A Space-Time GIS for Visualizing and Analyzing Clusters in Large Tracking Datasets

Hongbo Yu

Introduction

Large tracking datasets of moving objects are becoming increasingly available in various research fields due to recent advancements of information and location-aware technologies (Laube et al. 2007). Moving objects in general refer to objects whose location and/or shape may change over time (Erwig et al. 1999). In many tracking datasets, the concerned moving objects usually maintain their shape and identity while their locations change over time (Dodge et al. 2009). These moving objects, such as individual people, vehicles, and wild animals, are recognized as moving points (Erwig et al. 1999). Their movements then can be represented as trajectory lines. A tracking dataset of moving objects contains many such trajectories, which record the locations where each moving object visited and when the object was there. Such datasets provide a unique information source for researchers to explore spatiotemporal distribution of the observed objects. Information derived from such datasets can help researchers identify the locations where and when many observed objects cluster together. Such locations could be places where a group of people gather together for specific activities, bottleneck locations in a transportation network where traffic congestion occurs, or habitat areas of wild animals where they live or hunt for food. Being able to identify these locations and understand their spatiotemporal characteristics can contribute to the knowledge base of the related research areas, especially where such locations have not been clearly identified or well understood. Therefore, large tracking datasets provide promising opportunities for researchers to discover these key locations

H. Yu (✉)
Department of Geography, Oklahoma State University, Stillwater, OK 74078, USA
e-mail: hongbo.yu@okstate.edu

related to the observed moving objects. However, when a large number of objects are involved, it becomes very difficult to identify these locations and discern their spatiotemporal patterns (Dodge et al. 2009). Thus, analysis tools are needed to effectively represent the dataset, restructure the data, and derive useful information (Purves et al. 2014).

Time geography (Hägerstrand 1970), which supports an integrated space-time system for examining relationships between individual's activities and their spatiotemporal constraints, has a natural fit for representing individual-based tracking data and thus provides an elegant approach to studying individual's movements. However, limited geographic computational power in the past has constrained the development of an operational system of the time-geographic framework (Yuan et al. 2004). Recent advancements in computational technology have significantly increased the capability of geographic information system (GIS) to represent, process, and analyze spatial data. Hence, GIS has been suggested as a useful platform to support the implementation of the time-geographic framework and facilitate the management and analysis of tracking datasets. A number of early efforts (e.g., Miller 1991; Kwan and Hong 1998; Kwan 2000a) demonstrated the possibility of implementing the key concepts of time geography in a two-dimensional (2D) GIS environment. However, lacking an integrated time dimension in the design, current mainstream GIS falls short of providing an effective environment to handle tracking datasets which contain rich spatiotemporal information. The current GIS design needs to be extended to support the representation and manipulation of trajectories of moving objects. As there is a revived interest in time geography in the research community, a number of recent attempts have explored the possibility of using a three-dimensional (3D) GIS environment to simulate the space-time system of time geography and provide 3D visualization of space-time paths and prisms in GIS (Kwan 2000b; Buliung and Kanaroglou 2006; Yu 2006; Andrienko et al. 2007; Neutens et al. 2008; Yu and Shaw 2008; Shaw et al. 2008; Shaw and Yu 2009; Kveladze et al. 2015). These studies confirm the possibility of using a 3D GIS environment to operationalize the time-geographic framework, showcase the advantages of interactively visualizing time-geography concepts (e.g., space-time paths and prisms), and demonstrate the potential of using such an analysis environment to support the exploration of spatiotemporal relationships among moving objects. However, it remains a great research challenge to apply such an approach to large tracking datasets. Extending from the existing approaches, this study attempts to develop analysis approaches in a space-time GIS environment to help researchers investigate trajectories stored in large tracking datasets and explore the locations where the paths of the objects cluster in space and time.

The rest of this paper is organized into four sections. The next section includes discussions on related research topics, including moving objects, time geography and space-time GIS. Section 3 introduces the station concept and discusses how this concept can be used to help researchers investigate the spatial and temporal characteristics of places where the paths of objects converge. Several spatial and temporal aggregation methods are proposed to explore the spatial and temporal

extent of stations presented in moving objects datasets. A space-time GIS framework that can support the representation of trajectories and stations is introduced in Sect. 4 and some station analysis results based on a sample individual-based tracking dataset are reported. Finally, concluding remarks are provided in Sect. 5.

Related Research

Many objects in the world move across the space, with or without the change of their shapes. In the literature of spatiotemporal objects, such objects are identified as moving objects (Erwig et al. 1999; Laube et al. 2007; Dodge et al. 2009). Representing a common yet simple case, many moving objects change their locations with fixed shape and identity. In this case, moving objects are recognized as moving points and may be represented as point features. In many tracking datasets of moving objects, shape change does not happen or is not of concern. Therefore, a moving object's trajectory, which records the movement history of the object, can be constructed from a sequence of time-stamped point locations visited by the object. Rich spatiotemporal information of an object's movements is embedded in such a trajectory. However, when the number of objects in a tracking dataset increases, their trajectories may become so tangled and it can be a challenging task to discern anything meaningful from such a dataset. Many studies have attempted to untangle the twisted trajectories and reveal useful information hidden in the lines. A common approach is to develop certain movement descriptors which can be used to capture the characteristics of the object's movement trajectory and simplify the representation of the trajectories (Laube et al. 2007; Dodge et al. 2009). The descriptors can be for the overall shape of a trajectory (e.g., total distance, movement duration, average speed, straightness) or a range of movement properties along a trajectory (e.g., velocity, acceleration, moving direction). These descriptors then can be used to analyze similarities in movement behavior among the trajectories and identify certain movement patterns presented in the trajectory dataset (Laube and Purves 2006; Laube et al. 2007; Dodge et al. 2009; Long and Nelson 2013; Postlethwaite et al. 2013). In these approaches, individual trajectory usually is the focus for deriving the descriptors and the temporal property of a trajectory is either ignored, converted to a duration measure, or treated in a relative time manner. However, when investigating the spatiotemporal characteristics of places where the moving objects gather, it is important and necessary to deal with the original temporal property of the trajectories and analyze the spatiotemporal relationships among the trajectories. A system that embraces an integrated spatial and temporal representation therefore is needed to effectively model and analyze the trajectories to support the investigation.

