Classifying Pedestrian Movement Behaviour from GPS Trajectories using Visualisation and Clustering*

Gavin McArdle[1], Urška Demšar[2], Stefan van der Spek[3], and Seán McLoone[4]

[1] National Centre for Geocomputation, National University of Ireland Maynooth, Maynooth, Co. Kildare, Ireland
gavin.mcardle@nuim.ie
[2] Centre for GeoInformatics, School of Geography and Geosciences, University of St Andrews, Fife, Scotland, UK
urska.demsar@st-andrews.ac.uk
[3] Department of Urbanism, Faculty of Architecture, Delft University of Technology, Delft, the Netherlands
s.c.vanderSpek@tudelft.nl
[4] Department of Electronic Engineering, National University of Ireland Maynooth, Maynooth, Co. Kildare, Ireland
sean.mcloone@eeng.nuim.ie

Abstract. The quantity and quality of spatial data are increasing rapidly. This is particularly evident in the case of movement data. Devices capable of accurately recording the position of moving entities have become ubiquitous and created an abundance of movement data. Valuable knowledge concerning processes occurring in the physical world can be extracted from these large movement data sets. Geovisual analytics offers powerful techniques to achieve this. This article describes a new geovisual analytics tool specifically designed for movement data. The tool features the classic space-time cube augmented with a novel clustering approach to identify common behaviour. These techniques were used to analyse pedestrian movement in a city environment which revealed the effectiveness of the tool for identifying spatiotemporal patterns.
Keywords Geovisual Analysis, Clustering, Space-time Cube, Movement Data Analysis

1 Introduction

Moving objects are a common phenomena in the physical world. While moving objects include a variety of entities, in general terms, each moving object can be described as an entity whose position and/or geometric attributes change over time (Dodge et al., 2009). The ability to study moving entities has improved

* This is a preprint version of an article published in Annals of GIS. Please cite as follows: McArdle, G., Demšar, U., van der Spek, S., & McLoone, S. (2014). Classifying pedestrian movement behaviour from GPS trajectories using visualization and clustering. Annals of GIS, 20(2), 85-98.

due to the availability of positioning and tracking technologies. In many cases large data repositories of moving entities have been created. Although different technologies are used to collect movement data, the same fundamental process is involved: how objects move through the basic framework of our physical world, defined by geographic space and time.

The sheer volume of movement data creates opportunities to acquire knowledge about moving entities and how they interact with their environment. When this knowledge is related to human movement, it can be useful in location based services (Millonig and Gartner, 2011), emergency management (Zheng et al., 2009), traffic management (Castro et al., 2012), health (Landau et al., 2009) and urban planning (Van der Spek, 2010). In urban environments, we can gain particularly valuable knowledge from data on pedestrian movement. We can use information on the areas visited, the routes chosen and the time spent in different spaces to measure the vitality of a city. We can also estimate the economic value of interventions in the city landscape.By analysing the movement of individuals, we can identify repetitive behaviour, common routes and activities. We can use these patterns to plan modifications to the urban structure, such as new bridges or new pathways to assist individuals to reach their target locations in the optimal way. By considering the patterns of many individuals, designs which address the needs of the collective population can be produced, which is a key goal of urban design (Van der Spek, 2010; Van Schaick, 2011).

This article describes a geovisual analytics tool for exploring movement data that links three-dimensional visual representations, a virtual globe and new trajectory data mining algorithms. Spatial data is traditionally visualised in two dimensions on a static map, however, by utilising the third dimension, additional variables, in this case time, can be represented simultaneously adding extra layers of information into a single visualisation. The tool contains novel techniques to assist analysts in this respect by visually representing both the temporal and spatial aspects of movement. A Space-Time Cube (STC) visualisation, which allows us to identify patterns in movement, such as main routes and frequent stopping locations, is a central element of the tool. Within this approach, groups of similar trajectory patterns are identified using speed and acceleration in orthogonal directions as a proxy for behaviour. A case study using pedestrian data collected in the city of Delft in the Netherlands demonstrates these techniques.

The remainder of this article is organised as follows; the next section provides some related work in the area of movement data visualisation and analysis with a specific emphasis on pattern detection. Section 3 details the novel visualisation and analysis tool. A case study, describing how the tool is used to analyse pedestrian data is presented in section 4. The article concludes with a discussion and proposes future work in section 5.

2 Visualisation and Analysis of Trajectory Data

This section examines techniques used to represent, interpret, analyse and compare movement data. Common approaches used to visualise trajectories and their interaction with the underlying geographic space are discussed.

2.1 Trajectory Representation and Visualisation

The path of a moving object is usually represented as a trajectory which is a sequence of positions in a two-dimensional (2D) geographic environment with time stamps (Laube et al., 2005), i.e. $T = (x_1, y_1, t_1), ..., (x_n, y_n, t_n)$ for some n, such that (x_i, y_i) is the measured geographic location of the moving object at time t_i. Often trajectories have associated attribute data, which can be static if the attribute has the same value for the object regardless of its position, (e.g. object type) or dynamic if the attribute changes over time (e.g. speed and acceleration). Other interpretations conceptualise movement as a sequence of moves and stops (Ashbrook and Starner, 2003), or activity chains (Wilson, 2008). Dodge et al. (2008) propose a taxonomy for movement patterns, which includes two categories of patterns. Firstly, there are spatiotemporal patterns such as convergence, divergence and concurrence in space and time and secondly there are behavioural patterns including pursuit, play and flocking activities. These patterns can be used to classify trajectories.

For visual analysis, the complete trajectory can be represented on a 2D map, however, this removes the temporal aspect, as the sequence and relative times that locations were visited is not usually shown. While animation can address this issue, it is often not possible to interpret the entire spatiotemporal extent of trajectories using this technique. A series of separate maps representing distinct time periods, described by Andrienko and Andrienko (2008), can show the progression of moving entities, however a large number of maps are required to analyse data at a high temporal scale.

Trajectory-oriented visualisations such as the STC, which show complete trajectory paths, are valuable when studying movement patterns (Hägerstrand, 1970; Kraak, 2008). In a STC, the geographic coordinates appear on the x-axis and y-axis of the cube, while time is represented along the orthogonal z-axis. Time increases along this axis so that more recently occurring events appear above older events. Using this approach, a trajectory can be plotted as a three-dimensional (3D) polyline in the STC. In time geography, a trajectory can also be referred to as a space-time path and in this context, the STC represents the coordinate system in which the curves representing space-time paths are plotted.

A STC is suitable for simultaneously visualising a relatively low number of trajectories. When that number increases, the STC suffers from occlusion and cluttering, making it difficult to identify trends (Andrienko et al., 2007). Usually aggregation, for example, using kernel densities (Demšar and Virrantaus, 2010) or generating clusters of the most salient patterns (Andrienko and Andrienko, 2010) can resolve this. We extend the STC into a web-mapping tool and add a new approach for clustering trajectories based on the projection of trajectories on to the sides of the STC, as described in section 3.

2.2 Trajectory Similarity

In this section we review several existing methodologies for identifying similarity between trajectories. While this is a general overview, the details of our particular approach are explained in section 3.4.

