



**HAL**  
open science

## From Raw Sensor Data to Semantic Trajectories

Frédéric Bisone, Thomas Devogele, Laurent Etienne

► **To cite this version:**

Frédéric Bisone, Thomas Devogele, Laurent Etienne. From Raw Sensor Data to Semantic Trajectories. 5th ACM SIGSPATIAL International Workshop on the Use of GIS in Emergency Management (EM-GIS 2019), Nov 2019, Chicago, United States. 10.1145/3356998.3365777 . hal-02380444

**HAL Id: hal-02380444**

**<https://hal.science/hal-02380444>**

Submitted on 26 Nov 2019

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# From Raw Sensor Data to Semantic Trajectories

Frederick Bisone  
University of Tours  
Blois, France

frederick.bisone@univ-tours.fr

Thomas Devogele  
University of Tours  
Blois, France

thomas.devogele@univ-tours.fr

Laurent Etienne  
ISEN  
Brest, France

laurent.etienne@isen-ouest.yncrea.fr

## ABSTRACT

GPS position are useful to analyse movements of mobile objects. Unfortunately, the outcome can be unsatisfactory due to imprecision and signal lost. Several sensors (generic or specific ones depending on the type of vehicle) are now included into mobile objects. This article describes a new generic model that enhances the semantic trajectory model and a process to extract semantic from GPS and other sensors. This process is based on the raw data from these sensors which help identify how, why and when mobile objects are moving, in order to add semantic information to the trajectories and then design new applications such as smart GPS, reporting systems or Remote Maintenance System. This process was successfully applied to the Indre et Loire fire department and its connected ambulances.

## KEYWORDS

Semantic trajectory, Trajectory data mining, Emergency vehicles, Interactive traffic analysis with GPS and sensors, Semantic extraction process

## 1 INTRODUCTION

Nowadays, vehicles include several heterogeneous sensors such as GPS and gyroscopes. In addition, there are a lot of sensors that collect data from the internal network of the vehicle such as the handbrake, car doors and the engine itself just to name a few. Similarly, other specific hardware, for example sirens embedded with sensors are now added so as to fit a particular requirement and as a result transmit data which can be afterwards collected and utilized. The raw data collected from these various sensors can be used to design specific new applications depending on the particularity of the vehicles studied. One baseline of semantic trajectory modeling is the work of [8] which present the notion of contextual data repository and semantic enrichment. The authors present a conceptual model based on Stop and Move where trajectory are enhanced with annotations and/or one or several complementary segmentations. A complementary work of Parent et al. is presented in [3] where the conceptual model for semantic trajectories CONSTANT is explained. This model included many semantic information like the environment, the place, the event and also more complex

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from [permissions@acm.org](mailto:permissions@acm.org).

EMGIS'19, November 5, 2019, Chicago, IL, USA

© 2019 Association for Computing Machinery.

ACM ISBN 978-1-4503-6965-7/19/11...\$15.00

<https://doi.org/10.1145/3356998.3365777>

informations like goal, transportation mean and behavior. More recently, many conceptual models continue and complete the work initiated by CONSTANT in order to consider trajectories as semantic multi-dimensional sequences like MASTER [5]. In particular, such models make it easy to represent vehicle trajectories. Even better, their rich expressiveness makes it possible to take into account private vehicles such as ambulances equipped with many sensors. For instance, a smart GPS navigation system, which allows to determine an adapted trajectory depending on the type of vehicle and/or the state of the patient in the case of an ambulance. These kind of vehicles can move faster than speed limit. In addition, it provides the opportunity of designing reporting system with new indicators which enable a precise study of the trajectories such as the average time of a displacement and the average duration based on the type of activity.

