

# Fine-grained Stop-Move Detection in UWB-based Trajectories

Fatima Hachem<sup>1</sup>, Davide Vecchia<sup>2</sup>, Maria Luisa Damiani<sup>1</sup>, Gian Pietro Picco<sup>2</sup>

<sup>1</sup>University of Milano (Italy), {fatme.hachem, maria.damiani}@unimi.it

<sup>2</sup>University of Trento (Italy), {davide.vecchia, gianpietro.picco}@unitn.it

**Abstract**—Ultra-wideband (UWB) localization enables user tracking with high spatio-temporal resolution, whose exploitation for detecting higher-level mobility patterns is largely unexplored. We study whether *i*) existing detection techniques, developed for coarser-grained localization, apply also to UWB trajectories, and *ii*) the quantitative extent to which this enables finer-grained analyses. We focus on the well-known stop-move pattern, and offer a concrete use case of capturing visits in a real museum. We contribute a novel metric suited to the high UWB spatio-temporal resolution and use it to evaluate representative techniques. We deploy a UWB system in a  $25 \times 15$  m<sup>2</sup> museum area and base our analysis on 70000+ positions and 200+ ground-truth stops. These are very close in space and time, yet results confirm very accurate spatio-temporal estimation in the vast majority of cases.

**Index Terms**—Ultra-wideband, mobility pattern, trajectory.

## I. INTRODUCTION

The increasing availability of *spatial trajectories* [1], [2] from Global Navigation Satellite Systems (GNSS) and efficient techniques to process them [3], [4] enable the extraction of *mobility patterns*, application-specific abstractions of the movement of individuals [5], [6]. Existing approaches focus on the large-scale, outdoor settings germane to GNSS, yielding coarse spatio-temporal resolution. However, a recent technological wave targets sub-meter position accuracy even in indoor spaces, a powerful enabler in several applications [7].

Leading this wave, ultra-wideband (UWB) radios enable communication *and* accurate localization. WiFi and Bluetooth also offer both, but with a positioning error of meters [7]. UWB yields decimeter-level accuracy by relying on narrow pulses ( $\leq 2$  ns) improving time-of-arrival estimation and separation from multipath components. The increasing role of UWB in location-based applications is witnessed by the many Real-Time Location Systems (RTLS) based on it, and its recent inclusion in smartphones, albeit still awaiting public APIs.

**Research questions.** Existing mobility analysis techniques, limited by GNSS, target applications over large areas (m to km) and temporal intervals (hours to months). UWB trajectories, accrued with much higher spatial resolution and temporal frequency, should intuitively unlock finer-grained analyses, e.g., to decimeters and seconds. However, the increased spatio-temporal density of UWB positions, along with errors induced by indoor environments, may conflict with the operation of GNSS-based techniques or limit the resolution attainable. Unfortunately, UWB-based mobility analyses are largely unexplored, leaving crucial questions unanswered: *i*) *Are existing*

*techniques applicable to UWB trajectories, and ii*) *to what quantitative extent they improve spatio-temporal accuracy?*

**Focus: Stop-move detection.** We seek answers to these questions by focusing on the well-known stop-move [8] mobility pattern (§II), key to many applications, for which we illustrate representative techniques (§III). Broadly, a *stop* is an abstraction capturing a visit to a place; a *move* is a transition between stops. In a trajectory, stops are spatio-temporally disjoint: *i*) an individual visits one place at a time, and *ii*) visits to the same place are distinct stops. Stops may represent the home range of migrating animals, lasting months over a large area, or people visits to a point of interest, e.g., the workplace, lasting hours. This application-dependent spatio-temporal granularity differentiates our work from the literature, as we consider stops only few decimeters apart and lasting only few tens of seconds.

**Real-world requirements and setup: Science museum.** These challenging requirements stem from a motivating real-world use case also offering the concrete experimental setup where to distill findings. We are collaborating with the curators of the MUSE science museum (Trento, Italy), interested in the fruition of exhibits by visitors. This is hardly a novel topic; however, existing works [9]–[12] rely on Bluetooth and are limited to coarse spatial resolution, e.g., room-level. In contrast, our target museum area contains exhibits within few decimeters of each other (Fig. 1). Reliably discerning stops near them is basically impossible via Bluetooth. Further, the curators are interested in stop durations as short as 10 s, yielding precious insights on the visitors' behavior.

**Contributions and methodology.** Real-world validation entails comparing stop estimates vs. ground truth. This is often done qualitatively, or by using metrics (e.g., focusing on individual points [13]) that cannot faithfully capture the fine-grained spatio-temporal features we target (§VII). Therefore, we contribute a **novel metric** (§IV) that *i*) associates estimated stops to true ones on a per-stop level, and *ii*) quantifies their temporal overlapping with a novel indicator, S-score, along with the classic F-score summarizing precision and recall.

We base our results on experiments in the museum (§V). We track users wearing a UWB tag via time-difference-of-arrival (TDoA) localization [14] while recording their ground-truth movement via a user-operated smartphone application and cameras we deployed. The 9 trajectories we gather consist of 70000+ position samples and 209 stops over 100 minutes.

We exploit this dataset for a **quantitative analysis along several dimensions** of UWB-based stop-move detection

(§VI). We first confirm the expressiveness of our novel metric, then use it to compare techniques after selecting their best configuration in our context. We consider both the raw trajectories output by UWB and those “smoothed” via Kalman filters. Finally, we quantify spatio-temporal errors by sharply separating the contributions of segmentation and positioning.

Despite our challenging setup, the best technique correctly detects 186 stops out of 199, estimates their duration with an average error of 3.4 s and their position with an extra 3.1 cm error w.r.t. the larger ones from UWB localization, which we identify as the main source of spatial error. In absence of ground truth, the estimated stop can be correctly associated to the closest exhibit in 88.9% of the cases. Overall, these findings provide positive, quantitative answers to our research questions, pushing the applicability of mobility analysis to unprecedented fine-grained spatio-temporal resolution.

Our survey of related efforts (§VII) shows that this work is the first studying stop-move detection on UWB-based trajectories and, importantly, to offer an evaluation against systematically-acquired ground truth in a real-world environment. We end the paper with concluding remarks (§VIII).

