Exploring Pedestrian Movement Patterns

Daniel Orellana

Thesis committee

Thesis supervisor
Prof. dr. ir. A.K. Bregt
Professor of Geo-information Science
Wageningen University, The Netherlands

Thesis co-supervisors
Prof. dr. ir. M. Wachowicz
Professor of Geomatics Engineering
University of New Brunswick, Canada

Dr. ir. A. Ligtenberg
Assistant professor, Laboratory of Geo-Information Science and Remote Sensing
Wageningen University, The Netherlands

Other members

Prof. dr. V.R. van der Duim, Wageningen University, The Netherlands
Prof. dr. M.J. Kraak, University of Twente, Enschede, The Netherlands
Dr. Y. Saygin, Sabanci University, Turkey
Dr. H. Skov-Petersen, University of Copenhagen, Denmark

This research was conducted under the auspices of the C.T. de Wit Graduate School for Production Ecology and Resource Conservation (PE&RC).

Exploring Pedestrian Movement Patterns

Daniel Orellana

Thesis

submitted in fulfilment of the requirements for the degree of doctor
at Wageningen University
by the authority of the Rector Magnificus
Prof. dr. M.J. Kropff,
in the presence of the
Thesis Committee appointed by the Academic Board
to be defended in public
on Friday 11 May 2012
at 11 a.m. in the Aula

Orellana, Daniel

Exploring Pedestrian Movement Patterns

140 pages

PhD Thesis, Wageningen University, Wageningen, NL (2012)

With references. with summaries in English, Dutch and Spanish language.

ISBN 978-94-6173-271-2

*Lo importante no es llegar, lo importante es el camino**
–Fito Páez –

*The important thing is not to arrive; the important thing is the path

Contents

a Hernán

1

General Introduction

> *"Scientists, through a growing realization of the existence of pressing social and environmental problems and of the need to adapt their own roles, are becoming more entangled and more involved with the wider society. (...) If theory is to develop, and if we wish to influence policy decisions, the study of spatial behavioural patterns becomes a research priority, because the investigation of these patterns demonstrates the ways in people behave, as constrained by their access to resources, and also, therefore, points to some of the social and spatial inequalities in society." (Eyles, 1971 p242)*

1.1 Background

Occasionally in science, ideas and concepts are formed years before they are fully comprehended and can be applied. The quote at the beginning of this chapter raises the core idea behind a research topic that only started to attract increasing attention thirty or more years later. It also happens to come from the first document to contain a reference to a work that is now inspiring a new generation of geographers, economists, urban planners and computer scientists: Hägerstrand's "What about people in regional science?" (Hägerstrand, 1970). A historical analysis of citations of Hägerstrand's paper shows that it was hardly known outside the academic circle of human geography before 1995 (it had 56 citations in the period 1971–1995). It is only since 1996 that time geography, the main theoretical contribution of the work, has gained renewed momentum, inspiring a wide range of researchers and practitioners in different areas. In fact, the number of citations per year has been steadily increasing for the last fifteen years (Figure 1–1).

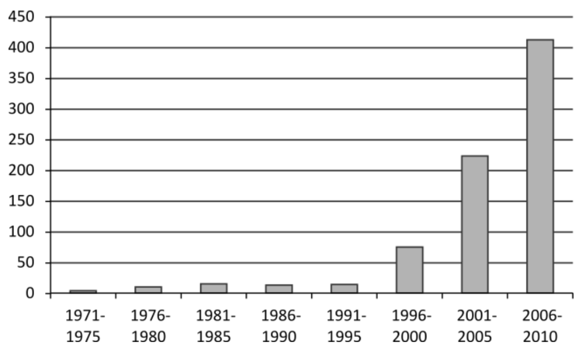

Figure 1–1. Number of citations of Hägestrand (1970) in five-year periods (Data from Google Scholar, retrieved 19 December 2011)

What has caused this renewed interest in human spatial behaviour? Why has it taken thirty years for the research priority pointed out by Eyles in 1971 to be taken up? The answer might be connected with recent technological advances and the development of more multidisciplinary approaches in science.

On the one hand, the combination of location-aware technologies and the increasing popularity of wireless technologies is producing a continuous flow of data about the location in space and time of users, with a spatial and temporal resolution hardly imaginable four decades ago (Miller, 2005; Yu and Shaw, 2007; Lee-Gosselin, Doherty and Shalaby, 2010; Wachowicz, 2010). On the other hand, the advances in databases, data models, algorithms and methods allow researchers to store, manage and analyse vast amounts of data on moving objects (Wolfson et al., 1998; Pfoser, Jensen and Theodoridis, 2000; Pelekis and Theodoridis, 2007; Andrienko and Andrienko, 2008; Spaccapietra et al., 2008). In fact, although movement data is relatively simple at the individual level (e.g., the movement of a person can be represented by a series of spatiotemporal locations), it may involve a substantial level of complexity at the collective level, where structures and patterns emerge as the result of interactions between individuals, collectives and their environment. Hence, the combination of the increasing flow of movement data and the availability of methodological and technological tools to deal with that flow is fuelling a promising research topic. Several areas, such as location-based services, transportation management, urban planning and recreation management, are already taking advantage of the renewed interest and advances in movement analysis.

The exploration and analysis of large movement datasets to detect movement patterns is a fertile research area. Despite the increasing interest in this topic, however, several issues still remain to be investigated. In this thesis, I identify some of these issues and consider them from a geographical point of view. In the following sections, I briefly describe the main elements of this thesis and the related gaps in knowledge in order to define the scope of my research.

1.2 Pedestrian movement in natural recreational areas

In 1970, the year that time geography was born, the influence of the environment on human behaviour was attracting the interest of an increasing number of psychologists who assumed that space is directly linked to behaviour (Ittelson, Rivlin, and Prohansky, 1970). Nowadays, almost no one would disagree with the idea that the surrounding space plays a critical role in the way people behave. Indeed, the U.S. National Library of Medicine (2010) defines "spatial behaviour" as the reactions of an individual or group to the surrounding area, including animate or inanimate objects.

In tourism studies, monitoring and assessing the movement of visitors in natural recreational areas (intra-site flow) is a key issue in understanding visitor behaviour, which in turn is directly applicable to the effective management of both conservational and recreational requirements (Muhar, Arnberger, and Brandenburg, 2002; McKercher & Lau, 2008). In order to understand these requirements and to design, implement and monitor sustainable management practices, detailed information about area usage and the preferences of different target groups is needed (Chiesura, 2004). Since different uses and activities can be related to different places and landscape configurations, the analysis of the spatial behaviour of visitors can provide insights into their preferences and purposes (Golicnik & Ward Thompson, 2010). For example, monitoring the movement of people during their visits to a recreational area can help to identify which are the most or the least visited places, how much time visitors spend in each place and which kind of attractions are preferred by different target groups. If managers know those preferences, they can segment the market and offer more diverse and focused options adapted to the wishes of specific groups of visitors (Holyoak & Carson, 2009). Moreover, monitoring and analysing the area usage and movement of visitors can provide information about potential crowding and conflicts between different groups (Manning & Valliere, 2001; Ostermann, 2009).

Walking is probably the most obvious aspect of spatial behaviour. One fact that is evident and yet frequently forgotten is that people walk not for its own sake, but rather with the intention of stopping. They move for a purpose, usually involving a social or economic transaction. "Spaces encourage stopping, stopping encourages moving" (Stonor, 2001 p3). The assumption here is that spaces should not be seen as areas of transition to be crossed as quickly as possible, but have to be understood as venues, and that the attractiveness and utility of places can be related to the number of people stopping there (Stonor, 2004). In 1984, Hiller had recognized the importance of "static people" as a crucial element of spatial culture. He argued that if people are stationary, it is because something has occurred to make them stop: they have seen something to look at, or have found a place to sit and rest, or have simply taken up a vantage point (Hillier, 1984). Just as silence is the keystone in the structure of music, stops are the building blocks of pedestrian movement.

Therefore, a central assumption in this thesis is that pedestrian movement is the result of the interactions between people and their environment. Consequently, we can explore and analyse pedestrian movement data to detect patterns that are evidence of such interactions.

Despite the recognized importance of the close relationship between pedestrian movement and space, how it should be represented and formalized is still an open issue. A conceptual framework that takes into account the main elements of pedestrian movement behaviour (patterns, interactions and environment) has not been developed yet. Such a framework would allow us to represent those elements and, more importantly, conceptualize the relationships between those elements to ultimately understand pedestrian movement from a geographical point of view.

1.3 Exploratory spatial data analysis

Besides the development of a conceptual framework, the study of pedestrian movement presents additional challenges. In fact, exploring and analysing large amounts of movement datasets constitute the core of the ongoing research. Exploratory Spatial Data Analysis (ESDA) (Anselin, 1993) focuses on the spatial aspects of the data, such as spatial association and spatial heterogeneity and aims to "describe spatial distributions, discover patterns of spatial association (spatial clusters), suggest different spatial regimes or other forms of spatial instability (non-stationarity) and identify atypical observations (outliers)" (Anselin, 1993, p. 114). Haining (2003, p. 183) stated that the primary goals of ESDA are "to summarize spatial properties, to detect patterns, and to formulate hypotheses from geographic data using methods that make minimal data assumptions and which are numerically and graphically resistant to the impact of isolated outliers".

Whereas traditional statistics are not suitable for spatial analysis due to the intrinsic spatial association of georeferenced data (spatial data is hardly independent), ESDA can take advantage of this property to explain the extent to which the data is autocorrelated in space (spatial autocorrelation). A number of indexes are available in the literature for exploring spatial autocorrelation in geographical data. Some examples include the Geary Ratio *c* (Cliff and Ord, 1972; 1981), Getis's *G* or *O* Index (Getis and Ord, 1992), Kulldorff's scan statistic (Kulldorff, 1997) and the most widely used, Moran's coefficient (*I*) (Moran, 1948). All of these indexes summarize the global properties of spatial autocorrelation in a dataset. In other words, they indicate the presence or absence of a spatial pattern for the entire dataset (i.e., global statistics). Anselin (1995) developed the Local Index of Spatial Association (LISA), which is based on the decomposition of Moran's *I* into its local version (i.e., local statistics). Because Moran's *I* is a global summation of individual statistics, LISA uses this property to evaluate the spatial association by calculating the Local Moran's *i* and evaluating the statistical significance of each unit. Therefore, LISA plays a critical role in ESDA since it allows for the exploration of data properties in a global context of a dataset.

Although ESDA, and specifically LISA, has been applied to different research areas, its feasibility for pedestrian movement analysis remains unexplored. Specifically, it is still not known whether or not spatial structures (i.e., spatial association) exist in pedestrian movement data. If they exist, we can use ESDA methods to detect them and formulate geographical hypotheses about movement. The detected structures in the data are movement patterns and the geographical hypothesis should aim to explain the origin of the patterns by examining the interaction between pedestrians and the geographical environment.

1.4 Movement patterns

Recent advances in human movement analysis suggest that despite the wide variety of potential movement behaviour, people usually follow simple and predictable movement patterns (Gonzalez, Hidalgo, and Barabasi, 2008; Song et al., 2010). Whereas it is accepted that these patterns may help in understanding the interactions between people and their environment (Batty, DeSyllas, and Duxbury, 2003; Bierlaire, Antonini, and Weber, 2007; Gudmundsson, Laube, and Wolle, 2009; Hoogendoorn and Bovy, 2005), there is no real consensus on what exactly movement patterns are and how they should be defined (Laube, 2009). Some examples of movement patterns reported in the literature are "flocking" (Gudmundsson, van Kreveld, and Speckmann, 2004), "encounter" (Laube, Kreveld, and Imfeld, 2005), "trend-setter" (Laube, Imfeld, and Weibel, 2005), and "leading and following" (Andersson et al., 2008). Looking at the large diversity of movement patterns reported in the literature, some authors have proposed formalization and classification systems (Dodge, Weibel, and Lautenschütz, 2008; Wood and Galton, 2009). Although these efforts have not been broadly adopted yet, they constitute an attempt to provide a systematic framework for the ongoing research.

A large proportion of the research in movement patterns comes from the field of computer sciences, where researchers are developing algorithms, models and tools to store and provide access to large movement datasets (Pfoser and Jensen, 2001; Güting and Schneider, 2005; Manco et al., 2008; Ortale et al., 2008; Renso et al., 2008), and to extract and visualize movement patterns (Andersson et al., 2008; G. Andrienko et al., 2008). Despite the increasing number of researchers working on movement pattern analysis, most of the current approaches are based on the geometric properties of trajectories, and little attention is paid to the fact that movement is essentially a spatial phenomenon and could be studied from a geographical point of view. Spatial properties of movement data, such as spatial association, remain unexplored. What is needed is a geographical approach to detecting and exploring movement patterns that aims to identify spatial structures in movement data (movement patterns) and to explore and explain the properties of these patterns, taking into account their geographical environment.

A geographical approach to movement patterns therefore implies that movement patterns can be detected and analysed as a geographical phenomenon. Consequently, I adopted the definition of movement patterns provided by Laube (2009) and adapted it to follow a geographical approach in the context of this research:

> *A movement pattern is a detectable structure in the data that constitutes a high-level description of the movement of an individual or a group of individuals resulting from their interactions with their environment.*

This definition entails some assumptions. The first assumption is that movement patterns are *observable*; this means that they can be detected or extracted from (raw) movement data through ESDA techniques. The second assumption is that movement patterns are *describable*, that their structure and properties can be described and represented. The third assumption is that patterns are *explainable*, in the sense that they can be explained and interpreted as the result of some interactions between the pedestrians and their environment. Using this definition, the scope of this thesis is defined as the study of movement patterns of pedestrians from a geographical point of view.

1.5 Objectives

The main objective of this thesis is to develop an approach for exploring, analysing and interpreting movement patterns of pedestrians interacting with the environment. This objective is broken down in sub-objectives related to four research questions. A case study of the movement of visitors in a natural area is used to develop and demonstrate the approach.

To achieve the objectives, four research questions were formulated:

- How can movement patterns evidencing the stopping behaviour of pedestrians be detected?
- What is the validity of the detected movement patterns for describing stopping behaviour of pedestrians?
- How can movement patterns be applied to study the movement behaviour of visitors in natural areas?
- How can movement patterns be formalized to represent the interactions between pedestrians and between pedestrians and their environment?

1.6 Structure

This thesis consists of six chapters, including this introductory chapter. The structure of the chapters is depicted in Figure 1-2.

Chapter 2 presents an exploratory spatial analysis approach to detect patterns of movement suspension using a Local Indicator of Spatial Association (LISA). These patterns are used to find places where pedestrians stop as a consequence of their interactions with geographical features.

Chapter 3 presents the results of a controlled experiment to investigate the validity of using Movement Suspension Patterns (MSPs) to represent the stopping behaviour of visitors in the Dwingelderveld National Park (the Netherlands). The detected MSPs are compared in space and time with a set of reference stops to assess the accuracy of the method.

Chapter 4 demonstrates how movement patterns can improve our understanding of the aggregated movement of visitors in natural recreational areas. The approach is demonstrated by detecting Suspension Patterns and Generalized Sequential Patterns in a dataset representing the movement of visitors in the Dwingelderveld National Park. Both patterns were analysed in their geographical context to characterize the aggregated flow of people to provide insights into visitors' movement behaviour.

Chapter 5 presents a framework to represent and formalize the main concepts of pedestrian movement (i.e., patterns, interactions and spatial behaviour). This framework constitutes an approach to formally representing those concepts and the relationships between them.

Chapter 6 discusses the results of this thesis in relation to the research questions. It also presents reflections on the implications of these results in the context of movement pattern analysis research.

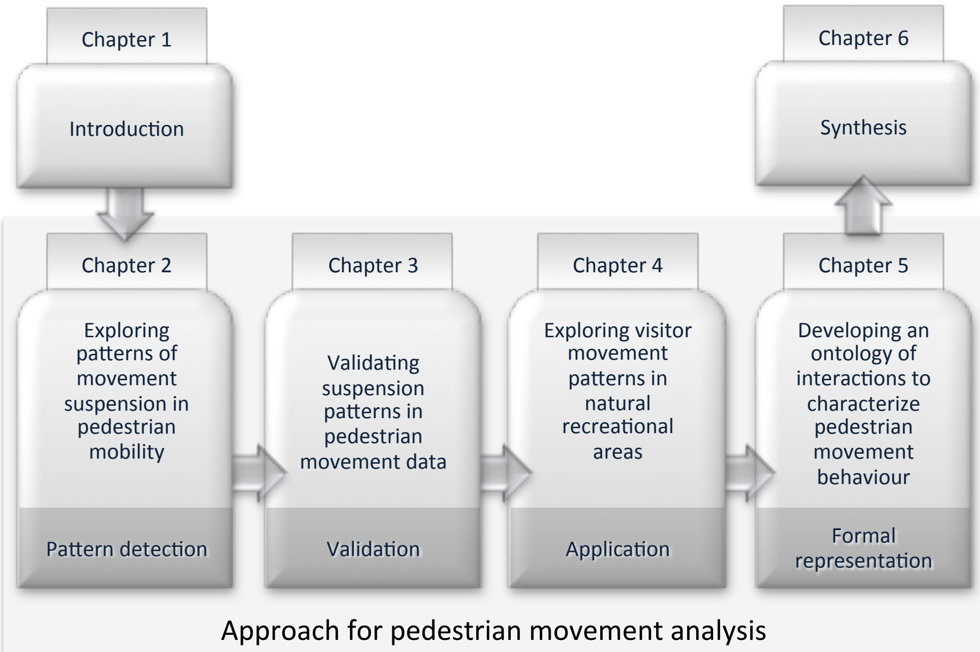

Figure 1–2. Outline of the thesis

2

Exploring Patterns of Movement Suspension in Pedestrian Mobility

Orellana, D. and Wachowicz, M. (2011)
Geographical Analysis 43(3): 241-260

Abstract: One of the main tasks in analysing pedestrian movement is to detect places where pedestrians stop, since those places usually are associated with specific human activities, and they can allow us to understand pedestrian movement behaviour. Very few approaches have been proposed to detect the locations of stops in positioning datasets, and they often are based on selecting the location of candidate stops as well as potential spatial and temporal thresholds according to different application requirements. However, these approaches are not suitable for analysing the slow movement of pedestrians where the inaccuracy of non-differential GPS commonly used for movement tracking is so significant that it can hinder the selection of adequate thresholds. In this paper, we propose an exploratory statistical approach in order to detect patterns of movement suspension using a local index of spatial association (LISA) in a vector space representation. Two different positioning data sets are used to evaluate our approach in terms of exploring movement suspension patterns that can be related to different landscapes: players of an urban outdoor mobile game, and visitors of a natural park. The results of both experiments show that patterns of movement suspension were located at places such as checkpoints in the game, and different attractions and facilities in the park. Based on these results, we conclude that using LISA is a reliable approach for exploring movement suspension patterns that represent the places where the movement of pedestrians is temporally suspended by physical restrictions (e.g., checkpoints of a mobile game, and the route choosing points of a park).

2.1 Introduction

The pervasive nature of GPS technology embedded in mobile phones, watches, PDAs, pedometers, and other wearable devices is producing personal mobility information such as routes taken, distances travelled, as well as the timing, duration and speed of movement. Usually this information is stored as a sequence of points over a time interval representing the trajectory followed by a particular moving entity. Traditionally, the trajectory of a moving entity has been modelled as a polyline in three-dimensional space. Trajectories have been used to represent geospatial lifelines (Hornsby and Egenhofer, 2002), space-time prisms

(Hägerstrand 1970; Miller, 2005) and paths (Pfoser et al., 2000, Andersson et al., 2008). A complementary approach also has been proposed to encompass a relative space view of movement, centred to the relative locations and relative velocities of neighbouring moving entities based on the characterization of elementary trajectories and trajectory transitions (Noyon et al., 2007).

The analysis of pedestrian mobility always attracted scientists to address a variety of issues, such as human behaviour in panic and evacuation situations (Galea, 2003; Helbing et al., 2005; Zheng et al., 2009), urban planning and architecture design (Fruin, 1971; Pauls, 1984; Horner and O'Kelly, 2001;van der Spek, 2006), transportation management (Hoogendoorn and Bovy, 2005; Daamen, 2004), diseases dispersion and epidemic studies (Bian, 2004; Colizza et al., 2007), and, more recently, location-based services (Mountain and Raper, 2001; Li and Hodgson, 2004; Millonig and Gartner, 2007). One of the main tasks in these analyses is to detect the locations where the movement of pedestrians come to a halt, because they usually represent the places where pedestrian motion is temporally suspended by physical, psychological, or social restrictions (e.g., traffic lights, cross roads, and decision making points).

In this paper, we propose a new approach to detect "movement suspension" in pedestrian positioning datasets. We use the term "movement suspension" in contrast to "stop" to point out the reduction of speed associated with stopping behaviour even when pedestrians are not completely still, or when their slow movement is indistinguishable from GPS inaccuracies. Our research premise is that the reduction of speed can be analysed based on a vector space representation. Most of the current approaches are based on the segmentation of the trajectories of a moving entity into "stops" and "moves," where stops are segments of the trajectory in which the movement ceases to occur. Spaccapietra et al. (2008) propose a semantic formalization for stops as a part of a trajectory where a moving entity does not move. The spatial range of a trajectory for a time interval is a single point, and an analyst needs to explicitly define this part of the trajectory. Three methods have been proposed to detect stops based on this semantic assumption. First, Alvares et al. (2007) propose a method called SMoT (Stops and Moves of Trajectories) for detecting stops based on the analysis of the intersection of trajectory segments with previous known geographical features that are candidates of stops within a particular application. In this case, if the duration of an intersection exceeds a pre-defined threshold, the trajectory segment is considered a stop (Figure 2–1a). Second, Rinzivillo et al. (2008) propose a similar approach where the stops are those segments of trajectories in which a moving entity keeps its position within a distance threshold for a minimum period of time (Figure 2–1b). Finally, Palma et al. (2008) propose a method called CB-SMoT (Clustering-Based Stops and Moves of Trajectories) that analyses each trajectory and generates stops when the speed value is lower than a given threshold for a minimal amount of time (Figure 2–1c).

In any of these conceptualizations of stops, some kind of spatial and/or temporal threshold is necessary depending on the application requirements. Moreover, some a priori knowledge about the collection of positioning data, the nature of human activities and landscape also is required to determine the candidate places where stops might occur. For example, when analysing commuter trips, stops could be considered as "Madrid" and "Toledo", "home" and "work place," or even "gas station" and "grocery store." In all of these cases, the thresholds must correspond with the minimum time that moving entities are expected to spend in these particular places. For moving in a car from Madrid to Toledo, or cycling from home to a work place, the meaning of stops and their respective space and time thresholds is significantly different.

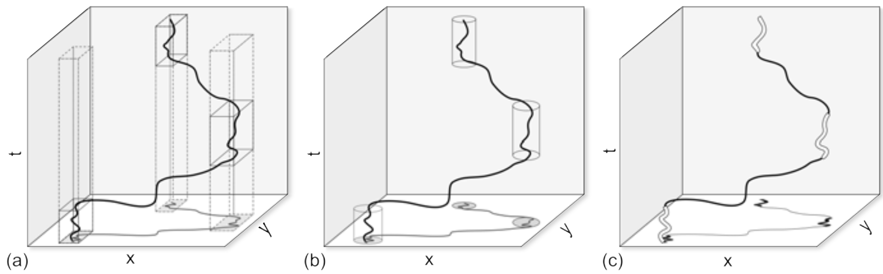

Figure 2–1. Three different conceptualizations of stops previously proposed in the literature: (a) intersection, (b) displacement, and (c) speed.

Additionally, because most positioning data are provided by GPS systems, a critical implementation issue for defining stops is related to movement parameters (i.e., speed, displacement and spatial position), which are highly sensitive to intrinsic measurement errors. Intensive pre-processing is required to eliminate such errors in positioning datasets, and, unfortunately, some errors for pedestrian movement are practically indistinguishable from actual slow movement. Despite some studies demonstrating that GPS receivers are sufficiently accurate for the computation of walking speed on relatively straight courses and controlled situations (Witte and Wilson, 2004; Terrier et al., 2000), these devices are less adequate for assessing walking speed on slow movements. For example, with a GPS recording of every 10 seconds and an average speed of 1 meters per second (m/s), the inaccuracy of a spatial position is higher than the actual travelled distance, leading to huge uncertainties.

Finally, consecutive GPS recordings of the position of a pedestrian standing still are quite unlikely to be in the same location, but rather will lie within an area defined by the GPS error circle (the circle inside of which the true horizontal coordinates of a position have a 50% percent probability of being located). Therefore, based on the tracking technology commonly used today, no real gaps of zero movement (speed = 0 or near to zero) exist in GPS recordings. In addition, some errors cause unusually high speeds that are hardly reachable for a pedestrian. Thus, any approach using spatial and or temporal thresholds will bring up the risk of an under- or over-estimation of stops (Figure 2–2).

Therefore, this paper focuses on two main conceptualizations: movement vectors, and their spatial association. The theory about vectors is well developed in physics, and using vectors to analyse the divergence, gradient and curl of vector field constructs in the context of pedestrian movement would be helpful (McQuistan, 1965; Maidment, 1993). The limited use of vector-based representations should be attributed to the scarcity of methods rather than the movement-awareness related applications that could benefit from their use (Li and Hodgson, 2004). One example is Wolfson et al. (1998), in which the authors have proposed a vector-based representation to deal with efficiency issues in moving object databases.

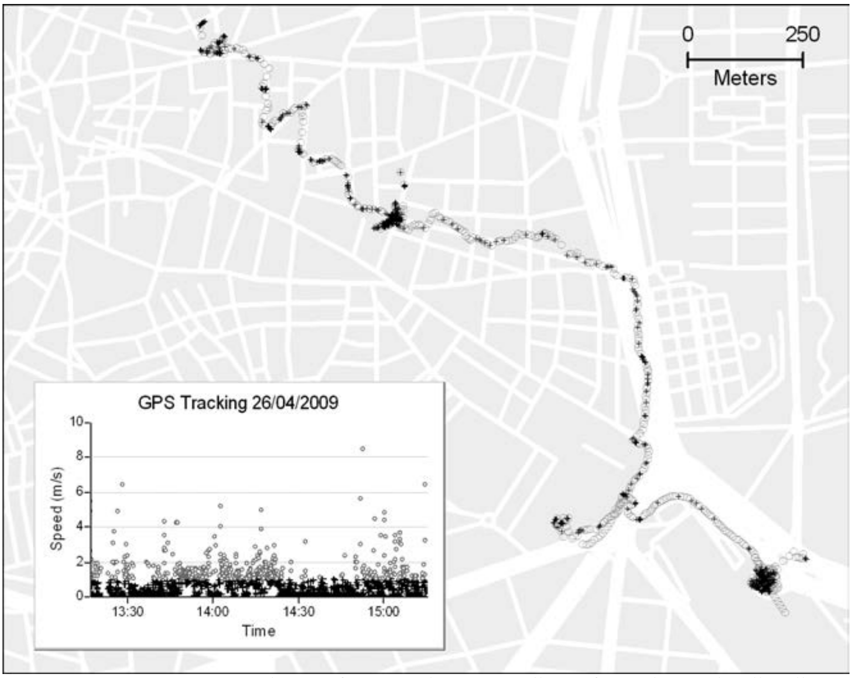

Figure 2–2. GPS recordings of one pedestrian walking for 2 hours hardly allow differentiating real stops from actual slow movement. Background data from www.OpenStreetMap.org.

Our new approach proposes the use of a univariate spatial association index for exploring "movement suspension" in pedestrians positioning datasets. A number of indexes are available in the literature for exploring univariate spatial autocorrelation in geographical data. The most popular ones include the Moran Coefficient (I) (Moran, 1948), the Geary Ratio (c) (Cliff and Ord, 1972; Cliff and Ord, 1981), the Getis's G or O Index (Getis and Ord, 1992), and the Kulldorff's scan statistic (Kulldorff, 1997). The purposes of these three indexes are very similar, and the most often used is the Moran's I because it tends to have the best statistical properties and has been applied for simultaneous measurements from many locations (i.e., global statistics). Anselin (1995) develops the Local Index of Spatial Association (LISA), which is based on the decomposition of Moran's I into its local version (i.e., local statistics).

This paper describes our results in using LISA as a local index of spatial association to compute the spatial association of speed values among movement vectors. The research challenge is to explore the spatial statistics of a vector representation in order to detect the locations of movement suspension patterns. Because our approach follows an exploratory analysis, no a priori knowledge about pedestrian movement is needed; neither spatial nor temporal thresholds are required to be defined based on an application's requirements.

2.2 Movement vectors

A formal definition of a movement vector is a directed line segment from an origin point in the Euclidean space specified by n-dimensions. Although movement vectors do not form a part of any trajectory (Figure 2–3a), they represent the collective movement that can be measured or sensed at one place at one time. They are the "grains" of the collective movement of pedestrians obtained by defining a magnitude that can be measured by a certain number of dimensions, such as spatial and temporal position, speed and orientation, and may be represented graphically by an arrow (Figure 2–3b). Movement vectors imply the conceptualization of movement as a property of space rather than a property of the trajectory of a particular movement entity. Although in this work we only use movement vectors that were observed at a particular location in time, the concept could be expanded using the field-based theory of Time Geography recently proposed by Miller and Bridwell (2009). In this case, vector fields could be generated for representing a continuous space of movement vectors for both observed and unobserved locations.

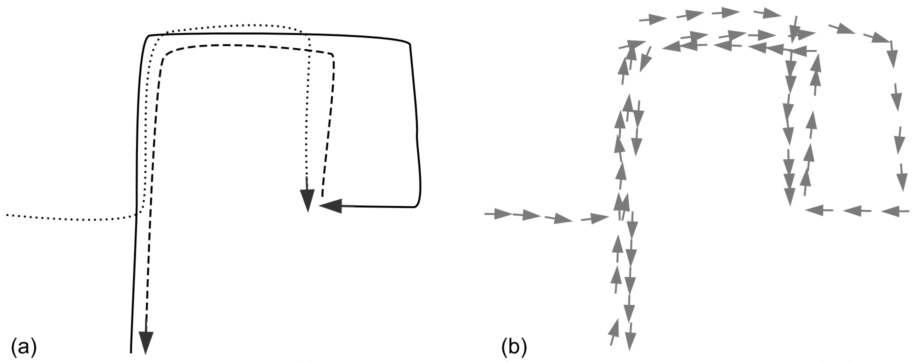

Figure 2–3. Two possible representations of movement: (a) trajectories, and (b) movement vectors.

Positioning datasets usually are recorded as a sequence of triplets of space-time coordinates (x, y, t) organized in subsets (i.e., tracks). Movement vectors comprise a recorded space-time location and computed movement parameters, which are displacement (absolute distance from previous point), time step (absolute time from previous point), speed (displacement over time), and bearing (angle between two consecutive points with respect to true North). Additional movement derivatives, such as acceleration or turn angle, also can be computed. Usually the logging frequency can be configured in three different ways: fixed time rate (a location each x seconds), fixed distance (a location each x meters), or automatic (depending on the speed and turn angle, fewer locations in straight and slow movements and more in sinuous and fast movements).

We implemented a simple computer procedure to read positioning data in diverse standard formats (e.g., gpx, csv, dbf, shp.), and compute movement vectors and store them in a database. The input parameters comprise paths to source files and declaration of four input fields: latitude, longitude, time, and track identification in the case that the data came from different devices. Additionally, the algorithm allows an optional splitting of tracks based on space and/or time. This option is particularly useful because the data usually have large time gaps due to loss of signal or simply because a device was switched off. The algorithm also includes functions to convert geographic data to different projections, and convert temporal data to the correct Time Zone. The implementation code is available at Orellana (2010).

2.3 Spatial association of movement vectors

"...the behaviour in a crowd strongly depends on the behaviour of other persons in the crowd." (Bierlaire et al., 2007, p84)

Spatial heterogeneity and spatial dependence are two well-known properties that are observed in most geographical phenomena. Spatial heterogeneity relates to the global variation of a condition over the Earth's surface, and refers to an observed attribute being expected to vary across locations. In contrast, spatial dependence refers to the persistence of an observed attribute at the local level despite the global variation (de Smith et al., 2009), and it is related to the fact that an observation in one location is also similar to other observations in near locations, and this similarity will decay with increasing the distance.

In pedestrian movement, these two properties can be used to explain the inherent spatial association that takes place when a pedestrian slows down due to a geographical feature in a landscape, or the proximity to other pedestrians. In order to determine spatial association, a local index is needed. In particular, Anselin (1995) developed the LISA. Applied to a set of movement vectors, this local statistic implies that each movement vector gives an indication of the extent of a significant spatial clustering of similar values around that movement vector. Moreover, the sum of local statistics for all movement vectors is proportional (or equal) to a corresponding global statistic (Anselin, 1993). Because Moran's I is a global summation of individual statistics, the LISA uses this property to evaluate the spatial association by calculating Local Moran's I and evaluating the statistical significance of each unit. The spatial relationship between vectors is represented using a matrix that stores the spatial structure of the weighted influence of a neighbourhood (de Smith et al., 2009). Under the assumption that this influence is only dependent on the distance, it can be conceptualized as an inverse distance weighted function.

However, a cut-off distance is necessary in order to include only the neighbours that can actually exert some influence, with the additional advantage of improving the computational performance and storage size of the spatial relationship matrix. Because movement vectors are hardly equally distributed in space, a minimum number of neighbours is established to overcome possible isolated observations and avoid the consequent invalidation of the statistical approach when no observations are present in the neighbourhood (Figure 2–4). Additionally, a row standardization of the spatial relationships matrix is also necessary in order to address this variability (de Smith et al., 2009). The resulting matrix represents the weighted standardized influence of every neighbour inside the cut-off distance for each movement vector.

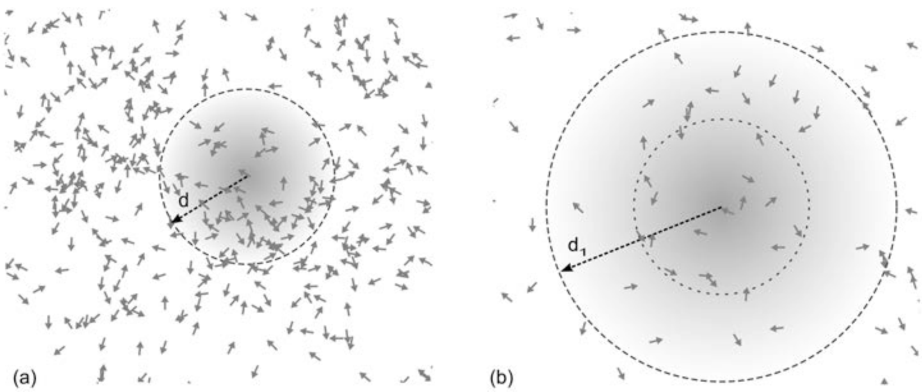

Figure 2–4. Spatial association is conceptualized as a standardized weighted inverse distance function: (a) A cut-off distance *d* limits the neighbour to a defined distance around a movement vector; (b) The neighbourhood can be expanded to *d1* to reach a minimum number *k* of neighbours.

Three values are computed for each movement vector in order to assess the local spatial association of speed values: the LISA, and both its probability *p* (ranging from 0 to 1) and *Z* values for a two-tailed test under a null hypothesis assumption (i.e., compared with a standard normal distribution). The value of a LISA indicates local association of speed values. For example, high positive values imply that a movement vector is surrounded by vectors with similar values; meanwhile high negative values indicate that a movement vector is surrounded by very different values. Both *p* and *Z* values are used to either reject or fail to reject the null hypothesis of no spatial association of nearby vectors using a certain statistical significance level.

The graphical version of the set of LISAs is a scatterplot of a data space constructed by plotting Z scores versus the movement speed values. It provides an exploratory visual tool for analysing the results that allows us to understand the relationship between the computed values of spatial association based on speed values of movement vectors. The scatterplot represent a common relation between movement speed and the LISA Z scores. This relation resembles a "saddle-point" shape in the scatterplot, and plays an important role in the classification of movement vectors into five different classes using a given significance level. Table 2–1 and Figure 2–5 show the classification of movement vectors using a 5% significance level.

Table 2–1. Classification of movement vectors according to movement speed, LISA sign, and LISA Z score using a 5% significance level.

Speed	LMi	LMi Z Score	Class
>Avg	+	> 1.96	High speed vector surrounded by other high speed vectors (high speed cluster). See zone 1 in Figure 2–5.
>Avg	-	< -1.96	High speed vector surrounded by low speed vectors (high speed outlier). See zone 2 in Figure 2–5.
<Avg	-	< -1.96	Low speed vector surrounded by high speed vectors (low speed outlier). See zone 3 in Figure 2–5.
<Avg	+	> 1.96	Low speed vector surrounded by other low speed vectors (low speed cluster). See zone 4 in Figure 2–5.
>Avg or <Avg	-/+	>-1.96 and < 1.96	All vectors with non-significant spatial association. See zone 5 in Figure 2–5.

