Real-Time Detection of Vehicle-Based Logistics Operations

Joel Ribeiro[1(✉)], Jorge Tavares[1,2], and Tânia Fontes[1]

[1] INESC TEC, Rua Dr. Roberto Frias, 4200-465 Porto, Portugal
{joel.ribeiro,tania.d.fontes}@inesctec.pt
[2] Faculty of Engineering, University of Porto,
Rua Dr. Roberto Frias, 4200-465 Porto, Portugal

Abstract. Geolocation data is fundamental to businesses relying on vehicles such as logistics and transportation. With the advance of the technology, collecting geolocation data become increasingly accessible and affordable, which raised new opportunities for business intelligence. This paper addresses the application of geolocation data for monitoring logistics processes, namely for detecting vehicle-based operations in real time. A stream of geolocation entries is used for inferring stationary events. Data from an international logistics company is used as a case study, in which operations of loading/unloading of goods are not only identified but also quantified. The results of the case study demonstrate the effectiveness of the solution, showing that logistics operations can be inferred from geolocation data. Further meaningful information may be extracted from these inferred operations using process mining techniques.

Keywords: Geolocation data · Event detection · Logistics operations

1 Introduction

Specific areas of the transportation sector such as fleet management, are being revolutionised by data and digital transformation. The use of telematics in fleets increased significantly in last years, from 48% in 2017 to 86% in 2019 [1]. As just 23% of fleets used big data analytics to guide strategic decision-making, new opportunities arise for the businesses to improve their fleets' operations.

Geolocation data rely on a geographic coordinate system such as the GPS (Global Positioning System), which define the positions on Earth. The inference of events using geolocation data can provide a valuable insight into business processes, especially in transportation and logistics, for example for the analysis of routes, vehicles and drivers [11]. The identification of travel events and motivations has been studied over the last few decades [14].

Focusing on public transportation, Pinelli et al. [9] proposed a methodology to detect the location of the scheduled and unscheduled bus stops using

A. L. Martins et al. (Eds.): INTSYS 2021, LNICST 426, pp. 192–205, 2022.
https://doi.org/10.1007/978-3-030-97603-3_14

the speed and acceleration variation, distinguishing then the type of stop relying on density-based clustering. Biagioni et al. [5] also identifies the location of bus stops with an automatic system which uses GPS traces collected from smartphones, identifying the routes and inferring schedules afterwards. Gong et al. [6] developed a methodology to identify activity stops in continuous GPS trajectories collected using mobile phones and combining clustering techniques with Support Vector Machines (SVM). Tavares et al. [13] defined and evaluated diverse approaches for identifying – from geolocation data – relevant locations where vehicles stopped.

Focusing on logistics, Pluvinet et al. [10] defined a GPS-based data collection technique and data processing algorithm to identify the delivery stops and route characterization, and Holguín-Veras et al. [7] developed a procedure to identify urban freight activity stops from raw GPS data using three parameters: GPS data points, the value of the cutoff acceleration (typically between 25 and 40 $\mathrm{km.h}^{-2}$), and the cutoff speed parameter (typically between 0 and 10 $\mathrm{km.h}^{-1}$). Yang et al. [15] proposed an algorithm to identify the delivery stops using second-by-second GPS data of different delivery tours according to the top speed and time between stops. Since a great number of stops are generated due to high road traffic, the delivery stops are distinguished from the remaining ones according to their features. Kinjarapu et al. [8] developed a heuristic-based model to identify and classify truck stops. These authors found that a combination of trucks' dwelling times and their entropy can be used to classify truck stops by purpose. Aziz et al. [4] analyzed GPS data of truck movement, identifying the truck stops, clustering the several stoppages, and characterizing the truck stops according to their arrival time and duration distributions. In order to overcome the limitations of the existing clustering practices in freight studies, Taghavi et al. [12] use a Hidden Markov Model to identify truck trip segments and extract activity and non-activity stops from GPS data while accounting for the spatio-temporal properties of GPS points.