Searching for patterns and trends within a dataset is an important exploration task. Identifying similarities between the attributes of objects and using this as a basis to group them is a common approach. The challenge is to determine a suitable attribute to analyse along with an appropriate similarity measure and several common metrics are discussed below.

Generally the geometric attributes of trajectory coordinates are used when examining movement data for similarity. This involves comparing the 2D space which defines a trajectory. Several approaches have been widely adopted. Global measures of similarity consider all points within a trajectory, while local measures utilise properties of sub-trajectories to determine similarity (Zheng and Zhou, 2011). Euclidean distance is a commonly used metric which produces a similarity score by summing the straight-line distance between all points in the trajectories being compared. Hu et al. (2007) extend this principle by using the average Euclidean distance between all trajectory points as a similarity score while the Hausdorrf metric (Zhang et al., 2006) only considers the largest of all the distances from a point in one trajectory to the closest point in another trajectory. The Euclidean distance is not always effective when comparing trajectories in the physical world due to the effects of the curvature of the earth. Edit Distance on Real Sequence overcomes this by taking a different approach. A similarity score is produced by summing the number of insert, remove and replace operations required to transform one trajectory into another trajectory (Chen et al., 2005). Additional penalties can be applied according to the type of operation. The Longest Common Sub-String (LCSS) metric is widely used in bioinformatics and has also been used for trajectory comparison. A similarity score is generated based on the length of the longest common pattern between two or more trajectories (Vlachos et al., 2002). Other measures examine the degree of overlap between the minimum bounding rectangles (MBRs) which surround complete or sub-trajectories (Elnekave et al., 2007).

Several similarity techniques which incorporate the temporal aspects of movement data have emerged. For example, a modified Hausdorrf measure has been proposed which preserves the sequential ordering of trajectories (Atev et al., 2006). Other approaches, such as the Frèchet Distance Metric, described by Buchin et al. (2006), take the location and ordering of trajectory points into account when determining similarity. Similarly, Dynamic Time Warping (DTW) techniques (Sakurai et al., 2005) incorporate the temporal aspects of trajectories into similarity measures. DTW stretches the time axis in order to identify similarities in trajectory shape. This allows a comparison of trajectories which span different time frames and spatial regions.

Some techniques have been exclusively designed for comparing geographic trajectories. For example Common Destination uses the end locations of trajectories as an indication of trajectory similarity. Route Similarity computes the mean distance between corresponding points in trajectories but also introduces a penalty distance for unmatched points (Andrienko et al., 2007). Other research focuses on comparing the stopping locations within trajectories (Ashbrook and Starner, 2003). This approach has also been extended to consider places where entities move slowly as significant elements of a trajectory (Palma et al., 2008).

While the above similarity measures focus on comparing the mathematical similarity between coordinates of the trajectories, other methods consider additional attributes of movement such as speed, acceleration, duration and direction (Dodge et al., 2009). The research presented in this article follows this trend. We augment a geovisual environment, based on the STC, with a new technique for identifying similar trajectories. This approach identifies similar movement behaviour by comparing the speed and acceleration of trajectories at each observation point.

2.3 Trajectory Clustering

After an appropriate similarity metric has been determined, objects which are deemed to be similar can be grouped together using a process known as clustering. Clustering is an unsupervised classification technique in which objects or data points are placed into clusters. Cluster assignment occurs so that objects within a cluster are more similar to each other than to objects placed in another cluster. The clusters can then be categorised using appropriate labels.

Many clustering techniques have been developed, each offering tailored improvements for specific cases. There are three classic approaches for clustering datasets, namely, hierarchical (Jain et al., 1999), partitioning (Kanungo et al., 2000) and density based (Kriegel et al., 2011). Within these categories, several algorithms such as DBSCAN (Ester et al., 1996) and OPTICS (Ankerst et al., 1999) have been developed specifically for trajectory data. The decision on which technique to utilise is a combination of the input data, the processing time and storage requirements.

Clustering algorithms suffer from performance and quality issues if the dimensionality of the dataset is large. Morris and Trivedi (2009) describe the application and performance of several similarity measures and clustering algorithms for trajectory data. In this study, we describe a trajectory in terms of the speed and acceleration at each observation point. Therefore the number of attributes is dependent upon factors such as the recording interval and the distance travelled. As a result, we utilise spectral clustering (Song et al., 2008), a partitioning algorithm which is effective for high dimensional data.

Spectral clustering is a hierarchical graph-based clustering technique. A graph is constructed by calculating a similarity matrix $S = s[ij]$ between all the objects in the attribute space. In our approach, each data point is a trajectory and the attribute space for clustering consists of speed and acceleration values at each trajectory point. Euclidean distance in this attribute space is used to determine similarity between two trajectories. An adjacency matrix S of the similarity graph, $G = (V, E)$ is produced, where V is the set of all data points (trajectories). This matrix describes pairwise similarity between two trajectories. Two vertices i and j in the graph are connected if the similarity between them (s_{ij}), is larger than a given threshold value. With this approach, the complexity of the original similarity matrix is reduced as only significant similarities are considered. Clustering then involves partitioning the similarity graph into k groups which can be determined through domain knowledge or analysis. Clusters are formed so that edges between groups have a low similarity score while edges within groups have a high similarity score.

Spectral clustering has previously been applied to trajectory analysis (Atev et al., 2010; Morris and Trivedi, 2009) where the similarity measure was based on trajectory location. This is in contrast to our approach where we use use spectral clustering with a similarity metric based on behaviour. Section 3.4 provides a complete description of the trajectory analysis techniques which we utilise in our approach.

3 Trajectory Exploration Environment

The environment introduced here uses the STC as the principle visual analysis component. To determine similarity between trajectories, the speed and acceleration attributes are used as they act as a proxy for movement behaviour. The similarity of trajectories is therefore determined by the mathematical similarity of their curves, which can be described in terms of the first and second derivatives. Since we are dealing with movement, these two derivatives are equal to speed and acceleration at each moment. We use spectral clustering to group trajectories exhibiting similar movement behaviour. The visual analysis techniques along with the similarity and clustering metrics are discussed below.

3.1 Web-based Space-time Cube

The visualisation component consists of a STC draped over a virtual globe (Google Earth), which provides a pseudo 3D view of the Earth. In the virtual globe, the latitude and longitude dimensions of the globe represent the spatial aspects of the Earth while the third dimension, z, which is perpendicular to the latitude and longitude axes, represents altitude. We manipulate altitude to represent time by placing a STC over the elevation surface. Trajectories therefore appear as lines floating above the surface of the Earth.

The z-axis combines two values: altitude and time. Depending on the terrain, the trajectories' position can be placed relative to ground level or relative to the underlying terrain. For example, in the case study in section 4, the underlying terrain is flat and therefore visualising the trajectories relative to ground level is effective. In mountainous regions, visualising trajectories in this way causes trajectories to bisect mountains, making analysis difficult. Therefore, in mountainous regions trajectories should be placed relative to the altitude of the terrain which needs to be considered during analysis.