## II. PROBLEM FORMULATION AND DEFINITIONS

A *trajectory*  $T = \{(p_1, t_1) \dots (p_n, t_n)\}$  is a sequence of positions  $p_k$  and associated timestamps  $t_k$ , sampling the movement in space of an entity, a user in our case. We call  $(p_k, t_k)$  a *trajectory unit*. A *stop* or *segment* is a sub-sequence  $s = \{(p_i, t_i) \dots (p_j, t_j)\} \subseteq T$  capturing the user permanence in the area represented by the centroid of positions  $\{p_i \dots p_j\}$  during the time interval  $[t_i, t_j]$ . A *segmentation* of  $T$  is a sequence of temporally disjoint stops. Abstractly, the problem is to extract from a trajectory  $T$  the segmentation  $S = \{s_1 \dots s_m\}$  of *estimated* stops via a detection operation  $m(T, \rho, \Pi)$ . Stops shorter than the application-dependent temporal threshold  $\rho$  are irrelevant and neglected; the set  $\Pi$  of configuration parameters depends on the technique at hand, as described next.

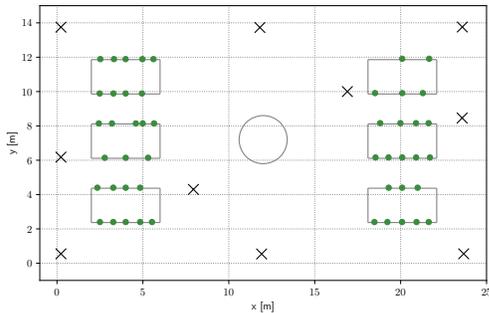


Fig. 1. Museum target area (top) and map (bottom). Dots are points of interest (POI) in front of exhibits; crosses are UWB anchors on the ceiling.

## III. SEGMENTATION TECHNIQUES

We summarize the stop-move detection techniques we compare (§VI) over our dataset, representative of state-of-the-art approaches (§VII) with different complexity and tradeoffs.

**Using spatial distance between units.** A stop  $s$  can be seen as the sequence of time-consecutive units  $\{(p_i, t_i) \dots (p_j, t_j)\}$  whose spatial distance from the start of the segment is smaller than an application-dependent threshold  $\delta \in \Pi$ , i.e.,  $|p_i - p_j| < \delta$ . Segments with duration  $[t_i, t_j] < \rho$  are ignored.

This approach was proposed in [15] as Stay Point Detection (SPD) to identify visits to point of interests in GPS trajectories, and is commonly used due to its simplicity. However, it is not well-suited when stops have different spatial size ( $\delta$ ) or are affected by outliers, e.g., due to positioning noise.

**Using unit density.** These limitations can be tackled by via density-based clustering. In DBSCAN [16], a *core point* in some abstract space has at least  $N$  neighbors within distance  $\epsilon$ ; a cluster contains these points and, transitively, those of neighboring core points. Unfortunately, when used for segmentation, these methods cannot guarantee the temporal separation of clusters except when spatially far from each other; recurrent stops, e.g., at the same exhibit, become indistinguishable.

SeqScan [17] overcomes this limitation by defining a stop as a cluster of units  $\{(p_i, t_i) \dots (p_j, t_j)\}$  where  $\{p_i \dots p_j\}$  is a DBSCAN cluster. Unlike SPD, clusters can thus have arbitrary spatial shape; unlike DBSCAN, SeqScan clusters have (disjoint) time intervals  $[t_i, t_j]$ ; those  $< \rho$  are ignored.

**Using user velocity (from Kalman filters).** Another way to look at stops is when user velocity is (nearly) zero. For generality, we do not determine it with sensors, rather compute it directly from UWB trajectories, whose noisy raw positions however induce unacceptable velocity jitter. Nevertheless, these trajectories are typically improved via Kalman filters (§V-A) whose operation already entails hidden state variables representing the velocity associated with units  $(p_i, t_i)$ . Segmentation then simply consists of identifying consecutive units whose velocity is greater than a threshold  $\theta \in \Pi$ , ignoring segments  $[t_i, t_j] < \rho$ . This Kalman-based velocity technique (hereafter, KBV) is, to our knowledge, novel in stop-move detection; it is interesting as a computationally cheap approach reusing the filtering often already applied to trajectories.

## IV. A NOVEL METRIC FOR STOP-MOVE DETECTION

Crucial to the practical application of segmentation techniques is a clear understanding of the quality they offer. This is important not only to compare alternatives but also their different parameter configurations. To provide reliable results, the output by a technique with a given configuration must be evaluated against a ground-truth segmentation; the question is how to measure the quality of the former vs. the latter.

**Baseline: Unit-centric metric.** A recent approach [13] is to quantify this quality at the *unit* level. A unit (§II) belonging to an estimated stop is a true positive (TP) if also part of a ground-truth stop (hereafter, true stop); a false positive (FP) if it is not. Dually, a unit belonging to no estimated stop is a false negative (FN) if part of some ground-truth stop; otherwise, it is

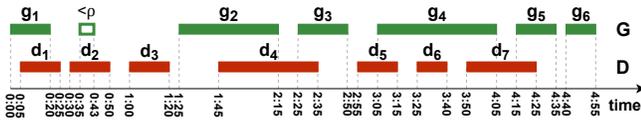


Fig. 2. An (artificial) example of segmentation.  $G$  is the set of time intervals of true stops,  $D$  the one for estimated stops.

a true negative (TN), hereafter disregarded. The values of TP, FP, FN can be used to compute popular metrics like precision  $P = \frac{TP}{TP+FP}$ , recall  $R = \frac{TP}{TP+FN}$ , and the  $F\text{-score} = 2 \frac{P \times R}{P+R}$  offering a single-value, concise quality indicator.

Unfortunately, this unit-centric approach is oblivious to the higher-level *structure* of the segmentation. By focusing on which *units* fall into the time interval of true stops, it does not capture properties of the estimated *stops* these units belong to. Even something as simple as the number of correctly identified stops is lost in the aggregated, unit-centric view.

**Contribution: A Stop-centric metric.** We propose an alternative metric that aims at directly *matching* each true stop with an estimated one. We use the same popular metrics above (TP, FP, FN, precision, recall, F-score), but defined at the level of stops rather than units. For instance, precision is directly the fraction of true stops over all estimated ones, rather than the fraction of units belonging to true stops over all units belonging to estimated stops. We argue, and confirm quantitatively (§VI), that this change in the “lens” used to analyze segmentations increases expressiveness and practical relevance. However, this approach also raises new problems, discussed next.

