7-2024

Development of a Rule-Based Monitoring System for Autonomous Heavy Equipment Safety

Amirpooya Shirazi
University of Nebraska-Lincoln

DEVELOPMENT OF A RULE-BASED MONITORING SYSTEM FOR AUTONOMOUS HEAVY EQUIPMENT SAFETY

by

Amirpooya Shirazi

A THESIS

Presented to the Faculty of

The Graduate College at the University of Nebraska

In Partial Fulfillment of Requirements

For the Degree of Master of Science

Major: Construction Engineering and Management

Under the Supervision of Professor Kyungki Kim

Lincoln, Nebraska

July, 2024

DEVELOPMENT OF A RULE-BASED MONITORING SYSTEM FOR AUTONOMOUS HEAVY EQUIPMENT SAFETY

Amirpooya Shirazi, M.S.

University of Nebraska, 2024

Advisor: Kyungki Kim

Roadway construction work zones are constantly exposed to interactions among construction equipment, workers, and vehicles. Furthermore, ensuring safety in these areas is considered a challenging task due to the complexity of the environment. As shown in the rising trend of fatal accidents in roadway work zones, current OSHA regulations in construction safety are insufficient in effectively detecting unsafe situations and mitigating the risks. Furthermore, best practices, such as internal traffic control planning (ITCP), exhibit critical limitations requiring continuous monitoring of active work zones as well as adjustments to the site coordination plans due to the dynamic nature of work zone environments. To overcome the stated challenges, this study proposes an innovative solution by integrating sensing and perception technologies of Autonomous Vehicles (AVs) to detect unsafe situations around heavy construction equipment by integrating vision-based sensors that can produce contextual information about the situation around the heavy equipment. To perceive such information, a Robotics Operating System (ROS) based algorithm has been developed, along with various PointCloud processing techniques aimed to identify and report the location of workers and other vehicles. Moreover, a simulation-based methodology was introduced aimed to devise an integrated sensor placement scheme, to facilitate a thorough sensor deployment strategy for monitoring unsafe zones using the designed ROS-based algorithms through an interconnected network of vision-based sensors. Furthermore, the designed sensor arrangement combined with the pipelines underwent three major experiments, illustrating the work zone within an isolated environment. Firstly, the experiments aimed to gauge the efficacy of both human and vehicle localization components by comparing reported locations with predetermined ground truths. Secondly, they involved delineating various human trajectory scenarios to analyze the tracking capabilities of the framework. Lastly, the experiments entailed simulating sequential entrances and exits into unsafe zones to assess the framework's sensitivity and accuracy in monitoring these designated zones. By providing the equipment with a precise understanding of its environment, the framework has proven its potential to

enhance safety protocols and prevent unforeseen and hazardous situations. Additionally, this study represents a critical step toward the integration of autonomous rules and technologies into roadway construction work zones.

To our friend, Ali Zolfaghari Sichani. Rest in peace.

Acknowledgments

First and foremost, I would like to thank my mother and my father for their invaluable support in helping me reach this stage. I am also profoundly grateful to my sister, Arezoo, for her unwavering support from the early stages of my studies until now, which paved the way for my academic success.

I would like to extend my heartfelt thanks to my companion, Mobina, whose indescribable emotional support helped me navigate one of the most challenging periods of my life. I am also grateful to my friends Mohammad Amin, Hossein, Pejman, and Hiva for their assistance in conducting various field tests, which significantly contributed to the progress of this study.

Lastly, I would like to thank my advisor, Dr. Kyungki Kim, for his invaluable guidance and support throughout my research.

This publication was made possible through cooperative agreement number U60-OH009762 from the National Institute for Occupational Safety and Health (NIOSH) to CPWR–The Center for Construction Research and Training. The contents are solely the responsibility of the authors and do not necessarily represent the official views of CPWR or NIOSH.

Contents

List of Tables

List of Figures

CHAPTER 1

Introduction

The construction industry has consistently been a prominent sector within the U.S. economy. As the national population has increased, the industry has experienced substantial growth in the total value of construction projects. Data from the U.S. Census Bureau indicates that from 2002 to 2023, the value of construction projects increased by over $1.2 trillion [1]. Furthermore, 2023 alone witnessed a remarkable surge, with the total value of construction projects rising by over $200 billion from the beginning to the end of the year. It should be noted that the roadway construction industry has also been part of this trend. According to 2023 data from the American Road & Transportation Builders Association (ARTBA), highway construction activities saw a 17 percent increase in value from November 2022 to November 2023 [2], which further rose to 19 percent by the end of December 2023 [3]. With significant growth over the past decades and recent years, the efficiency and effectiveness of safety rules and regulations within roadway projects have become increasingly important due to the corresponding rise in employment. According to recent data from the U.S. Bureau of Labor Statistics, employment within the heavy and civil engineering construction industry—which includes highway and roadway construction—reached over 1.1 million workers by the end of 2023 [4]. This represents an increase of 45,000 workers since the beginning of 2023 and 222,000 workers since 2002. This growth underscores the necessity of monitoring and regularly updating mandatory safety protocols to prevent an increase in injuries and fatalities.

Considering the roadway construction industry, several organizations are responsible for establishing and continually evaluating safety rules by analyzing trends in injuries and fatalities within work zones. To provide a better understanding of these agencies, it is essential to first introduce the

relevant organizations and examine their roles and the significance of their contributions to existing safety regulations.

1.1 Responsible Organizations

Several organizations monitor and analyze trends in transportation-related injuries and fatalities, particularly within roadway construction work zones involving motor vehicles, workers, and heavy equipment. The Occupational Safety and Health Administration (OSHA), under the Department of Labor (DOL), sets and enforces standards to protect worker health and safety, aiming to eliminate on-the-job injuries, illnesses, and fatalities. OSHA provides guidelines for traffic control, worker visibility, and protective measures, conducts inspections, and enforces compliance [5]. The National Institute for Occupational Safety and Health (NIOSH), under the CDC within the Department of Health and Human Services, analyzes statistical information from construction work zones and conducts research to prevent work-related injuries, illnesses, and fatalities [6]. NIOSH collaborates with OSHA, notably through the NIOSH/OSHA Alliance Program, to enhance occupational safety by providing information, guidance, and training resources [7]. The Federal Highway Administration (FHWA), part of the U.S. Department of Transportation, develops and promotes safety standards for highway construction zones, focusing on worker visibility and traffic control planning [8]. While its guidelines are suggestions from DOT studies, they serve as primary sources for OSHA to establish or update enforceable standards.

1.2 Current Safety Protocols

Among the relevant organizations, OSHA holds the primary authority to establish or modify mandatory regulations concerning roadway work zone safety. Among these regulations, several focus on the interaction between workers and vehicles, including both motor vehicles and construction equipment. These rules are crucial for managing the movement of heavy construction equipment and motor vehicles through the roadway work zone which are legislated on top of the Manual on Uniform Traffic Control Devices (MUTCD) guideline. The Manual on Uniform Traffic Control Devices (MUTCD) sets forth a uniform set of minimum standards established by the FHWA for the use of traffic control

devices, such as signs, barricades, flags, and other equipment. These standards are designed to ensure the safe movement of motor vehicles through construction work zones and to provide a secure environment for heavy construction equipment to operate. To put the MUTCD-related practices into play, OSHA introduced two sets of guidelines: External Traffic Control and Internal Traffic Control planning regulations [5]. External Traffic Control planning, also known as Temporary Traffic Control (TTC) planning, primarily focuses on providing a reasonably safe and efficient movement of road users through or around work zones while protecting workers [9]. Internal Traffic Control Planning (ITCP) rules serve as a tool for employers to coordinate the flow of construction vehicles, equipment, and workers on foot within the work zone, ensuring their safe and efficient movement in close proximity to one another [10], [11].

As FHWA and NIOSH conduct ongoing research on the impacts of the revisions on the regulations associated with the TTCs and ITCPs, OSHA also revises them in a way to ensure a safer environment for workers by keeping pace with evolving safety practices and technological advancements. However, there remains a lack of effective implementation of these safety protocols. Recent studies by NIOSH indicate an increasing trend in the number of injuries and fatalities within roadway work zones, primarily due to incidents involving workers and vehicles. While several studies have highlighted the positive impact of successful traffic control planning systems on reducing accident occurrences within roadway work zones [12], it is worth noting that continuously updating such plans within the job site can be time-consuming. This is primarily due to the dynamic nature of roadway construction work zones, which necessitates ongoing modifications to ensure their effectiveness. Furthermore, it is essential to have a competent person who will regularly monitor ongoing activities to identify any unsafe situations while implementing road planning regulations within the work zone [13], which may also be subject to human error.

According to a recent report published by the National Work Zone Safety Information Clearinghouse—a collaborative project funded by ARTBA, FHWA, and the Texas A&M Transportation Institute—fatal crashes in work zones have increased from 557 in 2012 to 874 in 2021. The percentage of all fatal crashes occurring in work zones has also risen slightly, from 1.8% in 2012 to 2.2% in 2021. Additionally, work zone fatalities have increased from 619 in 2012 to 956 in 2021, equating to nearly three persons per day being killed in work zones in 2021. Approximately four out of every five work zone fatalities involve a driver or passenger of a vehicle. Furthermore, estimated injuries in

work zones have risen from 31,000 in 2012 to 42,000 in 2021, which translates to approximately 112 injuries per day in 2021 [14]. Moreover, a recent report from NIOSH indicated that from 2003 to 2020, there were a total of 2,222 worker-related deaths reported at road construction sites, averaging 123 deaths per year [15]. Furthermore, the National Work Zone Safety Information Clearinghouse reported that despite a reduction in the total number of fatal injuries to roadway construction work zone workers, which decreased from its peak of 135 in 2019 to 94 in 2022, the percentage of fatalities involving workers on foot being struck by a vehicle increased from 44.4% to 53.2% during the same period. The report also noted that over 75% of the fatalities from 2020 to 2022 were associated with interactions between workers and motor vehicles [16].

These trends indicate that despite recent modifications to OSHA regulations, the effectiveness of these rules and regulations remains insufficient. Consequently, there is a pressing need for innovative approaches and technologies to enhance safety protocols within roadway construction work zones. This underscores the importance of state-of-the-art research focused on implementing additional safety standards in work zones and highlights how these research approaches are designed to improve worker safety.

1.3 State-of-the-art research

To address the escalating rates of injuries and fatalities within various types of construction work zones, including roadway zones, and the inadequacy of current practices and regulations provided by responsible organizations, several studies have proposed innovative approaches. These include 1) Risk mitigation and assessment methods tailored for construction work zones. 2) Implementation of automated monitoring systems utilizing non-perceptual sensing technologies (proximity solutions) or perceptual sensing technologies (perception-based solutions). 3) Adoption of simulation-based techniques integrating Building Information Modeling (BIM) simulation approaches with emerging technologies such as Virtual Reality (VR) or Augmented Reality (AR) to dynamically inspect work zones. 4) Implementation of automated inspection techniques through the integration of Unmanned Ground Vehicles (UGVs) and Unmanned Aerial Vehicles (UAVs) with perceptual sensing technologies.

1.3.1 Fundamental Studies: Risk Mitigation and Assessment

Since the advent of the construction industry, the issues of safety and hazard prevention have been the subject of extensive examination and analysis. These concerns are critical in construction projects due to the paramount importance of protecting workers' lives and the significant financial implications associated with accidents. Moreover, effective safety measures can substantially influence the contractor's reputation [17]. Therefore, several studies have aimed to improve safety protocols by identifying and mitigating the risks associated with various types of construction work zones.

Among the earliest research in this area, *Gambatese and Hinze* [18] provided pre-construction recommendations for specific tasks, supported by OSHA regulations, during the design phase of projects. These efforts aimed to reduce the number of exposures leading to hazardous situations. Although this study was a preliminary effort to enhance existing safety regulations, it represented an effective approach to identifying and analyzing the risks associated with various types of hazards by categorizing them based on the work zone type. Furthermore, *Gangollels et al.* [17] undertook an analysis of hazard exposure likelihood across various construction tasks, employing partial categorization. They devised a model aimed at quantifying the risk probabilities associated with these tasks, drawing upon the framework of expected values in their empirical investigation. This approach enabled them to assess the likelihood of exposure to various identified risks during the project's design phase.

Moreover, several studies have endeavored to classify various types of hazards within specific construction work zones and quantify their likelihood of occurrence by employing more sophisticated statistical models. These methodologies have facilitated the identification of the severity levels associated with different construction tasks, thereby highlighting the necessity for more comprehensive observation, particularly for those deemed most hazardous. As a result, *Hollowell and Gambatese* [19] proposed a Delphi-based methodology to assess the risk reduction attributed to manual safety inspections. They determined observation criteria for various types of hazards within this framework. They quantified the reduction rates by conducting a case study involving 13 experts who participated in multiple rounds of the Delphi process. This effort culminated in measuring risk reduction values, expressed as the ratio of the raw severity reduction score to the worker-hours per incident. Furthermore, *Zeng et al.* [20] introduced a model known as "Failure Modes and Effects Analysis

(FEMA)" for assessing the risk associated with OSHA predetermined hazards by categorizing various hazards associated with causes such as Use of Equipment, Falling Objects, Electrocution, and others to determine the Risk Priority Number (RPN) indicator. This metric is calculated as the product of occurrence, severity, and detection ratings, as validated through a localized case study conducted in China. In general, several studies have successfully approached the analysis of task hazards within construction work zones so far, exemplified by a recent study conducted by *Esmailli et al.* [21] which builds upon prior research conducted by *Hallowell and Gambatese* [19]. Therefore, such advancements represent the ongoing need to continuously conduct studies on risk assessment within construction work zones.

1.3.2 Automated Monitoring: Sensing Technologies

Generally, the automation of safety monitoring through sensing technologies involves developing a pipeline that integrates input data from various types of sensors, which are systematically selected and combined within a structured framework. Subsequently, this data undergoes processing through a series of algorithms, resulting in interpretable outputs in human language. These outputs can then be employed to verify adherence to established safety protocols requiring detailed monitoring measured by the risk assessment studies. As illustrated in Figure 1.1, pipelines employing sensing technologies are primarily categorized into two sections based on the types of devices they use: non-perceptual and perceptual sensing technologies. The devices in each section generate different types of data, which can significantly affect the outcomes. The nature of the data type determines the level of detail and comprehensiveness of the interpretation, which in turn necessitates more computational power to avoid delays and ensure real-time outcomes, a critical aspect in the safety monitoring sector. While Figure 1.1 provides an overview of the concept of sensing technologies, the subsequent subsections will offer a more detailed exploration of non-perceptual and perceptual sensing technologies.

Over the past two decades, technological advancements have significantly improved the understanding and application of wireless communication using low-frequency waves. This has led to the development of electromagnetic wave sensors that can transmit and receive signals over limited distances, facilitating the creation of algorithmic frameworks for interpreting distance-related data. Radio Frequency Identification (RFID) and Bluetooth-based sensors have become key tools in safety research, particularly in construction zones, to establish protocols responding to various hazardous

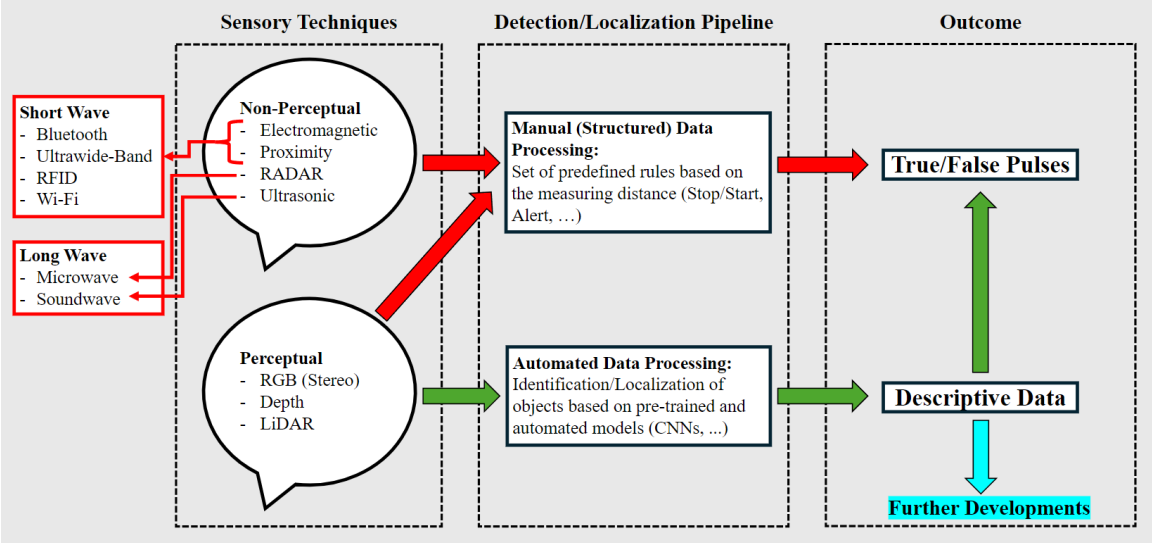

Figure 1.1: *Color-filled arrows visualize the potential directions sensory techniques can take. Green arrows indicate the direction of this study.*

scenarios. Notable studies include *Teizer et al.*'s [22] localization pipeline using Ultra-wideband technology with Bluetooth-based devices to pinpoint personnel positions on construction sites. *Roofigari-Esfahan et al.* [23], [24] developed a smart vest with Bluetooth and GPS for worker localization. *Park et al.* [25], [26] introduced a proximity-based pipeline using electromagnetic sensors to measure worker proximity to equipment, and a simulation-based pipeline to identify hazardous scenarios with Bluetooth tracking sensors. *Fang et al.* [27] combined RFID sensor deployment with simulation-based safety assessments within a Building Information Modeling (BIM) framework, enabling worker presence detection in hazardous zones.

On the other hand, Recent advancements in Artificial Intelligence (AI) and Unsupervised Machine Learning (ML), particularly Convolutional Neural Networks (CNNs), have enabled the analysis of complex and large-scale data from sensors like RGB or depth images. These technologies allow computers to generate descriptive information about their surroundings, including object existence, proximities, and interactions, using perceptual sensors like cameras, depth sensors, and LiDAR. Researchers have integrated these technologies into safety monitoring methodologies, creating automated systems that yield detailed outcomes. Innovative studies in this field include *Jeelani et al.*'s [28] hazard detection pipeline using an automated object detection algorithm to classify and segment construction equipment based on image-based RGB data, assessing worker proximity to hazards. *Tang et al.* [29] developed an image-based worker detection algorithm to identify worker

exposure to potential hazards, generating descriptive text about the worker's location, environment, and potential hazards. *Chen et al.* [30] utilized deep learning algorithms to detect and classify excavator activities, identifying hazards related to worker-equipment interactions, thereby enhancing safety analysis and preventive measures on construction sites.

1.3.3 Digital Twins: BIM, VR, and AR Applications

With the emergence of Building Information Modeling (BIM) and simulation-based construction methodologies in recent decades, numerous studies have sought to address the limitations of safety monitoring through the use of simulation-based approaches within work zones. Implementing systematic and simulated versions of entire projects, encompassing schedules and detailed task breakdowns, can facilitate the monitoring of safety and task execution. The dynamic nature of these simulations allows for modifications throughout the process, thereby enhancing robustness. This adaptability enables responsible personnel to observe tasks and inspect progress effectively within the work zone. Consequently, several studies have aimed to develop interactive monitoring systems that alleviate worker pressure while improving task execution accuracy.

A notable study by *Getuli et al.* [31] designed a safety analysis framework that leverages BIM and virtual reality (VR) devices to sequentially gather information from the workers' surrounding environment within the work zone. This framework modifies hazardous locations and protocols by updating safe and unsafe zones within the simulation, which are then communicated to the workers in real-time through their VR devices. Additionally, the generalized concept of utilizing simulation-based monitoring systems involves integrating and analyzing real-world data in real-time using various data collection devices to continuously modify responses. This approach has led researchers to introduce the concept of Construction 4.0 [32].

