Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 00 (2018) 000–000

Procedia Computer Science

www.elsevier.com/locate/procedia

The 9th International Conference on Ambient Systems, Networks and Technologies (ANT 2018)

Semantic Enrichment of Workers Trajectories for Safer Construction Sites

Abstract

Existing literature reveals that major reasons of fatalities on construction sites are related to the mobility related issues such as unsafe human behaviors, difficult site conditions, and workers falling from heights and striking against or being struck by moving objects. There exist many attempts in the literature to reduce these fatalities using information extracted from location based technologies such as GPS for tracking people and machineries on construction sites. However, a significant gap has been found that the usage of semantically enriched mobility data is still missing in worker`s safety monitoring scenarios for effective management of construction sites. To address this research gap, a system is proposed to enrich mobility data of workers with external data sources both openly available and private data related to construction sites. Three annotation techniques are used to transform raw spatio-temporal mobility data into semantic trajectories which are: (1) annotation with semantic points that maps site location identification to trajectory points; (2) annotation with semantic lines that relies on the speed based segmentation approach and infers the modes of transportation used in trajectory`s move episodes; and (3) annotation with semantic region for mapping a complete trajectory on an actual location. These semantic trajectories will help Health and Safety (H&S) managers in making improved decisions for planning, monitoring and controlling site activities by understanding workers behaviors that ultimately contributes in reducing the fatal accidents occurring on sites each year.
Keywords: Safety;workers;construction, spatio-temporal data

1. Introduction

Construction industry is very hazardous in nature because of harsh work environment and involvement of high safety risks. According to the Bureau of Labor Statistics[1], in 2015 out of 4,836 fatal work injuries, 19 % of fatalities were recorded from the U.S. construction industry. Whereas from the EU construction industry, 20.9% fatal accidents[2] were reported in 2014. The major reasons of fatalities were related to the unsafe human behaviors, difficult site conditions, and workers falling from heights and striking against or being struck by moving objects[3]. The above reasons highlight the need for more effective construction resources` mobility monitoring and safety planning methods. Construction resources such as workers, machineries and materials on sites are continuously changing their

Peer-review under responsibility of the Conference Program Chairs.

position and keep interacting with each other. These dynamic interactions raise serious safety concerns to health and safety (H&S) managers.

In the last few decades, to reduce safety hazards, dynamic interactions of workers and machineries are tracked in real-time using different sensor-based location tracking technologies such as Radio Frequency Identification (RFID), Global Positioning System (GPS), Ultra wideband (UWB) and vision based sensing systems[4]. Each one of them has its own benefits and limitations. Among all, the most prominent and widely used tracking technology for construction safety applications is the GPS. It is used to automatically record the spatio-temporal trajectory data. A trajectory is a series of location points in the form of x and y geographical coordinates with time stamps generated by a moving object in space[5]. It holds multi-faceted attributes for example; time, position of the object in geographical coordinate system, direction of an object, speed of an object, change in direction, acceleration and distance travelled. These attributes can be extracted from spatio-temporal trajectories by applying pre-processing techniques. However, most decision making applications requires additional information with the trajectory data from the application context[5]. After an extensive literature review, a significant gap is found in the existing literature that the application of semantic trajectories is still missing in worker`s safety monitoring scenarios for effective safety management of construction sites. In order to fill this gap, at first, a system was designed to pre-process raw spatio-temporal trajectories of construction workers in our previous research article[6]. It collects GPS data values, performs cleaning and various other pre-processing algorithms to computes multi-faceted attributes such as speed, direction, etc. (already discussed above). Based on these attributes, trajectory reduction techniques were applied in order to visualize the large trajectories of construction workers. However, processed trajectories were incomplete to give a complete picture of mobility patterns of construction resources to H&S managers. It was lacking semantic information and contextual domain knowledge. Considering the pre-processing the very first step to prepare spatio-temporal trajectories for semantic enrichment, generating processed trajectory data from this prototype system as a foundation, this article presents an extension of it. Processed spatio-temporal trajectories of workers are semantically enriched using external data sources both private data such as construction site information describing the work zones and public data such as OpenStreet and Google Maps to enhance the ability of an effective construction site monitoring and increasing the possibilities to carry safe construction operations.

