Smart and Sustainable Built Environment

Spatio-temporal Analysis of Trajectories for Safer Construction Sites

Journal:	*Smart and Sustainable Built Environment*
Manuscript ID	SASBE-10-2017-0047.R1
Manuscript Type:	Original Research Paper
Keywords:	safety, construction, workers, technology, health, Fatalities

Spatio-temporal Analysis of Trajectories for Safer Construction Sites

Muhammad Arslan, Christophe Cruz, Ana-Maria Roxin, Dominique Ginhac
Le2i, CNRS FRE 2005, Arts et Metiers, Univ. Bourgogne Franche-Comte, Batiment i3M rue Sully, Dijon France

ABSTRACT

Purpose – The purpose of this paper is to improve safety of construction workers by understanding their behaviors on construction sites using spatio-temporal (ST) trajectories.

Design/methodology/approach – A review of construction safety management literature and international occupational health and safety statistics shows that the major reasons of fatalities on construction sites are due 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. Consequently, literature is reviewed to find possible technological solutions to track mobility of construction workers to reduce fatalities. This examination has suggested that location acquisition systems such as Global Positioning System (GPS) has been widely used for real time monitoring and tracking of workers on construction sites for hazards prevention. However, the raw data captured from the GPS device is generally available as discrete points and do not hold enough information to understand the workers` mobility. As a solution, an application to transform a raw GPS data into ST trajectories using different preprocessing algorithms is proposed for worker safety on construction sites.

Findings – A proposed system preprocesses raw GPS data for stay points detection, trajectory segmentation and intersection of multiple trajectories to find significant places and movements of workers on a construction site to enhance the information available to H&S managers for decision making processes. In addition, it reduces the size of trajectory data before saving for future analyses.

Originality/value – Application of location acquisition systems for construction safety management is very well addressed in the existing literature. However, a significant gap has been found that the usage of preprocessed ST trajectories is still missing in workers' safety monitoring scenarios in the area of construction management. To address this research gap, our proposed system uses preprocessed ST trajectories to monitor workers` movements on a construction site to identify potential unsafe behaviors.

Keywords

Health and Safety (H&S); construction sites; fatal accidents; mobility

1. INTRODUCTION

International occupational health and safety statistics shows that the construction industry experiences one of the highest accident rates of all industries (Stats.bls.gov., 2017). According to the National Census of Fatal Occupational Injuries conducted by U.S. Bureau of Labor Statistics in 2015, out of 4,836 fatal work injuries 19 % of fatalities were recorded from the construction industry (Stats.bls.gov., 2017). 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. Despite numerous efforts to reduce fatalities on construction sites, such accidents continue to occur as shown in Figure 1.

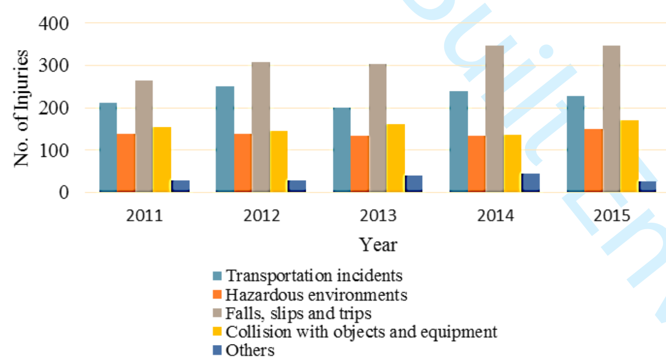

Figure 1. Fatal injuries in U.S. construction industry (Stats.bls.gov., 2017)

The advancements in location acquisition and mobile communication techniques based on GSM (Global System for Mobile Communications), Bluetooth, Wi-Fi and other wireless sensing technologies offers various opportunities to collect mobility data from construction sites for safety management (Neirotti *et al*., 2014). With the help of GPS equipped devices and vehicles, Radio Frequency Identification (RFID) tag tracking and location-aware wireless sensors, the possibilities to collect mobility data is increased significantly (Jin *et al*., 2014). These technologies can generate huge amount of data that contains time varying geographic positions of moving objects called trajectories (Patroumpas and Sellis, 2017). Such trajectories captured from moving objects have unique features; they are in the form of streaming data which means that data will keep on updating over time and includes spatial information (Patroumpas and Sellis, 2017). Trajectories with such information are known as ST trajectories. In recent years, it is technically more convenient and economically cheap to acquire ST trajectories of moving objects from a wider region in less time (Andrienko *et al*., 2013). The analysis of such trajectories can lead to establish better understanding and knowledge discovery about moving objects (Miller and Han, 2009).

The trajectories of moving objects are continuous in nature in real world scenarios (Šaltenis *et al*., 2000). Because of data collection and storage devices limitations, continuous movements of moving objects are collected and stored in a discrete format such as a sequence of GPS points having longitude, latitude and timestamp values. As tracking time of GPS devices increases, the trajectory data in applications will gradually grows. Such huge amount of data can sooner or later causes storage challenges. Therefore, trajectory data reduction should be done before trajectory data management (Patroumpas *et al*., 2016). The primary aim of trajectory reduction is to decrease the computational complexity and reducing the time to query the trajectory data (Patroumpas *et al*., 2016). Once trajectory data is processed and reduced, it is now ready to be saved in a trajectory database. Many applications managing mobility data require semantic understanding of data as well, that is not possible to extract directly from raw GPS trajectories. As the physical trajectories can easily be recorded using GPS devices but the semantic interpretation of the mobility data is still a big research challenge (Krisnadhi *et al*., 2016). In general, a trajectory is the path that is followed by a moving object in space as a function of time (Nardini *et al*., 2018). It is acquired as a series of location points having time stamps, denoted as follow;

$$\{\langle x_1, y_1, t_1\rangle, \langle x_2, y_2, t_2\rangle, \dots, \langle x_N, y_N, t_N\rangle\}$$

Here, x_i and y_i represent x and y geographical coordinates at time t_i, N is the total number of points in the trajectory data and it is usually followed by a sampling interval. There are many approaches to collect trajectory data. It can be based on the change in time, location, event or various combination of these approaches. As trajectory data holds multi-faceted characteristics, these are used to analyze and understand mobility (Zheng *et al*., 2017). Mobility related characteristics include: time (i.e., position of object mobility on the timescale), position of the object in geographical coordinate system, direction of the object, speed of the object, change in direction, acceleration (i.e., change in speed) and distance travelled. These characteristics can directly be computed from the raw trajectories and are well-suited for the applications that are performing localization of objects in motion. However, most decision making applications requires additional information with the trajectory data from the application context.

Other than location related characteristics of trajectories, there exists an important characteristic of mobility data that is the variety of various travel means by which mobility has taken place (Adey, 2017). This is of special interest in domains such as urban planning as selection of a location and a travel mean may be interconnected. For example, to analyze trajectories of a person going from home to office in a city, it is necessary to have some information about the city and its infrastructure such as information about shops and restaurants etc. Such data would help to visualize person going to office trajectories in more detail in terms of the point of interest locations rather than just by the geographical coordinates of the location. The process of supplementing the GPS trajectories with such additional data is known as semantic enrichment process (Shen and Cheng, 2017). This additional data categorizing stops and moves and distinguishing different types of moves is known as annotation that is attached to a trajectory either to some of its parts or as a whole (Yan *et al*., 2013). An annotation value is simply an attribute value that can be an "on-tram" or an "on-bus", a possible value for TransportationMeans annotation in case of a person going to an office scenario. An example of semantically enriched trajectory could be the following (Yan *et al*., 2013).

(Begin, home, 8am, -) → (move, road, 8am-8:30am, walk) → (move, road, 8:30am-9am, on-tram) → (stop, office, 9am-5pm, work) → (move, road, 5pm-5:30pm, on-tram) → (move, road, 5:30pm-6:00pm, walk) → (End, home, 6:00pm, -)

An above example includes generic movement characteristics (e.g., stops and moves), application specific geographical objects (e.g., office and work) and also additional behavioral context (e.g., work) (Mousavi, 2016). Once trajectories are annotated and stored in a trajectory database, it can be used to detect and analyze trends and behaviors of moving objects.

The paper is organized as follows: in section 2, research method is discussed and the existing applications of ST trajectories are presented. Section 3, is based on the preprocessing of raw trajectories using different algorithms for worker safety on a construction site application. Section 4 presents the discussion of the presented work and conclusion is discussed in section 5.

