Semantic Trajectory Insights for Worker Safety in Dynamic Environments

Muhammad Arslan
Univ. Bourgogne Franche-Comte, Batiment I3M rue Sully, Dijon, France
muhammad.arslan@u-bourgogne.fr

Christophe Cruz
Univ. Bourgogne Franche-Comte, Batiment I3M rue Sully, Dijon, France
christophe.cruz@ubfc.fr

Dominique Ginhac
Univ. Bourgogne Franche-Comte, Batiment I3M rue Sully, Dijon, France
dominique.ginhac@ubfc.fr

***Abstract*— Existing literature reveals that unsafe worker movement behaviors are one of the major reasons of construction site fatalities resulting in serious collisions with site objects and machinery. For capturing worker movements in dynamic construction environments which involves moving and changing objects, a solution based on semantic trajectories and Hidden Markov Model (HMM) is presented in the form of four subsystems. First, a real-time data collection and trajectory pre-processing subsystem is constructed for extracting multifaceted trajectory characteristics and stay locations of the workers using spatio-temporal data that will help in recognizing the important regions in the building for categorizing the worker movements. Second, to enable the desired semantic insights for better understanding the underlying meaningful worker movements using the contextual data related to the building environment, an ontology-based STriDE (Semantic Trajectories in Dynamic Environments) model is applied which has an ability to track information about the evolution of moving and changing building objects, and outputs semantic trajectories. The third subsystem uses the Hidden Markov Model (HMM) which is the most preferred probabilistic approach in the literature for describing the object behavior in time. An entire set of trajectories belonging to a stay location (a semantic region) is analyzed by categorizing the worker movements into four states using the HMMs along with the Viterbi algorithm. In the end, the output of the Viterbi algorithm is visualized using a BIM model for identifying the most probable high-risk locations involving sharp worker movements and rotations. The developed system will help safety managers in monitoring and controlling building activities remotely in dynamic environments by better understanding the worker behaviors for an improved safety management in day-to-day building operations as well as by preventing other workforce from accessing such hazardous locations which involve risky movements.**

Keywords—Semantic trajectories; Building Information Modeling (BIM), stay locations; GPS; mobility; Health and Safety (H&S); Hidden Markov Model (HMM)

I. INTRODUCTION

The construction industry is one of the largest industries and the major contributor to the economy [1]. Unfortunately, this industry is also known for the least safe sector as compared to the other work sectors because construction workers are frequently exposed to harsh environments involving high safety risks [2]. The safety risks arise from the uncertainty in the worker movements on construction sites as their activities are likely to deviate from a predefined planning due to the dynamic environments [3, 4]. These uncertain worker movements have a tendency to creating safety risks that can result in hazardous situations [5]. Such uncertain and dynamic working environments consequence in high occurrences of serious injuries or even deaths [6, 7]. International statistics logged by the Bureau of Labor Statistics (BLS) and Health and Safety Executive (HSE) show that the majority of fatal accidents occurred in the construction industry [8, 9]. According to the BLS, in 2016 out of 5,190 fatal occupational injuries, 937 were recoded from the U.S. construction industry. Whereas, the highest number, 30 fatal deaths representing almost 21% of total fatal accidents were recorded in 2017 by the HSE from the U.K. construction industry. Regardless of numerous efforts and more attention being paid to safety management practices in recent years [10, 11], rate of fatalities in the construction industry continues to be high [12]. The above statistics show that deployed site monitoring systems for worker safety are not sufficient to avoid fatalities. A closer look to the recent research reveals that one of the major reasons of construction site fatalities is because of unsafe worker behaviors resulting in serious collisions with site objects and machinery [8, 12]. For example, limited spatial awareness of the operating equipment involving sharp movements and rotations within the workers proximity due to blind spots and surrounding noise can lead to hazardous situations on sites [5, 13].

The most effective way of preventing such situations is to monitor the buildings and the construction sites in real-time to identify unsafe movements of the workers [4]. Existing literature [14] shows that the movements of moving objects can be monitored if step lengths and turning angles are periodically calculated from their trajectories. However, before extracting different users` movements, we need to identify specific regions in a building or on a construction site where the users are staying for a longer duration. These stay locations [15] which are termed as 'semantic regions' in our study are more critical to be monitored for extracting trajectory movement insights than moving locations as the workers are spending most of their time there. To achieve this, a scenario is initially constructed (see Fig. 2). The reason for using scenario-based methodology is because this method is built on the qualitative causal thinking that facilitates to communicate [16] and brings participants (application developers, building supervisors, H&S managers, etc.) together towards a shared understanding of the situation i.e. monitoring the buildings for unsafe behaviors of the construction resources [17]. In the generated scenario (see Fig. 2), capabilities of location data acquisition technology, semantic enrichment techniques and statistical methods are mapped to the information needs for real-time monitoring of the buildings and the construction sites. This scenario was followed by a prototype system development as a proof of concept.

The recent technological developments in location data acquisition systems based on Radio Frequency Identification (RFID), location estimation using 802.11, GSM and Bluetooth beacons, Global Positioning System (GPS) and Indoor Positioning System (IPS) have made very convenient to monitor sites for unsafe movement behaviors [18]. The data collected from a typical positioning system consists of spatio-temporal points having location information with timestamps [15]. Based on the application requirements, these raw points are transformed into finite meaningful episodes called trajectories [15]. In the prototype system, spatio-temporal points of the users are collected from the Bluetooth beacons [19] placed on different locations in a building. Then, pre-processing tasks are performed on the spatio-temporal trajectories and stay regions of the users are identified inside the building. Each stay region contains a set of trajectory segments of the users. While the information in such trajectory segments is only suitable for understanding movement dynamics of workers, it does not provide the contextual semantic information related to the building environment which is required to understand meaning behind each movement [20, 21]. For enriching trajectories with external data sources both openly available and private data related to construction sites, process of semantic enrichment is required [20]. Many approaches exist in the literature for semantic enrichment of trajectories. After the review [22], we have applied semantic enrichment process [20] using the STriDE model [23]. The STriDE model is based on ontologies and has an ability to generate semantic trajectories of the users inside the building in the context of a dynamic environment where the building objects are changing their shape, size, and attributes with time. The trajectory segments are semantically enriched in terms of their corresponding Points of Interest (POIs) and Region of Interest (ROIs) locations` information using the STriDE model.

The main research objective of this study is not only to construct semantic trajectories for an indoor dynamic environment but also to gain trajectory insights by recognizing the building areas where the unsafe worker movements are occurring that can lead to accidents. As for recognizing and categorizing the movements, many case studies are present in the literature based on machine learning algorithms using Bayesian dynamic models [24] and clustering techniques [25], state-based models such as simple Markov chains [3] and HMMs [4, 26] and patterns matching algorithms [27]. Among them, statistical HMMs have been applied widely in many works and proved to be the most appropriate choice for categorizing movements and extracting patterns [26]. However, before feeding a set of trajectories belonging to a semantic region into HMMs for categorizing the users` movements, it is required to define the hidden states to categorize different movements [28]. The variables ‘step length’ and ‘turning angle’ are taken from the research of Ilkovičová et al [14] to define the HMM states for categorizing movements. After defining the states and training the HMMs, Viterbi algorithm [28] is used to identify unsafe movements having long steps or many turnings. Later, the general fit goodness of the HMMs after using semantic trajectories [29] is analyzed using the pseudo-residual [30] method. For visualizing the categorized users` movements on a building model, the input trajectory streams of the Viterbi algorithm are enriched with their corresponding POIs` semantic information which is stored in the STriDE model for displaying the output of the Viterbi algorithm on a BIM model for identifying unsafe areas within a semantic region. BIM approach is chosen for generating visualizations because literature identifies it as a ‘future IT solution’ and preferred over traditional 3D CAD approaches as it is an efficient way of information management during the building lifecycle for safety analysis [31]. Above all, the BIM approach is becoming a construction industry standard in many countries [32]. Generated BIM-based visualization depicting different types of user movements can be used by the H&S managers that can lead to improved safety management intervention strategies by analyzing semantic trajectories and identifying building locations involving high-risk workers movements to prevent accidents. To understand the proposed insight process of trajectory preprocessing, semantic enrichment and movements’ categorization using HMMs, a brief survey is presented in the next section.

Our contributions are: (i) Real-time data collection and trajectories pre-processing subsystem: As trajectory data holds multifaceted characteristics include: time (i.e., position of the object on the timescale), position of the object in geographical coordinate system, direction of the object, speed of the object, change in the direction, acceleration (i.e., change in speed) and distance traveled [15, 20]. A system is developed to extract such characteristics to better understand the trajectories. After extracting the basic trajectories` characteristics, identification of the stay locations is achieved from workers` trajectories that will help in recognizing important regions in the building for categorizing movements. (ii) Semantic enrichment subsystem: To enrich trajectories with semantic information, a data model named STriDE is used. Our system is designed that includes application domain knowledge and geographic database information for transforming pre-processed trajectories into semantically enriched trajectories. (iii) Worker movements categorizing subsystem: After constructing the semantic trajectories, an entire set of trajectories belonging to a stay location (a semantic region) is further analyzed by categorizing the worker movements into four states using the probabilistic HMMs along with the Viterbi algorithm. (iv) BIM-based visualization subsystem: In the end, the output of the Viterbi algorithm is visualized using a BIM model for identifying the most probable high-risk locations involving sharp worker movements and rotations.

The paper is organized as follows. Section 2 introduces the background literature that is reviewed. First, it provides an overview of trajectories` pre-processing techniques. Second, it describes various state-of-the-art modeling approaches to construct semantic trajectories. Third, the reasons behind using the STriDE model are explained. Fourth, HMMs are defined along with their different application scenarios. Lastly, an introduction to BIM approach is presented. Section 3 presents the proposed solution to categorize worker movements from semantic trajectories using HMMs. Section 4 explains the selection of the HMM parameters and model checking. Limitations of the presented work, a conclusion and some future works are described in the section 5.

II. Background

A. Trajectory pre-processing

The literature review was initially performed to methodically collect information for identifying and understanding the problem domain, which is to identify unsafe movements [33] of construction workers from their semantically enriched trajectories after finding their stay locations. Nevertheless, trajectory pre-processing is the first step towards constructing semantically enriched trajectories [34]. The basic idea of pre-processing trajectories is to reduce the storage, processing and communication overheads without compromising the precision of a trajectory data [15]. Generally, there exist two modes to pre-process a trajectory; (a) online mode, where trajectory pre-processing algorithms are executed in real time, and (b) offline mode, where all the processes on trajectories are done in an offline mode [15]. The basic tasks of pre-processing include noise filtering, trajectory reduction, and stay point detection [15]. The objective of noise filtering is to remove noise from trajectories caused by weak signals of location acquisition systems. Trajectory reduction processes are used to reduce the size of the trajectory to minimize the computation overhead while keeping the usefulness of the trajectory [15]. Stay point detection techniques identify the location points where the moving object has spent some time by staying over there within a specified distance. A stay point location can be inside the building or outside such as a shopping mall, a restaurant or an office [21].

B. Existing data models for semantic trajectories

After pre-processing the trajectories, a data model is required to store the trajectories for the desired analysis. In the literature [34], there exists four types of trajectory data modeling approaches which are; data type-based modeling, design pattern-based modeling, ontology-based modeling and hybrid-based modeling. Frihida et al. [35] presented a model to represent trajectories as an Abstract Data Type (ADT). The model integrates spatial, temporal and thematic dimensions for representing and manipulating the trajectory data. The major limitation of abstract data type-based approach is the dependency of the trajectory data type on the application as model represents trajectories as a series of connected trips and activities. Data type-based models alone are not enough for constructing trajectory applications because it imposes the use of a generic data type to represent trajectories for all the applications [34]. To address this drawback, Parent et al. designed pattern-based model [36], that is based on Model Analysis and Decision Support (MADS). It supports spatial and temporal objects and relationships by providing a method for describing the spatial extent and lifespan of the trajectory. It represents trajectories as a series of 'stop' and 'move' segments having 'begin' and 'end' timestamps [36]. However, to define trajectory components in 'stop' and 'move' segments, there is a need to manually input the contextual information in the model according to the application [34]. Ontology is the conceptualization of a specific domain for showing relationships between concepts as a hierarchy and offers a multilayered model to represent trajectories [34]. An example of trajectory data modeling based on ontologies is presented by Yan and Spaccapietra [20] and Noël et al. [37]. If ontology-based modeling is compared with formally discussed two modeling approaches, it represents richer semantic information by integrating different types of information enrichment processes [34, 38]. Existing literature also presents solutions based on hybrid modeling approaches to combine the best features of different models discussed above for constructing trajectories [34]. An example of such model is discussed by Yan et al. [20] offering three different levels of data abstraction (raw trajectories, trajectory episodes in 'stop' and 'move' segments and semantically enriched trajectories) by encapsulating geometry and semantics of trajectories together.