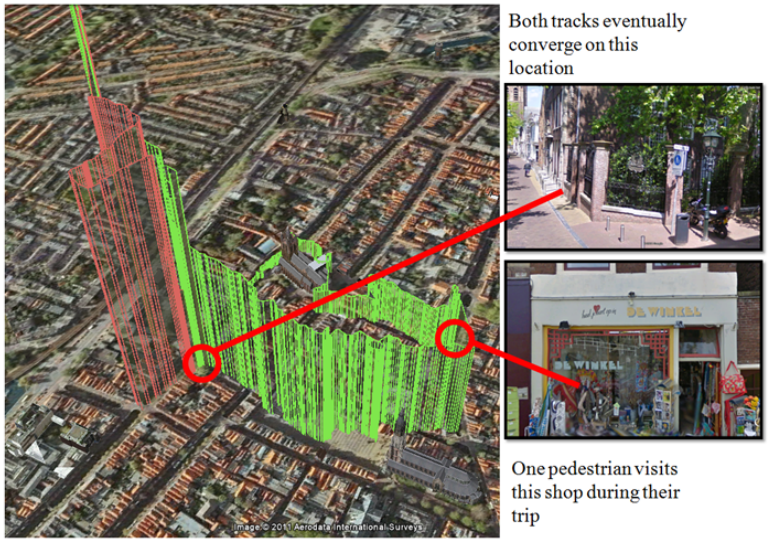

Fig. 1. The STC showing two trajectories with spatiotemporal concurrency. *Background aerial photography: Copyright Google, Aerodata International Surveys (2013).*

Figure 1 shows our visualisation interface of a STC superimposed on a virtual globe. The *fence* structures visible on the surface of the Earth represent trajectories. The points on the top of the *fences*, connecting to the ground along each trajectory are the points where the entity's location was recorded. The height of each point on the top of the *fence* is determined by the observed time of the corresponding trajectory point. Colours distinguish individual trajectories. The user can change the viewing angle, zoom level, interact with, query and identify spatial content. Using Google Earth, the analyst can take advantage of Google Street View and Google Maps to obtain additional spatial information about the study area. Through visual exploration, the analyst can obtain new information about the trajectories being studied. For example, if the observations occur at fixed time intervals, it is possible to visually gauge the speed of entities by measuring the distance between the observations. For faster moving entities the spatial gap between observations is larger. Similarly, the progress of entities using the same route can be compared by examining the distance from the Earth's surface to the polylines representing the entities at specific locations.

3.2 Stop Locations

In certain scenarios, identifying stopping positions is important (Palma et al., 2008). In order to assist analysts with this task, a component for determining and highlighting these stop locations in individual trajectories is included in the tool. As depicted in Figure 2, each trajectory is processed to identify sequential observations occurring in close spatial proximity. The temporal element is then considered. If the time difference between the first and last location of closely co-located recordings is larger than a given threshold then the area is considered a stop location. Visually, in the STC, a rectangular prism is placed at the centroid of the stopping position. While the base of the rectangular prism is placed at ground level, its height is determined by the duration of the stop (taller for longer stops). This is illustrated in section 4.3. Clustering can also be applied to this point dataset to identify common stopping locations.

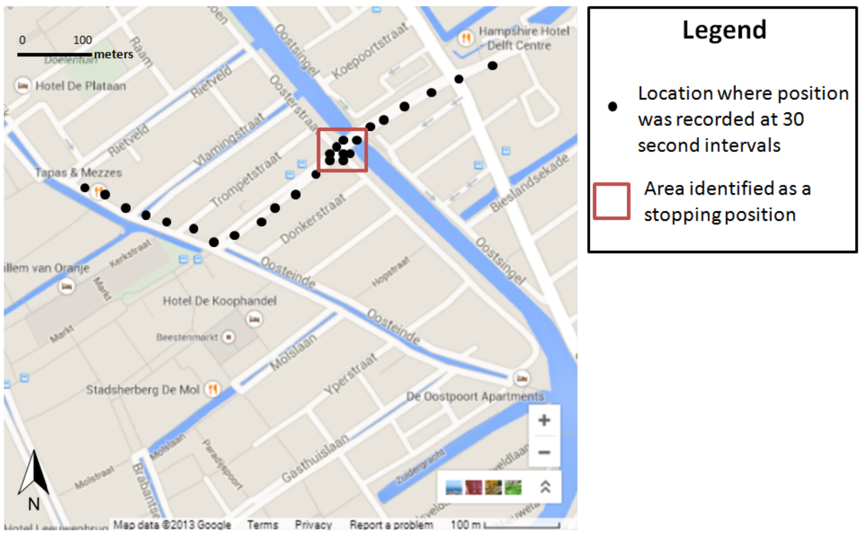

Fig. 2. An area where a stop occured is identified due to the absense of significant movement for a period of 4 minutes.

3.3 Physical decomposition of movement in the space-time cube

We are interested in using speed and acceleration as a proxy for behaviour and so it is necessary to analyse complete trajectories. The path of a moving object in the STC is approximated as a series of straight-line segments between sampled points (x_i, y_i, t_i) and $(x_{i+1}, y_{i+1}, t_{i+1})$ (Laube et al., 2005). Each trajectory is therefore represented by a continuous 3D polyline. We combine this representation with a principle from classical mechanics concerning the physical description of motion in an inertial frame of reference (Halliday et al., 1997). In classical mechanics, the inertial frame of reference is represented by a 3D Cartesian coordinate system located at an arbitrary starting point in space (denoted by 0) in which the object, that is not under the effect of any force, either appears to be at rest or in a state of uniform motion in a straight line. The position of the object is described with a position vector (x, y, z) stretching from the starting point 0 to the object. This vector is a function of time and for modelling purposes is decomposed into perpendicular directions of the coordinate system axes by projections onto relevant planes (x-z, y-z and x-y). Calculations describing various properties of movement are then completed separately for each plane defined by two of the axes.

We use this concept of decomposition and adapt it for the STC. We project each trajectory onto the x-t and y-t planes. Note that t here replaces the z in previous paragraph, as we use the STC rather than an arbitrary Cartesian coordinate system. These two projections are represented as two separate graphs that show how each object moves in one of the two geographic directions (figure 3). This approach is similar to the projection of time series data on to arbitrary planes developed within the GeoTime community (Kapler and Wright, 2005). In our case, decomposition is along the axes of latitude and longitude. It could be argued that this is an arbitrary choice, however, geographic data have been represented using the latitude/longitude or east/north axes for centuries.

The aim of the projections is twofold. Firstly, the projected graphs can be used to visually investigate similarities of movement in each coordinate direction. To enable this in the geovisual analytics environment, the projected graphs are linked to each other and to a STC representation of trajectories through colour brushing. The second purpose of the decomposition is to serve as the basis for calculating the similarity of shapes of trajectories in each projection plane. This is done using a mathematical comparison between the shapes of projected curves, as described in the next section.

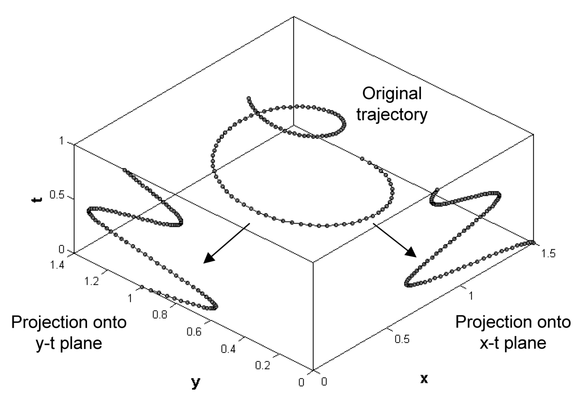

Fig. 3. Trajectories in a STC decomposed and projected on the x-t and y-t planes.