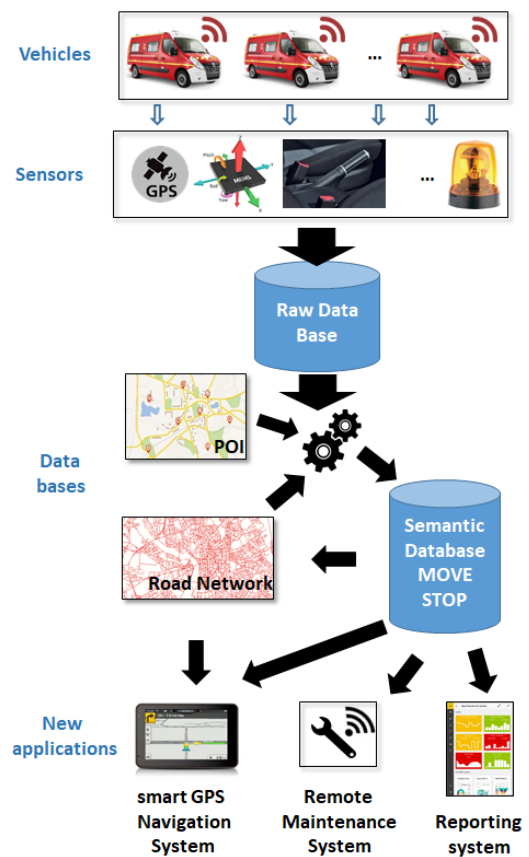


Figure 1: Overview of the use of data from sensors to new applications

It also facilitates the remote maintenance of the vehicle fleet by harmonizing the heterogeneous information provided by all the sensors of different vehicles, for example, detecting abnormal use of suspensions in regards of the type of roads taken, reduced battery capacity, etc. Anticipated maintenance is also possible by predicting car failure. As a matter of fact, due to their professional use (hard braking, repeated burst of speed, repeated door opening) car manufacturer's recommendations are insufficient to ensure car durability. The goal of this article is to define a process on how to extract these semantic behaviors from raw data and external database for specific vehicles. It should be acknowledge that Open Street Map is used as external database. Using large data set of raw data is very interesting to define precise semantic of specific movement. Nevertheless, handling large data set is complex. Databases have to manage large heterogeneous volume of data and CPU time to extract semantic information is consequent. Figure 1 summarizes the process and the data flow from row data to new applications. In this article, we focus on ambulance. Nevertheless, this semantic extraction process can be used for other kind of vehicles.

The reminder of this article is organized as follows. In section one, a new model of mobile object is defined. Section 2 describes how to identify stop and extract semantic information from them. This step is validated on large dataset from ambulances' sensors. In the third section, we enumerate how we identify and describe the semantic of move items. We conclude the paper with prospects and future work.

## 2 SEMANTIC MODEL OF MOBILE OBJECT

In this model, we improve the description of the existing items from [8]: MOVE and STOP. MOVE items are described by three different attributes: the kind of vehicles (KV), aim of the displacement (AD) and the type of the displacement (TD). For instance, an ambulance (KV) driving fast (TD) to transport a patient (AD) to a hospital or an ambulance (KV) driving normally (TD) to refuel (AD). STOP items are described by one attribute: activity during a stop (SA). STOP can also be linked with a POI. For instance, dropping off the patient to a hospital (AD) at the Bretonneau Hospital (POI) or taking care of a patient on scene (AD) without associated POI. In this model, there can be various consecutive STOPS without any MOVE in between if the semantic description of each stop is not the same. In the same way, several consecutive MOVES can be consecutive without any STOP in between if the semantic description changes. For example, figure 2 shows a timeline of an ambulance. Seven different MOVES are presented. The second column of the timeline summarizes the type of displacement (TD) and the aim of displacement (AD). We notice a particular case between the 3rd and the 4th MOVE where the ambulance is redirected (The ambulance on its way back to the fire station responded to an emergency call). The difference between the 2nd and 5th MOVES is that in the former situation, the ambulance is transporting a patient thus driving carefully while driving rapidly in the 5th one. Seven different STOPS are presented. Different activities are listed: waiting, taking care of a patient, dropping off a patient or logistic operation in the third column.

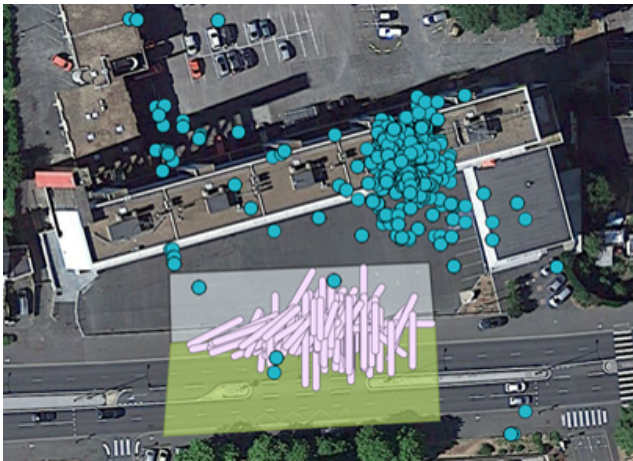
TIMELINE	MOVE: Type of displacement (TD)	STOP: Stop activity (SA)	STOP: POI: category	STOP: POI: name
				Tours Fire Station
				Trousseau hospital
				Cloche-ville hospital
				Total gas station
				Tours Fire Station