The analysis of the literature allowed to identify some gaps and limitations. First, even though procedures to identify freight activity stops from raw GPS data are already defined (e.g., [7,13,15]), such procedures use a set of variables as geolocation data, vehicle speed and acceleration. Second, to the best of our knowledge, such procedures were designed to work offline, which limit the use of the collected data for operational monitoring of vehicles and cargos or management of resources (e.g., work re-planing). Third, previous studies use an average cadence of geolocation entries lower than 60 s (e.g., [7,15]).

This paper addresses the application of geolocation data for monitoring logistics processes, namely for detecting vehicle-based operations in real time. This work includes the development of a methodology based on *stationary events* [11], which is designed to work simply on a stream of geolocation entries with arbitrary cadences, in real time and in an incremental fashion. A real-life logistics process is used as a case study to demonstrate the effectiveness of the methodology, namely the automatic detection of operations like the load/unload of goods.

The remainder of this paper is organized as follows. The context of the work is described in the next section. Then, the methodology for automatic detection of vehicle-based operations is defined. Next, a preliminary evaluation of the application of the methodology is presented, as well as a discussion of the results. Finally, the main conclusions and future work are outlined in the last section.

2 Context

This work is conducted to provide insight into the business processes of an international logistics company that operates mainly in the European market. The company relies on over 2000 vehicles, transporting – each year – 7 million *ton* of goods across 200 million *kms*. Each month, over 25000 distribution routes are performed to pick and transport about 4.5 million packages.

Generated by fleet tracking technologies, geolocation data is currently used for monitoring the state of vehicles in terms of positioning and navigation. Information about the execution of operations such as the start and conclusion of load/unload operations is generated by human resources, which has proven to be ineffective due to delayed, imprecise or missing inputs. As a consequence, not only the management of the logistics processes becomes more difficult but also the operations scheduling. Therefore, the exploitation of geolocation data for the detection of vehicle-based logistics operations in real time is an opportunity for improving the monitoring and management of logistics operations, namely the load/unload of goods. Also, this solution can be used to enhance the customer service, by providing means to negotiate more adjusted contracts to reality, and by enabling on-the-fly notifications to customers about their packages.

3 Automatic Detection of Vehicle-Based Operations

This section is structured in four main topics. First, the definition of geolocation entry and stationary event is presented (Sect. 3.1). Then, the operations detection (Sect. 3.2) and operations inference (Sect. 3.3) are described. The detection of vehicle-based operations consists of finding stationary events in a sequence of geolocation entries. Since a stationary event represents some abstract operation that occurred in some geolocation, the inference of logistics operations is necessary to assign a meaning to these events. The last section explains how vehicle-based logistics operations can be detected in real time (Sect. 3.4).

3.1 Geospatial Event Data

In the scope of this work, events consist of vehicle-based operations with a geospatial component. Geospatial events can be categorized as *stationary* and *non-stationary* events, depending on whether the location where the operation took place changed or not during its execution [11]. In a logistics process, operations like the load/unload of goods are stationary events, which can be automatically detected by analyzing the sequences of geolocation entries generated using the vehicle's GPS tracking device. For that, let us consider the following definition [11,13].

Definition 1 (Geolocation entry and stationary event). *Let the geolocation of a vehicle (v) at a specific time instant (t) be defined as the entry* $l = (v, t, lat, lon)$*, where lat and lon identifies the latitude and longitude of the vehicle's position on Earth. Given two geolocations* $l_1 = (lat_1, lon_1)$ *at time instant* t_1 *and* $l_2 = (lat_2, lon_2)$ *at time instant* t_2*:*

– *the function* $\mathbf{distance}(\mathbf{l_1}, \mathbf{l_2})$ *computes the orthodromic distance between the position of* l_1 *and* l_2*.*
– *the function* $\mathbf{speed}(\mathbf{l_1}, \mathbf{t_1}, \mathbf{l_2}, \mathbf{t_2})$ *computes the average speed of the movement from the position of* l_1 *to* l_2*.*

Let $L = [l_1, l_2, ..., l_n]$ *be a sequence (time series) of geolocation entries of the same vehicle. A stationary event E is a subsequence of L with at least two elements, in which all elements must be within a given range of distance, time, and/or average speed values. The following functions are defined for a given stationary event E:*