2. BACKGROUND

A literature review is initially done to methodically collect information to recognize and understand the problem domain. At first, a review of construction safety management literature and international occupational health and safety statistics shows that the major reasons of fatalities on construction sites are related to the unsafe human behaviors, difficult site conditions, and workers falling from heights and striking against or being struck by moving objects (Hsiao, 2016; Stats.bls.gov., 2017). Existing research shows that these reasons can be reduced if the mobility of workers and machineries are continuously monitored on construction sites for their behaviors (Guo *et al*., 2017). This capability of monitoring construction sites in real-time for decision support systems will lead to the development of an intelligent job site. Based on this literature, below mentioned competency questions have been formulated to enable intelligent monitoring of construction sites for building supervisors and H&S managers. Secondly, literature is also reviewed to find possible technological solutions to monitor mobility of construction workers. This examination has suggested that location acquisition systems such as Global Positioning System (GPS) has been used widely for real time monitoring and tracking of workers on construction sites for hazard preventions and safety management. However, the raw data captured from the GPS device is generally available as discrete points and do not hold enough information to understand the workers` mobility. As a solution, an application to transform a raw GPS data into ST trajectories using different preprocessing algorithms is proposed in order to address the competency questions that have been formulated through the literature review to reduce fatalities on construction sites.

Table 1. Selected competency questions

Competency questions related mobility
What are the stay points of workers or machineries on a construction site?
Which type of movements are done by the workers?
Which and how many workers or machineries are close to other workers or machineries?
How many workers are present in a specified area?
What is the closest point-pair distance between two workers or machineries?

In order to answer the questions as mentioned in Table 1, existing literature is reviewed for existing applications based on ST trajectories. However, before using ST trajectories in applications, it should be processed according to the applications` requirements. Trajectory data processing deals with the algorithms for trajectory reconstruction. For an initial preprocessing, research has been done to clean the data by removing the noise and outliers and then compressing it. Noise in trajectories is a small distortion that has the potential to affect speed, acceleration and the travel distance estimates. However, their impact can be reduced by smoothing techniques (Marketos *et al*., 2008). Marketos *et al*. proposed an online approach to filter outliers in a trajectory data by taking the maximum allowed speed of the moving object as a reference point (Marketos *et al*., 2008). In addition, Jun et al. proposed a modified version of Kalman filter to control the outliers in GPS data with more effective technique (Jun *et al*., 2006). Their algorithm demonstrates superior results as compared to other data smoothing techniques and it can be used for the real time data smoothing. There are also many research efforts discuss about the data compression methods of a trajectory data. For instance, Meratnia and de By proposed an algorithm that considers both temporal and spatial parameters in order to compress trajectories (Meratnia and Deby, 2004). In their proposed technique, distance and speed thresholds have been used for data compression. Although, choosing proper thresholds are totally dependent on the application and having a clear understanding of an object in motion will help in the selection of appropriate thresholds. Their proposed algorithms can be used for an offline as well as for an online data reduction. After data compression, segmentation is another important step to understand mobility data. Buchin et al. proposed a framework to segment a trajectory based on ST criteria (Buchin *et al*., 2011). Segmentation is achieved in such a way that each generated segment is made homogenous and it should fulfill the ST criteria. These criteria include location, speed, heading, velocity, curvature and shape. Under any on these criteria or any combination of these criteria, trajectory segmentation can be achieved in an efficient and optimal way.

After preprocessing of GPS trajectories data, it can now be used in variety of applications. As data acquired from GPS devices lacks in semantic information for the behavior analysis of a moving object. These GPS trajectories can be further annotated using tools and techniques offered by data mining and machine learning concepts. However, it will raise the requirement for a training data. Trajectories resulting from annotation processes have been used in various applications such as urban indoor activity detection and wild life monitoring (Hu *et al*., 2013; Urbano and Cagnacci, 2014). Such applications will help to understand the factors behind the movement of moving objects and these factors can be dependent on each other (Noël *et al*., 2015). In Vandecasteele et al, a prototype is developed for maritime surveillance system for better understanding the vessel trajectories (Vandecasteele *et al*., 2014). A major contribution of proposed system is to offer a reliable way to model semantic trajectories as well as semantic events. Geospatial trajectories have also been enriched with domain knowledge and contextual information to construct trajectory data-ware houses in order to support business intelligence processes in any organization (Brisaboa *et al*., 2017). These data-ware houses are used to offer context-aware services such as recommender systems or digital assistance providing higher level of personalization to people for travelling (Li *et al*., 2017). Once trajectories dataset is processed and maintained, hidden patterns can also be extracted. Moreover, this will not only benefit in the exploration of important aspects of semantic trajectories but can also help to automate the processes of clustering the trajectories based on the stored information using machine learning techniques (Lv *et al*., 2012). Table 2 presents the summary of ST applications that have been discussed above.

Application of ST trajectories for different systems is very well addressed in the existing literature as shown in Table 2. However, a significant gap is found that the usage of preprocessed ST trajectories is still missing in workers' safety monitoring scenarios in the area of construction management. To address this research gap, a prototype system is proposed that processes ST trajectories based on the data processing algorithms that are selected from the existing literature to monitor workers` movements on a construction site to identify potential unsafe behaviors in order to reduce fatalities.

3. CONSTRUCTION WORKERS SAFETY MONITORING APPLICATION

After an extensive review of applications, it is concluded that, ST trajectories can also be used to track workers on construction sites. For this, a prototype system application is discussed to achieve the goal of visualizing trajectory data for safety management of workers (Arslan *et al*., 2017). Building supervisor and H&S manager are the two roles have been identified from the literature for managing construction sites (Arslan *et al*., 2014). These roles have been taken into account for the development of a prototype system as mentioned in Figure 2. The prototype system application focuses on the following:

1. Capturing GPS data values from construction workers` handheld devices on a construction site.
2. Aggregating and saving the collected GPS data values on a cloud server having a centralized database management system configured.

Table 2. Applications of ST trajectories

Use cases	**Detecting semantic outliers from moving objects (Chakri *et al.*, 2017).**	**Semantic trajectory compression for representing urban movements (Richter *et al.*, 2012).**	**Compressing trajectory data by exploiting the semantic embedding of movement in a geographical context (Schmid *et al.*, 2009).**	**A geo-ontology design pattern for semantic trajectories (Hu *et al.*, 2013).**	**Semantic middleware for trajectories to enable annotating trajectories (Yan *et al.*, 2011).**	**Automated semantic trajectory annotation with indoor point-of-interest visits in urban areas (Wagner *et al.*, 2013).**	**Application of life trajectories for the modeling and analytics of human mobility in the urban context (Noël *et al.*, 2015).**
Building indoor/outdoor	Outdoor	Outdoor	Outdoor	Outdoor	Outdoor	Indoor	Outdoor
Dataset	Dataset of school buses dataset consists of 145 trajectories of two school buses. The second dataset is of are trucks which consists of 276 trajectories of 50 trucks.	Everyday trajectories reflecting movement between regularly visited places of four persons.	Data contains geometric representation of the path contains 115 points in space-time (115 tuples of (x,y,t)). It further comprises 52 events, i.e., 52 intersections and stops along the way.	An individual's trajectory data recorded by a handheld GPS receiver and an animal tracking data retrieved from the MoveBank, an online database.	(1) 3 million GPS records of two Lausanne taxis, collected over 5 months (2) 2 million GPS records of 17,241 private cars tracked in Milan during one week; (3) a GPS trace of 2-hour drive of a private car in Seattle.	Data from 132 mobile devices containing the trajectory id and fields of latitude, longitude, altitude, accuracy, speed, and timestamp.	Data of fifty people to assess the model.
Findings	Extracted outliers from semantic trajectories to understand the unusual behaviors.	Proposed algorithm compresses trajectory data with a higher compression rate (89.53%) and drastically reduces the amount of data to be stored.	STC algorithm achieves a high compression rate. Instead of the 115 original points, it ends up with only 6 items, which corresponds to a compression rate of 94.78%.	Proposed design pattern is used to semantically annotate trajectory data of navigation and wildlife monitoring.	A framework is proposed to support semantic enrichment of trajectories exploiting both the geometric properties of the stream and the background geographic and application data.	Proposed an algorithm for the automated detection of visited points-of-interest. It extracts the actual visited points-of-interest well in terms of precision for the challenging urban indoor activity detection.	Proposed model allows a better understanding the reasons why and circumstances in which people are moving, whether they depend on external or on internal factors.
Key components/ technologies	Weka-STPM toolkit	-	GPS	Web Ontology Language (OWL)	Java 6 platform, PostgreSQL 8.4 with spatial extension PostGIS 1.5.1	PhoneGap platform	-

