...m]$ is computed

using Equation (2.5.4) in $O(n \times m)$ time (Bozkaya et al., 1997):

$$D_{edit}(T_1, T_2) = f(n, m)$$
$$f(i,j) = \begin{cases} j & \text{if } i = 0 \\ i & \text{if } j = 0 \\ f(i-1, j-1) & \text{if } i, j > 0 \text{ and } T_1^i \text{ is equal to } T_2^j \\ 1 + min \begin{cases} f(i-1, j-1) \\ f(i-1, j) \\ f(i, j-1) \end{cases} & \text{otherwise} \end{cases} \tag{2.5.4}$$

where $i = 1, \ldots, n$ and $j = 1, \ldots, m$.

In the context of time series similarity analysis, Bozkaya et al. (1997) proposed a modified version of the edit distance, where the *change* operation is not allowed, together with an indexing procedure, in order to retrieve similar sequences of different length from a large dataset. In another attempt, Chen and Ng (2004) introduced the *Edit distance with Real Penalty (ERP)*, as an integration of the L_1 norm (*Manhattan distance*) and the edit distance to support local time shifting in the retrieval of similar time series. This approach uses the *insertion* operation with the cost of 1 for each added value (called *gap*) in order to compute the distance between *gaps* of two time series.

Longest Common Subsequence (LCSS)

LCSS was initiated from the concept of *edit distance*. The LCSS considers "two sequences to be similar, if they have enough non-overlapping time-ordered pairs of subsequences that are similar" (Agrawal et al., 1995, p. 491). Therefore, the LCSS allows the amplitude of sequences to be *scaled* by a threshold δ, whereas it prevents elements from rearranging. LCSS permits some elements to remain unmatched in contrast to DTW and Euclidean distance. This is useful when data contain *outliers* (Agrawal et al., 1995; Das et al., 1997; Vlachos et al., 2002a). Like edit distance and DTW, the computational complexity of LCSS using a *dynamic programing* approach is $O(n \times m)$. The LCSS distance between two sequences $T_1[1...n]$ and $T_2[1...m]$ is defined as follows (Vlachos et al., 2002a):

$$D_{lcss}(T_1, T_2) = f(n, m)$$
$$f(i,j) = \begin{cases} 0 & \text{if } i = 0 \text{ or } j = 0 \\ f(i-1, j-1) + 1 & \text{if } T_1^i = T_2^j \\ \max\left(f(i, j-1), f(i-1, j)\right) & \text{if } T_1^i \neq T_2^j \end{cases} \tag{2.5.5}$$

where $i = 1, \ldots, n$ and $j = 1, \ldots, m$.

Distances based on local features

These distances are a new concept of time series similarity measures that have been recently introduced in time series data mining, reflecting the dynamic trends of time series. That is, they use the *physical notion* of a pattern, such as the pattern of *amplitude* or slope variations in time series. Such measures have a clear *physical meaning*, and are more intuitive and usually require a simple calculation procedure (Yuelong et al., 2008).

- **Amplitude features**
  As one of the pioneering works, Aßfalg et al. (2006) introduced the notion of *threshold queries (TQuEST)* in time series databases. Based on this approach, a time series is decomposed into a sequence of segments, called *Threshold-Crossing Time Intervals*, with amplitude greater than a user defined threshold δ. Each segment is then transformed to a point in a two dimensional space using the start time and the end time of the segment, as the two dimensions. The similarity between two time series is then measured by the sum of minimum Euclidean distances between the points of the transformed sequences in the new space (Aßfalg et al., 2006). As an extension of this work, Aßfalg et al. (2008) introduced another similarity analysis approach for time series, called *Amplitude-Level Features*. This method is based on a decomposition of complex structured time series into a set of simpler structured segments, using several amplitude levels (instead of only one level in the previous method). The generated segments are then transformed into a set of feature vectors for the purpose of similarity search as described earlier (Aßfalg et al., 2008).
- **Slope features**
  In other studies, a different approach has considered the slope features of time series (Toshniwal and Joshi, 2005; Yuelong et al., 2008). These methods detect time series that exhibit similar variations in their slopes. The *slope-based distance* utilizes the Euclidean distance on the slopes of the ith interval of two time series (S^i), as given by Toshniwal and Joshi (2005):

$$D_{slope}(T_1, T_2) = \sqrt{\sum_{k=1}^{n} (S^i_{T_1} - S^i_{T_2})^2}, \quad |T_1| = |T_2| = n \tag{2.5.6}$$

  Toshniwal and Joshi (2005) further extended the concept of slope distance to the *time weighted slopes distance*, in order to avoid distortion due to the stretching of time series of different length in the similarity search process. The method scales the amplitude of the time series by applying a weight proportional to the time intervals between the consecutive points.

General comparison of time series similarity measures

In a comparative study, Ding et al. (2008b) investigated the effectiveness of a number of aforementioned measures, including *Euclidean distance*, *DTW*, *LCSS*, *ERP*, and similarity search based on *Threshold Queries (TQuEST)*. The study showed that on a small datasets, DTW, LCSS, and ERP can be significantly more accurate than the Euclidean distance. However, the accuracy of those measures is comparable to the Euclidean distance in larger datasets. The Euclidean distance and DTW provided better results in comparison with TQuEST on the subject datasets. Furthermore, the experimental results of their study suggested that the accuracy of edit based similarity measures such as LCSS and ERP is very close to the accuracy of the DTW method (Ding et al., 2008b).

The Euclidean distance is a metric measure and easy to compute in linear time (i.e. $O(n)$). However, this measure only works for sequences of same length and does not support *local time shifts* (i.e. time lags between similar trajectories). The effectiveness of the Euclidean distance decreases in the presence of noise and outliers. ERP is metric and supports local time shifting, whereas DTW and LCSS support local time shifting, yet are not metric (Chen and Ng, 2004).

2.5.3 Geometric Similarity Measures

The geometric techniques for matching and analyzing geometric shapes are of great interest in many disciplines such as computer vision, robotics, pattern recognition, molecular biology, and cartography. Among such techniques, *geometric shape matching* has been employed in many movement data mining efforts, specifically in the development of algorithms for finding *movement patterns* and *similarity* between trajectory data. These techniques aim to measure the similarity or distance between two shapes, often under certain transformations such as *translation*, *rotation*, or *scaling* (Goodrich et al., 1999; Alt and Guibas, 2000). In order to compute the distance between shapes, various geometric similarity measures have been used in the computational geometry literature.

This section presents two major classes of geometric similarity measures that are later extended in the development of movement similarity search methods, namely, the *Hausdorff distance* and *Fréchet distance*. Most of the subsequent research efforts were focused on developing efficient algorithms to speed up the computation of these two measures (Alt and Godau, 1995; Eiter and Mannila, 1994; Goodrich et al., 1999; Buchin et al., 2008; Alt, 2009).

Hausdorff distance

For two point sets A and B, representative of two geometric shapes (e.g. curves), the *Hausdorff distance* is defined as the maximum of the minimum Euclidean distances between each point of one set to the other. Accordingly, the *one-sided*

Hausdorff distance from A to B is computed as (Alt and Guibas, 2000):

$$D_H(A, B) = \max_{a \in A} \min_{b \in B} \parallel a - b \parallel \tag{2.5.7}$$

where $\parallel . \parallel$ is the Euclidean distance (see Equation (2.5.2)). The *bidirectional Hausdorff distance* is then defined as:

$$D_H(A, B) = \max(D_H(A, B), D_H(B, A)) \tag{2.5.8}$$

The computation cost of the Hausdorff distance is $O(n \times m)$, when A to B are composed of finite sets of n and m points, like *'point patterns'*. However, the problem becomes more complex when A to B are continuous curves or composed of sets of line segments. Furthermore, the Hausdorff distance does not take into account the order information of a curve. Therefore, curves that are perceived with completely different geometric shape might have a small Hausdorff distance (Alt, 2009).

Fréchet distance

The Fréchet distance is a measure of similarity between two curves that is defined on the continuous parameterizations of the curves ($\alpha, \beta : [0, 1] \longrightarrow \mathbb{R}^2$) as follows (Alt and Godau, 1995; Alt, 2009):

$$D_F(A, B) = \inf_{\sigma, \tau} \max_{t \in [0,1]} \parallel \alpha(\sigma(t)) - \beta(\tau(t)) \parallel \tag{2.5.9}$$

where $\parallel . \parallel$ is the Euclidean distance and $\sigma, \tau : [0, 1] \longrightarrow [0, 1]$ range over continuous and increasing functions.

An intuitive illustration of the Fréchet distance is to imagine that a man is walking his dog, both walking on their respective path curves. While both are allowed to control their speed, they are not allowed to move backward. The Fréchet distance of these two curves is defined as the minimal length of a leash that is necessary along the walk.

In comparison to the Hausdorff distance, the Fréchet distance is more difficult to compute. However, the Fréchet distance performs better in terms of capturing the similarity of shapes as perceived by human observers since it takes into account the order of the points traversed by the curves (Alt, 2009).

2.5.4 Trajectory Similarity Measures

Similarity analysis is a fairly new topic in the context of knowledge discovery and data mining in movement data. Recently, researchers developed a great interest into conceiving new methods to deal with the *trajectory similarity assessment*, which aims at finding *similar* trajectories of moving objects. Trajectory similarity measures proposed so far can be classified into two major categories:

- **spatial similarity measures**, which assess the similarity of trajectories based on their geometric shape, ignoring the temporal dimension.
- **spatio-temporal similarity measures**, which assess the similarity of trajectories considering both the spatial and temporal dimensions.

This section presents an overview on the available trajectory similarity literature. The review covers the previously proposed similarity measures for movement data and discusses the strengths and weaknesses of the methods as far as this research is concerned. A summary of the introduced measures is provided in Table 2.1.

Spatial similarity measures

Most of the available movement similarity analysis techniques address the spatial similarity problem, mainly using the time series similarity measures introduced earlier (section 2.5.2). The state of the art of the spatial similarity measures that have been proposed for movement data can be summarized as follows:

- **LCSS-based methods**
  Vlachos et al. (2002a,b) introduced $LCSS_{\delta,\epsilon}(T_1, T_2)$, an extension of LCSS, to compute the similarity between two trajectories (T_1, $|T_1| = n$ and T_2, $|T_2| = m$) of different durations and granularity in the presence of noise and outliers. The proposed technique allows stretching of the trajectories in time, based on a predefined threshold δ. Hence, the method does not preserve relative speed. The method applies a simple translation in space in both dimensions on the (x, y)-plane, using a predefined matching threshold ϵ. The similarity is computed in $O(\delta(n+m))$ time using dynamic programing. The second paper, Vlachos et al. (2002b), improved the computational cost of this approach by replacing the matching constant ϵ by a sigmoid matching function with a fixed matching width. The proposed distance measure is computed in $O((n+m)\delta)$ time, using dynamic programing.
- **Euclidean distance-based methods**
  Employing the Euclidean distance as a basic measure of similarity is very common in the trajectory similarity literature. Some examples include the following (more examples can be found in section 2.5.4):

  Yanagisawa et al. (2003) proposed a shape-based similarity search technique to support so-called *k-Nearest Neighbor Queries* in trajectory databases. The method considers the *Piecewise Linear Approximation* of trajectories and computes the average Euclidean distance between lines in space. In order to improve the computational cost, Yanagisawa et al. (2003) apply a *2-Dimensional Piecewise Aggregate Approximation* to approximate the trajectories' shapes.

  Later, Lin and Su (2005) introduced the *One Way Distance (OWD)* function based on the closeness of trajectories' shapes in space, ignoring the

time component. The function uses a piecewise linear representation of trajectories. OWD between two trajectories T_1 and T_2 is defined as the integral of the distance from points of T_1 to trajectory T_2 divided by the length of T_1. Consequently, the distance between two trajectories is computed as the average of their one way distances. This method can deal with trajectories of different length (i.e. different number of fixes). Moreover, Lin and Su (2005) presented two algorithms to compute the similarity between trajectories based on OWD for both field and object representations of trajectories in $O(n^2 \log n)$ and $O(n^2)$, respectively.

- **DTW-based methods**
  Vlachos et al. (2004) proposed a method to find similar trajectories under translation, scaling, and rotation. Based on this method, trajectories are first mapped to a space called *Angle/Arc Length space*, which is translation, scale and rotation invariant. In order to map a trajectory to the new space, the technique uses the *turning angle* (with respect to a reference vector) and the *Euclidean length* of the movement vector of trajectory fixes over time. Thereby, the spatial coordinates of a trajectory are transformed into a sequence of Angle/Arc length pairs. In this approach, the DTW distance is applied to compute the distance between trajectories in the Angle/Arc Length space. Since DTW is used, this method does not preserve relative speed of the trajectories and, in fact, is a spatial matching technique for trajectories of the same number of fixes.
- **Edit distance-based methods**
  In a different study, a new representation of trajectories is suggested in order to optimize similarity computation using the *edit distance* (Chen et al., 2004). They proposed a symbolic representation of a trajectory, called *movement pattern strings*, which encodes movement *distance* and *direction* information of a trajectory. In this approach, trajectories are first transformed to their corresponding *movement pattern strings*. The edit distance is then applied to compute the similarity between the generated movement pattern strings. A subsequent approach proposed the *Edit Distance on Real sequence* based on the concept of edit distance (Chen et al., 2005). The method seeks the minimum number of *edit operations* required to transform one trajectory into another. Based on this method, a pair of trajectory elements T_1^i and T_2^j that are located at a distance of less than ϵ of each other are considered to match, where ϵ is the matching threshold. In this approach, trajectories can be of different duration or may have a different number of fixes. Furthermore, trajectories may contain noise, outliers or gaps. In order to make the similarity measure invariant to the spatial scaling and shifting, the trajectories are first normalized using their mean coordinates (Chen et al., 2005).

Spatio-temporal similarity measures

Almost all the available spatio-temporal similarity measures proposed so far employ the notion of the *Euclidean distance*, either explicitly or implicitly. This section aims to present an overview of the most dominant techniques in the literature. Table 2.1 provides a structured summary of the measures.

As a pioneering work in the context of spatio-temporal similarity analysis, Sinha and Mark (2005) employed the average Euclidean distance to measure the similarity between regularly sampled trajectories. Their method considers trajectories as discrete space-time observations over the lifeline of moving objects and hence takes the temporal component of trajectories into account.

Later, van Kreveld and Luo (2007) extended this work and introduced a time-dependent similarity analysis approach to extract the exact or approximately most similar *subtrajectories* in polynomial time. The approach searches for the most similar parts of two trajectories with a particular start time and duration, allowing a specified local time shift from the start time. In this method, the time interval between two consecutive fixes along the trajectories is assumed to be regular (i.e. using a regular sampling rate) and the speed along an interval is assumed to be constant. Furthermore, van Kreveld and Luo (2007) presented four variations of this problem:

(a) fixed duration ($dt = \overline{dt}$) and no time shift ($t_{shift} = 0$) in $O(n)$ time;

(b) non_fixed duration and no time shift in $O(n^2)$ time, where a minimum duration dt_{min} needs to be specified;

(c) fixed duration with time shift in $O(n^3)$ time; and

(d) non_fixed duration and with time shift $O(n^3)$.

In a subsequent attempt, Buchin et al. (2009), extended this work for computing the most similar subtrajectories for the case where the duration is specified (i.e. cases (a) and (c)) and for the case when only a minimum duration is specified (i.e. case (b)). The method considers trajectories as a piecewise monotone function with n break points and hence can deal with trajectories with different sampling rates. Furthermore, an approximation algorithm is proposed to improve the computational complexity for the cases where time shift is allowed (i.e. (c) and (d)) (Buchin et al., 2009).

In a different study, Frentzos et al. (2007) proposed another measure called *DISSIM*, applying an approximation method. Based on their approach, the dissimilarity between two trajectories of the same length (T_1, $|T_1| = n$ and T_2, $|T_2| = n$) during the period of $[t_1, t_n]$ is defined as the sum of the definite integral of the Euclidean distance temporal function between the two trajectories over

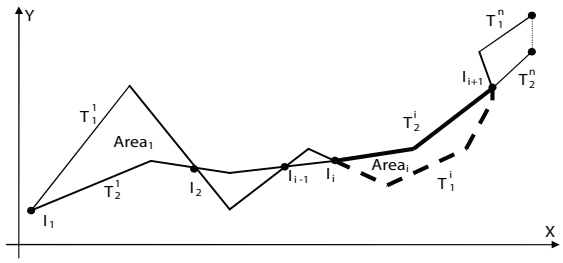


Figure 2.6: Locality In-between Polylines (Pelekis et al., 2007)

different time intervals, as presented in Equation (2.5.10) (Frentzos et al., 2007):

$$DISSIM(T_1, T_2) = \sum_{k=1}^{n-1} \int_{t_k}^{t_{k+1}} D_{T_1,T_2}(t)dt \tag{2.5.10}$$

where t_k are the timestamps of the recorded trajectory fixes, and D_{T_1,T_2} is the Euclidean distance (see Equation (2.5.2)) between the two trajectories.

In order to reduce the computational cost of the integral, an approximation method together with a pruning process were applied on the trajectories. This approach works with trajectories of same duration and same number of fixes. Trajectories with different sampling rates must first be resampled using linear interpolation. Since the technique uses the cumulative Euclidean distance, the computed distance is influenced by the length of the trajectories.

Pelekis et al. (2007) considered a slightly different approach with respect to the previous attempts. They introduced a set of distance operators based on the spatio-temporal coordinates of trajectories as well as movement parameters such as speed and direction. The distance measures rely on the *area* of the polygons formed between the intersection points created by the overlay of two trajectories in the (x, y)-plane (see Figure 2.6). The basic distance operator, called *'Locality In-between Polylines' (LIP)*, is defined as the weighted average of the created areas along the two overlaid trajectories (e.g. T_1 and T_2 in Figure 2.6), as given by Pelekis et al. (2007):

$$\begin{aligned} LIP(T_1, T_2) &= \sum_{polygon_i} Area_i.w_i \\ w_i &= \frac{Length_{T_1}(I_i, I_{i+1}) + Length_{T_2}(I_i, I_{i+1})}{Length_{T_1} + Length_{T_2}} \end{aligned} \tag{2.5.11}$$

where $i = 1, \ldots, n$.

In fact, as it can be seen in Equation (2.5.11), the proposed distance function does not take time into account and only computes the spatial similarity between two trajectories in $O(n + m)$ time. In order to compute the spatio-temporal similarity between trajectories, Pelekis et al. (2007) extended this measure by applying several weight factors to discover concurrent movement of objects that move closely at similar speeds, or direction. The new measures are computed in $O(n \log n)$ time. To some extent, the accuracy of these measures is influenced by the introduced penalty factors that need to be specified by the user. Since the fundamental element of the LIP-based distance operators is the area between intersection points, these measures work better for trajectories which follow the same route. The measures are not appropriate for winding trajectories with a lot of turns. Moreover, since the method relies on the intersection points, it requires an additional search process for finding such points and therefore is not efficient in real-time applications.

In a different study, Trajcevski et al. (2007) introduced the *Rigid Transformation Similarity Distance (RTSD)* employing the notion of Fréchet distance to compute the similarity between trajectories under rigid motion transformations (i.e. translation and rotation) in space and time. Their measure is defined as the minimal value of the maximum so-called *Euclidean time-uniform* distance between two trajectories applying a combination of translation and rotation transformations of one of the trajectories. Here, the Euclidean time-uniform $E_{ud}(t, T_1, T_2)$ is computed as the Euclidean distance between the two fixes on the trajectories T_1 and T_2 at a specific timestamp t. The method uses the discrete representation of trajectories over time. In order to find the similarity between two trajectories of different durations (e.g. $|T_1| \leq |T_2|$), Trajcevski et al. (2007) proposed the *Temporal-Containment Similarity Distance (TCSD)* on the basis of the RTSD measure. That is, the shorter trajectory T_1 is slid along the longer trajectory T_2 and the minimum computed RTSD is considered as TCSD. The computation costs of RTSD and TCSD are rather high and equate to $O(n + m)^2$ and $O(nm(n + m)^2)$, respectively, when approximation is applied.

In a similar attempt, Ding et al. (2008a) proposed a pseudo-metric spatio-temporal similarity measure called *w-constrained Fréchet distance (wDF)*. Their wDF constrains the discrete Fréchet distance by considering only pairs of trajectory fixes whose temporal distance is restricted by a given time window threshold. The size of the temporal matching window (ΔT_w) has an influence on the accuracy of the wDF distance as well as its computational time ($O(\Delta T_w n^2)$). In order to compute this measure more efficiently for longer trajectories, Ding et al. (2008a) propose two approximations for upper/lower bounding on a coarser representation of the trajectories using the concept of the minimum bounding box. The Fréchet distance based measures do not consider the relative speeds of the objects.

2.5.5 Summary

Similarity analysis in movement data is an important and challenging topic. This chapter presented a review of the state of the art of similarity analysis in the relevant domains (i.e. time series analysis, and computational geometry). Furthermore, the main trajectory similarity measures that have been proposed in the literature were introduced and summarized in Table 2.1.

From the state of the art review, it can be concluded that although there are several trajectory similarity search methods that are relatively well developed, most of them are restricted to geometric abstractions of the objects' movement path as a static curve (i.e. a time-ordered sequence of coordinates). And only a few of the available similarity analysis techniques take movement features and dynamic characteristics of movement parameters into account. Therefore, those measures are not appropriate for handling important dimensions of the spatio-temporal nature of trajectory data. Moreover, most methods work better for trajectories that follow similar routes, and the accuracy decreases when subject trajectories are complex or have a winding geometry. Many of the proposed methods have been tested only on simulated or artificial trajectories obtained from different levels of compression of a real trajectory. Therefore, the performance of these methods is not known on real movement datasets.

Table 2.1: Overview of the existing trajectory similarity measures

Similarity measure	Trajectory model	Type of matching	Similarity Dimension	Underling concept		Spacial features	Constraints
				time-series based	geometric based		
Vlachos et al. (2002a,b)	discrete	partial	spatial	LCSS	–	insensitive to noise, outliers, different durations, different sampling rates	non-metric, high complexity
Yanagisawa et al. (2003)	discrete	complete	spatial	Euclidean distance	–		sensitive to noise & outliers & same length trajectories
Vlachos et al. (2004)	discrete	complete	spatial	DTW	–	translation, scale and rotation invariant	sensitive to noise & outliers
Chen et al. (2004)	symbolic	complete	spatial	Edit distance	–	considering distance & direction, robust to noise, gaps and outliers & local time shift, different number of fixes & duration	
Chen et al. (2005) (EDR)	discrete	complete	spatial	Edit distance	–	robust to noise, gaps and outliers & local time shift, different number of fixes & duration	
Lin and Su (2005) (OWD)	continuous & discrete	complete	spatial	Euclidean distance	–	different number of fixes & duration	
Sinha and Mark (2005)	discrete	complete	spatio-temporal	Euclidean distance	–		sensitive to noise & outliers

Table 2.1: Overview of the existing trajectory similarity measures

Similarity measure	Trajectory model	Type of matching	Similarity Dimension	Underling concept		Spacial features	Constraints
				time-series based	geometric based		
van Kreveld and Luo (2007)	discrete	complete & partial	spatio-temporal	Euclidean distance	–	subtrajectory similarity, local time shift	specifying minimum duration, sensitive to noise & outliers
Frentzos et al. (2007) (DISSIM)	discrete with linear interpolation	complete	spatio-temporal	Euclidean distance	–		same duration, sensitive to noise & outliers
Pelekis et al. (2007) (LIP-based distances)	discrete	complete	spatio-temporal	area based	–	considering speed & direction	sensitive to the geometric shape & finding intersection points
Trajcevski et al. (2007) (RTSD/ TCSD)	discrete with linear interpolation	complete & partial	spatio-temporal	Euclidean distance	Fréchet distance		same duration & ignores speed & high complexity
Ding et al. (2008a) (wDF)	discrete	complete	spatio-temporal	–	Fréchet distance	allow local time shift	ignores speed & non-metric & sensitive to noise & outliers
Buchin et al. (2009)	continuous	complete & partial	spatio-temporal	Euclidean distance	–	subtrajectory similarity & allow local time shift	specifying the minimum duration & sensitive to noise & outliers

Chapter 3

Methods and Results

This chapter summarizes the overall research process of this thesis, introduced in chapter 1 (see section 1.2.2). It introduces the four papers corresponding to the three stages of the research process. For every stage, and for the respective paper, the objectives, methods, the main contributions, and the key findings are highlighted to provide a basis for the subsequent discussion chapter. This chapter, however, does not provide a substitute for reading the full papers. Hence, to obtain a comprehensive insight into the methods, problems and achievements of this work, studying the full papers provided in Part II is recommended.

3.1 Theoretical Framework Development

The first stage of this research aims at providing an overall understanding of movement itself and the patterns of movement, identifying the fundamental elements of movement patterns. The substance of this stage is presented in detail in Research Paper 1 (see Part II, Research Paper 1, page 81).

➤ **Research Paper 1:**

> Dodge, S., Weibel, R. and Lautenschütz, A.K. (2008). Towards a Taxonomy of Movement Patterns. *Journal of Information Visualization*, Vol. 7, pp. 240 – 252.

3.1.1 Objectives

The main objectives of this stage are captured in Objective 1 and Objective 2 of this thesis:

Objective 1: This research shall develop a conceptual framework, encapsulating essential elements that characterize the movement behavior of objects.

Objective 2: This research shall establish a comprehensive classification of movement patterns. The identified movement patterns shall be defined employing the elements of the conceptual framework.

Accordingly, Research Paper 1 aims at providing a framework identifying the fundamental elements of movement. Furthermore, it responds to the need for a catalog of movement patterns to facilitate the development of movement data mining techniques.

3.1.2 Methods and Results

We approached the objectives of this stage with a twofold process:

(a) *bottom-up approach*, by decomposing movement into its constituting elements. The aim here was to develop a conceptual framework of the movement of dynamic objects that could be used to build definitions of individual movement patterns. The developed framework consists of the following elements:

- Movement parameters are measurable quantities (i.e. primitives) and their derivatives such as position, time, speed, distance, direction. Movement parameters are categorized into *spatial*, *temporal*, *spatio-temporal* dimensions (Table 3.1).
- Number of moving objects involved (i.e. an individual object, a group of objects with a functional relationship, a cohort of objects with a common characteristic)
- Path type (i.e. continuous and discontinuous paths)
- Influencing factor such as intrinsic properties, spatial constraint, environment, and influence of other agents.
- Scale and granularity (i.e. spatial and temporal scales)

(b) *top-down approach*, with a survey of the research conducted prior to this study. The aim was to categorize patterns of movement proposed by other researchers and to discover commonalities and differences in the terminology and pattern types. The survey encompasses the available pertinent literature in geographic knowledge discovery, data mining, and visual analytics on movement data, as well as additional application specific references.

The result of this step is the classification of movement patterns presented in Figure 3.1. At the top level of this classification, two classes of *generic* and *behavioral* patterns, respectively, were identified. *Generic patterns* refer to the low-level patterns that are usually insufficient to explain specific movement behavior of a particular object. Hence, the proposed classification of movement patterns distinguishes the higher-level *behavioral patterns* for the domain-specific applications. The *generic patterns* are further categorized into two classes of *primitive patterns* and *compound patterns*. Furthermore, the classification distinguishes three dimensions for generic movement patterns: *spatial*, *temporal*, and *spatio-temporal* patterns.

Table 3.1: Movement parameters

Dimensions	Parameters		
	Primitives	***Primary derivatives***	***Secondary derivatives***
Spatial	Position (x, y)	Distance $f(posn)$ Direction $f(posn)$ Spatial extent $f(posn)$	Spatial distribution $f(distance)$ Change of direction $f(direction)$ Sinuosity $f(distance)$
Temporal	Instance (t) Interval (t)	Duration $f(t)$ Travel time $f(t)$	Temporal distribution Change of duration $f(duration)$
Spatio-temporal	_	Speed $f(x, y, t)$ Velocity $f(x, y, t)$	Acceleration $f(speed)$ Approaching rate

In order to facilitate the future development and consolidation of the proposed classification and pattern definitions into a complete taxonomy of movement patterns, a wiki[1] was set up. The wiki hosts more background information and more detailed definitions of movement patterns, and serves as an open discussion platform for the community.

3.1.3 Main Findings

- The occurrence of patterns can be influenced by various factors that impact or constrain the movement of objects. Examples of such factors include:
  - *intrinsic physical and behavioral properties of the moving object* (e.g. saccadic movement of mass-less eyes in contrast to the very slow and smooth movement of a massive body such as an elephant)
  - *spatial constraints* such as topography, road networks, and natural barriers (e.g. the occurrence of a moving cluster along a road).
  - *environment* (e.g. the evolution of hurricanes is influenced by the air and sea temperatures and the air pressure (Elsner and Kara, 1999))
  - *influences of other agents* can cause the emergence of certain patterns in domain-specific applications (e.g. attraction and competition among animals lead to particular behavioral patterns such as *courtship* or *fighting*, respectively)
- Movement patterns can be classified into two main categories ; (a) *generic patterns* and (b) *behavioral patterns* (Figure 3.1). The generic patterns represent the building blocks used to form the behavioral patterns.
  - Generic patterns can be found in any form of behavior that builds on movement of objects such as moving clusters, concurrence, and

[1] http://movementpatterns.pbworks.com/

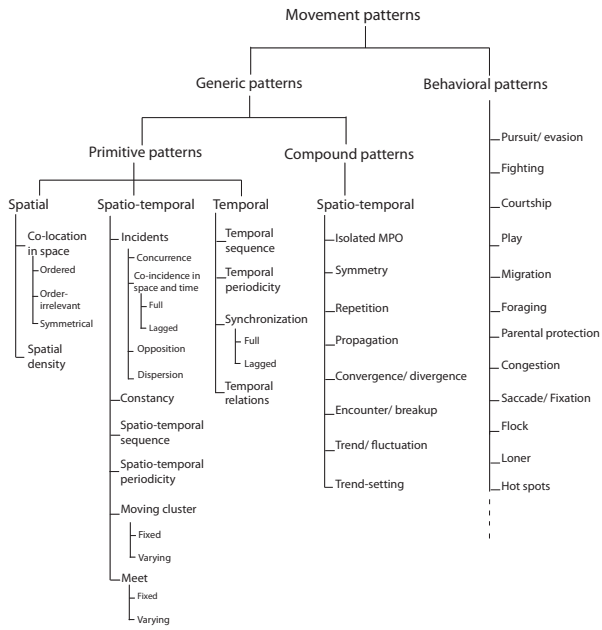


Figure 3.1: Classification of movement patterns

repetition patterns.

  - Behavioral patterns correspond to a particular behavior typical of a particular object such as pursuit, evasion, or foraging behaviors of animals.

- Generic movement patterns exhibit different degrees of complexity:
  - *Primitive patterns*, which are the basic patterns that are formed as a result of the similarity in the variations of a single movement parameter, such as *concurrency* in speed or direction.
  - *Compound patterns*, which are formed in the relations between multiple objects as a composition of several primitive patterns. For instance, the *convergence* of the movements of a set of objects to the same location, where the original movement direction of the involved object remains unchanged (Gudmundsson et al., 2007).
- *Primitive* patterns can occur purely in the spatial domain or in the temporal domain, or they can be mixed spatio-temporal. *Behavioral* patterns invariably involve both spatial and temporal dimensions.

3.1.4 Contributions

This work contributes to the development of a catalog for knowledge discovery and data mining algorithms by developing:

(a) a conceptual framework encapsulating the fundamental elements of the movement behavior of different objects; and

(b) a comprehensive classification and review of movement patterns (illustrated in Figure 3.1).

3.2 Movement Feature Extraction

The second stage of this research focuses on the development of a *trajectory decomposition* method. For the sake of consistency, from this point on the decomposition method is referred as *trajectory segmentation* in this thesis, just as in Stage III, Research Paper 3. The aim of segmentation is to decrease the complexity of movement data in preparation of subsequent analyses. The method allows extraction of the local movement features that are essential for modeling, simulating, and analyzing movement as well as discovering movement patterns. Research Paper 2 presents the substance of this stage extensively (see Part II, Research Paper 2, page 97).

➤ Research Paper 2:

> Dodge, S., Weibel, R. and Forootan, E. (2009). Revealing the physics of movement: Comparing the similarity of movement characteristics of different types of moving objects. *Computers, Environment and Urban Systems*, Volume 33, Issue 6, November 2009, pages 419 – 434.

3.2.1 Objectives

This stage primarily pursues the Objectives 3 and $5_{(a)}$ of this thesis (cf. chapter 1):

Objective 3: This research shall identify, and formalize important features characterizing the movement of objects from the parameters of movement. Quantitative methods shall be developed to extract such features from raw trajectory data, with the aim of transforming trajectories into a simpler structure, while still conveying the important movement features.

Objective 5: The applicability of the developed methods shall be evaluated in knowledge discovery tasks such as (a) *trajectory classification*, [...] in real movement datasets.

In order to investigate the applicability of the proposed feature extraction and segmentation methods, the Objective $5_{(a)}$ is expanded as follows:

Objective 5$_{(a)}$: *This research shall develop a trajectory classification technique using movement feature extraction. The developed classification technique shall enable classifying movement data generated by unknown moving objects and assigning them to the known types of moving objects.*

3.2.2 Methods and Results

In order to achieve the objectives of this stage, Research Paper 2 proposes a three-step methodology as illustrated in Figure 3.2. The key element of this methodology is the evolution function of the movement parameters (i.e. speed, acceleration, direction etc.) over time, called *movement parameter profile* (see Figure 3.3).

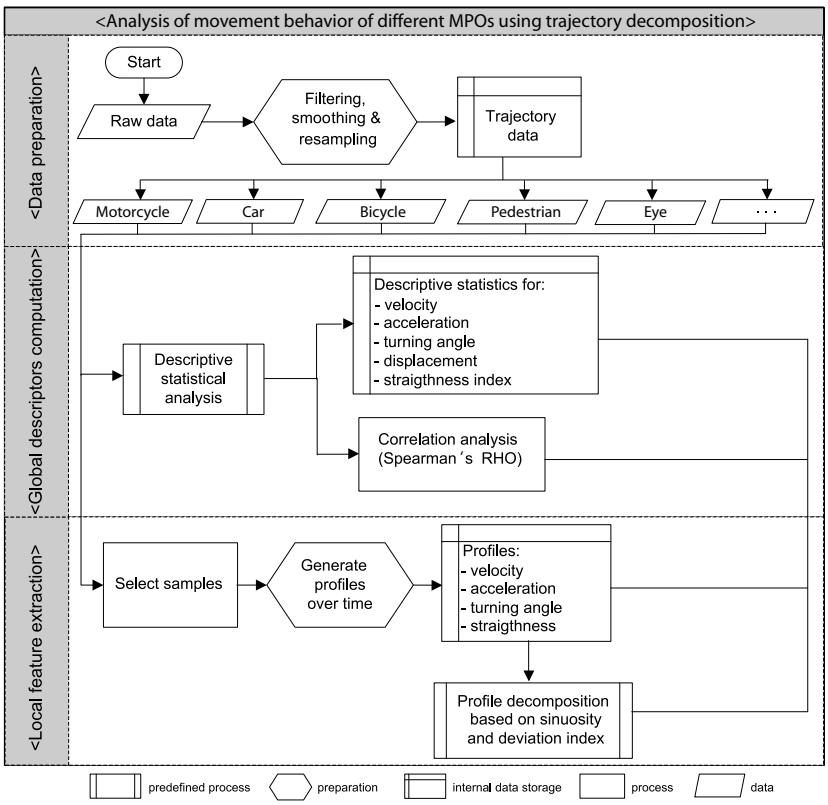


Figure 3.2: Movement feature extraction methodology

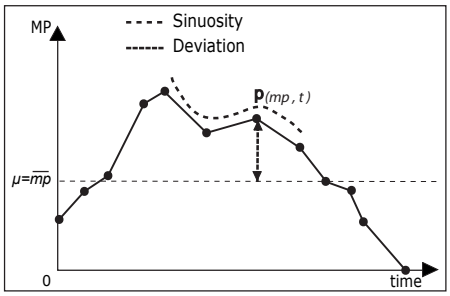


Figure 3.3: Movement parameter profile: The evolution function of a movement parameter over time

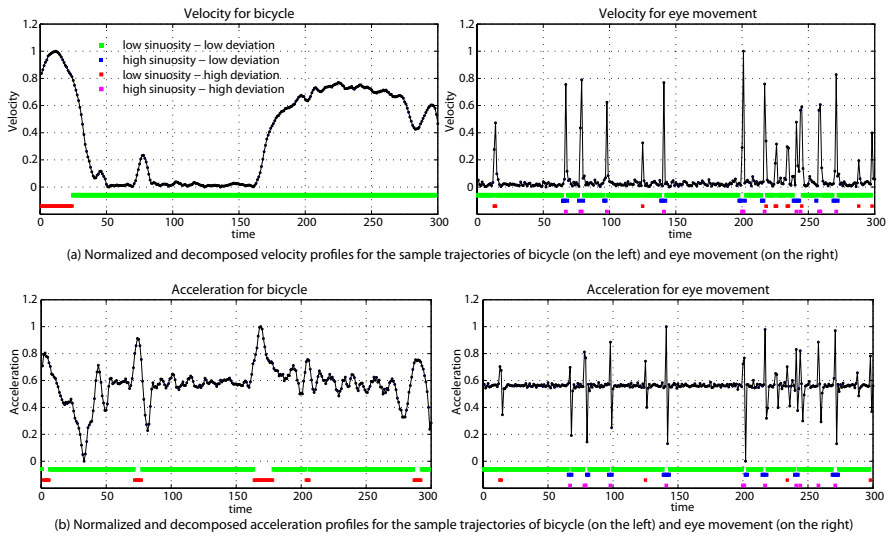
The developed methodology consists of the following processes (see Part II, Research Paper 2, Figure 3.2):

(1) *Trajectory data preparation*: consists of data cleaning and preprocessing steps in order to remove effects of noise and positioning errors.

(2) *Computation of global descriptors*: involves the extraction of global movement properties of objects (i.e. computation of the movement parameters and their descriptive statistics over the entire trajectory). In order to detect possible interrelationships, correlation analysis between the movement parameters is recommended.

(3) *Local feature extraction*: a trajectory segmentation approach is proposed to partition the movement parameter profiles into sections of homogeneous movement features. Important movement features are identified as the *frequency* and *amplitude* of variations of movement parameters. The frequency of variations is quantified by the *sinuosity* of the MP profile, while the amplitude of variations is measured by the *deviation* of the MP profile from the median (or mean) line (Figure 3.3). According to the magnitudes of sinuosity and deviation, each point of the MP profile is labeled with a certain sinuosity and deviation regime (later called *movement parameter class* in Research Paper 3). Here, four main MP *regimes* (or *MP classes*) representative of the local movement parameter features are distinguished, as seen in Figure 3.4 (the sequence of colored segments at the bottom of each graph):

- low sinuosity – low deviation
- low sinuosity – high deviation
- high sinuosity – low deviation
- high sinuosity – high deviation



(a) Normalized and decomposed velocity profiles for the sample trajectories of bicycle (on the left) and eye movement (on the right)

(b) Normalized and decomposed acceleration profiles for the sample trajectories of bicycle (on the left) and eye movement (on the right)

Figure 3.4: Different MP profiles exhibit different characteristics: Speed and acceleration profiles of the bicycle and eye movements exhibit different *amplitude* and *frequency* variations.

In response to Objective 5.a, Research Paper 2 suggests a trajectory classification strategy exploiting the results of the proposed feature extraction and segmentation methods. The classification strategy consists of *feature selection* and *dimension reduction* procedures using *Principal Component Analysis (PCA)* (Jolliffe, 1986), followed by *supervised classification* using *Support Vector Machines (SVM)* (Cortes and Vapnik, 1995). In a set of experiments it is shown how the developed methods can be used to label trajectories of unknown objects by similarity to previously learned moving objects. As an example, the classification strategy is applied on movement data from the transportation domain (e.g. pedestrians, motorcycles, bicycles, and cars) in order to extract the *mode of transport* of unknown trajectories.

3.2.3 Main Findings

- The proposed feature extraction and classification methods can be successfully applied to detect the mode of transport from unknown trajectories of people using different transportation means.
- The experiments suggest that the movement characteristics of a mass-less process such as eye movement are very different from full-body movement of humans and vehicles (e.g. as speed profiles in Figure 3.4 show). Hence,

such virtual movement data can not be used as a proxy of the movement of massive objects that produce a continuous path. In contrast, eye movement data could potentially be considered as a proxy of the movement of objects with a stop-and-go behavior (e.g. bees, butterflies).

3.2.4 Contributions

This study can contribute to knowledge discovery, modeling, simulation, and analyzing of movement data in the following aspects:

(a) To decrease the complexity of movement data by segmenting trajectories into sections of homogeneous movement characteristics and hence to facilitate knowledge discovery of massive movement data. The segmented trajectories convey information about the *frequency* and *amplitude* of variations of movement parameters, which are recognized as the important movement features in this thesis.

(b) To assess the similarity of the movement characteristics of proxy and simulated data to the movement of real objects.

(c) To automatically identify trajectories of unknown objects by applying the available knowledge about the movement of similar known objects.

3.3 Movement Similarity Assessment

The final stage of this research intends, firstly, to develop a methodology to assess the similarity of the movements of objects, and secondly, to evaluate the applicability of the developed methods in trajectory clustering and movement pattern discovery. In this study, the movements of objects are considered similar, if the objects exhibit *similar variations* in their movement parameters. This thesis suggests two *movement similarity assessment* approaches using the movement parameter profiles of the second stage. These approaches are presented in Research Paper 3 and Research Paper 4, respectively. This section summarizes the methods and the most relevant results of this stage, organized according to the corresponding papers.

3.3.1 Objectives

The final stage and its corresponding papers respond to Objective 4 and Objective $5_{(b,c)}$ of this thesis:

Objective 4: This research shall quantify similarity between movement behaviors of two individuals in space and time. The similarity measure shall consider dynamic properties of the movement parameters of objects. Accordingly, a similarity assessment approach shall be developed to investigate the similarity between both movement characteristics of several individuals and their patterns of movement.

Objective 5: The applicability of the developed methods shall be evaluated in [...], (b) *trajectory clustering*, and (c) *movement pattern discovery* in real movement datasets.

In order to investigate the applicability of the developed similarity assessment methods, Research Paper 3 (see Part II, page 115) pursues the Objective $5_{(b)}$ and Research Paper 4 (see Part II, page 143) addresses the Objective $5_{(c)}$:

Objective 5(b): *This research shall develop a trajectory clustering approach using the segmentation and similarity assessment methods. The developed clustering techniques shall enable grouping movement data according to the similarity in the movement characteristics of objects.*

Objective 5(c): *This research shall develop a movement pattern extraction technique based on the similarity analysis. The method shall enable discovery of concurrence and coincidence movement patterns.*

3.3.2 Similarity Assessment Using Trajectory Segmentation

The first similarity assessment method, introduced in Research Paper 3, embodies the main contribution of this research and relies on the segmentation technique developed in the previous stage (see section 3.2.2).

➤ Research Paper 3:

Dodge, S., Laube, P., and Weibel, R. (in revision, 2011). Movement Similarity Assessment Using Symbolic Representation of Trajectories. *International Journal of Geographic Information Science.*

Methods and Results

In response to the objectives of this stage, Research Paper 3 introduces a novel *movement similarity assessment* approach. The proposed method consists of two processes (see Figure 3.5 and Research Paper 3):

(i) *trajectory segmentation*: an extended version of the *trajectory segmentation* process (introduced in Research Paper 2 and section 3.2.2) is suggested to transform the trajectories into their corresponding *Movement Parameter Class (MPC)* sequence. An *MPC sequence* is a symbolic representation of a movement parameter profile derived from the trajectory of a dynamic object. The MPCs indicate the types of *amplitude* and *frequency* variations of a movement parameters (see Figure 3.6 and Research Paper 3), just as *MP regimes* in section 3.2.2. In addition to the four MP classes introduced in the previous stage, the *amplitude levels* are further classified into *positive* and *negative* values. Moreover, a class of *zero values* is also considered.