**Matching estimated and true stops (F-score).** Consider Fig. 2, an artificial example to illustrate all relevant cases, where  $G$  and  $D$  are the sets of time intervals associated to true stops and estimated stops, respectively. Our goal is to establish a one-to-one relationship between them, based on the intuition that if  $g \in G$  and  $d \in D$  represent the same stop, their time intervals must overlap (ideally, coincide). Still, an estimated stop can overlap multiple true stops and vice versa. How can the matching be performed?

We cast the problem as an unbalanced assignment optimization over the bipartite graph  $(G, D, E, W)$ , where  $E$  is the set of edges  $(g_i, d_j)$  linking overlapping stops, and  $W$  the set of associated weights  $w_{ij} = \frac{g_i \cap d_j}{g_i}$  capturing how much of the true stop  $g_i$  overlaps with the estimated one  $d_j$ . The optimal assignment yields the *matching set*  $M$  maximizing the sum of weights. Applying these concepts to Fig. 2 yields the graph in Fig. 3. Solving the assignment problem yields  $M = \{(g_1, d_1), (g_2, d_4), (g_4, d_6), (g_5, d_7)\}$ , from which  $TP = \{d_1, d_4, d_6, d_7\}$ ,  $FP = \{d_2, d_3, d_5\}$ ,  $FN = \{g_3, g_6\}$  can be derived, along with precision, recall, and F-score.

**Determining the nature of false detections.** This can be inferred for a segmentation technique based on the relationship between an estimated segment  $d$  and a true one  $g$ .

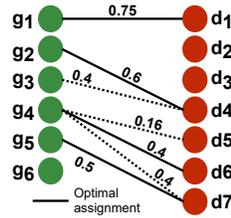


Fig. 3. Finding the matching set.

A false positive  $d$  is an actual *fake* if it does not overlap with any true stop  $g$  (e.g.,  $d_3$  in Fig. 2). Instead, when  $d \cap g \neq 0$  the FP is a *split* stop, i.e., a true stop is estimated as two or more separate ones (e.g.,  $g_4$  matched by  $d_6$  and overlapping with  $d_5$ ). Indeed, as  $d$  is a FP it does not match any  $g$  but some other  $d_i \neq d$  must, otherwise the assignment would not be optimal. A special case is a *short* stop, when the duration of  $g$  is  $< \rho$  (i.e., irrelevant, §II) but the one of  $d$  is not ( $d_2$ ).

Similarly, a false negative  $g$  is an actual *missing* stop ( $g_6$ ) when  $g \cap d = 0$ , otherwise it is a *merged* stop, i.e., two or more true stops estimated as a single one (e.g.,  $g_3$  lumped into the same estimate  $d_4$  matched to  $g_2$ ). Indeed,  $g$  is not matched by  $d$  although they overlap; thus,  $d$  must match another  $g_i \neq g$  otherwise, again, the assignment would not be optimal. As in the FP case, a FN short stop could capture a true  $g$  incorrectly estimated by  $d$  with duration  $< \rho$ . However, in practice, these are automatically filtered by segmentation techniques.

**Quantifying the similarity of matched stops (S-score).** In our example,  $M$  contains both  $(g_1, d_1)$  and  $(g_2, d_4)$ , whose temporal overlapping between estimated and true stops is remarkably different (Fig. 2). Still, another segmentation yielding the same  $M$  but with  $(g_1, d_1)$  and  $(g_2, d_4)$  perfectly temporally aligned would yield the same F-score. This indicator is therefore an expressive measure of *correctness*, but does not capture how *similar* matched stops are. Yet, accurate detection of temporal intervals is key to many analyses and also impacts the accuracy of the corresponding stop positions, as discussed later (§VI). Therefore, we complement the F-score with the

$$S\text{-score} = \frac{1}{|M|} \sum_{(g,d) \in M} \frac{g \cap d}{g \cup d}$$

The summation argument is the Jaccard index over the intervals  $g$  and  $d$  associated to true and estimated stops, respectively, a common way to capture their similarity. Its average over the matching set yields the S-score, whose value is in  $(0, 1]$ ; it cannot be 0 because stops belong to the matching set and is 1 only when they are perfectly aligned.

## V. DATASET ACQUISITION AND CHARACTERISTICS

Before applying the proposed metric to configure and compare segmentation techniques on our dataset, we describe the methodology used for collecting it and characterize its content.

### A. Collecting the Dataset

**UWB localization system.** We use TALLA [14], a state-of-the-art time-difference-of-arrival (TDoA) localization system. TDoA can support many users as a *single* UWB packet broadcast by the mobile tag enables position computation by the fixed localization anchors in the target area, based on their different packet time of arrival induced by signal propagation along different distances. However, positioning accuracy depends on the tight time synchronization of anchors, typically achieved via dedicated and costly wired infrastructure. TALLA can localize many mobile tags over large-scale areas spanning several anchors and with wireless-only time synchronization, all highly desirable features for adoption by the museum.

All nodes follow a common TDMA schedule. Anchors exchange beacon packets whose period (*epoch*) is key to time synchronization and thus positioning accuracy; beacons are also received by nearby tags, ensuring they are time-aligned. Each tag owns at least one slot where it broadcasts a location packet, whose time of arrival is timestamped at all anchors. These timestamps, along with the reference ones acquired during synchronization, are used by a localization server to estimate the tag position via a TDoA least-squares solver.

**Improving position estimates via Kalman filters.** UWB measurements are affected by errors, yielding noisy trajectories commonly “smoothed” via Kalman filters [7].

Noise exists both when the tag moves and when stationary; Kalman filters are usually optimized for either case. This clashes with stop-move detection, which requires efficient noise reduction in both cases *and* a fast switch between them to accurately determine stop start/end times. Therefore, we combine two Unscented Kalman filters (UKF) representing the tag mode (stopped or moving) under the framework of Interacting Multiple Models (IMM) [18]. The output position is a linear combination of both filter estimates, weighted by the probability of each filter to match the current tag behavior.

Hereafter, we refer to the trajectories output by TALLA as *raw* and to those post-processed via IMM-UKF as *filtered*. Fig. 4 exemplifies their difference, whose impact on stop positions (§V-B) and segmentations (§VI) we analyze later.

