2018

Segmenting human trajectory data by movement states while addressing signal loss and signal noise

Sungsoon Hwang
Cynthia VanDeMark
Navdeep Dhatt
Sai Yalla
Ryan Crews

Segmenting human trajectory data by movement states while addressing signal loss and signal noise

Sungsoon Hwang, Cynthia VanDeMark, Navdeep Dhatt, Sai V. Yalla & Ryan T. Crews

To cite this article: Sungsoon Hwang, Cynthia VanDeMark, Navdeep Dhatt, Sai V. Yalla & Ryan T. Crews (2018): Segmenting human trajectory data by movement states while addressing signal loss and signal noise, International Journal of Geographical Information Science, DOI: [10.1080/13658816.2018.1423685](https://doi.org/10.1080/13658816.2018.1423685)

To link to this article: <https://doi.org/10.1080/13658816.2018.1423685>

Published online: 12 Jan 2018.

Submit your article to this journal [↗](#)

View related articles [↗](#)

View Crossmark data [↗](#)

ARTICLE

Segmenting human trajectory data by movement states while addressing signal loss and signal noise

Sungsoon Hwang^a, Cynthia VanDeMark^a, Navdeep Dhatt^b, Sai V. Yalla^b
and Ryan T. Crews^b

^aDepartment of Geography, DePaul University, Chicago, IL, USA; ^bDr. William M. Scholl College of Podiatric Medicine, Rosalind Franklin University of Medicine and Science, North Chicago, IL, USA

ABSTRACT

This paper considers the problem of partitioning an individual GPS trajectory data into homogeneous, meaningful segments such as stops and trips. Signal loss and signal noise are highly prevalent in human trajectory data, and it is challenging to deal with uncertainties in segmentation algorithms. We propose a new trajectory segmentation algorithm that detects stop segments in a noise-robust manner from GPS data with time gaps. The algorithm consists of three steps that impute time gaps, split data into base segments and estimate states over a base segment. The state-dependent path interpolation was proposed as a framework for gap imputation to deal with locational and temporal uncertainties associated with signal loss. A spatiotemporal clustering-based trajectory segmentation was proposed to detect spatiotemporal clusters of any shape regardless of density to cut a trajectory into internally similar base segments. Fuzzy inference was employed to deal with borderline cases in determining states over base segments based on input features. The proposed algorithm was applied to detect stop/move episodes from raw GPS trajectories that were collected from 20 urban and 19 suburban participants. Sensitivity analysis was conducted to guide the choice of parameters such as the temporal and spatial definitions of a stop. Experimentation results show that the proposed method correctly identified 92% of stop/move episodes, and correctly estimated 98% of episode duration. This study indicates that a sequence of state-dependent gap imputation, clustering-based data segmentation and fuzzy-set-based state estimation can satisfactorily deal with uncertainty in processing human GPS trajectory data.

ARTICLE HISTORY

Received 15 March 2017
Accepted 31 December 2017

KEYWORDS

Uncertain trajectories; stop detection; GPS; path interpolation; fuzzy inference

1. Introduction

Tracking an individual's place history has become increasingly common as location-aware devices produce a large volume of GPS trajectory data, a set of temporally sequenced location and time values (x , y , t) recorded from GPS-enabled devices. The applicability of GPS trajectory data is as wide-ranging as transportation planning, location-based services, marketing, health care, animal ecology and governance. The value of GPS trajectory data, however, will not be fully realized without reliable GPS data

processing methods. GPS trajectory computing can be classified into: (a) preprocessing to make data clean, concise and accurate; (b) management to efficiently access and retrieve data; and (c) mining to uncover patterns and behavior (Zheng and Zhou 2011, Zheng 2015). Approaches differ by whether computing is conducted on (a) individual (Zhou *et al.* 2004, Stopher *et al.* 2008, Schuessler and Axhausen 2009) or collective trajectory (Li *et al.* 2008, Rocha *et al.* 2010, Yuan *et al.* 2013); (b) person (Zhou *et al.* 2004, Stopher *et al.* 2008, Zheng *et al.* 2008, Schuessler and Axhausen 2009), vehicle (Yuan *et al.* 2013) or animal trajectory (Handcock *et al.* 2009); and (c) with external data such as points of interest data (Liao *et al.* 2007), accelerometer data (Hemminki *et al.* 2013) or without external data (Schuessler and Axhausen 2009). This paper focuses on preprocessing techniques for an individual person-based GPS trajectory without external data.

GPS trajectory data is readily accessible, but it is time consuming to process large datasets. Hence, researchers would benefit from aggregating raw GPS track points into homogeneous segments to provide high-level representations of data, such as a sequence of stop and move episodes as a preprocessing step (Spaccapietra *et al.* 2008, Parent *et al.* 2013). In this paper, we propose a new method for segmenting a GPS trajectory data into stop/move episodes. A *stop episode* is a subsequence of a trajectory where a moving object exhibits little movement for at least the defined minimum time duration for a stop or *MinStopDur*. A stop episode can mean different things depending on intended applications (e.g. activity locations, change points). A *move episode* is a part of a trajectory that is delimited by two consecutive stops (Spaccapietra *et al.* 2008, Parent *et al.* 2013). Episodes possess spatiotemporal coordinates (e.g. arrival time, departure time, mean XY coordinates) in contrast to places that do not have temporal coordinates.

Human trajectories are characterized by a prevalence of signal loss and signal noise as significant portions of human activities occur indoors. Figure 1 shows how prevalent time gaps (data loss) are during two days' continuous tracking of a person's whereabouts. GPS signal is not completely lost indoors in many instances, but the signal quality is compromised due to structural interference. Those compromised GPS signals are typically expressed as back-and-forth movement (or indoor jitter) around a stop location. Figure 2 depicts GPS points that were logged for an hour when a person was staying in a single building. The map indicates that structural interference lowers both positional accuracy (>150 m) and attribute accuracy (e.g. false high-speed values).

Uncertainties of human trajectories make it difficult to detect where and when a trip starts and ends. A vast majority of literature on transport mode detection and trip detection does not address this ambiguity as those algorithms are calibrated against outdoor data (Das and Winter 2016, p. 6). For instance, the duration of a stop will be underestimated if a stop is missed due to signal loss, and the duration of a stop will be overestimated if movement in urban canyon is falsely classified as a stop due to signal shortage (Kerr *et al.* 2011, Meseck *et al.* 2016). In addition, validity of feature values calculated for segmentation (such as density, heading and speed) is put into question in sparse trajectories. For example, a stop detection algorithm based on density should be reengineered to accommodate the sparsity of GPS data. Furthermore, compromised GPS signals render feature values unreliable, thereby making existing algorithms ineffective. For instance, indoor jitter (Figure 2) will be misclassified as a move episode by a speed-based algorithm (due to false high-speed values) if the effect of signal noise is not taken

Figure 1. Time gaps in human trajectory data.