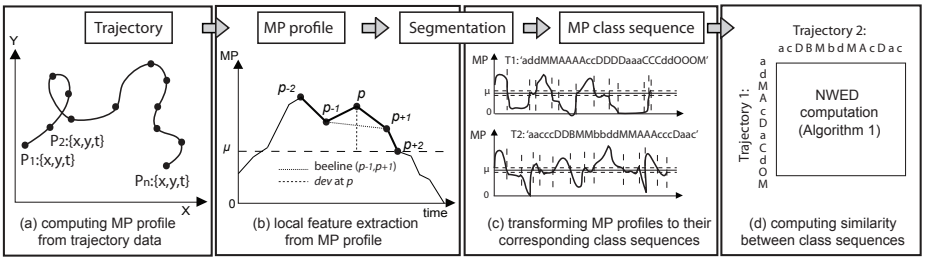
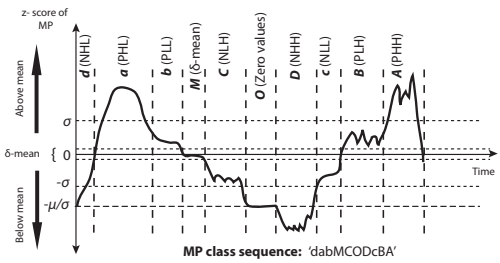


Figure 3.5: Process of trajectory segmentation and similarity computation

Hence, this stage recognizes total of nine movement parameter classes as shown in Figure 3.6.

(ii) *similarity computation*: a variation of the *edit distance* called *Normalized Weighted Edit Distance (NWED)* is defined as the *similarity measure* to quantify the similarity between two movement parameter class sequences. NWED is formalized as the minimum conversion cost between two sequences using weighted edit operations (i.e. insertion, deletion, and substitution). In this study, the substitution costs between the different movement parameter classes are identified according to the degree of similarity between the different MPCs.



MPC	Deviation	Sinuosity
PHH	positive high	high
PHL	positive high	low
PLH	positive low	high
PLL	positive low	low
M	δ-mean	—
NLH	negative low	high
NLL	negative low	low
NHH	negative high	high
NHL	negative high	low

Figure 3.6: MP classes and MP class sequence

In response to Objective $5_{(b)}$, Research Paper 3 presents two *trajectory clustering* strategies that are based on the developed trajectory segmentation and similarity computation approaches:

(1) MPC diversity based clustering: relying on the descriptive statistics computed on the MP classes generated from the segmentation process.

(2) MPC sequence based clustering: applying the NWED distance between the

MPC sequences.

In order to investigate the applicability of the developed techniques in clustering movement data, a set of comparative case studies on the following datasets were conducted:

- North Atlantic Hurricane trajectories, obtained from NOAA's Coastal Services Center [2].
- A GPS tracking dataset captured from the movements of couriers in Central London; obtained from the eCourier company [3] (located in London, UK).

Moreover, the NWED similarity measure suggested in this thesis was evaluated in comparison to another *edit distance-based* similarity analysis technique proposed by Chen et al. (2004) (introduced in section 2.5.4, page 30).

Main Findings

- The experimental results suggest that the proposed NWED similarity measure, together with the MP class sequence representation of trajectories, is more effective than the method introduced by Chen et al. (2004) to study the dynamic behaviors of moving objects such as hurricanes. That is, in contrast to the spatial similarity measures (e.g. EDM), NWED considers dynamic movement characteristics of objects rather than their movement path.
- The MPC diversity based clustering strategy works best for clustering trajectories with heterogeneous movement characteristics such as vehicle movements in a street network with a diverse traffic pattern.
- The MPC sequence based clustering strategy is better suited for clustering movement data with prominent sequential trends, such as hurricanes.

Contributions

This study contributes to knowledge discovery in movement data with the development of a novel *movement similarity assessment and trajectory* clustering methodology relying on the *variations* of movement parameters. The methodology uses the movement parameter sequence representation of trajectories, which embodies the important movement features of the objects. Moreover, this study introduces a new similarity measure as an extension of the edit distance, *named normalized weighted edit distance (NWED)*, that can be used to compute the similarity between the MPC sequence representation of trajectories. NWED is a generic measure and it can potentially be applied in time series similarity analysis techniques.

[2] http://csc-s-maps-q.csc.noaa.gov/hurricanes/
[3] http://www.ecourier.co.uk

3.3.3 Similarity Assessment in Movement Parameter Space

Research Paper 4 (see Part II, page 143) aims to introduce the second similarity analysis technique. The method is simpler compared to the first approach, however, it involves more dimensions in the similarity assessment using a *movement parameter space*. The main motivation for this work was: first, to assess trajectory similarity based on a combination of a set of movement parameters (including the spatial dimension); and second, to evaluate the applicability of the developed method in *movement pattern discovery* (cf. Objective $5_{(c)}$), specifically, in discovering *concurrency* and *coincidence* patterns (i.e. incidents of a set of objects showing the same values of movement parameters at a certain instant or duration of time (Laube, 2005)).

➤ Research Paper 4:

> Dodge, S., Weibel, R., and Laube, P. (2011). Trajectory Similarity Analysis in Movement Parameter Space. *GISRUK 2011*, April 27-29, 2011, University of Portsmouth, UK, pages 270 – 279, Short paper.

Methods and Results

The proposed method relies on a multidimensional representation of trajectories using a set of movement parameter profiles. The method consists of the following steps (illustrated in Figure 3.7):

(a) Generation of a *multidimensional movement parameter space* from a desired set of movement parameters. For instance, Figure 3.7.a shows a four dimensional MP space computed from speed, acceleration, and azimuth (direction) profiles over time.

(b) Similarity computation using the average Euclidean distance (see Equation (2.5.2), page 23). Figure 3.7.b illustrates the similarity computation between two trajectories in a three dimensional MP space.

Furthermore, Research Paper 4 applies the proposed method in order to discover *concurrence* and *coincidence* patterns from the North Atlantic Hurricanes dataset that was already used in the previous paper.

Main Findings

- The experimental results suggest that similarity analysis of the trajectories in the movement parameter space can help to discover similar movement behaviors of dynamic objects such as hurricanes.
- Specifically, the method facilitates the extraction of movement patterns originating from a similarity in the movement characteristics of objects, such as *concurrence* and *coincidence* patterns.

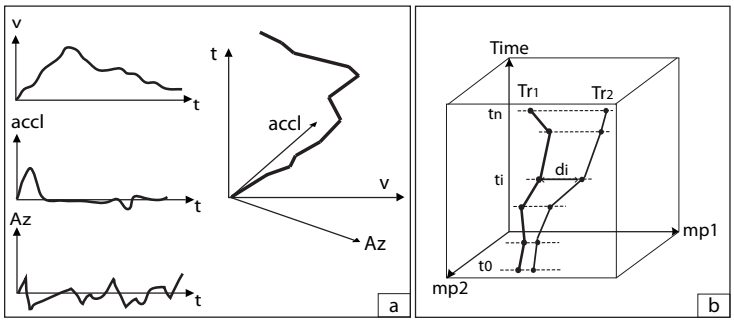


Figure 3.7: Trajectory similarity assessment in movement parameter space

Contributions

This study contributes to movement research with the development of a similarity assessment technique based on a multidimensional representation of trajectories using a set of movement parameters. The specifications of the proposed method are:

- According to the setting of the movement parameter space, the method incorporates both *spatial* and *temporal* dimensions of movement to assess the similarity of the trajectories of dynamic objects.
- The movement parameter space bears more dimensions, compared to the previous approach (i.e. similarity assessment based on the segmentation of a single movement parameter, cf. section 3.3.2).
- The method is based on the Euclidean distance. Hence, it requires less computational time (i.e. $O(n)$, where n is the number of observations along the trajectories) and is simpler than the edit distance in the previous approach.
- The method is very simple to implement.
- No control parameter is required.
- The method is limited to the trajectories of the same number of fixes in contrast to the edit distance that can handle trajectories of different lengths.

Chapter 4

Discussion

This chapter is presented in two contexts: First, the research questions identified in chapter 1 are revisited and discussed in the light of accepted knowledge and the obtained results. Second, a general discussion of the methods and outcomes of the thesis is provided in a broader perspective.

4.1 Revisiting the Research Questions

4.1.1 The fundamental Elements of Movement Patterns

Research Question 1:

❖ *What are the constituting components of movement, essential for defining movement patterns?*

In response to the first research question, the first paper (see Part II, Research Paper 1, page 81) suggests a *conceptual framework for movement* (introduced in section 3.1.2, page 38). The elements of this framework represent the primitives necessary for defining movement patterns. Furthermore, Research Paper 1 presents the relations between these elements with respect to the generic movement.

The developed framework identifies *movement parameters* as the fundamental components of movement patterns. Often, one or more movement parameters are required in the definition and formalization of movement patterns. For instance, Laube et al. (2004) defined the *flock pattern* as movements of entities, heading in the same *direction* at the same *time* while staying in close spatial proximity (i.e. proximate *position*) of each other. That is, three movement parameters (i.e. position, time, and direction) are involved in the formalization of *flock pattern*.

Laube et al. (2007) proposed four lifeline context operators in order to quantify movement parameters and the corresponding movement patterns. These operators include: *instantaneous* (i.e. at an infinitesimal instant in time), *interval* (i.e. in a moving temporal window), *episodal* (i.e. in a partition of the trajectory) and *total* (i.e. whole trajectory). In this context, the framework proposed in this thesis recognizes two groups of *primitive* and *derivative* (i.e *primary* and *secondary* derivatives) movement parameters. The *primary* and *secondary* derivatives such

as speed, acceleration, change of direction, and path sinuosity, can be quantified using *interval*, *episodal*, and *global* lifeline context operators of Laube et al. (2007). In contrast, *primitive* movement parameters such as position and time correspond to the *instantaneous* operators and can be directly measured using positioning devices and sensors. Nowadays, it is also possible to measure some of the *derivatives*, such as speed, acceleration, and direction at an instance of time using sensor technologies.

Contrary to this work, the conceptual model of trajectories proposed by Spaccapietra et al. (2008) does not include movement parameters. In their framework movement is modeled as a sequence of *move* and *stop* episodes and consists of two facets: *geometric facet* (i.e. spatio-temporal recording of the positions), and *semantic facet* (i.e. application-oriented and semantic characteristics of trajectories). Their conceptualization allows only the formalization of patterns using three *'begin of move'* (i.e from), *'end of move'* (i.e. to), and *'stop'* descriptors, as well as the *semantic* attributes attached to the stops and moves. For instance, their model is restricted to the identification of movement patterns such as *meet* or *convergence*, where entities share a common location (origin or destination). In contrast, the framework developed in this thesis takes not only position but other movement parameters into account and hence is more effective in defining rather complex patterns such as *concurrence* and *moving clusters* (e.g. flock). On the other hand, the *semantic facet* in their model facilitates the detection of *events* and *activities* along the trajectories. Moreover, such information may become useful in the identification of the *behavioral* movement patterns. The *influencing factors* considered in the proposed framework of this thesis can be related to the *semantic facet* of trajectories.

The definition and discovery of movement patterns are highly influenced by the *spatial scale* and *temporal granularity* (Wood and Galton, 2009). Laube (2005) distinguishes two types of granularity: *sampling granularity* and *analysis granularity*. Sampling granularity controls the trade-off between the amount of useful information and the autocorrelation in data. That is, undersampling causes information loss whereas oversampling may cause redundancy. The sampling granularity may even affect the accuracy of the computed movement parameters such as path sinuosity. On the other hand, at different levels of *analysis granularity* a pattern may be perceived differently. For instance, Figure 4.1.a illustrates the trajectories of a set of objects that form a *meet* pattern (i.e. a stationary cluster) over the period $[t_s, t_e]$, when the temporal granularity is considered as one day. However, from the same trajectories various patterns can be observed at a finer granularity (e.g. one hour) such as the *concurrence* pattern during $[t'_1, t'_2]$, and the *moving cluster* during $[t'_2, t'_3]$ (Figure 4.1.b). Therefore, as discussed in Laube (2005, p. 76), the interpretation of discovered patterns may vary depending on the chosen analysis granularity. Hence, this thesis includes *spatial scale and tem-*

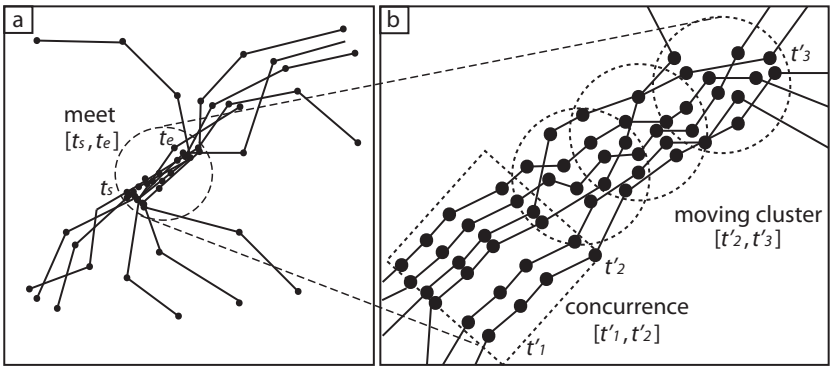


Figure 4.1: The effect of scale and granularity on movement patterns

poral granularity as one of the important facets of movement patterns that need to be considered in the identification and extraction of patterns in movement data.

4.1.2 Classes of Movement Patterns

Research Question 2:

❖ *How can we define a classification for different types of movement patterns?*

The rapid progress of research in movement pattern analysis triggered "a growing interest in categorizing of movement patterns in the GIScience community" (Laube, 2009, p. 53). The state of the art review carried out in this field (presented in Research Paper 1) suggested that there is little agreement on the relevant types of movement patterns and only few, isolated definitions of the movement patterns exist. Hence, this study highlighted the need for a comprehensive taxonomy of movement patterns. The development of a taxonomy of movement patterns has several advantages: *a*) the design and development of effective and efficient knowledge discovery and data mining techniques; *b*) the improvement of the interoperability of movement analysis methods among different disciplines; and *c*) an improved agreement and common understanding of different types of movement patterns, hence avoiding duplications of efforts.

As one of the early attempts, Laube et al. (2005) introduced three groups of movement patterns: (1) *Patterns over time* (i.e. movement parameters of an individual are compared over time), (2) *patterns over objects* (i.e. movement parameters of several objects are compared at an instance of time), and (3) *combined patterns over time and objects* (i.e. the interrelations of the movements of many objects at several time steps). Their study, however, did not aim to classify

movement patterns and hence only a list of relative movement patterns that belonged to each category were introduced and formalized. Hence, the relationships between different patterns of each category and dimensions of the patterns (i.e. spatial, temporal, spatio-temporal) were not investigated.

With a different approach, Andrienko and Andrienko (2007) recognized three categories of patterns as the top-level classes of patterns: *similarities*, *differences* and *arrangement*. Similar to the previous study, their classification provided a list of patterns for each category without introducing the interrelations between patterns of each category.

In this thesis, Research Paper 1 aimed at developing a rather comprehensive classification of movement patterns (see Figure 3.1). The top-level categories of this classification (i.e. *generic* versus *behavioral* patterns) discriminate the patterns that may be formed in the movement of any kind of dynamic object from the ones that are specific to a certain type of object. The classification proposed in this thesis exploits a hierarchy (see section 3.1.3 and Figure 3.1) only for the branch of *generic patterns*. Since the scope of the behavioral patterns is very broad and the patterns vary in different applications, this study does not inquire into the classification of behavioral patterns. The classification of such patterns requires domain-specific expert knowledge.

For several reasons, the top-level classes introduced by Andrienko and Andrienko (2007) were not used in the proposed classification: First, *similarity* and *difference* are merely antonyms and should not be used concurrently to avoid confusion. Second, in terms of similarity, we argue that similarity is never a pattern by itself but should be seen as a concept assisting the comparison between movement properties or patterns, as it was shown in this thesis. Third, many of the movement patterns (e.g., periodicity, repetition, propagation) originate from similarities between movement parameters. Finally, similarities can also be seen in the arrangement patterns proposed by Andrienko and Andrienko (2007). Moreover, both studies (Laube et al., 2005; Andrienko and Andrienko, 2007) identified a movement pattern named *change*. However, the classification proposed in this thesis does not consider this type of pattern. The reason being that *change* is a general term and cannot be considered as a specific pattern. For instance, movement itself is defined as a *change* in the spatial location of an object in time (cf. chapter 2).

Following our classification, Wood and Galton (2009) introduced a classification of collectives. Their study focused more on the relationship between the movement of collective phenomena as a whole and the movements of its individual members (e.g. an aggregated movement of a crowd, as opposed to the random movements of crowd members). In their classification, two levels of granularities were considered in the movements of collectives: *processes* (i.e an open-ended activity or a sequence of events) and *events* (i.e. initiation, termination, or a ho-

mogeneous part of a process). In this context, the classification proposed in this thesis recognizes different levels of complexity for movement patterns. That is, in addition to the high-level *behavioral* patterns, two classes of *primitive* versus *compound* patterns were identified at the level of *generic* patterns. The *process* can be related to the *behavioral* patterns introduced in our classification (e.g. a herd of cows *foraging* on a meadow) whereas some *events* may possibly be represented by the *generic* patterns (e.g. *concurrent* movements of cows foraging along a river on a meadow). In fact, the classification of collectives by Wood and Galton (2009) groups the behavior of a *specific* type of collectives. In contrast, the classification proposed in this thesis takes a broader perspective and considers the *generic patterns* that can be formed in movements of a single individual or collectives of any types. However, contrary to the classification of collectives by Wood and Galton (2009) the number of objects and the relationships between members of collectives have not been taken into account in the proposed classification of movement patterns.

Furthermore, Wood and Galton (2009) investigated the above mentioned classifications including ours, focusing on the level of *collectives*. Their study highlights a need for a link between the existing classifications of movement patterns and collectives.

❖ *To what extent are movement patterns generic to various types of moving objects?*

The review of the movement pattern analysis research carried out in Research Paper 1 suggests that there are a lot of commonalities in the definition of different types of movement patterns identified in the literature. In this regard, Laube (2005) investigated such commonalities specifically for certain patterns such as *constancy*, *concurrence*, *change*, and *trend-setter* in various fields such as biology, ecology, soccer scene analysis, and political science. Similarly, his studies suggested that such patterns can be identified in the movement of individuals from all the application fields studied. On the other hand, owing to the diversity in the terminology of movement patterns in different application domains, it is observed that the very same phenomenon is named differently in various contexts and there is little agreement on the relevant types of movement patterns in the literature. For example, a group of individuals moving close to each other is named *flock* in the ecology of some animals such as birds, sheep, and goats. However, different terms are observed in other domains for the same pattern. Examples include: *convoy* of cars, *school* of fish, *swarm* of bees (Laube, 2005). This suggests that there are patterns that can be generic to many moving objects. On the other hand, some patterns are only specific to a certain type of objects (e.g. foraging pattern of animals). Therefore, movement patterns can be divided into two categories of generic and domain-specific patterns. Research Paper 1 particularly seeks to achieve such a classification.

This thesis proposes the classification of movement patterns following the study by Laube (2005); however, in a broader domain. That is, our classification is based on a review of the literature from various application domains. In this study, a variety of patterns in movements of humans (e.g. meeting), vehicles (e.g. congestion), animals (e.g. leadership), and eye-tracking (e.g. saccades and fixations) were investigated and the corresponding *generic* patterns were sought, independent of the types of objects involved. Hence, the proposed classification suggests the distinction between *generic* and *behavioral* patterns. Accordingly, the generic patterns highlight the extent of the patterns that are general to any type of moving objects. In contrast, behavioral patterns refer to the domain-specific patterns that are either composed of several generic patterns, or are equal to single generic patterns; however, distinguished with specialized terms.

4.1.3 Trajectory Segmentation

Research Question 3:

❖ *How can we reduce the complexity of a trajectory, while preserving its important movement features for knowledge discovery applications?*

Simplification of objects, specifically linear features, is essential in many applications (e.g. map generalization, computational geometry, time series analysis etc.) due to the constraints of computational methods and/or human perception. Time series analysis and map generalization studies have introduced various *local measures* for characterization and simplification of linear features. For instance, *amplitude* and *frequency* parameters have widely been applied in time series analysis techniques such as *wavelets* and *spectral* approaches in order to analyze the localized variations of power within time series (Torrence and Compo, 1998). Furthermore, several studies have suggested using the *amplitude* level to characterize time series (Aßfalg et al., 2006, 2008) for similarity analysis. Likewise, in map generalization various line simplification methods have used local measures such as *curvature* (e.g. *sinuosity*), *amplitude* (e.g. *maxima*, *minima*) in order to characterize *salient points* of linear features (Plazanet et al., 1995; Dutton, 1999; Nakos and Mitropoulos, 2005). On the other hand, map generalization research suggests that *segmenting objects* can improve the line simplification methods, where each homogeneous part of objects can be expressed by its local characters and is processed by adequate algorithms (Plazanet et al., 1995). In this context, Dutton (1999) introduced a line segmentation method based on the *sinuosity* measure in order to characterize linear features for map generalization. His method segments a line into sequences of similar *sinuosity regimes*.

In response to this research question, this thesis proposed a *trajectory segmentation* method relying on the *line simplification approach* in order to reduce the complexity of trajectories. The proposed method integrates characterizing features introduced for both time series and linear geometric objects in the char-

acterization of movement. Following Dutton (1999)'s work, this research applies the *sinuosity* measure in order to quantify the important features that characterize the movement parameters of objects. Here, the *sinuosity* measure represents the *frequency* of variations in a movement parameter. Besides sinuosity, this research considers the *amplitude* of variations in movement parameters. The method transfers the movement parameter profiles derived from the trajectories into segments of homogeneous movement characteristics (see section 3.2.2 and Research Paper 2). These segments convey the important movement features (i.e. the *frequency* and *amplitude* patterns of the movement parameters) that are necessary to characterize movements for knowledge discovery applications such as trajectory classification, simulation, and similarity analysis.

The experimental results of this thesis suggest that the movement parameters of different kinds of objects may generate distinctive signatures (i.e. certain patterns of movement features) depending the objects' underling physics and behavior (see Research Paper 2). For example, such distinction in the speed and acceleration profiles of bicycle and eye movements can be seen in Figure 3.4. On the other hand, these signatures become similar when objects move similarly (i.e. exhibit similar movement characteristics) as discussed in Research Paper 3. Moreover, the results suggest that learning knowledge about the signatures that can be generated by different types of moving entities in the transportation domain (e.g. pedestrians and vehicles) can contribute to transport mode detection. In this context, the developed classification strategy based on the segmentation method suggested promising results.

4.1.4 Movement Similarity

Research Question 4:

❖ *How can we quantify and formalize the similarity between the movement characteristics of different objects in space and time?*

The state of the art of the similarity analysis techniques provided in this thesis (chapter 2) suggests that the conceptualization of *similarity* is application dependent and it varies according to the purpose of study. Likewise, it is observed that the notion of *movement similarity* is considered differently in various studies. For instance, some studies defined *movement similarity* as the resemblance of movement paths in space (Chen et al., 2004, 2005; Vlachos et al., 2002a,b, 2004; Lin and Su, 2005), whereas others recognized *movement similarity* as the closeness of the space-time observations over objects' lifelines (Sinha and Mark, 2005; van Kreveld and Luo, 2007; Trajcevski et al., 2007; Buchin et al., 2009). The available similarity measures mainly consider one or two dimensions of movement (i.e. spatial, temporal). So far little attention has been paid to other movement characteristics including the derivative movement parameters such as speed, direction, acceleration in the quantification of movement similarity.

This thesis proposed a new similarity measure, named *normalized weighted edit distance (NWED)* based on the notion of *edit distance*. Research Paper 3 introduces the NWED function and the costs of the each edit operation considered in the NWED. In fact, NWED quantifies the similarity between the *variation patterns of the movement parameters* derived from the movements of two objects over time. Hence, it does not measure the similarity of the shapes of the objects' trajectories. However, since the movement parameters are obtained as the result of the movements of objects in space and time, the NWED measure encompasses both spatial and temporal dimensions of movement. Namely, in addition to the temporal dimension, which is explicitly involved in the definition of NWED, the spatial dimension is implicitly incorporated through the definition of movement parameters. For instance, *speed* is a spatio-temporal movement parameter that is defined as the the rate of change of the object's *position* (see Research Paper 1 and Table 3.1).

The similarity assessment technique proposed in Research Paper 3 incorporates only one movement parameter to represent the movements of objects for the similarity assessment. This thesis therefore suggested a complementary approach in Research Paper 4, to consider more dimensions in the similarity assessment. In the second approach, the movement similarity was quantified as the average Euclidean distance between the multidimensional representations of objects' trajectories in the *movement parameter space*. This method can contribute to the computation of the *spatial similarity* as well as the *spatio-temporal similarity* between the movements of objects.

In order to investigate the applicability of the developed similarity assessment methods, this study conducted a set of experiments to investigate the similarity in the dynamic behaviors of two different types of movement data: (1) North Atlantic hurricane tracks (cf. Research Papers 3 and 4) and (2) GPS tracks of van and motorcycle couriers (cf. Research Paper 3). These experiments aimed to evaluate the developed similarity assessment methods according to the available knowledge about the behavior of these datasets. The aim was to test whether the developed methods can be applied to regenerate similar knowledge from the given datasets.

The following meteorological hypothesis was investigated for hurricanes: *the North Atlantic Hurricanes have similar dynamic characteristics based on the time of formation and their source locations*. The experimental results of this study showed that the developed movement similarity assessment technique was more successful in discovering the similarity of hurricanes in terms of their *speed* and extracting structures in the movement of hurricanes than the previous method by Chen et al. (2004). That is, our method could differentiate the classic types of *low-latitude* hurricanes versus *high-latitude* hurricanes as it is suggested in the literature (Elsner and Kara, 1999). Moreover, in Research Paper 4, we showed

how similarity analysis can be applied to discover hurricanes with *concurrent* or *coincident* movement patterns.

Furthermore, Research Paper 3 investigated the applicability of the developed methods in the extraction of mode of transport and diurnal traffic patterns from courier movement data. The study showed that the segmentation approach helps to discover the mode of transport from unknown GPS tracks. The experimental results suggested that NWED is better suited for the assessment of traffic patterns along a specific segment of the street network rather than the entire network as a whole. The reason being that the NWED looks at the *sequential order* of the variations of movement parameters, whereas sequence becomes irrelevant in a large street network with a heterogeneous structure. By contrast, segmentation together with the proposed diversity-based clustering is more relevant for discovering patterns in such heterogeneous datasets.

4.2 General Discussion

This section attempts to provide a broad discussion on the overall results of this thesis and to describe the strengths and weaknesses of the developed methods. The discussion of the strength and weaknesses primarily focuses on the trajectory segmentation method (presented in Research Paper 2) and the similarity assessment approaches introduced in Research Paper 3 and 4. We will first start with a discussion of Research Paper 1, however.

4.2.1 Relevance of the two Movement Frameworks

This section illustrates the relation between the movement model proposed by Nathan et al. (2008) and the conceptual framework of movement suggested in Research Paper 1.

The conceptual model proposed by Nathan et al. (2008), described in section 2.1, defines a general model of movement embedded in the domain of ecology. The movement framework suggested in this thesis explicitly identifies the principal components of movement, which are required to define a movement pattern (see section 3.1.2 and Research Paper 1). In other words, the conceptual framework suggested in Research Paper 1 enriches the general movement model proposed by Nathan et al. (2008) with the fundamental elements of movement patterns. The extensions primarily concern the following items:

- Movement characteristics in the extended version of the movement model (see section 2.1, Figure 2.2) encompass the *primitive* and *derivative* movement parameters of the conceptual framework proposed in this thesis (see section 3.1.2, Table 3.1).
- *Time*, an essential dimension of the movement, is embedded in the *movement characteristics* as one of the primitive movement parameters. How-

ever, the temporal dimension was not explicitly incorporated in the original movement model by Nathan et al. (2008).

- The movement path is considered as one of the principal elements of the movement model in this thesis. The developed framework distinguishes between two types of movement paths (i.e *continuous* and *discontinuous*). The type of movement path has an influence on the types of the patterns that can evolve in the movement of an object.

4.2.2 The Strengths of the Methods

Utilization of the movement parameters

The main element of our proposed movement similarity assessment approach is *the movement parameter profile*. Each type of *movement parameter* describes one dimension of the movement of an object and helps to reveal a particular characteristic of the dynamic objects. In the developed methods, the choice of movement parameter depends on the aim of the similarity assessment study; the research question; and the hypothesis that needs to be investigated. For instance, if the study aims to explore the traffic pattern of a section of a street network (i.e. a highway or a road segment), the *speed* and *acceleration* informations become relevant (Andrienko and Andrienko, 2008). However, applying the *direction* parameter does not lead to desired results in this case. On the other hand, for the study of the dynamic behavior of vehicles at a U-turn or a junction, it is important to investigate the similarity of the vehicles' movement based on their *turning angle* (i.e. change of direction), besides the *speed* and *acceleration*. Likewise, working on the movement of hurricanes, movement parameters such as *speed*, *acceleration*, *turning angle*, and *approaching rate* towards the landfall location seem appropriate according to the meteorological hypothesis (Elsner and Kara, 1999). Each of these parameters can give an insight into one dimension of the movement behavior of hurricanes.

Storage and retrieval of trajectories

The segmentation process transforms the trajectories into a symbolic representation. That is, each point of the trajectory is indicated by a *character* in the movement parameter class sequence. As Chen et al. (2004) pointed out, using a symbolic representation (i.e. sequence of characters) requires less storage space (up to 12 % less) in comparison to the trajectory representation (i.e. sequence of coordinates). Consequently, the pattern matching and retrieval costs of the MPC sequence is less than its corresponding trajectory. Therefore, the segmentation process helps to improve the efficiency of knowledge discovery in moving objects databases.

Real movement datasets

From the review of the available similarity analysis techniques, it becomes evident that most of the previously proposed movement similarity analysis techniques either remained in theory (van Kreveld and Luo, 2007; Buchin et al., 2009) or were only evaluated on synthetic and simulated datasets (Vlachos et al., 2002a; Chen et al., 2004, 2005; Frentzos et al., 2007; Pelekis et al., 2007). Furthermore, it can be observed that often similarity analysis methods have been evaluated in comparison between a set of real trajectories and their own synthetic derivatives obtained from applying different levels of compression or perturbation on the original trajectories (Trajcevski et al., 2007; Frentzos et al., 2007; Pelekis et al., 2007). Therefore, the effectiveness of such methods on the real movement data is not sufficiently determined.

Real movement datasets are usually very complex and contain a lot of noise, gaps, and outliers. Moreover, various datasets might exhibit different patterns and structures. Therefore, it is important to consider the real movement data in the evaluation of analysis and knowledge discovery approaches. In this thesis, we investigated the applicability of the proposed approaches on several real movement datasets from different application domains, such as transportation and meteorology.

Moreover, our proposed segmentation process has an advantage in analyzing data in the presence of systematic positioning error. The reason is that the segmentation procedure relies on *relative* movement parameters (i.e. speed, direction etc.) not the *absolute* attributes such as position that are obtained from the tracking devices. Therefore, the effect of systematic errors is reduced in the computations. Moreover, since the segmentation algorithm considers a set of user-defined thresholds to distinguish between different levels of *sinuosity* and *deviation* along the movement parameter profiles, the effect of noise in data can be moderated by these thresholds.

4.2.3 Open Problems and Weaknesses of the Methods

The effects of thresholds

The proposed segmentation and similarity assessment procedures rely on a set of thresholds that are required to be specified by the user prior to the analyses. These thresholds include:

- sinuosity threshold: to distinguish the *low sinuosity* from the *high sinuosity* segments in movement parameter profiles.
- deviation threshold: to distinguish the points with *low* versus *high* amplitude levels along the movement profile in the segmentation process.
- distance threshold: to determine similar entities in the similarity assessment and clustering techniques.

The determination of these thresholds depends on a variety of parameters such the spatial and temporal scale of the processes under study, spatial and temporal granularity (i.e. sampling interval), and the noise level of the movement data. Moreover, the purpose of similarity assessment and trajectory clustering affects the choice of the required thresholds. Therefore, a sensitivity analysis is suggested as a prerequisite of running the proposed methods, in order to adapt the thresholds to the subject data.

The effects of outliers

The segmentation process relies on the amplitude level of movement parameters. Therefore, the presence of outliers in the data affects the accuracy of the results. Hence, as described in Research Paper 2, a preprocessing step including trajectory filtering is required prior to the segmentation of the trajectories.

Limitations of the Trajectory Classification

The proposed trajectory classification using the segmentation technique (see Research Paper 2) relies on the distinctions between the movement characteristics (e.g. speed, acceleration etc.) of the various objects. However, the effectiveness of this method decreases when different types of objects exhibit similar movement characteristics. For instance, in the transport mode detection in very heavy traffic, different types of vehicles to a great degree exhibit similar movement characteristics (e.g. very slow movement). Therefore, it becomes difficult to extract the mode of transport of vehicles from movement data alone. This is also observed on the other for other transport mode detection techniques (Upadhyay et al., 2008).

Movement path similarity

The proposed similarity assessment technique determines the similarity between the movement characteristics of objects in terms of the variations of movement parameters. In this thesis, the similarity of the movement path generated by the dynamic objects was not the main focus of the methods. It was, however, investigated implicitly in the form of *coincidence movement patterns* in Research Paper 4. The available spatial similarity techniques such as the ones introduced in section 2.5.4 could be applied for this purpose.

Efficiency and Scalability

The edit-distance based similarity analysis approaches, including our proposed NWED measure, have a relatively high computational cost (i.e. $O(n^2)$, where n is the number of observations along a trajectory). Therefore, the efficiency of this measure decreases for very long trajectories in large movement datasets. Although the segmentation process improves the storage and pattern matching of

the trajectories, additional strategies should be considered to improve the computations of similarity in such datasets. In order to overcome this weakness, applying indexing, dimension reduction, and pruning approaches would be beneficial (Vlachos et al., 2002a, 2004; Chen et al., 2004, 2005; Ding et al., 2008b).

Chapter 5

Conclusions

This thesis presented research on the development of methods for knowledge discovery from movement data about dynamic objects and processes. The main motivation was to exploit the *movement parameters* (e.g. speed, acceleration, turning angle) in support of the study of the *collective* movement behavior of objects. Hence, the main focus was on the *dynamic behavior* of objects rather than the geometric specifications of the objects' lifelines (e.g. geometry of the movement paths). The general aim of this chapter is to highlight the most pertinent outcomes of this thesis and draw attention to some outlooks on future work.

5.1 Main Contributions

In response to the research objectives, this thesis sought to develop conceptual and methodological knowledge about *movement parameters* and their features. This research mainly contributes to knowledge discovery from movement data with the development of a quantitative approach that enables similarity assessment of the movements of dynamic objects. The methodology uses the *movement parameters* that can be either derived from the trajectory of objects or directly recorded by new generations of mobile sensor technologies.

This thesis pursued a three-stage research process towards the main goal. The achievements of each stage account for a portion of the contributions made in this thesis. Accordingly, this thesis brought about the following main achievements:

- **Stage I** developed a *conceptual framework of movement*. The framework encompasses the fundamental elements of the movement of objects. These elements are required in the identification and formalization of movement patterns. The conceptual framework formed the basis for the subsequent stages. Furthermore, the first stage of this research introduced *a comprehensive classification of movement patterns*. The proposed conceptual framework and classification should facilitate the development of pattern recognition algorithms that are required to be efficient, effective, and as generic as possible in a more systematic approach.
- **Stage II** developed a *feature extraction and segmentation* method to reduce the complexity of moving object trajectories for the purpose of analysis and knowledge discovery. In this thesis a segmentation technique was conceived

to generate a concise view of trajectories encapsulating the important movement features. Segmentation aims at decomposing data into parts of similar characteristics. Hence, the developed method can serve to assess the similarity of the movement characteristics of multiple objects in the classification (e.g. transport mode detection) and clustering of movement data. The developed segmentation algorithm forms the key element of the main movement similarity assessment approach in the subsequent stage.

- **Stage III** proposed two different *similarity assessment approaches* to extract the similarities between the *movement characteristics* of objects. Both methods detect objects whose *movement parameters* exhibit similar patterns. The main method applies the segmentation process, proposed in Stage II, on a single movement parameter profile. The NWED distance was introduced, based on the edit distance, as the similarity measure of the segmented profiles. The alternative, second method assesses the similarity of dynamic objects in a multidimensional movement parameter space using the average Euclidean distance. Furthermore, the final stage introduced *clustering* and *movement pattern extraction* strategies relying on the above similarity assessment methods. The applicability of the developed techniques was evaluated on clustering and movement pattern discovery from the movements of dynamic objects.

It is necessary to remark that this thesis investigated the applicability of the proposed methods on real movement datasets from different application domains. Sample data included, movement data from meteorology (i.e. hurricanes), different data sources in transportation (i.e. pedestrians, bicycles, motorcycles, cars), and eye-tracking. The experimental results of this thesis indicate that the proposed methods could be successfully applied in support of movement behavior studies of dynamic objects and processes. The methods were developed generically so that they could be applied to any kind of movement data from various application domains such as movement ecology of animals, urban ecology, and human mobility studies.

5.2 Insights

In response to the research questions, this thesis undertook an iterative process of conceptual design, prototype implementation, and experimentation. The actual research process helped to reveal valuable insight about the handling and discovering knowledge in real movement data from various application domains.

This research started off with the assumption that generic analytical and data mining techniques effective for handling all kinds of movement data can be developed. However, the experiments conducted in this research revealed an important insight that not every knowledge discovery method is suitable for any kind of movement data. Therefore, it is important to evaluate the proposed knowledge discovery techniques applying *relevant real movement data*, *feasible proxies*,

or *artificial simulated* movement data. The experiments provided evidence that there are similarities and differences in the movement characteristics of different types of moving objects in various domain. For instance, the physics of movement of virtually the mass-less movement of eyes is very different from the whole-body movement of objects that are governed by inertia such as vessels. Therefore, movement data generated from such objects cannot be a viable proxy for the other.

It is a major insight gained from this study that *movement parameters* can enrich exploratory and knowledge discovery techniques with valuable information in studying and understanding movement, where purely positional information becomes deficient. Learning such similarities and differences in the movement characteristics of objects can help in the *identification* or *simulation* of moving objects for different purposes. In this regard, this study showed how the *segmentation* approach applying *movement parameters* can help to extract typical patterns in the variations of movement parameters generated from movements of certain objects. Such knowledge can be applied to classify trajectories of unknown moving objects by similarity to the trajectories of previously learned objects or to assess the similarity of proxy or simulated data to the reference moving objects. Examples include: *detection of mode of transport* from unknown trajectories of people using different transportation means and *finding or simulating the most feasible proxies* for desired moving objects in various application domains. Simulation of dynamic processes as well as human and animal mobility behavior requires knowledge about their movement characteristics, the way that these objects move, and the patterns that their movement parameters can generate, rather than the purely positional information.

Furthermore, working with various movement datasets during this research, we observed that real movement datasets are often elaborately structured and hence each dataset has its own complications and pitfalls. Therefore, each dataset creates distinct challenges for preprocessing and may even require a new knowledge discovery technique. Finally, the research process of this thesis revealed the necessity of generating a set of benchmark datasets with detailed metadata, to be used in the development and validation of movement analysis methods.

5.3 Outlook and Challenges for Future Research

5.3.1 Taxonomy of Movement Patterns

This thesis proposed a classification of movement patterns. In this classification the *generic patterns* and *behavioral patterns* were recognized as the top level categories of the patterns to distinguish between the patterns that are generic to all moving objects from the domain-specific behavioral patterns. The proposed classification exhibits a hierarchical structure in the category of the *generic pat-*

terns. Further developments of the taxonomy should include the identification of *behavioral* movement patterns in individual application domains. Furthermore, it is necessary to extend the proposed classification with an ontological approach in collaboration with the domain experts into a complete taxonomy of movement patterns. The taxonomy should consist of the following elements:

- Classification of *generic patterns* (including possible future extensions).
- Identification and classification of *behavioral patterns* for domain specific applications.
- Accurate definition and formalization of the generic patterns.
- Accurate definition of the identified behavioral patterns.
- Description of the functional relationship between the behavioral patterns and/with their underlying generic patterns.

Our proposed classification can be considered as the first step towards a comprehensive taxonomy of movement patterns. The *movement pattern wiki*[1] provided in this study can serve as a common platform for knowledge sharing within the community and communication with research from other disciplines such as movement ecology, transportation, and marine science.

5.3.2 The Effect of Clustering and Classification

In the second and the third stages of this thesis, a set of trajectory classification and clustering strategies were introduced, respectively. These strategies served as the applications that motivated the experiments evaluating the developed segmentation and similarity assessment techniques. This thesis made use of the available standard classification and clustering techniques, such as *support vector machines* and *agglomerative hierarchical clustering*, respectively, in conjunction with our own segmentation and similarity assessment methods. This thesis did not investigate the sensitivity of the results to the classification and clustering techniques that were used. We suggest that future research should pursue a comparative analysis of the existing classification and clustering methods.

On the other hand, as indicated in chapter 2, several studies have been conducted on the development of trajectory classification and clustering approaches. However, to the best of our knowledge there is no benchmark study that provides a guideline about the strengths and weaknesses of such methods specifically for movement data. Thus, it is essential to investigate the influence of such techniques on the outcomes and if applicable provide a taxonomy and recommendation of use of these techniques according to different aspects of movement data.

[1] http://movementpatterns.pbworks.com/

5.3.3 Integration of the Two Similarity Assessment Methods

In this thesis we introduced a second similarity assessment approach that works in the movement parameter space. This technique served as a complementary approach to the main similarity assessment method using trajectory segmentation. However, in this study we did not investigate how the two methods actually complement each other and how they could be used in conjunction. An immediate extension of our work should consider the combination of these two methods in an integrated approach. The new approach might be capable of handling multiple dimensions of the movement data at the same time for the purpose of similarity analysis. That is, the integrated approach is hoped to enable similarity assessment of the movements of objects not only according to the behavior of their movement parameters, but also with respect to the geometric shape of their movement paths.

5.3.4 Development of Segmentation Methods for Movement Data

The results of this study suggested that segmentation is very helpful in the preparation of movement data for further analysis. Segmentation can be beneficial in various applications in movement research. This thesis showed how such technique can be applied in order to compress trajectories into a simplified and yet revealing structure. Segmentation can also be applied in order to partition trajectories into sequence of events or activities in movement behavior studies. A comprehensive survey is required to investigate the usability and limitations of the available segmentation methods . Future studies should investigate criteria (geometric and/or semantic) that shall be taken into account in the segmentation of movement data and if required to further develop the segmentation techniques with respect to such criteria.

5.3.5 Exploitation of Context

The methods proposed in this thesis rely only on movement data captured from dynamic objects and processes and, thus disregard the effects of external factors and geographic context (see Figure 2.2). Today, this is the case for almost all the available techniques in the domain of knowledge discovery of movement data. In order to extract valuable patterns and valid knowledge from the movement of objects, it is necessary to exploit the information about context within which the movement takes place. Also, adding context to the analysis would improve the evaluation, validation, and interpretation of the obtained results. For instance, incorporating information about the ocean dynamics, sea and air temperature, and wind speed would significantly support the study of the movement behavior of hurricanes. Therefore, a future extension of our proposed approach should consider the incorporation of contextual information in the assessment of similarity between the movement characteristics of objects. Enriching knowledge

discovery methods with contextual information would be a step forward towards gaining a better understanding of the movement behaviors of dynamic objects and processes.

Bibliography

Agrawal, R., Faloutsos, C., and Swami, A. (1993). Efficient Similarity Search in Sequence Databases. In *Foundations of Data Organization and Algorithms*, volume 730, pages 69–84. Springer Verlag.

Agrawal, R., Lin, K.-I., Sawhney, H. S., and Shim, K. (1995). Fast Similarity Search in the Presence of Noise , Scaling , and Translation in Time-Series Databases. In *VLDB '95: Proceedings of the 21th International Conference on Very Large Data Bases*, pages 490–501. Morgan Kaufmann Publishers Inc.

Alt, H. (2009). *The computational geometry of comparing shapes*, pages 235–248. Springer-Verlag.

Alt, H. and Godau, M. (1995). Computing the Fréchet distance between two polygonal curves. *International Journal of Computational Geometry and*, 5:75–91.

Alt, H. and Guibas, L. J. (2000). Discrete Geometric Shapes: Matching, Interpolation, and Approximation. In Sack, J.-R. and Urrutia, J., editors, *Handbook of Computational Geometry*, pages 121–153. North-Holland, Amsterdam.