– *the function* $\mathbf{location}(\mathbf{E})$ *identifies the centroid* $c = (lat_c, lon_c)$ *defined by the elements of the subsequence, which represents the geolocation of E.*
– *the function* $\mathbf{duration}(\mathbf{E})$ *computes the duration of E, which consists of the time difference between the first and last entries of the subsequence.*
– *the functions* $\mathbf{start}(\mathbf{E})$ *and* $\mathbf{end}(\mathbf{E})$ *identifies the time instants of the first and last entries of E.*

□

3.2 Operations Detection

Algorithm 1 describes how stationary events can be identified, in an incremental fashion. This algorithm assumes that, for each vehicle, there is a data structure that holds the history of stationary events as well as the current stationary event candidate.

In this work, a stationary event is defined by a sequence of geolocation entries such that every element must be less than 15 m from the elements' centroid (δ threshold), the movement from an element to another consecutive must be less than $1\,\text{km.h}^{-1}$ (v threshold), and the time difference between two consecutive elements must be less than 2 h (Γ threshold). These threshold values are the same as the ones considered in [13]. Since no logistics operation shorter than 1 min is expected to occur, $\theta = 1$ min is used as the minimum event duration in order to filter some noise in the data.

3.3 Operations Inference

The operations inference consists of linking stationary events to logistics operations. In the scope of this work, the execution of the logistics processes rely on some *work plan*, which is composed by an ordered list of *planned operations*.

Algorithm 1: Identification of stationary events

Input : A geolocation entry (v, t, lat, lon) as defined in Definition 1. As thresholds, δ is the maximum distance (default 30 m), ψ the maximum speed (default 1 $km.h^{-1}$), Γ the maximum time (default 2 h), and θ the minimum event duration (default 1 min).

Output: A stationary event, if identified.

Method

```
S ← null;                                  // the stationary event to be returned
E ← retrieve the current stationary event candidate of vehicle v;
if E ≠ null and end(E) = t then     // equal timestamps, discard entry
    return S;
end
if E ≠ null and distance(location(E),(lat, lon)) ≤ δ and
  speed(location(E),end(E),(lat, lon),t) ≤ ψ and (t − end(E)) ≤ Γ then
    // update E with geolocation (lat, lon) at time instant t
    append (v, t, lat, lon) to E;
else
    // the stationary event candidate is over
    if E ≠ null and duration(E) ≥ θ then
        add E as an executed event of vehicle v;
        S ← E;
    end
    // create a new stationary event candidate
    E ← new stationary event located in (lat, lon) with t as start time;
    set E as the current stationary event candidate of vehicle v;
end
return S;
```

Definition 2 (Planned operation and work plan). *Let v be a vehicle with a GPS tracking device.*

A planned operation p describes a future load/unload operation of a logistics process using vehicle v, which is expected to occur at a specific location. The function **location**$(\mathbf{p})$ *identifies the geolocation where p is supposed to occur. No time information (start time) is directly associated with planned operations.*

A work plan $W = [p_1, p_2, ..., p_m]$ is an ordered list of planned operations for vehicle v, which consists of a trip in which each event represents the trip's check-points. The function **start**$(\mathbf{W})$ *identifies the time instant when W is supposed to start. Only one work plan can be executed at a time, even though there are cases for which it is not possible to determine when a work plan ends and another starts. These cases happen when the first event of a work plan is at the same location of the last event of the previous work plan.*

□

An overview of the real-time monitoring of work plans (and the corresponding operations) is presented in Fig. 1. Comparing to the traditional definition of a business process [2], the work plans are process instances, while the operations are process events.

Work plan	State	Vehicle	Start time	*Operation 1*	*Operation 2*	*Operation 3*	*Operation 4*
Work plan 1	Active	V04	08:15	Location A 08:19 - 09:10	Location B *non-executed*	-	-
Work plan 2	Finished	V10	07:30	Location C 07:15 - 7:50	Location D 09:22 - 09:57	Location E *non-executed*	Location F 10:22 - 11:30
Work plan 3	Planned	V04	14:00	Location A *non-executed*	Location B *non-executed*	-	-
Work plan 4	Finished	V07	07:30	Location E 07:23 - 7:38	Location F 08:19 - 08:40	Location B 09:51 - 12:13	-

Fig. 1. Overview of the real-time monitoring of work plans.