Figure 2. Indoor jitter caused by structural interference.

into account. Therefore, effective methods for segmenting human trajectories should deal with uncertainty of trajectories.

There are limitations of existing approaches to dealing with signal loss and signal noise. Signal loss can be dealt with through gap imputation to some extent. A gap is typically imputed as a stop if gap duration exceeds the defined minimum stop duration *MinStopDur* (Li *et al.* 2008, Schuessler and Axhausen 2009). A path interpolation algorithm can be used (Idrissov 2012, Long 2016), but is not well suited to human trajectories of highly irregular sampling rates (Figure 1). Signal noise can be dealt with using moving centroids, accuracy thresholds, filtering calculated on sliding windows and limiting sampling to data collected outdoors, but these techniques come with limitations such as being sensitive to outliers and operational parameters (Zheng and Zhou 2011, Thalmann and Abdalla 2014). There are two challenges with inferring states (stop or move) from GPS data with signal loss. First, feature values are not distinctively different between stop and move, making it hard to classify. For instance, it would be difficult to determine whether a data point with speed 0.8 m/s is stop or move when signal noise is present. Therefore, an algorithm based on a simple threshold does not work well. Second, even if feature values are smoothed out over a moving window to buffer the effect of noise, it is difficult to estimate states accurately when the window size is chosen with disregard to internal similarity. Precision of local feature values over the moving window can be unfairly influenced by signal noise, thereby increasing the chance of false detection.

To overcome these limitations, the proposed method (a) imputes likely states (stop or move) during signal loss; (b) splits a trajectory into internally similar units (or base segments); and then (c) classifies states based on the overlapping membership degrees of local feature values that are calculated over base segments (not arbitrarily chosen segments). We achieve these three tasks using a state-dependent path interpolation algorithm, duration-based spatiotemporal clustering and fuzzy inference, respectively. Below, we review methods related to detecting stop episodes from GPS trajectory where signal loss and noise are relatively well considered.

2. Related work

Stop detection can be achieved as part of algorithms that detect modes of transportation or as a preliminary to mode detection (Zheng *et al.* 2008, Reddy *et al.* 2010, Biljecki *et al.* 2013). This paper focuses on the latter approach (detect stops first). We omit the review of the former approach (detect moves first) in this paper as it has already been reviewed (Prelipean *et al.* 2017). Researchers generally agree on what features should be used in detecting stop episodes without external data. They are namely speed, density, heading change, location and time (Buchin *et al.*, 2015). In contrast, there are variations of *how to* detect stop segments from GPS trajectory. Segmentation algorithms can be classified into four categories: rule-based, clustering-based, kernel-based and cost-function-based approach.

Rule-based approach detects segments using heuristic rules as to the association of input feature values with output values (stop or move). A set of consecutive GPS points are typically detected as a stop if those data points meet some of the following criteria: lack of location change (or high density), low speed and exceeding the minimum stop

duration MinStopDur (Stopher *et al.* 2008, Schuessler and Axhausen 2009). Li *et al.* (2008) proposed a staypoint-detection algorithm that marked a staypoint if the distance among two or more consecutive data points exceeded a distance threshold (200 m) and time span among those data points exceeded a time threshold (30 min); this effectively identifies a gap whose duration exceeds a time threshold as a stop regardless of spatial density. Heuristic rules were developed to impute likely states during signal loss; those rules were largely based on gap duration, average gap velocity and gap distance (Stopher *et al.* 2008, Schuessler and Axhausen 2009, Biljecki *et al.* 2013, Meseck *et al.* 2016). Suggested values for MinStopDur vary from 2 min to 30 min. Rule-based approach can be easily implemented, but segments detected are not robust to noise.

A clustering-based approach adapts a density-based clustering algorithm such as DBSCAN (Ester *et al.* 1996), to fit a temporal constraint of stops (i.e. should exceed MinStopDur). A density-based clustering was used to detect where taxis are most likely to be parked (Yuan *et al.* 2013); to detect important places (Zhou *et al.* 2004); and to measure community mobility (Hwang *et al.* 2013). However, gaps (signal loss) were not considered in these studies. Palma *et al.* (2008) adapted DBSCAN to detect regions with low speed while treating gaps exceeding MinStopDur as a stop. Density, direction change and time were taken into account within the framework of DBSCAN for detecting stop episodes (Gong *et al.* 2015). A clustering-based approach is useful in detecting stops of any shape that is robust to noise, but cannot detect stops reliably from a sparse trajectory.

A kernel-based approach determines states within a moving window (kernel) of the fixed size and merges data points based on homogeneity of feature values calculated across the kernel. The kernel can be point-based (e.g. ± 2 points before and after a current data point p) or time-based (e.g. ± 2 min. before after p). Dodge *et al.* (2009) partitioned a trajectory into trip segments using machine learning based on the deviation and sinuosity of local features (movement parameters) from the median line. Wan and Lin (2016) estimated probabilities that belong to three states (stop, slow move and fast move) for each individual GPS point using fuzzy inference, and aggregated GPS points into segments based on the smoothed probabilities within a kernel. Das and Winter (2016) estimated states over a kernel and merged consecutive data points using a set of heuristic rules. In this approach, gaps are typically used to split a trajectory into segments and states during signal loss are estimated based on similarity among data points adjacent to gaps. Although a kernel-based method can be applied to real-time data processing, it is difficult to accurately determine states over an arbitrarily chosen kernel.

A cost-function-based approach partitions a trajectory into segments by minimizing a cost function comprised of the conciseness of segments and intra-segment similarity that formalizes the principle of minimum description length (MDL). The cost function was defined in terms of the spatial distance for line segments (Lee *et al.* 2007), and the attribute distance between representative points and other points within segments (Júnior *et al.* 2015). The similar approach was proposed for noise filtering (Riyadh *et al.* 2017). This approach is useful in identifying characteristic points (e.g. where direction abruptly changes) in a trajectory, but gaps are not explicitly taken into account.

A path interpolation algorithm (Long 2016) can be used to estimate the 'unknown location at a given time during signal loss' $z(t_u)$ ($t_i < t_u < t_j$) as follows:

$$z(t_u) = z(t_i) + \bar{v} \Delta t \text{ where } \Delta t = t_u - t_i \quad (1)$$

where $\bar{v}$ is unknown velocity during Δt . It states that $z(t_u)$ is determined by how much a person moves from the last known location $z(t_i)$. $\bar{v}$ is a key parameter in estimating $z(t_u)$ since Δt is known. In a simple linear interpolation, a person is assumed to move at the average velocity between i and j (or $\bar{v}_{ij}$), calculated as gap distance Δz_{ij} divided by gap duration Δt_{ij} , following a linear path between i and j . Although this simplifies geographic representation of potential movement from i to j (Pfoser and Jensen 1999, Hornsby and Egenhofer 2002), the assumption of constant motion during signal loss is not usually held. Humans typically alternate a sequence of stop and move rather than constantly move. This is particularly likely if gap distance Δz_{ij} and gap duration Δt_{ij} are large. $z(t_u)$ can be better modeled as a function of states like stop and move.