The paper is organized as follows: in section 2, a literature on transforming spatio-temporal trajectories to semantic trajectories with existing applications is discussed. In section 3, a system is presented to use semantic trajectories for improving worker safety on a construction site. Section 4 presents the discussion on the presented work and conclusion is discussed in section 5.

2. Background

There exist four primary areas in the existing literature on semantic trajectories: constructing sematic trajectories, segmenting, annotating, and applications.

Construction of trajectories

There exists two modes to construct a trajectory: (a) online mode, where trajectories are constructed in real time, and (b) offline mode, where all trajectories` construction processes are done in offline mode. Although, literature on an online construction of trajectories is limited but, there are many offline trajectory construction methods present in the literature[8,9,10]. In these methods, GPS data is collected in advance. Once mobility data is collected, it will undergo various processing stages such as data cleaning, map matching and compression[5]. However, these methods are not suitable for real life applications, where mobility data of objects is continuously updating. In order to address the requirement of an online trajectory construction for real life applications, a real-time solution known as SeTraStream[11] is present that has the ability to process raw trajectories data within a controlled time window and generate semantic trajectories in an online mode. In online processing, server receives the data from GPS devices with a predefined window size. As soon as data is received, trajectory basic attributes such as speed, direction, and displacement, etc. are calculated. Consequently, pre-processing techniques are applied[11]. During this process, trajectories in different batches are compared to one another to form trajectory episodes. Now, these episodes carry a start and end time that corresponds to a specific geometry of a building.

Segmentation of trajectories

Segmentation is a process to divide a trajectory into episodes based on some predefined criteria. The very first data model proposed by the authors[10], in which segmentation process is used to divide a trajectory into a set of moves and stops. A stop is defined as a place where a moving object has spent some specific time. However, other than time threshold, segmentation can also be achieved by other attributes such as velocity, acceleration, direction, density, and geographic artefacts. Yan et al. have presented a technique to segment mobility data into single-mode episodes by adding the functionality to distinguish between 10 different transportation modes[12]. Similarly, there exists an extended segmentation framework to segment trajectories based on the movement states[13]. However, their framework depends on the mapping of each movement state with relevant spatio-temporal criteria based on the expert knowledge and manual user input. There also exists smart frameworks such as SeTra[14] to segment GPS trajectories. It distinguishes between trip and activity trajectories episodes and maps the transportation mode to the episodes from a list of 8 segmentation modes list (run, walk, bicycle, urban vehicle, train, aerial and naval). Moreover, Sankararaman et al. have presented an approach to distinguish between similar and dissimilar portions in trajectories[15]. After then, trajectories are segmented into fragments to extract contiguous portions of trajectories that are shared by many of the other trajectories. Furthermore, segmentation can be done on the basis of representativeness[16]. Such techniques takes input from the Moving Object Databases (MOD) and performs a global voting algorithm based on the local density and similarity of trajectories information. Then segmentation and sampling algorithms are executed over the resulting segments to extract more representative sub-trajectories in the MOD.

Annotation of trajectories

Once the segmentation of trajectories is completed, annotation techniques are applied to transform GPS trajectories into semantic trajectories[6]. In this process, trajectory`s segments having specific intervals are assigned meaningful information regarding their mobility tracks. As trajectories are already divided into stops and moves episodes during the segmentation process, now it's important to mark stops because stopping in a location means that something interesting is happened. The annotation process involves enrichment of episodes with the meaningful information such as; the mapping of places of interests (POI) that can be in the form of points, in the form of lines or the geographical regions (see Table 2 for existing trajectory annotation applications)[6].