Figure 2: Example of the timeline of an ambulance

## 3 AUTOMATIC IDENTIFICATION OF SEMANTIC STOPS

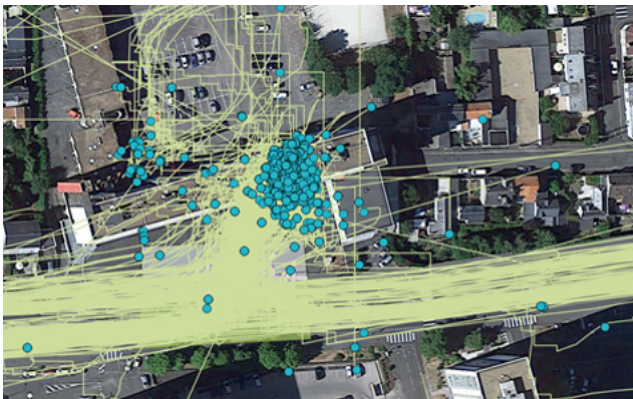
A semantic stop is modeled in three dimensions. The first one is the spatial dimension of the STOP (X and Y coordinates). Next, the temporal dimension (period, duration of the STOP). And the last dimension is the semantic one, a STOP is described by an activity and can be linked with a POI (Point of Interest). First of all, we need to identify the STOPS in the GPS raw data. To do so, several methods already exist. A classic one that only uses GPS data, is to extract from raw data, the period where the object isn't moving. However, this method is sensitive to signal lost and it is impossible to pinpoint the signal to a specific location. Indeed, when GPS positions are logged over time, the dispersion of points is high for vehicles stopped inside a building as garage or under the roof of the hospital's entrance [2, 4].

The second studied method is proposed by Etienne et al. [6] We create for each area of interest (Hospital, Fire station, Gas station) two polygons  $Z_1$  and  $Z_2$ . We consider that the mobile object has left a given zone when a trajectory which is first in the  $Z_1$  zone at a given moment  $t_0$  is then in the  $Z_2$  zone at time  $t_1$ . When the moving object passes from zone  $Z_2$  to  $Z_1$ , we consider that it has entered the zone. Between the entrance and exit, we consider it as a STOP. Figure 3 describe this method applied to our data.



**Figure 3: Scatter plot of stopped vehicles inside or close to the garage of Tours fire station (blue point), entrance and exit zones and line between points from entrance zone to exit zone.**

The blue points represent the raw data when the vehicle isn't moving (Speed = 0). Lines are created by linking the points for each trajectory between  $Z_1$  and  $Z_2$ . This method works well when the vehicle enters the fire station. However, due to the technical limitations of the GPS (Time to Fix, indoor) signal detection is not really efficient inside or near buildings such as garages or hospitals for exits. Indeed, as shown in figure 4, lots of exits are missed as the GPS isn't able to transmit/receive the signal before certain time. Furthermore, this method needs to create for each area of interest,



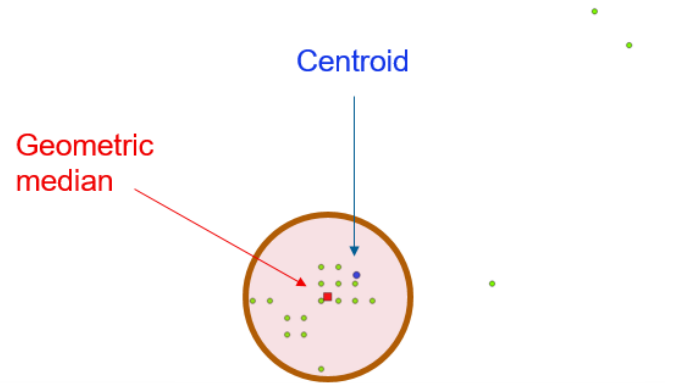
**Figure 4: Scatter plot of stopped vehicles inside or close to the garage of Tours fire station (blue point), entrance and exit zones and line between points from entrance zone to exit zone.**

two polygons in order to detect both the entry and exit, which is impossible for an ambulance as it can stop everywhere in a city.