Table 2. Applications of ST trajectories (continued)

Use cases	**Maritime surveillance system based on semantic events (Vandecasteele *et al.*, 2014).**	**Semantic trajectories in Mobile Workforce Management (MWM) applications (Brisaboa *et al.*, 2017).**	**Mob-warehouse, a trajectory data warehouse to enrich trajectory data with domain knowledge (Wagner *et al.*, 2013).**	**Personalized RoutE Guidance System (PaRE) to provide higher-quality navigation directions (Li *et al.*, 2017).**	**Efficient frequent sequence mining on taxi trip records using road network shortcuts (Zhang, 2014).**	**Extraction of hidden shared structure among human trajectories (Chen *et al.*, 2013).**	**A semantic trajectory data model to define important aspects of semantic trajectories (Bogorny *et al.*, 2014).**	**Discovering semantic places from GPS trajectories (Lv *et al.*, 2012).**
Building indoor/outdoor	Outdoor	Indoor and outdoor	Outdoor	Outdoor	Outdoor	Outdoor	Outdoor	Outdoor
Dataset	More than one million vessels positions with attributes such as the unique identification number of the vessel, the position, speed and rate of turn.	Data retrieved from the MWM system, the sensors of the mobile device and the GIS.	A trajectory dataset of people traveling by car in Milan (Italy), during one week. The dataset contains track of 16,946 cars and 48,906 trajectories for a total of 1,806,293 points.	Real trajectory dataset from Planet.gpx, which contains the GPS traces uploaded by OpenStreetMap users within 7.5 years.	17,558 taxi trip records in New York City over a month's period.	A collection of 230,000 GPS traces of taxi cabs in Beijing, China over a month's period.	A dataset of trajectories of tourists moving in Rome to visit the main city attractions.	Real GPS trajectories collected from 10 participants for nearly two months.
Findings	A prototype is developed to allow the maritime operator to better understand changes in the velocity of a vessel by analyzing semantic trajectories and semantic events.	Supporting business intelligence processes by collecting the information captured by the sensors of the mobile devices and analyzing and annotating the trajectory high-level activities using context information.	A model is proposed, where the ST component of trajectory data is properly integrated with context related information like transportation means, performed activities and mobility patterns.	Proposed system can reduce the number of navigation directions by more than 60% while still providing enough information for user to follow the route.	Results shows that runtimes of frequent sequence mining on shortcut sequences are orders of magnitude faster than on original road segment sequences.	"Pathlet Dictionary" is introduced to represent spatial regularities in a trajectory dataset and presented an effective algorithm to learn pathlet dictionaries from large collections of trajectories.	More semantics are added to raw trajectory data for real applications such as; tourism and animal behavior.	Hierarchical clustering algorithm is proposed to extract visit points from the GPS trajectories, and then these visit points can be clustered to form physical places.
Key components/ technologies	Triple store, Semantic Web Rules Language (SWRL), ST Inference engine	MWM system, OpenStreetMap and Android mobile devices	-	OpenStreetMap and Java platform	Spatial databases (PostgreSQL, PostGIS) and ArcGIS	-		

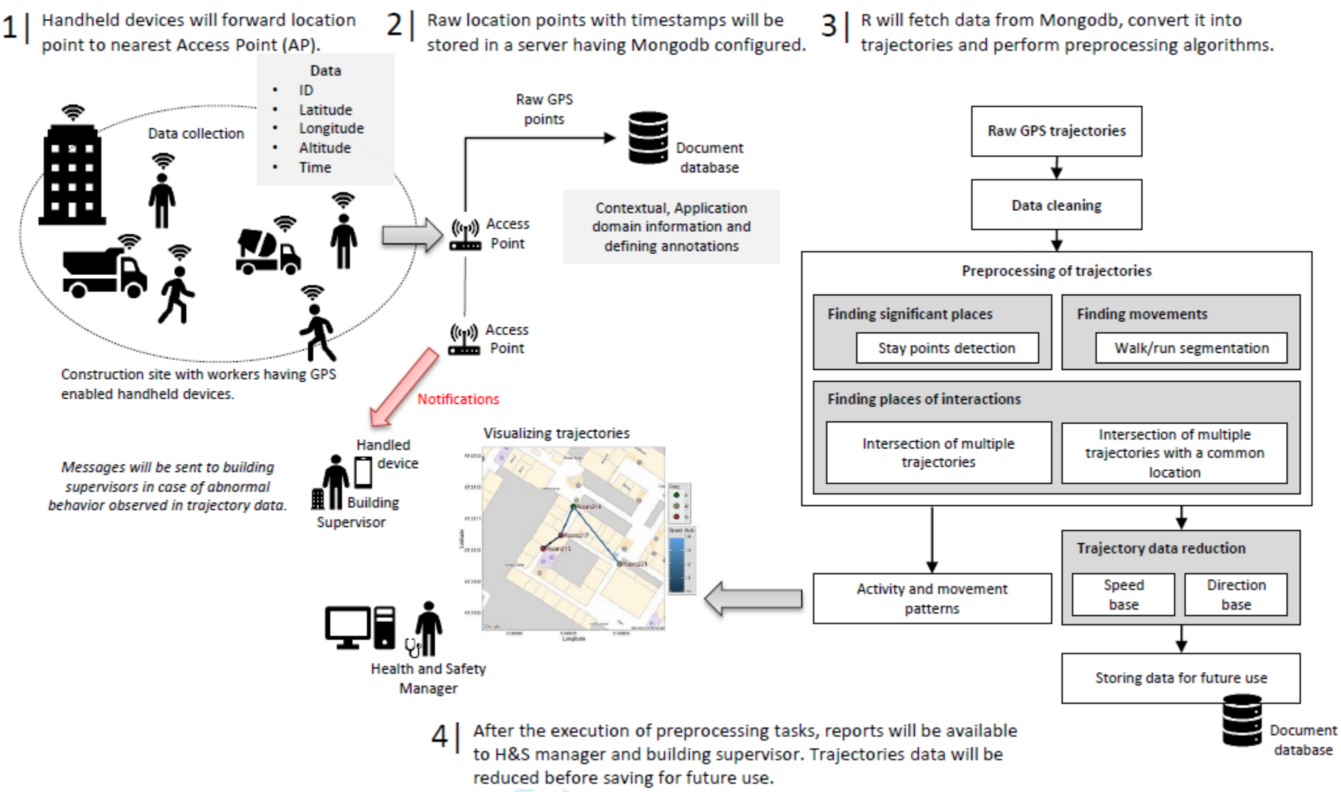

Figure 2 Preprocessing ST trajectories for safety management on construction sites

3. Transforming GPS data values into raw trajectories, executing data processing algorithms on a raw trajectory data to extract the movements of workers and reducing the trajectory data before storing for future analysis, once data processing tasks are completed.
4. Presenting trajectories visualizations to building supervisors and H&S manager for an effective construction site monitoring and ensuring safety at work.

The prototype system uses document oriented database such as Mongodb for data storage and R platform for processing the GPS data. Mongodb server is used because of open source cloud based solution, stores massive datasets efficiently and requires simple configurations with a programming platform. Whereas, R platform is used because of open source solution offers integrations with JSON based databases such as Mongodb and provides packages to support ST data for its processing and visualizations.

The prototype system can be divided into three main layers which are: GPS data acquisition, processing the GPS data and generation of its visualizations. The data acquisition layer consists of smartphones that sends GPS data of workers at a defined interval. A single GPS data record consists of a worker identification (ID), timestamp, altitude, latitude and longitude value. An application programming interface is designed to acquire GPS data from the workers` smart phone through wireless access points, aggregating and then storing it in a Mongodb server that is configured on a cloud. A data connection is established between a Mongodb and R studio to process the trajectories data. Mentioned below are the tasks that R studio will execute after retrieving data from a Mongodb server.