The proposed method, in principle, can fall into a clustering-based approach in that a spatial clustering algorithm is used to cut GPS points into segments of two types (cluster and noise). However, we attempt to overcome the limitation of a clustering-based approach whose performance is sensitive to sparsity of data by considering time duration in place of density. Moreover, our approach addresses a limitation of a kernel-based approach by calculating local feature values over homogeneous segments of varied size instead of a kernel of uniform size. Furthermore, the proposed method differs from a rule or cost-function-based approach that relies on discrete cutoff values as it addresses ambiguity associated with classification by modeling a continuous membership value (e.g. probability of being a stop is 50% if a speed is 0.8 m/s rather than all or nothing). Finally, we develop elaborate rules for imputing likely states between i and j (or S_{ij}) as any of stop, move or the combination of stop and move in contrast to a simple linear interpolation that assumes a constant motion between i and j .

3. The proposed method

The proposed method (outlined in Figure 3) classifies individual GPS points into stop or move without external data. The algorithm also generates a personal itinerary that lists the start and end time of stop and move episodes in sequence. The hallmark of the proposed method is to split a trajectory into internally homogeneous base segments using gap imputation and a clustering-based segmentation algorithm, and to update states over those base segments using fuzzy inference iteratively. Gap imputation is needed to deal with locational and temporal uncertainties associated with signal loss for segmenting a trajectory. A density-based clustering algorithm is adapted to detect spatiotemporal clusters of any shape *regardless of density* in cutting a trajectory into base segments. Fuzzy inference is employed to deal with borderline cases in determining states over base segments based on input features. In essence, the proposed method turns a large volume of a sparsely and discretely sampled GPS trajectory into a temporally continuous sequence of concise stop and move episodes.

The algorithm begins with removing inaccurate, redundant records, and outliers in order to suppress the influence of uncertainty and reduce processing time. Inaccurate records (GPS points) are defined as records with high Horizontal Dilution of Precision (≥ 4), low number of satellites used in calculation (< 4), no location fix and zero speed whenever information related to quality of GPS signal is available (Stopher *et al.* 2008). Redundant records are those that are within a small distance (< 10 m or less depending on a recording interval) from the previous record (Zhou *et al.* 2004). In dealing with

Figure 3. Outline of the proposed method.

outliers, filtering was not used because filtering alters original locations of raw GPS points, and the new location calculated after filtering is sensitive to outliers that filtering attempts to deal with (Zheng and Zhou 2011, p. 20–32). It was also shown that performance was sensitive to the type and size of sliding windows (Thalmann and Abdalla 2014). Instead, data points with extreme speed (>45 m/s) and data points with abrupt location change¹ were removed (Shekhar *et al.* 2003).

3.1. Impute gaps using state-dependent path interpolation

An imputable gap is defined as a period between i and j where the time span (or Δt_{ij}) is within MinStopDur and MaxStopDur for operational purposes. States during signal loss S_{ij} can vary depending on gap distance Δz_{ij} and average gap velocity $\bar{v}_{ij}$ (Figure 4). A gap can be imputed as a stop (e.g. staying indoors) if both gap distance and velocity are small; as a move (e.g. taking subway, passing tunnel) if both gap distance and velocity are large. In Figure 4, spatial size of stop (SR) serves as a cutoff value for gap distance, and the minimum velocity of move (MV) serves as a cutoff value for gap velocity. Large gap distance and small gap velocity (quadrant II in Figure 4) can be mapped to the occurrence of stop and move (SAM) since large gap distance indicates moving from i to j while small gap velocity indicates being stationary at some point. Small gap distance and large gap velocity (quadrant IV in Figure 4) can be regarded as the incidence of either stop or move (SOM) – move if a person is making a short trip ($\leq SR$) or stop if high-speed values ($\geq MV$) is an artifact of signal noise (Figure 2).

Once S_{ij} were imputed, state-dependent path interpolation (Figure 5) is used to estimate $z(t_u)$. It states that $z(t_u)$ is set to the location of the last known point $z(t_i)$ if S_{ij}

Figure 4. Decision tree for gap imputation.

If $\Delta z_{ij} \leq SR$ and $\bar{v}_{ij} < MV$: // $S_{ij} = \text{Stop}$
 $z(t_u) = z(t_i)$

If $\Delta z_{ij} > SR$ and $\bar{v}_{ij} \geq MV$: // $S_{ij} = \text{Move}$
 $z(t_u) = z(t_i) + \bar{v}_{ij} \Delta t$

If $\Delta z_{ij} > SR$ and $\bar{v}_{ij} < MV$: // $S_{ij} = \text{SAM}$
 If $\bar{v}_{ij} < \bar{v}_{>j}$: // $S_{ij} = \text{SAM-SM}$:
 If $S_{>j} = \text{Move}$:
 $z(t_u) = z(t_i)$ if $t_i \leq t_u < t_a$
 $z(t_u) = z(t_i) + \bar{v}_{>j} \Delta t$ if $t_a \leq t_u \leq t_j$
 where $\bar{v}_{>j}$ is average speed of data points within MinSegDur following a gap
 Else: // If $S_{>j} = \text{Stop}$
 $z(t_u) = z(t_i)$ if $t_i \leq t_u < t_a$
 $z(t_u) = z(t_i) + \bar{v}_m \Delta t$ if $t_a \leq t_u \leq t_j$
 where $\bar{v}_m$ is average speed on a route between i and j
 Else: // $S_{ij} = \text{SAM-MS}$:
 If $S_{<i} = \text{Move}$:
 $z(t_u) = z(t_i) + \bar{v}_{<i} \Delta t$ if $t_i \leq t_u < t_a$
 $z(t_u) = z(t_j)$ if $t_a \leq t_u \leq t_j$
 where $\bar{v}_{<i}$ is average speed of data points within MinSegDur preceding a gap
 Else: // If $S_{<i} = \text{Stop}$
 $z(t_u) = z(t_i) + \bar{v}_m \Delta t$ if $t_i \leq t_u < t_a$
 $z(t_u) = z(t_j)$ if $t_a \leq t_u \leq t_j$
 where $\bar{v}_m$ is average speed on a route between i and j

If $\Delta z_{ij} \leq SR$ and $\bar{v}_{ij} \geq MV$ // $S_{ij} = \text{SOM}$:
 If $S_{<i} = \text{Stop}$ and $S_{>j} = \text{Stop}$:
 $z(t_u) = z(t_i)$
 Else:
 $z(t_u) = z(t_i) + \bar{v}_{ij} \Delta t$

Figure 5. State-dependent path interpolation for gap imputation.

is imputed as stop. $z(t_u)$ is linearly interpolated between the path between i and j assuming a person moves at the speed of $\bar{v}_{ij}$ if S_{ij} is imputed as move. If S_{ij} is imputed as SAM, either the transition from stop to move (SM) or the transition from move to stop (MS) can occur. S_{ij} is mapped to SM if speed increases after a gap, and to MS otherwise.

