Video Article

Trajectory Data Analyses for Pedestrian Space-time Activity Study

Feng Qi¹, Fei Du²

¹School of Environmental and Life Sciences, Kean University

²Department of Geography, University of Wisconsin-Madison

Correspondence to: Feng Qi at fqi@kean.edu

URL: <http://www.jove.com/video/50130>

DOI: [doi:10.3791/50130](https://doi.org/10.3791/50130)

Keywords: Environmental Sciences, Issue 72, Computer Science, Behavior, Infectious Diseases, Geography, Cartography, Data Display, Disease Outbreaks, cartography, human behavior, Trajectory data, space-time activity, GPS, GIS, ArcGIS, spatiotemporal analysis, visualization, segmentation, density surface, density volume, exploratory data analysis, modelling

Date Published: 2/25/2013

Citation: Qi, F., Du, F. Trajectory Data Analyses for Pedestrian Space-time Activity Study. *J. Vis. Exp.* (72), e50130, doi:10.3791/50130 (2013).

Abstract

It is well recognized that human movement in the spatial and temporal dimensions has direct influence on disease transmission¹⁻³. An infectious disease typically spreads via contact between infected and susceptible individuals in their overlapped activity spaces. Therefore, daily mobility-activity information can be used as an indicator to measure exposures to risk factors of infection. However, a major difficulty and thus the reason for paucity of studies of infectious disease transmission at the micro scale arise from the lack of detailed individual mobility data. Previously in transportation and tourism research detailed space-time activity data often relied on the time-space diary technique, which requires subjects to actively record their activities in time and space. This is highly demanding for the participants and collaboration from the participants greatly affects the quality of data⁴.

Modern technologies such as GPS and mobile communications have made possible the automatic collection of trajectory data. The data collected, however, is not ideal for modeling human space-time activities, limited by the accuracies of existing devices. There is also no readily available tool for efficient processing of the data for human behavior study. We present here a suite of methods and an integrated ArcGIS desktop-based visual interface for the pre-processing and spatiotemporal analyses of trajectory data. We provide examples of how such processing may be used to model human space-time activities, especially with error-rich pedestrian trajectory data, that could be useful in public health studies such as infectious disease transmission modeling.

The procedure presented includes pre-processing, trajectory segmentation, activity space characterization, density estimation and visualization, and a few other exploratory analysis methods. Pre-processing is the cleaning of noisy raw trajectory data. We introduce an interactive visual pre-processing interface as well as an automatic module. Trajectory segmentation⁵ involves the identification of indoor and outdoor parts from pre-processed space-time tracks. Again, both interactive visual segmentation and automatic segmentation are supported. Segmented space-time tracks are then analyzed to derive characteristics of one's activity space such as activity radius etc. Density estimation and visualization are used to examine large amount of trajectory data to model hot spots and interactions. We demonstrate both density surface mapping⁶ and density volume rendering⁷. We also include a couple of other exploratory data analyses (EDA) and visualizations tools, such as Google Earth animation support and connection analysis. The suite of analytical as well as visual methods presented in this paper may be applied to any trajectory data for space-time activity studies.

Video Link

The video component of this article can be found at <http://www.jove.com/video/50130/>

Protocol

1. Getting Data

1. Trajectory data can be collected with handheld GPS units, GPS-enabled smart phone tracking applications, as well as A-GPS (assisted GPS) devices such as the one employed in our study, a commercial child tracker device.
2. Trajectory data is usually saved in terms of time-latitude-longitude records. A desired time interval should be set based on application needs. Often the most frequent interval is desired for space-time activity studies.
3. Convert the data to comma-separated values, or .csv files with separate columns for record id, latitude, longitude, and time, respectively. Then convert the .csv files into commonly-used Geographic Information Systems (GIS) file format (i.e. ESRI shapefile⁸).
4. Load in a shapefile of building polygons and another of the boundary of the study area with the trajectory analyzer. Set the "extrusion" of the buildings properly for a 3D display and set the "extrusion" and "transparency" of the boundary layer properly to display a space-time cube^{6,9} with the x,y dimensions representing space and the z dimension representing time.