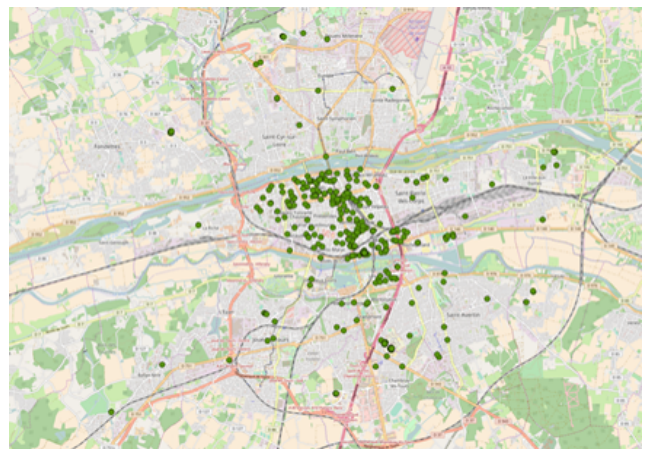
The last method uses the information provided by the park break sensor to identify STOPS. We detect when the park brake is on, and flag the row data as break. This method is not sensitive to GPS

typical problems because it only use GPS to identify where the stop is located. A STOP is a continuous period where all rows data are flagged as break which is during more than one minute. The next step afterwards is to compute the duration of the STOP and timestamp its beginning and ending. This methods have been tested on a small sample of data containing hundred STOPS, the first one using only GPS data identified all the stops, but also twenty one false positive. The second method identified all entrances in STOP areas, but it didn't get all the exits and only 82 STOPS where detected. Last method identified all the STOPS, this one is used in our process.

When a stop is finally identified, we need to represent it in space. It seems obvious that if a mobile object isn't moving, it must be represented by only a point. Due to the inaccuracy of the GPS, all the data for a STOP doesn't have the same position even if the object isn't moving. Small [9] gives a survey on central point. The geometric median [10] is chosen. This last one is less sensitive to outliers than other central position as centroid. Figure 5 shows the difference between the geometric median and the centroid.



**Figure 5: Geometric median and centroid for a set of positions of a STOP and the dispersion indicator : R95**



**Figure 6: STOPS dispersion in the city of Tours.**



In addition to the geometric median, two indicators are also computed to determine the quality of the representation. The first one is the percentage of lost data during the STOP while the second one R95, is the radius of a circle centered in the geometric median which satisfies the GPS' accuracy definition (At least 95% of the points are contained in the circle as shown in figure 5). The last dimension of a STOP is the semantic one. The activity of a STOP is determined by expert rules. For example, if the vehicle is parked inside a fire station, activity is waiting/to be on call/available for duty if necessary/waiting to respond to a call, and/or if the vehicle is at a gas station, activity is car refueling and so on. To test our method, we apply this extraction procedure on 5 millions rows from our data set. 2,1 million of them were flagged as break, and 807 STOPs were extracted. Figure 6 represents the STOPs on the map of the city of Tours. A matching process is then initiated to associate STOPs with external POIs. Beer et al. [1] describes a large number of point matching processes from different data sets. Currently, a simple process is employed. A maximum threshold of distance  $D_{max}$  is defined. If the nearest POI is at a greater distance than this threshold, then the STOP cannot be matched to it. Conversely, several STOPs can be associated with the same External POI. First study shows that around 80% of the STOPs are linked to a POI.

#### 4 AUTOMATIC DETECTION OF SEMANTIC MOVES

Automatic detection of semantic move starts by extracting trajectories between STOPs. Next, the Case based Reasoning (CBR) method proposed by [7] is used. In this process, several prototypical cases have to be defined. Similarity measures are used to associate real events to the forementioned cases.

Type of displacement	Previous POI	Previous Semantic	Displacement	Next POI	Next Semantic
Transport to hospital	*	Taking care of a patient on scene	Flashing lights 100%	Type : Hospital	Dropping off patient at hospital
Heading to intervention	Type : Fire station	Waiting	Flashing lights 100%	*	Taking care of a patient on scene
Logistical move	*	*	0% Flashing lights	Type : Logistic	Logistical activity
Back to firestation	*	*	0% Flashing lights	Type : Fire station	Attente caserne
Redirection to intervention	*	?	100% Flashing lights	*	Taking care of a patient on scene
Redirection canceled	*	?	0% Flashing lights	Type : Fire station	Waiting