3.1 Data Cleaning

Data cleaning is the first step that R will perfom on a data retrieved from a Mongodb. Real-life trajectories captured from GPS devices usually suffer from noise and to improve data quality, noise should be reduced. There can be various reasons for having noisy GPS trajectories, however sampling and measurement misadjustments, sensor battery outages and signal losses are some of them (Zheng *et al*., 2009). For noise reduction, there exists mean and median filters which are the most simplied forms of filters for smoothing the GPS data. For this research, median filter is used because of its robustness characteristic whereas, mean filter is not recommended because it is highly sensitive to outliers (Zheng *et al*., 2009). In a median filter, for a measured point z_i, the estimate of the unknown value is the median of z_i and its $n-1$ predecessors in time. The median filter is based on sliding window mechanism covers n temporally adjacent values of z_i as shown in below mentioned equation.

$$\hat{x}_i = median\{z_{i-n+1}, z_{i-n+2}, z_{i-n+3}, \dots, z_{i-1}, z_i\}$$

Choice of a median filter for our data cleaning is made also because of high sampling rate of our GPS data; that makes median filter a good option. Though, if a sampling rate of GPS data is too low than a median filter is not recommended and advanced filters such as Kalman filter can be considered for noise reduction. Figure 3 shows actual and filtered trajectory of a worker.

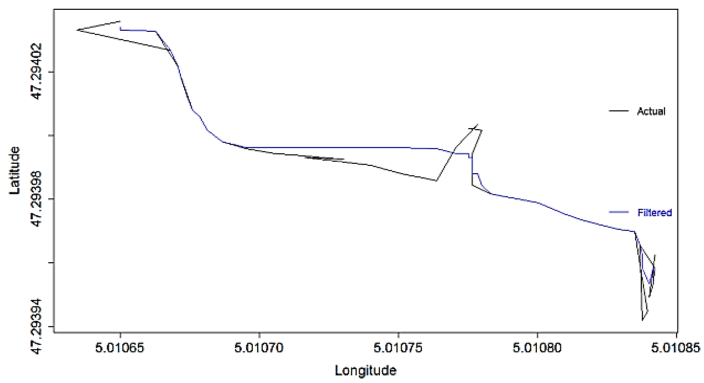

Figure 3. Actual and filtered trajectory of a worker

3.2 Stay Points Detection

After cleaning the GPS trajectories, our data processing algorithms in the R environment will calculate stay points of workers using the Algorithm 1. Stay points are the geographic location points where an individual has spent a significant time within a certain distance. The calculation of stay points depends on two parameters, a distance threshold (D_{thresh}) and a time threshold (T_{thresh}). The values of these both parameters are application dependent and set manually in an algorithm. Zheng *et al*. proposed an algorithm for finding stay points in trajectories (Zheng *et al*., 2009). A single stay point s can be treated as a virtual location point characterized by a set of successive GPS points $Z = \{z_m, z_{m+1}, z_{m+2}, \dots, z_n\}$, $\forall\, m < i \le n$, $Distance(z_m, z_i) \le D_{thresh}$ and $|z_n.T - z_m.T| \ge T_{thresh}$. Formally, conditioned by Z, D_{thresh} and T_{thresh}, a stay point $s = (Latitude, Longitude, arrivaltime, leavingtime)$. Where,

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

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

For our application, we need to find the locations on a construction site where workers have spent more time than required. Calculating such stay points will help building supervisors and H&S manager to monitor their level of mobility and safety plans can be implemented accordingly (see Figure 4).

Algorithm 1. Stay Points Detection

Input: GPS Trajectory $(G) \leftarrow \{G_{lat}, G_{lon}, G_t\}$, Distance Threshold ($D_{thresh}$), Time Threshold ($T_{thresh}$)
Output: Stay points $(S) \leftarrow \{S_{lat}, S_{arrv}, S_{leav}\}$

$i \leftarrow 0$, Total number of GPS points $(G_{num}) = |G|$
While $(i < G_{num})$
{
$j \leftarrow i + 1$
While $(j < G_{num})$
{
Distance ← distHaversine $\{(G_{lon}[j], G_{lat}[j]), (G_{lon}[i], G_{lat}[i])\}$
    If (Distance > D_{thresh})
    {
    deltaT ← $G_t[i+1] - G_t[i]$
    If (deltaT ≥ T_{thresh})
    {
    $M_{lon} \leftarrow$ mean $(G_{lon}[i], G_{lon}[i+1])$
    $M_{lat} \leftarrow$ mean $(G_{lat}[i], G_{lat}[i+1])$
    $S_{arrv} \leftarrow G_t[i]$
    $S_{leav} \leftarrow G_t[i+1]$
    }
    $i = j$; break;
}
j ← j + 1
}
}

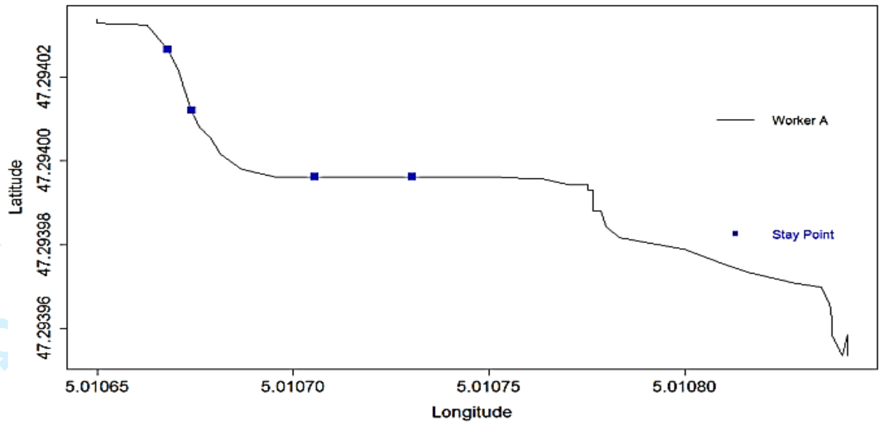

Figure 4. Stay points detection in a worker`s trajectory

For an average latitude and longitude of the collection Z, $s.arrivaltime = z_m.T$ and $s.leavingtime = z_n.T$ represent a worker`s arrival and leaving times on a stay point (s). Algorithm 1 first checks if the distance between a point under consideration and its successors in a trajectory larger than a specified threshold. Then it calculates the time interval between a point and the last successor that is within the distance threshold. If the time span is larger than a given threshold, a stay point is will be detected in a trajectory. For our application, we have set D_{thresh} to 5 meters and T_{thresh} to 20 meters for stay points detection. There are four stay points detected in a worker A trajectory as shown in Figure 4. However, number of stay points can be increased or decreased as these are totally dependent of the values of a distance threshold (D_{thresh}) and a time threshold (T_{thresh}).

3.3 Segmentation of a trajectory

For our application, we need to segment trajectories for finding types of movements which are being carried by the workers on a construction site. Segmentation is a process of dividing a trajectory into various segments to reduce the computational complexity but increasing the opportunities to mine richer knowledge (Zheng, 2015). Segmentation can be done based on: time interval, shape of a trajectory and semantics meaning of location points in a trajectory (Zheng, 2015; Potamias *et al*., 2006). As constuction sites are very dynamic in nature and its very important to monitor the type of movements a worker does on a site. Quick movements in workers trajectory data will be an alert for building supervisors and H&S manager highlighting abnormal situation occurred on a construction site.

Based on our application requirement, we have used segmentation method that is based on semantics. Zheng *et al*. proposed a walk-based segmentation method (Zheng *et al*., 2008). Their proposed algorithm calculates walk points and run points based on the point`s speed and acceleration as shown in Algorithm 2. The trajectory can then be divided into alternate walk and run segments as shown in Figure 5.