**System configuration and target area.** User tags are battery-powered and anchors mains-powered; both are DWM1001 nodes by Decawave (now Qorvo) equipped with the popular DW1000 UWB radio [19]. Each anchor is connected via USB to a Raspberry Pi, relaying TDoA data to the localization server. The UWB anchors are deployed on the ceiling of a  $25 \times 15$  m<sup>2</sup> area (Fig. 1). Those on the perimeter ensure low geometric dilution of precision (GDOP) and thus high positioning accuracy. The two near the center improve position diversity, mitigating the impact of radio signal occlusions.

We configured TALLA with 250 ms epochs (4 Hz time synchronization) and tags broadcasting 3 times/epoch (12 Hz position update rate), and the UWB radio with the recommended channel 5, a 64 MHz pulse repetition frequency (PRF) and a 128  $\mu$ s preamble.

**Data collection methodology and ground truth.** The target area contains a large globe surrounded by 6 tables hosting

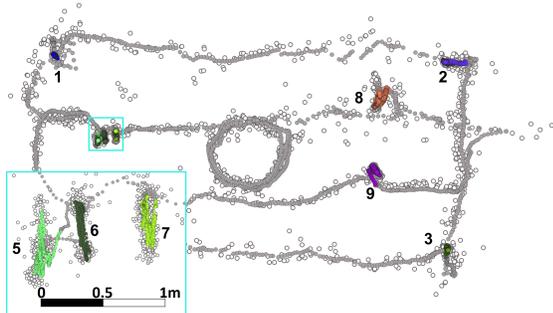


Fig. 4. Raw (white) vs. filtered (gray) trajectory. The colored points in the latter fall in ground-truth stop time intervals.

many small-size exhibits, of which 44 were visited. Members of the research team emulate the behavior of visitors by repeatedly moving in the area then stopping in front of some exhibit. Each user wears a necklace with a UWB tag on the chest, a natural option for a real use. The position of the tag is the one actually recorded by the UWB localization system.

Collecting reliable ground truth is challenged by mobility. We obtained accurate *spatial* data by placing floor stickers at all point of interests marking user stops (Fig. 1) and acquiring with a laser meter their position (hereafter, POI). As for *temporal* ground truth, a smartphone application enables each user to record arrival/departure times for each POI by toggling a button. We synchronized the smartphone and TALLA clocks, obtaining a common time reference for ground truth and UWB trajectories. Moreover, we placed 2 tripod-mounted cameras with 180° angle on opposite sides, covering the entire area, whose videos enabled validation of the smartphone data.

**Dataset content.** A UWB trajectory, the input for segmentation, contains units (§II) in the form  $((x, y), t)$ . For each timestamp  $t$  we collect both raw and filtered  $(x, y)$  positions (Fig. 4). The ground-truth segmentation is represented by a sequence of stops  $(ID, t_{arr}, t_{dep})$ , containing the arrival/departure times to/from the POI with identifier  $ID$ , separately associated to its ground-truth position. This unambiguously identifies the stop location and accounts for recurrent visits. We collected 9 trajectories of similar duration ( $\sim 11$  mins) for a total of 70090 units over 100.03 mins. The number of stops differs across trajectories, from 11 to 29, for a total of 209 stops.

### B. Characterizing the Dataset

**Spatio-temporal characteristics.** Fig. 6a shows *temporal* features via the cumulative distribution function (CDF) of stop durations in the ground truth. The chart substantiates the claim (§I) that our dataset contains very short stops, with a median of 12.4 s. The red line denotes the threshold below which durations become irrelevant for the application (§II), set to  $\rho=10$  s based on requirements by the museum curators. The 10 stops (<5%) below it should not be detected, leaving 199 true stops as the expected ideal segmentation output (§III).

As for *spatial* characteristics, exhibits (POIs, §V-A) are physically very close (Fig. 1). Fig. 6b shows that 80.6% of adjacent POIs are within 1 m, with a maximum of 1.59 m. Again, this challenging dataset demands *high spatial resolution* in discerning stops when segmenting trajectories. Still, the distance between *consecutive stops within a trajectory* varies significantly (POI-POI line, Fig. 6c) as exhibits are not necessarily visited in order. Fig. 5 shows the filtered trajectory in Fig. 4 with an extra time dimension. The user mimicking visitors mixes short strides to adjacent exhibits with longer ones (e.g., to join friends or avoid crowds), including one around the central globe (e.g., to observe it from all angles).

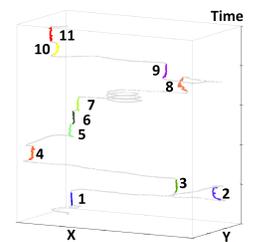


Fig. 5. Spatio-temporal view of Fig. 4 (filtered).

**Ground-truth stops vs. estimated UWB centroids.** Fig. 5 illustrates another key point: each UWB trajectory contains *several* positions (in color) for a stop, i.e., falling inside the interval  $[t_{arr}, t_{dep}]$  whose ground-truth value is reliably determined via smartphone and cameras. Ideally, the *centroid* of these UWB positions for a POI matches exactly the ground-truth one; in practice, this is not the case. The POI is very accurately measured with a laser meter in a fixed position; the positions yielding the centroid are measured for a moving tag and with larger UWB errors. Their main source is the user body, creating non-line-of-sight (NLOS) between the tag on the chest and the anchors behind the back. This is crucial when one of them is the main time reference; manually changing the latter when in NLOS reduces the mean positioning error by 25%. NLOS mitigation techniques, an active topic of research [20], [21], could be incorporated in TALLA and yield improvements; however, it is beyond the scope of this paper.

Nevertheless, the error between the POI ground-truth and centroid positions (Fig. 6d) remains sub-meter in 96.2% of the cases, i.e., significantly better than techniques based on WiFi and BLE, plagued by errors of several meters [7]. For both raw and filtered trajectories, the median and mean error are 42 cm and 46 cm, respectively; the commonly-used 75<sup>th</sup> percentile is 57 cm. Interestingly, these metrics are within few percents, i.e., *the smoothing induced by Kalman filters does not affect the position of the UWB stop centroid*.

**Trajectory structure: A key observation.** Fig. 6c compares the distance between consecutive stops computed using ground-truth POI vs. UWB centroids; their difference is negligible, despite the errors affecting the latter and the close placement of exhibits (Fig. 6b). Put differently, *the UWB positioning error does not modify the structure of trajectories*; the sequence of stops contained in each trajectory is essentially the same regardless of whether we express it via ground-truth positions or estimated UWB centroids. This fact has important implications in deriving our findings, as discussed next.