2. Pre-processing

1. Two options are available for pre-processing the noisy raw trajectory data. One may choose from the drop down list of the pre-processing menu.
2. If 'Interactive' is chosen, a 2D projection of the 3D trajectory is created for easy viewing and selection. Manipulate the 3D display to examine the raw trajectory in space and time. Identify errors in the data based on the shape, speed and/or topology of track segments. Usually track points (vertices) with unrealistic high speed or abrupt direction change signify errors. Select and remove them from the original trajectories. Select and remove them from either the 3D trajectory or its 2D projection.
3. A cluster of track points with spiky shapes (**Figure 1**) spatially and a long duration temporally signify errors that are most possibly caused by indoor locations where GPS signal is weak. If a group of these points is selected, the program can calculate the spatiotemporal centroid of the selected points and adjust the track to go through the centroid.
4. Alternatively, if 'Automatic' is chosen from the pre-processing menu, set the input and output locations as well as empirical parameters that determine the abnormal high speed and abrupt turning of points. The program searches through the loaded trajectory data and runs automatically based on an algorithm that mimics the visual error detection approach.

3. Trajectory Segmentation & Activity Space Characterization

1. Trajectory segmentation requires the building layer, so ensure the building shape file is loaded.
2. Click the segmentation tool in the toolbar to start the function. Set the input and output and located the building shape file as the reference layer. Use the building names to label the segmented trajectory. The algorithm identifies indoor segments based on set or default criteria such as speed, duration, etc. of track points, as well as the spatial topology with relation to buildings.
3. Click the activity space summarization tool to load in segmented trajectories and calculate selected summary attributes to characterize one's activity space, such as total activity radius, radius at a certain time period, ratio of total time spent indoors vs. outdoors, and so on.
4. The attributes can be exported to a spreadsheet for quantitative modeling uses.

4. Density Surface Mapping

1. Density surface shows the density of activities in space with the temporal dimension collapsed. Three options are available from the drop-down list of the density surface mapping menu.
2. If the 'Track point density' option is selected, fill in the dialog box with input and output information and choose to display in either 3D or 2D. All vertices from the trajectory data are used to calculate kernel densities of the points. **Figure 2** shows a density surface.
3. If 'Track path density' is selected, the algorithm calculates and displays density of individual paths traveled (**Figure 3**).
4. If the 'Re-sampled point density' option is selected, the algorithm re-samples the trajectory data using a set time interval and maps the densities of points spread evenly in time. This option is designed for tracking devices that collect tracking points in irregular time intervals due to varying sensitivity of the devices under various physical conditions or segmented trajectories. **Figure 4** shows the 2D and 3D density surfaces of segmented trajectories.
5. If 'Temporal focusing' is selected for any of the above options, temporal focusing¹⁰ can be performed to examine activity patterns at different time periods. For example, activity density surfaces at different times in a day may be visualized for easy identification of hot spots across time (**Figure 5**).

5. Density Volume Estimation and Volume Rendering

1. Density volume visualization uses the notion of a space-time cube as in the visualization of trajectories. The core of such visualization is the disaggregation of space into voxels¹¹. Our approach to visualizing density volume first estimates density volume in individual voxels by counting the number of space-time tracks that intersect with the voxels. One may click 'Density volume calculation' under the density volume visualization menu for this step.
2. The same three options are available for density volume visualization as for density surface visualization.
3. Next click 'Volume rendering' to launch the 3D volume visualization interface for interactive volume rendering¹². By setting the number of divisions along each axis, one may examine clusters at different scales. A z-factor is used to set the vertical exaggeration for better visualization. A reference layer such as the buildings can be loaded to aid visualization as well. The results of volume rendering can be interactively adjusted by manipulating the transfer function that controls the mapping from density to color. (**Figure 6**).

6. Other Exploratory Data Analyses (EDA) and Visualizations

1. A procedure is available to create animated series to be displayed in Google Earth. Under 'Other', click 'Export to KML for EDA' to access this procedure. It creates a kml¹³ file that opens in Google Earth for interactive animation of the trajectory.
2. One may follow the trajectory to travel the environment in time by scrolling along the timeline in Google Earth.
3. A procedure is available to visualize connections among places of interest through 'Connection analysis'. For example, connections among different buildings on a University campus are derived from segmented trajectory data that were collected by students (**Figure 7**).
4. Based on the derived connections, hotspots such as those buildings with the most outbound or inbound traffic and hubs that connect the most trafficked places may be identified.