Table 1: Example of the timeline of an ambulance

We employed CBR to precisely identify the type of displacement (TD). Some ordered set of sensors values during the MOVE are used as input. For this identification problem, CBR method is better than expert system. Indeed, identify prototypical cases with experts is easier than define precise rules with threshold values. The table 1 represents the list of the prototypical cases used in our automatic detection process. In this respect, to define a MOVE as heading to the intervention scene which is when the vehicle is moving from the fire station to an intervention point where the previous POI is a

fire station and activity is waiting. Flashing lights were switched on all along the MOVE. The next POI is ignored, and the next activity is taking care of a patient on scene. Moreover, to define a MOVE as a redirection to an intervention, which is when a vehicle is on its way back to the fire station, and receives a call to a new intervention. In this situation, the previous POI is ignored, previous activity is unknown because redirection isn't defined yet. Flashing lights were used all along the MOVE. The next POI is ignored, and the next activity is taking care of a patient on scene. To automatically label MOVES, we have to define a similarity measure between prototypical cases and real cases. Also, the position and the activity when a vehicle is redirected needs to be identified and defined.

#### 5 CONCLUSION AND FUTURE WORK

In this article we have presented a new model based on existing ones. This generic model enhances the existing semantic trajectory models. Moreover, several consecutive items can have the same type (MOVE or STOP) and displacements are described by their type (TD). We also presented a process to extract semantics STOPs and MOVES. This model is customized for ambulances and the process has been validated with real data of connected ambulances belonging to the Indre-et-Loire fire department in France.

As future work, we will study how to precisely identify the position where a vehicle is redirected and validate our process to semantic MOVES extraction. Furthermore, we will employ these semantic information to define new indicators for reporting systems in a short run. Then for a smart GPS and a remote maintenance system for specific vehicles.

#### ACKNOWLEDGMENTS

This work was funded and supported by Petit Picot by GRUAU, as well as the ANRT (Association Nationale de la Recherche et de la Technologie).

#### REFERENCES

- [1] Catriel Beer, Yaron Kanza, Eliyahu Safra, and Yehoshua Sagiv. 2004. Object fusion in geographic information systems. In *Proceedings of the Thirtieth international conference on Very large data bases-Volume 30*. VLDB Endowment, 816–827.
- [2] Boaz Ben-Moshe, Elazar Elkin, Harel Levi, and Ayal Weissman. 2011. Improving Accuracy of GNSS Devices in Urban Canyons.. In *CCCG*. 511–515.
- [3] V. Bogorny, C. Renso, A. Ribeiro de Aquino, F. de Lucca Siqueira, and LO. Alvares. 2014. CONSTAnT - A Conceptual Data Model for Semantic Trajectories of Moving Objects. *Transactions in GIS* 18 (2014), 66–88.
- [4] James W Cain III, Paul R Krausman, Brian D Jansen, and John R Morgart. 2005. Influence of topography and GPS fix interval on GPS collar performance. *Wildlife Society Bulletin* 33, 3 (2005), 926–934.
- [5] R. dos Santos Mello, V. Bogorny, LO. Alvares, L.HZ Santana, CA. Ferrero, AA. Frozza, GA. Schreiner, and C. Renso. 2019. MASTER: A multiple aspect view on trajectories. *Transactions in GIS* (2019), 1–20.
- [6] Laurent Etienne, Thomas Devogele, and Alain Bouju. 2012. Spatio-temporal trajectory analysis of mobile objects following the same itinerary. *Advances in Geo-Spatial Information Science* 10 (2012), 47–57.
- [7] Janet Kolodner. 2014. *Case-based reasoning*. Morgan Kaufmann.
- [8] Christine Parent, Stefano Spaccapietra, Chiara Renso, Gennady Andrienko, Natalia Andrienko, Vania Bogorny, Maria Luisa Damiani, Aris Gkoulalas-Divanis, Jose Macedo, Nikos Pelekis, et al. 2013. Semantic trajectories modeling and analysis. *ACM Computing Surveys (CSUR)* 45, 4 (2013), 42.
- [9] Christopher G Small. 1990. A survey of multidimensional medians. *International Statistical Review/Revue Internationale de Statistique* (1990), 263–277.
- [10] Endre Weiszfeld and Frank Plastria. 2009. On the point for which the sum of the distances to n given points is minimum. *Annals of Operations Research* 167, 1 (2009), 7–41.