Finally, once S_{ij} is imputed as SOM, $z(t_u)$ is set to $z(t_i)$ if a gap is surrounded by stops at both ends, or $z(t_u)$ is set to $z(t_i) + \bar{u}_{ij} \Delta t$ otherwise.

The following rules were used to estimate the velocity of a move segment when S_{ij} is imputed as SAM. For SM, $z(t_u)$ is presumed to be stationary up until t_a , where $t_i < t_a < t_j$, and then move at an unknown velocity $\bar{u}$ between t_a and t_j . $\bar{u}$ depends on whether $S >_j$ (state of data points following j) is stop or move. If $S >_j$ is move, then a person is assumed to move at the same speed as $\bar{u} >_j$ (average velocity of data points following j). If $S >_j$ is stop, no clue is provided with regard to $\bar{u}$ because a person has already completed a trip. In that case, the average speed on a route taken between i and j (or average walking or driving speed) is used to estimate $\bar{u}$. Similarly, for MS, $z(t_u)$ is presumed to move until t_a at the speed of $\bar{u} <_i$ (average velocity of data points preceding i), and be stationary after t_a if $S <_i$ is move. The same rule as SM applies if $S <_i$ is stop. $S <_i$ and $S >_j$ are determined using fuzzy inference, and data points within the minimum segment duration (MinSegDur) before i and after j are considered for calculating $S <_i$ and $S >_j$ (Section 3.3). Rules developed here are more elaborate than studies by Stopher *et al.* (2008) and Biljecki *et al.* (2013).

The gap imputation algorithm (Figure 5) is useful in demarcating spatiotemporal coordinates of stop and move episodes during signal loss. Gap imputation rules can be added as additional rules in a rule-based approach so that a gap can be imputed as one or more likely states rather than uniformly treated as a stop. Similarly, results of gap imputation (i.e. imputed states at any given time during signal loss) can be taken into account in calculating similarity metrics for merging or partitioning operation in a kernel-based and cost-function-based approach. To incorporate gap imputation into a clustering-based approach, synthetic data points were added at the estimated location at the start and end time of stop episodes such that those points would be explicitly detected as stop episodes by SpatioTemporal Clustering-based Stops and Moves of Trajectories (STC-SMoT) described below.

3.2. Partition a trajectory into base segments using STC-SMoT

DBSCAN checks whether data within a given search radius (Eps) from any unvisited point (core point) exceeds the minimum number of points (MinPts). If the condition above is met (or density reachable), the algorithm checks whether there are more than MinPts within Eps from each of all density-reachable points, to expand a cluster (to form 'density-connected' data points) (Ester *et al.* 1996). This process continues until the density-connected cluster is completely found. Although a clustering-based approach has been used in many occasions for detecting a stop episode, it does not work well with sparse trajectory data (Das and Winter 2016).

To address this problem, we propose a new algorithm STC-SMoT that detects spatiotemporal clusters of any shape and any size regardless of density. The idea is to check whether a *spatiotemporal neighbor* exceeds *MinStopDur* instead of checking whether a spatial neighbor (within Eps) has high density ($> \text{MinPts}$). Clusters C are identified from GPS trajectory D as follows (Figure 6). The algorithm visits an unclassified point P in dataset D and retrieves a spatiotemporal neighbor or $\text{STN}(P)$. P is marked as C if the longest duration of $\text{STN}(P)$ exceeds *MinStopDur*. If the longest duration of $\text{STN}(Q)$, where

```

STC-SMoT (D, SC, MinStopDur) {
  //Inputs:
  // D = {o1, o2, ..., on} Set of objects
  // SC: constant that captures the relationship between Eps and Interval
  // MinStopDur: minimum stop duration in sec.
  //Output:
  // S: set of stops
  // M: set of moves

  For each point P in dataset D:
    If P is not in cluster C:
      X = SpatioTemporalNeighbor(P)
      X.dur = X.MaxTime - X.MinTime
      If X.dur < MinStopDur:
        Mark P as noise
      Else:
        C = next cluster
        expandCluster(P, X, C, MinStopDur)
    Perform majority filtering on D
    S = C
    M = noise
  }

  expandCluster(P, X, C, MinStopDur) {
    Add P to cluster C
    For each point Q in dataset X:
      If Q is not in cluster C:
        Y = SpatioTemporalNeighbor(Q)
        Y.dur = Y.MaxTime - Y.MinTime
        If Y.dur >= MinStopDur:
          Add Q to cluster C
  }
}

```

Figure 6. STC-SMoT algorithm.

$Q \in \text{STN}(P)$, exceeds MinStopDur , then Q is added to C . Data points marked as C are labeled as a stop, and remainders are labeled as a move.

$\text{STN}(P)$ is defined as a set of points that are within Eps and are temporally adjacent. In most of clustering-based approach, Eps is set regardless of recording interval of a GPS trajectory data. This is problematic because Eps should be small if data is recorded in high frequency (small interval), and should be large if data is recorded in low frequency (large interval). To address this problem, Eps (in meter) is calculated as $\text{SC} \times \text{Interval}$, where SC is the constant that captures the relationship between Eps and Interval . The number of points considered for temporal adjacency is calculated as MinSegDur in seconds divided by Interval . For example, if data is recorded at every 10 s, SC is 5, and MinSegDur is 60 s, then 6 (60/10) data points before and after P within 50 (5 × 10) meters will be defined as $\text{STN}(P)$.

Since STC-SMoT does not consider density, sparse data points that are spatiotemporally adjacent can be also identified as a cluster. In contrast to DBSCAN used for trajectory segmentation that takes three parameters (MinPts, Eps and MinStopDur), the algorithm does not need to specify MinPts. To remove remaining noise in labeling, an existing value (noise) is replaced with the most common value over a sliding window consisting of three consecutive data points as the minimum number of data points that forms a segment is two (e.g. two synthetic data points added at the start and end time of a stop episode). In this process, move segments with too short of a duration ($< \text{MinSegDur}$) or length ($< \text{the walking distance for MinSegDur}$) were aggregated with adjacent stop segments. Consecutive and spatially close (≤ 250 m) stops were aggregated into a stop, where 250 m approximate the maximum diagonal length of a building (Wan and Lin 2016).