Andrienko, G. and Andrienko, N. (1999). Interactive maps for visual data exploration. *International Journal of Geographic Information Science*, 13(4):355–374.

Andrienko, G. and Andrienko, N. (2008). Spatio-temporal aggregation for visual analysis of movements. In *Visual Analytics Science and Technology, 2008. VAST'08. IEEE Symposium on*, pages 51–58, Andrienko2008. IEEE.

Andrienko, G. and Andrienko, N. (2009). Visual Analytics for Geographic Analysis, exemplified for different types of movement data. *Information Fusion and Geographic Information Systems*, pages 3–17.

Andrienko, N. and Andrienko, G. (2007). Designing Visual Analytics Methods for Massive Collection of Movement Data. *cartogeraphica*, 42(2):117–138.

Ankerst, M., Breunig, M. M., Kriegel, H.-p., and Sander, J. (1999). OPTICS : Ordering Points To Identify the Clustering Structure. *SIGMOD*, 28(2):49–60.

Aßfalg, J., Kriegel, H., Kröger, P., Kunath, P., Pryakhin, A., and Renz, M. (2006). Similarity search on time series based on threshold queries. In *Advances in Database Technology-EDBT 2006*, pages 276–294. Springer.

Aßfalg, J., Kriegel, H.-P., Kröger, P., Kunath, P., Pryakhin, A., and Renz, M. (2008). Similarity Search in Multimedia Time Series Data Using Amplitude-Level Features. In *Advances in Multimedia Modeling*, pages 123–133.

Bashir, F. I., Khokhar, A. A., and Schonfeld, D. (2007). Object Trajectory-Based Activity Classification and Recognition Using Hidden Markov Models. *IEEE transactions on image processing*, 16(7):1912–1919.

Bay, S. and Pazzani, M. (2001). Detecting group differences: Mining contrast sets. *Data Mining and Knowledge Discovery*, 5:213–246.

Berndt, D. and Clifford, J. (1994). Using dynamic time warping to find patterns in time series. In *AAAI-94 workshop on knowledge discovery in databases*, pages 359–370.

Bernstein, D. and Kornhauser, A. (1996). An Introduction to Map Matching for Personal Navigation Assistants.

Blythe, P. and Miller, G. (1996). Human simulation of adaptive behavior: Interactive studies of pursuit, evasion, courtship, fighting, and play. In *the fourth international conference on simulation of adaptive behavior*, pages 13–22.

Bozkaya, T., Yazdani, N., and Özsoyoğlu, M. (1997). Matching and indexing sequences of different lengths. In *Proceedings of the sixth international conference on Information and knowledge management - CIKM '97*, pages 128–135, New York, New York, USA. ACM Press.

Brakatsoulas, S., Pfoser, D., and Salas, R. (2005). On map-matching vehicle tracking data. In *VLDB'05: Proceedings of the 31st international conference on Very large data bases*, pages 853–864.

Buchin, K., Buchin, M., van Kreveld, M., and Luo, J. (2009). Finding long and similar parts of trajectories. In *Proceedings of the 17th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems - GIS '09*, pages 296–305, New York, New York, USA. ACM Press.

Buchin, K., Buchin, M., and Wenk, C. (2008). Computing the Fréchet distance between simple polygons. *Computational Geometry*, 41(1-2):2–20.

Çöltekin, A., Fabrikant, S. I., and Lacayo, M. (2010). Exploring the efficiency of users' visual analytics strategies based on sequence analysis of eye movement recordings. *International Journal of Geographical Information Science*, 24(10):1559–1575.

Chen, L. and Ng, R. (2004). On the marriage of Lp-norms and edit distance. In *VLDB '04: Proceedings of the Thirtieth international conference on Very large data bases*, pages 792–803. VLDB Endowment.

Chen, L., Özsu, M. T., and Oria, V. (2004). Symbolic Representation and Retrieval of Moving Object Trajectories. In *MIR '04: Proceedings of the 6th ACM SIGMM international workshop on Multimedia information retrieval*, pages 227–234. In Proc. of the ACM SIGMM international workshop on multimedia information retrieval.

Chen, L., Özsu, M. T., and Oria, V. (2005). Robust and fast similarity search for moving object trajectories. In *SIGMOD '05: Proceedings of the 2005 ACM SIGMOD international conference on Management of data*, pages 491–502, New York, NY, USA. ACM.

Cortes, C. and Vapnik, V. (1995). Support-vector networks. *Machine Learning*, 20(3):273–297.

Crochemore, M. and Rytter, W. (1994). *Text algorithms*. Oxford University Press, Inc., New York, NY, USA.

Das, G., Gunopulos, D., and Mannila, H. (1997). Finding similar time series. In *PKDD '97: Proceedings of the First European Symposium on Principles of Data Mining and Knowledge Discovery*, pages 88–100.

Ding, H., Trajcevski, G., and Scheuermann, P. (2008a). Efficient Similarity Join of Large Sets of Moving Object Trajectories. In *15th International Symposium on Temporal Representation and Reasoning*, pages 79–87. Ieee.

Ding, H., Trajcevski, G., Scheuermann, P., Wang, X., and Keogh, E. (2008b). Querying and mining of time series data: experimental comparison of representations and distance measures. *Proc. VLDB Endow.*, 1(2):1542–1552.

Dutton, G. (1999). *A Hierarchical Coordinate System for Geoprocessing and Cartography*. PhD thesis, University of Zurich.

Dykes, J. A. and Mountain, D. M. (2003). Seeking structure in records of spatio-temporal behaviour : visualization issues, efforts and applications. *Computational Statistics & Data Analysis*, 43:581 – 603.

Eiter, T. and Mannila, H. (1994). Computing discrete Fréchet distance. Technical report, Technische Universität Wien, Christian Doppler Laboratory for Expert Systems, TU Vienna.

Elsner, J. and Kara, A. (1999). *Hurricanes of the North Atlantic: Climate and society*. Oxford University Press, New York Oxford.

Erwig, M., Ting, R. G., and Schneider, M. (1999). Spatio-temporal data types: An approach to modeling and querying moving objects in databases. *GeoInformatica*, 296:269–296.

Ester, M., Kriegel, H.-p., Sander, J., and Xu, X. (1996). A Density-Based Algorithm for Discovering Clusters in Large Spatial Databases with Noise. In *Data Mining and Knowledge Discovery*, pages 226–231.

Etienne, L., Devogele, T., and Bouju, A. (2010). Spatio-Temporal Trajectory Analysis of Mobile Objects Following the Same Itinerary. In *Proceedings of the International Symposium on Spatial Data Handling (SDH)*, Hong Kong.

Fabrikant, S. I., Rebich-Hespanha, S., Andrienko, N., Andrienko, G., and Montello, D. R. (2008). Novel Method to Measure Inference Affordance in Static Small-Multiple Map Displays Representing Dynamic Processes. *Cartographic Journal, The*, 45(3):201–215.

Faloutsos, C., Jagadish, H., Mendelzon, A., and Milo, T. (1997). A signature technique for similarity-based queries. In *Proceedings. Compression and Complexity of*

SEQUENCES 1997 (Cat. No.97TB100171), pages 2–20. IEEE Comput. Soc.

Faloutsos, C., Ranganathan, M., and Manolopoulos, Y. (1994). Fast subsequence matching in time-series databases. In *SIGMOD '94: Proceedings of the 1994 ACM SIGMOD international conference on Management of data*, volume 23 (2), pages 419–429, New York, NY, USA. ACM.

Fayyad, U., Piatetsky-Shapiro, G., and Smyth, P. (1996). From Data Mining to Knowledge Discovery in Databases. *American Association for Artificial Intelligence*, pages 37–54.

Fraile, R. and Maybank, S. J. (1998). Vehicle Trajectory Approximation and Classification. In *British Machine Vision Conference*, pages 832–840.

Frank, A. (2001). *Socio-Economic Units: Their Life and Motion*, chapter 2, pages 21–34. Taylor & Francis.

Frentzos, E., Gratsias, K., and Theodoridis, Y. (2007). Index-based Most Similar Trajectory Search. In *ICDE 2007. IEEE 23rd International Conference on Data Engineering*, pages 816–825.

Fu, Z., Hu, W., and Tan, T. (2005). Similarity based vehicle trajectory clustering and anomaly detection. In *ICIP 2005: IEEE International Conference on Image Processing, 2005.*, pages 602–605. Ieee.

Galton, A. (2005). Dynamic collectives and their collective dynamics. In *Spatial Information Theory, Lecture Notes in Computer Science*, volume 3693, pages 300–315. Springer.

Giannotti, F. and Pedreschi, D. (2008). *Mobility, Data Mining and Privacy.* Springer-Verlag, Berlin Heidelberg.

González, M. C., Hidalgo, C. a., and Barabási, A.-L. (2008). Understanding individual human mobility patterns. *Nature*, 453(7196):779–82.

Goodrich, M. T., Mitchell, J. S. B., and Orletsky, M. W. (1999). Approximate Geometric Pattern Matching Under Rigid Motions. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 21(4):371–379.

Gudmundsson, J., Kreveld, M., and Speckmann, B. (2007). Efficient Detection of Patterns in 2D Trajectories of Moving Points. *Geoinformatica*, 11(2):195–215.

Holyoak, M., Casagrandi, R., Nathan, R., Revilla, E., and Spiegel, O. (2008). Trends and missing parts in the study of movement ecology. *Proceedings of the National Academy of Sciences of the United States of America*, 105(49):19060–5.

Hornsby, K. and Egenhofer, M. (2002). Modeling moving objects over multiple granularities. *Annals of Mathematics and Artificial Intelligence*, 36(1):177–194.

Imfeld, S. (2000). *Time, Point and Space- Towards a Better Analysis of Wildlife Data in GIS.* PhD thesis, Universitaet Zuerich.

Jolliffe, I. (1986). *Principal Component Analysis*. Springer, New York.

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:*, (1972):141–150.

Keogh, E., Chakrabarti, K., Pazzani, M., and Mehrotra, S. (2001). Dimensionality Reduction for Fast Similarity Search in Large Time Series Databases. *Knowledge and Information Systems*, 3(3):263–286.

Kisilevich, S., Mansmann, F., Nanni, M., and Rinzivillo, S. (2010). *Spatio-Temporal Clustering : a Survey*. ISTI-CNR, Italy.

Kraak, M. (1988). The space-time cube revisited from a geovisualization perspective. *Proceedings of the 21st International Cartographic*, (August):10 – 16.

Laube, P. (2005). *Analysing Point Motion, Spatio-Temporal Data Mining of Geospatial Lifelines*. PhD thesis, Universitz of Zurich.

Laube, P. (2009). *Progress in Movement Pattern Analysis*, volume 3 of *Behaviour Monitoring and Interpretation – BMI - Smart Environments*, pages 43–71. IOS Press BV.

Laube, P., Dennis, T., Forer, P., and Walker, M. (2007). Movement beyond the snapshot – Dynamic analysis of geospatial lifelines. *Computers, Environment and Urban Systems*, 31(5):481–501.

Laube, P., Imfeld, S., and Weibel, R. (2005). Discovering relative motion patterns in groups of moving point objects. *International Journal of Geographical Information Science*, 19(6):639–668.

Laube, P., Kreveld, M., and Imfeld, S. (2004). Finding REMO - detecting relative motion patterns in geospatial lifelines. In *Developments in Spatial Data Handling*, pages 201–215. Springer.

Laube, P. and Purves, R. (2006). An approach to evaluating motion pattern detection techniques in spatio-temporal data. *Computers, environment and urban systems*, 30(3):347–374.

Lee, J., Han, J., Li, X., and Gonzalez, H. (2008). TraClass: trajectory classification using hierarchical region-based and trajectory-based clustering. *VLDB '08*, (212):1081–1094.

Lee, J., Han, J., and Whang, K.-Y. (2007). Trajectory clustering: a partition-and-group framework. In *Proceedings of the 2007 ACM SIGMOD international conference on Management of data*, pages 593–604, New York, New York, USA. ACM.

Levenshtein, V. (1966). Binary codes capable of correcting deletions, insertions, and reversals. *Soviet Physics-Doklady*, 10:707–710.

Li, Z., Lee, J., Li, X., and Han, J. (2010). Incremental clustering for trajectories. In *Database Systems for Advanced Applications*, pages 32–46. Springer.

Lin, B. and Su, J. (2005). Shapes based trajectory queries for moving objects. *GIS '05: Proceedings of the 13th annual ACM international workshop on Geographic information systems*, pages 21–30.

Miller, H. J. (2005). A Measurement Theory for Time Geography. *Geographical Analysis*, 4(2):63–45.

Miller, H. J. and Han, J. (2009). *Geographic Data Mining and Knowledge Discovery.* Taylor & Francis Group, second edition.

Mountain, D. and Raper, J. (2001). Modelling human spatio-temporal behaviour: a challenge for location-based services. In *Proceedings of the 6th International Conference on GeoComputation*, University of Queensland, Brisbane, Australia.

Mountain, D. M. (2005). *Exploring mobile trajectories: An investigation of individual spatial behaviour and geographic filters for information retrieval.* PhD thesis, City University, London Department of Information Science School of Informatics.

Mouza, C. and Rigaux, P. (2005). Mobility Patterns. *GeoInformatica*, 9(4):297–319.

MOVE (2009). Knowledge discovery from moving objects (move). memorandum of understanding for the implementation of a european concerted research action designated as cost action ic0903. `http://w3.cost.eu/typo3conf/ext/bzb\_securelink/pushFile.php?cuid=253\&file=fileadmin/domain\_files/ICT/Action\_IC0903/mou/IC0903-e.pdf`.

Nakos, B. P. and Mitropoulos, V. C. (2005). Critical Point Detection Using the Length Ratio (LR) for Line Generalization. *Cartographica: The International Journal for Geographic Information and Geovisualization*, 40(3):35–51.

Nanni, M. and Pedreschi, D. (2006). Time-focused density-based clustering of trajectories of moving objects. *Intelligent Information Systems*, 27:267–289.

Nathan, R., Getz, W. M., Revilla, E., Holyoak, M., Kadmon, R., Saltz, D., and Smouse, P. E. (2008). A movement ecology paradigm for unifying organismal movement research. *Proceedings of the National Academy of Sciences of the United States of America*, 105(49):19052–9.

Pelekis, N., Kopanakis, I., Marketos, G., Ntoutsi, I., Andrienko, G., and Theodoridis, Y. (2007). Similarity Search in Trajectory Databases. In *TIME'07: Proceedings of the 14th International Symposium on Temporal Representation and Reasoning*, pages 129–140, Washington, DC, USA. IEEE Computer Society.

Plazanet, C., Affholder, J., and Fritsch, E. (1995). The importance of geometric modeling in linear feature generalization. *Cartography and Geographic Information Science*, (4):291–305.

Rinzivillo, S., Pedreschi, D., Nanni, M., Giannotti, F., Andrienko, N., and Andrienko, G. (2008). Visually driven analysis of movement data by progressive clustering. *Information Visualization*, 7(3-4):225–239.

Sakurai, Y., Yoshikawa, M., and Faloutsos, C. (2005). FTW: Fast Similarity Search under the Time Warping Distance. In *PODS '05: Proceedings of the twenty-fourth ACM SIGMOD-SIGACT-SIGART symposium on Principles of database systems*, pages 326–337, New York, New York, USA. ACM Press.

Sinha, G. and Mark, D. M. (2005). Measuring similarity between geospatial lifelines in studies of environmental health. *Journal of Geographical Systems*, 7(1):115–136.

Smyth, C. S. (2001). Mining Mobile Trajectories. In Harvey J. Miller, J. H., editor, *Geographic Data Mining and Knowledge Discovery*, chapter 14, pages 337–361. Taylor & Francis.

Spaccapietra, S., Parent, C., Damiani, M., De Macedo, J., Porto, F., and Vangenot, C. (2008). A conceptual view on trajectories. *Data & Knowledge Engineering*, 65(1):126–146.

Torrence, C. and Compo, G. (1998). A practical guide to wavelet analysis. *Bulletin of the American Meteorological Society*, 79(1):61–78.

Toshniwal, D. and Joshi, R. C. (2005). Similarity Search in Time Series Data Using Time Weighted Slopes. *Informatica*, 29:79–88.

Trajcevski, G., Ding, H., Scheuermann, P., Tamassia, R., and Vaccaro, D. (2007). Dynamics-aware similarity of moving objects trajectories. In *GIS '07: Proceedings of the 15th annual ACM international symposium on Advances in geographic information systems*, pages 1–8, New York, NY, USA. ACM.

Turchin, P. (1998). *Quantitative analysis of movement: measuring and modeling population redistribution in animals and plants.* Sinauer Associates, Sunderland, MA, USA.

Upadhyay, D., Axhausen, K. W., Flamm, M., and Kaufmann, V. (2008). Optimal parameter values for mode detection in GPS post-processing : An experiment.

van Kreveld, M. and Luo, J. (2007). The definition and computation of trajectory and subtrajectory similarity. In *GIS '07: Proceedings of the 15th annual ACM international symposium on Advances in geographic information systems*, pages 1–4, New York, NY, USA. ACM.

Vlachos, M., Gunopulos, D., and Das, G. (2004). Rotation invariant distance measures for trajectories. In *KDD '04: Proceedings of the 2004 ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 707 – 712, New York, New York, USA. ACM Press.

Vlachos, M., Gunopulos, D., and Kollios, G. (2002a). Discovering similar multidimensional trajectories. In *ICDE '02: Proceedings 18th International Conference on Data Engineering*, pages 673–684. IEEE Computer Societ.

Vlachos, M., Gunopulos, D., and Kollios, G. (2002b). Robust similarity measures for mobile object trajectories. In *DEXA'02: Proceedings of the 13th International Workshop on Database and Expert Systems Applications*, pages 721 – 728, Los

Alamitos, CA, USA. IEEE Computer Society.

Wood, Z. and Galton, A. (2009). Classifying Collective Motion. In Gottfried, B. and Aghajan, H., editors, *Behaviour Monitoring and Interpretation - Ambient Assisted Living*, pages 129–155. IOS Press BV.

Yanagisawa, Y., Akahani, J.-i., and Satoh, T. (2003). Shape-Based Similarity Query for Trajectory of Mobile Objects. In *MDM '03: Proceedings of the 4th International Conference on Mobile Data Management*, pages 63–77, London, UK. Springer-Verlag.

Yi, B.-K., Jagadish, H., and Faloutsos, C. (1998). Efficient retrieval of similar time sequences under time warping. In *ICDE '98: Proceedings 14th International Conference on Data Engineering*, pages 201–208. IEEE Computer Society.

Yuelong, Z., Shijin, L., Dingsheng, W., and Xiaohua, Z. (2008). A Novel Approach to the Similarity Analysis of Multivariate Time Series and Its Application in Hydrological Data Mining. In *2008 International Conference on Computer Science and Software Engineering*, pages 730–734. Ieee.

Part II

Research Papers

Research Paper 1

Dodge, S., Weibel, R. and Lautenschütz, A.K. (2008). Towards a Taxonomy of Movement Patterns. *Journal of Information Visualization*, Vol. 7, pp. 240 – 252.

Information Visualization (2008) 7, 240 – 252

www.palgrave-journals.com/ivs

Towards a taxonomy of movement patterns

Somayeh Dodge
Robert Weibel
Anna-Katharina Lautenschütz

Department of Geography, University of Zurich, Switzerland

Correspondence:
Somayeh Dodge, Department of Geography, University of Zurich, Winterthurerstrasse 190, 8057 Zurich, Switzerland.
Tel.: +41 44 6355131;
Fax: +41 44 6356848;
E-mail: somayeh.dodge@geo.uzh.ch

Abstract
Areview of research that has been carried out on data mining and visual analysis of movement patterns suggests that there is little agreement on the relevant types of movement patterns and only few, isolated definitions of these exist. Since the research interest in this area has recently started to soar, we believe that this is a good time to approach the definition of movement patterns in a more systematic and comprehensive way. This paper intends to contribute to the development of a toolbox of data mining algorithms and visual analytic techniques for movement analysis by developing firstly a conceptual framework for movement behavior of different moving objects and secondly a comprehensive classification and review of movement patterns. We argue that this is indispensable as a basis for the development of pattern recognition and information visualization algorithms that are required to be efficient (i.e. usable on massive data sets), effective (i.e. capable of accurately detecting patterns not artifacts), and as generic as possible (i.e. potentially applicable to different types of movement data). We demonstrate the utilization of our classification by answering the question as to what extent eye tracking data can be seen as a proxy of other types of movement data. We have set up a moderated discussion platform in order to facilitate the further evolution of our proposed classification towards a consolidated taxonomy in a consensus process.
Information Visualization (2008) **7,** 240 – 252. doi:10.1057/palgrave.ivs.9500182

Keywords: Movement pattern; moving object; spatio-temporal data mining; visual analytics; taxonomy; behavior

Introduction

Mobility is a key element of many processes and activities, and the understanding of movement is important in many areas of science and technology such as meteorology, biology, sociology, transportation engineering, to name but a few. Hence, increasing amounts of movement tracking data and other data about the dynamics of mobile objects or agents are being collected.

By definition, moving objects are entities whose positions or geometric attributes change over time. However, in many applications the dimension of the object is not as important as its position. Hence, moving objects are considered as moving points, whose trajectories (i.e. paths through space and time) can be visualized and analyzed. By all accounts, moving objects can be categorized into two major groups of geo-referenced *vs* non-geo-referenced dynamic objects. In other words, some are dynamic objects that move about in geographic space and may thus be geographically referenced such as humans, animals, or vehicles, while the other group includes dynamic phenomena that move in a non-geographic space, including gaze point movements in eye movement studies or particles in a bubble chamber. Each of these dynamic objects, to a varying degree, shares some similarities but also exhibits differences to the others in terms of the corresponding data structure, dynamic behavior and nature of movement.

Received: 16 April 2008
Revised: 5 June 2008
Accepted: 9 June 2008
Online publication date: 7 August 2008

In most cases, moving object data sets are rather large in volume and complex in the structure of movement patterns that they record. Therefore, it is necessary to develop efficient data mining algorithms and visual analytic techniques in order to extract useful and relevant information, regularities, and structure from massive movement data sets. Such tools can help researchers detecting movement patterns and exploring movement behavior of different entity types.

Moving object analysis has recently become the focus of many research projects in the area of geographic information science. A review of the literature on moving object data mining and visualization highlights the importance and significant progress of these studies. However, it also suggests that there is little agreement on the relevant types of movement patterns and only few, isolated definitions of these exist. Therefore, we believe that this is a good time to approach the definition of patterns in movement data more systematically and comprehensively.

The objectives of this article are firstly to propose a conceptual framework of the elements defining the movement behavior of different moving objects and secondly the development of a comprehensive classification and definitions of movement patterns. We restrict our proposal to *Moving Point Objects* (MPOs), ignoring the dimension of the objects. The third objective is a call for collaboration in consolidating the proposed classification and corresponding definitions into a complete taxonomy in a moderated participatory process among domain specialists.

There are several good reasons for a comprehensive taxonomy and accurate definitions of movement patterns. First, the design of efficient and effective data mining algorithms and objective visual analytic methods requires accurate formalization of the movement patterns and their properties. Second, most of the quoted work departs from the assumption that generic algorithms can be developed that will be suitable for different kinds of MPO data. However, this will only be possible if we know exactly the similarities but also the differences between different types of moving object data and the patterns inherent to them. Third, and related to the second point, is the argument of interoperability: Movement analysis and visualization extends across diverse disciplines and hence different people should be able to gain the same understanding of the same terms. This also applies to the 'translation' of natural language descriptions of movement patterns, as they may be collected in cognitive experiments.[1] Fourth, a classification and formalization of patterns is necessary to give guidelines for the development of visual and interactive methods that are expected to enable users to detect and explore patterns. Therefore it is a stepping-stone for the optimization of visualizations of movement data and renders pattern extraction comparable for various graphical representations and thus supports humans in their decision-making process. It also provides a starting point and interpretation guide for the visual analysis of movement data, supporting humans in identifying movement patterns. An agreement on pattern types beforehand allows an easier identification during the visual analysis process. And fifth and finally, an accurate definition of motion patterns and their constituents is also important for the evaluation of detected patterns by simulation.[2]

The remainder of the paper is organized as follows: We start with a brief summary of related work on movement pattern mining and visualization. We continue by explaining the methodology used to develop our classification of movement patterns. Next, we introduce the framework of movement that forms the conceptual basis of the classification. We then introduce the classification, starting with an overview, followed by descriptions of the various patterns that we have identified. We end with a discussion and conclusions, explaining how we envision evolving and consolidating the classification into a complete taxonomy through a moderated process.

Mining and visualizing movement patterns

Progress in positioning and tracking technologies is giving increasing access to huge amounts of spatio-temporal movement tracking data and other data about the movement of mobile objects. Non-stop generation of space–time trajectories from different kinds of moving entities provides the possibility to discover useful and interesting information about personal and vehicular mobile behavior, to find patterns, extract their meaning, and expand our knowledge about the mobile world.[3]

Data mining as a component of the KDD process is the application of specific algorithms for extracting patterns from data. In a moving object database, various data mining tasks, including exploratory data analysis, descriptive and predictive modeling, mining of association rules, and other pattern detection techniques can be applied on MPO trajectories in order to extract patterns of movement and discover spatio-temporal behavior of different types of moving objects.[2,4]

Generally, movement patterns include any recognizable spatial and temporal regularity or any interesting relationship in a set of movement data, whereas the proper definition (i.e. the instantiation) of 'pattern interestingness' depends on the application domain. Early work on movement pattern analysis includes the simulation study of human adaptive behavior[1] and the methods developed for the spatio-temporal analysis of wild animal movements by Imfeld[5]. Recent years have witnessed almost an explosion of research activities, triggered by the advent of cheap and ubiquitous positioning and data collection technology. Selected representatives of these more recent publications include the work on the extraction of movement patterns from trajectories generated by individual users of location-based services;[3,6] and the work on data mining of movement patterns in groups of moving objects.[7–10] Furthermore, visual analytic methods for exploratory analysis of movement data have been proposed by Andrienko and Andrienko.[11]

Visual analytics and information visualization serve the exploration of moving object data – in particular pattern extraction – and are built on the premise that humans are able to reason and learn more effectively in a visual setting. Information visualization research has produced a large variety of movement data representations. However, there are few empirical evaluation studies to assess how useful these representations are.

The above publications document significant progress of research over the past few years. These studies usually set out with fairly accurate definitions of the patterns they are looking for – as an indispensable prerequisite to visual analysis and data mining[4] – but they tend to be restricted to a selected, narrow set of patterns. Hence, we are still facing a fundamental problem and impediment to the development of a comprehensive toolbox of movement analysis techniques: There is neither agreement on the relevant types of movement patterns nor any comprehensive and systematic definition of these. Therefore, there is a need to create a systematic classification of patterns in movement data. Andrienko and Andrienko[11] probably come closest to what may be termed such a comprehensive taxonomy for this purpose. However, while their proposal forms an excellent point of departure for subsequent work, the taxonomy should be better rooted in the relevant literature and the associated definitions must become more detailed and accurate.

Methodology

We started off our work by a review of the research conducted so far, including the above references as well as review articles such as,[12,13] as well as additional application-specific references. The aim of this first step was to categorize patterns of movement proposed by other researchers and to discover commonalities and differences in terminology and pattern types. Furthermore, we wanted to avoid developing redundant, conflicting terminology. The result of this first step is the classification of movement patterns presented below (Figure 1).

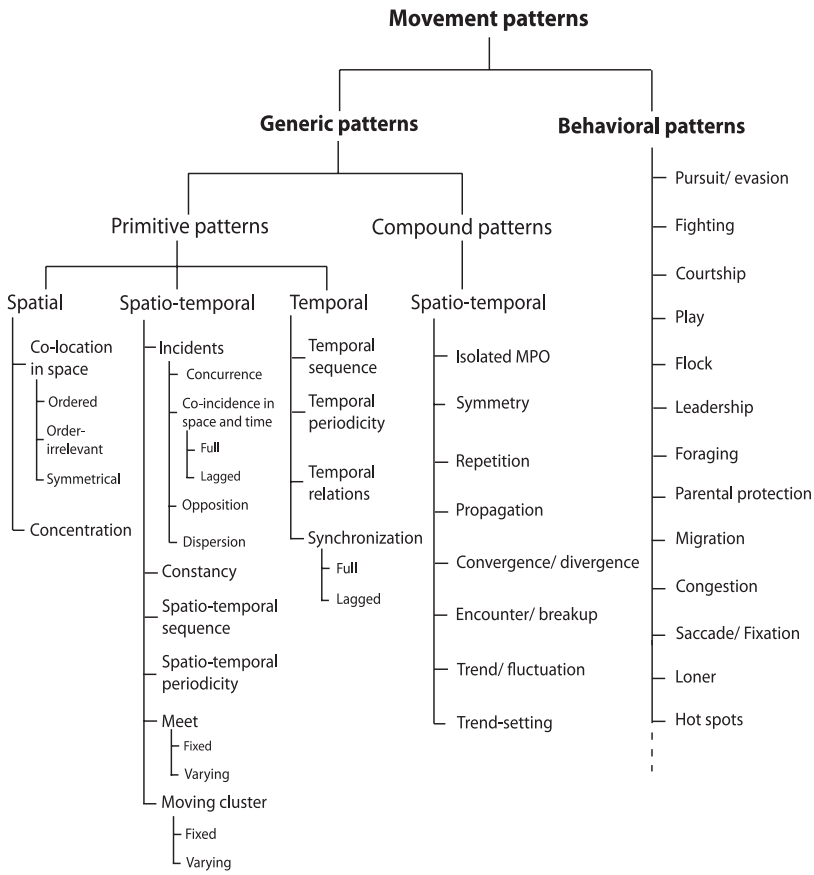


Figure 1 Classification of movement patterns.

While this first step essentially represents a top-down approach we also approached this problem from bottom-up, decomposing movement into its constituting elements such as movement parameters, influencing factors, etc. The aim here was to develop a conceptual framework of the movement of dynamic objects that could be used to build definitions of individual movement patterns.

The following sections present the proposed conceptual framework of movement and the classification of movement patterns, followed by descriptions of the various patterns constituting the classification.

Conceptual framework of movement

Generally speaking, the analysis of the movements of dynamic objects has formed the basis of physical science ever since the times of Galileo and Newton, in the 17th century. Movement as a physical phenomenon is change of position. In the geographic dynamic domain, movement is defined as a shift in location of an object over time while the object maintains the same identity.[14] Movement data can be treated as a function matching pairs (entity, time moment) with position in space.[11]

In order to study the movement behavior of dynamic objects, it is important to take a closer look at movement itself. In other words, it is necessary to know what exactly the parameters are that define movement, what external factors influence movement, and most importantly to understand what types of movement patterns can be composed from these primitives of movement. For this purpose, we tried to develop a conceptual framework of movement. The framework consists of the following elements discussed in the remainder of this section.

Movement parameters

We consider three major groups of movement parameters, *primitive parameters*, and derived parameters that we term *primary derivatives* and *secondary derivatives* (Table 1). Primitive and derived parameters may be organized into spatial, temporal, and spatio-temporal dimensions: For instance, primitive spatial movement parameters consist of the position of the object over consecutive timestamps and time *instance* (i.e. point in time) and time *interval* (i.e. temporal sampling rate) are primitive parameters in the temporal dimension. From these primitive parameters, several derived ones can be defined. Among the primary derivatives, *distance* and *direction* of movement are in the spatial dimension and solely a direct function of *position* (x, y). *Duration* is defined as a period of time in which a movement is observed. Duration is a direct function of time and consists of one or more time intervals. *Speed* (i.e. rate of change of the object's position) and *velocity* (i.e. rate of change of position *and* direction) are parameters that combine both space and time dimensions, and can be derived directly from spatial position and time instances. Higher-order parameters of movement such as acceleration can be derived from primary derivatives.[1,15]

Of the secondary derivatives, the definition of the spatial parameters is assumed to be commonly known. For instance, s*patial distribution* represents a snapshot of the positions of MPOs in the database at a specific time. *Sinuosity* is a function of *distance* and refers to the degree of windingness of an object's trajectory.[16,17] Among the temporal parameters, *temporal distribution* denotes the distribution of events along the time line. *Change of duration* (also termed *convexity* in finance) denotes the rate of change of the duration between different observations of the same movement behavior (e.g. rate of change of the migration duration of a species of animal). *Acceleration* (i.e. rate of change of the object's speed) represents a spatio-temporal parameter derived from *speed*. *Approaching rate* is a function of speed and distance and 'describes whether and how intensively moving object approaches its destination.'[16] Giannotti and Pedreschi[15] describes some other derivatives of movement parameters, in terms of overall characteristics and dynamics of movement.

In movement analysis, it is often preferable to define movement parameters not only in an *absolute* sense (i.e. with respect to the external reference system) but also in a *relative* sense, that is, in relation to the movement of other MPOs. Relative movement parameters are particularly useful in the analysis of movements of two objects[1] or in groups of MPOs.[10]

Table 1 Movement parameters

Parameters/Dimension	*Primitive*	*Primary derivatives*	*Secondary derivatives*
Spatial	Position *(x,y)*	Distance *f(posn)* Direction *f(posn)* Spatial extent *f(posn)*	Spatial distribution *f(distance)* Change of direction *f(direction)* Sinuosity *f(distance)*
Temporal	Instance *(t)* Interval *(t)*	Duration *f(t)* Travel time *f(t)*	Temporal distribution Change of duration *f(duration)*
Spatio-temporal *(x, y,t)*	—	Speed *f(x,y,t)* Velocity *f(x,y,t)*	Acceleration *f(speed)* Approaching rate

Table 2 Number of objects involved in a movement

Number of MPOs	*# Obj.*	*Relationship*	*Examples*
Individual	$N = 1$	—	Trajectory *of a person* over a day
Group	$N > 1$	Functional	Trajectories *of a flock* of sheep foraging on a meadow
Cohort	$N > 1$	Statistical	Eye movement trajectories *of all female participants* in an experiment

Number of moving objects involved

Obviously, moving objects show different behavior in different situations, depending on whether the MPOs are moving alone, or accompanied by other objects or in a group. Andrienko and Andrienko[11] distinguish between 'individual movement behavior' (IMB) on the one hand and two types of movement behaviors involving multiple objects on the other hand, termed 'momentary collective behavior' (MCB, comparing movement characteristics of multiple objects at a given instance) and 'dynamic collective behavior' (DCB, looking at movement characteristics of multiple objects over a duration). While we agree with the specific role of IMB, we argue that collective movement behavior should not be approached through arbitrary slicing across the time axis (MCB) or along the time axis (DCB), as that would ignore functional relationships that exist between moving objects. By 'functional relationships' we mean behaviorally relevant relationships between members of a group that are implied by the special nature of the group (e.g. between members of a family, players in a football team, cars in a convoy). Hence, in the analysis of collective movement we differentiate between *groups*, which share a behaviorally relevant functional relationship (e.g. a flock of sheep, a wolf pack) and *cohorts*, which merely have a factor in common that may be statistically relevant, such as the similar age (called age cohort) or sex (Table 2).

Path type

Paths of moving objects may take different forms. Some, perhaps most, moving objects travel more or less continuously, generating a *continuous path* (i.e. curvilinear path), exemplified by migrating elk, a pedestrian, or a car moving on a road. Such a continuous path is typically discretized into regular steps prior to computing the movement parameters.[9,17,18] Other moving objects will generate *discontinuous paths* of moves (i.e. steps) between stops.[17,18] Examples of these include the movement of a bee between flowers, or saccadic movements between fixations in eye movements. In this type of stop-and-go movement the stops themselves may become more important in explaining the movement than the displacements between stops, such as in eye tracking. The definition of stops is dependent on the application domain. For instance, in eye movement research fixations are identified by several methods, either by analysis of velocity, density, or duration.[19]

Influencing factors

Moving objects are influenced by various factors that impact and constrain their movements. We distinguish four groups of influencing factors, one group that represent intrinsic properties of the moving object, and three groups of extrinsic factors:

- *Intrinsic properties* of the moving object. Each moving object has its particular intrinsic physical and behavioral movement properties, such as top speed, acceleration behavior, etc.
- *Spatial constraints* (networks, barriers, etc.): These objects really act as a firm constraint to the movement, such as the road network constrains the movement of a car.
- *Environment* against which the movement takes place. Besides firm constraints, the environment is full of factors that are more or less attractive or repelling to moving objects. For instance, different vegetation influences animal movements, and the image in eye movement experiments acts as a target of attractors of varying degree.
- *Influences of other agents*. For instance, the movements of an animal are influenced by other members of its own or other species that induce competition, attraction, or disturbances.

Obviously, the above influencing factors vary between entity types. They are therefore non-generic and must be defined specifically for each entity type. Andrienko and Andrienko[11] have also proposed a classification of influencing factors. We believe, however, that our classes are better defined with respect to the behavioral characteristics of movement. For instance, we devote a specific class to influences of other agents.

Scale/granularity

Depending on the influencing factors acting on a particular mobile object, the resulting movement will take place at different *spatial and temporal scales*. Scale issues play an important role in producing and interpreting a specific movement pattern. The spatial scale of movement – ranging from the very local scale in the order of centimeters (e.g. eye movements) to global scale (e.g. air traffic) – dictates the requirements on the precision and accuracy of the positioning technology used to produce movement fixes. Likewise, the temporal scale varies between very short-term and long-term behavior. Again, this imposes requirements on the temporal sampling resolution in producing movement fixes. Since the sampling may be discontinuous (i.e. may exhibit sampling gaps), oversampling and undersampling may occur, and many

pattern analysis methods require fixes at regular time intervals, trajectories often need to be resampled prior to movement pattern detection. The *temporal granularity* that is generated in the resampling critically depends on the movement characteristics of the MPO in question. That is, setting the right temporal granularity requires at least some limited prior knowledge of the movement under study. Using the wrong temporal granularity may dramatically bias the resulting patterns. So, while it may still be possible to develop generic methods for data mining and visual analytics, the preprocessing of the trajectory data definitely is domain-specific. Laube *et al*.[10] provide a discussion of scale and granularity issues, quoting work by other authors as well.

Classification of movement patterns

In order to facilitate the detection of movement patterns, it is necessary to understand what types of patterns may exist in the data.[11] Having defined the elements of the above conceptual framework of movement, we are now in a position to attempt to develop a systematic classification of patterns in movement data. This classification and its elements should be applicable for all the common types of moving objects relevant to the geo-spatial domain, such as humans, animals, cars and eye movement data.

As explained above, our proposed classification is based on a review of the pertinent literature, in order to maximize reuse of already existing sufficiently accurate definitions, and minimize redundant, conflicting terminology. We have tried to align the definitions from the literature with the elements of our conceptual framework, and redefine and sharpen definitions where necessary. In some cases, additional pattern concepts had to be introduced to make the classification comprehensive.

In the remainder of this section, we first explain the overall organization of the classification, that is, the categories according to which the various movement patterns can be organized. This structure is also shown in Figure 1.

Individual *vs* group movement patterns

In our framework, one defining element of movement patterns is whether they occur in individuals or in multiple MPOs. In Table 2, we distinguished between movements of individuals on the one hand, and groups and cohorts as representatives of multiple MPOs on the other. For the sake of defining movement patterns, however, we can neglect the type of relationship that exists between the members of a collection of moving objects (which distinguishes groups from cohorts). The type of relationship between object collections only plays a role in the interpretation of movement patterns, not in defining them. Hence, we can simply distinguish *individual vs group* movement patterns.

Dimension of the patterns

Following the conceptual framework, depending on the dimension(s) used in studying the movement, patterns are categorized into three types of spatial, temporal and spatio-temporal patterns (Table 1).

Generic patterns *vs* behavioral patterns

'Movement is behavior.'[1] Yet, in order to develop data mining algorithms and visual analytics methods that are as widely applicable as possible, it is helpful to try to identify patterns that may be found in *any* form of behavior that builds on movement. These *generic patterns* are usually insufficient to explain specific behavior of particular moving objects. For instance, two animal species may exhibit periodicities in their movements, yet in one case the period may be diurnal and annual in the other. Hence, generic patterns represent the building blocks used to form higher-level movement patterns that correspond to a particular behavior typical of a particular MPO. These patterns are then called *behavioral patterns*. Behavioral patterns also include movement patterns that can solely be found in certain types of MPOs (e.g. certain animal species).

Primitive *vs* compound patterns

Generic movement patterns are simpler than behavioral patterns. However, even among the generic patterns, different levels of complexity can be distinguished. Hence, in our classification we distinguish between primitive and compound patterns among the generic movement patterns. *Primitive patterns* are the most basic forms of movement patterns, those where only a single movement parameter varies, such as *incidents* and *constancy*. In contrast, *compound patterns* are made up of several primitives involving complex inter-object relations, such as *trend-setting*, *convergence*, and *encounter*. In a similar way, Andrienko and Andrienko[11] proposed two major types of patterns for DCB, called descriptive and connectional patterns, respectively.

Describing the patterns

In the following two sections we give descriptions and examples of the patterns of the proposed classification of movement patterns; first the generic patterns, then the behavioral patterns. For lack of space, the descriptions are kept short and refer to the corresponding literature. Table 3 uses the organization of the classification of Figure 1 and plots the generic movement patterns against the parameters of Table 1 that can be used to describe a particular pattern. Note that full definitions, more examples, and illustrative graphics can be found on a wiki set up by the authors.[20] The reader is cordially invited to visit this moderated website in order to discuss the proposed definitions, and contribute to the development of a consolidated, comprehensive taxonomy.

Table 3 Comparing the generic patterns with the elements of the conceptual framework

	Primitive param.		*Primary derivatives*				*Secondary derivatives*		*Applicable to*		*Dimension*		
Generic patterns	*Position*	*Instance*	*Distance* $f(x, y)$	*Direction* $f(x, y)$	*Speed* $f(x, y, t)$	*Duration*	*Curvature*	*Acceleration*	*Individual MPO*	*Multiple MPOs*	*Spatial* (x)	*Temporal* (t)	*Spatio-temporal* (x, y, t)
Primitive patterns													
Co-location in space	x		x							x	x		
Concentration	x		x						x	x	x		
Concurrence	x	x	x	x	x	x	x	x		x			x
Co-incidence in space and time	x	x	x							x			x
Opposition	x	x	x	x	x	x	x	x		x			x
Dispersion	x	x	x	x	x	x	x	x		x			x
Constancy	x	x	x	x	x	x	x	x	x	x			x
Sequence	x	x				x			x	x		x	
Periodicity	x	x	x	x	x	x	x	x	x	x		x	x
Meet	x	x	x			x				x			x
Moving cluster	x	x	x	x	x	x	x	x		x			x
Temporal relations		x				x			x	x		x	
Synchronization	x	x	x	x	x	x	x	x		x		x	
Compound patterns													
Isolated object	x	x	x	x	x	x	x	x		x			x
Symmetry	x	x	x	x	x	x	x	x		x			x
Repetition	x	x	x	x	x	x	x	x	x	x			x
Propagation	x	x	x	x	x	x	x	x		x			x
Convergence/divergence	x	x		x		x				x			x
Encounter/breakup	x	x	x	x						x			x
Trend/fluctuation	x	x	x	x	x	x	x	x	x	x			x
Trend-setting	x	x	x	x	x	x	x	x		x			x

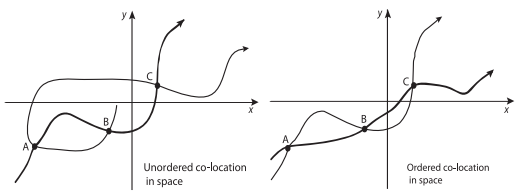


Figure 2 Co-location in space.

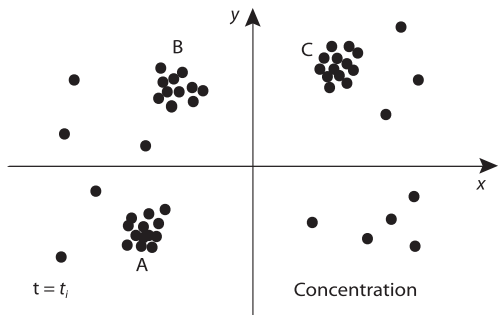


Figure 3 Concentration.