Algorithm 2. Trajectory Data Segmentation Based on Speed

Input: GPS Trajectory (G) ← {G_{lat}, G_{lon}, G_t}, Minimum Speed (S_{min}), Maximum Speed (S_{max})
Output: Run Segment (R) ← {R_{lat}, R_{lon}, R_t}, Walk Segment (W) ← {W_{lat}, W_{lon}, W_t}

1. $i \leftarrow$ 0, Total number of GPS points (G_{num}) ← $|G|$
2. While ($i < G_{num}$)
3. {
4. Distance ← distHaversine {($G_{lon}[j], G_{lat}[j]$), ($G_{lon}[i], G_{lat}[i]$)}
5. DeltaT← Append(($G_t[j] - G_t[i]$)*60)
6. Speed ← Append(Speed, Distance [i]/DeltaT[i])
7. If ((Speed [i] ≥ S_{min}) & (Speed [i] ≤ S_{max}))
8. {
9. W_{lon} ← Append($W_{lon}, G_{lon}[j]$)
10. W_{lat} ← Append($W_{lat}, G_{lat}[j]$)
11. W_t ← Append($W_t, G_t[j]$)
12. }
13. Else
14. {
15. R_{lon} ← Append($R_{lon}, G_{lon}[j]$)
16. R_{lat} ← Append($R_{lat}, G_{lat}[j]$)
17. R_t ← Append($R_t, G_t[j]$)
18. }
19. $z \leftarrow$ z + 1
20. $i \leftarrow$ j
21. $j \leftarrow$ j + 1
22. }

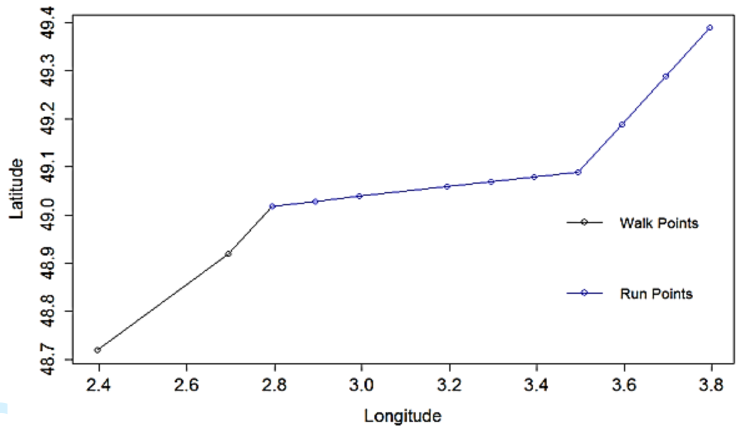

Figure 5. Segmentation of a worker`s trajectory based on speed

3.4 Intersection of two trajectories or Intersection between trajectories and a location

It is important for our application to calculate the workers movements that have been carried collectively. Such understanding will depicits that how many workers are working on a same activity and how many will be affected on the occurance of any safety hazard in that particular location. As trajectories are the traces in the form of chronologically ordered points, distance from one worker trajectory to an other worker trajectory or a distance between workers trajectories and a common location can be found using a Haversine distance formula as mentioned below.

$$d = 2r sin^{-1}\left(\sqrt{sin^2\frac{\emptyset_i - \emptyset_j}{2} + \cos\emptyset_i \cos\emptyset_j \, sin^2\frac{\varphi_i - \varphi_j}{2}}\right)$$

Where "r" is the Earth radius, $\emptyset$ and φ are the latitudes and longitudes of points "i" and "j". This formula is used to find points of intersection in trajectories to visualize the esembly points within the area of interest as shown in Figure 6 and 7.

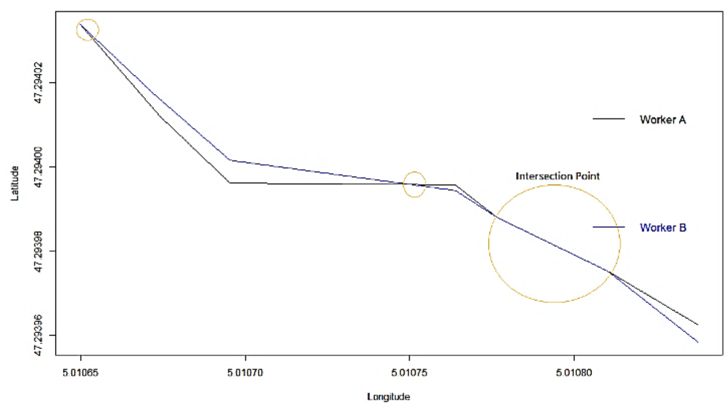

Figure 6. Points of intersection of two trajectories

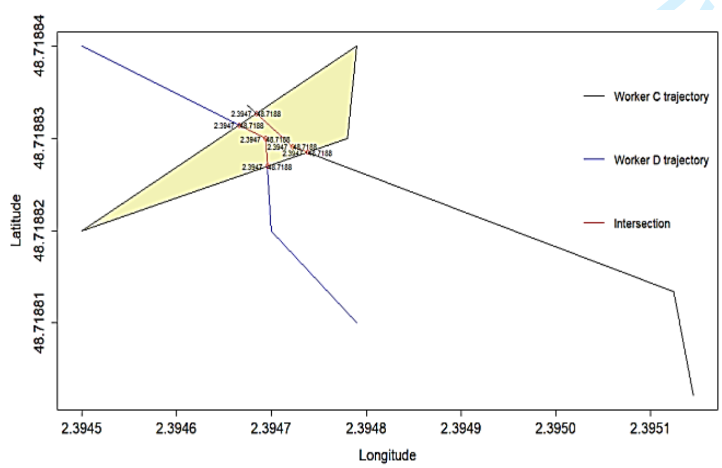

Figure 7. Points of intersection of two trajectories in a location

In addition, there can be some situations on construction sites where there is a need to find the closest point between two workers trajectories. For these scenarios, Haversine distance formula can also be used as shown in Figure 8.

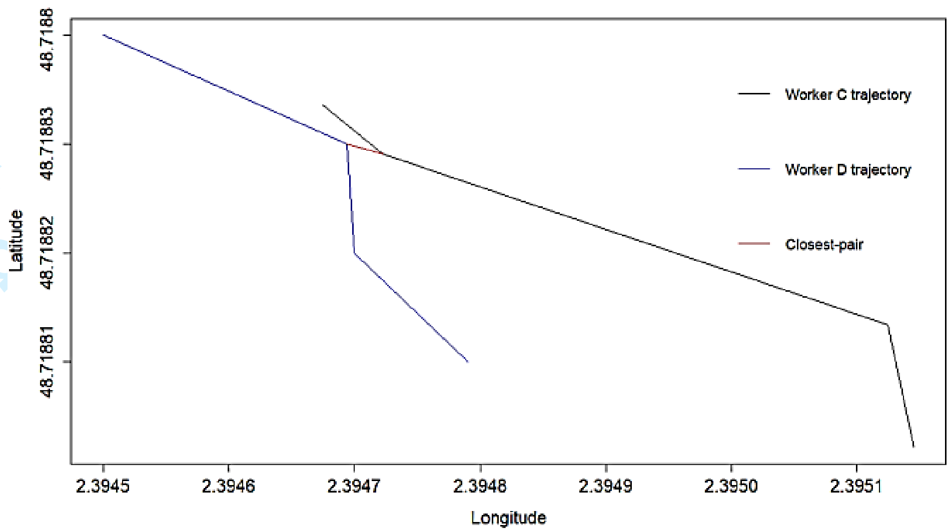

Figure 8. Calculating a closest point between two trajectories

3.5 Trajectory data reduction

Once data processing of trajectories is completed, it is now required to store these trajectories in a database for future use and analysis. Before storing, this data should be reduced in size. Potamias et. al discussed that a location point should be incorporated in a trajectory as long as it shows any change in the trajectory path (Zheng *et al*., 2008). As long as the location of an incoming GPS point can be predicted from previous points then this location point can be discarded safely as it will contribute very minute information. Batched compression and online data reduction are two methods to reduce a size of a trajectory (Zheng, 2015). Batch compression algorithms produce higher quality results when compared to online compression algorithms. For our application, we have used batched compression technique as our data is already been captured for a specific duration and its volume is kept limited for a deeper understanding. The most common form of batch compression algorithm for a data with higher sampling rate is the uniform sampling algorithm that is based on-line generalization mechanism. The main idea of line-generalization is to retain a fraction of the spatiotemporal data in the original trajectory without taking into account the redundancy of data points (Zheng, 2015; Potamias *et al*., 2006). It keeps every i^{th} data points (3^{rd}, 6^{th}, 9^{th}, etc.) and discards the remaining points. As results, in the future, reduced trajectories can be reconstructed that will be an approximation of an original trajectories with a coarser granularity. Figures 9 and 10 show the actual and reduced version of a trajectory using a uniform sampling algorithm.

