Missouri State University

BearWorks

MSU Graduate Theses

Spring 2017

Coupling Mobile Technology, Position Data Mining, and Attitude toward Risk to Improve Construction Site Safety

Khandakar Mamunur Rashid

As with any intellectual project, the content and views expressed in this thesis may be considered objectionable by some readers. However, this student-scholar's work has been judged to have academic value by the student's thesis committee members trained in the discipline. The content and views expressed in this thesis are those of the student-scholar and are not endorsed by Missouri State University, its Graduate College, or its employees.

Follow this and additional works at: https://bearworks.missouristate.edu/theses

 Part of the Business Administration, Management, and Operations Commons

Recommended Citation

Rashid, Khandakar Mamunur, "Coupling Mobile Technology, Position Data Mining, and Attitude toward Risk to Improve Construction Site Safety" (2017). *MSU Graduate Theses*. 3158.
https://bearworks.missouristate.edu/theses/3158

This article or document was made available through BearWorks, the institutional repository of Missouri State University. The work contained in it may be protected by copyright and require permission of the copyright holder for reuse or redistribution.
For more information, please contact BearWorks@library.missouristate.edu.

COUPLING MOBILE TECHNOLOGY, POSITION DATA MINING, AND ATTITUDE TOWARD RISK TO IMPROVE CONSTRUCTION SITE SAFETY

A Master's Thesis

Presented to

The Graduate College of

Missouri State University

In Partial Fulfillment

Of the Requirements for the Degree

Master of Science, Project Management

By

Khandakar M. Rashid

May 2017

Copyright 2017 by Khandakar Mamunur Rashid

COUPLING MOBILE TECHNOLOGY, POSITION DATA MINING, AND ATTITUDE TOWARD RISK TO IMPROVE CONSTRUCTION SITE SAFETY

Technology and Construction Management

Missouri State University, May 2017

Master of Science

Khandakar M. Rashid

ABSTRACT

Construction sites comprise constantly moving heterogeneous resources that interact in close proximity of each other. The sporadic nature of such interactions creates an accident prone physical space surrounding workers. Despite efforts to improve site safety using location-aware proximity sensing techniques, major scientific gaps still remain in reliably forecasting impending hazardous scenarios before they occur. In the research documented in this thesis, spatiotemporal data of workers and site hazards are fused with a quantifiable model of an individual's attitude toward risk to generate proximity-based safety alerts in real time. In particular, two trajectory prediction models, namely polynomial regression (PR) and hidden Markov model (HMM) are investigated and their effectiveness in predicting a worker's position given his or her past movement trajectory is evaluated. Next, HMM prediction is further improved and calibrated by factoring in a worker's risk profile, a measure of his affinity for or aversion to risky behavior near hazards. Finally, a mobile application is designed and tested in a series of field experiments involving trajectories of different shape and complexity to verify the applicability and value of the designed methodology in addressing construction safety-related problems. Results demonstrate that the developed risk-calibrated HMM-based motion trajectory prediction can reliably detect unsafe movements and impending collision events.

KEYWORDS: construction safety, trajectory prediction, real-time tracking, GPS, Markov model, risk attitude.

This abstract is approved as to form and content

_____

Amir H. Behzadan, PhD
Chairperson, Advisory Committee
Missouri State University

COUPLING MOBILE TECHNOLOGY, POSITION DATA MINING, AND ATTITUDE TOWARD RISK TO IMPROVE CONSTRUCTION SITE SAFETY

By

Khandakar M. Rashid

A Master's Thesis
Submitted to the Graduate College
Of Missouri State University
In Partial Fulfillment of the Requirements
For the Degree of Master of Science, Project Management

May 2017

Approved:

_____
Amir H. Behzadan, PhD

_____
Nebil Buyurgan, PhD

_____
Richard J. Gebken, PhD

_____
Hui Lui, PhD

_____
Julie Masterson, PhD: Dean, Graduate College

ACKNOWLEDGEMENTS

The present document is my final Master's Thesis. This Thesis describes my graduate research which began in Spring 2016 and concluded in Spring 2017. I would like to acknowledge my sincerest gratefulness to my advisor, Dr. Amir H. Behzadan for his guidance, endless patience, and perennial support and encouragement during my Master's studies and in the course of this research. I am also very thankful to Dr. Nebil Buyurgan, Dr. Richard J. Gebken, and Dr. Hui Lui for being members of my Thesis Committee. I thankfully acknowledge the technical assistance of Nipun D. Nath and Raiful Hasan. I would also like to appreciate the presence of Ananna Ahmed, a wonderful friend and a great source of motivation and emotional support during my hard days. I am deeply grateful to my parents, Khandakar Harunur Rashid and Shayla Rashid. Their love and patience have been always a source of support for me, particularly during the days I am far away from them.

I dedicate this thesis to my mother, Shayla Rashid, who always understands what I do not say, my father, Khandakar Harunur Rashid, who inspires me to keep moving in life and my brother, Khandakar Aminur Rashid, who always keeps surprising me.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

INTRODUCTION

Thesis Problem Statement

The construction industry is considered one of the most hazardous industries worldwide in terms of fatalities, serious injuries, lost work time, hospitalization and disability (Spangenberg et al. 2005; Hannerz et al. 2005; Wadick 2007; Loosemore and Andonakis 2007). Despite ongoing research and strict enforcement of regulatory systems and standards in occupational safety and health, this problem still persists. In the research presented in this Thesis, a systematic scientific methodology is laid out to help improve the safety conditions in the construction industry through proper implementation of proactive techniques that take advantage of the vast amount of site data captured and processed continuously by latest data collection, analysis, and storage technologies. The collected data, if properly used, can provide valuable insight into spatial interaction patterns between construction resources, which can in turn significantly enhance our understanding of and ability to predict similar future events, and ultimately preempt hazardous incidents from occurring. The use of such modern ubiquitous data sensing and mining methods can therefore ensure a universal and timely deployment of effective safety practices on the jobsite.

Research Motivation

Injury Statistics in Construction Industry. Inherent to the construction industry are high accident rates and hazardous activities that have resulted the industry to rank as a very dangerous industry worldwide (Sacks et al. 2009). In the U.S. alone, more than 17%

of all work-related deaths are related to construction (Tixier et al. 2016). According to the Bureau of Labor Statistics (BLS), approximately 925 fatal injuries happened in 2015 in the construction industry, and approximately $15 billion of revenue is lost each year due to construction injuries and fatalities (BLS 2016). BLS reported that there were approximately 2.9 million nonfatal workplace injuries and illnesses in the private sector in 2015, which occurred at a rate of 3 cases per 100 equivalent full-time workers. Since its establishment in 1971, the Occupational Safety and Health Administration (OSHA 2016) has aimed at creating safe working environments by enforcing safety regulations. While these regulations have resulted in an overall positive trend in workplace safety (Figure 1), the number of injuries and accidents is still very high. OSHA identifies fall, electrocution, struck by object, and caught in between as four major causes of construction injuries, named as "fatal four". As shown in Figure 2, among these four causes, fall, struck by object, and caught in between directly or indirectly relate to proximity of construction resources, and contribute to almost 51% of all construction-related fatalities.

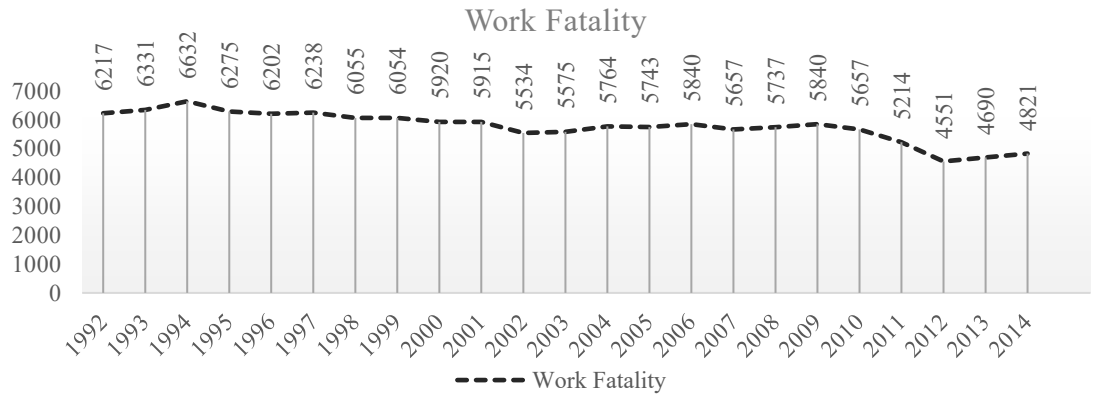


Figure 1: Number of fatal injuries 1992-2014 (BLS, 2016)

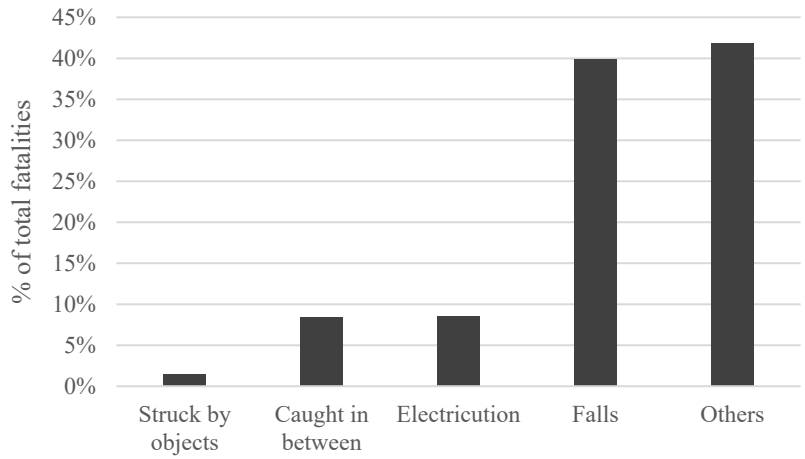


Figure 2: Four primary causes of construction worker fatalities (OSHA, 2016)

Hinze and Teizer (2011) conducted a research categorizing injuries and fatalities caused by vision or lack of visibility. Results showed that out of 659 equipment and visibility related fatalities, 521 cases were due to struck by moving equipment (Figure 3). Other factors included hit by equipment buckets, material being dropped or lowered by equipment, electrocution when equipment contacted power lines, and rollovers when equipment were operated on a steep slope. Another issue explored in the same study was the direction of move of a piece of equipment at the time an incident occurred. Results indicated that out of 594 equipment related incidents, 72.6% cases occurred when the equipment was travelling in reverse direction while only 18.5% cases resulted from equipment traveling forward (Figure 4). This study demonstrates a strong correlation between workplace accidents and proximity of construction resources.

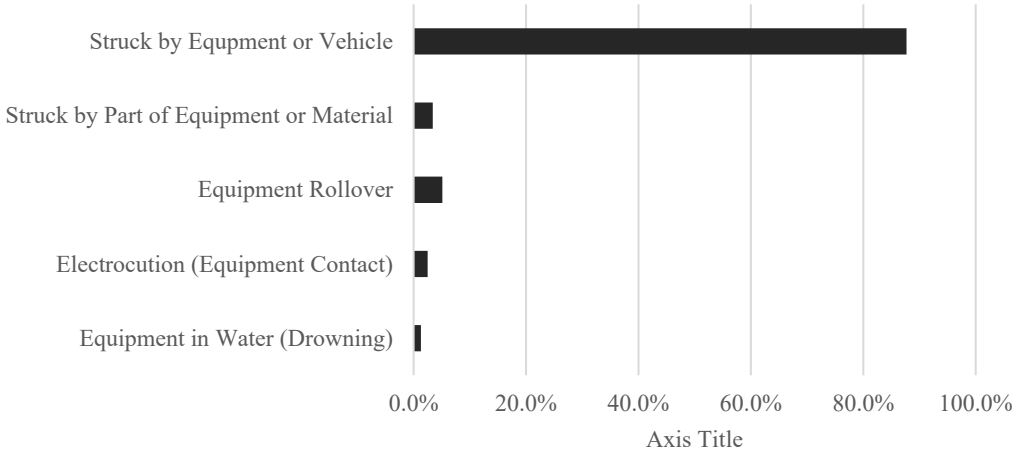


Figure 3: Distribution of construction fatalities due to visibility/awareness (N = 594)

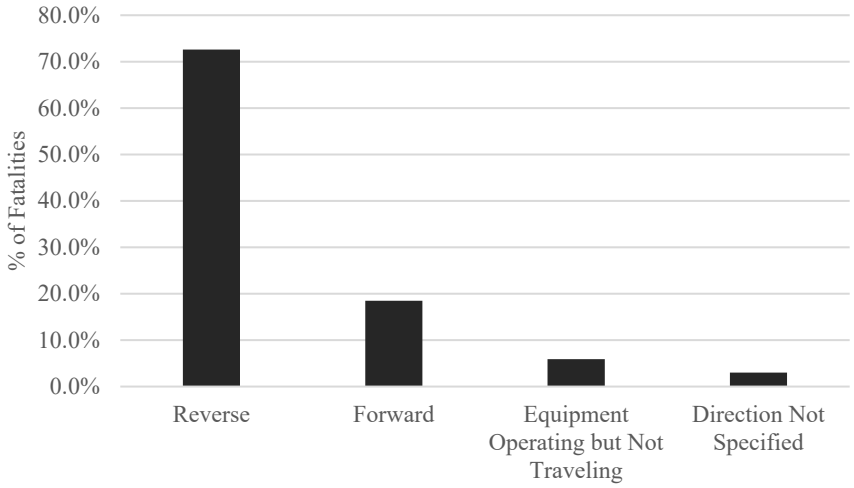


Figure 4: Equipment movement direction when accident occurs (N = 594)

There is significant cost associated with construction injuries and fatalities. In particular, the total cost of fatal and non-fatal injuries in the construction industry was estimated to be $11.5 billion in 2002 which was 15% of all injury and fatality costs in the private sector (Weahrer et al. 2007). Moreover, it was estimated that in 2002, each fatal or non-fatal construction injury cost an average of $27,000, compared to $15,000 in other

industries (Weahrer et al. 2008). Another study showed that between 2011 and 2013, the annual economic cost of construction-related fatalities was approximately $270 million in Illinois, $150 million in Indiana, and $125 million in Iowa (ILEPI 2015).

Existing Safety Practices. The main difference between construction and manufacturing environments or assembly lines is that almost no construction project occurs in a stationary work setting, since resources constantly move and frequently interact with one another in an unstructured manner. This makes addressing safety issues in construction even more challenging. Existing construction safety management practices are traditionally carried out in a fragmented manner (Benjaorana and Bhokhab 2010). To this end, it is worth noting that quite often the main focus of construction management which is productivity improvement (i.e. lower product cost and shorter completion time) is in clear contrast with workplace safety requirements (Benjaorana and Bhokhab 2010). Several researchers have proposed different approaches to integrate safety in construction design, planning and control (Navon and Kolton 2006; Hare et al. 2006). Ideally, safety measures must be taken into consideration from the design phase where designers can play an important role by implementing safer designs and directing the choice of construction means and methods to avoid or reduce hazardous situations on the jobsite (Benjaorana and Bhokhab 2010). However, due to the unpredictable and dynamic nature of construction field activities, it is challenging for designers to foresee each and every hazardous situation before the construction process begins. Following the design phase, the next step in a project lifecycle where safety precautions must be practiced is construction. This is normally done by checking and enforcing common industry safety regulations such as those of OSHA (Zhang et al. 2013). Previous research

has indicated that although complying with OSHA regulations contributes to an overall safe jobsite, such rules and regulations may not be adequate to avoid all incidents involving physical contacts between objects (a.k.a. contact collisions) (Teizer et al. 2010). The main underlying reason behind this is that OSHA mainly enforces the use of passive safety devices (e.g. hard hats, safety shoes, goggles, face shields, reflective clothing, hearing protection, wet weather gear, and filter masks) also known as Personal Protective Equipment (PPE), which are not capable of providing any kind of warning before a collision happens. The lack of education and experience in safety management has been identified as a major cause for many incidents (Le et al. 2015). Typically, knowledge about safety is conveyed through textbooks, specialized training, apprenticeship programs, and job experience (Gambatese 2003). These safety programs deliver information about site risks, hazards, and safe behaviors. For example, OSHA offers a 30-hour voluntary outreach class for personnel with supervisory authority over workplace safety and health, aiming to educate them about standards, procedures, policies with special emphasis on recognition, avoidance, abatement, and prevention of workplace safety hazards (Hardison et al. 2014). It has been stated that existing safety training programs are often not sufficiently engaging, offered within a very short period of time, and do not take advantage of active workers' participation (Le et al. 2015). In addition, very often, construction accidents happen due to errors, negligence, omissions, and misunderstandings of workers (Reese and Eidson 2006).

Technology in Construction Safety. As previously mentioned, information delivery methods deployed in traditional safety education may not actively engage workers (Le et al. 2015), and thus, researchers have tried to integrate emerging

technologies such as virtual reality (VR), augmented reality (AR), and mobile computing platforms to improve the effectiveness of safety teaching methods. Virtual environments have great potential both in terms of safety learning and teaching (Jamaludin 2009). A field safety training method based on VR for a steel erection site was presented by Abraham and Irizarry (2005). Also, Lin et al. (2011) used a 3D game environment to conduct a pilot study for construction safety education. In contrast to VR, AR enhances the real world with synthetic (computer-generated) data, essentially allowing real and virtual worlds to coexist in an augmented environment (Azuma 1997). Within the domain of safety training, Behzadan and Kamat (2011) developed an AR-based modeling environment to help students understand construction processes and operational safety. Le et al. (2015) presented a framework that uses mobile based-VR and AR technology for experiential safety education. Sampaio et al. (2010) demonstrated the importance of introducing computer-aided design (CAD), 3D modeling, and VR technology in architecture, engineering and construction (AEC) education for undergraduate students. Shirazi and Behzadan (2015) implemented and assessed the effectiveness of AR-based pedagogical tools to enhance the learning experience for construction students.

Within the same context, information and communication technologies such as building information modeling (BIM), virtual design and construction (VDC), along with geographical information system (GIS) are emerging tools in AEC that can facilitate the integration of safety measures in design, planning, and monitoring of field activities. For example, Zhang et al. (2013) presented a rule-based BIM-enabled engine to automatically analyze a building model, detect fall-related safety hazards, and provide preventive suggestions to the user. Hadikusumo and Rowlinson (2004) created a VR-based design-

for-safety process database which takes input from the building design phase and identifies safety hazards. An integrated system for construction safety management based on 4D CAD models and a rule-based algorithm was developed by Benjaoran and Bhokha (2010), which integrates safety measures at the early stages of design and planning to help all parties prepare for safety constraints the before actual work begins. Kang et al. (2011) proposed a 5D CAD system by combining 4D objects from progress schedule data with the risk data using analytical hierarchy process (AHP) analysis to visualize the risk level of each activity. Bansal (2011) developed a platform using GIS-based navigable 3D animation, linking information from the project schedule with the safety recommendation database to predict the places and activities having higher risk potentials. Video camera and time-lapsed photography is also used frequently to measure the overall safety conditions of construction sites and identify potential violations of regulations by workers or contractors (Bohn and Teizer 2009).

During the past several years, and with the invention of more robust sensing technologies, researchers have also studies the feasibility of real time proactive proximity safety warning systems for construction workers. For instance, Teizer et al. (2010) conducted an experiment in a realistic construction environment using a radio frequency (RF) system which gives audio-visual alerts to workers and equipment operators when they come to close proximity. Dinga et al. (2013) presented a safety management tool based on internet of things (IoT) which integrates fiber Bragg grating (FBG) sensors and a radio frequency identification (RFID) for labor tracking. The experiment was done in an underground tunnel construction site in China, and results indicated an improvement in real time monitoring, detecting, and warning of safety risks. Another research based on

location-aware technologies that combined wireless communication, global positioning system (GPS), and GIS, showed a significant potential of real time safety warning technology by automatically detecting the hazard, alerting drivers to avoid collision, and ultimately ensuring reliable navigation ofs construction equipment (Wu et al. 2013).

Automated Construction Resource Tracking. As stated earlier, contact collisions are the major cause of construction injuries and fatalities. In order to assess and understand the underlying characteristics of this kind of incidents, obtaining context-aware information including time-stamped positional data of construction resources is critical (Behzadan et al. 2008). However, despite recent developments in construction measurement and sensing technologies, collecting precise and timely location data from construction resources still remains a rather challenging task (Saidi et al. 2003). To address this challenge, several studies have investigated methods of automatically tracking resources (personnel, equipment, materials) in construction and facilities projects. In particular, several technologies for indoor and outdoor location tracking and remote sensing have been used. For example, the feasibility of automatically measuring labor input by tracking their position was explored by Navon and Goldschmidt (2002 and 2003). Using this approach, the time each worker spent on an activity was estimated with an accuracy of 10%-20%. The developed prototype labor control model, based on workers' GPS location was further integrated with a building project model (BPM) to compare labor performance by calculating the expected labor inputs based on planner labor rates and the work quantities required to complete each task. Sacks (2003) and Navon (2004) used GPS technology to track earthmoving equipment in regular intervals and convert location data into equipment productivity and material consumption. Other

researchers also used GPS to track construction equipment and materials (Hildreth et al. 2005; Caldas et al. 2005; Lu et al. 2007; Behzadan et al. 2008; Pradhananga and Teizer 2013). RFID is another location tracking technology used for tracking construction resources. In an early attempt, RFID was used to track high-value material on construction jobsites (Jaselskis, 1995). Song (2006) used RFID to track the delivery and receipt of fabricated pipe spools. Goodrum et al. (2006) developed a prototype tracking system to monitor hand tools in a mobile environment. In another research, RFID and GPS technologies were integrated to track precast concrete components in a storage yard thus minimizing labor input for locating these components (Ergen et al. 2007). Teizer et al. (2007) deployed ultra-wide band (UWB) technology as a data collection tool of real time location sensing and resource tracking for construction work zone safety and material tracking. Automated tracking has been also used in a number of indoor applications. For instance, in an indoor environment, where global navigation satellite system (GNSS) data is not available, indoor positioning technologies can be used. Indoor GPS, wireless local area network (WLAN), inertial navigation system (INS), Bluetooth, infrared, and ultrasonic are other available technologies for indoor tracking, and several applications have been developed based on these technologies (Behzadan et al. 2008; Khoury and Kamat 2009; Razavi and Moselhi 2011).

As the number of mobile phone users has been steadily growing with almost 2 billion smartphone users in the market by late 2015 (Kissonergis 2015), researchers in multiple disciplines have also directed their efforts toward utilizing a host of mobile embedded sensors (e.g. GPS, accelerometer, gyroscope, digital compass). In an early study, an android-based indoor/outdoor localization system was developed, taking

advantage of GPS and WiFi modules, to locate personnel carrying smartphones (Pereira 2011). Recently, the development of GPS and location-aware applications have gone beyond simply navigating through a route or obtaining the location of the phone. For instance, GPS data collected from cellular phones were used to gather large volumes of traffic information, process the collected data, and distribute it back to the phone users in real time for traffic analysis and monitoring (Handel 2014). In 2008, a traffic probing field campaign, known as Mobile Century field experiment, was carried out which involved 100 private cars carrying GPS-enabled Nokia N95 smartphones to collect real time measurements every third second while the vehicles repeatedly drove on 610-mile loops continuously for 8 hours on a freeway in the San Francisco Bay area (Herrera 2010). Another project, Mobile Millennium, was launched in 2008 with more than 2000 registered users, which demonstrated that infrastructure road data collection is feasible using cellular GPS (Jasper 2011). Bierlaire et al. (2010) proposed a model of route choice behavior from smartphone GPS data by developing a probabilistic path generation algorithm to generate a set of potential true paths. In another study, smartphone-based sensor fusion (accelerometer, gyroscope, magnetometer, GPS, video) was used to detect and record drivers' actions in order to recognize driving patterns (Johnson and Mohan 2011). Also, a smartphone-based navigation application was developed which alerts a visually-impaired pedestrian by an audible message at decision points prior to his or her arrival at a work zone (Chen-Fu 2014).

The application of smartphone-based location-aware technologies in healthcare is also explored in several studies. For instance, a generic android-based framework was developed to collect field data by epidemiologists and ecologists with the help of a web

database and GPS data (Aanensen 2009). The potential of location-aware smartphone technologies to help the elderly and visually-disabled persons was also described in several studies (Boulos et al. 2011).