Generic patterns: primitive patterns

Co-location in space: Occurs when the trajectories of moving objects have some positions in common.[11] There are three types of co-location in space: *ordered co-location* exists when the common positions are attained in the same order; *unordered co-location* if the common positions are attained in different orders; and *symmetrical co-location* when the common positions are attained similarly but in opposite orders.[11] For instance, co-location in space occurs during an eye movement experiment when different test subjects fixate on similar positions on the image; if the visiting order of fixation positions is the same, co-locations are ordered, and unordered otherwise. As another example, tourists visiting the same four places A to C in a city generate co-locations along their trajectories. If the order is from A to C in one case, and reverse in the other, then we have symmetrical co-location. Two types of co-location in space are shown in Figure 2.

Concentration: We dedicate a subclass of spatial patterns to spatial *concentration* of moving objects at a certain instance of time (e.g. A, B and C areas in Figure 3). As an example, congestion denotes a zone of high vehicle density on a transportation network. Likewise, fixations are spatially dense positions of eye movement tracks and represent concentration zones on the underlying image.

Incidents: Laube and Imfeld[21] introduced incidents as patterns occurring among multiple objects that can be further categorized as the following patterns:

- *Concurrence*: Is an incident of a set of entities showing the same values of motion attributes at a certain instant or duration. It happens when a set of objects exhibits a synchronous movement or at least similar motion parameter values over a certain duration (e.g., a flock of wild geese flying with similar motion azimuth).
- *Co-incidence in space and time*: Andrienko and Andrienko[11] introduced a specific kind of incidence

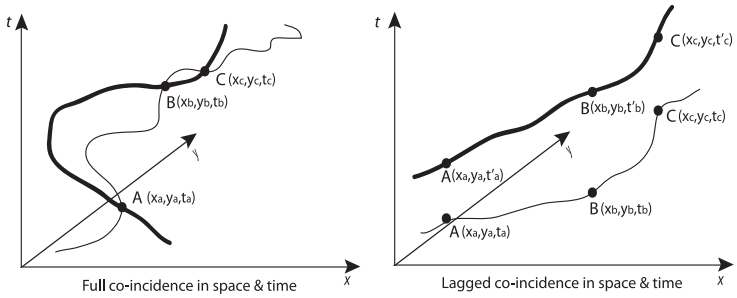


Figure 4 Co-incidence in space and time.

considering similar positions of moving objects (Figure 4). There are two variants of *co-incidence* patterns, *full vs lagged*. In the case of *full co-incidence*, the same positions are attained at the same time while in *lagged co-incidence*, it happens after a time delay. For instance, two different flocks of wild geese reach a particular pond at the same time or separated by a delay of 1 day.

- *Opposition*: A bi- or multi-polar arrangement of motion parameter values, such as the spatial splitting of a group of moving objects shown in a sudden appearance of two opposite motion directions. For instance, when flying geese are prompted to fly in opposite directions by a source of disturbance.
- *Dispersion*: It the opposite of concurrence. An evident pattern in a group of MPOs that is performing a non-uniform or random motion.

Constancy: When the movement parameters remain the same or change insignificantly for a particular duration,[12] for example, when a convoy of cars moves along a straight road, at a constant speed, speed and direction, and the derived parameters remain the same. Similarly, when a flock of wild geese is heading north on the spring migration or when football players execute a forward move.

Sequence: A *sequence* is an ordered list of visits to a series of locations. It consists of a contiguous series of *segments* with a known start and end point in space and time. A *spatio-temporal sequence* refers to an ordered subsequence of locations with their timestamps. As an example of sequential patterns, the tendency of tourists to visit a set of places A to C in a particular sequence A → B → C within specified duration could be mentioned.[6,22,23] Another example is the sequential order of fixations for several runs of an eye movement experiment.

Periodicity: Temporal periodic patterns indicate cyclical (e.g., yearly, weekly, or daily) phenomena.[6] Andrienko and Andrienko[11] introduced *spatio-temporal periodicity*, or regular repetition as occurrence of the same patterns or pattern sequences at regularly spaced time intervals (e.g., migrating geese follow the same path every year.)

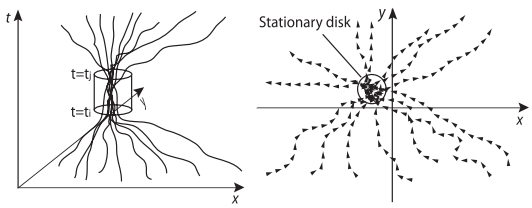


Figure 5 Meet.

Meet: A *meet* pattern consists of a set of MPOs that form a stationary cluster. That is, they stay within a cylinder of a certain radius in the space–time cube; in the projection to the plane, they stay within a stationary disk of specific radius in a certain time interval (Figure 5). There are two variants of meet, *fixed meet* and *varying meet*, depending on whether the objects that stay together for a certain duration are the same or change in the meeting region.[13,24] As an example for a fixed meet pattern, we mention families of geese that gather in the fall in a specific place to form a flock. An example for a varying meet is the rental car drop-off at an airport.

Moving cluster: Refers to a set of objects that stay close to each other while taking the same path for a specific duration. Nevertheless, it is not necessary that the objects participating in the pattern remain the same, but they may enter and leave, while the cluster is moving. A flock of migrating geese, a convoy of cars following the same route, and troops that move parallel on a military battlefield are different examples of moving clusters.[8,12,25] Based on the above definition, there are two variants of moving clusters, namely *fixed moving cluster* and *varying moving cluster*, depending on whether the participating entities stay the same or change during the observed period.[24]

A moving cluster pattern is also termed a *flock* by some authors.[7,9] However, we consider this a context-dependent term specifically related to biology, as in 'a flock of sheep.' Therefore, we prefer 'moving cluster' as

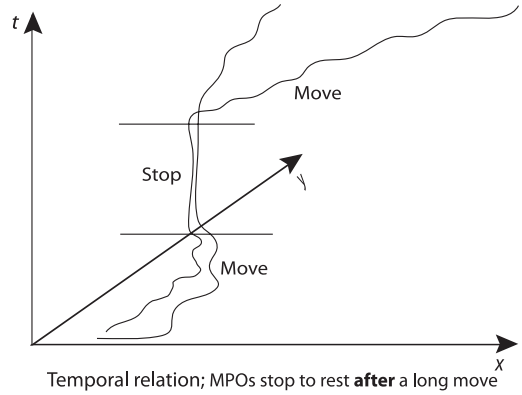


Figure 6 Temporal relations.

a term for the generic form of this pattern and 'flock' as a term for the specific behavioral pattern in biology. We will discuss *flock* further in the section about behavioral patterns.

Temporal relations: These include any temporal relation among various events on the time axis.[6] For instance, a flock of wild geese usually stops to rest *after* a long continuous flight (Figure 6).

Synchronization in time: According to Andrienko and Andrienko[11] there are two variants of synchronization patterns. *Full synchronization* happens when similar changes of movement parameters (e.g., speed, acceleration, direction, etc.) occur at the same time. In contrast, *lagged synchronization* happens when the changes of movement parameters occur after a time delay. For example, forwards in football (soccer) start moving in a similar direction synchronously, when their goalkeeper kicks the ball towards the opponent's side.

Generic patterns: compound patterns

The following patterns are all built from several primitive patterns described above, and hence termed compound patterns. Furthermore, since compound patterns consist of several primitives, they are invariably spatial *and* temporal.

Isolated object: Refers to an individual moving object (normally belonging to a group of MPOs) pursuing its own path, without any influence on or by the movement of other objects,[21] for example, when a goose misses the flock and travels alone.

Symmetry: Refers to sequences of patterns, where the same patterns are arranged in reverse order,[11] such as wild geese heading north in the spring, and south in the fall.

Repetition: Refers to the occurrence of the same patterns or pattern sequence at different time intervals.[11] For

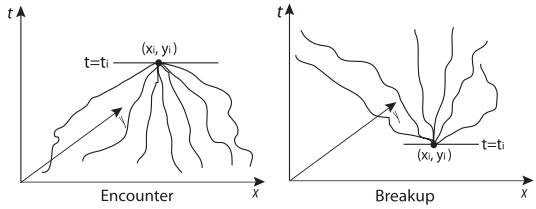


Figure 7 Encounter *vs* breakup.

instance, in a football match the wingers may repeatedly sprint along the sidelines or in an eye tracking experiment the test subjects may repeatedly scan the underlying image up and down.

Propagation: Propagation occurs when one object starts to show a certain movement parameter value, and little by little other objects start adopting the same pattern. By the same token, with every time step more objects are involved.[21] For instance, in the spring snow geese gradually start leaving at different times, depending on how far north they are migrating. The difference to the trend-setting pattern discussed below is that propagation always happens gradually and does not necessarily involve the influence of a 'trend-setter' object.

Convergence vs divergence: *Convergence* refers to the movement of a set of objects to the same location, while the original movement direction of the involved objects does not change. In other words, the motion azimuth vectors of the objects involved will be intersecting within a specific range and within a specific duration. The objects need not arrive at exactly the same time, however. For example, several flocks of snow geese may converge toward a lake to rest. *Divergence* is defined as the opposite pattern of convergence and describes a group of moving objects that disperse from a common location.[8,9]

Encounter vs breakup: Encounter refers to moving to and meeting at the same location. Encounter is a specific form of convergence pattern where the objects arrive at the same time (Figure 7, left). In an encounter pattern a set of MPOs have motion azimuth vectors that can be extrapolated from the current movement such that the vectors intersect within a specific range and the MPOs meet at the same time. Breakup is defined as the opposite of the encounter pattern (Figure 7, right) and describes a divergence, adding a temporal (concurrency) constraint.[8,9] In a football match, an encounter occurs when several players rush towards the ball and reach it at the same time. A breakup occurs when ducks flee from a pond after a gunshot is heard.

Trend vs fluctuation: Trend refers to consistent changes in the movement parameters of moving objects.[11] For example, for an airplane circling in a holding pattern the rate of change of the movement direction will

remain constant. Conversely, *fluctuation* refers to irregular changes in the movement parameters of moving objects,[11] for example, a flock of geese may change their flying formation between V-shape, irregular V-shape, or sometimes lines.

Trend-setting: The *trend-setting* pattern was introduced by.[9,21] Trend-setters are defined as objects that anticipate a certain movement pattern that is afterwards followed by a subset of the other moving objects. In another words, *trend-setters* are objects that influence the movement of others not necessarily in a spatial and temporal proximity.[9,21] For example, in a football match, a striker executing a sudden rush towards the adversary goal acts as a trend-setter, anticipating (or triggering) a similar movement direction by the defenders and his/her own teammates. There are two variants of trend-setting, *non-varying trend-setting* with a fixed subset of followers and *varying trend-setting*.[9,21] In the case of varying trend-setting, the subset of followers may change over the time intervals of the observation duration. Similarly to a moving cluster, in the trend-setting pattern objects move in the same direction or may have other similar movement characteristics such as same speed or acceleration.[8,26] The above authors use the term leadership to describe a specific kind of trend-setting pattern. Similar to the concept of a flock, we believe that leadership is a term mostly useful to describe animal and human movement behavior. For the same reason, we classified leadership as a type of behavioral pattern.

Behavioral patterns

As mentioned above, *behavioral patterns* are defined depending on the context and particular MPO type involved. Hence, for any type of moving object several behavioral patterns can be recognized that consist of generic movement patterns, yet take a particular meaning that is specific to a particular application domain. While we have tried to provide a complete list of generic movement patterns, we introduce only selected behavioral patterns for the sake of illustration. Clearly, there would be many more such patterns, as indicated in Figure 1, but we leave it to the domain specialists (e.g. behavioral ecologists, wildlife scientists) to define behaviorally relevant patterns using the generic patterns identified in this article as building blocks.

Pursuit/evasion: *Evasion* and *pursuit* belong together. Evasion refers to one animal (i.e. the prey) trying to move away and escape from a threatening, pursuing animal (i.e. the predator). They describe very high-speed movements combined with large amounts of turning and looping extending over a potentially large area of the environment.[1] *Pursuit* and *evasion* can be seen as a combination of leadership and trend-setting movements where the evader leads and affects the pursuer's movement parameters.

Fighting: Fighting is a combination of pursuit and evasion, attack and defense. Very high-speed movements are combined with large amounts of tightly intertwined turning, looping and frequent contact (where trajectories meet) in small distance between objects.[1] Fighting behavior consists of a complex combination of different generic patterns such as incidents, concurrence, repetition, co-location in space and time. If fighting occurs among a group of animals, other types of generic patterns such as convergence, divergence, encounter, breakup, and leadership might be involved.

Play: In animals, particularly juveniles, play is a form of practicing behaviors such as pursuit, evasion, fighting, or courtship. Hence, playing behavior consists of a combination of these movement behaviors, exhibiting looping, rapid dashes, and long still pauses. In play, animals repeatedly switch roles between pursuer and evader, or attacker and defender.[1]

Flock: The flock pattern describes a group of animals (representing the generic pattern of a moving cluster) moving in the same direction while staying close together,[8] for instance, a flock of sheep.

Leadership: The *leadership* pattern occurs when there is an individual that acts as the leader of a group for a specific duration. An individual can be said to be a leader if it does not follow anyone and is followed by sufficiently many other individuals at a proximate distance.[9,21] Leadership is a specific kind of the generic pattern *trend-setting* and is mostly associated with animal or human behavior. The difference to trend-setting, however, is that leadership requires spatial and temporal proximity, while in trend-setting this requirement is less stringent.

Congestion: Refers to movement with slower than usual speeds, longer trip times, and increased queuing. Extreme *congestion* in road traffic will lead to a traffic jam, with vehicles fully blocked for possible extended periods of time. This pattern can be seen as a combination of *meet* and *concurrence*, along with *constancy*. A convoy of cars is a *moving cluster* that may move at slow or near-zero speed and hence lead to congestion. Similarly, traffic jams form spots of *concentrations* (stationary clusters).

Saccade/fixation: In eye movement tracking studies, researchers typically analyze eye movements in terms of *fixations* (i.e. pauses over informative regions of interest) and *saccades* (rapid movements between fixations).[19] In a spatio-temporal sense, eye movement recordings represent a combination of *constancy* and *repetition* of fast high-speed movements between fixations. In a spatial sense, they can be seen as a sequence of *concentration*s, as fixations represent spots of high spatial density.

Mapping the classification to the conceptual framework

Table 3 gives a summary of how the introduced conceptual framework covers the proposed classification. As

mentioned before, movement patterns originate from changes in movement parameters of moving entities, and depending on the dimension of characteristics involved they are categorized into spatial, temporal, and spatio-temporal patterns. Table 3 shows the dependency of patterns on variations of each movement parameter and gives an overview of patterns regarding their dimension. It also compares patterns in terms of the number of moving objects involved (individual *vs* multiple MPOs). For instance, in co-incidence in space and time (Figure 4), objects have similar position at a certain instance or after a delay in time. In other words, in full co-incidence the distance between objects at a certain instance is near-zero. Consequently, the definition of the co-incidence in space and time pattern is based on three movement parameters: position, instance, and distance. Hence, this pattern is categorized as a spatio-temporal pattern (Table 3). It is also obvious that co-incidence in space and time involves multiple MPOs as it can only be defined by relations between two or more objects (Figure 4, Table 3).

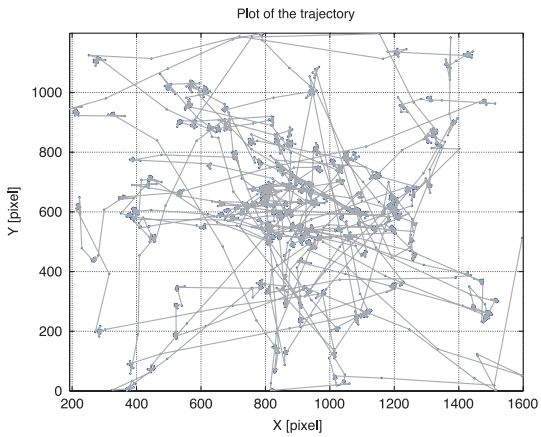


Figure 8 Sample trajectory of eye movement (data courtesy of Arzu Coltekin, The University of Zurich).

Discussion

Utilizing the classification: an example

Moving point analysis has recently become the focus of many research projects in the area of GIScience. Many spatio-temporal data models and analytical methods have been proposed at the theoretical level; however, only few of them have been implemented and applied in practice. A critical success factor for empirically based research is the availability of relevant movement data. The main problem is that data about MPOs are not easily available and accessible due to data security and privacy issues.[15] In order to overcome the problem of data scarcity, one may consider utilizing data that can act as proxies of 'physical' movement data, such as self-collected eye movement data from human subject experiments on graphic displays. Here, the main question is how similar are the movement behaviors of different MPOs. Specifically, to what extent can eye movement data be used as a proxy of other kinds of MPO data?

To answer this question, let us start with the hypothesis that the generic patterns defined above can be found in the movements of *any* MPO, but different MPOs will generate particular types of behavioral patterns. That is, the behavioral pattern that appears in the movement of an animal is assumed to differ from the patterns of movement of a human, and from those generated by eye movements. Here, we illustrate these similarities and differences utilizing the elements of the proposed conceptual framework and classification to compare eye movement with the whole-body movement of a human. Figures 8 and 9 show sample trajectories for these two types of movement data.

Movement parameters: All movement parameters are applicable to both eye movement and human movement data but they are different in terms of the values that they

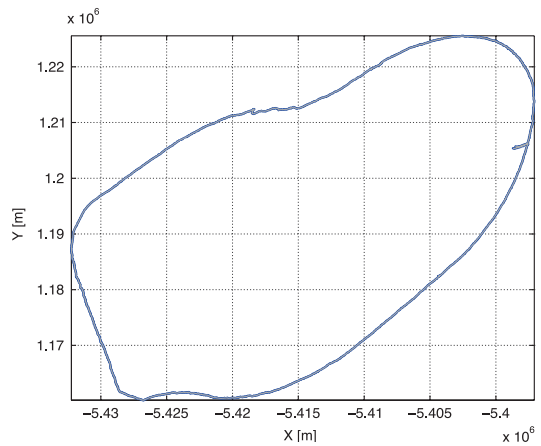


Figure 9 Sample trajectory of human movement (pedestrian) (data courtesy of www.openstreetmap.com).

can take. For instance, eyes can move quickly in fractions of a second from one end of a picture to the other in an almost mass-less movement somewhat akin of 'teleporting,' while the acceleration of human whole-body motion is dictated by greater mass and inertia.

Number of objects involved: Eye movement trajectories can be analyzed for individual test subjects or as cohorts with other subjects, which have a factor in common (e.g., the same sex, age, etc.). Human whole-body movement can be considered for an individual, in a group (e.g. family), or cohorts.

Path type: Owing to the different physics underlying the movement process, eye movements are best represented

as *discontinuous paths* with stop-and-go sequences, while human whole-body movement best matches the *continuous path* model, depending on the spatial and temporal study scale.

Influencing factors: Completely different, domain-specific factors affect eye movement and human whole-body movement. For eye movement, for example, the content of the test image *vs* physical obstacles, topography, and other persons in human movement.

Scale/granularity: Eye movement and human motion are very different with respect to the *spatial and temporal scales* involved. The spatial scale of eye movement tracking in laboratory experiments is in the order of centimeters; for humans it varies between several meters and global scale. Similarly, the temporal scale of movement differs from fractions of seconds for eyes, to minutes and days for humans.

Movement patterns: There are some similarities and differences between patterns generated from eye movement or human motion. For example, overlaid eye movement data from all subjects participating in an eye tracking experiment might (coincidentally) show a pattern similar to a moving cluster. Similarly, a group of people taking part in guided tour to an exhibition can be conceived as a moving cluster too. However, the group of exhibition visitors forms a functional group (somewhat akin to a 'flock'), while the collective movement behavior of the eyes of multiple test persons does not share a functional relationship, and would be simply the result of an 'overlay' of individual behaviors.

It is also conceivable that motion data of eyes and humans can be made comparable by stretching and compressing the spatial and temporal scales. Thus, for instance, the displacements of a person recorded over a day (between home, office, restaurant, sports club etc.) could be made to resemble to some extent the saccadic movements of eyes between fixations. Such transformations could help in exploiting the full potential of generic pattern detection procedures. However, clearly, the interpretation of the movement *behavior* (which is usually the ultimate goal of the analysis) must take place at the original scale(s) and context dictated by the application domain – and hence, the existence of an inverse transformation of scales absolutely indispensable.

So, we can reach a first conclusion: There are some commonalities between eye and human whole-body movements on the level of generic movement patterns – but only as far as they stay within the bounds of the different motion physics and scales involved. In terms of behavioral movement patterns, the two types of movement are very different.

The above discussion was purely conceptual. We are currently conducting empirical research to test our hypothesis in a more systematic way considering computational and visual data mining methods in pattern detection on the one hand, and perception and cognition experiments on the other.

Conclusions: a call for collaboration

With this paper we aim to contribute to the development of a comprehensive toolbox of data mining algorithms and visual analytics methods for movement analysis. We argue that thorough definitions for a commonly agreed set of movement patterns are indispensable as a basis for the development of pattern recognition algorithms that are required to be efficient, effective, and as generic as possible. We also believe that such work is needed to support the information visualization community in developing objective visual analytics methodologies. The generic patterns identified in our classification allow a domain-independent visualization of movement. It is therefore usable for movement researchers from various disciplines, because all generic patterns should be applicable to all moving datasets at all spatio-temporal resolutions. Bees, for instance, generate a cyclic pattern where they return to their hive, just like wolves. Both animals produce the same kind of pattern and the pattern can be visualized accordingly, although these patterns appear on different spatio-temporal resolutions. By classifying generic patterns, we also gain an advantage in identifying and exploring behavioral patterns as these consist of the generic pattern types.

This paper has contributed a conceptual framework of movement and a classification of movement patterns. The descriptions and definitions associated with the classification had to be kept short in this paper. We have therefore set up a wiki[20] for two reasons. First, the wiki hosts more background information and more detailed definitions. And even more importantly, we see this as a mechanism to facilitate the further development and consolidation of the proposed classification and pattern definitions into a complete taxonomy of movement patterns. We are convinced that generic pattern definitions must be based on the consensus of domain experts from various disciplines involved in the analysis of movement. Hence, a moderated revision process seems to have the best potential for developing a taxonomy of movement patterns that are useful on a broad range.

References

1 Blythe PW, Miller GF, Todd PM. Human simulation of adaptive behavior: interactive studies of pursuit, evasion, courtship, fighting and play. In: *Proceedings of the Fourth International Conference on Simulation of Adaptive Behavior*, MIT Press/Bradford Books: Cambridge, MA. 1996; 13–22.
2 Laube P, Purves R. An approach to evaluating motion pattern detection techniques in spatio-temporal data. *Computers, Environment and Urban Systems* 2006; **30**: 347–374.
3 Smyth CS. Mining mobile trajectories. In: Miller HJ and Han J (Eds). *Geographic Data Mining and Knowledge Discovery*. Taylor & Francis: London. 2001; 337–361.
4 Fayyad U, Piatetsky-Shapiro G, Smyth P. From data mining to knowledge discovery in databases. *AI Magazine* 1996; **17**(3): 37–54.

5 Imfeld S. Time, points and space – towards a better analysis of wildlife data. Ph.D. thesis, Department of Geography, University of Zurich, 2000.
6 Lee JW, Paek OH, Ryu KH. Temporal moving pattern mining for location-based services. *Journal of Systems and Software* 2004; **73**: 481–490.
7 Benkert M, Gudmundsson J, Hübner F, Wolle T. Reporting Flock Patterns. Lecture Notes in Computer Science, vol. 4168. Springer: Berlin. 2006; 660–671.
8 Gudmundsson J, van Kreveld MJ, Speckmann B. Efficient detection of motion patterns in spatio-temporal data sets. *Proceedings of the 12th Annual ACM International Workshop on Geographic Information Systems*, ACM: New York, NY, 2004; 250–257.
9 Laube P. Analysing point motion – spatio-temporal data mining of geospatial lifelines. Ph.D. thesis, Department of Geography, University of Zurich, 2005.
10 Laube P, Imfeld S, Weibel R. Discovering relative motion patterns in groups of moving point objects. *International Journal of Geographical Information Science* 2005; **19**(6): 639–668.
11 Andrienko N, Andrienko G. Designing visual analytics methods for massive collection of movement data. *Cartographica* 2007; **42**(2): 117–138.
12 Erwig M. Toward spatiotemporal patterns. In: Caluwe R de, Tré G de and Bordogna G (Eds). *Spatio-Temporal Databases – Flexible Querying and Reasoning*. Springer: Berlin. 2004; 29–54.
13 Gudmundsson J, Laube P, Wolle T. Movement patterns in spatiotemporal data. In: Shekhar S and Xiong H (Eds). *Encyclopedia of GIS*. Springer: Berlin. 2008;
14 Yuan M, Stewart Hornsby K. *Computation and Visualization for the Understanding Dynamics of Geographic Domains: A Research Agenda*. CRC Press: New York, NY, 2007.
15 Giannotti F and Pedreschi D (Eds). *Mobility, Data Mining and Privacy -- Geographic Knowledge Discovery*. Springer: Berlin. 2007;
16 Laube P, Dennis T, Forer P, Walker M. Movement beyond the snapshot – Dynamic analysis of geospatial lifelines Computers. *Journal of Environment and Urban Systems* 2007; **31**: 481–501.
17 Benhamou S. How to reliably estimate the tortuosity of an animal's path: straightness, sinuosity, or fractal dimension?. *Journal of Theoretical Biology* 2004; **229**: 209–220.
18 Turchin P. *Quantitative Analysis of Movement: Measuring and Modelling Population Redistribution in Animals and Plants*. Sinauer Publishers: Sunderland, MA, 1998.
19 Salvucci DD, Goldberg JH. Identifying fixations and saccades in eye-tracking protocols. In: *Proceedings of the Eye Tracking Research and Applications Symposium* (Palm Beach Gardens, FL), ACM Press: New York. 2000; 71–78.
20 URL of the Movement Pattern Wiki: http://movementpatterns.pbwiki.com/Patterns-of-Movement.
21 Laube P, Imfeld S. Analyzing relative motion within groups of trackable moving point objects. In: Egenhofer MJ and Mark DM (Eds). GIScience: Lecture Notes in Computer Science, vol. 2478. Springer: Berlin, 2002, 142–144.
22 Cao H, Mamoulis N, Cheung D. Mining frequent spatio-temporal sequential patterns. *ICDM'05: Proceedings of the fifth IEEE International Conference on Data Mining*, IEEE Computer Society, Washington, DC, vol. 8, 2005;
23 Agrawal R, Srikant R. Mining sequential patterns. In: Yu PS and Chen ASP (Eds). *Eleventh International Conference on Data Engineering*: (Taipei, Taiwan), IEEE Computer Society Press: Silver Spring, MD, 1995; 3–14.
24 Gudmundsson J, van Kreveld M. Computing longest duration flocks in trajectory data. In: *Proceedings of the 14th Annual ACM International Symposium on Advances in Geographic Information Systems (GIS '06)*: (Arlington, VA), ACM: New York, NY, 2006; 35–42.
25 Kalnis P, Mamoulis N, Bakiras S. Advances in Spatial and Temporal Databases. Lecture Notes in Computer Science, Springer: Berlin/Hidelberg. 2005.
26 Andersson M, Gudmundsson J, Laube P, Wolle T. Reporting leadership patterns among trajectories. *SAC '07: Proceedings of the 2007 ACM Symposium on Applied computing 2007*, (Seoul, Korea) ACM: New York, NY, 2007; 3–7.

Research Paper 2

Dodge, S., Weibel, R. and Forootan, E. (2009). Revealing the physics of movement: Comparing the similarity of movement characteristics of different types of moving objects. *Computers, Environment and Urban Systems*, Volume 33, Issue 6, November 2009, pages 419 – 434.

Computers, Environment and Urban Systems 33 (2009) 419–434

Contents lists available at ScienceDirect

Computers, Environment and Urban Systems

journal homepage: www.elsevier.com/locate/compenvurbsys

Revealing the physics of movement: Comparing the similarity of movement characteristics of different types of moving objects

Somayeh Dodge [a,*], Robert Weibel [a], Ehsan Forootan [b]

[a] Department of Geography, University of Zurich, Winterthurerstrasse 190, 8057 Zurich, Switzerland
[b] Institute of Biodiversity and Ecosystem Dynamics, University of Amsterdam, Nieuwe Achtergracht 166, 1018 WV Amsterdam, The Netherlands

ARTICLE INFO

Keywords:
Moving point data mining
Moving object
Movement parameters
Movement behavior
Trajectory decomposition
Trajectory classification

ABSTRACT

We propose a segmentation and feature extraction method for trajectories of moving objects. The methodology consists of three stages: trajectory data preparation; global descriptors computation; and local feature extraction. The key element is an algorithm that decomposes the profiles generated for different movement parameters (velocity, acceleration, etc.) using variations in sinuosity and deviation from the median line. Hence, the methodology enables the extraction of local movement features in addition to global ones that are essential for modeling and analyzing moving objects in applications such as trajectory classification, simulation and extraction of movement patterns. As a case study, we show how the method can be employed in classifying trajectory data generated by unknown moving objects and assigning them to known types of moving objects, whose movement characteristics have been previously learned. We have conducted a series of experiments that provide evidence about the similarities and differences that exist among different types of moving objects. The experiments show that the methodology can be successfully applied in automatic transport mode detection. It is also shown that eye-movement data cannot be successfully used as a proxy of full-body movement of humans, or vehicles.

1. Introduction

The analysis of trajectories of moving objects has recently become the focus of many research projects in the area of geographic information science (GIS), human–computer interaction (HCI), ecology, biology, social and behavioral sciences. Simulating human and animal mobility behavior, or studying human interaction with computers are emerging into an interesting area of research, which requires extracting knowledge about the dynamic behavior of different types of agents and thus challenges developing new exploratory data analysis methods on massive movement datasets. Therefore, many spatio-temporal data mining algorithms and analytical methods have been proposed at the theoretical level, however few of them have been implemented and applied in practice to date.

A critical success factor for empirically based research is the availability of relevant data. The main problem is that data about moving point objects (MPOs) are not easily available and accessible due to data cost, security and privacy issues (Giannotti & Pedreschi, 2007). In order to overcome the problem of data scarcity, one may consider utilizing data that can act as a proxy of 'physical' movement data or benefit from artificial, simulated movement data (Blythe, Miller, & Todd, 1996). For instance, bank note dispersals can be considered as a proxy for human movement given that money is carried by individuals (González, Hidalgo, & Barabasi, 2008), or mouse movement traces as a proxy of eye-movement data in HCI studies (Chen, Anderson, & Sohn, 2001; Cox & Silva, 2006). Similarly, eye-movement data from human subject experiments on graphic displays is potentially of interest to be used as a proxy of other types of moving objects, as it is relatively inexpensive to collect and usually not subject to particular privacy issues.

By the same token, the simulation of trajectories is used for diverse purposes, such as ecological modeling (Turchin, 1998), spatio-temporal database research (Pfoser & Theodoridis, 2003), agent-based pedestrian modeling (Batty, 2003), and in the evaluation of data mining algorithms (Laube & Purves, 2006). Therefore, detailed knowledge of the movement parameters of different MPOs is crucial in choosing the best representative proxy in trajectory simulation. The better the knowledge about the movement behavior of the particular objects that is simulated, the more realistic the simulation results will be. However, there are still some open research questions in the field of modeling and simulating trajectories of moving objects. For instance, how can we efficiently assess the similarity of the behavior of the simulated or proxy data in comparison to the original moving object? Is it possible to automatically identify trajectories of unknown objects by applying our knowledge about the movement behavior of similar known objects

* Corresponding author. Tel.: + 41 446355131; fax: +41 446356848.
E-mail addresses: somayeh.dodge@geo.uzh.ch, somayeh.dodge@gmail.com (S. Dodge), robert.weibel@geo.uzh.ch (R. Weibel), e.forootan@uva.nl (E. Forootan).

0198-9715/$ - see front matter © 2009 Elsevier Ltd. All rights reserved.
doi:10.1016/j.compenvurbsys.2009.07.008

whose movement characteristics have been previously learned by the system?

The above issues all point to a need for methods for analyzing the movement behavior of different MPOs, with the aim of determining the similarity of trajectories generated by different MPOs. Similarity search, that is, trying to find similar trajectories of moving objects, is a fairly new topic in spatial data mining. Most of the techniques proposed to date are looking for similarities of the geometric shape of the trajectories based on a distance function. Examples include the Edit Distance on Real sequence (EDR) (Chen, Özsu, & Oria, 2005), One-Way Distance (OWD) (Lin & Su, 2008), Euclidean and Time Wrapping distance and Longest Common Subsequence (LCSS) (Vlachos, Gunopulos, & Kollios, 2002). However, we are more interested in finding similarities in movement behavior of *different types* of moving objects. Therefore, our motivation is to take an analytical look at the movement characteristics and dynamic behavior of different types of dynamic objects such as humans, vehicles and eye movements and extract possible similarities among movement behavior of such objects. Consequently, we want to see whether we can predict the types of unknown MPOs by similarity to the trajectories of previously learned MPOs.

This article thus presents a methodology that allows extracting movement parameters from the trajectories of different types of moving objects. The key element of the methodology is an algorithm that decomposes the profiles generated for different movement parameters using variations in sinuosity and deviation from the median line, hence enabling the extraction of local movement features in addition to global ones.

Our proposed methodology is useful in several respects. It can inform developers of pattern recognition and data mining algorithms about similar and dissimilar types of moving objects, hence allowing to design rigorous algorithm evaluation strategies. It can help answer the question how similar simulated or proxy MPOs are to the corresponding reference MPOs. The proposed trajectory segmentation algorithm yields sub-trajectories that can facilitate similarity search. The methodology generates relevant movement attributes at the global level of the entire trajectory as well as at the local level of segments of homogeneous movement characteristics, enabling more differentiated parameterization of trajectory simulations. Thus, it can be used to answer to the above-mentioned research questions in simulation studies. And finally, it can be used to classify unknown moving objects into previously learned MPO types, in data mining operations on large trajectory databases or in real-time applications. For instance, it can be used in transportation research to detect the transport mode in anonymized trajectories of different transportation objects (e.g. cars, motorcycles, bicycles, pedestrian).

The remainder of the paper is organized as follows. We start in Section 2 with a brief introduction of moving point objects and a review of the relevant literature. We continue in Section 3 by explaining the proposed methodology for feature extraction of movement parameters. In Section 4, we propound some possible applications of the proposed methodology. In Section 5, we report the experiments conducted to validate the three steps of the methodology following the classification process. Section 6 provides a detailed discussion of the experimental results. We end in Section 7 with conclusions and an outlook.

2. Moving point objects (MPO)

We define moving objects as entities whose positions or geometric attributes change over time. In many applications moving objects are considered as moving points, ignoring the dimension of the object. In (Dodge, Weibel, & Lautenschütz, 2008), moving objects are categorized into two major groups of geo-referenced (i.e. dynamic objects that move about in geographic space) and non-geo-referenced (i.e. dynamic phenomena that move in a non-geographic space) dynamic objects. Accordingly, geographically referenced object such as humans, animals or vehicles belong to the first group, while gaze point movements in eye-movement studies can be mentioned as an example for the other group. Each of these dynamic objects, to a varying degree, shares some similarities but also exhibits differences to the others in terms of the corresponding data structure, dynamic behavior and nature of movement.

In general, the path of a moving object, named trajectory, is the subject of interest in moving object data analysis. A trajectory is defined as a sequence of successive positions of the moving object over a period of time and thus can be considered as a time series of spatial data in data mining tasks (Spaccapietra et al., 2008). In order to analyze or simulate the behavior of a moving object we need to have detailed information about the trajectory of the object as well as information about the environmental conditions related to the trajectory (Spaccapietra et al., 2008). In other words, it is necessary to extract differentiated movement parameters of a trajectory in order to analyze or simulate typical movement behavior of an object. In this regard many attempts have recently been carried out in the field of modeling and analyzing trajectories and moving object data mining. Giannotti and Pedreschi (2007) give an overview of the history of analyzing moving objects from the initial idea of time geography to the recent advances in knowledge discovery from moving objects using spatio-temporal data mining techniques, including latest attempts on data privacy and security issues. Batty (2003) applied agent-based modeling of individual and collective behavior of pedestrians to show how randomness and geometry are important to local movement and how individuals respond to locational patterns. Brillinger, Preisler, Ager, and Kie (2004) developed a stochastic differential equation-based model for exploratory data analysis of the trajectories of deer and elk to describe movement behavior of free-ranging animals. They tried to extract typical parameters of data obtained from animal telemetry studies. Laube and Purves (2006) considered modeling relative movement within groups of objects in order to evaluate extracted movement patterns by simulation through correlated random walk procedures. Hornsby and Cole (2007) focused on modeling moving objects from an event-based perspective and tried to detect movement patterns by analysis of different events. Other researchers have focused on differentiating and modeling moving objects in movement imagery databases, in order to describe and classify behavior of moving objects in computer vision systems using sequences of images (Agouris, Partsinevelos, & Stefanidis, 2003; Ozyildiz, Krahnstöver, & Sharma, 2002; Zheng, Dagan Feng, & Zhao, 2005). In Naftel and Khalid (2006) another approach for clustering and classification of object trajectory-based video clips using spatio-temporal function approximation has been proposed. Bashir, Khokhar, and Schonfeld (2007) present a classification algorithm for recognizing object activity using trajectory of objects. In the proposed classification method, trajectories are segmented at points of change in curvature and the sub-trajectories are represented by their principal component analysis (PCA) coefficients (Bashir et al., 2007). In Bay and Pazzani (2001) a search algorithm for mining contrast sets has been developed to differentiate between several contrasting groups (e.g. male or female students, or the same group over time) from observational multivariate data.

The above-mentioned modeling and classification techniques have mainly been applied on trajectories obtained from the same MPO types. Fewer studies exist on the classification and differentiation of trajectories of different kinds of moving objects. One domain where the comparison of trajectories from different moving objects is relevant is the field of transportation studies, specifically in the analysis of transport behavior in urban environment. In this

domain some researchers focused on extracting knowledge from raw GPS data to detect the mode of transport that people used, with the aim of understanding user behavior (Zheng, Liu, Wang, & Xie, 2008). For instance, Zheng et al. (2008) proposed an approach based on supervised learning to automatically learn the transportation mode, including walking, taking a bus, riding a bike and driving. Their method is comprised of a segmentation method based on change points (i.e. where the mode of transport presumably changes), an inference model (i.e. decision tree, support vector machine (SVM), Bayesian net, or conditional random field (CRF)), and a post processing method. In this study the four above-mentioned inference models have been evaluated. They show that the higher accuracy is obtained from the decision tree model. In another study, Tsui and Shalaby (2006) introduced a fuzzy logic approach. They applied a segmentation method based on three types of mode transfer points (MTP). In a similar study, Schüssler and Axhausen (2009) applied the same method based on speed and acceleration characteristics to distinguish five modes of transport (i.e. walk, bicycle, car, urban public transport, and rail). Moreover, Zheng et al. (2008) and Schüssler and Axhausen (2009) give a summary of other related research. To the best of our knowledge, almost all the proposed methods have difficulty distinguishing different transport modes in congestion or heavy traffic. They also do not seem effective in distinguishing the transport mode of vehicles with similar speed range. Finally, they appear having difficulties to detect the correct transport mode when people only take one kind of transport mode during a trip. Therefore, there is still a need for more research on more reliable approaches for transport mode detection.

In Dodge et al. (2008), Giannotti and Pedreschi (2007) and Laube, Dennis, Forer, and Walker (2007) parameters of a trajectory generated by a moving object are introduced such as speed, acceleration, duration of movement, sinuosity, traveled path, displacement, and direction. These descriptors form fundamental building blocks for characterizing the movement of an object and can be defined in an absolute sense (i.e. with respect to the external reference system) or in a relative sense, (i.e. in relation to the movement of other MPOs or to the previous states of the same MPO). Generally speaking, different types of moving objects, depending on the particular physics of their movement, to some degree exhibit different signatures of such movement descriptors. Each MPO has a typical dynamic behavior, which to some extent is similar for individuals of the same kind. Consequently, moving objects can be reproduced (simulated) according to the typical behavior of the similar sort of objects, or objects having the same dynamic behavior (Laube & Purves, 2006). Likewise, the typical behavior of different objects can be extracted from the particular parameters of their trajectories using the above-mentioned descriptors.

Therefore, we propose a methodology that allows extracting such movement parameters from the trajectories of different types of moving objects and classifying trajectories of unknown MPOs by similarity to the known trajectories. We focus on the characterization and classification of *different types* of moving objects and we conduct a comparative analysis and classification of the movement behavior of different objects, manifested through their trajectories. As a case study, we show how our model can be applied in the classification and prediction of transport mode of unknown trajectories of people using a supervised classification method. The following section describes our methodology in detail.

3. Methodology

Our methodology consists of three steps, shown graphically in Fig. 1 and expanded on in the remainder of this section: (1) trajectory data preparation; (2) global descriptors computation; and (3) local feature extraction. The products generated from applying this procedure can directly be used for other purposes, such as generating inputs for movement simulators, or trajectory classification as presented later in Section 4.

3.1. Trajectory data preparation

Raw data captured by movement tracking devices usually to some degree contain noise, outliers and gaps, depending on the nominal precision and accuracy of the tracker as well as other factors that influence the completeness, accuracy and reliability of fixes. The accuracy of GPS observations, especially in absolute positioning, is very sensitive to the existence of obstacles that block GPS signals, multi-path effects, ionospheric and tropospheric errors, etc. (Hoffmann-Wellenhof, Lichtenegger, & Collins, 2001). In kinematic GPS surveys used to generate trajectory data of the type used in this study, it seems reasonable to assume an accuracy of 5–10 m for practical purposes. Eye trackers have a higher accuracy (i.e. 0.5°) and sample eye movements at fine temporal granularity (e.g. about 20 ms). However, raw data generated by eye-trackers still contain a considerable amount of noise, outliers, and gaps, which should be remedied in order to achieve better results. Therefore, in order to remove effects of noise and positioning errors of the tracking devices and other factors, we recommend applying data cleaning and pre-processing procedures on the raw data to achieve more reliable trajectories. The pre-processing phase consists of three steps including filtering, re-sampling, and smoothing. During the filtering process outliers are removed from the raw data, namely those fixes that had a distance from the previous fix of more than three times the standard deviation (3σ) of the distances between consecutive fixes. The re-sampling procedure then generates a trajectory at regular intervals by linear interpolation along the trajectory. Finally, the smoothing step eliminates noise remaining in the data. In order to smooth raw GPS data several methods can be employed, such as least squares, spline approximation, moving average, kernel-based smoothing, and Kalman filtering (Eubank, 2005). In this regard, Jun, Guensler, and Ogle (2007) developed an analytical study of different smoothing methods and proposed a modified version of Kalman filtering to be applied for GPS data containing errors (see Section 5.2.1).

3.2. Computation of global descriptors

Movement parameters (i.e. speed, acceleration, turning angle, straightness, etc.) can be derived from the trajectory of an object and thus describe the dynamic behavior of the object. These descriptors are very different in terms of the values that they can take for each type of MPO. For instance, eyes can move quickly in fractions of a second from one end of a picture to the other in an almost mass-less movement, while the acceleration of human whole-body motion is governed by greater mass and inertia.

In order to evaluate the movement behavior inherent to the given trajectory data sets, various movement parameters can be computed for each point (fix) along a trajectory: for instance speed (i.e. rate of change of the object's position); acceleration (i.e. rate of change of the object's speed); turning angle (i.e. direction of the movement); displacement (i.e. the beeline connector distance between two consecutive points); traveled path (i.e. the path length along the trajectory); and straightness index (i.e. the ratio of the traveled path and displacement); giving an indication of the sinuosity of the trajectory at a specific point (Benhamou, 2004; Dodge et al., 2008; Laube et al., 2007).