3.4 Trajectory Analysis

To analyse the similarity of trajectories, we initially consider them as time series data (Altiparmak et al., 2006). Each time series is considered as a mathematical curve in a two-dimensional coordinate space with time on the x-axis. Two time series are of a similar form if the slopes of their curves are similar to each other at each moment in time. Here we use the term 'slope' in the mathematical sense, i.e. as the rate of change in values of a function in a mathematical graph. Slopes are calculated as the first derivatives on time at each moment of each time series.

A trajectory can be considered as a two-dimensional time series, since the response on each trajectory at each moment in time is a two-dimensional geographic location. After projecting the trajectories onto the two sides of the STC we calculate the slopes at each time by a numerical approximation of the first derivative. We also calculate the rate of change in slope at each time, which is the second derivative. Since we are dealing with movement and the length along each of our time series is the path covered from the starting point, we can consider the analogous meaning of these two derivatives from physics. There, the first derivative represents speed and the second derivative the acceleration of movement at that particular moment in time.

Similarity of trajectory curves is calculated by clustering the dataset which describes the shape of each projected trajectory and which consists of the speed and acceleration at each time. The first step is to invert the axes in each projection graph to form mathematical functions of the variable, where at each value of the independent variable there should be at most one value of the dependent variable. This is illustrated in figure 4a, where there are three values of the dependent variable t on the curve for one value of the independent variable x. Therefore it is necessary to invert the axis (figure 4b) to consider x as a function of t, i.e. $x = f(t)$

Similarity of two curves $x = f_1(t)$ and $x = f_2(t)$ in the t-x plane is then based on their slopes and rate of change in slope. Mathematically, the slope of a function $x = f(t)$ at a certain point t is represented as the slope of the tangent to f at that point, which is calculated as the first derivative at that point, i.e. slope $S = df(t)/dt$. Since we derive over time, in physical terms this is equal to the speed of movement in x direction. The rate of change in slope is calculated as the second derivative of the function, i.e. change in slope $C = df^2(t)/dt^2$. In physical terms, this is acceleration in the x direction. The shapes of two curves, $x = f_1(t)$ and $x = f_2(t)$, are more similar the more the slopes of the two curves (i.e. the derivatives) and the rate of change in slopes of the curves (i.e. the second derivatives) are similar to each other at each point t. In practical calculations of this similarity, we approximate derivations numerically with slopes of line segments starting at time t. Figure 4c shows this principle. There, for the given t, slope of function $f_1(t)$ (red) is very different to slopes of $f_2(t)$ (blue) and $f_3(t)$ (green), which are similar to each other. We repeat this comparison for all possible times as well as for the rate of change in slope at each time. The algorithm for calculation of the similarity is given by the following pseudocode:

1. Take a dataset of trajectories, given as sequences of (x_i, y_i, t_i) points and project them onto two planes.
2. Take one projected dataset, given as time series of points (t_i, x_i).

a) b)

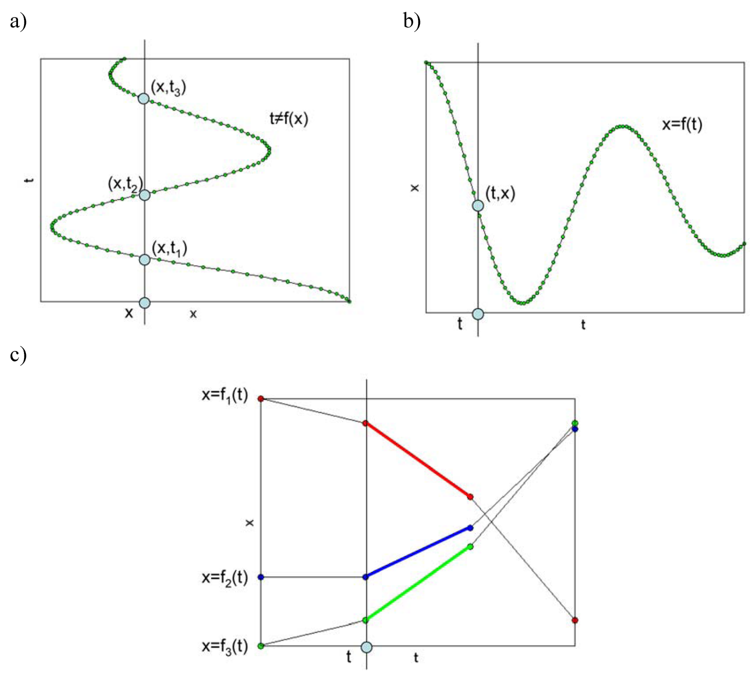

Fig. 4. a) Projection of the original trajectory in the x-t plane. b) Inversion of the coordinate system axes. c) Similarity in slopes of the lines.

3. Subsample times t_i to the union of all times across all trajectories in this dataset. This results in n different times, where n is the maximum number of observations in a single trajectory in the dataset.
4. Calculate missing x_i values on each trajectory using linear interpolation.
5. For each time t_i from t_1 to t_{n-1}, calculate the first derivative = slope of line segment between points (t_i, x_i) and (t_{i+1}, x_{i+1}).
6. For each time t_i from t_2 to t_{n-1}, calculate the second derivative = change in first derivatives between points (t_{i-1}, x_{i-1}) and (t_{i+1}, x_{i+1}).
7. Repeat calculation of derivatives in the second projection (steps 2-6), i.e. in time series of points (t_i, y_i).
8. Cluster the dataset of first and second derivatives from both projections.

Note that as a generalisation of this concept it is possible to select either direction only (i.e. x or y) or both (x and y) for the resulting dataset, depending on the type of movement space that the trajectories originate in. For example, if the original trajectories are sampled from entities moving in a block-like city centre (e.g. Manhattan), then looking at each coordinate axis makes sense. If however objects move in a more unconstrained manner (e.g. hurricanes), then a joint dataset of derivatives in both x and y directions would be a better option.

Based on this principle, curves can be clustered into groups of similarly-shaped curves using the dataset of the first and second derivatives over time at all sampling times. One of the benefits of this approach is its disregard of the actual values on the curves (i.e. particular locations). This means the result is independent of the physical location where the movement data was recorded. This allows analysts to focus on the behaviour and activities of the moving object, rather than focusing on its location and permits comparison of similar movement behaviour in different places. Furthermore, by normalising the temporal components, it is not necessary for the trajectories to be from the same time period.

In the last step, spectral clustering (Song et al., 2008), specifically the variation developed by Chen et al. (2011) is performed on the decomposed trajectories in order to group those exhibiting similar movement behaviour.

The groups resulting from the spectral clustering process, described in section 2.3 are then visualised to explore the particular movement behaviour represented by each cluster. Firstly, the trajectories, which form each group, are plotted. The resulting graphs can explain the groupings and help apply domain knowledge to classify entities based on the spatiotemporal movement patterns. The information is linked to the STC visualisation where individual clusters can be displayed for further visual analysis. The next section presents a case study which applies these techniques to a real dataset of pedestrian trajectories.