1.3.4 Dynamic Monitoring: Unmanned Vehicle's Application

Nowadays, Researchers have proposed the utilization of robots as an innovative approach to enhance safety monitoring and inspection within work zones, employing both manual and automated control systems. Robot manipulation encompasses a fusion of sensory techniques, predominantly perceptual sensory methods, simulation-based technologies, and the application of deep learning algorithms.

These elements collectively process input data from sensors affixed to the robot, enabling modification or cross-referencing with simulations to yield descriptive information. This information serves as the basis for implementing either a checking system or a monitoring framework throughout construction projects, while the robot is programmed to navigate safely within the work zone. Furthermore, various studies have explored the use of robots to perform diverse tasks while ensuring interaction with workers within a secure environment. Such utilization of robots can be broadly categorized into two main sections: aerial navigation employing Unmanned Aerial Vehicles (UAVs) or drones, and ground-based navigation using Unmanned Ground Vehicles (UGVs).

In the realm of Unmanned Aerial Systems (UAS), *Kim et al.* [33] developed an image-based pipeline aimed at localizing construction equipment such as excavators and loaders, along with workers, to assess the proximity of equipment to the workforce and thus evaluate safety protocols within the work zone. This methodology involved projecting images onto a global map to determine the spatial relationship between equipment and workers. UAVs were employed to capture aerial imagery, subsequently used to retrain an image-based object detection model known as *You Only Look Once (YOLO)* [34]. This model facilitated the identification of the aforementioned objects. By comparing the physical location of the UAV with predetermined global coordinates, the detected objects were localized, enabling the evaluation of safety protocols pertaining to distance. This approach not only highlighted the sensitivity of the situation but also facilitated the prediction of potential hazards. Furthermore, *Kim et al.* [35] devised a safety monitoring system founded on a digitalized iteration of an Internal Traffic Control Plan (ITCP) model tailored for a specific roadway work zone. This system aimed to gauge unsafe conditions by leveraging retrained YOLO to ascertain the positions of heavy construction equipment and workers. A manually controlled UAV was utilized to capture aerial imagery throughout the work zone, which was subsequently analyzed in conjunction with the pre-existing digitalized ITCP model. Moreover, *Jaycob-Loyola et al.* [36] devised an aerial inspection pipeline aimed at quantifying the completion percentage of structural projects. This methodology involved collecting both image and depth data utilizing a flying drone and analyzing these structured inputs alongside a pre-designed 3D model of the building. The integration of these elements facilitated the establishment of a robust physical progress monitoring system for building construction projects.

Despite the successful utilization of UAVs in enhancing safety protocols, recent research has

equally focused on the implementation of UGVs for monitoring, inspection, and task execution [37]. *Ibrahim et al.* [38] developed a Reinforcement Learning (RL) model aimed at optimizing the navigation of a 4-wheeled UGV robot within an indoor construction work zone. This model facilitated the generation of a high-quality map of the area by collecting depth data during the navigation process. Additionally, *Liu et al.* [39] introduced a pioneering robot-human interaction approach, presenting a pipeline for adapting the behavior of robots while navigating through construction work zones. This innovative method involves analyzing the body and mental state of existing workers by interpreting information gathered from sensory devices affixed to them. Such an approach represents a pioneering safety protocol that holds promise for implementation in work zones, facilitating collaboration between robots and human workers.

1.4 Thesis Organization

The rest of this study is structured as follows: Chapter 2 reviews relevant literature on implementing autonomous sensing systems, evaluating their processing models and algorithms as well as their outcomes accuracies which caused this study's direction. Chapter 3 outlines the problem statement, focusing on the essential elements needed for developing autonomous safety monitoring frameworks for construction equipment. Chapter 4 details the solution implemented to attain automated monitoring objectives. Chapter 5 presents a real-world demonstration of the methodology implementations within a case study provided by the Nebraska Department of Transportation (NDOT) in addition to the testing scenario's description. Chapter 6 represents the test results, evaluating the accuracy and safety of the developed framework through the designed testing scenarios. Chapter 7 provides a conclusion to the study and the potential future directions.

CHAPTER 2

Literature Review

During the last two decades, several studies have attempted to address the limitations associated with manual safety monitoring within construction work zones by utilizing various approaches. These efforts have resulted in multiple research directions, each with its advantages and disadvantages, which are described in this chapter.

2.1 The Importance of Automated Safety Monitoring

From the early stages of safety monitoring in construction job sites, numerous state-of-the-art research studies have addressed the inefficiency of existing regulations by exploring various directions. These efforts span from early risk assessment practices, which involved evaluating the sensitivity of construction tasks based on predetermined OSHA regulations [17], [18], to the use of statistical models that measure the importance and danger levels associated with these tasks [20], [21]. Recent studies have focused on implementing supervised models on enhanced datasets, which are created by converting expert opinion surveys into a computer-readable format [40]. This progression has led to an increasingly accurate understanding, rating, and scoring of dangers in various construction work zones, guiding safety experts in assessing the risks of construction tasks. However, the manual effort required to observe and address tasks identified as more dangerous remains a limiting factor. Additionally, human error in recognizing special or unforeseen situations can further constrain safety measures. Consequently, with the recent developments in AI and decision-making models, it is logical to consider semi- to fully automated frameworks for observing safety matters regardless of the

growing imperative to leverage such categorizing models within the safety sectors of the construction industry [41].

Furthermore, simulation-based research has emerged as one of the most successful and expanding directions in safety monitoring systems within construction work zones [31]. However, practical utilization of these technologies faces significant limitations. These include the extensive education and training required for professionals and workers to become proficient with such technologies. Wearable visualizer devices, such as VR and AR, are not only costly to implement but also present significant challenges in terms of compatibility for practitioners. As a result, the gap between traditional construction methods and the novel approach known as Construction 4.0 [32] remains substantial.

Additionally, the utilization of automated inspection techniques, such as implementing UAVs [34]–[36] or UGVs [37]–[39] in the construction environment to meet safety observation requirements seems to be challenging. This challenge primarily arises from workers' unfamiliarity with these novel technologies. Further studies are needed to develop a safe framework for interactions between automated equipment and the workforce to prevent distractions or destructive encounters that could lead to additional hazardous situations. Moreover, the urgency of safety monitoring is heightened by the fact that accidents can occur in mere seconds. This necessitates thorough inspections by automated systems, which in turn requires increasing the number of active robots within the work zone. However, this increase can lead to more encounters between robots and workers, potentially creating further safety issues.

Regardless of the aforementioned challenges and limitations, a key concept in utilizing robots (ground or aerial) for automatically observing, navigating, or executing hazardous construction tasks within work zones is the application of sensing technologies. Designing a computer-brain entity that can process information gathered from various sensors attached to it appears to be a highly effective solution for observation. This implies that a robot equipped with multiple types of sensors, such as RGB, depth, LiDAR, and others, can efficiently interpret and analyze environmental data using automated decision-making frameworks, including convolutional neural networks (CNNs), in its central processor. While this advancement enables robots to navigate, inspect, or execute tasks with varying levels of autonomy, similar capabilities can be achieved with heavy construction equipment. Autonomy in construction equipment can range from Level 1, where the operator maintains full

control without any automatic interaction, to Level 5, which involves executing fully automated construction tasks such as navigating, dumping, or drilling without causing hazardous interactions with the workforce [42]. This insight led to the idea of designing a sensory framework mounted on construction equipment. Such a framework aims not only to elevate the task execution of this equipment to higher levels of autonomy but also to create an automated monitoring system around the equipment. This system would prevent upcoming hazardous situations by facilitating meaningful interactions between the system and the equipment operator. Moreover, this approach provides a future-proof direction toward achieving the highest possible level of autonomy in heavy construction equipment tasks, which can significantly reduce the number of incidents and fatalities within roadway work zones.

2.2 Limitations of Non-Perceptual Sensing Technologies in Automation

Sensing Technologies can be categorized into two main groups: non-perceptual and perceptual sensors. Non-perceptual sensors, such as electromagnetic and proximity-based sensors, primarily use short-range waves like Ultrawide-band, Bluetooth, and Wi-Fi to measure proximity between transmitters and receivers. Perceptual sensors, including RGB, Depth, and LiDAR sensors, employ more sophisticated techniques to produce a multi-dimensional understanding of the environment, generating two-dimensional images and three-dimensional point cloud data. Several studies have attempted to utilize non-perceptual sensing technologies to determine the location of personnel [22]–[24] within work zones. These studies often combine sensor data from various senders into a unique cloud-based environment to identify potential hazards directly or analyze them within simulations [27]. Additionally, many studies have employed channels of senders and receivers using electromagnetic sensors or RFID-based sensors to measure the proximity of workers and equipment within a work zone to predict upcoming hazards [25], [26].

Meanwhile, several studies have attempted to recognize the movements and activities of construction equipment by attaching proximity-based devices to the machinery. in one of the novel studies, *Akhavian and Behzadan* [43] developed a pipeline to assess construction equipment activity based on detailed data and to interpret this data for use in simulation-based methodologies and applications. They utilized data from accelerometers, gyroscopes, and GPS to obtain the linear

and angular velocities of different parts of the construction equipment, as well as their location. A supervised machine learning classification algorithm was employed to identify the tasks performed by the equipment using the sensor data. More than that, A recent study conducted by *Ansaripour et al.* [44] focused on designing a truck pose estimation system within construction work zones by analyzing data gathered from proximity devices utilizing Ultrawide-band waves. These devices were attached to a moving truck and interacted with a set of stationary receivers within the work zone. By integrating the proximity data from these attached devices, the researchers were able to generate an aerial 2D pose of the truck as it moved within the work zone.

However, regardless of the ongoing advancements occurring within the domain of non-perceptual sensing technologies utilization within work zones, there are significant limitations associated with these methods. These limitations hinder researchers from exclusively developing frameworks based on such sensors. Firstly, the efficacy of these methodologies is often compromised by environmental intricacies, particularly when signal pathways encounter obstacles that obstruct transmission between transmitters and receivers. Furthermore, these modalities predominantly assess object proximity, thereby constraining their capacity to offer a perceptually interpretable understanding of environmental dynamics and to effectively identify diverse hazard typologies. Moreover, dynamic construction work zones, such as roadways, are highly susceptible to the entrance and exit of unrecognized objects, which may include workforce members, motor vehicles, and construction equipment not defined by tag-based sensors. Consequently, new hazardous situations can emerge from the interactions between the workforce and these unrecognized objects. Additionally, the increase in obstacles between existing senders and receivers can further disrupt the system's functionality, exacerbating the challenge of accurately identifying and mitigating hazards in the work zone. On the other hand, perceptual sensing technologies implemented in recent studies to design automated safety monitoring systems have provided significant advancements in the context of safety in construction work zones and roadway work zones. These techniques offer a more comprehensive understanding of environmental dynamics and enhance the capability to identify and mitigate diverse hazard typologies effectively.

2.3 Automated Safety Monitoring Systems in Construction

Perceptual sensing technologies are generally developed using structured (manually descriptive) or automated (CNN-based) algorithms applied to multidimensional data captured from sensors capable of generating more than a 1D signal. These sensors provide detailed observations of the environment. Two of the most commonly used sensors in such systems are image cameras (RGB) and depth cameras (short-range depth cameras or long-range depth sensors, known as LiDARs). By analyzing the color or depth images captured from the environment with various algorithms, researchers can identify and localize workforces, recognize their activities, and detect construction equipment and their interactions. This capability is crucial for monitoring hazardous situations and ensuring safety within construction work zones.

In one of the early studies of this concept, *Memarzadeh et al.* [45] developed a structured algorithm to detect construction equipment within video frame footage of a construction work zone using Histograms of Oriented Gradients and Colors (HOG+C). Compared to previous studies, their proposed algorithm was not only more efficient and accurate in detecting objects such as workers, excavators, and dump trucks within the construction site, but it was also capable of detecting these objects while they were idle, a capability that previous studies lacked. Furthermore, with advancements in the utilization of Convolutional Neural Networks (CNNs) for object recognition and classification, a novel study conducted by *Kim et al.* [46] designed a 2D construction equipment detection model by retraining a novel region-based convolutional network developed by *Dai et al.* [47], known as *R-FCN*, using the transfer learning technique. They achieved an accuracy of over 95% in detecting five different types of equipment. This advancement not only demonstrates the value of CNNs in the safety monitoring sector but also significantly contributes to the development of more accurate object detection and localization systems within work zones. As a subsequent study in this field, *Fang et al.* [48] aimed to detect workers and construction equipment using image data by retraining a recently published object detection CNN model called *Faster R-CNN* [49]. They incorporated an additional image size matching step within their proposed pipeline to achieve higher accuracy in detecting objects within the work zone which is noted as the *IFaster R-CNN* model. Their results demonstrated that the improved *Faster R-CNN* increased detection accuracy for five pre-trained equipment classes by 1% to 10%, depending on the detection class, compared to the

standard retrained *Faster R-CNN* as well as the standard *R-CNN* model [50].

As the utilization of image-based pipelines provided by the ongoing studies were developing more advanced workforce and equipment identification, segmentation, and activity recognition [51]–[53] through an image data, emerging technologies of depth cameras helped the researchers to localize the objects such as equipment and the workforce within the work zone by implementing manual data processing algorithms on top of the point cloud data to make the automated systems capable of earning a 3D understanding of the environment to then monitor the safety criteria in real-time. In a pioneering study that was among the first to use long-range depth sensors, *Wang and Cho* [54] designed a pipeline to visualize a 3D schematic of the construction work zone by employing data processing techniques to distinguish construction equipment in 3D unstructured environments. Their primary data processing technique involved using Convex Hull to filter the point cloud data by removing interior points and generating a surface connecting the outer points. This approach allowed them to visualize different construction equipment as structured 3D maps, providing professionals with a clear view of the work zone. This capability enables construction equipment operators to be aware of potential collisions with other objects, thereby helping to prevent incidents by pausing tasks when necessary. In a follow-up study, *Chen et al.* [55] designed an automated pipeline based on point cloud data obtained from a construction work zone to identify, classify, and localize heavy construction equipment. The pipeline begins by downsampling the point cloud data to create a uniform histogram, making it easier to process. Next, they used *Random Sample Consensus (RANSAC)* [56] to remove surface points, followed by clustering the remaining points and the *Euclidean Distance Metric* [57] to group points belonging to the same construction equipment. For categorizing different types of heavy construction equipment, they employed a pre-designed 3D object descriptor model called *Ensemble of Shape Functions (ESF)* [58]. Finally, a machine learning classifier was applied to label and classify each point cloud cluster, enabling the identification and localization of different types of heavy construction equipment using a laser scanner.

In addition to the aforementioned safety monitoring frameworks designed to perceive interactions between the workforce and construction equipment, few studies have addressed the challenge of observing interactions between workers and motor vehicles in construction work zones. These studies identified and localized vehicles using fast and efficient CNN models such as *Super Fast and*

Accurate 3D Object Detection (SFA3D) [59] to process point cloud data captured from a LiDAR sensor present at the job site [60]. This approach allows for real-time monitoring and enhances safety by quickly and accurately identifying potential hazards involving motor vehicles within the work zone. However, despite the accuracies and successes of perceptual sensing technologies, there are still limitations associated with them. A comprehensive safety monitoring framework must continuously monitor the entire construction work zone to detect potential hazards, which is challenging to achieve in indoor construction work zones and nearly impossible in outdoor zones such as roadways. This requirement necessitates numerous perceptual sensors placed throughout the environment. Additionally, a roadway construction work zone is highly dynamic in two major aspects: 1) The construction is ongoing, with machinery and workforces moving forward to the next section multiple times a day. 2) The continuous movement of construction equipment (e.g., mixer trucks, loaders) and motor vehicles passing through the work zone constantly creates new lines of sight. This dynamic nature requires sophisticated placement and frequent relocation of perceptual sensors to maintain the necessary safety monitoring criteria.

Lastly, there have been few studies in construction safety monitoring that combine image-based sensors (RGB) and depth sensors (depth cameras and LiDAR). This combined approach first uses RGB sensors to identify and segment objects within the work zone, and then utilizes depth data to localize these objects relative to the sensor's coordinates. This integrated framework has the potential to enhance the accuracy and reliability of safety monitoring by leveraging the strengths of both sensor types. Such frameworks have been successfully developed and tested in the context of Autonomous Vehicles (AVs) by designing pipelines to identify and coordinate pedestrians and other vehicles to avoid collisions and enable autonomous driving using perceptual sensing technologies. Adapting similar methodologies to construction work zones could be highly beneficial. The proposed approach would involve designing a pipeline to detect and localize workers and motor vehicles. By mounting these sensors on construction equipment, the system could localize workers and vehicles within the roadway work zone, enabling the equipment to stop tasks or alert workers to impending hazards. Additionally, this framework could serve as a preliminary study for implementing autonomous task execution of construction equipment within the work zone, potentially enhancing both safety and efficiency. The following section highlights several research studies that have developed pipelines for heavy construction equipment, serving as foundational support for the proposed concept in this

study.

2.4 Autonomy in Construction Equipment

Despite the extensive research conducted in the realm of Autonomous Vehicles (AV), few studies have addressed the development of Autonomous Construction Equipment (ACE). It is noteworthy that all these studies have employed automated algorithms to process data gathered from perceptual sensors, aiming to satisfy specific criteria such as obstacle avoidance, movement planning, execution planning, and safety observation. These advancements, along with their successful implementations, reinforce the concept of designing an automated safety framework for construction equipment to provide comprehensive safety monitoring within work zones.

In one of the early studies, *Kayhani et al.* [61] developed a path-planning pipeline for construction equipment to address the limitations associated with manual processes which tend to be more time-consuming and costly based on robotic obstacle avoidance techniques. Their methodology involved generating configuration space (C-Space) obstacles and using *Dijkstra*'s algorithm [62] for pathfinding. This system was designed for the planar motion of convex 2D objects, assuming continuous translational and discretized rotational movements. Although their study assumed a well-known plan of the construction work zone, it serves as a preliminary automated task planning pipeline for ACEs. Such autonomous navigation pipelines have been further developed within the field of robotics. Furthermore, *Zhang et al.* [63] designed an autonomous excavator capable of performing multiple tasks while addressing challenges such as collapse accidents, excessive dust, and extreme weather conditions. From a perceptual sensing technologies perspective, they utilized LiDAR sensors and RGB cameras to perform 2D and 3D parsing of the environment in front of the equipment. This allowed them to generate a 3D localization map to avoid obstacles and execute a task stop command when the cameras, using CNNs, detected humans or animals within the work site and make the excavator capable of performing tasks such as loading dump trucks, capturing and removing rocks, and clearing large piles. Moving forward, *You et al.* [64] proposed an autonomous bulldozer designed to perform ground elevation tasks autonomously. They employed RGB cameras and LiDAR sensors to detect and avoid obstacles during task execution. Additionally, they utilized *YOLO-v3* [34], [65] on their stereo camera data to detect and localize obstacles, particularly workers,

to prevent incidents while performing the task. In one of the most pioneering studies, *Khan et al.* [66] designed a predictive control model for bulldozer movement by simulating the dozer and its behavior with a modified Husky robot. This robot was equipped with LiDAR, RGB, and Depth sensors to avoid obstacles while performing navigation tasks.

In summary, it can be concluded that despite the varying success rates of perceptual sensing technologies, the utilization of such sensors to acquire detailed environmental data, combined with different types of CNNs to interpret this data, is highly effective in providing a 3D awareness for autonomous construction tasks [67]. Therefore, this study proposes the design of two parallel pipelines utilizing LiDAR and RGB-D sensors that not only identify hazardous situations around construction equipment by the required safety checks associated with the ITCP rules but also enable the equipment to perform tasks autonomously by implementing advanced navigation algorithms, ensuring safe task execution.