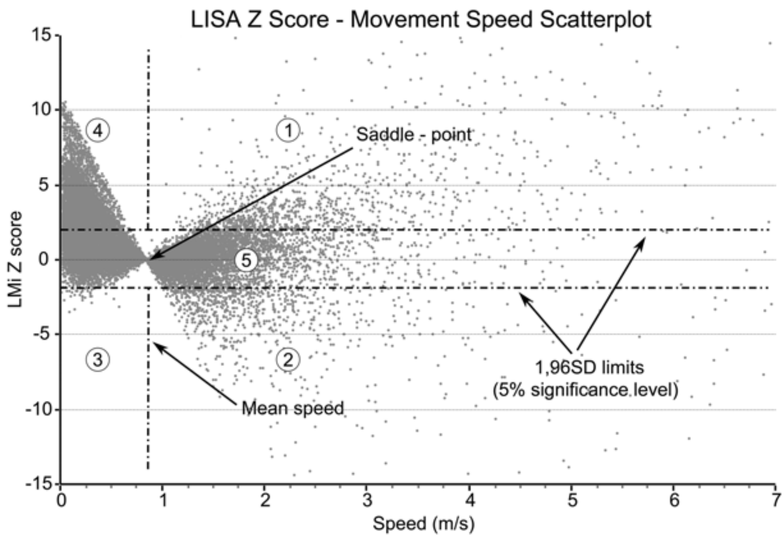

Figure 2–5. The saddle-point shape of an example of the LISA – movement speed scatterplot showing the five zones used for the classification. The movement vectors in zone 4 are classified as movement suspension.

We are especially interested in movement vectors located in zone 4, because these movement vectors have low speed values and high statistical significance of spatial association. They represent the spatial clusters of slow speed, and therefore they are the patterns of movement suspension.

2.4 The experiments

We performed two experiments to evaluate our proposed approach. These experiments consisted on analysing large pedestrian-positioning datasets for detecting the location of movement suspension patterns. We selected two datasets with substantial differences in order to test the suitability of our approach. The first dataset was collected during an educational urban mobile game with GPS-enabled mobile phones. The players walked around a city centre to find relevant historical places, and when they found one, they had to complete an assignment before proceeding to the next place. The second dataset was collected during a recreational study of visitors of a natural park area. Each visitor was asked to participate in a survey and provided with a GPS receiver to carry during his/her visit.

In both experiments, pedestrians walked and stopped at different places. Nevertheless, the conceptualization and meaning of stops were significantly different. In the urban mobile game, players mainly stopped to complete game assignments at historical places that were unknown to them. However, they also could stop at other places that were related with the urban configuration (e.g., pedestrian crossings, crossroads and traffic lights). In the park experiment, pedestrians usually stopped at points of special interest, depending on their visiting goals and activities, as well as at places where the park facilities are located, such as information boards and eating areas. Besides the different landscapes, the positioning datasets also varied in size, spatial extension, average travelled distance, and logging rate (See Table 2–2 for an overview).

Table 2–2. Main characteristics of the positioning datasets used in the experiments

Experiment	Pedestrians	Logging rate	Days	Covered Area (km^2)	Avg. Travelled Distance (m)	Tracking Points
Urban Outdoor Mobile Game	419	Fixed: 10s	10	1.2	1,248	61,782
Dwingelderveld National Park	372	Variable	7	45	5,576	141,826

2.4.1 Experiment 1: Urban outdoor mobile game

This experiment aimed at transforming 12-14 year-old students into Pilgrims of the medieval Amsterdam of 1550. The mobile game consists of using mobile phones and GPS technology for tracking the players through the city and performing location-based media-assignments about the city's history. As players move through the streets of Amsterdam, they interact with a historical map and virtual characters that provide information about historical locations and the riddles they must solve at specific checkpoints of the game. Meanwhile, they also compete against each other by placing traps on the medieval streets, and temporarily killing communication facilities with the headquarters. When players from different teams (i.e., medieval Orders) run into each other, a confrontation takes place. A set of rules determines who wins the confrontation and earns points (Waag Society, 2008).

Each player had a GPS-enabled mobile phone, and spatial coordinates were recorded in a fixed time rate (i.e., each 10 seconds during the actual game time of each player). Narrow streets sided by tall buildings, with some open spaces at squares characterizes the landscape (i.e., the city centre of Amsterdam). Most of the streets are shared by pedestrians and bikes, and some streets are allowed motorized public and private vehicles. This configuration is reflected in a relatively low quality of GPS signal that exhibits frequent unnatural jumps and fixes over buildings and channels (Figure 2–6). Moreover, frequent communication problems disturb communication between the players and the headquarters during the game, with the consequent partial loss of positioning data.

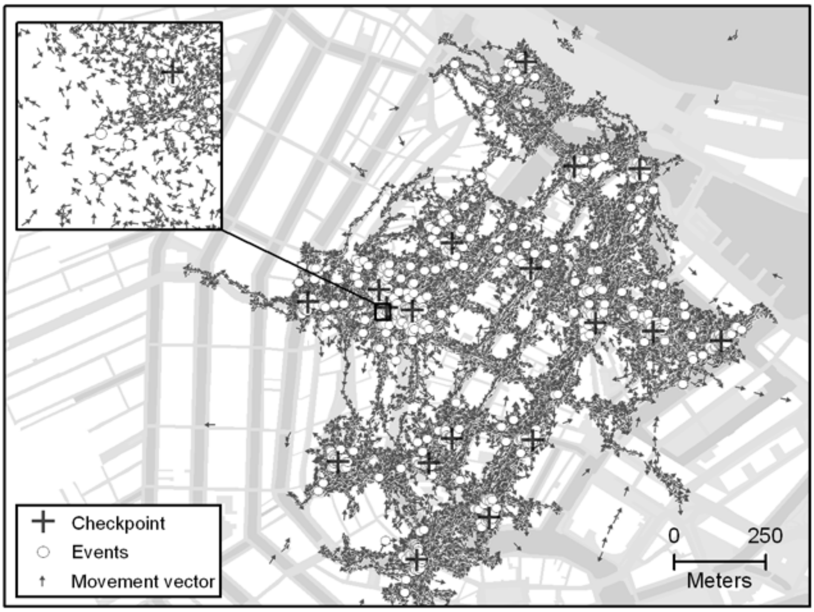

Figure 2–6. Overview of the GPS recordings of the urban mobile game experiment. The inset shows a detailed view of the observed movement vectors. The background map was obtained from www.OpenStreetMap.org.

The players were expected to explore the downtown in order to find the checkpoints for each assignment. Once they arrived at a checkpoint, they had to solve a riddle before continuing to the next checkpoint. Therefore, movement suspension patterns are expected to occur near the checkpoints of the assignments. During the game, the players of different teams meet each other at random locations in the city. Some of these events also are expected to cause the suspension of movement of the players. Therefore, in order to evaluate our approach, we compare the location of detected suspension patterns with the locations of those events.

2.4.2 Experiment 2: Dwingelderveld National Park

The Dwingelderveld National Park is a facility of about 3,700 ha in the North Eastern part of the Netherlands. It is a typical Dutch recreational area with an extensive network of short strolls (60 km of marked trails, each of less than 7 km in length) and long walks for cycling and horse riding. The landscape mainly consists of dry and wet heath lands, pine and deciduous forest, and an important complex of juniper shrubs (van Marwijk, 2009). Dwingelderveld is a very popular area that receives between 1.5 and 2 million visitors per year. Besides the wetlands, sheep farms, and some bird-watching lookouts that constitute the main tourist attractions, the park contains additional amenities for visitors, such as staffed and unstaffed information centres, a tea house, and some cultural spots such as a historical house and a radio-telescope (van Marwijk, 2009). Visitors enter and leave the park trough one of the five access points (where the car parks are located), and walk on the path network visiting one or more points of interest or performing different leisure activities.

In this experiment, information comes from three different sources. The first is a set of point coordinates captured by GPS receivers given to the visitors at the entrances where a visit starts (the beginning of the GPS track). This data collection was carried out during seven days (weekend and weekdays) in the spring and summer of 2006 (van Marwijk, 2009). The positioning dataset consists of about 142,000 GPS recordings with a variable time rate for 372 visitors (Figure 2–7). The second source is a map containing the path network and the locations of the park entrances. The third source is a collection of points of interest gathered from several specialized web pages containing recommendations and tips for visiting the Dwingelderveld National Park (Pol-Recreatie.nl, 2003; Natuurmonumenten, 2009).

The analysis task consists of computing movement suspension patterns and comparing them with the location of the points of interest. The assumption here is that the common visitor behaviour is walking around the park, and when visitors arrive at some attraction, their movement speed is reduced to reaching a halt. Obviously, the set of points of interest is not complete; and additional suspension patterns could be related to other events that are not related to visiting attractions.

2.5 Results and discussion

Results are shown using two main visual representations: data space and geographical space. Data space is represented as a scatterplot with LISA Z scores on the Y axis, and movement speed on the X axis. Geographical space is visualized as a dynamic choroplet map. The analysis task is to explore the relevant characteristics of spatial association values in both spaces, and identify the geographical places where the movement suspension patterns occurred. The computation of the LISAs was performed using ArcGIS, and resulting values were stored in a database as three new variables for each movement vector.

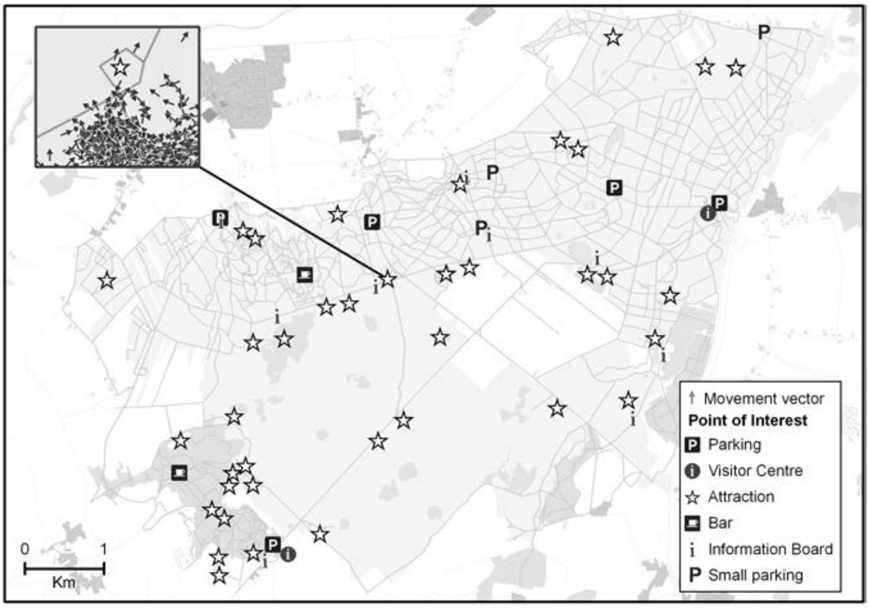

Figure 2–7. Map of Dwingelderveld National Park showing the path network, the positioning data, and several points of interest. The inset shows a detailed view of the observed movement vectors. The background map was obtained from www.OpenStreetMap.org.

The scatterplot diagrams for both experiments show a common relation between movement speed and the LISA Z scores, resembling the saddle-point depicted in Figure 2–5. Using a 5% significance level, the movement vectors in zone 4 of the scatterplot are classified as suspension patterns, as explained above. When these vectors are plotted in the map, they form clusters indicating the locations of patterns of suspension of movement. A visual exploration of these locations shows that they are mainly located at points of interest in both experiments, where pedestrians are expected to stop. These results are discussed in more detail in the next sections.

2.5.1 The urban outdoor mobile game

In this experiment, 17.84% (*N*=10,861) of the sample movement vectors are inside zone 4 of the LISA-movement speed scatterplot, and consequently they are classified into 55 clusters of movement suspension (Figure 2–8). From these vectors, 84.4% (*N*=9,162) belong to 18 clusters that are located at the corresponding 18 checkpoints of the game (Figure 2–8a, c). During the game, players should stop at the checkpoints to receive multimedia information and instructions that are necessary to solve a riddle and accomplish an assignment associated to that place. Moreover, 27 clusters are located at geographical places related to other events during the game, such as traps, confrontations, and the delivery of media to a player's mobile phone (Figure 2–8b). These clusters contained the 14.4% (*N*=1,568) of movement vectors classified as suspended.

Besides the clusters associated with the game activities, two clusters containing 0.5% (*N*=50) of the classified vectors are located at pedestrian crossings, and can be considered as evidence that other movement suspension occurs such as waiting for a traffic light (Figure 2–8e). The remaining eight clusters cannot be associated with any particular place or do not furnish evidence about some event of the game (Figure 2–8d). These clusters include 0.8% (*N*=81) of the classified vectors, and can be considered as false positives. Table 2–3 summarizes these results.

2.5.2 Dwingelderveld National Park

In the Dwingelderveld National Park experiment, 6.3% (*N*= 8,988) of the movement vectors are classified into 152 clusters of movement suspension (Figure 2–9). A visual exploration shows that five of the largest clusters are located at the car parks where visitors started and finished their visits (Figure 2–9b, e). These clusters include 45.9% (*N*=4,127) of all the movement vectors classified as movement suspension. Another 18 clusters, containing 26.3% (*N*=2,366) of the classified vectors, are associated with the locations of different attractions in the park, such as wetlands (Figure 2–9a), a sheep farm, and a radio-telescope (Figure 2–9c). These statistics clearly indicate a preference by visitors for these attractions. Moreover, seven additional clusters are associated with some facilities in the park, such as a teahouse, visitor centres (Figure 2–9 b, e), and information boards (Figure 2–9b). These clusters include 3.7% (*N*=303) of the classified vectors.

In addition to these clusters associated with typical attractions in the park, we found that 54 clusters are located at the cross-paths of the path network. These clusters contain 15.3% (*N*=1,376) of the vectors classified as suspension. We interpret these clusters as evidence of a route-choosing interaction during the movement of the visitors (Figure 2–9e). Overall, we are able to associate 90.2% (*N*=8,172) of the classified vectors with some relevant location in the park, allowing us to give a feasible interpretation to the detected suspension patterns. The remaining 9.1% (*N*=816) of the classified vectors grouped into 68 clusters, and cannot be associated with any known location of attractions or facilities in the park. Table 2–4 summarizes the results.

Table 2–3. Number of vectors classified as movement suspension and their association with the places of the urban mobile game experiment

Vectors classified as suspended	Associated with: Checkpoints	Associated with: Event Locations	Associated with: Crossings	True positives	False positives
10,861	9,162	1,568	50	10,780	81
	84.4%	14.4%	0.5%	99.2%	0.8%

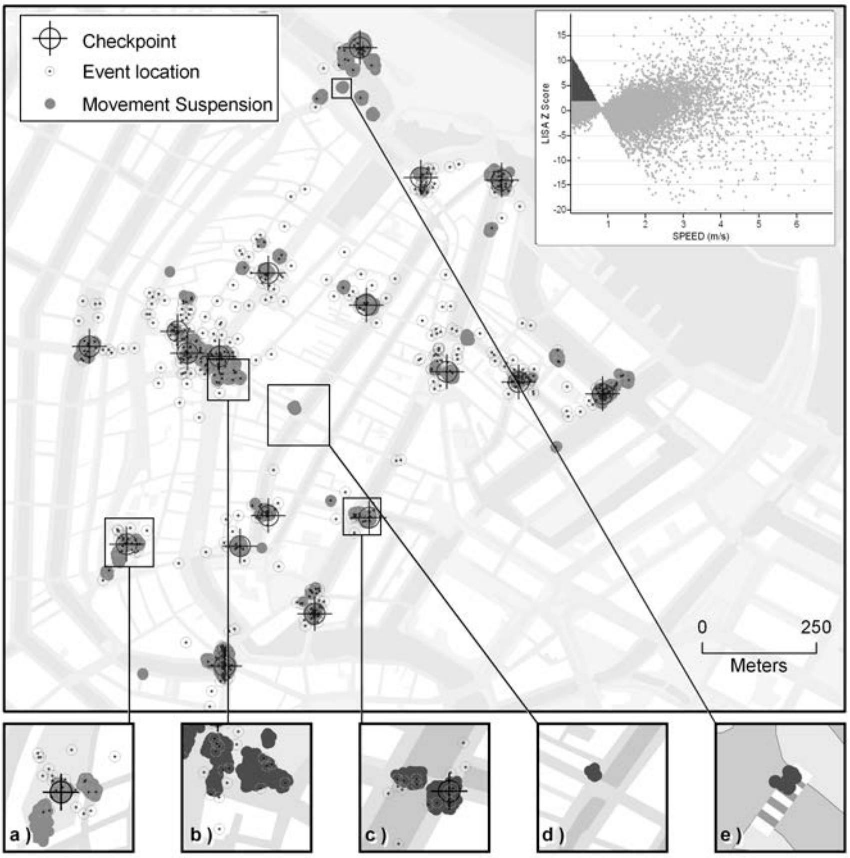

Figure 2–8. The clusters representing movement suspension patterns in the urban mobile game experiment are mainly associated with the locations of checkpoints (a, c) and events in the game (a, b, c). Two clusters are associated with pedestrian crossings; one is shown on (e). Eight clusters are not associated with any known game activity or player interactions during the game (d).

Table 2–4. Number of vectors classified as movement suspension and their associations with relevant places in the Dwingelderveld National Park.

Vectors classified as suspension	Associated with:				True positives	False positives
	Car parks	Attractions	Facilities	Cross paths		
8,988	4,127	2,366	303	1,376	8,172	816
	45.9%	26.3%	3.3%	15.3%	90.9%	9.1%

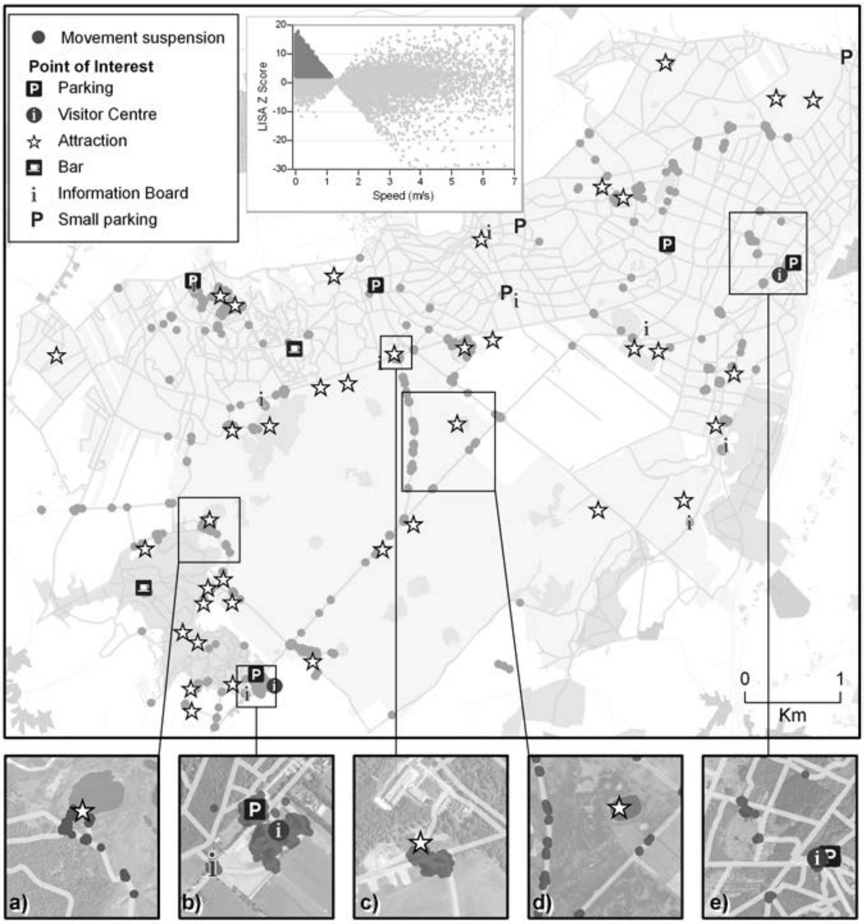

Figure 2–9. Overview of the clusters representing patterns of movement suspension found in the Dwingelderveld National Park experiment. Eighteen clusters are associated with several attractions, such as wetlands (a) and a radio-telescope (c). Another seven clusters are associated with park facilities, such as car parks and visitor centres (b, e). Small clusters also are detected at the cross paths (e). Reasons for some clusters remain elusive (d).

2.6 Conclusions

Collective pedestrian movement is a complex process influenced by a great variety of factors. In order to understand this process, one of the most interesting challenges is to explore pedestrian positioning datasets and detect the places where a suspension of movement to a complete halt occurs. However, available GPS tracking technologies are not accurate enough to capture the characteristics of movement at very low speeds; as a result, establishing reliable thresholds to distinguish stops from moves is not possible.

We propose a new approach to explore patterns of movement suspension. These patterns are represented as spatial clusters of slow speed movement vectors. We assume the existence of a spatial structure in the positioning datasets and compute the local spatial autocorrelation of movement vectors with LISAs. These indicators allow us to classify the movement vectors and identify the location of movement suspension patterns. We applied the proposed approach to two different datasets in order to compare suspension patterns with geographical places where pedestrians are most likely to stop.

The results show that most of the detected movement suspension patterns form spatial clusters located around places that represent some attraction in the context in which the pedestrians move. For players in the urban mobile game, suspension patterns are located at the checkpoints and a priori known locations of gaming events. For visitors to the Dwingelderveld National Park, suspension patterns are located around car parks and attractions, such as wetlands and a radio-telescope. Other patterns are associated with places where pedestrians wait for a traffic light in the city, or choose a route in a cross path in the park. In both experiments, over 90% of the vectors classified as movement suspension can be associated with relevant locations.

All places where pedestrians were expected to stop during the game (i.e., checking points) and during visits to the park (i.e., car parks), clearly are associated with the spatial clusters of movement suspension (see Table 2–5). However, the association of other places with the spatial clusters is not straightforward. Smaller association rates are found in places such as event locations (in the game) as well as attractions and facilities (in the park) (Table 2–5). This result suggests the suitability of the proposed approach to detect places that are independent of the context of an application.

Table 2–5. Number of places considered in each experiment, and the percentage of them associated with patterns of movement suspension

Experiment	Places	Number	% of places associated with suspension patterns
Urban outdoor mobile game	Checkpoints	18	100
	Event locations	579	53
Dwingelderveld National Park	Car parks	5	100
	Facilities	16	69
	Attractions	38	39

Implications of these results are twofold:

- Despite that very different values of movement vectors can be located at the same place, a spatial dependence exists among them, evidencing the influence of the environment on pedestrian movement. Some places attract or restrict human movement causing lower speeds, and, therefore, we can use local spatial association to discover such places in pedestrian positioning datasets.

- LISA is a reliable indicator for exploring movement suspension in pedestrian mobility.

Therefore, we propose the following consistent definition of suspension of movement based on the local spatial association of speed values:

> *Given a pedestrian positioning dataset, movement suspension patterns are spatial clusters of movement vectors that simultaneously fulfil the following three conditions: (a) having speed values below the average for a given dataset; (b) having a positive local spatial association of these speed values; and, (c) having a minimum statistical significance score of this association corresponding to an established significance level.*

This definition allows us to explore and interpret the patterns of movement suspension in positioning datasets having no a priori knowledge about the conditions in which the datasets were collected. An analyst can choose a level of significance based on the statistical properties of the datasets as well.

The main advantage of our approach is that very little knowledge is required about the context and characteristics of the data collection process, and, therefore, no need exists to establish ad hoc thresholds based on space, time, or speed. Consequently, the proposed approach can be considered as scale-independent. The only spatial parameter needed (i.e., the cut-off distance of the spatial relationships matrix) can be easily established as an additional step in the exploratory analysis, similar to a multi-band Ripley's K function. A further validation is still needed in order to assess the accuracy of the proposed approach in computing movement suspension patterns for a mixture of different moving objects, like vehicles and pedestrians.

Because our approach focuses on the analysis of collective movement using a vector-based representation, a comparison with existing methods is cumbersome; these other methods are designed to analyse individual trajectories. However, we found one example given by Palma et al. (2008) that uses the intersection-based (SMoT) and speed-based (CB-SMoT) approaches to analyse the same positioning dataset of our urban mobile game experiment. Their results show the sensitivity of the approaches to the selected space-time thresholds, having a large range of the number of detected stops, varying from 6 to 357 with SMoT, and from 37 to 182 with CB-SMoT.

Our approach can be applied to very noisy GPS tracking datasets, without necessity of cleaning positioning errors that could change the global distribution of speed values in a dataset and consequently have an effect on the local statistics. However, we are envisaging new controlled experiments to test the sensitivity of our approach to this kind of errors.

Finally, the results of our experiments show patterns of movement suspension that are the main evidence of a strong interaction between the environment and collective movement. These interactions can be used to understand pedestrian movement from a behavioural perspective. We are interested in further investigating the relation between movement suspension patterns and human activities in order to build a spatial knowledge representation for pedestrian movement. We also are interested in detecting suspension in real-time movement datasets. The main challenge arises from the computation of the spatial relationship matrix, a process that may require considerable computing time and distributed implementations to include the movement of many pedestrians. Despite these technical complications, the application possibilities are wide ranging, including ones for environmental management, urban planning, surveillance, and mobility aware services.

Acknowledgments: This research was funded by the IST - Information Society Technologies Program of the FP6 Framework under the GeoPKDD Project (www.geopkdd.eu). The authors thank Ramona van Marwijk, University of Wageningen, for the Dwingelderveld National Park data collection as well as Aske Hopman, Waag Society, for the Urban Mobile Outdoor Game data collection.

3

Validating Suspension Patterns in Pedestrian Movement Data

Orellana, D., Bregt, A.K., de Bruin, S., Ligtenberg, A. and Wachowicz, M.

Submitted to Professional Geographer

Abstract: Researchers in various fields are interested in studying movement patterns by analysing large GPS tracking datasets. One of the objectives of current research in movement analysis is to detect places where people stop. Movement Suspension Patterns (MSPs) have been proposed to identify places where movement is suspended. In this contribution, we validated the MSP approach using GPS data from a controlled experiment in which participants walked and stopped at designated locations in a natural area. We compared in time the occurrence of detected MSPs with a set of reference stops and found that the MSP approach detected 92 percent of the reference stops, with a false positive rate of $\alpha = 0.16$. We also compared the location and extent of places of movement suspension computed as a Percent Volume Contours (PVCs) in a Kernel Density Surface with a set of predefined stopping places and found that 96 percent of them lay within inside the areas delineated by the PVCs. These results show that MSP is a feasible approach for detecting the occurrence and location of stops in pedestrian movement data.

3.1 Introduction

The proliferation of mobile Global Positioning System (GPS) devices and wireless communication systems is generating a massive flow of data about the movement of people. These data represent both an opportunity and a challenge for researchers and specialists in various fields, such as transportation management, urban design, location-based services, tourism administration, and emergency management, who want to detect emergent movement patterns and extract useful information for studying the spatial behaviour of people (Nielsen and Hovgesen, 2004; Shoval and Isaacson, 2006; Laube, 2009). Experiments have been carried out to investigate the spatial behaviour of pedestrians in which volunteers were given GPS receivers to track their movement in different settings, such as shopping streets (van der Spek, 2006), touristic places (Modsching et al., 2006; Shoval, 2010), and natural areas (van Marwijk, 2009). Several projects are attempting to take advantage of data voluntarily collected by citizens to provide information about the geographical environment. Open Street Map, for example, contains nearly 2.2 billion GPS track points recorded by volunteers during their daily journeys and other trips made by various transportation modes to build an open, detailed map of the world (OpenStreetMap, 2011b).

One of the main tasks of movement analysis is to discover the locations where people stop, since these locations often represent important places in the geographical space and can be related to specific activities (Spaccapietra et al., 2008). For example, in tourism management applications the locations of stops can be used to analyse visitors' preferences for different features, and in urban planning applications the location of stops can provide indications of the value and function of public spaces (Stonor, 2004).

Several methods have been proposed for detecting stops in movement data. Most of these methods focus on separating individual trajectories into "stops" and "moves" using thresholds defined for a specific application. In (Alvares et al., 2007), for example, the authors proposed a method called SMoT for detecting stops in individual trajectories by analysing the intersection of trajectory segments with a set of previously known geographical features. If the duration of an intersection exceeds a predefined threshold, the trajectory segment is considered a stop. In the approach proposed by Wolf, Guensler, and Bachman (2001), the authors assumed that a stop is a part of a trajectory where the speed remains zero, or near to zero, for minimum amount of time (e.g., 1 minute). Similarly, Palma et al. (2008) presented a method called CB-SMoT that analyses individual trajectories and generates stops when the speed of the object remains below a threshold for a minimum amount of time. In Kang et al. (2004), the authors introduced a spatiotemporal clustering algorithm to detect significant places. The idea is that a significant place is created when a moving object does not move more than a specified distance for a minimum amount of time. Rinzivillo et al. (2008) also proposed that distance and time thresholds can be used to detect stops. Other authors (Marmasse and Schmandt, 2000; Ashbrook and Starner, 2003) assumed that stop places (e.g., home, work, shop) are usually located inside buildings and proposed exploiting the variability of GPS signals inside buildings to detect those places. In the approach proposed by Marmasse and Schmandt (2000), for example, a place is a position where the GPS signal is lost three or more times within a given radius. Ashbrook and Starner (2003) presented an improved approach that segments the trajectories by marking positions where the GPS receiver loses the satellite signal or indicates a speed continually below 1 mile per hour. These candidate positions are then merged using a variant of k-means clustering.

All these methods require a parameterization of space, time and/or speed thresholds defined in an ad-hoc way for each specific application. The problem of defining the maximum distance, the maximum speed and the minimum amount of time required for an object to be considered to have stopped is not trivial. This parameterization requires previous knowledge of the characteristics of the moving entities, the data collection techniques, and the characteristics of the environment, and it is highly dependent on the transportation mode, the granularity of the data, and the application objectives. Moreover, the results are highly sensitive to these thresholds. For example, in the method proposed by Palma et al. (2008), the number of stops detected in a dataset varied from 37 to 182 for different input thresholds, but no information was provided about the evaluation of the results. The problem with using thresholds for the determination of stops has also been recognized by Ashbrook and Starner (2003), Hightower et al. (2005), and G. Andrienko and N. Andrienko (2007).

An alternative approach is to use a local indicator of spatial association (LISA) to detect spatial patterns of stops, or Movement Suspension Patterns (MSPs) (Orellana and Wachowicz, 2011). In this approach, the places where people suspend their movement were detected by computing the local Moran's index (LMi) (Anselin, 1995) to evaluate the spatial association of speed values in a movement vector dataset. A movement vector consists of the spatiotemporal location of one observation together with the direction and speed of the movement. The local spatial association is computed for each movement vector and the result comprises two new variables: the LMi and a Z score representing the statistical significance of the spatial association (Anselin, 1995). If a vector has a low speed (below the mean speed of the dataset) and a high positive Z score (above a selected statistical significance level), the vector is classified as "suspension". When these suspension vectors are plotted on a map, they form spatial clusters (MSPs). These clusters are assumed to identify places associated with collective stopping behaviour. Since the approach relies on the spatial-statistical properties of the dataset, no application-dependent thresholds are required. Also, the MSP approach captures the suspension of movement even when moving objects do not come to a complete halt, which is common in many applications that record pedestrian movement because people are rarely completely motionless.

Although MSP is a promising approach for identifying stopping places in large movement datasets, its validity is still unknown. This is a common issue in the field of movement pattern analysis, in which the results of the studies are often not properly validated, for example by comparing them with reference data (Alvares et al., 2007; Palma et al., 2008; Rinzivillo et al., 2008). Although Orellana and Wachowicz (2011) compared the location of MSPs with a background map containing potential points of interest, they did not compare them with reference data (e.g., stops reported by the walkers). Moreover, since MSPs are clusters of points, the spatial location and extent of the places of movement suspension are not explicitly defined. This approach therefore needs to be properly assessed, for two main reasons. First, potential users will want to know how accurate the approach is. Second, the analytical methods, such as comparing the detected MSP with contextual geographic information, depend on the availability of explicit representations of the places of movement suspension.

In this article we evaluate the MSP approach for detecting stops in movement data. We analysed the data from a controlled experiment in which a group of participants walked in a natural area. MSPs were computed and evaluated by comparing them to a reference dataset in space and time. The evaluation sought to answer two questions: a) How accurately can MSPs be used to detect the occurrence of stops in pedestrian movement? and b) Do the MSP spatial clusters represent the places where people stop?

In the next section we explain the data and methods used for the evaluation. Section 3 presents the main results and related discussion. In Section 4 we draw general conclusions about the assessment and its implications for movement research.

3.2 Data and methods

To validate the Movement Suspension Patterns approach, it is necessary to compare its outcomes (i.e., MSPs computed in a movement dataset) with data on reference stops. This must be done for both the spatial and temporal dimensions of MSPs. The co-occurrence of MSPs with reference stops is evaluated in the temporal dimension, and the location and spatial extent of MSPs are compared with the location of the reference stops. The results of both comparisons are used to evaluate the accuracy of the approach.

Figure 3–1 shows the steps in the data collection and validation procedure. To obtain sound datasets, we designed a controlled experiment consisting of three phases. In the data collection phase, we tracked the movement a group of participants walking a route and stopping to take photographs at predefined places marked on a map. In the data processing phase, we detected MSPs in the GPS data and created a reference dataset containing the participants' stops. The time and duration of the reference stops were computed using the timestamp of the photographs, and their location was defined using the places marked on the maps. Finally, in the evaluation phase, we compared the spatial and temporal co-occurrence of the computed MSPs with the reference stops.

The experiment was designed to meet the following criteria: a) each stop made by the participants has to be recorded; b) to assess the approach for both individual and collective movement, the route must include places with different proportions of people stopping and walking; c) to obtain comparable results the landscape conditions and data collection must be similar to those in the original MSP research (Orellana and Wachowicz, 2011).

The evaluation makes use of the following quantitative concepts:

- *Movement Suspension Pattern* is a cluster of movement vectors of one participant. The state of suspension is defined by a speed value below the mean of the dataset and a Z score of the local Moran's index above 1.96 (corresponding to a significance level of 5 percent).
- *Reference stops* are stops with a temporal duration and spatial location computed from data independent of the MSPs.

- *Stopping ratio* is the number of participants stopping at each place divided by the total number of participants (stopping or walking) at that place.

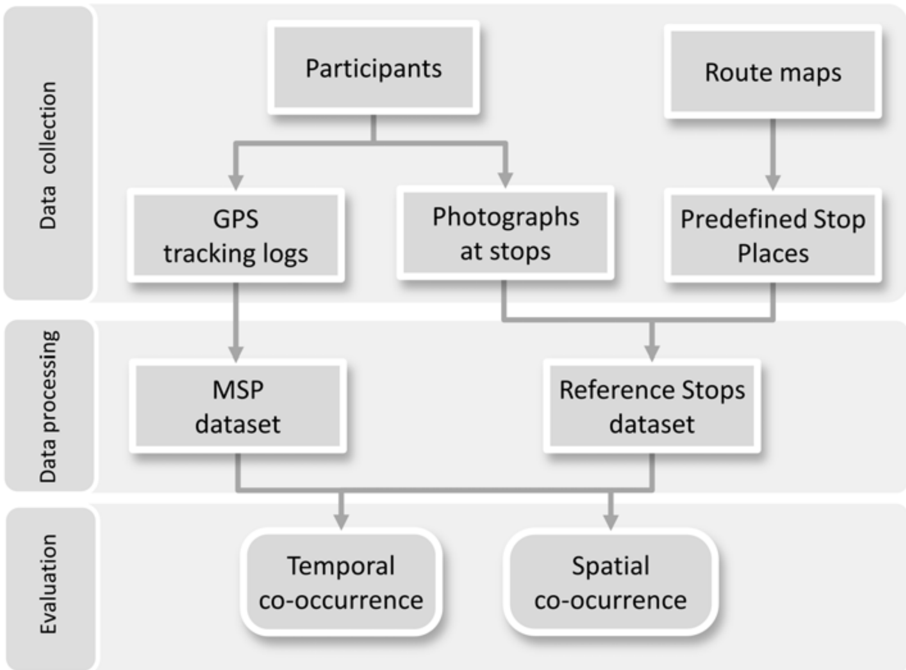

Figure 3–1. Steps in the method for the spatial and temporal evaluation of movement suspension patterns

3.2.1 Data collection

The data were collected during an experiment in the north-eastern part of the Dwingelderveld National Park (Figure 3–2) in December 2010. The movements of twenty-eight participants carrying GPS receivers were monitored. Participants were divided into twelve teams of two people, one team of three people, and one team with a single person. These fourteen teams each travelled a different route. The routes contained a total of sixty-two predefined places. The participants were provided with GPS receivers, cameras, report forms and route maps. The GPS receivers were configured to log the position of users every ten seconds. The routes and teams were designed to have different numbers of people stopping at each place, ranging from one to twenty-eight.