Worker Behavior Analysis. As evidenced in the literature and described above, within the construction domain, extensive research has been carried out to improve the safety environment for field workers. In particular to the topic of this Thesis, several studies have focused on ways to reduce contact collisions between workers and equipment by developing real time automated safety alert systems. However, one area which is not yet fully investigated in the construction safety domain is the influence of a worker's attitude toward risk in how he or she performs in the vicinity of site hazards. The interface of safety and human behavior has been the subject of previous studies. For instance, Salminen (2004) conducted a survey study that showed young workers under the age of 25 experience a higher injury rate than older workers. Gardner and Steinberg (2005) conducted an experimental study with 306 participants in three age groups – adolescents (13-16), youth (16-22), and adults (24 and older) to measure risk preference and risky decision making, and concluded that risk-taking and risky decision making decrease with age. Cooper (2003) stated that the risk-taking propensity of an individual depends on his or her perception of the situation, past experience, and personality. Gender differences in risk-taking attitude is also studied in several projects. For example, Charness and Gneezy (2012) showed that women are generally more risk averse in financial issues than men. Their study demonstrated that women make smaller investments in a risky asset than men do.

Technology	Research and Activity Level
GPS	Sacks et al. (2003): labor tracking Navon and Goldschmidt (2003): labor tracking Navon et al. (2004): equipment tracking Navon and Shpatnitsky (2005): equipment tracking Caldas et al. (2005): material tracking Hildreth et al. (2005): equipment tracking. Behzadan et al. (2008):user contextual data Song and Eldin (2012): heavy equipment tracking Pradhananga and Teizer (2013): labor and equipment tracking Zhang et al. (2015): labor tracking
RFID	Jaselskis et al. (1995): concrete supply, cost coding, material control Jaselskis and El-Misalami (2003): materials tracking Goodrum et al. (2005): small tool tracking Song et al. (2006): material tracking Chae and Yoshida (2010): heavy equipment tracking Montaser and Moselhi (2014): labor and material tracking Hubbard et al. (2015): labor and material tracking
UWB	Teizer et al. (2007): work zone safety and materials tracking Khoury and Kamat (2009): project information retrieval Cheng et al. (2011): labor, equipment and material tracking Yang et al. (2011): labor tracking Hwang (2012): equipment tracking
WLAN	Khoury and Kamat (2009): project information retrieval Behzadan et al. (2008): user contextual data Woo et al. (2011): labor tracking

Risk-taking propensity of men has been observed to be higher than women in other domains. For example, according to the U.S. Department of Transportation (USDOT 2004), male drivers in are three times more likely to be involved in fatal car accidents. Also, it has been reported that female drivers use seat belts substantially more often than men (Waldron et al. 2005). The same study showed that in the U.K., men

pedestrians are 80% more likely to be involved in accidents than female pedestrians, and men die much often from accidental poisoning or drowning than women (Waldron et al. 2005).

Trajectory Prediction. Trajectory prediction is a critical component of almost all spatial collision algorithms (Gong and McNally 2004). Researchers have studied a variety of trajectory prediction techniques in several fields such as robotics (Bennewitz et al. 2005), aerospace engineering (Gong and McNally 2004), maritime traffic management (Perera et al. 2012), physics and mechanics (Choi and Hebert 2006), and meteorology (Kim et al. 2015). Gong and McNally (2004) presented a methodology based on statistical analysis to improve the quality of trajectory prediction for decision support applications such as conflict detection and arrival metering for air traffic management. Perera et al. (2012) presented a methodology integrating intelligent features with vessel traffic monitoring and information system (VTMIS) to predict navigational vessel trajectory using extended Kalman filter (EKF) (Tanizaki 1996). One of the first attempts to collect global system for mobile communication (GSM) data was developed by Laasonen et al. (2004) who proposed a prediction model which took a sequence of recent cell transitions to find the most probable cell the user will enter next. Ashbrook and Starner (2003) used Markov model (Petrushin 2000) to predict a user's next location from his or her significant past locations extracted from GPS data. Mathew et al. (2012) designed a hybrid method to predict human mobility by training a hidden Markov model (HMM) using historical location clusters. Vasquez and Fraichard (2004) proposed a technique that learns the pattern of a moving object and applies a pairwise clustering algorithm to clustered trajectories to predict that object's future position. A hybrid

prediction model, coupling historical trajectory patterns and an object's recent motion was also explored (Jeung et al. 2008) and demonstrated accurate results than existing prediction models at that time. Monreale et al. (2009) proposed a trajectory pattern tree to predict the next location of a moving object using GPS data with a certain level of accuracy. Gambs et al. (2012) used the extended mobility Markov chain (MMC) theory to predict the next location of an individual using his previously visited locations. Kim et al. (2015) presented a destination prediction framework which detects a user's location via k-nearest neighbor (kNN) and decision trees, and predicts his or her future destination using HMM.

As will be discussed later in this Thesis, an effective and robust trajectory prediction method is of significant value to any construction safety alert system as it enables the prediction of future positions of workers and equipment given their immediate past movement patterns. When coupled with a reliable location tracking method, trajectory prediction can preempt potential contact collisions on the jobsite.

Research Contribution

As statistics show, the construction industry has one of the most hazard- and risk-prone working environments of all industries (Tixier et al. 2016). Research has revealed that within this environment, one of the major underlying reasons of jobsite accidents is contact collision. Arguably, the most hazardous encounters occur when two or more construction resources come within close proximity overlooking potential safety risks. Existing safety measures mainly emphasize on enforcing industry standards and OSHA regulations, which are mostly passive techniques and do not provide advance warnings

prior to an accident taking place. A real time proactive safety warning approach is thus necessary which can track the location of construction resources and generate safety alerts before resources get too close to each other. As previously described, several technologies such as GPS, WLAN, RFID, and UWB have been used by researchers to collect positional data of construction resources. However, the practical advantages of mobile wearable sensors (e.g. smartphones) for robust position tracking and to provide location-aware information for safety purposes has not been yet fully investigated in the construction industry. Some of the benefits of the mobile technology that makes it a viable area of research is that smartphone sensors do not rely on any preinstalled infrastructure in the construction site, are ubiquitous (almost everybody owns and knows how to operate a smartphone) and cost effective (as no extra device needs to be purchased), and cause minimum (if any) distraction and discomfort to the crew.

Although the mere position tracking of construction resources and giving timely proximity alerts could potentially improve current jobsite safety practices, the integration of trajectory prediction of moving resources can immensely benefit the same cause by enabling the prediction of future moving trajectory of an object based on its latest movement patterns. Trajectory prediction has been explored in different scientific and engineering disciplines. As previously stated, most trajectory prediction algorithms use clustering techniques, where prediction is made by analyzing the discovered historical motion patterns. However, cluster-based prediction models need large datasets and run computationally intensive data mining processes, and are generally not useful when there is no distinguishable motion pattern. In a construction jobsite, where resources move

randomly (do not follow any particular patterns) and frequently interact with each other, a slightly different trajectory prediction technique may yield better results.

In addition, most of existing studies on proactive construction safety focus on the context-awareness problem as related to construction resources. The attitude of individuals toward risk and safety is another important factor that must be also explored and properly integrated with context-awareness to further improve the safety environment on construction sites. As previously described, studies in different domains show a direct correlation between safety incident rates and one's risk-taking behavior with his or her age, gender, and experience level.

Considering the current state of research and practice in construction safety, and the abovementioned limitations in the body of knowledge, the work presented in this Thesis tries to investigate the value of mobile location-aware technologies to construction safety from a new perspective, by explaining the thought process, design, and implementation of a mobile safety alert framework which uses a combination of worker's positional data (obtained from built-in smartphone sensors), a robust trajectory prediction technique, and worker's attitude toward risk to detect imminent hazardous encounters and generate preemptive warnings before a contact collision occurs.

Research Objectives and Project Tasks

The overall objective of this research is to the design and implement a mobile safety alert system which uses a combination of worker's positional data (obtained from built-in smartphone sensors), a robust trajectory prediction technique, and a worker's risk profile to generate timely warning alerts before a contact collision occurs. In order to

achieve this objective, the following research tasks were identified and successfully completed:

- Collect historical trajectory data and design trajectory prediction models.
- Perform comparative analysis of various trajectory prediction models.
- Formulate and incorporate risk attitude into trajectory prediction, conduct several pilot tests, and perform precision, recall, and accuracy analyses.
- Design and develop a mobile application, conduct real time field experiments, and asses the feasibility of the designed system to address real-life safety situations.

Organization of the Thesis

The following Chapters of this Thesis are shaped around the concepts, details, and implementation of the research tasks listed above. This Thesis is divided into six Chapters. In particular:

- Introduction – This Chapter contains the thesis problem statement, identified research gaps that motivated this research, the novel approach that this study adopts to address the identified gaps, and the overall objective and tasks defined and accomplished in this project.
- Trajectory Prediction Models – This Chapter includes different trajectory prediction models, their applications and limitations, and a more in-depth analysis of two major prediction models adopted in this study.
- Machine Learning and Trajectory Prediction – This Chapter includes the literature review of hidden Markov model, its applications in different fields, and mathematical details of the designed HMM in this research to enhance construction safety.
- Incorporating Risk Attitude into Trajectory Prediction – This Chapter covers the incorporation of risk attitude in the previously developed HMM-based prediction model and the analysis of various pilot tests.
- Preemptive Construction Site Safety (PCS2) Application – This Chapter includes details and technical descriptions of the software architecture, and user interface

design of an Android-based mobile safety application. In addition, results from field experiments conducted using this application are presented.

- Conclusions and Future Work – This Chapter summarizes the key findings of the research presented in this Thesis, provides a thought provoking discussion of the identified gaps in knowledge and practice of construction safety, and describes how the developed research methodology in this research helps address these gaps. In addition, some potential directions of future work in this area are presented.

TRAJECTORY PREDICTION MODELS

Previous Work

Trajectory and motion prediction has been a key research topic in various fields such as robotics (Bennewitz et al. 2005), aerospace engineering (Gong and McNally 2004), maritime traffic management (Perera et al. 2012), meteorology (Kim et al. 2015), automation in manufacturing (Mainprice and Berenson 2013), traffic engineering (Houenou et al. 2013), and autonomous vehicles (Glaser et al. 2010). This Section summarizes previous research in trajectory prediction involving human and/or human-robot interactions.

In the domain of intelligent transportation systems, the development of self-driving vehicles has led to an increasing need for automated collision avoidance systems (CASs) that can predict the trajectories of neighboring vehicles to detect possible impending collision events. Houenou et al. (2013) proposed two trajectory prediction models for self-driving cars. The first model is based on constant yaw rate and acceleration (CYRA) which takes into account the instantaneous velocity and yaw rate of the car and is proven to be accurate for short term prediction. The second model presented by Houenou et al. (2013) combines the current maneuver recognition, road geometry, and CYRA model in order to provide accurate long term predictions. Another trajectory planning model for autonomous vehicles was presented by Glaser et al. (2009) who generated several preliminary trajectories using 4^{th} and 5^{th} order polynomial regression. Each trajectory was then further evaluated and ranked using a number of performance indicators such as risk, speed, consumption, comfort, and compliance with

driving rules. Finally, a weighted sum of all performance indicators was used to select the best trajectory. Persad-Maharaj et al. (2008) proposed an algorithm based on spatially-aware, geometric representations of user's historical trip data from a global positioning system (GPS)-enabled phone to predict an individual's real time travel path and destination in order to display customized advertising alerts upon entering a specific area. Mainprice and Berenson (2013) presented a framework allowing humans and robots to work simultaneously in a close proximity specifically in automated manufacturing settings. This framework was based on gesture recognition and workspace occupancy prediction by learned motion trajectories.

In this Thesis, two trajectory prediction models are adopted, validated, and compared for prediction accuracy: polynomial regression (PR) and hidden Markov model (HMM). As later explained, the model with more robustness and better accuracy will be further enhanced by integrating an individual's risk factor. The PR model is described in detail in the following Section.

Polynomial Regression (PR) Model

Construction ground crews often move in random patterns, especially in the absences of paved surfaces or marked paths. Therefore, to formulate movement patterns, a heuristic PR-based trajectory prediction technique is initially developed and tested. In particular, positional data obtained in the past s time intervals are used to predict an object's position in time $s+1$ (immediate future). Selecting the right value for s thus is an important aspect of developing this technique, as a large s may result in an unnecessarily complicated PR model, while a small s may yield a low-accuracy simplistic model. In

addition to selecting the most appropriate value for *s*, finding the right polynomial degree *n* is also of the essence. A high value for *n* can result in overfitting, while a low *n* value can lead to underfitting.

A hypothetical trajectory of a worker's movement is shown in Figure 5, where by visually inspecting the pattern of positional data, the first portion of the worker's past movements (labeled as segment 1) seems not to significantly contribute to his or her predicted future position. To address this and similar situations, only the last 60 seconds of a worker's movement are considered to predict its future position. The last 60-second segment is subsequently divided into four equal time frames (*T*) (last 15 seconds, last 30 seconds, last 45 seconds, and last 60 seconds) for calculating the best possible motion trajectory.

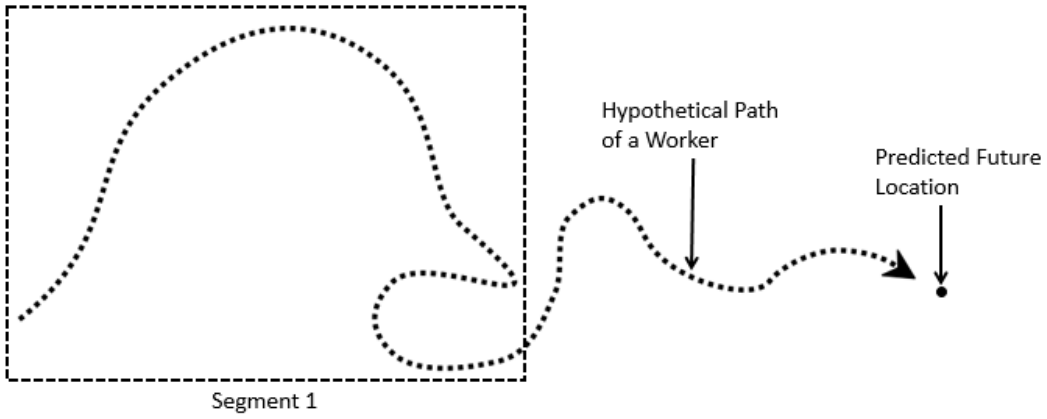


Figure 5**.** Hypothetical motion trajectory of a construction worker

As stated above, the second design aspect of the regression model is choosing the most proper value for the polynomial degree, or *n*. Equation (1) is the general polynomial formulation used in this research. In this Equation, *x* is the time frame (last 15 seconds,

last 30 seconds, last 45 seconds, and last 60 seconds) and y is the latitude or longitude of the worker's future position. In each iteration, for time stamp t, multiple polynomial equations are generated using combinations of T for time stamp t-1 and the polynomial degree, n (1, 2, 3, 4, 5), with possibility of testing higher values, as necessary. Initial predictions are made for time t using the generated polynomial equation.

$$y = a_0 + a_1 x + a_2 x^2 + \cdots + a_n x^n \qquad \text{Eq. (1)}$$

As shown in the process flow of Figure 6, calculated (predicted) values for y are then compared to the actual (collected) values at time t, and an absolute discrepancy factor (Δ) is calculated using Equation (2), and as illustrated in Figure 7. The combination of T and n that yields the minimum Δ is ultimately selected for use in the PR model to predict the latitude and longitude at time $t+l$ (l is the prediction horizon, or prediction time in advance). Results indicate that Δ tends to be larger for higher values of n. This can be attributed to the fact that predicted latitude and longitude values are expressed as functions of time. So, for a higher degree polynomial, a small error in prediction results in a significantly large Δ. For this reason, a linear regression model (n = 1) is ultimately selected to minimize Δ in the developed PR model.

$$\Delta = \left| y_{predicted} - y_{calculated} \right| \qquad \text{Eq. (2)}$$

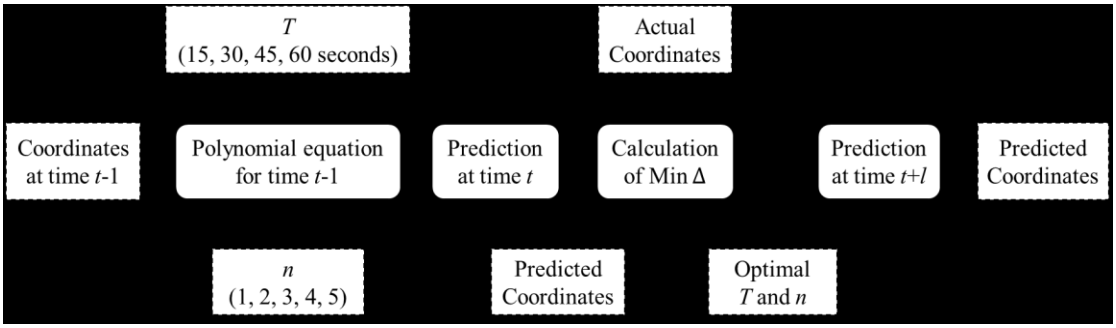


Figure 6. Input-process-output diagram of PR model at time *t*

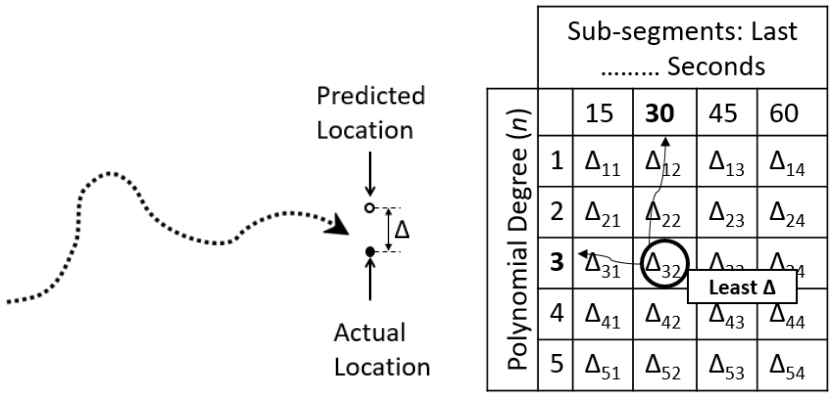


Figure 7. Selecting the best combination of *T* and *n* to minimize discrepancy (Δ)

Preliminary Evaluation of the PR Model

In order to validate the developed PR model, coordinates (latitude and longitude) of two separate trajectories (i.e. trajectory 1 and trajectory 2) are collected using an android-based application (GPS Logger) launched on an LG Nexus 5X mobile device. The frequency of the data is set at 1Hz (one coordinate data point per second). Before incorporating the collected data into the model, an instrument error analysis is conducted to calculate the absolute global error of the device's sensor. As shown in Figure 8, five benchmark locations are selected and their global coordinates (latitude and longitude) are captured and recorded by the device once a 3D fix is achieved (i.e. the device starts

communicating with at least 3 satellites). These coordinates are then compared with the absolute coordinates of the same points as obtained from accurate global maps (e.g. Google Map), and corresponding errors are calculated.

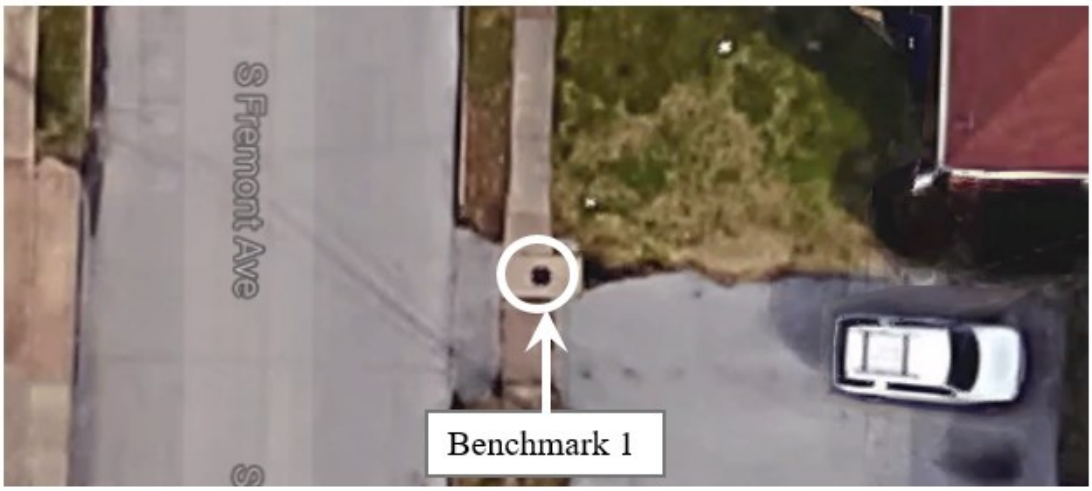


Figure 8. Sample benchmark selected for smartphone's GPS error calculation

Equation (3) (Haversine distance formula) is used to calculate error values. In this Equation, Θ_1 and Θ_2 are latitudes values of two points, k_1 and k_2 are longitudes values of the same two points, and R_E is the Earth radius. As Table 1 shows, average, minimum, and maximum errors for the five selected benchmarks are 3.63m, 2.17m, and 4.73m, respectively.

$$Distance = 2R_E\sqrt{sin^2\left(\frac{\Theta_2-\Theta_1}{2}\right) + cos\Theta_1 * cos\Theta_2 * sin^2\left(\frac{k_2-k_1}{2}\right)} \quad \text{Eq. (3)}$$

Table 1. Error calculation of Smartphone's (LG Nexus 5X) GPS sensor

Benchmark ID	From Google Map		From Smartphone		Absolute Error(m)
	Latitude	Longitude	Latitude	Longitude	
1	37.205550	-93.271045	37.205565	-93.271095	4.732
2	37.204928	-93.269155	37.204944	-93.269169	2.169
3	37.204923	-93.269637	37.20495	-93.269656	3.442
4	37.204948	-93.262435	37.204946	-93.262471	3.196
5	37.205407	-93.262353	37.205412	-93.262405	4.639

Trajectories 1 and 2 consist of 180 and 360 coordinate data points, respectively. These two trajectories represent two different levels of complexity and randomness. As shown in Figure 9, trajectory 1 represents a simple straightforward movement with no sharp turns or sudden changes in direction, while trajectory 2 contains several sharp turns and represents a more complex motion path. The approximate entropy (ApEn) analysis is conducted to measure the complexity and unpredictability for each trajectory. Entropy can be defined as a loss of information in a time series. ApEn is a measure of regularity to quantify the level of complexity within a time series and has been utilized in postural control, physical activity measurement, and human walking data analysis (Yentes et al. 2013). ApEn measures the "likelihood that runs of patterns that are close remain close on next incremental comparisons" (Pincus 1991). ApEn calculates the logarithmic likelihood that the next interval of the data will differ from the current interval. As a benchmark, a straight-line trajectory results in an ApEn value of zero. A smaller value of ApEn represents greater likelihood that the similar patterns of the trajectory will follow (i.e. simple trajectory). With the same token, a highly irregular trajectory is not likely to

follow similar patterns and will result a relatively high ApEn. To calculate the ApEn, the change of direction of every timestamp for both trajectories is initially calculated. As the change of velocity is not significant in human walking time series, it is safely assumed to be constant. Choosing a 5-second interval, and considering the change of direction as a time series data, the value of ApEn for trajectories 1 and 2 are calculated as 0.12 and 0.27, respectively.

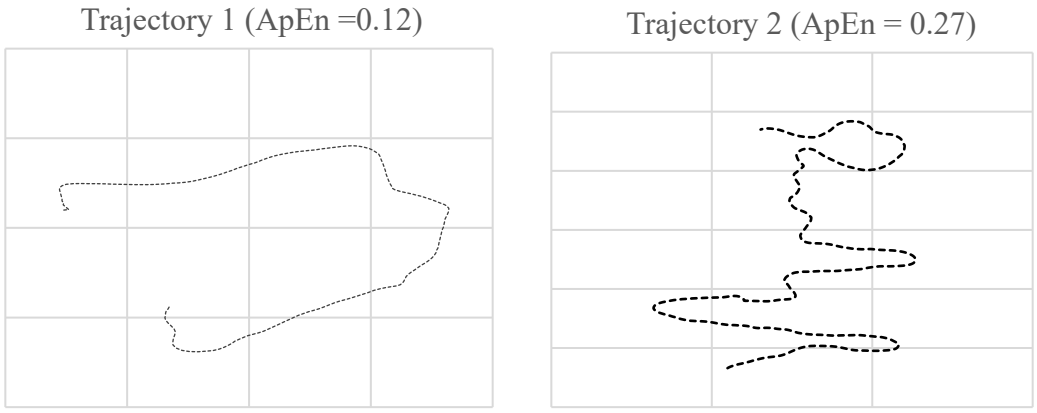


Figure 9. Trajectories 1 and 2 for preliminary validation of the PR model

Initially, both trajectories are used to predict future positions of the moving object using the PR model. The prediction horizon ranges from 1 second to 10 seconds, which implies that predictions are made as early as 10 and as late as 1 seconds in advance. The prediction error is then computed by calculating the linear distance between the predicted location and actual location for that time frame using the Haversine formula (Equation 3). It must be noted that since the minimum time frame is 15 seconds (i.e. the prediction is made using at least 15 data points collected in the last 15 seconds), at least 15 GPS data

points is required to start the PR prediction process. Figure 10 demonstrates the 95-percentile prediction error for both trajectories. It can be seen from this Figure that for both trajectories, the value of error increases with the prediction horizon. For example, the prediction error at 10 seconds in advance is higher than the error corresponding to a 5-second prediction horizon for both trajectories. The error diagrams closely resemble an exponential shape because the predicted latitude and longitude values are expressed as a function of time, and a higher value of prediction horizon results in a significantly larger error. Figure 10 also shows that the prediction made by the PR model is more accurate for less complex trajectories. For example, the 95-percentile error for a 5-second prediction horizon is 8.75m for trajectory 1, while it is 13.55m for trajectory 2.