CHAPTER 3

Problem Statement

The primary objective of this study is to develop a pipeline that can be implemented in any heavy construction equipment to provide autonomous safety monitoring through real-time observation of the surrounding environment. This objective represents a significant advancement in the design of automated monitoring systems within roadway work zones. Additionally, it can serve as a perception framework for task planning and execution in autonomous construction equipment. To achieve this goal, the study is structured around five main sub-objectives.

In the initial stage of this study, potential hazards and unsafe situations were identified based on the movements of construction equipment within the work zone. These hazardous situations were then categorized into two groups. The first group comprises situations arising from interactions between workers and construction equipment. To design a fully automated awareness system for monitoring these hazardous situations, each unsafe scenario was assigned to the corresponding construction equipment. The second group includes hazardous situations associated with interactions between other objects, primarily motor vehicles, and the workforce. Although these situations do not directly result from the movement of construction equipment, it is feasible for the existing construction equipment to observe and monitor them through a parallel framework alongside the one designed for monitoring the first category of hazards. By identifying and assigning hazards within work zones to each piece of construction equipment, we can develop a comprehensive framework to observe and mitigate such situations. However, it is essential to establish a method for translating these situations into a format understandable by computers before designing frameworks for monitoring hazards. This necessity arises from the inaccuracies that automated models, including

advanced Natural Language Processing models, have in interpreting descriptive data. To address this, traditional ITCP rules [68] were employed to convert each hazardous situation into a corresponding unsafe zone. Consequently, the primary objective of these frameworks is to accurately detect the presence and location of objects, primarily workers and motor vehicles, within the designated unsafe zones associated with them.

Subsequently, we designed localization frameworks capable of monitoring unsafe zones associated with construction equipment using RGB-D sensors and LiDAR sensors. While several previous studies have focused on detecting, localizing, and recognizing the activities of workers or construction equipment using a single RGB-D or LiDAR sensor, our approach utilizes multiple RGB-D sensors and LiDARs attached to construction equipment. This setup enables real-time monitoring of designated unsafe zones by simultaneously analyzing the captured data through different models. To achieve this, we developed frameworks based on automated data processing algorithms that require minimal computational power. This approach allows us to run and test the pipeline through a real-time case study, thereby demonstrating the efficiency of implementing autonomous techniques in construction equipment. We employed one of the fastest object detection models, *YOLO* [34], [65], in conjunction with point cloud processing libraries [69] based on the C++ programming language, and *SFA3D* [59], one of the fastest vehicle localization models. By utilizing LiDAR point cloud data, we were able to localize both the workforce and motor vehicles within the sensors' boundaries.

Furthermore, the placement of sensors and the fusion of sensory data were carefully considered to design an efficient arrangement and data combination of the aforementioned sensors. This was achieved using the ROS-based simulation application called *Gazebo* [70] which helped to ensure the elimination of any blind spots around the equipment from a sensory perspective. We simulated the construction equipment and the virtual representation of the sensors in the Gazebo application, then manually observed the simulated sensory data to design a placement configuration that avoids blind spots. The resulting sensor placement can be replicated on actual construction equipment. By running the pipeline with the designed sensor placement, we were able to effectively localize workers and motor vehicles moving around the equipment. Lastly, the proposed pipeline was validated through a series of testing scenarios within an isolated real-world environment having the algorithms tested on top of the designed sensor placement. These tests aimed to measure the framework's accuracy in localizing workers and motor vehicles around the equipment, identifying trajectory patterns around

the equipment, and assessing the pipeline's robustness in detecting entries and exits from designated unsafe zones.

CHAPTER 4

Methodology

The following sections will provide an overview of the proposed solutions for designing a system that ensures automated monitoring around heavy construction equipment.

4.1 Hazard Identification and Conversion to Unsafe Zone

Depending on the size and location of the roadway construction site, various types of workforces and construction equipment are engaged in multiple tasks. The forward movement inherent in constructing a roadway creates a dynamic environment while the project continues. Therefore, it is reasonable to assume that construction equipment performs regular and repetitive movements while completing tasks, as the ITCP regulations dictate, which delineate specific unsafe zones around the equipment. Similarly, motor vehicles passing through the work zone exhibit predictable movement patterns. Consequently, each vehicle, whether engaged in tasks or merely passing through the roadway work zone, follows a consistent directional movement pattern (Figure 4.1 B). Additionally, despite the workforce's regular movement patterns within the work zone, the repetitive nature of these movements cannot be conclusively determined due to the inherent uncertainty in human decision-making. Each worker may choose a particular movement direction based on the prevailing situation. Nevertheless, each workforce has a distinct task and purpose within the work zone, causing them to stay in specific locations while performing their duties. This results in minor and random movements within their designated areas (Figure 4.1 A). By assuming an offset threshold for each working crew's location and movements, multiple potential collisions can be identified

between the working crew's positions and the directional patterns of motor vehicles within the work zone. These collisions result in hazardous situations (Figure 4.1 C) that must be monitored by at least an automated safety monitoring system designed for construction equipment. Although single construction equipment cannot observe all hazardous situations due to obstacles creating several blind spots, each hazardous situation can be detected by at least one construction equipment. Therefore, each identified hazardous situation is converted into unsafe zones based on the directional movement pattern of the corresponding construction equipment (Figure 4.1 D) and the monitoring framework designed for the corresponding equipment calculates the relative location of each assigned object to the construction equipment within these unsafe zones. The implementation of such an automated monitoring framework is facilitated through a specific arrangement of perceptual sensors (Section 4.3) and an automated algorithm for detection, localization, and safety checking, which processes data received from these sensors (Section 4.2).

4.2 Perception Algorithms Pipeline

To ensure proper monitoring of unsafe zones through an automated perceptual framework, it is crucial to identify the types of objects that sensors need to target to assess the hazard level of a given situation. Generally, interactions occur between the workforce and construction equipment or passing motor vehicles. Therefore, when implementing a perception framework on construction equipment, it is imperative to detect and localize workers (or humans) in and around the equipment and even beyond the unsafe zones related to their interactions with the machinery. Additionally, the unpredictable nature of passing vehicles within work zones necessitates that construction equipment detect and track motor vehicles as they enter and leave the work zone. This development is vital for alerting workers to potential incidents involving motor vehicles.

To address these needs and following recent studies on automated safety monitoring, we developed two parallel pipelines for the workforce and motor vehicles which utilize perception technologies to localize the targeted objects within the unsafe zones and beyond. Initially, we tested the accuracy of each pipeline using a single sensor. Subsequently, we integrated them into the targeted construction equipment as a unified framework. This framework combines and analyzes data gathered from the corresponding sensor arrangement to determine whether the ongoing situation is hazardous, as

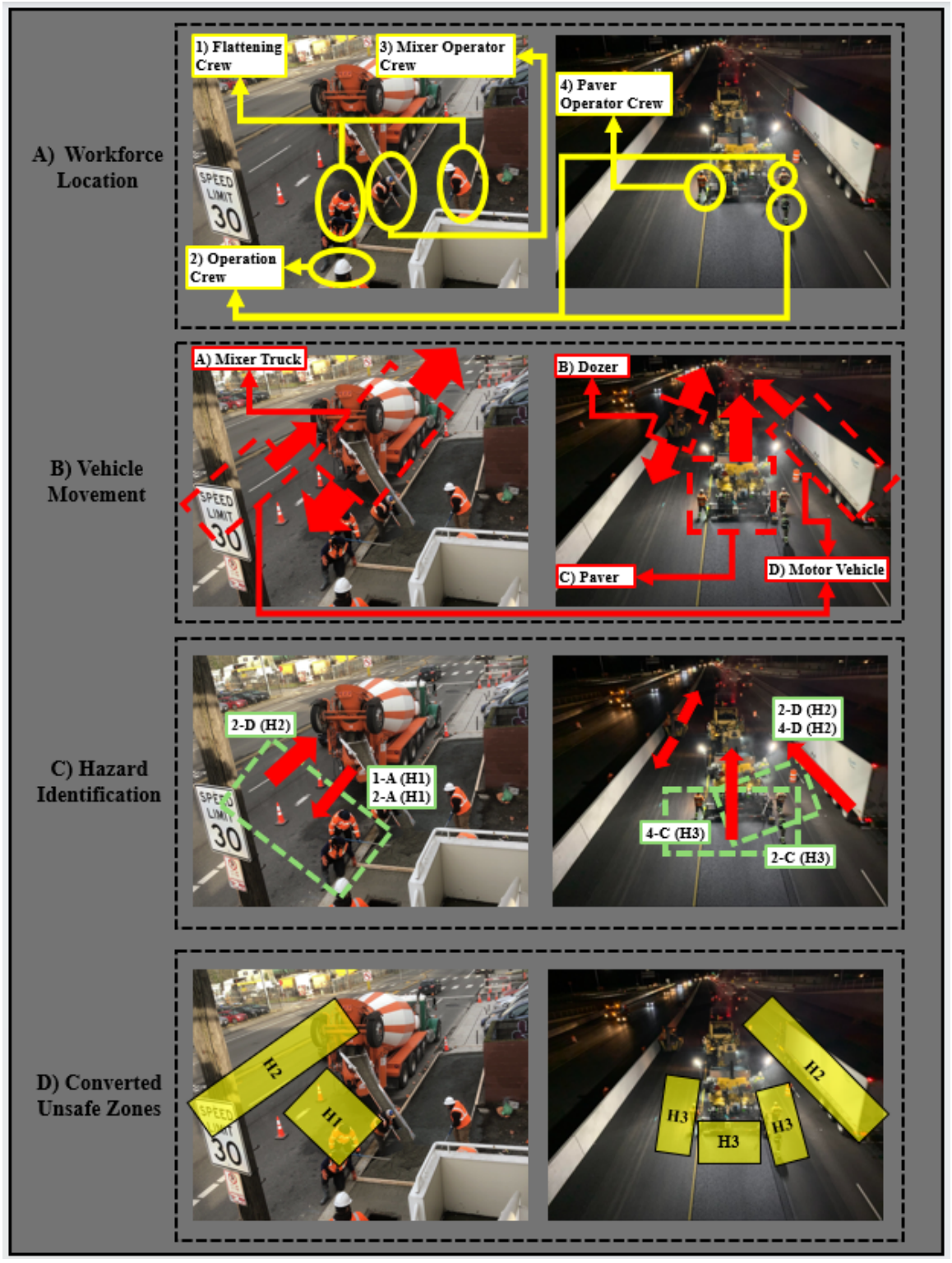

Figure 4.1: *A) Workforce identification, categorization, and localization. B) Vehicle identification and tracking. C) Hazard identification. D) Hazard conversion to unsafe zone.*

detailed in Section 4.3. The first pipeline employs a CNN-based object detection model, followed by point cloud data processing algorithms to ascertain the location of a person present in front of and within the image, as well as within the depth range of an RGB-D sensor. The second pipeline involves using a CNN-based vehicle localization model, which processes data captured from a LiDAR sensor.

4.2.1 Human Detection and Localization

In the initial stage of pipeline design for object detection and localization, particularly focusing on humans, we developed an automated system to detect and localize objects relative to the coordinates of a single RGB-D sensor. This pipeline is structured into four steps. Given the substantial GPU computing power required for the implementation of CNN models, our objective was to design a pipeline capable of localizing workers in real-time using not only a single RGB-D sensor but also multiple sensors simultaneously. This was achieved using a single laptop as our hardware setup, as detailed in the sensor placement and scenario testing sections.

As illustrated in Figure 4.2 A, the pipeline begins by capturing image data from the RGB-D sensor and inputting it into the *YOLO* object detection model. This model (which is a pre-trained model on the *COCO* [71] dataset) can identify and segment various objects, such as humans, within the image data using bounding boxes. Once the coordinates of the detected object's bounding box are identified within the image (Figure 4.2 B.1), the pipeline processes the point cloud data captured by the RGB-D sensor. It then modifies the point cloud data to remove the majority of points that are not associated with the detected object. In the first step of point cloud modification, we implemented a manual point cloud data removal algorithm. This algorithm operates based on the RGB-D sensor's elevation from the ground, with an offset of 20 centimeters, to eliminate points related to the ground surface (Figure 4.2 B.2). The ground's material and roughness can significantly affect its captured point cloud data which might hinder automated surface removal algorithms, such as Plane Segmentation, from accurately determining the features of the surfaces present in the point cloud data. Such algorithms operate by determining the surface normals of different clusters of points within the point cloud data to extract the flat-shaped clusters out of them since the roughness of the utilized material in the surfaces can result in huge noises within the capturing point cloud data (mostly ground surfaces). In addition to the manual ground removal technique, *Plane Segmentation* [69] was employed to segment and remove remaining flat surfaces (roofs or walls), based on the

normals value determination which has been described earlier. This approach prevents the remaining points associated with horizontal surfaces from being counted. By combining manual ground removal with Plane Segmentation, we successfully eliminated horizontal surfaces in both the foreground and background of the objects within the point cloud data.

In the subsequent step, using the dimensions and coordinates of the detected object bounding boxes within the image, we extracted the point cloud data of the detected object. This was achieved by trimming a frustum out of the modified point cloud data (Figure 4.2 B.3). The frustum's near plane corresponds to the image plane, while the far plane is set to the maximum distance from the RGB-D sensor at which the depth data is reliable, which we determined to be 6 meters. This distance was standardized across different sensor models, including the Astra camera, RealSense camera, and Microsoft Azure Kinect camera. This process was accomplished using a function from the Point Cloud Library (PCL) [69] called *Frustum Culling*, which trims a frustum of point cloud data from an input point cloud based on the provided horizontal and vertical Field of View (FOV). These FOVs were calculated based on the relative dimensions of the reported bounding box to the image's size (Figure 4.2 B.1), which were then converted to a proportion of the camera's depth FOV, both horizontally and vertically (angular). By inputting the calculated FOVs into the Frustum Culling function and setting the near and far planes as previously discussed, we extracted the trimmed point cloud data of the detected object identified by the *YOLO* model. In the last step, another point cloud data processing technique was applied to the trimmed data to remove outlier points, which are typically generated due to the sensor's inaccuracy, manifesting as sparse points around the solid object points (Figure 4.2 B.3). This was achieved using the Statistical Outlier Removal function from the PCL library, which employs a maximum distance threshold to determine whether a point is an outlier or an inlier of a 3D object within the trimmed point cloud. This function computes the distance of each targeted point to a fixed number of adjacent points around it.

By implementing these four major steps to the RGB-D sensor data and averaging the X, Y, and Z values of the remaining points using a *3D centroid* [69] calculation, we were able to determine the relative distance of the detected object from the RGB-D sensor.

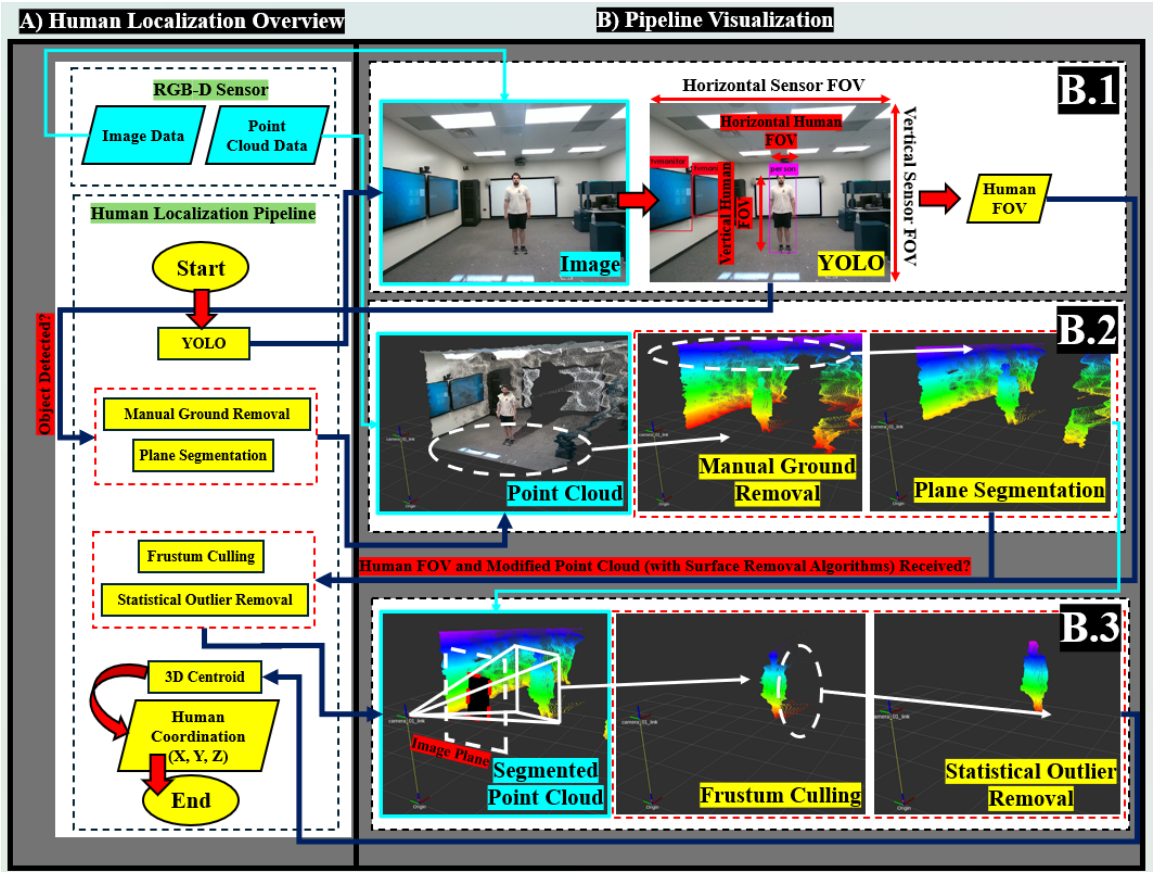

Figure 4.2: *A) Human Localization pipeline overview. B) Visualizing the results of the YOLO object detection model (B.1), surface removal (B.2), frustum culling, and statistical outlier removal (B.3).*

4.2.2 Motor Vehicle Localization

For the purpose of motor vehicle localization within a roadway construction work zone, it is essential to design a pipeline capable of detecting and tracking vehicles from the moment they approach from a distance until they exit the work zone area. This requires using sensors and algorithms that can receive data from greater distances than the typical RGB-D sensor's range. Additionally, the quality of the received data must be sufficient for the localization algorithms to effectively analyze it. To achieve this, we utilized the *SFA3D* model (which was trained on *KITTI* [72] dataset) to track the motor vehicles within the work zone. The *KITTI* dataset includes images, short-range point cloud data (Depth Sensor), and long-range point cloud data (LiDAR sensor) collected from a combination of sensors mounted on a moving vehicle operating in urban areas. Although many algorithms have been designed and trained using the *KITTI* dataset to detect and localize various objects through a combination of image and point cloud data, *SFA3D* stands out as one of the few 3D vehicle detection

algorithms (CNN models) [73] that has been specifically designed and trained to operate solely using LiDAR sensor data, which is long-range point cloud data.

To do so, we fed the point cloud data captured from a LiDAR sensor into the SFA3D model (Figure 4.3 B.1) to extract the coordinates of the detected vehicles (Figure 4.3 B.2) relative to the LiDAR sensor as they arrived and departed from the sensor's location.