The following instructions were given to all the participants: 1) Follow the assigned route on your map and stop *only* at the indicated places on the map. 2) When you arrive at an indicated place, stop for *at least* one minute at the location. 3) After the minute is over, record a mark (waypoint) in the GPS receiver and take a photograph of the GPS screen showing the GPS clock. 4) Take three photographs of the place. 5) Walk towards the next designated place following the route. 6) If something unexpected happens, report it using the provided forms. All the instructions were explained before the experiment started and provided on paper to the participants.

After the experiment, the GPS logs and the data extracted from the photographs and forms were stored in a geodatabase.

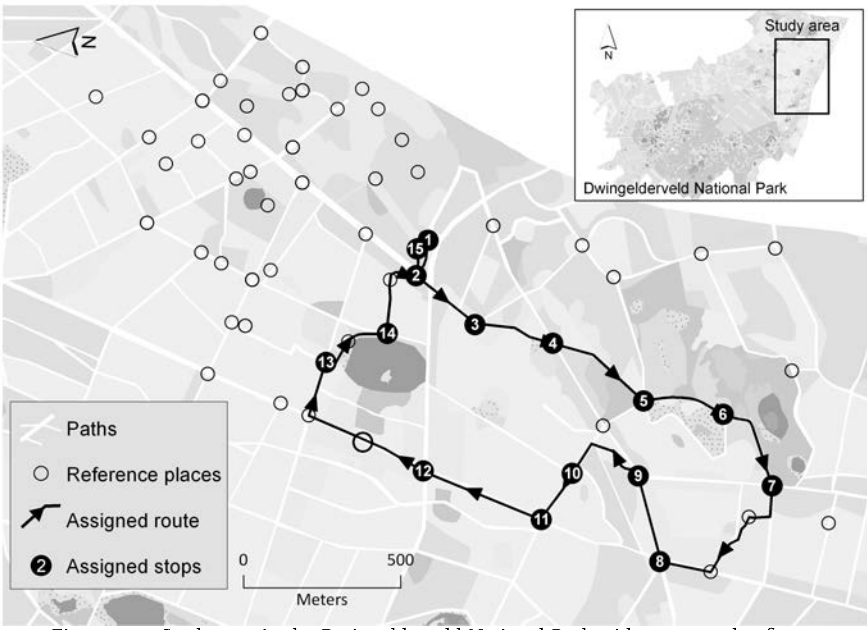

Figure 3–2. Study area in the Dwingelderveld National Park with an example of a route and stops.

3.2.2 Data processing

MSP dataset

The GPS logs were used to compute movement. Outliers were removed using spatial criteria (vectors outside the study area) and data criteria (vectors with speed > 5 metres/second). The LMi and Z scores of the movement vectors were then computed. The vectors with a speed below the mean and a Z score above 1.96 were classified as suspension.

Figure 3–3A illustrates how the vectors classified as suspension were aggregated into Movement Suspension Patterns. Each MSP consisted of a sequence of consecutive movement vectors of one participant that were classified as suspension. A tolerance of one observation was established to avoid unintended splitting of MSPs (e.g., a weak GPS signal may produce errors in the location and/or speed of the vector). Therefore, the duration of each MSP was determined by the first and last vector in the MSP. The resulting MSPs were stored in a new table in the geodatabase and related to the original vectors. The resulting MSP data consisted of a set of MSPs together with their location and duration.

Reference stops dataset

We computed the time and duration of reference stops using the timestamp of the photographs taken by the participants each time they stopped. Figure 3–3B illustrates how the times of the reference stops were computed. Each step begins 70 seconds before the first photograph (the photograph of the GPS screen) and ends 10 seconds after the last photograph. Since the internal clocks of the cameras are not accurate, they were synchronized with the GPS time using the photographs of the GPS screen. The JOSM software (OpenStreetMap, 2011a) was used for a semi-automatic synchronization.

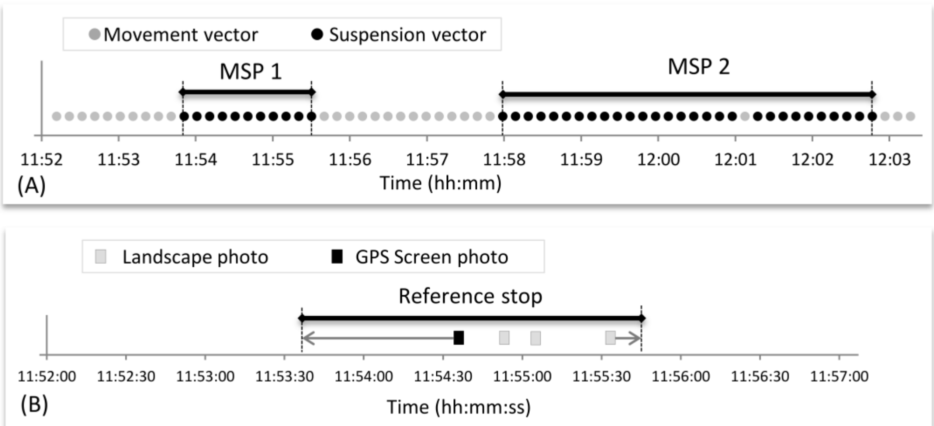

Figure 3–3. (A) Part of a timeline of movement vectors for one person showing two movement suspension patterns. A tolerance of one observation was used to aggregate the individual vectors into MSPs. (B) A validation stop starts 70 seconds before the photograph of the GPS screen was taken and ends 10 seconds after the third landscape photograph was taken.

The spatial locations of the reference stops were determined using the predefined places marked on the maps. Most of the predefined places (except those assigned to the single-person team) were located where paths crossed or joined to make them identifiable in the terrain. For the single person team, the places were located at identifiable features such as picnic benches and signposts. The photographs and report forms were used to select the places for the assessment. For example, if a team stopped at a different location than the predefined place, this information was used to exclude the place and the corresponding stops from the spatial analysis.

3.2.3 Evaluation

Temporal evaluation

The start and end times of the MSPs of each participant were compared with the reference stops. Figure 3–4 shows the temporal evaluation. If an MSP overlapped in time (completely or partially) with a reference stop, the MSP approach in that instance correctly detected the stop of the participant and the result was labelled as a True Positive. If an MSP did not overlap with a reference stop, it was marked as a False Positive. Each reference stops that was not overlapped by an MSP was marked as a False Negative. The intermediate, empty periods were marked as True Negative. Finally, the total numbers in each category were counted and used to compute the sensitivity and specificity of the evaluation.

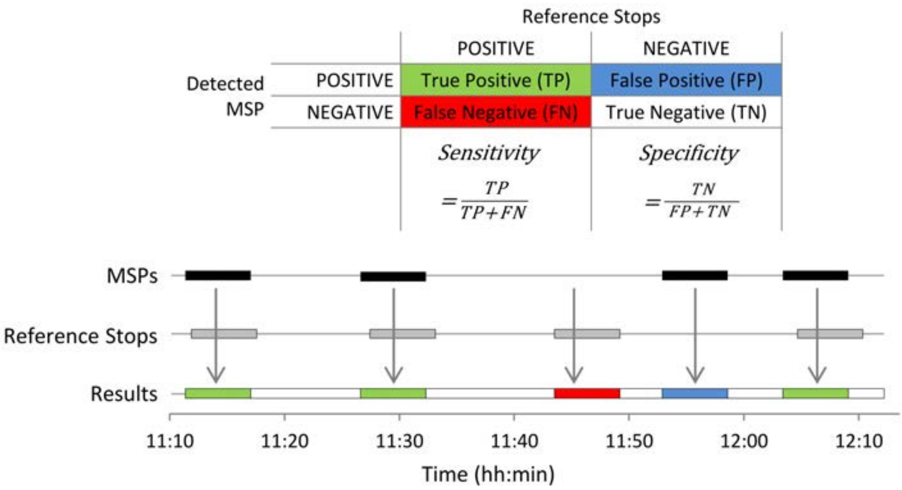

Figure 3–4. Temporal evaluation. Lines connecting movement suspension patterns represent detected negatives. Similarly, lines between reference stops are reference negatives.

Spatial evaluation

In the original approach, the spatial location and extent of MSPs were not explicitly defined, (i.e., MSPs were represented by the locations of vectors classified as suspension). To determine the location and extension of MSPs, we propose using a Kernel Density Estimator (KDE) to compute a density surface using the movement vectors classified as suspension. We used a quadratic function (3.1) (Silverman, 1986)

$$K = \begin{cases} \frac{1}{3}(1 - t^2), \ t = \frac{d}{h} \leq 1 \\ 0 \qquad , \ t = \frac{d}{h} > 1 \end{cases} \tag{3.1}$$

to calculate the density surface, since its critical parameter, the bandwidth distance *h*, can be established from the estimation of the accuracy of the observations (10 m for a single frequency GPS receiver without any signal correction), and because it is computationally more efficient than a Gaussian function (de Smith, Goodchild, and Longley, 2009). The KDE function produces a continuous raster surface representing the probability density values for each cell. On this surface, Percent Volume Contours (PVCs) represent the boundaries of the areas containing X percent of the volume of the probability density function. For example, using a PVC = 90 the delineated area contains on average 90 percent of the vectors that were used to generate the kernel density estimate (Beyer, 2010).

This approach implies a fuzzy representation of the location and extent of places where movement suspension occurred. The assumption is that the KDE function will produce a surface in which the inner PVCs represent the kernel of the place of movement suspension and the outer PVCs represent its boundary. Figure 3–5 shows an example of a kernel density estimate surface. The spatial evaluation was performed by counting the number of reference stops that fall within the PVCs for 10 percent intervals.

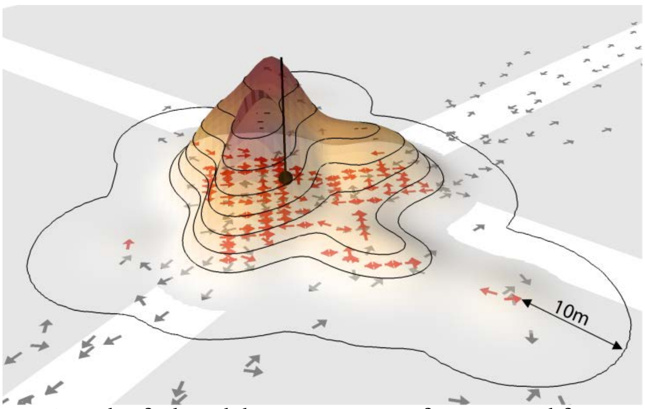

Figure 3–5. Example of a kernel density estimate surface computed from a Movement Suspension Pattern (red arrows). Contours correspond to percentiles of the density function.

3.3 Results

The movements of the twenty-eight participants were tracked and recorded in a geodatabase. The data on the beginning and the end of their walks were excluded from the analysis because they were not part of the experiment. The resulting dataset consisted of 20,225 GPS track points. The average walking time of the participants was 162 minutes and they covered on average a distance of 7.5 km. The mean speed was 0.7 metres/second. The LMi and Z score values were computed and 6,830 vectors (33.8 percent) were classified as suspension. These vectors were aggregated into 428 MSPs to create the MSP dataset. The participants took 1,611 photographs at 84 different places, and 387 reference stops were recorded. Figure 3–6 is a map of the results, which shows that most of the MSPs were located at the reference places. It is also evident that some participants lost their way and stopped at locations other than the predefined ones. This was confirmed by the participants in the report forms.

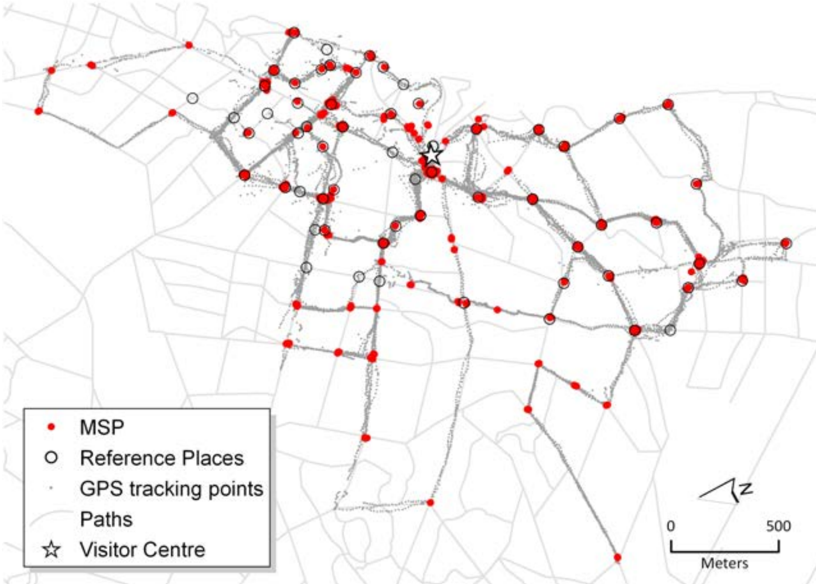

Figure 3–6. Map showing the spatial distribution of the MSPs and reference places. The star in the upper middle of the map shows the location of the visitor centre where all the routes started and ended.

3.3.1 Evaluation of results

The temporal analysis revealed that 355 of the stops were detected (True Positives), and 32 of them were not (False Negatives). Also, 73 MSPs did not correspond to any reference stops (False Positives). The assessment gave a sensitivity value of 0.92 (8 percent of the reference stops were not detected), and a specificity of 0.84 (16 percent of the detected MSP were False Positives).

Figure 3–7 presents the evaluation results in the form of timelines. The timelines reveal that some people stopped longer than the minimum time required. This finding was confirmed by the participants. The timelines also make it easier to identify the errors because all but one of the teams consisted of two or more people whose timelines are therefore similar. This is most evident in the results for the team that consisted of three participants (participant IDs 9, 10, and 11). The figure also reveals gaps in the GPS data in several timelines when the signal was lost.

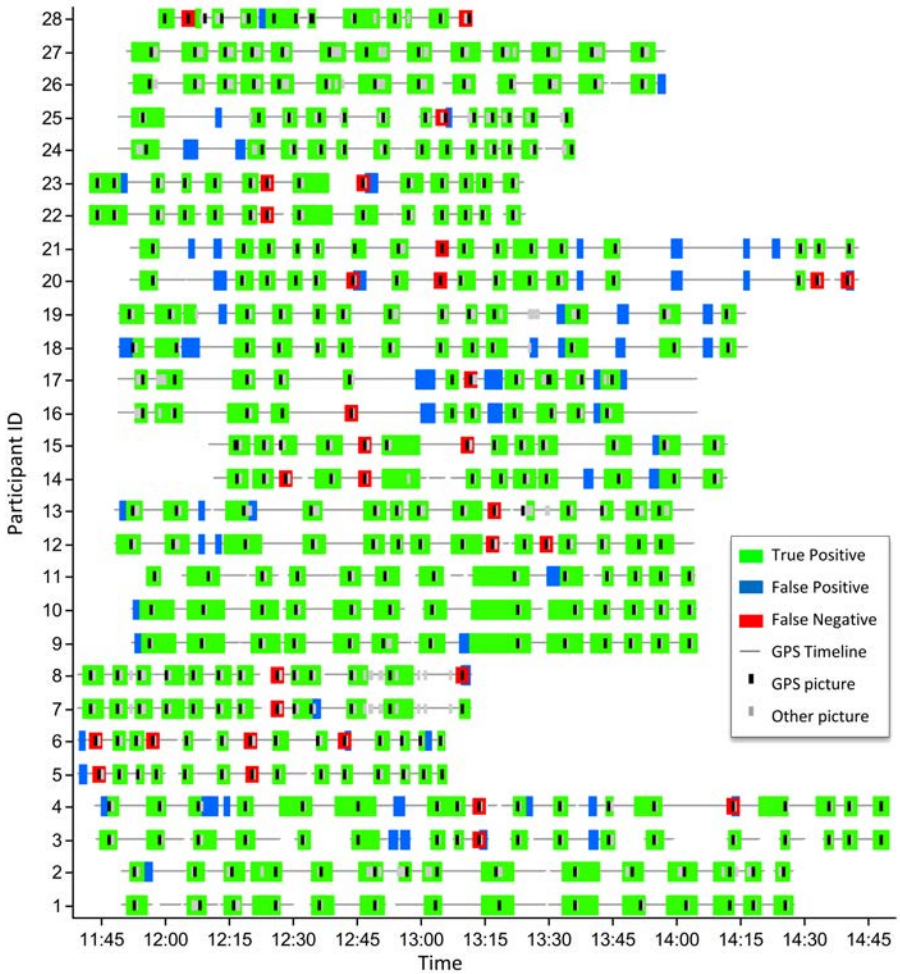

Figure 3–7. Timelines of the evaluation results. 92 percent of the reference stops were detected by the MSP approach (True Positives, in green); 16 percent of the MSPs did not correspond to any reference stop (False Positives, in blue); 8 percent of the reference stops were not detected (False Negatives, in red). The small black and grey rectangles represent the timestamps of the photographs to facilitate interpretation (black: photographs of the GPS screen; grey: landscape photographs).

The spatial evaluation was performed on a set of 143 reference stops in 47 different places. These reference stops were selected using information from the route maps, report forms, and photographs. We found that 137 stops (96 percent) were inside the areas delineated by the PVCs. Furthermore, over half of them were located inside areas corresponding to PVC = 30 percent, i.e., the core of the KDE function (see Figure 3–8D). These results are consistent with the design of the experiment, in which the participants were not motionless at one precise location, but stopped walking to take photographs of the place. These results also suggest that the KDE surface is a good approximation of the places of movement suspension. Moreover, the outer PVCs of each place can be used to represent the full extent of the place of movement suspension, whereas the inner boundaries represent the kernel of the place (Figure 3–8A, B, C).

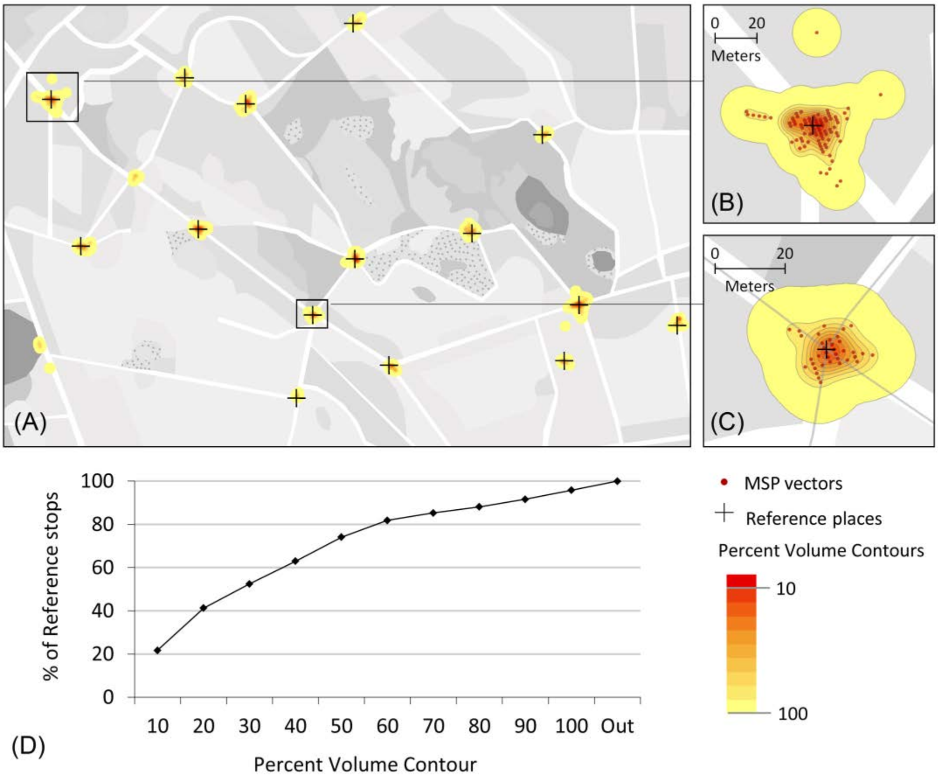

Figure 3–8. (A, B, C) Part of the study area with kernel density plots representing places of movement suspension. (D) 96 percent of the stops at the reference places were inside the PVCs. Half of those stops were inside the PVC=30 percent boundary.

3.3.2 Analysis of errors

An analysis of the 73 False Positives (MSPs that did not correspond to any reference stop) showed that the frequency distribution of their duration was different from the True Positives. This is illustrated in Figure 3–9. MSPs derived from a single vector with a duration corresponding to the granularity of the GPS recording (i.e., 10 seconds) accounted for 40 percent of the False Positives; moreover, 82 percent of the False Positives had duration of less than one minute. We may conclude, therefore, that most of these False Positives correspond to short stops not reported by the participants, since they lasted for less than one minute. Using the information from the report forms, twelve False Positives were associated with places and times where the corresponding participants reported losing their way. An example of this is shown in Figure 3–10B, where a cluster of False Positives is found at a junction of paths close to the visitor centre where many participants reported having problems finding the route. Other False Positives occurred when participants were waiting to cross the road that runs through the area (e.g., the False Positives of participants 16 and 17 around 13:00 hours in Figure 3–7).

Besides the information extracted from the report forms, the interpretation of the False Positives is also supported by the paired times and duration for each team. For example, in Figure 3–7 there is a False Positive at 13:30 hours for participant number 11 one minute before a True Positive, whereas the other two people in the team (participants 9 and 10) had a True Positive at the same time. After exploring the data in detail, it was determined that the GPS signal of participant 11 was lost for one minute, splitting the detected MSP into two parts. In general, nearly 85 percent of the False Positives could be explained by an actual suspension of the movement of the pedestrians. These results imply that the MSP approach also detected short stops that were not recorded by the participants.

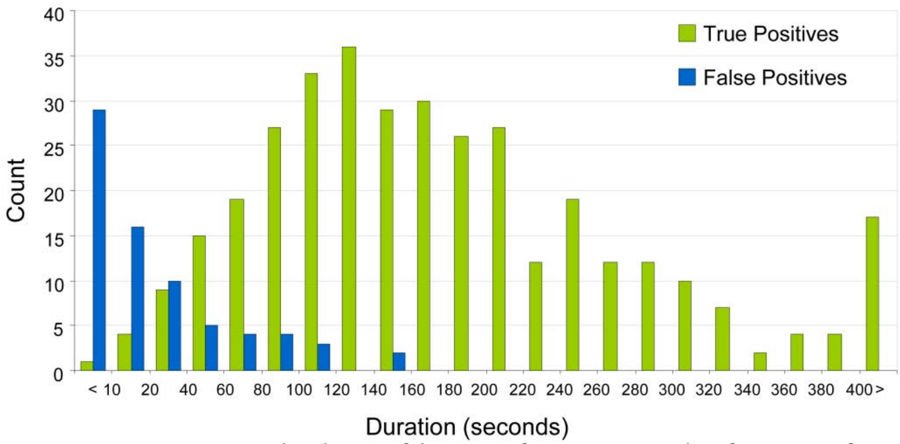

Figure 3–9. Frequency distribution of durations of positives. More than 80 percent of false positives had a duration of less than one minute.

When False Positives were plotted on the map, some were found at places where True Positives were also recorded. These can be seen in Figure 3–10. These False Positives occurred because the spatial association of low-speed vectors is high enough to classify them as suspension, even if participants slowed down but did not come to a complete halt. Although these False Positives may correspond to artefacts produced by the method, they are consistent with the original formulation of the MSP approach, which does not necessarily require pedestrians to be motionless.

The analysis of the 32 False Negatives (reference stops that were not detected) revealed that they were located at eighteen different places, eight of which had more than one False Negative. Our first assumption was that False Negatives would be located at places where the stopping ratio (the ratio between the number of stops and the total number of people at the place) was low, since the statistical significance of the spatial association of low-speed vectors at those places was also expected to be low. An example of this kind of error is shown in Figure 3–10C. On the left there is a place where the stopping ratio was 1 and the reference stops were correctly identified (True Positives); on the right there is a place where the stopping ratio was 0.1 and a cluster of four False Negatives were found. In total, 46 percent of the False Negatives were in places with a low stopping ratio. However, at least seven places with a low stopping ratio (< 0.025) were correctly detected by MSPs, which indicates that the stopping ratio is not strongly associated with False Negatives.

Other False Negatives were probably related to a degradation of the GPS signal. The effect of this degradation in the positional accuracy of movement vectors is illustrated in Figure 3–10D by standard deviational ellipses. The ellipse of the True Positive (green) is smaller than the ellipse of the False Negative due to the spatial dispersion of the corresponding movement vectors. This scattering was found in practically all False Negatives. Among the several causes that can be associated with the bad quality of the GPS signal, the most relevant in this dataset is the canopy coverage in some areas of the park. However, we found no conclusive evidence for this. The occurrence of False Negatives can therefore be associated with the MSP approach and the lack of accuracy of the tracking technology.

3.4 Conclusions

This article presents the results of a validation of the Movement Suspension Patterns approach to detect stops using data from a controlled experiment that tracked the movement of people in a natural area.

The results of the experiment suggest that the MSP approach is a feasible method for detecting stops in pedestrian movement. The temporal evaluation demonstrated that the approach correctly detected the occurrence of stops up to 92 percent of the time. However, 16 percent of the MSPs did not correspond to a reference stop. As most of these False Positives could be related to actual suspension of movement of the pedestrians, we can assume that the specificity can be higher. Therefore, it is reasonable to expect that most of the detected MSPs in the proposed approach will represent actual stops made by pedestrians.

Although the method failed to detect a small proportion of the stops (8 percent), this may be caused by degradation of the GPS signal. The scattering of the positions produced by a weak signal may strongly affect the computed speed and the local Moran's index and Z score values, which means that the corresponding movement vectors should not be classified as suspension. The GPS receivers used in the experiment work with a single frequency and do not use any signal correction. The use of more precise receivers and technical advances that enhance the accuracy of GPS and other positioning systems will have a positive impact on the MSP approach.

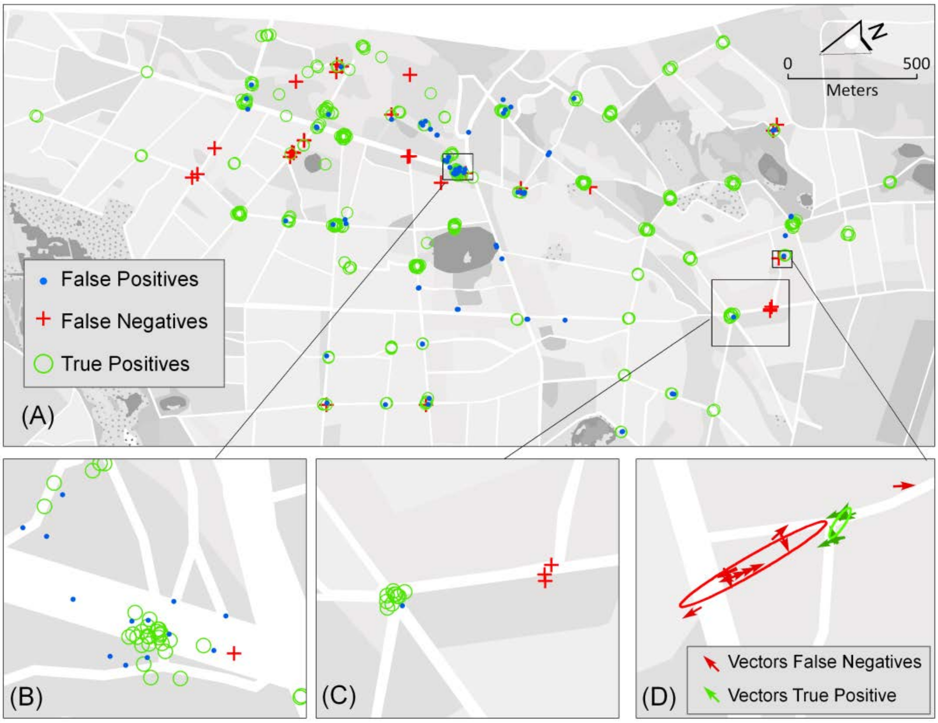

Figure 3–10. Map showing spatial distribution of results: (A) overview; (B) cluster of false positives (blue dots) located at crossing of paths near visitor centre where participants reported difficulties following the route; (C) false negatives located at location with stopping ratio 0.1 and true positives at place with stopping ratio 1; (D) scattered GPS locations from static position producing a false negative.

We can also confirm that the places where stops occur can be accurately represented using a density function surface computed for the MSP vectors, and that Percent Volume Contours (PVCs) can aid identification of the core and boundaries of these places. We found that 96 percent of the stops at reference places were inside the PVC boundaries. This representation can be useful for analytical purposes. For example, when analysing the movement data of visitors in the park, the detected places can be associated with a set of points of interest. The kernels of the places (inner PVCs) can be used to identify the corresponding points of interest and the outer PVCs can be used to identify the extent of the spatial influence of specific points on the behaviour of the visitors. In view of our results, we consider that the MSP approach accurately represents the occurrence and location of stops in pedestrian movement data.

Although the original formulation of MSPs was designed to detect collective stopping behaviour (i.e., places where many pedestrians stopped), the analysis in this article shows that the approach is also valid, at least partially, for individual stops. In fact, even at places where only a small proportion of pedestrians stopped (stopping ratio < 0.25), the MSP approach correctly detected stops 77 percent of the time.

Besides evaluating the MSP approach, this article presents a method and data (Orellana, 2010) that can be used by other researchers to evaluate their own algorithms or to compare them with existing ones, for example those presented in (Alvares et al., 2007; Palma et al., 2008; Rinzivillo et al., 2008).

This controlled experiment was specifically designed to evaluate the stops in pedestrian movement data. We can assume, however, that the method is also valid for other kinds of moving objects, such as vehicles or animals (Orellana et al., 2010). This means that the MSP approach can be used for a wide range of applications. In transportation management, for example, it is crucial to detect the location of traffic jams in historical data (Yoon, Noble, and Liu, 2007). In spatial ecology it is very important to detect the stopping behaviour of animals to understand their relation with the environment (Ganskopp and Johnson, 2007). Moreover, since the parameterization of MSPs does not depend on the application, but just the data itself, it offers a generic approach for the analysis of large datasets collected in crowdsourcing projects and location-based services. The discovery of places of interest in such datasets can be of great interest for those projects and services.

Acknowledgments: We are extremely grateful to the colleagues and friends who participated as volunteers in the experiment described in this article.

4

Exploring visitor movement patterns in natural recreational areas

Orellana, D., Bregt, A. K., Ligtenberg, A. and Wachowicz, M. (2012).

Tourism Management 33(3):672-682.

Abstract: GPS technology is widely used to produce detailed data on the movement of people. Analysing massive amounts of GPS data, however, can be cumbersome. We present a novel approach to processing such data to aid interpretation and understanding of the aggregated movement of visitors in natural recreational areas. It involves the combined analysis of two kinds of movement patterns: 'Movement Suspension Patterns' (MSPs) and 'Generalised Sequential Patterns' (GSPs). MSPs denote the suspension of movement when walkers stop at a place, and are used to discover places of interest to visitors. GSPs represent the generalised sequence in which the places are visited, regardless of the trajectory followed, and are used to uncover commonalities in the way that people visit the area. Both patterns were analysed in a geographical context to characterise the aggregated flow of people and provide insights into visitors' preferences and their interactions with the environment. We demonstrate the application of the approach in the Dwingelderveld National Park (The Netherlands).

4.1 Introduction

Monitoring and analysing the flow of visitors in natural recreational areas is key to understanding visitor behaviour, which in turn is needed for effective management that meets both conservation and recreational requirements (Muhar, Arnberger and Brandenburg, 2002; McKercher and Lau, 2008). To understand these requirements we need detailed information about area usage and the preferences of different target groups (Chiesura, 2004). Analysing the spatial behaviour of visitors by relating different uses and activities to different places and landscape configurations can provide insights into their preferences and purposes (Golicnik and Ward Thompson, 2010). One of the most important aspects of the spatial behaviour of visitors in recreational areas is their movement inside the area (intra-site flow). Monitoring the movement of people during their visits to a recreational area can help to identify which places they visit most or least, how much time they spend in each place and which kind of attractions different target groups prefer. Knowing those preferences, managers can segment the market and offer more diverse and focused options, adapted to the wishes of specific groups of visitors (Holyoak and Carson, 2009). Monitoring and analysing the movement of visitors and area usage can also provide information about potential crowding and conflicts between different groups (Manning and Valliere, 2001; Ostermann, 2009). The movement behaviour of visitors looking for solitude and relaxation may differ from visitors looking for social activities, such as playing and picnicking, and studying this can help us understand how different groups experience crowding. The study of intra-site flow of visitors can also provide information for conservation management. To assess the carrying capacity in sensitive areas, for example, we must know about the spatial and temporal distribution of visitors.

Traditionally, studies on visitors' use of space in recreational areas have been based on data and information collected from interviews, surveys and direct observation. Researchers have used geographic information systems (GIS) to analyse the spatial properties of these data to understand how the spatial behaviour of visitors is related to different places and landscape configurations (Golicnik and Ward Thompson, 2010). GIS has also been used to study how recreational areas are used by different groups to detect and understand processes of appropriation and exclusion (Ostermann, 2009).

To complement these techniques, location sensing technologies (e.g., GPS, mobile phones, PDA) are providing an inexpensive and unobtrusive way to collect massive datasets on the location in space and time of people in recreational areas (Nielsen and Hovgesen, 2004; Shoval and Isaacson, 2009; Taczanowska, Muhar and Brandenburg, 2008; van Schaick and van der Spek 2008; Xia, Arrowsmith, Jackson and Cartwright, 2008). To make sense of this new source of data, researchers are envisaging new methods and techniques for exploring and analysing vast amounts of positioning data to extract patterns that represent the movement of individuals and groups (Laube, 2009). Recent advances in the field suggest that despite the potential diversity of movement behaviour, people usually follow simple and predictable movement patterns (Gonzalez, Hidalgo and Barabasi, 2008; Song, Qu, Blumm and Barabasi, 2010). It is accepted that these patterns may provide information that will help to explain the interactions between moving entities and between those entities and the environment (Batty, DeSyllas and Duxbury, 2003; Hoogendoorn and Bovy, 2005; Bierlaire, Antonini and Weber, 2007; Gudmundsson, Laube and Wolle, 2009). Taking into account the diversity of movement patterns reported in the literature, some authors have proposed formalisation and classification systems to provide a systematic framework for ongoing research (Dodge, Weibel and Lautenschütz, 2008; Wood and Galton, 2009).

Spaccapietra et al. (2008) stated that in order to analyse movement data and detect useful patterns, the representation of the movement of an object must go beyond its raw spatiotemporal positions. In their work, the authors proposed a representation called 'semantic trajectories', in which the trajectory of the object is divided into semantic units called 'stops' and 'moves'. Stops are those segments of the trajectories where the object does not move. Among various methods proposed to implement this representation, Alvares et al. (2007) devised a method for detecting stops called IB-SMoT (Intersection-Based Stops and Moves of Trajectories), which is based on an analysis of the intersections of trajectories with user-defined geographical features for a minimal duration. Rinzivillo et al. (2008) proposed a similar approach, in which the stops are those segments of trajectories where a moving entity remains within a distance threshold for a minimum period of time. Palma et al., (2008) proposed a method called CB-SMoT (Clustering-Based Stops and Moves of Trajectories), which analyses each trajectory and generates stops when the speed value is lower than a given threshold for a minimal amount of time.

More recently, Bogorny, Heuserand Alvares (2010) suggested a general framework for modelling trajectory patterns during the conceptual design of a database. The authors provided a conceptual description of the framework, an implementation of IB-SMoT and SB-SMoT, and data-mining algorithms to extract three movement patterns (frequent patterns, sequential patterns and association rules) for semantic trajectories. They also provided examples of how to instantiate the model for different applications by parameterising the spatial and temporal dimensions. Other researchers have proposed methods for analysing aggregated movement data to learn more about the spatial behaviour of visitors. For example, Shoval (2010) proposed using a raster-based representation that divides the area of study into a regular grid of cells, and counting the number of GPS observations in each cell of the grid. Finally, some approaches focus on the aggregation of trajectories to improve the visual exploratory analysis of movement data (G. Andrienko and N. Andrienko, 2008; Demšar and Virrantaus, 2010; Scheepens et al., 2011).