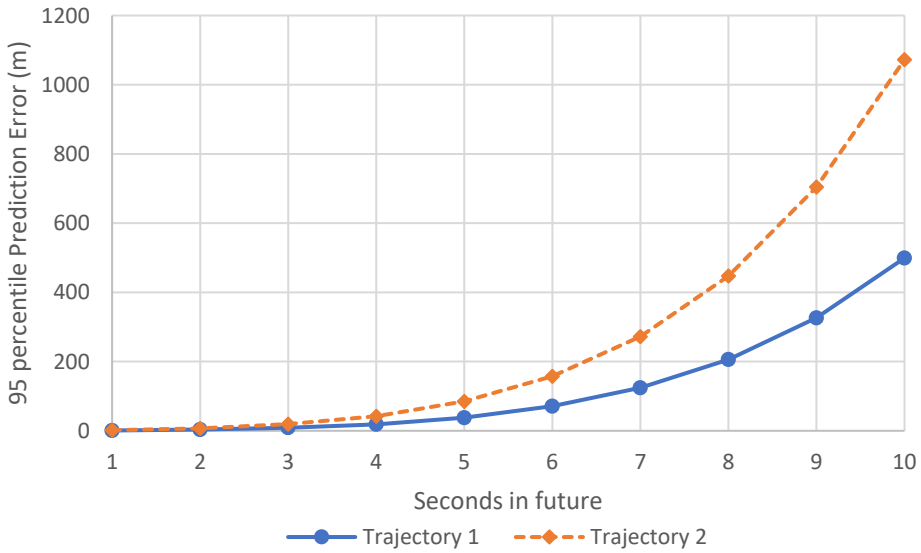
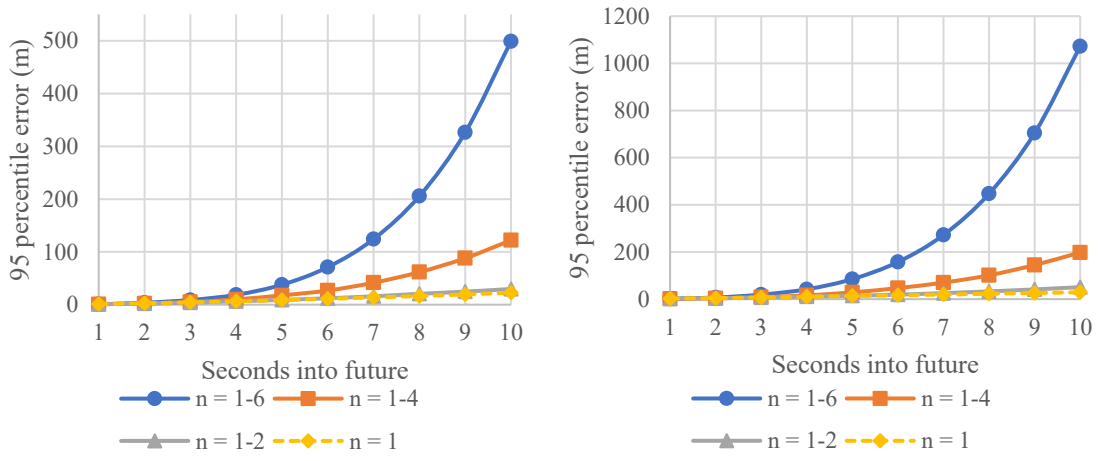


Figure 10. 95-percentile prediction error for trajectories 1 and 2

Next, different values of n are used in both trajectories to assess how the choice of polynomial degree affects the accuracy of the PR model. Figure 11 shows that a linear PR model ($n = 1$) yields a higher accuracy compared to nonlinear models ($n > 1$). For example, as shown in Figure 11(a), the 95-percentile error for trajectory 1 is the highest when n ranges from 1 to 6, and lowest when n is equal to 1. Also, it is observed that the shape of the error curve changes from exponential to (almost) linear as n decreases. Since latitude and longitude values vary over time, small changes in predicted latitude and longitude result in significantly larger errors for higher value of n. For example, for a prediction horizon of 5 seconds, while the 95-percentile error for trajectory 1 is 37.79m when n varies between 1 and 6, a linear PR ($n = 1$) leads to an error of only 8.75m.



(a) (b)

Figure 11. 95-percentile prediction error for different values of n for (a) trajectory 1 and (b) trajectory 2

Upon establishing the fact that a linear PR model results in the least error (i.e. highest possible accuracy), actual locations and 5-second advance predicted locations are plotted for both trajectories 1 and 2 using $n = 1$. Results are illustrated in Figure 12 and Figure 13. It can be seen from these Figures that the prediction error increases locally when there is a turn or a sudden change in direction. As an example, for trajectory 1 (Figure 12) the two highlighted sections demonstrate significant changes in direction. Consequently, the 95-percentile error for these two sections of trajectory 1 is 12.88m while the overall 95-percentile error for trajectory 1 is only 8.75m.

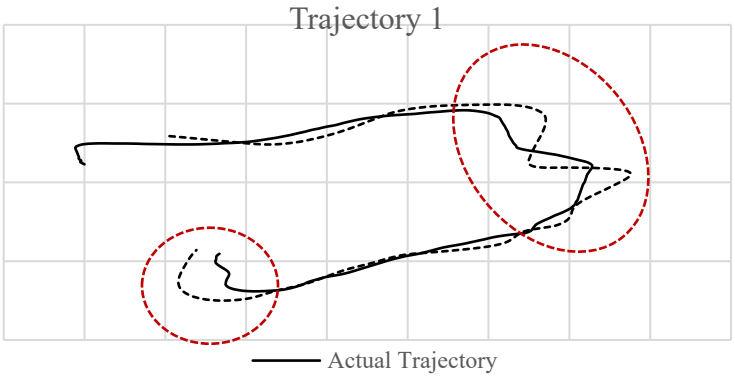


Figure 12. Actual vs. predicted trajectories (trajectory 1)

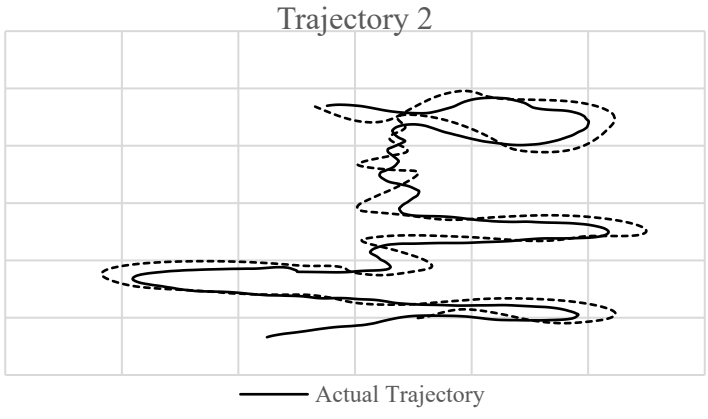


Figure 13. Actual vs predicted trajectories (trajectory 2)

Summary and Conclusions

Trajectory prediction has been an important field of research in many scientific areas. This Chapter explored polynomial regression (PR) as a trajectory prediction model to obtain the immediate future locations of construction workers on the jobsite given their past positional data collected by mobile sensors carried on their bodies. Prior to using the collected GPS data from each smartphone, an instrument error analysis was performed. Next, two test trajectories (one simple and one more irregular) were collected from the field and the developed PR model was applied to the collected data to predict future positions. Results indicated that a linear PR model is more accurate than nonlinear PR models. Also, it was observed that the prediction error increases locally when there is a turn or a sudden directional change in the trajectory.

MACHINE LEARNING AND TRJECTORY PREDICTION

The Problem of Context-Aware Information Delivery

Providing field personnel and construction crew with context-aware information about their surroundings can significantly improve their efficiency by saving the time needed to manually search for data and/or analyze work packages and tasks. To this end, the ability to predict the immediate future needs of a user, or foresee and preempt potential operational problems in advance can even further enhance this proposition. Among several prediction methods, those that rely on user's position (i.e. physical location on the jobsite) to provide context-aware information are generally referred to as geospatial information prediction techniques. When effectively coupled with mobility features, these methods can be used to gather contextual information of mobile users, process that information as necessary, and apply the results to new (yet contextually similar) settings. The most distinctive advantage of mobile context-awareness is its proactivity and robustness in predicting the future state of a user. A review of literature reveals that predicting the physical location of a constantly moving user or object is an inherently interesting and challenging problem in many domains and for different applications (Kononenko 2001, Joachims 2002; Peters et al. 2003; Lustreck and Kauluza 2009; Collobert et al. 2011; Nath et al. 2017).

One of the most widely used classes of context-aware prediction techniques in science and engineering is machine learning (ML). In a broad sense, ML is the field of study of algorithms that self-improve their efficiency automatically through obtaining new knowledge and experience. For this reason, ML is often categorized as a self-

adaptive method. ML algorithms have been extensively used in different fields of study including but not limited to search engines (Joachims 2002), medical diagnosis (Kononenko 2001), robotics (Peters et al. 2003), activity recognition (Lustreck and Kauluza 2009), ergonomics analysis (Nath et al. 2017) and natural language processing (Collobert et al. 2011). A precursor to ML is data mining (DM) which aims at discovering hidden patterns (a.k.a. features) in large datasets. ML algorithms can be generally grouped into supervised and unsupervised learning methods. Examples of supervised learning include classification and regression, while clustering, and association of rule discovery are examples of unsupervised learning methods.

Anagnostopoulos et al. (2007) proposed a ML-based contextual model that used two classification approached (call-based and cluster-based) to exploit the user's spatial and spatiotemporal data for predicting future movements. The methodology was validated with a 10-fold cross validation and results showed that the cluster-based approach was more accurate that the call-based approach. In another effort, a mixed Markov-chain model (MMM) was proposed to predict the future movement of pedestrian (Asahara et al. 2011). This model took into account the pedestrian's personality as an unobservable parameter in addition to the transition probability of simple Markov chain (MC). The result demonstrated 74.4% accuracy for MMM over 45% for simple MC. Mathew et al. (2012) presented a hybrid HMM to predict human mobility by clustering discrete historical location data and training the model for each cluster. A real-world experiment was conducted and in terms of geospatial distance between the true location of the user and predicted geospatial coordinates, the best result showed an average distance of 143.5 Kilometers and a median distance of 4.9 Kilometers. Vasquez and

Fraichard (2004) proposed a technique that learns the pattern of a moving object and applies a pairwise clustering algorithm to clustered trajectories to predict that object's future position.

This Chapter describes the design and implementation of a stochastic statistical trajectory prediction method based on HMM. The developed method uses classification and clustering of historical positional data in order to predict the future physical location of a mobile user (e.g. construction ground crew). The following Sections provide more detailed discussions about the underlying mathematical methods used in this research including the MC and HMM, followed by a thorough description of the developed trajectory prediction method.

Markov Chain (MC) and Hidden Markov Model (HMM)

A MC can be explained as a stochastic or random process with Markov property. If the conditional probability distribution of a future state of a stochastic process only depends on its present state, and not the sequence of states preceded it (a.k.a. limited horizon assumption), the process is called having a Markov property (Meyn and Tweedie 2012). The concept of MC is explained below through the use of a simple example.

Assume a large construction site with three work zones. Each day, the project manager visits the site and inspects one work zone randomly. In mathematical terms, this implies that the process of selecting and inspecting a work zone is a stochastic process that is not influenced by any external factor or the choice of the work zone in previous days. With the same token, this selection process may result in the project manager to visit the same work zone in consecutive days. Nevertheless, the project manager never

visits more than one work zone each day. Let's assume that in the last six days (indices -1 to -6) the sequence of work zones inspected by the project manager is $\{X_{-1} = z_2, X_{-2} = z_1, X_{-3} = z_1, X_{-4} = z_2, X_{-5} = z_2, X_{-6} = z_3\}$. Also, let's assume that at the present day, the project manager is visiting work zone 2; i.e. $X_0 = z_2$. Here, z_1, z_2, and z_3 represent work zone numbers, and X is a random variable describing the project manager's daily visit. According to the MC's limited horizon assumption, the probability of a work zone to be picked by the project manager for the tomorrow's visit (i.e. $X_1 = Z$) only depends on the project manager's today's choice of work zone (i.e. $X_0 = z_2$), and not on the project manager's past sequence of visits (i.e. $X_{-1} = z_2, X_{-2} = z_1, X_{-3} = z_1, X_{-4} = z_2, X_{-5} = z_2, X_{-6} = z_3$). Equation (4) formulates this statement,

$$P(X_1 = Z \mid X_0 = z_2, X_{-1} = z_2, X_{-2} = z_1, X_{-3} = z_1, X_{-4} = z_2, X_{-5} = z_2, X_{-6} = z_3) = P(X_1 = Z \mid X_0 = z_2) \quad \text{Eq. (4)}$$

A state diagram for this example is illustrated in Figure 14, in which a directed graph is used to show transitions between states. In this example, a state represents a work zone and therefore, is numbered 1, 2, or 3 at any given day. According to this Figure, zone 2 (z_2) is followed by zone 1 (z_1) 30% of the time, by zone 3 (z_3) 20% of the time, and by itself (z_2) 50% of the time. The transition matrix (A) for this example is shown in Equation (5) below,

$$A = \begin{array}{c} \\ z_1 \\ z_2 \\ z_3 \end{array} \begin{array}{c} \begin{array}{ccc} z_1 & z_2 & z_3 \end{array} \\ \begin{bmatrix} 0.1 & 0.2 & 0.7 \\ 0.3 & 0.5 & 0.2 \\ 0.3 & 0.3 & 0.4 \end{bmatrix} \end{array} \quad \text{Eq. (5)}$$

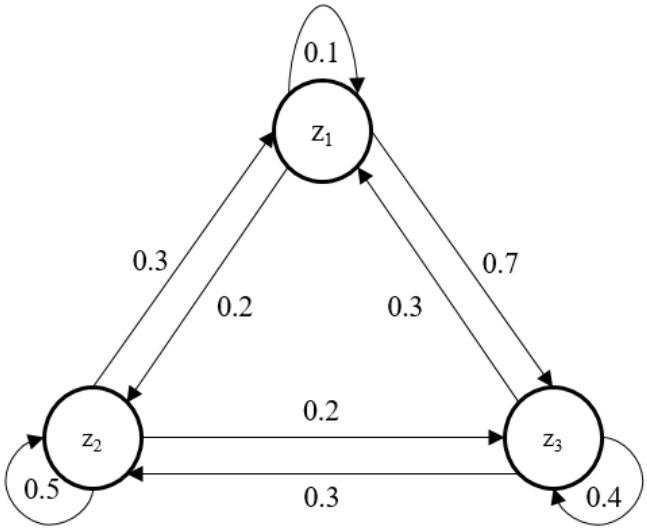


Figure 14. Directed graph representing the transition matrix

Now, let's assume that the project manager wants to determine the probability that she will be visiting work zone 2 in three days. This probability can be mathematically described by Equation (6),

$$P(X_3) = X_0 * A^3 \qquad \text{Eq. (6)}$$

Since the project manager has visited work zone 2 at the present day, X_0 can be described by the vector in Equation (7), where the probabilities of visiting work zones 1, 2, and 3 at the present day are 0%, 100%, and 0% respectively.

$$X_0 = [0 \quad 1 \quad 0] \qquad \text{Eq. (7)}$$

Now, let's slightly modify this example by installing video cameras on the jobsite to record site activities. This eliminates the need for a physical site visit by the project manager. Instead, a live video feed is transmitted to the project manager's office and she can observe each work zone from her office via the streaming video feeds. The previous

assumption still holds for the project manager only observes one work zone per day, and may even observe the same work zone in multiple consecutive days. At the end of each day, she prints out a card for each work zone to report on work progress in that particular zone, and submits all report cards through the company's online portal which is accessible to all employees. A report card can be blue or red. Clearly, unlike the project manager who can view the live video streams, other employees cannot observe the actual activities carried out in each work zone. The only way an employee can differentiate a work zone that was observed by the project manager from a work zone that was not, is by knowing that the color of the report card corresponding to each work zone is somehow correlated with whether that work zone was or was not observed by the project manager on a particular day. This situation can be conveniently modeled using a HMM.

A HMM is a Markov process that follows a Markov property, has unobserved (hidden) states, and provides correlated (and indirect) observations of those states. In the project manager example described above, the hidden states are the work zones (z_1, z_2, and z_3) that cannot be directly observed by employees, and the correlated outputs are the colors of the report cards (blue or red) that are correlated with the hidden states and are observable by employees. Each hidden state has a probability distribution over the observed outputs. Essentially, the sequence of the report card colors generated by the HMM gives information about the sequence of the states. Therefore, unlike MC, HMM has an observation probability parameter in addition to the transition probability. The term "hidden" in this context refers to the states of the process. Figure 15 demonstrates a directed graph showing the video feed observations made by the project manager as

hidden states, and the colors of the report cards as observed outputs together with their transition probabilities.

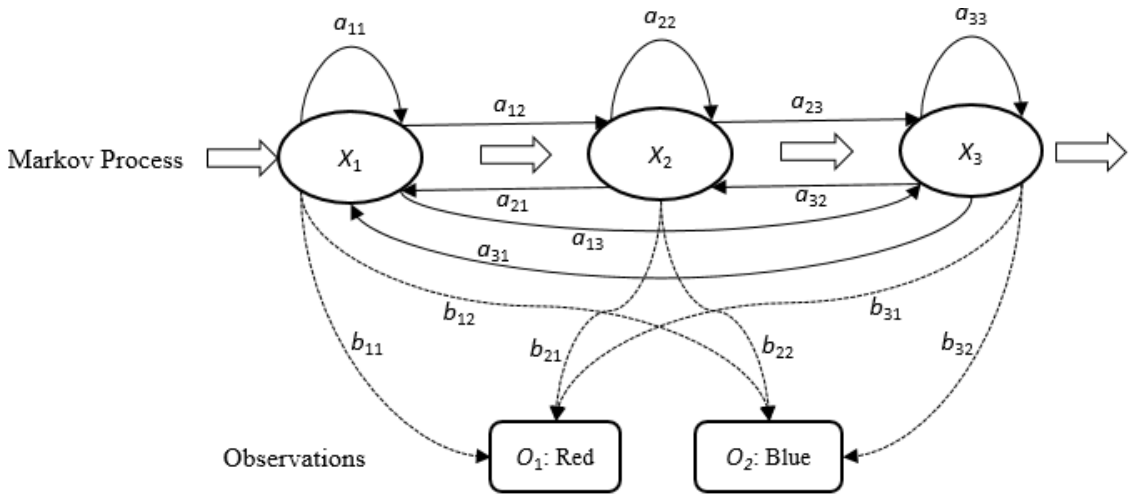


Figure 15. Probabilistic parameters of HMM

In Figure 15, *X* represents the hidden states (z_1, z_2, and z_3), *O* is the observations (red or blue), *a* is a transition probability (a_{ij} is the probability of transiting from state *i* to *j*) as formulated by Equation 8, and *b* is an observation probability (b_{jk} is the probability of observation *k* from hidden state *j*) as formulated by (Equation 9).

$$a_{ij} = P\big(state\ X_j\ at\ time\ t+1 \,\big|\, state\ X_i\ at\ time\ t\big) \quad \text{Eq. (8)}$$

$$b_{jk} = P(observation\ O_k\ at\ time\ t \mid state\ X_j\ at\ time\ t) \quad \text{Eq. (9)}$$

A HMM is denoted by defined by λ (Equation 10). In this Equation, *N* is the number of hidden states, *M* is the number of observations, *A* is the transition matrix, *B* is the observation matrix, and π is the initial state distribution. Matrices *A* and *B* are laid out

in Equation (11) and Equation (12), respectively (Stamp, 2004). In the project manager example described above, *N* is equal to 2 (two colors), *M* is equal to 3 (three work zones), and all work zones are equally likely to be selected for observation via video stream on any particular day. Therefore, the elements of π (initial state distribution) are equal to 1/3 (Equation 13).

$$\lambda = (N, M, A, B, \pi) \quad \text{Eq. (10)}$$

$$A = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix} \quad \text{Eq. (11)}$$

$$B = \begin{bmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \\ b_{31} & b_{32} \end{bmatrix} \quad \text{Eq. (12)}$$

$$\pi = [1/3 \quad 1/3 \quad 1/3] \quad \text{Eq. (13)}$$

When put together, the joint distribution of a sequence of states and observations can be described by Equation (14) (Ghahramani, 2001). In this Equation, the first term of the right-hand side represents π, the second term is the transition probability *A,* and the third term is the observation probability *B*. The joint HMM probability can be used to determine (predict) whether the project manager has observed a specific work zone given the color of the report card.

$$P\ (X_1, X_2 \ldots, X_N, O_1, O_2, \ldots, O_M) = P(O_1)[\prod_{n=2}^{N} P(X_n|X_{n-1})]\ \prod_{n=1}^{N} P(O_n|X_n)$$

Eq. (14)

Using a similar approach, a HMM-based trajectory prediction model is designed and implemented in this research. The developed model can predict the future position

(physical location) of a construction worker who is carrying a mobile device, given his or her past positions. The following Sections provide more insight into the designed trajectory prediction methodology.

Trajectory Prediction Using Hidden Markov Model (HMM)

The major shortcoming of the PR model described in previous Chapter is that for $n = 1$ (linear regression), it does not fully capture the randomness of the worker's trajectory, and for larger n values, the model shows high levels of instability in trajectory prediction. To overcome this problem, a trajectory prediction method based on HMM is designed and examined. In HMM, trajectories are treated as discrete stochastic processes (i.e. random walks). As shown in Figure 16, training trajectory data are first collected and stored in a trajectory database (DB). Next, statistical parameters are extracted from the dataset and a HMM it trained. New trajectory data is then collected from a target user (construction worker) and used as the input of the trained HMM to predict the worker's immediate future position. Clearly, since HMM is a trainable prediction method, with time and as more trajectory data come in, the model better adapts itself to the real-world movement patterns of construction workers and can provide more accurate predictions.

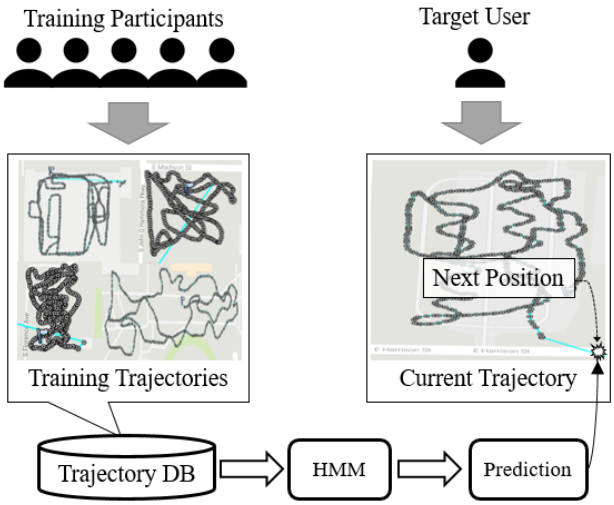


Figure 16. Main steps of the developed HMM-based trajectory prediction method

To initiate the training process, a trajectory is divided into a number of short trajectory sections, as shown in Figure 17. A group of short sections with common statistical features (e.g. mean, variance, and covariance) are bundled into one cluster which is represented by a single average section (a.k.a. latent segment) (Choi and Hebert 2006).