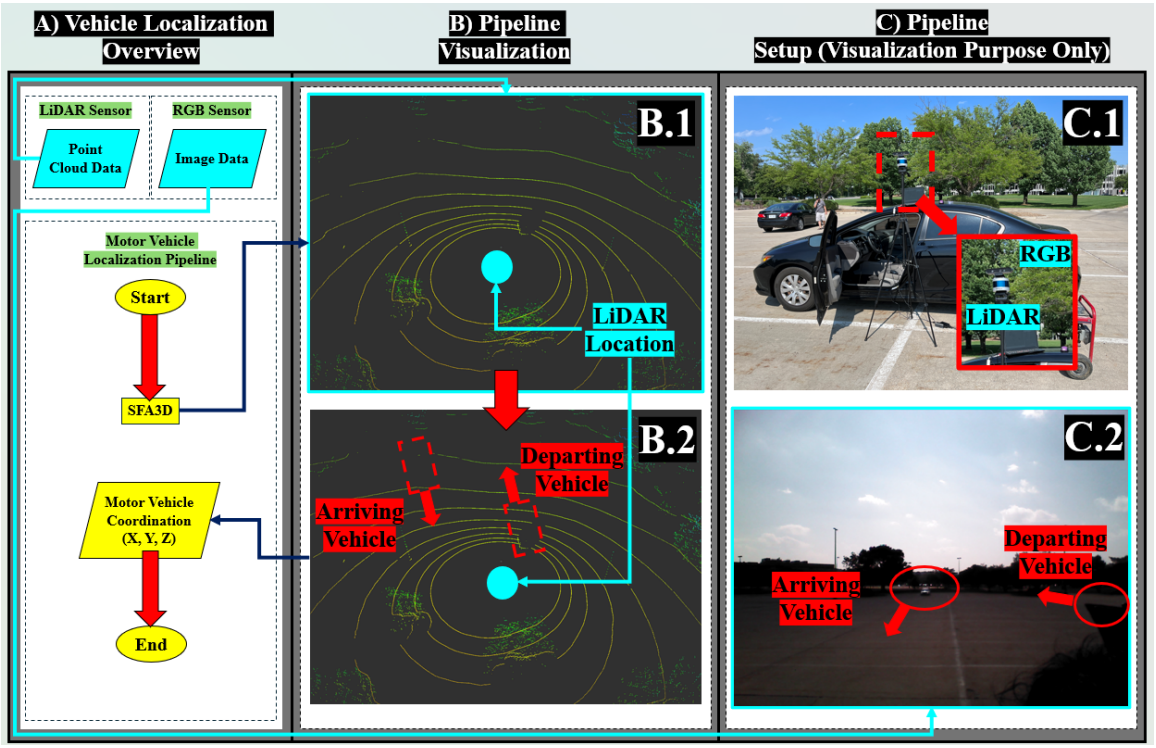

Figure 4.3: *A) Vehicle Localization Pipeline Overview. B) Visualization of SFA3D vehicle detection CNN. C) Overview of the sensor's setup (C.1) and motor vehicle's movement (C.2).*

4.3 Sensor Placement and Combination Through Simulation

As discussed in Section 4.1, various hazardous situations can be identified within a roadway work zone, influenced by the size of the work zone and the movements of the workforce and motor vehicles (both equipment and passing vehicles). These hazardous situations are converted into specific unsafe zones to enable the localization pipelines (detailed in Section 4.2) to detect, localize, and track target objects within or beyond these zones, facilitating further analysis of the associated hazard levels within the work zone.

The effectiveness of the localization pipelines described in Section 4.2 relies on the capabilities

of RGB-D and LiDAR sensors to capture real-time image and depth data of objects (for workforce localization) and real-time depth data of other objects (for motor vehicle localization). Therefore, it is crucial to design a sensor placement for each construction equipment, based on the assigned unsafe zones, as well as the equipment's shape and size. Although the assignment of unsafe zones is further elaborated in the Case Study chapter, the unsafe zones identified based on potential interactions between construction equipment and the workforce are assigned to the equipment causing the exposure. For unsafe zones caused by passing motor vehicles, the zones are assigned to the equipment with the highest possible elevation to minimize obstruction by other obstacles. Furthermore, the shape and size of the target equipment create various blind spots for the operator and wider areas of workforce occupation, resulting in larger unsafe zones. To cover these unsafe zones with the proposed pipelines, sensor placements must be capable of recording real-time image and depth data, dependent on the field of view (FOV) of the RGB-D sensor.

To achieve this goal, it is practical to generate a simulated environment using a ROS-based application (*Gazebo* [70]) to closely replicate the targeted work zone (Figure 4.4 A). This simulation illustrates the locations of existing workforces and construction equipment to demonstrate hazardous situations that need to be monitored. The shape and dimensions of the selected equipment should also be implemented in the simulation, ensuring coverage of potential blind spots within or beyond the designated unsafe zones. Furthermore, through a process of trial and replication by adding sensors (Figure 4.4 B), the captured image and depth data can be visualized to determine if the objects requiring detection within the sensor's corresponding unsafe zone are visible within the framework's field of view (FOV). By adding additional sensors or rotating the existing ones to cover blind spots (using RGB-D for close localization and LiDAR for distant localization) and visualizing the sensor data accordingly, an effective sensor arrangement to monitor the unsafe zones through the proposed pipelines can be achieved.

Subsequently, the localization pipelines can be employed to determine potential exposure by comparing the relative locations of the objects to the equipment and its movements. To do this, it is essential to calculate the relative location of each localized object (Figure 4.5 B) to the equipment rather than to the corresponding sensor. Therefore, once sensor placement is completed, the coordinates and orientation of each sensor should be calculated relative to a fixed axis (global axis) on the equipment. Finally, it is possible to transform the location of each localized object to the global

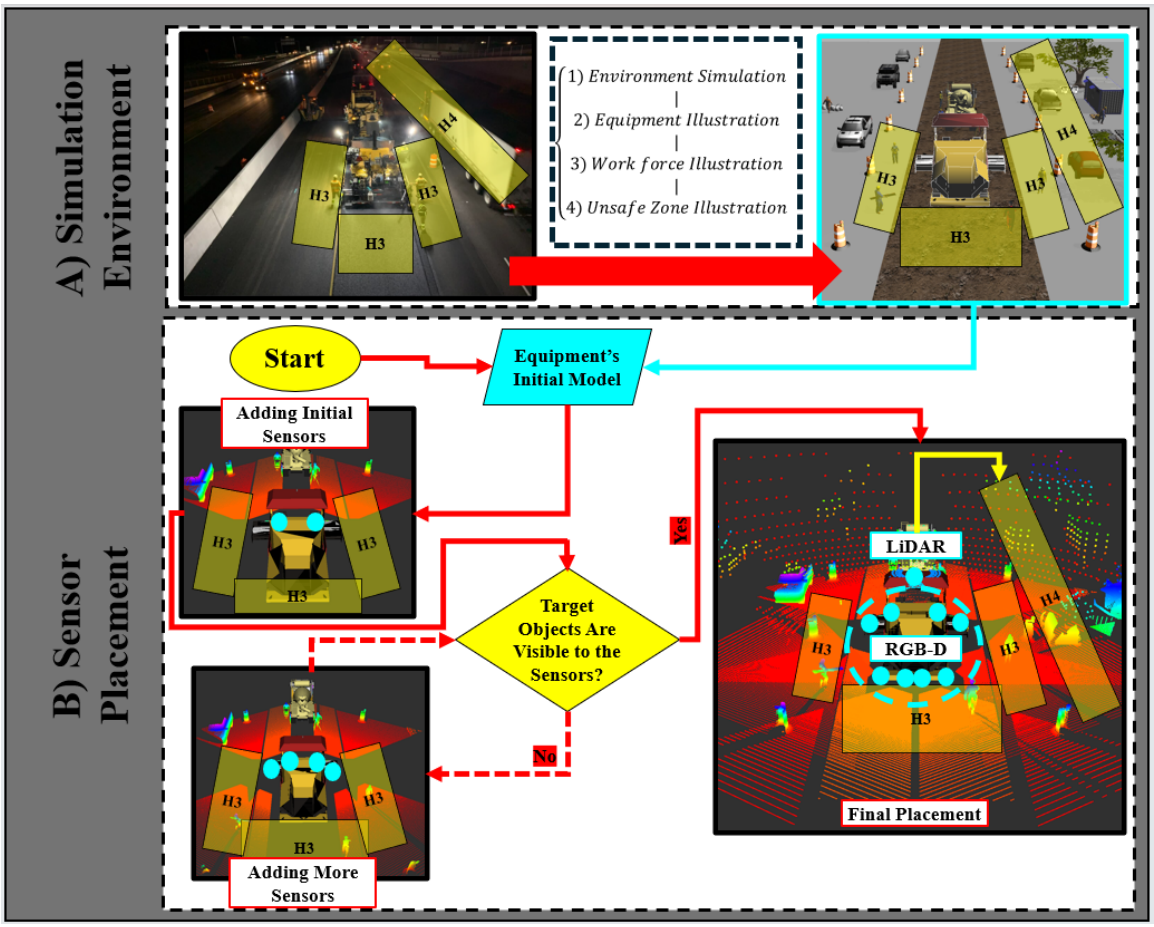

Figure 4.4: *A) Simulation environment in Gazebo. B) Sensor Placement pipeline.*

axis using transformation matrices (Figure 4.6), thereby determining exposures that cause unsafe situations and validating the framework's performance.

Once the final sensor placement has been achieved through the simulation and the global axis has been set on the equipment, the framework can be validated in terms of localization accuracy, tracking the localized objects (within and beyond the unsafe zones), and sensitivity in detecting entries or exits from the assigned unsafe zones. This validation is described in more detail in the case study and results chapters.

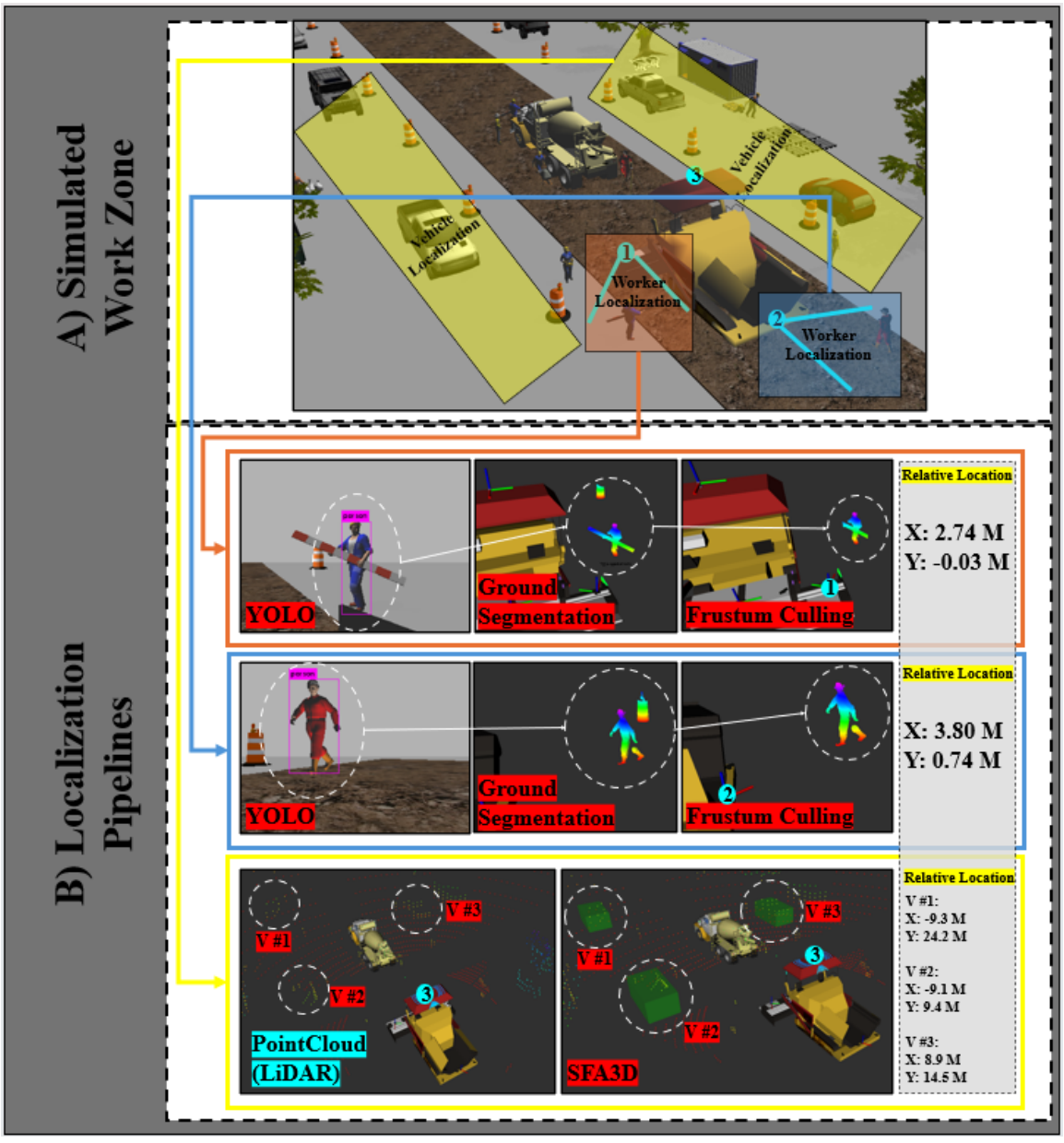

Figure 4.5: *A) Simulation environment in Gazebo. B) Localization Pipelines in the simulated environment to ensure the effectiveness of the proposed placement.*

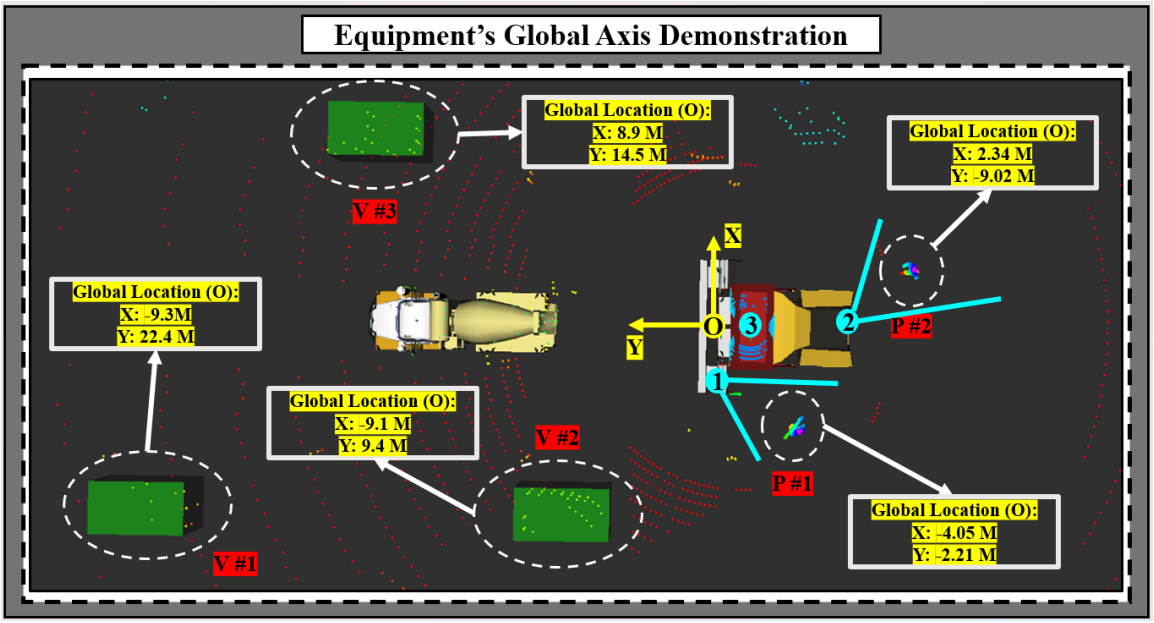

Figure 4.6: *Global location determination utilizing transformation matrices.*

CHAPTER 5

Case Study

To validate the feasibility and accuracy of the proposed autonomous monitoring system for heavy construction equipment, a real-world case study was conducted. This involved examining a simulated sensor placement, implementing the localization pipeline, and conducting a series of real-world validation scenarios in an isolated environment. A predetermined sensor arrangement (developed through simulation) and the actual localization pipelines were utilized to assess the system's accuracy in localizing objects and measuring its sensitivity in analyzing results. Specifically, the scenarios focused on determining the entrance, exit, and tracking of target objects (humans and motor vehicles) within and beyond the unsafe zones. The outcomes of these scenarios are detailed in the Results chapter.

5.1 Work Zone Analysis and Construction Equipment Selection

According to the case study presented in this research, which involves a set of aerial video frames captured from a roadway construction project managed by the U.S. Nebraska Department of Transportation (NDOT), we categorized the tasks and locations of the workers into four distinct groups: operation crew, mixer operator crew, paver operator crew, and flattening and curing crew. These groups are identified numerically from 1 to 4 in Figure 5.1 A. Additionally, we classified the equipment and vehicles present in the work zone into four categories: mixer truck, paver machine, curing machine, and motor vehicles, denoted by the letters A through D. Their respective directions of movement, aligned with their tasks, are indicated by red directional arrows in Figure 5.1 B. By

analyzing the interactions between each vehicle and the working crews, identified by their crew numbers and vehicle categories, we identified five primary hazardous situations (Figure 5.1 C). These situations should be monitored by implementing our automated framework on the construction equipment.

Subsequently, after assigning each hazard to the corresponding equipment or the equipment capable of monitoring it, we determined that four of these situations could be monitored by the mixer truck (indicated by yellow-colored hazards in Figure 5.2), and two by the paver machine (indicated by cyan-colored hazards in Figure 5.2). By examining the potential unsafe zones of each piece of equipment based on its movement direction (highlighted in yellow in Figure 5.1 C) and utilizing the ITCP and TTCP guidelines, we converted the detected hazardous situations into their corresponding unsafe zones for monitoring. Without setting a fixed dimension for each unsafe zone, we decided to extend the boundaries of the monitoring zones (unsafe zones) to the limit imposed by our sensor models, which exceeds the minimum distances specified by the regulations. Given the differences in the design of perceptual sensor arrangements on each construction equipment, we concluded that we should provide a design and a case study focused on the equipment responsible for the most hazards within the work zone. Although three types of construction equipment were identified in our case study (Figure 5.1 B), we chose the mixer truck as our test equipment since it is responsible for monitoring four hazardous situations. These hazardous situations are depicted as H1, H2, H3, and H5 in Figure 5.2 B, representing unsafe zones that require continuous monitoring through the designed sensor arrangement and the implementation of the proposed pipeline. Notably, the unsafe zone associated with H2 can manifest on either side of the target equipment, contingent upon the direction of concrete casting Therefore, this scenario creates a comprehensive 360-degree hazardous zone around the equipment, necessitating thorough surveillance.