Hägerstrand (1970) introduced time geography to study human activities and their constraints in a space-time context. In recent years, this framework has drawn great interests among researchers to study the trajectories of moving objects. Time

geography adopts a three-dimensional (3D) orthogonal system, with two dimensions for space and one dimension for time, to study individual's movements in space and time. The space-time path concept of time geography can be readily used to model the trajectories of moving objects (Long and Nelson 2013). Represented as a linear feature in the space-time system, a space-time path allows a continuous representation of the history of an object's changing positions. In general, a path can be constructed from a sequence of two types of segments: vertical segments and tilted segments. While a vertical segment represents an object's stay at a specific location, a tilted segment records an object's movements between two places.

The space-time system of time geography also offers an effective environment for analyzing various spatiotemporal relationships (e.g., co-location in time, co-location in space, and co-existence) among the trajectories when they are represented as a set of space-time paths (Parkes and Thrift 1980; Golledge and Stimson 1997). Among many defined spatiotemporal relationships of paths, the co-existence relationship, which exists when many paths reach and stay at the same location during the same time period, requires constraints in both space and time. Identifying locations where paths co-exist usually plays an important role in investigating spatiotemporal distribution of the observed objects (Yu 2006; Andrienko et al. 2007). In time geography, the concept of station has been used to describe a location where paths cluster in space and time (Pred 1977; Gollege and Stimson 1997; Miller 2005). A station is defined as a place where people can gather and participate in activities. At a station, many individuals will share some time together and their space-time paths will form a co-existence relationship. In the 3D space-time system, a station can be recognized at a place where the vertical segments of multiple paths bundle at a specific location and stay for a certain period of time (see Fig. 1). A tube is usually used to represent the existence of a station and describe its extent in the space-time system. The spatial and temporal extent of a station may vary significantly in different applications (Gollege and Stimson 1997). A station can be a building, a city, or a region in space and its lifespan can range from a couple of hours to decades or even longer. The tubes, which confine the bundled paths in the 3D space-time system, can effectively represent stations and portray their spatial and temporal extents. Therefore, the station concept provides an

Fig. 1 Space-time paths and station
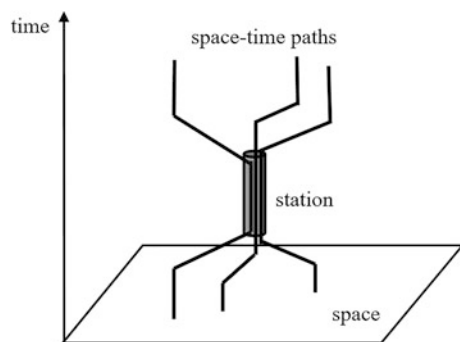
effective guidance for identifying where and when a large amount of space-time paths converge. With the 3D space-time system and the concepts of space-time path and station, time geography offers a useful theoretical foundation for exploring important locations where moving objects cluster in space and time.

Existing studies have shown the potential of GIS in managing moving objects datasets (e.g., Wolfson et al. 1998; Porkaew et al. 2001; Vazirgiannis and Wolfson 2001; Brinkhoff 2002; Dykes and Mountain 2003). These studies attempt to manage moving objects and their trajectories in a two-dimensional (2D) GIS framework, storing the temporal information of the observed objects as a non-spatial attribute in the table associated with the geographic layer. Such an approach works well for certain tasks such as maintaining the dataset and searching for records. However, without an integrated space-time system, the current 2D GIS framework cannot effectively model the rich spatiotemporal information stored in a large tracking dataset. Recent efforts have implemented the space-time system of time geography in GIS and developed a space-time GIS to support the visualization and analysis of space-time paths (see Güting et al. 2000; Kwan 2000b; Buliung and Kanaroglou 2006; Yu 2006; Shaw et al. 2008; Shaw and Yu 2009). Simulating the space-time system, these studies adopt a 3D GIS environment (2D space + 1D time) to support the representation, visualization, and analysis of paths. The space-time GIS environment allows an implementation of space-time path in a more straightforward manner to its original format. Moreover, it opens up further opportunities to operationalize other time geography concepts such as stations and support advanced spatiotemporal analysis applied to space-time paths.

A 3D space-time GIS also presents possibility to represent and visualize the station concept of time geography. In the classic time geography literature, a shape of tube has been used to describe the spatial and temporal extent of a station (Pred 1977; Gollege and Stimson 1997). Such tubes can be represented as space-time cylinders in the 3D space-time GIS. Several studies have attempted to implement the space-time cylinder approach for different types of spatiotemporal datasets. Kulldoff (2001) describes a space-time cylinder around a centroid of a census area as a geographic surveillance method for monitoring time periodic diseases. The height of the cylinder increases with increasing time and the width is determined by the radius based on the population at risk in the census area. The cylinder however is implemented by statistical methods and lacks an interactive visualization environment for spatiotemporal pattern detection and recognition. Onozuka and Hagihara (2007) employ a spatial scan statistic technique and use 3D cylindrical windows to study the geographic distribution and prediction of tuberculosis clusters in Japan. The base of the cylinder represents spatial extent and its height represents time. Both the spatial base and starting time for the cylinder are flexible and mutually independent. The methodology however does not capture fixed spatial units for which a centroid can be used as representative location. Rinner (2004) introduces a tool to model and visualize basic time geography concepts for exploring activity-travel patterns. Simple cylinder shapes with growing or shrinking cross-sections are used to represent and visualize stations. Even though the sizes of

cylinders are not directly related to the number of paths clustering at the locations, they do provide an effective visualization of the station concept.