4 Case Study

We apply our techniques to analyse pedestrian movement in the city of Delft, the Netherlands in order to demonstrate the effectiveness of the approach. Emphasis is placed on showing the knowledge gained through the STC, decomposed projection plots and the associated clustering procedure. The knowledge is used to infer how pedestrians utilise the city.

Data were collected over a four day period by distributing GPS receivers to pedestrians at car-park facilities in Delft city. The dataset consisted of 113 trajectories, with the sampling rate of five seconds. Data were cleaned and pre-processed and problematic data points were removed prior to analysis. The study area consists of an old city centre with pedestrian streets bordering canals and open town squares. The study area is approximately $4km^2$. From the origin, each participant made one journey on foot, represented by a single trajectory, from the origin, including stops, back to the origin which marks the end of a trip. Therefore, a trip can have multiple destinations and activities which are not known in advance.

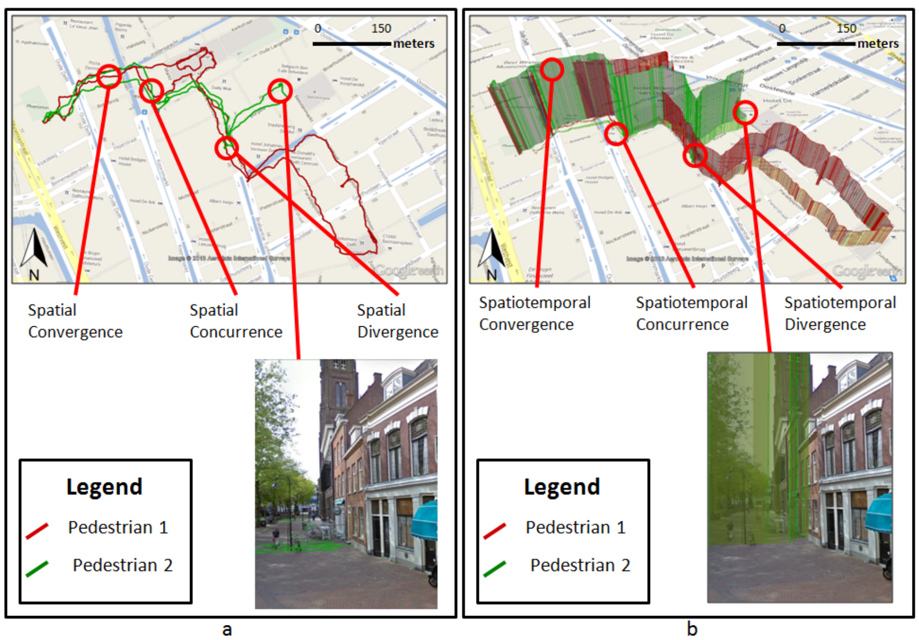

Fig. 5. Using 2D (a) and 3D (b) visualisations to display the path taken by 2 pedestrians in a city.

4.1 3D vs. 2D

We used the tool to identify temporal patterns in movement, such as the ones from the taxonomy by Dodge et al. (2008). We were interested in detecting objects that move together for a period of time, then split, then later joined together again. Many visualisation methods for trajectories are linked to two dimensions only and so the temporal characteristics of trajectories are only implied through location. In figure 5, we show an example of how a 3D STC visualisation can help identify spatiotemporal patterns that are invisible in 2D.

4.2 Route Analysis

The cleaned and validated set of trajectories representing the paths of 113 anonymous pedestrians was used to create the initial visualisation (figure 6) which highlights the spatial and temporal extent of the pedestrians' journeys. The data were collected in an area of flat terrain, so trajectories are placed relative to ground level. Each trip starts at ground level (time 0) which permits progress of individual pedestrians to be assessed.

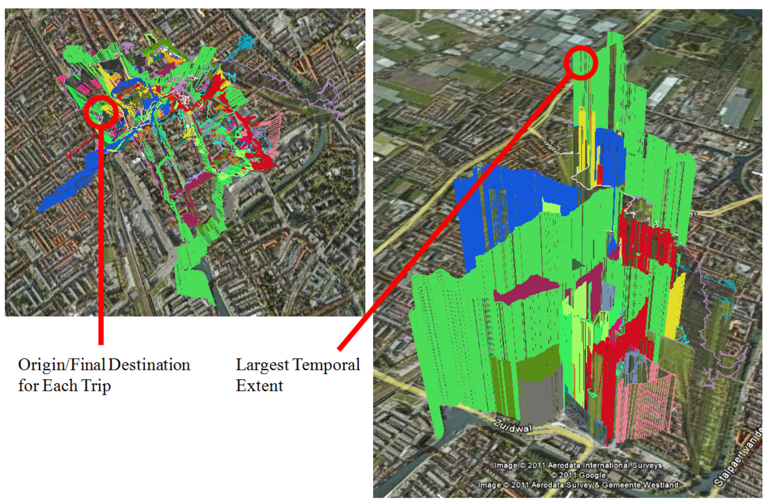

Fig. 6. All trajectories, viewed from above and from the side in the STC. *Background aerial photography: Copyright Google, Aerodata International Surveys (2013).*

The virtual globe facilitates interaction by zooming and panning. Individual trajectories can be queried, highlighted or removed from the display area which enables common spatial routes to be identified. For example, it can be seen that the majority of pedestrians travel east, while very few cross the barrier caused by railway tracks west of the origin. Temporal analysis, achieved by examining the height of the trajectories on the time axis, shows that pedestrians generally travel in an anti-clockwise direction, visiting areas to the south of the origin first before regions, which are parallel with, or north of the origin. Domain knowledge indicates that pedestrians therefore visit high street stores in the south of the city before visiting boutiques which are located in the northern area.

The inclusion of the temporal aspects of trajectories in the visualisation environment helps identify spatiotemporal concurrence which represents pedestrians travelling together. This is visually represented as two trajectories that follow the same curve through the three-dimensional space of the STC. For example, trajectories which follow similar spatial and temporal routes are likely to be related. Examining figure 5(b), we can see two trajectories which initially follow this pattern. These two trajectories commence at the same absolute time, therefore we can make the following comparison. Midway through the trip, there is spatiotemporal divergence; one pedestrian continues walking while the other pedestrian stops at a single location. By examining the spatial content at the stop location, also shown in figure 5(b), it appears to be an office; a long stop here could indicate a meeting or an appointment. The ability to detect the spatial feature in this way demonstrates the benefits of using Google Earth. Later, both trajectories exhibit spatiotemporal convergence and concurrence and return to the origin at a same time. Such behaviour suggests that these pedestrians are travelling together.

A further example of this can be seen in figure 1 where the spatial and temporal concurrence occurs at the end of a trip. One pedestrian goes directly to a specific location which has been identified as a meeting venue in Delft. Meanwhile a second pedestrian makes a separate trip around the city, stopping at one store, again identified using Google Street View in figure 1, before going to the same location as the the first pedestrian. The duration of the stop at each location can be identified to obtain a further insight into the pedestrian behaviour at certain locations. Both pedestrians leave the meeting venue simultaneously and return to the car park facility. This pattern of spatiotemporal concurrence gives an indication of the social connections in the city.