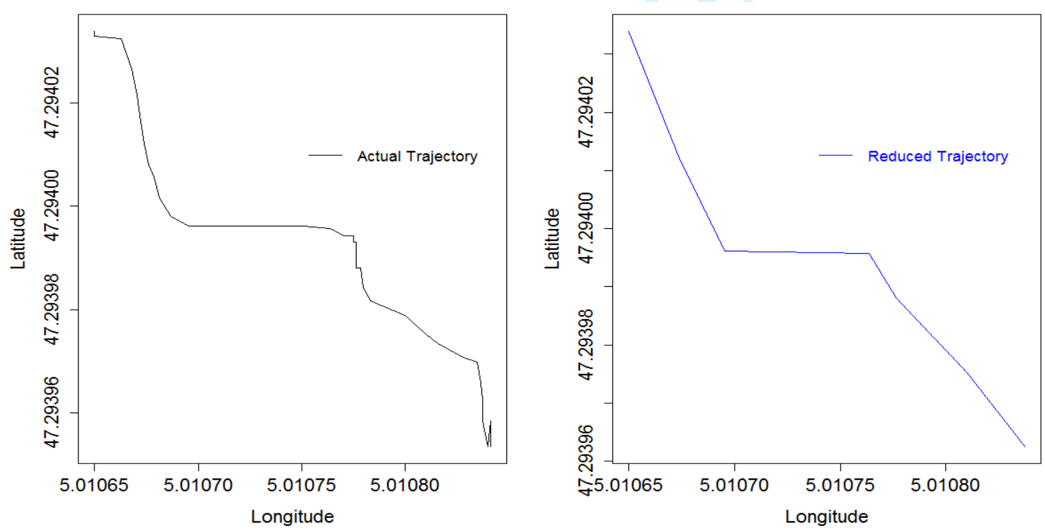

Figure 9. Actual and reduced worker trajectory

In addition to line-generalization reduction approach, changes in speed and direction can also be used to make a decision of whether a particular trajectory point should be included in a reduced trajectory or not (Zheng, 2015). For our case, threshold-guided sampling approach is used that has been proposed by Potamias *et al*. (Potamias *et al*., 2006). Safe zone is created based on the speed threshold, that is an allowable level of change in the speed of a worker. If the incoming location point resides within safe zone that is created by computing previous trajectory point`s speed and a speed threshold under consideration.Then it means, this particular point will not be contributing much information in a trajectory and hence can be discarded without compromising the accuracy of information in trajectory data. Same approach has been used to form a safe zone for a change in a direction scenario (see Figure 10). By changing the speed or a direction angle, more points will be deleted or added in a trajectory data.

Here, the purpose of the reduction is only to remove the redundant or insignificant points from the original trajectory because storing larger trajectory data is problematic as it can quickly overwhelm the available data storage capacity. However, discarding trajectory points randomly, uniformly or based on any criteria (as mentioned above) would cause information loss. Thus, for safety related critical applications, advanced trajectory reduction techniques (Muckell *et al*., 2011; Muckell *et al*., 2014) should be used that will reduce the storage requirements while maintaining an acceptable degree of error in the reduced trajectory data. In addition, this error threshold is totally dependent on the type of application in order to balance the tradeoff between storage size and accuracy of trajectories.

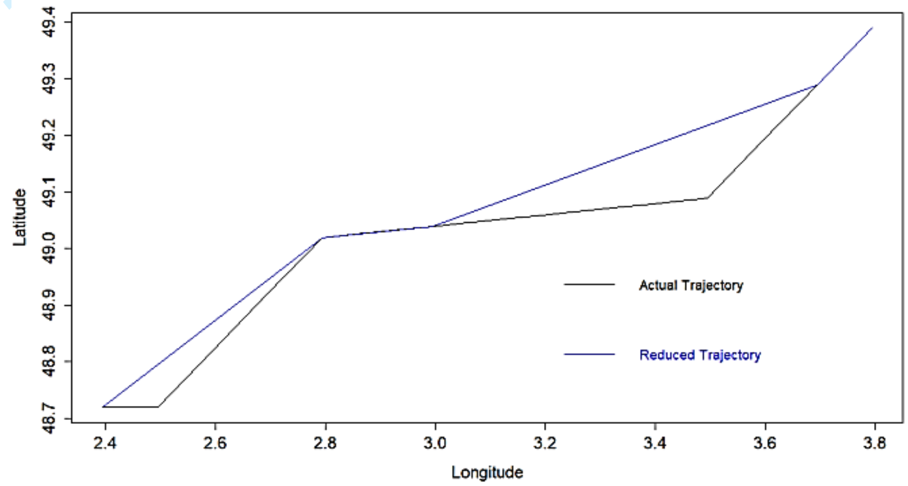

Figure 10. Reduction based on a speed 0-3.5 meter per second

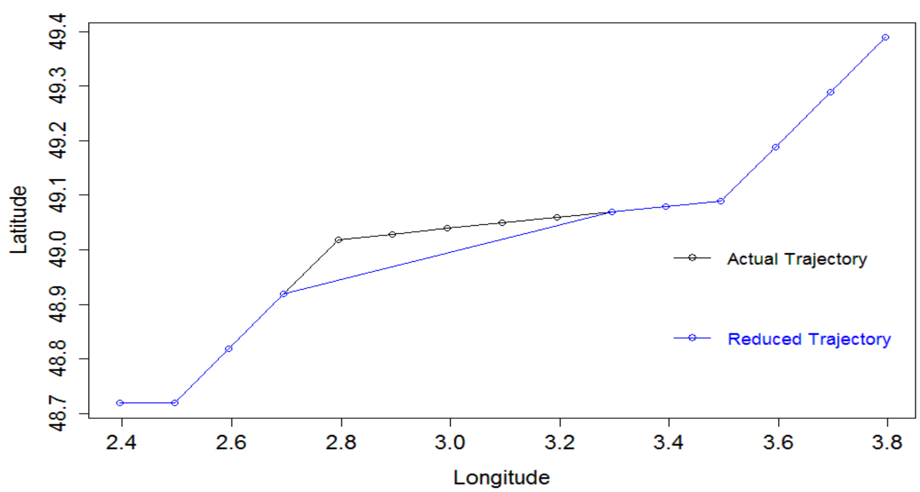

Figure 11. Reduction based on a direction angle 10°- 20°

4. DISCUSSION

The application based on ST trajectories as discussed in section 3 consists of different data processing algorithms that attempts to contribute in improving the worker safety on construction sites. The system application utlilizes GPS data from the construction site to understand mobilities of workers as well as machineries in order to analyze potential safety hazards on construction sites. The protoype system visualizations address the competency questions as shown in Table 3. This derived safety information will assist building supervisors and H&S managers in monitoring, controlling and implementing evacuation plans efficiently. Each algorithm opens a possibility to understand and explore different types of mobility more deeply that occur on the sites. This information can then be shared among various construction project stakeholders to improve existing work practices and to ensure safe construction operations, hence reducing the occurrences of fatal accidents.

However, there are some limitations of a discussed system application, are mentioned below;

Table 3. Prototype system addressing the selected competency questions

Competency questions related mobility	Derived safety information from a proposed system`s visualizations
What are the stay points of workers or machineries on a construction site?	Calculating the stay points as shown in Figure 4 will indicate the occurrence of an unexpected situation on a site if the duration of stay is greater or less than the required.
Which type of movements are done by the workers?	Segmenting the trajectories into walk and run points in a given time duration as shown in Figure 5 will indicate the occurrence of unexpected situation on a construction site.
Which and how many workers or machineries are close to other workers or machineries?	It will help to visualize that how many workers and which machineries are engaged for a certain project activity so that safety plans can be made accordingly (shown in Figure 6).
How many workers are present in a specified area?	Parallel movements of workers in a specified location as shown in Figure 7 will indicate that how many workers will be affected on the occurrence of any safety hazard in that particular location.
What is the closest point-pair distance between two workers or machineries?	Hazardous situations that occur on construction sites can be due to an excessive proximity between different workers, between a worker and an equipment or a dangerous material on site, as workers often carry out their work in a dynamic and complex manner. Calculating closed-pair distances as shown in Figure 8 will help to visualize the worker`s level of protection from the dangerous material or machinery as well as it shows the risk of injuring other workers with the equipment used in the activity.

1) The location data used for the proof of concept system application is not a real-time data. Formerly collected Comma Seperated Values (CSV) files of a GPS device is used to understand mobility by processing raw trajectories. In addtion, batch pre-processing algorithms have been implemented as discussed in section 3 to understand the trajectories occuring on a site. However, batch or offline data processing methods are only recommended when we have to visualize behavior patters from the existing data. But for the applications, where mobility data of people and objects is continuously updating and we need to find dynamic correlations between various trajectories then trajectories should be processed in real time. Real time trajectory processing solutions have the ability to process raw trajectory data within a controlled time window and generate processed trajectories in an online mode . In online processing, as soon as data is received, trajectory basic features such as speed, direction, and displacement, etc. are calculated (Zheng, 2015). Consequently, data filtering, smoothing and reduction techniques are applied before generating trajectory visualizations.