C. Reasons behind using the STriDE model

After doing the literature review of present data modeling approaches for constructing semantic trajectories, it was concluded that ontology-based models store trajectories with richer semantic information by providing flexible and self-contained annotations using heterogeneous data sources comprising geographic information and application domain knowledge as compared to the other approaches [20]. Conversely, for generating the semantic trajectories, the existing ontology-based models do not track the semantic information generated due to the change in the attributes (spatial, temporal and alphanumeric properties) of the building objects. As new walls and infrastructure supports are often added on construction sites, while others are detached. This opens more challenges to keep track of the changes in the attributes of the locations which involve with time for representing semantic trajectories. For example, a storage zone in the building is now a work zone having a different functionality. Another example would be, due to the placement of the hazardous construction material, the dumping zone on a site became a restricted zone. Moreover, the dimensions of the dumping zone are changed because of the construction of a new wall on a site. Alike situations occur very often on the construction sites. Such changes in the purpose or the position of the locations in the buildings or on construction sites will result in different movement patterns of the users. The updated semantic information about the locations along with the previous information is required to capture for studying the movement behaviors of the building objects including the users in detail with respect to the changes occurred in the building environment. Resulted user behaviors from the analysis can be used for different construction and built environment applications such as; construction resource monitoring for improved safety, managing building spaces based on their utilization, implementing worker access control system, etc. In order to address these requirements of the dynamic building environments which involve moving and changing objects, we have used our ontology graph-based STriDE model [23]. The STriDE model is derived from the Continuum model [39-43] and has the ability to store data of dynamic objects. To track changes in shape, size and attributes of dynamic objects (a user, a trajectory and a room), the STriDE model uses the concept of timeslices as shown in Fig. 1. A timeslice consists of four components: an identity, properties having alphanumeric values (semantic component), a time component indicating the validity of the timeslice and a spatial component depicting the geographical representation of the entity [23].

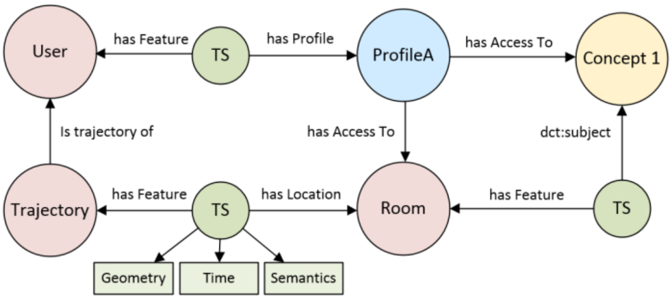

Fig. 1. STriDE data model representing dynamic entities (a user, a trajectory and a room) each with a timeslice (TS) and concepts are used for user profiling

In case of any change in the components of the timeslice, excluding the identity, a new timeslice is created inheriting the components of the last known state of the entity and then those changes are applied. This newly created timeslice is then linked with the previous one to keep the evolution of the entity during its lifecycle. The STriDE model involves mapping trajectory episodes to the meaningful information such as; the mapping of places of interest that can be in the form of Points of Interest (POIs) and the Regions of Interest (ROIs). This is achieved by annotating a trajectory with a meaningful geographic region that requires computation of topological correlations [44, 45]. These correlations are made with third party data sources containing spatial regions having semantic places [44]. For this, spatial data from an OpenStreetMap (OSM) file is extracted and stored in the STriDE model to make spatial joins i.e. $Trajectory\_epidode \bowtie_{\theta} Region$ for achieving annotations. Here, the parameter θ is very crucial for computation [34]. Calculation of the parameter θ can be based on a single or the combination of multiple topological spatial relations such as distance, displacement, etc. Once θ is computed, spatial joins are performed either with the boundary of the trajectory or with its center [34]. In this way, a region with its associated semantic metadata is annotated with the trajectory`s episodes. This process is used for findings the ROIs in the trajectory data. In the same way, a list of POIs exist inside the building which are the 'rooms' in our case are also defined in the STriDE model. After the semantic enrichment process, we categorized the worker movements using the HMMs within a semantic region that is identified as a stay location of the users.

D. Hidden Markov Models (HMMs)

Existing literature shows that a straightforward solution to describe the object behavior in time is the Markov chain (MC) model [46]. The MC is a state space model in which the probability of having the next state depends on the present state only [46]. HMMs are the direct generalization of MCs which assumes that we cannot directly observe the exact states of the system and only a stochastic function of these states is visible to us [28]. Generally, the HMM has three main properties that defines it [28]. The first, it assumes that the observation was generated by a process at time t whose state S_t is hidden from the observer. The second, it assumes that the state of this hidden process fulfills the Markov property. That is, the current state S_t of the process is only dependent on the only previous state S_{t-1} and independent of all the states prior to $t-1$. It means that to predict the future of the process, the hidden state encapsulates all the information we need to know from the process history [28]. The third supposition of the HMM is that the hidden state variable can only take T integer values $\{1, 2, 3, \dots, T\}$. In general, an HMM λ is described by a set of three parameters [28] which can be written as the 3-tuple $\lambda = (A, B, \pi)$. A is the transition probability matrix, B is the emission probability matrix and π is the vector of the initial state probabilities.

HMMs are used extensively in real-world applications to study the behavior patterns of moving objects as shown in Table 2. Numerous studies have been conducted based on the smart cameras, wearable sensors, and GPS devices to collect location data to track people, and other moving objects to investigate their different movement behaviors for better decision making. Understanding the building occupancy patterns, abnormal behavior identification and detection of dangerous activities of people, behavioral surveillance system for elderly care-and safety risk assessment on construction sites are the major HMM applications using the location data as shown in Table 1. For outdoor application scenarios, Global Navigation Satellite System (GNSS) technology and static cameras are widely used for locating people, where potential users were asked to enable GPS functionality of their handheld devices [2]. For developing indoor systems for behavioral analysis, wireless networks such as wearable sensors for meal-time and duration monitoring, accelerometer and gyroscope sensors for recognizing different human activities, CO2, humidity, temperature, motion and smart electric metering sensors are used for inferring building occupancy. The use of GPS technology is low cost as the necessary hardware is already present in most mobile devices. The same case with the deployment of wireless sensors as wireless communication infrastructure is already exists in majority of buildings. However, a high cost will be associated in mounting sensors in the building environment that is directly dependent on the size of the coverage area and type of functionalities required from the sensing devices. In addition, the deployment of cameras also requires extra hardware as well as advanced image processing techniques for higher accuracies in the data acquisition. From the literature review provided above, it is noted that communication technology for capturing user data should be carefully chosen before feeding it to the HMM for categorizing the behaviors. A selection must be made as to what is more crucial for designing a sensor network for a particular behavioral application that should be low cost, high precision and having low to moderate complexity in deployment.

In all the applications discussed above, there have been three fundamental scenarios for which HMMs are used [28]. These scenarios are the following;

TABLE. 1 NON-EXHAUSTIVE HMM APPLICATIONS FOR BEHAVIOR EXTRACTION LIST BY YEAR

No.	Use case	Year	Building indoor/outdoor	Key technology	State variables	Dataset
1	**[47] Abnormal behavior detection of crowd**	2018	Outdoor	CCTV cameras	Normal and abnormal behavior states	UMN and PETS datasets
2	**[48] Observing people's negative behaviors**	2018	-	Wi-Fi and Bluetooth devices	Emotional states	Proximity, location and call logs of 83 people
3	**[49] Estimating the feeding time of dairy cows**	2018	Outdoor	Ubisense positioning system	Feeding or not feeding states	50 dairy cows for 7 days
4	**[50] Meal-time and duration monitoring**	2017	Indoor	Wearable sensor system	Breathing, swallowing and talking states	34 hours 30 minutes of data of 14 people
5	**[51] Monitoring of potentially dangerous activities**	2017	Indoor	Kinect 2.0 sensor	Left handed, right handed and both handed actions	450 activity sequences
6	**[52] Recognizing human activities**	2017	Indoor	Accelerometer and gyroscope sensors	Walking, upstairs, downstairs, sitting, laying and standing	30 subjects actions at 50Hz frequency
7	**[53] Behavior classification of pasture-based cattle**	2017	Outdoor	GPS device	Activities (grazing, resting or walking)	150 hours of cows data sampled every 5 seconds
8	**[54] Household thermal profiling**	2017	Indoor	HVAC systems	Thermal load states	12 months data of thermal behavior
9	**[2] Risk behavior-based trajectory prediction**	2017	Outdoor	Smartphone GPS	Immediate future positions of a worker in the vicinity of site hazard.	26 random walks and each walk is of 15 min duration.
10	**[55] Animal behavior analysis**	2016	Outdoor	Static cameras	Unhealthy or poor health states	-
11	**[56] Occupancy detection and prediction**	2016	Indoor	Environmental monitoring sensors and energy meters	Number of occupants	Energy and environment data
12	**[57] Automatic meal intake montoring system**	2016	Indoor	MS kinect sensor	Types of meals (liquids and main meals)	447 valid human eating gestures
13	**[58]Time series data analysis to estimate home occupancy**	2016	Indoor	CO2, humidity and temperature sensors	Occupied or unoccupied states	1 month sensor data sampled at 5 minutes
14	**[59] Modeling behaviors of animals**	2016	Outdoor	-	Eating, drinking, resting and being milked	Ten consecutive days of cows` data
15	**[60] Characterisation of occupant behaviour**	2016	Indoor	Smart meters	Absent or asleep, medium and high consumption of the apartment.	Hourly weather data and smart metering data from 44 apartments.
16	**[61] Behavior prediction of disabled people**	2016	Indoor	-	Light, window, air conditioner, fan and curtains states	Users logs
17	**[62] Behavioral surveillance system for elderly care**	2015	Indoor	Security cameras	Falling from the bed, falling from the chair, collapsing, sitting and bending	35 different video sequences
18	**[63] Safety risk assessment for hydropower construction sites**	2014	Outdoor	GPS and Wi-Fi locating technology	Safety state space consisting; normal, dangerous and very dangerous states	20 minute trajectories
19	**[64] Behavioural change detection**	2014	Indoor	Activity logging	Activities (resting, sleeping, waking up, etc.)	Data logs for 3 months of 10 different users
20	**[65] Inferring occupancy from sensor data**	2014	Indoor	Motion sensors and smart electric meters	Number of people in the space and their identities	3 person house and 12 person lab occupancy data
21	**[66] Decision model for smart home environment**	2014	Indoor	CCTV cameras, occupancy and temperature sensors	Alarm, lighting and paging states	6000 events occurred in the building.
22	**[67] Recognizing human daily activities**	2011	-	Accelerometer sensor	Walking, standing, running, jumping, sitting and falling	492 samples from 13 subjects
23	**[68] Characterizing students behaviors**	2008	Indoor	Betty's Brain system	Edit map, ask query, request quiz.	Log files of students' interactions with the system
24	**[69] Understanding human behavior in a nursing center**	2007	Indoor	Static cameras	Walking, running, lying, sitting and standing	64 spatial segments
25	**[70] Recognizing behaviors from trajectories**	2005	Indoor	Static cameras	Short_meal, have_snack and normal_meal	45 observation sequences

1) Computing the probability of the observation sequence $P(O|\lambda)$ using the given HMM (λ) and the observation sequence $O = \{O_1, O_2, O_3, \dots, O_T\}$. 2) Extracting the most optimal hidden state sequence ($Q = q_1, q_2, q_3 \dots q_T$) which best explains the observations using the given HMM (λ) and the observation sequence $O = \{O_1, O_2, O_3, \dots, O_T\}$. 3) Adjusting the values of the state transition probabilities (A) and the output emission probabilities (B) of the HMM (λ) to maximize the probability of the observation sequence $O = \{O_1, O_2, O_3, \dots, O_T\}$. This process is called training as the observation sequence is used to train the HMM to make it best for observing real phenomena. However, this work focuses on the latter two scenarios.