Common locations where people frequently stop provide evidence of areas of specific importance (Palma et al., 2008). Stopping locations are determined by the frequency of observations at the same location (based on a temporal threshold), as described in section 3.1. Figure 7 shows all individual stopping locations of pedestrians in this case study. Notable stopping locations are the two town squares where there are cafés and restaurants. The height of the rectangular prism indicates the duration of a stop which can distinguish a long stop at a restaurant from a short stop to a shop. The colours of the prisms correspond to individual pedestrians and there are multiple stopping locations on each route. Figure 7 also identifies the main shopping district based on the frequency and duration of stops which are indicative of people entering and leaving stores.

4.3 Cluster Analysis

Interactive graphs show the decomposed trajectories and are used to identify similar behaviour among trajectories in each direction. As described in section 3.4, clustering is applied to the trajectories described in terms of speed and acceleration at each observation point, in each direction. Visual analysis, which involved the analyst examining the contents of clusters to ensure a balance be-

Fig. 7. The STC showing all identified stopping locations. *Background aerial photography: Copyright Google, Aerodata International Surveys (2013).*

tween sparse clusters and overcrowded clusters suggested that 6 clusters produce descriptive results without overlap or gaps.

Figures 8 - 13 show the graphs describing these clusters. Each graph has the same temporal extent to exemplify the differences between trip durations in each cluster. Time is represented on the x-axis while the location in either the longitude or latitude projection is shown on the y-axis. Each cluster is represented by two graphs (latitude and longitude). Figure 8 and figure 9 relate to cluster 1 and show the difference between the latitude and longitude projections of the cluster.

Both the latitude and longitude data were used in our analysis, however, for the purpose of presentation and limitations in the paper length, only a sample of graphs are shown here. There are two main characteristics visible in the graphs that define the separation of clusters (table 1). One of these characteristics is the duration of the trip, i.e. the length of the trajectory. Short trajectories are grouped together, medium-long trajectories are also grouped together and finally long trajectories are clustered together. The other distinguishing characteristic is the number and duration of stops on each trajectory. Stops can be visually identified as small oscillations around the horizontal baseline. These oscillations occur when the person carrying the GPS receiver enters a building causing the receiver to lose the GPS signal. A simultaneous loss of signal in both the latitude and longitude projections represents a stop. The algorithm separates clusters with different numbers and durations of stops into several groups, as per table 1. The decision to classify a trip or a stop as long, medium or short is determined by dividing the longest trip/stop length into thirds and assigning labels accordingly.

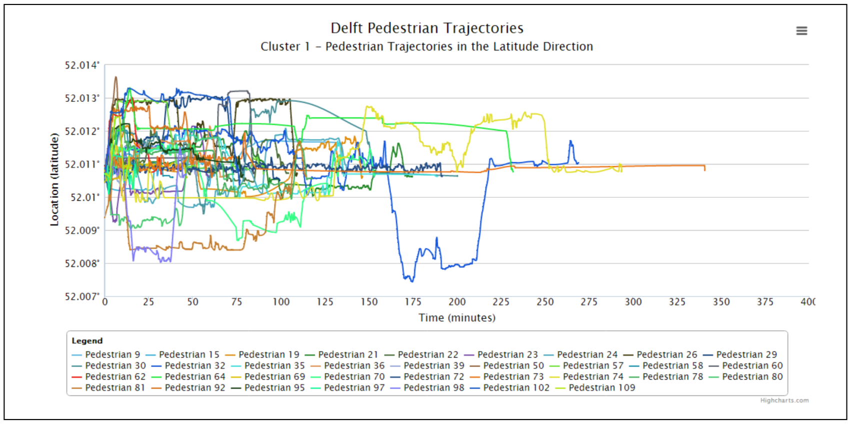

Fig. 8. Trajectories from cluster 1.

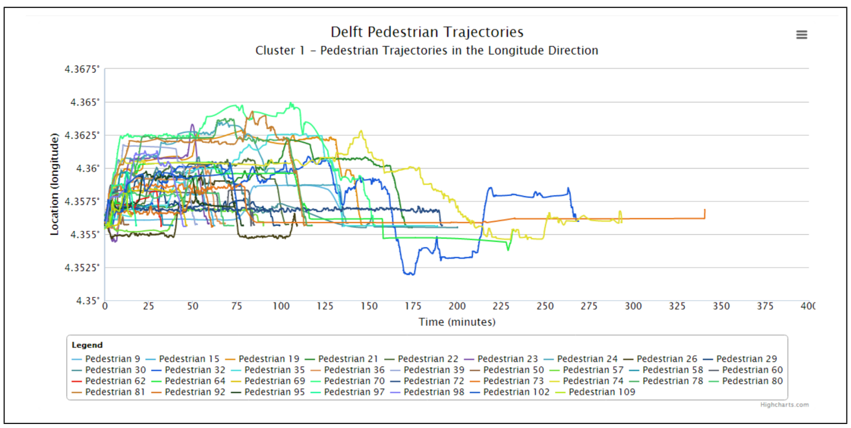

Fig. 9. Trajectories from cluster 1.

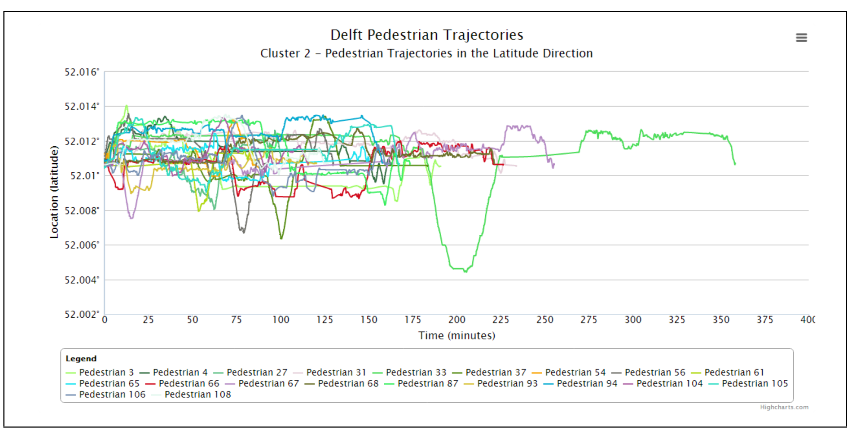

Fig. 10. Trajectories from cluster 2.

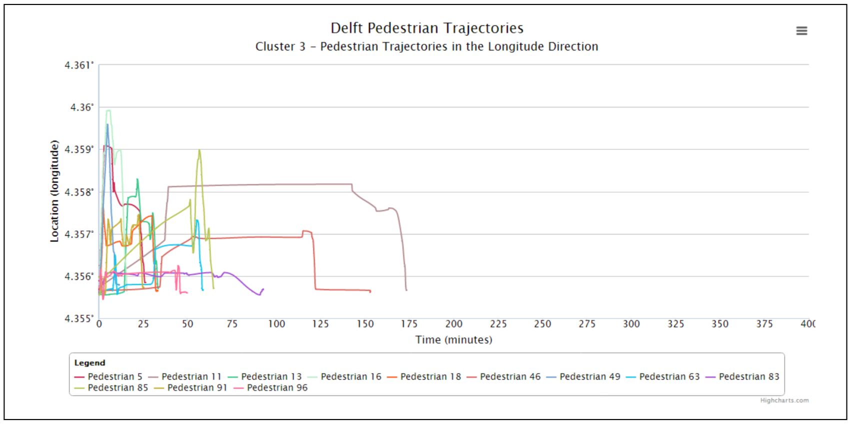

Fig. 11. Trajectories from cluster 3.