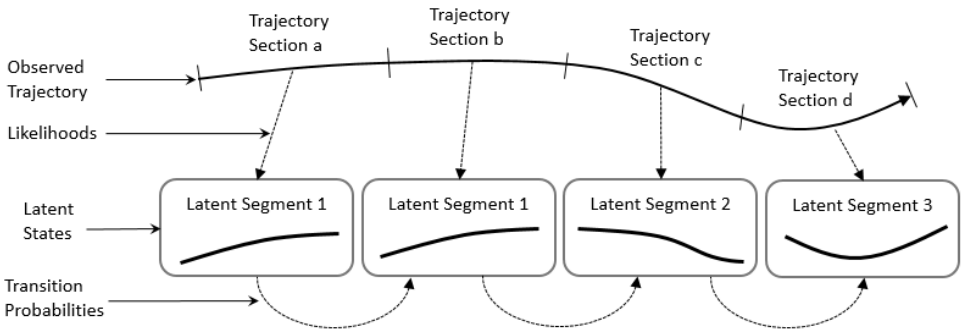


Figure 17. Key elements of the HMM prediction model

Considering the limited horizon assumption, which states that the probability of a future location depends only on the current location and not on the path by which the current location is achieved (Petrushin 2000), given a sequence of latent segments S_0, S_1, S_2, ….S_n, the probability of occurrence of a future latent segment S_{n+1} depends only on the current latent segment S_n, as stated in Equation (15). As previously stated, in HMM, these probabilities are termed transition probabilities and together create the transition matrix. Within a cluster of sections, the likelihood of a trajectory section to be generated from a specific latent segment is calculated from bivariate normal probability density function (pdf). These likelihood values are stored in the likelihood matrix. The HMM is trained to compute normalized trajectory sections, latent segments, and transition and likelihood matrices. The model first checks the likelihood matrix to find the latent segment that best resembles the observed trajectory section. Next, it determines the most probable future latent segment using the transition matrix, and finally provides the most likely trajectory section from the likelihood matrix.

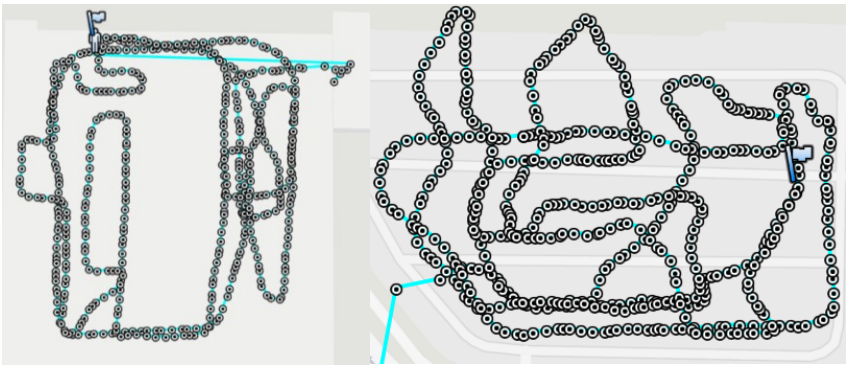
$$P(S_{n+1} \mid S_n, S_{n-1}, S_{n-2} \ldots\ldots S_0) = P(S_{n+1} \mid S_n) \qquad \text{Eq. (15)}$$

Training Data Collection and Processing

For the preliminary experiments conducted in this research, training data were collected from 26 participants. Each participant was asked to walk randomly for 15 minutes in an open field and log his/her GPS positional data recorded by a GPS data logger application launched on his/her smartphone device. This process was repeated three times (in three different locations) for each participant. The frequency of the recorded data was set at 1 Hz (one positional data per second). Individuals were allowed

to freely move and make any type of turns (90 degrees, U-turns) as they wished. The goal was to replicate the randomness of the walking paths of workers in a typical construction site. Training trajectories were stored in either text (.txt) or comma separated value (.csv) formats.

Figure 18 demonstrates samples of collected training data. Altogether, 78 positional datasets were collected from 26 participants each walking randomly for 15 minutes in 3 different locations. Before using the logged training data, the instrument error was checked to filter potential outliers. To this end, assuming an average human walking speed, the maximum possible distance between two consecutive positional data points is found to be 2.5m (Minetti 2000). While this value can be adjusted depending on the job type and work requirements, in this research, if the distance between two consecutive positional data points was more than 2.5m, the second data point would be automatically filtered out. Table 2 illustrates a portion of logged data that contains two adjacent data points with a distance of more than 2.5m. In this specific case, the linear distance between GPS coordinates collected in timestamps 20 and 21 is 5.67m, which is more than 2.5m. Thus, the second data point is deemed an instrument outlier and filtered out.



Item	Date	Elapse (s)	Latitude	Longitude	Altitude (m)	Heading	Speed (km/h)
1	45:38.0	0	37.197762°	-93.283498°	403.7	303.2	0
2	45:39.4	1.4	37.198040°	-93.283433°	403.2	106.4	81.5
3	45:41.1	3.1	37.198055°	-93.283343°	401.9	155.8	16.7
4	45:42.2	4.2	37.198076°	-93.283300°	402.4	173.2	15
5	45:44.1	6.1	37.198073°	-93.283278°	403.6	186.8	3.7
6	45:45.1	7.1	37.198071°	-93.283264°	403.9	155.5	4.5
7	45:47.1	9.1	37.198071°	-93.283243°	403.8	171.8	3.5
8	45:49.1	11.1	37.198057°	-93.283225°	403.7	190.7	4
9	45:50.1	12.1	37.198050°	-93.283219°	403.8	220.6	3.4
10	45:52.1	14.1	37.198034°	-93.283204°	403.9	173.1	4.1
11	45:53.1	15.1	37.198026°	-93.283196°	404	150.9	3.9
12	45:54.1	16.1	37.198019°	-93.283187°	403.9	147.2	3.9
13	45:56.1	18.1	37.198013°	-93.283165°	404.2	134.8	3.8
14	45:58.1	20.1	37.198014°	-93.283141°	404.1	136.3	3.9
15	45:59.1	21.1	37.198014°	-93.283133°	404.1	146.3	2.6

Figure 18. Snapshots of training data collection and logging

Table 2. Filtering instrument outlier from collected GPS training data

Time (s)	Latitude	Longitude	Distance (m)
…	…	…	…
…	…	…	…
19	37.208763	-93.274754	0.957
20	37.208776	-93.274756	0.906
21	37.20878	-93.274653	5.676
22	37.208789	-93.274757	1.132
23	37.208801	-93.274758	0.830
…	…	…	…
…	…	…	…

Learning Latent Segments

Each training trajectory dataset is divided into 12-second short sections. As shown in Figure 19, since trajectories have different start positions, directions, and velocities, they must be first normalized by a translation to the origin (0, 0), followed by a rotation so that the initial direction is (1, 0), and finally scaling so that the initial velocity is unit velocity. This results in a total of 4,662 normalized short trajectory sections extracted from the training data.

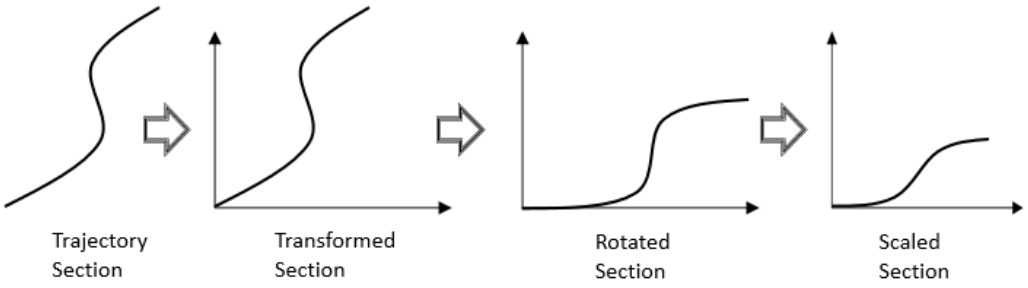


Figure 19. Normalization of trajectory sections

After normalizing, each short trajectory section is represented by a 22-dimensional vector consisting of x and y coordinates at 1 second intervals. Next, k-mean clustering (Choi and Hebert 2006) is applied to find statistically similar trajectory sections. k-mean clustering is an iterative process that groups data into k predetermined clusters by minimizing a cost function ζ, as shown in Equation (16). In other words, the goal is to minimize the sum of distance functions for each data point in a cluster relative to the center of that cluster. In Equation (16), C_j is the center of the jth cluster, and is the closest center to data point d_i, and n is the number of data points in the dataset. For

numerical datasets, a cluster center is characterized by the mean value of each attribute, the mean being calculated over all data points fitting to that cluster (Ahmad & Dey, 2007).

$$\zeta = \sum_{i=1}^{n} ||d_i - C_j||^2 \qquad \text{Eq. (16)}$$

For the training dataset collected in this research, eight clusters were found to best represent all short trajectory sections. Each cluster is described by a single latent segment the attributes of which are contained in a 5-by-12 characteristics matrix (*M*) as shown in Equation (17). In particular, five statistical features (means of *x* and *y* coordinates, variances of *x* and *y* coordinates, and covariance of *x, y* coordinate) are calculated for 12 data points (times $t_1, t_2, \ldots, t_{12}$) of all trajectory segments in a cluster. For time steps $t_1, t_2, \ldots, t_{12}$, the characteristics matrix (*M*) of latent segment *i* is given in Equation (5). Since the first two time steps of the trajectory segments are used as reference in the normalization process, the mean of *x* and *y* coordinates for the very first time step is $\binom{0}{0}$, and for the second tie step is $\binom{1}{0}$, and the variance for both time steps is zero. Therefore, eventually there are eight characteristics matrices for eight latent segments representing eight clusters identified by K-mean clustering.

$$M = \begin{bmatrix} \bar{x}_{it_1} & \bar{x}_{it_2} \ldots\ \ldots\ \ldots\ \ldots\ \ldots\ \ldots\ \bar{x}_{it_{12}} \\ \bar{y}_{it_1} & y_{it_2} \ldots\ \ldots\ \ldots\ \ldots\ \ldots\ \ldots\ \bar{y}_{it_{12}} \\ Var(x_{it_1}) & Var(x_{it_2}) \ldots\ldots\ldots\ldots\ldots Var(x_{it_{12}}) \\ Cov(x_{it_1}, y_{it_1}) & Cov(x_{it_2}, y_{it_2}) \ldots\ldots\ldots Cov(x_{it_{12}}, y_{it_{12}}) \\ Var(y_{it_1}) & Var(y_{it_2}) \ldots\ldots\ldots\ldots\ldots Var(y_{it_{12}}) \end{bmatrix} \qquad \text{Eq. (17)}$$

Training Markov Model for Trajectory Prediction

The overall training process of the HMM is demonstrated in Figure 20. This process involves calculating the transition probabilities between latent segments throughout all training trajectories. Transition probabilities are conditional probability distribution of a specific latent segment to be followed by other latent segments. As the sequence of short trajectory sections over the eight latent segments are known from training data, running all training trajectories through a HMM provides a probability matrix containing transition probability distribution of each latent state.

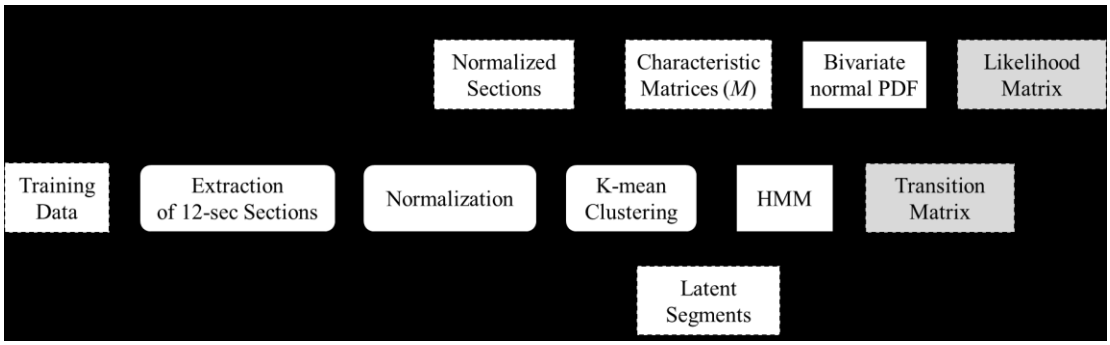


Figure 20. Input-process-output diagram of HMM training stage

As previously mentioned, these probabilities form an 8-by-8 transition matrix A. Equation (19) describes an element of A (i.e. a_{ij}) as a probability that a particular latent segment be followed by other latent segments. This process is implemented in MATLAB using the hmmestimate(seq,states)which returns the maximum likelihood estimate of transition probabilities of a HMM for sequence seq and known states (latent segments) states.

$$a_{ij} = P\big(latent\ segment\ S_j\ at\ time\ t+1 \,\big|\, latent\ segment\ S_i\ at\ time\ t\big) \qquad \text{Eq. (19)}$$

Next, the likelihood ($\mathcal{L}$) of normalized sections to be generated from a latent state is calculated by Equation (20), in which $\mathbb{N}$ is the bivariate normal pdf. This bivariate pdf can be defined using Equations (20) through (23) (Kotz et al. 2000). Here, μ is the mean, σ is the variance, and Σ is the covariance of the latent segments. The pdf of the bivariate normal distribution in MATLAB is implemented as mvnpdf(x,μ,Σ)which returns the density of the multivariate normal distribution with mean μ and covariance Σ. For the purpose of trajectory prediction, x is the vector containing latitude and longitude of a trajectory section, μ is the vector containing $\bar{x}$ and $\bar{y}$ of the characteristics matrix, and Σ is an identical 2-by-2 matrix containing covariance values on the diagonals and variances from the characteristics matrix.

$$\mathcal{L} = \prod_{n=t_1}^{t_{12}} \mathbb{N}\left(\begin{pmatrix} x_t \\ y_t \end{pmatrix} \mid \begin{pmatrix} \bar{x}_{it_n} \\ \bar{y}_{it_n} \end{pmatrix}, \begin{bmatrix} Var\left(x_{it_n}\right) & Cov\left(x_{it_n}, y_{it_n}\right) \\ Cov\left(x_{it_n}, y_{it_n}\right) & Var\left(y_{it_n}\right) \end{bmatrix}\right) \qquad \text{Eq. (20)}$$

$$P(x_1, x_2) = \frac{e^{-\frac{z}{2(1-\rho^2)}}}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} \qquad \text{Eq. (21)}$$

$$z = \frac{(x_1-\mu_1)^2}{{\sigma_1}^2} - \frac{2\rho(x_1-\mu_1)(x_2-\mu_2)}{\sigma_1\sigma_2} + \frac{(x_2-\mu_2)^2}{{\sigma_2}^2} \qquad \text{Eq. (22)}$$

$$\rho = Cor(x_1, x_2) = \frac{\Sigma_{12}}{\sigma_1\sigma_2} \qquad \text{Eq. (23)}$$

Considering the uniform prior distribution for latent states, from Bayes theorem, it can be said that posterior probability of a normalized trajectory section (l_i) generated from a latent state (S_i) is proportional to the likelihood of that section (l_i) generated from

the latent state (S_i) (Casella & Berger, 2002). So essentially, the latent state with the highest likelihood of generating a trajectory section is the one with the highest posterior probability, as mathematically described in Equation (24) and Equation (25). The likelihood of each trajectory section to be generated from a latent segment is calculated and stored in the likelihood matrix.

$$P(S_i \mid l_i) \propto \mathcal{L}(l_i \mid S_i)\, \mathrm{P}(S_i) \propto \mathcal{L}(l_i \mid S_i) \quad \text{Eq. (24)}$$

$$\mathrm{argmax}_i\, P(S_i \mid l_i) = \mathrm{argmax}_i\, \mathcal{L}(l_i \mid S_i) \quad \text{Eq. (25)}$$

Once the HMM is fully trained, the resulting transition matrix and the likelihood matrix are applied to future trajectory data for the purpose of trajectory prediction. In predicting the future location of an observed trajectory, at least 12 data points are required. As shown in Figure 21, the observed latest section (l_n), which contains 12 data points is first normalized. Next, the maximum likelihood of that section to be generated from a specific latent segment (S_n) is computed from the likelihood matrix. The latent segment with maximum likelihood is then used to compute the next most probable latent segment (S_{n+1}) from the transition matrix. Finally, the likelihood matrix is used to find the trajectory section which has the highest likelihood to be generated from that latent segment (S_{n+1}). The trajectory section is then denormalized and used as the predicted future trajectory (l_{n+1}). Since the first two points of l_{n+1} are patched to the existing trajectory, the HMM model can predict up to 10 seconds in advance.

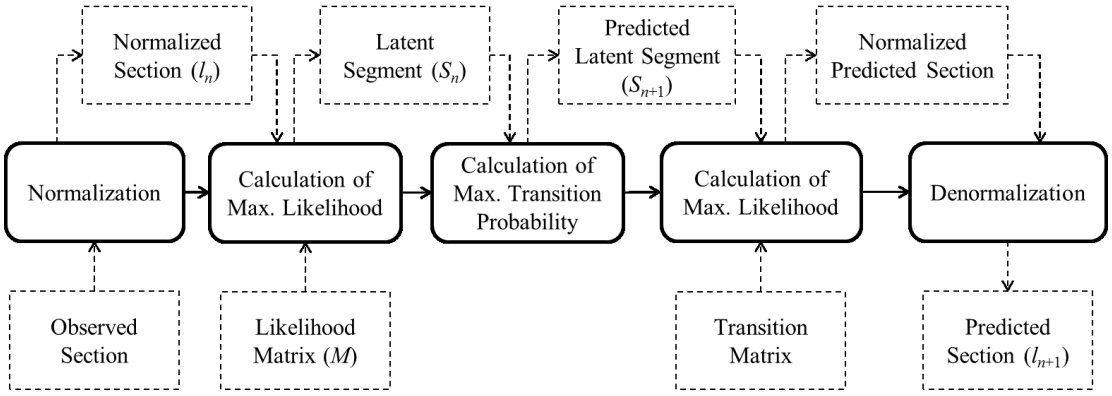
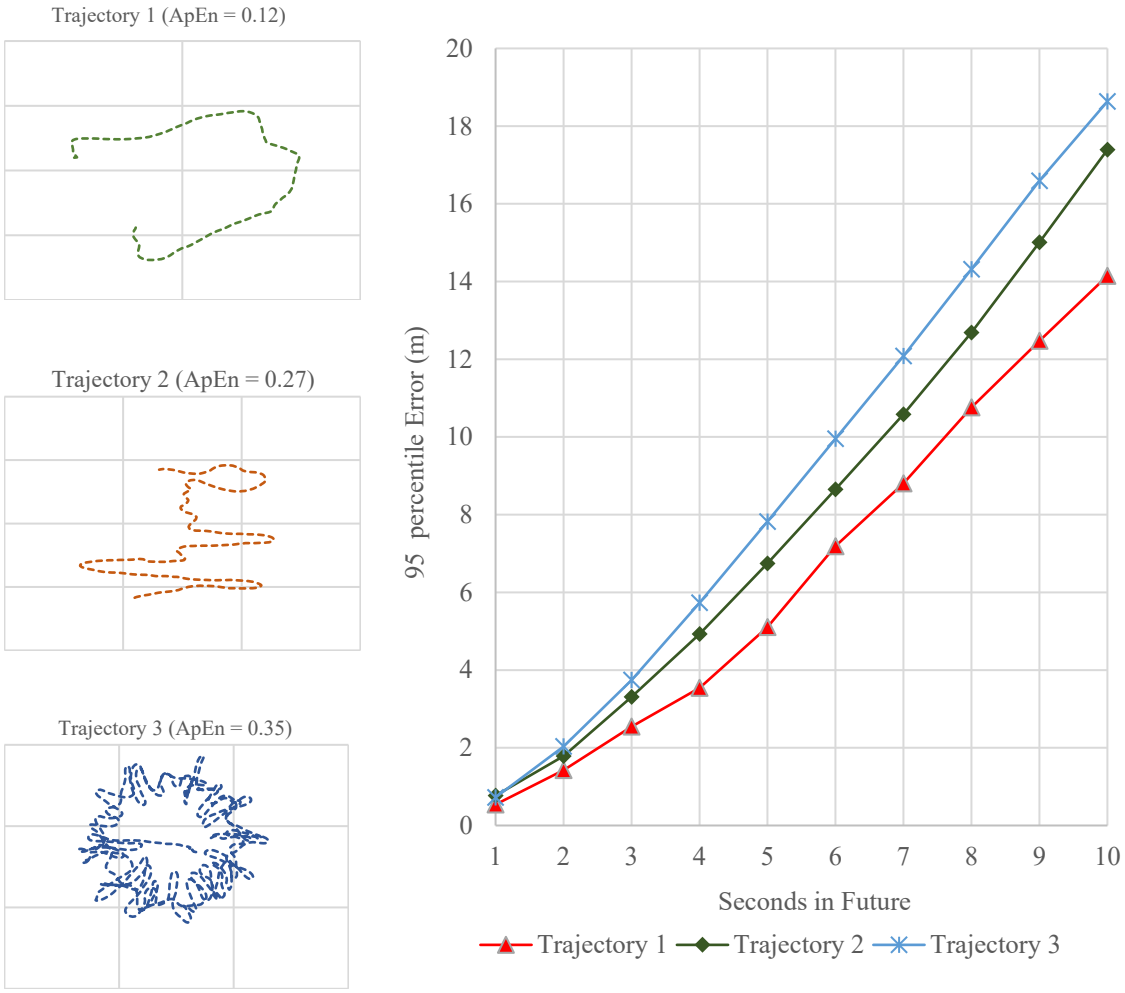


Figure 21. Input-process-output diagram of HMM prediction stage

Preliminary Evaluation of the HMM

To evaluate the robustness of developed HMM, three test trajectory datasets are collected each representing a different level of complexity and randomness. Trajectory 1 and trajectory 2 are the same trajectories used for evaluating the PR model in previous Chapter. In addition to these two trajectories, trajectory 3 is collected which contains more complex movement pattern and U-turns. The approximate entropy (ApEn) for trajectory 1 and trajectory 2 is 0.12 and 0.27, respectively, and trajectory 3 results in an ApEn value of 0.35. As shown in Figure 22(a), trajectory 1 represents a simple straightforward movement with no sharp turns or sudden changes in direction, trajectory 2 contains several sharp turns, and trajectory 3 is the most extreme example containing multiple frequent U-turns. While trajectories 1 and 2 consist of 180 and 360 coordinate data points, trajectory 3 consists of 2408 coordinate data points. All three trajectories are collected at a frequency of 1 Hz (1 data positional data per second).



(a) (b)

Figure 22. (a) Three different trajectories, (b) 95-percentile prediction errors using HMM

As stated before, a minimum of 12 data points are needed to initiate the HMM prediction as each short section is 12 seconds long and calculating the first likelihood requires at least 1 trajectory section. Figure 22(b) also shows the 95-percentile prediction errors for all three trajectories obtained from the HMM plotted against the prediction horizon (seconds in advance). It can be seen that for each scenario, the prediction error increases with prediction horizon. For example, for any given test trajectory, the

prediction error at 10 seconds in advance is higher than the error corresponding to a 5-second prediction horizon. Another observation from Figure 22(b) is that the error increases for more complex trajectories. For instance, the 95-percentile prediction errors for a 5-second prediction horizon for trajectories 1, 2 and 3 are 5.1m, 6.7m and 7.8m, respectively. To assess the sensitivity of HMM prediction to the presence of curves and sudden changes in the trajectory, the actual location and 5-second advance predicted location are plotted for trajectory 1, as shown in Figure 23. It is observed that the prediction error increases locally when there is a turn or a sudden change in direction. As an example, for trajectory 1 (Figure 23) the two highlighted sections demonstrate significant changes in direction. Consequently, the 95-percentile error for these two sections of trajectory 1 is 7.16m while the overall 95-percentile error for trajectory 1 is only 5.1m.

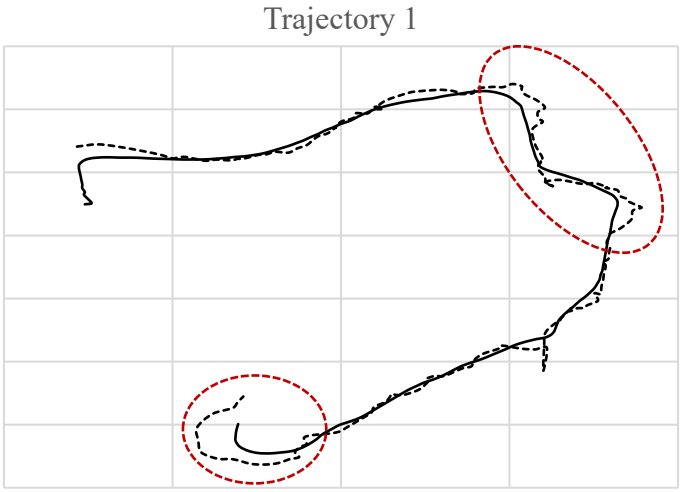


Figure 23. Actual vs. predicted trajectories (trajectory 1)

Comparison of PR and HMM Trajectory Prediction Models

To compare the robustness of the developed trajectory prediction models, the 95-percentile error values obtained from three test trajectories (trajectories 1, 2, and 3) are tabulated Table 3 and plotted in Figure 24. As previously stated, trajectory 1 represents a simple straightforward movement with no sharp turns or sudden changes in direction, trajectory 2 contains several sharp turns, and trajectory 3 is the most extreme example containing multiple frequent U-turns. Figure 24 shows prediction errors for all three trajectories using the developed PR and HMM methods. The 95-percentile error is plotted against the prediction horizon (seconds in advance). It can be seen that for each scenario, the prediction error increases with prediction horizon. For example, for any given test trajectory, the prediction error at 10 seconds in advance is higher than the error corresponding to a 5-second prediction horizon for both models. Another observation from Figure 24 is that in both models the error increases for more complex trajectories. For instance, using the PR model, the 95-percentile prediction errors for a 10-second prediction horizon for trajectories 1 and 3 are 22.5m, and 41.2m, respectively, while the same values using the HMM is 14.1m, and 18.6m, respectively. Since trajectories are considered random walks, it is difficult for both models to predict accurately in the presence of more sharp turns or sudden changes in direction. Also, as expected, the error lines for all three trajectories using HMM fall below the error lines using the PR model which means that the HMM produces more accurate predictions than PR.