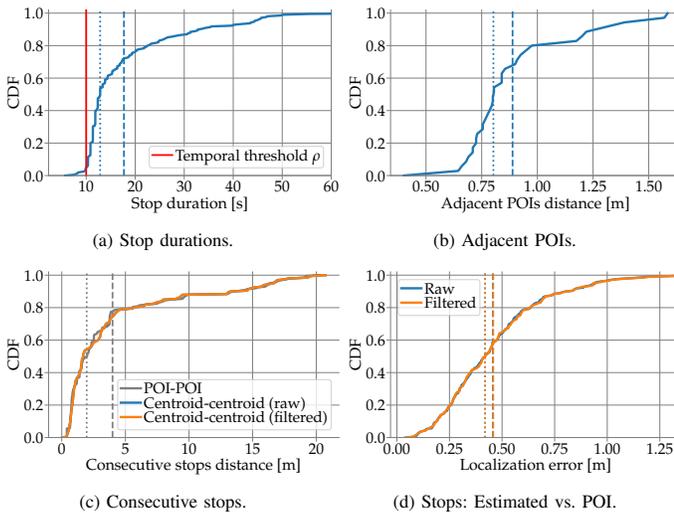


Fig. 6. Spatio-temporal characteristics of the dataset. Dashed lines represent the mean and dotted lines the median.

## VI. FINDINGS

We study the quality of reference segmentation techniques (§III) against the yardstick of our novel metric (§IV) using our dataset (§V) as the means to distill quantitative findings.

**What to compare?** The ground truth in our dataset consists of stops  $(ID, t_{arr}, t_{dep})$  where the times of arrival/departure from a known POI are accurately determined. However, segmentation techniques operate on UWB trajectories whose positions are less accurate than ground-truth POIs, with an error depending on the specific UWB localization system used.

Fig. 7 illustrates the methodological problem. The time interval  $[t_{arr}, t_{dep}]$  spent at a POI is known precisely from ground truth. A segmentation estimating it perfectly is still affected by spatial error: the distance (#1) between the centroid of units in  $[t_{arr}, t_{dep}]$  and the POI. However, real segmentations introduce errors w.r.t.  $[t_{arr}, t_{dep}]$  and thus the centroid computed over it, causing additional spatial error (#2).

Our metric focuses on the *temporal* dimension by matching estimated and ground-truth intervals and *directly* associating them to the *ID* of true POI positions. This may seem to neglect the spatial dimension of segmentation quality. However, we observed (§V-B) that UWB trajectories and ground-truth ones have the same spatial structure; the temporal structure is therefore key. Moreover, our choice sharply decouples the quality of segmentation from the one of localization; separating the errors they induce would be impossible otherwise.

Consequently, we first use our metric to compare the temporal errors of segmentations. Then, leveraging the above decoupling, we study their impact on both the position of the estimated centroid vs. the UWB one in the dataset *and* vs. the POI position. This accounts for both time and space, ultimately providing precious information for practical use.

**Novel metric: Is it worth?** We confirm the higher expressiveness of our stop-centric metric by observing that *segmentations with the same unit-centric quality can have different stop-centric quality, and vice versa*. For illustration only, we focus on SeqScan segmentations of sample trajectories.

Fig. 8 shows two segmentations of the same trajectory. The unit-centric metric assigns them the same F-score = 0.90; yet, the spatio-temporal maps show that they are very different. Segmentation  $\langle 15, 24 \rangle$  correctly detects all 28 true stops. Instead,  $\langle 15, 12 \rangle$  lumps 8 distinct true stops into 2 large estimated ones, representing incorrectly the user behavior. The unit-centric metric is oblivious to *structure*, as it considers only

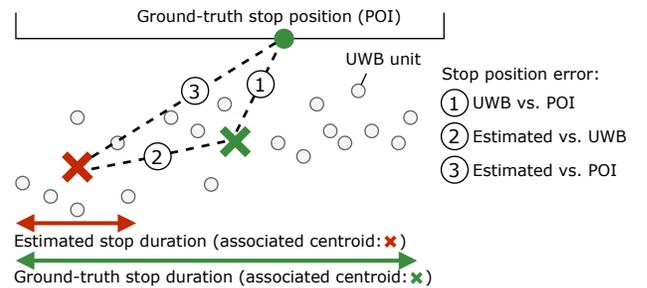


Fig. 7. UWB trajectories, ground-truth vs. estimated stops.

whether individual units belong to *any* stop. Conversely, our stop-centric metric accounts for the 6 false negatives in the second segmentation with a lower F-score than the first.

Fig. 9 illustrates the dual situation on another trajectory. The unit-based metric assigns a higher F-score to  $\langle 20, 36 \rangle$ ; one could infer it detects *more* stops than  $\langle 10, 36 \rangle$ . Instead, both detect 21 true stops with 1 FN;  $\langle 10, 36 \rangle$  actually yields more accurate duration estimates, as shown for one sample stop in the spatio-temporal maps. These aspects are faithfully captured by our metric via the F-score and S-score, respectively.

This higher expressiveness has practical implications. For instance, in our museum context, some analyses may focus on *how many* exhibits are visited, others on *how long* each visit is. Our metric clearly distinguishes the two aspects, guiding the choice of the most appropriate technique and/or configuration.

Based on all these considerations, hereafter we report only the results obtained with our proposed stop-centric metric.

**Segmentation: Which parameters for what quality?** We ascertained the impact on quality of several configurations for each technique. Table I shows results over the entire dataset for filtered trajectories. The highlighted best configurations are those with highest F-score and, when equal, highest S-score, e.g., as in  $\langle 10, 12 \rangle$  and  $\langle 15, 24 \rangle$  for SeqScan. Again, alternative criteria striking different quality tradeoffs are possible.

All methods yield good quality. KBV has the lowest one and is the most sensitive to its  $\theta$  parameter; yet, it is the cheapest computationally (§III). At the other extreme, SeqScan yields the highest quality and its two-parameter configuration increases flexibility. Table I also reports the unit-centric metric, confirming its lower expressiveness. This is evident for KBV, whose highest F-score is obtained with  $\theta=100$  that *i)* detects only 123 out of 199 true stops, yet *ii)* has nearly the same F-score of the best SPD configuration, detecting 189.

**Raw vs. filtered trajectories: Does it matter?** Filtered trajectories reduce spatial jitter vs. raw ones (Fig. 4) but also induce

TABLE I  
EXPLORING PARAMETERS FOR FILTERED TRAJECTORIES;  
 $\delta$  AND  $\epsilon$  ARE IN CM,  $\theta$  IN CM/S.