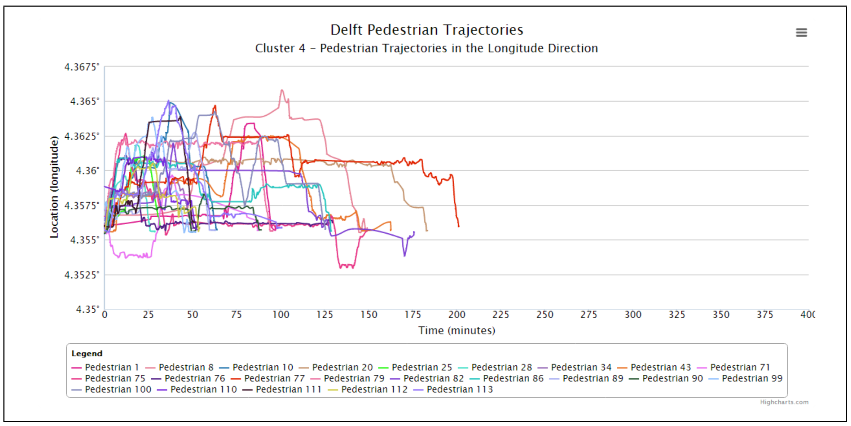

Fig. 12. Trajectories from cluster 4.

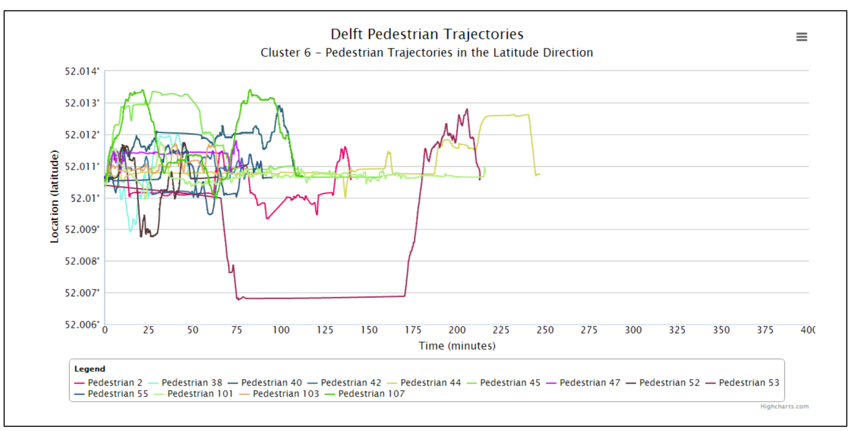

Fig. 13. Trajectories from cluster 6.

Figure No.	Cluster No.	Trip Duration	Number of Stops/ Stop Duration	Potential Explanation
8	1	Long	Few/Long Stops	Having lunch, a coffee or attending a meeting
10	2	Medium-Long	Many/Short Stops	Shopping
11	3	Short	No Stops	Wandering around the city for window shopping
12	4	Medium	Many/Mix of short and medium stops	Shopping combined with lunch/coffee
Not Shown	5	Short	One/Short	Specific purpose for trip - shopping
13	6	Medium	One/Long	Specific purpose for trip - meeting or lunch

Table 1. Summary of the clusters produced, along with a potential explanation of the identified movement behaviour.

5 Discussion & Conclusion

Movement is a fundamental aspect of the physical world. New technologies have increased access to movement data. Studying movement data and recognising patterns can be utilised in many domains to predict future movement, assess how spaces are used and identify unusual behaviour.

This article presents a novel geovisual environment for exploring movement data. The classic STC is used in conjunction with Google Earth to analyse the temporal and spatial aspects of trajectories simultaneously while adhering to the classic visual information seeking mantra of 'Overview first', followed by 'Zoom and filter', and 'Details-on-demand' (Shneiderman, 1996). A new similarity metric which compares speed and acceleration in perpendicular directions is used with spectral clustering to group trajectories exhibiting similar behaviour. Since similarity is based on behaviour, the technique does not require trajectories to be from the same geographic space.

The geovisual environment and clustering techniques were applied to pedestrian movement data collected in Delft in the Netherlands. The trajectories were recorded in the same region so the visual tool enabled spatiotemporal convergence and divergence of pedestrians to be identified. Clustering the trajectories, using the new behaviour-based similarity measures, resulted in a grouping of trajectories in which the speed of pedestrians, the number of stops and the duration of stops were similar. Given that the trajectories were collected in a city, it was possible to infer the activities of pedestrians based on the behaviour seen in the clusters. Six different activity types were identified, ranging from window shopping (no stops with periods of slow movement), to going to a café (consistent speed with a long single stop). The results can be combined with other data sources to understand how different types of individuals interact with each other and the city and to provide input into urban planning tasks.

The case study dataset consisted of pedestrian trajectories, where participants moved at their own pace within a mostly pedestrianised area. This means that the link between speed, acceleration and behaviour was not imposed by external rules. As such, it shows information on the characteristics of pedestrian flows in the city. The approach may not be as useful to analyse a more constrained dataset. For example, where the speed and acceleration of the moving objects are imposed from the outside such as the rules of the road, it would be difficult to deduce a solid meaning from individual behaviour patterns.

We use a data-driven approach for clustering trajectories. Our approach does not require knowledge of the underlying land-use or context, in which the trajectories originated, to produce meaningful clusters. Although not an input, context can subsequently be applied to the resulting clusters, using our visual interface or domain knowledge. The clusters produced in the case study in section 4 were labelled based on the common tasks of pedestrians in a city. Research in the field of context-aware trajectory analysis is now emerging. For example, Gschwend and Laube (2012) are examining the relationship between geographic context and movement while Buchin et al. (2012) are incorporating geographic context into similarity metrics. As the ability to associate geographic context with trajectories improves, we will seek to adapt our approach to utilise such context information as an input to the clustering process.

In general, determining the validity and quality of the clusters produced by any clustering technique is subjective and requires domain knowledge. While ground truth data can be used to assess this, in exploratory tasks, such data is often unavailable. In the case study presented here, the ground truth was of coarser detail than the categories produced by the clustering algorithm and so were not directly comparable. In future studies, we will collect such supplementary data alongside the raw movement data. This will enable the validity of the clustering results to be assessed. Additionally, we will assess the benefits of using this approach with movement data collected in other environments where speed and acceleration also act as a proxy for behaviour.

Acknowledgements: This work is a collaboration started through and performed under the European COST Action "MOVE -Knowledge Discovery from Moving Objects (IC0903). Irish authors are funded by a Strategic Research Cluster grant (07/SRC/I1168) by Science Foundation Ireland under the National Development Plan. We gratefully acknowledge both support sources.