Table 3. Side-by-side comparison of 95-percentile error (m) obtained from PR and HMM

Seconds in Advance	Trajectory 1		Trajectory 2		Trajectory 3	
	PR	HMM	PR	HMM	PR	HMM
1	1.20	0.54	2.16	0.77	3.07	0.72
2	2.63	1.42	4.54	1.79	6.43	2.03
3	4.42	2.55	7.18	3.31	10.17	3.74
4	6.47	3.54	10.29	4.93	14.57	5.73
5	8.75	5.11	13.55	6.74	19.20	7.82
6	11.11	7.19	16.61	8.65	23.53	9.95
7	13.66	8.80	19.68	10.58	27.88	12.08
8	16.39	10.76	22.92	12.69	32.46	14.32
9	19.38	12.47	26.09	15.01	36.96	16.59
10	22.56	14.15	29.10	17.39	41.22	18.63

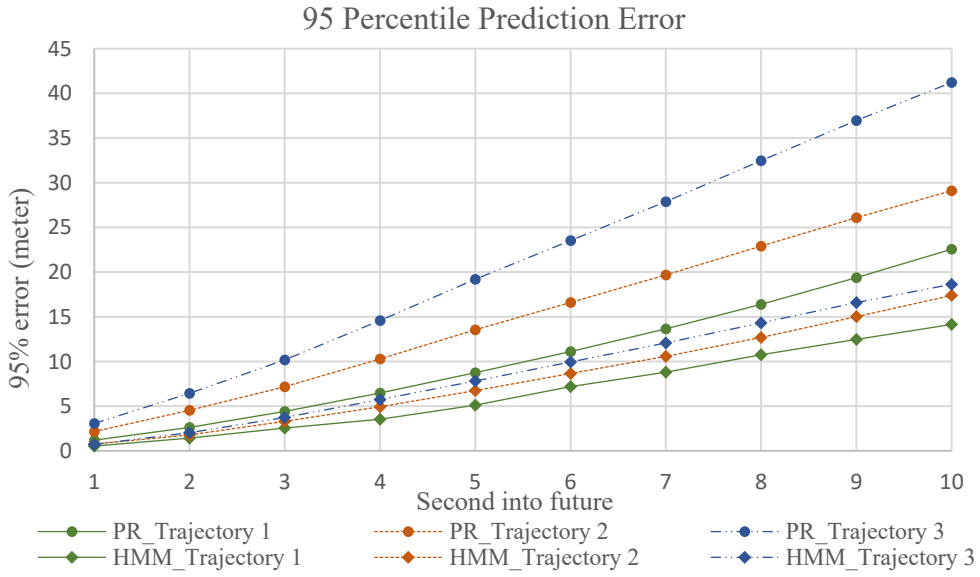


Figure 24. 95-percentile prediction errors using PR and HMM

With the average human walking speed of 1.38 m/s, a 5-second advance prediction results in a collision event prediction at ~7 meters away from hazard. For the case of 5-second advance prediction in Figure 25, HMM results in significantly lower errors compared to PR. In particular, HMM reduces the error by ~42% for trajectory 1, ~50% for trajectory 2, and ~59% for trajectory 3.

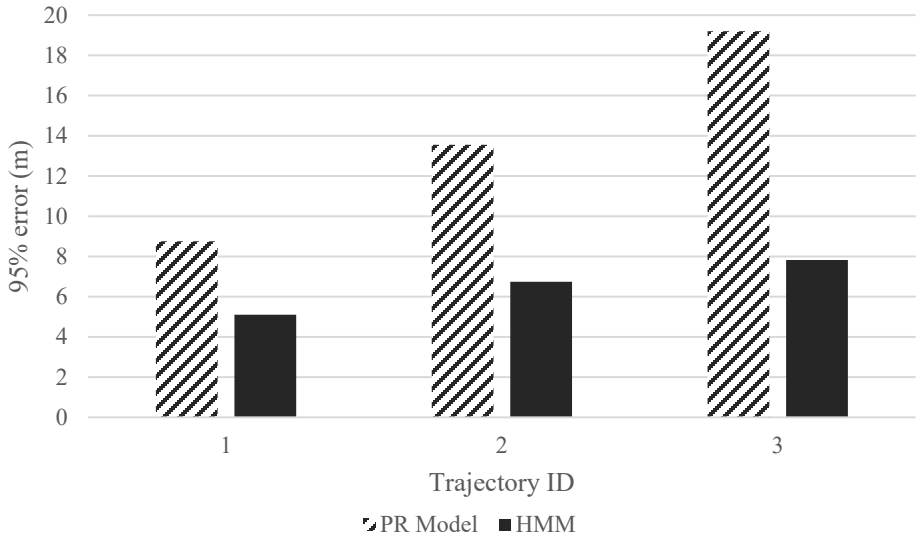


Figure 25. Comparison of 5th second prediction error between PR and HMM

Next, 50 random trajectories are selected from the 78 training datasets and both PR and HMM models are run through the selected trajectories in order to calculate the error for a 5-second prediction horizon and perform a more rigorous comparative analysis between the models. 95-percentile and median errors are calculated for each of the 50 trajectories, and are plotted in Figure 26. Although 95-percentile is a widely-used

parameter for data representation, median can be another good representation of errors in this particular case. It can be seen that in all cases, the error values obtained from HMM prediction is less than those obtained from the PR model. As an example, for trajectory 10, the 95-percentile error obtained from PR and HMM models are 11.1m, and 2.9m, and the median errors are 2.9m, and 0.6m, respectively.

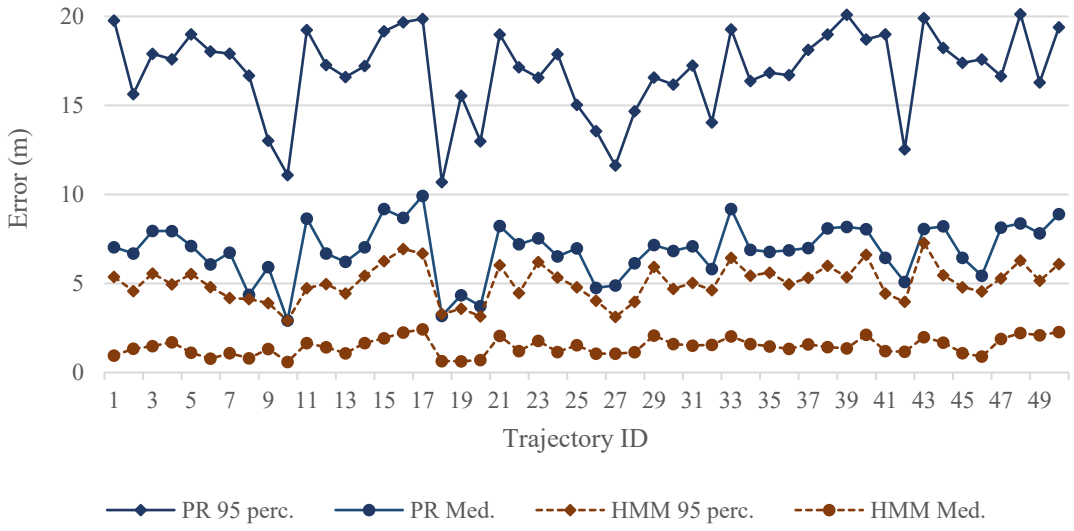


Figure 26. 95-percentile and median errors for 50 randomly selected trajectories

Adaptability of HMM

The analyses in previous Section clearly demonstrate that HMM outperforms PR in terms of prediction accuracy. In this Section, the performance of HMM is evaluated when applied to different types of trajectories with different levels of complexity. Figure 24 and Figure 26 show that HMM performs better for simple trajectories with fewer directional changes and/or randomness. A single trajectory (trajectory 4) from the

training dataset is randomly selected, and a new trajectory (trajectory 5) is collected from the field with similar movement patterns as in trajectory 4. HMM is then applied to both trajectories and prediction errors are calculated. Figure 27 shows 95-percentile errors of all five trajectories (including the three trajectories in Figure 24.

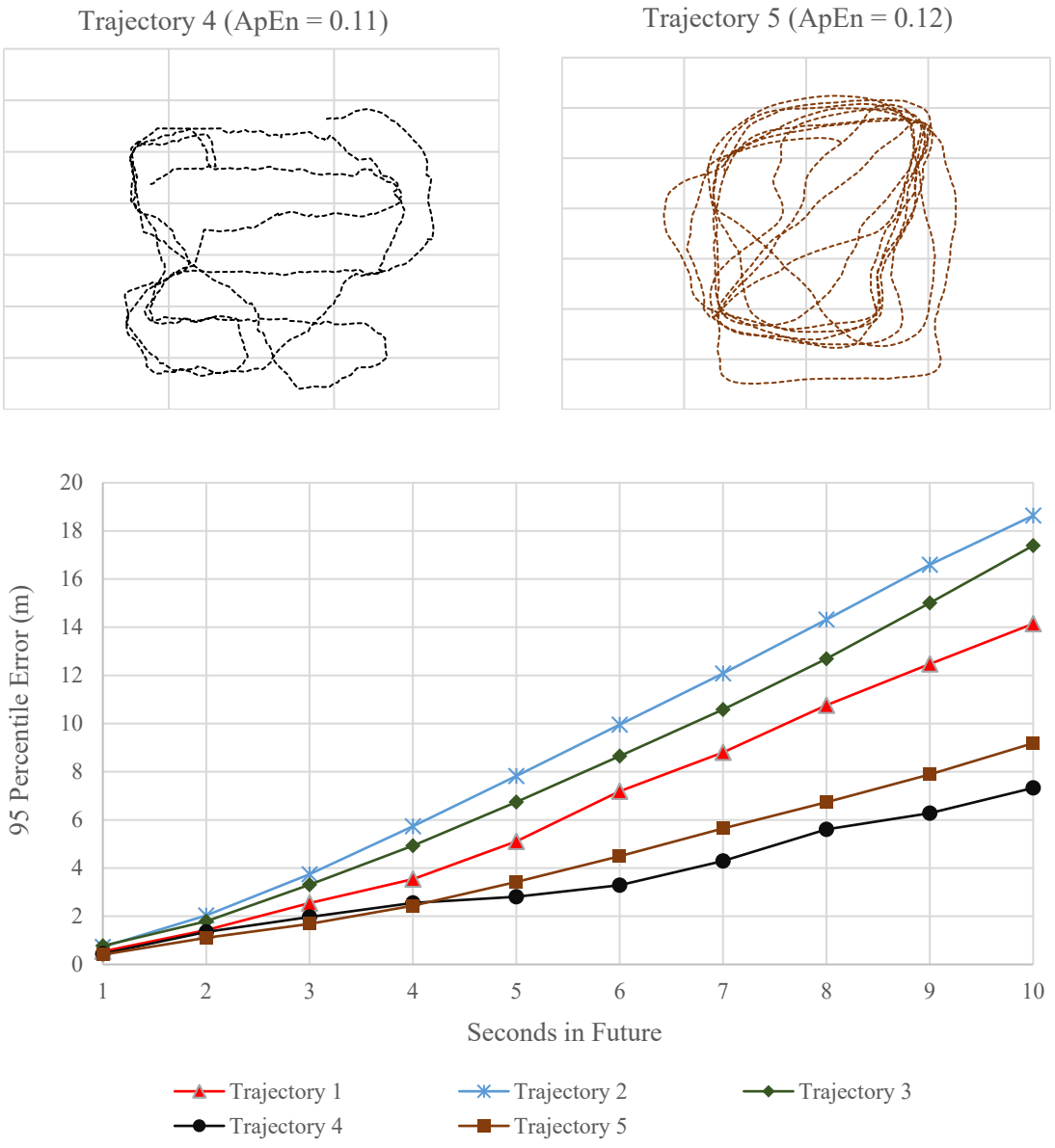


Figure 27. 95-percentile error of HMM for 5 different trajectories

It is observed that HMM performs best for trajectory 4, followed by trajectory 5, and the previously shown three trajectories. It is important to note that despite exhibiting similar ApEn values (0.11 for trajectory 4, 0.12 for trajectory 5, and 0.12 for trajectory 1), which is a measure of complexity and unpredictability, trajectories 4 and 5 outperform trajectory 1 in terms of prediction accuracy. For instance, at 6-second prediction horizon, the 95-percentile errors for trajectories 4 and 5 are 3.3m, and 4.5m, respectively, while the 95-percentile errors for trajectory 1 is 7.2m. This can be explained by recalling that unlike PR, HMM is a learning method capable of predicting more accurately when input data has similar characteristics to training data. This offers unique opportunity to adopt HMM as a self-learning prediction model that automatically updates itself over time with new data streams.

Summary and Conclusions

Machine learning tools, specifically Markov chain (MC) and hidden Markov model (HMM) have been previously used in context-aware applications. In this research,a trajectory prediction technique based on HMM was developed to predict the future location of construction workers on the jobsite. In order to train the HMM, 71 trajectories were collected. Each collected trajectorywas divided into short (12-second long) normalized trajectory sections and K-mean clustering was used to group all such sections into 8 clusters (a.k.a. latent segments). Statistical parameters such as mean, variance, and covariance were then extracted from these clusters. Next, bivariate normal pdf was used to calculate the likelihood of each section to be generated from a specific latent segment, and all calculated probabilities were stored in a likelihood matrix. Since

the sequence of sections over a latent segment is known from the training data, HMM was trained using all of the 71 trajectories to calculate the transition probabilities between states. During the prediction stage, observed sections were used to find the highest likelihood and transition probability, resulting in the most probable future latent segment. A trajectory section prediction is then made using the likelihood matrix and all previously stored trajectory sections. In the evaluation phase, comparative error analysis was conducted to compare the accuracy level of PR and HMM, and it was observed that HMM yielded more accurate predictions. It was also found that HMM could be used as an adaptive model as errors were least when the test trajectory was more similar to training trajectories. Therefore, HMM was ultimately selected for further improvement by incorporating risk attitude in trajectory prediction, as described in next Chapter.

INCORPORATING RISK ATTITUDE INTO TRAJECTORY PREDICTION

An Individual's Perception toward Risk

In recent years, several researchers have conducted several studies aimed at improving the safety environment of construction workers. Previous studies also focused on real-time location-aware techniques to reduce contact collisions between workers and equipment. However, despite some sparse work in other fields, one area that has not yet been fully explored within the construction domain is how a worker's attitude toward risk can influence his or her safety as well as the overall safety environment on the jobsite. Salminen (2004) conducted a survey study that showed that young workers under the age of 25 experienced a higher injury rate than older workers. Another study with 306 participants in three different age groups – adolescents (13-16), youth (16-22), and adults (24 and older) measured risk preferences and results demonstrated that risky decision-making decreases with age (Gardner and Steinberg 2005). A gender-based study showed that women are generally less likely to make risky decisions than men (Charness and Gneezy 2012). According to a USDOT (2004) report, U.S. male drivers are three times more likely to be involved in fatal car accidents than female drivers. Also, it has been reported that female drivers use seat belts substantially more often than men (Waldron et al. 2005). The same study showed that in the U.K., men pedestrians are 80% more likely to be involved in accidents than female pedestrians, and men die much often from accidental poisoning or drowning than women (Waldron et al. 2005).

Incorporating Risk-Taking Behavior into Trajectory Prediction

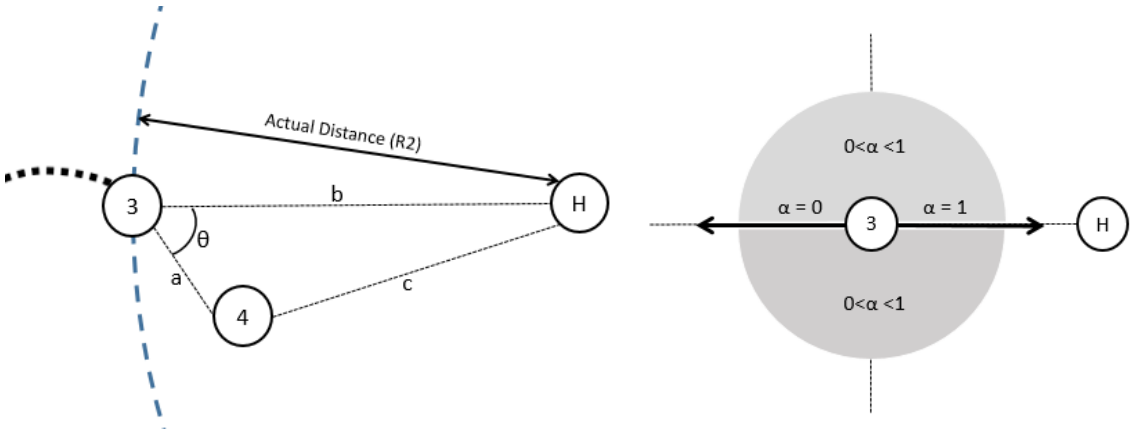
This research makes an effort to incorporate the risk attitude of workers with previously discussed HMM prediction model. An individual's attitude toward risk (represented by one's risk profile) has an important influence on his or her movement trajectory in the vicinity of site hazards. Therefore, it is imperative that a trajectory predicted solely based on mathematical principles must be adjusted to also reflect the extent to which a worker is inclined to take or avoid risks. In this research, the basic principle applied to incorporating risk behavior in trajectory prediction is that if a worker is risk-taker, his or her predicted future position is moved closer to the hazard zone since the worker is more likely to be on a collision course. In order to make the analysis more conservative, no calibration is made for a risk averse worker. In other words, the prediction position is not moved away from the hazard. Two types of risk factors are considered, namely the angular risk factor (α), and the linear risk factor (m). A worker's risk factor (r) is then calculated by multiplying angular and linear risk factors. In order for this approach to yield accurate results, it is important to properly quantify the risk attitude (a.k.a. the aggregate risk factor or μ in this paper) of each worker. To this end, a self-learning formulation is used to continuously calculate, store, and update the aggregate risk factor based on the history of a worker's movements in the vicinity of hazards.

The angular risk factor (α) is calculated based on the worker's actual trajectory. In Figure 28(a), point 3 is on the borderline of the safe distance (R_2) from a hazard center, H, thus initiating the trajectory prediction process. Point 4 represents the next location. It can be seen from this Figure that point 4 is moving with an angle of θ away from H. This

angle is calculated by Equation (26) in which *a*, *b*, and *c* are distances as marked in Figure 28(a) and can be calculated using the Haversine formula. Given θ, the angular risk factor (α) is then calculated using Equation (27).

$$\theta = \cos^{-1}\frac{a^2+b^2-c^2}{2ab} \quad \text{Eq. (26)}$$

$$\alpha = 1 - \frac{\theta}{180} \quad \text{Eq. (27)}$$



(a) (b)

Figure 28. (a) Angular risk factor based on actual locations, (b) Variation of the angular risk factor in different directions

From Equation (26) and Equation (27), it can be seen that within a certain vicinity of the hazard, if a worker is moving directly towards the hazard center, α is 1 or 100%, which means he or she is a full risk taker. In contrast, if a worker is moving in the opposite direction from the hazard center, α is 0, which implies that in that specific instance of time, he or she is a full risk averse. In cases where the direction of workers' movement is at any angle with the hazard center, α is between 0 and 1. Figure 28(b) demonstrates the relationship of α and the worker's movement direction. For a better

understanding of polar risk factor, a hypothetical trajectory is plotted in Figure 29 with a static hazard zone. The value of α at each position is superimposed in form of a circle on the corresponding coordinate at that point. The size of the circle represents the significance of α (how large or small its value is) in that particular position. As shown in this Figure, when the worker approaches the hazard directly, the circle grows, implying that the value of α approaches 1 (maximum possible). In this scenario, it can be inferred that the worker is exhibiting a risk-taking behavior. On the other hand, as the worker is passing by the hazard, the circle shrinks indicating that α approaches 0 (minimum possible). In this case, the worker demonstrates a risk-averse attitude.

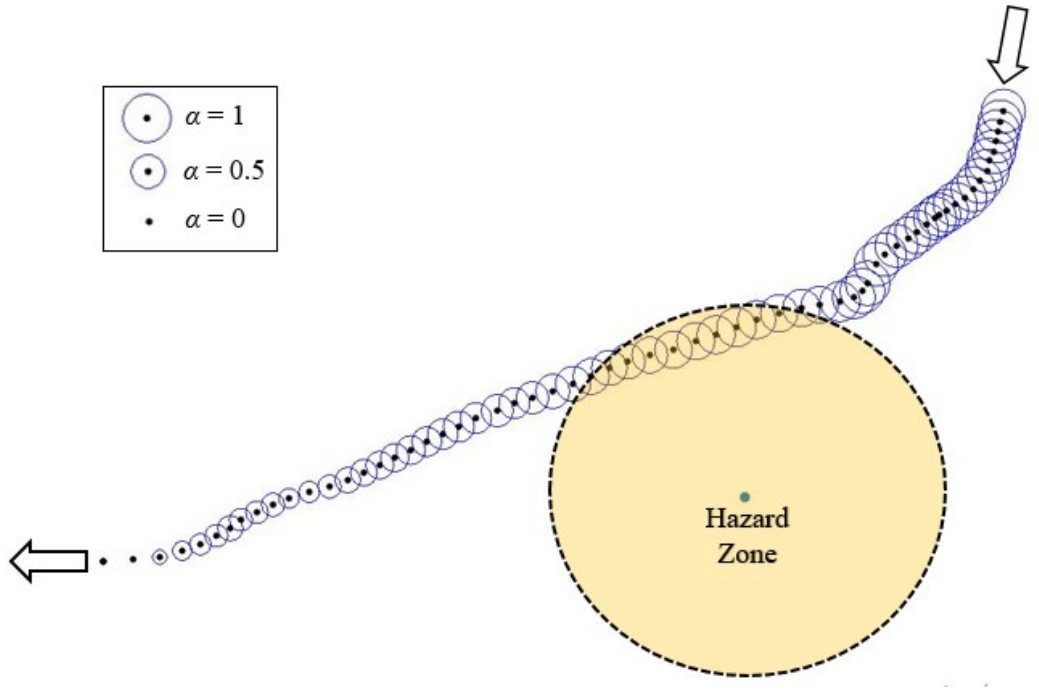


Figure 29. Variation of polar risk factor (α) in the vicinity of hazard zone

The linear risk factor (m), on the other hand, is based on the predicted and actual positions. It represents the radial error of the predicted position relative to the actual position, considering the hazard zone at the center of the circle. In Figure 30, the predicted future position, calculated by the trajectory prediction model is shown as point 4′. The linear distance (d) between the actual position (point 4) and the hazard center H, as well as the linear distance (d_1) between the predicted position (point 4′) and the hazard center H are calculated using the Haversine formula. The linear risk factor (m) represents the difference between d and d_1 as calculated by Equation (28). Knowing m, a risk factor, k can be calculated by Equation (29).

$$Linear\ Risk\ Factor\ (m) = d1 - \mathrm{d} \qquad \text{Eq. (28)}$$

$$Risk\ Factor\ (k) = \alpha m \qquad \text{Eq. (29)}$$

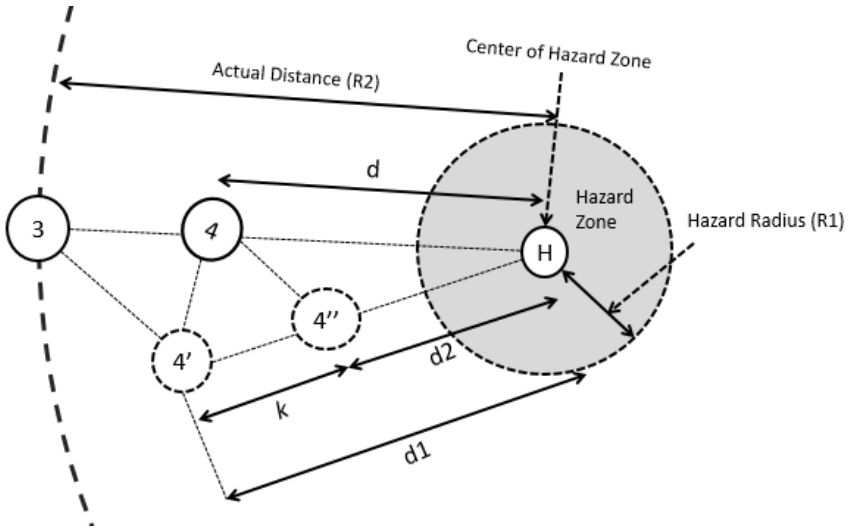


Figure 30. Linear risk factor and adjustment of prediction

At the initiation of the prediction process, a value of 0 (zero) is considered as the first estimate of the aggregate risk factor (μ). As a convention, all workers are initially assumed to be neutral (neither risk taker nor risk averse), and this will be gradually modified throughout time as more positional data is collected from each individual worker. For next iteration, point 4′ is shifted k linear distance units closer to H. The adjusted predicted location is labeled as point 4″. Next, given k and d_1, the modified linear distance d_2 (between points H and 4″) is calculated. This newly calculated distance (d_2) is then compared with a predefined proximity radius (R_1) which is determined based on the type of the hazard. Previous research has suggested values ranging from 8m to 12m depending on the nature and severity of the hazard (Teizer et al. 2015). If $d_2 \leq R_1$, the worker is deemed too close to the hazard, and thus a safety alert is generated. It is noted that, if in any case, $d>d_1$ the risk factor is assumed to be zero, so no adjustment is made (to yield conservative results). In other words, the proposed method adjusts prediction towards the hazard, but not away from the hazard. At this point of the iteration process, μ is refined using the weighted average of individual risk factors k_i calculated in previous steps of the iteration process. In essence, in step n of the iteration process, k values calculated in the previous $n-1$ steps will be used as shown in Equation (30) to calculate μ.

$$Aggregate\ Risk\ Factor\ (\mu) = \frac{\sum_{i=1}^{n-1} k_i(i)}{\sum_{i=1}^{n-1} i} \qquad \text{Eq. (30)}$$

Using this approach, the more recent risk factors receive a higher weight implying that an individual's most recent movements will have a higher influence on how his or

her aggregate risk factor is determined. For example, in step 5 of the iteration, μ_5 is calculated by Equation (31),

$$\mu_5 = \frac{4k_4 + 3k_3 + 2k_2 + 1k_1}{4+3+2+1} \quad \text{Eq. (31)}$$

As evidenced by this Equation, μ may take on a negative value in which case, a decision must be made as to whether to use this negative value and adjust the predicted position by moving it away from the hazard. As an alternative (and to yield more conservative output), one can consider *m* to be zero, when *k* is negative, resulting the value of μ to be always higher than zero. Table 5 shows two separate set of calculation, one considering both positive and negative values for *m* (referred to as the non-conservative design), and one with only positive values for *m* (referred to as the conservative design). Since the developed trajectory prediction method is aimed to improve workers' safety, adjusting the prediction by moving it away from the hazard zone may sometimes result in an impending collision to be overlooked, which is not desirable. Table 4 shows that for conservative design, when the value of *m* is negative, *k* is considered as zero, eventually not affecting μ for those certain time stamps. As an example, for time stamps 53, 54, and 55, μ remains the same as it was for time stamp 52 since in those three time stamps, *m* is negative. On the other hand, for non-conservative design, negative value of *m* is also considered and the value of μ can fall below zero (e.g. time stamps 54 through 60). Figure 31 illustrates the variations of μ from time stamp 51 to 60. As shown in this Figure, μ is always higher than 0 for the conservative design, and can be less than 0 for non-conservative design.