There are many annotation approaches present in the literature. Wu et al. have used historic social media data to map it with users` trajectories to understand the purpose of the trip[17]. Based on the location history, relevant words are extracted from the Twitter data according to the mobility records and user interests are retrieved for visiting a specific location at a specific time. In addition, user activities have also been used to annotate raw trajectories[18]. However, to cover a larger pool of POIs, there is a need of integration with more datasets such as Yellow Pages to enable tracking in large cities for the extraction of user activities[18]. Tales et al. have developed a new framework to annotate trajectories based on episodes[19]. It basically defines the environment where the user trajectory has been taken placed based on the Linked Open Data (LOD) cloud and OpenStreetMap. The ability of this framework of describing the spatial context of GPS trajectories can be used as a building block of future expert systems for trajectory exploration. Moreover, trajectories of slow and fast moving objects are also annotated at different levels of data abstraction using a multi-layer framework[20]. The locations where objects move provide with the information about their interests. At the same time, such behavioural analysis provides the information about the popularity of places[20]. A similar research effort is presented by Graaff et al., where an algorithm is proposed to combine existing multiple trajectory pre-processing methods to detect visited POIs to detect different indoor activities in urban settings[21]. Furthermore, ontological modelling approach has also been used to abstract trajectory data in a multilevel hierarchy using LOD collections and social media data[22]. It can enable to query trajectories for mobility analysis-

Table 1. Comparison of existing semantic annotations of trajectories

Use case	Findings	Key technologies/ methods used	Dataset	Annotation types comparison			
				Social media	Point	Region	Line
Semantic annotation of mobility data using social media[17]	Extracted purposes and interests of a user from his location history.	Kernel density estimation model	Geo-tagged 37,871,265 tweets from New York City, Chicago and Los Angeles	Y	Y	N	N
Inferring human activities from GPS tracks[18]	Automatically annotated raw trajectories with the activities performed by the users.	Gravity law	One year data of 28 users moving by car, for around 30,000 annotated trips	N	Y	N	N
Annotating semantic trajectories based on episodes[19]	Environments are identified where trajectories took place.	Linked Open Data cloud and OpenStreetMap	23 trajectories of a jogger collected in the city of Grenoble.	N	N	Y	N
Semantic annotation of heterogeneous trajectories[20]	Annotated trajectories for any kind of moving objects.	Java 6 (PostgreSQL)	3 million GPS records of two Lausanne taxis and 2 million GPS records of 17,241 private cars.	N	Y	Y	Y
Automated semantic trajectory annotation with indoor point-of-interest visits in urban areas[21]	Combined multiple trajectory pre-processing techniques to extract POIs.		23 trajectories from a wide range of device types, containing a total 30,500 trajectory samples.	N	Y	N	N
Automated semantic annotations based on existing knowledge bases[22]	Structuring and abstracting in a multilevel hierarchy of progressively detailed movement segments is achieved.	geoSPARQL and PostGIS	14,504 positions of 564 distinct users, was segmented in 2143 user's trail from Flickr and 57,099,806 tweets.	Y	Y	Y	Y
Dynamic semantic annotation of trajectories[23]	Annotation is achieved using contextual social media data.	Kernel density estimation model	50 million geo-tagged tweets	Y	N	N	N

-based on the domain and application related knowledge. In addition to automatic semantic annotations, there also exists dynamic semantic annotation methods based on contextual data in the literature[23]. Such methods calculate local density of words and maps words to each trajectory record, hence providing visualizations for trajectories exploration.