	Stop-centric							Unit-centric
	TP	FP	FN	Precision	Recall	F-score	S-score	F-score
$\delta$	SPD							
20	116	17	83	0.872	0.583	0.699	0.714	0.630
30	161	16	38	0.910	0.809	0.856	0.772	0.777
40	175	19	24	0.902	0.879	0.891	0.790	0.822
50	181	19	18	0.905	0.910	0.907	0.797	0.835
60	189	18	10	0.913	0.950	0.931	0.778	0.854
70	187	21	12	0.899	0.940	0.919	0.757	0.850
80	187	18	12	0.912	0.940	0.926	0.749	0.850
90	183	17	16	0.915	0.920	0.917	0.730	0.849
$\epsilon, N$	SeqScan							
10, 12	186	8	13	0.959	0.935	0.947	0.810	0.871
15, 12	176	11	23	0.941	0.884	0.912	0.774	0.872
20, 12	165	13	34	0.927	0.829	0.875	0.741	0.869
10, 24	176	8	23	0.957	0.884	0.919	0.815	0.842
15, 24	187	9	12	0.954	0.940	0.947	0.796	0.870
20, 24	177	11	22	0.941	0.889	0.915	0.770	0.871
10, 36	166	6	33	0.965	0.834	0.895	0.813	0.824
15, 36	183	8	16	0.958	0.920	0.938	0.803	0.859
20, 36	178	10	21	0.947	0.894	0.920	0.777	0.869
$\theta$	KBV							
10	31	4	168	0.886	0.156	0.265	0.594	0.42
20	110	6	89	0.948	0.553	0.698	0.729	0.658
30	154	4	45	0.975	0.774	0.863	0.766	0.719
40	168	11	31	0.939	0.844	0.889	0.781	0.801
50	175	10	24	0.946	0.879	0.911	0.781	0.831
60	168	10	31	0.944	0.844	0.891	0.751	0.844
70	162	9	37	0.947	0.814	0.876	0.724	0.846
80	142	7	57	0.953	0.714	0.816	0.684	0.847
90	129	9	70	0.935	0.648	0.766	0.650	0.851
100	123	11	76	0.918	0.618	0.739	0.629	0.855

the same stop-move structure (Fig. 6); it is unclear whether and how they affect segmentation quality. Therefore, we performed for raw trajectories the same parameter exploration of Table I, except for KBV. The best configurations for SPD ( $\delta=80$ ) and SeqScan ( $\epsilon=15, N=24$ ) respectively detect 8 and 10 fewer TP with an increase in FN, despite exploiting higher values of both distance ( $\delta, \epsilon$ ) and number of points ( $N$ ) to account for the higher position dispersion. This confirms that filtered trajectories yield higher quality, although the difference is not dramatic, as shown also by the other metrics (Table II).

**What is the nature of false positives/negatives?** The F-score of our stop-centric metric offers further insights on quality by expressing the *type* of false detections (§IV). The analysis in Table I shows that SPD detects nearly as many true stops as SeqScan (more, on raw trajectories) but with more false positives, lowering precision and F-score. Table II now clearly shows that the culprit are split stops, a known weakness of the method. Moreover, all techniques are equally sensitive to stops with duration shorter than  $\rho=10$  s; interestingly, this is the main source of mis-detection in SeqScan. Finally, fake stops are surprisingly rare, even absent with KBV.

Dually, missing stops are the main source of FN for all techniques. SeqScan and SPD achieve similar results; the latter is more sensitive to spatial resolution. A smaller  $\delta$  does not affect split and merge stops but increases missing ones; with  $\delta=20$  (not shown), they become the only FN source. In contrast, the two-parameter structure of SeqScan achieves

$\epsilon, N$	Stop-centric					Unit-centric			
	TP	FP	FN	F-score	S-score	TP	FP	FN	F-score
15, 24	28	0	0	1.00	0.79	4953	932	80	0.90
15, 12	22	0	6	0.88	0.74	5013	1087	20	0.90

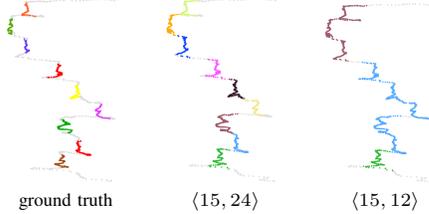


Fig. 8. Different stop properties, same unit quality.

$\epsilon, N$	Stop-centric					Unit-centric			
	TP	FP	FN	F-score	S-score	TP	FP	FN	F-score
10, 36	21	0	1	0.98	0.84	4348	483	1228	0.84
20, 36	21	0	1	0.98	0.83	5375	718	201	0.92



Fig. 9. Similar stop properties, different unit quality.

TABLE II  
SEGMENTATION TECHNIQUES AT A GLANCE: STOP-CENTRIC METRIC AND ADDITIONAL INSIGHTS.

Technique	Dataset	Main metric		TP	Nature of false detections				Spatio-temporal errors (in TP)								Correct POI		
		F-score	S-score		FP		FN		$t_{start}$		$t_{end}$		$\Delta t$		$\Delta s$		TP only	overall	
					split	short	fake	merged	missing	$\mu$	$\sigma$	$\mu$	$\sigma$	$\mu$	$\sigma$	$\mu$			$\sigma$
SPD	filtered	0.931	0.778	189	10	7	1	3	7	-0.83	4.60	1.16	3.61	2.00	6.65	4.11	6.12	93.1	88.4
	raw	0.903	0.764	181	13	7	1	5	13	-1.01	3.82	1.33	5.04	2.35	7.25	4.55	6.71	92.8	84.4
SeqScan	filtered	0.947	0.810	186	1	6	1	4	9	-1.64	2.53	1.78	4.05	3.43	4.93	3.12	4.45	95.2	88.9
	raw	0.921	0.779	176	1	5	1	8	15	-3.03	3.40	1.98	4.82	5.02	6.41	6.24	33.1	92.6	81.9
KBV	filtered	0.911	0.781	175	5	5	0	7	17	0.26	5.21	1.99	4.96	1.72	8.01	4.26	8.42	92.0	80.9

high quality with similar (or even lower) spatial resolution  $\epsilon$ , slightly increasing merged stops in other less-performant configurations (not shown). Finally, the many missing stops in KBV are due to its reliance on velocity rather than distance, frequently changing around the threshold  $\theta$ . This parameter crucially affects the nature of FN, dominant in KBV (Table I); a value of 100 cm/s yields a majority (40) of merged stops, while 10 cm/s yields all missing stops.