Predicting Impending Collision Events Using HMM

To evaluate the effectiveness of the developed HMM trajectory prediction technique in site safety applications, several experiments are conducted with both static (stationary) and dynamic (moving) site hazards. Static hazard zones are defined as hazards which remain in a fixed location. Examples include but are not limited to excavated trenches or wells, tower cranes, toxic material and chemical storage sites, and

Table 4. Calculation of μ by conservative and non-conservative designs

Time Stamp (s)	Conservative Design				Non-Conservative Design			
	α[1]	m[2]	k[3]	μ[4]	α	m	k	μ
51	0.83	14.49	12.03	5.69	0.83	14.49	12.03	2.24
52	0.76	3.26	2.48	5.44	0.76	3.26	2.48	2.25
53	0.75	-17.78	0.00	5.44	0.75	-17.78	-13.36	1.52
54	0.69	-81.95	0.00	5.44	0.69	-81.95	-56.91	-1.14
55	0.63	-33.73	0.00	5.44	0.63	-33.73	-21.39	-2.04
56	0.58	35.49	20.76	6.61	0.58	35.49	20.76	-1.05
57	0.51	1.33	0.68	6.18	0.51	1.33	0.68	-0.98
58	0.51	-6.20	0.00	6.18	0.51	-6.20	-3.19	-1.07
59	0.49	-7.27	0.00	6.18	0.49	-7.27	-3.59	-1.17
60	0.49	-2.67	0.00	6.18	0.49	-2.67	-1.31	-1.18

[1] Angular risk factor
[2] Linear risk factor
[3] Risk factor
[4] Aggregated risk factor

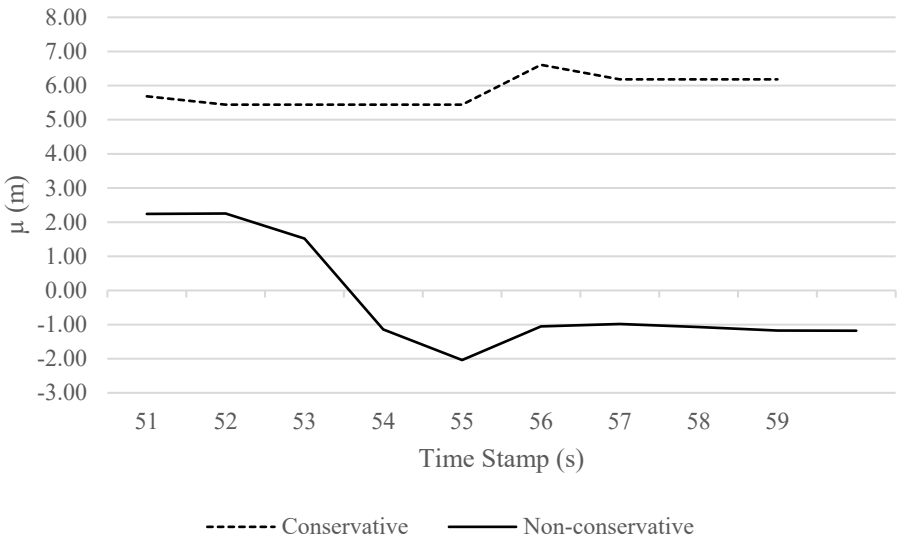


Figure 31. Value of μ in conservative and non-conservative designs

high voltage powerlines. Dynamic hazard zones, on the other hand, change position with time. Examples include but are not limited to heavy equipment such as excavators, crawler cranes, forklifts, and loaders. In the experiments conducted in this research it is assumed that the rough locations of all significant hazard areas are known ahead of time (for static site hazards), or can be determined in runtime (for dynamic site hazards). This is a realistic assumption considering the findings of previous work that among others, explored the feasibility of automatically identifying hazard zones on the site using inertial measurement units (IMUs) mounted on workers' bodies to monitor the bodily response data in the vicinity of the hazardous areas (Kim et al. 2016). The statistical analysis demonstrated that the bodily response of a worker is highly correlated with the hazard zone. In particular, near a hazard are, data from the IMU exhibited a more irregular distribution than in close vicinity of nonhazardous areas.

In the experiments conducted as part of this study, an event is triggered when the worker is inside the buffer radius (R_2), and every time the worker's distance to a hazard is less than the hazard radius (R_1) a collision event is logged. The values for R_1 and R_2 are initially selected by educated guess, and can be modified depending on the situation and type of hazard, equipment blind spots (Hefner and Breen 2004), or severity of a potential collision. For instance, Teizer et al. (2015) suggested 12m (maximum) and 8m (minimum) blind spots to be used for an excavator in safety applications. For experiments involving static hazards, worker's GPS position is obtained from his or her mobile phone and recorded over 15 minutes at a frequency of 1 Hz. The layout of the experiment is illustrated in Figure 30(a) where the hazard zone is marked as point *H*. For these experiments, R_1 and R_2 are selected to be 30.5m (100ft) and 45.7m (150ft), respectively. For experiments involving dynamic hazards, worker's GPS position is obtained from his or her mobile phone and recorded over 30 minutes at a frequency of 1 Hz. For these experiments, R_1 and R_2 are selected to be 10m (32.8ft) and 20m (65.6ft), respectively.

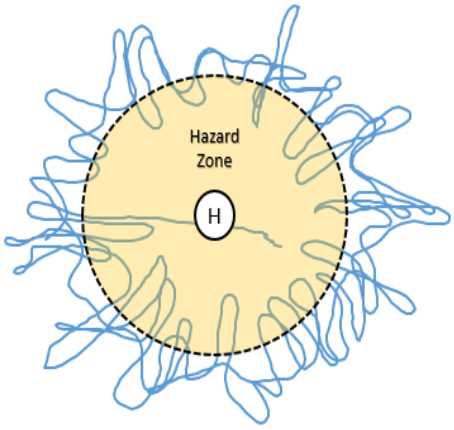
Initially, a prediction is made using the trained HMM to predict future positions of the worker 1 to 10 seconds in advance. Next, risk attitude is factored in to adjust the predicted position. A precision-recall analysis is then conducted to determine the reliability of the prediction model. Any data point inside the buffer zone is considered as an event, and four indicators namely true positive (TP), true negative (TN), false positive (FP), and false negative (FN) are used to calculate three performance measures: precision, recall, and accuracy. A TP event is when the prediction correctly identifies that a worker is inside the hazard zone, a TN event is when the prediction correctly identifies that the worker is outside the hazard zone, a FP event is when the prediction falsely

identifies that a worker is inside the hazard zone, and a FN event is when the prediction falsely identifies that a worker is outside the hazard zone. Recall is the measure of the TP detection over all collision events, precision represents the accurateness of the TP events over all positive events, and finally, accuracy is the measure of overall event detection (inside the hazard and buffer events) for all events. Table 5 summarizes the performance measures and metrics used in the experiments. Figure 32(a) shows the log of positional data collected from the worker in the vicinity of a static hazard. Figure 32(b), 32(c), and 32(d) show that recall, precision, and accuracy decrease with an increase in the prediction horizon. This can be due to the resulting increase in the uncertainty of trajectory prediction. Figure 32(b) shows that recall increases when risk calibration is factored in the prediction. By incorporating the risk factor, the predicted position is further moved toward or away from the hazard.

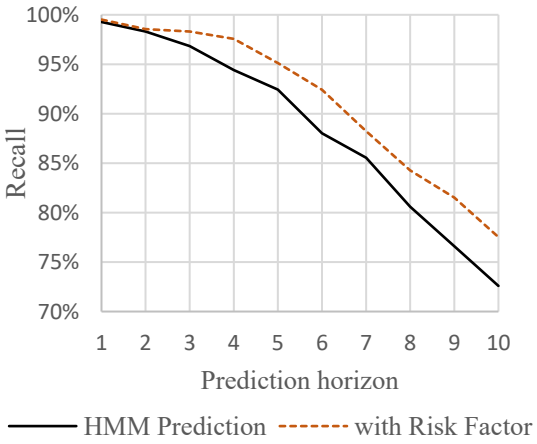
Due to this adjustment, more FP events are generated which can cause a slight decrease in precision as well as accuracy. For instance, for a risk-calibrated 5-second advance prediction, recall increases by 2.5% (from ~92.5% to ~95%) while accuracy decreases by 2% (from ~87% to ~89%).

Table 5: Performance measures and metrics used to evaluate the trajectory prediction model

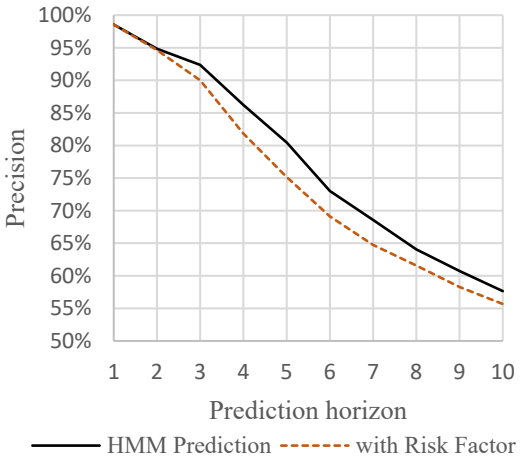
Measure	**Worker Is Inside the …**	
	Ground Truth	HMM Prediction
True Positive (TP)	Hazard zone	Hazard zone
False Negative (FN)	Hazard zone	Buffer zone
False Positive (FP)	Buffer zone	Hazard zone
True Negative (TN)	Buffer zone	Buffer zone
Performance Metric: $\boldsymbol{Recall = \frac{TP}{TP + FN}}$ $\boldsymbol{Precision = \frac{TP}{TP + TN}}$ $\boldsymbol{Accuracy = \frac{TP + TN}{TP + FN + FP + TN}}$		



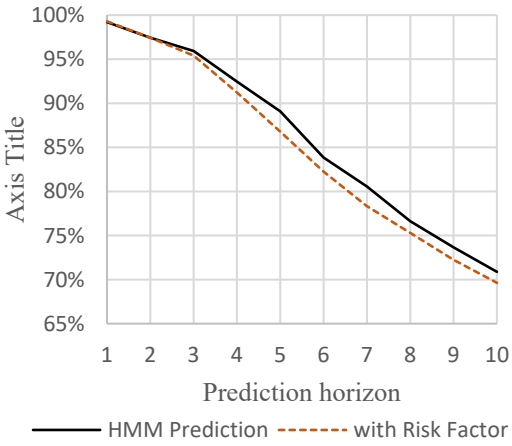
(a) Trajectory and Static hazard



(b) Recall vs. prediction horizon



(c) Precision vs. prediction horizon

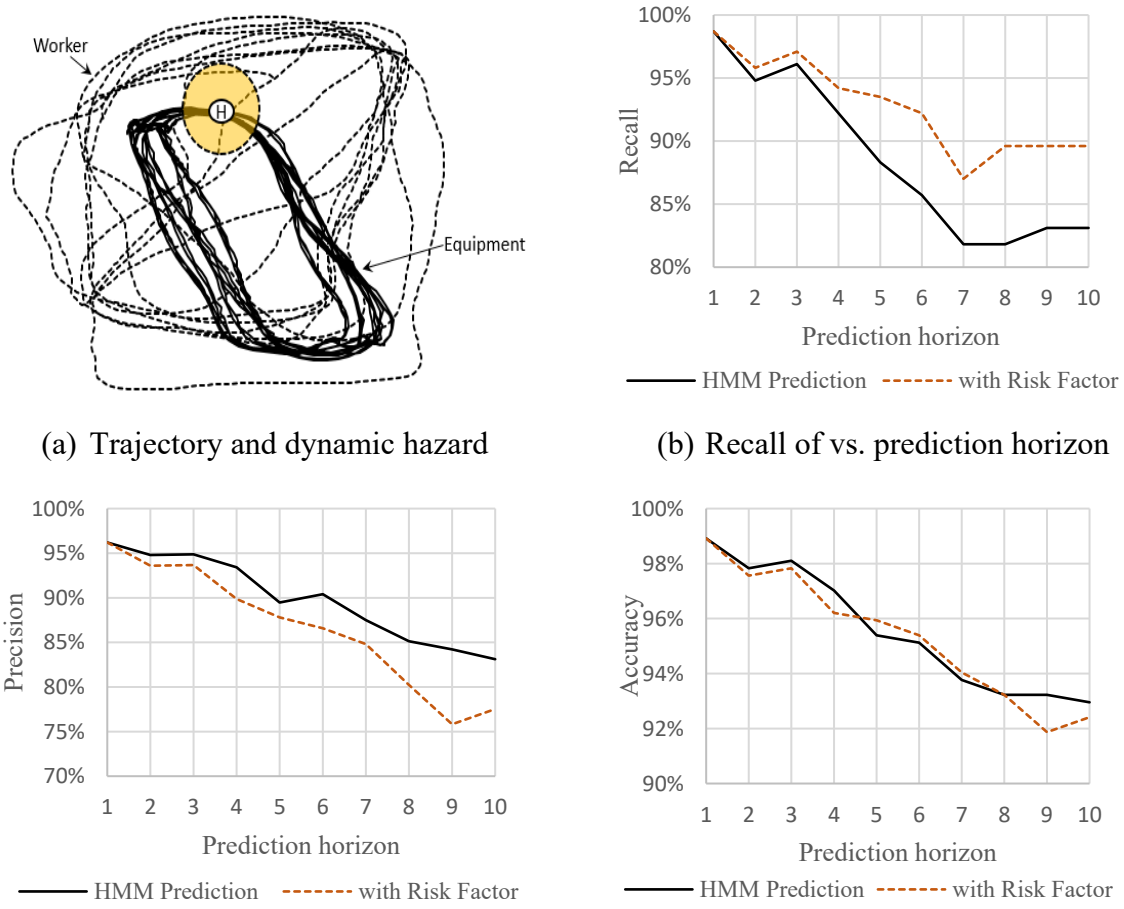


(d) Accuracy vs. prediction horizon

Figure 32. Recall, precision, and accuracy analyses for static hazard experiment

Figure 33(a) shows the log of positional data collected from the worker in the vicinity of a dynamic hazard. In this experiment, the dynamic hazard is a piece of heavy equipment moving between two points, while the worker is moving in the peripheral area occasionally crossing the equipment path. In this experiment, a total of 369 events are detected, with 77 potential collisions (worker inside the hazard zone). Figure 33(b), 33(c),

and 33(d) show that while precession, recall, and accuracy drop for farther prediction horizons, incorporating risk attitude results in a higher recall. For instance, for a risk-calibrated 5-second advance prediction, recall increases by 6% (from ~88% to ~94%) while precision falls only by 1% (from ~89% to ~88%).



(a) Trajectory and dynamic hazard

(b) Recall of vs. prediction horizon

(c) Precision vs. prediction horizon

(d) Accuracy vs. prediction horizon

Figure 33. Recall, precision, and accuracy analyses for dynamic hazard experiment

Calculation of the three measures (recall, precision and accuracy) was done by non-conservative design for both the experiment explained above. Further, for the static hazard experiment, same three measures were calculated based on only prediction and with risk factor; both conservative and unconservative to compare the results. Recall, precision and accuracy for five second prediction horizon is plotted in Figure 34. Figure shows that incorporating risk factor (both conservative and unconservative) with the HMM prediction increases the recall, but decreases precision and accuracy. It is important to mention that, in the safety realm recall is more significant measure than precision and accuracy as recall is the measure which represents the capability of the designed model to identify how many times worker is actually inside the hazard zone and model correctly detected it. Figure also shows that, conservative design increases the recall, but decreases precision and accuracy.

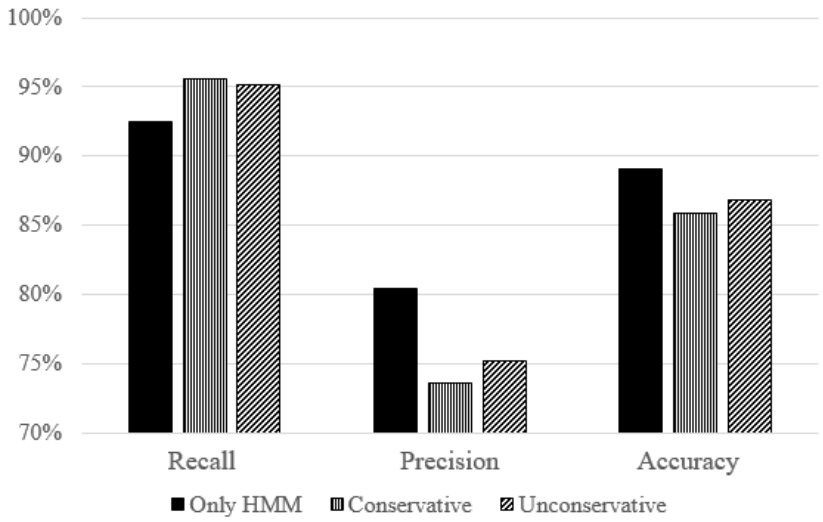


Figure 34. Recall, precision and accuracy comparison based on HMM and risk factor (μ)

Summary and Conclusion

Though extensive research has been done to improve the safety environment in construction, but one area that has not been fully researched is the effect of individuals' risk attitude in worker's safety. Researches demonstrates an existing correlation between individuals' risk attitude and his/her age, gender, experience. In this research risk, attitude of the worker in the vicinity of site hazards was incorporated to the HMM prediction model. A mathematical model was presented to calculate the risk factor of the worker and the HMM predictions were further adjusted based on the risk factors. Two field experiment were conducted, one with a static hazard zone, and another one with dynamic hazard zone. A precision, recall and accuracy analysis of both the experiments demonstrated that the recall value significantly increases after combining risk factor with HMM.

PREEMTIVE CONSTUCTION SITE SAFETY (PCS2) APPLICATION

Smartphones as Disruptive Technologies

In recent years, the development and wide adoption of smartphones has entirely changed the definition of mobile phone usage. Smartphones are not only used as a communication device, but also have become an integral part of the daily life. The number of mobile phone users has been steadily growing with almost 2 billion smartphone users in the market by late 2015 (Kissonergis 2015). In addition to its ordinary features and functionalities, almost all smartphones come with a rich set of embedded sensors such as the GPS, accelerometer, gyroscope, microphone, camera, and digital compass. Collectively, these sensors have inspired researchers to explore emerging applications of mobile phone sensing in multiple disciplines including healthcare (Consolvo et al. 2008), social networks (Miluzzo et al. 2008), safety and environmental monitoring (Mun et al. 2009), and transportation (Taigarajan et al. 2009)Smartphones are ideal mobile devices that support disrupting sensing applications such as sharing real-time location on social networks, tracking personal carbon footprint, or monitoring personal health (Lane et al. 2010). These mobile devices are programmable and come with internal storage and versatile computing resources which allow third party application developers to implement and incorporate a wide range of programs. In addition to standard features, phone manufacturers also provide online application stores (a.k.a. app stores) which allow users to access an extensive collection of application as well as reach out to other users with their own applications. Lastly, the computing cloud of smartphones permits programmers to store the content of their mobile services on

back-end severs, which facilitates large scale sensor data storage, transfer, and computation as well as supports analysis of big sensor data. These advances are all great assets for innovative research that can revolutionize a large number of existing practices, ultimately significantly impacting everyday lives. Location-based services (LBS) are defined as network-based services that use real-time geo-location data from a mobile device or smartphone in order to deliver context-specific information to a user (Xu and Gupta 2009). The most used LBS are the navigation services where users can plan their routes ahead of time and follow that in real-time while driving, cycling, or walking. Some other examples are automatically searching bus or train timetable as the user walks toward a stop or a station, and tracking user's carbon footprint as a result of using car or other means of transportation.

The Android Developer Platform

Android is a Linux-based operating system which uses the software stack architecture design pattern. As shown in Figure 35, the pattern consists of four major parts, Linux Kernel, Android Runtime and Libraries, Application framework and Applications (Zhao and Tian 2012). The Linux kernel is the foundation of Android platform and is created on top of Linux Kernel 2.6. For threading and low level memory management, the Android Runtime (ART) relies on Linux Kernel. Other Linux Kernel dependent features are Process Management, Driver Model, File System Management, Network Management, and User Account (Security). Android Runtime consists of Dalvik Virtual Machine (DVM). DVM is a kind of Java virtual machine specifically designed and optimized for Android systems. Starting from Android version 5.0 (API Level 21),

each application runs on its own instance of Android Runtime (ART). If an application runs well on ART, then it will work on Dalvik, but not necessarily the other way around. Android system native libraries connect Linux Kernel and Application Framework, and are developed by C/C++ language. Some of these libraries provide Java framework APIs to expose the functionality to applications through the Android platform. Key Android libraries include android.app, android.content, android.database, android.os, android.opengl, and android.text. The Android application framework layer provides many higher-level services to application layer through Java classes. The developers are allowed to access the entire API framework and make use of these services in their applications. This layer implements the concept that applications are constructed from reusable, interchangeable, and replaceable components. Applications layer is located at the top of the Android software stack, and consists of both native and third party apps. An average user would mostly interact in this layer through basic functions such as dialer, browser, and contact manager.