Since a planned operation is geolocated, the operations inference is achieved by checking whether a stationary event occurred nearby to that planned operation. In this work is considered that, if the orthodromic distance between a stationary event E and a planned operation O is no farther than 1000 m then E should represent the execution of O. If no planned operation satisfies the aforementioned condition for E, then it can be assumed that E represents a *negligible* operation (e.g., vehicle refueling or driver's resting). In the scope of this work, these unmatched stationary events are discarded.

Given a stationary event E and the list of work plans for some vehicle, Algorithm 2 describes how to identify the current active work plan (or plans, if one is ending in the same location as another starts). This algorithm assumes that there is a function that describes whether a planned operation was already executed or not. The *Radius* threshold defines the maximum orthodromic distance between E and the planned operations, which is set to 1000 m as explained before. The $minT$ and $maxT$ thresholds define the allowed time offset range for starting a new work plan, which are set to -5 h and $+12$ h of the planned starting time.

3.4 Real-Time Computation

Given a stream of geolocation entries and a list of work plans, vehicle-based logistics operations can be detected in real time using Algorithm 3. The geolocation entries are considered to compute stationary events by applying Algorithm 1. The stationary events are considered to identify the active work plans and planned operations by applying Algorithm 2. The non-executed planned operations of the active work plans are matched with the non-reported stationary events for checking the execution of operations. It is important to mention that the execution of a planned operations may be supported by more than one stationary event. A good example of this case is when a vehicle performs some check-in operation in one location prior to the load/unload of goods in another location a few hundred

Algorithm 2: Identification of active work plans and operations

Input : For a specific vehicle v, a list of work plans ($W = [w_1, w_2, ..., w_n]$) and a stationary event (E). [$\mathsf{minT}; \mathsf{maxT}$] is the allowed time offset range for starting a work plan (default $[-5\ h; 12\ h]$). Radius defines the area where operations must be performed (default 1000 m).

Output: The work plans and planned operations activated by E, if exist.

Method

`// Current and past work plans`

$B \leftarrow \{w$ in $W \mid w$ contains at least one planned operation that was executed already$\}$;

`// Future work plans that can be activated`

$C \leftarrow \{w$ in $W \mid w$ not in $B\ \wedge\ \mathsf{minT}\ \leq \mathtt{start}(E) - \mathtt{start}(w) \leq\ \mathsf{maxT}\}$;

`// Current work plan`

$A \leftarrow \{w\ in\ B \mid \forall_{x\ \neq\ w\ in\ W}[x$ not contains a planned operation which was executed after any executed event in $w]\}$;

`// Check whether the current work plan is still active`

if $\exists_{w\ in\ A}[\nexists_{p\ in\ w}[\ p\ is\ not\ executed \wedge$ $\mathtt{distance(location}\ (E)\mathtt{,location}\ (p)\mathtt{)} < \mathsf{Radius}]]\ \wedge$ $\exists_{w'\ in\ C}[\exists_{p'_x\ in\ w'}[\ p'_x\ is\ not\ executed \wedge x \leq 3 \wedge$ $\mathtt{distance(location}\ (E)\mathtt{,location}\ (p')\mathtt{)} < \mathsf{Radius}]]$ **then** $A \leftarrow \emptyset$;

`// Find the work plan and planned operation activated by` E

if $A = \emptyset$ **then**

  `// Current work plan is not active, find a new one`

  $X \leftarrow \{(w, p_1) \mid w \in C\ \wedge\ p_1\ in\ w\ \wedge\ \mathtt{distance}(\alpha,\beta) <\ \mathsf{Radius}\ \wedge$ $\forall_{w'\ \neq\ w\ \in\ C}[p'_1\ in\ w'\ \wedge\ \mathtt{distance}(\alpha,\gamma)\ \geq\ \mathtt{distance}(\alpha,\beta)]\}$, where $\alpha =$ $\mathtt{location}(E)$, $\beta = \mathtt{location}(p_1)$, and $\gamma = \mathtt{location}(p'_1)$;

  return $(X, \emptyset)$;

else

  `// Current work plan is active, find the activated operation`

  $X \leftarrow \{(w, p) \mid w \in A\ \wedge\ p\ in\ w\ \wedge\ p$ is not executed$\}$;