E. Building Information Modeling (BIM)

BIM is a term used for describing building activities in an object-oriented Computer Aided Design (CAD) for representing building elements in terms of their geometrical, functional attributes and relationships [71]. While the BIM technology holds building information and n-dimensional visualizations [72], it lacks the real-time data related to the building environment. To overcome this limitation and to make BIM models dynamic for displaying accurate data from deployed building sensors, plug-ins are developed using Application Programming Interface (API) for adding more functionalities in the BIM model for stimulating the concept of smart buildings [73]. For our study, Autodesk Revit software is used because of its open-sourced API support and extensive use in the industry [74]. To take benefits from the functionalities offered by the Revit API, a visual scripting tool Dynamo is used as a Revit plug-in. The main advantage of using the Dynamo is that it has the power to construct programmatic relationships using a graphical user interface without having any requirement to write long lines of code from scratch [75]. However, it offers a coding option using Python language to configure 'code blocks' which are special nodes in Dynamo for performing advanced user-customized operations if needed [75]. Dynamo is used alongside Revit, requires an active Revit document containing a BIM model and follows a simple execution structure that is; input, process, and output [75]. It means that firstly the relevant building geometry is taken from Revit into Dynamo, then desired processes are performed using pre-packaged nodes, and finally the Dynamo outputs the results back into Revit document. This principle is used for generating a BIM model for identifying the most probable high-risk locations involving sharp worker movements and rotations.

III. SEMANTIC TRAJECTORY INSIGHTS USING HMMS

From our knowledge, no previous work was done to understand workers` mobility patterns within the stay locations using HMMs on semantic trajectories. After reviewing the relevant literature [76-82] on the understanding users` behaviors in dynamic environments, using a scenario (see Fig. 2) a prototype system is constructed by taking into account the roles of building supervisors and H&S managers. The system is developed using the scenario that is applied to an already constructed building. Trajectory data of building users that are working inside the building is acquired in real-time for experimental analysis. However, the functionality of the prototype system will remain the same if deployed on a construction site. The developed system focuses on the following;

A. Trajectory data acquisition and pre-processing

To understand the user interactions in the dynamic environment, use of IPS technology is considered. For trajectory data acquisition, 200 Bluetooth beacons are placed in different building locations. Each beacon (Fig. 3) is configured to transmit a Bluetooth signal within the radius of 4 meters. To collect the location data, a mobile application is developed using the Android platform (see Fig. 4). As application is launched in the user`s mobile device, it will detect all the deployed beacons in its range. After detection, it will select best three beacons` signals based on their maximum signal strength. Using these signals, a triangulation technique will be performed to get a unique pair of building location coordinates having longitude, latitude value, and the floor number. This process of tagging is achieved by utilizing the stored spatial information residing in a database as a deployment map of the beacons (see Fig. 5). Using this method, in a day 13,223 location points are recorded across different locations with a sampling interval of 1 second. However, the mobile application has the capability to set different sampling intervals range from 0.5 seconds to 5 seconds. The recorded data fields are shown in the Fig. 6. As it can be seen in that a single data log consists of a user identification (ID), timestamp, floor, latitude and longitude value. The acquired location data is stored in a document database (Mongodb). After data acquisition, R studio is used to execute pre-processing tasks for transforming location data into trajectories, and storing back in Mongodb for semantic enrichment process.

Fig. 3. Bluetooth beacons

The first step is to use an appropriate filter to remove outliers from the location data. There exist many outliers removing filters to improve the data quality [15]. However, a median filter is used because it depicts robustness property in filtering (see Fig. 7) [15]. After removing the outliers, stay points of users are calculated to enrich a trajectory with semantic points in the form of stop and move segments using Zheng et al. approach [15]. A single stay point is treated as a virtual location point characterized by a set of successive GPS points. Such as $Z = \{z_m, z_{m+1}, z_{m+2}, \dots, z_n\}$, $\forall\, m < i \le n$, $\quad Distance(z_m, z_i) \le D_{thresh}$ and $|z_n.T - z_m.T| \ \ge T_{thresh}$. A stay point is described as $s = (Latitude, Longitude, arrivaltime, leavingtime)$. Where,

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

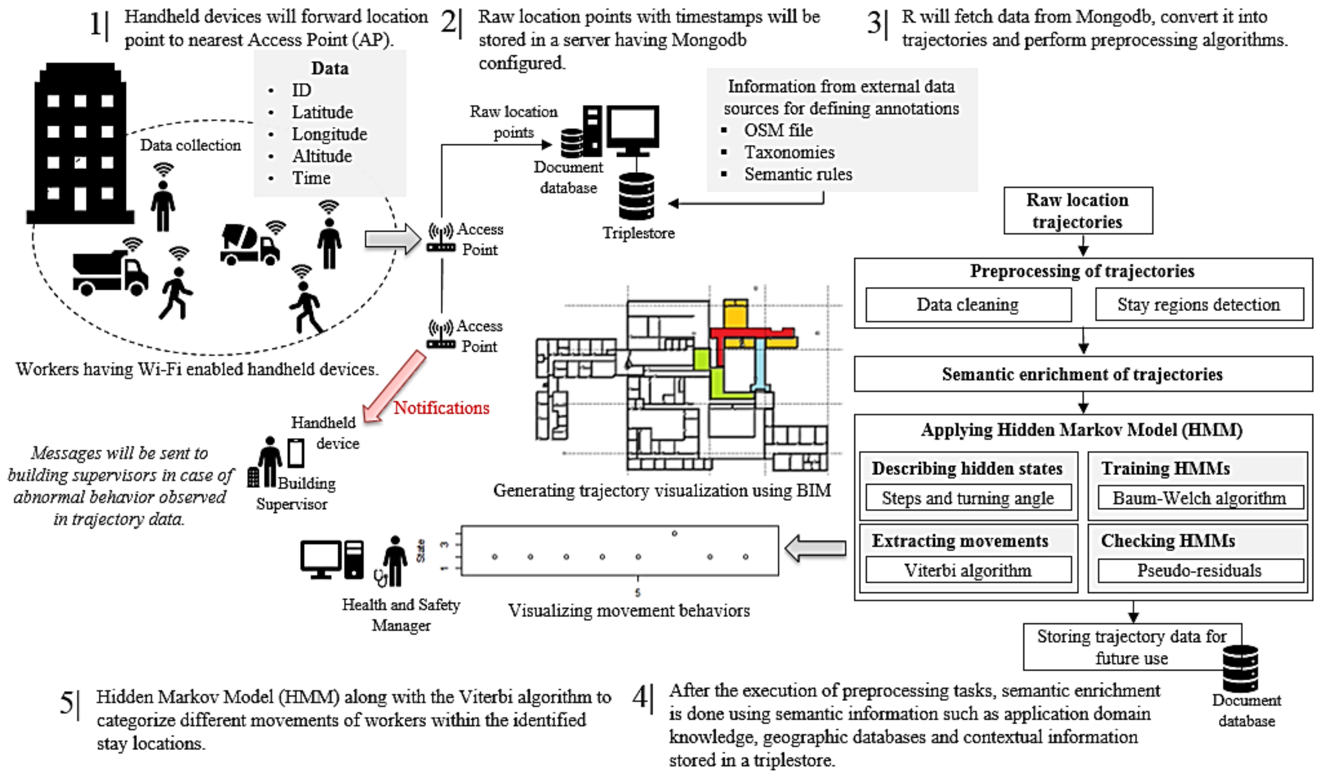

Fig. 2. Prototype system scenario

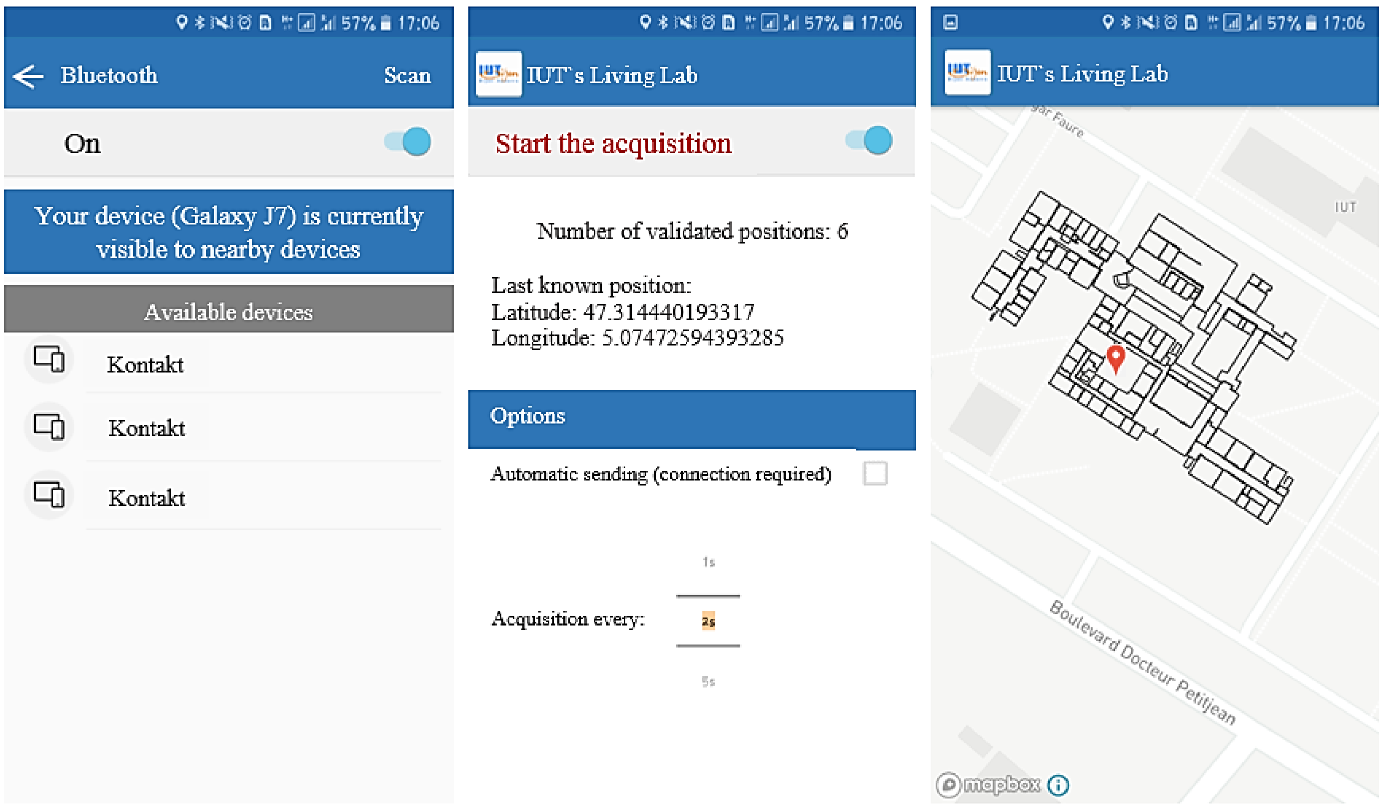

Fig. 4 Mobile application to collect beacon data for indoor geo-localization

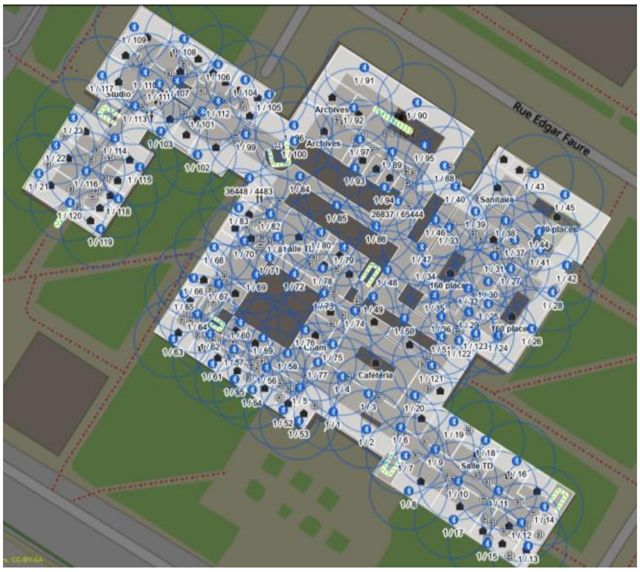

Fig. 5 Deployment plan of Bluetooth beacons


```
{
  "_id": {
    "$oid": "5ad5d55bafd0067343b27d1c"
  },
  "type": "Feature",
  "id": "0000-0000-0000-0000:1",
  "geometry": {
    "type": "Point",
    "coordinates": [
      5.0683130654777395,
      47.31075918540425
    ]
  },
  "properties": {
    "appuserid": "User1",
    "eventtype": "location",
    "devicedate": "2017-01-01T12:00:10",
    "dateString": "2017-01-01T11:00:10.000Z",
    "date": "2017-01-01T12:00:10",
    "location": {
      "building": "https://www.u-bourgogne.fr/stride#R1",
      "floor": "https://www.u-bourgogne.fr/stride#R2",
      "elevation": 1
    }
  }
},
```