3. Semantic enrichment system for improving worker safety

After reviewing the literature regarding all the steps involved in semantic enrichment of trajectories, a prototype system is discussed to monitor workers` mobility on construction sites using semantic spatio-temporal data. Building supervisor and H&S manager are the two roles have been identified from the literature[24] to develop this prototype system as mentioned in Fig.1. The developed prototype system focuses on the following:

1. It aggregates the GPS data from workers` handheld devices on a construction site. A single GPS data log consists of a user identification (ID), timestamp, floor, latitude and longitude value. An application programming interface is used to capture the GPS data through wireless access points, aggregating and then storing it in a document oriented database such as Mongdb. A data link is configured between a database system and R studio to pre-process the captured GPS data. There exists many noise filters, such as mean and median filter to improve the data quality. However, median filter is used on a captured GPS data (as shown

is Fig. 2) because it depicts robustness property in filtering and recommended for data with high sampling rate whereas, mean filter is not proposed because it is highly sensitive to outliers[5].

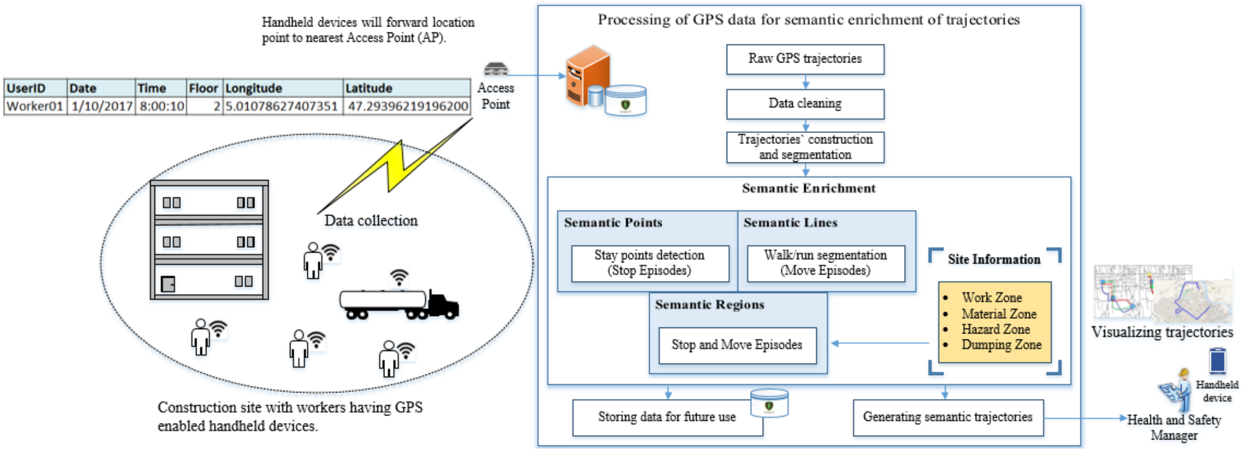

Fig. 1. Scenario for monitoring spatio-temporal data for worker safety on a construction site

2. After removing the outliers in a trajectory data, stay points of a worker are calculated to enrich a worker trajectory with semantic points in the form of stop and move segments. Stay points are the location points where a worker has spent a significant time within a specified distance. By manually setting the distance (D_{thresh}) and time threshold (T_{thresh}) values, stay points in a trajectory have been identified (as shown in Fig. 3) using Zheng et al. approach[5]. A single stay point 's' can be treated as a virtual location point characterized by a set of successive GPS points $Z = \{z_m, z_{m+1}, z_{m+2}, \dots, z_n\}$, $\forall\, m < i \leq n$, $Distance(z_m, z_i) \leq D_{thresh}$ and $|z_n.T - z_m.T| \geq T_{thresh}$. A stay point can be described as $s = (Latitude, Longitude, arrivaltime, leavingtime)$.
Where,

$$s.latitude = \sum_{i=m}^{n} {z_i.Latitude}/{|Z|}$$

$$s.lontitude = \sum_{i=m}^{n} {z_i.Longitude}/{|Z|}$$

The purpose of calculating stay points in a worker trajectory data is to find locations on a construction site where workers are spending more time than required. This information will help to track the occurrence of an unexpected situation on a site if a stay duration is greater or less than the required. After extracting the stay points, successive stay points on a same location are grouped together and aggregated stay duration in minutes is calculated (as shown in Fig. 4) in order to reduce the size of a trajectory.