To achieve differentiated results in the characterization of trajectories, we propose that the computation of movement parame-

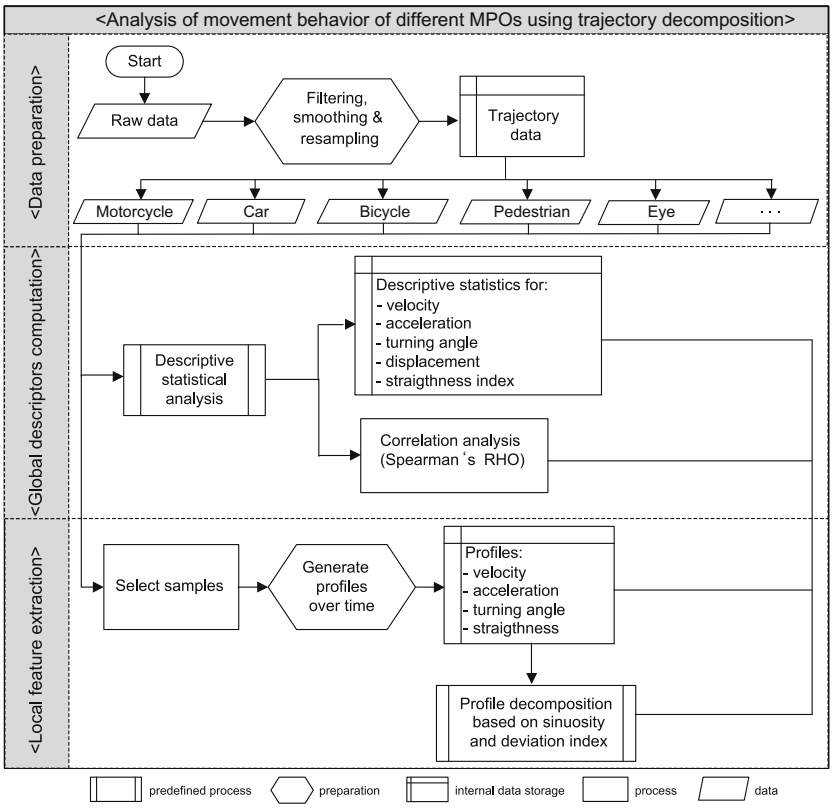


Fig. 1. Methodology for analyzing and extracting the movement behavior of different MPOs.

ters proceeds at consecutive levels of refinement. That is, the process should first take a global look, computing descriptive statistics for the entire trajectory. Then, it should zoom in to extract local information of the trajectories at finer resolutions. Finally, in order to reveal more detail in the movement behavior of the selected objects and make their trajectories comparable, we propose to decompose the computed profiles of movement parameters to a set of meaningful subsections (or segments). Sections 3.2.1 and 3.2.2 describe the computation of global descriptors; Section 3.3 describes the extraction of local movement descriptors and the profile decomposition.

3.2.1. Global descriptive statistics

In order to extract the global movement properties of a given MPO, the above-mentioned movement parameters are first derived from the entire trajectory of the object. Next, global descriptive statistics of the movement parameters are computed such as the minimum, maximum, mean, median, standard deviation, variance and skewness over the entire trajectory.

3.2.2. Correlation analysis

In order to assess potential interrelationships between movement parameters, a correlation analysis should be carried out after extracting the movement parameters of given MPOs. We recommend computing Spearman rank correlation (RHO) as a non-parametric measure of correlation, since it has the advantage of making no assumptions about the frequency distribution of the variables (Chatfield, 1989). It is used to test the direction and strength of the relationship between variables. High correlations between movement parameters suggest that some variables may be redundant.

3.3. Local feature extraction: profile decomposition

When a dynamic object moves about in space, its movement parameters (velocity, acceleration, turning angle, etc.) change over time. If we plot the evolution of a movement parameter over time, this will result in a profile or function, such as the one shown in Fig. 2. If we do this for different dynamic objects the resulting profiles will exhibit different amplitude and frequency variations, hence giving clues to the underlying movement physics and behavior. This has lead us to using the movement parameter profiles for extracting local features that could be used for trajectory simulation and classification, by decomposing profiles into segments (or sections) of 'similar movement character'. We propose to use two measures for characterizing movement from profiles: *deviation* from the median line of the profile gives an impression of the amplitude variation of a movement parameter over time, while *sinuosity* acts as a proxy of the frequency variation. In the following, we describe the computation of the deviation measure and the sinuosity measure that we use, as well as the proposed algorithm for profile decomposition. Fig. 2 provides supporting graphical illustrations and Algorithm 1 gives the pseudo-code of the profile decomposition algorithm.

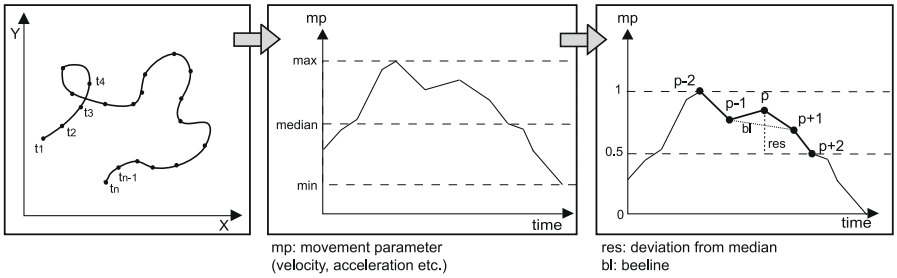


Fig. 2. Basic elements of movement parameter profiles.

Algorithm 1. Profile decomposition.

Inputs:
- *res*[]: residuals from median
- *st*: threshold to distinguish low sinuosity from high sinuosity

Outputs:
- *decomX*[]: classified and decomposed profile

Algorithm:

```
n ← the number of points on the profile
dtime ← time interval between consecutive points
for index of points i = 1–n do
    dres ← res_{i+1} − res_i
    sl[i] ← sqrt(dtime^2 + dres^2)
end for
dt ← standard deviation of res[]
sinuosity[] ← 0
for lag size k = 1–2 do
  for index of points i = (1 + k) to (n − 1 − k) do
    beeline_distance ← length of beeline connector of p_{i-k} and p_{i+k}
    profile_distance ← ∑sl of p_{i-k} to p_{i+k}
    sinuosity[i] ← profile_distance/beeline_distance + sinuosity[i]
  end for
end for
for index of points i = 1–n do
sinuosity[i] ← sinuosity[i]/2
sin_scaled ← scale sinuosity to the length of 1
if (sin_scaled < st) AND (res[i] < dt) then
decomX[i] ← 1   /* low sinuosity, low deviation
  elseif (sin_scaled ⩾ st) AND (res[i] < dt) then
decomX[i] ← 2   /* high sinuosity, low deviation
    elseif (sin_scaled < st) AND (res[i] ⩾ dt) then
decomX[i] ← 3   /* low sinuosity, high deviation
  elseif (sin_scaled ⩾ st) AND (res[i] ⩾ dt) then
decomX[i] ← 4   /* high sinuosity, high deviation
end if
end for
return decomX[]
```

Both deviation and sinuosity are defined for each point on a movement parameter profile. Before we compute these measures, we transform the profile data in the following way. First, we calculate the median of the particular movement parameter that was used to generate the profile. This median then can be seen to form a horizontal 'median line' that separates the movement parameter values into two halves. We then take the residuals from the median for each point along the original profile. And finally, in order to make the comparison across objects possible, we normalize all movement parameter profiles to a common interval [0, 1], as shown, for instance, in Fig. 2.

The deviation of a point p on a profile is easily established: it simply equates to its residual value from the median and has thus already been obtained when the residuals were calculated above. The measure of sinuosity for p is computed as a ratio of the distance $\pm k$ points along the profile to the length of the beeline connector centered at p, as follows:

$$Sinuosity_{p,k} = \frac{\sum_{i=p-k}^{i=p+k-1} d_{i,i+1}}{d_{p-k,p+k}}$$

where k is the lag parameter. This method was originally introduced by Dutton (1999) in order to classify the sinuosity of cartographic lines in map generalization. After some experimentation, in order to obtain a more reliable measure for the sinuosity, we considered both 1 and 2 for k as the lag value. Then, the final sinuosity at p is computed as the average of the $Sinuosity_{p,1}$ and $Sinuosity_{p,2}$:

$$Sinuosity_p = \frac{\sum_{k=1}^{k=2} Sinuosity_{p,k}}{2}$$

The sinuosity measure ranges from 1 (if profile points are collinear about the given point p) to infinity for a winding profile (i.e. a space-filling curve). The sinuosity values for all points are then transformed to the interval [0, 1], as proposed by (Dutton, 1999). Next, the profile points are classified into two regimes regarding the level of the corresponding sinuosity measure, *'low sinuosity'* and *'high sinuosity'*, separated by a user-defined threshold. The same is done with deviation, where the standard deviation of the residuals is used to separate '*low deviation*' from '*high deviation*'. The described procedure is summarized in Algorithm 1. The classified profile decomposes trajectory into the segments of homogeneous movement characteristics. The results of employing the Algorithm 1 on different movement parameter profiles (i.e. velocity, acceleration, etc.) can be used to compute local movement features for trajectory classification and simulation purpose.

4. Applications

We suggest that the above methodology, and in particular the trajectory decomposition algorithm, are useful for a variety of applications in movement data mining where finding similarities between the physical movement behavior of different objects is important. These include applications such as trajectory classification (e.g. transport mode detection in mobility analysis studies), movement pattern detection (e.g. fixation and saccade detection in eye-tracking research), and trajectory simulation (e.g. in human mobility behavior studies).

In the remainder of this section, we introduce a procedure for trajectory classification. In Section 5, we examine the applicability of the proposed methods in a series of classification experiments using transportation data as well as in fixation detection in eye-tracking data.

4.1. Trajectory classification

We are trying to classify trajectories of moving objects in a systematic way using the features (i.e. variables) extracted by the trajectory decomposition algorithm described above. This procedure aims at classifying trajectory data generated by unknown moving objects and assigning them to known types of moving objects, whose movement characteristics have been previously extracted and learned. That is, we are assuming to use a supervised classification algorithm. We are interested to find out whether trajectories of different kinds of MPOs can be classified distinctively. The following subsections introduce our trajectory classification process as shown in Fig. 3, which consists of two main steps: (1) Feature selection (i.e. choosing the variables that provide the input to the classification process) and dimension reduction using principal component analysis and (2) the actual classification using the support vector machine (SVM) classifier algorithm.

4.2. Feature selection and dimension reduction

A great number of global and local statistical descriptors can be computed for each trajectory. Each of these variables can potentially be selected as features for use in the classification process. However, as many of these features essentially describe similar characteristics, there are likely to exist correlations, suggesting that only a reduced set of features should in fact be used in the classification. Given the large number of global and local descriptors it would be very difficult to reduce the original set of features by correlation analysis, merely selecting a subset of the original features. Hence, we propose using principal component analysis (PCA) for reducing the number of original features, and hence dimensions in the feature space (Bozdogan, 2003; Guyon & Andre, 2006; Smith, 2002). PCA yields a (sub)set of synthetic, uncorrelated features called principal components, which contain the most important aspects of the original features.

4.3. Classification using SVM

The features that have been generated by the PCA for each MPO type are considered as a set of attribute categories that form the input for the final step of the classification procedure. This step has the aim of classifying trajectories by assigning them to different types of moving objects. Essentially, we are interested in two aspects. First, we would like to see whether it is possible to tell apart, that is, to discriminate the trajectories generated by different types of moving objects based on the movement parameters that we have extracted from the trajectory data. Second, assuming that this is possible, we are interested in classifying dynamic objects of unknown type to the correct object type, that is, we would like to be able to reveal the identity of unknown objects. For instance, in transportation studies analysts are interested in detecting different modes of transport from unknown GPS trajectories of people.

Given the latter objective, it is advisable to use a supervised classification method where a training (or learning) stage is followed by a classification (or testing) stage that applies the learned discriminating functions to classify the unknown objects. In principle, any supervised classification technique could serve our purposes, but we chose to use the support vector machine (SVM) approach (Cristianini & Shawe-Taylor, 2000; Duda, Hart, & Stork, 2001), which is widely used today in pattern recognition and data mining. The trajectory classification process then consists of the training stage where the SVM will learn from a set of trajectory samples (the training set) how to discriminate between MPO types by constructing separating hyperplanes in the multi-dimensional space formed by the input features; and a classification/testing stage that applies the learned hyperplanes on another set of trajec-

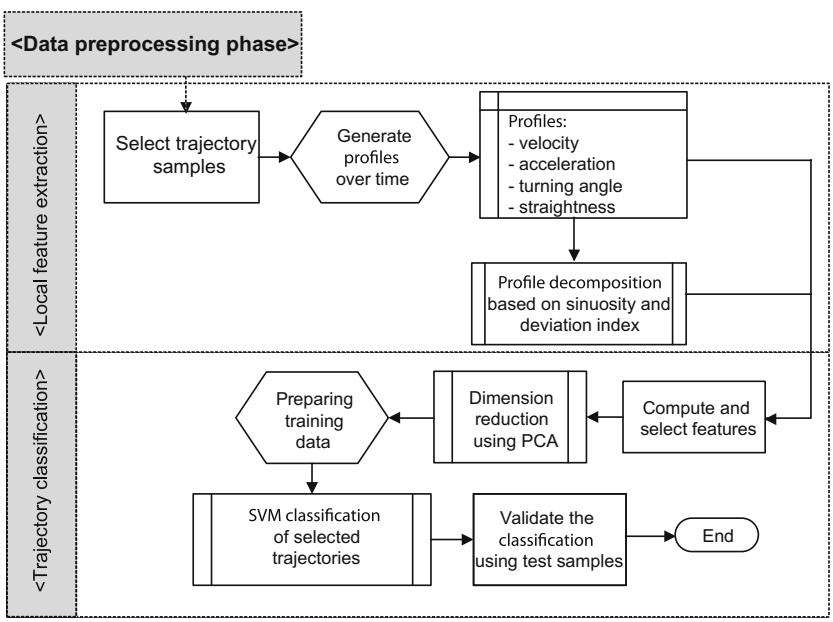


Fig. 3. Trajectory classification process.

tory samples (the testing set), thus predicting the object type of each of these unknown trajectories.

This step concludes our proposed overall methodology. After the SVM has been trained and validated, it is now ready for use in data mining operations to detect the MPO type of unknown dynamic objects from their trajectories. This could either take place off-line on large trajectory databases or in real-time.

5. Experiments: trajectory classification

In order to validate our methodology and demonstrate its applicability in the classification of trajectories of different MPOs, we have conducted a series of experiments that will be reported in this section and discussed in the next section. The experiments are designed to specifically investigate: (1) automatic mode detection in transportation analysis and (2) feasibility study of using eye-tracking data as a proxy for other MPOs. For these experiments, we considered different types of MPOs with varying physics and behavior of movement, expressed through different movement parameters (Dodge et al., 2008). We have therefore selected different samples of moving objects from both groups of dynamic objects introduced in Section 2. From the first group we have chosen movement data captured from pedestrians, bicycles, cars and motorcycles; from the second group we considered eye-movement data. Among these data, bicycles, motorcycles, and cars and to a lesser degree pedestrian movements are typically constrained to the transportation network.

5.1. Experiments – objective

5.1.1. Automatic transport mode detection

Two experiments were designed to validate the applicability of the proposed methodology using a supervised classification technique, with the aim of automatically assigning the correct transport mode to trajectories of unknown objects, after training with a sample of known objects.

5.1.1.1. Experiment #1: classification of objects of different speed range. For this experiment, we acquired various trajectories from openstreetmap.org of known object sources from the transportation domain, including tracks of pedestrians, bicycles, cars and motorcycles. Fig. 4 illustrates the 2-D plot of exemplar trajectories generated by the four object types. For each object type about 50,000 GPS fixes from 10 trajectories remained after data cleaning, filtering and re-sampling to a temporal sampling rate of 1 s.

Movements of different vehicles and pedestrians are performed at different ranges of speed. Therefore, classifying objects by simply taking the different speed range might seem as a straightforward solution. However, note that speed cannot be considered as the only parameter to classify objects in transportation since during rush hour all vehicles move at similar low speed. Therefore the proposed classification process takes variations and frequencies of changes of the other movement parameters (e.g. acceleration) into account, besides speed variations.

5.1.1.2. Experiment #2: classification of objects of similar speed range. This experiment aims to investigate detecting the transport mode of trajectories collected from objects of similar speed range, exemplified by cars and motorcycles. As mentioned earlier, speed plays an important role in simulating and classifying trajectories representing different object types. However, when the speed range is similar it is indispensable to inspect distinct variations of other movement parameters such as acceleration and also examine speed variations at finer detail, in order to be able to differentiate between object types. Therefore, this experiment is intended to demonstrate that the proposed classification process is sufficiently subtle to be able to classify trajectories obtained from very similarly behaving objects.

5.1.2. Using eye-tracking data as a proxy of other MPOs

5.1.2.1. Experiment #3: classifying trajectories of eyes vs. other object (non-eye). With this experiment we aimed to assess the suitability of eye-tracking data as a proxy of other types of moving objects. For this experiment, similar to the previous experiments, we classified eye-tracking data collected from an eye tracker against the data used in the first experiment. We intended to investigate whether it is possible to analytically tell apart trajectories generated by eye movement from those of other objects such as motorcycles, cars, bicycles and pedestrians that we subsume under the term "non-eye" objects. Specifically, we were interested to see whether it is feasible to use eye-tracking data in order to simulate other moving objects due to accessibility, privacy and data cost issues.

The eye-movement data set used here (Fig. 5) was contributed by Arzu Çöltekin (Eye Movement Laboratory, Department of Geography, University of Zurich) and consists of about 50,000 gaze points from two eye movement trajectories captured by a Tobii eye tracker at an interval of 16 ms during experiments on a 1600×1200 screen.

5.2. Experiments – workflow

For the three experiments we pursued our proposed 3-step methodology described in Section 3 followed by an additional phase of trajectory classification suggested in Section 4.1. The workflow of the three experiments is described in the following subsection in more detail.

5.2.1. Trajectory data preparation

First, the raw movement data were cleaned in order to remove outliers. In the case of eye-movement data, points that lay off the screen were considered as outliers and removed. The data were then re-sampled to a regular time interval, equal to the minimum sampling rate of the raw data (16 ms for eye-movement data and 1 s for the other objects). In order to fill gaps linear interpolation was used, as the underlying movement geometry didn't suggest the use of a more elaborate interpolation technique. Finally, we applied moving average smoothing (window size of 5 s) on the filtered, re-sampled data. For eye-movement data, only the filtering and re-sampling steps were applied. The reasons for not applying smoothing are the prevention of data loss and the potential creation of artifacts, as these types of trajectories exhibit a 'jagged' geometry that might be destroyed by the regularizing effect of trajectory smoothing. In the next step, from the entire dataset we selected our sample trajectories, each with a length of 300 points (i.e. with a duration of 5 min for the transportation objects). All the sample trajectories were taken from various overland roads and were visually checked to be consistent and to largely homogeneous in terms of their path geometry to prevent artifacts in the results of the trajectory classification. However, in the case of eye-tracking data it is impossible to avoid having 'jagged' geometries, as described earlier. Finally, the selected sample trajectories served as input data for the experiments.

In our study, we initially experimented with two methods for smoothing of raw GPS data, Kalman filtering (Eubank, 2005) and moving average smoothing. Both methods yielded similar results for our data, seemingly contradicting the results reported in Jun et al. (2007). However, the GPS data obtained from openstreetmap.org were captured by devices of unknown accuracy. Kalman filtering requires a model of movement, and not having solid

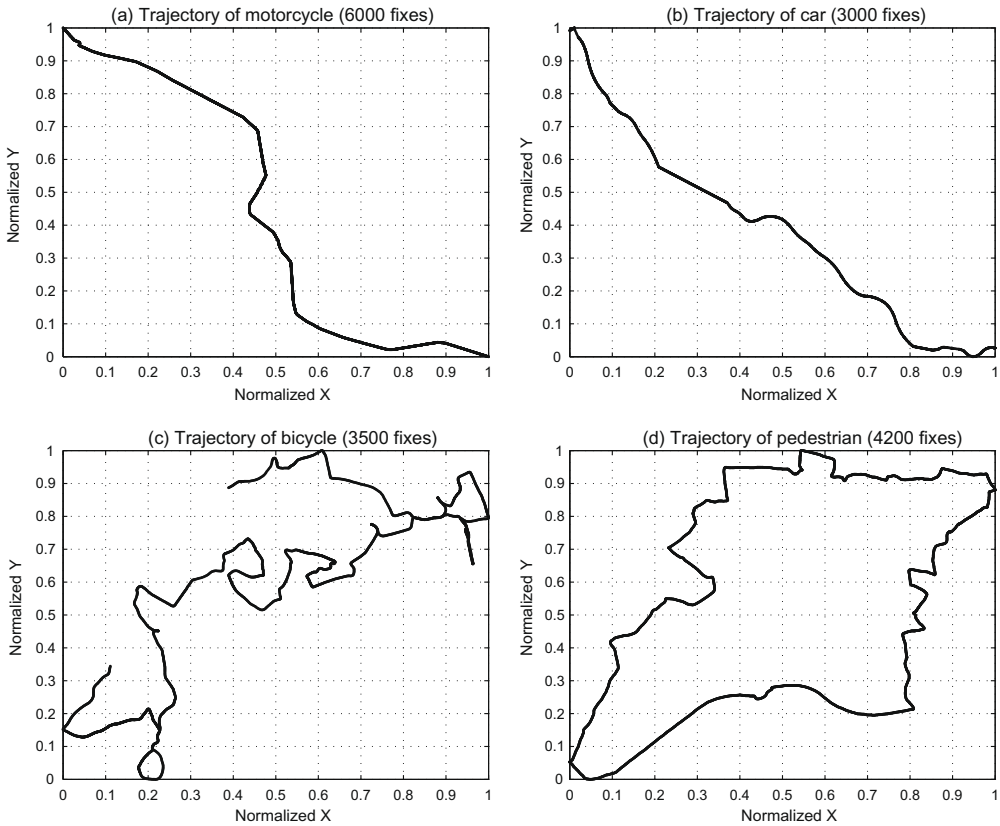


Fig. 4. Normalized trajectory data of exemplar moving objects.

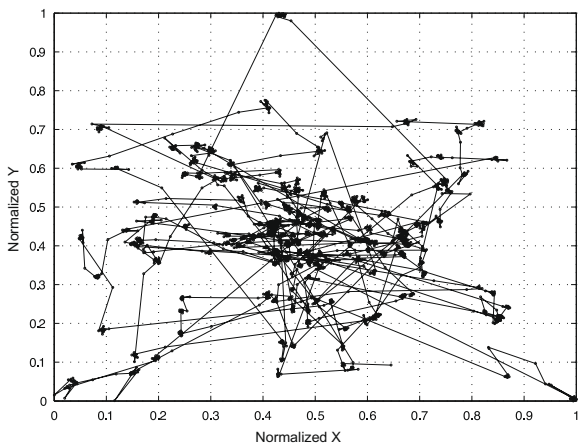


Fig. 5. Normalized sample trajectory of eye movement.

knowledge available about the movement of the objects under study has probably seriously impacted on the performance of this smoothing method. Further experiments indicated that Kalman filtering does indeed generate superior results when more accurate data are available, confirming the findings of Jun et al. (2007). Eventually, however, for reasons of practicability, we chose to use moving average smoothing, which is a reasonable smoothing method in the spatial domain.

5.2.2. Global descriptors

As mentioned before, Figs. 4 and 5 illustrate the 2-D plots of the trajectories of selected objects. From this figure it becomes obvious that the trajectory of the motorcycle (Fig. 4a), car (Fig. 4 b), bicycle (Fig. 4c), and pedestrian (Fig. 4d), are much smoother than the trajectories of eye movement (Fig. 5). Of course, temporal granularity of the sampling will influence the smoothness and length of the traveled path. For instance, the overall character of the car and motorcycle movement captured every second appears smoother and closer to the pedestrian and bicycle movement. However, with a lower sampling rate (e.g. every hours) the trajectory of the car and motorcycle movement to some degree would be probably closer to the eye movement captured every few milliseconds. Tables 1–3 present the descriptive statistics for the straightness index, velocity and displacement from the previous fix (or step length) as some examples of the movement parameters that were computed for the trajectories of the selected objects of Figs. 4 and 5.

5.2.3. Correlation analysis

For the four selected MPOs, Table 4 presents the results for the Spearman rank correlation coefficients for different pairs of movement variables. The straightness index is not used because it is a compound index using displacement. The results suggest a strong positive correlation between velocity and displacement from the previous fix for all studied objects. Moreover, there is no correlation identified between acceleration and turning angle for the selected objects. Outcomes show a negative weak correlation between velocity and turning angle for car, motorcycle, pedestrian and bicycle movement. However, for eye movement almost no correlation occurs (Table 4).

5.2.4. Locally extracted features

We generated movement parameter profiles for velocity, acceleration, turning angle, and straightness index for our trajectory data. Using Algorithm 1 we then decomposed the profiles into the four classes foreseen in the algorithm. After some initial experiments, we found threshold values that yielded consistent results over all trajectory samples. For sinuosity, we have set the threshold separating low from high sinuosity at 0.95. For deviation, we use the standard deviation of the residuals of a particular profile.

The results of the decomposition of the movement parameter profiles for four of the trajectory samples are depicted in Figs. 6 and 7. Fig. 6 illustrates the results of the decomposition process on a sample trajectory of a motorcycle on the left and a sample trajectory of a car on the right (from experiments #1 and #2). Similarly, Fig. 7 shows the results of the decomposition process on a sample trajectory of a bicycle on the left and a sample trajectory of eye movement on the right (from experiment #3). In order to save space, we do not visualize the sample result of the decomposition of a pedestrian trajectory, which looks very similar to the result for the bicycle trajectory. However, as mentioned earlier trajectory samples of pedestrians have been included in experiments #1 and #3. The individual graphs in Figs. 6 and 7 represent the normalized profiles of velocity (Figs. 6b and 7b) and acceleration (Figs. 6c and 7c), respectively. At the bottom of each graph the four decomposition classes are shown as follows:

- Green: low sinuosity – low deviation.
- Blue: high sinuosity – low deviation.
- Red: low sinuosity – high deviation.
- Magenta: high sinuosity – high deviation.

The above results form the input for the remaining steps and will be discussed in Section 6.

5.2.5. Feature selection and PCA

In our experiments, we selected a total set of 58 features from the movement parameters previously extracted on the global and local level from the trajectories, as summarized in Table 5. Following the correlation analysis conducted previously, we excluded displacement from the selection of features, as it correlates highly with velocity. From the global parameters, we further excluded turning angle, because it does not help to differentiate between objects. Consequently, we used three movement parameters (i.e. straightness index, velocity, and acceleration) to compute the mean and standard deviation at the global level, resulting in six selected global features (Table 5, top row).

The set of local features obtained from the four movement parameter profiles shown in Section 5.2.4 is made up of the mean and standard deviation of the segment length per decomposition class and per descriptor (resulting in 32 features); the number of changes of decomposition classes along the profile, computed for each descriptor (4 features); and the percentage that each decomposition class holds from the total number of points, per descriptor (16 features).

The above selected 58 features were input to a PCA to form uncorrelated linear combinations of the original features. Consequently, the number of features was reduced to 15 principal components for experiments #1 and #2 and 11 principal components for experiment #3, which formed the input for the trajectory

Table 1
Descriptive statistics for straightness index.

MPO	Min	Max	Mean	Median	Stddev	Skewness
Motorcycle	1.42	1.60	1.5	1.5	0.02	0.52
Car	1.48	1.60	1.49	1.49	0.11	8.21
Bicycle	1.07	3.3	1.5	1.5	0.08	4.21
Pedestrian	1.03	5.8	1.5	1.5	0.16	14.40
Eye	1	3141.6	8.77	2.60	89.69	26.99

Table 2
Descriptive statistics for velocity (eyes: [pixel/ms], other MPOs: [m/s]).

MPO	Min	Max	Mean	Median	Stddev	Skewness
Motorcycle	0	35.13	31.12	32.8	4.94	−3.11
Car	0	33.49	33.03	31.04	3.13	−3.04
Bicycle	0	15	5.29	5.18	2.29	0.5
Pedestrian	0	2.5	1.65	1.68	0.29	−1.97
Eye	0	20	1.18	0.48	2.36	4.13

Table 3
Descriptive statistics for displacement from the previous state (eyes: [pixel], and other MPOs [m]).

MPO	Min	Max	Mean	Median	Stddev	Skewness
Motorcycle	0	34.08	29.34	32.18	6.52	−1.94
Car	0	32.83	29.39	30.75	3.88	−2.77
Bicycle	0	17	3.34	2.69	2.48	3.34
Pedestrian	0	2.2	1.17	1.26	0.4	1.17
Eye	0	950	15.29	4.63	46.46	15.29

Table 4
Spearman rank correlation coefficients.

Correlation	Motorcycle	Car	Bicycle	Pedestrian	Eye
Velocity–acceleration	0.065	0.016	0.07	0.23	0.36
Velocity–turning angle	−0.38	−0.25	−0.25	−0.13	−0.06
Velocity–displacement	0.99	1	1	1	0.99
Acceleration–turning angle	−0.1	0.002	0.02	0.01	0.06
Acceleration–displacement	0.065	0.016	0.07	0.23	0.36
Displacement–turning angle	0.38	−0.25	0.25	−0.12	0.06

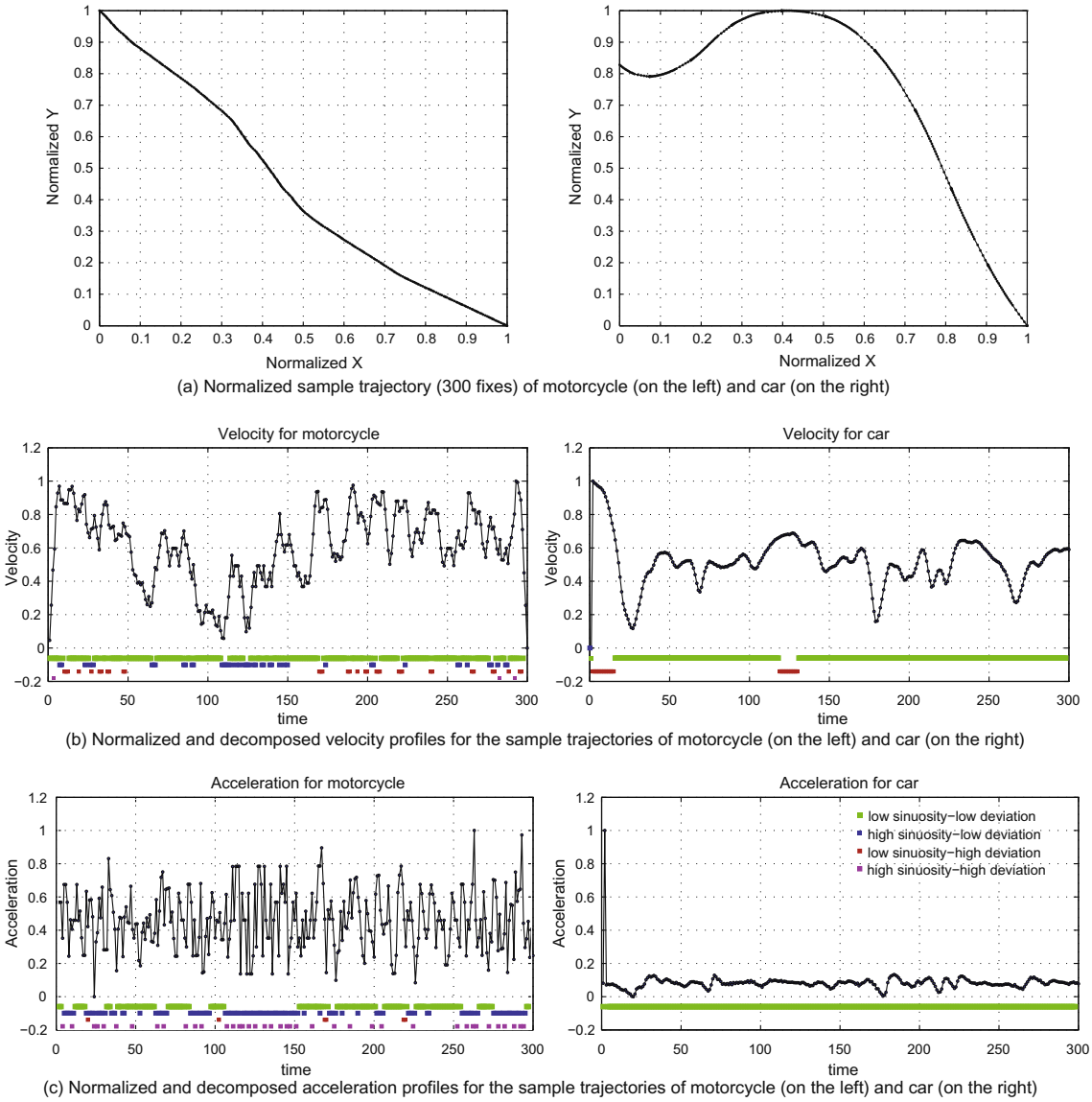


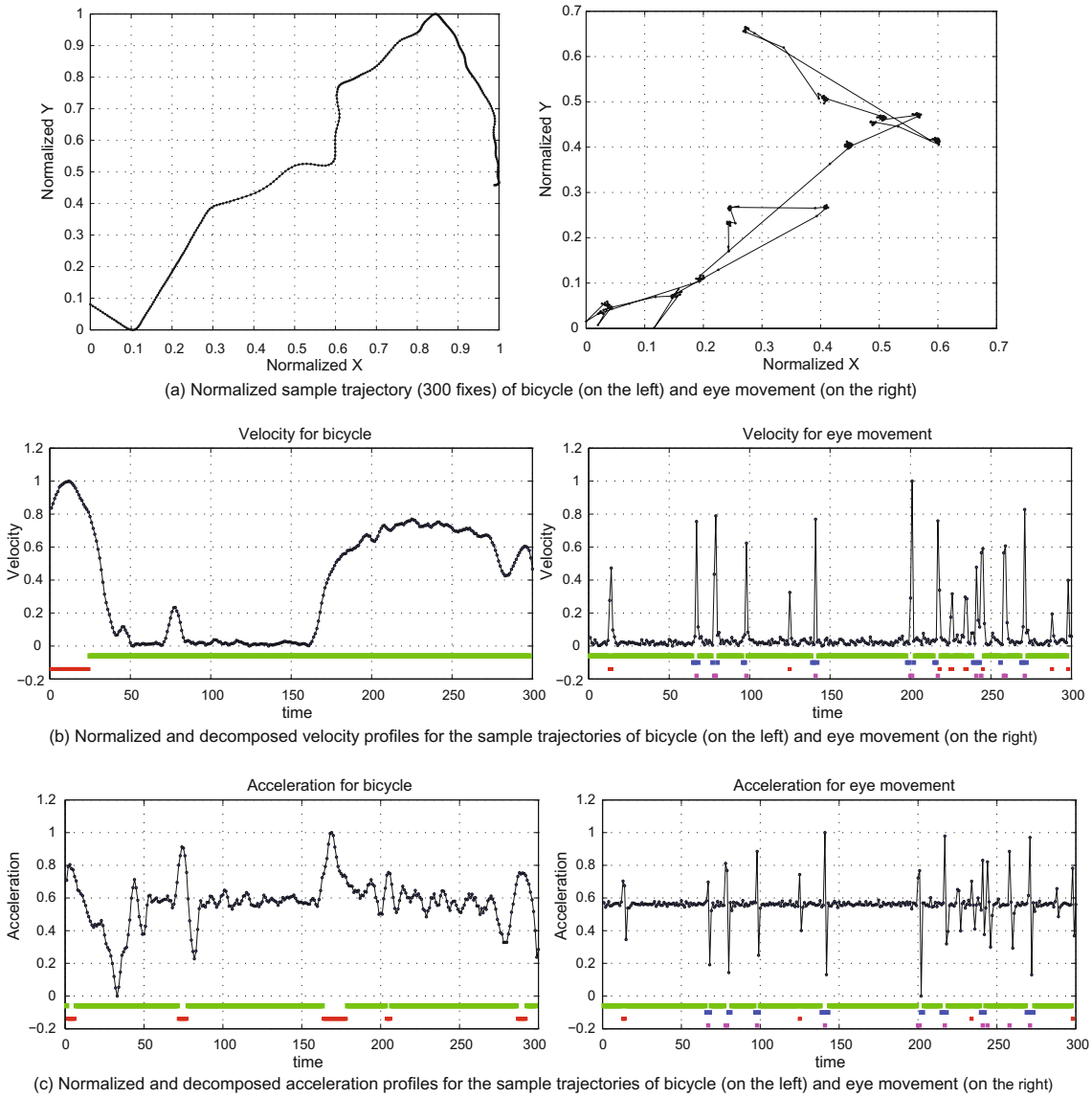
Fig. 6. Normalized and decomposed velocity and acceleration profiles for the sample trajectories of motorcycle and car.

classification step. Fig. 8 visualizes the 3-D plots of the first three principal components for the sample trajectories of the different objects used in these three experiments.

5.3. Experiments – results

For the classification stage of the proposed methodology, we randomly selected 165 samples of stretches consisting of 300 points from the various trajectories introduced in Section 5.1. 115 samples from eye movement trajectories, 165 from motorcycle trajectories, 165 from car trajectories, 165 from bicycle trajectories, and 165 from pedestrian trajectories. We then ran the decomposition algorithm for all the samples to compute the corresponding global and local movement properties. Three experiments were then conducted to evaluate the trajectory classification procedure.

The main objective of experiments #1 and #2 was to evaluate whether the proposed methodology could be applied in automatic detection of transportation mode. For experiment #1, we used 560 trajectory samples from the four pools of motorcycle, car, pedestrian and bicycle trajectories as a training set for SVM learning (i.e. 4×140 samples). The remaining 100 samples from the four pools (i.e. 4×25 samples) were used as a testing set to evaluate the performance of the classification. The aim of this experiment was to evaluate how well the different types of transportation MPOs could be differentiated using the proposed methodology in



(a) Normalized sample trajectory (300 fixes) of bicycle (on the left) and eye movement (on the right)

(b) Normalized and decomposed velocity profiles for the sample trajectories of bicycle (on the left) and eye movement (on the right)

(c) Normalized and decomposed acceleration profiles for the sample trajectories of bicycle (on the left) and eye movement (on the right)

Fig. 7. Normalized and decomposed velocity and acceleration profiles for the sample trajectories of bicycle and eye movement.

Table 5
Original features selected for the classification.

Descriptors		# of descriptors
Global	Mean and stddev at global level, per movement parameter (3)	6
Local	Mean and stddev of segment length, per decomposition class (4), per movement parameter (4)	32
	Number of decomposition class changes, per movement parameter (4)	4
	Percentage of each decomposition class (4), per movement parameter (4)	16

a multi-class classification mode. Conversely, experiment #2 had the objective of assessing a two-class classification. For this experiment, we used 280 trajectory samples from the two pools of motorcycle and car trajectories as a training set for SVM learning (i.e. 2×140 samples). The remaining 50 samples from the two pools (i.e. 2×25 samples) were used as a testing set to evaluate the

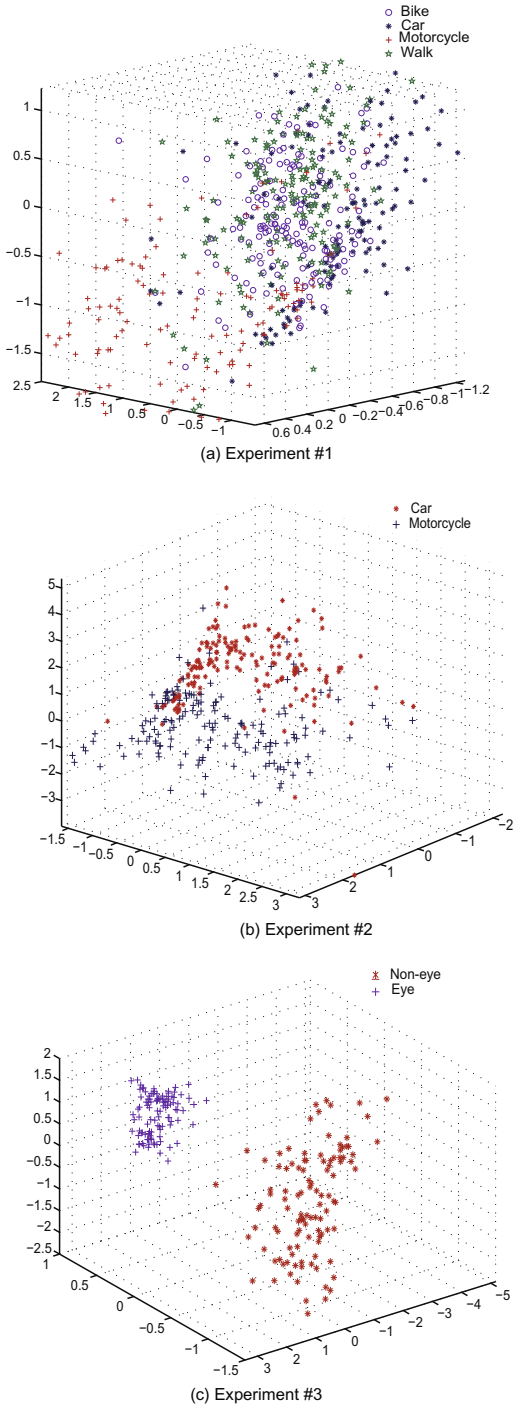


Fig. 8. 3-D plot of the first three principal components of the sample trajectories.

performance of the classification. More specifically, in this experiment we intended to assess how well trajectories of motorcycles and cars, as exemplars of MPOs of similar speed range, could be differentiated.

Finally, the intention of experiment #3 was to evaluate how similar (or different) trajectories generated by eye movement are from trajectories of non-eye objects from the transportation domain (i.e. motorcycles, cars, bicycles, and pedestrians) using the proposed methodology in a multi-class classification mode. Consequently, we ran the SVM learning process with a training set consisting of 90 eye movement trajectories and 90 non-eye movement trajectories (i.e. 25 motorcycles, 25 cars, 20 pedestrians and 20 bicycles trajectories). We tested the classification performance using a testing set of 25 eye movement trajectory samples, together with 25 non-eye movement trajectory samples (i.e. seven motorcycles, eight cars, five bicycles, five pedestrians).

In order to perform the experiments, we used the LIBSVM tool (Chang & Lin, 2001). We applied a radial basis function (RBF) kernel with two parameters: $c = 2$, which is a penalty function for misclassified sample points of training data; and $\gamma = 0.07$, which is an exponent factor in the RBF function (Cristianini & Shawe-Taylor, 2000). They were obtained by trying out different parameter combinations and evaluating the classification accuracy by means of cross-validation. The results of experiments #1 and #2 are presented in Table 6. From experiment #3, we achieved a classification accuracy of 100% cleanly separating all eye movement trajectories from the non-eye trajectories used in this study. Thus, we refrain from presenting this result in a table.

6. Discussion

In this section we discuss the results presented in the previous section. We first compare the characteristics of the 2-D trajectories as well as their associated movement parameters expressed in the profiles, then discuss the results of the three classification experiments, and finally take a brief look at efficiency considerations.

6.1. Global and local movement descriptors

6.1.1. Trajectories

Not surprisingly, the descriptive statistics of the straightness index and the 2-D plots of the trajectories (Table 1, Figs. 4 and 5) as well as the straightness index profiles for the trajectory samples suggest that the car movement with a mean straightness index value of 1.49 and standard deviation close to 0.11 represents the smoothest movements, while eye movement is the most unsteady movement, with a mean straightness index value of 8.77 and a standard deviation of 89.69.

The 2-D plots of the exemplar motorcycle, car, bicycle and pedestrian trajectories (Fig. 4) suggest that the geometry of such objects with a sampling rate of one second to some extent is comparable to each other. However, from the further numerical analysis and systematic classification that we have done in experiments #1 and #2, it can be concluded that these four moving objects behave differently in terms of the velocity, acceleration and straightness index of their paths (Tables 1–3; and Figs. 6 and 7, left side).

6.1.2. Velocity

As Figs. 6b and 7b and Table 7 show, the velocity of cars, bicycles and pedestrians lies in two classes of high (above 90%) and low (less than 10%) deviation from the median, always with low sinuosity. On the other hand, the velocity profile of motorcycle movement changes between all four decomposition classes. It mostly lies in two classes of high (72.48%) and low (15.1%) sinuosity, with low deviation from the median. This means that velocity undulates very closely around the median and does generally not deviate

Table 6
Results of the SVM classification for the experiments #1 and #2.