Fig. 6 Beacon data values stored in a document database

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

The main purpose of calculating stay points in trajectories is to find locations in the building where users are spending most of their time. This information will help to track the occurrence of unexpected situations in the building if the stay duration is greater or less than the required. By setting the distance (D_{thresh}) value to 3 meters and time threshold (T_{thresh}) value to 20 minutes, stay points in a trajectory are identified as shown in Fig. 8. The dataset is available online (https://github.com/ChristopheCruz/LivingLabSride).However, these thresholds (hyper-parameters) are totally application dependent and can be changed according to an indoor or

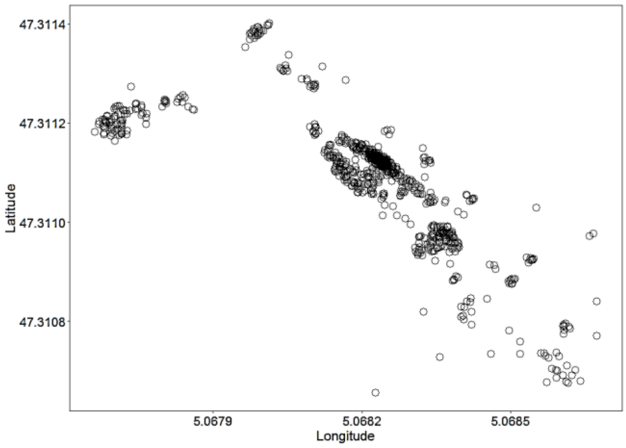

Fig. 7. Filtered trajectory data of a user.

outdoor environment. In addition, a frequency of visiting the same stay locations is also calculated to understand which stay locations are visited more often by the users. Once stay points are identified, each point is labelled with a building identification (ID) that corresponds to a building location such as Zone237. To tag an ID to a location point, we have utilized third-party data sources information. To do this, STriDE model is used that holds the information extract from an OSM file, taxonomies and semantic rules. The description of the STriDE model that is used for semantic enrichment process is stated in the next section.

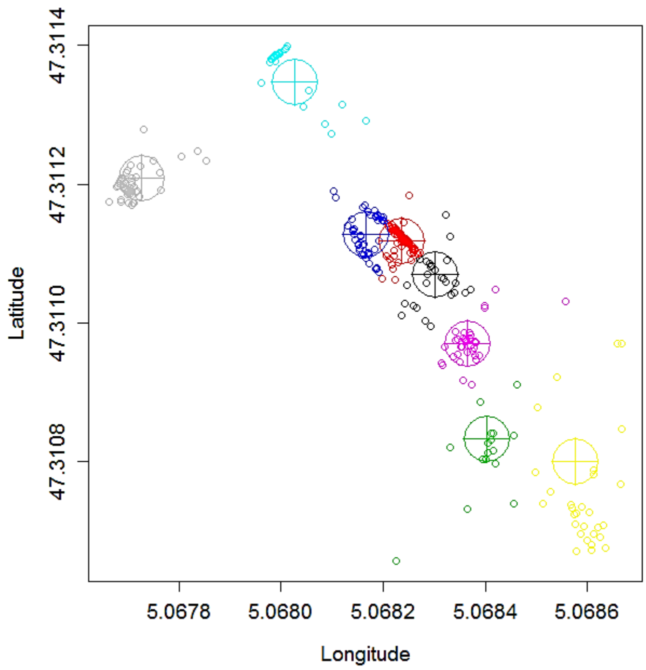

Fig. 8. Detection of stay points for a trajectory

B. *Trajectory semantic enrichment using the STriDE model*

For the semantic enrichment process, STriDE model requires an OSM file for a building under analysis, a set of semantic rules and a taxonomy as stated below (e.g. a BIM file [83]). OSM provides spatially rich geographic vector data in an XML format that is open and well-documented having the authorization for utilizing, copying or editing it [84]. OSM data model (see Fig.

9) contains nodes, ways and relations. In addition, an OSM entity can be tagged with single or multiple key-value pairs to add information. This information can be as simple (e.g. label) or complex (derived information from safety instructions) as per the application`s requirements. For example; one could specify that the "way" having an ID =237 is a corridor by adding a tag "highway=corridor" (see Fig. 11). These key-value pairs are clustered and maintained with the help of taxonomies created by the domain experts. In our case, an OSM file is used in which boundaries of the areas of the building zones are defined along with their links to each other. The file carries the complete building plan and its surroundings for defining the building structure.

```
<node id="2755" lat="47.523" lon= "5.214"/>

<way id="237" version="3" timestamp="2017-03-07T10:31:53Z" changeset="352">
  <nd ref="2755"/>
  <nd ref="2756"/>
  <nd ref="2757"/>
  <tag k="highway" v="corridor"/>
</way>
```

Fig. 9 OSM file structure

As shown in Fig. 9, "way id=237" is defined by a collection of nodes references (2755, 2756 and 2757) corresponding to different pairs of longitude and latitude. For labeling, each building location, taxonomy (see Fig. 10) is created in accordance with the building requirements. The created taxonomy is a hierarchy of concepts written as RDF triples using SKOS (Simple Knowledge Organization System) vocabulary. For example, in the Fig. 10, we have created a scheme of concepts named "ElementScheme" having a top concept (root) named "Element". This element has a narrower concept that is a "path", which itself has a narrower concept that is a "corridor". In addition, semantic rules are constructed in the form of a JSON file to link each OSM object with the taxonomy. For example, the rule shown in Fig. 11 states that any OSM entity having a tag whose key is "highway" and whose value is "corridor" is to be tagged with the concept "corridor" i.e. https://www.u-bourgogne.fr/stride#Corridor. Then, OSM file along with semantic rules are fed to a 2-step Java parser. Firstly, the parser will establish the mapping between each OSM entity with the Java object. Secondly, using semantic rules, these constructed objects are transformed into new Java objects according to the semantic definition (see Fig. 12). These processed objects are later stored in a Stardog (www.stardog.com), a triplestore to achieve the complete representation of the building environment. In the STriDE model (see Fig. 12), geometry is defined outside the main entity. Here, entity is describing the identity of the building location. For our scenario, on construction sites, the geometry of locations is changing over time. In order to log changes in shape, size and attributes of dynamic entities, we have used the concept of timeslices [23]. A timeslice consists of four components: an identity, properties having alphanumeric values, a time component indicating the validity of the timeslice and a geographical component depicting the spatial representation of the entity [23]. In case of any change in the components of the timeslice, excluding the identity, a new timeslice is created inheriting the components of the last known state of the entity and then those changes are applied. This newly created timeslice is then linked with the previous one to keep the evolution of the entity during its life cycle. Using stored semantic data from a triple store and multifaceted characteristics extracted during the pre-processing of trajectories, semantic trajectories are generated as shown in Fig. 13 (a and b). Semantically enriched processed trajectory is later used for the extraction of movement-related behavior information by applying the HMM probabilistic framework as discussed in the next section.

```
@prefix skos: <http://www.w3.org/2004/02/skos/core#> .
@prefix stride: <https://www.u-bourgogne.fr/stride#> .
stride:ElementScheme a skos:ConceptScheme ;
  skos:prefLabel "Thesaurus of the elements of a building"@en ;
  skos:hasTopConcept stride:Element .

stride:Element a skos:Concept ;
  skos:prefLabel "Element"@en ;
  skos:inScheme stride:ElementScheme .

  stride:Path rdfs:type skos:Concept ;
    skos:prefLabel "Path"@en ;
    skos:broaderTransitive stride:Element ;
    skos:inScheme stride:ElementScheme  .

      stride:Corridor rdfs:type skos:Concept ;
        skos:prefLabel "Corridor"@en ;
        skos:broaderTransitive stride:Path ;
        skos:inScheme stride:ElementScheme  .
```

Fig. 10. The script is a RDF Turtle https://www.w3.org/TR/turtle/ of the concept "Corridor" in the partially extracted STriDE schema

```
{
"type": "tag",
  "key": "highway",
  "value": "corridor",
  "compare": "equals",
  "conceptIRI": https://www.u-bourgogne.fr/stride#Corridor
}
```

Fig. 11. Semantic rule example

```
stride:W237 a stride:Entity ;
  skos:prefLabel "Corridor of floor 0" ;
  rdfs:comment "" .

stride:GEO-W237-0 a geo:Geometry ;
  geo:asWKT "LINESTRING (20 65, 15 65, 15 50, 20 50, 15 50, 15 15, 20 10, 15 15, 30 15, 30 20, 30 15, 50 15, 50 20)"^^geo:wktLiteral .

stride:W237-0 a stride:TimeSlice ;
  stride:hasStartDate "2018-01-01T00:00:00"^^xsd:dateTime ;
  stride:hasFeature stride:W237 ;
  geo:hasGeometry stride:GEO-W237-0 ;
  dct:subject stride:Corridor ;
  stride:hasEndDate "9999-12-31T23:59:59"^^xsd:dateTime .
```

Fig. 12. Parsed OSM file using the semantic rules and the taxonomy - The script is the RDF Turtle definition of an object of the kind "Corridor" identified by the value stride:W237. These values are used to define semantic trajectories.

SPARQL Results (returned in 55 ms)

traj	userName	location
stride:TrajOfMaintenance1-1	Maintenance 1	Outdoor pathway
stride:TrajOfMaintenance1-2	Maintenance 1	Storage room
stride:TrajOfUser1-1	User 1	Outdoor pathway
stride:TrajOfUser1-2	User 1	Corridor of floor 0
stride:TrajOfUser1-3	User 1	Office 1
stride:TrajOfUser1-4	User 1	Corridor of floor 0
stride:TrajOfUser1-5	User 1	Outdoor pathway
stride:TrajOfUser1-6	User 1	Storage room

Fig. 13(a). Semantic trajectories of two users

SPARQL Results (returned in 47 ms)

s	p	o
stride:TrajOfUser1-1	rdf:type	stride:TimeSlice
stride:TrajOfUser1-1	stride:hasStartDate	2018-01-02T09:00:00
stride:TrajOfUser1-1	stride:hasFeature	stride:TrajOfUser1
stride:TrajOfUser1-1	stride:hasEndDate	2018-01-02T09:05:00
stride:TrajOfUser1-1	stride:isTrajectoryOf	stride:User1
stride:TrajOfUser1-1	stride:hasLocation	stride:W1

Fig. 13(b). Timeslices of a trajectory; each consisting of an identity for its representation, valid period of time and a geometric component for the spatial representation

C. *Categorizing worker movements using HMMs*

To categorize the worker trajectory into different movement states for each stay location, an independent HMM is initially trained. According to the existing literature [14], the movement behavior of a moving object can easily be defined by calculating individual step length and turning angle. However, before feeding the trajectory sequences into an HMM, we need to visualize the data in order to set the hidden state values using measurement variables. For our case, we have used values of 'step length' (distance between two trajectory points) and 'turning angle' (change in direction in radians from the previous point to the current point), which were extracted during the pre-processing of trajectories. Step length (l_t) is calculated using the Haversine distance formula [14] between the locations (x_t, y_t) and (x_{t+1}, y_{t+1}) as below;

$$d = 2r\sin^{-1}\left(\sqrt{\sin^2\frac{x_{t+1}-x_t}{2} + \cos x_t \cos x_{t+1} \sin^2\frac{y_{t+1}-y_t}{2}}\right)$$