The existing literature indicates that a space-time GIS approach can be an effective and promising approach to examining clusters of trajectories. Building upon the existing space-time GIS design, this study will develop spatiotemporal analysis tools to facilitate the exploration of stations presented in large tracking datasets and support the spatiotemporal visualization of these stations in a GIS environment.

An Aggregation Approach to Deriving Stations from Space-Time Paths

In the past several decades, we have witnessed the growth of GIS in its capability of managing and representing spatial data. Since the interactive mapping environment of GIS allows more convenient visualization of geographic phenomena and their spatial relationship, it is not surprising that GIS have been frequently used to support geovisualization and exploratory data analysis (Gahegan 2000; Andrienko et al. 2003; Guo et al. 2005; Laube et al. 2007; Kveladze et al. 2015). A GIS design which is capable of accommodating an integrated space-time system will provide a useful environment to visualize and manipulate trajectory data that are represented as space-time paths. When only a small number of space-time paths are involved in a study, it is quite easy to identify the stations formed among the paths through visualizing the paths in a space-time GIS environment. However, when a large number of paths are involved, the visualization scene becomes cluttered and it is impossible to discern any useful patterns. Methods are needed to restructure the data and provide simplified and intuitive visualization of the data to help researchers explore the stations. In this section, an aggregation approach which implements the station concept to effectively aggregate paths is proposed to help researchers explore stations existing in large tracking datasets of moving objects. An intuitive visualization of stations is also provided in the space-time GIS, which is discussed in Sect. 4, to help researchers comprehend the spatiotemporal characteristics of derived stations.

From an analytical perspective, a station is a location where a number of objects can bundle in space and time for certain events. A larger number of objects or a longer total stay time duration of the objects at a location usually indicates a higher significance level of the location as a station (Andrienko et al. 2003). There are many measures to define the significance level of a location as a station. In this study, the significance level of a station is defined by a magnitude measure which is the accumulated duration of all objects staying at the same location during a certain time period. This magnitude measure is used to evaluate the significance level of locations and detect potential station sites where moving objects form clusters. As discussed in Sect. 2, a space-time path is composed of a sequence of vertical and

tilted line segments, which indicate an object's stays at specific locations and moves between locations respectively. Thus, the magnitude of a potential station can be derived by aggregating the vertical segments of multiple paths found at a specific location during a certain time period. Since the spatial and temporal extent of stations can vary significantly, choosing appropriate spatial and temporal resolution levels becomes essential to the exploration of stations. Several spatial and temporal aggregation systems are introduced in this study to provide more flexibility in exploring potential stations at various spatial and temporal structures and resolution levels.

As the first step to derive stations from path, the spatial extent for the aggregation process needs to be determined. Different tracking datasets may record the location information in different formats and at various spatial resolution levels. Also, researchers may have various familiarity levels on the movement patterns of the concerned moving objects. If researchers have developed good understanding of the movement patterns of the objects, they can develop a list of candidate sites for stations and investigate the clusters of trajectories at those sites. When researchers have little knowledge of the moving objects, they will have to rely on the dataset to explore the station sites. Therefore, different spatial aggregation methods may be needed under different circumstances. Three different spatial aggregation methods are proposed in this study to aggregate the recorded trajectories and explore where they cluster, including aggregating the paths based on (1) fixed spatial units, (2) spatial proximity defined by distance, and (3) spatial extent defined by kernel density estimation (KDE) analysis.

In some tracking datasets, a pre-defined fixed spatial unit system may be used in recording the movements of the observed objects. Such a fixed spatial unit system could be zip code tabulation areas, traffic analysis zones, counties, or a custom defined grid system. If it is meaningful to use a pre-defined fixed spatial unit to describe the distribution of the moving objects under investigation and represent the spatial extent of their stations, the fixed spatial units can be used to guide the aggregation of the paths. It is quite common that a pre-defined fixed spatial unit system may have a built-in hierarchical structure. For example, the administration boundary system contains several levels, including regions, states, and counties. The existing hierarchical structure of a fixed spatial unit system then can be used to support the exploration of the stations presented in a dataset at various spatial resolution levels.

In some cases, a pre-defined fixed spatial unit system may not be appropriate to guide the aggregation process. If prior knowledge of potential locations as stations exists, the distance-based spatial proximity method can be used to aggregate the paths. Under this approach, several places will be selected as the potential station sites first. Based on the existing knowledge of a potential site, a search radius can be determined to define the spatial extent for the potential station. A buffer zone will be calculated for each site and used to delineate the boundary of a potential station. All space-time paths that fall within the proximity defined by the buffer zone of a site will be aggregated to evaluate the significance of the site as a station. For instance, a half-mile circle centered at a city square can be used to define the spatial extent of

the square as an activity station. Different distances may need to be tested in order to help researchers find the appropriate spatial extent to describe the stations in a given application.