3. Once stay points are identified, we will map them to corresponding site locations having unique identification for actual site areas such as work zone, material zone, hazard zone and dumping zone as shown in Table 1. Construction sites typically have more work zones[25] however, we have restricted site zones to 4 for this article. To tag site identification (ID), trajectory location points are spatially joined with OpenStreetMap data file. In this way, we can annotate each position of a trajectory with the locations of the site where the worker has visited as shown in Fig. 5.

Table 1. Mapping of GPS data with the construction site information

GPS coordinates [Lon, Lat]	Geometry type	Site Area	Site ID	Building Name
{ [5.010756, 47.293998] , [5.010729, 47.294025], [5.010765, 47.294017] , [5.010756, 47.293998], [5.010710, 47.294010] }	Polygon	Workzone	W5768	Tevoyls

4. After mapping the spatio-temporal locations in a trajectory to site identifications, we can now enrich a trajectory with semantic lines in the form of movement types which are being carried by workers. Based on our application requirement, we have used walk-based segmentation method of Zheng et al. that is based on semantics[5]. Walk points and run points based on points` speed values are calculated then, a trajectory is divided into alternate walk and run segments as shown in Figure 6(a). Moreover, spend is also calculated between two segments in meters per second. Finally, to visualize the entire trajectory having stop and move segments with a semantic region, R studio is used with Google Maps as shown in Fig. 6(b).