**How temporal errors affect spatial ones?** We defined the S-score as a concise indicator of the temporal overlapping between true stops and estimated ones (TP). However, it does not account for the *absolute* error in the temporal alignment between the estimated and true stops. Here, we analyze this aspect along with the impact it bears on spatial error (Fig. 7).

Given an estimated segment  $[t_1, t_2]$  and a ground-truth one  $[t_a, t_b]$  of duration  $t_d=t_2-t_1$  and  $t_g=t_b-t_a$ , we consider the errors in duration  $\Delta t=t_d-t_g$ , start  $t_{start}=t_1-t_a$ , and end  $t_{end}=t_2-t_b$ . Moreover, we consider the corresponding spatial error  $\Delta s=|p_d-p_g|$ , i.e., the (absolute) distance (#2 in Fig. 7) between the centroids  $p_d$  and  $p_g$  of UWB positions falling in  $t_d$  and  $t_g$ . Table II reports their mean  $\mu$  and standard deviation  $\sigma$  in the best configurations; Fig. 10 shows the CDFs of  $\Delta t$  and  $\Delta s$  of filtered trajectories only, due to space limitations.

All techniques perform well, with errors of few seconds and centimeters. KBV is the most accurate temporally, with a mean error  $\mu=1.72$  s. Yet, its mean spatial error  $\Delta s$  is the highest among filtered trajectories; it is very near to SPD, whose median is however significantly worse. At the other extreme, SeqScan yields the worst duration estimates;  $\mu$  is nearly twice w.r.t. KBV, although the absolute difference is  $<2$  s. Nevertheless, it is the most accurate spatially—a counterintuitive result explained by observing that *i*) SeqScan is robust to outliers by design, intrinsically reducing spatial noise *ii*) temporal precision ( $\sigma$ ) is the highest *iii*)  $t_{start}$  is underestimated and  $t_{end}$  overestimated, both in median (not shown) and mean by nearly the same amount, which tends to center the true stop inside the estimated one, reducing the distance between the corresponding centroids (Fig. 7).

Fig. 10a also shows that SPD often severely underestimates stop duration, likely the culprit for the many stop splits (Table II). Nevertheless, its performance in terms of  $\Delta t$  and  $\Delta s$  does not change significantly when moving from filtered to raw trajectories. This is not the case for SeqScan, whose metrics for the latter (Table II) are nonetheless heavily affected by a *single* outlier, caused by the merging of distant stops, whose removal yields  $\mu=3.82$  cm and  $\sigma=7.74$  cm for  $\Delta s$ .

Anyway, this is in line with false detections (Table II); while SPD is prone to stop splitting, SeqScan is to merging.

**Can estimated stops be correctly associated to POIs?**

Among the distances in Fig. 7, we analyzed #1 in our dataset (Fig. 6d) and #2 by reporting  $\Delta s$  (Table II); we now investigate their combination, #3. Its value is not very informative, given that  $\Delta s$  is very small and so is the difference between #1 and #3; for SPD, the worst-case (filtered), is 46 cm median and 52 cm mean, only few centimeters more than in Fig 6d.

Instead, the crucial question is: *Can we correctly identify the POI visited by the user via the estimated centroid, i.e., without ground truth?* This of practical relevance, as it is how a real system would work. Intuitively, the association can consist simply of determining the POI *closest* to the estimated stop.

This is challenging in our setup with *i*) POIs close to each other (Fig. 6b) and *ii*) non-negligible UWB positioning error (Fig. 6d). Still, Table II shows that stops can be accurately associated to POIs. Considering only TP, the POI closest to the estimated stop is correct in  $>92\%$  of the cases, with a maximum of 95.2% for SeqScan. However, techniques differ in their ability to identify TP, reducing the fraction of overall correct associations, nonetheless always  $>80\%$ . KBV is the worst, due to its lowest TP; for the same reason, SeqScan is the best at 88.9%, although its accuracy on raw trajectories degrades below SPD, reasserting the importance of filtering.

VII. RELATED WORK

**Stop-move detection.** Trajectory segmentation has been applied to GPS trajectories for a long time [22]. We considered techniques (§III) representative of two main classes: *criteria-based* define stops based on, e.g., distance, time duration, velocity, as in SPD and KBV, but also [22], [23]; *cluster-based* include SeqScan and, e.g., [24]–[26]. Others are based on statistical models [13] or do not even rely on segmentation [27]. Our focus is not on exhaustive comparison, rather on *quantifying* the quality attainable w.r.t. ground truth when

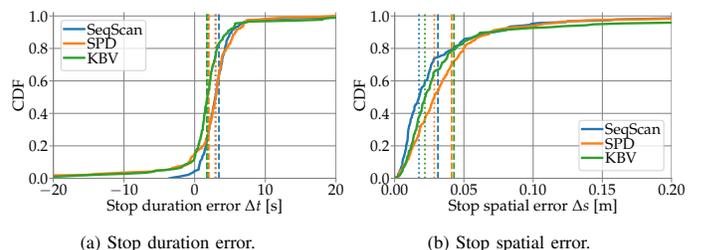


Fig. 10. Duration and spatial error for filtered trajectories.

applying stop-move techniques to UWB, and contributing a novel metric better eliciting tradeoffs and challenges.

**Ground truth and metrics.** Evaluation against ground truth is hindered by the complexity induced by mobile targets (e.g., a migrating animal). Validation is often only qualitative, e.g., through visual inspection of trajectories, or via simple metrics like the number of stops [24], [28]. This may suffice over large-scale areas with well-separated stops, but is difficult in our setup where stops are close both in space and time. The recent unit-centric approach [13] offers a quantitative metric, but without a real-world validation.

Our stop-centric metric (§IV) is significantly more expressive, as confirmed by the analysis of our dataset gathered in a real museum and with accurately acquired ground truth. The use of F-score on stops instead of units was also proposed in [25], [26]. However, these works do not state how a true stop is measured, likely demanding it to qualitative considerations. Instead, we provide definitions and methods enabling both automated quantitative analysis and interpretation of false detections, along with a measure of stop similarity, reuniting all relevant dimensions in a single methodological framework.