The previous two methods are helpful when researchers already have some knowledge of the potential station sites. However, it is quite often that there is little knowledge of the spatial distribution pattern of the recorded moving objects. In this case, kernel density estimation (KDE) analysis can be used to identify the potential station sites and determine their spatial extents. In this approach, all locations visited by the moving objects are included in the analysis. A visit to a location from a moving object can be defined as the object stays at the location for a certain duration to complete an activity. Then the total stay duration at each location can be calculated by adding up the stay durations from all objects that have visited the location during the observation time period. For instance, if a place is visited twice by an object for 15 and 45 min respectively, and visited once by another object for 30 min, the total stay duration of this place from both objects will be 90 min. The total stay duration is then assigned as the weight of the location in KDE analysis. A density surface can be generated with a proper search radius for KDE analysis and "hotspots" can be identified with a chosen threshold of density level. Different from the previous two methods which use arbitrarily determined boundaries to delimit the location and spatial extent of station sites, this method allows the data to present itself for identifying the station sites. As a result, the derived stations may vary significantly in term of their spatial extent sizes.

After the spatial location of a station is determined, two temporal aggregation methods—fixed time interval method and moving time window method—can be applied to the dataset to investigate the variation of the magnitude of a station over time. The fixed time interval method divides the time span of a dataset into several time periods based on a user-specified time interval, such as a one-day time interval or a five-year time interval. The moving time window method, on the other hand, starts with a time window chosen by a user and creates the next time window by replacing the earliest year in the current time window with the year following the last year in the current time window. For example, a three-year moving time window will create time periods such as 2010–2012, 2011–2013, 2012–2014, etc. Once the time periods are defined (by either the fixed time interval or the moving time window method), the vertical segments of all objects' paths that fall within the spatial extent of a station site and a specific time period are aggregated to calculate the magnitude of the station for that particular time period. The variation of a station's significance over time then can be examined via the sequence of magnitudes calculated for the station site at each defined time period. As a result, the moving time window method usually creates smoother transitions between the adjacent time periods. Different sizes of time intervals and time windows can be tested before a decision is made on an appropriate temporal resolution level for stations in an application.

In the approaches discussed so far, a space-first-and-time-second strategy is implied for exploring the stations. Following this strategy, the spatial extents of stations are defined first (using fixed spatial units, proximity defined by distance, or

spatial extent defined through KDE analysis) before the magnitude changes of these stations are examined. This strategy is based on an assumption that the spatial extent of a station remains unchanged through the observation time period. However, it is quite common that the spatial extent of a station changes over time. For example, the urbanized area of a city could expand over time through an urban sprawl process and the habitat area of a group of wild animals may migrate to different locations with seasonal changes. Therefore, it is also necessary to explore the spatial extent changes of stations while examining their magnitude variations. To achieve this goal, a time-first-and-space-second strategy can be implemented. Following this strategy, either the fixed time interval method or the moving window method is used to divide the tracking data first, and then a KDE analysis is applied to each subset of the tracking data in order to identify the spatial locations and extents of the hotspots (stations) for that particular time period. Different from the space-first-and-time-second strategy, this approach may produce hotspots with different locations and spatial extents for each time period. By assembling the hotspots from all time periods, it is now possible to examine the evolution of the identified stations in space and time.

Implementing the Station Concept in a Space-Time GIS Environment

This section introduces implementation of the proposed aggregation approaches and the time-geography station concept in a space-time GIS environment. A sample tracking dataset is used to demonstrate how such a GIS environment can reveal stations presented in the trajectories and support the visualization of the stations to help researchers comprehend the spatiotemporal characteristics of stations derived from the dataset. The 3D environment of ArcGIS, which is a product of the Environmental Systems Research Institute (ESRI), is adopted and adapted to simulate the space-time system of time geography. The third dimension (z) is used to represent the time dimension (t). The trajectory of a moving object then can be modeled as a 3D linear feature composed of a sequence of (x,y,t) triplets.

In this space-time GIS design, 3D cylinders are used to represent and visualize stations and a station is modeled as a sequence of cylinders for fixed spatial units or spatial extent defined by distance. All cylinders associated to a station will have their centers located at the same location, which can be either the exact location of a point station or the centroid location of an area station. The height of a cylinder indicates the duration of the defined time period used in the aggregation process, with the bottom surface of the cylinder located at the starting time of the period and the top surface at the ending time. The radius of a cylinder is used to represent the magnitude of the station in the specific time period, which is indicated by the position of the cylinder along the time dimension. A larger cylinder indicates more objects gathered at the site at the time, and the site is more likely to be an important

site for the observed objects. The varying size of cylinders captures the evolution of a site's significance level as a host location for events associated to the objects. Such a space-time GIS representation of stations offers intuitive and convenient visualization to help researchers comprehend the dataset and explore the important locations associated with the observed moving objects.

The aggregation methods are applied to a sample individual-based travel and activity survey data and the derived stations are visualized in the space-time GIS environment for a proof-of-concept study. The sample travel and activity survey data is a subset of the travel tracker survey data collected for northeastern Illinois between January 2007 and February 2008. This subset dataset contains detailed travel inventory of 658 individuals (a total of 3510 recorded trips and activities) who participated in the one-day survey and had at least one trip recorded. Even though the sample data size is not very large, it is used to showcase how the analysis functions and visualization of stations work in the space-time GIS environment, and the functions and visualization can be readily applied to a large dataset. Each record in the survey dataset contains information such as a unique ID for each individual, location visited by the person, when the person arrived at and left the location. Due to privacy concerns, the location information has been aggregated to the census tract level when the data was released to the public. In the data preparation stage, all records belonging to the same individual in the sample dataset are extracted and sorted by time. The location and temporal information in these records is then used to construct a space-time path for the individual. In this process, all locations visited by an individual and his/her stays at those locations during the day are connected in their temporal order to form a 3D linear feature which can be stored in a new dataset in the space-time GIS. As shown in Fig. 2, a total of 658 space-time paths are generated to represent the trajectories of the surveyed individuals. The paths are then used to support spatiotemporal analyses in the space-time GIS for exploring stations presented in the trajectories.