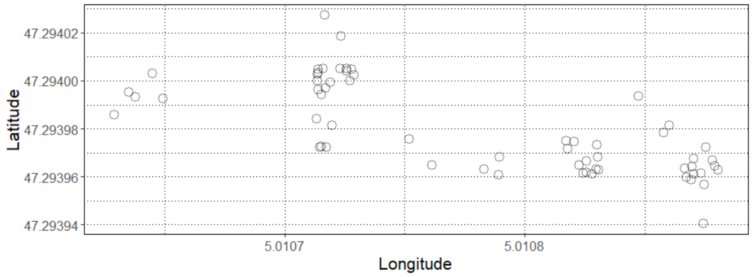

Fig. 2 Filtered raw trajectory data of a worker

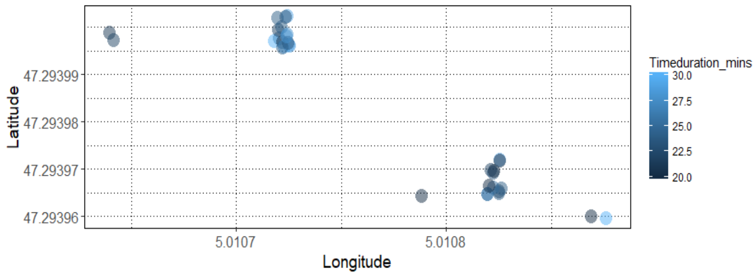

Fig. 3 Stay points extraction with time spent

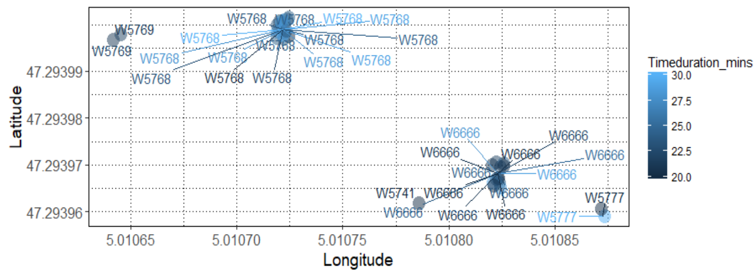

Fig. 4 Semantic enrichment of stay points with site location IDs

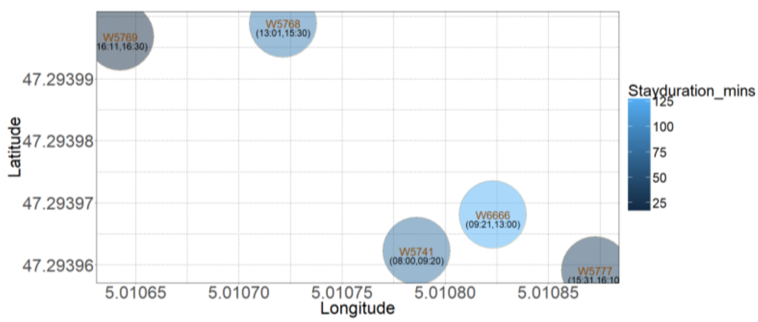

Fig. 5 Visualization of semantic stay points with stay durations calculated in minutes after aggregating successive stay points

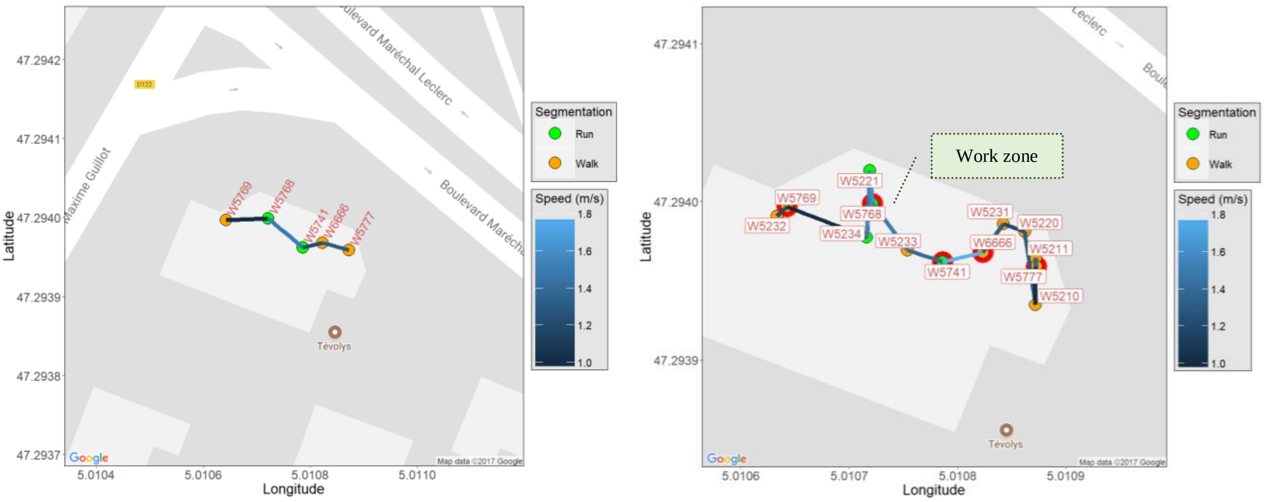

Fig. 6 (a) Enriching a worker trajectory with run and walk segments, and calculating speed between these segments in meters per second (b) Visualizing a complete worker` trajectory with semantic stay points in red colored circles (site identifications for e.g. W5768 for a workzone), semantic lines (segmentation into run and walk points) and a semantic region

4. Discussion and Conclusion

This work presented the application of a GPS data to track and analyse workers` trajectories for safer construction sites. The proposed system used construction site information to enrich pre-processed spatio-temporal GPS data of workers to have more meaningful visualizations of trajectories. Though, mobility data used in the proposed system is not processed in real-time but previously collected Comma Separated Value (CSV) file carrying location data is used. If the time spent by each worker is accurately calculated in different construction site zones, it will provide a good understanding of its utilization and safety plans for these site zones can be improved accordingly. In addition, visualization of workers` trajectories that have been generated using a developed system can help to monitor how often they enter in a specific site zone and how much time they have spent in each zone. Moreover, important information that H&S managers like to monitor is the traveling speed of the workers and machineries on site. Objects moving at a higher than a specified speed threshold inside a particular site zone would be considered unsafe. This information works as a leading indicator for worker safety and can help in the detection of potential hazardous zones on a site. Furthermore, people or machineries that are working close to each other can cause an accident. Visualizing complete trajectories mapped with site zones will help for proximity analysis and shows which, where and when the worker or a machinery is operating too close to another. However, further work is need to be done to fuse data from other emerging technologies such as Building Information Modeling (BIM) with spatio-temporal data to visualize processed trajectories with the building infrastructure context. In addition, further applications should be developed to incorporate pattern mining techniques to take complete advantage of fused information for deriving mobility behaviours of workers that will ultimately help H&S managers for better decision making in safety processes.