5.2 Sensor Selection in Real-World, Design, and Placement through Simulation

To develop a practical sensor arrangement for the target equipment (mixer truck), it is essential to select sensor models that minimize localization errors while meeting monitoring range requirements. For RGB-D sensors, several factors must be considered to determine their suitability. One of the

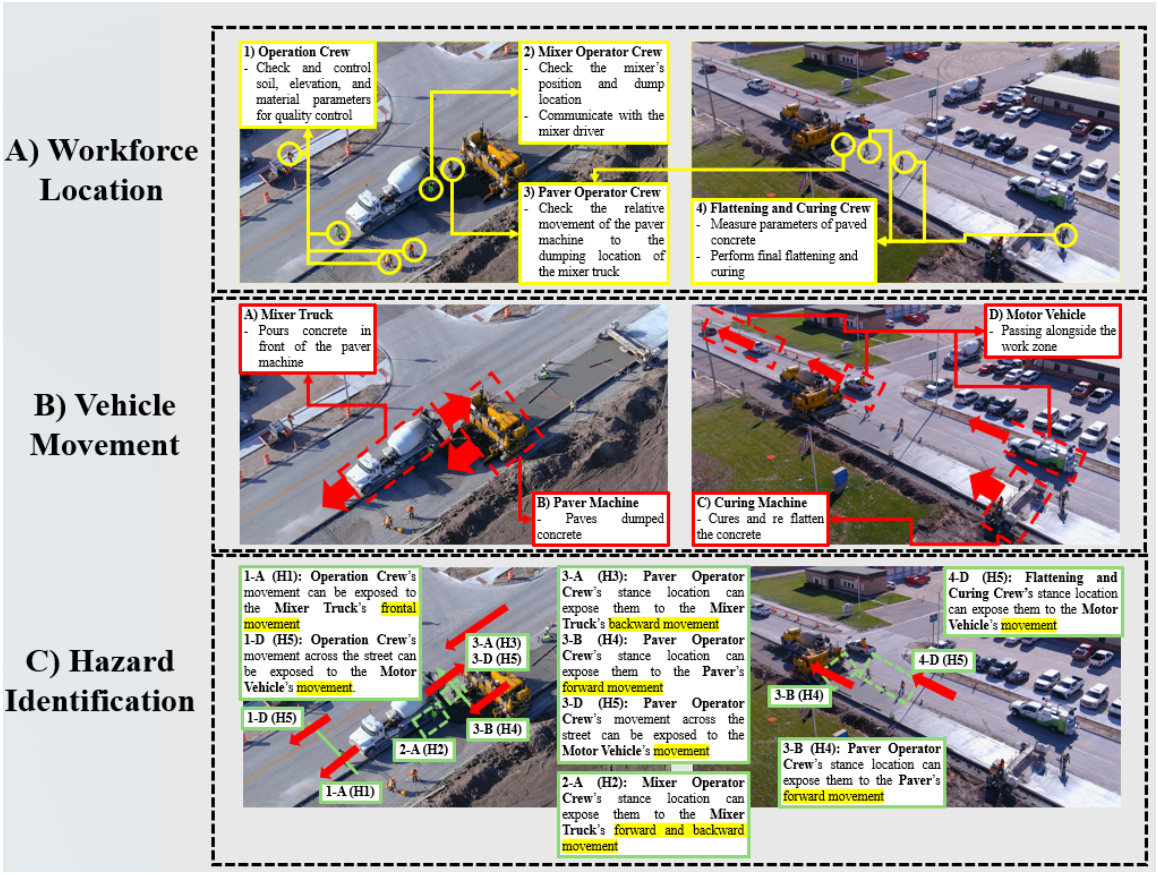

Figure 5.1: *A) Workforce type, task, and movement visualization (1 through 4). B) Construction equipment task and movement categorization (A through D). C) Hazard identification by assigning the interaction between the workforce and the equipment's movement.*

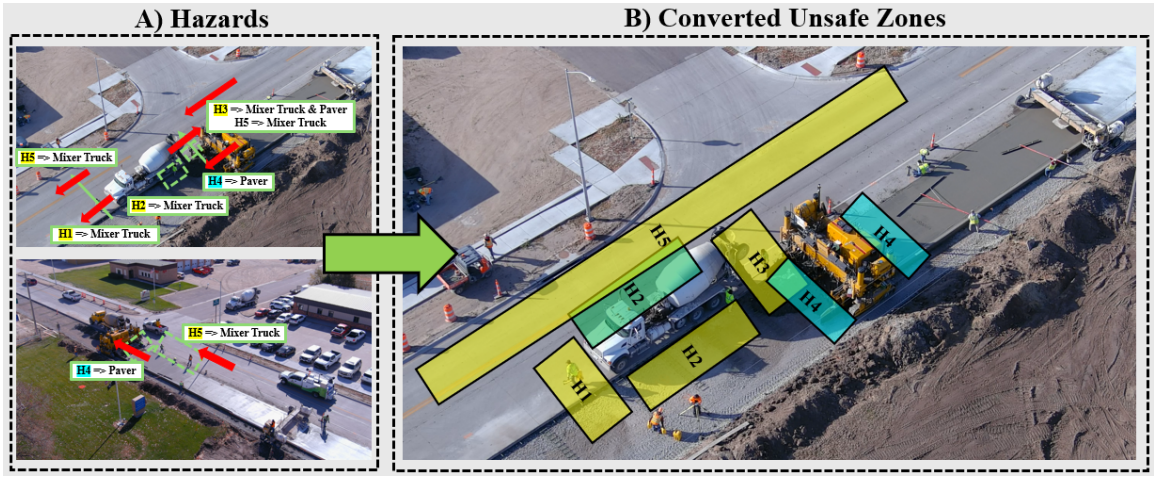

Figure 5.2: *A) Hazard assignment (Yellow-colored situations are assigned to the Mixer Truck and cyan-colored situations are assigned to the Paver Machine). B) Conversion of hazardous situations to unsafe zones.*

most crucial factors is the field of view (FOV), both vertical and horizontal, associated with their image and depth data. This can significantly influence the number of sensors needed in the proposed arrangement. Additionally, the minimum and maximum capturing distances for depth data are critical in determining the capability of the proposed pipelines to localize objects at varying distances, as these parameters differ among RGB-D sensors. Furthermore, the quality of depth data must be monitored to establish the reliable distance within which the sensor's depth data can be trusted, typically closer than the sensor's maximum depth capability. Finally, the sensor's data quality (both image and depth) should be evaluated under different conditions to ascertain its suitability for outdoor environments. Therefore, regarding the selection of RGB-D sensors, we examined three sensor models with varying capabilities: the Astra Camera, Microsoft Azure Kinect, and RealSense sensors.

Regarding the field of view (FOV), the Astra Camera had a horizontal FOV of 60 degrees and a vertical FOV of 75 degrees, while both the Azure Kinect and RealSense sensors had a horizontal FOV of 90 degrees and similar vertical FOVs. Given the importance of capturing real-time images of the entire unsafe zones for the designed localization pipeline, a wider horizontal FOV is beneficial as it reduces the number of sensors required. Therefore, we decided to eliminate the Astra Camera and consider only the Azure Kinect and RealSense sensors. In terms of the maximum and minimum depth capturing distances, the RealSense sensor offered the best performance, with a minimum range of 40 cm and a maximum range of up to 8 meters. This was comparable to the other sensors in terms of minimum range but exceeded the maximum range of the Azure Kinect, which was 5.5 meters. Consequently, we chose the RealSense sensor and designed a simulated version to develop our RGB-D sensor arrangement within the simulation. It is important to note additional reasons for the elimination of the Astra Camera and Azure Kinect sensors. The Astra Camera was incapable of capturing depth data under sunlight conditions, a significant drawback for an outdoor monitoring system. Additionally, the computational resources required to operate more than one Azure Kinect sensor exceeded the hardware configuration we used to test up to six RealSense sensors. Therefore, after validating several commercial and widely used RGB-D sensors, we decided to utilize the RealSense sensor. However, we limited the maximum depth-capturing distance of the RealSense sensor to 6 meters. This decision was due to the sparsity and waviness in the depth data at distances greater than 6 meters, which significantly reduced data quality and introduced noise and inaccuracies in the close object localization pipeline. Table 5.1 illustrates the close localization pipeline's accuracy

using the RealSense sensor.

Table 5.1: *Human detection and localization pipeline test results utilizing RealSense RGB-D sensor*

Actual X (M)	Actual Y (M)	Measured X (M)	Measured Y (M)	Error (M)
0	6	0.33	5.62	0.38
0	5	0.32	5.11	0.34
0	4	0.27	3.97	0.27
0	3	0.23	2.93	0.24
0	2	0.16	1.86	0.21
0	1	-0.09	0.94	0.11

In addition to selecting an appropriate RGB-D sensor, it is crucial to choose a suitable LiDAR sensor for the system design. Although the number of commercially available LiDAR sensors is limited, we tested the *SFA3D* algorithm in a simulated environment with simulated versions of two different LiDAR sensors provided by Velodyne company: the VLP-16 and the HDL-32E. By feeding the simulated sensor data from the simulated environment into the algorithm, we observed that the *SFA3D* was only capable of localizing vehicles using the data received from the VLP-16. This could be attributed to several factors, such as the lack of accuracy in simulating the sensors compared to the real ones, the difference in data quality between simulation and real-world scenarios, or the specific sensor type used to capture the *KITTI* dataset, which was used to train the *SFA3D* algorithm. As a result, the HDL-32E was found to be incompatible with the *SFA3D* algorithm. Based on these observations, we decided to utilize the VLP-16 as the LiDAR sensor for the system. This choice ensures compatibility with the *SFA3D* algorithm and provides practical functionality for the intended application.

In conclusion, we selected the RealSense model as the RGB-D sensor for the close localization pipeline and the VLP-16 model as the LiDAR sensor for the vehicle localization pipeline. Figures 4.2 and 4.3 illustrate the proposed pipelines utilizing these sensors.

Moving on to the design of the sensor placement, it is necessary to simulate the selected sensors within the Gazebo simulation application, as well as the work zone environment. As visualized in Figure 5.3, we simulated the environment using several SketchUp models of construction equipment and workforces, accurately depicting the unsafe zones that need to be monitored. The mixer truck was centered as the target equipment, and a specific sensor placement was designed to effectively observe the assigned unsafe zones. While the selected LiDAR sensor has a pre-designed simulated version

published by Velodyne, we used Gazebo plugins to create a simulated version of the RealSense sensor. This enabled us to begin the process of adding and re-coordinating the sensors to ensure 360-degree image and depth coverage of the close-range area around the target equipment. This coverage is essential for monitoring the aforementioned unsafe zones around the equipment. Additionally, we ensured that the depth coverage extended to long ranges at the highest possible elevation to make the vehicle localization pipeline work effectively.

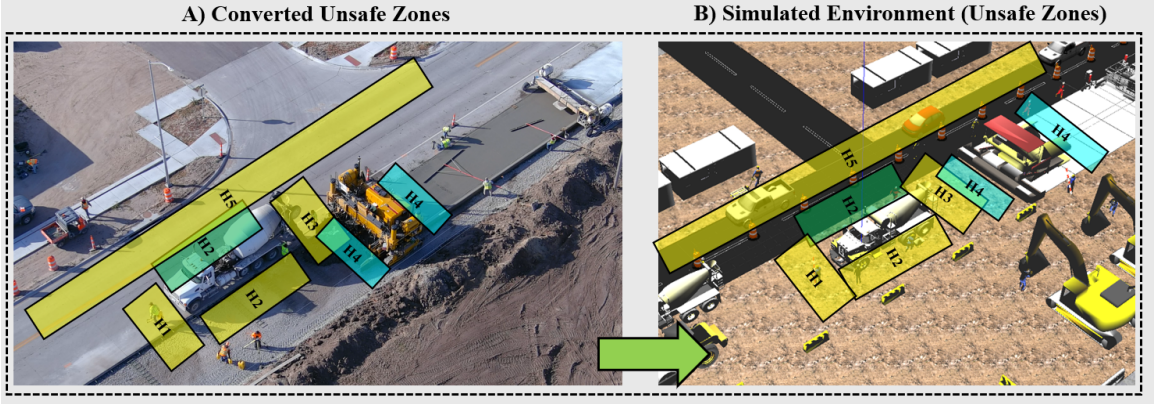

Figure 5.3: *Simulated environment visualization.*

To determine the optimal RGB-D sensor placement on the target equipment, we began by mounting four RealSense sensors on each side of the equipment to provide 360-degree real-time image and depth coverage of the unsafe zones (Figure 5.4 A). Due to the RealSense sensor's minimum depth capturing distance of 40 cm, we added three additional RealSense sensors on the equipment's hood and two on the back at higher elevations. Additionally, we placed two sensors on each side of the equipment, angled at 45 degrees to ensure the framework could localize objects closer than 40 cm to the equipment. This was necessary because the initially mounted sensors at lower elevations caused the *YOLO* model to fail in detecting humans within the image data when they were too close, and the sensors could not capture any point cloud data to measure the relative and global location of these humans to the equipment (Figure 5.4 B). Furthermore, we added an extra RealSense sensor on each corner of the equipment, angled at 45 degrees, to cover the remaining blind spots around the corners (Figure 5.4 C). By adjusting the positions of these sensors within their initial locations, we achieved a configuration capable of receiving image and depth data all around the target equipment and within and beyond the workforce-related unsafe zones. For the required LiDAR sensor placement for the vehicle localization pipeline, we decided to use two VLP-16 sensors. One was placed on

the cabin's roof and the other on top of the mixer's back to ensure full-depth coverage of passing vehicles, as illustrated in Figure 5.4 D.

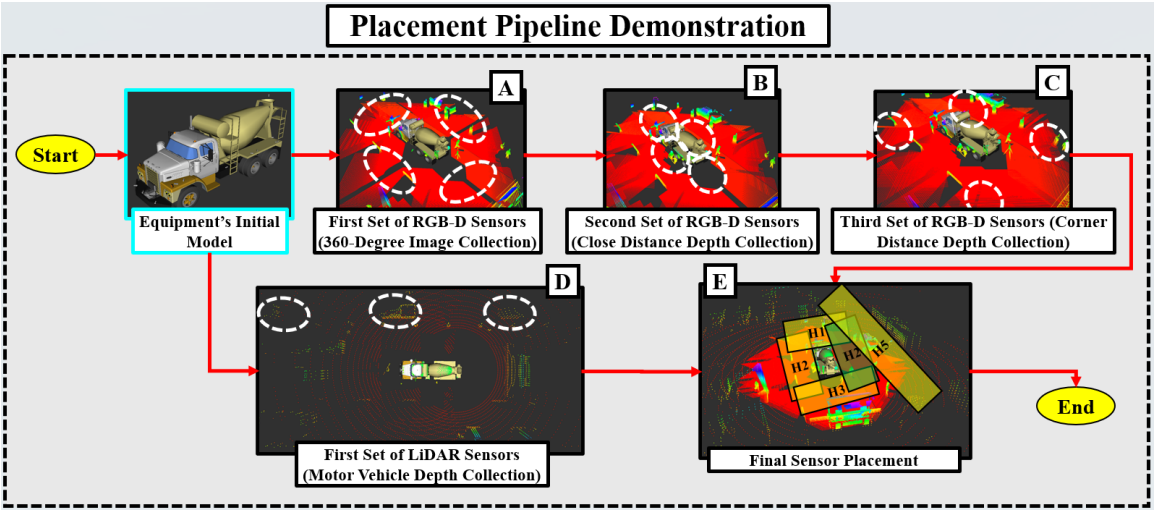

Figure 5.4: *Sensor placement process.*

By combining the placements of the RealSense sensors and the VLP-16 sensors and setting the equipment's origin in the middle of the front bumper and within the same directions as proposed in the methodology chapter, we designed a sensor arrangement to validate the localization pipelines for the mixer truck (Figure 5.4 E). This integrated setup ensures effective monitoring of the unsafe zones and accurate localization of objects and vehicles, thereby enhancing the overall safety and functionality of the system.

5.3 Validation Scenario Design and Description

To evaluate the performance of the proposed system, including the activation of localization pipelines on the designed sensor placement, we conducted three sets of experiments. These experiments assessed the system's accuracy, reliability, and sensitivity by partially implementing the proposed sensor placement on actual equipment within a controlled environment. This environment featured the equipment positioned centrally, with a surrounding area extending up to six meters and a safe passageway extending up to 30 meters in the direction of the mixer truck's stance, provided by the *Lyman-Richey* Corporation (Figure 5.5). The partial placement of the sensors involves dividing the equipment's unsafe zones to validate the system's performance in each division (sub-zone), with

sensors responsible for each specific sub-zone being placed and activated, and the corresponding experiments being performed within the three major sets.

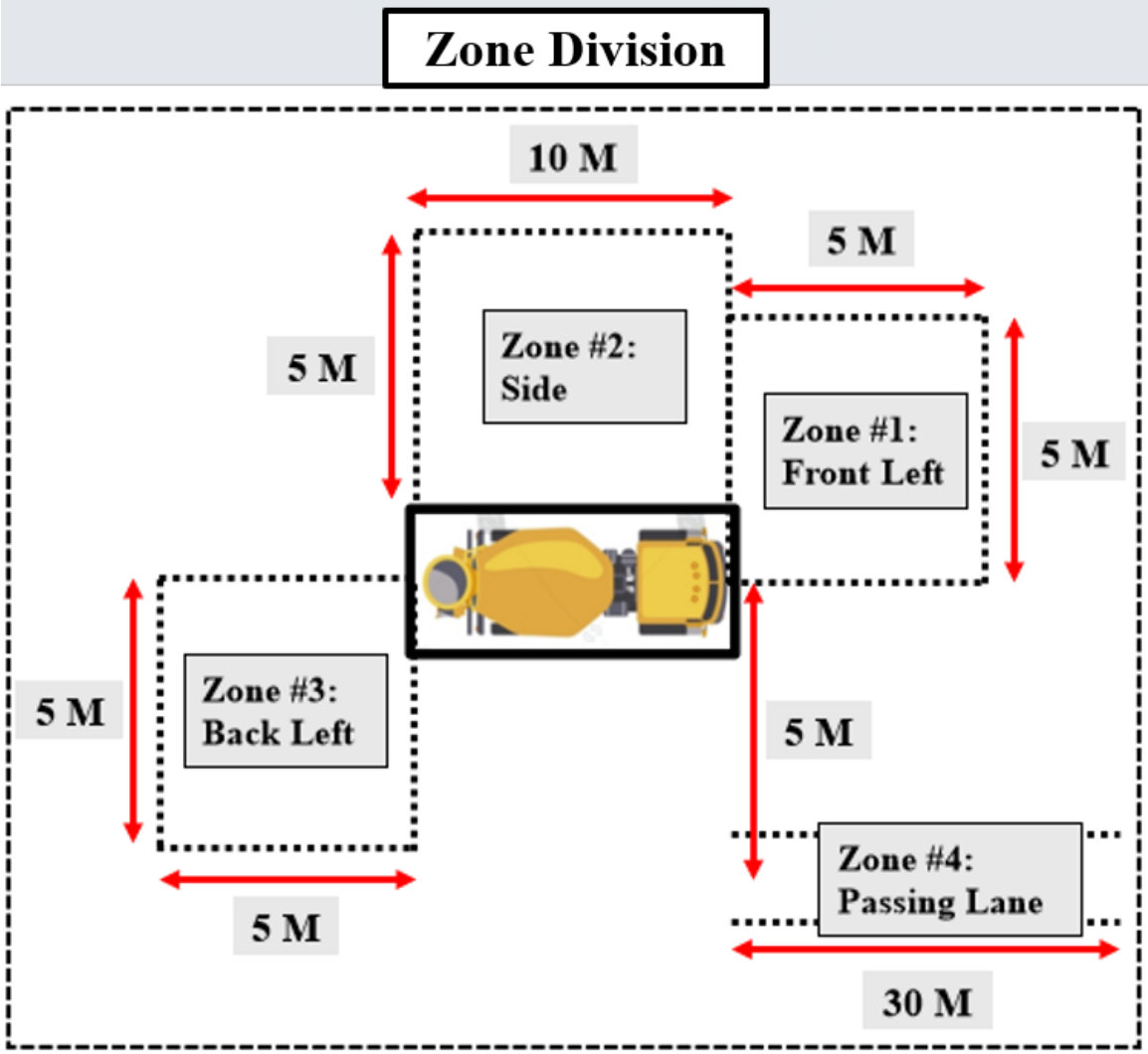

Figure 5.5: *Isolated environment and zone division around the target equipment.*

Given that the proposed sensor placement is symmetrical for the left and right parts of the equipment, we decided to divide the examined unsafe zones observed through this symmetric placement (H1 and H3) into equal leftward and rightward sub-zones. This approach simplifies the sensor placement and experiment execution process by assigning sensors specifically responsible for each sub-zone. Consequently, this reduces the number of testing sensors and the hardware requirements for parallel pipeline execution, as the testing scenarios were completed using only an *Alienware M16* laptop. Additionally, without dividing the side unsafe zone (H2), we considered it as a separate sub-zone and performed the experiments accordingly. Furthermore, since the

proposed pipeline for vehicle localization solely relies on data gathered from the VLP-16 sensors, we considered the unsafe zone associated with motor vehicles (H4) as an individual sub-zone requiring specific experimentation. Therefore, we divided the unsafe zone related to the interaction between motor vehicles and personnel (H4) into two sections: from the mixer truck forward and from the mixer truck backward. This division allowed us to perform the corresponding experiment using a single VLP-16 sensor mounted on the truck's cabin roof for just the forwarding sub-zone).