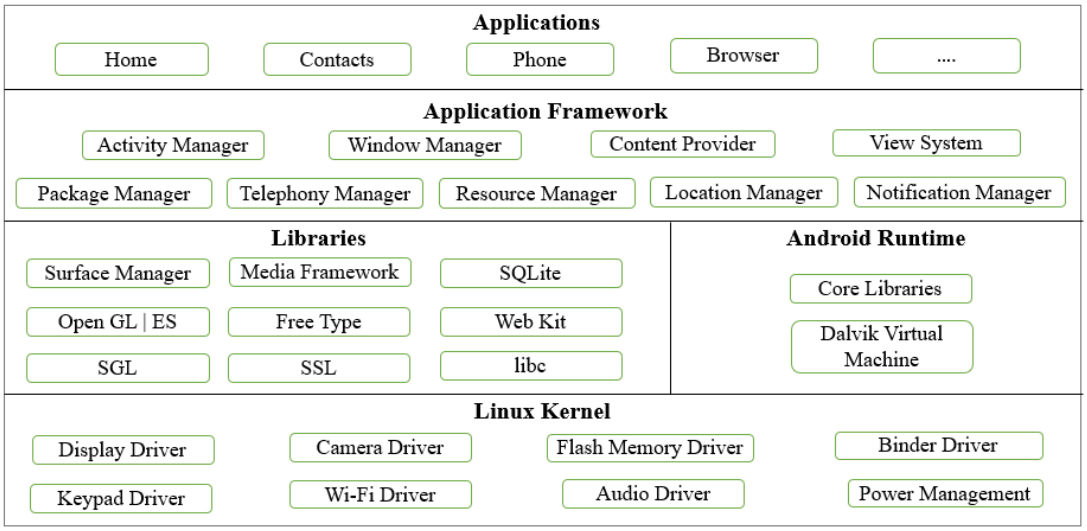


Figure 35. Android platform architecture

Preemptive Construction Site Safety (PCS2) Mobile Application

In this research, an Android implementation of the risk-incorporated HMM prediction model is developed to assess the practicality of the designed methodology. This mobile application, Preemptive Construction Site Safety (PCS2), relies on the built-in GPS sensor of a smartphone to locate a worker in the open space. A sensor fusion approach is adopted for triangulation using satellite, Wi-Fi, and cellular networks to obtain more accurate positional data. The process flowchart of PCS2 is illustrated in Figure 36. Once launched (node "Start"), PCS2 accesses the GPS of the smartphone and continuously collects user's position. If the user does not move by a minimum distance between two consecutive coordinates, the program considers the user stationary (i.e. not moving) and a null answer is generated in response to the "Walking?" decision node. On the other hand, if the user moves with a certain velocity, PCS2 stores his or her movements (GPS coordinates) in a SQLite database. If the user's current position is inside a buffer zone (i.e. trigger event), then PCS2's background service initiates the HMM algorithm. If the predicted coordinate is inside the hazard zone, the application generates and displays an alert message.

As shown in Figure 37, in addition to the background mathematical implementation, PCS2 has a user friendly graphical user interface (GUI). The GUI contains *Layout XML* files which are rendered as a set of *View* class objects. The "initial layout" of the GUI allows the user to manually enter input parameters such as the number of hazard zones and input their global coordinates, hazard and buffer radii, and the prediction horizon.

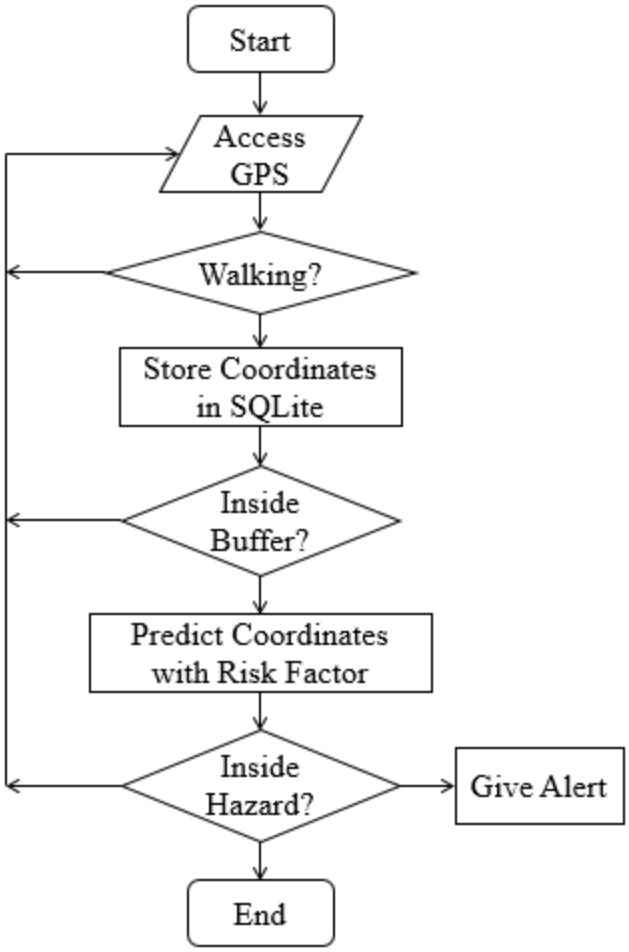
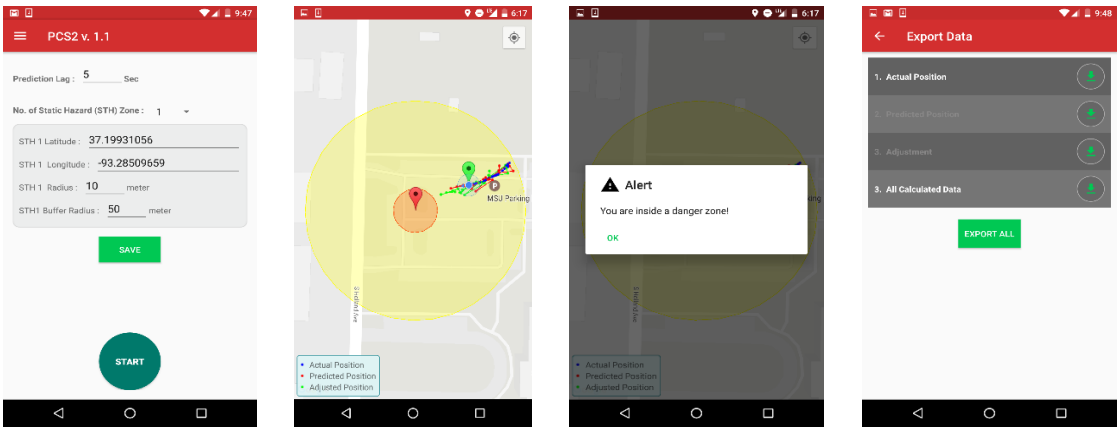


Figure 36. High-level flowchart of the PCS2 mobile application

The “operation layout” uses Google Map in the background for real time visualization. In this layout, hazard and buffer zones are superimposed on Google Map and shown as color-coded circles. The designed GUI also plots actual, predicted, and adjusted (i.e. risk-incorporated) positions. Once a collision event occurs (adjusted position falls inside the hazard zone), the application generates an alert dialogue box with ringtone and vibration in the “alert layout”. Finally, to support post-analysis of data, all collected data and calculated variables can be exported to the internal phone memory in .csv format using the “export layout”.



Initial layout Operation layout Alert layout Export layout

Figure 37. Screenshots of the PCS2 mobile application layouts

The PCS2 application consists of four major Android activities. The *Main Activity* initiates the application and contains the code which calls different methods from other activities. The *Map Activity* contains the bulk of the PCS2 code. Initially, it allows users to input prediction lag, latitude, longitude, buffer radius, and hazard radius values. Clicking the *"Save"* button saves these initial values to the application. When user clicks on the *"Start"* button, it implements a *Location Manager* service which allows the user to use both ACCESS_COARSE_LOCATION and ACCESS_FINE_LOCATION. ACCESS_COARSE_LOCATION provides the location of the user from the cellular network, and ACCESS_FINE_LOCATION provides the location based on GPS_Provider. Combining both methods generates a more accurate detection of users' geo-location. Figure 38 shows how permission to access both location managers are implemented inside the application code.

```
        if (android.os.Build.VERSION.SDK_INT >=
Build.VERSION_CODES.M) {

            int accessCoarsePermission =
ContextCompat.checkSelfPermission(this,
android.Manifest.permission.ACCESS_COARSE_LOCATION);
            int accessFinePermission =
ContextCompat.checkSelfPermission(this,
android.Manifest.permission.ACCESS_FINE_LOCATION);

            if (accessCoarsePermission !=
PackageManager.PERMISSION_GRANTED || accessFinePermission !=
PackageManager.PERMISSION_GRANTED) {

                // The Permissions to ask user.
                String[] permissions = new String[]
{android.Manifest.permission.ACCESS_COARSE_LOCATION,

android.Manifest.permission.ACCESS_FINE_LOCATION};
                // Show a dialog asking the user to allow the
above permissions.
                ActivityCompat.requestPermissions(this,
permissions, REQUEST_ID_ACCESS_COURSE_FINE_LOCATION);
```

Figure 38. Accessing network and GPS provider of smartphone

After collecting 12 coordinates points, the *Map Activity* implements three methods to normalize the trajectory section. First, tranSegemnt translates the 12-second trajectory to an origin of (0, 0), and allignSegment alings the section's initial velocity with the X direction. Finally, to complete the normalization process, scaleSegment scales the trajectory so that the initial velocity is (1, 0). If the current location is inside the buffer zone, this activity then initiates the HMM by first implementing getLikelihoodSection which obtains the likelihood value from the likelihood matrix, meaning it finds out to which latent segment does the current trajectory belong. Next, getTransitionSegment picks the predicted latent segment with the highest probability from the likelihood matrix. Then, it implements getLikelihoodSection again to identify the most likely trajectory section to be generated from the selected latent segment. Ultimately, denormalization is

implemented which basically reverses the normalization process to calculates the 10-second real-size predicted trajectory segment. Based on the initial user input for prediction lag, the *Map Activity* stores the latitude and longitude (e.g. if the user inputs 4 seconds in the prediction lag box, *Map Activity* only stores the 4th second predicted latitude and longitude). After the initial prediction, this activity implements a riskFactor method which uses the techniques explained in previous Chapter to calculate the risk factor based on the hazard position, and return the adjusted prediction. If the adjusted prediction is inside the hazard zone, it generates an alert by implementing alertDialogueBuilder, which also contains a VIBRATOR_PROVIDER to generate a physical device vibration. Beside the background mathematical calculations, the *Map Activity* implements addMarker and addPolylines to display the current location, predicted location, and adjusted location using colored markers and polylines. Figure 39 shows the implementation of the addPolylines method.

```
if (realdata.getPoints().size() > 1) {

    line_actual = mMap.addPolyline(realdata);
    line_prediction = mMap.addPolyline(prediction);
    line_adj = mMap.addPolyline(adjment);

}
```

Figure 39. Adding polyline to coordinates

The *Export Activity* initiates the *Export Layout* shown in Figure 37. The SQLite database is used to store the trained HMM matrices (transition matrix, likelihood matrix), and real-time, predicted, and adjusted user's positions. All variables such as the angular

risk factor (α), and linear risk factor (m), aggregated risk factor (μ) are also stored in the database to assist data analysis. The *Export Activity* implements a CSVwriter class to write and store the data in a .csv file. Figure 40 shows a section of code thatwrites the .csv file in PCS2.

```
            File file = new File(exportDir, filename);

            try {
                file.createNewFile();
                CSVWriter csvWrite = new CSVWriter(new
    FileWriter(file));
                csvWrite.writeNext(curCSV.getColumnNames());
                while (curCSV.moveToNext()) {
                    //Which column you want to exprort
                    String arrStr[] = {curCSV.getString(0),
    curCSV.getString(1), curCSV.getString(2), curCSV.getString(3)};
                    csvWrite.writeNext(arrStr);
                }
                csvWrite.close();
                curCSV.close();
```

Figure 40. Exporting data in .csv files

Field Experiment of PCS2

A field experiment is conducted at Missouri State University (MSU) campus to evaluate the effectiveness of the PCS2 application. Real-time coordinates of the user are collected and analyzed by the application and a prediction based on HMM is made. Further, the prediction is adjusted using the risk factor and if there is an impending collision, PCS2 provides a safety alert. As shown in Figure 41, the initial inputs are the coordinates of the hazard, prediction horizon, and hazard and buffer radii. A forklift is used as a hazard and is parked in the area surrounded by the Kemper Hall and Bear Park

North on MSU campus. Prediction horizon (lag) is set at 5 seconds, and hazard and buffer radii are set to be 10m and 20m, respectively. Figure 42 shows the experiment setup with the user, hazard, and hazard zone. The user carries the mobile device on which the PCS2 application is launched. Figure 43 illustrates user's trajectory captured during the field experiment.

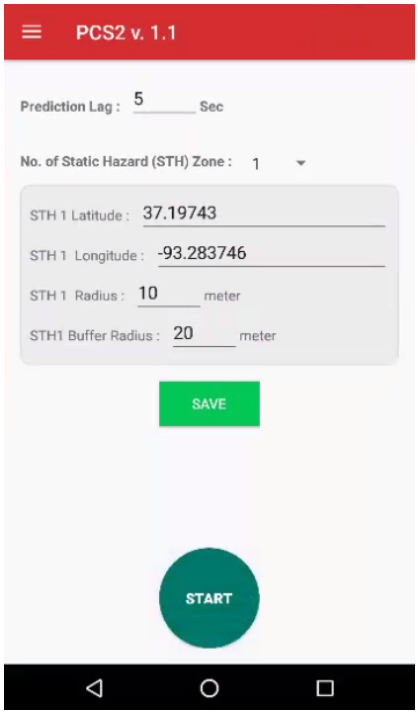


Figure 41. Initial inputs of PCS2 for the field experiment

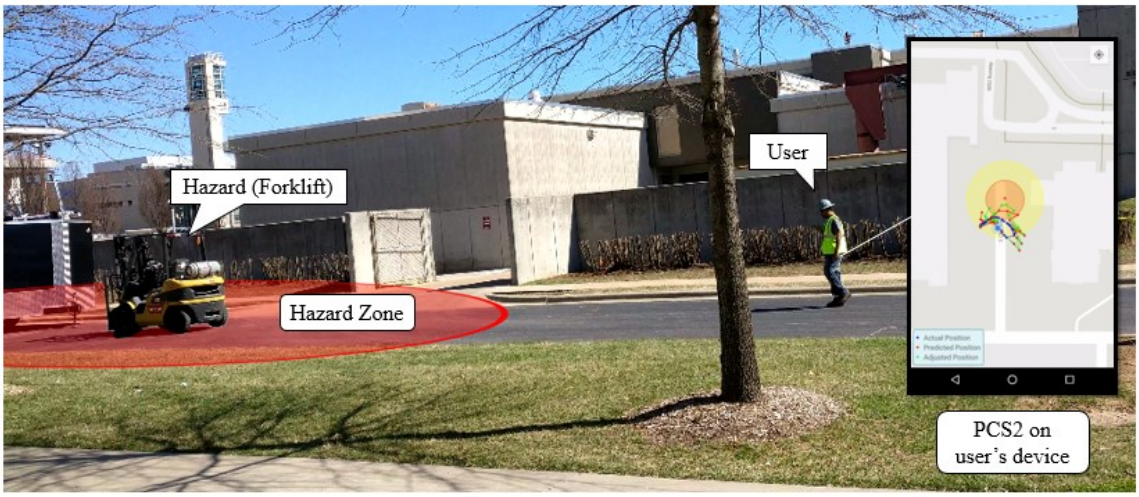


Figure 42. Experiment setup with the forklift as a site hazard

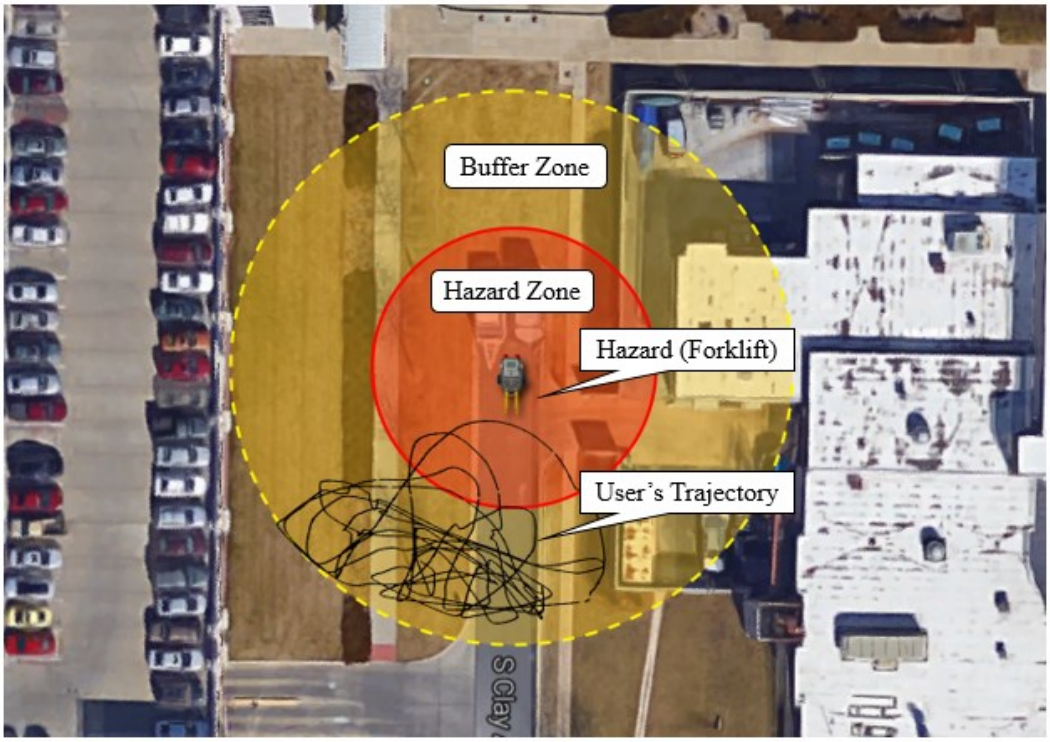


Figure 43. Collected user trajectory in the vicinity of the forklift

During the experiment, if there is any impending collision event detected by PCS2, the application gives an alert with vibration in advance so that the user has enough time to assess the surroundings and adjust his or her walking trajectory accordingly. Figure 44 illustrates a sample impending collision during the field experiment, where the user got too close to the hazard and the PCS2 application correctly predicted an imminent collision event, and provided a timely alert to the user.

Figure 44. User approaching a hazard zone during the field experiment

In total, 15 such alerts are generated by PCS2 during the field experiment. For each alert, the exact position is marked on the ground where the alerts are given. After the experiment, 15 distances each corresponding to an alert are measured. Considering average human walking speed (ranging between 0.5 m/s and 1.5 m/s), a 5-second advance prediction in theory should result in an alert within a distance of 12.5m to 17.5m from the hazard. If a generated alert is within this range, it is considered “timely”. Alerts

generated when the user was closer than 12.5m or farther away than 17.5m are considered “too late” or “too early”, respectively. Table 6 summarizes the results obtained from the field experiment in terms of the timeliness of the generated safety alerts. As seen in this Table, for this particular experiment, 10 out of the total 15 generated alerts are “timely”, 5 are “too late”, and no “too early” alert was generated.

Table 6. Results of PCS2 field test

Alert	Distance from Site Hazard (meter)	Alert Timeliness
1	10.14	Late
2	12.9	Timely
3	12.6	Timely
4	13.72	Timely
5	10.21	Late
6	12.92	Timely
7	13.1	Timely
8	13.32	Timely
9	8.54	Late
10	12.98	Timely
11	11.3	Late
12	9.15	Late
13	13.84	Timely
14	14.47	Timely
15	12.69	Timely

Summary and Conclusion

The advancement of information technology has initiated an emergence of new mobile devices equipped with a rich set of embedded sensors capable of high level computing, and opened new doors for researcher in multiple domains. In the same context, several operating systems (e.g. Android, iOS) for mobile devices have been developed and introduced. These operating systems take advantage of high computational power and online application stores allowing developers to access a large user population worldwide. In this research, smartphone's GPS location services and Android operating system were used to design and develop a native Android-based mobile application called "Preemptive Construction Site Safety" or PCS2 capable of real-time location tracking and predicting the future location of a user. The real advantage of PCS2 is that it incorporates risk attitude calculations described in previous Chapters for trajectory prediction in the vicinity of site hazards. In this Chapter, first the software architecture of PCS2 application was discussed at length. Next, a field experiment was carried out in which a static hazard (forklift) with buffer and hazard radii of 20m and 10m was used, and risk-incorporated trajectory predictions were made with a prediction horizon of 5 seconds. Result of the experiment showed that PCS2 could detect 10 impending collision events out of a total of 15. Given the promising preliminary results, there are several aspects of this prototype (e.g. software design, user interface, background mathematical formulae) that will be further improved as part of the future directions of this research so that PCS2 can be widely adopted in large-scale real-time construction site safety applications.

CONCLUSIONS AND FUTURE WORK

Conclusions

The construction industry is considered as one of the most hazardous and risk-prone working environments of all industries due to the presence of frequent and dynamic interactions between equipment, machinery and tools, personnel, and materials. In the past, researchers have worked to address the issue of jobsite safety from different perspectives including finding ways to reduce and ultimately eliminate proximity-related accidents in construction projects. To this end, several classes of location tracking technologies such as global positioning system (GPS), wireless local area network (WLAN), radio frequency identification (RFID), and ultra-wide band (UWB) have been used to collect positional data of construction resources and use the collected information to improve safety and health of project crews working on the jobsite. More recently, the explosion of mobile technologies integrated with a rich set of embedded sensors has opened the opportunity to explore their application to addressing long-standing problems within the construction domain including site safety. In particular, ubiquitous smartphone devices carried by almost everyone on the jobsite can be used to capture real time positional data of workers and equipment. A mobile position tracking framework equipped with this capability can be a great asset as it will allow the timely delivery of context-aware information such as safety alerts to workers, thus preempting potential hazardous encounters on the site. In this Thesis, a mobile location-aware safety framework was successfully designed and implemented. In addition to its ability to track a user's position in real time, the real value of the designed methodology is that it (1) uses

a novel hidden Markov model (HMM)-based trajectory prediction to estimate with high fidelity the future trajectory of a moving user from past positional data, and (2) allows the formulation and integration of a person's attitude toward risk into the trajectory prediction model which creates a customized solution for individual users based on their past behavior and movement history in the vicinity of site hazards.

In this Thesis, following an extensive literature review and identification of problem statement, the Chapter, "Trajectory Prediction Models", explored the efficiency and accuracy of polynomial regression (PR) to predict future trajectories of construction workers on the jobsite given their past positional data collected by mobile sensors carried on their bodies. An instrumental error analysis was performed prior to using the collected GPS data from each smartphone. Two trajectories were collected using smartphones and the developed PR model was implemented to predict future locations of workers who carried the phones. Next, a 95-percentile error analysis was performed and results indicated that the linear PR model outperformed polynomial PR models for both trajectories. Also, the prediction error increased both with the prediction horizon (seconds in advance) and when there was a sharp turn or a sudden directional change in the collected trajectory.

In the next Chapter, "Machine Learning and Trajectory Prediction", a hidden Markov model (HMM)-based prediction model was developed, trained, and tested to predict a worker's future locations given his or her past movement trajectory. In particular, 71 trajectories were collected, processed, and used to train the HMM which was then applied to find future positions given new trajectories. A comparative error analysis was also conducted which indicated that compared to the previously developed

PR model, HMM generates more accurate predictions. Another important finding was that errors resulted from HMM prediction were smaller when test trajectories were more similar to training trajectories. Therefore, HMM was ultimately selected for further improvement through incorporating an individual's risk attitude into the formulation.

Next Chapter, "Incorporating Risk Attitude into Trajectory Prediction", described a mathematical formulation to enhance the previously developed HMM trajectory prediction technique by calculating and integrating the behavioral risk factor of a construction worker in the vicinity of site hazards. Two field experiments were conducted, one with a stationary hazard, and another one with a moving hazard to systematically evaluate the value of incorporating risk attitude into the prediction model. In both experiments, precision, recall, and accuracy analyses were conducted and results demonstrated that in general, introducing risk attitude as a parameter in the prediction model increased recall, thus increasing the likelihood of detecting impending safety accidents prior to them taking place.

Finally, in Chapter, "Preemptive Construction Site Safety (PCS2) Application", the software architecture of an Android-based mobile application, PCS2, was described in detail. PCS2 uses a smartphone GPS location services to track real time positions of a worker (who is carrying the device) and accurately predict his or her future location in the immediate future using the developed risk-incorporated HMM-based trajectory prediction model. To test the efficiency and applicability of the designed mobile application, a field experiment was conducted with a forklift as a stationary hazard. Results of the experiment showed that PCS2 correctly detected 10 impending collision events out of a total of 15, and generated and displayed timely warnings to the user.

The following summarizes the main milestones of this research that have been successfully achieved:

- A polynomial regression (PR)-based trajectory prediction model was developed and implemented to predict future locations of construction workers.
- A hidden Markov model (HMM) was developed, trained, and applied for future location prediction of construction workers.
- Risk attitude of individual workers was formulated and integrated with the HMM prediction model, and several field experiments were performed which yielded satisfactory results in terms of precision, recall, and accuracy of the risk-incorporated prediction model.
- An Android operating system-based mobile application, PCS2 was designed and tested in the field to evaluate the practical values of the developed methodology in real world scenarios.