Acknowledgements

The authors thank the Conseil Régional de Bourgogne-Franche-Comté and the French government for their funding.

References

[1] Stats.bls.gov. (2017). "Census of Fatal Occupational Injuries (CFOI) - Current and Revised Data", available at: http://stats.bls.gov/iif/oshcfoi1.htm (Accessed 28 September 2017).
[2] http://ec.europa.eu/eurostat/statistics-explained/index.php/Accidents_at_work_statistics
[3] Li, H., Lu, M., Hsu, S.C., Gray, M. and Huang, T., 2015. Proactive behavior-based safety management for construction safety improvement. Safety science, 75, pp.107-117.

[4] Awolusi, I., Marks, E. and Hallowell, M., 2018. Wearable technology for personalized construction safety monitoring and trending: Review of applicable devices. Automation in Construction, 85, pp.96-106.

[5] Yu Zheng. 2015. Trajectory Data Mining: An Overview. ACM Trans. Intell. Syst. Technol. 6, 3, Article 29 (May 2015), 41 pages. DOI: http://dx.doi.org/10.1145/2743025

[6] Muhammad Arslan, Christophe Cruz, Ana-Maria Roxin and Dominique Ginhac, Using Spatio-temporal Trajectories to Monitor Construction Sites for Safety Management, ICIME, 2017.

[7] Albanna, B.H., Moawad, I.F., Moussa, S.M. and Sakr, M.A., 2015. Semantic trajectories: a survey from modeling to application. In Information Fusion and Geographic Information Systems (IF&GIS'2015) (pp. 59-76). Springer International Publishing.

[8] L. O. Alvares, V. Bogorny, B. Kuijpers, J. Macedo, B. Moelans, and A. Vaisman. A Model for Enriching Trajectories with Semantic Geographical Information. In GIS, 2007.

[9] M. Buchin, A. Driemel, M. V. Kreveld, and V. Sacristan. An Algorithmic Framework for Segmenting Trajectories based on Spatio-Temporal Criteria. In GIS, 2010.

[10] S. Spaccapietra, C. Parent, M. L. Damiani, J. A. de Macedo, F. Porto, and C. Vangenot. A Conceptual View on Trajectories. Data and Knowledge Engineering, 65(1), 2008.

[11] Yan, Z., Giatrakos, N., Katsikaros, V., Pelekis, N. and Theodoridis, Y., 2011. SeTraStream: semantic-aware trajectory construction over streaming movement data. Advances in Spatial and Temporal Databases, pp.367-385.

[12] Zhixian Yan, Dipanjan Chakraborty, Christine Parent, Stefano Spaccapietra, and Karl Aberer. 2011. SeMiTri: a framework for semantic annotation of heterogeneous trajectories. In Proceedings of the 14th International Conference on Extending Database Technology(EDBT/ICDT '11), Anastasia Ailamaki, Sihem Amer-Yahia, Jignesh Pate, Tore Risch, Pierre Senellart, and Julia Stoyanovich (Eds.). ACM, New York, NY, USA, 259-270. DOI=http://dx.doi.org/10.1145/1951365.1951398

[13] Buchin M., Kruckenberg H., Kölzsch A. (2013) Segmenting Trajectories by Movement States. In: Timpf S., Laube P. (eds) Advances in Spatial Data Handling. Advances in Geographic Information Science. Springer, Berlin, Heidelberg.