3.3. Estimate states over base segments using fuzzy inference

At this point, stop and move episodes (base segments) can still be incorrectly labeled due to signal noise. The radius of indoor jitter or outliers are often much larger than Eps, and as a result those data points are not marked as STN(P) even though they should be marked as a stop. To address this uncertainty, we considered aggregate values of speed, angle and circuitry over each base segment. Three aggregate feature values are calculated from data points $p_1, p_2, \dots, p_m$ that constitute a base segment: the average of point speed, the average of point angle and circuitry. The average of point angle is calculated as the mean of the angle ($< 180^\circ$) that is formed between a vector from p_{i-1} to p_i and a vector from p_i to p_{i+1} (Wan and Lin 2016). Circuitry is defined as the ratio of the distance between p_1 and p_m to the sum of length between two consecutive points. If a person walks or drives on a linear road, the average of point angle will be close to 180° and circuitry will be close to 1. In contrast, back-and-forth movement caused by structural interference (Figure 2) will have low point angle and high circuitry.

A stop is typically characterized by low average point speed, narrow average point angle and high circuitry whereas a move is characterized by medium to high average point speed, wide average point angle and low circuitry. Angle was previously considered to address locational uncertainties of a stop (Wan and Lin 2016). Circuitry was additionally considered to further distinguish between a stop and a move in the proposed method. Indoor jitter (stop) often has average point angle that is neither too wide nor narrow, making it difficult to distinguish it from a move. For instance, 20 GPS points in sequence (Figure 7) have average point angle 111° due to rather elongated shape of indoor jitter. A high circuitry value (5.92) can be instrumental in labeling this base segment as a stop.

Three input aggregate feature values were mapped to a fuzzy set membership degree because it is difficult to delineate boundaries of high and low values of features sharply. Three feature values were fuzzified using membership functions (Figure 8).² Membership functions are based on the fact that point speed for a move (stop) has the positive (negative) relationship with speed (Figure 8(a)); point angle of move tends to be wide, and that of stop tends to be narrow (Figure 8(b)); and actual human movement does not take highly circuitous form for efficiency and high circuitry is often an artifact of compromised GPS signal (Figure 8(c)).

Figure 7. A base segment with medium point angle and high circuitry.

Figure 8. Fuzzy set membership functions for speed, angle and circuitry.

The final aggregate membership degree to stop and move was calculated as a weighted average of membership values of three features following Sugeno-type model (Jassbi *et al.* 2006). That way, effects of erroneous values caused by signal noise in one or two features can be diluted in computing the final membership. Then the degree of final aggregate membership to stop and move was defuzzified to either stop or move based on maximum membership values. The final membership value was updated iteratively until no more update was needed (i.e. the updated membership

value over a reassigned segment did not differ from the membership value in the previous iteration any more).

Once states' update was completed, unique episodes of stop and move were generated as spatial data objects with Point and LineString geometry type, respectively. The mean center of data points that make up each stop episode was assigned as the XY coordinates of a stop episode. Finally, attributes such as duration and average speed were assigned to spatial data objects representing episodes. Any discontinuity between episodes within the duration of a recording interval was filled. This led to generating the personal itinerary that lists a continuous sequence of stop and move episodes.

4. Experimentation results and discussions

4.1. Data

The proposed method was applied to GPS trajectories that were collected from 39 participants (20 urban residents and 19 suburban residents). Participants were asked to wear a GPS data logger (QStarz Travel recorder BT-Q1000XT) for two days continuously. The dataset also had variability in terms of sampling rate (1, 5, 10 s), and logging mode (stationary mode on and off). A stationary mode stops logging if the device is stationary for 10 min. In total 271,129 GPS points were collected for 73.36 days of monitoring sometime between April and November of 2014.

Outputs generated from the proposed method (stop/move labeling for GPS points and a personal itinerary) were compared against baseline data for evaluation. To generate baseline data, two research personnel manually inspected the entire raw GPS trajectory data against high-resolution remotely sensed imagery in GIS, to code episode values. If any disagreement in coding raw data between personnel arose, those cases were reevaluated to minimize ambiguity. Coding results were verified against two reference datasets: a log of supervised activities recorded by research personnel at the start of GPS tracking and travel diaries completed by study participants every 2 h during the 48 h monitoring period.

4.2. Sensitivity analysis

Accuracy of the proposed method was assessed over a range of parameter values to guide the choice of parameters, and to examine whether the performance of the proposed method is robust to varying parameter values. Accuracy was measured as F1 score of stop and move as a harmonic measure of precision and recall calculated at a point level for sensitivity analysis. Overall accuracy was assessed with respect to two parameters for gap imputation: MV and SR (Figure 4). Performance peaked when MV reached 1.0 m/s (Figure 9) and SR reached 100 m (Figure 10) for both stop and move. Additionally, the percent of correctly classified duration (Carlson *et al.* 2015) was compared across different configurations of weights for speed, angle and circuitry used in calculating aggregate fuzzy set membership degrees (Figure 11). Performance was consistently high when circuitry was weighted relatively more ($\geq 40\%$) than other features (speed, angle), justifying the use of circuitry as an additional criterion for fuzzy classification.

Figure 9. Sensitivity analysis over the minimum move velocity (MV) for gap imputation.

Figure 10. Sensitivity analysis over the size of stop (SR) for gap imputation.

Accuracy was also assessed with respect to parameters for STC-SMoT: MinStopDur and SC. MinStopDur controls the temporal extent of potential stop episodes, and SC controls the spatial extent of potential stop episodes identified by STC-SMoT. Accuracy decreased as MinStopDur increased (Figure 12). This is expected as the longer MinStopDur the more likely actual stops will be missed. Conversely, a too short MinStopDur is likely to detect a stop that is not meaningful.

The larger SC (Figure 6), the larger search area became in retrieving STN(P). Small SC might help to delineate stop episodes precisely, but could lead to over-segmentation. Large SC might help to accommodate a large noise level of GPS measurement, but this is achieved at the expense of processing time. Sensitivity analysis showed that stops are most correctly inferred at SC = 5, and moves are most correctly inferred at SC = 10 (Figure 13), suggesting SC values between 5 and 10 as optimal values. So far, no differences were observed between urban and suburban data in choosing optimal parameter values. Contrastingly, Figure 14 shows that 50 m work well as search area

Figure 11. Choosing a vector of weights for calculating aggregate fuzzy set membership degrees.

Figure 12. Sensitivity analysis over MinStopDur for STC-SMoT.

(Eps) for urban areas, and 200 m work well for suburban areas if a recording interval is 10 s. This makes sense as a horizontal building size is typically larger in suburban areas than in urban areas.