2) GPS technology is discussed to acquire workers` location data from a construction site. However, this technology is recommended only in an outdoor environment. Whereas, construction sites are composed of an outdoor as well as an indoor environment. Using GPS devices inside the building will introduce large measurement errors in trajectories because of interference effects. This impact will result in partial to total loss of signal tracking to capture workers` mobility. Therefore, the accuracy of the GPS device becomes questionable for critical aspects such as H&S management. Hence, it is important to consider to use alternative technology to collect indoor positioning data.

3) As the scope of the paper is kept limited to only preprocessing the ST trajectories. Generated visualizations are still incomplete to give a complete picture of mobility patterns of workers to building supervisors and H&S manager for decision making as contextual information and application domain knowledge is not incorporated. Other technologies such as Building Information Modeling (BIM) can be used for visualizing the trajectories as it has been described as one of the most promising developments in the area of safety management (Cruz and Nicolle, 2006; Cruz, 2017). Another possible way to improve understanding of the generated visualizations is to enrich them with semantic information acquired from external data sources both private data such as construction site information describing the work zones and public data such as OpenStreetMap (OSM) and Google Maps. For this, existing semantic enrichment techniques (Yan et al., 2011) can be explored to construct semantically enriched trajectories which are: (1) enrichment with semantic points that maps site location identification to pre-processed trajectory points; (2) enrichment with semantic lines that relies on the speed based segmentation approach (as discussed in section 3) infers modes of transportation used in trajectory`s episodes; and (3) enrichment with semantic region for mapping a complete trajectory on an actual construction site zone. Semantically enriched trajectories (see Figure 12 as an example) will provide an intuitive and readily understandable pre-processed visualizations that will help H&S managers in making improved decisions for monitoring and controlling site activities by understanding workers behaviors that ultimately attempts to contribute in reducing fatal accidents occurring on construction sites each year.

5. CONCLUSION

In this paper, we have presented the importance of preprocessing ST trajectories captured from construction sites. Once trajectory data is preprocessed correctly, it can be used in variety of applications for construction safety monitoring and management. A significant gap has been found in the literature that the usage of preprocessed ST trajectory data is still missing in workers`s safety monitoring scenarios in the area of construction management. To address this research gap, our paper has discussed a system application that use preprocessed ST trajectories to monitor workers` movemenets on a construction site. GPS data acquired from workers` handheld devices has been used and it is then processed using various algorithms aims to provide effective visualizations to building supervisors and H&S manager. However, our proposed application is in the development stage lacking contextual and semantic information in the visualizations generated. Considering the preprocessing the very first step to prepare the ST trajectories for semantic enrichment, making this processed data as a base that can leads to improved understanding of mobility of workers on construction sites. In addtion, using appropriate data and pattern mining algorithms and techniques on a processed trajectory data will help to discover interesting patterns and rules and to extract unknown and nontrival behaviors for decision-making processes.

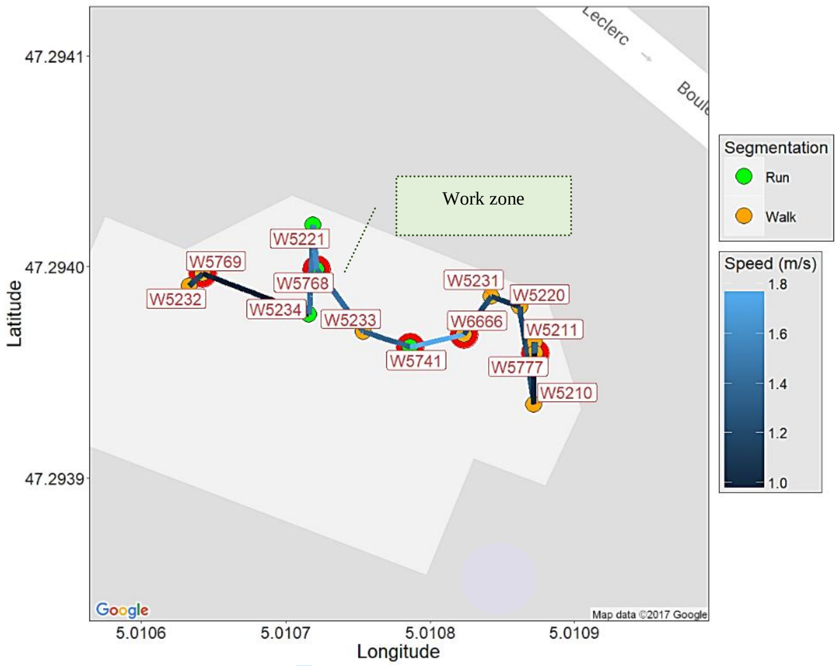

Figure 12. Visualizing a complete worker` trajectory with semantic stay points in red colored circles (site identifications for e.g. W5768 for a workzone), semantic lines (segmentation into run and walk points) and a semantic region

6. ACKNOWLEDGEMENT

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

7. REFERENCES

Adey, P. (2017), *Mobility*. Taylor & Francis.

Arslan, M., Cruz, C., Roxin, AM. and Ginhac, D. (2017), "Using spatio-temporal trajectories to monitor construction sites for safety management", in *Proceedings of the 9th International Conference on Information Management and Engineering (ICIME 2017)*, pp. 1-5.

Arslan, M., Riaz, Z., Kiani, A. and Azhar, S. (2014), "Real-time environmental monitoring, visualization and notification system for construction H&S management", *Journal of Information Technology in Construction*, 19, pp.72-91.

Andrienko, G., Andrienko, N., Bak, P., Keim, D. and Wrobel, S. (2013*), Visual analytics of movement*. Springer-Verlag Berlin Heidelberg.

Bogorny, V., Renso, C., Aquino, A. R., Lucca Siqueira, F. and Alvares, L. O. (2014), "Constant–a conceptual data model for semantic trajectories of moving objects", *Transactions in GIS*, 18(1), pp. 66-88.

Brisaboa, N. R., Luaces, M. R., Pérez, C. M. and Places, Á. S. (2017), "Semantic trajectories in mobile workforce management applications", in *Proceedings of the International Symposium on Web and Wireless Geographical Information Systems*, Shanghai, China, pp.100-115.

Buchin, M., Driemel, A., Kreveld, M. V. and Sacristan, V. (2010), "An algorithmic framework for segmenting trajectories based on spatio-temporal criteria", in *Proceedings of the 18th Annual ACM International Symposium on Advances in Geographic Information Systems (GIS'10)*, pp.202–211.

Buchin, M., Driemel, A., Van Kreveld, M. and Sacristán, V. (2011), "Segmenting trajectories: A framework and algorithms using spatiotemporal criteria", *Journal of Spatial Information Science*, 3, pp. 33-63.

Chakri, S., Raghay, S. and El Hadaj, S. (2017), "A Semantic Interpretation of Unusual Behaviors Extracted from Outliers of Moving Objects Trajectories", *International journal of advanced computer sciences and applications*, 8(3), pp. 119-127.

Chen, C., Su, H., Huang, Q., Zhang, L. and Guibas, L. (2013), "Pathlet learning for compressing and planning trajectories", in *Proceedings of the 21st ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems*, pp. 392-395.

Cruz, C. (2017), "Semantic Trajectory Modeling for Dynamic Built Environments", in *Proceeding of The 4th IEEE International Conference on Data Science and Advanced Analytics*.

Cruz, C. and Nicolle, C. (2006), “Active3D: Vector of collaboration, between sharing and data exchange”, *INFOCOMP Journal of Computer Science*, pp.1-8.

Hsiao, H. (2016), *Fall Prevention and Protection: Principles, Guidelines, and Practices*, CRC Press.

Hu, Y., Janowicz, K., Carral, D., Scheider, S., Kuhn, W., Berg-Cross, G., Hitzler, P., Dean, M. and Kolas, D. (2013), “A geo-ontology design pattern for semantic trajectories”, in *International Conference on Spatial Information Theory*, pp. 438-456.

Guo, h., Yu, Y. and Skitmore, M. (2017), “Visualization technology-based construction safety management: A review”, *Automation in Construction*, 73, pp.135-144.