The aforementioned divisions split the overall system design into four discrete segments that needed to be validated through a series of experimental scenarios without affecting each other's results. Figure 5.6 illustrates the zone division as well as the coordination and orientation of the launching sensors for each sub-zone examination. Additionally, we established a local origin axis for each sub-zone sensor launch, converting the reported locations to the corresponding local origin and subsequently converting the results for each sub-zone to a global origin (located on the equipment's front bumper) for visualization purposes in the results chapter.

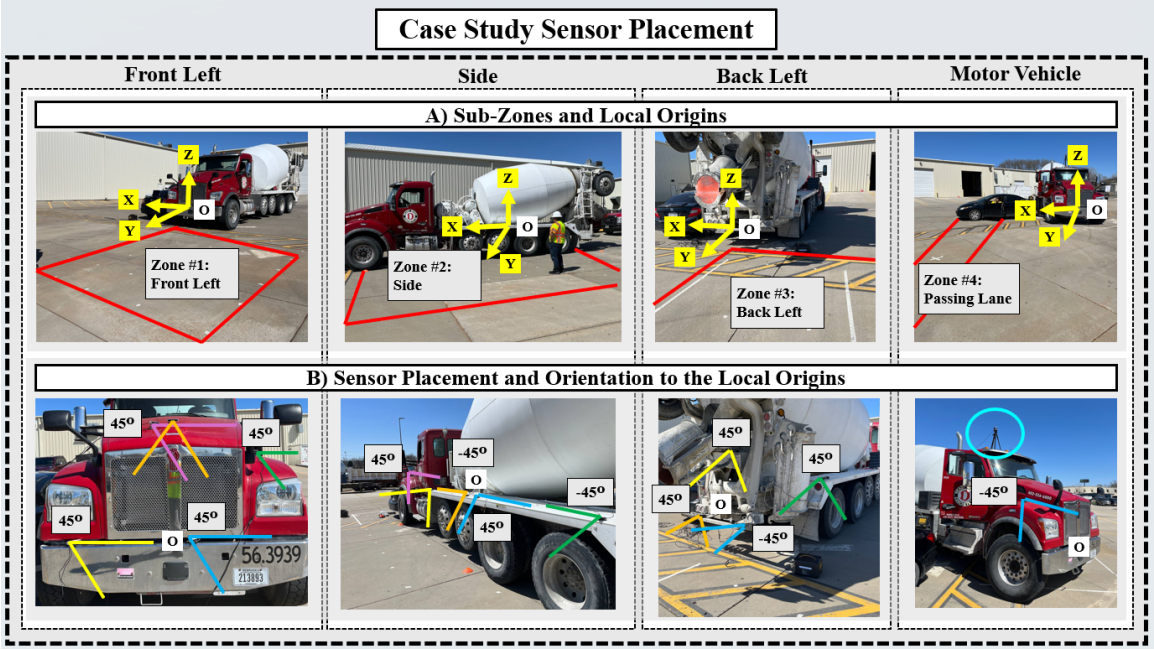

Figure 5.6: *A) Zone division and the corresponding local origin. B) sensor locations and coordination that are color-coded relative to the local origin (rotation degrees are based on the z-axis).*

The experiments performed for each sub-zone include three primary tests. First, we tested the partial placement error by comparing the reported locations from the launching pipeline with the marked ground truth within each sub-zone, measuring the error in reporting distances from the

system around the equipment. Second, we evaluated the partial placement's ability to track objects by executing a set of pre-determined trajectories within the validating sub-zone, from 6 meters to the closest possible location to the equipment on parallel lines. The reported trajectories were then compared with the performed trajectories' ground truth to measure the error. Third, we assessed the partial placement's ability to detect a series of entries and exits in a marked unsafe zone by performing multiple trajectories around the edges of the marked unsafe zone, executing a set of pre-determined exits and re-entries to and from the marked area, and evaluating the system's accuracy in detecting these movements.

Although we only tested the placement's error for motor-vehicle localization in terms of ground truth accuracy, we validated the close localization system by performing all three experiments for each determined sub-zone individually. The detailed procedures and results of these validations are demonstrated and described in the results chapter.

CHAPTER 6

Results

To validate the proposed system, we divided the monitoring zones around the equipment into four separate sub-zones: front-left, side, back-left, and motor-vehicle zones. This division enabled us to distinguish the sensors operating within the proposed pipelines both in parallel and as a unified framework, creating four individual sub-frameworks to be tested within their respective sub-zones. Additionally, We identified three major experiments for each sub-zone, which were performed and recorded using the corresponding sensors (sub-frameworks). Instead of recording raw data (images and point cloud data) from the running sensors for each scenario, we implemented the proposed pipeline during our tests and recorded only the essential data needed to validate the framework for each scenario, thereby avoiding the creation of large, cumbersome files. For each RealSense sensor within each sub-framework, we recorded the X and Y coordinates relative to the local origin (demonstrated in Figure 5.6) at the same timestamps as the other RGB-D sensors, along with the number of points in the resulting point cloud data (after performing the proposed point cloud processes for human localization) and the sensor number, which was uniquely assigned to each sensor in each sub-framework (as illustrated in Figure 5.6). Additionally, we recorded the X and Y values generated by the SFA3D algorithm, which was fed with the VLP-16's point cloud data relative to the testing local origin. For safety monitoring validation, we recorded a positive or negative pulse by comparing the final calculated location for each sub-framework to manually defined unsafe zones within each sub-zone, as described in detail in the last section.

Although the records for each scenario in each sub-zone were captured separately, we established the global origin as proposed in section 4.3, located on the front bumper which is the local origin

for the front-left sub-zone as well. Subsequently, we transformed the X and Y values of each record to this global origin to visualize the results of each designed experiment as a single outcome. This approach ensured the absence of errors, as the divided sub-frameworks for each sub-zone do not overlap in real-time monitoring—primarily point cloud data—due to their specific mounting coordination and orientation.

6.1 Location Report Accuracy

The initial experiment for each sub-zone involved drawing a 1-meter by 1-meter grid within each sub-zone and capturing the corresponding sub-framework outcomes to compare the reported locations with the predetermined ground truths. By conducting this test in the sub-zones around the equipment (zones 1 to 3), we generated a distance-based error map and calculated the average error in terms of distance around the equipment. For the last zone (zone 4), we marked the ground at intervals from 5 meters to 30 meters away from the equipment's bumper. Lastly, it is important to note that the distance-based error measurement in this section was derived by calculating the square root of the differences between the reported X and Y values and the X and Y values of the ground truth.

6.1.1 Human Localization

Although each sub-framework has no overlap in terms of recording data with other distinguished sub-frameworks, there is a record overlap within each sub-framework itself. This means that within each sub-zone, a single RealSense sensor observes a human's presence in its image and depth data at least, resulting in multiple location reports of the same person standing or walking in a unique location by different sensors observing the human's presence. Therefore, determining the most reliable reported location within a timestamp from the reports of several RealSense sensors requires careful consideration (Figure 6.1). To address this, we established two key assumptions for the framework. First, locations reported from different sensors are considered to represent a single person if the distances between them are 50 cm or less, based on an average error of 25 cm in individual RealSense reports (Table 5.1). Second, observations with fewer than 200 points in their point cloud data are not considered human locations. This threshold was determined by analyzing point cloud data from the human localization pipeline at various distances and angles in front of a single RGB-D

sensor, tested by both RealSense and Astra Camera. By applying these assumptions to the captured data from the testing experiments, we were able to analyze the framework's accuracy in terms of distance error, leading to findings used for subsequent validation tests.

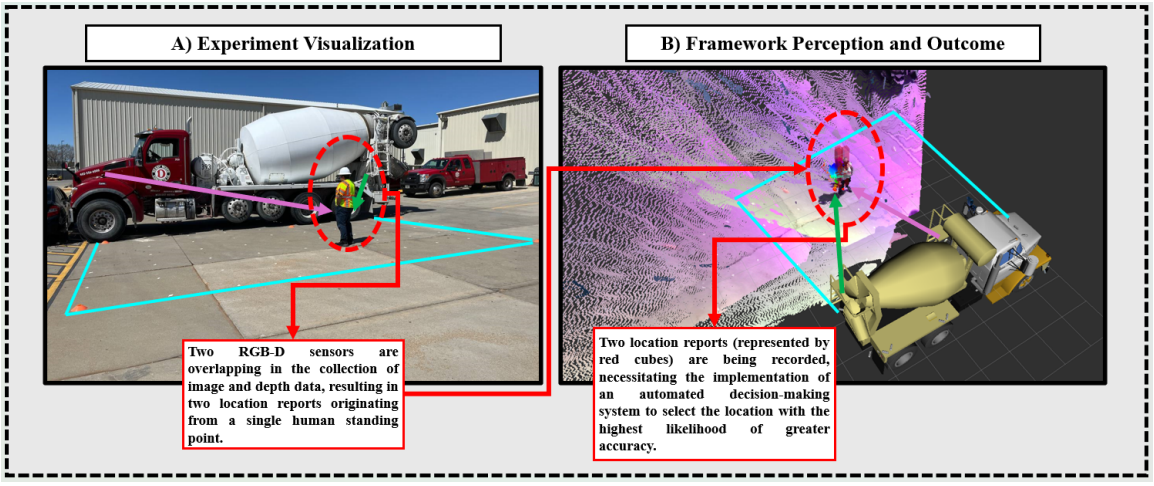

Figure 6.1: *A) Overview of the framework's human localization accuracy. B) Visualization of different location reports of the same human by different RealSense sensors.*

Figure 6.2 A visualizes an error-based color map within the zones requiring human localization, assuming that for each standing location, the report from the RealSense sensor with the least amount of error compared to the ground truth has been selected. This results in an average error of 27.1 cm. Although this error value is derived from manually selecting the best reporting location from all overlapping RealSense sensors, it indicates that the proposed framework can achieve a distance reporting error of 27.1 cm at its best performance. However, this does not qualify as an automated monitoring system since the reports were manually selected. Therefore, After removing reports with fewer than 200 points within their depth data, we tested two approaches for determining the final report for a specific location. The first approach involved averaging the reported locations within the same timestamp and within a range of 50 cm from each other (Figure 6.2 B). The second approach involved selecting the reported location with the maximum number of points within its point cloud data compared to other reported locations within the same timestamp (Figure 6.2 C). Among these two potential methods, averaging the reports resulted in an error of 34.1 cm, while selecting the report with the maximum number of points resulted in an average error of 32.8 cm. Although the error difference between these two methods is not significant, we decided to adopt the latter approach—selecting the report with the most points—as the third assumption of this study.

This decision was based on the rationale that a higher number of points within the reporting data of the human localization pipeline is more reliable. Temporary obstacles at closer distances can significantly reduce the accuracy of the human localization pipeline and decrease the number of points within the trimmed point cloud data. Consequently, we moved forward with this approach for the tracking and safety monitoring validation sections.

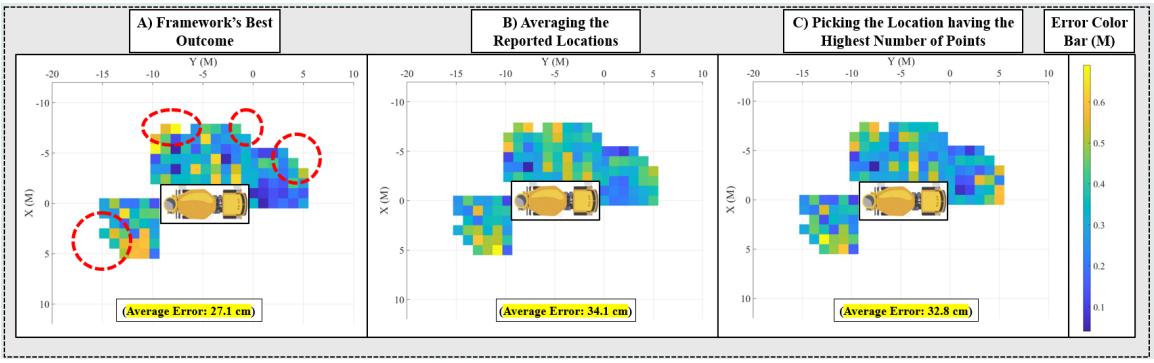

Figure 6.2: *Color-mapped plot of the location report's error: A) Least error pick. B) Averaging the coordination of reported locations. C) Picking the location having the highest number of points in its point cloud data.*

In conclusion, considering the aforementioned assumptions, the proposed system is capable of localizing workers within the areas surrounding the equipment with an average distance error of 32.8 cm. Additionally, the localization boundaries of this framework are limited in terms of the maximum distance at which it can localize workers. Specifically, it is capable of accurately localizing workers at distances of 5 meters or more from the equipment, as demonstrated by the red-dotted circles in Figure 6.2 A. This limitation is primarily due to the RealSense sensor's point cloud data capturing range, which is restricted to 6 meters, as discussed in Section 5.3.

6.1.2 Motor Vehicle Localization

In addition to the close localization framework for monitoring human-interacted unsafe zones, the motor vehicle localization pipeline's distance-based accuracy was also examined. To achieve this, we simulated a scenario where a motor vehicle approached the equipment within the passing lane, specifically in sub-zone 4 at a distance of 5 meters from the equipment (Figure 5.5). We compared the reported location of the pipeline to the pre-marked ground truth at every 5-meter interval from 30 meters to the equipment (Figure 6.3A). As a result, the motor vehicle localization pipeline

determined the location of the test motor vehicle with an average distance error of 1.05 meters (Figure 6.3 C). However, within the marked distances closer than 10 meters, the pipeline failed to localize any motor vehicle around the equipment. This failure was primarily due to the lack of depth data associated with the motor vehicle within the VLP-16's point cloud data. This issue arose from the LiDAR sensor's mounting height and its vertical field of view (FOV), which prevented the sensor from capturing point cloud data near ground level at distances closer than 9.87 meters—the last location where the motor vehicle pipeline successfully localized the vehicle (Figure 6.3 B). Consequently, we concluded that the LiDAR sensor's mounting height affects the closest distance from which the sensor can capture point cloud data, creating a dynamic blind spot for motor vehicle localization at close distances.

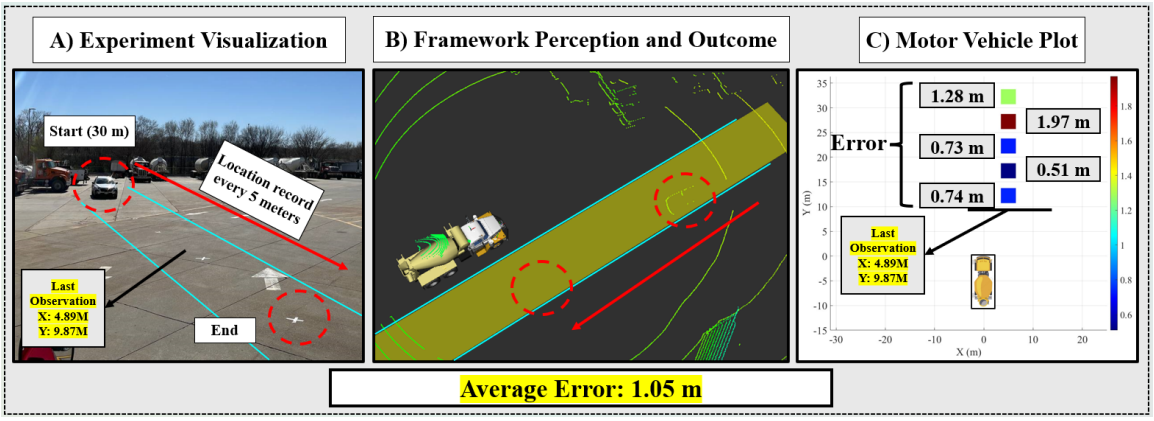

Figure 6.3: *A) Motor Vehicle validation test visualization. B) Pipeline's perception and outcome. C) color-mapped distance error of the recorded locations.*

Although the framework was capable of localizing motor vehicles approaching or leaving the roadway work zone from almost 10 meters to a distance of 30 meters in the direction of the work zone, which is sufficient to estimate the vehicle's approach time and monitor safety criteria, we also integrated the human localization pipeline with the motor vehicle pipeline. Instead of forcing the *YOLO* algorithm in the first step of human localization to further process images of cars, we allowed the human localization pipeline to follow the same procedure. This involved capturing, trimming, and averaging the coordinates of the extracted cluster from the RGB-D sensor's point cloud data to localize motor vehicles as well. This integration enables the system to efficiently monitor both human and vehicular movements, enhancing overall safety in the work zone. Therefore, despite the increased error due to differences in shape and size when processing motor vehicles,

running the RealSense sensor placed on the front right corner of the equipment (the blue-colored sensor in Figure 5.6 B) within the proposed framework alongside the VLP-16 sensor enabled the pipelines to localize the testing motor vehicles from 30 meters to 10 meters and from 5 meters to the equipment. By moving the motor vehicle from the 5-meter distance to the endpoint in 50 cm by 50 cm increments and recording the outcome (Figure 6.4 A), we achieved a distance-based accuracy of 50.1 cm (Figure 6.4 B) within the 5-meter range from the mixer truck. However, due to the length and shape of the target object, the error increased at closer distances. This increase occurred because the averaged value of the trimmed points from the depth data was captured from only the two visible sides of the motor vehicle, causing the values to be skewed toward the sensor rather than representing the exact locations.

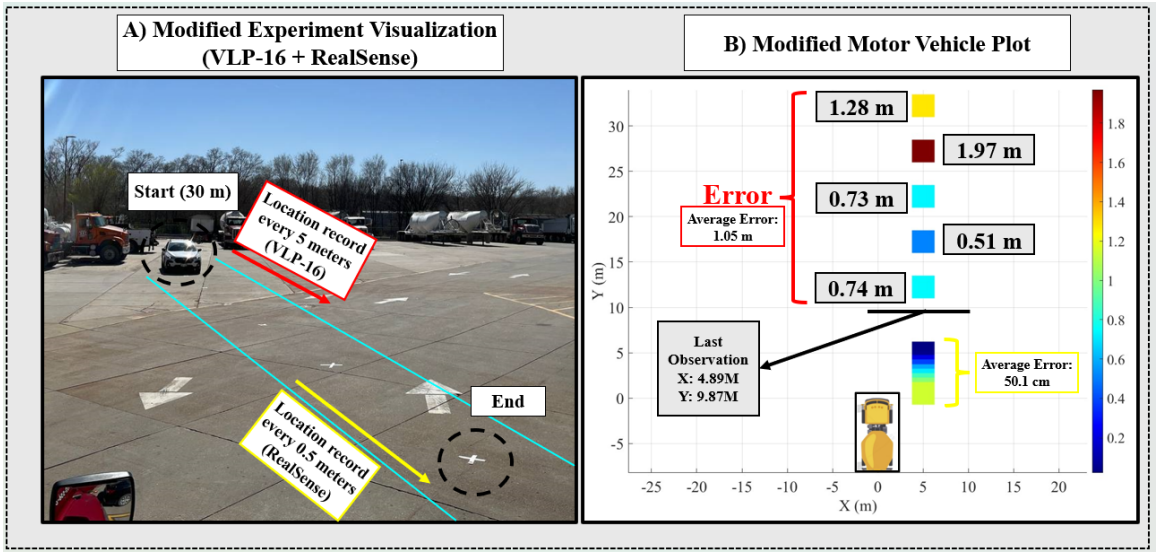

Figure 6.4: *A) Modified experiment visualization by adding the close localization pipeline parallel to the motor vehicle localization pipeline. B) color-mapped distance error of the recorded locations.*

To conclude, the modified system, which combines the close localization algorithm with the motor vehicle localization pipeline, enabled the framework to localize an approaching motor vehicle from a distance of 30 meters to approximately 10 meters with an average distance-based error of 1.05 meters. Additionally, from a distance of 5 meters (limited by the depth capturing capability of the RealSense sensor in the close localization pipeline) to the target equipment, the system achieved an average distance-based error of 50.1 cm.