[14] W. Balzano and M. R. D. Sorbo, "SeTra: A Smart Framework for GPS Trajectories' Segmentation," 2014 International Conference on Intelligent Networking and Collaborative Systems, Salerno, 2014, pp. 362-368.

[15] Sankararaman, S., Agarwal, P.K., Mølhave, T., Pan, J. and Boedihardjo, A.P., 2013, November. Model-driven matching and segmentation of trajectories. In Proceedings of the 21st ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems (pp. 234-243). ACM.

[16] Panagiotakis, C., Pelekis, N., Kopanakis, I., Ramasso, E. and Theodoridis, Y., 2012. Segmentation and sampling of moving object trajectories based on representativeness. IEEE Transactions on Knowledge and Data Engineering, 24(7), pp.1328-1343.

[17] Wu, Fei & Li, Zhenhui & Lee, Wang-Chien & Wang, Hongjian & Huang, Zhuojie. (2015). Semantic Annotation of Mobility Data using Social Media. 1253-1263. 10.1145/2736277.2741675.

[18] Barbara Furletti, Paolo Cintia, Chiara Renso, and Laura Spinsanti. 2013. Inferring human activities from GPS tracks. In Proceedings of the 2nd ACM SIGKDD International Workshop on Urban Computing (UrbComp '13). ACM, New York, NY, USA, Article 5, 8 pages. DOI: https://doi.org/10.1145/2505821.2505830

[19] Tales P. Nogueira, Reinaldo B. Braga, Carina T. de Oliveira, Hervé Martin, FrameSTEP: A framework for annotating semantic trajectories based on episodes, In Expert Systems with Applications, Volume 92, 2018, Pages 533-545, ISSN 0957-4174, https://doi.org/10.1016/j.eswa.2017.10.004.

[20] Zhixian Yan, Dipanjan Chakraborty, Christine Parent, Stefano Spaccapietra, and Karl Aberer. 2011. SeMiTri: a framework for semantic annotation of heterogeneous trajectories. In Proceedings of the 14th International Conference on Extending Database Technology(EDBT/ICDT '11), Anastasia Ailamaki, Sihem Amer-Yahia, Jignesh Pate, Tore Risch, Pierre Senellart, and Julia Stoyanovich (Eds.). ACM, New York, NY, USA, 259-270. DOI=http://dx.doi.org/10.1145/1951365.1951398

[21] Victor de Graaff, Rolf A. de By, and Maurice van Keulen. 2016. Automated semantic trajectory annotation with indoor point-of-interest visits in urban areas. In Proceedings of the 31st Annual ACM Symposium on Applied Computing (SAC '16). ACM, New York, NY, USA, 552-559. DOI: https://doi.org/10.1145/2851613.2851709.

[22] Renato Fileto, Cleto May, Chiara Renso, Nikos Pelekis, Douglas Klein, Yannis Theodoridis, The Baquara2 knowledge-based framework for semantic enrichment and analysis of movement data, In Data & Knowledge Engineering, Volume 98, 2015, Pages 104-122, ISSN 0169-023X, https://doi.org/10.1016/j.datak.2015.07.010.

[23] Fei Wu, Hongjian Wang, Zhenhui Li, Wang-Chien Lee, and Zhuojie Huang. 2015. SemMobi: A Semantic Annotation System for Mobility Data. In Proceedings of the 24th International Conference on World Wide Web (WWW '15 Companion). ACM, New York, NY, USA, 255-258. DOI: http://dx.doi.org/10.1145/2740908.2742837

[24] Mattila, M. and Hyödynmaa, M., 1988. Promoting job safety in building: An experiment on the behavior analysis approach. Journal of Occupational Accidents, 9(4), pp.255-267.

[25] Pradhananga, N. and Teizer, J., 2013. Automatic spatio-temporal analysis of construction site equipment operations using GPS data. Automation in Construction, 29, pp.107-122.