In the first attempt to explore the dataset, each census tract is considered as a potential station as it is the spatial unit for reporting activity locations in the sample dataset, and its magnitude change is examined over the survey time period. The census tract boundary (a fixed spatial unit) and five-minute time interval (a fixed time interval) are used to aggregate the paths that represent the individuals travel activity patterns during the survey day. In this analysis process, a co-existence spatiotemporal relationship is examined among the constructed space-time paths at each census tract using the method developed by Yu (2006). The magnitude of each census tract at each time interval is calculated by accumulating a total stay duration of all paths staying at this location. A cylinder with a radius representing the level of magnitude is then generated and positioned at the centroid of the census tract and the correct time location in the space-time GIS.

Figure 3 shows the space-time GIS visualization of the stations derived from aggregating the paths. Only census tracts with a significant magnitude level are included in the figure. The size of the cylinder associated with a station varies as the station's significance level increases or decreases over time. From the visualization, one can tell that a few census tracts located in the downtown area of Chicago are
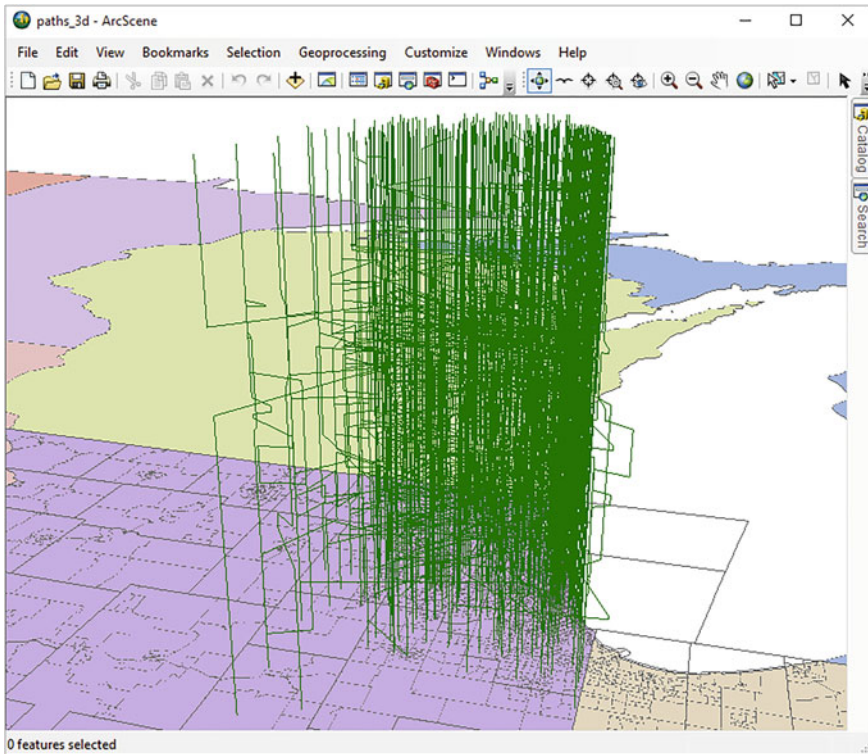
Fig. 2 Sample travel activity survey data represented as space-time paths

represented with a sequence of cylinders that show an increasing and then decreasing trend through the day. In the highlighted case located in the right of the figure, the magnitude level of this location (labeled as s1 in Fig. 3) starts to increase quickly around 8:40 am and maintains high through the day. Its magnitude level starts to decrease quickly at about 4:10 pm. This area is part of the central business district of Chicago, which has many jobs but fewer homes. Many individuals travel to this area for work during the day and leave for home in the evening. The magnitude level changes of this station correctly capture the characteristics of this location for work related activities. Many census tracts located in the suburbs are represented with a sequence of cylinders that have higher magnitude levels at both ends of the day. In the highlighted case (labeled as s2 in Fig. 3) located in the left of the figure, this census tract has high magnitude levels before 7:50 am and after 4:20 pm, and very low magnitude levels in between. As an area with many homes but fewer jobs, people leave this area for work during the day time and will not come back home until evening. The shape variation of the station depicts a typical place for home related activities. With restructured information and its visualization
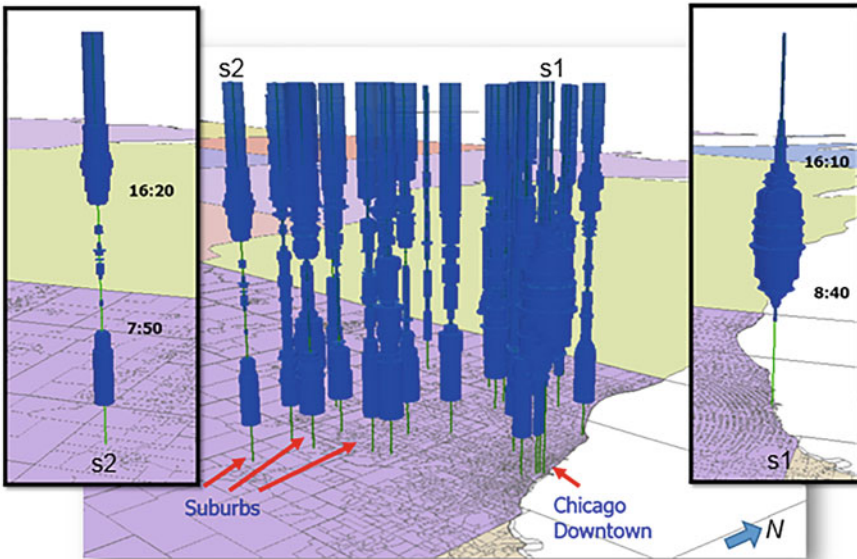
Fig. 3 Visualization of derived census tract-based activity stations in space-time GIS

in the space-time GIS, the activity patterns embedded in the tracking dataset can be easily visualized and comprehended.