A common feature of these approaches is that the conceptualisation of movement patterns requires a parameterisation of spatial and temporal dimensions, which makes the results highly dependent on the values assigned to those parameters. For example, in order to define a stop, the user must provide values for the minimum time, the minimum speed or the minimum distance to be used to determine whether an individual object has stopped, with the risk of overestimating or underestimating the number of stops. Similarly, to detect sequential patterns, the user must set the intervals for aggregating the temporal data in predefined periods (e.g., morning, afternoon, weekend). In the case of spatially aggregated data in raster-based representations, the size of the cell has a considerable effect on the summary statistics. The parameterisation of these values is not trivial and may be highly sensitive to the inherent GPS inaccuracy and to the spatial and temporal resolution of the observations (Palma et al., 2008). Moreover, the selection of parameters need is based on a priori knowledge of the dataset, and therefore may be not suitable for an exploratory approach.

In the present work, we propose a novel approach to explore the properties of the collective movement of visitors in recreational natural areas based on GPS tracking data. We define collective movement to be the aggregated properties of the movement of many people in a defined space and time, not the movement of specific groups of people moving together (i.e., collective movement rather than movement of collectives). Our approach relies on different methods of detecting movement patterns that represent the properties of collective movement. In this contribution, we focus on two kinds of movement patterns – Movement Suspension Patterns (MSPs) and Generalised Sequential Patterns (GSPs) – and demonstrate how they can be used to explore the collective movement of visitors in natural recreational areas.

The next section introduces the proposed approach and details the techniques used for the analysis. Section 3 details how the approach was implemented to analyse the flow of visitors in a national park in the Netherlands. Section 4 presents the results of the analysis and Section 5 discusses the most important findings. In the concluding section we briefly review the proposed approach and identify its current limitations and possible solutions.

4.2 The proposed approach

We want to represent the flow of visitors in a recreational area, defined as the aggregated movement of people visiting different places in a generalised sequence, regardless of the route followed by each individual (i.e., visitors may follow different routes, but a flow exist if the places are visited in a similar order). To represent this flow, we need to uncover spatial and temporal structures describing the visited places and how they are related in space and time. In other words, this flow is a quantitative and qualitative description of the aggregated spatial behaviour of the visitors. It can be graphically represented on a map by arrows between places (Tobler, 2003) and expanded using a space-time cube representation (Hägerstrand, 1970; Kwan, 2004), which we adapted to represent the sequential order in the Z-axis. This visual representation shows the general structure of the flow at the global level, as well as the local level of movement, the single elements of the flow representing the relations between the places. It aids the analysis of the way in which people use the area and interact with different geographical features.

We propose an exploratory approach to analysing the flow of visitors in natural areas using GPS data. The proposed approach has three aims: a) to determine the main places visited by the people in a recreational area by detecting Movement Suspension Patterns (MSPs); b) to establish the sequence in which each individual visited those places; and c) to detect commonalities in those sequences by extracting Generalised Sequential Patterns (GSPs).

Movement Suspension Patterns (Orellana and Wachowicz, 2011) denote the suspension of movement associated with places where people stop. MSPs are therefore spatial structures and are used to discover the places of interest to visitors. As MSPs are determined by the spatial-statistical properties of the whole dataset, no spatial or temporal thresholds are required. Generalised Sequential Patterns (Agrawal and Srikant, 1995) describe the sequence in which the places are visited, regardless of the trajectory followed. The term 'generalised' implies a relative order and not an absolute order: GSPs are temporal structures used to find commonalities in the order that places are visited. We believe that these two kinds of patterns aid understanding of the spatial behaviour of visitors. The patterns are used to analyse the overall flow of visitors and provide insights into how they interact with relevant places.

The movement of visitors is analysed in consecutive steps, using different techniques (e.g., spatial statistics, data mining and visual exploratory analysis). Each step results in a new dataset representing some specific characteristics of movement. The original data consist of a set of tuples representing the spatiotemporal coordinates of the people, which are recorded using positioning devices (e.g., GPS loggers). First, we compute movement parameters, such as speed and bearing, which generates a dataset of movement vectors representing the properties of movement observed in space and time. Second, we apply a spatial-statistical method to detect MSPs in the movement vectors dataset, which generates a set of spatial clusters representing the locations and times when visitors stopped. These data are used to find the points of interest to the visitors. Finally, we use a data-mining algorithm to extract GSPs that indicate the relative temporal sequence in which the places are visited. The result is a directed graph representing the frequent generalised sequences for those places. These patterns thus represent the aggregated flow of visitors in the study area. Figure 4–1 presents a graphical schema of the different steps of the process.

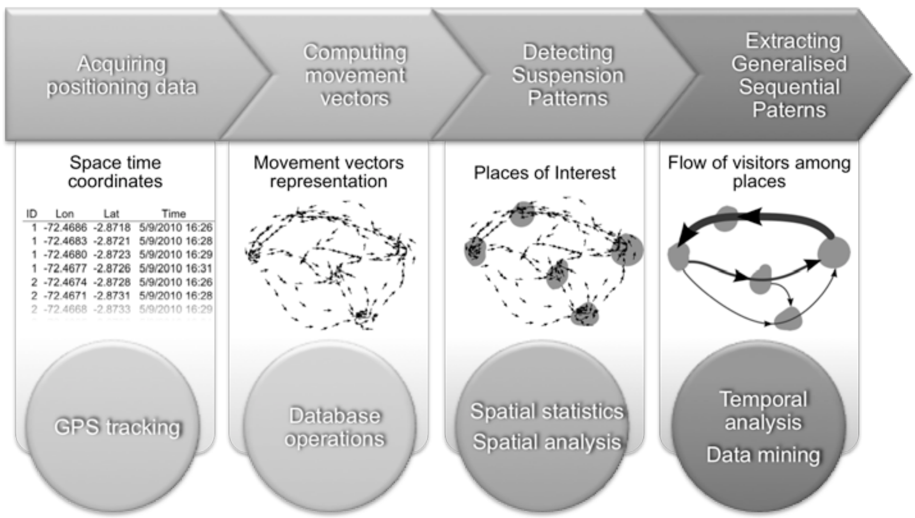

Figure 4–1. Schematic description of the proposed approach: First, movement data is captured using GPS devices and imported into a geodatabase. Movement vectors are then computed from observational data using a dedicated database procedure. Next, movement suspension patterns are obtained from spatial statistics and spatial analysis. Finally, generalised sequential patterns are extracted from the temporal sequences of suspension patterns using a data-mining algorithm.

4.2.1 Acquiring positioning data

The movement of visitors can be recorded using Global Navigation Satellite Technology (GNSS), such as GPS receivers and GPS-enabled smartphones. These devices are capable of capturing the location of users in space and time with a certain periodicity. The spatiotemporal positions of a large number of devices tracked in an area during a period of time provide a good basis for analysing collective movement in that area. Preconfigured GPS devices can be handed to the visitors at the entrances of the area. The captured data is stored in one of the standard formats, such as NMEA sentences (National Marine Electronics Association, 2010) or GPX (Topografix, 2010), and can be transmitted in real time to a server via a GPRS or 3G signal (for GPS-enabled smartphones) or imported later (for GPS receivers).

4.2.2 Computing movement vectors

Movement vectors represent individual observations of movement that can be measured or sensed at a particular place at a particular time. They are defined by the spatiotemporal coordinates of the observation coupled with a magnitude (e.g., speed, acceleration) and an angle (e.g., direction of movement), and can be represented graphically by an arrow. A vector space representation is a set of movement vectors of one or more entities moving in a defined spatiotemporal area. Movement is thus conceived as a spatial property (i.e., how movement is observed in space) rather than as a property of the trajectory of a particular entity (i.e., how does a specific person move).

The speed and bearing values of movement vectors can be derived in real time, depending on the capabilities of the device used to capture the space-time positions. Since not all the devices have those capabilities, a simple computer procedure can be used after the data is collected. The procedure takes two consecutive GPS observations and derives movement parameters (speed, distance, bearing and time step). If two consecutive observations are too separated in space or time (e.g., because the GPS signal is lost or because the GPS is turned off), this separation can be parameterised in the procedure to avoid errors in the computation of movement vectors. The procedure is publicly available online at http://ideasonmovement.wordpress.com. Movement vectors can be used to explore aggregated properties of movement, such as distribution and density and the global and local statistics of speed, direction and other movement parameters.

4.2.3 Detecting Movement Suspension Patterns

The local statistics of the set of movement vectors can be used to identify spatial clusters of low speed values. In Orellana and Wachowicz (2011) the authors demonstrated the use of a Local Indicator of Spatial Association (LISA) (Anselin, 1995) to find these clusters and detect Movement Suspension Patterns (MSPs). These patterns may indicate the location of geographical features associated with the reduction in speed that characterises the stopping behaviour of pedestrians.

Although MSPs are similar to the concept of 'stops' used in other approaches, they are essentially different. Whereas 'stops' are the parts of an object's trajectory where the object does not move, MSPs are spatial clusters of low speed vectors with a strong spatial association. This difference means that if only one individual stopped for a short time at a place where other individuals continued walking, this may not considered to be a MSP, because the spatial association of the speed values of movement vectors is not strong. This is an advantage for the analysis of collective spatial behaviour. The LISA method was selected because it tends to have the best statistical properties and requires few assumptions about the data. The time and duration of each MSP was obtained from the timestamps of the first and the last movement vectors of each visitor in each spatial cluster.

Figure 4–2 a shows an example of a vector-based representation of the movement of four visitors and contains four spatial clusters of movement suspension (numbered 1 to 4). Each cluster consists of the movement vectors of different visitors (numbered from A to D). Figure 4–2b depicts the temporal dimension of the MSPs. The vectors classified as suspension are plotted on the timeline of the corresponding spatial cluster, with different markers representing different visitors. The lines above each group of movement vectors represent the duration of each individual MSP.

4.2.4 Extracting Generalized Sequential Patterns

Generalised Sequential Patterns (GSPs) (Agrawal and Srikant, 1995) are the frequent generalised sequences that can be found in a timely-ordered set of events (in this case, the events are MSPs). They are 'generalised' because the MSPs occurred in a relative order rather than an absolute order. The assumption here is that, given a set of MSPs for a group of people, with their corresponding locations and times, there are structures in the relative order that characterise the collective spatial behaviour of the group.

Using data-mining techniques, GSPs are extracted from the original sequences of MSPs. Each GSP is an ordered list of MSPs (represented by the ID of the corresponding spatial cluster) together with a support value. The support value is the ratio between the number of sequences corresponding to the pattern and the total number of sequences in the dataset (Agrawal and Srikant, 1995). The example provided in Figure 2c shows the sequences of MSPs for each visitor (i.e., the order in which each visitor visited the different places). Figure 2d enumerates three GSPs with the corresponding support values. An interpretation of this example in terms of spatial behaviour is that there are four main places where people stop. In addition, the flow of visitors goes from the signpost (3) to the cafeteria (4) and then to the car park (1), with some visitors deviating via the monument. Moreover, all the people visited the monument (2) before stopping at the car park (1).

Since the potential number of extracted GSPs can be high, the most salient cases are selected during the exploratory analysis, which requires objective and subjective criteria of interestingness. Although support is the most frequently used objective measure of interestingness, there is some criticism that it does not provide the flexibility an exploratory approach needs, because it selects the most common cases and dismisses the uncommon, and probably interesting, cases (Laube, 2009). Depending on the application, potentially useful measures are complementary subjective criteria such as novelty and actionability (Han and Kamber, 2006). These subjective criteria rely not only on the input data, but also on the user examination of the pattern. In general, a pattern can be interesting if it is 'surprising' to users (novelty) or if they can 'do something' with the pattern (actionability) (Silberschatz and Tuzhilin, 1995). For example, a pattern can be considered interesting even when it has a low support value if it reveals a new, unexpected flow of visitors between places that were not initially considered. Similarly, a pattern may be interesting to a park manager if its discovery can be used to improve management practices. In our approach, we combined both objective (i.e., support) and subjective (i.e., novelty) criteria during the exploratory analysis of the flow of visitors.

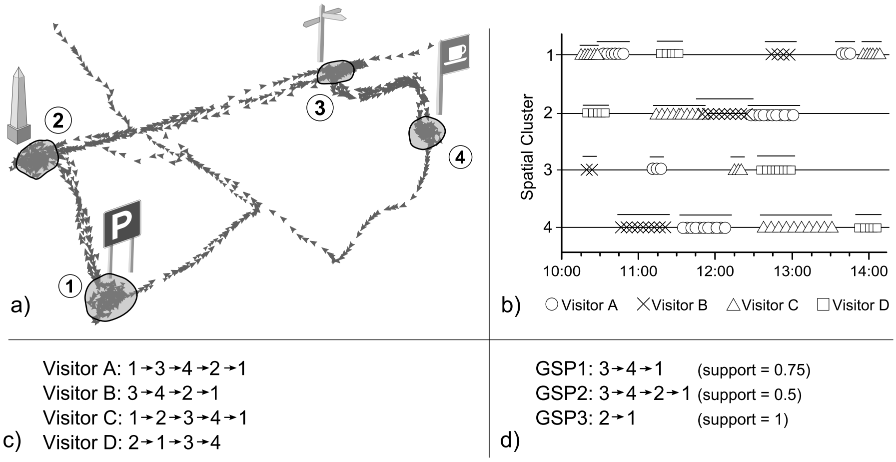

Figure 4–2. Examples of different representations of the movement of four visitors. a) Vector-based representation with spatial clusters of movement suspension located at four points of interest. b) Temporal duration of movement suspension patterns for each cluster. c) Sequence of suspension patterns for each visitor. d) Three examples of generalised sequential patterns and their corresponding support values.

4.3 Implementation

The proposed approach was used to explore the spatial behaviour of visitors in the Dwingelderveld National Park in the Netherlands. We analysed the positioning data recorded by GPS devices carried by 372 visitors during their visit to the park. We selected three research questions to illustrate the use of the approach: a) *What are the main visited places in the park?* b) *What are the visitor flows from the entrances to the main places?* and c) *What are the visitor flows between the main places?* The answers to these questions will provide a quantitative and qualitative description of the aggregated flow of visitors in the park.

Starting with a movement vector representation, the LISA approach was used to detect Movement Suspension Patterns (MSPs). The MSPs were then compared with a geographical dataset, allowing us to answer question a). In the next step, Generalised Sequential Patterns (GSPs) were computed using the BIDE+ data-mining algorithm (Wang, Han and Li, 2007) and compared for each of the five entrances in the park to answer question b). Finally, the GSPs between the main places were analysed to answer question c).

4.3.1 Study area

The Dwingelderveld National Park (DNP) is an area of about 3,700 ha in the north-east of the Netherlands. It is a typical Dutch recreational area with an extensive network of short strolls (60 km of marked trails, each less than 7 km in length) and long walks, as well as routes for cycling and horse riding. The landscape consists mainly of dry and wet heath lands, pine and deciduous forest, and an important complex of juniper shrubs. Dwingelderveld is a very popular area and receives between 1.5 and 2 million visitors each year. Besides the wetlands, sheep farms and some bird-watching hides, which are the main tourist attractions, the park contains additional amenities for visitors, such as staffed and unstaffed information centres, a tea house and some cultural attractions, including a historic house and a radio telescope (van Marwijk, 2009). Visitors enter and leave the park through one of the five access points (where car parks are located) and follow the paths to one or more points of interest or pursue various leisure activities.

Three different datasets were used. The first was a positioning dataset recorded by GPS receivers given to visitors at the entrances of the park (the beginning of the GPS track). Of the 461 visitors asked to participate, 400 agreed to carry a GPS device during their visit. An evaluation of the quality and completeness of the data led to the inclusion of 372 GPS tracks in the final dataset, which contained about 142,000 time-stamped geographical coordinates. This data was collected over a seven day period (weekend and weekdays) in the spring and summer of 2006 (details of the data collection can be found in van Marwijk, 2009). The second dataset was a map showing the path network and the locations of the park entrances. The third dataset was a collection of 271 points representing the locations of the attractions and facilities in the park, gathered from specialised web pages (Natuurmonumenten, 2009; Pol-Recreatie, 2003) and a field survey. Figure 4–3 is a map of the study area.

4.3.2 Finding the main visited places

The first task was to identify the places of interest in the park. We define 'main place' as a site where a movement suspension pattern is detected which can be associated with a relevant geographical feature that can explain the suspension of movement. The assumption here is that visitors are attracted to these places and temporally suspend their movement to perform some activity associated with the place. Some examples are visiting an interesting spot, reading an information board, eating or resting at a picnic bench, etc. These attractions affect the collective movement by shaping the flow of visitors.

We computed a set of movement vectors representing the movement of visitors using data from the positioning GPS dataset and stored them in a geodatabase. The Local Moran's index was computed for the movement vectors using an adaptive neighbourhood defined by a radius of 50 metres and a minimum of 15 observations. These parameters were defined in an exploratory analysis of the spatial distribution of the dataset. The movement vectors with speeds less than the mean and Z-scores above 1.96 (5% significance level) were classified as movement suspension and they showed spatial clusters when plotted on a map. To define the boundaries of the spatial clusters, we used a kernel-density function on the set of vectors classified as movement suspension to obtain a continuous surface representing a density estimate. Percent Volume Contours (PVCs) traced on this surface represent the boundaries of the areas that contains x% of the volume of the probability density distribution. For example, using a value of 99, the PVCs delineate the areas containing on average 99% of the vectors that were used to generate the kernel density estimate (Beyer, 2010). The bandwidth for the kernel-density function was determined by the estimated accuracy of the observation (e.g., 10 m for a GPS receiver under ideal conditions). These lines represent the boundaries of compact spatial clusters of suspension of movement. Moreover, the kernel-density function allowed us to find hotspots of movement suspension and differentiate outliers (e.g., clusters with only one movement vector).

GIS overlaying was used to compare the spatial clusters with a point dataset containing the attractions and facilities in the park. An attraction or facility was associated with a spatial cluster if its location lies inside the 99% PVC. However, some attractions were landscape elements outside a cluster but visible from within the cluster (e.g., a water body). In these cases, maps, aerial imagery interpretation and field verification were used to associate the cluster with the corresponding attraction. The relative importance of each place was evaluated using aggregated statistics for the cluster, such as the number of visitors and the number of MSPs. These results were used to answer the first question, What are the main visited places in the park?

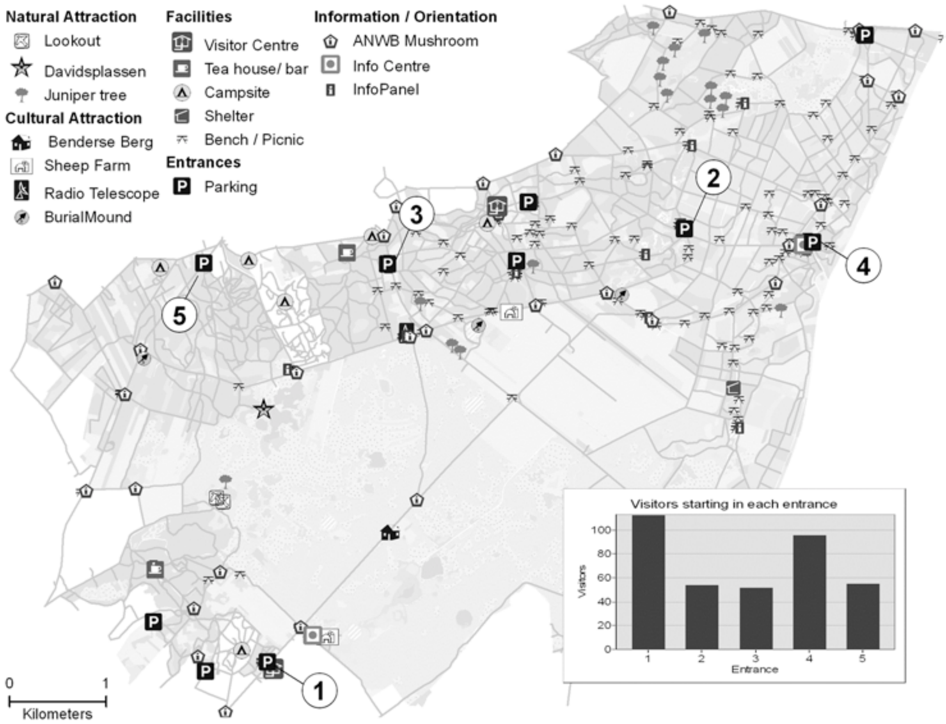

Figure 4–3. Map of the Dwingelderveld National Park showing the five entrances and the locations of the main attractions and facilities.

4.3.3 Exploring the aggregated flow of visitors

The second and third questions concern the aggregated flow of visitors. This flow is understood to be the aggregated collective movement of visitors between the places and is represented by Generalised Sequential Patterns (GSPs).

We implemented a database procedure in which the MSP dataset obtained in the previous step was analysed to determine the temporal sequences for each spatial cluster and visitor. The database was updated by adding to each MSP an integer indicating the position in the sequence and an identifier for the individual sequence. The result was a dataset of individual sequences of ordered MSPs. This dataset was analysed using the BIDE+ algorithm implemented in the Sequential Pattern Mining Framework, a publicly available JAVA code for analysing sequential patterns (Fournier-Viger et al., 2008). We selected this algorithm because it avoids redundancy in the results by extracting only ‘closed’ sequential patterns. ‘Closed patterns’ are sequences that are not contained in another sequence having the same support. A closed pattern induces an equivalence class of patterns sharing the same closure, and those patterns are partially ordered, e.g., according to the inclusion relation. The smallest elements in the equivalence class are called minimal generators, and the unique maximal element is called the closed pattern (Fournier-Viger et al., 2008). We modified the original code of the algorithm to produce a formatted output consisting of two tables that can be imported back into the geodatabase. The first table contained the structure of the GSP, in the form of an ordered list of spatial clusters where the MSPs occurred. The second table contained the properties of each GSP, consisting of the ID of the GSP and the values for support, frequency and size.

The results were represented in a dynamic map linked to the geodatabase to indicate the flow of visitors between the main places. This linkage enabled a dynamic visualisation by querying and filtering the datasets.

The second question, *What are the visitor flows from the entrances to the main places?* is answered by exploring the GSPs starting from each entrance in the park. We analysed the GSPs with the largest support values, representing the aggregated flow of visitors from the entrance to the main places. The results were shown on a map using arrows to indicate the direction of the aggregated flow of visitors to each place.

We took a similar approach to answer the last question, *What are the visitor flows between the main places?* The exploratory analysis was performed in a 3D sequential space-time cube, the base of the cube being a two-dimensional map of the park and the Z-axis representing the sequential time. The GSPs were rendered in the cube as three-dimensional polylines connecting the different places, ordered in time in the vertical axis, the first MSP at the bottom. The thickness and colour of the lines represent the support values of the GSPs and provided a visual cue for the analysis. To aid visual interpretation, vertical lines connect the places on the map to the MSPs. Moreover, each GSP was linked to a data-space representation, such as multi-dimensional scatter plots. This is a powerful exploratory tool and allowed us to interact with the data by querying, filtering and highlighting the GSP dataset in a multi-dimensional space. The resulting visualisations allowed a quick and effective interpretation of the GSPs and helped to uncover the structures in the flow of the visitors. For example, GSPs with relatively high support values implied that many people visited the places in that order, shaping a visible flow of visitors in the space-time cube.

4.4 Results

4.4.1 Most visited places in the park

Using the LISA index to classify the movement vectors, we found that 6.3% (n = 8,988) corresponded to movement suspension. We identified 184 spatial clusters defined by the 99% PVCs drawn on the kernel-density surface. These clusters contained in total 1,581 MSPs. By applying a spatial overlay function, we found that 158 spatial clusters (85.9%) could be associated to at least one relevant geographical feature in the park. These clusters contained 1,546 MSPs, or 97.8% of all the MSPs. This result allowed us to discover the places visited by the people participating in the data collection (Figure 4–4).

Assuming the degree of interest of a place can be related to the number of times it was visited (represented by the number of MSPs in the corresponding cluster), we analysed a subset containing the 10% (n = 18) largest clusters to provide an indication of the most interesting places for the visitors. Details of these clusters are reported in Table 4–1 and Figure 4–4.

The car parks at the five entrances were among the most visited places. In fact, all the visitors started and finished their visits there (e.g., Figure 4–4b, f). The remaining main visited places included attractions and amenities in the park. The most visited attraction in the park was the visitor centre close to Entrance 1 (70 visitors, Figure 4–4f), where the duration of the visits were typically long (more than 15 min.). Other main visited attractions were the radio telescope (41 visitors, Figure 4–4a), the sheep farm (32 visitors, Figure 4–4e), and some of the wetlands that characterise the landscape of the park and where picnic benches are located. Two of the wetlands that received a large number of visitors are Davidsplassen (31 visitors) and Smitsveen, also the site of an ancient burial mound (17 visitors, Figure 4–4c). The average duration of the visits was longer at the radio telescope and the sheep farm (about 7 min.) than at the wetlands and other attractions (less than 5 min.).

Table 4–1. The 10% largest clusters and their related geographical features. The figures indicate the number of different visitors stopping at the place, the number of Movement Suspension Patterns, the number of vectors classified as suspension and the average duration of each MSP.

Cluster (Id)	Associated Feature	Visitors (n)	MSP (n)	Vectors (n)	Avg. Duration (min:sec)
2	Car park	107	207	1212	03:01
1	Visitor centre	70	82	2011	16:25
19	Car park	60	92	201	07:19
4	Picnic/ANWB Mushroom	59	74	210	02:01
12	Radio Telescope	41	47	465	07:27
16	Car park	37	70	257	03:58
18	Snack Bar/Info Spier	37	46	84	01:32
15	Car park	33	57	198	05:35
3	Sheep farm	32	32	771	07:34
42	Information at Sheep Farm	31	47	321	04:37
10	Davidsplassen / Picnic	31	31	188	02:13
33	Cross path	27	28	60	00:19
21	Car park	25	45	152	07:38
5	Wetland / Picnic	25	25	132	04:09
7	Tea-house	24	24	126	27:50
6	ANWB Mushroom	20	20	35	00:37
69	ANWB Mushroom	19	19	102	01:05
110	Smitsveen / Burial mound	17	20	96	03:08

The most visited amenities were those providing information and orientation, including the information point nears the entrance at Spier (37 visitors, Figure 4–4b) and the one close to the sheep farm (31 visitors, Figure 4–4e). The other main visited places providing information were some of the mushroom-shaped ANWB signposts (e.g., Figure 4–4d), where the MSPs were of a short duration. Another frequently visited facility was the teahouse in the forest (Figure 4–4h), with the longest average duration of visits (27 min. 50 sec.).

It is interesting that, besides the attractions and amenities, one spatial cluster of MSPs is located at a path crossing (Figure 4–4g), where 27 visitors stopped. This pattern may indicate a specific spatial behaviour of visitors arriving at a crossing and choosing which direction to take. The short average duration of the MSP at this cluster (19 seconds) supports this interpretation.

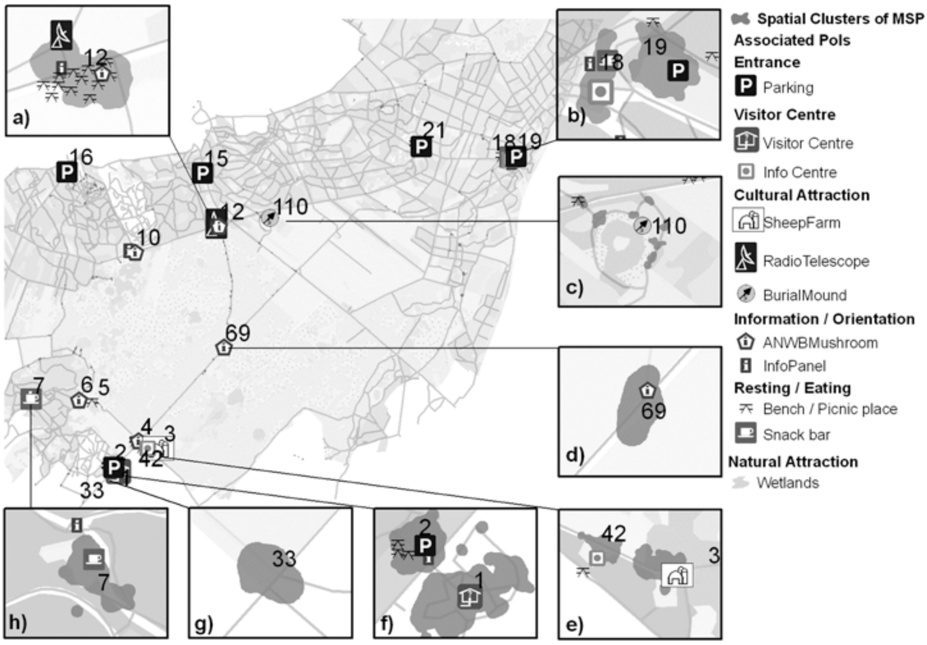

Figure 4–4. Map showing the spatial clusters of Movement Suspension Patterns and associated geographical features. The numbers are the IDs of the spatial clusters. Detailed examples are depicted in the insets.

4.4.2 Flow from the entrances to main places

The temporal analysis of the set of MSPs resulted in 282 sequences of MSPs, representing the relative order in which people visited each place. Since a sequence is an ordered set of MSPs, the number of sequences equals the number of people who visited at least two different places (i.e., at least two MSPs detected at different spatial clusters). The BIDE+ algorithm extracted 218 GSPs with a minimum support of 0.02. This value was low enough to capture uncommon GSPs to allow further filtering to select cases with higher support values. We used a query to extract GSPs in which the first element corresponded to an MSP located at an entrance and found 16 GSPs representing the flow of visitors from the entrances to the main places. This result is reported in Table 4–2 and Figure 4–5.

The shape of the flow of visitors from the entrances to the main attractions reflects the relative importance of Entrance 1 (Figure 4–5), where most of the people in the study started their visit to the park. Many of the main attractions are easily reachable from this entrance, which may also explain the large proportion of GSPs starting there. This interpretation is supported by the fact that the largest support values corresponded to the GSPs with the shortest lengths (measured as the Euclidean distance between the origin and destination of the flow). This implies that people usually choose an entrance near to their preferred places as the starting point of their visit to the park. For example, the flow to the teahouse, the sheep farm and the visitor centre came from Entrance 1. Likewise, the GSP representing the flow of visitors from Entrance 5 to the wetlands in Davidsplassen had larger support than the one from the more distant Entrance 1. We found two exceptions: the radio telescope and the wetlands at Davidsplassen and Smitsveen received visitors from two entrances. Another interesting finding is the flow from Entrance 1 to cluster 4 (Figure 4–5, lower inset). This cluster is located at a path crossing, which also has a signpost and a picnic bench. Visitors seemed to stop there before deciding which route to follow to continue the visit.

Table 4–2. Generalized Sequential Patterns from the entrances to the main places.

From (Entrance ID)	To (Cluster ID)	Geographical feature	Support	Frequency	Distance (m)
1	110	Wetland / Burial Mound	0.02	7	4099
1	10	Davidsplassen	0.03	9	3025
1	69	ANWB Mushroom	0.03	10	2283
1	12	Radio Telescope	0.04	15	3683
1	6	ANWB Mushroom	0.04	13	1057
1	7	Tea House	0.04	14	1483
1	33	Path Crossing	0.06	21	188
1	5	Wetland / Picnic	0.06	19	940
1	42	Information at Sheep Farm	0.07	25	576
1	3	Sheep Farm	0.08	27	741
1	4	Picnic / ANWB Mushroom	0.13	46	521
1	1	Visitor Centre	0.16	55	102
3	110	Smitsveen / Burial mound	0.02	6	1184
3	12	Radio Telescope	0.05	16	777
5	10	Davidsplassen	0.04	13	1483
4	18	Snack Bar / Info centre	0.06	20	75

4.4.3 Flow between the main places

The first visualisation created to explore the flow in the space-time cube was a display of all GSPs (Figure 4–6). This visualisation revealed that GSPs with high support values (*support* ≥ 0.1) usually consisted of only two or three MSPs. They all started at a car park, went to a nearby place and then came back to the starting point. The more MSPs the GSP had, the lower the support value was. This is an expected result because shorter GSPs represent relatively simple flows that many visitors may follow (e.g., Entrance 1 →Visitor Centre → Entrance 1). But GSPs are nested structures, and so shorter GSPs are parts of longer GSPs, which also explains the differences in support values. At this point, it is important to remember that GSPs imply a generalised order, and the example given captures the flow of all the visitors visiting the places in that order, regardless of whether they visited other places in between or not.

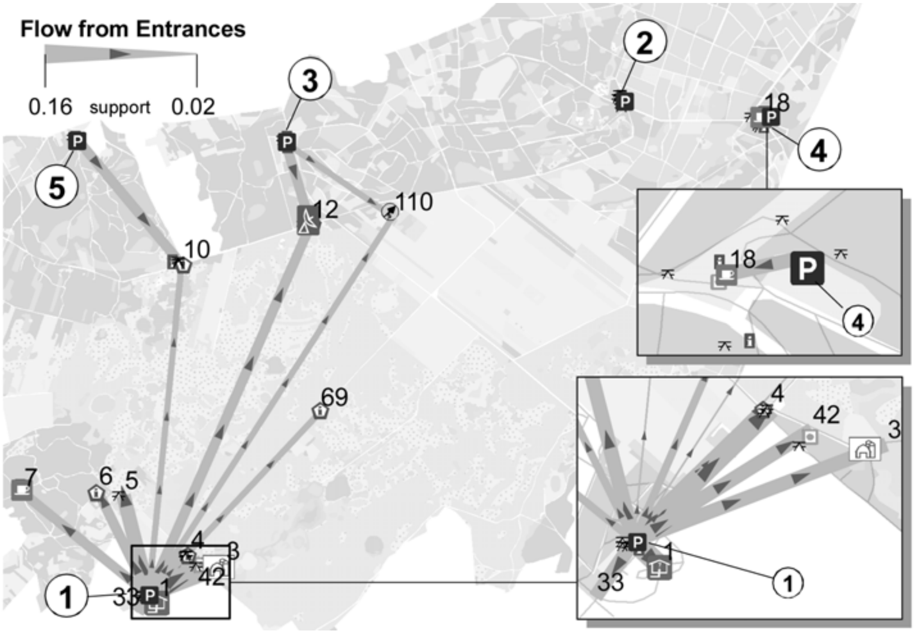

Figure 4–5. Flow from each entrance to the main visited places. The numbers in circles are the identifiers of the entrances and the plain numbers are the identifiers of the spatial clusters associated with each place. The thickness of the lines representing the flow is proportional to the corresponding support value.

The capabilities for dynamic interaction with the data allowed us to explore interesting structures and perform comparisons. For example, we found an interesting example of a GSP with many MSPs and relative high support values (highlighted in Figure 4–6). This example represents the flow of people who visited the places in the following sequence: Entrance 1 → Path crossing → Info Centre → Sheep farm → Info Centre → Entrance 1 (s = 0.05).

When we filtered the GSPs for specific places, some properties of the flow at those places were revealed. For example, from the GSPs associated with the radio telescope shown in Figure 4–7, it is possible to see that the tea house and Davidsplassen were typically visited after the radio telescope (flow lines go upwards from the radio telescope to those places), whereas the historic house and some of the wetlands were visited before the radio telescope (flow lines go upwards from those places to the radio telescope). An example of a GSP with more MSPs is Entrance 1 → Burial mound → Radio Telescope → Entrance 1. The support for this GSP was 0.02 (highlighted in Figure 4–7).

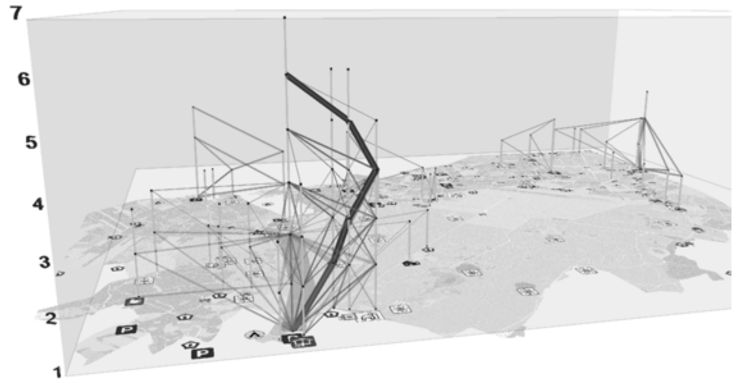

Figure 4–6. Sequential space-time cube representing Generalised Sequential Patterns with support 0.03. The vertical axis indicates the sequence in which the places in each GSP were visited. Dark thick lines indicate high support values; light thin lines indicate small support values. An interesting pattern is highlighted in black.