  $Y \leftarrow \{(w, p) \in X \mid \forall_{(w',p')\ \neq\ (w,p) \in X}[\mathtt{distance}(\alpha,\gamma)\ \geq\ \mathtt{distance}(\alpha,\beta)]\}$, where $\alpha = \mathtt{location}(E)$, $\beta = \mathtt{location}(p_1)$, and $\gamma = \mathtt{location}(p'_1)$;

  `// Find a next work plan for which the first event is at the same location of the current operation of the active plan`

  $Z \leftarrow \{(w, p_1) \mid w \in C \wedge p_1\ in\ w \wedge \exists_{(w',p') \in Y}[\mathtt{location}(p') = \mathtt{location}(p_1)]\}$;

  return (Y, Z);

end

meters away. In the scope of this work, all stationary events that represent the execution of a specific planned operation are aggregated. This means that, the logistics company is interested to simply know the time a vehicle remains at the location of some planned operation. Hence, the results consist of messages notifying and quantifying – in real time – the execution of planned operations.

Algorithm 3: Real-time detection of logistics operations

Input : A stream of geolocation entries (Input) and the list of work plans (WPs). Radius defines the area where planned operations must be performed (default 1000 m).

Output: A stream of detected logistics operations.

Method

1 Open Output as the stream of detected logistics operations;
2 $O \leftarrow \emptyset$, the stack of events associated with some unreported planned operations ; `// Activated planned operations to be reported`
3 **while** *stream* Input *is open* **do**
4 $(v, t, lat, lon) \leftarrow$ wait/get geolocation entry from Input;
5 $W \leftarrow$ get list of work plans of vehicle v from WPs;
6 **foreach** $(A, B, C) \in O$ **do**
7 $(w, p) \leftarrow$ the only element in A;
`// Check whether the vehicle left the operation area, so no more events can occur in there`
8 **if** distance(location(p), (lat, lon)) $>$ Radius $\times$ 2 **then**
`// Report the execution of the matched planned operations`
9 change the state of operation p to *executed*;
10 change the state of every event $e \in C$ to *reported*;
11 *start* $\leftarrow$ earliest start time of the events in C;
12 *end* $\leftarrow$ latest end time of the events in C;
13 **if** $B \neq \emptyset$ **then**
14 $(w', p') \leftarrow$ the only element in B;
15 *middle* $\leftarrow$ time instant that is equidistant to *start* and *end*;
16 change the state of p' to *executed*;
17 add $(v, w, p, start, middle)$ and $(v, w', p', middle, end)$ to Output;
end
18 **else** add $(v, w, p, start, end)$ to Output;
19 $O \leftarrow O \setminus (A, B, C)$; `// remove from the stack`
end
end
`// Check whether there is a new stationary event for vehicle` v
20 $e \leftarrow$ apply Algorithm 1 with (v, t, lat, lon);
21 **if** $e = null$ **then** break iteration (go to line 3);
`// Find the work plans and planned operations activated by` e
22 $(X, Y) \leftarrow$ apply Algorithm 2 with W and e;
23 **if** $X \neq \emptyset$ **then**
`// Add` e `to the stack of events, associated to the activated work plans and operations`
24 **if** $\exists_{(A,B,C) \in O}[A = X \wedge B = Y]$ **then** $O \leftarrow (X, Y, C \cup e) \cup O \setminus (A, B, C)$
25 **else** $O \leftarrow (X, Y, \{e\}) \cup O$;
end
end

4 Evaluation and Discussion

A preliminary evaluation of the application of the methodology presented in the previous section was conducted using real-life data from the logistics company. In this section, the characterization of the data used for the performed evaluation is presented. Also, the results of the Algorithms 1, 2 and 3 (implemented in Python) are assessed in terms of conformance and performance. The general limitations and impacts of the methodology are discussed in the end.