Representative Results

Trajectory data was collected by volunteering undergraduate students from Kean University (NJ, USA) in spring 2010. The purpose was to study activity patterns of students who caught influenza (diagnosed by doctor or self-diagnosed) in comparison to those who did not. In order to illustrate the methods and procedure presented in this paper we took the trajectories collected within the suburban campus area to generate representative results. Trajectories within the campus area are mostly pedestrian trajectories, with only a small portion resulting from driving between the various parking lots and outside of the campus.

The space-time cube representation of trajectories with reference to buildings on the university campus is shown in **Figure 1**. **Figure 1A** is the raw data collected by a student recording one day of his activity on campus using an AGPS device (a commercial child tracker). It is obvious that some long duration of indoor stays has resulted in noisy data (indicated by the spiky portion of the track). This is very common in pedestrian trajectory data. **Figure 1B** shows the pre-processed and segmented trajectory. **Figure 1C** shows the pre-processed and segmented trajectory with color-coded indoor and outdoor segments in the space-time cube.

Figures 2 to 4 illustrate density surface mapping of a set of trajectories. **Figure 2** shows the raw tracking points involved to perform a 'Track point density' mapping option (**Figure 2A**) and the resulting density map (**Figure 2B**). Instead of mapping densities of tracking points, **Figure 3** maps the densities of traveled paths. Density mapping is particularly useful when analyzing large amount of trajectories. **Figure 4A** displays a total of 470 trajectories. **Figures 4B** and **4C** show the density surface in a 2D (left) and 3D representations (right) using re-sampled points from these trajectories.

In addition to the interactive display of the temporal dimension in a space-time cube, the time variable can be processed through temporal focusing to examine spatial patterns at different time periods. **Figure 5** shows examples of such analysis using the sample data set that contains trajectory data collected by students during the flu season. It is obvious that their activities are centered around different locations throughout the day to lead eventually to the composite activity density map on the bottom.

An example of density volume rendering is illustrated in **Figure 6**. **Figure 6A** shows that it is hard to detect patterns if all the space-time tracks are visualized in space-time cube because of visual clutter. **Figure 6B** shows the corresponding density volume rendering results. The four illustrations represent different settings of the transfer function of our density rendering program, thus highlight density volumes at different frequency ranges.

Another way of finding hotspots is through connection analysis. **Figure 7** illustrates the result of such analysis with our sample data set. **Figure 7A** shows the straight line connections among all buildings on campus. The highlighted buildings are those with the highest outbound traffic volume. **Figure 7B** shows the same connections, with the most trafficked connections highlighted.

A

B

C

Figure 1. Pre-processing and segmentation of trajectory data. **A:** 2D view of a raw trajectory on the background of campus buildings; **B:** pre-processed trajectory; **C:** space-time cube representation of segmented trajectory. [Click here to view larger figure.](#)

A

B

Figure 2. Density surface mapping. A: raw track points of a trajectory data set; B: density surface derived from track points.

Figure 3. Density surface of trajectory paths.

A

B

C

Figure 4. Colored density surface mapping. **A:** a total of 470 trajectories; **B:** colored density surface in 2D; **C:** colored density surface in 3D.
[Click here to view larger figure.](#)

Figure 5. Temporal focusing for density mapping: student activity densities on campus at different time periods.

A

B

Figure 6. Density volume rendering and visualization. **A:** visual clutters resulting from raw trajectories; **B:** spatiotemporal clusters highlighted by visualizing density volumes at different frequency ranges. [Click here to view larger figure.](#)

A

B

Figure 7. Connection analysis results. **A:** straight line connections among all buildings on a university campus derived from trajectory data, with the most trafficked buildings highlighted; **B:** the most trafficked connections among buildings on campus. [Click here to view larger figure.](#)

ACTIVITY DIARY

Date: Nov. 18, 2010 GPRS #: 07

TIME (from _:_:_ to _:_:_)	PLACE (example: Science Building, library, outside UC, etc.)	ACTIVITY (example: entering campus, taking class, having lunch, etc.)
10:45-10:56	Sidewalks on campus	Walking to UC
10:56-12:00pm	University Center	Lunch
12:00-12:29	Outside of UC and sidewalk to Hennings Hall	Talking with friends on a bench and walking to Hennings
12:29-1:37	Hennings Hall	Class
1:37-1:41	Sidewalk	Walking to UC by CAS
1:41-1:46	University Center	Meeting a friend
1:46-1:50	Sidewalk	Walking to Vaughn-eams
1:50-1:56	In Vaughn-Eams	Found out class cancelled
1:58-2:03	Taking trolley bus	Trolley bus ride to parking lot on East campus

Figure 8. Portion of an activity diary recorded by a student.