Future Work

The findings of this research are sought to contribute to the body of knowledge by enhancing current understanding of how wearable technology and construction data analytics can be coupled in an integrated framework in support of more robust construction safety practices. Future steps in this research will include enabling risk-calibrated trajectory prediction in the presence of multiple site hazards and several workers, as well as creating the mathematical and technological foundations to allow for seamless transition between GPS sensor readings and indoor triangulation to guarantee that safety is not compromised as workers move between outdoor and indoor locations on a constant basis. To facilitate widespread accreditation and adaption of the designed methodology, large sets of trajectory data collected from ubiquitous mobile devices can be stored in a database and sorted by attributes such as geometrical shape and complexity, type of job, as well as worker's age, gender, and level of experience. When

used to train the prediction model, with time, such database of trajectories can significantly enhance the adaptability of the HMM to different types of projects, tasks, and workers, thus creating more reliable results. For example, a specific HMM trained using data from a large group of carpenters in multiple prefabrication shops can be used to predict the trajectory of a new group of carpenters on a jobsite in a different location. Preliminary results of PCS2 inspires future work to further improve several aspects of this application and the underlying scientific methodology to make it more practical in complex construction environments.

REFERENCES

Aanensen, D.M., Huntley, D.M., Feil, E.J. and Spratt, B.G. (2009). "EpiCollect: linking smartphones to web applications for epidemiology, ecology and community data collection." *PloS one*, 4(9), 6968.

Ahmad, A. and Dey, L. (2007). "A k-mean clustering algorithm for mixed numeric and categorical data." *Data & Knowledge Engineering,* 63(2), 503-527.

Akhavian, R. and Behzadan, A. H. (2012). "Remote monitoring of dynamic construction processes using automated equipment tracking." *Construction Research Congress 2012,*1360-1369.

Anagnostopoulos, T., Anagnostopoulos, C.B., Hadjiefthymiades, S., Kalousis, A. and Kyriakakos, M. (2007). "Path prediction through data mining." In *IEEE International Conference on Pervasive Services*. 128-135.

Asahara, A., Maruyama, K., Sato, A. and Seto , K. (2011). "Pedestrian-movement prediction based on mixed Markov-chain model." *Proceedings of the 19th ACM SIGSPATIAL international conference on advances in geographic information systems.* 25-33.

Ashbrook, D. and Starner, T. (2003). "Using GPS to learn significant locations and predict movement across multiple users." *Personal and Ubiquitous computing*, 7(5), 275-286.

Azuma, R. T. (1997). "A survey of augmented reality." *Presence: Teleoperators and virtual environments.* 6(4), 355-385.

Bansal, V.K. (2011). "Application of geographic information systems in construction safety planning." *International Journal of Project Management*, 29(1), 66-77.

Behzadan, A. H., Aziz, Z., Anumba, C. J. and Kamat, V. R. (2008). "Ubiquitous location tracking for context-specific information delivery on construction sites." *Automation in Construction,* 17(6), 737-748.

Behzadan, A.H. and Kamat, V.R. (2011). "Integrated information modeling and visual simulation of engineering operations using dynamic augmented reality scene graphs." *Journal of Information Technology in Construction (ITcon)*, 16(17), 259-278.

Benjaorana, V. and Bhokhab, S. (2010). "An integrated safety management with construction management using 4D CAD model." *Safety Science ,* 48(3), 395-403.

Bennewitz, M., Burgard, W., Cielniak, G. and Thrun, S. (2005). "Learning motion patterns of people for compliant robot motion." *The International Journal of Robotics Research*, 24(1), 31-48.

Bierlaire, M., Chen, J. and Newman, J. (2013). "A probabilistic map matching method for smartphone GPS data." *Transportation Research Part C: Emerging Technologies*, 26, 78-98.

BLS (Bureau of Labor Statistics), (2016). "Employer reported workplace injuries and illnesses – 2015." *Bureau of Labor Statistics*, USDL-16-2056. Online: https://www.bls.gov/news.release/pdf/osh.pdf

Bohn, J. S. and Teizer, J. (2009). "Benifits and barriers of construction project monitoring using high resolution automated cameras." *Journal of Construction Engineering and Management,* 136(6), 632-640.

Boulos, M. N., Anastasiou, A., Bekiaris, E. and Panou, M. (2011). "Geo-enabled technologies for independent living: examples from four European projects." *Technology and Disability,* 23(1), 7-17.

Caldas, C.H., Torrent, D.G. and Haas, C.T. (2005). "GPS technology for locating fabricated pipes on industrial projects." In *Computing in Civil Engineering,* 1-8.

Casella, G. and Berger, R. L. (2002). "Statistical inference (Vol. 2)." 2nd ed. CA: Duxbury Press.

Chae, S. and Yoshida, T. (2010). "Application of RFID technology to prevention of collision accident with heavy equipment." *Automation in Construction,* 19(3), 68-74.

Charness, G. and Gneezy, U. (2012). "Strong evidence for gender differences in risk taking." *Journal of Economic Behavior & Organization,* 83(1), 50-58.

Chen-Fu, L. (2014). "Development of a Navigation System Using Smartphone and Bluetooth Technologies to Help the Visually Impaired Navigate Work Zones Safely." Minnesota Department of Transportation, Report No 2014-12, Online: http://conservancy.umn.edu/bitstream/handle/11299/163204/2014-12.pdf?sequence=1&isAllowed=y

Cheng, T., Venugopal, M., Teizer, J. and Velab, P. A. (2011). "Performance evaluation of ultra wideband technology for construction resource location tracking in harsh environments." *Automation in Construction,* 20(8), 1173-1184.

Choi, P. P. & Hebert , M. (2006). "Learning and predicting moving object trajectory: a piecewise trajectory segment approach." *Robotic Institue*, 337.

Collobert, R., Weston, J., Bottou, L., Karlen, M., Kavukcuoglu, K. and Kuksa, P. (2011). "Natural language processing (almost) from scratch." *Journal of Machine Learning Research*, 12(August), 2493-2537.

Consolvo, S., McDonald, D.W., Toscos, T., Chen, M.Y., Froehlich, J., Harrison, B., Klasnja, P., LaMarca, A., LeGrand, L., Libby, R. and Smith, I. (2008). "Activity sensing in the wild: a field trial of ubifit garden." In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems,* 1797-1806.

Cooper, D., (2003). "Psychology, risk and safety. *Professional Safety*, 48(11), 39-46.

Ding, L.Y., Zhou, C., Deng, Q.X., Luo, H.B., Ye, X.W., Ni, Y.Q. and Guo, P. (2013). "Real-time safety early warning system for cross passage construction in Yangtze Riverbed Metro Tunnel based on the internet of things." *Automation in Construction*, 36, 25-37.

Ergen, E., Akinci, B. and Sacks, R. (2007). "Tracking and locating components in a precast storage yard utilizing radio frequency identification technology and GPS." *Automation in Construction,* 16(3), 354-367.

Gambatese, J. A. (2003). "Safety emphasis in university engineering and construction programs." *Construction safety education and training—A global perspective.*

Gambs, S., Killijian, M.O. and del Prado Cortez, M.N. (2012). "Next place prediction using mobility markov chains." In *Proceedings of the First Workshop on Measurement, Privacy, and Mobility*, 3.

Gardner, M. and Steinberg, L. (2005). "Peer influence on risk taking, risk preference, and risky decision making in adolescence and adulthood: an experimental study." *Developmental psychology*, 41.4, 625.

Ghahramani, Z. (2001). "An introduction to hidden Markov models and Bayesian networks." *International journal of pattern recognition and artificial intelligence.,* 15(1), 9-42.

Glaser, S., Vanholme, B., Mammar, S., Gruyer, D. and Nouveliere, L. (2010). "Maneuver-based trajectory planning for highly autonomous vehicles on real road with traffic and driver interaction." *IEEE Transactions on Intelligent Transportation Systems*, 11(3), 589-606.

Gong, C. and McNally, D. (2004). "A methodology for automated trajectory prediction analysis." In *AIAA Guidance, Navigation, and Control Conference and Exhibit*, 4788.

Goodrum, P. M., McLaren, M. A. and Durfee, A. (2006). "The application of active radio frequency identification technology for tool tracking on construction job sites." *Automation in Construction,* 15(3), 292-302.

Hadikusumo, B.H.W. and Rowlinson, S. (2004). "Capturing safety knowledge using design-for-safety-process tool." *Journal of construction engineering and management*, 130(2), 281-289.

Händel, P., Ohlsson, J., Ohlsson, M., Skog, I. and Nygren, E. (2014). "Smartphone-based measurement systems for road vehicle traffic monitoring and usage-based insurance." *IEEE Systems Journal*, 8(4), 1238-1248.

Hannerz, H., Spangenberg, S. and Tuchsen, F. (2005). "Disability retirement among former employees at the construction of the Great Belt Link." *Journal of Public Health,*119, 301-304.

Hardison, D., Behm, M., Hallowell, M. R. and Fonooni, H. (2014). "Identifying construction supervisor competencies for effective site safety." *Safety Science,* 65, 45-53.

Hare, B., Cameron, I. and Duf, A. R. (2006). "Exploring the integration of health and safety with pre-construction planning." *Journal of Engineering, Construction and Architectural Management,*13(5), 438-450.

Hefner, R. and Breen, P.J. (2004). "Construction vehicle and equipment blind area diagrams." *National Institute for Occupational Safety and Health,* Report: 200-2002, 00563.

Herrera, J.C., Work, D.B., Herring, R., Ban, X.J., Jacobson, Q. and Bayen, A.M. (2010). "Evaluation of traffic data obtained via GPS-enabled mobile phones: The Mobile Century field experiment." *Transportation Research Part C: Emerging Technologies*, 18(4), 568-583.

Hildreth, J., Vorster, M. and Martinez, J. (2005). "Reduction of short-interval GPS data for construction operations analysis." *Journal of construction engineering and management,* 131(8), 920-927.

Hinze, J.W. and Teizer, J. (2011). "Visibility-related fatalities related to construction equipment." *Safety science*, 49(5), 709-718.

Houenou, A., Bonnifait, P., Cherfaoui, V. and Yao, W. (2013). "Vehicle trajectory prediction based on motion model and maneuver recognition." In *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 4363-4369.

Hubbard, B., Wang, H., Leasure, M., Ropp, T., Lofton, T., Hubbard, S. and Lin, S. (2015). "Feasibility Study of UAV use for RFID Material Tracking on Construction Sites." In *Proceeding of ASCE Annual International conference*, 669-676.

Hwang, S. (2012). "Ultra-wide band technology experiments for real-time prevention of tower crane collisions." *Automation in construction,*22, 545-553.

ILEPI (2015). "The cost of construction injuries and fatalities in Illinois, Indiana, and Iowa." Illinois, USA: Illinois Economic Policy Institute, Economic Commentary 28.

Irizarry, J. and Abraham, D.M. (2005). "Application of virtual reality technology for the improvement of safety in the steel erection process." In *Computing in Civil Engineering,* 1-11.

Jamaludin, A., Chee, Y. S. and Ho, C. L. (2009). "Fostering argumentative knowledge construction through enactive role play in second life." *Journal of Computers and Education,* 52(2), 317-329.

Jaselskis, E.J., Anderson, M.R., Jahren, C.T., Rodriguez, Y. and Njos, S. (1995). "Radio-frequency identification applications in construction industry." *Journal of Construction Engineering and Management*, 121(2), 189-196.

Jaselskis, E.J. and El-Misalami, T. (2003). "Implementing radio frequency identification in the construction process." *Journal of Construction Engineering and Management*, 129(6), 680-688.

Jasper, K., Miller, S., Armstrong, C., and Golembiewski, G. (2011). "National evaluation of the safetrip-21 initiative: California connected traveler test bed final evaluation report: Mobile Millennium." Washington DC, USA: ITS Joint Program Office,

US Department of Transportation. Report: FHWA-JPO-11-013. Online: https://ntl.bts.gov/lib/38000/38500/38549/Final_MM_Report_2011_01_20.pdf

Jeung, H., Liu, Q., Shen, H.T. and Zhou, X. (2008). "A hybrid prediction model for moving objects." In *IEEE 24th International Conference on Data Engineering,* 70-79.

Joachims, T. (2002). "Optimizing search engines using clickthrough data." In *Proceedings of the eighth ACM SIGKDD international conference on Knowledge discovery and data mining*, 133-142.

Johnson, D. A. and Mohan, M. T. (2011). "Driving style recognition using a smartphone as a sensor platform." 14th International IEEE Conference, 1609-1615.

Kang, L.S., Moon, H.S., Kim, H.S., Choi, G.Y. and Kim, C.H. (2011). "Development of 5D CAD system for visualizing risk degree and progress schedule for construction project." In *Computing in Civil Engineering,* 690-687.

Khoury, H. M. and Kamat, V. R. (2009). "Evaluation of position tracking technologies for user localization in indoor construction environments." *Automation in Construction,* 18(4), 444-457.

Kim, Y. K., Han, J. & Park, H. (2015). "Trajectory Prediction for Using Real Data and Real Meteorological Data." *In Ubiquitous Computing Application and Wireless Sensor*, 89-103.

Kissonergis, P. (2015). "The HUBGlobal news and insights to inform, engage and educate." Online: http://www.smsglobal.com/thehub/smartphone-ownership-usage-and-penetration/

Kononenko, I. (2001). "Machine learning for medical diagnosis: history, state of the art and perspective." *Artificial Intelligence in medicine,* 23(1), 89-109.

Kotz, S., Balakrishnan, N. and Johnson, N. L. (2000). "Continuous multivariate distributions, vol. 1: models and applications." 2nd ed. NY: WIley.

Lane, N.D., Miluzzo, E., Lu, H., Peebles, D., Choudhury, T. and Campbell, A.T. (2010). "A survey of mobile phone sensing." *IEEE Communications magazine*, 48(9).

Laasonen, K. (2005). "Route prediction from cellular data." In *Workshop on Context-Awareness for Proactive Systems (CAPS), Helsinki, Finland*, 1617.

Le, Q.T., Pedro, A.K.E.E.M., Lim, C.R., Park, H.T., Park, C.S. and Kim, H.K. (2015). "A framework for using mobile based virtual reality and augmented reality for experiential construction safety education." *International Journal of Engineering Education*, 31(3), 713-725.

Lin, K.Y., Son, J.W. and Rojas, E.M. (2011). "A pilot study of a 3D game environment for construction safety education." *Journal of Information Technology in Construction (ITcon)*, 16(5), 69-84.

Loosemore, M. and Andonakis, N. (2007). "Barriers to implementing OHS reforms – The experience of small subcontractors in the Australian construction industry." *International Journal of Project Management,* 25, 579−588.

Lu, M., Chen, W., Shen, X., Lam, H.C. and Liu, J. (2007). "Positioning and tracking construction vehicles in highly dense urban areas and building construction sites." *Automation in construction*, 16(5), 647-656.

Lustreck, M. and Kauluza, B. (2009). "Fall detection and activity recognition with machine learning." *Informatica.,* 33(2), 197-204.

Mainprice, J. and Berenson, D. (2013). "Human-robot collaborative manipulation planning using early prediction of human motion." In *IEEE/RSJ International Conference on Intelligent Robots and Systems,* 299-306.

Mathew, W., Raposo, R. and Marins, B. (2012). "Predicting future locations with hidden Markov models." In *Proceedings of the 2012 ACM conference on ubiquitous computing,* 911-918.

Meyn, S. P. & Tweedie, R. L. (2012). "Markov chains and stochastic stability." Springer Science & Business Media..

Miluzzo, E., Lane, N.D., Fodor, K., Peterson, R., Lu, H., Musolesi, M., Eisenman, S.B., Zheng, X. and Campbell, A.T. (2008). "Sensing meets mobile social networks: the design, implementation and evaluation of the cenceme application." In *Proceedings of the 6th ACM conference on Embedded network sensor systems*, 337-350.

Minetti, A.E. (2001). "Invariant aspects of human locomotion in different gravitational environments." *Acta astronautica*, 49(3), 191-198.

Monreale, A., Pinelli, F., Trasarti, R. and Giannotti, F. (2009). "Wherenext: a location predictor on trajectory pattern mining." In *Proceedings of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining*, 637-646.

Montaser, A. and Moselhi, O. (2014). "RFID indoor location identification for construction projects." *Automation in Construction,* 39, 167-179.

Mun, M., Reddy, S., Shilton, K., Yau, N., Burke, J., Estrin, D., Hansen, M., Howard, E., West, R. and Boda, P. (2009). "PEIR, the personal environmental impact report, as a platform for participatory sensing systems research." In *Proceedings of the 7th international conference on Mobile systems, applications, and services*, 55-68.

Nath, N. D., Akhavian, R. and Behzadan, A. H. (2017). "Ergonomic analysis of construction worker's body postures using wearable mobile sensors." *Applied Ergonomics,* 62, 107-117.

Navon, R. and Goldschmidt, E. (2003). "Can labor inputs be measured and controlled automatically?." *Journal of Construction Engineering and Management*, 129(4), 437-445.

Navon, R., Goldschmidt, E. and Shpatnitsky, Y. (2004). "A concept proving prototype of automated earthmoving control." *Automation in Construction,* 13(2)*,* 225-239.

Navon, R. and Kolton, O. (2006). "Model for Automated Monitoring of Fall Hazards in Building Construction." *Journal of Construction Engineering and Management*, 132(7)*,* 733-740.

Navon, R. and Shpatnitsky, Y. (2005). "Field experiments in automated monitoring of road construction." *Journal of Construction Engineering and Management*, 131(4),487-493

OSHA (2016). United States Department of Labor. Online: https://www.osha.gov/oshstats/commonstats.html

Pereira, C., Guenda, L. and Carvalho, N.B. (2011). "A smart-phone indoor/outdoor localization system." In *International conference on indoor positioning and indoor navigation (IPIN)*, 21-23.

Perera, L. P., Oliveira, P. and Guedes, S. C. (2012). "Maritime traffic monitoring based on vessel detection, tracking, state estimation, and trajectory prediction." *Intelligent Transportation Systems, IEEE Transactions ,* 13(3), 1188-1200.

Persad-Maharaj, N., Barbeau, S.J., Labrador, M.A., Winters, P.L., Perez, R. and Georggi, N.L. (2008). "Real-time travel path prediction using GPS-enabled mobile phones." In *Proceeding of 15th World Congress on Intelligent Transportation Systems.*

Peters, J., Vijayakumar, S. and Schaal, S. (2003). "Reinforcement learning for humanoid robotics." In *Proceedings of the third IEEE-RAS international conference on humanoid robots,* 1-20.

Petrushin, V.A. (2000). Hidden markov models: Fundamentals and applications. In *Online Symposium for Electronics Engineer*.

Pincus, S.M. (1991). "Approximate entropy as a measure of system complexity." *Proceedings of the National Academy of Sciences*, 88(6), 2297-2301.

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

Razavi, S. and Moselhi, O. (2011). "Indoor construction location sensing using low cost passive RFID tags." In *The 2011 ASCE Construction Research Congress,* Ottawa, Ontario.

Reese, C. D. & Eidson, J. V. (2006). "Handbook of OSHA construction." In: *Handbook of OSHA construction.* Boca Raton, FL: CRC/Taylor & Francis.

Sacks, R., Navon, R., and Goldschmidt, E. (2003). "Building project model support for automated labor monitoring." *Journal of computing in civil engineering*, 17(1), 19-27.

Sacks, R., Rozenfeld, O. and Rosenfeld, Y. (2009). "Spatial and temporal exposure to safety hazards in construction." *Journal of Contruction Engineering and Management,* 135(8), 726-736.

Saidi, K.S., Lytle, A.M. and Stone, W.C. (2003). "Report of the NIST workshop on data exchange standards at the construction job site." In *Proceeding of 20th International Symposium on Automation and Robotics in Construction (ISARC)*, 617-622.

Salminen, S. (2004). "Have young workers more injuries than older ones? An international literature review." *Journal of safety research*, 35(5)*,* 513-521.

Shirazi, A. and Behzadan, A.H. (2015). "Content Delivery Using Augmented Reality to Enhance Students' Performance in a Building Design and Assembly Project." *Advances in Engineering Education*, 4(3), 1-24.

Sampaio, A.Z., Ferreira, M.M., Rosário, D.P. and Martins, O.P. (2010). "3D and VR models in Civil Engineering education: Construction, rehabilitation and maintenance." *Automation in Construction*, 19(7), 819-828.

Song, J., Haas, C.T., Caldas, C., Ergen, E. and Akinci, B. (2006). "Automating the task of tracking the delivery and receipt of fabricated pipe spools in industrial projects." *Automation in Construction*, 15(2), 166-177.

Song, L. and Eldin, N. N. (2012). "Adaptive real-time tracking and simulation of heavy construction operations for look-ahead scheduling." *Automation in Construction,* 27, 32-39.

Spangenberg, S., Hannerz, H. and Tuchsen, F. (2005). "Hospitalized injuries among bridge and tunnel construction workers." *Journal of Construction Management and Economics,* 23(3), 237-240.

Stamp, M. (2004). "A revealing introduction to hidden Markov models." Department of Computer Science, San Jose State University.

Tanizaki, H. (1996). "Nonlinear Filters: Estimations and Application." Newyork: Springer-Verlag.

Teizer, J., Ben, S. A., Clare, E. F. and Jimmie, H. (2010). "Autonomous pro-active real-time construction worker and equipment operator proximity safety alert system." *Automation in construction,* 19(5), 630-640.

Teizer, J., Lao, D. and Sofer, M. (2007). "Rapid automated monitoring of construction site activities using ultra-wideband." In *Proceedings of the 24th International Symposium on Automation and Robotics in Construction,* Kochi, Kerala, India, 19-21.

Thiagarajan, A., Ravindranath, L., LaCurts, K., Madden, S., Balakrishnan, H., Toledo, S. and Eriksson, J. (2009). "VTrack: accurate, energy-aware road traffic delay estimation using mobile phones." In *Proceedings of the 7th ACM Conference on Embedded Networked Sensor Systems*, 85-98.

Tixier, A. J., Hallowell, M. R., Rajagopalan, B. and Bowman, D. (2016). "Automated content analysis for construction safety: A natural language processing system to extract precursors and outcomes from unstructured injury reports." *Automation in Construction,* 62, 45-56.

Vasquez, D. and Fraichard, T. (2004). "Motion prediction for moving objects: a statistical approach." In *Proceedings of IEEE International Conference on Robotics and Automation,* 4, 3931-3936.

Wadick, P. (2007). "Safety culture among subcontractors in the NSW domestic housing industry." *Journal of Occupational Health and Safety in Australia and New Zealand,* 23(2), 143-152.

Waldron, I., McCloskey, C. and Earle, I. (2005). "Trends in gender differences in accident mortality: Relationships to changing gender roles and other societal trends." *Demographic Research,* 13, 415-454.

Woo, S., Jeong, S., Mok, E., Xia, L., Choi, C., Pyeon, M., and Heo, J. (2011). "Application of WiFi-based indoor positioning system for labor tracking at construction sites: A case study in Guangzhou MTR." *Automation in Construction*, 20(1), 3-13.

Wu, H., Tao, J., Li, X., Chi, X., Li, H., Hua, X. and Chen, N. (2013). "A location based service approach for collision warning systems in concrete dam construction." *Safety science*, 51(1), 338-346.

Xu, H. and Gupta, S. (2009). "The effects of privacy concerns and personal innovativeness on potential and experienced customers' adoption of location-based services." *Electronic Markets.,* 19(2-3), 137-149.

Yang, J., Cheng, T., Teizer, J., Vela, P. A., and Shi, Z. K. (2011). "A performance evaluation of vision and radio frequency tracking methods for interacting workforce." *Advanced Engineering Informatics*, 25(4), 736-747.

Yentes, J. M., Hunt, N., Schmid, K. K., Kaipust, J. P., McGrath, D., and Stergiou, N. (2013). "The appropriate use of approximate entropy and sample entropy with short data sets." *Annals of biomedical engineering*, 41(2), 349-365.

Zhang, S., Teizer, J., Lee, J. K., Eastman, C. M., and Venugopal, M. (2013). "Building information modeling (BIM) and safety: Automatic safety checking of construction models and schedules." *Automation in Construction*, 29, 183-195.

Zhang, S., Teizer, J., Pradhananga, N. and Eastman, C. M. (2015). "Workforce location tracking to model, visualize and analyze workspace requirements in building information models for construction safety planning." *Automation in Construction,* 60, 74-86.

Zhao, X., and Tian, D. (2012). “The architecture design of streaming media applications for Android OS.” In *IEEE 3rd International Conference on Software Engineering and Service Science (ICSESS),* 280-283.