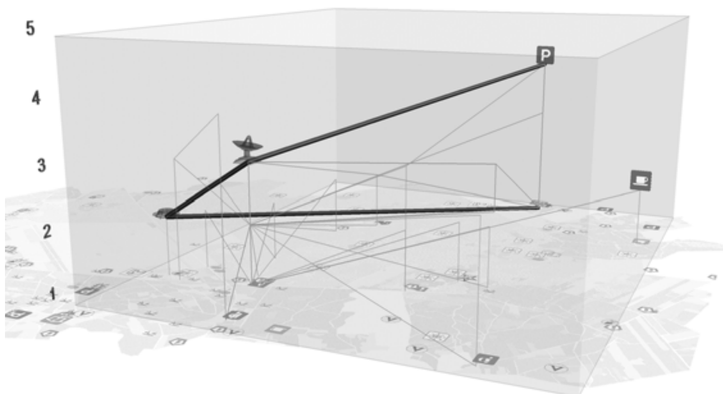

Figure 4–7. An example of the sequential space-time cube for the flow of visitors to the radio telescope. The vertical axis represents the temporal sequence, the thin diagonal lines represent the flow of visitors between the places, and the vertical lines link the clusters to the places on the base map. The highlighted Generalised Sequential Pattern represents the flow of visitors starting at Entrance 1, visiting the burial mound and then the radio telescope, before going back to the initial point (support = 0.02).

Using the filtering and linking functions, we found other interesting properties of the spatial behaviour of visitors. For example, the visitor centre was always visited before the sheep farm, but the nearest information centre was visited both after and before it. Also, of all the visitors going from Entrance 5 to Davidsplassen, 78% had first visited one of the wetlands near the entrance. Finally, of all the visitors that went from Entrance 1 to the teahouse, half went first to the visitor centre. There was no flow of visitors from the teahouse to the visitor centre.

In general, we found that GSPs with low support values were more common than those with high values. For example, almost 80% of the GSPs had a support value equal to or less than 0.04% (Figure 4–8a). This means that there were many different sequences followed by few visitors and only few sequences followed by many visitors (Figure 4–8a). The visitor flow is therefore made up of many different patterns. We also found that short GSPs (with three or four MSPs) were more common than GSPs with more MSPs (Figure 4–8b), indicating that the flow of visitors consists mainly of common, simple sequences.

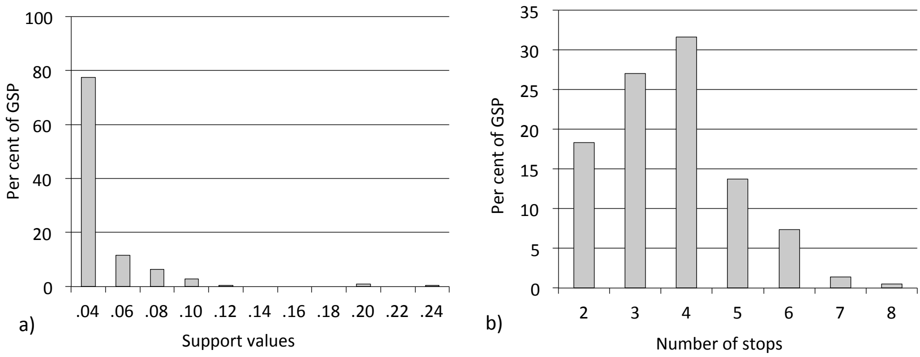

Figure 4–8. a) The large percentage of Generalised Sequential Patterns with low support values may indicate a large diversity in the flow of visitors (many sequences followed by few visitors and few sequences followed by many visitors). b) 77% of the GSPs had four Movement Suspension Patterns or less.

4.5 Discussion

The results obtained to answer the first question, What are the main visited places in the park? indicate that despite the large number of points of interest in Dwingelderveld National Park, only a limited number of them attracted significant numbers of visitors and many were hardly visited at all. The commonly visited places can be classified according to their functions as attractions and amenities. Attractions are places mainly associated with natural and cultural leisure activities, where people go to experience and enjoy the landscape and features in the park. Examples of these places are the radio telescope, the sheep farm and the wetlands in Davidsplassen and Smitsveen. Amenities are places with information facilities and services for visitors. Examples are the tea house, the information centres, the information boards and poles, and the picnic benches. Interestingly, besides these two kinds of places, we found that visitors also stopped at some path crossings. Our interpretation is that these are places where visitors temporary suspended their movement to decide which path to take.

The results obtained to answer the second question, What are the visitor flows from the entrances to the main places? show that the flow of visitors to the main places came mainly from the nearest entrance. Few places received a flow of visitors from two entrances. This indicates that the entrance selected as the starting point largely determines the places that are visited, suggesting that the flow of visitors may follow a gravity model.

The results obtained to answer the third question, What are the visitor flows between the main places? indicate a great diversity in the flow of visitors in the park. Although few places were visited often, they were hardly ever visited in the same order. This result is interesting since Dwingelderveld National Park, like many natural parks in the Netherlands, has predefined routes that are well marked on maps available to the visitors, as well as on the information boards and signposts. This result seems to contradict the findings of van Marwijk (2009) who reported that 66% of the visitors follow a predefined route in his experiment. Two facts should be borne in mind, though. First, visitors can follow the routes in both directions, changing the order of visited places. Second, the figures reported in van Marwijk's study are based on visitors' answers to a survey after they finished their visit, and not on GPS positioning data.

The answers to the three questions we posed at the beginning illustrate the suitability of our approach to analysing the flow of visitors. The proposed approach has some advantages over previous methods mentioned in the introduction. One advantage is that MSPs, which constitute the building blocks of the analysis, do not need spatial or temporal thresholds. In addition, using the BIDE+ algorithm to detect GSPs helps to avoid redundancy in the results, making it easier to explore and interpret them. Moreover, the combination of different methods provides the flexibility required for an exploratory approach. Our approach is also strongly related to the geographical context in which movement occurs, which helps with interpreting the meaning of the movement patterns. An additional advantage is that all the steps can be performed using publicly available GIS software and open source code – an important advantage for applying the approach in other areas.

The proposed approach can provide useful information for the design, implementation and monitoring of management practices in natural recreational areas. Park managers, for example, can use the proposed approach to assess the popularity of different places in their area and understand the flow of people from the park entrances to those places. This in turn can be used to evaluate the location of signs, design visitor routes and manage the flow of visitors to avoid crowding.

Although this analysis was restricted to the spatial behaviour of the visitors, their socio-economic backgrounds, purposes and motivations, can also be included to differentiate and compare different target groups. Moreover, the flows of different groups and their changes over time can be used as indicators of coping mechanisms in crowded areas (Manning and Valliere, 2001) and for studying possible processes of exclusion and domination (Ostermann, 2009).

The implications of these results are also of potential interest to tourism businesses. Holyoak and Carson (2009) identify several areas that can benefit from the study of the movement of visitors in a broader, regional context. Managers can segment the market and offer more diverse and focused options for specific groups of visitors. Information and marketing strategies can take into account the order in which the places are visited to provide relevant and appropriate information about the destinations. Moreover, researchers in recreational areas can study the flow of visitors to support the design of other data collection methods, such as deciding where and when conduct interviews, surveys and locate direct observers.

Although these results describe the flow of visitors in a relative small area, the approach can be used in larger areas and for longer periods of time. In fact, one of the advantages of Movement Suspension Patterns is that they are derived from the spatial-statistical properties of the dataset, and can therefore work at very different scales. Theoretically, the lower limit corresponds to the spatial accuracy of the data (i.e., it would be not possible to differentiate flows in an area of 10 metres radius for data collected with single-frequency GPS receivers without any signal correction) and the upper limit is set by the study design (the area in which the participants will be monitored). In this study, preconfigured GPS receivers were given to the visitors at the entrances of the park. Other portable devices, such as the visitors' own smartphones or tablets could be used to complement data collection. For example, visitors willing to participate in the research could install tracking applications in their devices, in return for receiving location-aware information during their visit to the area.

4.6 Conclusions

Understanding the spatial behaviour of visitors in recreational natural areas is a key issue for effective management. Using GPS tracking technology, managers and researchers can collect data on the routes followed by individuals to analyse how they interact with the geographical features in the area. When the movement of several people is analysed, some patterns may emerge indicating the existence of common structures in the spatial behaviour. In line with this idea, we suggested that movement patterns representing the flow of visitors could further our understanding of the collective spatial behaviour of the visitors.

In this article, we present a novel approach to exploring the flow of visitors in natural areas through the combined analysis of two kinds of movement patterns extracted from GPS positioning data. Movement Suspension Patterns were useful for uncovering the main attractions, while Generalised Suspension Patterns were helpful in understand common structures in the order that those places were visited. We demonstrated the application of the approach by analysing the movement of visitors in the Dwingelderveld National Park. The results suggest that the proposed approach helps us to understand the aggregated spatial behaviour of the visitors in the park.

In future research, we are envisaging new ways to build the sequential space-time cube to allow a better dynamic exploration of the detected GSPs. We are also developing a better method to establish the duration of the individual MSPs using hierarchical clustering.

Acknowledgments: The authors thank Ramona van Marwijk and Rene Jochem from Wageningen University, for the Dwingelderveld National Park data collection.

5

Developing an Ontology of Interactions for characterizing Pedestrian Movement Behaviour

Orellana, D., and Renso, C. (2010)

in Movement-Aware Applications for Sustainable Mobility: Technologies and Approaches. Wachowicz, M. (Ed) pp 62-86. IGI Global, New York.

Abstract: Since the introduction of Time Geography, the literature has witness a growing interest in representing and understanding human movement and its relationship with the environment. Although recent technology in personal tracking devices brought new potentialities in collecting and representing individual movements, methods to deal with the complexity and dynamism of collective movements are still lacking. This chapter introduces a spatial knowledge representation for the conceptualization of pedestrian movement as a complex system based on the interactions. Movement interactions are defined and classified to represented global characteristics of the movement as emergent properties other than as a set of individual properties. The devised approach is exemplified through a case study on characterizing visitor behaviour in the Dwingelderveld National Park in The Netherlands.

5.1 Introduction

The interest in representing the movement of people in order to understand their relationship with the environment dates back to the '70s, when Hägerstrand posed the basis of Time Geography. He studied the space-time path of individuals to identify the spatial and temporal constraints that characterize human movement (Hägerstrand, 1970). This idea has proved very suitable to represent the individual movement as an essential relationship between an individual and the environment, as well as between individuals. Moreover, analysis of pedestrian movement has been widely recognized to be essential in understanding human behaviour (Blythe, Miller and Todd, 1996). The space-time paths used by Hägerstrand (i.e., trajectories) have become the most common representation for human movement, due in part to its intuitive visualization and interpretation, but also to its feasibility to directly represent data from tracking technologies. As a result, several technologies and tools have been developed for the collection, storage and recovery of large trajectory data sets (Güting and Schneider, 2005; Manco et al., 2008; Ortale et al., 2008; Pfoser and Jensen, 2001; Renso et al., 2008), analysis (Andersson, et al., 2008) and visualization (G. Andrienko et al., 2008). Although trajectories have proved to be useful in representing the movement of individuals, they seem to be an inadequate representation to deal with the complexity and dynamism of collective movement and the interactions that are the cause and the consequence of it. For example, the spatio-temporal path of an individual would have different meanings depending on the presence or absence of other individuals as well as the relationship of the individual with the environment; the trajectory-based representation is not enough sophisticated to represent these different meanings. This idea will be further developed in the next sections.

Our research assumption is that this limitation arises from an underestimation of the representation's roles and implications. Davis stated in (Davis et al., 1993) that a "knowledge representation" has five roles that should be taken into account: a surrogate for the real world, a set of ontological commitments, a fragmentary theory of reasoning, a medium for efficient computation and a medium for human expression. We believe that the broad use of the trajectory-based representation has led to an underestimation of some of these roles. Mainly, the ontological commitments of this representation (i.e., what is represented and what is not) have been undervalued. For example, there is a commitment that the movement of an individual starts at the first point of the trajectory and ends at the last point, and that the whole movement can be represented as an interpolation of the intermediate points, which is not necessarily true. Another commitment is that intermediate points would represent either stops or moves, but this distinction, although intuitive, is not always precise, since individuals are hardly ever still and therefore the conceptualization of stops depends on the scale. Consider for example, a commuter's trajectory: It would have long stops (i.e., home, office) and short stops (i.e., kiosk, traffic lights), but the people are not completely static in these places and it would even be interesting to represent the movement inside those “stops”. As we can see, these commitments limit the possibilities of the representation.

The main contribution of this chapter is the description of a Spatial Knowledge Representation for conceptualization of pedestrian movement. Indeed, pedestrian movement can be conceptualized as a dynamic complex system based on the interactions that happen between individuals as well as between individuals and the environment in which they move. This conceptualization becomes useful since it allows for the representation of some global characteristics of the movement as emergent properties rather than as a set of individual properties. Moreover, it emphasizes the feedback effects of the movement of a pedestrian on other pedestrians and the environment. These characteristics and effects are the basis of the interactions based representation.

This representation is formalized according to the five previously mentioned roles. However, special emphasis will be placed on the first three roles, which allows us to answer the following questions:

- What are the movement interactions?
- What elements and relations are important to consider?

- How can we interpret them?

On the one hand, we propose a simple taxonomy of pedestrian interactions and we show how this taxonomy can be part of an ontology of interactions, where concepts like Context, Behaviour and Patterns are combined to offer a more general understanding of the essence of pedestrian movement behaviour. On the other hand, we have developed an ontology of pedestrian behaviour based on the proposed interactions. This last aspect benefits from the reasoning capabilities provided by formal ontologies in defining behaviour in terms of combination of interactions, movement patterns and contextual information. The viability of our approach to analyse the collective pedestrian movement will be demonstrated through the case study of visitors to the Dwingelderveld National Park in Netherlands.

The remaining parts of this chapter are organized as follows. Next section presents previous related research on pedestrian movement representation, patterns and behaviour. After, the focus is on interactions for pedestrian movements, where definitions of pedestrian interactions are given. We also propose an ontology for interactions where movement patterns, context, and interactions are combined to express pedestrian behaviour. Then, a case study about inferring the behaviour of visitors in a park is depicted, whereas conclusions are reported in the last section.

5.2 Related work

The tasks of detection, representation and interpretation of movement patterns have attracted the attention of several scientific communities. Important contributions have been made from such diverse areas as artificial intelligence, video surveillance, emergency management, environmental planning, human geography, transportation management, animal and human behaviour, and computational geometry, among others. Exhaustive review of research related to the representation of pedestrian movement exceeds the purpose of this chapter; nevertheless, we would like to mention some of the works that inspired us in the development of our approach.

The proposal of analysing people movement for the interpretation of their behaviour was first glimpsed in Time-Geography carried out by Hägerstrand, who showed how human activity can be determined not only by decisions made by individuals, but mainly by the spatial and temporal constraints they are subject to (Hägerstrand, 1970). Hägesrtrand's proposal of representing movement as spatio-temporal paths (the sequence of spatio-temporal positions of a moving object) still remains the foundation of most current approaches. Other authors are using different terms as "trajectories" (e.g., N. Andrienko et al., 2008) or "geospatial lifeline" (Hornsby and Egenhofer, 2002) for the same concept. Miller (2005) confronted the potential lack of rigour of Hägesrtrand's approach by extending the moving object database techniques to develop more rigorous definitions of basic concepts such as space-time path, prisms, bundles and intersections in the context of a measurement theory for Time Geography.

The most recent revision of Time-Geography was proposed by Yu and Shaw (2007), in which the authors propose integrating the representation of human activity together in both the physical and virtual worlds in order to study diverse types of interaction. Later, the authors recognized that trajectories themselves were not sufficient for the representation of patterns when using large movement data sets, and suggested the use of a generalized representation called Generalized Space-Time Paths (GSTP) (Shaw, Yu, and Bombom, 2008). More recently, they presented the implementation of this approach in a Space-Time GIS, where they demonstrated how the activities and interactions of individuals in the physical and virtual worlds could be suitably represented with Linear Reference Systems (LRS) in space-time paths (Shaw and Yu, 2009).

A different approach was proposed by Alvares et al. (2007) where authors addressed the complexity of analysing large trajectory data sets by enriching trajectories with semantic geographical information. In their work, the authors proposed the transformation of traditional raw trajectories into *semantic* trajectories, through a pre-processing of stops and moves (Spaccapietra et al., 2008). This promising approach is conceptually similar to that presented in this chapter; however we have expanded the formalization of movement patterns beyond stops and moves.

Daamen and Hoogendoorn (2003a, 2003b) described controlled experiments with a view to studying pedestrian behaviour in various situations. The authors represent movement at two levels: microscopic level (trajectories) and macroscopic level (flows). They also show how the former can be used to understand individual behaviour, while the latter are more useful for characterizing collective behaviour.

Moreover, a significant research project has been carried out on context-aware representation and reasoning of spatial knowledge. Location-Based Systems (LBS) use context to retrieve information that is relevant to the user, depending on their location. In an interesting study by Schlieder and Werner (2003), authors suggested that the location itself is not enough to determine the user's intentional behaviour, and proposed a location model that takes into consideration both the user's movement patterns and their context in order to infer behaviour. The main approach is based on the translation of movement patterns into sequences of intentions, assuming that these intentions are specific for each type of space and are characterized by the user's movement pattern.

Helbing et al. (1993) proposed a representation of pedestrian collective movement based on fluid dynamics models. The main assumption of this approach is that some of the fluids' properties could be also found in pedestrian collective movement. Therefore, the authors adapted the classic equations of fluid dynamics (particularly gas-kinetic) to integrate the effects of interactions between individuals and to represent movement as macroscopic flows. In a later work, the authors explored how non-linear interactions between individuals can result in surprisingly predictable emerging properties in collective movement, and this brought a new pedestrian behaviour model based on *social forces* where attraction and repulsion forces between pedestrians are represented (Helbing, et al., 2001). They have also explored interactions in which pedestrian movement both influences and is influenced by the environment, as happens, for example, in the formation of trails on deformable terrain.

The study of movement as the interaction between people and the environment is hardly new. Hillier et al. (1993) studied the effect of so-called "attractors", defined as points with a particular interest, or capable of generating flows of people movement between them. The authors determined that these attractors are only the multipliers of a more important effect caused by the configuration of travelable space, thus stressing the importance of representing the environment in order to understand people's movement. Following the same viewpoint Turner et al. (2001) approached the interactions between individuals and built-up environments through the study of visibility graphs, suggesting that some movement patterns, such as way-finding or route-choosing, are closely related to the visual perception of space.

In the field of computational geometry, several researchers have focused on studying techniques for pattern matching and recognition for movement data, proposing formalizations of patterns and developing algorithms for their detection (Gudmundsson et al., 2004; Andersson et al., 2008; Benkert et al., 2008). In particular, Laube, Imfeld, and Weibel (2005) proposed a complete approach in which the movement of individuals is represented as an analysis matrix having movement parameters of various individuals that are compared over the course of time. The authors also proposed a framework called REMO (RElative MOtion) for the formalization of possible movement patterns, as well as data mining algorithms for the detection of such patterns. Although these implementations were mainly carried out in data sets of animals' movement, this approach is perfectly valid for the analysis of human behaviour.

Data mining over trajectories is a new and promising research field aimed at investigating techniques to extract patterns from large datasets (Nanni et al., 2008). Clustering has been exploited to uncover a variety of global behavioural patterns, such as density-based clustering (Rinzivillo et al., 2008), moving clusters, and identifying groups of objects that move similarly and close to each other for a long time (Kalnis, Mamoulis and Bakiras, 2005). The identification of local patterns in movement data, i.e., of concise representations of interesting local behavioural patterns of moving objects, has been also a fertile area of research. Among them, trajectory pattern mining in (Giannotti et al., 2007) pursued to unveil sequences of temporally annotated spatial regions.

Despite the amount of research carried out on movement patterns, there appears to be no agreement on how they must be organized or represented. Recently

Dodge, Weibel, and Lautenschütz (2008) recognized this lack of organization and suggested a systemic representation of movement patterns. They proposed two main classes, namely *Generic Patterns* and *Behavioural Patterns,* and arranged subclasses of the first according to main movement parameters, whereas subclasses of the second were defined using the complex patterns in a specific domain. Dodge classification constitutes a first attempt to organize the existing definitions of movement patterns, but a more formal and comprehensive classification is still needed. Actually, Wood and Galton (2008) investigated the relationship between collective paths from Dodge classification and some examples of collective phenomena and detected some gaps in their approach; they have also concluded that the classification system could be formalized in First-Order Logic.

Finally, an important contribution for the representation of the dynamics of pedestrian movement has been provided by research on simulation and AI. This type of research seeks to create models that reproduce pedestrian behaviour and use them as the basis for the design and planning of transportation infrastructures, evacuation facilities and leisure areas. Various approaches have been suggested for these models; among them, the most noteworthy ones are based on Agent Based Modelling (Penn and Turner, 2001; Batty, 2003; Batty, De Syllas, and Duxbury, 2003; Bierlaire, Antonini, and Weber, 2007; Antonini, Bierlaire, and Weber, 2006), and Cellular Automata (Blue and Adler, 2001; Burstedde et al., 2001).

5.3 A representation of pedestrian movement based on interactions

Let us consider a recreation area with a large number of visitors freely walking around. If the movement of each person is represented using a trajectory, it will be possible to detect certain movement patterns such as the most widely used routes, the places where several people meet or the most frequently visited areas. However, it is worth noting that the higher the number of visitors, the less adequate those trajectories will be for movement representation. Indeed, consequently the whole space will be covered by them, thus making patterns more difficult to understand and the causes of these patterns more variable. Moreover, these patterns do not explain *per se* the behaviour of people, because the interpretation can vary. For example, if people meet by stopping in front of some attractive element or if, on the contrary, they all stop at some uninteresting area. As a result, the movement pattern may have been characterized by an interaction with the environment (e.g., to stop and look at some interesting element) or by an interaction between the pedestrians themselves (e.g., to stop for a chat). Therefore, it is obvious that these movement patterns can be interpreted by looking at the possible interactions that caused them and making them explicit through the representation of the context in which movement takes place. Pedestrian behaviour can be ultimately inferred from the analysis of interactions.

For example, the "socializing" behaviour in a pedestrian can be inferred from the interaction type they show when encountering other people in certain places.

Our approach suggests a new spatial knowledge representation for pedestrian movement through the exploitation of ontologies for the formalization of movement interactions and related concepts: patterns, context and behaviour. This interactions-based representation enables us to semantically represent movement patterns computed by different tools and to incorporate the knowledge of domain experts on pedestrian behaviour. Interactions operate therefore as a conceptual bridge between movement patterns and pedestrian behaviour.

As we mentioned in the introduction, the analysis approach in current research is based on the assumption of a univocal relationship between movement patterns and behaviour, or, as some authors have stated, *Movement is behaviour* (Blythe et al., 1996, p. 13; Dodge et al., 2008, p. 245). However, given that interpretation relies on context, we require a conceptual bridge in order to explicitly integrate that context. Our view is that *Movement is behaviour, but patterns are not*. Patterns are the evidence of the interactions that take place during pedestrian movement. For example, the pattern shown in Figure 5–1 can be interpreted as a group of individuals flocking at time t_3 and moving together (Gudmundsson et al., 2004).

Figure 5–2 shows different behaviours that can be inferred from the same movement pattern depending on the information we have about the context. Consider the first pattern on the left; as we saw, this could be interpreted as a flocking behaviour in which the individuals meet and move together intentionally. The second picture shows the context from which we can interpret the pattern as people following fixed pathways. Analogously, the third pattern discloses a typical type of behaviour of people seeking a goal (e.g., going to take the bus). Finally, the last pattern can be interpreted as a leadership behaviour in which the central individual guides the other ones. This contextual information allows us to understand the primary elements that influence the movement and produce the pattern. Therefore, the existence of a movement pattern is of course a necessary, but not sufficient condition for the interpretation of pedestrian behaviour.

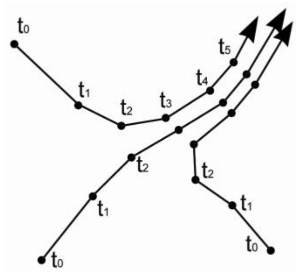

Figure 5–1. Flocking pattern.

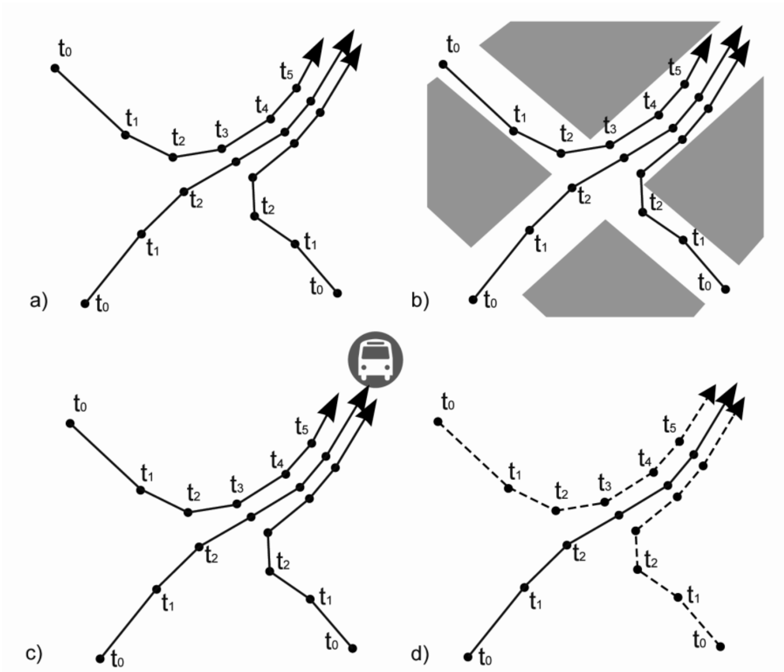

Figure 5–2. Different behaviours interpreted from the same movement pattern.

Our approach introduces the notion of interactions as the primary element of analysis and movement representation. Movement interactions can be defined as the result of active and reciprocal relationships between pedestrians themselves, and between pedestrians and their environment, interpreted in a given context. This definition is of considerable usefulness, as it allows us to integrate concepts from previous research and expands the possibilities of analysis and representation. Therefore, interactions act as a conceptual bridge between patterns and behaviour.

Thus, the traditional approach of direct mapping from patterns to behaviour is enriched with a formal context representation employed for the reasoning and interpretation of observed patterns (Figure 5–3). The behaviour can be ultimately disclosed by the analysis of a set of interactions and the relationships between them. In our opinion, this representation allows more formal differentiation between the various levels of movement interpretation, which in addition provides a higher expressivity and more independence from the application domain.

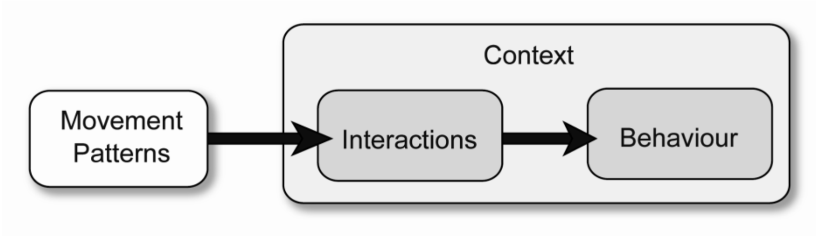

Figure 5–3. Interactions based representation.

5.3.1 Elements of representation

Next we will study the elements that form our proposal and the way they are represented.

Patterns

The most elementary unit of representation in our approach is the pedestrian movement pattern. As we have previously mentioned, a pattern is the evidence of the interaction between pedestrians and their environment. Patterns would be conceptualized as the building blocks of movement, each one with specific and observable characteristics in the spatial, temporal or spatio-temporal dimensions. These observable characteristics are used to detect the patterns in movement data sets. Besides the diversity of patterns, there is an increasingly wide range of tools and methods to detect these patterns in movement datasets. In our approach, we deal with this variability by establishing a set of properties for each pattern that defines its characteristics and computing methods. Once computed (using a specific tool or method), the pattern can be formalized with its properties and values as an instance of the ontology.

Context

The context is formed by a partial set of parameters and values that is never complete, precise nor objective, but it is useful for movement representation and for reasoning on it. Employing the metaphor suggested by Giunchiglia and Bouquet (1997), this can be illustrated as a box, inside which we find the expressions that establish or explain the domain or the phenomena to be studied, whereas outside the box we find a set of parameters and their values, which determine, at least partially, the interpretation of the expressions inside the box. Although it is obvious that any type of representation (and reasoning) will depend on the context, we have seen that most approaches either assume that the context is implicit, or rather exclude it. Thus, the interpretation of patterns is carried out only on the movement's own features.

By contrast, our approach seeks to explicitly and formally represent a context in order to sustain the reasoning and the interpretation of movement patterns, and to expand the possibilities of such interpretation. Moreover, representation of the context is enriched with the results of this interpretation, and can be used again for subsequent interpretations in an iterative process. This approach paves the way for a more dynamic form of representation and contextual reasoning about pedestrian interactions.

Considering again the example of a recreation area with a considerable amount of visitors: collective movement can show patterns of high pedestrian concentration in certain places, which will be interpreted as attraction areas. The location and extension of these attraction areas are used to enrich the representation of the context, allowing us to establish new interpretations of movement patterns: for example, visitors who stop by in these attraction areas can be interpreted as a visitor type dissimilar from those who do not.

Although context is usually linked to a geographic environment, there may be other parameters that can describe it, such as the purposes and intentions of visitors, temporal context, the degree of constriction in a given space, and relationships with other patterns. Interactions can become a part of the context themselves, thus interpreting new interactions.

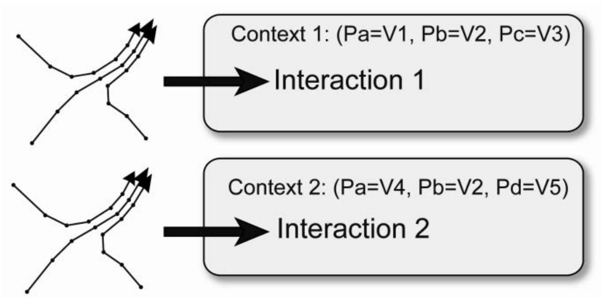

Figure 5–4. Contexts for movement representation. Each context is described by a set of parameters [P] and values [V].

Behaviour

Hoogendoorn, Bovy and Daamen (2002) suggested that pedestrian behaviour is determined by a hierarchic structure of decision-making at three levels. (i) A strategic level in which the pedestrian decides about his/her destination, the activities to be carried out or his/her aims (ii) a tactical level in which the pedestrian decides about the route to follow, the response to unexpected events or the avoidance of "unwanted" areas or spots; and finally, (iii) an operational level in which the pedestrian decides in which direction the next step will be taken, which means that they intuitively choose a direction and speed, depending on the immediate environment.

We must point out that each level carries an implicit conceptualization of the corresponding space: at the strategic level, space is conceptualized in terms of aims, needs and desires; at the tactical level, space is conceptualized in terms of assigned symbolic values of utility, opportunity, benefit, etc.; whereas at the operational level, conceptualization is only observational and based on perceived features. In our approach, operational and tactical behaviours can be represented by the simplest interactions, while representation of the strategic behaviour requires more complex sets of interactions. Consider for example a visitor who takes a detour (operational level) in order to avoid an overcrowded area (tactical level), and then discovers a new interesting spot that leads him/her to reprogram his/her activities (strategic level).

Analysts or domain experts can define diverse behaviour categories depending on the movement interactions expected in different levels. These definitions can be represented in an ontological formalization, with reasoning tasks being assigned in order to find examples of pedestrians performing every type of behaviour from a set of movement patterns.

Interactions

Although the notion of movement interactions is not new, in our approach we develop the idea of interaction as the key element for the representation of pedestrian movement. This idea is based on the assumption that there are different elements that play an active role, and which influence and are influenced by movement. These elements can be conceptualized as “agents” because of the active effect they have on movement interactions, and they would be “individual”, “collective” and “environmental” agents.

Movement interactions create specific dynamics of action and reaction between the agents in any given context. Consequently, pedestrian interactions must be represented bearing in mind the kind of agent participating, the context in which they occur, and the specific movement patterns that demonstrate them. The wide variety of possible interactions requires for a type of classification that allows us to organize them in an ontological representation. In our approach, we suggest a natural type of classification based on the kinds of agents involved.

- *Pair-wise Interactions*: They occur when the movement is determined by interactions between individuals. Examples include encounters, guiding-following and joint walking. Although pair-wise interactions would involve more than two individuals, the conceptualization remains as a relationship between pairs of pedestrians: the elements that determine the movement are individuals.

- *Environmental interactions*: They occur as a consequence of the relationship between pedestrians and features of the environment. These features have physical, social or psychological functions that have an influence on the movement attracting, repelling or restricting pedestrians when they are walking. Examples of environmental interactions include attraction, route following or trail formation. These interactions imply the conceptualization of environment as an active agent that influences and is influenced by pedestrian movement.

- *Collective Interactions*: They appear as a consequence of interactions between several individuals. Collective interactions differ from pair-wise interactions not (only) in number but in the appearance of observable systemic properties of the collective as a whole. These properties usually include a progressive synchronization of one or more movement parameters, resembling a moving herd. Although collectives would normally be defined by their member's characteristics, in our approach they are defined by the global properties of their movement, that is, collectives are conceptualized as the cause and consequence of movement. Examples of collective interactions include lane formation, incremental concentration and flocking, or circulating flows at intersections (Helbing et al., 2001).

This simple but intuitive classification offers several advantages: firstly, it focuses on the active role played by the environment, which influences and is influenced by pedestrian movement; secondly, it allows us to formally represent the systemic properties of collective movement reported by various authors (Helbing et al., 2005; Wood and Galton, 2008); thirdly, it allows for the representation of more than one type of interaction operating at the same time, and finally, there is the practical advantage in defining interaction types according to combinations of patterns and contexts for a wide variety of applications.

Moreover, this conceptualization represents pedestrian movement as an expressive collection of relationships between the different agents involved, relationships that are demonstrated by movement patterns in defined contexts. Finally, it is worth noting that movement actually involves more than one type of interaction at the same time. For example, an encounter at a point of interest would be represented as the combination of pair-wise and environmental interactions.

5.3.2 Examples of interactions

Since the potential number of possible interactions depends on the context, and therefore on the application, next we will describe some of the most common interactions that can be defined from a relatively small number of patterns, which in turn can be easily extracted with the available tools and methods. Therefore, it is not our intention to describe all the possible interactions, but to use some significant examples to how the interactions are used to represent pedestrian movement.

For the representation of the context we have used the "constriction level" as an example of a contextualizing parameter. The possible values to be assigned to that parameter could be: *constrained space*, *semi-constrained space*, and *open space*. Other contextual parameters are employed in order to evaluate the concepts of spatial and temporal proximity and threshold. Here, a number of spatial and temporal values are represented as threshold to indicate a short, medium or long value. These values make it possible to formalize the concepts related to distance and time. For example, when two people stop close to each other for a long period, the concepts “close to” and “long” are formalized as semantic concepts more than quantitative measurements of distance and time. An additional parameter is used to evaluate the concept of attractiveness of geographical features. The level of attractiveness indicates how much specific points or elements of environment tend to attract pedestrians. This attractiveness is represented as low, medium or high levels and an additional “neglected” level to indicate a negative attraction (i.e., repulsion).

Encounter

Main interaction type: Pair-wise interaction.
Associated movement pattern: Spatio-temporal coincidence.
Contextual parameters: Constriction level, spatio-temporal threshold.
Definition: The coincidence in space and time of two or more individuals who remain together for a long time-span. During the encounter, pedestrians may stop moving or may keep on walking together. Spatial context is used in order to evaluate the influence of spatial constriction or attraction elements as well as the spatio-temporal threshold conceptualization: If the duration of the pattern is long, then the restriction level does not have any influence. However, if the coincidence happens in an open space, a medium duration and spatial threshold are enough, since there is no influence from the environment that encourages the individuals to approach each other. If the interaction takes place in a location with high attractiveness, it cannot be inferred as an encounter since the pattern would be due the environment.

Approaching

Main interaction type: Pair-wise interaction.
Associated movement pattern: Spatio-temporal coincidence with relaxed temporal threshold.
Contextual parameters: Constriction level, spatio-temporal threshold.
Definition: The space-time coincidence of two or more individuals who do not remain together. Spatial context is used in order to evaluate the influence of spatial constriction or attraction elements. Contextual parameters also define the relaxation value of temporal proximity in order to differentiate crossings from encounters.