Where "r" is the radius of the earth. The reason for using the Haversine distance formula is because it is the one of the preferred method that calculates the geographic distance between two points on a sphere [85]. While turning angle ϕ_t is calculated as the change in bearing b_t as $b_t = atan2(y_{t+1} - y_t, x_{t+1} - x_t)$ between the time intervals $[t-1, t]$ and $[t, t+1]$ [86]. For visualizing the occurrences (frequency) of different values of the step length and turning angle in a trajectory during the sampling interval of 1 minute, time series and histograms are plotted in Fig. 14 to Fig 17. Time series plots will help us to

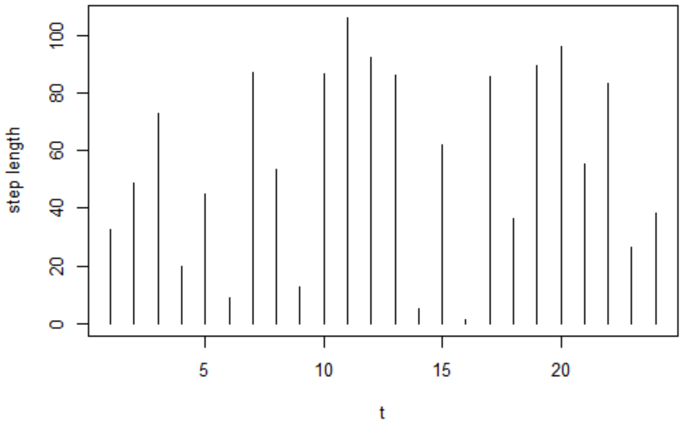

Fig. 14. Time series of step lengths of a single trajectory

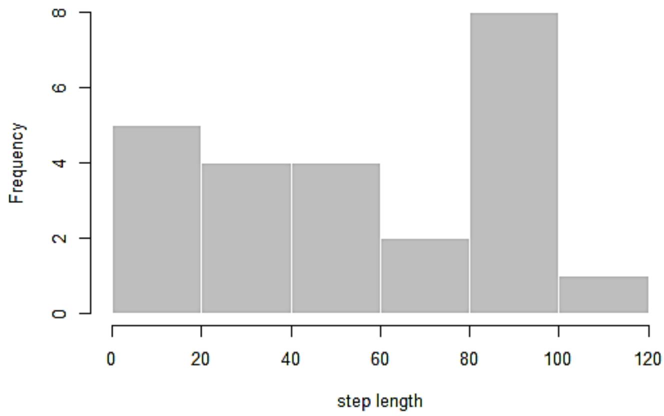

Fig. 15. Histogram of step lengths of a single trajectory

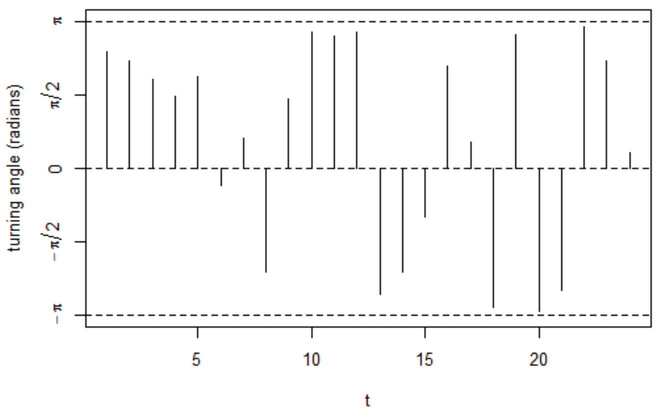

Fig. 16. Time series of turing angles of a single trajectory

detect any outliers or sudden shifts in the trajectory. Whereas, histograms provides us the frequency of the values of step length and turning angle in an entire trajectory. For movements' categorization of the workers within the stay locations, we need to select the hidden states. This is the most important step while training an HMM to generate most probable values of the hidden states. From the existing literature [20], the normal walking speed for an adult range from 1.0 to 1.6 meters per second (m/s). However, keeping an indoor environment into consideration, we have used the value of 1.4 m/s as a safe walking speed limit that will give us 84 steps per minute i.e. the sum of step lengths for

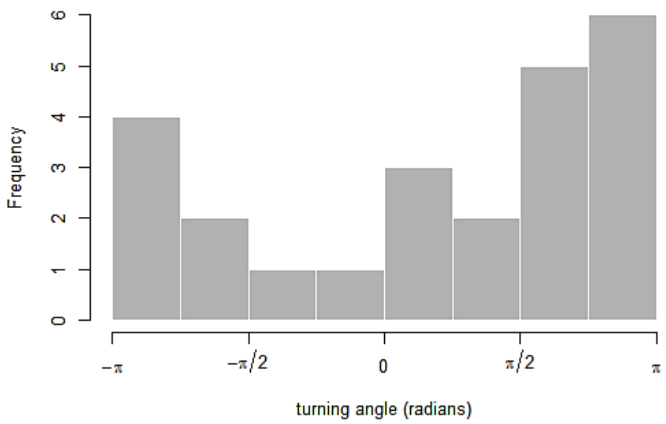

Fig. 17. Histogram of turning angles of a single trajectory

a minute. Four hidden states are defined using different values of 'step length' and 'turning angle' are shown in Table 3. However, the number of states can be decreased or increased according to the application requirements. The purpose of defining such hidden states is to identify unsafe worker movements within the stay location. For defining the hidden states in the HMM, we need to specify distributions according to the acquired trajectory data. For this, we have used Gamma distribution for step lengths and Von Mises distribution (also known as the circular normal distribution) for turning angles [87, 88] as shown in Table 4. For more details on these distributions, see Abramowitz and Stegun [87], and Gatto and Jammalamadaka [88] research.

TABLE 3 CRITERIA FOR DEFINING HIDDEN STATES

State symbol / description	Step length threshold (No. of steps in a minute)	Turning angle threshold (Radians)
S_1 Short steps and fewer turnings	steps ≤ 84	angle ≤ π/2
S_2 Short steps and many turnings	steps ≤ 84	π/2 < angle ≤ π
S_3 Long steps and fewer turnings	steps > 84	angle ≤ π/2
S_4 Long steps with many turnings	steps > 84	π/2 < angle ≤ π

TABLE 4 USED DISTRIBUTIONS FOR DEFINING STATES [87, 88]

Distribution	Parameters	Link function
Gamma	$mean > 0$	Logarithmic
	$standard\ deviation > 0$	Logarithmic
Von Mises	$mean\ \epsilon\ \{-\pi, \pi\}$	Logarithmic
	$concentration > 0$	Tangent $\left(\frac{mean}{2}\right)$

After defining the hidden states, we have used Baum-Welch algorithm [28] in the R studio that allows learning parameters of an HMM using the maximum likelihood approach and computes λ^* that maximizes the likelihood of the sample of training sequences $\chi = \{O^k\}_k^K$ namely $P(\chi|\lambda)$ (see Fig. 18). However, before training the model, we have to input initial probabilities. For our case, as already discussed above, there are four different states which are 'short steps and fewer turnings', 'short steps and many turnings', 'long steps and fewer turnings', and 'long steps and many turnings' are denoted as S_1, S_2, S_3 and S_4. Dividing the total probability equally for all the states. This will give us initially estimated states` probabilities as mentioned below;

$$\pi = \pi_1, \pi_2, \pi_3, \pi_4 = [1/4 \quad 1/4 \quad 1/4 \quad 1/4]$$

In order to fulfil the requirements of real construction processes, initial probabilities should be calculated based on the existing situation of construction sites, and using historical records for improved accuracy in prediction methods. After computing the transition probabilities of workers to move between different states using the Baum-Welch algorithm, the most probably occurring values of the hidden states are also extracted based on the output of learned parameters. For this, we have used Viterbi algorithm [28], a form of dynamic programming to extract the most probable sequence of states for a given trajectory. For each symbol o_T in the observation trajectory O, an algorithm calculates the probability of its emission for each possible hidden state. The algorithm starts by computing the initial probability of the emission of the symbol o_1 in all possible states. Then, it computes again the emission of the symbol o_2 for each state transition. This process is repeated for every observation symbol until the observation sequence ends that is at step T. Finally, having all possible paths covered, the Viterbi algorithm [28] look for the path, and output the most probable state sequence $Q = \{q_1, q_2, q_3 \ldots q_T\}$. Fig. 18 shows the most probable state sequence within a stay location using the trained HMM. The last three trailing rows in the Fig. 18 are revealing the probabilities of each hidden state individually at every observation index. In case of identification of unsafe movements (having longer steps and many turnings) in the extracted most probable path (see Fig. 18), the reason of occurrences of such movements can be figured out by taking into account its corresponding values of the step length and the turning angle in time series trajectory data using the observation indexes (see Fig. 14 and 16).

D. Visualizing user movements using a BIM model

To add the functionality for displaying categorized user movements, a Revit Architecture software along with a Dynamo (a Revit Plug-in) is used. The building that is used for the study, its BIM model doesn't exist. A simple building structure without architectural aesthetics is created in a Revit software utilizing the information from an OSM file of a building for demonstrating a proof of concept integration of systems. After constructing the BIM model of a building, Dynamo is used for generating the visualizations. Constructing a Dynamo script (also called a graph) consists of 'nodes' and 'wires'. Nodes are the blocks of code that aim to execute a single discrete function [75]. Whereas, wires are used for connecting nodes` outputs to nodes` inputs for building the data flow from left to right [75]. Dynamo graph constructed for the our study consists of the following key steps, which are; All the building locations which are tagged as 'rooms' in the Revit model are extracted by defining the category of elements as 'rooms' in the Dynamo. A room in the Revit denotes a three-

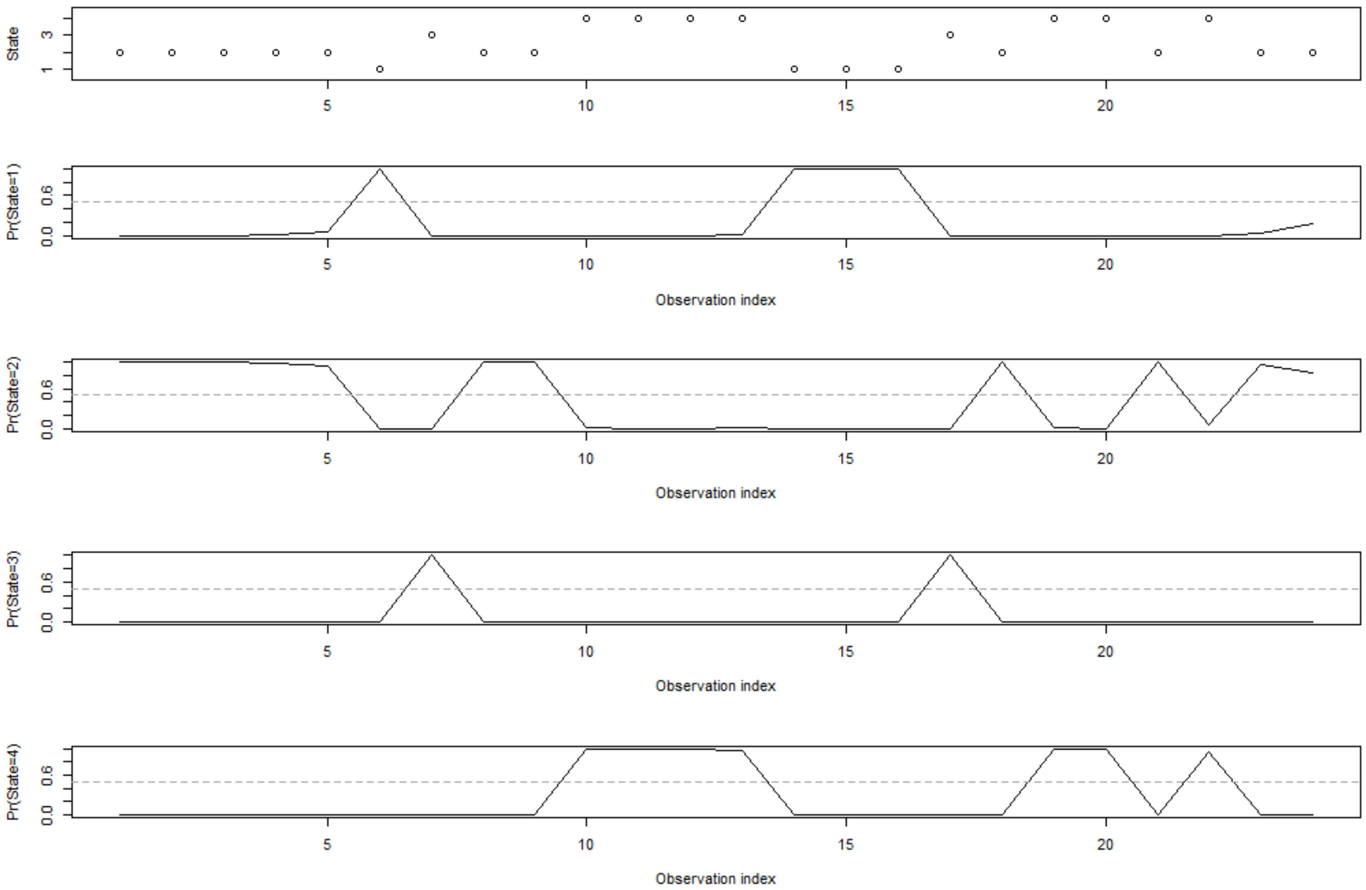

Fig. 18 Decoded states sequence (top row) and state probabilities of observations (last three rows)