The visualization of the aggregation results based on fixed spatial units offers an effective way to examine the activity patterns at the census tract level. However, the census tract boundary line may not be an appropriate choice to delineate the spatial extent of different types of activities, because the sizes of census tracts may vary significantly (much smaller in the downtown area and very large in the suburban area). In the second attempt, the spatial extent of major activity clusters in this area is determined by the activity location distribution presented in the dataset itself, instead of the fixed census tract boundaries. Therefore, the station boundary lines will not be limited to the census tracts. The second attempt will also examine the variations of the spatial extent of these stations over the survey time period. A 20-minute fixed time interval is used to divide the data into smaller subsets and KDE analysis is applied to each of the smaller subsets. A spatiotemporal dynamic segmentation method applied to space-time path (Yu 2006) is used to generate the sub-segments of the trajectories at the defined time interval. Each sub-segment is then converted to a set of 3D points at a finer temporal resolution (e.g., 5 min or 1 min) for KDE analysis. In order to run KDE analysis, a search radius needs to be determined for density calculation. Based on the results of several test runs, a search radius of 3.5 km is chosen as an appropriate radius for the analysis. This radius is about twice of the average size of census tracts in the surveyed area. After a density surface is generated, an equal interval classification method is used to classify the density values into several groups. A threshold value is then selected for identifying

the "hotspot" locations (i.e., potential stations). Later, a cylinder is generated for each defined "hotspot" location, with the cylinder base shaped as the spatial extent of the "hotspot" and its height as the defined time interval. The cylinders are placed in their corresponding positions in the space-time GIS to represent the spatiotemporal characteristics of the identified stations.

Figure 4 shows the visualization of the stations derived from the KDE analysis approach. As shown in this figure, these stations are not composed of strict cylinders (whose intersections are circles), but a sequence of broadly defined cylinders (whose intersections can be irregular shapes). Each of these broadly defined cylinders is derived by extruding the polygon which delineates the spatial extent of the station at a specific time period along the time dimension according to the pre-defined time interval. Similar to the strict cylinders used in Fig. 3, the bottom surface of a broadly defined cylinder is located at the starting time of the period and the top surface at the ending time mark. However, different from the cylinders in Fig. 3 where the size of a cylinder indicates the magnitude of a station at a specific time period, the size of a broadly defined cylinder in Fig. 4 shows the spatial extent of a station. The varying sizes of the cylinders portray the location changes of a station over time.

The results shown in Fig. 4 indicate that different number of stations can be identified in the study area by choosing different levels of density threshold. The higher density threshold level is chosen, the fewer stations are identified (see Fig. 4a–c).

As shown in Fig. 4a, a station may not exist for the entire observation time period and its spatial extent may vary over time. In comparison to the census tract-based station results, the station (labeled as s1 in Fig. 4a) in the Chicago Downtown area now has a larger and changing spatial extent. It only shows up during the day time and is not recognized as a "hotspot" activity location in the early morning and late evening times. Again, this captures the place as a heavy work-related activity location. There are several stations (labeled as s2 and s3 in Fig. 4a) appear only in the early morning and late evening times. They are located in places that have a heavy presence of residential homes, where home-related activities are the major theme. There is one identified station (labeled as s4 in Fig. 4a) whose life time spans the whole day. This station is in an area with mixed land use types (residential and commercial). The combination of home-related activities in the early morning and late evening times and work-related activities during day time makes this place occupied with a significant level of clusters of people through the day. As the density level is calculated and compared through the entire survey area, home activity locations in the suburban areas, which usually have a more spread-out distribution pattern, are not captured in the KDE-based approach due to the very large magnitude level of the downtown area.