4.1 Data Characterization

The data used in this work consists of one month of geolocation data for three vehicles, as well as the corresponding work plans (routes). A total of 82000 geolocation entries and 95 work plans were collected from the operational systems. This data captures a combination of urban and sub-urban operations between distribution centers and delivery points. The vehicles usually travel long distances (hundreds of kms per route) through highways to reach urban areas. The average cadence of data used is around 100 s. Vehicle V1 has the highest number of geolocation entries ($N = 29238$), but the lowest number of work plans ($N = 18$). Vehicles V2 and V3 have a similar number of geolocation entries ($N \approx 26000$). However, the volume of work plans is higher for vehicle V3 ($N = 56$) than for vehicle V2 ($N = 21$). On the other hand, vehicle V2 has the triple of stationary events ($N = 1676$), than the vehicle V3 ($N = 658$). This may be related with the level of complexity of work plans as vehicle V2 has, proportionally, more planned operations ($N = 67$) than vehicle V3 ($N = 113$). The input and output data characterization is presented in Table 1.

Table 1. Input and output data characterization.

Indicator	Vehicle V1	Vehicle V2	Vehicle V3
Geolocation entries	29328	26329	26335
in stationary event	15722 (53.6%)	17162 (65.2%)	21067 (80.0%)
(average cadence)	92 s	102 s	102 s
Stationary events	589	1676	658
with known location	155 (26.3%)	225 (13.4%)	368 (55.9%)
in work plan	210 (35.7%)	804 (48.0%)	431 (65.5%)
Work plans	18	21	56
fully fulfilled	17 (94.4%)	18 (85.7%)	52 (92.9%)
partially fulfilled	1 (5.6%)	2 (9.5%)	4 (7.1%)
Planned operations	36	67	113
with detected execution	35 (97.2%)	63 (94.0%)	108 (95.6%)

4.2 Conformance Checking

Conformance checking was performed to evaluate whether the work plans were executed according to the expected. On the one hand, start times were analyzed to identify and quantify delays. On the other hand, the detected operations were *parsed* in order to identify deviations to the work plan. These deviations can be either missing or swapped operations, such as the *alignment steps* for replaying event logs on process models [3]. The results of the conformance checking analysis are presented in Table 1.

4.3 Performance Analysis

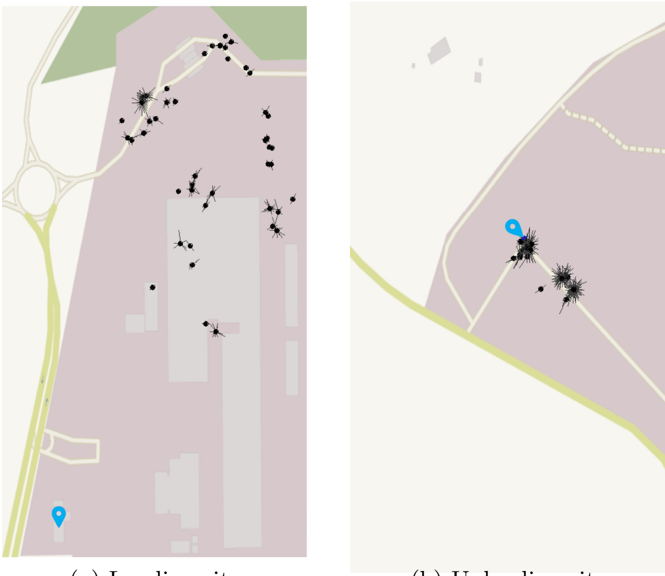
The performance of the execution of the logistics processes provides insight into the efficiency of the company. The performance analysis can be conducted taking into account different perspectives such as work plans, planned operations and vehicles. At the work plan level, we analyzed the throughput time, the start time, and the delay, while at the planned operations level we computed the execution time. Table 2 provides an overview of some process performance indicators obtained in this evaluation.

Table 2. Overview of the process performance analysis.

Indicator	Vehicle		
	V1	V2	V3
Work plans			
Average throughput time	12:14:33	05:14:19	03:20:02
Average start time (executed vs planned)	−03:14:16	00:12:55	−00:52:28
Average delay	00:07:38	02:47:23	00:39:48
Started on time	16 (88.9%)	15 (78.9%)	41 (74.5%)
Planned operations			
Average execution time	04:48:18	01:01:44	01:18:23

As can be observed, the throughput times range from 3 h to 12 h and the delays between a few minutes to a few hours. Vehicle V1 has the highest throughput time, but the lower delay. These results suggest that the work plans were performed quite efficiently.