Guiding - following

Main interaction type: Pair-wise interaction.
Associated movement pattern: Spatial coincidence with temporal delay.
Contextual parameters: Constriction level, spatio-temporal threshold.

Definition: A similar sequence of positions with a certain delay between one individual and the other. Spatial context is used in order to evaluate the influence of spatial constriction or attraction elements, the spatio-temporal threshold conceptualization and the temporal delay. This delay must be shorter and the duration longer in constrained spaces than in open spaces to be inferred as a guiding-following interaction.

Visiting

Main interaction type: Environmental interaction.
Associated movement pattern: Stopping.
Contextual parameters: Attractiveness level, spatial proximity.
Definition: A medium or long stop of an individual or a group of individuals in a geographic location with medium or high attractiveness level.

Route Choosing

Main interaction type: Environmental interaction.
Associated movement pattern: Short stops.
Contextual parameters: Constriction level, temporal duration.
Definition: A sequence of motion-stop-motion located in semi-constrained points in which pedestrians must choose a route from a limited number of options. Contextual parameters define the constriction level and the duration of stops.

Attraction

Main interaction type: Environmental interaction.
Associated movement pattern: High density / Movement suspension.
Contextual parameters: Attractiveness level.
Definition: The suspension of movement in locations containing interesting features. Contextual parameters define the attractiveness level.

Flocking

Main interaction type: Collective interaction.
Associated movement pattern: Collective coordination of relative motion parameters.
Contextual parameters: Constriction level, temporal threshold.
Definition: The formation of groups that move together in a similar way for a given time-span in an open space. Contextual parameters define the constriction level and the duration of the interaction.

Aggregation

Main interaction type: Collective interaction.
Associated movement pattern: Concentration (High-density pattern).
Contextual parameters: Constriction level.

Definition: The grouping effect produced by the concentration of pedestrians in defined locations, causing an increase in the density in a given time-span. Contextual parameters determine the constriction and attractiveness level. This interaction is equivalent to the environmental interaction "Attraction" but here, the attractive element is the collective interaction itself.

Trail formation

Main interaction type: Environmental interaction.
Associated movement pattern: Linear movement clusters.
Contextual parameters: Constriction level.
Definition: The formation of linear clusters in open spaces. The repeated passing of pedestrians leaves trails that, depending on the features of the area, may modify the space forming a trail that consequently acts as a linear attractor. Contextual parameters define the spatial constriction level and the time-span that has been calculated for the trail formation.

In the following section, we introduce the formalization of the concepts mentioned above in a taxonomy of interactions, namely the Interactions ontology. We give the conceptual view of the ontology in the next section, followed by a simple case study to disclose pedestrian behaviour in a park.

5.4 The Interactions Ontology

Following the Davis roles for a knowledge representation as mentioned in the introduction (Davis et al., 1993), this section introduces a formal representation of the classes defined above by means of a *formal ontology*. The definition given by Gruber (2008) is used to define formal ontology as "a technical term denoting an artefact that is *designed* for a purpose, which is to enable the modelling of knowledge about *some* domain, real or imagined." Such ontologies determine what can be represented and what can be inferred about a given domain, using a specific formalism of concepts. An ontology language is a formalism used to express such knowledge.

Web Ontology Language (OWL) is a well-known standard that came from the Semantic Web and it is now a W3C recommendation (W3C Consortium, 2004). OWL is based on a family of languages known as Description Logics (DL) that provide a deductive inference system based on formal well founded semantics (Baader et al., 2003). The basic components of DL are *concepts (classes)*, *properties (roles)*, and *instances (individuals)*. Concepts describe the common properties of a collection of instances and properties are binary relations between concepts. The special relation *is_a* represents the *specialization* property between two concepts and describes a taxonomy in the ontology, based on subsumption relation. Furthermore, a number of language constructs, such as intersection, union and role quantification, can be used to define new concepts by means of *axioms*. In other words, concepts may be intentionally defined by axioms, which express the properties that characterize such defined class. The Description Logic formalism comes with a number of primitive reasoning tasks, such as *classification* and *satisfiability*, *subsumption* and *instance checking*. Classification is the computation of a concept hierarchy based on subsumption, satisfiability checks if the ontology defined concepts are consistent, whereas instance checking verifies that an individual is an instance of a concept. In this chapter, we use OWL DL, the OWL sub-language that allows for the maximum expressiveness while retaining computational completeness, corresponding to Description Logics. Other OWL languages are OWL Lite and OWL Full. OWL Lite is a subset of OWL DL which restricts the logical operators allowed, thus resulting in an efficient inference checking system. On the contrary, OWL Full allows for the maximum expressiveness, however, it loses completeness.

The developed interactions ontology defines the four main concepts depicted in Figure 5–3, namely Movements Patterns, Interactions, Context and Behaviour. Figure 5–5 shows the top level classes and properties of the interactions ontology. The Interaction concept may be associated to a (pre-computed) movement pattern and may be located in a specific context (spatial and/or temporal). Furthermore, each Pedestrian may participate in an Interaction.

In the following we specify the taxonomy levels of the main concepts: Interaction, Movement Pattern and Context. The three kinds of interactions are specified by the following taxonomy (Figure 5–6).

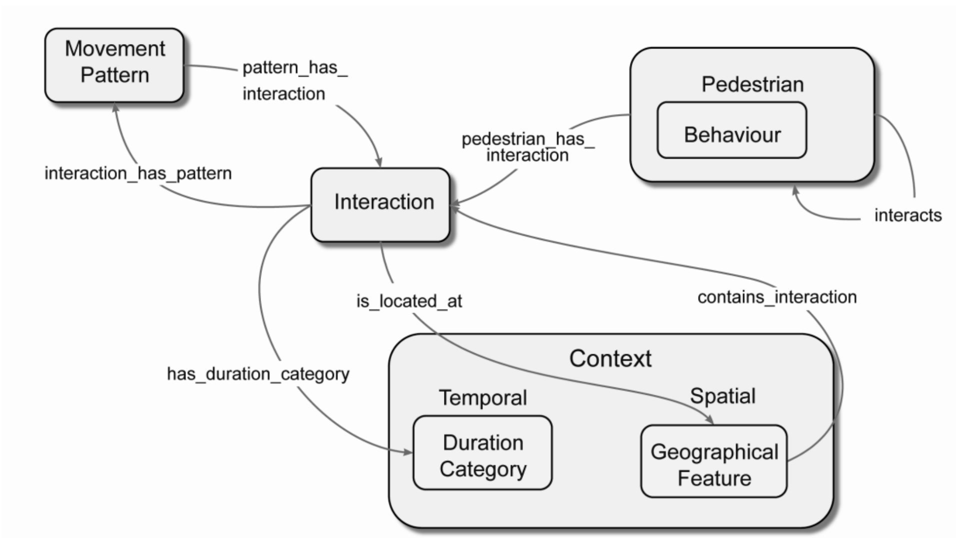

Figure 5–5. Main classes for the Interactions ontology.

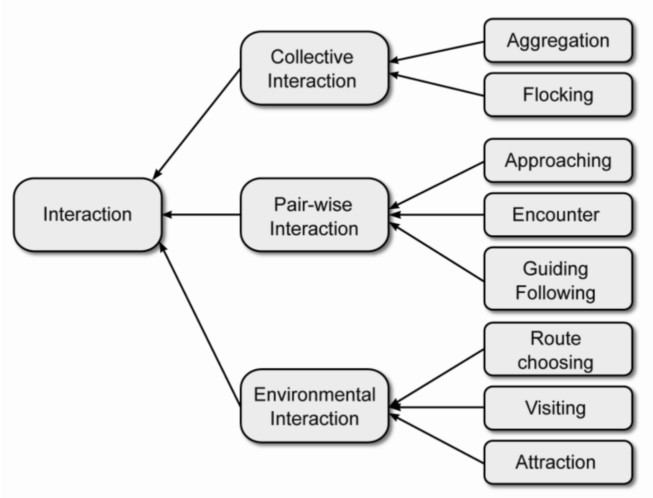

Figure 5–6. The interactions taxonomy.

Interaction concepts are defined in terms of movement patterns and context parameters. In other words, each interaction concept is a *defined class*, where a new concept is defined in terms of a logical combination of other concepts and/or properties. Consequently, the reasoning engine can automatically infer pedestrian movement interactions based on pre-computed movement patterns. As an example, consider the *Encounter* interaction. Encounter can be specified by the following OWL DL axiom:

```
Encounter ≡
     is_located_at some (Park_Feature and ((has_Attractiveness has
    Low_Attractiveness) or (has_Attractiveness has
    Medium_Attractiveness)))
 and
     (has_Duration_Class has DC_Long and inter_has_pattern some
    (SpatioTemporal_Coincidence and(has_Spatial_Threshold has
    Spatial_Threshold_Small) and (has_Temporal_Threshold has
    Temporal_Threshold_Short))
 or
    ((has_Duration_Class has DC_Medium) or (has_Duration_Class has
    DC_Long)and is_located_at some (Park_Feature and
    has_Constriction_Level has CL_Open)) and (has_Duration_Class has
    DC_Medium) or (has_Duration_Class has DC_Long)))
```

Intuitively, this axiom defines the Encounter class as a kind of Interaction that has been demonstrated by a spatial-temporal coincidence pattern if the duration is medium to long and it is not due to environmental features. The conceptualization of the necessary duration depends on the constriction level of the space. In contrast, the Approaching interaction is characterized by a spatio-temporal coincidence with short or medium temporal threshold and a small spatial threshold. Furthermore it is located within a constrained or semi-constrained space (e.g., a pathway or a square).

```
Approaching  ≡
      has_Duration_Class has Short
  And
     inter_has_pattern some (SpatioTemporal_Coincidence and ((or
     (has_Temporal_Threshold has Temporal_Threshold_Medium)) and
     (has_Spatial_Threshold has Spatial_Threshold_Small)) and is_located_at
     some (Geographical_Feature and ((has_Constriction_Level has
     Constrained) or (has_Constriction_Level has Semiconstrained)))
```

As we can notice in such definitions, each interaction is inferred from movement patterns, characterizing them in relation to the context features. In this sense, patterns capture the spatio-temporal essence of the interaction. We have identified a set of possible spatial and spatio-temporal patterns based on their role on interaction inference. An example of a movement pattern (here used on the encounter and approaching interaction) is the Spatio-Temporal Coincidence. This pattern describes a pair of trajectories that have some coincidence of time-space (they are in the same place at the same time). The thresholds that define the spatio-temporal coincidence have to be set by the analyst depending on the other contextual parameters (how close is “the same place”? how many meters? how long is a coincidence? Some seconds or some minutes?) The movement pattern taxonomy is depicted in Figure 5–7.

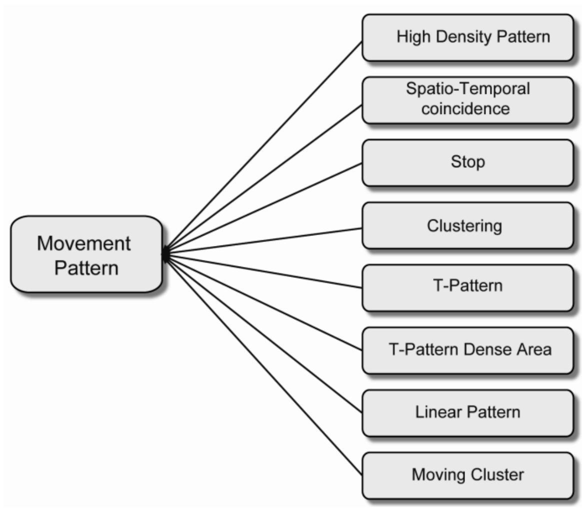

Figure 5–7. Movement patterns taxonomy.

Movement patterns can be computed using several techniques, ranging from spatial analysis tools (Orellana et al., 2009) and exploratory visual analytics (Thomas and Cook, 2005), to spatial and spatio-temporal data mining (Nanni et al., 2008). An example of pattern detection technique is REMO (RElative MOtion) that compares relative patterns among individuals in a pattern-matching process, focusing on identifying the similarity of one or more movement parameters that are depicted in a movement matrix (Laube, Kreveld and Imfeld, 2005; Weaver, 2008). Visual Analytics discovers movement patterns typically found by visual techniques, such as spatio-temporal coincidence (N. Andrienko et al., 2008), stops and moves (N. Andrienko and G. Andrienko, 2008). Data-Mining techniques extract common movements by analysing with a statistical technique a large number of trajectories. Examples of trajectory patterns are T-cluster (Rinzivillo et al., 2008), identified groups of similar trajectories give a similarity measure, and T-patterns (Giannotti et al., 2007) representing the frequently followed sequences of places. Other movement patterns can be discovered by spatial-statistical approaches thus finding linear patterns and dense areas for movement suspension (Orellana et al., 2009).

The other main concept of the interactions ontology is Context. In this work we have defined two different kinds of context: the temporal context, where we defined the time units of interest for the application, and the spatial context, defining the most interesting characteristics that describe the geographical locations where the movement takes place. Therefore, this representation strategy follows the approach known as “Compose-and-Conquer” (Bouquet et al., 2003) since it does not use global theories for context, but local theories that represent specific points of view (Figure 5–8).

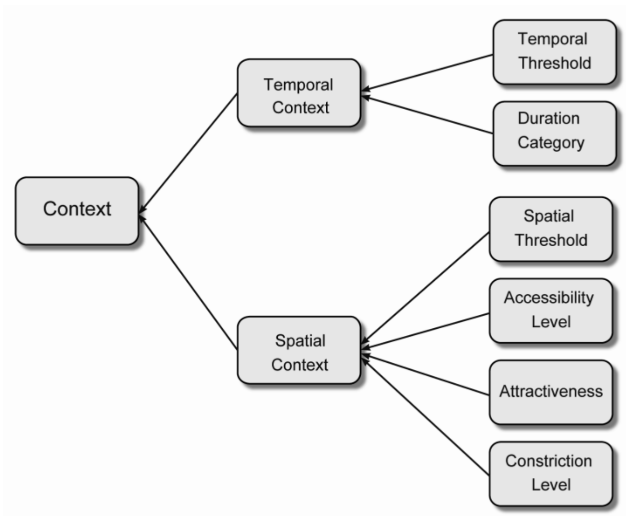

Figure 5–8. Context class and subclasses.

It is worth noting that the spatial context deals with geographical features of the environment, such as the constriction level (pedestrian moving in open space, constrained space such as a pathway, or a semi-constrained space like a square), the attraction areas, the accessibility levels, etc. Temporal context relates to the specification of duration and thresholds of interactions. Figure 5–9 shows the complete taxonomy of interactions ontology implemented in *Protégé Ontology Editor* v 3.4 (Stanford Centre for Biomedical Informatics Research, 2008).

5.5 Application Scenario

The viability of our approach to analysing collective pedestrian movement is experimented through an application scenario about visitor's behaviour in the Dwingelderveld National Park (DNP) in Netherlands. DNP is a natural park of about 3,700 ha in the North Eastern part of the Netherlands and receives 2million visitors a year. It is a typical Dutch recreational area with an extensive network of short strolls (60 km of marked trails, each of less than 7 km in length) and long walks for cycling and horse riding. To have a better understanding of how the park is used, park managers assign a specific behaviour to each visitor, depending on their movement behaviour. For example, a visitor who follows only marked trails is named as "follower" whereas a visitor that explicitly does not follow the marked trail is called a "browser" or "explorer" (van Marwijk and Pitt, 2008).

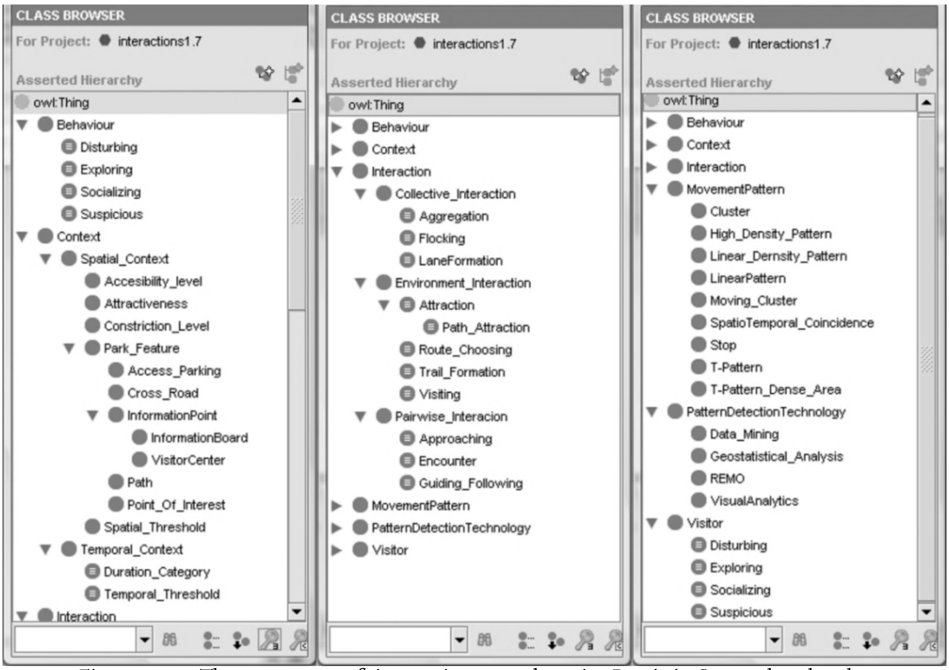

Figure 5–9. The taxonomy of interactions ontology in Protégé. Screenshot by the author.

The available dataset comes from three different information sources. A questionnaire that records visitor characteristics; a set of point coordinates captured by GPS receivers given to the visitors; and a GIS data-set containing the path network of DNP and the locations of access point. The questionnaire gives detailed information about the visitor, such as sex and age, and the reason for the visit (to have a walk, to take pictures, etc.) The park has five main entrances, from which the visit starts (the beginning of the GPS trajectories). This survey was carried out during 7 days (weekend and weekdays) in spring and summer 2006 for 461 hikers (van Marwijk and Pitt, 2008).

The objective of the current analysis is to characterize visitor behaviour in terms of the specific interactions that can be inferred from their movement patterns. In particular, we exploited the previously defined Interactions ontology, specialized in this particular case study, to represent and infer new knowledge about visitor behaviour.

The overall vision of the proposed approach is depicted in Figure 5–10. The problem we want to deal with can be stated as: *find visitors of the park whose behaviour can be classified as "explorer" (or follower, or other)*. This classification task is actually performed by the ontology-reasoning engine, where movement data and patterns are collected and an implicit class defined by OWL axioms specifies each kind of behaviour.

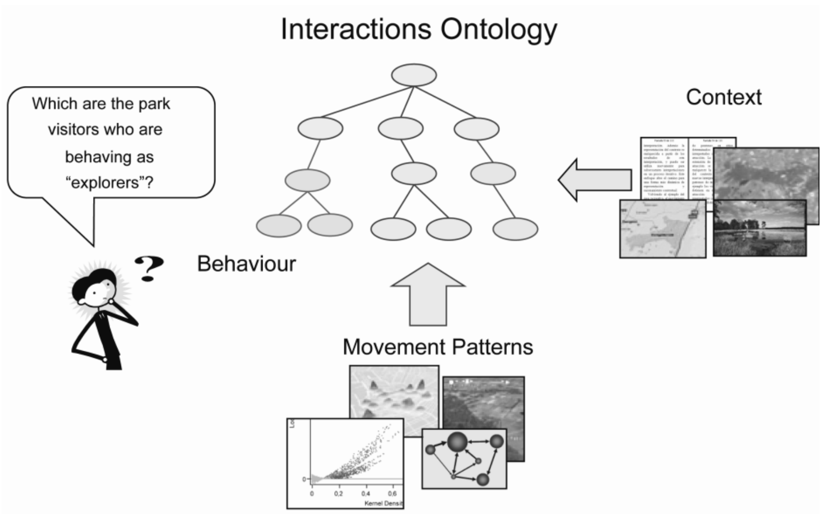

Figure 5–10. The proposed overall approach.

In this experiment, we used Protégé 3.4 (Stanford Center for Biomedical Informatics Research, 2008) with Pellet Reasoner (Sirin et al., 2007). As a very first step, the instances were introduced manually in Protégé. However, the feasibility of automatic systems was envisaged, in fact some previous work used ontologies for reasoning with trajectory data exploiting Oracle11g as an efficient reasoning engine over large datasets (Baglioni et al., 2008; 2009).

As an example, querying the ontology of the "explorer" behaviour expressed by visitors will give the identifier (ID) of the visitors whose movement has been classified by the ontology-reasoning engine (based on the previously defined axioms), as "explorer". In other words, the instances of the ontology class Visitor represent the individual visitor's movement. These visitors are classified by the reasoner into the appropriate classes based on the satisfiability of the axioms.

It is worth noting that pattern finding can be seen as a top-down and bottom-up process. Indeed, extracted patterns can be used in finding interactions, but also interaction definition can drive the analyst to find some new movement patterns. For example, the attraction areas can be some predefined points or regions (the campsite, the café, the radio-telescope, etc.) or can be detected by data mining patterns as the most frequently visited areas, or as the T-Patterns computed dense areas.

To acquire a better understanding of the approach, let us consider the exploring behaviour. As specified above, an explorer is a visitor who tends not to follow the marked trails. Therefore, which kind of interactions can be expected by "explorers"? For example, it is expected that an explorer will tend to stop at the cross-roads to decide which route to take and to consult information points such as the visitor centre or information boards (Figure 5–*11*). This definition can be rephrased as "visitor who has interactions of the kind *route-choosing* and *visiting* information points". This means that an *exploring behaviour* is a kind of visitor who can be disclosed by the following axiom:

```
Exploring ≡
     visitor_has_interaction some Route_Choosing visitor_has_interaction
     some (Visiting and (is_located_at some (Information_Point))
```

Analogously, a socializing behaviour can be identified by having encounters in places where people usually do not interact, such as the park trails and cross-roads.

```
Socializing  ≡
     visitor_has_interaction some (Encounter and (is_located_at some
     (Cross_Road or Path)))
```

Another example is the axiom below that defines a disturbing behaviour in terms of visiting forbidden park areas.

```
Disturbing  ≡
     visitor_has_interaction some (Visiting and (is_located_at some
     Park_Feature and (has_Accesibility_level has Forbidden)))
```

Another interesting feature of the ontology is that we can link visitors involved in a pair-wise interaction through a symmetric property called interacts_with. This makes it possible to define a "suspicious" behaviour as a visitor who interacts with someone who is inferred to be a "disturbing" visitor.

```
Suspicious  ≡
     Interacts_with some (Disturbing)
```

These examples can be redefined according to different domain experts, but interactions remain the building blocks used to represent visitor behaviour.

Figure 5–11. Part of the spatio-temporal path of a visitor on 06/08/2006. The time is represented by the height of the path; yellow bars represent route choosing interactions and red bars represent visiting interactions; green pins represent points of interest and (i) stands for an information point. The reasoning engine will classify the visitor as "explorer". Screenshot of Google Earth © by the author reproduced under the terms of "Fair Use".

The process outlined in this application scenario can be summarized by the following main steps:

- Visitor information is stored as ontology as instances of the Visitor's class.
- The basic spatio-temporal patterns are detected by means of a number of tools: GIS software such as ArcGIS (ESRI, 2008), or visual analytics tools such as Visual Analytics Toolkit (Andrienko, G. and Andrienko, N., 2008; IAIS, 2008), and data mining algorithms (Rinzivillo et al., 2008; Giannotti et al., 2007). In particular, here VA Toolkit has been used to find patterns for spatio-temporal coincidence and stops.
- Detected patterns are imported as instances into the ontology. This means that the main features of the patterns are represented as instances of the ontology along with the context information (spatial and temporal threshold, constrained levels in space and so on).
- Once patterns and contextual information are stored in the ontology as instances, the reasoning engine performs an instance-checking task to infer both interactions and movement behaviour.

In the specific case illustrated here, we have focused on *socializing* behaviour. We first detected the spatio-temporal coincidence patterns by means of the VA Toolkit (see, for example, Figure 5–11). We set the spatial threshold values as 5 meters for small, 10 meters for medium and 20 meters for long. The constrained space is represented by the set of paths, seen easily on the park map, whereas the cross roads and small areas represent the semi-constrained space; the remaining area is considered as open space. We populated the ontology with data referred to a part of the day 06/08/2006 and we ran the reasoning engine. We inferred four instances of visitors to be 'socializing' since they interacted with other visitors (Encounters) when they walked on the paths in the park. One visitor was interpreted as "Explorer" since they stopped and chose a route at several crossroads and stopped in two points that we marked as "Information Points". One visitor was interpreted as "Disturbing" since they were visiting an area that we marked as forbidden. In addition, we inferred one visitor to be "Suspicious" due to the fact that he or she encountered the Disturbing visitor. Figure 5–12 shows the relevant classes and instances that the reasoner engine used to perform the instance checking for some instances of Interaction and Visitor Classes. The reasoner engine inferred that I3 *is an instance* of Encounter and I6 *is an instance* of Visiting. Furthermore, visitor V9 was inferred as Disturbing since he or she visited a forbidden area and visitor V11 was inferred as Suspicious since he or she interacted with the disturbing V9.

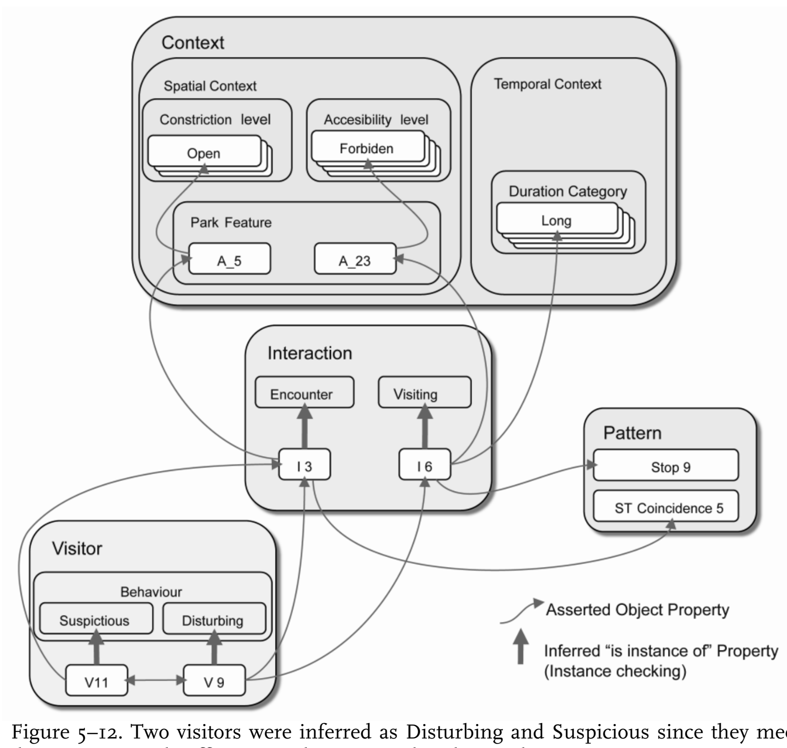

Figure 5–12. Two visitors were inferred as Disturbing and Suspicious since they meet the necessary and sufficient conditions stated in the ontology.

5.6 Conclusions

In this chapter we proposed a knowledge representation and reasoning approach to characterize pedestrian movement interactions and behaviour. The contribution of this work, in respect to previous literature proposals, is twofold. On the one hand, we have developed a taxonomy of interactions where a number of basic interactions, concepts and their relationships with movement patterns and pedestrian behaviour are defined within a context. On the other hand, we have enriched the introduced taxonomy with reasoning, where, exploiting the ontology formalism OWL, interactions are combined with context to define pedestrian movement behaviour. Furthermore, we have sketched a case study where we experiment the methodology with a real dataset recording visitors' movements in the Dwingelderveld National Park (DNP) in the Netherlands.

A crucial point in the implementation of this experiment is how to populate the ontology with data, such as movement patterns and visitor information. How to link each tool output (namely, an extracted pattern) to a specific ontology class? Common approaches to ontology population map database tables to ontology concepts. This means that at the first stage we have to store both data and extracted movement patterns in a database. The problem then becomes how to map each table to an ontology concept. Several approaches have been proposed in the literature following two main directions: to import each database record into the ontology or dynamically map the ontology to a database. These two approaches offer complementary benefits and drawbacks. Indeed, importing data into the ontology may become impractical when the dataset is very large, since current ontology systems are not scalable to large datasets. On the other hand, dynamic mapping from ontology to database cuts down the degree of expressiveness of the ontology language (Calvanese et al., 2007).

As a future task, we plan to design and implement an architecture where the patterns detected from the various tools are automatically or semi-automatically, inserted into the ontology. For this step we would benefit from approaches like Athena (Baglioni et al., 2009), a system built on top of Oracle 11g that allows for easily import of database tables storing trajectories and contextual information inside ontology concepts.

Although we consider that our approach can potentially achieve a high level of expressivity to represent pedestrian behaviour, we are aware of the shortcomings that it has at this stage. We are planning to further develop the concepts related to pedestrian behaviour applied to new and more interesting scenarios (e.g., including actual collective movement patterns and behaviour).

6

Synthesis

A road is a tribute to space. Every stretch of road has meaning in itself and invites us to stop. A highway is the triumphant devaluation of space, which thanks to it has been reduced to a mere obstacle to human movement and a waste of time. Before roads and paths disappeared from the landscape, they had disappeared from the human soul: man stopped wanting to walk, to walk on his own feet and to enjoy it. What's more, he no longer saw his own life as a road, but as a highway: a line that led from one point to another. (...) Time became a mere obstacle to life, an obstacle that had to be overcome by ever greater speed. Road and highway; these are also two different conceptions of beauty. In the world of highways, a beautiful landscape means: an island of beauty connected by a long line with other islands of beauty. In the world of roads and paths, beauty is continuous and constantly changing; it tells us at every step: "Stop!" (Kundera, 1991 p223).

6.1 Introduction

The main goal of this thesis is to develop an approach to detecting, analysing and interpreting movement patterns of pedestrians interacting with the environment. In the previous chapters, individual components of the approach were explained in detail. Here, the research objectives answering the research questions proposed in Chapter 1 are revisited and connected to provide a complete overview of the proposed approach.

The research work relied on the assumption that large observational datasets of the movement of many pedestrians contain information that can help us to understand their spatial behaviour. This information may not be obvious at first glance and several iterative steps may be required to extract it.

Figure 6-1illustrates the proposed approach to analysing pedestrian movement, which can be summarized in the following steps. First, GPS technology is used to track the movement of pedestrians and the data are stored in files containing sets of space-time coordinates. This raw data is preprocessed and the results are stored in a geodatabase containing movement datasets. These datasets are transformed to create different representations, such as movement vectors and trajectories. Spatial analysis methods and data-mining techniques are then used to detect movement patterns (i.e., Suspension Patterns and Generalized Sequential Patterns). These patterns are interpreted using contextual information about the environment and about the pedestrians to infer movement interactions that explain the movement behaviour giving rise to those patterns. Finally, the relations between patterns and movement behaviour are organized into an ontology.

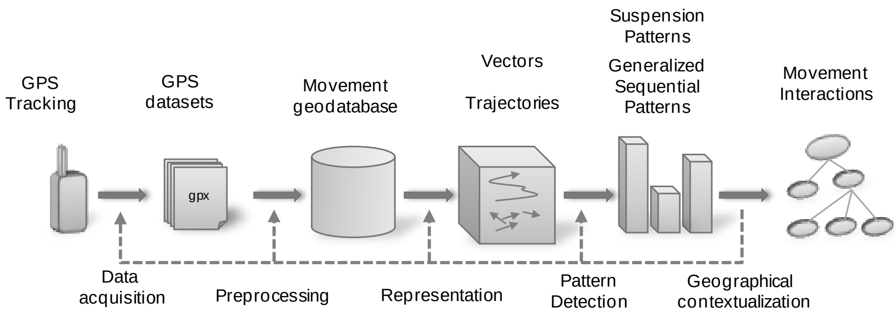

Figure 6-1. Overview of the approach to pedestrian movement analysis

Although these steps are presented as a linear process, it is important to point out that iterations are part of the overall approach (represented as dashed lines in Figure 6-1). For example, a movement pattern detected using spatial statistics has to be transformed into a new representation in order to apply further analysis. Therefore, instead of a straightforward workflow, the proposed approach implies an iterative process of exploration and experimentation in which the results of each step leads to a new loop of exploration and discovery to ultimately gain new knowledge about the movement behaviour of pedestrians. My approach is therefore an extension of the "knowledge discovery process" (Fayyad, 1996), which I adapted for the purpose of analysing pedestrian movement. Moreover, my approach allows the use of different representations of movement and the creation of synergy between spatial analysis and data-mining techniques.

The main motivation for developing the proposed approach was to gain knowledge about the spatial behaviour of pedestrians by analysing movement data. Some researchers have argued that "movement is behaviour" (Blythe, Miller and Todd, 1996; Dodge, Weibel, and Lautenschütz, 2008). However, an extensive review of the literature on movement pattern analysis revealed that in most of the studies the relation between movement patterns and behaviour tends to be implicit. In other words, it is not always clear how movement patterns should be interpreted (Galton 2005; Laube, 2009). The core idea of the proposed approach is that movement patterns are the result of individuals with similar goals showing some kind of collective response (interaction) to the geographical space. Therefore, in this research I make the relationship between movement patterns and spatial behaviour explicit. I propose that movement patterns are the evidence of pedestrians interacting with the environment.

This thesis makes two contributions to the field of movement pattern analysis: it introduces an approach to linking movement data, patterns, interactions and movement behaviour, and it presents a set of methods and techniques for analysing pedestrian movement.

The approach presented here is developed for use in studying the movement of pedestrians in natural recreational areas. These areas constitute an ideal setting for studying pedestrian movement since they allow us to frame the research in clearly defined geographical spaces and to focus on a set of possible movement interactions. Moreover, the study of visitors' movements in these areas is critical for effective area management. We are especially interested in the stopping behaviour of visitors, since this behaviour represents a strong interaction between visitors and places in natural recreational areas.

6.2 Main research findings

The process of developing the proposed approach was driven by four specific research objectives, presented in the introduction. Each objective has been addressed in the core chapters of this thesis. In this section, the research objectives are revisited and discussed.

6.2.1 Develop an approach to detect movement patterns evidencing the stopping behaviour of pedestrians

A novel approach to detecting patterns evidencing the stopping behaviour of pedestrians using GPS tracking data was introduced in Chapter 2. Using a Local Indicator of Spatial Association (LISA), pedestrian movement data was analysed to detect spatial clusters of low-speed vectors. Those clusters, called Movement Suspension Patterns (MSPs), were used to identify the places where people stopped.

The procedure for detecting MSPs consists of the following sub-steps: 1) create a movement vector representation from GPS data; 2) compute the LISA and Z score of the speed values for each movement vector; 3) classify the movement vectors with a speed below the mean and with a Z score above a defined significance level as movement suspension; 4) plot on a map the vectors classified as suspension to find the location of MSPs; 5) interpret the locations that might indicate places where people stop.

The MSP approach is conceptually and methodologically different from other methods, such as Intersection-Based Stops and Moves (IB-SMoT, Alvares et al., 2007), Clustering-Based Stops and Moves (CB-SMoT, Palma et al., 2008) and other parameter-based methods. The main conceptual difference is that previous methods are based on the properties of individual movement represented as trajectories (the conceptualization of a stop depends on the movement of each individual), whereas MSPs are based on the collective properties of movement represented as a set of movement vectors (the conceptualization of an MSP depends on the movement of each individual and other individuals). These collective properties are defined by global characteristics (e.g., mean speed) and local characteristics (e.g., local spatial association) and are the result of the stopping behaviour of people interacting with the environment.

The main methodological difference is that previous approaches require the definition of parameters that are highly application-dependent. For example, an analyst has to decide the minimum time or maximum speed for detecting a stop. To do so, he or she has to consider the activities of the pedestrians, their environment, the spatial scale and the quality and resolution of collected data. A variation of these parameters may produce an over- or under-estimation of the detected stops. In contrast, the detection of MSPs does not require application-dependent thresholds or detailed knowledge about the pedestrians and the characteristics of the data collection. The only required parameters (i.e., mean speed, cut-off distance for the spatial relationship matrix) can be derived directly from the dataset. This constitutes an important advantage over previous methods since the proposed approach can be directly implemented on any movement dataset without previous knowledge about the application.

6.2.2 Evaluate the detected movement patterns using a controlled experiment

Although several methods have been proposed for detecting stops in movement data, their validity for describing stopping behaviour of people remains unclear. Similarly, the MSP approach proposed in Chapter 2 required an experimental validation to assess the extent to which the detected patterns are the results of actual spatial behaviour of pedestrians. Such a validation is essential to both researchers and practitioners interested in movement analysis since they need to evaluate different methods, both to compare their performance for specific cases and to choose a method that is both accurate and easy to apply.