It is important to point out that these two approaches to implementing the station concept do not necessarily produce distinct results. Both the spatiotemporal cylinder and the KDE approaches yield a cluster of significant activity stations that
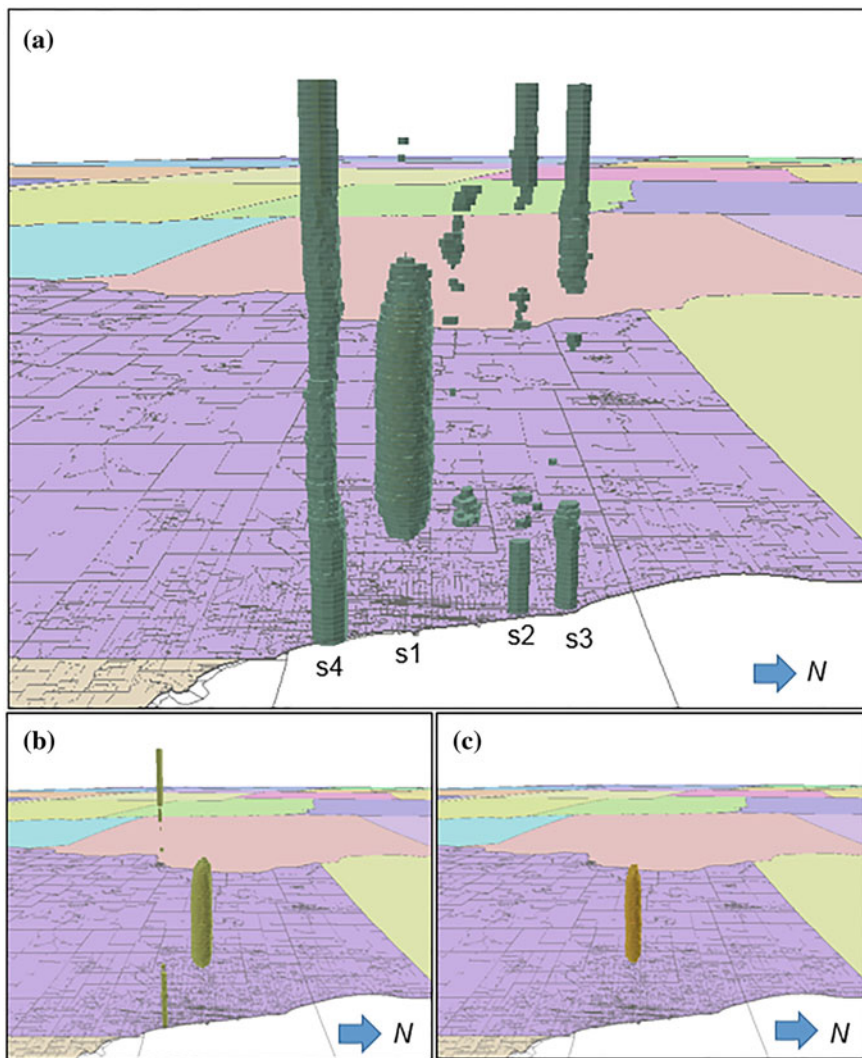
Fig. 4 Visualization of changing stations derived from KDE analysis in space-time GIS. Figure 4a shows stations defined at density level 2 and above; Fig. 4b shows stations defined at density level 3 and above; Fig. 4c shows stations defined at density level 4 and above

are centered in the Chicago Downtown area. Factors that may influence the choice of one approach over the other include whether or not the station sites are known and can be identified and whether the intended results are point-specific or area-based. Both approaches have demonstrated their usefulness to help researchers gain insight into patterns hidden in large individual tracking datasets.

Conclusions

Based on the station concept of time geography, this study proposes a system of aggregation methods in a space-time GIS environment to explore the important locations where the paths of many moving objects cluster in space and time. Several spatial and temporal aggregation methods are introduced to provide flexibility for researchers to manipulate the data and to identify stations with various spatial and temporal extents. Depending on whether researchers have some knowledge of the potential station sites, different representation approaches are proposed. When certain spatial locations have been identified as potential stations according to a priori knowledge, a sequence of cylinders centered at the centroid of an area spatial unit (e.g., a county) or at the exact location of a point unit (e.g., a city) are used to represent the stations. The varying sizes of the cylinders indicate the magnitude changes of the place as a station where the observed moving objects cluster. When little is known about the potential station sites of a group of observed moving objects, a sequence of broadly defined cylinders derived from extruding the polygons in the generated KDE surfaces are employed to visualize locations where space-time paths cluster. The proposed aggregation methods provide an effective approach to restructuring the trajectory data in large tracking datasets and exploring where and when the observed moving objects cluster. Representing stations as 3D objects, the space-time GIS design presents a useful and effective geovisualization environment to investigate the spatiotemporal characteristics of stations. With these capabilities, the proposed methods can benefit various research fields that utilize large tracking data sets for analysis.

At this moment, the proposed methods are designed to explore stations that are defined by the clusters of the vertical segments of space-time paths. In other words, the approach can only capture the bundles of space-time paths at fixed locations. As researchers have acknowledged, space-time paths may bundle either at fixed locations (e.g., buildings) or during movements (e.g., car-pooling). While the first scenarios are known as stationary bundles, the latter ones are referred as mobile bundles (Miller 2004). In order to explore mobile bundles, the tilted segments of space-time paths need to be included in the analysis. As the tilted segments of space-time paths may have numerous choices of directions in the space-time system and it becomes more complex when defining proximity among a titled space-time path segment in the space and time system, it presents an even more challenging research problem. However, being able to identify the mobile bundles among a large number of space-time paths is very important in some studies such as pinpointing where and when the vehicles on a road start to converge and form traffic congestion. For future development directions, the proposed methods need to be expanded so that they can be used to investigate both stationary and mobile bundles among space-time paths and provide enhanced analysis power to explore the spatiotemporal clusters of trajectories in large tracking datasets.

References

Andrienko, G., Andrienko, N., & Gritis, V. (2003). Interactive maps for visual exploration of grid and vector geodata. *Photogrammetry and Remote Sensing, 9*(2), 380–389.

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

Brinkhoff, T. (2002). A framework for generating network-based moving objects. *GeoInformatica, 6*(2), 153–180.

Buliung, R. N., & Kanaroglou, P. S. (2006). A GIS toolkit for exploring geographies of household activity/travel behavior. *Journal of Transport Geography, 14,* 35–51.

Dodge, S., Weibel, R., & 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.

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

Erwig, M., Güting, R., Schneider, M., & Vazirgiannis, M. (1999). Spatio-temporal data types, an approach to modeling and querying moving objects in databases. *GeoInformatica, 3*(3), 269–296.