A common work plan is given as an example for exploiting the spatial aspect of the results. The work plan, which is executed in a regular basis, consists on just two operations: (1) the loading (of goods) in location A and (2) the unloading in location B. The road distance between A and B is around 200 km, which can be driven in 3 h. Figure 2 depicts – on a map – the history of stationary events associated with these operations. Details about the operations' performance are provided in Table 3. Note that the high time variations described by the standard deviations (event duration and operation execution time) are due to the drivers' resting nearby the location of the operation.



(a) Loading site (b) Unloading site

Fig. 2. History of stationary events of two logistics operations. The blue markers represent the operations' expected geolocation, while the black circles represent the detected stationary events.

Table 3. Performance analysis of a specific work plan.

Indicator	Loading of goods	Unloading of goods
	avg (*std*)	*avg* (*std*)
Stationary events		
Location offset (distance)	396 m (63 m)	35 m (28 m)
Location offset (azimuth)	117.6° (9.6°)	61.0° (31.4°)
Duration	00:19:41 (00:30:47)	01:26:27 (01:29:27)
Logistics operations		
Aggregated events	6.9 (1.5)	4.6 (4.5)
Execution time	02:00:25 (01:11:34)	08:11:30 (09:17:23)

Note: *avg* and *std* stand for average and standard deviation.

4.4 Discussion

The logistics processes considered in this work can be considered rather structured. This means that, there is neither a high variability in the workflow nor too much unexpected behavior in the execution of the processes. So, once that the focus of this work is simply the load/unload of goods, the application of process discovery techniques would not provide much new knowledge about the logistics processes. The application of conformance checking, however, is useful to verify the correct and complete execution of the work plans. Non-conforming cases may be due to either data issues (e.g., noise or missing data) or work performed in an unexpected manner (e.g., unfulfilled operations).

The real-time computation poses a challenge to the detection of logistics operations. In this work, the execution of an operation is assumed to be completed if the vehicle exits the area where the operation is supposed to be performed. If, for some reason, the vehicle has not exited the area permanently, then the detection would be erroneous. The methodology applied in this work addresses this issue, being able to correctly detect the execution of around 95% of the logistics operations (load/unload of goods).

The accuracy of the geolocation of the logistics operations is also a critical factor for the application of the methodology. The geolocation reference is often computed either using the postal address or the street entrance, which may be several hundred meters away of the location of the logistics operations. This issue is even worse when two or more distinct sites are located close by. In this work, a constant distance value (1000 m) is used to check whether a stationary event represents a logistics operation. However, a dynamic approach would be more adequate, especially because the layout and dimension of the sites where logistics operations occur vary enormously. The location offset, as presented in Table 3, can be used to adjust the location of the logistics operations.

5 Conclusions

This paper addresses the application of geolocation data for monitoring operations of an international logistics company, namely the load/unload of goods. The implementation of the methodology as well as the main issues for its application are described and discussed throughout the paper. The effectiveness of the solution was demonstrated using real-life geolocation data.

The exploitation of geolocation data for the detection of vehicle-based logistics operations in real time is an opportunity for improving the monitoring and management of logistics operations. The scope of this work is primarily to enhance the customer service of the logistics company, by providing means to negotiate more adjusted contracts to reality, and by enabling on-the-fly notifications to customers about their packages. It is acknowledged, however, that this solution has potential to provide new insight into the execution of the existing logistics processes. Process mining techniques should be a valuable complement for that.

As a future work, we envision the extension of the methodology in order to detect all kinds of vehicle-based operations instead of simply the load/unload of goods. This extension will require the classification of events that occur at unknown locations, which might be driven by the points of interest (POIs) in the surroundings of the events' location.