Figure 9. Activity density patterns of two groups of students. A: activity density patterns of students with mild flu symptoms during a flu season; **B:** activity density patterns of students with more notable flu symptoms.

Figure 10. Connection analysis results based on trajectory data of students who had notable flu symptoms during a flu season. A: straight line connections among buildings with the most trafficked buildings highlighted; **B:** the most trafficked connections taken by students with flu. [Click here to view larger figure.](#)

Discussion

We used add-in mechanism of ArcGIS to develop the interface. All the interactive operations were implemented using C++. All the automatic processing and analysis functions were developed using Python.

AGPS data, or GPS data collected by pedestrian presents unique challenge in preprocessing as the errors can be massive due to adjacency to buildings and frequent indoor stops. Moreover, the focus of preprocessing should not be data reduction as what is usually done for vehicle GPS trajectory data due to the already obvious scarcity of tracking points. The obvious error patterns in pedestrian trajectory data, however, provide unique solution to preprocessing. Instead of using standard preprocessing algorithms¹⁴, we developed the heuristic method (2.3) that mimics a manual visual error detection approach mentioned in 2.2) and cleans up errors in the trajectory data. Specifically, it calculates attributes (speed and direction change) for each track point in a trajectory first. Track points with unrealistic high speeds and/or direction changes are removed. It then re-calculates attributes (duration and direction change) for each remaining track point and detects clusters of track points with spiky shapes (a series of track points with abrupt direction changes). Finally the spatiotemporal centroid of each cluster is calculated and the trajectory is reduced and adjusted to go through the centroid.

The automatic pre-processing and trajectory segmentation algorithms have been evaluated using traditional activity diary data. Ten students were recruited to each carry an AGPS to collect the trajectory data and at the same time were asked to actively record their stops and movements. Portion of a typical activity diary is illustrated in **Figure 8**. A three-day experiment generated 30 trajectories. Pre-processed and segmented trajectories were compared to the diary data. The results indicated that 1) the processed trajectories captured majority of indoor activities; 2) time recorded in the trajectory data is more accurate as diary takers often write down a rough estimate of time; 3) the trajectory data

captured all details of walkway paths while only straight line connections could be obtained from diary data; and 4) some activities are missing from the diary data as participants often skip records due to the burden. But one limitation of our approach is that the segmented trajectory data sometimes mislabel an indoor segment in a wrong building, especially when two buildings are connected to each other, which is the case with some buildings in our experiment. Improvement on this aspect of the algorithm is needed.

Density surface mapping is an effective tool to explore activity patterns, especially when large amount of trajectory data is involved. **Figure 4** shows that a large number of trajectories leads to apparent visual clutter if displayed in its original form while density mapping reveals interesting patterns. A simple application of this was conducted using the data set collected on Kean university campus during the 2010 flu season. Students who caught the flu and students who didn't generated two sets of trajectories. Students were also interviewed regarding the severity of their symptoms. **Figure 9** illustrates the activity density patterns of two groups of students, one showing only mild symptoms (**Figure 9A**) and another showing more notable ones (**Figure 9B**). It is noted that the real sick students' activity space tends to cluster around a particular building. Further investigation could thus be conducted to determine the causes of such clustering. This experiment indicates that the method has the potential to reveal hidden patterns in trajectory data.

The above density surface maps, however, collapses the temporal dimension. Density volume visualization uses the notion of a space-time cube and represents both spatial and temporal dimensions. **Figure 6** indicates that such visualization is effective in dealing with visual cluttering problems. Once made interactive, it allows one to manipulate the rendering to highlight different frequency ranges in the data to detect patterns. One limitation of our current approach, however, is that the rendered volume does not appear to be completely smooth. We are in the process of improving the density estimation algorithm to deal with the issue. One consideration is that Kernel density estimation may improve the visual effect, but the computation time would become much longer. Sequential kernel density estimation¹⁵ could be another option that we would investigate.