**Stop-move and museums.** Museums are a natural application for stop-move detection, crucial to understanding visitor behavior. However, reported experiences are based on Bluetooth, whose poor positioning accuracy [7] forces coarser trajectory models whose units are entry/exit or proximity events instead of positions. In [9], nodes in key areas of the Louvre enabled analysis of the overall stay in the museum (hours) and frequency of visits to areas. Similar experiences determined the stay in a room [12] or “hotspots” and other macro-level indicators [10]. These approaches extract spatio-temporal features from the Bluetooth signal, increasing complexity and reducing accuracy. In contrast, we showed quantitatively how the higher accuracy of UWB directly translates in a much greater spatio-temporal resolution in discriminating stops.

## VIII. CONCLUSIONS

The high spatio-temporal accuracy of UWB localization intuitively enables fine-grained detection of stop-move patterns, key to many applications. Yet, this opportunity has not been studied, let apart quantitatively and experimentally. This is our goal, exploiting a museum deployment with accurate ground truth, also rare in the literature. The findings, albeit not directly generalizable, concretely inform about the quality attainable in practice in a real-world setting.

We define a novel, expressive metric relating estimated and true stops, enabling the configuration and comparison of segmentation techniques originally targeting coarser-grained scenarios. We show that, once applied to UWB trajectories, they induce only small spatio-temporal errors of few centimeters and seconds. Therefore, we identify the UWB localization system, not segmentation, as the main source of spatial error, likely mitigated by continuous progress in UWB research. Nevertheless, estimated stops are correctly associated to true ones in the vast majority of cases, enabling fine-grained UWB-based stop-move detection in this and other practical contexts.

## ACKNOWLEDGMENTS

We are deeply grateful to Michele Lanzingher and Vittorio Cozzio for making this study possible at MUSE. We thank A. Bacchiega, M. Fenu, A. Giovannone, T. Istomin, and D. Molteni for their help on various technical and experimental issues. This work is partially supported by the Italian government via the NG-UWB project (MIUR PRIN 2017).

## REFERENCES

- [1] S. Wang *et al.*, “A Survey on Trajectory Data Management, Analytics, and Learning,” *ACM Computing Surveys*, 2021.
- [2] Y. Zheng and X. Zhou, *Computing with spatial trajectories*. Springer, 2011.
- [3] R. H. Güting *et al.*, “A foundation for representing and querying moving objects,” *ACM Trans. on Database Systems*, 2000.
- [4] E. Zimányi *et al.*, “MobilityDB: A Mobility Database Based on PostgreSQL and PostGIS,” *ACM Trans. on Database Systems*, 2020.
- [5] Y. Zheng, “Trajectory data mining: an overview,” *ACM Trans. Intelligent Systems and Technology*, 2015.
- [6] C. Parent *et al.*, “Semantic trajectories modeling and analysis,” *ACM Computing Surveys*, 2013.
- [7] F. Zafari *et al.*, “A Survey of Indoor Localization Systems and Technologies,” *IEEE Comm. Surveys & Tutorials*, vol. 21, no. 3, 2019.
- [8] S. Spaccapietra *et al.*, “A conceptual view on trajectories,” *Data & Knowledge Eng.*, 2008.
- [9] Y. Yoshimura *et al.*, “Noninvasive Bluetooth monitoring of visitors’ length of stay at the Louvre,” *IEEE Pervasive Computing*, 2017.
- [10] C. Martella *et al.*, “Visualizing, clustering, and predicting the behavior of museum visitors,” *Pervasive and Mobile Computing*, 2017.
- [11] A. Kontarinis *et al.*, “Towards a semantic indoor trajectory model: application to museum visits,” *Geoinformatica*, 2021.
- [12] P. Centorrino *et al.*, “Managing crowded museums: Visitors flow measurement, analysis, modeling, and optimization,” *J. of Computational Science*, 2021.
- [13] L. Bermingham and I. Lee, “A probabilistic stop and move classifier for noisy GPS trajectories,” *Data Mining and Knowledge Discovery*, 2018.
- [14] D. Vecchia *et al.*, “TALLA: Large-scale TDoA Localization with Ultra-wideband Radios,” in *Proc. of IPIN*, 2019.
- [15] Y. Zheng *et al.*, “Recommending friends and locations based on individual location history,” *ACM Transactions on the Web*, 2011.
- [16] M. Ester *et al.*, “A density-based algorithm for discovering clusters in large spatial databases with noise,” in *Proc. of KDD*, 1996.
- [17] M. L. Damiani *et al.*, “Cluster-based trajectory segmentation with local noise,” *Data Mining and Knowledge Discovery*, 2018.
- [18] H. A. Blom and Y. Bar-Shalom, “The interacting multiple model algorithm for systems with Markovian switching coefficients,” *IEEE Trans. on Automatic Control*, 1988.
- [19] DecaWave Ltd., “DW1000 Data Sheet, version 2.19,” 2017.
- [20] K. Yu *et al.*, “A novel NLOS mitigation algorithm for UWB localization in harsh indoor environments,” *IEEE Trans. on Vehicular Tech.*, 2018.
- [21] C. Di Franco *et al.*, “Calibration-Free Network Localization using Non-Line-of-Sight Ultra-Wideband Measurements,” in *Proc. of IPSN*, 2017.
- [22] J. Kang *et al.*, “Extracting places from traces of locations,” in *Proc. of ACM Int. Wkshp. on Wireless Mobile Applications and Services on WLAN hotspots*, 2004.
- [23] M. Buchin *et al.*, “Segmenting trajectories: a framework and algorithms using spatiotemporal criteria,” *J. of Spatial Information Science*, 2011.
- [24] A. T. Palma *et al.*, “A clustering-based approach for discovering interesting places in trajectories,” in *Proc. of ACM SAC*, 2008.
- [25] S. Hwang *et al.*, “Segmenting human trajectory data by movement states while addressing signal loss and signal noise,” *Int. J. of Geographical Information Science*, 2018.
- [26] T. Luo *et al.*, “An Improved DBSCAN Algorithm to Detect Stops in Individual Trajectories,” *ISPRS Int. J. of Geo-Information*, 2017.
- [27] Z. Li *et al.*, “MoveMine: mining moving object data for discovery of animal movement patterns,” *ACM Trans. Intelligent Systems and Technology*, 2011.
- [28] S. Hwang *et al.*, “Detecting Stop Episodes from GPS Trajectories with Gaps,” in *Seeing Cities Through Big Data*. Springer, 2017.