The results in Chapter 2 suggested that MSPs were located at places where it was likely that pedestrians stopped, and a statistical significance level was used to select MSPs with low probability to be originated by some random process. However, as a thorough validation was still missing, a controlled experiment was designed to evaluate MSPs both in space and time. In this experiment, described in Chapter 3, a group of volunteers carrying GPS receivers were asked to follow predefined routes and stop to take pictures at predefined places indicated on a map. Participants had to record the times they stopped. The GPS data were later analysed to detect MSPs and the results were compared with the location and time of the stops reported by the participants.

The results showed that the method detected up to 92% of predefined stops. It was also found that 16% of the detected MSPs were false positives, although further analysis indicated that these could be related to actual suspension of movement by the participants. The evaluation also showed that the location of the detected MSPs largely corresponded with the location of the predefined reference stop places. Taking into account these results, it was confirmed that MSPs are a feasible approach for detecting the stopping behaviour of pedestrians.

The MSPs depend not only on the movement of each pedestrian (individual properties of movement), but also on the movement of other pedestrians and the places where they stop (collective properties of movement). For example, if a pedestrian stopped at a place where everybody else was walking, the spatial association of the speed values of the movement vectors could be not statistically significant, and therefore would not be detected as an MSP. However, if other pedestrians also stopped there (or the same pedestrian stopped at other times), the spatial association would be high enough and an MSP would be detected.

To investigate this property, the experiment was designed to include places with different proportions of participants stopping and walking (i.e., stopping ratio, Chapter 3, section 3.2). It was expected that places with a low stopping ratio would not be detected. The results, however, were not conclusive since some places with a stopping ratio of <0.025 were correctly detected. Looking closely at the spatial association values of movement vectors at these places, it is possible to see that their Z score is near the significance level. This implies that the differentiation between individual and collective properties of movement in terms of spatial association is not a crisp line, but rather a fuzzy region. Further controlled experiments must be conducted to investigate this assumption.

6.2.3 Demonstrate the applicability of the approach to studying the movement of visitors interacting with places in natural areas

A case study was presented in Chapter 4 to demonstrate the usefulness and applicability of the proposed approach. In this study, two kinds of movement patterns were analysed to study the aggregated flow of visitors in a natural park in the Netherlands. These patterns were Movement Suspension Patterns (MSPs) and Generalized Sequential Patterns (GSPs).

MSPs were used to discover places of interest for the visitors, under the assumption that those places are associated with the stopping behaviour of the visitors. GSPs denote the generalized order in which people visit the places regardless of the route followed and were used to discover commonalities in that order. Together, both patterns were used to explore and understand several aspects of the movement behaviour of the visitors. For example, they helped to determine which of the attractions in the park where the most visited, which facilities were used, which places were visited from each park entrance, and the orders in which those places were visited.

This case study demonstrated how tourism researchers and park managers could use the proposed framework to analyse the movement of visitors in natural areas and to evaluate and improve management practices. For example, park managers can analyse movement data collected with GPS devices to determine the location of facilities and services, to improve the information elements such as boards and signposts, and to design specific routes for different kinds of visitors. A long-term analysis of movement patterns can also help to evaluate the implementation of these practices. Besides, movement analysis methods can be used with surveys, interviews and other traditional methods to better understand the goals and motivations of visitors. This understanding becomes critical in natural areas, where the spatial behaviour of visitors is directly related to environmental impacts.

6.2.4 Develop an ontology to represent relations between movement patterns and behaviour

Chapter 5 introduced an ontological formalization to represent relations between movement patterns, interactions and movement behaviour. In this formalization, pedestrian movement was conceptualized as a set of interactions between pedestrians and between pedestrians and their environment. These interactions were defined in terms of movement patterns and context and formalized using the Web Ontology Language (OWL). Concepts and relations between concepts were both represented. The viability of the approach was demonstrated in a recreational application where some examples of concepts about movement behaviour of visitors, such as “exploring” or “disturbing” were formalized. Then the interactions related to these kinds of behaviour were defined using movement patterns and including semantics about the context in which patterns occur. It was also demonstrated how the interpretation of patterns and interactions depends on the contextual information.

Although it is in an early stage, the field of ontological representation is changing fast. In the last two years, other approaches to formal representations of movement behaviour have been introduced. For example, Van Hage, Wielemaker and Schreiber (2010) demonstrated how SWI-Prolog, a declarative semantic language, could be extended to support spatial reasoning using the Simple Event Model (Van Hage et al., 2009). In their work, the authors demonstrated the application of their approach by inferring movement behaviour of ships using predefined movement patterns. Similarly, Baglioni et al. (2009) presented a model for the conceptual representation and deductive reasoning of trajectory patterns obtained from mining raw trajectories. Here, the authors used a process called “semantic-enrichment” to annotate movement data with contextual geographic information using a sub-language of OWL called OWLPRIME (Oracle, 2010) for the implementation. More recently, Andrienko et al. (2011) presented a conceptual framework to describe the types of information that can be extracted from movement data as well as a taxonomy of analytical techniques for movement analysis. All these examples have similarities and differences to the approach presented in Chapter 5. All of them aim to bridge the gap between raw movement data and concepts about movement behaviour by interpreting movement patterns, but these new approaches are not specifically focused on pedestrian movement.

6.3 Moving beyond the patterns

Whereas a significant amount of research on movement analysis has been focused on pattern detection, the research described in this thesis attempted to look beyond the patterns. I approached the topic from a geographical point of view by explicitly linking the pedestrians and their environment. The geographical perspective also brings the possibility (and necessity) of different representations of movement. In the proposed approach, data are transformed to represent movement as a set of movement vectors to detect MSPs. Then the results are transformed again to represent movement as trajectories linking the places where MSPs occur. Finally, GSPs are used to represent movement as flows.

The geographical perspective on movement analysis therefore had two important implications. On the one hand, it enabled the production of multiple representations of movement to explore new ways to analyse movement data, or, more precisely, to create synergy between longstanding geographical techniques such as LISA or density functions and data-mining methods to analyse a new kind of data (i.e., GPS movement data). On the other hand, the geographical perspective raised an important conceptual implication: the close relationship between movement and place. In fact, an underlying assumption in this approach is that people move from one place to another, suspending their movement at those places. Places therefore “emerge” from movement, whose location and extent are indicated by the MSPs. The places where people suspend their movement play a central role in understanding pedestrian movement. They are more than locations where a person stops; they indicate the existence of an interaction between people and the environment (Stonor, 2004). This has its literary counterpart in the opening passage of this chapter.

6.4 Further research

Based on the results presented in this thesis, I suggest a number of directions for future research.

First, the current LISA method uses a matrix to represent the spatial relationship between observations. This spatial relationship was conceptualized as a simple distance-decay function in the Cartesian plane. However, more complex relationships can be explored and more dimensions can be included. For example, instead of Euclidean distance, cost-distance or network distance can be used to build the spatial relationship matrix to take into account the influence of other environmental factors. Moreover, it is possible to include the temporal dimension along with the spatial dimensions to build a matrix in which the spatiotemporal relationship between observations is represented. In turn, this matrix allows the spatiotemporal clusters of movement suspension to be identified.

Second, although the controlled experiment detailed in Chapter 3 was successful in demonstrating the validity of MSPs, further experiments will help to evaluate the temporal accuracy of the patterns. The moments when movement suspension begins and ends still have to be determined, taking into account the temporal granularity of the observations.

A third research line is related to the formal representation. Whereas the approach presented in Chapter 5 constitutes a first step towards the formal representation of movement patterns, several challenges still remain. The current ontology does not go beyond a proof of concept and is limited to some examples of interactions. More complex ontologies should be considered to extend the approach. For example, the taxonomy of collective phenomena proposed by Wood and Galton (2009) can be used as a starting point for a comprehensive representation of pedestrian movement behaviour.

Finally, although the research presented in this thesis was focused on pedestrian movement in recreational areas, the approach can be also useful in different applications in which pedestrian movement is involved. For example, urban planners can evaluate the attractiveness of public space by taking into account the places where movement of people is suspended (Stonor, 2004). Also, movement patterns can be used to improve indicators of social sustainability in public areas (Ostermann, 2009). In mobility management, exploring and understanding movement patterns is essential to the design and implementation of better mobility and transportation systems. Moreover, the suitability of the approach to other domains such as transportation management or movement ecology is still to be investigated.

References

Agrawal, R. and Srikant, R. (1995). Mining sequential patterns. In *Proceedings of the Eleventh International Conference on Data Engineering.*

Alvares, L. O., Bogorny, V., Kuijpers, B., Macedo, J., Moelans, B., and Vaisman, A. A. (2007). A model for enriching trajectories with semantic geographical information. In *Proceedings of the 15th annual ACM international symposium on Advances in geographic information systems*. November 7 - 9.

Andersson, M., Gudmundsson, J., Laube, P., and Wolle, T. (2008). Reporting Leaders and Followers among Trajectories of Moving Point Objects. *GeoInformatica* 12:497-528.

Andrienko, G. and Andrienko, N. (2008). A Visual Analytics Approach to Exploration of Large Amounts of Movement Data. In Sebillo, M., Vitiello, G. and Schaefer, G., (Eds.) *Visual Information Systems, 10th International Conference, VISUAL 2008,* 1-4. Springer-Berlin.

Andrienko, G. and Andrienko, N. (2007). Extracting patterns of individual movement behaviour from a massive collection of tracked positions. In *Workshop on Behaviour Modelling and Interpretation (BMI)*.

Andrienko, G., Andrienko, N., Bak, P., Keim, D., Kisilevich, S., and Wroblel, S. (2011). A conceptual framework and taxonomy of techniques for analyzing movemen. *Journal of Visual Languages and Computing* 22:213-232.

Andrienko, G., Andrienko, N., Kopanakis, I., Ligtenberg, A., and Wrobel, S. (2008). Visual Analytics Methods for Movement Data. In Giannotti, F. and Pedreschi, D., (Eds.) *Mobility, Data Mining and Privacy,* 375-410. Berlin: Springer-Verlag.

Andrienko, N., Andrienko, G., Pelekis, N., and Spaccapietra, S. (2008). Basic Concepts of Movement Data. In Giannotti, F. and Pedreschi, D., (Eds.)*Mobility, Data Mining and Privacy,* 15-38. Berlin: Springer-Verlag.

Andrienko, N., Andrienko, G., Wachowicz, M., and Orellana, D. (2008). Uncovering Interactions between Moving Objects. In *Geographic Information Science, Extended Abstracts of 5th International Conference, GIScience 2008*. Park City, Utah, USA. September 23-26, 2008.

Anselin, L. (1995). Local indicators of spatial association — LISA. *Geographical Analysis* 27:93-115.

Anselin, L. (1993). The Moran Scatterplot as an ESDA Tool to Assess Local Instability in Spatial Association. In Fischer, M. M., Scholten, H. J. and Unwin, D. J., (Eds.)*Spatial Analytical Perspectives on GIS,* 111-127. London: Taylor & Francis.

Antonini, G., Bierlaire, M., and Weber, M. (2006). Discrete choice models of pedestrian walking behavior. *Transportation Research Part B: Methodological* 40:667-687.

Ashbrook, D. and Starner, T. (2003). Using GPS to learn significant locations and predict movement across multiple users. *Personal and Ubiquitous Computing* 7:275-286.

Baader, F., Calvanese, D., McGuinness, D., Nardi, D., and Patel-Schneider, P. (2003). The description logic handbook: theory, implementation and applications. Cambridge: Cambridge University Press.

Baglioni, M., Macedo, J., Renso, C., Trasarti, R., and Wachowicz, M. (2009). Towards a semantic interpretation of movement behaviour. In Sester, M., Bernard, L. and Paelke, V., (Eds.)*Advances in GIScience,* 271-288. Berlin: Springer.

Baglioni, M., Macedo, J., Renso, C., and Wachowicz, M. (2008). An Ontology-Based Approach for the Semantic Modelling and Reasoning on Trajectories. In *ER '08: Proceedings of the ER 2008 Workshops on Advances in Conceptual Modeling*. Barcelona, Spain.

Batty, M. (2003). Agent-based pedestrian modelling. In Batty, M. and Longley, P. A., (Eds.)*Advanced Spatial Analysis: The CASA book of GIS,* 81-106. Redlands: Esri Press.

Batty, M., DeSyllas, J., and Duxbury, E. (2003). The discrete dynamics of small-scale spatial events: agent-based models of mobility in carnivals and street parades. *International Journal of Geographical Information Science* 17:673-697.

Benkert, M., Gudmundsson, J., Hübner, F., and Wolle, T. (2008). Reporting flock patterns. *Computational Geometry* 41:111-125.

Bettini, C., Brdiczka, O., Henricksen, K., Indulska, J., Nicklas, D., Ranganathan, A., and Riboni, D. (2010). A survey of context modelling and reasoning techniques. *Pervasive and Mobile Computing* 6:161-180.

Beyer, H. L. (2010). Geospatial Modelling Environment v0.4. Available from http://www.spatialecology.com (last accessed 10/11 2010).

Bian, L. (2004). A conceptual framework for an individual-based spatially explicit epidemiological model. *Environment and Planning B* 31:381-396.

Bierlaire, M., Antonini, G., and Weber, M. (2007). Behavioral dynamics for pedestrians. In Axhausen, K., (Ed.)*Moving through nets: The physical and social dimensions of trave,* 81-105. Amsterdam: Elsevier.

Blue, V. J. and Adler, J. L. (2001). Cellular automata microsimulation for modeling bi-directional pedestrian walkways. *Transportation Research Part B* 35:293-312.

Blythe, P. W., Miller, G. F., and Todd, P. M. (1996). Human simulation of adaptive behavior: Interactive studies of pursuit, evasion, courtship, fighting, and play. In *From animals to animats 4: Proceedings of the Fourth International Conference on Simulation of Adaptive Behavior*.

Bogorny, V., Heuser, C., and Alvares, L. (2010). A Conceptual Data Model for Trajectory Data Mining. In Fabrikant, S. I., Reichenbacher, T., Kreveld, M. V. and Schlieder, C., (Eds.)*Geographic Information Science.6th International Conference, GIScience 2010,* 1-15. Springer.

Bouquet, P., Ghidini, C., Giunchiglia, F., and Blanzieri, E. (2003). Theories and uses of context in knowledge representation and reasonin. *Journal of pragmatics* 35:455-484.

Burstedde, C., Klauck, K., Schadschneider, A., and Zittartz, J. (2001). Simulation of pedestrian dynamics using a two-dimensional cellular automaton. *Physica A: Statistical Mechanics and its Applications* 295:507-525.

Calvanese, D., Lembo, D., Lenzerini, M., Poggi, A., and Rosati, R. (2007). Ontology-based database access. In *Proc. of the 15th Italian Conf. on Database Systems (SEBD)*.

Chiesura, A. (2004). The role of urban parks for the sustainable city. *Landscape and Urban Planning* 68:129-138.

Cliff, A. and Ord, J. K. (1972). Testing for spatial autocorrelation among regression residuals. *Geographical Analysis* 4:267-284.

Cliff, A. and Ord, J. K. (1981). Spatial Processes: Models & Applications. London: Pion Ltd.

Colizza, V., Barrat, A., Barthelemy, M., Valleron, A. J., and Vespignani, A. (2007). Modeling the worldwide spread of pandemic influenza: baseline case and containment interventions. *PLoS medicine* 4:95-110.

Connell, J. and Page, S. (2008). Exploring the spatial patterns of car-based tourist travel in Loch Lomond and Trossachs National Park, Scotland. *Tourism Management* 29:561-580.

Daamen, W. (2004). Modelling Passenger Flows in Public Transport Facilities. Amsterdam: IOS Press.

Daamen, W. and Hoogendoorn, S. P.). Controlled experiments to derive walking behaviour. *European Journal of Transport and Infrastructure Research* 3:39-59.

Daamen, W. and Hoogendoorn, S. P. (2003). Experimental research of pedestrian walking behavior. *Transportation Research Record* 1828:20-30.

Davis, R., Shrobe, H., and Szolovits, P. (1993). What is A Knowledge Representation? *AI Magazine* 14:17-33.

de Smith, M. J., Goodchild, M. F., and Longley, P. A. (2009). Geospatial Analysis - A comprehensive guide. Leicester: Matador.

Demšar, U. and Virrantaus, K. (2010). Space–time density of trajectories: exploring spatio-temporal patterns in movement data. *International Journal of Geographical Information Science* 24:1527-1542.

Dodge, S., Weibel, R., and Lautenschütz, A. K. (2008). Towards a taxonomy of movement patterns. *Information Visualization* 7:240-252.

ESRI. (2008). ArcGIS 9.2.

Eyles, J. (1971). Pouring New Sentiments into Old Theories: How Else Can We Look at Behavioural Patterns? *Area* 3:242-250.

Fayyad, U., Piatetsky-Shapiro, G., and Smyth, P. (1996). From Data Mining to Knowledge Discovery in Databases. *AI Magazine* 17(3):37-54.

Fournier-Viger, P., Nkambou, R., and Nguifo, E. (2008). A knowledge discovery framework for learning task models from user interactions in intelligent tutoring systems. *MICAI 2008: Advances in Artificial Intelligence* 5317/2008:765-778.

Fruin, J. J. (1971). Designing for pedestrians. A level-of-service concept. *Highway Research Record* 377:1-15.

Galea, E. R., (Ed.).(2003). Pedestrian and Evacuation Dynamics. London: CMS Press.

Galton, A. (2005). Dynamic collectives and their collective dynamics. *Lecture notes in computer science* 3693:300.

Ganskopp, D. C. and Johnson, D. D. (2007). GPS error in studies addressing animal movements and activities. *Rangeland Ecology & Management* 60:350-358.

Getis, A. and Ord, J. K. (1992). The analysis of spatial association by use of distance statistics. *Geographical Analysis* 24:189-206.

Giannotti, F., Nanni, M., Pinelli, F., and Pedreschi, D. (2007). Trajectory pattern mining. In *KDD '07: Proceedings of the 13th ACM SIGKDD international conference on Knowledge discovery and data mining*. San Jose, California, USA.

Giunchiglia, F. and Bouquet, P. (1997). Introduction to contextual reasoning. An Artificial Intelligence Perspective. *Perspectives on Cognitive Science* 3:138-159.

Golicnik, B. and Ward Thompson, C. (2010). Emerging relationships between design and use of urban park spaces. *Landscape and Urban Planning* 94:38-53.

Gonzalez, M. C., Hidalgo, C. A., and Barabasi, A. (2008). Understanding individual human mobility patterns. *Nature* 453:779-782.

Gruber, T. (2008). *Ontology*. In Liu, L. and Özsu, M. T., (Eds.) *Encyclopedia of DatabaseSystems,* Springer-Verlag.

Guarino, N. (1998). Formal ontology in information systems. In *Proceedings of FOIS'98*. Trento, Italy. 6-8 June 1998.

Gudmundsson, J., Laube, P., and Wolle, T. (2009). Movement patterns in spatio-temporal data. In Shekhar, S. and Xiong, H., (Eds.) *Encyclopedia of GIS,* 726-732. New York: Springer-Verlag.

Gudmundsson, J., van Kreveld, M., and Speckmann, B. (2004). Efficient detection of motion patterns in spatio-temporal data sets. In *Proceedings of the 12th annual ACM international workshop on Geographic information systems.*

Güting, R. H. and Schneider, M. (2005). Moving Objects Databases. London: Morgan Kaufmann.

Hägerstrand, T. (1970). What about people in Regional Science? *Papers in Regional Science* 24:6-21.

Haining, R. P. (2003). Spatial data analysis: theory and practice. Cambridge: Cambridge University Press.

Hall, T. and Cole, D. N. (2000). An expanded perspective on displacement: A longitudinal study of visitors to two wildernesses in the Cascade Mountains of Oregon. In *Wilderness science in a time of change conference-Volume 4: Wilderness visitors, experiences, and visitor management.* Missoula, USA. 23-27 May 1999.

Han, J. and Kamber, M. (2006). Data mining: concepts and techniques. Amsterdam: Morgan Kaufmann.

Helbing, D., Buzna, L., Johansson, A., and Werner, T. (2005). Self-Organized Pedestrian Crowd Dynamics: Experiments, Simulations, and Design Solutions. *Transportation Science* 39:1-24.

Helbing, D., Molnár, P., Farkas, I. J., and Bolay, K. (2001). Self-organizing pedestrian movement. *Environment and Planning B* 28:361-384.

Helbing, D., Twining, C. J., Bulsari, A., Oliker, S., Furst, M., and Maimon, O. (1993). A Fluid-Dynamic Model for the Movement of Pedestrians. *Complex Systems* 6:391-415.

Hightower, J., Consolvo, S., LaMarca, A., Smith, I., and Hughes, J. (2005). Learning and recognizing the places we go. In *UbiComp 2005: Ubiquitous Computing, 7th International Conference.* Tokyo. September 11-14.

Hillier, B. (1984). Mansion Square report inquiry – proof of evidence. London: University College London.

Hillier, B., Penn, A., Hanson, J., Grajewski, T., and Xu, J. (1993). Natural movement: or, configuration and attraction in urban pedestrian movement. *Environment and Planning B: Planning and Design* 20:29-66.

Holyoak, N. and Carson, D. (2009). Modelling self-drive tourist travel patterns in desert Australia. In *32 Australasian Transport Research Forum - ATRF 2009 (Auckland).* Auckland, New Zealand. 29 September - 1 October 2009.

Hoogendoorn, S. P. and Bovy, P. H. L. (2005). Pedestrian travel behavior modeling. *Networks and Spatial Economics* 5:193-216.

Hoogendoorn, S. P., Bovy, P. H. L., and Daamen, W. (2002). Microscopic pedestrian wayfinding and dynamics modelling. In Schreckenberg, M. and Sharma, S., (Eds.) *Pedestrian and Evacuation Dynamics,* 123-155. Berlin: Springer.

Horner, M. W. and O'Kelly, M. E. (2001). Embedding economies of scale concepts for hub network design. *Journal of Transport Geography* 9:255-265.

Hornsby, K. S. and Egenhofer, M. J. (2002). Modeling Moving Objects over Multiple Granularities. *Annals of Mathematics and Artificial Intelligence* 36:177-194.

IAIS. (2008). CommonGIS.

Ittelson, W. H., Rivlin, L. G., and Prohansky, H. M. (1970). The use of behavioural maps in environmental psychology. In Proshansky, H. M., Ittelson, W. H. and Rivlin, L. G., (Eds.) *Environmental psychology: Man and his physical setting,* 658-668. New York: Holt, Rinehart and Winston.

Kalnis, P., Mamoulis, N., and Bakiras, S. (2005). On discovering moving clusters in spatio-temporal data. In Bauzer-Medeiros, C., Egenhofer, M. J. and Bertino, E., (Eds.) *Advances in Spatial and Temporal Databases,* 364-381. Springer.

Kang, J. H., Welbourne, W., Stewart, B., and Borriello, G. (2004). Extracting places from traces of locations. In *Proceedings of the 2nd ACM international workshop on Wireless mobile applications and services on WLAN hotspots*. Philadelphia, PA, USA.

Kulldorff, M. (1997). A spatial scan statistic. *Communications in Statistics - Theory and Methods* 26:1481-1496.

Kundera, M. (1991). Immortality. London: Faber & Faber.

Kwan. (2004). GIS Methods in Time-Geographic Research: Geocomputation and Geovisualization of Human Activity Patterns. *Geografiska Annaler Series B Human Geography* 86:267.

Laube, P. (2009). Progress in Movement Pattern Analysis. In Gottfried, B. and Aghajan, H. K., (Eds.) *Behaviour Monitoring and Interpretation. BMI, Smart Environments,* 43-71. Amsterdam: IOS Press BV.

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

Laube, P., Kreveld, M. V., and Imfeld, S. (2005). Finding REMO — Detecting Relative Motion Patterns in Geospatial Lifelines. In Fisher, P., (Ed.) *Developments in Spatial Data Handling,* 201-215. Springer Berlin Heidelberg.

Lee-Gosselin, M., Doherty, S. T., and Shalaby, A. (2010). Data Collection on Personal Movement Using Mobile ICTs: Old Wine in New Bottles? In Wachowicz, M., (Ed.) *Mobility-aware applications for Sustainable Mobility. Technologies and Approaches,* 1-14. New York: IGI Global.

Li, X. and Hodgson, M. E. (2004). Vector field data model and operations. *GIScience & Remote Sensing* 41:1-24.

Maidment, D. R. (1993). GIS and hydrologic modeling. In Goodchild, M. F., Parks, B. O. and Steyaert, L. T., (Eds.) *Environmental Modeling with GIS* 147-167. New York: Oxford University Press.

Manco, G., Baglioni, M., Giannotti, F., Kuijpers, B., Raffaetà, A., and Renso, C. (2008). Querying and Reasoning for Spatiotemporal Data Mining. In Giannotti, F. and Pedreschi, D., (Eds.) *Mobility, Data Mining and Privacy,* 335-374. Berlin: Springer-Verlag.

Manning, R. E. and Valliere, W. A. (2001). Coping in Outdoor Recreation: Causes and Consequences of Crowding and Conflict Among Community Residents. *Journal of leisure research,* 33:410-426.

Marmasse, N. and Schmandt, C. (2000). Location-aware information delivery with commotion. In Thomas, P. and Gellersen, H. W., (Eds.) *Handheld and Ubiquitous Computing,* 157-171. Berlin Heidelberg: Springer.

Mckercher, B. and Lau, G. (2008). Movement Patterns of Tourists within a Destination. *Tourism Geographies: An International Journal of Tourism Space, Place and Environment* 10:355-374.

McQuistan, R. B. (1965). Scalar and vector fields: a physical interpretation. New York: John Wiley & Sons.

Miller, H. J. (2005). A Measurement Theory for Time Geography. *Geographical Analysis* 37:17-45.

Miller, H. J. and Bridwell, S. (2009). A Field-Based Theory for Time Geography. *Annals of the Association of American Geographers* 99:49-75(27).

Millonig, A. and Gartner, G. (2007). Monitoring pedestrian spatio-temporal behaviour. In *Workshop on Behaviour Monitoring and Interpretation BMI 07*. Osnabrück.

Modsching, M., Kramer, R., Gretzel, U., and Hagen, K. (2006). Capturing the beaten paths: A novel method for analysing tourists' spatial behaviour at an urban destination. In Hitz, M., Sigala, M. and Murphy, J., (Eds.) *Information and Communication Technologies in Tourism,* 75-86. Viena: Springer.

Moran, P. A. P. (1948). The interpretation of statistical maps. *Journal of the Royal Statistical Society B* 10:243-251.

Mountain, D. and Raper, J. (2001). Positioning techniques for location-based services (LBS): Characteristics and limitations of proposed solutions. *Aslib Proceedings* 53:404-412.

Muhar, A., Arnberger, A., and Brandenburg, C. (2002). Methods for Visitor Monitoring in Recreational and Protected Areas: An Overview. In *Monitoring and Management of Visitor Flows in Recreational and Protected Areas*. Vienna. anuary 30 – February 02, 2002.

Nanni, M., Kuijpers, B., Korner, C., May, M., and Pedreschi, D. (2008). Spatiotemporal Data Mining. In Giannotti, F. and Pedreschi, D., (Eds.) *Mobility, Data Mining and Privacy,* 267-296. Berlin: Springer-Verlag.

National Marine Electronics Association. (2010). NMEA Standards. Available from http://www.standards.nmea.org (last accessed 10/11 2010).

Natuurmonumenten. (2009). Natuurmonumenten Web Site. Available from http://www.natuurmonumenten.nl (last accessed October/01 2009).

Nielsen, T. and Hovgesen, H. (2004). GPS in pedestrian and spatial behaviour Surveys. In *to Walk21-V Cities for People, The Fifth International Conference on. Walking in the 21st. Century*. Copenhagen. 9-11 June 2004.

Noyon, V., Claramunt, C., and Devogele, T. (2007). A Relative Representation of Trajectories in Geogaphical Spaces. *GeoInformatica* 11:479-496.

OpenStreetMap. (2011a). JOSM. v 3773.

OpenStreetMap. (2011)b. OpenStreetMap. Available from http://www.openstreetmap.org (last accessed 01/04/2011).

Oracle Co. (2010). Information Management with Oracle Database 11g. 1-13.

Orellana, D. (2010). Ideas on Movement. Available from http://ideasonmovement.wordpress.com (last accessed 11/10 2010).

Orellana, D., Bregt, A. K., Ligtenberg, A., and Wachowicz, M. (2012). Exploring visitor movement patterns in natural recreational areas. *Tourism Management* 33(3):672-682.

Orellana, D. and Renso, C. (2010). Developing an Interactions Ontology for Characterising Pedestrian Movement. In Wachowicz, M., (Ed.) *Movement-Aware Aplications for Sustainable Mobility,* 62-86. New York: Information Science Reference.

Orellana, D. and Wachowicz, M. (2011). Exploring patterns of movement suspension in pedestrian mobility. *Geographical Analysis* 43(3):241-260.

Orellana, D., Wachowicz, M., Andrienko, N., and Andrienko, G. (2009). Uncovering Interaction Patterns in Mobile Outdoor Gaming. In *Advanced Geographic Information Systems & Web Services, 2009. GEOWS '09*. Cancún. 1 - 7 Feb. 2009.

Orellana, D., Wachowicz, M., Knegt, H. d., Ligtenberg, A., and Bregt, A. K. (2010). Uncovering Patterns of Suspension of Movement. In *GIScience 2010 extended Abstracts*. Zurich. 14-17th September, 2010.

Ortale, R., Manco, G., Ritacco, E., Pelekis, N., Trasarti, R., Costa, G., Giannotti, F., Manco, G., Renso, C., and Theodoridis, Y. (2008). DAEDALUS: Towards Progressively Querying and Mining Movement Data. ICAR-CNR.

Ostermann, F. O. (2009). Indicators for Socially Sustainable Park Use – Results from a Case Study. In *Proceedings of the 14th Conference on Regional Planning*. Sitges, Spain. 22-25 April 2009.

Palma, A. T., Bogorny, V., Kuijpers, B., and Alvares, L. O. (2008). A clustering-based approach for discovering interesting places in trajectories In *Proceedings of the 2008 ACM symposium on Applied computing*.

Pauls, J. (1984). The movement of people in buildings and design solutions for means of egress. *Fire technology* 20:27-47.

Pelekis, N., and Theodoridis, Y. (2007). An Oracle Data Cartridge for Moving Objects. Report UNIPI-ISL-TR-2007-04. Information Systems Lab., University of Piraeus, Piraeus, Hellas.

Penn, A. and Turner, A. (2001). Space syntax based agent simulation. In Schreckenberg, M. and Sharma, S., (Eds.) *Pedestrian and Evacuation Dynamics,* 99-114. Springer.

Pfoser, D. and Jensen, C. S. (2001). Querying the trajectories of on-line mobile objects. In *MobiDe '01: Proceedings of the 2nd ACM international workshop on Data engineering for wireless and mobile access*. Santa Barbara, California, United States.

Pfoser, D., Jensen, C. S., and Theodoridis, Y. (2000). Novel Approaches to the Indexing of Moving Object Trajectories. In *Proceedings of the 26th International Conference on Very Large Data Bases*.

Pol-Recreatie. (2003). Recreatieverblijf 't Polletje. Available from http://www.pol-recreatie.nl/ (last accessed October/02 2009).

Renso, C., Puntoni, S., Frentzos, E., Mazzoni, A., Moelans, B., Pelekis, N., and Pini, F. (2008). Wireless Network Data Sources: Tracking and Synthesizing Trajectories. In Giannotti, F. and Pedreschi, D., (Eds.) *Mobility, Data Mining and Privacy,* 73-100. Berlin: Springer-Verlag.

Riboni, D. and Bettini, C. (2011). OWL 2 modeling and reasoning with complex human activities. *Pervasive and Mobile Computing* In Press, Corrected Proof:.

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* 7:225-239.

Scheepens, R., Willems, N., van de Wetering, H., and van Wijk, J. J. (2011). Interactive visualization of multivariate trajectory data with density maps. In *Pacific Visualization Symposium (PacificVis), 2011 IEEE*. Hong Kong, China. 1-4 March 2011.

Schlieder, C. and Werner, A. (2003). Interpretation of Intentional Behavior in Spatial Partonomies. In Freksa, C., Brauer, W., Habel, C. and Wender, K. F., (Eds.) *Spatial Cognition III,* 1035-1035. Berlin: Springer.

Shaw, S. and Yu, H. (2009). A GIS-based time-geographic approach of studying individual activities and interactions in a hybrid physical–virtual space. *Journal of Transport Geography* In Press, Corrected Proof:.

Shaw, S., Yu, H., and Bombom, L. (2008). A Space-Time GIS Approach to Exploring Large Individual-based Spatiotemporal Datasets. *Transactions in GIS* 12:425-441.

Shoval, N. (2010). Monitoring and Managing Visitors Flows in Destinations using Aggregative GPS Data. In Gretzel, U., Law, R. and Fuchs, M., (Eds.) *Information and Communication Technologies in Tourism 2010,* 171-183. Vienna: Springer Vienna.

Shoval, N. and Isaacson, M. (2009). Tourist Mobility and Advanced Tracking Technologies. Routledge.

Shoval, N. and Isaacson, M. (2006). Application of Tracking Technologies to the Study of Pedestrian Spatial Behavior. *The Professional Geographer* 58:172-183.

Silberschatz, A. and Tuzhilin, A. (1995). On Subjective Measures of Interestingness in Knowledge Discovery. In *The First International Conference on Knowledge Discovery & Data Mining*. Montreal, Quebec, Canada. August 20–21, 1995.

Silverman, B. W. (1986). Density Estimation for Statistics and Data Analysis. Chapman & Hall/CRC.

Sirin, E., Parsia, B., Grau, B. C., Kalyanpur, A., and Katz, Y. 2007. Pellet: A practical owl-dl reasoner. *Web Semantics: Science, Services and Agents on the World Wide Web* 5:51-53.

Song, C., Qu, Z., Blumm, N., and Barabasi, A. 2010. Limits of Predictability in Human Mobility. *Science* 327:1018-1021.

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

Stanford Center for Biomedical Informatics Research. 2008. Protégé.

Stonor, T. (2004). Stop that person. Strategic value and the design of Public Spaces. *Locum Destination Review* 16:28-31.

Stonor, T., de Arruda Campos, M. B., and Smith, A. 2001. Towards a walkability index. In *Walk 21.3rd International conference. Steps towards liveable cities*. Donostia-San Sebastián. 9th - 10th May 2001.

Taczanowska, K., Muhar, A., and Brandenburg, C. (2008). Potential and limitations of GPS tracking for monitoring spatial and temporal aspects of visitor behaviour in recreational areas. In *Fourth International Conference on Monitoring and Management of Visitor Flows in Recreational and Protected Areas*.

Terrier, P., Ladetto, Q., Merminod, B., and Schutz, Y. (2000). High-precision satellite positioning system as a new tool to study the biomechanics of human locomotion. *Journal of Biomechanics* 33:1717-1722.

Thomas, J. J. and Cook, K. A. (2005). Illuminating the path: The research and development agenda for visual analytics. IEEE.

Tobler, W. (2003). Movement Mapping. (last accessed April/9 2011).

Topografix. (2010). GPX 1.1 Schema Documentation. (last accessed April/9 2011).

Turner, A., Doxa, M., O'Sullivan, D., and Penn, A. (2001). From isovists to visibility graphs: a methodology for the analysis of architectural space. *Environment and Planning B: Planning and Design* 28:103-121.

U. S. National Library of Medicine. (2010). Medical Subject Headings. Available from http://www.nlm.nih.gov/cgi/mesh/2010 (last accessed 11/10 2010).

van der Spek, S. (2006). Pedestrian oriented design, designing the sustainable city - a tool to optimize pedestrian routing In Wang, V., Sheng, Q. and Sezer, C., (Eds.) *Modernization and Regionalism. Re-inventing Urban Identity* 693-697. Delft: TU Delft.

Van Hage, W. R., Malaisé, V., de Vries, G., Schreiber, G., and van Someren, M. (2009). Combining ship trajectories and semantics with the simple event model (SEM). In *Proceedings of the 1st ACM international workshop on Events in multimedia*. Beijing, China. 23 October 2009.