6.2 Tracking Accuracy

Tracking moving objects, particularly the workforce, within unsafe zones is crucial as well due to the dynamic nature of the environment. This tracking not only localizes objects around the equipment but also predicts unforeseen situations and analyzes them from a safety perspective. To achieve this, we designed a second validation scenario to compare the framework's records in terms of auto-capturing a human's location continuously and mapping their movement around the equipment while following a predetermined movement pattern. These predetermined patterns were marked as straight lines starting from 5 meters away from the equipment and ending upon reaching it. By following the same trajectory at intervals of 1 meter within each sub-zone, we collected observations from each RealSense sensor responsible for the corresponding unsafe zone (Figure 6.5 A). Considering the timestamps of each observation for the trajectories, we applied the assumption of removing locations calculated with fewer than 200 points within their depth data and selecting the location associated with the RealSense sensor that collected the most points within its finalized and trimmed point cloud data (maximum number of point cloud assumption). We then extracted the final trajectory record from the recording data and compared the accuracy of each trajectory by averaging the distance error of the selected location for each timestamp to the actual trajectory line. Each performed trajectory was assigned a number to illustrate its error (Figure 6.5 B).

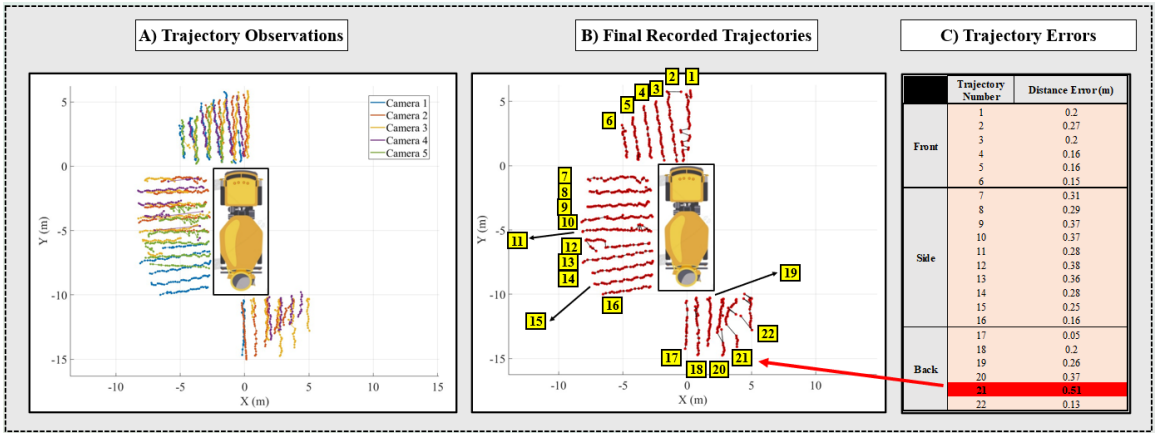

Figure 6.5: *A) Recorded Trajectories. B) Assumption implemented trajectories. C) Trajectory errors.*

Additionally, figure 6.5 C demonstrates the calculated error of each performed trajectory. While the error value for most trajectories remained under 50 cm, one trajectory (number 21) exhibited higher error values (51 cm) and was not properly observed within the framework. This issue is

visually apparent through the jumps observed in this trajectory in Figure 6.5 B, representing errors in observing a solid, directional, and continuous movement pattern. We concluded that such errors are due to the need for further RGB-D sensor calibration before the experiment, specifically in terms of orientation. Misalignment of sensor coordination relative to the local and global origin can result in captured trajectories that do not overlap with the intended trajectory line (or the trajectory ground truth). The collected observations (Figure 6.5 B) indicate that, within each trajectory, observations from different sensors are recorded as semi-diagonal lines. These lines are centered at closer distances and act as overlapping lines on the trajectory line. Thus, the placement of sensors in each sub-framework had an angular error relative to the local origin, resulting in diagonal lines and finalized lines that incorrectly depict the performed trajectories with random jumps. This occurs because the assumption of selecting the observation with the most point clouds causes the system to choose a different observation from a different sensor at each timestamp, creating significant gaps and jumps in the final trajectories, especially when only two sensors are observing that trajectory (such as in the aforementioned trajectories).

Therefore, we can conclude that the proposed system is capable of tracking human movements around the equipment with an average error of 28 cm. This implies that each observed trajectory should have a 28 cm offset to its side (left and right) if it is to be used for measuring interactions for safety monitoring criteria. However, this error can be significantly reduced by enhancing the accuracy of sensor calibration, specifically by further verifying their relative location and alignment to the predetermined origins.

6.3 Safety Monitoring Validation

As the final set of experiments, the framework's sensitivity should be evaluated in recognizing the entrance and exit of several pre-defined unsafe zones around the equipment. To date, the framework has been assessed in terms of location report accuracy for both humans and motor vehicles and tracking accuracy for humans within the sensors' boundaries. These boundaries extend up to 5 meters from the equipment due to RealSense's depth data quality and up to 30 meters due to the limitations of the SFA3D algorithm. Although the sub-zones marked around the equipment cover a monitoring area exceeding the requirements set by ITCP rules, it is crucial to measure the framework's accuracy

in detecting the workforce's entry into and exit from these unsafe zones. The size of these zones may vary depending on the work zone's situation. For our final validation test, we defined unsafe zones based on predetermined hazardous situations caused by workforce movements within these zones. These zones include the area surrounding the mixer truck, extending up to 3 meters from it on each side (Figure 6.6 A). To validate the framework's sensitivity, we conducted three sets of trajectories for each unsafe zone. This involved performing movement patterns along the boundary of the unsafe zone and slightly within it while making exits and re-entries perpendicular to the boundary line at fixed locations (represented by colored dots in Figure 6.6). During these tests, we recorded the reported locations and the duration of movements (as the time window). By knowing the starting and ending timestamps of each exit and re-entry performed in the tests, we measured the duration of these movements within the performed trajectories for each set of tests (three time windows for each sub-zone and each test). By calculating the number of location reports indicating a positive value (being outside) among all reported locations, we measured the framework's accuracy (in percentage) in detecting the exit and entrance of workforces within our predefined unsafe zones (Table 6.1).

Additionally, figure 6.7 A illustrates the combined recorded trajectories of all RealSense sensors in each experiment set (each set of colored exits and re-entries in Figure 6.6) within the unsafe zones. This figure assumes the removal of records with fewer than 200 data points. Additionally, Figure 6.7 B shows the results of implementing the second assumption of selecting the location with the highest number of points within each timestamp. In this figure, locations outside the unsafe zones are represented by green dots, while those within the unsafe zones are represented by red dots.

According to Table 6.1, the framework was able to recognize entrances and exits in 16 out of 24 scenarios with an accuracy of over 90%. For the remaining scenarios, it achieved an accuracy of over 80%, except for one scenario where the accuracy was 36% (exit and re-enter number 4 within the second test in Figure 6.6 B). This lower accuracy is mainly due to the side sub-framework's monitoring environment, which should not exceed the front-left corner of the equipment (the sub-framework associated with the front-left sub-zone is responsible for that particular area). Since the timestamps associated with each exit and re-enter are different, it is important to reduce the weight of percentage values that have more timestamps (duration) compared to those with less duration. We distinguished the sensitivity percentage for each unsafe zone and re-calculated the framework's sensitivity percentage for each sub-zone (by combining the positive/negative values of all tests within

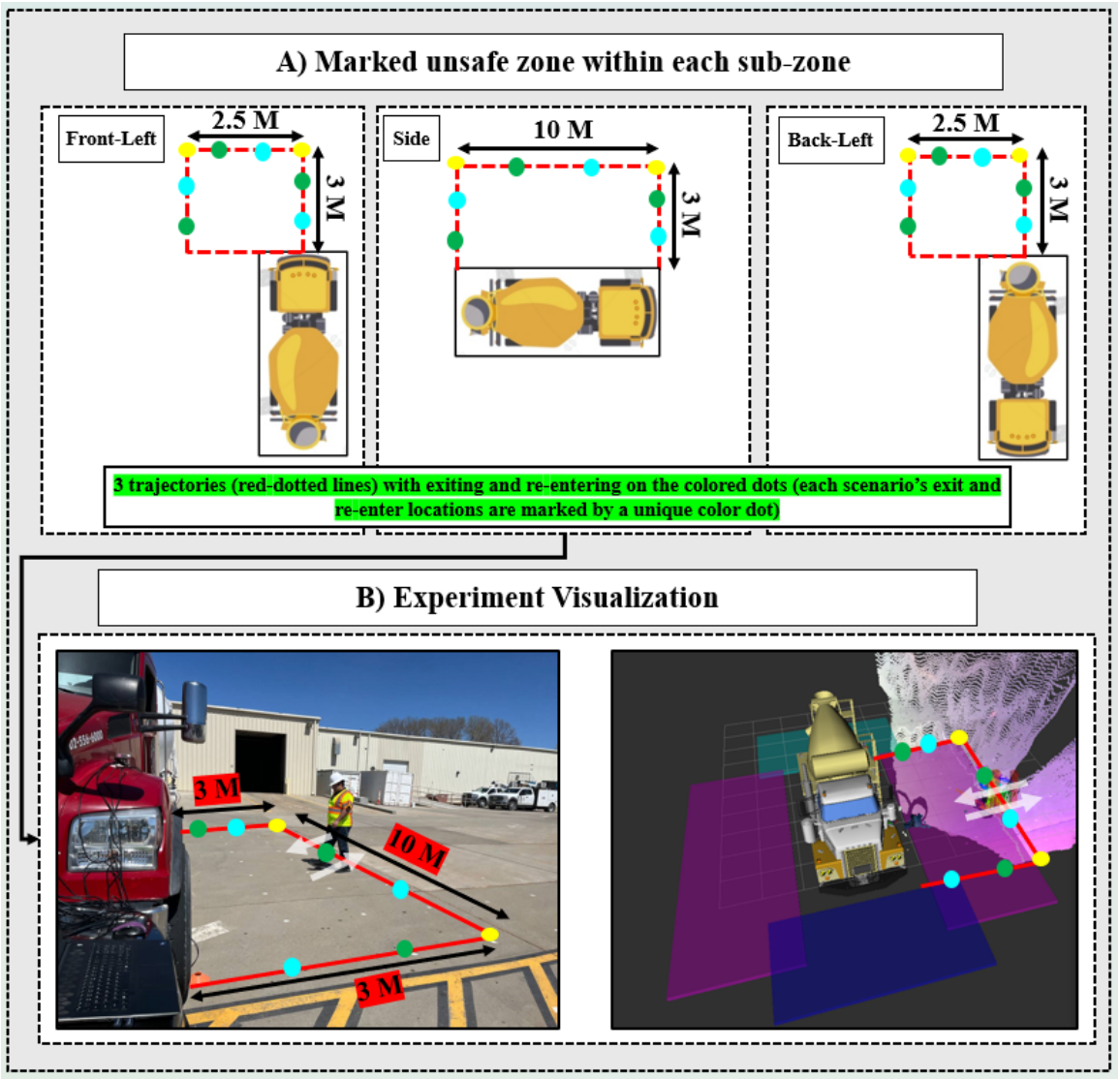

Figure 6.6: *Overview of the safety test scenario (B) by visualizing the unsafe zone's dimensions and the trajectory details.*

each sub-zone separately) using the Inversed Weighted Averaging method. As a result, we achieved a sensitivity percentage of 91.73% for the front-left unsafe zone, 82.86% for the side unsafe zone, and 90.18% for the back-left unsafe zone. This means that for each unsafe zone, the framework can detect the entrance or exit of the workforce with the corresponding percentage accuracy. It is worth mentioning that the sensitivity accuracies are likely to improve by launching the system as a single framework without dividing it into sub-frameworks, which could resolve the low sensitivity accuracy in the side scenario. Additionally, the issue of the RealSense sensor's orientational calibration is evident in Figure 6.7 A, as the recorded trajectories of the sensors do not overlap to represent a

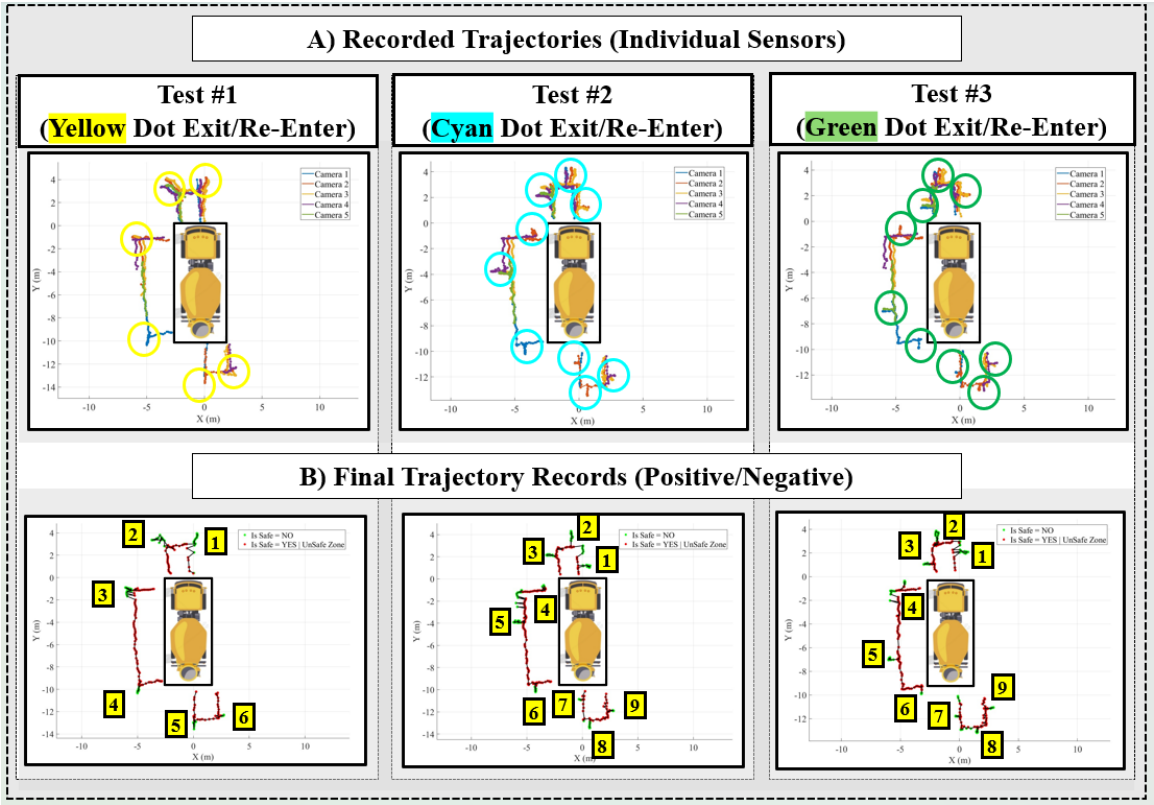

Figure 6.7: *Results of plotting the recorded trajectories (A) and the assumption implementations (B).*

unique trajectory. This misalignment caused several incorrect recognitions of being inside or outside the unsafe zones, especially within the front-left and side unsafe zones. There are extra green-colored reports that should have been reported within the unsafe zone rather than outside of it.

Table 6.1: *Success percentage of the trajectories (exit/re-enters).*

Scenario	Place	Out/In No.	Starting Timestamp	Ending Timestamp	Observation Period (Timestamps)	Wrong Responses	Percentage of Right Responses
Test #1	Front Left	1	65	109	44	2	95%
	Front Left	2	130	159	29	1	97%
	Side	3	16	64	48	0	100%
	Front Left	4	146	166	20	2	90%
	Back Left	5	20	35	15	2	87%
	Back Left	6	52	65	13	0	100%
Test #2	Front Left	1	17	45	28	0	100%
	Front Left	2	64	88	24	4	83%
	Side	3	111	129	18	1	94%
	Side	4	14	28	14	9	36%
	Side	5	108	144	36	0	100%
	Side	6	209	228	19	1	95%
	Back Left	7	7	25	18	0	100%
	Back Left	8	49	88	39	2	75%
	Back Left	9	76	88	12	2	83%
Test #3	Front Left	1	12	43	31	1	97%
	Front Left	2	67	90	23	1	96%
	Front Left	3	115	134	19	2	89%
	Side	4	126	142	16	2	86%
	Side	5	184	198	14	3	79%
	Side	6	142	158	16	1	94%
	Back Left	7	11	42	31	1	97%
	Back Left	8	48	58	10	1	90%
	Back Left	9	90	101	11	1	91%

In conclusion, the framework's sensitivity in detecting the entrance and exit of workforces to

and from a rectangular unsafe zone (the most common type of unsafe zone) within the surrounding environment of the target equipment is 88.26%. This indicates that the framework can detect instant entries or exits with less than 12% uncertainty. This error can be further reduced by improving sensor calibration and testing the framework as a single unified arrangement of sensors.

CHAPTER 7

Thesis Conclusion and Future Directions

To summarize, this study aimed to propose a novel safety monitoring system utilizing the current emerging autonomous vehicle (AV) technologies being industrialized these days. This system is designed to detect unforeseen unsafe situations by aiming to localize objects within various unsafe zones utilizing visual-based sensors. Furthermore, it seeks to bridge the gap between controversial safety observation techniques and pioneering ideas of how to autonomously perceive and react to sudden changes in different fields of research studies. Additionally, the capability to reduce inaccuracies, as determined and described in the results chapter, through more sophisticated localization pipelines can be considered the second reason for the efficiency of implementing such automated safety monitoring systems in heavy construction equipment.

Regardless of the accuracy and response speed of the utilized pre-trained detection models (CNNs) within the proposed pipelines of this study, the analysis of vision-based sensor data (image and point cloud) can be approached in different and novel directions. These approaches could result in outcomes as accurate as, or even more accurate than, those of this study, with varying computational speeds. This is primarily due to the variety of architectures and approaches developed for designing CNN models, as well as their training procedures, which result in detection models with different response times and accuracies (tested for each model on its training dataset). Therefore, the lack of accuracy in selecting models not being able to fully distinguish the workers or motor vehicle objects within the images can be considered as the first limitation of this study. The utilization of new versions of YOLO is expected to be one of the main enhancements in localization pipelines, as the identification and segmentation of workers can be performed with greater accuracy compared

to older versions. This improvement will lead to more precise trimming of the point cloud data corresponding to the detected workers. The primary objective of this enhancement is to reduce inaccuracies in capturing and trimming point cloud data that are incorrectly assigned to workers and thus improperly trimmed. Additionally, the analysis of point cloud data is another critical area that requires attention. Compared to image data, point cloud data has a more complex structure, necessitating more investigation and innovation in its analysis. This can be achieved through testing more efficient and accurate algorithms or developing new CNN models specifically for point cloud data, such as the SFA3D, to improve processing. This also represents the second limitation of this study, aimed at enhancing or redesigning the proposed pipelines in terms of effectiveness.

In addition to the accuracy of the localization pipelines presented in this study, the placement and selection of sensors on the equipment to ensure high-quality environmental data were constrained by the limited availability of RGB-D and LiDAR sensor models. As commercial products in AV technologies continuously optimize sensor placement and models to enhance the quality of surrounding data, which results in more accurate outcomes, the selection of the most efficient sensors and their optimal placement becomes crucial. This optimization should consider blind spot prevention, cost efficiency, and time efficiency of the implemented system. Therefore, the selection and placement of sensors are essential areas for follow-up research to complement the findings of this study.