4.3. Evaluation

The proposed method was implemented over datasets described in [Section 4.1](#) with parameter values as follows: weight for speed, angle and circuitry = [0.2, 0.4, 0.4], MV = 1 m/s, SR = 100 m, MinStopDur = 5 min, MaxStopDur = 2 days and SC = 5. Accuracy was measured in terms of correctly classified duration in seconds as well as the number of correctly classified segments (stop/move episodes). Accuracy by duration was assessed to examine effects of gap imputation as it was designed to improve the estimation of episode duration by marking arrival and departure time ([section 3.1](#)). Results were compiled as a confusion matrix ([Table 1](#)). Overall accuracy by duration was 97.98%, and accuracy by segment was 92.40%. Results of an independent *t*-test

Figure 13. Sensitivity analysis over constant for Eps (SC) for STC-SMoT.

Figure 14. Sensitivity analysis for urban versus suburban group over constant for Eps (SC) for STC-SMoT.

show that there is no difference in accuracy in terms of place category (urban vs. suburban), logging mode (stationary yes/no) and recording interval. This indicates that the proposed method performs reliably across heterogeneous data.

To evaluate classification quality, we calculated several indicators of accuracy for each of 39 datasets (Zhou *et al.* 2004, Tran *et al.* 2013, Gong *et al.* 2015, Wan and Lin 2016). The F1 score of stop (move) calculated at a segment level was 93.12% (93.05) on average. Precision and recall of stop (move) was 90.52% (88.86) and 96.92% (99.06) on average.³ Over-detection was much more common than under-detection because indoor jitter (stop) beyond Eps was misclassified into short moves (thereby creating redundant stops). Given this, adding a heuristic rule that removes short moves with the same origin and destination can improve performance. Mean duration deviation of stop (move) was 2.29 (2.32) min. Ratio of number of stops (moves) found over correct stops (moves) was 1.09 (1.14).

To assess the quality of the proposed gap imputation algorithm (Figure 5), we randomly generated artificial gaps with varying minimum duration (from 5 min to 2 h) in the data described above. We then compared imputed S_{ij} within those gaps to actual

Table 1. Segmentation accuracy (% correctly classified duration and segments) in confusion matrix.

A. The percent of correctly classified duration			
	Ground truth	Stop	Move
Classification results			
Stop		5,888,533 sec (92.95%)	41,650 sec (0.66%)
Move		86,126 sec (1.36%)	319,041 sec (5.04%)
Overall accuracy			6,207,574 (97.98%)
Kappa index: 82.2%			
B. The percent of correctly classified segments			
	Ground truth	Stop	Move
Classification results			
Stop		376 (46.66%)	6 (0.76%)
Move		54 (6.84%)	353 (44.74%)
Overall accuracy			729 (92.40%)
Kappa index: 84.8%

Figure 15. Assessment of the proposed gap imputation algorithm by cross-validation.

S_{ij} . Results (Figure 15) show that performance decreased as gap duration increased as expected, and over 90% of duration were correctly inferred for relatively short gaps (<30 min). Additionally, accuracy was compared between the gap imputation algorithm (Figure 5) and the method that imputes S_{ij} as stop. Accuracy by duration was higher (mean 98.17 > 97.95; median 98.2 > 97.78) and less variable (StDev 1.14 < 1.16) in the former than in the latter. A paired t -test shows that the former outperforms the latter (p -value = 0.016).

Finally, we compared the proposed method to two similar algorithms – namely staypoint detection (SPD) algorithm (Li *et al.* 2008, p. 3) and spatiotemporal clustering on trajectories (STC) algorithm (Das and Winter 2016, p. 13). The two algorithms are comparable to the proposed method in that they have relative focus on stop detection (move detection as a byproduct), and use a similar set of parameters (distance

Table 2. Comparison of the proposed method to other stop detection algorithms.

Method*	A_d	A_p	Precision(Stop)	Recall(Stop)	Precision(Move)	Recall(Move)
A. Comparison of methods at large distance and time threshold (200 m, 20 min)						
Test	98.26%	94.20%	92.65%	93.92%	83.94%	89.96%
SPD	82.25%	83.57%	97.83%	82.45%	38.19%	85.59%
STC	63.82%	88.11%	74.59%	78.31%	76.81%	81.06%
Method**	A_d	A_p	Precision(Stop)	Recall(Stop)	Precision(Move)	Recall(Move)
B. Comparison of methods at small distance and time threshold (50 m, 5 min)						
Test	98.01%	94.78%	96.59%	92.15%	80.78%	95.40%
SPD	83.34%	85.77%	87.06%	75.08%	60.15%	95.65%
STC	52.54%	90.87%	89.82%	92.44%	78.49%	83.23%

* Distance threshold (or Eps) = 200 m; SC = 20, Time threshold (or MinStopDur) = 20 min.

**Distance threshold = 50 m; SC = 5, Time threshold = 5 min.

Test: proposed method; SPD: Stoppoint Detection; STC: Spatiotemporal Clustering; A_d : accuracy by duration (average of % correctly classified duration over 39 data); A_p : accuracy by point (average of % correctly classified GPS points over 39 data).

threshold/Eps/SC and time threshold/MinStopDur). The proposed method outperforms other methods (Table 2). Several things are notable. First, STC led to significantly low accuracy by duration (<64%). This is because STC fails to detect stops with long duration when data are sparse due to signal loss. Second, SPD does not detect moves well (precision < 40%) due to its focus on stoppoint. In general, the proposed method correctly inferred stop and move in a relatively balanced manner.

Outputs of the algorithms for one participant are depicted in Figure 16. Five stop episodes and four move episodes were extracted from a participant's data. Each episode has attributes such as duration, average speed, arrival time and departure time. For instance, a trip with STCID 6 started at 8:03:08 AM and ended at 8:27:09 AM on 24 April, 2014. Then this person stayed at a location with STCID 7 for over 2 h. The personal itinerary shows the sequence of stop (loctype = 0) and move episodes. The proposed method was useful in reducing a volume of data, turning on average 6952 raw GPS points to about 10 trips per participant.

5. Conclusions

This paper proposes a fully automated approach to partitioning sparse GPS trajectory data into a continuous subsequence of stop and move episodes without external data. The current study provides proof of concept that a sequence of state-dependent gap imputation, clustering-based data segmentation and fuzzy-set-based state estimation can satisfactorily deal with uncertainty in processing human GPS trajectory data. This study shows how path interpolation can be extended to model the motion during signal loss as a function of states, which is more suitable for human trajectories of irregular sampling rate. The proposed clustering-based approach detects stops regardless of density, and hence works well with sparse trajectories unlike other clustering-based methods. The segmentation is greatly improved by using STC-SMoT as it processes GPS trajectory into internally homogeneous segments more accurately than other approaches. Ambiguity with distinguishing stop and move was addressed by fuzzy inference based on speed, angle and circuitry.