Experiment	Object	# Train traj.	# Test traj.	# Correct class	Error of commission	Error of omission	Kappa coefficient	% Correct class
Exp. #1	Motorcycle	140	25	23	0.041	0.08	0.76	82
	Car	140	25	21	0.043	0.12		
	Bicycle	140	25	19	0.34	0.24		
	Pedestrian	140	25	18	0.25	0.28		
Exp. #2	Motorcycle	140	25	23	0.042	0.08	0.88	94
	car	140	25	24	0.077	0.04		

Table 7
Summary table of the velocity profile decomposition of the sample trajectories.

Obj#	Mean	Stddev	Low Sinuosity–low deviation			hiGh Sinuosity–low deviation			Low Sinuosity–high deviation			High Sinuosity–high deviation		
			% class	Mean length	Stddev length	% class	Mean length	Stddev length	% class	Mean length	Stddev length	% class	Mean length	Stddev length
Motorcycle	28.73	6.91	72.48	7.85	6.41	15.1	2.31	0.95	5.37	2.23	0.43	1	1.67	0.58
Car	10.95	3.07	91.27	90.67	84.18	0	0	0	8.72	13	1.41	0	0	0
Bicycle	4.56	3.88	91.94	274	0	0	0	0	8.05	24	0	0	0	0
Pedestrian	3.25	0.56	97.65	97	127.47	0	0	0	2.34	3.5	2.12	0	0	0
Eye	308.34	617.03	73.87	7.42	6.80	17.59	3.18	1.40	0	0	0	8.54	2	0

greatly from the trajectory (i.e. only 5.37% of profile points are classified as high deviation). The results indicate that the velocity profiles of the bicycles and pedestrians have the least variations between classes and the highest proportion of low sinuosity-low deviation points. However, the velocity profiles of the motorcycle, car, bicycle and pedestrian trajectories have some small perturbations that can be attributed to the limited accuracy of the GPS and random noise. In comparison, the profile of eye movement velocity suddenly increases at certain points (Fig. 7b on the right) when a saccade (i.e. rapid movement of the eyes) happens, although it stays close to the median (like the pedestrian movement) for the remaining part of the profile at fixation points, where the eyes fixate (Salvucci & Goldberg, 2000). This points to the potential of using our approach to detect fixations and saccades from eye-movement protocols. As shown in Fig. 9, long segments of low deviation indicating fixations can be nicely extracted from short segments of high sinuosity-high deviation with a length of only 1 or 2 points in saccades. This behavior is distinctly different from the velocity variation of the other objects under study.

6.1.3. Acceleration

In terms of the profile decomposition classes, the acceleration profiles of the five objects share similarities with the corresponding velocity profiles (Figs. 6 and 7, Table 8). For instance, the acceleration profile of cars (and similarly for bicycles and the pedestrians) mostly varies very close to its median, with only 0.33% of points showing a higher deviation. All profiles show a higher proportion of high sinuosity-low deviation points than the corresponding velocity profiles. In the case of motorcycle, car, bicycle and pedestrian movement, there are some small perturbations that cause higher sinuosity on the corresponding acceleration profiles, which are due to the accuracy of the GPS devices used as well as random noise. This noise could be removed by curve fitting to profiles (instead of simply smoothing the trajectories). In the case of the eye movement and motorcycle movement, it is interesting to see that despite the noise, the high sinuosity-high deviation points are also picked up in the acceleration profiles. For eye movement, the match is even perfect; some segments are slightly shorter but they all occur at the same spot as in the velocity profiles. Therefore,

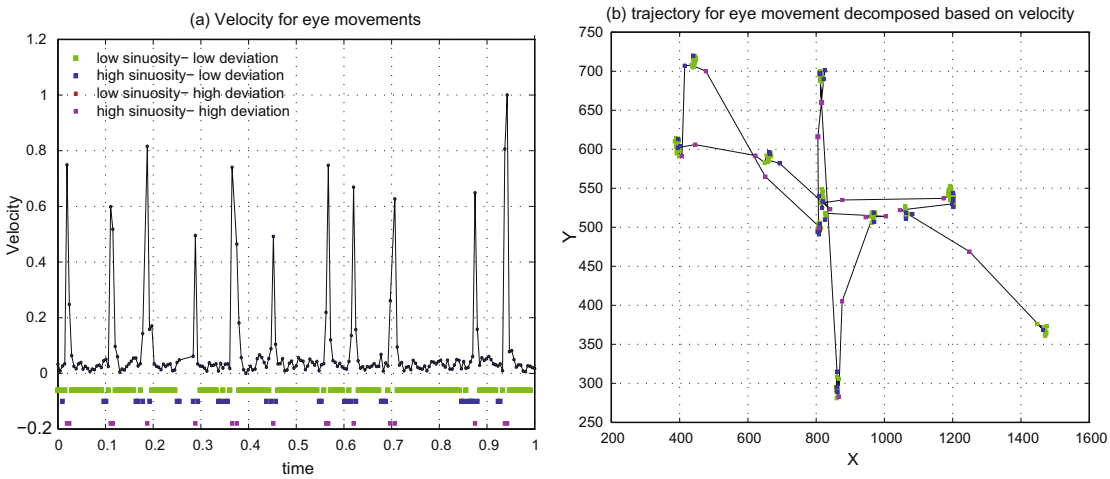


Fig. 9. Extracting saccades and fixations from trajectories of eye movement.

Table 8
Summary table of the acceleration profile decomposition of the sample trajectories.

Obj#	Mean	Stddev	Low sinuosity–ow deviation			High sinuosity–low deviation			Low sinuosity–high deviation			High sinuosity–high deviation		
			% Class	Mean length	Stddev length	% Class	Mean length	Stddev length	% Class	Mean length	Stddev length	% Class	Mean length	Stddev length
Motorcycle	0.002	7.4	45.97	8.68	6.45	38.59	6.67	8.05	2.01	1.75	0.5	13.42	1.97	0.16
Car	0.04	1.3	99.32	148	207.89	0.33	1	0	0.33	1	0	0	0	0
Bicycle	0.01	0.22	88.59	52.8	46.67	0	0	0	11.4	4.5	3.25	0	0	0
Pedestrian	0	0.14	90.60	38.57	34.85	0.67	1.5	0.7	8.39	4.16	1.6	0.33	1	0
Eye	-0.44	41.15	70.35	7.68	6.64	23.62	2.52	1.34	0.50	1	0	5.52	1.9	0.32

as stated earlier, the proposed segmentation algorithm can be employed successfully on velocity and acceleration profiles of eye movement trajectories as a fixation detection method. The acceleration profile of the motorcycle movement shows longer periods of high deviation than the eye movement and a more intermittent pattern of changes between the four different decomposition classes than any other profile (Fig. 6b).

6.1.4. Straightness index and turning angle

The decomposition results for the straightness index profiles are not shown graphically, in order to save space. A summary of decomposition classes is given in Table 9. The results indicate that motorcycle, car, bicycle and pedestrian movement are very smooth, with about 98% of the profile points assigned to the low sinuosity class. In the case of cars, bicycles and pedestrians the profile mostly stays close to the median (about 98% in the high deviation class). However, the motorcycle profiles lie in 10% of the cases in the high deviation class. In contrast, from the decomposition results it is obvious that the path of eye movement trajectories is more sinuous.

By the same token, the decomposition results of the turning angle profiles (not shown here to save space) demonstrated that the turning angle profiles of eye movement are very rough and exhibit an irregular, almost violent behavior, in contrast to the turning angle profiles of the other objects.

6.2. Trajectory classification

For experiment #1, the multi-class classification of motorcycle, car, bicycle and pedestrian trajectories, we achieved an overall accuracy of 82% and a Kappa coefficient of 0.76 (Table 6). One of the motorcycle sample trajectories was classified as a car trajectory and another one was classified as a bicycle trajectory. The same happened in the case of car movements (three misclassifications). The other misclassifications were due to pedestrian trajectories classified as bicycle trajectories, and vice versa. As the discussion of movement parameter profiles above shows, these misclassifications were due to the fact that motorcycle and car movements on the one hand, and bicycle and pedestrian movements on the other hand, are indeed quite similar. The two confusions of motorcycle trajectories with a bicycle and a car, respectively, were related to movement samples at lower speed.

For experiment #2, the motorcycle vs. car classification, we reached an overall accuracy of 94% and a Kappa coefficient of 0.88 (Table 6). One car movement sample was classified as motorcycle, and two motorcycle samples classified as car. These misclassifications were again due to the fact that these particular samples happened to fall into an extended period of low speed movement.

For experiment #3, the eye vs. non-eye classification, we achieved an overall accuracy of 100%. This is clearly due to the fact that the non-eye MPOs used in this experiment are lacking the typical saccadic movement patterns of eyes. Hence, we can conclude that generating movement parameters similar to those of other moving objects is not possible using eye-movement data, and hence eye-movement data are not suitable as a proxy of other movement data that are examined in this study.

The above findings are further illustrated in Fig. 8, which shows a 3-D plot of the first three principal components computed on the trajectory samples used in the three experiments. Fig. 8.a shows how the bicycle and the pedestrian samples take the middle ground between the car and the motorcycle movement samples. Fig. 8b illustrates the separation of the car and the motorcycle movement samples. Fig. 8c then illustrates how the eye movement samples clearly stay apart from the non-eye movement observations (motorcycle, car, bicycle and pedestrian samples).

From the outcomes of the experiments it can be concluded that the amplitude and variation of velocity and acceleration are the most essential features in recognizing a certain travel mode or object type. For instance, the following rules, which can also be discovered from Figs. 6 and 7, are learned by the SVM to classify the trajectories: if the velocity and acceleration profiles are rather smooth and mostly composed of low sinuosity-low deviation segments, then the profile may belong to a trajectory of a car or bicycle. If the velocity and acceleration profiles contain a number of points with high sinuosity, then they may belong to a motorcycle trajectory. If the velocity and acceleration profiles have a jagged geometry consisting of a set of low sinuosity-low deviation segments interrupted by a set of high sinuosity-high deviation points, then the profiles are indicating the saccadic movement of eyes.

6.3. Efficiency

In order to be useful for data mining our proposed methodology has to be reasonably efficient for massive databases or for real-

Table 9
Summary table of the straightness index profile decomposition of the sample trajectories.

Obj#	Mean	Stddev	Low sinuosity–low deviation			High sinuosity–low deviation			Low sinuosity–high deviation			High sinuosity–high deviation		
			% Class	Mean length	Stddev length	% Class	Mean length	Stddev length	% Class	Mean length	Stddev length	% Class	Mean length	Stddev length
Motorcycle	1.5	0.2	89.26	53.2	56.66	0.33	1	0	10.4	2	0	0	0	0
Car	1.5	0.15	97.99	97.67	88.99	0.67	2	0	1.34	2	1.41	0	0	0
Bicycle	1.48	0.18	96.64	96	61.73	2.01	3	1.41	0.33	1	0	1	1.67	0.58
Pedestrian	1.49	0.16	96.98	96.33	87.75	0.33	1	0	2.68	4	1.41	0	0	0
Eye	5.32	9.27	54.27	5.27	4.15	35.17	3.71	2.55	0.5	1	0	10.05	1.94	0.23

time applications. Due to lack of a large trajectory database it was not possible to empirically assess the computational performance of our methodology under these conditions. Nevertheless, we would like to briefly touch on efficiency issues here in order to support the argument that our methodology indeed has the potential to be used with massive datasets or in a real-time setting.

First, all parts of the methodology including the profile decomposition algorithm run in linear time, except the PCA and the SVM classification. Second, the training stage of the SVM classifier, which is known to have slow computational performance, is run off-line and on a subset of the data. And finally, it is possible to replace the PCA and the SVM classifier by simpler and computationally more efficient techniques.

6.4. Test data used

The test data sets used are relatively large: 660 (4×165) transportation tracks for experiments #1 and #2, and another 115 eye movement tracks for experiment #3. We believe our experiments to be sufficient to establish the *feasibility* of the proposed methodology. However, the test data are restricted to movement on overland and suburban roads (i.e., no urban traffic included) and they were originally sampled at a similar temporal interval (around 1 s). In order to make conclusive statements about the *scope of applicability* of the proposed methodology, the experiments would have to be extended to data sets of very different moving objects; to traffic movement in urban situations; and possibly to data that have been sampled at different temporal resolutions and may contain gaps.

While such experiments still need to be carried out, we expect that the methodology should be capable of handling tracks with different transportation modes due to the decomposition of trajectories into segments of homogenous character based on change points (Zheng et al., 2008). Also, the decomposition algorithm used is based on simple principles and does not use any extra knowledge, which is why we expect it to be robust also for different moving object types. The performance of the decomposition, and thus of the overall methodology, might decrease for very short trajectories or tracks with similar movement parameters, for instance in congested traffic situations. However, by considering the history of the entire trajectories, such track sections may be classified more accurately. For instance, knowing the velocity characteristics in uncongested parts of the trajectories involved in a congestion, bicycles may be distinguished from cars or motorcycles.

7. Conclusions

We have presented a comprehensive, three-stage methodology that allows extracting movement parameters from the trajectories of different types of moving objects. As one of the application of the proposed methodology, we showed how to classify trajectories of unknown MPOs by similarity to the trajectories of previously learned MPOs. We have then conducted a series of experiments that not only demonstrated the feasibility of the proposed methodology but also provided interesting empirical results. Our experiments provide evidence about the similarities and differences that exist among different types of moving objects in the transportation domain. The results show that using our methodology we can successfully detect the mode of transport from unknown trajectories of people using different transportation means. It was also shown that eye-movement data cannot be successfully used as a proxy of full-body movement of humans, or vehicles. The physics of movement of virtually mass-less movement processes, such as eye movement (and possibly also computer mouse movement), is very different from the movement of objects that are governed by inertia to a much greater extent. Nevertheless, the methodology can contribute to finding the most feasible proxies for desired moving objects in various application domains (e.g. biology, ecology). For instance, eye movement could potentially be considered a proxy of some objects that have a stop-and-go movement behavior such as bees and butterflies.

We see potential for future work in three directions. First, there is plenty of room for more experiments aiming to further exploit, enhance and consolidate the proposed methodology. For instance, experiments with different trajectory datasets; other MPO types; different transportation data (e.g. movement on urban roads); different sets of movement parameters; fine-tuning of the SVM classifier (e.g. kernel tuning); and other classification techniques (e.g. decision trees). Since we have set up the methodology in a streamlined, automated process, we are in a good position to conduct such further experiments. From the point of view of real-time processing, experiments with a simpler classifier than SVM, which is known to have a high computational complexity, may be warranted. Finally, the proposed methodology could be developed further to set up an automatic transport mode detection system in transportation applications.

Second, we are interested in further exploring the method and results of movement parameter profile decomposition. For instance, as we discussed in Section 6.1, we believe that there is a potential in using the decomposition algorithm as an alternative technique for fixation detection in the analysis of eye-tracking data. Also, we are interested in using the results of the profile decomposition algorithm for trajectory similarity analysis as well as for more differentiated parameterization of movement simulators.

Third, our methodology currently does not take into account the context and constraints that influence movement. Further studies therefore have to consider how to involve movement context.

Acknowledgement

This research is funded by the University of Zurich. We would like to express our gratitude to Patrick Laube and Patrick Lüscher (University of Zurich), and David Mountain (City University of London) for their helpful comments and valuable suggestions, as well as Arzu Cöltekin (University of Zurich) for providing us with eye movement data.

References

Agouris, P., Partsinevelos, P., & Stefanidis, A. (2003). Differentiating and modeling multiple moving objects in motion imagery datasets. *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*. XXXIV-5/W10.

Bashir, F. I., Khokhar, A. A., & Schonfeld, D. (2007). Object trajectory-based activity classification and recognition using hidden Markov models. *IEEE Transactions on Image Processing, 16*(7), 1912–1919.

Batty, M., (2003). *Agent-based pedestrian modelling. Working paper*. Centre for Advanced Spatial Analysis, University College London, UK [ISSN: 1467-1298].

Bay, S., & Pazzani, M. (2001). Detecting group differences: Mining contrast sets. *Data Mining and Knowledge Discovery, 5*(3), 213–246.

Benhamou, S. (2004). How to reliably estimate the tortuosity of an animal's path: Straightness, sinuosity, or fractal dimension? *Journal of Theoretical Biology, 229*, 209–220.

Blythe, P. W., Miller, G. F., & Todd, P. M. (1996). Human simulation of adaptive behavior: Interactive studies of pursuit, evasion, courtship, fighting, and play. In *Proceedings of the fourth international conference on simulation of adaptive behavior* (pp. 13–22). Cambridge, MA: MIT Press/Bradford Books.

Bozdogan, H. (2003). *Statistical data mining and knowledge discovery*. CRC Press, Inc..

Brillinger, D. R., Preisler, H. K., Ager, A. A., & Kie, J. G. (2004). An exploratory data analysis (EDA) of the paths of moving animals. *Journal of Statistical Planning and Inference, 122*(1–2), 43–63.

Chatfield, C. (1989). *The analysis of time series: An introduction* (4th ed.). London UK: Chapman and Hall.

Chang, C. -C., & Lin, C. -J. (2001). LIBSVM: A library for support vector machines. <http://www.csie.ntu.edu.tw/~cjlin/libsvm>.

Chen, L., Özsu, M. T., & Oria, V. (2005). Robust and fast similarity search for moving object trajectories. In *Proceedings of the ACM SIGMOD international conference on management of data* (pp. 491–502) ACM.

Chen, M. C., Anderson, J. R, & Sohn, M. H. (2001). What can a mouse cursor tell us more? Correlation of eye/mouse movements on web browsing. In *Proceedings of the ACM conference on human factors in computing systems (CHI'01)* (pp. 281–282). Seattle, Washington, USA.

Cox, A. L., & Silva, M. M. (2006). The role of mouse movements in interactive search. In *Proceedings of the 28th annual CogSci conference*, Vancouver, Canada (pp. 1156–1162).

Cristianini, N., & Shawe-Taylor, J. (2000). *An introduction to support vector machines and other kernel-based learning methods*. Cambridge: Cambridge University Press.

Dodge, S., Weibel, R., & Lautenschütz, A.-K. (2008). Towards a taxonomy of movement patterns. *Information Visualization*. doi:10.1057/palgrave.ivs.9500182.

Duda, R. O., Hart, P. E., & Stork, D. G. (2001). *Pattern classification*. New York: John Wiley & Sons, Inc.

Dutton, G. (1999). Scale, sinuosity, and point selection in digital line generalization. *Cartography and Geographic Information Science, 26*, 33–53.

Eubank, R. (2005). *A Kalman filter primer (statistics: Textbooks and monographs)*. Chapman & Hall/CRC.

Giannotti, F., & Pedreschi, D. (Eds.). (2007). *Mobility, data mining and privacy – geographic knowledge discovery* (pp. 90–99). Berlin Heidelberg: Springer-Verlag.

González, M. C., Hidalgo, C. A., & Barabási, A.-L. (2008). Understanding Individual Human Mobility Patterns. *Nature, 453*, 779–782.

Guyon, I., & Andre, E. (2006). An Introduction to feature extraction. In I. Guyon, S. Gunn, M. Nikravesh, & L. A. Zadeh (Eds.), *Feature extraction: Foundations and applications studies in fuzziness and soft computing* (pp. 1–25). Springer-Verlag.

Hoffmann-Wellenhof, B., Lichtenegger, H., & Collins, J. (2001). *Global positioning system: Theory and practice* (5th ed.). Wien: Springer.

Hornsby, K. S., & Cole, S. (2007). Modeling moving geospatial objects from an event-based perspective. *Transactions in GIS, 11*, 555–573.

Jun, J., Guensler, R., & Ogle, J. (2007). Smoothing methods to minimize impact of Global Positioning System random error on travel distance, speed, and acceleration profile estimates. *Transportation Research Record, 1972*, 141–150.

Laube, P., & Purves, R. S. (2006). An approach to evaluating motion pattern detection techniques in spatio-temporal data. *Computers, Environment and Urban Systems, 30*, 347–374.

Laube, P., Dennis, T., Forer, P., & Walker, M. (2007). Movement beyond the snapshot – Dynamic analysis of geospatial lifelines. *Computers, Environment and Urban Systems, 31*, 481–501.

Lin, B., & Su, J. (2008). One way distance: For shape based similarity search of moving object trajectories. *GeoInformatica, 12*, 117–142.

Naftel, A., & Khalid, S. (2006). Classifying spatiotemporal object trajectories using unsupervised learning in the coefficient feature space. *Multimedia Systems, 12*, 227–238.

Ozyildiz, E., Krahnstöver, N., & Sharma, R. (2002). Adaptive texture and color segmentation for tracking moving objects. *Pattern Recognition, 35*, 2013–2029.

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

Tsui, S. Y. A., & Shalaby, A. (2006). An enhanced system for link and mode identification for GPS-based personal travel surveys. *Transportation Research Record*, 1972 (pp. 38–45).

Salvucci, D. D., & Goldberg, J. H. (2000). Identifying fixations and saccades in eye-tracking protocols. In *Proceedings of the eye tracking research and applications symposium* (pp. 71–78). New York: ACM Press.

Schüssler, N., & Axhausen, K.W. (2009) Processing GPS raw data without additional information. In *Paper presented at the 88th annual meeting of the transportation research board*, Washington, DC, January 2009.

Smith, L. I. (2002). *A tutorial on principal components analysis*. USA: Cornell University.

Spaccapietra, S., Parent, C., Damiani, M. L., de Macedo, J. A., Porto, F., & Vangenot, C. (2008). A conceptual view on trajectories. *Data and Knowledge Engineering, 65*(1), 126–146.

Turchin, P. (1998). *Quantitative analysis of movement: Measuring and modelling population redistribution in animals and plants*. Sunderland, MA: Sinauer Publishers.

Vlachos, M., Gunopulos, D., & Kollios, G. (2002). Robust similarity measures for mobile object trajectories. In *Proceedings 13th international workshop on database and expert systems applications* (pp. 721– 726). Washington, DC, USA: IEEE Computer Society.

Zheng, J. -B., Dagan Feng, D., & Zhao, R. -C. (2005). Trajectory matching and classification of video moving objects. In *Proceedings 7th IEEE workshop on multimedia signal processing* (pp. 1–4).

Zheng, Y., Liu, L., Wang, L., & Xie, X. (2008). Learning transportation mode from raw GPS data for geographic applications on the web. In *Paper presented at the 17th World Wide Web conference*, Bejing, April 2008.

Research Paper 3

Dodge, S., Laube, P., and Weibel, R. (in revision, 2011). Movement Similarity Assessment Using Symbolic Representation of Trajectories. *International Journal of Geographic Information Science*.

International Journal of Geographical Information Science
Vol. 00, No. 00, Month 200x, 1–25

RESEARCH ARTICLE

Movement Similarity Assessment Using Symbolic Representation of Trajectories

Somayeh Dodge*, Patrick Laube, and Robert Weibel

University of Zurich, Winterthurerstrasse 190, 8057, Zurich, Switzerland.

(*revisions received June 2011*)

This paper describes a novel approach for finding similar trajectories, using trajectory segmentation based on movement parameters such as speed, acceleration, or direction. First, a segmentation technique is applied to decompose trajectories into a set of segments with homogeneous characteristics with respect to a particular movement parameter. Each segment is assigned to a *movement parameter class*, representing the behavior of the movement parameter. Accordingly, the segmentation procedure transforms a trajectory to a sequence of class labels, that is, a symbolic representation. A modified version of edit distance, called *Normalized Weighted Edit Distance* (NWED) is introduced as a similarity measure between different sequences. As an application, we demonstrate how the method can be employed to cluster trajectories. The performance of the approach is assessed in two case studies using real movement datasets from two different application domains, namely, North Atlantic Hurricane trajectories and GPS tracks of couriers in London. Three different experiments have been conducted that respond to different facets of the proposed techniques, and that compare our NWED measure to a related method.

Keywords: Movement similarity; trajectory segmentation; movement parameter; movement patterns; trajectory clustering.

*Corresponding author. Email: somayeh.dodge@geo.uzh.ch

ISSN: 1365-8816 print/ISSN 1362-3087 online

DOI: 10.1080/1365881YYxxxxxxxx
http://www.informaworld.com

1. Introduction

Understanding the dynamic behavior of moving objects (e.g., humans, animals, vehicles, etc.) or processes (e.g., hurricanes, oil spills) is quickly becoming a key task in many GIScience application domains. In movement behavior studies, it is essential to take into account the key parameters that characterize the movement of objects, so-called *movement parameters (MP)* such as speed, acceleration, or direction (Dodge *et al.* 2008). For example, in intelligent transport systems it is important to know the speed patterns of vehicles across the street network at different times of the day in order to detect traffic anomalies and incidences (Miller and Han 2009). Likewise, in order to predict the position or time of a hurricane's landfall it is crucial to know the speed, acceleration and direction behavior of hurricanes before and close to the landfall (Elsner and Kara 1999). Movement parameters either can be derived from the trajectories of objects or recorded directly by sensors. Today, with the emergence of new sensor technologies such as accelerometers, gyroscopes, and recent advances of in-vehicle sensors, a variety of movement parameters of mobile objects can be registered as an object moves. The development of analysis techniques that are capable of exploiting these new sources of information thus appears to be a logical step forward for knowledge discovery from movement datasets.

Similarity analysis as an exploratory tool in movement research is an important and challenging topic. Recently, similarity analysis has become the focus of many studies in mobility data mining (Giannotti and Pedreschi 2008, Miller and Han 2009). A review of the relevant literature suggests that although there are several trajectory similarity search methods that are relatively well developed, most of them are restricted to geometric abstractions of the objects' movement path as a static curve (i.e. a time-ordered sequence of coordinates). And only a few of the available similarity analysis techniques take the variations of movement parameters into account. However, in many applications spatial similarity alone may not be appropriate to detect objects with similar movement characteristics. For instance, trajectories of vehicles that move on the same route are similar in terms of geometric shape (i.e. spatial similarity). However, the speed variations of vehicles might exhibit different patterns over time that cannot be discovered with purely spatial similarity assessment. As another example, "history has shown that many of the hurricanes that have struck New England over the last 100 years share very similar characteristics"[1]. For instance, in 1954 two hurricanes, Carol and Edna, made landfall on Cape Cod, MA. and exhibited very similar movement characteristics insofar as they have been known as identical hurricanes in the meteorological literature in terms of their evolution (Malkin and Holzworth 1954). However, since their movement paths exhibit a distinctively different geometry (see Figure 1), these two hurricanes would not be extracted as similar using the available spatial movement similarity measures. By the same token, hurricane Edna and hurricane Isaac, that their movement paths follow a similar geometry as seen in Figure 1, would be extracted as similar, although their speeds exhibit a very different pattern. Therefore, in addition to geometric similarity of trajectories, it seems inevitable to investigate the similarity of the *variation of movement parameters* over time.

The *objective* of this paper is to contribute to trajectory similarity search, by proposing a new approach that allows seeking for trajectories in movement datasets that exhibit *common patterns in the variation of their movement parameters over time*. The *main contribution* of this research is the introduction of a novel technique for spatio-temporal

[1] http://www.hurricanescience.org/history/storms/1950s/carol/

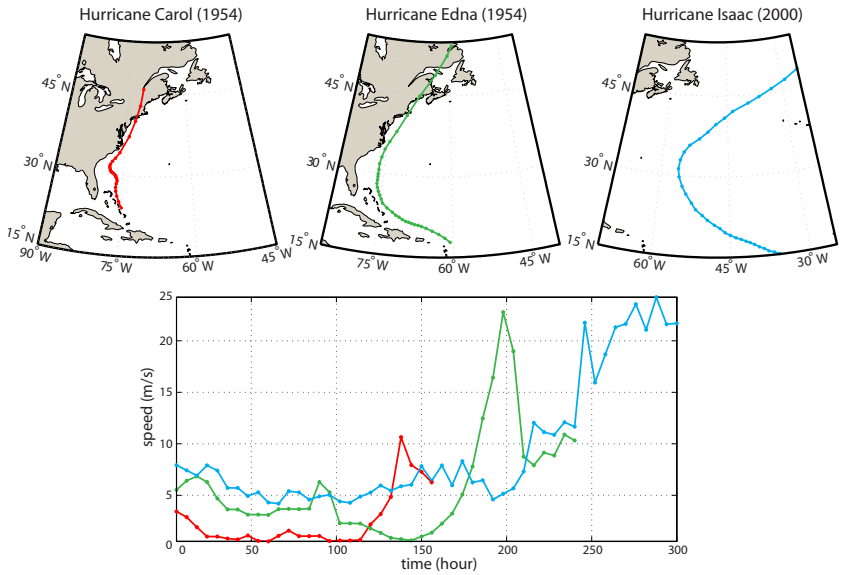


Figure 1. Hurricane Edna (green), exhibited a very similar behavior to her predecessor, Carol (red). Hurricane Isaac (blue) exhibited a different speed pattern and yet similar geometry to hurricane Edna. Data source: http://www.nhc.noaa.gov/

trajectory similarity assessment that relies on trajectory segmentation based on the movement parameters of the objects under study, yielding segments of homogeneous movement characteristics. The segmentation process leads to a simplified, compressed representation of trajectories, called *movement parameter class (MPC) sequence*, which converts movement parameter profiles derived from trajectories into a symbolic representation. In this representation, the important movement characteristics are preserved. We propose a modified version of *edit distance*, termed *Normalized Weighted Edit Distance (NWED)* as the measure of similarity between trajectories. We evaluate NWED in comparison to a relevant similarity measure, called EDM, proposed by Chen *et al.* (2004). Similar to our approach, EDM applies the *edit distance* on *a symbolic representation of trajectories* obtained from the segmentation of two movement parameters, distance and direction between fixes. However, there are distinctive differences in how the movement parameter classes and the edit distance are computed, as will be discussed later.

Additionally, as an application example of the proposed method we present two clustering strategies: (1) MPC diversity-based clustering, using descriptive statistics of the trajectories' MP classes; and (2) MPC sequence-based clustering using NWED. These strategies are presented as demonstrations to show how our proposed segmentation and similarity analysis techniques can be applied for clustering trajectories according to similarities in the spatio-temporal variation of their movement parameters.

The remainder of this paper is organized as follows: Section 2 gives a structured overview of previous research on similarity analysis, segmentation of movement data and trajectory clustering. Section 3 introduces the methodology that is applied in this research. Section 4 describes our approach for trajectory clustering as one of the applications of the developed method. Section 5 presents three experiments based on two case studies using tracking data of hurricanes and couriers in an urban setting, respectively.

Section 6 discusses the key findings of the experiments, the strengths, and limitations of our method in comparison to existing techniques. Finally, Section 7 presents the concluding remarks and outlines directions for future work.

2. State of the Art

2.1. *Movement Similarity Assessment*

Similarity between two objects is quantified as the cost of transforming one entity into another or the distance between the two objects, using a *similarity measure* (Faloutsos *et al.* 1997). So far, a variety of similarity measures has been developed in order to address various aspects of movement similarity assessment problems. The existing movement similarity assessment techniques can be divided into two classes of (1) *spatial similarity*, that is, finding trajectories with similar geometric shape, ignoring the temporal dimension; and (2) *spatio-temporal similarity*, focusing on both spatial and temporal aspects of movement data. Up to now, the proposed movement similarity assessment methods mostly originate from either *time series similarity measures* such as Euclidean distance, edit distance, Longest Common Subsequence (LCSS), and Dynamic Time Warping (DTW) (Ding *et al.* 2008b), or *geometric shape matching* techniques such as Fréchet distance or Hausdorff distance (Alt 2009).

2.1.1. Spatial Measures of Movement Similarity

Most of the recent works on similarity search in trajectory data address the spatial similarity problem (Vlachos *et al.* 2002a,b, Yanagisawa *et al.* 2003, Chen *et al.* 2005, Lin and Su 2005). Euclidean distance based approaches are usually less complex than the other methods and only work on trajectories of the same duration and granularity. The efficiency of such methods decreases when the movement data contain noise, outliers or gaps. The proposed techniques based on LCSS or edit distance are more robust in this respect. The latter approaches can be applied for trajectories of different durations or granularity, albeit at higher computational cost.

The above techniques use the trajectory representation of movement. In contrast, some methods have been proposed that represent trajectories using movement parameters. For example, Little and Gu (2001) apply DTW on separate *path* and *speed* curves of trajectories. In their approach, in order to simplify the process of similarity analysis, a local geometric feature extraction technique is applied using *curvature* information of the path and speed curves. Curvature is invariant to scaling and rigid motion transformations. Another DTW based approach proposed by Vlachos *et al.* (2004) to find similar trajectories of the same granularity under translation, scaling, and rotation transformations. In this approach, a different representation of trajectories based on *turning angle* and *distance* of trajectory fixes over time is applied. DTW based methods allow trajectories to be stretched or compressed and hence do not preserve relative speed of the trajectories. With a different approach, Chen *et al.* (2004) introduced a new representation of trajectories, called *movement pattern strings (MPS)*, in order to optimize similarity computation using an extension of the edit distance, called *Edit Distance on Movement Pattern Strings (EDM)*. MPS is a symbolic representation of a trajectory using *movement direction*, and *distance ratio* information derived from the original trajectory.

The latter methods are relevant to our proposed approach in terms of incorporating *movement parameters* in similarity assessment of moving objects. However, these measures mainly look into the spatial aspect of movement data by taking geometric

movement parameters such as distance and direction. Among the aforementioned techniques, the method by Chen *et al.* (2004) is directly comparable to our method since it also applies an *edit distance* on a *symbolic representation of trajectories*. In contrast, Little and Gu (2001) and Vlachos *et al.* (2004) use DTW as similarity measure on metric (non-symbolic) values of movement parameters. Moreover, since Chen *et al.* (2004) uses the relative movement parameters (i.e. distance and direction between two consecutive fixes), the temporal dimension of movement is implicitly involved in the similarity computation when data is sampled at a regular sampling rate. As it is experimentally shown in Chen *et al.* (2004, 2005), and Ding *et al.* (2008b), the edit distance is more accurate than the other commonly used trajectory similarity measures such as Euclidean distance, DTW, LCSS, particularly in the presence of noise in the data. Hence, Experiment #2 in Section 5.1.3 will be devoted to empirically comparing NWED to the method by Chen *et al.* (2004).

2.1.2. Spatio-temporal Measures of Movement Similarity

As pioneering studies in the context of spatio-temporal similarity, Sinha and Mark (2005), Frentzos *et al.* (2007) employed the Euclidean distance for regularly sampled trajectories. Later, van Kreveld and Luo (2007), Buchin *et al.* (2009) improved such techniques to extract the most similar subtrajectories using an approximation from a set of trajectories with different granularity. Pelekis *et al.* (2007) considered a slightly different approach and proposed a family of distance measures, *Locality in In-between Polylines (LIP)*. LIP relies on the *area* of the polygons formed between the intersection points created by the overlay of two trajectories. In order to compute the spatio-temporal similarity between trajectories, different weight factors are applied to support the detection of concurrent movement of objects that move closely at similar speeds, or directions (Pelekis *et al.* 2007). The accuracy of LIP based measures is influenced by penalty factors that need to be specified by user. On the other hand, because the fundamental element of this approach is the area between intersection points, these measures work better for trajectories which follow the same route and are not appropriate for winding trajectories with a lot of turns. Moreover, an additional search process is required to find the intersection points prior to the similarity assessment. Another study by Trajcevski *et al.* (2007) suggested a Rigid Transformation Similarity Distance employing the notion of Fréchet distance to compute the similarity between trajectories under translation and rotation transformations. In a similar attempt, Ding *et al.* (2008a) proposed the *w-constrained Fréchet distance (wDF)*, which constrains the discrete Fréchet distance by a given temporal threshold. The computational cost of these methods is rather high, especially for long trajectories. These measures are not restricted to similar geometric routes, however, they do not consider the speed of the objects either.

2.2. *Trajectory Segmentation*

Segmentation is essential in many applications where the subject of the analysis is a complex and heterogeneous phenomenon (e.g. map generalization, time series analysis etc.). In the study of movement, segmentation facilitates finding patterns and structures in movement data and hence can help to understand the behavior of objects. Trajectory segmentation refers to decomposing a trajectory into segments of homogenous characteristics. Segmentation has recently been applied in several studies in the domain of moving object data analysis in order to simplify trajectories for several purposes, such as indexing and efficient data handling (Anagnostopoulos *et al.* 2006), event and activity

recognition along the geospatial lifelines of objects (Yan *et al.* 2010), and classification of movement data (Dodge *et al.* 2009).

In this study, we use segmentation in order to convert trajectories into revealing structures for extracting similarities in the movement characteristics of objects. Therefore, we extend the segmentation method proposed in an earlier paper by Dodge *et al.* (2009). Their method applies feature extraction techniques from map generalization and time series analysis in order to decompose trajectories into sequences of homogeneous movement characteristics. Buchin *et al.* (2010) recently proposed a similar segmentation approach, however, applying different criteria on movement parameters of objects (e.g. using ranges of speed or turning angle) on a continuous representation of trajectories.

2.3. *Trajectory clustering*

Trajectory clustering is an exploratory data mining technique that, similar to segmentation, facilitates studying movement data and understanding their structure by reducing its complexity. Miller and Han (2009) and Kisilevich *et al.* (2010) provide a survey of the well-known clustering algorithms such as hierarchical clustering, k-means, DBSCAN, OPTICS, BIRCH, TRACLUS, and TOPTICS that have been used or proposed for clustering trajectory data. Most of these approaches do not consider the semantics inherent to the trajectories and often treat movement data as point clouds in a space-time cube and cluster points based on their spatial density over time. In contrast, Etienne *et al.* (2010) proposed a filtering approach to extract a group of trajectories with similar spatio-temporal patterns among moving objects following the same itinerary. Their method takes the semantics of trajectories into consideration in terms of the planned or optimal itinerary and schedule of movement.

3. Methodology

When an object moves about in space over time (Fig. 2.a), the evolution of its movement parameters can be represented as a function or profile over time (Fig. 2.b). We refer to this function as *movement parameter profile* (or short MP profile). The amplitude and frequency variation of such functions can be quite different for different types of moving objects (Dodge *et al.* 2009). Even more so, the differences may pertain to different episodes of a single object's lifeline, owing to the diversity of the underlying physics of movement and the behavior of the object. However, when two or more objects move in a similar way, the corresponding functions will most likely also express similarity. This has led us to using movement parameter profiles for extracting similarities among trajectories, by decomposing the trajectories into segments (i.e. sections) exposing similar movement characteristics.

An overview of the main methodology used in this work is illustrated in Figure 2 and explained in detail in the subsequent sections. Our approach relies on a discrete trajectory representation (Fig. 2.a), modeling trajectories as a sequence of coordinates over time (Laube *et al.* 2007). The methodology consists of two main processes: (1) trajectory segmentation (Fig. 2.b-c), and (2) similarity computation (Fig. 2.d). The key component of the segmentation algorithm is the movement parameter profile. The segmentation procedure makes use of the amplitude and frequency variations of movement parameters over time and will be explained next. The similarity computation uses a variation of the edit distance as described in Section 3.2.

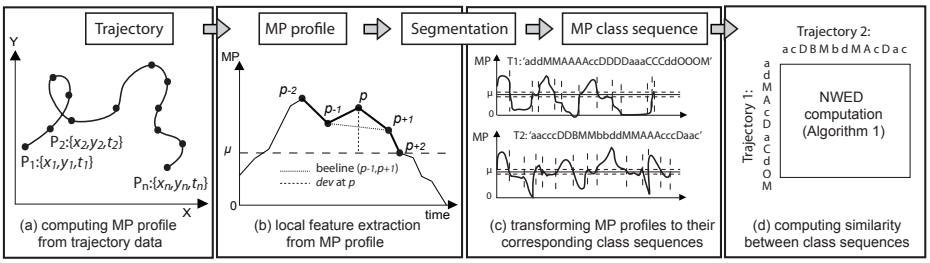


Figure 2. Overview of the trajectory segmentation and similarity computation process.

3.1. *Trajectory Segmentation*

For the purpose of segmentation, an MP profile (e.g. the profile of the movement parameter *speed*) is first generated from trajectory data (Fig. 2.a,b). Then, the MP profile is segmented into sections of homogeneous movement characteristics using an extension of the segmentation approach introduced in Dodge *et al.* (2009). There are several extensions, however, which will be explained in Section 3.1.2. The aim of segmentation is to reduce the complexity of trajectories while conserving their important movement features.

3.1.1. Extracting Local Features from an MP Profile

In order to measure movement characteristics from MP profiles, two measures are used: *Deviation* from the mean value and *sinuosity* of MP profiles, respectively. Deviation gives an impression of the amplitude variation of a movement parameter over time, while sinuosity acts as a proxy of the frequency variation of movement parameters. The amplitude and frequency of variation of an MP profile describe the important features in the evolution function of the corresponding movement parameter over time, and hence identify the main characteristics of movement of an object (Dodge *et al.* 2009).

Figure 2.b provides the supporting illustrations for the computation of the deviation measure and the sinuosity measure on an MP profile. Both measures are defined for each point on the MP profile. The MP profile is first standardized using its standard score ($z = \frac{x-\mu}{\sigma}$) to make it dimensionless and comparable with other profiles. The deviation of a point p on an MP profile equates to its residual value from the mean line (i.e. *dev* in Fig. 2.b). The z-scores indicate the deviation of MP values from the mean line of the original MP profile and therefore have a mean of $\mu = 0$ and standard deviation of $\sigma = 1$. The measure of *sinuosity* for p is computed as a ratio of the distance $\pm k$ points along the profile to the length of the beeline connector centered at the point (i.e. beeline at p for $k = 1$ in Figure 2.b). Where k is the *lag* parameter and is considered as $k = 2$ for GPS observations as discussed. The final sinuosity at p is computed as the average of the computed sinuosity values with different k, as shown in Equation (1). In the end, the sinuosity values are transformed to the interval $[0, 1]$, as proposed in Dodge *et al.* (2009). That is, if profile points are collinear about the given point p the sinuosity measure equals 0 and for a winding profile (i.e. a space-filling curve) it comes to 1.

It should be noted that the lag parameter depends on the temporal granularity, spatial scale, as well as the noise level of the observations. The higher k is set, the bigger the window size gets in the sinuosity computation. And hence, the sinuosity results are smoothed over more points. For instance, for the macro scale hurricane dataset with a temporal granularity of some hours (i.e. 6 hours) and little noise, k can be set to

1. In contrast, for micro scale observations such as eye-tracking data with a temporal granularity of some milliseconds and a high amount of tremors, k can take higher values to reduce the effect of noise.

$$Sinuosity_{p,k} = \frac{\sum_{i=p-k}^{i=p+k-1}(d_{i,i+1})}{d_{p-k,p+k}} \tag{1}$$

$$Sinuosity_p = \frac{\sum_{j=1}^{j=k} Sinuosity_{p,j}}{|k|}$$

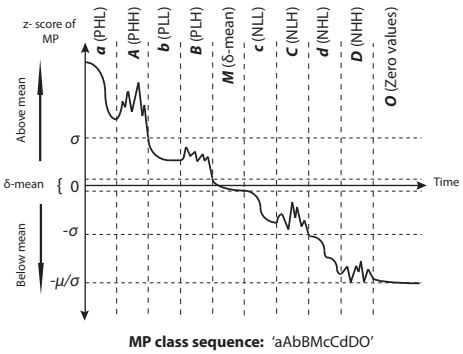
3.1.2. Transforming an MP Profile to an MP Class Sequence

With the movement parameter segmentation procedure, the profile points are classified into two regimes regarding the level of the corresponding sinuosity measure, *low sinuosity* and *high sinuosity*, separated by a user-defined threshold. The same is done with deviation to separate *low deviation* from *high deviation* . In addition to these classes that were introduced in Dodge *et al.* (2009), the position of the points with respect to the mean line is also used to distinguish *positive deviation* (i.e. above the mean) from *negative deviation* (i.e. below the mean). The reason being that since in this study the interest is to detect similarity in the movements of objects over time, it is essential to know whether the amplitude levels of movement parameters are increasing or decreasing. In contrast, Dodge *et al.* (2009) aimed merely at the classification of moving objects according to their intrinsic movement properties. Therefore, the properties of movement parameter profiles were of greater importance than their relative values (e.g. values below or above the mean). Accordingly, the number of *sinuosity* and *deviation* classes is doubled, compared to Dodge *et al.* (2009). Moreover, an additional class of MPs is considered for values within an acceptable error threshold δ from the mean, providing a further facility to deal with noise. Consequently, nine classes are extracted from the segmentation of MP profiles, called *movement parameter class (MPC)* (Fig. 3). *An MPC indicates the type of amplitude and frequency variations of a movement parameter (i.e. speed, acceleration etc.).*

The number of classes is sought to be small, yet describing the main features of an MP profile. In addition to the aforementioned main movement parameter classes, segments with zero values of MP profiles and with the z-score equal to $\frac{-\mu}{\sigma}$ are tagged as a separate and optional class 'O'. This class obtains importance in application domains like transportation where zero values represent stops (i.e. movement at zero speed) and are treated accordingly.