Moreover, as noted in the results chapter, the accuracy of sensor placement (sensor calibration) significantly impacts the reported locations of detected objects. Errors in the locational and orientational placement of each sensor can result in consistent local location reports but varied global location reports. This discrepancy reduces the accuracy of the system's synchronization of captured locations from parallel pipelines. Issues such as treating a single person as multiple individuals, recording the trajectories of people walking around the equipment in patterns diagonal to their actual paths, or producing wavy scattered lines offset from the actual paths are consequences of sensor calibration errors. These issues can be effectively addressed by dedicating sufficient time to achieve millimeter-level accuracy in calibration.

Furthermore, it is worth mentioning that in this study, the utilization of vision-based sensors was limited to localization purposes. While we received multi-dimensional data from the surrounding environment of the equipment, it was only used to determine the location of target objects. However,

this perception data can be further exploited to generate a virtual 3D model of the environment. This model could be utilized not only by the AV systems mounted on the equipment (acting as a robotic perception system for autonomous task execution) but also to enhance real-time situational awareness for the workforce using AR devices. Such visualization could mitigate the limitations of human perception—such as field of view constraints, blind spots, and hearing impairments in noisy environments—thereby reducing the likelihood of exposure to hazardous situations. This potential application represents another important future direction for this study, as many researchers have recognized the positive impact of employing VR/AR devices to improve workforce understanding of work zones. Moreover, the technical implementations proposed in this study aim to provide a proof of concept for the further integration of sensing technologies into the autonomous operation of heavy construction equipment. The outcomes of this study, or similar studies, can facilitate advancements by incorporating ITCP rules as fundamental assumptions for defining safety protocols for both workers and autonomous heavy construction equipment. These protocols are typically established for autonomous vehicles to focus on identifying and addressing high-risk zones, thereby ensuring the equipment can pause, shut down, or restart tasks with heightened attention to safe navigation within the work zone. By integrating these predefined rules with visualization technologies such as AR devices, it will be possible to illustrate safe and unsafe zones around heavy construction equipment. This approach aims to create an automated awareness system that notifies workers of impending dangers, thereby enhancing safety and preventing accidents.

Although this study aimed to address the limitations of existing safety monitoring regulations and systems, its scope was limited to hazardous scenarios occurring within roadway work zones, particularly those resulting from interactions between workforces and motor vehicles. It is important to note that further research is needed to implement such autonomous systems in equipment used in various types of work zones. This is due to the differing nature of exposures that cause hazardous situations in different work environments.

Bibliography

[1] United States Census Bureau, *Business and industry: Time series / trend charts*, https://www.census.gov/econ/currentdata/?programCode=VIP&startYear=2002&endYear=2024&categories[]=AXXXX&dataType=T&geoLevel=US&adjusted=1&notAdjusted=0&errorData=0, Accessed: May 21, 2024, 2024.

[2] American Road & Transportation Builders Association, *Continued growth in bridge and highway construction - artba*, https://www.artba.org/news/continued-growth-in-bridge-and-highway-construction/, Jan. 2024.

[3] American Road & Transportation Builders Association, *Another big year for transportation construction - artba*, https://www.artba.org/news/another-big-year-for-transportation-construction/, Feb. 2024.

[4] U.S. Bureau of Labor Statistics, *Employment, hours, and earnings from the current employment statistics survey (national)*, https://data.bls.gov/timeseries/CES2023700001, Accessed: May 21, 2024, 2024.

[5] Occupational Safety and Health Administration, *Work zone hazards workbook*, https://workzonesafety.org/publication/work-zone-hazards-workbook/, Accessed: May 21, 2024, 2008.

[6] National Institute for Occupational Safety and Health, *Highway work zone safety*, https://www.cdc.gov/niosh/topics/highwayworkzones/bad/default.html, Accessed: May 21, 2024, 2022.

[7] National Institute for Occupational Safety and Health and National Hearing Conservation Association, *Concluded national alliances*, https://www.osha.gov/alliances/niosh-nhca/niosh-nhca, Accessed: May 21, 2024, 2008.

[8] Federal Highway Administration and U.S. Department of Transportation, *Worker safety - fhwa work zone*, https://ops.fhwa.dot.gov/wz/workersafety/index.htm, Accessed: May 21, 2024, Jan. 2024.

[9] Federal Highway Administration, *Temporary traffic control plans*, https://mutcd.fhwa.dot.gov/htm/2003r1r2/part6/part6c.htm, Accessed: May 22, 2024, Nov. 2023.

[10] Federal Highway Administration, *Internal traffic control plans for work zones*, https://ops.fhwa.dot.gov/publications/fhwahop17046/fhwahop17046.pdf, Accessed: May 22, 2024, 2017.

[11] American Road & Transportation Association and Occupational Safety and Health Administration, *Internal traffic control planning (instructor guide)*, https://www.osha.gov/sites/default/files/2018-11/fy13_sh-24861-sh3_ITCPInstructorGuide.pdf, Accessed: May 22, 2024, 2013.

[12] J. L. Graham and R. Burch, "Internal traffic control plans and worker safety planning tool," *Transportation Research Record: Journal of the Transportation Research Board*, vol. 1948, no. 1, 2006. DOI: 10.1177/0361198106194800107.

[13] Occupational Safety and Health Administration, *Competent person - standards*, https://www.osha.gov/competent-person/standards, Accessed: Jul. 26, 2023, 2023.

[14] National Work Zone Safety Information Clearinghouse, *Work zone traffic crash trends and statistics*, https://workzonesafety.org/work-zone-data/work-zone-traffic-crash-trends-and-statistics/, Accessed: May 22, 2024, 2022.

[15] National Institute for Occupational Safety and Health and U.S. Bureau of Labor Statistics, *Highway work zone safety*, https://www.cdc.gov/niosh/topics/highwayworkzones/default.html, Accessed: May 22, 2024, Aug. 2022.

[16] National Work Zone Safety Information Clearinghouse, *Worker fatalities and injuries at road construction sites*, https://workzonesafety.org/work-zone-data/worker-fatalities-and-injuries-at-road-construction-sites/, Accessed: May 22, 2024, 2023.

[17] M. Gangolells, M. Casals, N. Forcada, X. Roca, and A. Fuertes, "Mitigating construction safety risks using prevention through design," *Journal of Safety Research*, vol. 41, no. 2, 2010. DOI: 10.1016/j.jsr.2009.10.007.

[18] J. Gambatese and J. Hinze, "Addressing construction worker safety in the design phase designing for construction worker safety," *Automation in Construction*, vol. 8, no. 6, 1999. DOI: 10.1016/S0926-5805(98)00109-5.

[19] M. R. Hallowell and J. A. Gambatese, "Construction safety risk mitigation," *Journal of Construction Engineering and Management*, vol. 135, no. 12, 2009. DOI: 10.1061/(asce)co.1943-7862.0000107.

[20] S. X. Zeng, C. M. Tam, and V. W. Y. Tam, "Integrating safety, environmental and quality risks for project management using a fmea method," *Engineering Economics*, no. 1, 2010.

[21] B. Esmaeili, M. R. Hallowell, and B. Rajagopalan, "Attribute-based safety risk assessment. i: Analysis at the fundamental level," *Journal of Construction Engineering and Management*, vol. 141, no. 8, 2015. DOI: 10.1061/(asce)co.1943-7862.0000980.

[22] J. Teizer, M. Venugopal, and A. Walia, "Ultrawideband for automated real-time three-dimensional location sensing for workforce, equipment, and material positioning and tracking," *Transportation Research Record*, no. 2081, 2008. DOI: 10.3141/2081-06.

[23] N. R. Esfahan, S. Du, C. Anumba, and S. Razavi, "Smart tracking of highway construction projects," in *Congress on Computing in Civil Engineering, Proceedings*, 2017. DOI: 10.1061/9780784480830.024.

[24] N. Roofigari-Esfahan, E. E. White, M. A. Mollenhauer, and J. P. T. Vilela, *Development of a connected smart vest for improved roadside work zone safety*, https://vtechworks.lib.vt.edu/bitstream/handle/10919/104207/Safe-D_Report_Smart-Vest_04-104.pdf?sequence=1&isAllowed=y, 2021.

[25] J. Park, E. Marks, Y. K. Cho, and W. Suryanto, "Performance test of wireless technologies for personnel and equipment proximity sensing in work zones," *Journal of Construction Engineering and Management*, vol. 142, no. 1, 2016. DOI: 10.1061/(asce)co.1943-7862.0001031.

[26] J. Park, K. Kim, and Y. K. Cho, "Framework of automated construction-safety monitoring using cloud-enabled bim and ble mobile tracking sensors," *Journal of Construction Engineering and Management*, vol. 143, no. 2, 2017. DOI: 10.1061/(asce)co.1943-7862.0001223.

[27] Y. Fang, Y. K. Cho, S. Zhang, and E. Perez, "Case study of bim and cloud–enabled real-time rfid indoor localization for construction management applications," *Journal of Construction Engineering and Management*, vol. 142, no. 7, 2016. DOI: 10.1061/(asce)co.1943-7862.0001125.

[28] I. Jeelani, K. Asadi, H. Ramshankar, K. Han, and A. Albert, "Real-time vision-based worker localization hazard detection for construction," *Automation in Construction*, vol. 121, 2021. DOI: 10.1016/j.autcon.2020.103448.

[29] S. Tang, D. Roberts, and M. Golparvar-Fard, "Human-object interaction recognition for automatic construction site safety inspection," *Automation in Construction*, vol. 120, 2020. DOI: 10.1016/j.autcon.2020.103356.

[30] C. Chen, Z. Zhu, and A. Hammad, "Automated excavators activity recognition and productivity analysis from construction site surveillance videos," *Automation in Construction*, vol. 110, 2020. DOI: 10.1016/j.autcon.2019.103045.

[31] V. Getuli, P. Capone, A. Bruttini, and S. Isaac, "Bim-based immersive virtual reality for construction workspace planning: A safety-oriented approach," *Automation in Construction*, vol. 114, 2020. DOI: 10.1016/j.autcon.2020.103160.

[32] J. Sawhney, A. Riley, and M. Irizarry, *Construction 4.0—An Innovation Platform for the Built Environment*. 2020.

[33] D. Kim, M. Liu, S. H. Lee, and V. R. Kamat, "Remote proximity monitoring between mobile construction resources using camera-mounted uavs," *Automation in Construction*, vol. 99, 2019. DOI: 10.1016/j.autcon.2018.12.014.

[34] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You only look once: Unified, real-time object detection," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2016. DOI: 10.1109/CVPR.2016.91.

[35] K. Kim, S. Kim, and D. Shchur, "A uas-based work zone safety monitoring system by integrating internal traffic control plan (itcp) and automated object detection in game engine environment," *Automation in Construction*, vol. 128, 2021. DOI: 10.1016/j.autcon.2021.103736.

[36] N. Jacob-Loyola, F. M.-L. Rivera, R. F. Herrera, and E. Atencio, "Unmanned aerial vehicles (uavs) for physical progress monitoring of construction," *Sensors*, vol. 21, no. 12, 2021. DOI: 10.3390/s21124227.

[37] H. Oyediran, A. Shirazi, M. Peavy, L. Merino, and K. Kim, "Human-aware safe robot control and monitoring system for operations in congested indoor construction environment," in *Construction Research Congress 2024*, 2024. DOI: 10.1061/9780784485262.082.

[38] A. Ibrahim, W. Torres-Calderon, and M. Golparvar-Fard, "Reinforcement learning for high-quality reality mapping of indoor construction using unmanned ground vehicles," *Automation in Construction*, vol. 156, 2023. DOI: 10.1016/j.autcon.2023.105110.

[39] Y. Liu, M. Habibnezhad, and H. Jebelli, "Brainwave-driven human-robot collaboration in construction," *Automation in Construction*, vol. 124, 2021. DOI: 10.1016/j.autcon.2021.103556.

[40] M. O. Sanni-Anibire, A. S. Mahmoud, M. A. Hassanain, and B. A. Salami, "A risk assessment approach for enhancing construction safety performance," *Safety Science*, vol. 121, 2020. DOI: 10.1016/j.ssci.2019.08.044.

[41] I. Okpala, C. Nnaji, and A. A. Karakhan, "Utilizing emerging technologies for construction safety risk mitigation," *Practice Periodical on Structural Design and Construction*, vol. 25, no. 2, 2020. DOI: 10.1061/(asce)sc.1943-5576.0000468.

[42] S. International, *Taxonomy and definitions for terms related to driving automation systems for on-road motor vehicles j3016*, 2018.

[43] R. Akhavian and A. H. Behzadan, "Construction equipment activity recognition for simulation input modeling using mobile sensors and machine learning classifiers," *Advanced Engineering Informatics*, vol. 29, no. 4, 2015. DOI: 10.1016/j.aei.2015.03.001.

[44] A. Ansaripour, M. Heydariaan, K. Kim, O. Gnawali, and H. Oyediran, "Viper+: Vehicle pose estimation using ultra-wideband radios for automated construction safety monitoring," *Applied Sciences*, vol. 13, no. 3, 2023. DOI: 10.3390/app13031581.

[45] M. Memarzadeh, M. Golparvar-Fard, and J. C. Niebles, "Automated 2d detection of construction equipment and workers from site video streams using histograms of oriented gradients and colors," *Automation in Construction*, vol. 32, 2013. DOI: 10.1016/j.autcon.2012.12.002.

[46] H. Kim, H. Kim, Y. W. Hong, and H. Byun, "Detecting construction equipment using a region-based fully convolutional network and transfer learning," *Journal of Computing in Civil Engineering*, vol. 32, no. 2, 2018. DOI: 10.1061/(asce)cp.1943-5487.0000731.

[47] J. Dai, Y. Li, K. He, and J. Sun, "R-fcn: Object detection via region-based fully convolutional networks," in *Advances in Neural Information Processing Systems*, 2016.

[48] W. Fang, L. Ding, B. Zhong, P. E. D. Love, and H. Luo, "Automated detection of workers and heavy equipment on construction sites: A convolutional neural network approach," *Advanced Engineering Informatics*, vol. 37, 2018. DOI: 10.1016/j.aei.2018.05.003.

[49] S. Ren, K. He, R. Girshick, and J. Sun, "Faster r-cnn: Towards real-time object detection with region proposal networks," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 39, no. 6, 2017. DOI: 10.1109/TPAMI.2016.2577031.

[50] K. He, G. Gkioxari, P. Dollár, and R. Girshick, "Mask r-cnn," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 42, no. 2, 2020. DOI: 10.1109/TPAMI.2018.2844175.

[51] H. Luo, M. Wang, P. K. Y. Wong, and J. C. P. Cheng, "Full body pose estimation of construction equipment using computer vision and deep learning techniques," *Automation in Construction*, vol. 110, 2020. DOI: 10.1016/j.autcon.2019.103016.

[52] J. Kim, S. Chi, and J. Kim, "3d pose estimation and localization of construction equipment from single camera images by virtual model integration," *Advanced Engineering Informatics*, vol. 57, 2023. DOI: 10.1016/j.aei.2023.102092.

[53] A. Assadzadeh, M. Arashpour, H. Li, R. Hosseini, F. Elghaish, and S. Baduge, "Excavator 3d pose estimation using deep learning and hybrid datasets," *Advanced Engineering Informatics*, vol. 55, 2023. DOI: 10.1016/j.aei.2023.101875.

[54] C. Wang and Y. K. Cho, "Smart scanning and near real-time 3d surface modeling of dynamic construction equipment from a point cloud," *Automation in Construction*, vol. 49, 2015. DOI: 10.1016/j.autcon.2014.06.003.

[55] J. Chen, Y. Fang, and Y. K. Cho, "Automated equipment recognition and classification from scattered point clouds for construction management applications," in *ISARC 2016 - 33rd International Symposium on Automation and Robotics in Construction*, 2016.

[56] R. Schnabel, R. Wahl, and R. Klein, "Efficient ransac for point-cloud shape detection," *Computer Graphics Forum*, vol. 26, no. 2, 2007. DOI: 10.1111/j.1467-8659.2007.01016.x.

[57] R. B. Rusu, "Semantic 3d object maps for everyday manipulation in human living environments," *KI - Kunstliche Intelligenz*, vol. 24, no. 4, 2010. DOI: 10.1007/s13218-010-0059-6.

[58] W. Wohlkinger and M. Vincze, "Ensemble of shape functions for 3d object classification," in *2011 IEEE International Conference on Robotics and Biomimetics, ROBIO 2011*, 2011. DOI: 10.1109/ROBIO.2011.6181760.

[59] N. M. Dung, *Super-fast-accurate-3d-object-detection-pytorch*, https://github.com/maudzung/Super-Fast-Accurate-3D-Object-Detection, Accessed: Aug. 17, 2023, 2020.

[60] A. Darwesh, D. Wu, M. Le, and S. Saripalli, "Building a smart work zone using roadside lidar," in *IEEE Conference on Intelligent Transportation Systems, Proceedings, ITSC*, 2021. DOI: 10.1109/ITSC48978.2021.9564527.

[61] N. Kayhani, H. Taghaddos, and S. BehzadiPour, "Construction equipment collision-free path planning using robotic approach," in *ISARC 2018 - 35th International Symposium on Automation and Robotics in Construction and International AEC/FM Hackathon: The Future of Building Things*, 2018. DOI: 10.22260/isarc2018/0169.

[62] R. Jonker and A. Volgenant, "A shortest augmenting path algorithm for dense and sparse linear assignment problems," *Computing*, vol. 38, no. 4, 1987. DOI: 10.1007/BF02278710.

[63] L. Zhang, A. Boularias, A. Dogar, K. Fragkiadaki, J. Borreguero, and A. Johnson, "An autonomous excavator system for material loading tasks," *Science Robotics*, vol. 6, no. 55, 2021. DOI: 10.1126/scirobotics.abc3164.

[64] K. You, L. Ding, C. Zhou, Q. Dou, X. Wang, and B. Hu, "5g-based earthwork monitoring system for an unmanned bulldozer," *Automation in Construction*, vol. 131, 2021. DOI: 10.1016/j.autcon.2021.103891.

[65] M. Bjelonic, *Yolo ros*, https://github.com/leggedrobotics/darknet_ros, Accessed: May 16, 2023, 2018.

[66] S. Khan, J. Guivant, and X. Li, "Design and experimental validation of a robust model predictive control for the optimal trajectory tracking of a small-scale autonomous bulldozer," *Robotics and Autonomous Systems*, vol. 147, 2022. DOI: 10.1016/j.robot.2021.103903.

[67] A. Shirazi and K. Kim, "Development of a rule-based safety checking system for autonomous heavy construction equipment," in *Construction Research Congress 2024*, 2024. DOI: 10.1061/9780784485262.098.

[68] Work Zone Safety Consortium, *Developing internal traffic control plans (itcps) for work zones*, https://www.workzonesafety.org/files/documents/training/courses_programs/rsa_program/RSP_Guidance_Documents_Download/RSP_ITCP_Guidance_Download.pdf, Accessed: Jul. 05, 2023.

[69] R. B. Rusu and S. Cousins, "3d is here: Point cloud library (pcl)," in *Proceedings - IEEE International Conference on Robotics and Automation*, 2011. DOI: 10.1109/ICRA.2011.5980567.

[70] N. Koenig and A. Howard, "Design and use paradigms for gazebo, an open-source multi-robot simulator," in *2004 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2004. DOI: 10.1109/iros.2004.1389727.

[71] T. Y. Lin, M. Maire, S. Belongie, *et al.*, "Microsoft coco: Common objects in context," in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 2014. DOI: 10.1007/978-3-319-10602-1_48.

[72] A. Geiger, P. Lenz, and R. Urtasun, "Are we ready for autonomous driving? the kitti vision benchmark suite," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2012. DOI: 10.1109/CVPR.2012.6248074.

[73] B. Li, "3d fully convolutional network for vehicle detection in point cloud," in *IEEE International Conference on Intelligent Robots and Systems*, 2017. DOI: 10.1109/IROS.2017.8205955.