The proposed method was applied to GPS trajectory data that differ in category of sampling frequencies, logging modes and built environments that were collected by 39

Figure 16. Algorithm outputs: stop/move episodes and personal itinerary.

participants. No significant difference was observed in performance by different categories, suggesting that the proposed method can be used in diverse data. An experiment suggests that the state-dependent path interpolation algorithm can correctly impute likely states (stop or move) for a short gap (<1 h) at least 85% of time. With a rational choice of parameters, the proposed method correctly classifies 98% of the monitoring duration, and 92% of segments on average. High level of accuracy by duration makes the proposed method suitable for use in exposure assessment as dwelling time is important in measuring the extent of exposure.

Human errors in creating baseline data can be acknowledged as a limitation of the study. The use of multiple parameters could also limit the use of the proposed method. In particular, the effect of the interaction among parameter values on the outcomes was not thoroughly examined in this paper. In general, the choice of parameter values should be dictated by data quality, application domains and analysis scale. For instance, a large SC (or Eps) and window size would work well with data with high noise level. It makes sense to use small MinStopDur and SR (stop radius) for detecting sublocations with short duration of visits, and to use large MinStopDur and SR for identifying important places at a rough geographic scale. Accuracy can be improved by refining rules for gap imputation and fuzzy inference, and adding heuristic rules for reconsidering misclassified cases. As alternative strategies, efficacy of different approaches to handling outliers (e.g. filtering vs. outlier removal) can be assessed; issues related to compromised GPS signals can be dealt with upfront to improve both accuracy and efficiency of a stop detection algorithm. The

proposed gap imputation algorithm can be further improved with the use of accelerometer data (Hemminki *et al.* 2013). It is hoped that the proposed method will aid in improving ways to quantify environmental exposure, and understand human mobility in many emerging application domains such as e-Health and smart cities.

Notes

1. To detect data points with abrupt location change, a location deviation score is calculated as a ratio of the distance between the current point p and the representative location of p 's temporal neighbor to standard distance of p 's temporal neighbor. The higher the location deviation score, the more the current location is deviated from its temporal neighbor. Data points with a location deviation score >2.4 are marked as outliers based on experimentations.
2. A sigmoid function was used because it would allow for controlling a range of possible values based on central values and level of uncertainty. Parameters for a central value and slope in the sigmoid function were set based on summary statistics of 15% random sample of collected data. Comparison of membership functions (Figure 8) to logistic regression indicates that defined functions perform fairly well (84% vs. 64%). The automated process for optimizing membership function was described in Biljecki *et al.* (2013, p. 20).
3. To calculate precision and recall, inferred segments within 100 m (SR) and 1 min (MinSegDur) from actual segments are considered to be matched.

Disclosure statement

No potential conflict of interest was reported by the authors.

Funding

This study was supported by a joint institutional grant from DePaul University and Rosalind Franklin University of Medicine and Science [400011].

References

- Biljecki, F., Ledoux, H., and Van Oosterom, P., 2013. Transportation mode-based segmentation and classification of movement trajectories. *International Journal of Geographical Information Science*, 27 (2), 385–407. doi:[10.1080/13658816.2012.692791](https://doi.org/10.1080/13658816.2012.692791)
- Buchin, M., *et al.*, 2015. Segmenting trajectories: a framework and algorithms using spatiotemporal criteria. *Journal Spatial Information Sciences*, 33–63. doi:[10.5311/josis.v0i3.66](https://doi.org/10.5311/josis.v0i3.66)
- Carlson, J.A., *et al.*, 2015. Validity of PALMS GPS scoring of active and passive travel compared with SenseCam. *Medicine & Science in Sports & Exercise*, 47, 662–667. doi:[10.1249/MSS.0000000000000446](https://doi.org/10.1249/MSS.0000000000000446)
- Das, R.D. and Winter, S., 2016. Automated urban travel interpretation: a bottom-up approach for trajectory segmentation. *Sensors*. 16. doi:[10.3390/s16111962](https://doi.org/10.3390/s16111962)
- 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*, Spatial Data Mining–Methods and Applications, 33, 419–434. doi:[10.1016/j.compenvurbsys.2009.07.008](https://doi.org/10.1016/j.compenvurbsys.2009.07.008)
- Ester, M., *et al.*, 1996. A density-based algorithm for discovering clusters in large spatial databases with noise. In: *Proceedings of the Second International Conference on Knowledge Discovery and Data Mining, KDD'96*. Portland, Oregon: AAAI Press, 226–231.