dimensional volume for representing a real building space. Building spaces that need to be identified as individual rooms should be properly bounded by walls before placing 'room tags' on them. These tags are called as annotation elements and can be changed manually. Each tagged room in Revit carries the semantic information in the form of a set of parameters such as room number, room name, physical area, etc. that are used for viewing or editing that particular room. In our case, the parameter 'room name' will act as a unique identification for each space representing a POI for visualizing the problematic locations in the Revit model using the output of the Viterbi algorithm. The naming convention of Revit rooms is set according to the tagging of semantic points as described in the STriDE model. For visualizing the users` movements on a BIM model, the input trajectory streams of the Viterbi algorithm are semantically enriched with their corresponding POIs` information which is stored in the STriDE model. This information of the POIs is further mapped with the Revit`s room list extracted from an active Revit document for displaying the output of the Viterbi algorithm (the probabilities) using the different colors for representing each movement state of a user. The changes in the color of the Revit rooms are achieved using the 'Element: OverrideColorInView' node. This node changes the color of the Revit element (room in our case) in the active view irrespective of the current element display properties as shown in Fig. 19

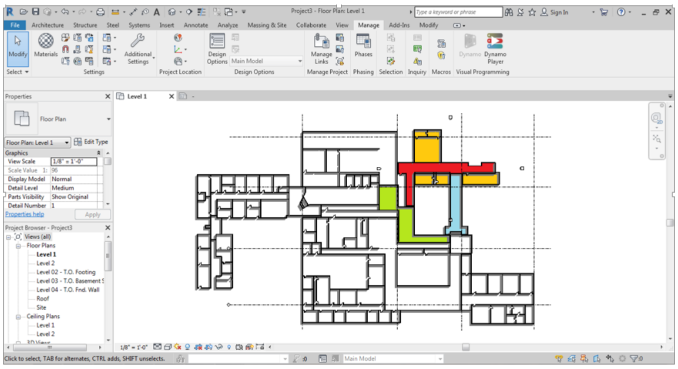

Fig. 19 BIM model constructed for plotting output of the Viterbi algorithm based on the most probable observed states` values as; 'short steps and fewer turnings' in Green, 'short steps and many turnings' in Blue, 'long steps and fewer turnings' in Orange, and 'long steps and many turnings' in Red color.

IV. Discussion

The experimentation is conducted using batch processing techniques in which location data of building users in a day is first acquired using Bluetooth beacons and then pre-processing algorithms are executed in an offline mode. However, for achieving the real-time insights of worker movements on construction sites, stream processing techniques can be used to pre-process the worker trajectories in an online mode. However, the HMM model computations should remain a batch process. After collecting a required dataset of trajectories, workers` stay regions are identified with their stay duration in that particular day. The reason for extracting such locations in a trajectory is to identify the semantic regions (ROIs) and their associated POIs. The ROIs are the wider areas of the building (e.g. work-zone237, etc.) which consist of multiple geographical POIs (e.g. outdoor pathway, storage room, office1, etc.) labelled as 'rooms' in our model. The idea of dividing a building into different semantic regions is taken from the Pradhananga and Teizer, 2013 research [76]. As on a typical construction site, there exist different zones [80] such as material zone, dumping zone, loading zone, work zone, etc. which are specified by the building supervisors for dividing a site into different regions for well-organized site monitoring and management. The developed system has identified a number of stay locations in users` trajectories as shown in Fig.10 which relate to different semantic regions. For mapping the trajectories with their corresponding semantic regions, spatial joins are performed as discussed in the section III (b). Figures 20 and 21 are showing the extraction of the semantic regions mapped to a set of users` trajectories. Figuring out the amount of time spent by the users in each identified semantic region will give a good understanding to know the rate of utilization of the location. In the similar way, on the sites, trajectories of machinery can be tracked using semantic regions which covers wider geographical areas for analyzing how often machinery enter a particular site location and for how long it stays over there. This information will help building supervisors and H&S managers in managing the building or site resources effectively in real-time by knowing the locations and machinery` usage.

```
select ?user ?location ?concept ?start ?end ?duration{
?traj stride:isTrajectoryOf ?u.
?u skos:prefLabel ?user.
?traj stride:hasLocation ?loc.
?loc skos:prefLabel ?location.
?traj stride:hasStartDate ?start.
?traj stride:hasEndDate ?end.
?locTS stride:hasFeature ?loc.
?locTS dct:subject ?concept.
?locTS stride:hasStartDate ?locStart.
?locTS stride:hasEndDate ?locEnd.
bind( minutes(?end-?start) as ?minutes ).
bind( hours(?end-?start) as ?hours ).
bind( 60*?hours+?minutes as ?duration ).
FILTER(?duration > 20 && ?start >= ?locStart && ?start
<= ?locEnd && ?end != "9999-12-31T23:59:59"^^xsd:dateTime)
}order by ?user ?start
```

Fig. 20. Querying a triplestore for extracting semantic regions of a user

user	location	concept	start	end	duration
Maintenance 1	Storage-zone-300	stride:Storage	2018-01-02T09:05:00	2018-01-02T12:00:00	175
User 1	Work-zone-237	stride:SafeWorkzone	2018-01-02T09:06:00	2018-01-02T09:59:00	53
User 1	Work-zone-237	stride:DangerousWorkzone	2018-01-02T10:01:00	2018-01-02T11:15:00	74
User 1	Storage-zone-300	stride:Storage	2018-01-02T11:21:00	2018-01-02T12:00:00	39

Fig. 21. Semantic regions of a user

For further semantic annotations, a semantic region named 'work-zone237' is taken into consideration and trajectories belonging to two users (User1 and Maintenence1) are broken down into multiple timeslices using the STriDE model as shown in Fig. 13(a). The process of constructing multiple timeslices is achieved based on the detection of a change in the components (spatial, temporal and semantics) of a user trajectory`s timeslice as shown in Fig. 13(b). After extracting the semantic stay regions, an entire set of trajectory points belonging to a particular stay region is further analyzed using HMMs for identifying unsafe user movements involving long steps and many turnings per minute within different POIs. Though, the insights can be extracted on the data sampled per second to understand worker movements more closely. The same process can be repeated for each identified stay region to have a complete picture of different movements occurring inside all the critical building locations. The purpose here is to understand the movements within the trajectories with respect to identified POIs corresponding to different building zones. As it can be seen from the most probable values of the hidden states (see Fig. 18) that there are fluctuations in the movements observed during different times in a day within a building zone named as 'work-zone237'. There are unsafe movements identified having long steps or many turnings as the 4th state in the Fig. 18. These hidden states` values correspond to different building locations are used for plotting the building model for visualizing the most probable user movements in different regions of the building within the identified stay location as shown in Fig. 19. For further tracking the unsafe movements, the corresponding speed and turning angle values associated with such movements can be traced from the time series plots as shown in Fig 14 and 16 for more analysis.

Visualizing categorized movements using step lengths and turning angles using BIM software will help H&S managers to monitor critical building and site locations for identifying accident prone scenarios in the buildings or on the construction sites, and enabling them to take quick actions in case of detection of unsafe worker movements. In addition, the system will also help building supervisors in maintaining safe distances between the workers and the construction machinery operating at high speed or many turnings, and this is achieved by increasing the level of spatial awareness about sites using semantic trajectories. Such real-time visualizations can result in preventing the occurrence of accidents because of an excessive proximity between the workers and the machinery. Moreover, if the construction equipment is not being operated in the safe speed limits on a construction site, such visualizations of unsafe movements using a BIM model will help H&S managers in identifying the workers who are not complying the safety regulations and the requirement for machinery operators` trainings will be determined accordingly. Furthermore, understanding different movements of the workers and the machinery will act as a pro-active measure for preventing fatalities and the planning of the future construction operations can be readjusted by the building supervisors to avoid unsafe situations.

In the developed prototype system, we have used 4 hidden states for categorizing the user movements using different values of step length and turning angle. The reason for choosing such number of states is based on the Akaike Information Criteria (AIC) [87]. AIC has been used extensively for excessing the relative quality of statistical models such as HMMs, and aims to find the best approximating model to the true data generating process [87, 88]. AIC is calculated using the following formula;

$$AIC = -2\ln(L) + 2p$$

Where, L is the likelihood model function and p is the number of estimated model parameters. For our prototype system, to test the model performance, we chose the number of states from 2 to 6 to keep the model simple. However, increasing the number of states will also increase the number of model parameters (values for step and angle) and thus increasing the complexity of the HMM. Using the above formula, AIC values are calculated in Table 5.

TABLE 5 AIC VALUES FOR FITTING HMM

No. of states	No. of parameters	AIC value
2	4	1529.911
4	8	1506.903
6	12	1507.910

A lower AIC value indicates a better performance of an HMM relative to a given trajectory data [89]. In the Table 5, a difference between AIC value for 4-state HMM and 6-state HMM is not significant. Hence, the 4-state HMM is investigated further to categorize worker trajectory into four different states. To evaluate the general goodness of fit of the 4-state trained model, we have used ordinary pseudo-residuals to check if the trained HMM is a true data generating process of a trajectory. It will also help us to detect extreme deviations from the observed data. The pseudo-residuals should follow a standard normal distribution if the trained model is a true data-generating process [89]. It means that, if the model fits the data well, the points in the qq-plot will be closer to the straight line and a deviation from normality will indicate a lack of fit [90, 91]. Moreover, an absolute value of residual increases with increased deviation from the straight line and specific observations can be known in less time. For more details on pseudo-residuals, see Zucchini et al. [91]. The pseudo-residuals of the 4-state model fitted to the trajectory data are displayed in Fig. 22. The last row of the Fig. 22 shows quantile-quantile plot pseudo-residuals, with the theoretical quantiles on the horizontal axis. The plot shows that the trained model is well fitted for the observation dataset, and has few deviations from the normality. However, the goodness of fit of the model can be improved by adjusting the hidden states [92]. The Fig. 23 shows correlogram i.e. an autocorrelation (ACF) plot of the 4-state HMM model. Such plot shows that the correlation of the time series with itself, lagged by x time units [93]. In short, ACF function shows if the present value depends consistently on previous lagged values of a time series. The plot scale ranges from -0.4 to 0.9 called as the 'correlation coefficients'. Whereas, dotted lines at -0.4 to 0.4 showing the 'significance levels' of the plot. We have a correlogram till lag 20, however we can make it until any other desired value. As it can be observed that, the only value is the spike at lag 1, showing the high correlation but correlation value is decreasing drastically after a few lags showing that the lags do not have any significant effect. Hence from this plot it appears that the amount of residual autocorrelation is low and thus indicating that the trained 4-state HMM model have tried to capture most of the relevant correlation structures in the provided dataset.