In addition to detecting activity clusters (hot spots) in time and space, the method we introduced in 7.2 detects another kind of hot spots that is related to connections among places. Again with our data set collected by students during a flu season, connection analysis was conducted to identify strong connections among campus buildings based on all trajectory data (**Figure 7**) as well as those representing only students who demonstrated notable flu symptoms (**Figure 10**). Comparing the two figures, we see that one connection (between a building called the University Center and another called the CAS building) appears to be a strong connection in general (**Figure 7**) but is missing from the set of strong connections identified for sick students (**Figure 10**). The two strong connections that remain in the latter are one between the University Center and the Science Building and another among Science Building, Hennings Hall and Hutchinson Hall. Knowledge regarding these campus buildings indicates that the University Center is the most heavily trafficked stop on campus with a cafeteria and recreation rooms inside. It could be a potential high risk hub during flu season when students interact with each for a long period of time in a crowded space. It is also learned that the three buildings involved in the second connection are all attached to each other with indoor pathways. These buildings have classrooms where students may spend many hours indoors taking classes without having to go outside of a building. These buildings are also relatively old constructions with aged ventilation systems that could increase risks of respiratory disease transmission. The CAS building that appears in the connection in **Figure 7** but not in **Figure 10**, on the other hand, is a brand new building and stands by itself in large open space. Good ventilation and the fact that student activity has to involve outdoor time period when taking classes elsewhere both could lead to lower risks. These, are of course speculations but proves that such analysis, like other methods presented in this paper can be a useful exploratory analysis tool to reveal hidden patterns. This package, however, by no means includes all possible methods useful for trajectory data analyses. We are keeping our efforts going to develop and incorporate more analytical as well as visual functions into our system.

Disclosures

No conflicts of interest declared.

Acknowledgements

This work is funded by NIH grant 1R03AI090465.

References

1. Stoddard, S.T., Morrison, A.C., *et al.* The role of human movement in the transmission of vector-borne pathogens. *PLoS Negl. Trop. Dis.* **3** (7), e481, doi:10.1371/journal.pntd.0000481 (2009).
2. Morens, D.M., Folkers, G.K., *et al.* The challenge of emerging and re-emerging infectious diseases. *Nature*. **430**, 242-249 (2004).
3. Viboud, C., Bjornstad, O.N., *et al.* Synchrony, waves, and spatial hierarchies in the spread of influenza. *Science*. **312**, 447-451 (2006).
4. Shoval, N. & Isaacson, M. The Application of tracking technologies to the study of pedestrian spatial behaviour. *The Professional Geographer*. **58** (2), 172-183 (2006).
5. Yu, H. Spatio-temporal GIS design for exploring interactions of human activities. *Cartography and Geographic Information Science*. **33** (1), 3-19 (2006).
6. Kwan, M. Interactive geovisualization of activity-travel patterns using three-dimensional geographical information systems: a methodological exploration with a large data set. *Transportation Research Part C*. **8**, 185-203 (2000).
7. Demšar, U. & Verrantaus, K. Space-time density of trajectories: exploring spatio-temporal patterns in movement data. *International Journal of Geographical Information Science*. **24** (10), 1527-1542 (2010).
8. Environmental Systems Research Institute, Inc., ESRI shape file technical description. <http://www.esri.com/library/whitepapers/pdfs/shapefile.pdf>, (1998).
9. Kraak, M. & Koussoulakous A. A visualization environment for the space-time cube. In: *Proceedings of 11th International Conference on Developments in Spatial Data Handling*, Fisher, P., ed., Springer, Berlin, 189-200 (2004).

10. MacEachren, A.M., Polsky, C., *et al.* Visualizing spatial relationships among health, environmental, and demographic statistics: interface design issues. *Proceedings of 18th International Cartographic Conference.*, 880-887 (1997).
11. Levory, M., Display of surfaces from volume data. *IEEE Computer Graphics and Application.* **8** (5), 29-37 (1998).
12. Drebin, R.A., Carpenter, L., *et al.* Volume Rendering. *Computer Graphics*, SIGGRAPH88, (1998).
13. Open Geospatial Consortium, Inc. OGC kml documentation. <http://www.opengeospatial.org/standards/kml/>, (2012).
14. Lee, W. & Krumm, J. Trajectory preprocessing. In: *Computing with Spatial Trajectories.*, Zheng, Y., Zhou, X., eds., Springer, Bucher, 3-34 (2011).
15. Han, B., Comaniciu, D., *et al.* Sequential kernel density approximation and its application to real-time visual tracking. *IEEE Transactions on Pattern Analysis and Machine Intelligence.* 0.1109/TPAMI.2007.70771., (2007)