- Gong, L., et al., 2015. Identification of activity stop locations in GPS trajectories by density-based clustering method combined with support vector machines. *Journal of Modern Transportation*, 23, 202–213. doi:[10.1007/s40534-015-0079-x](https://doi.org/10.1007/s40534-015-0079-x)
- Handcock, R.N., et al., 2009. Monitoring animal behaviour and environmental interactions using wireless sensor networks, GPS collars and satellite remote sensing. *Sensors*, 9, 3586–3603. doi:[10.3390/s90503586](https://doi.org/10.3390/s90503586)
- Hemminki, S., Nurmi, P., and Tarkoma, S., 2013. Accelerometer-based transportation mode detection on smartphones. In: *Proceedings of the 11th ACM Conference on Embedded Networked Sensor Systems, SenSys '13*. New York, NY: ACM, 13:1–13:14. doi:[10.1145/2517351.2517367](https://doi.org/10.1145/2517351.2517367)
- Hornsby, K. and Egenhofer, M.J., 2002. Modeling moving objects over multiple granularities. *Annals of Mathematics and Artificial Intelligence*, 36, 177–194. doi:[10.1023/A:1015812206586](https://doi.org/10.1023/A:1015812206586)
- Hwang, S., Hanke, T., and Evans, C., 2013. Automated extraction of community mobility measures from GPS stream data using temporal DBSCAN. In: B. Murgante, et al., eds. *Computational science and its applications – ICCSA 2013, lecture notes in computer science*. Berlin, Heidelberg: Springer, 86–98. doi:[10.1007/978-3-642-39643-4_7](https://doi.org/10.1007/978-3-642-39643-4_7)
- Idrissov, A.Y., 2012. *A data cleaning framework for trajectory clustering*. Alberta, Canada: University of Alberta.
- Jassbi, J.J., et al., 2006. A comparison of Mandani and Sugeno inference systems for a space fault detection application. In: *2006 World Automation Congress. Presented at the 2006 World Automation Congress*, 1–8. doi:[10.1109/WAC.2006.376033](https://doi.org/10.1109/WAC.2006.376033)
- Júnior, A.S., et al., 2015. GRASP-UTS: an algorithm for unsupervised trajectory segmentation. *International Journal of Geographical Information Science*, 29, 46–68. doi:[10.1080/13658816.2014.938078](https://doi.org/10.1080/13658816.2014.938078)
- Kerr, J., Duncan, S., and Schipperjin, J., 2011. Using global positioning systems in health research: a practical approach to data collection and processing. *American Journal of Preventive Medicine*, 41, 532–540. doi:[10.1016/j.amepre.2011.07.017](https://doi.org/10.1016/j.amepre.2011.07.017)
- Lee, J.-G., Han, J., and Whang, K.-Y., 2007. Trajectory clustering: a partition-and-group framework. In: *Proceedings of the 2007 ACM SIGMOD international conference on management of data, SIGMOD '07*. New York, NY: ACM, 593–604. doi:[10.1145/1247480.1247546](https://doi.org/10.1145/1247480.1247546)
- Li, Q., et al., 2008. Mining user similarity based on location history. In: *Proceedings of the 16th ACM SIGSPATIAL international conference on advances in geographic information systems, GIS '08*. New York, NY: ACM, 34, 1–34, 10. doi:[10.1145/1463434.1463477](https://doi.org/10.1145/1463434.1463477)
- Liao, L., Fox, D., and Kautz, H., 2007. Extracting places and activities from GPS traces using hierarchical conditional random fields. *The International Journal of Robotics Research*, 26, 119–134. doi:[10.1177/0278364907073775](https://doi.org/10.1177/0278364907073775)
- Long, J.A., 2016. Kinematic interpolation of movement data. *International Journal of Geographical Information Science*, 30, 854–868. doi:[10.1080/13658816.2015.1081909](https://doi.org/10.1080/13658816.2015.1081909)
- Meseck, K., et al., 2016. Is missing geographic positioning system data in accelerometry studies a problem, and is imputation the solution? *Geospatial Health*, 11, 403. doi:[10.4081/gh.2016.403](https://doi.org/10.4081/gh.2016.403)
- Palma, A.T., et al., 2008. A clustering-based approach for discovering interesting places in trajectories. In: *Proceedings of the 2008 ACM symposium on applied computing, SAC '08*. New York, NY: ACM, 863–868. doi:[10.1145/1363686.1363886](https://doi.org/10.1145/1363686.1363886)
- Parent, C., et al., 2013. Semantic trajectories modeling and analysis. *ACM Computation Survey*, 45, 42, 1–42, 32. doi:[10.1145/2501654.2501656](https://doi.org/10.1145/2501654.2501656)
- Pfoser, D. and Jensen, C.S., 1999. Capturing the uncertainty of moving-object representations. In: R. H. Güting, D. Papadias, and F. Lochovsky, eds. *Advances in spatial databases, lecture notes in computer science*. Berlin, Heidelberg: Springer, 111–131.
- Prelipcean, A.C., Gidófalvi, G., and Susilo, Y.O., 2017. Transportation mode detection—an in-depth review of applicability and reliability. *Transport Reviews*, 37 (4), 442–464. doi:[10.1080/01441647.2016.1246489](https://doi.org/10.1080/01441647.2016.1246489)
- Reddy, S., et al., 2010. Using mobile phones to determine transportation modes. *ACM Transactions on Sensor Networks (TOSN)*, 6 (2), 13.
- Riyadh, M., et al., 2017. ONF-TRS: on-line noise filtering algorithm for trajectory segmentation based on MDL threshold. *Journal of Artificial Intelligence*, 10, 42–48. doi:[10.3923/jai.2017.42.48](https://doi.org/10.3923/jai.2017.42.48)

- Rocha, J.A.M.R., et al., 2010. DB-SMoT: a direction-based spatio-temporal clustering method. In: *Intelligent Systems (IS), 2010 5th IEEE international conference. Presented at the Intelligent Systems (IS), 2010 5th IEEE international conference*, 114–119. doi:[10.1109/IS.2010.5548396](https://doi.org/10.1109/IS.2010.5548396)
- Schuessler, N. and Axhausen, K., 2009. Processing raw data from global positioning systems without additional information. *Transportation Research Record: Journal of the Transportation Research Board*, 2105, 28–36. doi:[10.3141/2105-04](https://doi.org/10.3141/2105-04)
- Shekhar, S., Lu, C.-T., and Zhang, P., 2003. A unified approach to detecting spatial outliers. *Geoinformatica*, 7, 139–166. doi:[10.1023/A:1023455925009](https://doi.org/10.1023/A:1023455925009)
- Spaccapietra, S., et al., 2008. A conceptual view on trajectories. Data Knowl. Eng., Including Special Section: Privacy Aspects of Data Mining Workshop (2006) - Five Invited and Extended Papers 65, 126–146. doi:[10.1016/j.datak.2007.10.008](https://doi.org/10.1016/j.datak.2007.10.008)
- Stopher, P., FitzGerald, C., and Zhang, J., 2008. Search for a global positioning system device to measure person travel. *Transportation Research Part C: Emerging Technologies*, 16, 350–369. doi:[10.1016/j.trc.2007.10.002](https://doi.org/10.1016/j.trc.2007.10.002)
- Thalmann, T. and Abdalla, A., 2014. Assessing the influence of preprocessing methods on raw GPS-data for automated change point detection. In: J. Huerta, et al., eds. *Connecting a digital europe through location and place*. Berlin, Heidelberg: Springer, 123–139.
- Tran, L.H., Dang, T.K., and Thoai, N., 2013. Hybrid stop discovery in trajectory records. In: *2013 24th international workshop on database and expert systems applications. Presented at the 2013 24th international workshop on database and expert systems applications*. 9–14. doi:[10.1109/DEXA.2013.6](https://doi.org/10.1109/DEXA.2013.6)
- Wan, N. and Lin, G., 2016. Classifying human activity patterns from smartphone collected GPS data: a fuzzy classification and aggregation approach. *Transactions in GIS*, 20, 869–886. doi:[10.1111/tgis.12181](https://doi.org/10.1111/tgis.12181)
- Yuan, N.J., et al., 2013. T-finder: a recommender system for finding passengers and vacant taxis. *IEEE Transactions Knowl Data Engineering*, 25, 2390–2403. doi:[10.1109/TKDE.2012.153](https://doi.org/10.1109/TKDE.2012.153)
- Zheng, Y., et al., 2008. Understanding mobility based on GPS data. In: *Proceedings of the 10th international conference on ubiquitous computing, UbiComp '08*. New York, NY: ACM, 312–321. doi:[10.1145/1409635.1409677](https://doi.org/10.1145/1409635.1409677)
- Zheng, Y., 2015. Trajectory data mining: an overview. *ACM Transactions Intelligent Systems Technological*, 6, 1–41. doi:[10.1145/2743025](https://doi.org/10.1145/2743025)
- Zheng, Y. and Zhou, X., eds., 2011. *Computing with spatial trajectories*. New York, NY: Springer New York.
- Zhou, C., et al., 2004. Discovering personal gazetteers: an interactive clustering approach. In: *Proceedings of the 12th annual ACM international workshop on geographic information systems, GIS '04*. New York, NY: ACM, 266–273. doi:[10.1145/1032222.1032261](https://doi.org/10.1145/1032222.1032261)