In this study, the deviation threshold is set to the standard deviation of the z-scores (equal to 1) and the sinuosity threshold is set to 0.80, respectively, following the experience made in Dodge *et al.* (2009). Similarly, the δ threshold, which is the acceptable error threshold, is set to 0.01.

With the segmentation process each trajectory is transformed into a *symbolic representation*, or *string* composed of a sequence of movement parameter classes, called *movement parameter class sequence* or short *class sequence* (Fig. 3). A *movement parameter class sequence* is composed of a string of MP classes representing the transition pattern of movement parameters along a trajectory. The domain of the string is $\Sigma = \{A, a, B, b, C, c, D, d, M\}$. Each character of the string represents a specific class (i.e. A: PHH, a: PHL, B: PLH, b: PLL, C: NLH, c: NLL, D: NHH, d: NHL, and M: δ-mean). This new representation of the trajectories is then employed for the purpose of



MPC	Deviation	Sinuosity	Assigned letter
PHH	positive high	high	A
PHL	positive high	low	a
PLH	positive low	high	B
PLL	positive low	low	b
M	δ-mean	—	M
NLH	negative low	high	C
NLL	negative low	low	c
NHH	negative high	high	D
NHL	negative high	low	d

Figure 3. MP classes and MP class sequence.

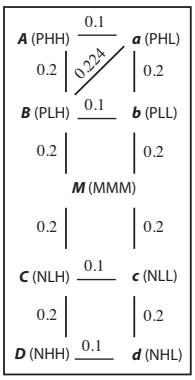
similarity computation as described in Section 3.2.

3.2. *Similarity Computation Between MP Class Sequences*

The class sequence representation of MP profiles is now exploited for assessing trajectory similarity. In order to detect similar movement behaviors of objects, our method searches for similar transitions of the MP classes along the trajectories. To do so, the raw trajectories are first transformed to their respective class sequences. Subsequently, in order to calculate the similarity between the sequences we introduce a modified version of the edit distance as a similarity measure, called *Normalized Weighted Edit Distance (NWED)*. The edit distance is used since it is related to the concept of string matching and is a metric to measure the difference between two sequences, in our case MPC sequences. In fact, NWED extends the Levenshtein distance as a well-known form of edit distance. The Levenshtein distance is defined as the smallest number of insertions, deletions, and substitutions required to convert one string into another (Levenshtein 1966). The edit distance and its variations have been widely used in bioinformatics and speech recognition and recently in similarity analysis of movement data as discussed in Section 2.1.1. Computing the distance between strings with the edit distance has a complexity $O(m \times n)$, where m and n are the lengths of the two strings. Nevertheless, the efficiency of the edit distance can be improved using fast string matching techniques, indexing, or pruning approaches (Du Mouza *et al.* 2007, Ding *et al.* 2008b).

The *Normalized Weighted Edit Distance (NWED)* computes the weighted and normalized cost of converting one MP class sequence into another using *edit operations* (i.e. insertion, deletion, substitution). In comparison to the original edit distance, the modification of NWED concerns the costs of the insertion, deletion, and substitution operators, which all are equal to 1 in the original version of the edit distance. In contrast, we define the substitution costs differently for each specific pair of MP classes. That is, the costs are weighted based on the degree of dissimilarity between different classes. Here, insertion and deletion are given the maximum cost, which is defined as 1. Since inserting or deleting points changes the length of MP profiles (i.e. duration of movement), we consider such operations to be more severe than the substitution of different MP classes. Moreover, as it is illustrated in figures 3 and 4, we assign *four* degrees of differences between the considered *amplitude levels*. In order to normalize the costs to the domain $[0, 1]$, all substitution costs are divided by 5 (i.e. 4 amplitude degrees + 1 insertion/deletion

cost), leading to 0.2 for each substitution cost of two consecutive amplitude levels. Moreover, the substitution cost between *low deviation* and *high deviation* classes is assigned twice the substitution cost than between *low sinuosity* and *high sinuosity* classes. The reason being that the amplitudes (i.e. deviations) give an indication of the long term variation of movement parameters and hence a considerable cost (i.e. time, energy) is required to transit from one level of amplitude to another. In contrast, the frequencies (i.e. sinuosities) capture the short term variation and local features of the MP profiles, hence, the transition between the two classes of sinuosities at the same level of amplitude involve less cost. Accordingly, MP values of the same level of deviation but with a different sinuosity regime are considered more similar than the ones that deviate more yet have the same sinuosity regime. For instance, MP values of class $'a'$ are considered more similar to an $'A'$ segment than to a $'b'$ segment (see Figure 3 and Figure 4). By the same token, $'C'$ is more similar to $'c'$ than to $'B'$. Accordingly, the other costs are determined using the Pythagorean theorem as shown in Figure 4. These costs, summarized in $COST\_Matrix$ in Equation (2), indicate the degree of dissimilarity between the different trajectory segments that belong to different MP classes.



$d(x_i, \phi) = 1$ insertion/deletion cost

$d(B, C) = d(b, c) = 2/5 = 0.4$

$d(B, c) = d(b, C) = 2.06/5 = 0.412$

$d(A, D) = d(a, d) = 4/5 = 0.8$

$d(A, d) = d(a, D) = 4.03/5 = 0.806$

$d(A, B) = d(B, M) = d(M, C) = d(C, D) = 1/5 = 0.2$

$d(a, b) = d(b, M) = d(M, c) = d(c, d) = 1/5 = 0.2$

$d(A, a) = d(B, b) = d(C, c) = d(D, d) = 0.5/5 = 0.1$

$d(A, b) = d(B, a) = d(C, d) = d(D, c) = 1.12/5 = 0.224$

$d(A, c) = d(a, C) = d(b, D) = d(B, d) = 3.04/5 = 0.608$

$d(A, C) = d(a, c) = d(b, d) = d(B, D) = 3/5 = 0.6$

$d(A, M) = d(a, M) = d(d, M) = d(D, M) = 2/5 = 0.4$

Figure 4. Costs (i.e. dissimilarities) of the MP classes employed in NWED computed using the Pythagorean theorem.

$$\text{COST\_Matrix} = \begin{array}{c} \\ PHH \\ PHL \\ PLH \\ PLL \\ MMM \\ NLH \\ NLL \\ NHH \\ NHL \end{array} \begin{array}{c} \begin{array}{ccccccccc} PHH & PHL & PLH & PLL & MMM & NLH & NLL & NHH & NHL \end{array} \\ \begin{pmatrix} 0 & 0.1 & 0.2 & 0.224 & 0.4 & 0.6 & 0.608 & 0.8 & 0.806 \\ 0.1 & 0 & 0.224 & 0.2 & 0.4 & 0.608 & 0.6 & 0.806 & 0.8 \\ 0.2 & 0.224 & 0 & 0.1 & 0.2 & 0.4 & 0.412 & 0.6 & 0.608 \\ 0.224 & 0.2 & 0.1 & 0 & 0.2 & 0.412 & 0.4 & 0.608 & 0.6 \\ 0.4 & 0.4 & 0.2 & 0.2 & 0 & 0.2 & 0.2 & 0.4 & 0.4 \\ 0.6 & 0.608 & 0.4 & 0.412 & 0.2 & 0 & 0.1 & 0.2 & 0.224 \\ 0.608 & 0.6 & 0.412 & 0.4 & 0.2 & 0.1 & 0 & 0.224 & 0.2 \\ 0.8 & 0.806 & 0.6 & 0.608 & 0.4 & 0.2 & 0.224 & 0 & 0.1 \\ 0.806 & 0.8 & 0.608 & 0.6 & 0.4 & 0.224 & 0.2 & 0.1 & 0 \end{pmatrix} \end{array} \quad (2)$$

Algorithm **1** presents the computation process of the NWED between two class sequences T and P. In order to compute the dissimilarity between two segmented trajectories (i.e. two MP class sequences), one is considered as *subject trajectory* (i.e. $T[1 \ldots n], \quad |T| = n$) and the second one is considered as a *pattern* or *template* (i.e. $P[1 \ldots m], \quad |P| = m$). A $n \times m$ dissimilarity matrix (i.e. WED_Matrix) is then formed based on the costs between segments of the two trajectories, obtained from $COST\_Matrix$ (Equation (2)), applying Equation (3a) (Bozkaya *et al.* 1997). The element $WED\_Matrix(n, m)$ indicates the cost of conversion between the two sequences

or the dissimilarity between the two trajectories. Finally, in order to remove the effect of varying length of trajectories on the similarity results, we normalize the total dissimilarity between T and P, as proposed by Yujian and Bo (2007) using Equation (3b). The NWED obtained from Equation (3b) is metric, as proven by Yujian and Bo (2007).

$$WED\_Matrix_{T,P} = C_{0\ldots n, 0\ldots m} \tag{3a}$$

$$C_{i,j} = \begin{cases} j & \text{if } i = 0 \\ i & \text{if } j = 0 \\ C_{i-1,j-1} & \text{if } i, j > 0 \text{ and } T_i = P_j \\ COST\_Matrix\,(T_i, P_j) + \min\,(C_{i-1,j-1}, C_{i-1,j}, C_{i,j-1}) & \text{otherwise} \end{cases}$$

$$NWED_{T,P} = \frac{2 \times C_{n,m}}{n + m + C_{n,m}} \tag{3b}$$

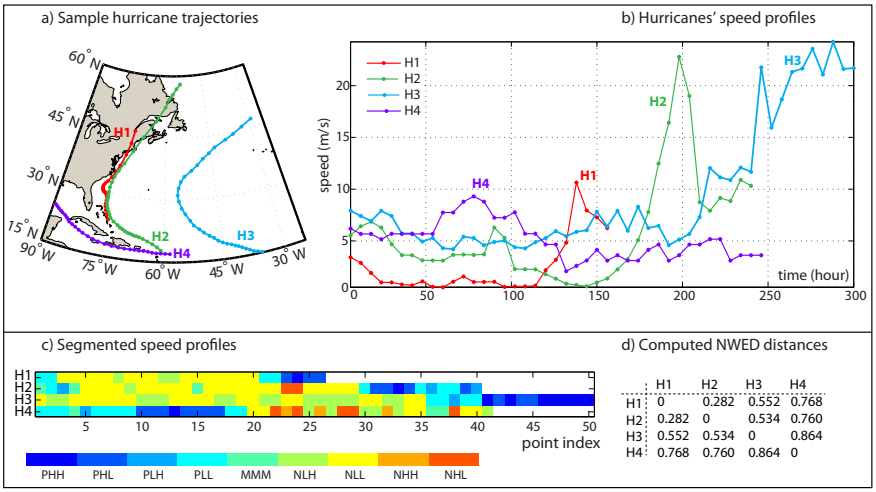
Algorithm 1 NWED

Require: input: two sequences T and P, lengths of T and P
Require: input: $COST\_Matrix$ (weighted costs between MP classes)
Ensure: output: NWED between two sequences T and P

```
n ← |T|
m ← |P|
sumLength ← (n + m)
for i = 0 to n do
    for j = 0 to m do
        if i = 0 then
            WED_Matrix(i, j) ← j
        else if j = 0 then
            WED_Matrix(i, j) ← i
        else if T_i = P_j then
            WED_Matrix(i, j) ← WED_Matrix(i − 1, j − 1)
        else
            mpcCOST ← COST_Matrix(T − i, P − j)
            minCOST ← min { WED_Matrix(i − 1, j − 1)
                            WED_Matrix(i − 1, j)
                            WED_Matrix(i, j − 1) }
            NWED_Matrix(i, j) ← mpcCOST + minCOST
        end if
    end for
end for
WED_{T,P} ← WED_Matrix(n, m)
NWED_{T,P} ← 2×WED_{T,P} / (sumLength+WED_{T,P})
return NWED_{T,P}
```

Figure 5 provides an example of running the segmentation and NWED algorithms on four hurricane trajectories of different durations (i.e. of unequal length), including hurricanes Carol (H1), Edna (H2), and Isaac (H3) mentioned in the Introduction (Fig. 5.a). Hurricanes Carol and Edna exhibit a rather similar increase-decrease speed pattern in their evolution from formation to decaying after landfall, although their movement paths differ. Hurricane Isaac is the longest and its movement path exhibits a relatively similar curve to hurricane Edna. Contrary to the other selected hurricanes, the speed profile of Isaac (H3) shows a variable, yet overall an increasing pattern, since the hurricane does not make landfall. On the other hand, hurricane H4 (San Zacarias, 1910) moves along a relatively straight path but its speed varies more frequently. As the figure illustrates, first of all the speed profiles are computed from trajectories (Fig. 5.b). Then, the speed

profiles are converted into their corresponding class sequences using the segmentation method (Fig. 5.c). Figure 5.d presents the computed pairwise NWED between the four hurricanes. As it can be seen from the figure, the NWED between hurricanes Edna (H1) and Carol (H2) is computed as 0.282, which is smaller with respect to the other pairwise NWED distances, since both hurricanes show a relatively similar sequence in their speed variations. The computed NWED between Isaac (H3) and Edna (H2), 0.534, indicates that these two hurricanes are relatively dissimilar in terms of their speed patterns. Hurricane H4 is obtained as the most dissimilar hurricane ($NWED > 0.7$), since its speed exhibits a distinctively different pattern with respect to the other hurricanes.



	H1	H2	H3	H4
H1	0	0.282	0.552	0.768
H2	0.282	0	0.534	0.760
H3	0.552	0.534	0	0.864
H4	0.768	0.760	0.864	0

Figure 5. Computation of NWED on four hurricane trajectories including hurricanes Carol (H1), Edna (H2), and Isaac (H3).

4. Application: Trajectory Clustering

The proposed methodology can be applied in different application domains, whenever the aim is to discover common behaviors of dynamic entities in space and time. These include applications such as trajectory clustering (e.g. grouping trajectories with similar speed patterns) and movement pattern detection (e.g. discovering concurrence patterns). Below, we suggest two strategies for trajectory clustering, as a demonstration of the usefulness of our proposed segmentation method and similarity measure. Since they are both based on the trajectory segmentation, they are able to capture the variations of the movement parameters over time, rather than simply clustering the trajectories based on geometric properties. However, a filtering process such as the one presented in Etienne *et al.* (2010) (section 2.3) can be applied as a preprocessing step in order to ensure route similarity. Both strategies can in principle be applied in conjunction with any standard clustering techniques such as hierarchical clustering, K-means, or DBSCAN. However, clustering results may vary according to the clustering techniques. In this study, we applied the complete-linkage agglomerative hierarchical clustering technique (Miller and Han 2009).

4.1. *Strategy 1: MPC Diversity-Based Clustering of Trajectories*

The first approach is based on descriptive statistics computed on the MP classes resulting from the segmentation process. For each MP, the number of transitions of different classes and the percentage, standard deviation and average length of each MPC of the domain (i.e. $\Sigma = \{A, a, B, b, C, c, D, d, M\}$) in the MPC sequence is computed. The number of transitions of each class gives an indication of the diversity in the dynamic trend of the MP and the percentage of each class in the MP profile indicates the frequency of the contribution of each class of Σ to the total trend of the MPC sequence. Therefore, for each segmented trajectory the following features are computed: percentage of the number of class alterations (i.e. 1 feature), percentage of contribution of each class (i.e. 9 features), the mean and standard deviation of each class length (i.e. 18 features). Thus, the total number of features that are fed to the clustering process is 28. Following that, a standard hierarchical clustering approach can be used to cluster the segmented trajectories (i.e. sequences) from the computed features. This strategy focuses more on the *diversity and frequency* (i.e. variability) of the variations of movement parameters rather than the sequence of the transitions.

4.2. *Strategy 2: MPC sequence-based Clustering of Trajectories*

The second approach uses the NWED distance function introduced in section 3.2 for similarity-based clustering of trajectory data. Hence, a distance matrix is computed here from the pairwise NWED between segmented trajectories (i.e. sequences). This distance matrix is then used as an input for the clustering process, along with the complete-linkage agglomerative hierarchical clustering to group trajectories with similar trend in the transition of MP variations. In comparison to the previous approach (Section 4.1), which considers more the variability of the segments, this method focuses on the transition sequence of the classes of the MP profiles. Therefore, this method is recommended for applications where the evolution of the movement parameters is important, such as movement behavior study of hurricanes, and homing pigeons (Laube *et al.* 2007).

5. Experiments

We conducted two case studies in order to assess the applicability of the proposed methods. In these two case studies, the proposed methods were applied on two distinct types of movement data, from different application domains and with rather different dynamic behaviors. From the domain of meteorology, we considered tracks of North Atlantic hurricanes, which express rather smooth and predictable trajectories. In contrast, from the transportation domain, GPS trajectories of couriers were analyzed. The latter dataset involves very diverse dynamic behaviors. Additionally, we compared the outcomes of our similarity measure with the one introduced by Chen *et al.* (2004). To do so, we conducted a comparative experiment employing hurricane tracking data since we have the background information from meteorological literature to validate the computed similarities. Table 1 summarizes the objectives of the conducted experiments.

Table 1. Overview of the Experiments

Case Study	Data	Exp. No	Objective
I	Hurricanes	Exp. #1	Assessment of the diversity-based versus sequence-based clustering methods in seeking structures in hurricanes.
		Exp. #2	Evaluation of NWED in comparison to the method by Chen *et al.* (2004) (i.e. EDM) on hurricanes.
II	Couriers	Exp. #3	Assessment of the diversity-based versus sequence-based clustering methods in exploring traffic patterns.

5.1. *Case Study I - Clustering North Atlantic Hurricanes*

According to meteorological studies, a hurricane develops gradually in different phases, from formation to decay after landfall (Elsner and Kara 1999). During each phase of the lifecycle, hurricanes to a great extent have similar characteristics. Apart from meteorological prerequisites, the *season* and the *geographic latitude* of the hurricanes' origin are two important factors influencing the dynamic behavior of hurricanes. Elsner and Kara (1999) distinguish between 'classic' *low-latitude* hurricanes originating south of about 20° N and *high-latitude* hurricanes originating north of about 20° N. Furthermore, the authors state that North Atlantic Hurricanes have similar dynamic characteristics based on the time of formation and their source locations. That is, it is observed that hurricanes of similar season (i.e. fall or summer) with similar characteristics usually originate in a spatial proximity (i.e. in the same quadrants w.r.t. (19° N, 80° W)). This case study aims at evaluating our similarity assessment and clustering approaches by confirming these findings. Therefore, we primarily sought for extracting two clusters in the historical tracks of hurricanes to see whether the results correspond with the categorization of the *low-latitude* and *high-latitude* hurricanes. Moreover, we were interested to assess whether the clusters relate to the season in which hurricanes occurred or to the distance of their origin to the US coastline.

5.1.1. Data

Since destructions caused by hurricanes mainly happen after landfall, it is most important to investigate the dynamic behavior of hurricanes that reach the coastline. Hence, this case study uses trajectories of 397 hurricanes that made landfall at the East or South East coastline of the United States between 1907 and 2007[1]. The temporal sampling rate of the observations is 6 hours. Since the raw hurricane movement dataset obtained from NOAA contains little noise and is regularly sampled, no preprocessing was required prior to the segmentation and clustering procedures.

5.1.2. Experiment #1: Assessment of Diversity-Based vs. Sequence-Based Clustering Methods

The aim of this experiment is to evaluate the usefulness of the proposed clustering approaches for seeking structure in a movement dataset with sequential behavior. The similarity of hurricane trajectories was assessed based on their *speed* and *turning angle* (change of movement direction) behavior, since such MPs are key parameters that affect the time and location of a hurricane' landfall.

[1] from NOAA's Coastal Services Center (http://csc-s-maps-q.csc.noaa.gov/hurricanes/)

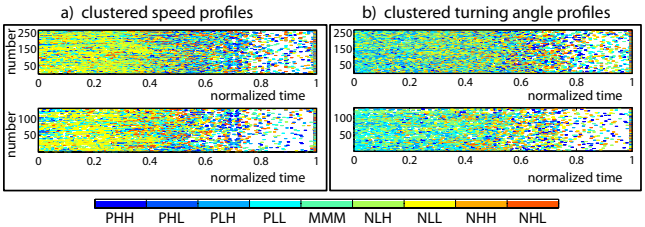


Figure 6. Experiment #1 (diversity-based clustering): Class sequences of the two clusters of a) speed, and b) turning angle.

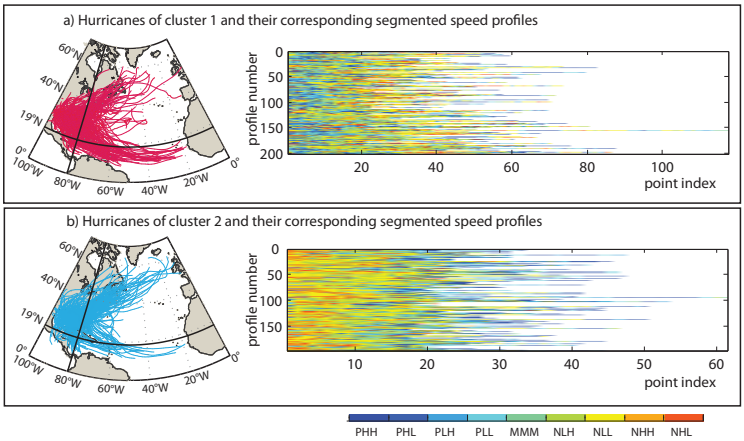


Figure 7. Experiment #1 (sequence-based clustering of speed profiles): The map view and segmented speed profiles of the two clusters of North Atlantic Hurricane trajectories.

For diversity-based clustering (cf. section 4.1), 28 descriptive statistics features were derived for 397 trajectories, for both segmented MP profiles, and used for hierarchical clustering. Two clusters for speed profiles and two clusters for turning angle profiles were generated. The clustered MP profiles are presented in Figure 6. As Figures 6.a illustrates, both speed clusters exhibit a relatively smooth transition from NLL (yellow) to PHH (dark blue). Similarly, a gradual transition from MMM (aquamarine) to PHL (light blue) can be observed in the turning angle profiles (Figure 6.b). Compared to the speed profiles, more variability can be observed in the turning angle behavior.

For sequence-based clustering (cf. section 4.2), NWED is used to compute distance matrices between segmented MP profiles. Two distance matrices of extent 397×397 were computed, one for speed and one for turning angle, which were then used separately for clustering. Two clusters were generated for both speed and turning angle profiles. Figure 7 illustrates the map view and the class sequence representations of the two clusters for speed only. The obtained clusters also reveal the gradual increase-decrease pattern in the speeds of hurricanes. However, the trend of this pattern varies in the two clusters: From NLH over NLL to PHH in cluster 1 (Figure 7.a), and from NLL to PHL in cluster 2 (Figure 7.b).

In order to test whether clusters reveal significant difference between hurricanes of different months or hurricanes originating from different locations, we applied Mann-Whitney U tests. The resulting p-values (see Table 2, Exp. #1) show which attributes

Table 2. Experiment #1: Computed p-values of the Mann-Whitney U test on the clustering results for hurricanes (* $p < 0.05$)

Exp.	**Method**		**No. of clusters**	**Lat. of origin**	**Long. of origin**	**Low vs. high Lat. origin**	**Origin west or east of 80°W**	**Season**
#1	diversity-based	speed	2	0.107	0.356	0.249	0.068	0.019*
		turning angle	2	0.354	0.323	0.088	0.187	0.497
	NWED	speed	2	.000*	.000*	.000*	.000*	0.001*
		turning angle	2	.000*	.000*	.000*	.000*	0.196
#2	EDM		2	.000*	.000*	.000*	0.025*	0.281

do or do not explain the two generated clusters: latitude of origin, longitude of origin, low-latitude vs. high-latitude of origin, origin west or east of 80° W, season (fall or summer), and month of the year. The results indicate no significant difference in the source location of hurricanes in the obtained clusters from diversity-based clustering (p-values above 0.05). Since the hurricanes generally show a gradual increasing-decreasing speed trend (speeding up after formation, slowing down after landfall) (Elsner and Kara 1999), their speed profiles exhibit little variability and rather a sequential behavior. Therefore, the effectiveness of the diversity-based clustering decreases for clustering such sequential behavior. In contrast, the clusters obtained from sequence-based clustering reveal a significant difference between trajectories originating from the latitudes north and south of 19° N ($p < .001$, Table 2, Exp. #1; Figure 7). Hence, the sequence-based clustering method (using NWED) was capable of distinguishing between *low-latitude* and *high-latitude* hurricanes. Moreover, with respect to longitude, clusters obtained from sequence-based clustering differentiate hurricanes originating east and west of 80° W ($p < .001$, Table 2, Exp. #1). This distinction reflects the distance of the hurricanes' origin to the US coastline (see map view in Figure 7).

On the other hand, the clusters obtained from both clustering methods on speed profiles suggest a significant difference between hurricanes in different seasons (i.e. summer and fall) ($p < 0.05$ in Table 2, Exp.#1). In cluster 1 obtained from sequence-based clustering, we find a tendency for hurricanes in summer (i.e. May to August) and originating from the southeastern quadrant of (19° N, 80° W) (Figure 8.a). In contrast, cluster 2 predominantly shows hurricanes in fall (i.e. September to December), originating from the northwestern quadrant of (19° N, 80° W) (Figure 8.b). This observation complies with results from the meteorology literature (Elsner and Kara 1999).

The clusters obtained from turning angle profiles suggest similar outcomes (Table 2, Exp. #1). The result of sequence-based clustering suggests that hurricanes that originate in the same region (quadrant) to a certain extent tend to follow a similar change in their direction or geometric shape (i.e. spatial similarity). However, the trend of this variations does not significantly differ in time (i.e $p = 0.196$ for different seasons). In contrast, we could not find such structure from the results of diversity-based clustering of the turning angles profiles ($p > 0.05$).

5.1.3. Experiment #2: Comparing NWED and EDM

This experiment compares NWED with EDM introduced by Chen *et al.* (2004). For this comparative study we implemented the EDM similarity measure (described in Section

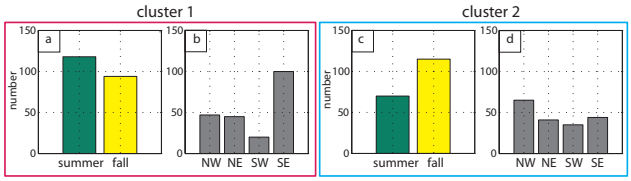


Figure 8. Experiment #1 (sequence-based clustering of speed profiles): Properties of resulting two clusters. Histograms of clusters formed, by season and source location of hurricanes. Source locations are given as quadrants NW, NE, SW, and SE of 19° N latitude and 80° W longitude.

2.1.1), which computes the spatial similarity between *movement pattern string (MPS)* representations of trajectories (Chen *et al.* 2004). The main motivation for this comparative study is that EDM is one of the few available techniques similar to our approach. In detail, EDM and NWED are comparable as they share the following specifications:

- Both consider movement parameter informations in trajectory similarity assessment (e.g. movement direction);
- both measures are an extension of edit distance, and compute distance between symbolic representations of trajectories; and
- both have a similar computational complexity (i.e. $O(n^2)$).

Chen *et al.* (2005) have already investigated the performance of edit distance in comparison with the other similarity measures such as LCSS, DTW, and Euclidean distance. Their study suggests that edit distance is more accurate and robust, specially in the presence of noise and time lags between similar trajectories (local time shifts). Therefore, here we specifically only compare the similarity results of NWED (i.e. sequence-based clustering from Experiment #1) to EDM on clustering the hurricane trajectories, but not the computational complexity. However, in order to make both measures comparable, we normalized EDM to the scale $[0, 1]$ as it is described for NWED in Section 3.2. As recommended, we used a distance threshold of $\epsilon - dis = 0.125$ and a direction threshold $\epsilon - dir = \pi/4$ to generate the 8×8 (movement direction and movement distance) quantization map (Chen *et al.* 2004). Next, movement pattern string (MPS) sequences had to be derived for all hurricane trajectories using the quantization map (Chen *et al.* 2004). Finally, a 397×397 distance matrix for EDM was computed. Just as in Experiment #1, the EDM distance matrix was used for trajectory clustering (again complete-linkage). It is necessary to remark that since hurricane data are sampled at a regular interval, *movement distance* in EDM implicitly represents the speed information of hurricanes, and hence, EDM is indeed comparable to our NWED measure for this case study.

As in Experiment #1 two distinct clusters were generated based on EDM. We applied a Mann-Whitney U test on the clusters, as is described in Experiment #1. The resulting p-values suggest that (see Table 2): The two clusters show significant differences on all attributes related to the location of hurricanes' origin (Lat. and Long. of origin; low vs. high Lat. origin; origin west or east of 80°W). By contrast, the clusters do not significantly differ regarding time-related attributes (i.e $p = 0.281$, and $p = 0.345$ for different seasons and months, respectively). These findings are similar to the results from the clusters obtained from turning angle based on NWED (Experiment #1). The reason being that both MPS and turning angle sequences capture the geometric shape of the hurricanes.

Furthermore, we aimed at examining the difference between distances obtained from EDM in comparison to NWED (based on speed). To do so, we first grouped all hurricanes based on the time of formation (i.e. fall or summer), and then based on the location

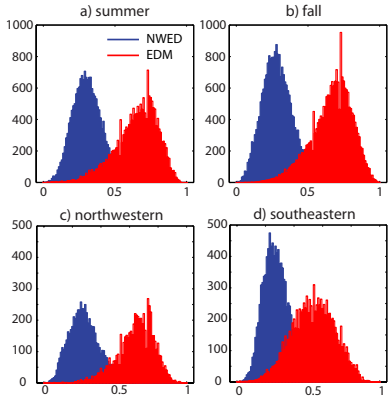


Figure 9. Experiment #2: Histograms of the NWED and EDM distance matrices of hurricanes in a) summer, b) fall, c) northwestern of (19°N, 80°W), and d) southeastern of (19°N, 80°W)

Table 3. Experiment #2: Descriptive statistics and Mann-Whitney U test on NWED and EDM distances

Groups	**No. of distances in groups**	**NWED**				**EDM**				***p*-values**
		Mean	**Std.**	**Median**	**Skewness**	**Mean**	**Std.**	**Median**	**Skewness**	
summer	17578	0.30	0.10	0.3	0.29	0.65	0.14	0.67	-0.71	.000
fall	21736	0.28	0.10	0.27	0.37	0.66	0.14	0.68	-0.63	.000
northwest	6216	0.28	0.11	0.27	0.36	0.64	0.13	0.66	-0.68	.000
southeast	10298	0.27	0.09	0.27	0.47	0.53	0.15	0.53	-0.15	.000

of their origin (i.e. northwestern and southeastern quadrants of (19° N, 80° W)). We then computed descriptive statistics for the NWED and EDM distance distributions for the two groupings (Figure 9 and Table 3). We applied the Mann-Whitney U test on the histograms of each group separately, in order to see if the histograms were the same or different. The resulting p-values (i.e. $p < .001$) indicate for all four cases that the distribution of the NWED and EDM significantly differs (Figure 9 and Table 3). Moreover, having a closer look at the histograms of NWED, we could infer that the hurricanes originating in a spatial and temporal proximity of each other (hurricanes of each group) exhibit a similar speed behavior (mean NWED distances ≤ 0.3). This observation confirms the hypothesis of this case study. By contrast, the results obtained from EDM distances do not reveal such similarity (mean EDM distances > 0.5). Hence, EDM seems to be insufficient in studying the similarity of the speed patterns of hurricanes since it can only capture the spatial similarity.

5.2. *Case Study II - Clustering Courier Trajectories*

In traffic management, it is important to understand the traffic patterns on a given street network over space and time. Quantifying the similarity of vehicles moving on a specific section of a street network can help to distinguish normal and abnormal traffic patterns. This second case study is based on trajectories of couriers captured in Central London

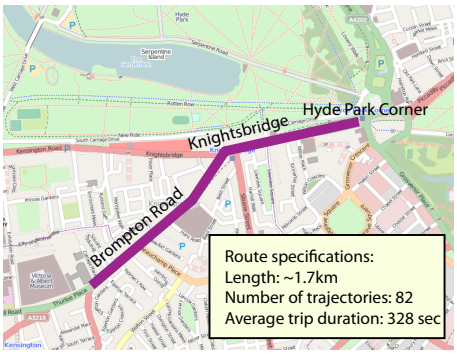


Figure 10. The selected route of courier trajectories (basemap: OpenStreetMap.org)

by the eCourier company[1] during the month of November 2009. The proposed similarity measures shall be used to discover traffic patterns of vehicles moving on a particular section of the street network, based on their speed behaviors.

5.2.1. Data

The raw GPS data have a temporal sampling rate of approximately one fix per 10 seconds. Two subsets were extracted from all courier trajectories. The first subset, named *ZoneData*, covers the Congestion Zone of London[1]. The aim was to evaluate the performance of our approach on a large transportation dataset with diverse behaviors. The second subset, named *RouteData*, contains sets of trajectories that follow a given route. The selected route leads from Hyde Park Corner to the end of Brompton Road, via Knightsbridge (Figure 10). The reason to restrict the study area to a specific route is first to remove the effect of the geometric shape of the road network on the similarity computation from trajectories; and second to render the courier trajectories comparable to the hurricane trajectories (which have a relatively similar geometric shape).

The courier data required an elaborate pre-processing procedure including various filtering and resampling techniques. Here, however, we only list the most important pre-processing steps. First, outliers (speed over 20 ms^{-1}) and stops were removed from the raw GPS tracking data. Next, stops (speed below 1 ms^{-1}) representing deliveries or stops at traffic lights were filtered from the remaining data as suggested in Doherty *et al.* (2001), since stops can contain errors due to loss of signal. Finally, the cleaned trajectories were resampled using linear interpolation to achieve trajectories with a temporal granularity of exactly 10 seconds. Since the raw trajectories contain information about temporal properties and movement parameters, in order to maintain that information no additional smoothing or map matching that could change the geometry of the trajectories was applied.

The average movement phase between two deliveries is approximately 15 minutes. For that reason, for the *ZoneData* we partitioned the pre-processed trajectories into subtrajectories of 15 minutes duration. Eventually, 100 random samples of *small van* trajectories per hour between 8 AM to 8 PM during weekdays were selected (i.e. a total of 1200 subtrajectories of 15 min duration). For *RouteData*, subtrajectories that followed the aforementioned selected route were then extracted from the entire pre-processed

[1] http://www.ecourier.co.uk
[1] http://www.tfl.gov.uk/tfl/roadusers/congestioncharge/whereandwhen/

dataset applying geometric curve matching within a threshold distance of 30 m. Overall, a total of 71 trajectories (i.e. 35 motorbike and 36 vans) with an average duration of 323 seconds were obtained on the selected route between 8 AM to 20 PM during weekdays.

5.2.2. Experiment #3: Exploring Traffic Patterns over Time

Following the pre-processing procedure, the speed profiles speed were derived for both *ZoneData* and *RouteData*. The segmentation and clustering process was applied on the generated profiles, just as in Experiment #1.

First, we aimed at testing the applicability of our approach for finding trajectories with similar speed patterns along a specific route, where the effect of geometry is limited. We applied sequence-based clustering using NWED on the speed profiles of courier trajectories from *RouteData*. Two clusters were generated. The clustered, segmented speed profiles of courier trajectories as well as the space-time cube representation of the corresponding trajectories are illustrated in Figure 11. As it can be seen from the segmented profiles in Figure 11.b, the two obtained clusters show different speed behaviors of the courier vehicles. Specifically, the first cluster represents fast movements with a smooth transition from NLL (yellow) to PHH classes (dark blue). In contrast, the second cluster features diverse movement behaviors, including stop-and-go movements (i.e. repetitive sequences of PLL-NLL-NHH-NHL classes) at the beginning of the trajectory (Figure 11.c). The movement behavior represented by the obtained clusters to a great extent reveals the major traffic light about midway on the route (see Figure 11.a,c). The results suggest that the sequence-based method is capable of detecting the traffic behavior along particular segment of a street network.

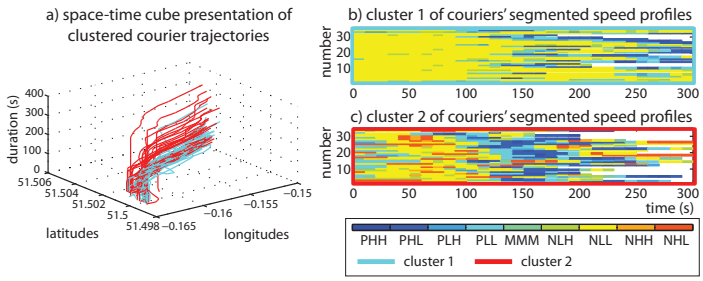


Figure 11. Experiment #3 (sequence-based clustering, RouteData, speed): a) space-time plot, z-axis represents duration, all trajectories are synchronized to start at time 0 s. b) speed class sequences of the resulting clusters

Next, we applied both clustering strategies on *ZoneData*, which contain a larger extents of the street network in comparison to the *RouteData*. The aim was to test whether our methods can detect clusters corresponding to the traffic peak and off-peak hours using speed profiles of the couriers. We were interested to find diurnal time windows corresponding to three categories of *low speed* (i.e. slow movements), *medium speed* (i.e. moderate movements), and *high speed* (smooth movements). Therefore, three clusters were generated from the segmented speed profiles of *ZoneData* using the diversity-based clustering approach (Figure 12). We used the Kruskal-Wallis test in order to first test whether clusters are significantly different in terms of the mean speed. And second, to test whether clusters represent specific diurnal periods. Both tests were resulted in $p < .001$. Hence, the results confirmed that the mean speed of 3 clusters are significantly different ($p < .001$, see box-and-whisker plots in Figure 12.a), although we did not use mean speed

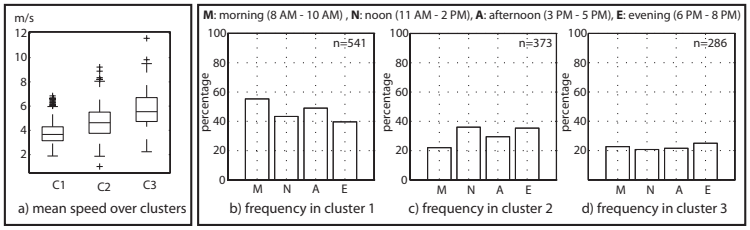


Figure 12. Experiment #3 (diversity-based clustering of speed profiles, ZoneData): Mean speed and frequencies per hour of van trajectories in the three clusters

as a feature in the clustering. Furthermore, the diurnal composition significantly differs between the three clusters ($p < .001$, see Figure 12.b-d). As it can be seen in Figure 12, cluster 1, which represents very slow movement (mean = 3.79 ms^{-1}), has higher frequency during the morning peak (8 - 10 AM) and in the afternoon (3 - 5 PM). In contrast, cluster 2 with medium speed (mean = 4.72 ms^{-1}) represents trajectories at noon (11 AM to 2 PM) and in the evenings (after 6 PM). Cluster 3 with mean speed of 5.7 ms^{-1} did not represent any specific time period. However, it shows a slightly higher percentage in the evenings. These results to a some extent are comparable to the available traffic information about the London Congestion Zone (i.e. AM peak (7 - 10 AM), inter-peak (10 AM - 4 PM), and PM peak (4 - 7 PM))[1].

Just as in Experiment #1, we also applied the sequence-based clustering approach on the pair-wise NWED distances between segmented speed profiles of van trajectories for *ZoneData*. However, we could not relate the obtained clusters to any specific time periods. This can be explained with the fact that the *ZoneData* trajectories were obtained on a street network heterogeneous with respect to geometry and traffic. Therefore, the interpretation of a sequential movement behavior would not be valid for such networks. In addition, in the contrary to the *RouteData*, we did not have any information about the geographic context of the trajectories (i.e. traffic lights, delivery locations, type of street etc.) in order to make plausible assumptions to validate the results.

6. Discussion

Based on the results of our comparative study, we conclude:

(1) When the movement characteristics of objects are highly inconsistent over time, the variability in the segmentation results is high, resulting in a large number of short segments. Therefore, the MPC diversity-based clustering strategy (cf. Section 4.1) from the descriptive statistics of the segments works best for clustering trajectories with heterogeneous movement characteristics, that is, when objects exhibit high diversity in the variation of their movement parameters over time (courier data, in our case).

(2) In contrast, the MPC sequence-based clustering strategy (cf. Section 4.2) is better suited for clustering movement data with pronounced sequential movement behavior (as the hurricanes, in our case). The reason is that here the adapted string matching technique helps to detect similar sequences among segmented profiles.

[1] Transport for London (http://www.tfl.gov.uk/)

Furthermore, the experiments suggest that applying the NWED similarity measure together with the MP class sequence representation of trajectories is better suited to study spatio-temporal behavior of moving objects in comparison to related spatial similarity measures (i.e. the EDM measure proposed by Chen *et al.* 2004). That is, as showed in the first case study, our approach is more effective in extracting the similarities in movement data w.r.t the evolution patterns of the objects' movement parameters over time (e.g. the speed behavior of hurricanes) in comparison to the method by Chen *et al.* (2004). Although in the experiments by using hurricane data at a regular sampling rate, we implicitly involved time in the computation of EDM to make it more comparable to NWED, the study showed that NWED is better suited to capture the spatio-temporal aspects of hurricane evolution. On the other hand, applying the geometric parameters such as turning angle, NWED provides comparable results to EDM, both being geometric similarity measures.

Representing trajectories with a symbolic representation such as movement parameter class sequences (our approach), or movement pattern strings (MPS, Chen *et al.* 2004) significantly reduces the storage costs of trajectory data (e.g. to 12.5 % as shown in Chen *et al.* 2004) Moreover, the sequence representation is invariant to rotation and spatial transformations since the proposed segmentation algorithm relies on relative movement parameters computed between consecutive fixes along trajectories. The advantage of our proposed representation over MPS is that the number of classes and hence the domain of the sequences is much smaller (i.e. 9 classes in our case, 64 classes in MPS). As shown in Du Mouza *et al.* (2006), the pattern matching and retrieval costs are less for strings that are represented with a small number of characters (i.e. class labels in our case), especially in very large datasets. Moreover, similar to other edit-distance based approaches, our approach can deal with trajectories of unequal length and unequal sampling rate as well.

On the other hand, the string matching process is relatively slow and depends on the length of the profiles (i.e. $O(n^2)$). Therefore, the proposed methodology is computationally expensive for very large trajectory datasets with long trajectories. This issue is common to all edit distance-based approaches (e.g. NWED and EDM). In order to reduce the computational cost of similarity computation, Chen *et al.* (2004) proposed a *Modified Frequency Distance (MFD)* for frequency vectors that are obtained from movement pattern strings. This method is similar to our diversity-based clustering approach. However, in the diversity-based approach we employ more features in addition to the frequency of classes to describe characteristics of movement parameters. Besides, similar to diversity-based clustering, MFD does not preserve the sequence of a movement. Another strategy to overcome this problem is to apply pruning approaches prior to the similarity computation (Chen *et al.* 2005).

Our proposed approach is originally developed for movement parameter profiles. However, the presented methodology can be applied for similarity analysis and clustering of other types of time series (since MP profiles can be seen as a specific type of time series). A similar method has been proposed for addressing threshold queries as well as similarity analysis in time series databases by Aßfalg *et al.* (2008). However, their method only considers the deviation (amplitude) of the time series. In contrast, our approach can handle the frequency of variations by considering the sinuosity of time series.

Finally, our study suggests there can hardly be a universally applicable similarity measure for movement trajectories. Depending on the application at hand, the best suited similarity measure and the adequate movement parameter must be chosen. Background knowledge about the investigated movement process helps making an informed choice. Such knowledge can come in the form of knowledge about the geographic context em-

bedding the movement (as in the case of the traffic light in Experiment #3).