Gahegan, M. (2000). The case for inductive and visual techniques in the analysis of spatial data. *Journal of Geographical Systems, 2,* 77–83.

Golledge, R., & Stimson, R. (1997). *Spatial Behavior: A Geographic Perspective*. New York: The Guilford Press.

Guo, D., Gahegan, M., MacEachren, A. M., & Zhou, B. (2005). Multivariate analysis and geovisualization with an integrated geographic knowledge discovery approach. *Cartography and Geographic Information Science, 32,* 113–132.

Güting, R., Böhlen, M., Erwig, M., Jensen, C., Lorentzos, N., Schneider, M., et al. (2000). A foundation for representing and querying moving objects. *ACM Transactions on Database Systems, 25*(1), 1–42.

Hägerstrand, T. (1970). What about people in regional science? *Papers of the Regional Science Association, 24,* 7–21.

Kulldoff, M. (2001). Prospective time periodic geographical disease surveillance using a scan statistic. *Journal of the Royal Statistical Society Series A, 164,* 61–72.

Kveladze, I., Kraak, M. J., & Van Elzakker, C. P. (2015). The space-time cube as part of a GeoVisual analytics environment to support the understanding of movement data. *International Journal of Geographical Information Science, 29*(11), 2001–2016.

Kwan, M.-P. (2000a). Human extensibility and individual hybrid-accessibility in space-time, a multi-scale representation using GIS. In D. Janelle & D. Hodge (Eds.), *Information, place, and cyberspace, issues in accessibility* (pp. 241–256). Berlin, Germany: Springer-Verlag.

Kwan, M.-P. (2000b). Interactive geovisualization of activity-travel patterns using three dimensional geographical information systems, A methodological exploration with a large data set. *Transportation Research C, 8,* 185–203.

Kwan, M.-P., & Hong, X. (1998). Network-based constraints-oriented choice set formation using GIS. *Geographical Systems, 5,* 139–162.

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

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

Long, J. A., & Nelson, T. A. (2013). A review of quantitative methods for movement data. *International Journal of Geographical Information Science, 27*(2), 292–318.

Miller, H. (1991). Modeling accessibility using space-time prism concepts within geographical information systems. *International Journal of Geographical Information Systems, 5,* 287–301.

Miller, H. (2004). Activities in space and time. In D. Hensher, K. Button, K. Haynes, & P. Stopher (Eds.), *Handbook of transport 5, transport geography and spatial systems* (pp. 647–660). London, UK: Elsevier Science.

Miller, H. (2005). A measurement theory for time geography. *Geographical Analysis, 37*(1), 17–45.

Neutens, T., Van de Weghe, N., Witlox, F., & De Maeyer, P. (2008). A three-dimensional network-based space-time prism. *Journal of Geographical Systems, 10*(1), 89–107.

Onozuka, D., & Hagihara, A. (2007). Geographic prediction of tuberculosis clusters in Fukuoka, Japan, using the space-time scan statistic. *BMC Infectious Diseases, 7,* 26–34.

Parkes, D., & Thrift, N. (1980). *Times, spaces, and places: A chronogeographic perspective*. New York: Wiley.

Porkaew, K., Lazaridis, I., & Mehrotra, S. (2001). Querying mobile objects in spatio-temporal databases. *SSTD, 2001,* 59–78.

Postlethwaite, C. M., Brown, P., & Dennis, T. E. (2013). A new multi-scale measure for analysing animal movement data. *Journal of Theoretical Biology, 317,* 175–185.

Pred, A. (1977). The choreography of existence: Comments on Hägerstrand's time-geography and its usefulness. *Economic Geography, 53*(2), 207–221.

Purves, R. S., Laube, P., Buchin, M., & Speckmann, B. (2014). Moving beyond the point: An agenda for research in movement analysis with real data. *Computers, Environment and Urban Systems, 47,* 1–4.

Rinner, C. (2004). Three-dimensional visualization of activity-travel patterns. In M. Raubal, A. Sliwinski & K. Kuhn (Eds.), Geoinformation und Mobilität [Geoinformation and Mobility], Proc. of the Münster GI Days, 1–2 July 2004, Münster, Germany, IfGIprints series No. 22. Verlag Natur and Wissenschaft, Solingen, Germany, pp. 231–237. http://www.ryerson.ca/∼crinner/pubs/rinner-3d-vis_full.pdf. Accessed on Jan 8, 2008.

Shaw, S.-L., & 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, 17*(2), 141–149.

Shaw, S.-L., Yu, H., & Bombom, L. S. (2008). A space-time GIS approach to exploring large individual-based spatiotemporal datasets. *Transactions in GIS, 12*(4), 425–441.

Vazirgiannis, M., & Wolfson, O. (2001). A spatiotemporal model and language for moving objects on road networks. *SSTD, 2001,* 20–35.

Wolfson, O., Xu, B., Chamberlain, S., & Jiang, L. (1998). Moving objects databases, issues and solutions. *Proceedings of SSDB Conference, 1998,* 111–122.

Yu, H. (2006). Spatio-temporal GIS design for exploring interactions of human activities. *Cartography and Geographic Information Science, 33*(1), 3–19.

Yu, H., & Shaw, S.-L. (2008). Exploring potential human activities in physical and virtual spaces, a spatio-temporal GIS approach. *International Journal of Geographical Information Science, 22*(4), 409–430.

Yuan, M., Mark, D., Egenhofer, M., & Peuquet, D. (2004). Extensions to geographic representations. In R. McMaster & E. Usery (Eds.), *A research agenda for geographic information science* (pp. 129–156). Boca Raton, FL: CRC Press.