V. CONCLUSION AND FUTURE WORK

The study explained the functionality of the developed prototype system that uses an IPS technology and HMM based probabilistic framework for visualizing categorized worker movements within the stay region using the BIM approach. Visualizing the results of HMMs using the BIM after analyzing the semantic trajectories will enable the system users (H&S managers and building supervisors) to see the exact locations where the unsafe movements are occurring in real-time during construction and facility management operations. For the trajectory data acquisition, Bluetooth beacons are used. Though, Bluetooth beacons are not recommended as an indoor tracking solution because of less precision in the accuracy of determining the locations. These errors were noticeable in our experiments as collected trajectory data points did not completely joined spatially with the semantic point information extracted from an OSM file. Consequently, the closest possible semantic points were mapped. For extracting the stay locations, the subsequent trajectory data points at the same location are grouped together. For grouping, a time threshold of 20 minutes and distance threshold of 3 meters are used. It means that subsequent stops between which the time gap is equal to 20 minutes within an area of 3 meters are grouped together. The reason for the 20 minutes threshold is that intervals smaller than 20 minutes are likely to represent a user that was just visiting the stay location for a short period of time. After stay locations` detection, trajectory enrichment of a set of trajectory segments belonging to a corresponding semantic region is achieved using the STriDE model. Then, HMMs along with the Viterbi algorithm are used for categorizing movements within the stay region. In the end, BIM approach is used for identifying the unsafe user movements in different POIs within a semantic region. The reason of using BIM approach is because, it is becoming a standard worldwide in the Architecture, Engineering & Construction industry. Our system functionality of analyzing semantic trajectories that is integrated with the BIM works as a plug-in that aims to enhance the present capabilities of a BIM software by providing a reliable and up-to-date information about the different types of users` movements occurring within the facility throughout the building lifecycle to AEC team members for conducting further studies, simulations or operations using building occupants` mobility data.

The developed system will not only serve as a semantic trajectory visualization platform for H&S managers but will provide the most probable occurring movement patterns in real-time that can improve building monitoring processes by identifying unsafe worker behaviors. As the proposed system prototype is still in the development phase, and this article presents some initial results. Further work needs to be done to take a complete advantage of the STriDE model for modeling, and testing the dynamic environment scenarios where the purpose and the nature of locations are constantly changing with the passage of time. For example, a work-zone237 is now a dangerous zone. See Fig. 21 that how STriDE keeps this evolution about the nature of the location by utilizing the definitions of "concepts". Here, a 'concept' is defined as an object with a set of attributes which are maintained using the concept of timeslices in our model. Hence, the change in the nature of the building region will results in different stay regions having different users` mobility behaviors. These dynamic changes, and interactions can be efficiently managed by the STriDE model that is designed in a way to capture the information of changing as well as moving objects in a dynamic building scenarios. Moreover, future research needs to focus to empirically test the system on real construction sites as at present a system is modelled for an already constructed building. Taking information from the real-time site cases will improve the training of HMMs. It will result in improved, and precise mobility behavior extraction for identifying fast moving and turning construction resources such as workers and machineries to reduce accidents on sites.

ACKNOWLEDGMENTS

The authors thank the Conseil Régional de Bourgogne-Franche-Comté, the French government for their funding, SATT Grand-Est, and IUT-Dijon (http://iutdijon.u-bourgogne.fr) for allowing us to conduct a real-time case study in their laboratory. The authors also want to thank Orval Touitou for his technical assistance to this research work.

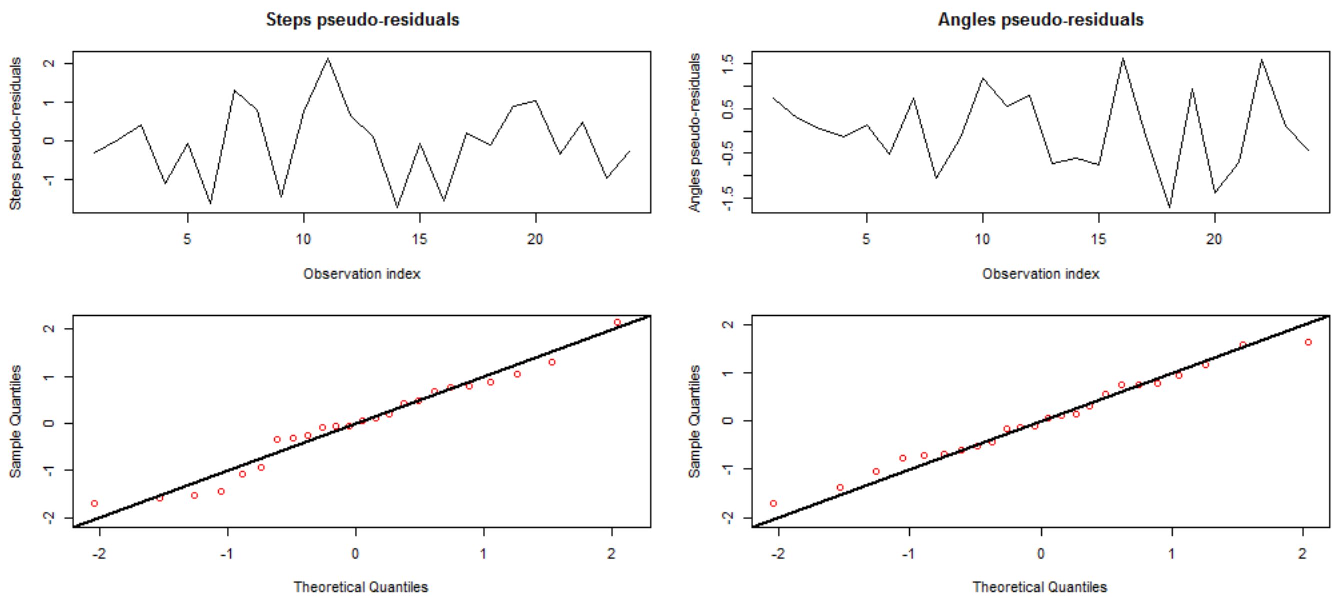

Fig. 22 Time series (top row) and qq-plots (bottom row) of the pseudo-residuals of the 4-state HMM

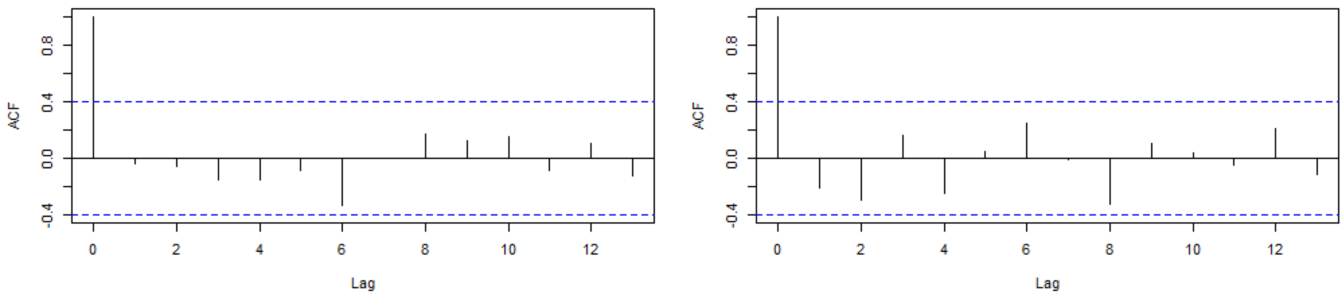

Fig. 23 ACF plot of the 4-state HMM against lagged values of a traejctory

REFERENCES

[1] R. Stasiak-Betlejewska, M. Potkány, Construction costs analysis and its importance to the economy, Procedia Economics and Finance 34 (2015) 35-42.

[2] K. M. Rashid, A. H. Behzadan, Risk behavior-based trajectory prediction for construction site safety monitoring, J. of Construction Engineering and Management 144(2) 04017106 (2017) 1-10.

[3] H. Li, X. Yang, F. Wang, T. Rose, G. Chan, S. Dong, Stochastic state sequence model to predict construction site safety states through real-time location systems, Safety science 84 (2016) 78-87.

[4] R. Akhavian, A. H. Behzadan, Smartphone-based construction workers' activity recognition and classification, Automation in Construction 71 (2016) 198-209.

[5] J. Park, X. Yang, Y. K. Cho, J. Seo, Improving dynamic proximity sensing and processing for smart work-zone safety, Automation in Construction 84 (2017) 111-120.

[6] B. Y. Jeong, Occupational deaths and injuries in the construction industry, Applied ergonomics 29(5) (1998) 355-360.

[7] R. A. Haslam, S. A. Hide, A. G. Gibb, D. E. Gyi, T. Pavitt, S. Atkinson, A. R. Duff, Contributing factors in construction accidents, Applied ergonomics 36(4) (2005) 401-415.

[8] Stats.bls.gov.2017, "Census of Fatal Occupational Injuries (CFOI) - current and revised data," available at: http://stats.bls.gov/iif/oshcfoi1.htm (Last accessed 28 March 2018).

[9] HSE 2017, "Fatal injuries arising from accidents at work in Great Britain 2017," available at:http://www.hse.gov.uk/statistics/pdf/fatalinjuries.pdf (Last accessed 07 June 2018).

[10] Y. Li, F. W. Guldenmund, Safety management systems: A broad overview of the literature, Safety Science 103, (2018) 94-123.

[11] R. Gao, A. P. Chan, S. Lyu, H. Zahoor, W. P. Utama, Investigating the difficulties of implementing safety practices in international construction projects, Safety Science 108 (2018) 39-47.

[12] X. S. Dong, J. A. Largay, X. Wang, C. T. Cain, N. Romano, The construction FACE database—Codifying the NIOSH FACE reports, J. of Safety Research 62 (2017) 217-225.

[13] J. Teizer, T. Cheng, Proximity hazard indicator for workers-on-foot near miss interactions with construction equipment and geo-referenced hazard areas, Automation in Construction 60 (2015) 58-73.

[14] Ľ. Ilkovičová, P. Kajánek, A. Kopáčik, Pedestrian indoor positioning and tracking using smartphone sensors step detection and map matching algorithm, International Symposium on Engineering Geodesy (2016) 20-22.

[15] Y. Zheng, Trajectory data mining: An overview, Transactions on Intelligent Systems and Technology 6(3) (2015) 1-41.

[16] K. Heijden, Scenarios: The Art of Strategic Conversation, 2nd ed. John Wiley & Sons, 2005.

[17] J. M. Carroll, Scenario-based design: envisioning work and technology in system development, John Willey & Sons, Inc., New York, 1995.

[18] H. Li, G. Chan, J.K.W. Wong, M. Skitmore, Real-time locating systems applications in construction, Automation in Construction 63 (2016) 37-47.

[19] Kontakt 2018, "Beacons," available at:https://kontakt.io/ble-beacons-tags/ (Last accessed 28 March 2018).

[20] Z. Yan, S. Spaccapietra, Towards semantic trajectory data analysis: a conceptual and computational approach, VLDB PhD Work 15(2), 2009.

[21] F. Wu, H. Wang, Z. Li, W.C. Lee, Z. Huang, SemMobi: A semantic annotation system for mobility data, Proceedings of the 24th International Conference on World Wide Web (WWW '15 Companion) (2015) 255-258.

[22] M. Arslan, C. Cruz, D. Ginhac, Semantic enrichment of spatio-temporal trajectories for worker Safety on construction sites, Procedia Computer Science 130 (2018) 271-278.

[23] C. Cruz, Semantic trajectory modeling for dynamic built environments, IEEE International Conference on Data Science and Advanced Analytics (DSAA) (2017) 468-476.

[24] Z. Z. Wang, C. Chen, Fuzzy comprehensive Bayesian network-based safety risk assessment for metro construction projects, Tunnelling and Underground Space Technology 70 (2017) 330-342.

[25] J. Seo, S. Han, S. Lee, H. Kim, Computer vision techniques for construction safety and health monitoring, Advanced Engineering Informatics 29(2) (2015) 239-251.

[26] F. Liu, D. Janssens, J. Cui, G. Wets, M. Cools, Characterizing activity sequences using profile Hidden Markov Models, Expert Systems with Applications 42(13) (2015) 5705-5722.

[27] Q. Fang, H. Li, X. Luo, L. Ding, T. M. Rose, W. An, Y. Yu, A deep learning-based method for detecting non-certified work on construction sites, Advanced Engineering Informatics 35 (2018) 56-68.

[28] L. R. Rabiner, A tutorial on Hidden Markov Models and selected applications in speech recognition, Proceedings of the IEEE 77(2) (1989) 257-286.

[29] S. Spaccapietra, C. Parent, M. L. Damiani, J.A. de Macedo, F. Porto, C. Vangenot, A conceptual view on trajectories, Data and Knowledge Engineering 65(1) (2008) 126-146.