References

Altiparmak, F., Ferhatosmanoglu, H., Erdal, S., and Trost, D. C. (2006). Information mining over heterogeneous and high-dimensional time-series data in clinical trials databases. *IEEE Transactions on Information Technology in Biomedicine*, 10(2):254–263.

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

Andrienko, G. and Andrienko, N. (2010). Poster: Dynamic time transformation for interpreting clusters of trajectories with space-time cube. In *Proceedings of the IEEE Symposium on Visual Analytics Science and Technology (VAST 2010)*, pages 213–214.

Andrienko, G., Andrienko, N., and Wrobel, S. (2007). Visual analytics tools for analysis of movement data. *ACM SIGKDD Explorations Newsletter*, 9(2):38–46.

Ankerst, M., Breunig, M. M., Kriegel, H.-P., and Sander, J. (1999). Optics: ordering points to identify the clustering structure. *ACM SIGMOD Record*, 28(2):49–60.

Ashbrook, D. and Starner, T. (2003). Using gps to learn significant locations and predict movement across multiple users. *Personal and Ubiquitous Computing*, 7(5):275–286.

Atev, S., Masoud, O., and Papanikolopoulos, N. (2006). Learning traffic patterns at intersections by spectral clustering of motion trajectories. In *Proceedings of the IEEE Conference on Intelligent Robots and Systems*, pages 4851–4856. IEEE.

Atev, S., Miller, G., and Papanikolopoulos, N. (2010). Clustering of vehicle trajectories. *IEEE Transactions on Intelligent Transportation Systems*, 11(3):647–657.

Buchin, K., Buchin, M., and Wenk, C. (2006). Computing the fréchet distance between simple polygons in polynomial time. In *Proceedings of the 22nd Annual Symposium on Computational Geometry*, pages 80–87. ACM.

Buchin, M., Dodge, S., and Speckmann, B. (2012). Context-aware similarity of trajectories. In *Geographic information science*, pages 43–56. Springer.

Castro, P. S., Zhang, D., and Li, S. (2012). Urban traffic modelling and prediction using large scale taxi gps traces. In *Pervasive Computing*, pages 57–72. Springer.

Chen, L., Özsu, M., and Oria, V. (2005). Robust and fast similarity search for moving object trajectories. In *Proceedings of the 2005 ACM SIGMOD International Conference on Management of Data*, pages 491–502. ACM.

Chen, W., Song, Y., Bai, H., Lin, C., and Chang, E. (2011). Parallel spectral clustering in distributed systems. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 33(3):568–586.

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(10):1527–1542.

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

Dodge, S., Weibel, R., and Lautenschütz, A. (2008). Towards a taxonomy of movement patterns. *Information Visualization*, 7(3-4):240–252.

Elnekave, S., Last, M., and Maimon, O. (2007). A compact representation of spatio-temporal data. In *Proceedings of the International Conference on Data Mining Workshops (ICDM 2007)*, pages 601–606. IEEE.

Ester, M., Kriegel, H.-P., Sander, J., and Xu, X. (1996). A density-based algorithm for discovering clusters in large spatial databases with noise. In *Proceedings of the 2nd International Conference on Knowledge Discovery and Data Mining*, page 226. Amer Assn for Artificial.

Gschwend, C. and Laube, P. (2012). In *Proceedings of GISRUK 2012*.

Hägerstrand, T. (1970). What about people in regional science? *Papers in Regional Science*, 24(1):6–21.

Halliday, D., Resnick, R., and Walker, J. (1997). *Fundamentals of Physics*. Wiley.

Hu, W., Xie, D., Fu, Z., Zeng, W., and Maybank, S. (2007). Semantic-based surveillance video retrieval. *IEEE Transactions on Image Processing*, 16(4):1168–1181.

Jain, A. K., Murty, M. N., and Flynn, P. J. (1999). Data clustering: a review. *ACM Computing Surveys (CSUR)*, 31(3):264–323.

Kanungo, T., Mount, D., Netanyahu, N., Piatko, C., Silverman, R., and Wu, A. (2000). The analysis of a simple k-means clustering algorithm. In *Proceedings of the 16th Annual Symposium on Computational Geometry*, pages 100–109. ACM.

Kapler, T. and Wright, W. (2005). Geotime information visualization. *Information Visualization*, 4(2):136–146.

Kraak, M. (2008). Geovisualization and time–new opportunities for the space–time cube. *Geographic Visualization: Concepts, Tools and Applications*, pages 293–306.

Kriegel, H.-P., Kröger, P., Sander, J., and Zimek, A. (2011). Density-based clustering. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, 1(3):231–240.

Landau, R., Werner, S., Auslander, G., Shoval, N., and Heinik, J. (2009). Attitudes of family and professional care-givers towards the use of gps for tracking patients with dementia: an exploratory study. *British Journal of Social Work*, 39(4):670–692.

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

Millonig, A. and Gartner, G. (2011). Identifying motion and interest patterns of shoppers for developing personalised wayfinding tools. *Journal of Location Based Services*, 5(1):3–21.

Morris, B. and Trivedi, M. (2009). Learning trajectory patterns by clustering: Experimental studies and comparative evaluation. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2009 (CVPR 2009)*, pages 312–319. IEEE.

Palma, A., Bogorny, V., Kuijpers, B., and Alvares, L. (2008). A clustering-based approach for discovering interesting places in trajectories. In *Proceedings of the 2008 ACM Symposium on Applied Computing*, pages 863–868. ACM.

Sakurai, Y., Yoshikawa, M., and Faloutsos, C. (2005). Ftw: fast similarity search under the time warping distance. In *Proceedings of the 24th ACM SIGMOD-SIGACT-SIGART Symposium on Principles of Database Systems*, pages 326–337. ACM.

Shneiderman, B. (1996). The eyes have it: A task by data type taxonomy for information visualizations. In *Proceedings of the IEEE Symposium on Visual Languages*, pages 336–343. IEEE.

Song, Y., Chen, W., Bai, H., Lin, C., and Chang, E. (2008). Parallel spectral clustering. *Machine Learning and Knowledge Discovery in Databases*, pages 374–389.

Van der Spek, S. (2010). Tracking tourists in historic city centres. *Information and Communication Technologies in Tourism 2010*, pages 185–196.

Van Schaick, J. (2011). Timespace matters-exploring the gap between knowing about activity patterns of people and knowing how to design and plan urban areas and regions.

Vlachos, M., Kollios, G., and Gunopulos, D. (2002). Discovering similar multi-dimensional trajectories. In *Proceedings of the 18th International Conference on Data Engineering*, pages 673–684. IEEE.

Wilson, C. (2008). Activity patterns in space and time: calculating representative hagerstrand trajectories. *Transportation*, 35(4):485–499.

Zhang, Z., Huang, K., and Tan, T. (2006). Comparison of similarity measures for trajectory clustering in outdoor surveillance scenes. In *Proceedgings of the International Conference on Pattern Recognition (ICPR 2006)*, volume 3, pages 1135–1138. IEEE.

Zheng, X., Zhong, T., and Liu, M. (2009). Modeling crowd evacuation of a building based on seven methodological approaches. *Building and Environment*, 44(3):437–445.

Zheng, Y. and Zhou, X. (2011). *Computing with spatial trajectories.* Springer.