Van Hage, W. R., Wielemaker, J., and Schreiber, G. (2010). The Space Package: Tight Integration between Space and Semantics. *Transactions in GIS* 14:131-146.

van Marwijk, R. (2009). These routes are made for walking: understanding the transactions between nature, recreational behaviour and environmental meanings in Dwingelderveld National Park, the Netherlands. Dissertation, Wageningen University.

van Marwijk, R. and Pitt, D. G. (2008). Where Dutch recreationists walk: path design, physical features and walker usage. In *The Fourth International Conference on Monitoring and Management of Visitor Flows in Recreational and Protected Areas,* *Management for Protection and Sustainable Development*. Montecatiny Terme, Italy. 14-19 October 2008.

van Schaick, J. and van der Spek, S., (Eds). (2008). Urbanism on track: application of tracking technologies in urbanism. Amsterdam: IOS Press BV.

W3C Consortium. (2004). OWL Web Ontology Language.

Waag Society. (2008). Frequentie 1550. Available from http://www.waag.org/project/frequentie (last accessed 11/07/ 2008).

Wachowicz, M., (Ed.). (2010). Movement-Aware Aplications for Sustainable Mobility. New York: Information Science Reference.

Wang, J., Han, J., and Li, C. (2007). Frequent Closed Sequence Mining without Candidate Maintenance. *IEEE Transactions on Knowledge and Data Engineering* 19:1042-1056.

Weaver, C. (2008). Cross-dimensional Visual Queries for Interactive Animated Analysis of Movement. In *Proceedings of Geospatial Visual Analytics Workshop at GIScience 2008*. Utah. September 2008.

Witte, T. H. and Wilson, A. M. (2004). Accuracy of non-differential GPS for the determination of speed over ground. *Journal of Biomechanics* 37:1891-1898.

Wolf, J., Guensler, R., and Bachman, W. (2001). Elimination of the travel diary: Experiment to derive trip purpose from global positioning system travel data. *Transportation Research Record: Journal of the Transportation Research Board* 1768:125-134.

Wolfson, O., Xu, B., Chamberlain, S., and Jiang, L. (1998). Moving objects databases: Issues and solutions. In *Proceedings of the 10th International Conference on Scientific and Statistical Database Management*.

Wood, Z. and Galton, A. (2009). A taxonomy of collective phenomena. *Applied Ontology* 4:267-292.

Wood, Z. and Galton, A. (2008). Collectives and How They Move: A Tale of Two Classifications. In *Proceedings of the 2nd Workshop on Behaviour Monitoring and Interpretation, BMI'08*. Kaiserslautern.

Xia, J., Arrowsmith, C., Jackson, M., and Cartwright, W. 2008. The wayfinding process relationships between decision-making and landmark utility. *Tourism Management* 29:445-457.

Yoon, J., Noble, B., and Liu, M. (2007). Surface street traffic estimation. In *Proceedings of the 5th international conference on Mobile systems, applications and services*. San Juan. June 14-19, 2007.

Yu, H. and Shaw, S. (2007). Revisiting Hägerstrand's time-geographic framework for individual activities in the age of instant access. In Miller, H. J., (Ed.) *Societies and Cities in the Age of Instant Access,* 103-118. Dordrecht: Springer.

Zheng, X., Zhong, T., and Liu, M. (2009). Modeling crowd evacuation of a building based on seven methodological approaches. *Building and Environment* 44:437-445.

Summary

Walking is one of the main activities in natural recreational areas. Its importance goes far beyond the simple goal of travelling from one point to another; it provides pleasure and wellbeing to visitors looking for a closer connection with nature. People walking in these areas are in direct relationship with their environment, interacting with geographic features that in turn affect and modulate people's movements. Therefore, the movement of visitors in natural areas can be studied from a geographical point of view to understand the interactions between people and the environment in which they move. The understanding of such movement interactions can help park managers to meet the challenges caused by increasing use of recreational areas and to implement and evaluate management practices to reduce the impact of a growing number of visitors. It also provides new insights into human spatial behaviour, allowing researchers to increase their knowledge about the relationship between people and the environment.

The exploratory analysis of these "movement interactions" of visitors walking in recreational areas is the central theme of this thesis. I approach the topic by creating a synergy between spatial analysis and data-mining techniques to analyse massive movement datasets collected using GPS technology. Under the assumption that these data enclose evidence of spatial interactions between people and environment, I develop an approach for pedestrian movement analysis to uncover these interactions and gain knowledge about the spatial behaviour of pedestrians in natural areas. The proposed approach, based on a "knowledge discovery process", goes through successive steps of analysis from raw observational data towards concepts about movement interactions by detecting and analysing movement patterns.

The main goal of this thesis, therefore, is to develop an approach to exploring, analysing and interpreting movement patterns of pedestrians interacting with the environment. To accomplish the main objective, four specific goals were formulated: a) develop an approach to detecting movement patterns evidencing the stopping behaviour of pedestrians; b) evaluate the detected movement patterns using a controlled experiment; c) demonstrate the applicability of the approach to studying the movement of visitors interacting with places in natural areas; and d) develop an ontology to represent relations between movement patterns and behaviour.

Chapter 2 presents an exploratory spatial analysis approach to detecting patterns of movement suspension using a Local Indicator of Spatial Association (LISA). Patterns of movement suspension are used to find places where pedestrians stop as a consequence of their interactions with geographical features of the environment usually associated with specific human activities. These allow us to understand pedestrian movement behaviour. The proposed approach consists of

the following steps: a) collect data about pedestrian movement from GPS receivers; b) compute movement vectors to create a vector-based representation of movement; c) compute the LISA value for the speed values of movement vectors; d) classify the results using the average speed and the Z score of the LISA values; and finally e) plot the classified vectors on a map to find places where Suspension Patterns occur. Two different positioning datasets are used to demonstrate the approach: players of an urban outdoor mobile game and visitors to a natural park. The results of both experiments show that patterns of movement suspension were located at places such as checkpoints in the game and different attractions and facilities in the park. The results suggest that LISA is an appropriate indicator for exploring Movement Suspension Patterns (MSPs) representing the places where the movement of pedestrians is suspended by geographical features such as attractions or obstacles.

Chapter 3 presents the results of a controlled experiment to investigate the validity of using MSPs to represent the stopping behaviour of visitors in a natural recreational area. In the experiment, participants walked and stopped at designated locations in the Dwingelderveld National Park (the Netherlands), carrying GPS receivers to collect data. The collected data were analysed to detect MSPs using the method described in Chapter 2. The accuracy of the detected MSPs was evaluated in space and time. The occurrence of detected MSPs was compared in time with a set of reference stops and the MSP approach was found to detect 92% of them, with a false positive rate of $\alpha = 0.16$. When the location and extent of places of movement suspension computed as Percent Volume Contours (PVCs) in a Kernel Density Surface were compared with a set of predefined stopping places, 96% of them lay inside the areas delineated by the PVCs. These results confirmed that MSPs are feasible representations of the occurrence and location of stops in pedestrian movement data.

Chapter 4 is of interest to park managers and other tourism researchers and practitioners and demonstrates how movement pattern analysis can be used to improve understanding of the aggregated movement of visitors in natural recreational areas. The chapter presents a case study in which the movement of visitors in the Dwingelderveld National Park (the Netherlands) was studied through the combined analysis of two kinds of movement patterns: Movement Suspension Patterns (MSPs) and Generalized Sequential Patterns (GSPs). MSPs representing the movement suspension occurring when walkers stop at a location are used to discover places of interest to visitors. GSPs representing the generalized sequence in which the places are visited (regardless of the route followed), are used to uncover commonalities in the way that people visit the area. Both patterns were analysed in a geographical context to characterize the aggregated flow of people and provide insights into visitors' preferences and their interactions with the environment.

Chapter 5 introduces a spatial knowledge representation for the main concepts about pedestrian movement (i.e., patterns, interactions and spatial behaviour).

This constitutes a proof of concept for a formal representation about those concepts and the relationships between them. The representation consists of an ontology in which both concepts and relationships between concepts are formally represented. The ontology is implemented in the Web Ontology Language (OWL) using the Protégé framework. Finally, a set of instances of movement patterns is imported into the ontology to test some basic reasoning tasks, such as instance checking and classification.

Chapter 6 argues, on the basis of the findings of the preceding chapters, that the process for knowledge discovery in movement data is a feasible geographical approach to pedestrian movement analysis. The geographical perspective to movement analysis had two important implications. First, it enabled the synergy of multiple representations of movement to explore new ways of analysing movement data, and second, the geographical perspective stressed the importance of analysing movement to study the relationship between people and places. In fact, in the proposed approach the places "emerge" from movement and their location and extent are indicated by the MSPs. The places where people suspend their movement play a central role in understanding pedestrian movement. They are more than locations where people stop; they indicate the existence of an interaction between people and the environment.

Samenvatting

Wandelen is een van de belangrijkste activiteiten in natuurlijke recreatiegebieden. Dit gaat verder dan het jezelf simpel verplaatsen van A naar B. Het geeft plezier en zorgt voor een gevoel van welbevinden bij bezoekers die op zoek zijn naar een hechtere band met de natuur. Mensen die door natuurgebieden wandelen hebben een directe band met hun omgeving en interacteren met ruimtelijke verschijnselen. Dit beïnvloed de verplaatsing van mensen door het gebied. Daarom kan verplaatsing van mensen in natuur gebieden bestudeerd worden vanuit een geografische oogpunt, om zo interacties tussen mens en de omgeving waardoor hij zich beweegt, beter te begrijpen. Het begrijpen van deze interacties kan beheerders van natuurgebieden helpen om de uitdagingen, die samengaan met een intensiever gebruik door recreanten van natuurgebieden, aan te gaan. Daarnaast levert het nieuwe inzichten op over het ruimtelijke gedrag van mensen waardoor onderzoekers hun kennis over relaties tussen mens en omgeving vergroten.

Het centrale thema van deze thesis is een exploratieve analyse van deze verplaatsingen van bezoekers van natuurgebieden. Ik benader dit thema door een combinatie van ruimtelijke analyse en “data-mining” technieken te gebruiken, om zo de enorme datasets met verplaatsingsdata, ingewonnen via GPS technologie, te kunnen analyseren. Er van uitgaande dat deze data aanwijzingen bevat voor interacties tussen mens en omgeving, heb ik een werkwijze ontwikkeld om verplaatsingsgedrag van wandelaars zodanig te analyseren dat deze interacties duidelijk worden. Deze werkwijze is gebaseerd op een “knowledgediscoveryprocess” en loopt via een aantal opeenvolgende stappen; van de analyse van de ruwe data tot aan het detecteren en analyseren van verplaatsingspatronen.

De belangrijkste doelstelling van deze thesis is daarom het ontwikkelen van een werkwijze om verplaatsingspatronen van wandelaars in interactie met hun omgeving te kunnen verkennen, analyseren en interpreteren. Om deze doelstelling te kunnen bereiken zijn vier specifieke doelstellingen geformuleerd: a) ontwikkel een werkwijze voor het detecteren van stopgedrag van wandelaars in verplaatsingspatronen, b) evalueer de gedetecteerde verplaatsingspatronen via een gecontroleerd experiment, c) demonstreer de toepasbaarheid van deze werkwijze voor het bestuderen van interacties tussen verplaatsingen van wandelaars en plaatsen in natuurlijke gebieden, en d) ontwikkel een ontologie om relaties tussen patronen van verplaatsingen en gedrag te kunnen representeren.

Hoofdstuk 2 presenteert een benadering voor exploratieve ruimtelijke analyse om “Movement Suspension Patterns”(vertragingspatronen) te kunnen detecteren door gebruik te maken van een lokale index voor ruimtelijke associaties (LISA). Patronen die vertraging van verplaatsing weergeven zijn gebruikt om plaatsen te ontdekken waar wandelaars stoppen als gevolg van interactie met ruimtelijke verschijnselen die gewoonlijk geassocieerd worden met specifieke activiteiten. Dit type patronen laat ons het verplaatsingsgedrag van wandelaars beter begrijpen. De voorgestelde benadering bestaat uit de volgende stappen: a) verzameldata over verplaatsingen van wandelaars via GPS ontvangers, b) bereken de vectoren van verplaatsingen om zo een vector gebaseerd representatie van verplaatsing te verkrijgen, c) bereken de waarden voor de LISA op basis van de snelheid, d) classificeer de resultaten door gebruik te maken van de gemiddelde waarde, en de waarden voor de Z-scores van de LISA en uiteindelijk e) verwaardig een kaart met de geclassificeerde vectoren om zo plaatsen waar vertragingen in verplaatsing optreden te ontdekken. Twee verschillende datasets zijn gebruikt om deze werkwijze te demonstreren: een dataset van spelers van een spel via mobile telefoons en een dataset van bezoekers van een natuurgebied. Het resultaat van beide experimenten toont aan dat vertragingspatronen optreden op herkenbare plaatsen zoals de controleposten in het spel en de verschillende activiteiten en voorzieningen in het natuurgebied. Het resultaat laat zien dat de LISA een geschikte index is voor het exploreren van vertragingspatronen die plaatsen representeren waar de verplaatsing van wandelaars wordt vertraagt door ruimtelijke verschijnselen zoals attracties of obstakels.

Hoofdstuk 3 presenteert de resultaten van een gecontroleerd experiment met als doel het onderzoeken van de validiteit van de vertragingspatronen voor het representeren van stopgedrag van bezoekers van natuurlijke recreatie gebieden. Gedurende het experiment wandelden en stopten de deelnemers, voorzien van GPS apparatuur, op vooraf bepaalde locaties in het nationale park Dwingelderveld. De verzamelde data werd geanalyseerd volgens de methode beschreven in hoofdstuk 2. De nauwkeurigheid van de gedetecteerde vertragingspatronen werd zowel ruimtelijk als temporeel beoordeeld. De gedetecteerde vertragingspatronen werden vergeleken met een set met referentie stops. Het bleek dat 92% van de referentiestops correct gedetecteerd konden worden via de vertragingpatronen, met een vals-positief waarde van = 0.16. De locatie en omvang van plaatsen waar vertragingen optreden zijn berekend als“Percent Volume Contours” (PVC) in een “KernelDensitySurface” en vergeleken met de set van referentiestops. Het bleek dat 96% van de referentiestops binnen de door PVCs gedefinieerde gebieden liggen. Deze resultaten bevestigden dat vertragingspatronen bruikbaar zijn voor representeren van het optreden en de locatie van “stops” in data met verplaatsingsgevens van wandelaars.

Hoofdstuk 4 richt zich op beheerders van nationale parken en onderzoekers en uitvoerders van toeristische activiteiten. Er wordt gedemonstreerd hoe de analyses van verplaatsingspatronen toegepast kunnen worden om geaggregeerde verplaatsingen van bezoekers van natuurlijke gebieden te begrijpen. Dit hoofdstuk presenteert een case studie waarbij het verplaatsingsgedrag van bezoekers in het nationaal park Dwingelderveld werd bestudeerd via een gecombineerde analyse van twee soorten verplaatsingspatronen: “Movement Suspension Patterns

(MSPs)" en "GeneralisedSequentialPatterns (GSPs)". MSPs, als representatie van vertraging in de verplaatsing die optreedt wanneer wandelaars stoppen, zijn gebruikt om plaatsen te ontdekken die interessant zijn voor bezoekers. GSPs representeren de gegeneraliseerde volgorde waarin plaatsen zijn bezocht (onafhankelijk van de gevolgde trajecten) en worden gebruikt om overeenkomsten aan het licht te brengen in de wijze waarop mensen het gebied verkennen. Beide type patronen werden geanalyseerd in een geografische context om de geaggregeerde beweging van bezoekers te typeren en inzicht te geven in voorkeuren van bezoekers en hun interacties met de omgeving.

Hoofdstuk 5 introduceert een kennis representatie voor de belangrijkste concepten van verplaatsingen van wandelaars (i.c. patronen, interacties en ruimtelijk gedrag). Dit is bedoelt als een aanzet voor een formele representatie van deze concepten en de relaties ertussen. De representatie bestaat uit een ontologie waarin zowel concepten als relaties tussen deze concepten formeel worden beschreven. De ontologie is geïmplementeerd in "Ontological Web Language (OWL)" gebruikmakend van het Protégé"framework". Uiteindelijk is er een set met verplaatsingspatronen geïmporteerd in de ontologie om een aantal basis redeneer taken, zoals "instancechecking" en classificatie te testen.

Hoofdstuk 6 bepleit, op basis van de resultaten van voorgaande hoofdstukken, dat het "knowledgediscoveryprocess" toegepast op data van verplaatsingen, een uitvoerbare geografische benadering isom verplaatsing van wandelaars te analyseren. De analyse van verplaatsingen vanuit het geografische perspectief had twee belangrijke implicaties. Ten eerste, maakt de synergie tussen meerdere representaties van verplaatsingen nieuwe manieren voor analyse van data van verplaatsingen mogelijk en ten tweede het geografische perspectief benadrukte het belang van het analyseren van verplaatsingsgedrag om inzicht te krijgen in de relatie tussen mensen en plaatsen. De plaatsen waar mensen vertragen spelen een centrale rol in het begrijpen van verplaatsing van wandelaars. Plaatsen zijn meer dan de locaties waar mensen stoppen; ze duiden op het bestaan van interacties tussen mensen en hun omgeving.

Resumen

Caminar es una de las principales actividades en áreas naturales y recreacionales. Su importancia se extiende más allá del simple objetivo de trasladarse de un punto a otro, pues provee bienestar y placer a los visitantes que buscan una conexión con la naturaleza. Las personas que caminan en éstas áreas se relacionan directamente con su entorno e interactúan con elementos del paisaje que afectan y modulan su movimiento. Por lo tanto, el movimiento de visitantes en las áreas naturales puede ser estudiado desde un punto de vista geográfico para entender las interacciones entre las personas y el entorno en el que se mueven. Entender estas interacciones puede ayudar a los gestores de parques y áreas naturales tanto a enfrentar los retos de un creciente uso de los espacios recreacionales, como a implementar y evaluar medidas de manejo para reducir el impacto de un número de visitantes cada vez mayor. Además, puede proveer nuevos elementos para comprender el comportamiento espacial de las personas, permitiendo así a los investigadores ampliar su conocimiento sobre las relaciones entre las personas y el medio ambiente.

El análisis exploratorio de estas "interacciones de movimiento" de los visitantes de áreas recreacionales es el tema central de esta tesis. El enfoque que propongo consiste en crear una sinergia entre análisis espacial y técnicas de minería de datos para analizar cantidades masivas de datos recolectados con tecnología GPS, asumiendo que estos datos guardan evidencia de las interacciones espaciales entre las personas y el ambiente. Además, presento una metodología para el análisis del movimiento de caminantes para descubrir tales interacciones y mejorar el conocimiento acerca del comportamiento espacial de las personas en áreas naturales. La metodología se basa en un proceso de "descubrimiento de conocimiento" que implica diferentes pasos de análisis desde la observación de datos, la detección de patrones de movimiento, y hasta la representación de conceptos sobre interacciones de movimiento.

El objetivo principal de esta tesis es por lo tanto, desarrollar un enfoque para la exploración, análisis e interpretación de patrones de movimiento de caminantes interactuando con el entorno geográfico. Para lograr este objetivo general, se han propuesto cuatro objetivos específicos: a) desarrollar un método para detectar patrones de movimiento que identifiquen las paradas durante el movimiento de los caminantes; b) evaluar el método a través de un experimento controlado, c) demostrar la aplicabilidad del método para el estudio del movimiento de visitantes en áreas naturales; y d) desarrollar una representación basada en ontologías para los conceptos de relacionados con patrones de movimiento y comportamiento espacial.

Esta tesis está estructurada en seis capítulos. El Capítulo 1 presenta la motivación, los principales conceptos y los objetivos de mi investigación.

El Capítulo 2 presenta un enfoque metodológico para detectar Patrones de Suspensión de Movimiento (MSPs) utilizando un Indicador Local de Asociación Espacial (LISA). Los MSPs detectados se utilizan para encontrar lugares donde los caminantes se detienen como consecuencia de su interacción con elementos geográficos del ambiente, usualmente asociados con actividades humanas específicas, permitiéndonos entender el comportamiento espacial de los caminantes. La metodología propuesta consiste en los siguientes pasos: a) recolectar datos de caminantes utilizando receptores GPS; b) calcular vectores de movimiento para crear una representación del movimiento basada en vectores; c) calcular los valores LISA para la variable de velocidad de los vectores de movimiento; d) clasificar los resultados utilizando la velocidad promedio y el puntaje Z de LISA; y finalmente e) representar los vectores clasificados en un mapa para encontrar los lugares donde ocurren los MSPs. Dos conjuntos de datos de movimiento fueron utilizados para demostrar la metodología: el movimiento de jugadores de un juego de rol urbano en la zona central de una ciudad y el movimiento de visitantes en un parque nacional. Los resultados de los dos experimentos mostraron que los MSPs estuvieron localizados en lugares significativos, como los puntos de control del juego y las diferentes atracciones y servicios en el parque. Los resultados sugieren que LISA es un indicador apropiado para explorar y descubrir los lugares donde el movimiento de los caminantes se suspende debido a elementos geográficos como atracciones u obstáculos.

El Capítulo 3 presenta los resultados de un experimento controlado para investigar la validez de utilizar MSPs para representar el comportamiento espacial de los caminantes cuando se detienen. En este experimento, varios participantes caminaron por diversas rutas y se detuvieron en puntos previamente designados en el Parque Nacional Dwingelderveld (Países Bajos). Los participantes llevaron receptores GPS que automáticamente registraban sus sucesivas localizaciones. Los datos recolectados fueron analizados para detectar MSPs utilizando el método descrito en el Capítulo 2. La exactitud de los MSPs detectados fue evaluada en el espacio y en el tiempo. La ocurrencia de los MSPs detectados fue comparada en el tiempo con un conjunto de paradas de referencia y se estableció que el método fue capaz de detectar el 92% de estas paradas con una tasa de falsos positivos de α = 0.16. Cuando la localización y extensión de los MSPs, calculada utilizando los contornos de porcentaje de volumen sobre una superficie de función de densidad Kernel, fue comparada con un conjunto predeterminado de lugares de referencia se encontró que el 96% de los MSPs estaban dentro de las áreas delineadas por los contornos de porcentaje de volumen. Estos resultados confirmaron que los MSPs son una representación adecuada de la ocurrencia y localización de las paradas de caminantes en datos de movimiento capturados con GPS.

El Capítulo 4 demuestra cómo el análisis de patrones de movimiento puede ser utilizado para mejorar el conocimiento de los flujos de visitantes en áreas naturales recreacionales. Este capítulo presenta un caso de estudio en el cual el movimiento de visitantes en el Parque Nacional Dwingelderveld fue estudiado utilizando el análisis combinado de dos tipos de patrones: Patrones de Suspensión de Movimiento (MSPs) y Patrones Secuenciales Generalizados (GSPs). Los MSPs representan la suspensión del movimiento que ocurre cuando los caminantes se detienen en un lugar y son utilizados para descubrir los sitios de interés para los visitantes. Los GSPs representan la secuencia generalizada en la cual se visitan los lugares (sin importar la ruta que siguen) y son utilizados para descubrir aspectos comunes en la manera en la que los visitantes visitan el área. Los dos tipos de patrones fueron analizados en el contexto geográfico para caracterizar el flujo agregado de visitantes y dilucidar aspectos de las preferencias de los visitantes y sus interacciones con el ambiente. Por lo tanto, este capítulo es de especial interés para gestores de parques e investigadores en turismo en áreas naturales interesados en ampliar o extender los sistemas de monitoreo y manejo de visitantes.

El Capítulo 5 presenta una representación del conocimiento espacial de los principales conceptos sobre movimiento de caminantes, como por ejemplo patrones, interacciones y comportamiento espacial. Esto constituye una prueba de concepto de una representación formal de estos conceptos y las relaciones entre ellos. La representación consiste en una ontología en la cual tanto los conceptos como las relaciones entre ellos son representados formalmente. La ontología es implementada en el Lenguaje de Ontologías para la Web (OWL) utilizando la plataforma Protégé. Finalmente, un conjunto de instancias de patrones de movimiento son importadas en la ontología para experimentar el funcionamiento de algunas tareas básicas de razonamiento automatizado, tales como comprobación de instancias y clasificación.

Finalmente en el Capítulo 6 se argumenta, en base a los hallazgos presentados en los capítulos precedentes, que el proceso de "descubrimiento de conocimiento" en datos de movimiento es un enfoque geográfico adecuado para el análisis del movimiento de caminantes. La perspectiva geográfica del análisis del movimiento tiene dos implicaciones importantes. La primera, es que crea una sinergia de múltiples representaciones del movimiento para explorar nuevas formas de análisis. La segunda, es que la perspectiva geográfica resalta la importancia del análisis del movimiento para estudiar las interacciones entre las personas y los lugares. De hecho, en el enfoque propuesto, los lugares "emergen" como consecuencia del movimiento (o más precisamente de su suspensión), y cuya localización y extensión son detectadas y representadas por los MSPs. Los lugares donde la gente suspende su movimiento juegan un papel central en la comprensión del movimiento de los caminantes. Esos lugares son algo más que meras localizaciones en el espacio, ellos indican la existencia de la interacción entre las personas y el ambiente.

Acknowledgments

This book contains the outcomes of four years of my PhD research. Together with the defence, it represents the last stop in a long path in which personal and scientific challenges, difficulties and satisfactions appeared all the time. However I am lucky because I never walked alone. At every step there were people who encouraged me to keep moving forward, people who provided scientific and technical help, constructive criticism, warm friendship and love. I am extremely grateful to all of them.

I want to express my gratitude to my supervisors Monica Wachowicz, Arnold Bregt and Arend Ligtenberg. Their supervision and experience played a key role in completing this thesis, as well as for the development of my scientific skills. Their combination of experience, vision, patience and scientific rigorousness left a mark on me that will always help me find the way in future scientific ventures.

The first part of this research was carried out in the Technical University of Madrid (Spain) thanks to the support of Miguel Angel Bernabé who offered me the first chance to pursue a PhD. He encouraged me to walk the first steps to build a multidisciplinary scientific network that became crucial during the next stages of my research. Thanks also to all the staff from LatinGEO and the Mercator Research Group.

Several persons were of great help and support during the development of this thesis. I had the privilege of working and publishing with high quality researchers, and I want to express my gratitude to Chiara Renso from CNR (Italy), Maribel Santos and Adriano Moreira from Universidade do Minho (Portugal), and Sytze de Bruin from Wageningen University (the Netherlands). It was also a pleasure to work with Joan Batista Company and Francesc Sardá from Consejo Superior de Investigaciones Científicas CSIC who encouraged me to pursue a PhD. I want to extend my gratitude to the rigorous and helpful, anonymous reviewers of the papers included in this thesis.

Many people helped not only with scientific discussions, but also with support and friendship throughout the length of the trip. Gaby, Dan, Omar, Govi, Iria and Alicia you were of great support during my stay in Madrid. Later, in the Netherlands, science, sports, and scientific discussions were part of the daily life with the fellows from the Centre for Geo-Information. I had the pleasure to share these years with Roberto, Valery, Lucía, Richard, Laure, Sidney, Lukaz, Jochem, Gerd, Titia, Erika, Kim, Rogier, Manos, Arun, Hans, Noriko, Jan, Michael, Jie, and Johannes. I am especially grateful to Matthijs, Horacio, Estelle, Harmke, Malu, Natalia, Debora, and Cris whose support and friendship were as deep as delightful.

During the last seven years I never felt too far from home. The support, help and love from my parents Paúl y María Augusta, my siblings Santiago, Anisa, Juanse y

Bernardo, my grandmother Ruth and all the members of my family provided me with the strength and enthusiasm to follow this path, especially my grandfather Hernán who introduced me in the world of Science Fiction and who was always interested in the advances of this research. Thanks also to the members of my extended family: Pablo, Suco, Elsa, Julio, Andrés, Pablo, Galo, Pedro, David, Sebas, Sebitas, Sebastián, Paúl, Sandy, Monica and all the others who were always there to encourage me to keep walking.

All this time there was always a person who was everything for me: colleague, mate, friend, namesake, advisor, therapist, and subject of love and admiration. Daniela, my beloved wife, you taught me that patience, dedication and effort are the best compass and map for venturing in this life. You are part of this, you are part of me.

About the author

Daniel Orellana was born on April 9th, 1977 in Cuenca, Ecuador. In 1996, he obtained his high school degree at Instituto Técnico Superior Salesiano, and later in 2001 obtained a bachelor degree in Environmental Biology. Convinced of the importance of the relationship between people and environment, he studied a specialisation degree on Geographic Information Systems oriented to environmental management, working at the same time in sustainable development programs at NGOs and the "Universidad del Azuay". In 2004, he moved to Europe with an AECI scholarship and a partial FUDACYT scholarship helped him to obtain a MSc in Geographic Information Systems at the "Universitat Politècnica de Catalunya" with a internship at the "*Consejo Superior de Investigaciones Científicas de España*". Another scholarship allowed him to start his PhD at the "Universidad Politécnica de Madrid" obtaining in 2009 the Diploma of Advanced Studies in geography under the supervision of Miguel Angel Bernabé and Monica Wachowicz. During this period, he found support and inspiration in his colleague and future wife Daniela Ballari. He then moved to Wageningen University to finish its PhD program under supervision of Arnold Bregt, Monica Wachowicz and Arend Ligtenberg. During this period he had the opportunity to collaborate with researchers from Universidade do Minho (Portugal), Consiglio Nazionale delle Ricerche (Italy), and Fraunhofer Institute IAIS (Germany).

Presently, he is appointed as theme leader of Human Systems Resarch Area at the Charles Darwin Foundation in Galapagos, Ecuador, where he conducts research on monitoring the relationship between people and environment in the islands.

List of publications

Peer-reviewed

Orellana, D., Bregt, A. K., Ligtenberg, A. and Wachowicz, M. (2012). Exploring visitor movement patterns in natural recreational areas. *Tourism Management* 33:672-682.

Orellana, D. and Wachowicz, M. (2011). Exploring patterns of movement suspension in pedestrian mobility. *Geographical Analysis* 43(3):241-260.

Moreira, A., Santos, M., Wachowicz, M., and Orellana, D. (2010). The impact of data quality in the context of pedestrian movement analysis. In Painho, M., Santos, M. and Pundt, H., (Eds.) Geospatial Thinking, 61-78. Springer-Verlag.

Orellana, D. and Renso, C. (2010). Developing an Interactions Ontology for Characterising Pedestrian Movement. In Wachowicz, M., (Ed.) Movement-Aware Aplications for Sustainable Mobility, 62-86. New York: Information Science Reference.

Sardà, F., Company, J. B., Bahamon, N., Rotllant, G., Flexas, M., Sánchez, J., Zúñiga, D., Coenjaerts, J., Orellana, D., Jordà, G., Puigdefábregas, J., Sánchez-Vidal, A., Calafat, A., Martín, D., and Espino, M. (2009). Relationship between environment and the occurrence of the deep-water rose shrimp Aristeus antennatus (Risso, 1816) in the Blanes submarine canyon (NW Mediterranean). *Progress in Oceanography* 82:227-238.

Conference Proceedings

Orellana, D., Wachowicz, M., Knegt, H. d., Ligtenberg, A., and Bregt, A. K. (2010). Uncovering Patterns of Suspension of Movement. In *GIScience 2010 extended Abstracts*. Zurich. 14-17th September, 2010.

Orellana, D., Wachowicz, M., Andrienko, N., and Andrienko, G. (2009). Uncovering Interaction Patterns in Mobile Outdoor Gaming. In *Advanced Geographic Information Systems & Web Services, 2009. GEOWS '09*. Cancún. 1 - 7 Feb. 2009.

Andrienko, N., Andrienko, G., Wachowicz, M., and Orellana, D. (2008). Uncovering Interactions between Moving Objects. In *Geographic Information Science, Extended Abstracts of 5th International Conference, GIScience 2008*. Park City, Utah, USA. September 23-26, 2008.

Wachowicz, M., Orellana, D., Renso, C., and Parada, J. (2008). The spatial knowledge representation of players' movement in mobile outdoor gaming. In *Management for Protection and Sustainable Development. Proceedings of the 4th International Conference on Monitoring and Management of Visitor Flows in Recreational and Protected Areas*. Montecatini, Italy. November, 2008.

Bahamon, N., Sardà, F., Company, J. B., Sánchez, J., Orellana, D., Coenjaerts, J., Martín, D., Zúñiga, D., Calafat, A., and Flexas, M. (2007). Non-linear response of deep-sea red shrimp (Aristeus antennatus) catches to environmental forcings in a submarine canyon in NW Mediterranean Sea. In *Cascades, Chains, and Webs: Alternative Approaches to Understanding Complex System. Special Poster Session of the 55th North American Benthological Society Annual Meeting*. Columbia, South Carolina, USA. June 3-8, 2007.

Other publications

Orellana, D. (2011). Dime cómo te mueves y te diré quién eres: La movilidad como huella del comportamiento espacial de las personas. En Vaquero, J.R. *Movilidad: Retos y oportunidades para los profesionales de la información y la comunicación: 32–45*. Plasencia: Fundación Ciencias de la Documentación.

Orellana, D. (2010). Gaia: La historia de un planeta. Ecuador Terra Incógnita N63.

Orellana, D. and Ballari, D. (2009). La GeoWeb y su evolución. Universidad Verdad 49:25-49.

Wachowicz, M., Vazquez, A., Ballari, D., Orellana, D., and Rodriguez de Castro, A. (2009). Human Mobility Patterns: A Source of Geospatial Knowledge. *European Journal for the Informatics Professional* 10:16-24.

Orellana, D. (2006). Modelamiento de Datos Submarinos en el Cañón de Blanes con Sistemas de Información Geográfica. In Conferencia de Usuarios ESRI-España. Madrid. 2006.

Orellana, D. (2001). Gestión ambiental en el biocorredor del Río Tomebamba: de la confrontación hacia la participación. In Participación Ciudadana en la Gestión Ambiental. Red Cántaro. Cuenca, Ecuador. 2001.

Research Blog:

http://ideasonmovement.wordpress.com

PE&RC PhD Education Certificate

With the educational activities listed below the PhD candidate has complied with the educational requirements set by the C.T. de Wit Graduate School for Production Ecology and Resource Conservation (PE&RC) which comprises of a minimum total of 32 ECTS (= 22 weeks of activities)

Review of literature (5.6 ECTS)

- Literature review on spatio-temporal models in geometrics; presented to an evaluation tribunal, Universidad Politécnica de Madrid (2008)

Writing of project proposal (4.5 ECTS)

- Developing a spatial knowledge representation for pedestrian movement based on interactions (2009)

Post-graduate courses (5.2 ECTS)

- Remote sensing for environmental applications; Technical U of Madrid (2007)
- Environmental cartographic models; Technical U of Madrid (2007)
- Spatial data infrastructures; Technical U of Madrid (2007)
- Map publishing on the Internet; Technical U of Madrid (2007)

Laboratory training and working visits (0.9 ECTS)

- Data-mining techniques for clustering pedestrian movement data; Universidade do Minho, Portugal (2009)
- Knowledge representation for movement data; VU, Amsterdam (2010)

Invited review of (unpublished) journal (3 ECTS)

- International Journal of GIS: human movement modelling (2009)
- International Journal of GIS: pedestrian mobility (2010)
- Computers, Environment and Urban Systems: speed profiles (2011)

Deficiency, refresh, brush-up courses (6 ECTS)

- Spatial modelling and statistics (2010)

Competence strengthening / skills courses (3.8 ECTS)

- Publishing scientific papers in English; UPM (2008)
- Scientific writing; WUR (2010)
- PhD Competence assessment; WGS (2010)
- Experiments to innovate during great transformations; WUR (2010)

PE&RC Annual meetings, seminars and the PE&RC weekend (1.5 ECTS)

- PE&RC Introduction weekend (2009)
- PE&RC Day (2009)
- PE&RC Day (2010)

Discussion groups / local seminars / other scientific meetings (3.3 ECTS)

- GeoPKDD Workshops and scientific meetings (2008-2009)
- Maths and Stats discussion group (2009)
- Spatial Methods discussion group (2009-2010)

International symposia, workshops and conferences (8.5 ECTS)

- Presentation in Conference: Monitoring and Management of Visitor Flows in Recreational and Protected Areas (2008)
- Presentation in Conference: Advanced in Geographic Information Systems and Web Services (2009)
- Presentation in Conference: AGILE (2010)

- Presentation in Conference: GIScience (2010)
- Movement Patterns Analysis Workshop (2010)

Lecturing / supervision of practical 's / tutorials; 10 days (3 ECTS)
- Graduation course: spatial analysis tools and applications; by invitation from Universidad del Azuay, Ecuador (2007)

Supervision of 2 MSc students; 10 days (6 ECTS)
- Visualization of suspension patterns in human movement; MSc. Thesis: Anastasia Petrenko, Geo-information, WUR (2010)
- Analyzing the temporal dimension of movement suspension patterns on pedestrian movement; MSc. Thesis: Michail Marinakis, Geo-information, WUR (2010)