7. Conclusions and Future Work

In this paper, we introduce a new methodology for trajectory similarity detection, which moves similarity analysis beyond considering merely the geometric similarity of trajectories, towards considering movement dynamics. The method bases on the segmentation of movement parameter profiles of an object over time, which can be derived from trajectory data or directly observed using data from tracking sensors.

In our paper, we present a comparative evaluation, to show the usefulness of our approach in clustering both hurricane and courier trajectories. Besides, we experimentally evaluate our approach in comparison to a relevant method by Chen *et al.* (2004). The experiments show that the proposed NWED similarity measure together with the MP class sequence representation of trajectories can be successfully applied in movement behavior analysis for finding structure in movement datasets. Particularly, when objects exhibit a common movement path, our approach becomes more effective in comparison to the available spatial similarity measures. We demonstrate that taking into account the domain and frequency variation of movement parameters can help identifying interesting patterns in the movement of objects.

Contrary to most existing work, the proposed similarity assessment approach focuses on the parameters describing the dynamic characteristics of movement, and does not deal with any geo-spatial or geometry-based similarity. As part of our future work, we plan to develop a combined approach that will integrate the proposed methodology with spatial/geometrical similarity analysis. Depending on the requirements of a particular application, one technique could be used as a filtering/pruning stage for the other. For example, in the second case study, we pruned courier subtrajectories to those along a selected route prior to the segmentation. Furthermore, this study shows that segmentation is a useful technique in extracting the structure of trajectories and assist knowledge discovery in movement data. However, our segmentation applies only one movement parameter at a time. As a future extension of our approach, we intend to develop a trajectory segmentation technique using multiple movement parameters. Moreover, we also intend to enrich the developed approach by incorporating contextual data. Finally, to alleviate the computational complexity of the proposed method, we are currently developing a multi-scale pruning procedure, working from coarse to finer spatial and temporal granularities.

References

Alt, H., 2009. The Computational Geometry of Comparing Shapes. *In*: S. Albers, H. Alt and S. Näher, eds. *Efficient Algorithms.*, Vol. 5760 of *Lecture Notes in Computer Science* Springer Berlin, Heidelberg, 235–248.

Anagnostopoulos, A., *et al.*, 2006. Global distance-based segmentation of trajectories. *In*: *Proceedings of the 12th ACM SIGKDD international conference on Knowledge discovery and data mining* ACM, 34–43.

Aßfalg, J., *et al.*, 2008. Similarity Search in Multimedia Time Series Data Using Amplitude-Level Features. *In*: *Advances in Multimedia Modeling*, 123–133.

Bozkaya, T., Yazdani, N., and Özsoyolu, M., 1997. Matching and indexing sequences of

different lengths. *In*: *Proceedings of the sixth international conference on Information and knowledge management - CIKM '97* New York, New York, USA: ACM Press, 128–135.

Buchin, K., *et al.*, 2009. Finding long and similar parts of trajectories. *In*: *Proceedings of the 17th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems - GIS '09* New York, New York, USA: ACM Press, 296–305.

Buchin, M., *et al.*, 2010. An algorithmic framework for segmenting trajectories based on spatio-temporal criteria. *In*: *Proceedings of the 18th SIGSPATIAL International Conference on Advances in Geographic Information Systems*, GIS '10 New York, NY, USA: ACM, 202–211.

Chen, L., Özsu, M.T., and Oria, V., 2004. Symbolic Representation and Retrieval of Moving Object Trajectories. *In*: *MIR '04: Proceedings of the 6th ACM SIGMM international workshop on Multimedia information retrieval* In Proc. of the ACM SIGMM international workshop on multimedia information retrieval, 227–234.

Chen, L., Özsu, M.T., and Oria, V., 2005. Robust and fast similarity search for moving object trajectories. *In*: *SIGMOD '05: Proceedings of the 2005 ACM SIGMOD international conference on Management of data* New York, NY, USA: ACM, 491–502.

Ding, H., Trajcevski, G., and Scheuermann, P., 2008a. Efficient Similarity Join of Large Sets of Moving Object Trajectories. *In*: *15th International Symposium on Temporal Representation and Reasoning*, Jun.. Ieee, 79–87.

Ding, H., *et al.*, 2008b. Querying and mining of time series data: experimental comparison of representations and distance measures. *Proc. VLDB Endow.*, 1 (2), 1542–1552.

Dodge, S., Weibel, R., and Forootan, E., 2009. Revealing the physics of movement: Comparing the similarity of movement characteristics of different types of moving objects. *Computers, Environment and Urban Systems*, 33 (6), 419–434.

Dodge, S., Weibel, R., and Lautenschütz, A.K., 2008. Towards a taxonomy of movement patterns. *Information Visualization*, 7 (3-4), 240–252.

Doherty, S., *et al.*, 2001. Moving beyond observed outcomes: Integrating Global Positioning Systems and interactive computer-based travel behaviour surveys. *Transportation Research E-Circular, C*, 26, 449–466.

Du Mouza, C., Rigaux, P., and Scholl, M., 2006. On-line Aggregation and Filtering of Pattern-based Queries. *In*: *Scientific and Statistical Database Management, 2006. 18th International Conference on* IEEE, 333–342.

Du Mouza, C., Rigaux, P., and Scholl, M., 2007. Parameterized pattern queries. *Data & Knowledge Engineering*, 63 (2), 433–456.

Elsner, J. and Kara, A., 1999. *Hurricanes of the North Atlantic: Climate and society.* New York Oxford: Oxford University Press.

Etienne, L., Devogele, T., and Bouju, A., 2010. Spatio-Temporal Trajectory Analysis of Mobile Objects Following the Same Itinerary. *In*: *Proceedings of the International Symposium on Spatial Data Handling (SDH)*, Hong Kong.

Faloutsos, C., *et al.*, 1997. A signature technique for similarity-based queries. *In*: *Proceedings. Compression and Complexity of SEQUENCES 1997 (Cat. No.97TB100171)* IEEE Comput. Soc, 2–20.

Frentzos, E., Gratsias, K., and Theodoridis, Y., 2007. Index-based Most Similar Trajectory Search. *In*: *ICDE 2007. IEEE 23rd International Conference on Data Engineering*, Apr.., 816–825.

Giannotti, F. and Pedreschi, D., 2008. *Mobility, Data Mining and Privacy.* Berlin Heidelberg: Springer-Verlag.

Kisilevich, S., *et al.*, 2010. *In*: *Spatio-Temporal Clustering : a Survey.*, Italy.

Laube, P., *et al.*, 2007. Movement beyond the snapshot Dynamic analysis of geospatial lifelines. *Computers, Environment and Urban Systems*, 31 (5), 481–501.

Levenshtein, V., 1966. Binary codes capable of correcting deletions, insertions, and reversals. *Soviet Physics-Doklady*, 10, 707–710.

Lin, B. and Su, J., 2005. Shapes based trajectory queries for moving objects. *GIS '05: Proceedings of the 13th annual ACM international workshop on Geographic information systems*, 21–30.

Little, J.J. and Gu, Z., 2001. Video retrieval by spatial and temporal structure of trajectories. *In*: *Proceedings of SPIE, the International Society for Optical Engineering* Spie, 545–552.

Malkin, W. and Holzworth, G.C., 1954. Hurricane Edna, 1954. *Monthly Weather Review*, 82 (9), 267–279.

Miller, H.J. and Han, J., 2009. *Geographic Data Mining and Knowledge Discovery.* Second Taylor & Francis Group.

Pelekis, N., *et al.*, 2007. Similarity Search in Trajectory Databases. *In*: *TIME'07: Proceedings of the 14th International Symposium on Temporal Representation and Reasoning*, Jun.. Washington, DC, USA: IEEE Computer Society, 129–140.

Sinha, G. and Mark, D.M., 2005. Measuring similarity between geospatial lifelines in studies of environmental health. *Journal of Geographical Systems*, 7 (1), 115–136.

Trajcevski, G., *et al.*, 2007. Dynamics-aware similarity of moving objects trajectories. *In*: *GIS '07: Proceedings of the 15th annual ACM international symposium on Advances in geographic information systems* New York, NY, USA: ACM, 1–8.

van Kreveld, M. and Luo, J., 2007. The definition and computation of trajectory and subtrajectory similarity. *In*: *GIS '07: Proceedings of the 15th annual ACM international symposium on Advances in geographic information systems* New York, NY, USA: ACM, 1–4.

Vlachos, M., Gunopulos, D., and Kollios, G., 2002a. Discovering similar multidimensional trajectories. *In*: *ICDE '02: Proceedings 18th International Conference on Data Engineering* IEEE Computer Societ, 673–684.

Vlachos, M., Gunopulos, D., and Das, G., 2004. Rotation invariant distance measures for trajectories. *In*: *KDD '04: Proceedings of the 2004 ACM SIGKDD international conference on Knowledge discovery and data mining* New York, New York, USA: ACM Press, 707 – 712.

Vlachos, M., Gunopulos, D., and Kollios, G., 2002b. Robust similarity measures for mobile object trajectories. *In*: *DEXA'02: Proceedings of the 13th International Workshop on Database and Expert Systems Applications* Los Alamitos, CA, USA: IEEE Computer Society, 721 – 728.

Yan, Z., *et al.*, 2010. A Hybrid Model and Computing Platform for Spatio-Semantic Trajectories. *In*: *The Semantic Web: Research and Applications* Springer, 60–75.

Yanagisawa, Y., Akahani, J.i., and Satoh, T., 2003. Shape-Based Similarity Query for Trajectory of Mobile Objects. *In*: *MDM '03: Proceedings of the 4th International Conference on Mobile Data Management* London, UK: Springer-Verlag, 63–77.

Yujian, L. and Bo, L., 2007. A normalized levenshtein distance metric. *IEEE transactions on pattern analysis and machine*, 29 (6), 1091–1095.

Research Paper 4

Dodge, S., Weibel, R., and Laube, P. (2011). Trajectory Similarity Analysis in Movement Parameter Space. *GISRUK 2011*, April 27-29, 2011, University of Portsmouth, UK, pages 270 – 279, Short paper.

Trajectory Similarity Analysis in Movement Parameter Space

Somayeh Dodge, Robert Weibel, Patrick Laube

University of Zurich, Winterthurerstrasse 190, CH-8057 Zurich, Switzerland
Tel. +41 44 63 55131, Fax +41 44 63 56848
{somayeh.dodge | robert.weibel | patrick.laube}@ geo.uzh.ch

ABSTRACT: This paper introduces a similarity analysis method for moving object trajectories. The proposed method assesses the similarity between a set of trajectories in a multidimensional space, whose dimensions are formed by different movement parameters (e.g. position, speed, acceleration, direction), plus time. We investigate the applicability of the proposed method in finding relative movement patterns such as *coincidence* and *concurrence* in the movement of North Atlantic hurricanes.

KEYWORDS: Trajectory similarity analysis, movement parameters, movement patterns, knowledge discovery, moving objects.

1 Introduction

In many domains of science and technology, understanding the collective movement behaviour of dynamic objects (i.e. humans, animals, vehicles, etc.) or processes (e.g. hurricanes) is very important. Nowadays, the advances in positioning technologies provide access to massive amounts of movement data in diverse application domains. The availability of such valuable repositories of movement data requires the development of new knowledge discovery tools in order to extract meaningful information and discover patterns of movement behaviours of mobile objects.

In order to study the dynamic behaviour of objects, it is necessary to observe the movement characteristics along the objects' geospatial lifelines, in addition to the positional information. These characteristics, so called *'movement parameters'* (MP) (Dodge et al., 2008), are divided into two types of 'instantaneous' parameters (i.e. detectable at individual moments) such as position, speed, and acceleration and 'relative' parameters (i.e. measurable over time intervals) such as relative speed, direction, and path sinuosity (Laube et al., 2007; Giannotti and Pedreschi, 2008).

However, in spite of the recent progress in the field of knowledge discovery and data mining (Giannotti and Pedreschi, 2008; Miller and Han, 2009), most of the existing spatio-temporal analysis techniques for moving object data deal only with the *positional* information of the tracked objects over time (i.e. with trajectory *geometry*), and very little attention has been paid to other movement characteristics. The same can be observed in the available literature on similarity analysis of movement data (Vlachos et al., 2002; Chen et al., 2005; Trajcevski et al., 2007; Pelekis et al., 2007; Buchin et al., 2009).

Similarity analysis is crucial in the process of knowledge discovery from movement data. The results of similarity analysis can significantly contribute to other important mobility data

mining tasks, such as *trajectory classification* (Dodge et al., 2009), *trajectory clustering* (Zhang et al., 2006; Chen et al., 2005), or *movement pattern detection* (Gudmundsson and van Kreveld, 2006; Buchin et al., 2008).

The aim of this paper is to propose a spatio-temporal similarity analysis method with the perspective of detecting trajectories with similar dynamic behaviour. That is, the method assesses the similarity of the evolution of objects' movement parameters over time. The technique uses the Euclidean distance in a multidimensional space of movement parameters and can be applied for the detection of the movement patterns *coincidence* (i.e. similar positions over time) and *concurrence* (i.e. similar movement parameters over time). Such patterns occur when a set of objects exhibits a synchronous movement or at least similar movement parameters over a certain duration (Dodge et al., 2008). Similar to our approach, a number of previous studies used the Euclidean distance to assess the similarity of movements of objects (e.g. Yanagisawa et al., 2003; Buchin et al., 2009). However, these methods are based on the *positional* information of trajectories in the space-time cube. Therefore, these techniques can only detect *coincidence* patterns. To the best of our knowledge, no other authors so far have used the Euclidean distance in an n-dimensional movement parameter space. By doing so, however, our method is capable of detecting both *coincidence* and *concurrence* patterns.

2 Methodology

When an object moves about in space, the evolution of its movement parameters over time can be seen as functions over time, so-called *'movement parameter profiles'* (Dodge et al., 2009). The properties of these profiles can be quite different for different object types: Some may be rather smooth, others may express diversity in their evolution. However, when multiple moving objects form particular movement patterns such as *concurrence* or *coincidence*, their movement characteristics to a certain degree exhibit similar trends. Therefore, we may exploit information about the movement parameters of a given type of dynamic object for extracting spatio-temporal similarities among trajectories. Accordingly, in this paper, in comparison between two or more trajectories, *the movement characteristics of objects are considered similar when the evolution of their movement parameters resembles each other over a given period time.*

The methodology used in this study consists of four steps as presented in the following sections: 1) trajectory pre-processing, 2) set up a multidimensional movement parameter space, 3) similarity computation in the MP space, and 4) movement pattern detection.

2.1 Trajectory Pre-processing

Due to the nominal precision and accuracy of the positioning technologies used as well as the influence of the environment and other external factors, usually the raw movement data to some degree contain noise, gaps, and outliers. Therefore, to obtain more reliable trajectories for the purpose of similarity analysis, an initial stage of data cleaning and pre-processing (i.e. filtering, smoothing, resampling, etc.) is recommended, tailored to the peculiarities of the specific application domain. The aim is to eliminate the effects of noise, outliers, and other positioning errors from the raw movement data, and to generate regularly sampled trajectories of the same and synchronised duration.

2.2 Trajectory Representation in the Multidimensional MP Space

In the second stage, in order to assess the similarity between the movements of a set of objects, a required set of movement parameters for all fixes along the trajectories is computed. As an example, Figure 1.a illustrates the computed MP profiles from a sample set of speed (v), acceleration ($accl$), and direction (Az) for a trajectory. Then, a multidimensional MP space is generated from the computed movement parameters for each trajectory (Fig. 1.b). Each of

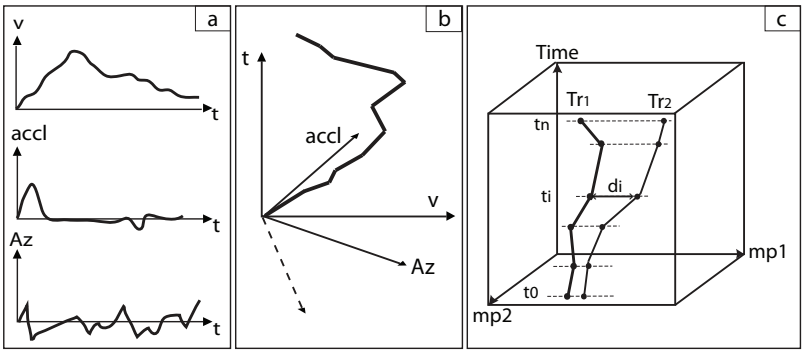


Figure 1: a) Computed MP profiles; b) the MP space; c) similarity computation in a 3D MP space

the movement parameters constitutes one dimension of the multidimensional feature space for trajectories. This provides a multidimensional profile (time series) for each trajectory.

The selection of movement parameters depends on the purpose of the similarity search study, and the type of movement characteristics that one wishes to compare among objects. For instance, speed (i.e. the rate of change of an object's position) and acceleration (i.e. the rate of change of an object's speed) give an indication of how slow or fast, smooth or jerky is the movement. Azimuth (i.e. direction of the movement) and turning angle (i.e. the change of direction) indicate the geometric shape of the trajectory, and the straightness index (i.e. the ratio of the length of the travelled path and the straight-line displacement) gives an indication of the sinuosity of the trajectory at a specific point.

2.3 Similarity Computation in the MP Space

In order to quantify the similarity between two trajectories of the same duration, the average Euclidean distance between the two multidimensional profiles is applied as the similarity measure. Equation 1 presents the computation of the distance between two trajectories Trj_1 and Trj_2, where $|Trj_1| = |Trj_2| = n$:

$$D(Trj_1, Trj_2) = \frac{\sum_{i=1}^{n} d_i}{t_n - t_0} \tag{1}$$

with

$$d_i = \sqrt{\sum_{j=1}^{k-1} (mp_j^{Trj_1} - mp_j^{Trj_2})^2} \tag{2}$$

For the sake of simplicity, Figure 1.c illustrates computing the similarity between two trajectories in a three-dimensional MP space. As shown in the figure, the MP space is generated from a set of two arbitrary movement parameters (i.e. mp_1, mp_2) over time. The distance between two trajectories at timestamp t_i is d_i and is computed as the Euclidean distance between the two points in the k-dimensional MP space (Fig. 1.c and Equation 2). Where k is the number of movement parameters that are considered, with the time dimension added (i.e. movement parameters account for k-1 dimensions, plus time).

The proposed method applies the Euclidean distance, since the complexity of computing this measure is linear (i.e. $O(n)$). The measure is easy to implement and requires no control parameter. Moreover, as shown in Ding et al. (2008), the Euclidean distance can compete with more complex measures such as edit distance and LCSS in very large datasets. However, this measure is sensitive to noise and outliers. Therefore, a pre-processing step is recommended to alleviate this problem.

It is necessary to remark that prior to the similarity computation, movement parameter profiles have been normalised to the scale of one (i.e. $mp_j \in [0-1]$). Therefore, the total dissimilarity value, $D(Trj_1, Trj_2)$, is between 0 - 1. That is, if the total distance equates to one, the subject trajectories are at the maximum dissimilarity or least similarity. In contrast, dissimilarity values less than a small distance threshold (i.e. $0 < D \leq \epsilon$) indicate that the trajectories to a large extent resemble each other. The distance threshold (ϵ) is required in a majority of similarity assessment techniques, is typically application-specific and may vary depending on the purpose of the queries.

2.4 Movement Pattern Detection based on Similarity

Most movement patterns emerge from similarity in one or more movement parameters (Laube, 2005). For example, *concurrence* and *coincidence* movement patterns appear from trajectories of objects that exhibit similar movement characteristics over time (like the ones inside the 'tube' in Fig. 2). Here we suggest using the proposed similarity measure in the detection of such patterns in the MP space. *Concurrence* is defined over a period $[t_0, t_n]$ in the attribute space (e.g. (speed, azimuth, t) in Fig 2.a), whereas *coincidence* is defined over a period $[t_0, t_n]$ in the space-time cube (i.e. (lat , lon, t) in Fig 2.b). The distance between identical trajectories equals zero.

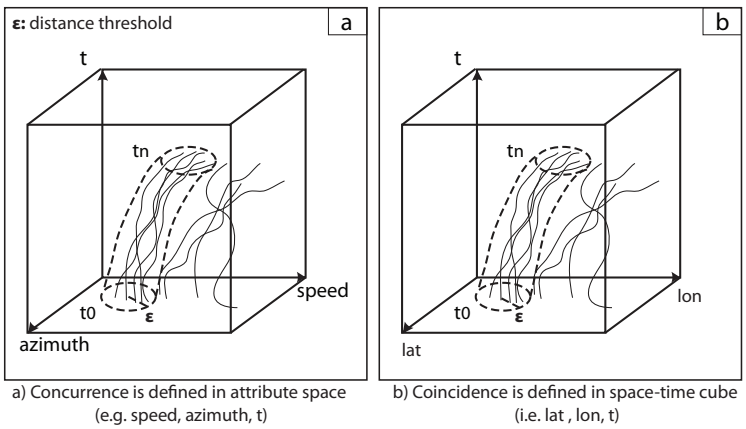


Figure 2: Concurrence and coincidence patterns: trajectories that stay within the threshold region (tube) for a given time period $[t_0, t_n]$ are identified as patterns.

3 Case Study

The proposed methodology can be exploited for clustering trajectory data, and the discovery of spatio-temporal movement patterns, whenever the interest is to detect common movement characteristics of objects over time. In this study, we evaluate the proposed method on 100 years of historical trajectories of North Atlantic hurricanes[1] that occurred between 1907 and 2007.

3.1 Objective

This study uses hurricanes as a test bed for an exemplary proof of concept that the developed method is capable of identifying patterns of similar trajectories such as concurrence and coincidence movement patterns. Hence, this case study intends to seek for such similarity patterns in

[1] from NOAA's Coastal Services Centre (http://csc-s-maps-q.csc.noaa.gov/hurricanes/)

the movement characteristics of hurricanes, specifically, around the time of landfall. According to the meteorological literature, the most critical moment of the movement of a hurricane is at the time of recurvature (i.e. change to a more northerly direction). Moreover, the destruction caused by hurricanes happens at the time and location of the landfall (Elsner and Kara, 1999). Therefore, it is essential to gain knowledge about the behaviour of hurricanes around these two points in their evolution.

3.2 Pre-processing the Hurricane Dataset

For this study, 397 trajectories of North Atlantic hurricanes that made landfall were considered. From each hurricane trajectory a subtrajectory starting from four days before the time of landfall to 1 day after landfall was extracted. Out of the 397 only 167 hurricanes were long enough and coincided with the selected time window. 167 hurricane subtrajectories of the duration of 5 days were thus obtained. The reason for the selection of such data was that we wanted to investigate the hurricane movement patterns from around the recurvature point to shortly after the landfall. Figure 3 shows the obtained subtrajectories in dark blue, which were then used for the similarity analysis experiments. The original hurricane tracks obtained from NOAA (shown in light blue) contain little noise and are regularly sampled (i.e. at a 6 hours interval). Therefore, the pre-processing stage did not involve resampling or smoothing.

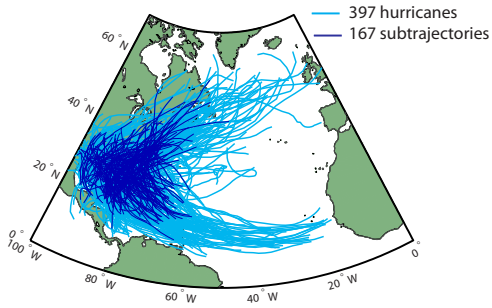


Figure 3: 397 hurricanes with landfall between 1907 and 2007; and the selected 167 hurricane subtrajectories of 5 days duration.

3.3 Similarity Assessment and Movement Pattern Detection

At this stage, for each hurricane subtrajectory the movement parameter space was generated from the speed, acceleration, azimuth, and turning angle profiles. Four (167×167) matrices of the pair-wise distances between trajectories were then computed using the proposed similarity function with different settings of the movement parameters:

1. Latitude – longitude – time: to find hurricanes that followed similar movement path (geometric shape).
2. Speed – azimuth – time: to find hurricanes that moved with similar direction and speed.
3. Speed – turning angle – time: to find hurricanes that generated similar curvature at similar speed.
4. Speed – acceleration – turning angle – time: to find hurricanes that generated similar curvature at similar speed and acceleration.

These distances were then used for the discovery of *concurrence* and *coincidence patterns*. For the purpose of pattern discovery, one arbitrarily selected sample hurricane subtrajectory was

considered as a 'reference (or query) pattern'. Then, the hurricanes which attained a distance less than a small threshold to the 'reference pattern' subtrajectory were extracted.

The *coincidence* patterns were discovered on the 3D space-time cube (Fig. 2.b), computed from the geographic (lat/lon) coordinates of hurricanes over time (i.e. setting (1)), whereas the following settings were investigated for the extraction of the *concurrence* patterns:

3. Concurrence of speed and azimuth over time (Fig. 7)
4. Concurrence of speed and turning angle over time (Fig. 8)
5. Concurrence of speed, acceleration, and azimuth over time (Fig. 9)

Figures 4 and 5, respectively, illustrate the *coincident* subtrajectories (in light blue) that were extracted for two different reference patterns (shown in dark blue). For this case, after running a set of experiments the optimum threshold was set at 0.07. Figure 5 shows the effect of changing the threshold: By reducing the distance threshold from 0.07 to 0.06 four trajectories (the ones that are labelled with $*$) do not match to the reference pattern. This confirms the visual impression that their shape is less similar to the reference pattern.

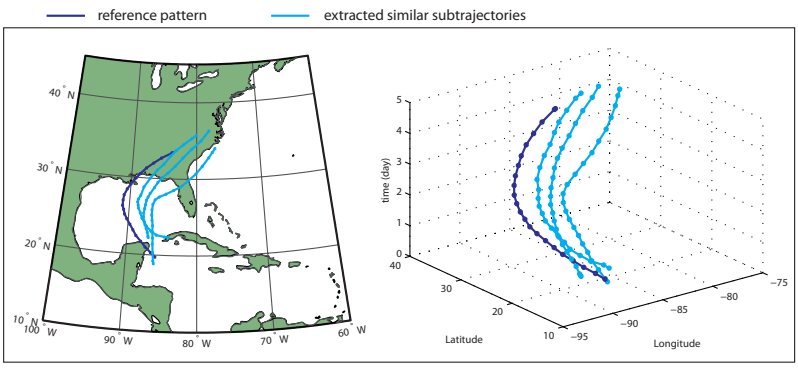


Figure 4: Extracted *coincidences* for a first reference pattern.

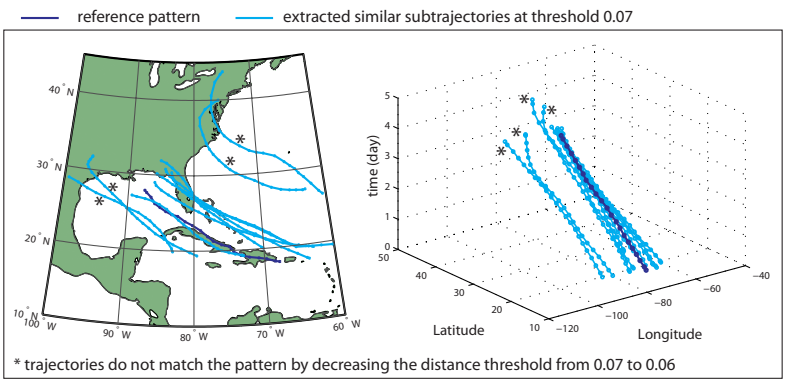


Figure 5: Extracted *coincidences* for a second reference pattern.

Figure 6 illustrates in magenta 10 % of the subtrajectories that did not match the reference pattern of Figure 5 within the selected threshold. This small subset of 10 % of all 'dissimilar' subtrajectories was chosen arbitrarily and simply to avoid over-crowding of the display. As

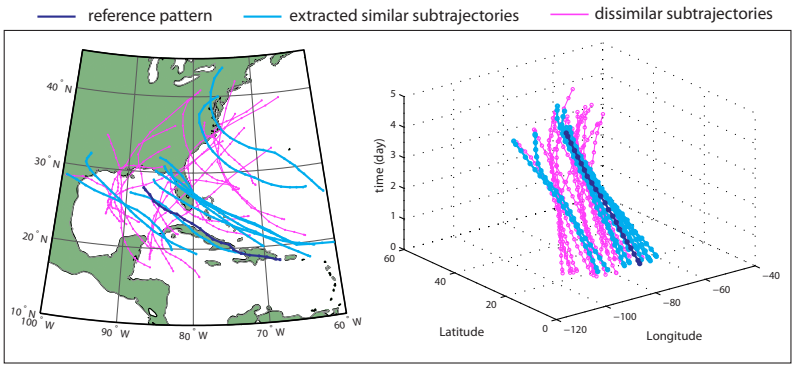


Figure 6: Extracted *coincidences* for the reference pattern of Figure 5.

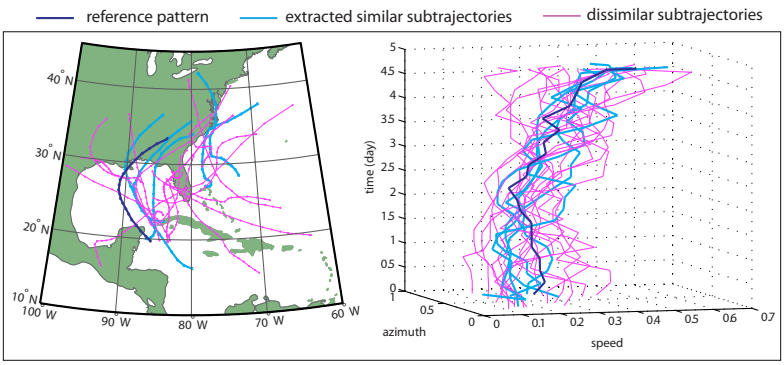


Figure 7: Extracted *concurrences* of 'speed – azimuth – time' for a reference pattern.

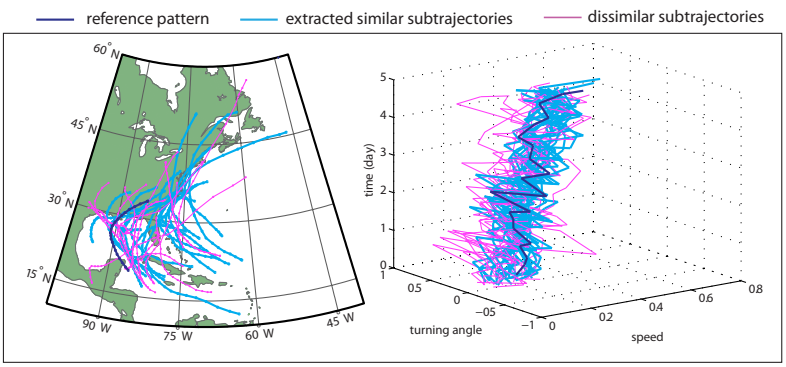


Figure 8: Extracted *concurrences* of 'speed – turning angle – time' for a reference pattern.

it can be observed, the method could successfully distinguish and excludes the subtrajectories with a geometry that is different from the sample pattern. Hence, the results suggest that the proposed similarity assessment method is useful for the detection of *coincidence* patterns. Also, the method facilitates discovering different types of *concurrence* patterns only by changing the setting of the MP space (Figures 7-9). The subtrajectories shown in magenta in the figures are an arbitrary 10 % of the subtrajectories that did not match to reference patterns. Figure 9 does not

visualise the corresponding movement parameter space, since the space was four-dimensional.

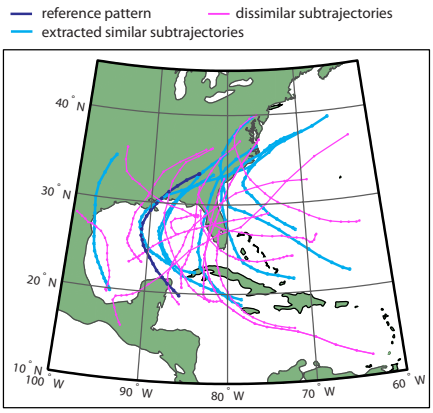


Figure 9: Extracted *concurrences* of 'speed – acceleration –turning angle – time' for a reference pattern.

According to the meteorological literature (Elsner and Kara, 1999), hurricanes that originate from proximate spatial latitudes exhibit similar movement characteristics. As a consequence, meteorologists distinguish two classes of hurricanes: 'low-latitude hurricanes' and the 'high-latitude hurricanes' with respect to the 20° N latitude. The same observation has been obtained from the hurricanes of the same season (Elsner and Kara, 1999). That is, the hurricanes of the early season (April, May, June, and July) to some extent have a dynamic behaviour that sets them apart from the late season hurricanes (August, September, October, and November).

In order to assess whether the obtained patterns confirm the meteorological hypotheses, we computed counts for the extracted hurricane subtrajectories and related these to the time of formation and the latitude of the origin of the hurricanes. The two reference patterns used in this study belonged to two arbitrary late season hurricanes (i.e. occurred in September) with origin locations below the 20° N. Table 1 summarises the counts for the extracted subtrajectories. The outcomes suggest that the extracted hurricanes of similar movement characteristics to a great extent also share the same attributes in terms of their formation time and their locations of origin. This further demonstrates the utility of the proposed technique in hurricane research.

Table 1: Counts of the extracted patterns w.r.t. the time of formation and latitude of origin of the corresponding hurricanes.

	No. extracted similar subtrajectories	**late season**	**latitude $\leq$ 20° N**
(latitude, longitude, time) shown in Figure 4	3	2	2
(latitude, longitude, time) shown in Figure 5 where threshold is 0.06	5	4	5
(speed, azimuth, time) shown in Figure 7	5	3	4
(speed, turning angle, time) shown in Figure 8	19	18	11
(speed, acceleration, turning angle, time) shown in Figure 9	8	7	5

4 Conclusions and Outlook

This paper proposed a simple, yet effective method for assessing the spatio-temporal similarity of the movement of dynamic objects and processes. We evaluated the applicability of the proposed technique through a set of experiments using trajectories of North Atlantic hurricanes. The results suggest that the proposed method can successfully reproduce the existing meteorological knowledge about the movements of hurricanes by the extraction of different movement patterns such as *coincidence* and *concurrence*. The strategy for future work is twofold: First, instead of using arbitrary query patterns, we will use the suggested similarity measures for a systematic search for coincidence and concurrence patterns (i.e. extraction of frequent movement patterns in a large dataset). Second, the influence of the distance threshold will be investigated in a systematic sensitivity study.

References

Buchin, K., Buchin, M., Gudmundsson, J., Löffler, M., and Luo, J. (2008). Detecting commuting patterns by clustering subtrajectories. In *Proceedings of the 19th International Symposium on Algorithms and Computation*, number 642, pages 644–655, Gold Coast, Australia. Springer.

Buchin, K., Buchin, M., van Kreveld, M., and Luo, J. (2009). Finding long and similar parts of trajectories. In *Proceedings of the 17th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems - GIS '09*, pages 296–305, New York, New York, USA. ACM Press.

Chen, L., Özsu, M. T., and Oria, V. (2005). Robust and fast similarity search for moving object trajectories. In *SIGMOD '05: Proceedings of the 2005 ACM SIGMOD international conference on Management of data*, pages 491–502, New York, NY, USA. ACM.

Ding, H., Trajcevski, G., Scheuermann, P., Wang, X., and Keogh, E. (2008). Querying and mining of time series data: experimental comparison of representations and distance measures. *Proc. VLDB Endow.*, 1(2):1542–1552.

Dodge, S., Weibel, R., and Forootan, E. (2009). Revealing the physics of movement: Comparing the similarity of movement characteristics of different types of moving objects. *Computers, Environment and Urban Systems*, 33(6):419–434.

Dodge, S., Weibel, R., and Lautenschütz, A.-K. (2008). Towards a taxonomy of movement patterns. *Information Visualization*, 7(3-4):240–252.

Elsner, J. and Kara, A. (1999). *Hurricanes of the North Atlantic: Climate and society.* Oxford University Press, New York Oxford.

Giannotti, F. and Pedreschi, D. (2008). *Mobility, Data Mining and Privacy.* Springer-Verlag, Berlin Heidelberg.

Gudmundsson, J. and van Kreveld, M. (2006). Computing Longest Duration Flocks in Trajectory Data. In *GIS '06: Proceedings of the 14th annual ACM international symposium on Advances in geographic information systems*, pages 35–42, New York, NY, USA. ACM.

Laube, P. (2005). *Analysing Point Motion, Spatio-Temporal Data Mining of Geospatial Lifelines.* PhD thesis, Universitz of Zurich.

Laube, P., Dennis, T., Forer, P., and Walker, M. (2007). Movement beyond the snapshot Dynamic analysis of geospatial lifelines. *Computers, Environment and Urban Systems*, 31(5):481–501.

Miller, H. J. and Han, J. (2009). *Geographic Data Mining and Knowledge Discovery.* Taylor & Francis Group, second edition.

Pelekis, N., Kopanakis, I., Marketos, G., Ntoutsi, I., Andrienko, G., and Theodoridis, Y. (2007). Similarity Search in Trajectory Databases. In *TIME'07: Proceedings of the 14th International Symposium on Temporal Representation and Reasoning*, pages 129–140, Washington, DC, USA. IEEE Computer Society.

Trajcevski, G., Ding, H., Scheuermann, P., Tamassia, R., and Vaccaro, D. (2007). Dynamics-aware similarity of moving objects trajectories. In *GIS '07: Proceedings of the 15th annual ACM international symposium on Advances in geographic information systems*, pages 1–8, New York, NY, USA. ACM.

Vlachos, M., Gunopulos, D., and Kollios, G. (2002). Discovering similar multidimensional trajectories. In *ICDE '02: Proceedings 18th International Conference on Data Engineering*, pages 673–684. IEEE Computer Societ.

Yanagisawa, Y., Akahani, J.-i., and Satoh, T. (2003). Shape-Based Similarity Query for Trajectory of Mobile Objects. In *MDM '03: Proceedings of the 4th International Conference on Mobile Data Management*, pages 63–77, London, UK. Springer-Verlag.

Zhang, Z., Huang, K., and Tan, T. (2006). Comparison of Similarity Measures for Trajectory Clustering in Outdoor Surveillance Scenes. In *ICPR'06: 18th International Conference on Pattern Recognition*, volume 3, pages 1135–1138, Los Alamitos, CA, USA. IEEE Computer Society.

Biography

Somayeh Dodge is a Ph.D. student and research assistant at the Department of Geography, University of Zurich. Her main research interests are mobility data mining and trajectory similarity analysis. She has been recently awarded a grant for a postdoctoral research on 'context-dependent similarity analysis of movement' at the University of Zurich.

Robert Weibel is a professor of GIScience at the Department of Geography, University of Zurich. He serves as chair of the European research coordination action COST IC0903 Knowledge Discovery from Moving Objects (MOVE), as well as on various journal editorial boards and program committees of conferences in GIScience.

Patrick is a lecturer with the Department of Geography, University of Zurich. Patrick's main research interests are movement analysis, spatio-temporal data mining, and most recently decentralised spatial computing for geosensor networks.

Part III

Appendices

List of Publications

- **Dodge, S.**, Laube, P., and Weibel, R. (in revision, 2011). Movement Similarity Assessment Using Symbolic Representation of Trajectories. *International Journal of Geographic Information Science.*
- **Dodge, S.**, Weibel, R., and Laube, P. (2011). Trajectory Similarity Analysis in Movement Parameter Space. *GISRUK 2011*, April 27-29, 2011, University of Portsmouth, UK, pages 270 – 279, Short paper.
- **Dodge, S.**, Weibel, R. and Laube, P. (2009). Exploring Movement- Similarity Analysis of Moving Objects. The SIGSPATIAL Special, Volume 1, Number 3, November 2009, 11-16. *17th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems.* November 4-6, 2009.
- **Dodge, S.**, Weibel, R. and Forootan, E. (2009). Revealing the physics of movement: Comparing the similarity of movement characteristics of different types of moving objects. *Computers, Environment and Urban Systems*, Volume 33, Issue 6, November 2009, pages 419 – 434.
- **Dodge, S.**, Weibel, R. (2008): Moving Object Trajectory Mining - Trajectory Decomposition Algorithm. *Dagstuhl Seminar on Representation, Analysis and Visualization of Moving Objects*, November 2-7, 2008, Dagstuhl, Germany. Abstract.
- **Dodge, S.**, Weibel, R. and Lautenschütz, A.K. (2008). Towards a Taxonomy of Movement Patterns. *Journal of Information Visualization*, Vol. 7, pp. 240 – 252.
- **Dodge, S.**, Weibel, R. and Lautenschütz, A.K. (2008).Taking a Systematic Look at Movement: Developing a Taxonomy of Movement Patterns. *GeoVisualization of Dynamics, Movement and Change, Workshop at the AGILE 2008 Conference*, May 5, 2008, Girona, Spain. Extended Abstract.
- **Dodge, S.**, Weibel, R. and Fabrikant, S.I. (2007). Spatio-Temporal Pattern Analysis of Moving Point Objects Utilizing Eye Movement Data, *GI-Days 2007*, Munster, Germany. Abstract.

Curriculum Vitae

SOMAYEH DODGE

Research assistant
Geographic Information Systems Division
Department of Geography
University of Zürich
somayeh.dodge @ geo.uzh.ch

EDUCATION

- 2007 - 2011:
  Ph.D. studies, GIS Division, Department of Geography, University of Zürich, Zürich, Switzerland.
- 2003 - 2005:
  Master of Science, GIS Engineering, Department of Geomatics and Geodesy Eng., Khaje Nasir Toosi University of Technology (KNT), Tehran, Iran. M.Sc. thesis: Evaluating and Extending Spatio-Temporal Database Functionalities for Moving Objects, supervised by Dr. Ali A. Alesheikh.
- 1998 - 2003:
  Bachelor of Science, Surveying Engineering, Department of Geomatics and Geodesy Eng., Khaje Nasir Toosi University of Technology (KNT), Tehran, Iran.
- 1994 - 1998:
  Diploma in Mathematics and Physics, Abadan Petrol Refinery Company High School, Abadan, Iran.

SPECIAL QUALIFICATIONS

- **Summer school**, the "1st Summer School on Mobility, Data Mining, and Privacy", co-organized by the FP7/ICT project MODAP "Mobility, Data Mining, and Privacy" (www.modap.org) and the COST Action IC0903 MOVE "Knowledge Discovery from Moving Objects" (http://move-cost.info/), August 27th - November 1st, 2010, Rhodos Island, Greece.

- **Visiting Scholar**, visiting predoctoral fellow in the Department of Electrical Engineering and Computer Science, Northwestern University, Evanston, IL, USA, November 2009.

AWARDS, SCHOLARSHIPS

- Travel grant from the Zürich Graduate School in Geography to attend the ACM SIGSPATIAL GIS 2009 conference in Seattle, November 4-6, 2009.
- Travel grant from the Mentoring Project at the University of Zürich for a scholar visit to the Department of Electrical Engineering and Computer Science, Northwestern University and the Department of Computer Science at the University of Illinois, November 2009.
- Travel grant from the the COST Action IC0903 MOVE "Knowledge Discovery from Moving Objects" (http://move-cost.info/) to attend the "1st Summer School on Mobility, Data Mining, and Privacy", August 27th - November 1st, 2010, Rhodos Island, Greece.
- **Forschungskredit 2010**, Research Awards, from "University of Zürich, Funding for Individuals and Projects" program, for project "CONSIST: CONtext dependent SImilarity Search in Traveler movement", for a Postdoctoral Research at the GIS division of the University of Zürich, from January 1st, 2011 to December 30th 2011.

RECENT POSITIONS

- January 2007 - December 2010:
  Research Assistant, GIS Division, Department of Geography, University of Zürich.
- March 2006 - December 2006:
  Lecturer, Department of Geography, Azad University of Shahre Rey, Tehran, Iran.
- November 2005 - February 2006:
  Teaching Assistant , Principals of GIS module, JKIP (Joint KNT and ITC Program)[2], Department of Geomatics and Geodesy Eng., Khaje Nasir Toosi University of Technology (KNT), Tehran, Iran.
- October 2005 - December 2005:
  Teaching Assistant, GIS softwares, Department of Geomatics and Geodesy Eng., Khaje Nasir Toosi University of Technology (KNT), Tehran, Iran.

[2] JKIP is an international cooperating M.Sc program between Tehran Khaje Nasir Toosi University of technology and Netherlands ITC institute in GIS and Remote Sensing.