Acknowledgements. This work is financed by the European Regional Development Fund through the Operational Programme for Competitiveness and Internationalisation - COMPETE 2020 Programme and by National Funds through the

Portuguese funding agency, FCT - Fundação para a Ciência e a Tecnologia within project PTDC/ECI-TRA/32053/2017 - POCI-01-0145-FEDER-032053. Tânia Fontes also thanks FCT for the Post-Doctoral scholarship SFRH/BPD/109426/2015.

References

1. Telematics Benchmark Report 2019 - US Edition, Teletrac Navman. https://www.teletracnavman.com/resources/resource-library/articles/telematics-benchmark-report. Accessed 01 Sept 2021
2. van der Aalst, W.: Process Mining: Discovery, Conformance and Enhancement of Business Processes. Springer, Berlin (2011). https://doi.org/10.1007/978-3-642-19345-3
3. van der Aalst, W., Adriansyah, A., van Dongen, B.: Replaying history on process models for conformance checking and performance analysis. Wiley Interdiscip. Rev. Data Min. Knowl. Discov. **2**(2), 182–192 (2012). https://doi.org/10.1002/widm.1045
4. Aziz, R., et al.: Identifying and characterizing truck stops from GPS data. In: Perner, P. (ed.) ICDM 2016. LNCS (LNAI), vol. 9728, pp. 168–182. Springer, Cham (2016). https://doi.org/10.1007/978-3-319-41561-1_13
5. Biagioni, J., Gerlich, T., Merrifield, T., Eriksson, J.: EasyTracker: automatic transit tracking, mapping, and arrival time prediction using smartphones. SenSys **2011**, 68–81 (2011). https://doi.org/10.1145/2070942.2070950
6. Gong, L., Sato, H., Yamamoto, T., Miwa, T., Morikawa, T.: Identification of activity stop locations in GPS trajectories by density-based clustering method combined with support vector machines. J. Mod. Transp. **23**(3), 202–213 (2015). https://doi.org/10.1007/s40534-015-0079-x
7. Holguin-Veras, J., Encarnacion, T., Pérez-Guzmán, S., Yang, X.: Mechanistic identification of freight activity stops from global positioning system data. Transp. Res. Rec. **2674**(4), 235–246 (2020). https://doi.org/10.1177/0361198120911922
8. Kinjarapu, A., Demissie, M.G., Kattan, L., Duckworth, R.: Applications of passive GPS data to characterize the movement of freight trucks-A case study in the calgary region of Canada. IEEE Trans. Intell. Transp. Syst. (2021). https://doi.org/10.1109/TITS.2021.3093061
9. Pinelli, F., Calabrese, F., Bouillet, E.P.: Robust bus-stop identification and denoising methodology. In: IEEE Conference on Intelligent Transportation Systems (ITSC), pp. 2298–2303 (2013). https://doi.org/10.1109/ITSC.2013.6728570
10. Pluvinet, P., Gonzalez-Feliu, J., Ambrosini, C.: GPS data analysis for understanding urban goods movement. Procedia Soc. Behav. Sci. **39**, 450–462 (2012). https://doi.org/10.1016/j.sbspro.2012.03.121
11. Ribeiro, J., Fontes, T., Soares, C., Borges, J.: Process discovery on geolocation data. Transp. Res. Procedia **47**, 139–146 (2020). https://doi.org/10.1016/j.trpro.2020.03.086
12. Taghavi, M., Irannezhad, E., Prato, C.G.: Identifying truck stops from a large stream of GPS data via a hidden Markov chain model. In: 2019 IEEE Intelligent Transportation Systems Conference (ITSC), pp. 2265–2271. IEEE (2019). https://doi.org/10.1109/ITSC.2019.8917156
13. Tavares, J., Ribeiro, J., Fontes, T.: Detection of vehicle-based operations from geolocation data. Transp. Res. Procedia (2021, in press)

14. Van Dijk, J.: Identifying activity-travel points from GPS-data with multiple moving windows. Comput. Environ. Urban Syst. **70**, 84–101 (2018). https://doi.org/10.1016/j.compenvurbsys.2018.02.004
15. Yang, X., Sun, Z., Ban, X.J., Holguín-Veras, J.: Urban freight delivery stop identification with GPS data. Transp. Res. Rec. J. Transp. Res. Board **2411**(1), 55–61 (2014). https://doi.org/10.3141/2411-07