[30] W. Zucchini, An introduction to model selection, J. Math. Psychol. 44(1) (2000) 41-61.

[31] Z. Riaz, M. Arslan, A.K. Kiani, S. Azhar, CoSMoS: A BIM and wireless sensor based integrated solution for worker safety in confined spaces. Automation in construction 45 (2014) 96-106.

[32] N. Nawari, Standardization of structural BIM. Computing in Civil Engineering (2011) 405-412.

[33] J. Teizer, U. Mantripragada, M. Venugopal, Analyzing the travel patterns of construction workers, Proceedings of the 25th International Symposium on Automation and Robotics in Construction, Vilnius, Lithuania (2008) 26-29.

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

[35] A. Frihida, D. Zheni, H. B. Ghezala, C. Claramunt, Modeling trajectories: A spatio-temporal data type approach, 20th International Workshop on Database and Expert Systems Application DEXA'09 (2009) 447-451.

[36] C. Parent, S. Spaccapietra, E. Zimányi, Conceptual modeling for traditional and spatiotemporal applications: the MADS approach, Springer-Verlag Berlin Heidelberg, New York, 2006.

[37] D. Noël, M. Villanova-Oliver, J. Gensel, P. Le Quéau, Design patterns for modelling life trajectories in the semantic web, International Symposium on Web and Wireless Geographical Information Systems (2017) 51-65.

[38] C. Parent, S. Spaccapietra, C. Renso, G. Andrienko, N. Andrienko, V. Bogorny, M.L Damiani, A. Gkoulalas-Divanis, J. Macedo, N.Pelekis, Y.Theodoridis, Semantic trajectories modeling and analysis, ACM Computing Surveys (CSUR) 45(4) (2013) 1-37.

[39] B. Harbelot, H. Arenas, C. Cruz, LC3: A spatio-temporal and semantic model for knowledge discovery from geospatial datasets, Web Semantics: Science, Services and Agents on the World Wide Web 35(1) (2015) 3-24.

[40] B. Harbelot, H. Arenas, C. Cruz, Continuum: A spatiotemporal data model to represent and qualify filiation relationships, Proceedings of the 4th ACM SIGSPATIAL International Workshop on GeoStreaming (IWGS) (2013) 76-85.

[41] H. Arenas, B. Harbelot, C. Cruz, A Semantic analysis of moving objects using as a case study maritime voyages from eighteenth and nineteenth centuries, 6th International Conference on Advanced Geographic Information Systems, Applications, and Services (2014), 45-50.

[42] B. Harbelot, H. Arenas, C. Cruz, Using semantic web technologies to follow the evolution of entities in time and space, International J. On Advances in Intelligent Systems 6 (2013) 256-265.

[43] B. Harbelot, H. Arenas, Helbert, C. Cruz, LC3: A spatio-temporal and semantic model for knowledge discovery from geospatial datasets, Web Semantics: Science, Services and Agents on the World Wide Web 35 (2015) 3-24.

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

[45] B. Zhong, C. Gan, H. Luo, X. Xing, Ontology-based framework for building environmental monitoring and compliance checking under BIM environment, Building and Environment 141 (2018) 127-142.

[46] G. Hévízi, M. Biczó, B. Póczos, Z. Szabó, B. Takics, A. Lorincz, Hidden Markov model finds behavioral patterns of users working with

a headmouse driven writing tool, IEEE International Joint Conference on Neural Networks (2004) 669-674.

[47] T. Wang, M. Qiao, Y. Deng, Y. Zhou, H. Wang, Q. Lyu, H. Snoussi, Abnormal event detection based on analysis of movement information of video sequence, Optik-International Journal for Light and Electron Optics 152 (2018) 50-60.

[48] Z. Du, Y. Yang, Q. Cai, C. Zhang, Y. Bai, Modeling and inferring mobile phone users' negative emotion spreading in social networks, Future Generation Computer Systems 78 (2018) 933-942.

[49] M. Pastell, L. Frondelius, A Hidden Markov Model to estimate the feeding time of dairy cows based on indoor positioning data, bioRxiv (2018) 1-4.

[50] B. Dong, S. Biswas, Meal-time and duration monitoring using wearable sensors, Biomedical Signal Processing and Control 32 (2017) 97-109.

[51] A. Postawka, Real-time monitoring system for potentially dangerous activities detection, IEEE 22nd International Conference on Methods and Models in Automation and Robotics (MMAR) (2017) 1005-1008.

[52] C. A. Ronao, S. B. Cho, Recognizing human activities from smartphone sensors using hierarchical continuous Hidden Markov Models, International J. of Distributed Sensor Networks 13(1) (2017) 1-16.

[53] M. L. Williams, W. P. James, M. T. Rose, Fixed-time data segmentation and behavior classification of pasture-based cattle: Enhancing performance using a Hidden Markov Model, Computers and Electronics in Agriculture 142 (2017) 585-596.

[54] A. P. Ulmeanu, V. S. Barbu, V. Tanasiev, A. Badea, Hidden Markov Models revealing the household thermal profiling from smart meter data, Energy and Buildings 154 (2017) 127-140.

[55] T. T. Zin, I. Kobayashi, P. Tin, H. Hama, A general video surveillance framework for animal behavior analysis, Third International Conference on Computing Measurement Control and Sensor Network (CMCSN) (2016) 130-133.

[56] S. H. Ryu, H. J. Moon, Development of an occupancy prediction model using indoor environmental data based on machine learning techniques, Building and Environment 107 (2016) 1-9.

[57] L. Costa, P. Trigueiros, A. Cunha, Automatic meal intake monitoring using Hidden Markov Models, Procedia Computer Science 100 (2016) 110-117.

[58] J. Chaney, E. H. Owens, A. D. Peacock, An evidence based approach to determining residential occupancy and its role in demand response management, Energy and Buildings 125 (2016) 254-266.

[59] T. Šabata, T. Borovicka, M. Holena, Modeling and clustering the behavior of animals using Hidden Markov Models, CEUR Workshop Proceedings 1649 (2016) 172–178.

[60] J. Liisberg, J. K. Møller, H. Bloem, J. Cipriano, G. Mor, H. Madsen, Hidden Markov Models for indirect classification of occupant behaviour, Sustainable Cities and Society 27 (2016) 83-98.

[61] E. Wu, P. Zhang, T. Lu, H. Gu, N. Gu, Behavior prediction using an improved Hidden Markov Model to support people with disabilities in smart homes, IEEE 20th International Conference on Computer Supported Cooperative Work in Design (CSCWD) (2016) 560-565.

[62] A. Farzad, R. N. Asli, Recognition and classification of human behavior in Intelligent surveillance systems using Hidden Markov Model, International J. of Image Graphics and Signal Processing 7(12) (2015) 31-38.

[63] H. Jiang, P. Lin, Q. Fan, M. Qiang, Real-time safety risk assessment based on a real-time location system for hydropower construction sites, The Scientific World Journal (2014) 1-14.

[64] A. R. M. Forkan, I. Khalil, Z. Tari, S. Foufou, A. Bouras, A context-aware approach for long-term behavioural change detection and abnormality prediction in ambient assisted living, Pattern Recognition 48(3) (2015) 628-641.

[65] L.Yang, K. Ting, M.B. Srivastava, Inferring occupancy from opportunistically available sensor data, IEEE International Conference on Pervasive Computing and Communications (PerCom) (2014) 60-68.

[66] A. Babakura, M. N. Sulaiman, N. Mustapha, T. Perumal, HMM-based decision model for smart home environment, International J. of Smart Home 8(1) 129-138. (2014)

[67] J. Wang, R. Chen, X. Sun, M.F. She, Y. Wu, Recognizing human daily activities from accelerometer signal, Procedia Engineering 15 (2011) 1780-1786.

[68] H. Jeong, A. Gupta, R. Roscoe, J. Wagster, G. Biswas, D. Schwartz, Using Hidden Markov Models to characterize student behaviors in learning-by-teaching environments, International conference on intelligent tutoring systems, Springer Berlin Heidelberg (2008) 614-625.

[69] P.C. Chung, C.D. Liu, A daily behavior enabled Hidden Markov Model for human behavior understanding, Pattern Recognition 41(5) (2008) 1572-1580.

[70] N. T. Nguyen, D. Q. Phung, S. Venkatesh, H. Bui, Learning and detecting activities from movement trajectories using the hierarchical Hidden Markov Model, IEEE Computer Society Conference on Computer Vision and Pattern Recognition (2005) 955-960.

[71] R. Vanlande, C. Nicolle, C. Cruz, IFC and Buildings Lifecycle Management, Automation in Construction 18(1) (2008) 70-78.

[72] Y. Peng, J. R. Lin, J. P. Zhang, Z. Z. Hu, A hybrid data mining approach on BIM-based building operation and maintenance, Building and Environment 126 (2017) 483-495.

[73] J. Chen, T. Bulbul, J.E. Taylor, G. Olgun, A case study of embedding real-time infrastructure sensor data to BIM, In Construction Research Congress: Construction in a Global Network (2014) 269-278.

[74] M. Röck, A. Hollberg, G. Habert, A. Passer, LCA and BIM: Visualization of environmental potentials in building construction at early design stages. Building and Environment 140 (2018) 153-161.

[75] Autodesk. (2018). [online] Available at: http://paulaubin.com/_downloads/2017_AU/BIM128338-Aubin-AU2017.pdf [Accessed 13 Sep. 2018].

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

[77] Z. Zhu, M. W. Park, C. Koch, M. Soltani, A. Hammad, K. Davari, Predicting movements of onsite workers and mobile equipment for enhancing construction site safety, Automation in Construction 68 (2016) 95-101.

[78] A. Carbonari, A. Giretti, B. Naticchia, A proactive system for real-time safety management in construction sites, Automation in Construction 20(6) (2011) 686-698.

[79] A. A. Oloufa, M. Ikeda, H. Oda, Situational awareness of construction equipment using GPS, wireless and web technologies, Automation in Construction 12(6) (2003) 737-748.

[80] T. Cheng, J. Teizer, Real-time resource location data collection and visualization technology for construction safety and activity monitoring applications, Automation in Construction 34 (2013) 3-15.

[81] S. Sun, J. Zhao, Q. Gao, Modeling and recognizing human trajectories with beta process hidden Markov models, Pattern Recognition 48(8) (2015) 2407-2417.

[82] Z. Dashdorj, S. Sobolevsky, S. Lee, C. Ratti, Deriving human activity from geo-located data by ontological and statistical reasoning, Knowledge-Based Systems 143(1) (2018) 225-235.

[83] C. Nicolle, C. Cruz, Semantic building information model and multimedia for facility management, International Conference on Web Information Systems and Technologies Springer, Berlin, Heidelberg (2010) 14-29.

[84] Z. Dashdorj, S. Sobolevsky, S. Lee, C. Ratti, Deriving human activity from geo-located data by ontological and statistical reasoning, Knowledge-Based Systems. 143(1) (2018) 225-235.

[85] S. Harkema, A. Behrman, H. Barbeau, Chapter 16 - Evidence-based therapy for recovery of function after spinal cord injury, Handbook of Clinical Neurology Elsevier 109, (2012).

[86] C. F. Karney, Algorithms for geodesics, Journal of Geodesy 87(1) (2013) 43-55.

[87] M. Abramowitz, I. A. Stegun, Handbook of Mathematical Functions. New York: Dover. Chapter 6: Gamma and Related Functions (1972).

[88] R. Gatto, S. R. Jammalamadaka, The generalized von Mises distribution. Statistical Methodology 4(3) (2007) 341-353.

[89] H. Akaike, Information theory and an extension of the maximum likelihood principle, Springer Series in Statistics (Perspectives in Statistics) Springer, New York (1998) 199-213.

[90] J. Liisberg, J. K. Møller, H. Bloem, J. Cipriano, G. Mor, H. Madsen, Hidden Markov Models for indirect classification of occupant behaviour, Sustainable Cities and Society 27 (2016) 83-98.

[91] W. Zucchini, An Introduction to Model Selection, J. Math. Psychol. 44(1) (2000), 41-61.

[92] H. Bozdogan, Model Selection and Akaike's Information Criterion (AIC): The general theory and its analytical extensions, Psychometrika 52(3) (1987) 345-370.

[93] A. V. Metcalfe, P. S. Cowpertwait, Introductory time series with R. (2009)