Jin, J., Gubbi, J., Marusic, S. and Palaniswami, M. (2014), “An Information Framework for Creating a Smart City Through Internet of Things”, *IEEE Internet of Things Journal*, 1(2), pp.112-121.

Jun, J., Guensler, R., and Ogle, J. (2006), “Smoothing methods to minimize impact of global positioning system random error on travel distance, speed, and acceleration profile estimates”, *Journal of the Transportation Research Board*, 1, pp.141–150.

Krisnadhi, A.A., Hitzler, P. and Janowicz, K. (2016), “A spatiotemporal extent pattern based on semantic trajectories”, in *Proceedings of the Workshop on Ontology and Semantic Web Patterns*, Kobe, Japan, pp.1-6.

Li, Y., Su, H., Demiryurek, U., Zheng, B., He, T. and Shahabi, C. (2017), “PaRE: A System for Personalized Route Guidance”, in *Proceedings of the 26th International Conference on World Wide Web*, pp. 637-646.

Lv, M., Chen, L. and Chen, G. (2012), “Discovering personally semantic places from GPS trajectories” in *Proceedings of the 21st ACM international conference on Information and knowledge management*, pp. 1552-1556.

Marketos, G., Frentzos, E., Ntoutsi, I., Pelekis, N., Raffaetà, A. and Theodoridis, Y. (2008), “Building real world trajectory warehouses”, in *Proceedings of the 7th ACM International Workshop on Data Engineering for Wireless and Mobile Access (MobiDE’08)*, pp. 8–15.

Meratnia, N. and Deby, R. A. (2004), “Spatiotemporal compression techniques for moving point objects”, in *Proceedings of the 9th International Conference on Extending Database Technology (EDBT’04)*, pp.765–782.

Miller, H. and Han, J. (2009), *Geographic data mining and knowledge discovery*, Boca Ratón, CA: CRC Press.

Mousavi, A. (2016), Ontology Based Semantic Knowledge Discovery for Movement Behaviors, Doctoral dissertation, University of Calgary.

Muckell, J., Hwang, J.H., Patil, V., Lawson, C.T., Ping, F. and Ravi, S.S. (2011), “SQUISH: an online approach for GPS trajectory compression”, in *Proceedings of the 2nd International Conference on Computing for Geospatial Research & Applications*, pp. 13.

Muckell, J., Olsen, P.W., Hwang, J.H., Lawson, C.T. and Ravi, S.S. (2014), “Compression of trajectory data: a comprehensive evaluation and new approach”. *GeoInformatica*, 18(3), pp.435-460.

Nardini, F.M., Orlando, S., Perego, R., Raffaetà, A., Renso, C. and Silvestri, C. (2018), “Analysing trajectories of mobile users: From data warehouses to recommender systems”, in *A Comprehensive Guide Through the Italian Database Research Over the Last 25 Years*, Springer International Publishing, pp. 407-421.

Neirotti, P., De Marco, A., Cagliano, A., Mangano, G. and Scorrano, F. (2014), “Current trends in Smart City initiatives: Some stylised facts”, *Cities*, 38, pp.25-36.

Noël, D., Villanova-Oliver, M., Gensel, J. and Le Quéau, P. (2015), “Modeling semantic trajectories including multiple viewpoints and explanatory factors: application to life trajectories”, in *Proceedings of the 1st International ACM SIGSPATIAL Workshop on Smart Cities and Urban Analytics*, pp. 107-113.

Patroumpas, K. and Sellis, T. (2017), “Managing Trajectories of Moving Objects as Data Streams”, in *Proceedings of the Second Workshop on SpatioTemporal Database Management*, Toronto, Canada pp.41-48.

Patroumpas, K., Alevizos, E., Artikis, A., Vodas, M., Pelekis, N. and Theodoridis, Y. (2016), “Online event recognition from moving vessel trajectories”, *GeoInformatica*, 21(2), pp.389-427.

Potamias, M., Patroumpas, K. and Sellis, T. (2006), “Sampling Trajectory Streams with spatiotemporal criteria”, in *Proceedings of the 18th International Conference on Scientific and Statistical Database Management*, Vienna, Austria, pp. 275–284.

Richter, K. F., Schmid, F. and Laube, P. (2012), “Semantic trajectory compression: Representing urban movement in a nutshell”, *Journal of Spatial Information Science*, 4, pp.3-30.

Šaltenis, S., Jensen, C.S., Leutenegger, S.T. and Lopez, M.A. (2000), “Indexing the positions of continuously moving objects”, in *Proceedings of the 2000 ACM SIGMOD International Conference on Management of Data*, Texas, USA, pp. 331-342.

Schmid, F., Richter, K. F. and Laube, P. (2009), “Semantic trajectory compression”, *Advances in Spatial and Temporal Databases*, pp. 411-416.

Shen, J. and Cheng, T. (2017), “Semantic enrichment of interesting regions with POI data”, in *Proceedings of the GISRUK conference, Manchester*, pp. 1-5.

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).

Urbano, F. and Cagnacci, F. (2014). “Spatial Database for GPS Wildlife Tracking Data: A Practical Guide to Creating a Data Management System with PostgreSQL/PostGIS and R”, *Springer Science & Business Media*.

Vandecasteele, A., Devillers, R. and Napoli, A. (2014), “From movement data to objects behavior using semantic trajectory and semantic events”, *Marine Geodesy*, 37(2), pp.126-144.

Wagner, R., de Macedo, J. A. F., Raffaetà, A., Renso, C., Roncato, A. and Trasarti, R. (2013), “Mob-warehouse: A semantic approach for mobility analysis with a trajectory data warehouse”, in *International Conference on Conceptual Modeling*, pp. 127-136.

Yan, Z., Chakraborty, D., Parent, C., Spaccapietra, S. and Aberer, K. (2013), “Semantic trajectories: Mobility data computation and annotation”, *ACM Transactions on Intelligent Systems and Technology (TIST)*, 4(3), pp. 1-38.

Yan, Z., Chakraborty, D., Parent, C., Spaccapietra, S. and Aberer, K. (2011), “SeMiTri: a framework for semantic annotation of heterogeneous trajectories”, in *Proceedings of the 14th international conference on extending database technology*, pp. 259-270.

Zhang, J. (2014), “Efficient frequent sequence mining on taxi trip records using road network shortcuts”, *Big Data: Techniques and Technologies in Geoinformatics*, HA Karimi, Ed, pp.193-206.

Zheng, W., Huang, X. and Li, Y. (2017), “Understanding the tourist mobility using GPS: Where is the next place?”, *Tourism Management*, 59, pp. 267-280.

Zheng, Y. (2015), “Trajectory data mining: an overview”, *ACM Transactions on Intelligent Systems and Technology (TIST)*, 6(3), pp. 1-44.

Zheng, Y., Liu, L., Wang, L. and Xie, X. (2008), “Learning transportation mode from raw GPS data for geographic application on the Web”, in *Proceedings of the 17th International Conference on World Wide Web*, Beijing, China, pp. 247–256.

Zheng, Y., Zhang, L., Xie, X. and Ma, W.Y. (2009), “Mining interesting locations and travel sequences from GPS trajectories”, in *Proceedings of the 18th International Conference on World Wide Web*, Madrid, Spain, pp. 791-800.

Reviewer 01	Comments	Status	Comments
1	page2/ line 37,38, 39: The section's numbers describing the organization do not correspond to the article's section numbering.	Done	
2	page 4/ line 43: where it says Table 1 for the ST applications, it may refer to Table 2.	Done	
3	Page 11/ line 47: "..as discussed in section 5 .." refer to section 3 instead.	Done	
Reviewer 02			
1	Kindly elaborate upon how and in which form the acquired data shall be communicated in an intuitive and readily understandable fashion (you could also provide some examples of what could be a speculative scenario) to safety managers on site.	Done	Page 12 (Figure 12 is added and lines 40-47 are added)
2	Some clarity pertaining to on how worker based ST Trajectory vs various typologies of machinery (static vs dynamic) co-relate and if this can be extended towards moving vehicles and thus the activities which surround loading, unloading of goods etc (which would dynamically change typical ST trajectory maps in time), should be explained a bit more clearly.	Done	Page 12 (Lines 26 -30 are added)
3	Actual vs Reduced trajectory section displays quite a change in the values of longitudes and latitudes....kindly specify how much does this impact safety related decisions pertaining to spatial location.	Done	Page 11 (Lines 6-11 are added)